In [10]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
from sklearn.model_selection import train_test_split
```

In [20]:
```
#read the dataset
salary_data = pd.read_csv('Salary_dataset.csv')

#print the head and tail to know the data
salary_data.head()
```

Out[20]:

	YearsExperience	Salary
0	1.2	39344
1	1.4	46206
2	1.6	37732
3	2.1	43526
4	2.3	39892

In [21]:
```
salary_data.tail()
```

Out[21]:

	YearsExperience	Salary
25	9.1	105583
26	9.6	116970
27	9.7	112636
28	10.4	122392
29	10.6	121873

In [23]:
```
salary_data.describe()
```

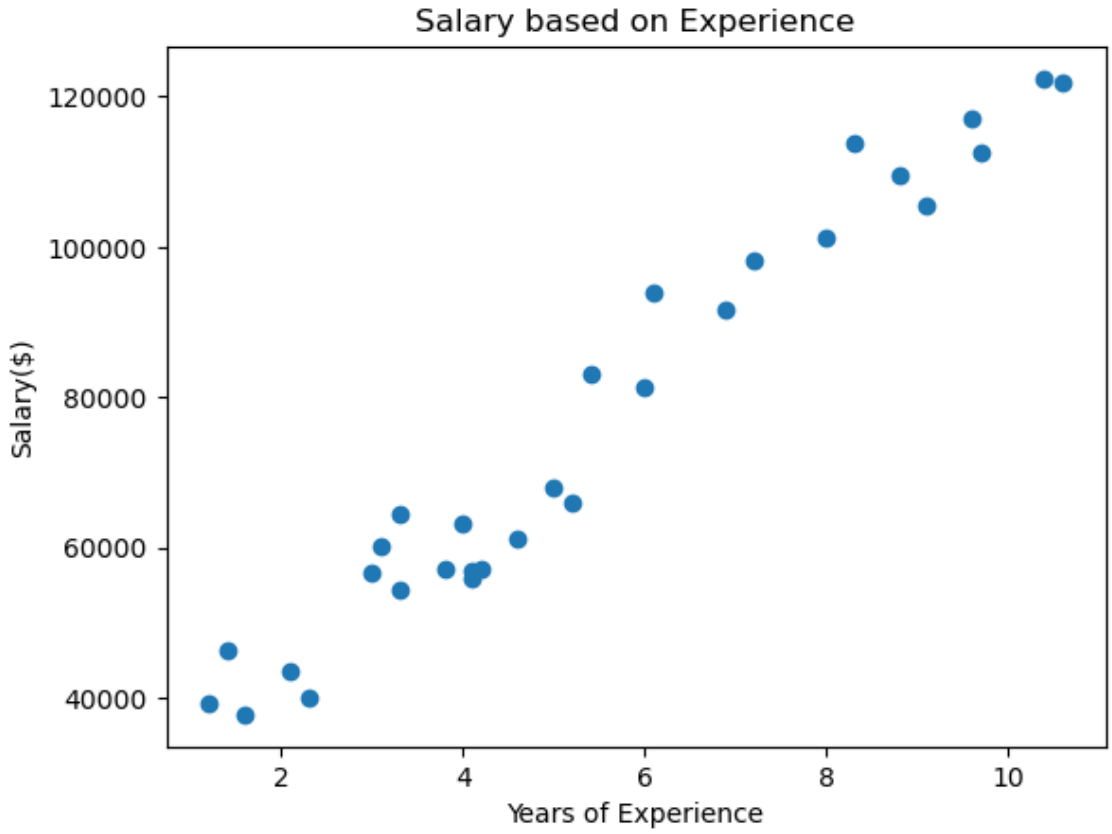
Out[23]:

	YearsExperience	Salary
count	30.000000	30.000000
mean	5.413333	76004.000000
std	2.837888	27414.429785
min	1.200000	37732.000000
25%	3.300000	56721.750000
50%	4.800000	65238.000000
75%	7.800000	100545.750000
max	10.600000	122392.000000

In [24]:
```
#create independent and dependent variables
indepent_var = salary_data['YearsExperience'].values.reshape(-1,1)
dependent_var = salary_data['Salary']
```

In [25]:
```
#create a scatter plot
plt.scatter(indepent_var,dependent_var)
plt.xlabel("Years of Experience")
plt.ylabel("Salary($)")
plt.title("Salary based on Experience")
```

Out[25]:
```
Text(0.5, 1.0, 'Salary based on Experience')
```



In [169]:
```
#assign the independent and dependent var with train_test_split
#dividing the training data and testing data, assi
train_x,test_x,train_y,test_y = train_test_split(indepent_var,dependent_va
```

In [170]:
```
lin_reg_model = LinearRegression()
```

In [171]:
```
lin_reg_model.fit(train_x,train_y)
```

Out[171]:
```
▾ LinearRegression

LinearRegression()
```

In [172]:
```
#with test_x, we give the machine values and let the model predict our valu
predictive_val = lin_reg_model.predict(test_x)
predictive_val
```

Out[172]:
```
array([ 36144.62176044,  34238.05465324,  66649.69547576,  59023.4270469
3,
        91435.06786946,  80948.94877982, 101921.1869591 ,  52350.4421717
1,
        42817.60663567])
```

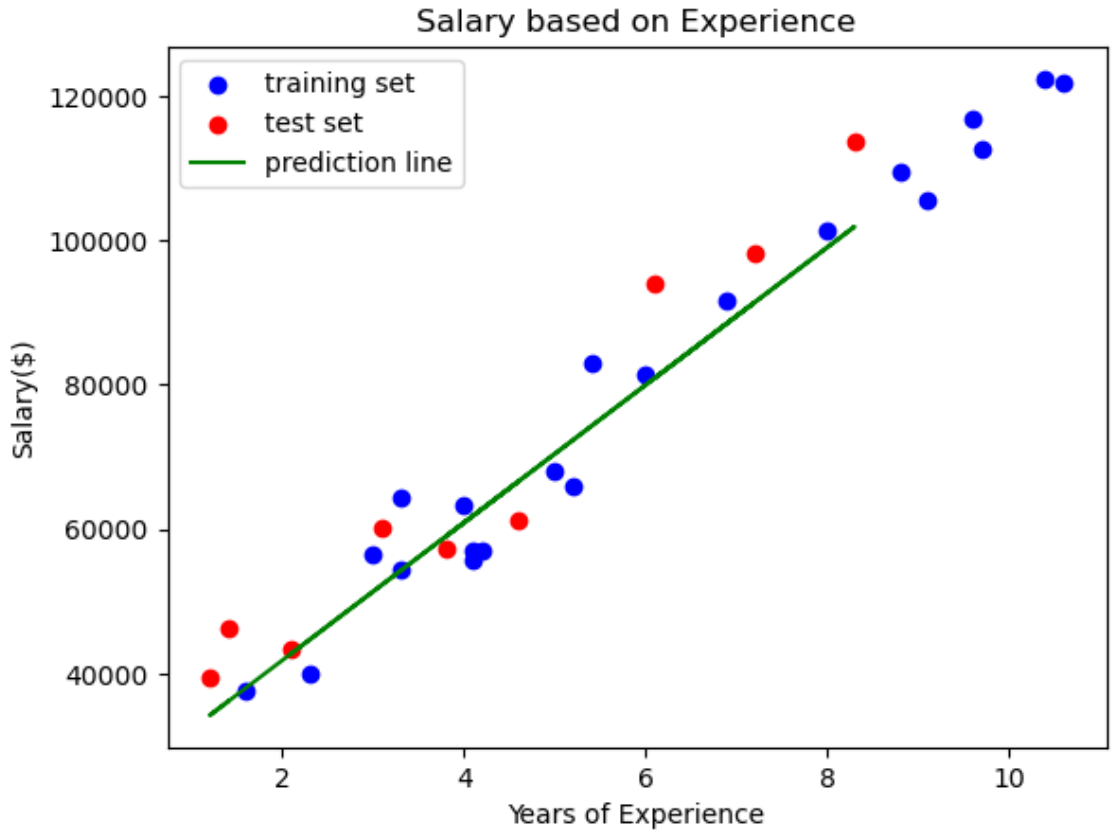
In [173]:
```
#calculate mean square error
mse = mean_squared_error(predictive_val, test_y)
```

In [174]:
```
mse
```

Out[174]:
```
64406629.3852596
```

In [175]:
```
#plot the values
plt.scatter(train_x,train_y,color='blue',label='training set')
plt.scatter(test_x,test_y,color='red',label='test set')
plt.plot(test_x,predictive_val,color="green",label="prediction line")
plt.legend()
plt.xlabel("Years of Experience")
plt.ylabel("Salary($)")
plt.title("Salary based on Experience")
```

Out[175]:
```
Text(0.5, 1.0, 'Salary based on Experience')
```



In [ ]:

In [ ]: