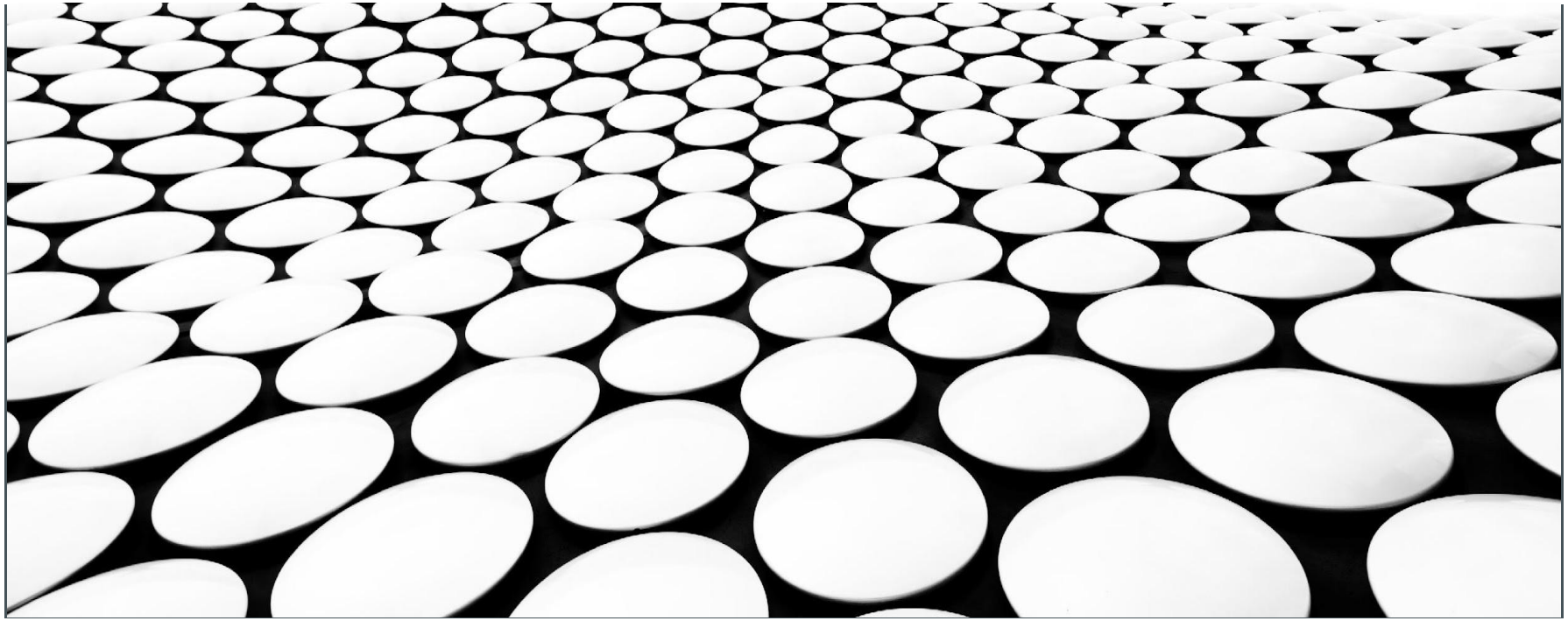
CRUDE OIL PRICE PREDICTION

MENTORS: KARTHIK MUSKULA &
DHANYAPRIYA SOMASUNDARAM



n Members

na Vishnu
u Nandhini
dindla Divya
Naveen Kumar
Pradeep Kumar
t Patil

Business Problem:

EXCELR
Raising Excellence

With a global trading price of 88$ per barrel, oil prices have hit an all-time high in the last seven years. Due to rise in oil price the economy are getting affected vary badly.

Objective:

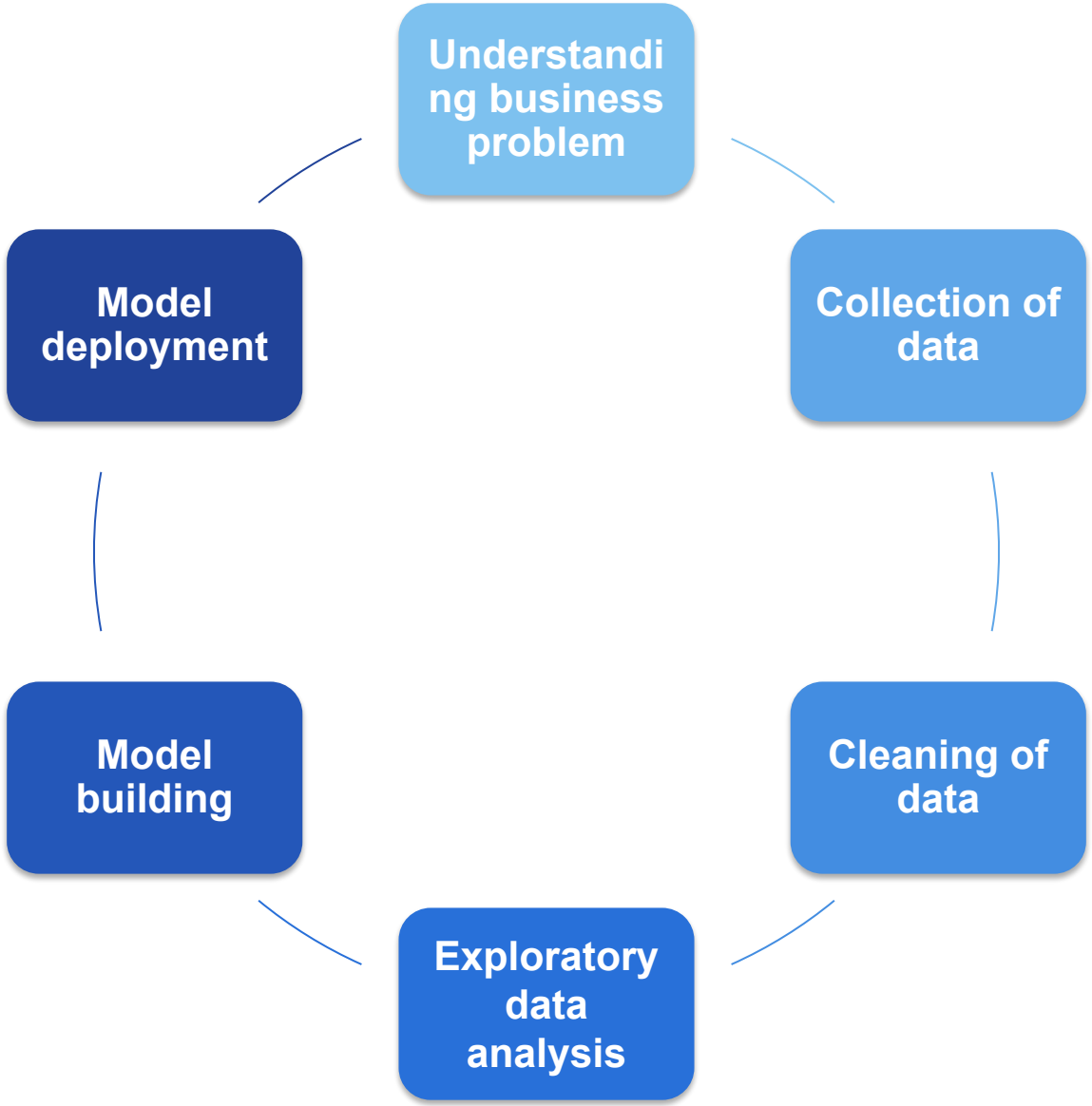
Oil is a product that goes completely in a different direction for a single market event as the oil prices are rarely based on real-time data, instead, it is driven by externalities making our attempt to forecast it even more challenging

As the economy will be highly affected by oil prices our model will help to understand the pattern in prices to help the customers and businesses to make smart decisions.

Factors That Affect the Price of Oil

affected by most of the factors , but the one’s that affect the most ar
and supply

Project Architecture / Project Flow



taset

wnloaded the crude oil price dataset from NASDAQ website.

aset has total 9345 rows and 2 columns which includes the crude oil price d
n 02-01-1986 to 21-12-2022.

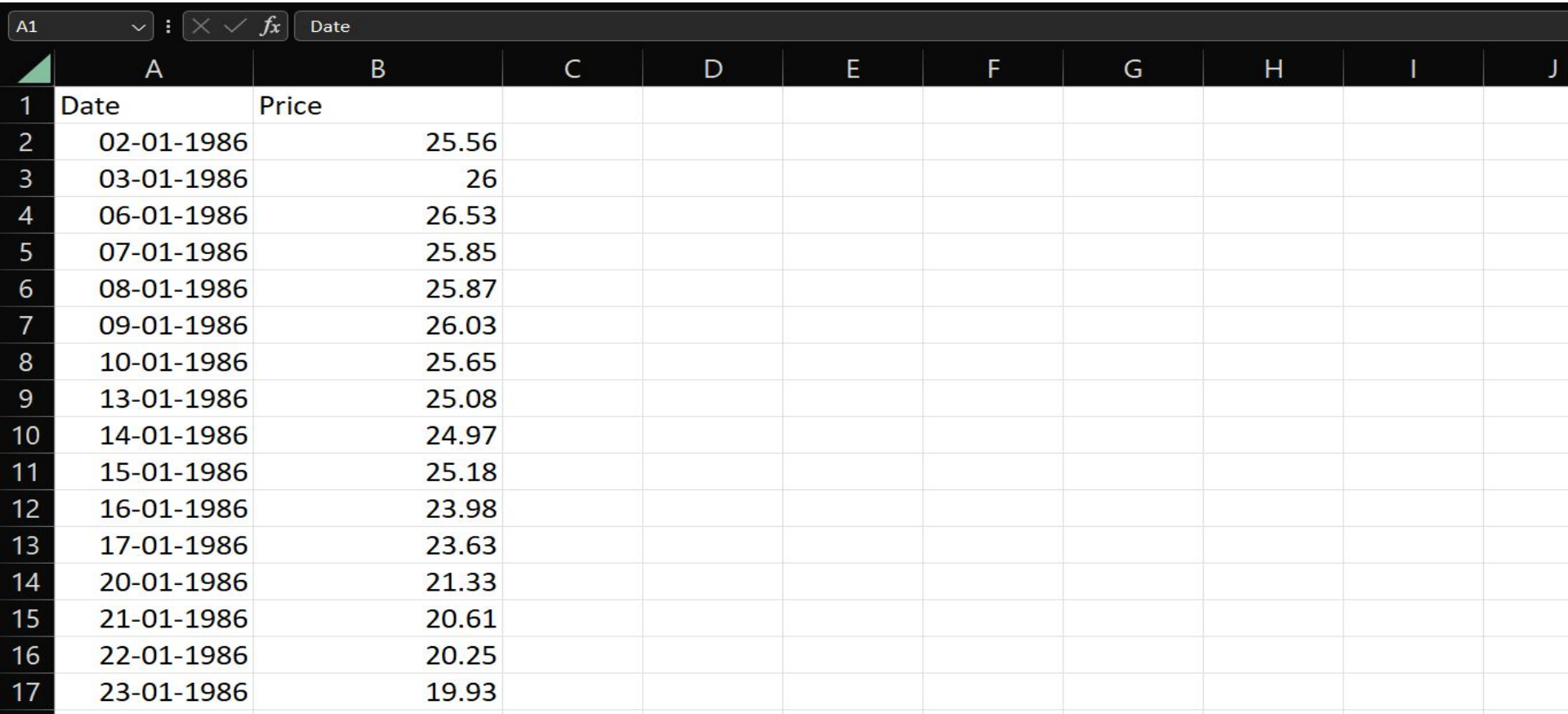
Exploratory Data Analysis (EDA)

Dataset Details

Dataset contains 9345 rows and 2 columns

Datatypes are DataTime and Float

Data ranges from 02-01-1986 to 21-12-2022.

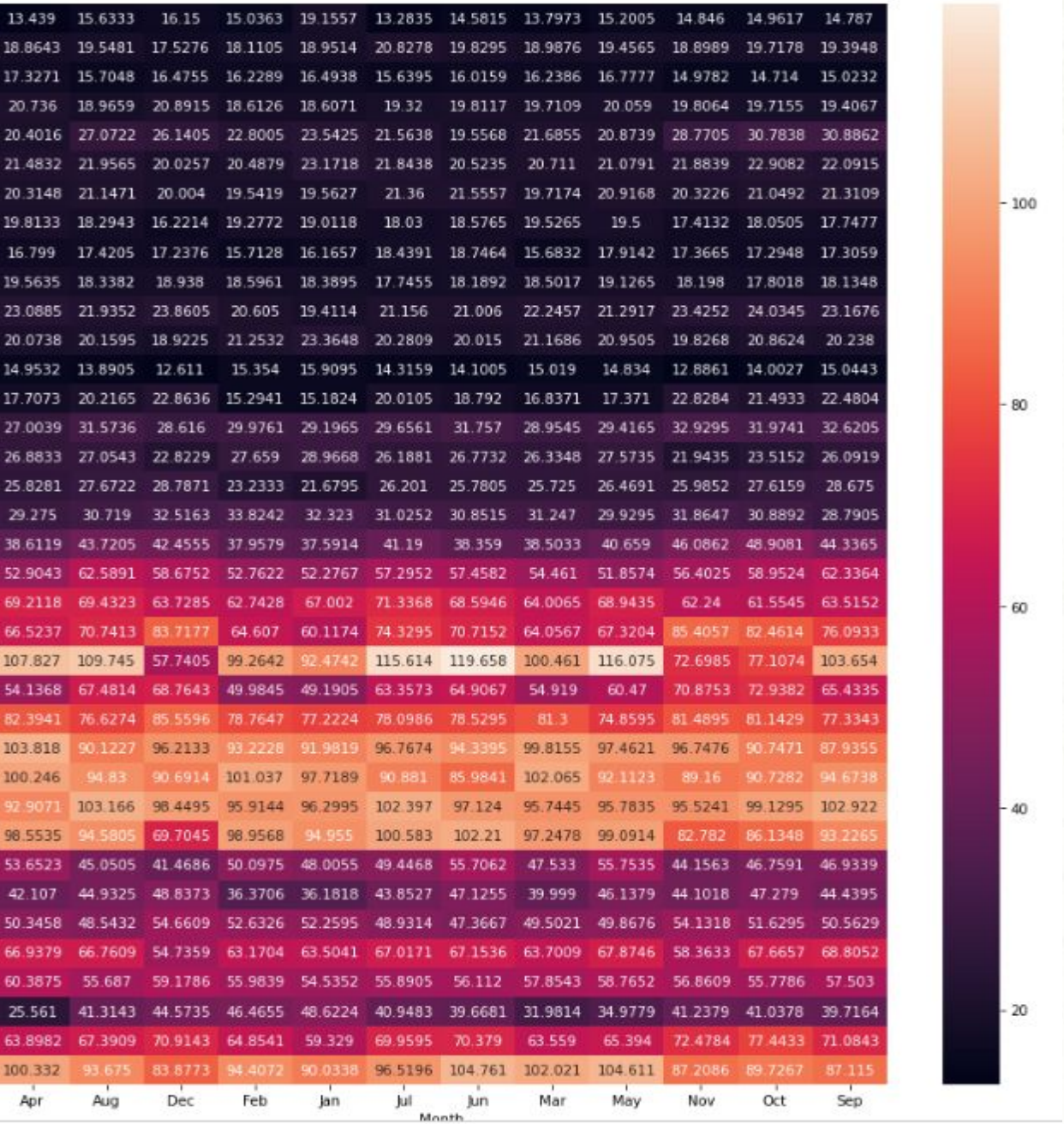


A1 | Date

	A	B	C	D	E	F	G	H	I	J
1	Date	Price								
2	02-01-1986	25.56								
3	03-01-1986	26								
4	06-01-1986	26.53								
5	07-01-1986	25.85								
6	08-01-1986	25.87								
7	09-01-1986	26.03								
8	10-01-1986	25.65								
9	13-01-1986	25.08								
10	14-01-1986	24.97								
11	15-01-1986	25.18								
12	16-01-1986	23.98								
13	17-01-1986	23.63								
14	20-01-1986	21.33								
15	21-01-1986	20.61								
16	22-01-1986	20.25								
17	23-01-1986	19.93								

Exploratory Data Analysis (EDA)

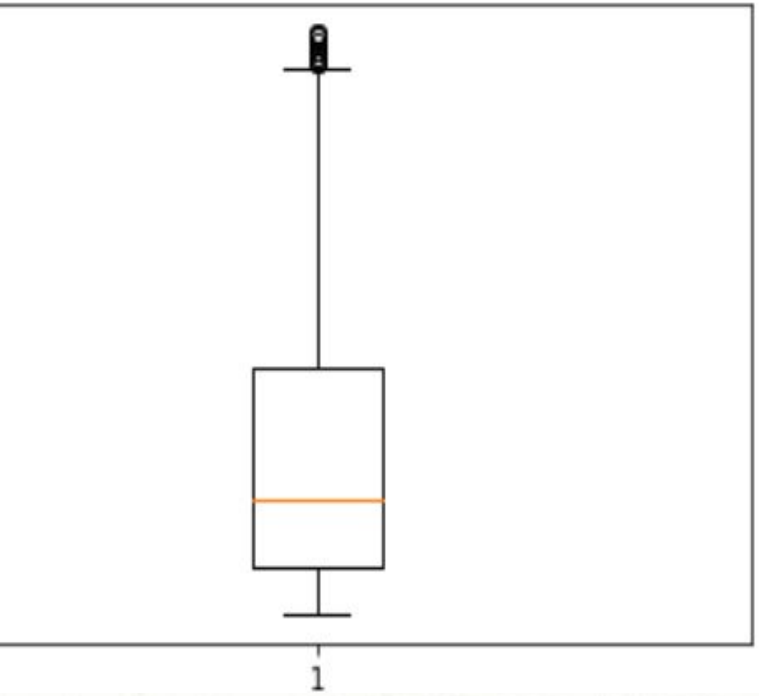
- Dataset has total 9345 rows and 2 columns which includes the crude oil price data from 02-01-1986 to 21-12-2022.
- Calculated Mean, Median, Variance, Skewness, Kurtosis. Mean = 46.028511, Median = 36.060000, Variance = 871.16243,Kurtosis = -0.553953 (Flat peak or Platykurtic), Skewness = 0.761338 (Positive Skewness or Right skewed).
- We found that the dataset has 6 null values. We applied mean imputation technique to remove the null values from the data.
- Calculated the correlation between Price and Year. Correlation = 0.723501 (Positive correlation)
- We used outlier detection technique (Histogram, Boxplot, Describe function) and found that the data has some outliers.
- We used boxplot formula to calculate the upper extreme and lower extreme values and removed the outliers using drop function.
- Finally, null values and outliers have been removed from the data.

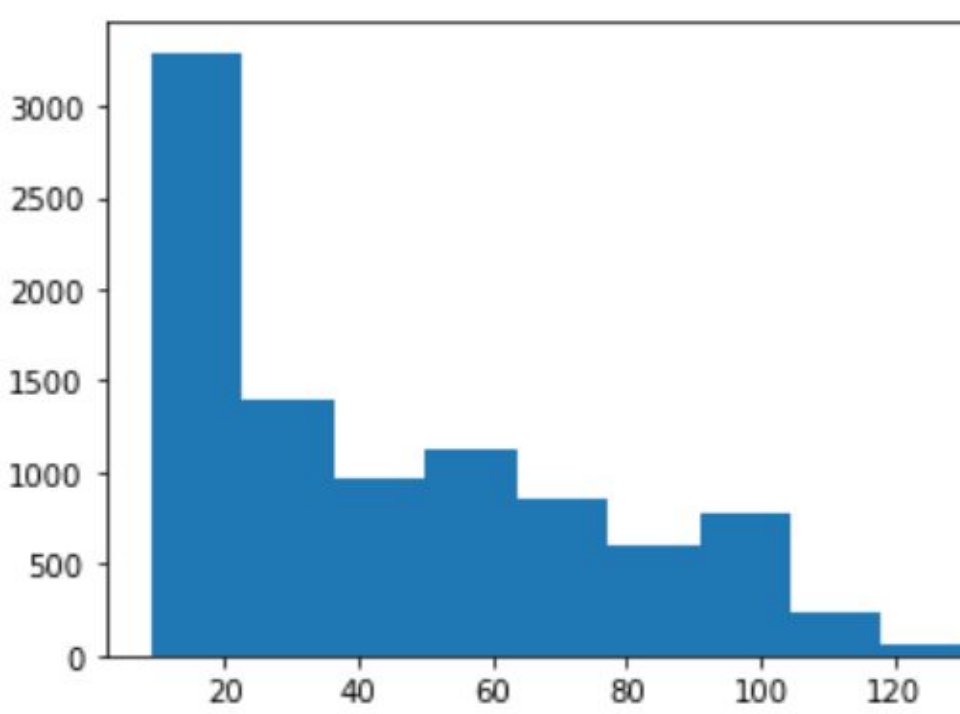


Inferences:

Since Heatmap provide us easy tool to understand the correlation between two entities, we used the heatmap to represent the correlation between the two variables year & month.

tlier Detection

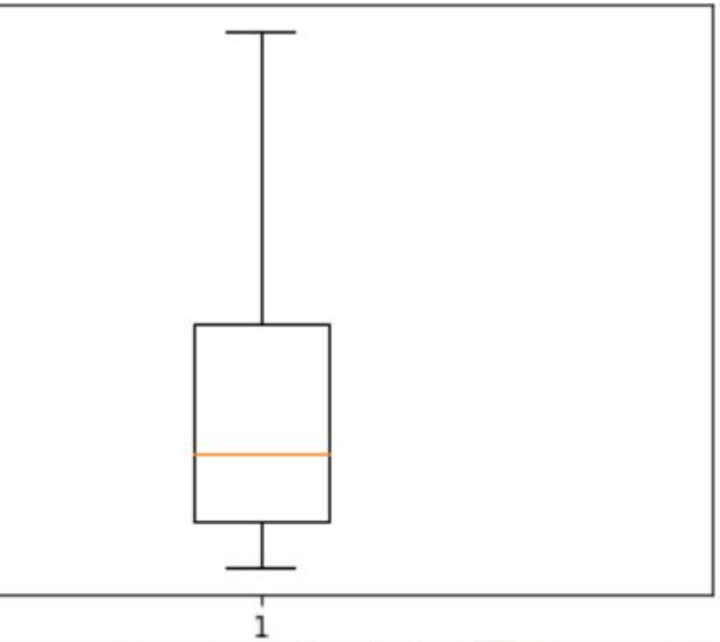


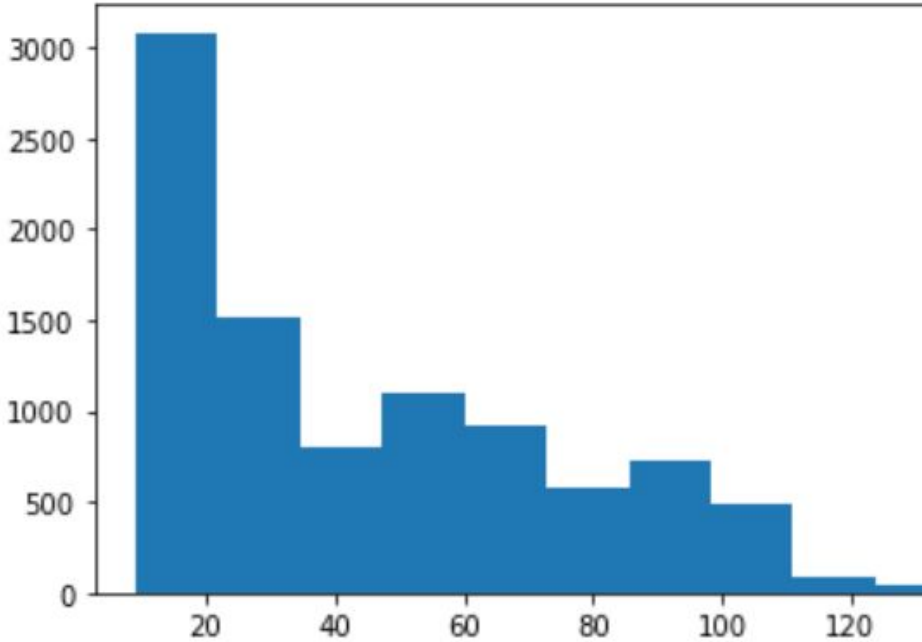


nces:

ere are outliers above the positive upper extreme whisker.
ere are no outliers below the negative side of the lower whisker.
e the data doesn't follow Normal Distribution. Right skewed data.

bing outliers



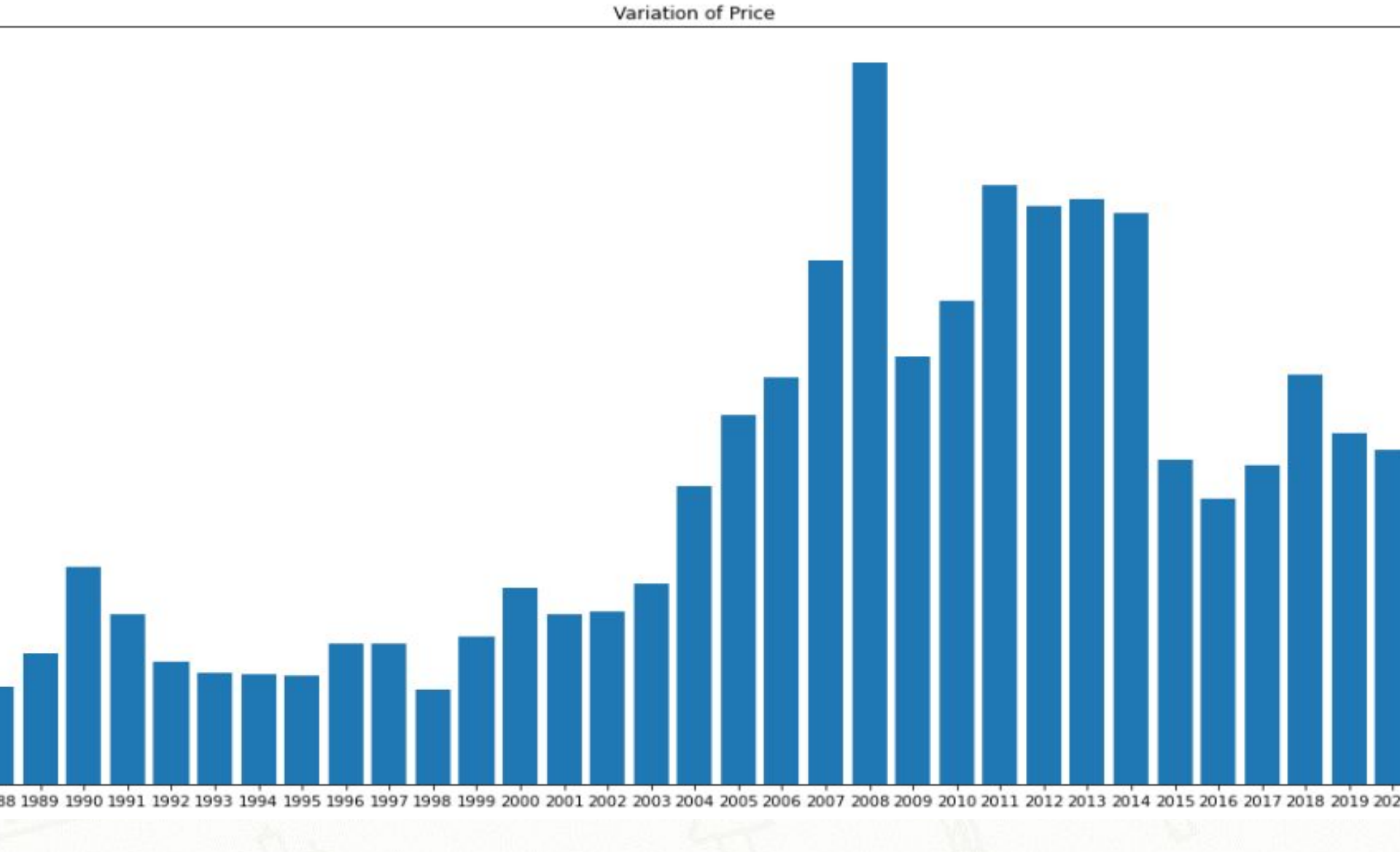


:

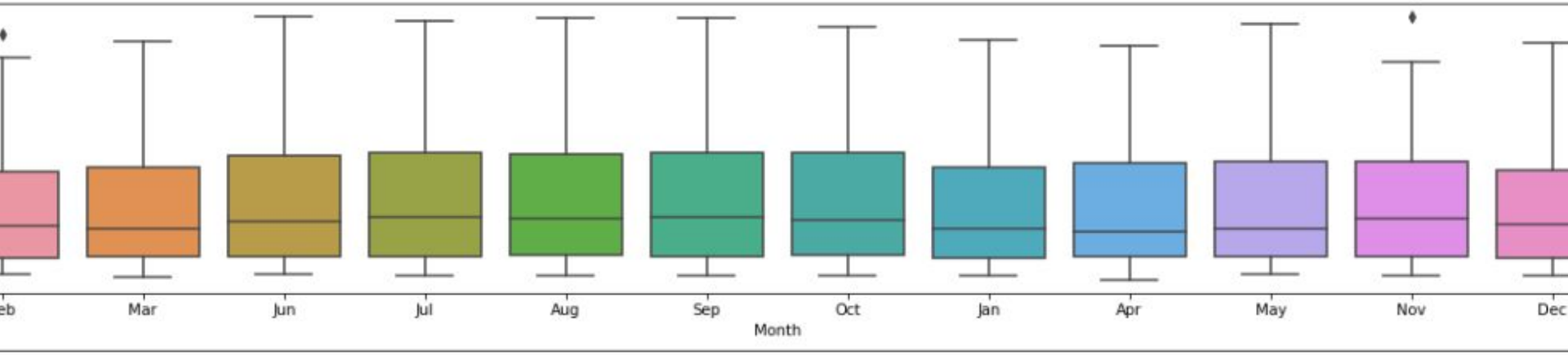
e no outliers above the positive upper
whisker.

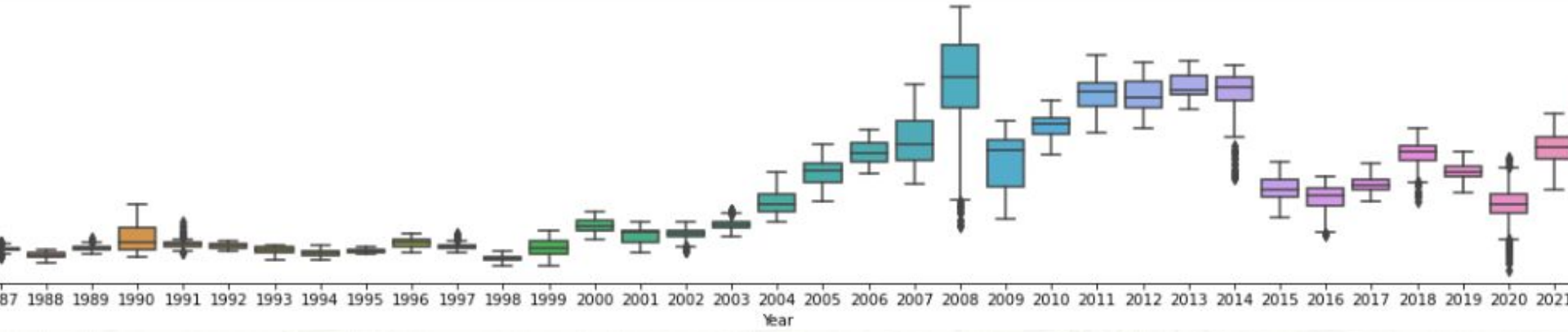
Data Visualization

Bar Plot of Variation of Price over Years



Box plot



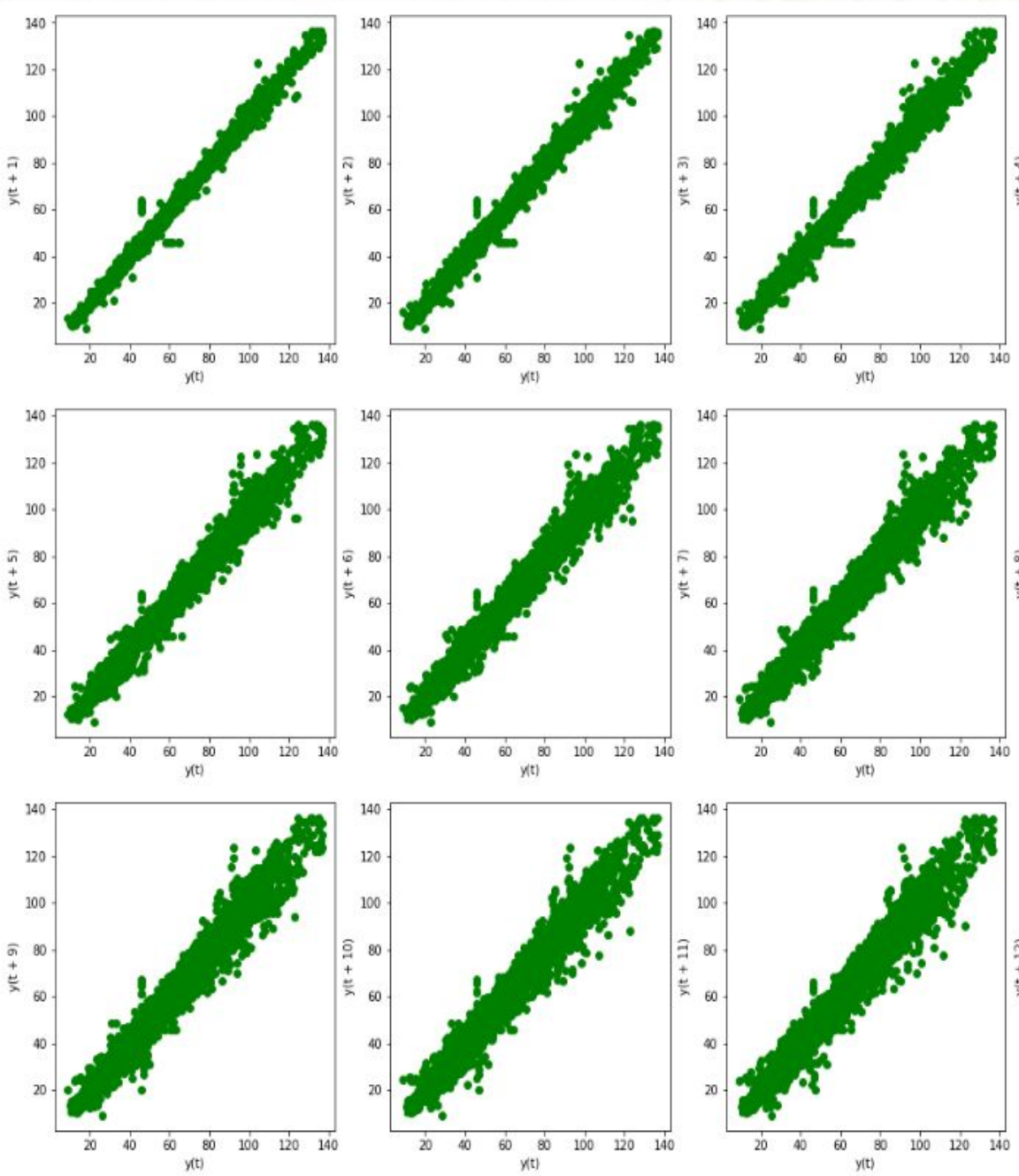


PLOT : Inferences :- Years 2018 and 2020 show outliers. 2018 The price of oil dropped in November 2018 becaus
umber of factors, including "rising petro-nations' oil production, the U.S. shale oil boom, and swelling in North
ican oil inventories," according to Market Watch. 2020 As rising Covid cases prompted fears of demand
own, due to that the price of the oil dropped.

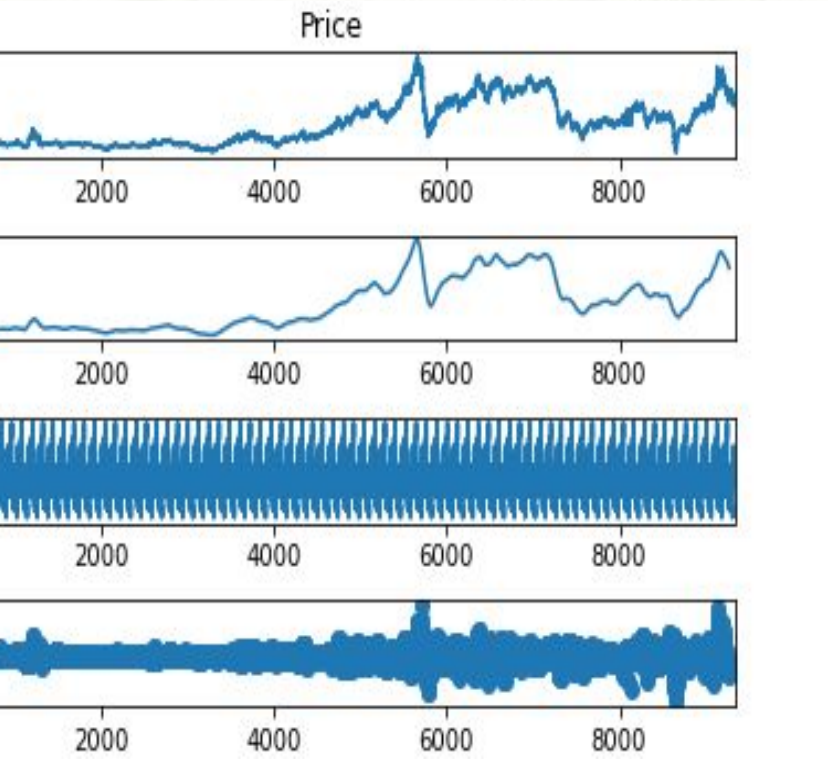
zation
arly)

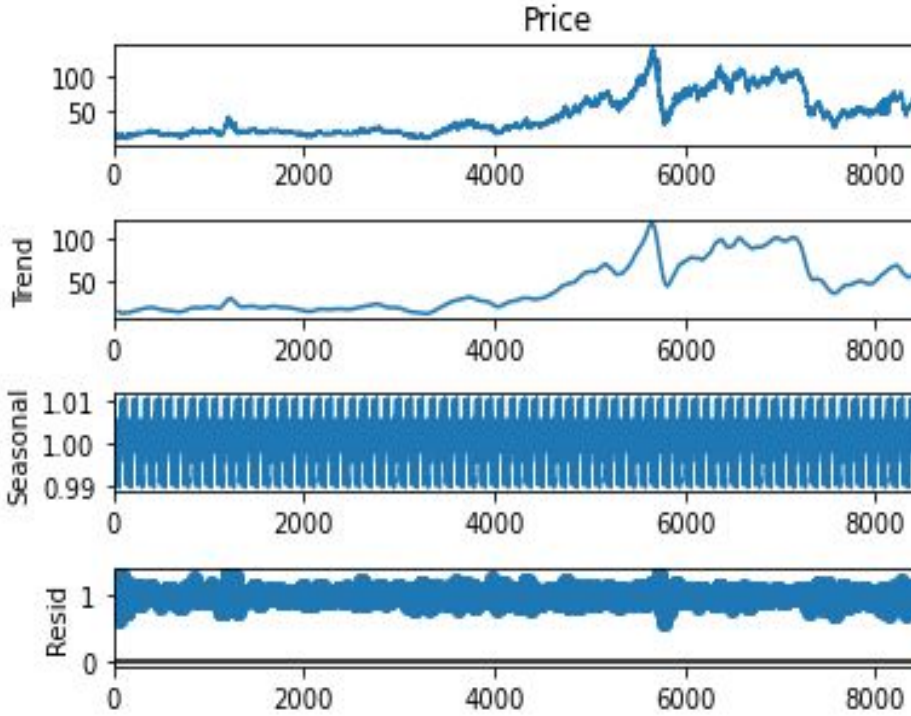
pecial type of scatter plot in
s represents the dataset with
behind or ahead as compared
e difference between these
ed lag or lagged and it is
k.

series is showing autoregressive
s good for time series modelling
gives a Positive correction
nce they can be modelled.

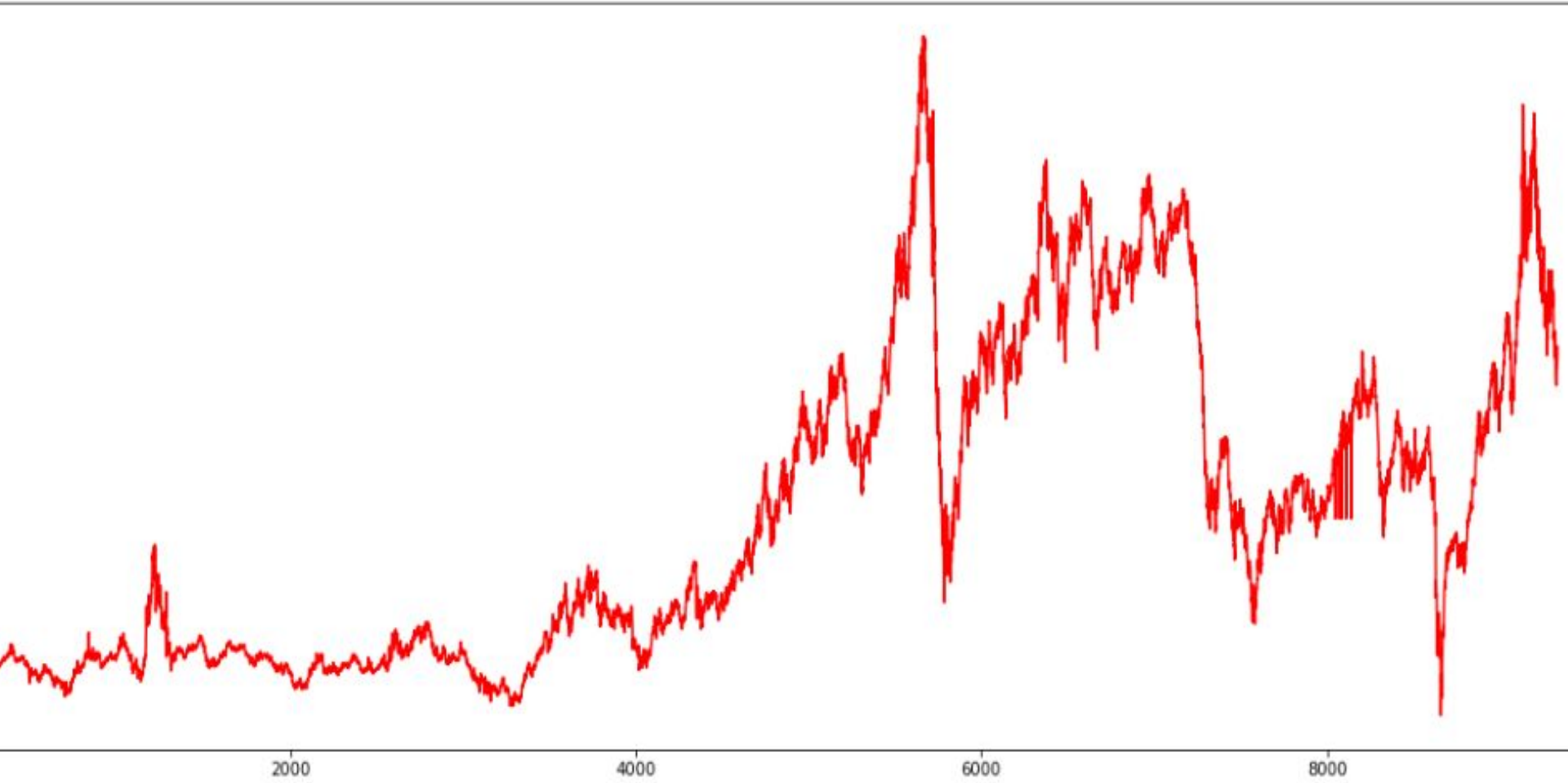


Time series Decomposition Method





es Smoothing



xponential trend in Moving average & it shows orginal Oil Price data.

Model Building

MODEL BASED METHODS

	MODEL	RMSE_Values	MAPE_Values
0	Linear_model	2.918314e+01	48.066785
1	Exp_model	4.420679e+01	92.698980
2	Quad_model	3.435598e+97	150.729517
3	Add_sea_model	4.425814e+01	48.859428
4	Add_sea_quad_model	9.655196e+01	150.898063
5	Mult_sea_model	4.904882e+01	94.686302
6	Mult_add_sea_model	4.423942e+01	92.699036

ARIMA Model

SARIMAX Results

Dep. Variable:	y	**No. Observations:**	9331
Model:	SARIMAX(0, 0, 2)	**Log Likelihood**	-16089.679
Date:	Mon, 09 Jan 2023	**AIC**	32185.357
Time:	15:32:38	**BIC**	32206.781
Sample:	0	**HQIC**	32192.634
	- 9331		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ma.L1	-0.1213	0.003	-40.770	0.000	-0.127	-0.115
ma.L2	-0.0155	0.006	-2.703	0.007	-0.027	-0.004
sigma2	1.8418	0.007	257.533	0.000	1.828	1.856

Ljung-Box (L1) (Q):	0.00	**Jarque-Bera (JB):**	356642.45
Prob(Q):	0.99	**Prob(JB):**	0.00
Heteroskedasticity (H):	11.12	**Skew:**	-0.67
Prob(H) (two-sided):	0.00	**Kurtosis:**	33.26

Splitting the data into

Train data = 66.6%
Test data = 33.4%

Model with order

Data Driven Methods

	MODEL	RMSE_Values	MAPE_Values
0	Ses_model	1.147909e+33	37.926884
1	Holt_model	inf	1181.995312
2	Hwe_add_add_model	4.066831e+44	56.706486
3	Hwe_mult_add_model	2.841681e+44	56.403096

FB Prophet Model

Train2 = data11.iloc[:len(data11)-150] # Total data observations

Test2 = data11.iloc[len(data11)-150:] # Only last 1

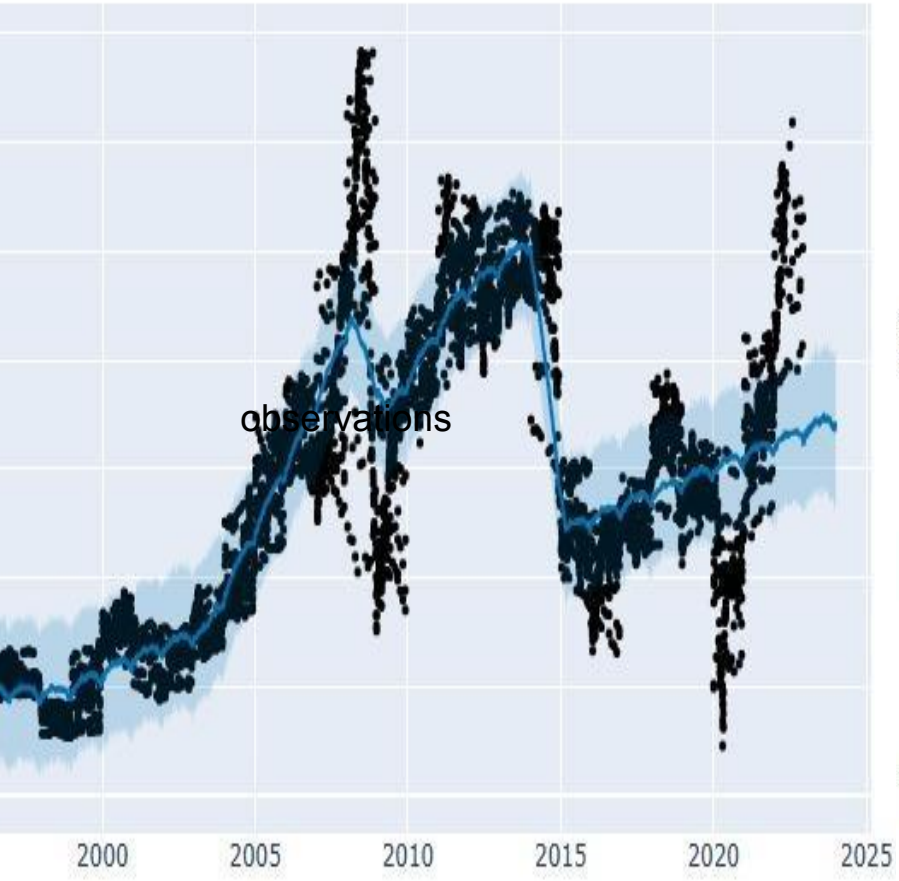
Result:

rmse: 9.83

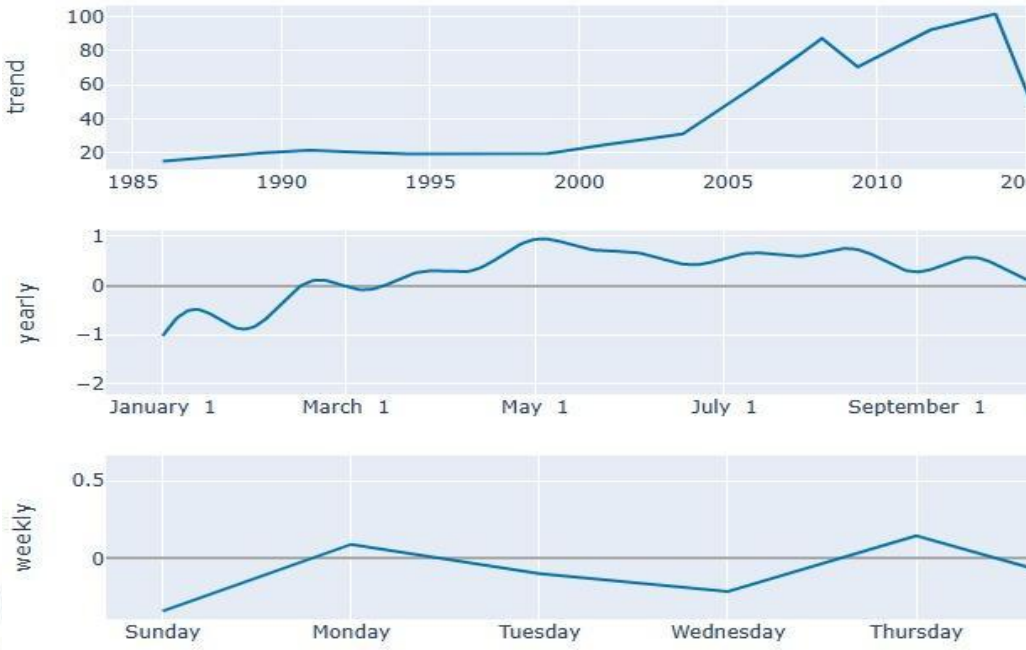
MAE: 6.16

MSE: 96.54

R-squared Score: 0.88

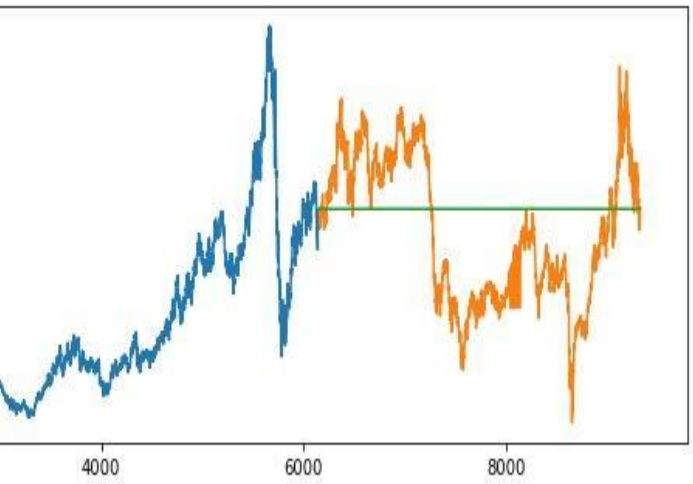


Components



el (Last Sample Method)

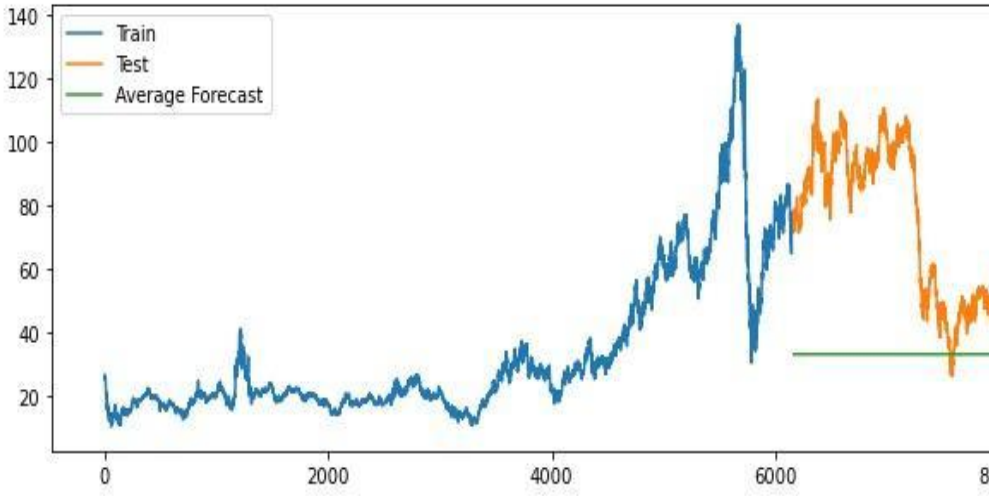
ead(6160)
l(3172)



.389729471063823e+33

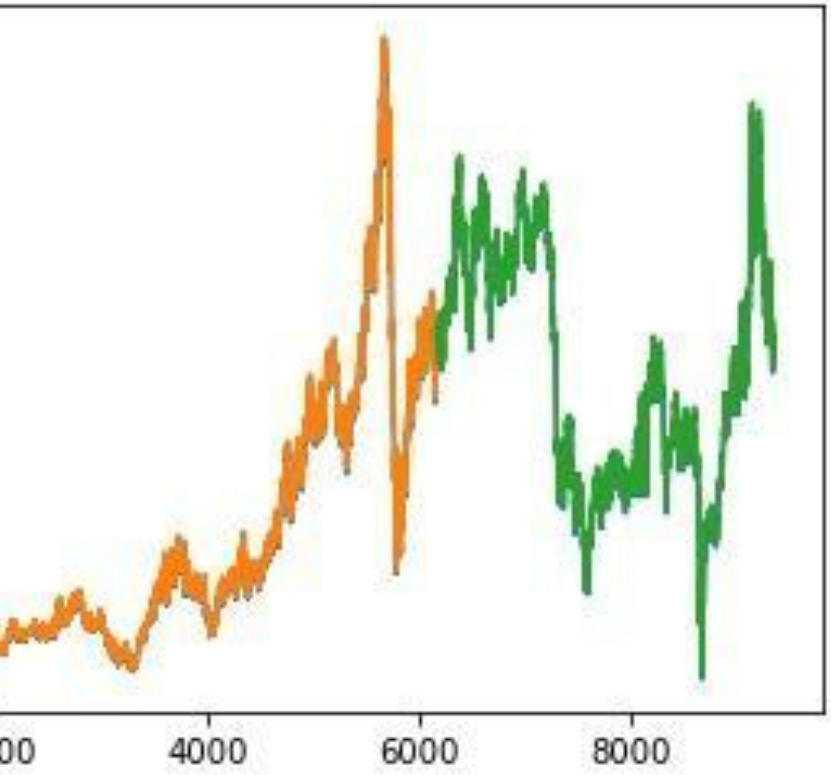
Simple Averages Met

Train: 6159
Test: 3173



rmse:2247105882177780.16

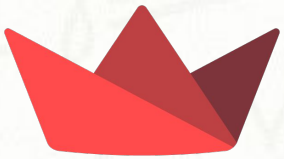
Long Short Term Memory (LSTM) Model



Train Score: 1.19 RMSE
Test Score: 2.04 RMSE

Train-R2 score: 0.9973721176476127
Test-R2 Score: 0.99215850176137

Model Deployment using Streamlit



Streamlit

Crude Oil Price Forecasting

This data app uses Facebook's open-sources prophet library to automatically genera
future forecast values from imported dataset. You'll be able to import your data set fr
a CSV file, visualize trend and feature, analyze forecast performance, and finally
download the created forecast

Import Data

Upload here

Drag and drop file here

Limit 200MB per file • CSV

Browse files

Select Forecast Period

How many periods would you like to forecast into the future?

1 − +

Uploading the data

Here, the user should uploa
and enter the number of for
for which he want the predic

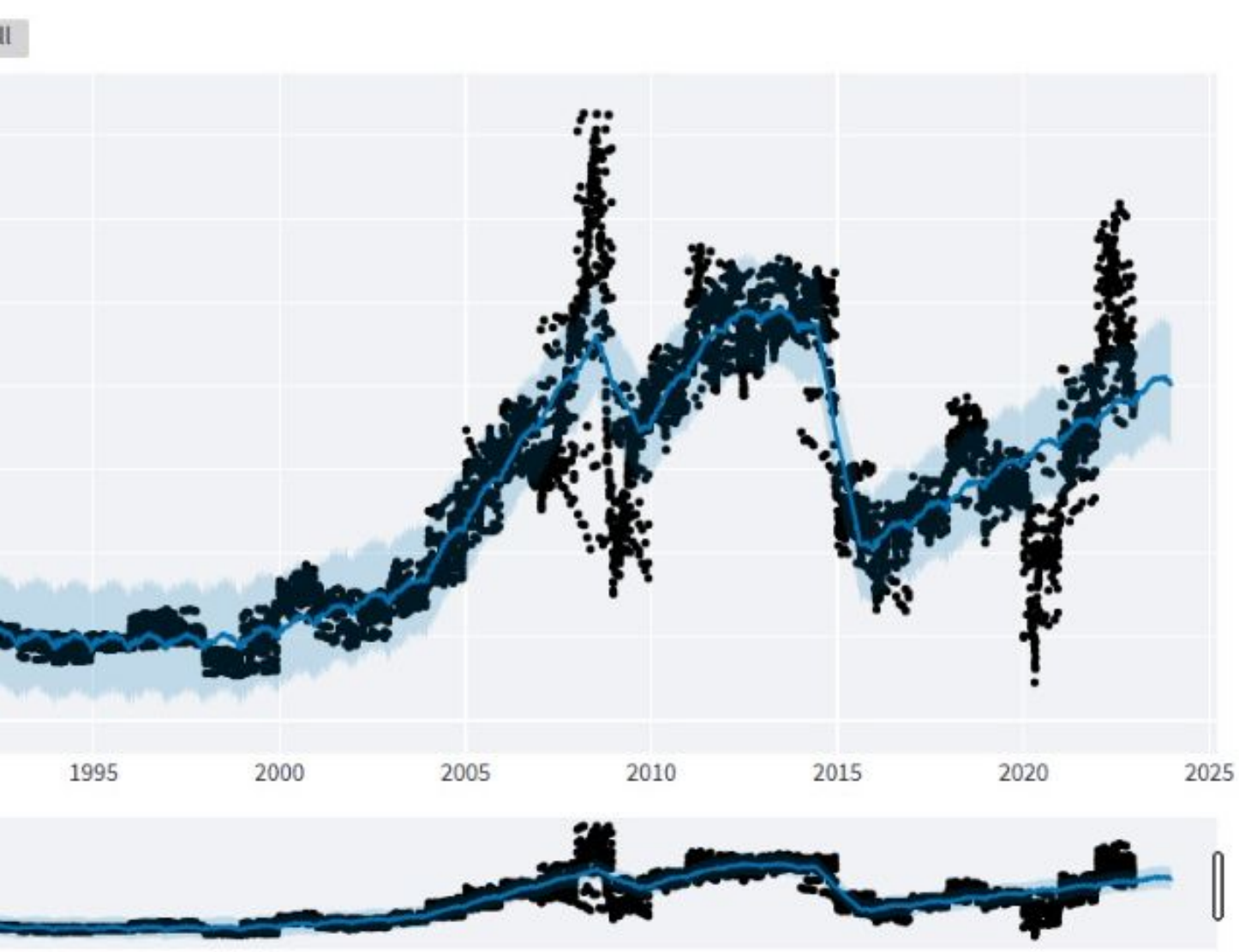
Visualized forecast data

	ds	yhat	yhat_lower	yhat_upper
9700	2023-12-12T00:00:00	80.2908	66.3457	93.8470
9701	2023-12-13T00:00:00	80.0786	66.1881	95.3446
9702	2023-12-14T00:00:00	80.5899	67.2124	94.7726
9703	2023-12-15T00:00:00	80.1658	65.1791	93.7202
9704	2023-12-16T00:00:00	80.9960	66.8745	96.0558
9705	2023-12-17T00:00:00	80.1438	65.8343	94.2432
9706	2023-12-18T00:00:00	80.5524	66.8627	94.5782
9707	2023-12-19T00:00:00	80.3275	65.6418	94.3695
9708	2023-12-20T00:00:00	80.2018	65.6389	94.9570
9709	2023-12-21T00:00:00	80.8024	65.8381	94.8606

This visual shows future predicted values of the next 365 days. "yhat" is th predicted value, and the upper and lower limits are (by default) 80% confidence intervals.

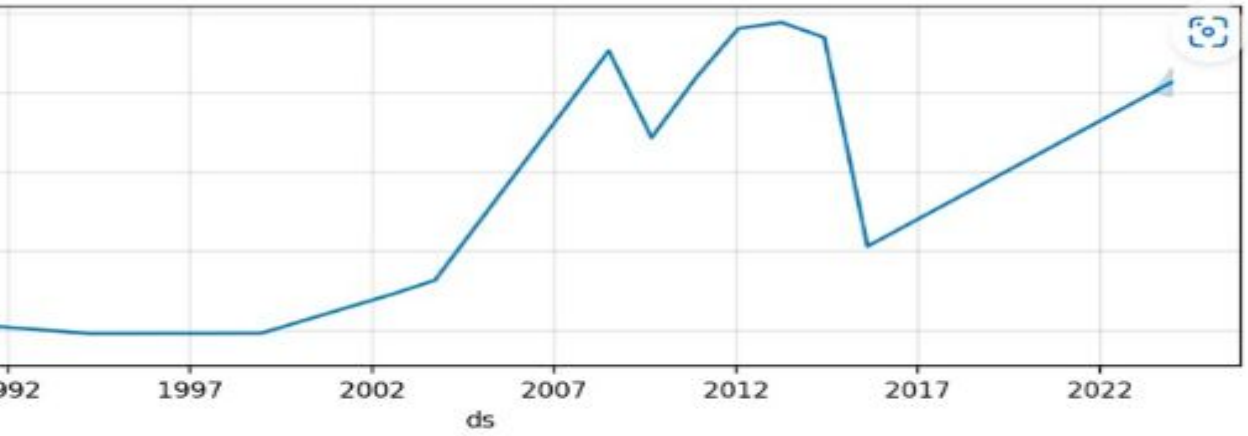
Visualizations

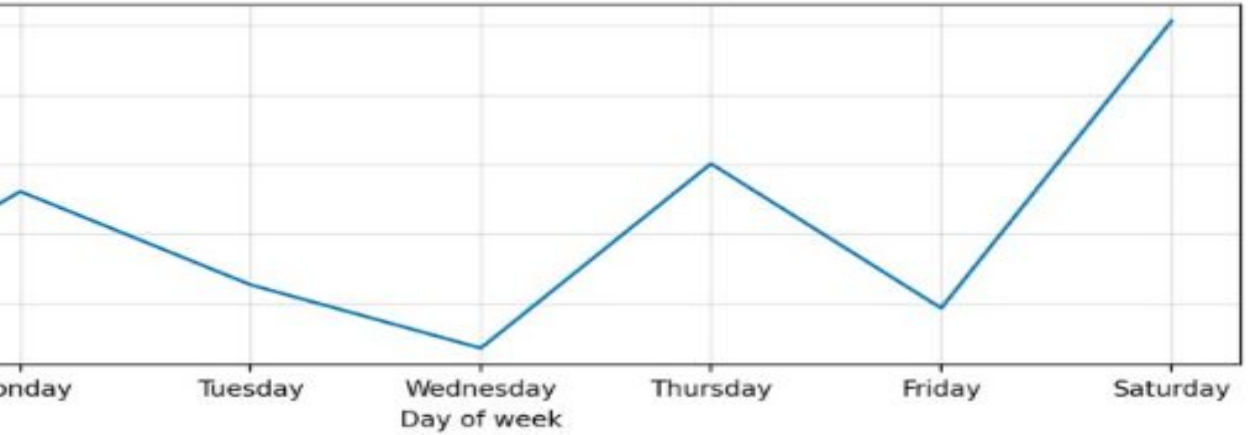
t

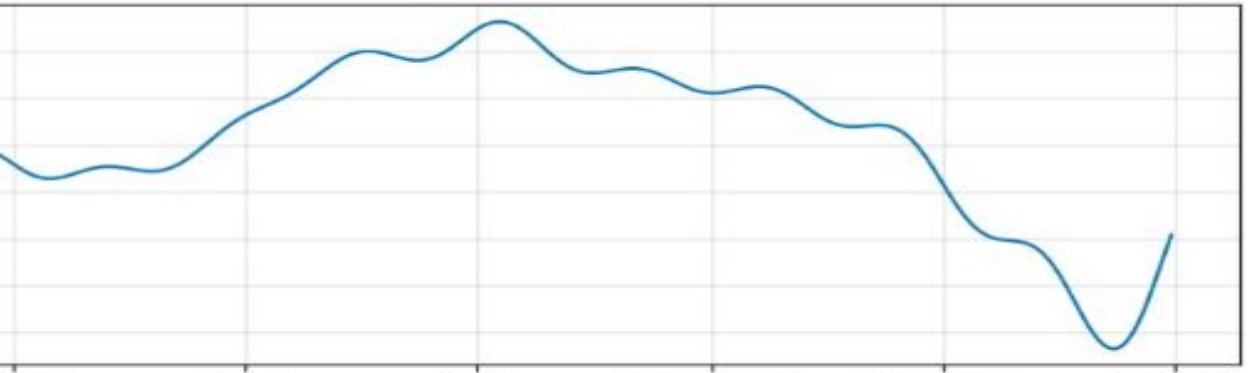


- This visualiz
  shows the ac
  values (“**Bla**
  and predicte
  (“**Blue line**”)
  time.

l components







- This visuals shows a
  trend of predicted val
  week trends, and yea
  dataset covers multip
  The blue shaded are
  upper and lower conf
  interval

Challenges faced?

The part of the project where we have faced the difficulty was in the model building stage deploying the final model. We ran into many errors in the process.

How did you overcome?

How did you overcome: We finished the deployment for our project on time, even after facing many errors because of our team work and with the help of our mentors and projects team.

Thank you