**Report**

This project uses the car price dataset to predict the car price auction using linear regression models. This dataset contains a car detail e.g., color, odometer, year car type and others. We can use this information to find a relationship between the car price and others and feed it to the method to get the result of car price prediction.

**Dataset Analysis**

## This is a dataset that we use for analysis. Before we get the final dataset, we need to pass the preparation process. e.g., remove a null value, fill in missing data, clean unneeded data, etc.

Background pattern

Description automatically generated

## Below is a summary of the statistics of the dataset.

Table

Description automatically generated

This is the shape of the final dataframe to be used for the modelling. (91768 rows, 50 columns)

**Graphical user interface, text, application

Description automatically generated**

**EDA, Feature observation and hypothesis**

we are using two features from the car price dataset: Odometer and MMR. We could predict the following for each feature:

Chart, scatter chart

Description automatically generated

* The car with more MMR (higher MMR value) will be better. MMR is an indicator of wholesale prices. MMR is high means the market has more demand from the people. So, this will makes a car price is high as well.
* A car with a lower odometer will make the price down. The odometer will show the number of distances that the car was used. When the car is used a lot, the engine and the parts wear out more than the cars that are used less. This is reasonable for the car will cheaper when the odometer is more.

**Simple Linear Regression Report**

We use two methods to create a linear regression model which are polynomial feature and MinMax Scaler.

* **Polynomial features** are the creation of new input features based on the existing features.
* **MinMax Scaler** is transform features by scaling each feature to the same scale range between zero and one [0,1]

Both methods are standard methods to generate and scale data to be the same standard and can use prepared data to predict the result.

**Linear Regression with Lasso/Ridge Report**

* **Lasso Report:** We use alphas values between 1 and 0.001, The positional argument is 15. We decided to use the lowest RMSE value as below.

Table

Description automatically generated

Below is a result of Lasso from those parameters.

Table

Description automatically generated

* **Ridge Report**: We use alphas values between 1 and 0.001, The positional argument is 15. We decided to use the lowest RMSE value as below.

Table

Description automatically generated

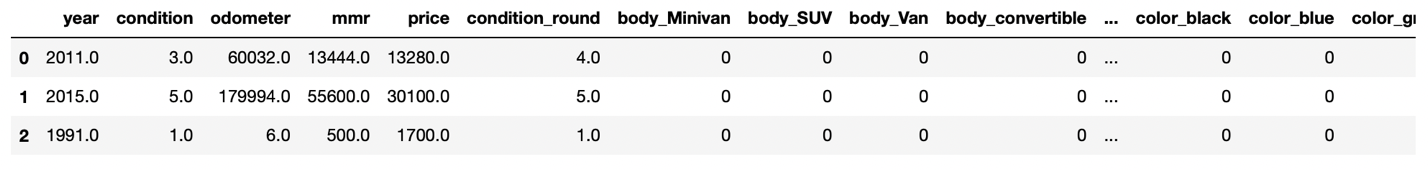
Below is a result of Lasso from those parameters.

Table

Description automatically generated

**Summary**

Here is the summary used from the newly created dataframe in step 9. When we process this dataframe with the selected model to predict the result.



The predicted price is somewhat inconsistent with the actual price. Because we don’t have enough data to predict. In my opinion, we need more data in the dataset to predict a better result.

Chart, scatter chart

Description automatically generated