

# Report

## Summary

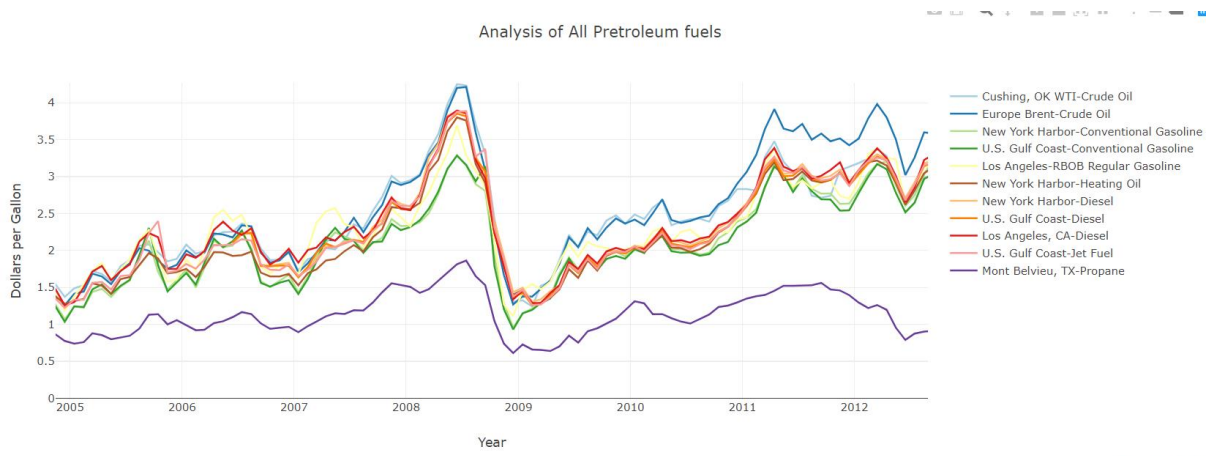
My project included the following sequence (Time taken for each is mentioned on the side):

- 1) Importing Datasets (10 min)
- 2) Data Cleaning (20 min)
- 3) Creating column to find price changes month over month for each petroleum products (20 min)
- 4) Creating a summarized price and price movement by year for each petroleum products (15 min)
- 5) Exploratory Analysis
  - a. Finding Insights (60 min)
  - b. Identifying the correlation between Conventional Gasoline prices between New York Harbor and U.S. Gulf Coast (10 min)
  - c. Finding time period in which prices of all Products were lower than prior data points (25 min)
- 6) Predicting the crude oil prices for both cities for next 6 months using SARIMA model (60 min)

**Note:** In depth Analysis of each step can be seen in the html file provided. Below, we can find the findings and snapshots of the data manipulation and prediction

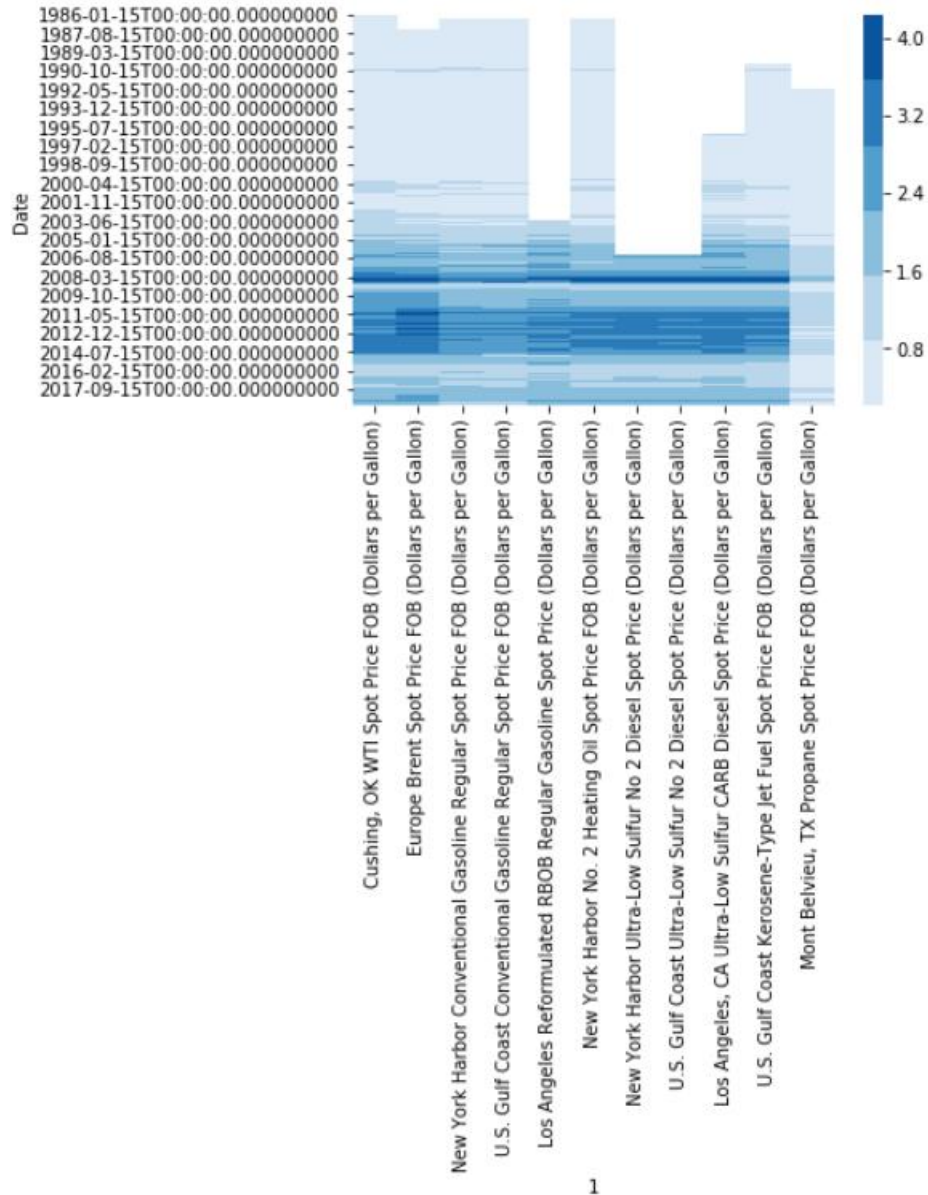
## Hidden Insights and their possible external factors

- 1) The highest prices for almost all the petroleum products were in August 2008 and the prices for all decreased a lot in the same year towards the end. This can be seen in the figure below from the analysis:



External Factor for the steep decrease in the price was **recession** which caused demand for energy to shrink

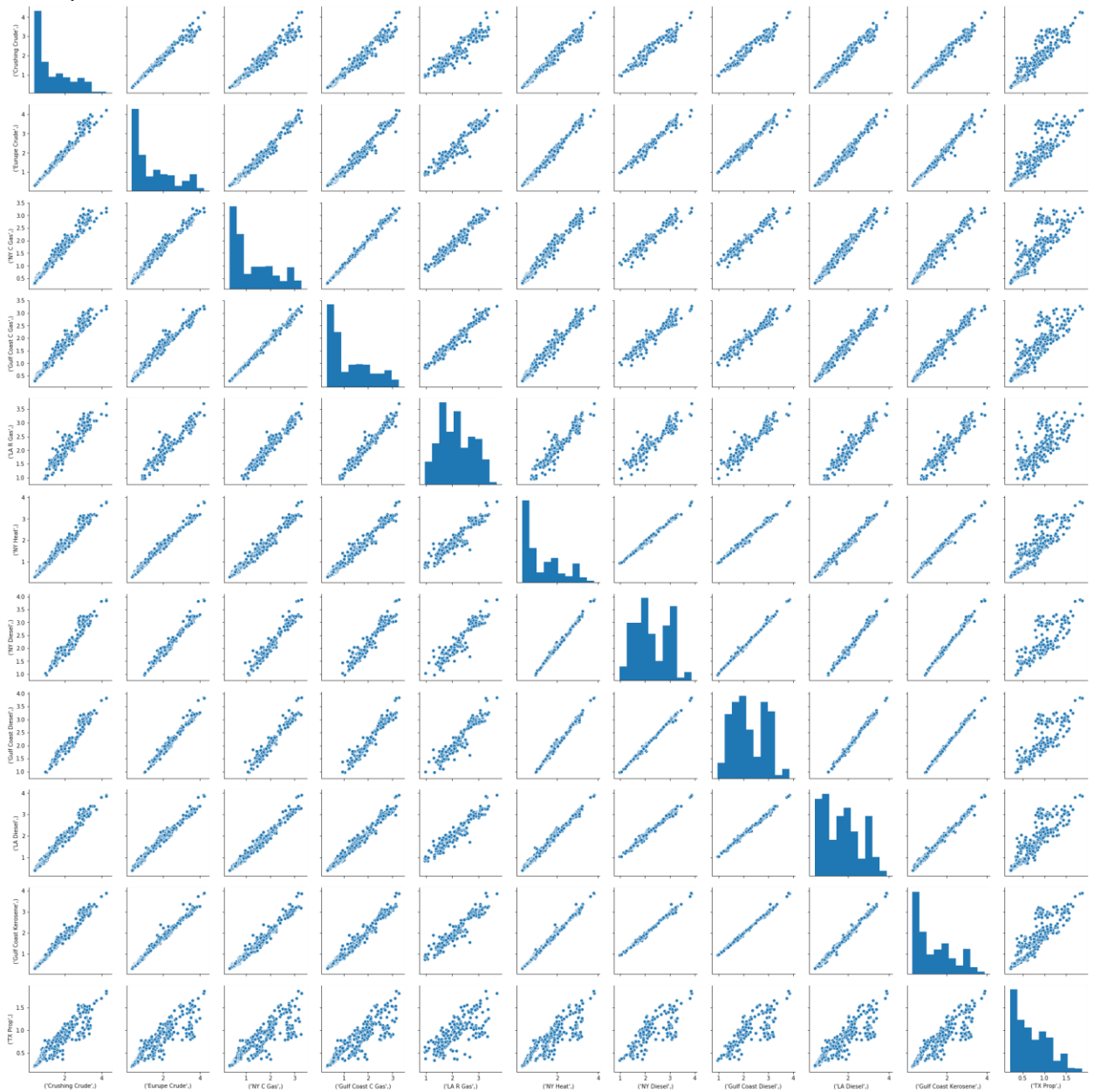
- 2) From almost mid of 2004, the prices for all petroleum products started to increase and move to higher ends till 2008. This can be clearly seen from the heatmap below.



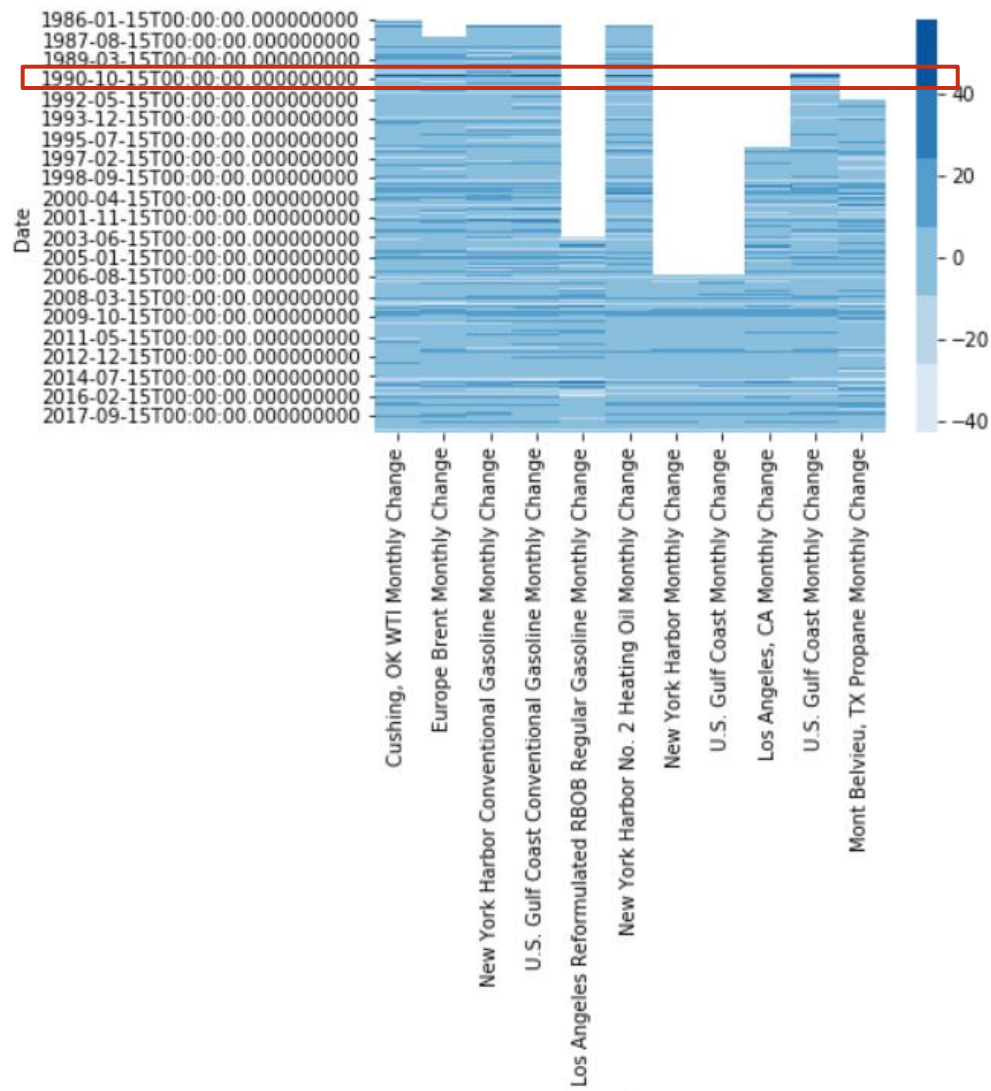
1

*External Factor: Hurricane Katrina caused severe damage to U.S. refinery and production capacity in the Gulf of Mexico which caused the oil prices to spike.*

- 3) The below correlation graph showed that prices of all petroleum products were positively correlated with each other. Only, propane showed slight positive correlation with the other products.

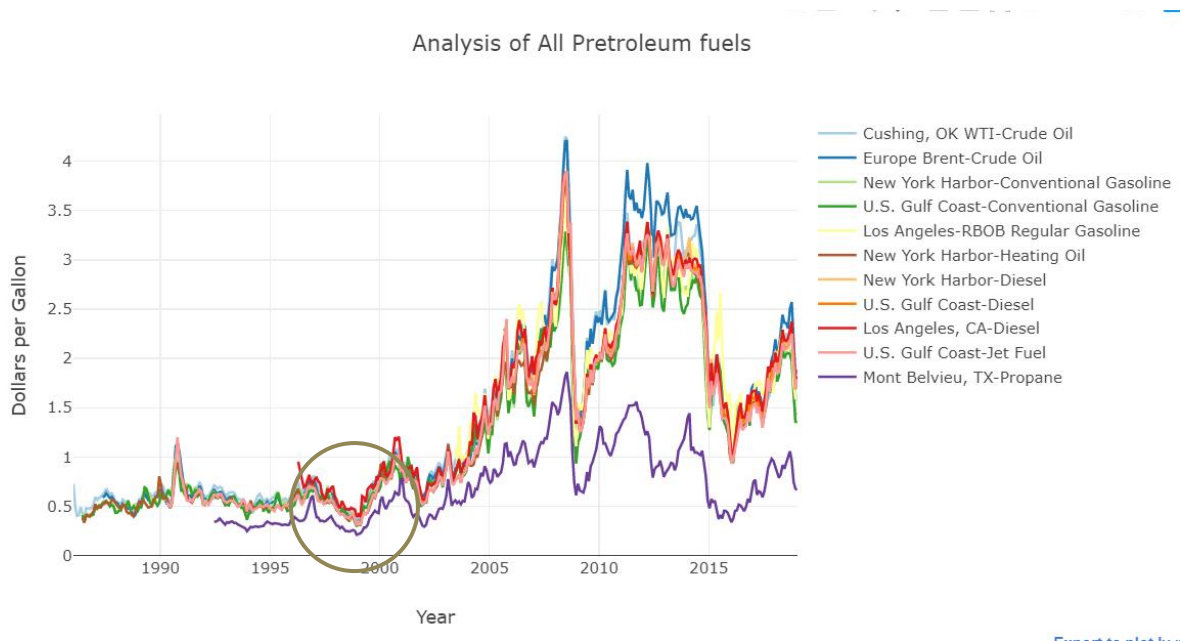


- 4) There was almost 50+ percent increase in oil prices from July 1990 to August 1990.



*External Factor: United Nations forbade the two of the world's biggest oil producers, Iraq and Kuwait from importing oil from which caused the oil price increase*

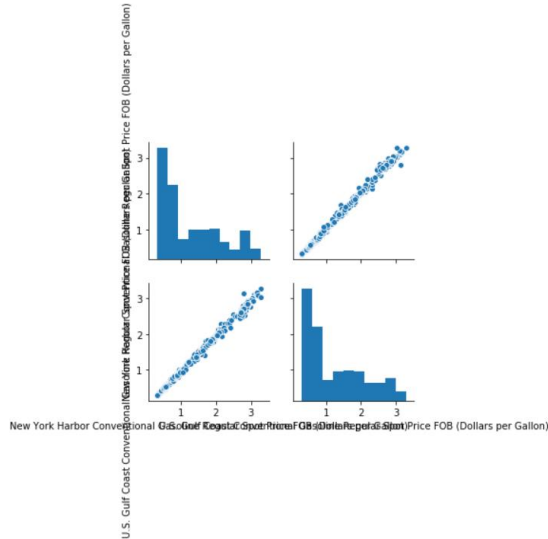
- 5) The least prices for most of the petroleum products were seen in the year 1990 in the month of December



Possible External Factor: Rapid technological advances pushed the cost of finding, developing and producing crude oil outside the Middle East. Maybe that is why a minimum prize was seen.

### **Correlation between Conventional Gasoline prices between New York Harbor and U.S. Gulf Coast**

A positive correlation was found between the prices of Conventional Gasoline of the two cities. The karl pearson's coefficient of correlation was found to be 0.998 which also says that both the prices are strongly and positive correlated to each other.



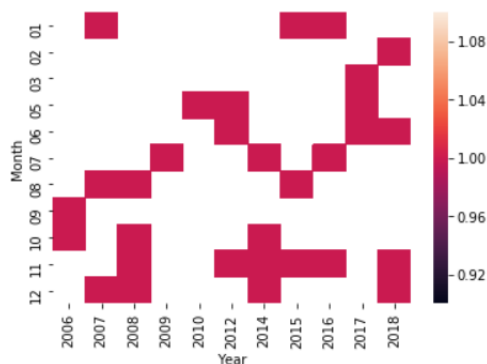
```
np.corrcoef(df_cGasoline.iloc[:,0],df_cGasoline.iloc[:,1])
array([[1.          , 0.99802942],
       [0.99802942, 1.          ]])
```

We can see strong positive correlation between both the cities' prices for conventional Gasoline

## Months where prices of all Products were lower than prior data points

From the below heatmap created, we can see the months where the prices of the previous months of all the products were more as compared to the current month

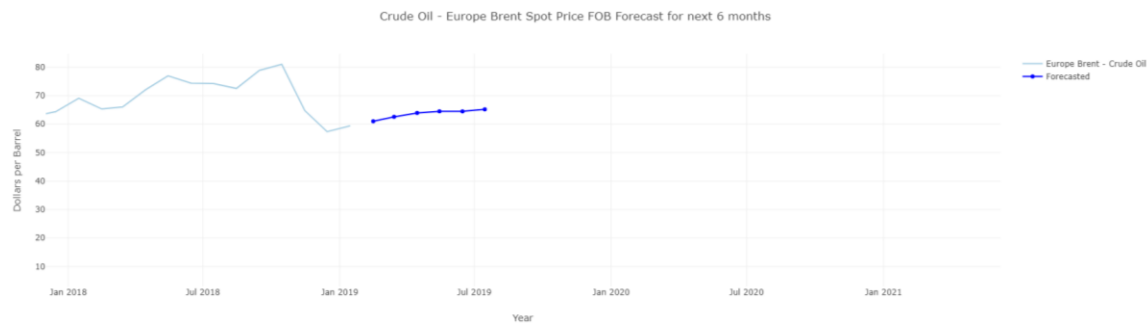
Plot to represent the times when all Products prices were lower than their prior data



From this plot we can conclude that in the year 2008 and 2014, during the October, November and December; we saw a constant downfall in the prices as compared to its previous months. The other can be judged from the plot above.

*We also see that the only month which did not see a downfall as compared to its previous months in all the years was the month of April*

## Crude Oil Price Predictions made by S-ARIMA for the Next 6 months



## Snapshot of the addition column to find price changes month over month

Date	New York Harbor Ultra-Low Sulfur No 2 Diesel Spot Price (Dollars per Gallon)	U.S. Gulf Coast Ultra-Low Sulfur No 2 Diesel Spot Price (Dollars per Gallon)	Los Angeles, CA Ultra-Low Sulfur CARB Diesel Spot Price (Dollars per Gallon)	New York Harbor Monthly Change	U.S. Gulf Coast Monthly Change	Los Angeles, CA Monthly Change
1996-04-15	NaN	NaN	0.950	NaN	NaN	NaN
1996-05-15	NaN	NaN	0.870	NaN	NaN	-8.421053
1996-06-15	NaN	NaN	0.745	NaN	NaN	-14.367816
1996-07-15	NaN	NaN	0.692	NaN	NaN	-7.114094
1996-08-15	NaN	NaN	0.705	NaN	NaN	1.878613

## Snapshot of summarized price and price movement by year

Out[18]:

1	Cushing, OK WTI Spot Price FOB (Dollars per Barrel)	Europe Brent Spot Price FOB (Dollars per Barrel)	Cushing, OK WTI Monthly Change	Europe Brent Monthly Change
Date				
1986-12-31	15.036667	NaN	-1.711560	NaN
1987-12-31	19.171667	18.522500	0.780099	-1.146253
1988-12-31	15.982500	14.948333	-0.239611	-0.582507
1989-12-31	19.640833	18.251667	2.261132	2.389460
1990-12-31	24.467500	23.675833	3.418130	4.598114
1991-12-31	21.502500	20.010833	-2.465033	-3.154075
1992-12-31	20.563333	19.305833	0.044119	-0.033064

## Future Scope of the Project

- 1) The crude oil time series data was not stationary. It could have been made stationary
- 2) Different machine learning model as Vector Auto Regression could be applied to time series data for forecasting and comparing the current seasonal ARIMA model to it
- 3) Insights from each petroleum products could be found out and explained in depth
- 4) For the yearly summary table, the percent change of the price of oil in Jan to price of oil in Dec could have been consider and visualized to see the insights.
- 5) Percent change of Years or Quarters could have been check by taking percent change as mentioned in the above point.