**MACHINE LEARNING ASSIGNMENT-4**

Jayanth Sri Sai Dulla

700734068

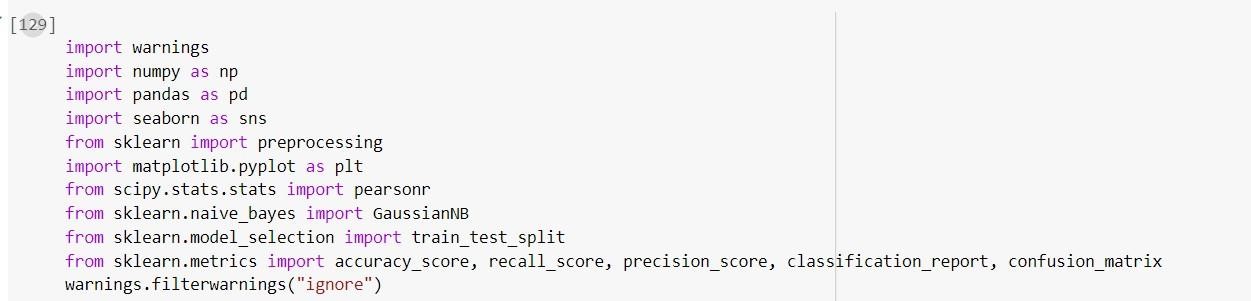
Video link: https://drive.google.com/file/d/19fgUgHuRgNMDQ6QclqV7pBnM69Rw28LR/view?usp=share\_link

Github link: https://github.com/Jayanthsrisaidulla1998/ML-ASSIGNMENT-4

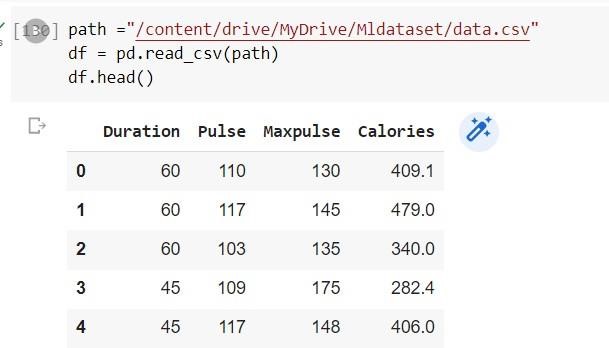
1.Pandas:

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

https://drive.google.com/drive/folders/1h8C3mLsso-R-sIOLsvoYwPLzy2fJ4IOF?usp=sharing

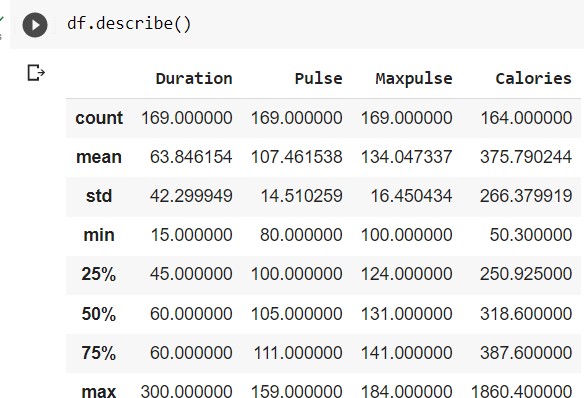


Importing all the required libraries.



The pd.read csv() function of the Pandas library reads a CSV file and creates a DataFrame object from it.

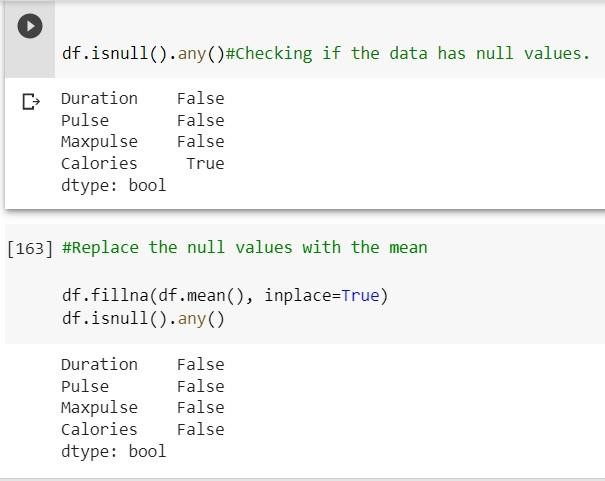
2. Show the basic statistical description about the data.



Pandas has a built-in function called df.describe() that generates descriptive statistics for the DataFrame df.

3. Check if the data has null values.

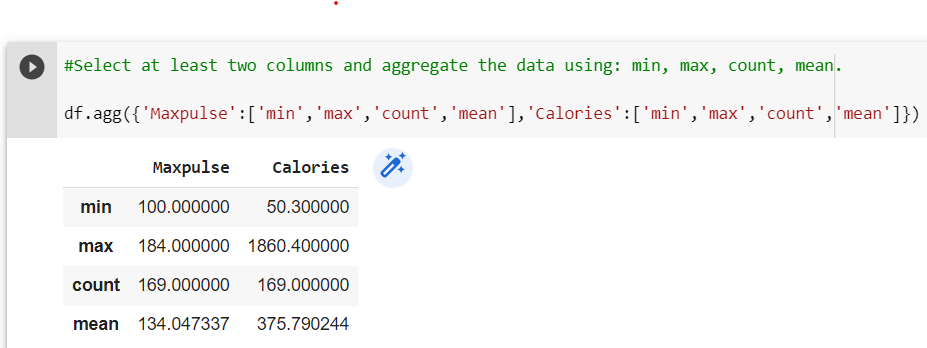
a. Replace the null values with the mean



Any missing values in the DataFrame df are filled in by the df.fillna() method using the mean value of each column.

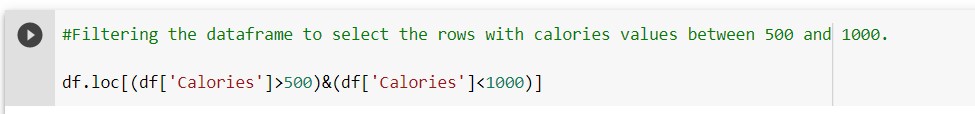
A df.isnull ().any() is used to determine whether the DataFrame still contains any missing values.

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



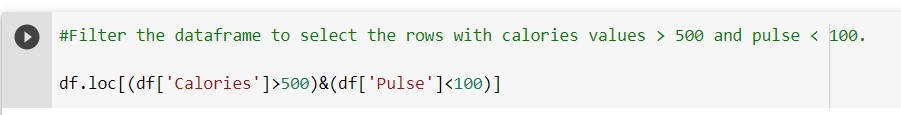
The agg() method is used to do the aggregate computations of the Maxpulse and Calories on the dataframe df.

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



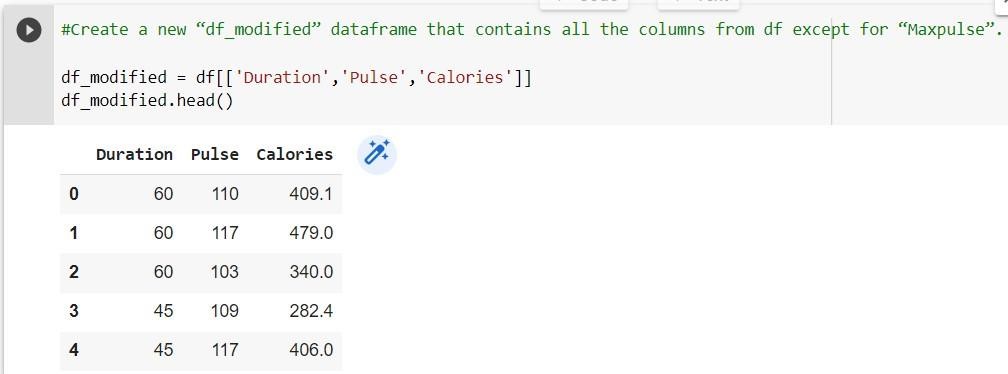
To choose the rows with calorie values between 500 and 1000, use the df.loc() function.

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



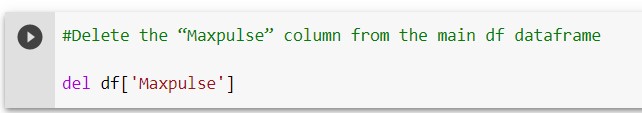
The df.loc() function is used to select the rows with calories values that are greater than 500 and the pulse value less than 100.

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

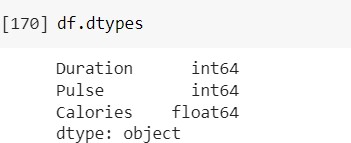


Just the Duration, Pulse, and Calories columns of the original DataFrame df are included in the newly formed DataFrame df modified.

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

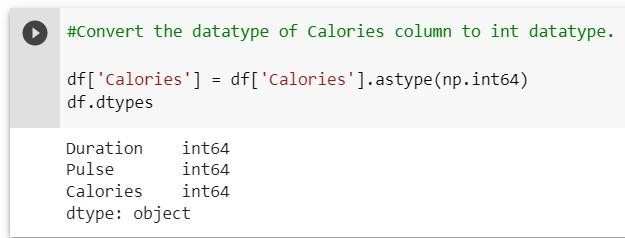


The Maxpulse column is removed from the dataframe df using del df[‘MaxPulse’]



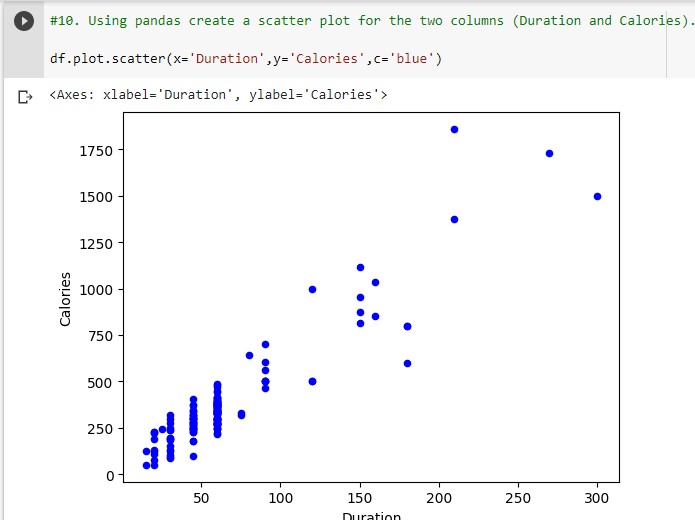
The df.dtypes is used to display the data types of each column in the DataFrame.

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

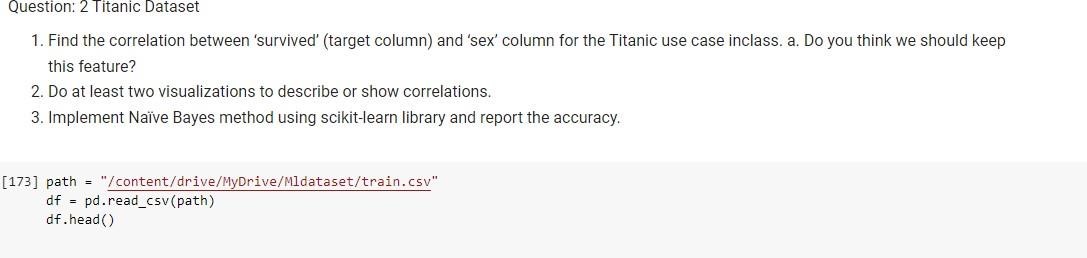


To convert a column in a DataFrame to a particular data type, use the astype() method. Here, the np.int64 data type is used to convert the calories column to a 64-bit integer data type.

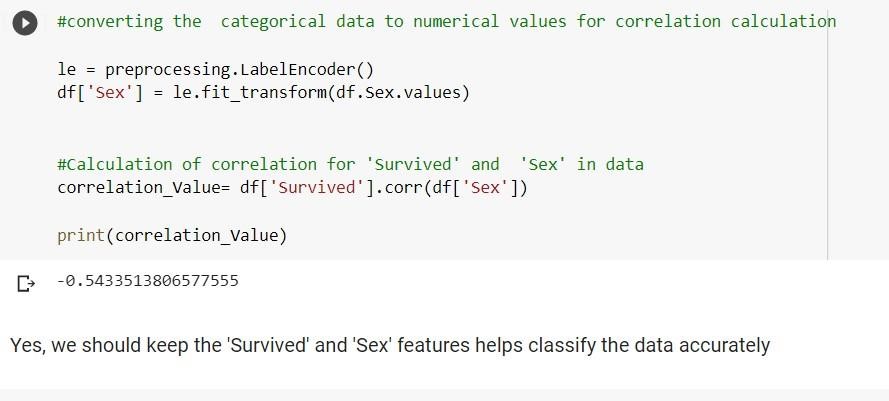
10. Using pandas create a scatter plot for the two columns (Duration and Calories)

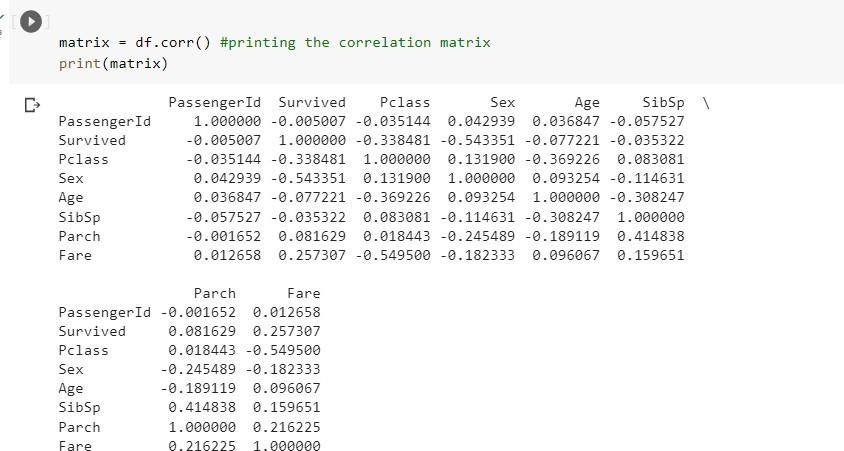


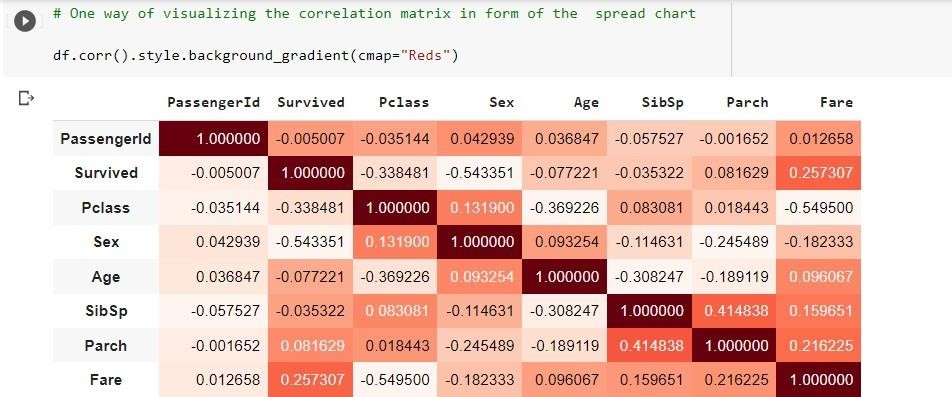
The scattered plot of the Duration and Calories columns in the DataFrame df is displayed.



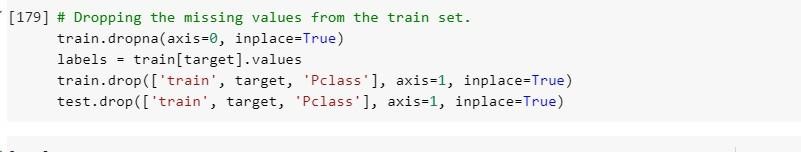


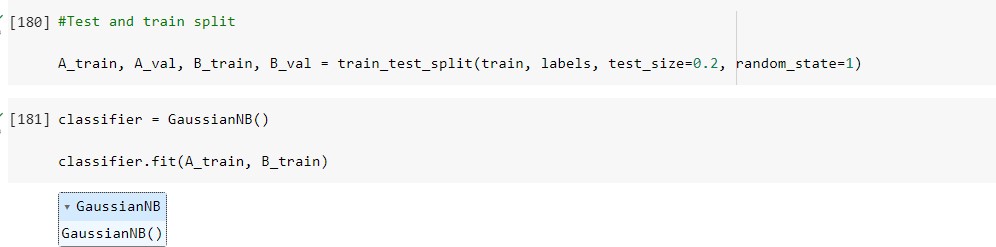


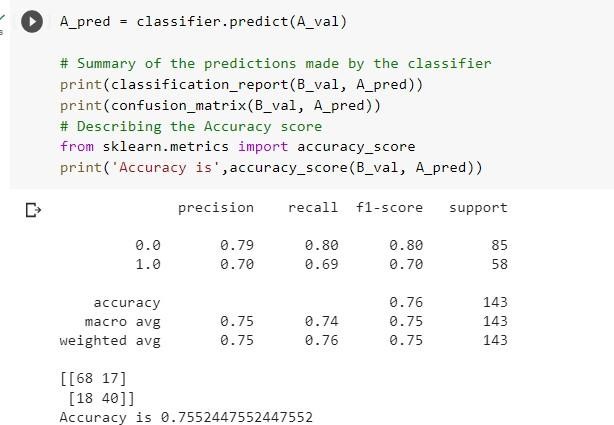












Question-3:

