

In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:

```
df1=pd.read_csv(r'C:\Users\user\Downloads\17_student_marks.csv')  
df1
```

Out[2]:

	Student_ID	Test_1	Test_2	Test_3	Test_4	Test_5	Test_6	Test_7	Test_8	Test_9	Test_10
0	22000	78	87	91	91	88	98	94	100	100	100
1	22001	79	71	81	72	73	68	59	69	59	60
2	22002	66	65	70	74	78	86	87	96	88	80
3	22003	60	58	54	61	54	57	64	62	72	60
4	22004	99	95	96	93	97	89	92	98	91	90
5	22005	41	36	35	28	35	36	27	26	19	20
6	22006	47	50	47	57	62	64	71	75	85	80
7	22007	84	74	70	68	58	59	56	56	64	70
8	22008	74	64	58	57	53	51	47	45	42	40
9	22009	87	81	73	74	71	63	53	45	39	40
10	22010	40	34	37	33	31	35	39	38	40	40
11	22011	91	84	78	74	76	80	80	73	75	70
12	22012	81	83	93	88	89	90	99	99	95	80
13	22013	52	50	42	38	33	30	28	22	12	20
14	22014	63	67	65	74	80	86	95	96	92	80
15	22015	76	82	88	94	85	76	70	60	50	50
16	22016	83	78	71	71	77	72	66	75	66	60
17	22017	55	45	43	38	43	35	44	37	45	30
18	22018	71	67	76	74	64	61	57	64	61	50
19	22019	62	61	53	49	54	59	68	74	65	50
20	22020	44	38	36	34	26	34	39	44	36	40
21	22021	50	56	53	46	41	38	47	39	44	30
22	22022	57	48	40	45	43	36	26	19	9	10
23	22023	59	56	52	44	50	40	45	46	54	50
24	22024	84	92	89	80	90	80	84	74	68	70
25	22025	74	80	86	87	90	100	95	87	85	70
26	22026	92	84	74	83	93	83	75	82	81	70
27	22027	63	70	74	65	64	55	61	58	48	40
28	22028	78	77	69	76	78	74	67	69	78	60
29	22029	55	58	59	67	71	62	53	61	67	70
30	22030	54	54	48	38	35	45	46	47	41	30
31	22031	84	93	97	89	86	95	100	100	100	90
32	22032	95	100	94	100	98	99	100	90	80	80
33	22033	64	61	63	73	63	68	64	58	50	50
34	22034	76	79	73	77	83	86	95	89	90	90
35	22035	78	71	61	55	54	48	41	32	41	40
36	22036	95	89	91	84	89	94	85	91	100	100

	Student_ID	Test_1	Test_2	Test_3	Test_4	Test_5	Test_6	Test_7	Test_8	Test_9	Test_10
37	22037	99	89	79	87	87	81	82	74	64	74
38	22038	82	83	85	86	89	80	88	95	87	81
39	22039	65	56	64	62	58	51	61	68	70	70
40	22040	100	93	92	86	84	76	82	74	79	77
41	22041	78	72	73	79	81	73	71	77	83	81
42	22042	98	100	100	93	94	92	100	100	98	98
43	22043	58	62	67	77	71	63	64	73	83	77
44	22044	96	92	94	100	99	95	98	92	84	81
45	22045	86	87	85	84	85	91	86	82	85	81
46	22046	48	55	46	40	34	29	37	34	39	44
47	22047	56	52	54	47	40	35	43	44	40	51
48	22048	42	44	46	53	62	59	57	53	43	51
49	22049	64	54	49	59	54	55	57	59	63	70
50	22050	50	44	37	29	37	46	53	57	55	60
51	22051	70	60	70	62	67	67	68	67	72	60
52	22052	63	73	70	63	60	67	61	59	52	51
53	22053	92	100	100	100	100	100	92	87	94	100
54	22054	64	55	54	61	63	57	47	37	44	44
55	22055	60	66	68	58	49	47	39	29	39	44

In [3]:

```
df=df1.head(100)  
df
```

Out[3]:

	Student_ID	Test_1	Test_2	Test_3	Test_4	Test_5	Test_6	Test_7	Test_8	Test_9	Test_10
0	22000	78	87	91	91	88	98	94	100	100	100
1	22001	79	71	81	72	73	68	59	69	59	60
2	22002	66	65	70	74	78	86	87	96	88	80
3	22003	60	58	54	61	54	57	64	62	72	60
4	22004	99	95	96	93	97	89	92	98	91	90
5	22005	41	36	35	28	35	36	27	26	19	20
6	22006	47	50	47	57	62	64	71	75	85	80
7	22007	84	74	70	68	58	59	56	56	64	70
8	22008	74	64	58	57	53	51	47	45	42	40
9	22009	87	81	73	74	71	63	53	45	39	40
10	22010	40	34	37	33	31	35	39	38	40	40
11	22011	91	84	78	74	76	80	80	73	75	70
12	22012	81	83	93	88	89	90	99	99	95	80
13	22013	52	50	42	38	33	30	28	22	12	20
14	22014	63	67	65	74	80	86	95	96	92	80
15	22015	76	82	88	94	85	76	70	60	50	50
16	22016	83	78	71	71	77	72	66	75	66	60
17	22017	55	45	43	38	43	35	44	37	45	30
18	22018	71	67	76	74	64	61	57	64	61	50
19	22019	62	61	53	49	54	59	68	74	65	50
20	22020	44	38	36	34	26	34	39	44	36	40
21	22021	50	56	53	46	41	38	47	39	44	30
22	22022	57	48	40	45	43	36	26	19	9	10
23	22023	59	56	52	44	50	40	45	46	54	50
24	22024	84	92	89	80	90	80	84	74	68	70
25	22025	74	80	86	87	90	100	95	87	85	70
26	22026	92	84	74	83	93	83	75	82	81	70
27	22027	63	70	74	65	64	55	61	58	48	40
28	22028	78	77	69	76	78	74	67	69	78	60
29	22029	55	58	59	67	71	62	53	61	67	70
30	22030	54	54	48	38	35	45	46	47	41	30
31	22031	84	93	97	89	86	95	100	100	100	90
32	22032	95	100	94	100	98	99	100	90	80	80
33	22033	64	61	63	73	63	68	64	58	50	50
34	22034	76	79	73	77	83	86	95	89	90	90
35	22035	78	71	61	55	54	48	41	32	41	40
36	22036	95	89	91	84	89	94	85	91	100	100

	Student_ID	Test_1	Test_2	Test_3	Test_4	Test_5	Test_6	Test_7	Test_8	Test_9	Test_10
37	22037	99	89	79	87	87	81	82	74	64	81
38	22038	82	83	85	86	89	80	88	95	87	91
39	22039	65	56	64	62	58	51	61	68	70	75
40	22040	100	93	92	86	84	76	82	74	79	77
41	22041	78	72	73	79	81	73	71	77	83	91
42	22042	98	100	100	93	94	92	100	100	98	91
43	22043	58	62	67	77	71	63	64	73	83	77
44	22044	96	92	94	100	99	95	98	92	84	81
45	22045	86	87	85	84	85	91	86	82	85	81
46	22046	48	55	46	40	34	29	37	34	39	41
47	22047	56	52	54	47	40	35	43	44	40	41
48	22048	42	44	46	53	62	59	57	53	43	41
49	22049	64	54	49	59	54	55	57	59	63	71
50	22050	50	44	37	29	37	46	53	57	55	61
51	22051	70	60	70	62	67	67	68	67	72	61
52	22052	63	73	70	63	60	67	61	59	52	51
53	22053	92	100	100	100	100	100	92	87	94	100
54	22054	64	55	54	61	63	57	47	37	44	41
55	22055	60	66	68	58	49	47	39	29	39	41

In [4]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 56 entries, 0 to 55
Data columns (total 13 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Student_ID  56 non-null    int64
1   Test_1      56 non-null    int64
2   Test_2      56 non-null    int64
3   Test_3      56 non-null    int64
4   Test_4      56 non-null    int64
5   Test_5      56 non-null    int64
6   Test_6      56 non-null    int64
7   Test_7      56 non-null    int64
8   Test_8      56 non-null    int64
9   Test_9      56 non-null    int64
10  Test_10     56 non-null    int64
11  Test_11     56 non-null    int64
12  Test_12     56 non-null    int64
dtypes: int64(13)
memory usage: 5.8 KB
```

In [5]:

```
df.describe()
```

Out[5]:

	Student_ID	Test_1	Test_2	Test_3	Test_4	Test_5	Test_6
count	56.000000	56.000000	56.000000	56.000000	56.000000	56.000000	56.000000
mean	22027.500000	70.750000	69.196429	68.089286	67.446429	67.303571	66.000000
std	16.309506	17.009356	17.712266	18.838333	19.807179	20.746890	21.054043
min	22000.000000	40.000000	34.000000	35.000000	28.000000	26.000000	29.000000
25%	22013.750000	57.750000	55.750000	53.000000	54.500000	53.750000	50.250000
50%	22027.500000	70.500000	68.500000	70.000000	71.500000	69.000000	65.500000
75%	22041.250000	84.000000	83.250000	85.000000	84.000000	85.250000	83.750000
max	22055.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000

In [6]:

```
df.columns
```

Out[6]:

```
Index(['Student_ID', 'Test_1', 'Test_2', 'Test_3', 'Test_4', 'Test_5',  
      'Test_6', 'Test_7', 'Test_8', 'Test_9', 'Test_10', 'Test_11',  
      'Test_12'],  
      dtype='object')
```

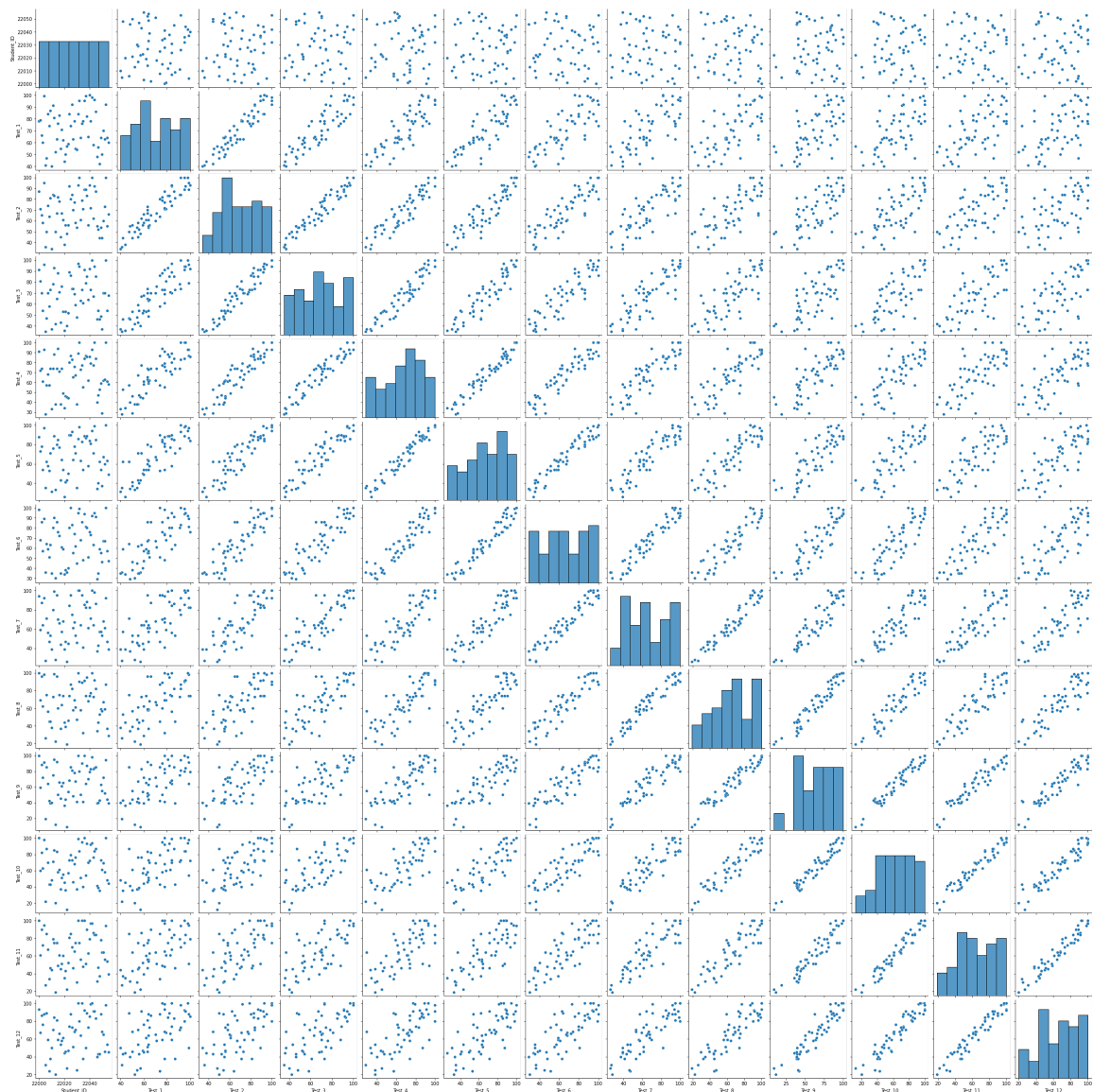


In [7]:

```
sns.pairplot(df)
```

Out[7]:

&lt;seaborn.axisgrid.PairGrid at 0x2f5838b09a0&gt;



In [8]:

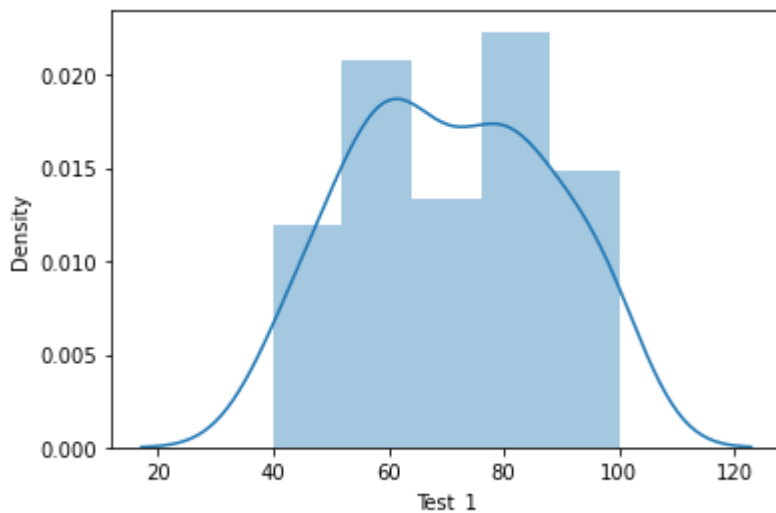
```
sns.distplot(df['Test_1'])
```

C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

```
warnings.warn(msg, FutureWarning)
```

Out[8]:

```
<AxesSubplot:xlabel='Test_1', ylabel='Density'>
```

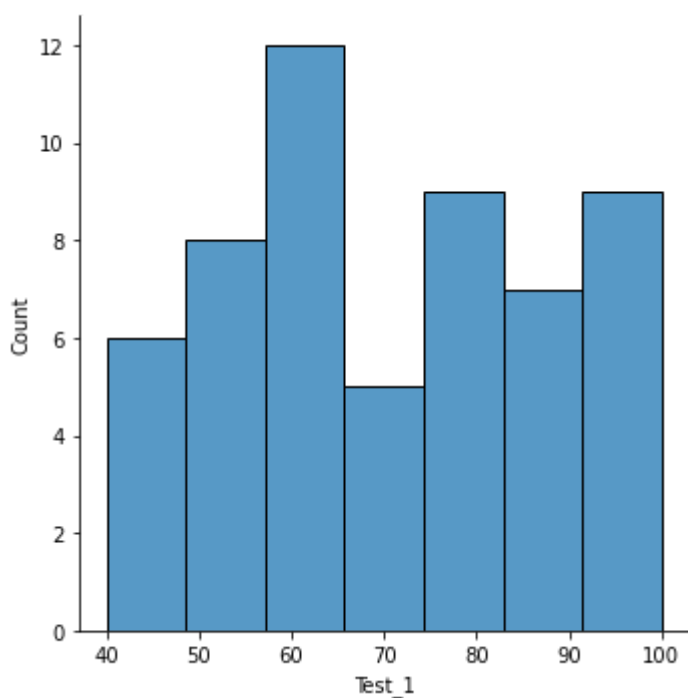


In [9]:

```
sns.displot(df["Test_1"])
```

Out[9]:

```
<seaborn.axisgrid.FacetGrid at 0x2f5ff960460>
```



In [10]:

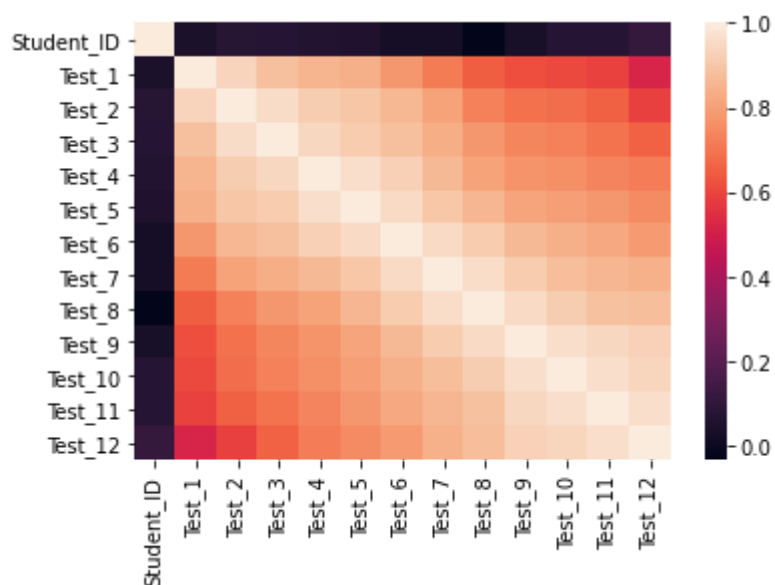
```
df1=df[['Student_ID', 'Test_1', 'Test_2', 'Test_3', 'Test_4', 'Test_5',  
        'Test_6', 'Test_7', 'Test_8', 'Test_9', 'Test_10', 'Test_11',  
        'Test_12']]
```

In [11]:

```
sns.heatmap(df1.corr())
```

Out[11]:

<AxesSubplot:>



In [12]:

```
df2=df.dropna(axis=1)  
df2
```

Out[12]:

	Student_ID	Test_1	Test_2	Test_3	Test_4	Test_5	Test_6	Test_7	Test_8	Test_9	Test_10
0	22000	78	87	91	91	88	98	94	100	100	100
1	22001	79	71	81	72	73	68	59	69	59	60
2	22002	66	65	70	74	78	86	87	96	88	80
3	22003	60	58	54	61	54	57	64	62	72	60
4	22004	99	95	96	93	97	89	92	98	91	90
5	22005	41	36	35	28	35	36	27	26	19	20
6	22006	47	50	47	57	62	64	71	75	85	80
7	22007	84	74	70	68	58	59	56	56	64	70
8	22008	74	64	58	57	53	51	47	45	42	40
9	22009	87	81	73	74	71	63	53	45	39	40
10	22010	40	34	37	33	31	35	39	38	40	40
11	22011	91	84	78	74	76	80	80	73	75	70
12	22012	81	83	93	88	89	90	99	99	95	80
13	22013	52	50	42	38	33	30	28	22	12	20
14	22014	63	67	65	74	80	86	95	96	92	80
15	22015	76	82	88	94	85	76	70	60	50	50
16	22016	83	78	71	71	77	72	66	75	66	60
17	22017	55	45	43	38	43	35	44	37	45	30
18	22018	71	67	76	74	64	61	57	64	61	50
19	22019	62	61	53	49	54	59	68	74	65	50
20	22020	44	38	36	34	26	34	39	44	36	40
21	22021	50	56	53	46	41	38	47	39	44	30
22	22022	57	48	40	45	43	36	26	19	9	10
23	22023	59	56	52	44	50	40	45	46	54	50
24	22024	84	92	89	80	90	80	84	74	68	70
25	22025	74	80	86	87	90	100	95	87	85	70
26	22026	92	84	74	83	93	83	75	82	81	70
27	22027	63	70	74	65	64	55	61	58	48	40
28	22028	78	77	69	76	78	74	67	69	78	60
29	22029	55	58	59	67	71	62	53	61	67	70
30	22030	54	54	48	38	35	45	46	47	41	30
31	22031	84	93	97	89	86	95	100	100	100	90
32	22032	95	100	94	100	98	99	100	90	80	80
33	22033	64	61	63	73	63	68	64	58	50	50
34	22034	76	79	73	77	83	86	95	89	90	90
35	22035	78	71	61	55	54	48	41	32	41	40
36	22036	95	89	91	84	89	94	85	91	100	100

	Student_ID	Test_1	Test_2	Test_3	Test_4	Test_5	Test_6	Test_7	Test_8	Test_9	Test_10	Test_11	Test_12
37	22037	99	89	79	87	87	81	82	74	64	81	81	81
38	22038	82	83	85	86	89	80	88	95	87	81	81	81
39	22039	65	56	64	62	58	51	61	68	70	70	70	70
40	22040	100	93	92	86	84	76	82	74	79	79	79	79
41	22041	78	72	73	79	81	73	71	77	83	83	83	83
42	22042	98	100	100	93	94	92	100	100	98	98	98	98
43	22043	58	62	67	77	71	63	64	73	83	83	83	83
44	22044	96	92	94	100	99	95	98	92	84	84	84	84
45	22045	86	87	85	84	85	91	86	82	85	85	85	85
46	22046	48	55	46	40	34	29	37	34	39	39	39	39
47	22047	56	52	54	47	40	35	43	44	40	40	40	40
48	22048	42	44	46	53	62	59	57	53	43	43	43	43
49	22049	64	54	49	59	54	55	57	59	63	63	63	63
50	22050	50	44	37	29	37	46	53	57	55	55	55	55
51	22051	70	60	70	62	67	67	68	67	72	72	72	72
52	22052	63	73	70	63	60	67	61	59	52	52	52	52
53	22053	92	100	100	100	100	100	92	87	94	94	94	94
54	22054	64	55	54	61	63	57	47	37	44	44	44	44
55	22055	60	66	68	58	49	47	39	29	39	39	39	39

In [13]:

```
x=df2[['Student_ID', 'Test_2', 'Test_3', 'Test_4', 'Test_5',
        'Test_6', 'Test_7', 'Test_8', 'Test_9', 'Test_10', 'Test_11',
        'Test_12']]
y=df2[['Test_1']]
```

In [14]:

```
from sklearn.model_selection import train_test_split
```

In [15]:

```
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.3)
```

In [16]:

```
from sklearn.linear_model import LinearRegression

lr=LinearRegression()
lr.fit(x_train,y_train)#ValueError: Input contains NaN, infinity or a value too large for
```

Out[16]:

```
LinearRegression()
```

In [17]:

```
print(lr.intercept_)
```

[-293.85876146]

In [18]:

```
coef= pd.DataFrame(lr.coef_)  
coef
```

Out[18]:

	0	1	2	3	4	5	6	7	8	
0	0.013996	0.803786	-0.28695	0.35118	0.170395	0.001338	-0.47368	0.251711	0.1184	-0.15

In [19]:

```
print(lr.score(x_test,y_test))
```

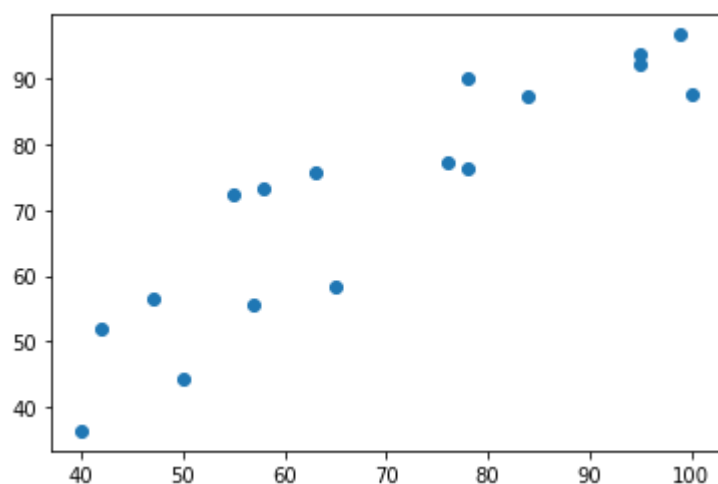
0.8014413261344475

In [20]:

```
prediction = lr.predict(x_test)  
plt.scatter(y_test,prediction)
```

Out[20]:

<matplotlib.collections.PathCollection at 0x2f58d874d60>



In [21]:

```
lr.score(x_test,y_test)
```

Out[21]:

0.8014413261344475

In [22]:

```
lr.score(x_train,y_train)
```

Out[22]:

0.9178851779544543

In [23]:

```
from sklearn.linear_model import Ridge,Lasso
```

In [24]:

```
rr=Ridge(alpha=10)  
rr.fit(x_train,y_train)
```

Out[24]:

Ridge(alpha=10)

In [25]:

```
rr.score(x_test,y_test)
```

Out[25]:

0.8027266438910517

In [26]:

```
la=Lasso(alpha=10)  
la.fit(x_train,y_train)
```

Out[26]:

Lasso(alpha=10)

In [27]:

```
la.score(x_test,y_test)
```

Out[27]:

0.8620248081816229

## ELastic Net

In [28]:

```
from sklearn.linear_model import ElasticNet  
en = ElasticNet()  
en.fit(x_train,y_train)
```

Out[28]:

ElasticNet()



In [29]:

```
print(en.coef_)
```

```
[-0.          0.76631227 -0.23005619  0.28371152  0.19065254  0.
 -0.38590481  0.18147112  0.08167798 -0.06998803  0.3461974  -0.3273816 ]
```

In [30]:

```
print(en.intercept_)
```

```
[14.36825564]
```

In [31]:

```
prediction=en.predict(x_test)
print(prediction)
```

```
[89.91178715 75.35359851 37.80041678 51.77644091 55.00078557 87.64348345
 87.60438006 71.60426465 77.85647859 58.73582642 75.2001484  96.89496285
 57.21782084 44.86429709 94.60074178 91.81636841 71.83888536]
```

In [32]:

```
print(en.score(x_test,y_test))
```

```
0.8136880363118768
```

## Evaluation Metrics

In [33]:

```
from sklearn import metrics
```

In [34]:

```
print("Mean Absolute Error:",metrics.mean_absolute_error(y_test,prediction))
```

```
Mean Absolute Error: 6.845760698493046
```

In [35]:

```
print("Mean Squared Error:",metrics.mean_squared_error(y_test,prediction))
```

```
Mean Squared Error: 72.02859196867342
```

In [36]:

```
print("Root Mean Squared Error:",np.sqrt(metrics.mean_squared_error(y_test,prediction)))
```

```
Root Mean Squared Error: 8.486966004920335
```