import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

# Load the dataset

dataset = pd.read\_csv(r'D:\Samsom - All Data\Naresh IT Institute\New folder\Salary\_Data.csv')

# Check the shape of the dataset

print("Dataset Shape:", dataset.shape) # (30, 2)

# Feature selection (independent variable x and dependent variable)

x = dataset.iloc[:, :-1] # Years of experience (Independent Variable)

y = dataset.iloc[:, -1] # Salary (Dependent variable)

# Split the dataset into training and testing sets (80% training)

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\_state=0)

# Reshape x\_train and x\_test into 2D arrays if they are single

x\_train = x\_train.values.reshape(-1, 1)

x\_test = x\_test.values.reshape(-1, 1)

# Predicting the results for the test set

from sklearn.linear\_model import LinearRegression

from sklearn.model\_selection import train\_test\_split

regressor = LinearRegression()

regressor.fit(x\_train, y\_train)

y\_pred = regressor.predict(x\_test)

# Compare predicted and actual salaries from the test set

comparison = pd.DataFrame({'Actual': y\_test, 'Predicted': y\_pred})

print(comparison)

# Visualizing the Training set results

plt.scatter(x\_test, y\_test, color = 'red') # Real salary

plt.plot(x\_train, regressor.predict(x\_train), color = 'blue')

plt.title('Salary vs Experience (Training set)')

plt.xlabel('Years of Experience')

plt.ylabel('Salary')

plt.show()

m\_slope = regressor.coef\_

print(m\_slope)

c\_intercept = regressor.intercept\_

print(c\_intercept)

y\_12 = m\_slope\*12+c\_intercept

print(y\_12)