ML ASSIGNMENT

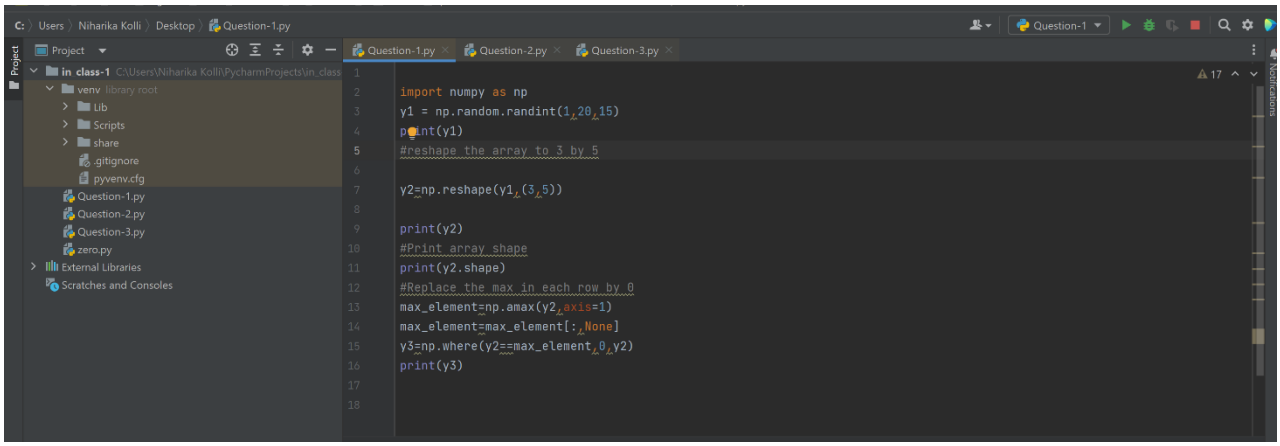
NAME: Sai Pavana Niharika Kali

STUDENT ID: 700740603

GITHUB LINK:
https://drive.google.com/file/d/1R4r0oe7QP7qEdHELWk5D8yX7GcQzrDa0/view?usp=sharing

1. NumPy:

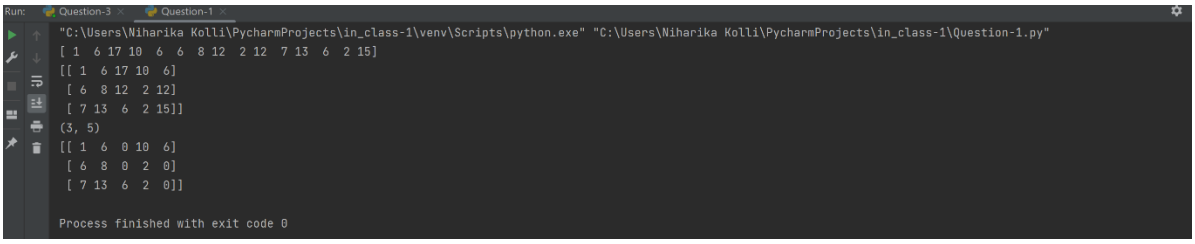
Using NumPy create random vector of size 15 having only Integers in the range 1-20. 1. Reshape the array to 3 by 5 2. Print array shape. 3. Replace the max in each row by 0



EXPLAINATION:

1.initially we are importing NumPy
2.then were using NumPy and creating a random vector of size 15 having only integers in the range 1-20
3.and then reshaping the array to 3 by 5
4.printing the array's shape
5.and then were replacing the max value in each row with 0

OUTPUT:



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.

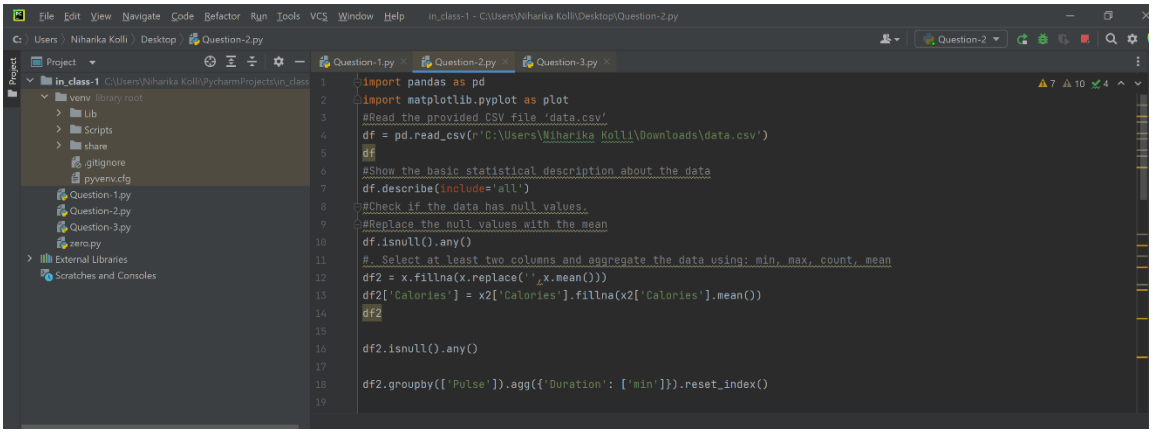
6. Filter the dataframe to select the rows with calories values > 500 and pulse < 100.

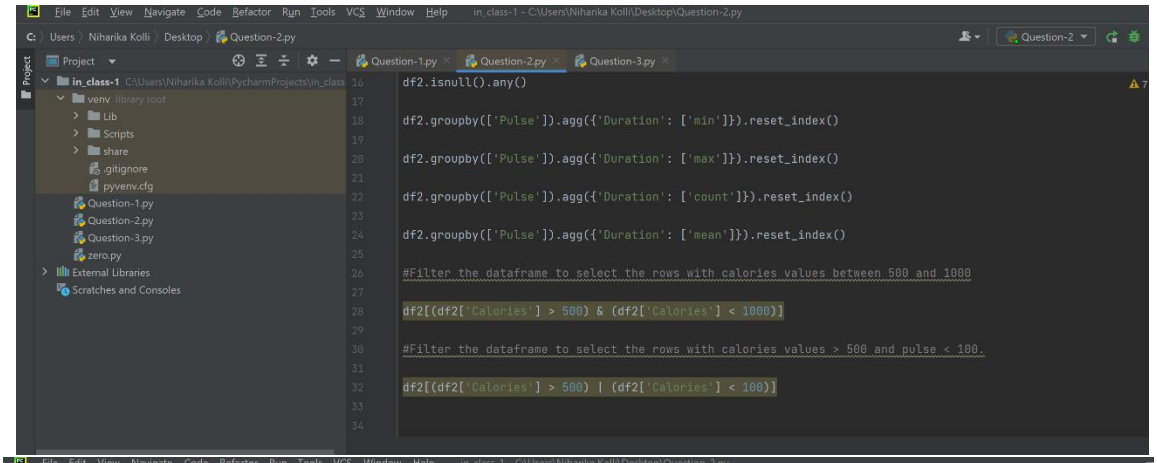
7. Create a new "df_modified" dataframe that contains all the columns from df except for "Maxpulse".

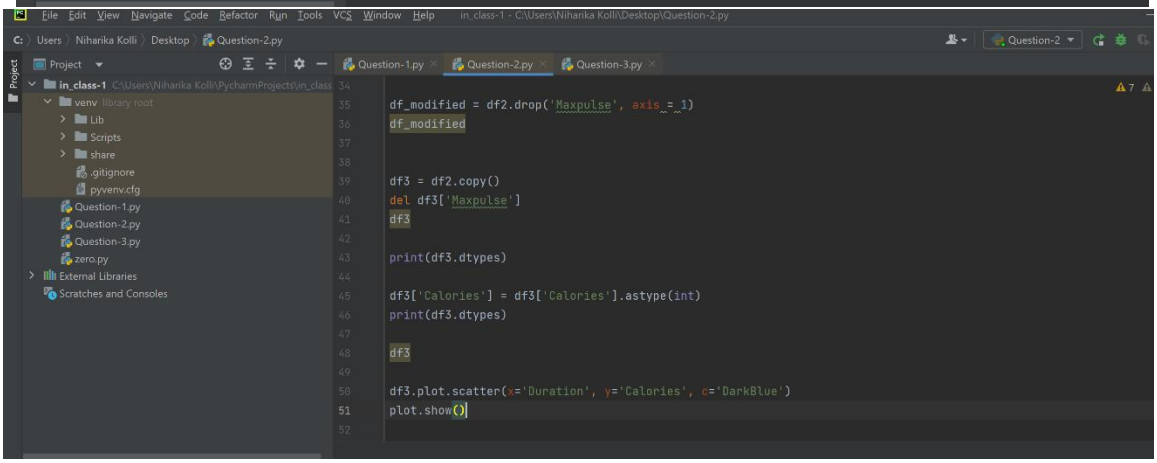
8. Delete the "Maxpulse" column from the main df dataframe

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

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



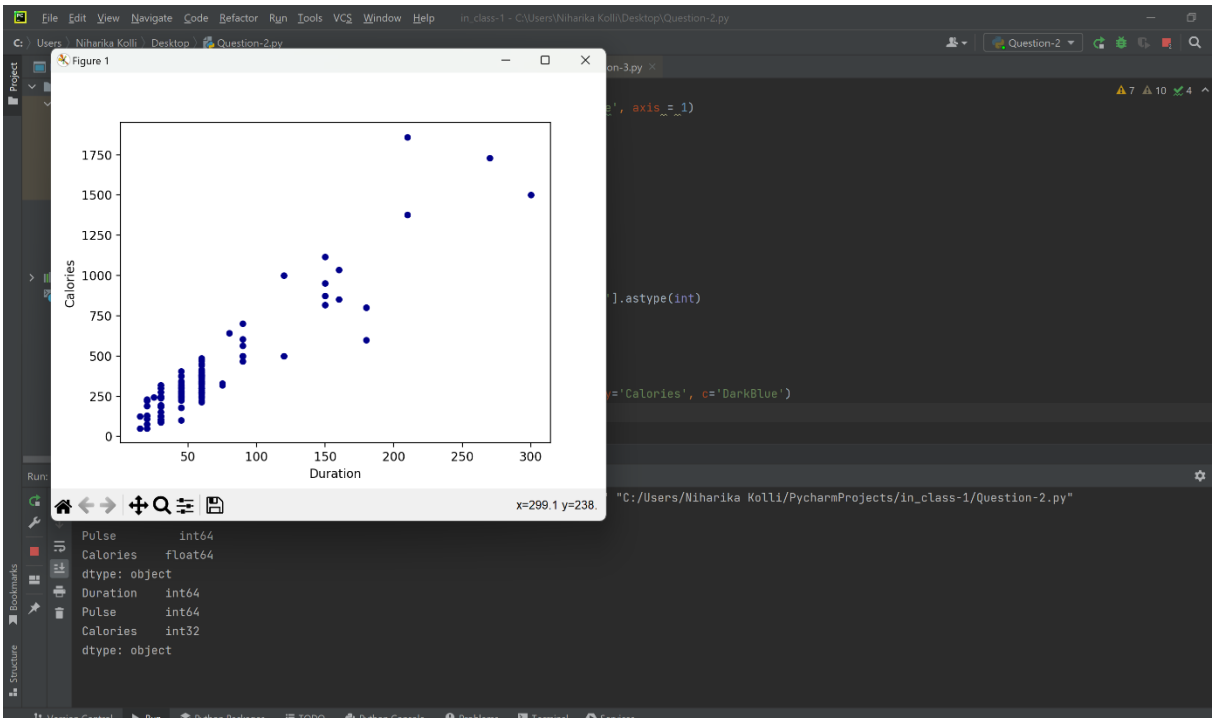




EXPLAINATION:

1.importing pandas
2.we're reading the dataset
3.then we are showing the basic statistical description about the data
4.and then we check if the data has null values
5.and replace the null values with the mean value
6.and select at least 2columns and aggregate the data using:min,max,count,mean
7.filter the dataframe to select the rows with calories value between 500,1000
8.filter the dataframe to select rows with calories values more than 500 and pulse less than 100
9.create a new modified dataframe which is x-modified that contains all the columns of data frame except that of maxpulse
10.delete maxpulse column from main data frame
11.convert the data type calorie column to int data type
12.using the pandas create a scatter plot for the 2 columns

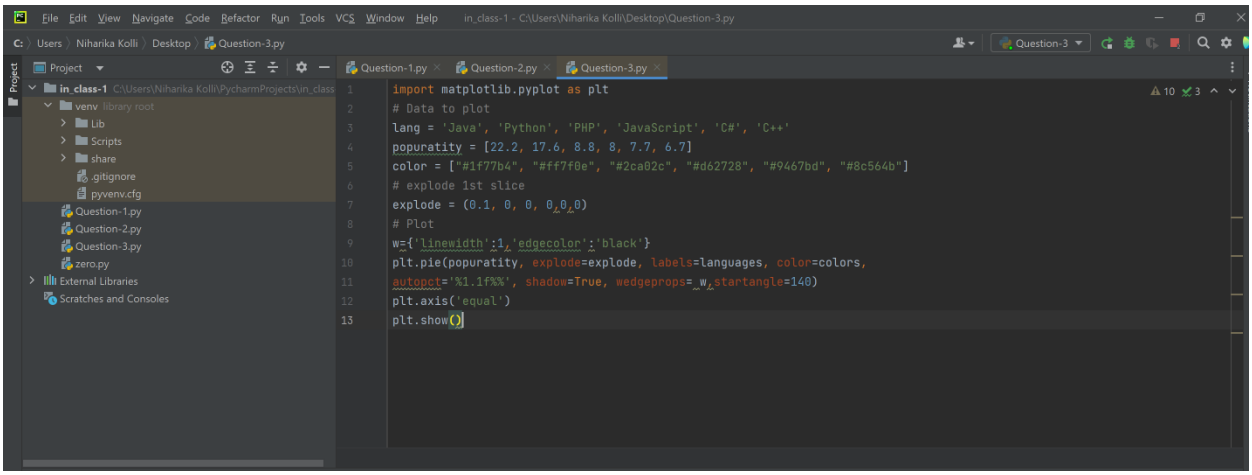
OUTPUT:



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



EXPLAINATION:

1.Initially were importing the package matplotlib as plt

2.Next we are considering the data to plot

3.Assigning the lang variable with languages Color variable with the colour codes

4.And then were exploding first slice of the pie chart

5.then we slice the axis equally in the pie chart and then display the chart

OUTPUT: