

# Predicción de Precio de Autos Usados

# Introducción

Mediante un conjunto de datos con información sobre el nombre del automóvil usado, los kilómetros recorridos, el tipo de combustible, el tipo de transmisión, el año de fabricación, el CC del motor y el número de asientos, se realizará la predicción del precio del automóvil por medio de un modelo de regresión.

# Manipulación de Datos

La base de datos contiene 10 columnas, la cual nos provee solo 3 características: fuel\_type, transmission y Seats. Por lo tanto, para esta limpieza tenemos el reto asignar la tipología correspondiente a cada variable, pero lo mas retador es identificar si todas las variables nos aportan valor.

	car_name	car_prices_in_rupee	kms_driven	fuel_type	transmission	ownership	manufacture	engine	Seats
0	Jeep Compass 2.0 Longitude Option BSIV	10.03 Lakh	86,226 kms	Diesel	Manual	1st Owner	2017	1956 cc	5 Seats
1	Renault Duster RXZ Turbo CVT	12.83 Lakh	13,248 kms	Petrol	Automatic	1st Owner	2021	1330 cc	5 Seats
2	Toyota Camry 2.5 G	16.40 Lakh	60,343 kms	Petrol	Automatic	1st Owner	2016	2494 cc	5 Seats
3	Honda Jazz VX CVT	7.77 Lakh	26,696 kms	Petrol	Automatic	1st Owner	2018	1199 cc	5 Seats
4	Volkswagen Polo 1.2 MPI Highline	5.15 Lakh	69,414 kms	Petrol	Manual	1st Owner	2016	1199 cc	5 Seats

Se realiza limpieza a los datos, donde no se identifican valores nulos, pero si valores duplicados y se requiere homologación de estos correspondiente a cada columna.

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 5436 entries, 0 to 5483
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   car_name               5436 non-null   object
1   car_prices_in_rupee    5436 non-null   float64
2   kms_driven             5436 non-null   int64
3   fuel_type              5436 non-null   object
4   transmission           5436 non-null   object
5   ownership              5436 non-null   object
6   manufacture            5436 non-null   object
7   engine                 5436 non-null   int64
8   Seats                  5436 non-null   int64
dtypes: float64(1), int64(3), object(5)
memory usage: 424.7+ KB
None
```



```
#transformar datos de crore a lakhs
filtro= df[df.car_prices_in_rupee.str.contains('Crore')]
filtro['car_prices_in_rupee']= filtro['car_prices_in_rupee'].str.replace(' Crore', '').astype(float)
filtro['car_prices_in_rupee']= filtro['car_prices_in_rupee']* 100
filtro.info()
```

```
#Quitar Crore y limpiar car_prices
df.drop(df[df.car_prices_in_rupee.str.contains('Crore')].index, inplace=True)
df['car_prices_in_rupee']= df['car_prices_in_rupee'].str.replace(',','.')
df['car_prices_in_rupee']= df['car_prices_in_rupee'].str.replace(' Lakh', '').astype(float)
df.info()
```

```
#Limpiar variables
```

```
#kms_driven
```

```
df['kms_driven']= df['kms_driven'].str.replace(',','')
df['kms_driven']= df['kms_driven'].str.replace(' kms', '').astype(int)
```

```
#Seats
```

```
df['Seats']= df['Seats'].str.replace(' Seats', '').astype(int)
```

```
#engine
```

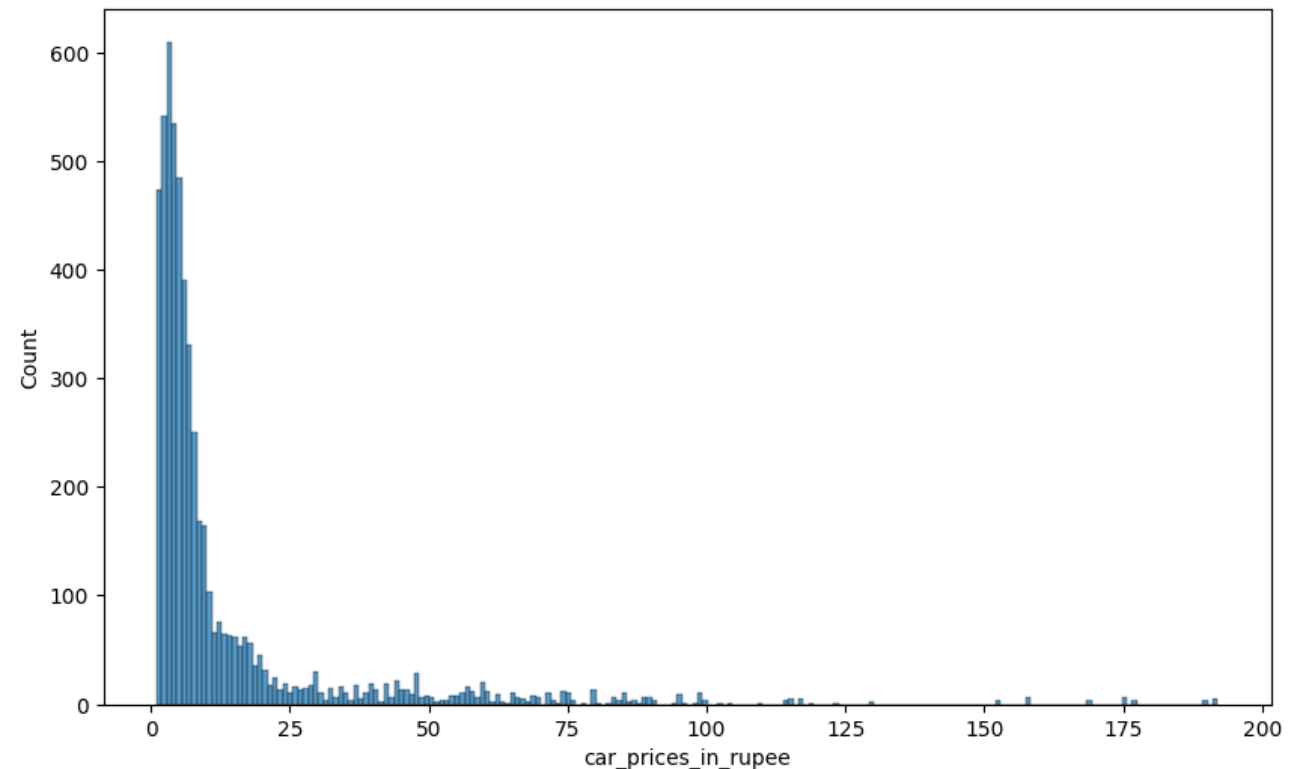
```
df['engine']= df['engine'].str.replace(' cc', '').astype(int)
```

```
#manufacture
```

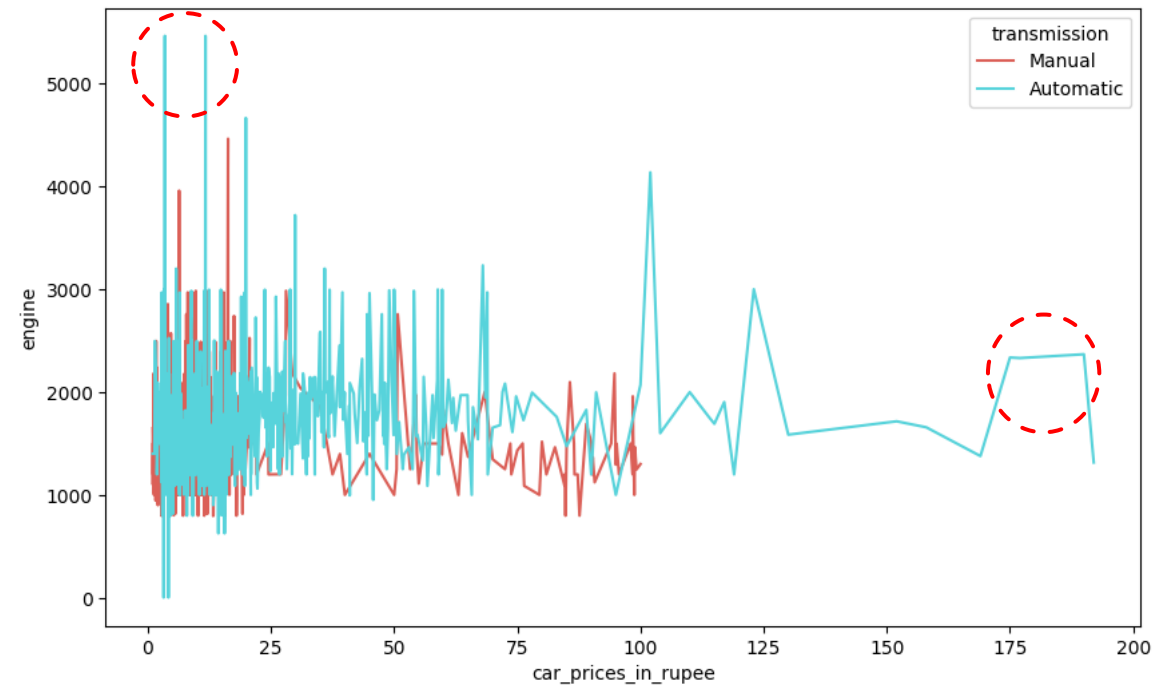
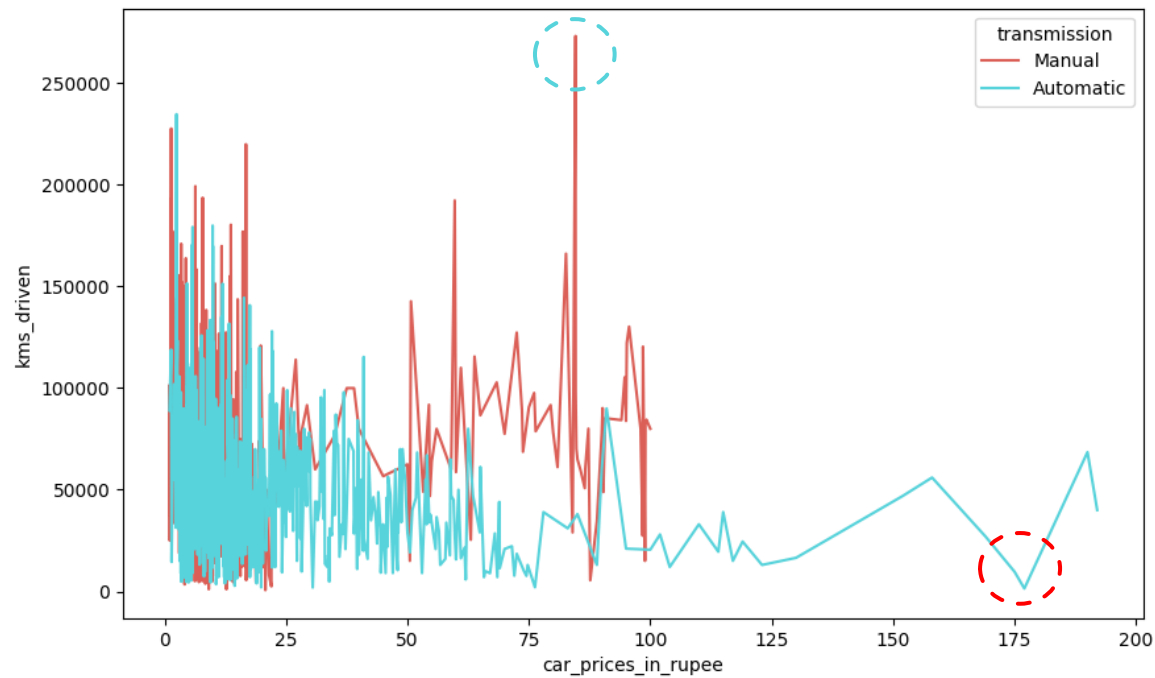
```
df['manufacture']= df['manufacture'].astype(str)
```

# Visualizaciones

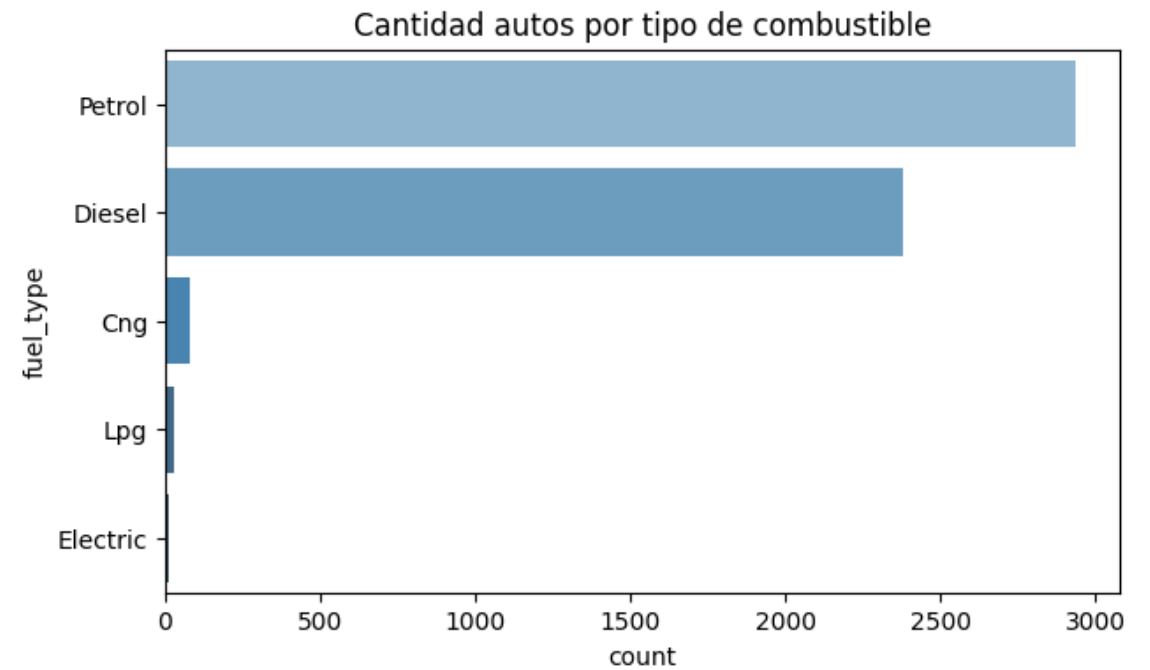
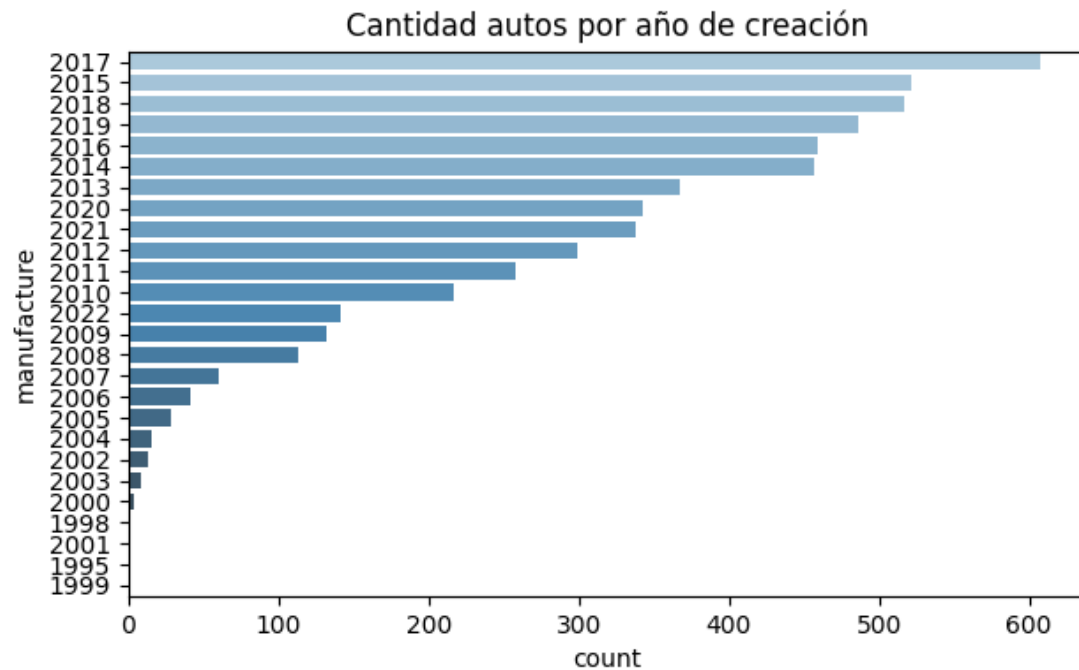
La variable objetivo para el modelo es el precio del automóvil, el cual podemos observar que su frecuencia esta concentrada a la izquierda, y su cola a la derecha muestra una gran desviación positiva de los datos, por lo cual revisaremos como se comportan los otros datos en relación al precio.



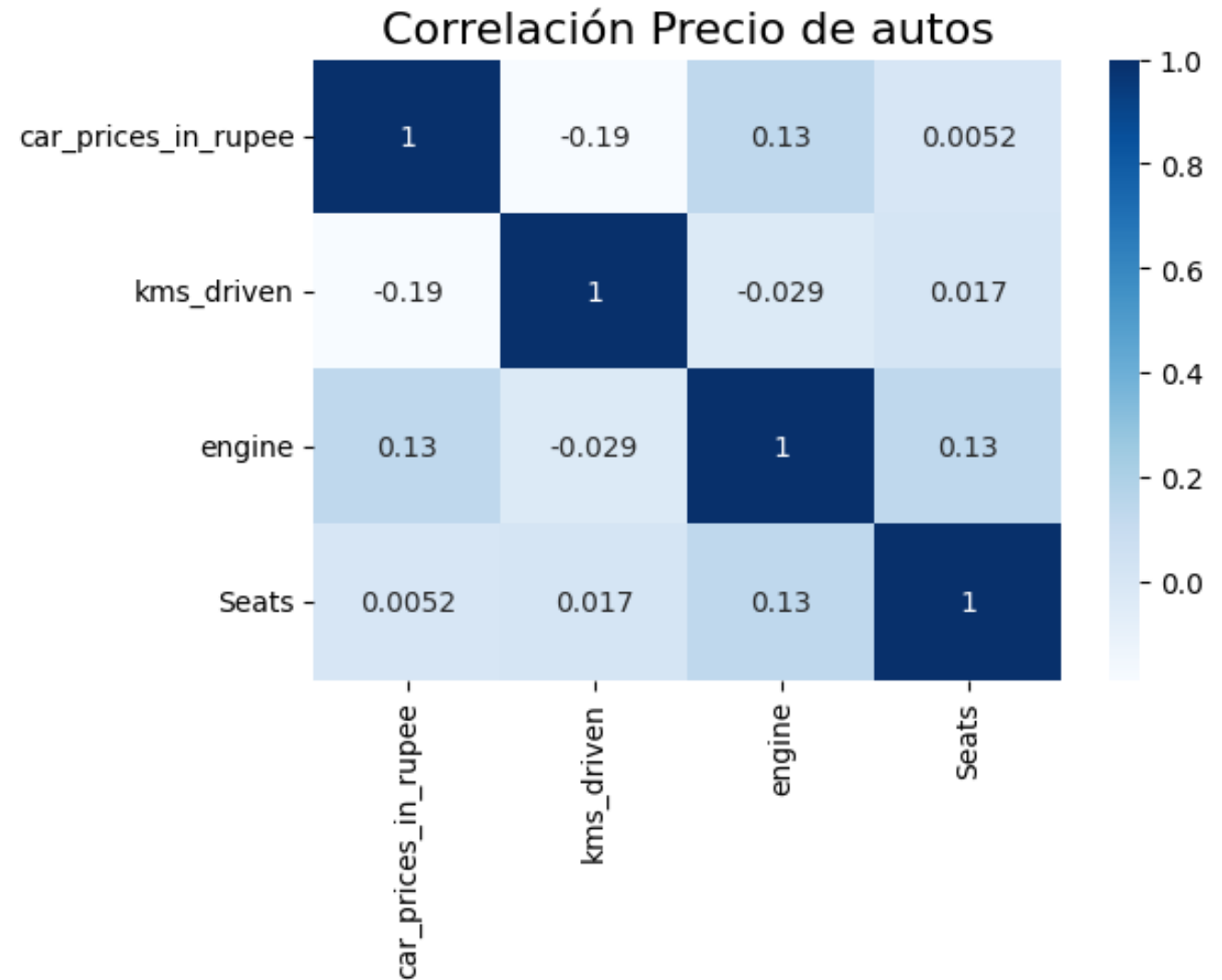
A continuación podemos observar que la mayoría de autos tienen un kilometraje entre 0 y 100.000, además de un motor entre 1000 y 3000, para los cuales los mas costosos son de transmisión automática, y entre menor sea el kilometraje y se tenga un motor mas pequeño, mas costoso será el auto.



Otras características que pueden aportar valor los autos es su año de creación y tipo de combustible, lo cual vemos que la mayor cantidad de estos funciona con Petrol y son de los años 2018, 2015 y 2017.



Finalmente podemos observar que variables impactan al precio de los autos, donde aunque las correlaciones son bajas, tenemos que el kilometraje y el motor son las mas representativas de las variables numéricas.





# Fuente de Datos

- ▶ <https://www.kaggle.com/datasets/milanvaddoriya/old-car-price-prediction>
- ▶ <https://www.cardekho.com/used-car-details#usedCarCity>