



Step 1: Install and import libraries

```
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 r2_score, mean_absolute_error

from google.colab import files
import zipfile, io
```

Step 2: Upload ZIP file

```
print("📁 Upload the rainfall dataset ZIP file...")
uploaded = files.upload()
```

```
zip_file_name = list(uploaded.keys())[0] # get zip file name
```

Step 3: Extract CSV from ZIP

```
with zipfile.ZipFile(io.BytesIO(uploaded[zip_file_name]), 'r') as z:
    csv_files = z.namelist()
    print("Files found in ZIP:", csv_files)
```

Load the rainfall dataset CSV (auto-detect file containing 'rainfall')

```
target_file = [f for f in csv_files if 'rainfall' in f.lower()][0]
```

```
print("📄 Using file:", target_file)
```

```
with z.open(target_file) as f:
```

```
    df = pd.read_csv(f)
```

```
# ✅ Step 4: Show first 5 rows
print("\n🚀 Dataset Preview:")
print(df.head())

# ✅ Step 5: Select input and output columns
X = df[['JAN', 'FEB', 'MAR', 'APR', 'MAY', 'JUN', 'JUL', 'AUG', 'SEP', 'OCT', 'NOV', 'DEC']]
y = df['ANNUAL']

# ✅ Step 6: Train-Test Split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# ✅ Step 7: Train Model
model = LinearRegression()
model.fit(X_train, y_train)

# ✅ Step 8: Predict
y_pred = model.predict(X_test)

# ✅ Step 9: Print Accuracy
print("\n📊 Model Performance:")
print("R² Score :", r2_score(y_test, y_pred))
print("MAE :", mean_absolute_error(y_test, y_pred), "mm")

# ✅ Step 10: Plot Graph
plt.figure(figsize=(8,5))
plt.plot(y_test.values, label="Actual Rainfall")
plt.plot(y_pred, label="Predicted Rainfall")
plt.title("Actual vs Predicted Rainfall")
plt.xlabel("Index")
```

```
plt.xlabel('Index')  
plt.ylabel('Rainfall (mm)')  
plt.legend()  
plt.show()
```

archive (1).zip

archive (1).zip(application/x-zip-compressed) - 191592 bytes, last modified: 11/5/2025 - 100% done

Saving archive (1).zip to archive (1).zip

Files found in ZIP: ['district wise rainfall normal.csv', 'rainfall in india 1901-2015.csv']

Using file: district wise rainfall normal.csv

Dataset Preview:

	STATE_UT_NAME			DISTRICT	JAN	FEB	MAR	APR	\
0	ANDAMAN And	NICOBAR	ISLANDS	NICOBAR	107.3	57.9	65.2	117.0	
1	ANDAMAN And	NICOBAR	ISLANDS	SOUTH ANDAMAN	43.7	26.0	18.6	90.5	
2	ANDAMAN And	NICOBAR	ISLANDS	N & M ANDAMAN	32.7	15.9	8.6	53.4	
3		ARUNACHAL	PRADESH	LOHIT	42.2	80.8	176.4	358.5	
4		ARUNACHAL	PRADESH	EAST SIANG	33.3	79.5	105.9	216.5	

	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	\
0	358.5	295.5	285.0	271.9	354.8	326.0	315.2	250.9	2805.2	165.2	
1	374.4	457.2	421.3	423.1	455.6	301.2	275.8	128.3	3015.7	69.7	
2	343.6	503.3	465.4	460.9	454.8	276.1	198.6	100.0	2913.3	48.6	
3	306.4	447.0	660.1	427.8	313.6	167.1	34.1	29.8	3043.8	123.0	
4	323.0	738.3	990.9	711.2	568.0	206.9	29.5	31.7	4034.7	112.8	

	Mar-May	Jun-Sep	Oct-Dec
0	540.7	1207.2	892.1
1	483.5	1757.2	705.3
2	405.6	1884.4	574.7
3	841.3	1848.5	231.0
4	645.4	3008.4	268.1



Model Performance:

R^2 Score : 1.0

MAE : 1.1192425109027314e-12 mm

Actual vs Predicted Rainfall

