University of Chicago
MS Applied Data Science

GLOBAL GOLD PRICE PREDICTION

03/11/2025

Amy Kim, Cicily Mathew, Olivia Rumere

Course: Time Series Analysis and Forecasting Instructor: Fan Yang

AGENDA

Key takeaways:

- Business Problem and Scope
- Data Description
- Time Series Analysis
- Model Development
- Conclusion

BUSINESS PROBLEM

&

SCOPE

BUSINESS PROBLEM

Gold has long been a valuable asset and a key indicator of economic stability. This project aims to address the critical business challenge of accurately predicting gold prices with key external factors that influence price trends**.** By forecasting gold prices **for short-term period,** our model aims to **enhance investment decision-making, improve risk management strategies, and support more informed financial planning for stakeholders.**

SCOPE & OBJECTIVES

- Collect & process relevant data
- Macroeconomic factors as regressors
- Analyze the gold price time series
- Utilize different time series models
- Evaluate accuracy of different models

DATA OVERVIEW

Data Used: Gold price, Crude Oil Price, US Dollar Index, and S&P 500 index from **yfinance package**

Time span: From 2014-01-02 to 2025-02-28 (11 years), 2804 observations

	date	Gold Price	Oil Crude	US Dollar Index	S&P 500
		0049	69.7		00000
1	2025-02-27	2883.199951	70.349998	107.239998	5861.569824
2	2025-02-26	2916.800049	68.620003	106.419998	5956.060059
3	2025-02-25	2904.500000	68.930000	106.309998	5955.250000
4	2025-02-24	2947.899902	70.699997	106.599998	5983.250000
5	2025-02-21	2937.600098	70.400002	106.610001	6013.129883
6	2025-02-20	2940.000000	72.570000	106.339996	6117.520020
7	2025-02-19	2919.399902	72.250000	107.169998	6144.149902
8	2025-02-18	2931.600098	71.849998	107.050003	6129.580078
10	2025-02-14	2883.600098	70.739998	106.570000	6114.629883

Target Variable

Regressors

Details

Gold Price: US dollars per ounce

Oil Crude: US dollars per barrel

US Dollar Index: The strength of US dollar against a basket of other currencies

S&P 500: Stock market index value

PRE-PROCESSING

04 ***Train - Test Split***

Train-Test Split with **80:20** ratio

03 ***Data Transformation***

Implement
Log-Transformation & differencing

02 ***Data Exploration***

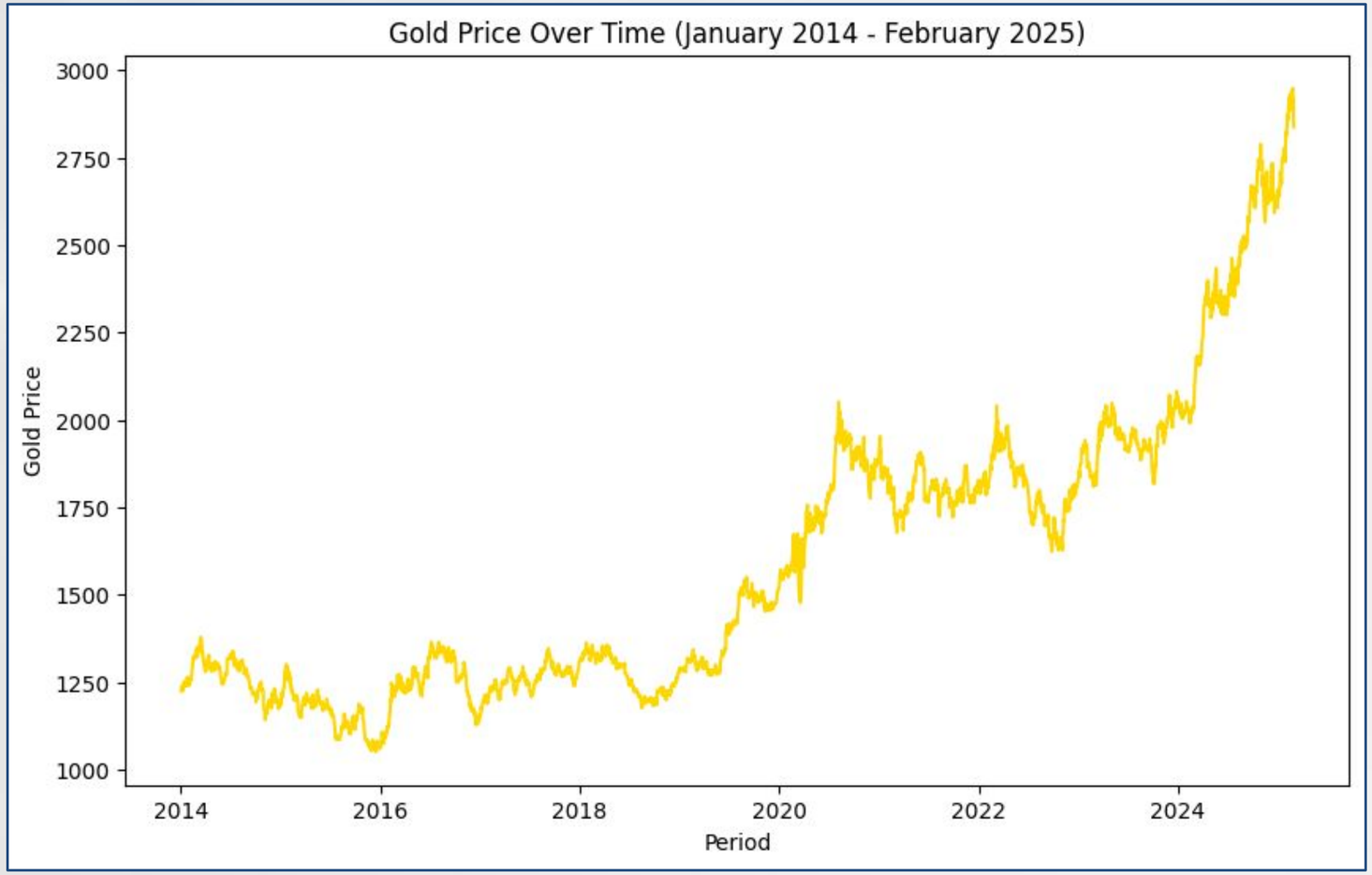
Assumption check on **stationarity** (ADF Test) and **normality** (QQ-Plot)

01 ***Data Collection***

Retrieve & merge relevant data from ***Yahoo Finance***

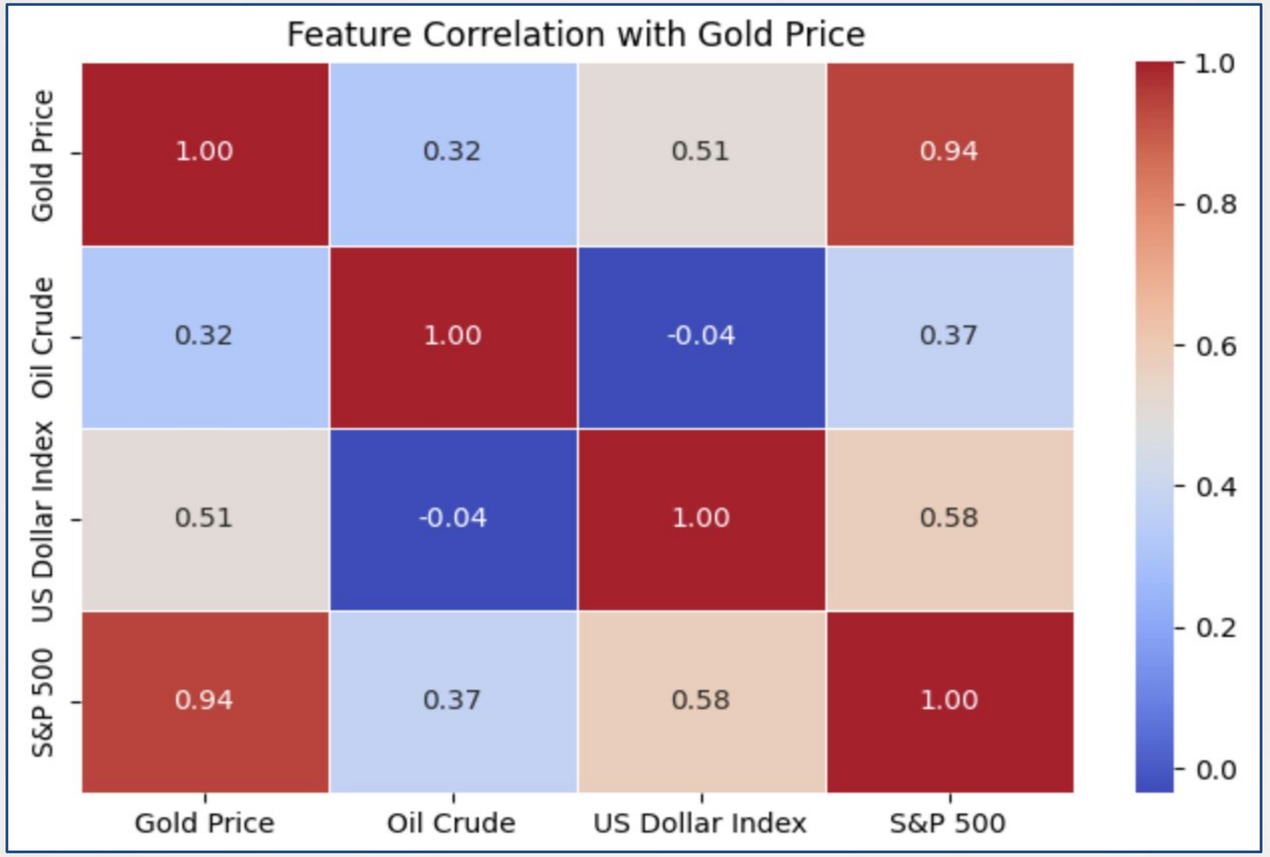
TIME SERIES ANALYSIS

TIME SERIES PLOT



- Shows a steady upward trend in gold prices over time, with significant growth accelerating especially after 2020.

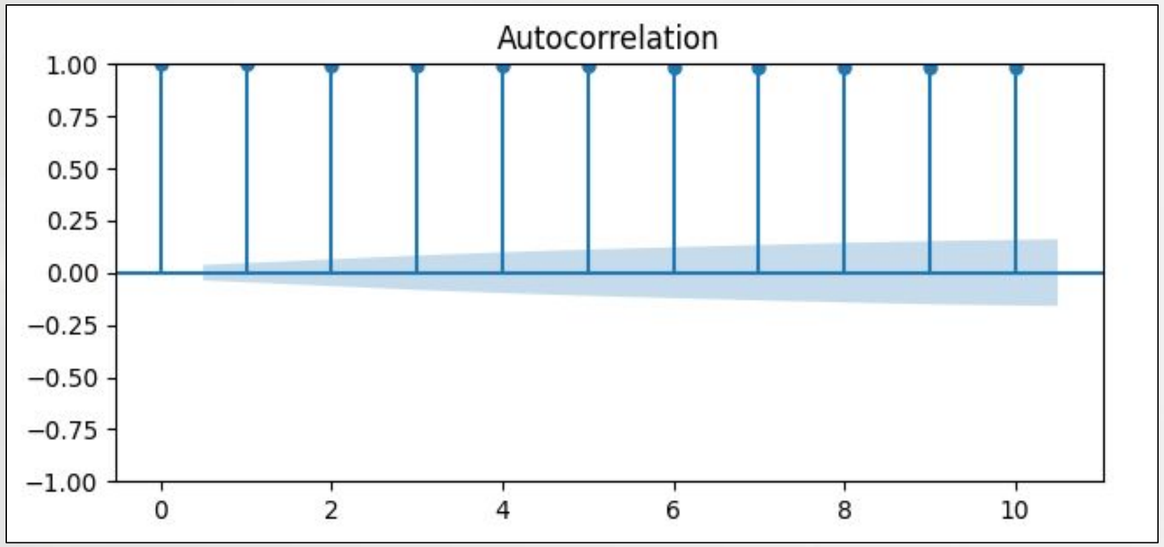
CORRELATION ANALYSIS on Gold Price & Macroeconomic Variables

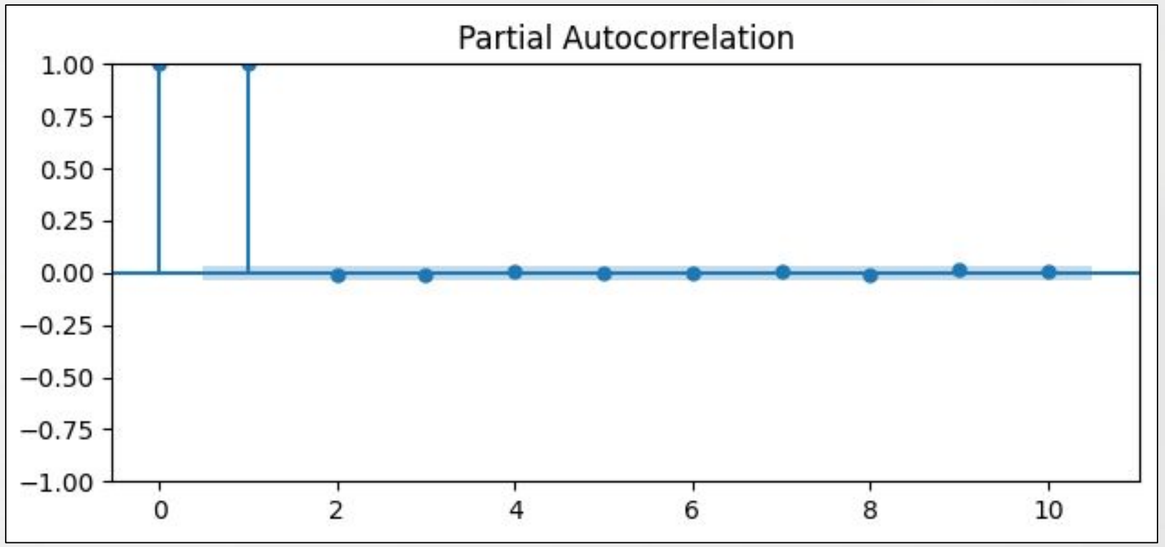


S&P 500 and Gold Prices are highly correlated

→ Develop & compare models using **S&P 500** vs **all features** as regressors

ACF & PACF PLOT





- **ACF:** Strong positive correlation at all lags indicates **non-stationarity** and a **persistent trend** in the data

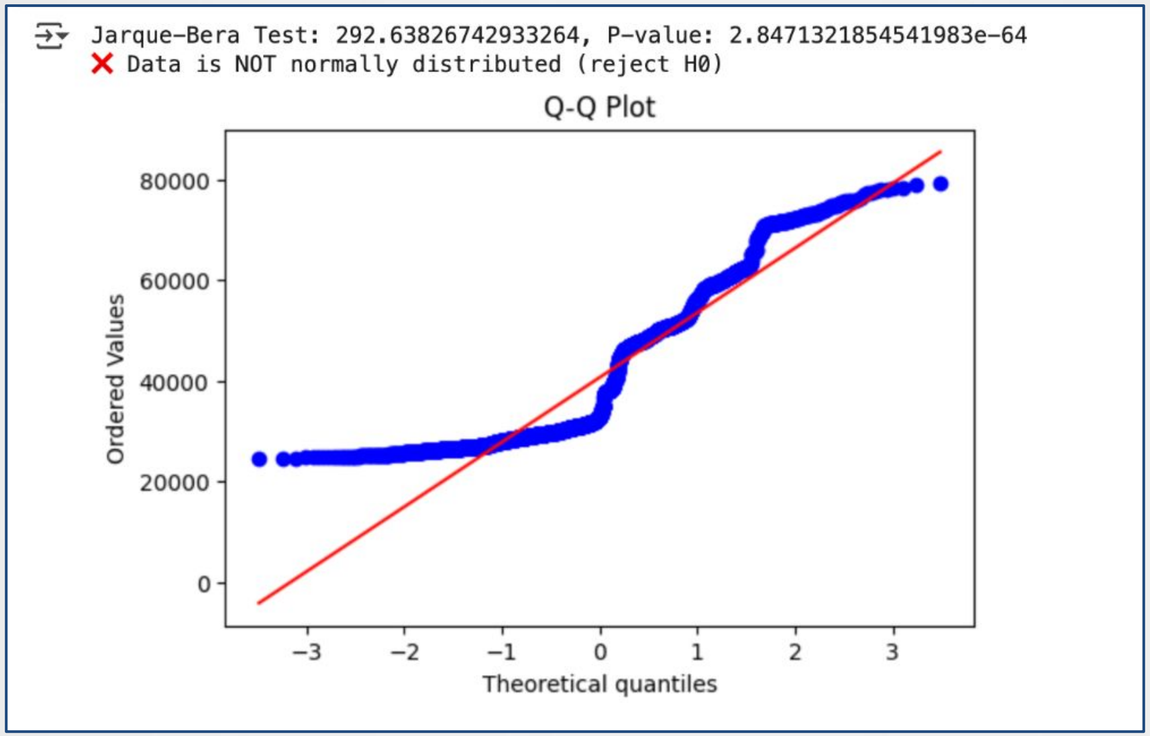
- **PACF:** Only the first lag has a significant direct impact, implying a **possible AR(1) process.**

STATIONARITY

ADF Statistic: 2.136
p-value: 0.999
Critical Values:
1%: -3.433
5%: -2.863
10%: -2.567
The time series is **non-stationary**.

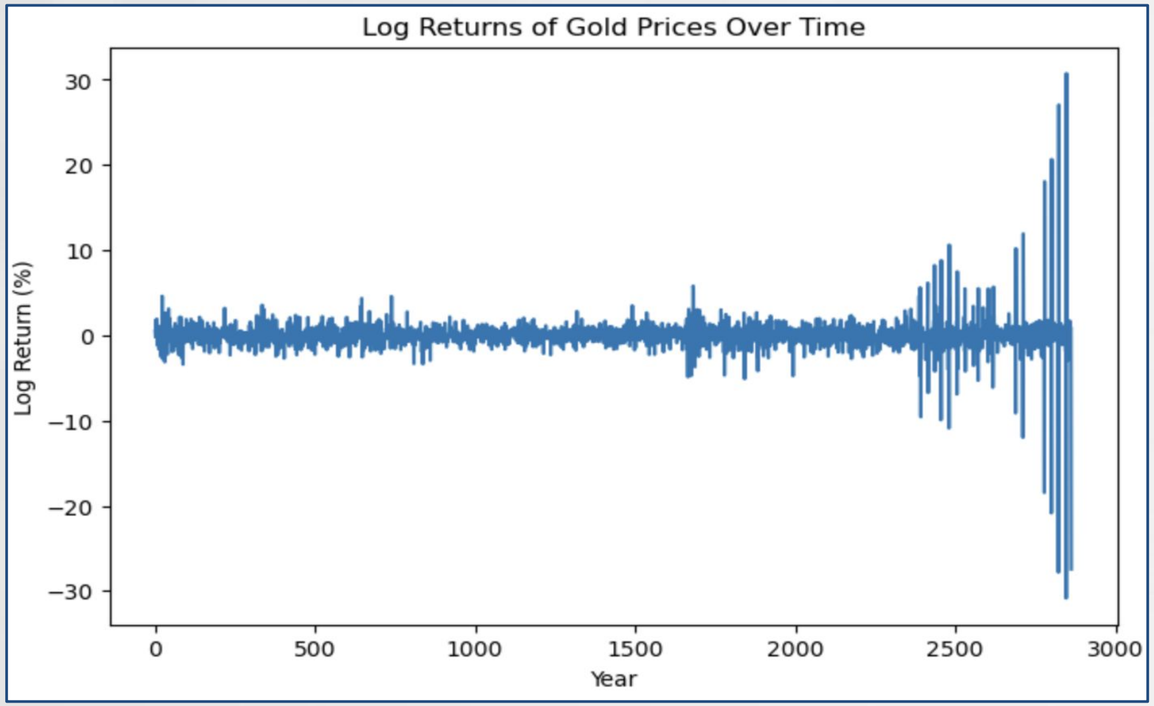
➢ The high ADF statistic and p-value (0.99) indicate that we fail to **reject the null hypothesis,** confirming that the time series is **non-stationary.** This aligns with the strong autocorrelation pattern in the ACF plot, suggesting the need for differencing to achieve stationarity before modeling.

NORMALITY

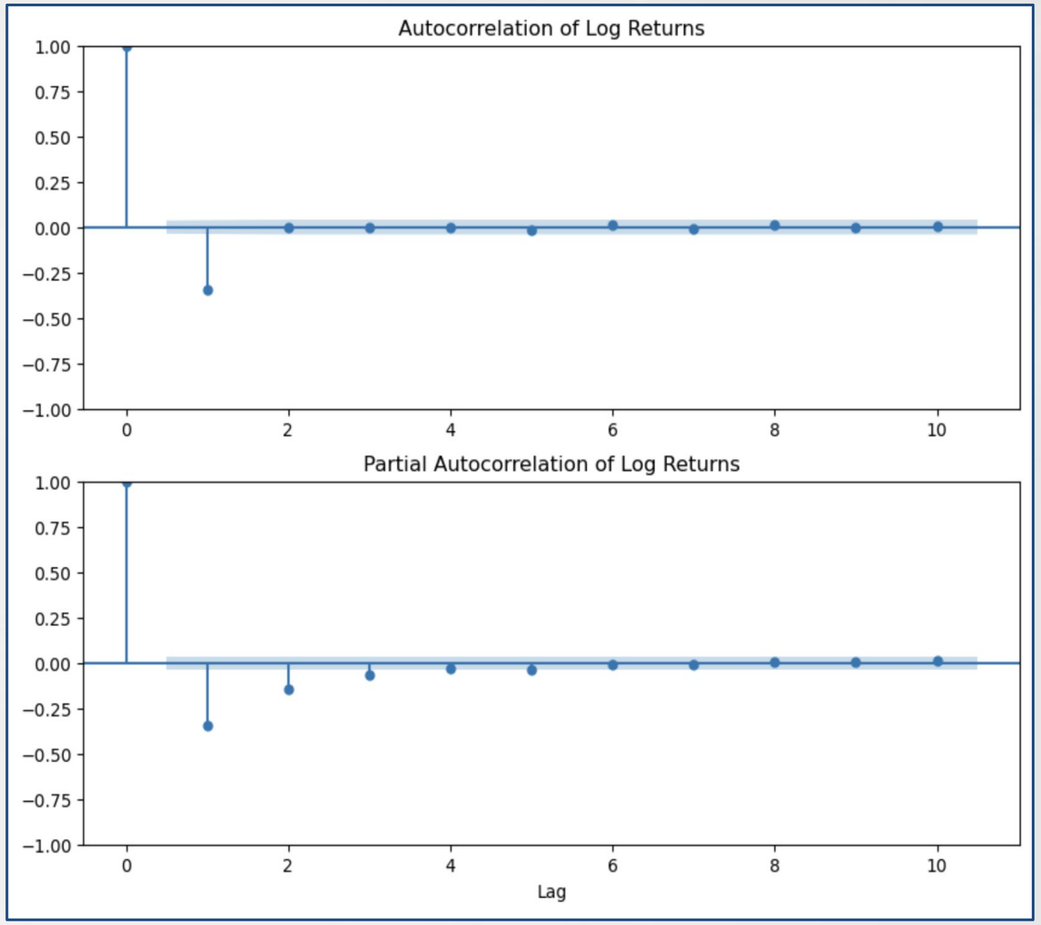


➢ The Q-Q plot shows that the data deviates significantly from the normal distribution, particularly at the tails, indicating heavy-tailed behavior and non-normality. This suggests that **transformations like log transformation might be needed** for modeling.

Volatility Analysis



Significant clustering of **volatility is not evident** over most of the period, except for some spikes toward the end.



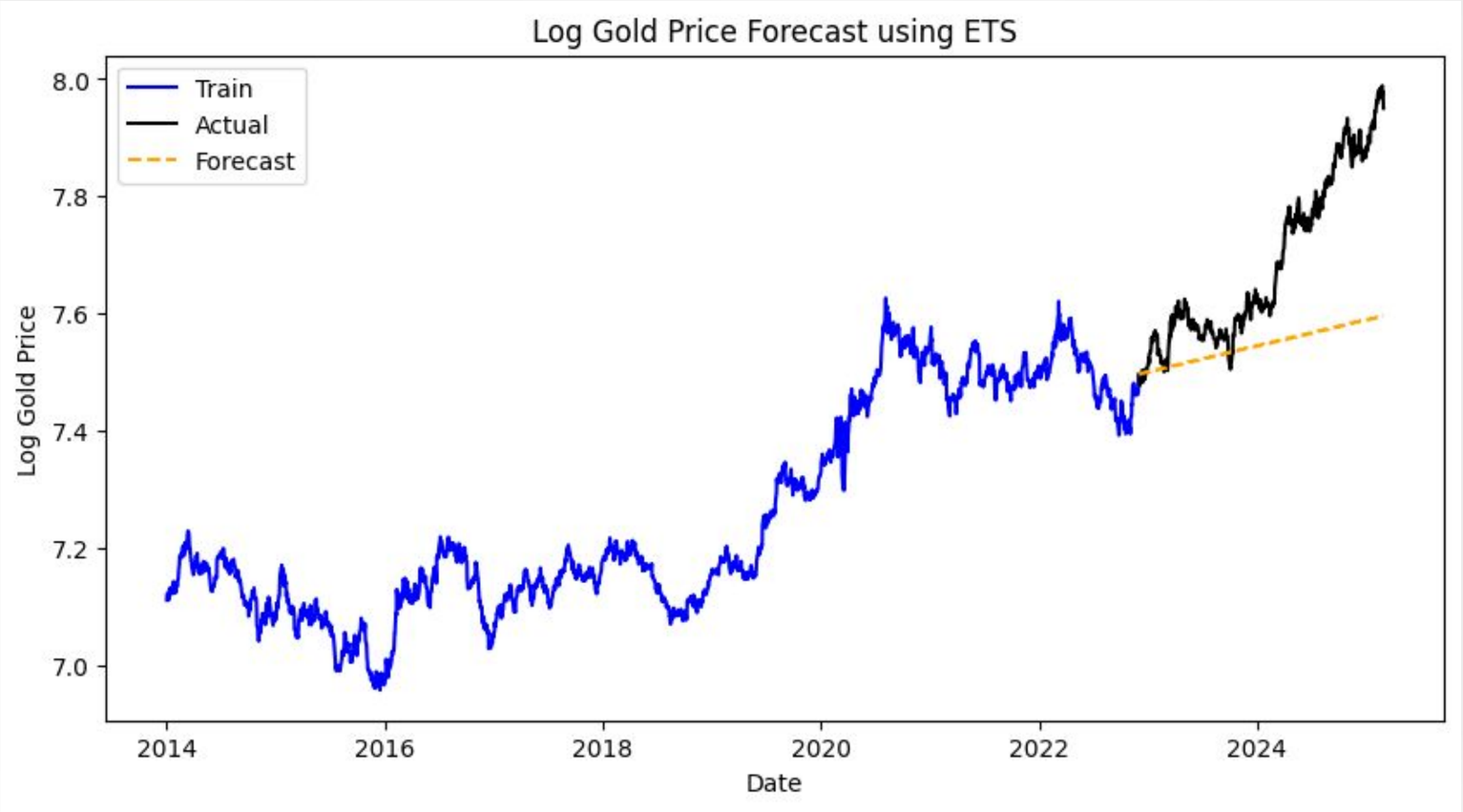
No strong dependence in log returns suggests that **past volatility does not significantly impact future volatility.**

➢ **Given the absence of persistent volatility clustering, ARCH or GARCH models are not necessary.**

MODEL DEVELOPMENT

PREDICTION USING GOLD PRICE ONLY (1/3)

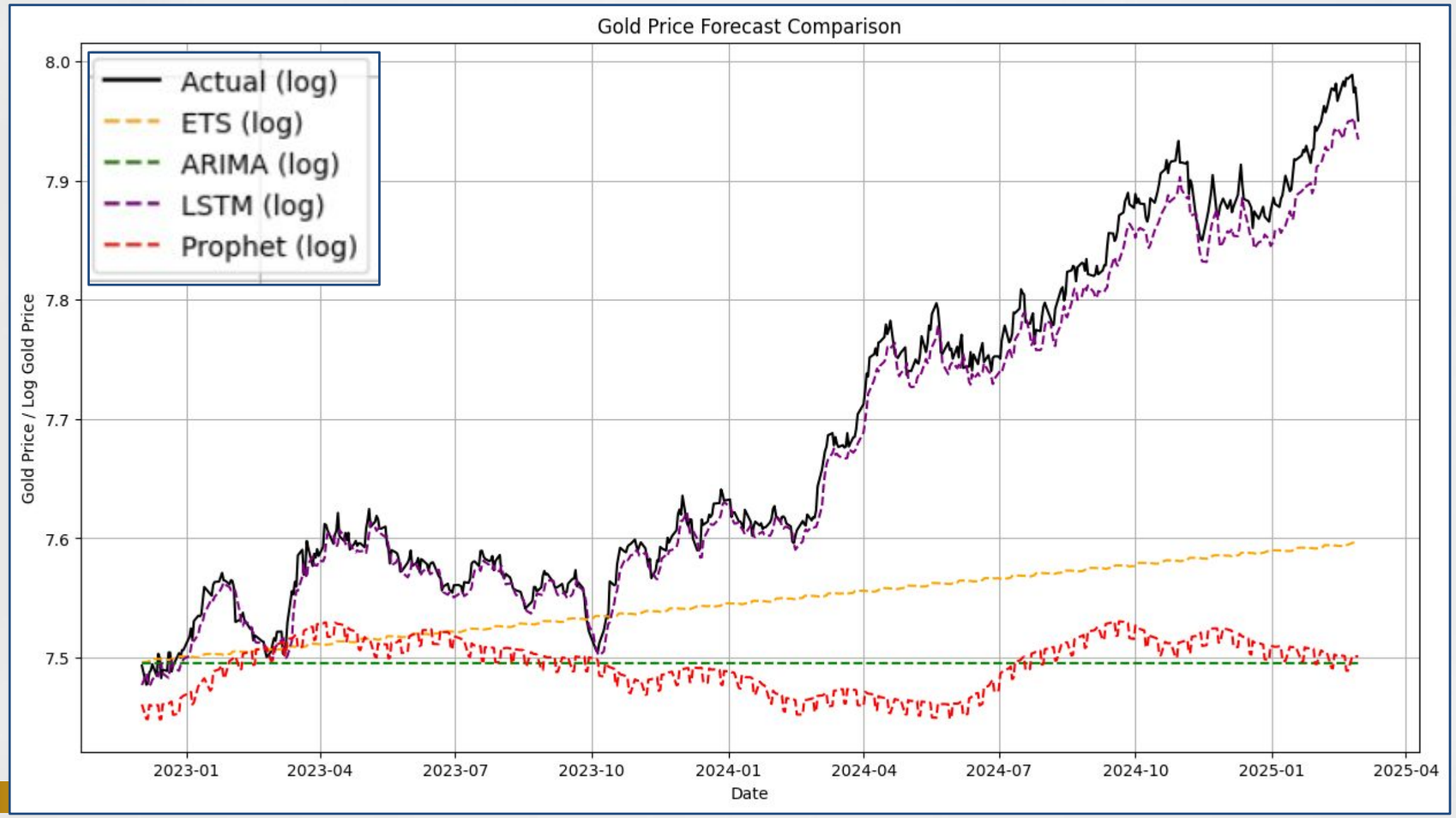
ETS MODEL AS A BASELINE



➢ **A baseline model using past gold price data only helps us identify the direction of the gold price.** Shows the upward trend of the gold price

PREDICTION USING GOLD PRICE ONLY (2/3)

FORECAST RESULTS FROM MULTIPLE MODELS



PREDICTION USING GOLD PRICE ONLY (3/3)

RESULT & VALIDATION

Model	Specification	AIC (Training Data)	RMSE (Test Data)
ETS	Log gold price	-20971.170	0.180
ARIMA (1,1,0)	log gold price	-14620.618	0.236
LSTM	gold price	100147	50.89
LSTM	log gold price	21533.777	0.0337
PROPHET	gold price	34013.460	548.546
PROPHET	log gold price	34025.647	0.236
NN (MLP)	gold price	-8470.135	0.096

➢ **LSTM** (log-transformed) performed best, while **Prophet** on raw prices had the worst RMSE

Is historical gold price alone enough for accurate forecasting?

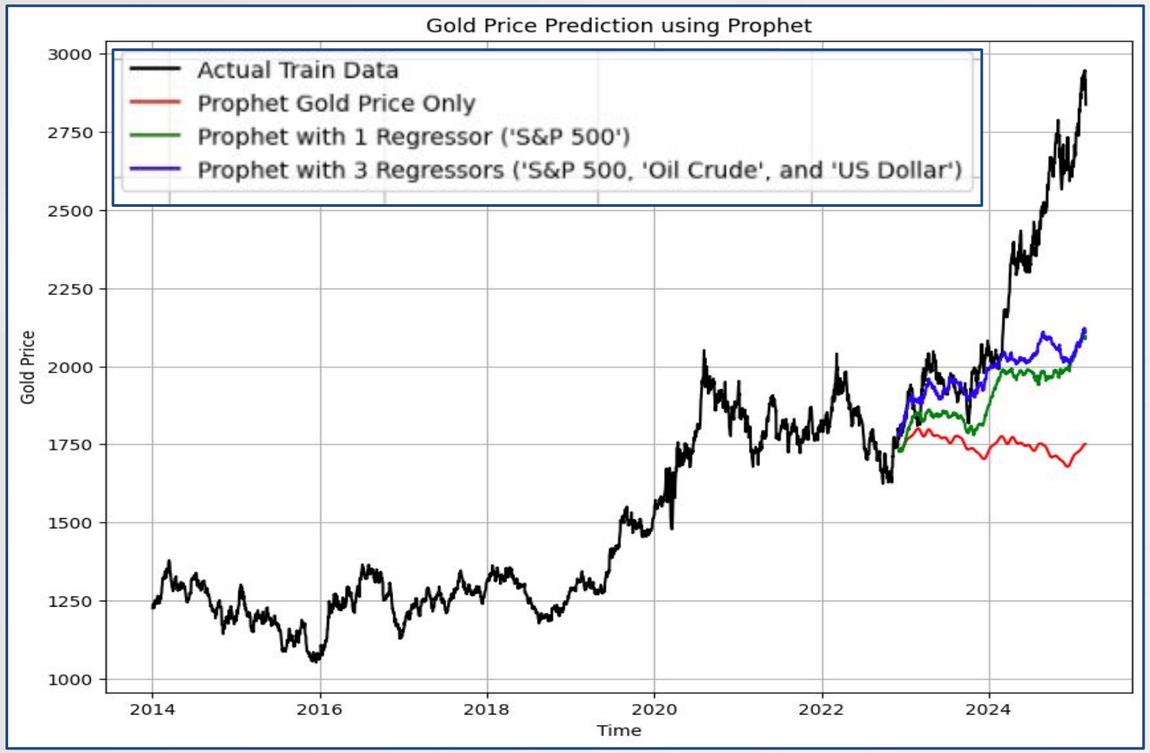
Do macroeconomic factors enhance predictive power?

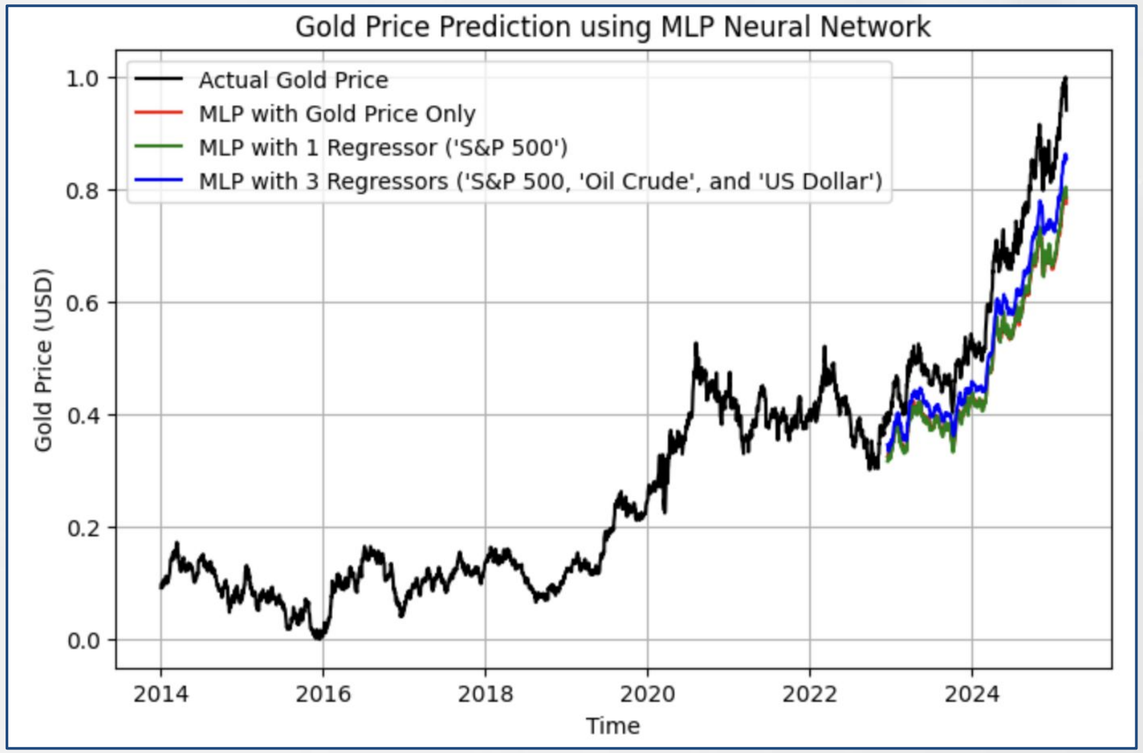
Train & Compare different models with:

- **One macroeconomic regressor (S&P 500)** - the most correlated feature
- **Three macroeconomic regressors (S&P 500, Oil Crude, and US Dollar index)**

IMPACT OF MACROECONOMIC REGRESSORS (1/3)

PROPHET & NEURAL NETWORK (MLP) MODEL

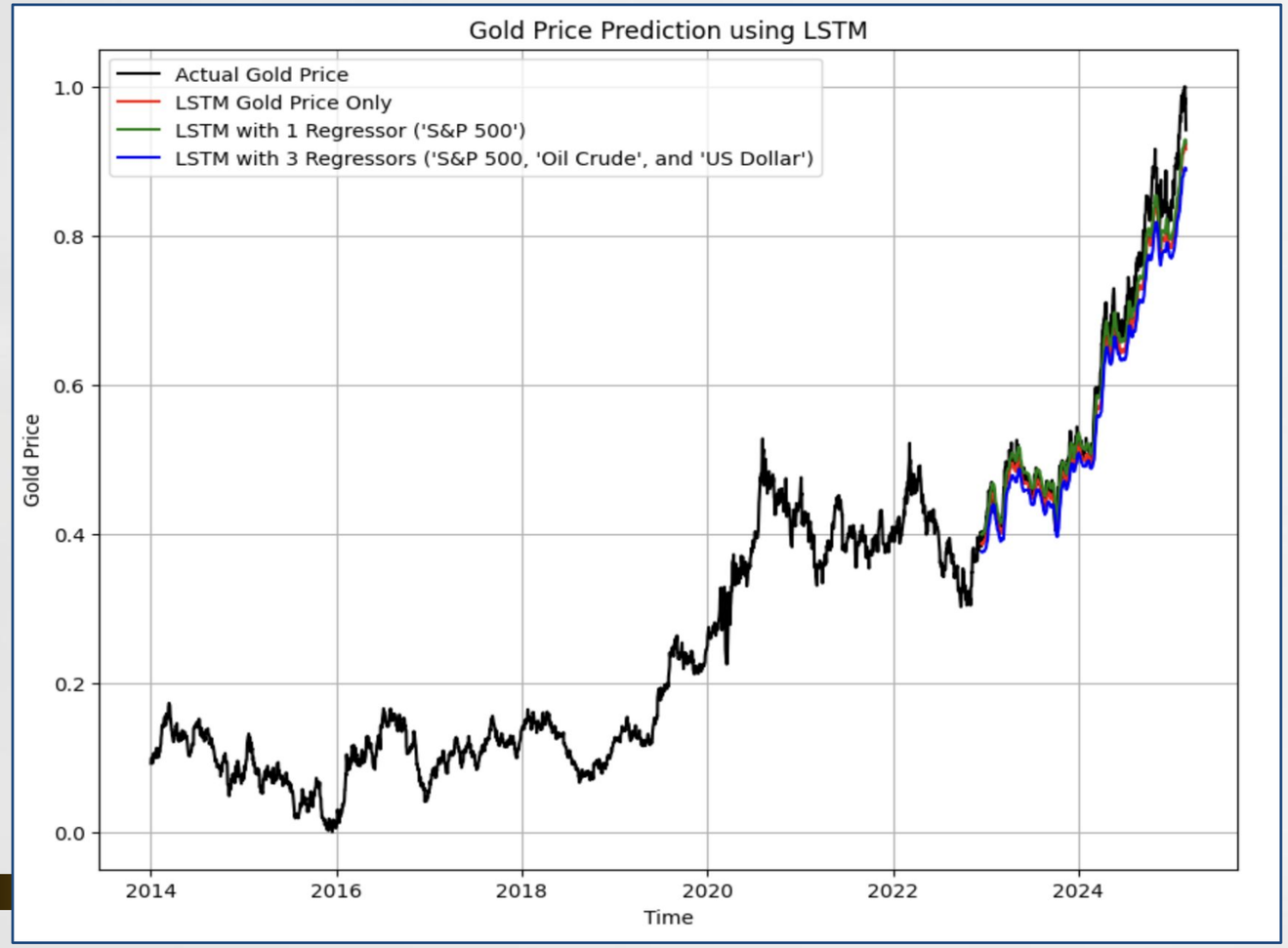




- Incorporating economic indicators (S&P 500, Oil Crude, and US Dollar) **enhances predictive performance** of both models
- **Neural Network outperforms Prophet** in capturing the trend and noise of the gold price

IMPACT OF MACROECONOMIC REGRESSORS (2/3)

LSTM MODEL



- Incorporating **1 best correlated economic indicator (S&P 500)** had the best performance
- Using all three regressors also improved accuracy
- **LSTM best captures trends and fluctuations** in gold prices

→ **Our Best Model!**

IMPACT OF MACROECONOMIC REGRESSORS (3/3)

RESULT & VALIDATION

Model	Specification	AIC (Train Data)	RMSE (Test Data)
LSTM	log gold price + S&P 500	12915.578	0.028
	log gold price + S&P 500 + Crude Oil + US Dollar Index	25915.495	0.049
PROPHET	log gold price + S&P 500	34026.338	493.064
	log gold price + S&P 500 + Crude Oil + US Dollar Index	34019.344	444.249
NN (MLP)	log gold price + S&P 500	-7066.348	0.079
	log gold price + S&P 500 + Crude Oil + US Dollar Index	-5679.741	0.101

- **Neural Network** best fits the training data, but might be slightly overfitting, capturing too much noise from the training set.
- **LSTM** has higher AIC due to its complexity, but it has the best generalization performance.

CONCLUSION

Do macroeconomic factors enhance predictive power?

Model	Specification	AIC (Training Data)	RMSE (Test Data)
LSTM	gold price	22915.495	0.0456
	gold price + S&P 500	12915.578	0.028
	gold price + S&P 500 + Crude Oil + US Dollar Index	25915.495	0.049
NN (MLP)	gold price	-8470.135	0.124
	gold price + S&P 500	-7066.348	0.110
	gold price + S&P 500 + Crude Oil + US Dollar Index	-5679.741	0.091

→ YES!

CONCLUSION

Our analysis demonstrates that incorporating **macroeconomic factors** significantly enhances the predictive power of all gold price forecasting models. Among all models, **LSTM with the S&P 500 as a regressor achieves the best performance**, making it the most reliable choice for predicting future gold prices. With an **RMSE of 0.028**, our model demonstrates strong accuracy and generalizability. By leveraging our model, stakeholders can make data-driven investment decisions, and gain a competitive edge in financial planning. From our model, we predict today's gold price as 2907.86 (actual value is 2900.04).

FUTURE IMPROVEMENTS

- Explore **additional macroeconomic indicators** (e.g., interest rates, inflation, or geopolitical events)
- Incorporate **external events** and **sentiment analysis**, such as news and financial reports, to capture sudden market fluctuations and its impact on gold prices.
- Include **real-time market data** to enhance dynamic prediction capabilities.

THANK YOU

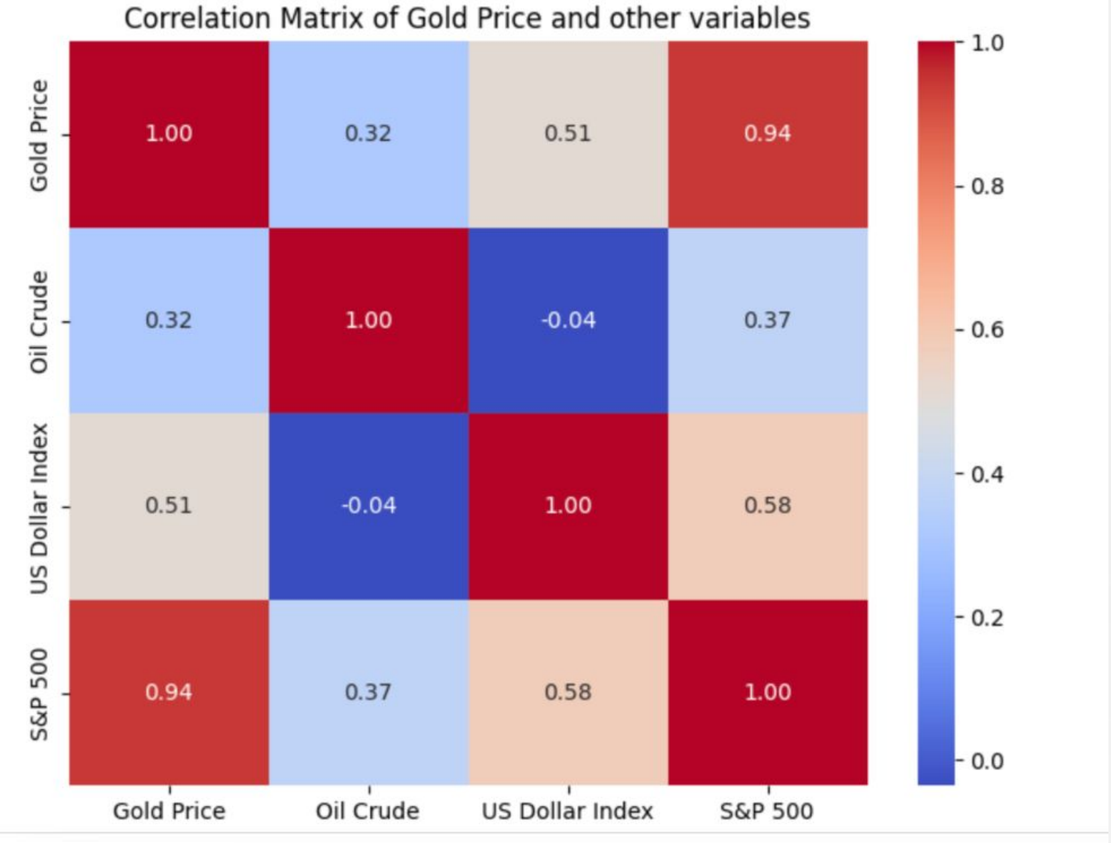
Amy Kim, Cicily Mathew, Olivia Rumere

How accurate we predict gold price this month!

Date	Actual (Yfinance)	LSTM
3 March 2025	2890.20	2864.43
4 March 2025	2909.60	2871.97
5 March 2025	2915.30	2879.29
6 March 2025	2916.60	2885.89
7 March 2025	2804.70	2891.48
10 March 2025	2891.00	2904.16
11 March 2025	2900.40	2907.86

Cross-Correlation

```
Cross-correlation between Gold Price and Oil Crude: 0.3188918421366752
Cross-correlation between Gold Price and US Dollar Index: 0.5069039619624995
Cross-correlation between Gold Price and S&P 500: 0.9427601964061753
```



Short-run Cointegration Test Results:

	Variable	ADF Statistic	p-value	Short-run Cointegrated
0	Oil Crude	0.550591	0.986332	False
1	US Dollar Index	-0.883417	0.793384	False
2	S&P 500	-3.009054	0.034058	True

Model Selection

Model	Type	Data Requirements	Strengths	Weaknesses
ETS (Exponential Smoothing)	Classical Time Series	Works well with trend & seasonality	Simple, interpretable, captures trend/seasonality	Cannot handle external variables, poor for nonlinear patterns
ARIMA (Auto-Regressive Integrated Moving Average)	Classical Time Series	Requires **stationarity** (differencing needed)	Strong for short-term predictions, interpretable coefficients	Assumes linear relationships, not great for long-term forecasting
LSTM (Long Short-Term Memory - Deep Learning)	Neural Network	Large datasets, needs scaling	Captures long-term dependencies, can handle complex patterns	Computationally expensive, requires tuning
Prophet (Facebook's Forecasting Model)	Additive Model	Works well with missing data & seasonality	Easy to use, handles seasonality well	Struggled with gold price trends, not great for irregular trends
MLP (Multi-Layer Perceptron)	Neural Network	Works better with transformed data (e.g., log)	Captures hidden relationships in data	Computationally expensive, requires tuning

GOLD PRICE 30 DAYS

GOLD PRICE ONLY DATA

	Date	Predicted Gold Price
0	2025-03-01	2858.115479
1	2025-03-02	2858.228516
2	2025-03-03	2864.432373
3	2025-03-04	2871.967041
4	2025-03-05	2879.293213
5	2025-03-06	2885.893555
6	2025-03-07	2891.477783
7	2025-03-08	2896.344727
8	2025-03-09	2900.427490
9	2025-03-10	2904.159668
10	2025-03-11	2907.857666
11	2025-03-12	2911.362549
12	2025-03-13	2914.832520
13	2025-03-14	2918.255127
14	2025-03-15	2921.617432
15	2025-03-16	2924.905762
16	2025-03-17	2928.108398
17	2025-03-18	2931.211182
18	2025-03-19	2934.214844
19	2025-03-20	2937.117920
20	2025-03-21	2939.924561
21	2025-03-22	2942.634277
22	2025-03-23	2945.256592
23	2025-03-24	2947.788330
24	2025-03-25	2950.238037
25	2025-03-26	2952.606689
26	2025-03-27	2954.898193
27	2025-03-28	2957.112793
28	2025-03-29	2959.255127
29	2025-03-30	2961.323242