Data visualization with Matplolib & pandas

First of all download latest version of python :-
from website
https://www.python.org/downloads/
& install python.

Step 1:- create folder okpython in your d: drive then open command prompt and type following commands as shown below:-

```
C:\Users\software>d:

D:\>cd okpython
```

step 2:- Setup Virtual Environment

D;\okpython> python -m venv env

Step 3:- **to activate virutal enviornment**

```
D:\okpython>env\Scripts\activate
```

Step 4:- To install matplotlib type command **pip install matplotlib** as shown below :-

```
(env) D:\okpython>pip install matplotlib
Collecting matplotlib
  Using cached matplotlib-3.6.2-cp311-cp311-win_amd64.whl (7.2 MB)
Collecting contourpy>=1.0.1
  Using cached contourpy-1.0.6-cp311-cp311-win_amd64.whl (163 kB)
Collecting cycler>=0.10
  Using cached cycler-0.11.0-py3-none-any.whl (6.4 kB)
Collecting fonttools>=4.22.0
  Using cached fonttools-4.38.0-py3-none-any.whl (965 kB)
Collecting kiwisolver>=1.0.1
  Using cached kiwisolver-1.4.4-cp311-cp311-win_amd64.whl (55 kB)
Collecting numpy>=1.19
  Using cached numpy-1.23.5-cp311-cp311-win_amd64.whl (14.6 MB)
Collecting packaging>=20.0
  Downloading packaging-22.0-py3-none-any.whl (42 kB)
     ---------------------------------------- 42.6/42.6 kB 2.2 MB/
```

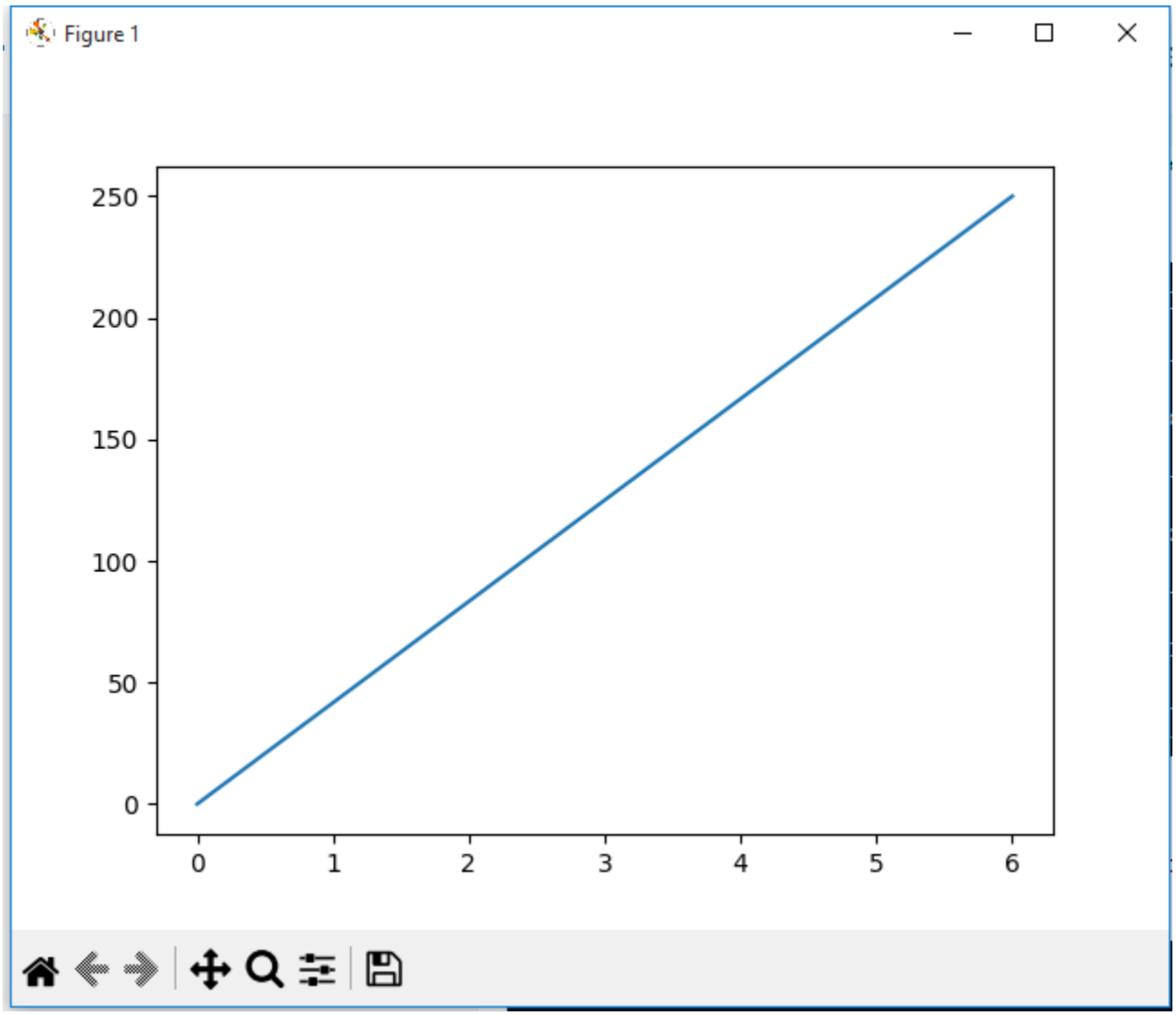
step 5:- open your editor (visual studio code) and type following code as shown below and save file name as m1.py :-

```
import matplotlib.pyplot as plt
import numpy as np

xpoints = np.array([0, 6])
ypoints = np.array([0, 250])

plt.plot(xpoints, ypoints)
plt.show()
```

step 6:- in your command prompt type following command to run this file as shown below :-

```
(env) D:\okpython>py m1.py
```

Output you will see :-



after this to read csv file we need to install pandas as shown below

using command **pip install pandas**

```
(env) D:\okpython>pip install pandas
Collecting pandas
  Using cached pandas-1.5.2-cp311-cp311-win_amd64.whl
Requirement already satisfied: python-dateutil>=2.8.1
Collecting pytz>=2020.1
  Using cached pytz-2022.6-py2.py3-none-any.whl (498 k
Requirement already satisfied: numpy>=1.21.0 in d:\okp
Requirement already satisfied: six>=1.5 in d:\okpython
.0)
Installing collected packages: pytz, pandas
Successfully installed pandas-1.5.2 pytz-2022.6
```

after it download csv file form link :-

https://geocities.ws/ommauryasir/python/data.csv

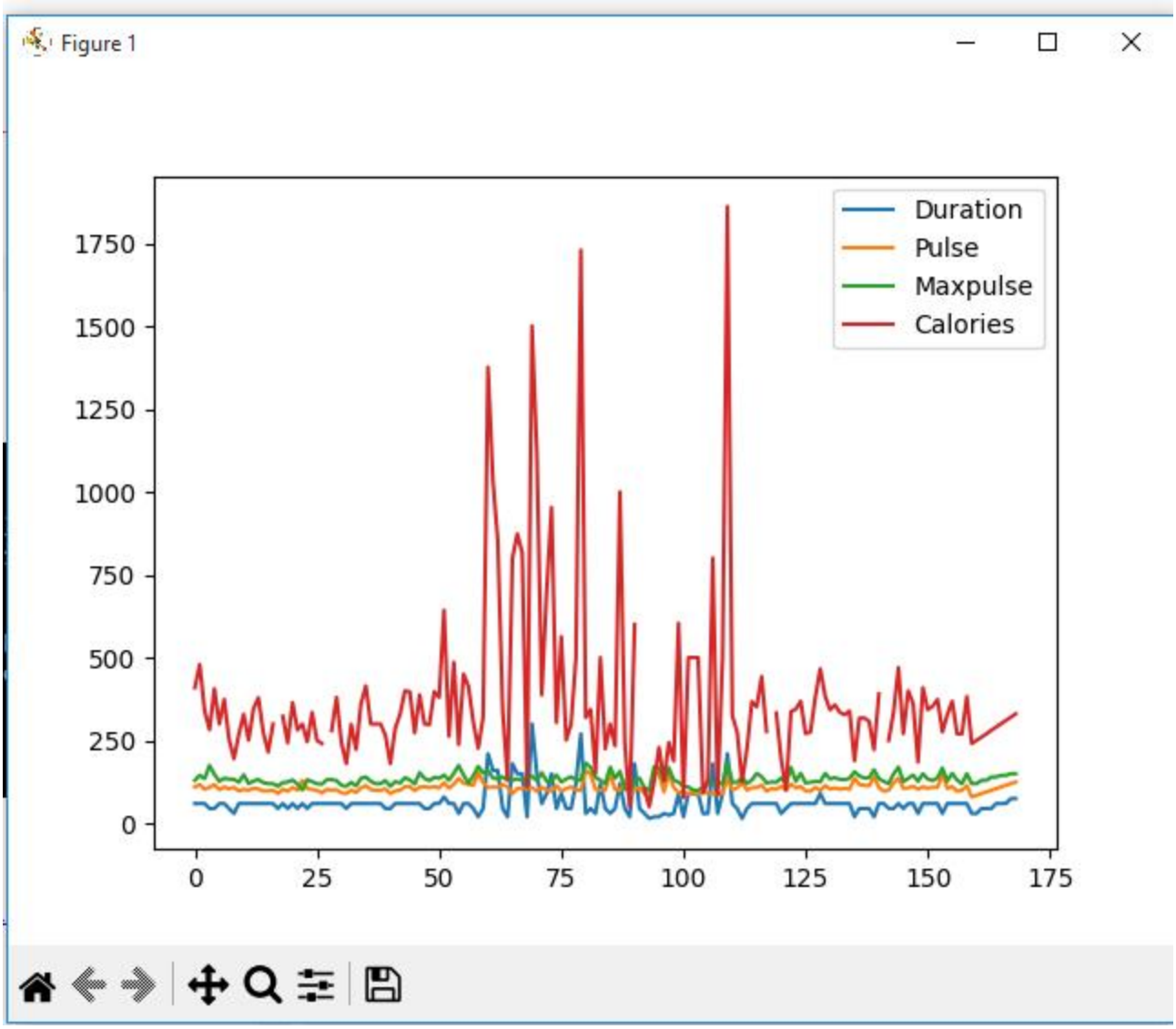
and keep this data.csv file under your okpython folder

and write following code in filename p1.py inside your **okpython** folder :-

p1.py file code :-

```
import pandas as pd
import matplotlib.pyplot as plt

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

df.plot()

plt.show()
```

output:-



Scatter Plot :-

Specify that you want a scatter plot with the `kind` argument:

```
kind = 'scatter'
```

A scatter plot needs an x- and a y-axis.

In the example below we will use "Duration" for the x-axis and "Calories" for the y-axis.

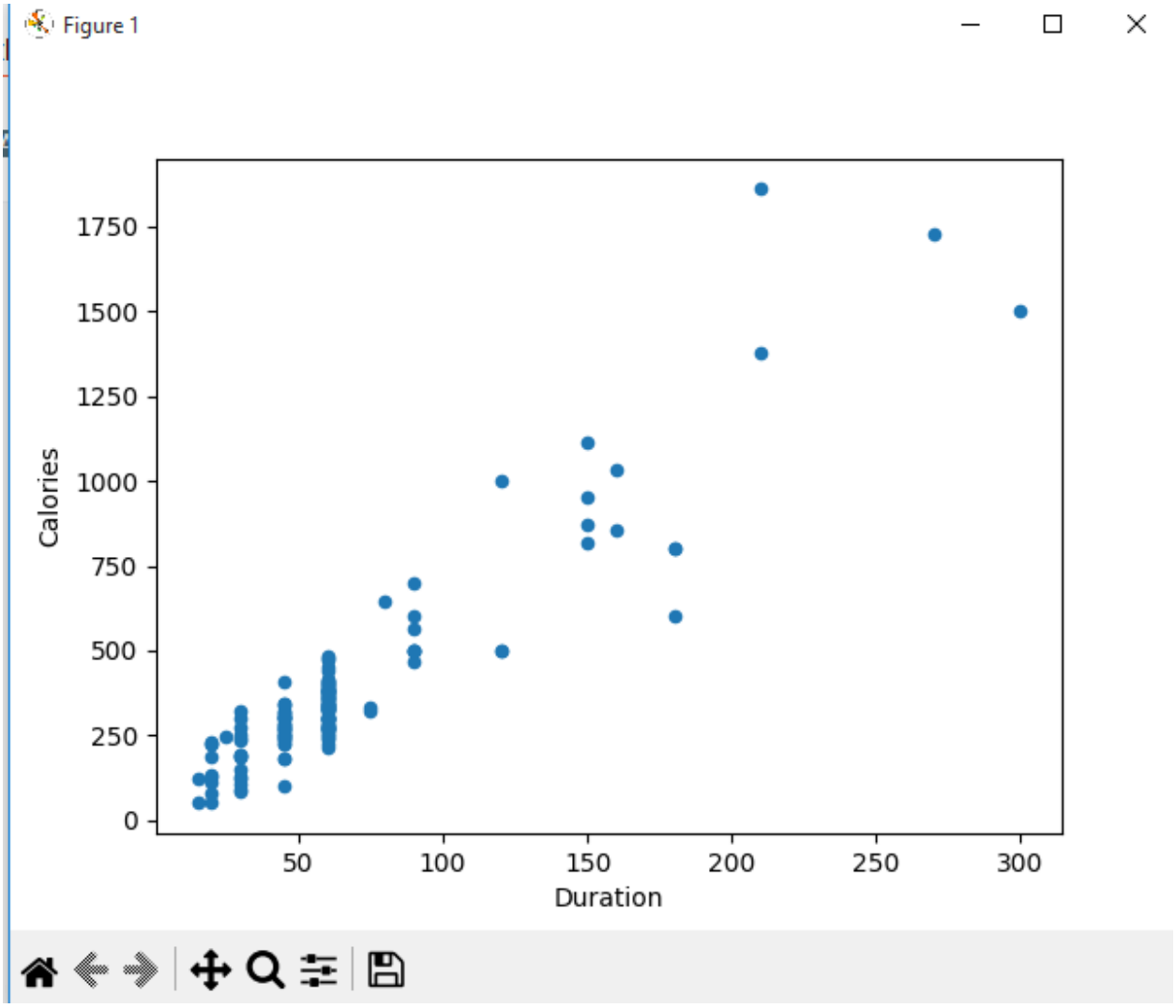
Include the x and y arguments like this:

```
x = 'Duration', y = 'Calories'
```

write following example code py2.py file :-

```
import pandas as pd
import matplotlib.pyplot as plt

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

df.plot(kind = 'scatter', x = 'Duration', y = 'Calories')

plt.show()
```

output:-



Histogram:-

Use the `kind` argument to specify that you want a histogram:

```
kind = 'hist'
```

A histogram needs only one column.

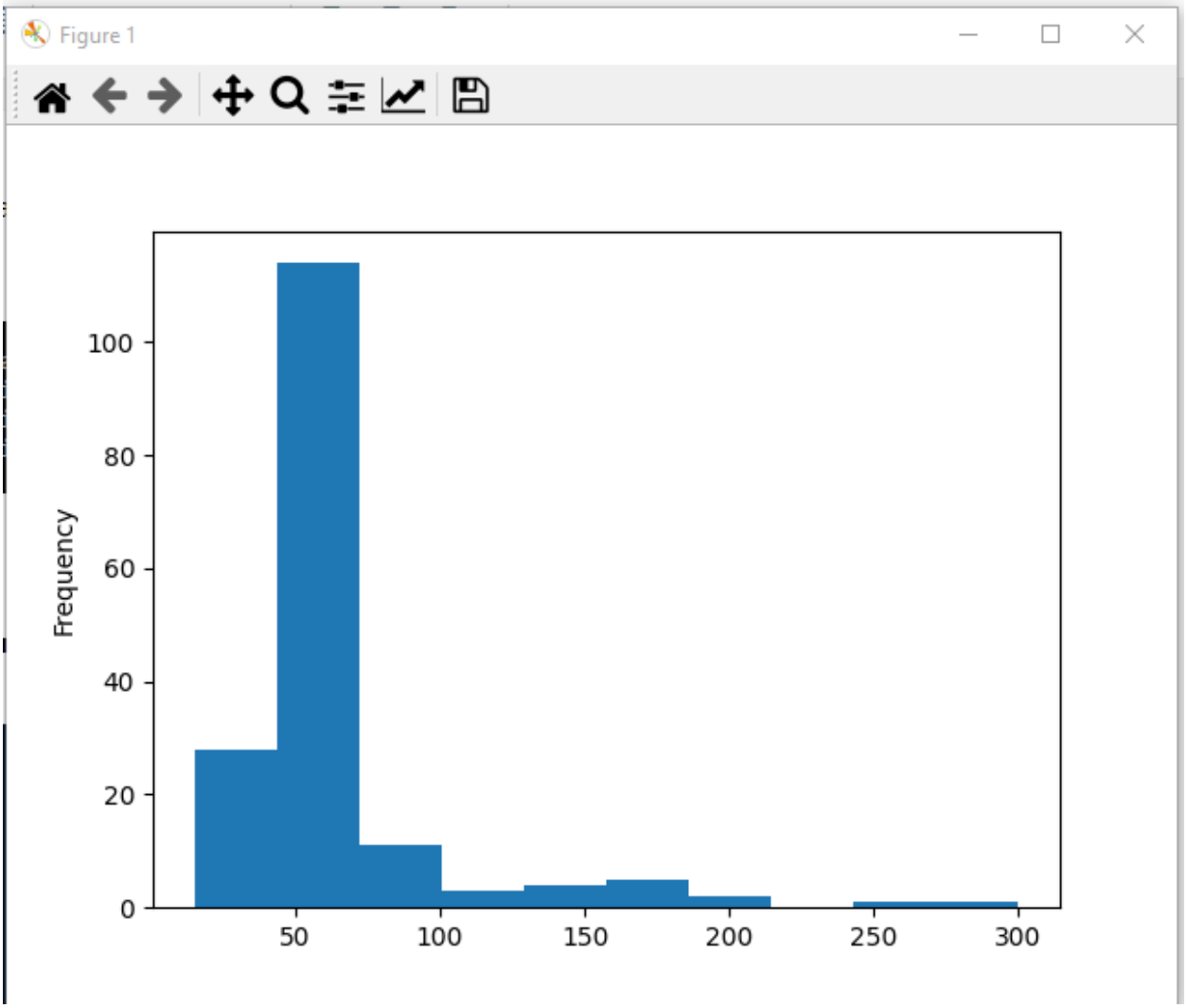
A histogram shows us the frequency of each interval, e.g. how many workouts lasted between 50 and 60 minutes?

In the example below we will use the "Duration" column to create the **histogram:-**

Now write code for p3.py file as shown below :-

```
import sys
import matplotlib
matplotlib.use('qtagg')

import pandas as pd
import matplotlib.pyplot as plt

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

df["Duration"].plot(kind = 'hist')

plt.show()
```

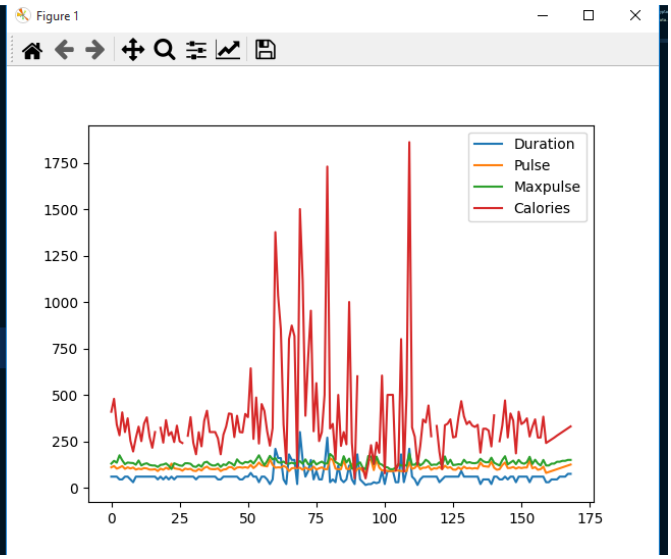
Output:-



Example to run in your visual studio code as shown below are :-

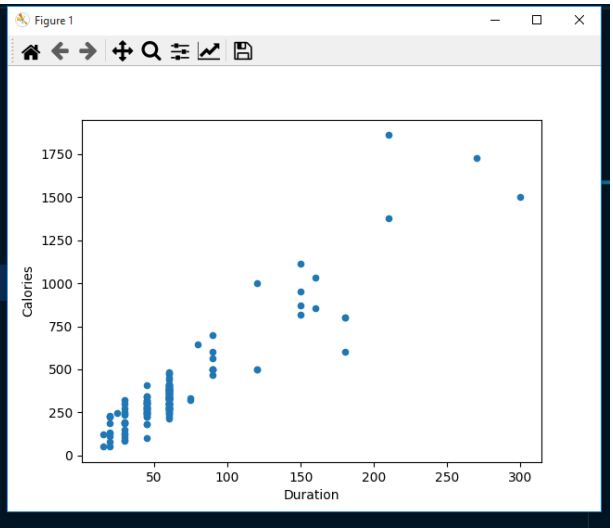
1)

```
import pandas as pd
import matplotlib.pyplot as plt

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

df.plot()

plt.show()
```



2)

```
import pandas as pd
import matplotlib.pyplot as plt

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

df.plot(kind = 'scatter', x = 'Duration',

plt.show()
```



3)

p3.py > ...

```python
import sys
import matplotlib
matplotlib.use('qtagg')

import pandas as pd
import matplotlib.pyplot as plt

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

df["Duration"].plot(kind = 'hist')

plt.show()
```