



Car's Sale Price prediction using the Car Dekho Dataset

Pinal Patel

Academic Xi



Process

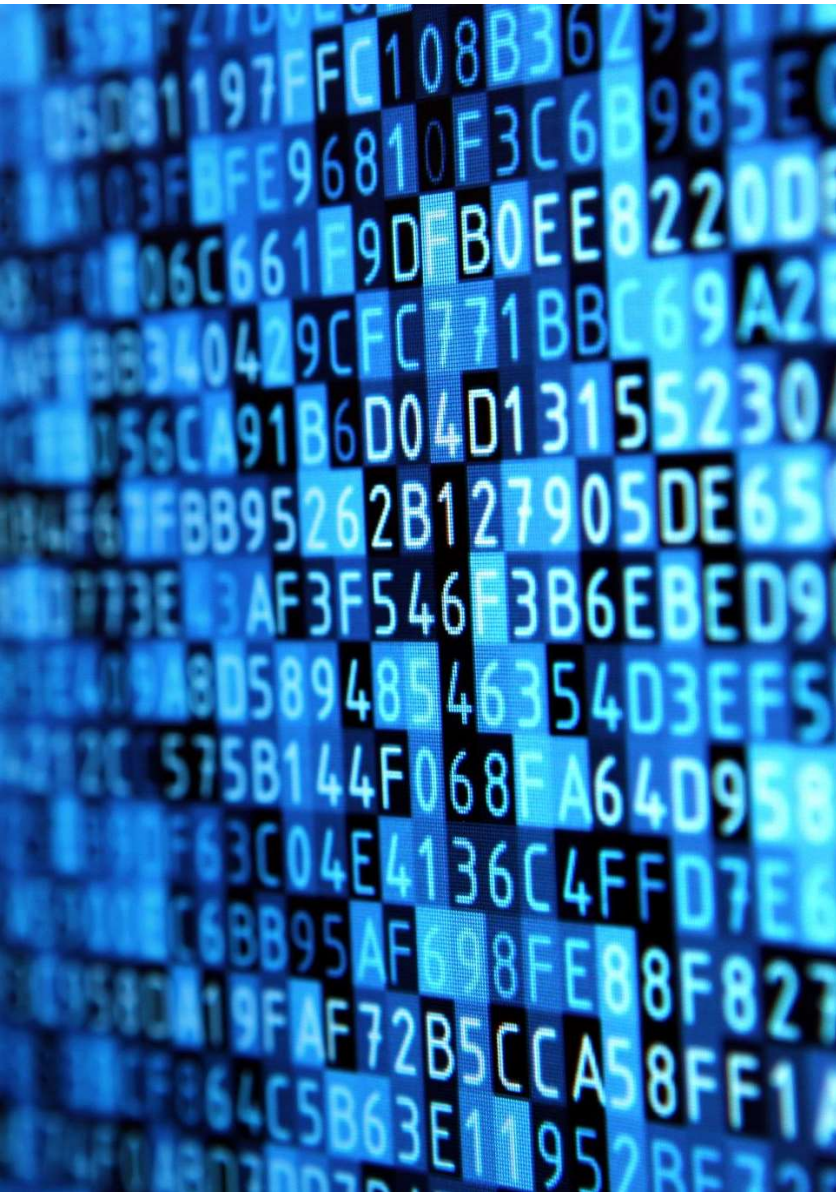
- The Data
- Data cleaning and pre-processing.
- Exploring the data
- Building regression models
- Validation of the model
- Result



Business Problem

- Car Dekho is one of the agency who used to buy second-hand cars and performs any necessary repairs, and then sells the vehicles. Now, The agency wanted to improve their cars sales, but they don't know that which features needs to improved, based on customer interest and wanted to know on what basis customers buy cars, and the predictors will affect client purchases.

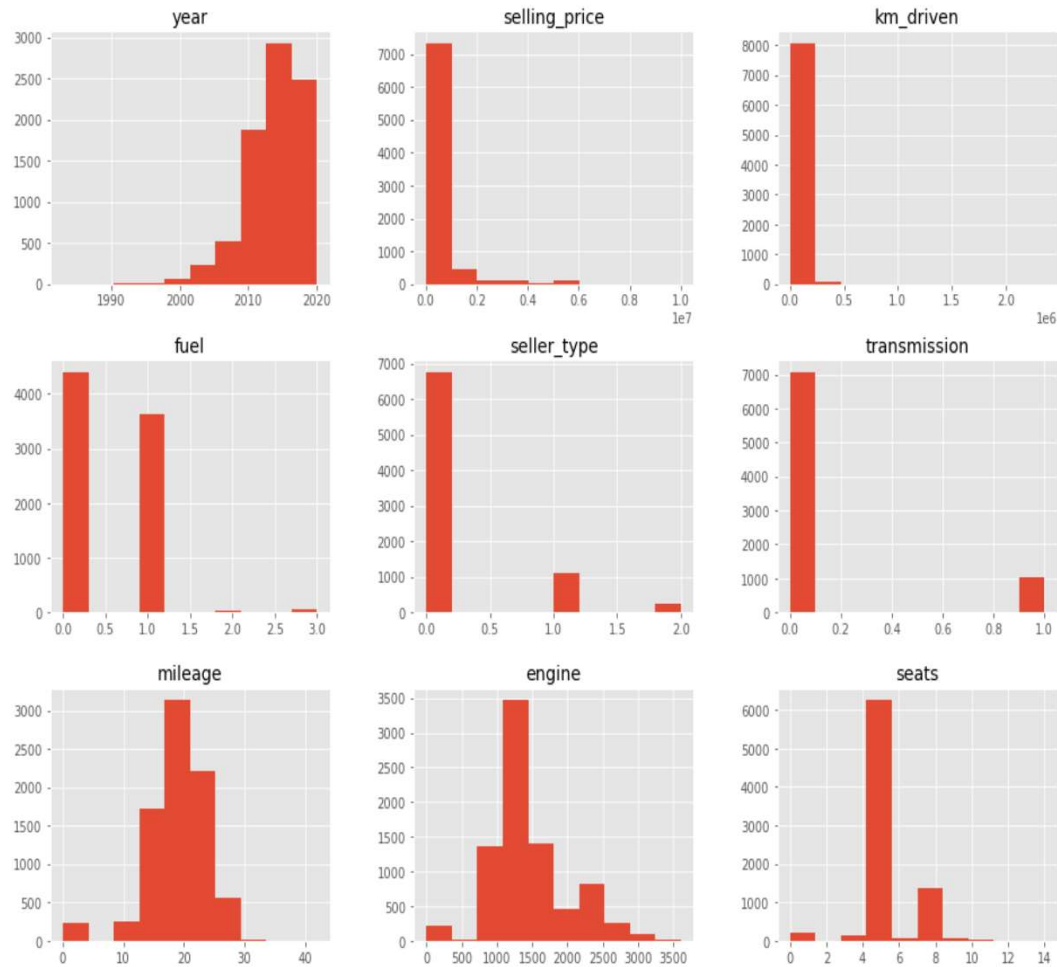


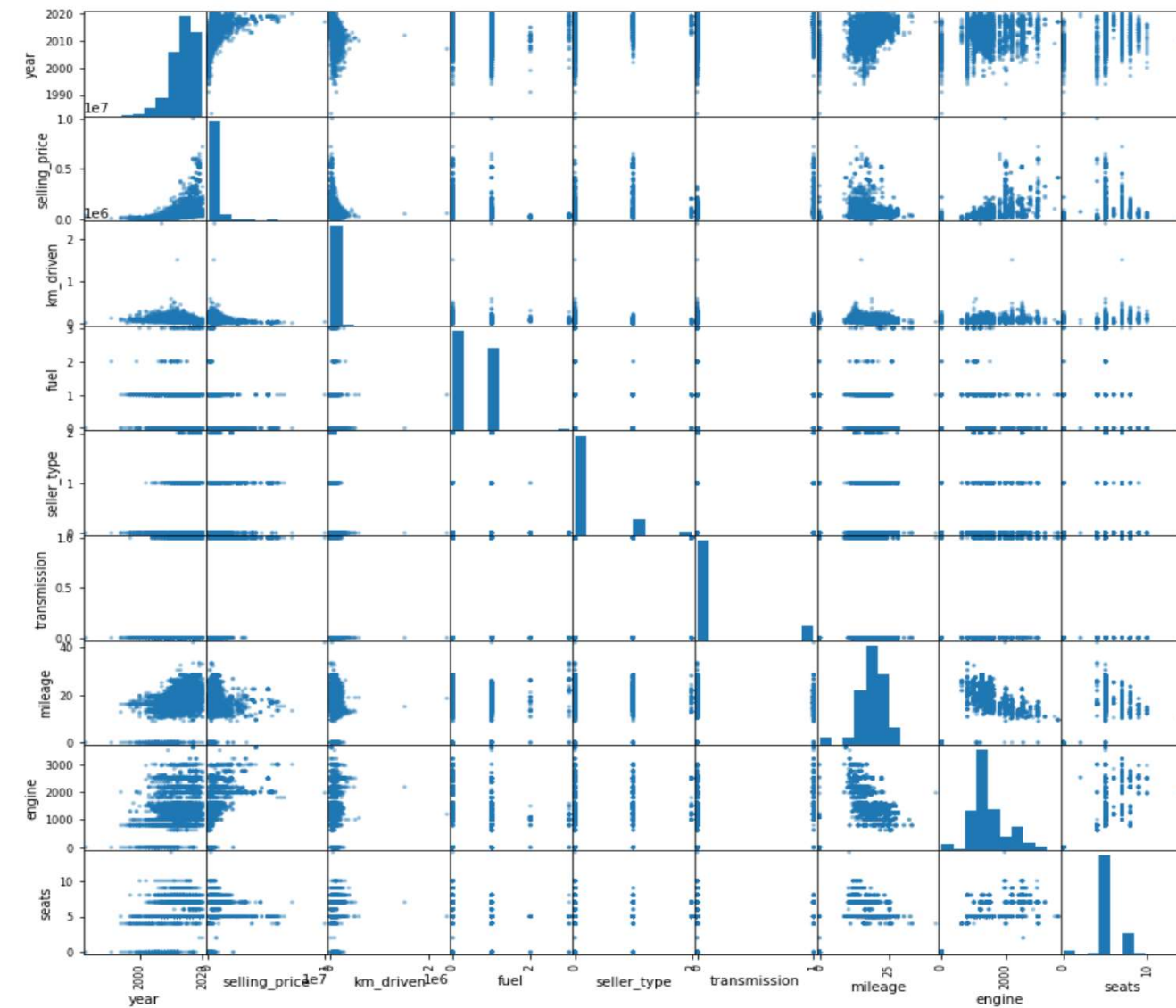


About the Data

- The data is in a CSV file which includes the following columns: model, year, selling price, showroom price, kilometers driven, fuel type, seller type, transmission, and number of previous owners, mileage, engine, Maxpower and seats.

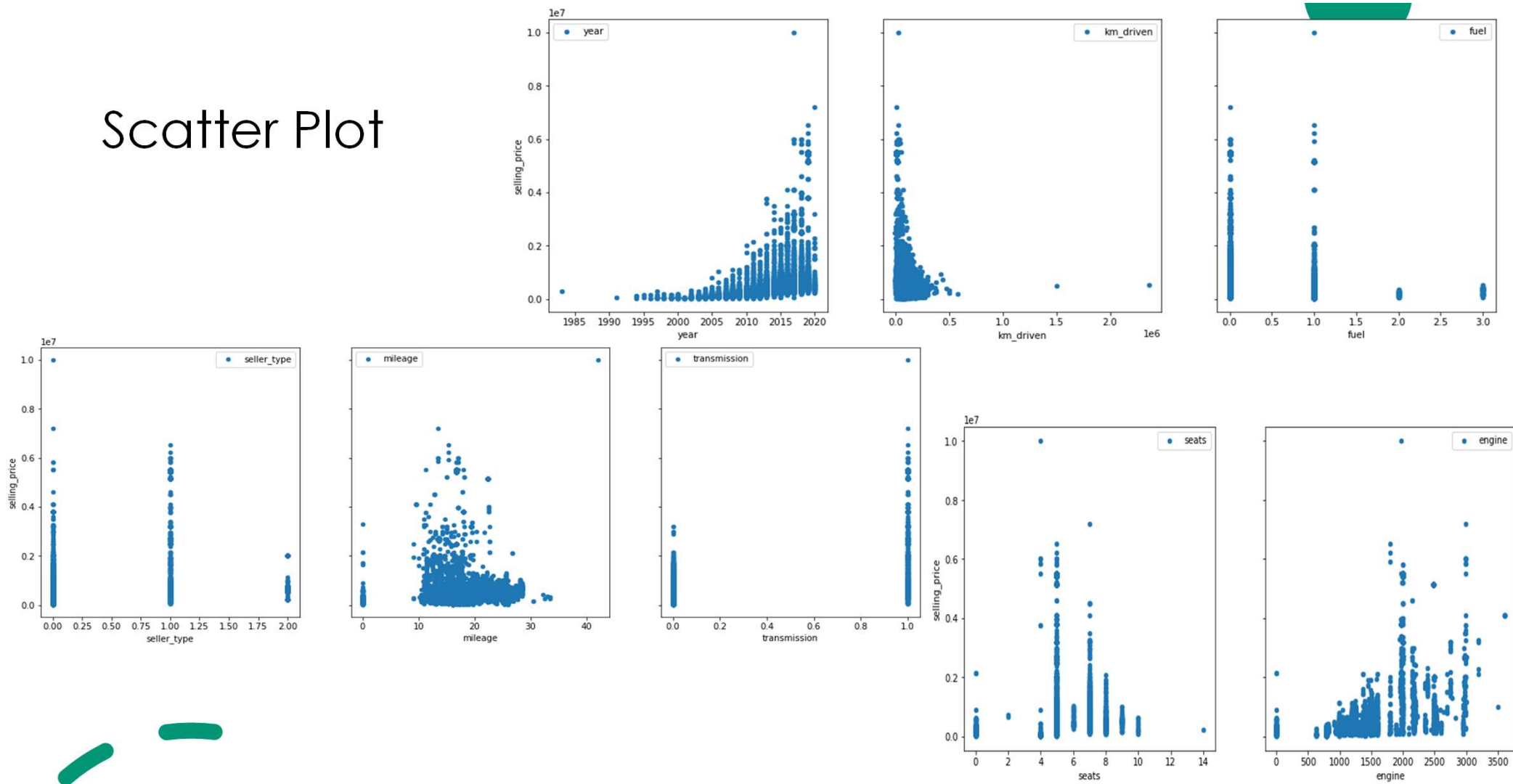
Let's have a
closer look of
the histogram.



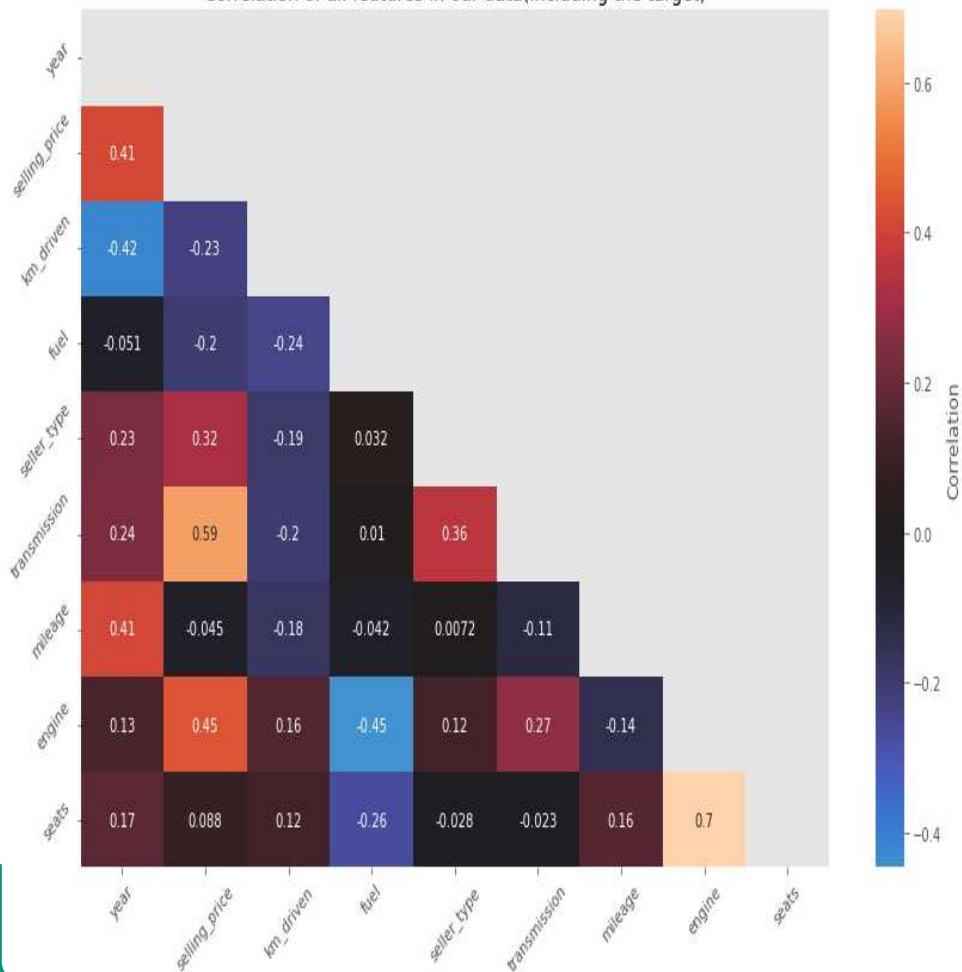


scatter
matrix

Scatter Plot

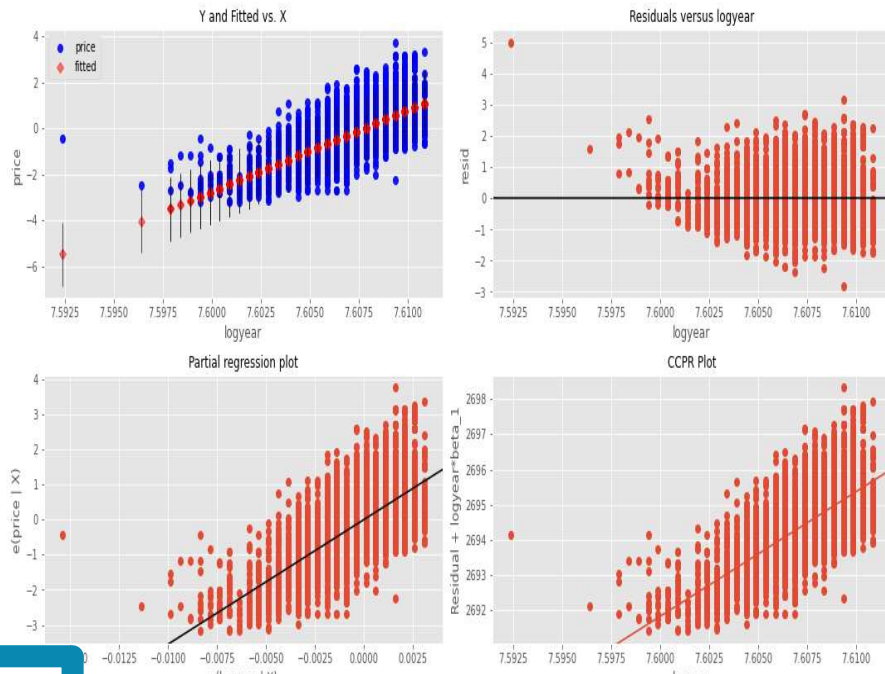


Correlation of all features in our data(including the target)



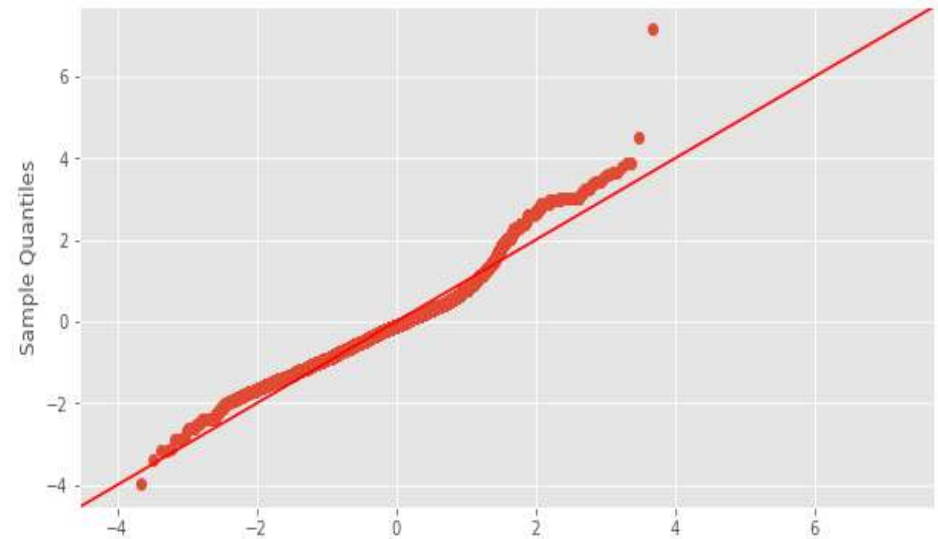
- Selling price 1.000000
- transmission 0.588845
- engine 0.445985
- year 0.414092
- seller_type 0.323047
- seats 0.088110
- mileage -0.045405
- fuel -0.202916
- km_driven -0.225534
- Name: selling price, dtype: float64

Regression Plots for logyear



Final Model

- R-squared: 0.507
- For all the variable P-value: 0.00





Model Validation

- Train Mean Squared Error:
 $3.6033202806429144e-32$
- Test Mean Squared Error:
 $3.8452496119908856e-32$



An aerial photograph of a parking lot filled with numerous cars, mostly in shades of blue, white, and grey, arranged in neat rows. A semi-transparent white box is overlaid on the right side of the image, containing text and a small green graphic.

Conclusions

- The best predictors which affected car prices is Year.
- And the second one is km driven.

The background features several abstract geometric shapes. On the left, there are two vertical teal bars, a teal square outline, and a dashed teal line. In the upper center, there is a teal circle. To its right is a large teal checkmark. Further right is a teal semi-circle. The bottom right portion of the image is dominated by a large green semi-circle.

Thank you

<https://github.com/pinalnikhil/cardekhomu-ltilinearregression-.git>



Any Question