**Project 1**

Project Title: Weather Data Analysis (Entry Level)

1. Descriptive Statistics: - Compute and present basic statistics (mean, median, standard deviation) for Min Temp, Max Temp, Rainfall, and Evaporation.

MEAN: -

Minimum Temperature: 7.27

Maximum Temperature: 20.55

Rainfall: 1.43

Evaporation: 4.52

MEDIAN: -

Minimum Temperature: 7.45

Maximum Temperature: 19.65

Rainfall: 0.00

Evaporation: 4.20

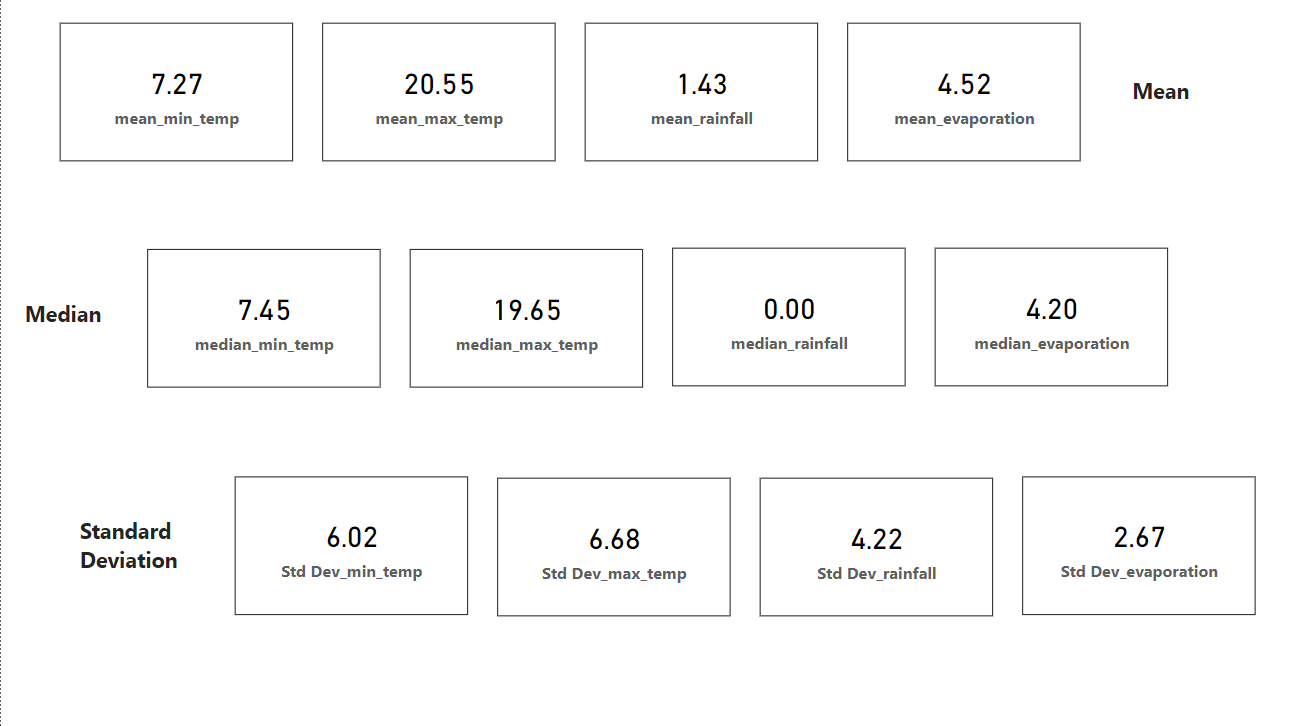
STANDARD DEVIATION: -

Minimum Temperature: 6.02

Maximum Temperature: 6.68

Rainfall: 4.22

Evaporation: 2.67



2. Time Series Visualization: - Create a line chart to show the variations in weather variables over time, identifying trends or patterns.

For this time series visualization, I had taken 4 weather variables they are

1. Temperature (9am and 3pm)

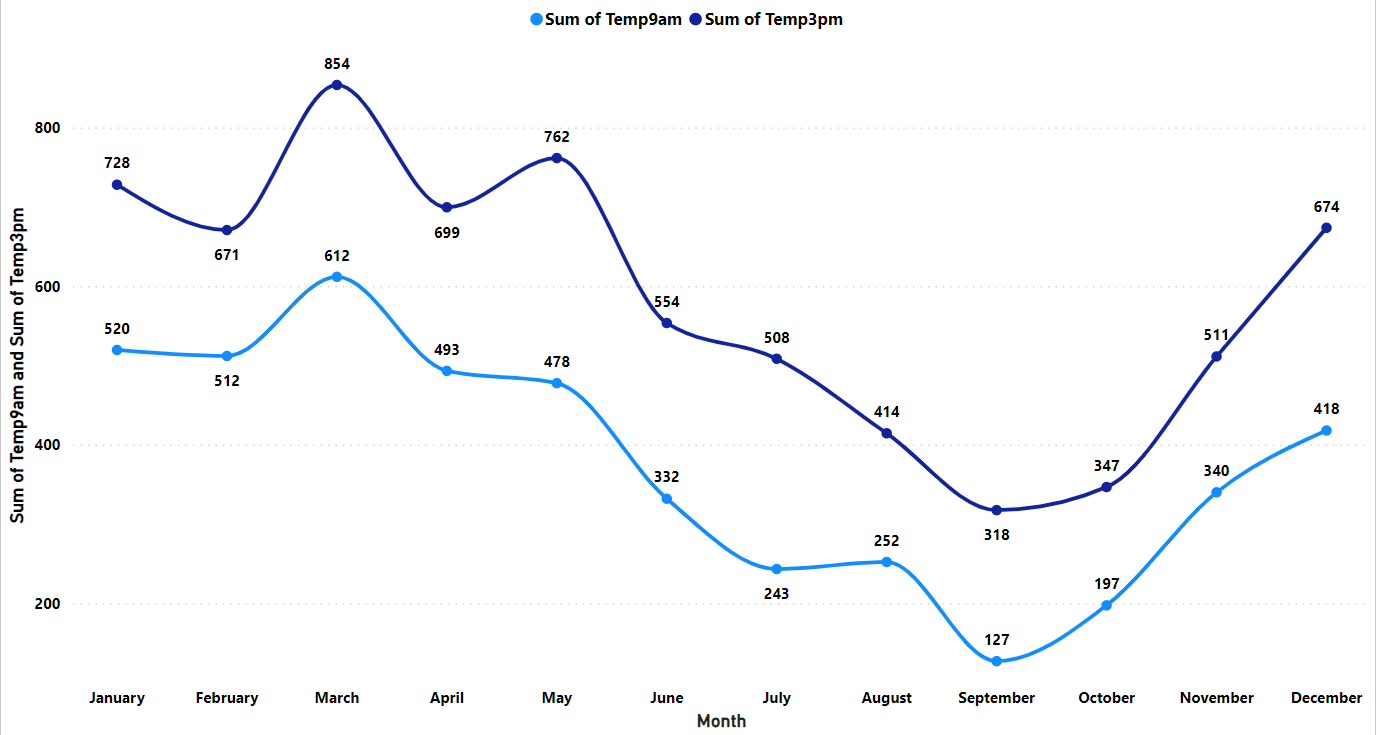
2. Pressure (9am and 3pm)

3. Humidity (9am and 3pm)

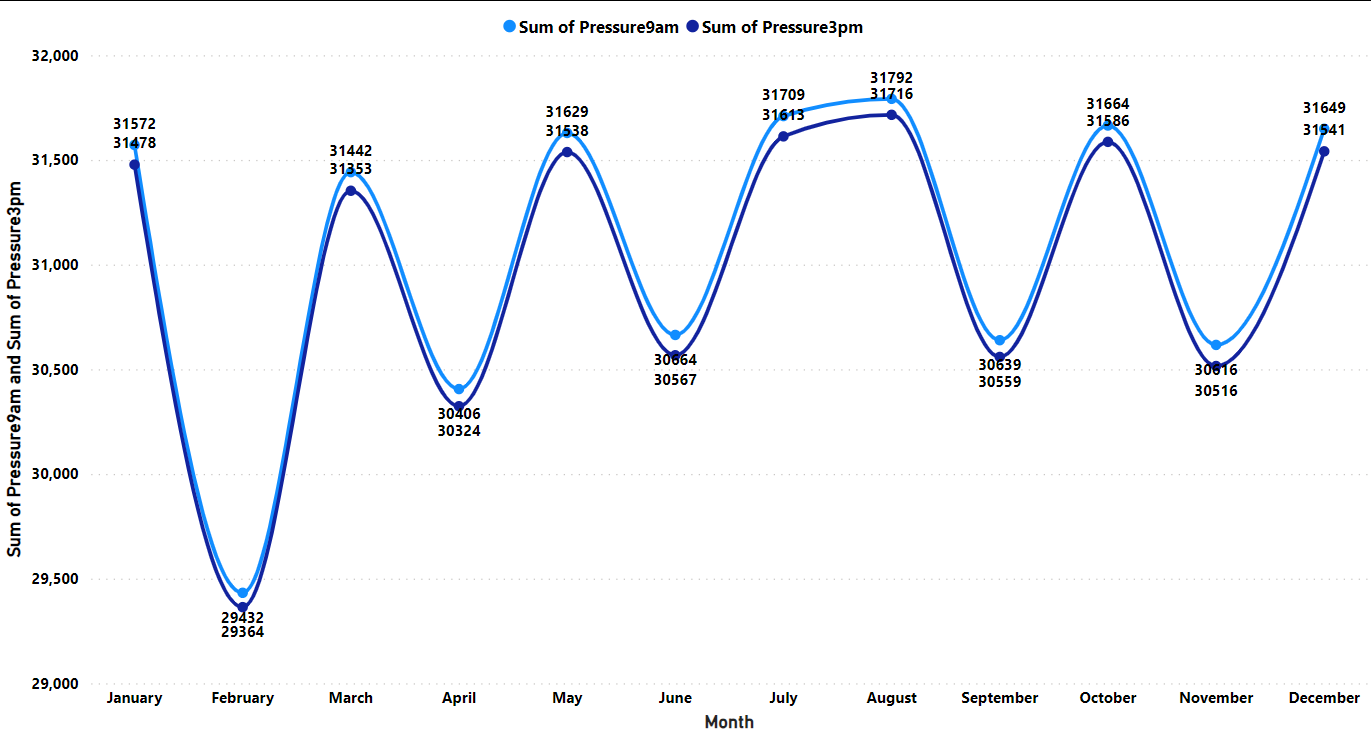
4. Cloud (9am and 3pm)

To achieve this, I created a date table by using power query and extracted month from the date table and showed variation of weather variables over time with respected to month and used them as my x axis for the line chart

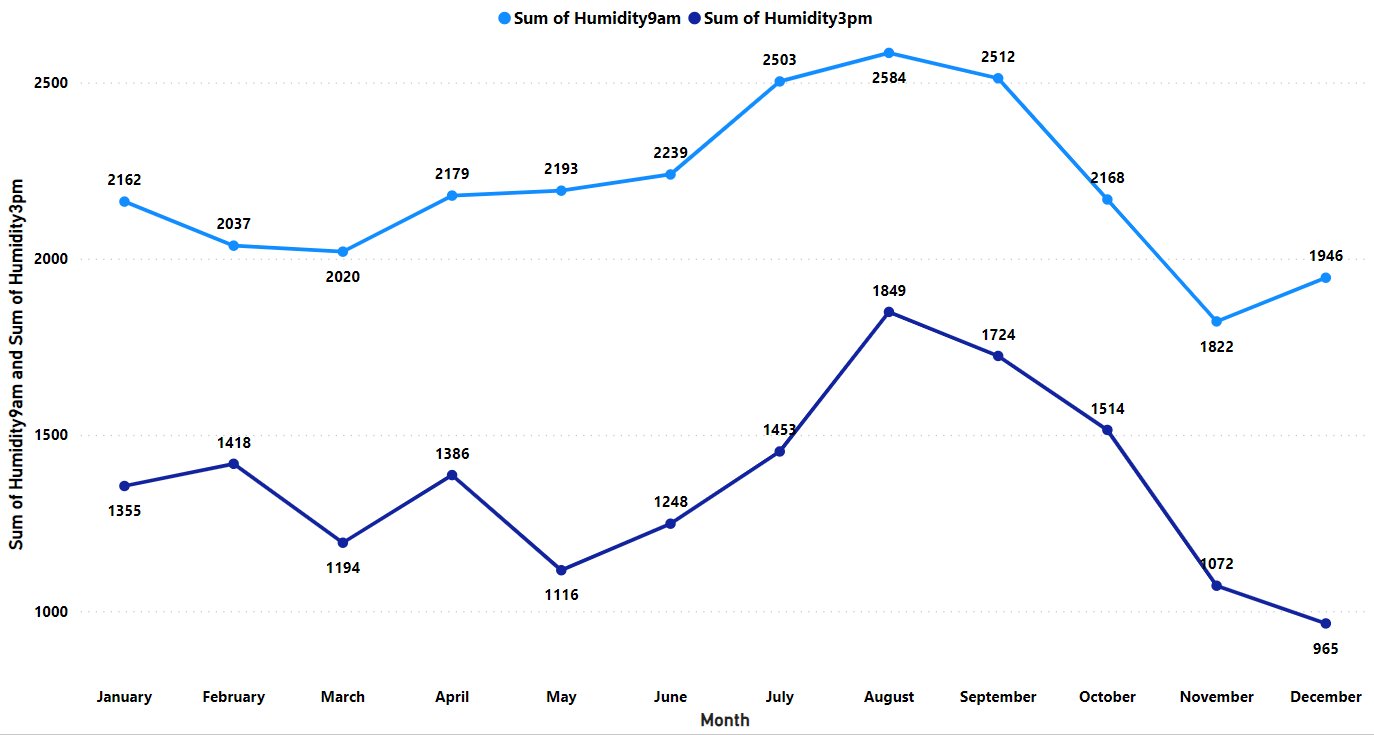
1.Temperature (9am and 3pm).



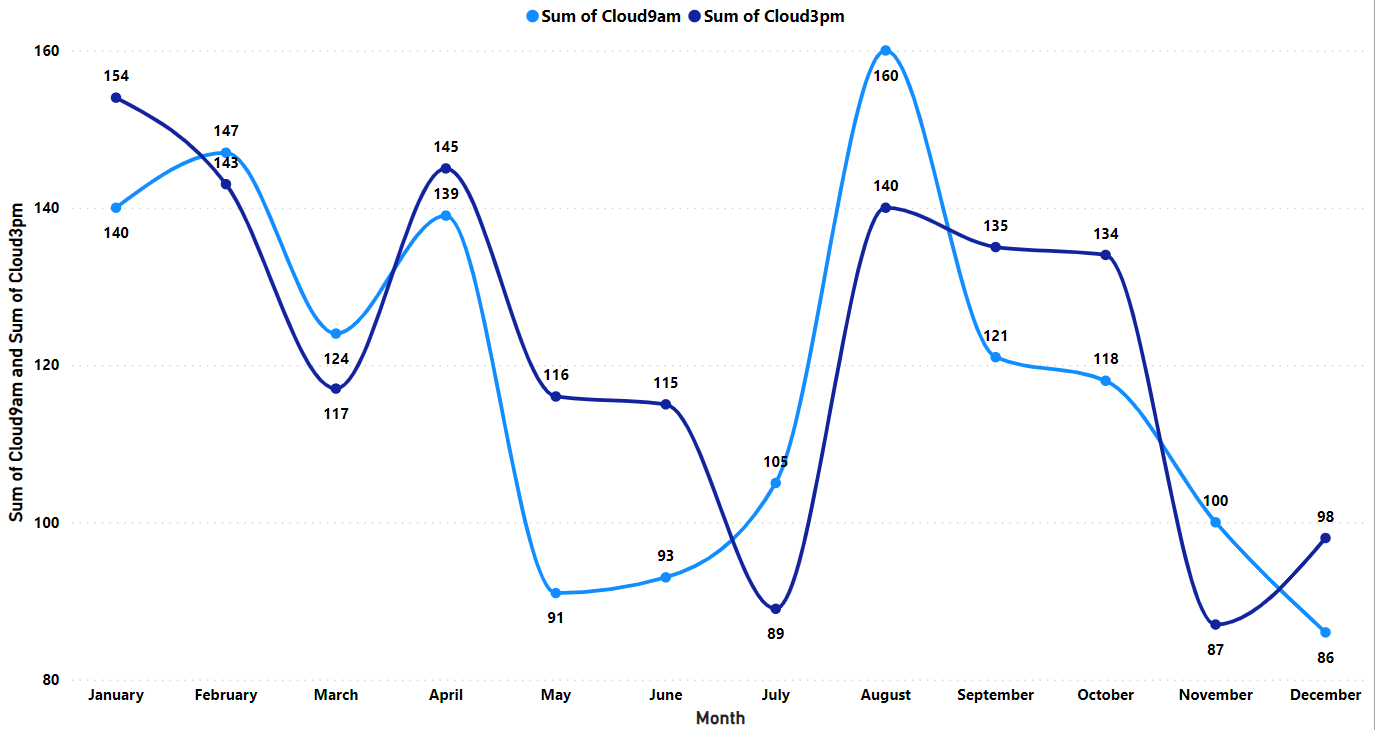
2. Pressure (9am and 3pm)



3.Humidity (9am and 3pm)



4. Cloud (9am and 3pm)



3. Correlation Analysis: - Calculate and visualize correlations between Min Temp, Max Temp, Rainfall, and Evaporation using a heatmap.

For this correlation analysis I have used Python visual which is embedded in power BI. I had written python to code to calculate correlation analysis between Min temp, Max temp, Rainfall, Evaporation

Python Code:

# dataset = pandas.DataFrame(MinTemp, MaxTemp, Evaporation, Rainfall)

# dataset = dataset.drop\_duplicates()

# Paste or type your script code here:

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sn

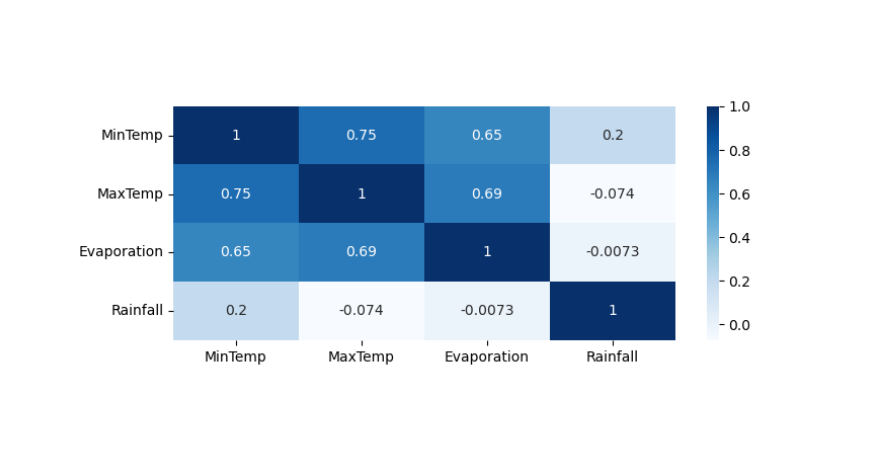
print(dataset.corr())

plt.figure(figsize=(8,3))

dataplot = sn.heatmap(dataset.corr(),cmap="Blues",annot=True)

plt.show()

Heatmap:



4. Rainfall Distribution: - Illustrate the distribution of rainfall through a histogram or kernel density plot, highlighting common levels and outliers.

For this RainfalI Distribution i have used Python visual which is embedded in power BI. I had written python to code to Illustrate Rainfall Distribution by using histogram highlighting common levels and outliers.

Python Code:

# dataset = pandas.DataFrame(Rainfall)

# dataset = dataset.drop\_duplicates()

# Paste or type your script code here:

import matplotlib.pyplot as plt

import seaborn as sns

import numpy as np

plt.figure(figsize=(7,5))

sns.histplot(dataset.Rainfall, bins=30, kde=True, color='blue')

common\_levels = [25, 50, 75]

for level in common\_levels:

    plt.axvline(x=level, color='green', linestyle='--', linewidth=2, label=f'Common Level {level}')

outlier\_threshold = 30-50

outliers = dataset.Rainfall[dataset.Rainfall > outlier\_threshold]

plt.scatter(outliers, [0] \* len(outliers), color='red', marker='o', label='Outliers')

plt.xlabel('Rainfall')

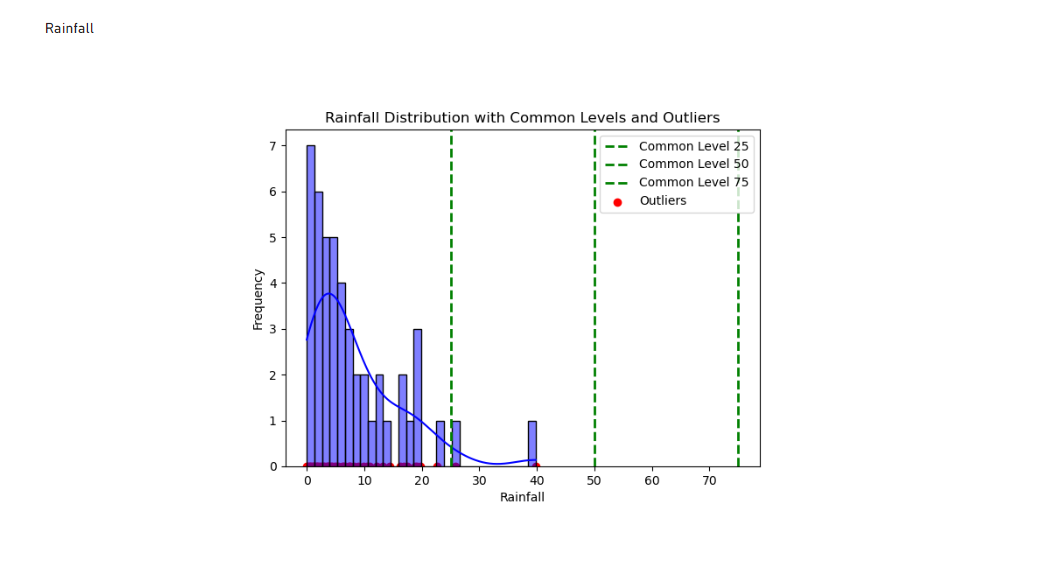
plt.ylabel('Frequency')

plt.title('Rainfall Distribution with Common Levels and Outliers')

plt.legend()

plt.show()

Histogram:



5. Seasonal Analysis: - Analyze average values of weather variables across different seasons and visualize seasonal patterns with bar graphs.

For this time series visualization, I had taken 6 weather variables they are

1. Minimum Temperature

2. Maximum Temperature

3. Evaporation

4. Rainfall

5. Sunshine

6. Wind Gust speed

To achieve this, I created a date table by using power query and extracted month from the date table and used a conditional column to categorize the months with seasons and showed variation of weather variables over time with respected to Seasons and used them as my x axis for the Bar graph

