1. Introduction

Crude oil prices are influenced by macroeconomic indicators such as the strength of the U.S. dollar, inflation rates, and the U.S. Consumer Price Index (CPI). These factors serve as key indicators in understanding how oil prices move relative to broader economic conditions.

Crude oil, being a globally traded commodity, is sensitive to changes in the U.S. dollar index, which reflects the dollar's relative value to major world currencies. Inflation and CPI are additional indicators that help gauge economic stability and purchasing power, which can impact the global demand for oil.

Problem Statement

Crude oil prices are highly volatile and influenced by macroeconomic factors like the U.S. dollar strength, inflation rates, and the Consumer Price Index (CPI), making accurate forecasting is challenging for businesses and policymakers. A predictive model that effectively integrates these indicators is needed to improve forecasting accuracy, mitigate financial risks, and enhance strategic decision-making in industries reliant on oil.

Dataset:

The dataset covers the relationship between crude oil prices and other external economic factors such as the strength of U.S. dollar, inflation rates, and the U.S. Consumer Price Index (CPI) from the year 2014 to 2024.

Data Dictionary:

Column Name	Data Type	Description
Date	Date/Time	The date of the observation in MM/DD/YY format.
Crude_Price	Float	The price of crude oil (USD per barrel) at the given date.
USD_Currency_Index	Float	A measure of the value of the U.S. dollar relative to a basket of foreign currencies.
US_Inflation_Rate	String	The inflation rate in the U.S. as a percentage (e.g., "2.1%").
US_Inflation	Float	The inflation factor, a decimal representation of the inflation rate (e.g., 0.021 for 2.1%).
US_CPI	Float	The U.S. Consumer Price Index, which measures changes in the price level of consumer goods.

2. Background Work

The data in the dataset is collected from Bloomberg. First the crude oil price and then the exogenic factors which affect the crude oil price is collected and added into the dataset.

Crude oil prices are influenced by various external factors, particularly the U.S. dollar, inflation rates, and economic indicators like the Consumer Price Index (CPI). These factors impact global supply and demand dynamics, causing price fluctuations.

2.1 U.S. Dollar and Oil Prices

Since crude oil is traded globally in U.S. dollars, the **USD Currency Index** plays a crucial role. A strong dollar typically leads to lower oil prices as oil becomes more expensive for other countries, reducing demand.

2.2 Inflation and CPI

Inflation, measured through the **US Inflation Rate** and **US CPI**, affects purchasing power and consumer behavior. Higher inflation can reduce energy demand, leading to lower crude oil consumption and price changes.

2.3 Prior Research

Studies show an inverse relationship between oil prices and the U.S. dollar. Inflation and CPI also influence oil prices, particularly during periods of economic uncertainty. These relationships are essential for understanding and forecasting crude oil price trends.

3. Need for the Solution

Understanding and forecasting crude oil prices are vital for global economies, businesses, and policymakers, as oil is a critical energy source and its price volatility can disrupt markets. However, predicting crude oil prices remains a challenge due to the complexity of external factors such as currency fluctuations, inflation, and economic conditions.

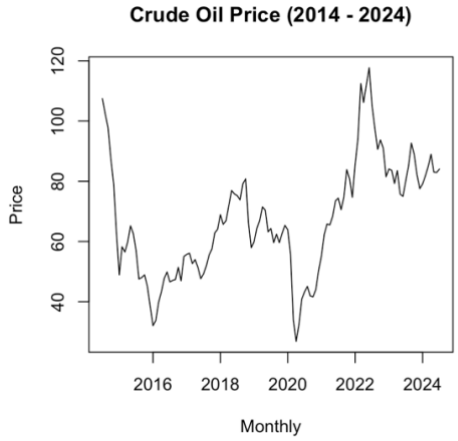
There is a growing need for robust models that can accurately analyze how exogenous factors, like the **USD Currency Index**, **US Inflation Rate**, and **CPI**, impact crude oil prices. Traditional forecasting methods often overlook these interconnected relationships, leading to gaps in decision-making for governments, investors, and industries reliant on oil. A solution that integrates these external factors can provide a more comprehensive understanding of price movements and help mitigate risks associated with crude oil price volatility.

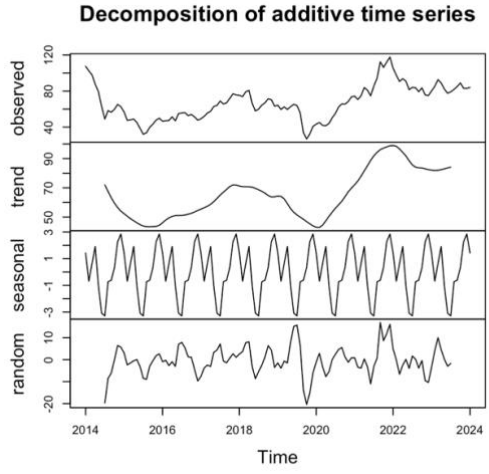
This solution is particularly crucial for industries like energy, transportation, and manufacturing, where fluctuations in oil prices can lead to increased operational costs, reduced profits, and market instability. Additionally, policymakers can use such a solution to make informed decisions regarding energy regulations, subsidies, and future economic planning.

4. Proposed Solution

The proposed solution involves developing a predictive model that integrates key exogenous factors—such as the **USD Currency Index**, **US Inflation Rate**, and **US Consumer Price Index (CPI)**—to forecast crude oil prices more accurately. By leveraging advanced statistical techniques or machine learning algorithms, this model can capture the complex relationships between these economic indicators and crude oil price movements.

Time Series Plot

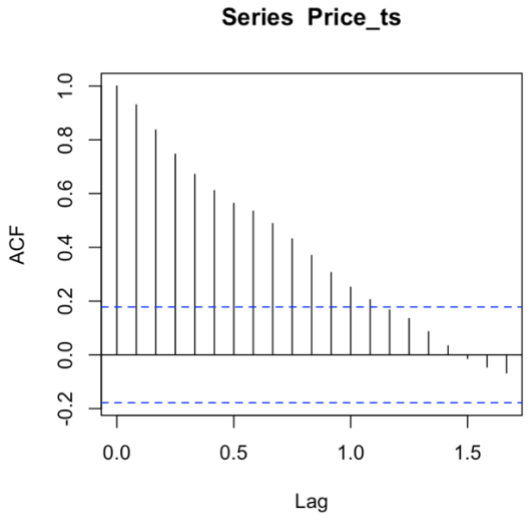




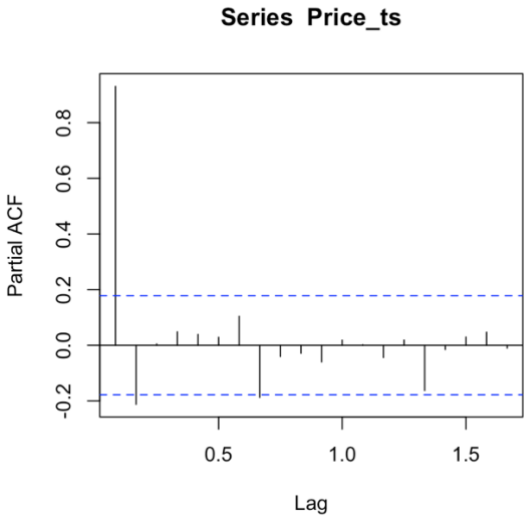
The graph shows that crude price has both trend and seasonality

1.ARIMA

ACF Test



PACF Test



Graph is dying down. so p value in p,d,q in 0. One spike is above blue line, so q is 1

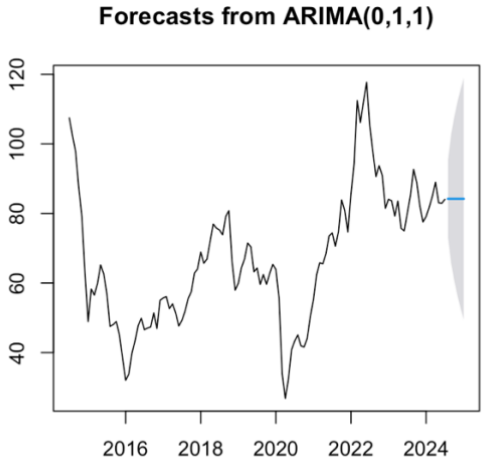
ADF Test

```
        Augmented Dickey-Fuller Test

data:  Price_ts
Dickey-Fuller = -3.2192, Lag order = 4, p-value = 0.08802
alternative hypothesis: stationary
```

p value is 0.088 >0.05 so it is not stationary, so make it stationary. no. of differentiation is 1 ie, d=1 is needed to make the series stationary

Forecasted Value Plot

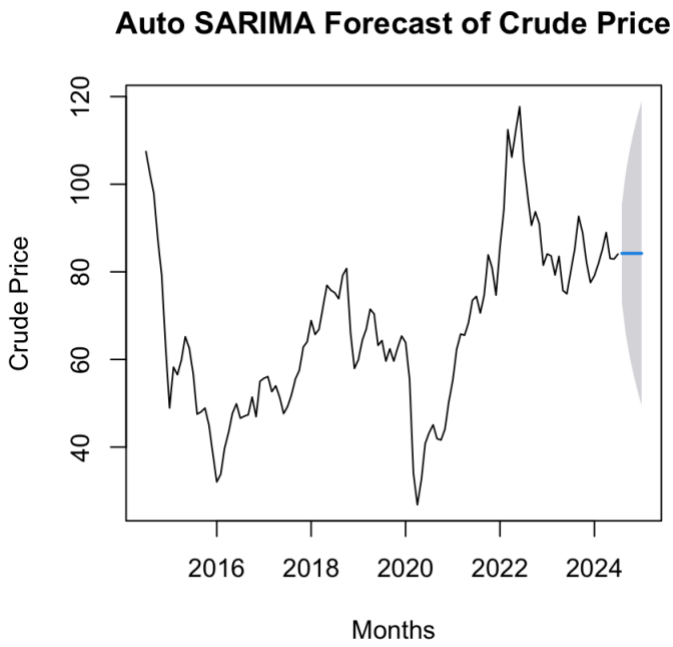


Leader Board

Model	RMSE	AIC	BIC
ARIMA	5.669	762.065	767.640

2.SARIMA

Forecasted Value Plot



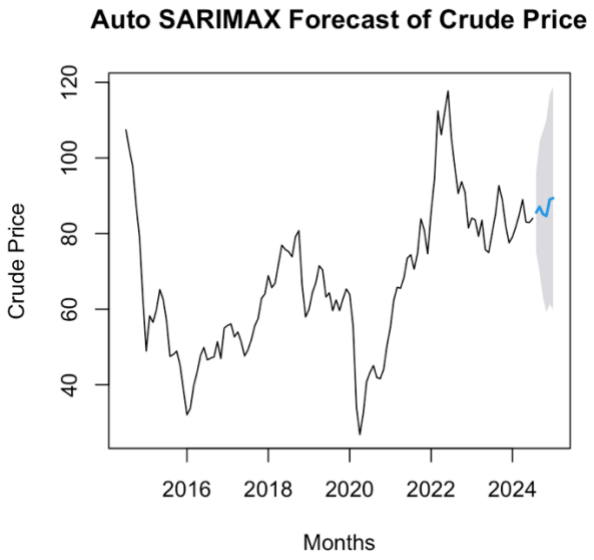
Leader Board

Model	RMSE	AIC	BIC
ARIMA	5.669	762.065	767.640
SARIMA	5.441	721.997	735.408

3.SARIMAX

Here all the three exogenic variables such as USD_Currency_Index, US_Inflation and US_CPI are considered

Forecasted Value Plot

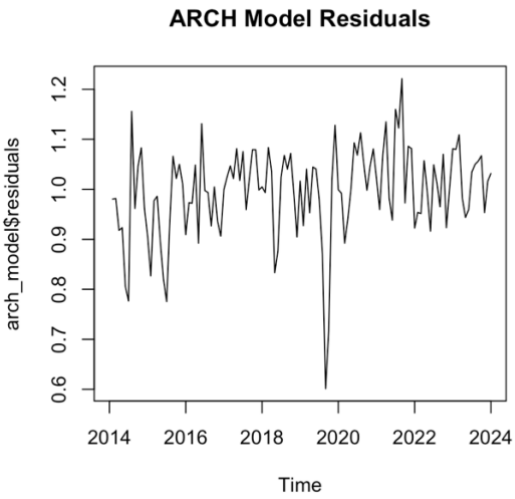


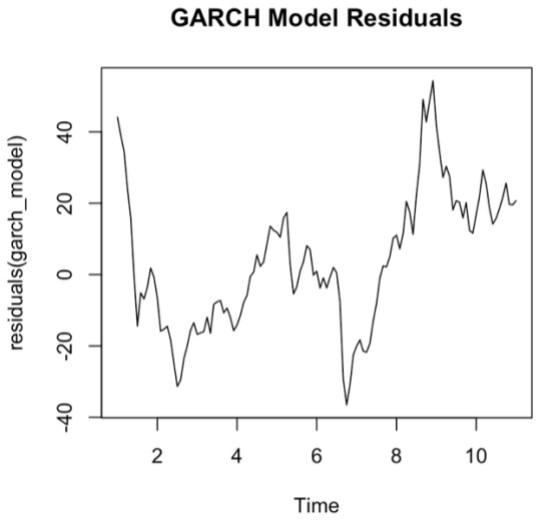

```
> print(Crude_Price_Auto_SARIMAX_Forecast)
         Point Forecast    Lo 95     Hi 95
Aug 2024       85.59017 75.11156  96.06879
Sep 2024       87.07149 69.91092 104.23205
Oct 2024       85.14233 63.19234 107.09233
Nov 2024       84.67607 59.34566 110.00649
Dec 2024       89.01488 61.29783 116.73192
Jan 2025       89.38108 59.96882 118.79334
```

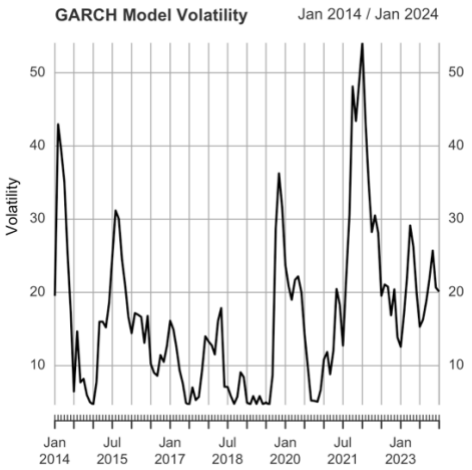
Leader Board

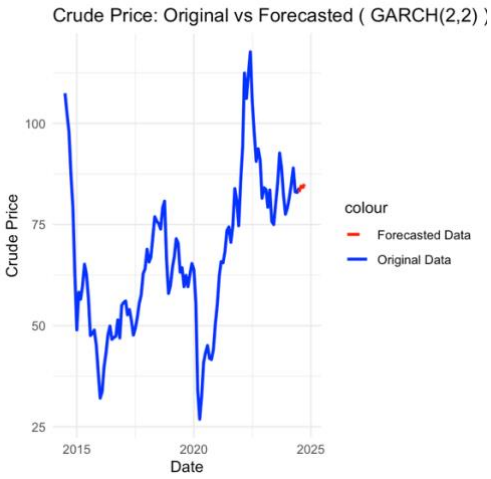
Model	RMSE	AIC	BIC
ARIMA	5.669	762.065	767.640
SARIMA	5.441	721.997	735.408
SARIMAX	5.189	761.798	784.164

4.ARCH and GARCH









ARCH Model Residuals:

- **Plot Characteristics**: The residuals appear to fluctuate around a mean value, showing periods of high and low volatility.
- **Interpretation**: The residuals don't seem to exhibit strong volatility clustering (where large residuals follow large residuals, and small follow small), though there are a few periods of high spikes. The volatility seems to be somewhat stable over time, indicating the ARCH model might capture some of the variability in the data, but the volatility dynamics seem less prominent.

GARCH Model Residuals:

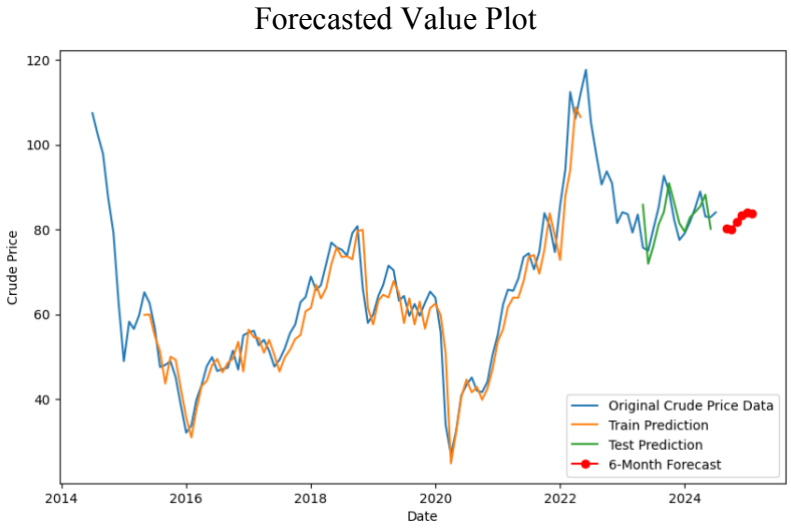
- **Plot Characteristics**: The residuals show clearer signs of volatility clustering, especially toward the latter part of the time series. There is an evident period where the residuals increase dramatically, indicating that the GARCH model captured periods of increasing volatility more effectively.
- **Interpretation**: The GARCH model is designed to account for both short-term shocks and long-term volatility, and the graph shows larger fluctuations that cluster over time. The model might be capturing volatility more accurately than the ARCH model, as it shows more structure, especially around periods of increasing residual variance.

- The **ARCH model** seems to provide a relatively stable volatility pattern, though there is less evidence of strong clustering.
- The **GARCH model** shows stronger evidence of capturing volatility clustering, with clearer patterns of residual variability. This suggests the GARCH model might be better suited for your dataset in terms of modeling time-varying volatility.

Leader Board

Model	RMSE	AIC	BIC
ARIMA	5.669	762.065	767.640
SARIMA	5.441	721.997	735.408
SARIMAX	5.189	761.798	784.164
ARCH	53.067	994.080	1002.47
GARCH	53.299	994.795	1005.98

5.LSTM



Leader Board

Model	RMSE	AIC	BIC
ARIMA	5.669	762.065	767.640
SARIMA	5.441	721.997	735.408
SARIMAX	5.189	761.798	784.164
ARCH	53.067	994.080	1002.47
GARCH	53.299	994.795	1005.98
LSTM	5.360	63849.013	84235.581

Model Comparison

1. Root Mean Square Error (RMSE):

- RMSE measures the average difference between the predicted and actual values, with lower values indicating better model accuracy.
- **SARIMAX** has the lowest RMSE (5.189), suggesting it provides the most accurate predictions among the models.
- **ARIMA** and **SARIMA** also perform well with RMSE values of 5.669 and 5.441, respectively.
- **ARCH** and **GARCH** models have very high RMSE values (53.067 and 53.299), indicating poor predictive accuracy.
- **LSTM** has a competitive RMSE of 5.360, but its AIC and BIC values (are extremely high.

2. Akaike Information Criterion (AIC):

- AIC measures model quality, balancing goodness of fit with model complexity. Lower AIC values indicate better models.
- **SARIMA** has the lowest AIC (721.997), making it the best model by this criterion.
- **SARIMAX** (761.798) and **ARIMA** (762.065) are close but have slightly higher AIC values.
- **ARCH** and **GARCH** have the highest AIC values (994.080 and 994.795), further confirming their poor performance.
- **LSTM** has a significantly high AIC (63849.013), likely due to its complex neural network architecture.

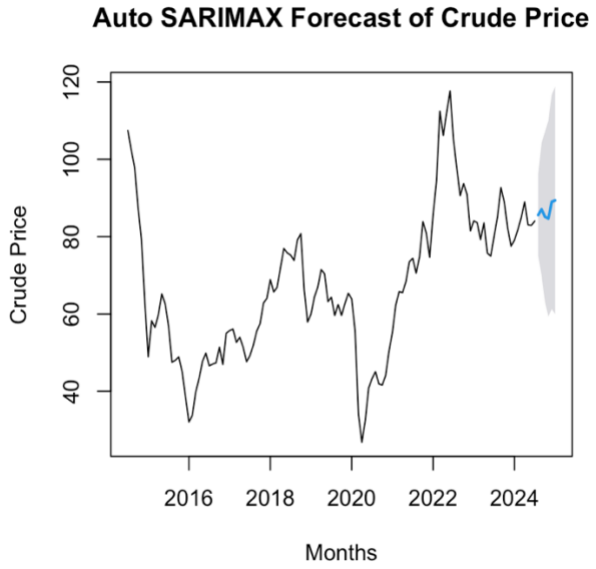
3. Bayesian Information Criterion (BIC):

- BIC is similar to AIC but penalizes more for model complexity. Again, lower values indicate better models.
- **SARIMA** has the lowest BIC (735.408), reinforcing that it's a strong candidate for the best model.
- **ARIMA** and **SARIMAX** follow, with BIC values of 767.640 and 784.164, respectively.
- **ARCH** and **GARCH** have poor BIC values (1002.47 and 1005.98), confirming their complexity and poor fit.
- **LSTM** has a dramatically high BIC (84235.581), further suggesting it may be too complex for this dataset.

Choosing the Best Model

- **Best Model**: **SARIMAX** has the lowest RMSE (most accurate predictions), but **SARIMA** has the best balance between fit and complexity with the lowest AIC and BIC.
- **Worst Models**: **ARCH** and **GARCH** models have high RMSE, AIC, and BIC, indicating poor performance.
- **LSTM** performs well in terms of RMSE but has much higher AIC and BIC, suggesting overfitting or excessive complexity for this dataset.

5. Analysis of the Solution



1. **Historical Data (2014–2024)**:

- The solid black line represents historical crude oil prices from 2014 to 2024. The prices exhibit significant volatility, with notable peaks around 2018 and 2022, reaching as high as approximately 120 USD per barrel. Periods of decline are also visible, with lows around 40 USD per barrel.
- The price movements reflect real-world events, such as economic changes, geopolitical tensions, and fluctuations in global oil supply and demand.

2. **Forecast Period (Aug 2024–Jan 2025)**:

- The blue line represents the SARIMAX forecast for crude oil prices starting from August 2024 to January 2025. The values are:

Month	**Crude Price**
Aug 2024	85.59 USD
Sep 2024	87.07 USD
Oct 2024	85.14 USD
Nov 2024	84.68 USD
Dec 2024	89.01 USD
Jan 2025	89.38 USD

- These predicted values suggest a moderate range for crude prices between 84 and 89 USD during this forecast period. The prices remain relatively stable with slight fluctuations, which is indicative of a possible stabilization phase for crude oil.

3. Prediction Confidence Interval**:**

- The shaded gray area around the forecasted values represents the **confidence interval**, indicating the range within which the actual crude oil prices are likely to fall with a certain level of confidence (usually 95%).
- The widening of the gray band towards the end of the forecast period (late 2024 to early 2025) reflects increasing uncertainty in predictions as time progresses. This is typical for time-series models, where forecast accuracy tends to decrease the further out into the future it predicts.

4. General Trend**:**

- The forecast suggests that crude oil prices will hover around 85–89 USD per barrel through late 2024 into early 2025. This indicates a relatively stable period after the volatility seen in previous years.

The **SARIMAX model** predicts that crude oil prices are likely to remain steady, with minor fluctuations between 84 and 89 USD in the latter half of 2024. The widening confidence interval highlights the inherent uncertainty of predicting commodity prices, especially in the long term. Nonetheless, the model suggests no drastic spikes or drops in prices during this period, which could signal a period of stabilization in the crude oil market.

6. Recommendations

1. **Hedge Against Price Fluctuations**: Lock in crude oil prices (84–89 USD) to stabilize costs and minimize risks from minor fluctuations.
2. **Monitor Economic Indicators**: Track the USD, inflation, and CPI regularly to adjust strategies based on external economic changes.
3. **Scenario Planning**: Prepare for uncertainties by simulating different oil price scenarios to guide decision-making.
4. **Diversify Energy Sources**: Invest in renewable energy to reduce dependence on oil and mitigate long-term risks.
5. **Leverage Short-Term Stability**: Take advantage of stable crude oil prices for investments like oil futures for consistent short-term returns.
6. **Evaluate and Adjust Pricing Strategies**: For businesses directly affected by oil prices, such as transportation or manufacturing, consider adjusting pricing strategies for products and services to reflect anticipated price stability or fluctuations.
7. **Review and Adjust Capital Expenditure Plans**: Given the projected stability in crude oil prices, consider accelerating or delaying capital expenditures related to oil-dependent operations. Use the period of relative stability to reassess long-term investment plans and align them with future market expectations.

7. Conclusion

The forecasted stability of crude oil prices between 84 and 89 USD per barrel from late 2024 to early 2025 provides a window of opportunity for businesses, investors, and policymakers to make informed decisions. Despite this period of projected steadiness, the inherent uncertainty in commodity markets requires continuous monitoring of key economic indicators such as the USD index, inflation rates, and the Consumer Price Index (CPI).

Implementing strategies such as hedging, optimizing supply chain management, diversifying energy sources, and leveraging advanced analytics can help mitigate risks and capitalize on the current stability in oil prices. By adopting a proactive approach that integrates scenario planning, risk management, and policy advocacy, stakeholders can better prepare for potential market shifts and safeguard against future volatility.

Ultimately, while the SARIMAX model suggests a moderate and stable range for crude oil prices, stakeholders should remain vigilant and adaptive to changes in the global economic environment to ensure resilience and sustained growth in the face of ongoing uncertainties.