

Car pice predictrion Using Linear Regression

This is formatted as code

1) Import Library

```
import pandas as pd
```

```
import numpy as np
```

2) Import CSV as Dataframe

```
df = pd.read_csv(r'https://github.com/YBI-Foundation/Dataset/raw/main/Car%20Price.csv')

```

3) Get first five rows of data Frame

```
df.head()
```

	Brand	Model	Year	Selling_Price	KM_Driven	Fuel	Seller_Type	Transmission
0	Maruti	Maruti 800 AC	2007	60000	70000	Petrol	Individual	Manual
1	Maruti	Maruti Wagon R LXI Minor	2007	135000	50000	Petrol	Individual	Manual
	Hyundai							

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4340 entries, 0 to 4339
Data columns (total 9 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Brand           4340 non-null  object
1   Model           4340 non-null  object
2   Year            4340 non-null  int64
3   Selling_Price   4340 non-null  int64
4   KM_Driven       4340 non-null  int64
```

```

5   Fuel      4340 non-null  object
6   Seller_Type  4340 non-null  object
7   Transmission  4340 non-null  object
8   Owner       4340 non-null  object
dtypes: int64(3), object(6)
memory usage: 305.3+ KB

```

```
df.describe()
```

	Year	Selling_Price	KM_Driven
count	4340.000000	4.340000e+03	4340.000000
mean	2013.090783	5.041273e+05	66215.777419
std	4.215344	5.785487e+05	46644.102194
min	1992.000000	2.000000e+04	1.000000
25%	2011.000000	2.087498e+05	35000.000000
50%	2014.000000	3.500000e+05	60000.000000
75%	2016.000000	6.000000e+05	90000.000000
max	2020.000000	8.900000e+06	806599.000000

▼ Get catagories and counts of catagorical variables

```
df[['Brand']].value_counts()
```

```

Brand
Maruti      1280
Hyundai     821
Mahindra    365
Tata        361
Honda       252
Ford        238
Toyota      206
Chevrolet   188
Renault     146
Volkswagen  107
Skoda       68
Nissan       64
Audi        60
BMW         39
Fiat        37
Datsun      37
Mercedes-Benz 35
Mitsubishi  6
Jaguar      6
Land        5
Ambassador  4
Volvo       4
Jeep        3
OpelCorsa   2

```

```

MG                2
Isuzu             1
Force            1
Daewoo           1
Kia              1
dtype: int64

```

```
df.columns
```

```

Index(['Brand', 'Model', 'Year', 'Selling_Price', 'KM_Driven', 'Fuel',
      'Seller_Type', 'Transmission', 'Owner'],
      dtype='object')

```

```
df[['Model']].value_counts()
```

```

Model
Maruti Swift Dzire VDI          69
Maruti Alto 800 LXI            59
Maruti Alto LXi                47
Hyundai EON Era Plus          35
Maruti Alto LX                 35
..
Mahindra KUV 100 G80 K4 Plus    1
Mahindra KUV 100 mFALCON D75 K8 1
Mahindra KUV 100 mFALCON D75 K8 AW 1
Mahindra KUV 100 mFALCON G80 K2 Plus 1
Volvo XC60 D5 Inscription      1
Length: 1491, dtype: int64

```

```
df[['Fuel']].value_counts()
```

```

Fuel
Diesel      2153
Petrol      2123
CNG         40
LPG         23
Electric     1
dtype: int64

```

```
df.columns
```

```

Index(['Brand', 'Model', 'Year', 'Selling_Price', 'KM_Driven', 'Fuel',
      'Seller_Type', 'Transmission', 'Owner'],
      dtype='object')

```

```
df[['Seller_Type']].value_counts()
```

```

Seller_Type
Individual      3244
Dealer          994
Trustmark Dealer 102
dtype: int64

```

```
df[['Transmission']].value_counts()
```

```
Transmission
Manual      3892
Automatic   448
dtype: int64
```

```
df[['Owner']].value_counts()
```

```
Owner
First Owner      2832
Second Owner     1106
Third Owner       304
Fourth & Above Owner  81
Test Drive Car    17
dtype: int64
```

```
df.shape
```

```
(4340, 9)
```

► Get encoding Of categorical features

```
[ ] ↳ 6 cells hidden
```

► Define independent variable(X) and dependent variable(Y)

```
[ ] ↳ 5 cells hidden
```

► Get train test split

```
[ ] ↳ 4 cells hidden
```

► Get model train

```
[ ] ↳ 6 cells hidden
```

► Model Evaluation

```
[ ] ↳ 5 cells hidden
```

► Get Visualisation of Actual Vs Predicted Results

[] ↳ 1 cell hidden

► Get Future Predictions

[] ↳ 6 cells hidden