ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN1176

Year Wise Analysis and Prediction of Gold Rates

Sonali Alpeshbhai Thakkar[1]; Kush Dineshbhai Jain[2]; Jayesh Jayantilal Parikh[3]
Assistant Professor[1,2,3]
Silver Oak College of Computer Application,
Silver Oak University, Ahmadabad, Gujarat

Abstract:- Gold is the only commodity which maintains its value even in the economic and financial crisis. The future gold price prediction becomes the warning system for the investors to predict risk in the market. Given historical price data up to and including a given day's opening price, closing price, attempted to analyze what variation is made between 2 or 5 years and what price ratio would be in given next 2-5 years.

Keywords:- *Prediction Of Gold Rate, Data Analysis and Data Set, Price, Consumer*

I. INTRODUCTION

India is the largest consumer of gold in the world, accounting for almost a quarter of the world's total consumption. It has, since long, maintained this position and, unlike countries like China, India uses gold primarily in the form of jewelry and investments. It is viewed as a solid instrument for investments and even traders, who are into commodities trading, invest in gold bullion. These investments are usually dictated by the gold rates prevailing in the economy at that time.

Gold rate in India change on a daily basis, with a number of factors impacting their price in a particular place on a given day. Demand and supply, global market conditions and currency fluctuations are some of the most critical factors which go into determining the rate of gold in a country, with prices changing every day.

A. Project Profile

- Project Title: Year wise analysis and prediction of gold rate.
- Project Type: Machine Learning.
- Tools: Python, Tableau
- Libraries: Matplotlib, Pandas, Numpy

B. Problem Statements

- Predicting price to go high/low.
- Reasons that are influencing the gold supply chain from the past empirical research findings.
- Tools are established to predict the future price of Gold.

II. DATA ORGANIZATION

A. Data Source

The data-set was collected from well-known website Kaggel.com.

B. Data Category

Given data set is in csv (excel) format and it can be categorized in gold rate dataset.

C. Data Description

➢ *It has Rows as:*

- Date - Date field has dates from 1979 to 2018.
- Price - It contains price of gold with reference to date.
- Open - It contains opening market price of gold.
- High - It Contains Max price reached referenced to date.
- Low - It Contains Min price reached referenced to date.

D. No of Rows and Columns

➢ *Data Set has 5 Columns and 9771 Rows.*

Table 1: Data Set

Filed	Description
Date	Shows a particular date.
Open	Open columns shows the gold open price on this particular day.
High	High column represent the maximum gold price of the day.
Low	Low column represent the Minimum gold price of the day.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN1176

Table 2: Data Visualization

Box Plot	Is the visual representation of the depicting groups of numerical data .A box plot consisting of 5 things. 1) Minimum 2) First Quartile or 25% 3) Median Second Quartile) or 50% 4) Third Quartile or 75% 5) Maximum.
Histogram	A histogram is a plot of the frequency distribution of numeric array by splitting it into small equal sized bins. We can investigate the distributions of the data by reviewing histograms. 1) A histogram for each variable in the time series. 2) A histogram of active power consumption for the two full years of data.

- Price Compared to 2005 to 2010 is very high.
- Ratio Between open and close price is fluctuating on 2012.

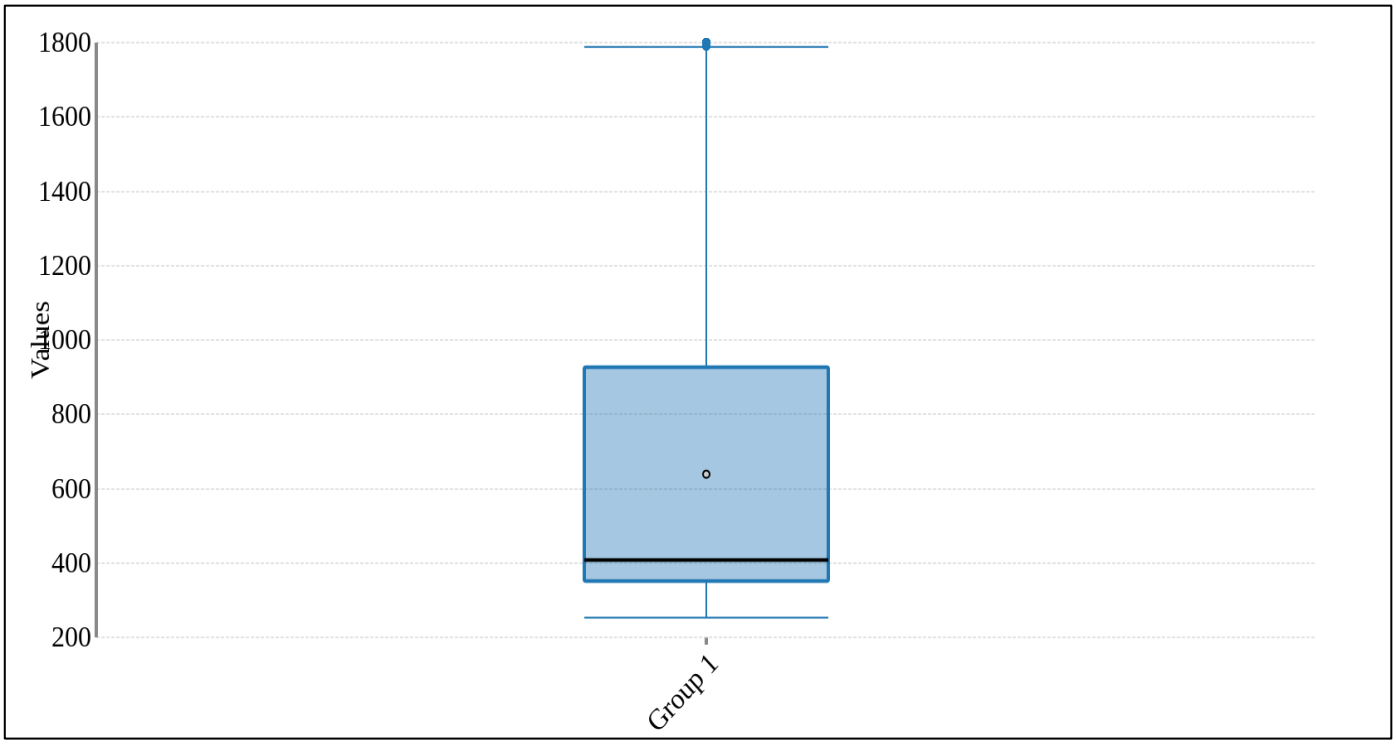

Fig 1: Box Plot for Price List

- **Description**: The above box plot depicts groups of numerical data through their quartiles for the price field from the dataset that is the gold rate provide is in range of 200 to 1800 and from year of 1999 to 2005.

Table 3: Data Summary

Data Summary								
Groups	N	Min	Q_1	Median	Q_3	Max	Mean	SD
Group 1	9771	253	351.15	407.9	926.75	1888.7	638.9222	419.7768

- **Description**: The summary have the data about N, min, Q1, median, Q3, Max and mean. It describes the value of the above field.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN1176

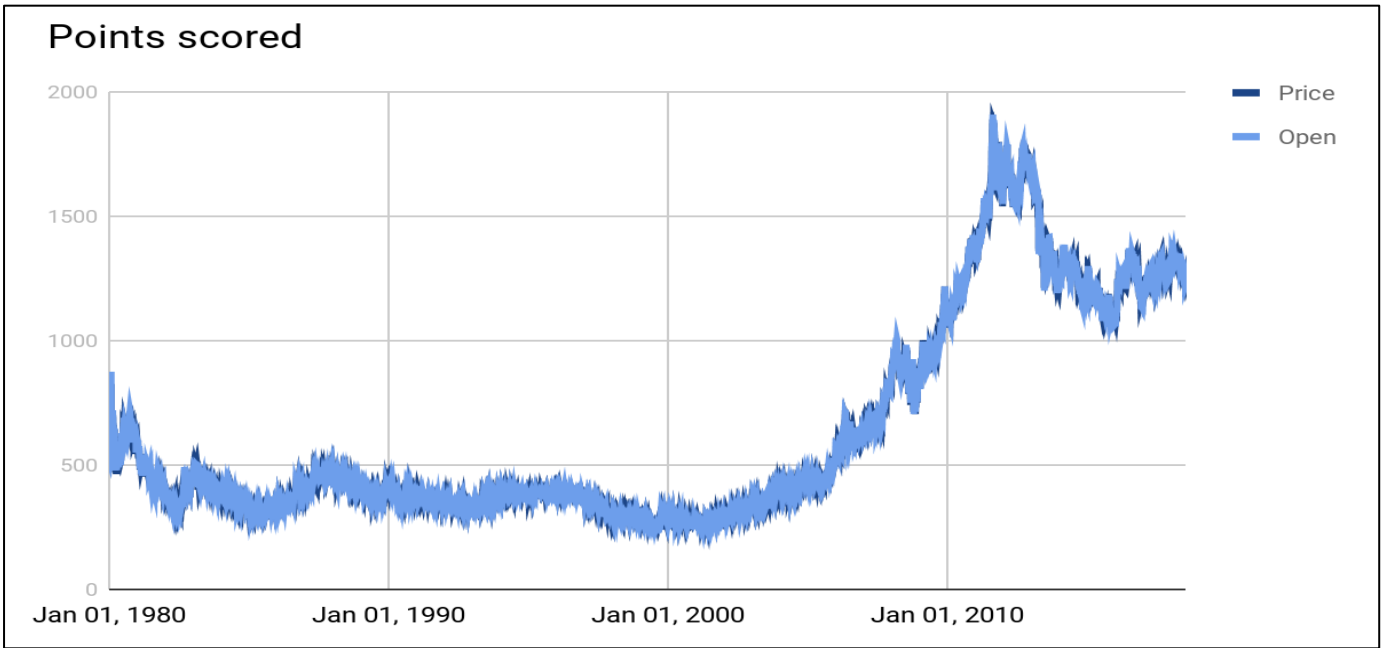

Fig 2: Scatter Plot for Opening and Closing Price of Gold Ratio

- **Description**: The scatter diagram is to analysis opening and closing price ratio or the difference between opening & closing price from 1980 to 2010.

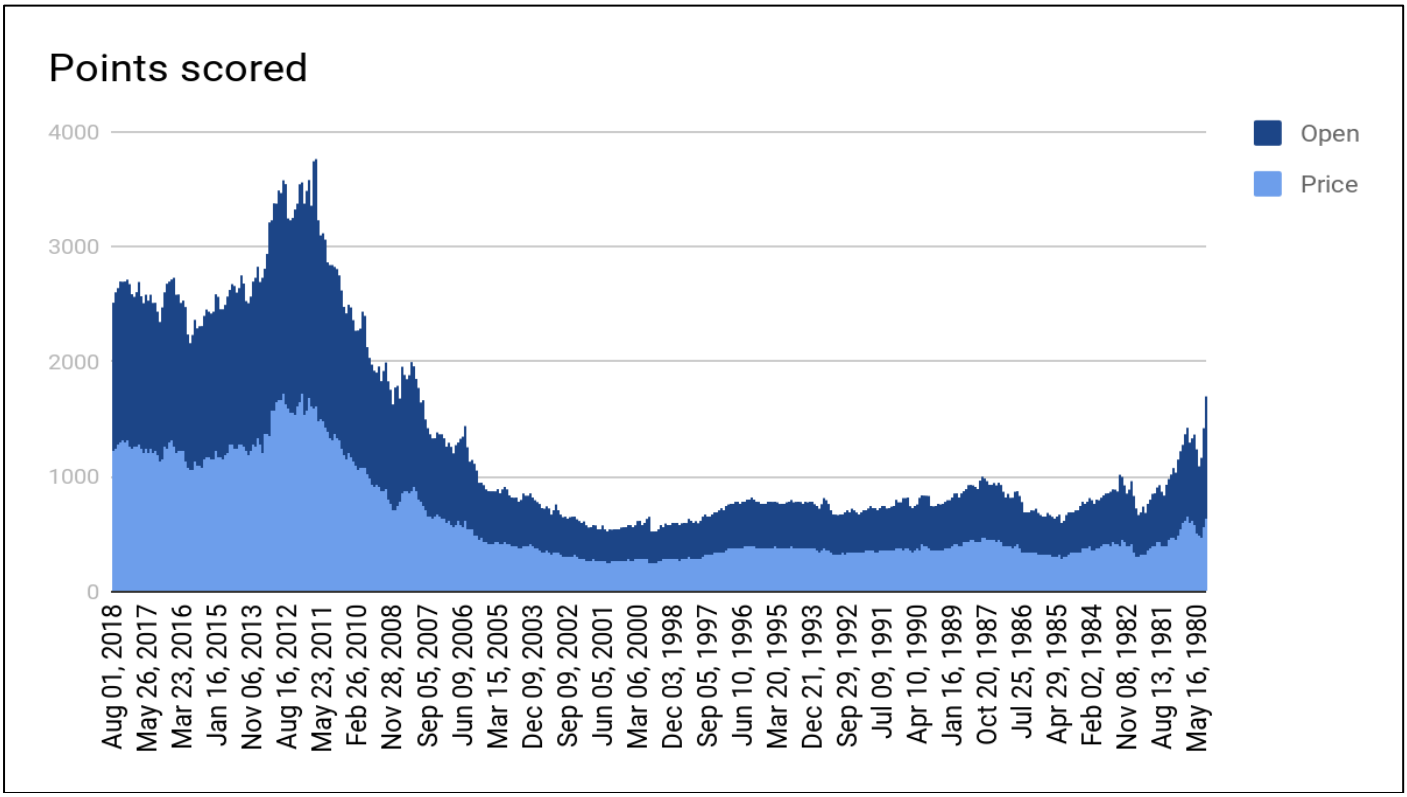

Fig 3: Scatter Diagram Represents Difference between Opening and Closing Price Of Gold Rate

- **Description**: The scatter diagram represents difference between opening and closing price with graphical representation.

III. DATA CLEANING

Data Cleaning is not necessary for this dataset because of cleaned dataset which does not have any null or missing values and we can filter out required values by simply using filter functionality in excel.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN1176

Table 4: Data Summary

Date	Price	Open	High	Low
Apr 01,	Min. : 253.0	Min. : 252.9	Min. : 252.9	Min. : 252.5
Apr 01, 1981: 1	1st Qu.: 351.1	1st Qu.: 351.2	1st Qu.: 352.0	1st Qu.: 350.5
Apr 01, 1982: 1	Median : 407.9	Median : 407.8	Median : 409.0	Median : 406.3
Apr 01, 1985: 1	Mean : 638.9	Mean : 639.1	Mean : 641.9	Mean : 635.9
Apr 01, 1986: 1	3rd Qu.: 926.8	3rd Qu.: 926.0	3rd Qu.: 932.5	3rd Qu.: 920.5
Apr 01, 1987: 1	Max. :1888.7	Max. :1909.0	Max. :1911.6	Max. :1864.0

- **Description:** The above table shows detailed description of gold opening price, maximum price, and minimum price and also mentions the date. That describes the on which date the changes have been made in the gold rate.

IV. HYPOTHESIS

Predicting Price for next 4 year ratio based on previous year's ratio, that price should increase by 2004 to 2009. In Festival month or Marriage season price should be increasing because of Indian culture after year 2014. Opening price ratio should high in 2017 with respect to closing price ratio.

➢ *Models*

```
import pandas as pd
#from sklearn.cross_validation import *;
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.cross_validation import train_test_split
import numpy as np
from sklearn import metrics

dataset=pd.read_csv("C://Users//admin//Desktop//goldx.csv")
print(dataset)
#x=dataset["Interest rate(%) X"].reshape(-1,1)
x=dataset.iloc[:,0:1]
#x = x.reshape((x.shape[0], 1))
y=dataset["Gold Price"]
y=y.replace(['Yes','No'],[1,0])
#y= pd.factorize(dataset['Wear a jacket'])[0]
print(y)
X_train, X_test, y_train, y_test = train_test_split(x, y, test_size=0.01, random_state=123)
model = LinearRegression()
model = model.fit(X_train, y_train)
y_pred = model.predict(X_test)
y_pred_val=model.predict(18)
print(y_pred_val)
if(y_pred_val > 0.5):
    print("Yes")
else:
    print("No")
plt.scatter(X_train,y_train, color = 'red')
plt.plot(X_train, model.predict(X_train))
print('Mean Absolute Error:', metrics.mean_absolute_error(y_test, y_pred))
print('Mean Squared Error:', metrics.mean_squared_error(y_test, y_pred))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))
```

Fig 4: Linear Regression

➢ *Calculation Summary*

- Sum of X = 635090.33
- Sum of Y = 630935.28
- Mean X = 1272.7261
- Mean Y = 1264.3994
- Sum of squares (SS_X) = 1202725.5465
- Sum of products (SP) = 1204596.9834
- Regression Equation = $\hat{y} = bX + a$

✓ $b = SP/SS_X$ = 1204596.98/1202725.55 = 1.00156
✓ $a = M_Y - bM_X$ = 1264.4 - (1*1272.73) = -10.30711
✓ $\hat{y}$ = 1.00156X - 10.30711

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN1176

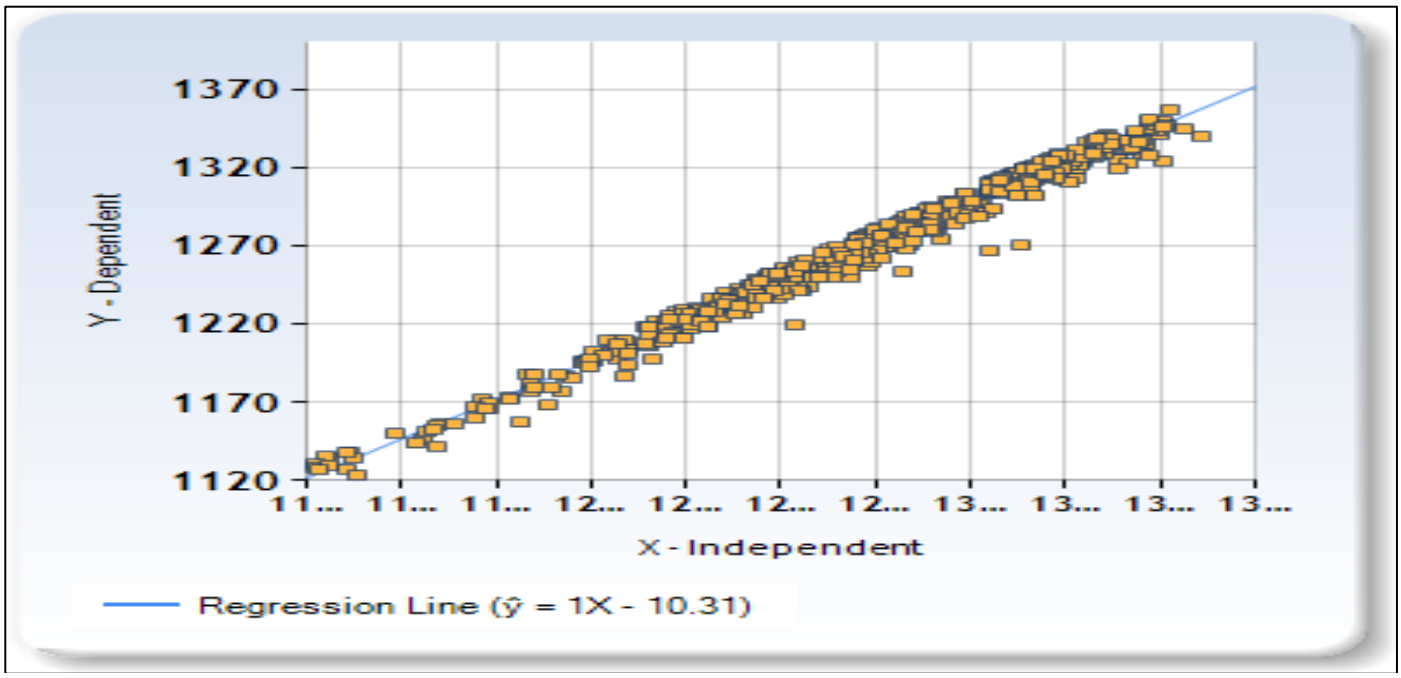

Fig 5: Linear Regression

- **Description**: Linear regression analysis is used to predict the value of a variable based on the value of another variable. To determine relation between opening price and closing price from data from 2017-2018, Linear Regression is required.

```
import numpy as np
import pandas as pd
from sklearn.metrics import confusion_matrix
from sklearn.cross_validation import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score


data = pd.read_csv("C://Users//Tirth//Desktop//goldx.csv")
#print(data)
predictors = data.iloc[:, 1:4]#seggregating the predictor variables
target = data.iloc[:, 4]#Seggregating the target/class variable

# Spliting the dataset into train and test
predictors_train,predictors_test,target_train,target_test = train_test_split(predictors, target, test_size=0.3, random_state = 123)
naive_bayes_classifier=GaussianNB()
model=naive_bayes_classifier.fit(predictors_train,target_train)
prediction=naive_bayes_classifier.predict(predictors_test)
print(naive_bayes_classifier.predict([0.9,0.9,0.9,0,1,0.7,0.9]))
print(accuracy_score(target_test,prediction,normalize=True))
print(confusion_matrix(target_test,prediction))
```

Fig 6: Naive Bayes Algorithm

- **Description**: Naïve bayes algorithm is supervised learning algorithm, which is based on bayes theorem and used for solving classification problems. Here the data is taken from csv file which name is goldx.csv.here we are predicting a a gold rate depending upon features of the gold.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN1176

➢ *Graphical Description of Data*

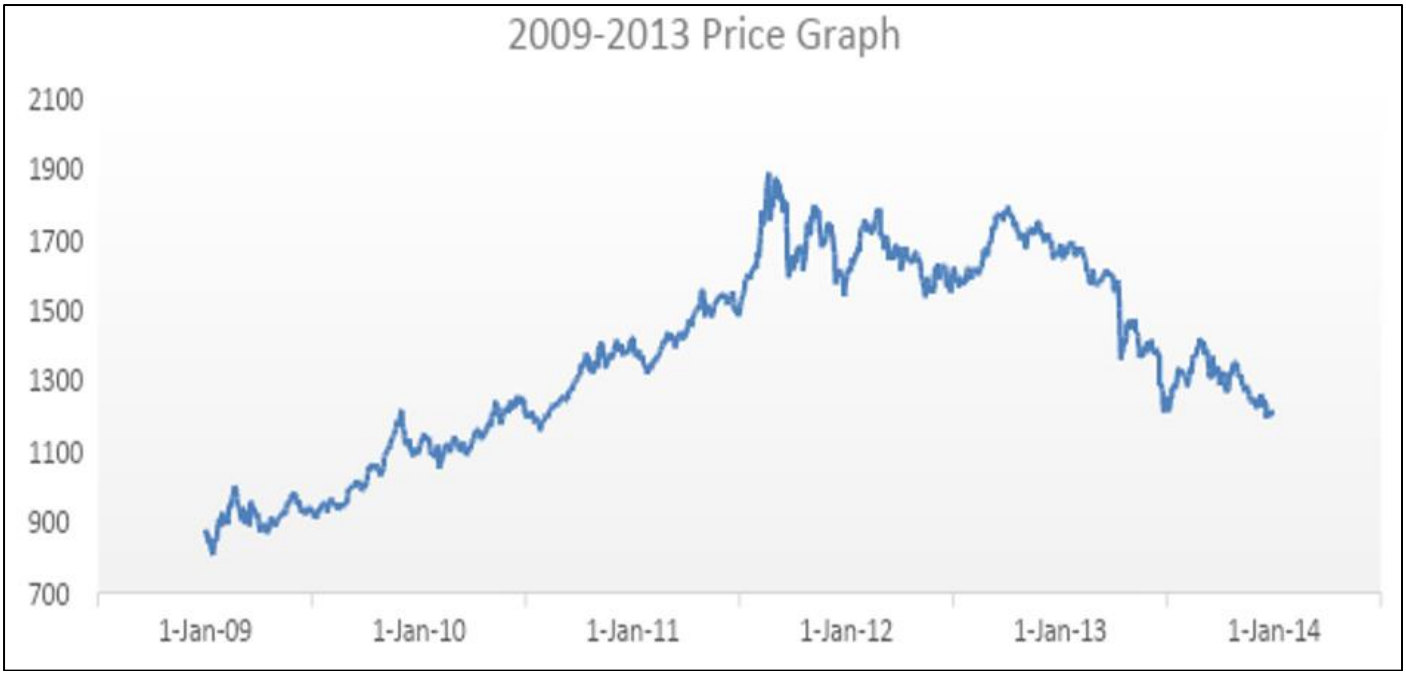

Fig 7: The Data Analysis about year 2009 to 2013

- **Description**: We are going to analyze the opening price of 2009-2013 and 1999-2003 to predict future price ratio.

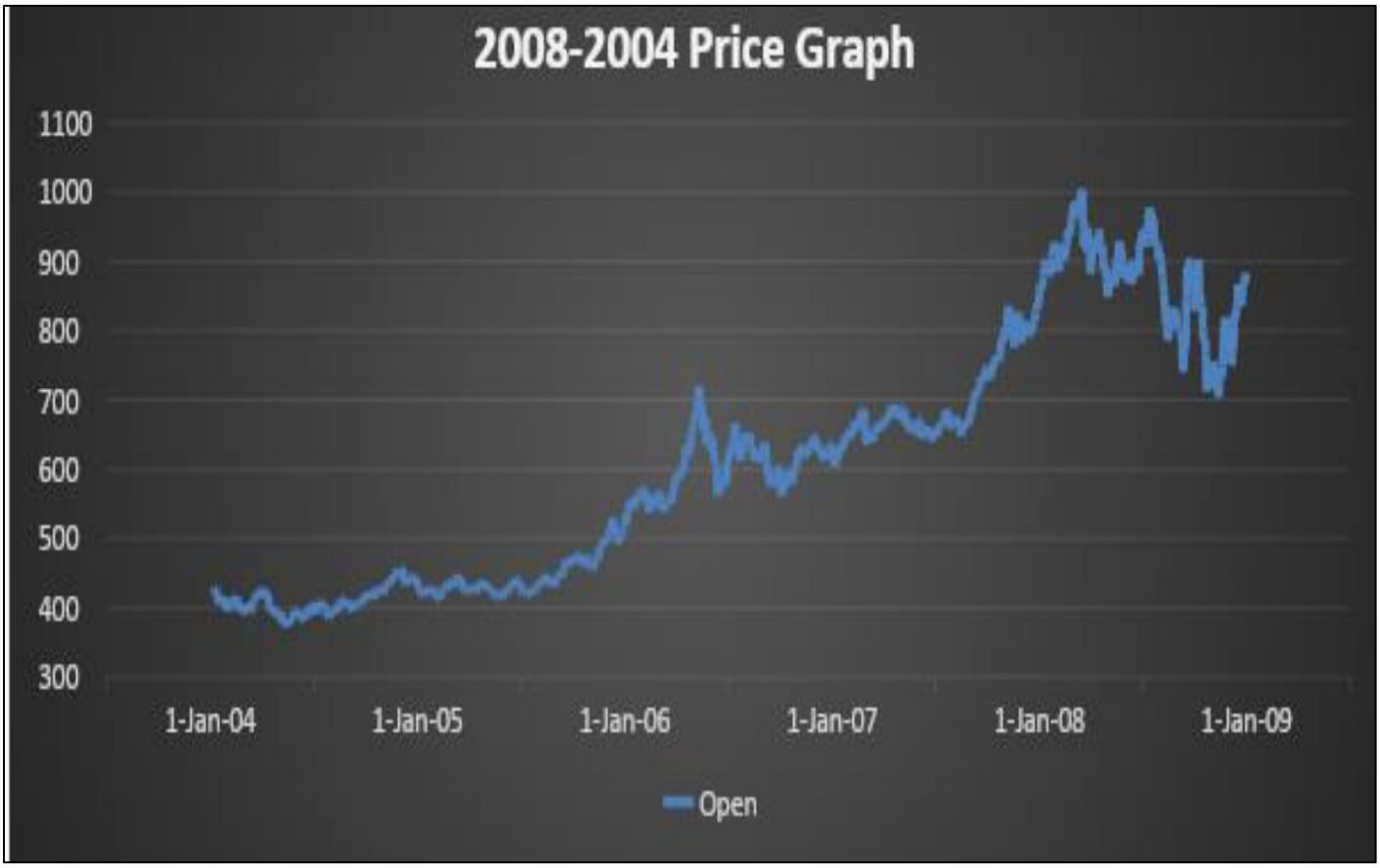

Fig 8: The Data Analysis about Year 2004 to 2008

- **Description**: We are going to analyze the opening price of 2004-2008 to predict future gold price ratio

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN1176

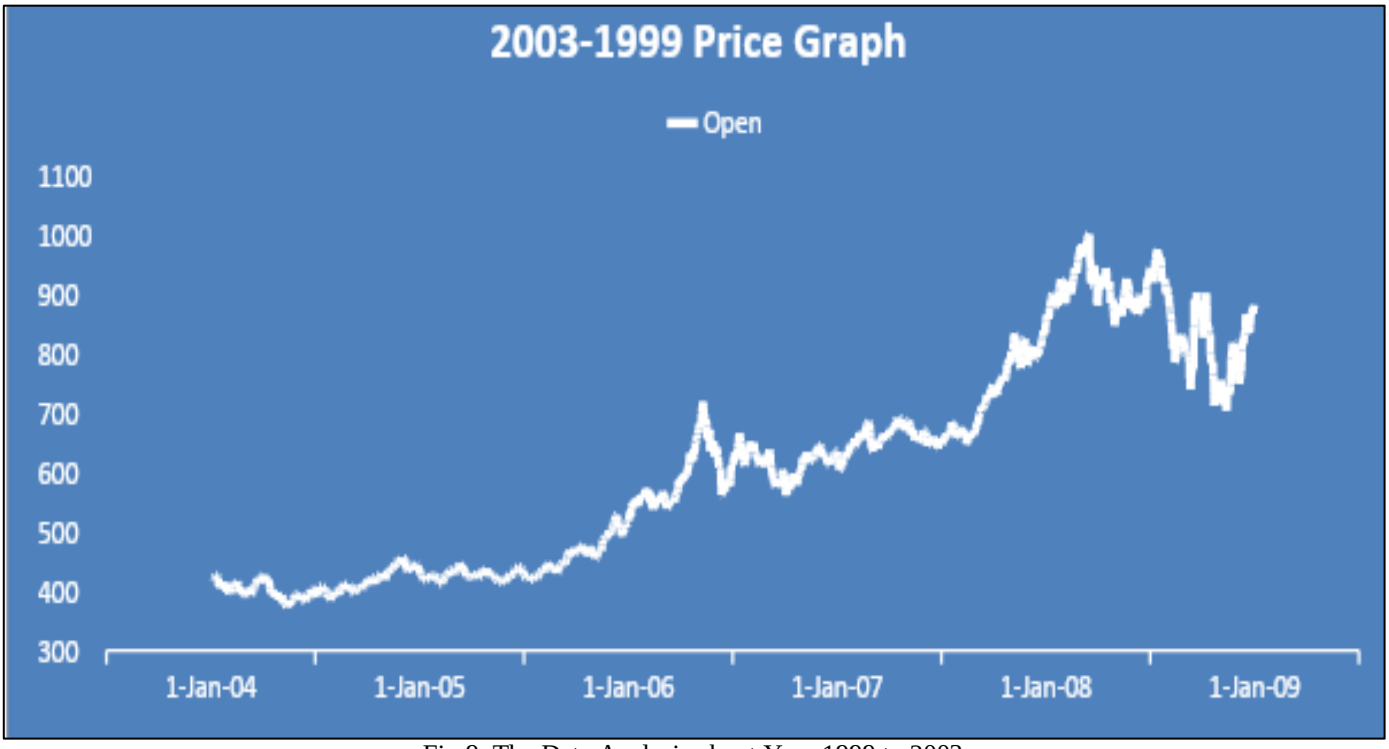

Fig 9: The Data Analysis about Year 1999 to 2003

- **Description**: We are going to analyze the opening price of 1999-2008 to predict future gold price ratio

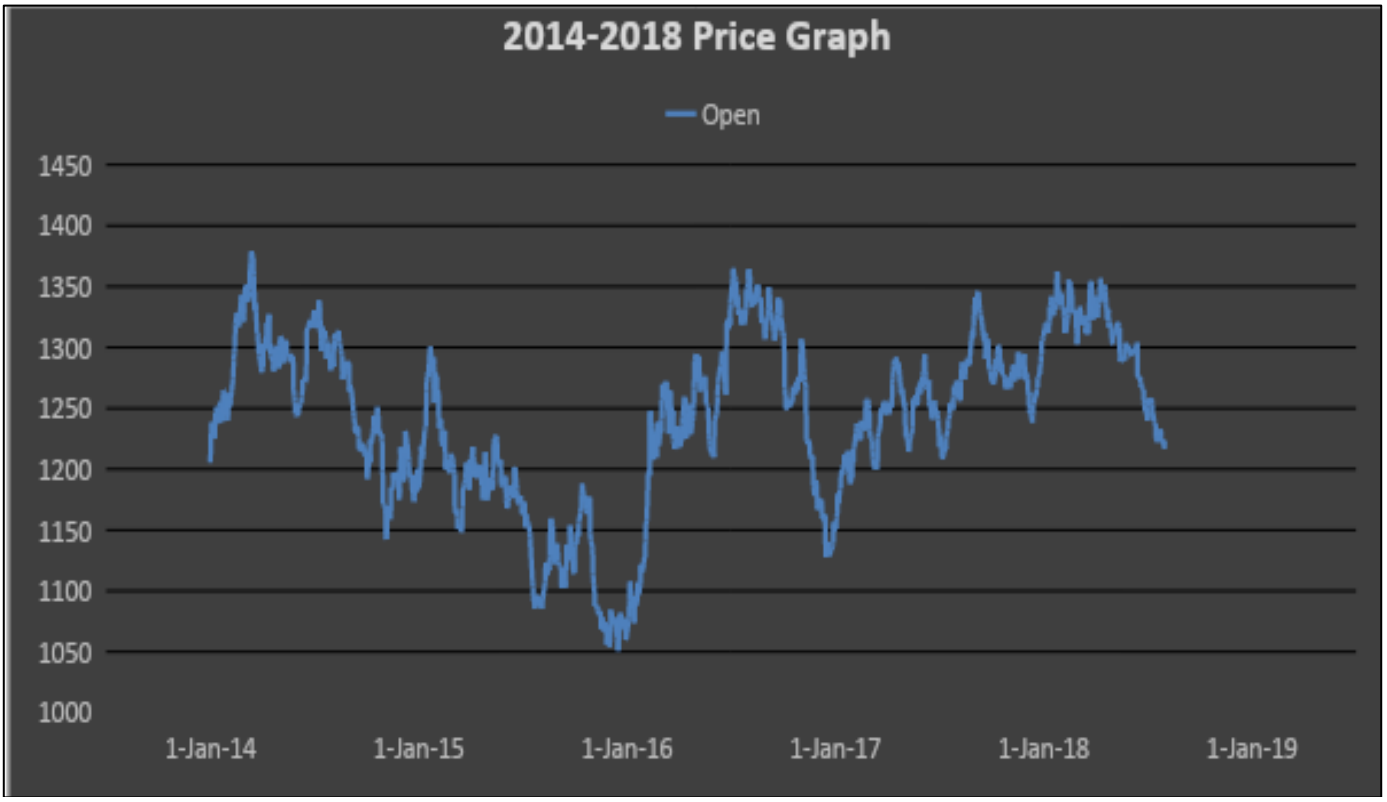

Fig 10: The Data Analysis about Year 2014 to 2018

- Here we are working with opening price of 2014-2018 to prove our hypothesis. And to do so we are analyzing price ratio with the help of the given graph.
- Here we are going to put all pricing data with graphical representation of how the price conflict or changed in past years to predict our future price on a monthly basis.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUN1176

Fig 11: The Data Analysis about Year 2017

- **Description**: We are going to analyze the points scored in year 2017

V. CONCLUSION

From the report of analyzed model linear regression over price of gold we came to conclusion that price confliction of gold 2017-2018 ratio have not high differences so we can get go ahead on developing or analyzing price through graphical representation. As our hypothesis and our report suggests that we come to conclusion that our hypothesis comes true and the price of gold goes high in years 2004-2009. As per our hypothesis for year after 2014 we failed on our hypothesis according to graph that suggest on January month of wedding month price of gold goes down. Opening Price is high at some moth of 2017-2018 but for all over month report our hypothesis fails that compared the price of 2017-2018 year's opening and closing are not as per hypothesis.

REFERENCES

➢ *Web Link:*

[1]. https://kaggel.com/
[2]. https://www.javatpoint.com/gold-price-prediction-using-machine-learning
[3]. https://www.kaggle.com/code/sid321axn/gold-price-prediction-using-machine-learning
[4]. https://www.kaggle.com/datasets/sid321axn/gold-price-prediction-dataset
[5]. https://www.w3schools.com/python/matplotlib_histograms.asp
[6]. https://www.geeksforgeeks.org/
[7]. https://en.wikipedia.org/wiki/Box_plot#:~:text=A%20boxplot%20is%20a%20standardized,the%20first%20and%20third%20quartiles.&text=First%20quartile%20(Q1%20or,lower%20half%20of%20the%20data set.
[8]. https://www.geeksforgeeks.org/box-plot-visualization-with-pandas-and-seaborn/
[9]. https://www.geeksforgeeks.org/plotting-histogram-in-python-using-matplotlib/
[10]. https://www.google.com/search?q=linear+regression&oq=linear+regression&gs_lcrp=EgZjaHJvbWUyDAgAEEUYORixAxiABDIHCAEQABiABDINCAIQABiDARixAxiABDINCAMQABiDARixAxiABDIMCAQQABhDGIAEGIoFMg0IBRAAGIMBGLEDGIAEMgoIBhAAGLEDGIAEMgcIBxAAGIAEMgcICBAAGIAEMgcICRAAGIAE0gEINDkwOWowajeoAgCwAgA&sourceid=chrome&ie=UTF-8
[11]. https://www.javatpoint.com/machine-learning-naive-bayes-classifier