# Fall 2022: CS5710 – Machine Learning In-Class Programming Assignment-2 – 700734184

**1. Numpy: Using NumPy create random vector of size 15 having only Integers in the range 1-20.**

**1. Reshape the array to 3 by 5**

**2. Print array shape.**

**3. Replace the max in each row by 0**

Graphical user interface, application

Description automatically generated

**2. Pandas**

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

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

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

Graphical user interface, text, application, chat or text message

Description automatically generated

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

Graphical user interface, application

Description automatically generated

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

Table

Description automatically generated with low confidence

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

Background pattern

Description automatically generated with medium confidence

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

Graphical user interface, text, application

Description automatically generated

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

Graphical user interface, application

Description automatically generated

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

Graphical user interface, text, application

Description automatically generated

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

Graphical user interface, text, application

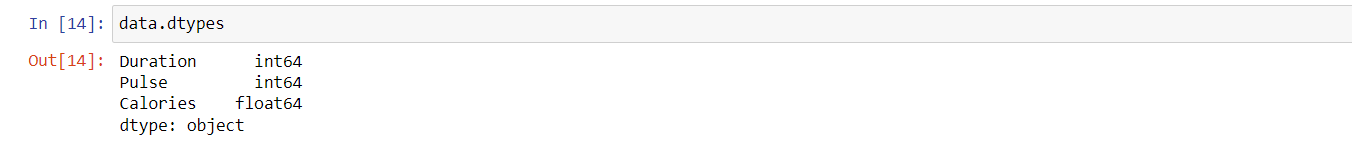
Description automatically generated

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

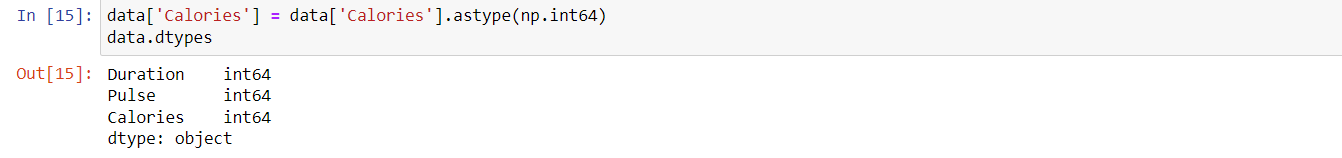


Shape

Description automatically generated with low confidence



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



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

Graphical user interface

Description automatically generated

**3. Matplotlib**

**1. Write a Python programming to create a below chart of the popularity of programming Languages.**

**2. Sample data:**

**Programming languages: Java, Python, PHP, JavaScript, C#, C++**

**Popularity: 22.2, 17.6, 8.8, 8, 7.7, 6.7**

Chart

Description automatically generated with medium confidence