

HAMZA PATEL - 2305215

```
In [1]: import pandas as pd
import numpy as np
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, r2_score
```

```
In [2]: df = pd.read_csv("scores2.csv")
print(df.head())
```

	Student_ID	Math	Science	English
0	1	78	72	80
1	2	85	89	87
2	3	62	60	65
3	4	90	92	91
4	5	55	58	61

```
In [3]: X = df.iloc[:, :-1]
y = df.iloc[:, -1]

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_st
```

```
In [4]: model = LinearRegression()
model.fit(X_train, y_train)

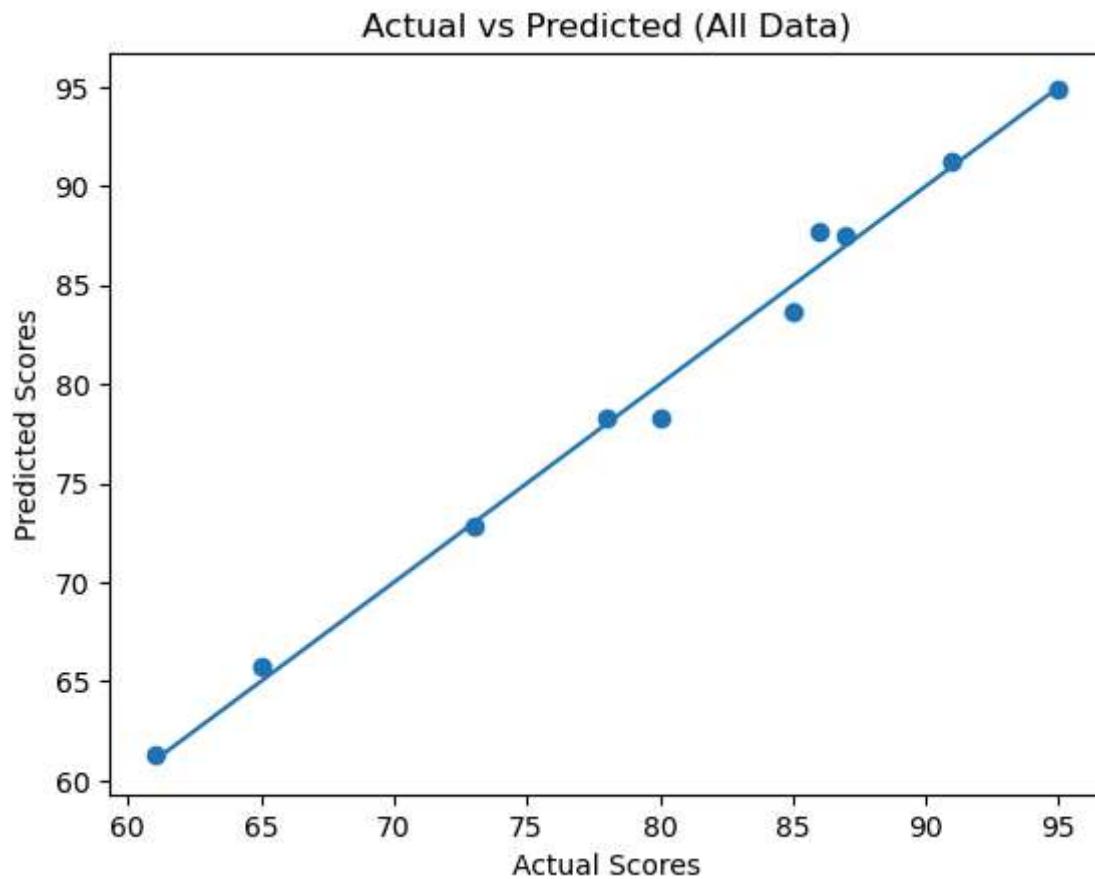
y_pred = model.predict(X_test)

print("MSE:", mean_squared_error(y_test, y_pred))
print("R2 Score:", r2_score(y_test, y_pred))
```

MSE: 0.13842878210794277
R2 Score: 0.9913482011182536

```
In [5]: y_all_pred = model.predict(X)

plt.figure()
plt.scatter(y, y_all_pred)
plt.plot([y.min(), y.max()], [y.min(), y.max()])
plt.xlabel("Actual Scores")
plt.ylabel("Predicted Scores")
plt.title("Actual vs Predicted (All Data)")
plt.show()
```



In []: