# 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**

**Description of the dataset:**

The dataset consists of a time-series of crude oil and gasoline prices from 1986 to 2016. There are several variables in the main dataset, including the year, price of crude oil(Cushing, OK WTI Spot Price FOB (Dollars per Barrel), Europe Brent Spot Price FOB (Dollars per Barrel)), and price of Gasoline(New York Harbor Conventional Gasoline Regular Spot Price FOB (Dollars per Gallon), U.S. Gulf Coast Conventional Gasoline Regular Spot Price FOB (Dollars per Gallon)). Other datasets like Crude Oil Production, Baltic Dirty Tanker Index, S&P 500 Historical Data will also be used.

**Defining the Challenges, We Aim to Tackle:**

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. We also want to explore the impact of external factors such as political events, economic indicators, and industry-specific events on crude oil and gasoline prices. Finally, we want to evaluate the accuracy of our forecasts and identify any potential sources of error.

Some of the questions we plan to address:

1. Are there any seasonal patterns or trends in the data?
2. Can we accurately predict future prices of crude oil and gasoline based on historical data? What is the level of uncertainty in our predictions?
3. How do political events, such as wars or sanctions, affect crude oil and gasoline prices? Can we quantify the impact of these events on the price series?
4. Can we identify any outliers or anomalies in the data, and if so, what caused them?

**Planned Methods for Potential Implementation:**

To address these problems, we plan to use various time-series analysis techniques, such as ARIMA models, exponential smoothing, or regression analysis. We may also explore machine learning algorithms, such as neural networks or decision trees, to see if they can improve the accuracy of our forecasts. We will also need to perform exploratory data analysis to identify any trends, patterns, or outliers in the data. We may also use statistical tests to evaluate the significance of any relationships or correlations between variables.

**Plan for workload split among team members:**

Arun Srivatsan Swaminathan: Model Fitting, Specification and Forecasting, documentation
Kannan Walter: Data cleaning and preprocessing, plotting the seasonal factors monthly
Mason Zepnick: Initial data visualization, ACF and PACF charts, ordinary differencing
Prageeshwar Chandran: Decomposing the time series,  CCF tests, presentation preparation.

We will also schedule regular team meetings to discuss progress and address any issues that arise. Each team member will be responsible for documenting their work and sharing their results with the rest of the team.

**Comments and/ or concerns:**

One potential concern is the accuracy of the data. Crude oil and gasoline prices can be affected by many factors, and it may be difficult to account for all of them in our analysis. We will need to carefully consider which variables to include in our models and how to handle missing values or outliers. We also need to be mindful of the assumptions underlying the various time-series analysis techniques we use and ensure that our models are appropriate for the data we have. Finally, we need to be prepared to adjust our methods or assumptions as we learn more about the data and the problem we are trying to solve.