**import** numpy **as** np *# linear algebra*

**import** pandas **as** pd *# data processing, CSV file I/O (e.g. pd.read\_csv)*

**import** matplotlib.pyplot **as** plt

**import** seaborn **as** sns

**import** random

*# Input data files are available in the "../input/" directory.*

*# For example, running this (by clicking run or pressing Shift+Enter) will list all files under the input directory*

**import** os

**for** dirname, \_, filenames **in** os**.**walk('/kaggle/input'):

**for** filename **in** filenames:

print(os**.**path**.**join(dirname, filename))

In [7]:

data **=** pd**.**read\_csv('/content/rainfall\_India\_2017 (1).csv')**.**rename(columns**=**str**.**lower)

data**.**head(3)

Out[7]:

|  | **subdivision** | **year** | **jan** | **feb** | **mar** | **apr** | **may** | **jun** | **jul** | **aug** | **sep** | **oct** | **nov** | **dec** | **annual** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **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 |
| **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 |
| **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 |

In [5]:

data**.**info()

RangeIndex: 4188 entries, 0 to 4187

Data columns (total 15 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 subdivision 4188 non-null object

1 year 4188 non-null int64

2 jan 4184 non-null float64

3 feb 4185 non-null float64

4 mar 4182 non-null float64

5 apr 4184 non-null float64

6 may 4185 non-null float64

7 jun 4183 non-null float64

8 jul 4181 non-null float64

9 aug 4184 non-null float64

10 sep 4182 non-null float64

11 oct 4181 non-null float64

12 nov 4177 non-null float64

13 dec 4178 non-null float64

14 annual 4162 non-null float64

dtypes: float64(13), int64(1), object(1)

memory usage: 490.9+ KB

In [6]:

print('Dataset comprises of {} observations and {} characteristics'**.**format(data**.**shape[0],data**.**shape[1]))

print('\nUnique Values: ',data**.**nunique())

print('\nMissing Values: ',data**.**isna()**.**sum())

Dataset comprises of 4188 observations and 15 characteristics

Unique Values: subdivision 36

year 117

jan 808

feb 902

mar 989

apr 1247

may 1751

jun 2754

jul 3093

aug 2950

sep 2664

oct 1958

nov 1245

dec 810

annual 3770

dtype: int64

Missing Values: 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

dtype: int64