

Machine Learning

▾ Step-1 Import Libraries

```
1 pip install scikit-learn
```

```
↳ Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/  
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.10/dist-packages (1.2.2)  
Requirement already satisfied: numpy>=1.17.3 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.22.4)  
Requirement already satisfied: scipy>=1.3.2 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.10.1)  
Requirement already satisfied: joblib>=1.1.1 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (1.2.0)  
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from scikit-learn) (3.1.0)
```

```
1 import numpy as np  
2 import pandas as pd  
3 import seaborn as sns  
4 import matplotlib.pyplot as plt  
5 import sklearn
```

▾ Step-2 Import Data

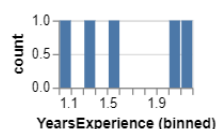
```
1 df=pd.read_csv('salary_data.csv')  
2 df.head()
```

	YearsExperience	Salary
0	1.1	39343
1	1.3	46205
2	1.5	37731
3	2.0	43525
4	2.2	39891

```

1 import altair as alt
2 from google.colab import altair
3 df_5778746289481977118 = altair.get_registered_df('df_5778746289481977118')
4
5 def histogram(df, colname, maxbins=20, width=100, height=50):
6     return (alt.Chart(df).mark_bar()
7             .encode(
8                 x=alt.X(colname, bin=alt.Bin(maxbins=maxbins)),
9                 y=alt.Y('count()', axis=alt.Axis(title='count')),
10            )
11            .properties(width=width, height=height)) # Pixels.
12
13 chart = histogram(df_5778746289481977118, *['YearsExperience'], **{})
14 chart

```



▼ Selecting input and Output variable

```

1 X=df[["YearsExperience"]]
2 y=df["Salary"]

```

▼ Step-4 Making Linear Regression Model

```

1 from sklearn.linear_model import LinearRegression
2 model = LinearRegression()

```

▼ Step-5 Fitting the Model

```

1 plt.xlabel('X')
2 plt.ylabel('y')
3 model=model.fit(X,y)
4 model

```

LinearRegression
LinearRegression()

Step-6 Predicting the Model

```
1 model.predict([[20],[25],[30]])
2 model
```

```
-----
AttributeError                                Traceback (most recent call last)
<ipython-input-42-413c418e1508> in <cell line: 1>()
----> 1 model.predict([[20],[25],[30]])
      2 model
```

AttributeError: 'list' object has no attribute 'predict'

SEARCH STACK OVERFLOW

Step-7 Evaluating the Model

```
1 from sklearn.model_selection import train_test_split
2 X_train,y_train,X_test,y_test=train_test_split(X,y,test_size=0.2,random_state=0)
3 # Making and Fitting the model
4 model.fit(X_train,X_test)
```

LinearRegression
LinearRegression()

```
1 plt.scatter(X, y, color='blue', label='Actual Data')
2 plt.plot(X, model.predict(X), color='red', label='Linear Regression')
3 plt.xlabel('X')
4 plt.ylabel('y')
5 plt.legend()
6 plt.show()
```

