**CS 5710 Machine Learning**

**ASSIGNMENT 2**

**Submitted by:**

**SRAVYA CHEVUTUKUR**

**Student Id: 70075406**

**CRN: 30527**

**Github Link :** [**https://github.com/SravyaChevutukur/ML\_Assignment2**](https://github.com/SravyaChevutukur/ML_Assignment2)

**Video Link:** [**https://drive.google.com/drive/folders/1uTxwx7xIkWPoP1rmc70wDQ6JoZ2tcq4S?usp=drive\_link**](https://drive.google.com/drive/folders/1uTxwx7xIkWPoP1rmc70wDQ6JoZ2tcq4S?usp=drive_link)

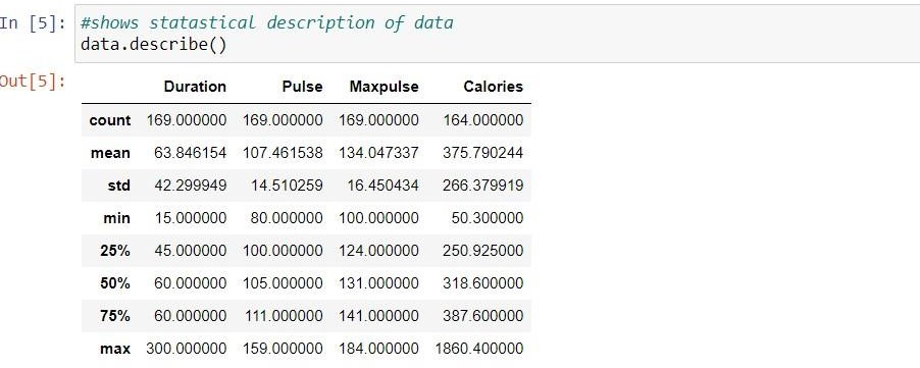
**Q1) Pandas**

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

**A screenshot of a computer

Description automatically generated with medium confidence**

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

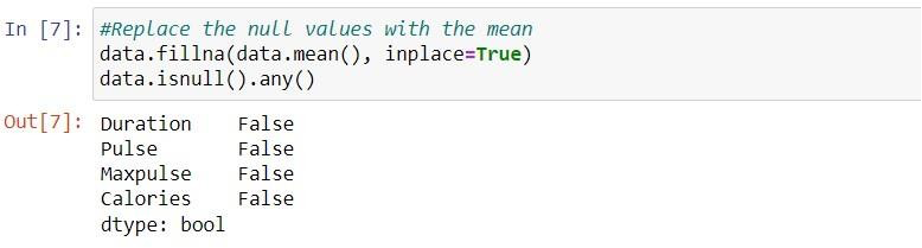
****

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

**A picture containing text, font, line, screenshot

Description automatically generated**

**a. Replace the null values with the mean**

****

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

**A screenshot of a computer

Description automatically generated with low confidence**

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

**A screenshot of a calories chart

Description automatically generated with low confidence**

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

**A screenshot of a computer

Description automatically generated with medium confidence**

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

**A screenshot of a computer

Description automatically generated with low confidence**

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

**A screenshot of a computer

Description automatically generated with low confidence**

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

**A screenshot of a computer

Description automatically generated with medium confidence**

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

**A screen shot of a graph

Description automatically generated with low confidence**

**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.**

**A screenshot of a computer

Description automatically generated with medium confidence**

**2.** **2. Evaluate the model on testing part using score and**

**Classification\_report(y\_true,y\_pred)**

**A screenshot of a computer

Description automatically generated with medium confidence**

**A screenshot of a computer screen

Description automatically generated with low confidence**

**Do at least two visualizations to describe or show correlations in the Glass Dataset. Which algorithm you got better accuracy? Can you justify why?**

**A screen shot of a computer code

Description automatically generated with low confidence**

**A picture containing text, screenshot, square, diagram

Description automatically generated**

**A picture containing screenshot, text, diagram, line

Description automatically generated**