**Fall 2022: CS5710 – Machine Learning**

**In-Class Programming Assignment-2**

**ID: 700743830**

**NAME: SAI KRISHNA REDDY CHEVUTUKUR**

**USERNAME: SXC38300**

**GIT HUB LINK:**

<https://github.com/csk17/ML-ASSIGNMENT-2>

**SOURCE CODE LINK:**

<http://localhost:8891/notebooks/700743830%20Assignment%202.ipynb>

**VIDEO LINK:**

<https://drive.google.com/file/d/1WKLWJtqjEqtjcgv9IJ7VfGHakF1sThKX/view?usp=sharing>

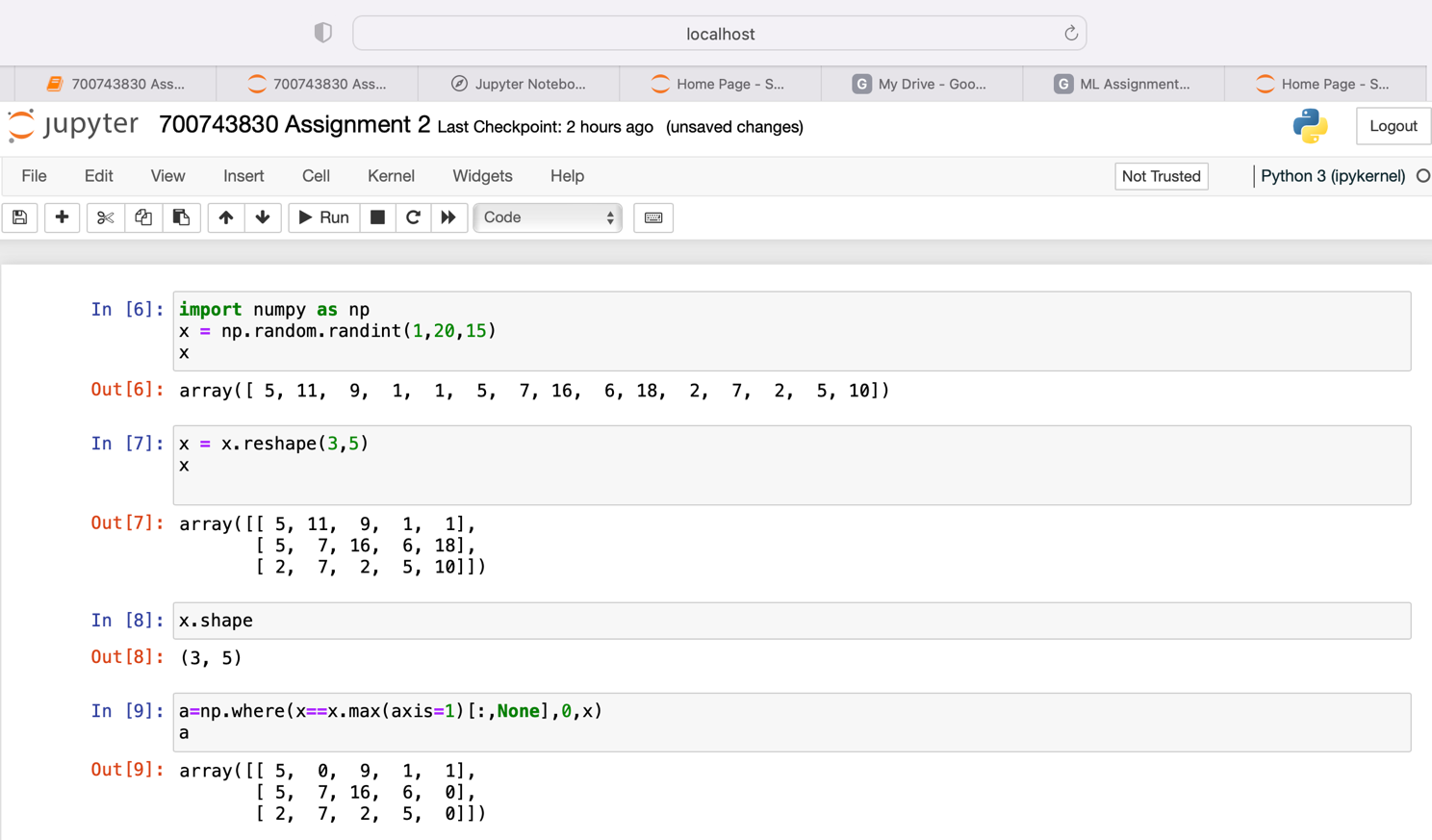
**1.Numpy:**

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**

**SOLUTION:**

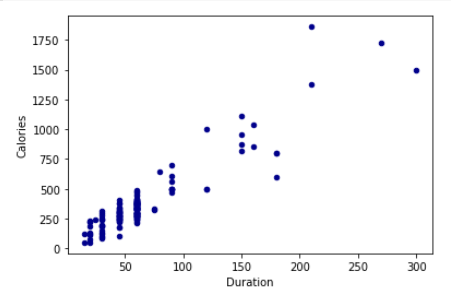
* Here, we are creating random vector of size 15 having only integers in the range 1-20 by importing numpy as np with variable x.
* Reshaping the array to 3 by 5, where number of rows are 3 and coloumns are 5 by reshape() function.
* Now print the array shape by function shape().
* Now, replace the max value in each row by 0 using max() function equal to 0. We get the updated array as output a.

****

**2. 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.
   1. 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).

a. Example:



**SOLUTION:**

* Here, we get the data frame from the above given link and read the file by importing pandas.
* And we found the null values in Calories attributes as those need to be replaced by mean value by obtaining the mean of Calories by using the function mean().
* We have selected two coloumns Duration and Pulse from the data frame and found the aggregate of max, min, mean and count.
* We filter the dataframe to select the rows of calories between 500 to 1000 by using ‘&’ operator.
* We filter the dataframe to select the rows of calories > 500 and pulse<100 by using ‘&’ operator.
* We have modified the dataframe by excepting the maxpulse attribute by using drop function.
* Similarly, we have deleted Maxpulse attribute from the original dataframe.
* Finally, converted the datatype of calories from float to integer using astype() function.
* Created a scatter plot for duaration and calories by importing matplot library as plt.

**Graphical user interface

Description automatically generated**

**Graphical user interface

Description automatically generated**

**Table

Description automatically generated with medium confidence**

**Graphical user interface, text, application, email

Description automatically generatedGraphical user interface

Description automatically generatedGraphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application, email

Description automatically generatedChart, scatter chart

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

**SOLUTION:**

* A New dataframe is created by using the sample data in python language. And, a pie chart is created for the following data and autopct function is used to display the value in a particular region with percentage.

**Graphical user interface, text, application, email

Description automatically generated**