Name: Shreyas Kamath
Roll: 07
Sec: B
Reg-No: 190905022
Lab: 02

Solved Exersises (Solved in Class)

```
import numpy as np
import pandas as pd
s=pd.Series([3,9,-2,10,5])
print(s.sum())
print(s.min())
print(s.max())
```

```
student@V310Z-000:~/Documents/190905022_dslab/week2$ python sample.py
25
-2
10
```

```
import pandas as pd
data=[['dinesh',10],['shreyas',12],['Raj',13]]
df=pd.DataFrame(data,columns=['Name','Age'])
print(df)
```

```
student@V310Z-000:~/Documents/190905022_dslab/week2$ python3 s1.py
      Name  Age
0   dinesh   10
1  shreyas   12
2      Raj   13
```

```
import pandas as pd

data={'Name':['Kavitha','sudha','Raju','Vignesh'],'Age':[28,34,29,42]}
df=pd.DataFrame(data,index=['rank1','rank2','rank3','rank4'])
print(df)
```

```
student@V310Z-000:~/Documents/190905022_dslab/week2$ python sample.py
       Age     Name
rank1   28  Kavitha
rank2   34    sudha
rank3   29     Raju
rank4   42  Vignesh
```

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

```
student@V310Z-000:~/Documents/190905022_dslab/week2$ python3 s1.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
```

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

print(df1.head())
print(df1.tail())
print(df1.describe())
```

```
student@V310Z-000:~/Documents/190905022_dslab/week2$ python3 s1.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
(4, 3)
           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
           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
         B
count  4.0
mean   3.0
std    0.0
min    3.0
25%    3.0
50%    3.0
```

```
import pandas as pd
import numpy as np

dates=pd.date_range('20130101',periods=100)

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

print(df.tail())
print(df.head())

print(df.index)

print(df.columns)

df.T
print(df.T)

df.sort_index(axis=1,ascending=False)
print(df)
df.sort_values(by='B')

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

```
student@V310Z-000:~/Documents/190905022_dslab/week2$ python sample.py
                   A         B         C         D
2013-04-06 -0.422489  2.551101  0.045432  0.160338
2013-04-07 -0.260260 -1.243564 -0.415989 -0.092309
2013-04-08  0.146545  1.808143  0.710867 -0.715895
2013-04-09  0.801257 -2.078110 -0.888695  0.517201
2013-04-10 -1.583524  1.114768  1.497705  1.532570
                   A         B         C         D
2013-01-01 -0.775936 -0.100750  0.779962  1.575277
2013-01-02 -0.400234  1.242758  1.011418  0.578207
2013-01-03 -0.489189  1.385622 -0.961877 -0.314943
2013-01-04 -0.685464 -1.191260  0.428329 -0.035071
2013-01-05 -1.462577 -0.318984 -0.780645 -0.271261
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',
```

```
Index([u'A', u'B', u'C', u'D'], dtype='object')
   2013-01-01  2013-01-02  2013-01-03  ...  2013-04-08  2013-04-09  2013-04-10
A    0.636024   -0.243795    0.647504  ...    0.417890   -2.026812    1.220225
B   -1.798056    1.460423    0.150956  ...    0.693125    1.415365   -0.063331
C   -0.103079    0.218760   -0.994054  ...    0.928888    1.514047    1.925075
D   -0.274215    0.009759    2.997407  ...    1.337150   -1.769977    0.725991

[4 rows x 100 columns]
                   A         B         C         D
2013-01-01  0.636024 -1.798056 -0.103079 -0.274215
2013-01-02 -0.243795  1.460423  0.218760  0.009759
2013-01-03  0.647504  0.150956 -0.994054  2.997407
2013-01-04  0.218276  0.195297  0.399782  0.534067
2013-01-05  0.654916 -0.003493  0.722732 -1.486053
2013-01-06 -1.088096 -0.281169 -0.104729 -1.517349
2013-01-07 -0.318490  0.622933 -0.132408  0.878398
2013-01-08 -0.077669 -1.592769 -0.673673  0.011755
2013-01-09 -1.709003  1.350271 -1.032003  0.471892
2013-01-10 -1.168341 -0.635973  0.495871  0.051844
2013-01-11 -0.839495 -0.153006 -1.254133  0.828176
2013-01-12  0.607046  0.944160  1.355778  1.738568
2013-01-13  2.548360 -0.418590 -0.137570  1.453927
2013-01-14 -0.840869  0.121800 -0.077830  0.374365
2013-01-15  0.514252 -0.860169  0.884045  1.166975
2013-01-16 -0.284684 -0.812792 -2.466582  2.429035
2013-01-17 -1.200213 -0.767096  1.637072  0.727382
2013-01-18  0.900130  1.583399 -0.715910  0.360752
2013-01-19  1.259096 -1.327698  0.004762  0.566217
2013-01-20  0.545712  0.236771  0.319087 -1.403607
```

```
[100 rows x 4 columns]
2013-01-01    0.636024
2013-01-02   -0.243795
2013-01-03    0.647504
2013-01-04    0.218276
2013-01-05    0.654916
2013-01-06   -1.088096
2013-01-07   -0.318490
2013-01-08   -0.077669
2013-01-09   -1.709003
2013-01-10   -1.168341
2013-01-11   -0.839495
2013-01-12    0.607046
2013-01-13    2.548360
2013-01-14   -0.840869
2013-01-15    0.514252
2013-01-16   -0.284684
2013-01-17   -1.200213
2013-01-18    0.900130
2013-01-19    1.259096
2013-01-20    0.545712
2013-01-21   -0.424974
2013-01-22    0.578928
2013-01-23   -1.326453
2013-01-24   -0.591839
2013-01-25   -1.080408
2013-01-26   -0.141881
2013-01-27    0.769946
2013-01-28    2.222005
2013-01-29    1.638878
```

```
2013-03-12 -1.762340  0.462061
2013-03-13  0.366423 -1.289116
2013-03-14 -1.214460 -1.275386
2013-03-15  0.376203 -0.394172
2013-03-16  0.194865  1.018148
2013-03-17  0.919653 -0.102399
2013-03-18  2.027834 -0.170079
2013-03-19 -0.171475  1.561854
2013-03-20 -1.428691 -1.067315
2013-03-21 -0.324278 -1.448144
2013-03-22  1.736760  0.523338
2013-03-23  0.662068  1.120210
2013-03-24 -1.652288  0.720678
2013-03-25  1.402294  0.582122
2013-03-26  1.710701  0.496772
2013-03-27  0.636174 -0.508617
2013-03-28  0.922040 -1.180518
2013-03-29  0.716985  0.941910
2013-03-30  0.108162 -0.791903
2013-03-31 -0.702479  0.684689
2013-04-01  1.317144  1.777658
2013-04-02 -1.647949  0.228275
2013-04-03 -0.195562  0.593098
2013-04-04  1.334488 -0.012341
2013-04-05  0.881573  0.259834
2013-04-06  0.546425 -0.214801
2013-04-07  0.453023 -0.286726
2013-04-08  0.417890  0.693125
2013-04-09 -2.026812  1.415365
2013-04-10  1.220225 -0.063331
```

```
[100 rows x 2 columns]
                   A         B
2013-01-01  0.636024 -1.798056
2013-01-02 -0.243795  1.460423
2013-01-03  0.647504  0.150956
2013-01-04  0.218276  0.195297
2013-01-05  0.654916 -0.003493
```

Boolean indexing

selecting positive values

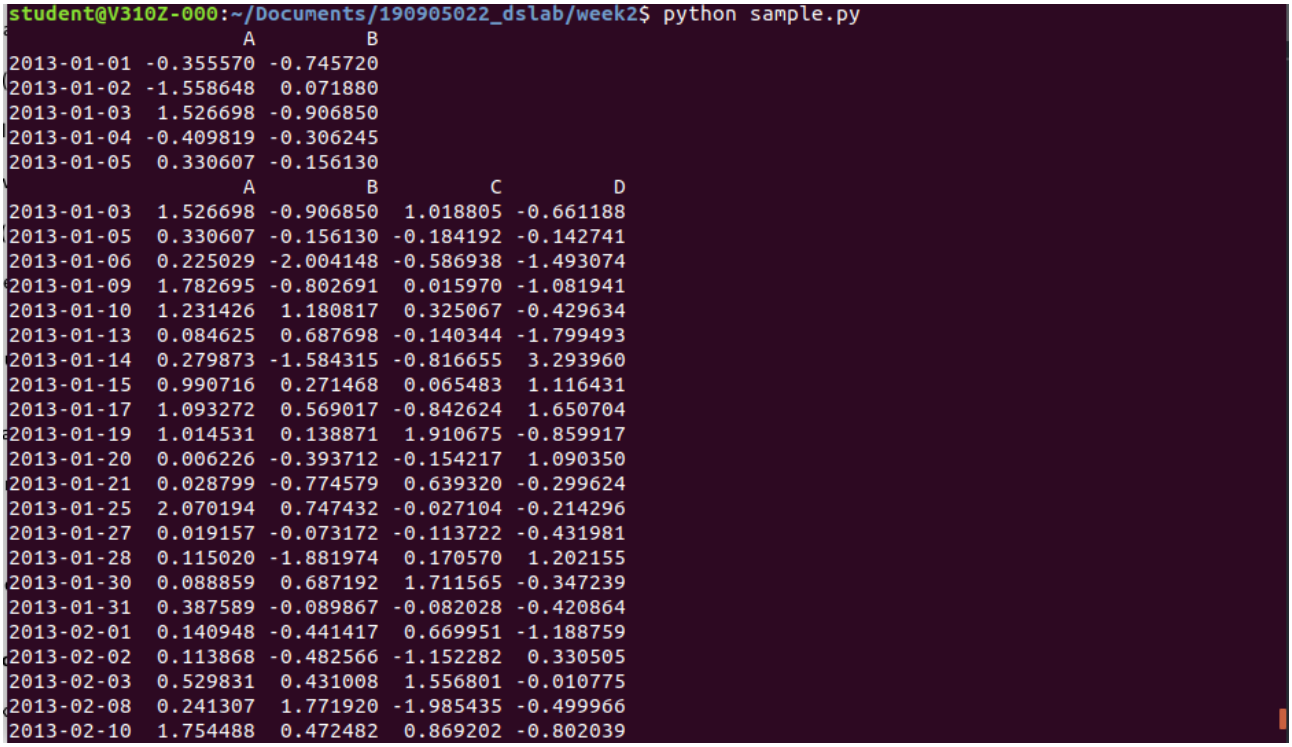
```
import pandas as pd
import numpy as np

dates=pd.date_range('20130101',periods=100)

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


df.T
```

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

df.sort_values(by='B')

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

print(df[df.A>0])



import pandas as pd
import numpy as np

dates=pd.date_range('20130101',periods=100)

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

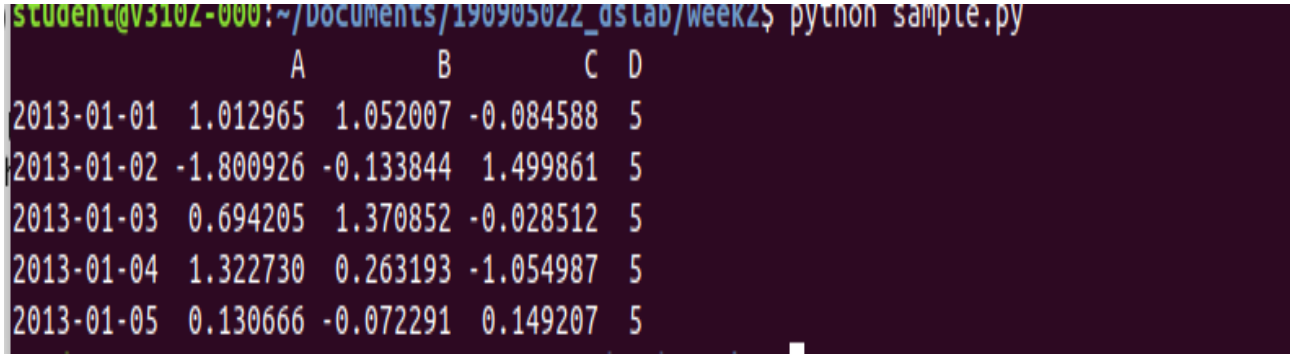
df.T

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

df.sort_values(by='B')

df.loc[:,'D']=np.array([5]*len(df))

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



import pandas as pd
import numpy as np

dates=pd.date_range('20130101',periods=100)

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

df.T

df.sort_index(axis=1,ascending=False)
df.sort_values(by='B')

df.loc[:,'D']=np.array([5]*len(df))

df.drop('C',axis=1,inplace=True)
print(df[:10])
df.drop(df.index[1:2], axis =0, inplace=True)

print(df[:10])

```
student@V310Z-000:~/Documents/190905022_dslab/week2$ python sample.py
                   A         B  D
2013-01-01  1.806032  1.550688  5
2013-01-02 -0.502459  0.630446  5
2013-01-03 -0.846956 -0.719034  5
2013-01-04  0.044196  0.001763  5
2013-01-05  0.865882  1.493340  5
2013-01-06  0.131096  0.315774  5
2013-01-07 -1.717306  0.120522  5
2013-01-08  0.579298 -1.214472  5
2013-01-09 -0.019564 -1.249152  5
2013-01-10 -0.724667 -0.432452  5
                   A         B  D
2013-01-01  1.806032  1.550688  5
2013-01-03 -0.846956 -0.719034  5
2013-01-04  0.044196  0.001763  5
2013-01-05  0.865882  1.493340  5
2013-01-06  0.131096  0.315774  5
2013-01-07 -1.717306  0.120522  5
2013-01-08  0.579298 -1.214472  5
2013-01-09 -0.019564 -1.249152  5
2013-01-10 -0.724667 -0.432452  5
2013-01-11  0.567638 -0.396599  5
```

concatinate

```
import pandas as pd
import numpy as np

dates=pd.date_range('20130101',periods=10)

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

df2=pd.DataFrame(np.random.randn(10,3),index=dates,columns=list('ABC'))

Df_new=pd.concat((df1,df2),axis=1)

print(Df_new)
print(Df_new.shape)
```

```
student@V310Z-000:~/Documents/190905022_dslab/week2$ python sample1.py
                   A         B         C         D         E         A         B         C
2013-01-01 -0.568673  0.992769 -0.004202  0.277238  0.224792  1.156317  0.622590  2.252316
2013-01-02  0.066814  1.880162  0.631122 -0.720810 -1.465164  1.548985  0.834571 -0.988120
2013-01-03  0.041403  0.785294 -0.564383 -0.958841 -0.976001 -0.499922 -0.140062  1.346770
2013-01-04 -0.996490  1.390782  1.296198  0.188554  0.058785  1.640849 -1.535615 -0.092101
2013-01-05 -1.012787 -1.046151 -1.216767 -1.751929  0.414850  1.121580  0.776597  0.857540
2013-01-06  0.376308 -0.855052  1.677788  1.321952  0.509866  1.797511  0.128367  0.696187
2013-01-07  0.338316 -2.275686 -0.177180 -0.243272  0.108262 -0.982292  0.384080 -0.415235
2013-01-08  0.206484 -2.728132 -1.757991 -0.606445 -3.619257  0.365566 -0.880899  0.711043
2013-01-09  0.981182 -0.128858  0.631890 -0.659232 -0.098791 -0.606735 -0.313258  0.901039
2013-01-10  0.795489  0.283606 -0.382681 -1.852843  0.204227 -0.725435  0.335619 -0.975663
(10, 8)
```

```
import pandas as pd
import numpy as np


df1=pd.DataFrame(np.random.randn(10,5))

df2=pd.DataFrame(np.random.randn(15,5))

Df_new=pd.concat((df1,df2),axis=0)

print(Df_new)
print(Df_new.shape)
```

```
student@V310Z-000:~/Documents/190905022_dslab/week2$ python sample1.py
           0         1         2         3         4
0  -1.421090 -0.641525 -0.440978 -0.816031 -1.332856
1   0.923953 -0.890647  0.116895  0.986705 -1.189054
2  -2.804611 -1.006509 -1.477749 -0.166934 -1.940861
3   1.041083  1.574614  2.072933 -3.014366  0.286408
4  -2.639666  0.880171 -1.224833 -0.209007  0.023201
5  -0.925583 -0.195023 -0.981321  0.226989 -0.762903
6  -0.227847 -1.277712 -0.011632 -1.370734  0.332452
7  -1.268663  0.217040 -0.389965  1.087192  0.543145
8   0.480754 -0.060478 -0.362671 -0.175721  1.488431
9  -0.823546 -1.201603  0.353002 -0.439864  1.383286
0  -0.199895 -1.879134  0.372048  0.117624  1.979234
1   1.066428  2.066992  0.330216 -1.148260 -0.628658
2  -0.587658  0.287394  0.053801  1.439447  0.160243
3   0.716904  0.307085 -0.003284 -1.404545 -0.131953
4   1.717588 -0.646074  0.807112 -0.507413 -0.297678
5  -0.333708 -0.870172 -0.915215 -0.618669 -0.132153
6  -0.920966  0.857333  0.598381 -1.512017  0.627389
7   2.033254 -0.069334  0.691520 -0.246981  0.358470
8   0.277652  1.127500  1.566464 -1.714824 -2.840642
9  -0.755064 -0.886931  0.594848  0.644633 -1.044854
10 -1.227604 -0.104807  1.254421  0.433839  2.136237
11  0.645065  0.720376 -1.153137  1.869562  0.356594
12 -0.573856 -1.158064 -0.234770  0.378619  1.693877
13  0.284629  0.084078 -0.688257  0.146161 -1.034675
14  1.743673  0.235009 -0.044535 -0.594768  0.972436
(25, 5)
```

```
import pandas as pd
import numpy as np
G=pd.read_excel('German Credit_for_Week2.xlsx',sheet_name='Sheet1',engine='openpyxl')
print(G.head())
D= np.loadtxt('diabetics.csv',delimiter=",")
D[:5,:]
print("Boxplot")
df.boxplot(column=['age'],by='class')
plt.show()
print()
print(D)
```

set2

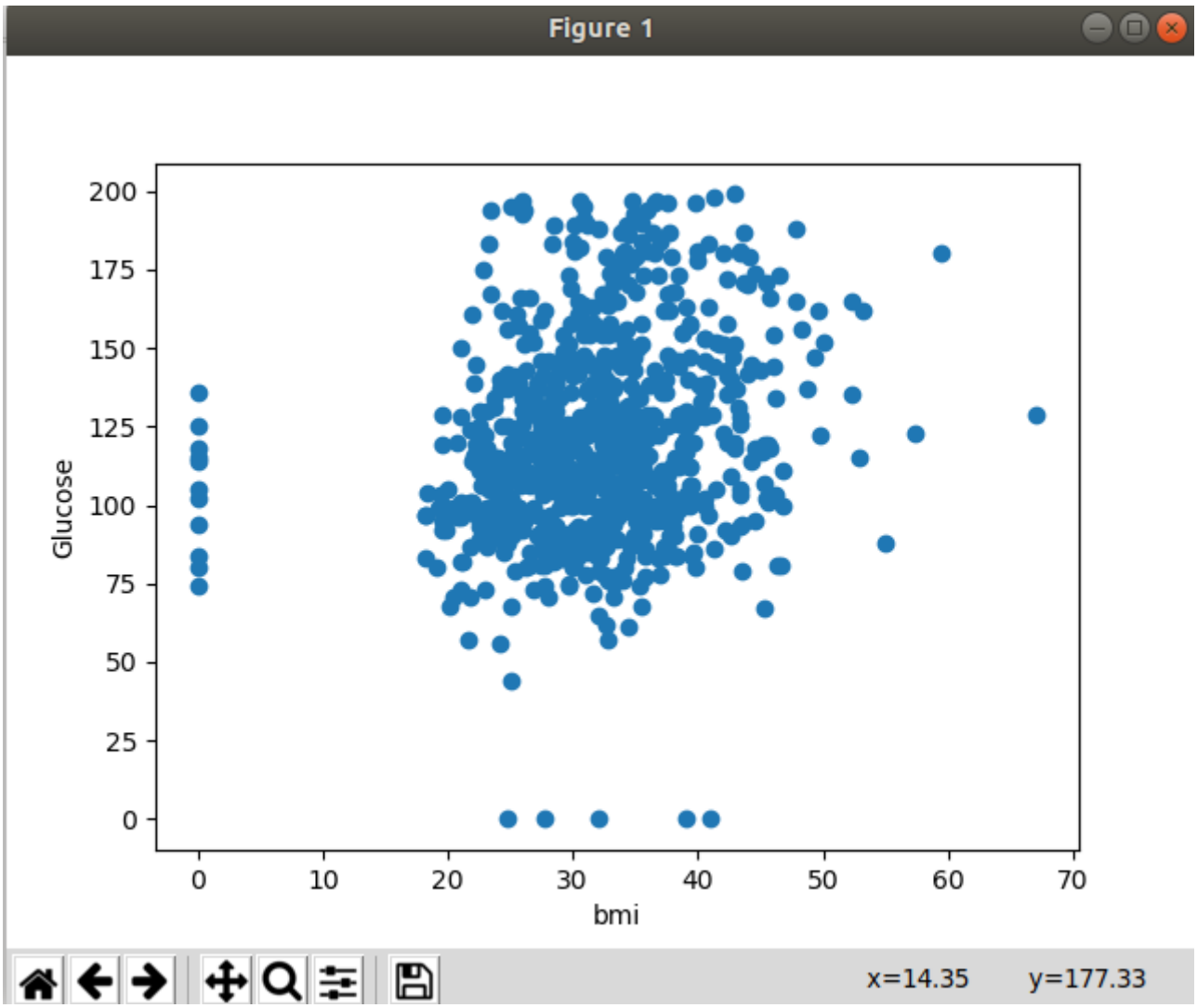
```
import pandas as pd
df=pd.read_csv('diabetics.csv',header=None)
print(df.head())
print(df.tail())
```

```
student@V310Z-000:~/Documents/190905022_dslab/week2/set2$ python3 sample.py
   0    1   2   3    4     5      6   7  8
0  6  148  72  35    0  33.6  0.627  50  1
1  1   85  66  29    0  26.6  0.351  31  0
2  8  183  64   0    0  23.3  0.672  32  1
3  1   89  66  23   94  28.1  0.167  21  0
4  0  137  40  35  168  43.1  2.288  33  1
      0    1   2   3    4     5      6   7  8
763  10  101  76  48  180  32.9  0.171  63  0
764   2  122  70  27    0  36.8  0.340  27  0
765   5  121  72  23  112  26.2  0.245  30  0
766   1  126  60   0    0  30.1  0.349  47  1
767   1   93  70  31    0  30.4  0.315  23  0
```

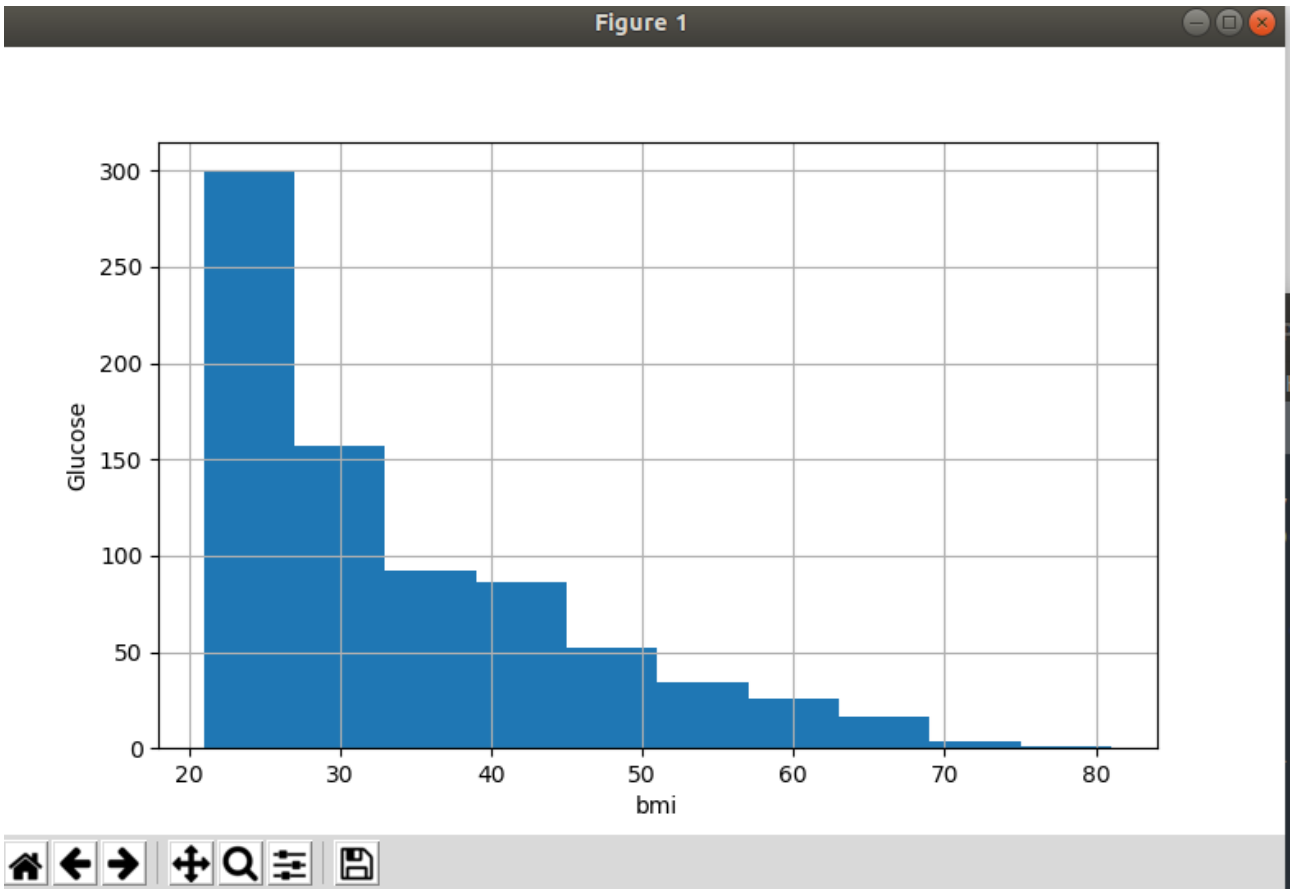
```
import pandas as pd
df=pd.read_csv('diabetics.csv',header=None)

df.columns=['preg','glu','bp','sft','ins','bmi','dpf','age','class']
print(df.head())
print(df.tail())
```

```
student@V310Z-000:~/Documents/190905022_dslab/week2/set2$ python3 sample.py
   preg  glu  bp  sft  ins   bmi    dpf  age  class
0     6  148  72   35    0  33.6  0.627   50      1
1     1   85  66   29    0  26.6  0.351   31      0
2     8  183  64    0    0  23.3  0.672   32      1
3     1   89  66   23   94  28.1  0.167   21      0
4     0  137  40   35  168  43.1  2.288   33      1
     preg  glu  bp  sft  ins   bmi    dpf  age  class
763    10  101  76   48  180  32.9  0.171   63      0
764     2  122  70   27    0  36.8  0.340   27      0
765     5  121  72   23  112  26.2  0.245   30      0
766     1  126  60    0    0  30.1  0.349   47      1
767     1   93  70   31    0  30.4  0.315   23      0
```

```
import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_csv('diabetics.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()
```




```
import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_csv('diabetics.csv',header=None)


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

plt.xlabel('bmi')
plt.ylabel('Glucose')
df['age'].hist()
plt.show()
```




```
import pandas as pd
import numpy as np
W=pd.read_csv('wine_for_Week2.xls',header=None)
print(W.head())
#D=np.loadtxt('sample.data',delimeter=",")
```

```
student@V310Z-000:~/Documents/190905022_dslab/week2/set2$ python3 s1.py
   0      1     2     3     4   ...    9     10    11    12    13
0   1  14.23  1.71  2.43  15.6  ...  2.29  5.64  1.04  3.92  1065
1   1  13.20  1.78  2.14  11.2  ...  1.28  4.38  1.05  3.40  1050
2   1  13.16  2.36  2.67  18.6  ...  2.81  5.68  1.03  3.17  1185
3   1  14.37  1.95  2.50  16.8  ...  2.18  7.80  0.86  3.45  1480
4   1  13.24  2.59  2.87  21.0  ...  1.82  4.32  1.04  2.93   735

[5 rows x 14 columns]
```

```
import pandas as pd
import numpy as np
G=pd.read_excel('German Credit_for_Week2.xlsx',sheet_name='Sheet1',engine='openpyxl')
print(G.head())
```

```
student@V310Z-000:~/Documents/190905022_dslab/week2/set2$ python3 s1.py
   Creditability  CreditAmount  DurationOfCreditInMonths
0              1          1049                        18
1              1          2799                         9
2              1           841                        12
3              1          2122                        12
4              1          2171                        12
```

```
import pandas as pd
import numpy as np

D= np.loadtxt('diabetics.csv',delimiter=",")
D[:5,:]
print(D)
```

```
student@V310Z-000:~/Documents/190905022_dslab/week2/set2$ python3 s1.py
   Creditability  CreditAmount  DurationOfCreditInMonths
0              1          1049                        18
1              1          2799                         9
2              1           841                        12
3              1          2122                        12
4              1          2171                        12
[[  6.    148.     72.    ...   0.627  50.      1.   ]
 [  1.     85.     66.    ...   0.351  31.      0.   ]
 [  8.    183.     64.    ...   0.672  32.      1.   ]
 ...
 [  5.    121.     72.    ...   0.245  30.      0.   ]
 [  1.    126.     60.    ...   0.349  47.      1.   ]
 [  1.     93.     70.    ...   0.315  23.      0.   ]]
```