



Facultad de ciencias Económicas Empresariales

Trabajo de
fin de grado

Toma de decisiones en el
consumo cinematográfico.

Jorge Pazó Zurrón

Junio 2024

Resumen

El propósito de este trabajo es analizar cómo influyen las características personales de los individuos, y de los bienes a elegir, en la toma de decisiones. Para lograr este objetivo se ha desarrollado un ejemplo empírico sobre el consumo cinematográfico, donde el eje de la toma de decisiones será la satisfacción aportada por la película consumida. La investigación realizada partirá de una metodología analítica recurriendo a modelos tanto econométricos, como de machine learning, aplicados a un data set de cinco variables objeto de análisis con un millón de observaciones. Los modelos estudiados permitirán entender las relaciones que explican cómo se asignan las satisfacciones, y, por lo tanto, como se puede ver afectado el patrón de consumo general. Finalmente, tras comparar y discutir los hallazgos resultantes de la aplicación de los modelos, el trabajo concluye con las principales reflexiones extraídas y las líneas de investigación futuras. Indicar, que el conjunto de este trabajo está compuesto por 9.859 palabras.

Índice

Resumen	2
Índice de abreviaturas	5
Índice de tablas, gráficos o figuras	6
Introducción.....	7
1 Marco teórico.....	8
1.1 Toma de decisiones.....	8
1.2 Objetivos.	9
2 Datos y metodología.....	9
2.1 Obtención y Preparación de datos.....	10
2.1.1 Dataset: Estructura y Variables	11
2.1.2 Variables de los usuarios	11
2.1.3 Variables de las películas.....	12
2.1.4 Variable dependiente de los usuarios y las películas	13
2.2 Métodos estadísticos	13
2.3 Modelización	13
2.3.1 Modelos econométricos.....	14
2.3.2 Machine learning.....	16
3 Análisis Empírico	18
3.1 Estadísticos descriptivos	19
3.2 Tablas cruzadas	22
3.3 Modelos empíricos.....	25
3.3.1 Regresión lineal	25
3.3.2 Logit binomial	29
3.3.3 Decision tree.....	35
3.4 Comparación de Modelos y Discusión general.	37
Conclusiones y ampliación.....	40
Bibliografía.....	43

Índice de abreviaturas

Término completo	Abreviatura
Decision making	DM
Standard Process for Data Mining	CRISP-DM
Método de máxima verosimilitud	MLE
Hipótesis 1	H1
Hipótesis 2	H2
Hipótesis 3	H3
Hipótesis 4	H4

Índice de tablas, gráficos o figuras

FIGURA 1, METODOLOGÍA PROPUESTA	10
FIGURA 2, CONJUNTO DE VARIABLES	11
FIGURA 3, SEXO	19
FIGURA 4, EDAD	19
FIGURA 5, OCUPACIÓN.....	20
FIGURA 6, GÉNERO CINEMATográfico	21
FIGURA 7, PUNTUACIÓN.....	21
FIGURA 8, ÁRBOL DE CLASIFICACIÓN	35

TABLA 1, TABLA CRUZADA: PUNTUACIÓN-SEXO.....	22
TABLA 2, TABLA CRUZADA: PUNTUACIÓN-EDAD	22
TABLA 3, TABLA CRUZADA: PUNTUACIÓN-EDAD	23
TABLA 4, TABLA CRUZADA: PUNTUACIÓN-GÉNERO CINEMATográfico	24
TABLA 5, MODELO DE REGRESIÓN LINEAL.....	26
TABLA 6, PRUEBA DE BREUSH-PAGAN	29
TABLA 7, MODELO LOGIT BINOMIAL.....	31
TABLA 8, EFECTOS MARGINALES	33

Introducción

La toma de decisiones por parte de agentes económicos racionales constituye uno de los temas clave en el estudio de la economía y particularmente es la base de la microeconomía. Desde los modelos de elección discreta, que se apoyan en los Random Utility Models, los cuales permiten cuantificar la utilidad derivada de cada alternativa, tomando en cuenta factores no observables y transformándolos en probabilidades (Briceño, 2000), hasta el innovador machine learning, (usado principalmente para predecir la probabilidad de conseguir un determinado resultado) se pretende analizar las utilidades y satisfacciones de los agentes económicos, todo para entender y prever su toma de decisiones.

Expuesto ya el marco de referencia sobre el que se trabaja, se plantea en esta investigación la necesidad de estudiar a nivel empírico, a partir de un experimento con datos reales, la toma de decisiones de los agentes económicos a través de modelos econométricos. Incluyendo además una introducción a los árboles de decisión.

Para desarrollar este experimento y aplicar la teoría a un contexto real, se utilizará un ejemplo de consumo cinematográfico. Dicho experimento tomará como fuente la base de datos proporcionada por Movielens en 2003, la cual contiene aproximadamente un millón de calificaciones de películas, así como datos demográficos de los usuarios que las evaluaron.

El trabajo se estructurará en varias partes diferenciadas, comenzando con el marco teórico, donde se explorará el concepto de toma de decisiones. A continuación, se presentará la metodología, que incluye la descripción de la base de datos utilizada y la definición de los tres modelos analíticos. Posteriormente, se analizarán los resultados obtenidos a partir de estos modelos, y finalmente, se presentarán las conclusiones del estudio.

1 Marco teórico

1.1 Toma de decisiones

En el corazón de la economía se encuentra la toma de decisiones, un proceso fundamental que determina cómo los recursos escasos son asignados entre múltiples usos alternativos. Asimismo, la teoría de la elección racional, profundamente arraigada en el análisis microeconómico, proporciona un marco para entender y predecir el comportamiento de individuos y organizaciones en un mundo de recursos limitados. De este modo, como “homo economicus” se presupone que obramos bajo el paraguas de la “elección racional”, es decir, como tomadores de decisiones tenemos una función de utilidad completa y consistente, trataremos de conocer todas las alternativas posibles y se puede calcular el valor esperado de la utilidad asociada a cada alternativa, todo ello para escoger aquella opción que maximiza la utilidad esperada (Griskevicius y Kenrick, 2013).

Así, la teoría de elección racional se basa en el supuesto de que el individuo buscará una combinación de decisiones que le otorgue la mayor satisfacción posible. Dicha toma de decisiones dependerá de las preferencias de los consumidores, siendo estas diferenciadas para cada uno de ellos. De este modo, una misma opción aportará un distinto grado de satisfacción a cada individuo dependiendo de la utilidad que éste le asigne a dicha opción. Consecuentemente, un consumidor al elegir una decisión sobre otra revelará sus preferencias y el nivel de utilidad que le reporta. Entender, por lo tanto, la racionalidad de los agentes permitirá estimar sus posibles decisiones futuras (Varían, 2010).

Para trabajar bajo este nuevo marco conceptual de decision making, recurrimos a la metodología microeconómica de Daniel McFadden, que marcó un cambio revolucionario en este campo al transformar las utilidades en probabilidades, ofreciendo un marco matemático para representar cómo los agentes económicos toman decisiones cuando se les presentan múltiples alternativas (Atrio, 2024). De igual manera, el avance de las tecnologías de la información y el desarrollo de técnicas de machine learning permiten a los economistas y analistas manejar grandes volúmenes de datos y detectar patrones complejos que no son evidentes mediante métodos tradicionales.

En resumen, la teoría de la elección racional y la toma de decisiones son pilares fundamentales para comprender el comportamiento en los mercados, mientras que, de manera práctica, la combinación de enfoques econométricos con innovaciones modernas en machine learning, proporciona un marco robusto para analizar y prever decisiones.

1.2 Objetivos.

La revisión del marco teórico ha motivado el principal objetivo de la investigación: “analizar como las características personales y de los bienes que se escogen pueden afectar a la satisfacción esperada de los individuos, y, en consecuencia, a sus decisiones”. Para lograr este objetivo, se plantean los siguientes subobjetivos:

Subobjetivo 1: Analizar las metodologías de regresión lineal, regresión logística y árbol de decisión como modelos de elección racional para poder evaluar los resultados obtenidos y así explicar la asignación de utilidades y toma de decisiones en el ejemplo empírico.

Subobjetivo 2: Comparar los resultados de las metodologías (modelos econométricos y machine learning) que se utilizarán para analizar la toma de decisiones. Este análisis comparativo permitirá demostrar empíricamente las características, ventajas y limitaciones de cada enfoque.

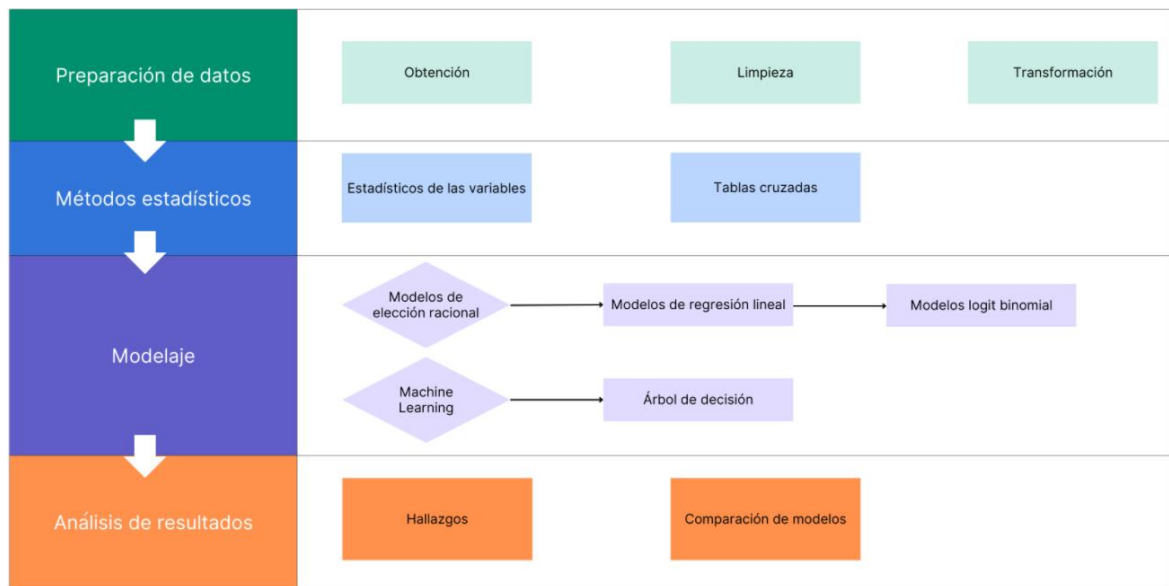
Subobjetivo 3: Finalmente, al aplicar estos métodos al conjunto de datos, también se tiene por objetivo desarrollar una herramienta práctica y útil. Esta herramienta podría ser utilizada dentro de una empresa para mejorar su toma de decisiones, optimizar estrategias comerciales y entender mejor el comportamiento del mercado.

2 Datos y metodología

Para extraer información que se pueda convertir en conocimiento, se emplea una metodología similar a la usada en minería de datos, denominada Cross Industry Standard Process for Data Mining (CRISP-DM), con un enfoque menos empresarial donde la implementación y retroalimentación no tienen relevancia. (IBM, 2015)

Con los objetivos de investigación definidos, en la etapa de desarrollo nos enfocaremos en el análisis de la toma de decisiones (DM) mediante un ejemplo práctico relacionado con el consumo de películas. Antes de avanzar a dicha etapa, se procederá a describir exhaustivamente el conjunto de datos utilizado en este estudio, incluyendo las variables y características presentes en la muestra. Posteriormente, se describirá la metodología de modelización que se aplica para obtener resultados significativos en el capítulo de desarrollo.

Figura 1, Metodología propuesta



Fuente: Elaboración propia

2.1 Obtención y Preparación de datos

La base de datos utilizada para desarrollar la metodología propuesta proviene del laboratorio de investigación del Departamento de Ingeniería y Ciencias de la Computación de la Universidad de Minnesota, conocido como GroupLens. Este conjunto de datos incluye 1.000.209 calificaciones anónimas de aproximadamente 3.900 películas, realizadas por 6.040 usuarios de MovieLens que se unieron a la plataforma en el año 2000.

Siguiendo el modelo CRISP-DM, en este estudio, la tarea de preparación de la base de datos se ha efectuado en el programa R-Studio versión 4.2, permitiendo la integración de la base de datos que incluye las características de los usuarios, con aquella que recoge las valoraciones asignadas a cada película por dichos usuarios. La dificultad de esta tarea radica en la conexión de los atributos de los usuarios con las valoraciones que dichos usuarios efectúan sobre las películas. Esta tarea constituye una de las labores más costosas en los modelos de minería de datos, asignándole en la planificación de este trabajo los primeros tres meses de dedicación. Asimismo, para definir los géneros de las películas, en los casos donde una película tiene múltiples géneros, se seleccionó el género principal, eliminado también la variable que indicaba el nombre de la película con el objetivo de construir un data-set organizado y ordenado.

2.1.1 Dataset: Estructura y Variables

Antes de proceder al apartado de modelización, es necesario describir las variables presentes en el conjunto de datos que se utilizarán para extraer conclusiones. Como se mencionó anteriormente, la base de datos objeto de estudio contiene aproximadamente un millón de observaciones, que corresponden a todas las valoraciones realizadas por los 6.040 usuarios. Además de las calificaciones asignadas en cada valoración, el conjunto de datos incluye características demográficas y ocupacionales de los usuarios que evaluaron las películas, así como el género cinematográfico de las mismas. Estas variables se consideran fundamentales para el posterior análisis de la toma de decisiones recogido en el marco teórico de la elección racional.

Figura 2, Conjunto de variables

Nombre	Descripción
Género cinematográfico de la película evaluada	"Acción", "Aventura", "Animación", "Infantil", "Comedia", "Crimen", "Documental", "Drama", "Fantasía", "Cine Negro", "Terror", "Musical", "Misterio", "Romance", "Ciencia Ficción", "Suspense", "Bélica", "Western"
Puntuación de la película	Valores del 1 al 5 (n)
Edad	"Por debajo de los 18", "18-24", "25-34", "35-44", "45-49", "50-55", "56+"
Sexo	Femenino ("F"), Masculino ("M")
Ocupación	"otro" o no especificado, "académico/educador", "artista", "administrativo", "estudiante universitario/postgrado", "atención al cliente", "médico/cuidado de la salud", "ejecutivo/gerencial", "agricultor", "ama de casa", "estudiante de primaria/secundaria", "abogado", "programador", "jubilado", "ventas/marketing", "científico", "autónomo", "técnico/ingeniero", "artesano", "desempleado", "escritor"

Fuente: Elaboración propia en base a los datos aportados por GroupLens

2.1.2 Variables de los usuarios

En este apartado se analizarán las variables que caracterizan a los tomadores de decisiones y que, teóricamente, contribuirán a definir sus elecciones.

2.1.2.1 Edad

Desde una perspectiva microeconométrica, la edad afecta las preferencias individuales debido a variaciones en gustos, experiencias culturales acumuladas y etapas de vida. Diferentes grupos etarios pueden mostrar propensiones distintivas hacia ciertos géneros cinematográficos y estilos narrativos. Además, la edad está correlacionada con factores como el ingreso disponible y el tiempo libre, que también pueden modular las decisiones de consumo en el ámbito del entretenimiento. (Mares y Sun, 2010)

2.1.2.2 Sexo

Masculino y femenino, según como el individuo se identifique. Estudios previos han demostrado que existen variaciones notables en los géneros cinematográficos preferidos por cada sexo, así como en la sensibilidad a ciertos temas y estilos narrativos. Además, estas diferencias pueden estar relacionadas con factores socioculturales y de socialización que afectan las elecciones individuales (Dolan, 2017).

2.1.2.3 Ocupación

La ocupación laboral puede influir notablemente en la toma de decisiones al escoger una película debido a las diferencias en intereses, disponibilidad de tiempo y niveles de estrés asociados con distintas profesiones. Las preferencias cinematográficas pueden variar según el entorno laboral, los ingresos y las rutinas diarias de los individuos. Por ejemplo, profesionales en campos creativos podrían mostrar una inclinación hacia películas artísticas o innovadoras, mientras que aquellos en ocupaciones de alta presión podrían preferir géneros que ofrezcan mayor relajación o distracción. (Mohamed, 2017)

2.1.3 Variables de las películas

Se procederá, a continuación, a analizar las características del objeto de decisión del agente económico, es decir, las películas a consumir.

2.1.3.1 Género de la película

Uno de los factores más importantes a estudiar será el género cinematográfico. Este aspecto será fundamental en el análisis de elección, ya que refleja claramente cómo los usuarios priorizan ciertos géneros sobre otros al tomar decisiones. El género cinematográfico, por lo tanto, se constituirá como la característica principal que diferencia los bienes (películas) entre los cuales los individuos eligen. Su análisis

permitirá observar las diferencias en las utilidades percibidas y las decisiones de consumo, basadas en las variaciones en los gustos y expectativas de los consumidores.

2.1.4 Variable dependiente de los usuarios y las películas

2.1.4.1 Puntuación

Se trata de las calificaciones que los usuarios han asignado a las películas que han visto y decidido evaluar. Volviendo a al marco teórico, gracias a estas valoraciones se obtiene una aproximación bastante precisa de la utilidad percibida por el consumidor a partir de su elección cinematográfica. Este aspecto será crucial para el análisis posterior de la elección racional del usuario, que se llevará a cabo en la modelización.

2.2 Métodos estadísticos

Basándonos en las características de la base de datos detalladas en la Figura 2, expondremos los resultados principales del análisis descriptivo, los cuales respaldarán la elección de variables en secciones posteriores.

Dado el predominio de variables cualitativas en el conjunto de datos, los métodos estadísticos convencionales relacionados con medidas de tendencia central o dispersión no son aplicables, a excepción de la variable de puntuaciones. Por consiguiente, nos enfocaremos en extraer información mediante la moda y las frecuencias relativas para las variables cualitativas. Para las puntuaciones, también se considerará el coeficiente de variación y la media.

Como complemento al análisis descriptivo de las variables, antes de proceder con la modelización, se optará por crear tablas cruzadas entre aquellas variables que parecen tener una relación inicial. Esta aproximación proporcionará una primera impresión de las relaciones entre características que afectan la toma de decisiones, las cuales serán más evidentes en los modelos subsiguientes.

2.3 Modelización

En este apartado se definirán los modelos que se tiene como herramientas de análisis y que permitirán entender la utilidad esperada y su consecuente efecto en el DM de los usuarios. Se pretende con ello analizar el comportamiento de los consumidores como conjunto y las elecciones racionales que maximizarán sus utilidades esperadas. Para explicar estos modelos, se empleará la última edición del manual "An Introduction to Statistical Learning" de Gareth James. (Gareth et al, 2023)

2.3.1 Modelos econométricos

2.3.1.1 Modelo de regresión lineal

En microeconometría, los modelos de regresión lineal múltiple se utilizan para analizar la relación entre una variable dependiente y múltiples variables independientes, la forma general de un modelo de regresión lineal múltiple es:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

Donde:

- Y es la variable dependiente que queremos explicar o predecir.
- X_i son las variables independientes que usamos para explicar Y .
- β_0 es el intercepto del modelo, que representa el valor esperado de Y cuando todas las X_i son cero.
- β_k son los coeficientes de regresión que representan el cambio esperado en Y por un cambio unitario en cada X_i , manteniendo constantes las otras variables.
- ε es el término de error, que captura la variación en Y no explicada por las X_i .

Estimación

La estimación de los coeficientes β se realiza típicamente mediante el método de mínimos cuadrados ordinarios (OLS, por sus siglas en inglés). Este método encuentra los valores de β que minimizan la suma de los cuadrados de los residuos (la diferencia entre los valores observados y los valores predichos por el modelo).

Aplicación en Decision Making

En el contexto del decision making, los modelos de regresión lineal múltiple pueden utilizarse para cuantificar la utilidad que un usuario asigna a diferentes opciones. Al incluir variables que representen características de las opciones y del usuario, se puede estimar cómo cada característica afecta la utilidad percibida, esto permitirá predecir las preferencias.

Este enfoque proporciona una herramienta adecuada para analizar y predecir comportamientos en base a datos cuantitativos y cualitativos, apoyando la toma de decisiones basadas en evidencia.

2.3.1.2 Modelo logit binomial

El modelo logit binomial es una técnica de regresión utilizada para modelar una variable de respuesta binaria (que toma dos valores posibles, como "sí" o "no", "éxito" o "fracaso"). Este modelo es especialmente útil cuando se quiere estimar la probabilidad

de ocurrencia de un evento en función de una o más variables predictoras, con el objetivo de clasificar.

La Función Logística

El modelo logit binomial se basa en la función logística que se asegura de que las probabilidades estimadas estén en el rango de 0 a 1 y se define como:

$$p(Y) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p}}$$

Donde

- $p(Y)$ es la probabilidad de que el evento de interés ocurra, dado el conjunto de predictores $X = (X_1, X_2, \dots, X_p)$.
- β_0 es el término de intersección o constante del modelo.
- $\beta_1, \beta_2, \dots, \beta_p$ son los coeficientes asociados a cada una de las variables predictoras X_1, X_2, \dots, X_p .

Para simplificar la relación entre la probabilidad y las variables predictoras, se transforma la probabilidad usando la función logit, que es el logaritmo de las probabilidades (odds).

$$\text{logit}(p(Y)) = \log\left(\frac{p(Y)}{1-p(Y)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

Estimación de Coeficientes

Los coeficientes $\beta_1, \beta_2, \dots, \beta_p$ del modelo logit se estiman utilizando el método de máxima verosimilitud (Maximum Likelihood Estimation, MLE). Este método busca los valores de los coeficientes que maximicen la probabilidad de observar los datos reales.

La intención detrás de la MLE en el contexto de la regresión logística es encontrar los valores de $\beta_1, \beta_2, \dots, \beta_p$ que hagan que las probabilidades predichas para cada individuo se acerquen lo más posible a las observaciones reales de la variable de respuesta $\hat{p}_i(Y)$ (ya sea 0 o 1). En términos matemáticos, esto implica maximizar la siguiente función de verosimilitud:

$$L(\beta) = \prod_{i=1}^n p(X_i)^{y_i} (1 - p(X_i))^{1-y_i}$$

Donde:

- n es el número de observaciones.
- y_i es la respuesta observada para la i -ésima observación.
- X_i es el vector de predictores para la i -ésima observación.

Interpretación de Coeficientes

Cada coeficiente β_i en el modelo de regresión logística múltiple representa el cambio en el logaritmo de las probabilidades (log-odds) de la respuesta en función de un cambio unitario en la variable predictora X_i manteniendo constantes las demás variables. En términos de odds ratio, un coeficiente positivo indica que un aumento en la variable predictora está asociado con un aumento en la probabilidad del evento de interés.

Este análisis permite analizar si las variables explicativas afectan de una manera positiva o negativa a la probabilidad del evento de interés, pero para poder cuantificar en cuanto afectan, se debe recurrir a los efectos marginales.

Efectos marginales

Estos efectos se asocian a una variable predictora X_i en el modelo logístico, y son el cambio en la probabilidad de que ocurra el evento de interés ($Y = 1$) debido a un pequeño cambio en X_i .

$$\text{Efecto Marginal} = \frac{\partial P(Y = 1 | X)}{\partial X_i}$$

En otras palabras, es la derivada parcial de la probabilidad $P(Y = 1 | X)$ con respecto a X_i que permite estimar cuanto una variable explicativa afecta a la probabilidad del evento de interés.

Aplicación en Decision Making

Así, el modelo logit binomial proporciona una herramienta efectiva para explicar la toma de decisiones al estimar la probabilidad de que un evento ocurra o no. Al modelar la relación entre las variables explicativas y la variable dependiente binaria, como las decisiones de los consumidores, el modelo logit binomial permite entender cómo diferentes factores afectan la probabilidad de elegir una opción sobre otra.

2.3.2 Machine learning

Dentro del contexto de la modelización, también se incorporará una metodología innovadora que depende de tecnologías emergentes. Los modelos de machine learning constituyen un subcampo de la inteligencia artificial que se centra en el desarrollo de algoritmos que permiten a las computadoras aprender y hacer predicciones a partir de datos. Estos algoritmos identifican patrones y regularidades en los datos, que pueden utilizarse para predecir resultados en nuevos conjuntos de datos. Un ejemplo de la aplicación del machine learning para prever las utilidades esperadas de los consumidores y, por ende, su posible toma de decisiones es el uso de árboles de decisión. (IBM,2024)

2.3.2.1 Árboles de decisión

En contraste con los modelos econométricos tradicionales, basados en ecuaciones matemáticas para cuantificar la relación entre variables, los árboles de decisión representan gráficamente las interacciones mediante divisiones recursivas de los datos en subconjuntos más pequeños, dependiendo de los valores de las características. La debilidad de los árboles de decisión se centra en que, a pesar de contribuir a realizar explicaciones sencillas, carecen de capacidad explicativa completa al no poder cuantificar como las características individuales influyen en dicha decisión. (Gareth James et al, 2023). En este trabajo se pondrá el enfoque específicamente en los árboles de clasificación, los cuales proporcionan una respuesta cualitativa:

El primer paso en la creación de un árbol de clasificación es la división recursiva binaria aplicada al conjunto de datos. Para realizar este método primero se selecciona el predictor X_j y el punto de corte s , de modo que, al dividir el espacio del predictor en las regiones, se conduzca a la mayor reducción posible de las impurezas (medidas con el índice de Gini). La expresión matemática será la siguiente:

$$R_1(j, s) = \{X|X_j < s\} \quad \text{y} \quad R_2(j, s) = \{X|X_j \geq s\}$$

Siendo $\{X|X_j < s\}$ la región del espacio predictor en la que X_j toma un valor menor que s (en el caso de que X_j fuese una variable cualitativa no tendríamos menores/mayores, si no, igualdades o desigualdades según las modalidades de X_j , las cuales adoptan la sigla k en vez de s). Consecuentemente se buscan los valores de k y j que minimicen la siguiente ecuación (índice de Gini):

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

Siendo \hat{p}_{mk} la proporción de observaciones de entrenamiento en la m -ésima región que son de la k -ésima clase.

Finalmente, al repetir este proceso de búsqueda de la mejor división de forma recursiva, se obtendrá el conjunto de datos segmentado en distintas regiones según las variables que aparecen en los nodos. Asimismo, se predecirá la probabilidad de la variable respuesta de cada región según la moda de los datos de entrenamiento aportados que caen en dicha región. (Orellana, 2018)

Dentro de del enfoque de DM, se podrá obtener, gracias a esta metodología, una manera visual de predecir la toma de decisiones según las características que definen a los usuarios y a los bienes entre los que escogen.

3 Análisis Empírico

Una vez definidos los objetivos y la metodología claramente, se pasará de la teoría al enfoque práctico, en el que se analizará empíricamente los datos para obtener información que explique y estime la decisión de consumo.

Como se ha manifestado en apartados anteriores, los datos incluyen variables asociadas a las películas y variables asociadas a los usuarios. Sin embargo, también existe una variable que depende de ambos y que será el eje central del análisis: “las puntuaciones”.

Según la teoría de la utilidad esperada en la elección racional del consumidor, los individuos toman decisiones basadas en la maximización de su utilidad esperada (Varían, 2010). En este contexto, las puntuaciones asignadas por los usuarios reflejan una medida aproximada a la satisfacción o utilidad obtenida de cada película. Por lo tanto, estas valoraciones pueden ser utilizadas como un sustituto válido para las utilidades, permitiendo analizar y predecir las decisiones de consumo de manera consistente con los principios de la teoría de la elección racional.

Para ejemplificar estos aspectos y siguiendo el supuesto de racionalidad, estaríamos considerando que se va a preferir aquella película a la que se le ha asignado una valoración de 4 por encima de aquella a la que se le ha asignado una de 3 (Sanchez-Cuenca, 2009).

$$(4) > (3)$$

Debido a que

$$u(4) > u(3)$$

Con esta perspectiva, centrada en las puntuaciones de las películas, se pretende analizar qué grupos de usuarios obtienen mayor satisfacción del consumo cinematográfico, lo que podría condicionar su elección futura. Igualmente, este proceso también se realizará analizando los atributos del bien consumido (películas) para identificar los géneros cinematográficos que proporcionan mayor satisfacción, lo que incrementa la probabilidad de ser seleccionados por los usuarios.

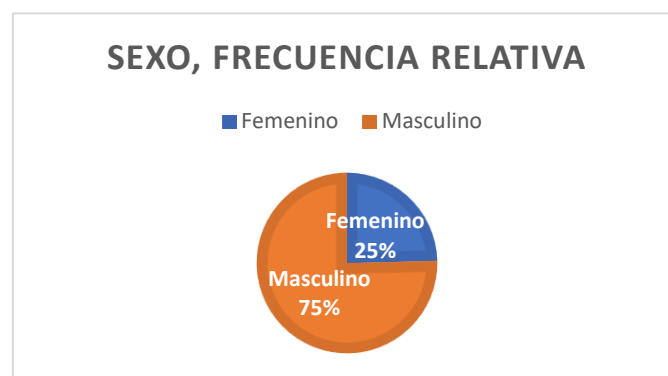
Esta aproximación, aplicada al mercado en su totalidad y sin detenerse en las utilidades individuales, permitirá estimar el éxito general de las películas basándose en las

características tanto del usuario que elige como de la película en cuestión. De este modo, se ayudará a prever cómo el conjunto de usuarios realiza sus elecciones cinematográficas, proporcionando una visión integral de los patrones de consumo.

3.1 Estadísticos descriptivos

Siguiendo con este enfoque, tal y como se ha detallado en la metodología, se realizará una descripción estadística individual de cada una de las variables antes de examinar su relación con las puntuaciones.

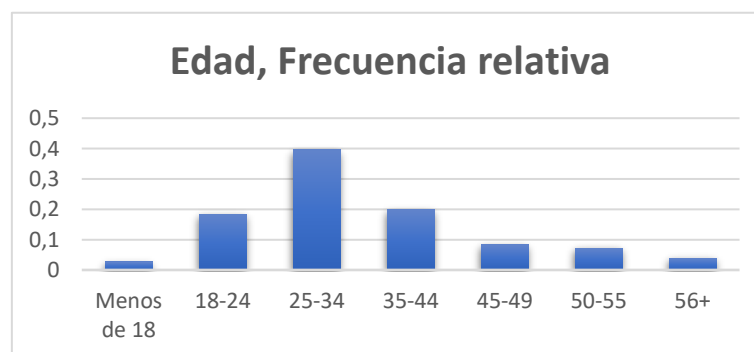
Figura 3, Sexo



Fuente: Elaboración propia en base a los datos aportados por GroupLens

La muestra analizada presenta un sesgo de género, estando el género masculino significativamente más representado que el femenino. Sin embargo, debido al elevado número de participantes, estimamos que se podrá obtener conclusiones representativas y robustas para ambos géneros.

Figura 4, Edad



Fuente: Elaboración propia en base a los datos aportados por GroupLens

Como se puede observar en la figura anterior, destaca significativamente que el grupo de usuarios que más películas a puntuado se encuentra en el rango de edad de 25 a 34 años. Asimismo, los intervalos adyacentes a él, es decir, los rangos de 18 a 24 años y de

35 a 44 años, le siguen como los más frecuentes. Al combinar estos tres intervalos, se observa que aproximadamente el 78% de los usuarios se encuentra en el rango de edad de 18 a 44 años, lo que encaja aproximadamente con una distribución poblacional normal.

Aun teniendo intervalos de edad realmente poco frecuentes, debido que la muestra cuenta con casi un millón de observaciones y 6000 usuarios consideramos que tenemos representación para todas las edades, similar a como pasaba con el sexo.

Figura 5, Ocupación

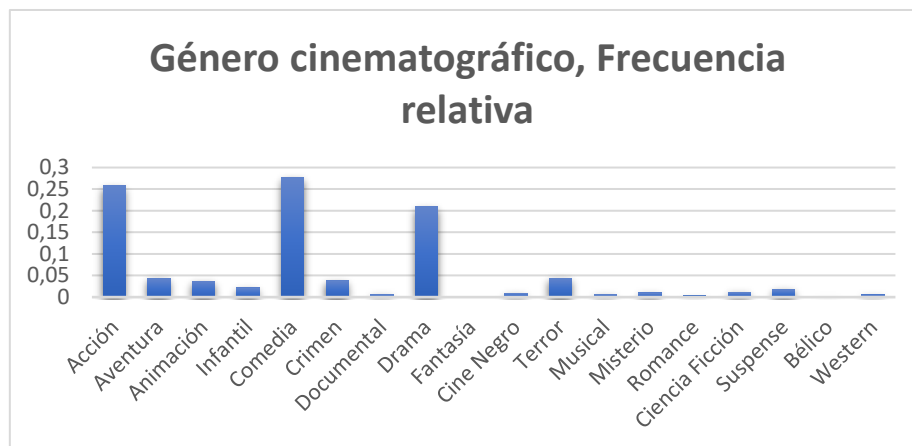


Fuente: Elaboración propia en base a los datos aportados por GroupLens

Se ve representado de una manera clara como en la muestra las ocupaciones no son homogéneas. Empleos como el de agricultura o artesanía apenas tienen representación en comparación con la cantidad de estudiantes o académicos que si puntúan películas.

Es esperable que las categorías menos específicas presenten mayores porcentajes. La categoría "otro o no especificado" tiene la segunda frecuencia más alta, similar a "ejecutivo/gerencial". Estas categorías, al no ser tan específicas como "abogado", exhiben frecuencias elevadas debido a sus límites imprecisos, lo que permite abarcar a una mayor cantidad de trabajadores.

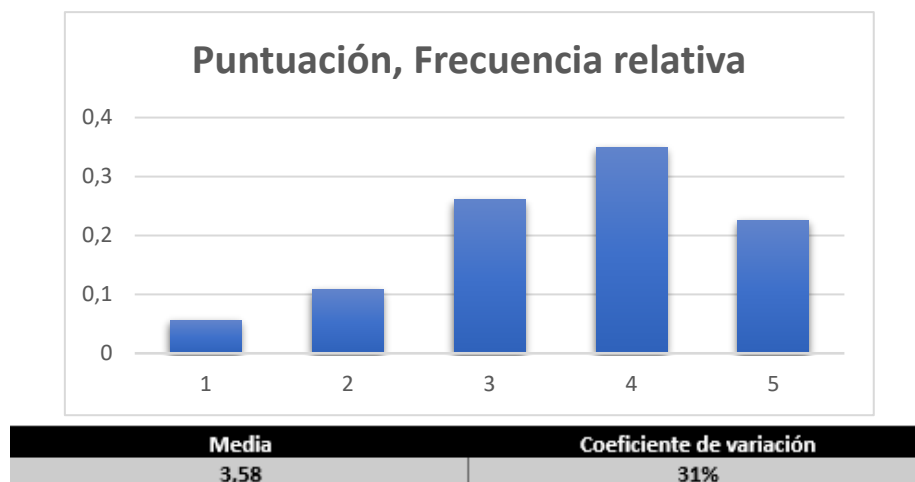
Figura 6, Género cinematográfico



Fuente: Elaboración propia en base a los datos aportados por GroupLens

Los datos analizados son indudablemente heterogéneos, con una clara predominancia de tres géneros específicos. Acción, comedia y drama son, por un amplio margen, los géneros más populares, representando el 70% de las películas valoradas. En contraste, ninguno de los otros géneros supera el 5%.

Figura 7, Puntuación



Fuente: Elaboración propia en base a los datos aportados por GroupLens

Antes de extraer conclusiones de la gráfica de frecuencias relativas, se observa que el valor medio es de 3,58, lo que explica que el valor modal en la tabla de frecuencias relativas sea de 4 estrellas. Además, la variable presenta una distribución relativamente homogénea, con un coeficiente de variación del 31%, lo que indica una variabilidad moderada. Por lo tanto, estimamos que no enfrentaremos problemas significativos de representatividad.

3.2 Tablas cruzadas

Como se ha observado, cada una de estas variables aporta información valiosa de manera individual. Sin embargo, al analizarlas de manera conjunta, se puede comprender mejor cómo ciertas características influyen en la propensión a diferentes decisiones de consumo cinematográfico. Siguiendo con la perspectiva utilitarista del marco teórico, contrastaremos las puntuaciones con los atributos de los usuarios (edad, sexo y ocupación) y de las películas (género cinematográfico). Para ello, se utilizarán tablas cruzadas de frecuencias relativas, interpretadas todas por columnas, ya que por filas no resultaba significativo.

Tabla 1, Tabla cruzada: Puntuación-sexo

	Femenino	Masculino
1	0.054	0.056
2	0.099	0.110
3	0.255	0.262
4	0.352	0.347
5	0.237	0.222

Fuente: Elaboración propia en base a los datos aportados por GroupLens

Es notable que ambos géneros tienden a ser reacios a asignar bajas calificaciones. Sin embargo, los hombres parecen ser más críticos, mostrando frecuencias mayores en las puntuaciones de 1 y 2 en comparación con las mujeres. A pesar de esto, se concluye que las distribuciones de calificaciones entre ambos géneros son muy similares, con los valores más comunes siendo 3 y 4. Por lo tanto, el género del evaluador no parece influir significativamente en la calificación asignada.

Tabla 2, Tabla cruzada: Puntuación-edad

	18-24	25-34	35-44	45-49	50-55	56+	Menos 18
1	0.071	0.060	0.045	0.040	0.040	0.040	0.082
2	0.120	0.113	0.101	0.101	0.082	0.077	0.109
3	0.259	0.263	0.266	0.266	0.254	0.236	0.234
4	0.328	0.345	0.361	0.362	0.365	0.368	0.323
5	0.221	0.216	0.224	0.228	0.256	0.277	0.250

Fuente: Elaboración propia en base a los datos aportados por GroupLens

Se puede observar que, en todas las edades, la calificación más frecuentemente otorgada es de 4 estrellas, lo cual es coherente con las frecuencias generales de las puntuaciones previamente analizadas. Cabe destacar que los grupos de mayor edad (50-55 y 55+) son menos críticos, ya que sus puntuaciones de 4 y 5 estrellas son más altas

que las frecuencias generales. En contraste, los grupos más jóvenes (18-24 y menores de 18) parece que tienden a recibir de manera general menos satisfacción por su consumo cinematográfico.

No obstante, la forma en que los individuos de diferentes grupos de edad distribuyen sus puntuaciones, sin considerar otras características, no varía significativamente. Por lo tanto, se anticipa que la edad no será un factor muy diferenciador en la asignación de calificaciones, ya que no se aprecia ningún sesgo notable en el que un grupo de edad específico otorgue sistemáticamente una calificación mucho más alta o baja.

Tabla 3, Tabla cruzada: Puntuación-edad

	Académico educador	Artista	Administrativo	Universitario/postgrado	Servicio de consumidor
1	0.055	0.061	0.048	0.068	0.049
2	0.106	0.109	0.097	0.117	0.116
3	0.261	0.258	0.251	0.252	0.295
4	0.358	0.337	0.354	0.329	0.325
5	0.218	0.233	0.247	0.231	0.213

	Medico/Cuidado de la salud	Ejecutivo/gerencial	Agricultor	Am@ de casa	Estudiante primaria/secundaria
1	0.055	0.048	0.065	0.034	0.088
2	0.089	0.104	0.108	0.092	0.110
3	0.243	0.266	0.302	0.280	0.231
4	0.362	0.362	0.342	0.367	0.319
5	0.249	0.218	0.180	0.224	0.250

	Abogado	Otro o no especificado	Programador	Jubilado	Ventas/Marketing
1	0.065	0.060	0.045	0.029	0.047
2	0.100	0.112	0.099	0.085	0.105
3	0.234	0.271	0.251	0.238	0.258
4	0.354	0.338	0.365	0.367	0.358
5	0.246	0.216	0.238	0.279	0.230

	Autónomo	Técnico/Ingeniero	Artesano	Desempleado	Escritor	Científico
1	0.044	0.046	0.048	0.095	0.069	0.038
2	0.108	0.098	0.107	0.123	0.123	0.099
3	0.268	0.273	0.302	0.267	0.260	0.242
4	0.363	0.358	0.350	0.300	0.336	0.372
5	0.215	0.222	0.191	0.213	0.211	0.246

Fuente: Elaboración propia en base a los datos aportados por GroupLens

Es notable que los jubilados, científicos y programadores asignan calificaciones altas (4 y 5) con mayor frecuencia en comparación con otros grupos ocupacionales. En contraste, los agricultores artesanos y desempleados tienden a ser más críticos en sus

evaluaciones, como se evidencia en sus frecuencias más altas para las calificaciones de 1 y 2, y más bajas para la calificación de 5.

Asimismo, al igual que con la edad y el sexo, la distribución de calificaciones según la ocupación es muy similar a la distribución general de las puntuaciones. Por lo tanto, se estima que estos atributos personales no influirán significativamente en la asignación de calificaciones y, en consecuencia, en la satisfacción percibida.

Tabla 4, Tabla cruzada: Puntuación-género cinematográfico

	Acción	Aventura	Animación	Infantil	Comedia	Crimen	Documental
1	0.064	0.067	0.045	0.095	0.057	0.025	0.034
2	0.122	0.130	0.075	0.138	0.112	0.068	0.057
3	0.274	0.290	0.246	0.291	0.270	0.219	0.184
4	0.336	0.326	0.371	0.308	0.349	0.375	0.383
5	0.202	0.185	0.260	0.166	0.209	0.310	0.339

	Drama	Fantasia	Cine negro	Terror	Musical	Misterio	Romance
1	0.032	0.079	0.012	0.142	0.032	0.023	0.059
2	0.080	0.160	0.041	0.170	0.070	0.076	0.123
3	0.238	0.265	0.152	0.280	0.221	0.248	0.274
4	0.377	0.310	0.353	0.267	0.353	0.387	0.348
5	0.270	0.183	0.440	0.138	0.322	0.263	0.193

	Ciencia Ficción	Suspense	Bélico	Western
1	0.083	0.050	0.014	0.027
2	0.141	0.113	0.053	0.055
3	0.254	0.266	0.236	0.232
4	0.270	0.369	0.421	0.407
5	0.249	0.200	0.274	0.277

Fuente: Elaboración propia en base a los datos aportados por GroupLens

Comparando cómo los géneros de las películas afectan las calificaciones recibidas por parte del conjunto de usuarios, se observan variaciones significativas en las frecuencias, lo que sugiere una posible relación significativa.

Se puede observar que ciertos géneros tienden a recibir calificaciones más altas, como se refleja en sus porcentajes de puntuaciones de 4 y 5 estrellas. Estos géneros incluyen: animación, crimen, documental, cine negro, bélico y western. En contraste, otros géneros se asocian con utilidades más bajas, evidenciado por sus puntuaciones más frecuentes, como aventura, infantil, fantasía y terror.

Un aspecto destacable en todos los análisis realizados es que no se presentan resultados polarizados, esto es, las frecuencias están por debajo o encima de la media sin valores extremo, siguiendo una tendencia normal.

Como conclusión de este primer análisis, es esperable que, entre todas las variables comparadas con la distribución de puntuaciones, los géneros cinematográficos sean la única que provocará cambios significativos. En el apartado 3.3 a través de los modelos analizados contrastaremos las hipótesis derivadas del análisis descriptivo, siendo estas las siguientes:

- **H1: El sexo no es significativo en la elección**
- **H2: La edad no es significativa en la elección**
- **H3: La ocupación no es significativa en la elección**
- **H4: El género de la película que condiciona la elección**

3.3 Modelos empíricos

Habiendo realizado un análisis preliminar de las relaciones entre las variables a través de tablas cruzadas, se propone ahora avanzar en la construcción de modelos que permitan entender y cuantificar de manera más precisa las interacciones entre las puntuaciones otorgadas a las películas y las características de los usuarios y de los géneros cinematográficos.

Este enfoque permitirá no solo identificar patrones generales, sino también realizar un análisis más detallado y robusto que considere las múltiples dimensiones de los datos, proporcionando así una comprensión más profunda y matizada de las decisiones generales de consumo cinematográfico, que permita rechazar, o no, las hipótesis anteriormente mencionadas.

3.3.1 Regresión lineal

Como se detalló en la metodología, comenzaremos utilizando un modelo de regresión lineal. Este modelo permitirá analizar la relación entre la puntuación de una película (considerada como la utilidad percibida por el usuario) y diversas variables explicativas, siendo estas: la edad, el sexo, la ocupación laboral de los usuarios y el género cinematográfico de las películas. A través de este análisis, se busca cuantificar el impacto de cada una de estas variables en la utilidad asignada por los usuarios, proporcionando así una base sólida para entender los factores que influyen en sus decisiones de consumo generales.

$$Puntuación_i = \beta_0 + \beta_1 Sexo_i + \beta_2 Edad_i + \beta_3 Ocupación_i + \beta_4 Género\ cinematográfico_i + \varepsilon_i$$

Al tratarse de variables categóricas se implementarán como variables ficticias de control, usando la codificación “dummy”. Esto quiere decir que en cada una de ellas se dividirá en tantas variables (que solo podrán adoptar el valor 0 si no ocurre o 1 si sí ocurre) como modalidades, menos una, con el objetivo de no generar multicolinealidad perfecta.

Tabla 5, Modelo de regresión lineal

Variable	Coefficiente	Pr(> t)
Ordenada en el origen	3.41	***
Sexo: Masculino	-0.034	***
Edad: 25-34	0.028	***
Edad: 35-44	0.096	***
Edad: 45-49	0.112	***
Edad: 50-55	0.183	***
Edad: 56+	0.202	***
Edad: Menos 18	0.090	***
Ocupación: Artista	0.027	***
Ocupación: Administrativo	0.099	***
Ocupación: Universitario/postgrado	0.057	***
Ocupación: Servicio al consumidor	0.038	***
Ocupación: Medico/Cuidado de la salud	0.082	***
Ocupación: Ejecutivo/Gerencial	0.047	***
Ocupación: Agricultor	-0.091	***
Ocupación: Am@ de casa	0.070	***
Ocupación: Estudiante primaria/secundaria	0.004	0.689
Ocupación: Abogado	0.057	***
Ocupación: Otro no especificado	0.007	0.135
Ocupación: Programador	0.133	***
Ocupación: Jubilado/a	0.092	***
Ocupación: Ventas/marketing	0.086	***
Ocupación: Científico	0.157	***
Ocupación: Autónomo	0.032	***
Ocupación: Técnico/Ingeniero	0.091	***
Ocupación: Artesano	0.010	*
Ocupación: Desempleado	-0.089	***
Ocupación: Escritor	-0.055	***
Género: Aventura	-0.063	***
Género: Animación	0.238	***
Género: Infantil	-0.176	***
Género: Comedia	0.049	***
Género: Crimen	0.385	***
Género: Documental	0.448	***
Género: Drama	0.272	***
Género: Fantasía	-0.140	***
Género: Cine negro	0.660	***
Género: Terror	-0.393	***
Género: Musical	0.344	***
Género: Misterio	0.288	***
Género: Romance	-0.013	0.523
Género: Ciencia Ficción	-0.038	***
Género: Suspense	0.061	***
Género: Bélica	0.354	***
Género: Western	0.341	***

Residual standard error:	1.1 on 998491 degrees of freedom
Multiple R-squared:	0.03139
Adjusted R-squared:	0.03135

Fuente: Elaboración propia en base a los datos aportados por GroupLens

Nota: 0 = *** 0.001 = ** 0.01 = *

Este modelo realizado con RStudio, pretende cuantificar las tendencias generales de consumo cinematográfico para una muestra considerada representativa, estimando que un nivel mayor de satisfacción hará más probable el consumo. Con todo ello se pretende comprender la utilidad esperada de una película en función de su género y las características utilizadas para segmentar a los consumidores en distintos grupos.

Por lo tanto, las características de los individuos con coeficientes positivos en el modelo sugieren una mayor propensión al consumo cinematográfico entre aquellos que las poseen. Simultáneamente, los coeficientes de los géneros cinematográficos indicarán qué géneros son más propensos a ser seleccionados, al proporcionar mayores niveles de satisfacción a los consumidores.

Analizando los resultados, se observa como en la ordenada en el origen se están teniendo en cuenta los efectos fijos de referencia de que la película sea vista por una mujer, de que el consumidor sea un académico y finalmente que pertenezca al grupo de edad 19-25. Esta ordenada en el origen representa también la puntuación media base antes de considerar cualquier otra variable.

Comenzando por la variable género, observamos en el coeficiente "Género: Masculino" un valor de -0,034, lo que indica que, en promedio y ceteris paribus, los hombres tienden a percibir una utilidad ligeramente menor en las películas en comparación con las mujeres. Sin embargo, esta diferencia es tan reducida que podría indicar que no se presenten cambios significativos en los patrones de consumo por sexo, tal como habíamos teorizado en las tablas cruzadas, no habiendo indicios suficientes para rechazar la H1.

En cuanto a las diferentes categorías de edad, las hipótesis de partida basadas en el análisis estadístico descriptivo sugerían que, al igual que con el sexo, la edad del consumidor no influiría significativamente en la satisfacción. Al comprobar esta hipótesis con los resultados del modelo, observamos en los coeficientes de las distintas modalidades de la variable "Edad" (usando como referencia el grupo de edad 18-24) que los grupos de mayor edad (35-44, 45-49, 50-55, 56+) muestran incrementos progresivos en las puntuaciones, con los mayores de 56 años otorgando puntuaciones 0,202 unidades más altas. Asimismo, los usuarios menores de 18 años también tienden a dar puntuaciones más altas, con un incremento de 0,090 unidades. De esta manera, los usuarios entre 18-24 son los que otorgan las peores puntuaciones. Podemos concluir que, a partir de los 25 años, cuanto mayor es el grupo etario, mayor es la satisfacción percibida por las películas, haciendo que estos estén más dispuestos a consumir más. No obstante, aunque esta relación entre edad y satisfacción no es muy intensa, es lo suficiente como para refutar la hipótesis H2 en la que se planteaba la edad como un elemento no significativo en la elección de películas.

Analizando la tercera variable, ocupación, la hipótesis 3 planteaba que la ocupación de los usuarios tendría un impacto mínimo en la utilidad percibida de las películas. Para verificar esto, se examinan los resultados del modelo. Al analizar el conjunto de coeficientes de la variable "Ocupación", observamos diversos impactos en las puntuaciones en comparación con el grupo de referencia (académicos):

- Ocupaciones como "Administrativo", "Universitario/Postgrado", "Médico/Cuidado de la salud", "Ejecutivo/Gerencial", "Programador", "Jubilado", "Ventas/Marketing", y "Científico" muestran incrementos en las puntuaciones.
- En contraste, ocupaciones como "Agricultor", "Desempleado" y "Escritor" tienden a otorgar puntuaciones más bajas.

No obstante, consideramos que esta relación es poco intensa. Los científicos, que son los más generosos en sus calificaciones, otorgan solo una décima y media más en comparación con los académicos, mientras que los agricultores, quienes otorgan las calificaciones más bajas, difieren en solo -0,091 unidades respecto a los académicos. Aun así, estos resultados son concluyentes para poder rechazar la hipótesis H3 en la que se planteaba que la ocupación no afectaría a la toma de decisiones (aunque afecte en muy poca medida).

En resumen, habiendo visto de manera general como afectan los atributos de los individuos las utilidades, se ha observado que la satisfacción que en general estos perciben por el conjunto de películas es muy similar, no se espera que entre ellos tengan patrones de consumo muy diferentes solo por ser su sexo, edad u ocupación, ya que estas cualidades no afectan en gran medida a la satisfacción que consumir películas les aporta.

Centrándonos en las propiedades de las opciones que pueden escoger, se observará como efectivamente los géneros cinematográficos, al tener coeficientes significativamente mayores que el resto de las variables, provocan más cambios en las notas, y por lo tanto en las elecciones de los consumidores, aceptando así la H4.

Utilizando el género "Acción" como referencia, se observa que hay una preferencia generalizada hacia géneros como "Documental" (coeficiente de 0.448), "Cine negro" (0.660), "Musical" (0.344) y "Bélica" (0.354), los cuales reciben puntuaciones significativamente más altas. Esto sugiere que existe una mayor satisfacción general hacia estas películas, lo que probablemente se traduzca en un mayor consumo de estas. Por otro lado, géneros como "Terror" (-0.393), "Infantil" (-0.176) y "Fantasía" (-0.140) muestran puntuaciones significativamente más bajas, lo que indica que estas películas no causan gran satisfacción al conjunto de usuarios, o al menos no tanto como las películas de acción.

Deteniéndonos a analizar el modelo desde un punto de vista formal, vemos como la bondad del ajuste, con un valor R-cuadrado igual a 0,031, demuestra que el porcentaje de la variación en la variable de respuesta que es explicado por el modelo es bajo. En todo caso, este nos es un aspecto relevante en la estimación de una variable no continua. Consecuentemente, si los coeficientes son significativos, aún representan el cambio medio en la respuesta para una unidad de cambio en el predictor mientras se mantienen constantes los otros predictores del modelo (Minitab, 2019).

Es por esto, que se analizará la significatividad de los coeficientes de las variables para ver si son prescindibles o no, para ello se recurre a la última columna de la tabla, donde aparece el valor del estadístico p obtenido a partir de una distribución T-Student. Precisamente en nuestro modelo, de 45 variables, solo tres tienen un coeficiente no significativo, haciéndolo a primera vista un modelo robusto.

Aun así, la regresión lineal asume una varianza constante de los errores en todos los niveles de las variables predictoras. En el caso de una variable explicada no continua, puede haber varianzas desiguales para diferentes categorías, lo que viola la suposición de homocedasticidad. Se tendrá en este caso un problema de heterocedasticidad que podría invalidar nuestros estimadores al estos perder su calidad de óptimos. Igualmente, los contrastes de significatividad pasarían a depender de una matriz de varianzas-covarianzas que no se calcula adecuadamente, volviéndolos también incorrectos (Oddi et al, 2020). Para comprobar si la perturbación no se distribuye normalmente, se realiza una prueba de Breush-Pagan (Galán, 2016), que efectivamente demuestra la existencia de heterocedasticidad.

Tabla 6, Prueba de Breush-Pagan

Test Breusch-Pagan	
p-value	2.2e-16 < 0,05

Fuente: Elaboración propia en base a los datos aportados por GroupLens

Y es que, utilizar un modelo lineal para estimar una variable discreta con solo cinco valores posibles (las puntuaciones de las películas) no es apropiado. Sería más adecuado emplear un modelo de regresión no lineal, como el Probit ordenado, que corrige los problemas de heterocedasticidad. Sin embargo, dado que el modelo Probit ordenado aumenta la complejidad en la interpretación de los efectos de las variables, se ha optado en el siguiente apartado por la categorización de la variable Y_t (calificación de las películas) en dos categorías, lo que permitirá emplear la regresión logística binomial, que tiene una capacidad explicativa mucho más decisiva.

3.3.2 Logit binomial

Como se acaba de mencionar, para aplicar la metodología de regresión logística binomial y así obtener probabilidades sobre las posibles elecciones de los individuos, es necesario primero simplificar la variable utilizada como proxy de la utilidad, es decir, las

puntuaciones. Procederemos a dicotomizar esta variable en dos categorías: aprobado (valores 4 y 5) y suspenso (valores 1, 2 y 3). Esta simplificación, de acuerdo con el marco teórico, permitirá afirmar que cuanto mayor sea la probabilidad de una nota "aprobada", mayor será la probabilidad de que dicha película sea elegida, o de que el dicho usuario decida consumir cine.

Similar al modelo previamente realizado, las variables explicativas que aportarán información sobre la probabilidad de obtener una nota aprobada serán las características que definen a los individuos (edad, sexo y ocupación) y el género de las películas a elegir.

$$P(\text{Aprobado})_i = \beta_0 + \beta_1 \text{Sexo}_i + \beta_2 \text{Edad}_i + \beta_3 \text{Ocupación}_i + \beta_4 \text{Género cinematográfico}_i + \varepsilon_i$$

Al igual que con el modelo lineal, se implementa la metodología de variables ficticias para incorporar todas las categorías de las variables cualitativas. Se tendrá así, como efectos fijos en la ordenada en el origen las categorías omitidas: que la película sea vista por una mujer, que el consumidor sea un académico, que pertenezca al grupo de edad 19-25 y finalmente que la película sea de acción.

Tabla 7, Modelo logit binomial

Variable	Coefficiente	Pr(> t)
Ordenada en el origen	0.057	***
Sexo: Masculino	-0.058	***
Edad: 25-34	0.043	***
Edad: 35-44	0.134	***
Edad: 45-49	0.148	***
Edad: 50-55	0.268	***
Edad: 56+	0.327	***
Edad: Menos 18	0.151	***
Ocupación: Artista	0.025	*
Ocupación: Administrativo	0.140	***
Ocupación: Universitario/postgrado	0.081	***
Ocupación: Servicio al consumidor	-0.033	*
Ocupación: Medico/Cuidado de la salud	0.150	***
Ocupación: Ejecutivo/Gerencial	0.062	***
Ocupación: Agricultor	-0.189	***
Ocupación: Am@ de casa	0.055	***
Ocupación: Estudiante primaria/secundaria	0.033	0.143
Ocupación: Abogado	0.132	***
Ocupación: Otro no especificado	-0.018	*
Ocupación: Programador	0.206	***
Ocupación: Jubilado/a	0.120	***
Ocupación: Ventas/marketing	0.119	***
Ocupación: Científico	0.252	***
Ocupación: Autónomo	0.032	**
Ocupación: Técnico/Ingeniero	0.109	***
Ocupación: Artesano	-0.029	0.135
Ocupación: Desempleado	-0.149	***
Ocupación: Escritor	-0.083	***
Género: Aventura	-0.116	***
Género: Animación	0.388	***
Género: Infantil	-0.257	***
Género: Comedia	0.080	***
Género: Crimen	0.626	***
Género: Documental	0.811	***
Género: Drama	0.439	***
Género: Fantasía	-0.190	***
Género: Cine negro	1.169	***
Género: Terror	-0.523	***
Género: Musical	0.536	***
Género: Misterio	0.452	***
Género: Romance	-0.012	0.739
Género: Ciencia Ficción	-0.086	***
Género: Suspense	0.120	***
Género: Bélica	0.610	***
Género: Western	0.595	***

Fuente: Elaboración propia en base a los datos aportados por GroupLens

Nota: 0 = *** 0.001 = ** 0.01 = *

La estimación de este modelo logit ha sido realizado a través del programa Rstudio. Los resultados de esta estimación presentan una limitación inherente de la regresión logística: al no ser un modelo lineal, los estimadores no pueden interpretarse de manera directa. La única información que se puede extraer de ellos es el signo de la relación. Para determinar cuantitativamente el efecto de cada variable en la probabilidad, es necesario recurrir a los efectos marginales. Dichos efectos permiten interpretar su valor como incrementos (o descensos) de puntos porcentuales en la probabilidad de que ocurra el suceso recogido en nuestra variable explicada (Gareth James et al, 2023).

Tabla 8, Efectos marginales

Variable	Df/Dx	P> z
Sexo: Masculino	-0.014	***
Edad: 25-34	0.010	***
Edad: 35-44	0.032	***
Edad: 45-49	0.035	***
Edad: 50-55	0.064	***
Edad: 56+	0.077	***
Edad: Menos 18	0.036	***
Ocupación: Artista	0.006	*
Ocupación: Administrativo	0.033	***
Ocupación: Universitario/postgrado	0.019	***
Ocupación: Servicio al consumidor	-0.008	*
Ocupación: Medico/Cuidado de la salud	0.036	***
Ocupación: Ejecutivo/Gerencial	0.015	***
Ocupación: Agricultor	-0.046	***
Ocupación: Am@ de casa	0.013	***
Ocupación: Estudiante primaria/secundaria	0.008	0.143
Ocupación: Abogado	0.031	***
Ocupación: Otro no especificado	-0.004	*
Ocupación: Programador	0.049	***
Ocupación: Jubilado/a	0.029	***
Ocupación: Ventas/marketing	0.029	***
Ocupación: Científico	0.060	***
Ocupación: Autónomo	0.007	**
Ocupación: Técnico/Ingeniero	0.026	***
Ocupación: Artesano	-0.007	0.135
Ocupación: Desempleado	-0.036	***
Ocupación: Escritor	-0.020	***
Género: Aventura	-0.028	***
Género: Animación	0.091	***
Género: Infantil	-0.063	***
Género: Comedia	0.019	***
Género: Crimen	0.142	***
Género: Documental	0.177	***
Género: Drama	0.104	***
Género: Fantasía	-0.047	***
Género: Cine negro	0.238	***
Género: Terror	-0.129	***
Género: Musical	0.123	***
Género: Misterio	0.105	***
Género: Romance	-0.003	0.739
Género: Ciencia Ficción	-0.021	***
Género: Suspense	0.029	***
Género: Bélica	0.138	***
Género: Western	0.135	***

Fuente: Elaboración propia en base a los datos aportados por GroupLens

Nota: 0 = *** 0.001 = ** 0.01 = *

De manera similar al modelo lineal, las características individuales con efectos marginales positivos indicarán que los individuos que las poseen tendrán una mayor probabilidad de asignar calificaciones altas, lo que sugiere una mayor propensión a consumir cine. Igualmente, los efectos marginales asociados a los géneros

cinematográficos identificarán aquellos géneros que deberían ser seleccionados con mayor o menor frecuencia.

Comenzando con la variable del sexo, se observa que los hombres tienen, en promedio, una probabilidad menor de otorgar una calificación positiva, como lo indica el efecto marginal de "Sexo: Masculino", que evidencia una disminución del 1.4% en la probabilidad de otorgar una buena calificación en comparación con las mujeres. Esta relación, aunque presente, es mínima, como se observó en el modelo lineal, no se puede rechazar H1.

Al examinar las diferentes categorías de edad, también se observa un patrón similar al del modelo lineal. Con la referencia en el grupo de edad de 18 a 24 años, los demás grupos tienen una mayor probabilidad de otorgar buenas calificaciones a las películas. Se puede afirmar que, a partir de los 25, a medida que la edad del individuo aumenta, también aumenta la probabilidad de otorgar una buena calificación, sin embargo, la intensidad de esta relación es baja, con efectos marginales que no superan el 10%. Aun así, se rechazará a la hipótesis H2, observándose diferencias significativas entre los diferentes grupos de edad que pueden afectar a la toma de decisiones.

En cuanto a las ocupaciones, con la referencia en la categoría "Académico", se observa que ocupaciones como científico o programador aumentan la probabilidad de otorgar una buena calificación en aproximadamente un 5%, mientras que ocupaciones como agricultor o desempleado la reducen en un 4%. Se observa así una similitud con el modelo lineal en cuanto a la baja importancia de la relación entre la ocupación y la probabilidad de otorgar una buena calificación, sin embargo, se considera la relación suficiente intensa como para afectar al consumo, pudiendo así rechazar la H3.

Finalmente, los géneros que definen la película que se consume. Estudiando los efectos marginales (teniendo en cuenta que la modalidad de referencia es el género acción) se observa como las características del bien a elegir son lo que realmente va a afectar a la probabilidad de tener una buena nota, tal y como se había teorizado en H4. Por lo tanto, si mantenemos las características de los usuarios fijas, habrá géneros que tengan mayor probabilidad de recibir una buena nota (demostrando que aportan gran satisfacción) que otros, haciendo que su consumo sea más frecuente.

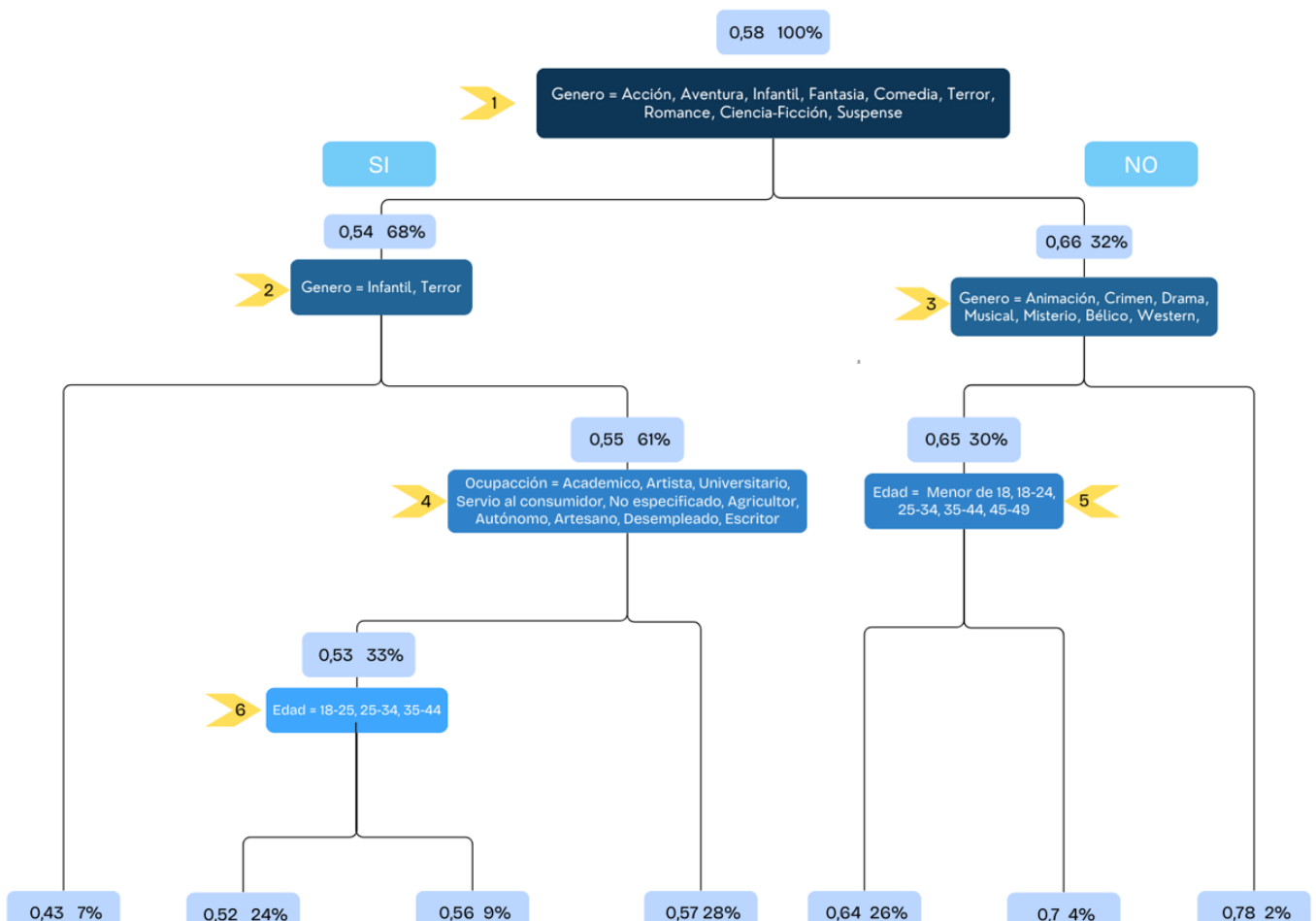
- Documental, Cine Negro, Crimen, Musical: Estos géneros tienen los efectos marginales más altos (17.78%, 23.86%, 14.25%, y 12.30% respectivamente), indicando que las películas dentro de estos géneros tienen una alta probabilidad de recibir buenas calificaciones, y, por lo tanto, la satisfacción que aportan suele ser mayor.
- Terror, Infantil, Ciencia Ficción: Estos géneros muestran efectos marginales negativos, sugiriendo una menor probabilidad de recibir buenas calificaciones y, consecuentemente, una menor probabilidad de ser seleccionadas de manera general.

3.3.3 Decision tree

Concluimos los modelos con este subapartado donde se procederá a construir un árbol de decisión para continuar explicando las relaciones entre las características de los individuos, los géneros cinematográficos y las calificaciones asignadas a las películas. Este enfoque complementará los modelos de regresión logit binomial, proporcionando una representación gráfica más interpretable de cómo interactúan estas variables para influir en la probabilidad de que una película reciba una buena calificación y, en consecuencia, sea seleccionada.

El árbol de decisión permitirá visualizar las reglas y patrones derivados de los datos, destacando las combinaciones específicas de características que conducen a diferentes resultados en las calificaciones de las películas. Así, se obtendrá una comprensión más detallada y práctica de los factores que inciden en la toma de decisiones de los usuarios en el ámbito cinematográfico.

Figura 8, Árbol de clasificación



Fuente: Elaboración propia en base a los datos aportados por GroupLens

Nodos Superiores y Variables de Mayor Influencia

Sabemos que los nodos superiores en un árbol de decisión representan las bifurcaciones basadas en las condiciones de las variables que explican la mayor variabilidad en la variable dependiente. En este caso, los nodos superiores están dominados por el género cinematográfico, confirmando la H4 sobre la importancia de esta variable en la asignación de calificaciones y corroborando lo que ya validamos a través de la información de los modelos econométricos analizados en el apartado 3.2.

Las siguientes bifurcaciones se realizan en función de la ocupación y la edad del usuario, indicando que estas variables, aunque tienen influencia, es mucho menor que la del género cinematográfico. El sexo, en cambio, no aparece como un nodo decisor, lo que indica una baja relevancia en la probabilidad de asignar una buena calificación. Estas afirmaciones sobre la jerarquía de las variables encajan con los bajos valores que observamos en los efectos marginales de "Ocupación", "Edad" y "Sexo" de los modelos de regresión logística.

Probabilidad de Buena Calificación, Nodo a Nodo

En el nodo raíz, la probabilidad base de obtener una buena calificación es de 0,58. Esto implica que, sin considerar ninguna variable específica, es más probable obtener una buena calificación que una mala.

Siguiendo con la primera división, para películas de los géneros Acción, Aventura, Infantil, Fantasía, Comedia, Terror, Romance, Ciencia-Ficción y Suspense, la probabilidad de buena calificación desciende a 0,54. Para los géneros de Terror e Infantil, la probabilidad es aún menor (0,47), indicando que es mayor la probabilidad de obtener una mala calificación que la de una buena. La manera en la que al pertenecer a estos géneros la satisfacción se espera menor, encaja con los efectos marginales de dichos géneros, observados en la Figura 5, ya que muchos apenas sumaban a la probabilidad de buena nota y algunos incluso eran negativos.

Dentro del subconjunto de géneros: Acción, Aventura, Fantasía, Comedia, Romance, Ciencia-Ficción y Suspense (excluyendo Terror e Infantil), la probabilidad de buena calificación es de 0,55. Siguiendo por la misma rama, la ocupación del usuario ofrece mayor segmentación (nodo 4): si la ocupación es Académico, Artista, Universitario, Servicio al Consumidor, No Especificado, Agricultor, Autónomo, Artesano, Desempleado o Escritor, la probabilidad de una buena calificación disminuye a 0,53; para otras ocupaciones, aumenta a 0,57. De manera similar a los géneros cinematográficos, observamos que las modalidades de "Ocupación" que mostraban efectos marginales

cercanos a 0 o negativos, en el árbol de clasificación se las asignan a grupos con menores probabilidades de otorgar una buena calificación.

Para usuarios dentro del rango de 18 a 44 años, en combinación con los géneros mencionados y las ocupaciones específicas (nodo 6), la probabilidad de una buena calificación es de 0,52. Para aquellos fuera de este rango, la probabilidad es de 0,56. Esto es consistente con las conclusiones derivadas de los modelos econométricos, los cuales indican que los usuarios menores de 18 años y aquellos mayores de 44 años tienden a experimentar una mayor satisfacción con las películas, reflejada en puntuaciones más altas en términos generales.

Si ahora se analiza la rama derecha del árbol nos fijaremos en el nodo 3, donde se segmenta según si la película pertenece, o no, a los géneros: Animación, Crimen, Drama, Musical, Misterio, Bélico o Western. Si la respuesta es afirmativa la probabilidad de una buena calificación es significativamente superior (0,66) respecto a los géneros del nodo 1. Si la respuesta es negativa, esto quiere decir que el género es Cine Negro o Documental, haciendo que la probabilidad sea aún mayor, alcanzando 0,78.

Finalmente, en el nodo número 5, si el usuario tiene más de 50 años y la película pertenece a los géneros del nodo 3, la probabilidad de buena calificación es de 0,70. Para usuarios menores de 50 años, esta probabilidad es de 0,64.

Se concluye así que el árbol de decisión proporciona una herramienta predictiva decisiva que segmenta el mercado cinematográfico de manera efectiva. La estructura jerárquica del árbol permite identificar cómo las variables género cinematográfico, ocupación y edad influyen en la probabilidad de que una película reciba una buena calificación, proporcionando una visión detallada y precisa de las dinámicas subyacentes en las decisiones de calificación de los usuarios.

3.4 Comparación de Modelos y Discusión general.

Este estudio ha implementado tres enfoques distintos para establecer relaciones entre la variable utilizada como estimación de la utilidad (las puntuaciones) y las características de los bienes a elegir e individuos.

En primer lugar, la regresión lineal permitió, mediante los coeficientes de mínimos cuadrados, cuantificar el impacto de las características de los individuos y las películas en la puntuación esperada. En segundo lugar, la regresión logística, proporcionó la posibilidad de estimar la probabilidad de que una película consumida genere una alta

cualificación, lo que indicaría una mayor satisfacción. Finalmente, el árbol de decisión facilitó la segmentación de los datos según la probabilidad de asignar una puntuación alta.

Aunque estas tres metodologías han proporcionado información valiosa de manera individual, es crucial señalar que la regresión lineal presenta problemas de heterocedasticidad y en general, no es la más adecuada para este tipo de variables categóricas. Sin embargo, aunque aplicar una regresión lineal a una variable dependiente no continua no es teóricamente correcto, en la práctica los resultados son extremadamente similares a los obtenidos con modelos probit o logit (Ferrer-i-Carbonell y Frijters 2004), tal y como se ha comprobado a lo largo de esta investigación

Después de esta aclaración, la aplicación de estas tres metodologías (regresión lineal, modelo logit y árbol de decisión) ha llevado a resultados muy similares. Esta coherencia en los hallazgos a través de diferentes técnicas analíticas refuerza la fiabilidad y coherencia de las conclusiones obtenidas a partir de los datos empíricos. A continuación, se presentan los hallazgos:

- **Probabilidad de Satisfacción Cinematográfica:** Los usuarios tienen una mayor probabilidad de reportar una satisfacción adecuada con su consumo cinematográfico que de no hacerlo, lo que indica una tendencia general hacia una valoración positiva de las películas.
- **Impacto de los Atributos del Usuario:** Excepto sexo (donde H1 no se puede rechazar), las variables demográficas utilizadas para diferenciar entre usuarios (edad y ocupación) permiten rechazar la hipótesis H2 y H3 de no existencia de diferencias significativas en términos de satisfacción reportada. Mientras tanto, el género cinematográfico de las películas seleccionadas es el factor que más influye en las calificaciones obtenidas y por lo tanto en los patrones de consumo, aceptándose así H4.
- **Criticidad por Género:** Los datos sugieren que los hombres, en general, son más críticos en sus calificaciones, reportando menor satisfacción en comparación con las mujeres, pero no creando diferencias en los patrones de consumo.
- **Efecto de la Edad:** A partir de los 25 años, se observa que a medida que aumenta la edad de los usuarios, tienden a otorgar mejores calificaciones a las películas que consumen, haciendo que se muestren más propensos a consumir cine.

- **Variabilidad en las Ocupaciones:** Las ocupaciones muestran efectos variados en las calificaciones, pero no se puede concretar un patrón común entre las ocupaciones que suelen puntuar mejor o peor.
- **Influencia del Género Cinematográfico:** El género cinematográfico, como principal diferenciador en las calificaciones, muestra valores muy dispersos. Los géneros que generalmente son más probables de ser elegidos, ya que, reciben mejores calificaciones son: Animación, Crimen, Drama, Musical, Misterio, Bélico, Western, Cine Negro y Documental. Por otro lado, los géneros que tienden a recibir peores calificaciones incluyen: Acción, Aventura, Infantil, Fantasía, Comedia, Terror, Romance, Ciencia-Ficción y Suspense.

Conclusiones y ampliación.

Teniendo en cuenta que el objetivo de este trabajo era demostrar el uso de modelos de elección racional para analizar la toma de decisiones en un ejemplo empírico, se ha concluido que, en efecto, estos son capaces de aportar resultados relevantes.

Al ejemplo estudiado de consumo cinematográfico se le ha aplicado desde la más simple regresión lineal, pasando por los modelos de regresión logística, hasta el árbol de decisión propio del Machine learning, todo para poder comprender como el conjunto de agentes económicos deciden asignar en general satisfacciones que afectan a los patrones de consumo.

Entrando en profundidad en cada modelo vemos que cada uno de ellos ha presentado ciertas características y matices sobre la información que aportan, haciendo relevante que se haya estudiado este fenómeno desde tres ópticas diferentes.

Primero, el modelo de regresión lineal proporcionó una visión cuantitativa que facilitaba la comprensión del impacto de las características del usuario y del bien elegido en la satisfacción esperada. Mientras tanto, la regresión logística binomial, con su enfoque en la elección discreta, permitió identificar y clasificar las variables que influían en la probabilidad de asignar altas o bajas probabilidades de satisfacción. Finalmente, el árbol de clasificación ofreció una representación comprensible de las características que polarizan la asignación de satisfacción, permitiendo segmentar a los usuarios en grupos con mayor o menor probabilidad de otorgar calificaciones altas.

Todos estos enfoques convergen en pautas comunes que indican la adecuación y robustez del análisis para explicar las decisiones de consumo. Una de ellas, es que los usuarios de manera general tienen una mayor probabilidad de reportar una satisfacción adecuada con su consumo cinematográfico, destacando que el género de la película es el factor más influyente en las calificaciones obtenidas, superando en relevancia a las variables demográficas como edad, sexo y ocupación.

Pero si el enfoque es variable a variable, se han establecido cuatro hipótesis generales que han sido validadas o rechazadas del mismo modo por los tres modelos.

La primera variable, relativa al sexo, una vez analizada afirma que los hombres, en general, tienden a ser más críticos y reportan menor satisfacción comparados con las mujeres, pero no hemos podido rechazar la hipótesis nula de diferencias significativas

(H1) y por lo tanto no se esperan cambios en los patrones de consumo. Con relación a la segunda variable, relativa a la edad, a medida que los usuarios envejecen, especialmente después de los 25 años, tienden a otorgar mejores calificaciones a las películas, lo que hace más probable que sigan consumiendo, siendo esta relación lo suficientemente intensa como para rechazar H2. Continuando con las ocupaciones, estas muestran efectos variados en la satisfacción reportada lo suficientemente relevantes como para rechazar H3, demostrando así una variabilidad significativa en el consumo, pero no se presenta un patrón claro sobre aquellas que lo incentivan y las que no. Finalmente, los géneros cinematográficos por su relevancia tienen una gran importancia en la toma de decisiones (aceptándose así H4), soliendo recibir mejores calificaciones, y por lo tanto más probables de ser consumidos los géneros: Animación, Crimen, Drama, Musical, Misterio, Bélico, Western, Cine Negro y Documental, mientras que: Acción, Aventura, Infantil, Fantasía, Comedia, Terror, Romance, Ciencia-Ficción y Suspense tienden a obtener peores calificaciones, haciendo que de manera general aporten menos satisfacción y se consuman menos.

Se obtiene así, gracias a la metodología seguida, una herramienta útil, que podría permitir a un estudio cinematográfico predecir según las características del individuo y género de la película, si su filme va a recibir una cualificación alta, haciendo que aporte más satisfacción en general y por lo tanto tenga más éxito en el mercado.

Aun así, al analizar los modelos y su capacidad para explicar la satisfacción y la toma de decisiones en el consumo cinematográfico, es importante destacar que, aunque el enfoque abordado proporciona resultados valiosos, hay espacio para futuras líneas de investigación.

Primero, se ha partido de la premisa de que una mayor satisfacción aumenta la probabilidad de que una película sea escogida. De manera similar, se asume la proposición teórica de que los consumidores que otorgan calificaciones más altas en promedio tienden a consumir más películas. El enfoque teórico presentado relaciona las calificaciones con la satisfacción lo cual llevaría en última instancia, a incrementar la probabilidad de que el usuario finalmente consuma la película, esto es, una vez definidos cuales son los tipos de usuarios y géneros con mayor puntuación, estos deberían de ser los que finalmente obtienen una mayor cuota de mercado

No obstante, esta premisa no siempre sucede en la realidad. Por ejemplo, observamos que géneros cinematográficos altamente valorados, como documentales y western, deberían, en teoría, ser los más consumidos debido a su alta probabilidad de recibir buenas calificaciones. Sin embargo, los datos de la Figura 6 que muestra el porcentaje de valoraciones por género cinematográfico, indican que esto no se corresponde con la realidad.

Una posible explicación podría ser la sobresaturación del mercado con películas de géneros que son más sencillos de producir para los estudios cinematográficos. En este escenario, los usuarios no tendrían una variedad suficiente de películas para consumir todo lo que desearían de los géneros mejor valorados. Esto nos lleva a realizar

recomendaciones de programación a los estudios, puesto que si el objetivo de la cadena, o estudio, es la maximización de la satisfacción del usuario, deberían enfocarse en los géneros que más probabilidad de dar una buena nota tienen, y quizás de este modo aumentar el número de usuarios de este ámbito, puesto que es esperable que dada su satisfacción, sean mayores consumidores, siendo de interés para las productoras realizar análisis más profundos de las tendencias de consumo por géneros cinematográficos.

Sin embargo, existen limitaciones técnicas, como el hecho de que las utilidades ofrecidas por diferentes géneros cinematográficos pueden no ser directamente comparables debido a que las experiencias que proporcionan son diferentes. Esto implicaría, tal y como señalamos en el punto anterior, la necesidad de un análisis más profundo de las características psicológicas de esos usuarios que se comportan de modo diferente en relación a los que se podrían denominar “grupos” o “categorías” de géneros.

Para avanzar en esta línea de investigación se sugiere el uso de metodologías más complejas como pueden ser los modelos Naive Bayes en su modalidad de recomendadores, ya que con ellos se podría predecir no solo la nota, si no el género que más le puede gustar elegir al consumidor.

Se concluye así que gracias al enfoque práctico de esta investigación se podrá comprender dinámicas que permiten predecir la satisfacción de los consumidores de manera más precisa, lo que puede ser importante para estrategias de marketing y producción cinematográfica. Tal y como se predecía en el apartado de objetivos, esta información es particularmente útil para que las plataformas de streaming ajusten su oferta a las preferencias y decisiones del mercado de manera óptima.

Bibliografía

Ada Ferrer-i-Carbonell, & Paul Frijters. (2004). How important is methodology for the estimates of the determinants of happiness? *The Economic Journal*, 114, 641–659.

Daniel McFadden. (1980). Econometric models for probabilistic choice among products. *The Journal of Business*, 53(3) <https://www.jstor.org/stable/2352205>

Eric Dolan. (2017, May 24.). *Study: Gender stereotypes about movie preferences are mostly accurate.* <https://www.psypost.org/>. https://www.psypost.org/study-gender-stereotypes-movie-preferences-mostly-accurate/#google_vignette

Facundo José Oddi, Fernando Miguez, Guido Benedetti, & Lucas Alejandro Garibaldi. (2020). Cuando la variabilidad varía: Heterocedasticidad y funciones de varianza. *Ecología Austral*, 30(3) <https://doi.org/http://dx.doi.org/10.25260/EA.20.30.3.0.1131>

Gareth James, Daniela Witten, Trevor Hastie, & Robert Tibshirani. (2023). *An introduction to statistical learning* (2° ed.)

Grouplens. (2024, *MovieLens 1M dataset*). <https://grouplens.org/>.
<https://grouplens.org/datasets/movielens/1m/>

Hal Ronald Varían. (2010). *Microeconomía intermedia: un enfoque actual*: (9 ed.).

Barcelona, Spain: Antoni Bosch editor.

IBM. (2024, *¿Qué es un árbol de decisión?* IBM.com. Recuperado 18/05/2024, de

<https://www.ibm.com/es-es/topics/decision-trees>

IBM. (2015, *Conceptos básicos de ayuda de CRISP-DM*. IBM.com. Recuperado

18/05/2024, de <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp->

[help-overview](https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview)

Ignacio Sánchez-Cuenca. (2009). *Teoría de juegos (Colección Cuadernos*

Metodológicos, 34): (2 ed.). CIS - Centro de Investigaciones Sociológicas.

<https://elibro-net.ezbusc.usc.gal/es/lc/busc/titulos/52034>

Javier Galán, Jorge Feregrino, Lucía A. Ruíz Galindo, Luis Quintana Romero, Miguel

Ángel Mendoza González, & Roldán Andrés Rosales. (2016). *Econometría*

aplicada utilizando R (1º ed.)

http://saree.com.mx/econometriaR/sites/default/files/Ebook_econometriaR.pdf

Johana Orellana. (2018, *Arboles de decision y random forest*. Recuperado 12/04/2024,

de <https://bookdown.org/content/2031/>

Marie-Louise Mares, & Ye Sun. (2010). The multiple meanings of age for television

content preferences. *Human Communication Research*,

36(3)<https://doi.org/10.1111/j.1468-2958.2010.01380.x>

Minitab. (2019). *Análisis de regresión: ¿Cómo puedo interpretar el R-cuadrado y*

evaluar la bondad de ajuste? Recuperado 18/03/2024 de

<https://blog.minitab.com/es/analisis-de-regresion-como-puedo-interpretar-el-r-cuadrado-y-evaluar-la-bondad-de-ajuste>

Robert Stephen Pindyck, & Daniel Lee Rubinfeld. (2009). *Microeconomics* (7th ed.). Pearson.

Rodrigo Briceño. (2000). Principales contribuciones de los premios nobel de economía del año 2000. *Economía Y Sociedad*, 5(14)

<https://repositorio.una.ac.cr/bitstream/handle/11056/19170/Principales%20contribuciones%20de%20los%20premio%20nobel%20de%20econom%c3%ada.pdf?sequence=1&isAllowed=y>

Vldas Griskevicius, & Douglas Kenrick. (2013). How evolutionary needs influence consumer behavior. *Journal of Consumer Psychology*, 23(3), 372–386.

<https://doi.org/10.1016/j.jcps.2013.03.003>

Yago Atrio Lema. (2024). *Modelling tourists' decisionmaking from a cross-cultural perspective* (Economics and Business).

Yesuf Ebrahim Mohammed. (2017). *Factors that determine film genre preference of audiences: The case of three selected government owned cinemas in Addis Ababa*

