



```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
```

```
In [2]: df = pd.read_csv("/content/drive/MyDrive/Hours and Scores.csv")
print(df.head())
```

```
Unnamed: 0  Hours  Scores
0      2.5,21    2.5     21
1      5.1,47    5.1     47
2      3.2,27    3.2     27
3      8.5,75    8.5     75
4      3.5,30    3.5     30
```

```
In [3]: print(df.isnull().sum())
```

```
Unnamed: 0      0
Hours          0
Scores         0
dtype: int64
```

```
In [5]: X = df[['Hours']]
y = df['Scores']
```

```
In [6]: X_train, X_test, y_train, y_test = train_test_split(
        X, y, test_size=0.2, random_state=42
    )
```

```
In [8]: model = LinearRegression()
model.fit(X_train, y_train)
```

```
Out[8]: ▾ LinearRegression ⓘ ?  
LinearRegression()
```

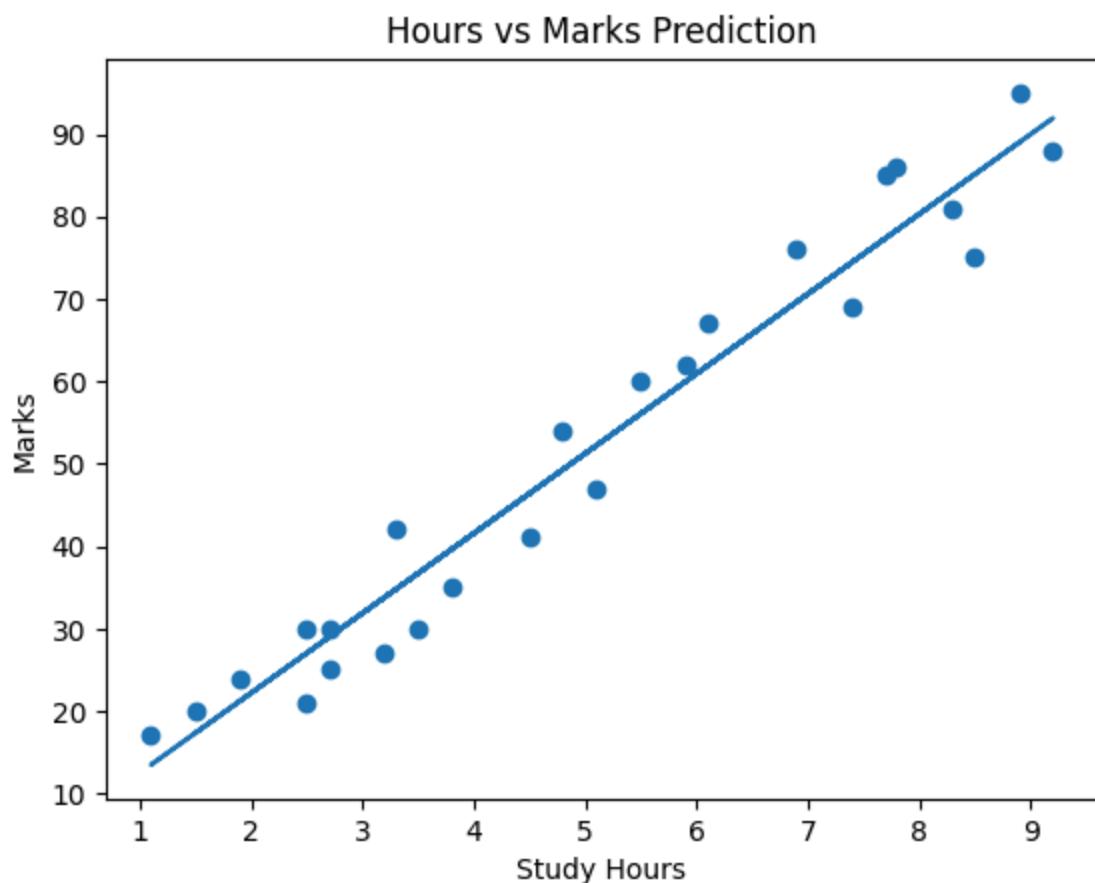
```
In [9]: y_pred = model.predict(X_test)
```

```
In [10]: mse = mean_squared_error(y_test, y_pred)
print("Mean Squared Error:", mse)
```

```
Mean Squared Error: 18.943211722315272
```

```
In [11]: plt.scatter(df['Hours'], df['Scores'])
plt.plot(df['Hours'], model.predict(df[['Hours']]))

plt.xlabel("Study Hours")
plt.ylabel("Marks")
plt.title("Hours vs Marks Prediction")
plt.show()
```



```
In [14]: hours = float(input("Enter study hours: "))
predicted_score = model.predict([[hours]])
print("Predicted Marks:", predicted_score[0])
```

```
Enter study hours: 10
Predicted Marks: 99.6476738984567
```

```
/usr/local/lib/python3.12/dist-packages/sklearn/utils/validation.py:2739: UserWarning: X does not have valid feature names, but LinearRegression was fitted with feature names
    warnings.warn(
```