In [37]:
```
import pandas
mydataset = {
'Person': ["Bibek", "Umang", "Piyush"],
'Rollno': [20051722,20051836, 20051802]
}
myvar = pandas.DataFrame(mydataset)
print(myvar)
print()
```

```
   Person    Rollno
0   Bibek  20051722
1   Umang  20051836
2  Piyush  20051802
```

In [36]:
```
import pandas
data = {
'Person': ["Bibek", "Umang", "Piyush"],
'Roll No.': [20051722, 20051836, 20051802]
}
df = pd.DataFrame(data, index = ["1", "2", "3"])
print(df)
```

```
   Person  Roll No.
1   Bibek  20051722
2   Umang  20051836
3  Piyush  20051802
```

In [3]:
```
import pandas as pd
df = pd.read_csv(r"B:\6th sems\T & T lab\Lab-7\data.csv")
print(df)
print(pd.options.display.max_rows)
```

```
     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 x 4 columns]
60
```

In [61]:
```
print(df.to_string())
```

```
5          60    102       127     300.0
6          60    110       136     374.0
7          45    104       134     253.3
8          30    109       133     195.1
9          60     98       124     269.0
10         60    103       147     329.3
11         60    100       120     250.7
12         60    106       128     345.3
13         60    104       132     379.3
14         60     98       123     275.0
15         60     98       120     215.2
16         60    100       120     300.0
17         45     90       112       NaN
18         60    103       123     323.0
19         45     97       125     243.0
20         60    108       131     364.2
21         45    100       119     282.0
22         60    130       101     300.0
23         45    105       132     246.0
```

In [38]:
```
import pandas as pd
df = pd.read_csv(r"B:\6th sems\T & T lab\Lab-7\data.csv")
print(df.tail(10))
```

```
     Duration  Pulse  Maxpulse  Calories
159        30     80       120     240.9
160        30     85       120     250.4
161        45     90       130     260.4
162        45     95       130     270.0
163        45    100       140     280.9
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
```

In [4]:
```
print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 169 entries, 0 to 168
Data columns (total 4 columns):
 #   Column    Non-Null Count  Dtype  
---  ------    --------------  -----  
 0   Duration  169 non-null    int64  
 1   Pulse     169 non-null    int64  
 2   Maxpulse  169 non-null    int64  
 3   Calories  164 non-null    float64
dtypes: float64(1), int64(3)
memory usage: 5.4 KB
None
```

In [5]:
```
print(df.head(10))
```

```
   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
5        60    102       127     300.0
6        60    110       136     374.0
7        45    104       134     253.3
8        30    109       133     195.1
9        60     98       124     269.0
```

In [6]:
```
new_df = df.dropna()
print(new_df.to_string())
print()
df.fillna(300,inplace = True)
print(df.to_string())
df["Calories"].fillna(300, inplace = True)
print(df.to_string())
```

```
     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
5          60    102       127     300.0
6          60    110       136     374.0
7          45    104       134     253.3
8          30    109       133     195.1
9          60     98       124     269.0
10         60    103       147     329.3
11         60    100       120     250.7
12         60    106       128     345.3
13         60    104       132     379.3
14         60     98       123     275.0
15         60     98       120     215.2
16         60    100       120     300.0
18         60    103       123     323.0
```

In [8]:
```
x = df["Calories"].mean()
df["Calories"].fillna(x, inplace = True)
print(df.to_string())
```

```
     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
5          60    102       127     300.0
6          60    110       136     374.0
7          45    104       134     253.3
8          30    109       133     195.1
9          60     98       124     269.0
10         60    103       147     329.3
11         60    100       120     250.7
12         60    106       128     345.3
13         60    104       132     379.3
14         60     98       123     275.0
15         60     98       120     215.2
16         60    100       120     300.0
17         45     90       112     300.0
```

In [7]:
```
x = df["Calories"].median()
df["Calories"].fillna(x, inplace = True)
print(df.to_string())
```

```
85          30    151       170     300.0
86          45    102       136     234.0
87         120    100       157    1000.1
88          45    129       103     242.0
89          20     83       107      50.3
90         180    101       127     600.1
91          45    107       137     300.0
92          30     90       107     105.3
93          15     80       100      50.5
94          20    150       171     127.4
95          20    151       168     229.4
96          30     95       128     128.2
97          25    152       168     244.2
98          30    109       131     188.2
99          90     93       124     604.1
100         20     95       112      77.7
101         90     90       110     500.0
102         90     90       100     500.0
103         90     90       100     500.4
104         30     92       108      92.7
```

In [9]:
```
x = df["Calories"].mode()[0]
df["Calories"].fillna(x, inplace = True)
print(df.to_string())
```

```
     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
5          60    102       127     300.0
6          60    110       136     374.0
7          45    104       134     253.3
8          30    109       133     195.1
9          60     98       124     269.0
10         60    103       147     329.3
11         60    100       120     250.7
12         60    106       128     345.3
13         60    104       132     379.3
14         60     98       123     275.0
15         60     98       120     215.2
16         60    100       120     300.0
17         45     90       112     300.0
```

In [46]:
```
import pandas as pd
df = pd.read_csv(r"B:\6th sems\T & T lab\Lab-7\data1.csv")
x = df["Calories"].median()
df["Calories"].fillna(x, inplace = True)
print(df.to_string())
```

```
    Duration  Pulse  Maxpulse  Calories        Date
0         60    110       130     409.1  01-12-2020
1         60    117       145     479.0  02-12-2020
2         60    103       135     340.0  03-12-2020
3         45    109       175     282.4  04-12-2020
4         45    117       148     406.0  05-12-2020
5         60    102       127     300.0  06-12-2020
6         60    110       136     374.0  07-12-2020
7         45    104       134     253.3  08-12-2020
8         30    109       133     195.1  09-12-2020
9         60     98       124     269.0  10-12-2020
10        60    103       147     329.3  11-12-2020
11        60    100       120     250.7  12-12-2020
12        60    106       128     345.3  13-12-2020
13        60    104       132     379.3  14-12-2020
14        60     98       123     275.0  15-12-2020
15        60     98       120     215.2  16-12-2020
16        60    100       120     300.0  17-12-2020
17        45     90       112     300.0  18-12-2020
18        60    103       123     323.0  19-12-2020
19        45     97       125     243.0  20-12-2020
20        60    108       131     364.2  21-12-2020
21        45    100       119     282.0  22-12-2020
22        60    130       101     300.0  23-12-2020
23        45    105       132     246.0  24-12-2020
24        60    102       126     334.5  25-12-2020
25        60    100       120     250.0  26-12-2020
26        60     92       118     241.0  27-12-2020
27        60    103       132     300.0  28-12-2020
28        60    100       132     280.0  29-12-2020
29        60    102       129     380.3  30-12-2020
```

In [54]:
```
df['Date'] = pd.to_datetime(df['Date'])
print(df.to_string())
```

```
    Duration  Pulse  Maxpulse  Calories       Date
0         60    110       130     409.1 2020-01-12
1         60    117       145     479.0 2020-02-12
2         60    103       135     340.0 2020-03-12
3         45    109       175     282.4 2020-04-12
4         45    117       148     406.0 2020-05-12
5         60    102       127     300.0 2020-06-12
6         60    110       136     374.0 2020-07-12
7         45    104       134     253.3 2020-08-12
8         30    109       133     195.1 2020-09-12
9         60     98       124     269.0 2020-10-12
10        60    103       147     329.3 2020-11-12
11        60    100       120     250.7 2020-12-12
12        60    106       128     345.3 2020-12-13
13        60    104       132     379.3 2020-12-14
14        60     98       123     275.0 2020-12-15
15        60     98       120     215.2 2020-12-16
16        60    100       120     300.0 2020-12-17
17        45     90       112       NaN 2020-12-18
18        60    103       123     323.0 2020-12-19
19        45     97       125     243.0 2020-12-20
20        60    108       131     364.2 2020-12-21
21        45    100       119     282.0 2020-12-22
22        60    130       101     300.0 2020-12-23
23        45    105       132     246.0 2020-12-24
24        60    102       126     334.5 2020-12-25
25        60    100       120     250.0 2020-12-26
26        60     92       118     241.0 2020-12-27
27        60    103       132       NaN 2020-12-28
28        60    100       132     280.0 2020-12-29
29        60    102       129     380.3 2020-12-30
```

In [11]:
```
import pandas as pd
df = pd.read_csv(r"B:\6th sems\T & T lab\Lab-7\data1.csv")
print()
df.dropna(subset=['Date'], inplace = True)
print(df.to_string())
```

```
    Duration  Pulse  Maxpulse  Calories        Date
0         60    110       130     409.1  01-12-2020
1         60    117       145     479.0  02-12-2020
2         60    103       135     340.0  03-12-2020
3         45    109       175     282.4  04-12-2020
4         45    117       148     406.0  05-12-2020
5         60    102       127     300.0  06-12-2020
6         60    110       136     374.0  07-12-2020
7         45    104       134     253.3  08-12-2020
8         30    109       133     195.1  09-12-2020
9         60     98       124     269.0  10-12-2020
10        60    103       147     329.3  11-12-2020
11        60    100       120     250.7  12-12-2020
12        60    106       128     345.3  13-12-2020
13        60    104       132     379.3  14-12-2020
14        60     98       123     275.0  15-12-2020
15        60     98       120     215.2  16-12-2020
16        60    100       120     300.0  17-12-2020
17        45     90       112       NaN  18-12-2020
18        60    103       123     323.0  19-12-2020
19        45     97       125     243.0  20-12-2020
20        60    108       131     364.2  21-12-2020
21        45    100       119     282.0  22-12-2020
22        60    130       101     300.0  23-12-2020
23        45    105       132     246.0  24-12-2020
24        60    102       126     334.5  25-12-2020
25        60    100       120     250.0  26-12-2020
26        60     92       118     241.0  27-12-2020
27        60    103       132       NaN  28-12-2020
28        60    100       132     280.0  29-12-2020
29        60    102       129     380.3  30-12-2020
```

In [12]:
```
new_df = df.dropna()
print(new_df.to_string())
```

```
    Duration  Pulse  Maxpulse  Calories        Date
0         60    110       130     409.1  01-12-2020
1         60    117       145     479.0  02-12-2020
2         60    103       135     340.0  03-12-2020
3         45    109       175     282.4  04-12-2020
4         45    117       148     406.0  05-12-2020
5         60    102       127     300.0  06-12-2020
6         60    110       136     374.0  07-12-2020
7         45    104       134     253.3  08-12-2020
8         30    109       133     195.1  09-12-2020
9         60     98       124     269.0  10-12-2020
10        60    103       147     329.3  11-12-2020
11        60    100       120     250.7  12-12-2020
12        60    106       128     345.3  13-12-2020
13        60    104       132     379.3  14-12-2020
14        60     98       123     275.0  15-12-2020
15        60     98       120     215.2  16-12-2020
16        60    100       120     300.0  17-12-2020
18        60    103       123     323.0  19-12-2020
19        45     97       125     243.0  20-12-2020
20        60    108       131     364.2  21-12-2020
21        45    100       119     282.0  22-12-2020
22        60    130       101     300.0  23-12-2020
23        45    105       132     246.0  24-12-2020
24        60    102       126     334.5  25-12-2020
25        60    100       120     250.0  26-12-2020
26        60     92       118     241.0  27-12-2020
28        60    100       132     280.0  29-12-2020
29        60    102       129     380.3  30-12-2020
```

In [13]:
```
df.dropna(inplace = True)
print(df.to_string())
```

```
    Duration  Pulse  Maxpulse  Calories        Date
0         60    110       130     409.1  01-12-2020
1         60    117       145     479.0  02-12-2020
2         60    103       135     340.0  03-12-2020
3         45    109       175     282.4  04-12-2020
4         45    117       148     406.0  05-12-2020
5         60    102       127     300.0  06-12-2020
6         60    110       136     374.0  07-12-2020
7         45    104       134     253.3  08-12-2020
8         30    109       133     195.1  09-12-2020
9         60     98       124     269.0  10-12-2020
10        60    103       147     329.3  11-12-2020
11        60    100       120     250.7  12-12-2020
12        60    106       128     345.3  13-12-2020
13        60    104       132     379.3  14-12-2020
14        60     98       123     275.0  15-12-2020
15        60     98       120     215.2  16-12-2020
16        60    100       120     300.0  17-12-2020
18        60    103       123     323.0  19-12-2020
19        45     97       125     243.0  20-12-2020
20        60    108       131     364.2  21-12-2020
21        45    100       119     282.0  22-12-2020
22        60    130       101     300.0  23-12-2020
23        45    105       132     246.0  24-12-2020
24        60    102       126     334.5  25-12-2020
25        60    100       120     250.0  26-12-2020
26        60     92       118     241.0  27-12-2020
28        60    100       132     280.0  29-12-2020
29        60    102       129     380.3  30-12-2020
```

In [14]:
```
df.fillna(230, inplace = True)
print(df.to_string())
```

```
    Duration  Pulse  Maxpulse  Calories        Date
0         60    110       130     409.1  01-12-2020
1         60    117       145     479.0  02-12-2020
2         60    103       135     340.0  03-12-2020
3         45    109       175     282.4  04-12-2020
4         45    117       148     406.0  05-12-2020
5         60    102       127     300.0  06-12-2020
6         60    110       136     374.0  07-12-2020
7         45    104       134     253.3  08-12-2020
8         30    109       133     195.1  09-12-2020
9         60     98       124     269.0  10-12-2020
10        60    103       147     329.3  11-12-2020
11        60    100       120     250.7  12-12-2020
12        60    106       128     345.3  13-12-2020
13        60    104       132     379.3  14-12-2020
14        60     98       123     275.0  15-12-2020
15        60     98       120     215.2  16-12-2020
16        60    100       120     300.0  17-12-2020
18        60    103       123     323.0  19-12-2020
19        45     97       125     243.0  20-12-2020
20        60    108       131     364.2  21-12-2020
21        45    100       119     282.0  22-12-2020
22        60    130       101     300.0  23-12-2020
23        45    105       132     246.0  24-12-2020
24        60    102       126     334.5  25-12-2020
25        60    100       120     250.0  26-12-2020
26        60     92       118     241.0  27-12-2020
28        60    100       132     280.0  29-12-2020
29        60    102       129     380.3  30-12-2020
```

In [15]:
```
df["Calories"].fillna(230, inplace = True)
print(df.to_string())
```

```
    Duration  Pulse  Maxpulse  Calories        Date
0         60    110       130     409.1  01-12-2020
1         60    117       145     479.0  02-12-2020
2         60    103       135     340.0  03-12-2020
3         45    109       175     282.4  04-12-2020
4         45    117       148     406.0  05-12-2020
5         60    102       127     300.0  06-12-2020
6         60    110       136     374.0  07-12-2020
7         45    104       134     253.3  08-12-2020
8         30    109       133     195.1  09-12-2020
9         60     98       124     269.0  10-12-2020
10        60    103       147     329.3  11-12-2020
11        60    100       120     250.7  12-12-2020
12        60    106       128     345.3  13-12-2020
13        60    104       132     379.3  14-12-2020
14        60     98       123     275.0  15-12-2020
15        60     98       120     215.2  16-12-2020
16        60    100       120     300.0  17-12-2020
18        60    103       123     323.0  19-12-2020
19        45     97       125     243.0  20-12-2020
20        60    108       131     364.2  21-12-2020
21        45    100       119     282.0  22-12-2020
22        60    130       101     300.0  23-12-2020
23        45    105       132     246.0  24-12-2020
24        60    102       126     334.5  25-12-2020
25        60    100       120     250.0  26-12-2020
26        60     92       118     241.0  27-12-2020
28        60    100       132     280.0  29-12-2020
29        60    102       129     380.3  30-12-2020
```

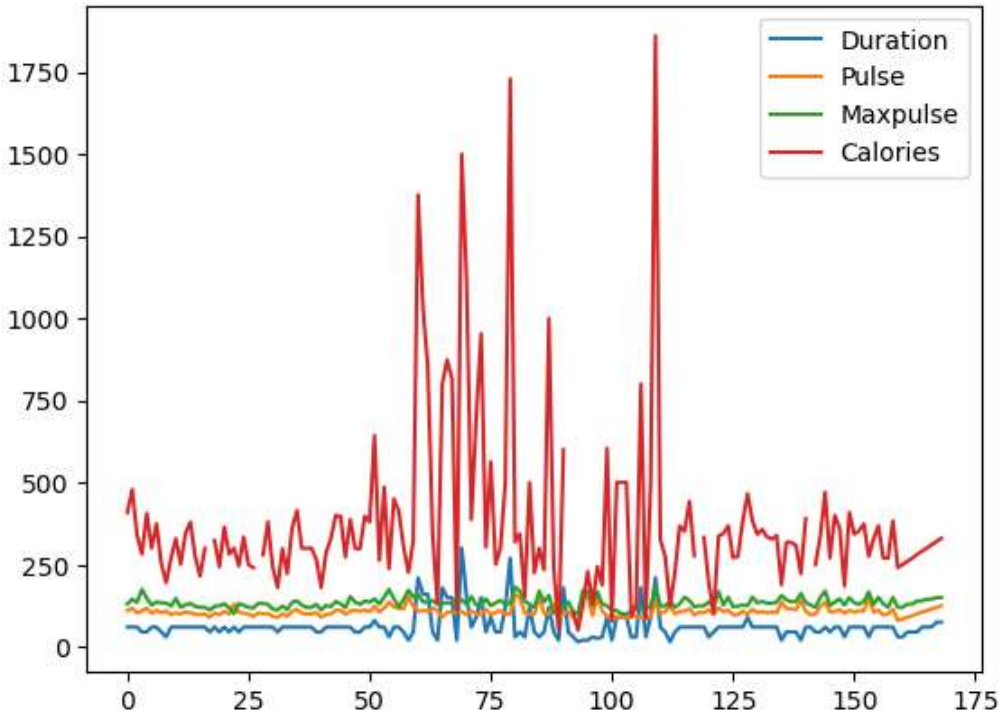
In [16]:
```
df.loc[7, 'Duration'] = 50
print(df)
```

```
    Duration  Pulse  Maxpulse  Calories        Date
0         60    110       130     409.1  01-12-2020
1         60    117       145     479.0  02-12-2020
2         60    103       135     340.0  03-12-2020
3         45    109       175     282.4  04-12-2020
4         45    117       148     406.0  05-12-2020
5         60    102       127     300.0  06-12-2020
6         60    110       136     374.0  07-12-2020
7         50    104       134     253.3  08-12-2020
8         30    109       133     195.1  09-12-2020
9         60     98       124     269.0  10-12-2020
10        60    103       147     329.3  11-12-2020
11        60    100       120     250.7  12-12-2020
12        60    106       128     345.3  13-12-2020
13        60    104       132     379.3  14-12-2020
14        60     98       123     275.0  15-12-2020
15        60     98       120     215.2  16-12-2020
16        60    100       120     300.0  17-12-2020
18        60    103       123     323.0  19-12-2020
19        45     97       125     243.0  20-12-2020
20        60    108       131     364.2  21-12-2020
21        45    100       119     282.0  22-12-2020
22        60    130       101     300.0  23-12-2020
23        45    105       132     246.0  24-12-2020
24        60    102       126     334.5  25-12-2020
25        60    100       120     250.0  26-12-2020
26        60     92       118     241.0  27-12-2020
28        60    100       132     280.0  29-12-2020
29        60    102       129     380.3  30-12-2020
```

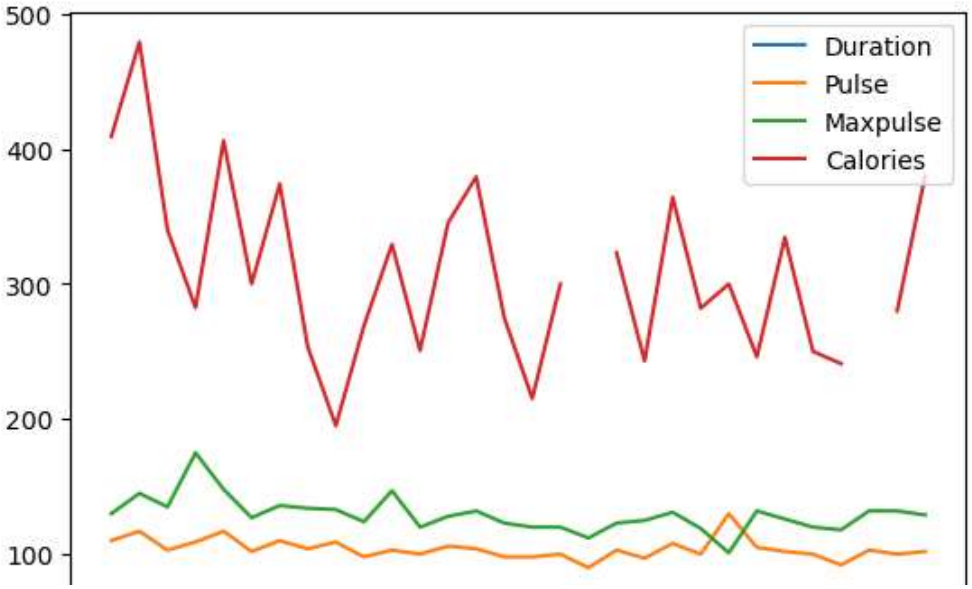
In [47]:
```
import pandas as pd
df = pd.read_csv(r"B:\6th sems\T & T lab\Lab-7\data.csv")
for x in df.index:
    if df.loc[x, "Duration"] > 120:
        df.loc[x, "Duration"] = 120
print(df.to_string())
```

```
    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
5         60    102       127     300.0
6         60    110       136     374.0
7         45    104       134     253.3
8         30    109       133     195.1
9         60     98       124     269.0
10        60    103       147     329.3
11        60    100       120     250.7
12        60    106       128     345.3
13        60    104       132     379.3
14        60     98       123     275.0
15        60     98       120     215.2
16        60    100       120     300.0
17        45     90       112       NaN
```

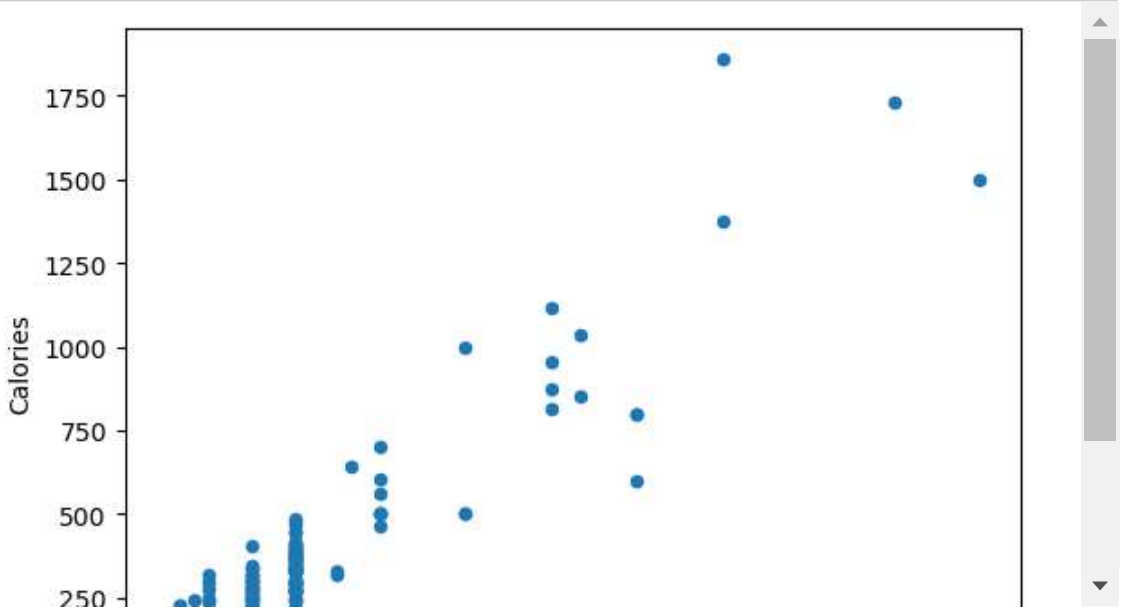
In [32]:
```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv(r'B:\6th sems\T & T lab\Lab-7\data.csv')
df.plot()
plt.show()
```



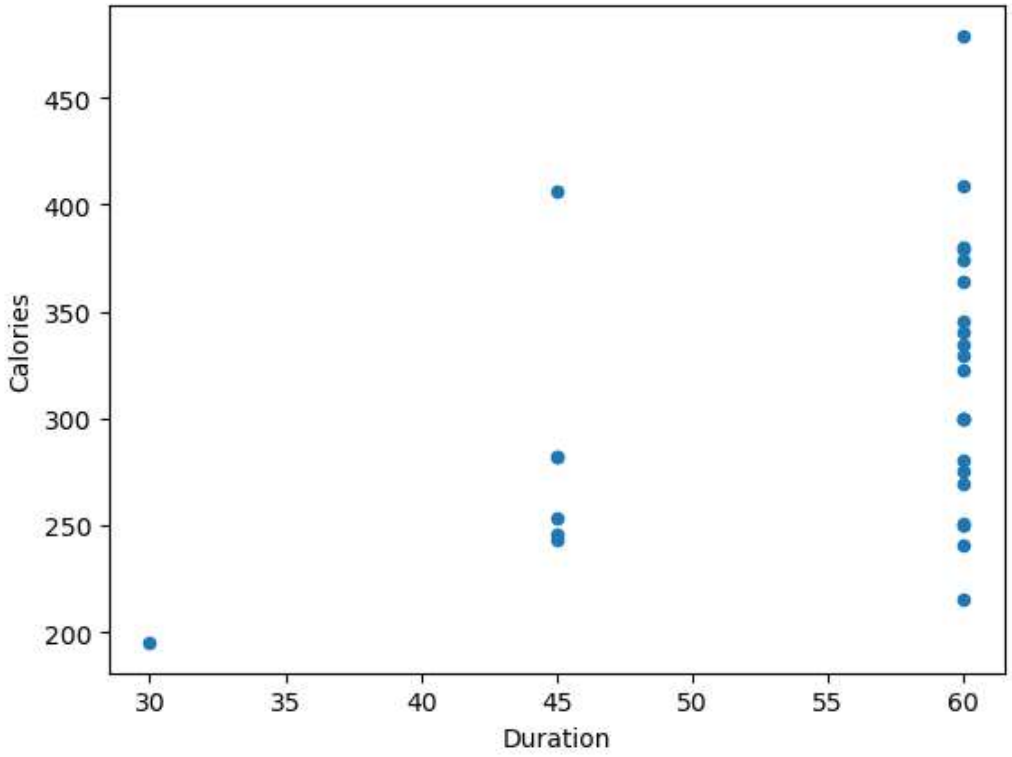
In [19]:
```
df = pd.read_csv(r'B:\6th sems\T & T lab\Lab-7\data1.csv')
df.plot()
plt.show()
```



In [20]:
```
df = pd.read_csv(r'B:\6th sems\T & T lab\Lab-7\data.csv')
df.plot(kind = 'scatter', x = 'Duration', y = 'Calories')
plt.show()
```

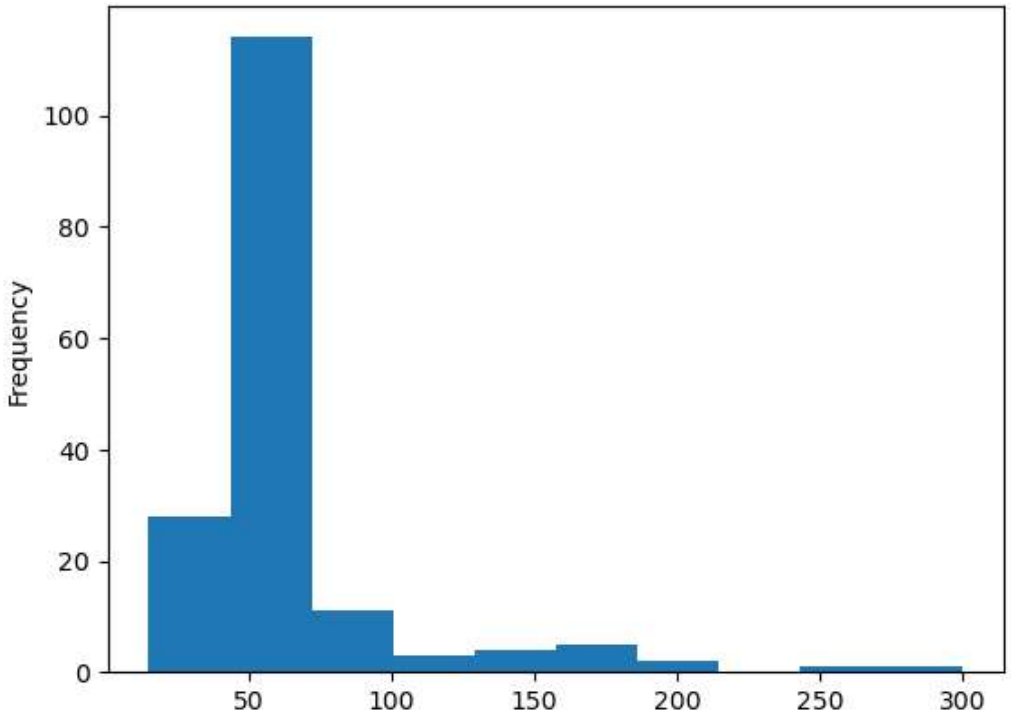


In [21]:
```
df = pd.read_csv(r'B:\6th sems\T & T lab\Lab-7\data1.csv')
df.plot(kind = 'scatter', x = 'Duration', y = 'Calories')
plt.show()
```



In [22]:
```
df = pd.read_csv(r'B:\6th sems\T & T lab\Lab-7\data.csv')
df["Duration"].plot(kind = 'hist')
```

Out[22]: <AxesSubplot:ylabel='Frequency'>



In [23]:
```
df = pd.read_csv(r'B:\6th sems\T & T lab\Lab-7\data1.csv')
df["Duration"].plot(kind = 'hist')
```

Out[23]: <AxesSubplot:ylabel='Frequency'>