



SERGIO TECH

PREDICCIÓN DEL PRECIO DE COCHES

SERGIO CHALELA

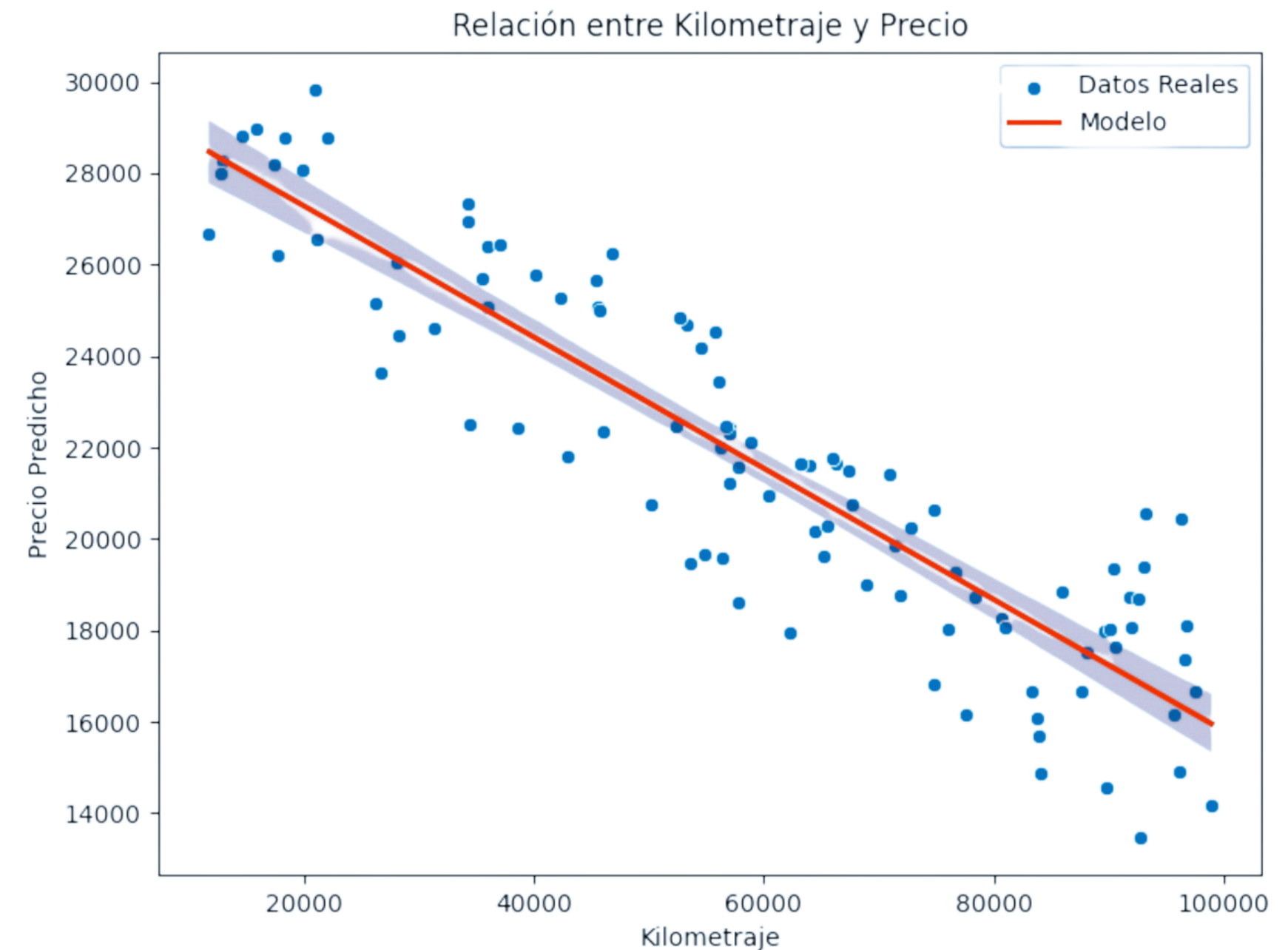


QUE PROBLEMA RESOLVEMOS?

Nuestro objetivo principal fue construir un modelo de Machine Learning que permita predecir el precio estimado de un vehículo basado en sus características, de manera eficiente y confiable.

Este enfoque también destaca por la combinación de:

- **Modelos supervisados:** Para entrenar el modelo en un dataset robusto.
- **Clustering no supervisado:** Para encontrar patrones ocultos que mejoren la precisión.

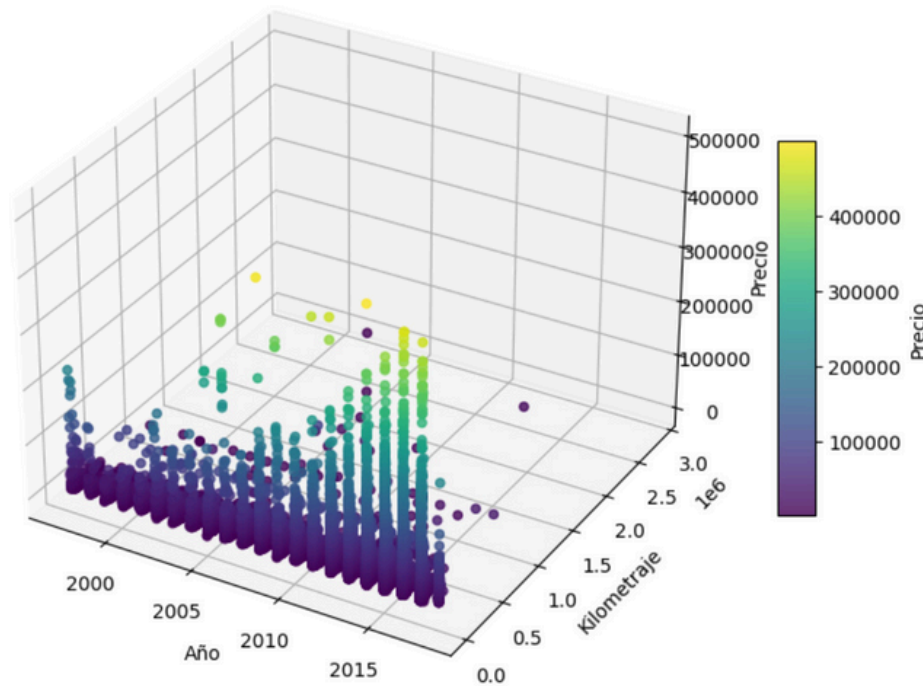




SERGIO Tech

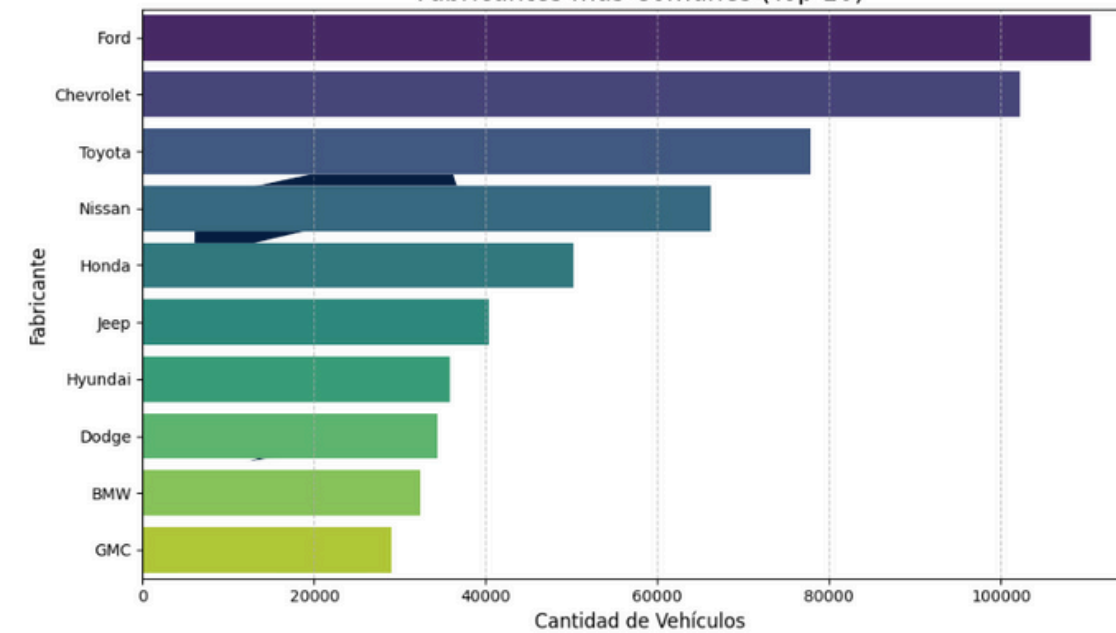
EXPLORACIÓN INICIAL DE DATOS

Relación entre Año, Kilometraje y Precio



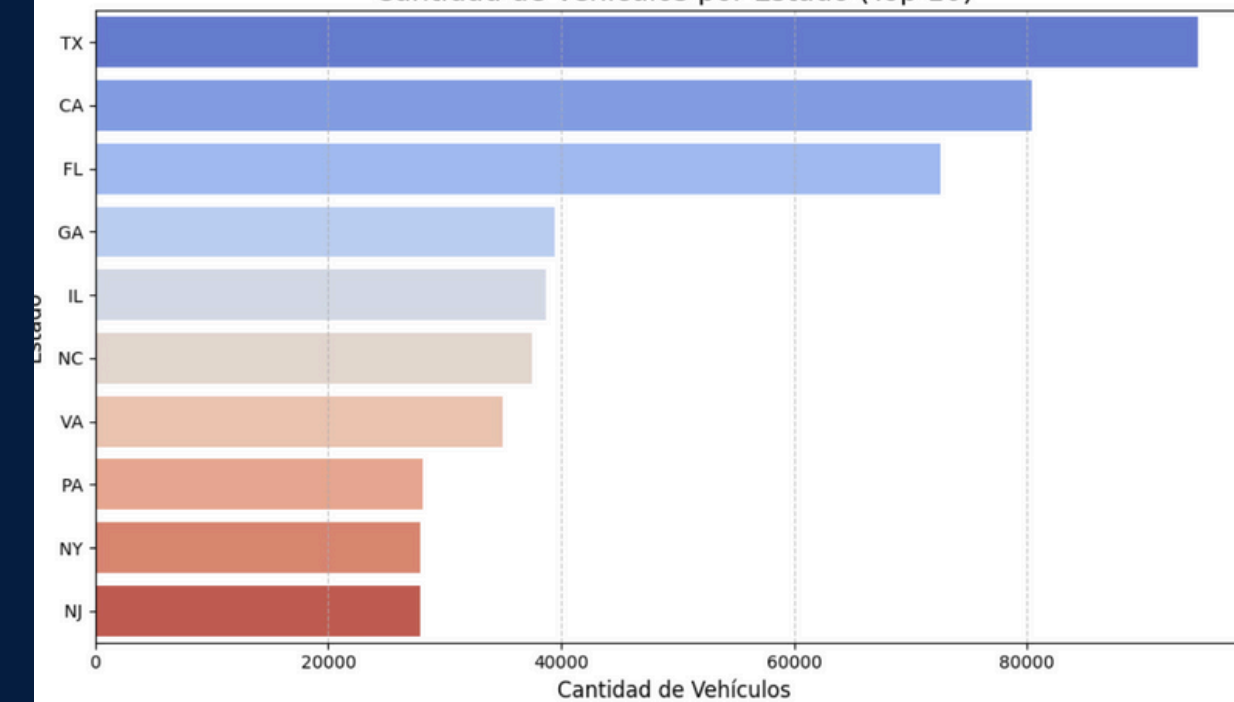
KM vs Año vs Precio

Fabricantes más Comunes (Top 10)



Fabricantes más Comunes

Cantidad de Vehículos por Estado (Top 10)

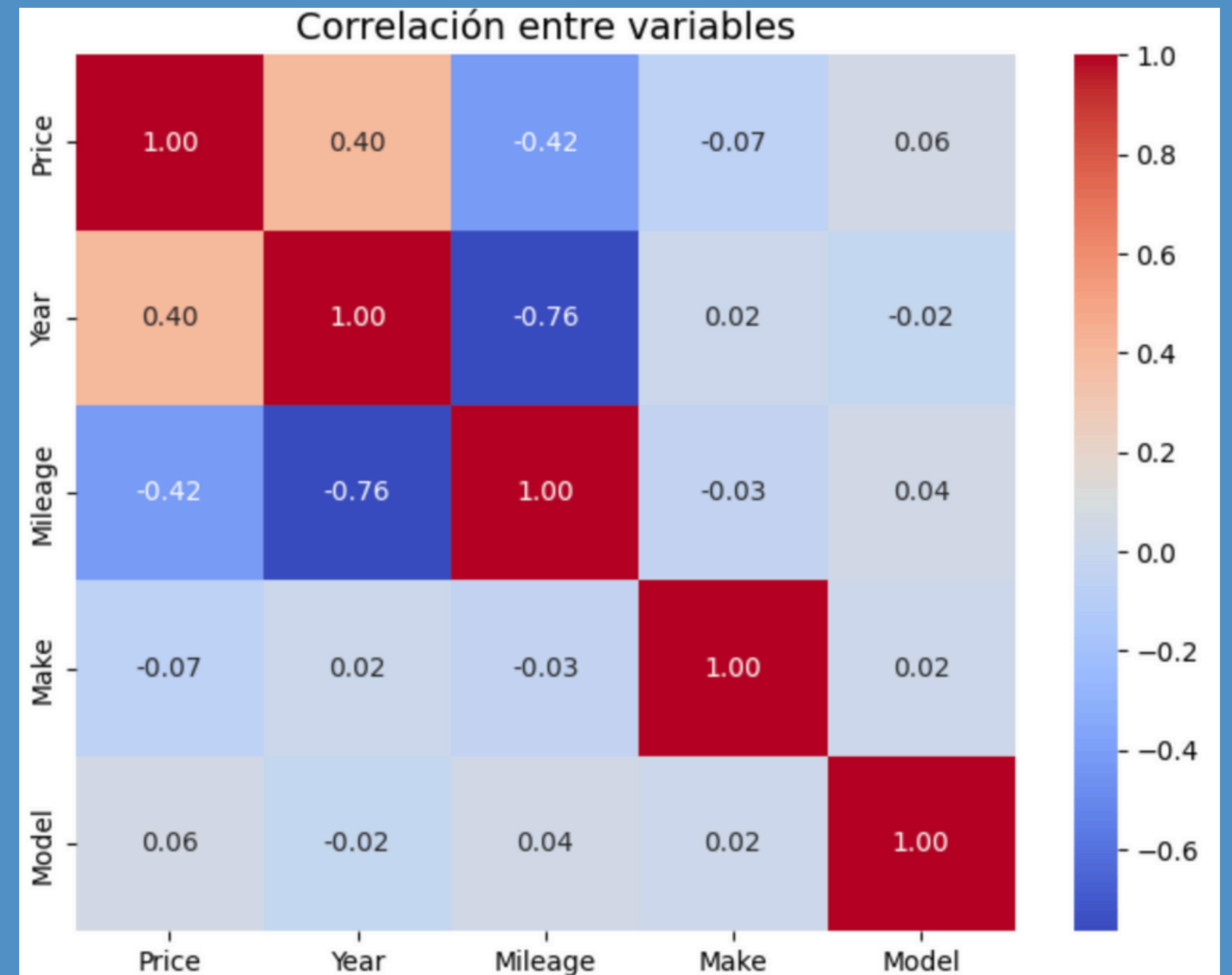
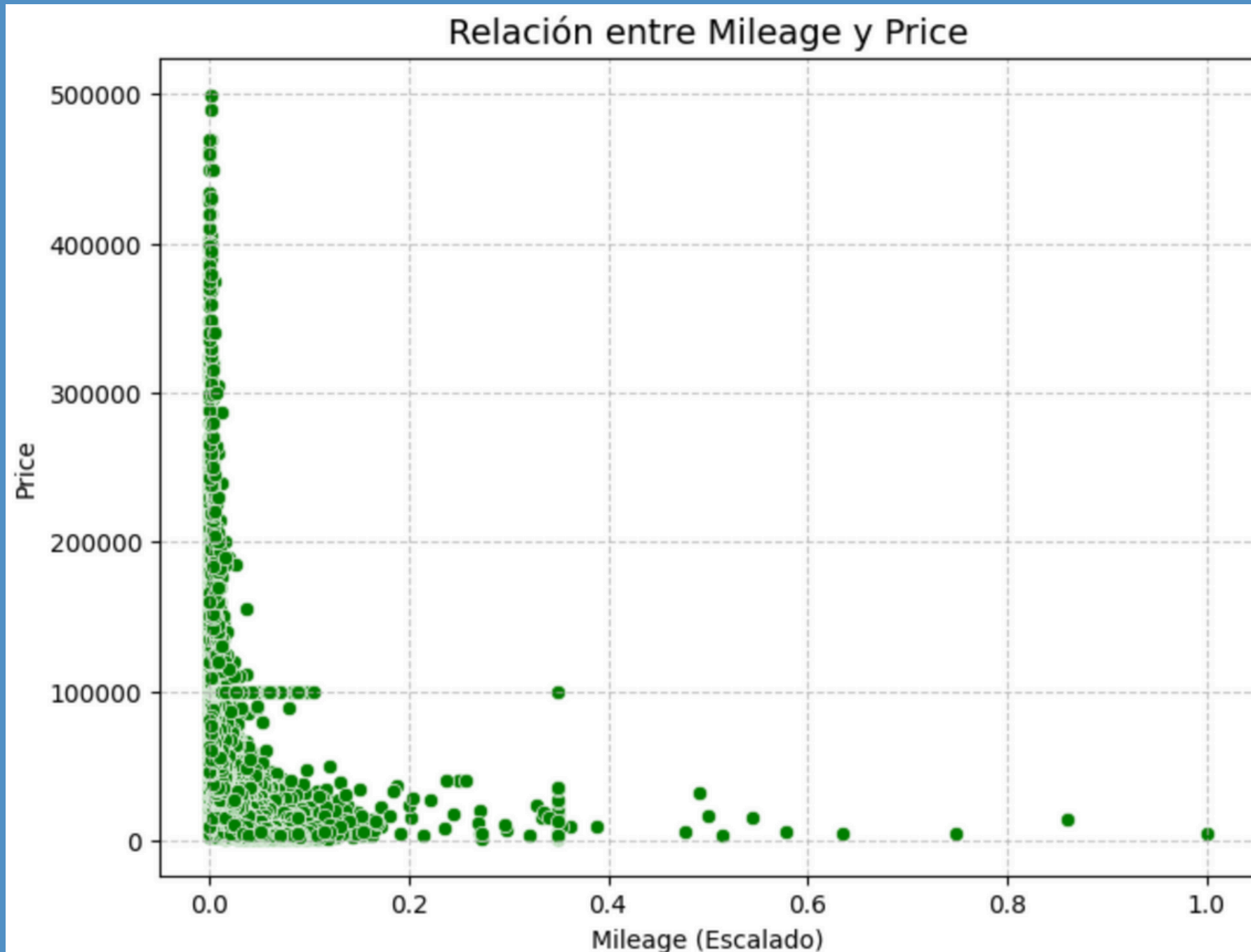


Cantidad de Vehiculos por Estado

EXPLORACIÓN DE DATOS PROCESADOS



SERGIO Tech

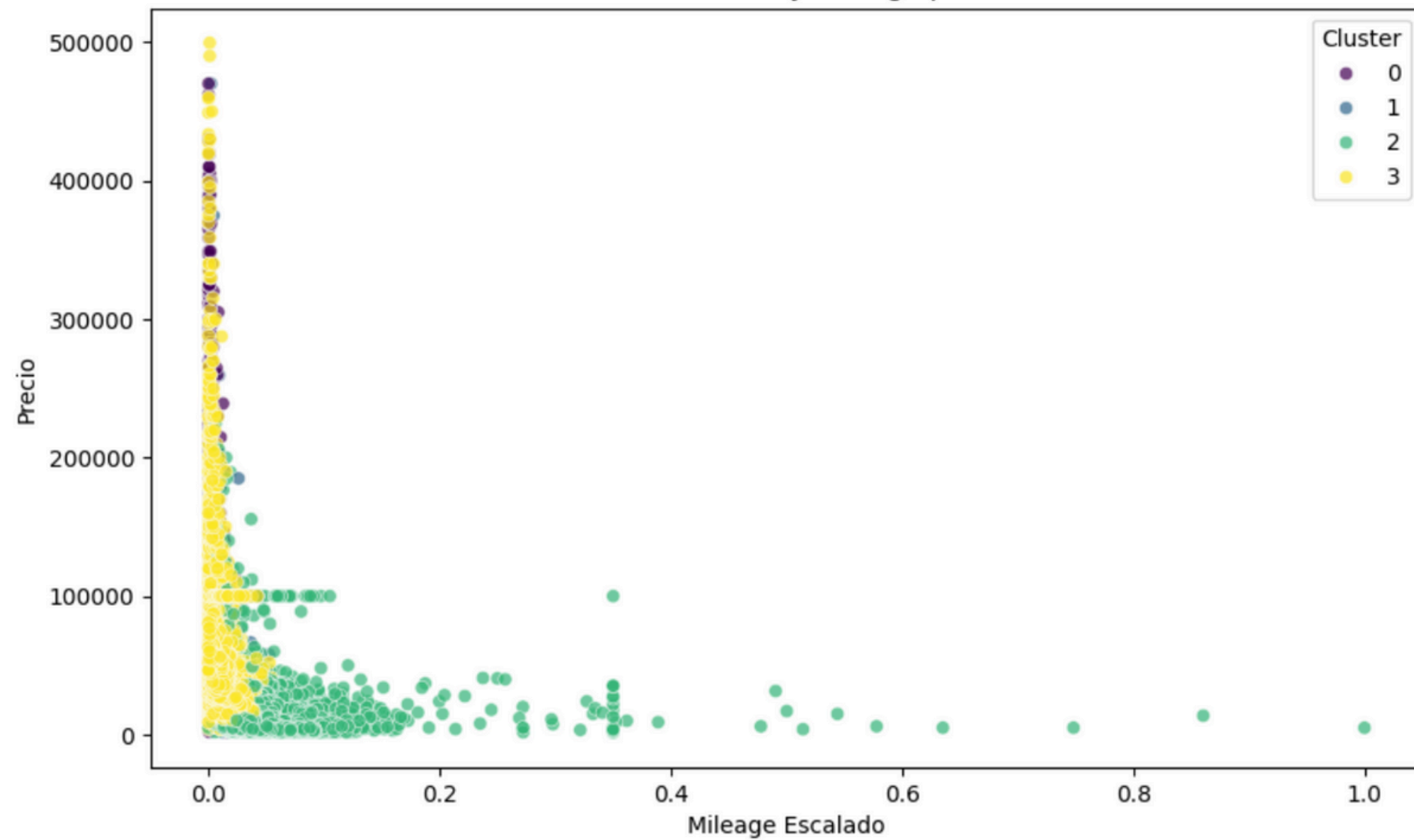


EXPLORACION DE DATOS CLUSTERIZADOS

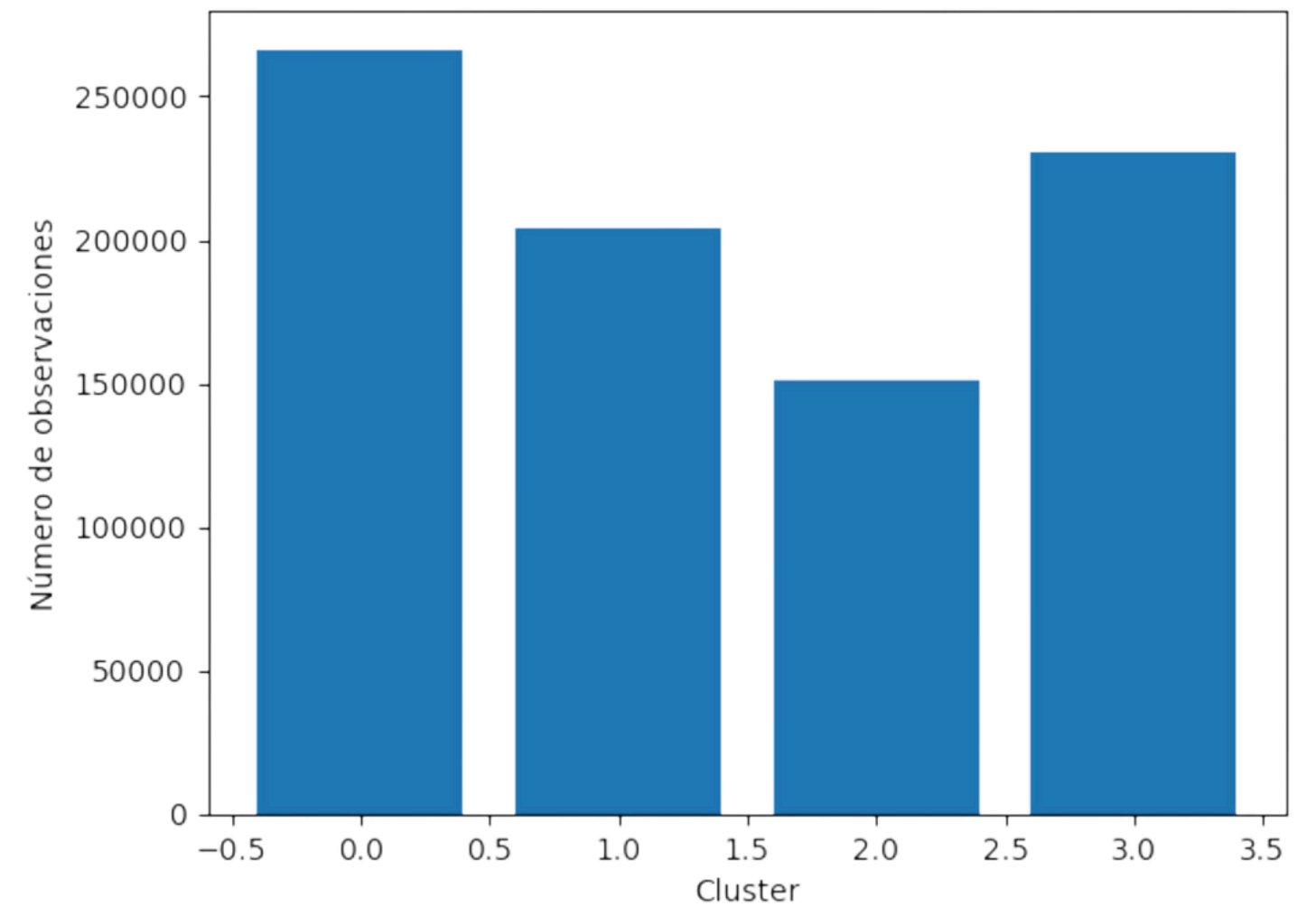


SERGIO Tech

Relación entre Precio y Mileage por Cluster



Distribución de Clusters



MODELOS SUPERVISADOS



SERGIO Tech

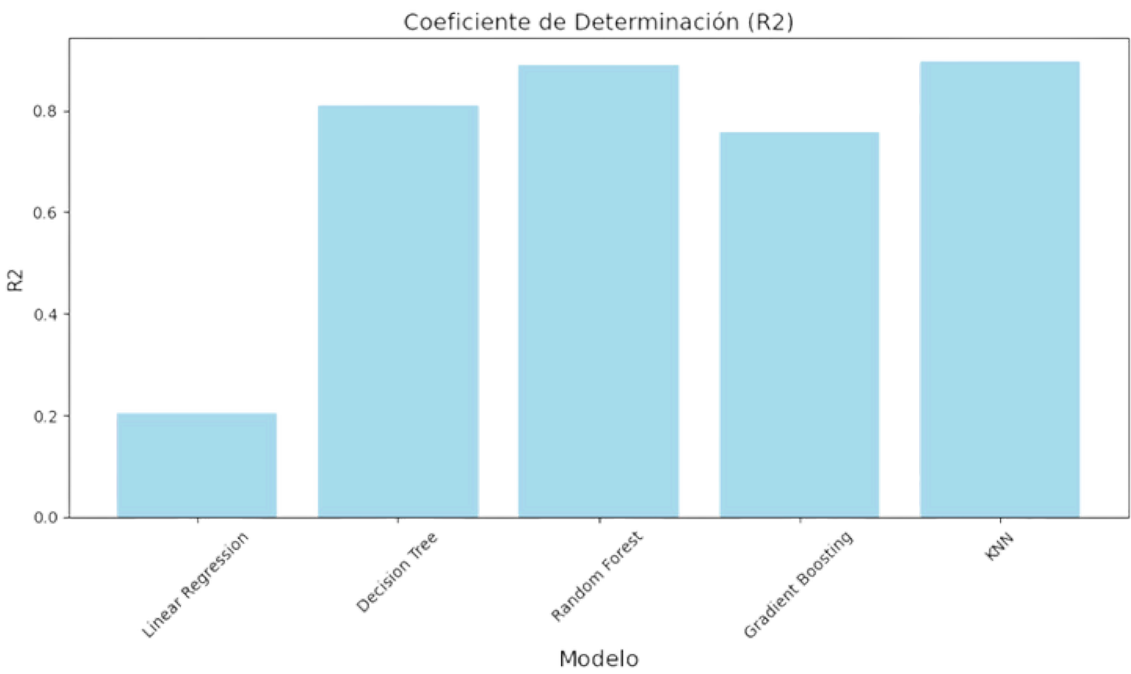
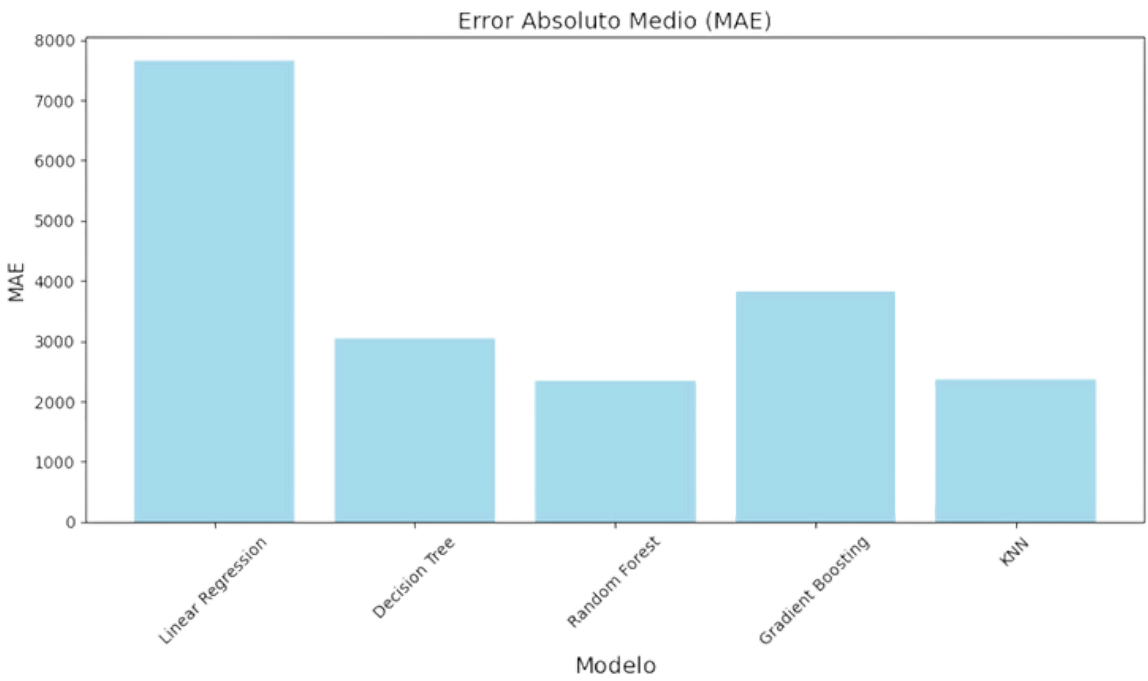
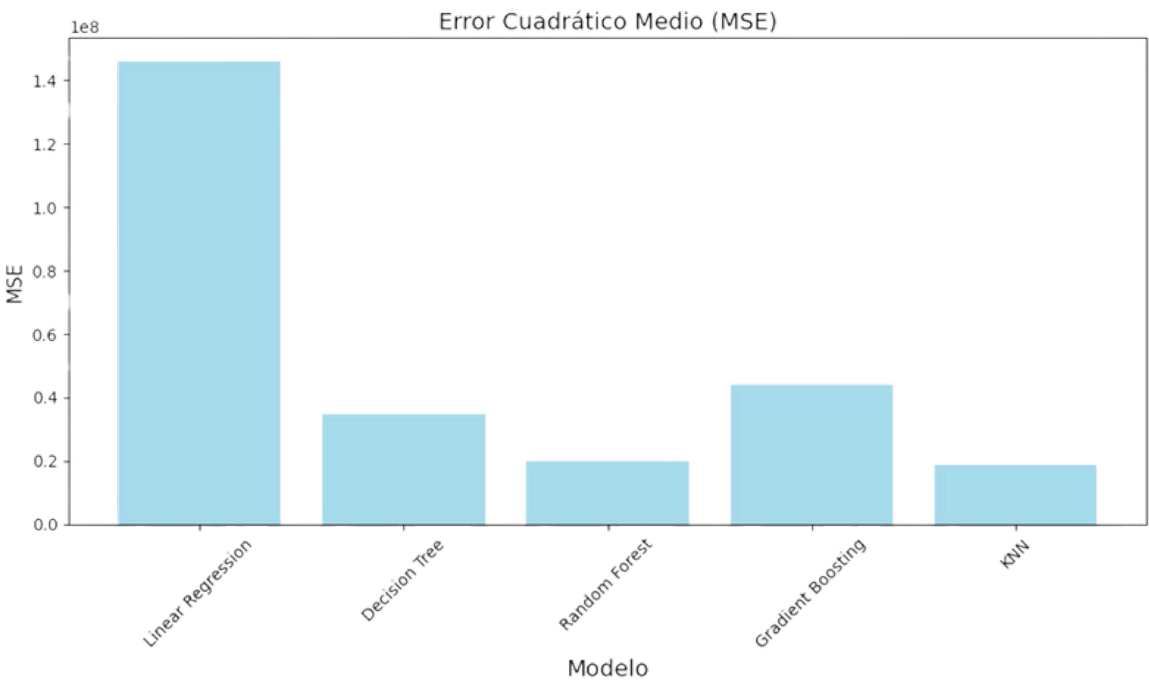
Regresión Lineal

Decision Trees

Random Forest

Gradient Boosting

KNN



MODELO NO SUPERVISADO



SERGIO Tech

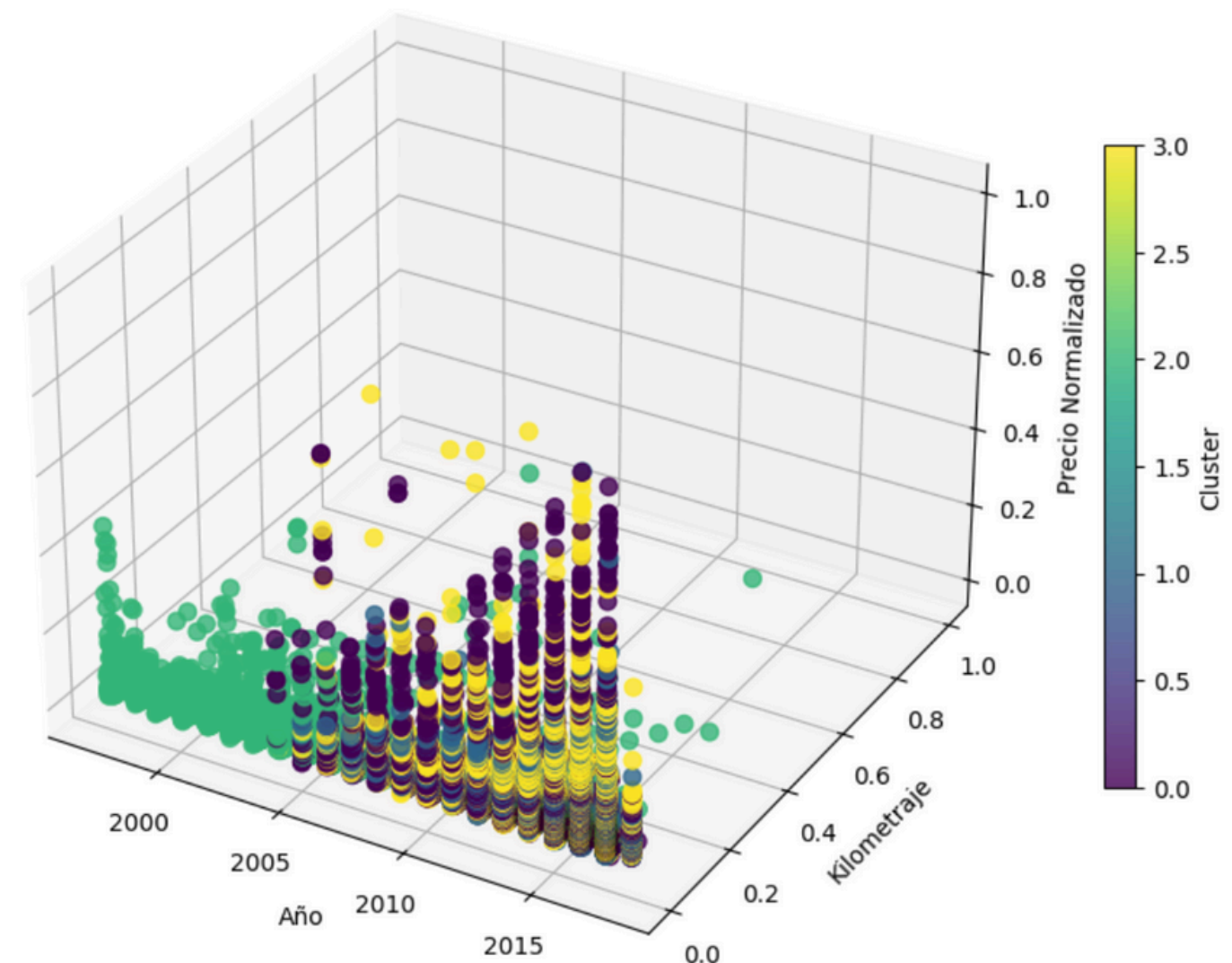
Porque K-Means

- Aplicamos K-Means para segmentar los datos en grupos homogéneos.
- Los clusters revelan patrones latentes que no eran evidentes en el análisis inicial

Resultado

- El algoritmo identificó X clusters basados en variables como Año, Kilometraje, y otras características.
- Cada cluster representa un grupo de vehículos con características similares, lo que mejora la precisión del modelo supervisado.

Distribución de Clusters (Precio Normalizado)



MODELO FINAL



SERGIO Tech

hyperparameters:

```
algorithm: auto  
metric: minkowski  
n_neighbors: 5  
weights: uniform
```

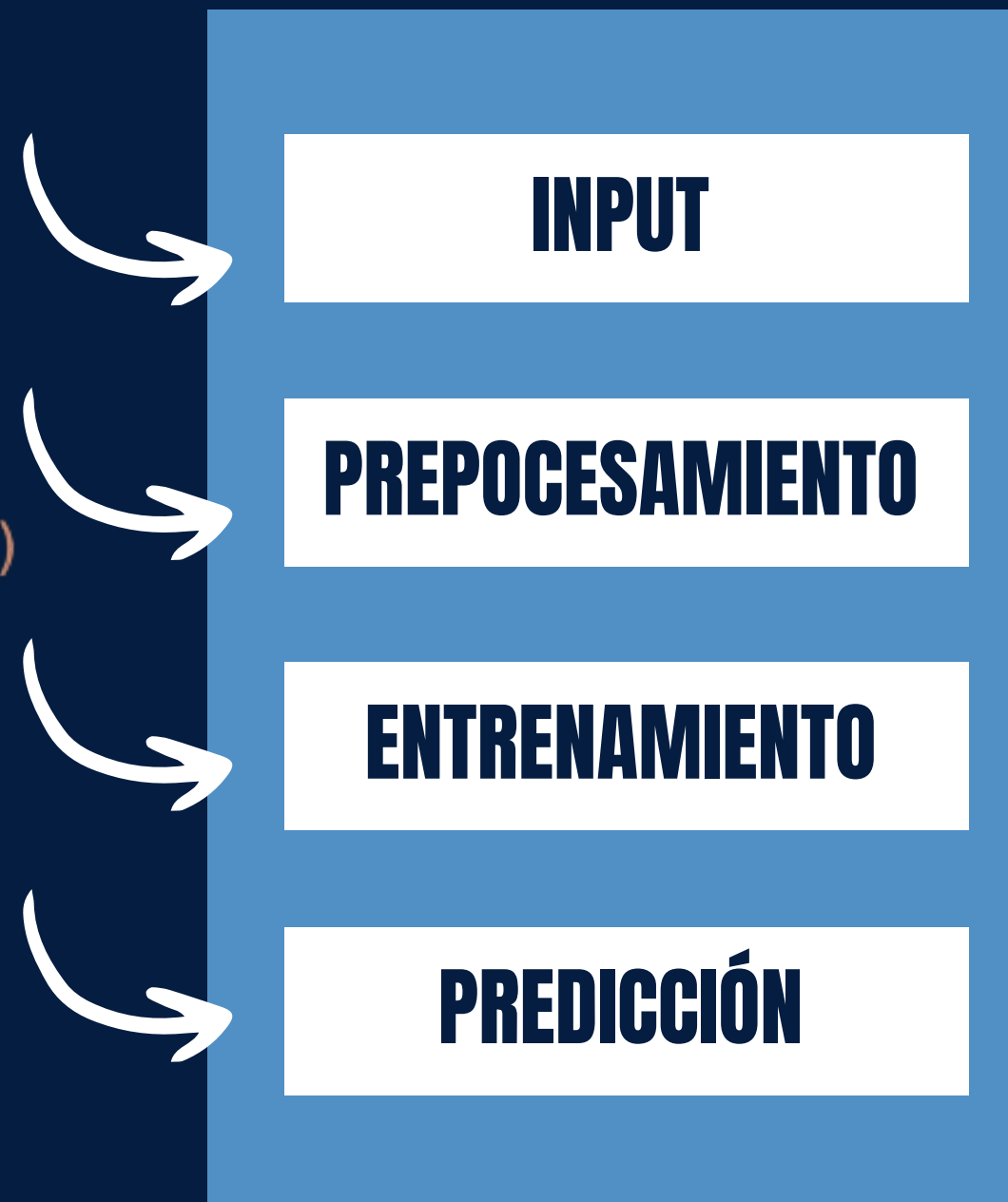
model_name: K-Nearest Neighbors (KNN)

performance:

```
MAE: 2368.780735220772  
MSE: 18919220.428636175  
R2: 0.8970978654746379
```

preprocessing:

```
clustering: KMeans (k=4)  
encoding: LabelEncoder  
scaling: MinMaxScaler
```



PUNTOS CLAVE

- Facil de interpretar
- Tiempo de entrenamiento eficiente
- Mejora en las metricas al incorporar “clustering”





SERGIO Tech

¡GRACIAS!



+31 691 726 256



<https://github.com/SergioChalela>



ingenierosergio@ML.com