



MOTORALPES

Presentación de resultados

MOTORALPES 2023

Juan Jose Osorio - 202021720

Juan Sebastián Hoyos - 201822167

Alejandro Guatibonza - 202014393

Presentación y preparación de datos

Para poder analizar correctamente la información de motorAlpes, es necesario hacer una limpieza de los datos. Para esto, se revisa cual es el porcentaje de datos nulos en la tabla según la característica. Como se puede ver, la columna selling_price tiene un 5% de error. Debido a que estos son datos de prueba, con los cuales se quiere entrenar el modelo, es necesario que esta columna no tenga nulos. Por tanto, se eliminan.

Porcentaje de valores vacíos

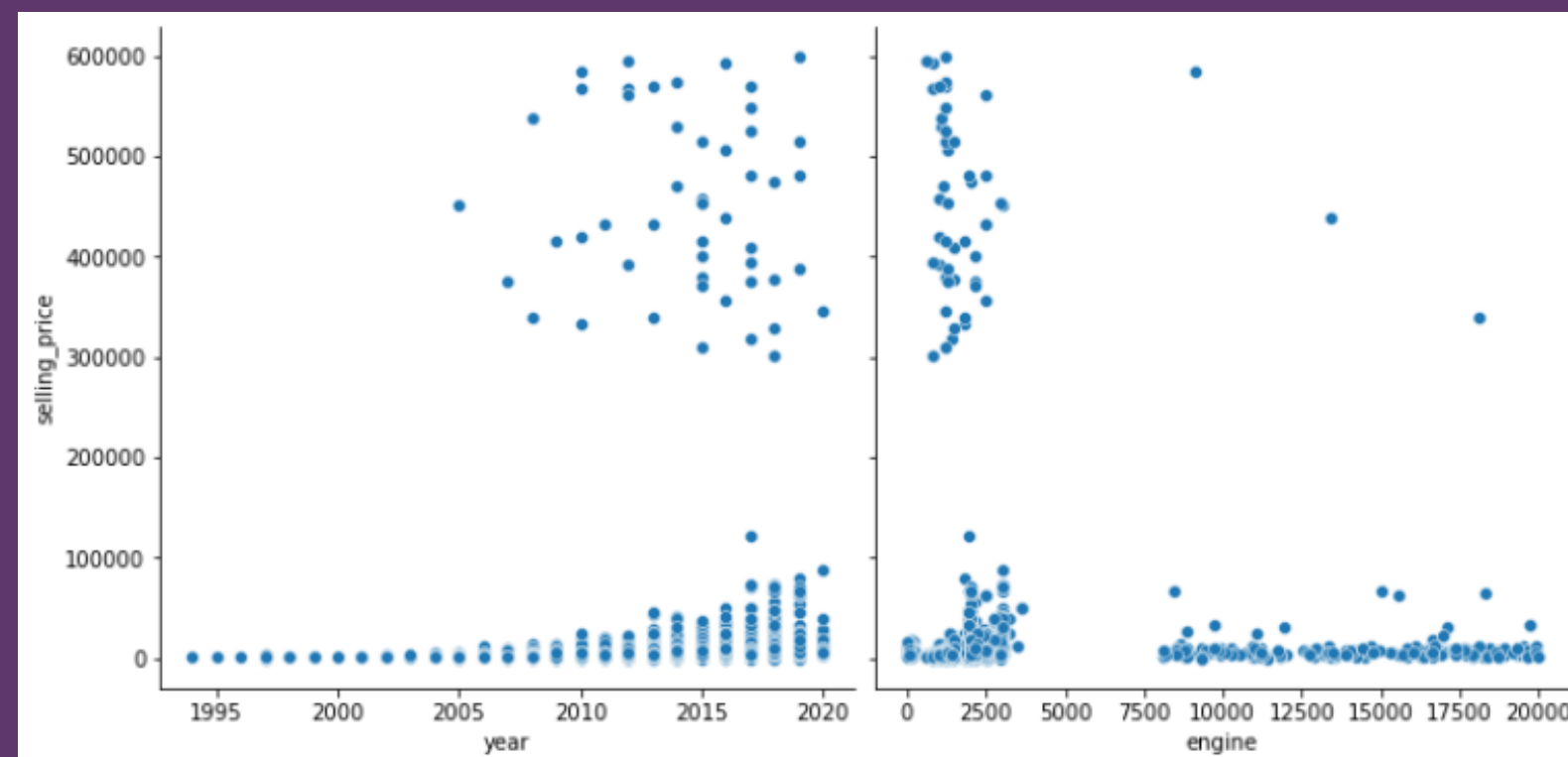
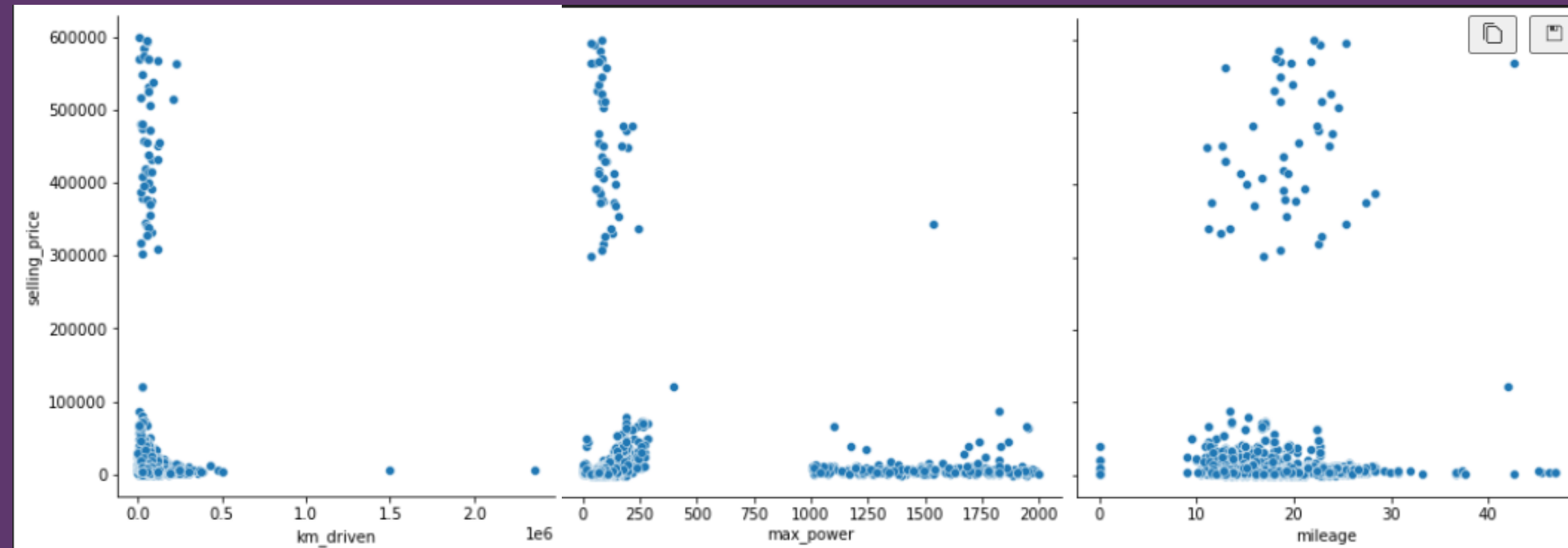


selling_price	5.635980
engine	3.935348
max_power	3.766690
year	3.359100
owner	3.359100
km_driven	2.782853
mileage	2.782853
seller_type	0.000000
seats	0.000000
fuel	0.000000
transmission	0.000000

Relación variables y precio de venta



MOTORALPES



A partir de los datos anteriores se puede observar que hay varios datos candidatos para el modelo, tales como year, km_driven y engine.

Coeficientes

- Coeficientes determinan influencia de las características sobre el precio de venta
- Las columnas Year y mileage son las mas influyentes en el precio de venta
- Entre mas kilometraje tenga, es menor el precio de venta del carro
- Entre mas nuevo sea el carro, mas dinero se paga por el

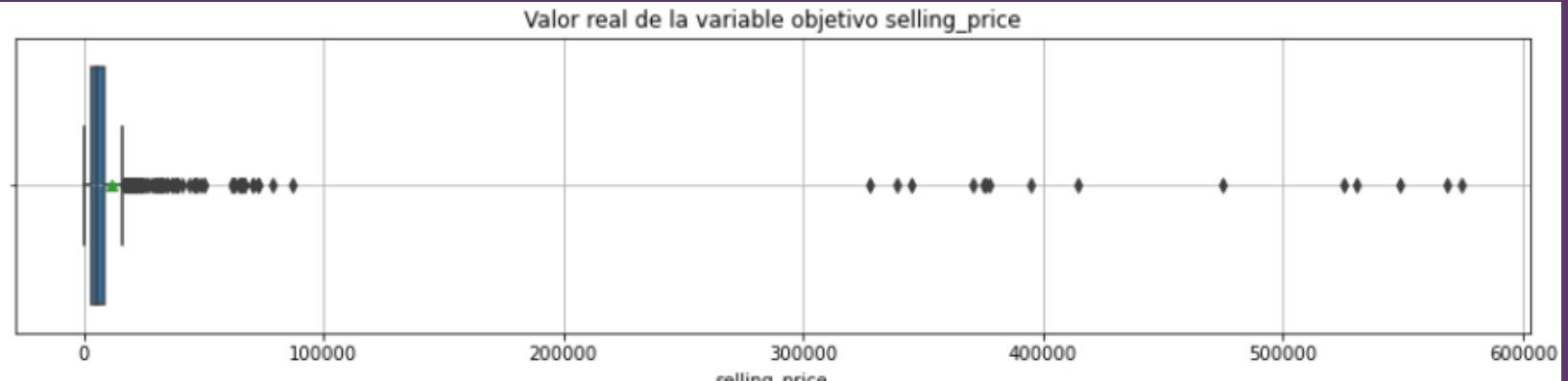


MOTORALPES

	columns	coeficientes
0	year	1231.506865
1	engine	0.699109
2	max_power	0.228390
3	mileage	-668.223865
4	km_driven	-0.009640

Evaluación cuantitativa

- Las métricas de error para el dataset de entrenamiento y prueba son muy similares.
- los valores de la variable objetivo están centrados en 11324 dolares, con una desviación estandar de 40844 dolares
- El 50% de los errores de estimación del modelo se encuentran en +- 5440\$ dolares



count	1814.000000
mean	11324.929719
std	40844.939588
min	28.500000
25%	3149.990000
50%	5451.900000
75%	8480.740000
max	574092.810000



count	1814.000000
mean	10040.154796
std	39331.192114
min	4.969506
25%	3003.427445
50%	5440.073452
75%	8254.518230
max	575740.289209

MAE

Train: 10204.409067607368
Test: 10040.154796196848

RMSE

Train: 1790463592.829145
Test: 1646894601.6025705

Evaluación cualitativa

- Se normaliza para que todas las características esten en un mismo rango
- El año es la característica que mas aporta al precio del carro
- Los errores se disminuyeron haciendo la normalización
- Se espera que en la regresión lineal los errores sigan una distribución normal
- Como no la sigue, hay que tratar los errores

Normalización



	columns	coef
0	year	4783.810777
1	engine	1645.263952
2	max_power	59.585075
3	mileage	-2853.823056
4	km_driven	-611.745876

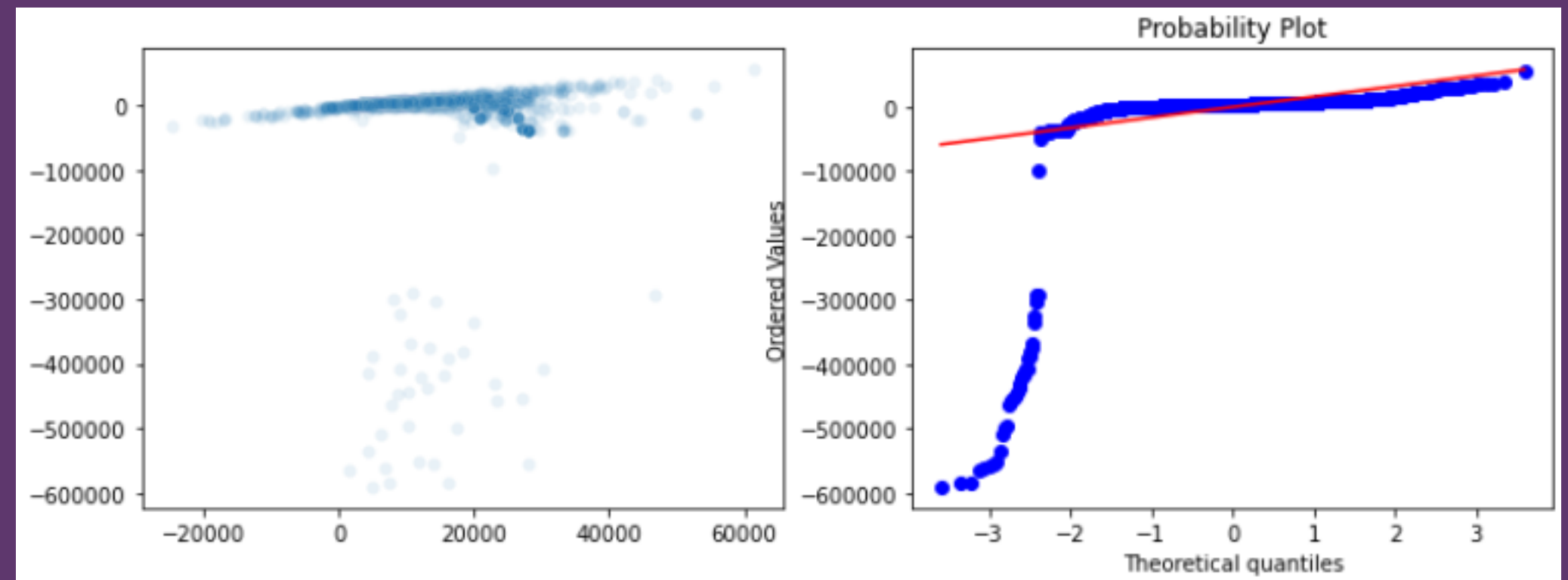
MAE

Train: 9578.8809
Test: 9536.90133

RMSE

Train: 42039.5057
Test: 40302.81942

Normalidad en errores



Evaluación cualitativa

- Se quitan los datos atipicos, es decir, los que están entre el 25 y el 75% de los datos.
- El error de los datos se reduce considerablemente.
- La varianza debe permanecer constante con el cambio de la variable objetivo.
- Una mala grafica indica que se necesita una transformación o una variable extra.
- La mejor grafica es la que tiene a year, km_driven y mileage. Por tanto, se toman estas variables.

Remoción outliers



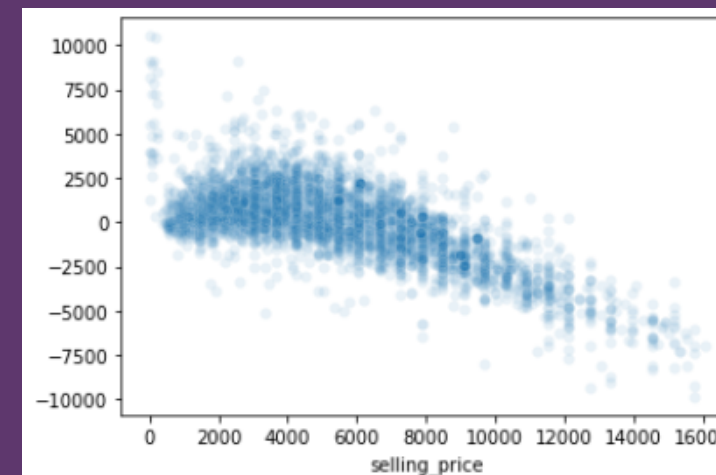
MAE

Train: 1579.9754
Test: 1585.1174

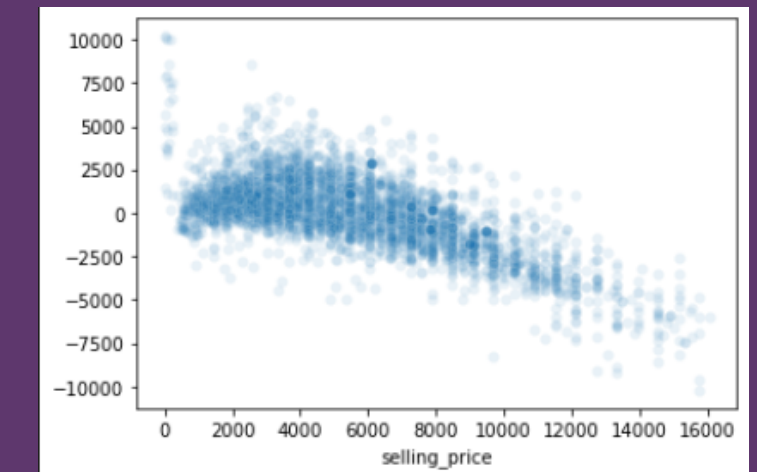
RMSE

Train: 2139.5167
Test: 2236.28638

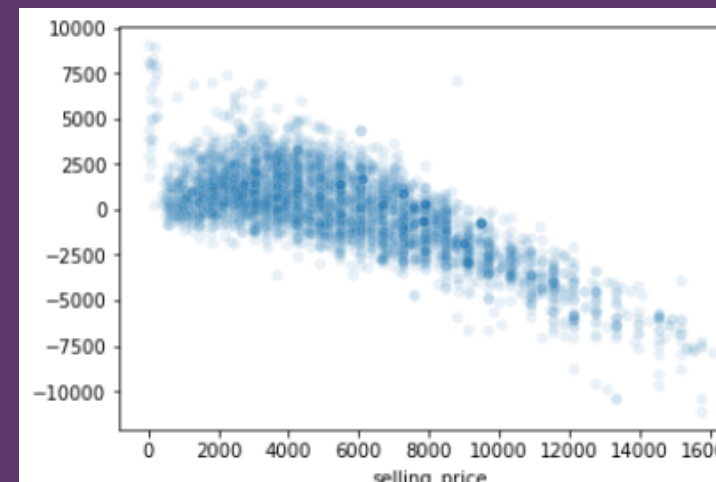
Varianza constante



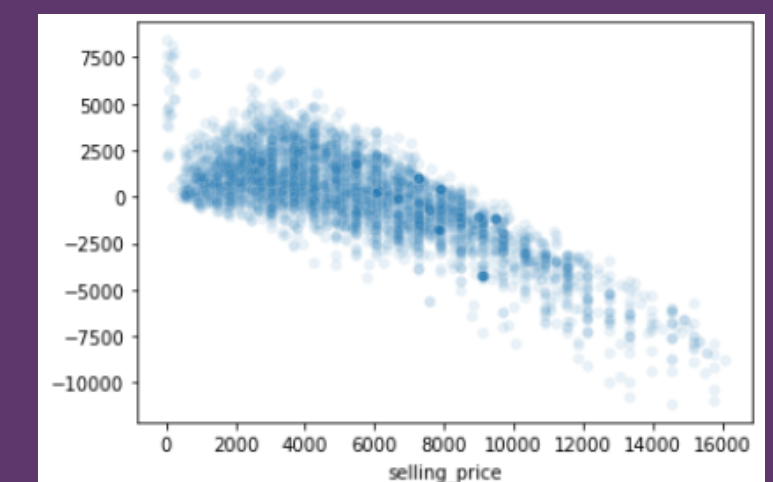
Variables iniciales



variables iniciales y seats



year, km_driven y mileage



year y km_driven

Conclusiones

- Se obtienen menos errores entre menos datos atipicos se tengan
- Los datos atipicos son aquellos carros cuyo precio es muy distinto a aquellos carros que tienen características en comun .
- Para obtener el precio correctamente es necesario conocer otras características de los carros.
- En general, se obtuvo que las características mas influyentes fueron año, kilometraje y kilometros conducidos
- Entre mas reciente sea el modelo del carro, su costo se va a elevar mas
- A mayor kilometraje o kilometros conducidos, el precio del carro tiende a ser mas bajo
- Los otros datos aportan al modelo, pero de una manera muy leve