

Week 2 - PYTHON BASIC PRACTICE – II

NAYAKANTI SAI MIHIRNATH – 210905368

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

O/p :-
25 -2 10

```
data = [['Dinesh',10],['Nithya',12],['Raji',13]]
df = pd.DataFrame(data,columns=['Name','Age'])
print(df)
```

O/p :-

	Name	Age
0	Dinesh	10
1	Nithya	12
2	Raji	13

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

O/p :-

	Name	Age
rank1	Kavitha	28
rank2	Sudha	34
rank3	Raju	29
rank4	Vignesh	42

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

O/p :-

	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
```

```
print(df1.shape)
```

```
print(df1.dtypes)
```

```
print(df1.head())
```

```
print(df1.tail())
```

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

O/p :-

(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

	A	B
count	4	4.0
mean	2013-01-02 00:00:00	3.0
min	2013-01-02 00:00:00	3.0
25%	2013-01-02 00:00:00	3.0
50%	2013-01-02 00:00:00	3.0
75%	2013-01-02 00:00:00	3.0
max	2013-01-02 00:00:00	3.0
std	NaN	0.0

```

print(df.head())
print(df.tail())
print(df.index)
print(df.columns)
print(df.T)
print(df.sort_index(axis=1,ascending=False))
print(df.sort_values(by='Name'))
print(df[0:3])
print(df.iloc[0:2])

```

O/p :-

	Name	Age
rank1	Kavitha	28
rank2	Sudha	34
rank3	Raju	29
rank4	Vignesh	42

	Name	Age
rank1	Kavitha	28
rank2	Sudha	34
rank3	Raju	29
rank4	Vignesh	42

	Name	Age
rank1	Kavitha	28
rank2	Sudha	34
rank3	Raju	29
rank4	Vignesh	42

Index(['rank1', 'rank2', 'rank3', 'rank4'], dtype='object')

Index(['Name', 'Age'], dtype='object')

	rank1	rank2	rank3	rank4
Name	Kavitha	Sudha	Raju	Vignesh
Age	28	34	29	42

	Name	Age
rank1	Kavitha	28
rank2	Sudha	34
rank3	Raju	29

rank4 Vignesh 42

	Name	Age
rank1	Kavitha	28
rank3	Raju	29
rank2	Sudha	34
rank4	Vignesh	42

	Name	Age
rank1	Kavitha	28
rank2	Sudha	34
rank3	Raju	29

	Name	Age
rank1	Kavitha	28
rank2	Sudha	34

```
print(df[df['Age']>30])  
df['Gender']=['Male','Female','Female','Male']  
print(df)
```

O/p :-

	Name	Age
rank2	Sudha	34
rank4	Vignesh	42

	Name	Age	Gender
rank1	Kavitha	28	Male
rank2	Sudha	34	Female
rank3	Raju	29	Female
rank4	Vignesh	42	Male

```
df.drop('Gender',axis=1,inplace=True)
print(df)
df.drop('rank1',axis=0,inplace=True)
print(df)
```

O/p :-

	Name	Age
rank1	Kavitha	28
rank2	Sudha	34
rank3	Raju	29
rank4	Vignesh	42

	Name	Age
rank2	Sudha	34
rank3	Raju	29
rank4	Vignesh	42

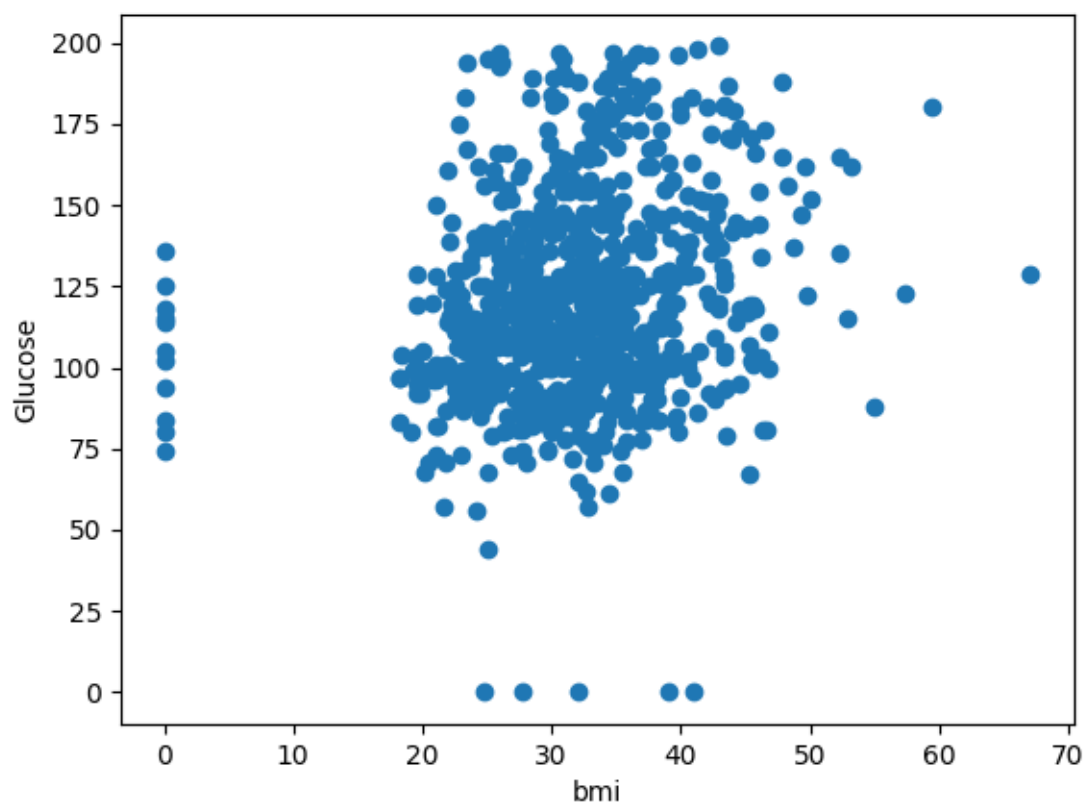
```
df = pd.read_csv('Lab2-files/xyz.csv',header=None)
df.columns=['preg','glu','bp','sft','ins','bmi','dpf','age','class']
```

O/p :-

	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

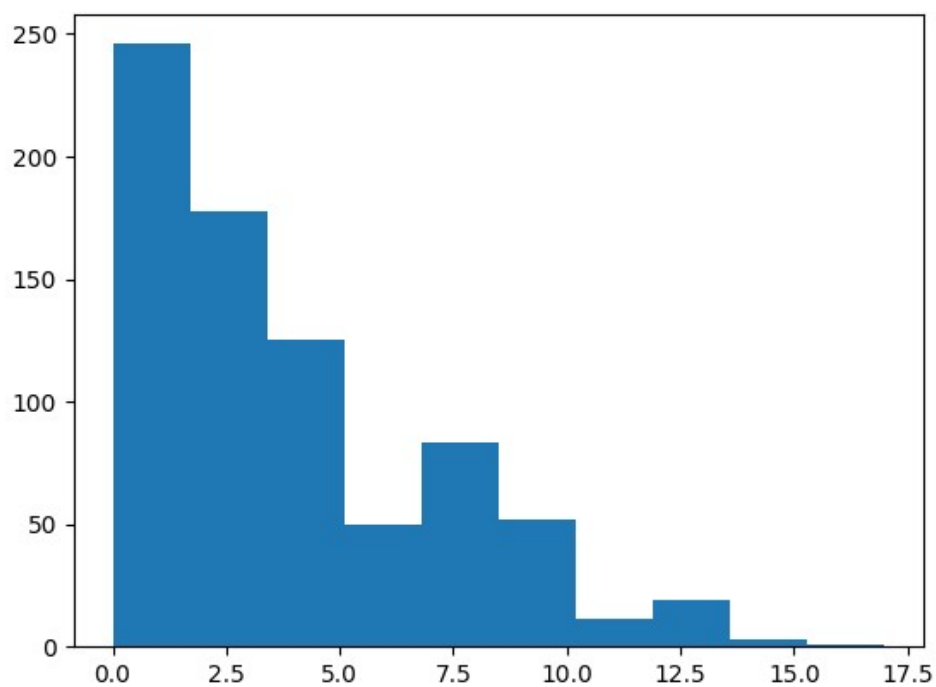
	preg	glu	bp	sft	ins	bmi	dpf	age	class
763	10	101	76	48	180	32.9	0.171	63	0
764	2	122	70	27	0	36.8	0.340	27	0
765	5	121	72	23	112	26.2	0.245	30	0
766	1	126	60	0	0	30.1	0.349	47	1
767	1	93	70	31	0	30.4	0.315	23	0

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

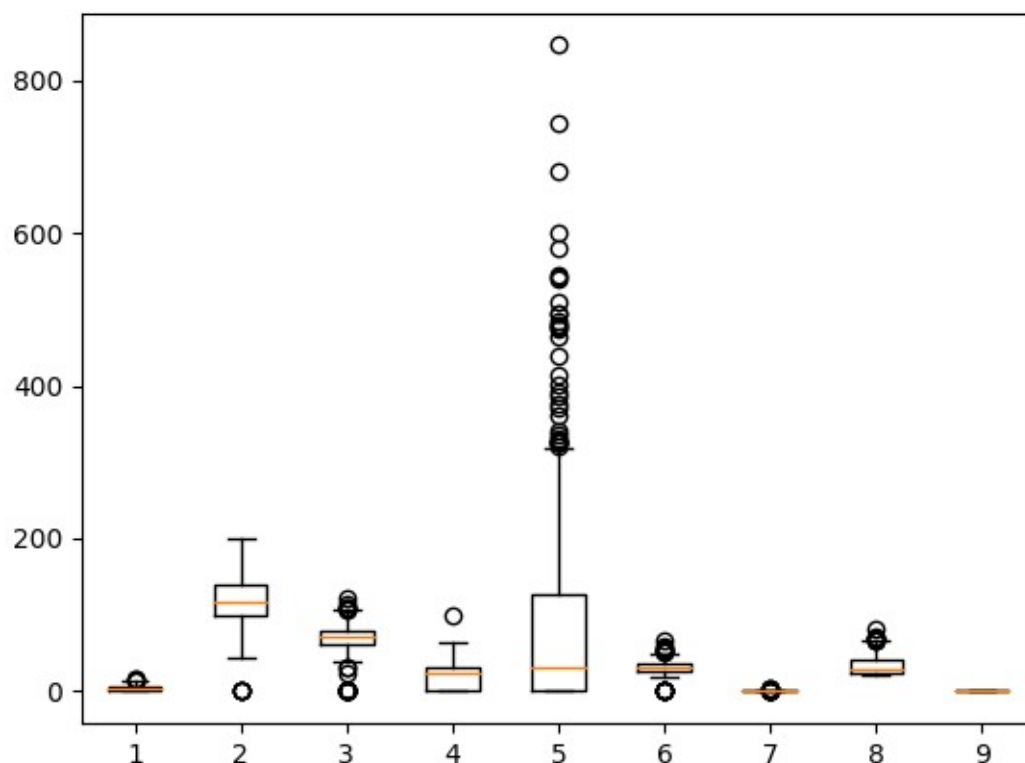


```
plt.hist(df['preg'])
```

```
plt.show()
```



```
plt.boxplot(df)
plt.show()
```



```
W = pd.read_csv('Lab2-files/wine_for_Week2.xls',header=None)
print(W.head())
```

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	1	14.23	1.71	2.43	15.6	127	2.80	3.06	0.28	2.29	5.64	1.04	3.92	1065
1	1	13.20	1.78	2.14	11.2	100	2.65	2.76	0.26	1.28	4.38	1.05	3.40	1050
2	1	13.16	2.36	2.67	18.6	101	2.80	3.24	0.30	2.81	5.68	1.03	3.17	1185
3	1	14.37	1.95	2.50	16.8	113	3.85	3.49	0.24	2.18	7.80	0.86	3.45	1480
4	1	13.24	2.59	2.87	21.0	118	2.80	2.69	0.39	1.82	4.32	1.04	2.93	735

```
G=pd.read_excel('Lab2-files/
German Credit for Week2.xlsx',sheet_name='Sheet1')
print(G.head())
```

	Creditability	CreditAmount	DurationOfCreditInMonths
0	1	1049	18
1	1	2799	9
2	1	841	12
3	1	2122	12
4	1	2171	12

```
D = np.loadtxt('Lab2-files/xyz.txt',delimiter=',')
print(D[:5,:])
```

```
[[1000.  200.  500.]
 [ 500.   45.   56.]
 [2000. 2200. 4500.]
 [5070.  465.  556.]
 [8000. 8200. 5800.]]
```

```
B = pd.read_html('Lab2-files/Test runs-1.html')
print(B)
```

Too Big Output

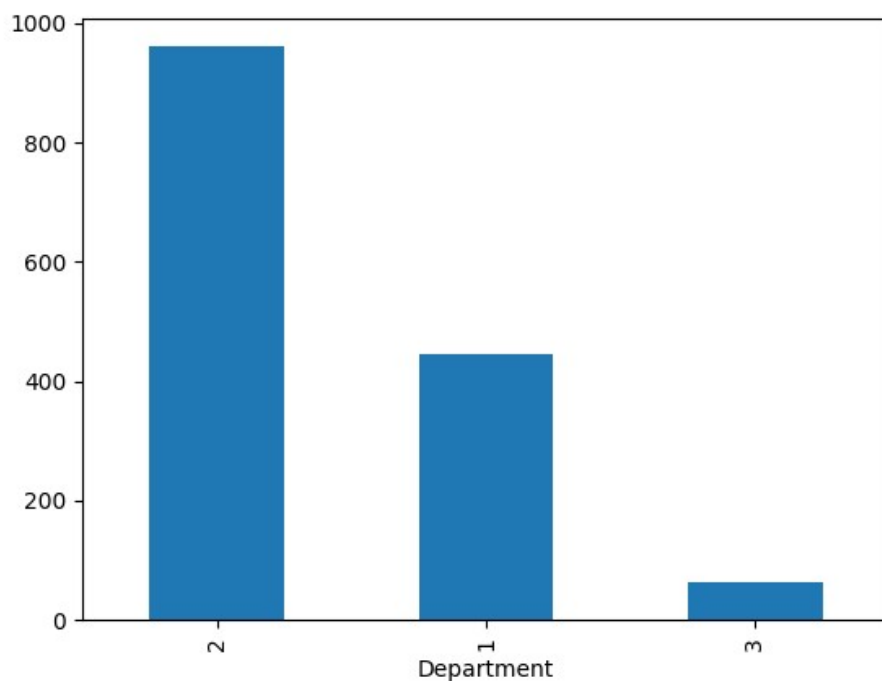
```
H = pd.read_table('Lab2-files/HR.txt')
print(H.head())
```

	Individual	Attrition	...	YearsSinceLastPromotion	YearsWithCurrManager
0	Ind1	Yes	...	0	5
1	Ind2	No	...	1	7
2	Ind3	Yes	...	0	0
3	Ind4	No	...	3	0
4	Ind5	No	...	2	2

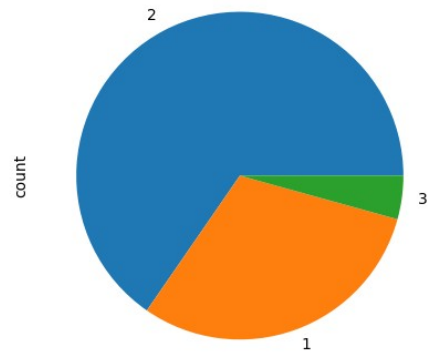
```
f = H['Department'].value_counts()
print(f)
```

```
Department
2    961
1    446
3     63
```

```
f.plot(kind='bar')
plt.show()
```




```
f.plot(kind='pie')  
plt.show()
```



```
pip3 install lxml html5lib openpyxl
```