**Crude Oil Price Forecasting and Visualisation Project**

**Overview**

This project focuses on predicting WTI crude oil prices for the period August 2023 to December 2024 and forecasting the price for January 2025, using machine learning models (Random Forest and XGBoost) and visualising the results in an interactive Tableau dashboard.

The goal was not just to forecast prices but to connect them with real-world geopolitical and economic events, supply-side movements (like OPEC production), and model accuracy, creating a full storytelling experience.

**Objectives**

* Forecast crude oil prices using ML models based on macroeconomic, supply, and sentiment indicators.
* Analyse model performance with clear error metrics and visualisations.
* Map key global events alongside price movements to show market impacts.
* Visualise OPEC production trends as a supply-side factor.
* Deliver everything through a clean, professional Tableau dashboard.

**Data Used**

* WTI Crude Oil Spot Prices (Yahoo Finance, EIA)
* Brent Crude Oil Prices
* OPEC Crude Oil Production Levels
* Macroeconomic Indicators:
* US Consumer Price Index (CPI)
* Federal Reserve Interest Rates
* US Dollar Index (DXY)
* News Sentiment Score (based on major global news coverage)

**Process and Methodology**

1.⁠ ⁠Data Preparation

* Cleaned missing values and normalized where required.
* Created additional features like lagged prices, rolling averages, and percentage changes.
* Ensured a time-based train-test split (80% train, 20% test) to simulate real-world forecasting (future unseen data).

**2.⁠ ⁠Modelling**

* Trained two machine learning models: Random Forest and XGBoost Regressor.
* Focused on predicting WTI prices based on the macro indicators and engineered features.
* Evaluated performance based on prediction errors (RF\_Error and XGB\_Error).

**3.⁠ ⁠Predictions**

* Predicted WTI crude oil prices from August 2023 to December 2024.
* Compared model predictions against actual prices to analyze accuracy.

**4.⁠ ⁠Visualization (Tableau Dashboard)**

Built a professional dashboard to tell the story beyond the numbers.

Included:

* Price trends over time (WTI and Brent Crude)
* Model comparison (Actual vs Predicted WTI prices)
* Supply movement (OPEC Production Trend)
* Model error visualizations (Random Forest and XGBoost)

Annotated key global events like:

* OPEC Production Cuts
* Israel–Hamas Conflict
* Fed Rate Pause
* U.S. SPR (Strategic Petroleum Reserve) refill
* China’s Economic Stimulus
* U.S. Election Campaigns and Results

**Final Dashboard Sections**

* Crude Oil Price Movement with Major Events
* OPEC Production Trend
* Actual vs Predicted WTI Price Comparison (Random Forest & XGBoost)
* Random Forest Error Distribution (Histogram)
* XGBoost Mean Absolute Error (Simple KPI)

**Published on Tableau Public:** [<https://public.tableau.com/app/profile/supriyo.banerjee8197/viz/CrudeOil_ForecastingModel_Dashboard/Dashboard1?publish=yes>]

**Key Highlights**

* Accurately captured price behavior trends and market influences.
* Forecasting models stayed relatively stable across high-volatility periods.
* Clear linkage between macroeconomic events and crude oil price movements.
* Created a storytelling dashboard that goes beyond just showing numbers — it explains why and how things happened.

**Tools and Technologies**

* Python (Pandas, Scikit-Learn, XGBoost)
* Jupyter Notebook
* Tableau Public
* Microsoft Excel (for minor data formatting)

**How to Navigate the Project**

Explore the Tableau Dashboard (link above) for the full experience.

Review the Python code (available on request or GitHub for full transparency into the modelling process.

Analyse how real-world events were integrated directly into the visual storytelling.

**Future Enhancements**

* Experiment with ARIMA or LSTM-based models for more sequential forecasting.
* Add live updating capability for real-time dashboard refresh.
* Expand feature engineering to include more granular oil market variables (refinery throughput, storage levels, etc.)

**About This Project**

This project was independently conceptualized, modeled, and visualized to demonstrate a full end-to-end machine learning + analytics + visualization pipeline in the context of a dynamic real-world market like crude oil.

The focus was not only on technical modeling but also on making the insights accessible and business-relevant.