

Analysis Report for Seattle Weather and Global Weather Data

Prepared by: Almas Ali Pinto (23123639)

1. Introduction

Understanding weather patterns is crucial for various industries, from agriculture to urban planning. This analysis focuses on two key datasets: the Seattle Weather Prediction Dataset and 380,000 Weather Data Points with specific states. I aim to explore, analyze, and uncover significant trends or patterns to better understand weather dynamics in Seattle and other USA locations.

Objectives

1. Analyze trends in weather data (temperature, precipitation, snowfall.).
 2. Compare weather patterns across geographical locations.
 3. Provide a correlation heatmap of weather data.
-

Used Data

Seattle Weather Prediction Dataset

- **Source:** <https://www.kaggle.com/api/v1/datasets/download/petalme/seattle-weather-prediction-dataset>
- **Structure:**
 - Variables: Date, Temperature, Precipitation, Humidity, Wind Speed, etc.
 - Domain: Weather metrics specific to Seattle over multiple years.
- **License:** Ensure compliance with dataset-specific usage rights.

380,000 Weather Data Points

- **Source:** <https://www.kaggle.com/api/v1/datasets/download/pinto391/380000-weather-data>
- **Structure:**
 - Variables: Date, Location, Temperature, Wind Speed, Precipitation, etc.
 - Domain: Weather metrics from various USA locations over an extended timeframe.
- **License:** Adhere to data license terms.

Data Cleaning and Preparation:

- Cleaned missing data by filling numeric fields with medians and removing redundant entries.
- Standardized temperature units (Celsius/Fahrenheit).
- Unified date formats using `pandas.to_datetime`.
- Normalized categorical labels for "Weather Condition" fields across datasets.

Analysis

Methods

1. Temperature analysis:

- Identified the month that has the highest and lowest average temperature.

2. Weather analysis:

- Analyzed the average(temperature, precipitation, and snowfall) by state.

3. Precipitation analysis:

- Analyzed the USA vs Seattle average precipitation trends.

4. Correlation weather data:

- Visualized a correlation heatmap between various weather aspects.

Results

Temperature analysis:

The warmest months are July and August where the average temperature is 27.5° C and the coolest month is January where the average temperature is 7.5° C.

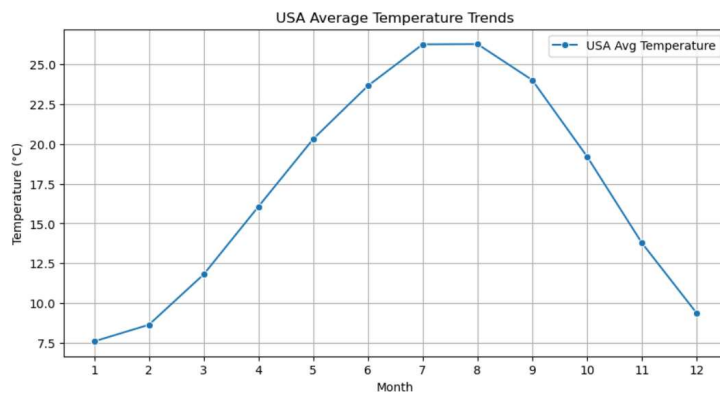


Figure 1: USA average temperature trends.

Weather analysis: The warmest states are Miami and Houston (the average temperature is around 27°C).

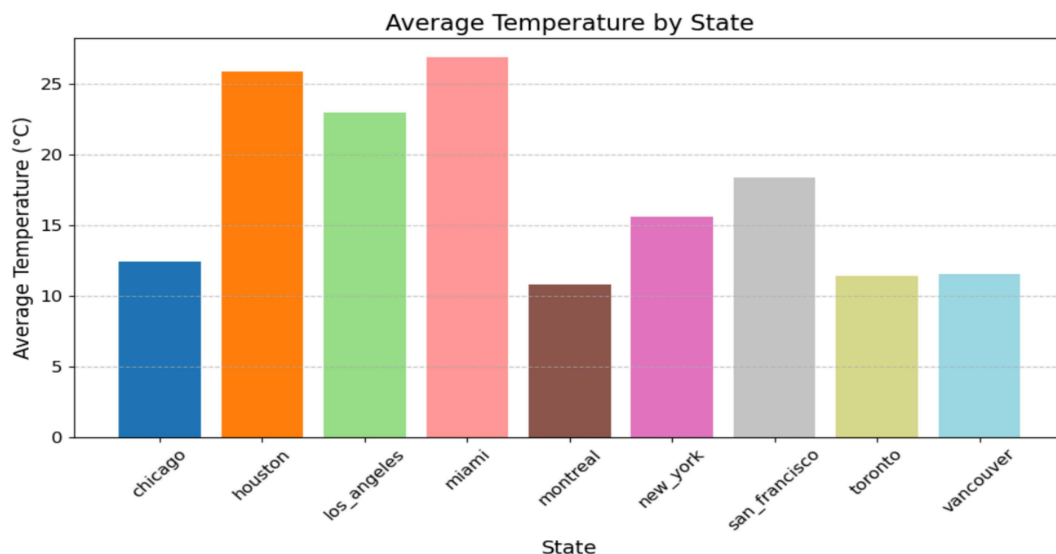


Figure 2: Average temperature by state.

The most rainy state is Vancouver (the average precipitation is around 3.25mm)

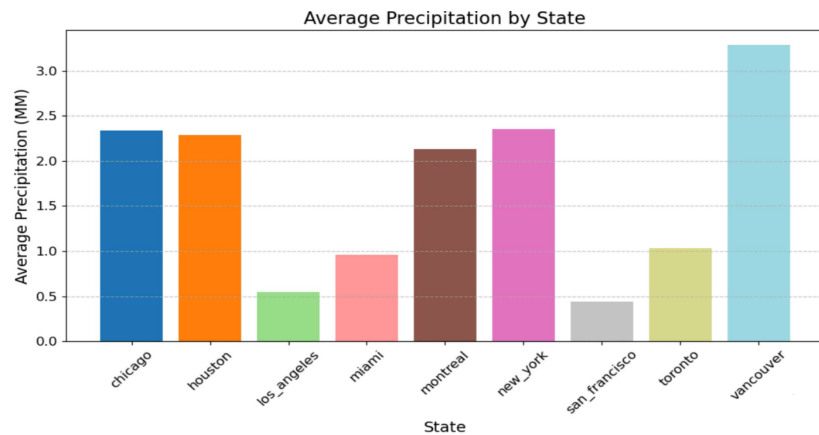


Figure 3: Average precipitation by state.

The most snowfall state is Montreal (the average snowfall is around 0.25cm)

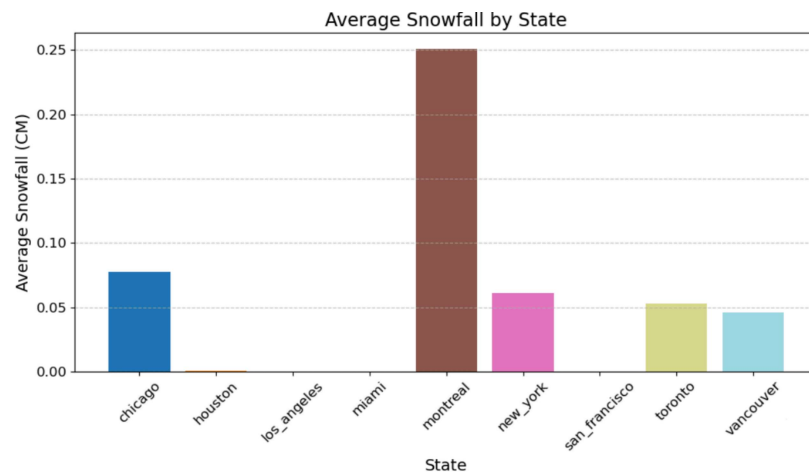


Figure 4: Average Snowfall by state.

Precipitation analysis: The overall USA has less precipitation trends than Seattle state.

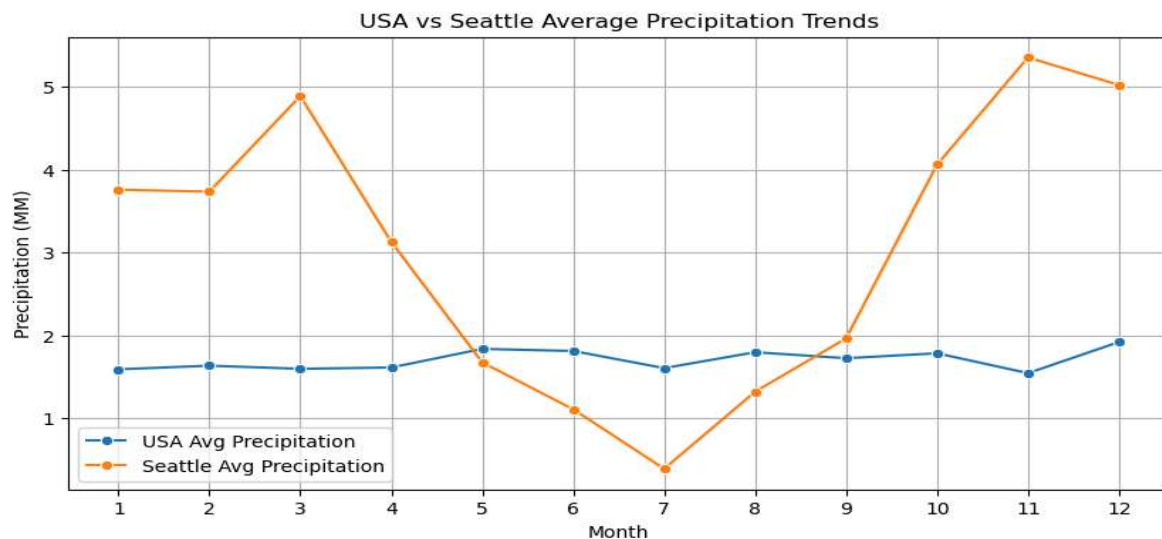


Figure 5: Overall USA vs Seattle average precipitation trends.

Correlation weather data:

The heatmap highlights strong correlations between temperature variables, confirming data consistency. It also reveals a mild relationship between humidity, precipitation, and visibility, aligning with typical weather patterns. Overall, the weak correlations for wind speed and snowfall suggest localized or independent influences on these variables.

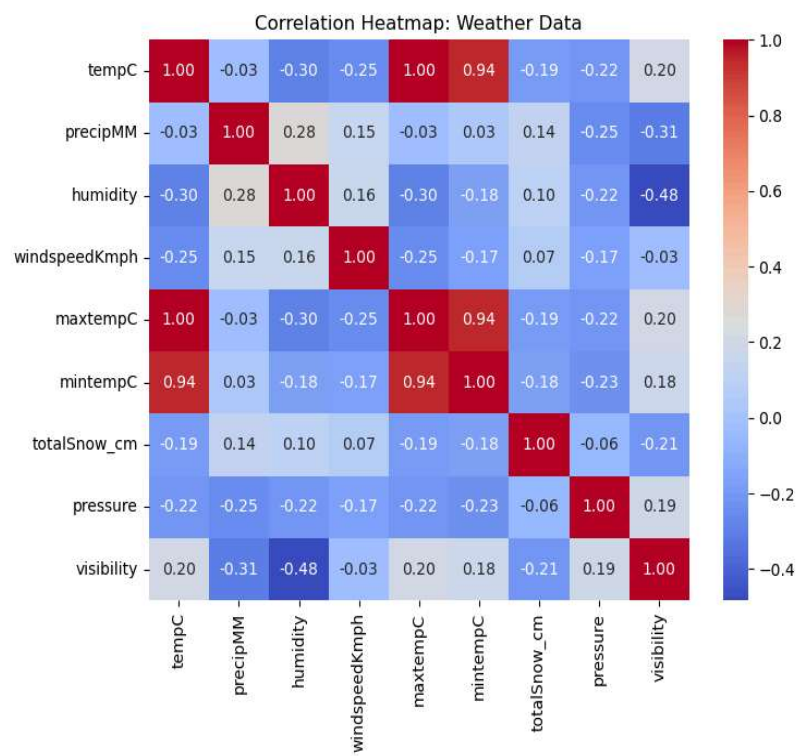


Figure 6: Correlation heatmap of weather.

Conclusions

Key findings:

1. Seattle experiences consistent precipitation throughout the year and also has higher overall precipitation compared to the USA average.
2. Montreal experiences the most snowfall, with an average of 0.25cm, aligning with its colder, snow-prone climate.
3. Seasonal trends indicate significant variability in temperature and precipitation across states, underscoring the importance of localized climate analysis.
4. Miami and Houston are the warmest states in the dataset, with an average temperature of around 27°C, reflecting their tropical climates.

Limitations:

- Predictive modeling might miss microclimatic influences due to dataset granularity.
- Long-term trends like climate change impacts are outside the scope of this dataset.