Final Project Report
Introduction to Data Analytics

Project Title:
Gas Prices Prediction

Prepared by:
Vrushank Amin (N01452389)

ITE 5201 – Summer 2022
Humber College

1. Problem Statement:

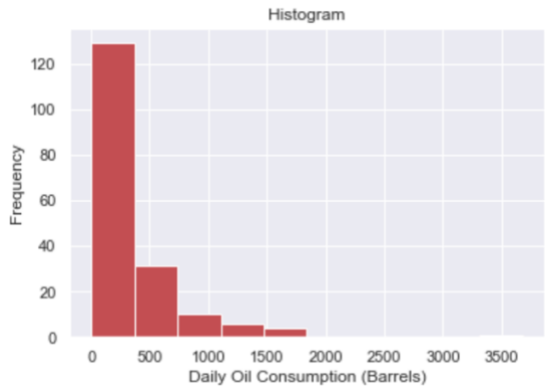
- Prediction of Gas Prices per litre (USD) based on several features in dataset.

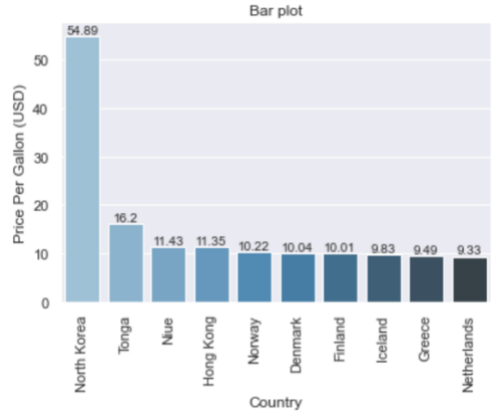
2. Dataset Description

- Gas prices per litre in USD using the world wide gas prices dataset contains daily usage and price information of gas such as world share of gas consumption by country, price per gallon, (USD), price per litre (USD), daily oil consumption (barrels), yearly gallons per capita

3. Dataset Analysis and Observations

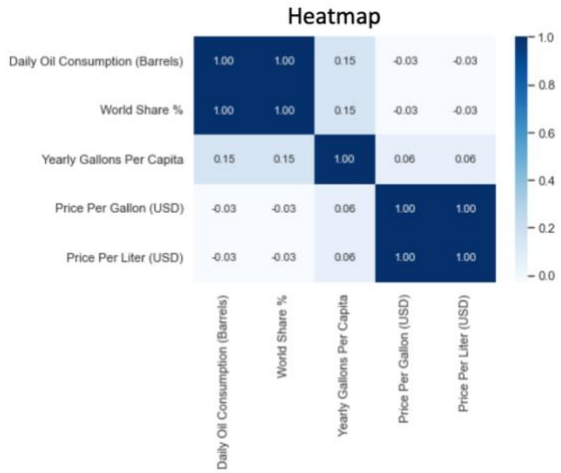
- For dataset analysis, I used the histogram, distribution plot, bar chart for univariate analysis, for bivariate analysis I have used the scatter plot, and heatmap for finding the correlation coefficient rank.
- I got the Daily Oil consumption frequency using histogram, and Top 10 countries having highest Gas Price per Gallon (USD) which are North Korea, Tonga, Niue, Hong Kong, Norway and more.



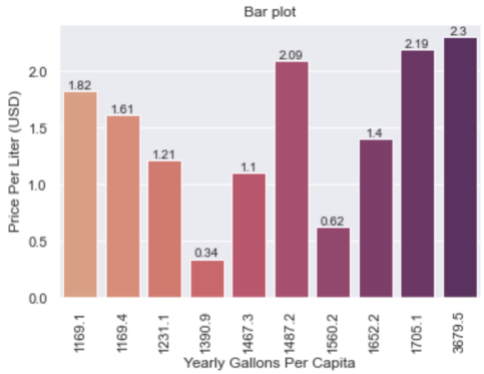


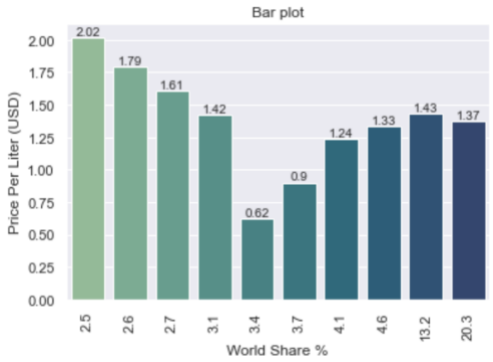
4. Proposed Analytical/Prediction Model

- Based on heatmap correlation coefficient rank I have selected the features.
- As per the analysis, I have some of the features on which output variable which is price per litre (USD) is dependent. So that's why I have used linear regression model for modelling the relationship between output variables and explanatory variables.

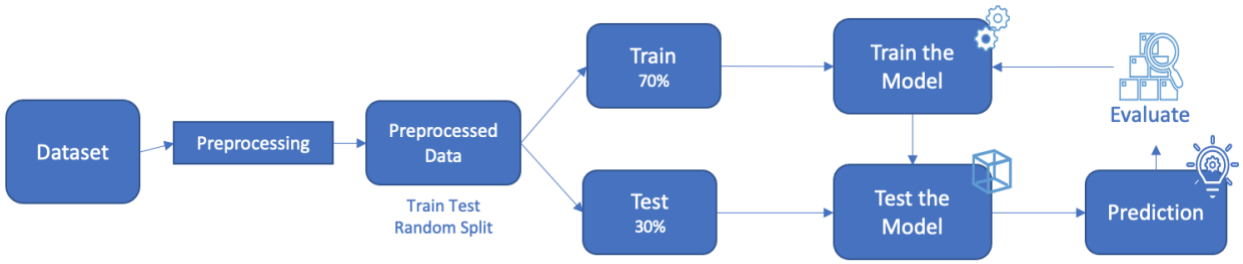


Relationships between inputs and outputs:



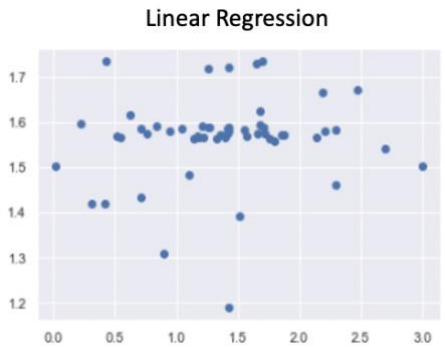


Model and design:



5. Results and Discussions

- I have trained both Linear and Polynomial Regression Model.
- Generated the scatter plot for the both predictive model,
- Using Mean Absolute Error(MAE), Mean Square Error(MSE), and Root-Mean Square Error, I can compare the both model efficiency and determine the best model for the dataset.




```
MAE: 0.5059378820057605
MSE: 0.418856542492768
RMSE: 0.6471912719534836
```


```
MAE: 0.5499413025185136
MSE: 0.5389062321839719
RMSE: 0.734102330866734
```

Conclusion: Based on the MAE, MSE and RMSE value, minimal error and more data accuracy found using Linear Regression Model. Linear Regression Model is better fit for the dataset compared to the Polynomial Regression Model.