

- 1 Loaded the given data.csv file.
- 2 Used describe() function to display the data in statistical form

```
import pandas as pd

dataSet = pd.read_csv('data.csv')

# Showing basic statistical description of the data
print(dataSet.describe())
```

Output:

	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 isnull() function used to display the null values and any() function used to display in Boolean format

```
# Checking if the data has null values
dataSet.isnull().any()
```

Output:

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

3.a fillna() function is used to fill the null values with the mean values

```
# Replacing the null values with the mean
dataSet.fillna(dataSet.mean(), inplace=True)
dataSet.isnull().any()
```

Output:

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

4 Aggregated the data using agg() function with min, max, count and mean

```
# Aggregating the data using min, max, count, and mean
agg_dataSet = dataSet[['Duration', 'Calories']].agg(['min', 'max', 'count', 'mean'])
print(agg_dataSet)
```

Output:

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

5 Filtered the given dataset as shown below with the conditions

```
# Filtering the dataframe to select rows with calories between 500 and 1000
filtered_dataSet = dataSet[(dataSet['Calories'] >= 500) & (dataSet['Calories'] <= 1000)]
print(filtered_dataSet)
```

Output:

	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 Filtered the given dataset as shown below with the conditions.

```
# Filtering the dataframe to select rows with calories > 500 and pulse < 100
filtered_dataSet2 = dataSet[(dataSet['Calories'] > 500) & (dataSet['Pulse'] < 100)]
print(filtered_dataSet2)
```

Output:

	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 Creating a new dataframe without Maxpulse column by dropping the Maxpulse column using drop() function()

```
# Creating a new df_modified dataframe without Maxpulse column
dataSet_modified = dataSet.drop(columns=['Maxpulse'])

# Printing the created new dataframe
print(dataSet_modified)
```

Output:

	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 x 3 columns]

8 Deleting the Maxpulse column from the original dataframe

```
# Deleting Maxpulse column from the main dataframe
dataSet.drop(columns=['Maxpulse'], inplace=True)
print(dataSet)
```

Output:

	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 x 3 columns]

9 Converting the datatype of Calories using astype()

```
# Converting datatype of Calories column to int
dataSet['Calories'] = dataSet['Calories'].astype(int)
dataSet.dtypes
```

Output:

```
Duration    int64
Pulse       int64
Calories    int64
dtype: object
```

GitHub Link:

<https://github.com/shruthikatkam26/MachinelearningAssign4.git>

Video Link:

https://drive.google.com/file/d/1GftEN5A-xc4cRreYyXuHN9SlwgLvnc_3/view?usp=sharing