**MACHINE LEARNING ASSIGNMENT - 2**

# Name: Bommareddy Eswar reddy Student ID: 700748146

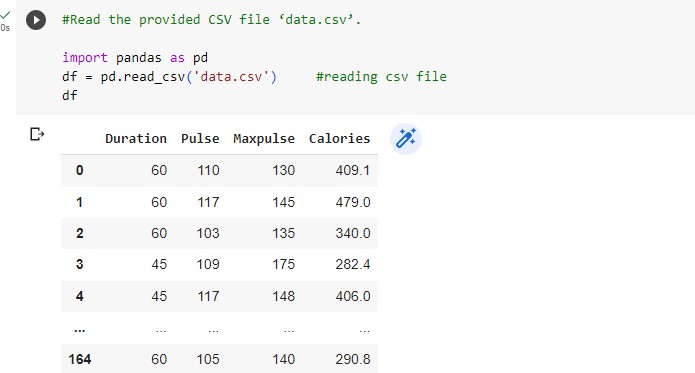
**Git hub Link :- https://github.com/SiddarthPaladi/ML\_Assignment2.git**

# Video Link: -

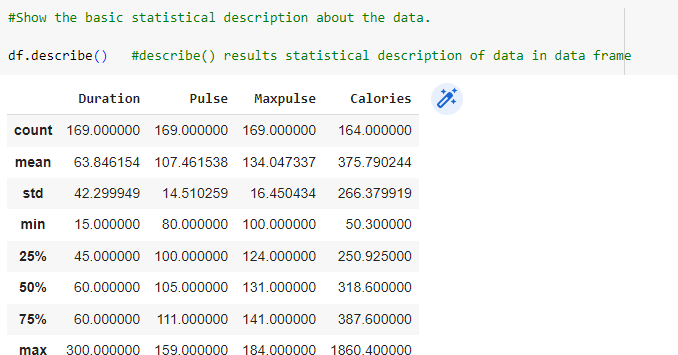
<https://drive.google.com/file/d/1leeQyCmlPXVN-T7aYqnOuFpVEZ5xOnVk/view?usp=sharing>

# Pandas

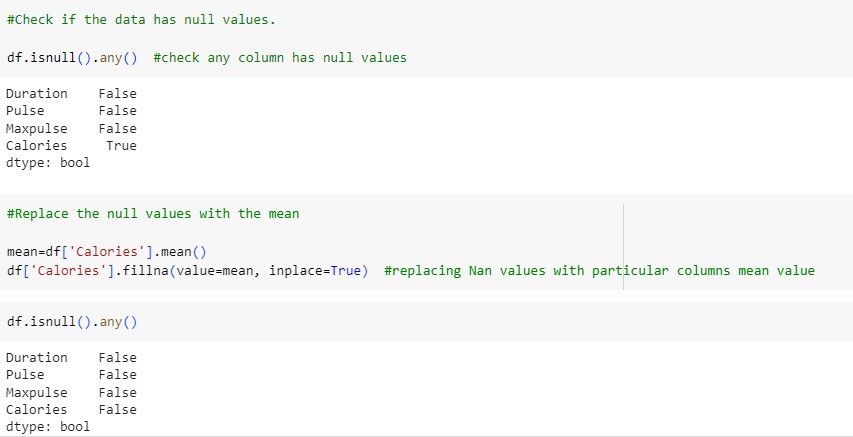
* 1. **Read the provided CSV file ‘data.csv’. [https://drive.google.com/drive/folders/1h8C3mLsso-R-sIOLsvoYwPLzy2fJ4IOF?usp=sharin](https://drive.google.com/drive/folders/1h8C3mLsso-R-sIOLsvoYwPLzy2fJ4IOF?usp=sharing) [g](https://drive.google.com/drive/folders/1h8C3mLsso-R-sIOLsvoYwPLzy2fJ4IOF?usp=sharing)**

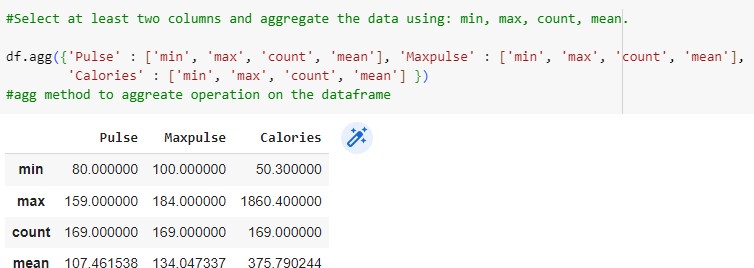


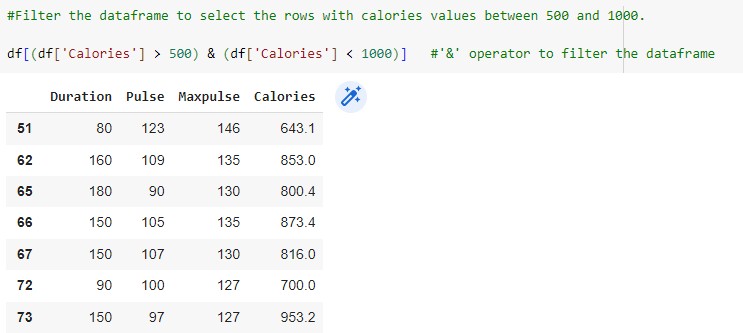
* 1. **Show the basic statistical description about the data.**

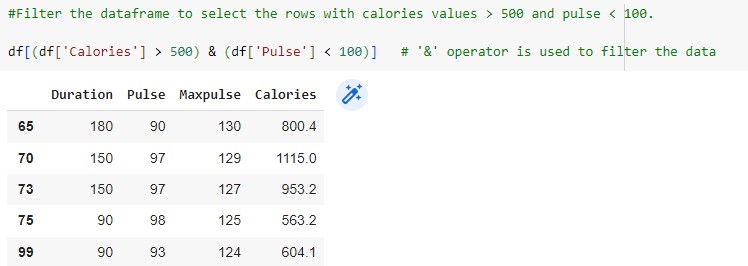
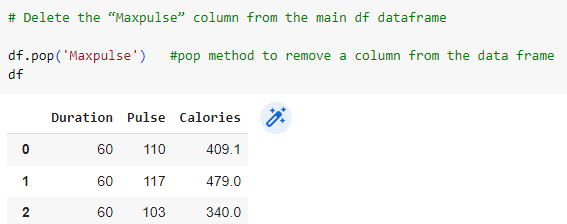


* 1. **Check if the data has null values. a. Replace the null values with the mean**

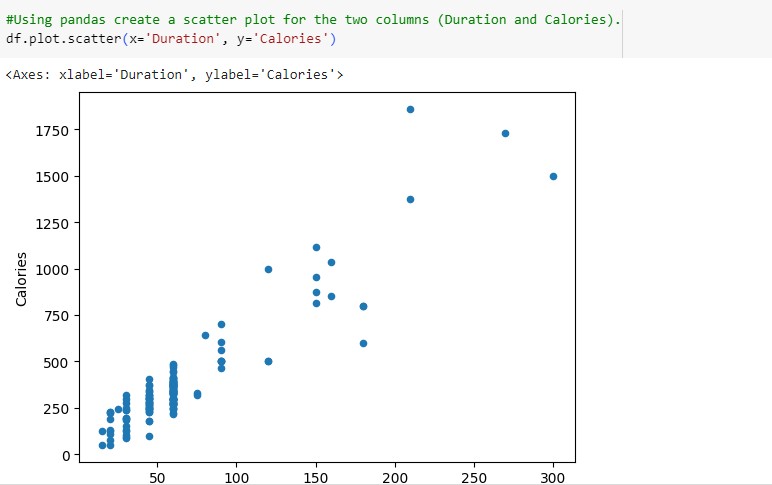


* 1. **Select at least two columns and aggregate the data using: min, max, count, mean.**
  2. **Filter the dataframe to select the rows with calories values between 500 and1000. 6. Filter the dataframe to select the rows with calories values > 500 and pulse <100.**



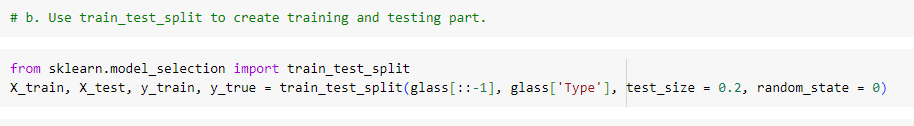
1. **Create a new “df\_modified” dataframe that contains all the columns from df except for “Maxpulse”.**
2. **Delete the “Maxpulse” column from the main df dataframe**
3. **Convert the datatype of Calories column to int datatype.**



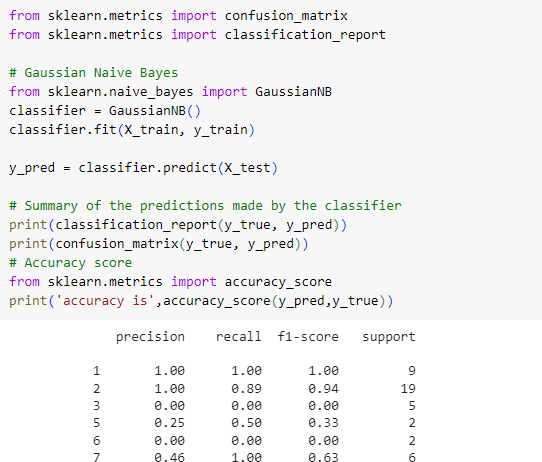
1. **Using pandas create a scatter plot for the two columns (Duration and Calories).**

# Scikit-learn

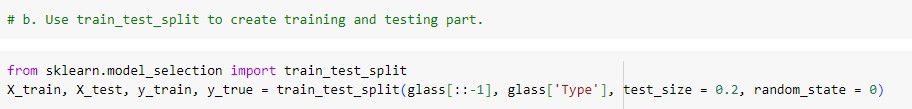
* 1. **Implement Naïve Bayes method using scikit-learnlibrary.**

a. Use the glass dataset available in Link also provided in your assignment. b. Use train\_test\_split to create training and testing part.

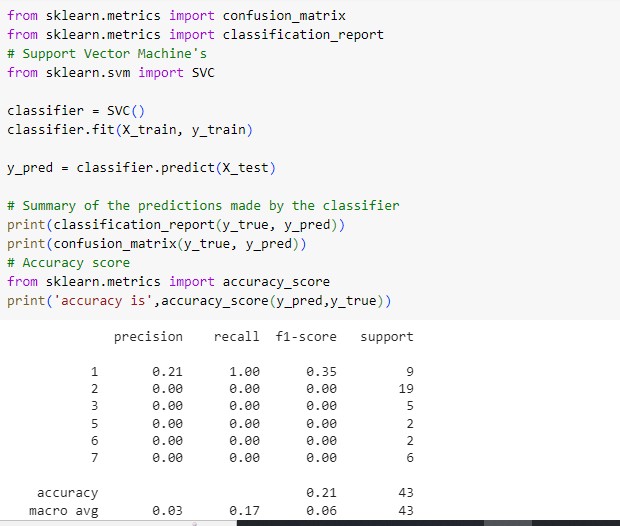
* 1. **Evaluate the model on testing part using score and**



1. **Implement linear SVM method using scikit 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.

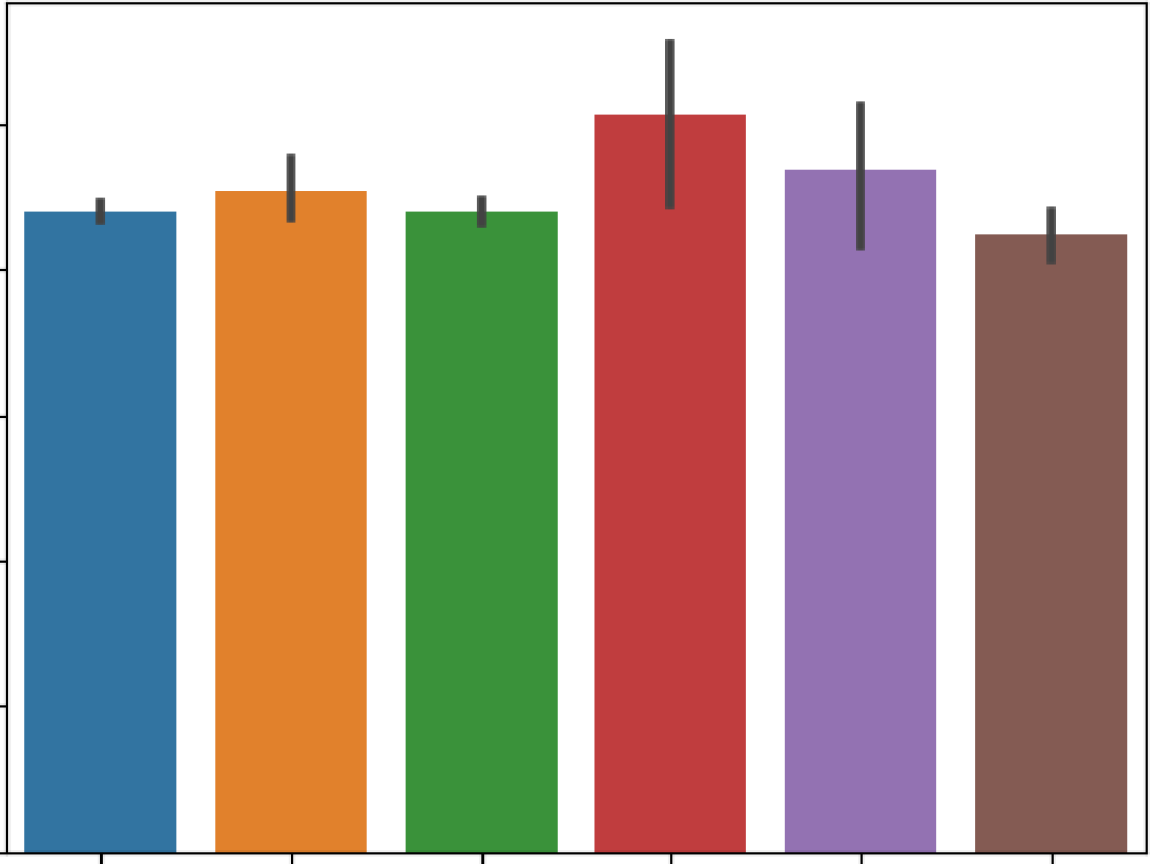
1. **Evaluate the model on testing part using score and**



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

import seaborn as sns #For Visualisation import seaborn library impo t matplotlib.pyplot as plt

sns. barplot(x = glass[·Type' J, y = g1ass[ ·ca·J)

<Axes: xlabe1="Type·', ylabel='Ca'>

10

8

uro 6

4

2

0

1 2 3 5 6 7

Tvoe

sns.catplot(data=glass, x="Type", y="K")

<seaborn.axisgrid.FacetGrid at 0x7fb 0938ab00>

-

6

5

4

::..:: 3

•

2

1 ..

0

ft

•

• ••

•

•t

-.:

't •

•

.•.

•

-

•

•

sns. regplot(x="Type", y="Fe",, data=g1ass);

0.5

•

I

•

I

•

•

...

••

••

•

I

•

•

I

•

•I

I

•

•

I

••

l

•

2

•

3

**4**

•

5

•

6

••

I

7

*a:.*

**0.4**

**0.3**

**0.2**

**0.1**

**0.0**

Type