```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split as tts
from sklearn.linear_model import LinearRegression as LR
from sklearn.metrics import mean_squared_error
import math
```

```
sal_data = pd.read_csv("/content/Salary_Data - Salary_Data.csv")
sal_data
```

| | YearsExperience | Salary |
|---|---|---|
| **0** | 1.1 | 39343 |
| **1** | 1.3 | 46205 |
| **2** | 1.5 | 37731 |
| **3** | 2.0 | 43525 |
| **4** | 2.2 | 39891 |
| **5** | 2.9 | 56642 |
| **6** | 3.0 | 60150 |
| **7** | 3.2 | 54445 |

```
x = sal_data.drop('Salary',axis = 1)
y = sal_data.Salary
```

```
train_x, test_x, train_y, test_y = tts(x,y,test_size=0.3, random_state=42)
```

| | |
|---|---|
| **11** | 4.0 55794 |

```
train_x.head()
```

| | YearsExperience |
|---|---|
| **0** | 1.1 |
| **4** | 2.2 |
| **16** | 5.1 |
| **5** | 2.9 |
| **13** | 4.1 |

```
train_x, test_x, train_y, test_y = tts(x,y,test_size=0.3, random_state=30)
```

| | |
|---|---|
| **21** | 7.1 98273 |

```
train_x.head()
```

| | YearsExperience |
|---|---|
| **4** | 2.2 |
| **11** | 4.0 |
| **22** | 7.9 |
| **18** | 5.9 |
| **19** | 6.0 |

```
test_x.head()
```

| | YearsExperience |
|---|---|
| **15** | 4.9 |
| **28** | 10.3 |
| **26** | 9.5 |
| **10** | 3.9 |
| **24** | 8.7 |

```
train_y.head()
```

```
4      39891
11     55794
22    101302
18     81363
19     93940
Name: Salary, dtype: int64
```

```
test_y.head()
```

```
15     67938
28    122391
26    116969
10     63218
24    109431
Name: Salary, dtype: int64
```

```
model = LR()
```

```
model.fit(train_x,train_y)
```

```
LinearRegression()
```

```
model.coef_
```

```
array([9846.5759533])
```

```
model.intercept_
```

```
23278.822101930637
```

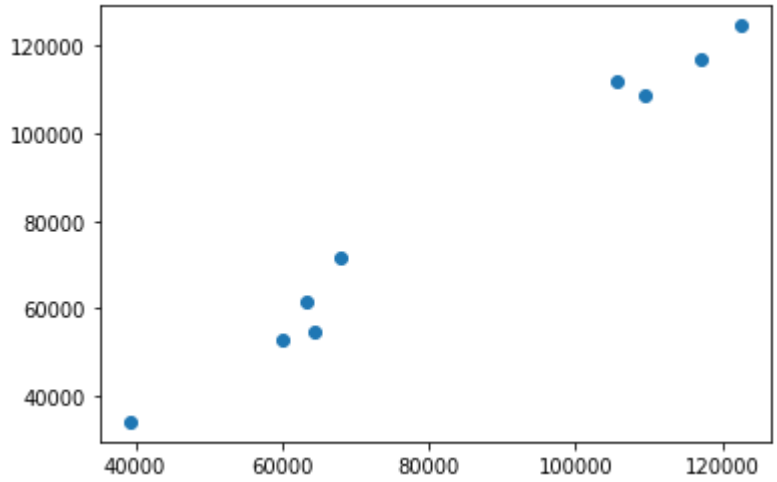
```
train_Y_hat = model.predict(train_x)
test_Y_hat = model.predict(test_x)
```

```
print('Train MSE' , math.sqrt(mean_squared_error(train_y,train_Y_hat)))
print('Test MSE' , math.sqrt(mean_squared_error(test_y,test_Y_hat)))
```

```
Train MSE 5954.982711949841
Test MSE 5111.153940719106
```

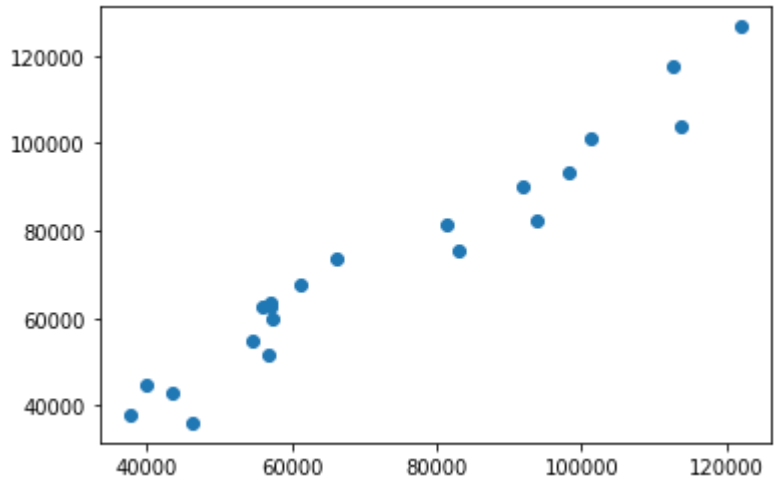
```
plt.scatter(test_y,test_Y_hat)
```

```
<matplotlib.collections.PathCollection at 0x7f7e3cdbe2d0>
```



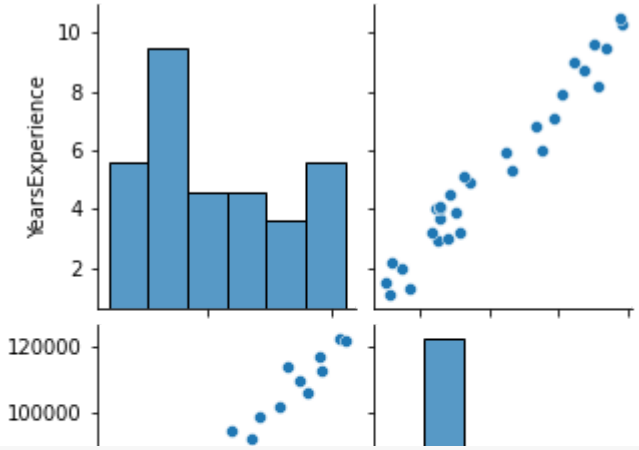

```
plt.scatter(train_y,train_Y_hat)
```

```
<matplotlib.collections.PathCollection at 0x7f7e3c2f6290>
```




```
sns.pairplot(sal_data)
```

```
<seaborn.axisgrid.PairGrid at 0x7f7e3bc6b7d0>
```




```
plt.figure(figsize = (14,10))
sns.heatmap(sal_data.corr(),annot=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7e3bc68450>
```