**Name: Deepika Rani Gokavarapu**

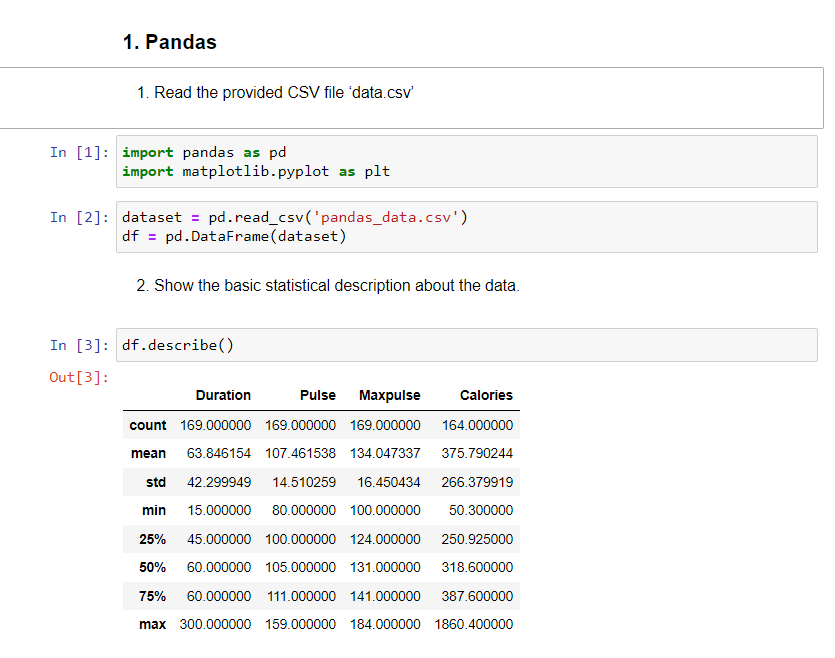
**700#: 700740599**

**Programming Assignment 2**

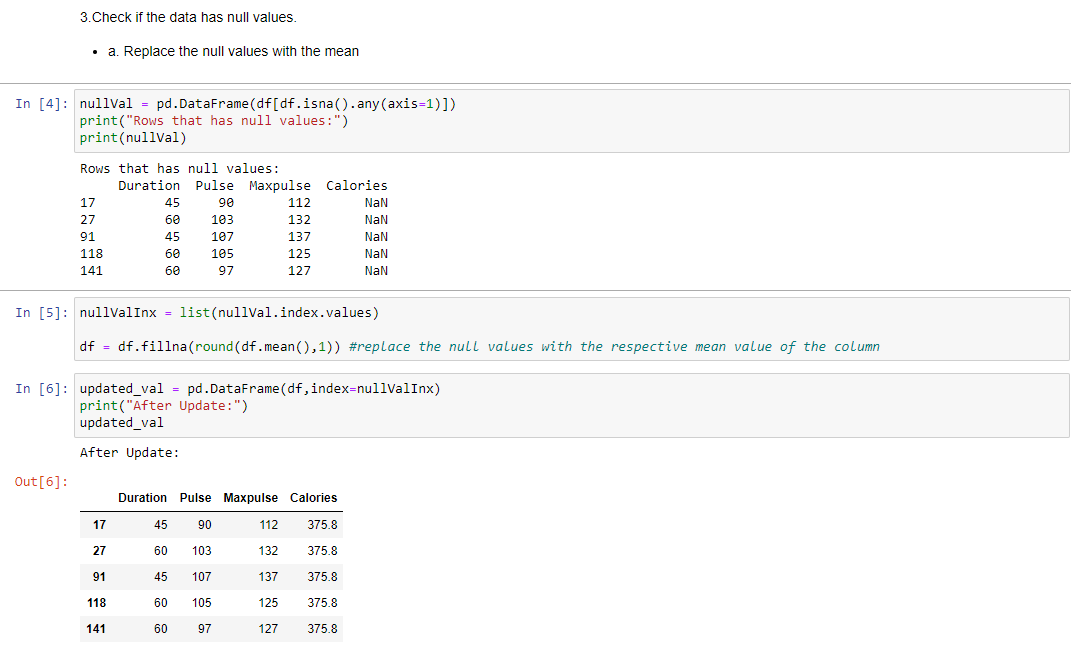
GitHub Link: <https://github.com/GDeepikarani59/MachineLearning_Assignment2>

Video Link: <https://drive.google.com/file/d/1pwfFRTxinMaEkBu8gJQnYa2Vv8ENXhOD/view?usp=drive_link>

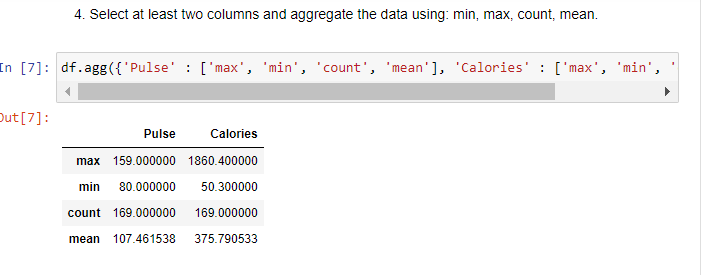
Question 1:



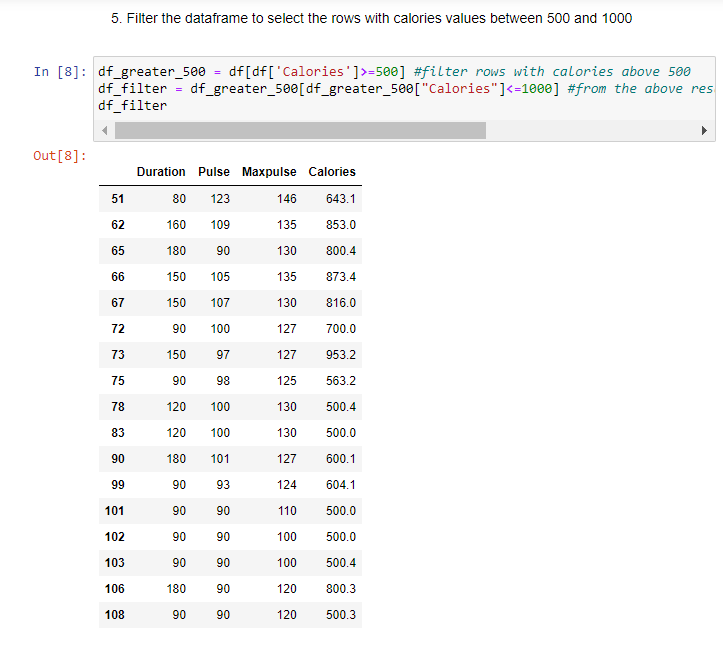
Import the pandas libraries. Read the csv file containing the data sets and display the basic statistical descriptions of the dataset.



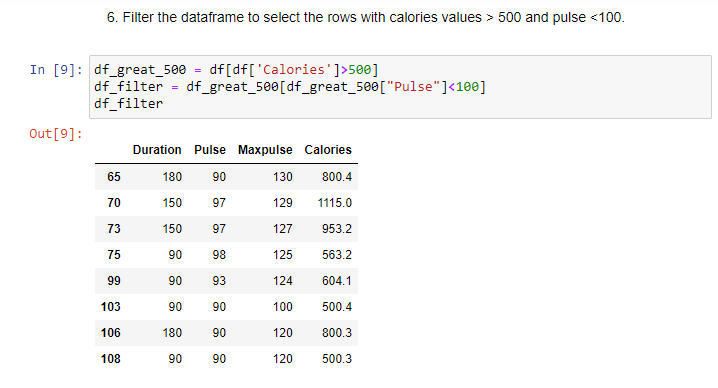
Collect the rows that have any of their column values set to null, then copy the indexes of those rows into a list. To observe how our data will seem after updating the null values, we are saving the indexes here. Replace any null values with their corresponding means. We can now view the rows' updated appearances using the row indexes.



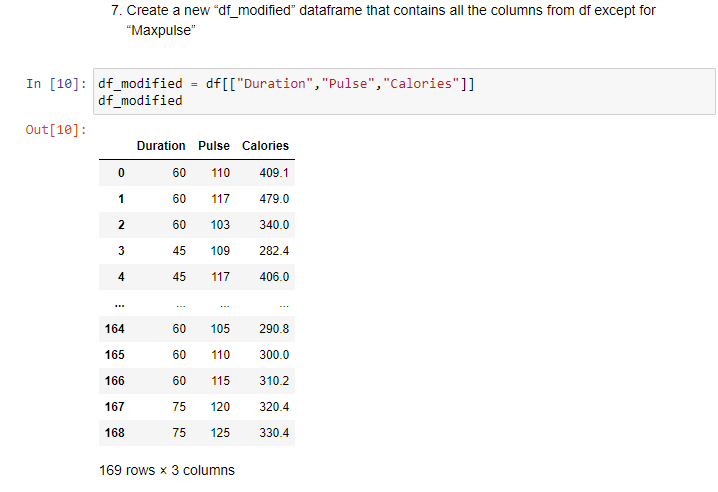
Here, the maximum value, minimum value, count, and mean of the two columns, pulse and calories, are combined and displayed.



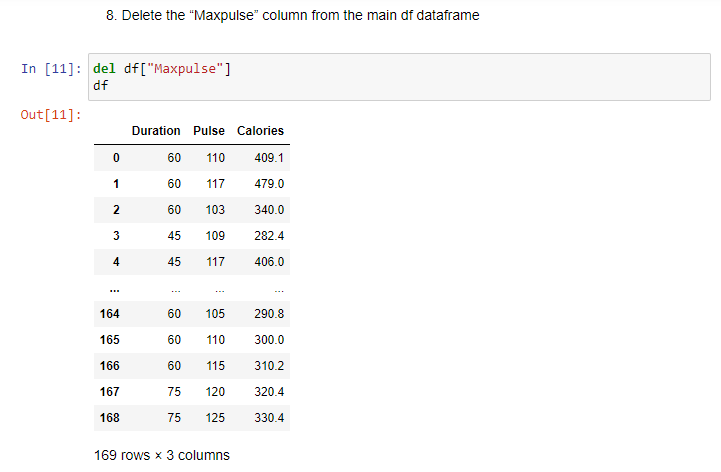
The rows with Calories column values between 500 and 1000 are shown below. There are two steps to this. We first filter out numbers greater than 500 and record the results. After that, only the data with values less than 1000 will be kept.



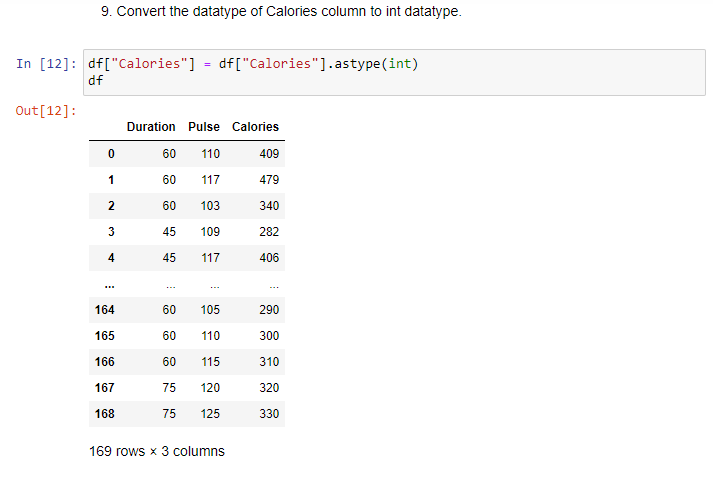
The rows with Calories column values greater than 500 and Pulse column values less than 100 are shown here. There are two steps to this. In the beginning, we filter out calories with values more than 500 and store them. Then, filter the output data to exclude Pulse values below 1000.



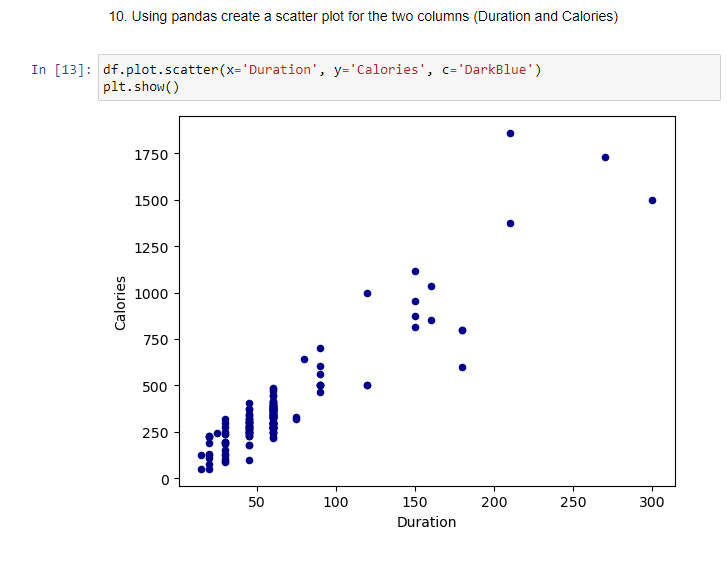
Creating a new data frame containing the all the columns except Maxpulse



From the main data frame, removed the Maxpulse column.

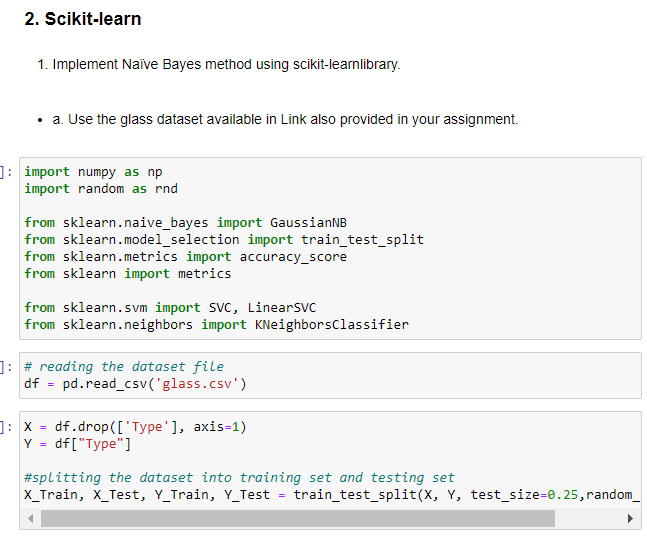


changing the calorie column's float datatype to an int datatype.

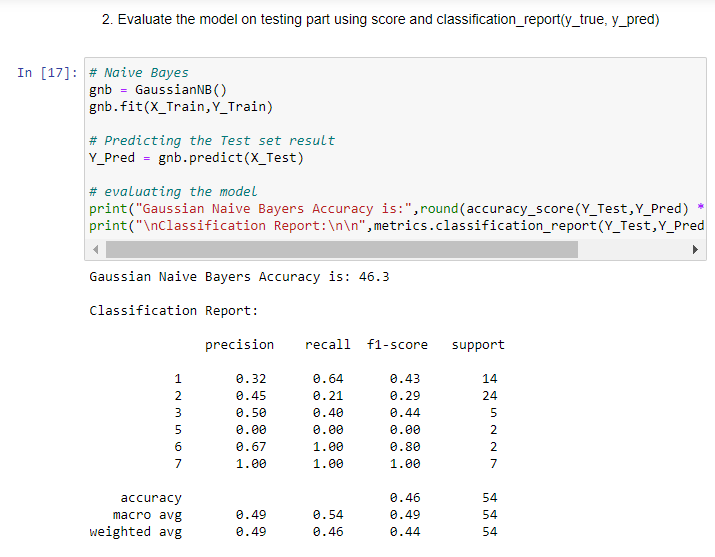


Created a scatter plot for Duration and Calories.

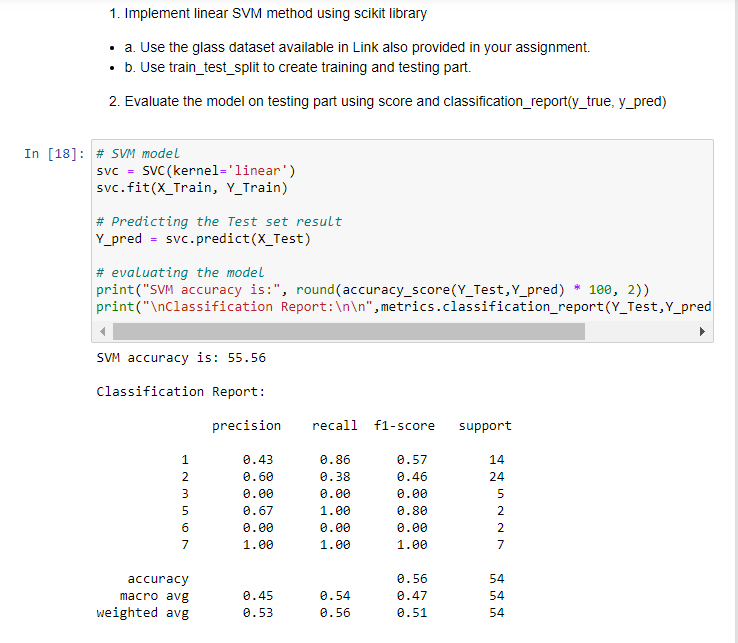
**Question 2:**

****

Imported necessary libraries. Downloaded glass.csv file from the provided link and converted into a data frame using pandas and removed the column ‘type’. Splitted the dataset into training and testing dataset in 75 25 ratio respectively using train\_test\_split method.

****

After splitting the data we have given training data to the naive bayes model. After that we predicted independent variable using test data and trained naive bayes model. Evaluated the model on testing part using score and classification\_report.

****

After splitting the data we have given training data to the SVM model. After that we predicted independent variable using test data and trained the SVM model. Evaluated the model on the testing part using score and classification\_report.