

MotorAlpes

Modelo predictivo MotorAlpes

Mariana Díaz

Juan Diego Yepes

Esteban Gonzalez

Información General



MotorAlpes recolectó información general de todo tipo de vehículos usados junto a los precios que se consideran apropiados.

7115 filas y 11 columnas

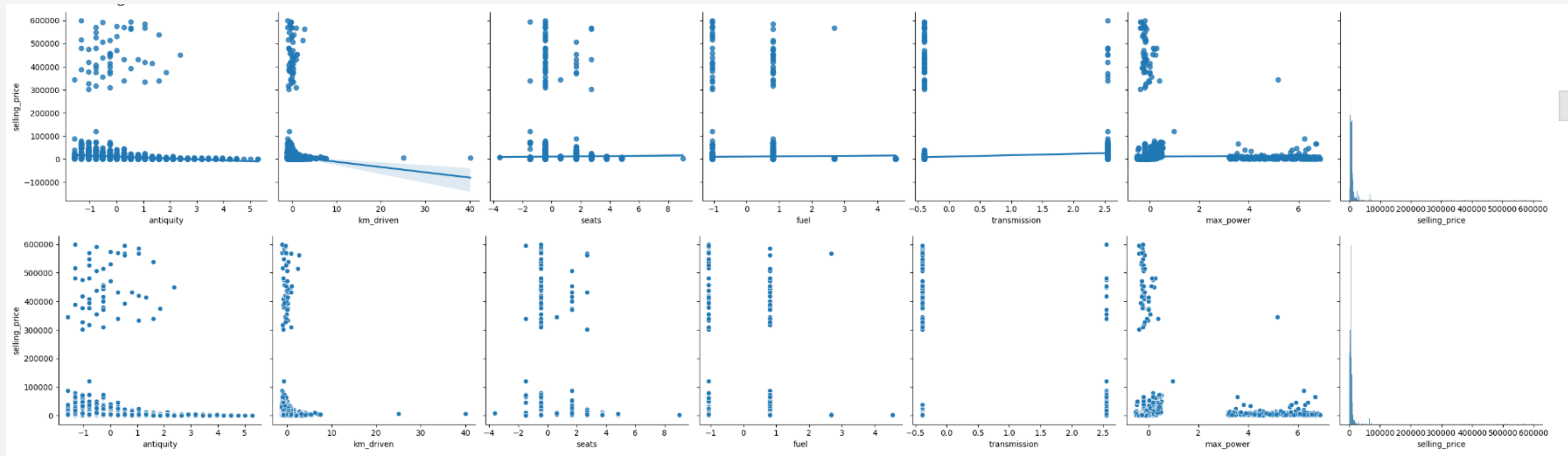
Se realizó un análisis de los datos para verificar la relevancia e integridad de los datos y se verificó que los datos cumplieran con:

Compleitud
Unicidad
Consistencia
Validez



ANÁLISIS EXPLORATORIO

Preparación de los datos: Agregar nuevas columnas relevantes, llenar valores vacios con alguna metrica determinada y reemplazar los valores categoricos por representaciones numericas



Datos incialmente comparados contra el "selling_price" para ver tendencias y ayudar en la identificacion de features

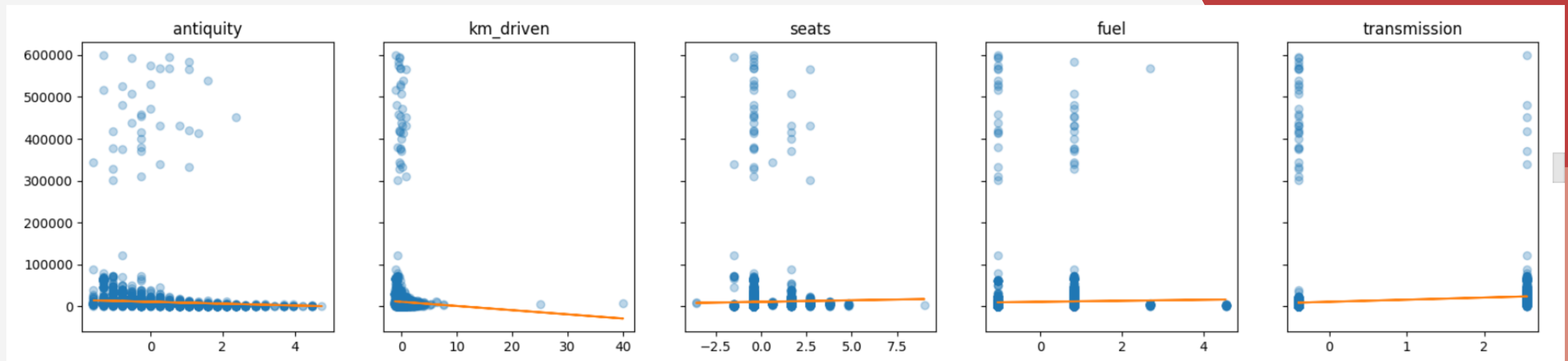
ANÁLISIS EXPLORATORIO

Al analizar los coeficientes entre las columnas se puede tener una impresión inicial que cada una de ellas tiene sobre selling_price

	antiquity	km_driven	seats	fuel	transmission	max_power	selling_price
antiquity	1.000000	0.412650	0.013623	-0.018283	-0.239609	-0.053672	-0.102322
km_driven	0.412650	1.000000	0.224160	0.253972	-0.192115	-0.007598	-0.057280
seats	0.013623	0.224160	1.000000	0.305467	-0.073186	0.022073	0.013726
fuel	-0.018283	0.253972	0.305467	1.000000	-0.041470	0.037757	0.026568
transmission	-0.239609	-0.192115	-0.073186	-0.041470	1.000000	0.072900	0.149686
max_power	-0.053672	-0.007598	0.022073	0.037757	0.072900	1.000000	0.015804
selling_price	-0.102322	-0.057280	0.013726	0.026568	0.149686	0.015804	1.000000

MODELAMIENTO

Se obtuvieron muestras para datos de entrenamiento y de prueba y a partir de ellos, se aplicó la regresión para ver su impacto sobre selling_price. Se pudo observar que una de las variables que mas lo afectan es km_driven



MODELAMIENTO

Tras realizar un modelo predictivo se debían evaluar los errores en los datos de entrenamiento y los de prueba.

Mean Absolute Error Train:
8482.988867562579

Mean Absolute Error Test:
8569.891106196956

Podemos analizar que los errores al ser tan similares nos indican que nuestro modelo es preciso, pues está generando valores muy parecidos a los valores reales de acuerdo a los atributos que determinan el precio de un carro usado.

CONCLUSIONES

Nuestro modelo puede predecir correctamente los costos de los carros usados, ya que, los valores obtenidos son muy similares a los entregados por la empresa

Las variables de antiquity, km_driven, seats, fuel, transmission y max_power efectivamente si tienen un impacto sobre el precio calculado y fueron bien elegidas.

Se realizó un pipeline que hace la predicción para cualquier otro set de datos que será entregado a la empresa para ser evaluado y usado de acuerdo a los resultados exitosos obtenidos,



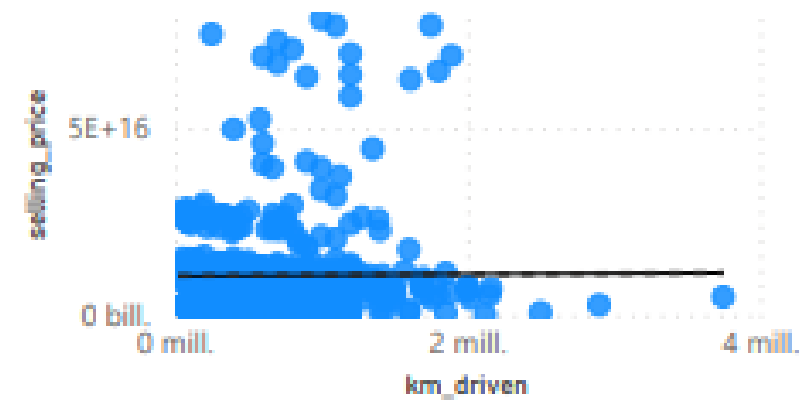
Dashboard



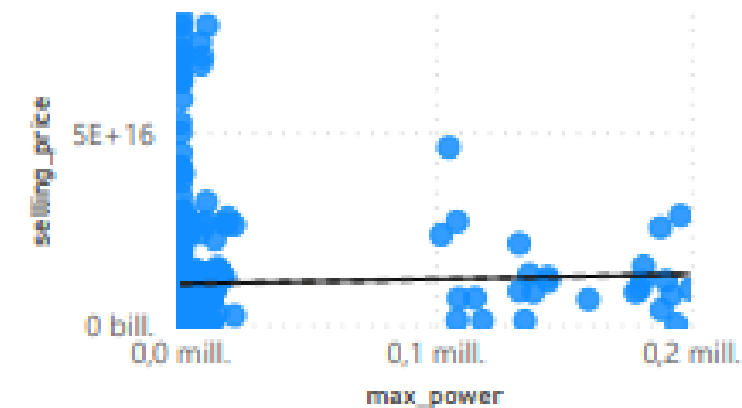
MotorAlpes

Relación entre las variables y el precio de venta

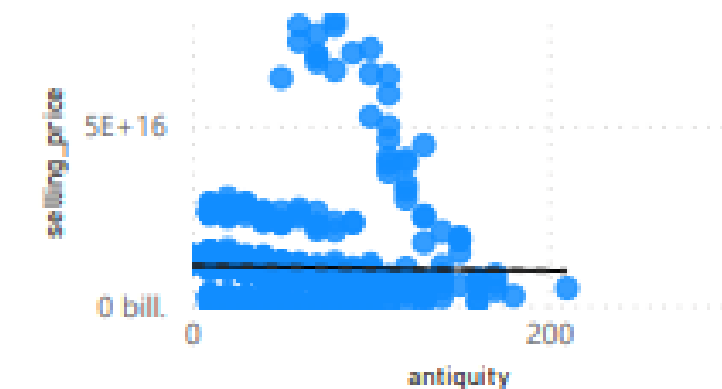
km_driven y selling_price



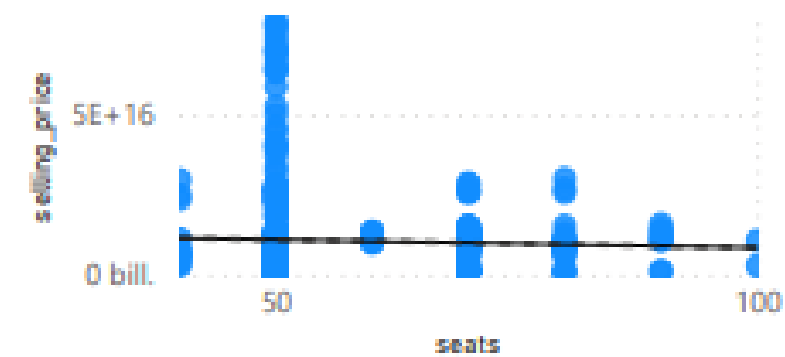
max_power y selling_price



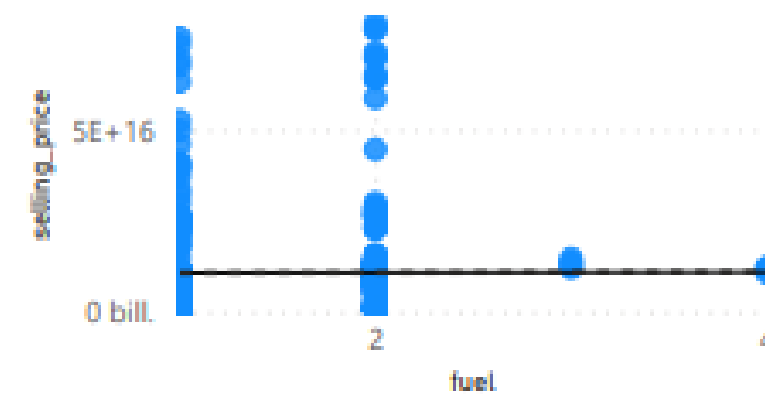
antiquity y selling_price



seats y selling_price



fuel y selling_price



transmission y selling_price

