Gold Price Forecasting

(*Univariate time series forecasting using Facebook Prophet algorithm*)

Data Collection Strategy:

Data has been collected from the secondary source - world gold council https://www.gold.org/

The website contains gold price in different currencies. As we are interested in forecasting gold price (INR), we have collected gold price data in INR from the year1979 till 2020.

Exploratory Data Analysis:

The data is collected at daily level. The data is not continuous (i.e. it contains gold price of every week days). As the basic properties of time series requires the data to be continuous, appropriate treatment has been done to make the data continuous.

Actual Data

date	rupee
1979-01-08	1841.33
1979-01-09	1822.52
1979-01-10	1803.50
1979-01-11	1814.57
1979-01-12	1782.62
1979-01-15	1782.01
1979-01-16	1796.52
1979-01-17	1842.06
1979-01-18	1861.74
1979-01-19	1940.75

Interpolation

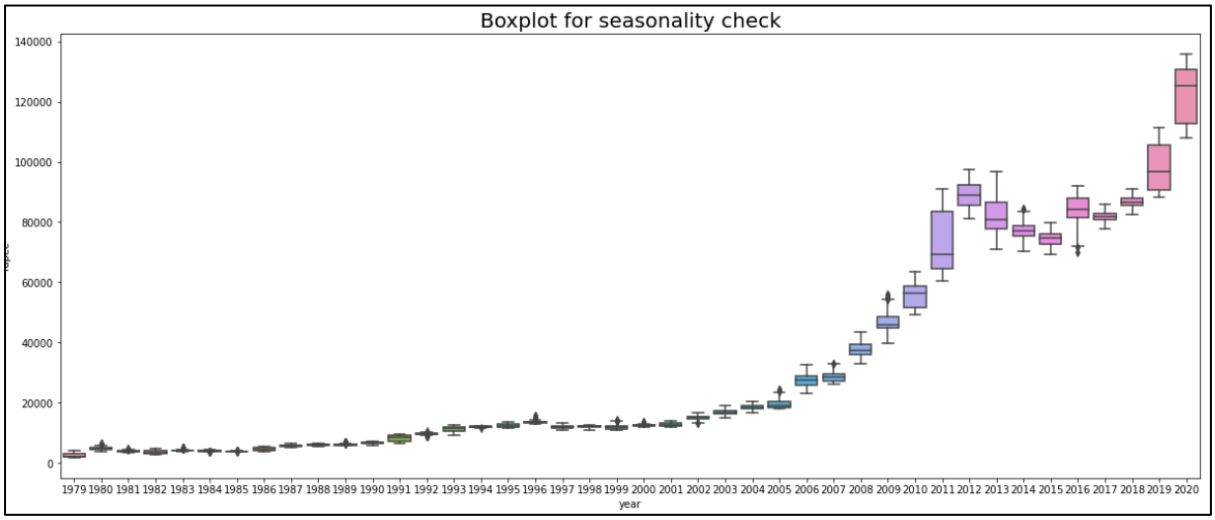
Continuous Data

date	rupee
1979-01-08	1841.330000
1979-01-09	1822.520000
1979-01-10	1803.500000
1979-01-11	1814.570000
1979-01-12	1782.620000
1979-01-13	1782.416667
1979-01-14	1782.213333
1979-01-15	1782.010000
1979-01-16	1796.520000
1979-01-17	1842.060000

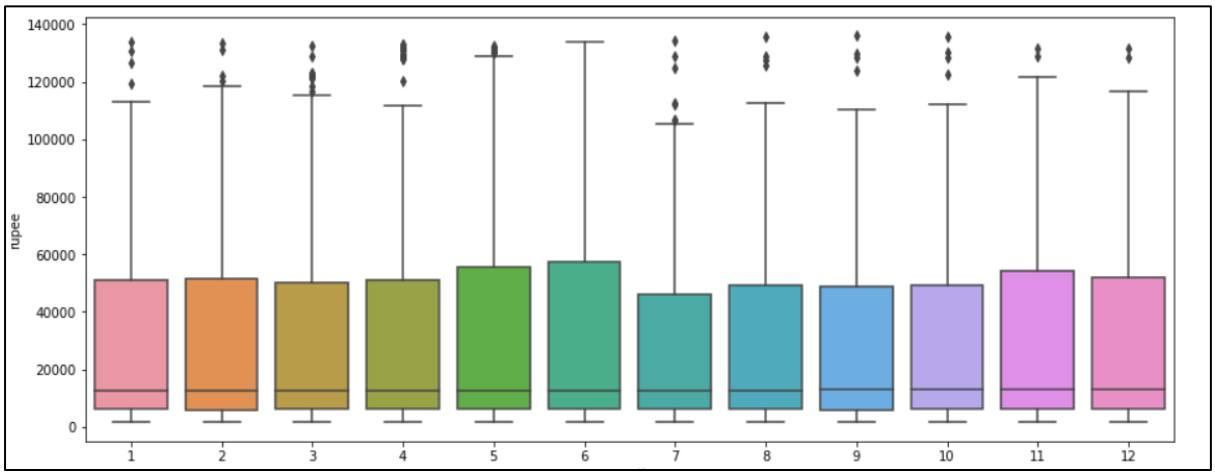
Time Series Plot

Linear growth of gold price till 2005, henceforth an exponential growth is visible with lots of seasonal ups and down.

Seasonality Plot

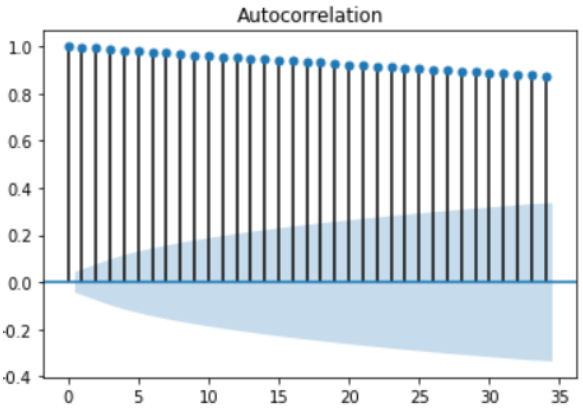


Month wise distribution of gold price

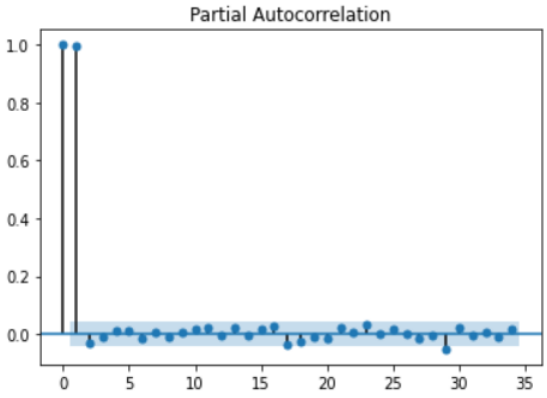


The gold price tends to increase during middle and end of the year

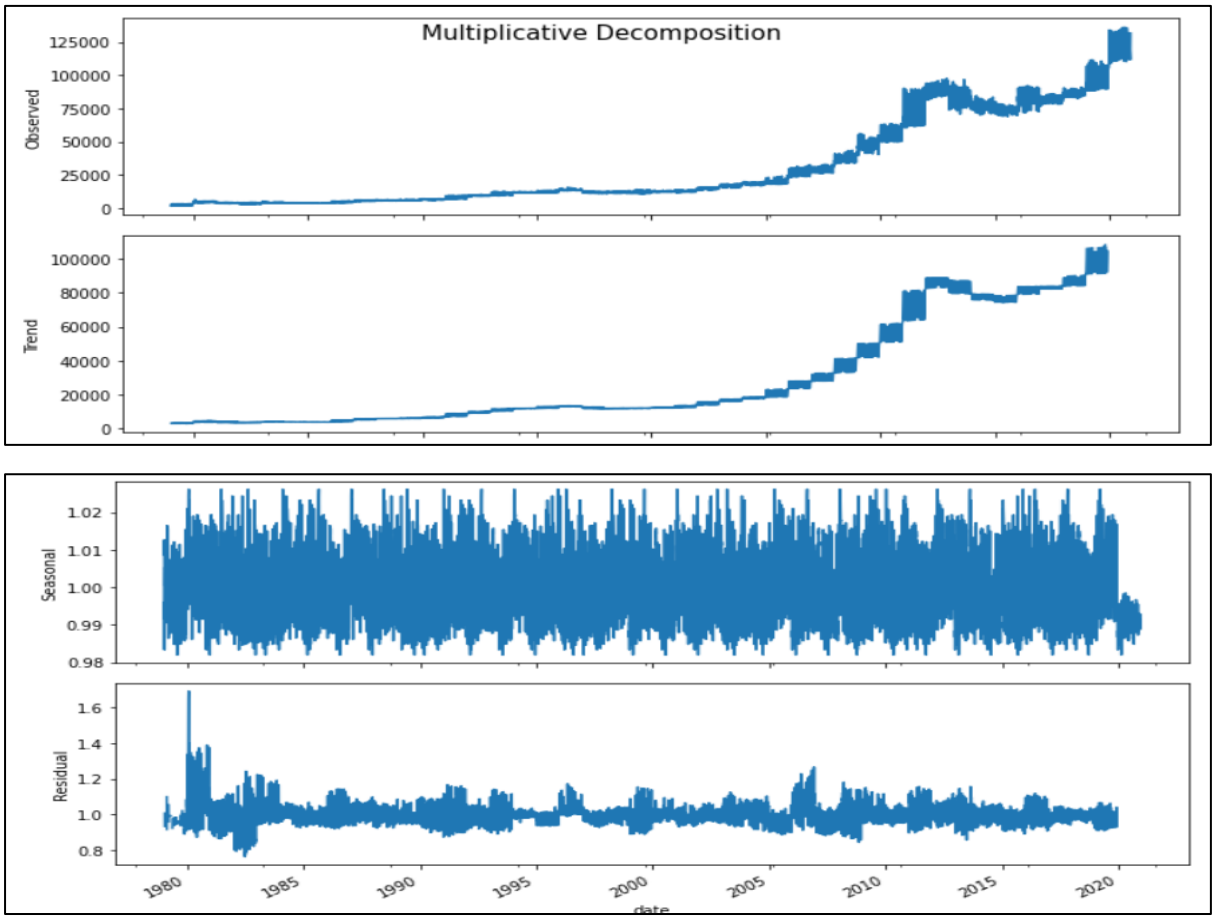
ACF Plot



PACF Plot

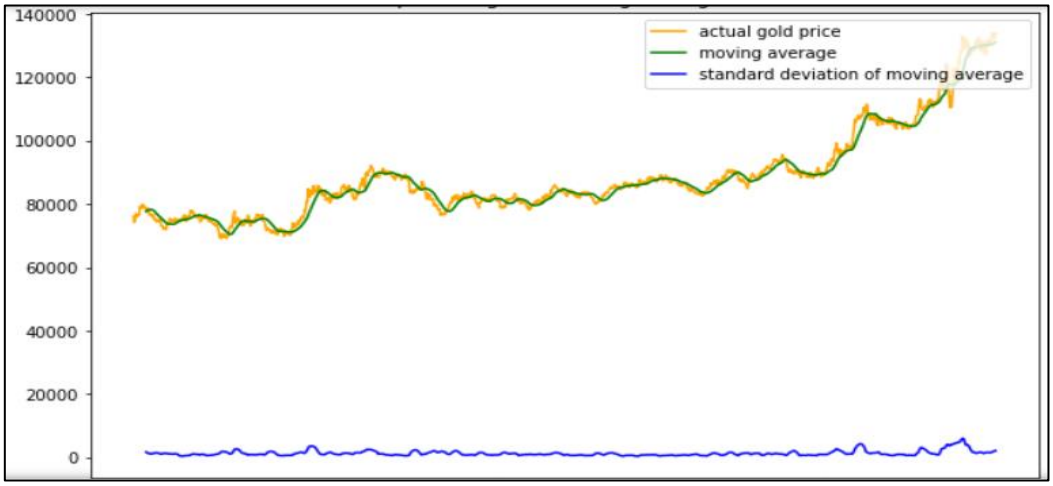


Time Series Decomposition:



Check for Stationarity:

- **Visual Inspection:**



From the visual inspection of time series along with its moving average and standard deviation, it is pretty clear that the time series is non-stationary. Confirming the same with ADF test

- **ADF test:**

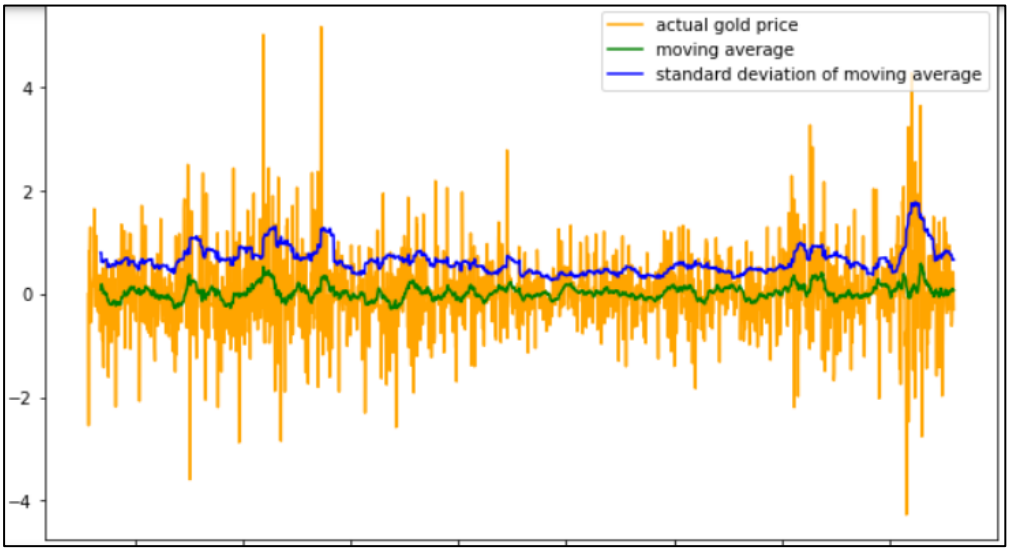
```
Augmented-Dickey_Fuller of gold price(actual series)

ADF Statistic: 1.389120
p-value: 0.997064
Critical Values:
        1%: -3.434
        5%: -2.863
        10%: -2.568
```

P-value greater than 0.05 indicate the rejection of alternate hypothesis. Hence the time series is non-stationary.

Transforming the data into stationary series:

- Differencing: First differencing is done to make the data near stationary. Below is the plot and statistical test for the same.




```
Augmented-Dickey_Fuller(transformed series)

ADF Statistic: -10.188356
p-value: 0.000000
Critical Values:
        1%: -3.434
        5%: -2.863
        10%: -2.568
```

p-value is less than 0.05, hence we can reject null hypothesis which means the transformed series is stationary.

Prophet Algorithm:

Facebook-Prophet describe a time series forecasting model designed to handle the common features of business time series. Importantly, it is also designed to have intuitive parameters that can be adjusted without knowing the details of the underlying model.

Prophet decomposable time series model with three main model components: ***trend, seasonality, and holidays***. They are combined in the following equation:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t.$$

Here g(t) is the trend function which models non-periodic changes in the value of the time series, s(t) represents periodic changes (e.g., weekly and yearly seasonality), and h(t) represents the effects of holidays which occur on potentially irregular schedules over one or more days. The error term ∈t represents any idiosyncratic changes which are not accommodated by the model. The algorithm also makes the parametric assumption that ∈t is normally distributed.

- **The Trend Model**
  Prophet has the option to choose two trend models that cover almost all applications: a saturating growth model, and a piecewise linear model.

  Linear growth model

$$g(t) = (k + \mathbf{a}(t)^\intercal \boldsymbol{\delta})t + (m + \mathbf{a}(t)^\intercal \boldsymbol{\gamma}),$$

  Logistic growth model

$$g(t) = \frac{C}{1 + \exp(-k(t - m))},$$

- **Seasonality Model**
  Business time series often have multi-period seasonality as a result of the human behaviors they represent. For instance, a 5-day work week can produce effects on a time series that repeat each week, while vacation schedules and school breaks can produce effects that repeat each year. To fit and forecast these effects the model must specify seasonality models that are periodic functions of t. Prophet rely on Fourier series to provide a flexible model of periodic effects.

  Let P be the regular period, we expect the time series to have (e.g. P = 365 for yearly data or P = 7 for weekly data, when we scale our time variable in days). We can approximate arbitrary smooth seasonal effects with

$$s(t) = \sum_{n=1}^{N} \left(a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right)$$

- **Holidays and Events**
  Incorporating list of holidays into the model is a game changer phenomenon. The model is made straightforward by assuming that the effects of holidays are independent. For each holiday i, let Di be the set of past and future dates for that holiday. Prophet add an indicator function representing whether time t is during holiday i, and assign each holiday a parameter ki which is the corresponding change in the forecast. This is done in a similar way as seasonality by generating a matrix of regressors.

$$Z(t) = [\mathbf{1}(t \in D_1), \ldots, \mathbf{1}(t \in D_L)]$$

and

$$h(t) = Z(t)\boldsymbol{\kappa}.$$

As with seasonality, we use a prior k ~Normal (0, σ^2).

Model Building:

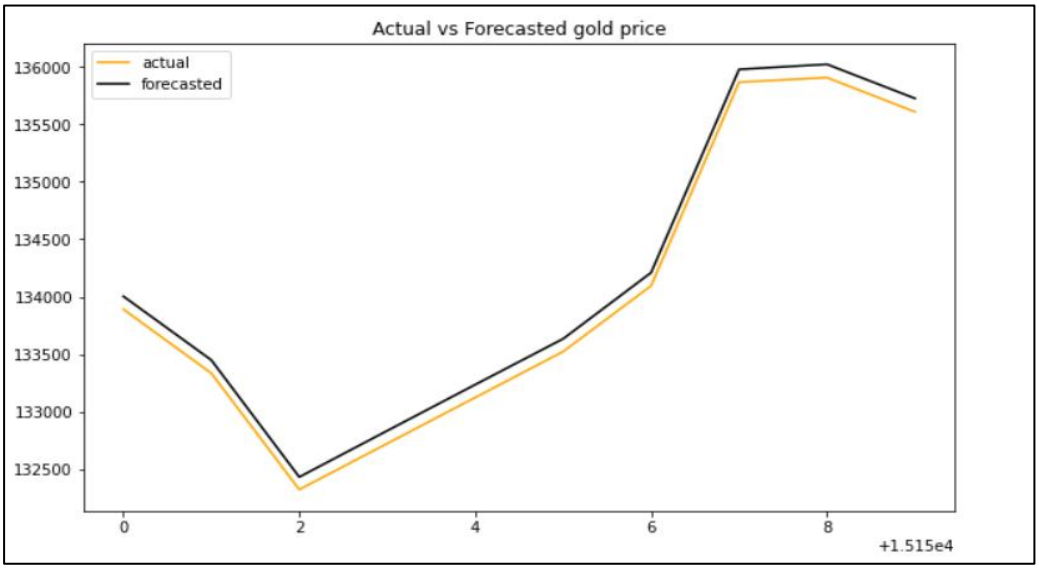
Data points since Jan,2015 till June,2020 has been considered for building the time series model. Recent 10 days data points has been kept as test data and the rest of the data has been used for building the model. Check for over fitting is done during model building. Appropriate parameters are set for trend, seasonality and irregularity components.

Forecasting

- Forecasting on the test data (horizon=10 days)

date	rupee	Forecasted_GoldPrice
2020-07-01	133891.39	134003.924217
2020-07-02	133335.41	133449.906123
2020-07-03	132324.81	132433.634952
2020-07-04	132724.46	132835.679094
2020-07-05	133124.11	133236.325935
2020-07-06	133523.76	133634.518196
2020-07-07	134095.45	134209.702695
2020-07-08	135864.20	135975.600917
2020-07-09	135904.16	136019.545563
2020-07-10	135606.66	135723.044193

Plot of test data vs actual gold price



Model Metrics:

Model Name	Test RMSE	Test MAPE
Prophet	112.76	0.084 %

***Note**: As an experiment the model was build with forward fill, backward fill and interpolation technique. It has been found that the interpolation technique is giving a better forecasting model.*

Also, different strategies for stationarity has been applied to the data out of which differencing is the best.

Thank you