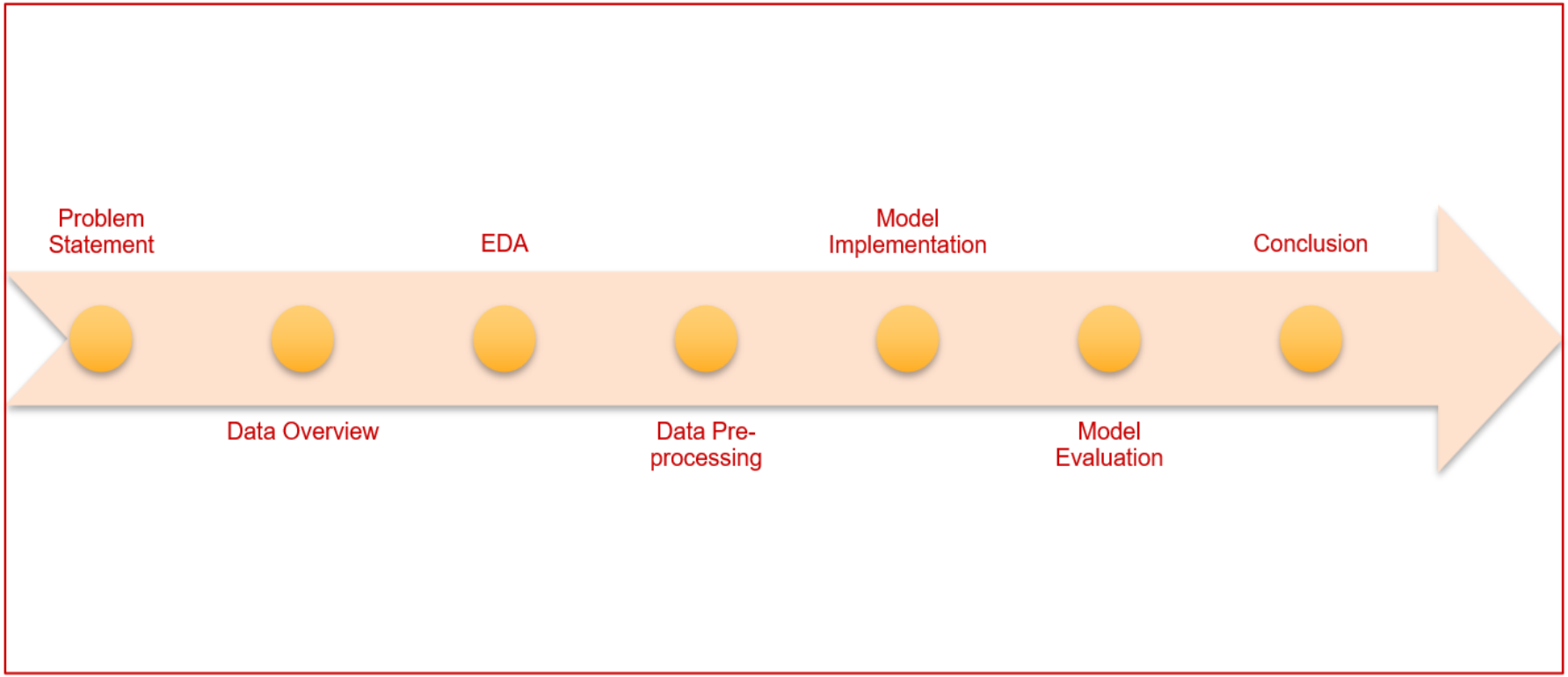
Capstone Project-2

Yes Bank Stock Closing Price Prediction

Prepared By
Akshay Nikam

Points for Discussion



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. This dataset has monthly stock prices of the bank since its inception and includes closing, starting, highest, and lowest stock prices of every month. The main objective is to predict the stock's closing price of the month.

- Perform regression analysis using multiple models and predict the closing price of the stock and in the end will compare the evaluation metrics for all different models to find the best model.

Data Overview

In this dataset We have only 5 columns and 135 rows which includes Date, Open Price, High, Low, Closing Price.

Columns

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

Data Overview

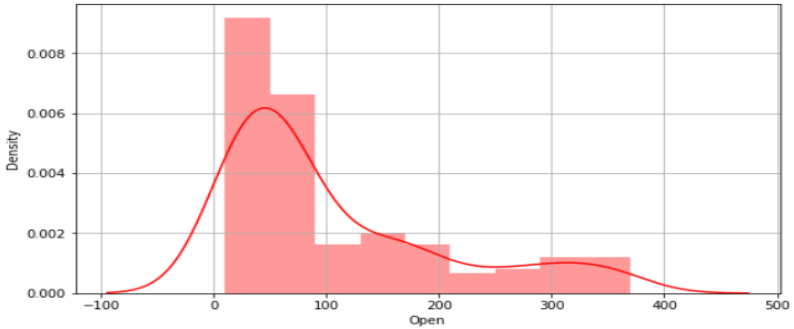
Data Summary:

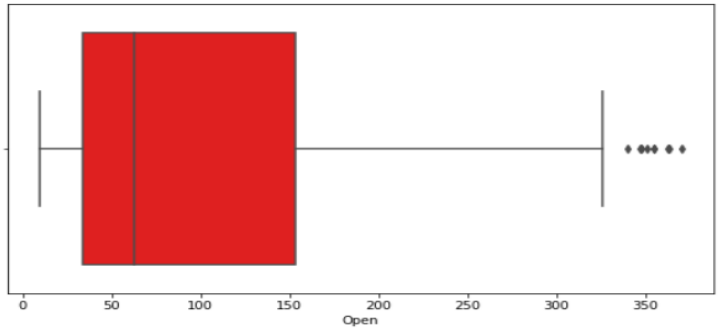
Data Summary

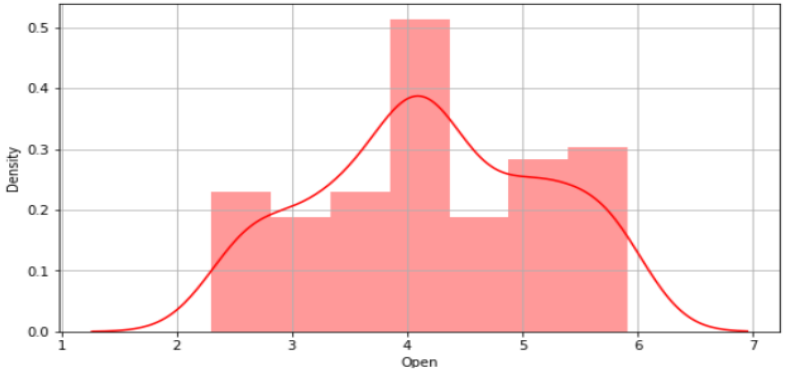
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 185 entries, 0 to 184
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Date    185 non-null    object
 1   Open    185 non-null    float64
 2   High    185 non-null    float64
 3   Low     185 non-null    float64
 4   Close   185 non-null    float64
dtypes: float64(4), object(1)
memory usage: 7.4+ KB
```

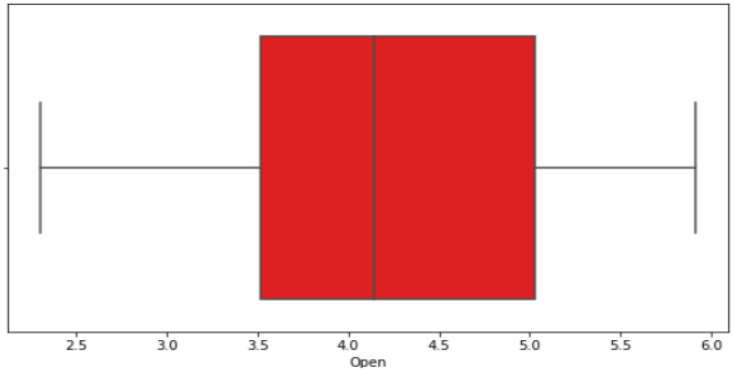
EDA

Univariate Analysis : Independent variable (Open)



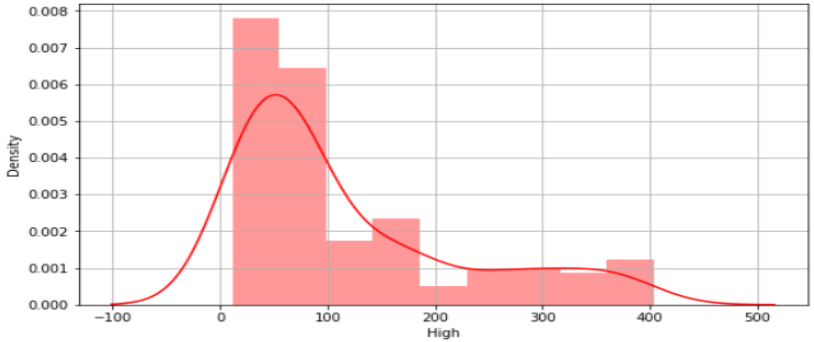


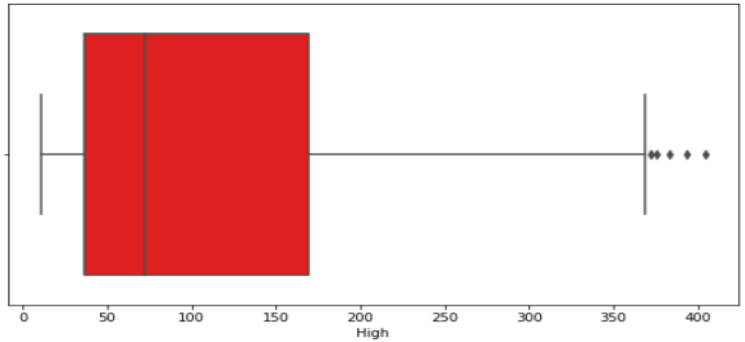


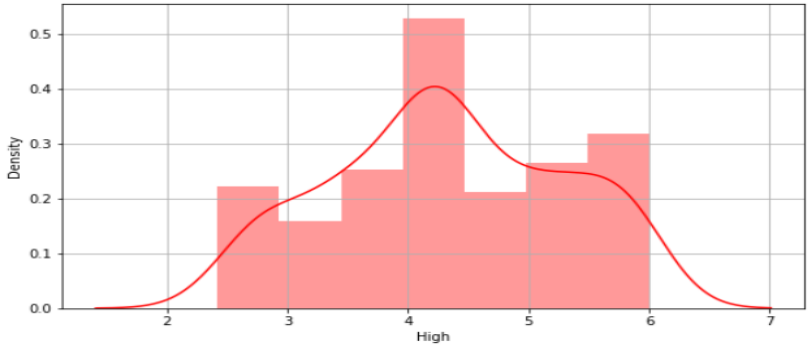


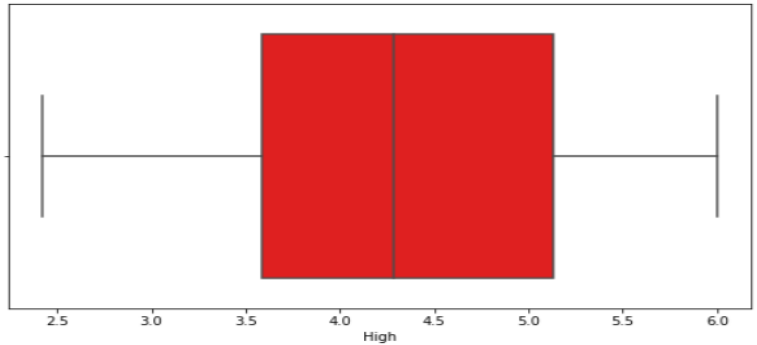
EDA

Univariate Analysis : Independent variable (High)



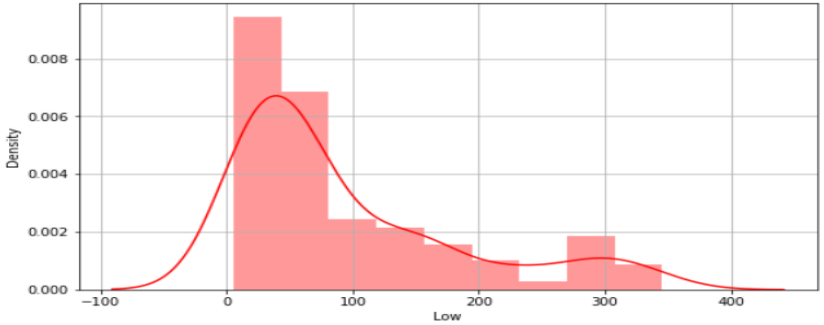


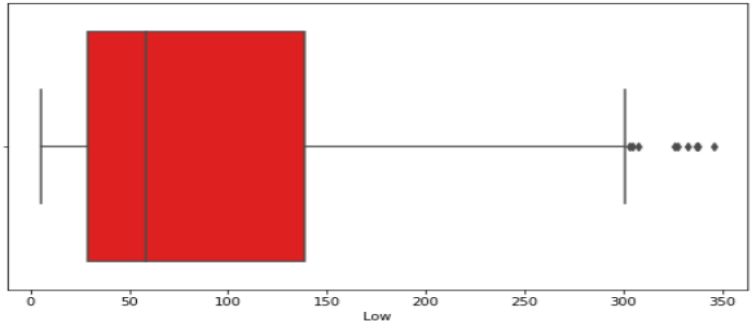


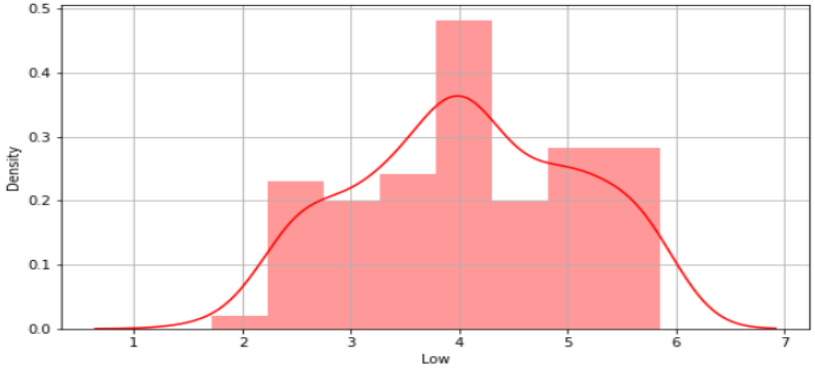


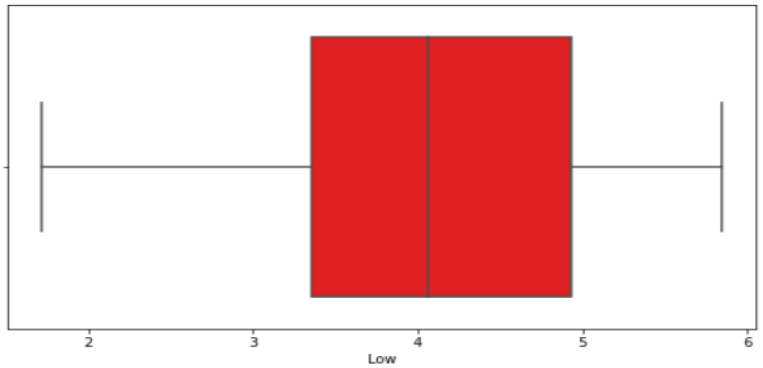
EDA

Univariate Analysis : Independent variable (Low)



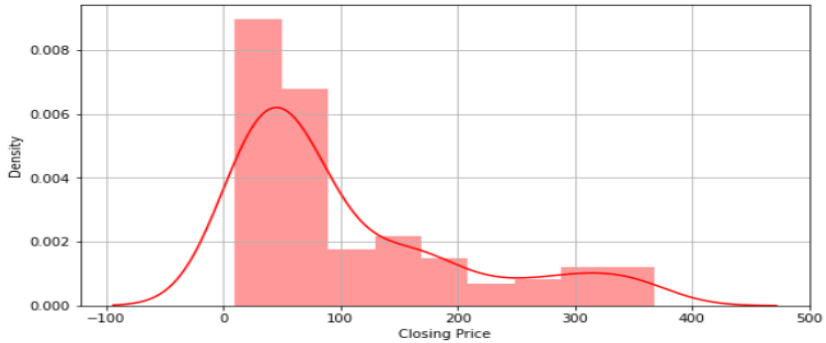


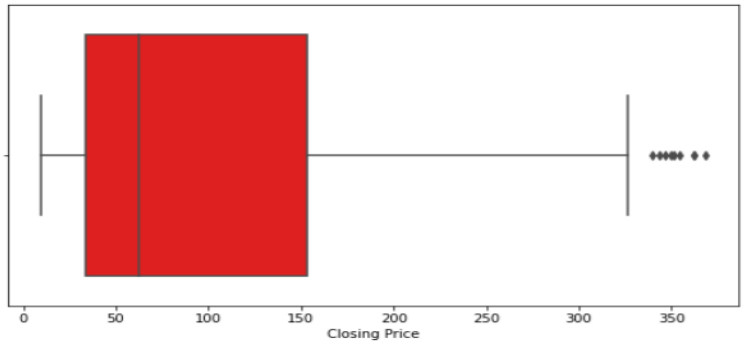


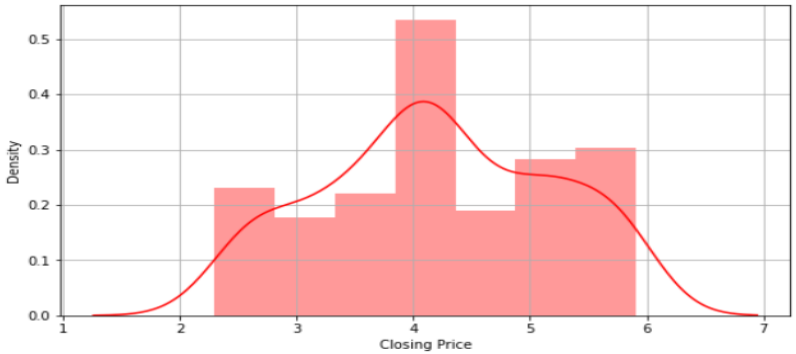


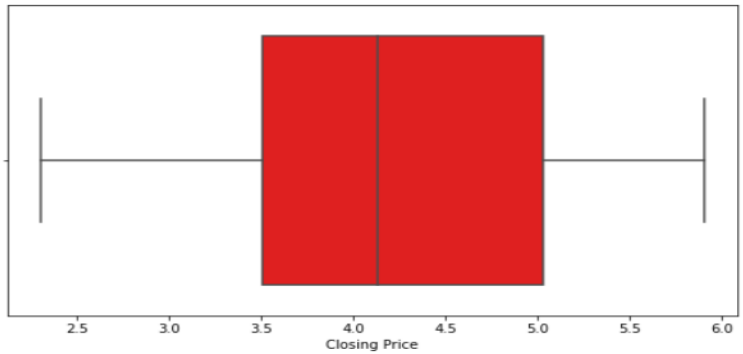
EDA

Univariate Analysis : dependent variable (Closing Price)







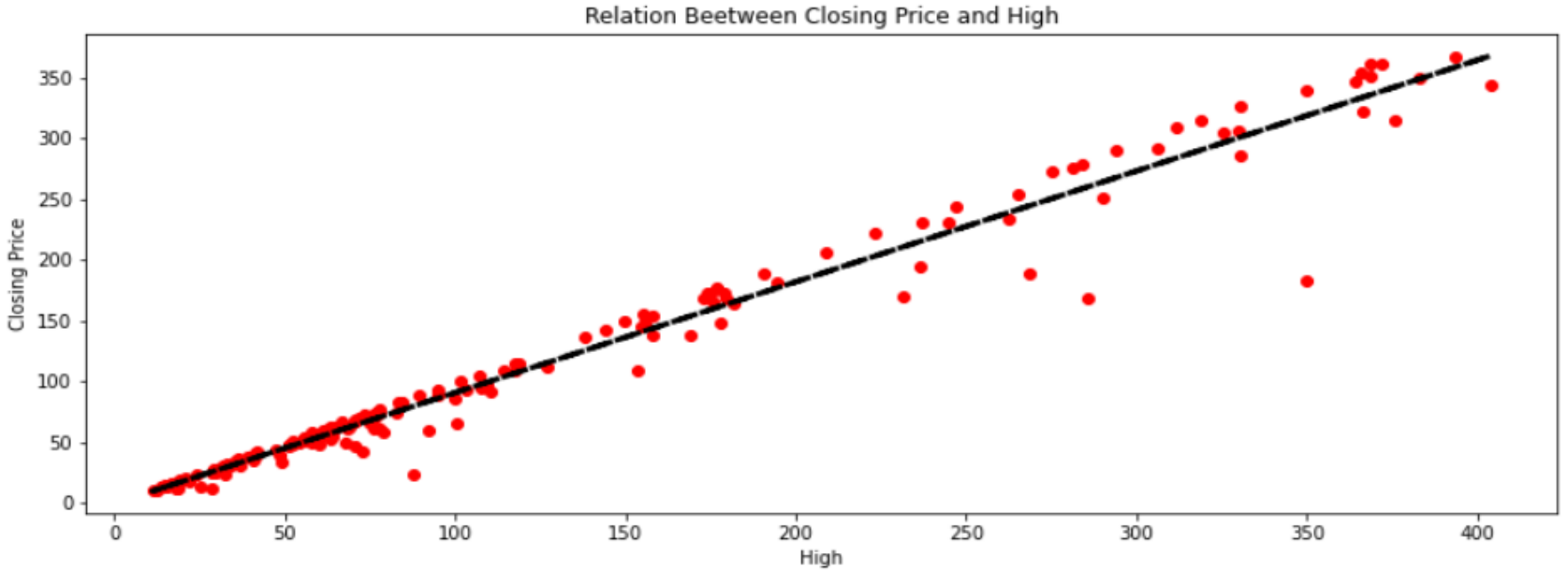


EDA

Bivariate Analysis : Relation of Opening price with Closing Price

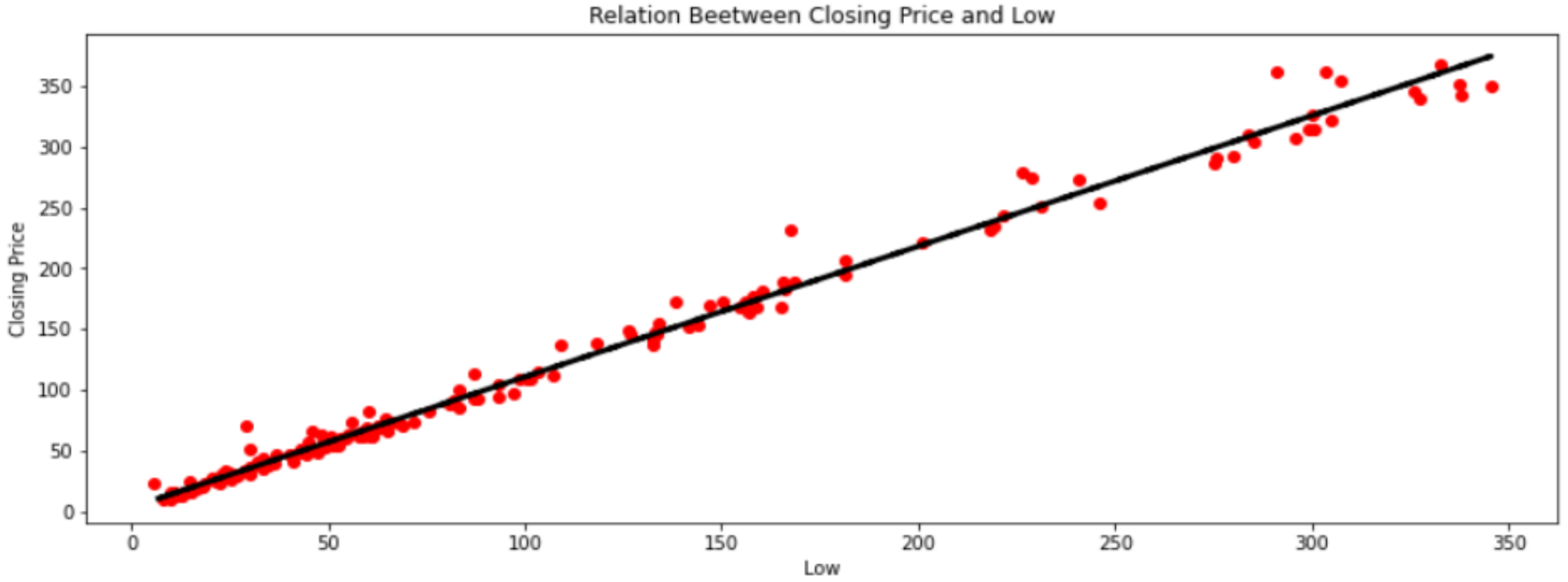
EDA

Bivariate Analysis : Relation of High price with Closing price.



EDA

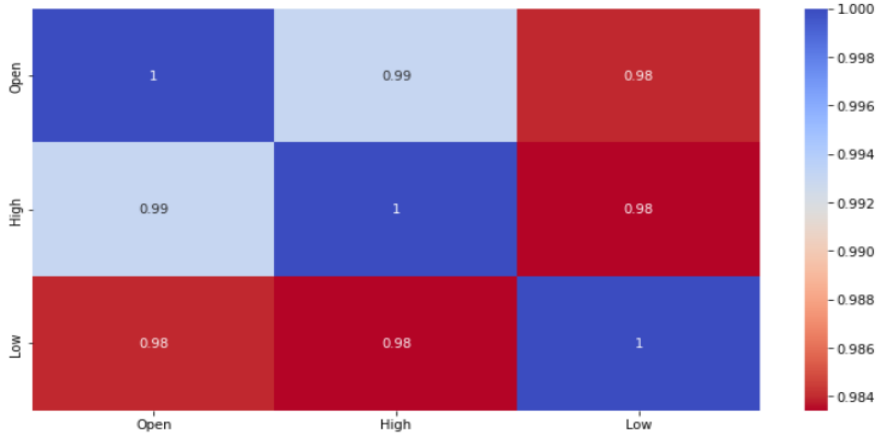
Bivariate Analysis : Relation of Low price with Closing price.



Data Pre-processing

Correlation and VIF analysis :

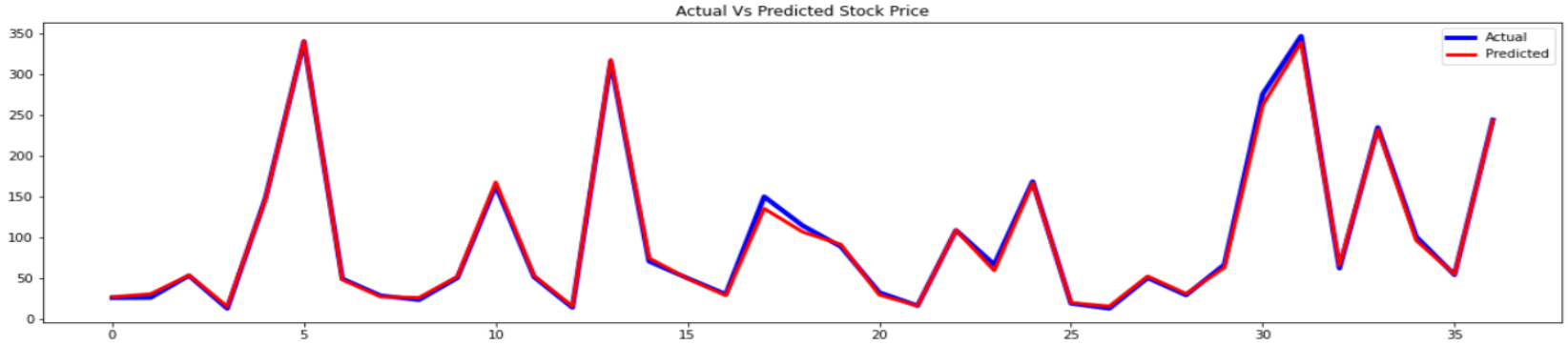
	variable	vif
0	Open	175.185704
1	High	167.057523
2	Low	71.574137

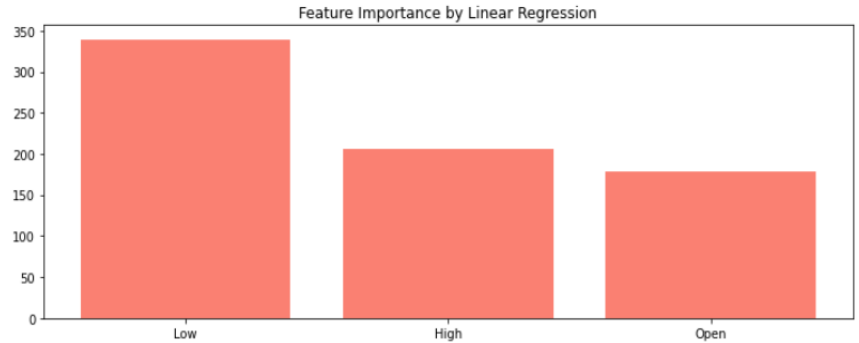


- As we can see that VIF scores are very high for all our features which means there is a multicollinearity between our independent variables.
- also we know that 'Open', 'High', 'Low' all three variables are important for predictions for that reason I am not performing any feature engineering here in this case.

Model Implementation

Linear Regression:

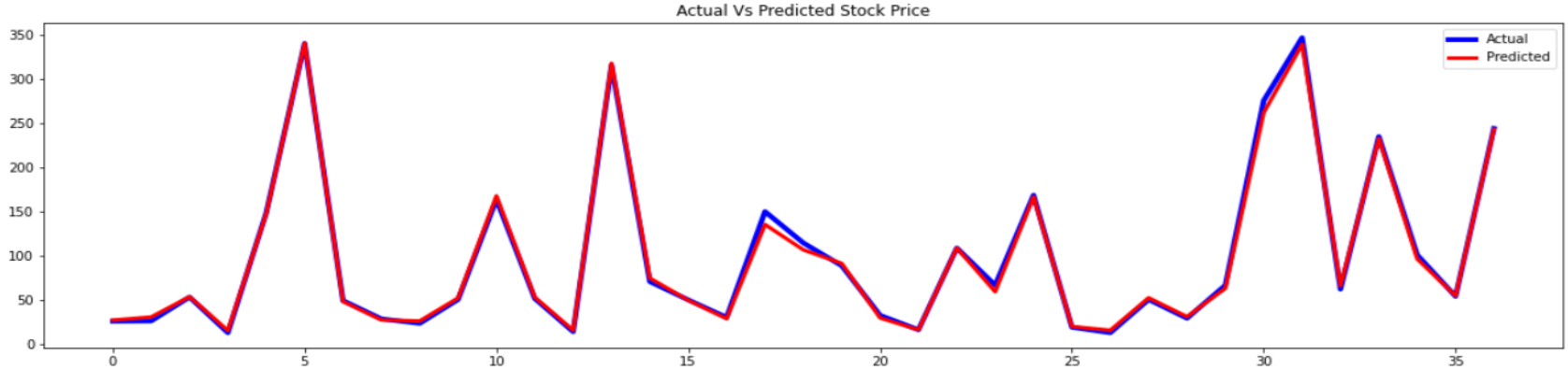


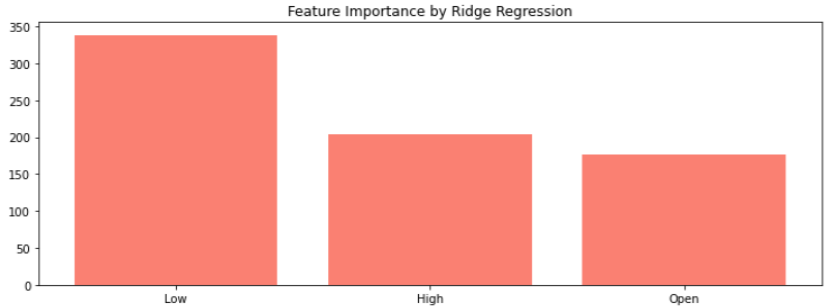


- MSE : 19.988578593595
- RMSE : 4.470858820584139
- R_squared : 0.9978412541225983

Model Implementation

Ridge Regression:

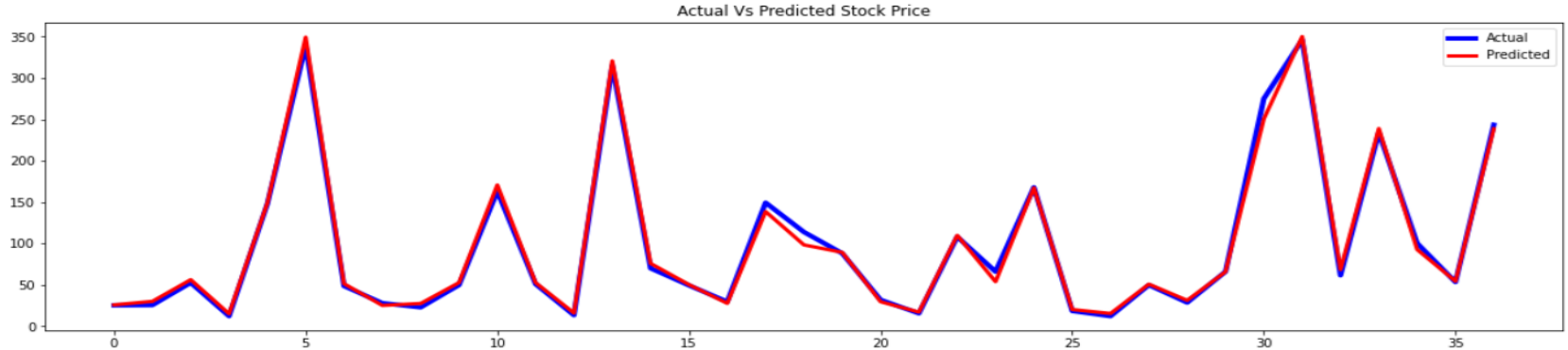


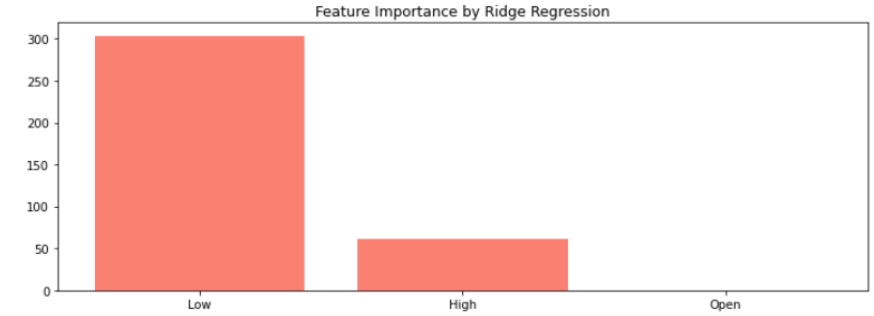


- MSE : 20.095425485603744
- RMSE : 4.482792152844446
- R_squared : 0.9978297147684337

Model Implementation

Lasso Regression:

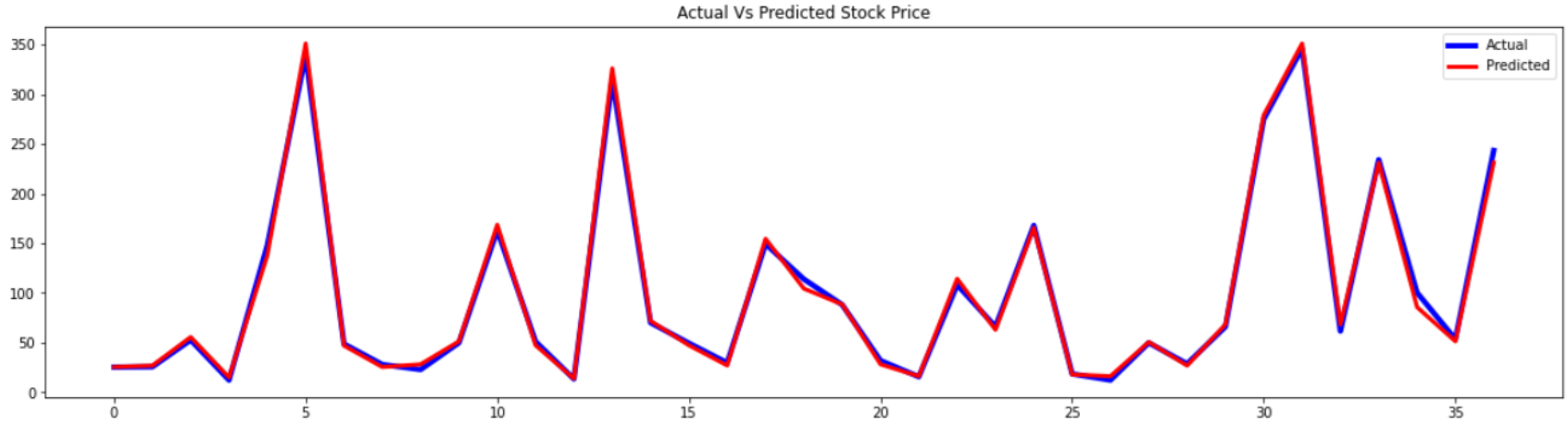




- MSE : 43.955793347527035
- RMSE : 6.629916541520491
- R_squared : 0.995252819642348

Model Implementation

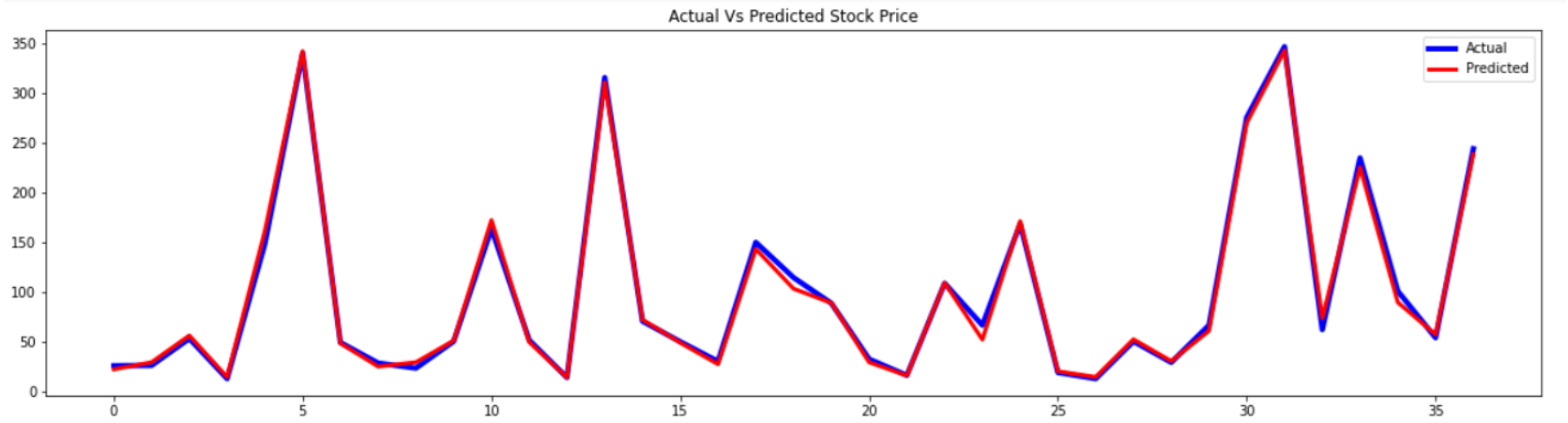
K-Nearest Neighbors Regression:



MSE : 30.80357567567563
RMSE : 5.550096906872494
R_squared : 0.9966732455893385

Model Implementation

XG-Boost Regression:

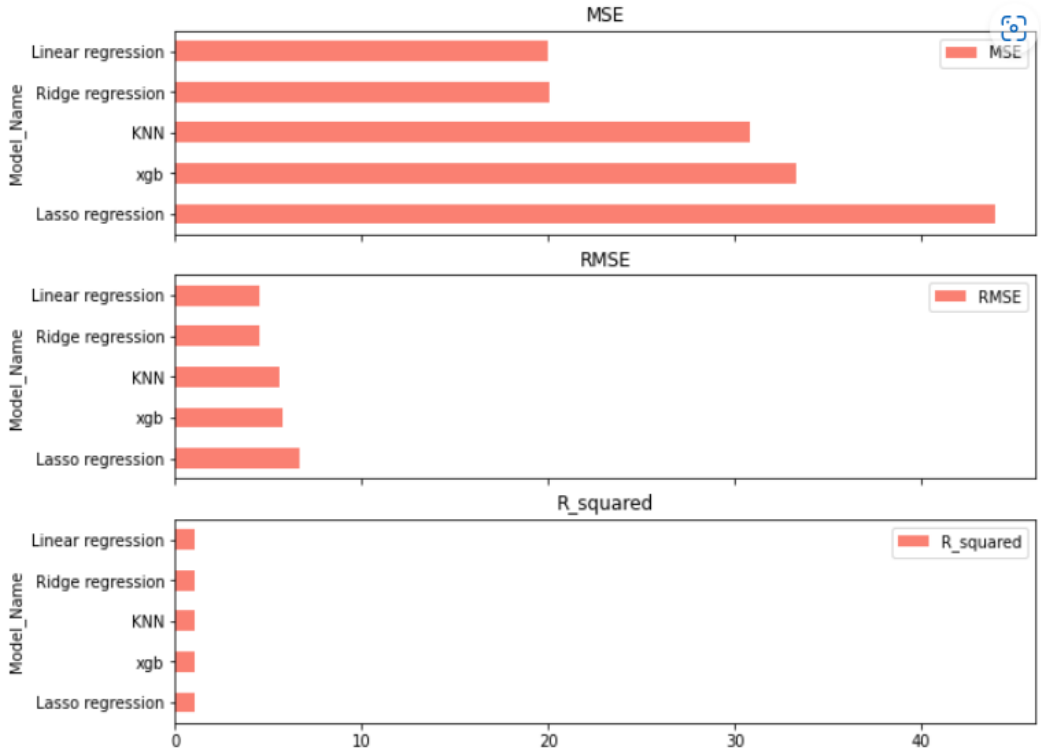


- MSE : 33.318703967055924
- RMSE : 5.7722356125729934
- R_squared : 0.9966732455893385

Model Evaluation

Model Evaluation :

	Model_Name	MSE	RMSE	R_squared
2	Lasso regression	43.9558	6.6299	0.9953
4	xgb	33.3187	5.7722	0.9967
3	KNN	30.8036	5.5501	0.9967
1	Ridge regression	20.0954	4.4828	0.9978
0	Linear regression	19.9886	4.4709	0.9978



Conclusion

Conclusion :

- All the input variables are highly contributing for output prediction.
- The accuracy score for each model is more than 95%.
- Linear Regression has given the best results with lowest MSE and RMSE and highest R-square value scores out of other three algorithms because of linear relation of dataset.
- KNN and XG Boost not performed well with high MSE and RMSE score.
- We see lasso with low performance because it penalizes the coefficient and thus shrinks them to zero. whereas in this case all the features were important for prediction purpose so, it ended up with poor results.

Thank You