**NAME :**Leela Sekhar Chowdary Mannava

**ID(700#) :**700740979

**E-Mail :**lxm09790@ucmo.edu

**Assignment :**4

**Tools :**Google Colab

**Programming:** Python

**GitHub**

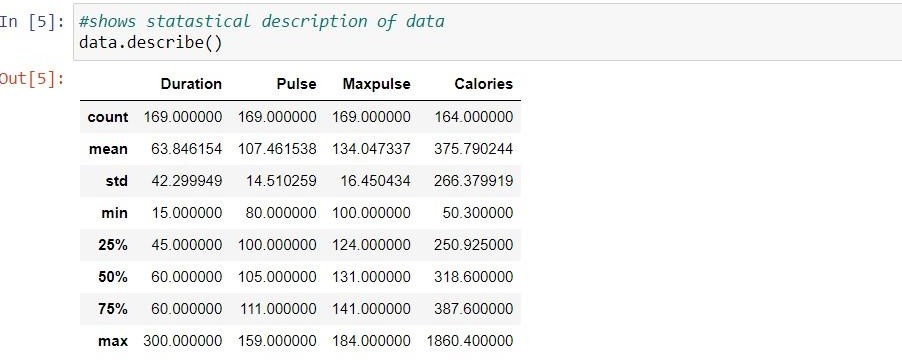
**Link:** [**https://github.com/LXM09790/Assignment-4.git**](https://github.com/LXM09790/Assignment-4.git)

**Q1) Pandas**

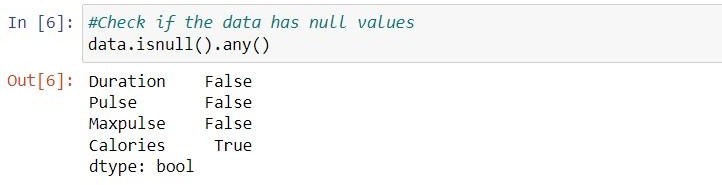
1. **Read the provided CSV file ‘data.csv’.**



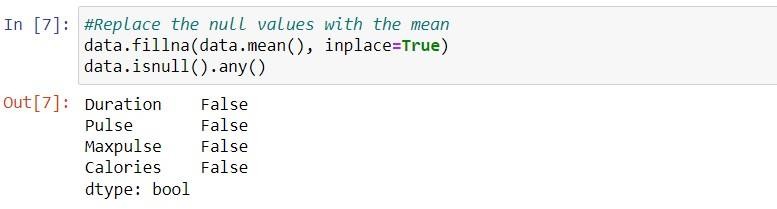
1. **Show the basic statistical description about the data.**



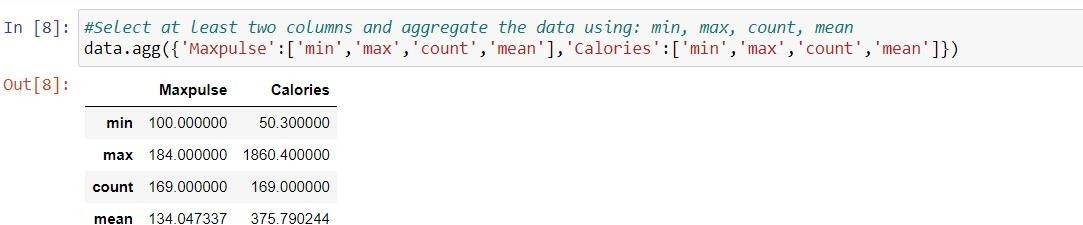
1. **Check if the data has null values.**



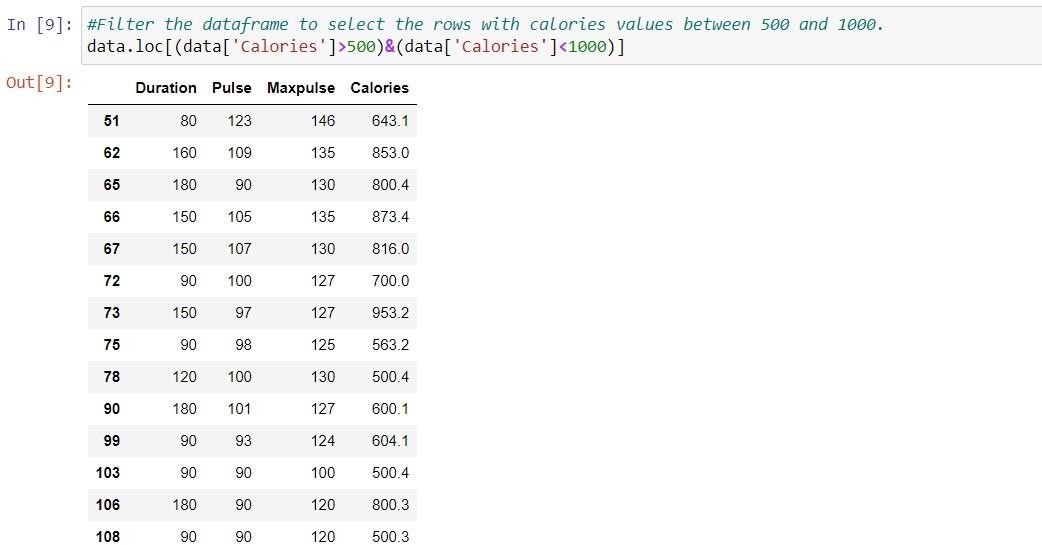
a. Replace the null values with the mean



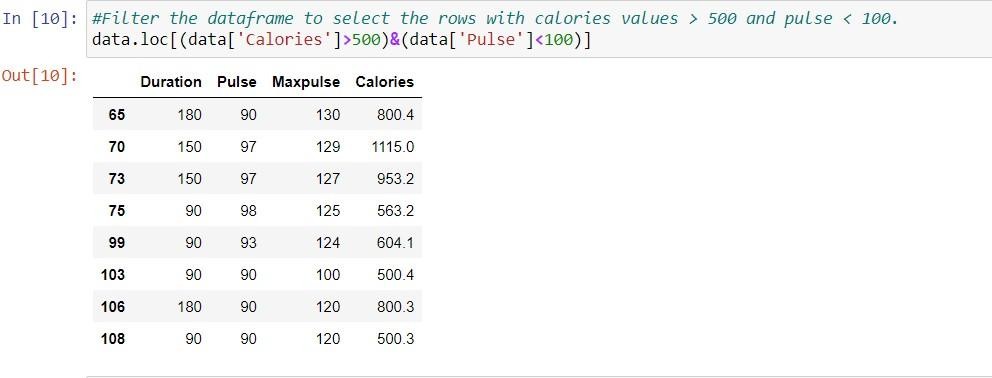
1. **Select at least two columns and aggregate the data using: min, max, count, mean**



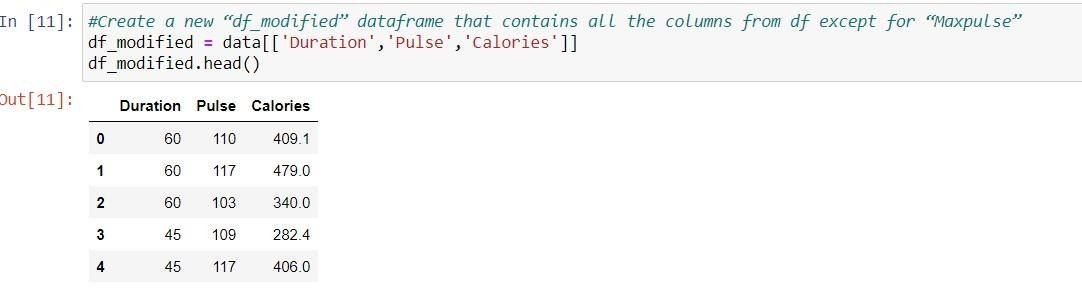
1. **Filter the data frame to select the rows with calories values between 500 and 1000.**



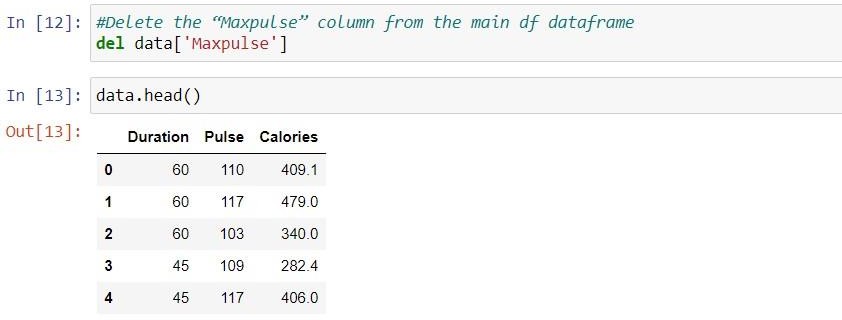
1. **Filter the data frame to select the rows with calories values > 500 and pulse < 100**



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



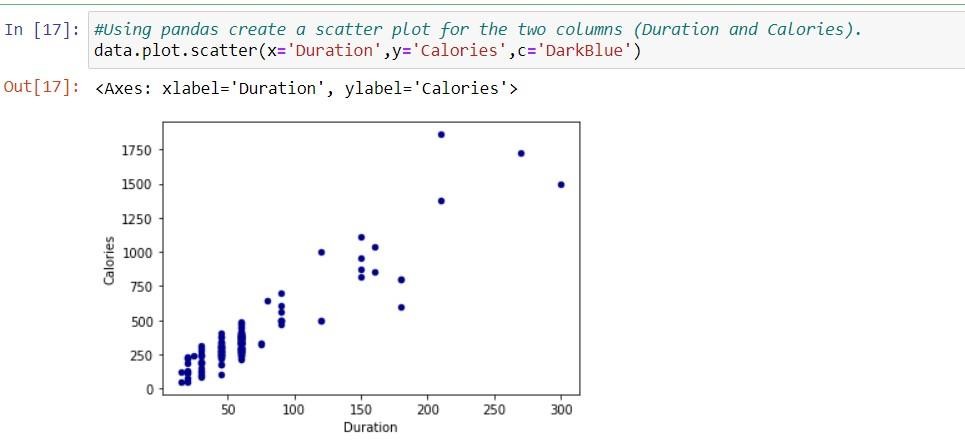
1. **Delete the “Maxpulse” column from the main df dataframe**



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



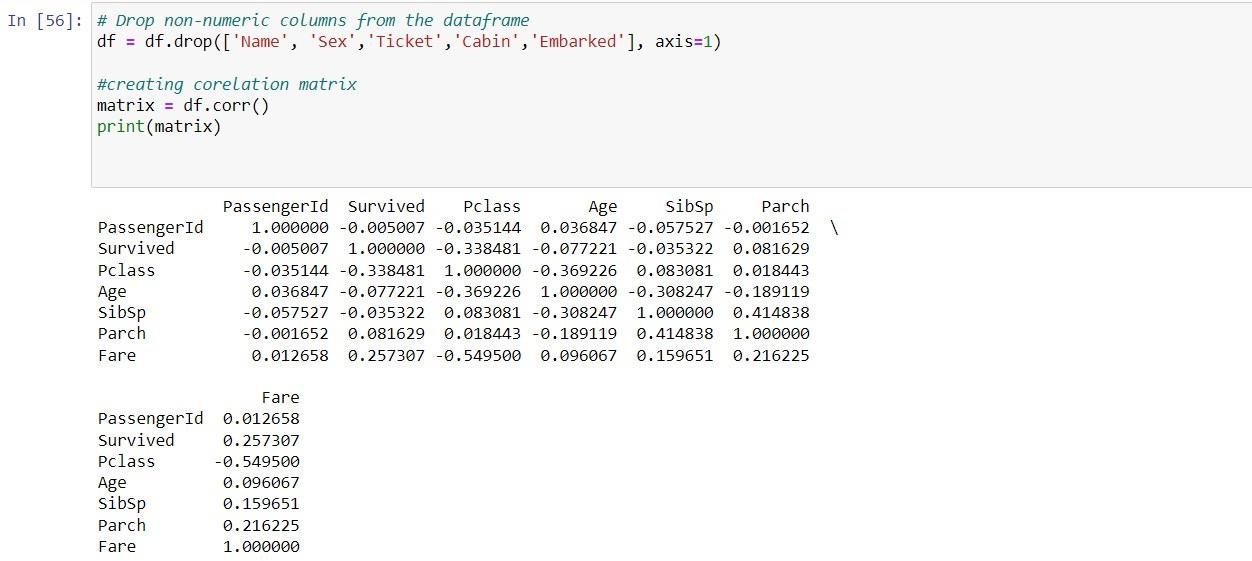
1. **Using pandas create a scatter plot for the two columns (Duration and Calories).**

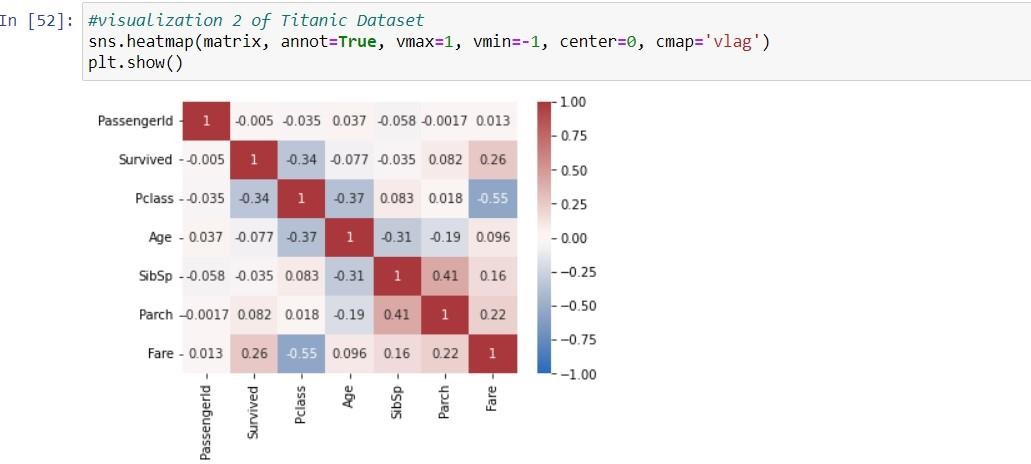


1. **(Titanic Dataset) 1. Find the correlation between ‘survived’ (target column) and ‘sex’ column for the Titanic use case in class**
2. **Do you think we should keep this feature?**

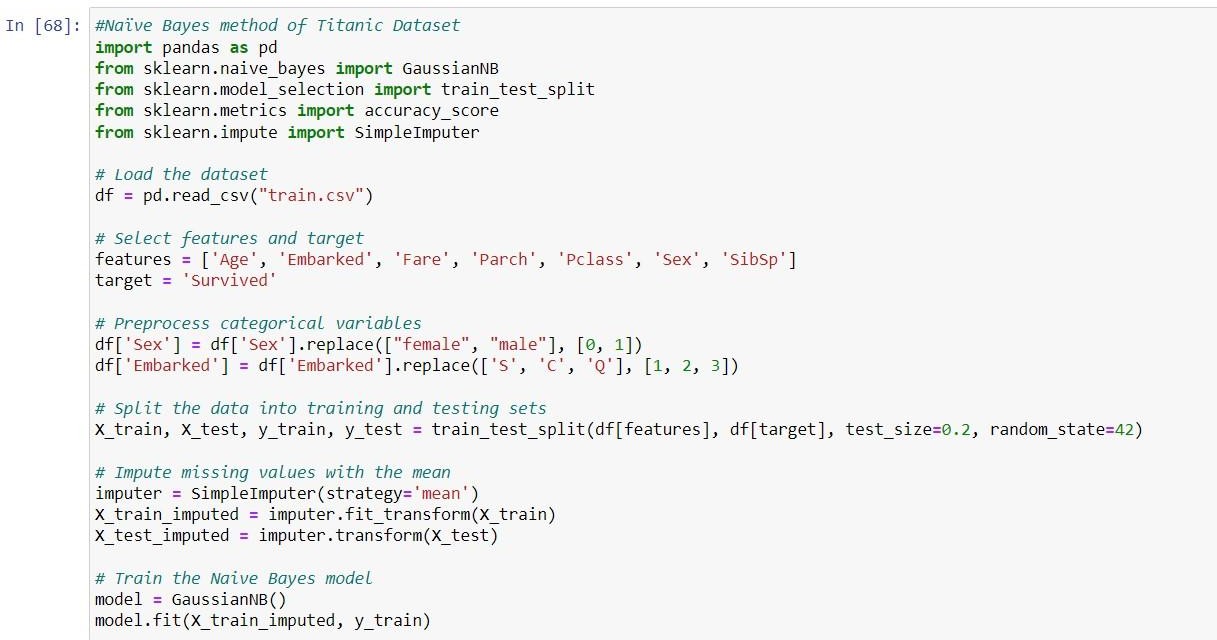
Ans. No, because the accuracy is just 54% only. So we should not keep this feature.

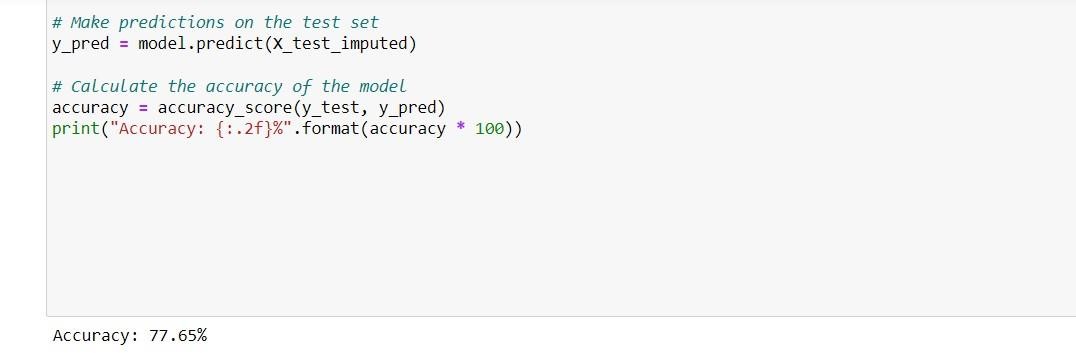
1. **Do at least two visualizations to describe or show correlations.**





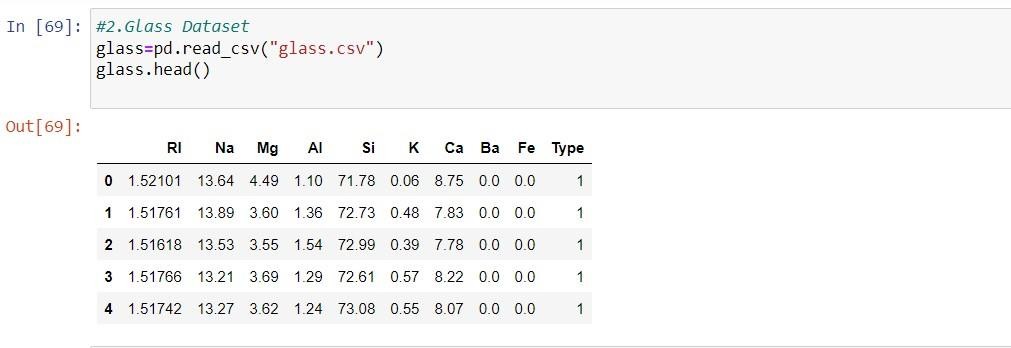
1. **Implement Naïve Bayes method using scikit-learn library and report the accuracy.**

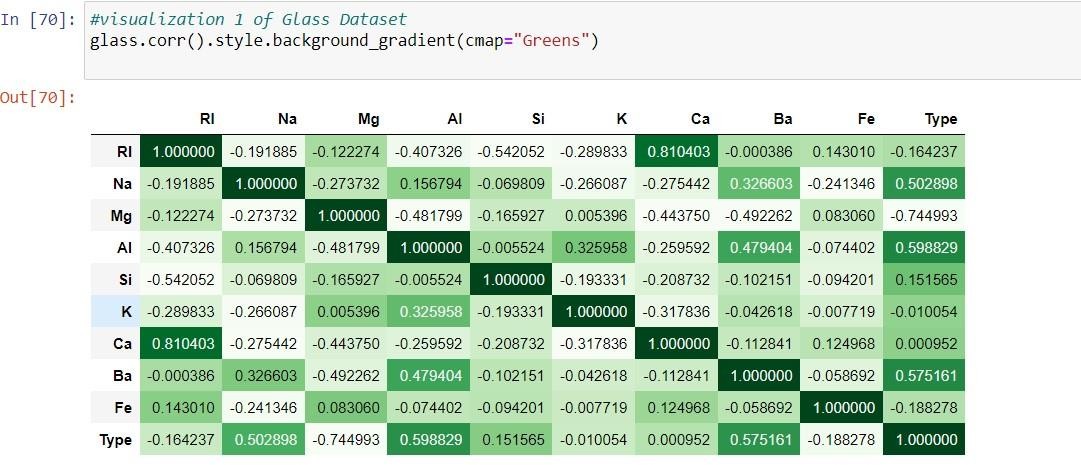


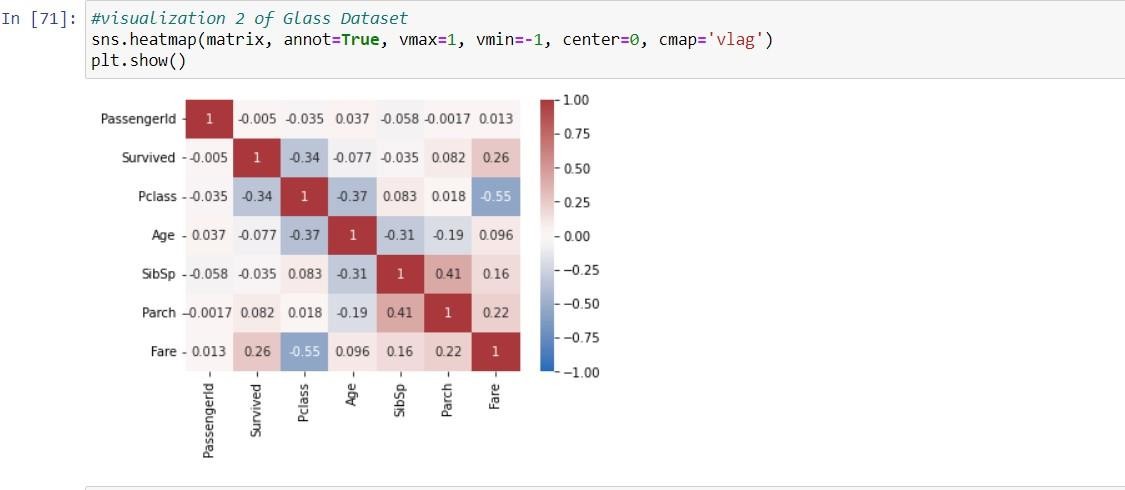


1. **(Glass Dataset)**

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

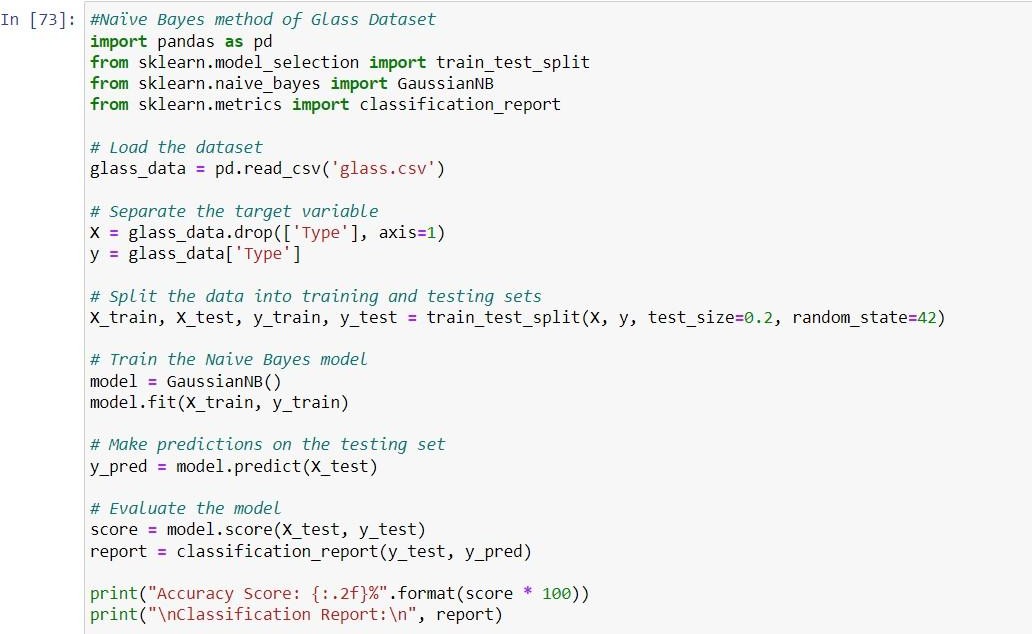


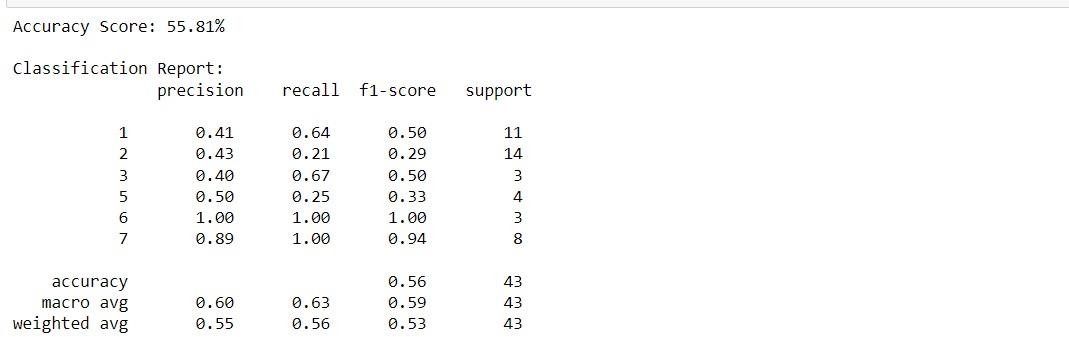




* 1. **Implement Naïve Bayes method using scikit-learn library.**

1. **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)**

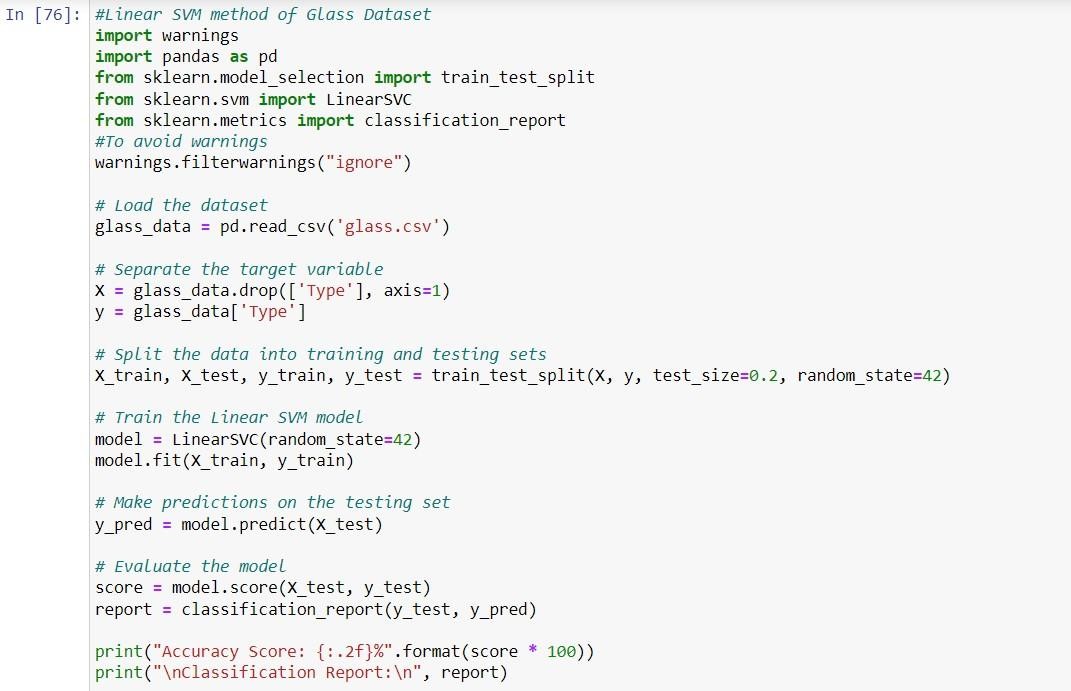


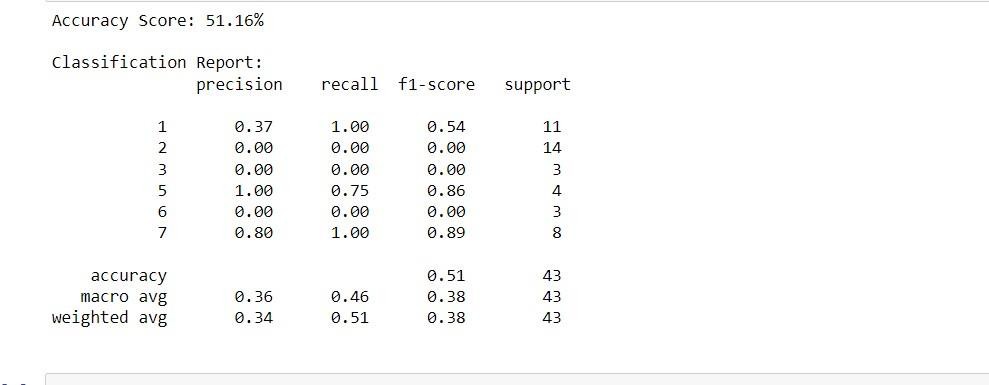


1. **Implement linear SVM method using scikit library**

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)**





MACHINE LEARNING ASSIGNMENT 4

1. **Which algorithm you got better accuracy? Can you justify why?**

Ans. Naïve Bayes algorithm got better accuracy than Linear SVM algorithm because naïve bayes need small amount of training data Where as Linear SVM need large amount of training data and SVM is more expensive than Naïve bayes algorithm. But it will give output easily.And naïve bayes is good for spam detection and text classification problems

So Naïve bayes is better than Linear Svm.