CAR PRICE PREDICTION

Submitted by

Team Name: Pandavs

Members: 1. Kishore Kunal (MT2018051)

2. Nilesh Singh (MT2018069)

3. Karuna Nidhan Upadhyay (MT2018048)

Abstract

Observing the relation between the marked price(this price is different from actual selling price) which the automobile manufacturers gives to the car produced , with the different features of the car (ie automobile company,car model,year of production,engine horsepower etc) and tried to predict the marked price of the cars produced by the different automobile companies.

Data URL: https://www.kaggle.com/CooperUnion/cardataset

Variables Used For Analysis

This data set contains the following features:

- Make : Automobile manufacturers
- Model : particular model of a car
- Year : year of manufacturing
- Engine Fuel Type
- Engine_HP : horse power of engine
- Engine Cylinders : number of cylinder
- Transmission Type
- Number of Doors
- Vehicle Size
- Highway MPG: miles with 1 gallon car travel at highway
- City MPG: miles with 1 gallon car travel in city
- Popularity

Target Variable:

- MSRP : Manufacture Suggested Retail Price

Exploratory Data Analysis (EDA)

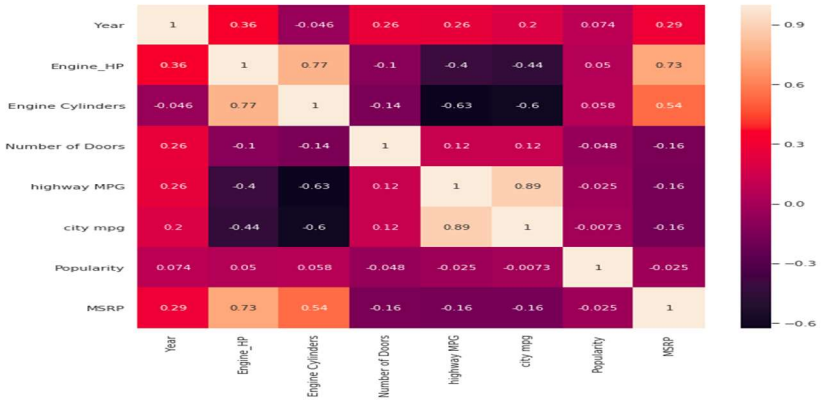

Fig: Correlation between various features and MSRP

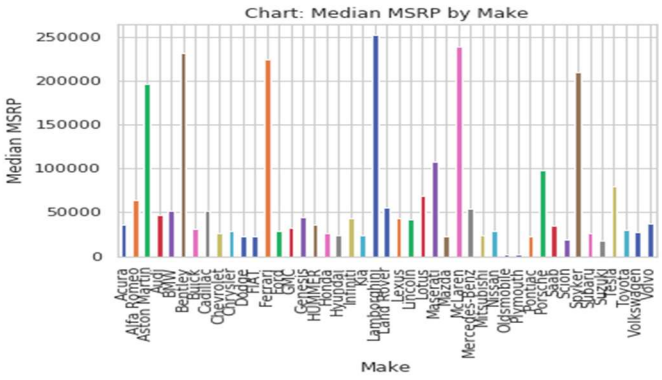

Fig:Bar Chart between Median of MSRP and Make

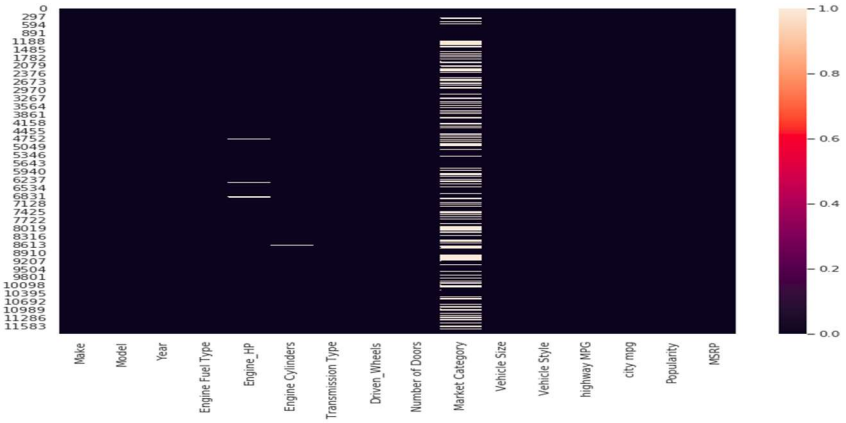

Fig:Null value estimation

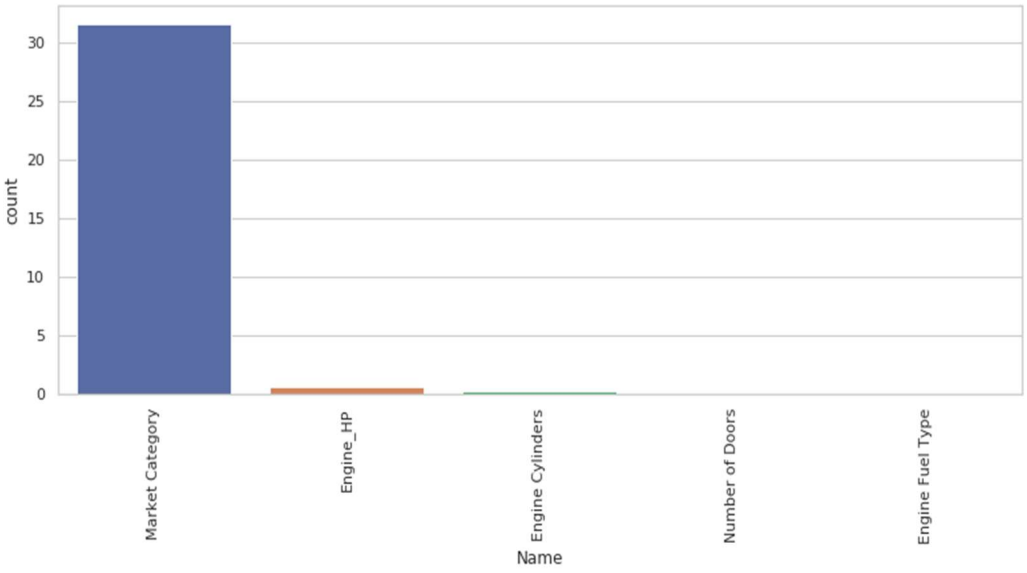

Fig: Count of null values in features

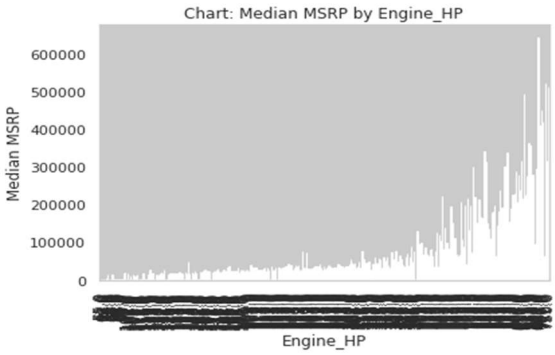

Fig: correlation between Engine hp and MSRP

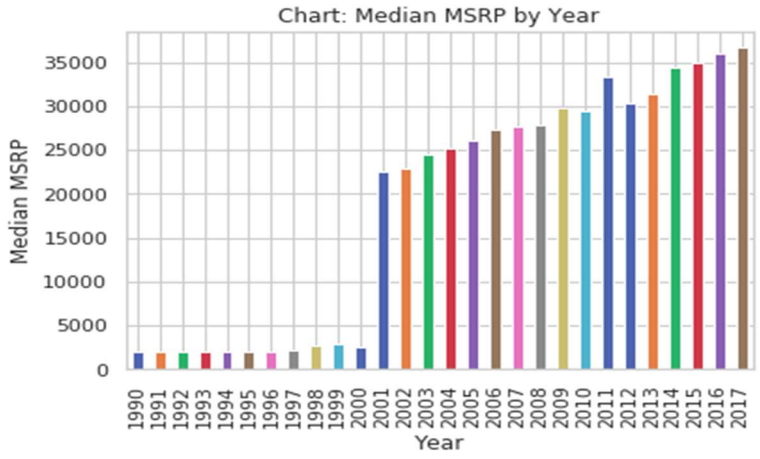

Fig: Correlation between Median Msrp and Year

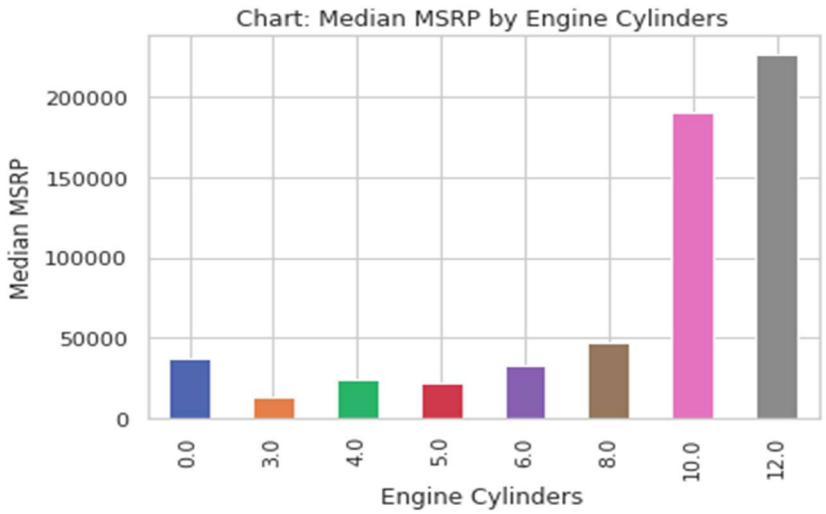

Fig:Correlation between Median MSRP and Engine Cylinders

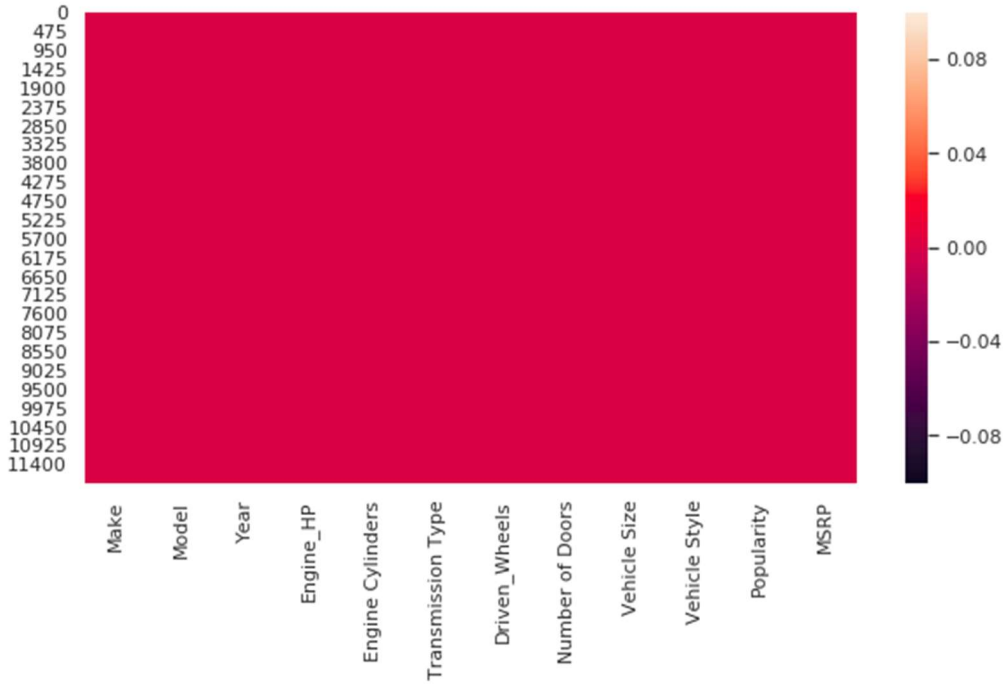

Fig: heat map after null removal

Approach

- First we found missing values and visualized that features to observe it's pattern and accordingly fill the missing values.
- We visualize relationship between features by finding correlation matric.
- We observe outliers using correlation metric and boxplot and drop some of the columns or tried to remove outliers.
- We even observed the pattern using grouping on single feature and observe it's pattern.
- Then, we applied linear models and then XGBoost.
- We performed parameter tuning to find best parameters and applied cross validation to better fit the model.

Conclusion

1. With linear regression the marked price prediction of the car manufactured have the accuracy of 72.16 and RMSE value of 21975.
2. With ridge and lasso the marked price prediction of the car manufactured have the accuracy of 70 and RMSE value of 21900
3. With XGBoost the marked price prediction of the car manufactured have the accuracy of 98.49 and RMSE value of 5206
4. Even after applying cross validation and parameter tuning Ridge,Lasso doesn't prove to be good model.