

Reporte

Introducción:

Este reporte final presenta un análisis detallado de diversas categorías de productos vendidos en Amazon, abordando aspectos clave como los precios, descuentos, compromiso del cliente y puntuaciones en reseñas.

En la primera parte, se exploran las diferencias en los precios antes y después de los descuentos en distintas categorías y subcategorías, proporcionando una visión sobre cuáles de ellas resultan más rentables o cuentan con los mayores descuentos.

También se analizan métricas de compromiso del cliente, como el número promedio de reseñas y calificaciones, lo que permite identificar patrones en la satisfacción y fidelidad del cliente hacia ciertos productos.

En la segunda sección del informe, se introduce un modelo de regresión lineal Robusto, diseñado para predecir el precio con descuento a partir del precio real de los productos.

Este modelo, ajustado con errores estándar robustos, ha demostrado un alto nivel de precisión y es evaluado mediante métricas como el coeficiente de determinación (R^2), el error cuadrático medio (MSE) y la media de errores absolutos (MAE). A través de este modelo, se busca optimizar la estrategia de precios en base a predicciones precisas del precio final al consumidor.

Con base en los hallazgos de ambos análisis, el reporte ofrece recomendaciones estratégicas que apuntan a maximizar los ingresos, mejorar las políticas de descuentos y aumentar la satisfacción del cliente en Amazon.

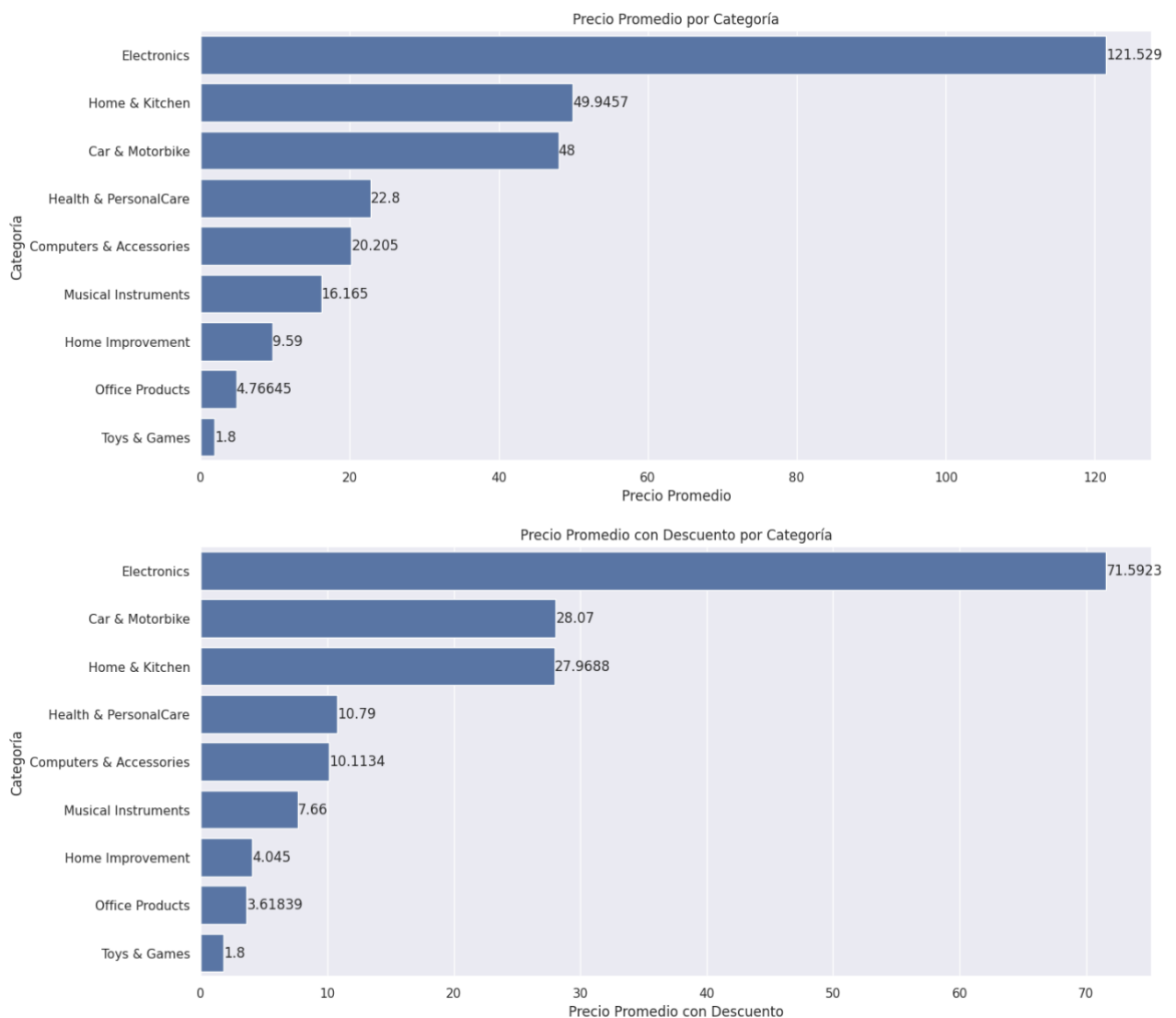
La categoría "Electrónica" tiene el precio medio más alto por producto antes y después del descuento.

El segundo y tercer precio medio más alto por producto antes de los descuentos son **las categorías "Hogar y cocina" y "Coche y moto"**.

En promedio, ambas categorías tienen productos con precios similares.

Después de que se hayan aplicado los descuentos, las categorías "Hogar, cocina" y "Coche, moto" cambian de lugar.

La categoría "Hogar y cocina" puede tener descuentos más altos en promedio en comparación con la categoría "Coches y motos".



Las subcategorías "Ordenadores portátiles" y "Tabletas" tienen el precio promedio más alto por producto antes y después de los descuentos.

Aunque "Ordenadores portátiles" y "Tabletas" tienen el precio medio más alto por producto, la categoría "Ordenadores y accesorios" tiene un pequeño precio medio por producto. Esto se debe a que **las subcategorías "Ordenadores portátiles" y "Tabletas" constituyen un total del 0,14 % de la lista de productos del dataframe.**

Con la adición de "Monitores" (subcategoría con el cuarto precio medio de producto más alto),

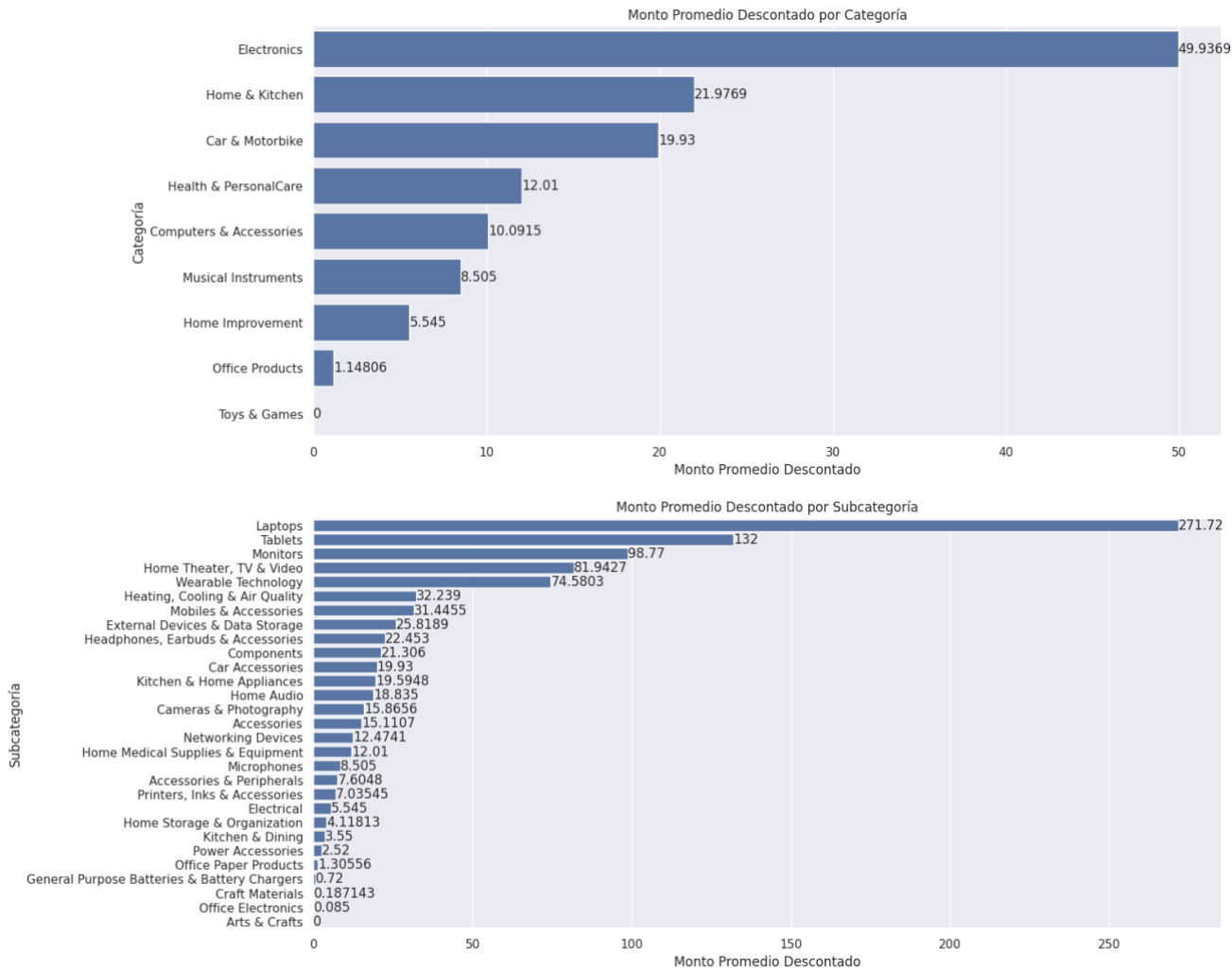
la categoría "Ordenadores y accesorios" solo constituye el 0.28 % de la lista de productos.

Las 3 subcategorías que pertenecen a la categoría "Electrónica" con el precio promedio más alto son "Cine en casa, TV y vídeo", "Móviles y accesorios" y "Tecnología portátil".

Estas subcategorías constituyen el 27,24 % de la lista de productos.

Las 2 subcategorías que pertenecen a la categoría "Hogar y cocina" con el precio promedio más alto son "Calefacción, refrigeración y calidad del aire" y "Cocina y electrodomésticos".

Estas subcategorías constituyen el 28,94 % de la lista de productos.



Descuentos por Categoría:

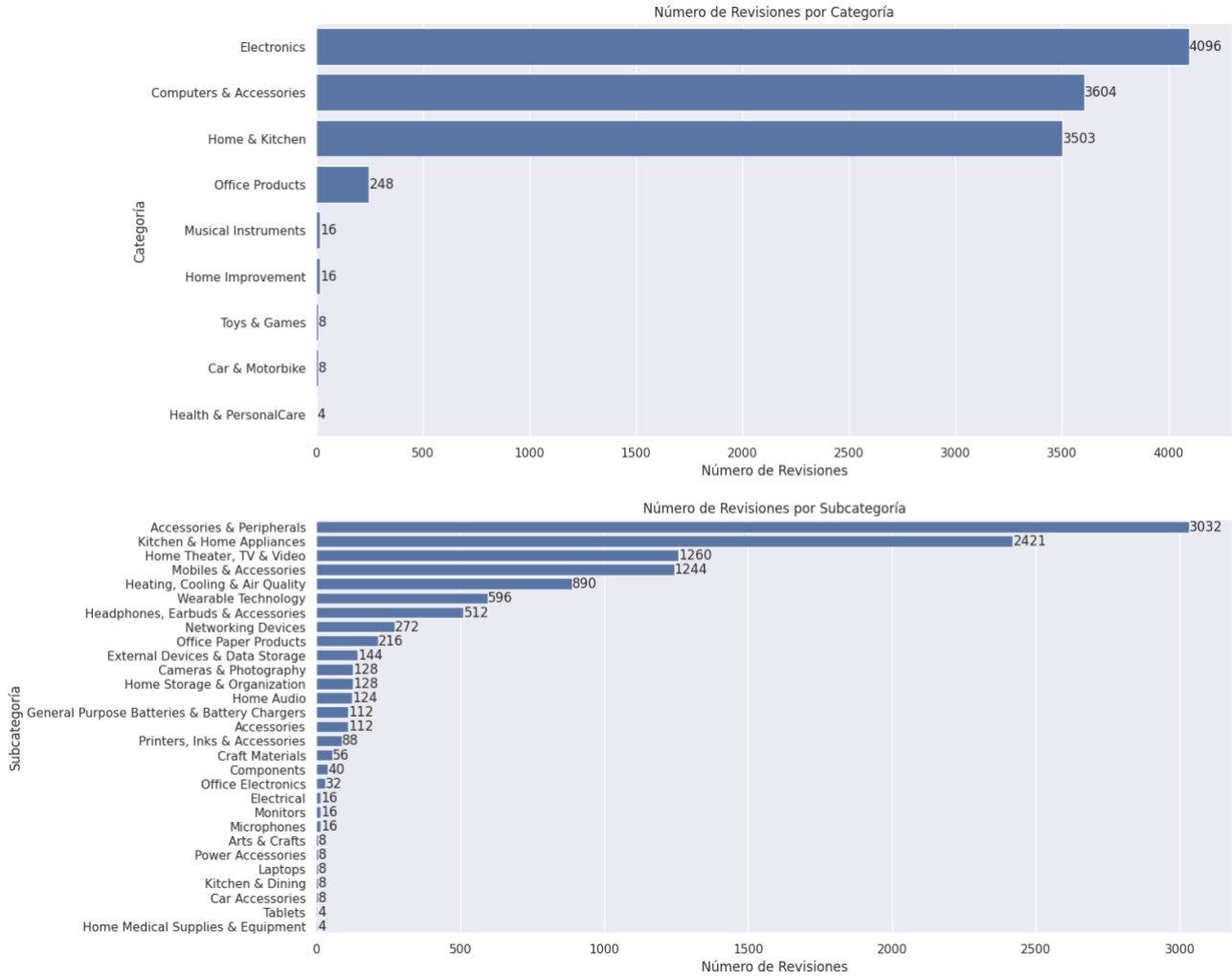
- **Computers&Accessories** ofrece el mayor descuento promedio del **54.02%**.
- **Electronics** ofrece un descuento promedio del **50.83%**.
- La mayoría de los descuentos están en el rango del **40% al 70%**.
- Las subcategorías como **Tecnología portátil, Auriculares y accesorios** tienen los mayores porcentajes de descuento.

Compromiso del Cliente:

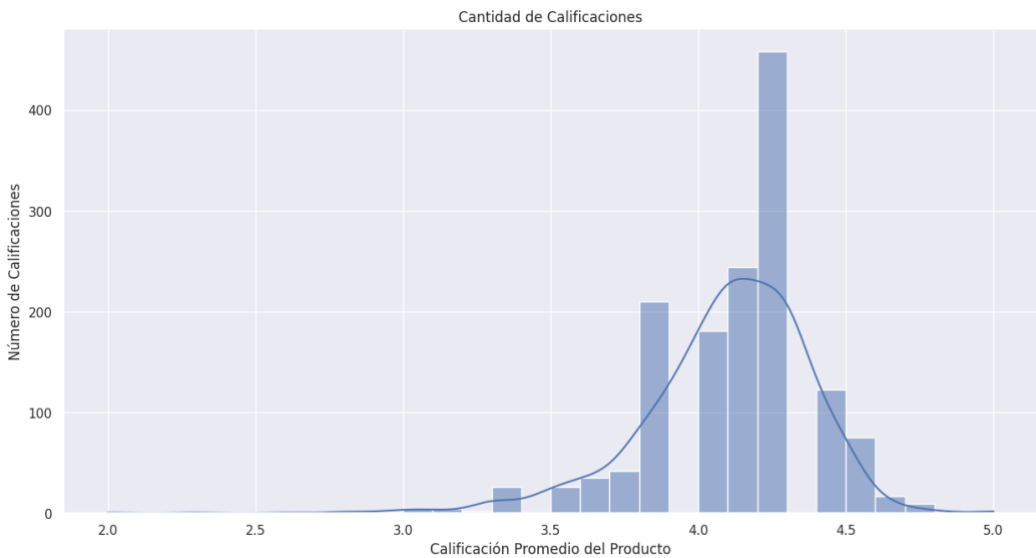
Las categorías con el mayor número promedio de reseñas son las categorías "Electrónica", "Ordenadores y accesorios" y "Hogar y cocina".

Las subcategorías con el mayor número promedio de reseñas son las subcategorías "Accesorios y periféricos" y "Cocina y electrodomésticos".

Este patrón también coincide con el patrón que se muestra en los gráficos de Número de Productos por Categoría y Número de Productos por Subcategoría. El número de reseñas está fuertemente correlacionado con el número de productos diferentes que proporciona una categoría o subcategoría.



- La mayoría de los productos tienen una calificación de **4 a 4.5 estrellas** y los clientes generalmente permanecen anónimos al dejar reseñas.



Recomendaciones

1. Estrategias de Marketing y Promoción:

- Continuar invirtiendo en **Electronics** y **Home&Kitchen**, las categorías más lucrativas, para mantener y aumentar los ingresos.
- Evaluar la efectividad de las políticas de descuento en **Computers&Accessories** y **Electronics** para optimizar la rentabilidad.

Optimización de Descuentos:

- Analizar si los altos descuentos en **Computers&Accessories** y **Electronics** están atrayendo suficiente volumen de ventas para justificar la reducción de precios.
- Experimentar con diferentes estrategias de descuento para encontrar un equilibrio óptimo entre volumen de ventas y márgenes de beneficio.

Compromiso del Cliente:

- Mantener la calidad de productos y servicios en **OfficeProducts** y **Juguetes y juegos**, donde el compromiso del cliente es alto.
- Investigar y mejorar la calidad de productos en **Car&Motorbike** y otras categorías con calificaciones más bajas para aumentar la satisfacción del cliente.

Análisis de Reseñas:

- Realizar análisis de texto de las reseñas para identificar áreas específicas de mejora y oportunidades de innovación.
- Implementar un sistema de retroalimentación que permita a los clientes expresar sus opiniones de manera más efectiva y responder rápidamente a sus necesidades.

El análisis de los datos de ventas de más de 1000 productos vendidos en Amazon ha revelado que **Electronics** y **Home&Kitchen** son las categorías más rentables, mientras que **Computers&Accessories** ofrece los mayores descuentos.

La categoría de **OfficeProducts** tiene la calificación promedio más alta, indicando un fuerte compromiso del cliente. Implementar las recomendaciones mencionadas puede ayudar a maximizar los ingresos, optimizar las políticas de precios y mejorar la satisfacción del cliente en las diversas categorías de productos vendidos en Amazon.

Modelo de Regresión Lineal con errores estándar robustos para la Predicción del Precio con Descuento

```
Número de observaciones en entrenamiento: 1172
Número de observaciones en prueba: 293

Valores de VIF para cada variable predictora (Entrenamiento):
      Feature      VIF
0      const  1.253367
1 actual_price  1.000000

Resumen del modelo de regresión lineal con errores estándar robustos (HC3) - Entrenamiento:
=====
OLS Regression Results
=====
Dep. Variable:    discounted_price    R-squared:            0.927
Model:            OLS                Adj. R-squared:       0.927
Method:           Least Squares       F-statistic:         1191.
Date:             Sat, 02 Nov 2024    Prob (F-statistic):   1.53e-180
Time:             22:05:56            Log-Likelihood:      -5304.8
No. Observations: 1172               AIC:                 1.061e+04
Df Residuals:     1170               BIC:                 1.062e+04
Df Model:          1
Covariance Type:  HC3
=====
              coef      std err          t      P>|t|      [0.025      0.975]
-----
const         -2.8037      0.788      -3.557      0.000      -4.350      -1.257
actual_price    0.6126      0.018     34.505      0.000      0.578      0.647
=====
Omnibus:            387.243    Durbin-Watson:           2.046
Prob(Omnibus):       0.000    Jarque-Bera (JB):       29299.734
Skew:                0.591    Prob(JB):               0.00
Kurtosis:            27.466    Cond. No.               163.
=====

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)

Parámetros del modelo robusto (Entrenamiento):
const          -2.803659
actual_price    0.612577
dtype: float64

Métricas de Evaluación del Modelo en el Conjunto de Prueba:
R²: 0.9186
MSE: 592.8684
RMSE: 24.3489
MAE: 9.5791
Durbin-Watson: 2.0456
```

Introducción

Este reporte presenta el desarrollo y evaluación de un **modelo de Regresión Lineal con errores estándar robustos** diseñado para predecir el **precio con descuento** (discounted_price) de productos en Amazon. El objetivo principal es optimizar la estrategia de precios mediante la predicción precisa del precio final al consumidor basado en el precio real del producto.

El análisis entre actual_price y discounted_price muestra una relación fuerte y lineal. El modelo desarrollado para predecir el precio con descuento (Y) basado en el precio real (X) resulta en la ecuación **Y = 0.613X + -2.804**. Esto indica que **por cada incremento de 1 Euro en el precio real, el precio con descuento aumenta en 0.613 Euros**.

Descripción de los Datos

El análisis se realizó utilizando un conjunto de datos que incluye las siguientes variables:
Se seleccionó una única variable independiente:
actual_price: Precio real del producto en € euros

después de aplicar el descuento (variable objetivo). La variable dependiente es: **discounted_price**: Precio final del producto

El conjunto de datos se dividió en dos subconjuntos:

- **Entrenamiento**: 80% (1,172 observaciones)
- **Prueba**: 20% (293 observaciones)

Evaluación de Multicolinealidad

Se calculó el **Factor de Inflación de la Varianza (VIF)** para asegurar la ausencia de multicolinealidad significativa entre las variables independientes.

Feature	VIF
const	1.253367
actual_price	1.000000

Interpretación: Todos los valores de VIF son inferiores a 5, indicando una **ausencia de multicolinealidad significativa** entre las variables independientes.

Ajuste del Modelo de Regresión Lineal con errores estándar robustos

Se ajustó un modelo de Regresión Lineal con errores estándar robustos utilizando el método de **Mínimos Cuadrados Ordinarios (OLS)** y se aplicaron **errores estándar robustos (HC3)** para manejar posibles heterocedasticidades.

Interpretación:

- **R² (0.9186)**: El modelo explica aproximadamente el 91.86% de la variabilidad en el precio con descuento, lo que indica un **ajuste muy bueno**.
- **MSE (592.8684) y RMSE (24.3489)**: Representan el error cuadrático medio y su raíz, respectivamente, reflejando la precisión de las predicciones del modelo.
- **MAE (9.5791)**: Indica que, en promedio, las predicciones del modelo están a aproximadamente 9.58 Euros del valor real.
- **Durbin-Watson (2.0456)**: Un valor cercano a 2 sugiere **ausencia de autocorrelación** en los residuos.

Diagnóstico del Modelo

Las pruebas realizadas confirman la presencia de **heterocedasticidad** en los residuos. No obstante, el uso de **errores estándar robustos (HC3)** garantiza la validez de las inferencias sobre los coeficientes del modelo.

Además, la estadística de **Durbin-Watson** cercana a 2 respalda la integridad del modelo al indicar **ausencia de autocorrelación** en los residuos.

Conclusiones sobre el Modelo

Basándonos en los resultados del modelo de regresión lineal Robusto, muestra un desempeño estadístico sólido. Los altos valores de R^2 en ambos conjuntos de datos (0.927 en entrenamiento y 0.9186 en prueba) indican que el modelo explica aproximadamente el 92% de la variabilidad en el precio con descuento basándose en el precio real. Esto sugiere que el modelo tiene una buena capacidad predictiva.

El coeficiente de `actual_price` es 0.6126, lo que significa que por cada incremento de 1 € en el precio real, el precio con descuento aumenta en aproximadamente 0.613 €. Este incremento menos que proporcional puede ser interesante desde una perspectiva empresarial, ya que indica que los descuentos no son lineales en relación con el precio original.

Aplicabilidad Empresarial:

- **Estrategias de Precios:** Si una empresa está buscando establecer políticas de descuento basadas en el precio original, este modelo puede ayudar a predecir cómo ajustar los descuentos en función de los cambios en los precios.
- **Segmentación de Mercado:** Comprender esta relación puede ser útil para segmentar productos o clientes según su sensibilidad al precio y optimizar las ofertas promocionales.
- **Análisis de Competitividad:** El modelo puede servir para analizar cómo los precios con descuento se comparan con los de la competencia, permitiendo ajustar estrategias para mejorar la posición en el mercado.

Consideraciones Adicionales:

- **Variables Omitidas:** Si bien el modelo muestra un buen ajuste, podría beneficiarse de la inclusión de otras variables que afecten el precio con descuento, como promociones especiales, estacionalidad, o características del producto.
- **Validación Continua:** Es importante mantener una validación continua del modelo para asegurarse de que sigue siendo relevante en condiciones cambiantes del mercado.

El resultado obtenido ofrece una visión valiosa que puede ser aplicada en la toma de decisiones empresariales. Dado su alto poder explicativo y la

relevancia del coeficiente obtenido, el modelo puede ser una herramienta útil para optimizar estrategias de precios y descuentos.

Modelo Guardado

El modelo final ha sido **guardado** en el archivo model.pkl para su uso futuro en predicciones o integraciones en sistemas de producción.