MACHINE LEARNING ASSIGNMENT 4

STUDENT ID: 700766162

NAME: SRAVANI MANNURU

Github link: https://github.com/sravs2031/Machine-Learning-Assignment-4.git

Video link: https://drive.google.com/file/d/1y557Hgo0EB3yCWkdykb1N-teAx_ulyuQ/view?usp=drive_link

Screenshots:

1)

```
In [1]: import pandas as pd
        #Reading the data present in the csv file and storing it in a variable named data
        df=pd.read_csv('data.csv')
        df
```

Out[1]:

	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
166	60	115	145	310.2
167	75	120	150	320.4
168	75	125	150	330.4

2)

```
In [2]: #describing the basic statistical description about the data
        df.describe()
```

Out[2]:

	Duration	Pulse	Maxpulse	Calories
count	169.000000	169.000000	169.000000	164.000000
mean	63.846154	107.461538	134.047337	375.790244
std	42.299949	14.510259	16.450434	266.379919
min	15.000000	80.000000	100.000000	50.300000
25%	45.000000	100.000000	124.000000	250.925000
50%	60.000000	105.000000	131.000000	318.600000
75%	60.000000	111.000000	141.000000	387.600000
max	300.000000	159.000000	184.000000	1860.400000

3)

```
In [3]: #checking null values in data
        print(df.isnull().any())
        #replacing null values using mean
        df.fillna(df.mean(),inplace=True)
        #checking null values after replacing
        print("\nData after replacing null with mean value:\n{}".format(df.isnull().any()))
```

```
Duration    False
Pulse       False
Maxpulse    False
Calories     True
dtype: bool

Data after replacing null with mean value:
Duration    False
Pulse       False
Maxpulse    False
Calories    False
dtype: bool
```

4)

```
In [4]: #Selecting the two columns duration and calories and aggregating them with the computation values of m
        df.agg({'Duration':['min','max','count','mean'],'Calories':['min','max','count','mean']})
```

Out[4]:

	Duration	Calories
min	15.000000	50.300000
max	300.000000	1860.400000
count	169.000000	169.000000
mean	63.846154	375.790244

5)

```
In [5]: df[(df['Calories']>=500) & (df['Calories']<=1000)]
```

Out[5]:

	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
83	120	100	130	500.0
90	180	101	127	600.1
99	90	93	124	604.1
101	90	90	110	500.0
102	90	90	100	500.0
103	90	90	100	500.4
106	180	90	120	800.3
108	90	90	120	500.3

6)

```
In [6]: df[(df['Calories']>500) & (df['Pulse']<100)]
```

Out[6]:

	Duration	Pulse	Maxpulse	Calories
65	180	90	130	800.4
70	150	97	129	1115.0
73	150	97	127	953.2
75	90	98	125	563.2
99	90	93	124	604.1
103	90	90	100	500.4
106	180	90	120	800.3
108	90	90	120	500.3

7)

```
In [7]: df_modified=df.drop("Maxpulse",axis=1)
        df_modified
```

Out[7]:

	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

8)

```
In [8]: df=df.drop("Maxpulse",axis=1)
        df
```

Out[8]:

	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

168	75	125	330.4

169 rows × 3 columns

9)

```
In [9]: print("Datatype of the Calories column before changing it to int:",df['Calories'].dtypes)
        df['Calories']=df['Calories'].astype(int)
        print("Datatype of the Calories column after changing it to int:",df['Calories'].dtypes)
```

```
Datatype of the Calories column before changing it to int: float64
Datatype of the Calories column after changing it to int: int32
```