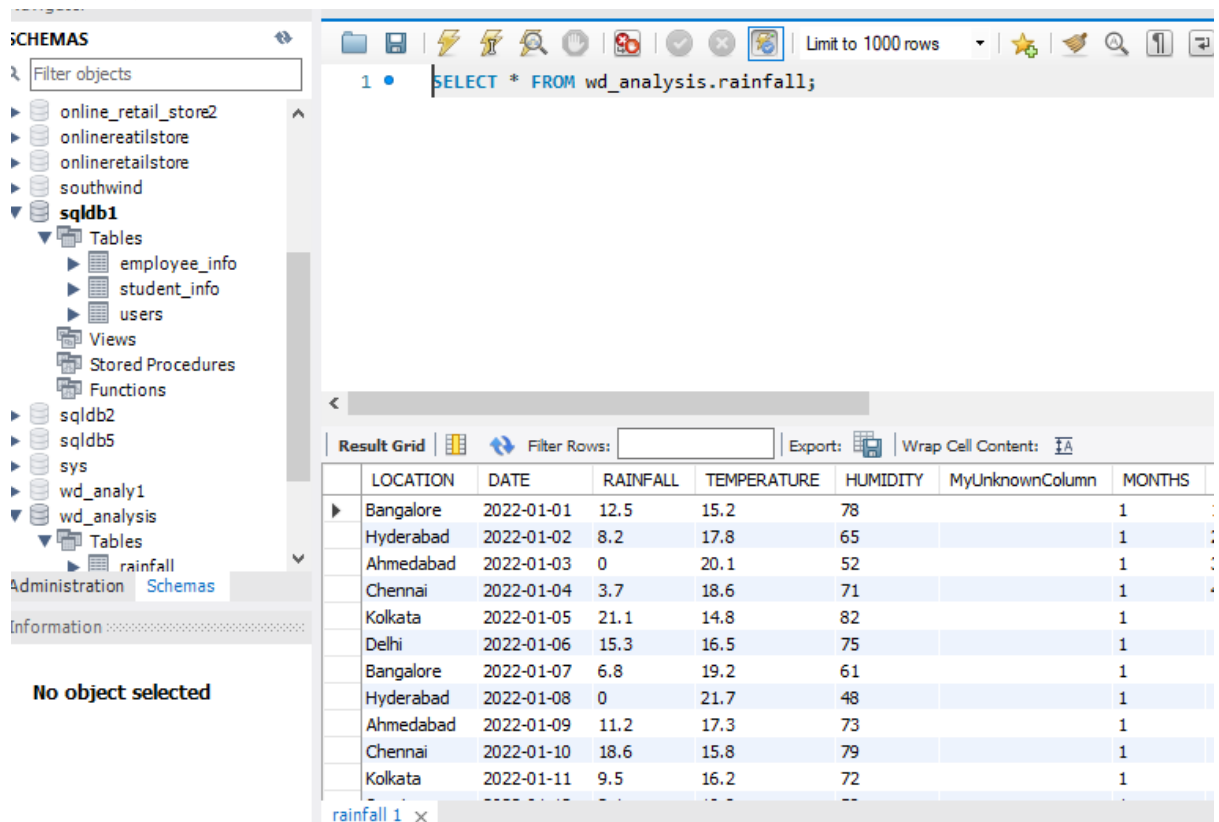


WEATHER DATA ANALYSIS

Weather data analysis allows us to understand patterns, trends, and anomalies in weather conditions over time. We will explore how to analyze weather data using python language. We will use a dataset containing various weather parameters such as region, date, temperature, rainfall, humidity.

➤ Data Collection(MYSQL)

We used rainfall.csv dataset containing date, temperature, rainfall, and humidity.



The screenshot shows a MySQL database interface. On the left, the 'SCHEMAS' panel lists various databases, including 'wd_analysis'. The main query editor displays the SQL query: `SELECT * FROM wd_analysis.rainfall;`. Below the query editor, the 'Result Grid' shows the data returned by the query. The data is presented in a table with columns: LOCATION, DATE, RAINFALL, TEMPERATURE, HUMIDITY, MyUnknownColumn, and MONTHS. The table contains 11 rows of data, showing weather information for various locations in India over a period of 11 days in January 2022.

LOCATION	DATE	RAINFALL	TEMPERATURE	HUMIDITY	MyUnknownColumn	MONTHS
Bangalore	2022-01-01	12.5	15.2	78		1
Hyderabad	2022-01-02	8.2	17.8	65		1
Ahmedabad	2022-01-03	0	20.1	52		1
Chennai	2022-01-04	3.7	18.6	71		1
Kolkata	2022-01-05	21.1	14.8	82		1
Delhi	2022-01-06	15.3	16.5	75		1
Bangalore	2022-01-07	6.8	19.2	61		1
Hyderabad	2022-01-08	0	21.7	48		1
Ahmedabad	2022-01-09	11.2	17.3	73		1
Chennai	2022-01-10	18.6	15.8	79		1
Kolkata	2022-01-11	9.5	16.2	72		1

Monthly data from the database using MYSQL

The screenshot displays two instances of the MySQL Workbench interface, each showing a query and its results in a Result Grid.

Top Instance:

- Query 1:** `#SELECT * FROM wd_analysis.rainfall;`
- Query 2:** `SELECT TEMPERATURE, RAINFALL, monthname(DATE) FROM wd_analysis.rainfall;`
- Result Grid:** Displays 12 rows of data for January.

TEMPERATURE	RAINFALL	monthname(DATE)
15.2	12.5	January
17.8	8.2	January
20.1	0	January
18.6	3.7	January
14.8	21.1	January
16.5	15.3	January
19.2	6.8	January
21.7	0	January
17.3	11.2	January
15.8	18.6	January
16.2	9.5	January
18.6	9.6	January

Bottom Instance:

- Query 1:** `#SELECT * FROM wd_analysis.rainfall;`
- Query 2:** `SELECT TEMPERATURE, RAINFALL, monthname(DATE) FROM wd_analysis.rainfall;`
- Result Grid:** Displays 12 rows of data for February and March.

TEMPERATURE	RAINFALL	monthname(DATE)
18.6	9.6	February
21.3	0	February
16.8	7.4	February
15.2	21.6	February
17.8	17.2	February
20.1	15.4	February
18.6	16.7	February
14.8	18.9	February
16.5	22.1	February
19.2	17.8	March
21.7	14.7	March
18.6	9.6	March

➤ DATA CLEANING AND PREPROCESSING

The data will be cleaned and preprocessed by MS Excel

Step1: We have gathered the information about weather data in an excel.

Step2: Handling missing values:- for this we can use ISBLANK, we have to fill all those rows that are empty or use proper substitutes.

Step3: Removing duplicates;- For this we have to select all the data and go to the DATA tab and find the Remove Duplicates button.

Now we have the cleaned and monthly average Temperature and Rainfall

	A	B	C	D	E	F	G	H	I	J
	LOCATION	DATE	RAINFALL	TEMPERATURE	HUMIDITY		MONTHS	unique_months	average_temp	average_rainfall
2	Bangalore	2022-01-01	12.5	15.2	78		1	1	17.88064516	9.361290323
3	Hyderabad	2022-01-02	8.2	17.8	65		1	2	17.86071429	10.76071429
4	Ahmedab	2022-01-03	0	20.1	52		1	3	18.01290323	15.82258065
5	Chennai	2022-01-04	3.7	18.6	71		1	4	17.725	10.81666667
6	Kolkata	2022-01-05	21.1	14.8	82		1			
7	Delhi	2022-01-06	15.3	16.5	75		1			
8	Bangalore	2022-01-07	6.8	19.2	61		1			
9	Hyderabad	2022-01-08	0	21.7	48		1			
0	Ahmedab	2022-01-09	11.2	17.3	73		1			
1	Chennai	2022-01-10	18.6	15.8	79		1			
2	Kolkata	2022-01-11	9.5	16.2	72		1			
3	Surat	2022-01-12	2.1	19.8	58		1			
4	Pune	2022-01-13	0	22.4	45		1			
5	Jaipur	2022-01-14	7.4	17.9	69		1			
6	Lucknow	2022-01-15	14.9	15.1	81		1			
7	Surat	2022-01-16	19.2	16.8	77		1			
8	Pune	2022-01-17	5.6	18.5	63		1			
9	Jaipur	2022-01-18	0	21.2	51		1			
0	Lucknow	2022-01-19	10.8	17.6	75		1			
1	Kanpur	2022-01-20	16.3	14.5	84		1			
2	Nagpur	2022-01-21	3.9	16.1	71		1			
3	Indore	2022-01-22	0	19.4	59		1			
4	Ahmedab	2022-01-23	7.1	18.2	67		1			

DATA VISUALIZATION(POWER BI)

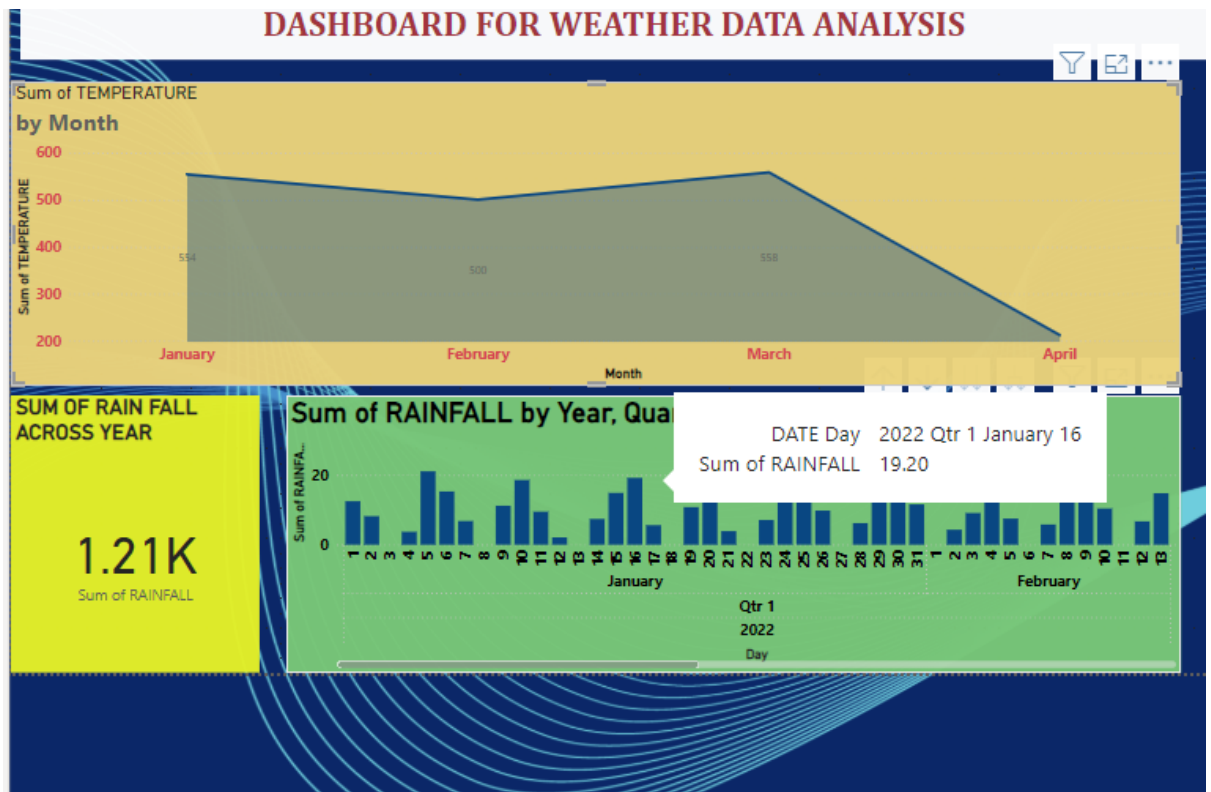


Temperature trends over month:



Compare rainfall data across years

A line chart will be used to visualize overall temperature trends over month and bar chart is used to compare rainfall data across years.

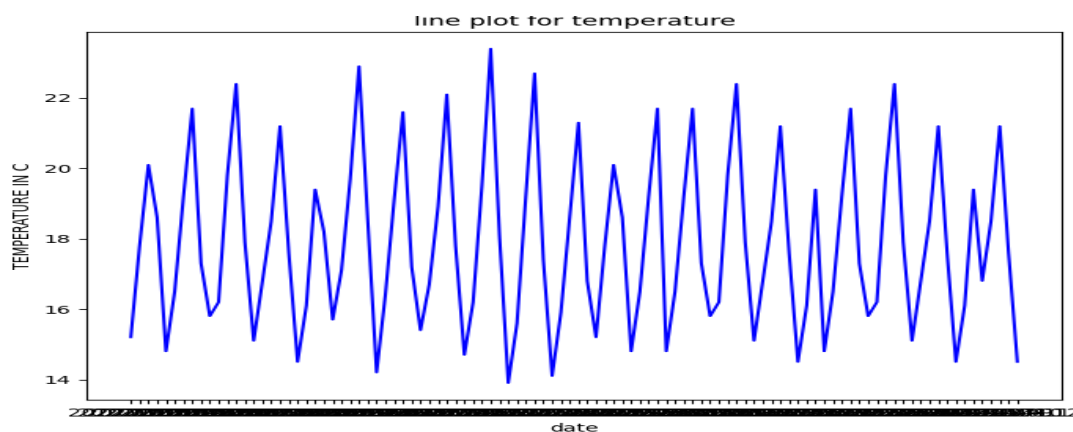


The line graph displays the average monthly temperature over a year, March month consistently maintains higher average monthly temperature throughout the year and April month maintains lowest average monthly temperature throughout the year. A bar graph compares the average rainfall data across year.

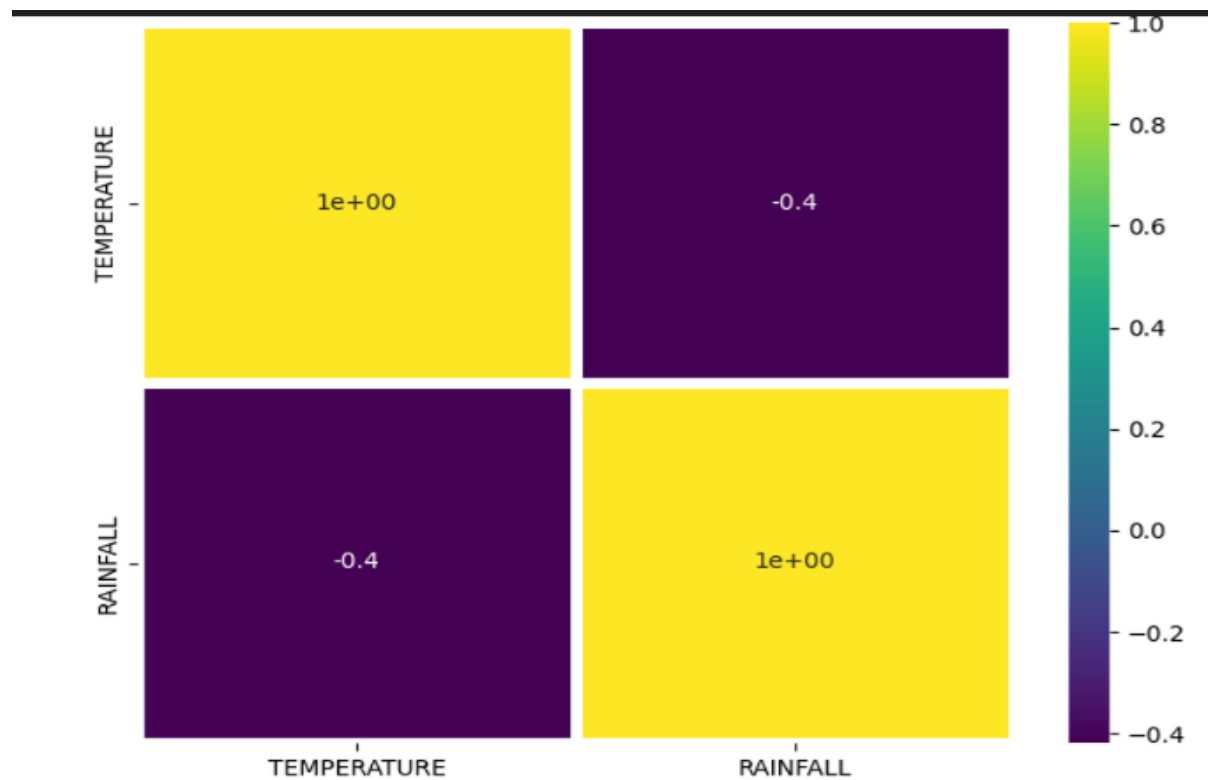
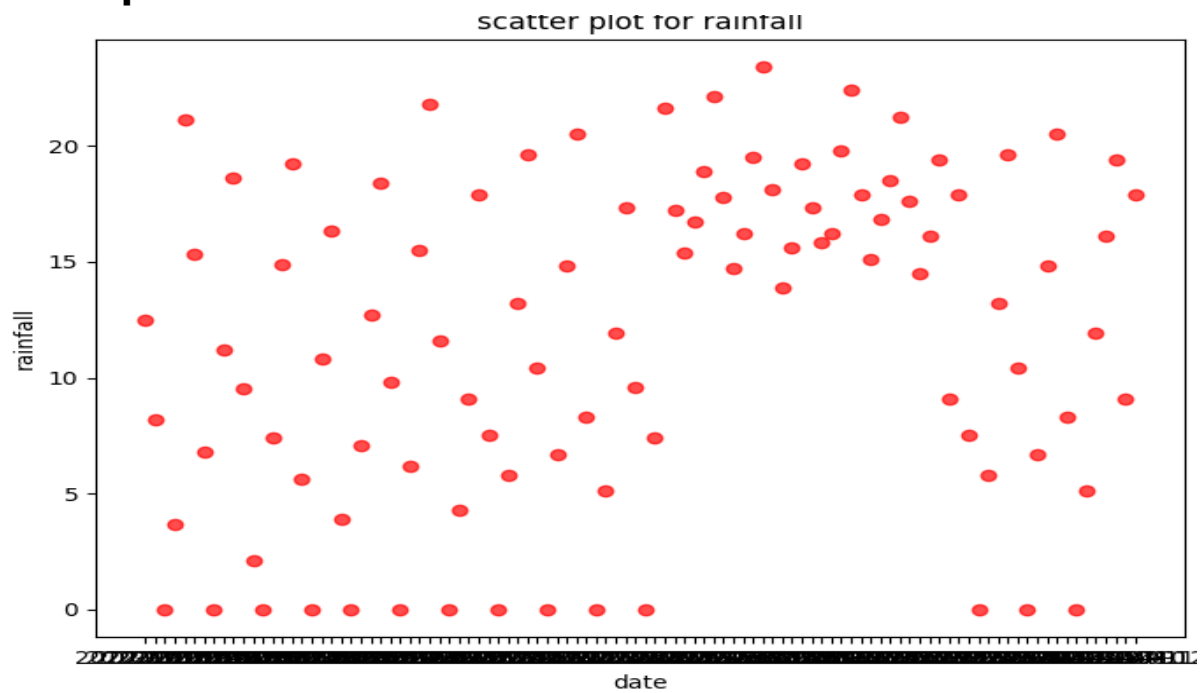
➤ Exploratory Analysis(Python)

EDA is an important step in any data analysis to understand the distribution of features, relationships between variables.

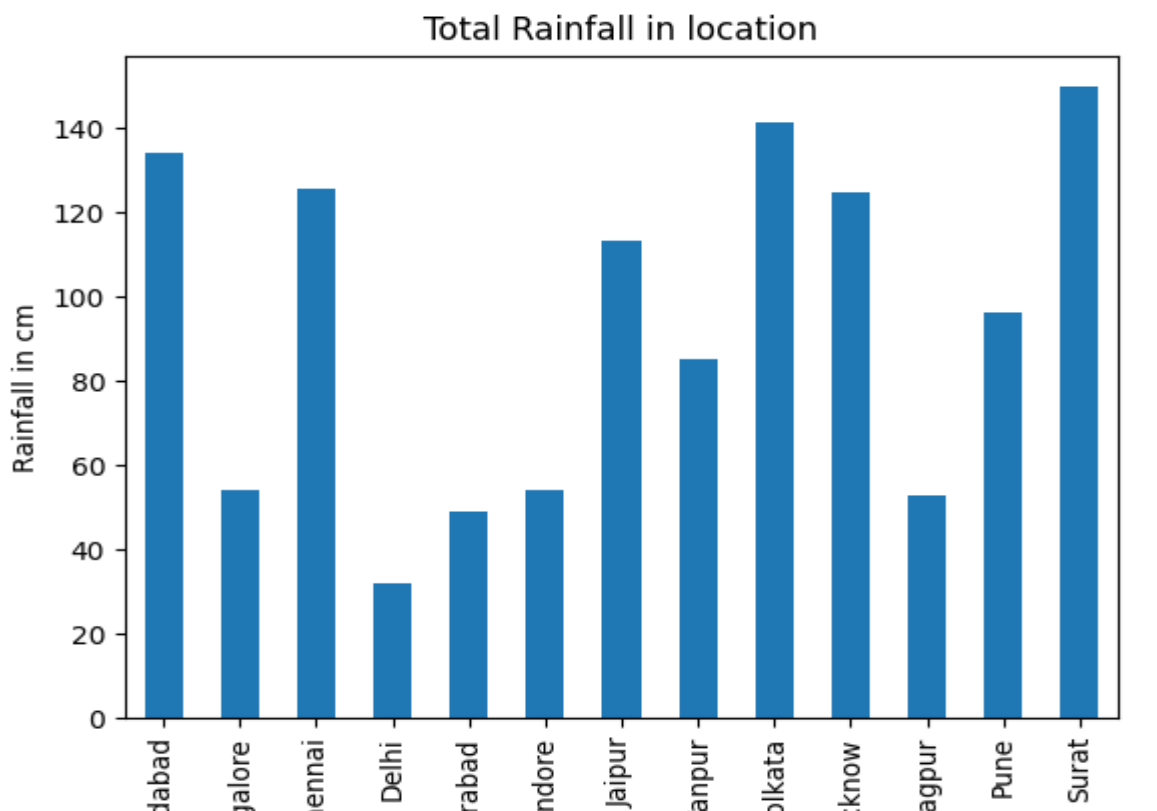
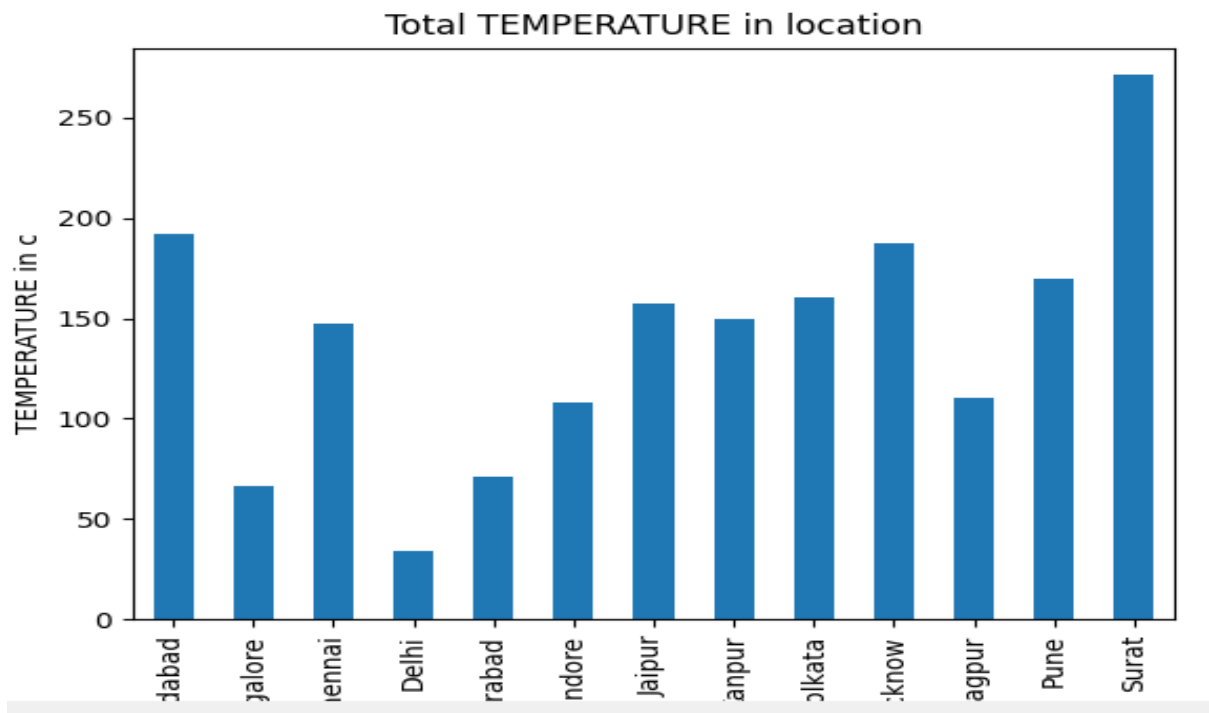
Line plot for Temperature:



Scatterplots for Rainfall:



A heatmap where each cell represents the correlation coefficient between temperature and rainfall in the weather dataset. The color intensity indicates the strength and direction of the correlation, with yellow indicates the positive correlation and purple represents negative correlation.



Conclusion:

This analysis provides a comprehensive overview of the weather conditions in major cities, highlighting the temperature ,rainfall, and humidity. By using visualization like line chart, scatter plot and heatmaps we were able to extract meaningfull insights into data. Further analysis could involve comparing these trends with historical weather patterns or exploring predictive modeling to forecast future weather trends.