Time Series Analysis Gold Prices Prediction

Hindi Vidya Prachar Samiti's

R. J. COLLEGE

of Arts, Science & Commerce

(Autonomous College)

NAAC Re-Accredited 'A' Grade (CGPA 3.50)

DEPARTMENT OF STATISTICS

GROUP PROJECT

(2021 – 2022)

CLASS: M.Sc. - I **GROUP NUMBER: 7**

TITLE: Time Series Analysis

Project Group Members		
UID	Name	Email ID
05	Asma Shaikh	asmarjcollege@gmail.com
22	Amruta Bandopadhyay	amrutabandopadhyay20@gmail.com
28	Nikita Rahi	nikitarahi30@gmail.com
30	Shubham Upadhyay	shubhamupadhyayrjcollege@gmail.com
32	Pallavi Sawant	pallavisawant990@gmail.com
34	Pradnya More	pradnyarjcollege@gmail.com
52	Vedant Bhalekar	vedantrjcollege@gmail.com

PROF. MAYUR MORE

PROFESSOR IN-CHARGE

PROF. CHHAYA PINGE

HEAD OF STATISTICS DEPARTMENT

I N D E X

Dataset information in detail

We used the secondary data from Kaggle : https://www.kaggle.com/code/dewashyadubey/trends/data

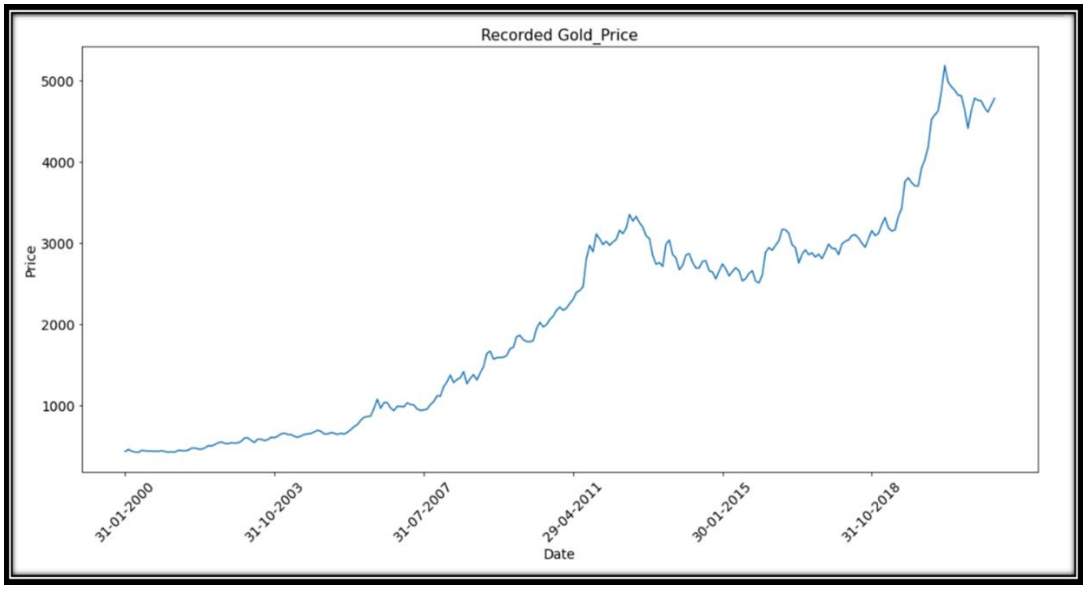
Monthly gold data mentioned as the last date of the month.

Price in Rupees per 10 gram.

There are total 263 observations.

From year 2000 to 2021.

Data Visualization



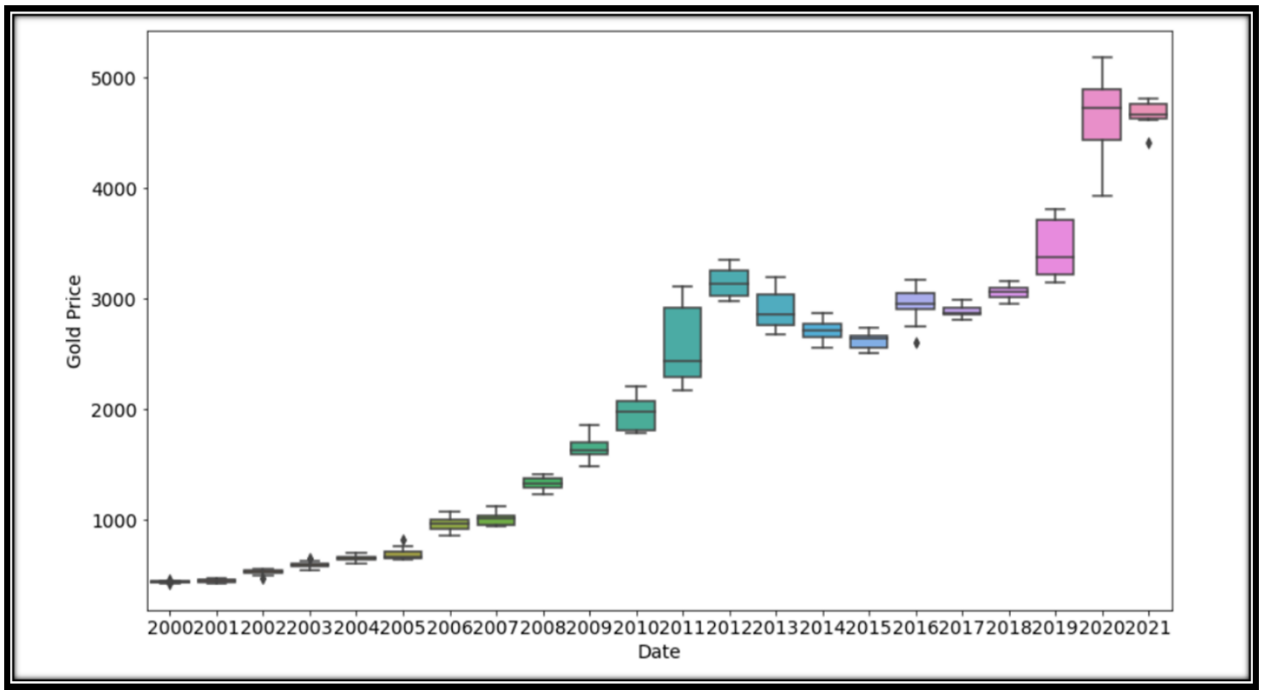
Objective : To predict future gold prices using Time Series

Methodology

- **Data Visualization**
- **Train and Test Data**
- **Arima**
- **Exponential smoothing**
- **MAPE values**

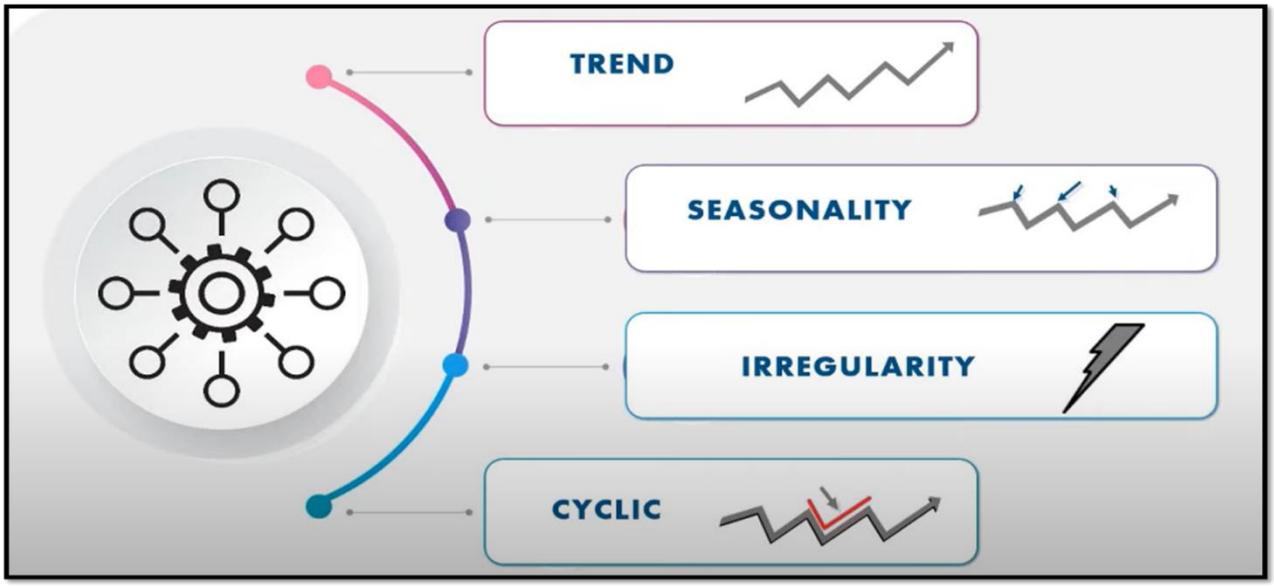
Analysis

> Yearly "Box-Plot" for Gold Price dataset

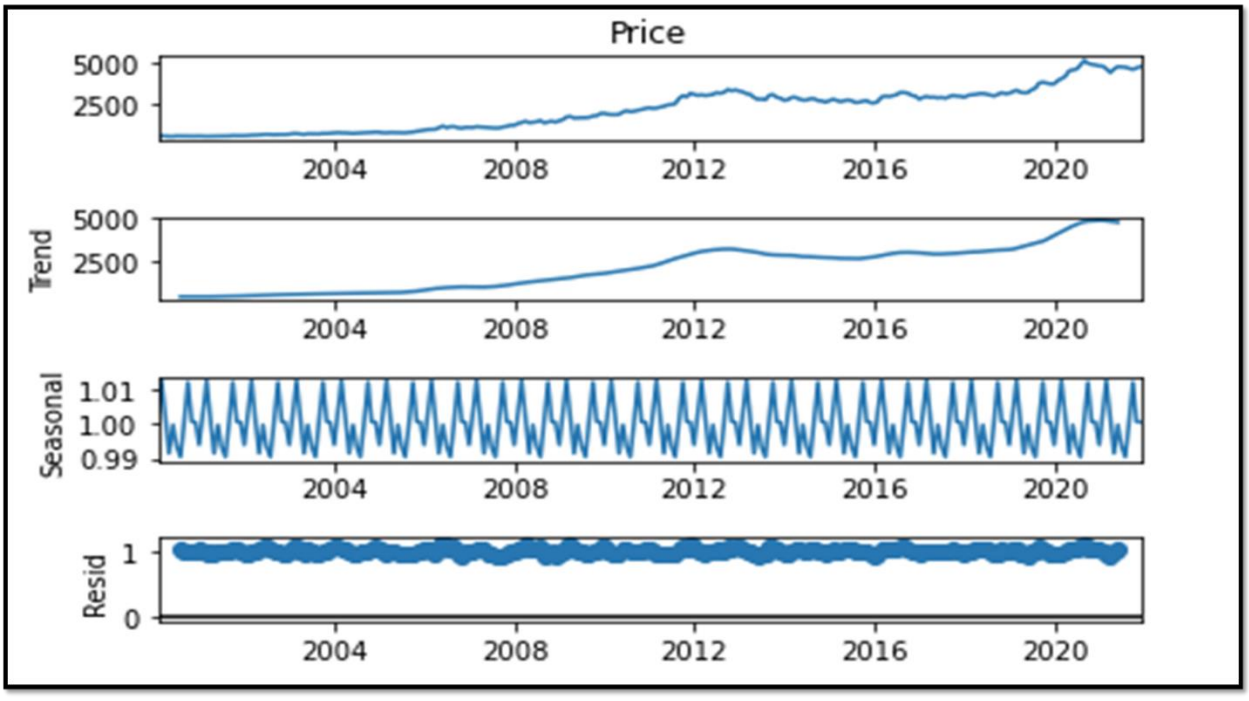


Conclusion: Here we can observe that the variation of gold prices in the year 2020 is much more than any other year.

> Components of Time Series



> Multiplicative Decomposition



<u>Conclusion :</u> Here we can see that graph of our data set showing a upward trend. And it is also showing a seasonality. If we will combine these all decomposition we will get original plot of our data set.

> Testing for Stationarity

Augmented Dickey-Fuller test:

_____The null hypothesis for the test is that the data is not stationary.

_____The alternate hypothesis for the test is that the data is stationary.

ADF Statistic: 0.6057144242539166

p-value: 0.9877556362370769

Conclusion: With the help of ADF test we can conclude that p-value is greater than 0.05 l.o.s. then we do not reject H0 and our data is not stationary.

Kwiatkowski-Phillips-Schmidt-Shin:

_____The null hypothesis for the test is that the data is stationary.

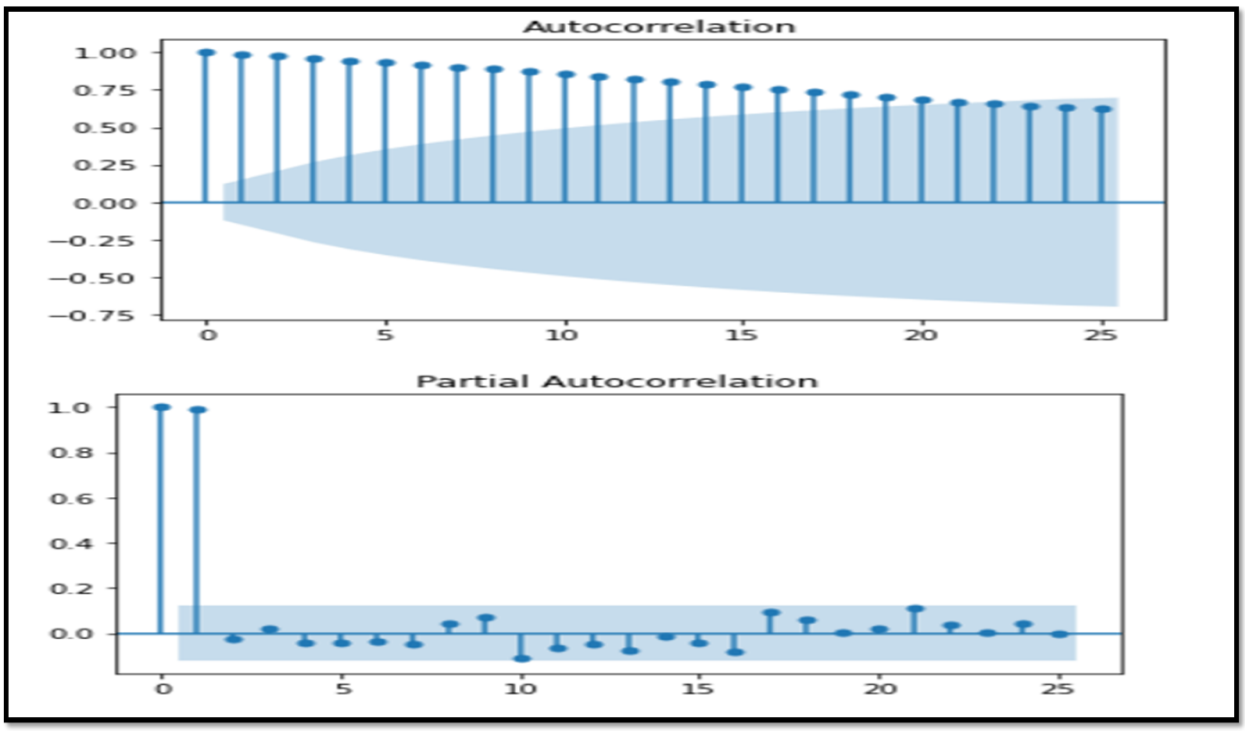
_____The alternate hypothesis for the test is that the data is not stationary.

KPSS Statistic: 2.3045598973941983

p-value: 0.01

Conclusion: Similarly by kpss test also we get that data is not stationary as p-value is less than level of significance (0.05) then we reject H0 i.e we accept H1.

> Autocorrelation and Partial-Autocorrelation



Conclusion:

In given graph of acf most of spikes are crossing the statistical limit i.e. confidence intervals then we can say that our data is not stationary.

From pacf graph 1 spike crosses the threshold which say's our data is not stationary.

> Train and Test :

To prevent the model from overfitting and to accurately evaluate the model

We use train data to fit the model and test accuracy on the test data

<u>Train Data :</u> Used to estimate any Parameters of forecasting methods.

We have selected First 210 observations for train data set. (i.e. 80%)

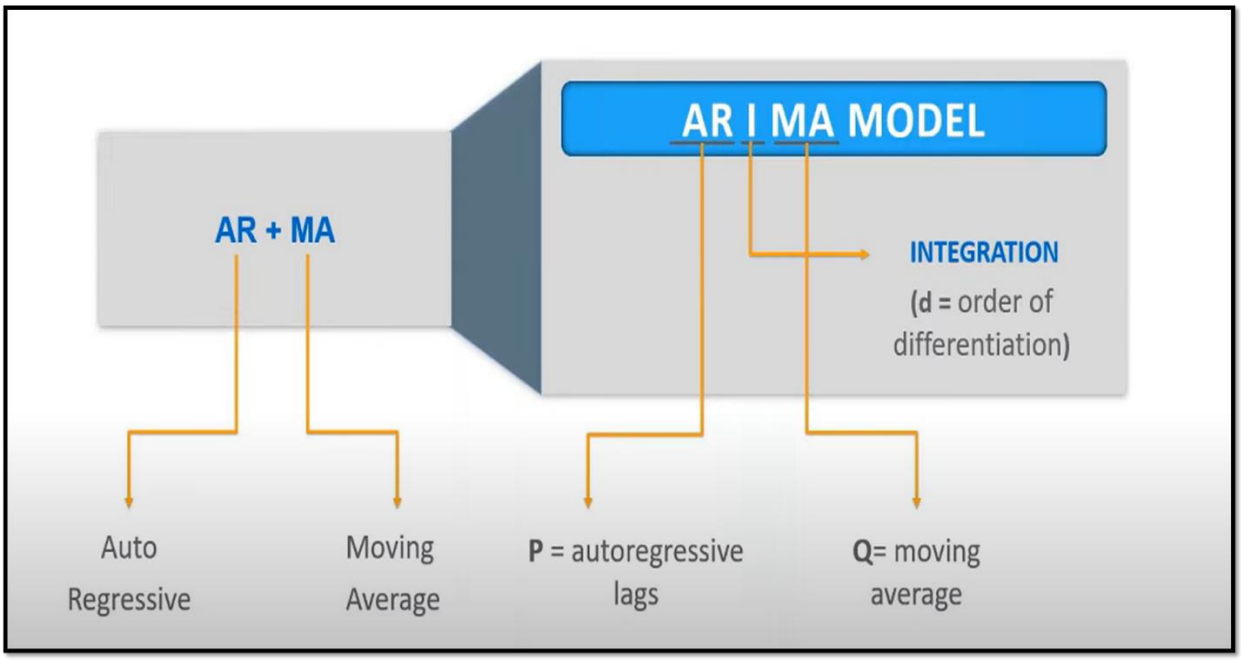
<u>Test Data :</u> Used to evaluate its accuracy.

The test set should be the most recent part of data.

53 observations left for test data set. (i.e. 20%)

We check whether our predicted value based on train dataset match the actual values of test dataset. If it is same which means our forecast model is a good fit.

> ARIMA (Auto Regressive Integrated Moving Average) Model



> ARIMA best fit

Performing stepwise search to minimize aic

ARIMA(2,1,2)(0,0,0)[0] intercept : AIC=3075.195, Time=0.67 sec

ARIMA(0,1,0)(0,0,0)[0] intercept : AIC=3079.106, Time=0.03 sec

ARIMA(1,1,0)(0,0,0)[0] intercept : AIC=3073.405, Time=18.65 sec

ARIMA(0,1,1)(0,0,0)[0] intercept : AIC=3071.565, Time=0.16 sec

ARIMA(0,1,0)(0,0,0)[0] : AIC=3086.765, Time=0.02 sec

ARIMA(1,1,1)(0,0,0)[0] intercept : AIC=3071.200, Time=0.22 sec

ARIMA(2,1,1)(0,0,0)[0] intercept : AIC=3073.192, Time=0.53 sec

ARIMA(1,1,2)(0,0,0)[0] intercept : AIC=3073.190, Time=0.47 sec

ARIMA(0,1,2)(0,0,0)[0] intercept : AIC=3071.571, Time=0.21 sec

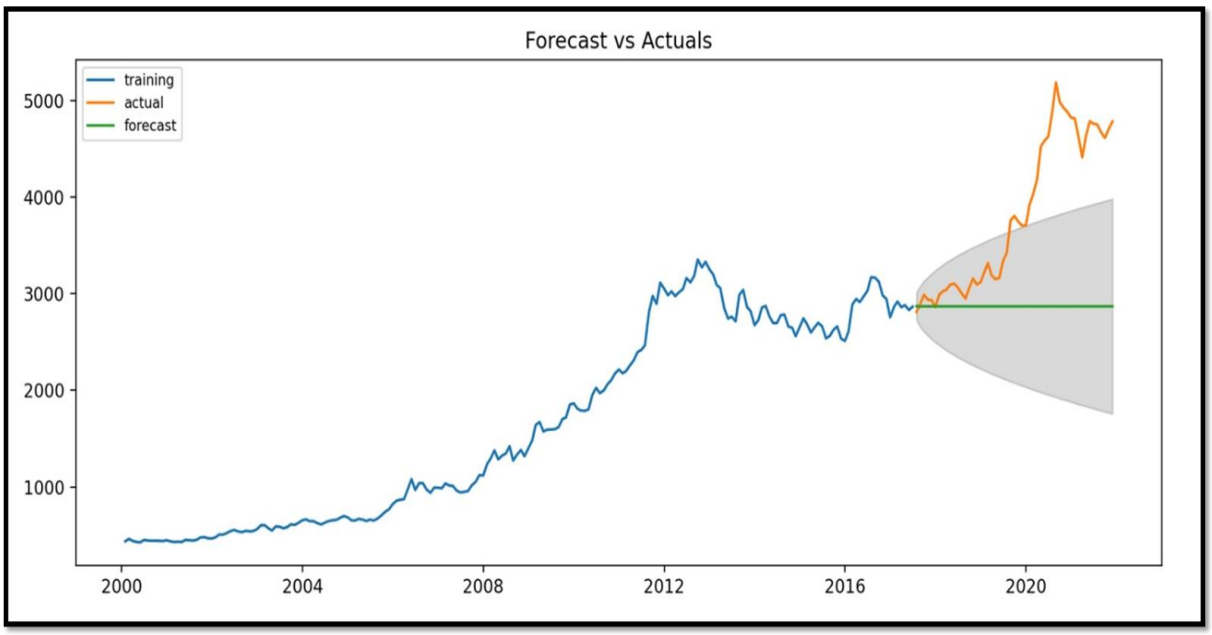
ARIMA(2,1,0)(0,0,0)[0] intercept : AIC=3072.522, Time=0.05 sec

ARIMA(1,1,1)(0,0,0)[0] : AIC=3076.900, Time=0.28 sec

> *Best model:*

ARIMA(1,1,1)(0,0,0)[0] intercept

ARIMA's model see if they are white noise or not that's why pick the best model with well behaved residuals then we forecast the model



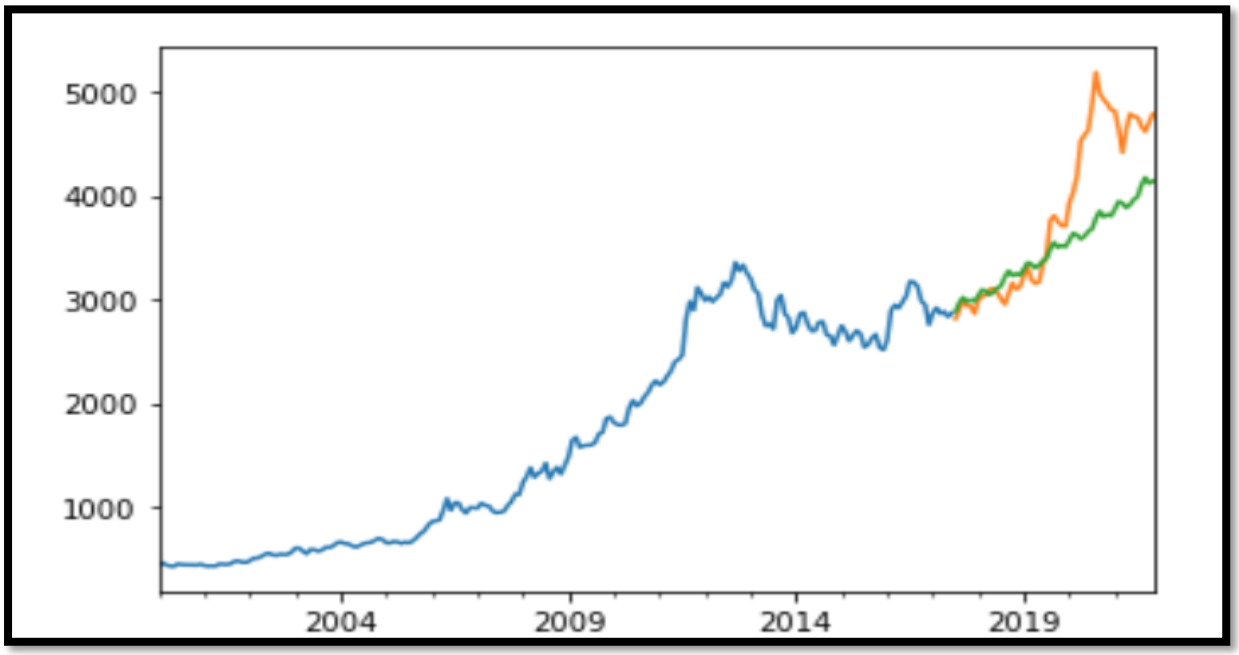
Conclusion :

We can observe our graph and it is not giving a proper prediction according to test data set.

Hence ARIMA model is not good fit.

> Exponential Smoothing

3rd Stage (Triple Exponential Smoothing)



Conclusion :

As a seasonality and trend is present in our dataset so third stage exponential smoothing gave us better results forecasts.

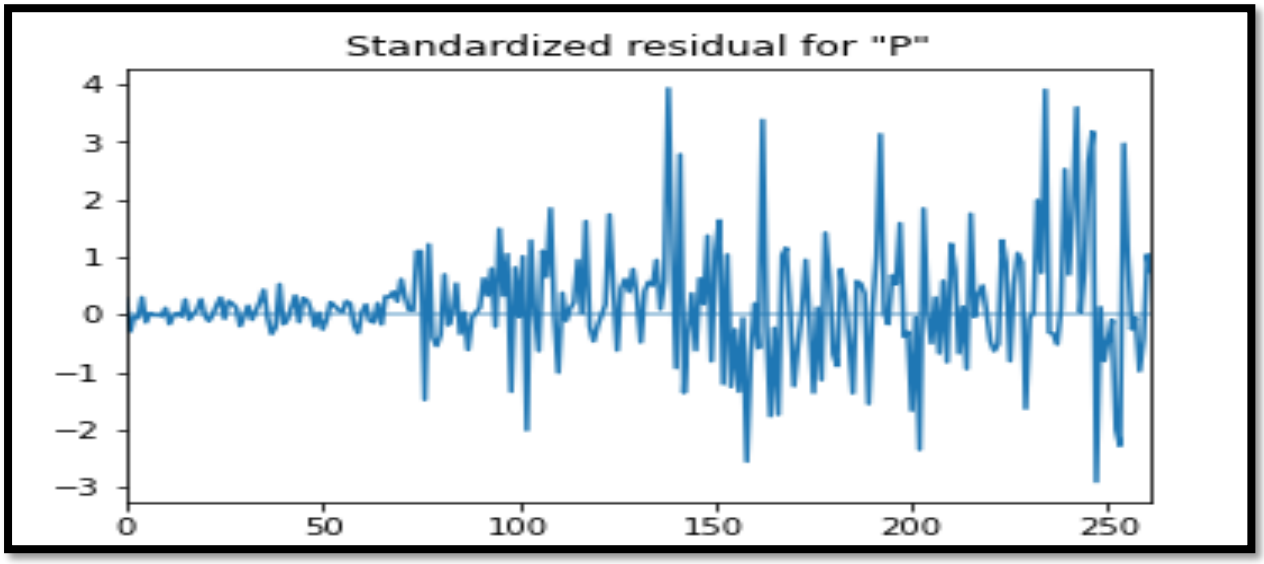
> **Comparison of MAPE values**

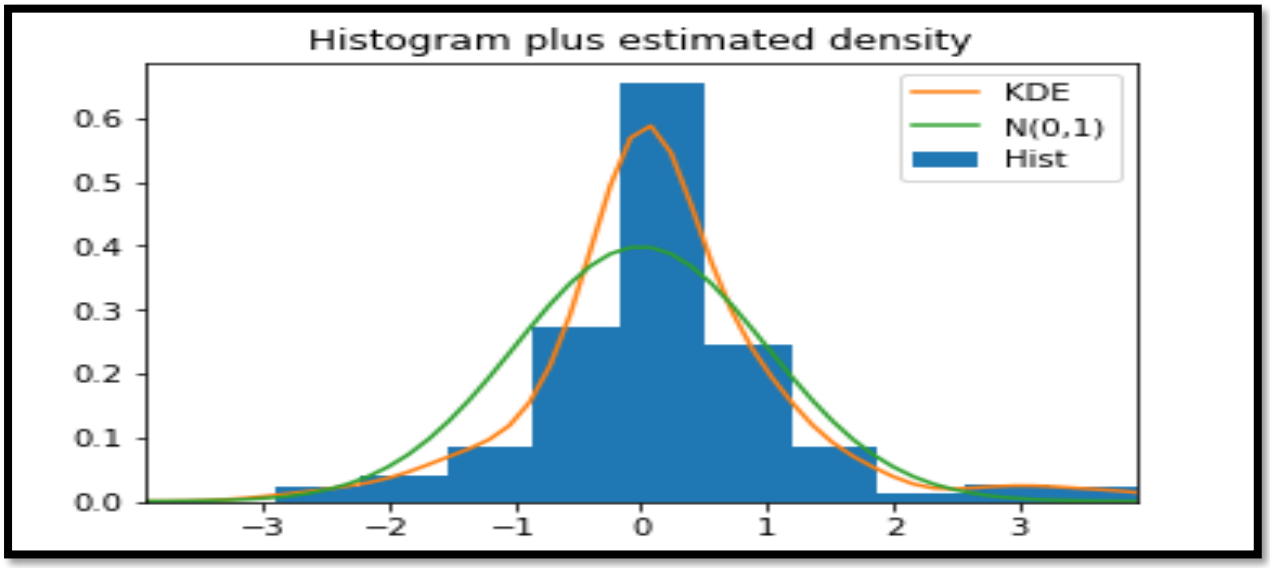
MODELS	MAPE VALUES
Autoreggressive Integrated Moving Average(ARIMA)	22
Simple Exponential Smoothing	22
Double Exponential Smoothing	20
Triple Exponential Smoothing (Additive)	20
Triple Exponential Soothing (Multiplicative)	17

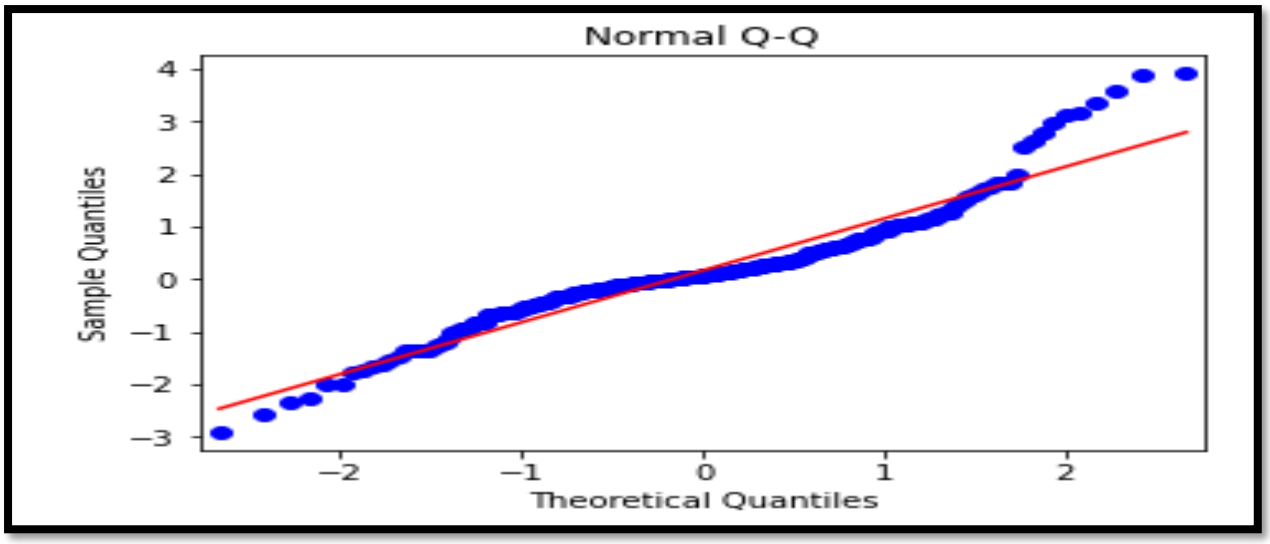
Conclusion :

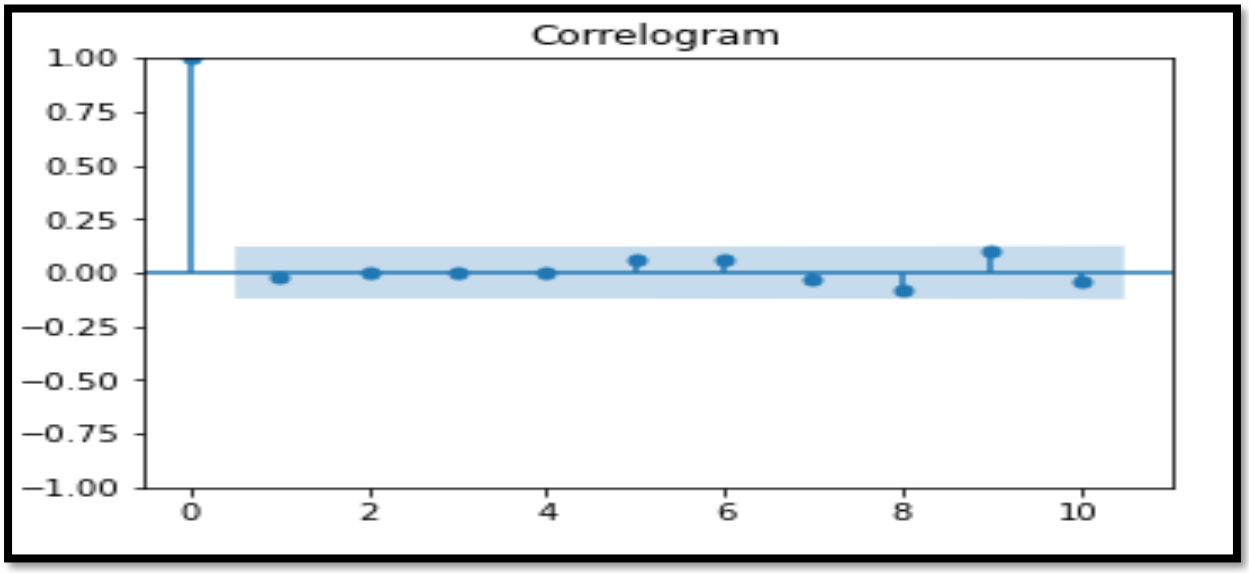
From this table we can see that the leader mape value we got for triple exponential smoothing and so we will proceed our forecasting using triple exponential smoothing model only.

> Model adequacy



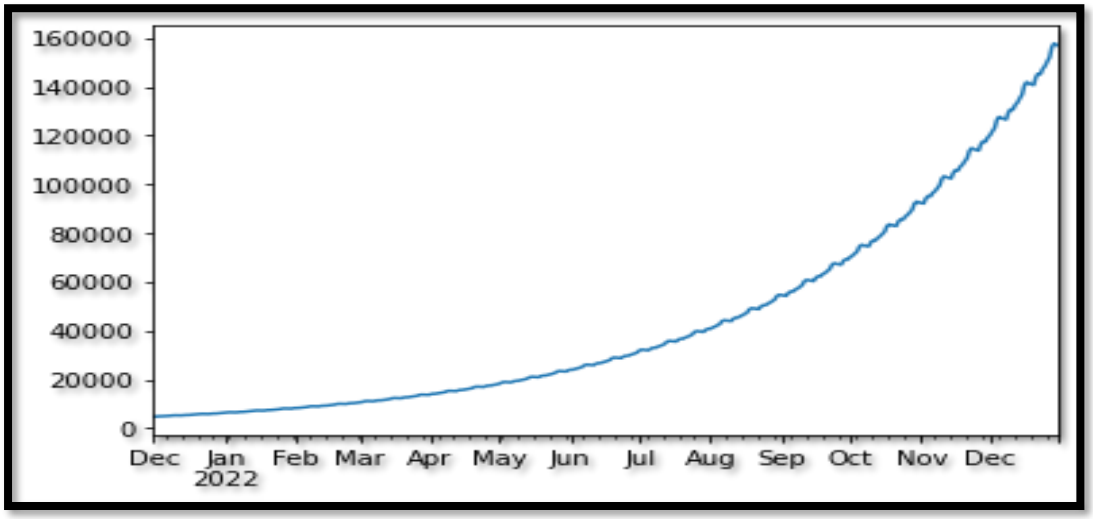






We checked for model adequacy and here we can see that our model is adequate.

> Prediction Plot



2021-12-01	4764.013221
2021-12-02	4870.531071
2021-12-03	4923.916539
2021-12-04	4916.787216
2021-12-05	4989.800006
	...
2022-12-27	152063.817284
2022-12-28	156214.126221
2022-12-29	157785.256596
2022-12-30	157154.616079
2022-12-31	156758.147963

Conclusion :

We have predicted for the next one year and we conclude that gold prices will increase exponentially

Reference used : **https://www.youtube.com/watch?v=8FCDpFhd1zk**

Software used : Python

Required packages for "time series analysis"

numpy

pandas

matplotlib

statsmodel.tsa.seasonal

Code used

file:///C:/Users/HP/Downloads/Untitled%20(2).html

file:///C:/Users/HP/Downloads/time_series_project_%20(2)%20(1).html