# Untitled

January 22, 2026

```python
[1]: import numpy as np
     import pandas as pd
     import matplotlib.pyplot as plt
     from sklearn.linear_model import LinearRegression
```

```python
[2]: filename="Salary_dataset.csv"
```

```python
[3]: df=pd.read_csv(filename,usecols=['YearsExperience','Salary'])
```

```python
[4]: df['sqr_sal'] = df['Salary'] ** 2
     df['sqr_yrs'] = df['YearsExperience'] ** 2
     df['xy'] = df['YearsExperience'] * df['Salary']


     n = len(df)


     m = (n * df['xy'].sum() - df['YearsExperience'].sum() * df['Salary'].sum()) / \
         (n * df['sqr_yrs'].sum() - (df['YearsExperience'].sum() ** 2))


     b = (df['Salary'].sum() - m * df['YearsExperience'].sum()) / n


     equation_label = f'y = {m:.2f}x + {b:.2f}'


     x = np.linspace(df['YearsExperience'].min(), df['YearsExperience'].max(), 100)
     y = m * x + b
```
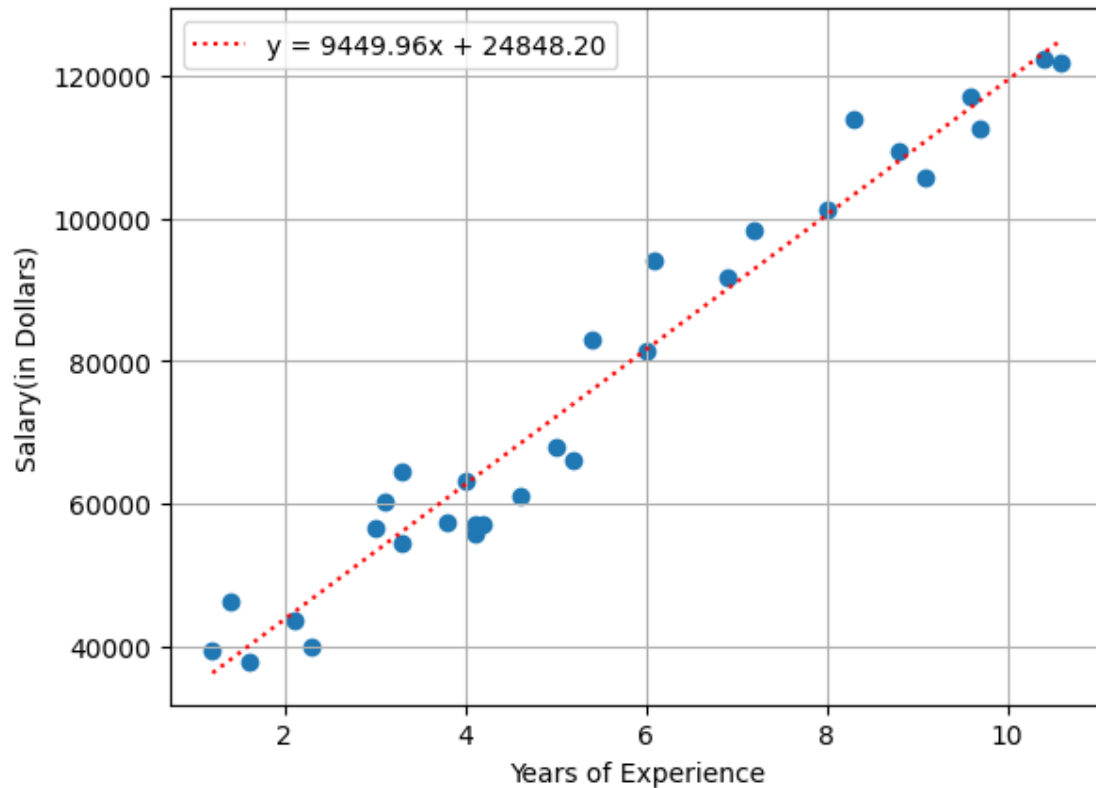
```python
[5]: plt.plot(x, y, label=equation_label, color='red' ,linestyle=":")

     plt.xlabel("Years of Experience")
     plt.ylabel("Salary(in Dollars)")
     plt.scatter(df['YearsExperience'],df['Salary'])
     plt.legend()
     plt.grid(True)


     plt.show()
```

Figure: Scatter plot of Salary (in Dollars) vs Years of Experience with fitted line $y = 9449.96x + 24848.20$.

```
[6]: x_trained = 5
```

```
[7]: y_trained = m * x_trained + b
```

```
[8]: print(int(y_trained), " is the expected pay")
```

72098   is the expected pay

```
[9]: from sklearn.linear_model import LinearRegression

     # Features and target
     X = df[['YearsExperience']]    # 2D
     y = df['Salary']               # 1D

     # Train model
     model = LinearRegression()
     model.fit(X, y)

     print("Coefficient:", model.coef_)
     print("Intercept:", model.intercept_)

     x_trained = 5
```

```
prediction = model.predict([[x_trained]])
print(f"Predicted value for {x_trained} years:", int(prediction[0]))
```

```
Coefficient: [9449.96232146]
Intercept: 24848.203966523208
Predicted value for 5 years: 72098
```