Daily Rainfall Data in District

Project Overview and Objective:

To analyze daily rainfall data to identify maximum and minimum of rainfall with days. To identify district – wise difference in rainfall and visualize rainfall data using tools like Excel and Power BI. The **objective of the rainfall dataset** is to **analyse historical rainfall data** to understand rainfall patterns, seasonal variations, and extreme events, and to **support better planning and decision-making** in areas such as water management, agriculture, and disaster preparedness.

Data Source:

Source Description and Timeline: India data Portal

Source Link: https://indiadataportal.com/p/climate-data/r/wris-daily_rainfall_district-dt-dl-imd

Dataset Topic: Daily Rainfall Dataset in District-wise (2007 to 2024)

Format: CSV file

Problem Statement:

- The purpose of this project is to analyze daily rainfall data to understand rainfall patterns, seasonal behavior and extreme events using data analytics tools.
- The insights obtained can support better planning in agriculture, water management and disaster mitigation.

Attribute (Column /Features) Details:

Attribute Name	Data type	Description
Date	Date (DD/MM/YY)	Represents the day when rainfall was measured in each district.
State Code	Numeric (Integer)	Identify states uniquely

Daily Rainfall Data in District

State Name	String (Text)	Name of the States
District Code	Numeric (Integer0	Identify district uniquely
District Name	String (Text)	Name of the District
Actual Rainfall (millimetre)	Number (Decimal)	Measured using rain Gauges (or) Shows the observed Rainfall
Rainfall Storage	Number (Decimal)	The quantity of rainwater stored in water bodies (or) ground water system during a specific period
Normal Rainfall	Number (Decimal)	The long-term average rainfall for a specific location and period, based on historical records
Percentage Deviation from Normal	Number (Decimal)	Indicates excess or deficit rainfall, help understand weather anomalies

Tools and Technologies:

- **Excel:** Data cleaning, transformation, and Pivot Tables.
- **Power BI:** Data modelling, DAX calculations, visualization, and interactive dashboard creation.

Data Pre-Processing (Excel / Power Query):

Task Performed:

- Data Cleaning and Transformation
- Remove Duplicates, handle missing values on Actual, Storage , Normal Rainfall Column Using **AVERAGE** function.
- State name and District name using **TRIM** and **CLEAN, PROPER** function
- Percentage Deviation from Normal column using formula:

  Percentage Deviation = (Actual Rainfall – Normal Rainfall)/Normal Rainfall * 100

Daily Rainfall Data in District

- **Filtering and Sorting:** Organized data to focus on Relevant (State name) Records.
- Using XLOOKUP function.
- Convert to data into **one fact table** and **two dimensions table** (State and District).
- **Pivot Tables:**

  State vs Count of Actual

  State vs sum of Normal (Date)

  Date(yr) vs District Name
- **Descriptive Statistics:**

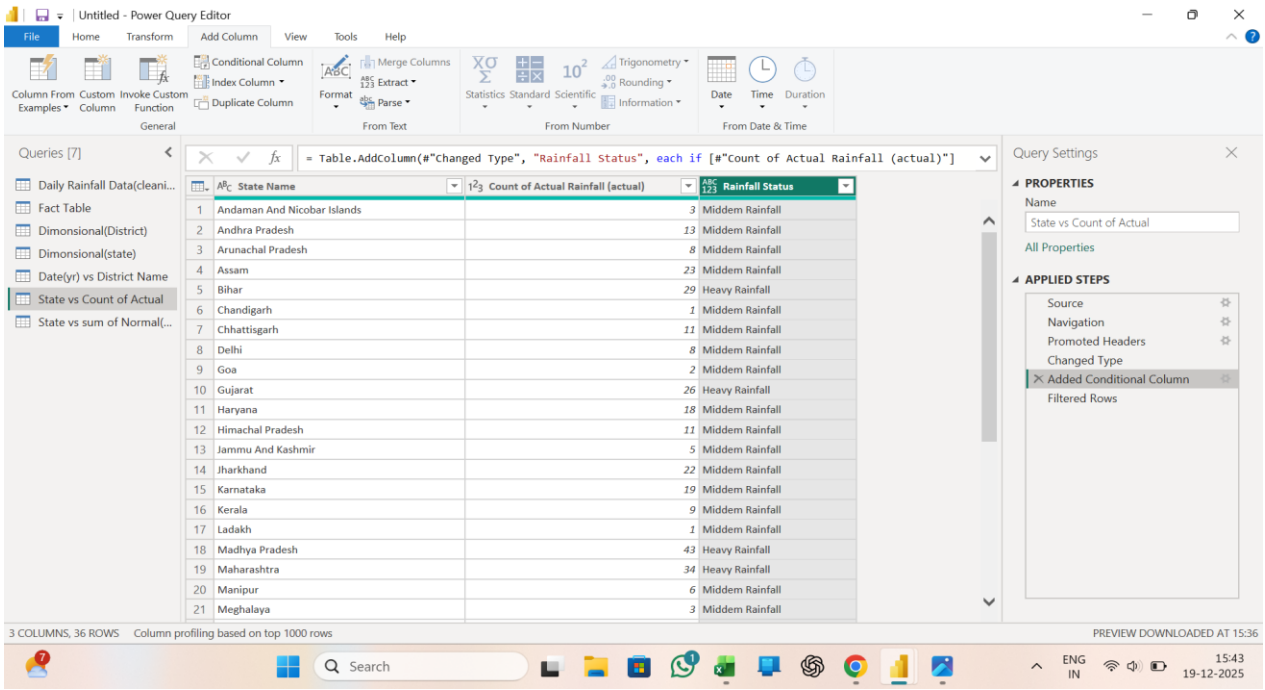
  Descriptive (Normal Rainfall)

  Descriptive (Rainfall Storage)

Data Modelling and DAX (Power BI):

Date model:

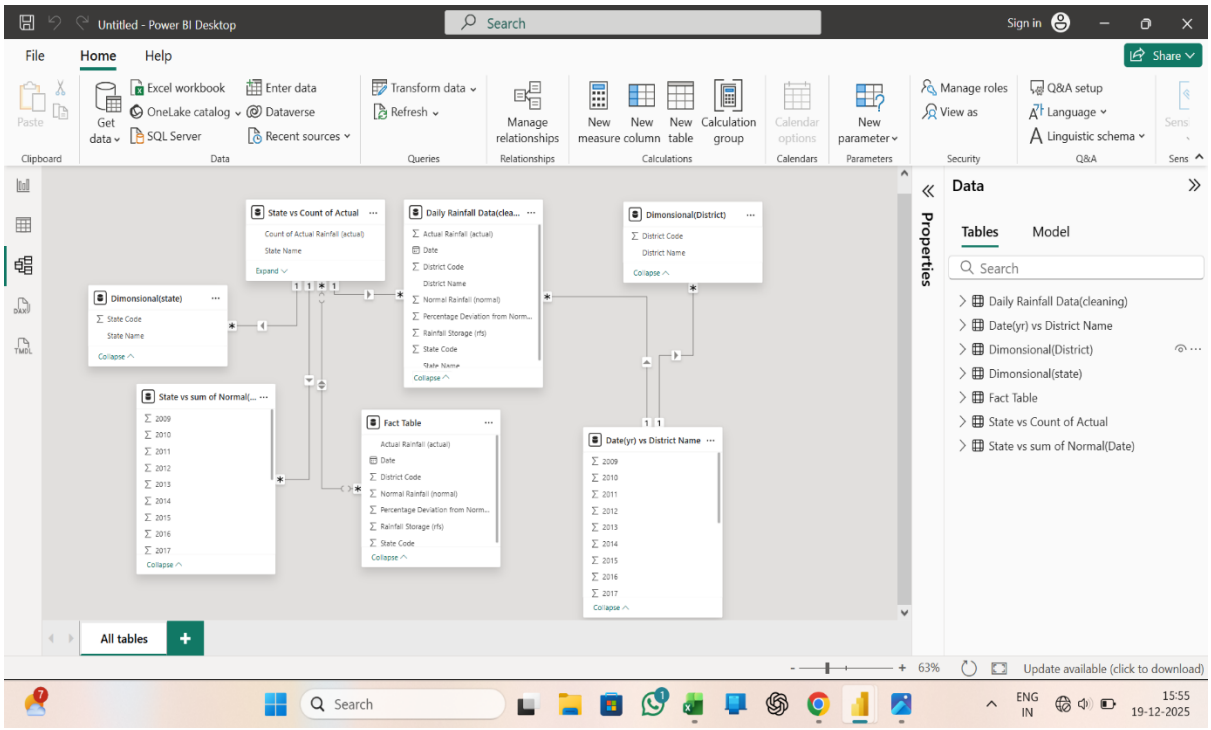
- Get date → from excel → Select Daily Rainfall (cleaning), Date (yr) vs District Name, Dimensional (State), Dimensional (District), Fact Table, State vs Sum of Normal (Date) and State vs Count of Actual Rainfall → load and Transform Data
- Select Date(yr) vs District Name →change type → Go to transform→use first row in headers
- Select State vs Count of Actual Rainfall →Add column→Conditional column→ New column→Name: **Rainfall Status**



Daily Rainfall Data in District

- **Mange Relationship:**

Goto model view → Mange relationship→use **one to many, many to many** functions.



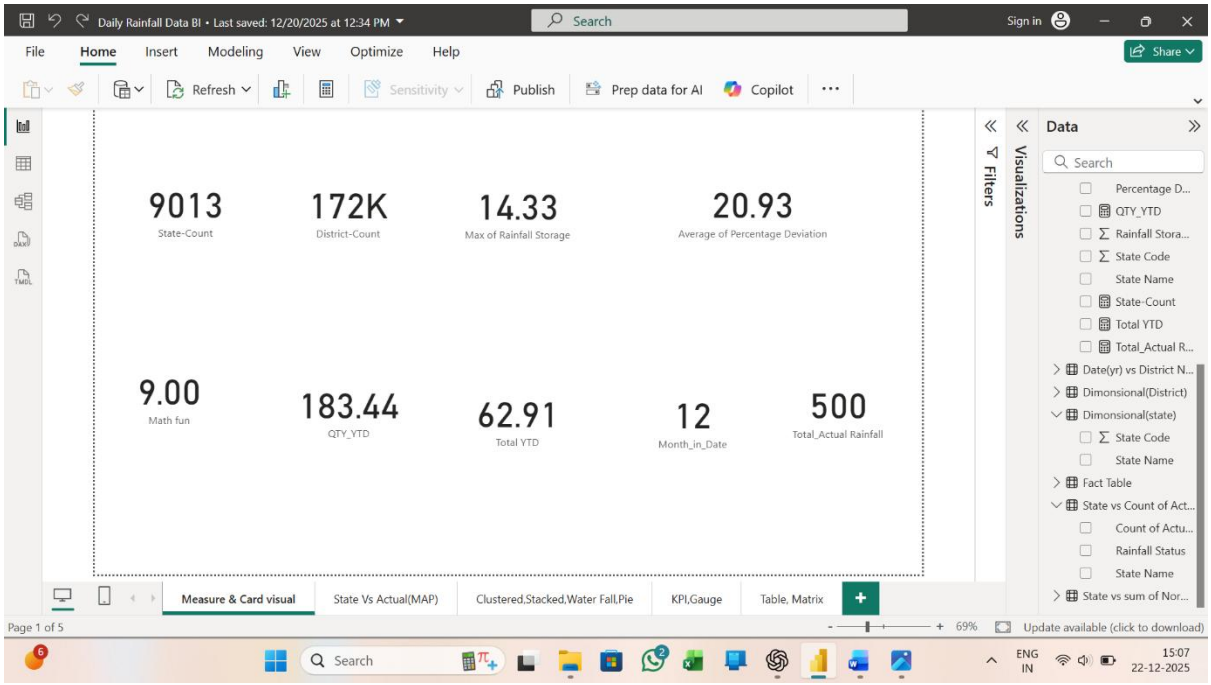
Calculated New Measure Formula's:

Goto Report view→ Select Daily rainfall data(cleaning)→ Table tools→New measure→Formula bar using formulas:

- State count→" State-Count = sum ('Daily Rainfall Data(cleaning)'[State Code])"
- District count→" District-Count = sum ('Daily Rainfall Data(cleaning)'[District Code])"
- Max of Rainfall storage→" Max of Rainfall Storage = MAX ('Daily Rainfall Data(cleaning)'[Rainfall Storage (rfs)])"
- Average of percentage Deviation→" Average of Percentage Deviation = AVERAGEA ('Daily Rainfall Data(cleaning)'[Percentage Deviation from Normal (deviation)])"
- Mathematical function→" Math fun = power (3,2)"
- QTY-YTD Function→" QTY_YTD = CALCULATE (sum ('Daily Rainfall Data(cleaning)'[Rainfall Storage (rfs)]),'Daily Rainfall Data(cleaning)'[Date])"
- Total-YTD →" Total YTD = TOTALYTD ('Daily Rainfall Data(cleaning)'[Average of Percentage Deviation],'Daily Rainfall Data(cleaning)'[Date])"

Daily Rainfall Data in District

- Month-in-date→" Month in date = MONTH ("23-12-2020")"
- Total actual rainfall→" Total Actual Rainfall = COUNT ('Daily Rainfall Data(cleaning)'[Actual Rainfall (actual)])"

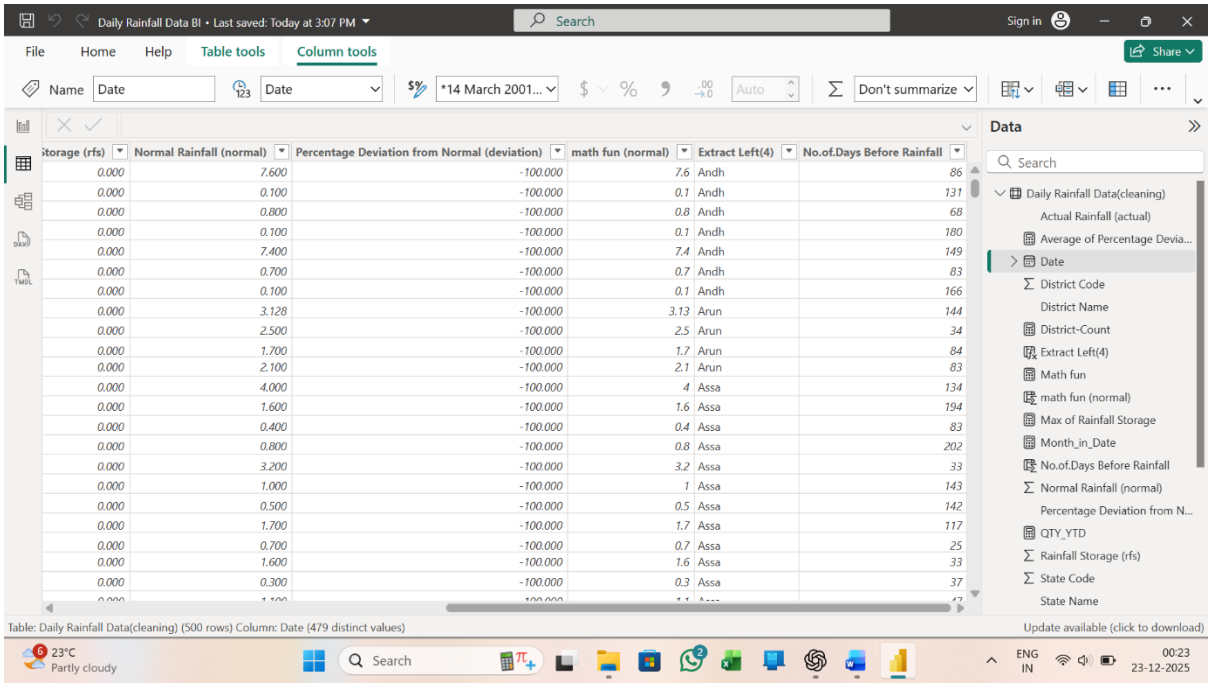


Calculated New column Formula's:

Goto Report view→select Table daily rainfall data(cleaning)→Table tools→New column→Formula bar using Formulas:

- Mathematical function (Normal Rainfall):" math fun(normal) = ROUND('Daily Rainfall Data(cleaning)'[Normal Rainfall (normal)],2)"
- Extract Left (4):" Extract Left (4) = left ('Daily Rainfall Data(cleaning)'[State Name],4)"
- Number of days before Rainfall (Date):" No of Days Before Rainfall = DATEDIFF ('Daily Rainfall Data(cleaning)'[Date], TODAY (), MONTH)"

Daily Rainfall Data in District



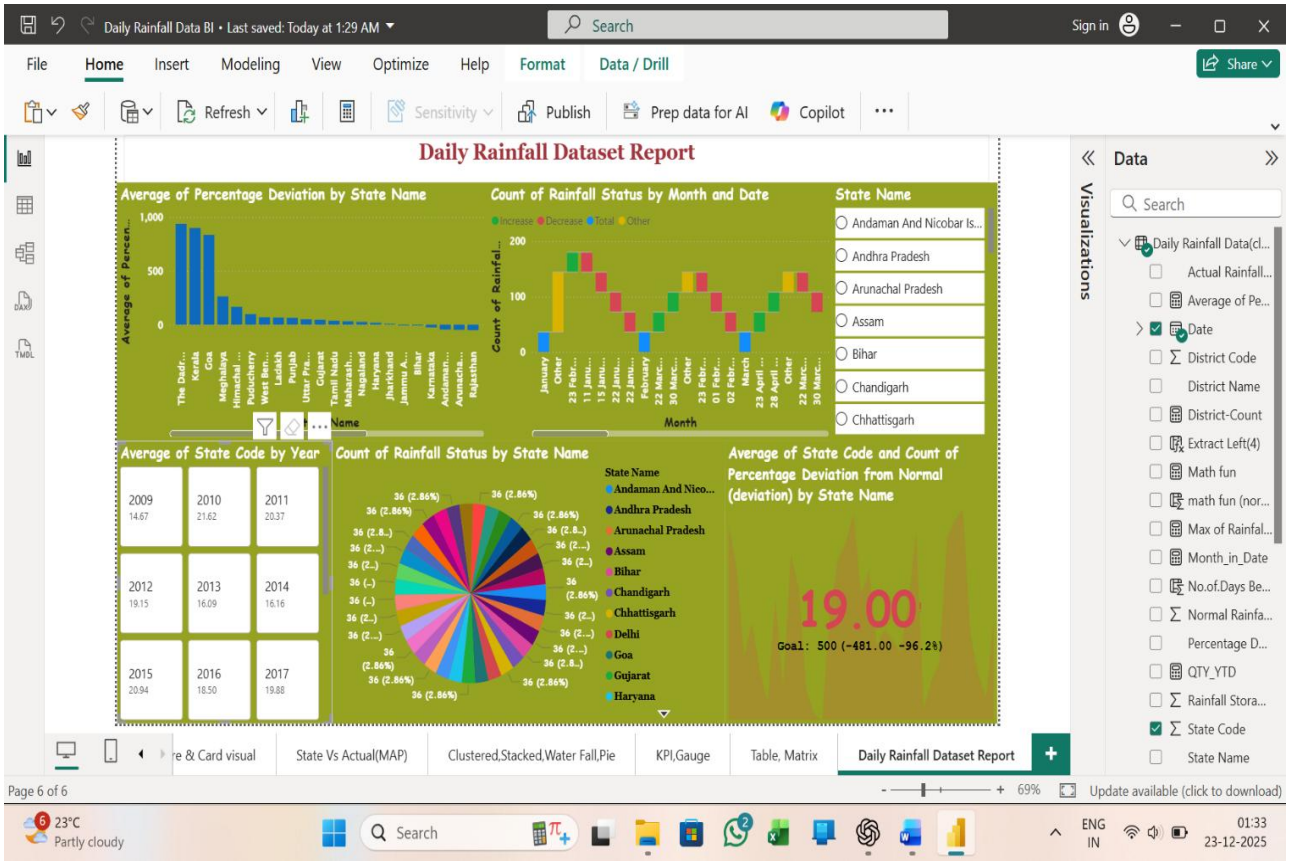
Storage (rfs)	Normal Rainfall (normal)	Percentage Deviation from Normal (deviation)	math fun (normal)	Extract Left(4)	No.of.Days Before Rainfall
0.000	7.600	-100.000	7.6	Andh	86
0.000	0.100	-100.000	0.1	Andh	131
0.000	0.800	-100.000	0.8	Andh	68
0.000	0.100	-100.000	0.1	Andh	180
0.000	7.400	-100.000	7.4	Andh	149
0.000	0.700	-100.000	0.7	Andh	83
0.000	0.100	-100.000	0.1	Andh	166
0.000	3.128	-100.000	3.13	Arun	144
0.000	2.500	-100.000	2.5	Arun	34
0.000	1.700	-100.000	1.7	Arun	84
0.000	2.100	-100.000	2.1	Arun	83
0.000	4.000	-100.000	4	Assa	134
0.000	1.600	-100.000	1.6	Assa	194
0.000	0.400	-100.000	0.4	Assa	83
0.000	0.800	-100.000	0.8	Assa	202
0.000	3.200	-100.000	3.2	Assa	33
0.000	1.000	-100.000	1	Assa	143
0.000	0.500	-100.000	0.5	Assa	142
0.000	1.700	-100.000	1.7	Assa	117
0.000	0.700	-100.000	0.7	Assa	25
0.000	1.600	-100.000	1.6	Assa	33
0.000	0.300	-100.000	0.3	Assa	37

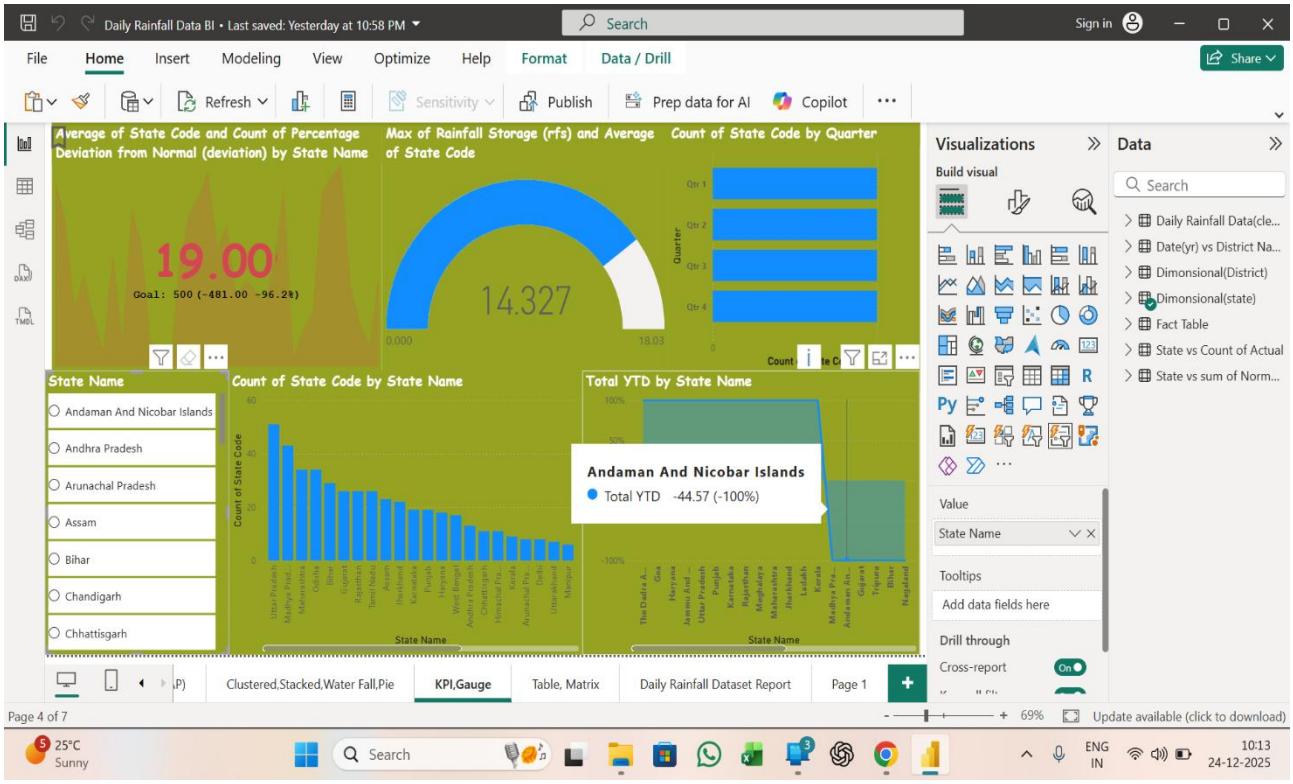
Analysis and Visualizations (Power BI):

Multiple Visualizations based on problem statement:

- State vs Actual Rainfall using Map visual
- Average of percentage deviation by State name using clustered chart
- Count of rainfall status by month and date using waterfall chart
- Count of rainfall status by state name using pie chart
- This visual using button slicer
- Average of state code and count of percentage deviation from normal by state name using KPI chart
- Max of Rainfall storage and average of State code using Gauge
- Count of state code by quarter, month and day using Clustered bar chart
- This visual using List slicer and using bookmark
- Create Table and Matrix using Button slicer and drill through
- Finally Create **Daily Rainfall Data Report.**

Daily Rainfall Data in District





Daily Rainfall Data in District

Insights and Conclusions:

Key Findings:

- Highest rainfall is observed during the monsoon season (June–September).
- Lowest rainfall occurs during summer months (March–May).
- Rainfall peaks are **strongly influenced by seasonal patterns**, especially the southwest monsoon.
- **Flood preparedness measures** are required in high rainfall districts.
- Farmers should **plan crop cycles based on predicted rainfall trends.**
- To understand high actual rainfall distribution over **time (05/October/2009)** and **location (Maharashtra/Mumbai Suburban).**
- To identify normal rainfall levels (**0.00**) and extreme values (**29.800).**
- Compare high-rainfall (27/July/2016) and (Maharashtra/Ratnagiri), actual rainfall (6.910) but normal rainfall (29.800).
- Number of rainy days **212**
- Number of dry days **288** (0 mm rainfall)

Water Resource Management

- Decide how much water to store or release
- Plan reservoir and dam operations

Flood Risk Management

- Identify high-risk periods and locations
- Plan preventive measures

Agricultural Planning

- Decide crop selection and sowing time
- Optimize irrigation schedules

Conclusions:

Overall, this study demonstrates that daily rainfall data is valuable for **weather forecasting, agricultural planning, water conservation, and disaster management**, supporting informed decision-making and sustainable planning. The analysis of the daily rainfall dataset reveals that rainfall patterns are **highly seasonal and strongly influenced by monsoon period.**