**CS 5710**

**Machine Learning (Assignment # 4)**

E-mail: [gxm04390@ucmo.edu](mailto:gxm04390@ucmo.edu)

Name: Mogalapu Gideon Jaideep Sudarshan

Course: CS 5710

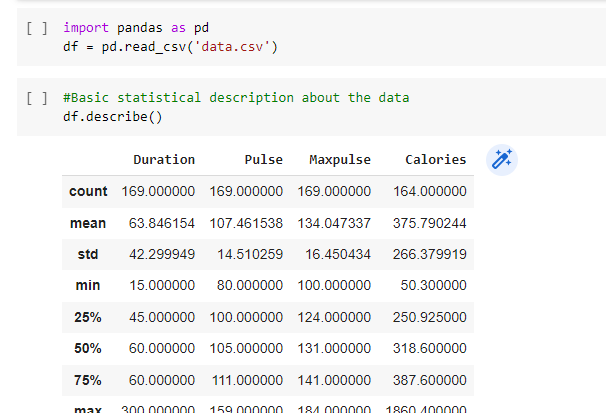
Assignment: Assignment 4

Git Hub: <https://github.com/JaideepMogalapu/Assignment4>

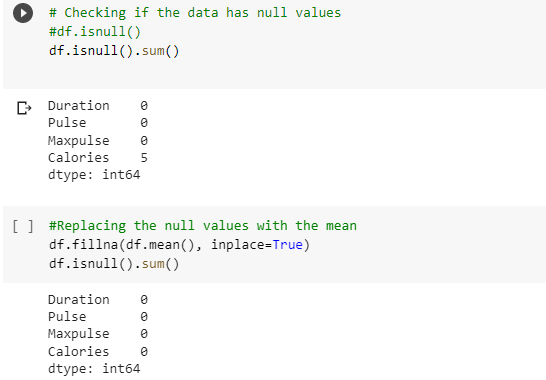
Video Link: <https://drive.google.com/drive/folders/1jpC7moc6gbPz-m7hMIa4Zyp_NN1JzpVE?usp=share_link>

**Question 1**

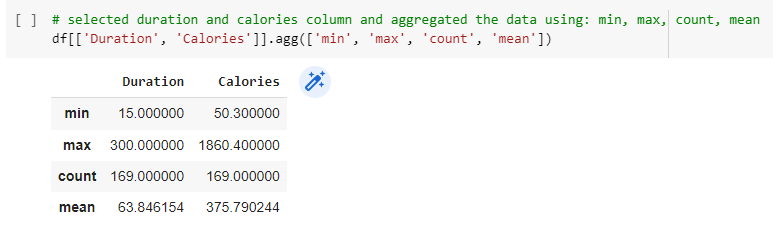
**Description:** . Reading the provided CSV file ‘data.csv’. we also printed basic statistical description about the data



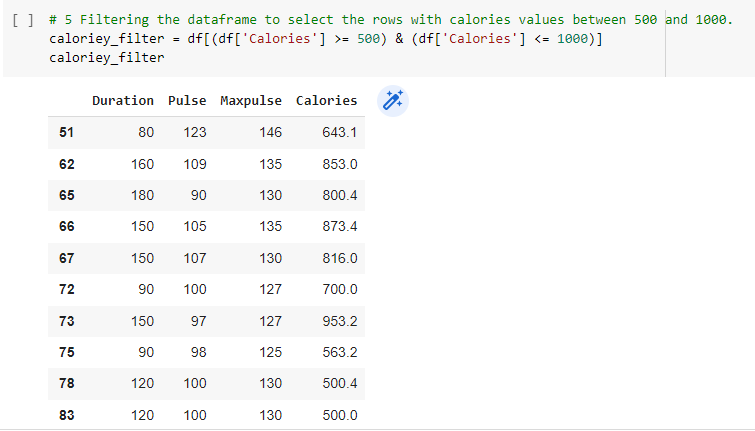
**Description:** . 3. Check if the data has null values. Replace the null values with the mean



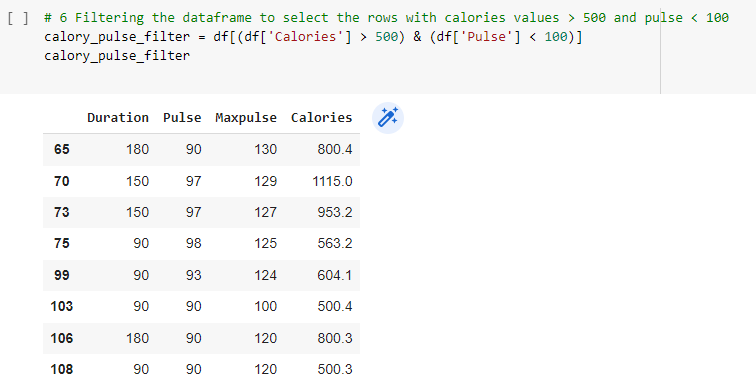
**Description:** . Select at least two columns and aggregate the data using: min, max, count, mean



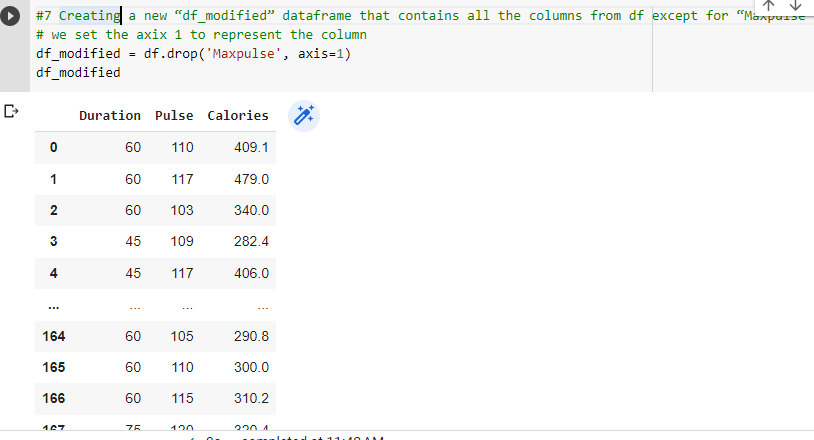
**Description:** . Filter the dataframe to select the rows with calories values between 500 and 1000.



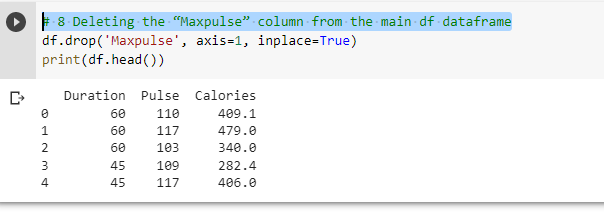
Filter the dataframe to select the rows with calories values > 500 and pulse < 100



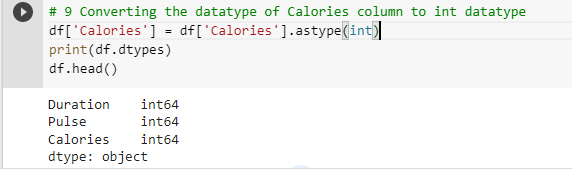
**Description:**Creating a new “df\_modified” dataframe that contains all the columns from df except for “Maxpulse” .we set the axix 1 to represent the column



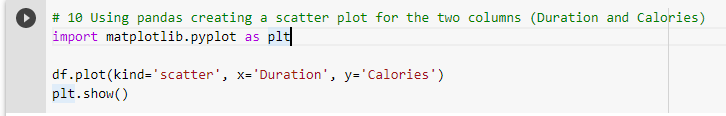
**Description:**  Deleting the “Maxpulse” column from the main df dataframe

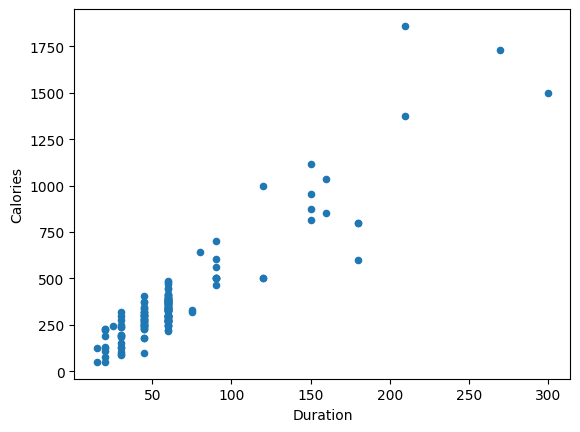


**Description:**Converting the datatype of Calories column to int datatype



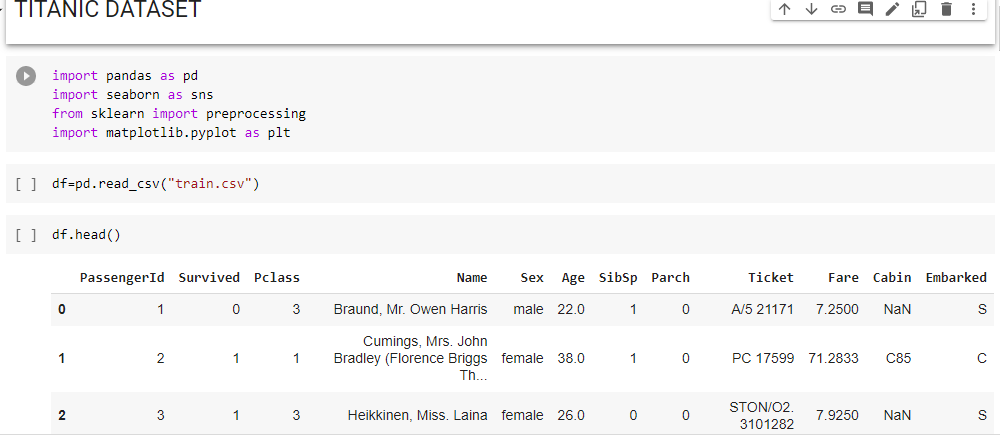
**Description:**Using pandas creating a scatter plot for the two columns (Duration and Calories)

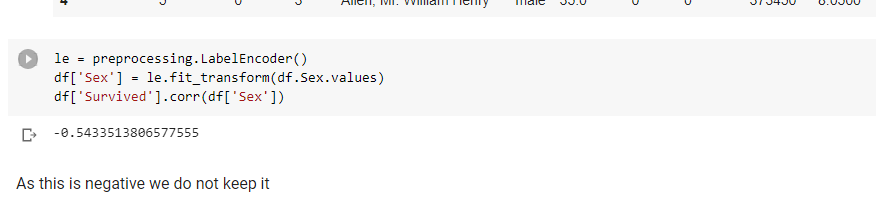




**2) Titanic Dataset**

**Description:** Imported the necessary libraries. As we got negative correlation, we can remove these two variables

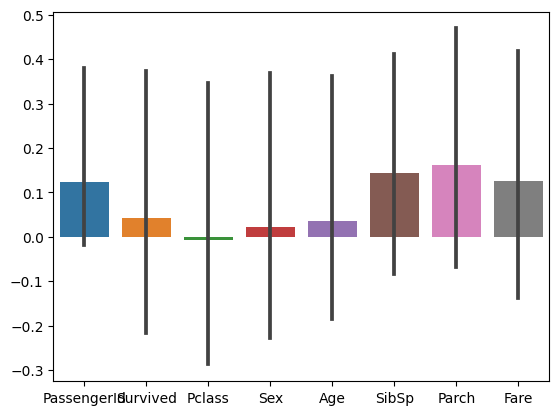




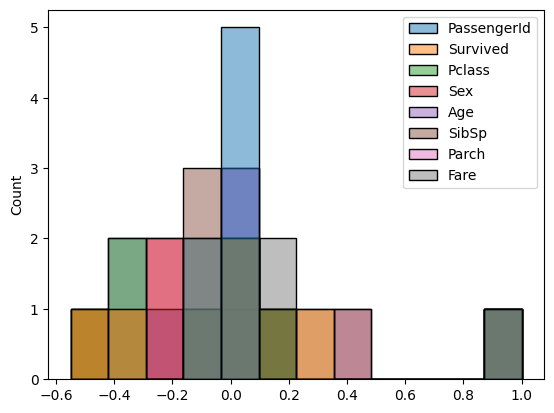


**Description :**  We have plotted the histogram and bar chart for the above data set

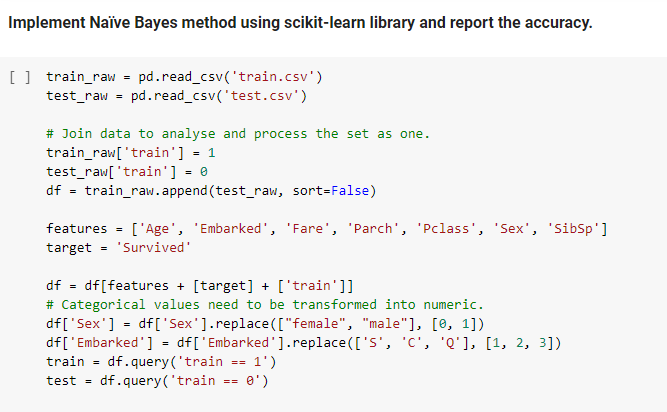


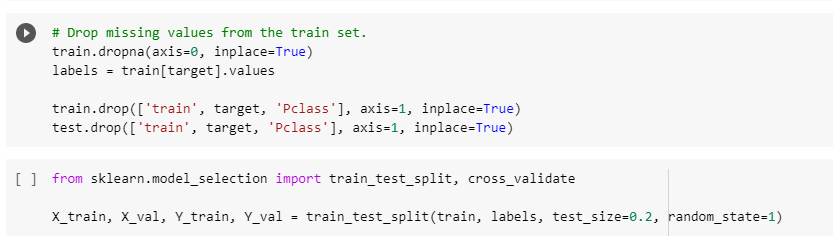




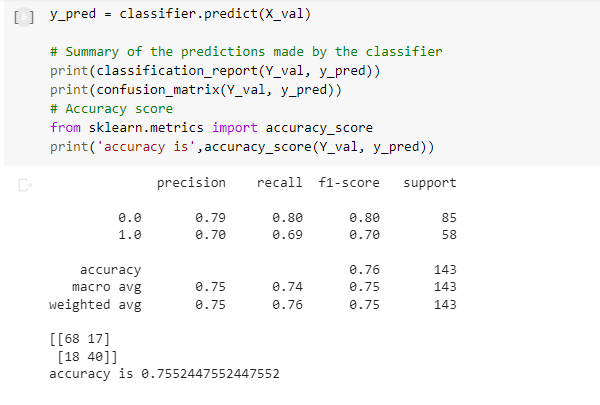


**Description** : Implemented Naïve Bayes method using scikit-learn library and reported the accuracy.



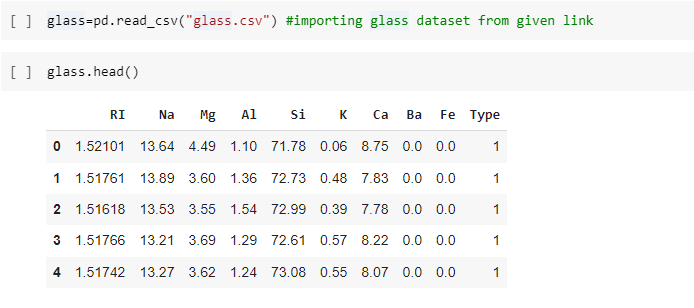


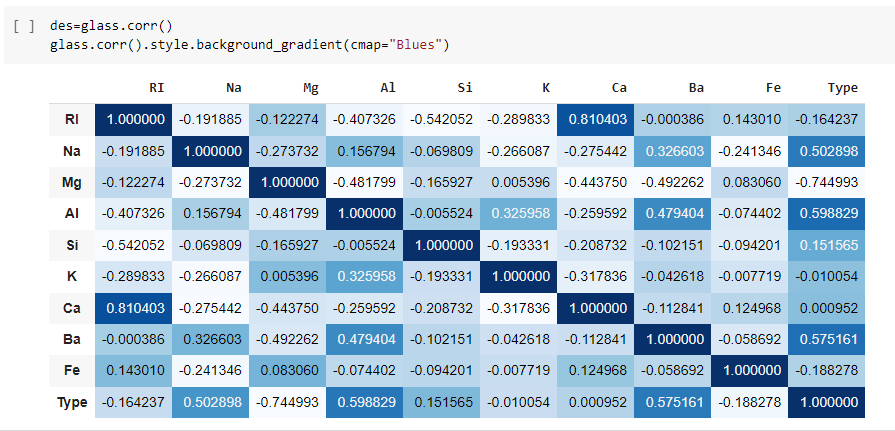




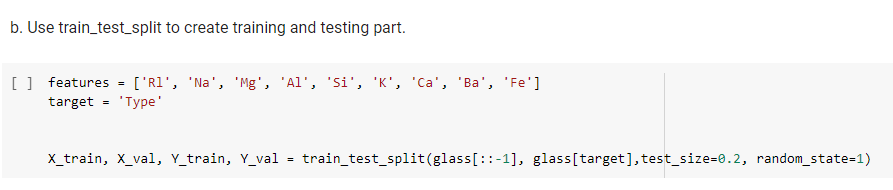
**3) Glass Dataset**

1. Used the glass dataset available in Link also provided in your assignment:



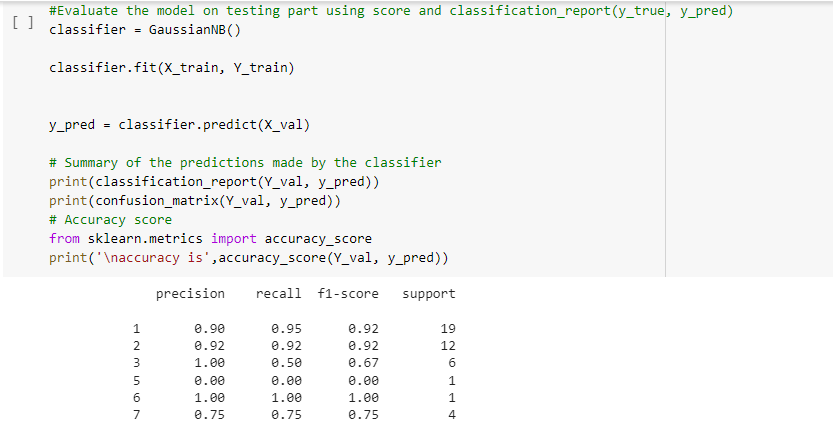


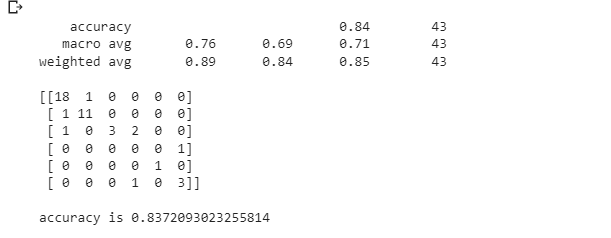
B:Use train\_test\_split to create training and testing part:



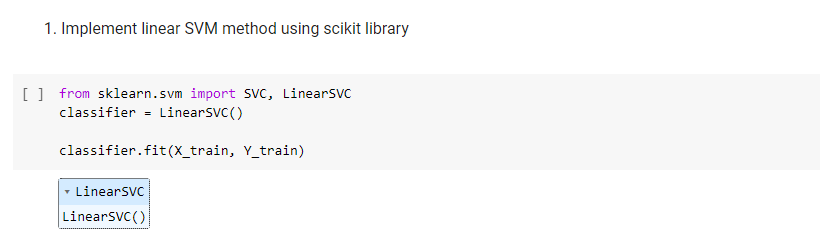
2. Evaluate the model on testing part using score and classification\_report(y\_true, y\_pred)

Output :

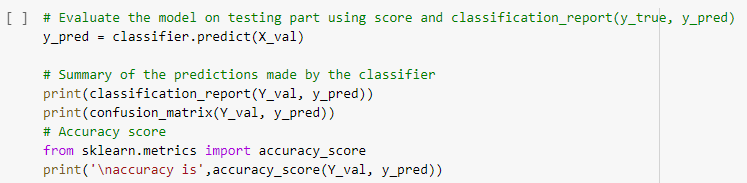


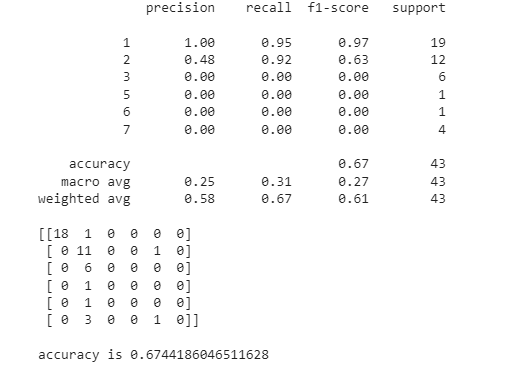


Implement linear SVM method using scikit library

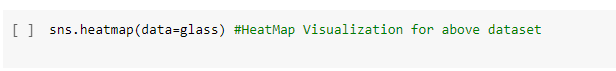


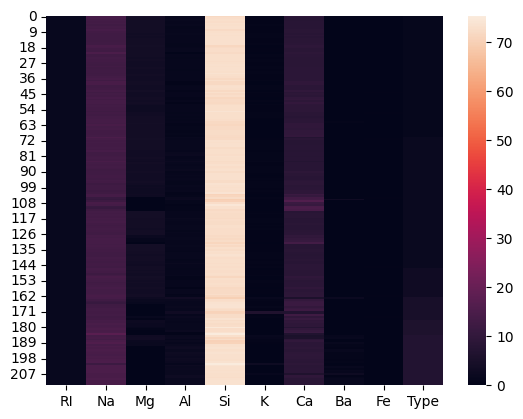
1. Evaluate the model on testing part using score and classification\_report(y\_true, y\_pred) :



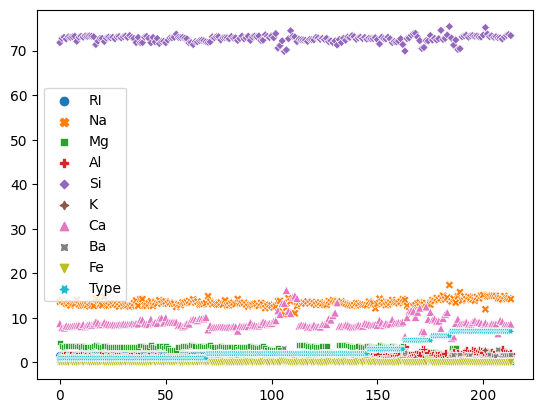


Do at least two visualizations to describe or show correlations in the Glass Dataset:









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

Naïve Bayes accuracy is 83.7% which got more than SVM algorithm which has only 60.46%. While naive Bayes can handle both categorical predictor variables and continuous hence, therefore SVM algorithm can handle only continuous predictors.