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:

	index	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL	Jan-Feb	Mar-May	Jun-Sep	Oct-Dec
0	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	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	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	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
4	4	ANDAMAN & NICOBAR ISLANDS	1905	1.3	0.0	3.3	26.9	279.5	628.7	368.7	330.5	297.0	260.7	25.4	344.7	2566.7	1.3	309.7	1624.9	630.8
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4111	4111	LAKSHADWEEP	2011	5.1	2.8	3.1	85.9	107.2	153.6	350.2	254.0	255.2	117.4	184.3	14.9	1533.7	7.9	196.2	1013.0	316.6
4112	4112	LAKSHADWEEP	2012	19.2	0.1	1.6	76.8	21.2	327.0	231.5	381.2	179.8	145.9	12.4	8.8	1405.5	19.3	99.6	1119.5	167.1
4113	4113	LAKSHADWEEP	2013	26.2	34.4	37.5	5.3	88.3	426.2	296.4	154.4	180.0	72.8	78.1	26.7	1426.3	60.6	131.1	1057.0	177.6
4114	4114	LAKSHADWEEP	2014	53.2	16.1	4.4	14.9	57.4	244.1	116.1	466.1	132.2	169.2	59.0	62.3	1395.0	69.3	76.7	958.5	290.5
4115	4115	LAKSHADWEEP	2015	2.2	0.5	3.7	87.1	133.1	296.6	257.5	146.4	160.4	165.4	231.0	159.0	1642.9	2.7	223.9	860.9	555.4

4116 rows × 20 columns

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

	SUBDIVISION	ANNUAL
109	ANDAMAN & NICOBAR ISLANDS	2904.6
206	ARUNACHAL PRADESH	2767.5
321	ASSAM & MEGHALAYA	2470.9
436	NAGA MANI MIZO TRIPURA	1922.4
551	SUB HIMALAYAN WEST BENGAL & SIKKIM	2518.6
666	GANGETIC WEST BENGAL	1530.3
781	ORISSA	1210.1
896	JHARKHAND	1081.8
1011	BIHAR	872.7
1126	EAST UTTAR PRADESH	603.3
1241	WEST UTTAR PRADESH	582.7
1356	UTTARAKHAND	1247.6
1471	HARYANA DELHI & CHANDIGARH	435.3
1586	PUNJAB	510.8
1701	HIMACHAL PRADESH	1210.5
1816	JAMMU & KASHMIR	1572.8
1931	WEST RAJASTHAN	458.4
2046	EAST RAJASTHAN	650.7
2161	WEST MADHYA PRADESH	1042.3
2276	EAST MADHYA PRADESH	939.2
2391	GUJARAT REGION	622.9
2506	SAURASHTRA & KUTCH	441.7
2621	KONKAN & GOA	2082.0
2736	MADHYA MAHARASHTRA	644.5
2851	MATATHWADA	532.2
2966	VIDARBHA	993.8

STEP 4: ANALYSIS TECHNIQUES

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

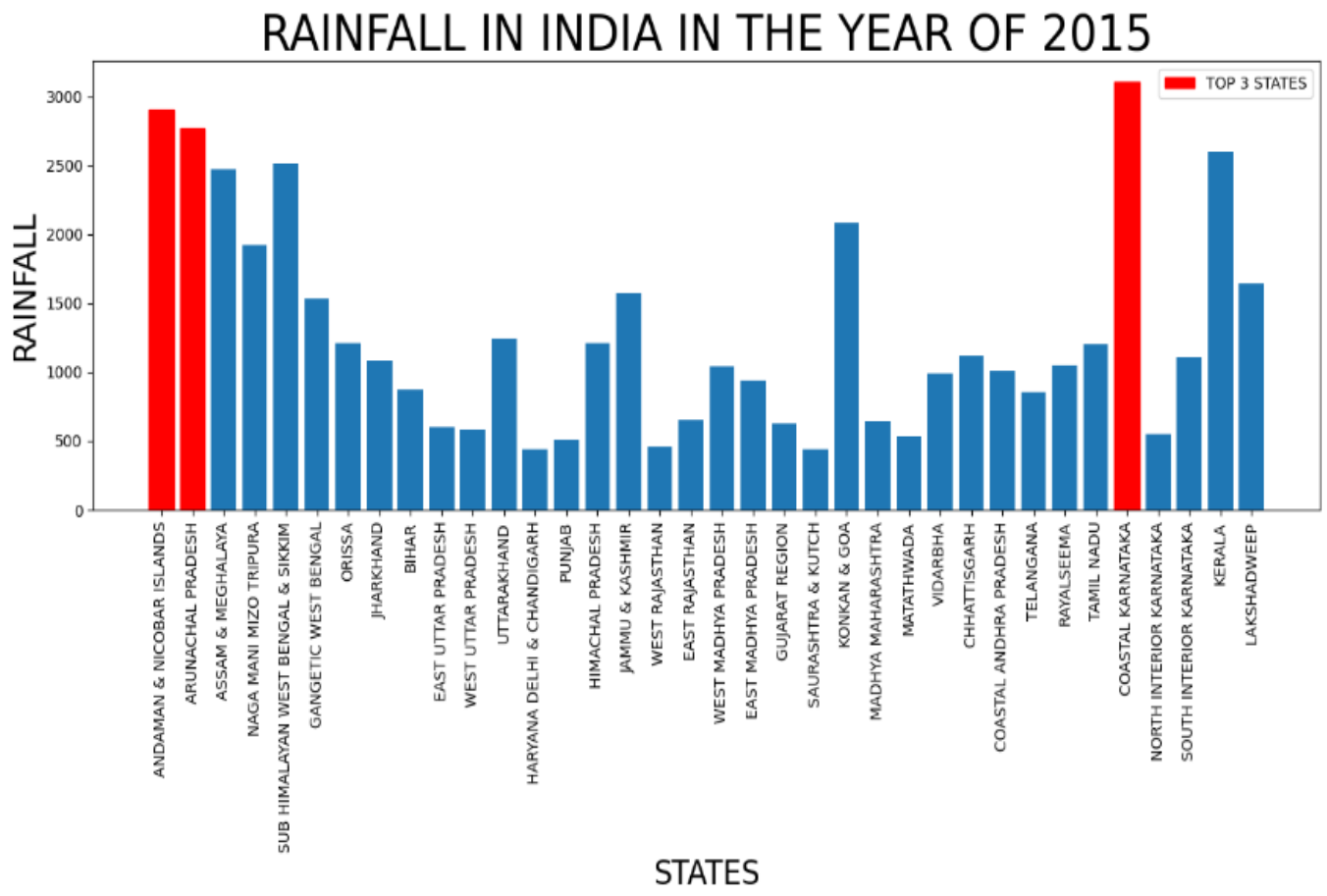
Our View: But in this step, I don't go to any model I only virtualize and understand the model.

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.