

Análisis de Opiniones de Productos **en Retail/E-commerce**

Identificación de Patrones de Satisfacción y Áreas de Mejora

c19-119-n-data-bi



Team leader : Johanna Tarabelli

Roles del Equipo:

Franco Gabriel Iribarne: Data Scientist Principal / Líder de Proyecto

Maribel Mesa Yepes: Especialista en Visualización de Datos / Científica de Datos

Diego Euler: Ingeniero de Datos / Analista de Datos



Descripción del Proyecto

Objetivo del Proyecto:

Analizar las opiniones de los clientes sobre productos vendidos en una plataforma de e-commerce para identificar patrones de satisfacción del cliente y áreas de mejora.



Contexto:

Este tipo de análisis posee mucho potencial y puede generar gran impacto en áreas como e-commerce, campañas políticas, y análisis de nuevos productos.



Proceso de Análisis de Opiniones de Productos



Herramientas Utilizadas:

Python, Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Jupyter Notebooks.



Exploración de Datos:

Descripción de los Datos:

El dataset `data_review_cleaned.csv` contiene 93,221 filas y 4 columnas:

`rating`: Calificación del producto (1 a 5).

`title`: Título de la opinión.

`review`: Texto de la opinión.

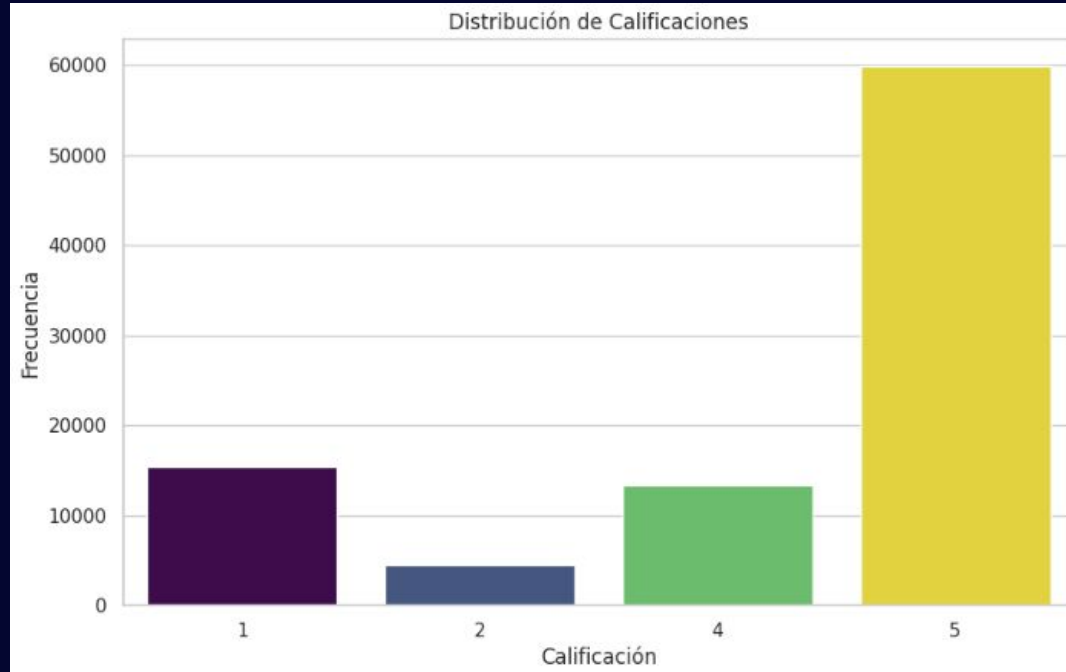
`target`: Variable objetivo (0 para opiniones negativas y 1 para opiniones positivas).

	rating	title	review	target
0	5Am I in Midgard anymore ?	I was so disappointed to find she has not writ...	1.0
1	1	Zip wax new formula is just horrible.	I've been using zip wax for many years. It was...	0.0
2	5	How do you know if you're a narcissist? Alternati...	My narcissist husband has been a narcissist for...	1.0
3	1	Dream of C.	It is to bad such talented illustrations are i...	0.0
4	1	It doesnt suck.	this product doesnt vaccum out the air I was d...	0.0
5	5
99995	4	girls from the big forest	Värttinä first became popular in our country i...	1.0
99996	1	The book keeps on promising to become a 1000 star...	however, it is indeed too general, too much to...	NaN
99997	5	Jogging Down The Road	If I could have the 7 years of The Golden Road...	1.0
99998	1	thou shalt not commit false witness.	I have never written a review, but this book w...	0.0
99999	4	Gorgone	I really enjoyed listening to track 4 the most...	1.0
100000 rows x 4 columns				



Análisis Exploratorio de Datos (EDA):

Distribución de las Calificaciones:

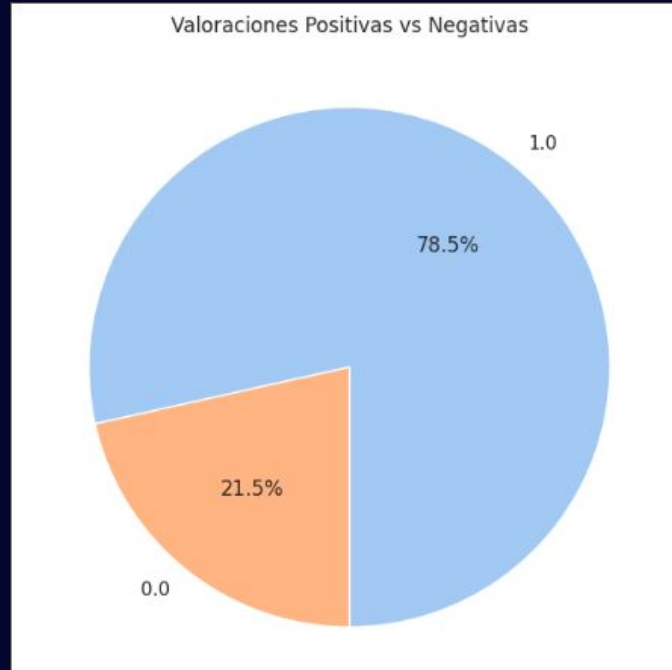


Las calificaciones varían de 1 a 5, con la mayoría de las opiniones concentradas en los extremos (1 y 5).



Análisis Exploratorio de Datos (EDA):

Distribución del Sentimiento (Target):

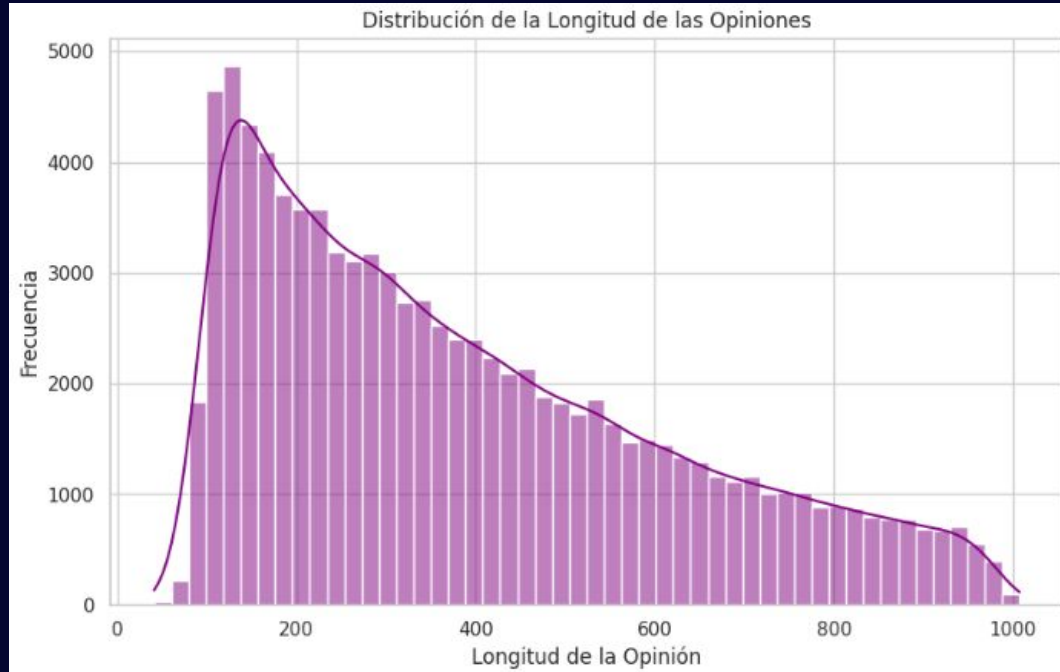


La variable target parece ser binaria, con valores de 0 y 1. La distribución está bastante equilibrada entre opiniones positivas (1) y negativas (0).



Análisis Exploratorio de Datos (EDA):

Longitud de las Opiniones:



La longitud de las opiniones varía ampliamente, con un pico notable en opiniones más cortas (alrededor de 50 a 100 caracteres).



Vectorización del Texto y Modelado de Sentimientos :

Vectorización del Texto: Convertir el texto de las opiniones en vectores numéricos utilizables para modelos de machine learning.

Entrenamiento del Modelo: Utilizar un modelo de machine learning para clasificar las opiniones en categorías de satisfacción (positiva o negativa).



Evaluación Modelo de Sentimientos :

El modelo alcanzó una precisión (accuracy) del 92.3%, con las siguientes métricas detalladas en el reporte de clasificación:

```
Accuracy: 0.923
Classification Report:

```

	precision	recall	f1-score	support
0.0	0.92	0.92	0.92	9315
1.0	0.93	0.93	0.93	12430
accuracy			0.92	21745
macro avg	0.92	0.92	0.92	21745
weighted avg	0.92	0.92	0.92	21745

Precisión (Precision): Proporción de opiniones clasificadas correctamente como positivas o negativas.

Recall: Proporción de opiniones positivas o negativas que fueron correctamente identificadas.

F1-Score: Media armónica de la precisión y el recall, proporcionando una medida equilibrada de la exactitud del modelo.



Interpretación Modelo de Sentimientos :

- La alta precisión del modelo indica que es eficaz para clasificar las opiniones de los clientes.
- Las métricas de precisión, recall y F1-score son similares, lo que sugiere que el modelo no tiene un sesgo significativo hacia opiniones positivas o negativas.
- Este modelo puede ser utilizado para automatizar el análisis de opiniones y ayudar a identificar rápidamente áreas problemáticas y aspectos positivos.



Conclusiones :

- **Patrones de Satisfacción:** La polarización en las calificaciones sugiere que los clientes tienden a tener experiencias muy positivas o muy negativas. Esto puede indicar inconsistencias en la calidad del producto o servicio.
- **Áreas de Mejora:** Se deben analizar las opiniones negativas más comunes para identificar y abordar las principales quejas de los clientes.
- **Potencial de Mejora Continua:** Utilizando el modelo de clasificación, la plataforma puede monitorear continuamente las opiniones y ajustar sus estrategias de acuerdo con los comentarios de los clientes.



Recomendaciones :

- **Monitoreo Continuo:** Implementar el modelo de clasificación para monitorear continuamente las opiniones y detectar problemas emergentes rápidamente.
- **Análisis de Opiniones Negativas:** Realizar un análisis más profundo de las opiniones negativas para identificar causas raíz y tomar acciones correctivas.
- **Fomento de Opiniones Detalladas:** Incentivar a los clientes a proporcionar opiniones más detalladas para obtener insights más específicos y útiles.





¡Gracias por su atención!

