

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

df=pd.read_csv("/content/Salary_Data.csv")
df.head()
```

	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

```
X=df.iloc[:, :-1].values
Y=df.iloc[:, -1].values
```

```
print(X)
```

```
[[ 1.1]
 [ 1.3]
 [ 1.5]
 [ 2. ]
 [ 2.2]
 [ 2.9]
 [ 3. ]
 [ 3.2]
 [ 3.2]
 [ 3.7]
 [ 3.9]
 [ 4. ]
 [ 4. ]
 [ 4.1]
 [ 4.5]
 [ 4.9]
 [ 5.1]
 [ 5.3]
 [ 5.9]
 [ 6. ]
 [ 6.8]
 [ 7.1]
 [ 7.9]
 [ 8.2]
 [ 8.7]
 [ 9. ]
 [ 9.5]
 [ 9.6]
 [10.3]
 [10.5]]
```

```
print(Y)
```

```
[ 39343.  46205.  37731.  43525.  39891.  56642.  60150.  54445.
64445.
 57189.  63218.  55794.  56957.  57081.  61111.  67938.  66029.
83088.
 81363.  93940.  91738.  98273. 101302. 113812. 109431. 105582.
116969.
112635. 122391. 121872.]
```

#Since there are no null values will directly proceed to train test split

```
from sklearn.model_selection import train_test_split
X_train,X_test,Y_train,Y_test=train_test_split(X,Y,test_size=0.25,random_state=0)
```

```
print(X_train)
```

```
[[ 5.3]
 [ 7.9]
 [ 2.9]
 [ 5.1]
 [ 3.2]
 [ 4.5]
 [ 8.2]
 [ 6.8]
 [ 1.3]
 [10.5]
 [ 3. ]
 [ 2.2]
 [ 5.9]
 [ 6. ]
 [ 3.7]
 [ 3.2]
 [ 9. ]
 [ 2. ]
 [ 1.1]
 [ 7.1]
 [ 4.9]
 [ 4. ]]
```

```
print(X_test)
```

```
[[ 1.5]
 [10.3]
 [ 4.1]
 [ 3.9]
 [ 9.5]
 [ 8.7]
 [ 9.6]
 [ 4. ]]
```

```

print(Y_train)

[ 83088. 101302.  56642.  66029.  64445.  61111. 113812.  91738.
 46205.
 121872.  60150.  39891.  81363.  93940.  57189.  54445. 105582.
 43525.
  39343.  98273.  67938.  56957.]

print(Y_test)

[ 37731. 122391.  57081.  63218. 116969. 109431. 112635.  55794.]

#Building the Linear Regression model
from sklearn.linear_model import LinearRegression
regressor=LinearRegression()
regressor.fit(X_train,Y_train)

LinearRegression()

y_pred=regressor.predict(X_test)

#Visualizing training set
plt.scatter(X_train,Y_train,color='red')
plt.plot(X_train,regressor.predict(X_train))
plt.title('Salary Vs Experience(Training set)')
plt.xlabel('No.of years of experience')
plt.ylabel('Salary')
plt.show()

```



```
#Visualizing test set  
plt.scatter(X_test,Y_test,color='red')  
plt.plot(X_train,regressor.predict(X_train))  
plt.title('Salary Vs Experience(Test set)')  
plt.xlabel('No. of years of experience')  
plt.ylabel('Salary')  
plt.show()
```



```
from sklearn.metrics import r2_score

score=r2_score(Y_test,y_pred)
print(score)

0.9779208335417602

#Predicting results for someone who has worked for 12 years
regressor.predict([[12]])

array([139543.21722009])

#Checking the regressor coefficient and intercept

print(regressor.coef_)
print(regressor.intercept_)

[9379.71049195]
26986.691316737248

# @title Default title text
#Therefore the equation stands as
#y=9379.71049195*x+26986.691316737248
```

#Here x is the number of years of experience