

# Proyecto Data Science:

## AUTOMOTORA ANACONDA

### Hito 2

---

**Integrantes:**

- Paulo González
- Jorge Guerrero
- Daniel Mardones
- Gonzalo Rojas
- Esteban Sánchez

**Profesor:**

- Daniel Beffermann

**Tutor:**

- Jhon Poma



Jueves 02 de febrero 2023



## Resumen Hito 1



### Motivación

Generar herramienta de apoyo a la toma de decisiones de los equipos de compra y venta de vehículos.



### Hipótesis

Determinar mediante un modelo de regresión en base a las especificaciones de un vehículo si su compra es viable para realizar negocios de reventa.



### Propuesta inversion

Comprar vehículos que tengan un precio estimado menor al precio de mercado para poder revenderlos.



### Segmentar gama de vehículos

Columna que clasifique la marca del vehículo por generalista y premium.



# Resumen Hito 1



## Modelo

Modelo de **regresión** con técnicas de Machine Learning



## Vector objetivo

Variable **“Price”**, que muestra el precio en USD de cada vehículo.



## Métricas

R2 score, Median Absolute Error (MAE), Root Mean-Squared Error (RMSE)

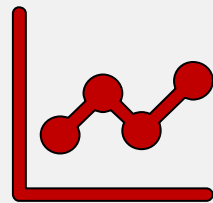


## Enriquecimiento data set

A través de webscrapping aumentar la información de los vehículos.

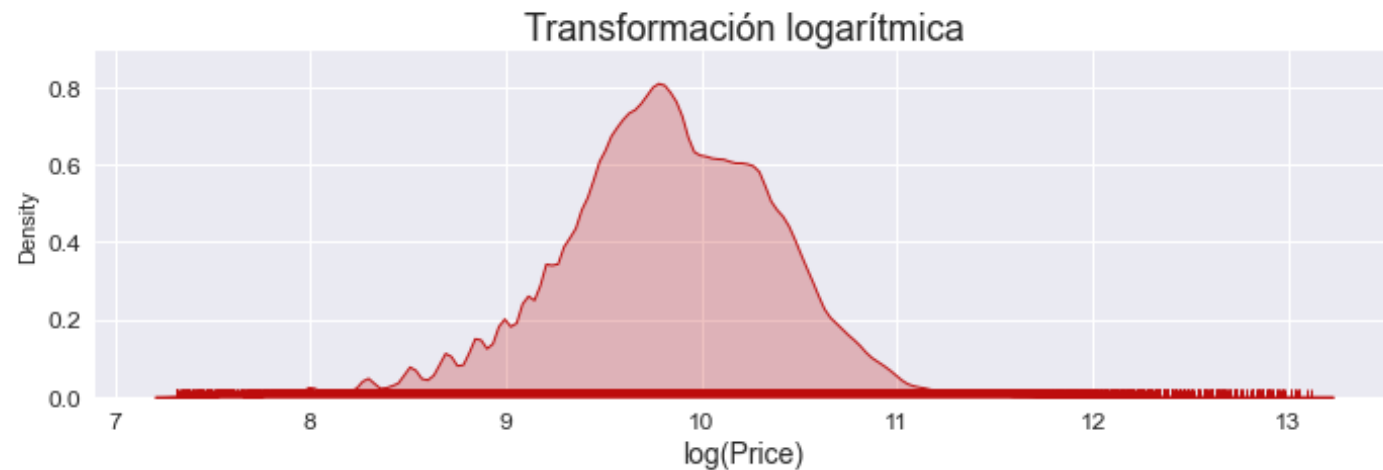
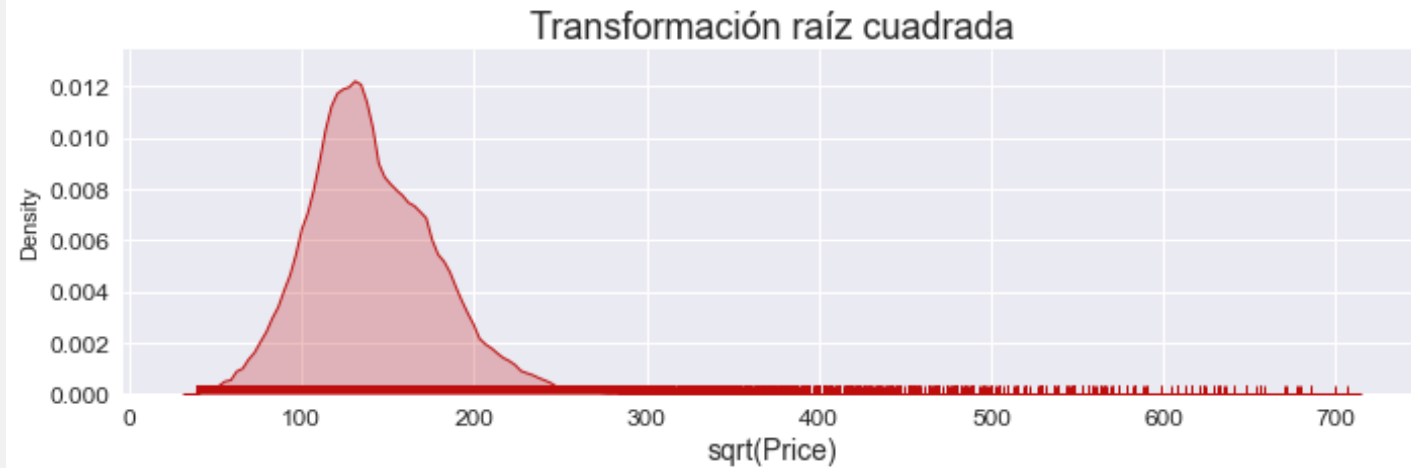
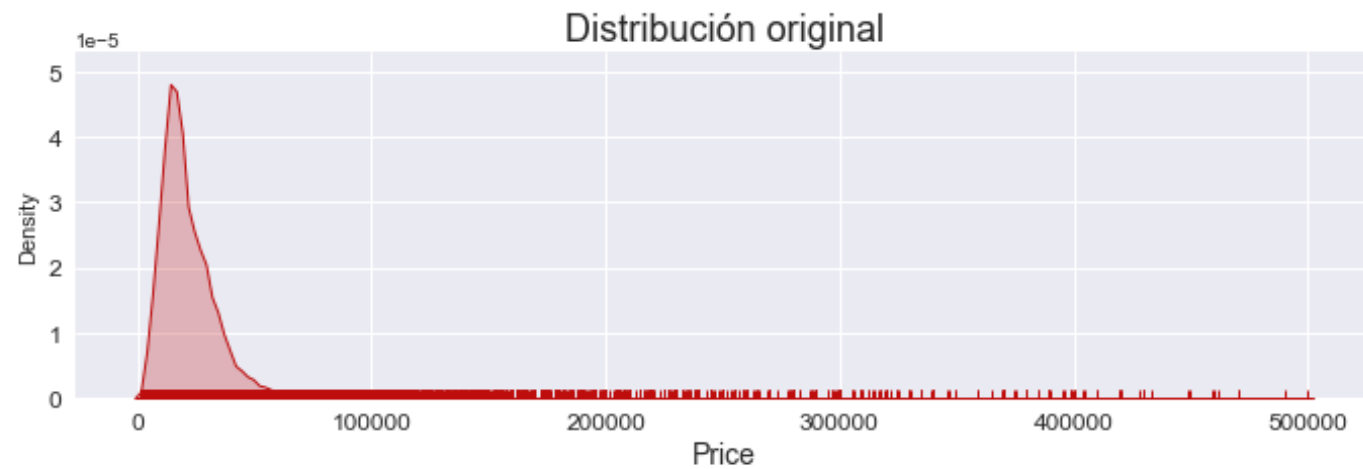
---

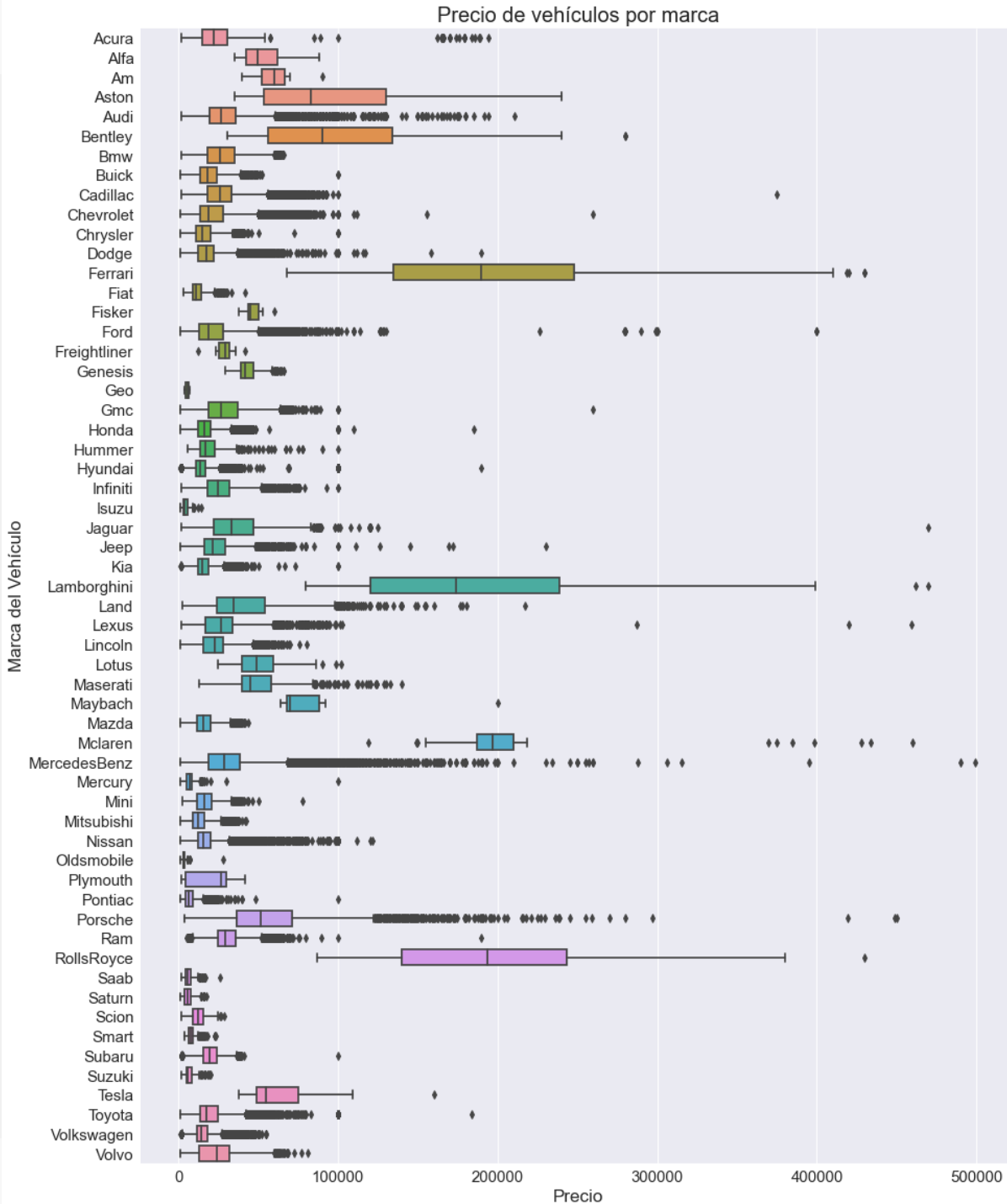
# Análisis Exploratorio Descriptivo



# Comportamiento **vector** objetivo “**Price**”

- Distribución original:
  - Sesgo
  - Outliers
- Raíz Cuadrada:
  - Sesgo
  - Outliers
- Logarítmica:
  - Outliers
  - Tendencia normal





# Precio de vehículos por marca

Presencia de outliers

Pocas marcas que superan los 200 mil dólares

Marcas de elevado costo, no presentan precios muy altos debido a la presencia de muchos outliers

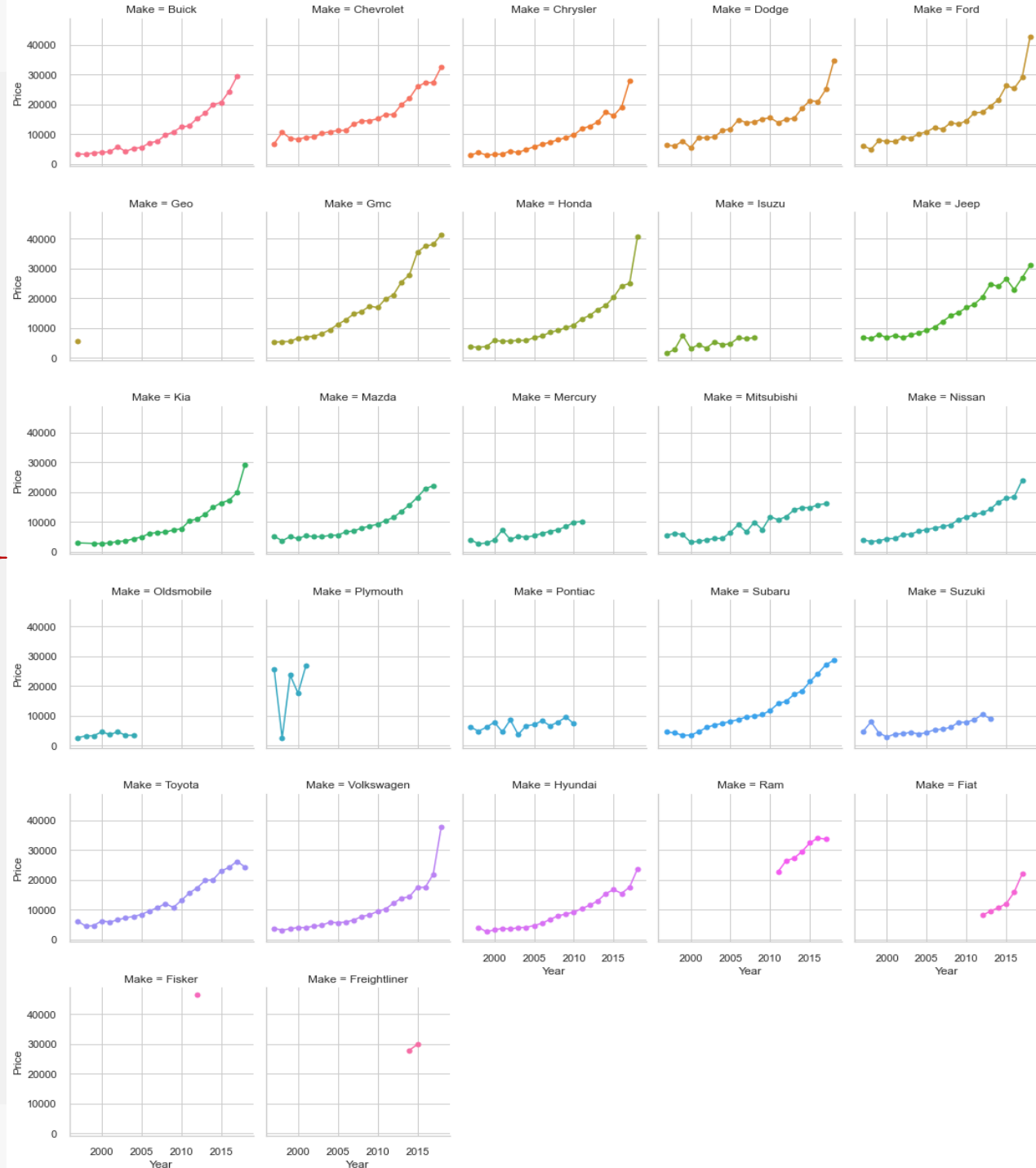
División de la muestra cada 10.000 dólares para bajar outliers

Outliers > 90.000 dólares se mantiene  
Outliers < 90.000 dólares bajan

# Estrategia de venta de vehículos por gama

Gama: generalista

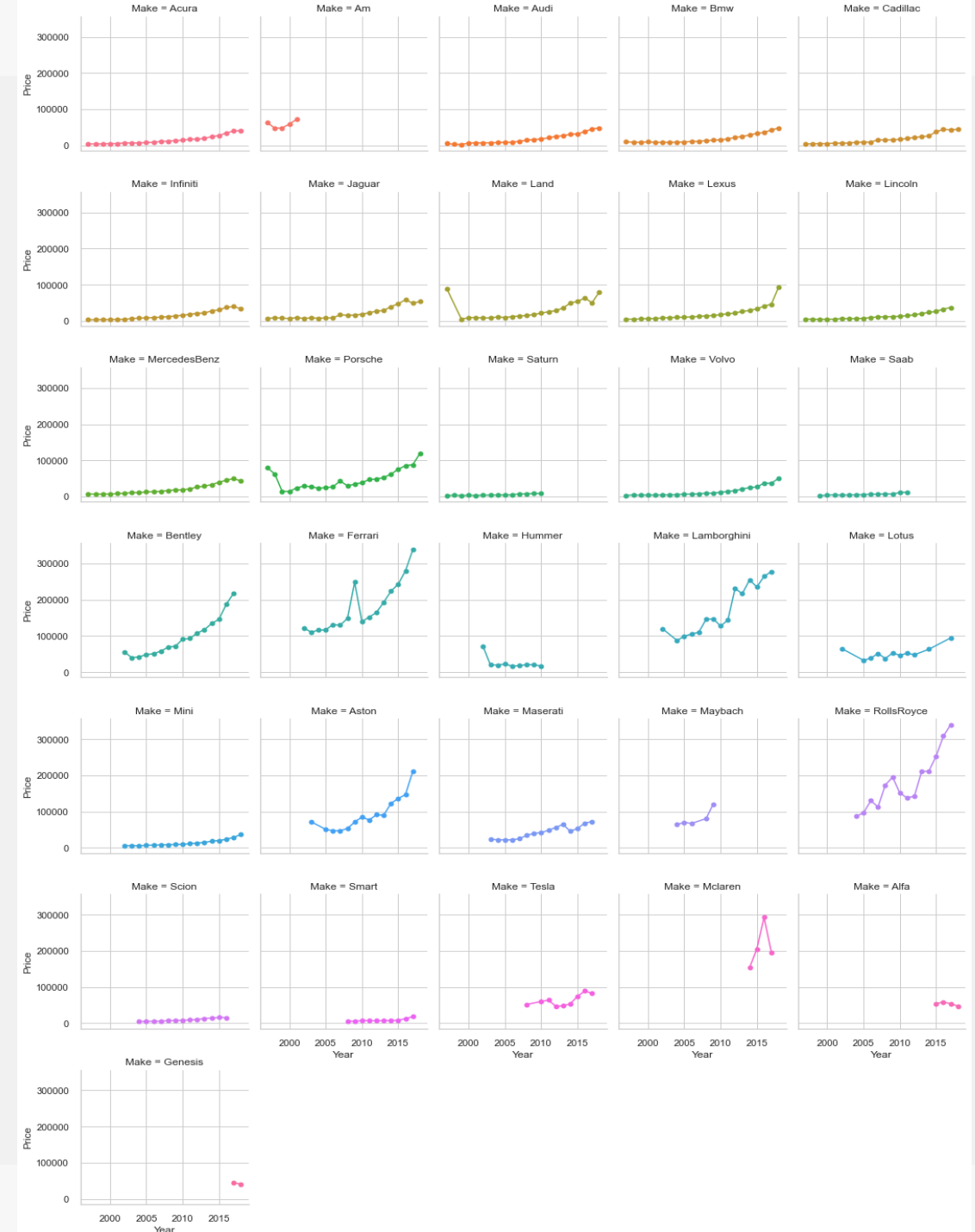
Esta división contiene demasiado outliers, lo que perjudicará el desempeño del modelo a entrenar



# Estrategia de venta de vehículos por gama

Gama: premium

Esta división contiene demasiado outliers, lo que perjudicará el desempeño del modelo a entrenar







# Conclusión:

- La división por gama no aporta positivamente al desempeño del modelo por la presencia de muchos outliers
- La división de la muestra se hará por rango de precio, quedando de la siguiente manera:





## División de muestra para disminuir influencia de outliers basado en columna 'Price'

| Rango de precio (en dólares) | Categoría   |
|------------------------------|-------------|
| 0 – 25.000                   | Generalista |
| 25.001 – 35.000              | Premium 1   |
| 35.001 – 45.000              | Premium 2   |
| 45.001 – 55.000              | Premium 3   |
| 55.001 – 65.000              | Premium 4   |
| 65.001 – 75.000              | Premium 5   |
| 75.001 – 85.000              | Premium 6   |
| 85.001 – 95.000              | Premium 7   |
| > 95.001                     | Premium 8   |

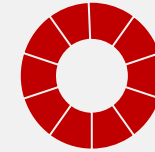
# Pre-Procesamiento



Binarización



Limpieza data set



División  
vehículos por  
precio

A través de Base de Datos Postgres  
(Pg-Admin)





”

**MODELOS CANDIDATOS**

## Modelos entrenados para categoría generalista

Se realizaron entrenamientos de modelos “vainilla” para observar su rendimiento, y como afectaban los outliers en las métricas. Las mejores métricas se consiguieron en los siguientes modelos:

| Modelo           | Métricas      | Justificación  |
|------------------|---------------|--|
| LinearRegression | MAE: 2275     | Asume una relación lineal entre las variables (Km, año, marca, etc.) |
|                  | RMSE: 3448,75 |  |
|                  | R2: 56,28%    |  |
| LassoCV          | MAE: 2575     | Selecciona automáticamente las características más importantes       |
|                  | RMSE: 3870,04 |  |
|                  | R2: 44,94%    |  |

Modelos como DecisionTreeRegressor y RandomForestRegressor no obtuvieron buenos resultados





# Gracias!

¿Dudas?  
¿Consultas?  
¿Retroalimentación?

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**

”

