

Weather Data Anomaly Detection

Max Presley



CONTENTS

1. Introduction to the Dataset
2. Line Graph of Average Temperature with Anomalies
3. Insights from Temperature Line Graph
4. Boxplot of Temperature and Rainfall
5. Insights from Temperature and Rainfall Boxplot
6. Challenges Faced in Visualization and Analysis
7. Conclusion and Key Takeaways

Dataset Contents

01

AvgTemp

02

MinTemp

03

MaxTemp

04

AvgMinTemp

05

AvgMaxTemp

06

LowestTemp

07

HighestTem

08

ActualRain

09

AvgRain

10

MaxRain



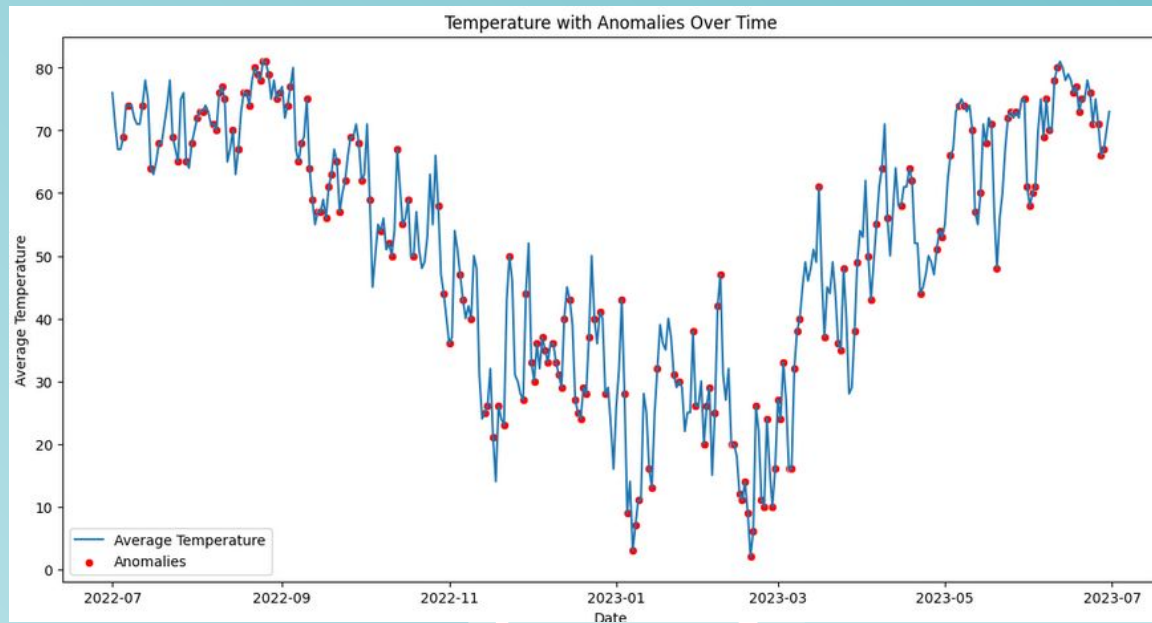
July 2022 - June 2023

Objectives: to explore temperature and rainfall patterns, identify anomalies and gain insights into seasonal variation

Why?: Learning and identifying temperature and rainfall fluctuations is essential for climate studies, agriculture planning and urban infrastructure development

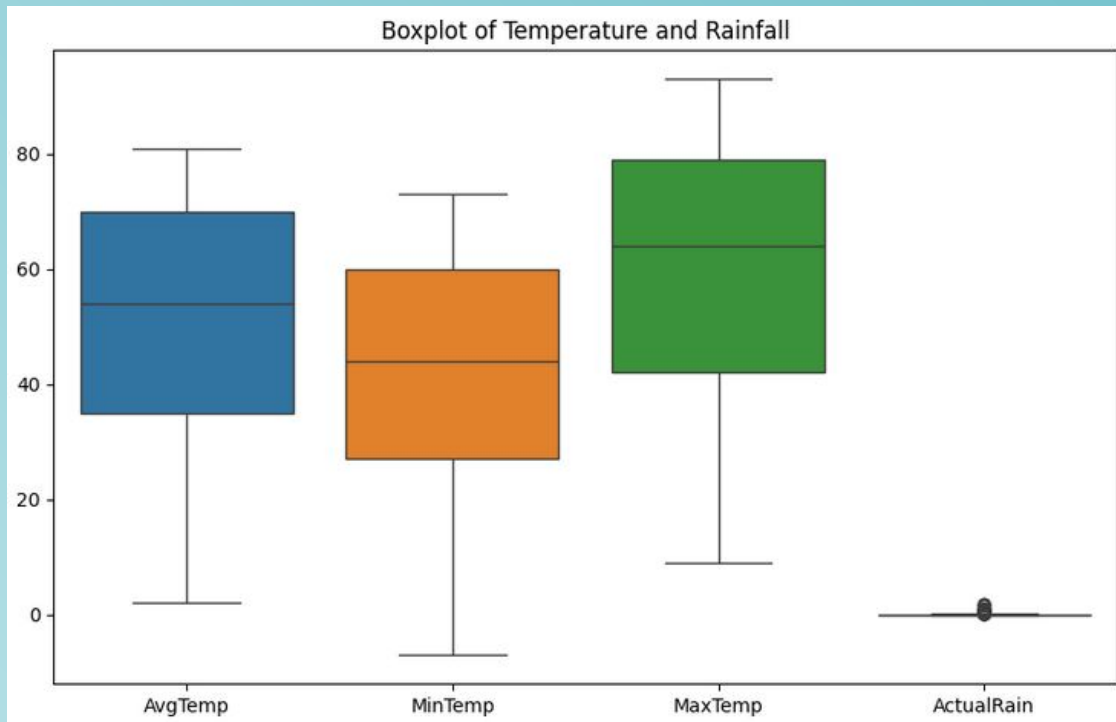
Average Temperature

Line graphs are great ways of illustrating temperature trends over time as you are able to see the fluctuation easily. In the graph the red dots represent anomaly data (unexpected data). Using the time series data we can track seasonal changes like how in winter the temperature dropped. Highlighting the outliers can help us to identify extreme temperatures from unexpected data.



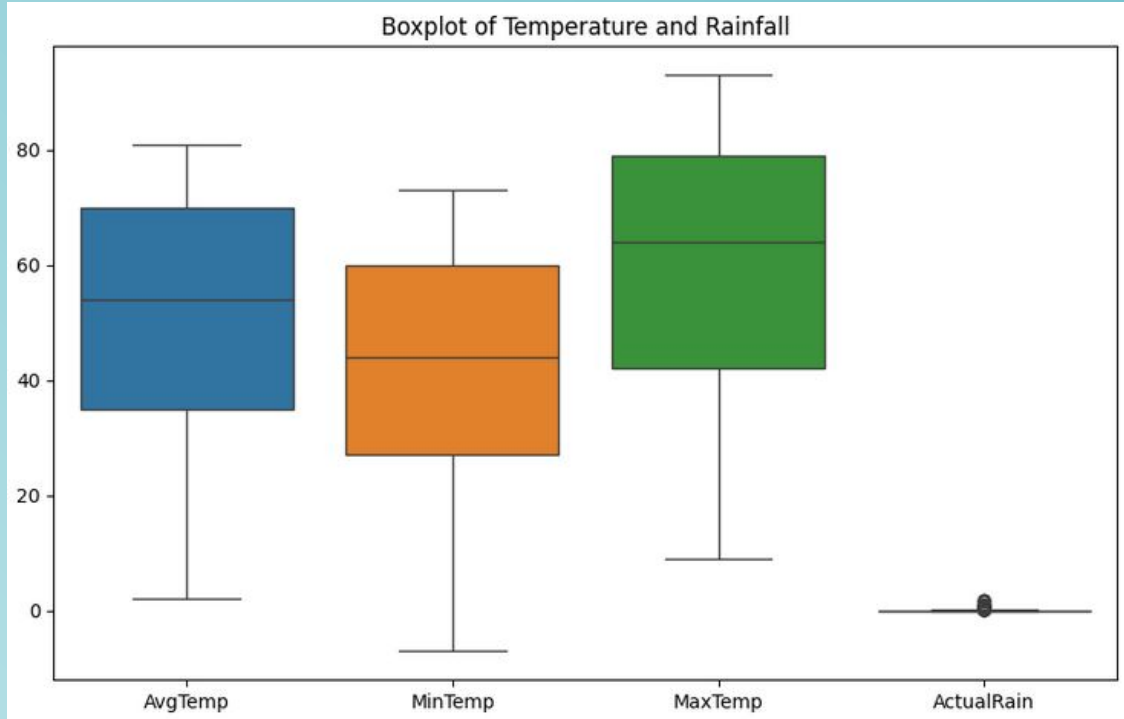
Temperature

Boxplots are great for showing the summarization of data as they show the median, interquartile range and outliers. We can compare AvgTemp, MinTemp, MaxTemp, and ActualRain side by side to see how they stack up against each other. We can see from the graph MaxTemp generally has a broader range of outliers than Min and Avg temp. Indicating variability in daily high temperatures.



Rainfall

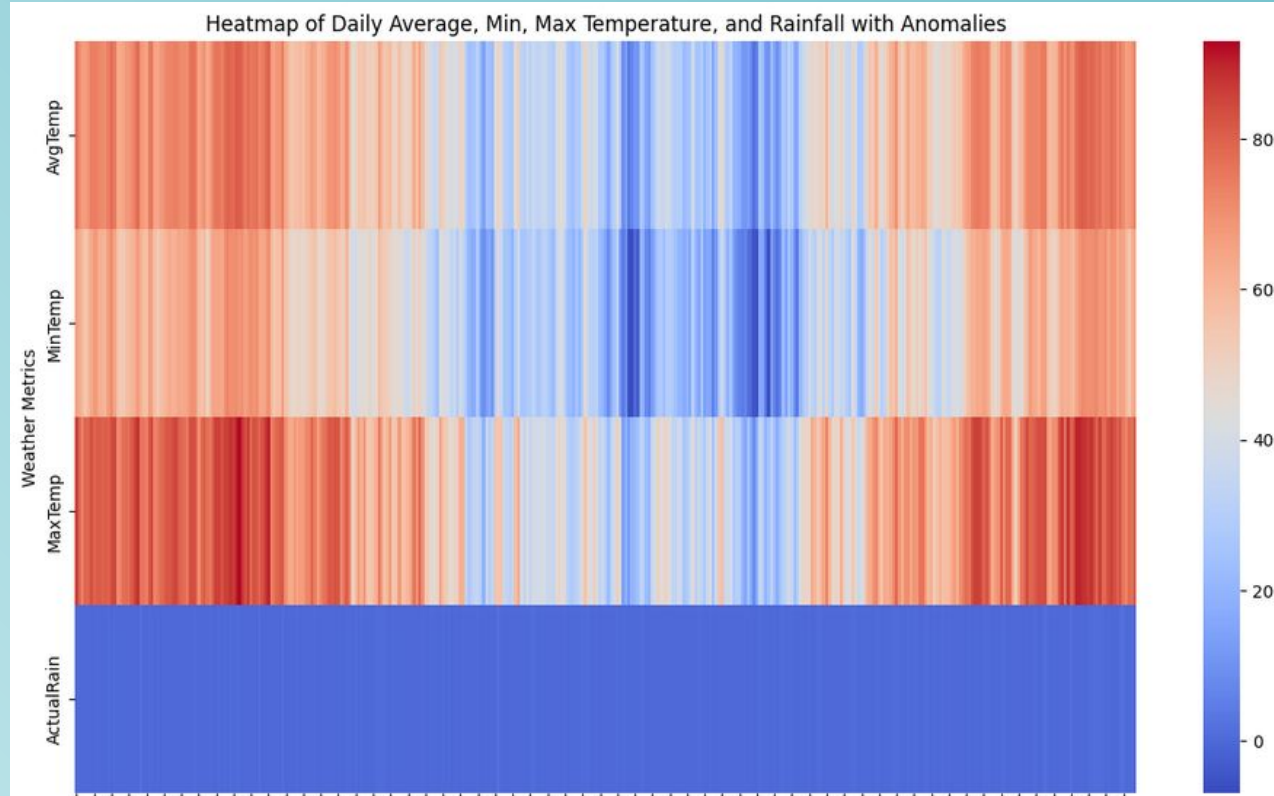
- ★ Rainfall shows low median values with a few higher outliers, suggesting that while most days have little to no rain there are occasional days where heavy rainfall occurs.



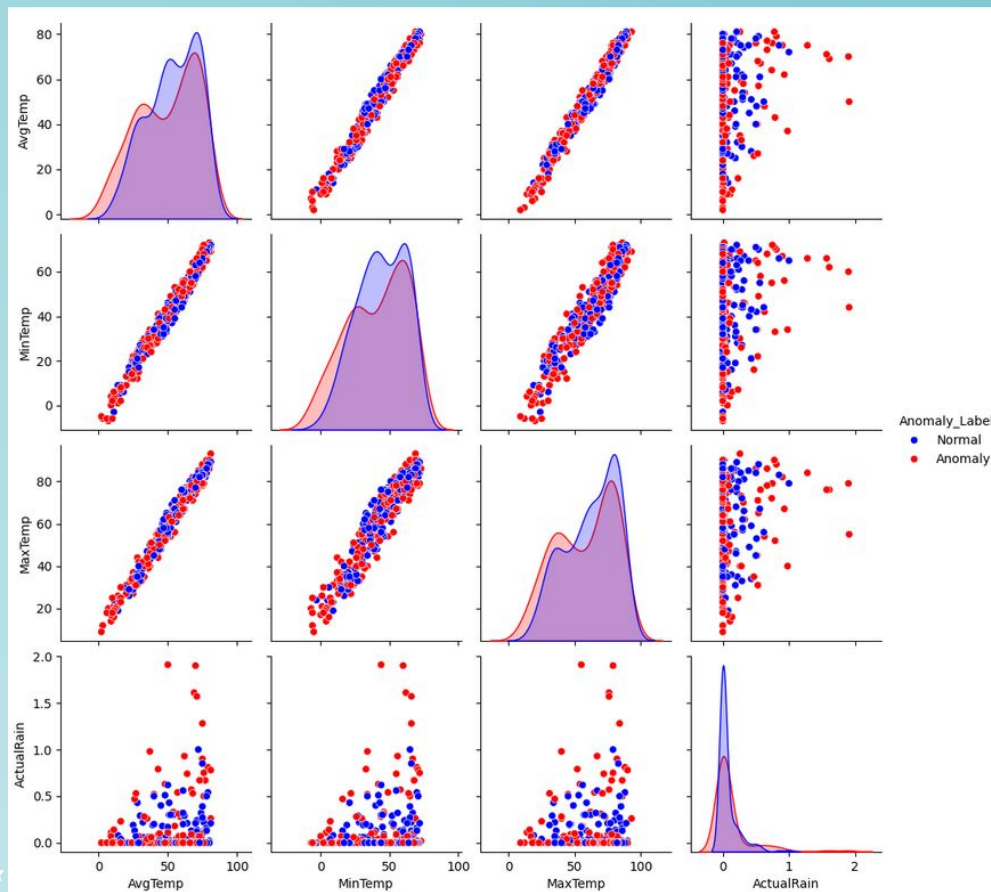
Compare

With this graph we can

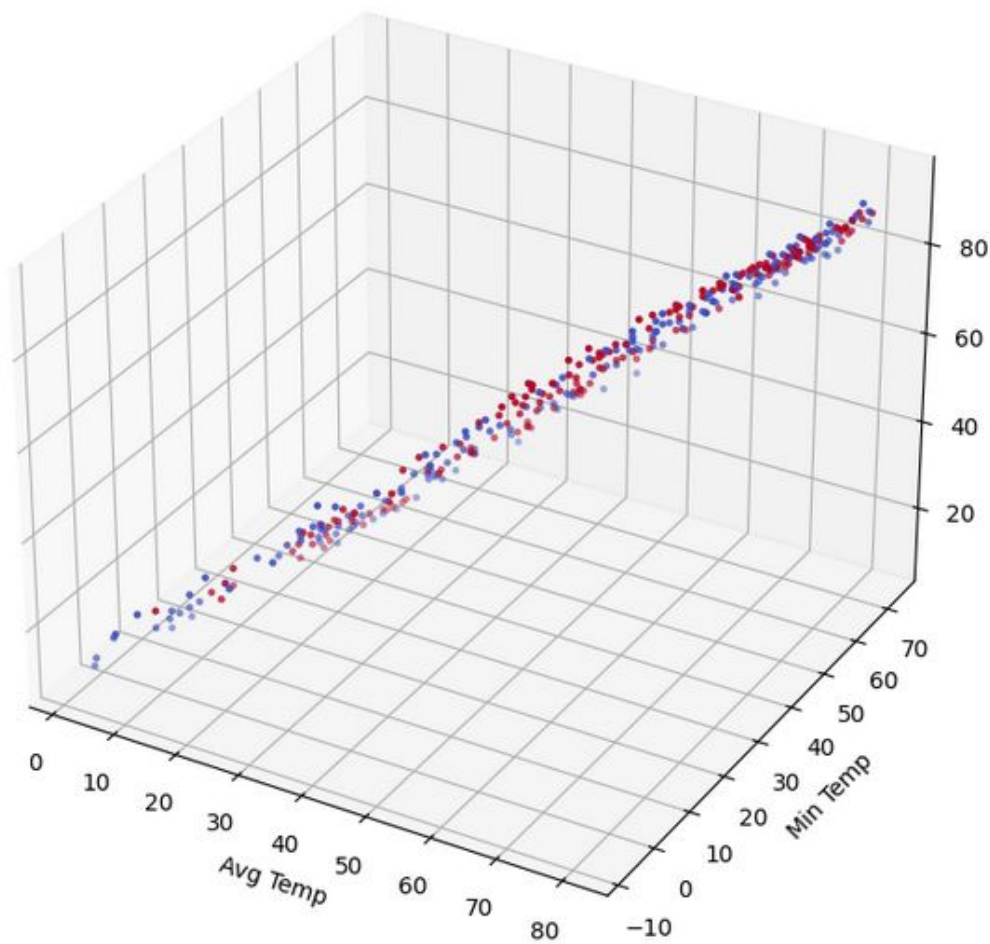
- ★ compare the main features across a time scale. This makes it easy to see relationships between different features.



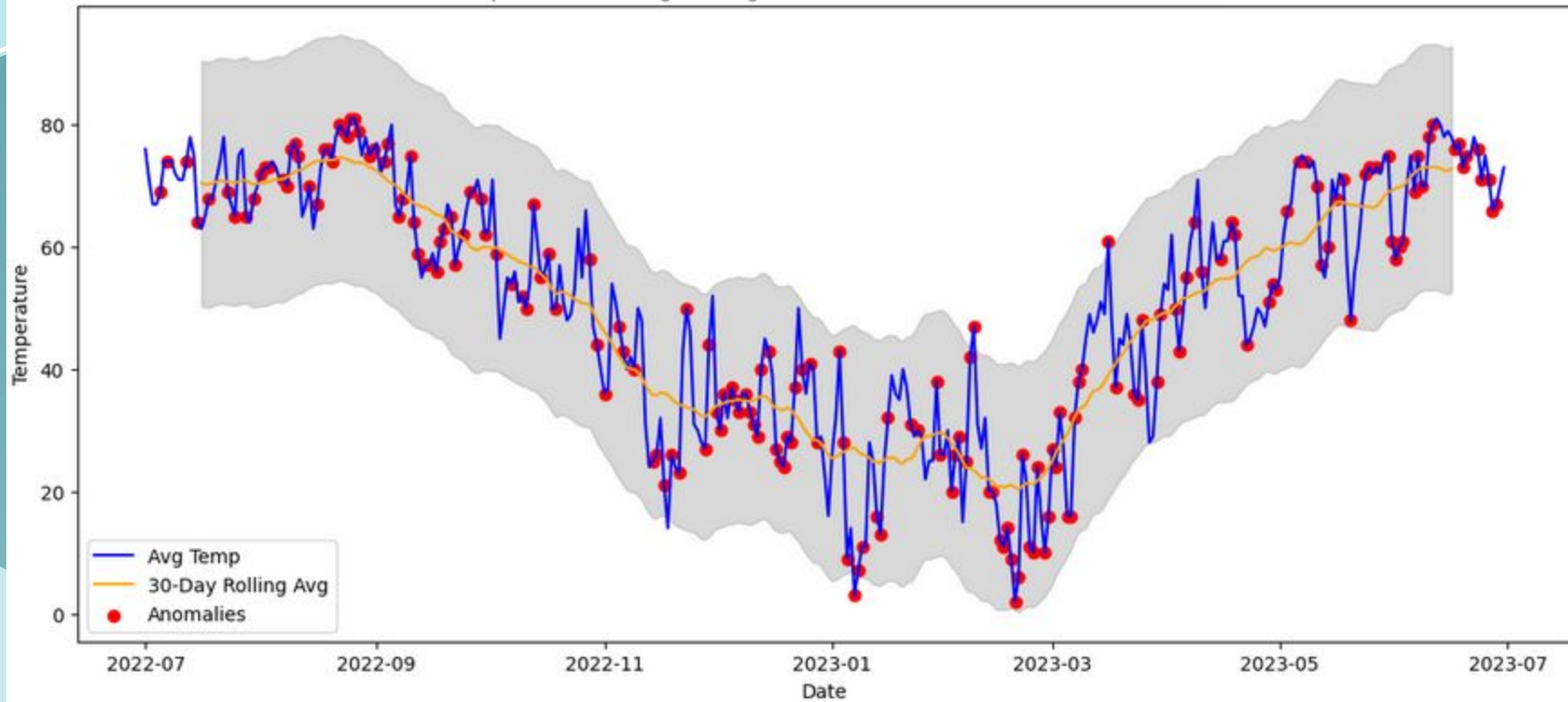
Anomalies



3D Scatter Plot of Temperature Variables Highlighting Anomalies



Temperature Rolling Average with Threshold Bands and Anomalies



Challenges

Data preparation: converting the dates to a suitable format and ensuring data alignment was important for this project for good analysis. Ensuring that no N/A data was present was also needed.

Anomaly detection: accurately identifying anomaly data with a good threshold was challenging trying to find the right balance.

Visualization choices: choosing the best way to represent the data was challenging as there are a wide range of options.

Conclusion

- Temperature shows a clear seasonal trend
- Anomalies highlight days with unusual temperature possibly due to extreme weather events.
- Rainfall data shows low daily amounts with occasional significant rainfall events.
- Visualizations make complex datasets interpretable, allowing patterns and outliers to be easily identified
- Showing the relationships between the data is clear with the graphs chosen.

It is obvious how visualizations can be utilized to easily understand what would be complex data.

