# Forecasting Crude Oil and Gasoline Prices: Analysis of Time-Series Data from 1986-2016


## Project Group 6

**Arun Srivatsan Swaminathan, Kannan Walter, Mason Zepnick, Prageeshwar Chandran**

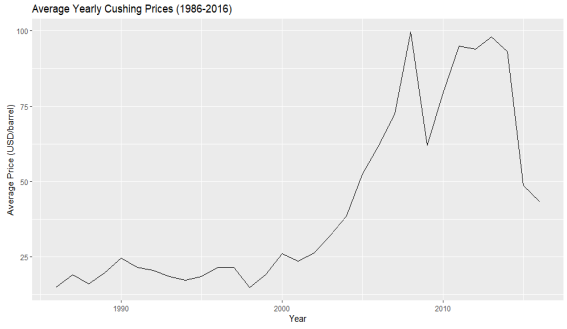

**Problem Statement:**
Crude oil and gasoline prices are critical factors in the global economy, as they impact various industries and have significant geopolitical implications. The primary goal of the analysis is to forecast future crude oil and gasoline prices based on historical data. This will require identifying any patterns, trends, or seasonality in the data and selecting appropriate forecasting methods to model these patterns. Finally, we want to evaluate the accuracy of our forecasts and identify any potential sources of error. We will analyse whether there are any seasonal patterns or trends in the data. We will check whether we can accurately predict future crude oil and gasoline prices based on historical data. In this report we will analyze crude oil and gasoline prices and visualize the trends using the R programming language.

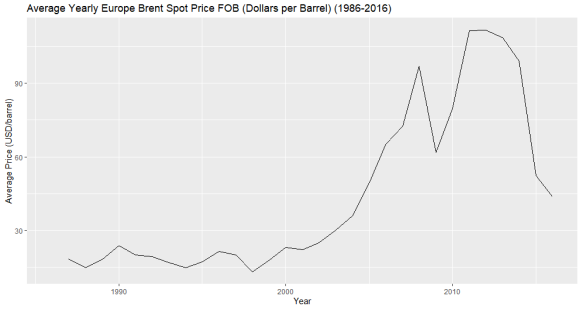
**Data and Methodology:**

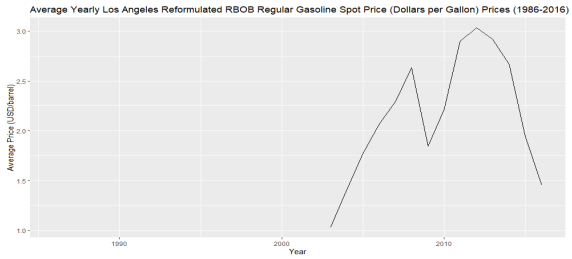
The data used in this analysis is loaded into R using the "read.csv" function, and the "tidyverse" and "forecast" libraries are loaded for data manipulation, visualization, and time series modeling. The data consists of yearly average prices for different types of crude oil and gasoline, including Cushing, Europe Brent Spot Price FOB, New York Harbor Conventional Gasoline Regular Spot Price FOB, U.S. Gulf Coast Conventional Gasoline Regular Spot Price FOB, and Los Angeles Reformulated RBOB Regular Gasoline Spot Price.

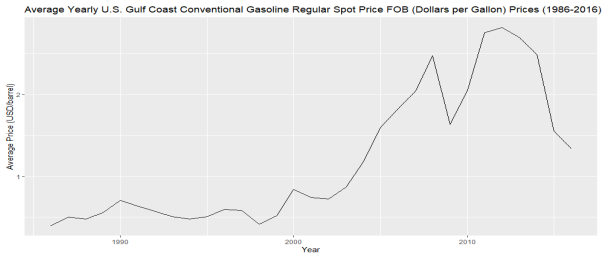
**Trend Analysis:**

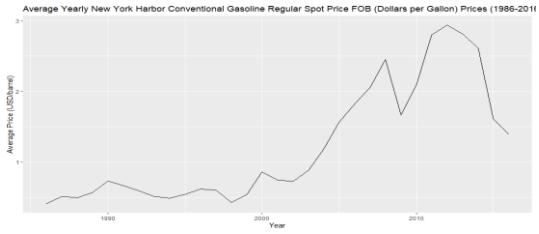
We start the analysis by obtaining summary statistics of the crude oil prices using the "summary" function in R. This provides us with key measures such as mean, median, minimum, maximum, and quartiles for each type of crude oil. Next, we visualize the trends in crude oil and gasoline prices using line plots. We use the "ggplot" function in R to create line plots for each type of crude oil and gasoline, with the years on the x-axis and the average prices on the y-axis. The line plots show the trends in crude oil and gasoline prices over the years, providing insights into price movements and patterns.



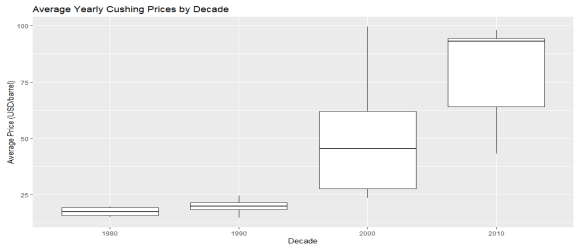


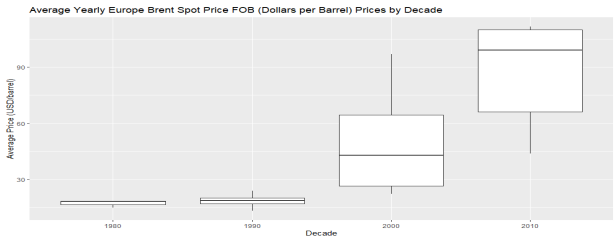



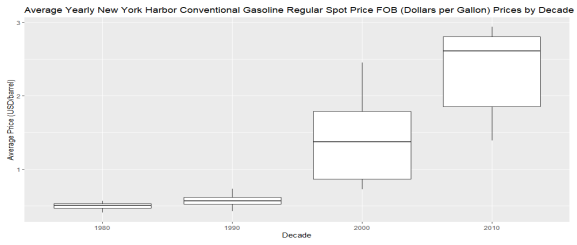



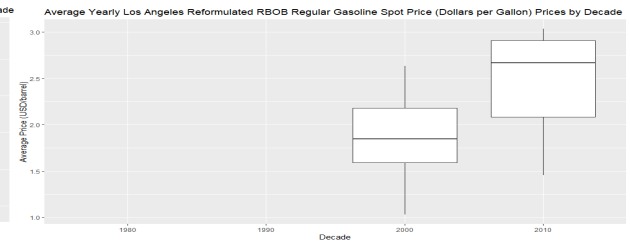

## Decade-wise Analysis:

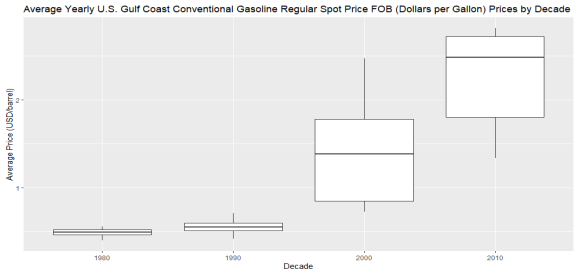
We further analyze the crude oil and gasoline prices by grouping them into decades using the "Decade" variable created in R. We use box plots to compare the average prices of crude oil and gasoline for each decade. The box plots provide information on the median, quartiles, and outliers of the data, allowing us to identify any significant differences in prices between decades.



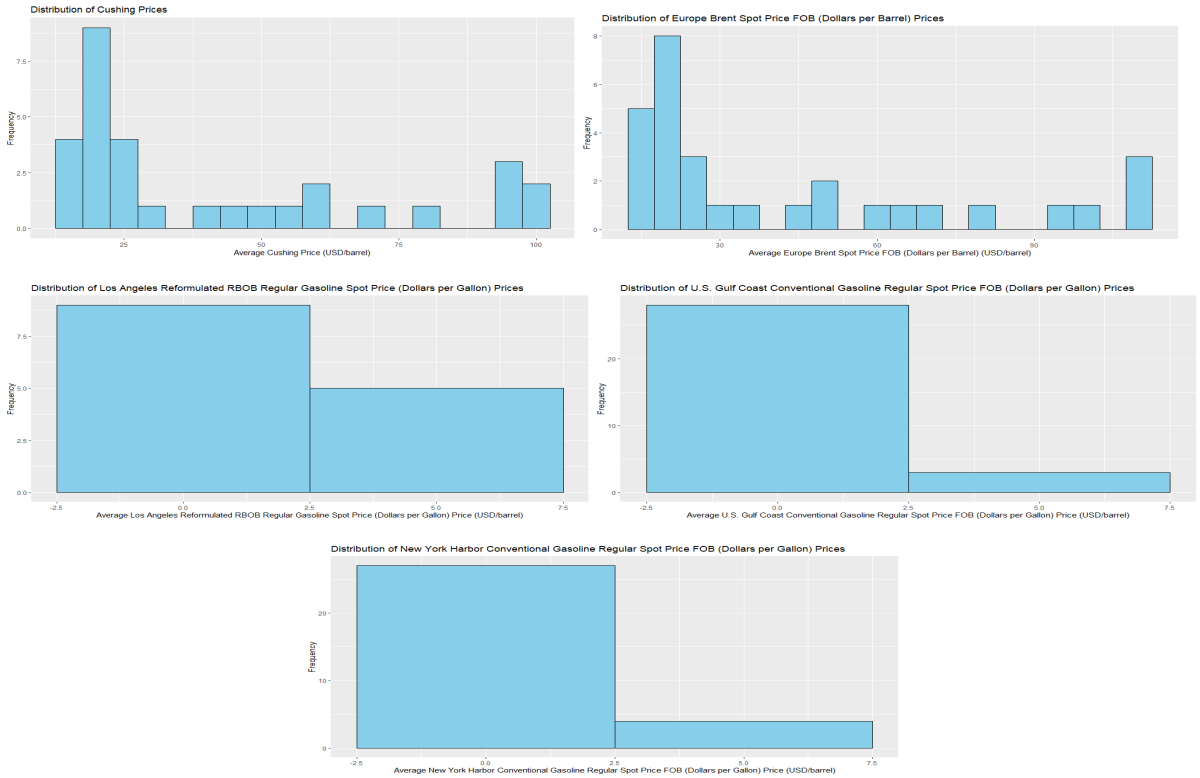




## Distribution Analysis:

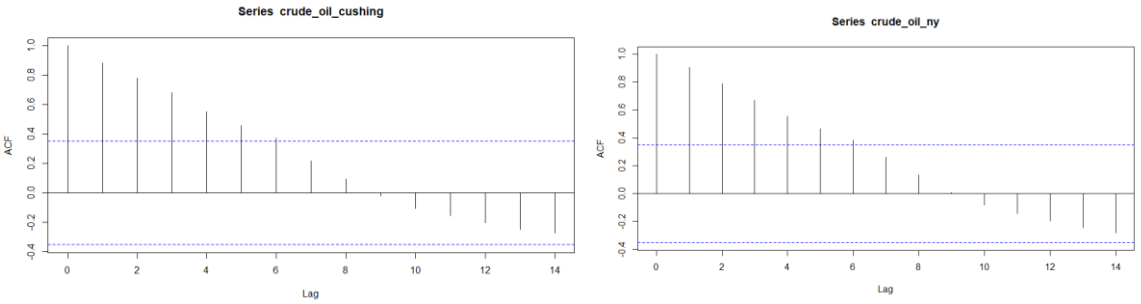
We also analyze the distribution of crude oil and gasoline prices using histograms. We create histograms for each type of crude oil and gasoline, showing the frequency of prices within certain price ranges. The histograms provide insights into the distribution of prices and can help identify any skewness or asymmetry in the data.

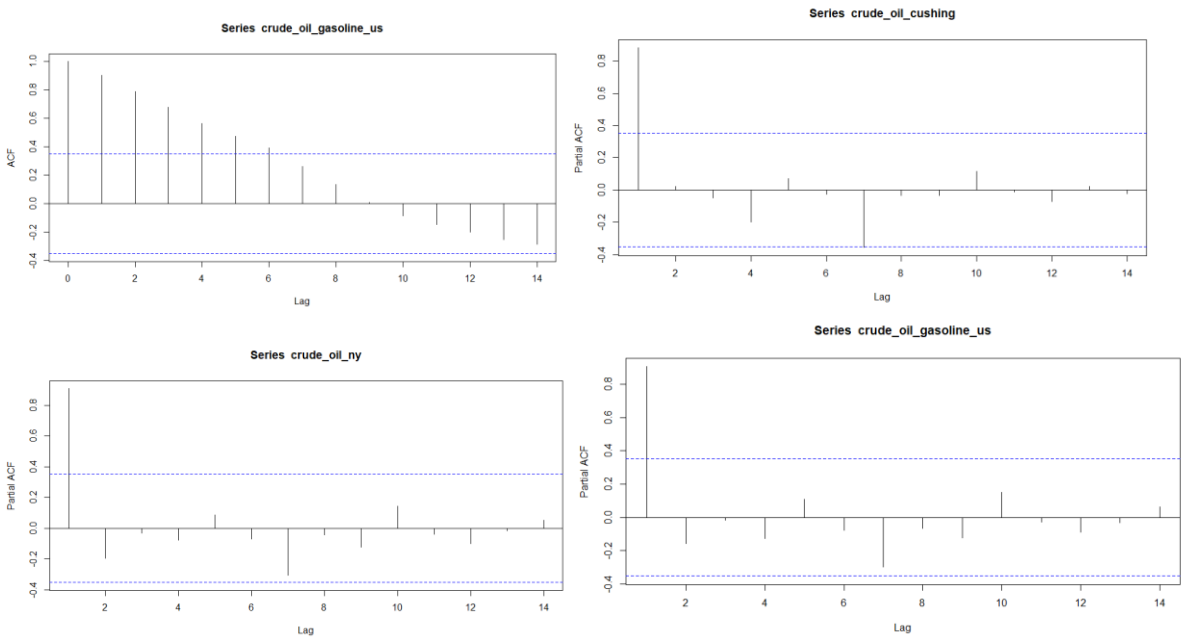



## ACF and PACF:

ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) are important tools used in time series analysis to identify and interpret the correlation patterns within time series data. To check the correlation in our dataset we also used these functions to create the plots in R to determine the appropriate ARIMA models.

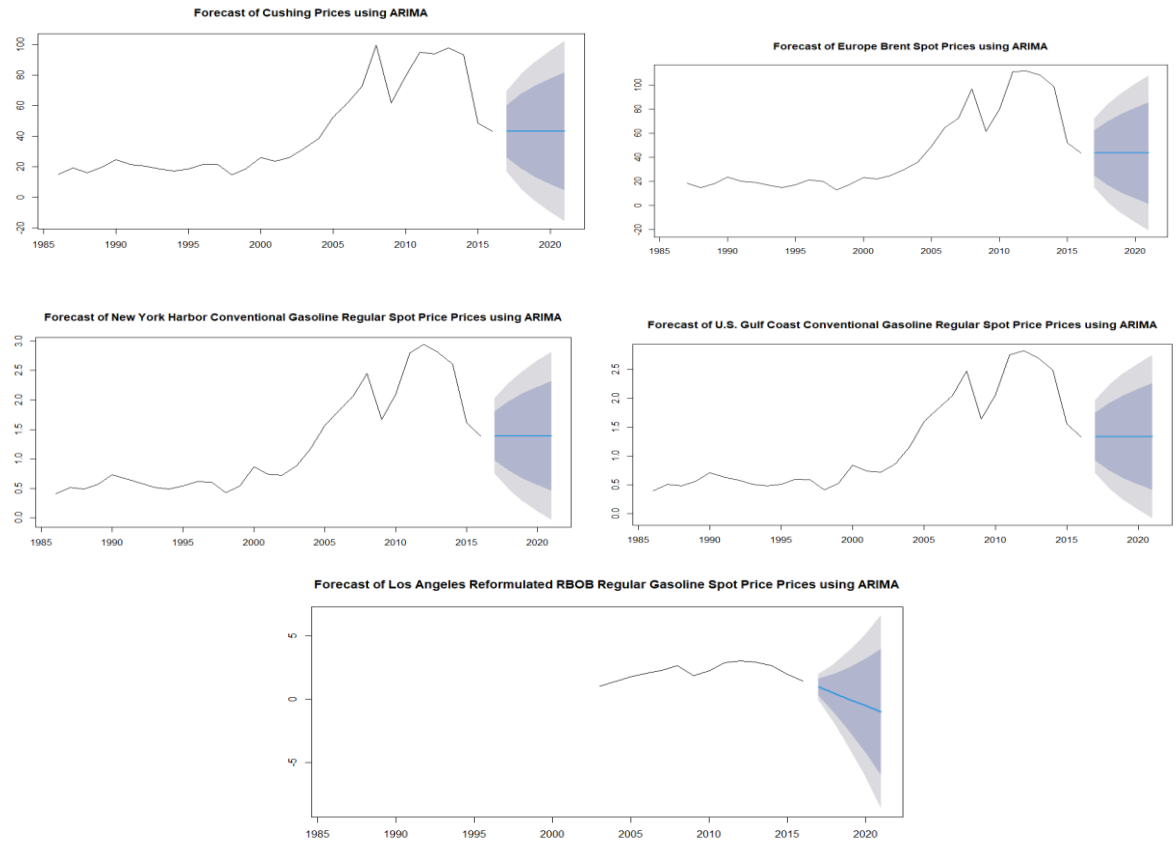
ACF and PACF are statistical tools used in time series analysis to measure the correlation between a time series data point and its lagged values. ACF identifies auto-correlation by measuring the similarity between current and lagged values, while PACF accounts for the influence of intermediate lags to determine the direct influence of each lag on the current value. ACF and PACF plots display correlation coefficients against lags on the x-axis and can indicate positive or negative correlation. Significant coefficients outside the confidence interval suggest the presence of autocorrelation or partial autocorrelation, respectively, providing insights into the underlying structure of the time series data for accurate forecasting.






## ARIMA Modeling:

We apply the ARIMA model to forecast crude oil and gasoline prices for the next year. The ARIMA model is a popular time series forecasting method that can capture patterns and trends in the data. We use the "auto.arima" function in the "forecast" library to automatically select the best ARIMA model based on the data. We then fit the selected ARIMA model to the data and generate forecasts for the next year.



**Conclusion:**

In conclusion, this analysis provides insights into the trends, decade-wise changes, and distribution of crude oil and gasoline prices for different types. The visualizations created using R help us understand the historical movements and patterns in crude oil prices, while the ARIMA modelling allows us to forecast future prices. These findings can be useful for decision-making in industries and policy-making at a macroeconomic level. Further analysis, including advanced time series modelling and incorporating external factors, can be conducted to refine the forecasts and gain deeper insights into crude oil and gasoline price dynamics and their implications.

**Future Works:**

In the upcoming part of the project, a crucial step involves assessing the accuracy of the predicted values generated by the implemented forecasting model. To achieve this, several performance metrics will be utilized to provide a comprehensive evaluation of the model's performance. These metrics include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), Forecast Accuracy, and R-squared (R2).

Furthermore, we will make efforts to improve the accuracy of the model. This may involve adjusting model parameters, exploring different algorithms, or incorporating additional data features. The goal is to continuously refine and enhance the forecasting model to ensure accurate and reliable predictions, thereby providing valuable insights for decision-making purposes.