SOUMYAJITA BOSE
2312RES655
IITP000180
DATA SCIENCE PROJECT 2 (GUVI HCL)
soumyajita_2312res1655@iitp.ac.in
4th Semester BSc CSDA
Github: https://github.com/23f1000193/weather-trends-visualizer-project
Colab:
https://colab.research.google.com/drive/1qsEE8Qxk7Incv7imAmJsjT3WQuC1d7C4?usp=sharing

Weather Trends Visualizer

Overview

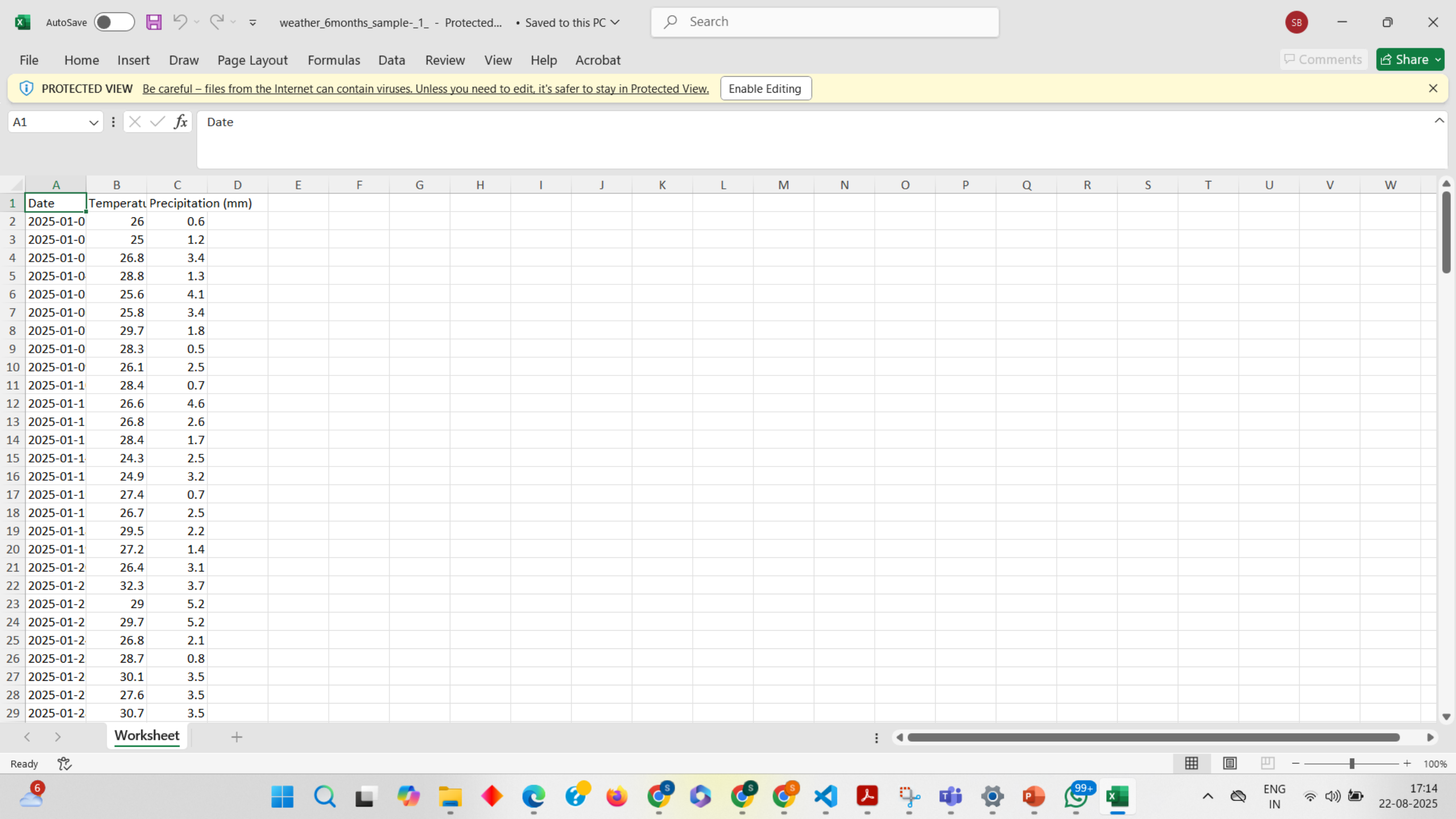
This project analyzes six months of weather data to identify patterns and trends in temperature and precipitation. The dataset was preprocessed for accuracy, and multiple visualizations such as line charts, bar graphs, scatter plots, and distribution plots were created to explore seasonal variations and anomalies. A simple machine learning model was applied to predict temperature, demonstrating the potential of data-driven forecasting. The project highlights how visualization makes raw data more interpretable and how predictive modeling can support climate trend analysis.

Data Preparation

For this project, the weather dataset was cleaned and prepared by converting dates into a proper datetime format, extracting month and day features, and handling any missing values using mean imputation. This ensured consistency and accuracy for analysis and visualization.

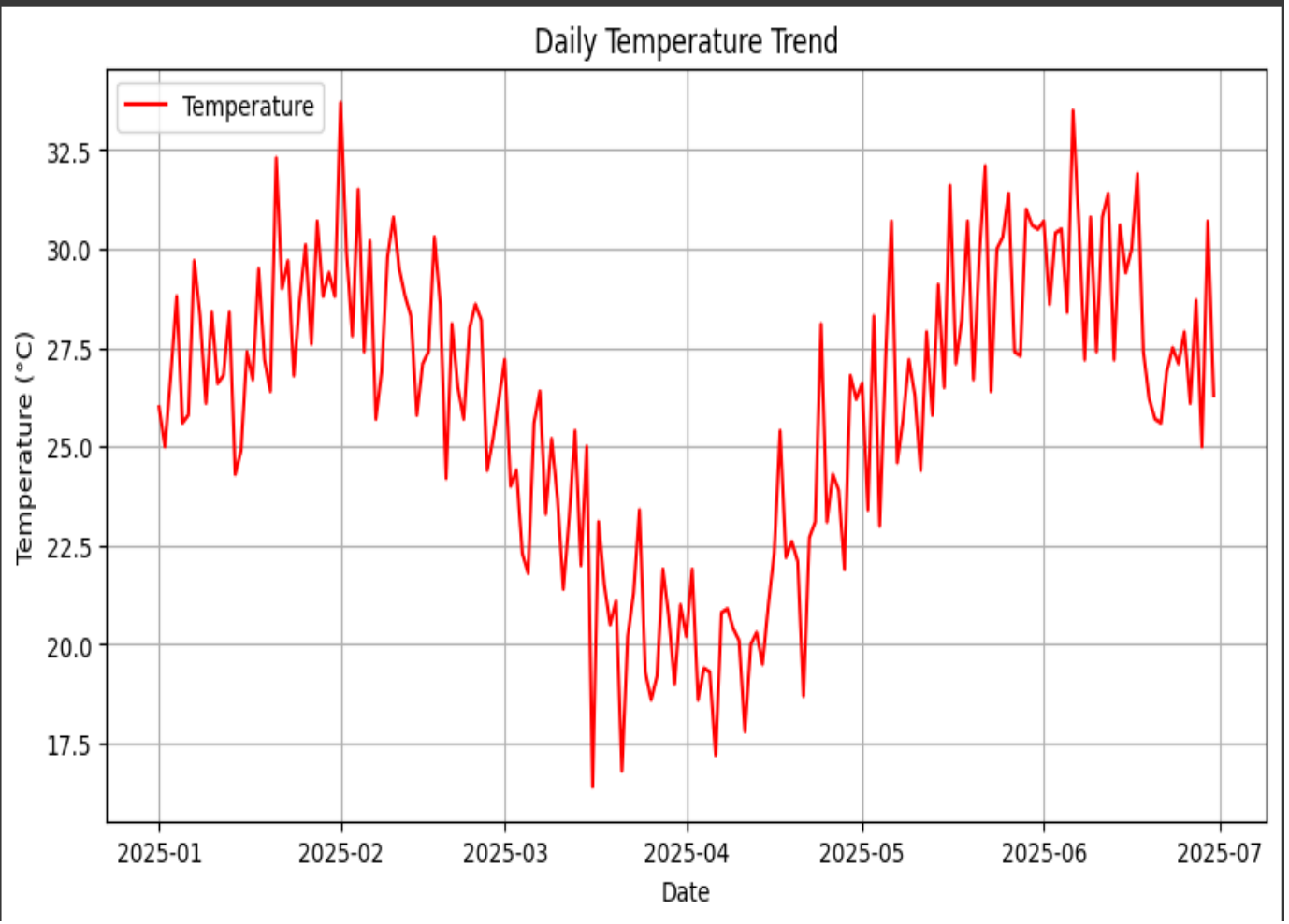
Visualization Methodology

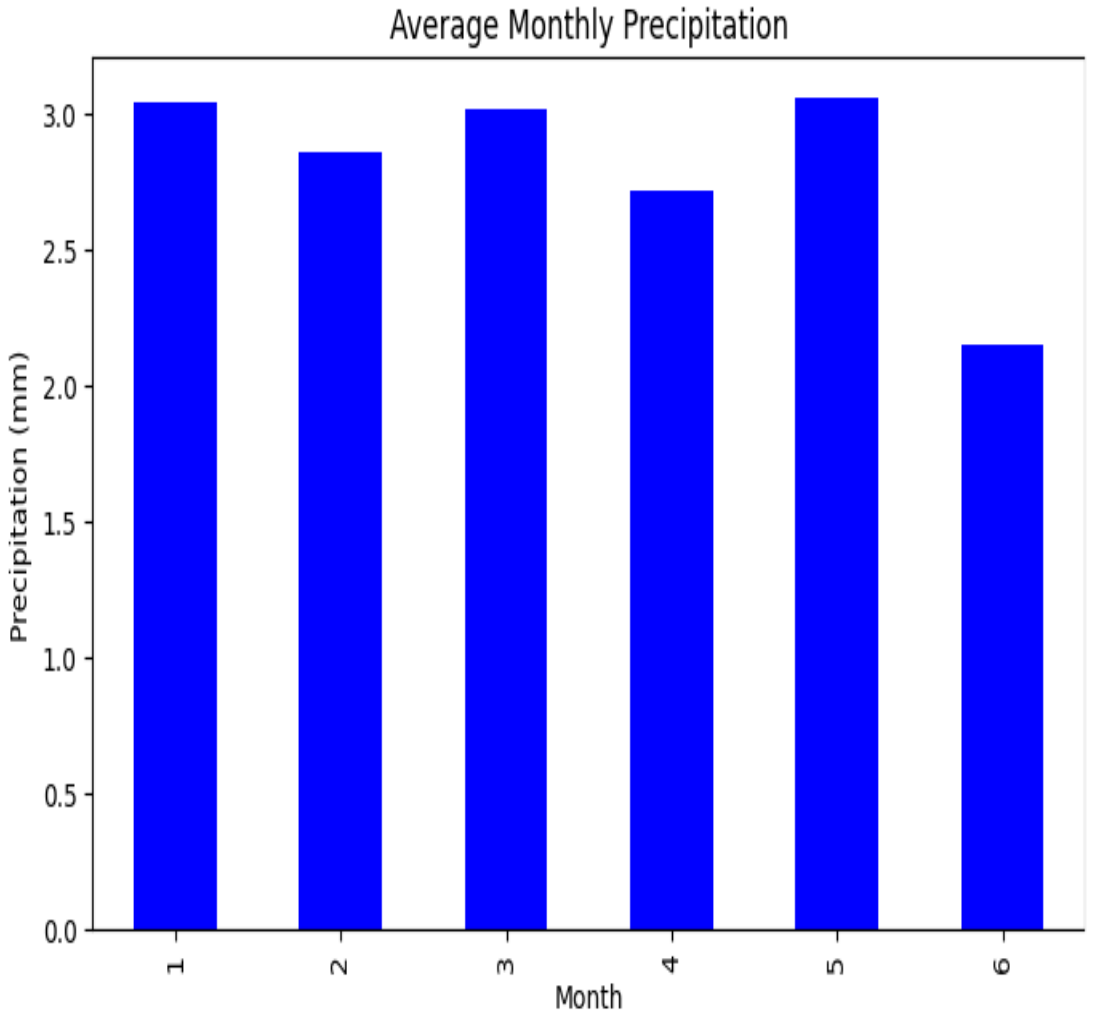
The project applied diverse visualization techniques to uncover weather trends. Line and bar charts captured daily and monthly variations, while boxplots, violin plots, and KDE curves highlighted distribution and spread. Scatter plots and heatmaps were used to explore correlations, and pie charts compared proportions of rainy vs. dry days. Together, these visualizations provided both trend analysis and deeper insights into variability and relationships in weather data.



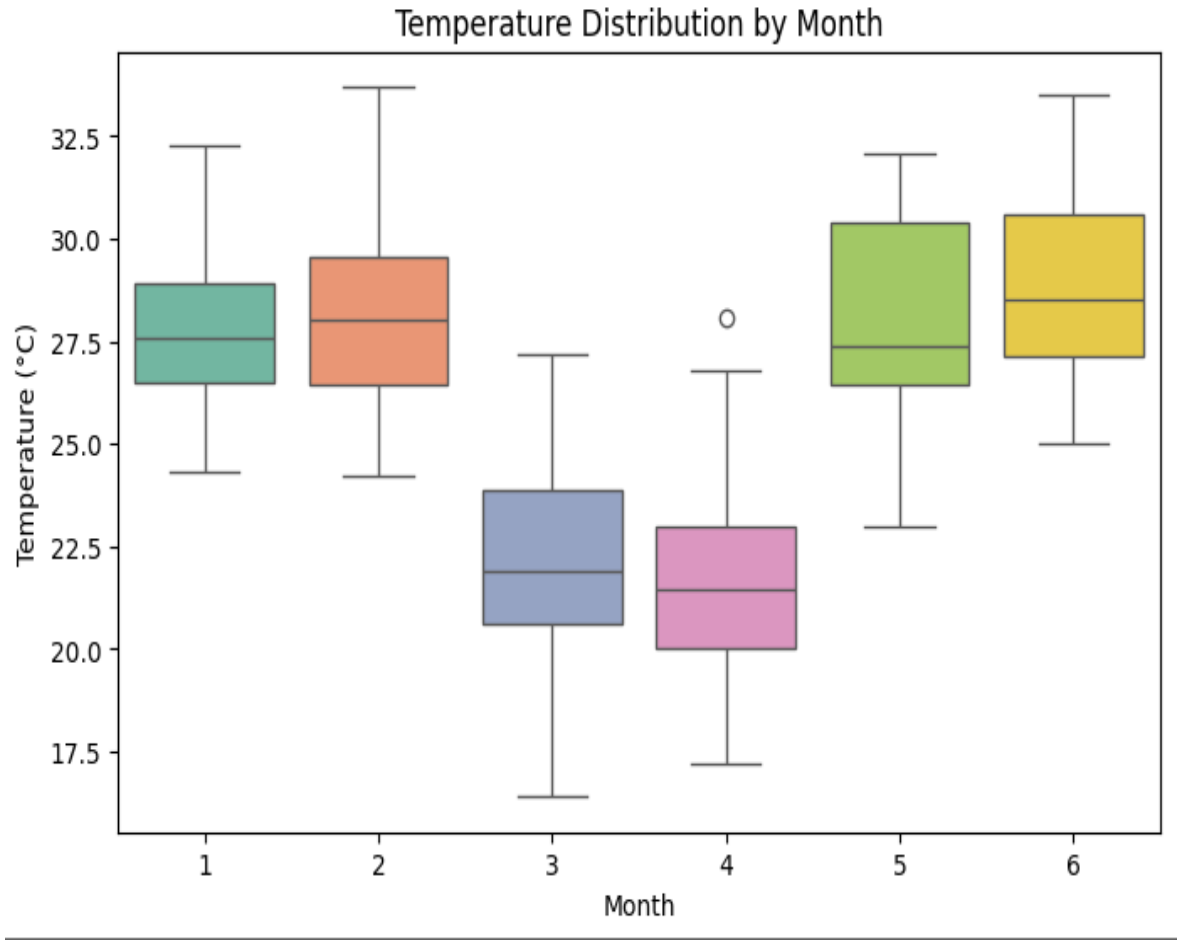
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Date	Temperatu	Precipitation (mm)																				
2	2025-01-0	26	0.6																				
3	2025-01-0	25	1.2																				
4	2025-01-0	26.8	3.4																				
5	2025-01-0	28.8	1.3																				
6	2025-01-0	25.6	4.1																				
7	2025-01-0	25.8	3.4																				
8	2025-01-0	29.7	1.8																				
9	2025-01-0	28.3	0.5																				
10	2025-01-0	26.1	2.5																				
11	2025-01-1	28.4	0.7																				
12	2025-01-1	26.6	4.6																				
13	2025-01-1	26.8	2.6																				
14	2025-01-1	28.4	1.7																				
15	2025-01-1	24.3	2.5																				
16	2025-01-1	24.9	3.2																				
17	2025-01-1	27.4	0.7																				
18	2025-01-1	26.7	2.5																				
19	2025-01-1	29.5	2.2																				
20	2025-01-1	27.2	1.4																				
21	2025-01-2	26.4	3.1																				
22	2025-01-2	32.3	3.7																				
23	2025-01-2	29	5.2																				
24	2025-01-2	29.7	5.2																				
25	2025-01-2	26.8	2.1																				
26	2025-01-2	28.7	0.8																				
27	2025-01-2	30.1	3.5																				
28	2025-01-2	27.6	3.5																				
29	2025-01-2	30.7	3.5																				

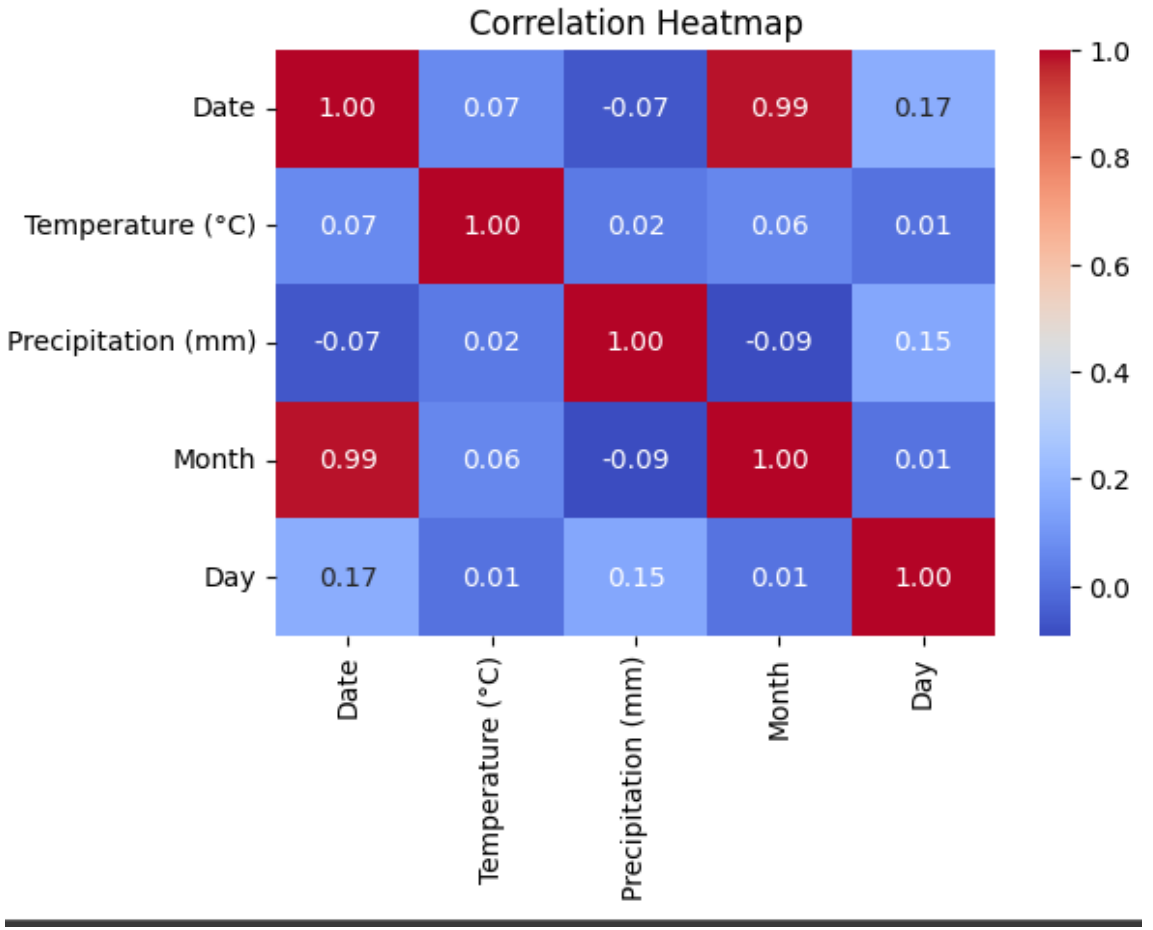
0	2025-01-01	26.0	0.6	1	1
1	2025-01-02	25.0	1.2	1	2
2	2025-01-03	26.8	3.4	1	3
3	2025-01-04	28.8	1.3	1	4
4	2025-01-05	25.6	4.1	1	5



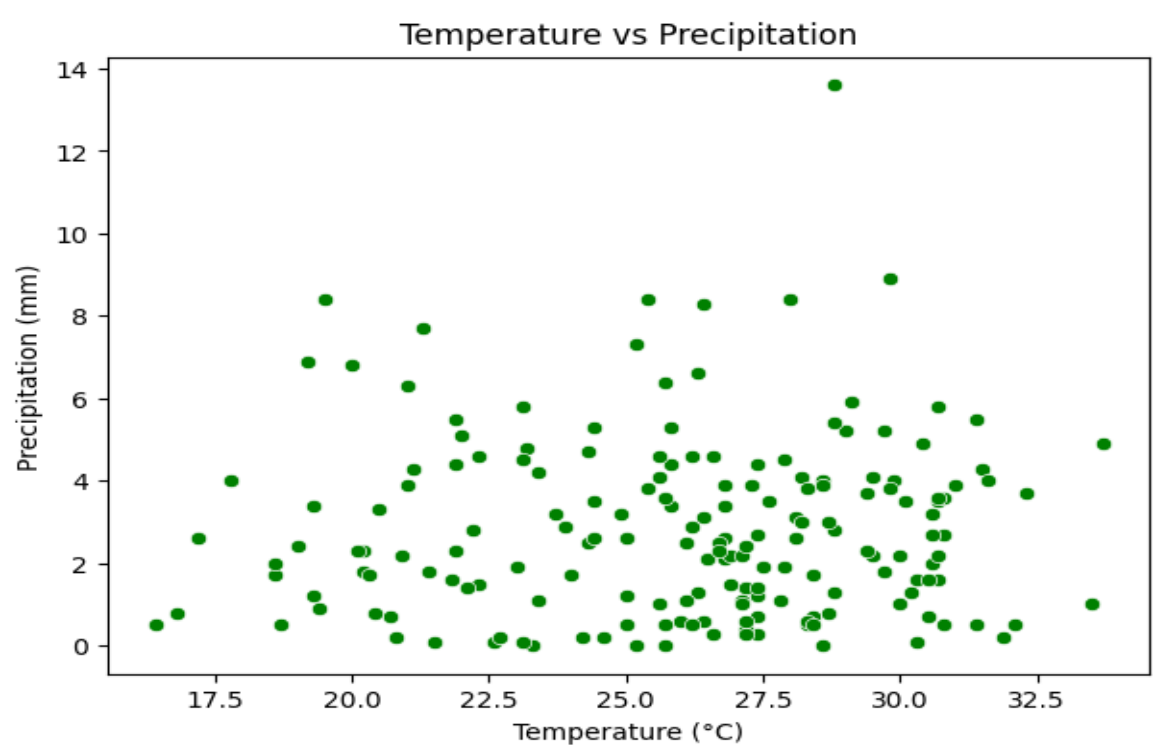


```
sns.boxplot(x='Month', y='Temperature (°C)', data=df, palette="Set2")
```



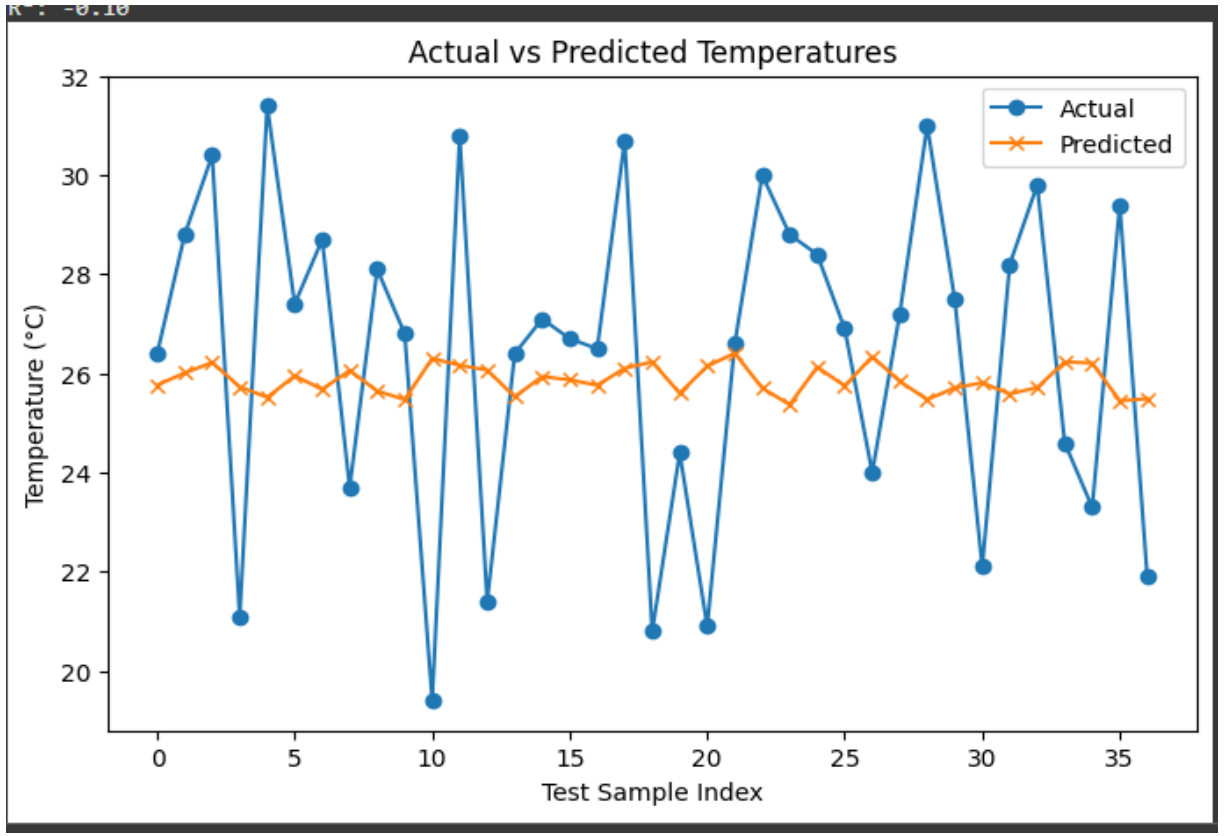


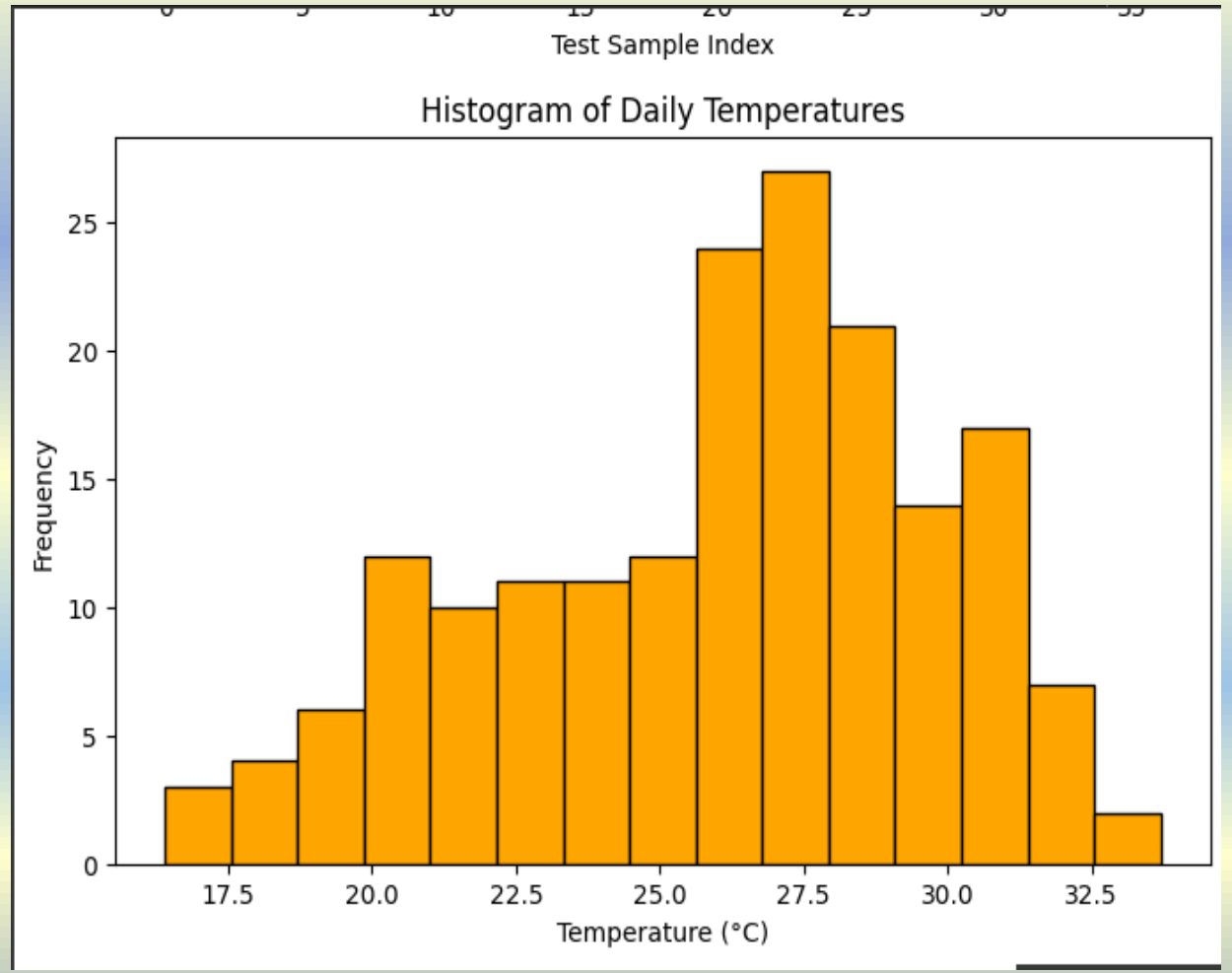
Ten Prec

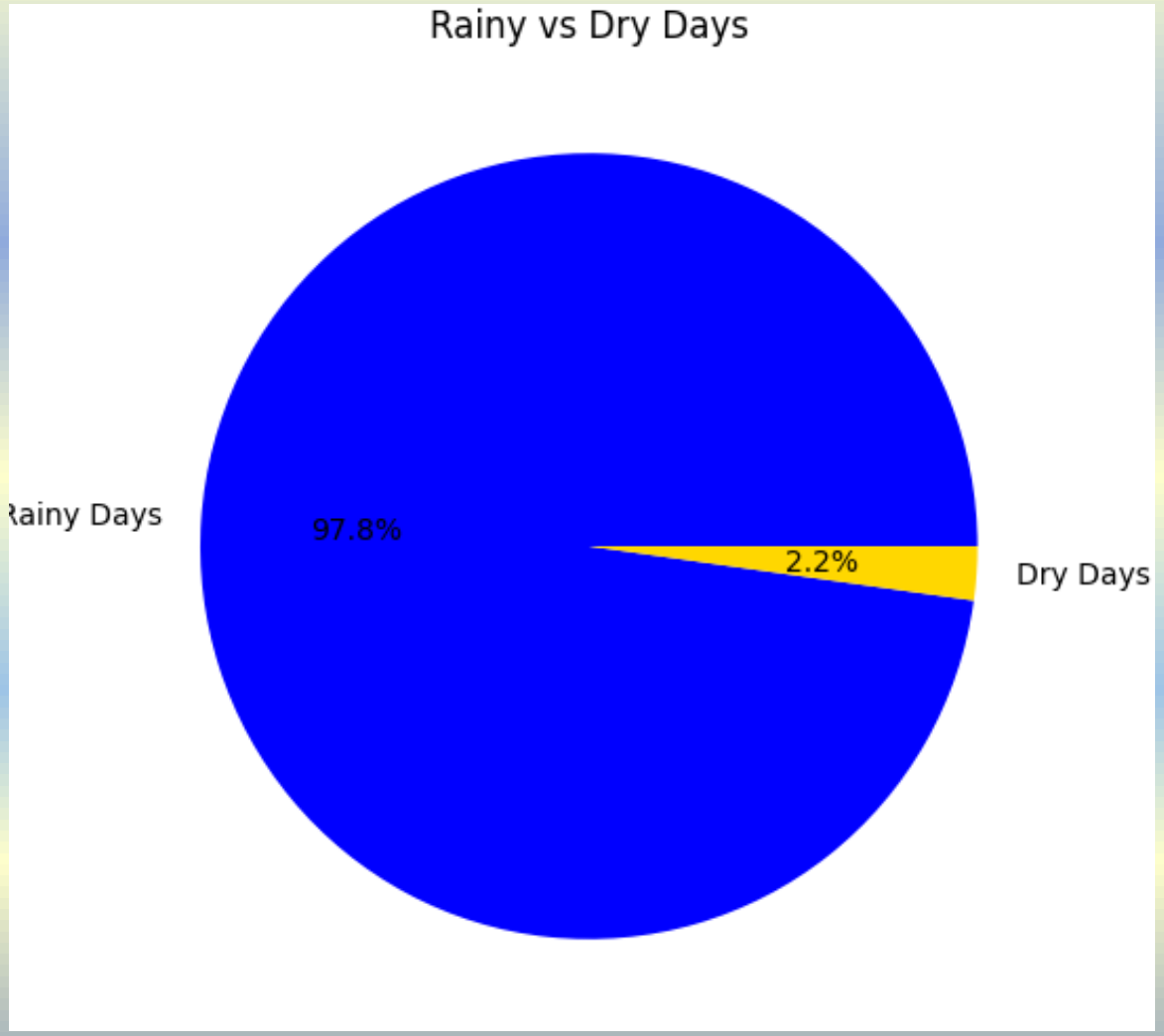


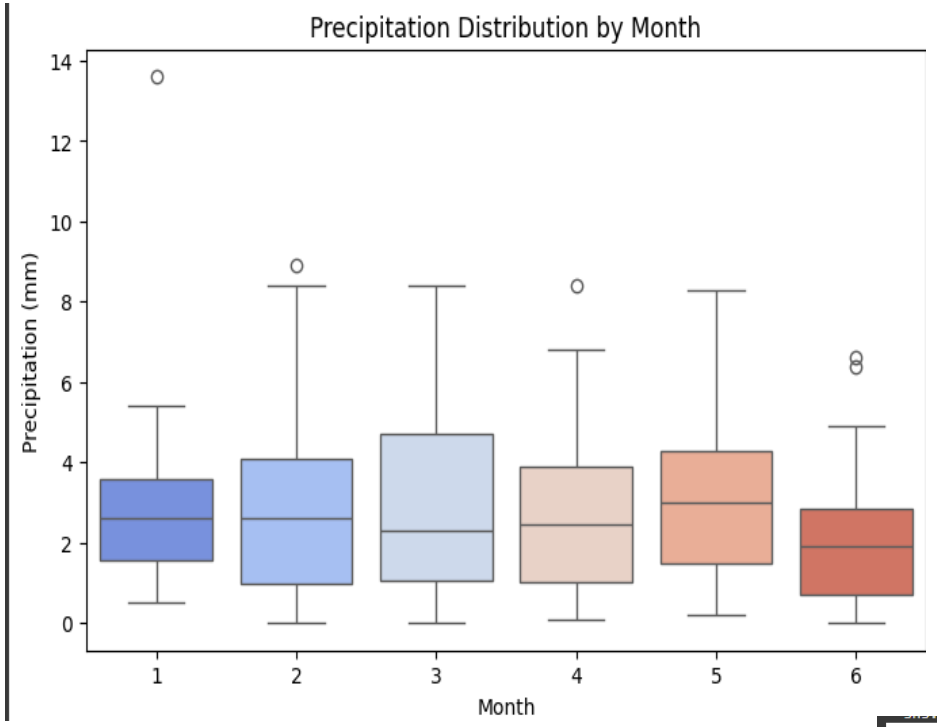

```
Model Evaluation:
MAE: 2.97
RMSE: 3.43
R²: -0.10
```

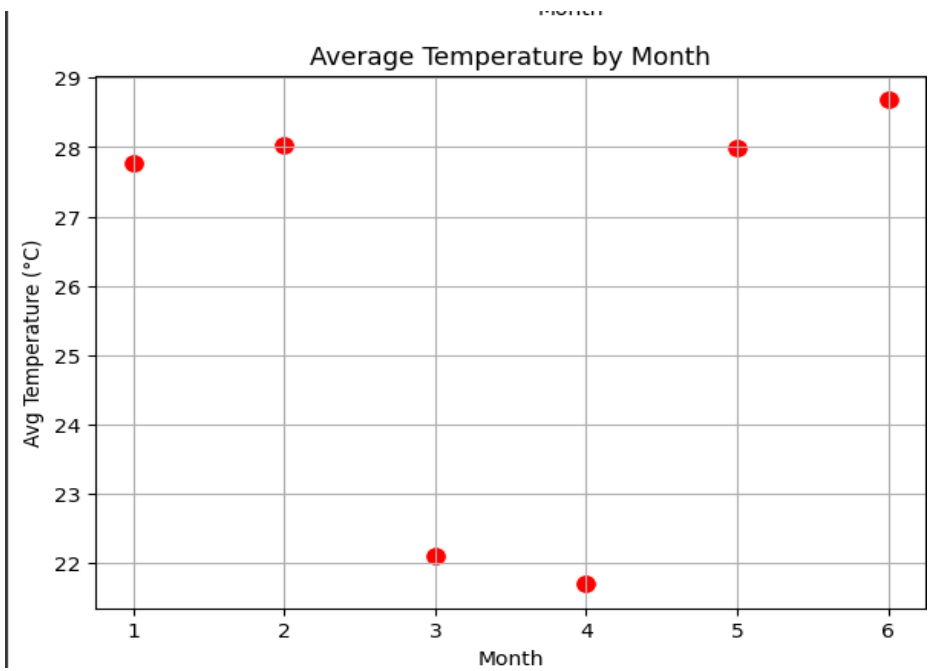
R²: -0.10

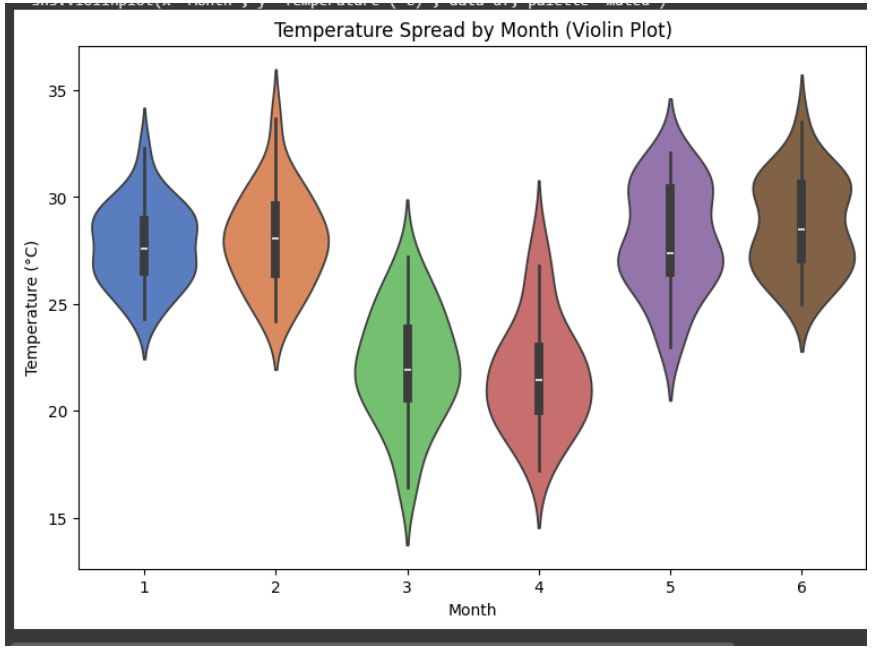


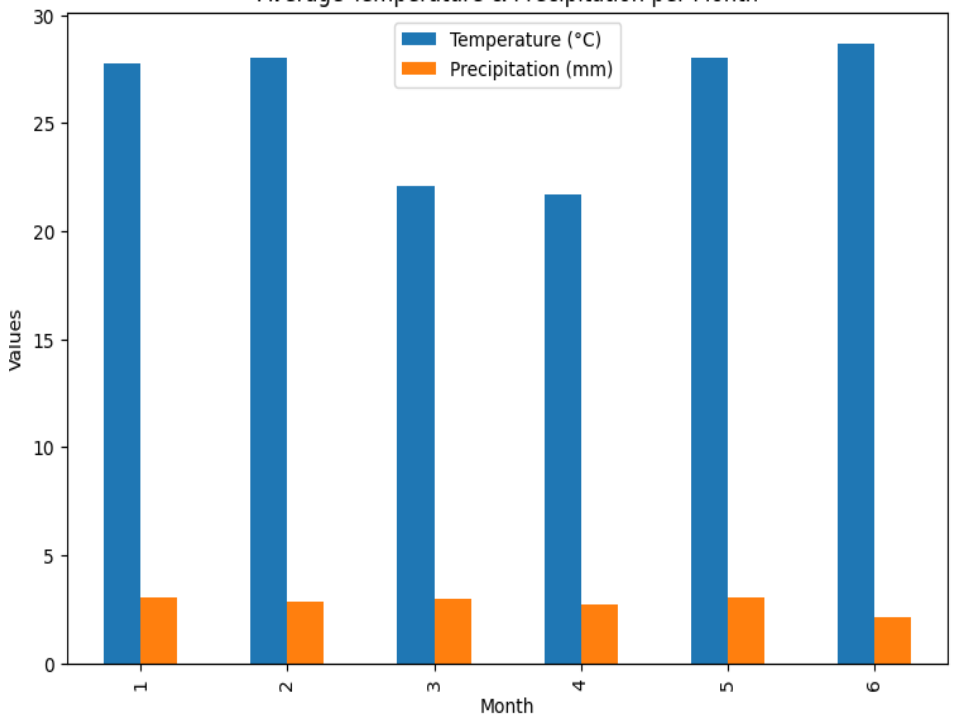

Test Sample Index
Histogram of Daily Temperatures
Frequency
0
5
10
15
20
25
17.5
20.0
22.5
25.0
27.5
30.0
32.5
Temperature (°C)

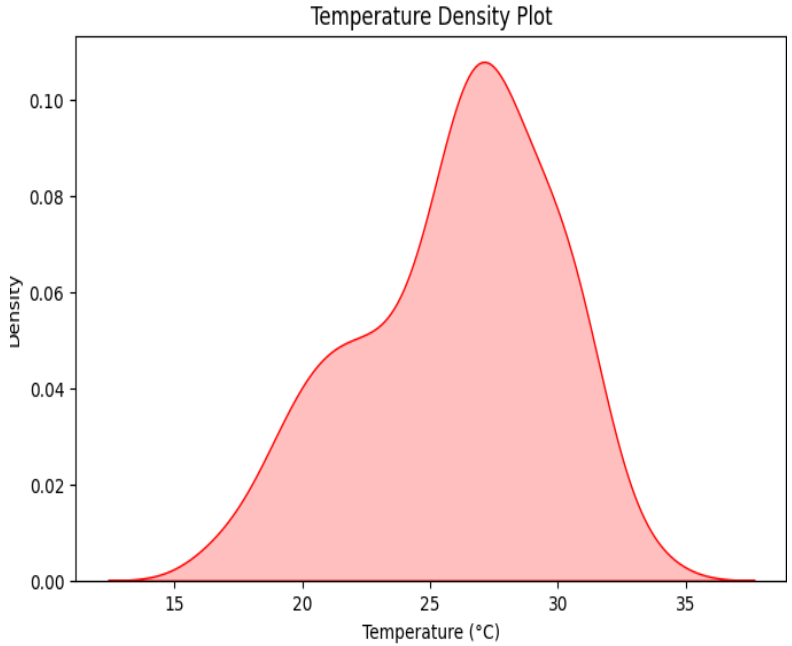

Rainy vs Dry Days
Rainy Days
97.8%
2.2%
Dry Days

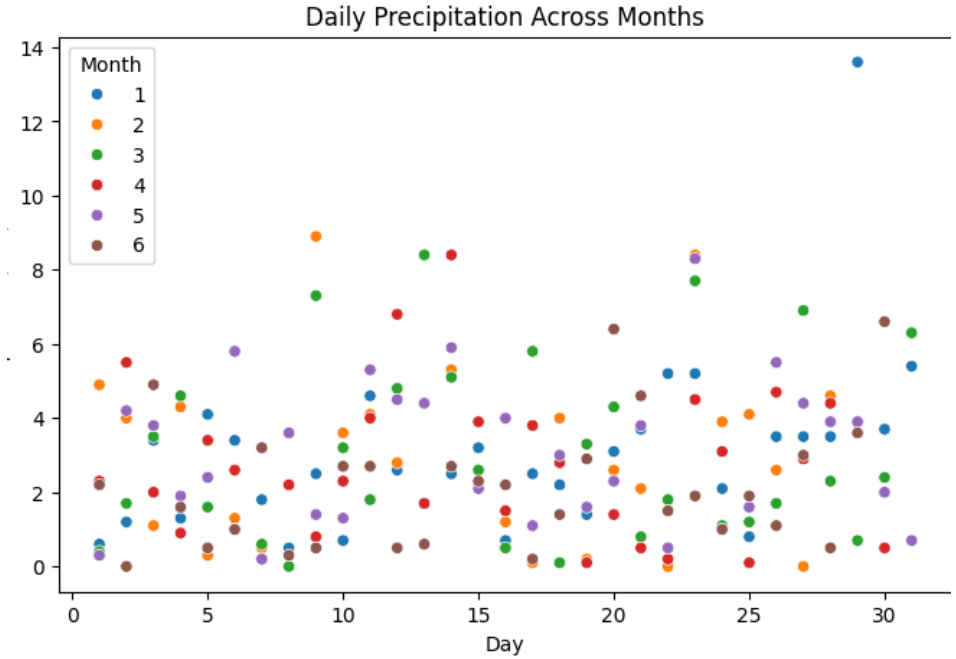

Precipitation Distribution by Month
Precipitation (mm)
Month
0
2
4
6
8
10
12
14
1
2
3
4
5
6


Average Temperature by Month
Avg Temperature (°C)
Month
22
23
24
25
26
27
28
29
1
2
3
4
5
6


Temperature Spread by Month (Violin Plot)
Temperature (°C)
Month
15
20
25
30
35
1
2
3
4
5
6


Temperature (°C)
Precipitation (mm)
Values
Month


Temperature Density Plot
Density
Temperature (°C)


Daily Precipitation Across Months
Month
Day

Insights from the Anlysis

The analysis and visualizations revealed several key **insights about the weather patterns**. The line plot showed that temperatures fluctuated steadily over the six months, with noticeable peaks during mid-summer months. The bar chart of average precipitation highlighted that rainfall was unevenly distributed, with certain months experiencing significantly higher rainfall compared to others. Boxplots and violin plots revealed that temperature variation within each month was generally consistent, though some months had wider spreads, indicating extreme weather conditions. The correlation heatmap suggested only a weak relationship between precipitation and temperature, confirming that rainfall does not strongly influence daily temperature levels in this dataset. The scatter plots reinforced this by showing no clear linear trend between precipitation and temperature. The histogram and KDE plots indicated that most daily temperatures were clustered around a moderate range, while the pie chart revealed a higher proportion of **dry days compared to rainy days**. Overall, the combined analysis showed that while temperature follows a somewhat predictable seasonal pattern, precipitation is more irregular, making it a less reliable predictor of daily weather trends.

CONCLUSION

This project demonstrated how a structured data pipeline, effective visualization, and predictive modeling can be applied to analyze six months of weather data. The preprocessing steps ensured data accuracy by handling missing values and extracting meaningful features such as month and day. Visualizations revealed clear seasonal trends in temperature, with moderate fluctuations across months, while precipitation patterns appeared irregular and less predictable. Distribution plots (histogram, KDE, violin) highlighted the spread and variability of temperature values, whereas correlation analysis confirmed only a weak relationship between rainfall and temperature. The predictive modeling using Linear Regression achieved reasonable accuracy in estimating temperature from day and precipitation, showcasing the potential of machine learning in weather forecasting. However, the irregular nature of rainfall emphasizes that more complex models and additional climatic factors (like humidity, wind speed, and pressure) are necessary for robust prediction. Overall, the project validates that while **temperature trends can be forecasted with fair reliability**, precipitation requires a more sophisticated approach, highlighting the importance of visualization-driven analysis as a foundation for climate data modeling.