In [1]:
```
import pandas as pd
```

In [2]:
```
df = pd.read_csv("my_data.csv")
```

In [3]:
```
df.head()
```

Out[3]:

	Years Experience	Salary
0	1.1	39343
1	1.3	46205
2	1.5	37731
3	2.0	43525
4	2.2	39991

Step2: splitting dataset into training and testing data

In [5]:
```
X = df[['Years Experience']]
y=df['Salary']
```

In [6]:
```
X
```

Out[6]:

	Years Experience
0	1.1
1	1.3
2	1.5
3	2.0
4	2.2
5	2.9
6	3.0
7	3.2
8	3.7
9	3.9
10	4.0
11	4.0
12	4.1
13	4.5
14	4.9
15	5.1
16	5.3
17	5.9
18	6.0

In [7]:
```
y
```

Out[7]:
```
0     39343
1     46205
2     37731
3     43525
4     39991
5     56642
6     60150
7     54445
8     64445
9     57189
10    63218
11    55794
12    56957
13    57081
14    61111
15    67938
16    66029
17    83088
18    93940
Name: Salary, dtype: int64
```

Step2: import Library

In [8]:
```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test =train_test_split(X,y,test_size = 0.2, random
# always write train first then test
```

Step3: Fit- Linear Regression Model

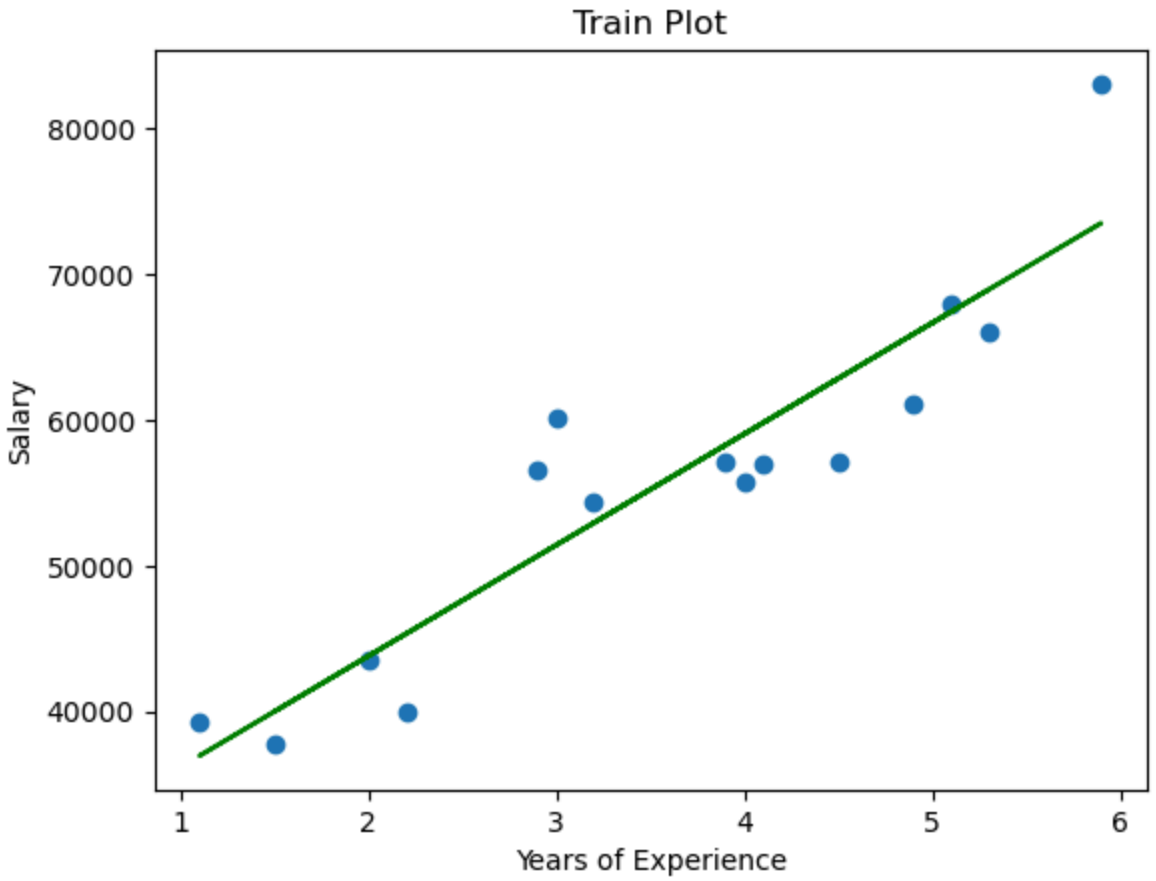
In [9]:
```
from sklearn.linear_model import LinearRegression
model =LinearRegression().fit(X_train, y_train)
model
```

Out[9]:
```
LinearRegression()
```

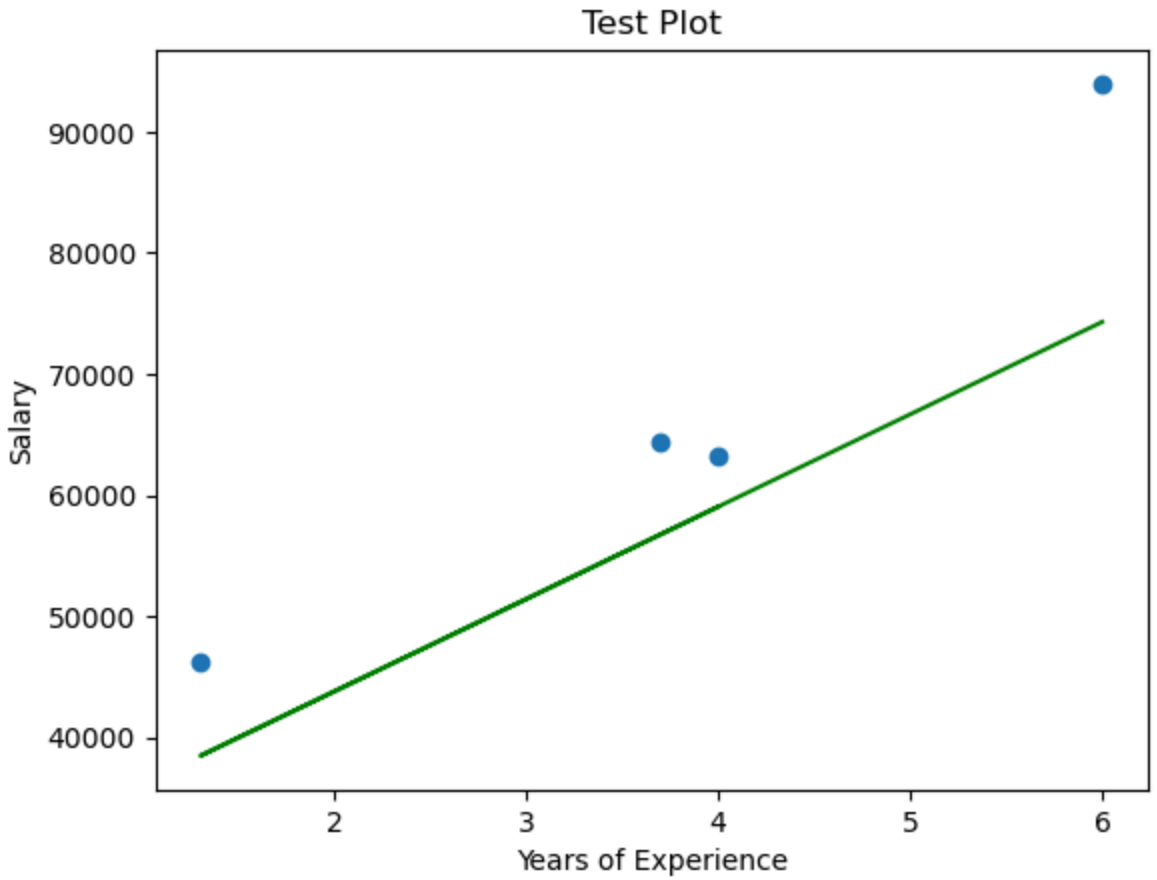
Step 4: Plotting

In [10]:
```
import matplotlib.pyplot as plt
```

In [25]:
```
plt.scatter(X_train, y_train)
plt.plot(X_train, model.predict(X_train), color="green")
plt.xlabel("Years of Experience")
plt.ylabel("Salary")
plt.title("Train Plot")
plt.show()
#plt.scatter(X_test, y_test)
```



In [11]:
```
plt.scatter(X_test, y_test)
plt.plot(X_test, model.predict(X_test), color="green")
plt.xlabel("Years of Experience")
plt.ylabel("Salary")
plt.title("Test Plot")
plt.show()
```



Step5: Testing and Evaluation Model

In [12]:
```
#model fitness
model.score(X_test,y_test)
```

Out[12]: 0.5571840287254162

In [13]:
```
model.score(X_train,y_train)
```

Out[13]: 0.8389395127764611

In [14]:
```
print("score for train data", model.score(X_test,y_test))
print("score for train data",model.score(X_train,y_train))
```

```
score for train data 0.5571840287254162
score for train data 0.8389395127764611
```

Step6: Prediction of Unknown Value

In [15]:
```
#how much salary for 5 years of experience, bcz we find salary on the basis of
model.predict([[5]])
```

```
C:\Users\lenovo\anaconda3\lib\site-packages\sklearn\base.py:450: UserWarning:
X does not have valid feature names, but LinearRegression was fitted with fea
ture names
  warnings.warn(
```

Out[15]:
```
array([66671.58179853])
```

In [17]:
```
model.predict(X_test)
```

Out[17]:
```
array([59051.96838835, 38479.01218084, 56766.08436529, 74291.19520872])
```

In [18]:
```
#How much salary for 3 different experience
model.predict([[5],[10],[12]])
```

```
C:\Users\lenovo\anaconda3\lib\site-packages\sklearn\base.py:450: UserWarning:
X does not have valid feature names, but LinearRegression was fitted with fea
ture names
  warnings.warn(
```

Out[18]:
```
array([ 66671.58179853, 104769.64884946, 120008.87566983])
```

In [20]:
```
#How much salary for 3 different experience--another way
x = ([5],[10],[12])
model.predict(x)
```

```
C:\Users\lenovo\anaconda3\lib\site-packages\sklearn\base.py:450: UserWarning:
X does not have valid feature names, but LinearRegression was fitted with fea
ture names
  warnings.warn(
```

Out[20]:
```
array([ 66671.58179853, 104769.64884946, 120008.87566983])
```

In [ ]:
```

```