Linear Regression

For used car market in UK

Presented by:
Ammer saleh
Naif alsulihem
Nasser Alshehri

December 2021

I. Introduction

The history of transportation starts from the age of humans and has continued to change over some time. The first means of transportation was the human foot. The history of transport has undergone a radical change with the introduction of wheels. Existing means of transportation were continuously improved after that, for example, horse-drawn vehicles (carts or karts). Until the gas and fuel cars were invented Cars are one of the most important necessities for every family, as they are the most popular means of transportation these days.

Use the predict to help us to figure out the UK used car market, and the data was collected from Kaggle.

II. Study Methodology

The methodology of this project is as follows, extracting data from Kaggle using cars in the UK and it was more than 33,000 rows and 9 columns.

In the cleanup step, we dropped the tax column because it was not included in our study, and we dropped missing values in the dataset and outliers.

We explored the data, applied comprehensive analysis methods to the data, and extracted important information from the data.

We will forecast UK used car market prices, by engine size, car model, transmission type, etc., by linear regression.

III. Data Description

The data set is provided in .csv format, contains information of price, transmission, model, mileage, fuel type, road tax, miles per gallon (mpg), and engine size.

The data set was extracted from Kaggle

Variables	Description
Model	In this column, the model of the car shows us the name of the car.(e.g., ford Kuga.)
Year	In this column, shows the year of manufacture of the car (e.g., 2012)
Price	This column shows the cost of car (e.g., 60000)
Transmission	It shows us the type of car transmission (e.g., Manual) .
Mileage	In this column, It shows us the distance the car has traveled in miles
Fuel type	in this column, it shows us the type of fuel (e.g., Petrol)
Engine	Size In this column, it shows us the engine capacity of the car
Mpg	In this column, it shows us the distance covered by the car per gallon
Tax	In this column, the tax amount charged on the car

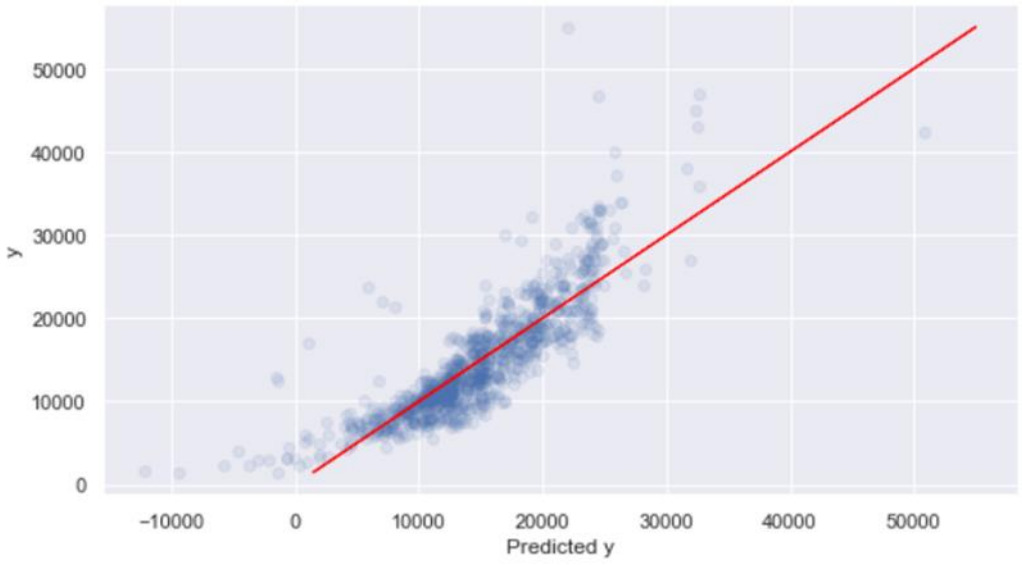
IV. Tools and Libraries:

- Python.
- Jupyter Notebook.
- PowerPoint.
- Excel
- NumPy.
- Pandas.
- Matplotlib
- sciket learn

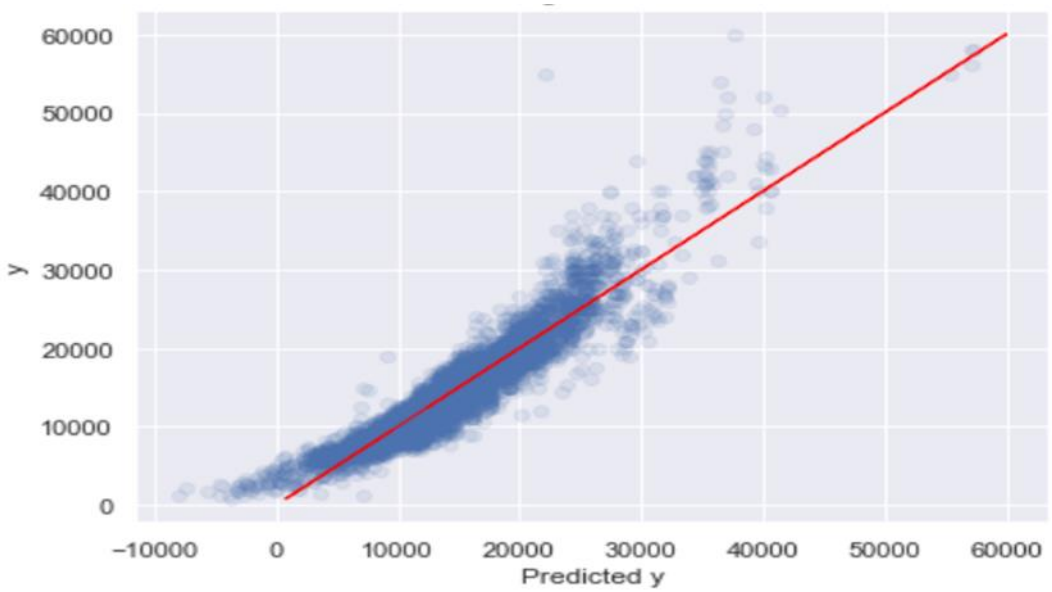
V. Linear Regression Model

Exp	Train	Val	Test
Baseline	75.21%	73.65%	74.11%
Dummy	88.30%	88.11%	87.95%
Log	92.45%	92.36%	92.21%
Coxbox	92.57%	92.37%	92.37%
STD	92.57%	92.38%	90.52%

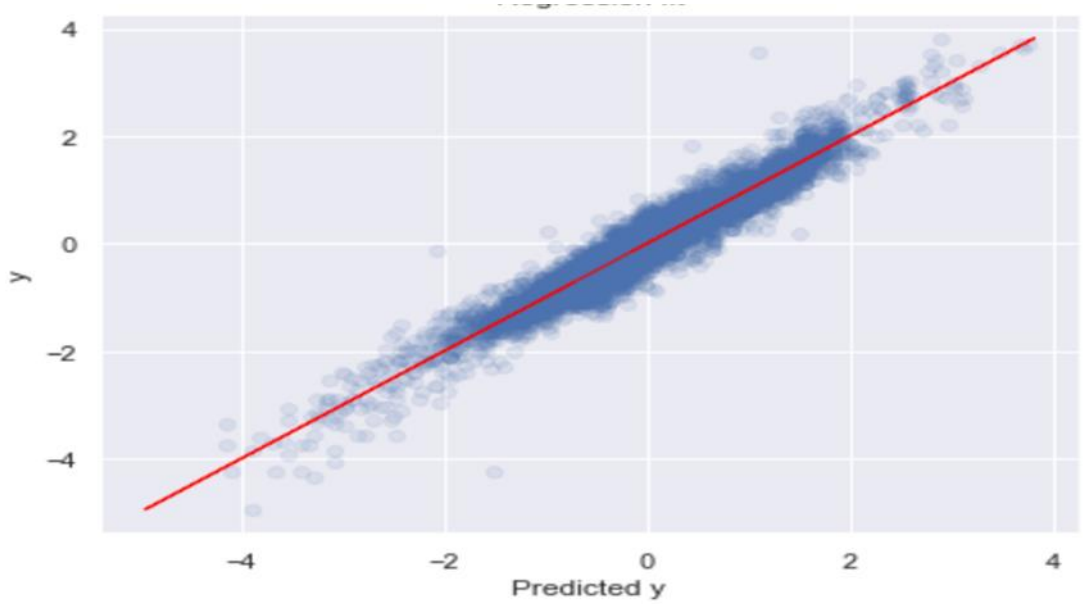
Baseline model plot :



Dummy model plot :



Coxbox model plot :



VI. Summary

- The model is improved after using the **Dummy Variable** function.
- We find The best model is **Coxbox Model.**
- So , our model ready to predict the prices of used car.

Reference :

[1] https://www.kaggle.com/adityadesai13/used-car-dataset-ford-and-mercedes