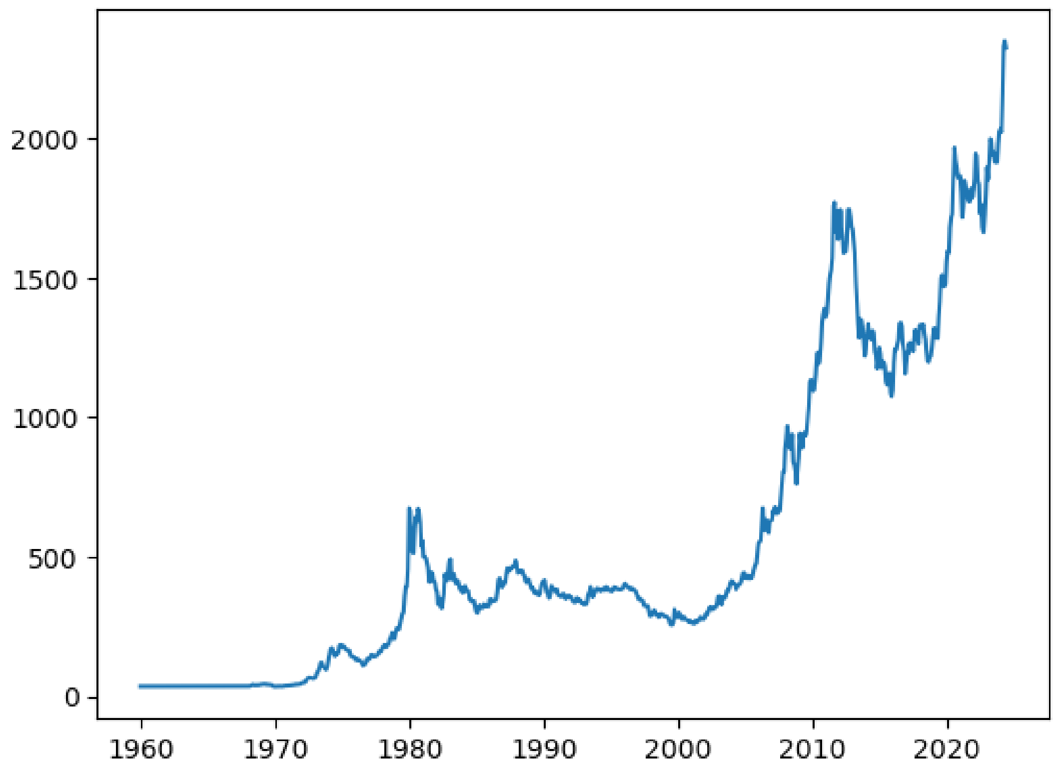
PART B – PYTHON – SECTION B



Interpretation of the Gold Price Time Series Plot

Description:

- The graph displays the historical gold prices from the 1960s to 2020.
- The x-axis represents the years, while the y-axis represents the gold price in dollars per troy ounce.

Analysis:

1. **Initial Stability (1960s to 1970s):**
   - Gold prices remained relatively stable during the 1960s and early 1970s, indicating low volatility and minimal market fluctuation.
2. **First Major Spike (Late 1970s to Early 1980s):**
   - A significant increase in gold prices is observed during the late 1970s, peaking around 1980.

    - This spike can be attributed to various economic factors such as high inflation, geopolitical tensions, and changes in monetary policies.
3. **Price Decline and Stability (1980s to Early 2000s):**
    - Post-1980, gold prices experienced a decline and a prolonged period of relative stability with minor fluctuations.
    - This period indicates a market adjustment and stabilization phase.
4. **Second Major Spike (2000s to Early 2010s):**
    - Starting from the early 2000s, there is a noticeable and sharp increase in gold prices, peaking around 2011-2012.
    - Factors contributing to this rise include the financial crisis of 2008, increased demand for safe-haven assets, and monetary easing policies.
5. **Recent Trends (2010s to 2020):**
    - After the 2011 peak, gold prices experienced a correction but remained relatively high compared to historical levels.
    - From 2018 onwards, another upward trend is observed, culminating in a new peak around 2020.
    - This recent surge can be linked to global economic uncertainty, trade tensions, and the COVID-19 pandemic driving demand for gold as a safe asset.

Conclusions:

- The gold market has experienced significant volatility, particularly during periods of economic uncertainty and geopolitical tensions.
- The long-term trend indicates that gold remains a valuable asset, with prices generally increasing over extended periods.
- Recent trends suggest continued interest in gold as a hedge against economic instability and inflation.

```
Single LSTM with hidden Dense...
Model: "sequential"
```

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 64)	17,920
dense (Dense)	(None, 32)	2,080
dense_1 (Dense)	(None, 1)	33

```
Total params: 20,033 (78.25 KB)
Trainable params: 20,033 (78.25 KB)
Non-trainable params: 0 (0.00 B)
```

Interpretation of the LSTM Model Architecture

6. **LSTM Layer**
   - **Purpose:** Captures temporal dependencies in the input data.
   - **Output Shape:** (None, 64)
   - **Parameters:** 17,920
7. **First Dense Layer**
   - **Purpose:** Reduces dimensionality of LSTM output.
   - **Output Shape:** (None, 32)
   - **Parameters:** 2,080
8. **Output Dense Layer**
   - **Purpose:** Outputs a single prediction value.
   - **Output Shape:** (None, 1)
   - **Parameters:** 33
9. **Overall Model**
   - **Total Parameters:** 20,033 (all trainable)
   - **Complexity:** Moderately complex, suitable for time series forecasting.

Key Points

- The LSTM layer learns patterns over time.
- Dense layers refine the output for prediction.
- The model is designed to predict a single continuous value, such as gold prices.

Train Score: 18.28 RMSE

Test Score: 219.62 RMSE

Train Score: 18.28 RMSE

- **Meaning:** The Root Mean Squared Error (RMSE) on the training data is 18.28.
- **Interpretation:** The model performs well on the training data, with relatively low error. This indicates that the model has learned the patterns in the training data effectively.

Test Score: 219.62 RMSE

- **Meaning:** The RMSE on the test data is 219.62.
- **Interpretation:** The model has a much higher error on the test data compared to the training data. This suggests that the model may be overfitting, meaning it performs well on training data but fails to generalize to unseen data.

Stationarity and Cointegration Tests

ADF Test Results

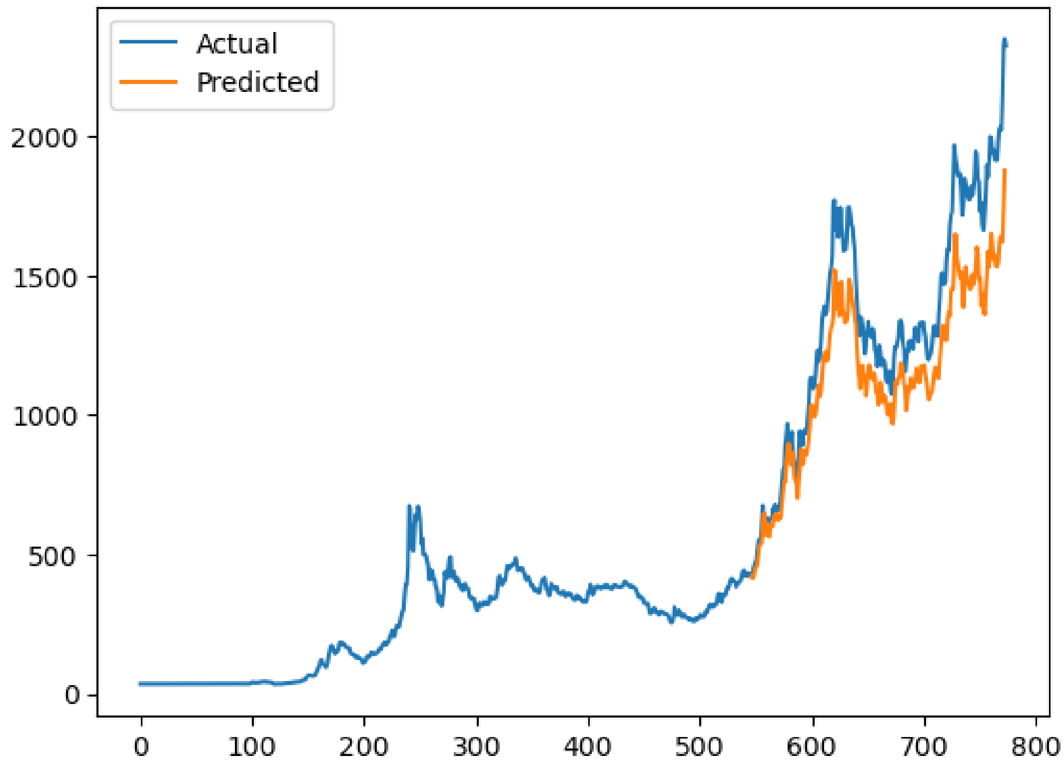
Interpretation:

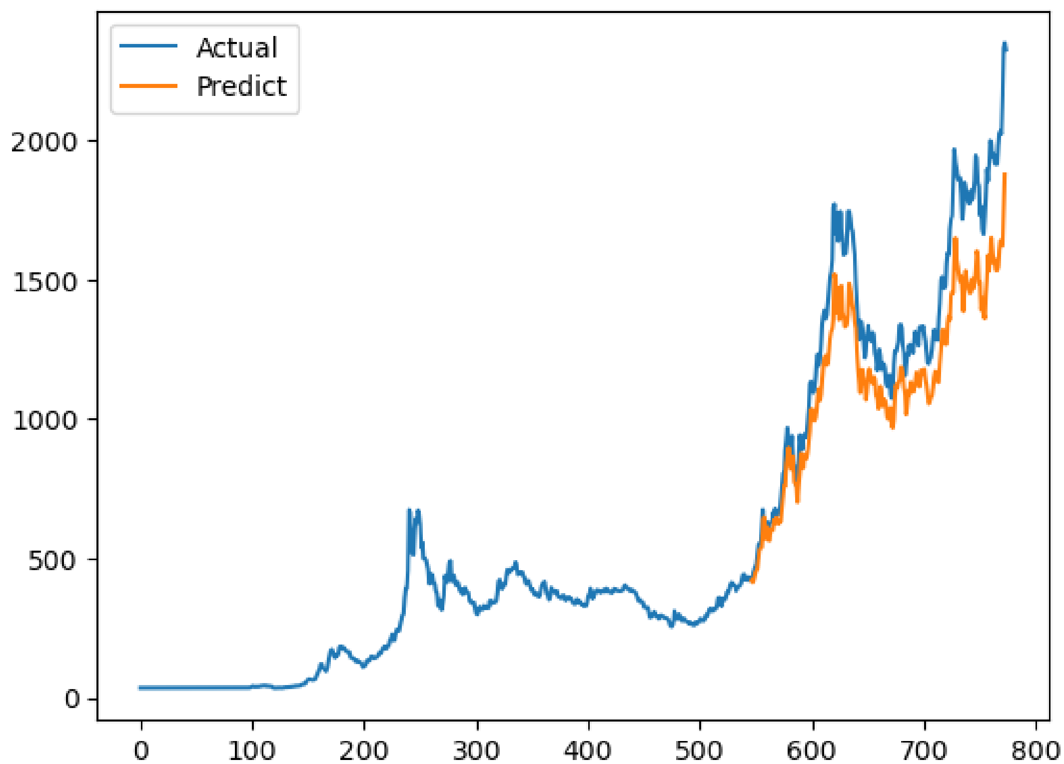
• Non-stationary time series data indicate that their statistical properties, such as mean
and variance, change over time, making it challenging to model and forecast accurately without transformation (e.g., differencing).

Johansen Cointegration Test

Interpretation:

• Cointegration suggests that although the individual commodity prices are non-stationary, they move together in the long run, maintaining a consistent relationship.
This is useful for building multivariate time series models like VAR (Vector AutoRegression), which can exploit these relationships for better forecasting accurac


Actual
Predicted
2000
1500
1000
500
0
0
100
200
300
400
500
600
700
800


Actual
Predict
2000
1500
1000
500
0
0
100
200
300
400
500
600
700
800

Graph 1: Actual vs. Predicted (Labeled "Actual" and "Predicted")

Description:

- The blue line represents the actual gold prices over time.
- The orange line represents the predicted gold prices from the model.

Analysis:

- The predicted prices (orange line) generally follow the trend of the actual prices (blue line), indicating that the model captures the overall pattern of gold price movements.
- However, there are noticeable discrepancies where the predicted prices do not match the actual prices exactly, especially during periods of rapid change in the actual prices.
- The model seems to perform better during stable periods but struggles to predict sharp spikes and drops accurately.

Graph 2: Actual vs. Predicted (Labeled "Actual" and "Predict")

Description:

- Similar to the first graph, the blue line represents the actual gold prices, and the orange line represents the predicted prices.

Analysis:

- The patterns observed are consistent with the first graph: the model's predictions closely follow the actual prices but with some noticeable differences.

- The model captures the overall trend well but has difficulty matching the exact values during periods of significant price changes.
- Both graphs show the model's limitations in predicting exact price points, particularly during high volatility.

Section A – Part B

Logistic Regression Model

Model Summary: The logistic regression model estimates the relationship between the predictor
variables and the log odds of the binary outcome variable y (whether the client subscribes to a term
deposit).

• **Statistical Significance:**

o Significant predictors ($p < 0.05$) include jobretired, educationuniversity.degree,
contacttelephone, monthaug, monthjun, monthmar, monthmay, poutcomesuccess, and
emp.var.rate.

Interpretation:

• The logistic regression model demonstrates strong predictive performance with high accuracy
and an excellent AUC-ROC score.

• Key predictors include contact method, month of contact, previous campaign outcomes, and
employment variation rate.

• The model suggests targeted strategies, such as focusing on specific months and previous
successful campaigns, to improve subscription rates.

Decision Tree Model

Model Structure: The decision tree splits the data based on the most significant predictor variables at
each node to predict the outcome.

• **Root Node:**

o nr.employed >= 5088: Splits the data into two branches.

▪ Left Branch: Clients with nr.employed >= 5088 (most likely not to subscribe)

▪ Further splits based on duration < 606

▪ Splits again based on euribor3m >= 1.4 and duration < 835

▪ Right Branch: Clients with nr.employed < 5088 (more likely to subscribe)

▪ Further splits based on duration < 166

▪ Splits again based on poutcome = failure, nonexistent and
duration < 249

• **Subsequent Nodes:**

o duration: Call duration is a critical factor in multiple splits.

o euribor3m, poutcome: Additional splits are influenced by these variables.

Interpretation:

• The decision tree model provides a clear, interpretable structure of decision rules.• Important factors influencing the decision include nr.employed, duration, euribor3m, and
poutcome.

• While the model's accuracy is comparable to the logistic regression model, its AUC-ROC is
lower, suggesting it is less effective at distinguishing between clients who will and will not
subscribe.

• The decision tree's interpretability makes it useful for understanding the hierarchy of factors
influencing subscription decisions.

Conclusion

• **Logistic Regression Model:** Shows slightly better performance with a higher AUC-ROC
value, indicating stronger discriminative power. It identifies key predictors and their influence
on subscription likelihood, useful for targeted marketing strategies.

• **Decision Tree Model:** Offers clear interpretability and insights into decision-making
processes but has slightly lower discriminative ability. It highlights important factors in a
hierarchical manner, which can be valuable for understanding the sequential importance of
predictors.

Decision Tree Model

1. **Clear Decision Rules for Sales Teams:**

o **Nr. Employed:** The decision tree identifies nr.employed (number of
employees) as a critical factor. Banks can instruct their sales teams to consider
this metric when prioritizing contacts, focusing more on clients from
companies with higher employment numbers, which might indicate economic
stability.

2. **Personalized Follow-Up Strategies:**

o **Duration of Calls:** The duration of previous calls plays a significant role in
the likelihood of subscription. Sales teams can be trained to spend more time

with potential clients, ensuring that all their questions are answered and building a stronger relationship, thereby increasing the chances of subscription.

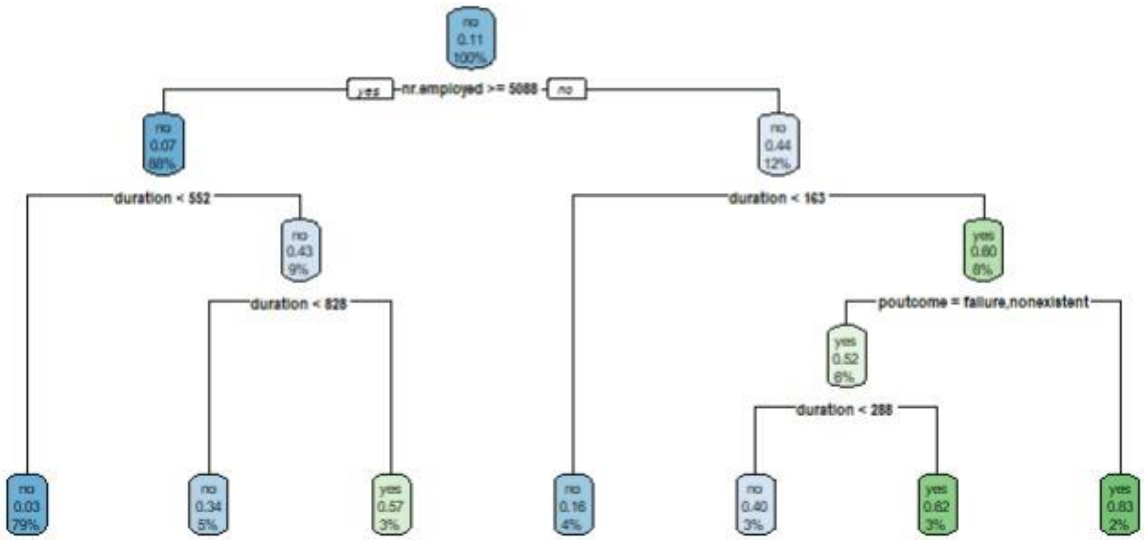
3. **Outcome of Previous Contacts:**

o **Previous Campaigns:** The tree shows that clients with poutcome of "failure"
or "nonexistent" in previous campaigns have a different likelihood of subscription. This insight helps banks to re-engage these clients with new offers or follow-up strategies tailored to address past objections or non-responsiveness.

4. **Economic and Market Conditions:**

o **Euribor Rate:** The model indicates that the euribor3m rate influences subscription decisions. Banks can monitor market conditions and adjust their
marketing strategies accordingly, offering more attractive rates or benefits during periods when the Euribor rate is high to compensate for client concerns.

Decision Tree Structure



AUC-ROC Curves