INTERNATIONAL DOCTORAL
SCHOOL OF THE USC

Yago
Atrio Lema

PhD Thesis

Modelling tourists' decision-making from a cross-cultural perspective

Santiago de Compostela, 2024

Doctoral Programme in Economics and Business

DOCTORAL THESIS

MODELLING TOURISTS' DECISION-MAKING FROM A CROSS-CULTURAL PERSPECTIVE

Author

Yago Atrio Lema

Supervisor/s: Isabel Neira Gómez and Eduardo Sánchez Vila

Tutor: Isabel Neira Gómez

PHD PROGRAMME IN ECONOMICS AND BUSINESS

SANTIAGO DE COMPOSTELA

One of the big lessons from behavioural Economics is that we make decisions as a function of the environment we're in.

Dan Ariely

I can always choose, but I ought to know that if I do not choose, I am still choosing.

Jean Paul Sartre

AGRADECIMIENTOS

La elaboración de mi tesis doctoral se ha presentado como un proceso de aprendizaje constante, tanto a nivel personal como académico, me gustaría reconocer el esfuerzo de todas las personas que me han apoyado en esta travesía, tanto a nivel personal como académico, sin las cuales no podría haber finalizado esta tesis.

Me gustaría mostrar mi más sincera gratitud a mi directora de tesis, Isabel Neira, para quien no tengo palabras suficientes que describan su influencia en mi modo de entender la ciencia y el mundo. Todo lo que pueda decir no se compara a la realidad, podría decir que es una persona humanista, dado que cree en las personas, también podría decir que es una científica, dado que estudia los fenómenos sociales de manera crítica, pero no cientificista, dado que es humilde al conocer la complejidad de los fenómenos sociológicos, lo que sí que puedo decir es que es mi referente en cuanto a dignidad y generosidad intelectual, tratando de influir en la mejora de la calidad de vida de las personas desde las instituciones, y compartiendo sus conocimientos con todas las personas con las que interactúa.

A mi director de tesis, Eduardo Sánchez le agradezco el rigor y precisión científica aportados a la tesis, su interdisciplinariedad y su visión global del proceso investigador. El interés que me ha suscitado en el estudio de la toma de decisiones lo conservaré para siempre.

Quiero agradecer a aquellos que me han ayudado a comenzar mi carrera investigadora, al equipo de investigación GEM, en particular a Loreto Fernández y Ana Fernández por enseñarme a transferir la investigación a la sociedad, a Nuria Calvo y Ariadna Monje por enseñarme a llevar un proyecto e investigar en emprendimiento, al grupo de investigación Valfinap (Valoración Financiera Aplicada), por acogerme como a uno de sus miembros a todos los niveles, especial mención a Marta Portela por ayudarme en mi pasión por el análisis de datos y la programación; así como a los profesores Sara Fernández, Luis Otero, Celia López y Lucía Rey por el apoyo recibido, a las profesoras Marisa Chas y Ánxela Troitiño del Departamento de Economía Cuantitativa, por iniciarme en la práctica docente, a las profesoras Elvira Vieira, Ana Borges y Rosa Conde por tratarme como un miembro más de su equipo de investigación en la estancia en el Instituto Superior de Administração e Gestão, a Marisa del Río por ayudarme ante cualquier duda sobre turismo, a José Ramón por apoyarme en la estancia en University of

Greenwich y a Noelia y a Don Daniel Lorenzo por facilitarme el uso de la base de datos de la Oficina de Acogida al Peregrino, esencial para realizar la investigación.

Me gustaría agradecer a mis amigos de toda la vida, Juan y Alexis, que ya estabais antes de iniciar la tesis doctoral y habéis entendido los períodos de ausencia. A mis amigos conocidos durante el periodo de investigación doctoral Vanessa, Karen, Sugey, Guillermo, Napoleón, Freddy, Angélica, José Manuel, Jazmín y Jackyline por el apoyo moral sin el que probablemente no habría sido capaz de finalizar esta tesis.

Finalmente agradezco el apoyo incondicional de mi familia, en particular a mi padre Alfredo y a mi madre María, que son los mejores padres a los que un hijo podría aspirar, fuentes de cariño, generosidad, seguridad y enseñanzas inagotables, así como a mis abuelos, modelos de humildad y esfuerzo.

Table of contents

ABSTRACT

Decision-making models are widely used tools that economists employ to study how economic agents select their preferred features on conflictive alternatives. Despite the importance given to individual preferences in decision-making with the universalization of Random Utility Models, cognitive biases, emotions or psychological influences were not studied in the realm of economics till the 70s, with rise of behavioural economics, opening a new field of study. However, contextual factors like culture from the field of social-psychology were less studied in the area of economics. To explore the influence of culture on decision-making, we use the case of tourism, where the primary hypothesis is that culture affects decision-making. Thus, the aim of the thesis is to merge the knowledge from decision-making, behavioural economics and social-psychology in a tourism context to check if culture affect decision-making, and how.

We introduce several alternative methodologies to include culture as an explanatory factor of the decision-making process, studying both discrete choice models and machine learning models. To contrast our models, we employ Pilgrims Welcome Office database to assess how useful are our models at explaining and predicting decision-makers behaviour, proving that it is still possible to improve widely used decision-making models, just changing researcher's perspective and integrating insights from different fields of study.

RESUMO

A toma de decisións é un fenómeno que os seres humanos están a realizar de maneira continuada (que roupa vestir, que carreira universitaria elixir, a onde ir en vacacións etc.), motivo polo cal é estudiado por unha gran variedade (Cortez & Embrechts, 2013) de ciencias, en concreto, a toma de decisións por parte de axentes económicos racionais constitúe a base da base da microeconomía. Os Modelos de Utilidade Aleatoria desenvolvidos por McFadden (1973), revolucionaron o campo da investigación en microeconomía, ó poder representar como os axentes económicos deciden entre múltiples alternativas. A utilización deste novo tipo de modelos permitiu cuantificar a utilidade derivada de cada alternativa para cada axente, tendo en conta tamén os factores non observables (Mcfadden, 1980), transformando esta utilidade en probabilidades, e creando así os modelos discretos de toma de decisións.

A inclusión de factores psicolóxicos nos modelos de toma de decisións, concretamente os relacionados con nesgos cognitivos, emocións ou influencias, é un tema esencial na rama da economía coñecida como economía conductual, que complementa a teoría económica tradicional facéndoa máis realista. Con todo, o papel do contexto cultural, concretamente a tradición dentro do ámbito da socioloxía foi menos desenvolvida e constitúe unha area de análise de gran interese na nova economía. Unha rama da psicoloxía social, como a investigación transcultural, puxo de relevo a importancia de incluír variables que expliquen o contexto no que viven os individuos, para estudar mellor como os individuos de diferentes culturas amosan diferentes comportamentos en canto á toma de decisións. A motivación desta investigación é a de fusionar os coñecementos da rama da toma de decisións, da economía conductual e da psicoloxía social a través da utilización de metodoloxías á vangarda da investigación nestas área, aplicadas ao caso de uso turístico para contribuír á comprensión da toma de decisións neste ámbito, elemento clave á hora de construír recomendadores de produtos e servicios, cuxo uso en economía é recente e que presenta unha novidosa liña de investigación en turismo. Empregaremos metodoloxías da familia dos métodos econométricos empregados na economía e técnicas de analítica avanzada empregadas no ámbito de Big Data. O caso de uso para o que se desenvolven os modelos será o Camiño de Santiago, camiño composto por un conxunto de rutas que percorre boa parte de España, Portugal, e parte de Europa coa

finalidade de chegar a Santiago de Compostela, que presenta unha gran utilidade ao ser aplicables a calquera ámbito de toma de decisións económicas. Estas rutas alternativas teñen os nomes de: inglés, primitivo, do norte, francés, portugués de costa, portugués e a vía da prata. Estas alternativas permiten testar as nosas hipóteses de xeito que podemos analizar non so as características individuais das persoas que percorren os camiños, senón as das alternativas en si mesmas, dispoñendo dunha mostra de mais e 3 millóns de peregrinos procedentes dunha gran variedade de nacións e culturas, que chegaron a Santiago entre o ano 2003 e 2022. Con relación á análise descritiva da base de datos, pódese observar como o número de peregrinos que percorren o Camiño de Santiago, aumentou constantemente dende o ano 2003, salvo en 2020 e 2021, debido á pandemia da Covid-19. Ademais de poderse observar un patrón nos "Anos Santos", cando hai un incremento de 100,000 chegadas de peregrinos, como nos anos 2004 e 2010. Analizando os datos por continente, pódese observar que a maioría dos peregrinos que realizan o Camiño de Santiago son europeos, de Norte América e de Sur América.

Despois de realizar a revisión do estado da arte, seleccionouse o obxectivo principal da tese, analizar o rol da cultura na toma de decisións, para o que se desenvolven novas metodoloxías para predicir, clasificar e recomendar produtos e servizos, completando catro obxectivos específicos:

- Analizar o rol da cultura na toma de decisións usando a cultura supra-nacional como factor de empuxe no modelo de empuxe-arrastre turístico.
- Crear un novo modelo causal, no que a cultura inflúe sobre as preferencias polos atributos das alternativas dos decisores, afectando ó modo de resolver o dilema de compensación entre as alternativas conflitivas.
- Analizar os constructos e variables dos marcos conceptuais da análise transcultural en procesos de toma de decisións.
- Atopar a mellor metodoloxía para utilizar nun sistema de recomendación.

Para lograr os obxectivos, baseados na revisión do estado do arte e nas motivacións, exploramos detidamente as preguntas de investigación e hipóteses, onde a meta principal é a de abordar e resolver as brechas nos coñecementos previos.

En primeiro lugar localizamos unha brecha na literatura de turismo e toma de decisións, onde os avances presentados por Sirakaya & Woodside (2005) e Uysal et al. (2008), clasificaron aos motivos de empuxe do modelo de empuxe-arrastre coma variables internas relacionadas cos desexos e preferencias dos individuos, sabendo que a cultura é capaz de influír nas preferencias dos individuos, propoñemos unha serie de preguntas e hipóteses:

Pregunta 1 (P1): Cal é o efecto de incluír a cultura supra-nacional no proceso de toma de decisións turísticas? Hipótese 1 (H1): A cultura supra-nacional condiciona o proceso de toma de decisións (capítulo 4 e 5).

En segundo lugar localizamos unha brecha no uso da cultura supra-nacional na toma de decisións no ámbito do turismo. As medidas culturais empregadas na actualidade en turismo, non incluían á cultura supra-nacional como unha medida cultural habitual, sendo esta medida amplamente empregada na análise transcultural e na socioloxía. A nosa investigación trata de pechar esta brecha introducindo a cultura supra-nacional, entendida como unha agregación de valores nacionais nos modelos de toma de decisións turísticas, propoñendo unha serie de preguntas e hipóteses:

Pregunta 2 (P2): Como se pode incluír a cultura e en particular a nación de cultura supra-nacional no modelo de empuxe-arrastre? Hipótese 2 (H2): O modelo logit multinomial con variables específicas das alternativas pode explicar as preferencias dos turistas (capítulo 4 e 5).

Pregunta 3 (P3): Como se pode formalizar ó modelo de empuxe-arrastre con variables culturais? Hipótese 3 (H3): A cultura é un factor de empuxe que inflúe nas preferencias dos individuos (capítulo 4 e 5).

A revisión do estado da arte guiounos cara a identificación da brecha no estudo la cultura supra-nacional no contexto de toma de decisións turísticas, así como o seu uso nos modelos de toma de decisións discretas. Polo tanto, comparamos o rendemento destes modelos cos de aprendizaxe automático, que acostuman a ser considerados como os mellores para predicir, o que xera unha serie de preguntas e hipóteses:

Pregunta 4 (P4): Son os modelos de toma de decisións discretas válidos para tarefas de predición? Hipótese 4 (H4): Os modelos de toma de decisións discretas son válidos para tarefas de predición (capítulo 6).

Pregunta 5 (P5): É a cultura a variable que mellor predí os resultados dun modelo de toma de decisións? Hipótese 5 (H5): A cultura é a variable mais importante á hora de explicar as tomas de decisións individuais (capítulo 6).

Pregunta 6 (P6): Cales son os mellores modelos para estudar a toma de decisións dos turistas? Hipótese 6 (H6): Os mellores modelos para predicir a toma de decisións dos turistas son os de aprendizaxe automático, particularmente os métodos de ensamble (capítulo 6).

Como cuarta brecha no estado da arte temos que o modelo causal actual de toma de decisións e cultura, non incorpora as preferencias dos individuos. A falta desta característica provoca que os modelos non poidan cuantificar as preferencias dos individuos de maneira directa nun contexto cultural, de tal maneira que un observador externo (por exemplo un investigador) é incapaz de deducir estas preferencias. Para pechar esta brecha, e superar as limitacións da liña de investigación actual, proponse o deseño dun novo modelo causal para explicar a toma de decisión, de tal maneira que a cultura afecta ás preferencias dos individuos, o que a súa vez afecta á fase de compensación das características conflitivas entre as alternativas. O uso desde novo modelo causal xera dúas pregunta e dúas hipóteses:

Pregunta 7 (P7): Existe un problema co modelo causal actual de cultura en toma de decisións? Como podemos solucionar ese problema? Hipótese 7 (H7): A cultura afecta a como os individuos resolven os dilemas de compensación. Hipótese 8 (H8): Os individuos de diferentes nacións e culturas supra-nacionais resolven o dilema de compensación de forma diferente (capítulo 5).

No capítulo 4 da tese desenvolvemos a análise da "Cultura supra-nacional e a toma de decisións do turista", co obxectivo de sustentar a análise do rol da cultura supra-nacional na toma de decisións turística. Neste capítulo partimos da definición xenérica de cultura baseada no marco teórico do análise transcultural, na que a cultura é un fenómeno colectivo que consiste nun conxunto de crenzas, valores, coñecementos e normas aceptados por un grupo social, que se relaciona estreitamente con factores xeográficos, climáticos e socioeconómicos. Posteriormente realizase unha revisión da literatura da definición de cultura no ámbito do turismo, detectando que se utilizan 4 conceptos de forma recorrente para incorporar á cultura nas análises de toma de decisións: i) cultura nacional do turista; ii) cultura internalizada polo turista; iii) cultura do país de destino, e iv) distancia cultural, o concepto máis estudado en turismo e que actúa como aproximación das diferenzas culturais entre os países de orixe e de destino.

Realizamos unha revisión da literatura do modelo de empuxe-arrastre, que é o marco máis empregado para entender a toma de decisións en turismo. Baixo este marco, os motivos de empuxe estarían relacionados coas necesidades e os desexos dos turistas, mentres que os motivos de arrastre serían os relacionados coas características dos destinos turísticos. Na creación do modelo empuxe-arrastre por parte de Crompton (1979) e Dann (1976), os factores de empuxe e de arrastre non interactuaban entre si, porén Sirakaya & Woodside (2005) e Uysal

& Jurowski (1994) estenderon o modelo de empuxe-arrastre a finais dos anos noventa, de tal maneira que segundo o seu novo modelo, os turistas séntense atraídos polas características dos destinos, entendidos estes como factores externos aos individuos, cando teñen factores internos ou psicolóxicos que os empuxan cara a ese destino. A novidade da análise presentada na nosa investigación é a incorporación dun concepto teórico procedente da análise transcultural, como é o nivel de interpretación supra-nacional, entendido coma un conxunto de valores que comparten os individuos que pertencen a áreas xeográficas máis amplas que as nacións, e que se calculan por medio da agregación das valoracións culturais nacionais, que non se tivera usado con asiduidade no campo do turismo. Para incorporar á cultura supra-nacional no modelo de empuxe-arrastre de turismo, empregamos o mapa supra-nacional de Inglehart (2006), baseado nas agrupacións de Huntington (1993), formalizando a relación por medio dun modelo logit multinomial con variables específicas das alternativas, para poder estudar as preferencias dos individuos polas mesmas. Unha vez definida a nova variable de cultura supra-nacional, incorporámola no modelo de empuxe-arrastre de turismo, parándonos a estudar os factores de arrastre no turismo, como son as características de cada unha das alternativas (rutas) do Camiño de Santiago, onde se inclúen tres variables específicas destas alternativas: i) o seu estatus como Patrimonio da Humanidade outorgado pola UNESCO, ii) a popularidade baseada no número de peregrinos que chegan a Santiago percorrendo cada unha das rutas dispoñibles e, iii) a intensidade do camiño, relacionada coa temperatura, precipitacións e dificultade media das etapas.

Os resultados obtidos no capítulo 4 destacan a relevancia da cultura supra-nacional (medida a través dos 8 clúster culturais definidos por Inglehart) como factor de empuxe na toma de decisións, identificando diferenzas nas preferencias dos distintos grupos culturais. Tamén atopamos que a metodoloxía logit multinomial con variables específicas das alternativas permítenos entender en mellor medida como afectan as características das alternativas ás decisións. Este novidoso avance abre un novo abano de investigacións nos modelos de toma de decisións, non só no ámbito turístico, senón en todas as decisións entre alternativas que impliquen a individuos de diferentes grupos culturais.

Motivado por esta nova liña de investigación que abrimos no capítulo 4, seguimos avanzando na investigación tratando de comprender como os individuos de cada un dos ámbitos culturais toman as súas decisións, para iso o capítulo 5 céntrase en “Como as culturas resolven os dilemas de compensación?”. Onde temos como obxectivos principais crear un novo modelo causal para toma de decisións e analizar os diferentes constructos dos autores máis importantes da análise transcultural, tratando deste xeito de avanzar en dous ámbitos, na medición da cultura e no seu emprego para resolución de conflitos de compensación na toma de decisións. O novo modelo, introduce as preferencias no modelo causal de toma de decisións incluíndo a cultura á que pertencen os individuos, así como o modo en como resolven a compensación a partir das características das alternativas en conflito. En primeiro lugar definimos á compensación das alternativas en conflito como unha fase na toma de decisións, onde individuos ou grupos atópanse con múltiples alternativas e ningunha domina ás outras en todas as características. Detallando o modelo de toma de decisións actual en cultura, os expertos en toma de decisións, Chu & Spires (2008) consideran ás compensacións como estratexias individuais que varían en función do tomador de decisións, pero onde o contexto cultural tamén pode afectar, porén, non inclúen no seu modelo as preferencias dos individuos. Yates & de Oliveira (2016) identifican dúas limitacións principais na liña de investigación actual sobre a toma de decisións e cultura: i) os estudos actuais adoitan comparar só 2 ou 3 países ou culturas como EE. UU., China ou Xapón, o que pode xerar nesgos ao xeneralizar a outras culturas; ii) as comparacións céntranse en países asiáticos e occidentais, onde poucas variables culturais poden explicar gran parte da variación na toma de decisións, pero isto podería non se aplicar ao incluír un rango máis amplo de grupos culturais, especialmente cando estes grupos teñan características similares.

A primeira contribución deste capítulo ven dada pola creación dun novo modelo causal en cultura, que introduce as preferencias na toma de decisións. A segunda contribución está relacionada coa utilización do modelo logit multinomial con variables específicas das alternativas para explicar o novo modelo causal, permitindo analizar con alta capacidade explicativa o dilema da compensación, e resolvendo as dúas limitacións mencionadas por Yates e Oliveira, podendo comparar un gran número de grupos culturais con varios dos constructos culturais máis utilizados en análise transcultural. Os constructos culturais nacionais estudados serán os de Inglehart, empregado no capítulo 4, comparando os seus resultados cos dos constructos culturais de Hofstede (1980) e Schwartz (1992), onde temos dous niveis de interpretación de resultados, o nivel nacional, agregando datos individuais e o nivel de grupo supra-nacional, agregando datos nacionais. Unha vez analizados os resultados dos modelos, eliximos seguir utilizando o marco supra-nacional de análise de Inglehart e Welzel, amplamente recoñecido na literatura, ademais de ter demostrado a súa relevancia no capítulo 4. Ao analizar os resultados dos 10 países con máis chegadas polo Camiño de Santiago, notamos que cada un deles solucionou o dilema de compensación de maneira diferente, polo que nos preguntamos se as culturas supra-nacionais, como as de Inglehart, tamén difiren neste aspecto. Analizando os resultados das culturas supra-nacionais puidemos comprobar que tamén estaban a solucionar o dilema de compensación de maneira diferente entre si. Deste xeito, a nosa investigación destaca a importancia de incorporar os constructos de análise transcultural e todas as súas variables asociadas ao estudo dos dilemas de compensación. Estes resultados son relevantes non so para a solución de conflitos, senón para avanzar no perfilado estatístico de peregrinos neste caso particular e de turistas nun ámbito mais xeral.

Unha vez identificada a cultura supra-nacional como un factor clave nos modelos de toma de decisións e verificado o seu efecto no dilema de compensación. No capítulo 6 estudamos diferentes modelos empíricos para afinar este constructo orientado a incluírse nun sistema recomendador, atopando as mellores metodoloxías e especificacións para utilizar en modelos explicativos e de predición. Os modelos estudados son os de elección discreta, onde se inclúen os dous modelos estudados nos capítulos anteriores; mentres que doutra banda estudamos ós modelos baseados en modelos de aprendizaxe automático, máis en concreto os modelos de árbores de decisións, que segmentan o espazo de predición en ramas, baseado nas variables preditivas. Este capítulo presenta un caso de uso onde predicimos a ruta ou alternativa do Camiño de Santiago que elixiron os peregrinos, comparando por medio de diversas métricas os resultados dos modelos con diferentes especificacións culturais e constructos culturais, revisados nos capítulos 4 e 5, como son os de Inglehart, Hofstede e Schwartz.

Sendo o obxectivo do capítulo o de comparar a capacidade preditiva dos modelos nunha serie de especificacións culturais, en primeiro lugar analizamos a importancia das variables explicativas, atopando que, para todas as especificacións e modelos, as variables culturais son capaces de explicar en torno ao 70-90% da variabilidade atribuíble a cada modelo. Ao analizar cada constructo cultural por separado, no caso de Inglehart, as dimensións tradicional contra secular e supervivencia contra expresión persoal, son ambas esenciais e igual de importantes, no caso do constructo cultural de Hofstede, as dimensións máis importantes son as de individualismo contra colectivismo, evitar a incerteza e indulxencia contra contención, mentres que no caso de Schwartz, as dimensións culturais mais relevantes serían as de igualitarismo e mestría. Calculamos os nosos modelos utilizando parámetros óptimos, que búscamos por medio da técnica de busca en malla, e despois realizamos validación cruzada 10-fold, determinando que a capacidade preditiva é similar en todos os modelos e especificacións, atopándose a precisión de case todos os modelos entre o 50 e o 55%. Os modelos baseados en árbores de clasificación destacan lixeiramente en capacidade preditiva sobre os modelos de clasificación discretos en todas as especificacións, porén, utilizando criterios de selección de modelos para sistemas de recomendadores, que se basean en flexibilidade, complexidade computacional e interpretabilidade os mellores modelos serían a árbore de clasificación simple e o modelo logit multinomial con variables específicas das alternativas. A menor complexidade computacional da árbore de clasificación decantaría a elección da mesma para realizar predicións, mentres que a capacidade de estudar o dilema de compensación entre as características das alternativas do modelo logit multinomial con variables específicas das alternativas concede unha maior interpretabilidade. Por iso, o uso do recomendador final sería o que decantaría a decisión sobre o emprego de un ou de outro modelo. A gran preponderancia pola elección de camiño francés na mostra proposta sería un factor clave que inflúe sobre os resultados, indicando que neste caso o valor engadido de modelos mais complexos é limitado. Este achado ten consecuencias relevantes en dous ámbitos i) na necesidade de analizar o conxunto de alternativas, o seu peso e as características das mesmas a hora de establecer complexos modelos nos sistemas de recomendación, ii) a posibilidade de comparar estes resultados con mostras mais equilibradas e comparar a ganancia observada dos diferentes modelos dependendo das mostras, nesgos, etc...

En conclusión, a través da investigación estudamos a rol da cultura no proceso de toma de decisións, centrándonos especialmente no concepto de cultura supra-nacional, que foi pouco estudada no campo do turismo. Ó longo do traballo demostramos que o concepto de cultura supra-nacional é esencial para entender as preferencias dos turistas ante problemas de toma de decisións, podendo ser incorporado ao modelo de empuxe-arrastre de turismo, e recomendando a inclusión deste concepto ao abano de técnicas que utilizan ós investigadores deste ámbito de estudo. A inclusión do concepto de cultura supra-nacional nos modelos de toma de decisión turística como factor de empuxe foi alcanzado por medio da utilización dos modelos logit multinomiais con variables específicas das alternativas, os cales destacaron pola súa capacidade explicativa ao coñecer as valoracións das características das alternativas por parte dos individuos, abordando o primeiro sub-obxectivo da tese.

A utilización dos modelos logit multinomiais con variables específicas das alternativas permitiunos coñecer con precisión as preferencias dos individuos sobre estas características como investigadores externos ó problema de decisión. Así, a inclusión de preferencias no novo modelo causal con respecto ao antigo modelo, configuraron un modelo no que as preferencias se ven afectadas pola cultura , influenciando o xeito de solucionar os dilemas de compensación entre os atributos conflitivos das alternativas, afectando globalmente ao proceso de toma de decisións, cumprindo o sub-obxectivo 2 da tese.

A creación e utilización do novo modelo causal de toma de decisións permitiunos resolver as brechas localizadas na revisión da literatura en canto ao antigo modelo de toma de decisións e cultura, consistente nunha análise mais limitada, orientada a un reducido número de países e culturas de maneira simultánea, así como a utilización de escasas variables culturais cando se consideraban problemas referentes á solución do dilema de compensación entre características das alternativas. Logramos resolver esta brecha por medio da utilización dun novo modelo causal, que nos permite analizar unha gran variedade de culturas, países e constructos culturais, ó mesmo tempo que coñecemos ás súas preferencias sobre as características das alternativas, dándonos coñecemento sobre como cada grupo soluciona o dilema de compensación, acadando o sub-obxectivo 3 da tese.

Ó longo da tese demostramos que case todas as variables e constructos culturais orixinais da análise transcultural empírica, así como transformacións das mesmas son relevantes nos modelos de toma de decisións. Unha vez avaliados os modelos de toma de decisións discretas e os de aprendizaxe automático, comprobamos que que os modelos de árbores de clasificación e o modelo logit multinomial con variables específicas das alternativas, presentan un mellor rendemento en canto ó seu balance entre interpretabilidade, flexibilidade e complexidade computacional, completando así o sub-obxectivo 4 da tese.

A investigación confirma as hipóteses relacionadas co obxectivo principal da tese, estudar a influencia da cultura na toma de decisións dos individuos. Os resultados calculados co modelo logit multinomial con variables específicas das alternativas amosan que as culturas supra-nacionais ás que pertencen os individuos afectan ás preferencias dos individuos, e as preferencias a como resolven os dilemas de compensación sobre as diferentes características das alternativas, evidenciado polo uso do novo modelo causal proposto neste estudio. A investigación tamén confirma que os individuos de diferentes culturas, tanto a nivel nacional como supra-nacional difiren na súa solución dos dilemas de compensación. Os modelos de toma de decisión discretos foron validados en tarefas preditivas baseadas no caso de uso, amosando un rendemento similar ao dos modelos de aprendizaxe automático, onde a cultura aparece como a variable máis importante para explicar á variabilidade dos modelos. A pesar da confirmación de todas estas hipóteses anteriormente comentadas, a hipótese de que os modelos de aprendizaxe eran os de mellor rendemento predicindo a toma de decisións dos turistas non se confirma, dado que o seu rendemento foi similar ó calculado para os modelos discretos de toma de decisións.

As implicacións sociais do estudio virían dadas polo feito de que os grupos culturais supra-nacionais inflúen no comportamento dos membros que os compoñen, en relación á toma de decisións turísticas, de tal maneira que é posible agregar os valores individuais en oito niveis supra-nacionais para ser utilizados en modelos de toma de decisións. No ámbito do márketing, recomendamos o uso da cultura supra-nacional para segmentar os mercados en lugar de realizar modelizacións individuais para cada un dos países sobre os que queiran actuar, co obxectivo de optimizar a asignación dos recursos das empresas e poder chegar a un maior número de mercados con un custe menor, ao centrarse nas características culturais concretas de cada unha das agrupacións de países. Na formulación de políticas públicas, a investigación permite comprender as preferencias dos consumidores a través da cultura supra-nacional, nun modelo de toma de decisións que axuda a deseñar políticas específicas para rexións, mellorando á xestión dos destinos adaptando as estratexias políticas en función dos impactos sociais, ambientais e económicos, e aplicando taxas ou descontos dirixidos a grupos culturais específicos.

Finalmente, en relación coa transferencia á sociedade da tese, esperamos que esta contribúa á mellora dos sistemas de recomendación das empresas, axudando a autoridades turísticas e investigadores no desenvolvendo de modelos de toma de decisións máis sofisticados ao integrar a dimensión cultural, tan comunmente pasada por alto no eido da economía.

ABBREVIATIONS

Abbreviation	Name
AEMET	State Meteorological Agency
AR	Autoregression
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
D	Distance
DF	Degrees of Freedom
Dif	Difficulty
DifI	Difficulty Index
DM	Decision-making
DW	Distance Weight
EES	Salary Structure Survey
EUC_Hofstede	Euclidean distance of Hofstede's national values
EUC_Inglehart	Euclidean distance of Inglehart's national values
EUC_Schwartz	Euclidean distance of Schwartz's national values
GDP	Gross Domestic Product
GIS	Geographical Information System
Hof_idv	Individualism vs Collectivism dimension of Hofstede's national values
Hof_ivr	Indulgence vs Restraint dimension of Hofstede's national values

Abbreviation	Name
Hof_ltowvs	Long-Term Orientation vs Short-Term Orientation dimension of Hofstede's national values
Hof_mas	Masculinity vs Femininity dimension of Hofstede's national values
Hof_pdi	Power Distance vs Closeness dimension of Hofstede's national values
Hof_uai	Uncertainty Avoidance vs Acceptance dimension of Hofstede's national values
IGN	National Geographic Institute
II	Intensity Index
INE	National Statistical Institute
IQR	Interquartile Range
KS_Hofstede	Kogut and Singh distance of Hofstede's national values
KS_Inglehart	Kogut and Singh distance of Inglehart's national values
KS_Schwartz	Kogut and Singh distance of Schwartz's national values
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
Max	Maximum
Min	Minimum
MSE	Mean Squared Error
ONET	Occupational Information Network

Abbreviation	Name
P	Precipitation
PI	Precipitation Index
Q1	Quartile 1
Q3	Quartile 3
RMSE	Root Mean Squared Error
SARIMA	Seasonal Autoregressive Integrated Moving Average
Sch_aff_auton	Affective Autonomy dimension of Schwartz's national values
Sch_egalitar	Egalitarianism dimension of Schwartz's national values
Sch_embedded	Embedded dimension of Schwartz's national values
Sch_harmony	Harmony dimension of Schwartz's national values
Sch_hierarchy	Hierarchy dimension of Schwartz's national values
Sch_intel_auton	Intellectual Autonomy dimension of Schwartz's national values
Sch_mastery	Mastery dimension of Schwart's national values
SD	Standard Deviation
Surv_Self	Survival vs Self-Expression dimension of Inglehart's national values
T	Temperature
TBATS	Trigonometric Seasonality, Box-Cox transformation, ARMA errors, Trend and Seasonal components
TI	Temperature Index
Trad_Sec	Traditional vs Secular dimension of Inglehart's national values

Abbreviation	Name
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNWTO	United Nations World Tourism Organization
WTCC	World Travel and Tourism Council
WVS	World Values Survey

DEFINITIONS

Compostela: is a certificate granted by the Pilgrim's Welcome Office to pilgrims who have completed Saint James Way, where there is a distance requirement. If the pilgrimage is done on foot or horseback, it's necessary to cover 100km to obtain this credential, or 200km if completed by bicycle.

Pilgrim's Welcome Office: is the office under the authority of the Cathedral of Santiago de Compostela within the Archdiocese of Santiago de Compostela, where pilgrims arriving in the city of Santiago de Compostela are received and obtain the final stamp of Saint James Way from the Cathedral of Santiago, thus issuing the "Compostela".

Holy Year: also known as a "Xacobeo" in Galicia, is a Catholic celebration that occurs when July 25th, the day of Saint James the Apostle, falls on a Sunday. During this time, anyone who crosses the Holy Door of the Cathedral of Santiago de Compostela in that year receives the jubilee. Besides the appeal this holds for Catholic pilgrims, both the Galician regional government (Xunta de Galicia) and the City council of Santiago actively promote Saint James Way during Holy Years, specially oriented to Spanish population, leading to an increase in arrivals of pilgrims from this country.

CHAPTER 1: INTRODUCTION

1.1 MOTIVATION OF THE INVESTIGATION

Decision-making (DM) by rational economic agents constitutes one of the key topics in the study of economics and particularly is the basis of microeconomics. In the field of economics, there was a discovery known as Random Utility Models (McFadden, 1973) which revolutionized the area of microeconomics by being able to represent how economic agents make decisions among multiple alternatives through the study of their preferences. The use of this new type of models allowed for the quantification of the utility derived from each alternative for each agent, taking into account unobservable factors (Mcfadden, 1980), transforming it into probabilities, thus creating discrete decision-making models within the economic domain. The inclusion of psychological factors in DM, specifically those related to cognitive biases, emotions, or influences, is a main subject on a branch of economics commonly known as behavioural economics, which complement traditional economic theory by making it more realistic. However, the role of context, specifically the tradition within the realm of sociology, has been less developed and constitutes an area of analysis of great interest for the new economy. Thus, a branch of social-psychology, such as cross-cultural research (Hofstede, 1980; Schwartz, 1992; Weber & Hsee, 2000), has highlighted the importance of including variables that explain the context in which individuals live, to better study how individuals from different cultures exhibit different behaviours regarding DM. To introduce this concept of culture into individuals' DM, we resort to several empirical definitions, both at the level of study and cultural construct. This involves conducting an in-depth review of the concept of supra-national culture from sociology and integrating this concept into the economic and decision-making domain. In order to explore the influence of sociological factors like supra-national culture on decision-making, we examine a highly relevant economic domain: tourism. This allows studying how individuals' decisions are shaped by cultural backgrounds within a broader context, shedding light on sociological impacts on decision-making processes.

The motivation for this research is to merge knowledge from decision-making, behavioural economics and social-psychology, aimed at contributing at the cutting edge of research. This aims to open new interdisciplinary areas of work and contribute to the understanding of individual decisions within their context. Hence, it's interesting to comprehend cultures to better define economic decisions.

After reviewing the state of the art and the relevant concepts, we analyse the candidate methodologies to include cultural variables as explanatory factors. Thereafter, we consider the cultural variables that might be useful in DM models. To contrast our models, we employ the Pilgrims Welcome Office Registry, which we found as a suitable testbed to validate our models and variables. This registry records the arrivals of pilgrims that complete Saint James Way and receive the "Compostela" certificate. The database comprises over four million records of pilgrims, who choose among the seven main ways that lead to the city of Santiago de Compostela. The pilgrims come from more than 50 different countries that can be grouped into different cultural constructs. The rich dataset allows us to address the research objectives that we present in the next section.

1.2 STATE OF THE ART

The current state of the art highlights our focus on three paradigms: decision-making, culture, and tourism. The exploration led us to understand individual choices in various contexts, recognizing the influential role of culture in DM processes. This review aims to summarize the intersection of these paradigms, especially prominent in the domain of tourism, where the interplay between DM and cultural dynamics significantly influences tourist's choices.

1.2.1 Decision-making

Decision-making (DM) is a fundamental concept in economics and comprises the process of selecting a course of action or choosing from among various alternatives based on a set of preferences, values, information, and potential outcomes (Yates & de Oliveira, 2016). The main goal of DM is to understand the behaviour of economic agents in the field of microeconomics (Arrow, 1959; Kreps, 2020; Varian, 2010). The introduction of Random Utility Models (McFadden, 1973) marked a revolutionary shift in the field, offering a mathematical framework to depict how economic agents make decisions when presented with multiple alternatives. This innovative approach enabled the introduction of uncertainty in the quantification of the utility derived from each alternative, by means of taking into account the unobserved factors (Mcfadden, 1980). Consequently, these models transform utilities into probabilities, giving rise to discrete DM models within the economic domain.

The estimation of utilities involve the application of trade-offs to weight the relevancy of the alternatives' attributes Chu et al., (1999, 2005) and Chu and Spires (2008) proposed that culture can influence trade-offs, and these trade-offs, can resolve the conflict, which can be represented as the following causal relationship: [Culture -> Trade-off -> Conflict]. However, this model lacks the explicit consideration of individual preferences. The concept of choice under conflict, or a trade-off dilemma (Tversky & Shafir, 1992), emerges when individuals or groups are faced with alternatives, none of which is superior in all desirable characteristics (Yates & de Oliveira, 2016). During this phase of DM, decision-makers struggle with the challenge of making trade-offs between conflicting attributes, estimating the relevance of each attribute in relation to others. While trade-offs are individual strategies that vary among decision-makers, cultural backgrounds can shape how they resolve the conflict. The rationale of Chu and Spires (2008) suggests that a decision-maker's culture may determine their trade-off strategy, serving as the mechanism to resolve conflicts encountered during the choice process.

Despite all the contributions in the field of culture and DM, Yates & de Oliveira (2016) highlighted a series of limitations in the current state of the art. Firstly, they focus on comparing only two or three national cultures, particularly Eastern cultures against Western cultures (Vijver & Leung, 2021; Wang, 1996; Wright & Phillips, 1980), introducing potential bias and restricts the generalization of results. Secondly, the reduced number of studied cultures may generate a bias on the set of cultural factors capable of capturing most of the differences in DM between cultures, neglecting the influences of other cultural factors in this process.

1.2.2 Culture

As seen in the review of the state of the art in DM, culture can play a key role as an explanatory factor in DM models. Thus, we delve deeper into defining the concept of culture, a widely debated topic in several social sciences (Baldwin et al., 2006; Kroeber & Kluckhohn, 1952). Most definitions of culture from cross-cultural analysis focus on culture being a collective phenomenon (Hofstede et al., 2005), consisting of a set of beliefs, values, knowledge, and norms accepted by a social group (Schwartz, 1999; Inglehart, 1997). From the empirical branch of cross-cultural analysis arises cultural constructs, which are conceptual frameworks, used for identifying and comparing different cultures. In what follows, we review the three most relevant paradigms.

- Inglehart's national values: drawn from seven rounds of the World Values Survey spanning from 1981 to 2022. The data reveals two dimensions, traditional versus secular and survival versus self-expression. Typically presented in cultural map from, Inglehart's work introduces the concept of cultural clusters or supra-national cultures.
- Schwartz's national values: stem from rounds of values surveys conducted between 1988 and 2005 for teachers and students across various countries. The resulting seven dimensions—Harmony, Embeddedness, Hierarchy, Mastery, Affective Autonomy, Intellectual Autonomy, and Egalitarianism, offering a deep insight into societal variations and trends.
- Hofstede's national values: gathered from questionnaires administered to IBM employees in the late 1960s, the initial four dimensions—Individualism versus Collectivism, Power Distance, Masculinity, and Uncertainty Avoidance—have been enriched in the 2000s with two additional dimensions, Long-term Orientation, and Indulgence versus Restraint.

Once the national values of the main authors of cross-cultural analysis has been reviewed, the question arises about how cultures can be identified or defined. Authors from different fields arrived at different solutions, sparking extensive debates on the most effective way to empirically identify the culture to which individuals belong. Firstly we have the national values, proposing a solution by aggregating individual cultural values at the national level, widely used in social-psychology, sociology and cross-cultural research (Hofstede, 1980; Inglehart, 1997; Schwartz, 1994). Secondly, supra-national culture measure involves aggregating national values into a broader framework. Empirical studies by Akaliyski et al. (2021) and Michael Minkov & Hofstede (2012) support the validity of using supra-national culture, demonstrating that most variability in cultural values of individuals can be explained by affiliation with different supra-national cultures. Despite abundant research on national cultures, the exploration of supra-national cultures is limited, with advocates from cross-cultural analysis and sociology like Inglehart (1997b), Welzel (2012), and Beugelsdijk & Welzel (2018) using this approach. The identification of can be complemented with the study of discrimination between them, with cultural distance playing a key role in this process, by acting as a proxy indicator of cultural differences between countries (Akaliyski, 2017; Beugelsdijk et al., 2018; Kaasa & Minkov, 2020; Kandogan, 2012).

1.2.3 Tourism

The most popular paradigm to understand tourism DM is the push-pull model (Crompton, 1979; Dann, 1976; Pestana et al., 2020; Sirakaya & Woodside, 2005; Uysal & Jurowski, 1994). In its initial phase, the push-pull model asserted that push and pull factors operated independently of each other (Crompton, 1979; Dann, 1976). However, this viewpoint has evolved over time, scholars like Cha et al. (1995) and Uysal and Jurowski (1994) proposed that tourists embark on a journey only when push and pull factors interact. This implies that, for tourists to be attracted to a destination attribute, they must evaluate that destination positively. Subsequent developments in the push-pull model, presented by (Sirakaya & Woodside (2005) and Uysal et al. (2008) classified push motives as internal variables and pull motives as external variables, which is the prevailing perspective nowadays. Push motives, including the desire for escape, rest, prestige, and social interaction, encourage travel, while destinations with attractive features act as pull motives. External variables related to destination characteristics, such as cuisine and history, contribute to tourists' DM (Correia et al., 2013; Pestana et al., 2020; Uysal & Jurowski, 1994). Thus, push-pull model can be used to study a wide variety of situations in which tourists must make decisions: transportation modes, analysing decisions related to traveling to and within destinations (Cheng et al., 2018; Iseki et al., 2018; Walters et al., 2019), selection of products, services (Ismoilov, 2017; Jung et al., 2015; Kim & Park, 2017; Tung & Soo, 2004) and experiences during tourism activities, investigating decisions made before, during, or after engagement, the destination (Lacher et al., 2013; Masiero & Qiu, 2018; Oppewal et al., 2015).

Tourist’s DM is also influenced by the culture to which the tourists belong. Evidence has been found that culture particularly affects: perception of trip satisfaction (Crotts & Erdmann, 2000; Huang & Crotts, 2019; Li et al., 2010), the selection of a destination (Hagag et al., 2015; Hsu et al., 2013; Litvin et al., 2004; Ng et al., 2007; Qian et al., 2018; Woodside et al., 2011; Yang et al., 2019), the perception of risk associated with a destination (Kozak et al., 2007; Yang & Wong, 2012) and the competitiveness of tourist destinations (Kumar & Dhir, 2020).

Recalling the issue of how to identify/define cultures (see previous section), we found that tourism literature has extensively employed four cultural measures, all of which happened to be variants of the interpretation of national culture. The first measure focuses on the tourist's national culture (Crompton, 1979; Kozak, 2002; H. Liu et al., 2018; Ng et al., 2007), treating culture as a collective phenomenon encompassing beliefs, values, knowledge, and norms within a social group. The second measure is the tourist’s internalized culture at the individual level, including personal values and beliefs, influencing destination selection and trip characteristics. The third measure involves destination’s culture, shaping how international tourists perceive the nation in which their chosen destination (McKercher & Cros, 2003; Rinschede, 1992) is situated and influencing the decision to travel. Finally, cultural distance is explored as a proxy for cultural gaps between countries (H. Liu et al., 2018; Manosuthi et al., 2020; Yang et al., 2019), introduces risk and uncertainty into tourism dynamics, affecting DM processes, where cultural distance is usually assessed through differences in cultural values. None of them included the concept of supra-national culture, thus opening an interesting line of research in the field of tourism research.

The comprehensive state of the art review has provided a thorough understanding of the current landscape on the intersection between DM, culture, and tourism. This synthesis of perspectives enhances our appreciation of the complexities within the domain and sets the stage for the formulation of the use case, research goals, research questions, and hypotheses, all of these are based on the review of the state of the art.

1.3 USE CASE: TOURISM

According to the WTO (2023), tourism has become one of the activities with the most sustained economic growth in the last decades. Throughout history, significant migrations have occurred for economic, security, familial, or religious reasons. However, mass tourism, as we know it today, emerged after World War II due to increased union power, securing the right to paid vacations in most countries (Green, 1997), rising worker purchasing power (Lucas, 1988), and opening of new transportation ways (Comín, 2014), enabling the development of mass tourism. The growth of this sector has stimulated other sectors of the economy due to its dragging effect, attracting significant foreign direct investment as has proved to be a highly lucrative industry. The development of this activity has led to a positive feedback loop, where destinations hosting tourists gain extraordinary income they wouldn't receive otherwise, while simultaneously both the origin and destination countries of tourists engage in positive cultural exchanges.

The economic importance of the tourism sector is essential, accounting for 10.3% of the world's GDP and 10% of generated employment (WTTC, 2020). This sector has been developing and flourishing in Spain since the 1960s, currently contributing to approximately 12.5% of Spain's GDP and 12.7% of its employment in 2019, as per the Tourism Satellite Account data (INE, 2023). In Galicia, this sector represented 10.4% of its GDP and generated 11% of the employment in 2017 (Impactur, 2017). The sector's significance for the Galician economy justifies the use case of studying DM system of pilgrims, related to Saint James Way, constituting one of the main attractions when choosing the Autonomous Community of Galicia as a tourist destination (Fernández Méndez et al., 2019), being one of the most important pilgrimage sites for the Catholic religion since the 10^{th} century (Murray & Graham, 1997). The importance of Saint James Way measured by tourism received by the city of Santiago de Compostela has been increasing substantially in the last three decades, this evidenced at the Pilgrim's Welcome Office Registry (Pegregrino, 2023), receiving an average of 25,000 pilgrims from 1987 and 1997 to over 4,000,000 pilgrims from 2003 to 2022, with 440,000 arriving in 2022, whose estimated direct expenditure exceeded 300 million euros (Rodríguez, 2019) from 2017 onwards.

1.4 RESEARCH GOALS

The review of the state of the art has motivated the main research goal, which is to analyse the role of culture in tourists' DM. This encompasses the three main themes of the thesis, DM, culture, and tourism. To achieve this, we will develop new methodologies for predicting and classifying products or services consumption by integrating DM models, from the field of econometrics and machine learning, fulfilling four specific goals:

Specific Goal 1: Analyse the role of culture in DM, specifically focusing on using supra-national culture as push factor in the push-pull model, employing choice models to formalize the relationship (Chapter 4 and 5).

Specific Goal 2: Build a new causal model, such that culture influences the preferences for the attributes of alternatives by decision-makers, influencing how they resolve trade-off dilemmas between alternatives due to the conflict among them (Chapter 5).

Specific Goal 3: Analyse the constructs and variables of the conceptual frameworks of founders of empirical cross-cultural analysis in DM processes (Chapter 5).

Specific Goal 4: Find the best methodology to implement a recommendation system including several cultural constructs (Chapter 6).

1.5 RESEARCH QUESTIONS AND HYPOTHESES

In this section we explore research questions and hypotheses related to the research goals, which are grounded in the open issues and gaps found in the state of the art.

One of the open issues in Tourism DM literature is the role of culture in the push-pull model. The advances presented by (Sirakaya & Woodside, 2005; Uysal et al., 2008) have shaped the current state of push-pull model, they classified push motives as internal variables related to individuals' desires or preferences, and pull motives as external variables. Based on the review of the state of art, we know that culture can influence people's preferences, thus we pose a question in this regard:

Question 1 (Q1): Which is the effect of including supra-national culture in tourism decision-making? Hypothesis 1 (H1): Supra-national culture condition the decision-making process (Chapter 4 and 5).

Another issue regards with the interpretation of culture in Tourism DM. The current cultural measures employed in tourism literature do not consider supra-national culture, a measure used in cross cultural psychology and sociology that relate to a group bigger than the nation. Our research aims to address this by introducing supra-national culture into a tourism DM model, thus we pose several questions in this regard:

Question 2 (Q2): How can culture, and particularly the notion of supra-national culture, be incorporated into the push-pull model? Hypothesis 2 (H2): Multinomial choice models can explain tourist's preferences (Chapter 4 and 5).

Question 3 (Q3): How can we formalize the push-pull model with cultural variables? Hypothesis 3 (H3): Culture as a push factor due to the influence in preferences of individuals (Chapter 4 and 5).

The review of the state of the art led us to the identification of a gap in the study of supra-national culture in DM in tourism setting, as well as in the use of discrete choice models.

Therefore, we compared the predictive performance of these models with machine learning models, usually considered as the best in this setting, generating a series of questions:

Question 4 (Q4): Are discrete choice models valid for prediction tasks? Hypothesis 4 (H4): Discrete choice models are valid for prediction tasks (Chapter 6).

Question 5 (Q5): Is culture the best variable to predict decision-making outcomes? Hypothesis 5 (H5): Culture is the most important variable at explaining individuals' decision-making (Chapter 6).

Question 6 (Q6): Which are the best models for studying tourist's decision-making? Hypothesis 6 (H6): The best models for predicting tourists' decision-making are those from machine learning, particularly ensemble methods (Chapter 6).

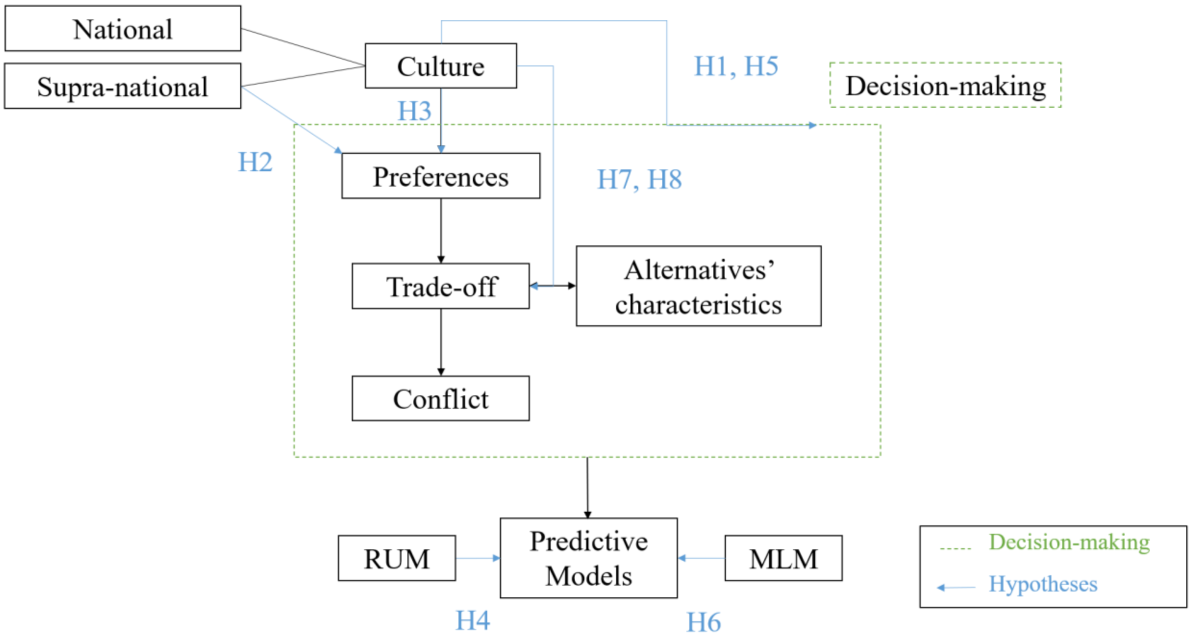
The current causal model lacks the incorporation of individual preferences. This absence results in a deficiency of models that explicitly quantify preferences in DM in cultural context, in such a way that an external observer (i.e. a researcher) can't be estimated. To address this gap and overcome the limitations of current research line, the proposal involves allowing preferences to be represented through utility functions. This approach enables the assessment of alternative features during the trade-off phase, allowing individuals to order alternatives based on their preferences when the process concludes.

We design a new causal model presented on Figure 1 oriented to explain DM, such that, culture (supra-national culture) affects individual's preferences, which affect the trade-off phase of DM between conflicting alternatives' characteristics. Once we confirmed the usefulness of this causal model, through the study of the model along with discrete choice models, we performed a series of models from the family of discrete DM models and machine learning to generate predictions with a touristic recommender system.

The introduction of preferences results in the update of the current causal model into the new causal model: Culture -> Preferences -> Trade-off -> Conflict. The use of the new causal model raises a series of questions:

Question 7 (Q7): Is there a problem with the causal model of culture in decision-making? How can we solve the problem? Hypothesis 7 (H7): Culture shapes how individuals resolve trade-offs dilemmas. Hypothesis 8 (H8): Individuals from different cultural background differ in how they resolve the trade-off dilemma (Chapter 5).

Figure 1. New causal model on decision-making with hypotheses



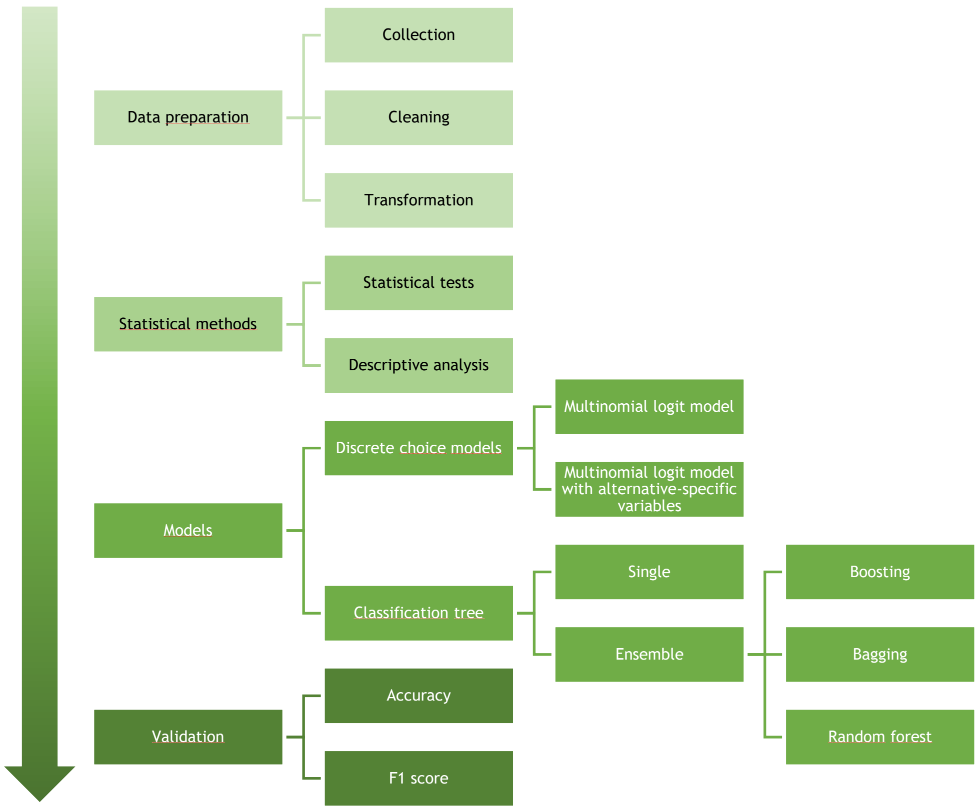
Source: own elaboration

We plan to address the research goals by examining the questions and hypotheses considering the results from chapters 4, 5, and 6, with the causal model presented in Figure 1. Our focus is on deriving meaningful conclusions and implications from these findings, ensuring an alignment with the research goals.

CHAPTER 2. METHODOLOGY

This section outlines the main methods used in the thesis, illustrated in methodological workflow diagram of Figure 2. The workflow follows the CRISP-DM (IBM, 2012) methodology used in the research, which was published to standardize data mining processes and remains the reference methodology in data analytics projects. Our research required Data preparation techniques, Statistical methods, Models, and Validation procedures, whose details are provided in the following sections.

Figure 2. Methodological workflow diagram



Source: own elaboration

2.1 DATA PREPARATION

The initial phase involved to retrieve data from the Welcome Pilgrim's Office Registry. This was carried out by stablishing a connection to the Pilgrim's Office MariaDB Database using an ODBC connector. Afterwards, we downloaded and pre-processed the available data. In parallel, multiple meetings were required to address and resolve problems.

The workflow then proceeded towards to the creation of intermediate datasets and the development of a comprehensive relational structure between databases, needed to construct the initial exploratory data analysis and visualization in Power BI. Data was retrieved on a periodic basis and each of the annual (2003-2022) individual databases were recompiled in text format for later usage. Data analytics of the 4,000,000 pilgrim's profiles as well as the attributes of the chosen ways were carried out using R and Python scripts.

Simultaneously, a long and effortful data cleaning process was conducted, which involved both the main database (Pilgrim's Welcome Office) and the rest of the data sources used in this thesis. The process started with the identification of issues related to missing data, incomplete information, and coding errors within each of the data sources, and proceeded with specific requests to the data managers of the Pilgrim Office to solve these data problems. A feedback loop was therefore stablished to repeat the process back and forth through the development of this thesis. The tools used for data cleaning were the R packages: tidyverse 2.0.0 (Wickham et al., 2019) and fastDummies 1.7.1. (Kaplan, 2020) from R version 4.3.2.). After data cleaning, a process of encoding (i.e. string values storing numerical values could be converted into numerical variables) or construction (i.e. cultural distance of a pilgrim's national culture to Spanish national culture) was carried out. Finally, a feature selection process was followed in order to identify relevant variables, where relevancy was determined by both theoretical criteria and the relationship with the research objectives of the thesis.

2.2 STATISTICAL METHODS

Once the data preparation was completed, we employed statistical methods assessing the descriptive characteristics of the variables as well as the relationships between them.

2.2.1 Descriptive statistics

We started descriptive analysis conducting the statistical analysis of numerical variables (minimum, maximum, range, mean, quartile 1, median, quartile 3, interquartile range, mode, standard deviation, and variance) and categorical variables (frequencies), performed with packages tidyverse (Wickham et al., 2019), psych (Revelle, 2015), rstatix (Kassambara, 2020), summarytools (Comtois, 2016) and corrplot (Wei et al., 2017) for R version 4.3.2, and Pandas (McKinney, 2011) and Numpy (Harris et al., 2020) for Python version 3.8.5, this descriptive analysis can be seen in Annex 4. Subsequently, we conducted correlation analysis of the numerical variables, and finally, we performed a data visualization with package tidyverse from R version 4.3.2 and a descriptive commentary, that can be seen in section 3.3.

2.2.2 Statistical test

Pearson's chi-squared test (Rao & Scott, 1984) is a method used to assess the degree of association between two categorical variables, where it compares the observed frequencies in a contingency table with the expected frequencies. Posing a null hypothesis that there is no significant association between variables:

$$H_0: O_i = E_i \tag{1}$$

That we compare with the value of the statistic of Pearson's chi-squared:

$$\chi^2 = \sum_{i}^{n} \frac{(O_i - E_i)^\wedge 2}{E_i} \tag{2}$$

where $\boldsymbol{\chi^2}$ is Pearson's cumulative test statistic, $\boldsymbol{O_i}$ the number of observations of type $\boldsymbol{i}$ and $\boldsymbol{E_i}$ the expected number of observations of type $\boldsymbol{i}$. The statistical results are presented in standard form: "*" means $p < 0.1$, "**" means $p < 0.05$, "***" means $p < 0.01$ and "****" means $p < 0.001$.

We employ Pearson's chi-squared test in chapter 4 to check if different supra-national clusters make different decisions, measured by the estimated probability of selecting each way, verifying whether two or more probability distributions come from the same distribution by measuring the discrepancy between them.

2.3 MODELS

2.3.1 Econometric models of DM

DM involves choosing among a set of alternatives, usually with more than two alternatives, thus, individuals or groups needs to deal with multiple alternatives simultaneously, a complexity that is not considered in the simple choice models as binary logit (Berkson, 1944) or binary probit (Bliss, 1934) that need a generalization to be used as multinomial models (Train, 2009).

2.3.1.1 Multinomial logit model

The multinomial logit model can be defined as a model in which a set of decision-makers have to choose among more than two alternatives, the utility function can be constructed in two ways, using individual-specific variables, and using alternative-specific variables. Starting with the definition of the simplest multinomial model, where the form of utility is that of a multinomial logit model, and where we need to predict the probability that every individual chooses each of the alternatives. The choice probability of the model can be written following Train (2009) notation as:

$$P_{nj} = \frac{e^{\beta_n * X'_{nj}}}{1 + \sum_{j=1}^{J} e^{\beta_n * X'_{nj}}} \tag{3}$$

where P is the probability, n refers to the individual, j to the alternative, β is the vector of parameters, and X is the matrix of observed features.

With the goal of studying how individuals make decisions, it is relevant to know which the main characteristics of the alternatives are and how this influences the preferences in their DM process. So, we use an extension of multinomial models by adding the alternative-specific variables on top of the individual-specific variables, thus using multinomial models with alternative-specific variables, also known as multinomial choice models, whose probability that can be written as follows:

$$P_{nj} = \frac{e^{\beta_n * X'_{nj} + \alpha_j * Y'_j}}{1 + \sum_{j=1}^{J} e^{\beta_n * X'_{nj} + \alpha_j * Y'_j}} \quad (4)$$

where P is the probability, n refers to the individual, j to the alternative, β is the vector of parameters, X is the matrix of observed features, α vector of parameters of the alternatives and Y the matrix of observed features of the alternatives.

Searching the best econometric models to explain the relationship between culture and tourists' DM, we found that if we wanted to predict a discrete multi categorical variable, we needed to resort to the use of multinomial econometric techniques for discrete choice modelling, as the multinomial logit model is an expansion of the classical logit model (McFadden, 1973), which allow us to include more than two categories to predict (Train, 2009), and that we perform using mlogit (Croissant, 2020) package in R.

Once we have defined these models, we found the necessity of investigating the particular characteristics of each alternative, to address this, we resort to the use of multinomial choice models (Train, 2009), which are suitable to deal with the complexities of addressing the preferences of individuals for the characteristics of multiple alternatives. Thus, examining the DM problem exemplified by our use case (exploring how culture impacts pilgrims' DM among the 7 alternatives of Saint James Way), these models are deemed the most suitable at explaining the phenomenon, performed on chapter 4, 5 and 6. The use of these methodologies will serve us to validate hypothesis related to discrete choice models, namely H1, H2, H3, H4, H6, H7 and H8.

2.3.2 Machine learning models of DM

In the field of Artificial Intelligence and Machine Learning, different types of DM models have been developed (Almomani et al., 2023; Fernández-Delgado et al., 2014; van Cranenburgh et al., 2022). These models provide highly valuable for classifying and predicting outcomes, finding applications in recommender systems as alternatives to discrete-choice models.

2.3.2.1 Single tree

Starting with single trees, these are a type of model that segments the predicted variable using binary recursive partitioning, splitting the features into regions forming a tree-like structure (Gareth et al., 2021). The recursive binary splitting involves the consideration of the splitting variable X_j and a split point s, ensuring that the splitting at the specified split point is:

$$R_1(i,s) = \{X \mid X_j \geq s\} \; and \; R_2(i,s) = \{X \mid X_j < s\} \tag{5}$$

where R_J are the non-overlapping regions, and the tree can be formally described as:

$$f(X) = \sum_{m=1}^{M} c_m 1(X \in R_m) \tag{6}$$

where c_m is a constant assigned to each terminal node that uses classes to predict the categorical response. As the variable to be explained is discrete, classification trees will provide us with the mode or majority on the terminal node (Almomani, 2020), performed with the package Scikit-Learn (Pedregosa et al., 2011) from Python version 3.8.5.

2.3.2.2 Ensembles

An ensemble model is an aggregation of single models or predictors. In our case, we have built ensembles of multiple classification trees. Although single tree models have an advantage in terms of interpretability over tree-based ensemble models, the latter generally improve upon the former in terms of performance and robustness.

Boosting

Combines multiple trees in such a way that the trees grow sequentially, based on the information from the previous trees. The aggregated prediction of the boosting method can be written as follows:

$$f(x) = \sum_{b=1}^{B} \lambda f^{b}\,(x) \tag{7}$$

where B is the number of trees, λ the shrinkage parameter (small positive number), and f^{b}(x) the prediction of each single tree.

Bagging

Generates multiple single classification trees using Bootstrap and then aggregates them. The result of the bagging method can be written as follows:

$$f(x) = \frac{1}{M} \sum_{m=1}^{M} f^{b}\,(x) \tag{8}$$

where M are the different bootstrapped training datasets and $f^{b}(x)$ the prediction of each tree, that it is going to be aggregated.

Random forest

Selects random explanatory variables within the database using Bootstrap and averages the results afterwards. The random forest method can be written as follows:

$$f(x) = \frac{1}{M} \sum_{m=1}^{M} f^{b}\,(x) \tag{9}$$

where M are the different bootstrapped training datasets with random explanatory variables and $f^{b}(x)$ the prediction of each tree, that it is going to be aggregated.

Given that the main focus of the comparison between discrete choice models and machine learning models is their ability to infer or predict, thus, we find that parametric models provide much greater interpretability than more complex methods like ensembles, Support Vector Machines (SVM), and neural networks (Gareth et al., 2021). These latter methods are often more challenging to interpret, with some even being considered black boxes, or grey boxes (Cortez & Embrechts, 2013), due to their lack of interpretability but in theory they can provide better performance regarding predictive capability. To assess our discrete choice models predictions we use the most popular and easiest to interpret machine learning model (classification tree and ensembles of classification trees) as a benchmark.

The comparison of these methodologies on chapter 6 will serve to validate the use of discrete choice models in predictive problems, employing a branch of machine learning models (classification trees and ensembles) as a benchmark for good performance in this task. This will enhance research's credibility and ensure the methodological suitability of the methods developed in the use case.

2.4 VALIDATION

Predictive models in general, and DM in particular require a validation to assess their performance on generating predictions with new data. The performance metrics can be used as well to compare the predictive power of several models. Cross-validation (Gareth et al., 2021; Jones et al., 1996) is a popular method to provide a systematic evaluation of models' performance by dividing the dataset into multiple datasets and training and testing different combinations of the samples. Two metrics were chosen to assess the models' performance: Accuracy and F1 score (Sasaki, 2015).

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + False\ Positive + True\ Negative + False\ Negative} \tag{10}$$

$$F1\ score = 2 * \frac{\frac{True\ Positive}{True\ Positive + False\ Positive} * \frac{True\ Positive}{True\ Positive + False\ Negative}}{\frac{True\ Positive}{True\ Positive + False\ Positive} + \frac{True\ Positive}{True\ Positive + False\ Negative}} \tag{11}$$

To perform the validation for our use case, we employed 10-fold cross-validation (Fushiki, 2011) due to being a standard in cross-validation, where 10 subsets of the same size were generated, creating training (75% of the sample) and testing (25% of the sample) sets, then we performed each model in each training dataset, and the models and specifications were tested with the testing set, subsequently, we calculated the prediction capability of each model and specifications.

CHAPTER 3. DATA SOURCES

This section justifies the use of tourism as a domain to develop DM models, exploring the datasets and variables for this analysis.

3.1 DATASETS

Pilgrim's Welcome Office Dataset: data related to the sociodemographic variables of the pilgrims, as well as those related to the characteristics of Saint James Way that they have completed, come from the Pilgrim's Welcome Office dataset. This dataset was provided by the Pilgrim's Welcome Office through a confidentiality and exploitation agreement exclusively for scientific purposes. The time interval analysed ranges from 2003 to 2022.

Inglehart's national values**:** compiles data on cultural values and attitudes across multiple countries, enabling the study of social trends and differences over time. In this thesis we use the national aggregated values from year 2021.

Schwartz's national values: gathers information on cultural values and attitudes across nations, facilitating the study of societal variations and trends. In this thesis we use the national aggregated values from year 2006.

Hofstede's national values: compiles data on cultural dimensions across different countries, aiding the study of societal values, behaviours, and variations within and between nations. In this thesis we use the national aggregated values from year 2015.

AEMET: Collect average values of precipitation and temperature for various weather stations in Spain. In this thesis we use the average value from the period 1981-2020.

EES: Collects data on wages, benefits, sociodemographic variables, and disparities across sectors in Spain. In this thesis we use data from 2018 EES to group professionals by approximated salary.

ONET: offers comprehensive occupational information including job tasks, skills, and employment outlooks for various professions. In this thesis we use these data to group professionals by approximated salary.

IGN: is a comprehensive guide providing several maps and ways that we can download to represent GIS data and obtain topographical information.

3.2 DATASETS VARIABLES DESCRIPTION

Pilgrim's Welcome Office Dataset description

Table 1. Pilgrim's Welcome Office Dataset

Name	Description
Way	French (1), Portuguese Coastal (2), English (3), North (4), Silver (5), Portuguese (6), Primitive (7)
Gender	Female (1), Male (0)
Age	Age of pilgrim (n)
Transport mode	Walking (0), Not walking (1)
Profession	Artists (1), Athletes (2), Civil Servants (3), Employees (4), Farmers (5), Homemakers (6), Laborers (7), Liberals (8), Managers (9), Member of religious order (10), Priests (11), Retirees (12), Seafarers (13), Students (14), Teachers (15), Technicians (16), Unemployed
Professional Group	Low income (1), Medium income (2), High income (3)
Motivation	Religious (1), Religious and others (2), Not religious (3)
Autonomy	Andalusia (1), Aragon (2), Asturias (3), Balearic Islands (4), Basque Country (5), Canary Islands (6), Cantabria (7), Castille and Leon (8), Castilla-La Mancha (9), Catalonia (10), Ceuta (11), Community of Madrid (12), Extremadura (13), Galicia (14), La Rioja (15), Melilla (16), Navarre (17), Region of Murcia (18), Valencian Community (19)
Continent	Africa (1), Asia (2), Europe (3), North America (4), Oceania (5), South America (6)
Start (distance)	Stage of departure
Date	Certificate collection date

Source: own elaboration based on Pilgrim's Welcome Office dataset

Inglehart's national values description

Table 2. Inglehart's national values description

Name	Description
Traditional vs Secular	Relate to societies in which religion, authority and family are very important topics
Survival vs Self-expression	Relate to societies shaped by existential insecurity, focused on economic and physical security with constrains on human autonomy

Source: own elaboration based on Inglehart's national values

Schwartz's national values description

Table 3. Schwartz's national values description

Name	Description
Harmony	Considers the degree to which a society supports integration with the surrounding environment
Embeddedness	Considers the degree to which a society manages its relationships as a group
Hierarchy	Considers the importance of maintaining an unequal distribution of power.
Mastery	Considers a society's emphasis on success, ambition, or competitiveness
Affective Autonomy	Considers the degree of desirability for individuals to have a happy independent affective life
Intellectual Autonomy	Considers the degree to which it is desirable for a society that individuals think independently
Egalitarianism	Considers the emphasis of a society on promoting the welfare of others

Source: own elaboration based on Schwartz's national values

Hofstede's national values description

Table 4. Hofstede's national values description

Name	Description
Individualism vs Collectivism	Considers the extent to which society is integrated into groups, as well as its dependence on them
Power Distance vs Closeness	Considers the degree to which certain societies can tolerate inequity or hierarchy in decision-making
Masculinity vs Femininity	Considers the degree to which a society supports traditional social roles
Uncertainty Avoidance vs Acceptance	Considers the degree of acceptance by members of a society for unstructured or repeated situation
Long-Term Orientation vs Short-Term Orientation	Considers the extent to which a society values short-term results or plans for the long-term
Indulgence vs Restraint	Considers a society's preference in satisfying primal instincts

Source: own elaboration based on Hofstede's national values

AEMET description

Table 5. AEMET dataset description

Name	**Description**
Precipitation	Mean precipitation by month (mm)
Temperature	Mean temperature by month (ºC)

Source: own elaboration based on AEMET data

3.3 DESCRIPTIVE ANALYSIS OF THE SAMPLE

We generate a specific analytic section for analysis of the evolution of descriptives of the main variables on Pilgrim's Welcome Office variables, developing a comprehensive examination of them by exploring distribution, trends, interrelationships between variables, relation with the economic reality.

3.3.1 Main characteristics

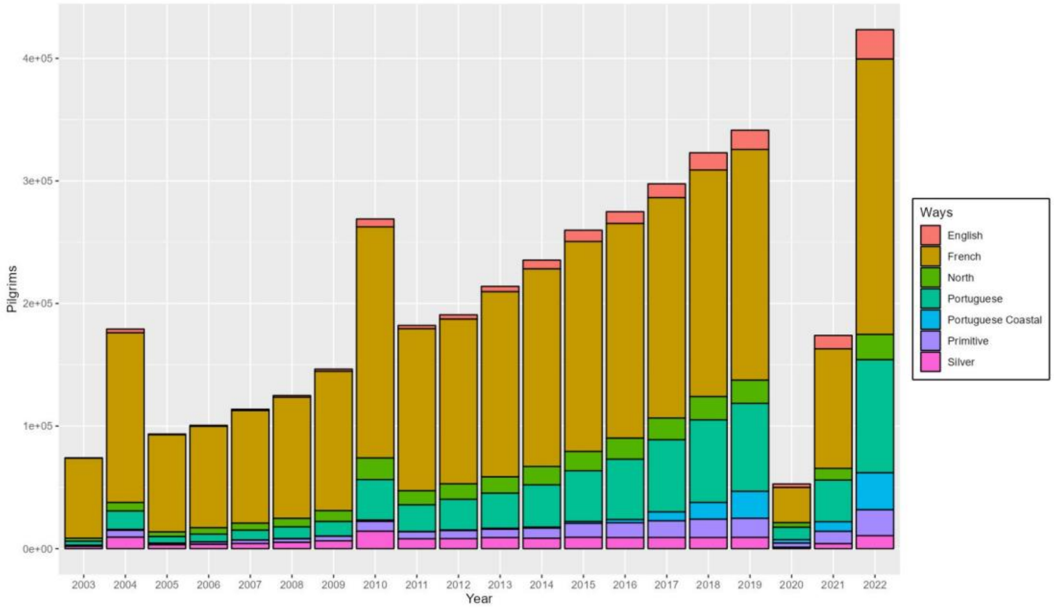
Following the characteristics from the database and outlined in Table 1, we present the main findings of the descriptive analysis, which will justify the selection of variables in subsequent chapters.

Way

The number of pilgrims undertaking Saint James Way, particularly the French Way, has shown a steady growth since 2003, except for the years 2020 and 2021, which were affected by the COVID pandemic health crisis. Throughout the analysed series, a general increase in pilgrim arrivals across all ways can be observed, except for the Silver Way, which appears stagnant and reaches its peak in 2010. Despite all the Ways presenting clear rise in pilgrim arrivals, it is more noticeable the increase on the English, Portuguese, and notably the Coastal Portuguese ways. The latter went from having fewer than 1,000 arrivals before 2015 to becoming the third most frequented way since 2019, recording over 30,000 arrivals in 2022. Overall, it can be observed that the number of pilgrims during Holy Years exceeds by 100,000 those in the surrounding years, this change during Holy Years is attributed to the increased arrival of Spanish pilgrims, usually the target of advertising during the Xacobeo Year. Thus, while at the beginning of the series, the French Way represented 88% of the arrivals in the city of Santiago de Compostela, with occasional relative drops during Holy Years, the significant relative increase in other ways has led, in 2022, to its representation being around 50% of the arrivals. There is a clear trend indicating a further decline in the future, given the remarkable growth of the Portuguese Coastal and Primitive Ways. For the second consecutive time, these ways have surpassed the Northern

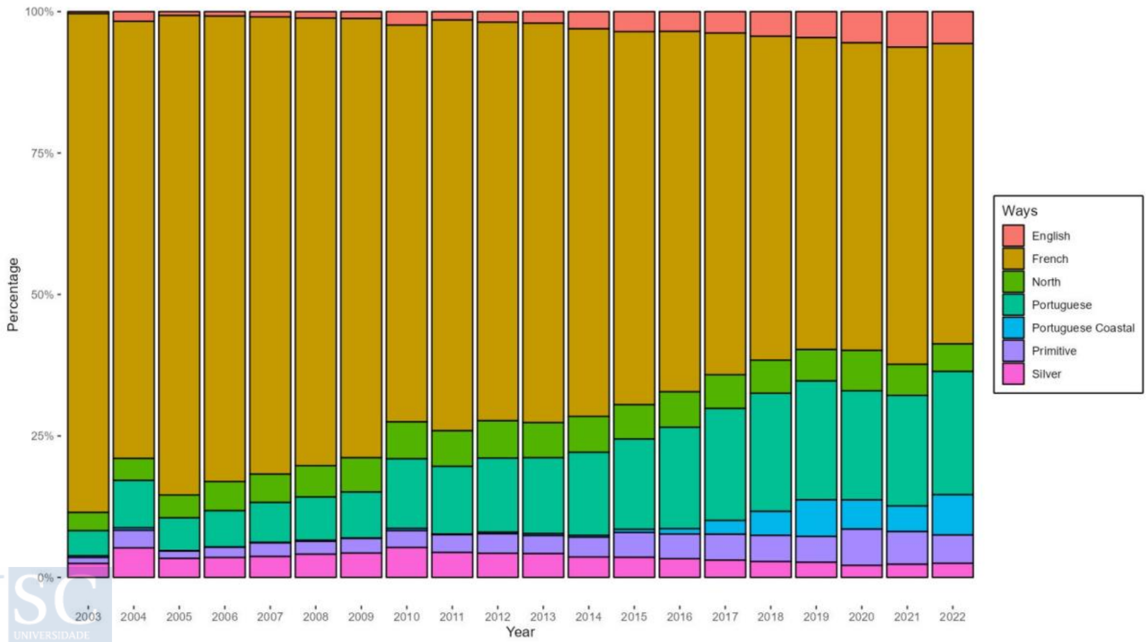
Way, the second way to have been granted the UNESCO World Heritage certificate alongside the French Way.

Figure 3. Number of pilgrims by Way and year



Source: own elaboration based on Pilgrim's Welcome Office dataset

Figure 4. Percentage of pilgrims by Way and year

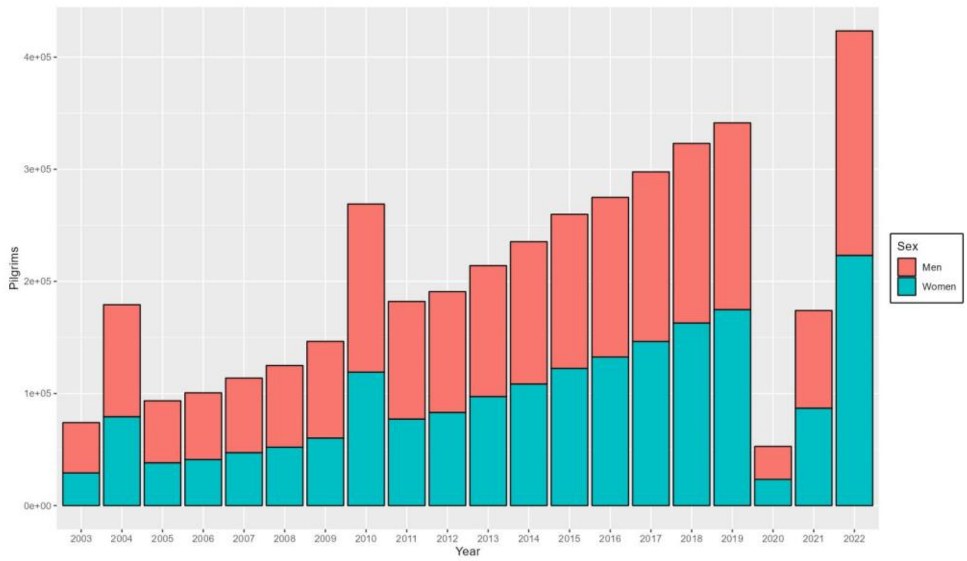


Source: own elaboration based on Pilgrim's Welcome Office dataset

Gender

Over the years, there has been an increase in the percentage of women undertaking Saint James Way. This trend becomes particularly noticeable during the Holy years, where the proportion of women completing Saint James Way surpasses that of men. This holds true for all years except for 2020 and 2021, which were marked by the coronavirus crisis. Since 2018, it has been observed that the number of women has exceeded that of men, potentially influenced by a higher proportion of female foreign pilgrims compared to Spanish ones, however, even when women are becoming the preponderating sex on Saint James Way, both sexes haven growing steadily from 2003 to 2022 as can be seen on Figure 5.

Figure 5. Number of pilgrims by gender and year

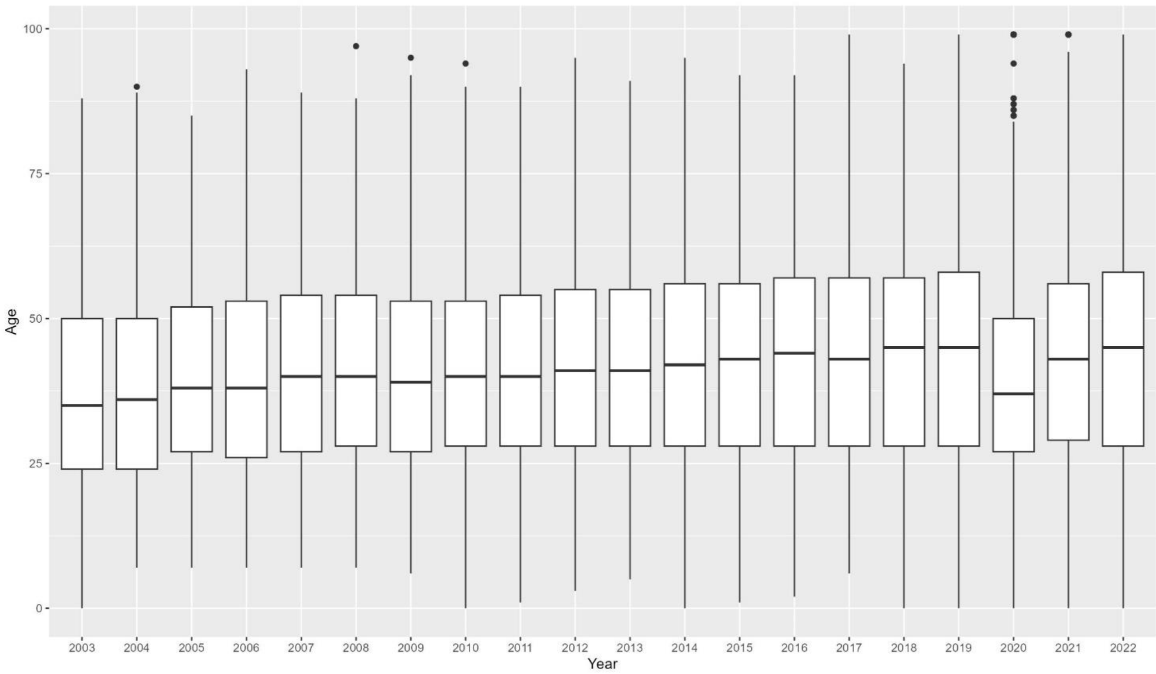


Source: own elaboration based on Pilgrim's Welcome Office dataset

Age

Since the beginning of the series, the median age of individuals completing Saint James Way has increased steadily, along with its variability, except for the years 2020 and 2021 that occurred due to the COVID health crisis, which relatively deterred older individuals from participating.

Figure 6. Boxplot of pilgrims by age and year

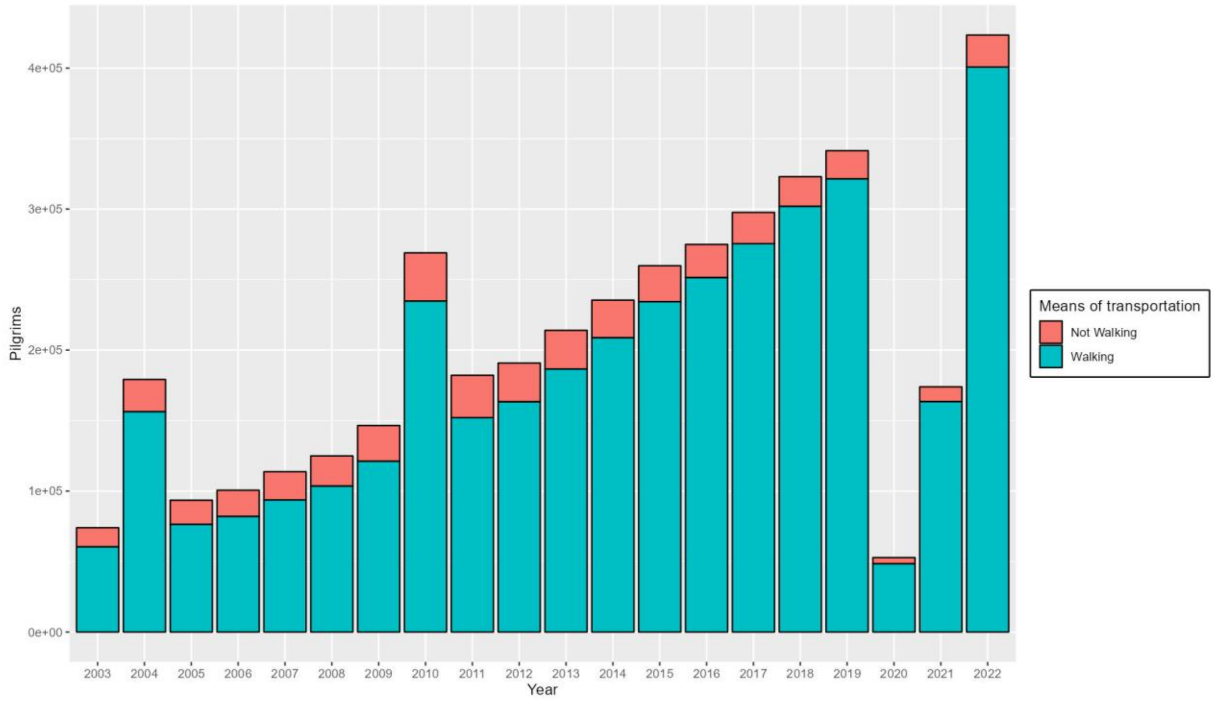


Source: own elaboration based on Pilgrim's Welcome Office dataset

Transport mode

Since 2003, there has been a gradual decline in the arrival of pilgrims traveling Saint James Way by means of transportation other than walking (mostly by bicycle). This decline is particularly notable during the Holy years. It might be due to the peak saturation of pilgrims, as those cycling on Saint James Way prefer not to encounter as many pedestrians. This holds true to such an extent that not only does the proportion of pilgrims using means other than walking decrease, but there is also an absolute decline in numbers since 2010, with the exception of 2022, as can be seen in Figure 7.

Figure 7. Number of pilgrims by means of transport and year

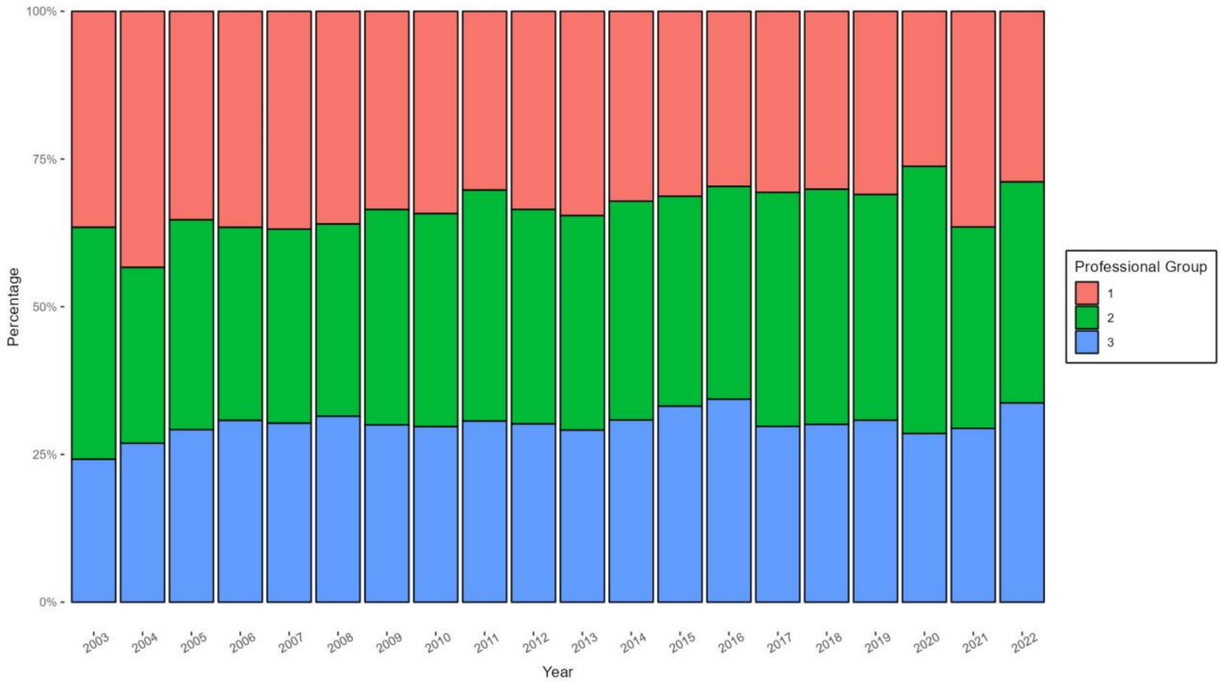


Source: own elaboration based on Pilgrim's Welcome Office dataset

Profession

Throughout the entire series, there is a trend of relative stability in the behaviour of the three professional groups. Initially, there was a higher prevalence of individuals from professional group 1, linked with lower-income individuals. However, as the years progressed, there emerged a greater predominance of individuals from professional groups 2 and 3, which can be associated with higher-income individuals. This shift suggests a transition from group 1 dominance, with high importance of students towards an increased representation of higher-income individuals in groups 2 and 3 by the end of the series.

Figure 8. Percentage of pilgrims by professional group and year

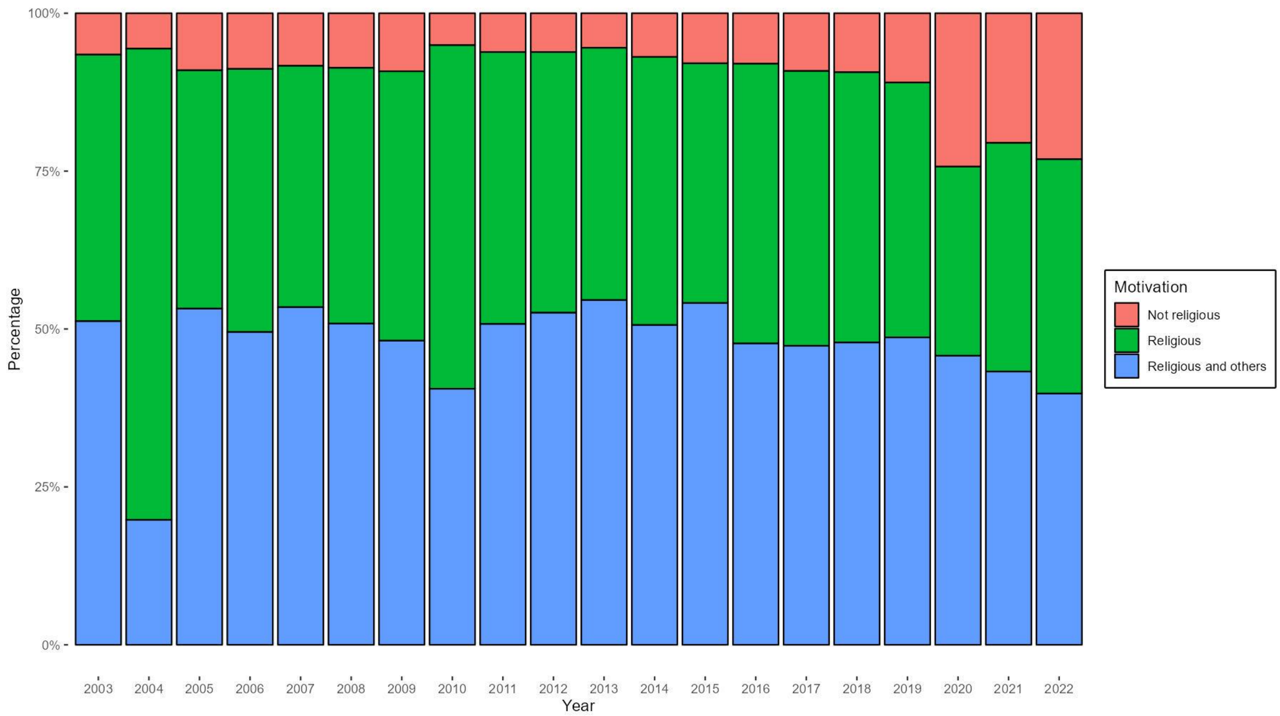


Source: own elaboration based on Pilgrim's Welcome Office dataset

Motivation

There has been a reduction in the significance of religious motives and others, gradually equating the importance of non-religious motives with that of religious ones. This decline in religious and other motives is linked to a relative decrease in the arrivals of religious workers and students, as well as a relative increase in pilgrims coming from less religious countries, bringing motivations that are less religious than at the beginning of the series. It can be observed that during a Holy year, pilgrims tend to respond more frequently that their motive is religious. This trend is noticeable in 2004 and 2010; however, it is not as apparent in 2021 and 2022. Hence, it is evident that over time, Holy years have become less religiously oriented, and their influence in relative terms has diminished.

Figure 9. Percentage of pilgrims by motivation and year

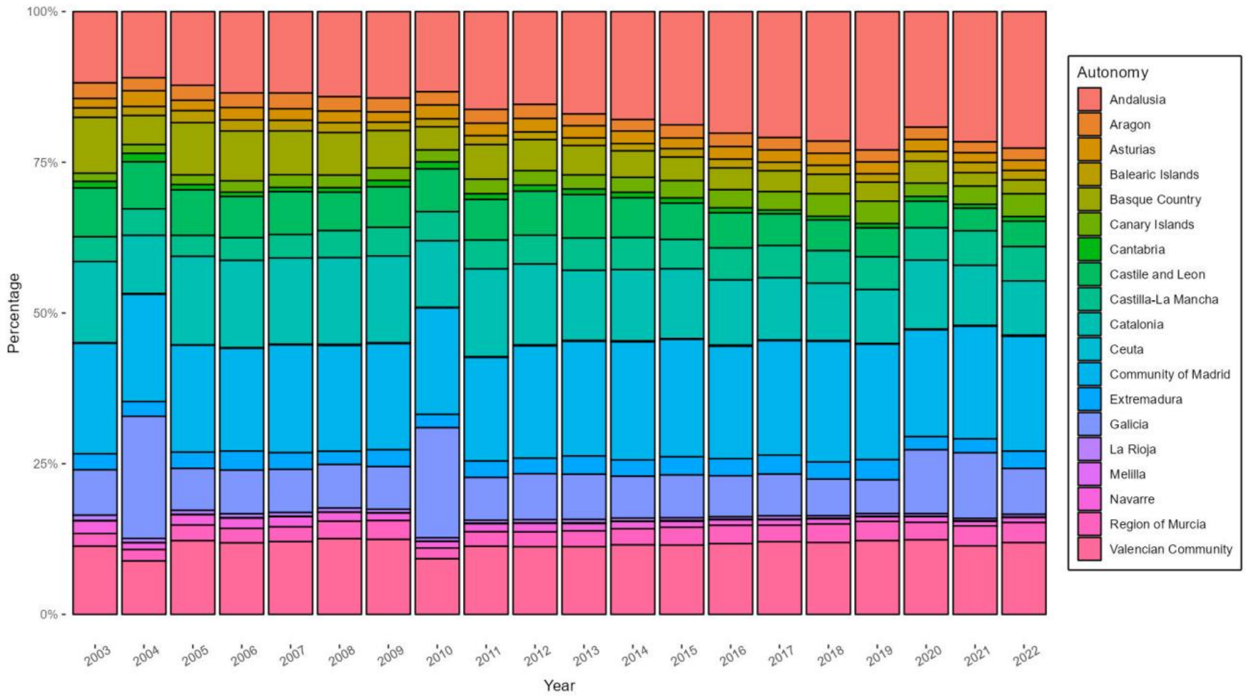


Source: own elaboration based on Pilgrim's Welcome Office dataset

Origin of Spanish pilgrims

Analysing the evolution of Spanish pilgrims, it can be observed that most regions have demonstrated a relatively stable pattern over the years. Notably, there has been a significant increase in the relative importance of Andalusia, while the Basque Country and Castile and Leon have experienced a decline. During Holy years, the percentage of Galician pilgrims deciding to undertake Saint James Way notably tripled, although in the last Holy year, their percentage only doubled compared to the total of all autonomous communities.

Figure 10. Percentage of pilgrims by autonomous community and year

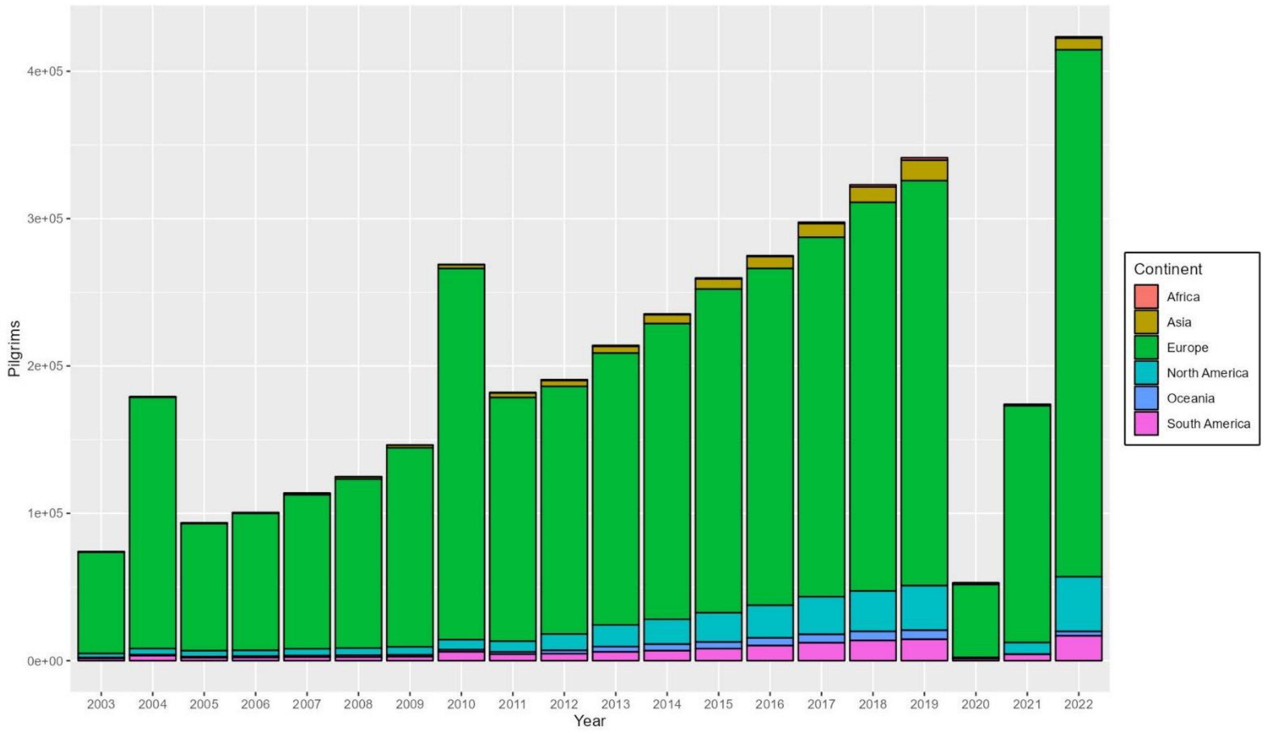


Source: own elaboration based on Pilgrim's Welcome Office dataset

Continent

Since the beginning of the historical series, it has been evident that most pilgrims completing Saint James Way are Europeans. However, a notable increase in pilgrims from other continents such as Asia, North America, and Latin America has been observed, multiplying their numbers by over 15 times.

Figure 11. Number of pilgrims by continent and year

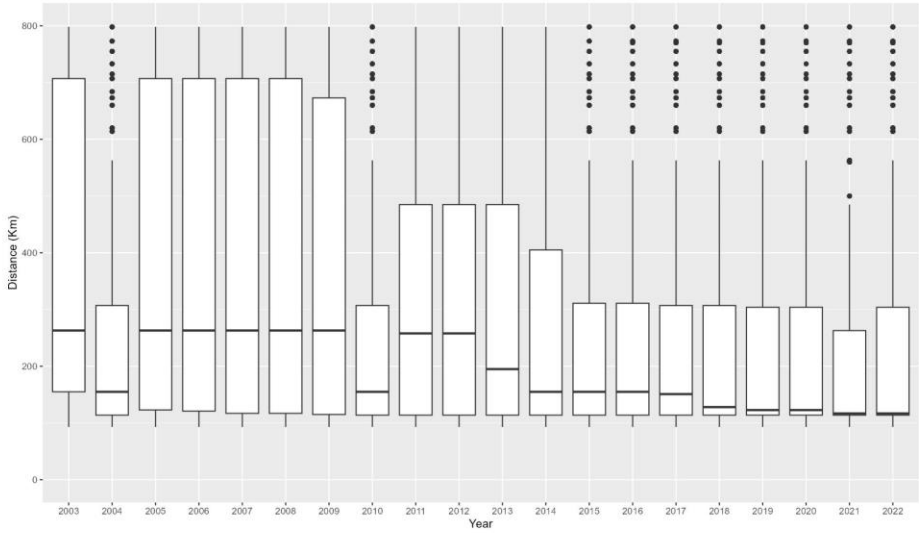


Source: own elaboration based on Pilgrim's Welcome Office dataset

Distance

Analysing the evolution of the distance travelled by pilgrims, it can be observed that in Holy Years such as 2004, 2010, or 2021-2022, the distance covered by pilgrims decreases in comparison with baseline years. This could be due to the composition of pilgrims or that in those years, there are people who would not be willing to complete a long journey but rather wish to receive the jubilee by entering through the Holy Door at the Cathedral of Santiago de Compostela. Not only does the arrival of pilgrims decrease across all quartiles, but the median also decreases in these Holy Years to be just above the minimum distance required to obtain the Compostela certificate. In addition to the pattern regarding Holy Years, we can also observe that the trend is towards shorter ways for Saint James Way over time, while 25% of people who started longer ways in 2003 walked over 700 km, in 2022 this figure was 300 km.

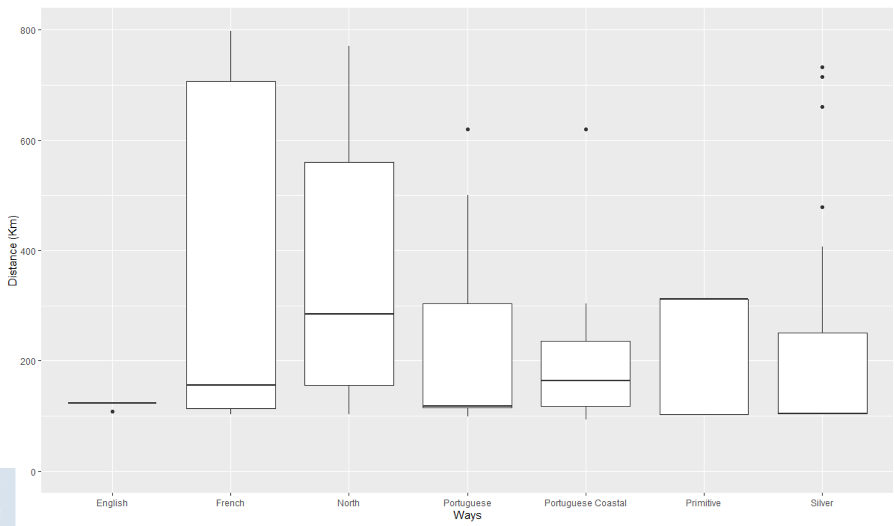
Figure 12. Boxplot of pilgrims by distance and year



Source: own elaboration based on Pilgrim's Welcome Office dataset

In terms of the distances covered by pilgrims along each way, we detect a distinct pattern on English Way, where there are only two stages exceeding 100 km, so that they are the only ones that can be selected. Among the other ways, the French and Northern Ways exhibit higher variability, evident in their interquartile range. Meanwhile, the Northern Way and the Primitive Way are the paths where pilgrims cover the longest distances as can be seen on the median distance, with over 280 km, coinciding with ways known for their greater technical difficulty.

Figure 13. Boxplot of pilgrims by distance and Way



Source: own elaboration based on Pilgrim's Welcome Office dataset

CHAPTER 4. SUPRA-NATIONAL CULTURE AND TOURIST'S DECISION-MAKING

4.1 INTRODUCTION

The role of culture in shaping tourists' decision-making (DM) has been a popular topic of research in the field of tourism. It has been reported that culture affects the DM process by conditioning: (1) the likelihood of revisiting certain tourist destinations based on trip satisfaction (Crotts & Erdmann, 2000; Huang & Crotts, 2019; Li et al., 2010), (2) the choice of a specific tourist destination over others at any phase of the decision-making system that alters consumer behaviour (Hagag et al., 2015; Hsu et al., 2013; Litvin et al., 2004; Ng et al., 2007; Qian et al., 2018; Woodside et al., 2011; Yang et al., 2019), (3) the perception of risk associated with a country, which generally negatively affects the intention to visit a destination (Kozak et al., 2007; Yang & Wong, 2012), (4) or the competitiveness of tourist destinations, altering the probability of choosing one destination over another (Kumar & Dhir, 2020).

Four concepts of culture have been reported in the literature review performed by Ng et al. (2007) on the measures that are likely to impact tourist's destination choice: tourist's national culture, tourist's internalized culture, destination's culture and cultural distance between tourist and destination. The first one is the most popular interpretation in other sciences and considers culture as a national phenomenon. The second one assumes that individuals personalize cultural values based on their specific needs. The third one focuses on the culture of tourist's destination and how this factor both impacts the perception of the destination and persuades tourists to travel. The last one has been studied extensively in the tourism domain as a proxy of cultural gaps between countries, where gaps represent negative factors for tourism, such as risk and uncertainty.

In the push-pull model framework, the study of decision-making process has consisted on the analysis of the four aforementioned culture definitions, we propose a fifth cultural definition, which makes the contribution of this chapter threefold: (1) by proposing culture as a supra-national concept (Akaliyski et al., 2021), (2) by using supra-national cultures as push factors in the push-pull theory (Crompton, 1979; Sirakaya & Woodside, 2005; Uysal & Jurowski, 1994), and (3) by applying choice models (Mcfadden, 1980), to formalize the push-pull theory. The concept of culture as a supra-national phenomenon is a new approach in the field of tourism. Based on the evidence found in other fields, we believe that the application of the supra-national culture can be a key explanatory factor of tourist's behaviour. In summary, the objective of our work is to demonstrate whether supra-national cultures influence individual decisions or not by means of testing push-pull models.

The chapter is organized as follows. Section 2 provides a literature review. Section 3 overviews our approach, setting the Inglehart-Welzel cultural clusters as our supra-national cultures. Section 4 describes the methodology, including the case-study, the data, and the choice models used to formalize the push-pull approach. Section 5 provides the results. Section 6 discusses the findings and the overall contributions. Finally, Section 7 stablish the conclusions.

4.2 LITERATURE REVIEW

4.2.1 Culture in tourism

As observed in the literature review concerning the definition of culture (Baldwin et al., 2006), this concept is complex and has been studied by a wide variety of sciences. Within the tourism realm, the impact of culture has been extensively studied, and the review performed by Ng et al. (2007), enable us to categorize this extensive body of literature. So, four cultural concepts have been studied:

Tourist's national culture. Culture is understood as a collective phenomenon (Hofstede et al., 2005) consisting of the set of beliefs, values, knowledge and norms accepted by a social group (Schwartz, 1999; Inglehart, 1997). National culture has been used in the fields of DM and tourism to compare the decisions of tourists from two different countries (Crompton, 1979; Kozak, 2002; H. Liu et al., 2018; Ng et al., 2007). It has also been used to explain the DM styles that tourists may apply when making choices (Correia et al., 2011).

Tourist's internalized culture (individual level). Individuals personalize cultural values based on their specific needs. Individual culture therefore affects how tourists select their destinations and the characteristics of their trips (Ng et al., 2007). Tourists' internalized culture can be measured by their individual values or beliefs as defined by Hofstede, Schwartz or Inglehart.

Destination's culture. The culture of the destination to which the tourist travels affects the way international tourists perceive the nation in which their destination is located (McKercher & Cros, 2003; Rinschede, 1992). This factor persuades tourists to make the trip (Qian et al., 2018).

Cultural distance. Cultural distance has been studied extensively in the literature as a proxy of cultural gaps between countries. Such gaps introduce risk and uncertainty, with negative implications for tourism. This distance is commonly measured in two ways. The first is the cultural distance that tourists perceive, measured by the values they report on a Likert scale (Bogardus, 1925; H. Liu et al., 2018; Ng et al., 2007). The second is the difference between the cultural values of national cultures (H. Liu et al., 2018; Manosuthi et al., 2020; Yang et al., 2019). One of the most frequently used variables to estimate cultural distance is Hofstede's (1989) dimension of the Uncertainty Avoidance Index (UAI) (Crotts, 2004; Kozak et al., 2007; Qian et al., 2018; Yang et al., 2019; Yang & Wong, 2012). Other studies use Schwartz's and Hofstede's cultural values (Hsu et al., 2013; Manosuthi et al., 2020; Yang et al., 2019). In short, cultural distance inversely affects tourists' DM process (Crotts, 2004; Hsu et al., 2013; Litvin et al., 2004; H. Liu et al., 2018; Pizam & Sussmann, 1995) and is thus inversely related to tourist flows (H. Liu et al., 2018; Manosuthi et al., 2020; Zhang et al., 2019). This effect becomes less negative, however, both over time (Yang et al., 2019) and when the main motivation for visiting a country is to learn about its heritage (H. Liu et al., 2018).

4.2.2 Push-pull model in tourism

The DM process is broad and complex, and has been studied in various domains (Banerjee, 1992; Charles et al., 2013; Jarvenpaa et al., 1999; Kahneman & Tversky, 1984). In the field of tourism, the most popular paradigm for explaining why do tourists make decisions is the push-pull model (Yiamjanya & Wongleedee, 2014), considering this theory as essential to explain why individuals engage in touristic activities (Crompton, 1979; Dann, 1976; Pestana et al., 2020; Sirakaya & Woodside, 2005; Uysal & Jurowski, 1994). Since the model's inception, pull factors included all the destination's features, and push factors the tourist's needs and desires, however, extensive study from different points of view has refined and organised it, leading to its development through several phases (Uysal et al., 2008).

In its original phase, the push-pull theory argued that push and pull factors did not interact with each other (Crompton, 1979; Dann, 1976). This view has evolved over time, authors such as Cha et al. (1995) and Uysal and Jurowski (1994) proposed that tourists travel only when push and pull factors interact. That is, for tourists to be attracted to a destination attribute, they must evaluate that destination positively. Subsequently, Sirakaya and Woodside (2005) and Uysal et al. (2008) proposed extension of the push-pull classification of factors that included push motives among the internal variables of the individual and pull motives within the external variables of the individual. Since internal variables would thus include push motives related to needs, socio-psychological motivations, or desires of individuals (intrinsic to the individual), it is impossible for tourist destinations to drive the creation of push motives. Push motives (e.g., escape, rest, relaxation, prestige, health, adventure and social interaction) are needs and desires that encourage tourists to travel, while destinations have a set of desirable characteristics that attract them (Uysal & Jurowski, 1994). External variables, on the other hand, include pull motives related to the characteristics of each touristic destination and thus extrinsic to the individual. Whether perceived or real, these attributes attract tourists to a destination. The following pull motives are often used to explain the DM process: cuisine, history, natural resources, destination culture, availability of various activities, presence of friends and family (Correia et al., 2013; Pestana et al., 2020; Uysal & Jurowski, 1994).

4.2.3 Choice models in Tourism

The application of choice models in the field of tourism as tools to describe the DM process can be used in a variety of phenomena related to individual’s behaviour, as can be seen in a classic literature review performed by Crouch & Louviere (2001) or the newest literature review performed by Li et al. (2023), realizing that there are four main types of variables to explain:

Transportation mode: They are used to study how tourists choose among the different modes of transportation available for traveling to the destination or for getting around within the tourist destination (Cheng et al., 2018; Iseki et al., 2018).

Products and services: They analyse how tourists choose the products, services, and experiences they decide to purchase before, during, or after engaging in a tourism-related activity (Ismoilov, 2017; Jung et al., 2015; Kim & Park, 2017; Tung & Soo, 2004).

Destination: Revolve around analysing how tourists choose among different cities, countries, regions, or specific attractions (Lacher et al., 2013; Masiero & Qiu, 2018; Oppewal et al., 2015).

Travel packages: They analyse how tourists design their travel itineraries, or how they choose among various plans created by travel agencies (Walters et al., 2019).

4.3 OUR PROPOSAL

4.3.1 Theoretical approach

Three main points: (1) proposing culture as a supra-national concept (Akaliyski et al., 2021), (2) using supra-national cultures as push factors in the push-pull model (Crompton, 1979; Sirakaya & Woodside, 2005; Uysal & Jurowski, 1994), and (3) using choice models (McFadden, 1973; Train, 2009) to formalize the push-pull model.

The concept of supra-national culture has been theoretically used since the pioneering work of Haire et al. (1966), but it wasn't until the 1980s, with the research by Ronen & Shenkar (1985), that this supra-national culture was quantified. This quantification involved the aggregation of countries' demographic, physical, economic, or political characteristics (Akaliyski, 2017).

Various authors have investigated the national culture of tourists and its influence on their DM system (Correia, 2011) in tourism. However, to date, we have not encountered any studies examining how tourists are impacted by the supra-national context, which we refer to as "supra-national culture" throughout this chapter. Therefore, we will draw upon authors from fields beyond tourism to justify its inclusion. Thus, authors from cross-cultural analysis domain, such as Akaliyski (2021) and Michael Minkov (2012), have justified practically and theoretically the potential of aggregation of individuals' cultural values, initially at the country level and subsequently at the supra-national level. They sustained this through a practical demonstration by studying the variability in people's cultural values, illustrating that most of the variability in individuals' responses stemmed from their belonging to different supra-national cultures.

Despite the wide array of methods available for aggregating cultural characteristics of individuals, this chapter chooses to utilize the conceptual framework developed by Inglehart (1997), due to its stability over the years, as evidenced in the successive waves of the World Values Survey. In his subsequent works, Inglehart (2000), using a variation of Huntington's (1993) classification of world cultures, has consistently demonstrated that many patterns of behaviour on people's DM process can be explained based on the culture to which they belong. Thus, we select this supra-national construct to test our hypothesis.

Akaliyski et al. (2021) demonstrates that a significant portion of variability among individuals can be explained based on their supra-national cluster, empirically confirming Inglehart and Baker (2000, pag 49) assertion that "Economic development tends to transform a given society in a predictable direction, but the process and path are not inevitable. Many factors are involved, so any prediction must be contingent on the historical and cultural context of the society in question".

Conducting a brief literature review to justify the use of supra-national cultures, we could find two types of articles: On the one hand we found several articles that use Inglehart's clusters to study the influence of the pertaining to a different Inglehart cluster on specific political or sociological factors, domains in which these constructs were created, on the other hand, we found articles that focused on aspects related to the focus of our chapter, where it was observed that Inglehart's clusters or supra-national clusters affect economic or business-related factors. Analysing the articles in political science and sociology: Sandholtz & Taagepera (2005) build upon Inglehart's conceptual framework of clusters to demonstrate that the culture of large groups of countries, specifically communism as a supra-national cultural factor, increases corruption; Rossteutscher (2004) employs his own cluster to explain the social or political preferences of individuals, comparing it with Inglehart's, both being useful for explaining the phenomenon; Welzel (2013) examines how supra-national cultures evolve over time in relation to their emancipatory and secular values, concluding that this evolution varies among them. Analysing the articles related with economics and business: In the field of social capital Portela et al. (2013) construct clusters of social capital that encompass a significant part of cultural dimensions, noting that pertaining to a different social capital cluster resulted in individuals having a different level of subjective well-being; Gaston-Breton & Martín (2011) create clusters based on Inglehart's and Eurobarometer surveys to groups of countries, demonstrating that market segmentation by cultural clusters should be considered when designing effective marketing campaigns, instead of relying solely on market potential studies; Popov et al. (2018) studied the impact of the autonomy of society members and gender equality on social entrepreneurship, with the former being related to self-expression values and the latter to secular-rational values. Subjective wellbeing can be a precursor to satisfaction, as the second author emphasizes the importance of segmenting consumers by cultural clusters, and the third article suggests that the Inglehart cluster to which individuals belong may influence the likelihood of engaging in social entrepreneurship. Thus, these three recent articles justify the use of supra-national clusters in the context of DM in tourism. We validate Inglehart's cluster theory in the context of DM thus opening an important line of analysis in the tourism domain.

Hence, authors argue that it is advisable to incorporate the country's culture and Inglehart supra-national culture into tourism marketing, allowing for a targeted focus on specific cultures, enhancing the user experience of products and services for tourists. Otherwise, implementing public policies or marketing campaigns not aimed to cultural groups by destination managers and hospitality firms, may be ineffective due to cultural differences (Hagag et al., 2015; Hsu et al., 2013; Kumar & Dhir, 2020).

We propose to include supra-national cultures in the DM process by means of the push-pull model. Specifically, we consider supra-national cultures as push factors due to its influence on the psychological profile of people, which determines the desire to relax, the desire to escape, the importance of the family, etc., which are the internal factors that push tourists to travel (Uysal & Jurowski, 1994).

4.3.2 Formal approach

The push-pull model is implemented as a choice model, and more specifically as a multinomial choice model. This formalization enables precise study of tourists' DM system and thus collection of pull motivations (alternative's characteristics). Based on random utility models (McFadden, 1980), this technique enables the inclusion of the characteristics of each alternative while analysing several alternatives simultaneously. The proposed model also enables us to analyse the tourist preferences identified, thus overcoming the limitation of the absence of psychological features in the dataset.

4.4 METHODOLOGY

4.4.1 Case study: Saint James Way

We choose Saint James Way as case study because of its relevancy on the tourism sector of the autonomous community of Galicia in Spain. We are going to test if supra-national culture as a push factor might explain the choice based on the attributes of each way. The data cover the period from 2003 to 2020 and compile information on more than 3,470,000 pilgrims from 59 countries who made the pilgrimage to Santiago de Compostela via one of the seven major Ways, named: French, Portuguese Coastal, English, North, Silver, Portuguese, and Primitive. After filtering the data for missing information, we obtained a dataset from 3,180,856 pilgrims.

4.4.2 Push factors: supra-national culture and pilgrim variables

4.4.2.1 Supra-national culture

Using supra-national variables enables us to simplify international comparisons by dividing the world into large areas of study. This approach is of great empirical utility, as eliminates the need to include country control variables to an explanatory construct by allowing us to analyse and compare individuals' and alternatives' characteristics in terms of culture. In Table 6, one can observe the two cultural values from Inglehart's supra-national culture definition.

Table 6. Inglehart cultural values and description

Factor	Description
Traditional vs Secular	Relate to societies in which religion, authority and family are very important topics
Survival vs Self-expression	Relate to societies shaped by existential insecurity, focused on economic and physical security with constrains on human autonomy

Source: own elaboration based on Inglehart-Welzel Cultural Map

The cultural factors provide us with 8 different cultural clusters, the main explanatory variable to which the individual's country belongs, identified using Inglehart's World Cultural

Map scores. The 2018 Inglehart-Welzel clusters in our dataset, analysing separately the destination country would be as follows:

West and South Asia (1): Malaysia, Singapore, Israel, Philippines. Mid secular and mid self-expression

Catholic Europe (2): Andorra, Malta, Latvia, Estonia, Croatia, Slovenia, Lithuania, Slovakia, Hungary, Austria, Belgium, Czech Republic, Poland, France, Portugal, Italy. Mid-high secular and mid-high self-expression.

Latin America (3): Paraguay, Guatemala, Costa Rica, Peru, Puerto Rico, Ecuador, Uruguay, Venezuela, Colombia, Argentina, Mexico, Brazil, Chile. Low secular and mid self-expression.

Orthodox Europe (4): Ukraine, Bulgaria, Romania, Russia. Mid secular and low self-expression.

Protestant Europe (5): Iceland, Norway, Finland, Switzerland, Sweden, Denmark, Netherlands, Germany. High self-expression and high secular.

Confucian (6): Taiwan, China, Korea, Japan. Mid self-expression and high secular.

Destination Country (7): Spain. Mid-high secular and mid-high self-expression.

English-Speaking (8): New Zealand, Canada, Australia, Ireland, United Kingdom, United States. High self-expression and mid-high secular.

E.g. a Protestant individuals would assign little importance to values related to religion, authority, and family, while placing great importance on personal autonomy. Conversely, individuals from Latin America would exhibit the opposite trend, prioritizing values associated with religion, authority, and family while placing higher emphasis on personal security.

4.4.2.2 Pilgrim's variables

The pilgrims are described using the following variables or attributes:

Professional group: (1) farmers, homemakers, students, seafarers, labourers, unemployed, members of religious orders and priests; (2) artists, athletes, employees, civil servants, and teachers; and (3) managers, retirees, liberals, and technicians. Group 1 is assigned the lowest values and Group 3 the highest. With proxy wages being estimated according to "ONET" (Development, 2022) and "EES" (INE, 2022), the former a database of occupational characteristics and information on worker requirements and the latter a database of the labour characteristics of employed people in Spain.

Means of transportation: walking and not walking (biking, sailing, horseback, or wheelchair).

Finally, we included pilgrim's gender, due to evidence of its effects on pilgrims' DM in cultural choices (Kozak et al., 2007; Lefebvre & Franke, 2013; H. Liu et al., 2018; Manosuthi et al., 2020).

4.4.3 Pull factors: Way variables

To fill in the push-pull framework, we must introduce the pull factors of the model, in our case the characteristics describing the ways. In what follows, a general description of the seven main ways:

French: It is the most popular way, crossing vast forests and mountains, passing through many historically significant cities through inner Spain, and it is one of the two that obtained the World Heritage credential.

Portuguese coastal: It is a way that largely runs along the North Portuguese and Galician coast, passing through significant cities and completing an itinerary of low technical difficulty.

English: It is a lesser-known way that begins its itinerary through urban and coastal areas and finish crossing forests.

North: It is a way that stretches along the northern coast of Spain, being a path of high technical difficulty and oceanic climate; it is the second way to have obtained the World Heritage credential.

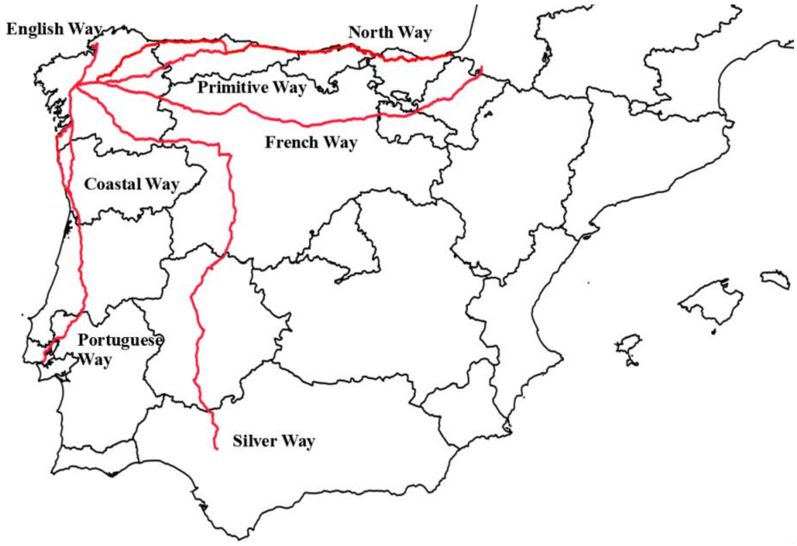
Silver: It is a way that crosses Spain from south to north, traversing areas with high technical difficulty, sweltering summer climate, and cities with historical value.

Portuguese: It is a highly popular way that runs through the interior of Portugal and Galicia, following paths of low technical difficulty and historical cities.

Primitive: It is a less popular way that runs through the interior of Galicia and Asturias, through mountainous areas with technical terrain and low levels of urbanization.

We plot the main ways to reach Santiago de Compostela through Saint James Way routes over a map of the Iberian Peninsula on Figure 14 with data from IGN (2023).

Figure 14. Map of main Saint James Ways through Iberian Peninsula



Source: own elaboration based on IGN data

The descriptions of the Ways can be summarized as a set of characteristics or attributes:

World Heritage Status: UNESCO grants special World Heritage credentials to places with exceptional artistic and cultural value. The French and North Ways have achieved this status and might therefore be more attractive, increasing the probability of pilgrims choosing them. World Heritage Status is a binary variable indicating whether the Way has the credential.

Popularity: Evaluation of the Ways' popularity depends greatly on the pilgrim's mindset. Popularity can be considered as positive in terms of safety (Gil et al., 2017; Powell et al., 2017) or socialization, or as negative when pilgrims prefer to focus on introspection. We measured popularity by number of people who walked the path between 2003 and 2020, dividing the variable into three categories. Paths travelled by more than 500,000 pilgrims take the value High, paths travelled by 150,000 - 500,000 the value Medium, and paths travelled by fewer than 150,000 the value Low.

Intensity: When we consider the pilgrimage as a physical activity, intensity of the Way is an important feature conditioning the DM process. Weather factors (Hall & Ram, 2021) contribute to this variable, such as temperature (Smith, 1993), precipitation (Amelung et al., 2007), both based on AEMET (2023) data as well as the difficulty of the stages (based on average length and slope and deduced from Gronze (2023)). We used a set of equations to estimate the variable intensity (for equations, see ANNEX 1).

Table 7 summarises the values the attributes popularity, intensity, and World Heritage for each of the seven routes of Saint James Way.

Table 7. Alternative-specific variable values for each Way

Alternatives / Variables	French	Portuguese Coastal	English	North	Silver	Portuguese	Primitive
High Popularity	1	0	0	0	0	0	0
Medium Popularity	0	0	0	0	0	1	0
Low Popularity	0	1	1	1	1	0	1
Intensity	1	0	0	1	1	0	1
World Heritage	1	0	0	1	0	0	0

Source: own elaboration

4.4.4 Choice models

Choice models study individuals' DM based on revealed or stated preferences in a particular context (Ben-Akiva& Lerman, 1985). Following this line of research, we analyse the impact of cultural clusters in a choice model to determine how belonging to a specific cultural area influences individuals' preferences for different ways/alternatives.

Choice models are one of the most used tools in creating a recommendation system—a computer-based structure that supplies valuable suggestions especially suitable for choosing touristic ways or products and services.

Empirically, a choice model can be described by considering a decision-maker c_n:

$$C = c_1, \dots, c_n, \dots, c_N \tag{12}$$

who chooses an alternative a_j from a choice set A, composed of all the alternatives.

$$A = \{a_j, J = 1\,, \dots, j, \dots, J\} \tag{13}$$

The most popular form of probability is the logit model (McFadden, 1973), where following from this point onwards we use Train (2009) notation, the representative utility is:

$$V_{nj} = \beta X'_j \tag{14}$$

where $\beta$is a vector of unknown parameters and X_ja matrix of regressors. We can write the closed-form expression for the logit choice probability as:

$$P_{ni} = \frac{e^{V_{nj}}}{\sum_{j=1}^{J} e^{V_{nj}}} \tag{15}$$

If we substitute the values for representative utility in the logit choice probability for model (I) equation, it becomes:

$$P_{nj} = \frac{e^{\beta * X'_{nj}}}{\sum_{j=1}^{J} e^{\beta * X'_{nj}}} \tag{16}$$

Alternative a_j is chosen when the corresponding choice probability P_{nj} is higher than the choice probabilities of the other alternatives a_j in A, resolving the maximization of response probabilities for a logit model:

$$Max\ (P_n(a_j = 1), \dots, P_n(a_j = j), \dots, P_n(a_j = J)) \tag{17}$$

For our choice problem, the decision-maker is a pilgrim n choosing way a_j from the choice set A composed of all seven ways described in Section 4.4.3. The observed utilities V_{nj} are specified as linear in the set of explanatory variables. Thus, the representative utility for multinomial logit model with alternative specific variables becomes:

$$V_{nj} = \beta_n * X'_{nj} + \alpha_j * y_j \tag{18}$$

where β continues to be a vector of parameters for the individual n, $X_{nj}$the matrix of observed features of pilgrim n, y_j the vector of observed features of Ways a_j, and α_j the vector of parameters of the Way (alternative).

Assuming we have J possible outcomes as independent variables (Ways), X'_{nj} denotes the set of variables and k each of the observations. We must run N-1 logit estimations, such that the estimate α_i when J=1 is inferred from the others (Wooldridge, 2010) would denote the alternative-specific variables that characterize the outcomes (Train, 2009). Thus, response probabilities for multinomial logit model with alternative-specific variables takes the form:

$$P_{nj,}\, j = 0,1,2, \dots, J, = \frac{e^{\beta_n * X\prime_{nj} + \alpha_j * y\prime_j}}{1 + \sum_{j=1}^{J} e^{\beta_n * X\prime_{nj} + \alpha_j * y\prime_j}} \tag{19}$$

where the set of β_{nj} and α_n are the model coefficients to be estimated by fitting Equation 19 to the dataset of choices described in section 4.4.

To fit the multinomial model, we resolve to the use of a maximization of the log-likelihood function like the one presented on Equation 20, where we need to employ a Newton-Raphson technique for numerical optimization to find the values that maximize said function, and where the probabilities have the form of Equation 19.

$$l(P) = log(N! \prod_{j=1}^{m} \frac{p_n^{x_n}}{x_n!}) = log\, N! + \sum_{j=1}^{m} x_n \, log\, p_n - \sum_{j=1}^{m} log\, x_n! \tag{20}$$

This involves setting the derivatives (Lagrange multiplier) of the log-likelihood function while imposing the constraint that the sum of all probabilities must be equal to 1, obtaining at the end of the optimization process the probability distribution that maximizes the likelihood.

We specify two specific choice models. Model I seek to explain the effect of the variable cultural cluster on choice probability. In this case, we fit seven models to estimate choice probability for each Way in A. Each model is a simple binary model that considers only two probabilities: probability of choosing vs. probability of not choosing Way a_j. Equation 21 represents the utility considered for Model I, where we need only substitute the values of the regressors to calculate the model:

$$V_{nj} = \beta_0 + \beta_1 x'_{cultural\ cluster} + \beta_2 x'_{man} + \beta_3 x'_{professional\ group} + \beta_4{}' x_{not\ walking} \tag{21}$$

Model II focuses on understanding the distribution of choices for each cluster. We thus fit eight models, one for each cluster. In this case, each model is a multinomial logit model with alternative-specific variables, as it considers all seven ways as possible alternatives. Equation 22 represents the utility considered for Model II. We thus need only substitute the values of the regressors to calculate the model:

$$\begin{aligned} V_{nj} = \beta_0 + \beta_1 x'_{man} + \beta_2 x'_{professional\ group} + \beta'_3 x'_{not\ walking} + \alpha_1 + \alpha_1 y_{Highpopularity} \\ + \alpha_2 y_{Mediumpopularity} + \alpha_2 y_{intensity} + \alpha_3 y_{World\ Heritage} \end{aligned} \tag{22}$$

4.4.5 Software tools

We used software R version 4.2.3 for the estimations and the base package version 4.2.3 for Model I. For Model II, we used the package mlogit version 1.1.1 (Croissant, 2020).

4.5 RESULTS

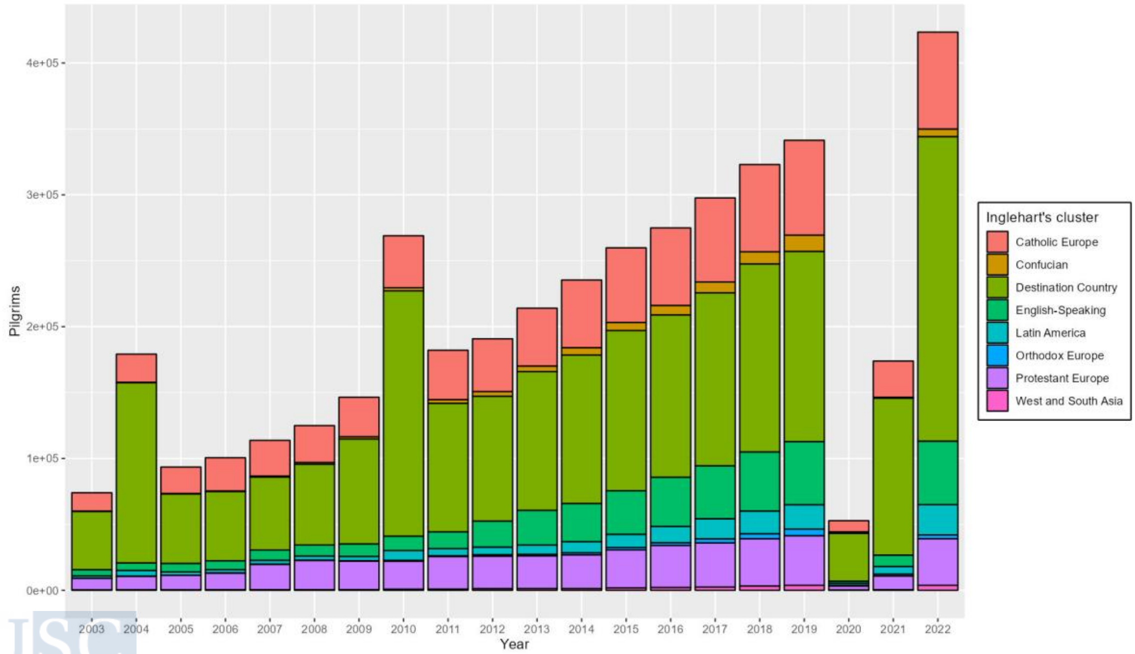
In this results section, we delve into the descriptive analysis of Inglehart-Welzel cultural construct in relation with pilgrim's variables. Additionally, a thorough examination is conducted to test the effect of cultural clusters on the choice of Way, studying how preferences vary based on the supra-national cluster. Finally, we analyze the results of the two choice models presented in the methodology.

4.5.1 Inglehart-Welzel cultural construct

Once we have defined the concept of supra-national culture, we conduct a descriptive analysis of this variable along with the main variables of the dataset from the Pilgrim's Welcome Office studied in Chapter 3 of data sources.

We start analysing the evolution of the number of pilgrims across Inglehart’s clusters, revealing that from 2003 to 2019, all clusters experienced significant increment in the number of arrivals. Therefore, while the destination country remains the main source of pilgrims, the increase in the rest of the clusters has been higher in relative numbers, causing a decline in its significance. However, the COVID-19 crisis in 2020 and 2021 disrupted the series, leading only two clusters; Catholic Europe and the destination country (Spain) to maintain a portion of their arrivals. Notably, clusters like West and South Asia and the Confucian cluster failed to reach their 2019 figures by 2022. Presently, four clusters stand out prominently: Catholic Europe, the destination country (Spain), English-speaking, and Protestant Europe, however, there has been a notable increase in Latin American pilgrims in recent years, all clusters have generally witnessed a remarkable rise. It is worth noting that during Holy Years, the destination country cluster, experienced a surge of around 100,000 individuals, a trend not observed in any other cluster.

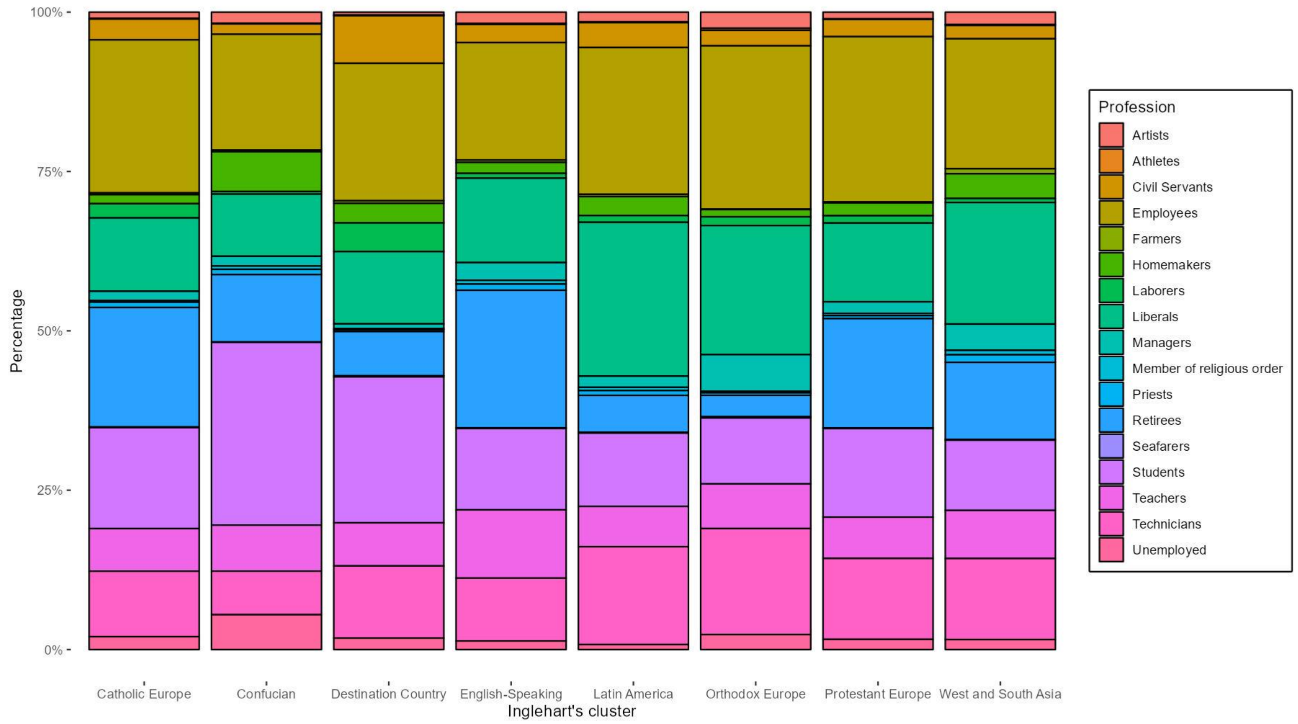
Figure 15. Number of pilgrims by Inglehart’s cluster and year



Source: own elaboration based on Pilgrim’s Welcome Office dataset

Analysing the percentage of pilgrims with certain professions within each cluster, we can see that for most countries, the predominant professions are employees, liberals, retirees, students, and technicians. Nevertheless, there are distinct patterns in the composition of the clusters, where Latin America and Orthodox Europe have a low number of retirees and students but a high number of liberal professionals and technicians, likely due to the average income in those regions. A journey of several weeks might only be feasible for wealthier individuals. Additionally, we observe that Spanish and Confucian individuals have a higher percentage of students, being this influenced by the low cost of Saint James Way for Spaniards and the recent popularity of Saint James Way in Confucian countries and appeal to the younger demographic. Furthermore, the substantial number of retired pilgrims from Catholic Europe, English-speaking countries, and Protestant Europe impact the average age of individuals in these groups, raising it significantly.

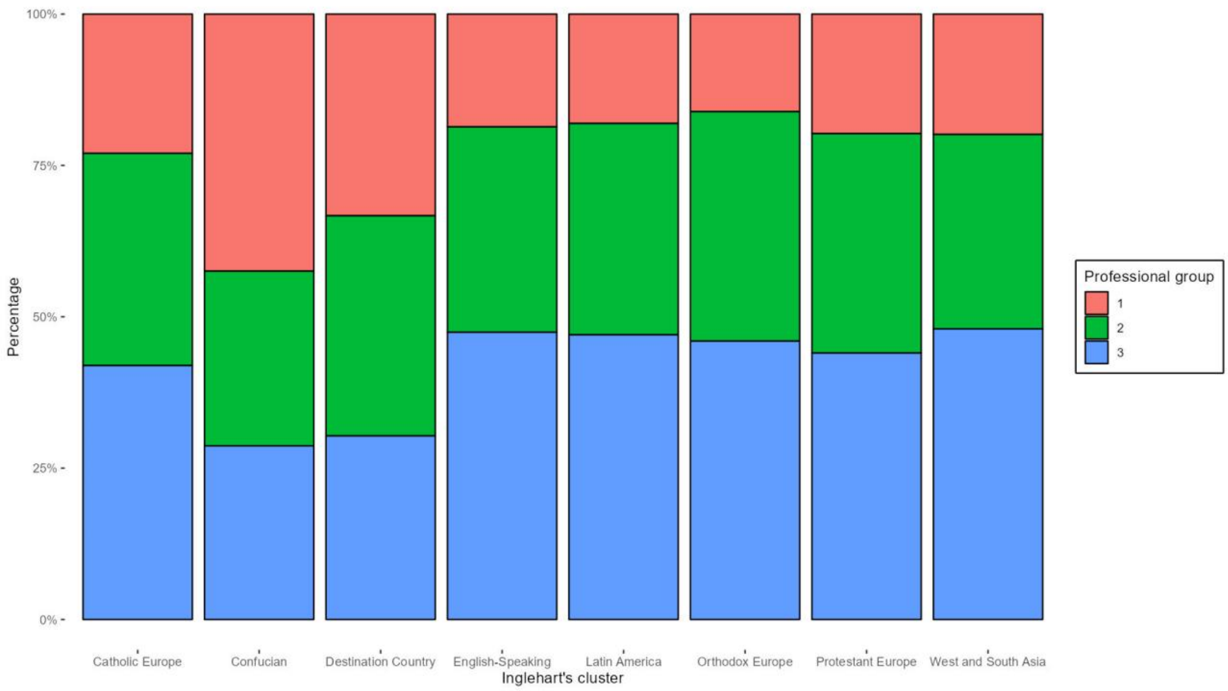
Figure 16. Percentage of pilgrims by profession and Inglehart's clusters



Source: own elaboration based on Pilgrim's Welcome Office dataset

The percentage of pilgrims that belong to professional groups that we associate with high income (professional group 3), or middle income (professional group 2) is around 80% for most clusters, without much variability, which is expected because these higher income individuals are those from other countries that can travel around the world. In the case of Confucian pilgrims, the reason for having a higher percentage of individuals from a professional group associated with low income is because they have a larger number of students than the other clusters.

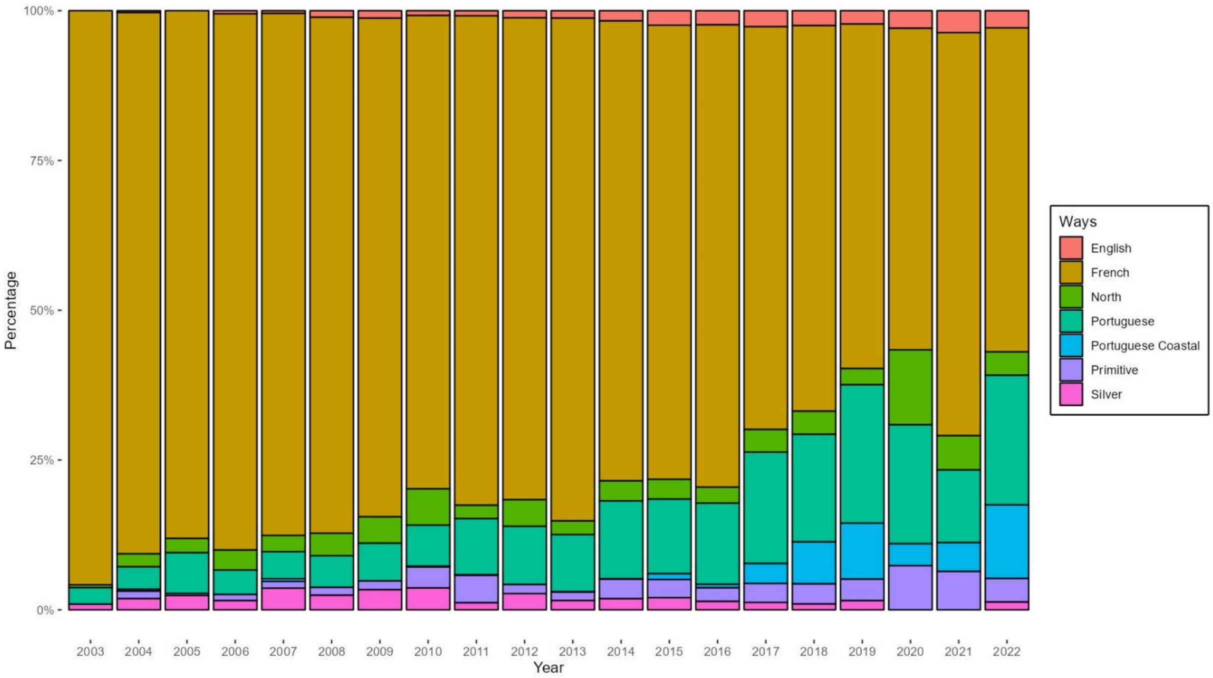
Figure 17. Percentage of pilgrims by professional group and Inglehart's clusters



Source: own elaboration based on Pilgrim's Welcome Office dataset

Analysing the evolution of the percentage of pilgrims that select each route of Saint James Way based on the Inglehart cultural cluster they pertain to, we can start with the pilgrims from West and South Asia. We find a similar evolution to total number of pilgrims on Figure 4, except that in the case of individuals from West and South Asia, it takes longer to start the trend on the decline of the predominance of the French Way, to then accelerate this process towards the end of the series. This results in the percentage of pilgrims traversing each way being similar to the Spanish in the last years, but with a greater importance of the Portuguese Coastal Way and a lesser percentage of pilgrims in the other ways besides the Portuguese and the French, which would be at the same level.

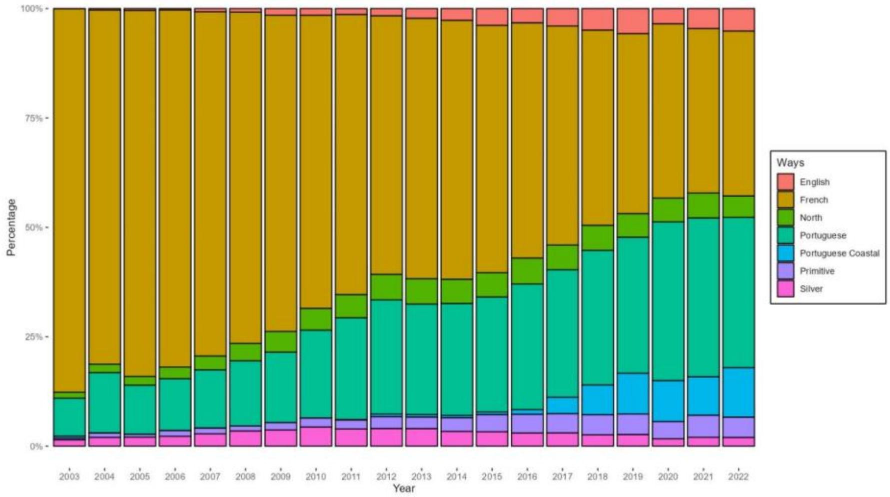
Figure 18. Percentage of West and South Asian pilgrims by Way and year



Source: own elaboration based on Pilgrim's Welcome Office dataset

The evolution of way selection of Catholic Europe pilgrims has shown a clear trend since the first year of the series, where the French Way has gradually decreased in importance. In contrast, the rest of the ways have grown in significance, notably the Portuguese Way and the Portuguese Coastal Way. Consequently, since the year 2020, the combined percentage of these two paths has surpassed the percentage represented by the French Way. While it is true that this trend seems to have halted in the last two years, this could indicate a new shift in direction or a maintenance of the percentages in future years.

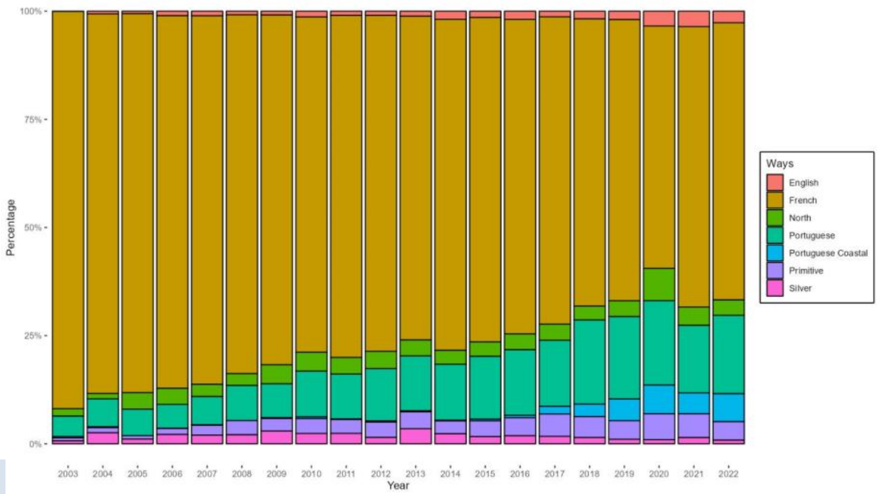
Figure 19. Percentage of Catholic Europe pilgrims by Way and year



Source: own elaboration based on Pilgrim's Welcome Office dataset

The percentage of pilgrims from Latin America completing the French Way has been declining over the years in relative numbers, however, it can be observed that it remains the most important Way in the year 2022, accounting for around 2/3 of pilgrims arrivals. The increasing importance of the other ways is similar for all of them in relation to the French Way, apart from the Portuguese Coastal Way, which had little to no importance in 2003 and becomes the third most important in 2022, as well as the Portuguese Way, contributing approximately 20% of the total pilgrims from the Latin American cluster since 2018.

Figure 20. Percentage of Latin America pilgrims by Way and year

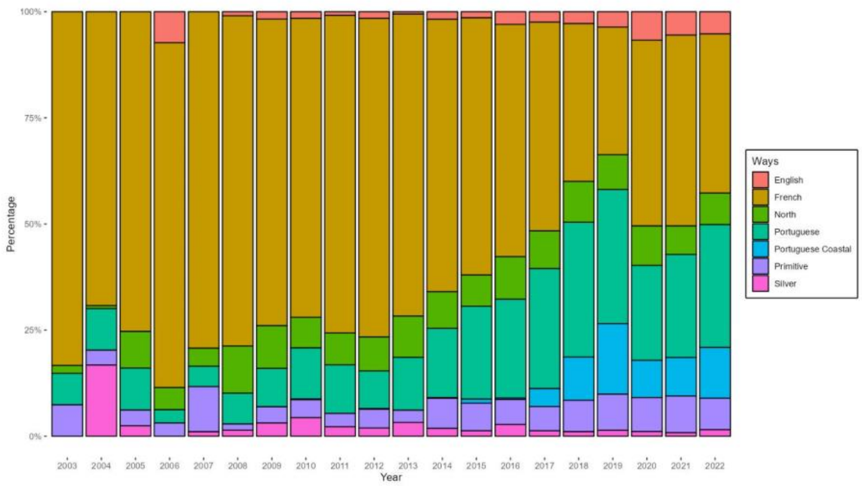


USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Source: own elaboration based on Pilgrim's Welcome Office dataset

The percentage of pilgrims from Orthodox Europe completing the French Way has been declining since the beginning of the series. Consequently, the importance of the other ways has remained consistent, but there has been a substantial increase in the percentage of pilgrims choosing to undertake the Portuguese Way or the Portuguese Coastal Way since 2012, while 6 years later, the combined percentage of these two ways has surpassed the percentage of pilgrims on the French Way.

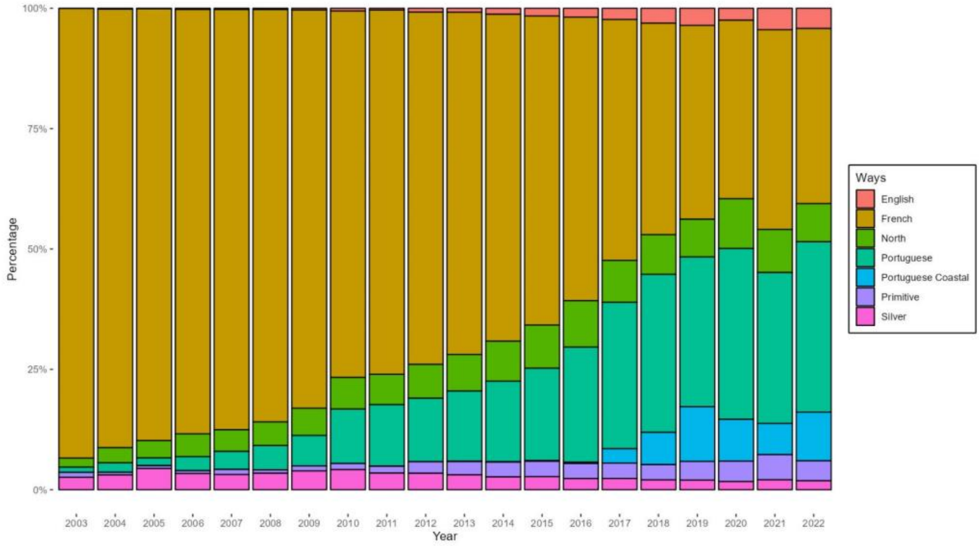
Figure 21. Percentage of Orthodox Europe pilgrims by Way and year



Source: own elaboration based on Pilgrim's Welcome Office dataset

The percentage of pilgrims from Protestant Europe completing the French Way has been declining since the beginning of the series, coinciding with the rise in the percentage of pilgrims choosing to undertake the Portuguese Way, Portuguese Coastal Way, and the Northern Way. Consequently, since 2019, the combined percentage of the Portuguese ways has surpassed the percentage of pilgrims on the French Way, and by 2022, the Portuguese Way has reached the numbers of the French Way, suggesting that this trend is unlikely to reverse in the near future.

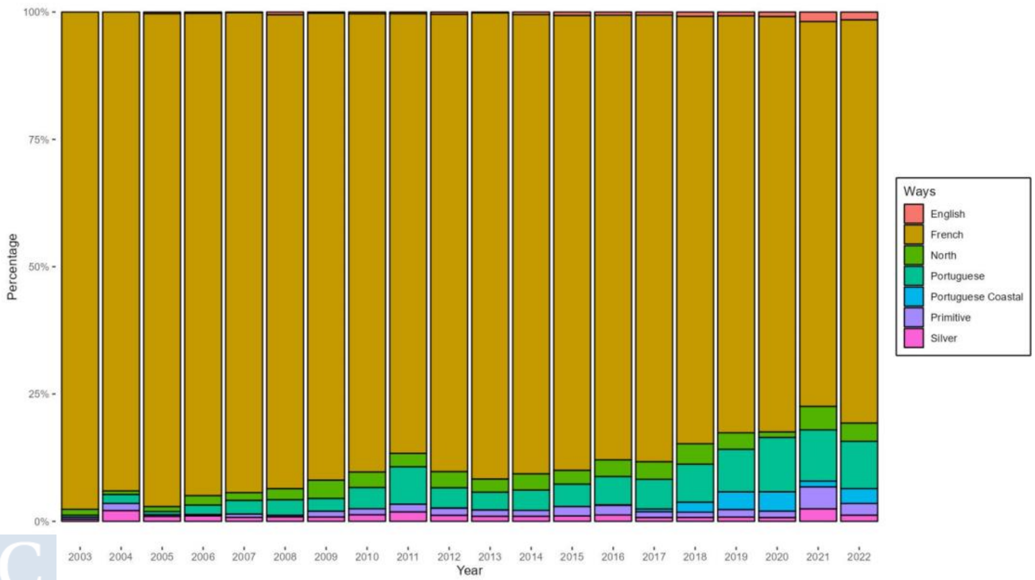
Figure 22. Percentage of Protestant Europe pilgrims by Way and year



Source: own elaboration based on Pilgrim's Welcome Office dataset

The behaviour of the Confucian cluster regarding the chosen pilgrimage path seems to be the only one that doesn't show abrupt changes in the historical series. It can be observed that they maintain an absolute predominance, consistently representing over 75% of the total pilgrimage paths throughout all the years in the series, despite a gradual increase in the rest of the paths in recent years, especially the Portuguese way.

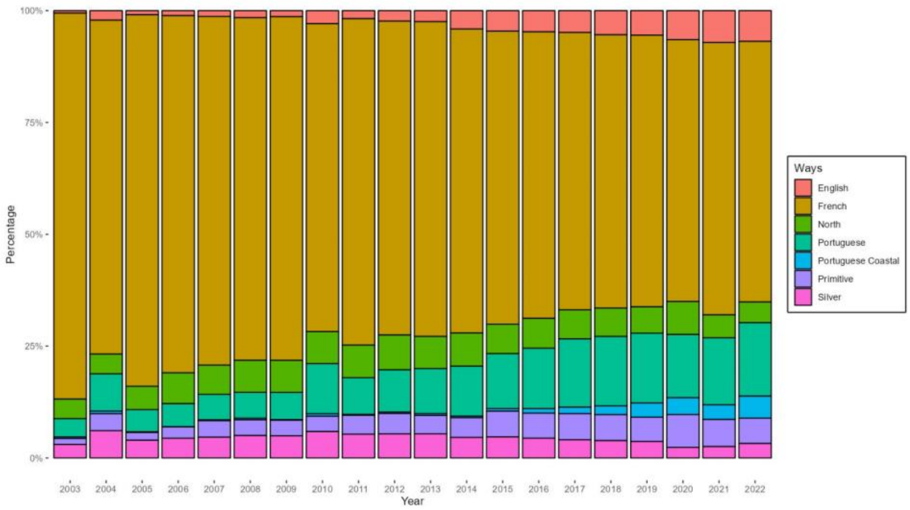
Figure 23. Percentage of Confucian pilgrims by Way and year



Source: own elaboration based on Pilgrim's Welcome Office dataset

The percentage of Destination country pilgrims (Spanish) completing the French Way declined from the beginning of the series until 2016. However, contrastingly with the other clusters, from this year onwards, it stabilized around the same figures. This decline was primarily caused by the rise of the Portuguese, English, and Primitive Ways. There was also a decrease in the importance of the Silver Way since 2010. Additionally, a pattern can be observed during the Holy years of 2004 and 2010, when pilgrims chose the more popular way, the French Way, in a smaller proportion and instead embarked on different Ways from the French way, probably to avoid the saturation of the French way on said years.

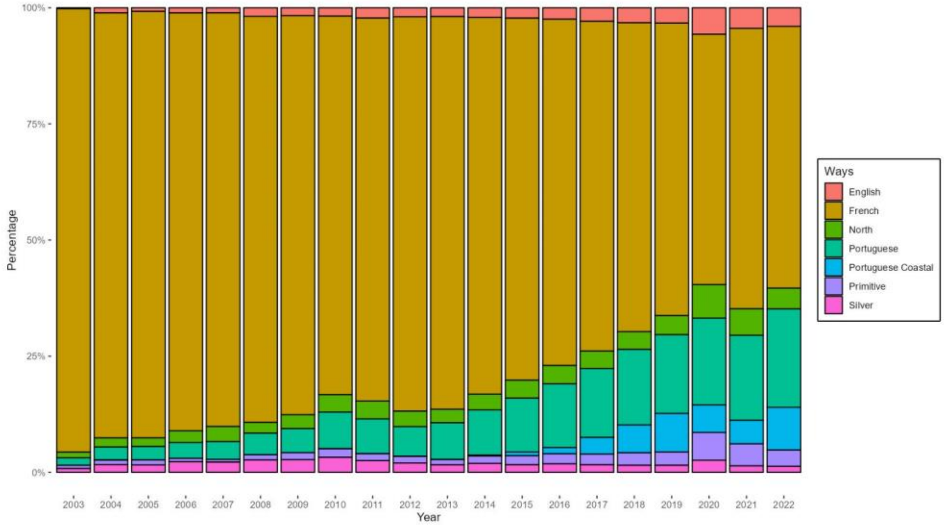
Figure 24. Percentage of Destination country (Spanish) pilgrims by Way and year



Source: own elaboration based on Pilgrim's Welcome Office dataset

The percentage of pilgrims from the English-Speaking cluster completing the French Way has been declining over the years. However, it can be observed that it remains the most important way in terms of arrivals in the last year of the series, accounting for over 50% of pilgrim arrivals. The rise in importance of the ways is similar for all concerning the French Way, apart from the Portuguese Coastal Way, which had little significance in 2003 and has become the third most important way by 2022.

Figure 25. Percentage of English-Speaking pilgrims by Way and year



Source: own elaboration based on Pilgrim's Welcome Office dataset

4.5.2 Testing the effect of cultural clusters on choice of Way

We test the hypothesis that the proportion of selected alternatives differs for each Way and cluster based on the descriptive analysis of the sample, that being explained by different clusters do not sharing the same preferences for the alternatives. We applied Pearson's chi-squared test for categorical variables to assess whether each cluster presents the same distribution (Gibbons & Chakraborti, 2020; Crawley, 2012)(see notes in ANNEX 2).

The contingency table on Table 8 displays the distribution of choices per cultural cluster, showing that the distribution differs among the cultural clusters and that the French and Portuguese Ways are by far the most popular.

Table 8. Contingency table for Way and cluster

Way	Cluster							
	1	2	3	4	5	6	7	8
Portuguese Coastal	3.22%	2.34%	1.45%	6.45%	2.24%	1.07%	.93%	2.59%
French	73.15%	6.28%	74.74%	5.55%	65.19%	87.38%	7.76%	76.84%
English	1.93%	3.11%	1.44%	2.51%	1.55%	.60%	3.46%	2.40%
North	3.01%	4.62%	3.07%	7.93%	6.91%	2.99%	5.39%	3.22%
Portuguese	14.43%	23.33%	13.85%	24.58%	18.70%	5.77%	11.05%	11.42%
Primitive	2.61%	3.26%	3.76%	6.37%	2.46%	1.32%	4.29%	1.93%
Silver	1.65%	3.07%	1.69%	1.61%	2.95%	.88%	4.13%	1.61%

Source: own elaboration based on Pilgrim's Welcome Office dataset

Table 9. Pearson's chi squared for way and cluster

	X-squared	df	p-value
Pearson's Chi-squared	118641	42	2,2E-16 ***

*** Significance 1%; ** Significance 5%; * Significance 10%

Source: own elaboration based on Pilgrim's Welcome Office dataset

The Pearson's chi-squared test for all clusters on Table 9 rejects the null hypothesis (proportion of choices is similar for each cluster) and suggests that choices are in fact distributed differently across cultural clusters.

4.5.3 Estimation of Model I

Our first estimation seeks to show that the cultural cluster of the supra-national culture conditions the chosen alternative. To test whether this is the case, we fit seven binary logit models, one for each Way. The results reveal that pilgrims have different preferences for different Ways, even after controlling the effect of their individual characteristics. Model I contain a seven-specification model with one specification per Way. This model analyses the choice of one Way relative to the other six alternatives, using pilgrim's cultural cluster as an explanatory variable on their election of the alternative.

Table 10. Coefficients of Model I for each Way

Variables	French	Portuguese Coastal	English	North	Silver	Portuguese	Primitive
Cluster (Ref: Spain)							
West and South Asia	.121 ***	1.151 ***	-.684 ***	-.488 ***	-.842 ***	.266 ***	-.517 ***
Catholic Europe	-.468 ***	.906 ***	-.122 ***	-.139 ***	-.294 ***	.894 ***	-.290 ***
Latin America	.200 ***	.393 ***	-.925 ***	-.533 ***	-.876 ***	.245 ***	-.148 ***
Orthodox Europe	-.865 ***	1.882 ***	-.393 ***	.530 ***	-.863 ***	.923 ***	.425 ***
Protestant Europe	-.252 ***	.806 ***	-.898 ***	.363 ***	-.256 ***	.583 ***	-.598 ***
Confucian	1.060 ***	.058 **	-1.880 ***	-.526 ***	-1.479 ***	-.742 ***	-1.225 ***
English-Speaking	.319 ***	.939 ***	-.467 ***	-.419 ***	-.865 ***	-0,001	-.848 ***
Gender: Male	-.056 ***	-.291 ***	.011 *	.256 ***	.282 ***	-.108 ***	.267 ***
Professional Group (Ref: Professional Group 3)							
Professional Group 1	.012 ***	-.105 ***	-.039 ***	.120 ***	.045 ***	-.014 ***	-.164 ***
Professional Group 2	-.050 ***	.102 ***	.089 ***	-.031 ***	-.142 ***	.086 ***	.042 ***
Means of transportation: Not Walking	.166 ***	-.615 ***	-3.694 ***	.617 ***	.665 ***	-.382 ***	-.699 ***
Constant	.908 ***	-4.436 ***	-3.218 ***	-3.149 ***	-3.392 ***	-2.013 ***	-3.161 ***
Log Likelihood	-1,952,238	-259,201	-410,767	-628,902	-456,883	-1,285,883	-485,702
Akaike Inf. Crit.	3,904,499	518,426	821,559	1,257,829	913,790	2,571,791	971,429
Wald test of cluster variables X^2 (DF=7)	50825.5	10663.5	7135.0	7499.4	6927.0	64647.5	8005.5

*** Significance 1%; ** Significance 5%; * Significance 10%

Source: own elaboration based on Pilgrim's Welcome Office dataset

Table 10 displays the results of Model 1. Rows indicate the cluster and individual variables, and columns present the seven models for each Way. We include two measures of goodness of fit, Log-likelihood and Akaike Information Criteria (Akaike, 2011). Spaniards (reference estimation cluster) prefer the English, Silver, and Primitive Ways more strongly than do the other clusters, except for Orthodox Europeans, who prefer the Primitive Way more strongly than Spaniards do. All clusters except Confucians, in contrast, prefer the Portuguese Coastal and Portuguese Ways. None of the other clusters shows a clear preference for any way, indicating that their tastes differ.

High and medium-income individuals (Professional Groups 2 and 3) prefer the Portuguese Coastal, English, Portuguese, and Primitive Ways, whereas low and medium-income individuals prefer the Northern and Silver Ways.

We can conclude that women prefer the English, Northern and Silver Ways, whereas men prefer the others. Pilgrims completing Saint James Way with a means of transport other than by feet prefer the French, Northern and Portuguese Ways. This preference may be explained by the fact that the Northern and Portuguese Ways have the highest proportion of cyclists relative to total number of pilgrims.

All coefficients of the cultural cluster variables are statistically significant, indicating a robust fitting procedure. Most importantly, the coefficients differ for each cluster, a finding that supports the first hypothesis advanced above.

To confirm whether the cultural cluster variables are statistically significant, we performed a Wald test (Wooldridge, 2010, 2016). We assume that the coefficients of the clusters would be equal to zero for all Ways, whereas the alternative hypothesis would show at least one of these coefficients to be different from zero:

Since the cluster coefficients differ from 0 for all Ways, we reject the null hypothesis and conclude that pilgrims from different cultural clusters have different preferences for each Way, in line with our first interpretation of the results of Model I.

The results confirm that supra-national culture affects pilgrims' DM process about the chosen alternative. In line with this first result, we propose Model II, which enables us to test whether the path and individual characteristics condition the DM process based on supra-national culture.

4.5.4 Estimation of Model II

Table 11. Coefficients of model II for each cluster

	Dependent variable: Way	West and South Asia	Catholic Europe	Latin America	Orthodox Europe	Protestant Europe	Confucian	Spain	English-Speaking
Common variables for all Ways	Popularity (Ref: Low Popularity)								
	High Popularity	3.384 ***	2.606 ***	3.563 ***	2.019 ***	2.451 ***	**3.289 *** **	2.901 ***	3.514 0***
	Medium Popularity	1.697 ***	1.979 ***	2.347 ***	1.659 ***	2.314 ***	**1.786 *** **	1.554 ***	1.517 ***
	Intensity	-.347 ***	.035 ***	.356 ***	-.308 ***	.182 ***	**.155 * **	.277 ***	-.684 ***
	World Heritage	.263	.491 ***	.017	.489 ***	.776 ***	**.796 *** **	.091 ***	.472 ***
Portuguese Coastal	Gender: Male	-.008	**-.285 *** **	-.157 ***	.063	-.253 ***	-.277 ***	-.714 ***	-.204 ***
	Professional Group (Ref: Professional Group 3)								
	Professional Group 1	-.103	-.175 ***	-.118 *	-.625 ***	.043	-.351 ***	-.560 ***	-.666 ***
	Professional Group 2	.263 ***	.288 ***	.094 **	.269 ***	.337 ***	.087	-.693 ***	.077 ***
	Means of transportation: Not Walking	-.137	-.292 ***	.330 ***	-1.10 ***	-1.233 ***	.056	-1.338 ***	-.089
English	Gender: Male	-.176 *	.073 ***	.073 *	-.494 ***	-.201 ***	-.171 *	.198 ***	.133 ***
	Professional Group (Ref: Professional Group 3)								
	Professional Group 1	-.122	.151 ***	.311 ***	-.454 ***	-.444 ***	-.943 ***	.247 ***	-.427 ***
	Professional Group 2	-.096	.486 ***	.045	-.586 ***	-.070 ***	-.703 ***	.295 ***	-.112 ***
	Means of transportation: Not Walking	-3.285 ***	-3.807 ***	-4.016 ***	-16.681	-5.391 ***	-2.281 **	-3.636 ***	-3.025 ***
North	Man	.359 ***	.153 ***	.332 ***	.222 ***	.256 ***	.129 ***	.270 ***	.431 ***
	Professional Group (Ref: Professional Group 3)								
	Professional group 1	-.031	-.122 ***	.345 ***	-.182 ***	.312 ***	-.284 ***	.140 ***	.121 ***
	Professional Group 2	-.071	-.069 ***	.121 ***	.032	.013	-.164 ***	-.023 ***	.163 ***
	Means of transportation: Not Walking	.725 ***	.043 **	.578 ***	.791 ***	.070 ***	.429 ***	.678 ***	.745 ***

	Dependent variable: Way	West and South Asia	Catholic Europe	Latin America	Orthodox Europe	Protestant Europe	Confucian	Spain	English-Speaking
Silver	Gender: Male	.013	.353 ***	-.005	-.533 ***	.550 ***	.097	.120 ***	.424 ***
	Professional Group (Ref: Professional Group 3)								
	Professional Group 1	-.315 **	-.450 ***	.783 ***	-.744 ***	-.679 ***	-.856 ***	.141 ***	-.189 ***
	Professional Group 2	-.317 ***	.273 ***	-.699 ***	-.732 ***	-.422 ***	-.635 ***	-.084 ***	-.218 ***
	Means of transportation: Not Walking	.644 ***	.768 ***	-.283 ***	.261 *	.711 ***	.637 ***	.533 ***	.467 ***
Portuguese	Gender: Male	.053	-.002	-.071 ***	-.101 ***	-.186 ***	-.169 ***	-.107 ***	.025 **
	Professional Group (Ref: Professional Group 3)								
	Professional Group 1	-.187 ***	.133 ***	-.191 ***	-.481 ***	-.171 ***	-.384 ***	.016 **	-.489 ***
	Professional Group 2	.017	.267 ***	.033 *	-.090 **	.036 ***	-.112 ***	.044 ***	-.116 ***
	Means of transportation: Not Walking	-.248 ***	.564 ***	-.500 ***	-.585 ***	-1.398 ***	.067	-1.602 ***	-.290 ***
Primitive	Gender: Male	.460 ***	.252 ***	.503 ***	.388 ***	.141 ***	-.015	.392 ***	.345 ***
	Professional Group (Ref: Professional Group 3)								
	Professional Group 1	.196 *	-.074 ***	.647 ***	.056	-.210 ***	-.173 **	-.134 ***	.315 ***
	Professional Group 2	.250 ***	.307 ***	.405 ***	.432 ***	.001 ***	-.308 ***	.070 ***	.278 ***
	Means of transportation: Not Walking	-1.716 ***	-1.205 ***	-1.198 ***	-1.453 ***	-1.564 ***	-1.743 ***	-.563 ***	-1.003 ***

*** Significance 1%; ** Significance 5%; * Significance 10%

Source: own elaboration based on Pilgrim's Welcome Office dataset

Estimation of Model II involves an eight-specification scheme, in which the supra-national cultures serve to segment the dataset into eight simples, which results can be seen in Table 11. These specifications enable us to determine whether the characteristics of the different Ways and the pilgrims' characteristics condition the DM process depending on the supra-national culture to which the pilgrims belong.

The coefficients of the Way's variables are easy to interpret, as they are common to the eight models. Consider the coefficients for the Confucian cluster. If we recall the meaning of the popularity variables the coefficients 3.289 for "Popularity High" and 1.786 for "Popularity Medium" indicate that the pilgrims from the cluster "Confucian" prefer the most popular ways. The positive and significant (* and ***, respectively) coefficients of intensity (.155), and World Heritage Status (.796) on the other hand, indicate that Confucian pilgrims prefer the most intense ways, as well as the ways recognized by the UNESCO as World Heritage Sites.

Determining the meaning of the coefficients for the pilgrim's variables is more complex, as the implementation of the multinomial logit model with alternative- specific variables requires fitting the model for J-1 alternatives. The results must thus be interpreted by comparison to a reference alternative that we have eliminated. In our case, the reference alternative is the French way. As this process complicates interpretation of the coefficients estimated, we provide detailed explanation of one individual coefficient and its relation to individuals' preferences. For the Catholic Europe cluster, the variable "man" has a negative (-.285) and significant (***) coefficient for the "Portuguese Coastal" Way. A negative coefficient means that women from the "Catholic Europe" cluster prefer the "Portuguese Coastal" Way more than do men from this cluster. We thus clarify interpretation of the coefficient by understanding it as a preference relationship. Since we can quantify a preference relationship through utility theory, we present the relation as a comparison of utility ratios between men and women as:

$$\frac{Utility(Portuguese\ Coastal\ Way\ Men)}{Utility(French\ Way\ Men)} < \frac{Utility(Portuguese\ Coastal\ Way\ Women)}{Utility(French\ Way\ Women)} \quad (23)$$

We can put numbers to those utilities by estimating the number of times each Way has been chosen by either men or women. We thus finally obtain:

$$\frac{2.0\%}{59.2\%} = .034 < \frac{2.8\%}{61.2\%} = .046 \tag{24}$$

which indicates a higher utility ratio for the "Portuguese Coastal" Way for women than for men, and thus a stronger preference, where these findings must be taken as relative to the French Way.

As Table 11 shows, pilgrims from all clusters except the Latin American and West and South Asian (which have no significant coefficients), prefer both the most popular and World Heritage awarded Ways. However, preference for intensity that varies among the clusters. Half the clusters (Catholic Europe, Latin America, Protestant Europe, Confucianists and Spain) prefer high-intensity Ways, whereas the other half prefer lower intensity.

Women from clusters of Latin America, Protestant Europe, Confucianism and Spain prefer the French, Portuguese Coastal and Portuguese Ways, whereas men from these clusters prefer the Northern and Silver Ways. Gender does not clearly affect preferences for the four remaining clusters.

Low-income individuals in most clusters prefer the Silver Way to the others, whereas high-income individuals prefer the Portuguese Coastal and Portuguese Ways. We find dissimilar patterns among the other clusters, as they show no clear preference based on income level.

Pilgrims undertaking Saint James Way by a means of transport other than by foot prefer the French, Northern and Portuguese Ways for most clusters. Individuals from Catholic Europe prefer the Primitive over the Portuguese Way, individuals from Latin America prefer the Portuguese and Coastal Ways, and Confucians prefer the Silver Way.

4.6 DISCUSSION

Based on the gap on the literature review of push-pull theory, we have considered supra-national cultures as a relevant push factor that may explain tourists' preferences for any form of touristic experience. We tested the hypothesis by choosing the supra-national cultural clusters developed by Inglehart and Welzel to explain why pilgrims choose alternative routes of Saint James Way. In what follows, the main findings and the benefits of using supra-national culture and choice models in the field of tourism are presented.

4.6.1 Main findings of the case study

The findings suggest that the main hypothesis of this chapter, i.e. tourists' supra-national culture is a relevant push factor that explain the tourist's DM processes, is correct. More specifically, we tested the relationship between pilgrims' cultural cluster and their choice of the Way to reach the city of Santiago de Compostela. Firstly, the results of Pearson's chi-squared test performed in section 4.5 and shown in Table 9 rejects the null hypothesis, i.e. the proportion of choices is similar for each cluster. This indicating that the choices of the Way are in fact distributed differently across supra-national cultural clusters. Secondly, the findings of Model I shown in Table 10 indicated that the "Destination Country" cultural cluster prefers the English, Silver, and Primitive Ways more strongly than do the other clusters. This is true for all clusters except for Orthodox Europeans, who prefer the Primitive Way more strongly than do Spaniards. In contrast, all clusters except Confucians prefer the Portuguese Coastal and Portuguese Ways. The differences found on the taste of each cultural cluster are robust as all coefficients of the cultural cluster variables are statistically significant. Finally, the coefficients differ for each cluster, evidence that supports the potential of supra-national cultural clusters not only to explain the choices, but also to predict and discriminate the DM of the pilgrims. Thirdly, the results of Model II, as shown in Table 11, are like those obtained in Model I. Thus, the main difference between both models is that the second one has enabled us to understand how the characteristics of the pilgrims and the alternatives influence the choice of ways by each supra-national culture.

4.6.2 Benefits of using supra-national cultures as explanatory variables in the field of tourism

Regarding the most popular concepts of culture used in the field of tourism (cultural distance and national culture), our proposal relies on the supra-national construct of culture. Two main benefits can be obtained from the use of this variable: aggregation and accuracy. When using the national culture variable, the volume of available data per country can be an issue. On many occasions, there is not enough data per country to fit or train a quantitative/predictive model. In these cases, the supra-national cultural variable can overcome the problem by facilitating the aggregation of data of different countries into a single group. On the other hand, when using the cultural distance, the issue comes in the form of ambiguity. This happens when two countries with opposing cultural values could be labelled with identical cultural distances from the destination country. For example, if we take Spain as the destination country, Malaysia and Sweden obtain quite similar cultural distances of 1.316 and 1.153, respectively (measured with Kogut & Singh (1988) distance), but a comparison performed with Inglehart and Welzel cultural map, shows that the two countries' supra-national culture is quite different. According to Inglehart and Welzel's Cultural Map, these tourists would belong to different, even opposing, cultures (Malaysia to West and South Asia, Sweden to Protestant Europe). To avoid these ambiguity problems, the "supra-national culture" variable provides a better accuracy on the identification of real differences between cultures.

4.6.3 Benefits of using choice models to implement the push-pull theory

The application of choice models allows to estimate coefficients for each explanatory variable, which therefore is the mean to determine the importance of push and pull factors of the DM process. This way, policymakers and companies in the sector can design campaigns targeted at groups "that are likely to be different to from that of another, offering strategies that may apply to countries with the same supra national culture" (Kumar & Dhir, 2020). The specific application of the multinomial logit with specific alternative variables in this chapter has allowed us to capture the impact of individuals observed and unobserved factors on utility for a variety of supra-national cultures.

4.7 CONCLUSIONS

Traditionally, DM analyses in the field of tourism have been based on the study of push and pull factors that influence individuals' motivations to consume specific tourism products, including socio-economic variables to control for spurious behaviour. It has also been confirmed that the supra-national context in which individuals have developed their lives condition their satisfaction or their perception of marketing campaigns. Despite these insights in DM literature, the context has only been measured with national-level factors. The results obtained in this study lead us to believe that the analysis of the supra-national context, widely validated in the field of sociology, is of great interest for DM models, particularly those empirically developed through choice models. Thus, we found that several implications can be drawn from the study:

Societal implications: Based on our findings, we can conclude that supra-national cultural groups determine similar behavioural patterns within its members and differentiation from those outside their group. This interpretation of culture, as a collective phenomenon raises a complex theoretical problem: which is the best way to measure this characteristic empirically? In this chapter, we propose a solution to this question; aggregate the individual cultural values of members of the different national contexts into 8 different cultural clusters. Thus, this conception of group behaviour allows for a new interpretation of complex behaviour in a globalized society.

Marketing lessons: The application of supra-national culture as a means for market segmentation may optimize the usage of resources in marketing campaigns. It allows reaching a larger number of markets with the same resources by designing marketing strategies targeted at 8 specific market segments instead of every country that they want to target. This approach reduces the cost of campaigns and enables to focus on the specific characteristics of each of the 8 groups defined by Inglehart's cultural map in cross-cultural research. The use of supra-national culture also allows the search of target markets since there are sociological characteristics that differ, and some cultures may be more aligned with the values of companies and internationalization policies of the companies may benefit from this approach, reaching broader regions rather than specific countries, as is commonly done. From a consumer's perspective, the recommendation of new tourism products or services may be enriched by focusing not only on the classic individual characteristics of consumers, but also on group level characteristics. This allows a better understanding of consumer preferences as a member of the society they belong to, enabling consumers to receive even more personalized experiences, and therefore, more satisfying.

Policymaking lessons: using the concept of supra-national culture helps to better understand consumer preferences, enabling the design of specific policies for municipalities, regions, or countries, thereby enhancing destination management. This improvement in destination management could imply encouraging or discouraging the arrival of tourists based on whether they are known to have low or high social, environmental, or economic impact in the area balancing pros and cons, perhaps even defining rates, taxes, or discounts for individuals belonging to specific cultural groups.

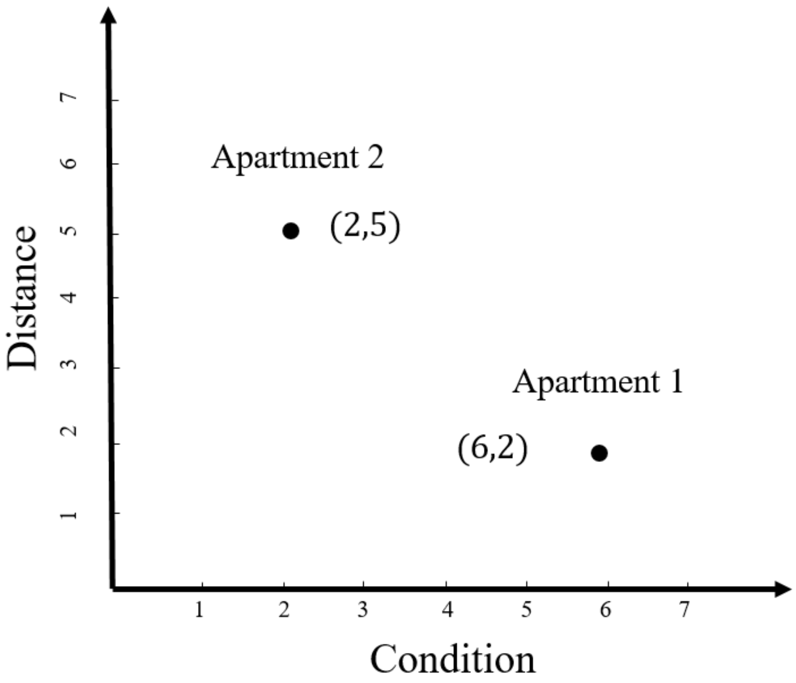
CHAPTER 5. HOW DO CULTURES RESOLVE TRADE-OFF DILEMMAS?

5.1 INTRODUCTION

Choice under conflict reffers to a choice problem or choice situation with a conflict between attributes or features of the alternatives (Tversky & Shafir, 1992). This concept arises when individuals are unable to compare the pros and cons of each alternative to choose from or find it very difficult, making this process arduous as the number of non-dominated alternatives to compare increases. Also referred to as a trade-off dilemma, is a phase of the decision-making (DM) process, in which individuals or groups of individuals confront a series of possible alternatives, in which none of them is superior in all desirable characteristics to the others (Yates & de Oliveira, 2016). When faced with conflict, the decision maker needs to make trade-offs between the attributes. This means that each decision-maker must estimate the relevance of the attributes by exchanging the value of some of them for the value of others. This exchange can be formally represented through preferences, thereby illustrating a representation of trade-offs as an expression of these preferences.

Analysing Figure 26, depicting a trade-off dilemma, we can see that Apartment 1 is in better condition but located further, while Apartment 2 is in worse condition and located closer. The fact that neither of the apartment's alternatives dominate over the other, requires individuals to choose which set of features they prefer.

Figure 26. Conflict and trade-off between the attributes of alternatives



Note: Figure adapts Fig 1 of Tversky & Shafir (1992), depicting the valuation of characteristics of example in Section 3.8. of Yates & de Oliveira (2016), where Apartment 1 is in better condition than Apartment 2, having a higher valuation, and Apartment 2 is at a shorter distance than Apartment 1, such that the former has a superior valuation to the latter in terms of distance

Source: own elaboration

Trade-offs are individual estrategies that vary among decision-makers. However, the cultural context and background of each individual may shape how the trade-offs emerge. Chu et al., (1999, 2005) and Chu and Spires (2008) are the only researchers that have been able to demonstrate that different cultures (China, Taiwan, Japan and United States of America) resolve their DM through different strategies (Hammond et al., 1998; Retief et al., 2013). They found that countries that use compensatory strategies, where the negative characteristics of the alternative offset the positive, tend to resolve trade-off dilemmas. While societies that follow non-compensatory strategies (Chu et al., 1999; Kreps, 2020), tend to use attribute dominance strategies in such a way that one negative attribute can eliminate an alternative from the choice set. In summary, the rationale followed by Chu and Spires (2008) is that the decision-maker's culture impact on their trade-off strategy, which is the mechanism that solves the conflict during the choice process. The following causal model can synthetize the line of thinking: [Culture -> Trade-off -> Conflict].

The current line of research presents a series of limitations that are worth considering:

Analysing the effect of affiliation to certain cultures on the DM process of individuals has only been focused on two or three national cultures (Levinson & Peng, 2007; Yates & de Oliveira, 2016). Authors usually benchmark the DM processes of East Asian (China or Japan) against Western individuals (USA) (Vijver & Leung, 2021; Wang, 1996; Wright & Phillips, 1980).This can lead to bias on the results of the studies, neglecting the generalization to other cultural groups.

Researchers of culture and DM tend to rely on the concept of collectivism vs individualism, the problem is that other cultural factors that can affect the DM process are often neglected (Yates & de Oliveira, 2016). This bias may originate from the tendency of comparing Asian countries against Western countries on cross-cultural analysis research. This makes the ratings on these values very different, with tendency towards collectivism being more typical of Asian countries, while individualism more typical of Western countries. However, if we extend this analysis to culturally similar groups, the range of explanatory factors that can explain the differences in DM may be larger, as they need a wider variety of factors to explain the difference in the DM process.

In the field of cross-cultural analysis, individual's preferences have not been generally introduced through rational choice theory (Arrow, 1959; Kreps, 2020; Varian, 2010). The authors of these studies do not use the conceptual framework of microeconomics or judgment and DM, resulting in a lack of models that explicitly quantify preferences in cross-cultural analysis.

So, we have a missing point: how can trade-offs be estimated by an external observer (i.e., a researcher)? In this chapter, we will put ourselves in the position of a researcher trying to understand how a series of individuals solve the trade-off between the attributes of alternatives in a decision problem, deducing the strategy they follow. Therefore, we will attempt to complete the work of Chu et al. (1999, 2005) and Chu & Spires (2008), as they were able to deduce the strategies individuals tend to use to solve trade-offs between attributes of the alternatives, but they were unable to quantify the valuation that individuals give to each of said attributes. We propose the use of rational choice theory to solve said limitation, representing preferences through relations or utility functions (Kreps, 2020). Thereby, individuals assess each of the features of the available alternatives on the trade-off phase, choosing the importance they assign to each one of them (Yates & de Oliveira, 2016), so that when the process ends, they can order the alternatives according to said preferences (Mcfadden, 1980).

This chapter makes a significant contribution to the field of cross-cultural analysis and DM by addressing two research gaps and providing valuable insights into them. The first contribution is the creation of a new causal model, where individuals' preferences explain the trade-offs between attributes. The consequence of introducing preferences is the update of the causal formula: [Culture ->**Preferences** -> Trade-off -> Conflict]. Furthermore, as evidenced in the literature review, certain cultures tend to eliminate an alternative from the DM process when it has a negative characteristic (Chu et al., 1999, 2005; Chu & Spires, 2008). However, in this chapter, we will demonstrate that even cultures that theoretically employ such strategies engage in trade-offs between attributes, resolving a conflict when selecting their preferred alternatives.

This chapter uses multinomial choice models, based on the Random Utility Models (Mcfadden, 1980), to explain how culture affects individuals' preferences and how these preferences impact their trade-off resolution. The methodology allows the inclusion of alternative-specific characteristics, in a model that enable us to capture the impact of individual's observed and unobserved factors on utility (Train, 2009; Wooldridge, 2010), studying several alternatives. The methodology enables the selection of relevant cultural variables for the trade-off process and facilitates the comparison of preferences for alternatives across each country and supra-national cultures.

To resolve the two limitations of the current research line on cross-cultural analysis and DM, we propose the use of the full conceptual framework of the founders of empirical cross-cultural analysis (Hofstede, 1980; Inglehart, 1997; Schwartz, 1994), comparing not two but a diverse set of national cultures. We do not overlook any of the factors considered in their analysis, since we select the significant cultural variables through a statistical analysis of the models, to test whether these cultural variables affect the trade-off phase.

The chapter is organized in a logical manner in order to ensure maximun clarity. It starts with a literature review on the characterization of cultures explaining the cultural factors, national cultures and supra-national cultures. The methodology section explains the research approach and the model. Results are presented sistematically following the structure of the aforementioned sections and finally the discussion analyses and interprets the results, relating them to the research objectives, literature review and conclusions.

5.2 LITERATURE REVIEW

All definitions of culture indicate to a greater or lesser degree that it is a collective phenomenon that identifies groups of people (Akaliyski et al., 2021). The interpretation of culture as a collective phenomenon raises a complex theoretical problem, widely debated in the literature; which is the best way to measure this characteristic empirically? The founders of empirical cross-cultural analysis (Hofstede, 1980; Inglehart, 1997; Schwartz, 1992), came to a solution, aggregate to the national level, the individual cultural values of members of different national contexts.

5.2.1 Cultural constructs

Cultural constructs are conceptual frameworks used to compare different cultures, based on the values they attribute to specific items that characterize human behaviour. In this section, we will provide an overview of the cultural dimensions proposed by Hofstede, Schwartz, and Inglehart.

Hofstede

Hofstede developed questionnaires that he administered to IBM employees in fifty countries in the late 1960s, from which he obtained four dimensions (Individualism versus Collectivism, Power Distance, Masculinity and Uncertainty Avoidance), after performing a factor analysis. In the 2000s he added two cultural dimensions (Long-term Orientation and Indulgence versus Restraint) with data from the World Values Survey (Inglehart et al., 2014).

Schwartz

Schwartz conducted values surveys between 1988 and 2005 for teachers and students of several countries, from which he obtained seven cultural value orientations (Harmony, Embeddedness, Hierarchy, Mastery, Affective Autonomy, Intellectual Autonomy and Egalitarianism) after performing a factor analysis.

Inglehart

Inglehart carried seven rounds of the World Values Survey between 1981 and 2022 for adult population of several countries, from which he obtained two dimensions (traditional versus secular and survival versus self-expression) after performing a factor analysis on ten questions of his survey. On Table 12 we can see the overview of cultural dimensions depicting theoretical framework and dimensions of aforementioned conceptual frameworks.

Table 12. Overview of cultural dimensions depicting theoretical framework, name of dimension and description of dimension of the three most popular cultural dimensions: Hofstede, Schwartz and Inglehart

Theoretical framework	Name of dimension	Description of dimension
Hofstede	1. Individualism vs Collectivism	1. Considers the extent to which society is integrated into groups, as well as its dependence on them
	2. Power Distance vs Closeness	2. Considers the degree to which certain societies can tolerate inequity or hierarchy in decision-making
	3. Masculinity vs Femininity	3. Considers the degree to which a society supports traditional social roles
	4. Uncertainty Avoidance vs Acceptance	4. Considers the degree of acceptance by members of a society for unstructured or repeated situation
	5. Long-Term Orientation vs Short-Term Orientation	5. Considers the extent to which a society values short-term results or plans for the long-term
	6. Indulgence vs Restraint	6. Considers a society's preference in satisfying primal instincts
Schwartz	1. Harmony	1. Considers the degree to which a society supports integration with the surrounding environment
	2. Embeddedness	2. Considers the degree to which a society manages its relationships as a group
	3. Hierarchy	3. Considers the importance of maintaining an unequal distribution of power.
	4. Mastery	4. Considers a society's emphasis on success, ambition, or competitiveness
	5. Affective Autonomy	5. Considers the degree of desirability for individuals to have a happy independent affective life
Schwartz	6. Intellectual Autonomy	6. Considers the degree to which it is desirable for a society that individuals think independently
	7. Egalitarianism	7. Considers the emphasis of a society on promoting the welfare of others
Inglehart	1. Traditional vs Secular	1. Relate to societies in which religion, authority and family are very important topics
	2. Survival vs Self-expression	2. Relate to societies shaped by existential insecurity and rigid constrains on human autonomy

Source: own elaboration based on Hofstede, Schwartz and Inglehart national values

5.2.2 National cultures

National cultures are described using the cultural dimensions mentioned above. These dimensions provide information about the beliefs (Schwartz, 1999), values, knowledge (Inglehart, 1997) or social norms that individuals from different countries collectively accept (Berry, 1997). To exemplify national values, we will characterize two countries, namely Germany and UK, which are widely recognized across the world.

Table 13. Cultural dimensions of Hofstede, Schwartz and Inglehart: Germany and UK

Name of dimension	Germany	UK
Individualism vs Collectivism	67	89
Power Distance	35	35
Masculinity	66	66
Uncertainty Avoidance	65	35
Long-Term Orientation	83	51
Indulgence vs Restraint	40	69
Harmony	4.62	3.91
Embeddedness	3.03	3.34
Hierarchy	1.87	2.33
Mastery	3.86	4.01
Affective Autonomy	4.11	4.26
Intellectual Autonomy	4.99	4.62
Egalitarianism	5.07	4.92
Traditional vs Secular	.97	.48
Survival vs Self-expression	2.16	2.35

Source: own elaboration based on Hofstede, Schwartz and Inglehart national values

As we can observe in Table 13, most cultural dimensions are quite similar between the two national cultures. They are only dissimilar on some dimensions of Hofstede's values, namely Individualism vs Collectivism, Uncertainty Avoidance, Long-Term Orientation, and Indulgence vs Restraint and from Inglehart's Survival vs Self-expression. In the methodology section, we will construct and present a "toy example" based on these two national cultures.

5.2.3 Supra-national cultures

Several authors have provided theoretical justification for the analysis of supra-national cultures by incorporating the cultural values of each individual, first within national cultures and then aggregating these national cultures into supra-national cultures. Akaliyski et al. (2021) and

Minkov & Hofstede (2012) proved that the aggregation of individual values is an adequate construct to study culture, showing that most of the variability in the cultural values of individuals is explained by their belonging to different supra-national cultures at an empirical level. This statement is supported by Page & Shapiro (2010), whom analyse the variability within and between countries, finding that more than 50% of within countries variability is due to not controlling for spurious variability. Therefore, the uncontrolled variability would be masking differences between countries by assuming that most of the variability comes from one source.

The theoretical concept of supra-national culture has been used since the work of Haire et al. (1966), but it was not until the 1980s, with the pioneering work of Ronen & Shenkar (1985), that supra-national culture values began to be measured, by means of the aggregation of countries' demographic, physical, economic or political characteristics (Akaliyski, 2017). Several authors indicate that in order to create these groups, it is necessary to study the similarity within the characteristics (Saxena et al., 2017; Shenkar, 2001) or the cultural values of the national cultures (Inglehart & Baker, 2000; Schwartz, 1992).

5.2.3.1 Overview of three supra-national cultures

Analyses of national cultures are abundant in the literature, which cannot be said for supra-national cultures, where authors like Inglehart (1997b), Welzel (2012), and Beugelsdijk & Welzel (2018) are the main advocates of this type of research. Therefore, we will provide a brief overview of their supra-national cultures. Inglehart obtained two dimensions after performing a factor analysis on ten questions the World Values Survey (Inglehart, 1997). Welzel (2012) uses a series of external and internal indicators of the WVS to test whether the supra-national cultural cultures he has theoretically created correspond to the values of his questionnaire. Beugelsdijk & Welzel (2018) uses all rounds of the World Values Survey in conjunction with the European Values Survey between the years 1981 and 2014 to select 15 items that adequately represent Hofstede's cultural dimensions, having a correlation of at least 0.5 with at least one of them. Beugelsdijk applies factor analysis to reduce the number of dimensions, identifying 3 dimensions, for which he attempts to measure their evolution across 5 supra-national cultures. In Table 14, we can observe the overview of supra-national cultures.

Table 14. Overview of supra-national cultural groupings.

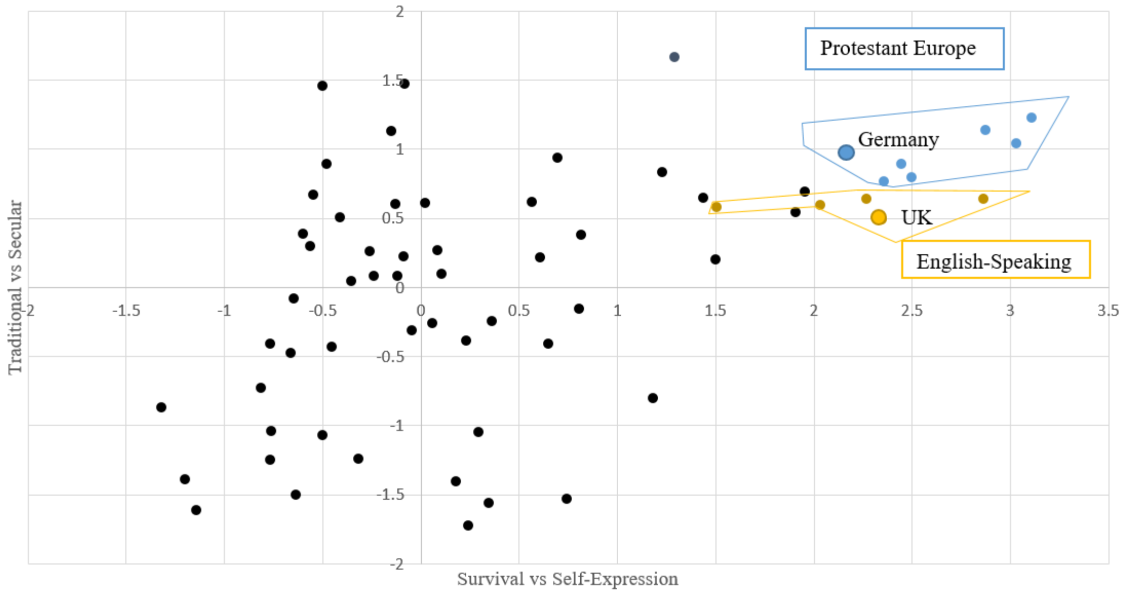
Name of authors of supra-national culture grouping	List of involved factors	Theoretical key points
Inglehart-Welzel (1997)	1. Traditional vs Secular 2. Survival vs Self-Expression	1. Relate to societies in which religion, authority and family are very important topics 2. Relate to societies shaped by existential insecurity and rigid constrains on human autonomy
Welzel (2013)	1. Knowledge Index 2. Society's civil liberties and political rights 3. Society's human rights.	1. Is a summary of society's production of information, communication technology and education 2. Relate to societies degree of individual freedom 3. Relate to the degree of protection of the fundamental human rights
Beugelsdijk (2018)	1. Collectivism-Individualism 2. Duty-Joy 3. Distrust-Trust	1. Relate to societies that place high importance on the relationship within the group 2. Relate to societies in which self-restraint and long-term orientation are encouraged 3. Relate to societies that are not confident on people's intentions

Note: Table depicts name of authors, list of involved factors and theoretical key points

Source: own elaboration based Inglehart, Welzel and Beugelsdijk supra-national cultural groupings

Due to the length constraints of the chapter, we have chosen to include in the methodology section the supra-national culture grouping by Inglehart and Welzel, which is the most validated by the literature, also demonstrating its usefulness in chapter 4, furthermore, it is the only supra-national culture grouping that generates specific cultural clusters. In Figure 27, we can see the Inglehart-Welzel cultural map with an example of supra-national culture grouping for Germany and UK. Here, the authors' cultural dimensions are plotted on a two-dimensional graph, with the Survival vs. Self-Expression dimension on the horizontal axis and the Traditional vs. Secular dimension on the vertical axis. We can observe that there are three types of clusters of countries in this classification (Welzel, 2012); there are clusters that exhibit high values in the Traditional and Survival dimensions, such as the African-Islamic cluster, there are countries with high values in the Secular and Self-Expression dimension, such as Protestant European, English-speaking and some Catholic European countries, and finally there are a majority of countries that are in a transitional phase, from Traditional-Survival values to Secular-Self-Expression values (Inglehart, 1997).

Figure 27. Scatter plot of Inglehart and Welzel's national values



Note: We highlight Protestant Europe and English-speaking clusters in blue and yellow respectively, to which Germany and UK respectively belong

Source: own elaboration based on Inglehart-Welzel Cultural Map

5.3 METHODS

In this study, we aim to investigate the influence of cultural factors on the trade-off phase of the DM process. To do this, we motivate the case study, and then we propose a model that allows us to address the questions raised in the introduction. We analyse the different aggregations of cultural values and indicate the tools and software used.

5.3.1 The choice problem

As in previous chapters, the use case we are going to explore involves explaining the DM process of pilgrims in relation to the Way they have chosen on Saint James Way. The pilgrims interested in Saint James Way face the problem of choosing a unique Way, starting from several points in Spain, France, Portugal, and other European countries to arrive to the destination: the city of Santiago de Compostela. The data was collected from pilgrims that come from forty-eight countries that received the official credential of completing Saint James Way in 2022, certified by Pilgrim's Welcome Office. 420,474 pilgrims were included in the final sample, including seven possible Ways or alternatives: French, Coastal, English, North, Silver, Portuguese, and Primitive that correspond with the most important Ways.

We considered the following features or attributes of the alternatives as relevant to identify each Way: Popularity, Intensity and World Heritage Status, represented on Table 15.

Table 15. Attributes of Saint James Way alternatives, depicting names, values, and descriptions

Name	Values	Description
Popularity	High Medium Low	Number of people who have completed the Way in 2022 coded as three factors.
Intensity	Intense Not Intense	Physical difficulty of completing each Way coded as two factors
World Heritage Status	World Heritage Status Not World Heritage Status	The Way has been granted World Heritage Status by the UNESCO coded as two factors

Source: own elaboration

The popularity of the Way is a feature whose evaluation depends on the pilgrim's mind set. We created the variable with three categories, so that Popularity measures the number of people who have walked each way in 2022. In such a way, that the Ways that have received more than 150,000 pilgrims take value High, if they have received between 150,000 and 50,000 it takes value Medium and takes value Low when these have been less than 50,000.

UNESCO grants special World Heritage credentials to places with exceptional artistic and cultural value. The French and North Ways have achieved this status and could therefore be more attractive, increasing the probability of pilgrims choosing them. World Heritage Status is a binary variable indicating whether the Way have said status.

The intensity of the Way is a main feature that conditions the DM process. Many factors contribute to this variable, such as temperature (Smith, 1993), precipitation (Amelung et al., 2007) and difficulty of the stages (based on average length and slope). We used a set of equations to estimate the variable intensity, which can be seen in ANNEX 1 to generate a dichotomous variable.

Decision-makers also referred as pilgrims in this study are described using the attributes represented on Table 16. Gender is coded as a dichotomous variable with two values, male, and female. Age is coded as a categorical variable with four values (under 18 years, between 18 and 45 years, between 46 and 65 years and over 65 years), selecting the ranges in such that there are a similar number of people between 18 and 45 years old and between 46 and 65 years old, also leaving a similar number of minors and over 65 years, retirement age in Spain. Group is coded as a categorical variable with two values: when the individual arrives at the Pilgrim's Welcome Office and a certificate is issued for more than one pilgrim at the same time, he/she is considered to have completed Saint James Way in a group, he/she is considered to arrive as an individual when this is not the case.

Table 16. Description of the attributes of the decision-maker, depicting names, values, and descriptions.

Name	Values	Description
Gender	Female Male	Gender of the decision maker coded as two factors
Age	<18 years 18-45 years 46-65 years >65 years	Age of the decision maker coded as four factors
Group	Individual Group	Number of individuals that have completed the Way coded as two factors

Source: own elaboration based on Pilgrim's Welcome Office dataset

5.3.2 The choice model

To explain how culture affects individuals' preferences and how these preferences impact their trade-off resolution, we apply multinomial choice models, also known as multinomial model with alternative-specific variables, to see this models further developed, go to section 4.4.4.

Analysis of cultural factors

To ensure the relevance of each cultural dimension in the multinomial model with alternative-specific variables, we perform one model for each of the theoretical frameworks we are interested in analysing, thus we perform three multinomial logit models on the whole sample of individuals, focusing on determining which Hofstede, Schwartz and Inglehart's cultural variables are relevant in solving the trade-off dilemma. We can write a generic model for the tree specifications:

$$\begin{aligned} V_{nj} = \beta_0 + \beta_1 x'_{\text{man}} + \beta_2 x'_{\text{age}} + \beta_3' x'_{\text{group}} + \beta_4' x'_{\text{cultural variables}} \\ + \alpha_1 y_{High\ popularity} + \alpha_2 y_{Medium\ popularity} + \alpha_3 y_{intensity} \\ + \alpha_4 y_{World\ Heritage\ Status} \end{aligned} \tag{25}$$

Analysis of national and supra-national cultures

To further explore how individuals from different national and supra-national cultures resolve the trade-off dilemma in DM problems, we perform a multinomial choice model on each subsample (countries and supra-national cultures), thus focusing on determining whether national cultures and Inglehart cultural clusters, as well as individual attributes, are relevant in solving the trade-off dilemma. We can write a generic model for the specifications, where we delete the cultural variables as explanatory variable and use it as subsampling criteria.

$$\begin{aligned} V_{nj} = \beta_0 + \beta_1 x'_{\text{man}} + \beta_2 x'_{\text{age}} + \beta_3' x'_{\text{group}} + \alpha_1 y_{High\ popularity} \\ + \alpha_2 y_{Mediumpopularity} + \alpha_3 y_{intensity} + \alpha_4 y_{WorldHeritageStatus} \end{aligned} \tag{26}$$

In our case, we will use the Inglehart cultural map grouping to segment each subsample and observe how each of them resolves the trade-off between alternative's attributes, where equation 26 represent the generic equation for each cultural subsample. We use said clustering approach based on the conclusions obtained in chapter 4 and on the popularity of the construct in the literature. Table 17 displays the countries included in the Inglehart cultural map.

Table 17. **Description of the countries within each Inglehart cultural cluster**

Cluster	Countries
West and South Asia	Malaysia, Singapore, Israel, Philippines, Chile, and South Africa. *Mid secular and mid self-expression*
Catholic Europe	Andorra, Malta, Latvia, Estonia, Croatia, Slovenia, Lithuania, Slovakia, Hungary, Austria, Belgium, Czech Republic, Poland, France, Portugal, and Italy *Mid-high secular and mid-high self-expression*
Latin America	Paraguay, Guatemala, Costa Rica, Peru, Puerto Rico, Ecuador, Uruguay, Venezuela, Colombia, Argentina, Mexico, and Brazil *Low secular and mid self-expression*
Orthodox Europe	Ukraine, Bulgaria, Romania, and Russia *Mid secular and low self-expression*
Protestant Europe	Iceland, Norway, Finland, Switzerland, Sweden, Denmark, Netherlands, and Germany *High self-expression and high secular*
Confucian	Taiwan, China, Korea, and Japan *Mid self-expression and high secular*
Destination Country	Spain *Mid-high secular and mid-high self-expression*
English-Speaking	New Zealand, Canada, Australia, Ireland, United Kingdom, and United States *High self-expression and mid-high secular*

Note: We include the destination country (Spain), as a Distinct Group and relation with list of involved factors reviewed on Table 14

Source: own elaboration based on Inglehart-Welzel Cultural Map

5.3.3 Tools and software for data analysis

We performed multinomial choice models using R software in its 4.3.2 version and "mlogit" package 1.1.1. version (Croissant, 2020). Data cleaning was carried out with "tidyverse" (Wickham et al., 2019) 2.0.0. version.

5.4 RESULTS

5.4.1 Relevancy of cultural factors in the choice process

In this section, we present the results of the multinomial logit model for the full sample described in section 5.3.1. of this chapter. As can be seen in Table 18, the effect of cultural variables on the alternatives are presented, determining that all the cultural dimensions from the conceptual frameworks of Inglehart, Hofstede, and Schwartz are significant, justifying their potential inclusion in subsequent analyses.

Table 18. Cultural variables on multinomial logit model

Variables	Coastal	English	North	Silver	Portuguese	Primitive
Inglehart National values						
Survival vs Self-Expression	.125 ***	-.067 ***	.328 ***	-.094 ***	.024 ***	.117 ***
Traditional vs Secular	-.401 ***	.133 ***	.020	.179 ***	-.290 ***	.081 ***
Hofstede National values						
Individualism vs Collectivism	-.012 ***	-.006 ***	.006 ***	-.001	-.021 ***	.001
Power Distance vs Closeness	-.016 ***	-.022 ***	-.002	-.011 ***	-.031 ***	.003 **
Masculinity vs Femininity	.020 ***	-.001	.001	-.013 ***	.006 ***	.001
Uncertainty Avoidance vs Acceptance	-.010 ***	-.002 **	.003 ***	.010 ***	-.009 ***	.001
Long-Term Orientation	-.016 ***	-.012 ***	.014 ***	-.001	-.015 ***	-.003 ***
Indulgence vs Restraint	-.032 ***	-.032 ***	.003 ***	-.013 ***	-.035 ***	-.014 ***
Schwartz National values						
Harmony	2.177 ***	.408 ***	1.109 ***	.110	2.660 ***	.825 ***
Embedded	-.824 ***	-1.560 ***	-.959 ***	-2.360 ***	-1.128 ***	-.185
Hierarchy	-1.336 ***	-.913 ***	-.548 ***	-.021	-1.133 ***	-.437 ***
Mastery	2.170 ***	1.091 ***	-1.122 ***	.631 ***	3.729 ***	-.570 ***
Affective Autonomy	.846 ***	-.487 ***	.880 ***	-.577 ***	.895 ***	.048
Intellectual Autonomy	-2.781 ***	-.878 ***	-1.029 ***	-.483 ***	-2.942 ***	-.464 ***
Egalitarianism	-1.492 ***	.258 ***	-1.648 ***	.487 ***	-.675 ***	-.833 ***

Note: Alternative's coefficients need to be interpreted with French Way as reference

*** Significance 1%; ** Significance 5%; * Significance 10%

Source: own elaboration based on Pilgrim's Welcome Office dataset

5.4.2 Preferences and choices for different national and supra-national cultures

As we have seen in the previous subsection, individuals from different countries make different decisions about alternatives because they have different cultural values, and these cultural values are significant when solving a DM problem with multiple alternatives.

5.4.2.1 Simple example: comparison between two countries (Germany vs UK)

Given that the objective of this chapter is to address the question of whether different cultures solve the trade-off dilemma in the same way and considering that this is a challenging problem, we will first present the feature conflict and trade-off dilemma between two countries. Continuing with the example from Table 13, we make a comparison between Germany and

UK, and we adjust the model from section 5.3.2, to determine if these two countries solve the trade-off between the characteristics of the alternatives differently.

Table 19. Example of feature conflict and trade-off dilemma based on the attributes of the French and Portuguese Way

Attributes / Alternatives	High Popularity	Medium Popularity	Low Popularity	Intensity	World Heritage Status
French Way	1	0	0	1	1
Portuguese Way	0	1	0	0	0

Source: own elaboration

Table 19 shows an example of the feature conflict between French Way and Portuguese Way, where the former offers the attractiveness of its heritage at the cost of an intense experience and high popularity, while the second shows opposite values for intensity and World Heritage Status as well as a lesser popularity.

Table 20. Results of the simple example comparing the coefficients of the multinomial choice model for Germany and UK.

Attributes / Clusters	High Popularity	Medium Popularity	Low Popularity	Intensity	World Heritage Status
Germany	1.278 ***	1.495 ***	Reference	-.936 ***	1.118 ***
UK	2.143 ***	.595 ***	Reference	.030	-.435 **

Note: The coefficients of the attributes of the alternatives and decision-makers can be verified on Table 37, Table 38 and Table 39 in the Annex

*** Significance 1%; ** Significance 5%; * Significance 10%

Source: own elaboration based on Pilgrim's Welcome Office dataset

Analysing the results of the simple example from Table 20, we can see that the tree alternative's variables; Popularity, Intensity, and World Heritage Status constitute the attributes that each of Saint James Ways possesses, where the sign and coefficient of these variables indicate the preference of German and British pilgrims for each of the attributes of the Way. We can observe significant differences between German and British pilgrims: while Germans prefer Ways with moderate popularity, the British prefer highly popular Ways; regarding intensity, Germans dislike it, whereas the British are indifferent to it; lastly, Germans prefer Ways that have World Heritage Status, whereas the British negatively value this characteristic.

The choice models reveal that different countries show different preferences over the Way's attributes, where the values are the result of the different preferences of each country over the attributes of the alternatives. Such that in Germany, the negative preference for intensity and the preference for medium popularity compensate for their preference for World Heritage Status in such a way that they prefer the Portuguese Way over the French Way. Conversely, the opposite can be said for the British, as the preference for highly popular ways dominates over the insignificant effect of intensity and the negative preference for the World Heritage Status as can be seen on Table 21.

This simple example highlights the effectiveness of employing supra-national culture instead of national values to elucidate differences in individual preferences. Despite Germany and the UK sharing similar national values, they fall into distinct cultural clusters. Consequently, it becomes evident that the disparities in preferences among individuals from these countries find better explanation through supra-national culture rather than their respective national cultures.

Table 21. Empirical choice frequencies for Germany and UK

Country \ Alternatives	French Way	Portuguese Way
Germany	.305	.395
UK	.535	.191

Note: Empirical choice frequencies are the actual observed choice frequencies in the database

Source: own elaboration

5.4.2.2 Comparison between countries

To delve deeper into how people in different countries deal with the trade-off between different alternatives, we present on Table 22, the feature conflict and trade-off dilemma for the seven alternatives, that we use to compare how the top 10 countries by number of pilgrims in 2022 resolve their trade-off.

Table 22. Feature conflict and trade-off dilemma based on the attributes of the seven main ways

Attributes / Alternatives	High Popularity	Medium Popularity	Low Popularity	Intensity	World Heritage Status
French Way	1	0	0	1	1
Coastal Way	0	0	1	0	0
English Way	0	0	1	0	0
North Way	0	0	1	1	1
Silver Way	0	0	1	1	0
Portuguese Way	0	1	0	0	0
Primitive Way	0	0	1	1	0

Source: own elaboration

The values of the attributes for the seven Ways are shown on Table 22, the French Way is the most popular Way (High Popularity), followed by the Portuguese Way (Medium Popularity), the North Way offers the attractiveness of its heritage as the French Way, but at the cost that they both have High intensity, same as Silver and Primitive Way.

Table 23. Coefficients of countries of the multinomial choice model

Attributes / Countries	High Popularity	Medium Popularity	Low Popularity	Intensity	World Heritage Status
Spain	2.270 ***	.869 ***	Reference	-.214 ***	.049 *
Italy	3.021 ***	.910 ***	Reference	-.973 ***	-.198 *
USA	2.297 ***	.509 ***	Reference	-1.112 ***	.814 ***
Germany	1.278 ***	1.495 ***	Reference	-.936 ***	1.118 ***
Portugal	2.262 ***	1.704 ***	Reference	-1.447 ***	-1.397 ***
France	1.294 ***	.784 ***	Reference	-.651 ***	1.921 ***
UK	2.143 ***	.595 ***	Reference	.030	-.435 **
Ireland	2.840 ***	1.378 ***	Reference	-.544 **	.169
Mexico	1.886 ***	.701 ***	Reference	-.269	1.052 ***
South Korea	4.023 ***	.965 ***	Reference	-1.228 ***	1.135 **

Note: The coefficients of the attributes of the alternatives and decision-makers can be verified on Table 37, Table 38 and Table 39 in the Annex

*** Significance 1%; ** Significance 5%; * Significance 10%

Source: own elaboration based on Pilgrim's Welcome Office dataset

As can be seen on Table 23 Ireland is indifferent to the World Heritage Status, while Italy, UK and Portugal dislike the Ways that have this status. UK and Mexico are indifferent to the intensity of the way, while the rest of countries dislike the intensity of the Way. All the countries prefer High or Medium popularity Ways over low popularity Ways, however, Germans prefer medium popularity Ways over High popularity Ways.

Table 24. Empirical choice frequencies for top 10 countries

	French	Coastal	English	North	Silver	Portuguese	Primitive
Spain	.583	.049	.069	.046	.032	.164	.057
Italy	.522	.053	.079	.041	.014	.214	.047
USA	.578	.097	.030	.044	.009	.207	.035
Germany	.305	.119	.042	.084	.015	.395	.040
Portugal	.088	.152	.052	.008	.028	.649	.024
France	.685	.036	.012	.105	.025	.105	.033
UK	.535	.081	.074	.086	.031	.191	.043
Ireland	.608	.087	.054	.030	.007	.188	.027
Mexico	.714	.053	.025	.048	.009	.107	.046
South Korea	.856	.021	.007	.026	.009	.073	.008

Note: Empirical choice frequencies are the actual observed choice frequencies in the database

Source: own elaboration based on Pilgrim's Welcome Office dataset

Observing the choice probabilities on Table 24, it is clear that the preferences on Table 23 reflects reality, given that the actual probability of choosing each of the alternatives for each cluster, can be deduced by taking the exponential of the data from said table, and calculating the estimated probabilities of choice for each alternative by each country.

5.4.2.3 Comparison between supra-national cultures

Once we have verified that countries resolve the trade-off differently, we need to determine whether supra-national cultures, also address the trade-off dilemma in a distinct manner, standing out Inglehart and Welzel approach as the most recognized and supported on the study of supra-national cultures. We adjust the model from section 0 to determine if these eight clusters resolve the trade-off between the characteristics of the alternatives in the same way.

Table 25. Coefficients of supra-national clusters of the multinomial choice model

Attributes / Clusters	High Popularity	Medium Popularity	Low Popularity	Intensity	World Heritage Status
West and South Asia	2.130***	.523***	Reference	-.869***	.478*
Catholic Europe	2.284***	1.302***	Reference	-1.236***	.057
Latin America	2.430***	.831***	Reference	-.512***	.582***
Orthodox Europe	2.809***	.790***	Reference	-.840***	-.450
Protestant Europe	1.527***	1.322***	Reference	-.958***	.998***
Confucian	3.231***	.751***	Reference	-.425	.853***
Spain	2.272***	.870***	Reference	-.210***	.046*
English-Speaking	2.327***	.817***	Reference	-.643***	.414***

Note: The coefficients of the attributes of the alternatives and decision-makers can be verified on Table 36 in the Annex

*** Significance 1%; ** Significance 5%; * Significance 10%

Source: own elaboration based on Pilgrim's Welcome Office dataset

We can see on Table 25 that Catholic and Orthodox Europe cluster, share a preference for Ways that are highly popular, yet have low intensity, while also valuing less the World Heritage Status. Protestant Europe cluster does not have a clear preference for high or medium popularity Ways but shares the dislike for intensity and values the inclusion of World Heritage Status. The Confucian cluster exhibits a preference for Ways with high popularity, while also valuing the World Heritage Status. Similarly, the Spanish cluster leans towards high popularity Ways, downplaying the importance of intensity and appreciating to a lesser extend World Heritage Status. Lastly, the West and South Asia, English-Speaking and Latin America cluster shows a preference for high popularity Ways, a dislike for intensity, and an active pursuit of destinations with World Heritage Status.

Table 26. Empirical choice frequencies for Inglehart clusters

	French	Coastal	English	North	Silver	Portuguese	Primitive
West and South Asia	.541	.123	.029	.039	.013	.216	.039
Catholic Europe	.377	.113	.051	.048	.020	.345	.045
Latin America	.634	.063	.027	.037	.008	.188	.043
Orthodox Europe	.375	.120	.053	.074	.015	.290	.074
Protestant Europe	.364	.101	.042	.079	.019	.354	.042
Confucian	.792	.029	.015	.036	.012	.093	.023
Spain	.583	.049	.069	.046	.032	.164	.057
English-Speaking	.563	.092	.040	.045	.013	.212	.035

Note: Empirical choice frequencies are the actual observed choice frequencies in the database

Source: own elaboration based on Pilgrim's Welcome Office dataset

By examining the choice probabilities presented in Table 26 it becomes evident that the preferences outlined in Table 25 reflects reality. This assertion is supported by the fact that the actual probability of selecting any given alternative within each cluster can be derived by taking the exponential of the data from Table 25, thus calculating the probabilities of choice for each alternative within every cluster.

5.5 CONCLUSIONS

Considering our results, we can confirm that all the cultural dimensions of the three most used conceptual frameworks in cross-cultural analysis are relevant in determining why individuals choose one alternative over another. We have been able to verify that individuals are influenced by their cultural values, requiring the inclusion of as many cultural variables as possible that have a significant effect on the DM system , as can be seen in the results of section 5.4.1., not only including individualism vs collectivism dimension.

Comparing the results of the multinomial choice model among different national cultures, a series of differences in terms of preferences have been identified. Thus, countries in southern Europe and the British Isles show a lack of interest in Ways that have World Heritage Status, unlike the rest of the analysed national cultures. The United Kingdom and Mexico are indifferent to the intensity of the Way, whereas the remaining national cultures dislike this characteristic. Unlike the other compared countries, Germans prefer Ways of medium popularity instead of highly popular ones. It can be observed that individuals from different national cultures use all the variables of the alternatives to solve the trade-off problem, apart from the UK and Mexico regarding intensity, and Ireland regarding World Heritage Status, as they are indifferent to the presence of these characteristics in the alternatives.

By comparing the results of the multinomial choice model among various supra-national cultures, a range of preference differences has been identified. The Catholic Europe, Orthodox Europe, and English-Speaking cluster exhibit a preference for highly popular Ways with low intensity, while placing less emphasis on the inclusion of World Heritage Status. Protestant Europe cluster does not display a clear preference for Ways of high or medium popularity but shares the aversion to intensity and values the inclusion of World Heritage Status. The Confucian cluster shows a preference for highly popular Ways without considering intensity, while also valuing World Heritage Status. Similarly, the Spanish cluster leans towards highly popular Ways, downplaying the importance of intensity and placing a somewhat lesser emphasis on World Heritage Status. Lastly, the West and South Asia and Latin America cluster demonstrates a preference for highly popular Ways, a dislike for intensity, and actively pursues destinations with World Heritage Status. It can be observed that individuals from various supra-national cultures utilize all the variables of the alternatives to address the trade-off problem, except for the Catholic Europe and Orthodox Europe clusters in relation to World Heritage Status, and the Confucian cluster concerning intensity, as they display indifference towards the presence of these attributes in the alternatives. This is consistent with what was observed in the model comparing national cultures.

Our work shed light on the debate surrounding choice between national or supra-national cultures frameworks at explaining DM and specifically, the trade-off dilemma phase, even though both concepts are useful depending on the research question the advantages of the supra-national cultures become apparent:

Information: Using the principle of maximum entropy, which consists of representing the grouping of individuals through the probability distribution that provides the highest entropy (Guiasu & Shenitzer, 1985), resulting more neutral and less biased choice. We calculate the average information entropy of supra-national cultures and national cultures over the probability of selecting each route of Saint James Way, and we observe a value of 1.91 for the former and a value of 1.76 for the latter, indicating a greater abundance of information within the supra-national clusters, thus justifying the use of de supra-national culture by the principle of maximum entropy.

Sample: Having larger groups increases the likelihood of having sufficient data to calculate models such as multinomial logit with alternative specific variables, allowing for a deeper understanding of the trade-off phase of DM.

We have been able to verify that all cultural dimensions are relevant in determining the alternative chosen by individuals. Additionally, individuals from different national cultures and supra-national cultures have revealed distinct preferences and resolutions for the trade-off, which justifies the use of a model like multinomial choice models that allow us to study said trade-off whenever possible. From this, we can conclude that the most appropriate unit of analysis will depend on the research question. Analysing national cultures allows us to study differences between countries, while supra-national cultures are useful when studying differences across broader regions. Let’s use an example of the utility of the technique developed in this chapter; we could have a tourist, to whom we plan to recommend a product, thus we develop a recommender system, however, there might not be enough customers from certain countries to make it meaningful to use this feature in a model. However, we could use the supra-national culture they belong to for making recommendations. This could be the case for tourists from Switzerland or New Zealand in Saint James Way, who are crucial due to their high daily spending and length of stay, but their numbers are not abundant enough to build a separate model for each country. Thus, the level of granularity required for the analysis will depend on the scope of the analysis posed by business, policymakers, and researchers.

CHAPTER 6. COMPARING THE PERFORMANCE OF CHOICE-BASED MODELS: A CULTURAL EVALUATION

6.1 INTRODUCTION

In this chapter we compare several DM paradigms such as discrete choice models and tree-based models in order to select the best techniques for implementing a recommendation system. We will consider flexibility, computational complexity, and interpretability, on each of the 10 combinations of measurements of culture on three cultural constructs developed by Inglehart, Hofstede and Schwartz.

The chapter is organized as follows: in Section 2, we provide a literature review on DM models. In Section 3, we explain the case study, describing cultural constructs and cultural measures. In Section 4, we characterize the methodology, including the models and evaluation measures. In Section 5 we analyse the feature performance and performance of the models. Finally, in Section 6, we present the discussion of the chapter.

6.2 LITERATURE REVIEW

In the field of DM, there are several paradigms or approaches that are used to address different problems. One of the main ones, which we will develop in this section, is the study of the recommendation problem, as a choice prediction task (Almomani et al., 2022; Saavedra et al., 2016).

6.2.1 Discrete choice models

Ben-Akiva & Lerman (1985) and Mcfadden (1973,1980) explored choice models, examining how users select items based on their interactions. Under this framework, a discrete choice model is a model that maximizes the utility provided by each item for each user, aiming to predict the probability that a user chooses a specific item (Train, 2009). To fit in a discrete choice framework, the set of items needs to have three characteristics: mutually exclusivity, completeness, and finitude of items.

Multinomial model: is an extension on the basic logistic regression model, used for prediction in cases where the variable to be explained has a categorical multinomial distribution, rather than a dichotomous distribution.

Multinomial model with item specific variables (choice): is an extension of the multinomial model, where variables that determine the characteristics of the items are incorporated, facilitating justification of the users' choice of items.

6.2.2 Machine learning models

Machine learning is a branch of artificial intelligence that involves the development of algorithms and models that enabling computers to learn patterns from data and make decisions or predictions based on the optimization of a performance criterion (Alpaydin, 2010). The techniques associated with this branch of artificial intelligence has experienced significant development thanks to the new computational capabilities achieved in the last 20 years, which have allowed their widespread use (Cortes & Vapnik, 1995; Hochreiter & Schmidhuber, 1997).

Unlike the more open-ended purpose of machine learning models, the main objective of discrete choice models is the analysis of discrete decisions, such as consumer preferences or user behaviour, requiring the specification of a utility function that make them more interpretable specially in relation to complex machine learning models usually known as "black boxes" (Cortez & Embrechts, 2013). While machine learning models are generally more adaptable (Gareth et al., 2021), especially when nonlinear relations arise, not requiring explicit specification of the underlying DM process, discrete choice models provide a more explicit and interpretable structure for addressing specific decision-making problems.

In this section, we explain how the three most popular types of machine learning models work: support vector machines, neural networks, and tree-based models.

6.2.2.1 Support Vector Machines

The Support Vector Machine (SVM) is a supervised learning technique used for classification and regression tasks (Gareth et al., 2021). The main idea behind SVM is to find a hyperplane in a high-dimensional space that can separate the maximum number of different classes or categories while maximizing the margin of the hyperplane (Alpaydin, 2010; Cortes & Vapnik, 1995; Guyon & Elisseeff, 2003).

6.2.2.2 Neural networks

A neural network is a machine learning algorithm, that can be trained using both supervised and unsupervised data, employed in pattern recognition, classification or regression problems (Alpaydin, 2010). The design of a neural network is inspired by the structure of a brain, comprising layers of interconnected nodes. It processes information by adjusting weights during training to optimize predictions (Gareth et al., 2021; Rosenblatt, 1958).

6.2.2.3 Tree-based models

Decision tree

A tree-based model is a supervised machine learning algorithm used for classification and regression purposes, that recursively segments the predictor variable into a number of regions based on the explanatory variables that lead to a tree structure. It serves to predict the mean or mode for the observations in each region, divided usually by high-dimensional rectangles (Gareth et al., 2021).

Tree-based ensembles

Single tree-based models have one main advantage over other models, namely interpretability, however, ensembles overcome several of their weaknesses. An ensemble is a technique that combines or aggregates several single trees to obtain a model that can improve both performance and robustness, by reducing bias, variance, and overfitting (Parmar et al., 2019).

Boosting: is an ensemble technique that combine several trees, in such a way that trees are grown sequentially, based on information from previously grown trees, resampling the original data (Polikar, 2006; Ricci et al., 2022).

Bagging: is an ensemble technique that involves producing several trees using several bootstrapped training sets, producing complete trees that are finally averaged (Gareth et al., 2021).

Random forest: is an ensemble technique that selects random variables from the dataset and performs a bootstrap to generate subsamples, which are finally averaged.

6.3 CASE STUDY

This chapter presents a case study that relates DM of pilgrims (users) to the chosen pilgrimage Way (items). The case study offers an in-depth comparison of several classification methods that originate from discrete choice modelling and machine learning methods, in relation with several cultural measures.

The methodology and results section contains the description of the dataset, the variables used to calculate the models in each of the chosen methodologies and the evaluation of the models.

Therefore, our goal is to compare several models regarding their predictive ability concerning which way each individual in the dataset will attempt to complete, using common techniques in the field of classifiers, such as those based on random utility models, trees, and their variations. The specifications of the models we'll compare are selected based on the cultural variables we'll introduce into each of them, segmenting these variables by primary cultural constructs and cultural measures.

Due to the characteristics of our case study, where we aim to compare the predictive capacity of various models for specifications distinguished by the cultural construct utilized and its measure, while maintaining the same control variables for individuals, we will explore both the cultural constructs and the specific measures we will employ to quantify each of them when incorporating them into the models.

6.3.1 Cultural constructs

Before explaining what an empirical cultural construct is, it is worth considering the definition of culture in the first place. The definition of culture is one of the most debated topics within the field of social sciences, as evidenced in several comprehensive compilations of its definitions such as those found in (Baldwin et al., 2006) or (Kroeber & Kluckhohn, 1952). Therefore, an operational definition of culture would have to combine several of the most common definitions of culture into one, such as stating that culture is a collective phenomenon (Hofstede et al., 2005) consisting on the set of beliefs (Schwartz, 1999), values, knowledge (Inglehart, 1997) and norms, accepted by a social group that closely relates (Berry, 1997) to geographical, climatic and socio-economic factors (Akaliyski et al., 2021; Minkov & Hofstede, 2013; Schwartz, 2014; Solomon & Panda, 2004; Vijver & Leung, 1997).

Empirical cultural constructs refer to systematically measurable and quantifiable elements within cultures, both at the national and supra-national levels, these constructs are derived from theories or conceptual frameworks aimed at quantitatively analysing diverse cultures. Among the variety of cultural constructs that exist, we will focus on three of the most popular ones: those of Hofstede (1980), Schwartz (1994), and Inglehart (1997b). For further details on their definitions, refer to Table 12 and section 5.2.1.

6.3.2 Cultural measures

Once the cultural constructs to be analysed have been described, we need to outline the specific cultural measures that will be used to empirically analyse them. Three types of analyses will be employed, namely national values analysis, supra-national values analysis, and cultural distance analysis.

6.3.2.1 National values

National values are constructed as the aggregation to the national population of the values of each of the individuals surveyed by the creators of the cultural constructs. Thus, a national value provides country average information on beliefs (Schwartz, 1999), values, knowledge (Inglehart, 1997) or social norms that individuals from different countries collectively accept (Berry, 1997). The use of national values is empirically straightforward as it doesn't require any intermediate calculations. Instead, it only requires inputting the specific values directly into each model and specification that we aim to test.

6.3.2.2 Supra-national culture

In chapter 1 and 2 we've referenced several articles recommending the use of supra-national cultures and we reaffirmed said utility with our discussion and conclusions on the chapters. Thus, the use of supra-national cultures such as Inglehart's, which is the one we will use in the chapter, enables us to divide the world into large areas of study, eliminating the need to include country control variables as in national cultures, simply by assigning the cultural cluster to which each individual belongs. For further details check Table 17.

6.3.2.3 Cultural distance

Cultural distance serves as a widely used measure that serves as a proxy of cultural disparities (Beugelsdijk et al., 2018) between countries, which introduce risk and uncertainty for consumers and suppliers of tourism goods and services, altering the DM process (Crotts, 2004; Hsu et al., 2013; Litvin et al., 2004; H. Liu et al., 2018; Pizam & Sussmann, 1995). Thus, empirically, to employ cultural distance, the standard procedure involves obtaining national values and calculating a measure of cultural distance, where the most used ones are those by Kogut and Singh distance and the Euclidean distance. Authors like Kogut & Singh (1988) and Ronen & Shenkar (1985) justify the use of this measure, arguing that simplifies the problem of analysing specific definitions of cultural values, making easier the examination of cultural differences by focusing on a single proxy factor of risk.

Empirically, the Kogut and Singh formula is a derivative of the Mahalanobis distance (Kandogan, 2012; Mahalanobis, 1936) between the cultural values of countries, adjusted for variance, assuming a covariance of 0, and attributing equal weight to each value. Mahalanobis distance is usually measured for each pair of countries that we want to compare, thus, given the characteristics of our research question, we only need to compute all cultural distances from each country in our database to Spain, which is the destination country.

Following Kogut and Singh's original article, we can create their distance:

$$KS_{ij} = \frac{1}{n}\sum_{d=1}^{n}\frac{\left(I_i^d - I_j^d\right)^2}{V^d} \tag{27}$$

where KS_{ij} is the Kogut and Singh distance between country i and country j, I_i^d is the index of the cultural value d for country i, I_j^d is the index of the cultural value d for country j and $V^d$is the variance of the index in the cultural value d.

To calculate the Euclidean distance, which is the simplest way to measure the distance between two points in a two-dimensional space, we would need to follow the following formula:

$$EUC_{ij} = \sqrt{\sum_{i=1}^{n} \left(I_i^d - I_j^d\right)^2} \tag{28}$$

where EUC_{ij} is the Euclidean distance of the value index between country i and country j, and I_i^d is the index of the cultural value d for country i and I_j^d is the index of the cultural value d for country j. To check cultural distance values for each country check Table 42.

6.3.3 Dataset

The dataset we'll use on the case study corresponds to data from the year 2022 collected by the Pilgrim's Welcome Office, which includes 420,474 pilgrims who have completed Saint James Way to the city of Santiago de Compostela. Due to the dataset's structure, which doesn't allow us to identify whether pilgrims have completed the journey more than once, the dataset is only suitable for use with collaborative filtering, not content-based filtering.

6.4 METHODOLOGY

The methodology section presents a comparative analysis of several classification models, examining their performance on 10 different specifications encompassing 3 cultural constructs and 4 cultural measures. Our goal is to discern which classification models perform the best considering its limitations, flexibility, and interpretability. Table 27 presents the 10 specifications proposed, changing the combination of cultural construct with the level of cultural aggregation, so that the different specifications can be compared using the different models.

Table 27. Specifications by cultural construct and measure

Specifications	Cultural construct	Measure
Specification 1	Inglehart	Kogut and Singh distance
Specification 2	Inglehart	Euclidean distance
Specification 3	Inglehart	National values
Specification 4	Inglehart	Cultural cluster
Specification 5	Hofstede	Kogut and Singh distance
Specification 6	Hofstede	Euclidean distance
Specification 7	Hofstede	National values
Specification 8	Schwartz	Kogut and Singh distance
Specification 9	Schwartz	Euclidean distance
Specification 10	Schwartz	National values

Source: own elaboration

6.4.1 Models

Once the objective of the methodology has been presented, we proceed to develop the models included in the literature review, such as those belonging to the discrete choice models family, simple tree-based models, and tree-based ensembles.

6.4.1.1 Discrete choice models

Discrete choice models have been developed in Chapter 4 of this thesis 'Supra-national culture and tourist's decision-making', so for a detailed implementation of these models, it is recommended to refer to section 4.4.4.

6.4.1.2 Tree-based models

A tree-based model is a type of machine learning model that recursively segments the predictor variable into regions based on explanatory variables, creating a tree-like structure. This structure allows predicting the mode of these regions in classification problems.

Classification tree

Following Gareth et al. (2021), we have a dependent variable Y that has several categories that we want to classify. We divide the predictor space $X = X_1, X_2, \dots, X_i, \dots, X_n$ (set of possible explanatory variables) into J non-overlapping regions $R_1, R_2, \dots R_j, \dots, R_J$. For each observation within region R_j, we make the same prediction, determined by the mode of response values within R_j of the dependent variable Y. In classification trees, the easiest criterion used for making binary splits is the classification error rate, that represents the percentage of categories of the dependent variable that have been misclassified by imposing the mode as the prediction. However, this classification error rate is not sensitive enough for tree growth. Therefore, we use an alternative called the Gini index, defined as the total variance across all classes of the variable to predict. A small value in a node indicates that observations predominantly belong to the same category. This Gini index can be written as:

$$G = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk}) \tag{29}$$

where $\hat{p}_{mk}$ represents the proportion of observations in the m th region and Gini index is a measure of total variance across the K classes.

Pruning process aims to address the potential issue of overfitting to the data, which could result in poor performance, as the resulting classification tree from the process might become excessively large. If this occurs, removing nodes and branches from the classification tree can lead to reduced overfitting, reducing the variance at the expense of introducing some bias. In our case study, to choose the pruning parameters for the models, we opt to tune the alpha parameter, which enables us to implement cost complexity pruning (Gareth et al., 2021). By testing various complexity parameters, we select the one that maximizes the prediction accuracy, given a series of alpha parameters.

Boosting

Boosting aims to enhance the predictive performance of models by sequentially training weak learners or simpler trees that collectively outperform a larger single tree. In this specific application, the trees grow sequentially, with each tree containing information from previously created trees, rather than adjusting data to multiple datasets created with bootstrap, as seen in bagging technique. In this case, we fit a decision tree to the residuals of the model, and recursively add these new decision trees into the fitted function to update the residuals (Almomani et al., 2020; Gareth et al., 2021). This process involves assigning higher weights to incorrectly predicted instances, enabling subsequent models to prioritize and rectify these errors, ultimately refining the overall predictive accuracy.

Bagging

Bagging aims to enhance the predictive performance of classification trees by constructing a large number of trees based on multiple datasets created using bootstrap (Breiman, 1996; Polikar, 2006). By combining the results of these unpruned trees, each built on a different sample, it can be observed that they exhibit high variability but low bias. When performing the aggregation or averaging of classifications, the objective is to maintain this low bias without significantly increasing it.

Random Forest

Random Forest aims to enhance the predictive performance of classification trees like bagging by generating numerous trees based on multiple datasets created using bootstrap. However, in Random Forest, a distinguishing factor is that a random subset of predictors (typically the square root) is chosen as split candidates from the full set of predictors. This method differs from bagging, where when a dataset is relatively homogeneous, the trees tend to be very similar due to the strong predictors being likely the same variables. In contrast, Random Forest forces each split to consider a fraction of these predictors, reducing the correlation between trees (Almomani et al., 2022). This approach ensures that the average of the trees is even more reliable as they compensate better among themselves.

6.5 RESULTS

In this results section, we delve into the analysis of the feature importance of the specifications and machine learning models, we perform a grid search to find optimal parameters to finally analyse the performance of each specification and model with accuracy and f1 score metrics.

6.5.1 Feature importance

As a first metric to compare the ten proposed specifications for each of the tree-based models we chose to represent the feature importance.

Table 28. Feature importance of the classification tree for each specification

	Variables		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
			Specifications									
Culture	DC KS Inglehart		.70	-	-	-	-	-	-	-	-	-
	DC EUC Inglehart		-	.70	-	-	-	-	-	-	-	-
	Inglehart National Values	Surv_Self	-	-	.21	-	-	-	-	-	-	-
		Trad_Sec	-	-	.50	-	-	-	-	-	-	-
	Cluster Inglehart		-	-	-	.65	-	-	-	-	-	-
	DC KS Hofstede		-	-	-	-	.71	-	-	-	-	-
	DC EUC Hofstede		-	-	-	-	-	.71	-	-	-	-
	Hofstede National Values	PDI	-	-	-	-	-	-	.01	-	-	-
		IDV	-	-	-	-	-	-	.28	-	-	-
		MAS	-	-	-	-	-	-	.01	-	-	-
		UAI	-	-	-	-	-	-	.30	-	-	-
		LTOWVS	-	-	-	-	-	-	.09	-	-	-
		IVR	-	-	-	-	-	-	.03	-	-	-
	DC KS Schwartz		-	-	-	-	-	-	-	.72	-	-
	DC EUC Schwartz		-	-	-	-	-	-	-	-	.72	-
	Schwartz National Values	Harmony	-	-	-	-	-	-	-	-	-	.03
		Embedded	-	-	-	-	-	-	-	-	-	.01
		Hierarchy	-	-	-	-	-	-	-	-	-	.05
		Mastery	-	-	-	-	-	-	-	-	-	.10
		Affective Autonomy	-	-	-	-	-	-	-	-	-	.06
		Intellectual Autonomy	-	-	-	-	-	-	-	-	-	.04
		Egalitarianism	-	-	-	-	-	-	-	-	-	.42
Gender	Man	-	.04	.03	.02	.04	.02	.02	.02	.04	.02	.03
	Woman	-	.02	.02	.03	.02	.03	.03	.03	.02	.02	.02
Group	Group	-	.03	.01	.03	.02	.02	.02	.04	.02	.02	.02
	Individual	-	.02	.04	.03	.05	.03	.03	.01	.02	.02	.04
Age	18-45	-	.04	.04	.04	.04	.03	.03	.04	.03	.03	.04
	46-65	-	.02	.02	.02	.01	.01	.02	.02	.01	.02	.02
	>65	-	.02	.02	.02	.03	.02	.02	.02	.02	.02	.02
	<18	-	.04	.05	.04	.05	.04	.05	.04	.04	.05	.04
Mean of transport	Not Walking	-	.01	.01	.05	.01	.02	.02	.05	.01	.01	.02
	Walking	-	.06	.06	.01	.09	.06	.06	.01	.07	.07	.06

Source: own elaboration

Table 29. **Feature importance of the boosting and bagging for each specification**

	Variables		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
			Specifications									
Culture	DC KS Inglehart		.91	-	-	-	-	-	-	-	-	-
	DC EUC Inglehart		-	.90	-	-	-	-	-	-	-	-
	Inglehart National Values	Surv_Self	-	-	.54	-	-	-	-	-	-	-
		Trad_Sec	-	-	.36	-	-	-	-	-	-	-
	Cluster Inglehart		-	-	-	.88	-	-	-	-	-	-
	DC KS Hofstede		-	-	-	-	.90	-	-	-	-	-
	DC EUC Hofstede		-	-	-	-	-	.90	-	-	-	-
	Hofstede National Values	PDI	-	-	-	-	-	-	.16	-	-	-
		IDV	-	-	-	-	-	-	.09	-	-	-
		MAS	-	-	-	-	-	-	.09	-	-	-
		UAI	-	-	-	-	-	-	.09	-	-	-
		LTOWVS	-	-	-	-	-	-	.18	-	-	-
		IVR	-	-	-	-	-	-	.29	-	-	-
	DC KS Schwartz		-	-	-	-	-	-	-	.91	-	-
	DC EUC Schwartz		-	-	-	-	-	-	-	-	.90	-
	Schwartz National Values	Harmony	-	-	-	-	-	-	-	-	-	.10
		Embedded	-	-	-	-	-	-	-	-	-	.09
		Hierarchy	-	-	-	-	-	-	-	-	-	.06
		Mastery	-	-	-	-	-	-	-	-	-	.18
		Affective Autonomy	-	-	-	-	-	-	-	-	-	.19
		Intellectual Autonomy	-	-	-	-	-	-	-	-	-	.09
		Egalitarianism	-	-	-	-	-	-	-	-	-	.19
Gender	Man	-	.01	.01	0	.02	0	0	0	0	.01	.01
	Woman	-	0	0	.01	-	.01	.01	.01	.01	0	0
Group	Group	-	.02	.02	0	-	.01	.01	0	.01	0	.01
	Individual	-	0	0	.02	.02	.01	.01	.02	.01	.02	.01
Age	18-45	-	.01	.01	.01	.01	.01	.01	.01	.01	.01	0
	46-65	-	.01	.01	.01	.01	.01	.01	.01	0	.01	.01
	>65	-	.01	.01	.01	.02	.02	.02	.01	.02	.02	.02
	<18	-	.01	.02	.02	.02	.01	.01	.02	.01	.01	.02
Mean of transport	Not Walking	-	.02	.01	0	.01	.01	.01	.01	.01	.01	.02
	Walking	-	0	.01	.02	.01	.01	.01	.01	.01	.01	0

Source: own elaboration

Table 30. Feature importance of the random forest for each specification

			Specifications									
	Variables		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Culture	DC KS Inglehart		.90	-	-	-	-	-	-	-	-	-
	DC EUC Inglehart		-	.90	-	-	-	-	-	-	-	-
	Inglehart National Values	Surv_Self	-	-	.43	-	-	-	-	-	-	-
		Trad_Sec	-	-	.44	-	-	-	-	-	-	-
	Cluster Inglehart		-	-	-	.83	-	-	-	-	-	-
	DC KS Hofstede		-	-	-	-	.90	-	-	-	-	-
	DC EUC Hofstede		-	-	-	-	-	.90	-	-	-	-
	Hofstede National Values	PDI	-	-	-	-	-	-	.07	-	-	-
		IDV	-	-	-	-	-	-	.14	-	-	-
		MAS	-	-	-	-	-	-	.13	-	-	-
		UAI	-	-	-	-	-	-	.22	-	-	-
		LTOWVS	-	-	-	-	-	-	.12	-	-	-
		IVR	-	-	-	-	-	-	.13	-	-	-
	DC KS Schwartz		-	-	-	-	-	-	-	.90	-	-
	DC EUC Schwartz		-	-	-	-	-	-	-	-	.90	-
	Schwartz National Values	Harmony	-	-	-	-	-	-	-	-	-	.10
		Embedded	-	-	-	-	-	-	-	-	-	.11
		Hierarchy	-	-	-	-	-	-	-	-	-	.12
		Mastery	-	-	-	-	-	-	-	-	-	.15
		Affective Autonomy	-	-	-	-	-	-	-	-	-	.08
		Intellectual Autonomy	-	-	-	-	-	-	-	-	-	.07
		Egalitarianism	-	-	-	-	-	-	-	-	-	.17
Gender	Man	-	.01	.01	.01	.01	.01	.01	.02	.01	.01	.02
	Woman	-	.01	.01	.01	.01	.01	.01	.02	.01	.01	.02
Group	Group	-	.01	.01	.01	.01	.01	.01	.02	.01	.01	.02
	Individual	-	.01	.01	.01	.01	.01	.01	.02	.01	.01	.02
Age	18-45	-	.01	.01	.02	.02	.01	.01	.02	.01	.01	.02
	46-65	-	.01	.01	.01	.01	.01	.01	.02	.01	.01	.02
	>65	-	.01	.01	.01	.02	.01	.01	.02	.01	.01	.02
	<18	-	.02	.03	.03	.04	.02	.02	.03	.02	.02	.03
Mean of transport	Not Walking	-	.01	.01	.01	.01	.01	.01	.02	.01	.01	.02
	Walking	-	.01	.01	.01	.02	.01	.01	.02	.01	.01	.02

Source: own elaboration

The metrics exposed in Table 28 to Table 30, evaluate the importance of each feature of the decision classifier, used to analyse which of them have the most impact on the dependent variable (Guyon & Elisseeff, 2003; Sarwar et al., 2000). The method used to calculate it is the Gini impurity, which would be a number that indicates the likelihood of misclassifying a random data point based on the distribution of the classes of the dataset, this process must be repeated recursively for each split of the classification tree (Xia et al., 2008). The most important features would be those that reduce the Gini impurity the most.

Table 28 shows the feature importance of the 10 specifications that have been calculated with the classification tree. The importance of the cultural variables is always close to 70%, while the importance of the individual variables is around 30%, with the importance of gender and group in the classification being 5%, means of transport 7% and age 13%. Table 29 and Table 30 shows the feature importance of boosting, bagging, and Random Forest respectively, in which we see that the importance of culture increases to 90%, reducing the importance of the others. Thus, we can determine that the different specifications of cultural variables cause a similar aggregated feature importance (all variables). However, it can be observed that in the case of Hofstede national values, the most relevant variables of the construct for explaining the election of the alternatives are UAI, with the least importance placed on PDI. Conversely, in the case of Schwartz, the most relevant variables are Egalitarianism and Mastery, while the least important ones are Affective Autonomy and Intellectual Autonomy, being both of Inglehart's values similar in their exploratory power.

6.5.2 Optimal parameters

To carry out parameter optimization for the models belonging to the classification tree family, we chose to preselect a series of commonly used values from the literature for the parameters of the four models, this optimization being necessary for accuracy, prevention of overfitting and control of complexity. Thus, we employed a manual grid search (Bergstra & Bengio, 2012), where we specified a range of values for each parameter, and then trained and evaluated each model using all possible combinations of parameters and values to find the one that offers the best performance. The parameter space can be seen on Table 31.

Table 32 reflect the optimal parameters selected for each specification and model based on which was the parameters that minimized the prediction error, allowing us to find the best combination for each model. To do this, we performed a grid search by systematically exploring the parameter space, enabling us to fine-tune each model to maximize its predictive power. Analysing the optimal parameters, we can say that the prune model resulted in a full tree, as the optimal alpha value is near 0, as the learning rate of the boosting is 0.05, potentially having an optimal value near 0.

Table 31. Parameter space to optimize for each model

Model	Parameters	Values
Single-Pruning	Alpha	0.01, 0.035, 0.055, 0.1
Ensemble-Boosting	N Estimators	50, 100, 150, 200, 250
	Learning Rate	0.05, 0.1, 0.5, 1.0
Ensemble-Bagging	N Estimators	50, 100, 150
	Max Samples	0.5, 0.75, 1.0
	Max Features	0.5, 0.75, 1.0
Ensemble-Random Forest	N Estimators	50, 100, 150
	Max Depth	5, 10, 15
	Max Features	Square Root, Log2

Source: own elaboration

Table 32. Optimal parameters for each specification and model

Specification	Single-Pruning	Ensemble-Boosting	Ensemble-Bagging	Ensemble-Random Forest
Specification 1	Alpha: 0.01	Learning rate: 0.05 N Estimators: 250	Max features: 1.0 Max samples: 0.75 N Estimators: 50	Max depth: 15 Max features: log2 N Estimators: 100
Specification 2	Alpha: 0.01	Learning rate: 0.05 N Estimators: 250	Max features: 1.0 Max samples: 0.5 N Estimators: 50	Max depth: 15 Max features: sqrt N Estimators: 150
Specification 3	Alpha: 0.01	Learning rate: 0.05 N Estimators: 150	Max features: 0.5 Max samples: 1.0 N Estimators: 150	Max depth: 10 Max features: sqrt N Estimators: 50
Specification 4	Alpha: 0.01	Learning rate: 0.05 N Estimators: 100	Max features: 1.0 Max samples: 0.75 N Estimators: 50	Max depth: 10 Max features: sqrt N Estimators: 50
Specification 5	Alpha: 0.01	Learning rate: 0.05 N Estimators: 250	Max features: 0.75 Max samples: 0.75 N Estimators: 100	Max depth: 15 Max features: log2 N Estimators: 150
Specification 6	Alpha: 0.01	Learning rate: 0.05 N Estimators: 250	Max features: 1.0 Max samples: 0.5 N Estimators: 100	Max depth: 15 Max features: sqrt N Estimators: 50
Specification 7	Alpha: 0.01	Learning rate: 0.1 N Estimators: 100	Max features: 0.5 Max samples: 1.0 N Estimators: 100	Max depth: 5 Max features: log2 N Estimators: 150
Specification 8	Alpha: 0.01	Learning rate: 0.1 N Estimators: 50	Max features: 1.0 Max samples: 1.0 N Estimators: 50	Max depth: 15 Max features: sqrt N Estimators: 100
Specification 9	Alpha: 0.01	Learning rate: 0.05 N Estimators: 50	Max features: 1.0 Max samples: 1.0 N Estimators: 150	Max depth: 15 Max features: sqrt N Estimators: 50
Specification 10	Alpha: 0.01	Learning rate: 0.1 N Estimators: 150	Max features: 0.5 Max samples: 0.75 N Estimators: 150	Max depth: 10 Max features: log2 N Estimators: 150

Source: own elaboration

6.5.3 Performance

To evaluate the performance of our multiple models across each of the 10 specifications, we chose to analyse the accuracy of prediction and F1 score for each combination. Thus, we conducted our models using the optimal parameters and 10-fold cross-validation, due to being a standard number of folds to perform in the literature, ensuring that the original observations can appear in both the training and test sets. 10-fold cross-validation involved splitting the data into 10 subsets and using each subset in turn as a validation set while training the model on the remaining 9 subsets. Averaging the results, we obtained a robust estimation of the model performance on the data (Fushiki, 2011). Only optimal parameters were used to calculate single (tree and pruned tree) and ensemble (boosting, bagging and Random Forest) models. On Table 33 we can view the accuracy, determining the proportion of correct predictions made by the model divided by the total number of predictions, and on Table 34 we can see the F1 score, that is a metric that combines precision and recall, providing a single score that balances them, particularly important on imbalanced datasets (Goutte & Gaussier, 2005). Both metrics were performed with optimal parameters from Table 32 and 10-fold cross-validation for single (tree and pruned tree) and ensemble (boosting, bagging and Random Forest) models, and calculating the scores through weighted averaging for each class separately, and being the weights proportional to the number of instances in each outcome. As anticipated from the study of feature importance, it's evident that the ten specifications exhibit a similar explanatory power studied by their performance on accuracy and F1 score, except for the fourth specification, utilizing Inglehart's Cultural cluster, which shows less performance in classification tree models than in multinomial and choice models.

Table 33. Accuracy of prediction for each specification and model

Specification	Multinomial	Choice	Single-Tree	Single-Pruning	Ensemble-Boosting	Ensemble-Bagging	Ensemble-Random Forest
Specification 1	0.5312	0.5310	0.5530	0.5370	0.5570	0.5569	0.5570
Specification 2	0.5312	0.5255	0.5568	0.5468	0.5554	0.5558	0.5555
Specification 3	0.5316	0.5308	0.5582	0.5461	0.5551	0.5557	0.5554
Specification 4	0.5321	0.5364	0.5253	0.5228	0.5259	0.5259	0.5259
Specification 5	0.5306	0.5311	0.5541	0.5482	0.5581	0.5582	0.5581
Specification 6	0.5325	0.5307	0.5571	0.5470	0.5574	0.5572	0.5572
Specification 7	0.5349	0.5282	0.5562	0.5604	0.5574	0.5594	0.5602
Specification 8	0.53153	0.5301	0.5585	0.5348	0.5570	0.5568	0.5566
Specification 9	0.5304	0.5329	0.5544	0.5545	0.5577	0.5577	0.5577
Specification 10	0.5327	0.4993	0.5577	0.5371	0.5568	0.5593	0.5588

Source: own elaboration

Table 34. F1 score of prediction for each specification and model

Specification	Single-Tree	Single-Pruning	Ensemble-Boosting	Ensemble-Bagging	Ensemble-Random Forest
Specification 1	0.4601	0.4601	0.4527	0.4602	0.4605
Specification 2	0.4581	0.4581	0.4500	0.4587	0.4593
Specification 3	0.4588	0.4588	0.4529	0.4594	0.4593
Specification 4	0.4353	0.4353	0.4353	0.4389	0.4389
Specification 5	0.4569	0.4569	0.4497	0.4579	0.4576
Specification 6	0.4587	0.4587	0.4488	0.4586	0.4590
Specification 7	0.4589	0.4589	0.4561	0.4594	0.4590
Specification 8	0.4592	0.4592	0.4421	0.4599	0.4599
Specification 9	0.4577	0.4577	0.4402	0.4581	0.4581
Specification 10	0.4593	0.4593	0.4542	0.4601	0.4602

Source: own elaboration

6.6 DISCUSSION

With the goal of offering the optimal algorithm for explaining which classification model performs the best, considering all limitations regarding flexibility, computational complexity, and interpretability, we conducted a comparative analysis of several classification models on ten specifications, examining their performance. As a baseline for comparison, we used tree-based models as well as ensembles. Comparing the feature importance of classification tree family models, it was observed, as hypothesized, that cultural variables are the most explanatory for individuals' chosen alternatives, explaining around 70-90% of all variability. After conducting a thorough comparison of the accuracy and F1-scores of the models and specifications we observed minimal discrepancies similar to other classification problems (Almomani et al., 2023; Delic et al., 2018; Y. Y. Liu et al., 2011; Ngo et al., 2015), caused by the slightly better performance of the classification tree and ensembles of classification trees (boosting, bagging and random forest). Thus, the ten specifications displayed a similar performance, suggesting that all the cultural constructs used were effective in predicting the alternatives chosen by individuals. However, Specification 4, which examined Inglehart’s cultural cluster, exhibited poorer performance when integrated into the classification tree models. This indicates that while most constructs were reliable in predicting choices, the specific framework related to Inglehart's cultural clusters might not be as suitable within the classification tree models compared to other specifications.

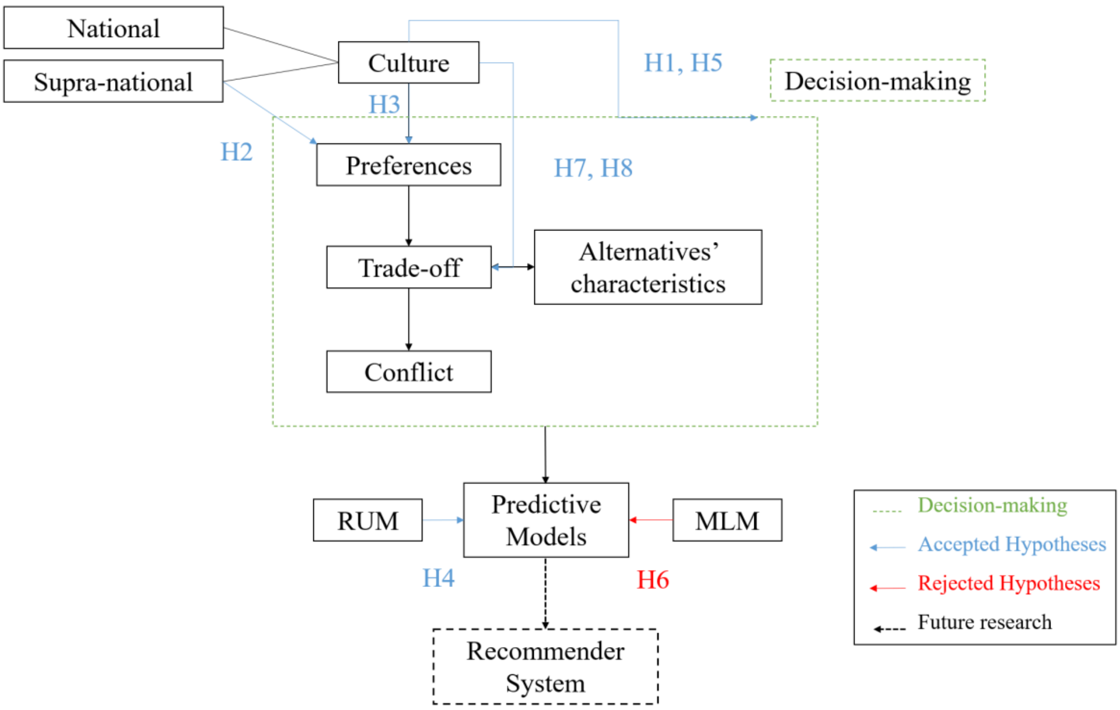
Considering that models are relative similar in their ability to classify outcomes of the dependant variable, choosing between them may come down to other factors such as the flexibility, computational complexity, and interpretability (Herlocker et al., 2004; Sarwar et al., 2000). Based on these model selection criteria, the two most favoured models in this case study would be the simple classification tree and the multinomial choice model. On the one hand, the simple classification tree stands out for its straightforward implementation, interpretability, and flexibility, appealing for practical usage on classification problems. On the other hand, the multinomial choice model, though highly computationally costly, offers a substantial advantage by supplying comprehensive insights into evaluations linked to each alternative, thus making it useful for policy making.

CHAPTER 7. DISCUSSION, CONCLUSIONS AND CONTRIBUTIONS

7.1 DISCUSSION

This thesis has been guided by the questions and hypotheses presented in Chapter 1. As can be seen in Figure 28, the research has provided empirical evidence supporting the first seven hypotheses. However, the sixth hypotheses needed to be refuted.

Figure 28. New causal model on decision-making and evaluation of research hypotheses



Source: own elaboration

Relating hypotheses with results from different chapters:

- Supra-national culture conditions the decision-making process (H1) has been confirmed, as each supra-national culture exhibits distinct preferences over the alternatives, as can be seen in chapters 4 and 5.
- Multinomial logit model with alternative-specific variables can explain tourist's preferences (H2) has been confirmed. The results of this model allowed us to understand the preferences of individuals over the characteristics of the alternatives, and thus over the alternatives themselves, as can be seen in chapters 4 and 5.
- Culture as a push factor due to the influence in preferences of individuals (H3) has been confirmed. More specifically, we confirmed that subjects from different supra-national culture have different preferences about the characteristics of the alternatives. This results supported by the significance of each alternative-specific coefficient, seen in chapters 4 and 5.
- Discrete choice models are valid for prediction tasks (H4) has been confirmed for our dataset. The prediction metrics (accuracy and F1 score), showed comparable performance of discrete choice models with machine learning models, as can be seen in chapter 6.
- Culture is the most important variable at explaining individuals' decision-making (H5) has been confirmed. Culture has been confirmed as the main explanatory variable for machine learning models, explaining 80% of the variability in the model, seen in chapter 6.
- The best models for predicting tourists' decision-making are those from machine learning, particularly ensemble methods (H6). This hypothesis could not be confirmed, due to the similar performance of machine learning models compared with discrete decision-making models, seen in chapter 6.
- Culture shapes how individuals resolve trade-offs dilemmas (H7) has been confirmed. The use of the new causal decision-making model with culture explanatory variables, proposed in this thesis has demonstrated that culture affects individuals' preferences, influencing how they resolve the trade-off dilemma between the conflictive characteristics of the alternatives, seen in chapter 5.
- Individuals from different cultural background differ in how they resolve the trade-off dilemma (H8) has been confirmed. The significance of the trade-off evaluations by

tourists from different cultural groups and constructs regarding the characteristics of the alternatives has been proved to differ, as can be seen in chapter 5.

7.2 CONTRIBUTIONS

Having addressed the main research goal of analysing the role of culture in tourists' decision-making (DM), we relate gaps noticed in the literature review with the four specific goals stablished on the introduction, presenting the contributions to science developed during the research.

Throughout the research we studied the role of culture in DM, specifically we elaborate on the concept of supra-national culture, scarcely explored in the field of tourism, that we prove helpful to explain tourist’s preferences on DM problems. This closes the first gap noticed in tourism DM literature (current cultural measures employed in tourism literature do not consider supra-national culture). We determine the utility of incorporating the supra-national culture in the push-pull model frameworks, benefiting from the inclusion of such variables as motives for traveling, closing the second gap noticed in the literature review (the limited study on the role of culture in the push-pull model). Thus, justifying the incorporation of supra-national culture into researcher’s toolbox, enabling the use of this new concept on DM problems. The inclusion of supra-national culture into DM models was achieved using multinomial logit with alternative-specific variables, demonstrating that the inclusion of cultural variables is significant when facing alternative selection problems, addressing specific goal 1.

The use of multinomial logit models with alternative-specific variables allowed us to precisely understand individuals' preferences for alternatives, as external researchers to the DM process. On our new causal model, culture is shaping individuals’ preferences, and how they resolve the trade-off dilemma between conflicting attributes of alternatives. This globally impacts the DM process, justifying the results, that we have addressed specific goal 2 and that we have closed the third gap noticed in the literature review.

The creation of the new causal DM model allowed us to resolve some of the gaps noticed in the literature of culture and DM, consisting of the limited analysis of countries or cultures and the utilization of few cultural variables in models when addressing the trade-off dilemma. We were able to address this gap using the new causal model. Which combined with multinomial choice models allowed us to analyse a wide variety of cultures, countries, and cultural constructs from the founders of cross-cultural analysis. This simultaneously explained individuals' preferences regarding the characteristics of alternatives, significant in DM models. This provided us with insights into how each group resolves the trade-off dilemma, addressing specific goal 3.

After evaluating discrete DM models and machine learning models, we observed that classification tree models and multinomial logit models with alternative-specific variables performed best in terms of their balance between interpretability, flexibility, and computational complexity. They all exhibited similar predictive capability, completing specific goal 4.

Our interdisciplinary work, combining paradigms from different fields, is what has allowed us to contribute to the field of tourism DM. We believe that the results justify the use of these new conceptual frameworks into future works in the field, holding the findings importance for authors, business, and policymakers.

7.3 IMPLICATIONS

The study's implications extend beyond academic realms into practical applications with significant potential impact. On the social realm, the recognition of supra-national cultural influence on tourist DM introduces understanding over how individuals within these cultural groups resolve DM problems. This understanding facilitates the aggregation of individual values into different supra-national levels, enabling the creation of robust DM models. Such models have the potential to explain intricate cultural dynamics and their role in shaping consumer behavior.

In the realm of marketing, the implication of using supra-national culture for market segmentation presents a strategic shift. Instead of relying on individual country models, businesses are encouraged to adopt a supra national culture approach, grouping nations into culturally cohesive groups bigger than the nation. This approach optimizes resource allocation, allowing companies to target a broader array of markets while being cost efficient. Thus, we expect that business that use specific cultural characteristics shared among supra-national groups can become more culturally aware, resonating their marketing strategies effectively with diverse consumer bases.

Our study's findings offer valuable insights into public policy formulation. Understanding consumer preferences through the lens of supra-national grouping allows policymakers to tailor specific interventions for different cultures, instead of the classical country-oriented interventions. This approach enhances destination management by acknowledging and adapting to the unique social, environmental, and economic impacts of policies. The study suggests the potential implementation of targeted fees or discounts directed at specific cultural groups, generating new policy frameworks.

Ultimately, the study aspires to bridge the gap between academic research and real-world applications, transferring our acquired knowledge to businesses, policymakers, and researchers alike. We recommend a culturally informed approach, recognizing that culture play a main role in shaping economic decisions and consumer behaviour.

Finally, in terms of the thesis's societal impact, we hope that it contributes to the improvement of company recommendation systems, aiding tourism authorities, and assisting researchers in developing more sophisticated DM models by integrating the cultural dimension, which is usually overlooked in the field of economics.

7.4 LIMITATIONS

As a first limitation of the research, we found that the Pilgrims Welcome Office dataset lacks several key variables that could enrich the analysis of the relationship between culture and pilgrims' DM. This is due to the design of the dataset as a registry of pilgrims, with secondary intention of conducting analyses. Thus, variables related to pilgrims' motivations, individual cultural variables and psychological factors (availability and polarity of information about the Ways and possible anchoring effects due to price of touristic packages offered by travel agencies), useful at understanding DM, as demonstrated by Kahneman & Tversky (1984), are missing.

As a second limitation, we have the lack of knowledge about the individual's emotional state at the time of decision-making.

As a third limitation we have the lack of comparisons between supra-national cultures, however, we believe it's worth utilizing the cultural cluster generated by Inglehart in the 1990s as our objective demonstrating the effect of construct that has been validated and employed by other authors in the field of cross-cultural analysis.

As a final limitation, we have not been able to fully evaluate the predictive models with new pilgrims, comparing them with their actual choices. Additionally, we lack the comparison of the predictive capabilities of all machine-learning algorithms.

7.5 FUTURE RESEARCH

Future research should improve Pilgrims Welcome Office questionnaire by integrating crucial variables absent in its initial design, such as pilgrims' motivations, psychological and cultural determinants, key explanatory variables of DM. This expansion of the questionnaire of Pilgrims Welcome Office aim to integrate aforementioned variables, so that we can understand the intricate mechanism of DM processes, offering more comprehensive insights into this domain. Despite potential improvements of the Pilgrims Welcome Office database, the number of questions that can be included will be limited, consequently, we plan to conduct in-person questionnaires along the Portuguese ways of Saint James, collaborating with Portuguese researchers. This will enable us to compare declared preferences of the pilgrims with their observable behaviour, focus of this research. Specifically, we will incorporate a greater number of control sociodemographic variables, along with psychological and cultural variables, motivations, and satisfaction with the experiences, studying the underlying factors influencing the DM process.

We are currently developing an article investigating how cultural aggregations impact on pilgrim's preferences along the Portuguese ways of Saint James Way. This study aims to unveil relationships between cultural aggregation levels and the selection of alternatives, where we seek to understand how diverse cultural aspects shape pilgrims' DM processes along Saint James Way.

Our future research will extend its scope to include a broader range of models, not studied in this thesis, deepening our comprehension of the DM process of individuals by examining several techniques, thereby ensuring a comprehensive analysis of outcomes. Incorporating additional models, our objective is to refine classification methodologies and enhance performance of our models. We aim to ensure that the recommendation tool that reaches production state, is the best balancing interpretability, flexibility, and computational complexity, thus enlarging the benefits for business and users.

Finally, this thesis may contribute to society, with the transfer of the research, collaborating with different agents such as Xunta de Galicia, Spanish and Portuguese government and technological companies related to tourism sector. This collaboration aims to achieve better management of Saint James Way hotspots. Thus, we are going to develop an analytic tool, deploying an integrated predictive system that generates alerts for each potential overcrowding issue, aiding each administration to reach a proactive management of the potential problems.

CHAPTER 8. BIBLIOGRAPHY

AEMET. (2021). *Datos y Estadísticas*. http://www.aemet.es/es/lineas_de_interes/datos_y_estadistica

AEMET. (2023). *Valores climatológicos normales*. AEMET. https://www.aemet.es/es/serviciosclimaticos/datosclimatologicos/valoresclimatologicos

Akaike, H. (2011). Akaike's information criterion. In *International encyclopedia of statistical science* (pp. 25–25). Springer Berlin Heidelberg. https://doi.org/https://doi.org/10.1007/978-3-642-04898-2_110

Akaliyski, P. (2017). Sources of societal value similarities across Europe. *Comparative Sociology*, *16*(4), 447–470. https://doi.org/10.1163/15691330-12341432

Akaliyski, P., Welzel, C., Bond, M. H., & Minkov, M. (2021). On "Nationology": The Gravitational Field of National Culture. *Journal of Cross-Cultural Psychology*, *52*(8–9), 771–793. https://doi.org/10.1177/00220221211044780

Almomani, A. (2020). *Ensembles of choice-based models for recommender systems*. Universidad de Santiago de Compostela.

Almomani, A., Monreal, C., Sieira, J., Graña, J., & Sánchez, E. (2020). Rational, emotional, and attentional models for recommender systems. *Expert Systems*, *May*, 1–13. https://doi.org/10.1111/exsy.12594

Almomani, A., Saavedra, P., Barreiro, P., Durán, R., Crujeiras, R., Loureiro, M., & Sánchez, E. (2022). Application of choice models in tourism recommender systems. *Expert Systems*, *April*, 1–18. https://doi.org/10.1111/exsy.13177

Almomani, A., Saavedra, P., Barreiro, P., Durán, R., Crujeiras, R., Loureiro, M., & Sánchez, E. (2023). Application of choice models in tourism recommender systems. *Expert Systems*, *40*(3), 1–18. https://doi.org/10.1111/exsy.13177

Alpaydin, E. (2010). *Introduction to machine learning* (2nd ed.). MIT press.

Amelung, B., Nicholls, S., & Viner, D. (2007). Implications of global climate change for tourism flows and seasonality. *Journal of Travel Research*, *45*(3), 285–296. https://doi.org/10.1177/0047287506295937

Arrow, K. J. (1959). Rational Choice Functions and Orderings. *Economica*, *26*(102), 121–127.

Baldwin, J. R., Faulkner, S., Lindsley, S. L., & Hecht, M. L. (2006). *Redefining culture: perspectives across the disciplines* (J. R. Baldwin, S. Faulkner, S. L. Lindsley, & M. L. Hecht (eds.)). Lawrence Erlbaum Associates.

Banerjee, A. V. (1992). A simple model of herd behavior. *The Quaterly Journal of Economics*, *107*(3), 797–817.

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis*. MIT press.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, *13*, 281–305.

Berkson, J. (1944). Application of the Logistic Function to Bio-Assay. *Journal of the American Statistical Association*, *39*(227), 357–365.

Berry, J. (1997). *An ecocultural approach to the study of cross-cultural industrial/organizational psychology*. The New Lexington Press/Jossey-Bass Publishers.

Beugelsdijk, S., Ambos, B., & Nell, P. C. (2018). Conceptualizing and measuring distance in international business research: Recurring questions and best practice guidelines. *Journal of International Business Studies*, *49*(9), 1113–1137. https://doi.org/10.1057/s41267-018-0182-4

Beugelsdijk, S., & Welzel, C. (2018). Dimensions and Dynamics of National Culture: Synthesizing Hofstede With Inglehart. *Journal of Cross-Cultural Psychology*, *49*(10), 1469–1505. https://doi.org/10.1177/0022022118798505

Bliss, C. I. (1934). The method of probits. *Science*, *79*(2037), 38–39.

Bogardus, E. S. (1925). Measuring social distance. *Journal of Applied Sociology*, *9*, 299–308.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*, 123–140. https://doi.org/10.3390/risks8030083

Cha, S., McCleary, K. W., & Uysal, M. (1995). Travel motivations of Japanese overseas travelers: a factor-cluster segmentation approach. *Journal of Travel Research*, *34*(1), 33–39. https://doi.org/10.1177/004728759503400104

Charles, C., Gafni, A., & Whelan, T. (2013). Decision making in the physician-patient encounter: Revisiting the shared treatment decisionmaking model. *The Sociology of Health and Illness: A Reader*, *49*, 229–241. https://doi.org/10.4324/9781315013015-30

Cheng, G., Zhao, S., & Huang, D. (2018). Understanding the effects of improving transportation on pilgrim travel behavior: Evidence from the Lhasa, Tibet, China. *Sustainability (Switzerland)*, *10*(10), 1–13. https://doi.org/10.3390/su10103528

Chu, P. C., & Spires, E. E. (2008). The cost-benefit framework and perceptions of decision strategies: A comparison of China and the United States. *Journal of Cross-Cultural Psychology*, *39*(3), 303–308. https://doi.org/10.1177/0022022108314549

Chu, P. C., Spires, E. E., Farn, C. K., & Sueyoshi, T. (2005). Decision processes and use of decision AIDS: Comparing two closely related nations in East Asia. *Journal of Cross-Cultural Psychology*, *36*(3), 304–320. https://doi.org/10.1177/0022022104273653

Chu, P. C., Spires, E. E., & Sueyoshi, T. (1999). Cross-cultural differences in choice behavior and use of decision aids: A comparison of Japan and the United States. *Organizational Behavior and Human Decision Processes*, *77*(2), 147–170. https://doi.org/10.1006/obhd.1998.2817

Comín, F. (2014). *Historia Económica Mundial*. Alianza Editorial.

Comtois, D. (2016). *Package 'summarytools.'* https://cran.r-project.org/package=summarytools

Correia, A., Kozak, M., & Ferradeira, J. (2011). Impact of culture on tourist decision-making styles. *International Journal of Tourism Research*, *13*(5), 433–446. https://doi.org/10.1002/jtr.817

Correia, A., Kozak, M., & Ferradeira, J. (2013). From tourist motivations to tourist satisfaction. *International Journal of Culture, Tourism, and Hospitality Research*, *7*(4), 411–424. https://doi.org/10.1108/IJCTHR-05-2012-0022

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, *20*, 273–297.

Cortez, P., & Embrechts, M. J. (2013). Using sensitivity analysis and visualization techniques to open black box data mining models. *Information Sciences*, *225*, 1–17. https://doi.org/10.1016/j.ins.2012.10.039

Crawley, M. (2012). *The R Book* (2nd ed.). John Wiley & Sons Inc.

Croissant, Y. (2020). Estimation of Random Utility Models in {R}: The {mlogit} Package. *Journal of Statistical Software*, *95*(11), 1–41. https://doi.org/10.18637/jss.v095.i11

Crompton, J. L. (1979). Motivations for pleasure vacation. *Annals of Tourism Research*, *Oct/Dec*(December), 408–424.

Crotts, J. C. (2004). The Effect of Cultural Distance on Overseas Travel Behaviors. *Journal of Travel Research*, *43*(August), 83–88. https://doi.org/10.1177/0047287504265516

Crotts, J. C., & Erdmann, R. (2000). Does national culture influence consumers' evaluation of travel services? A test of Hofstede's model of cross-cultural differences. *Managing Service Quality: An International Journal*, *10*(6), 410–419. https://doi.org/10.1108/09604520010351167

Crouch, G. I., & Louviere, J. J. (2001). A review of choice modelling research in tourism, hospitality and leisure. In J. A. Mazanec, G. I. Crouch, J. R. Brent Ritchie, & A. G. Woodside (Eds.), *Consumer psychology of tourism, hospitality and leisure. Volume 2* (pp. 67–86). CABI Publishing. https://doi.org/10.1079/9780851995359.0019

Dann, G. (1976). «The holiday was simply fantastic». *The Tourist Review*, *31*(3), 19–23. https://doi.org/10.1108/eb057732

Delic, A., Neidhardt, J., Nguyen, T. N., & Ricci, F. (2018). An observational user study for group recommender systems in the tourism domain. *Information Technology and Tourism*, *19*(1–4), 87–116. https://doi.org/10.1007/s40558-018-0106-y

Development., N. C. for O. (2022). *ONET online*. https://www.onetonline.org/

Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, *15*, 3133–3181.

Fernández Méndez, D., Fernández, Fernández, M., & Riveiro García, D. (2019). *Impacto socioeconómico del Camino de Santiago*. https://www.eixoatlantico.com/images/SIE/181129-sie-turismo-barcelos/informe_estudio_camino_santiago-esp_laura-dopeso.pdf

Fushiki, T. (2011). Estimation of prediction error by using K-fold cross-validation. *Statistics* *and Computing*, *21*(2), 137–146. https://doi.org/10.1007/s11222-009-9153-8

Gareth, J., Witten, D., Hastie, T., & Robert, T. (2021). *An Introduction to Statistical Learning: With Applications in R* (Second). Springer Science+ Business Media.

Gaston-Breton, C., & Martín, O. M. (2011). International market selection and segmentation: A two-stage model. *International Marketing Review*, *28*(3), 267–290. https://doi.org/10.1108/02651331111132857

Gibbons, J. D., & Chakraborti, S. (2020). *Nonparametric statistical inference* (6th ed.). CRC Press.

Gil A., L., Dwivedi, A., & W. Johnson, L. (2017). Effect of popularity and peer pressure on attitudes toward luxury among teens. *Young Consumers*, *18*(1), 84–93. https://doi.org/http://dx.doi.org/10.1108/YC-10-2016-00639

Goutte, C., & Gaussier, E. (2005). A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. *Lecture Notes in Computer Science*, *3408*, 345–359. https://doi.org/10.1007/978-3-540-31865-1_25

Green, F. (1997). Union recognition and paid holiday entitlement. *British Journal of Industrial Relations*, *35*(2), 243–255. https://doi.org/10.1111/1467-8543.00050

Gronze. (2021). *El Camino de Santiago*. https://www.gronze.com/

Gronze. (2023). *El Camino de Santiago*. Gronze. https://www.gronze.com/

Guiasu, S., & Shenitzer, A. (1985). The principle of maximum entropy. *The Mathematical Intelligencer*, *7*(1), 42–48. https://doi.org/10.1007/BF03023004

Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, *3*, 1157–1182. https://doi.org/10.1016/j.aca.2011.07.027

Hagag, W., Clark, L., & Wheeler, C. (2015). A framework for understanding the website preferences of Egyptian online travel consumers. *International Journal of Culture, Tourism, and Hospitality Research*, *9*(1), 68–82. https://doi.org/10.1108/IJCTHR-03-2013-0008

Haire, M., Ghiselli, E., & Porter, L. W. (1966). *Managerial thinking: an international study*. John Wiley & Sons Inc.

Hall, C. M., & Ram, Y. (2021). Weather and climate in the assessment of tourism-related walkability. *International Journal of Biometeorology*, *65*(5), 729–739. https://doi.org/10.1007/s00484-019-01801-2

Hammond, J. S., Keeney, R. L., & Raiffa, H. (1998). Even swaps: a rational method for making trade-offs. *Harvard Business Review*, *76*(2).

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with {NumPy}. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Herlocker, J., Konstan, J., Terveen, L., & Riedl, J. (2004). Evaluating Collaborative Filtering

Recommender Systems. *ACM Transactions on Information Systems*, *22*(1), 5–53. https://doi.org/10.1561/1100000009

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*, 1735–1780.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values.* SAGE Publications.

Hofstede, G. (1989). Organising for cultural diversity. *European Management Journal*, *7*(4), 390–397.

Hofstede, G., Hofstede, G. J., & Minkov, M. (2005). *Cultures and organizations: Software of the mind*. Mcgraw-hill.

Hsu, S., Woodside, A. G., & Marshall, R. (2013). Critical Tests of Multiple Theories of Cultures' Consequences: Comparing the Usefulness of Models by Hofstede, Inglehart and Baker, Schwartz, Steenkamp, as well as GDP and Distance for Explaining Overseas Tourism Behavior. *Journal of Travel Research*, *52*(6), 679–704. https://doi.org/10.1177/0047287512475218

Huang, S. (Sam), & Crotts, J. (2019). Relationships between Hofstede's cultural dimensions and tourist satisfaction: A cross-country cross-sample examination. *Tourism Management*, *72*(November 2018), 232–241. https://doi.org/10.1016/j.tourman.2018.12.001

Huntington, S. P. (1993). The Clash of Civilizations? *Foreign Affairs*, *72*(3), 22. https://doi.org/10.2307/20045621

IBM. (2012). *IBM SPSS modeler CRISP-DM guide* (pp. 1–53). IBM Corporation. https://www.ibm.com/docs/en/spss-modeler/18.1.1?topic=spss-modeler-crisp-dm-guide

IGN. (2023). *Caminos de Santiago en Europa*. IGN. https://nco.ign.es/CaminoSantiago/

Impactur. (2017). *Impactur. Impacto del turismo activo en Galicia*. https://www.exceltur.org/wp-content/uploads/2019/01/IMPACTUR-Galicia-2017.pdf

INE. (2022). *Encuesta de estructura salarial*. https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177025&menu=ultiDatos&idp=1254735976596

INE. (2023). *Cuanta satélite del turismo de España*. INE. https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736169169&menu=ultiDatos&idp=1254735576863

Inglehart, R. (1997). *Modernization and postmodernization: Cultural, economic, and political change in 43 societies*. Princeton university press.

Inglehart, R. (2006). Mapping Global Values. *Comparative Sociology*, *5*(2), 115–136. http://www.worldvaluessurvey.org/

Inglehart, R., & Baker, W. E. (2000). Modernization , Cultural Change , and the Persistence of Traditional Values. *American Sociological Review*, *65*(1), 19–51.

Inglehart, R., C. Haerpfer, A., A. Moreno, C., & C. Welzel, K. (2014). *World Values Survey:*

Round One - Country-Pooled Datafile Version. JD Systems Institude.

Iseki, H., Liu, C., & Knaap, G. (2018). The determinants of travel demand between rail stations: A direct transit demand model using multilevel analysis for the Washington D.C. Metrorail system. *Transportation Research Part A: Policy and Practice*, *116*(March), 635–649. https://doi.org/10.1016/j.tra.2018.06.011

Ismoilov, J. (2017). *Stated and revealed preferences on gastronomic tourism in Santiago de Compostela*. https://dialnet.unirioja.es/servlet/tesis?codigo=146695&info=resumen&idioma=SPA

Jarvenpaa, S., Tractinsky, N., & Saarinen, L. (1999). Consumer trust in an internet store: A cross-cultural validation. *Journal of Computer-Mediated Communication*, *5*(2). https://doi.org/https://doi.org/10.1111/j.1083-6101.1999.tb00337.x

Jones, M. C., Marron, J. S., & Sheather, S. J. (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, *91*(433), 401–407. https://doi.org/10.1080/01621459.1996.10476701

Jung, J. M., Sydnor, S., Lee, S. K., & Almanza, B. (2015). A conflict of choice: How consumers choose where to go for dinner. *International Journal of Hospitality Management*, *45*, 88–98. https://doi.org/10.1016/j.ijhm.2014.11.007

Kaasa, A., & Minkov, M. (2020). Are the World's National Cultures Becoming More Similar? *Journal of Cross-Cultural Psychology*, *51*(7–8), 531–550. https://doi.org/10.1177/0022022120933677

Kahneman, D., & Tversky, A. (1984). Choices, Values, and Frames. *American Psychologist*, *April*, 341–350.

Kandogan, Y. (2012). An improvement to Kogut and Singh measure of cultural distance considering the relationship among different dimensions of culture. *Research in International Business and Finance*, *26*(2), 196–203. https://doi.org/10.1016/j.ribaf.2011.11.001

Kaplan, J. (2020). fastDummies: Fast creation of dummy (binary) columns and rows from categorical variables. *R Package Version*, *1*(3).

Kassambara, A. (2020). rstatix: Pipe-friendly framework for basic statistical tests. *R Package Version 0.6. 0*. https://cran.r-project.org/web/packages/rstatix/index.html

Kim, D., & Park, B. J. (Robert). (2017). The moderating role of context in the effects of choice attributes on hotel choice: A discrete choice experiment. *Tourism Management*, *63*, 439–451. https://doi.org/10.1016/j.tourman.2017.07.014

Kogut, B., & Singh, H. (1988). The effect of national culture on the choice of entry mode. *Journal of International Business Studies*, 411–433. https://doi.org/10.1177/030639688102200403

Kozak, M. (2002). Comparative analysis of tourist motivations by nationality and destinations. *Tourism Management*, *23*(3), 221–232. https://doi.org/10.1016/S0261-5177(01)00090-5

Kozak, M., Crotts, J. C., & Law, R. (2007). The Impact of the Perception of Risk on

International Travellers. *International Journal of Tourism Research*, *9*, 233–242. https://doi.org/10.1002/jtr.607

Kreps, D. M. (2020). *A course in microeconomic theory*. Princeton university press.

Kroeber, A. L., & Kluckhohn, C. (1952). Culture: A critical review of concepts and definitions. *Papers. Peabody Museum of Archaeology \& Ethnology. Cambridge*, *47*(1).

Kumar, S., & Dhir, A. (2020). Associations between travel and tourism competitiveness and culture. *Journal of Destination Marketing and Management*, *18*(October), 1–23. https://doi.org/https://doi.org/10.1016/j.jdmm.2020.100501

Lacher, R. G., Oh, C. O., Jodice, L. W., & Norman, W. C. (2013). The Role of Heritage and Cultural Elements in Coastal Tourism Destination Preferences: A Choice Modeling-Based Analysis. *Journal of Travel Research*, *52*(4), 534–546. https://doi.org/10.1177/0047287512475215

Lefebvre, R., & Franke, V. (2013). Culture matters: Individualism vs. collectivism in conflict decision-making. *Societies*, *3*(1), 128–146. https://doi.org/10.3390/soc3010128

Levinson, J. D., & Peng, K. (2007). Valuing cultural differences in behavioral economics. *ICFAI Journal of Behavioral Finance*, *4*, 32–47.

Li, C. L., Chick, G. E., Wu, H. C., & Yen, T. M. (2010). Examining the dimensionality of values for culturally diverse customers in parks and recreation. *World Leisure Journal*, *52*(2), 80–93. https://doi.org/10.1080/04419057.2010.9674633

Li, S., Chark, R., Bastiaansen, M., & Wood, E. (2023). A review of research into neuroscience in tourism: Launching the annals of tourism research curated collection on neuroscience in tourism. *Annals of Tourism Research*, *101*. https://doi.org/10.1016/j.annals.2023.103615

Litvin, S. W., Crotts, J. C., & Hefner, F. L. (2004). Cross-cultural Tourist Behaviour : a Replication and Extension Involving Hofstede's Uncertainty Avoidance Dimension. *International Journal of Tourism Research*, *6*, 29–37. https://doi.org/10.1002/jtr.468

Liu, H., Robert, X., Cárdenas, D. A., & Yang, Y. (2018). Perceived cultural distance and international destination choice : The role of destination familiarity , geographic distance , and cultural motivation. *Journal of Destination Marketing & Management*, *9*, 300–309. https://doi.org/10.1016/j.jdmm.2018.03.002

Liu, Y. Y., Yang, M., Ramsay, M., Li, X. S., & Coid, J. W. (2011). A Comparison of Logistic Regression, Classification and Regression Tree, and Neural Networks Models in Predicting Violent Re-Offending. *Journal of Quantitative Criminology*, *27*(4), 547–573. https://doi.org/10.1007/s10940-011-9137-7

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, *22*(1), 3–42.

Mahalanobis, P. C. (1936). *On the generalized distance in statistics*. National Institute of Sciences of India.

Manosuthi, N., Lee, J., & Han, H. (2020). Impact of distance on the arrivals , behaviours and attitudes of international tourists in Hong Kong : A longitudinal approach. *Tourism*

Management, *78*, 103963. https://doi.org/10.1016/j.tourman.2019.103963

Masiero, L., & Qiu, R. T. R. (2018). Modeling reference experience in destination choice. *Annals of Tourism Research*, *72*(June), 58–74. https://doi.org/10.1016/j.annals.2018.06.004

Mcfadden, D. (1980). Econometric Models for Probabilistic Choice Among Products. *The Journal of Business*, *53*(3), S13–S29.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In *Analysis of qualitative choice behavior* (pp. 105–142). Institute of Urban and Regional Development, University of California Oackland. https://doi.org/10.1080/07373937.2014.997882

Mckercher, B., & Cros, H. (2003). Testing a Cultural Tourism Typology. *International Journal of Tourism Research*, *5*, 45–58. https://doi.org/10.1002/jtr.417

McKinney, W. (2011). Pandas: a foundational Python library for data analysis and statistics. *Python for High Performance and Scientific Computing*, *14*(9), 1–9.

Minkov, M., & Hofstede, G. (2012). Is national culture a meaningful concept? Cultural values delineate homogeneous national clusters of in-country regions. *Cross-Cultural Research*, *46*(2), 133–159.

Minkov, M., & Hofstede, G. (2013). *Cross-cultural analysis*. SAGE Publications Inc.

Murray, M., & Graham, B. (1997). Exploring the dialectics of route based tourism: the Camino de Santiago. *Tourism Management*, *18*(8), 513–524.

Ng, S. I., Lee, J. A., & Soutar, G. N. (2007). Tourists ' intention to visit a country : The impact of cultural distance. *Tourism Management*, *28*, 1497–1506. https://doi.org/10.1016/j.tourman.2006.11.005

Ngo, F. T., Govindu, R., & Agarwal, A. (2015). Assessing the Predictive Utility of Logistic Regression, Classification and Regression Tree, Chi-Squared Automatic Interaction Detection, and Neural Network Models in Predicting Inmate Misconduct. *American Journal of Criminal Justice*, *40*(1), 47–74. https://doi.org/10.1007/s12103-014-9246-6

Oppewal, H., Huybers, T., & Crouch, G. I. (2015). Tourist destination and experience choice: A choice experimental analysis of decision sequence effects. *Tourism Management*, *48*, 467–476. https://doi.org/10.1016/j.tourman.2014.12.016

Page, B. I., & Shapiro, R. Y. (2010). *The rational public: Fifty years of trends in Americans' policy preferences*. University of Chicago Press.

Parmar, A., Katariya, R., & Patel, V. (2019). A Review on Random Forest: An Ensemble Classifier. *Lecture Notes on Data Engineering and Communications Technologies*, *26*, 758–763. https://doi.org/10.1007/978-3-030-03146-6_86

Pedregosa, F., Varoquaux, G., Gramfort, A., & Michel, V. (2011). Scikit-learn: Machine Learning in {P}ython. *Journal of Machine Learning Research*, *12*(2825–2830).

Pegregrino, O. del. (2023). *Información estadística Oficina PEregrino para 2022 hasta diciembre*. Estadísticas. https://oficinadelperegrino.com/estadisticas-2/

Pestana, M. H., Parreira, A., & Moutinho, L. (2020). Motivations, emotions and satisfaction: The keys to a tourism destination choice. *Journal of Destination Marketing and Management*, *16*(December 2018), 100332. https://doi.org/10.1016/j.jdmm.2018.12.006

Pizam, A., & Sussmann, S. (1995). Does nationality affect tourist behavior? *Annals of Tourism Research*, *22*(4), 901–917. https://doi.org/10.1016/0160-7383(95)00023-5

Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, *6*(3), 21–44. https://doi.org/10.1109/MCAS.2006.1688199

Popov, E., Veretennikova, A., Naumov, I., & Kozinskaya, K. (2018). Non-Formal Institutional Environment of Social Entrepreneurship. *Economic and Social Changes: Facts, Trends, Forecast*, *11*(4 (58)), 217–234. https://doi.org/10.15838/esc.2018.4.58.14

Portela, M., Neira, I., & Salinas-Jiménez, M. del M. (2013). Social Capital and Subjective Wellbeing in Europe: A New Approach on Social Capital. *Social Indicators Research*, *114*(2), 493–511. https://doi.org/10.1007/s11205-012-0158-x

Powell, D., Yu, J., DeWolf, M., & Holyoak, K. J. (2017). The Love of Large Numbers: A Popularity Bias in Consumer Choice. *Psychological Science*, *28*(10), 1432–1442. https://doi.org/10.1177/0956797617711291

Qian, J., Law, R., Wei, J., Qian, J., Law, R., & Wei, J. (2018). Effect of cultural distance on tourism : A study of pleasure visitors in Hong Kong Effect of cultural distance on tourism : A study of pleasure. *Journal of Quality Assurance in Hospitality & Tourism*, *19*(2), 1–16. https://doi.org/10.1080/1528008X.2017.1410079

Rao, T. N. K., & Scott, A. J. (1984). On Chi-Squared Tests for Multiway Contingency Tables with Cell Proportions Estimated from Survey. *Institute of Mathematical Statistics*, *12*(1), 46–60.

Retief, F., Morrison-Saunders, A., Geneletti, D., & Pope, J. (2013). Exploring the psychology of trade-off decision-making in environmental impact assessment. *Impact Assessment and Project Appraisal*, *31*(1), 13–23. https://doi.org/10.1080/14615517.2013.768007

Revelle, W. (2015). Package 'psych.' *The Comprehensive R Archive Network*, *337*(338).

Ricci, F., Rokach, L., & Shapira, B. (2022). Recommender Systems Handbook. In *Recommender Systems Handbook* (3rd ed.). Springer Science. https://doi.org/10.1007/978-1-0716-2197-4_14

Rinschede, G. (1992). Forms of religious. *Annals of Tourism Research*, *19*, 51–67.

Rodríguez, M. (2019). *El impacto económico de los peregrinos supera por primera vez los 300 millones - Faro de Vigo*. Faro de Vigo. https://www.farodevigo.es/galicia/2019/02/18/impacto-economico-peregrinos-supera-primera-15770580.html

Ronen, S., & Shenkar, O. (1985). Clustering Countries on Attitudinal Dimensions: A Review and Synthesis. *Academy of Management Review*, *10*(3), 435–454. https://doi.org/10.5465/amr.1985.4278955

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386–408.

https://doi.org/10.1037/h0042519

Rossteutscher, S. (2004). Explaining politics: An empirical test of competing value measures. *European Journal of Political Research*, *43*(5), 769–795. https://doi.org/10.1111/j.0304-4130.2004.00174.x

Saavedra, P., Barreiro, P., Durán, R., Crujeiras, R., Loureiro, M., & Vila, E. S. (2016). Choice-Based Recommender Systems. *RecTour@ RecSys*, *1685*, 38–46.

Sandholtz, W., & Taagepera, R. (2005). Corruption, culture, and communism. *International Review of Sociology*, *15*(1), 109–131. https://doi.org/10.1080/03906700500038678

Sarwar, B., Karypis, G., Constan, J., & Riedl, J. (2000). *An introduction to statistical learning* (2nd ed.). Springer.

Sasaki, Y. (2015). The truth of the F-measure The truth of the F-measure. *International Conference on Information Retrieval, January 2007*, 1–6.

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., Er, M. J., Ding, W., & Lin, C. T. (2017). A review of clustering techniques and developments. *Neurocomputing*, *267*, 664–681. https://doi.org/10.1016/j.neucom.2017.06.053

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology*, *25*(C), 1–65. https://doi.org/10.1016/S0065-2601(08)60281-6

Schwartz, S. H. (1994). *Beyond individualism/collectivism: New cultural dimensions of values*. SAGE Publications Inc.

Schwartz, S. H. (1999). A Theory of Cultural Values and Some Implications for Work. *Applied Psychology: An International Review*, *48*(1), 23–47.

Schwartz, S. H. (2014). Rethinking the Concept and Measurement of Societal Culture in Light of Empirical Findings. *Journal of Cross-Cultural Psychology*, *45*(1), 5–13. https://doi.org/10.1177/0022022113490830

Shenkar, O. (2001). Revisited : and Conceptualization Measurement Cultural. *Journal of International Business Studies*, *32*(3), 519–535.

Sirakaya, E., & Woodside, A. G. (2005). Building and testing theories of decision making by travellers. *Tourism Management*, *26*(6), 815–832. https://doi.org/10.1016/j.tourman.2004.05.004

Smith, K. (1993). The influence of weather and climate on recreation and tourism. *Weather*, *48*(12), 398–404. https://doi.org/10.1002/j.1477-8696.1993.tb05828.x

Solomon, M. R., & Panda, T. K. (2004). *Consumer behavior, buying, having, and being*. Pearson Education India.

Train, K. E. (2009). Discrete choice methods with simulation, second edition. In *Discrete Choice Methods with Simulation, Second Edition* (Vol. 9780521766). https://doi.org/10.1017/CBO9780511805271

Tung, H. W., & Soo, V. W. (2004). A personalized restaurant recommender agent for mobile E-service. *Proceedings - 2004 IEEE International Conference on e-Technology, e-*

Commerce and e-Service, *EEE 2004*, 259–262. https://doi.org/10.1109/eee.2004.1287319

Tversky, A., & Shafir, E. (1992). Choice under conflict:The Dynamics of Deferred Decision. *Psychological Science*, *3*(6), 358–361.

Uysal, M., & Jurowski, C. (1994). Testing the push and pull factors. *Annals of Tourism Research*, *21*(4), 844–846. https://doi.org/10.1016/0160-7383(94)90091-4

Uysal, M., Li, X., & Sirakaya-turk, E. (2008). Push – pull dynamics in travel decisions. In *Handbook of Hospitality Marketing Management* (Oxford, pp. 412–439). Butterworth-Heinemann. https://doi.org/https://doi.org/10.1016/B978-0-08-045080-3.50018-4

van Cranenburgh, S., Wang, S., Vij, A., Pereira, F., & Walker, J. (2022). Choice modelling in the age of machine learning - Discussion paper. *Journal of Choice Modelling*, *42*(December 2021). https://doi.org/10.1016/j.jocm.2021.100340

Varian, H. (2010). *Intermediate Microeconomics* (8th Editio). W.W.Norton & Company.

Vijver, V. de, & Leung, K. (1997). *Methods and data analysis of comparative research*. Allyn & Bacon.

Vijver, V. de, & Leung, K. (2021). *Methods and data analysis for cross-cultural research*. Cambridge University Press.

Walters, G., Wallin, A., & Hartley, N. (2019). The Threat of Terrorism and Tourist Choice Behavior. *Journal of Travel Research*, *58*(3), 370–382. https://doi.org/10.1177/0047287518755503

Wang, X. T. (1996). Framing Effects: Dynamics and Task Domains able models of human decision making. *Organizational Behavior and Human Decision Processes*, *68*(2), 145–157.

Weber, E., & Hsee, C. (2000). Culture and Individual Judgment and Decision Making. *Applied Psychology: An International Review*, *49*(1), 32–61.

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., Zemla, J., & others. (2017). Package 'corrplot.' *Statistician*, *56*(316), e24.

Welzel, C. (2012). Freedom rising: Human empowerment and the quest for emancipation. In *Freedom Rising: Human Empowerment and the Quest for Emancipation*. https://doi.org/10.1017/CBO9781139540919

Welzel, C. (2013). *Freedom Rising: Human Empowerment and the Quest for Emancipation*. Cambridge University Press.

Wickham, H., Averick, M., Bryan, J., & Mcgowan, L. (2019). Welcome to the tidyverse. *J. Open Source Softw.*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Woodside, A. G., Hsu, S. Y., & Marshall, R. (2011). General theory of cultures' consequences on international tourism behavior. *Journal of Business Research*, *64*(8), 785–799. https://doi.org/10.1016/j.jbusres.2010.10.008

Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data. *Booksgooglecom*, *58*(2), 752. https://doi.org/10.1515/humr.2003.021

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). The MIT Press.

Wooldridge, J. M. (2016). *Introductory Econometrics* (5th ed.). South-Western.

Wright, G. N., & Phillips, L. D. (1980). Cultural Variation. *International Journal of Psychology, 15*, 239–257. https://doi.org/10.4324/9781003081685-3

WTO. (2023). *Re-globalization for a secure, inclusive and sustainable future What is the World Trade Report?* www.wto.org

WTTC. (2020). Travel & Tourism. Global Economic Impact Trends 2020. *World Travel & Tourism Council*, *June*, 1–20. https://wttc.org/Portals/0/Documents/Reports/2020/Global Economic Impact Trends 2020.pdf?ver=2021-02-25-183118-360

Xia, F., Zhang, W., Li, F., & Yang, Y. (2008). Ranking with decision tree. *Knowledge and Information Systems*, *17*(3), 381–395. https://doi.org/10.1007/s10115-007-0118-y

Yang, Y., Liu, H., & Li, X. (2019). The World Is Flatter? Examining the Relationship between Cultural Distance and International Tourist Flows. *Journal of Travel Research, 58*(2), 224–240. https://doi.org/10.1177/0047287517748780

Yang, Y., & Wong, K. (2012). The influence of cultural distance on China inbound tourism flows: A panel data gravity model approach. *Asian Geographer*, *29*(1), 21–37.

Yates, J. F., & de Oliveira, S. (2016). Culture and decision making. *Organizational Behavior and Human Decision Processes, 136,* 106–118. https://doi.org/10.1016/j.obhdp.2016.05.003

Yiamjanya, S., & Wongleedee, K. (2014). International tourists' travel motivation by push-pull factors and the decision making for selecting thailand as destination choice. *Social, Behavioral, Educational, Economic and Management Engineering*, *8*(5), 1335–1340.

Zhang, Y., Li, X. R., & Wu, T. S. (2019). Current Issues in Tourism The impacts of cultural values on bilateral international tourist flows : a panel data gravity model. *Current Issues in Tourism*, *22*(8), 967–981. https://doi.org/10.1080/13683500.2017.1345870

ANNEX

ANNEX 1 INTENSITY MEASURE

Precipitation Index

$$PI = \frac{1 - P_i}{1 - MAX(P_i)} \tag{30}$$

Temperature Index

$$TI = \frac{1 - ABS(20 - T_i)}{1 - MAX(ABS(20 - T_i))} \tag{31}$$

Distance Weight

$$DW = \begin{cases} 1, & D < 140\ km \\ 1.2, & 141km < D < 235\ km \\ 1.4, & 236km < D \leq 400\ km \\ 1.6, & D > 400\ km \end{cases} \tag{32}$$

Difficulty Index

$$DifI = \frac{(1 - Dif_i * DW_i)}{(1 - MAX(Dif_i * DW_i))} \tag{33}$$

Intensity Index

$$II = \frac{N\ of\ \frac{stages(2 * DifI + PI + TI)}{4}}{N\ of\ stages} \tag{34}$$

Intensity

$$Intensity = \begin{cases} 0, & II < .65 \\ 1, & II \geq .65 \end{cases} \tag{35}$$

P denotes Precipitation (AEMET, 2021) (measured in average mm per month), PI Precipitation Index, T temperature (AEMET, 2021) (measured in average degrees Celsius), TI Temperature Index, D Distance, DW Distance Weight, Dif (Gronze, 2021) Difficulty (measured as average difficulty of the stages of each way) , DifI the Difficulty Index, and II the Intensity Index. The concept of "stage" corresponds to the typical distance a pilgrim covers each day during the pilgrimage. This study categorizes pilgrims into four distinct ranges of distances, with the shortest (< 140 km) specifically targeting those who complete the minimum required for a Compostela certificate, while the other ranges of distances are designed to achieve similar-sized groups in terms of pilgrim arrivals between 2003-2022. We assessed intensity as the absolute difference from a reference temperature (20 ºC), recognizing that significant deviations, either exceeding or falling below, pose challenges to outdoor sporting activities as well as the precipitation, acknowledging its adverse effects on the enjoying of outdoor activities (Hall & Ram, 2021).

ANNEX 2 TESTS

Pearson's chi-squared

This test is applied to determine whether two or more probability distributions come from the same distribution by measuring the discrepancy between them. The Pearson's chi-squared statistic (Wooldridge, 2002) takes the following form:

$$\chi^2 = \sum_{i}^{n} \frac{(O_i - E_i)^\wedge 2}{E_i} \tag{36}$$

where O_i is the frequency of the O alternative and E_i the frequency of the E alternative. Once the χ^2statistic is estimated and the degrees of freedom established, we calculate the p-value, which determines the statistical significance of the coefficient.

Wald test

This test is applied to determine the statistical significance of each independent variable. For each binary variable, the Wald statistic (Wooldridge, 2016) takes the following form:

$$W = \frac{(\theta - \theta_0)^2}{var(\theta)} \tag{37}$$

where θ is the estimation that maximizes the unconstrained likelihood function and θ_0 the constrained likelihood function. Once the W statistic is estimated and the degrees of freedom set, we calculate the p-value, which determines the statistical significance of the coefficient.

ANNEX 3 MODELS

Table 35. Control variables results

Variables	Inglehart National Values	Hofstede National Values	Schwartz National Values
High Popularity	2.284 ***	3.629 ***	2.249 ***
Medium Popularity	.988 ***	2.882 ***	1.010 ***
Intensity	-.294 ***	-3.354 ***	-.303 ***
World Heritage Status	.010	-1.941 ***	.028
Gender**:** Female **(2)**	.198 ***	.245 ***	.201 ***
Gender**:** Female **(3)**	.005	-.009	-.001
Gender**:** Female **(4)**	-.334 ***	-.326 ***	-.329 ***
Gender**:** Female **(5)**	-.511 ***	-.360 ***	-.516 ***
Gender**:** Female **(6)**	.098 ***	.139 ***	.112 ***
Gender**:** Female **(7)**	-.226 ***	-.281 ***	-.215 ***
Age: **<18** **(2)**	-.833 ***	-.841 ***	-.845 ***
Age: <18(**3**)	-.206 ***	-.193 ***	-.248 ***
Age: <18**:** **(4)**	-.687 ***	-.633 ***	-.660 ***
Age: <18(**5**)	-.035	.253 ***	-.065 **
Age: <18(**6**)	-.367 ***	-.394 ***	-.406 ***
Age: <18(**7**)	-.693 ***	-.798 ***	-.699 ***
Age: **46-65** (**2**)	-.116 ***	-.048 ***	-.103 ***
Age: 46-65(**3**)	-.042 ***	-.016	-.021
Age: 46-65(**4**)	-.254 ***	-.245 ***	-.240 ***
Age: 46-65(**5**)	.028	.249 ***	.052 **
Age: 46-65(**6**)	-.001	.072 ***	.017 *
Age: 46-65(**7**)	-.003	-.052 ***	.008
Age: **>65** (**2**)	-.121 ***	.058 **	-.099 ***
Age: >65(**3**)	-.121 ***	-.025	-.092 ***
Age: >65 (4)	-.038	-.065 **	-.003
Age: >65(**5**)	.343 ***	.596 ***	-.370 ***
Age: >65(**6**)	-.091 ***	.126 ***	-.056 ***
Age: >65(**7**)	.030	-.017	.048 *
Group: **individuals 2**	-.281 ***	-.225 ***	-.231 ***
Group: individuals **3**	-.140 ***	-.161 ***	-.131 ***
Group: individuals **4**	.260 ***	.222 ***	.290 ***
Group: individuals **5**	-.632 ***	-.481 ***	-.678 ***
Group: individuals **6**	.028 ***	.073 ***	.048 ***
Group: individuals **7**	.176 ***	.057 ***	.152 ***

*** Significance 1%; ** Significance 5%; * Significance 10%

Source: own elaboration based on Pilgrim's Welcome Office dataset

Table 36. Control variables for each cluster

	Dependent variable: Way	West and South Asia	Catholic Europe	Latin America	Orthodox Europe	Protestant Europe	Confucian	Spain	English-Speaking
Coastal	Gender: Female	.312 ***	.423 ***	.314 ***	.005	.461 ***	.528 ***	.101 ***	.298 ***
	Age (Ref: 18-45 years)								
	< 18 years	.427	-.694 ***	-.183	-.495	.356 **	-.352	-.804 ***	-.327 ***
	46-65 years	.307 **	-.071 ***	.009	-.107	-.333 ***	-.653 ***	-.147 ***	.280 ***
	>65 years	.499 ***	-.685 ***	.186 *	-1.624 **	-.663 ***	-.143	.075 **	.385 ***
	Group: Individual	-.381 ***	-.346 ***	-.082	.534 ***	.260 ***	.266	-.454 ***	-.208 ***
English	Gender: Female	-.385 ***	-.158 ***	-.467 ***	-.180	.049	.208	.043 ***	-.416 ***
	Age (Ref: 18-45 years)								
	< 18 years	2.215 **	-.815 ***	-.056	-.200	-.330 *	.073	-.136 ***	-.474 ***
	46-65 years	-.790 ***	-.196 ***	-.579 ***	-.156	-.402 ***	-.258	.058 ***	-.356 ***
	>65 years	-1.911 ***	-.806 ***	-.515 ***	-17.095	-.620 ***	-.033	.161 ***	-.404 ***
	Group: Individual	-.374 **	-.722 ***	-.071	-.232	-.335 ***	-.272	-.153 ***	.014
North	Gender: Female	-.474 ***	-.142 ***	-.448 ***	-.098	-.234 ***	-.070	-.406 ***	-.345 ***
	Age (Ref: 18-45 years)								
	< 18 years	-.502	-1.155 ***	-.791 ***	-.653	-.321 **	.186	-.556 ***	-1.131 ***
	46-65 years	-.393 **	-.121 ***	-.419 ***	.005	-.702 ***	-.152	-.187 ***	-.345 ***
	>65 years	-.005	-.089 *	-.329 **	-.249	-.659 ***	.051	.235 ***	-.514 ***
	Group: Individual	-.078	.497 ***	.071	1.484 ***	.505 ***	.283	.072 ***	.471 ***
Silver	Gender: Female	-.389	-.643 ***	-.917 ***	-.871 ***	-.702 ***	-.426 *	-.440 ***	-.834 ***

	Age (Ref: 18-45 years)								
	< 18 years	-16.353	-.900 ***	-1.303 **	-1.279	-.230	-15.738	.041	-1.801 ***
	46-65 years	.275 ***	.257 ***	-.644 ***	-.193	.335 ***	.227	.065 ***	-.834 ***
	>65 years	-2.134 **	.306 ***	-.472 **	-.447	.794 ***	1.090 ***	.409 ***	-.448 ***
	Group: Individual	-1.753 ***	-.544 ***	-.450 ***	-.241	-.472 ***	-.019	-.773 ***	-.281 ***
Portuguese	Gender: Female	.115	.153 ***	.091 **	.088	.258 ***	.480 ***	.073 ***	.060 **
	Age (Ref: 18-45 years)								
	< 18 years	.316	-.417 ***	-.087	-.200	-.248 ***	-.887 *	-.170 ***	-.469 ***
	46-65 years	.247 **	-.191 ***	.123 ***	.004	-.299 ***	-.341 ***	.146 ***	.060 **
	>65 years	-.258 *	-.977 ***	.225 ***	-.165	-.552 ***	-.296 *	.234 ***	.328 ***
	Group: Individual	.009	-.227 ***	.445 ***	.537 ***	.338 ***	.780 ***	-.203 ***	.157 ***
Primitive	Gender: Female	.091	-.116 ***	.044 **	.044	-.271 ***	.272	-.203 ***	-.381 ***
	Age (Ref: 18-45 years)								
	< 18 years	-1.289 *	-.858 ***	.254	.640 **	-.314	.137	-.673 ***	-.454 ***
	46-65 years	1.008 ***	-.224 ***	-.345 ***	-.390 **	-.184 ***	-1.041 ***	.244 ***	-.381 ***
	>65 years	-1.041 ***	-.561 ***	-.165	-.071	-.792 ***	-.925 **	.549 ***	-.768 ***
	Group: Individual	.754 ***	.530 ***	.682 ***	.997 ***	.759 ***	.834 ***	-.019	.716 ***

*** Significance 1%; ** Significance 5%; * Significance 10%

Source: own elaboration based on Pilgrim's Welcome Office dataset

Table 37. Control variables for each country (I)

	Dependent variable: Way	Spain	Italy	USA	Germany	Portugal
Coastal	Gender: Female	.101 ***	.416 ***	.236 ***	.559 ***	.238 ***
	Age (Ref: 18-45 years)					
	<18 years	-.805 ***	-.556 ***	.303 **	-.516 ***	-2.289 ***
	46-65 years	-.147 ***	-.052	.252 ***	-.300 ***	-.050
	>65 years	.074 **	.126	.350 ***	-.615 ***	-.391 ***
	Group: individual	-.456 ***	-.288 ***	-.264 ***	.412 ***	.069
English	Gender: Female	.043 ***	.112 **	-.513 ***	.041	-.230 ***
	Age (Ref: 18-45 years)					
	<18 years	-.137 ***	-.462 ***	-.037	-.680 ***	-2.306 ***
	46-65 years	0.058 ***	.262 ***	-.743 ***	-.277 ***	-.382 ***
	>65 years	.161 ***	.170 **	-.704 ***	-.563 ***	-.573 ***
	Group: individual	-.155 ***	-.225 ***	-.192 ***	-.264 ***	-.665 ***
North	Gender: Female	-.406 ***	-.376 ***	-.236 ***	-.160 ***	-.303 *
	Age (Ref: 18-45 years)					
	<18 years	-.557 ***	-.708 ***	-.464 **	-.414 **	-3.513 ***
	46-65 years	-.187 ***	.120 *	-.561 ***	-.431 ***	.427 **
	>65 years	.235 ***	.136	-.483 ***	-.605 ***	.217
	Group: individual	.070 ***	.751 ***	.285 ***	.392 ***	-.067
Silver	Gender: Female	-.438 ***	-1.216 ***	-.456 ***	-.977 ***	-.231 **
	Age (Ref: 18-45 years)					
	<18 years	.042	-.862 ***	-.656 *	-.194	-1.945 ***
	46-65 years	.068 ***	.310 ***	-.707 ***	.152	.432 ***
	>65 years	.412 ***	.874 ***	-.643 ***	.284 **	.125
	Group: individual	-.774 ***	-.617 ***	-.426 ***	-.282 **	-.423 ***
Portuguese	Gender: Female	.073 ***	.168 ***	0.153 ***	.302 ***	.209 ***
	Age (Ref: 18-45 years)					
	<18 years	-.172 ***	.092	.186 *	-.338 ***	-1.875 ***
	46-65 years	.146 ***	-.013	.267 ***	-.334 ***	-.320 ***
	>65 years	.232 ***	-.135 **	.353 ***	-.680 ***	-.772 ***
	Group: individual	-.203 ***	-.035	.218 ***	.330 ***	.019
Primitive	Gender: Female	-.206 ***	-.245 ***	-.185 ***	-.283 ***	-.441 ***
	Age (Ref: 18-45 years)					
	<18 years	-.674 ***	-.725 ***	.480 ***	-.254	-2.688 ***
	46-65 years	.239 ***	-.051	-.375 ***	-.093	-.085
	>65 years	.545 ***	-.026	-.672 ***	.843 ***	-.922 ***
	Group: individual	-.014	.644 ***	.854 ***	.704 ***	.181 **

*** Significance 1%; ** Significance 5%; * Significance 10%

Source: own elaboration based on Pilgrim's Welcome Office dataset

Table 38. Control variables for each country (II)

	Dependent variable: Way	France	UK	Ireland	Mexico	South Korea
Coastal	Gender: Female	.151	.243 ***	.366 ***	.211 *	.523 **
	Age (Ref: 18-45 years)					
	<18 years	.924 ***	.128	-1.481 ***	.155	-.089
	46-65 years	-.072	-.160 *	.631 ***	-.023	-1.486 ***
	>65 years	-.011	.124	.544 ***	.015	-.035
	Group: individual	-.635 ***	-.332 ***	-.066	-.103	.281
English	Gender: Female	.032	-.224 ***	-.305 ***	-.753 ***	-.169
	Age (Ref: 18-45 years)					
	<18 years	-.155	.400	-1.115 ***	-.198	-.180
	46-65 years	-.982 ***	-.172 *	.203 *	-.553 ***	-.990 ***
	>65 years	-1.320 ***	.078	.338 **	-.324	-.945
	Group: individual	-1.054 ***	-.031	.334 ***	-.060	-.297
North	Gender: Female	-.143 **	-.536 ***	-.818 ***	-.513 ***	-.279
	Age (Ref: 18-45 years)					
	<18 years	-.852 ***	-.117	-3.595 ***	-1.026 **	.022
	46-65 years	-.415 ***	-.780 ***	-.825 ***	-.355 ***	-.033
	>65 years	-.454 ***	-.786 ***	-.192	-.283	.581 **
	Group: individual	-.227 ***	.538 ***	.986 ***	-.589 ***	.770 ***
Silver	Gender: Female	-.814 ***	-.999 ***	-2.161 ***	-1.026 ***	-.702 **
	Age (Ref: 18-45 years)					
	<18 years	-1.757 *	-1.939 *	-18.186	-1.155	-15.712
	46-65 years	.172	-.611 ***	-.953 ***	-.732 ***	.349
	>65 years	.668 ***	-.610 ***	-.240	-.633	1.269 ***
	Group: individual	-.017	-.601 ***	.220	-.448 **	.829 **
Portuguese	Gender: Female	.149 **	-.158 ***	-.224 ***	.019	.423 ***
	Age (Ref: 18-45 years)					
	<18 years	.344 *	-.354	-1.405 ***	-.017	-1.188
	46-65 years	-.253 ***	.025	.111	-.033	-.417 ***
	>65 years	-.021	.218 **	.293 ***	.326 **	-.285
	Group: individual	-.095	.185 **	.119	.061	.570 ***
Primitive	Gender: Female	-.090	-.621 ***	-.918 ***	-.524 ***	-.219
	Age (Ref: 18-45 years)					
	<18 years	.420	.018	-2.696 ***	.496 *	-15.848
	46-65 years	-.220	-1.014 ***	-.512 ***	-.649 ***	-.282
	>65 years	-.106	-.843 ***	-.757 **	-.957 ***	-.034
	Group: individual	.324 **	.159	.951 ***	.932 ***	.762 *

*** Significance 1%; ** Significance 5%; * Significance 10%

Source: own elaboration based on Pilgrim's Welcome Office dataset

Table 39. Control variables for each country (III)

Attributes of the Way		Switzerland	New Zealand
High Popularity		1.010 ***	1.594 ***
Medium Popularity		1.127 ***	2.216 ***
Intensity		-1.071 ***	-1.673
World Heritage Status		1.479 ***	1.811
Dependent variable: Way			
Coastal	Gender: Female	.369 **	1.164 ***
	Age (Ref: 18-45 years)		
	<18 years	.676	2.770 *
	46-65 years	-.436 **	-.336
	>65 years	-.576 **	2.050 ***
	Group: individual	-.160	-.753
English	Gender: Female	.095	.649
	Age (Ref: 18-45 years)		
	<18 years	.730	-17.927
	46-65 years	-.594	-.022
	>65 years	-.382	-.476
	Group: individual	-.804 ***	-.112
North	Gender: Female	-.114	.712 *
	Age (Ref: 18-45 years)		
	<18 years	.574	-19.299
	46-65 years	-.673 ***	-1.547 ***
	>65 years	-1.106 ***	-1.194 *
	Group: individual	.188	-.153
Silver	Gender: Female	.328	-.117
	Age (Ref: 18-45 years)		
	<18 years	-15.024	-16.207
	46-65 years	.685 *	-.610
	>65 years	1.898 ***	-16.407
	Group: individual	-1.655 ***	16.500
Portuguese	Gender: Female	.098	.346
	Age (Ref: 18-45 years)		
	<18 years	-.212	-19.783
	46-65 years	-.071	-.261
	>65 years	-.420 **	.437
	Group: individual	-.172	-.326
Primitive	Gender: Female	-.124	.067
	Age (Ref: 18-45 years)		
	<18 years	.169	-17.885
	46-65 years	.048	-1.830 ***
	>65 years	-.380	-.725
	Group: individual	.216	17.972

*** Significance 1%; ** Significance 5%; * Significance 10%

Source: own elaboration based on Pilgrim's Welcome Office dataset

ANNEX 4 DESCRIPTIVE STATISTICS

Table 40. National culture variables (I)

Country	Surv_Self	Trad_Sec	Sch_harmony	Sch_embedded	Sch_hierarchy	Sch_mastery	Sch_aff_auton	Sch_intel_auton
Argentina	0,23	-0,38	3,98	3,52	2,1	3,92	3,73	4,34
Australia	2,27	0,64	3,99	3,59	2,29	3,97	3,86	4,35
Austria	1,95	0,69	4,31	3,11	1,75	3,92	4,29	4,9
Belgium	1,5	0,2	4,35	3,25	1,69	3,84	3,94	4,64
Brazil	-0,05	-0,31	4,03	3,62	2,37	3,93	3,52	4,27
Bulgaria	-0,55	0,67	4,13	3,87	2,68	4,02	3,47	4,29
Canada	2,03	-0,6	3,83	3,46	2,09	4,12	4	4,5
Chile	-0,09	0,22	4,33	3,64	2,25	3,78	3,03	4,32
China	-0,13	0,6	3,78	3,74	3,49	4,41	3,3	4,18
Colombia	0,35	-1,56	3,66	3,86	2,9	4,03	3,61	4,3
Costa Rica			4,13	3,49	2,29	4,01	3,49	4,37
Croatia	0,11	0,1	4,02	4	2,55	4,05	3,92	4,35
Czech Republic	2,88	1,13	4,27	3,59	2,22	3,75	3,49	4,62
Denmark	2,88	1,13	4,16	3,19	1,86	3,91	4,3	4,77
Slovakia	0,56	0,62	4,47	3,82	2	3,83	2,99	4,29
Estonia	0,69	0,94	4,31	3,81	2,04	3,79	3,36	4,23
Finland	2,45	0,89	4,34	3,37	1,8	3,66	3,96	4,93
France	1,91	0,54	4,21	3,2	2,21	3,72	4,39	5,13
Germany	2,16	0,97	4,62	3,03	1,87	3,86	4,11	4,99
Hungary	0,02	0,61	4,34	3,6	1,94	3,73	3,63	4,57
Ireland	1,18	-0,8	3,77	3,41	2,09	4,04	4,05	4,54
Israel	0,61	0,22	3,28	3,61	2,51	4,02	3,79	4,54
Italy	0,82	0,38	4,62	3,46	1,6	3,81	3,3	4,91
Japan	1,29	1,66	4,21	3,49	2,65	4,06	3,76	4,78
Latvia	-0,48	0,89	4,46	3,83	1,8	3,75	3,48	4,22
Malaysia	-0,36	0,05	3,65	4,35	2,25	3,91	2,98	4,15
Mexico	0,29	-1,05	4,5	3,9	2,13	3,9	2,83	4,36
Netherlands	2,5	0,8	4,05	3,19	1,91	3,97	4,13	4,85
New Zealand	2,86	0,64	4,03	3,27	2,27	4,09	4,21	4,65
Norway	3,03	1,04	4,4	3,45	1,49	3,85	3,69	4,68
Peru	-0,5	-1,07	3,71	3,92	2,76	4,08	2,98	4,3
Philippines	0,18	-1,4	4,04	4,03	2,68	3,76	3	3,95
Poland	0,64	-0,41	3,86	3,86	2,51	3,84	3,32	4,31
Portugal	0,36	-0,24	4,27	3,43	1,89	4,11	3,62	4,53

Country	Surv_Self	Trad_Sec	Sch_harmony	Sch_embedded	Sch_hierarchy	Sch_mastery	Sch_aff_auton	Sch_intel_auton
Puerto Rico	0,74	-1,53						
Romania	-0,77	-0,41	4,11	3,78	2	4,06	3,45	4,61
Russia	-0,6	0,39	3,9	3,81	2,72	3,96	3,51	4,3
Singapore	-0,12	0,09	3,76	4	2,82	3,88	3,3	3,86
Slovenia	1,23	0,83	4,45	3,71	1,62	3,71	3,72	4,88
South Africa	0,06	-0,26	3,86	4,03	2,59	3,89	3,48	3,85
South Korea	-0,5	1,45	3,57	3,68	2,9	4,21	3,46	4,22
Spain	1,44	0,65	4,47	3,31	1,84	3,8	3,67	4,99
Sweden	3,11	1,22	4,46	3,12	1,83	3,81	4,24	5,09
Switzerland	2,36	0,76	4,4	3,04	2,06	3,74	4,33	5,32
Taiwan	-0,15	1,13	4,12	3,82	2,69	4	3,27	4,36
UK	2,35	0,49	3,91	3,34	2,33	4,01	4,26	4,62
Ukraine	-0,41	0,5	3,87	3,93	2,56	3,99	3,49	4,08
Uruguay	0,8	-0,15						
USA	1,51	0,58	3,46	3,67	2,37	4,09	3,87	4,19
Venezuela	0,24	-1,72	3,99	3,74	2,09	4,01	3,26	4,44

Source: own elaboration based on Hofstede, Schwartz and Inglehart national values

Table 41. National culture variables (II)

Country	Sch_egalitar	Hof_pdi	Hof_idv	Hof_mas	Hof_uai	Hof_ltowvs	Hof_ivr
Argentina	4,96	49	46	56	86	20	62
Australia	4,79	38	90	61	51	21	71
Austria	4,89	11	55	79	70	60	63
Belgium	5,2	65	75	54	94	82	57
Brazil	4,89	69	38	49	76	44	59
Bulgaria	4,13	70	30	40	85	69	16
Canada	4,8	39	80	52	48	36	68
Chile	5,06	63	23	28	86	31	68
China	4,23	80	20	66	30	87	24
Colombia	4,69	67	13	64	80	13	83
Costa Rica	4,85	35	15	21	86		
Croatia	4,6	73	33	40	80	58	33
Czech Republic	4,45	57	58	57	74	70	29
Denmark	5,03	18	74	16	23	35	70
Slovakia	4,58	104	52	110	51	77	28
Estonia	4,58	40	60	30	60	82	16
Finland	4,9	33	63	26	59	38	57
France	5,05	68	71	43	86	63	48

Country	Sch_egalitar	Hof_pdi	Hof_idv	Hof_mas	Hof_uai	Hof_ltowvs	Hof_ivr
Germany	5,07	35	67	66	65	83	40
Hungary	4,51	46	80	88	82	58	31
Ireland	4,9	28	70	68	35	24	65
Israel	4,77	13	76	47	81	38	
Italy	5,27	50	76	70	75	61	30
Japan	4,36	54	46	95	92	88	42
Latvia	4,32	44	70	9	63	69	13
Malaysia	4,41	104	26	50	36	41	57
Mexico	4,73	81	30	69	82	24	97
Netherlands	5,03	38	80	14	53	67	68
New Zealand	4,94	22	79	58	49	33	75
Norway	5,12	31	69	8	50	35	55
Peru	4,84	64	16	42	87	25	46
Philippines	4,59	94	32	64	44	27	42
Poland	4,48	68	60	64	93	38	29
Portugal	5,21	63	27	31	104	28	33
Puerto Rico							
Romania	4,48	90	30	42	90	52	20
Russia	4,38	93	39	36	95	81	20
Singapore	4,6	74	20	48	8	72	46
Slovenia	4,56	71	27	19	88	49	48
South Africa	4,52	49	65	63	49	34	63
South Korea	4,42	60	18	39	85	100	29
Spain	5,23	57	51	42	86	48	44
Sweden	4,9	31	71	5	29	53	78
Switzerland	5,06	34	68	70	58	74	66
Taiwan	4,31	58	17	45	69	93	49
UK	4,92	35	89	66	35	51	69
Ukraine	4,31					86	14
Uruguay		61	36	38	100	26	53
USA	4,68	40	91	62	46	26	68
Venezuela	4,77	81	12	73	76	16	100

Source: own elaboration based on Hofstede, Schwartz and Inglehart national values

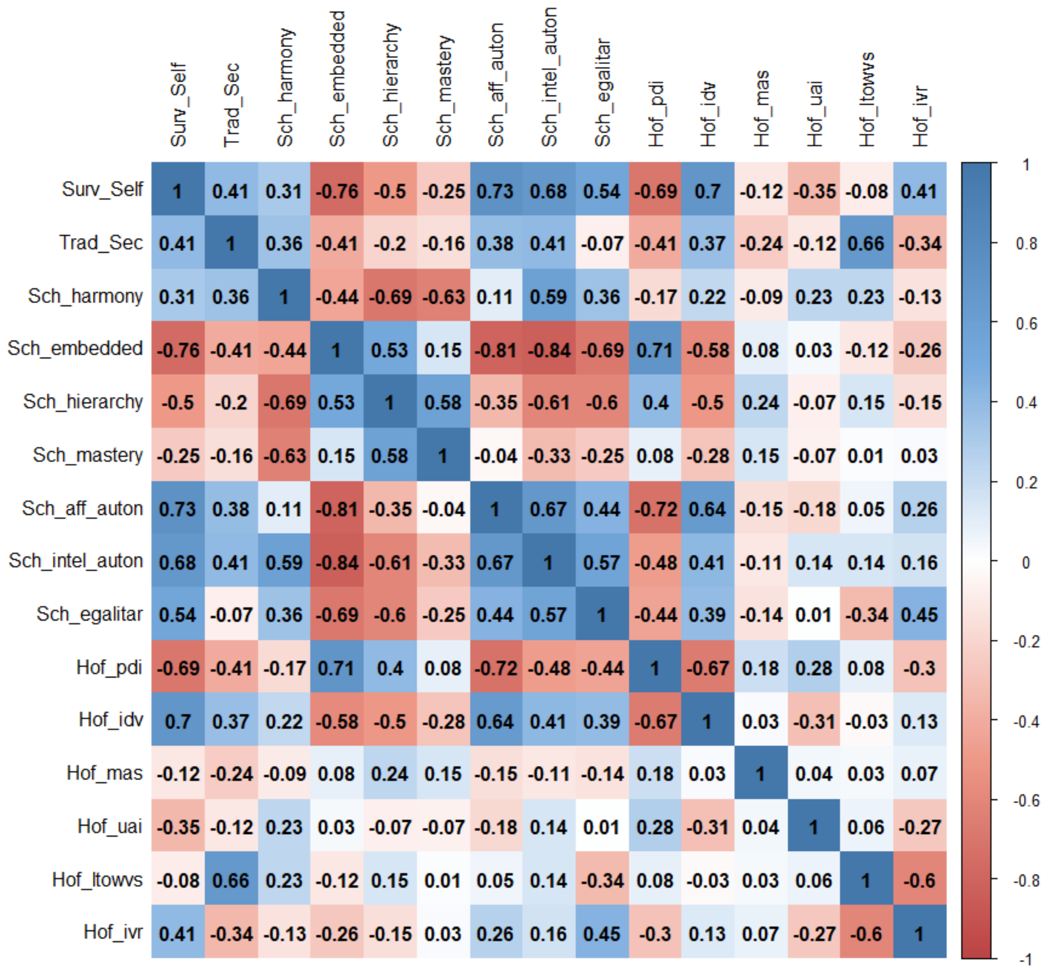
Table 42. Cultural distance from each country to Spain

Country	EUC_Inglehart	KS_Inglehart	EUC_Hofstede	KS_Hofstede	EUC_Schwartz	KS_Schwartz
Argentina	1.59	1.20	36.88	1.30	0.93	3.92
Australia	0.83	0.23	70.33	4.78	1.09	5.51
Austria	0.52	0.09	65.50	4.56	0.77	2.14
Belgium	0.45	0.13	46.85	2.00	0.49	0.74
Brazil	1.77	1.36	26.84	0.72	1.12	5.09
Bulgaria	1.99	1.32	42.89	1.76	1.71	14.88
Canada	1.38	1.15	58.80	3.38	1.06	7.02

Country	EUC_Inglehart	KS_Inglehart	EUC_Hofstede	KS_Hofstede	EUC_Schwartz	KS_Schwartz
Chile	1.59	0.90	43.49	1.84	1.09	3.35
China	1.57	0.82	84.74	6.94	2.36	27.83
Colombia	2.46	3.65	69.40	4.66	1.70	12.38
Costa Rica	1.58	0.97			0.98	4.60
Croatia	1.44	0.79	29.03	0.80	1.45	9.57
Czech R.	1.52	0.85	33.44	1.09	1.02	5.78
Denmark	1.52	0.85	86.85	7.57	0.78	2.01
Slovakia	0.88	0.26	95.54	9.91	1.29	6.40
Estonia	0.80	0.24	55.88	2.99	1.19	6.17
Finland	1.04	0.38	44.54	2.01	0.49	1.47
France	0.48	0.08	28.14	0.72	0.89	2.15
Germany	0.79	0.24	54.87	2.91	0.57	0.98
Hungary	1.42	0.67	57.89	3.67	0.90	4.79
Ireland	1.47	1.43	73.99	5.48	1.04	5.89
Israel	0.93	0.35	63.18	4.08	1.56	12.67
Italy	0.68	0.18	44.30	2.03	0.50	0.63
Japan	1.02	0.69	67.13	4.81	1.28	9.21
Latvia	1.93	1.27	59.42	3.67	1.32	8.88
Malaysia	1.89	1.32	75.03	5.56	1.95	15.85
Mexico	2.05	2.36	72.14	5.35	1.34	5.91
Netherlands	1.07	0.39	63.76	4.09	0.70	2.39
New Zealand	1.43	0.68	69.41	4.79	0.98	4.87
Norway	1.64	0.95	61.30	3.91	0.51	0.87
Peru	2.59	3.23	42.25	1.59	1.73	11.81
Philippines	2.41	3.34	66.23	4.37	1.83	11.91
Poland	1.32	0.95	32.18	1.12	1.51	10.04
Portugal	1.40	0.92	39.23	1.45	0.61	3.01
Puerto Rico	2.29	3.33				
Romania	2.44	2.37	46.09	2.11	1.09	7.50
Russia	2.05	1.43	57.06	3.12	1.61	11.99
Singapore	1.66	1.02	89.14	7.67	1.94	14.34
Slovenia	0.28	0.04	36.36	1.37	0.82	4.05
South Africa	1.65	1.19	51.33	2.67	1.82	13.39
South Korea	2.10	1.69	63.65	3.56	1.88	18.24
Sweden	1.77	1.15	83.00	7.11	0.69	1.52
Switzerland	0.93	0.29	59.73	3.61	0.84	1.82
Taiwan	1.66	1.00	59.51	3.13	1.60	11.92
Ukraine	1.86	1.16	131.79	17.18	1.73	14.24
United Kingdom	0.92	0.29	76.08	5.74	1.09	5.10
Uruguay	1.02	0.56	31.75	0.92		
USA	0.10	0.01	70.51	4.82	1.58	13.04
Venezuela	2.65	4.22	85.85	7.39	1.10	5.79

Source: own elaboration based on Hofstede, Schwartz and Inglehart national values

Figure 29. Pairwise Correlation of National Culture Variables



Source: own elaboration based on Hofstede, Schwartz and Inglehart national values

Table 43. Summary of descriptive statistics for numeric variables

	Min	Max	Range	Mean	Q1	Median	Q3	IQR	Mode	Variance	SD
Age	0	98	98	41.58	28	42	55	27	50	278.80	16.70
Stages	12.56	37	33	12.56	5	7	14	9	5	106.58	10.32
Distance	100	1073.3	973.3	294.46	114.1	151.8	312.2	198.1	114.1	66774.89	258.41A

Source: own elaboration based on Pilgrim's Welcome Office dataset

Table 44. Frequency Distribution and Pilgrim Count of Age in 4 Ranges

	Pilgrims	Frequency
Less than 18	352,110	8.65%
Between 18 and 45	1,923,349	47.25%
Between 46 and 65	1,495,267	36.73%
More than 65	299,986	7.37 %

Source: own elaboration based on Pilgrim's Welcome Office dataset

Table 45. Frequency Distribution and Pilgrim Count of the Number of Stages

	Pilgrims	Frequency
Less than 6	1,606,503	43.40%
Between 6 and 13	1,092,808	29.53%
More than 13	1,001,947	27.07%

Source: own elaboration based on Pilgrim's Welcome Office dataset

Table 46. Frequency Distribution and Pilgrim Count of Professions

	Pilgrims	Frequency
Artists	30,425	.86%
Athletes	3,591	.10%
Civil servants	185,931	5.28%
Employees	783,651	22.24%
Students	666,024	18.91%
Farmers	12,028	.34%
Homemakers	88,926	2.52%
Laborers	106,130	3.01%
Liberals	431,245	12.24%
Managers	46,482	1.32%
Member of religious order	10,069	.29%
Priests	18,693	.53%
Retirees	424,935	12.06%
Seafarers	4,680	.13%
Teachers	250,884	7.12%
Technicians	395,202	11.22%
Unemployed	64,024	1.82%

Source: own elaboration based on Pilgrim's Welcome Office dataset

Table 47. Frequency Distribution and Pilgrim Count of Professional Group

	Pilgrims	Frequency
Group 1	970,574	27.55%
Group 2	1,254,482	35.61%
Group 3	1,297,864	36.84%

Source: own elaboration based on Pilgrim's Welcome Office dataset

Table 48. Frequency Distribution and Pilgrim Count of Autonomies

	Pilgrims	Frequency
Andalusia	374,828	17.77%
Aragon	45,106	2.14%
Asturias	42,059	1.99%
Balearic Islands	31,081	1.47%
Basque Country	92,114	4.37%
Canary Islands	56,534	2.68%
Cantabria	18,421	.87%
Castile and Leon	127,142	6.03%
Castilla-La Mancha	106,055	5.03%
Catalonia	237,822	11.28%
Ceuta	3,036	.14%
Community of Madrid	391,594	18.57%
Extremadura	58,017	2.75%
Galicia	191,844	9.10%
La Rioja	10,971	.52%
Melilla	1,795	.09%
Navarre	23,182	1.10%
Region of Murcia	57,656	2.73%
Valencian Community	239,886	11.37%

Source: own elaboration based on Pilgrim's Welcome Office dataset

Table 49. Frequency Distribution and Pilgrim Count of Years

	Pilgrims	Frequency
2003	74,022	1.81%
2004	179,152	4.40%
2005	93,516	2.30%
2006	100,604	2.47%
2007	113,729	2.79%
2008	124,944	3.07%
2009	146,482	3.60%
2010	268,966	6.61%
2011	182,122	4.48%
2012	190,809	4.69%
2013	214,009	5.26%
2014	235,402	5.78%
2015	259,804	6.38%
2016	274,927	6.75%
2017	297,636	7.31%
2018	323,000	7.93%
2019	341,386	8.39%
2020	42,817	1.30%
2021	173,943	4.27%
2022	423,442	10.10%

Source: own elaboration based on Pilgrim's Welcome Office dataset

Table 50. Frequency Distribution and Pilgrim Count of Month

	Pilgrims	Frequency
January	17,402	.43%
February	24,697	.61%
March	87,756	2.16%
April	277,348	6.81%
May	444,857	10.93%
June	528,401	12.98%
July	679,829	16.70%
August	879,871	21.61%
September	585,407	14.38%
October	408,715	10.04%
November	95,086	2.34%
December	41,343	1.02%

Source: own elaboration based on Pilgrim's Welcome Office dataset

Table 51. Frequency Distribution and Pilgrim Count of Motivations

	Pilgrims	Frequency
Religious	1,749,429	10.18%
Religious and others	1,906,868	42.98%
Not religious	414,415	46.84%

Source: own elaboration based on Pilgrim's Welcome Office dataset

Table 52. Frequency Distribution and Pilgrim Count of Ways

	Pilgrims	Frequency
English	131,203	3.22%
French	2,687,943	66.03%
North	232,068	5.70%
Portuguese	627,856	15.42%
Portuguese Coastal	92,873	2.28%
Primitive	156,225	3.84%
Silver	142,544	3.50%

Source: own elaboration based on Pilgrim's Welcome Office dataset

Table 53. Frequency Distribution and Pilgrim Count of Continents

	Pilgrims	Frequency
Africa	11224	.28%
Asia	83913	2.06%
Europe	3540723	86.98%
North America	258131	6.34%
Oceania	50445	1.24%
South America	126276	3.10%

Source: own elaboration based on Pilgrim's Welcome Office dataset

Table 54. Frequency Distribution and Pilgrim Count of Inglehart clusters

	Pilgrims	Frequency
Catholic Europe	805,954	19.80%
Confucian	74,065	1.82%
Destination Country	2,127,750	52.27%
English-Speaking	409,398	10.06%
Latin America	158,188	3.89%
Orthodox Europe	26,784	.66%
Protestant Europe	442,687	10.87%
West and South Asia	25,886	.64%

Source: own elaboration based on Pilgrim's Welcome Office dataset

Table 55. Frequency Distribution and Pilgrim Count of Sex

	Pilgrims	Frequency
Men	2,167,216	53.24%
Women	1,903,495	46.76%

Source: own elaboration based on Pilgrim's Welcome Office dataset

Table 56. Frequency Distribution and Pilgrim Count of Mean of transport

	Pilgrims	Frequency
Walking	3,635,669	89.31%
Not walking	435,043	10.69%

Source: own elaboration based on Pilgrim's Welcome Office dataset

Table 57. Difficulty and distance of stages by Ways (I)

Stages to Santiago	Coastal		Portuguese		French		English	
	Distance (km)	Difficulty	Distance (km)	Difficulty	Distance (km)	Difficulty	Distance (km)	Difficulty
1	24.3	2	24.3	2	19.4	2	15.7	1
2	18.6	2	18.6	2	19.3	2	24.4	2
3	21.1	1	21.1	1	28.5	3	24	3
4	19.6	2	19.6	2	24.8	2	19.7	2
5	16	1	31.6	3	22.2	2	27.7	3
6	22.6	2	19.1	1	25	3		
7	16	1	17.4	3	20.8	2		
8	23.5	2	34.5	3	27.8	4		
9	26.8	3	27.3	2	24.2	2		
10	20.8	2	26.5	2	26.8	3		
11	24.5	2	15.1	1	25.8	2		
12	14	1	19	1	23.7	2		
13	24.5	2	28.8	3	24.6	2		
14			15.8	1	18.5	1		
15			23.5	2	26.3	2		
16			24.8	2	23.2	2		
17			11.3	1	26.3	2		
18			26.1	2	18.8	1		
19			22.7	2	24.7	2		
20			31.7	3	19.9	2		
21			31.3	3	21	2		
22			30.7	3	25.8	2		
23			33	3	23.9	2		
24			20	1	22	1		
25			18.8	1	20.7	1		
26			21.7	2	29	2		
27					27.6	2		
28					21.3	2		
29					21.6	2		
30					23.9	2		
31					20.4	1		
32					21.4	3		
33					24.2	5		

Source: own elaboration based on Gronze and IGN

Table 58. Difficulty and distance of stages by Ways (II)

Stages to Santiago	Primitive		North		Silver	
	Distance (km)	Difficulty	Distance (km)	Difficulty	Distance (km)	Difficulty
1	19.4	2	19.4	2	16.7	1
2	19.3	2	19.3	2	33.1	3
3	20	2	22	2	32.8	4
4	26.5	2	39.8	4	21.2	3
5	29.5	3	18.5	1	21.8	1
6	24.3	3	20.8	2	33	4
7	25.2	3	16.6	2	34	4
8	20.4	3	21.3	1	23.7	3
9	24.1	4	30.5	3	29.6	4
10	15.9	2	15.3	2	31.7	3
11	19.8	3	18.5	3	36.1	4
12	22.1	2	15.3	2	22.9	2
13	25.2	3	23.2	2	25.1	2
14			25.4	2	22.6	2
15			29.8	4	19.2	1
16			17.2	2	31.6	3
17			20.1	2	36.4	4
18			30	3	23.6	2
19			22.9	2	27.9	3
20			28	3	20.2	2
21			22.2	2	21.8	3
22			36.3	3	38.3	4
23			15.3	1	30.2	3
24			28.7	3	20	2
25			25.7	3	33.8	3
26			25.4	3	37.4	4
27			19.2	2	35.9	4
28			10.8	2	15.3	1
29			20.8	3	26.7	2
30			24.6	3	19.8	1
31			24	4	24.2	2
32			21.8	3	20.7	2
33			22.2	3	33.6	4
34			24.8	3	28.2	3
35			15.4	2	18.3	2
36			29	1	21.6	2

Source: own elaboration based on Gronze and IGN

Table 59. Mean intensity by stage

Way	Mean difficulty	Precipitation	Temperature	Intensity	Intense
Coastal 1	1.57	94.68	14.60	0.49	1
Coastal 2	1.81	94.68	14.60	0.58	1
Coastal 3	1.77	97.94	14.64	0.63	1
Coastal 4	1.77	97.94	14.64	0.68	1
French 1	2.33	81.96	13.59	0.61	0
French 2	2.50	81.96	13.59	0.71	0
French 3	2.25	51.90	12.45	0.70	0
French 4	2.12	49.04	12.73	0.72	0
English 1	2.20	101.23	14.12	0.62	0
North 1	2.17	81.96	13.59	0.59	1
North 2	2.10	81.96	13.59	0.64	1
North 3	2.24	83.02	13.68	0.73	1
North 4	2.42	92.02	14.05	0.85	1
Silver 1	2.40	71.27	15.01	0.54	0
Silver 2	2.88	71.27	15.01	0.70	0
Silver 3	2.86	47.72	13.81	0.77	0
Silver 4	2.69	36.15	15.40	0.73	0
Portuguese 1	1.83	94.68	14.60	0.53	0
Portuguese 2	2.11	94.68	14.60	0.63	0
Portuguese 3	1.88	97.94	14.64	0.65	0
Portuguese 4	2.00	89.60	15.40	0.69	0
Primitive 1	2.33	81.96	13.59	0.61	0
Primitive 2	2.60	81.96	13.59	0.73	0
Primitive 3	2.62	83.02	13.68	0.81	0

Note: To check how to measure the intensity and intense variables, see ANNEX 1.The suffix for each way indicates if the range of stages are shorter than 140 km, between 141 km and 235 km, between 236 km and 400 km and more than 400 km

Source: own elaboration based on Gronze, IGN and AEMET

ANNEX 5 QUESTIONNAIRE

Online questionnaire: Individual and group Pilgrim registration

1. **Way**. Select a Way from Table 60.
2. **Conveyance**. Select a conveyance from Table 61.
3. **Reasons for pilgrimage**. Select a reason for pilgrimage from Table 64.
4. **Sex**. Select a Sex from Table 62.
5. **Profession**. Select a profession from Table 63.
6. **Start of the way**. Select a Start of the Way from Table 65.
7. **Nationality**. Select a nationality from Table 66.
8. **Autonomy**. If nationality is Spanish, select an autonomy form Table 67.
9. **Name_______**
10. **Surname________**
11. **Age_______**
12. **Start date________**
13. **Arrival date** (approximate)**_____**
14. **Group name________**

Table 60. Way
Camino de Geira-Arrieros
Camino de Invierno
Camino de Barbanza
Camino del Mar
Camino del Norte
Camino de Muros-Noia
Camino de San Rosendo
Camino de San Salvador
Camino Francés
Camino Inglés
Camino Olvidado
Camino Portugués
Camino Portugués por la Costa
Camino Primitivo
Miñoto Ribeiro
Muxía-Finisterre
Otros Caminos
Vía Céltica
Vía de la Plata

Table 61. Conveyance
Bicycle
By Horse
Sailing
Walking
Wheelchair

Table 62. Sex
Women
Man
Others

Table 63. Profession
Artists
Athletes
Employees
Farmers
Housewives
Managers
Officials
Priests
Professionals
Religious
Retirees
Sailors
Students
Teachers
Technicians

Table 64. Reasons for pilgrimage
Not religious
Others
Religious

Table 65. Start of the Way

2115	Castilla la Mancha otros
2116	Reino Unido C.F.
2117	Castilla La Mancha VP
2118	Vega de Valcarce
2119	Vezelay
2120	Arles
2121	Republica Checa
2122	Murcia
2123	Valencia O.C.
2124	Artieda
2125	Astorga
2126	Austria
2127	Avilés
2128	Badajoz
2129	Barcelona
2130	Bélgica
2131	Benavente
2133	Bilbao
2134	Borres
2135	Braga
2136	Burgos
2137	Cáceres
2139	Canfranc
2141	Carrión de los Condes
2142	Cast. la Mancha - C.F.
2143	Castrojeriz
2144	Cataluña - C.F.
2146	Cebreiro
2147	Com. Valenciana - C.F.
2148	Córdoba
2150	Covelo
2151	Dinamarca
2152	El Escamplero
2153	Estella
2154	Ferrol
2155	Finisterra
2156	Cataluña - O.C.
2157	Finlandia
2160	Fonsagrada - C.P.
2161	Francia - C.F.
2162	Frómista
2163	Gándara
2164	Gijón
2166	Grado
2167	Granada
2169	Grandas de Salime - C.P.
2170	Grecia
2173	Hendaya
2174	Holanda
2175	Hospital de Orbigo
2176	Huelva
2177	Inglaterra C.F.
2178	Irlanda C.F.
2179	Irún
2180	Italia
2181	Jaca
2182	La Bañeza
2183	La Mesa
2184	Lalín
2185	Laza
2186	Le Puy
2187	León
2188	Leyre
2189	Lisboa
2190	Logroño
2191	Lourdes
2193	Lugo - C.P.
2194	Luxemburgo
2195	Madrid - V.P.
2196	Madrid - C.F.
2198	Mérida
2199	Mondoñedo
2200	Muxia
2201	Nájera
2202	Neda
2203	Oporto
2205	Ortigueira
2206	Ourense
2207	Oviedo - C.N.
2208	Oviedo - C.P.
2210	Pamplona
2213	París
2214	Peñaseita
2216	Ponferrada
2219	Porriño
2220	Puebla de Sanabria
2221	Puente la Reina
2222	Rábade
2223	Resto Asturias - C.N
2224	Resto Asturias - C.P.
2225	Resto C. León C.F.
2226	Resto C. León - V.P.
2227	Resto Cantabria
2228	Resto de Extremadura
2229	Resto País Vasco - C.N.
2230	Resto Portugal
2231	Ribadeo
2232	Roma
2233	Roncesvalles
2234	S. Jean P. Port
2235	Sahagún
2236	Salamanca

2237	Salas
2238	Samos
2239	San Juan de Ortega
2240	Santander
2241	Sarria
2242	Sevilla
2246	Somport
2247	Sto. Domingo de la Calzada
2248	Suiza
2250	Tineo - C.P.
2251	Triacastela
2252	Tui
2253	Valcarlos
2254	Valença do Minho
2255	Valladolid
2256	Vegadeo
2258	Verín
2259	Vigo
2260	Vilabade
2261	Vilafranca
2262	Vilalba
2263	Vincios
2264	Viveiro
2265	Zamora
2266	Granja de Moreruela
2267	Malaga
2268	Rabanal del Camino
2269	Ponte de Lima
2270	Xunqueira de Ambia
2271	Resto Andalucia
2272	Lourenzá
2273	Resto Europa
2274	Hungría
2275	Jerusalem
2276	Zaragoza
2277	Polonia
2278	Cadavo
2279	R.Pais Vasco C.F.
2280	Gudiña
2281	Alemania
2283	A Guarda
2284	Com. Valenciana - O.C.
2285	Chaves-Portugal
2286	Francia - C.N
2287	Andorra
2288	Egipto
2289	Allariz
2290	Irlanda C. Ing
2291	Reino Unido C.Ing

2292	Abadin
2293	Rusia
2294	Resto Africa
2295	San Sebastián
2296	Montserrat
2297	Resto Galicia
2298	Baamonde
2299	Monforte de Lemos
2300	Chantada
2301	A Rúa
2302	Las Médulas
2303	O Barco de Valdeorras
2304	Quiroga
2306	Baiona
2308	Navarra
2309	La Rioja
2311	Molinaseca
2313	Ponferrada. C.Inv.
2314	Povoa de Varzim
2315	Barcelos
2316	Fatima
2317	Faro
2318	Aveiro
2319	Coimbra
2320	Rates
2321	Viana do Castelo
2322	Esposende
2323	Guimaraes
2324	Viseu
2325	Eslovaquia
2326	Cruz de Ferro
2327	Puente de Domingo Flórez
2328	Jaén
2329	Aragón
2330	Resto de Galicia
2331	Bayonne
2332	R. Castilla León C.Inv
2333	Australia
2334	Caminha
2335	Oporto Costa
2336	Vila do Conde
2337	A Coruña
2338	Lires
2339	Castroverde
2340	Resto Europa C. Ing.
2341	Fonfría
2342	Biarritz
2343	Oia
2344	Resto C. León
2345	Resto de Portugal

2346	Lisboa Costa
2347	Resto de Portugal
2348	Barbadelo
2349	Coimbra Costa
2350	Xinzo de Limia
2351	Ferreiros
2352	Luarca
2353	Resto de Europa
2354	Suecia
2355	Murcia V.P.
2356	Noruega
2357	Berán C.G.A.
2358	Ribadavia C.G.A.
2359	Cortegada C.G.A.
2360	Lobios C.G.A.
2361	Entrimo C.G.A.
2362	Gerês C.G.A.
2363	T. Bouro C.G.A.
2364	Braga C.G.A.
2365	Castro Leboreiro C.G.A.
2366	Laguna de Castilla
2367	C.G.A resto de Portugal
2368	Bulgaria
2369	Madrid
2370	Resto Galicia
2371	almeria
2372	Alemania. O.C.
2389	RESTO DE PORTUGAL C.G.A
2421	Muros
2422	Porto do Son
2423	Corrubedo
2424	Ribadeo C. del Mar
2425	Barreiros
2427	San Ciprián (Cervo)
2428	Viveiro
2429	Ponte do Porto (Mañón)
2430	Mera (Ortigueira)
2431	San Andrés de Teixido
2432	Porto do Cabo (Cedeira)
2433	Braga C.G.R
2434	Mosteiro Rendufe C.G.A
2435	Santa Cruz C.G.A
2436	Terras de Bouro C.G.A
2437	Campo do Gerês C.G.A
2438	Portela do Homem C.G.A
2439	Lobios C.G.A
2440	Entrimo C.G.A
2441	Ameixoeira C.G.A

2442	Castro Leboreiro C.G.A
2443	Azoreira C.G.A
2444	San Amaro C.G.A
2445	Cortegada C.G.A
2446	Río Arnoia C.G.A
2447	Ribadavia C.G.A
2448	Berán C.G.A
2450	Laxe
2451	Ponteceso
2452	Corme
2453	Malpica
2454	Caión
2455	Barizo
2456	Fisterra
2457	Muxía
2458	Cereixo
2459	Ponte do Porto
2460	Camariñas
2461	Braga
2462	Braga
2463	Santa Cruz
2464	Gerês
2465	Lobios
2466	Bande
2467	Celanova
2468	Fazouro
2469	Vilaverde.

Table 66. Nationality

1	Belice
2	Islas Feroe
3	Albania
4	Alemania
5	Andorra
6	Angola
7	Antigua y Barbuda
8	Antillas Holandesas
9	Arabia Saudita
10	Argelia
11	Argentina
12	Armenia
13	Aruba
14	Australia
15	Austria
16	Azerbaiyán
17	Bahamas
18	Bahrein
19	Bangladesh
20	Barbados
21	Bélgica
22	Belice
23	Benin
24	Bermudas
25	Bhután
26	Bielorrusia
27	Bolivia
28	Bosnia
29	Botswana
30	Brasil
31	Brunei
32	Bulgaria
33	Burkina Faso
34	Burundi
35	Cabo Verde
36	Camboya
37	Camerún
38	Canadá
39	Colombia
40	Comores
41	Congo
42	Costa de Marfil
43	Costa Rica
44	Croacia
45	Cuba
46	Chad
47	Chile
48	China
49	Chipre
50	Dinamarca
51	Djibouti
52	Dominica
53	Ecuador
54	Egipto
55	El Salvador
56	Emiratos Arabes Unidos
57	Eritrea
58	Eslovaquia
59	Eslovenia
60	España
61	Estados Unidos
62	Estonia
63	Etiopía
64	Fiji
65	Filipinas
66	Finlandia
67	Francia
68	Gabón
69	Gambia
70	Georgia
71	Ghana
72	Granada
73	Grecia
74	Groenlandia
75	Guadalupe
76	Guam
77	Guatemala
78	Guayana Francesa
79	Guinea
80	Guinea - Bissau
81	Guinea Ecuatorial
82	Guyana
83	Haití
84	Herzegovina
85	Holanda
86	Honduras
87	Hungría
88	I. Reunión
89	India
90	Indonesia
91	Irak
92	Irán
93	Irlanda
94	Islandia
95	Islas Caimán
96	Islas Malvinas
97	Islas Marshall
98	Islas Salomón
99	Islas Vírgenes
100	Israel
101	Italia
102	Jamaica
103	Japón
104	Jordania

105	Kazajistán
106	Kenya
107	Kirguistán
108	Kiribati
109	Kuwait
110	Laos
111	Lesotho
112	Letonia
113	Líbano
114	Liberia
115	Libia
116	Liechtensein
117	Lituania
118	Luxemburgo
119	Macedonia
120	Madagascar
121	Malasia
122	Malawi
123	Maldivas
124	Malta
125	Marruecos
126	Martinica
127	Mauricio
128	Mauritania
129	México
130	Micronesia
131	Moldavia
132	Mónaco
133	Mongolia
134	Mozambique
135	Myanmar
136	Namibia
137	Naurú
138	Nepal
139	Nicaragua
140	Níger
141	Nigeria
142	Noruega
143	Nueva Zelanda
144	Omán
145	Pakistán
146	Palau
147	Panamá
148	Papúa - Nueva Guinea
149	Paraguay
150	Perú
151	Polonia
152	Portugal
153	Puerto Rico
154	Qatar

155	Reino Unido
156	Rep. Centroafricana
158	Rep. Dominicana
160	República Checa
161	Rumania
162	Rusia
163	Rwanda
164	Samoa Occidental
165	San Cristóbal y Nevis
166	San Marino
167	San Pedro y Miquelón
168	San Vicente y Las Granadinas
169	Santa Lucía
170	Santo Tomé y Príncipe
171	Senegal
172	Serbia
173	Seychelles
174	Sierra Leona
175	Singapur
176	Siria
177	Somalia
178	Sri Lanka
179	Sudáfrica
180	Sudán
181	Suecia
182	Suiza
183	Surinam
184	Swazilandia
185	Tailandia
186	Taiwán
187	Tanzania
188	Tayikistán
189	Togo
190	Tonga
191	Trinidad y Tobago
192	Túnez
193	Turkmenistán
194	Turquía
195	Tuvalu
196	Ucrania
197	Uganda
198	Urbekistán
199	Uruguay
200	Vanuatu
201	Vaticano
202	Venezuela
203	Vietnam
204	Yemen
206	Zaire
207	Zambia

208	Zimbabwe
209	Corea
210	Timor Oriental
211	Afganistán
212	Mali
213	Montenegro
214	Palestina
215	Brunei
216	Nueva Caledonia
217	Kosovo
218	Hong Kong
219	Jersey

Table 67. Autonomy

0 Extranjero
1 Andalucía
2 Aragón
3 Asturias
4 Baleares
5 Canarias
6 Cantabria
7 Castilla La Mancha
8 Castilla León
9 Cataluña
10 Ceuta
11 Extremadura
12 Galicia
13 La Rioja
14 Madrid
15 Melilla
16 Murcia
17 Navarra
19 País Vasco
20 Comunidad Valenciana
21 Residentes Extranjero

Index of tables

Index of figures

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Decision-making models are widely used tools developed to study how agents select their preferred features on conflictive alternatives. Despite the importance given to individual preferences in decision-making, cognitive biases, emotions or psychological influences were not studied in the realm of economics till the 70s, with rise of behavioural economics. However, contextual factors like culture from the field of social psychology were less studied in the area of economics. To explore the influence of culture on decision-making, we use the case of tourism, where the primary hypothesis is that culture affects decision-making. Thus, the aim of the thesis is to merge the knowledge from decision-making, behavioural economics and social-psychology in a tourism context to check if culture affect decision-making, and how.