LAB 2
HRITIK AJAY BANSAL
CSE A, 15
180905105

Practice Questions 1:

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

```
sum=  25


min=  -2


max=  10
```

2)

```
import pandas as pd
data=[['Hritik',21],['Vijay',39],['Raman',56]]
df=pd.DataFrame(data,columns=['name','age'])
print(df)
```

```
In [2]: import pandas as pd
        data=[['Hritik',21],['Vijay',39],['Raman',56]]
        df=pd.DataFrame(data,columns=['name','age'])
        print(df)

     name  age
0  Hritik   21
1   Vijay   39
2   Raman   56
```

3)

```
import pandas as pd
data = {'Name':['Hritik','Vijay','Raman'],'age':[21,39,56]}
df=pd.DataFrame(data,index=['rank1','rank2','rank3'])
print(df)
```

```
In [3]: import pandas as pd
        data = {'Name':['Hritik','Vijay','Raman'],'age':[21,39,56]}
        df=pd.DataFrame(data,index=['rank1','rank2','rank3'])
        print(df)

          Name  age
rank1  Hritik   21
rank2   Vijay   39
rank3   Raman   56
```

4)

```
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('\n')
print(df1.shape)
print('\n')
print(df1.dtypes)
print('\n')
print(df1.head())
print('\n')
print(df1.tail())
print('\n')
print(df1.T)
```

```
           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    datetime64[ns]
B             int32
C          category
dtype: object


           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


                    0                   1                   2  \
A  2013-01-02 00:00:00  2013-01-02 00:00:00  2013-01-02 00:00:00
B                   3                   3                   3
C                Male              Female                Male

                    3
A  2013-01-02 00:00:00
B                   3
C                Female
```

5) import pandas as pd
import numpy as np
data = {'Name':['Hritik','Vijay','Raman'],'age':[21,39,56]}
df=pd.DataFrame(data,index=['rank1','rank2','rank3'])
#print(df)
df1=pd.DataFrame({'A':pd.Timestamp('20130102'),'B':np.array([3]*4,dtype='int32'),
'C':pd.Categorical(['Male','Female','Male','Female'])})
print(df1)
print('\n')
print(df1.head())
print('\n')
print(df1.tail())
print('\n')
print(df1.index)
print('\n')
print(df1.columns)
print('\n')
print(df1.T)

```
           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


           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


RangeIndex(start=0, stop=4, step=1)


Index(['A', 'B', 'C'], dtype='object')


                    0                   1                   2  \
A  2013-01-02 00:00:00  2013-01-02 00:00:00  2013-01-02 00:00:00
B                   3                   3                   3
C                Male              Female                Male

                    3
A  2013-01-02 00:00:00
B                   3
C                Female
```

6) Boolean indexing:

```
In [43]: df[df.A>0]
```

Out[43]:

	A	B	C	D
2013-01-03	0.407945	-0.057386	-0.263448	-0.212229
2013-01-04	1.349356	0.156570	-1.537562	-0.983377
2013-01-05	1.169310	0.401390	0.519456	-1.183804
2013-01-06	0.882671	0.393707	1.797434	-0.746562
2013-01-08	0.318633	-1.070645	-0.671766	0.057286
2013-01-09	0.972629	0.560891	0.086486	-0.947404
2013-01-10	0.109558	-0.248898	0.443874	-0.250383
2013-01-11	0.465274	-0.139257	0.669039	-1.335782
2013-01-15	0.545144	-0.631945	-0.530343	-0.970961
2013-01-19	1.081066	-0.591033	0.036577	-1.617320
2013-01-20	0.517175	0.274999	-0.396060	1.242740
2013-01-21	0.032702	-0.557427	-0.264580	-0.765678
2013-01-22	0.253274	0.898953	-0.849968	-0.842999
2013-01-24	0.064738	-0.108003	1.471862	0.260146
2013-01-26	1.058019	-0.217231	-1.026181	0.458218
2013-01-27	1.592748	0.067281	-1.054839	0.650854
2013-01-29	0.812700	-0.401304	-0.575509	-0.556427
2013-01-30	1.646523	1.115095	-1.033942	-0.533655
2013-01-31	0.212639	1.917912	0.097612	-1.208774
2013-02-02	0.395549	1.251152	-0.148185	1.063857
2013-02-03	0.822031	0.612275	-1.361615	-0.586021
2013-02-05	0.156389	0.676953	-0.677920	1.184788
2013-02-07	0.065426	-0.915180	0.212536	-0.032875
2013-02-09	0.055502	-0.392300	1.376076	-0.974575
2013-02-10	0.816739	0.785343	-0.168560	-0.277842

```
list=['Male']
for i in range(99):
   if(i%2==1):
      list.append('female')
   else:
      list.append('male')
df['F']=list
```

Output:

```
In [53]: df['F']=list
```

```
In [54]: df
```

Out[54]:

	A	B	C	D	F
2013-01-01	-0.558857	-1.726804	0.542055	5	Male
2013-01-02	-1.286936	0.303464	-1.519532	5	male
2013-01-03	0.407945	-0.057386	-0.263448	5	female
2013-01-04	1.349356	0.156570	-1.537562	5	male
2013-01-05	1.169310	0.401390	0.519456	5	female
...	...	...	...	...	...
2013-04-06	-0.088096	0.922608	-0.414948	5	male
2013-04-07	-0.638060	0.233246	-2.526144	5	female
2013-04-08	0.224488	-0.974740	-1.203296	5	male
2013-04-09	-0.402442	-1.213243	-0.223837	5	female
2013-04-10	-0.384780	1.116736	-0.311550	5	male

100 rows × 5 columns

Setting by assigning with a numpy array:
df.loc[:,'D']=np.array([5]*len(df))

```
In [49]: df.loc[:,'D']=np.array([5]*len(df))
         df.head()
```

Out[49]:

	A	B	C	D
2013-01-01	-0.558857	-1.726804	0.542055	5
2013-01-02	-1.286936	0.303464	-1.519532	5
2013-01-03	0.407945	-0.057386	-0.263448	5
2013-01-04	1.349356	0.156570	-1.537562	5
2013-01-05	1.169310	0.401390	0.519456	5

Deleting a row or column:
df.drop ('A', axis =1, inplace=True)

```
In [57]: df.drop ('F', axis =1, inplace=True)
         df
```

Out[57]:

	A	B	C	D
2013-01-01	-0.558857	-1.726804	0.542055	5
2013-01-02	-1.286936	0.303464	-1.519532	5
2013-01-03	0.407945	-0.057386	-0.263448	5
2013-01-04	1.349356	0.156570	-1.537562	5
2013-01-05	1.169310	0.401390	0.519456	5
...	...	...	...	...
2013-04-06	-0.088096	0.922608	-0.414948	5
2013-04-07	-0.638060	0.233246	-2.526144	5
2013-04-08	0.224488	-0.974740	-1.203296	5
2013-04-09	-0.402442	-1.213243	-0.223837	5
2013-04-10	-0.384780	1.116736	-0.311550	5

100 rows × 4 columns

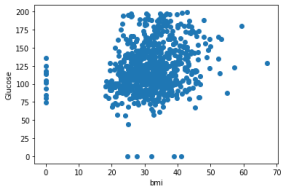
7) import pandas as pd
data=[['Hritik',10],['Vijay',12],['Raman',13]]
df1=pd.DataFrame(data,columns=['Name','Age'])
data2 = [['vinay',10],['Ravi',12],['raj',13],['donald',22],['ram',45]]
df2=pd.DataFrame(data2,columns=['Name','Age'])
print(df1.shape)
print(df2.shape)
Df_new= pd.concat ((df1, df2), axis=1)
print(Df_new.shape)
Df_new2=pd.concat((df1,df2),axis=0)
print(Df_new2.shape)
print(df1.sort_values(by='Age'))

```
In [7]: import pandas as pd
data=[['Hritik',10],['Vijay',12],['Raman',13]]
df1=pd.DataFrame(data,columns=['Name','Age'])
data2 = [['vinay',10],['Ravi',12],['raj',13],['donald',22],['ram',45]]
df2=pd.DataFrame(data2,columns=['Name','Age'])
print(df1.shape)
print(df2.shape)
Df_new= pd.concat ((df1, df2), axis=1)
print(Df_new.shape)
Df_new2=pd.concat((df1,df2),axis=0)
print(Df_new2.shape)
print(df1.sort_values(by='Age'))

(3, 2)
(5, 2)
(5, 4)
(8, 2)
     Name  Age
0  Hritik   10
1   Vijay   12
2   Raman   13
```

Practice Questions 2:
1) #numpy practice
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

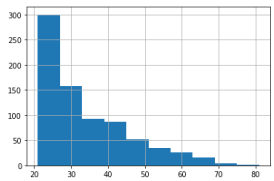
```
#Reading a CSV file
df=pd.read_csv('prima_indians_diabetes.csv',header=None)
print(df.head())
print('\n')
print(df.tail())
df.columns=['preg','glu','bp','sft','ins','bmi','dpf','age','class']
plt.scatter(df['bmi'],df['glu'])
plt.xlabel('bmi')
plt.ylabel('Glucose')
```

```
   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

Out[1]: Text(0, 0.5, 'Glucose')
```



2) df['age'].hist()

```
In [3]: df['age'].hist()
Out[3]: <AxesSubplot:>
```



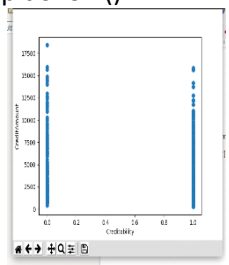
3) import pandas as pd
import numpy as np
import xlrd
import matplotlib.pyplot as plt

df=pd.read_excel('\\wsl$\Ubuntu-18.04\home\hritik\LabsSem6\DS LAB\WEEK2\German Credit.xlsx')

plt.scatter(df['Creditability'],df['CreditAmount'])

plt.xlabel('Creditability')
plt.ylabel('CreditAmount')

plt.show()



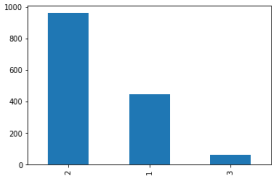
4)import pandas as pd

H = pd.read_table('HR.txt')
H.head()
f=H['Department'].value_counts()
f

```
In [4]: import pandas as pd

        H = pd.read_table('HR.txt')
        H.head()
        f=H['Department'].value_counts()
        f

Out[4]: 2    961
        1    446
        3     63
        Name: Department, dtype: int64
```

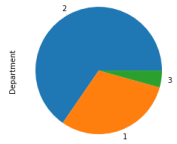
5) f.plot(kind='bar')

```
In [6]: f.plot(kind='bar')
Out[6]: <AxesSubplot:>
```



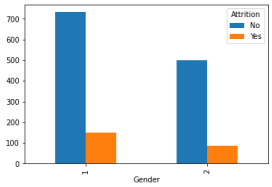
6) f.plot(kind='pie')

```
In [7]: f.plot(kind='pie')
Out[7]: <AxesSubplot:ylabel='Department'>
```



7) fa=pd.crosstab(H['Gender'],H['Attrition'])
fa.plot(kind='bar')

```
In [8]: fa=pd.crosstab(H['Gender'],H['Attrition'])
        fa.plot(kind='bar')

Out[8]: <AxesSubplot:xlabel='Gender'>
```



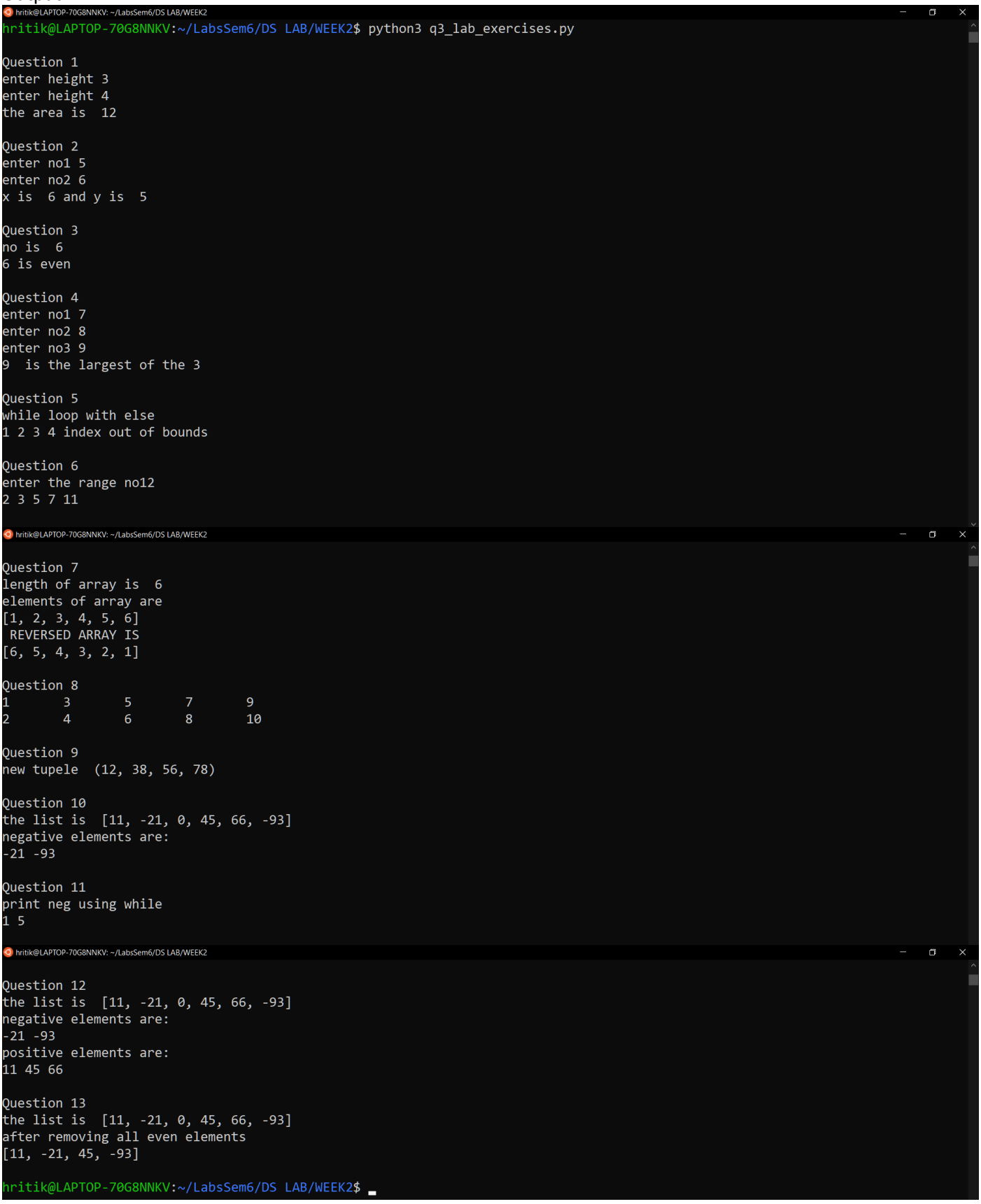
LAB EXERCISES:
Q1-12)

```
print("\nQuestion 1")
x=int(input("enter height "))
y=int(input("enter height "))
print("the area is ",x*y)

print("\nQuestion 2")
x=int(input("enter no1 "))
y=int(input("enter no2 "))
x=x+y
y=x-y
```

```
x=x-y
print("x is ",x,"and","y is ",y)

print("\nQuestion 3")
print("no is ",x)
if x%2==0:
	print(x,"is even")
else:
	print(x," is odd")

print("\nQuestion 4")
a=int(input("enter no1 "))
b=int(input("enter no2 "))
c=int(input("enter no3 "))
a=a if a>b and a>c else b if b>a and b>c else c
print(a," is the largest of the 3")

print("\nQuestion 5")
print("while loop with else")
a=[1,2,3,4]
i=0
while(i<len(a)) :
	print(a[i],end=" ")
	i=i+1
else:
	print("index out of bounds")

print("\nQuestion 6")
x=int(input("enter the range no"))
j=2
while (j<x):
	z=2
	count=0
	while(z<j):
		if j%z==0:
			count=1
		z=z+1
	j=j+1

	if count==0:
		print(z,end=" ")
print()

print("\nQuestion 7")
a=[1,2,3,4,5,6]
print("length of array is ",len(a))
print("elements of array are ")
print(a)

a.reverse()
print(" REVERSED ARRAY IS ")
print(a)

print("\nQuestion 8")
a= tuple((1,3,5,7,9,2,4,6,8,10))
for i in  a[0:int(len(a)/2)]:
		print(i,end="     ")
```

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

print("\nQuestion 9")
a=[]
b=tuple((12,7,38,56,78))
for i in b:
	if i%2==0:
		a.append(i)
a=tuple(a)
print("new tupele ", a)

print("\nQuestion 10")
a=[11,-21,0,45,66,-93]
print("the list is ",a)
print("negative elements are: ")
for i in a:
	if i<0:
		print(i,end=" ")
print()

print("\nQuestion 11")
a=[11,-21,0,45,66,-93]
print("print neg using while")
i=0
while(i<len(a)):
	if a[i]<0:
		print(i,end=" ")
	i=i+1
print()

print("\nQuestion 12")
a=[11,-21,0,45,66,-93]
print("the list is ",a)
print("negative elements are: ")
for i in a:
	if i<0:
		print(i,end=" ")
print()
print("positive elements are: ")
for i in a:
	if i>0:
		print(i,end=" ")
print()

print("\nQuestion 13")
a=[11,-21,0,45,66,-93]
print("the list is ",a)
print("after removing all even elements")
b=a
for i in b:
	if i%2==0:
		a.remove(i)
print(a)
print()
```

Output:



Q14) import pandas as pd
data = {'Name': ['Hritik', 'Harsheet', 'Ravi', 'Akash'] , 'Height': [177, 158, 166, 167],
'Qualification': ['PhD', 'MS', 'MBA', 'BCom.']}
df = pd.DataFrame.from_dict(data)
addr_list = ['Mumbai', 'Delhi', 'Chennai', 'Pune']
df['Address'] = addr_list

print(df.head())

Output:

```
hritik@LAPTOP-70G8NNKV:~/LabsSem6/DS LAB/WEEK2$ python3 q13.py
       Name  Height Qualification  Address
0    Hritik     177           PhD   Mumbai
1  Harsheet     158            MS    Delhi
2      Ravi     166           MBA  Chennai
3     Akash     167         BCom.     Pune
```

Q15) import pandas as pd
data = {'Name': ['Hritik', 'Harsheet', 'Ravi', 'Akash'] , 'Height': [177, 158, 166, 167],
'Qualification': ['PhD', 'MS', 'MBA', 'BCom.']}
df = pd.DataFrame.from_dict(data)
cols = [21, 14, 60, 30]
df.insert(3, 'Age', cols)
print(df.head())

output:

```
hritik@LAPTOP-70G8NNKV:~/LabsSem6/DS LAB/WEEK2$ python3 q14.py
       Name  Height Qualification  Age
0    Hritik     177           PhD   21
1  Harsheet     158            MS   14
2      Ravi     166           MBA   60
3     Akash     167         BCom.   30
```