Weather Pattern Analysis: Exploring Atmospheric Data Using Python

Introduction and Background

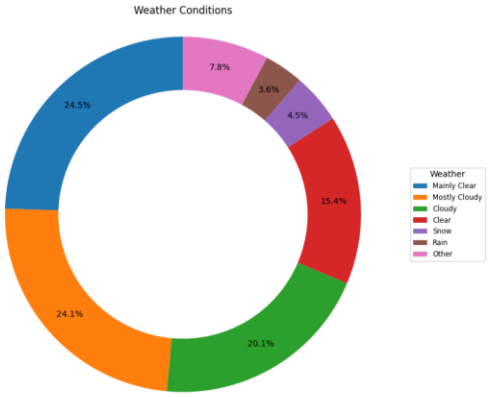
Weather patterns significantly influence public planning, transportation, health, and environmental decisions. In this project, we will focus on exploring hourly weather data collected over a full year (2012) to uncover patterns among temperature, humidity, visibility, and wind speed, and how these are associated with different weather conditions like clear skies, fog, and snow. Using Python's Pandas library, we conducted data cleaning, transformation, and analysis to understand trends in atmospheric behavior. The project emphasizes the importance of descriptive analytics as a foundation before developing predictive machine learning models. Data analysis not only enhances our ability to read and interpret large datasets but also supports the creation of evidence-based assumptions for future decision-making in climate-related fields.

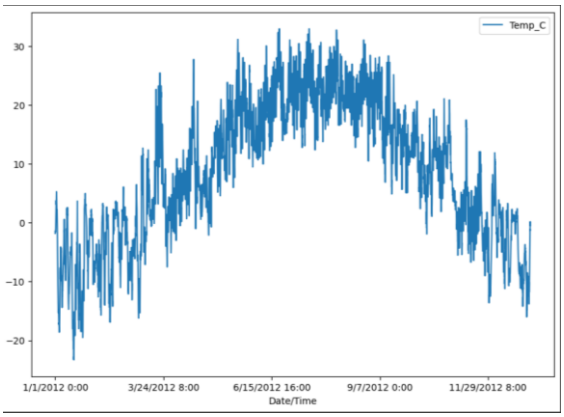
Hypothesis

Our central hypothesis is that specific weather conditions—such as snow, fog, or clear skies—are consistently associated with measurable environmental factors like wind speed, humidity, and visibility. For example, snow is expected to occur at low visibility and high humidity, while clear skies likely align with moderate temperatures and low humidity. The objective was to use data analysis techniques to determine whether these expected patterns truly emerge from the dataset or if surprising correlations exist. Throughout the project, this hypothesis remained consistent, and we aimed to confirm or challenge our assumptions by examining filtered subsets, group statistics, and visual patterns. By doing so, we sought to turn subjective expectations into data-driven insights supported by empirical evidence.

Analysis and Implication

The dataset, sourced from Kaggle, contains hourly weather records for the year 2012. It includes temperature, humidity, wind speed, visibility, pressure, and a textual weather description. We first cleaned the data by removing null values and renamed the “Weather” column to “Weather Condition.” Using Pandas, we explored value distributions, calculated statistical summaries (mean, variance), and identified unique weather types.





Visualizations played a key role. A doughnut chart was created which shows the proportion of weather conditions. Then a line graph was created which illustrates temperature variation over time. Filtering techniques helped us identify the conditions when snow occurred (low visibility, high humidity) and when wind speeds exceeded 30 km/h with above-zero temperatures.

We also grouped the dataset by weather condition and computed averages for variables like pressure, temperature, and visibility. These group-wise comparisons revealed that fog was associated with low visibility and high humidity, while clear skies showed higher visibility and moderate temperatures.

Although this phase emphasized descriptive statistics, it laid the groundwork for future predictive models. Potential response variables include the “Weather Condition,” with predictor variables being wind speed, visibility, and humidity. This foundational analysis offers insight into atmospheric interactions and sets the stage for time-based or categorical predictions.

4. Conclusion

This project supports our hypothesis that consistent relationships exist between weather conditions and key environmental variables. By applying data exploration techniques in Python, we uncovered patterns such as snow aligning with high humidity and low visibility, and clear weather appearing under low humidity and high visibility. These insights confirm our assumptions and highlight the value of exploratory data analysis.

Challenges included handling missing data, interpreting ambiguous weather labels, and managing rare conditions with limited instances. Despite these obstacles, our analysis showed promising results, indicating that larger or multi-year datasets could provide

even deeper insights. Future work could include building classification models to predict weather types based on environmental variables or extending the analysis to other geographical regions.

References

Kaggle Weather Dataset (2012). https://www.kaggle.com

GitHub Repository: https://github.com/shayan139/CS2704-Project.git