



PRESENTACIÓN PROYECTO

Venta de Automoviles -
Mercado Libre

Alumno: Gabriel Cuenca

Profesor: Norberto González

Tutor: Álvaro Galindo

Tabla de contenidos:

Presentación de la empresa -----	3
Definición del problema -----	4
Objetivo y preguntas de la investigación -----	5
Data Acquisition -----	6
Variables -----	7
Decisiones tomadas -----	8
Análisis Univariable -----	9
Análisis Bivariable -----	10
Predicción del precio justo del vehículo -----	11
Evaluación de los resultados -----	12

Presentación de la empresa

Mercado Libre es un marketplace, es decir, un sitio de compra y venta que funciona como un centro comercial virtual. Al principio de sus actividades, Mercado Libre atendía a vendedores únicos u ocasionales que querían ofrecer algo que ya no necesitaban.

Fundada por Marcos Galperín, Mercado Libre es la compañía más valiosa del país y la única latinoamericana en este Top 100. Además, es la primera nacida en Argentina en acceder al ranking Kantar BrandZ desde que fue publicado por primera vez en 2006.



Definición del problema

En un contexto de distorsión de precios en Argentina, a veces representa un problema importante determinar el precio justo en la venta de un automóvil.

Por lo tanto, el modelo ayudará tanto a los vendedores como a la empresa, asignando al vehículo un precio justo en base a las características principales del mismo, garantizando una venta más rápida.



Objetivo y preguntas de la investigación

Objetivo:

- Construir un modelo de regresión basado en datos para determinar el precio justo de venta de un automóvil.

Preguntas de la investigación:

- ¿Qué datos son necesarios?
- ¿Qué variables están correlacionadas?
- ¿Cuáles son las variables que van a determinar el precio?

Data Acquisition

Este dataset contiene datos reales de la oferta de vehículos en Argentina.

El dataset posee 510 filas, cada una de las cuales es un vehículo publicado en la plataforma, con datos del año 2022. El dataset contiene además 12 variables entre las que se describen las características del vehículo, y su valor, ya sea en pesos o en dólares.



Variables

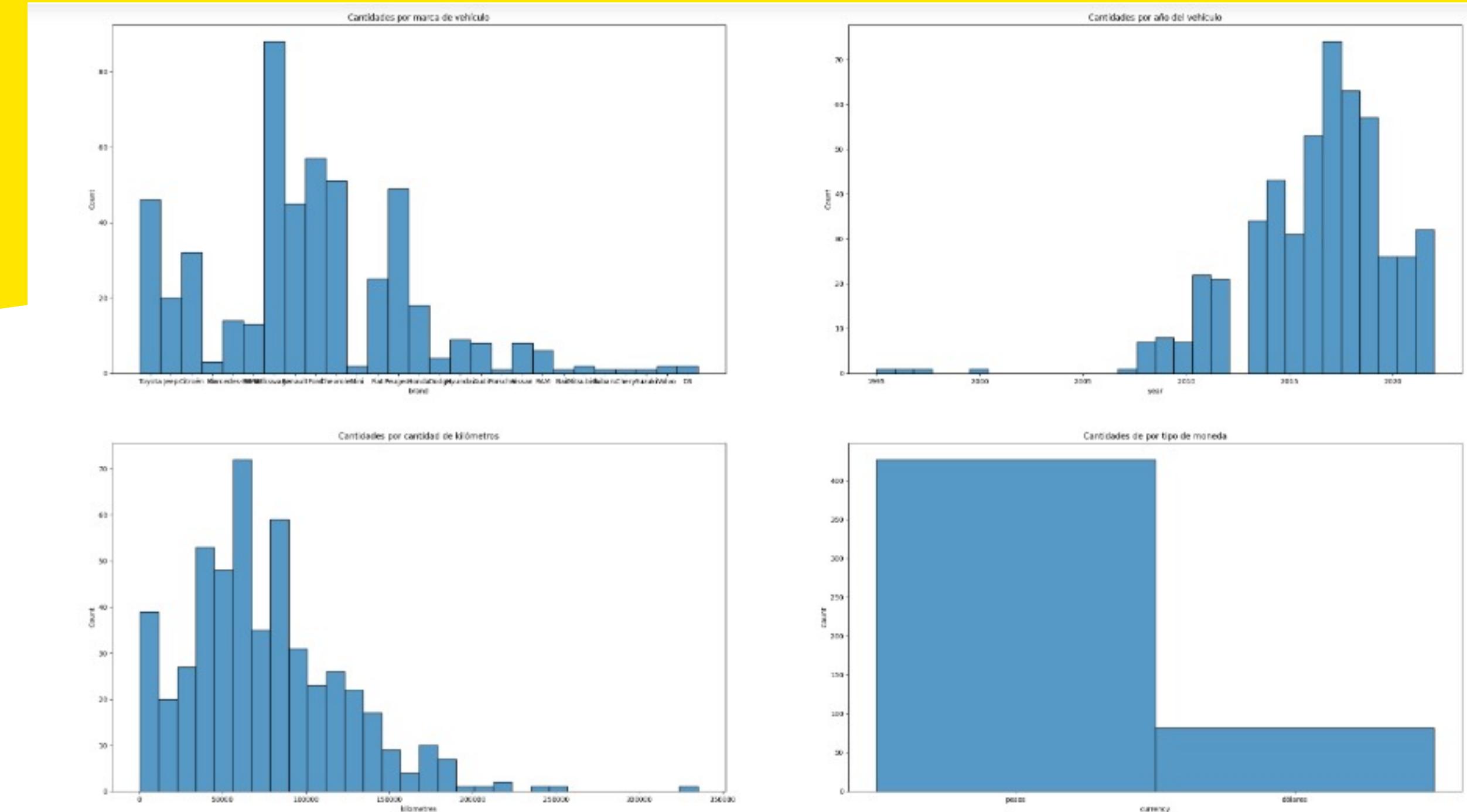
- Money: Valor del vehículo (puede ser en pesos o en dólares)
 - Brand: Marca del vehículo
 - Model: Modelo del vehículo
 - Year: Año de fabricación del vehículo
 - Color: Color del vehículo
- Fuel_type: Tipo de combustible del vehículo
- Door: Cantidad de puertas del vehículo
- Gear: Transmisión del vehículo
- Long: Motor del vehículo
- Body_type: Tipo de Carrocería
- Kilometres: Cantidad de kilómetros del vehículo
- Currency: Tipo de cambio (Pesos o dólares)

Decisiones tomadas

- Se eliminó la variable 'color', por no ser determinante en el precio del vehículo.
- Ya que habían publicaciones en dólares y otras en pesos, se creó columna 'precio' para poder unificar el valor del vehículo en pesos, asignando un valor al dólar (que puede ser modificada).
- Se completó la mayor cantidad de nulos posibles (incluso de forma manual), ya que se trataba de una base pequeña.
- Se eliminaron outliers (como precios muy por encima del mercado, y cantidad de kilómetros del vehículo), para evitar distorsionar los resultados del modelo.

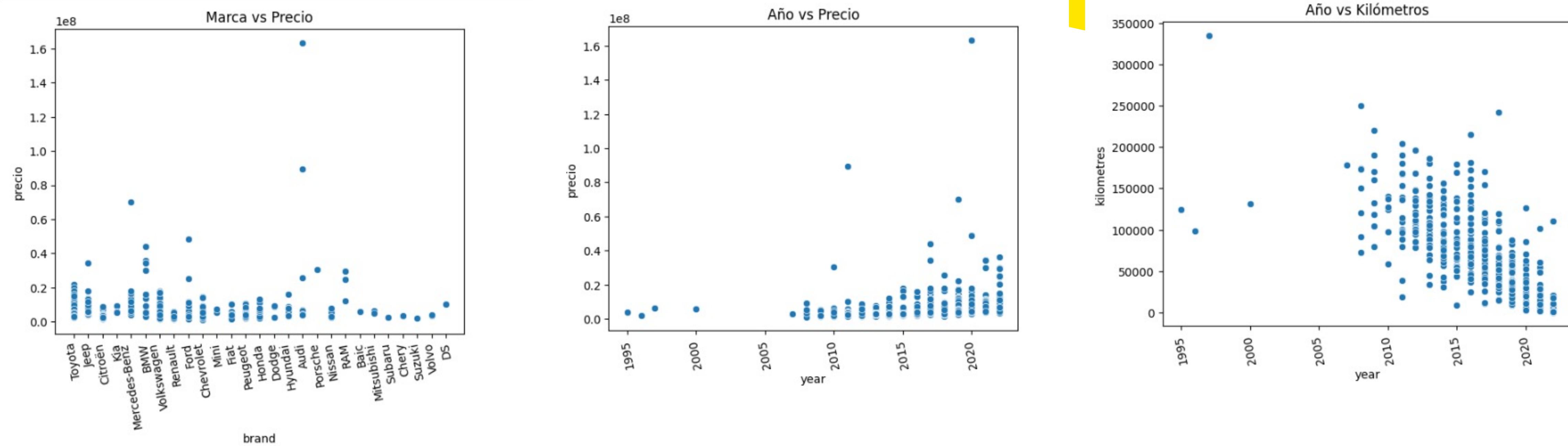


Análisis Univariable



Se analizó las cantidades por marca, por año y por cantidad de kilómetros del vehículo, y por tipo de moneda de la publicación.

Análisis Bivariante



Se analizó la cantidad de kilómetros vs año del vehículo, año vs precio, y marca vs precio. Quitando alguno outliers, hay una clara relación entre las variables.

Predictión del precio justo del vehículo

Performance de los algoritmos

Algoritmo	R2	MSE	MAE
AdaBoost	0.7254650989732581	579179360401,552	1212782,56880733
GradientBoosting	0.7894752860815418	4441386821857,490	888256,31217884
LightGBM	0.588901516940091	8672841070122,940	1572645,07711008
XGBoost	0.8523002507209745	3115984379361,760	756576,5

Evaluación de los resultados

De acuerdo a las métricas, el XGBoost resulta el mejor modelo a seguir para este caso de regresión.

r2: 0.8523002507209745

MSE: 3115984379361.7656

MAE: 756576.5