Github Link : https://github.com/LANKESAGAR/ML-Assignment-2

Question- 1

Description : In this Question I have created random vector of size 15 having only Integers in the range 1-20 using numpy and performed the actions like reshaping the array to 3 by 5, printing the array and replacing the max in each row by 0.

Source Code : This is the source code along with the output.

1. Numpy

In [109…
```
import numpy as np
```

In [110…
```
a=np.random.randint(1,20,15)
```

In [111…
```
a=a.reshape(3,5)
a
```

Out[111…
```
array([[13, 19,  9, 18,  6],
       [15, 14,  4,  5,  8],
       [ 3,  4,  6,  9,  9]])
```

In [112…
```
a.shape
```

Out[112…
```
(3, 5)
```

In [113…
```
a[np.isin(a,np.max(a,axis=1))]=0
print(a)
```

```
[[13  0  0 18  6]
 [ 0 14  4  5  8]
 [ 3  4  6  0  0]]
```

Question- 2

Description : In this Question I have used pandas and did the following operations on the given data.csv file.

Source Code : This is the source code along with the output.

2. Pandas

In [9]:
```
import pandas as pd
```

In [10]:
```
df=pd.read_csv("D:/Fall2022/ML Class/data.csv")
```

In [11]:
```
mean_value=df['Calories'].mean()
```

In [12]:
```
df['Calories'].fillna(value=mean_value,inplace=True)
```

In [13]:
```
df.head(25)
```

Out[13]:

	Duration	Pulse	Maxpulse	Calories
0	60	110	130	409.100000
1	60	117	145	479.000000
2	60	103	135	340.000000
3	45	109	175	282.400000
4	45	117	148	406.000000
5	60	102	127	300.000000
6	60	110	136	374.000000
7	45	104	134	253.300000
8	30	109	133	195.100000
9	60	98	124	269.000000
10	60	103	147	329.300000
11	60	100	120	250.700000
12	60	106	128	345.300000
13	60	104	132	379.300000
14	60	98	123	275.000000
15	60	98	120	215.200000
16	60	100	120	300.000000
17	45	90	112	375.790244
18	60	103	123	323.000000
19	45	97	125	243.000000
20	60	108	131	364.200000
21	45	100	119	282.000000
22	60	130	101	300.000000
23	45	105	132	246.000000
24	60	102	126	334.500000

In [32]:
```
df.Pulse.describe()
```

Out[32]:
```
count    169.000000
mean     107.461538
std       14.510259
min       80.000000
25%      100.000000
50%      105.000000
75%      111.000000
max      159.000000
Name: Pulse, dtype: float64
```

In [33]:
```
df.Pulse.describe()
```

Out[33]:
```
count    169.000000
mean     107.461538
std       14.510259
min       80.000000
25%      100.000000
50%      105.000000
75%      111.000000
max      159.000000
Name: Pulse, dtype: float64
```

In [16]:
```
df[(df['Calories']>500) & (df['Calories']<1000)]
```

Out[16]:

	Duration	Pulse	Maxpulse	Calories
51	80	123	146	643.1
62	160	109	135	853.0
65	180	90	130	800.4
66	150	105	135	873.4
67	150	107	130	816.0
72	90	100	127	700.0
73	150	97	127	953.2
75	90	98	125	563.2
78	120	100	130	500.4
90	180	101	127	600.1
99	90	93	124	604.1
103	90	90	100	500.4
106	180	90	120	800.3
108	90	90	120	500.3

In [17]:
```
df[(df['Calories']>500 & (df['Pulse']<100))]
```

Out[17]:

	Duration	Pulse	Maxpulse	Calories
0	60	110	130	409.1
1	60	117	145	479.0
2	60	103	135	340.0
3	45	109	175	282.4
4	45	117	148	406.0
...	...	...	...	...
164	60	105	140	290.8
165	60	110	145	300.0
166	60	115	145	310.2
167	75	120	150	320.4
168	75	125	150	330.4

169 rows × 4 columns

In [18]:
```
df_modified=df.drop("Maxpulse",axis=1)
```

In [19]:
```
df_modified
```

Out[19]:

	Duration	Pulse	Calories
0	60	110	409.1
1	60	117	479.0
2	60	103	340.0
3	45	109	282.4
4	45	117	406.0
...	...	...	...
164	60	105	290.8
165	60	110	300.0
166	60	115	310.2
167	75	120	320.4
168	75	125	330.4

169 rows × 3 columns

In [20]:
```
df=df.drop("Maxpulse",axis=1)
```

In [21]:
```
df
```

Out[21]:

	Duration	Pulse	Calories
0	60	110	409.1
1	60	117	479.0
2	60	103	340.0
3	45	109	282.4
4	45	117	406.0
...	...	...	...
164	60	105	290.8
165	60	110	300.0
166	60	115	310.2
167	75	120	320.4
168	75	125	330.4

169 rows × 3 columns

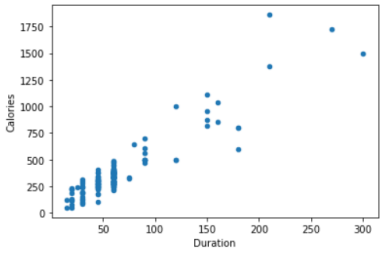
In [22]:
```
df['Calories']=df['Calories'].astype(int)
```

In [35]:
```
df['Calories'].dtypes
```

Out[35]:
```
dtype('int32')
```

In [36]:
```
df.plot.scatter( x = 'Duration', y = 'Calories')
```

Out[36]:
```
<AxesSubplot:xlabel='Duration', ylabel='Calories'>
```



Question- 3

Description : In this Question I have used Matplotlib and did the following.

Source Code : This is the source code along with the output.

3. Matplotlib

In [25]:
```
prgmng_df=pd.DataFrame({"popularity":[22.2, 17.6, 8.8, 8, 7.7, 6.7]},index=['Java','Python','PHP','JavaScript', 'C#', 'C++'])
prgmng_df
```

Out[25]:

	popularity
Java	22.2
Python	17.6
PHP	8.8
JavaScript	8.0
C#	7.7
C++	6.7

In [31]:
```
prgmng_df.plot.pie(y='popularity',autopct='%1.1f%%')
```

Out[31]:
```
<AxesSubplot:ylabel='popularity'>
```