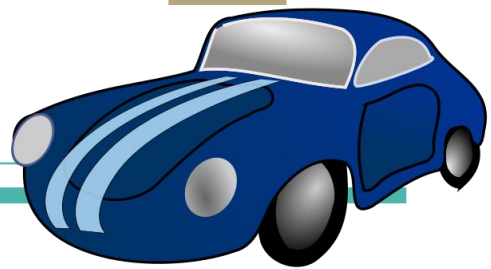
Used Car Price Predictions

By: Anu Joshi



Background

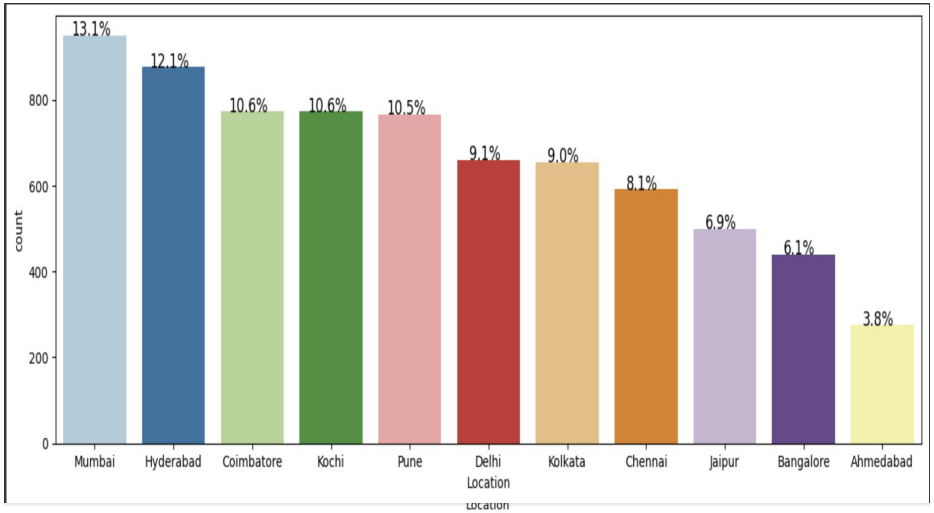
This model takes data from 1996 to 2019 which is a significant period in car sales in particular used cars in the Indian Market. For the year 2018 -2019 new car sales were recorded at 3.6 million units and around 4 million pre-owned cars were sold. The models suggest that pre-owned car sales will keep growing. It is recommended that stakeholders consider car sale, price, fuel type, mileage, brand of the car and city that it was sold in when building a long term model. All these factors can influence the price of a pre owned car.

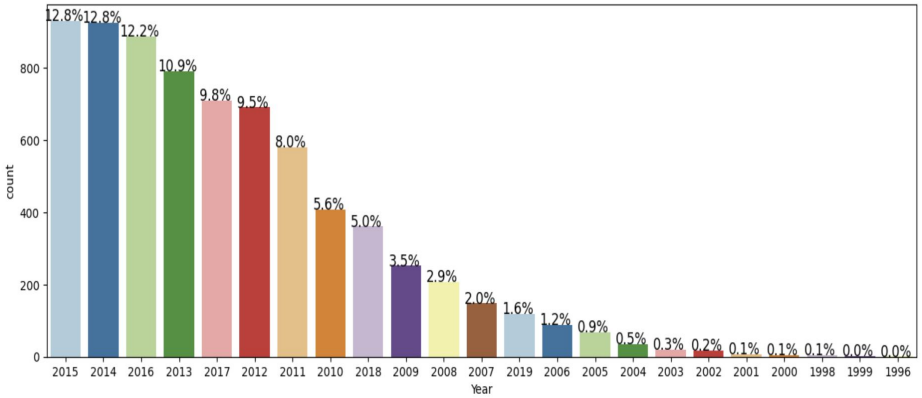
In India the middle class is growing and so it the amount of cars being purchased. Also there is more accessibility to vehicle financing for used cars.

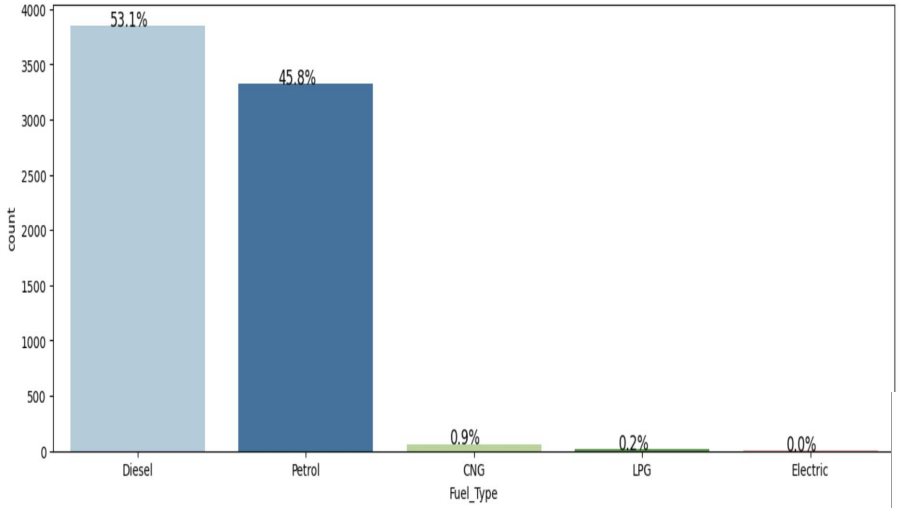
Exploratory Data Analysis

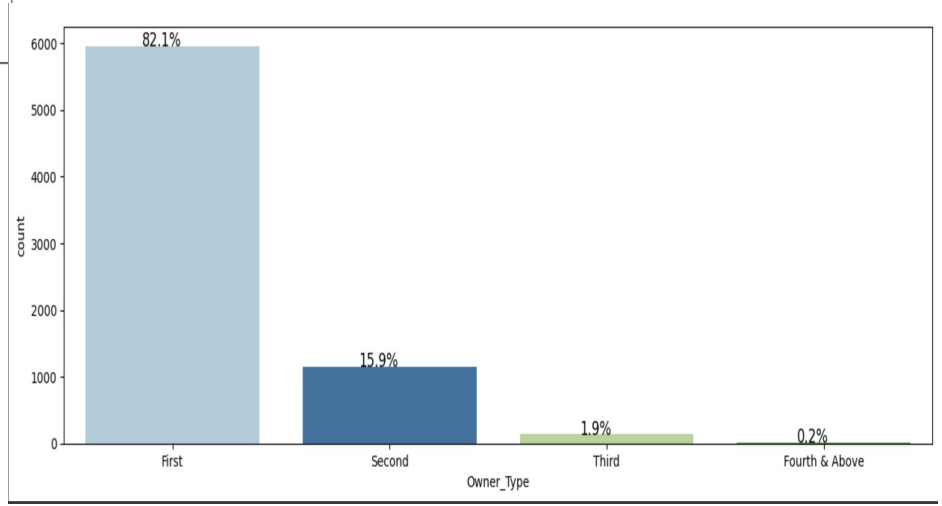
Top 5 car brands sold

1. Maruti
2. Hyundai
3. Honda
4. Toyota
5. Mercedes-benz.



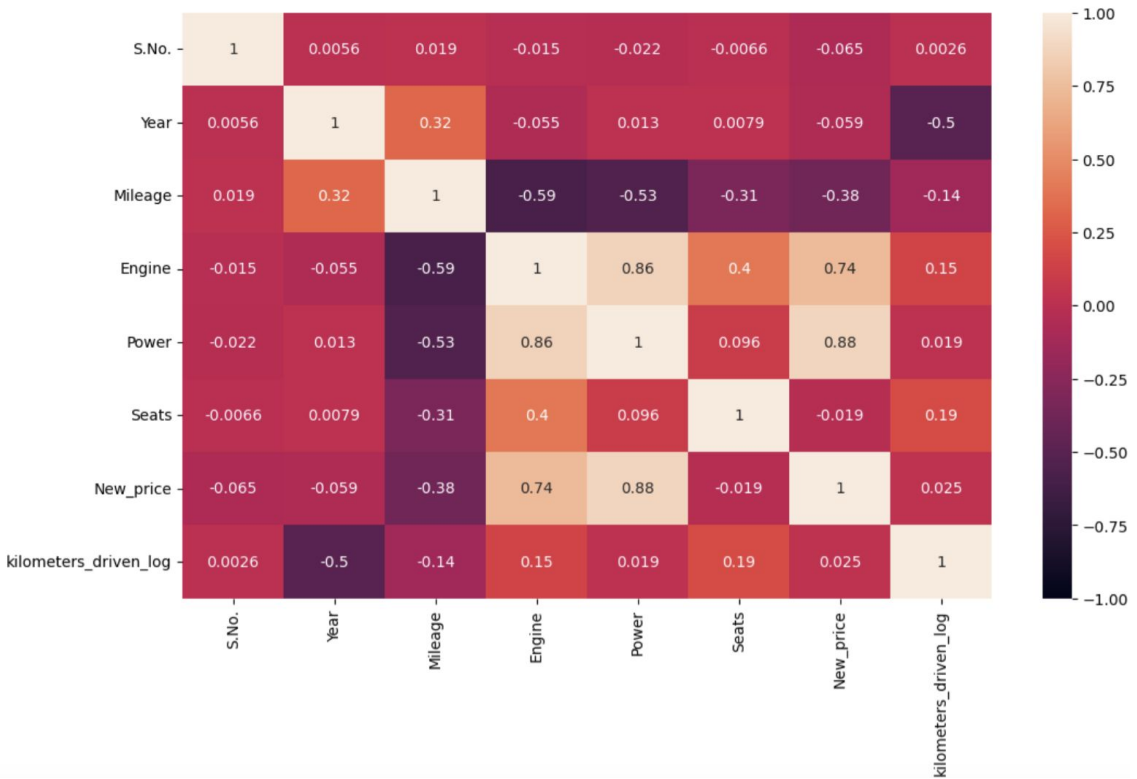



4000
3500
3000
2500
2000
1500
1000
500
0
count
53.1%
45.8%
0.9%
0.2%
0.0%
Diesel
Petrol
CNG
LPG
Electric
Fuel_Type


6000
5000
4000
3000
2000
1000
0
count
82.1%
15.9%
1.9%
0.2%
First
Second
Third
Fourth & Above
Owner_Type

Heat map

Mileage and year are positively correlated, which would be due to technology advancement.



Fitting the linear model

- R squared test and training data are similar
  - No overfitting of the data
- RMSE score is high

```
R-sqaure on training set :  0.9015176992395979
R-square on test set :  0.8969837054583856
RMSE on training set :  4.476108975988494
RMSE on test set :  4.1208018789612675
```

Lasso

- R squared value is poor and indicates underfitting
- RMSE is high

```
R-sqaure on training set :  0.17056604251242058
R-square on test set :  0.663796997874941
RMSE on training set :  12.990099096533363
RMSE on test set :  7.4443976999332095
```

Decision Tree

- R-squared on training set is 0.999 which is almost perfect.
- R-squared on testing set is 0.88 which is less than training set, which is suggesting little overfitting.
- RMSE on training set is 0.005 which is very good.
- RMSE on test set is 3.5 is more than training set, confirming overfitting.
-

```
R-sqaure on training set :  1.0
R-square on test set :  0.9294288004616825
RMSE on training set :  3.7593045898076 49e-15
RMSE on test set :  3.41069215608893
```

Random Forest

- R squared values are the same.
- Lowest RMSE value score out of all the test models.
- Hypertune the models further

```
R-sqaure on training set :  0.9732463824943884
R-square on test set :  0.9741118599476957
RMSE on training set :  2.3329896665032606
RMSE on test set :  2.0657568303333953
```

Conclusion

- The most bought car in India is the Maruti due to how affordable it is in India to Hyundai,Toyota, Volvo, Jaguar. The company Cars4U should focus more car dealerships in bigger cities where there are more people like Mumbai. There are people who are more likely to buy a second hand car than a new one. That information might not be readily available in the database.
- Location data and different car types should be compared in K- medoids cluster analysis to see if more trends can be noticed.