190905104
Parth Shukla
Lab 2

Exercise questions

1)
Area of rectangle

```
l = int(input('Enter length l: '))
w = int(input('Enter width w: '))

print("Area is " + str(l*w*0.5))
```

```
student@dslab-12:~/190905104_DS/lab2$ python exercise.py
Enter length l: 42
Enter width w: 35
Area is 735.0
```

2)

```
x = int(input('Enter x: '))
y = int(input('Enter y: '))
temp =x
x = y
y = temp

print("x is " + str(x))
print("y is " + str(y))
```

```
student@dslab-12:~/190905104_DS/lab2$ python exercise.py
Enter x: 23
Enter y: 12
x is 12
y is 23
```

3)

```
x = int(input('Enter x: '))

if x % 2 == 0:
print("x is " + str("even"))
else:
print("x is " + str("odd"))
```

```
student@dslab-12:~/190905104_DS/lab2$ python exercise.py
Enter x: 55
x is odd
```

4)

```
x = int(input('Enter x: '))
y = int(input('Enter y: '))
```

```
z = int(input('Enter z: '))

if x >= y and x >= z :
print("x is " + str("largest"))

elif y >= x and y >= z :
print("y is " + str("largest"))

elif z >= x and z >= y :
print("z is " + str("largest"))
```

```
student@dslab-12:~/190905104_DS/lab2$ python exercise.py
Enter x: 3
Enter y: 6
Enter z: 5
y is largest
```

5)

```
y = int(input('Enter a number: '))

while(y > 0):
if y%5 ==0:
print( str(y) + " divisible by 5" )
else:
print( str(y) + " is not divisible by 5" )
y= y-1
```

```
student@dslab-12:~/190905104_DS/lab2$ python exercise.py
Enter a number: 7
7 is not divisible by 5
6 is not divisible by 5
5 divisible by 5
4 is not divisible by 5
3 is not divisible by 5
2 is not divisible by 5
1 is not divisible by 5
```

6)

```
a = ["A", "B"]
print(a)
a.append("C")
print(a)
a.extend(["D", "E"])
print(a)
a.insert(0, "F")
print(a)
a.insert(2, "G")
print(a)
a.remove("A")
print(a)
```

```
student@dslab-12:~/190905104_DS/lab2$ python exercise.py
['A', 'B']
['A', 'B', 'C']
['A', 'B', 'C', 'D', 'E']
['F', 'A', 'B', 'C', 'D', 'E']
['F', 'A', 'G', 'B', 'C', 'D', 'E']
['F', 'G', 'B', 'C', 'D', 'E']
```

7)

```
t = (1,2,3,4,5,6,7,8,9,10)

for i in range(int(len(t)/2)):
print(t[i],end = " ")
print("\n")


for i in range(int(len(t)/2),int(len(t))):
print(t[i],end = " ")
```

```
student@dslab-12:~/190905104_DS/lab2$ python exercise.py
1 2 3 4 5

6 7 8 9 10 student@dslab-12:~/190905104_DS/lab2$
```

8)

```
t = (12,7,38,56,78)
t2 = []
for i in range(int(len(t))):
if t[i]%2 == 0:
t2.append(t[i])
tuple_t2 = tuple(t2)
print(tuple_t2)
print(type(tuple_t2))
```

```
(12, 38, 56, 78)
<class 'tuple'>
```

9)

```
l = [11,-21,0,45,66,-93]

[print(i) for i in l if i < 0]
```

```
student@dslab-12:~/190905104_DS/lab2$ python exercise.py
-21
-93
```

10)

```
l = [11,-21,0,45,66,-93]
```

```
i = len(l) - 1
while( i > 0):
if l[i] < 0:
print(l[i])
i = i-1
```

```
student@dslab-12:~/190905104_DS/lab2$ python exercise.py
-93
-21
```

11)

```
l = [11,-21,0,45,66,-93]

pos = 0
neg = 0

for i in l:
if i > 0 :
pos = pos +1
elif i < 0 :
neg = neg + 1

print(l)
print("positive numbers " + str(pos) )
print("negative numbers " + str(neg) )
```

```
student@dslab-12:~/190905104_DS/lab2$ python exercise.py
[11, -21, 0, 45, 66, -93]
positive numbers 3
negative numbers 2
```

12)

```
l = [11,-21,0,45,66,-93]

[l.remove(i) for i in l if i%2 == 0]

[print(i,end = " ") for i in l ]
```

```
11 -21 45 -93
```

13)

```
import pandas as pd

dic = {'name': ['Parth', 'Matt', 'Dan'], 'height': ['2', '1', '3'], 'qualification': ['cool', 'very cool', 'not cool']}
df = pd.DataFrame.from_dict(dic)
print(df)
```

```
addr = ['Manipal', 'Mumbai', 'Delhi']

df['address'] = addr
print(df)
```

```
student@dslab-12:~/190905104_DS/lab2$ python exercise.py
    name height qualification
0  Parth      2          cool
1   Matt      1     very cool
2    Dan      3      not cool
    name height qualification  address
0  Parth      2          cool  Manipal
1   Matt      1     very cool   Mumbai
2    Dan      3      not cool    Delhi
```

14)

```
import pandas as pd

dic = {'name': ['Parth', 'Matt', 'Dan'], 'height': ['2', '1', '3'], 'qualification': ['cool', 'very cool', 'not cool']}
df = pd.DataFrame.from_dict(dic)
print(df)

df.insert(3, "lastcol", ['data1', 'data2', 'data3'])
print(df)
```

```
student@dslab-12:~/190905104_DS/lab2$ python exercise.py
    name height qualification
0  Parth      2          cool
1   Matt      1     very cool
2    Dan      3      not cool
    name height qualification lastcol
0  Parth      2          cool   data1
1   Matt      1     very cool   data2
2    Dan      3      not cool   data3
```

Intro to pandas

1)

```
import pandas as pd
import numpy as np

s = pd.Series([3, -2, 6, 8, 5])

print(s.min())
print(s.max())
print(s.sum())
```

```
student@dslab-12:~/190905104_DS/lab2$ python3 script.py
-2
8
20
```

2)

```
import pandas as pd
data = [['Parth',20],['Matt',22],['Dan',23]]
df = pd.DataFrame(data,columns=['Name','Age'])

print(df)
```

```
student@dslab-12:~/190905104_DS/lab2$ python3 script.py
    Name  Age
0  Parth   20
1   Matt   22
2    Dan   23
```

3)

```
import pandas as pd

data = {'Name':['Parth', 'Sudha', 'Raju'],'Age':[20,34,29]}
df = pd.DataFrame(data, index=['rank1','rank2','rank3'])
print(df)
```

```
student@dslab-12:~/190905104_DS/lab2$ python3 script.py
        Name  Age
rank1  Parth   20
rank2  Sudha   34
rank3   Raju   29
```

4)

```
df1=pd.DataFrame({'A':pd.Timestamp('20130102'),'B':np.array([3]*4,dtype='int32'),
'C':pd.Categorical(['Male','Female','Male','Female'])})

print(df1)
```

```
student@dslab-12:~/190905104_DS/lab2$ python3 script.py
           A  B       C
0 2013-01-02  3    Male
1 2013-01-02  3  Female
2 2013-01-02  3    Male
3 2013-01-02  3  Female
```

5)

```
df1=pd.DataFrame({'A':pd.Timestamp('20130102'),'B':np.array([3]*4,dtype='int32'),
'C':pd.Categorical(['Male','Female','Male','Female'])})
print('First 2 rows')
print(df1.head(2))
```

```
print('Last 2 rows')
print(df1.tail(2))
```

```
print(df1.describe())
```

```
student@dslab-12:~/190905104_DS/lab2$ python3 script.py
First 2 rows
           A  B       C
0 2013-01-02  3    Male
1 2013-01-02  3  Female
Last 2 rows
           A  B       C
2 2013-01-02  3    Male
3 2013-01-02  3  Female
         B
count  4.0
mean   3.0
std    0.0
min    3.0
25%    3.0
50%    3.0
75%    3.0
max    3.0
```

6)

```
dates=pd.date_range('20130101', periods=100)
df = pd.DataFrame(np.random.randn(100,4), index=dates, columns=list('ABCD'))
```

```
print(df.head())
```

```
print(df.index)
```

```
student@dslab-12:~/190905104_DS/lab2$ python3 script.py
                   A         B         C         D
2013-01-01 -0.675542 -1.101237 -0.724333  0.284915
2013-01-02 -0.087000  2.011004  0.729667 -0.092953
2013-01-03 -1.357185  1.529013 -0.234619  0.631925
2013-01-04 -0.369263  0.262422 -1.566752  0.100116
2013-01-05  0.031329 -0.276000  0.479649  1.242280
```

```
student@dslab-12:~/190905104_DS/lab2$ python3 script.py
DatetimeIndex(['2013-01-01', '2013-01-02', '2013-01-03', '2013-01-04',
               '2013-01-05', '2013-01-06', '2013-01-07', '2013-01-08',
               '2013-01-09', '2013-01-10', '2013-01-11', '2013-01-12',
               '2013-01-13', '2013-01-14', '2013-01-15', '2013-01-16',
               '2013-01-17', '2013-01-18', '2013-01-19', '2013-01-20',
               '2013-01-21', '2013-01-22', '2013-01-23', '2013-01-24',
               '2013-01-25', '2013-01-26', '2013-01-27', '2013-01-28',
               '2013-01-29', '2013-01-30', '2013-01-31', '2013-02-01',
               '2013-02-02', '2013-02-03', '2013-02-04', '2013-02-05',
               '2013-02-06', '2013-02-07', '2013-02-08', '2013-02-09',
               '2013-02-10', '2013-02-11', '2013-02-12', '2013-02-13',
               '2013-02-14', '2013-02-15', '2013-02-16', '2013-02-17',
               '2013-02-18', '2013-02-19', '2013-02-20', '2013-02-21',
               '2013-02-22', '2013-02-23', '2013-02-24', '2013-02-25',
               '2013-02-26', '2013-02-27', '2013-02-28', '2013-03-01',
               '2013-03-02', '2013-03-03', '2013-03-04', '2013-03-05',
               '2013-03-06', '2013-03-07', '2013-03-08', '2013-03-09',
               '2013-03-10', '2013-03-11', '2013-03-12', '2013-03-13',
               '2013-03-14', '2013-03-15', '2013-03-16', '2013-03-17',
               '2013-03-18', '2013-03-19', '2013-03-20', '2013-03-21',
               '2013-03-22', '2013-03-23', '2013-03-24', '2013-03-25',
               '2013-03-26', '2013-03-27', '2013-03-28', '2013-03-29',
               '2013-03-30', '2013-03-31', '2013-04-01', '2013-04-02',
               '2013-04-03', '2013-04-04', '2013-04-05', '2013-04-06',
               '2013-04-07', '2013-04-08', '2013-04-09', '2013-04-10'],
              dtype='datetime64[ns]', freq='D')
```

```
print(df.columns)
```

```
student@dslab-12:~/190905104_DS/lab2$ python3 script.py
Index(['A', 'B', 'C', 'D'], dtype='object')
```

```
print(df.T)
```

```
student@dslab-12:~/190905104_DS/lab2$ python3 script.py
   2013-01-01  2013-01-02  2013-01-03  2013-01-04  2013-01-05  ...  2013-04-06  2013-04-07  2013-04-08  2013-04-09  2013-04-10
A   -0.539078    0.939313   -0.559948   -0.089752   -1.556837  ...    1.031146   -0.018921   -0.919047    0.623761    1.286103
B    1.382097   -1.905341   -0.220232    0.482165    0.535605  ...    0.690239   -0.398830   -0.351780    0.586489   -0.422121
C    1.572518   -0.986951    0.424774    0.452059    1.033029  ...   -0.058239   -1.273623    1.424389   -1.210888   -0.361333
D   -0.928857    0.208984   -1.891328   -2.376855   -1.046181  ...   -0.455801    0.624872    1.399976    0.363846    0.364807

[4 rows x 100 columns]
```

```
print(df.sort_index(axis=1, ascending=False) )
```

```
student@dslab-12:~/190905104_DS/lab2$ python3 script.py
                   D         C         B         A
2013-01-01 -0.024313  0.759672 -0.046301  0.804864
2013-01-02 -2.040299  0.147212  1.563354 -1.322301
2013-01-03 -0.379823  0.517726 -1.494301 -0.543223
2013-01-04  0.340134 -1.003598 -0.277621 -0.379904
2013-01-05  0.779578 -0.223029  0.644031  0.070089
...              ...       ...       ...       ...
2013-04-06 -0.677025  0.778731  0.012434 -0.431493
2013-04-07 -1.178883  1.116184 -0.670255  0.422219
2013-04-08 -0.826369  0.026847  0.404022  0.699582
2013-04-09 -0.826461  0.590436  1.893385  0.261047
2013-04-10 -1.570556  2.139014  0.206243  1.083684

[100 rows x 4 columns]
```

```
print(df.sort_values(by='B'))
```

```
student@dslab-12:~/190905104_DS/lab2$ python3 script.py
                   A         B         C         D
2013-02-14  0.703089 -3.337957 -0.292366 -1.189996
2013-01-13  0.360657 -2.468972  0.436076 -1.424446
2013-02-05 -1.193088 -1.894973 -0.149242  0.592118
2013-03-14 -0.621309 -1.474254 -0.139266 -1.100982
2013-02-07 -0.331278 -1.438995  1.105630 -0.778491
...              ...       ...       ...       ...
2013-01-20 -1.837231  1.839984  1.572317  0.412579
2013-01-02 -0.168887  1.921490  0.519160  0.942068
2013-03-29  2.207714  1.986890 -0.506933 -0.189475
2013-03-28  0.626393  2.089835 -0.567804  1.765032
2013-03-03 -0.161848  2.379935 -0.119812  1.545374

[100 rows x 4 columns]
```

```
print(df[0:3])
```

```
student@dslab-12:~/190905104_DS/lab2$ python3 script.py
                   A         B         C         D
2013-01-01  0.184034  0.760374 -1.250177 -0.295391
2013-01-02 -0.004499  0.632744 -0.929525  0.667386
2013-01-03 -1.155616 -0.051348 -1.157979  0.681448
```

```
print(df['20130105':'20130110'] )
```

```
student@dslab-12:~/190905104_DS/lab2$ python3 script.py
                   A         B         C         D
2013-01-05  1.349338  0.546032 -0.945738  0.362859
2013-01-06  0.129282  0.222750 -0.021711  0.144728
2013-01-07  0.089755 -0.389804  0.159009  1.365910
2013-01-08 -0.353758 -0.602043 -1.233626 -0.431123
2013-01-09 -1.365402 -0.970264 -0.502474  0.036164
2013-01-10  2.028221  1.736034  1.012697 -0.252455
```

```
print(df.iloc[0])
```

```
student@dslab-12:~/190905104_DS/lab2$ python3 script.py
A    1.035751
B   -0.629473
C   -0.287010
D    0.228532
Name: 2013-01-01 00:00:00, dtype: float64
```

```
print( df.iloc[0,:2])
```

```
student@dslab-12:~/190905104_DS/lab2$ python3 script.py
A   -0.189255
B    1.952795
Name: 2013-01-01 00:00:00, dtype: float64
```

```
print(df[['A','B']][:5])
```

```
student@dslab-12:~/190905104_DS/lab2$ python3 script.py
                   A         B
2013-01-01 -0.072484 -1.857473
2013-01-02 -0.010504  2.361396
2013-01-03 -0.311103  0.197702
2013-01-04 -0.498058  0.423065
2013-01-05 -0.403549 -0.616584
```

7) Boolean indexing

```
print(df[df.A>0])
```

```
student@dslab-12:~/190905104_DS/lab2$ python3 script.py
                   A         B         C         D
2013-01-02  0.569049 -0.033800  0.811761  0.774076
2013-01-03  0.164219  0.430351  1.159527  0.540081
2013-01-04  0.417359 -0.023139 -0.449346 -1.165163
2013-01-06  0.211703  0.087400 -0.023205 -0.195963
2013-01-07  0.725637  0.411814 -0.054427 -1.030651
2013-01-10  0.048610 -0.826534 -0.830960 -0.843953
2013-01-11  0.195841  0.052492 -0.098065  0.790641
2013-01-13  0.821585 -1.512436 -0.211592  0.415321
2013-01-14  0.261011 -0.353866  1.126577 -0.843941
2013-01-15  0.440873  0.153600 -1.614943  0.098175
2013-01-17  0.628822  1.648735  0.622613 -0.938515
2013-01-19  0.498700  0.497938  0.269903 -1.004091
2013-01-21  0.265631  0.795552  1.692459 -0.968352
```

```
genders = ['M', 'F', 'NB']
df['F'] = np.random.choice(genders, len(df))
```

```
                   A         B         C         D   F
2013-01-01  1.274465  0.957498  0.449968 -0.710568  NB
2013-01-02 -1.565072  1.175652  0.068870 -1.280112   F
2013-01-03  0.372865 -1.974254  1.407981  2.414408   M
2013-01-04 -1.053805 -0.590624 -0.987818 -1.802933   F
2013-01-05 -0.349506 -2.455996  0.140945  1.855518   F
...              ...       ...       ...       ...  ..
2013-04-06 -0.857001 -0.276207 -0.120491  0.748328   M
2013-04-07 -0.299276 -0.048452 -0.538772  0.646515   F
2013-04-08  0.520052  0.939581  0.501035  0.508976   F
2013-04-09 -0.374454  0.120753 -0.331423 -0.496277  NB
2013-04-10 -0.454771  2.586094 -1.045410 -0.070157  NB
```

8) Drop

```
df.drop('B', inplace=True, axis = 1)
df.drop('20130102', axis = 0, inplace=True, )
```

9) Concatenate dataframes

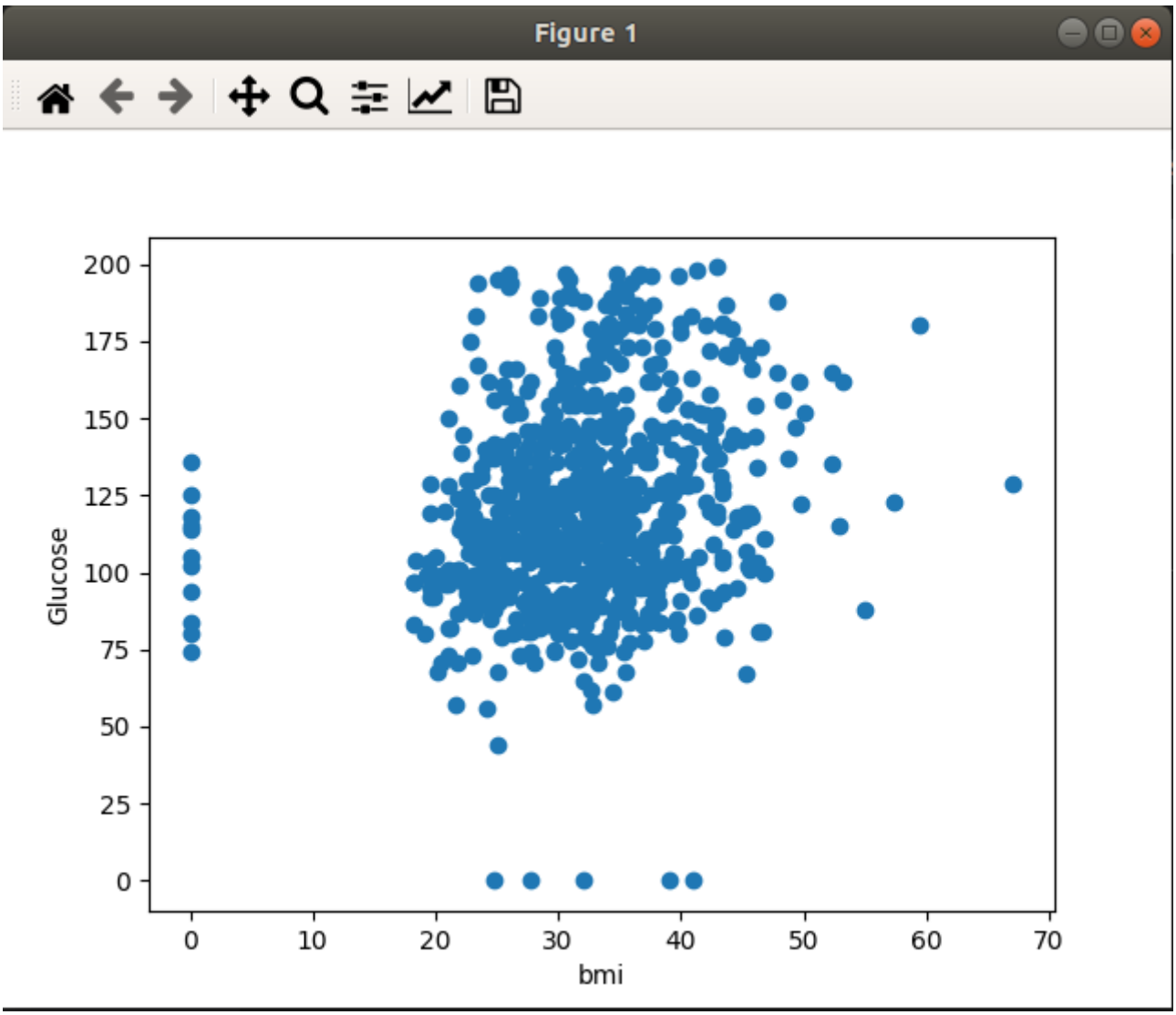
```
dates=pd.date_range('20220101', periods=10)
df2 = pd.DataFrame(np.random.randn(10,3), index=dates, columns=list('XYZ'))

dates=pd.date_range('20130101', periods=10)
df1 = pd.DataFrame(np.random.randn(10,5), index=dates, columns=list('ABCDE'))
```

```
df_new = pd.concat([df1, df2])
print(df_new)
```

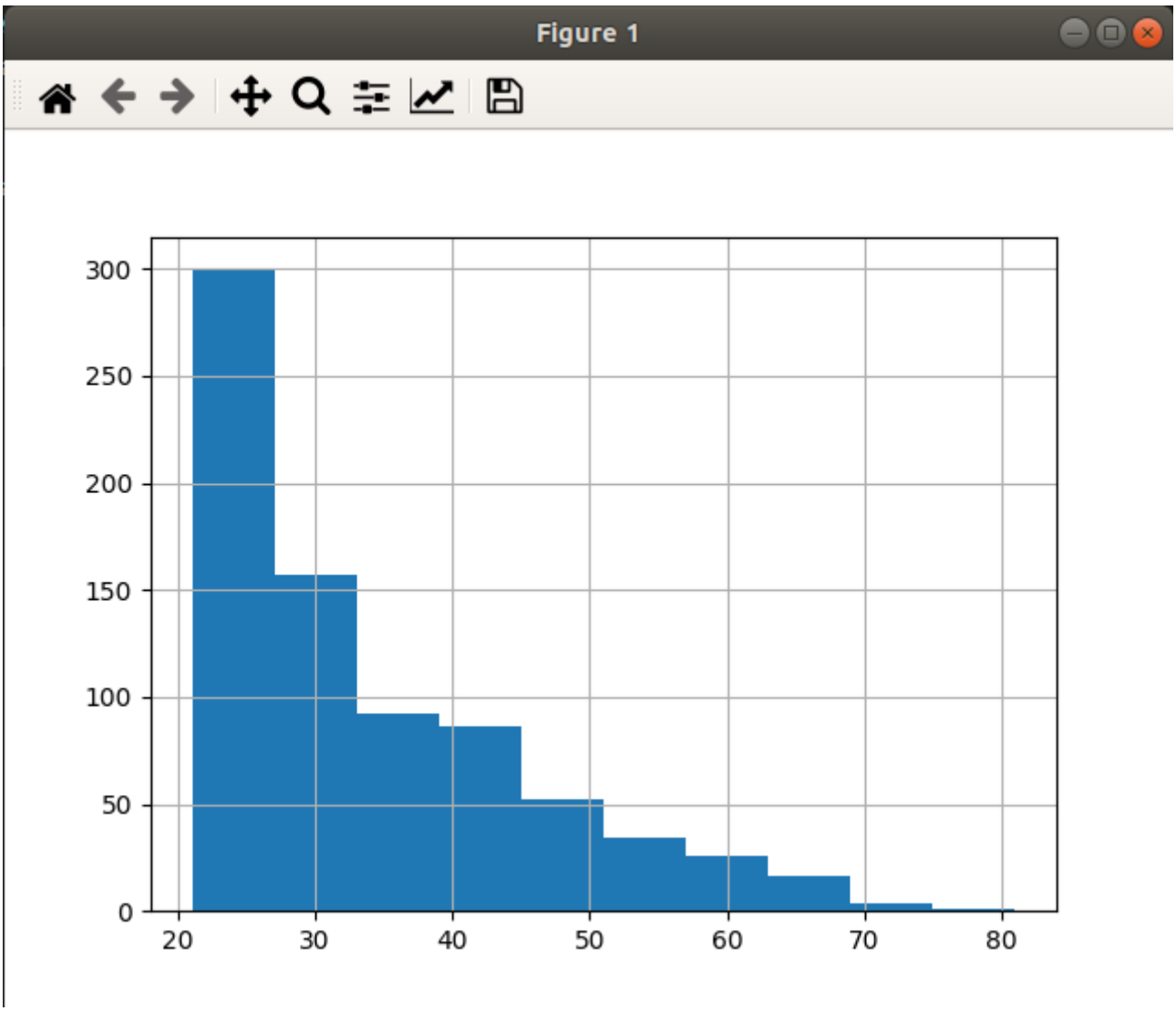
```
student@dslab-12:~/190905104_DS/lab2$ python3 script.py
                   A         B         C         D         E         X         Y         Z
2013-01-01 -0.007793 -0.113755 -0.167071 -0.222423 -0.710486       NaN       NaN       NaN
2013-01-02  0.350383 -0.199335 -0.874095  0.599177 -0.221561       NaN       NaN       NaN
2013-01-03  1.496752 -0.679251  2.232947 -0.760742  1.214697       NaN       NaN       NaN
2013-01-04  1.719527 -0.662335  0.610854 -0.265787 -0.195554       NaN       NaN       NaN
2013-01-05  0.408159  0.731179 -2.006198 -0.034437 -0.625654       NaN       NaN       NaN
2013-01-06 -0.298380  0.219497 -2.031651  1.159191 -0.112844       NaN       NaN       NaN
2013-01-07 -2.016627  0.601354  0.193538  1.022315 -0.536406       NaN       NaN       NaN
2013-01-08  0.819938  0.895650 -0.132478 -1.581577 -1.019130       NaN       NaN       NaN
2013-01-09  0.954148  0.843752 -0.200342 -2.158209 -0.234638       NaN       NaN       NaN
2013-01-10  0.728611 -0.921391 -1.116660 -0.868779  0.914801       NaN       NaN       NaN
2022-01-01       NaN       NaN       NaN       NaN       NaN -1.074829 -0.963077  0.807474
2022-01-02       NaN       NaN       NaN       NaN       NaN  0.314931 -0.852455  1.055514
2022-01-03       NaN       NaN       NaN       NaN       NaN -1.167112 -0.362150 -0.075295
2022-01-04       NaN       NaN       NaN       NaN       NaN  1.852903 -0.102454 -0.617004
2022-01-05       NaN       NaN       NaN       NaN       NaN  0.004818  0.630298 -0.079847
2022-01-06       NaN       NaN       NaN       NaN       NaN  0.867987 -0.123674  0.837254
2022-01-07       NaN       NaN       NaN       NaN       NaN -0.094669  1.153133 -0.289412
2022-01-08       NaN       NaN       NaN       NaN       NaN -0.541776  1.248340 -1.108088
2022-01-09       NaN       NaN       NaN       NaN       NaN -0.982245 -0.496337  0.362475
2022-01-10       NaN       NaN       NaN       NaN       NaN -0.336266  1.646784  0.418664
student@dslab-12:~/190905104_DS/lab2$
```

Matplotlib

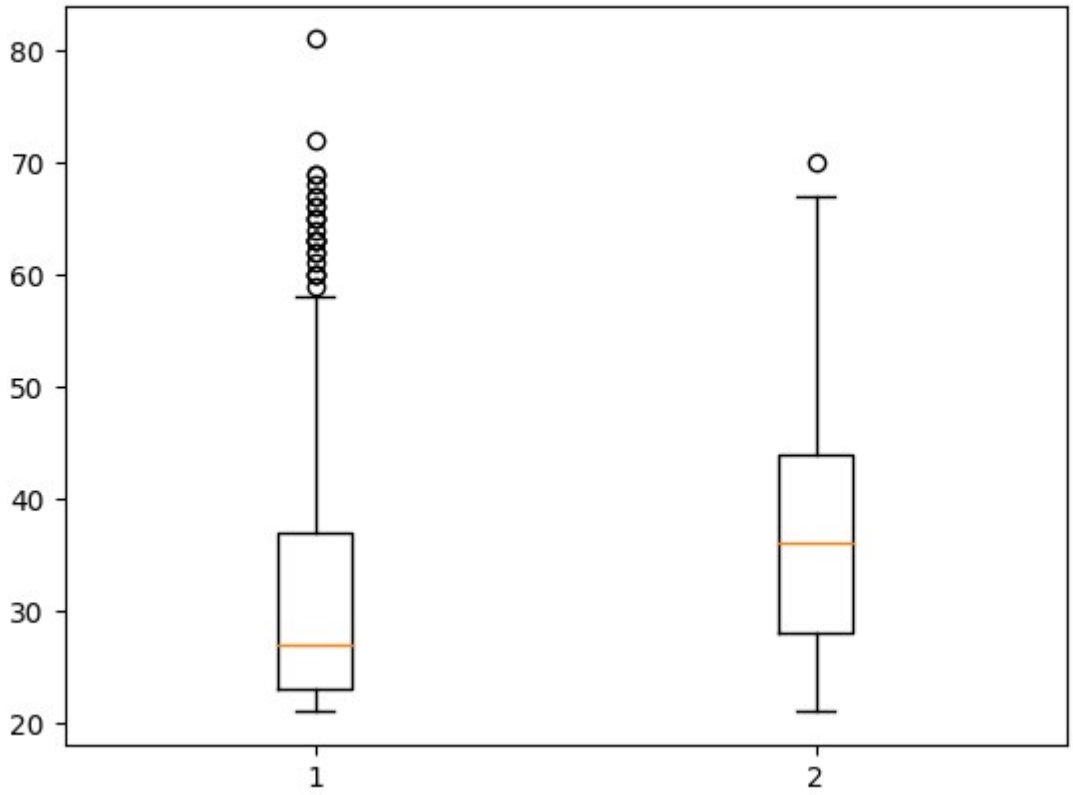
```
import matplotlib.pyplot as plt
import pandas as pd

df = pd.read_csv('prima_indians_diabetes_for_Week2.csv', header=None)

df.columns=['preg','glu','bp','sft','ins','bmi','dpf','age','class']

plt.scatter(df['bmi'],df['glu'])
plt.xlabel('bmi')
plt.ylabel('Glucose')

plt.show()
```




```
df['age'].hist()
plt.show()
```




```
plt.boxplot([df.loc[df['class']==0]['age'], df.loc[df['class']==1]['age']])
plt.show()
```




```
G=pd.read_excel('German Credit_for_Week2.xlsx',sheet_name='Sheet1',engine='openpyxl')
print(G.head())
```

```
student@dslab-12:~/190905104_DS/lab2$ python plotting.py
   Creditability  CreditAmount  DurationOfCreditInMonths
0              1          1049                        18
1              1          2799                         9
2              1           841                        12
3              1          2122                        12
4              1          2171                        12
```