

**Name- Pallavi Harishchandra Meher**

**700727681**

## **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**

Answer-

```
b= a.reshape(3,5) #reshaping vector into 3 by 5 array
```

```
print(b)
```

### **2. Print array shape.**

Answer

```
import numpy as np
```

```
a = np.random.randint(1,20,(15)) #random and randint function to create vector with random numbers
```

```
print(a)
```

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

Answer

```
c =np.where(b == [[i]for i in np.amax(b,axis=1)],0,b) #replaced max values of each row with 0
```

```
print(c)
```

## **2. Pandas**

### **1. Read the provided CSV file 'data.csv'.**

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

Answer-

```
import pandas as pd
```

```
data=pd.read_csv("data.csv") #reading csv file using pandas
```

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

Answer-

```
data.describe() #basic description of the data
```

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

**Answer-**

```
data.isnull() #checking for null values in data
```

a. Replace the null values with the mean

**Answer-**

```
data.fillna(data.mean(),inplace=True) #filling the null values with mean value
```

```
data.head(20)
```

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

**Answer-**

```
data.agg({'Duration': ['min', 'max','count','mean'], 'Pulse': ['min', 'max','count','mean']})
```

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

**Answer-**

```
data.loc[(data['Calories'] >=500) & (data['Calories'] <=1000)] #filtering dataframe with calories between 500 and 1000
```

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

**Answer-**

```
data.loc[(data['Calories'] >500) & (data['Pulse'] <100)] #filtering dataframe with calories >500 and pulse < 100
```

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

**Answer-**

```
df_modified = data.drop('Maxpulse',axis=1) #creating new dataframe without Maxpulse
```

```
df_modified
```

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

**Answer-**

```
del data["Maxpulse"] #deleted the Maxpulse  
  
data
```

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

**Answer-**

```
data = data.astype({'Calories':'int'}) #changed the datatype of calories  
  
print(data.dtypes)
```

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

**Answer-**

```
ax1 = data.plot.scatter(x='Duration', y='Calories') #created a scatter plot for the two columns (Duration  
and Calories).
```

```
ax1
```

### **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**

**Answer-**

```
import matplotlib.pyplot as plt  
  
languages = 'Java', 'Python', 'PHP', 'JavaScript', 'C#', 'C++'  
  
popurativity = [22.2, 17.6, 8.8, 8, 7.7, 6.7]  
  
colors = ["#1f77b4", "#ff7f0e", "#2ca02c", "#d62728", "#9467bd", "#8c564b"]  
  
explode = (0.1, 0, 0, 0, 0, 0)  
  
plt.pie(popurativity, explode=explode, labels=languages, colors=colors,  
autopct='%1.1f%%', shadow=True, startangle=140)
```

```
plt.axis('equal')
```

```
plt.show()
```