The data set that will be used for this assignment is from Kaggle. The data set contains daily weather data dating back to January 2, 1833. There are different weather features such as average temperature, precipitation, snowfall, and windspeed. Using this data, an analysis can be conducted on the cities with the highest temperatures and precipitation. With that, another analysis can be done on the correlation between highest temperatures and precipitation. The other variables can be analyzed as well to see where there is a strong correlation.

A graph with numbers and symbols

Description automatically generatedTo begin analysis, the daily data will be condensed. In condensing the data, the weather information can be grouped by city and the average can be used for the records of each city. These averages can be compared against each other to identify the cities with the highest average temperature and highest precipitation. To being analyzing the data, a boxplot will provide information briefly.

A graph of a graph

Description automatically generated with medium confidenceThe box plot above shows the range of values for each numeric column. While the temperature features do not have many outliers, many of the other columns do. The next analysis will be the identification of cities with highest temperature and precipitation. This will be the initial exploration conducted.

The histogram created above shows a distribution of temperatures and which are seen more frequently. The cities that will have the highest temperature will most likely be around the 30-degree area while most of the data lies between 10 and 25 degrees.

A graph with blue bars

Description automatically generatedThe next histogram shows precipitation values. While most values lie on the lower end, there are instances of where precipitation is around 23-24 MM, our highest city will be around this value. Most of the other instances are between 0 and 5 MM. Next, the top 5 cities with highest temperature and precipitation will be identified.

A graph of blue rectangular bars

Description automatically generatedWith the data sorted, the cities with the highest average temperature can be seen in a bar graph. As expected, these cities are all near 30 degrees Celsius. The warmest city being Faya-Largeau.

The cities with the highest precipitation are displayed above. These cities have precipitation levels ranging from 13 MM to 24 MM. The city with the most precipitation being Sittwe.A graph of blue bars

Description automatically generated Next, an analysis is conducted on the correlation between the two. This can be done using a scatter plot.

A graph of blue dots

Description automatically generated

As seen in the scatter plot, there is a slight correlation between temperature and precipitation. The points are ascending and shifting to the right side in a positive direction. To further understand correlation, all the columns can be analyzed to understand their relation to one another.

A screenshot of a graph

Description automatically generated

The data provides a representation of weather of different cities over time, and it shows that the warmest city is Faya-Largeau and the city with the most precipitation is Sittwe. This is based on the average over time with the data provided. The bar graphs created demonstrates this for both cities. When understanding the correlation between temperature and precipitation, the scatterplot shows that there is a slight correlation, and the correlation matrix confirms this. Since the correlation coefficient of temperature and precipitation is 0.47, this indicates a moderate correlation between the two.