Assignment 1

Subject: Data Mining

Course Code: CSI0803

Date: 5th January, 2024

*Problem Statement*

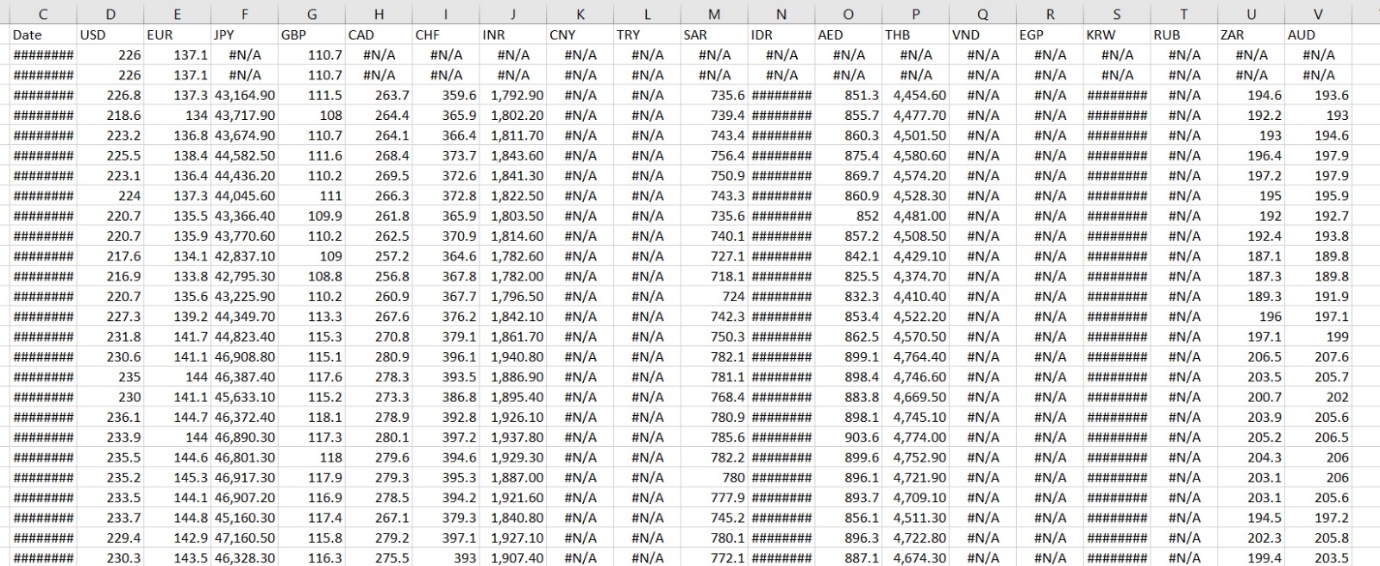
The current volatility of the gold market poses a challenge for investors. Can we build a time series forecasting model that accurately predicts the gold price in USD for the next one year using historical data, with an average error below 10% to enable profitable investment strategies?

*Approach*

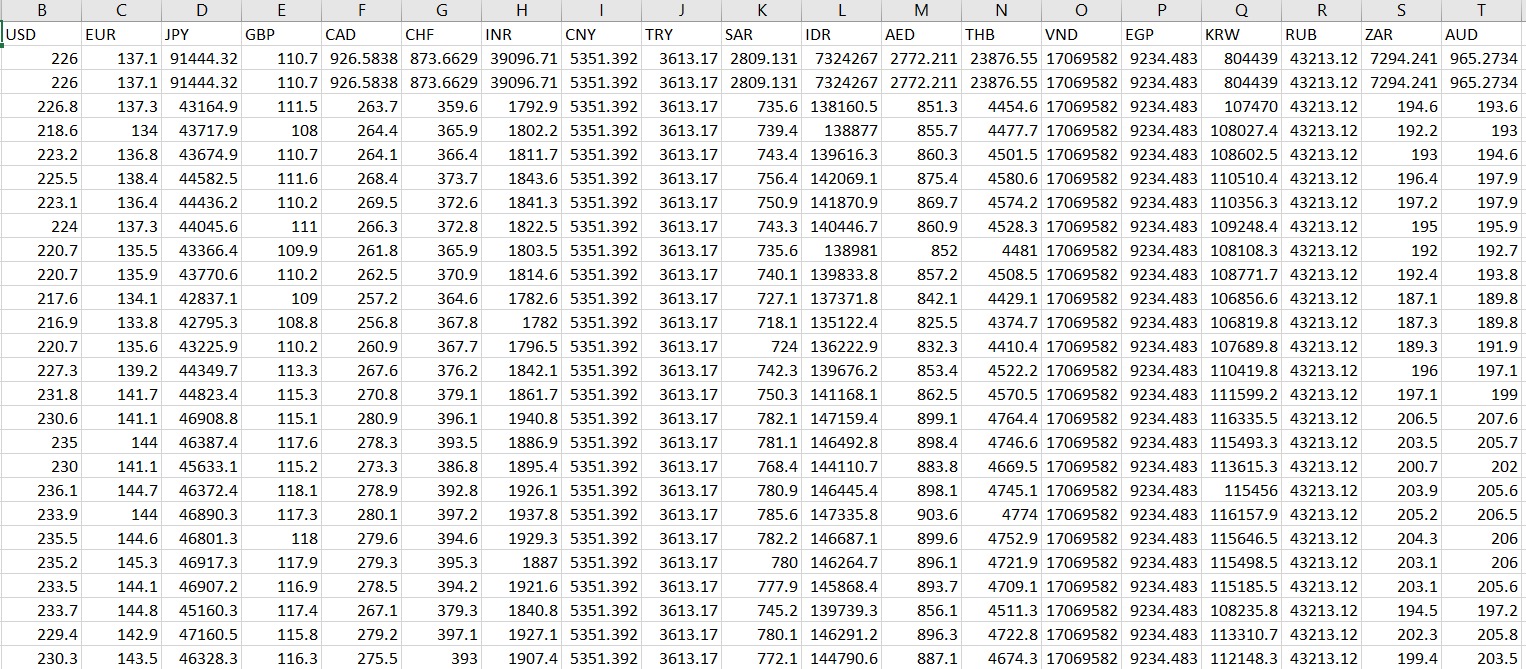
* Preprocessing

Firstly, we preprocess the data by filling the missing values by taking the mean of the entire column.

Original Dataset:



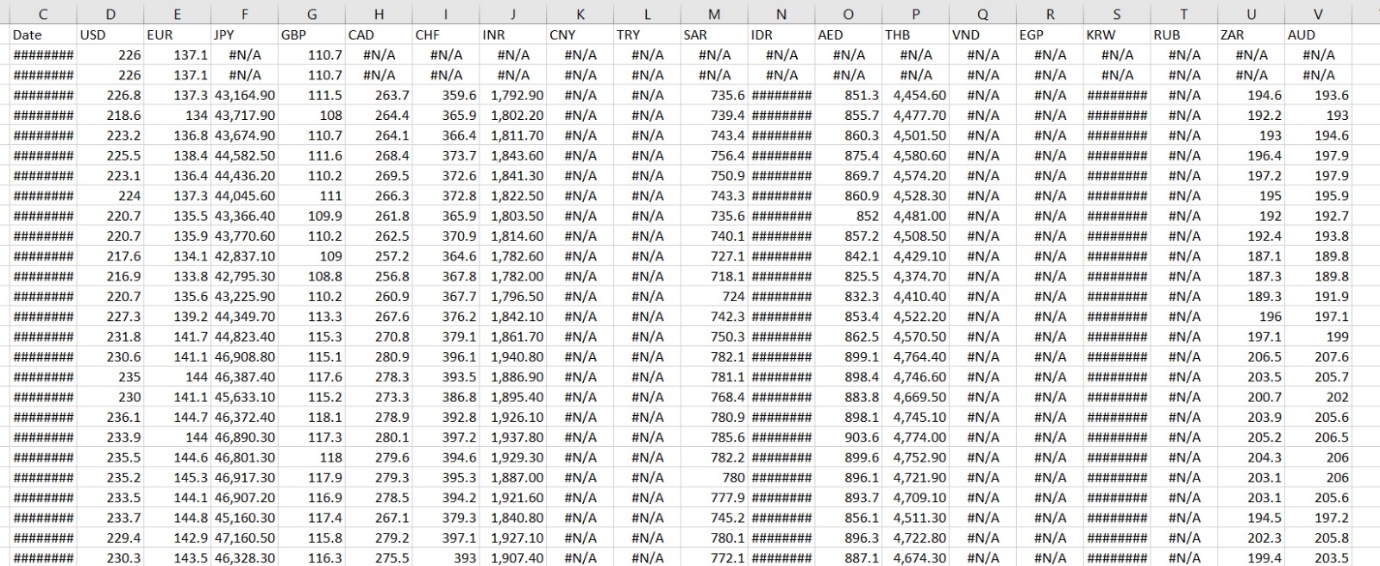
Pre-processed dataset:



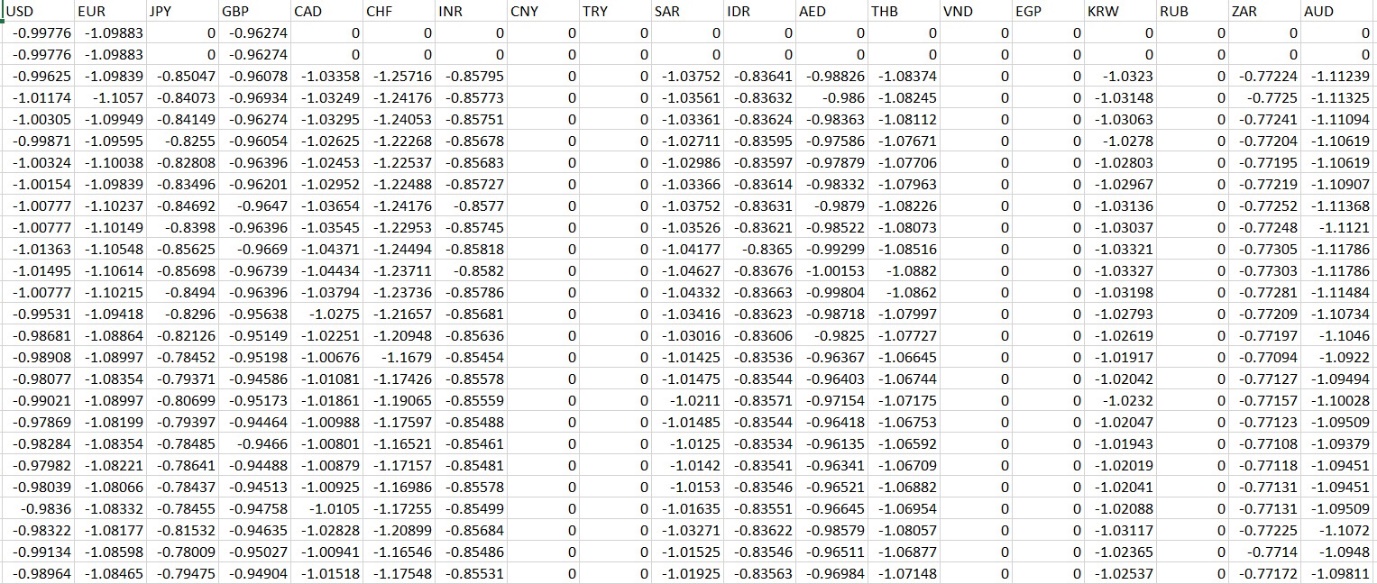
* Normalization

Now, we normalize the entire dataset using Z- Score Normalization.

Original Dataset:



Normalized dataset:



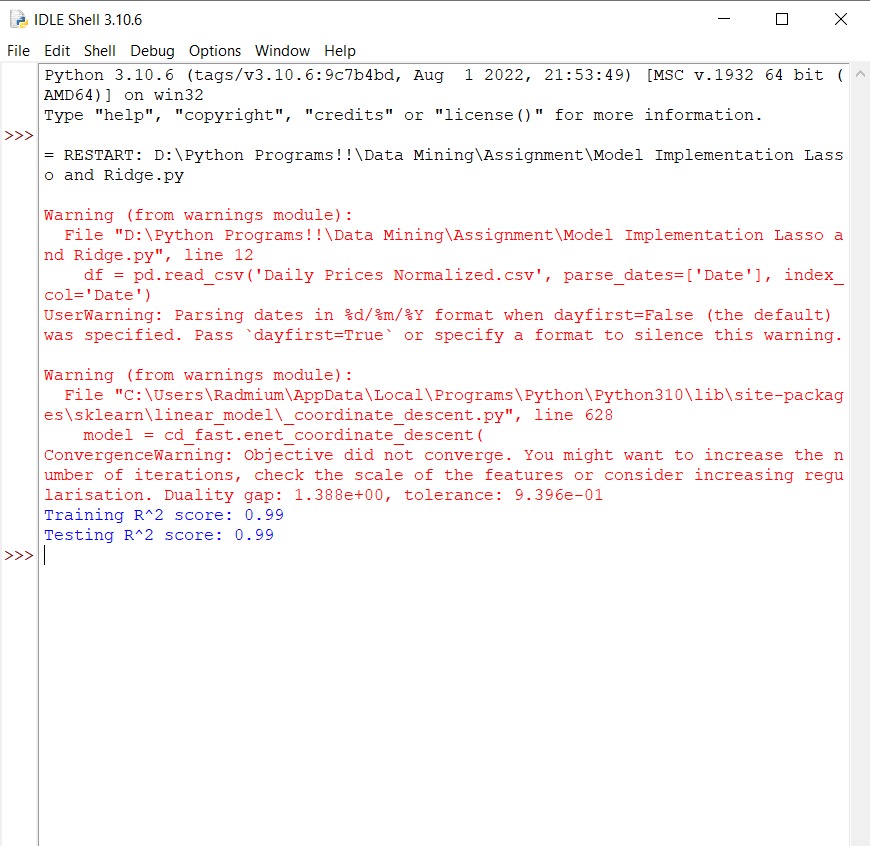
* Machine Learning Model

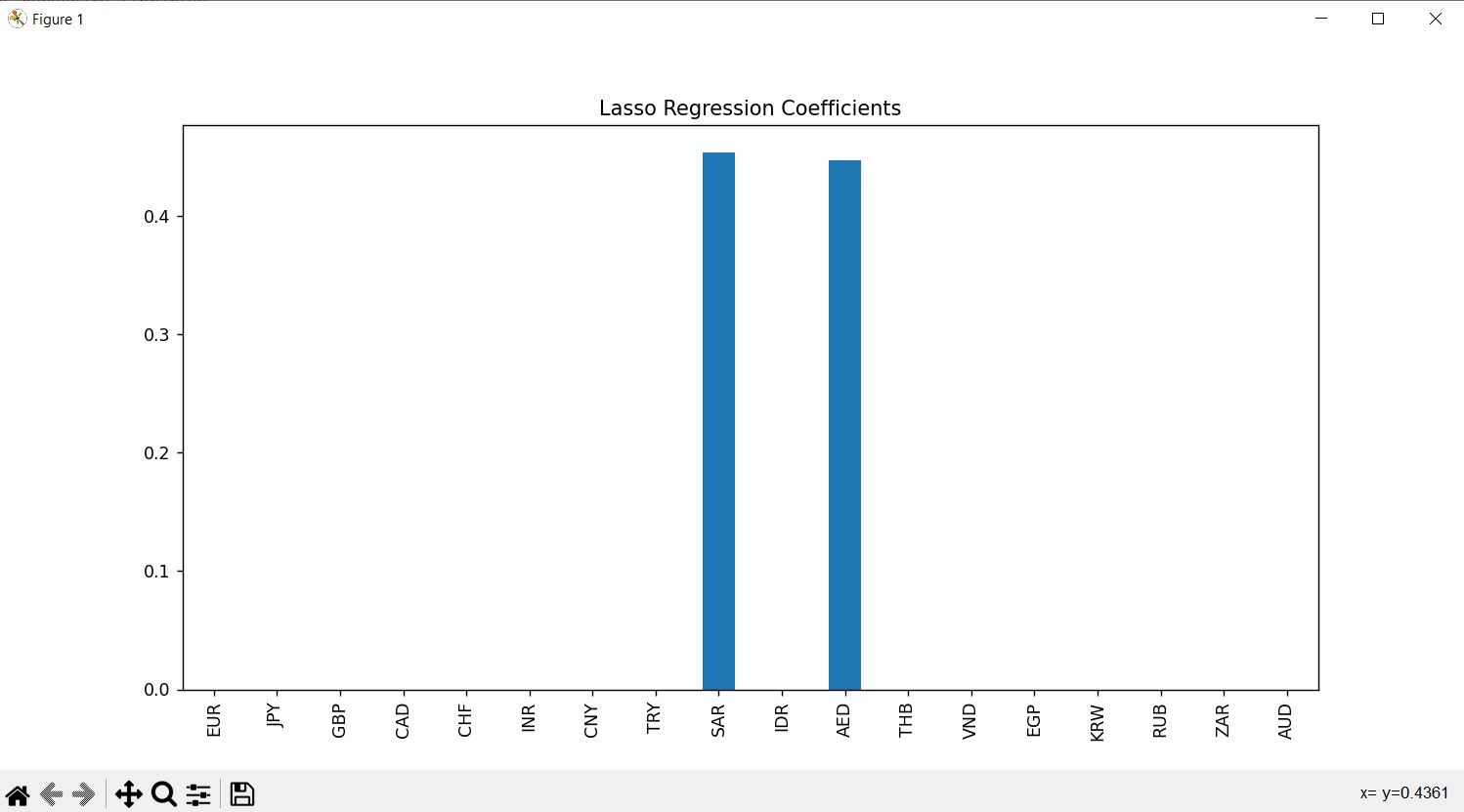
We have used the Polynomial Regression Model for the time series analysis of the Gold Prices for one year.

We tried other approaches such as:

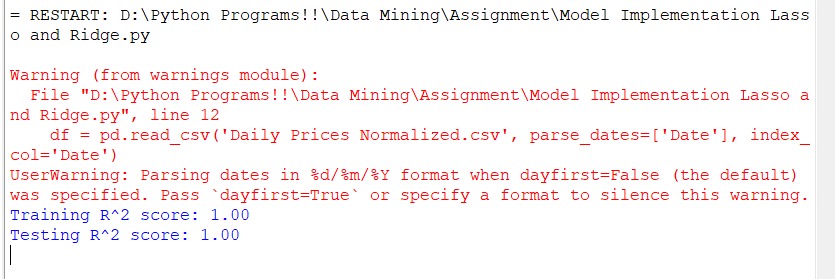
Lasso Regression and Ridge Regression, but in those cases due to the problem of overfitting the accuracy of model significantly reduced and for better results we would have to Regularize the dataset.

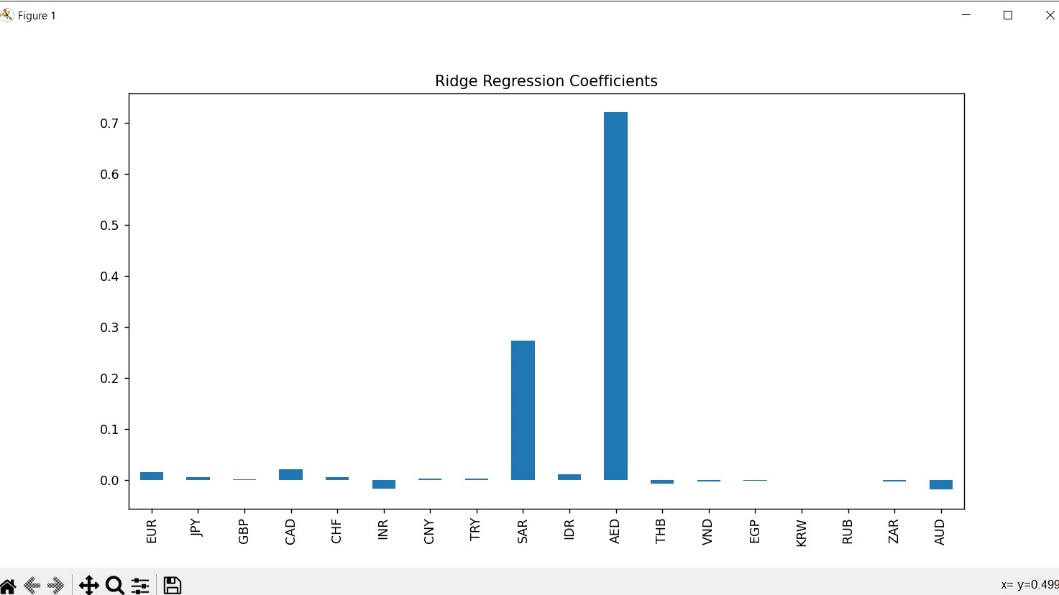
Lasso Regression





Ridge Regression



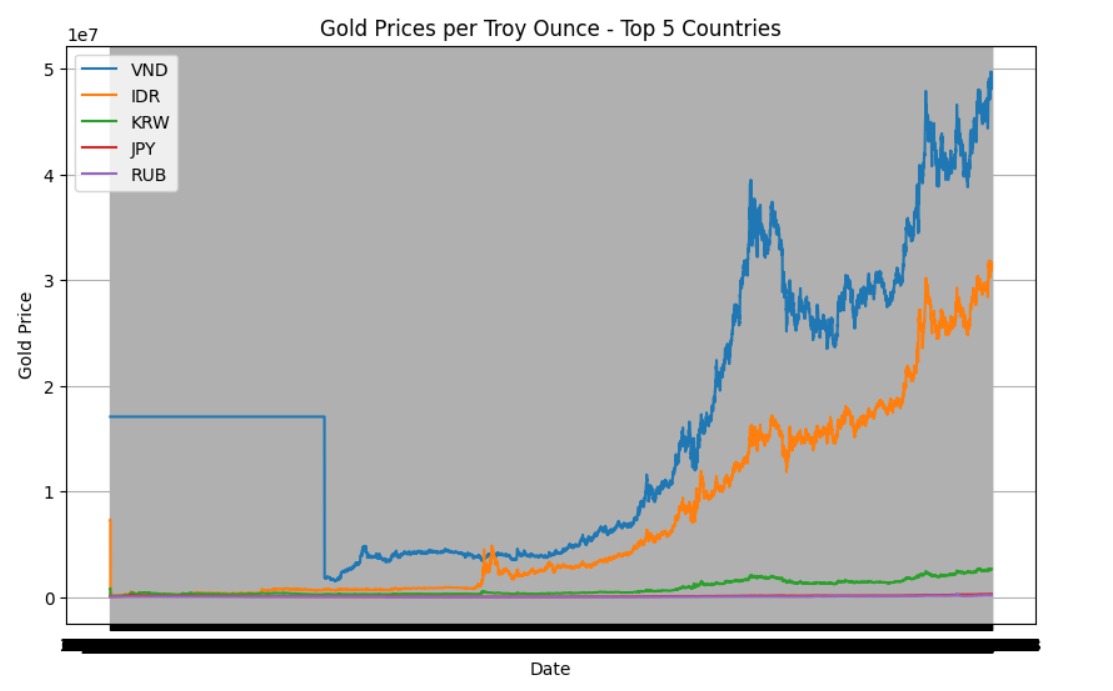


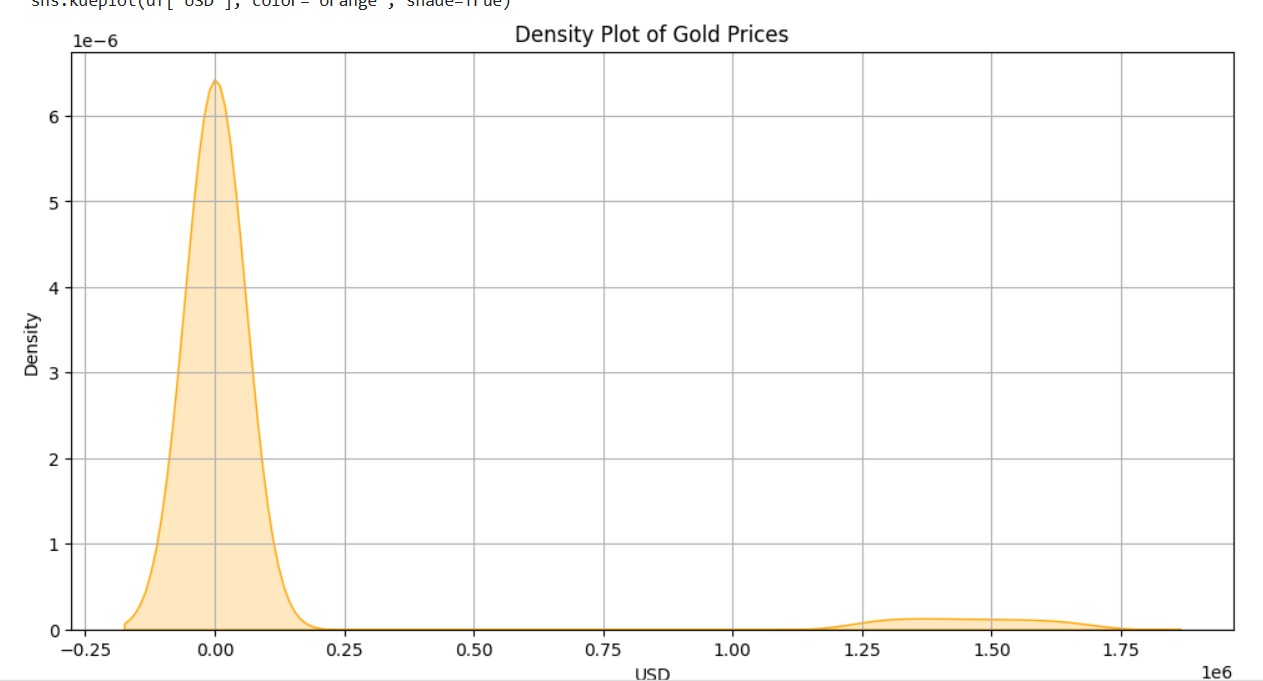
* Accuracy and Results

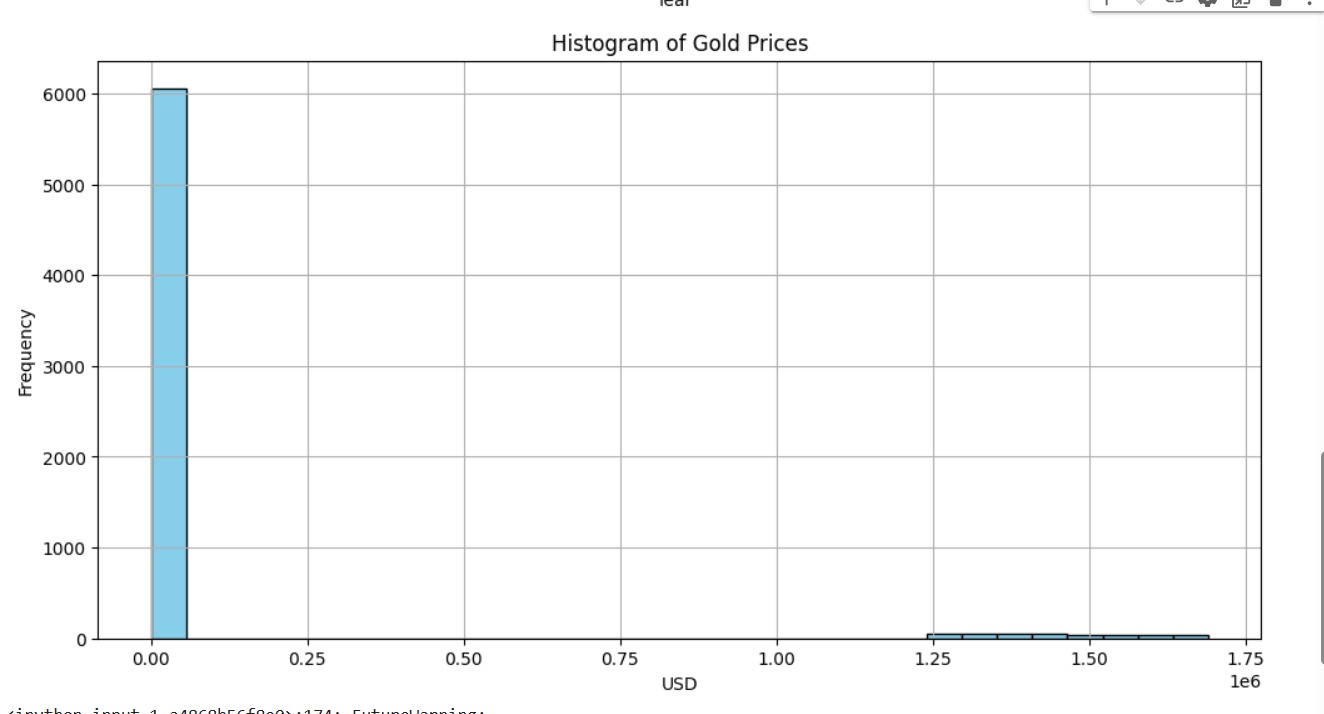
In the Polynomial Regression model, the best degree is calculated which is !! in our case. We have calculated the accuracy by using r2 score, the final accuracy of our model is 91.7%.

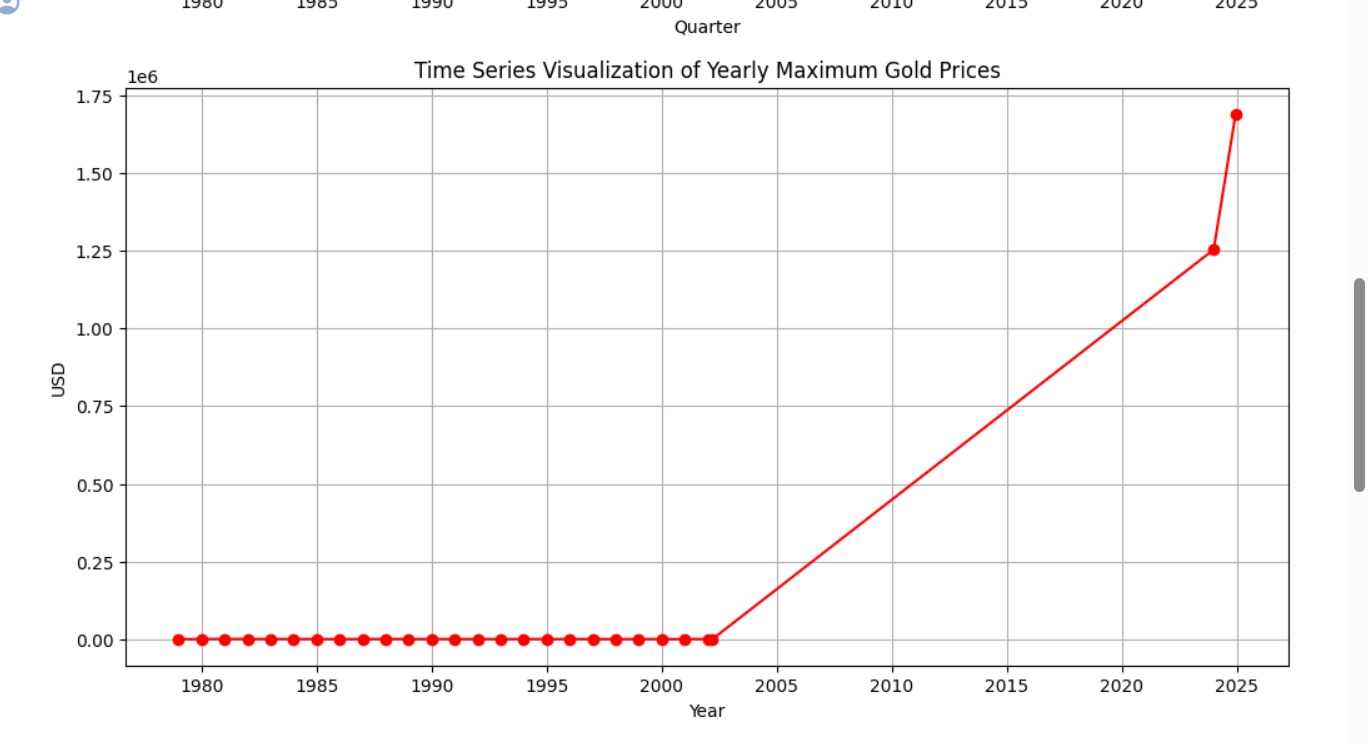
* Data Visualization

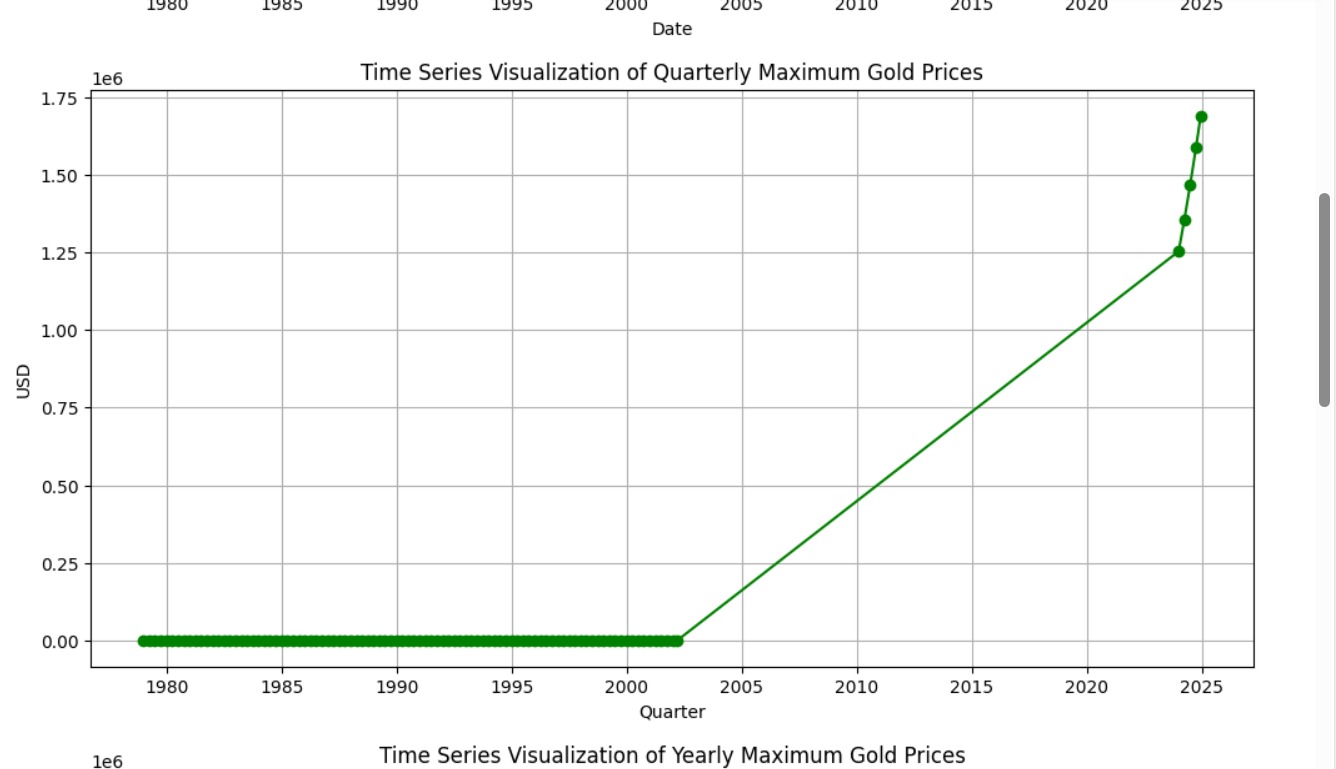
We have plotted various graphs to visualize the obtained results.

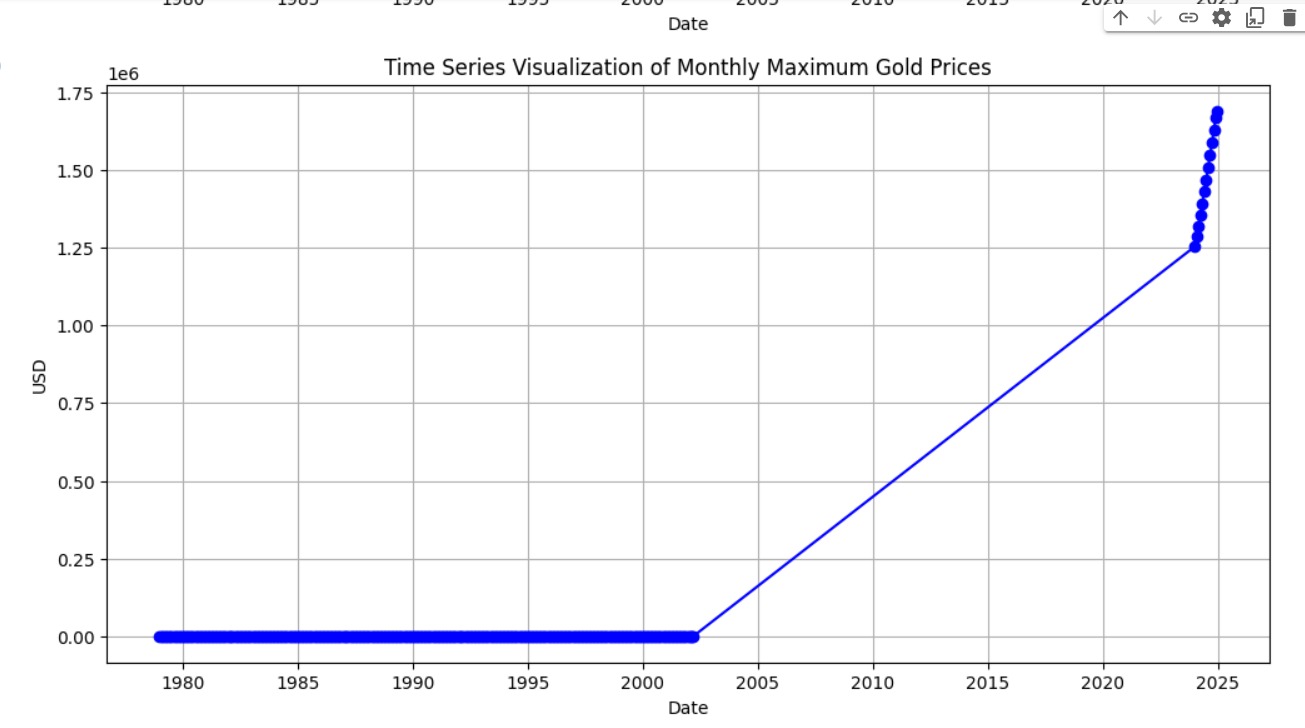


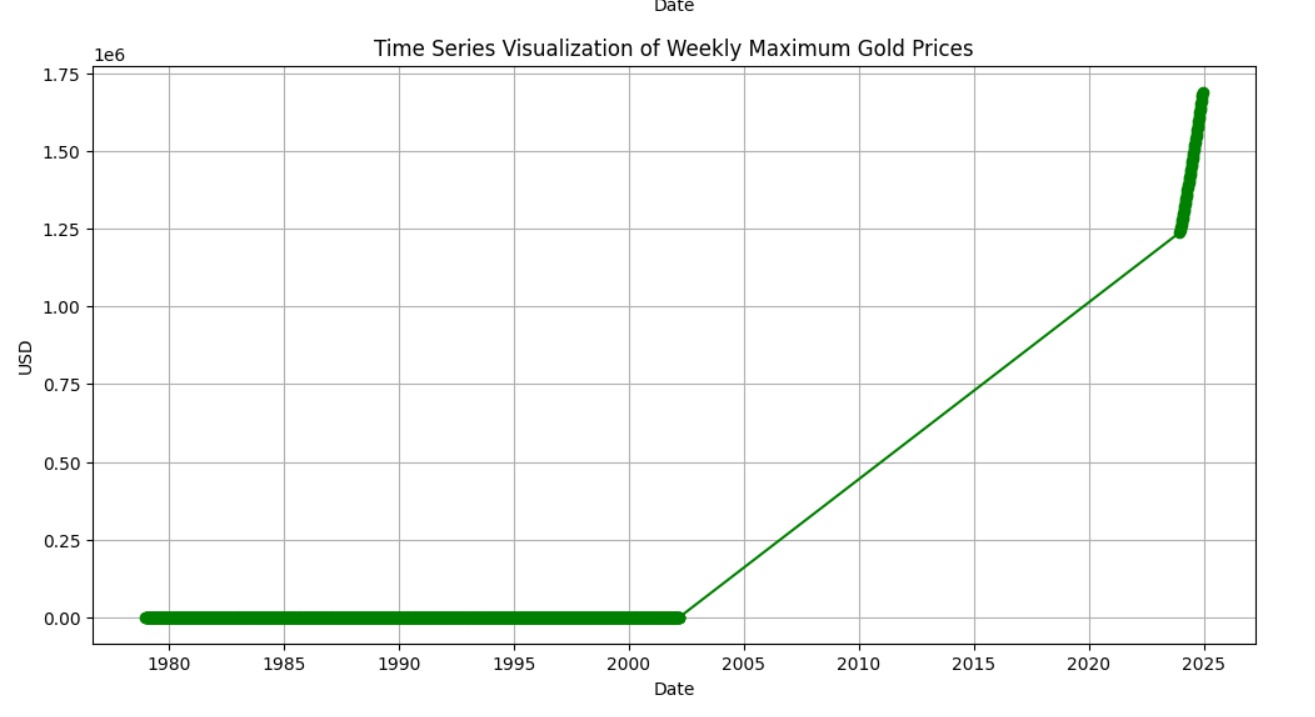


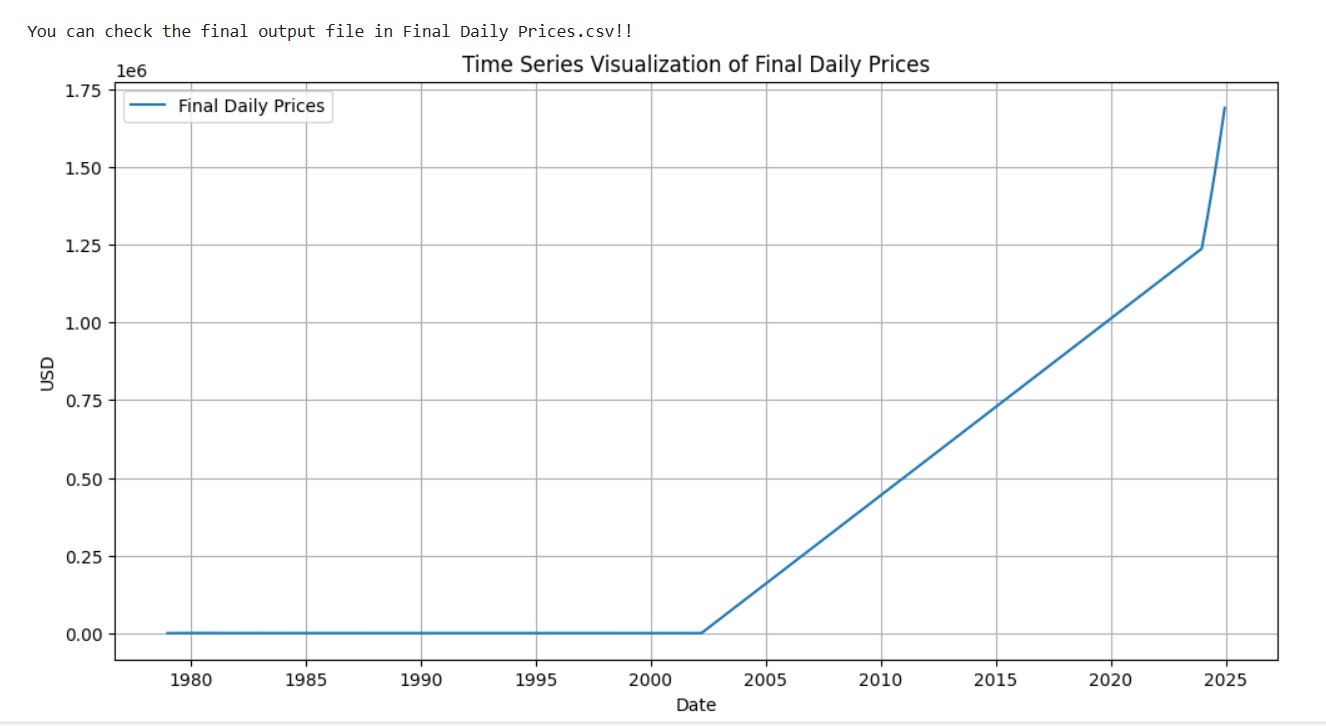












* Conclusion

This one-year time series forecasted data of gold Prices which has an error rate less than 10% will be very beneficial for the investors or anyone who is interested in buying or selling gold.

* Team members

1. Sanchit Kumar (21bcm058)
2. Tanishka Munot (21bcm062)
3. Amaan Haamid (21bcm077)