In [25]:
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
```

In [26]:
```python
d=pd.read_csv("Salary_Data.csv")
print(d)
```

```
    YearsExperience    Salary
0               1.1   39343.0
1               1.3   46205.0
2               1.5   37731.0
3               2.0   43525.0
4               2.2   39891.0
5               2.9   56642.0
6               3.0   60150.0
7               3.2   54445.0
8               3.2   64445.0
9               3.7   57189.0
10              3.9   63218.0
11              4.0   55794.0
12              4.0   56957.0
13              4.1   57081.0
14              4.5   61111.0
15              4.9   67938.0
16              5.1   66029.0
17              5.3   83088.0
18              5.9   81363.0
19              6.0   93940.0
20              6.8   91738.0
21              7.1   98273.0
22              7.9  101302.0
23              8.2  113812.0
24              8.7  109431.0
25              9.0  105582.0
26              9.5  116969.0
27              9.6  112635.0
28             10.3  122391.0
29             10.5  121872.0
```

```
In [27]: x=d.iloc[:,:-1]
         print("Independent variable\n",x)
```

```
Independent variable
      YearsExperience
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              3.2
9              3.7
10             3.9
11             4.0
12             4.0
13             4.1
14             4.5
15             4.9
16             5.1
17             5.3
18             5.9
19             6.0
20             6.8
21             7.1
22             7.9
23             8.2
24             8.7
25             9.0
26             9.5
27             9.6
28            10.3
29            10.5
```

```
In [28]: y=d.iloc[:,-1:]
         print("Dependent variable\n",y)
```

```
Dependent variable
        Salary
0     39343.0
1     46205.0
2     37731.0
3     43525.0
4     39891.0
5     56642.0
6     60150.0
7     54445.0
8     64445.0
9     57189.0
10    63218.0
11    55794.0
12    56957.0
13    57081.0
14    61111.0
15    67938.0
16    66029.0
17    83088.0
18    81363.0
19    93940.0
20    91738.0
21    98273.0
22   101302.0
23   113812.0
24   109431.0
25   105582.0
26   116969.0
27   112635.0
28   122391.0
29   121872.0
```

```
In [29]: x_train,x_test,y_train,y_test=train_test_split(x,y,random_state=0)
         print("x_train\n",x_train)
```

```
x_train
     YearsExperience
17            5.3
22            7.9
5             2.9
16            5.1
8             3.2
14            4.5
23            8.2
20            6.8
1             1.3
29           10.5
6             3.0
4             2.2
18            5.9
19            6.0
9             3.7
7             3.2
25            9.0
3             2.0
0             1.1
21            7.1
15            4.9
12            4.0
```

```
In [30]: regressor=LinearRegression()
         regressor.fit(x_train,y_train)
```

```
Out[30]: LinearRegression()
```

```
In [31]: y_pred=regressor.predict(x_test)
         print("Predicted values of x_test\n",y_pred)
```

```
Predicted values of x_test
 [[ 41056.25705466]
 [123597.70938378]
 [ 65443.50433372]
 [ 63567.56223533]
 [116093.94099022]
 [108590.17259667]
 [117031.91203942]
 [ 64505.53328452]]
```
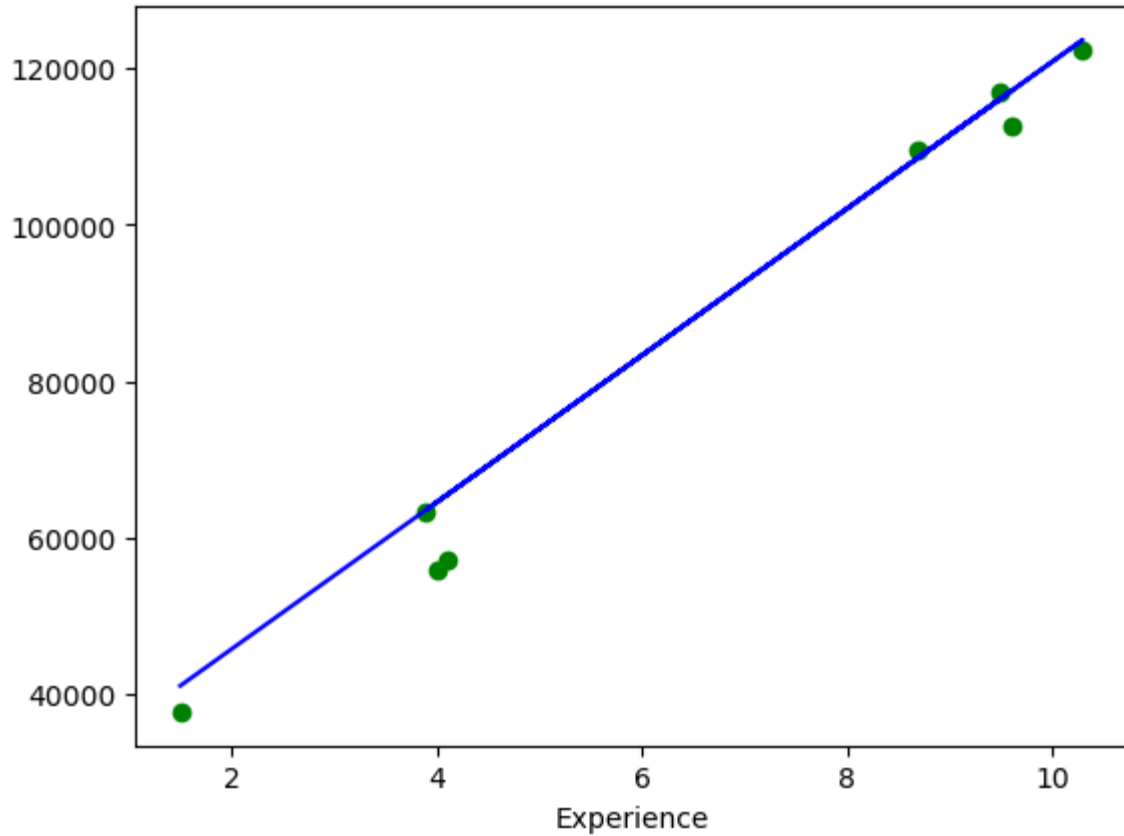
```
In [32]: print("--------Graph for x_train--------")
         plt.scatter(x_train,y_train,color='red')
         plt.plot(x_train,regressor.predict(x_train),color='blue')
         plt.xlabel('Experience')
         plt.show()
```

--------Graph for x_train--------
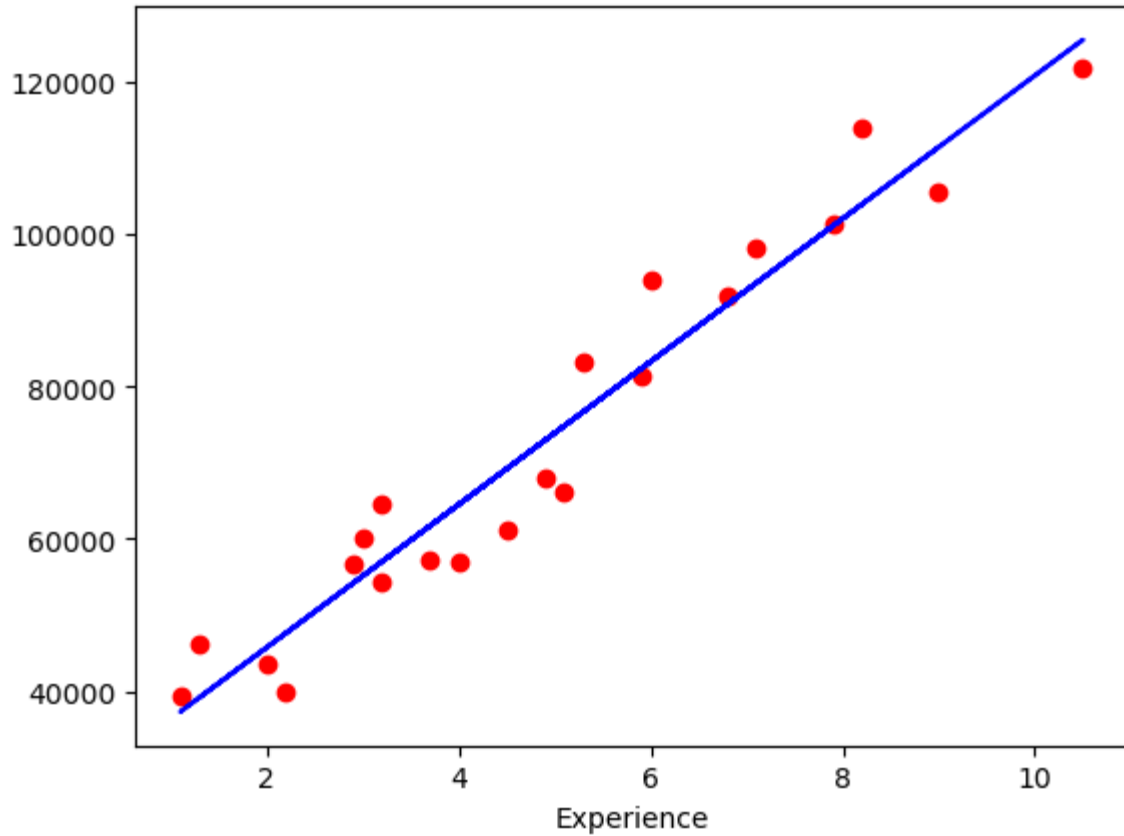
In [33]:
```python
print("--------Graph for x_test--------")
plt.scatter(x_test,y_test,color='green')
plt.plot(x_test,regressor.predict(x_test),color='blue')
plt.xlabel('Experience')
plt.show()
```

--------Graph for x_test--------

In [ ]: