**Big Data Analysis with IBM Cloud Databases**

**PROBLEM DEFINITION:**

The project involves delving into big data analysis using IBM Cloud Databases. The objective is to extract valuable insights from extensive datasets, ranging from climate trends to social patterns. The project includes designing the analysis process, setting up IBM Cloud Databases, performing data analysis, and visualizing the results for business intelligence.

Let us see the solution for the above problem,

**STEP 1: DATA SELECTION**

This is the very fundamental step for every data analytics. You have to collect the dataset from various resources like Kaggle, Dataworld, Datahub, Big Query and check that dataset is suitable for our projects and it is relevant to our topic or ideas.

In our project they mentioned two popular dataset names. They are

* + - * Climate Data
      * Social Media Trends

Our wish is to choose the climate data topic and collect the dataset from the kaggle

DATASET SOURCE:

<https://www.kaggle.com/datasets/thedevastator/annual-subdivision-wise-rainfall-in-india-1901-2>

**STEP 2: DATABASE SETUP**

In this first we need to create an account in the IBM Cloud, put all the necessary details they want and create the account. Then only you put the dataset to the IBM Cloud and also take the dataset to virtualize and analyze it.

**Our View:** First, I will create the IBM Cloud account and put my dataset to the IBM Cloud Database.

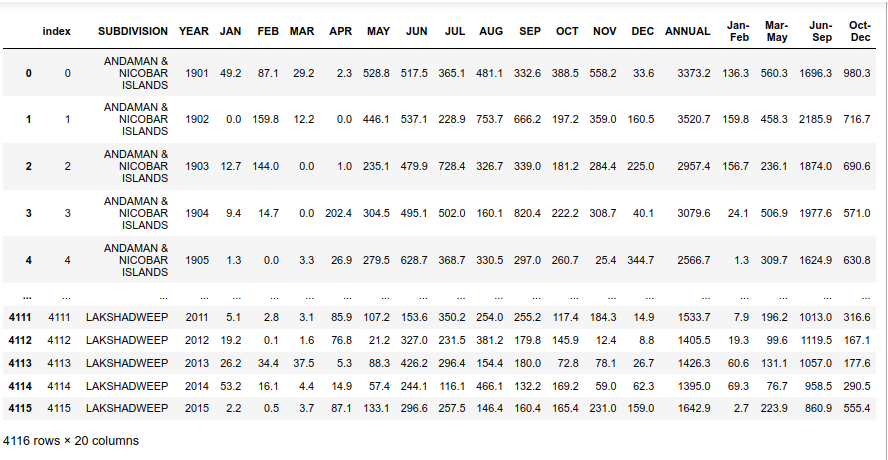
**STEP 3: DATA EXPLORATION**

In this step we explore the dataset using queries and scripts and drop the irrelevant data in our dataset.

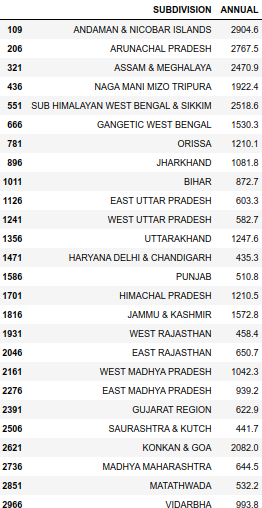
**Our View:**

I am using the python programming language to explore the data using the pandas library. The pandas library is mainly used to framing the data, exploring the data, analyzing the data and many features are there. I use this library to drop the irrelevant columns for my analysis and take the relevant columns.

**BEFORE DROPING:**



**AFTER DROPING:** We got only two columns what we visualize.



**STEP 4: ANALYSIS TECHNIQUES**

In this step we need to apply statistical analysis or machine learning techniques to uncover the insights.

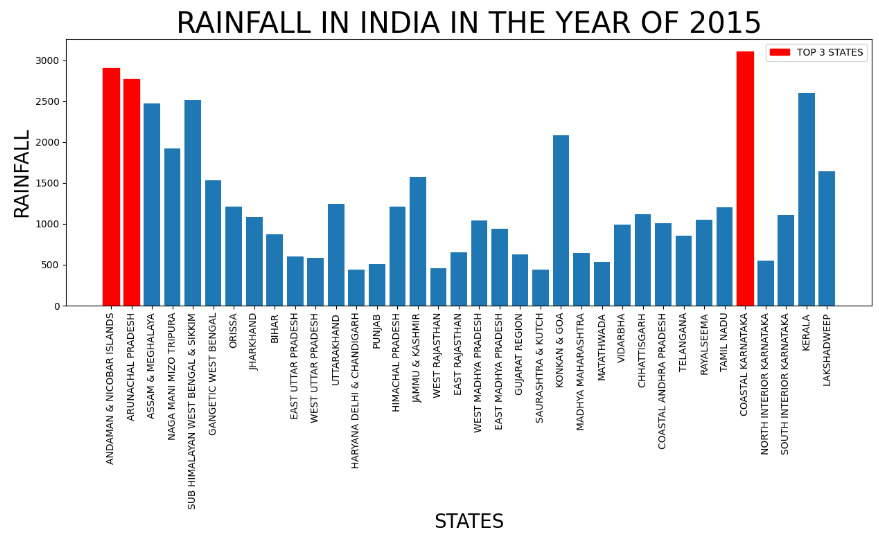
**Our View:** But in this step, I don’t go to any model I only virtualize and understand the model.

**STEP 5: VISUALIZATION**

Using python, we have to use the visualize the dataset to find the insights and trends to understand the dataset and conclude our result.

**Our View:**

I am using the matplotlib library to visualize the dataset and understand the dataset and conclude the result.



**STEP 6: BUSINESS INSIGHTS (or) CONCLUSION**

According to the above graph we conclude that the coastal Karnataka, Arunachal Pradesh and Andaman and Nicobar have the highest rainfall in the year of 2015.

These are all the steps to follow and find the insights in the dataset.