

```
/kaggle/input/linear-regression-dataset/Linear Regression - Sheet1.csv
```

```
[161]: df=pd.read_csv('/kaggle/input/linear-regression-dataset/Linear Regression - Sheet1.csv')
df.head()
```

| [161... | X | Y |
|---------|---|----------|
| 0 | 1 | 3.888889 |
| 1 | 2 | 4.555556 |
| 2 | 3 | 5.222222 |
| 3 | 4 | 5.888889 |
| 4 | 5 | 6.555556 |

```
[162]: df.describe()
```

| | X | Y |
|-------|------------|------------|
| count | 300.000000 | 300.000000 |
| mean | 150.500000 | 102.215556 |
| std | 86.746758 | 57.842711 |
| min | 1.000000 | 1.888889 |
| 25% | 75.750000 | 52.388889 |
| 50% | 150.500000 | 102.222222 |
| 75% | 225.250000 | 152.055556 |
| max | 300.000000 | 201.888889 |

```
[163]: total_null=df.isnull().sum().sort_values(ascending=False)
percent=((df.isnull().sum()/df.isnull().count())*100).sort_values(ascending=False)
print("total=",df.shape[0])
```

```
total= 300
```

```
[164]: missing_data=pd.concat([total_null,percent,round(2)],keys=['total miss','percent'])
missing_data.head(10)
```

```
[164... total miss X    0.0
          Y    0.0
        percent X    0.0
          Y    0.0
        dtype: float64
```

```
[165]: ##No Null values
```

[]:

```
[166]: from sklearn.model_selection import train_test_split
X_train, X_test, Y_train, Y_test = train_test_split(df[['X']], df[['Y']], test_size=0.33, random_state=42)
```

Linear Regression

```
[167]: X_train=X_train.values.reshape(-1,1)
Y_train=Y_train.values.reshape(-1,1)
X_test=X_test.values.reshape(-1,1)
Y_test=Y_test.values.reshape(-1,1)
```

```
[168]: from sklearn.linear_model import LinearRegression
model=LinearRegression()
```

```
[ ]:
```

```
[169]: lr=model.fit(X_train,Y_train)
lr.score(X_train,Y_train)
```

```
[169...] 0.8838472271064208
```

```
[ ]:
```

Find Coefficient and Predict

```
[170]: lr.coef_
```

```
[170...] array([[0.62791701]])
```

```
[171]: lr.intercept_
```

```
[171...] array([7.22340726])
```

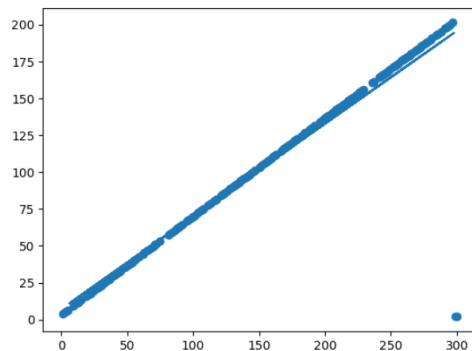
```
[172]: Y_pred=lr.predict(X_test)
```

Plot for linear regression

```
[173]: import matplotlib.pyplot as plt
```

```
[174]: plt.scatter(X_train,Y_train)
plt.plot(X_test,Y_pred)
```

```
[174...] [<matplotlib.lines.Line2D at 0x7f74ab6c6050>]
```



Accuracy

```
[175]: test=pd.read_csv('/kaggle/input/linear-regression-dataset/Linear Regression - Sheet1.csv')
```

```
[176]: X_test=np.array(test.iloc[:, :-1].values)
Y_test=np.array(test.iloc[:, 1].values)
```

```
[177]:
```

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

0.9209585680911518

[]:



+ Code

+ Markdown

