**CS 5710 Machine Learning**

**ASSIGNMENT 1**

**Submitted by:**

**Vinod Yadav Kannaboina**

**Student Id: 700748732**

**CRN: 23921**

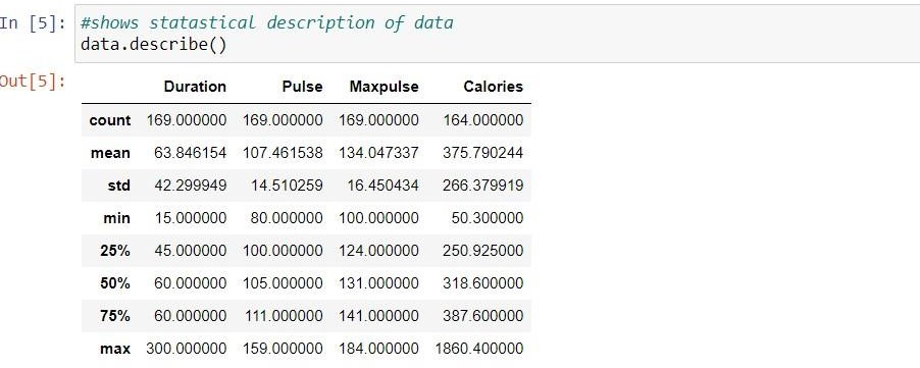
**Video Link:** [**https://drive.google.com/file/d/1EPiKW39dVB9Ba3fdDAKKyB1A-ZNguEhN/view?usp=share\_link**](https://drive.google.com/file/d/1EPiKW39dVB9Ba3fdDAKKyB1A-ZNguEhN/view?usp=share_link)

**Q1) Pandas**

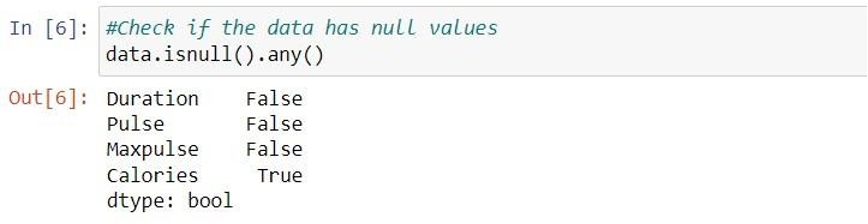
**1.** **Read the provided CSV file ‘data.csv’.**

****

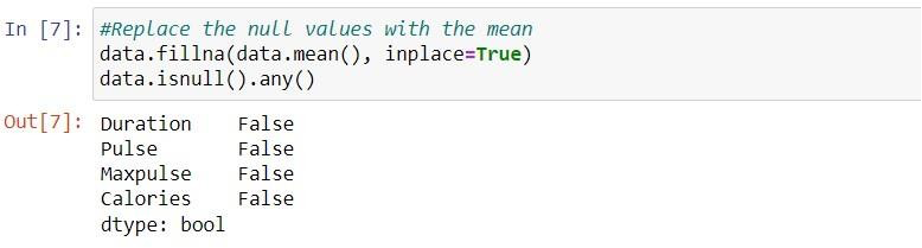
**2.** **Show the basic statistical description about the data.**

****

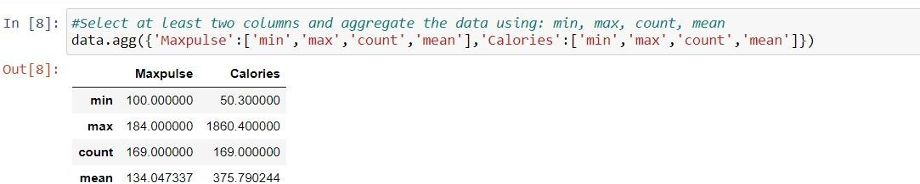
**3.** **Check if the data has null values.**

****

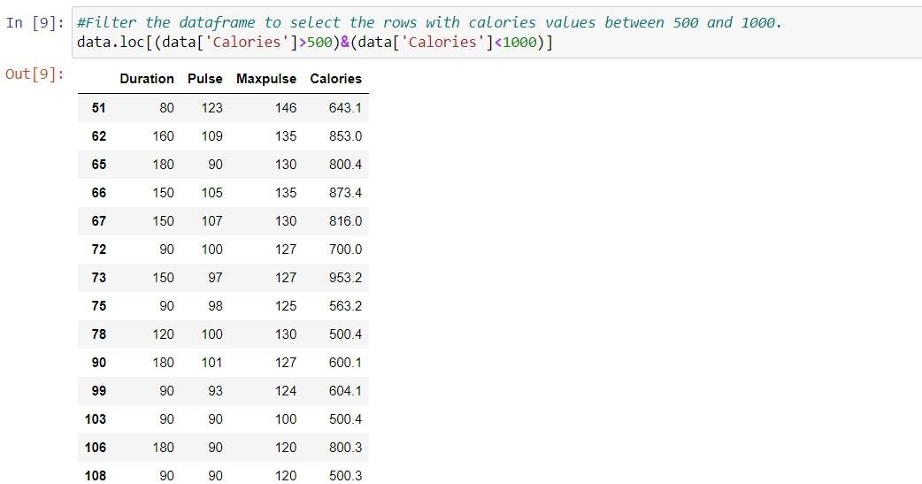
**a. Replace the null values with the mean**

****

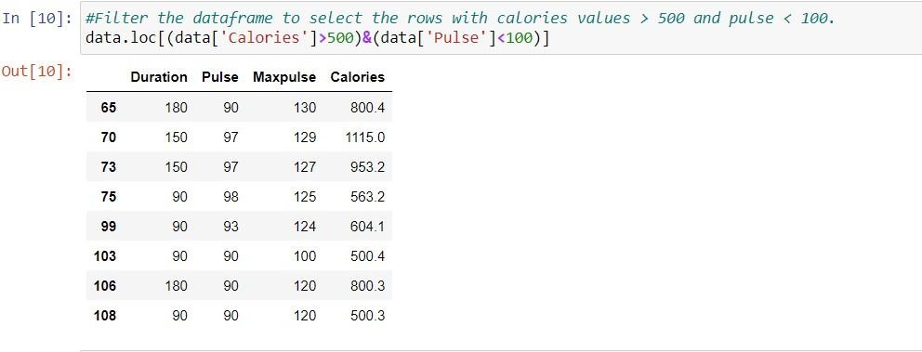
**4.** **Select at least two columns and aggregate the data using: min, max, count, mean**

****

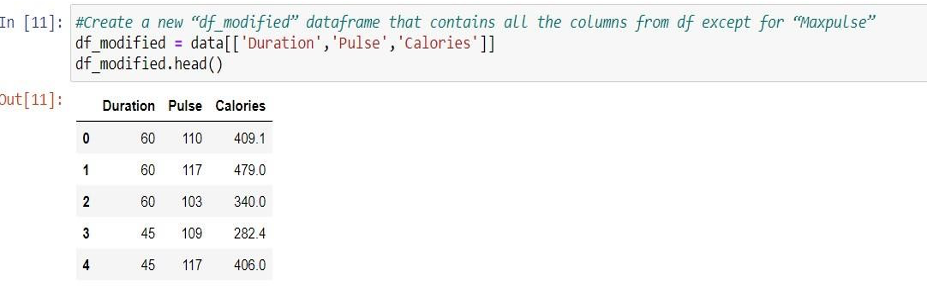
**5.** **Filter the dataframe to select the rows with calories values between 500 and 1000.**

****

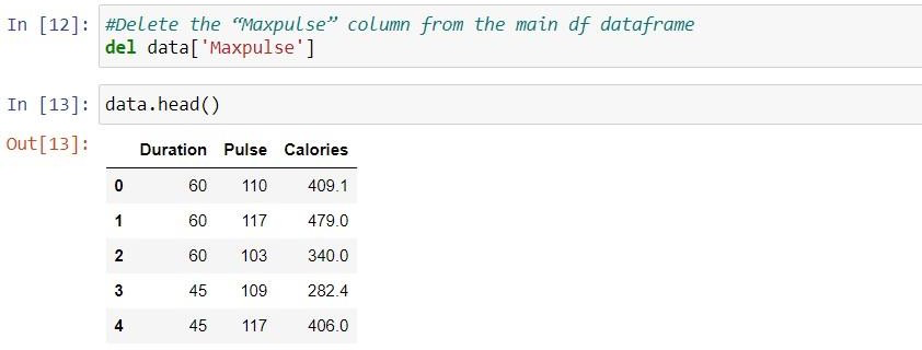
**6.Filter the dataframe to select the rows with calories values > 500 and pulse < 100**

****

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

****

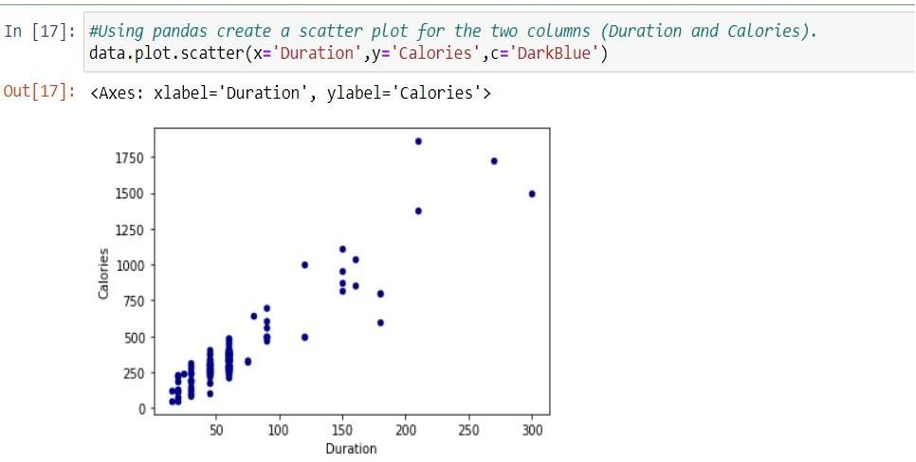
**8.Delete the “Maxpulse” column from the main df dataframe**

****

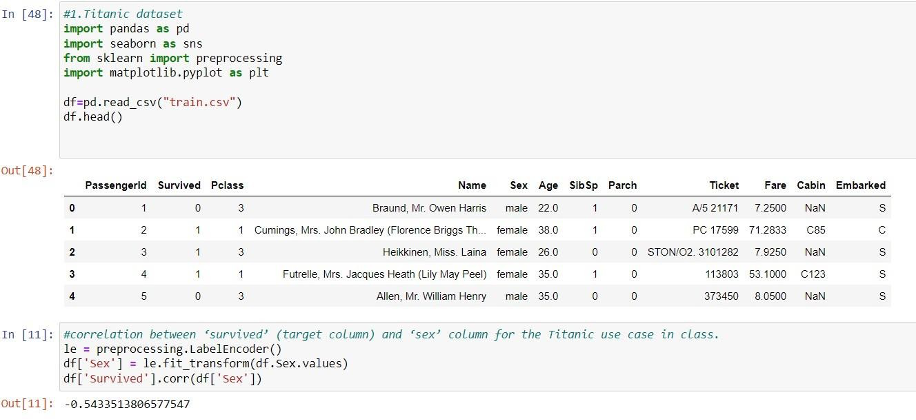
**9.** **Convert the datatype of Calories column to int datatype.**

****

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

****

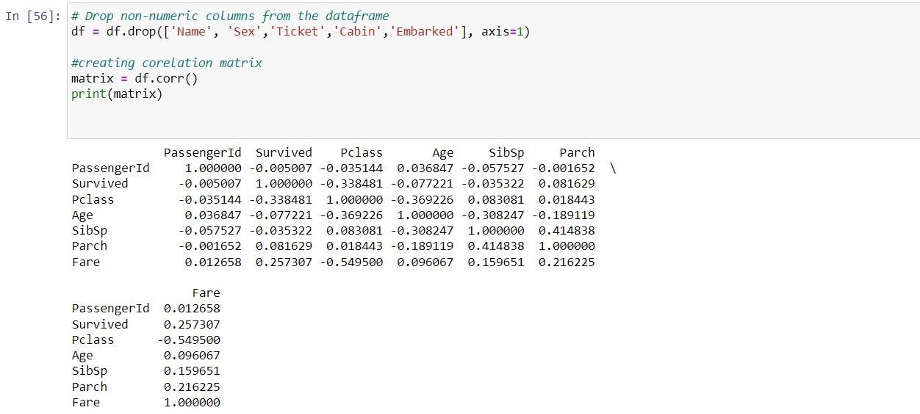
**1.** **(Titanic Dataset) 1. Find the correlation between ‘survived’ (target column) and ‘sex’ column for the Titanic use case in class**

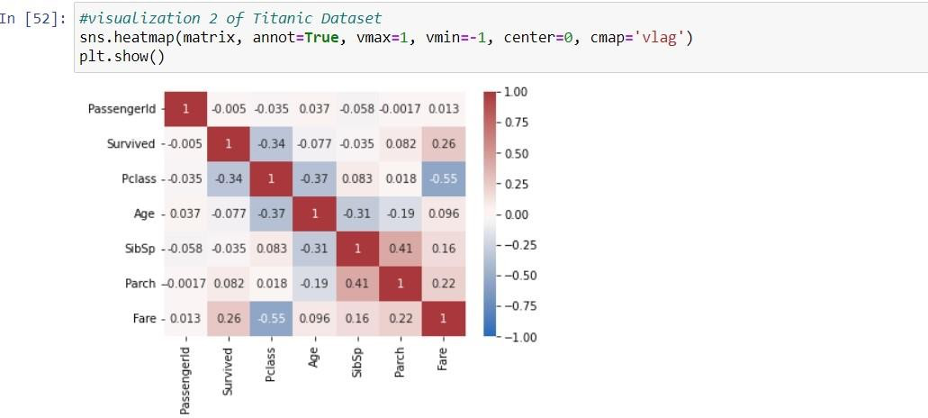
****

**a.** **Do you think we should keep this feature?**

**Ans. No , we shouldn’t keep this feature as the accuracy obtained is only 54%.**

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

****

****

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

****