

# Proyecto productos de Amazon

Sergio Andrés Villamil Clavijo

## Índice

<b>1. Introducción</b>	<b>1</b>
<b>2. Objetivos</b>	<b>1</b>
2.1. Objetivo general . . . . .	1
2.2. Objetivos específicos . . . . .	1
<b>3. Metodología</b>	<b>2</b>
<b>4. Análisis exploratorio</b>	<b>2</b>
<b>5. Conclusiones</b>	<b>6</b>

## 1. Introducción

Este proyecto analiza una base de datos con información sobre productos de Amazon, donde se incluyen detalles como la categoría, el precio, la cantidad de reseñas, la calificación, un identificador único para cada producto, una variable booleana que indica si es un best seller y un enlace URL a la imagen correspondiente de cada artículo.

## 2. Objetivos

### 2.1. Objetivo general

Explorar y analizar la base de datos con el fin de identificar y comprender las relaciones existentes entre las diferentes variables contenidas.

### 2.2. Objetivos específicos

- Calcular el precio total de los productos de Amazon por categoría, calificación y número de reseñas.
- Encontrar relaciones entre el precio total y la calificación asociada.
- Analizar la relación entre los best sellers y su calificación o número de estrellas.
- Identificar el top 10 de categorías con mayor precio total.
- Determinar las mejores categorías según criterios definidos.

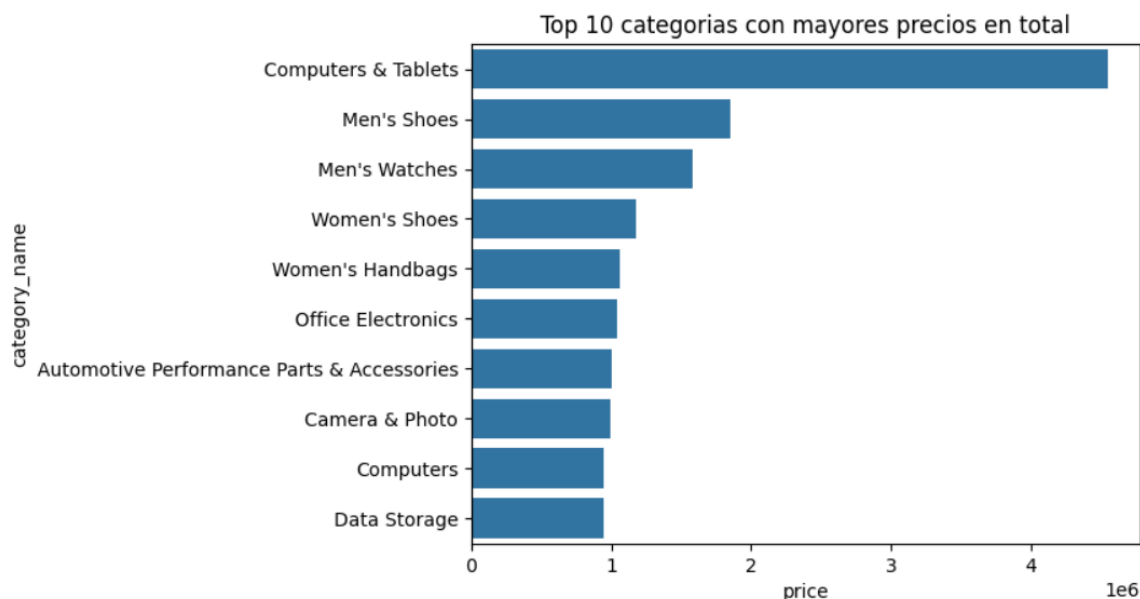


Figura 1: Categorías con mayores precios en total

### 3. Metodología

La base de datos fue obtenida a través de la plataforma Kaggle. Amazon es una empresa que ofrece una amplia variedad de productos y servicios. Las herramientas utilizadas en este análisis fueron: Python, SQLite y Power BI. Se llevó a cabo el siguiente proceso con los datos:

1. Limpieza y transformación de los datos con Python.
2. Creación de la base de datos utilizando SQLite.
3. Análisis exploratorio con Python para identificar relaciones y obtener información relevante de los datos.
4. Generación de diversas visualizaciones y tablas mediante Python.
5. Análisis exploratorio a través de sentencias SQL.
6. Diseño del dashboard utilizando Power BI.

### 4. Análisis exploratorio

En la Figura 1 se observa que la categoría de computadores y tablets tiene un precio total superior al de las demás categorías por más de dos millones de dólares, en comparación con la segunda categoría. Sin embargo, para determinar la "mejor categoría" se utiliza el número de reseñas (reviews). Al tabular estos datos, se encuentra que varias categorías cuentan con cero reseñas, como se muestra en la Figura 2, lo que sugiere que la base de datos podría estar incompleta. Por esta razón, se procede a explorar las relaciones entre la categoría y la calificación en estrellas, así como entre la categoría y la cantidad de productos best sellers.

Al analizar la cantidad de best sellers por categoría, se obtiene la tabla mostrada en la Figura 3. Considerando las mayores cantidades, se presenta la tabla de la Figura 4; cabe destacar que en

	category_name	price	reviews
54	Computers & Tablets	4553201.97	0
137	Men's Shoes	1855241.35	3294939
138	Men's Watches	1581441.04	0
242	Women's Shoes	1181527.85	0
240	Women's Handbags	1060822.95	6322447
...	...	...	...
75	Gift Cards	3790.99	0
74	Garment Bags	3303.19	0
192	Smart Home Thermostats - Compatibility Checker	3285.33	0
202	Smart Home: Voice Assistants and Hubs	3232.91	116572
150	Online Video Game Services	870.54	0

248 rows × 3 columns

Figura 2: Relación entre el precio y las reseñas por categoría

	category_name	price	reviews	stars	isBestSeller
205	Sports & Fitness	178143.97	0	4.451996	483
100	Industrial & Scientific	88294.71	0	4.554726	404
10	Automotive Replacement Parts	199199.11	0	4.483158	368
119	Kitchen & Dining	128970.61	0	4.544674	279
207	Sports & Outdoors	77066.10	0	4.499774	259
217	Tools & Home Improvement	57883.34	0	4.462653	240
220	Toys & Games	620199.47	10020872	4.191442	240
239	Women's Clothing	611848.12	0	4.183384	164
152	Outdoor Recreation	102438.17	0	4.456028	150
12	Automotive Tools & Equipment	401365.51	7981819	4.404137	149

Figura 3: Relación entre el precio, las reseñas, las estrellas y la cantidad de bestsellers por categoría, ordenadas según la mayor cantidad de bestsellers

ambas tabulaciones el valor de reseñas es mayormente cero. A partir de estas tablas, se identifican los elementos comunes que ocupan las primeras posiciones en cada una, lo que representa las categorías con mayor calificación en estrellas y mayor cantidad de best sellers, reflejadas en la Figura 5.

Además, se analiza la relación entre la calificación en estrellas y la cantidad de bestsellers por categoría, centrándose en las ocho mejores categorías, como se muestra en la gráfica de la Figura 6. Al ampliar este análisis a todos los datos, no se observa ninguna relación significativa, debido a que varias categorías presentan cero bestsellers y/o calificación en estrellas.

Finalmente, en el dashboard realizado en Power BI se puede separar las diferentes realciones considerando si son o no bestSeller, de donde se obtienen las Figuras 7 y Figura 8.

	category_name	price	reviews	stars	isBestSeller
75	Gift Cards	3790.99	0	4.832374	2
86	Health & Household	11074.05	0	4.567541	54
100	Industrial & Scientific	88294.71	0	4.554726	404
98	Household Supplies	32165.62	0	4.545314	51
119	Kitchen & Dining	128970.61	0	4.544674	279
69	Food Service Equipment & Supplies	133744.68	0	4.522902	50
62	Electrical Equipment	196385.90	0	4.518131	58
167	Power Tools & Hand Tools	142791.78	0	4.502181	48
207	Sports & Outdoors	77066.10	0	4.499774	259
63	Electronic Components	46317.26	0	4.487720	61

Figura 4: Relación entre el precio, las reseñas, las estrellas y la cantidad de bestsellers por categoría, ordenadas según la mayor cantidad de estrellas

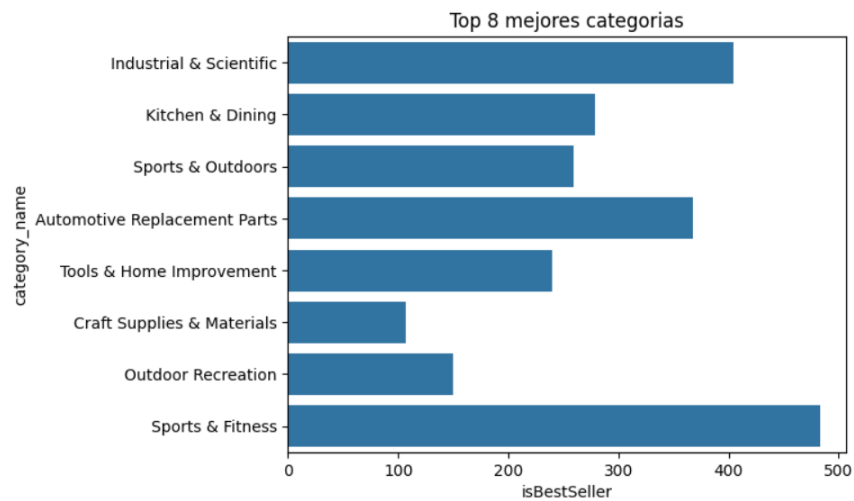


Figura 5: Top ocho de categorías con mayor calificación en estrellas y mayor cantidad de best sellers

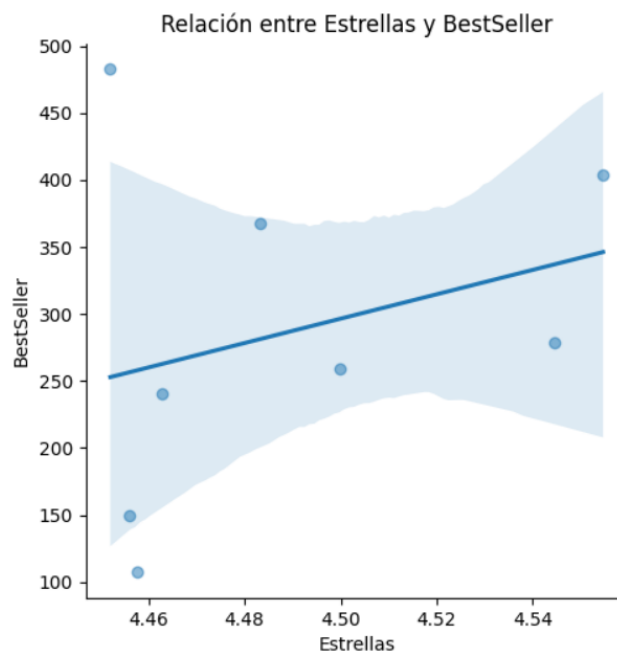


Figura 6: relacion entre la calificacion de estrellas y la cantidad de bestseller por categoría, para las ocho mejores categorías

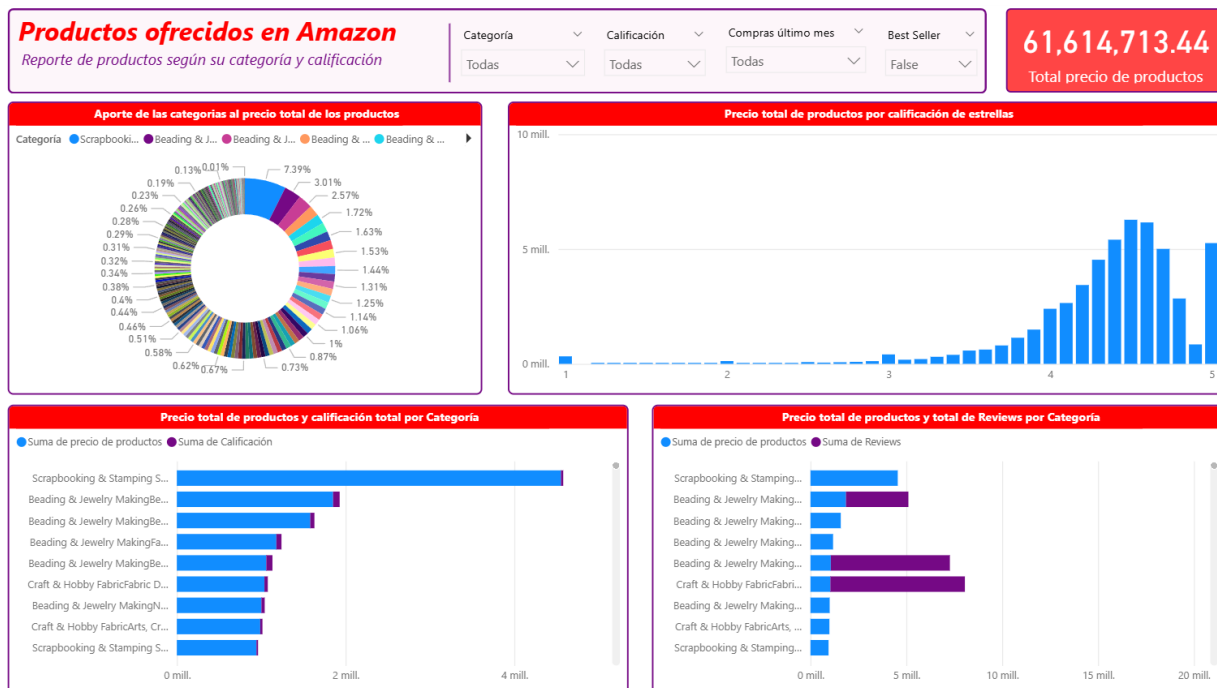


Figura 7: Análisis de diferentes relaciones para las categorías considerando únicamente productos que no son bestseller

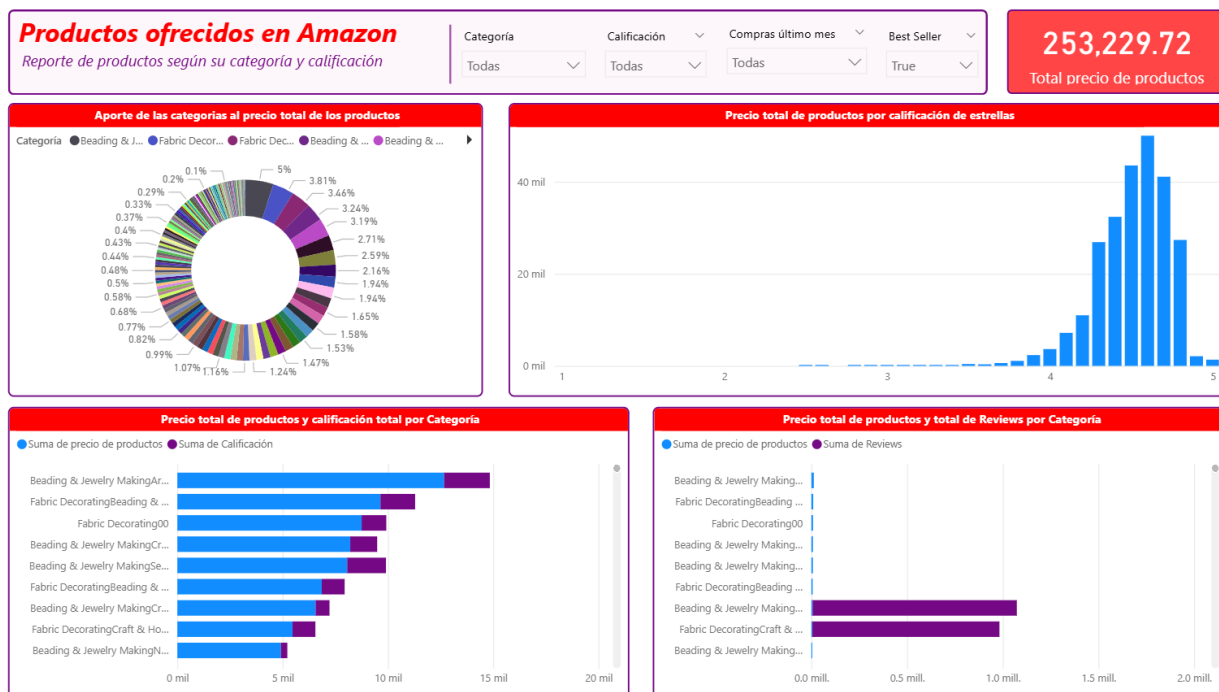


Figura 8: Análisis de diferentes relaciones para las categorías considerando únicamente productos que son bestseller

## 5. Conclusiones

El análisis de los datos reveló que la información utilizada está incompleta. Además, se observó que los productos bestseller suelen tener un promedio de calificaciones principalmente entre 4 y 5 estrellas, mientras que los productos que no son bestseller presentan una distribución más amplia de calificaciones, con un pico notable en las 5 estrellas.