

Rainfall Prediction using Machine Learning

Team: 615851-U813CPU7

About Dataset

| rainfaLLIndia-2 | | | | | | |
|--------------------------|------|-------|-------|-------|-------|---------|
| subdivision | YEAR | JUN | JUL | AUG | SEP | JUN-SEP |
| ANDAMAN & NICOBAR ISLAND | 1901 | 517.5 | 365.1 | 481.1 | 332.6 | 1696.3 |
| ANDAMAN & NICOBAR ISLAND | 1902 | 537.1 | 228.9 | 753.7 | 666.2 | 2185.9 |
| ANDAMAN & NICOBAR ISLAND | 1903 | 479.9 | 728.4 | 326.7 | 339 | 1874 |
| ANDAMAN & NICOBAR ISLAND | 1904 | 495.1 | 502 | 160.1 | 820.4 | 1977.6 |
| ANDAMAN & NICOBAR ISLAND | 1905 | 628.7 | 368.7 | 330.5 | 297 | 1624.9 |
| ANDAMAN & NICOBAR ISLAND | 1906 | 733.3 | 247.7 | 320.5 | 164.3 | 1465.8 |
| ANDAMAN & NICOBAR ISLAND | 1907 | 305.2 | 443.9 | 377.6 | 200.4 | 1327.1 |
| ANDAMAN & NICOBAR ISLAND | 1908 | 693.6 | 481.4 | 699.9 | 428.8 | 2303.7 |
| ANDAMAN & NICOBAR ISLAND | 1910 | 472.7 | 264.3 | 337.4 | 626.6 | 1701 |
| ANDAMAN & NICOBAR ISLAND | 1911 | 649 | 253 | 187.1 | 464.5 | 1553.6 |
| ANDAMAN & NICOBAR ISLAND | 1912 | 549.8 | 468.9 | 370.3 | 386.2 | 1775.2 |
| ANDAMAN & NICOBAR ISLAND | 1913 | 530 | 280.8 | 205.8 | 580.1 | 1596.7 |
| ANDAMAN & NICOBAR ISLAND | 1914 | 383.3 | 792.8 | 520.5 | 310.8 | 2007.4 |
| ANDAMAN & NICOBAR ISLAND | 1915 | 334.7 | 269 | 317.2 | 429.8 | 1350.7 |
| ANDAMAN & NICOBAR ISLAND | 1916 | 450.1 | 317.3 | 425 | 561.2 | 1753.6 |
| ANDAMAN & NICOBAR ISLAND | 1917 | 301.1 | 394.8 | 437.4 | 471.8 | 1605.1 |
| ANDAMAN & NICOBAR ISLAND | 1918 | 710.8 | 200.9 | 455.4 | 303.3 | 1670.4 |
| ANDAMAN & NICOBAR ISLAND | 1919 | 542.5 | 246.5 | 259.8 | 170.7 | 1219.5 |
| ANDAMAN & NICOBAR ISLAND | 1920 | 546.9 | 294.4 | 467.4 | 505.4 | 1814.1 |
| ANDAMAN & NICOBAR ISLAND | 1921 | 282.7 | 487.1 | 330 | 581.2 | 1681 |
| ANDAMAN & NICOBAR ISLAND | 1922 | 506.1 | 425.8 | 307.4 | 511.7 | 1751 |
| ANDAMAN & NICOBAR ISLAND | 1923 | 808.4 | 636.9 | 182.2 | 560.5 | 2188 |
| ANDAMAN & NICOBAR ISLAND | 1924 | 261.2 | 493.3 | 290.9 | 251.2 | 1296.6 |
| ANDAMAN & NICOBAR ISLAND | 1925 | 663.8 | 241.8 | 278.2 | 201.9 | 1385.7 |

| | | | | | | |
|------------|------|-------|-------|-------|-------|--------|
| LAKSHDWEEP | 1996 | 427.1 | 335.3 | 197.3 | 230.4 | 1190.1 |
| LAKSHDWEEP | 1997 | 307 | 459.5 | 216.7 | 144 | 1127.2 |
| LAKSHDWEEP | 1998 | 382 | 388.8 | 196.7 | 274.7 | 1242.2 |
| LAKSHDWEEP | 1999 | 279.6 | 459.3 | 133.8 | 73.4 | 946.1 |
| LAKSHDWEEP | 2000 | 225.3 | 95.5 | 319.9 | 164.5 | 805.2 |
| LAKSHDWEEP | 2001 | 283.9 | 198.9 | 144.3 | 213.5 | 840.6 |
| LAKSHDWEEP | 2002 | 261.8 | 81.3 | 143.9 | 50 | 537 |
| LAKSHDWEEP | 2003 | 364.5 | 400.6 | 92.1 | 84.3 | 941.5 |
| LAKSHDWEEP | 2004 | 251.2 | 280.8 | 169.5 | 200 | 901.5 |
| LAKSHDWEEP | 2005 | 248.5 | 378.9 | 102.4 | 278 | 1007.8 |
| LAKSHDWEEP | 2006 | 286.9 | 172.3 | 150.7 | 318.5 | 928.4 |
| LAKSHDWEEP | 2007 | 573.4 | 427.3 | 294.7 | 457.5 | 1752.9 |
| LAKSHDWEEP | 2008 | 254.6 | 363.9 | 206.6 | 108.9 | 934 |
| LAKSHDWEEP | 2009 | 401.2 | 266.4 | 185 | 145 | 997.6 |
| LAKSHDWEEP | 2010 | 318.9 | 336.7 | 335.1 | 161.5 | 1152.2 |
| LAKSHDWEEP | 2011 | 153.6 | 350.2 | 254 | 255.2 | 1013 |
| LAKSHDWEEP | 2012 | 326.9 | 231.5 | 381.2 | 179.8 | 1119.4 |
| LAKSHDWEEP | 2013 | 426.2 | 296.4 | 150.8 | 180.4 | 1053.8 |
| LAKSHDWEEP | 2014 | 244.1 | 115.4 | 466.1 | 130.6 | 956.2 |
| LAKSHDWEEP | 2015 | 296.7 | 257.4 | 146.7 | 157.5 | 858.3 |
| LAKSHDWEEP | 2016 | 321.1 | 262.6 | 86.2 | 75.6 | 745.5 |
| LAKSHDWEEP | 2017 | 521.9 | 164.2 | 206.2 | 216 | 1108.3 |
| LAKSHDWEEP | 2018 | 203.2 | 139.7 | 124.4 | 74.4 | 541.7 |
| LAKSHDWEEP | 2019 | 242.5 | 489.4 | 338.4 | 156.7 | 1227 |
| LAKSHDWEEP | 2020 | 265.4 | 476.2 | 269.6 | 334.3 | 1345.5 |
| LAKSHDWEEP | 2021 | 160 | 156.5 | 319.5 | 154.9 | 790.9 |

Lenth - 4332

Subdivisions - 36

Total:

Max: 45082.6 [1961]

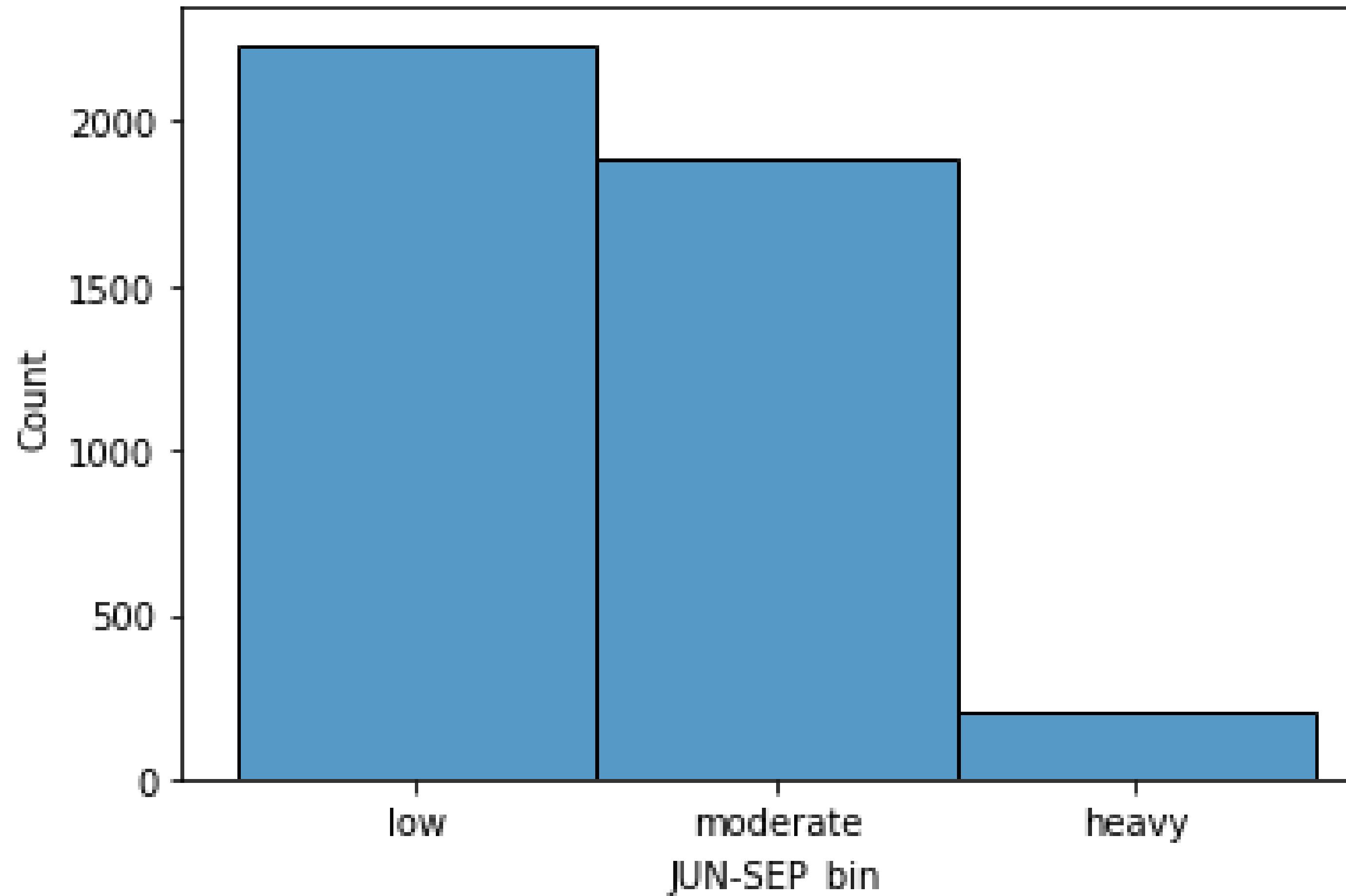
Min: 29426.6 [1972]

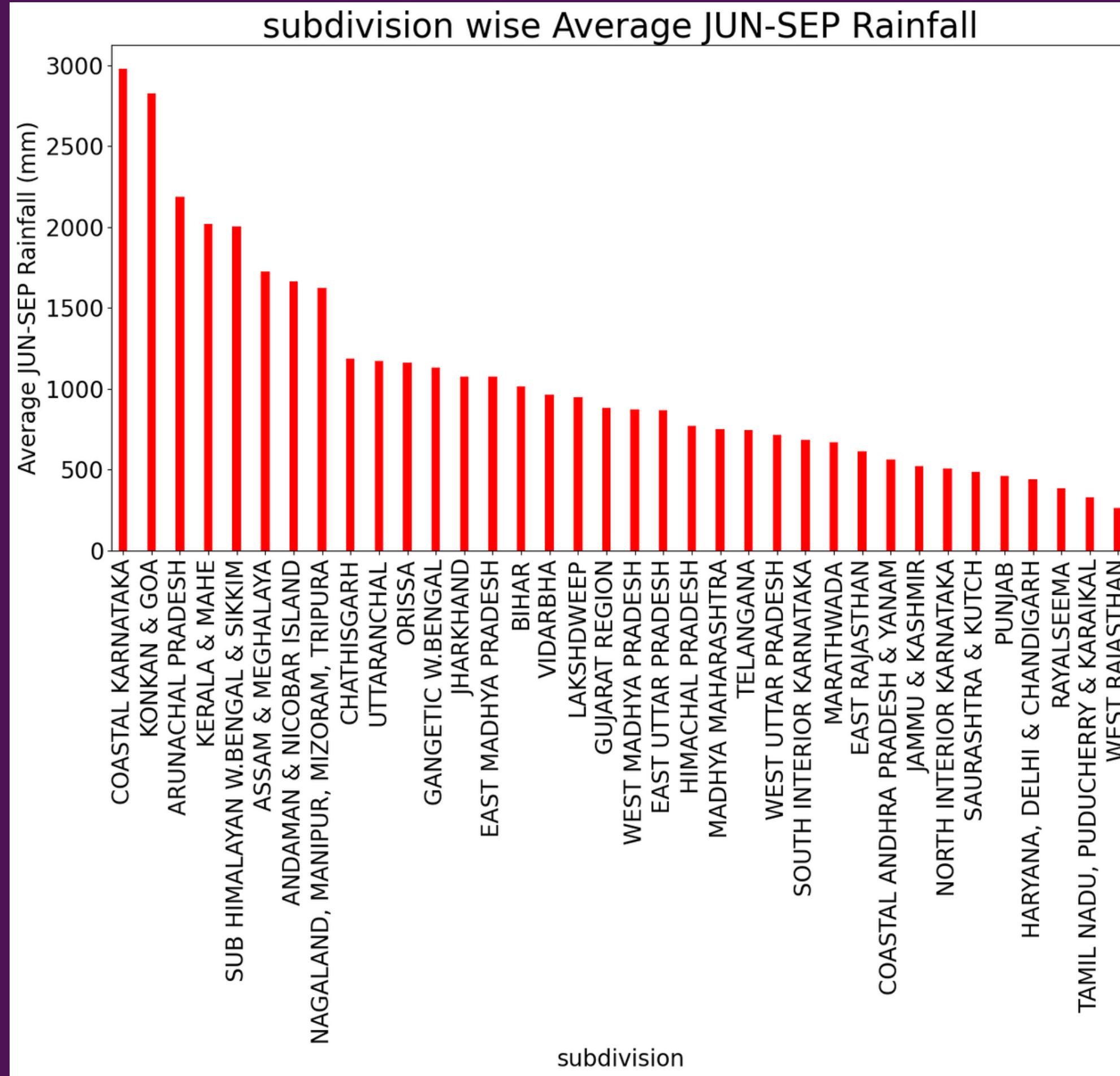
Mean: 37846

Time Frame:

1901 - 2021

Classification of rainfall:

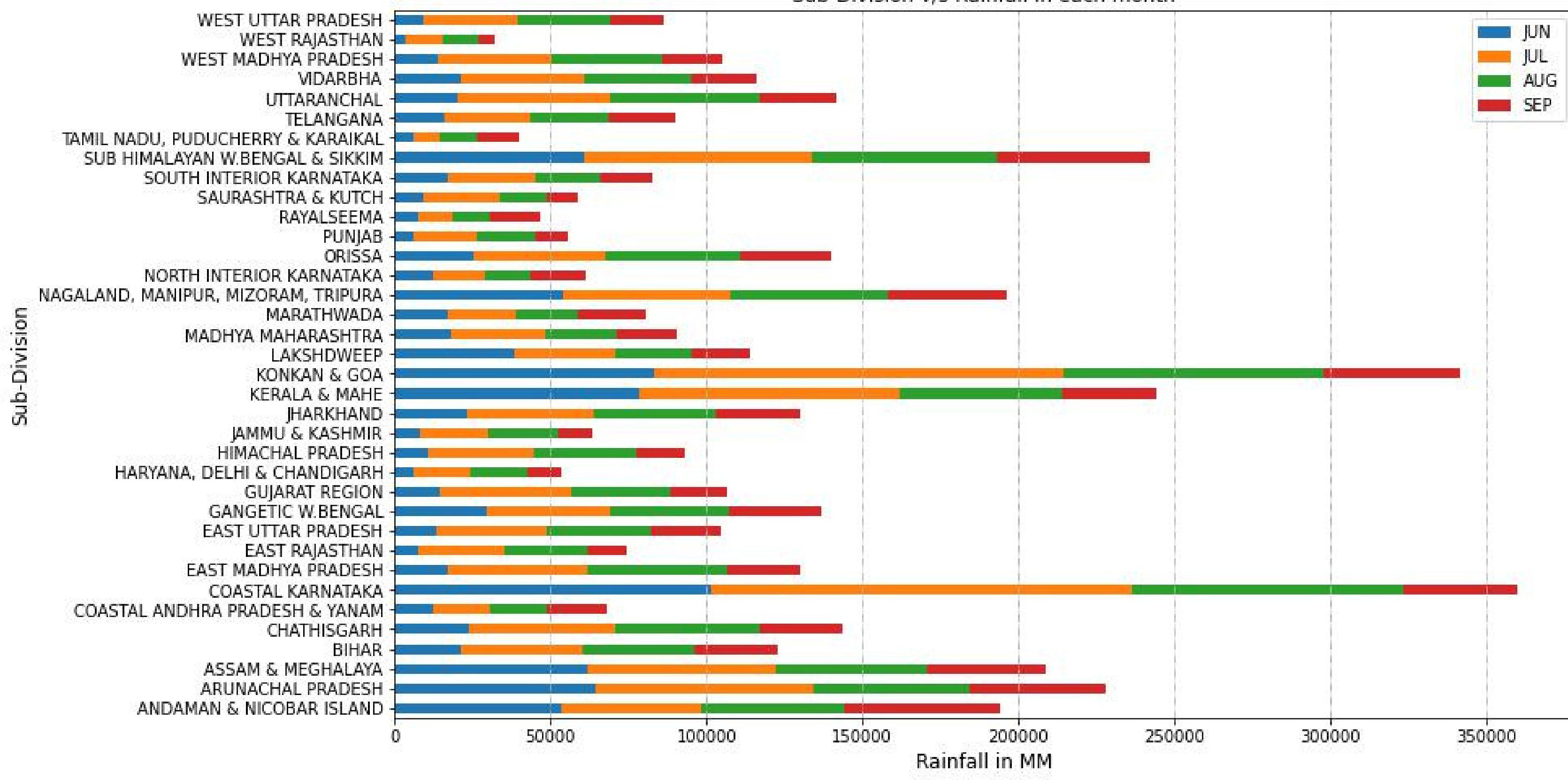


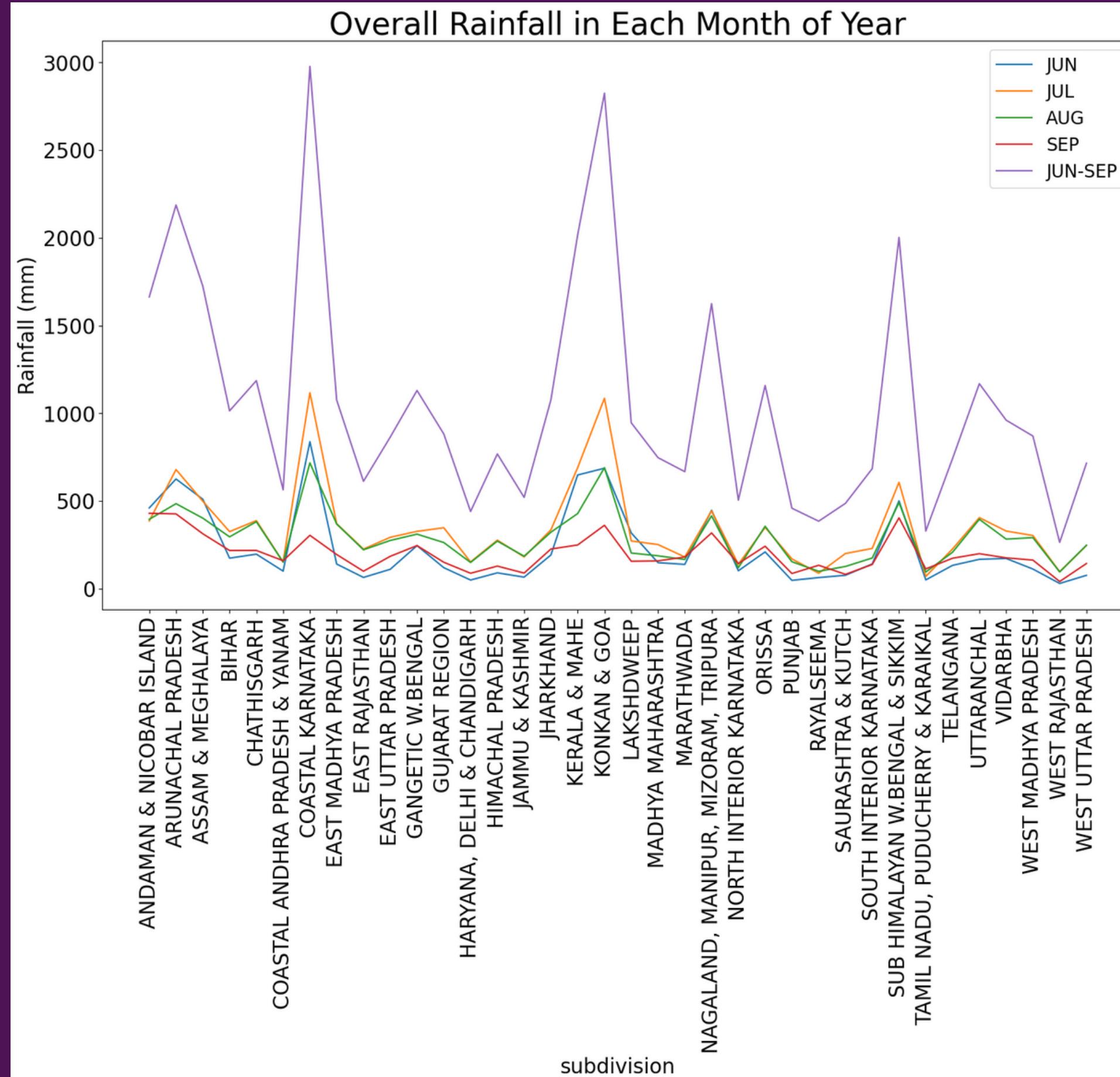


Area with least rainfall:
Western Rajasthan (292mm)
 [monsoon branch enters through Gujarat which runs parallel to the Aravalli ranges]

Area with most rainfall:
Coastal Karnataka (3408mm)
 [due to its location near the Western Ghats mountain range and its exposure to two monsoons.]

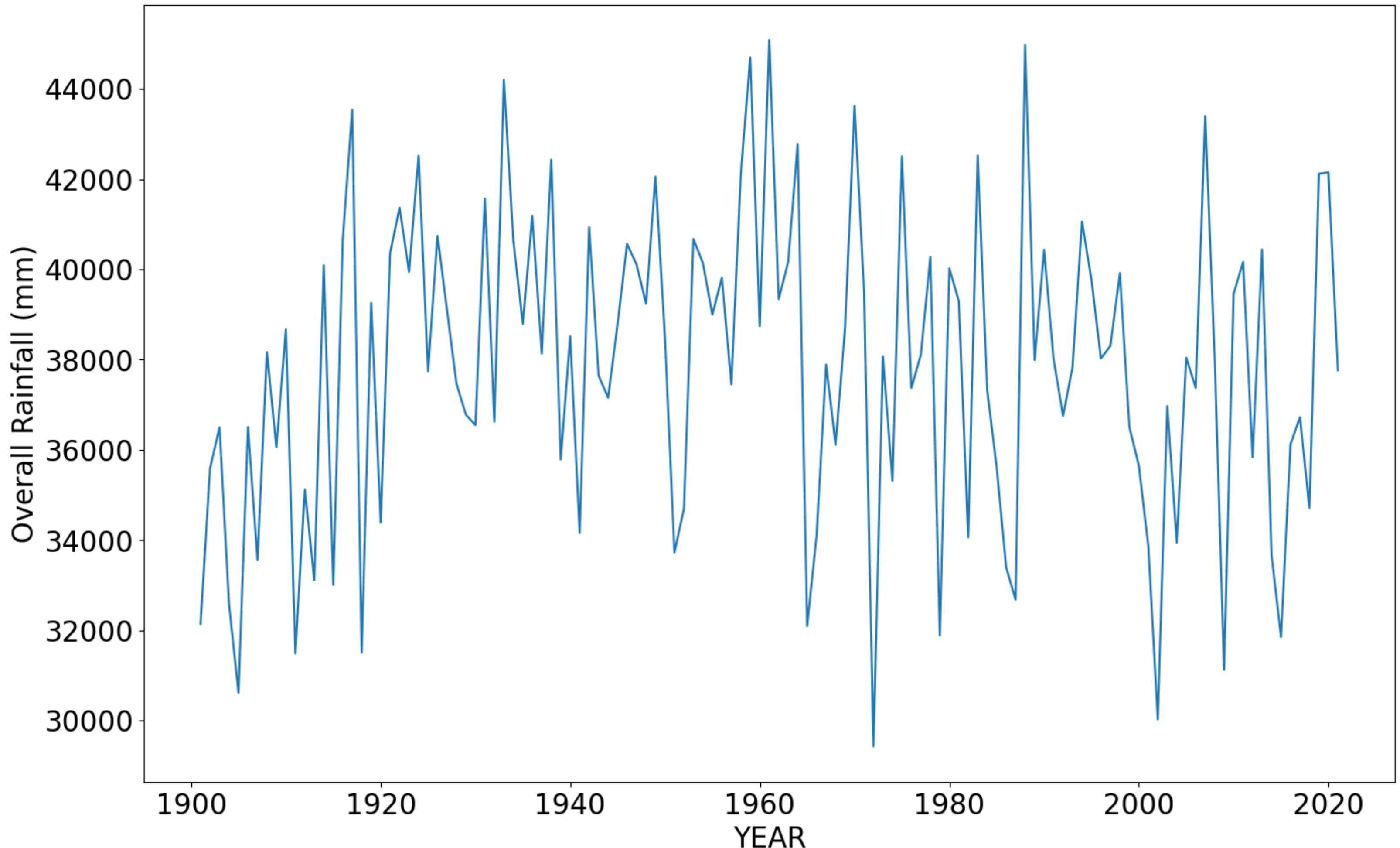
Sub-Division v/s Rainfall in each month





It can be observed
that most of the
subdivision receive
maximum rainfall in
the month of July.

Overall Rainfall in Each Year



Tech Stack

Data Cleaning

We used pandas for basic data cleaning like importing dataset, dropping values and/or columns and grouping subdivisions using pandas's functions like *read_csv*, *dropna* and *groupby*.

Data Visualisation

We used python's *Matplotlib* library to generate various types of plots to get a better understanding of the data and gain insights about the data.
We also Used *Tableau* to create to a visualisation dashboard to represent a few corelation in our dataset.

Data Preprocessing

Linear Regression

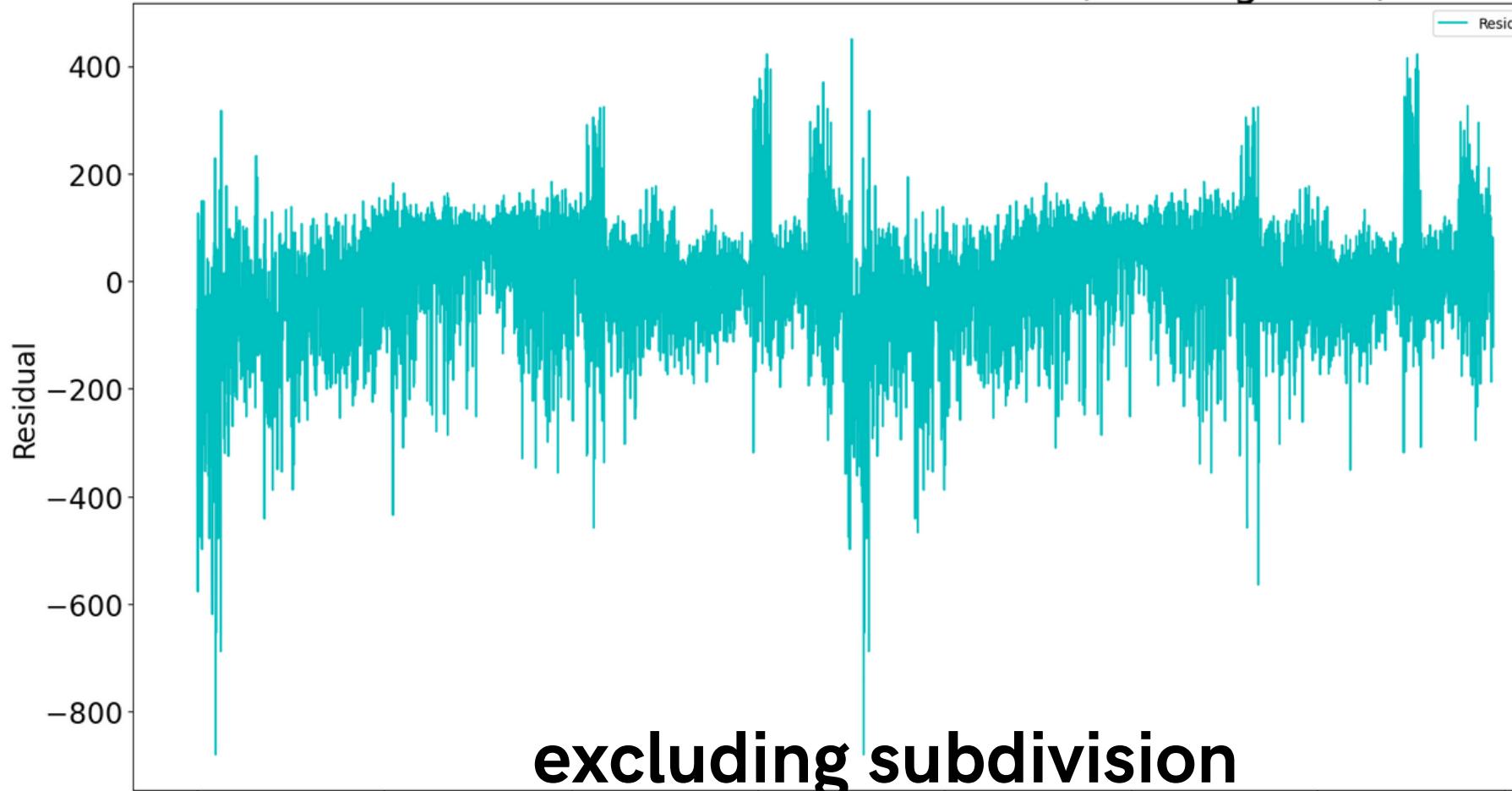
Tech Stack

We used some methods from sklearn's preprocessing class such as MinMaxScaler, SimpleImputer

We performed linear regression on the dataset. Calculated the r value using the given r squared value in order to see the strength of correlation that we had. Used the f score to predict if the data that we got had statistical significance.

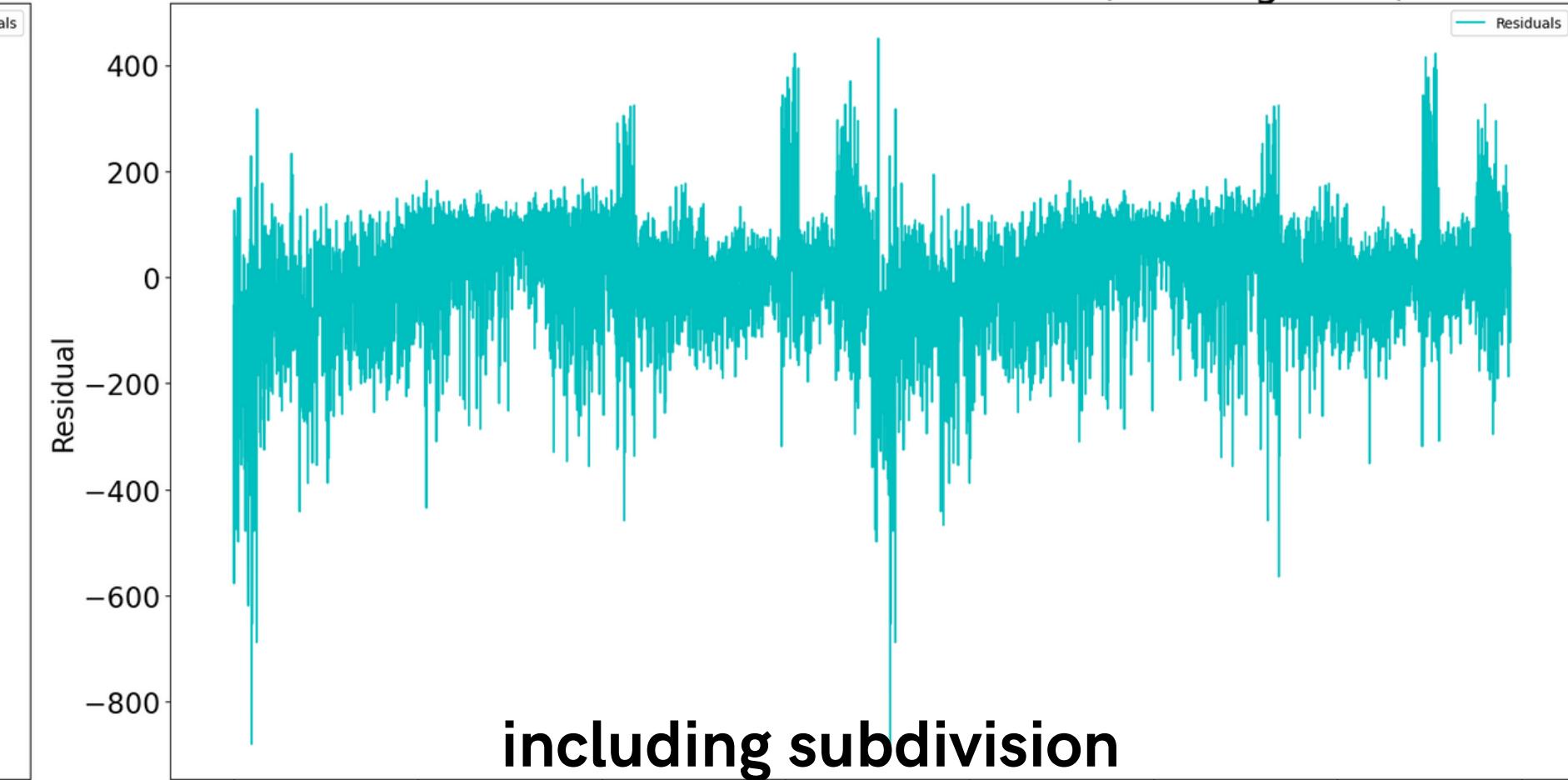
Training Accuracy: 43.9
Testing Accuracy: 32.5

Difference b/w Actual and Predicted (Training Data)



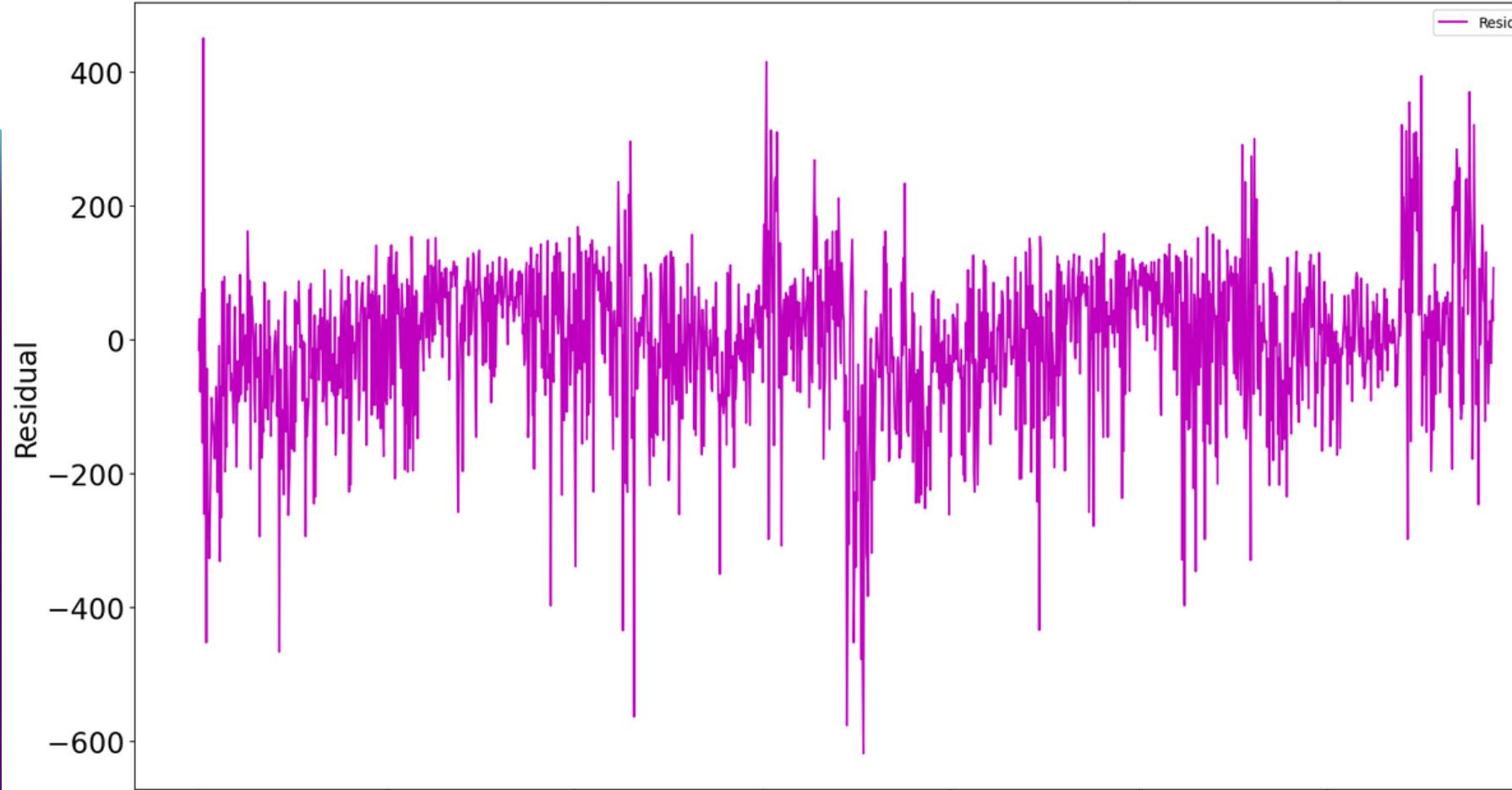
excluding subdivision

Difference b/w Actual and Predicted (Training Data)

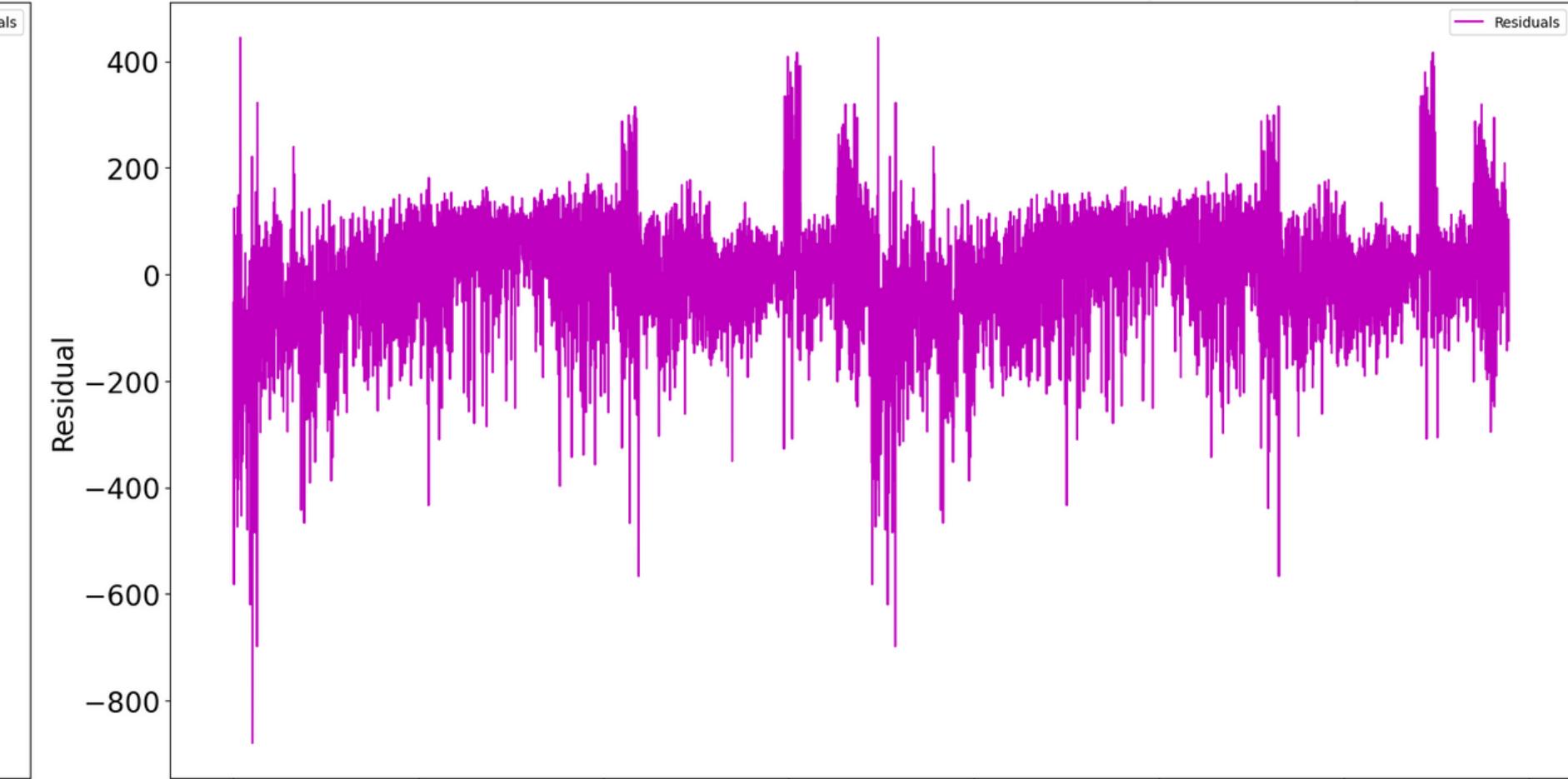


including subdivision

Difference b/w Actual and Predicted (Test Data)



Difference b/w Actual and Predicted (Test Data)



Tech Stack

Lasso Model

Training Accuracy: 28.2
Testing Accuracy: 27.6

SVM Model

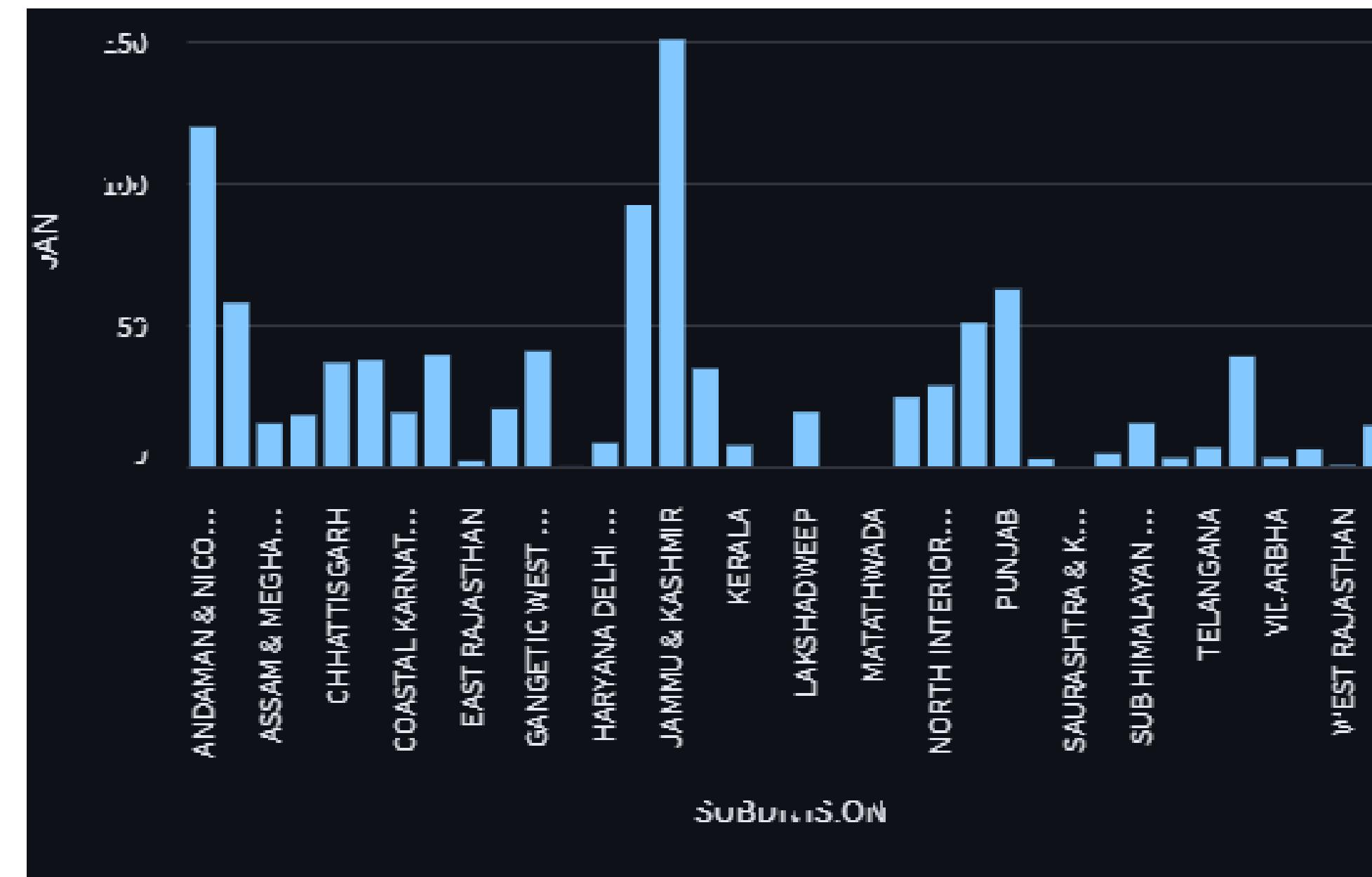
Training Accuracy: 10
Testing Accuracy: 11.6

Random
Forest Model

Training Accuracy: 80-95 %
Testing Accuracy: 75-85 %

As you can see for yourself, after utilizing various models random forest appeared to be best performing.

Predictions -



Jan 2022 Data

Predictions -

Rainfall Prediction using RandomForest

Subdivision

ARUNACHAL PRADESH

Year

2022

Select Duration

Monthly

Annual

Moderate rainfall of 2634.93mm is predicted

Accuracy: 78.6000000000001%

Predictions -

Rainfall Prediction using RandomForest

Subdivision

ARUNACHAL PRADESH

Year

2023

Select Duration

- Monthly
- Annual

Moderate rainfall of 2620.54mm is predicted

Website

<https://rainfall-prediction-sigfest.streamlit.app/>

Future Plans

RNN - LSTM

Long Short Term Memory networks are a special kind of Recurrent Neural Network, capable of learning long-term dependencies.

We also tried using RNN to make an artificial neural network as it does well with sequential data, but in our testing it had a high variance, which might be due to our limited dataset.

To continue this analysis further, we would try adding features to our dataset, while at the same time implementing LSTM, to reduce errors.

Thank You