

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn import metrics
from sklearn.svm import SVC
from xgboost import XGBClassifier
from sklearn.linear_model import LogisticRegression
from imblearn.over_sampling import RandomOverSampler

import warnings
warnings.filterwarnings('ignore')

data=pd.read_csv('/content/rainfall in india 1901-2015.csv')
data.head()
```

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan- Feb	Mar- May	Jun- Sep	Oct- Dec
0	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6	388.5	558.2	33.6	3373.2	136.3	560.3	1696.3	980.3
1	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.2	0.0	446.1	537.1	228.9	753.7	666.2	197.2	359.0	160.5	3520.7	159.8	458.3	2185.9	716.7
2	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.0	1.0	235.1	479.9	728.4	326.7	339.0	181.2	284.4	225.0	2957.4	156.7	236.1	1874.0	690.6
3	ANDAMAN & NICOBAR ISLANDS	1904	9.4	14.7	0.0	202.4	304.5	495.1	502.0	160.1	820.4	222.2	308.7	40.1	3079.6	24.1	506.9	1977.6	571.0

```
data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4116 entries, 0 to 4115
Data columns (total 19 columns):
#   Column          Non-Null Count  Dtype
---  -
0   SUBDIVISION      4116 non-null   object
1   YEAR             4116 non-null   int64
2   JAN              4112 non-null   float64
3   FEB              4113 non-null   float64
4   MAR              4110 non-null   float64
5   APR              4112 non-null   float64
6   MAY              4113 non-null   float64
7   JUN              4111 non-null   float64
8   JUL              4109 non-null   float64
9   AUG              4112 non-null   float64
10  SEP              4110 non-null   float64
11  OCT              4109 non-null   float64
12  NOV              4105 non-null   float64
13  DEC              4106 non-null   float64
14  ANNUAL           4090 non-null   float64
15  Jan-Feb          4110 non-null   float64
16  Mar-May          4107 non-null   float64
17  Jun-Sep          4106 non-null   float64
18  Oct-Dec          4103 non-null   float64
dtypes: float64(17), int64(1), object(1)
memory usage: 611.1+ KB

data.dropna(inplace=True)

data.isnull().sum()

data.head()
```



	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan- Feb	Mar- May	Jun- Sep	Oct- Dec
0	ANDAMAN & NICOBAR ISLANDS	1901	49.2	87.1	29.2	2.3	528.8	517.5	365.1	481.1	332.6	388.5	558.2	33.6	3373.2	136.3	560.3	1696.3	980.3
1	ANDAMAN & NICOBAR ISLANDS	1902	0.0	159.8	12.2	0.0	446.1	537.1	228.9	753.7	666.2	197.2	359.0	160.5	3520.7	159.8	458.3	2185.9	716.7
2	ANDAMAN & NICOBAR ISLANDS	1903	12.7	144.0	0.0	1.0	235.1	479.9	728.4	326.7	339.0	181.2	284.4	225.0	2957.4	156.7	236.1	1874.0	690.6

```

telangana = data.loc[data['SUBDIVISION'] == 'TELANGANA']
x = telangana['YEAR']
x.drop(columns=['YEAR'])
y = telangana['ANNUAL']
y.head()

```

```
x.shape, y.shape
```

```
import matplotlib.pyplot as plt
```

```

plt.scatter(x, y)
plt.plot()

```

After reshaping the 'x' and 'y' variables to match the expected input shape for the LinearRegression model, the code fits the LinearRegression model to the data. It calculates the intercept and coefficient of the regression line.

```

from sklearn.linear_model import LinearRegression
model = LinearRegression()
#x.head()
x = np.array(x).reshape(-1, 1)
y = np.array(y).reshape(-1, 1)
model.fit(x, y)

```

The code visualizes the scatter plot of the 'x' and 'y' data points and overlays the regression line on the plot.

```
b = model.intercept_
```

```
m = model.coef_
```

```

plt.scatter(x, y)
plt.scatter(x, m*x+b)
plt.plot()

```

