**CS 5710 Machine Learning**

**ASSIGNMENT 1**

**Submitted by:**

**Divya Chilukuri**

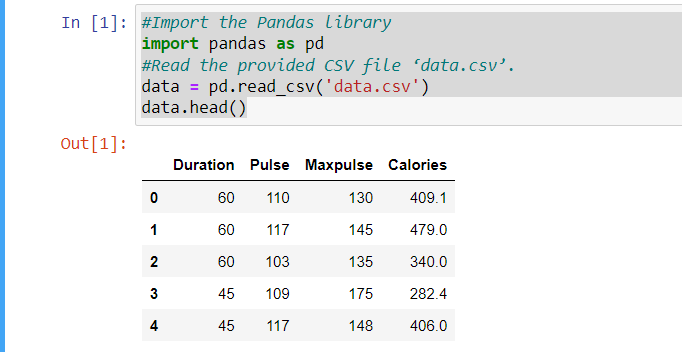
**Student Id: 700744169**

Programming elements:

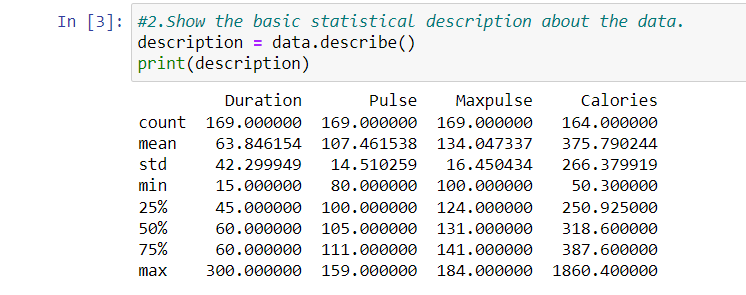
Classification

1. **Pandas**

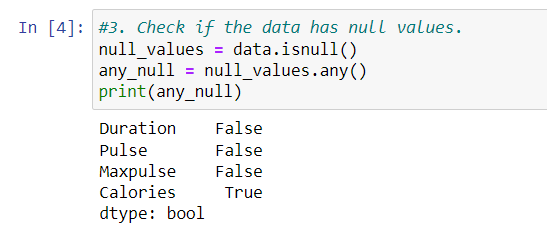
1. Read the provided CSV file ‘data.csv’.



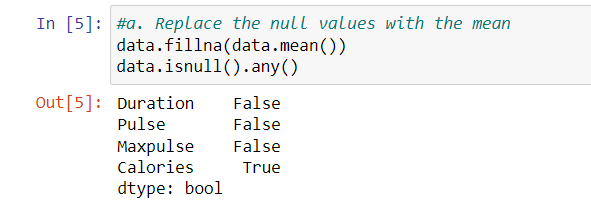
1. Show the basic statistical description about the data.



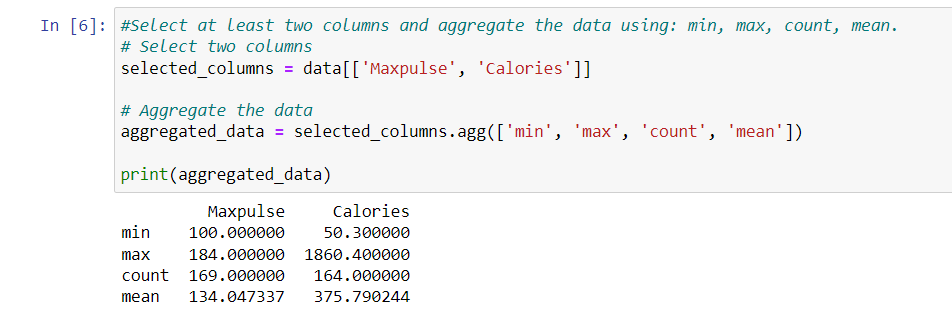
1. Check if the data has null values.



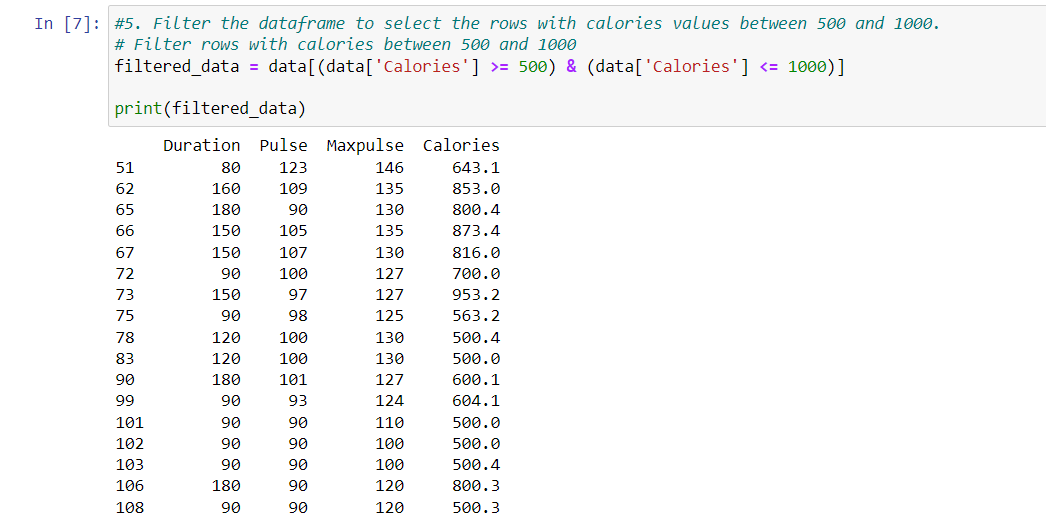
1. Replace the null values with the mean



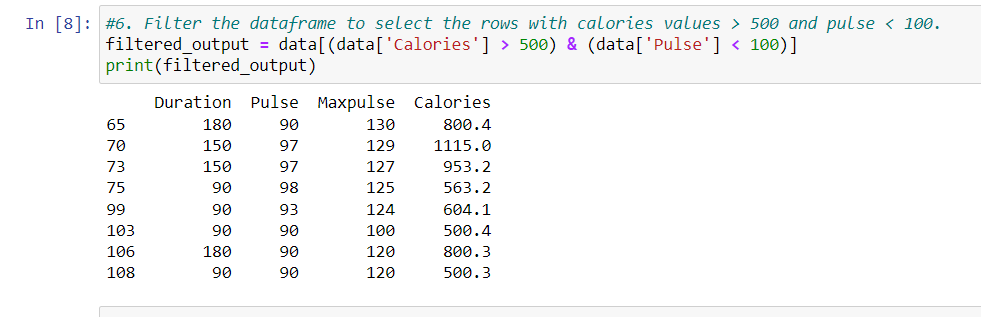
1. Select at least two columns and aggregate the data using: min, max, count, mean.



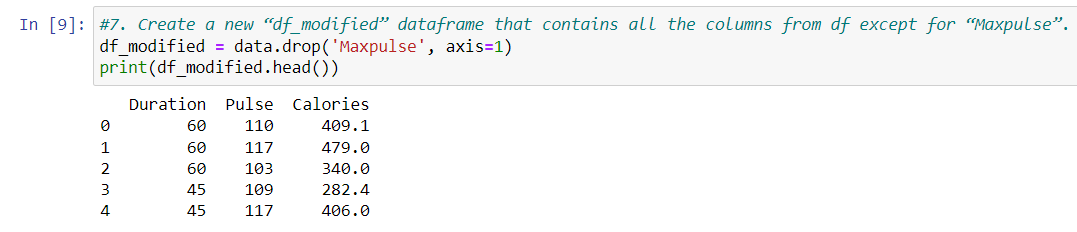
1. Filter the dataframe to select the rows with calories values between 500 and 1000.



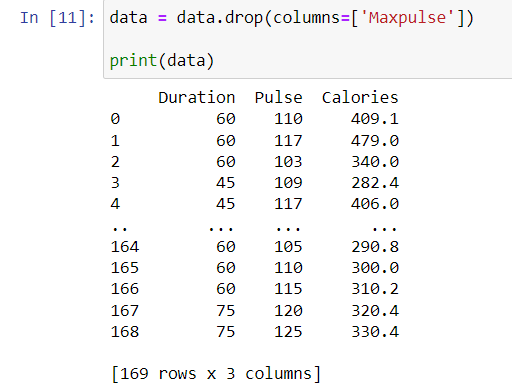
1. Filter the dataframe to select the rows with calories values > 500 and pulse < 100.



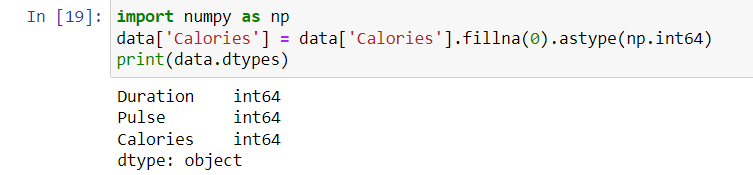
1. Create a new “df\_modified” dataframe that contains all the columns from df except for “Maxpulse”.



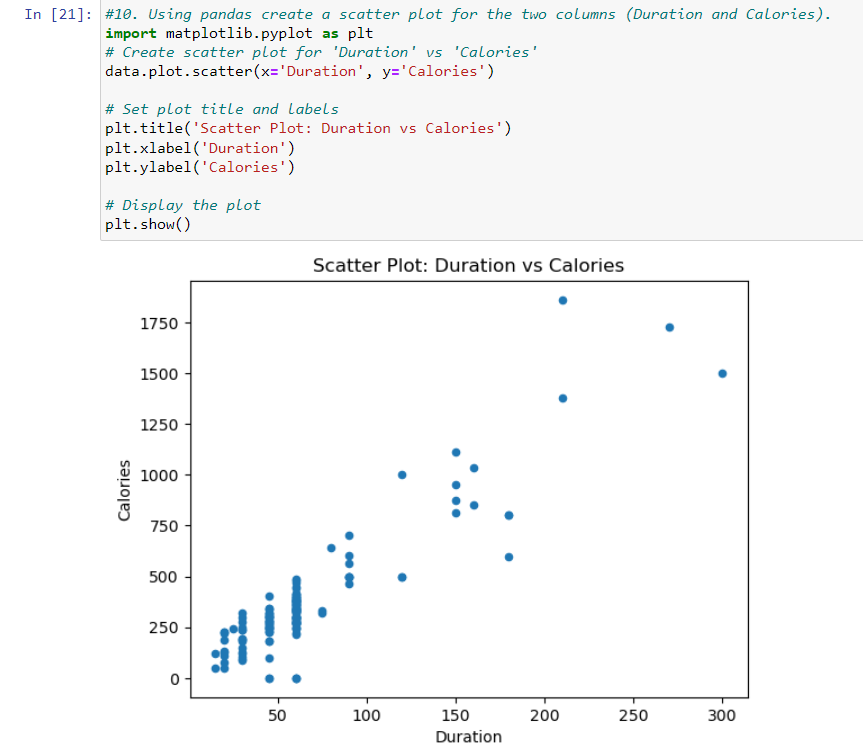
1. Delete the “Maxpulse” column from the main df dataframe



1. Convert the datatype of Calories column to int datatype.



1. Using pandas create a scatter plot for the two columns (Duration and Calories).

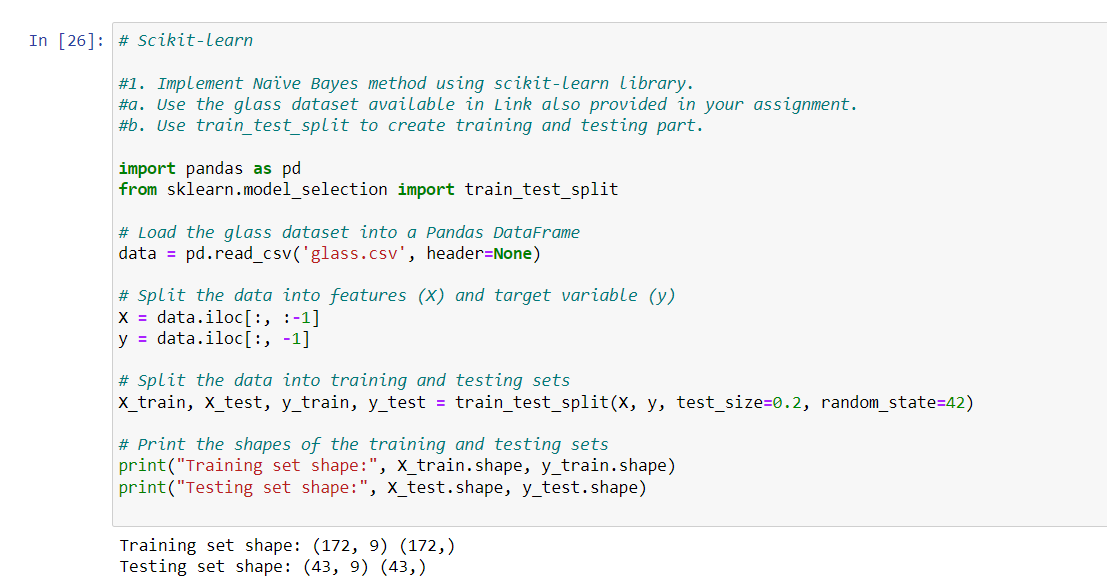


**2. Scikit-learn**

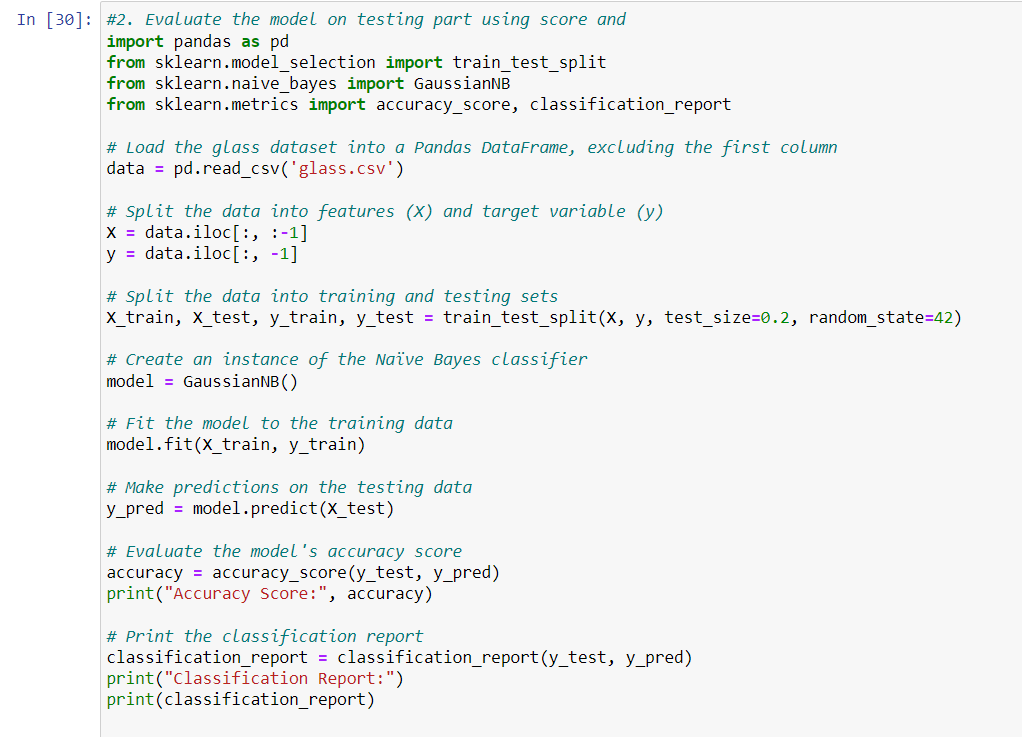
1. Implement Naïve Bayes method using scikit-learn library.

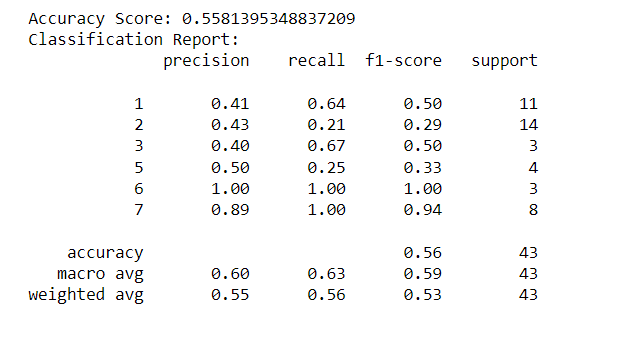
a. Use the glass dataset available in Link also provided in your assignment.

b. Use train\_test\_split to create training and testing part.



1. Evaluate the model on testing part using score and classification\_report(y\_true, y\_pred)



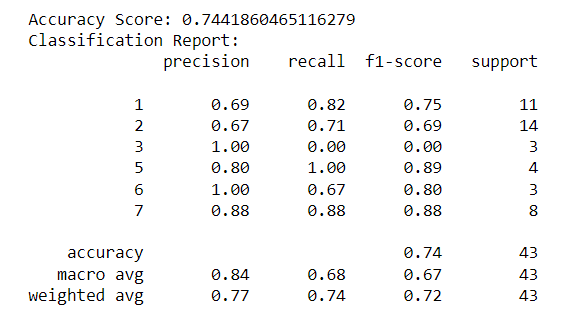


1. Implement linear SVM method using scikit library

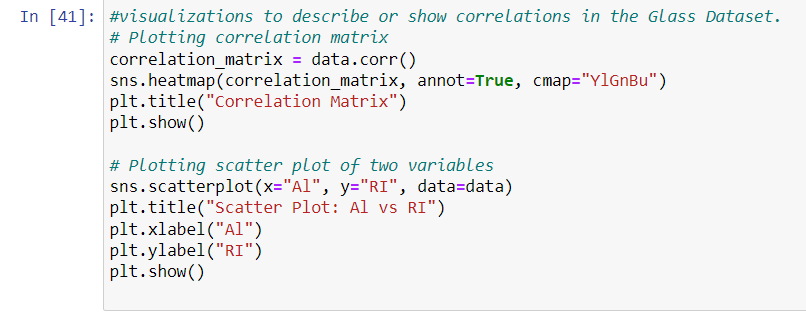
a. Use the glass dataset available in Link also provided in your assignment.

b. Use train\_test\_split to create training and testing part.

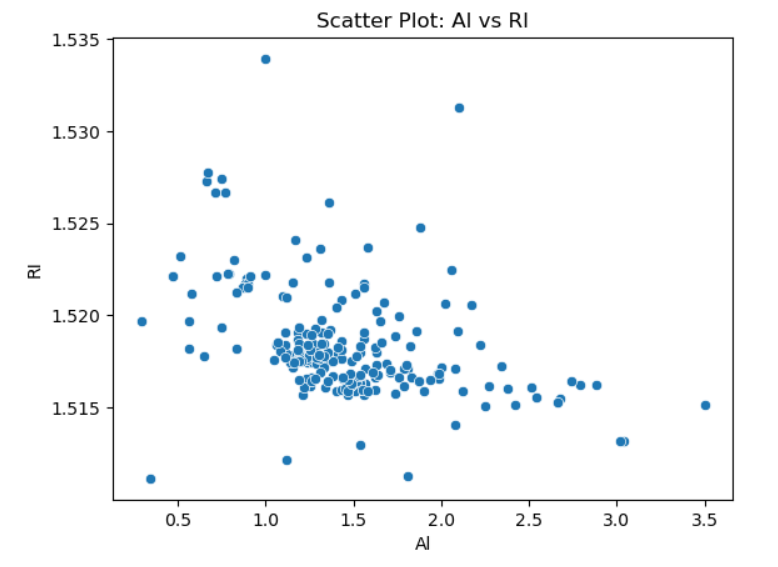




Do at least two visualizations to describe or show correlations in the Glass Dataset.







Which algorithm you got better accuracy? Can you justify why?

To determine which algorithm has better accuracy, we have to compare the accuracy scores obtained from different algorithms.

In this case, we have implemented linear SVM using the Linear SVC algorithm. we can also try other algorithms such as K-Nearest Neighbors (KNN), Decision Tree, or Random Forest and compare their accuracy scores.

The algorithm with the highest accuracy score can be considered as the better-performing algorithm for this particular dataset. In this case Linear SVM has the better accuracy rate, However, it's important to note that the performance of different algorithms may vary depending on the dataset and problem at hand