Capstone Project

MACHINE LEARNING - REGRESSION

YES BANK STOCK CLOSING PRICE PREDICTION

TEAM MEMBERS	EMAIL
Pranav Rajmane	rajmanepranav19@gmail.com
Sayali Mandhare	saileemandhare98@gmail.com
Prathamesh Kalambe	Prathmeshkalambe122@gmail.com
Pavan Malvay	pmalvay@gmail.com

AI

Table of Content

- **Problem Statement**
- **Knowing the data set**
- **Data Wrangling**
- **Data Cleaning**
- **Data Analysis and Visualization**
- **Data Transformation**
- **Model creation**
- **Summary**
- **Conclusion**

Problem Statement
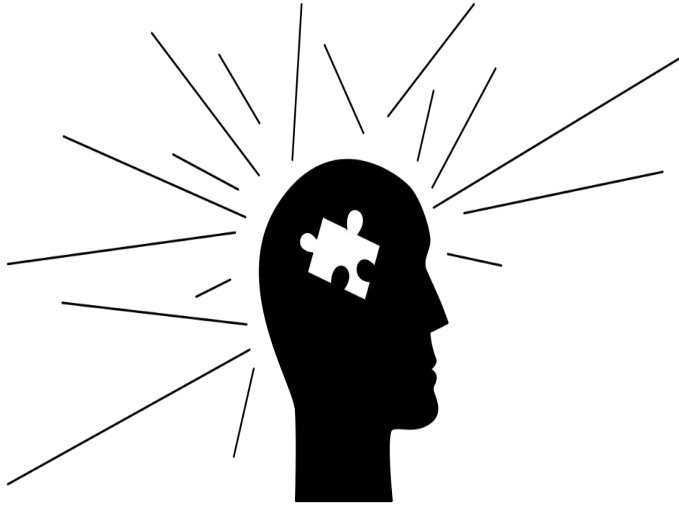
Yes Bank is a well-known bank in the Indian financial domain. Since 2018, it has been in the news because of the fraud case involving Rana Kapoor.

Owing to this fact, it was interesting to see how that impacted the stock prices of the company and whether Time series models or any other predictive models can do justice to such situations and to predict the closing values of the stocks.

KNOWING THE DATASET

In this dataset, we have been provided with data which consists of total **5 columns** and **185 rows** of variables including 'Date', 'Open', 'High', 'Low', 'Close'.

- '**Date**' variable provides the time period of data which is dated from month of July 2005 to November 2020.
- '**Open**' is the price at which the financial security opens in the market when trading begins.
- '**High**' is the highest price at which a stock trades over the course of a trading session.
- '**Low**' is the lowest price at which a stock trades over the course of a trading session.
- '**Close**' is the last price at which a stock trades during a regular trading session.

Data Wrangling

Data Wrangling was carried out on given dataset which involved the process of cleaning, organizing, structuring, and enriching the raw data to make it more useful for analysis and visualization purposes. During this process we changed the format and data type of date feature into datetime.

Date	Open	High	Low	Close
Jul-05	13.00	14.00	11.25	12.46
Aug-05	12.58	14.88	12.55	13.42
Sep-05	13.48	14.87	12.27	13.30
Oct-05	13.20	14.47	12.40	12.99
Nov-05	13.35	13.88	12.88	13.41

Date	Open	High	Low	Close
2005-07-01	13.00	14.00	11.25	12.46
2005-08-01	12.58	14.88	12.55	13.42
2005-09-01	13.48	14.87	12.27	13.30
2005-10-01	13.20	14.47	12.40	12.99
2005-11-01	13.35	13.88	12.88	13.41

Data Cleaning

We analysed the dataset for duplicate values, null values, missing values and it was observed the the data showed absence of any such data which means that there were no duplicate, missing or null values present in the given dataset.

|	
Date	0
Open	0
High	0
Low	0
Close	0
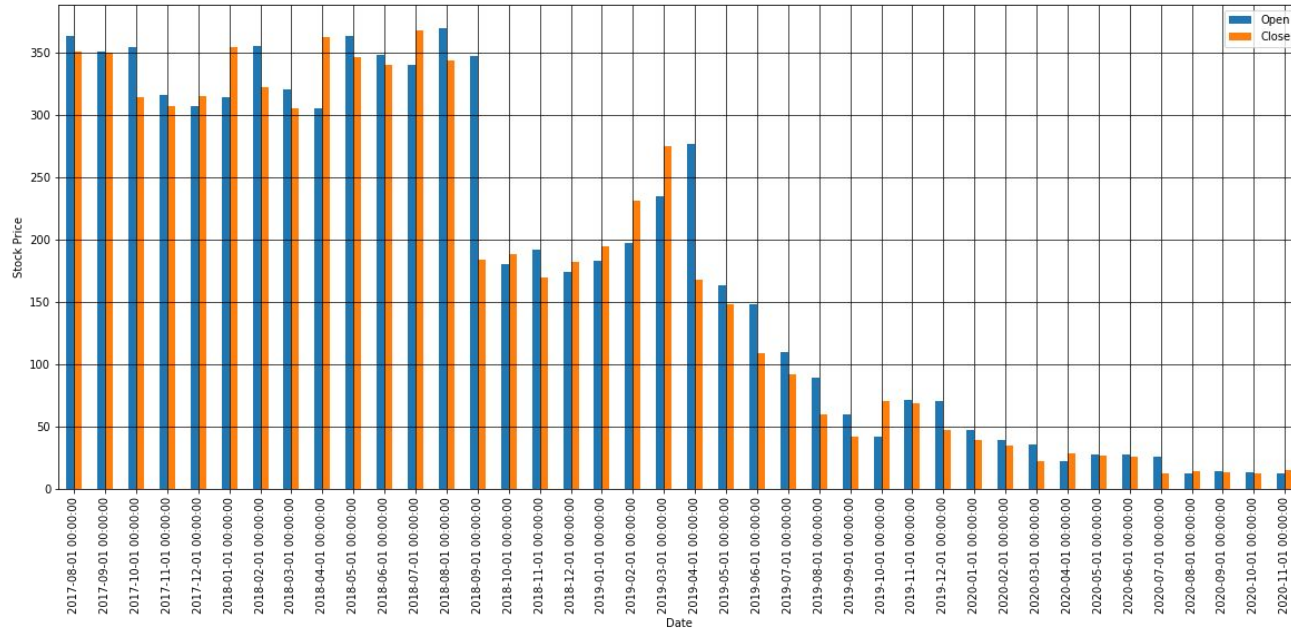
From the above chart we can understand that 6 times opening price had reached its peak of above 350 Rs in the month of August 2017, September 2017, October 2017, February 2018, May 2018, August 2018.

The closing price reached its peak 4 times of above 350 Rs, in the month of August 2017, January 2018, April 2018, July 2018.

The opening and closing prices were at lowest during these 4 months - August 2020, September 2020, October 2020, November 2020.

Data Visualization

Yes Bank opening price
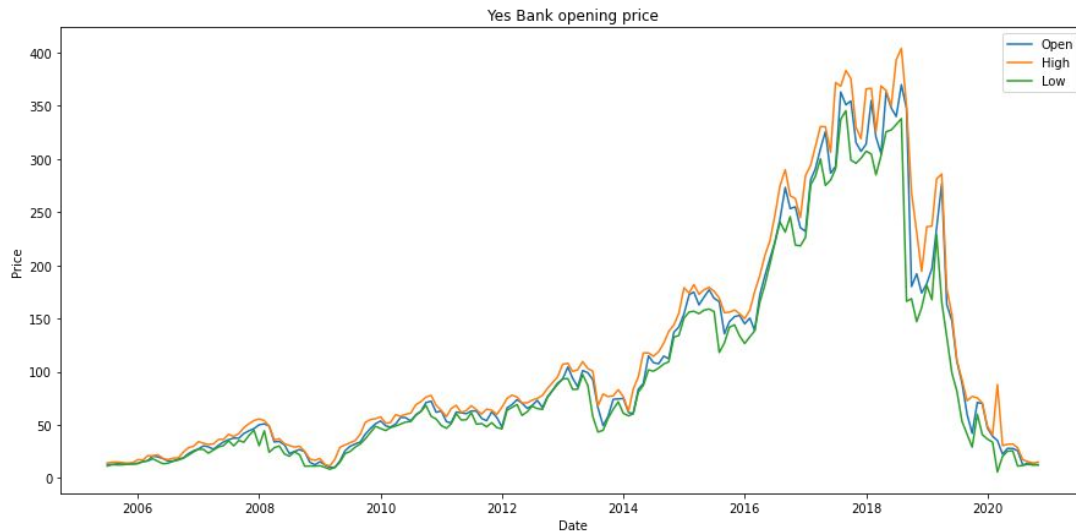
From the above chart we can see that there was a good price rise from July 2005 till 2018 and it was a good time to invest as there was a up trend. The price between 5 Rs - 10 Rs in around July 2005 and reached a peak price of above 350 Rs in the year 2018.

After Rana Kapoor fraud which was done in 2018 the stock price started to fall and there wa a severe down trend, in around 2020 the price reached its July 2005 rate of around 5 Rs - 10 Rs.
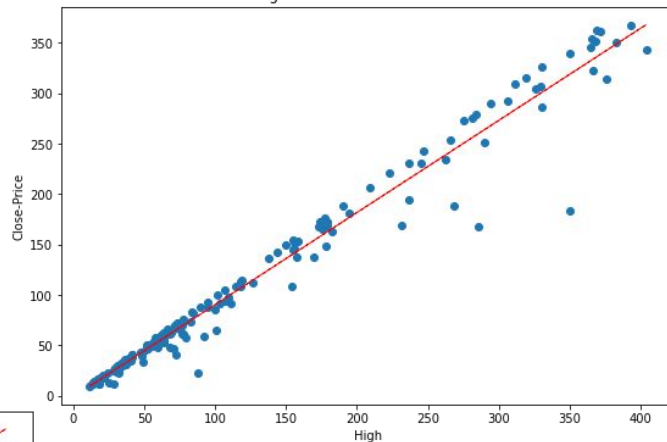
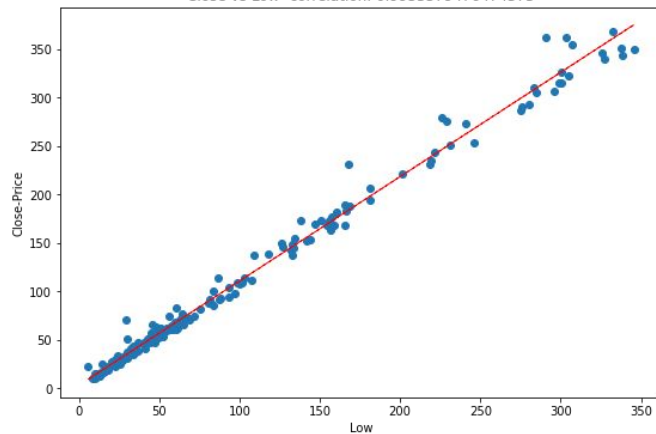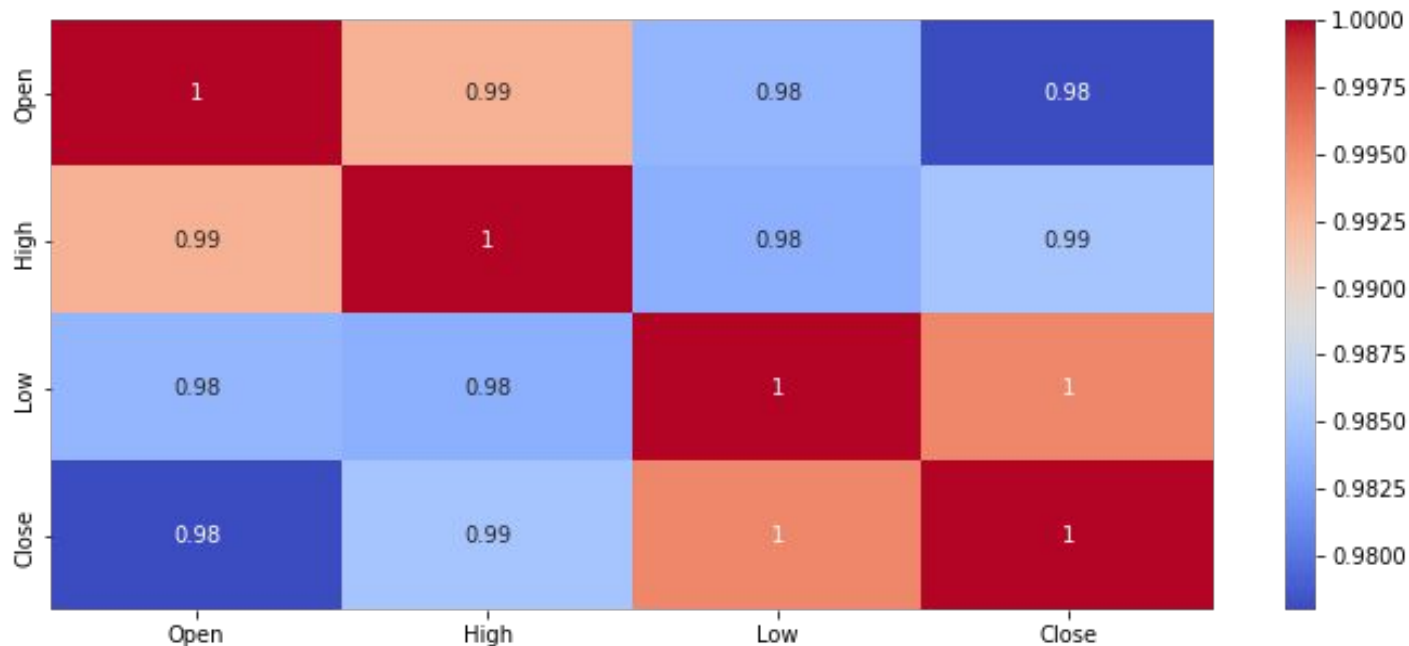# Scatter plot to find correlation between different attributes of given dataset:
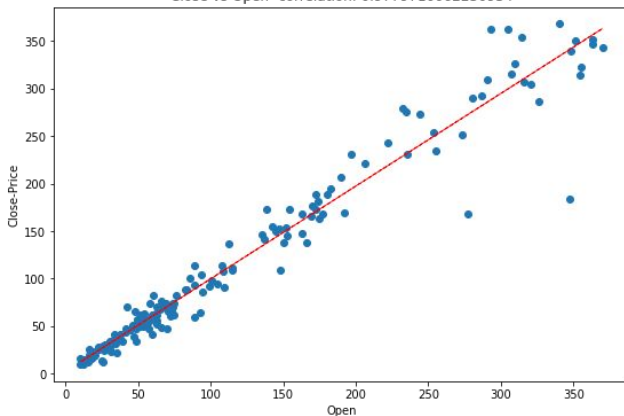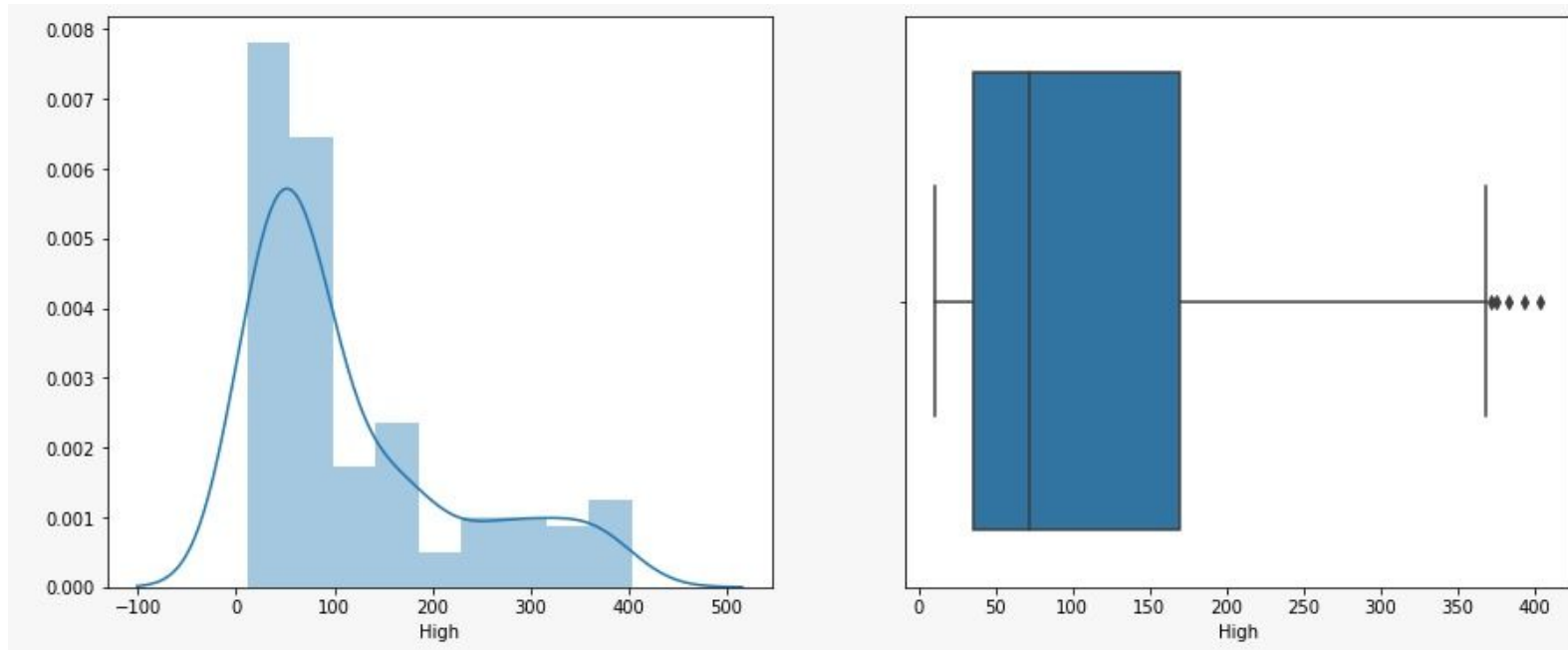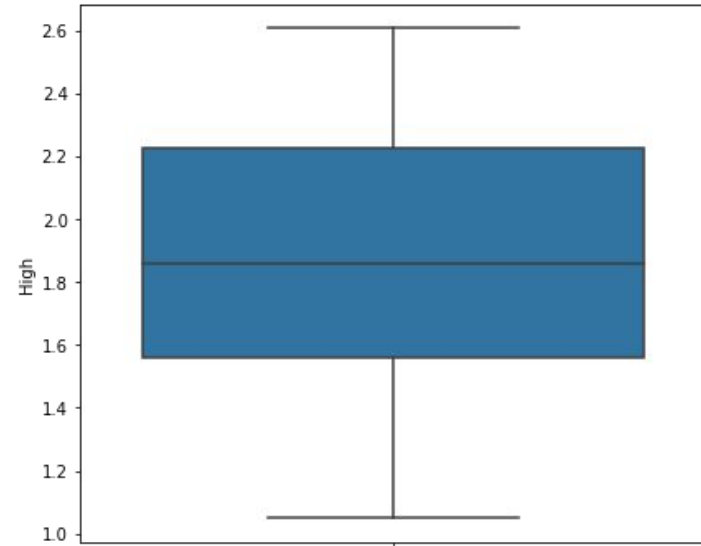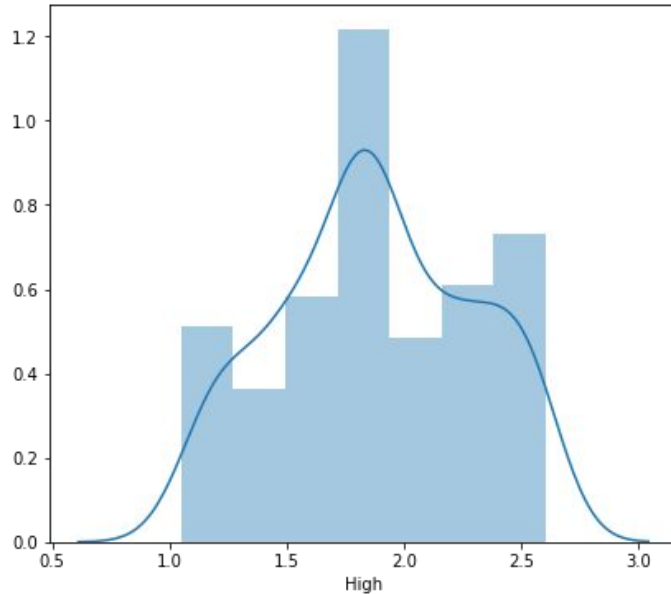
Correlation (Heatmap)

By analysing the heatmap we can say that the relation between low and close is the highest, followed by high and close, and least is open and close.
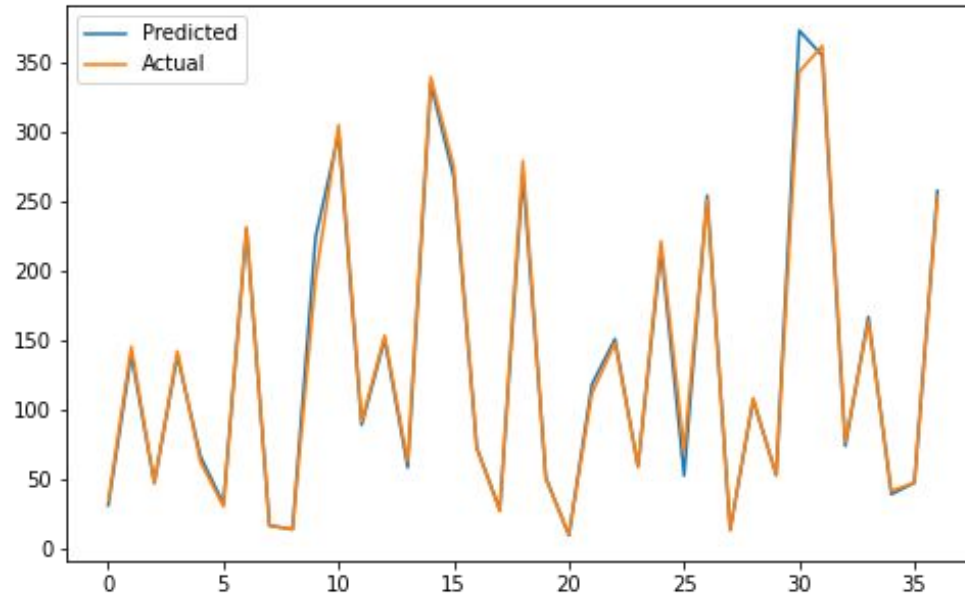
Data Transformation

Data transformation of distribution was carried out using log transformation as the data was moderately right skewed.

After Transformation

After transformation we can see our all dataset features high,low and open are now normally distributed with no outliers as well and are ready for fitting in our model.

Model Creation

Testing accuracy was determined using three different models of which Linear regression and Ridge regression were more accurate and gave 99% as a result, while Lasso regression showed less accuracy of 67%.

Summary

At first we did the data wrangling followed by data cleaning and data transformation and then the Modeling part.

In EDA part of our dataset we conclude that:

- The graphs for Yes BANK opening price and closing price has same result.

- Stock's closing price decreased after year 2018 which was mainly because of Rana Kapoor case which hit the stock price badly.

- The analysis points out that the stock price of the YES BANK falls down after the year 2018 and it is not beneficial for investors to invest their money.

- From scatter plot we can concluded that bivariate analysis shows high correlation of close price with other features.

- All histogram plot shows that all are right skewed.

- From heatmap we can conclude that all the features showing high correlation between each other. We implemented linear regression and the accuracy of our linear regression model is 99.78%.

- For data transformation we used log Transformation because it was moderately right skewed.

After that we visualised the performance of our linear regression model and the graph shows that we achieved the almost best fit model for our dataset.

Conclusion

From the above data analysis we can conclude that the stocks were closing at a high price until the fraud case by Rana Kapoor happened in 2018.

Till then it was a great time period for people to invest in stocks of YES BANK but after that a huge downtrend took place which resulted in decline of the closing prices of the stocks. From our linear regression model and the graph we conclude that it can predict future stock price with almost 99% accuracy.

THANK YOU!!!