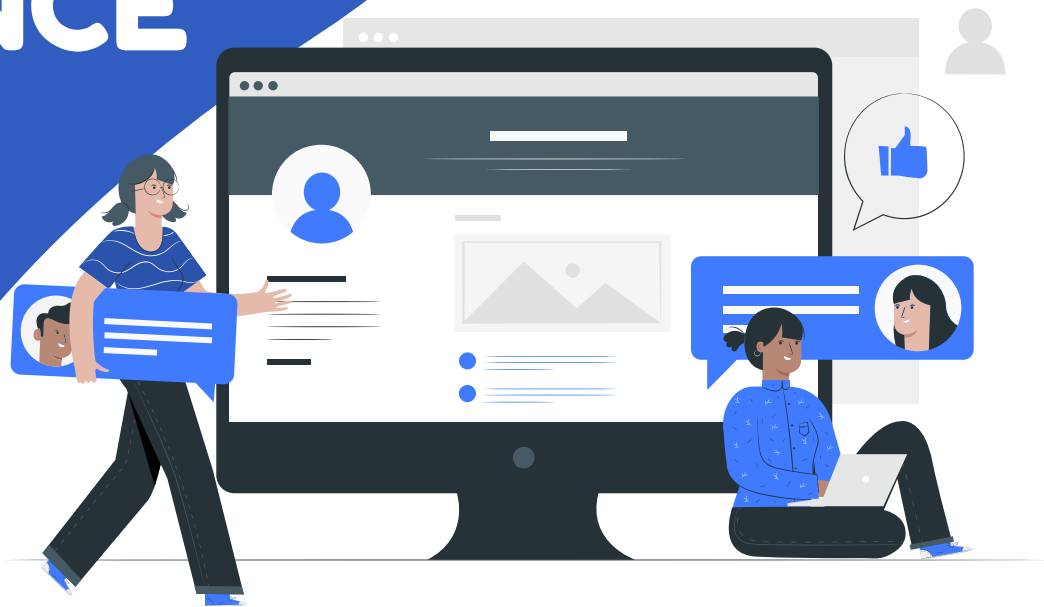


# PROYECTO DATA SCIENCE

Silva Franco - CoderHouse



**En el siglo XXI, nuestras vidas están entrelazadas con la tecnología, y en el corazón de esta revolución digital se encuentran las laptops. Estos dispositivos no solo son herramientas, sino socios en nuestro día a día, transformando cómo trabajamos, aprendemos y nos conectamos.**



# Contenidos

**01**

**Introducción**

**02**

**Desafío**

**03**

**Estrategia**

**04**

**Nuestros datos**

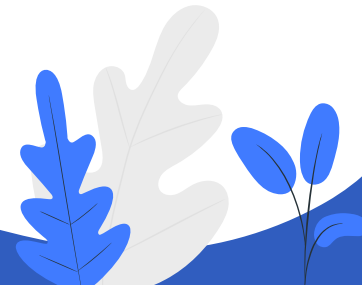
**05**

**Metodología**

**06**

**Qué sigue**

# 01 Introducción



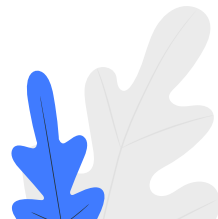
# Proyecto: Laptops

Esta presentación explora el análisis de datos de laptops con el objetivo de entender las variables que afectan el precio. Al identificar patrones y tendencias, buscamos proporcionar información valiosa para consumidores, fabricantes y entusiastas de tecnología. Descubriremos cómo estos insights pueden beneficiar a compañías y aficionados de la tecnología.

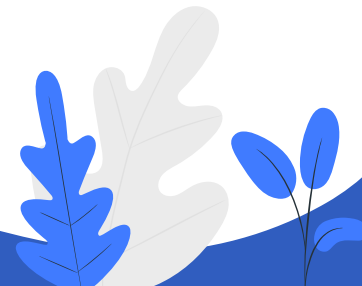


Esta presentación se sumerge en el fascinante mundo del análisis de datos de laptops, explorando las motivaciones detrás de la creciente importancia de estos dispositivos en nuestras vidas diarias. Desde el auge del trabajo remoto hasta la revolución en la educación en línea, la toma de decisiones informadas sobre laptops se ha vuelto esencial. Nuestro análisis busca desentrañar las complejidades de las especificaciones de laptops y su impacto en las decisiones de compra.

Dirigido a una audiencia diversa, desde consumidores hasta fabricantes y analistas de datos, el objetivo es proporcionar insights valiosos que transformen la manera en que percibimos y elegimos laptops. Más allá de la mera relación entre especificaciones y precios, estamos explorando cómo este análisis puede inspirar el desarrollo futuro de tecnologías, influir en estrategias de mercado y contribuir a la evolución continua de la industria de laptops.



## 02 Desafío



# Las dudas luego de obtener una laptop



## Elección

¿Es esta la mejor que encontré?



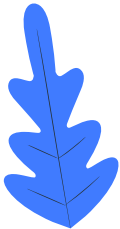
## Precio

¿La obtuve al mejor precio?



## Expectativas

¿Cumple con mis requisitos diarios?





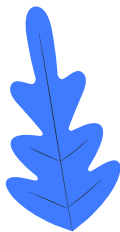
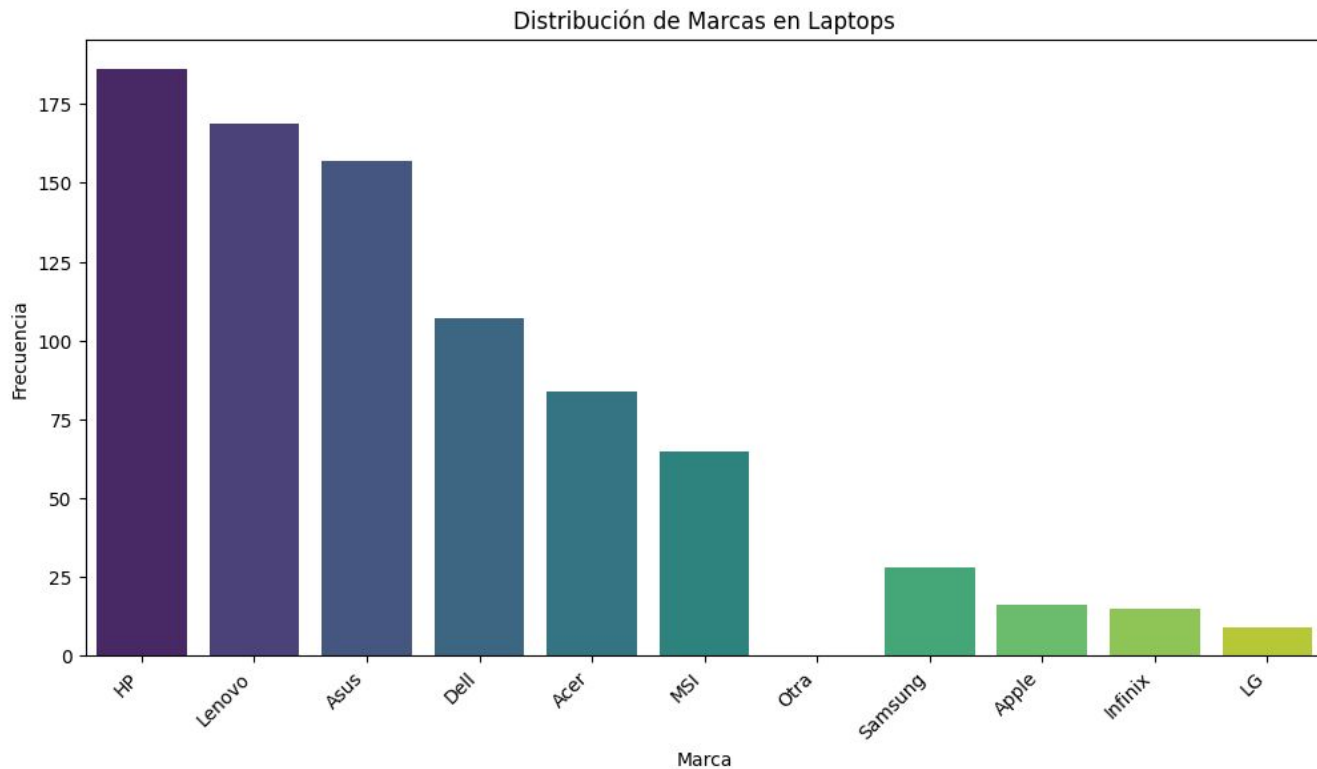
Guiados por la curiosidad, exploramos preguntas clave como

- "¿Cuáles son las especificaciones técnicas más influyentes en el precio de una laptop?"
- "¿Cómo varía el precio en función de la potencia del procesador?"
- "¿Existe una relación clara entre la capacidad de almacenamiento y el precio de las laptops?"
- "¿Influyen factores como la memoria RAM en la percepción del valor por parte de los consumidores?"
- "¿Cómo se distribuyen las preferencias de los consumidores en términos de marcas de laptops?"
- "¿Cómo se relacionan las características técnicas con las preferencias de los consumidores?"

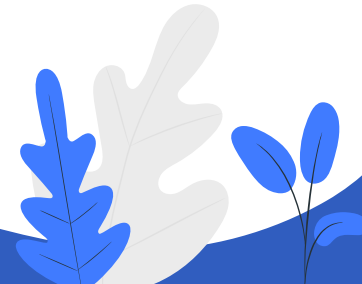
Estas mejoras buscan agregar claridad, detalles y un tono más orientado a la audiencia para que la presentación sea más efectiva y cautivadora. Recuerda adaptar los textos según la realidad específica de tu proyecto.



# ¿Cuáles son las marcas más frecuentes?



## 03 Estrategia

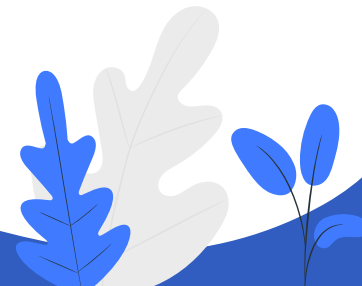


# Modelo de predicción

Con las habilidades obtenidas a lo largo del curso seremos capaces de predecir en base a sus especificaciones el precio de los equipos.



# **04 Nuestros Datos**



# DataFrame 'laptop'

Contamos con una amplia base de datos de diversos tipos de laptops en las cuales disponemos de sus especificaciones técnicas como lo son:

- Modelo, Marca
- Almacenamiento, CPU, RAM
- Medidas, Resolución
- PRECIO

# Nuestro DataFrame

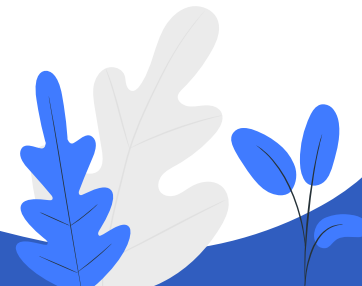
laptop.shape  
(893, 18)

	brand	name	price	spec_rating	processor	CPU	Ram	Ram_type	ROM	ROM_type	GPU	display_size	resolution_width	resolution_height	OS	warranty
0	HP	Victus 15-fb0157AX Gaming Laptop	49900	73.000000	5th Gen AMD Ryzen 5 5600H	Hexa Core, 12 Threads	8GB	DDR4	512GB	SSD	4GB AMD Radeon RX 6500M	15.6	1920.0	1080.0	Windows 11 OS	1
1	HP	15s-fq5007TU Laptop	39900	60.000000	12th Gen Intel Core i3 1215U	Hexa Core (2P + 4E), 8 Threads	8GB	DDR4	512GB	SSD	Intel UHD Graphics	15.6	1920.0	1080.0	Windows 11 OS	1
2	Acer	One 14 Z8-415 Laptop	26990	69.323529	11th Gen Intel Core i3 1115G4	Dual Core, 4 Threads	8GB	DDR4	512GB	SSD	Intel Iris Xe Graphics	14.0	1920.0	1080.0	Windows 11 OS	1
3	Lenovo	Yoga Slim 6 14IAP8 82WU0095IN Laptop	59729	66.000000	12th Gen Intel Core i5 1240P	12 Cores (4P + 8E), 16 Threads	16GB	LPDDR5	512GB	SSD	Intel Integrated Iris Xe	14.0	2240.0	1400.0	Windows 11 OS	1
4	Apple	MacBook Air 2020 MGN03HN Laptop	69990	69.323529	Apple M1	Octa Core (4P + 4E)	8GB	DDR4	256GB	SSD	Apple M1 Integrated Graphics	13.3	2560.0	1600.0	Mac OS	1

Ejemplo de registro:

Marca, Nombre del modelo, Precio, Calificación de las especificaciones, Procesador, Detalles del CPU, RAM, Tipo de RAM, Almacenamiento, Tipo de almacenamiento, Gráficos, Pulgadas de pantalla, Resolución ancho, Resolución alto, Sistema operativo y Garantía.

# 05 Metodología





# Cómo desarrollamos el modelo

## EDA

Exploración de Datos y análisis



## Entrenamiento

Modelos de Machine Learning para hallar patrones



## Validación

Ajuste y puesta en marcha del modelo

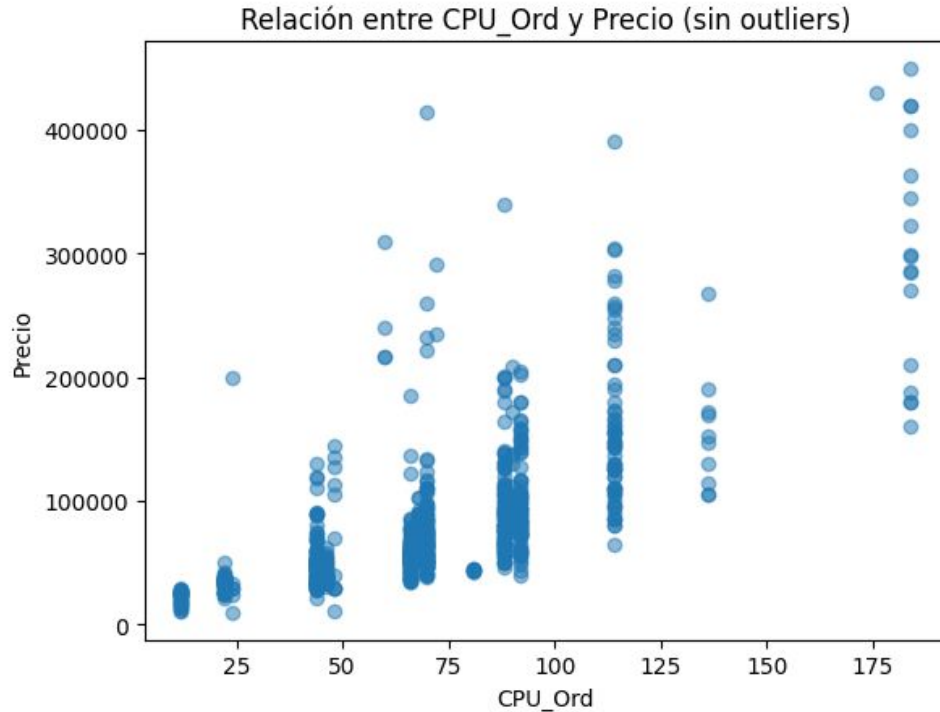


## Preprocesamiento

Lidiar con datos faltantes, conversiones



# EDA



Por ejemplo, gracias a uno de los pasos en el proyecto, el EDA, descubrimos que el tipo de CPU, sus hilos y núcleos, se encuentra estrictamente relacionado con el precio del producto.

# Preprocesamiento

## Transformaciones

```
# Droppeo IDs innecesarios
laptop = laptop.drop(columns = ['Unnamed: 0.1', 'Unnamed: 0'])

# RAM -> int
laptop['RAM_int'] = laptop['Ram'].str.extract(r'(.+?)GB').astype(int)

# ROM -> tb a gb
laptop['ROM'] = laptop['ROM'].replace({'1TB': '1000GB', '2TB': '2000GB'})

# ROM -> int
laptop['ROM_int'] = laptop['ROM'].str.extract(r'(.+?)GB').astype(int)

# ROM_type -> bool
laptop['ROM_is_SSD'] = laptop['ROM_type'].map({'SSD': 1, 'Hard-Disk': 0})

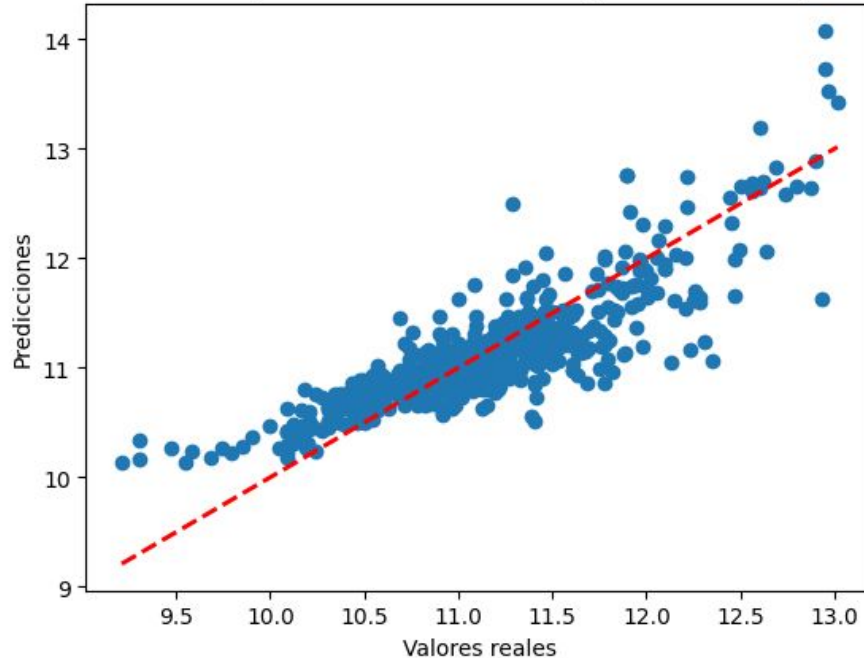
# Uniformización de Brands
# Calculo su frecuencia
frec_marcas = laptop['brand'].value_counts()
porc_marcas = frec_marcas / len(laptop) * 100
resultados_marcas = pd.DataFrame({'Categoría': frec_marcas.index,
                                  'Frecuencia': frec_marcas.values,
                                  'Porcentaje': porc_marcas.values})

# Me quedo con el TOP 10 (>1%) y agrupo las restantes en 'Otra'
marcas_principales = resultados_marcas[resultados_marcas['Porcentaje'] > 1]['Categoría'].tolist()
laptop['Brand_Ord'] = laptop['brand'].apply(lambda x: x if x in marcas_principales else 'Otra')
```

Gracias a varias transformaciones, hemos llegado a un dataset mucho más limpio y ordenado para ponernos manos a la obra.

# Entrenamiento

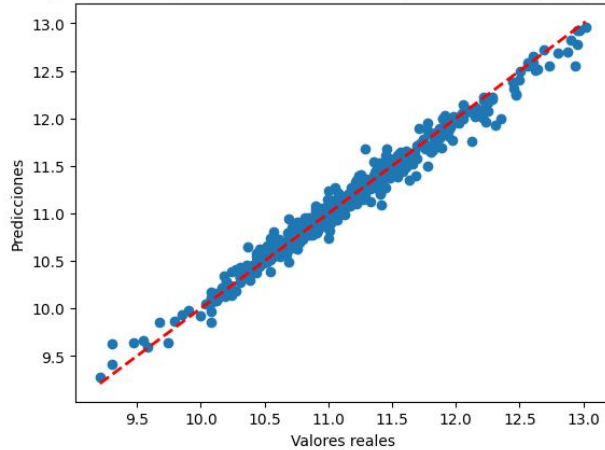
Regresión Lineal - Comparación de Predicciones y Datos Reales (Entrenamiento)



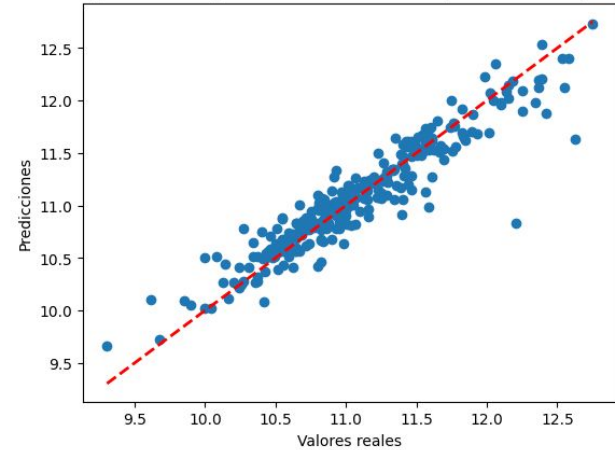
Uno de los modelos empleados en este proyecto fue el de Regresión Lineal, el cual tuvo resultados buenos, pero perfectamente mejorables.

# Validación

Random Forest Regressor - Comparación de Predicciones y Datos Reales (Conjunto de Entrenamiento)

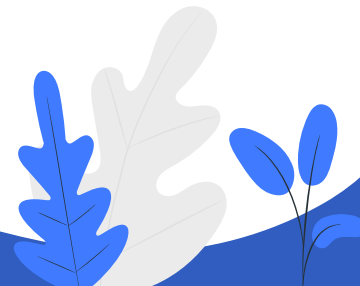


Random Forest Regressor - Comparación de Predicciones y Datos Reales (Conjunto de Prueba)



Comprobamos que las predicciones de los modelos sean similares entre muestras de entrenamiento y testeo

## 06 Qué sigue?



# Puesta en marcha



En este proyecto, hemos explorado diferentes enfoques de modelado para predecir los precios de laptops. Después de un análisis exhaustivo, hemos decidido utilizar el modelo de Random Forest Regressor como nuestra principal herramienta de predicción.

El modelo de Random Forest Regressor es una técnica de aprendizaje automático poderosa y versátil que se basa en el concepto de "ensamblaje de árboles de decisión". Utiliza múltiples árboles de decisión para hacer predicciones y combina sus resultados para obtener una predicción más robusta y precisa.

# Resultados del Modelo

El modelo Random Forest Regressor fue seleccionado como el enfoque principal para predecir los precios de laptops debido a sus excelentes resultados durante la fase de entrenamiento y evaluación.

**Precisión del Modelo:** El modelo logró una alta precisión en la predicción de los precios de las laptops, como se refleja en las métricas de evaluación obtenidas, tales como el bajo Error Absoluto Medio (MAE) y el alto Coeficiente de Determinación ( $R^2$ ).

**Flexibilidad y Robustez:** El Random Forest Regressor es conocido por su capacidad para manejar conjuntos de datos complejos con múltiples características. En nuestro caso, el modelo demostró su flexibilidad al adaptarse eficazmente a la variabilidad en los precios de las laptops, que pueden estar influenciados por una amplia gama de factores.

**Resistencia al Overfitting:** A pesar de su complejidad inherente, el modelo Random Forest Regressor mostró una resistencia notable al sobreajuste (overfitting), lo que significa que pudo generalizar bien a partir de los datos de entrenamiento para hacer predicciones precisas en datos no vistos.

**Versatilidad y Escalabilidad:** El enfoque de Random Forest es altamente versátil y escalable, lo que lo hace adecuado para conjuntos de datos de diferentes tamaños y complejidades. Además, la capacidad de paralelización del algoritmo permite un procesamiento eficiente de grandes volúmenes de datos.

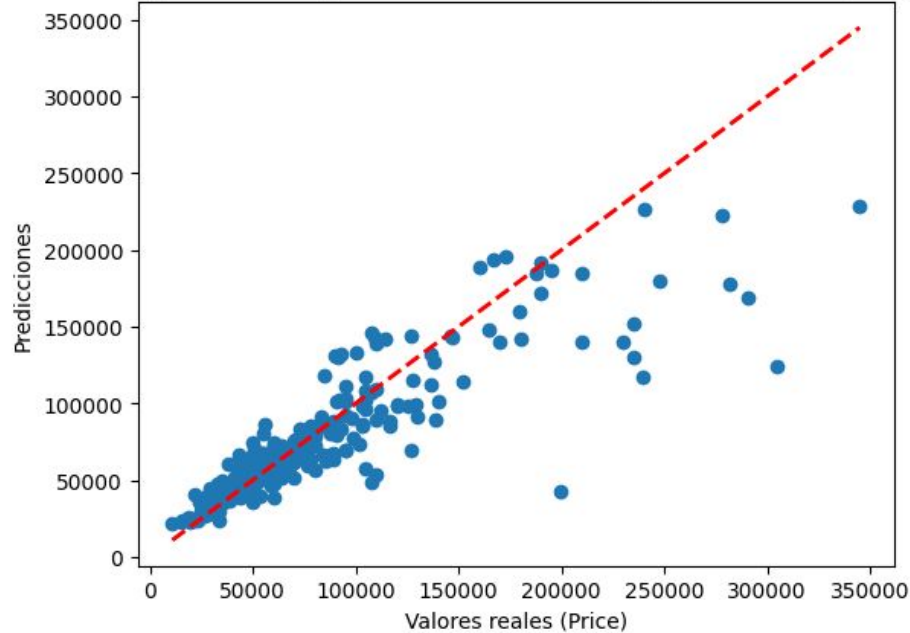
**Interpretación de Resultados:** A través de técnicas como la importancia de las características, podemos entender mejor qué características contribuyen más a las predicciones del modelo, lo que proporciona información valiosa para la toma de decisiones.



# Comparación de resultados

**74%**  
Precisión

Random Forest Regressor - Comparación de Predicciones y Datos Reales (Conjunto de Prueba)



En este proyecto, hemos desarrollado un modelo de regresión utilizando Random Forest para predecir los precios de laptops. Este modelo ha demostrado ser una herramienta eficaz para estimar los precios de productos tecnológicos basados en sus especificaciones.

Las predicciones generadas por nuestro modelo podrían ser de gran utilidad en diversos escenarios, como:

- Industria minorista: Ayudar a los minoristas a fijar precios competitivos y optimizar sus estrategias de precios.
- Comercio electrónico: Facilitar a los consumidores la toma de decisiones de compra informadas al ofrecer estimaciones de precios precisas en línea.
- Fabricantes de tecnología: Apoyar en la planificación de precios y el desarrollo de nuevos productos al comprender mejor la relación entre las especificaciones y los precios.

En resumen, nuestro modelo de predicción de precios de laptops ofrece una herramienta valiosa para una amplia gama de aplicaciones en la industria minorista, el comercio electrónico y la fabricación de tecnología.

