**SPRINT 1**

|  |  |
| --- | --- |
| Team ID | PNT2022TMID24764 |
| Project Name | Exploratory Analysis Of Rainfall Data In India For Agriculture |

**DATA READ AND PREPROCESSING**

df = pd.read\_csv(r"C:/Users/ARUN/Downloads/rainfall.csv")

df = df.fillna(df.mean())

df.info()

OUTPUT:

<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 4116 non-null float64

3 FEB 4116 non-null float64

4 MAR 4116 non-null float64

5 APR 4116 non-null float64

6 MAY 4116 non-null float64

7 JUN 4116 non-null float64

8 JUL 4116 non-null float64

9 AUG 4116 non-null float64

10 SEP 4116 non-null float64

11 OCT 4116 non-null float64

12 NOV 4116 non-null float64

13 DEC 4116 non-null float64

14 ANNUAL 4116 non-null float64

15 Jan-Feb 4116 non-null float64

16 Mar-May 4116 non-null float64

17 Jun-Sep 4116 non-null float64

18 Oct-Dec 4116 non-null float64

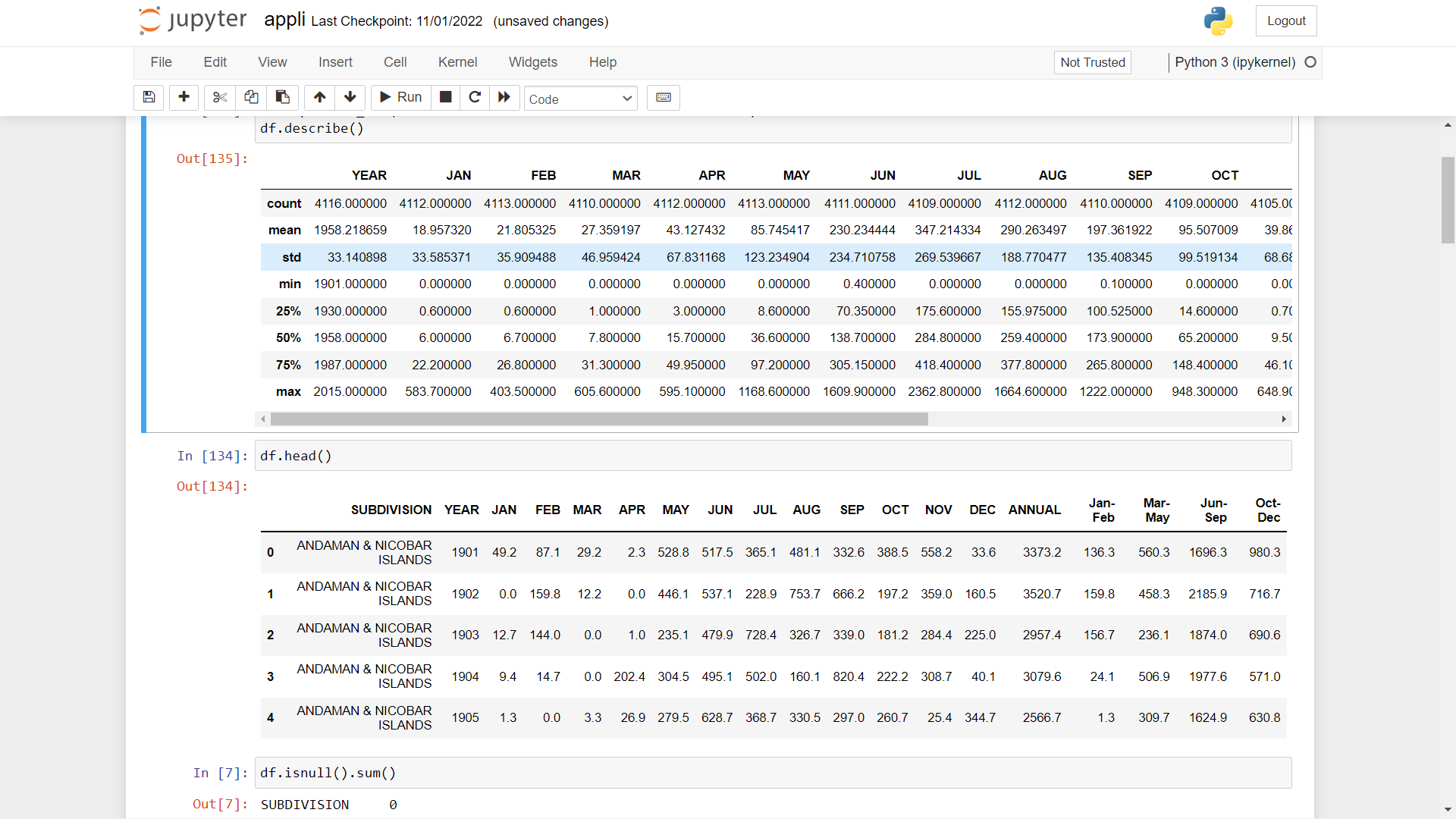
dtypes: float64(17), int64(1), object(1)

memory usage: 611.1+ KB

df.head()

df.describe()

**OUTPUT:**



**PREPROCESSING THE DATASET:**

**TO CHECK FOR NULL VALUES AND FILLING THEM:**

df.isnull().sum()

SUBDIVISION 0

YEAR 0

JAN 4

FEB 3

MAR 6

APR 4

MAY 3

JUN 5

JUL 7

AUG 4

SEP 6

OCT 7

NOV 11

DEC 10

ANNUAL 26

Jan-Feb 6

Mar-May 9

Jun-Sep 10

Oct-Dec 13

dtype: int64

df=df.fillna(df.mean(numeric\_only=True).round(1))

**DATA VISUALIZATION**

1)

df[["SUBDIVISION","ANNUAL"]].groupby("SUBDIVISION").sum().sort\_values(by='ANNUAL',ascending=False).plot(kind='barh',stacked=True,figsize=(18,15))

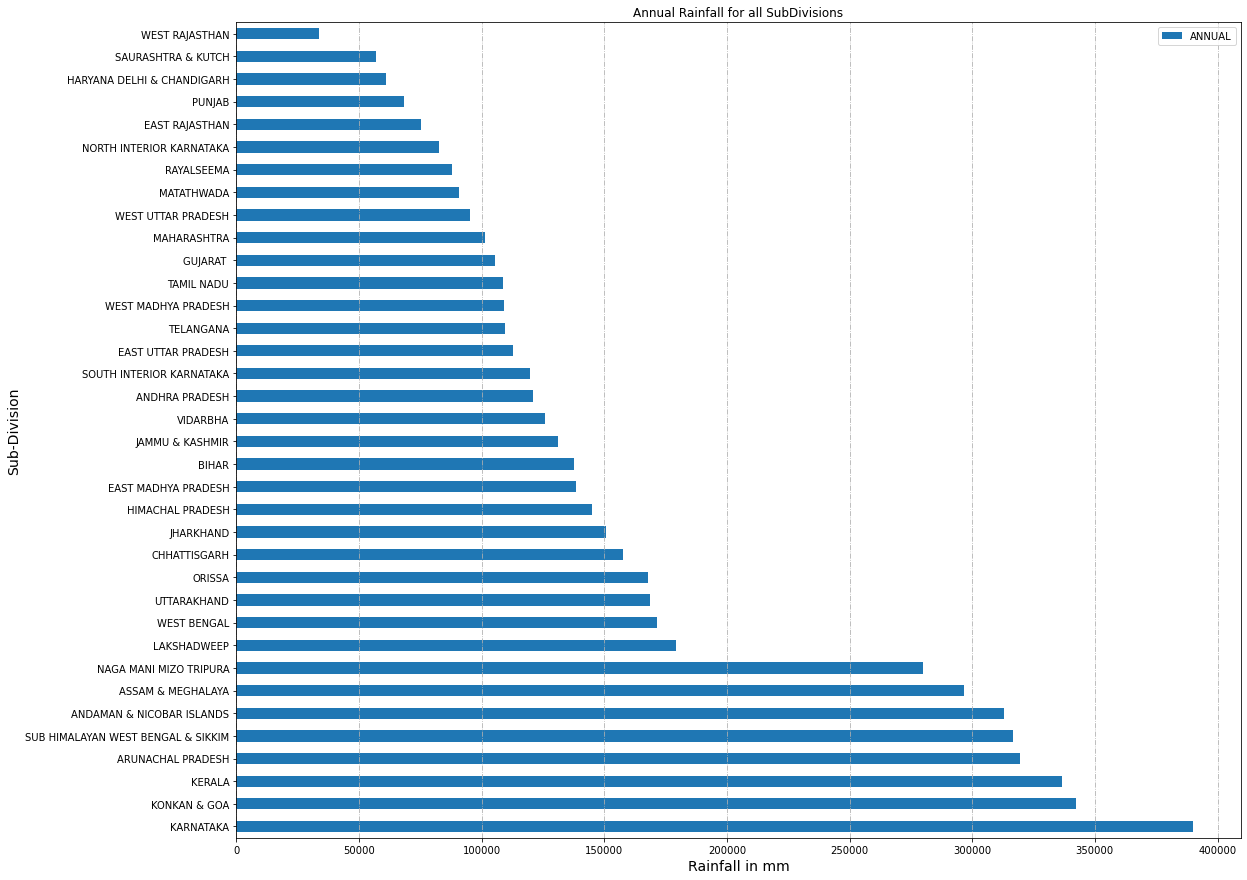
plt.xlabel("Rainfall in mm",size=14)

plt.ylabel("Sub-Division",size=14)

plt.title("Annual Rainfall for all SubDivisions")

plt.grid(axis="x",linestyle="-.")

plt.show()



2)

plt.figure(figsize=(15,8))

df.groupby("YEAR").sum()['ANNUAL'].plot(kind="line",color="g",marker="\*")

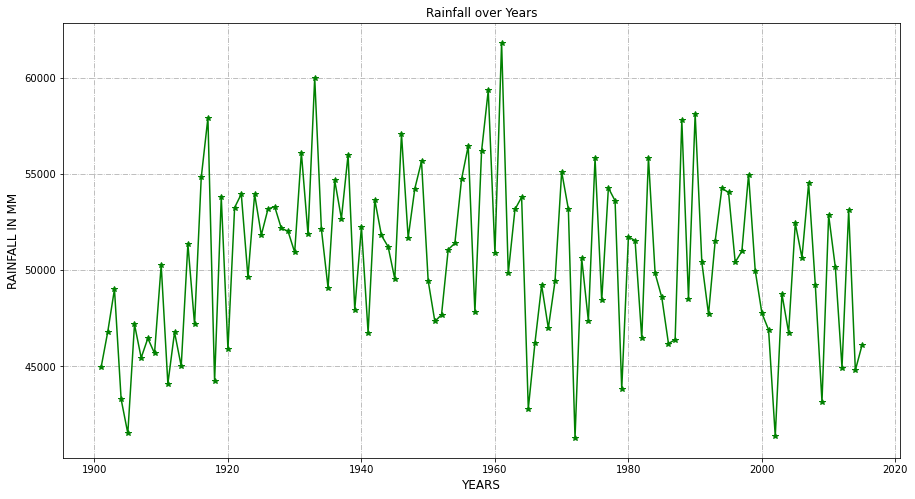
plt.xlabel("YEARS",size=12)

plt.ylabel("RAINFALL IN MM",size=12)

plt.grid(axis="both",linestyle="-.")

plt.title("Rainfall over Years")

plt.show()



3)

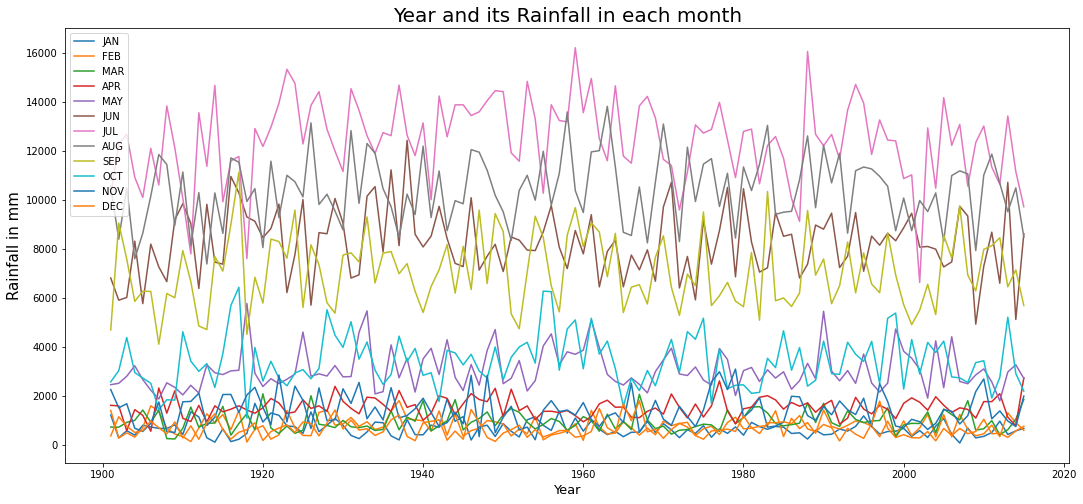
df[['YEAR', 'JAN', 'FEB', 'MAR', 'APR', 'MAY', 'JUN', 'JUL','AUG', 'SEP',

      'OCT', 'NOV', 'DEC']].groupby("YEAR").sum().plot(kind="line",figsize=(18,8))

plt.xlabel("Year",size=13)

plt.ylabel("Rainfall in mm",size=15)

plt.title("Year and its Rainfall in each month",size=20)

lt.show()

4)

plt.figure(figsize=(16,6),dpi=80)

plt.xticks(rotation=90)

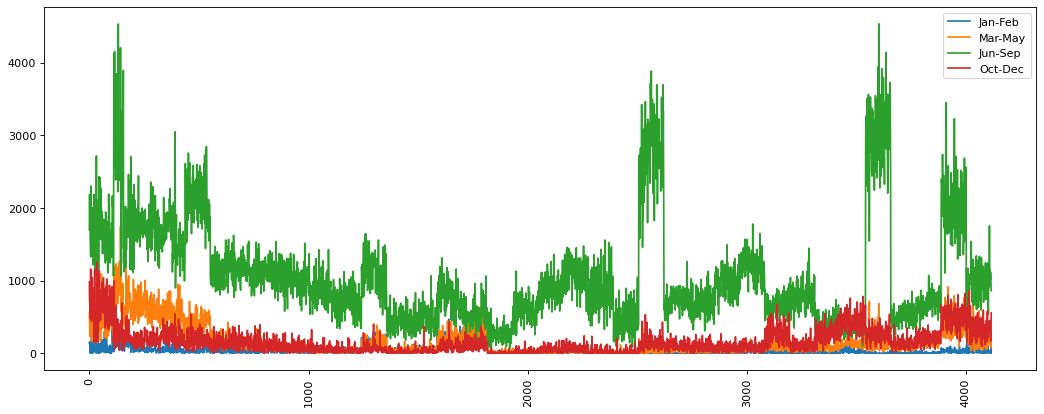
plt.plot(df['Jan-Feb'],label='Jan-Feb')

plt.plot(df['Mar-May'],label='Mar-May')

plt.plot(df['Jun-Sep'],label='Jun-Sep')

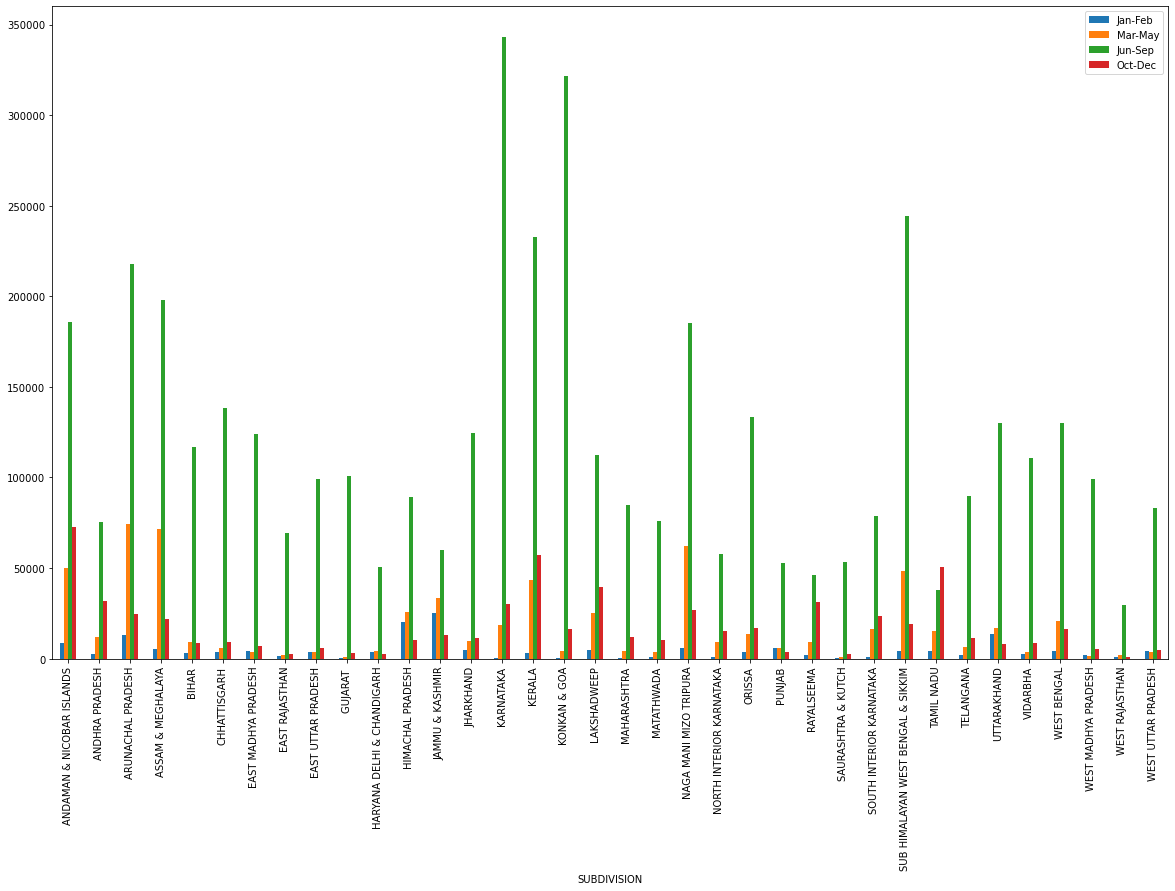
plt.plot(df['Oct-Dec'],label='Oct-Dec')

plt.legend(loc='best')



5)

ax = df[['SUBDIVISION', 'Jan-Feb', 'Mar-May','Jun-Sep', 'Oct-Dec']].groupby("SUBDIVISION").sum().plot.bar(stacked=False,figsize=(20,12))



6)

df[['SUBDIVISION', 'JAN', 'FEB', 'MAR', 'APR', 'MAY', 'JUN', 'JUL',

       'AUG', 'SEP', 'OCT', 'NOV', 'DEC']].groupby("SUBDIVISION").sum().plot(kind="barh",stacked=True,figsize=(13,8))

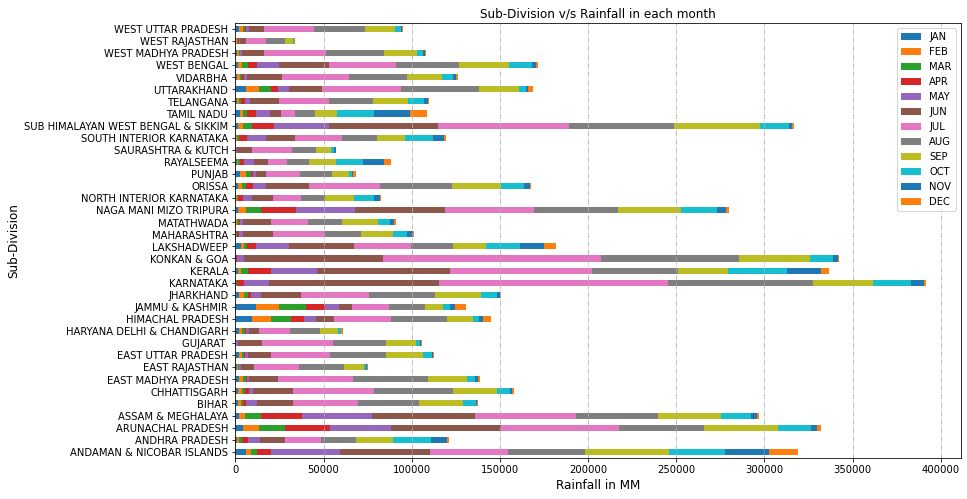
plt.title("Sub-Division v/s Rainfall in each month")

plt.xlabel("Rainfall in MM",size=12)

plt.ylabel("Sub-Division",size=12)

plt.grid(axis="x",linestyle="-.")

plt.show()



7)

#Highest rainfall receiving regions

plt.figure(figsize=(15,8))

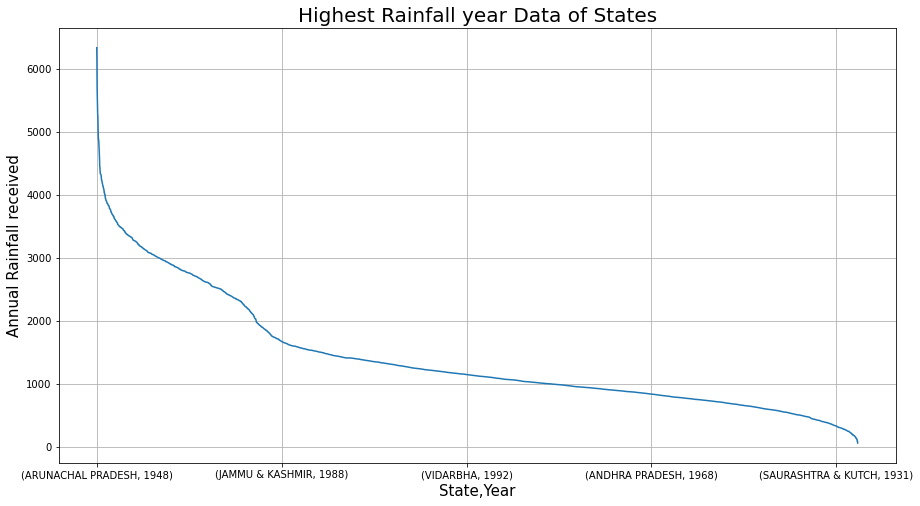
df.groupby(['SUBDIVISION','YEAR'])['ANNUAL'].sum().sort\_values(ascending=False).plot()

plt.grid()

plt.xlabel("State,Year",fontsize=15)

plt.ylabel("Annual Rainfall received",fontsize=15)

plt.title('Highest Rainfall year Data of States',fontsize=20)



8)

#Month with highest rainfall

plt.figure(figsize=(10,7))

df[['JAN', 'FEB', 'MAR', 'APR', 'MAY', 'JUN', 'JUL', 'AUG',

       'SEP', 'OCT', 'NOV', 'DEC']].mean().plot(kind= 'bar', color='red')

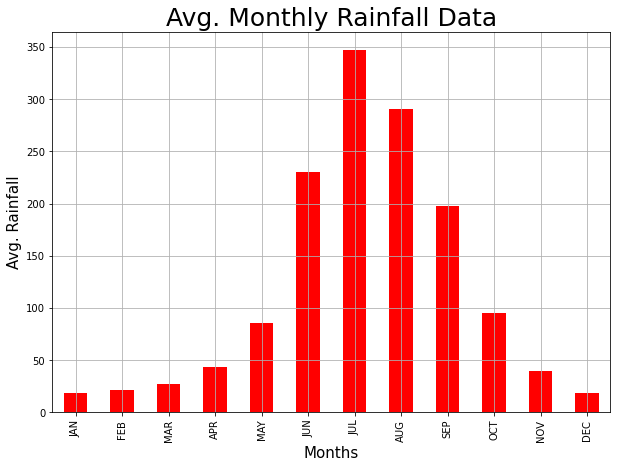
plt.xlabel('Months',fontsize=15)

plt.ylabel('Avg. Rainfall',fontsize=15)

plt.title('Avg. Monthly Rainfall Data',fontsize=25)

plt.grid()

plt.show()



These are the visualizations that we have made for our analysis of rainfall in India. These predictions show the highest rainfall region, average rainfall, the rainfall in each month with respect to the region, etc.