In [ ]: ▶|

In [1]: ▶|
```python
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
from sklearn.linear_model import LinearRegression
```

In [2]: ▶|
```python
df = pd.read_csv("ml_data_salary.csv")
df.rename(columns={"age":"Age", "distance":"Distance", "YearsExperience":"Yea
df.head()
```

Out[2]:

	Age	Distance	Years of Experience	Salary
0	31.1	77.75	1.1	39343
1	31.3	78.25	1.3	46205
2	31.5	78.75	1.5	37731
3	32.0	80.00	2.0	43525
4	32.2	80.50	2.2	39891

In [3]: ▶|
```python
x = np.reshape(df[["Age", "Distance", "Years of Experience"]], newshape=(30,-
y = df[["Salary"]]
```

In [4]: ▶|
```python
X_train, X_test, y_train, y_test = train_test_split(x,y, train_size=0.8, rand
```

In [5]: ▶|
```python
lm = LinearRegression().fit(X_train, y_train)
```

In [6]: ▶ | `train_pre = lm.predict(X_train)`
         `train_pre`

Out[6]: array([[116032.],
                [63936.],
                [76032.],
                [100288.],
                [53696.],
                [74176.],
                [56384.],
                [68608.],
                [102976.],
                [90048.],
                [38784.],
                [124480.],
                [54592.],
                [47104.],
                [81600.],
                [82560.],
                [61056.],
                [56384.],
                [110528.],
                [45312.],
                [36864.],
                [92736.],
                [72384.],
                [63936.]])

In [22]:
```python
plt.scatter(X_train['Age'], y_train, label='Age')
plt.scatter(X_train['Distance'], y_train, label='Distance')
plt.scatter(X_train["Years of Experience"], y_train, label="Years of Exprienc
plt.legend()
plt.plot(X_train, lm.predict(X_train), color="Red")
plt.ylabel("Salary")
plt.grid(b='True')
plt.title("Train Data Graph")
```

C:\Users\MohHu\AppData\Local\Temp\ipykernel_16768\1206406881.py:7: Matplotl
ibDeprecationWarning: The 'b' parameter of grid() has been renamed 'visibl
e' since Matplotlib 3.5; support for the old name will be dropped two minor
releases later.
    plt.grid(b='True')

Out[22]: Text(0.5, 1.0, 'Train Data Graph')
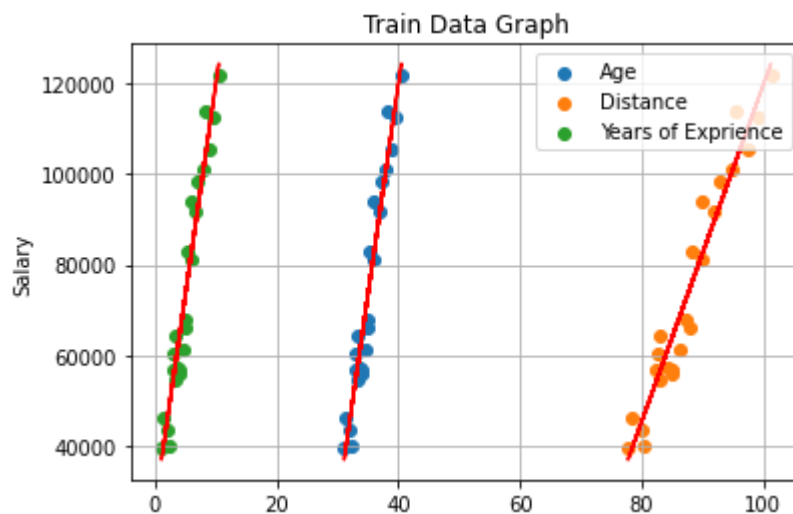
In [25]:
```python
plt.scatter(X_test["Age"], y_test, label="Age")
plt.scatter(X_test["Distance"], y_test, label="Distance")
plt.scatter(X_test["Years of Experience"], y_test, label="Years of Experience
plt.legend( loc="upper right")
plt.plot(X_test, lm.predict(X_test), color="Red")
plt.ylabel("Salary")
plt.title("Test Data Prediction Graph")
plt.grid(b="True")
```

```
C:\Users\MohHu\AppData\Local\Temp\ipykernel_16768\2139387764.py:8: Matplotl
ibDeprecationWarning: The 'b' parameter of grid() has been renamed 'visibl
e' since Matplotlib 3.5; support for the old name will be dropped two minor
releases later.
  plt.grid(b="True")
```
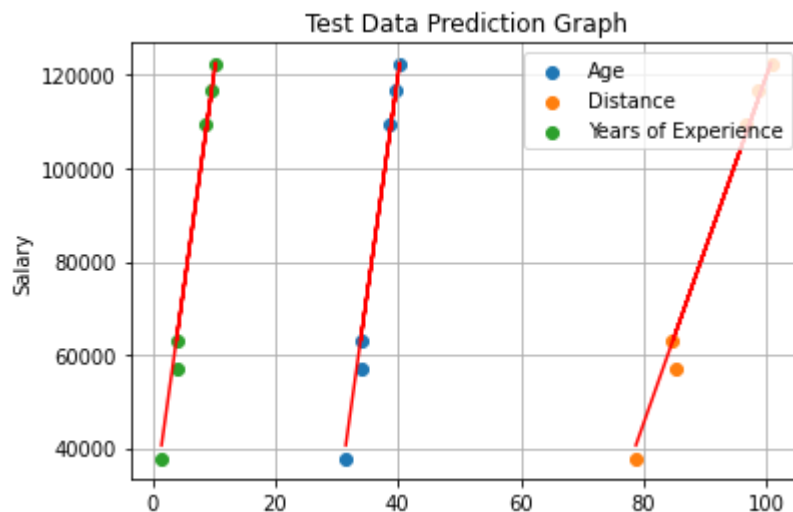

In [12]:
```python
X_train.columns
```

Out[12]: Index(['Age', 'Distance', 'Years of Experience'], dtype='object')

In [13]:
```python
Coff = pd.DataFrame(np.reshape(lm.coef_, (3,-1)), X_train.columns ,columns=["
Coff
```

Out[13]:

	Coffeicnet
Age	-1.243181e+16
Distance	-2.605358e+14
Years of Experience	1.308315e+16

In [14]:
```python
lm.score(X_train, y_train)
```

Out[14]: 0.9409532368371482

In [15]:
```python
lm.score(X_test, y_test)
```

Out[15]: 0.988401541985491

In [16]: ▶| `df.head()`

Out[16]:

	Age	Distance	Years of Experience	Salary
0	31.1	77.75	1.1	39343
1	31.3	78.25	1.3	46205
2	31.5	78.75	1.5	37731
3	32.0	80.00	2.0	43525
4	32.2	80.50	2.2	39891

In [17]: ▶| `lm.predict([[31.1, 77.75, 1.1]])`

```
c:\Users\MohHu\AppData\Local\Programs\Python\Python310\lib\site-packages\sk
learn\base.py:450: UserWarning: X does not have valid feature names, but Li
nearRegression was fitted with feature names
  warnings.warn(
```

Out[17]: `array([[36864.]])`

In [ ]: ▶|

In [ ]: ▶|