

Import Library

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

Import Dataset

```
dataset = read_csv('salary_data.csv')
print(dataset)
```

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

print(X)
print(Y)
```

```
plt.xlabel('area')
plt.ylabel('price')
plt.title('Area vs. Price')
plt.scatter(X, Y, color='red', marker='+')
```

Split the data

```
from sklearn.model_selection import train_test_split
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.3)
print(X_train.shape, X_test.shape)
```

Fitting Simple Linear Regression

```
from sklearn.linear_model import LinearRegression
model = LinearRegression()
model.fit(X_train, Y_train)
```

Predicting the test results

```
Y_pred = model.predict(X_test)
print(Y_pred)
print(Y_train)
```

Calculate Accuracy

```
accuracy = model.score(X_test, Y_test)
print(accuracy)
```

Plot the regression curve

```
plt.scatter(X_train, Y_train, color='red', marker='+')
plt.xlabel('YearsExperience')
plt.ylabel('Salary')
plt.title('YearsExperience vs Salary (Traning Data) ')
plt.plot(X_train, model.predict(X_train), color='blue')
plt.show()

plt.scatter(X_test, Y_test, color='red', marker='+')
plt.xlabel('YearsExperience')
plt.ylabel('Salary')
plt.title('YearsExperience vs Salary (Test DataSet)')
```

```
plt.plot(X_test, model.predict(X_test), color='blue')  
plt.show()
```

✓ Predict on external data

```
test_data = np.array([[3300]])  
test_pred = model.predict(test_data)  
print(test_pred)
```

✓ Predict data from user through manual input

```
WorkExperience = int(input("How many years have you worked in this field of expertise?"))  
SalaryEarned = model.predict(np.array([[WorkExperience]]))  
print('Estimated salary ' + str(WorkExperience) + 'years of experience are' + str(SalaryEarned[0]))
```