For

1CS201

INTRODUCTION TO AI & ML

B. Tech. Semester II

Report Prepared By

***VEDANT THUMBADIYA(23BTM061)***

***SHYAM RANIPA(23BCE325)***

School of Engineering

Institute of Technology

Nirma University

Ahmedabad

382481

**INDEX**

**1.** Introduction(Description of Topic)

**2.** Code

**3.** Output

**4.** Conclusion

**INTRODUCTION**

**Selection of data domain:**

The dataset selected for analysis of Tatasteel . This dataset aims to predict the type of symbol(TISCO or TATASTEEL) on the basis of series, prev close, open, high, turnover, volume and many more

**Selection of Fields(Data types):**

The fields in this dataset consist of a mix of string and numerical data types:

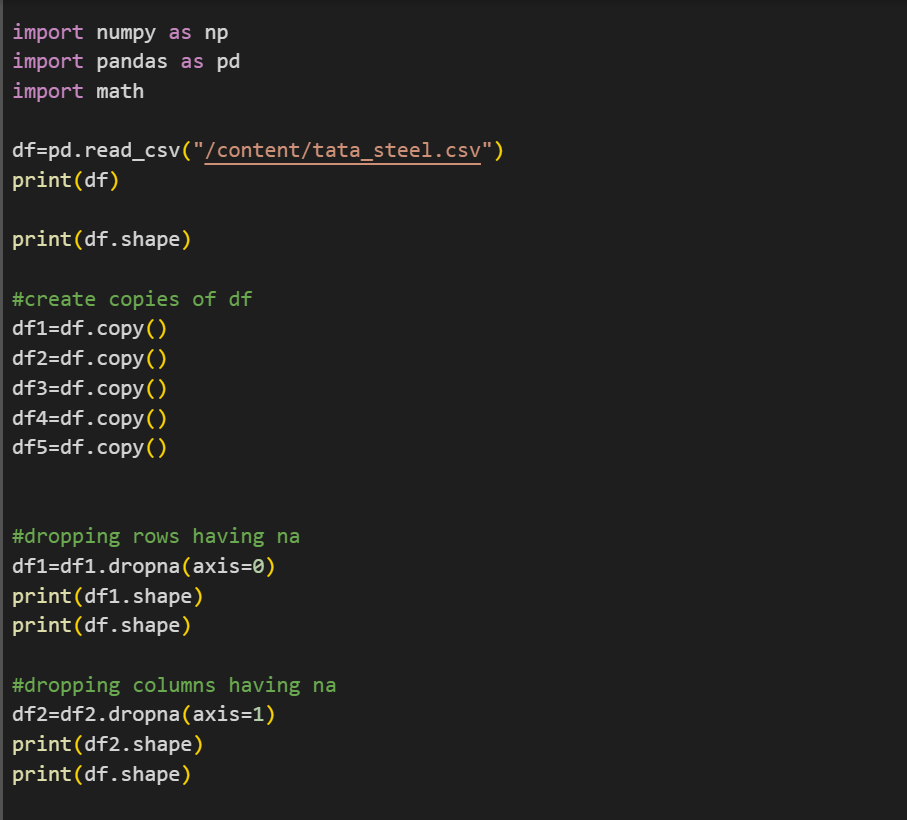
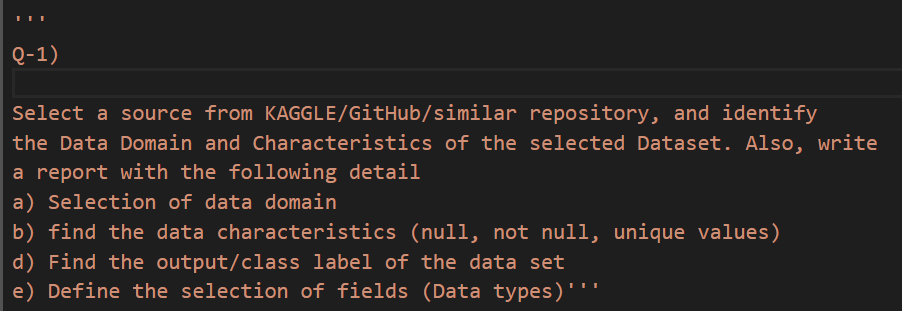
● Numerical variables:

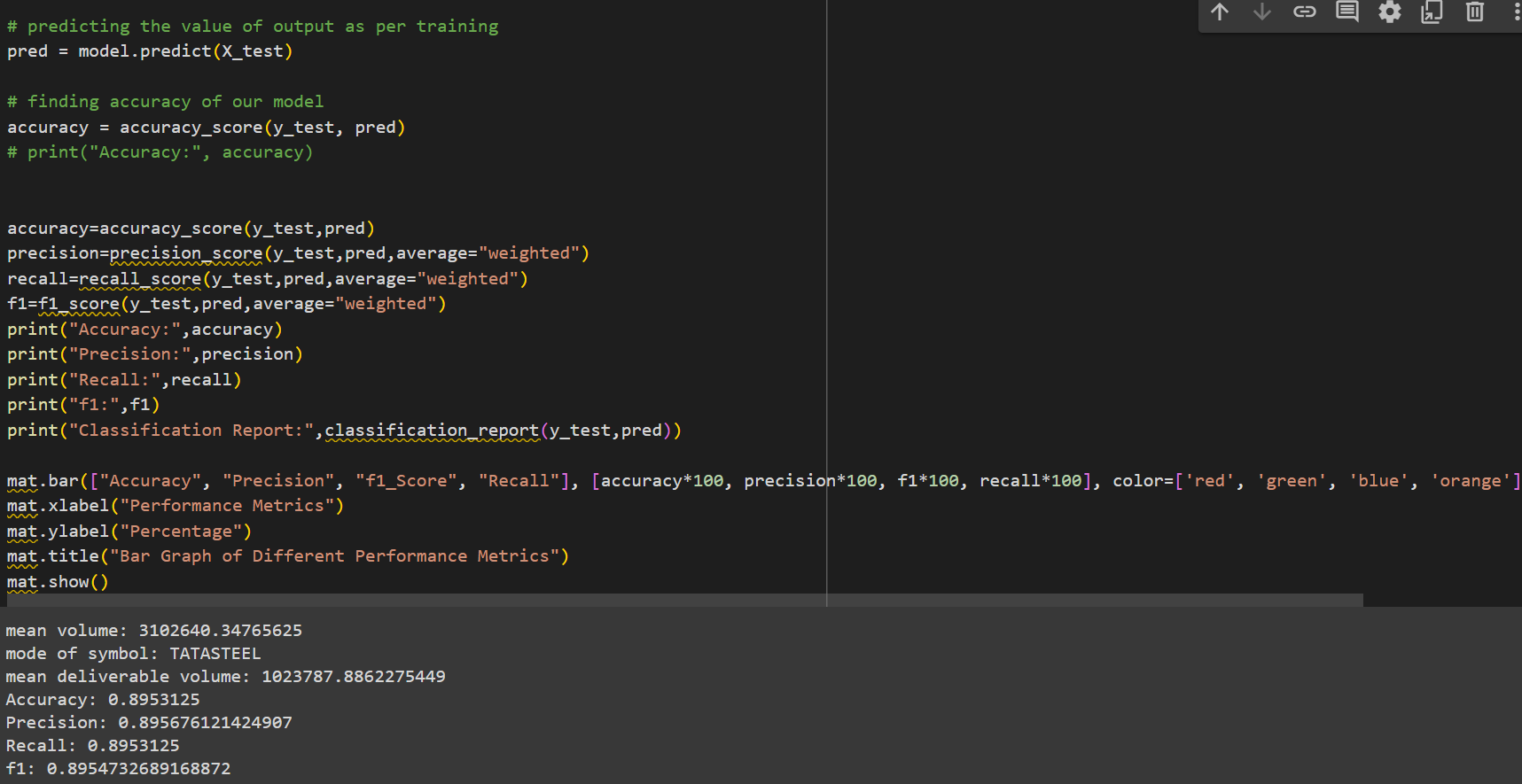
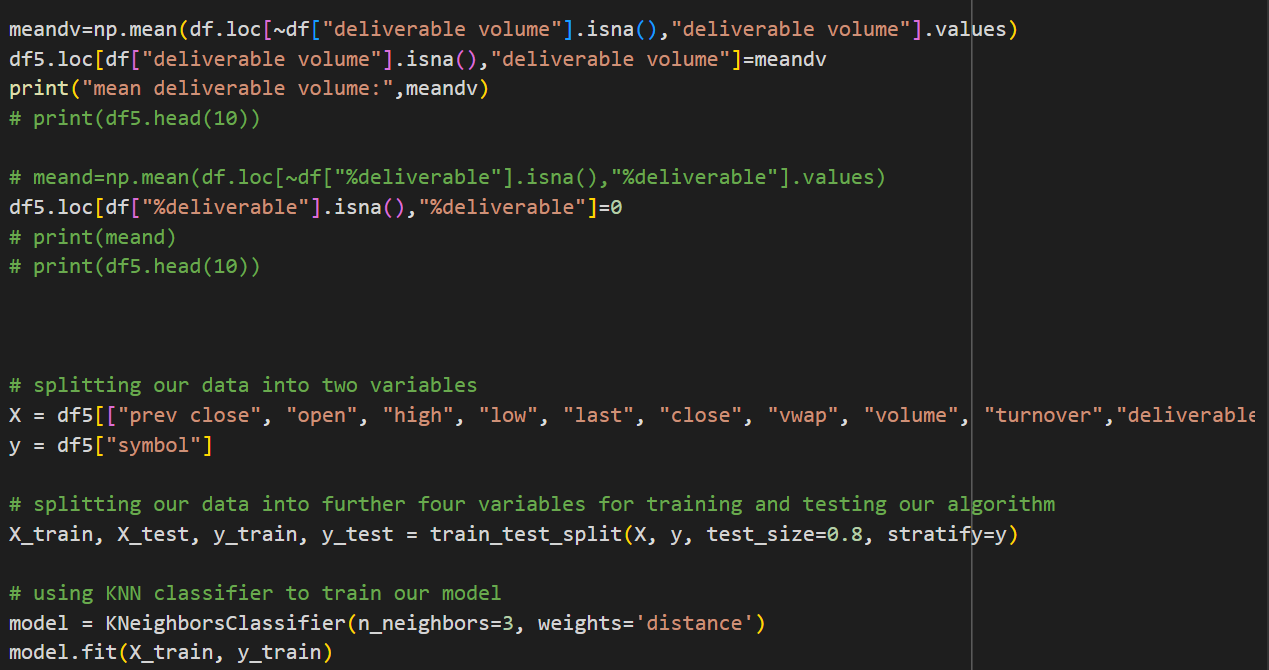
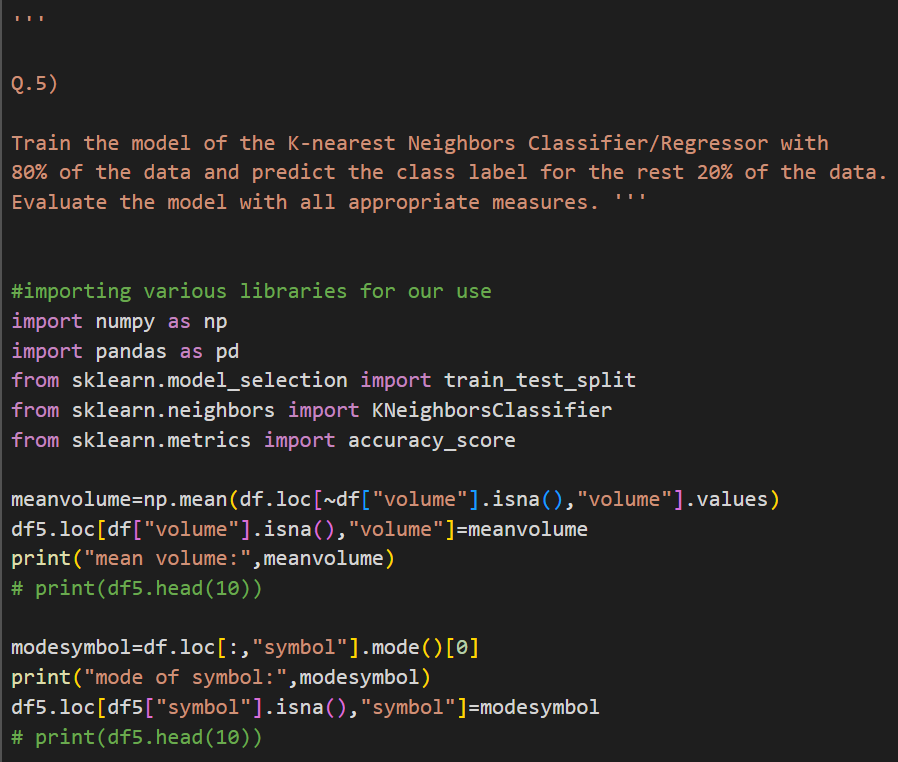
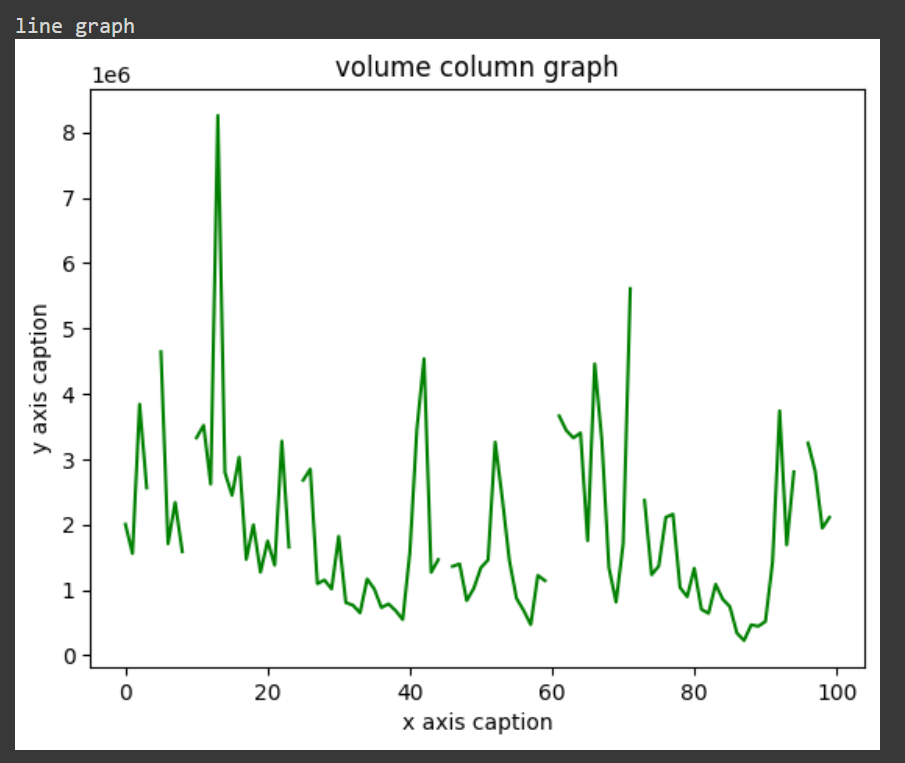
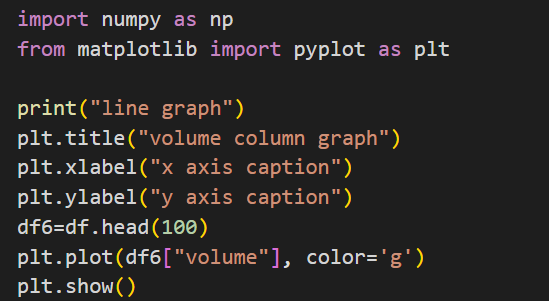
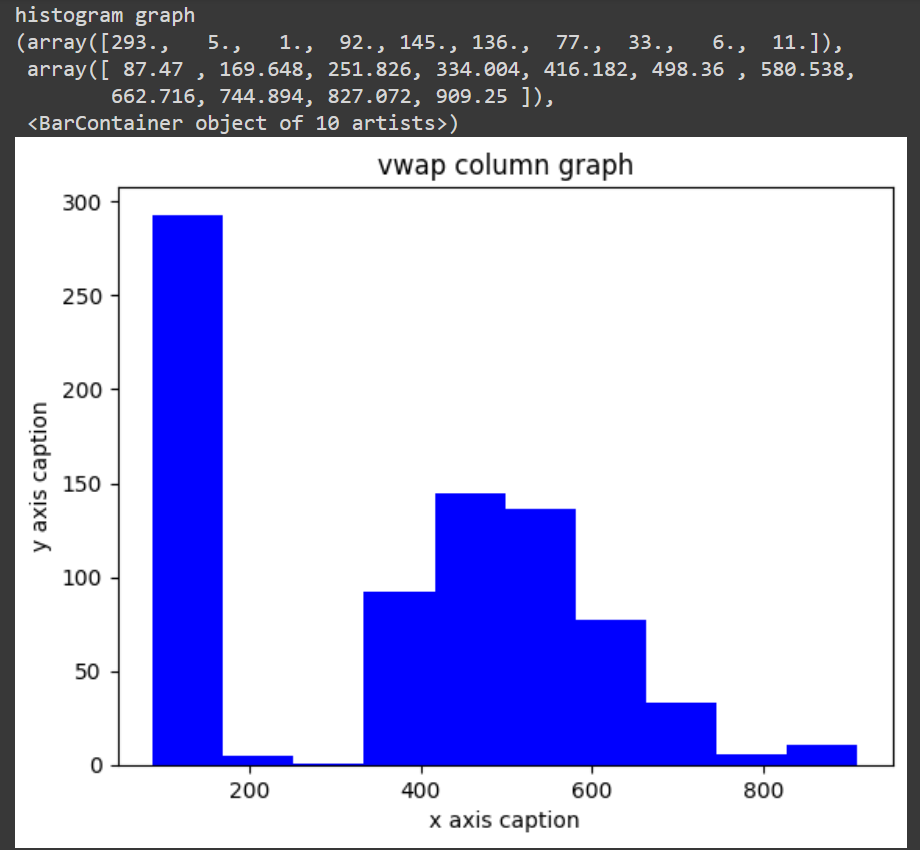
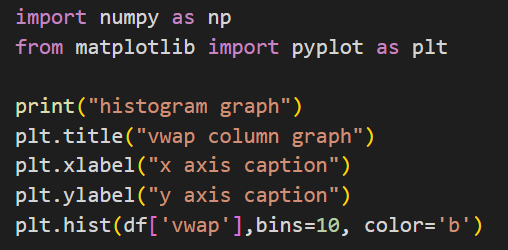
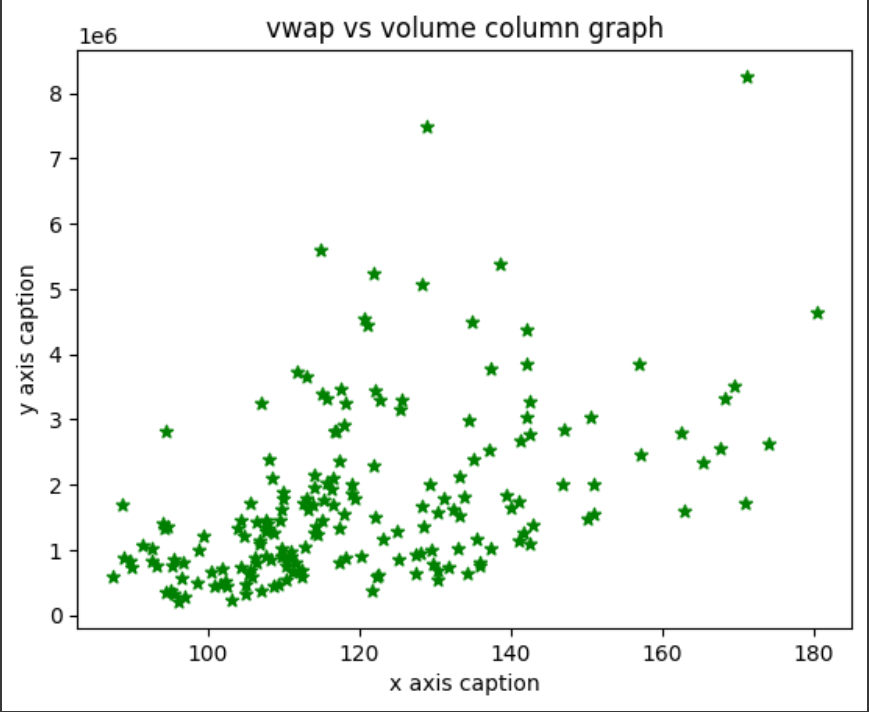
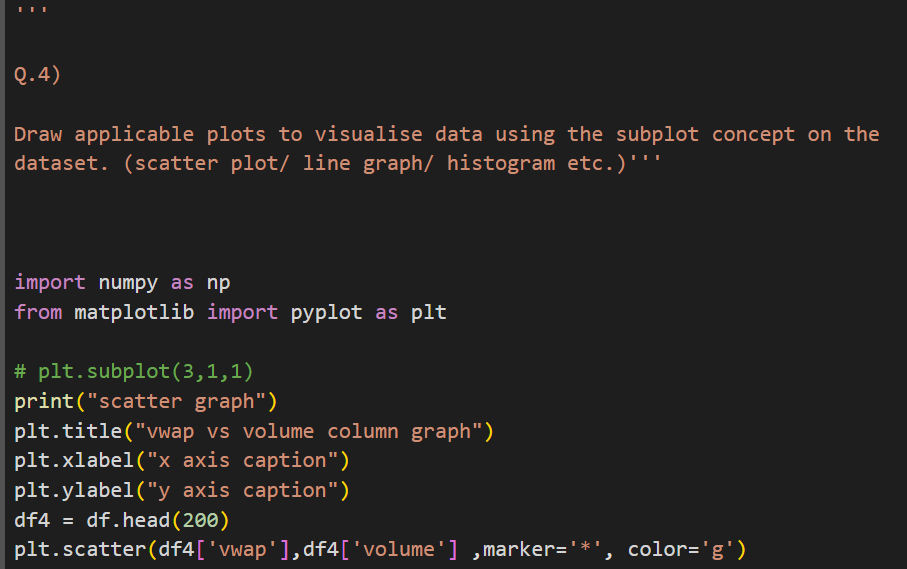
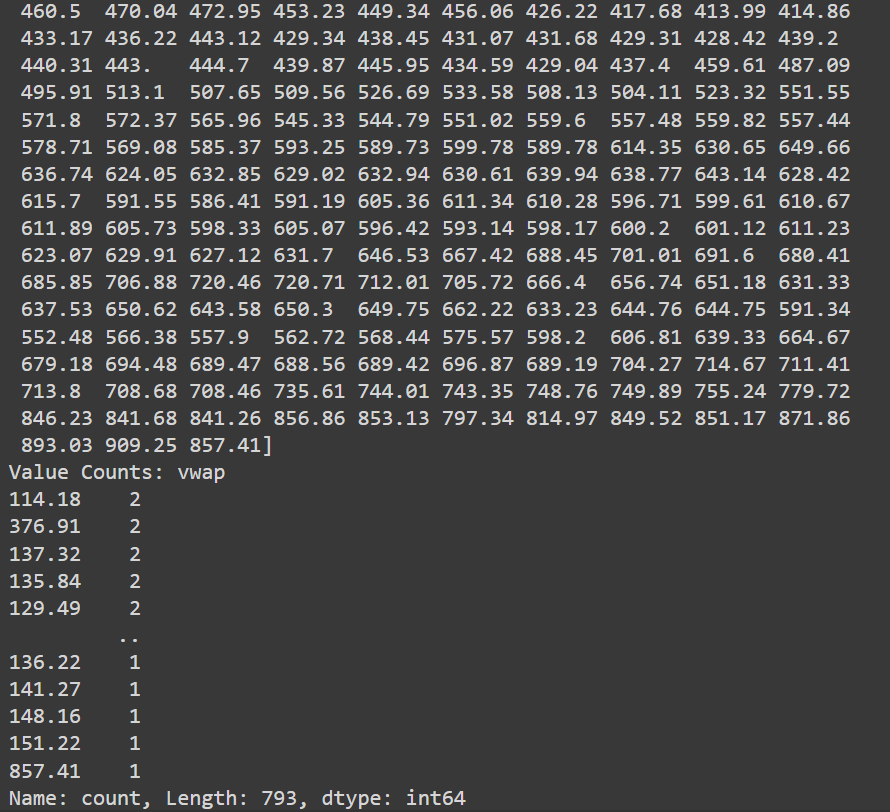
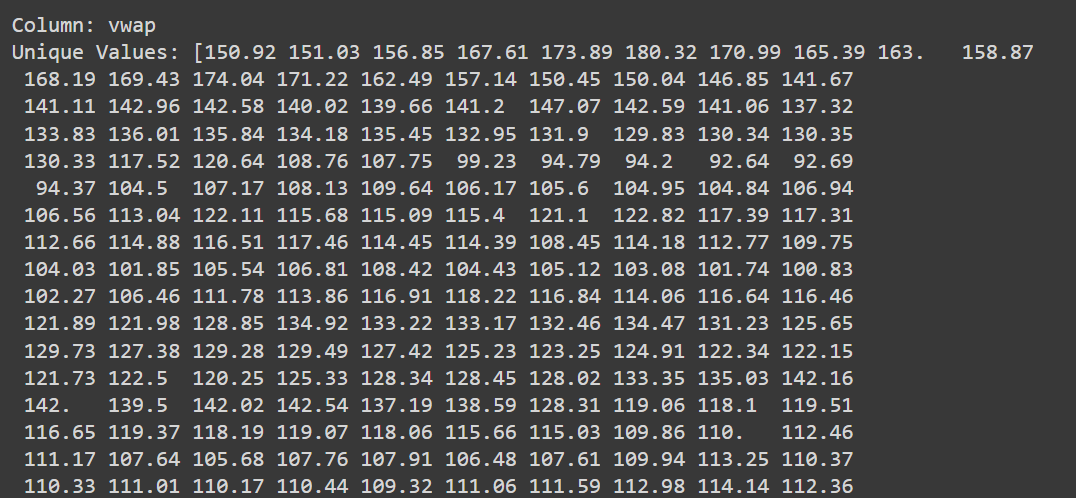
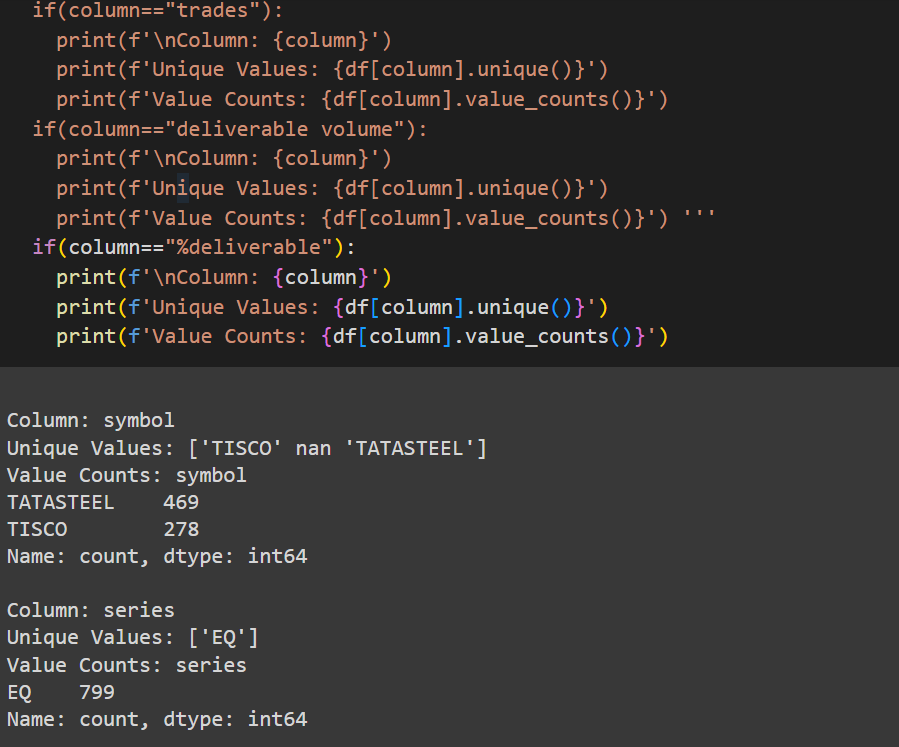
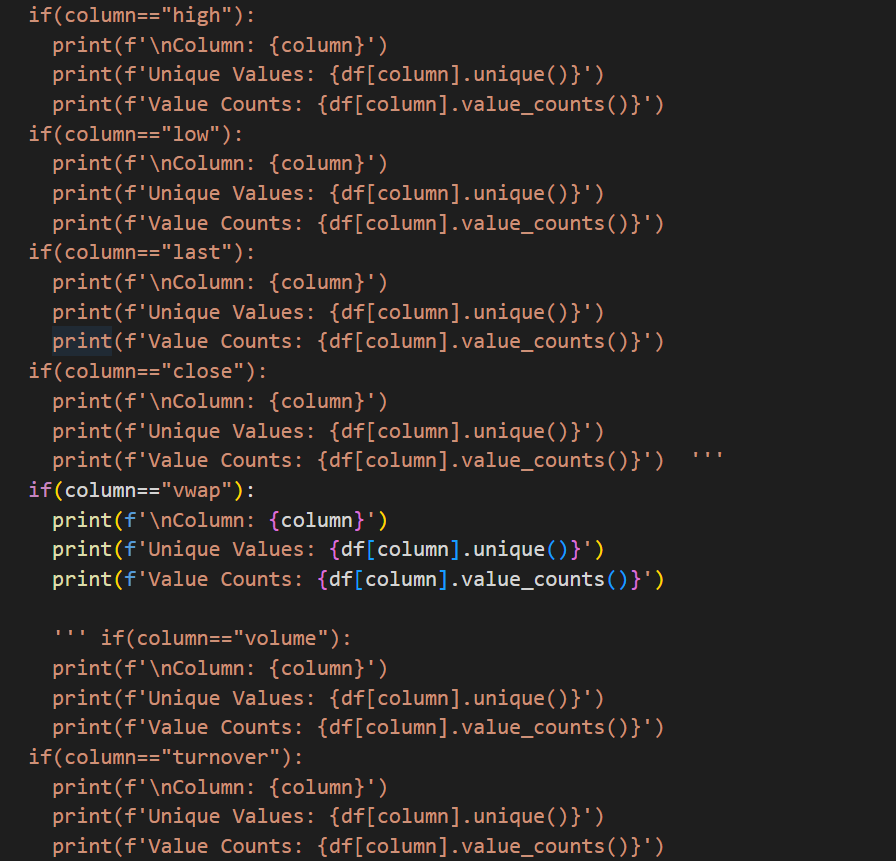
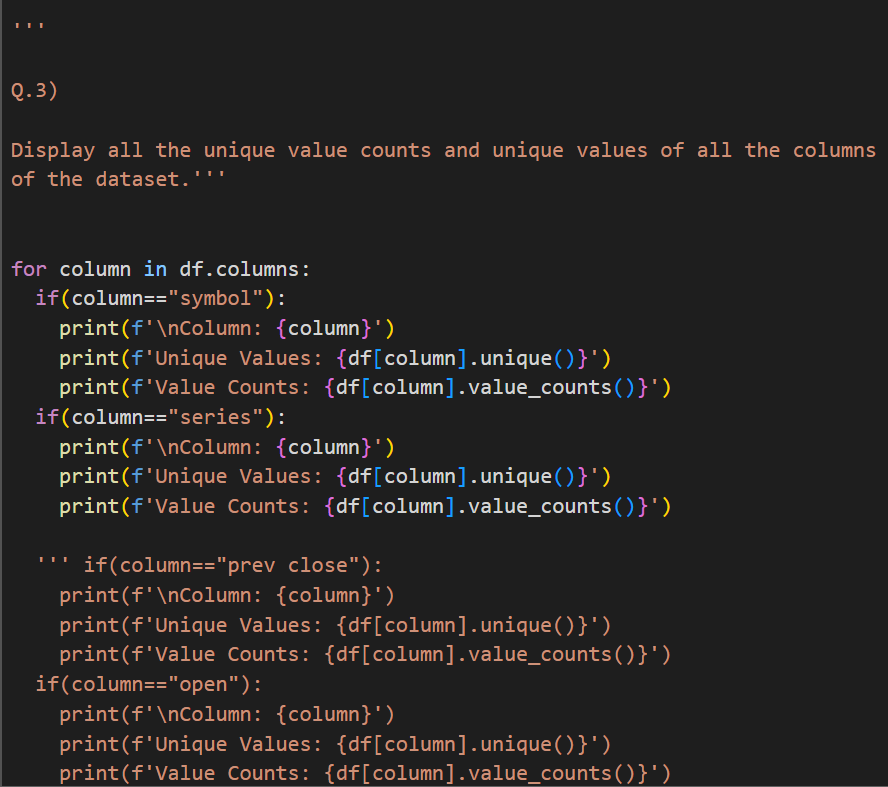
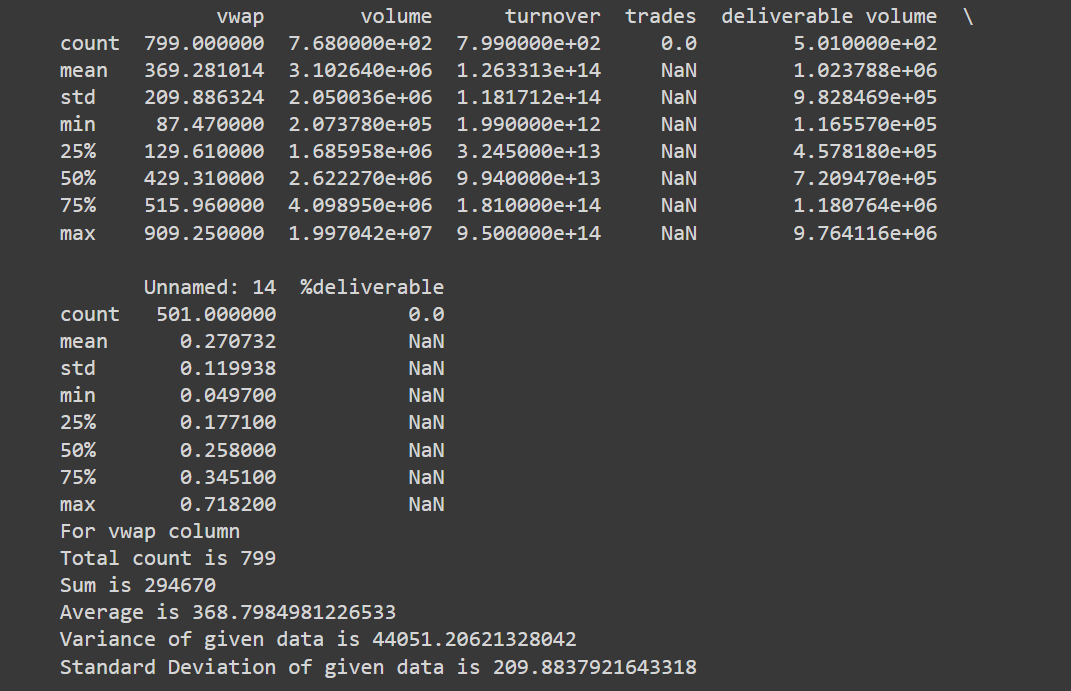
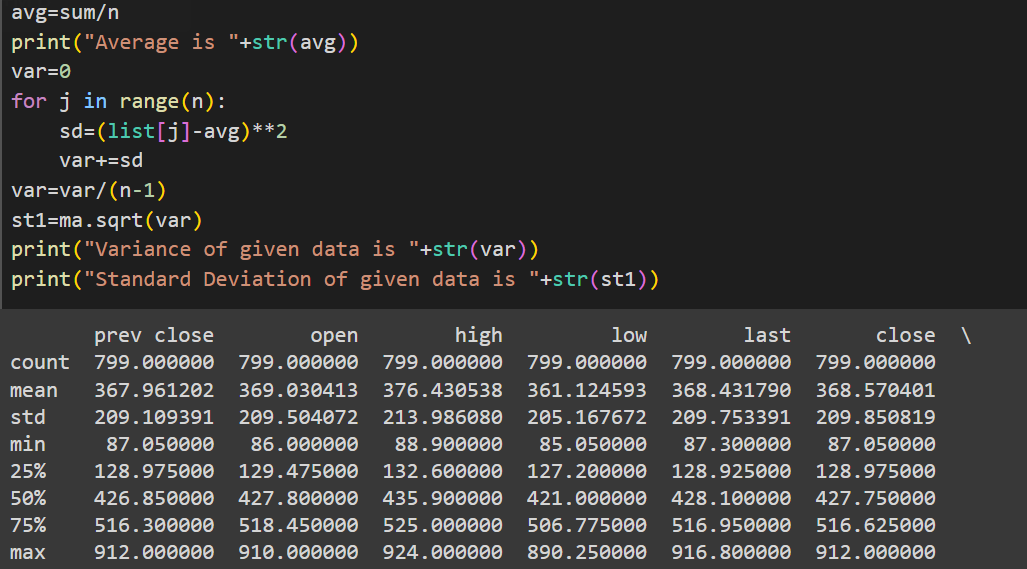
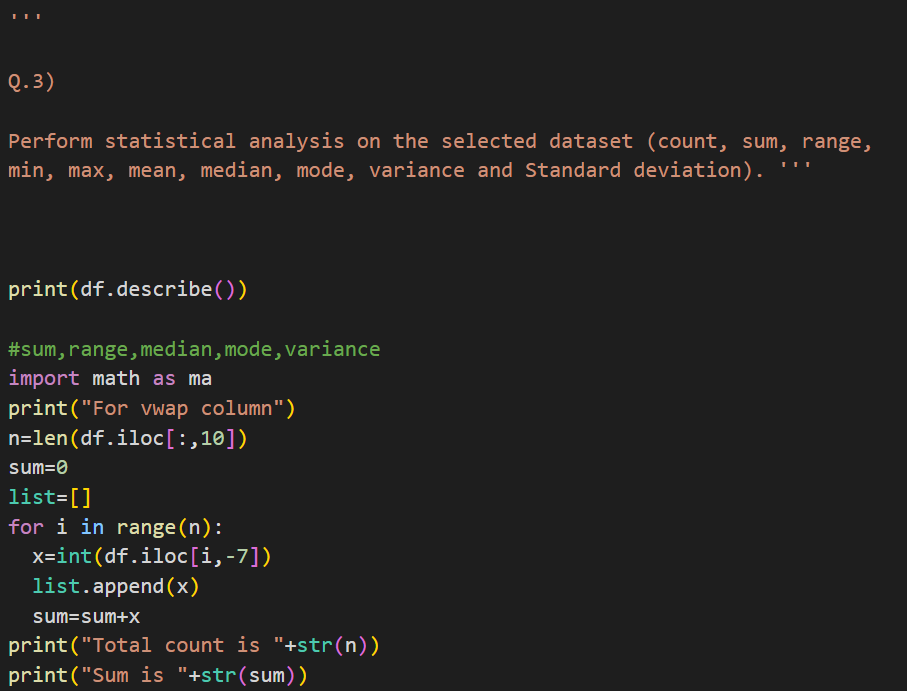
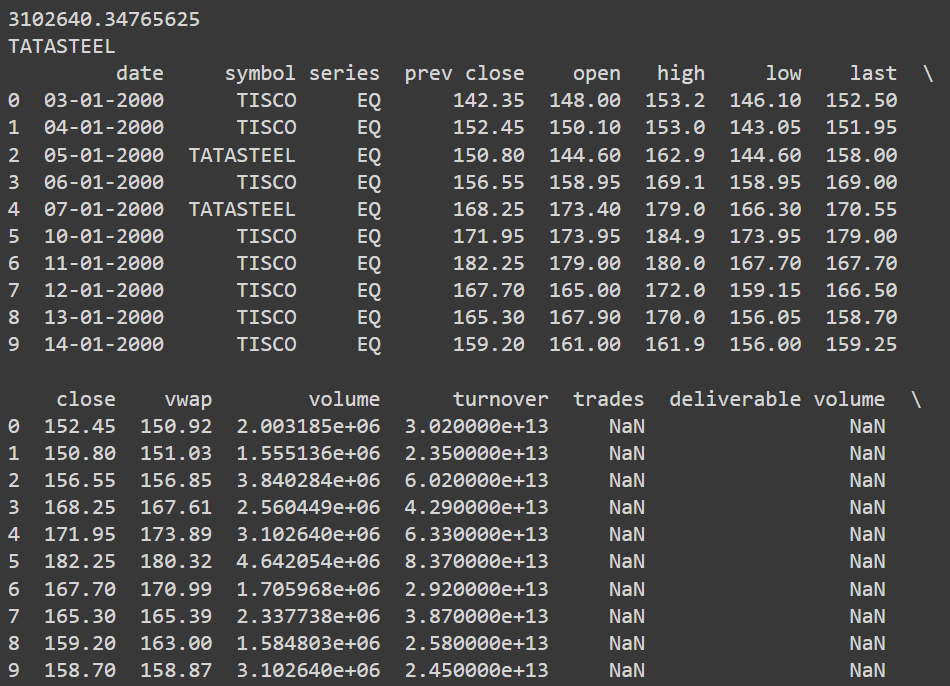
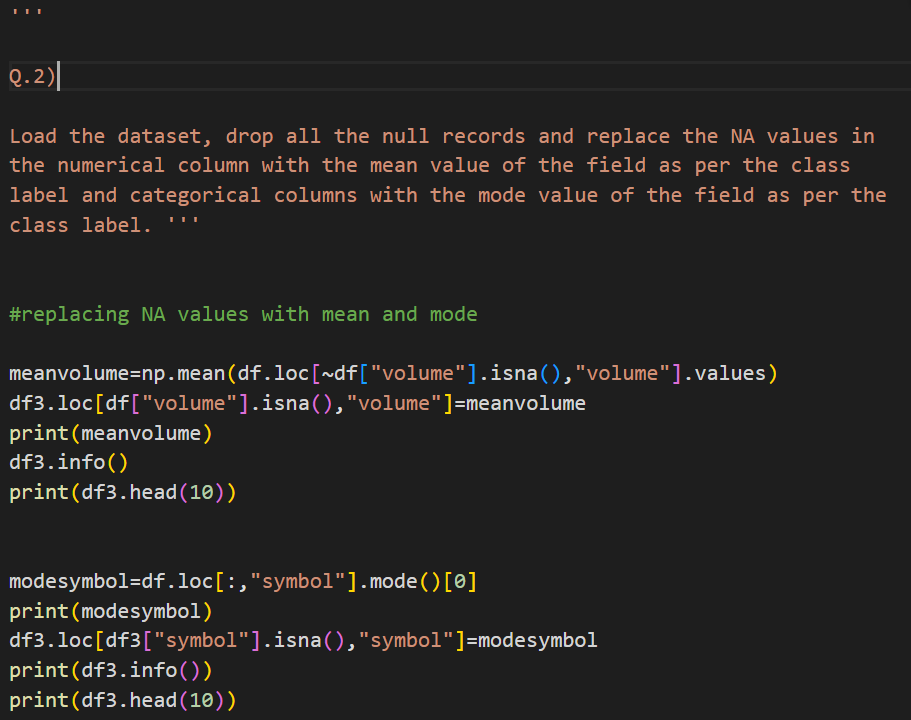
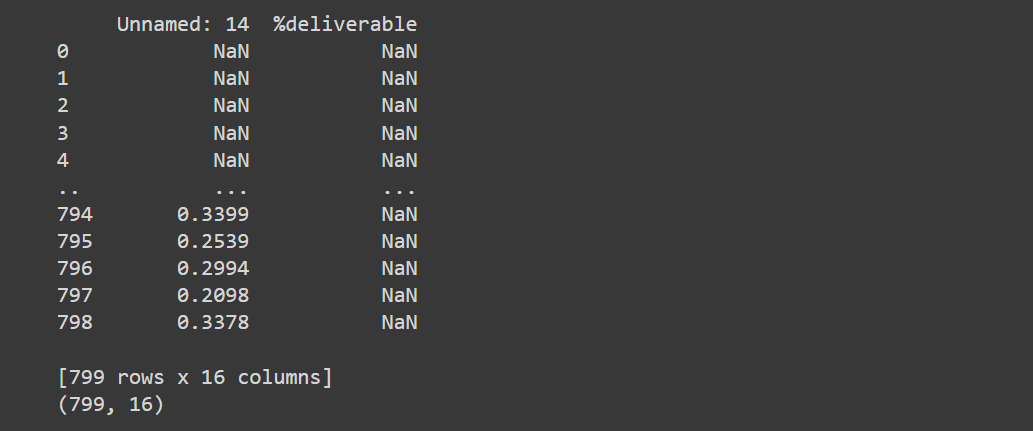
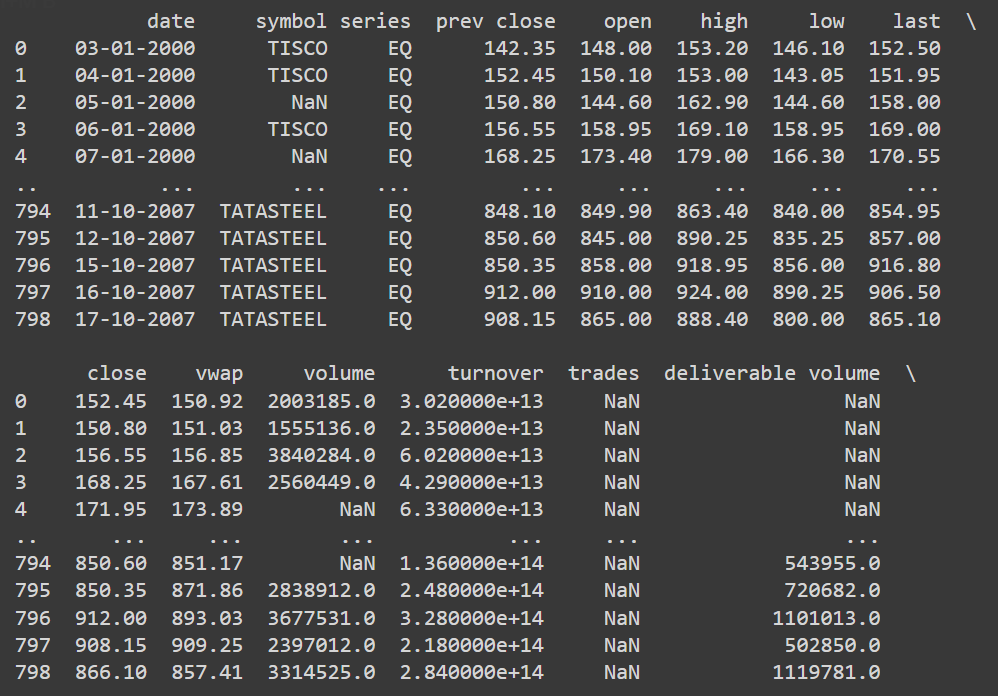
1. prev close, 12. symbol(string)
2. open, 13. Series(string)
3. high,
4. low,
5. last,
6. close,
7. volume,
8. vwap,
9. turnover,

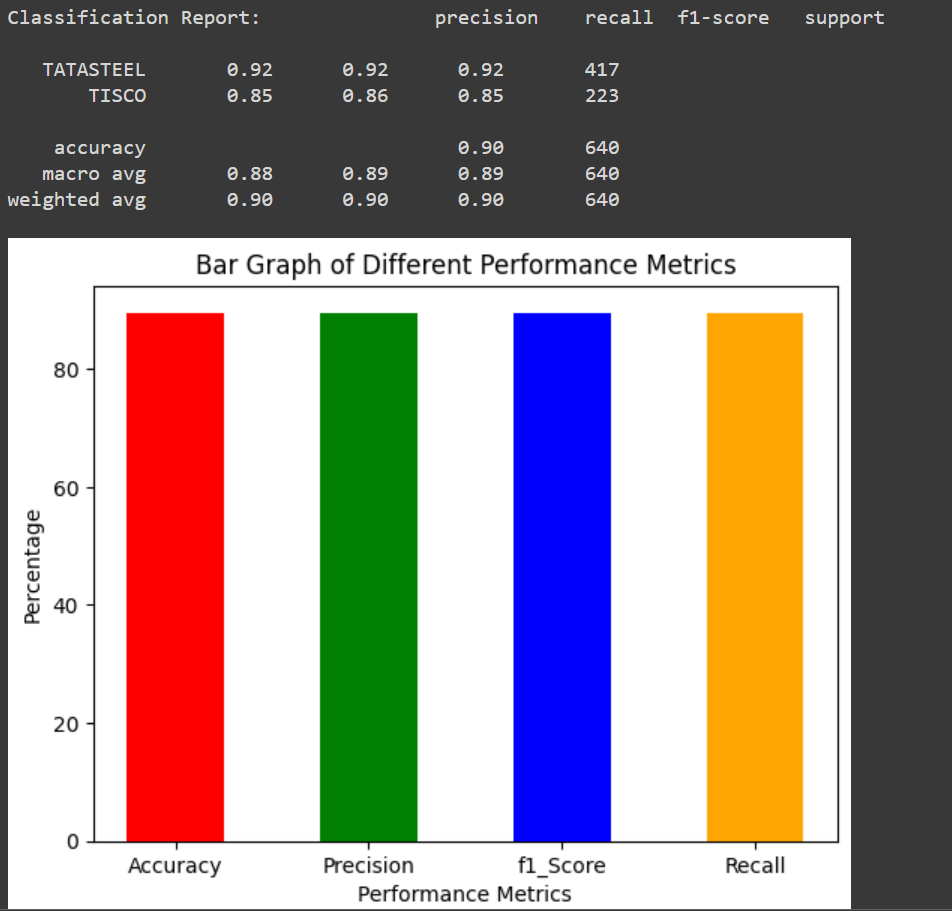
10.deliverable volume,

11.% deliverable.

**CODE AND OUTPUT**







**CONCLUSION**

Conclusively, the dataset analysis and K-nearest Neighbors (KNN) model training have yielded significant insights into the data and its predictive potential. The main takeaways and conclusions from the analysis are listed below:

**1. Data Domain and Characteristics:**

The selected dataset belongs to the domain of Tatasteel and Tisco and the dataset consists of a list of 800 entries of different financial symbols of tatasteel

**2. Data Preprocessing:**

Null entries were removed from the dataset, and missing values in numerical columns were replaced by the mean value and there are no NA values in categorical columns. By preprocessing, we ensure that our dataset is complete before any further analysis.

**3. Statistical Analysis:**

The dataset was analyzed using statistical measures such as count, sum, range, min, max, mean, median, mode, variance, and standard deviation to gain a full picture of the data distribution and central tendencies.

**4. Unique Value Counts:**

By displaying the unique value counts and values of every column, we uncovered the different characteristics and distribution of data within each attribute.

**5. Data Visualization:**

Applicable plot such as bar graph was drawn, allowing for visual exploration and interpretation of the dataset's patterns and relationships.

**6. Model Training and Evaluation:**

The K-nearest Neighbors (KNN) Regressor model was trained with 80% of the data and evaluated using the remaining 20%. The model achieved an accuracy of 88.5%. These metrics indicate that the model performs reasonably well in regression.