

ML ASSIGNMENT

NAME: Sai Pavana Niharika Kali

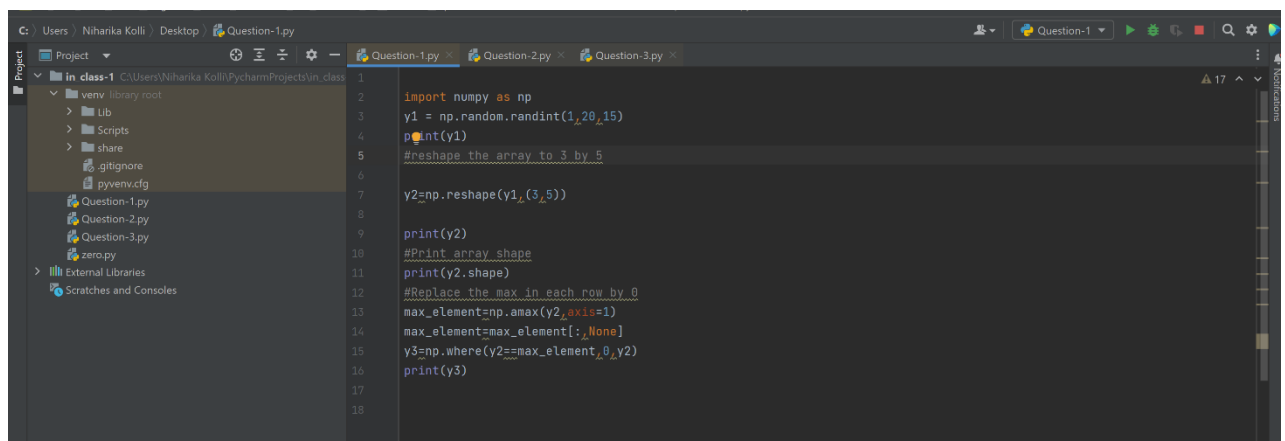
STUDENT ID: 700740603

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

A screenshot of a PyCharm IDE window. The left sidebar shows a project structure with folders like 'venv', 'Lib', 'Scripts', 'share', and files like '.gitignore', 'pyvenv.cfg', and 'Question-1.py'. The main editor area shows a Python script for 'Question-1.py'. The script imports numpy as np, creates a random vector y1 of size 15 with integers from 1 to 20, prints y1, reshapes it to a 3x5 array y2, prints y2, prints the array shape, finds the maximum value in each row, and replaces those maximum values with 0 to create y3, which is then printed.

```
1 import numpy as np
2 y1 = np.random.randint(1,20,15)
3 print(y1)
4 #Reshape the array to 3 by 5
5 y2=np.reshape(y1,(3,5))
6
7 print(y2)
8 #Print array shape
9 print(y2.shape)
10 #Replace the max in each row by 0
11 max_element=np.amax(y2,axis=1)
12 max_element=max_element[:,None]
13 y3=np.where(y2==max_element,y2,0)
14 print(y3)
15
16
17
18
```

EXPLANATION:

- 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:

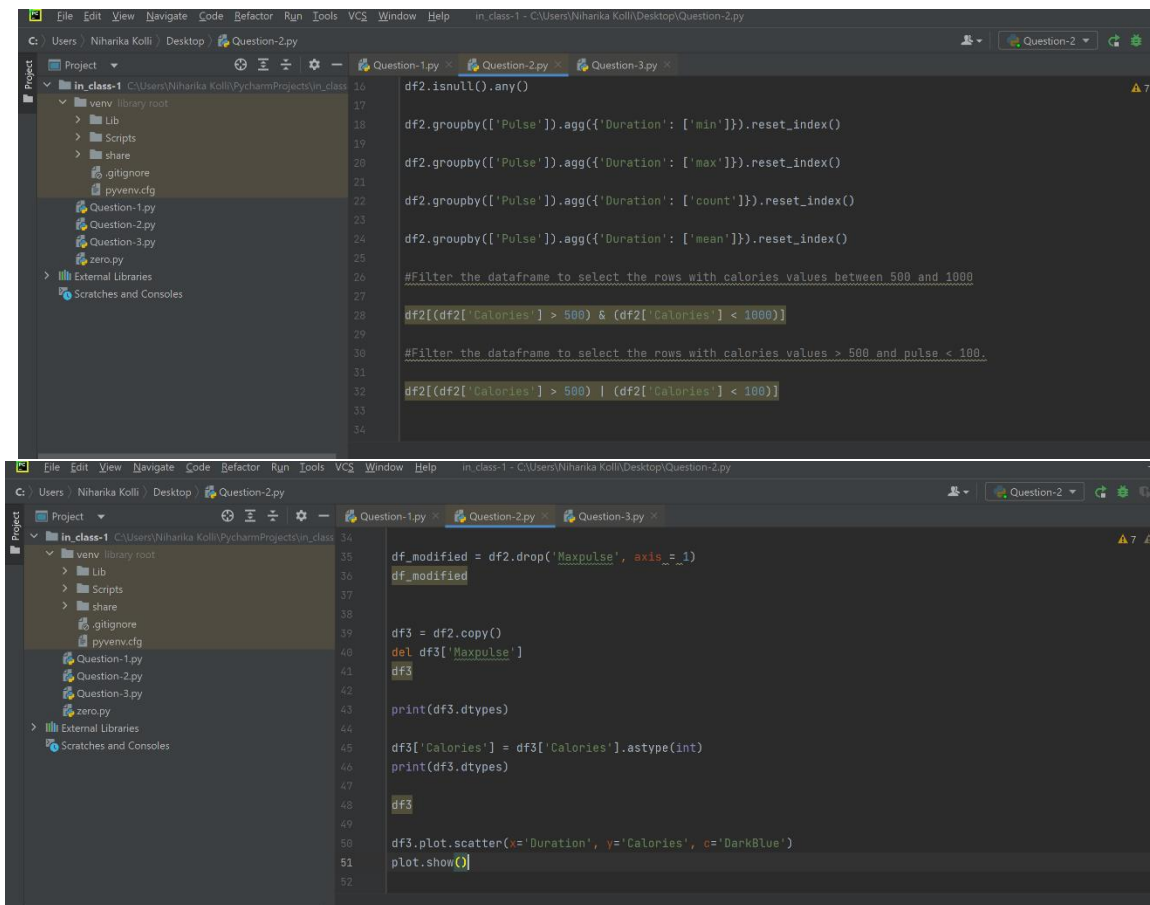
```
Run: Question-3 x Question-1 x
"C:\Users\Niharika Kolli\PycharmProjects\in_class-1\venv\Scripts\python.exe" "C:\Users\Niharika Kolli\PycharmProjects\in_class-1\Question-1.py"
[ 1  6 17 10  6  6  8 12  2 12  7 13  6  2 15]
[[ 1  6 17 10  6]
 [ 6  8 12  2 12]
 [ 7 13  6  2 15]]
(3, 5)
[[ 1  6  0 10  6]
 [ 6  8  0  2  0]
 [ 7 13  6  2  0]]
Process finished with exit code 0
```

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).

```
File Edit View Navigate Code Refactor Run Tools VCS Window Help in_class-1 - C:\Users\Niharika Kolli\Desktop\Question-2.py
Project
  in_class-1 C:\Users\Niharika Kolli\PycharmProjects\in_class-1
    venv library root
    Lib
    Scripts
    share
    gitignore
    pyenv.cfg
  Question-1.py
  Question-2.py
  Question-3.py
  xamp.py
External Libraries
Scratches and Consoles

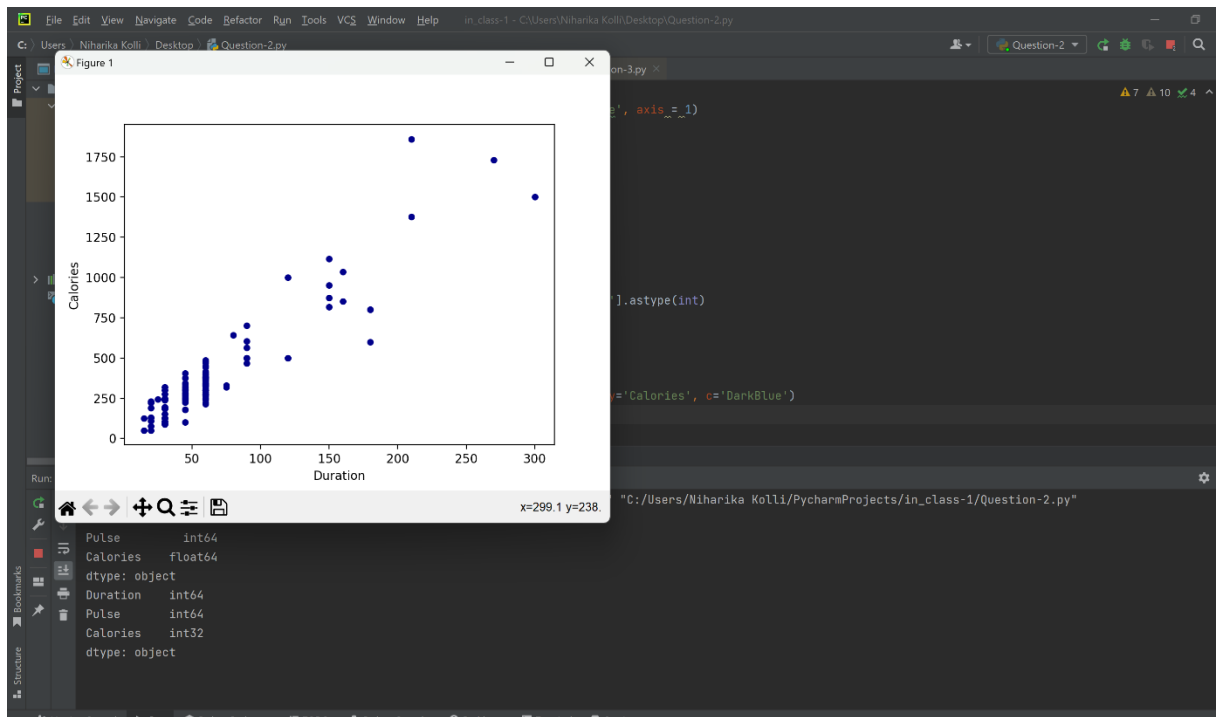
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 #Read the provided CSV file 'data.csv'
4 df = pd.read_csv(r"C:\Users\Niharika Kolli\Downloads\data.csv")
5 df
6 #Show the basic statistical description about the data
7 df.describe(include='all')
8 #Check if the data has null values.
9 #Replace the null values with the mean
10 df.isnull().any()
11 #. Select at least two columns and aggregate the data using: min, max, count, mean
12 df2 = x.fillna(x.replace(' ',x.mean()))
13 df2['Calories'] = x2['Calories'].fillna(x2['Calories'].mean())
14 df2
15
16 df2.isnull().any()
17
18 df2.groupby(['Pulse']).agg({'Duration': ['min']}).reset_index()
19
```



EXPLANATION:

- 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

```
File Edit View Navigate Code Refactor Run Tools VCS Window Help in_class-1 - C:\Users\Nihanka Kolli\Desktop\Question-3.py
Project
  in_class-1 C:\Users\Nihanka Kolli\PycharmProjects\in_class-1
    venv library root
      Lib
      Scripts
      share
      gitignore
      pyvenv.cfg
    Question-1.py
    Question-2.py
    Question-3.py
    zero.py
  External Libraries
  Scratches and Consoles
Question-3.py
1 import matplotlib.pyplot as plt
2 # Data to plot
3 lang = 'Java', 'Python', 'PHP', 'JavaScript', 'C#'
4 popularity = [22.2, 17.6, 8.8, 8, 7.7, 6.7]
5 color = ['#1f77b4', '#ff7f0e', '#2ca02c', '#d62728', '#9467bd', '#8c564b']
6 # explode 1st slice
7 explode = (0.1, 0, 0, 0, 0, 0)
8 # Plot
9 w = {'linewidth':1, 'edgecolor':'black'}
10 plt.pie(popularity, explode=explode, labels=languages, color=colors,
11         autopct='%1.1f%%', shadow=True, wedgeprops=w, startangle=140)
12 plt.axis('equal')
13 plt.show()
```

EXPLANATION:

- 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:

