# Weather Prediction in Seattle: An Exploratory Data Analysis

Group Members:

1. Khushi Rampurkar-2023UG000120
2. Kunal R. Shrivastav-2023UG000132
3. Pruthviraj Shinde-2023UG000103
4. Madhuri Sable-2023UG000122
5. Preethi Verma-
6. Priyanshu Bhargava-

# Table of Contents

---

## Introduction

Weather prediction is a critical aspect of modern life, influencing decisions in agriculture, transportation, energy consumption, and daily planning.This report presents an Exploratory Data Analysis (EDA) of weather data from Seattle, Washington, spanning from 2012 to 2015. The analysis aims to uncover patterns and relationships between various weather parameters, including temperature, precipitation, wind, and weather conditions, to inform the development of a robust weather prediction model. By examining seasonal trends, correlations, and conducting cluster analysis, the report highlights key insights into Seattle's climate and offers recommendations for improving predictive accuracy. This exploration serves as a foundation for creating a reliable weather forecasting system tailored to Seattle's unique weather patterns.

## Problem Statement

The objective of this project is to develop a reliable weather prediction model for Seattle based on historical weather data from 2012 to 2015. By performing an exploratory data analysis (EDA) of key weather parameters such as temperature, precipitation, and wind speed, the goal is to identify significant patterns and relationships that can improve forecasting accuracy. This will address Seattle's unique weather conditions, characterized by frequent rainfall and seasonal temperature variations, and support better decision-making in sectors like agriculture, transportation, and daily planning.

## Data Acquisition

The dataset used in this analysis was obtained from a CSV file containing weather information for Seattle from 2012 to 2015. The data includes daily records of the following parameters:

1. Date
2. Precipitation
3. Maximum temperature
4. Minimum temperature
5. Wind speed
6. Weather conditio

### Data Overview

Seattle, located in the Pacific Northwest of the United States, is known for its mild, temperate climate. The city's weather is heavily influenced by its proximity to the Pacific Ocean and the Puget Sound, as well as its position between the Olympic Mountains to the west and the Cascade Range to the east.

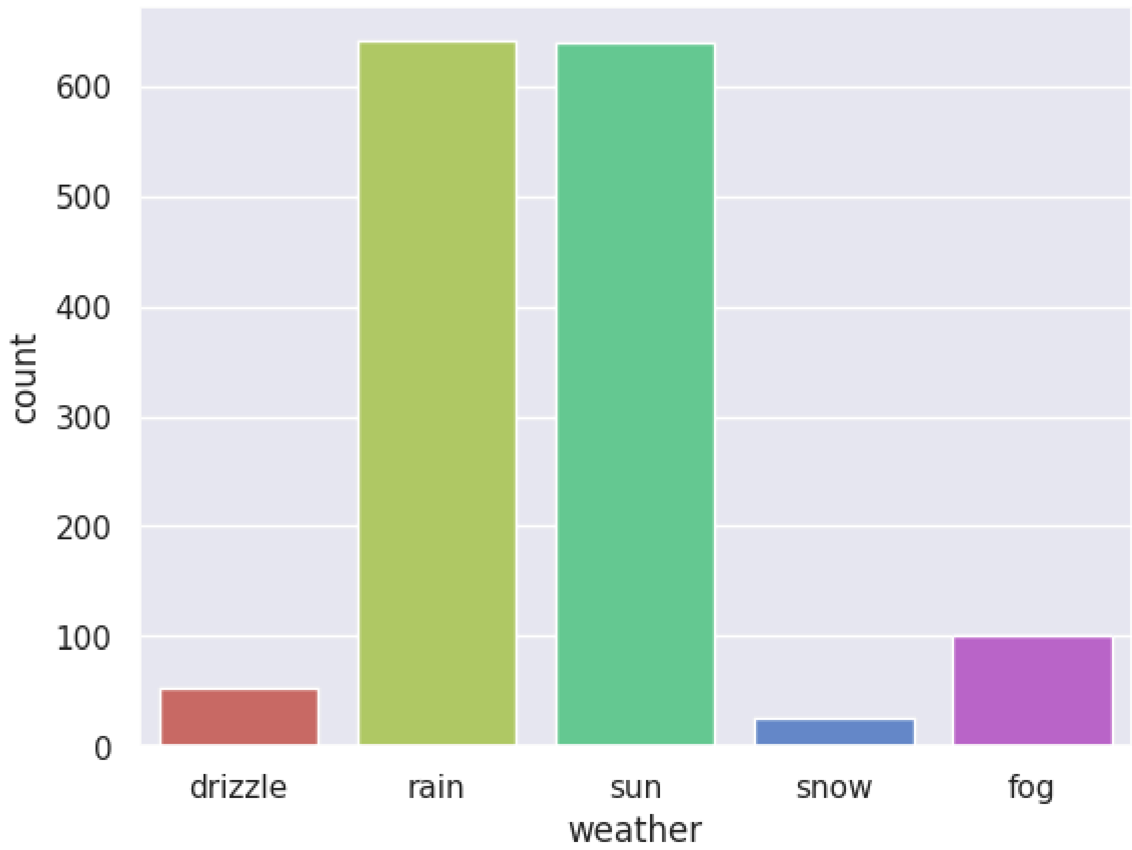
Key characteristics of Seattle's weather include:

1. Mild temperatures year-round, with cool winters and warm summers
2. Frequent cloud cover and light rain, especially during fall, winter, and spring
3. A reputation for rainfall, although the actual precipitation amount is less than many other U.S. cities
4. A dry season from July to early October

Our dataset spans from 2012 to 2015, providing a good representation of Seattle's recent weather patterns.

## Exploratory Data Analysis

### Weather Distribution

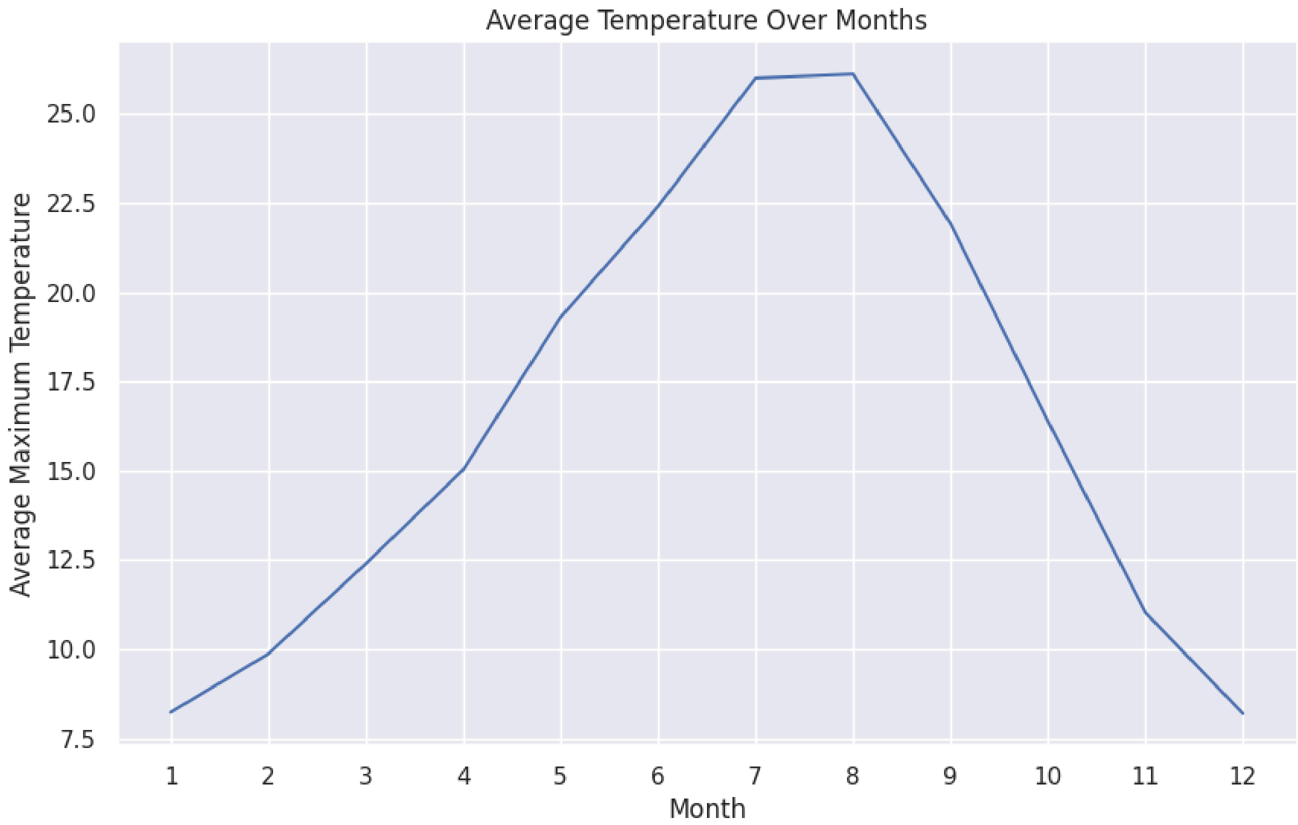



The bar plot above illustrates the distribution of weather conditions in Seattle:

Rain and sun are the most common weather conditions, each accounting for about 44% of the data. Fog occurs in about 7% of the cases.
Drizzle and snow are relatively rare, with snow being the least common at less than 2% of occurrences.

This distribution aligns with Seattle's reputation for frequent rainfall and overcast conditions, but also highlights that clear, sunny days are equally common. The rarity of snow events is notable and may present challenges for prediction models due to the limited number of samples.

# Temperature Pattern

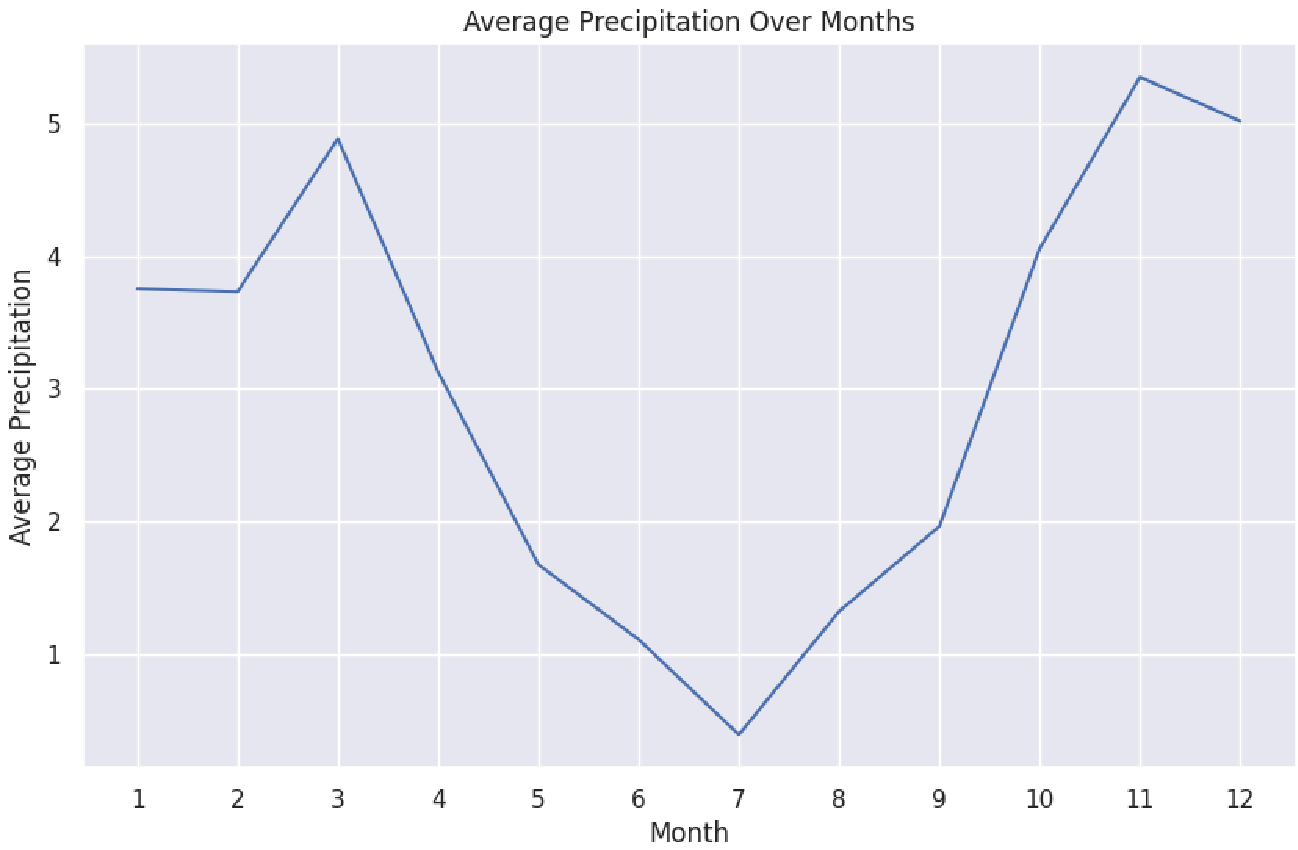



The line plot above shows the average temperature trends throughout the year:

There's a clear seasonal pattern in temperature, with peaks in summer (July-August) and troughs in winter (December-January).
The temperature range is approximately 15°C, from about 7°C in winter to 22°C in summer.
The temperature increase from spring to summer appears more gradual than the decrease from fall to winter.

These patterns reflect Seattle's temperate climate, with relatively mild winters and cool summers compared to many other
U.S. cities. The gradual warming in spring and rapid cooling in fall could be related to oceanic influences and changing atmospheric patterns.

# Precipitation Patterns

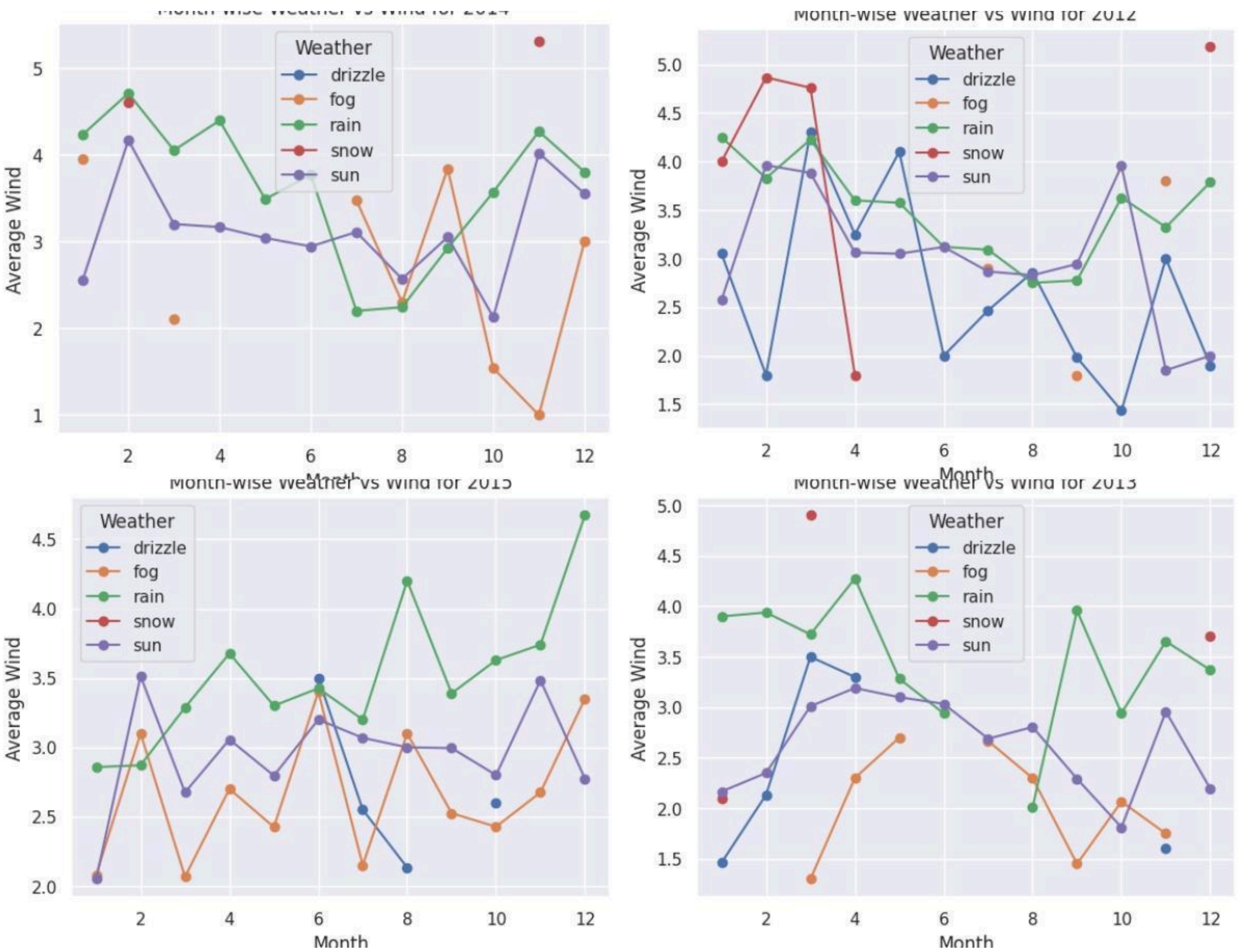



The line plot above demonstrates the average precipitation levels throughout the year:

There's a noticeable seasonal pattern in precipitation.
The wettest months appear to be November and December.
The driest months are typically in the summer, around July and August.

This aligns with Seattle's known dry season in late summer and increased rainfall in late fall and winter. The precipitation pattern is likely influenced by the Pacific Northwest's atmospheric river events, which bring significant moisture to the region in the cooler months.

# Wind Patterns

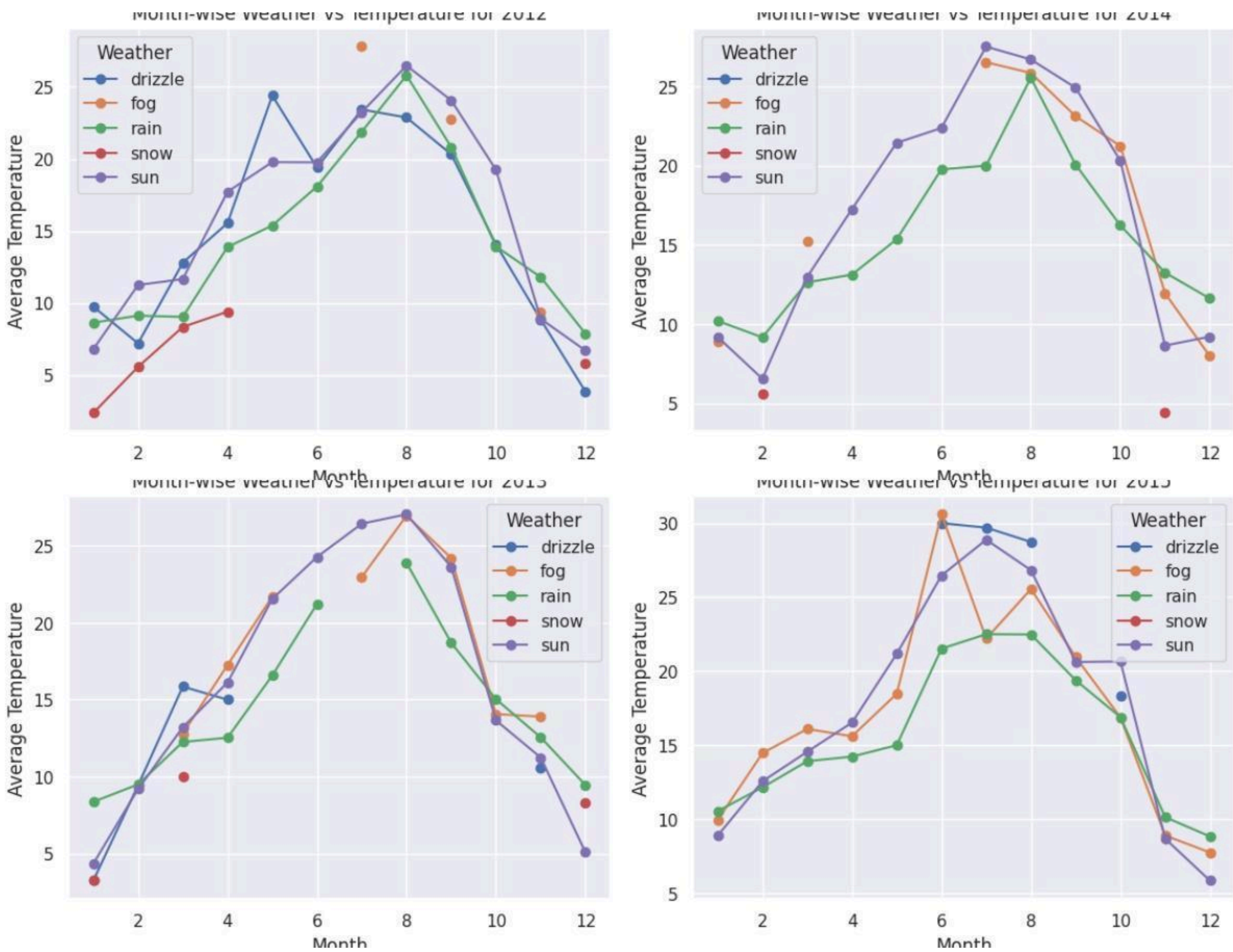



The multiple line plots for each year show how wind speeds relate to different weather conditions:

Wind speeds seem to be highest in winter months, particularly for rainy and snowy conditions. Summer months generally have lower wind speeds across all weather types.
Foggy conditions are often associated with lower wind speeds.

These patterns suggest a seasonal variation in wind speeds, with winter storms likely contributing to higher wind speeds during colder months. The association of fog with lower wind speeds is typical, as fog often forms under calm conditions.

# Weather-Temperature Relationships

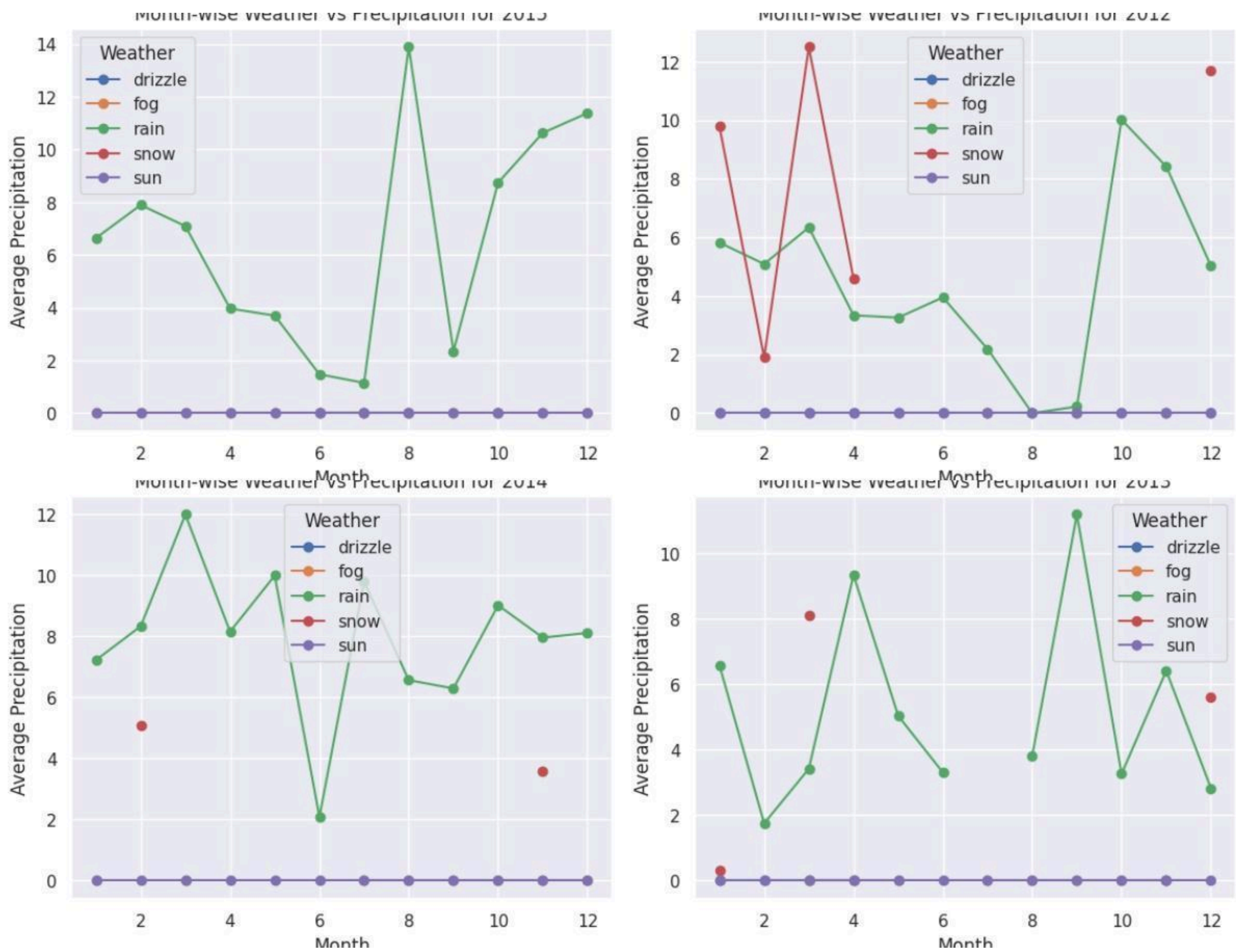



The box plot above illustrates the relationship between weather conditions and temperatures:

Snow consistently occurs at the lowest temperatures across all years. Rain and drizzle tend to occur at moderate temperatures.
Sun is associated with the highest temperatures, especially in summer months. Fog seems to occur across a wide range of temperatures.

These relationships provide valuable insights for prediction models, showing clear temperature ranges for different weather conditions. The wide temperature range for fog suggests it can occur in various seasons, possibly due to different formation mechanisms (e.g., radiation fog vs. advection fog).

# Weather-Temperature Relationships

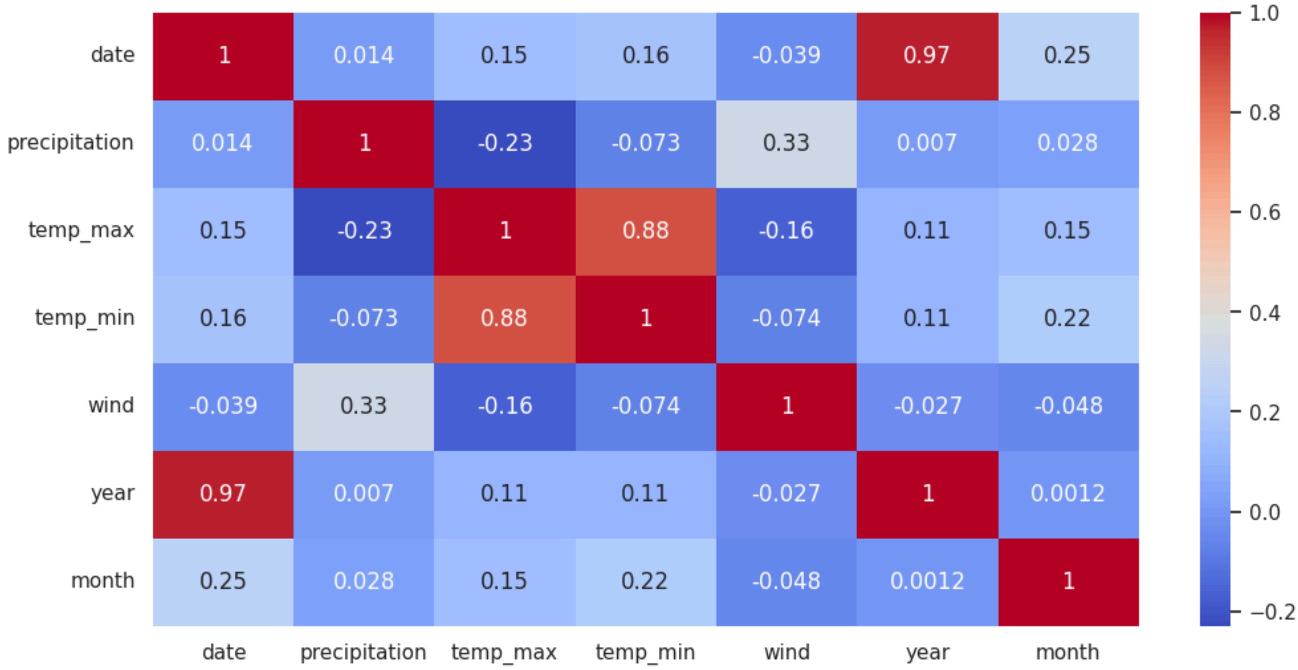



The multiple line plots for each year demonstrate how different weather types relate to precipitation amounts:

Rain is consistently associated with the highest precipitation levels across all years. Snow, when it occurs, is also associated with significant precipitation. Sun and fog are generally associated with lower precipitation levels.

These patterns are crucial for understanding the interplay between precipitation and weather conditions. The high precipitation associated with snow events, despite their rarity, highlights the intensity of these occurrences when they do happen.

The multiple line plots for each year demonstrate how different weather types relate to Correlations

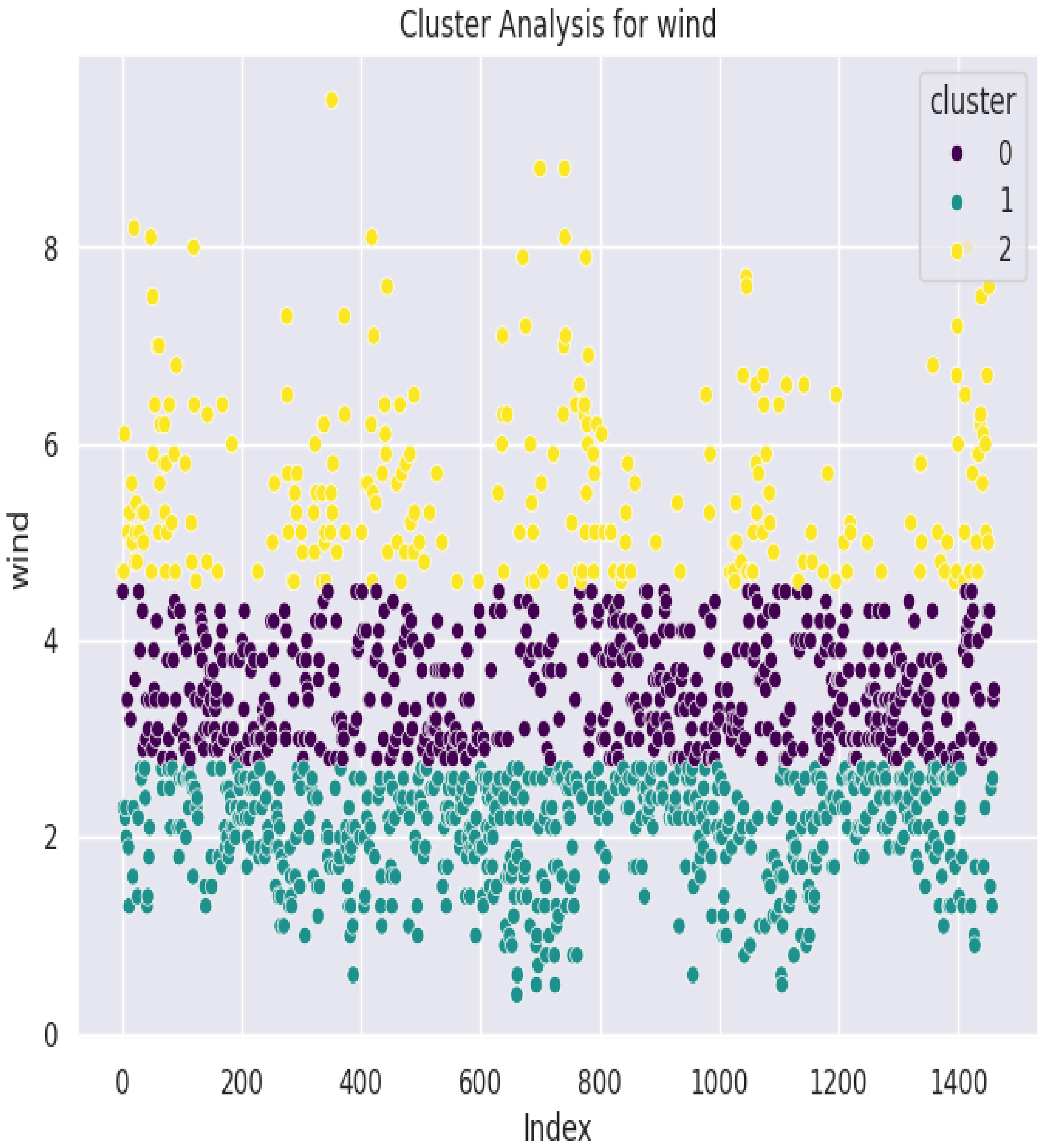


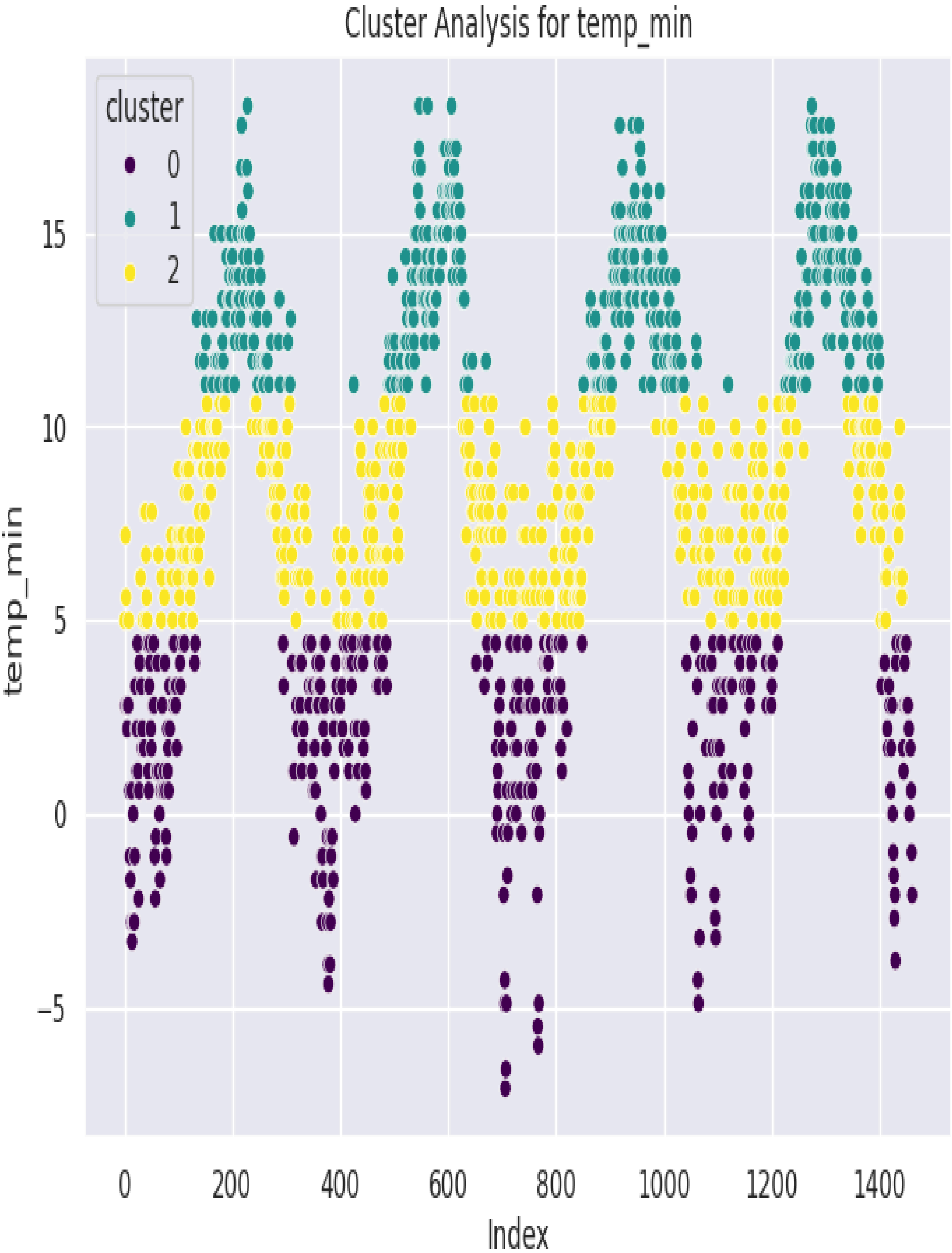

The heatmap above shows the correlations between numerical variables in the dataset:

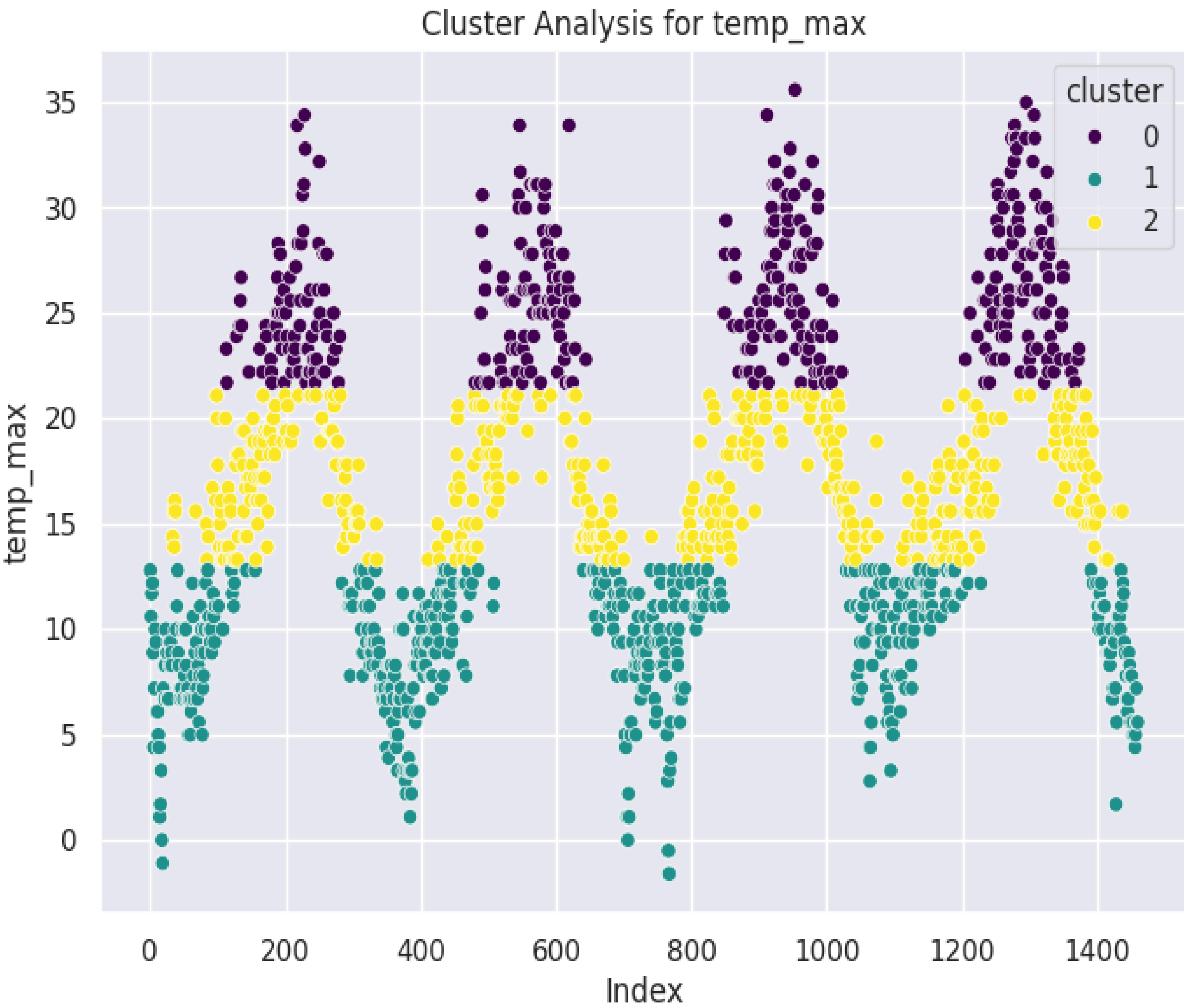
There's a strong positive correlation between maximum and minimum temperatures.
Precipitation shows a weak negative correlation with temperature, suggesting slightly cooler temperatures during rainy periods.Wind speed has very weak correlations with other variables, indicating it might be an independent factor in weather patterns.

These correlations provide insights into the relationships between different weather parameters. The strong correlation between max and min temperatures suggests that daily temperature ranges remain relatively consistent. The weak correlations of wind speed with other variables indicate that it might be influenced by factors not captured in this dataset.

Cluster Analysis for precipitation, temp_max, temp_min, wind.




Cluster Analysis for temp_min
cluster
0
1
2
temp_min
15
10
5
0
−5
Index
0
200
400
600
800
1000
1200
1400




## Cluster Analysis Insights

**1. Precipitation:**

- The clusters likely represent different levels of precipitation.
- Cluster 0: Represents days with low precipitation.
- Cluster 1: Represents days with moderate precipitation.
- Cluster 2: Represents days with high precipitation.
- By analyzing the distribution of other variables within each cluster, further insights can be gained, such as the relationship between precipitation and temperature or wind.

**2. Temperature (temp_max, temp_min):**

- The clusters likely represent different temperature ranges.
- Cluster 0: Represents days with low temperatures.
- Cluster 1: Represents days with moderate temperatures.

- Cluster 2: Represents days with high temperatures.
- Exploring the relationship between temperature clusters and seasonal patterns or weather conditions can provide additional insights.

**3. Wind:**

- The clusters likely represent different wind speed ranges.
- Cluster 0: Represents days with low wind speeds.
- Cluster 1: Represents days with moderate wind speeds.
- Cluster 2: Represents days with high wind speeds.
- Investigating whether wind speed clusters are associated with specific weather events or seasons may be insightful.

## Recommendations for Improvement:

1. Additional Data Collection: Gather data on humidity, air pressure, and cloud cover to potentially improve prediction accuracy.

2. Extended Time Frame: Increase the dataset's time span beyond 2012-2015 to capture long-term climate trends.

3. Address Class Imbalance: Use oversampling techniques or collect additional data for underrepresented weather conditions, particularly snow events.

4. Advanced Modeling Techniques: Experiment with ensemble methods (e.g., Random Forests, Gradient Boosting) or deep learning approaches (e.g., LSTM networks for time series prediction).

5. Incorporate External Data: Include broader regional weather patterns or climate indices (e.g., El Niño/La Niña) that might influence Seattle's weather.

6. Feature Engineering: Create new features such as season, day of the week, or holiday indicators to capture additional patterns.

7. Granularity of Data: Obtain hourly data instead of daily data to capture more detailed weather patterns.

8. Geospatial Data: Incorporate elevation data or distance from water bodies to account for local weather variations within the Seattle area.

## Conclusion

This exploratory data analysis has revealed significant insights into Seattle's weather patterns from 2012 to 2015. The clear seasonal trends in temperature and precipitation, the distribution of weather conditions, and the relationships between various weather parameters provide a strong foundation for developing a weather prediction model.

By implementing the recommended improvements and leveraging these insights, we can develop a more accurate and robust weather prediction model for Seattle. This improved model could provide valuable information for various sectors including urban planning, agriculture, and daily life in the Puget Sound region.

The next steps would involve implementing some of these recommendations, particularly addressing the class imbalance and incorporating additional relevant features. With these enhancements, we can move towards building and testing predictive models, ultimately working towards a reliable weather forecasting system for Seattle.

---

# **Reference**:

## GitHub Link:

https://github.com/KunalShrivastav98/Weather_Prediction-EDA-Project.git