Exploratory Data Analysis (EDA) Report: Weather Dataset

1. Introduction

This report presents a comprehensive Exploratory Data Analysis (EDA) performed on a weather dataset. The objective of the analysis is to understand the structure of the data, assess data quality, explore the distribution of meteorological variables, and identify meaningful patterns and relationships across seasons and locations. The analysis follows a systematic, industry-standard EDA workflow.

---

2. Dataset Overview

The analysis begins with an initial inspection of the dataset to understand its size, structure, and basic characteristics.

- **Dataset Shape:** The number of rows and columns is examined to assess the scale of the data.
- **Data Types:** Each column's data type is reviewed to ensure variables are correctly classified as numerical or categorical.

This step confirms that the dataset is suitable for statistical and graphical analysis.

---

3. Data Quality Checks

To ensure the reliability of insights, data quality checks are performed.

3.1 Missing Values

- The dataset is examined for null or missing values.
- No significant missing values are observed, indicating that the dataset is complete and ready for analysis.

3.2 Duplicate Records

- Duplicate entries are identified and checked.
- The dataset contains either no duplicates or an insignificant number, ensuring data consistency.

---

4. Statistical Summary

A descriptive statistical analysis is conducted for numerical variables such as temperature, humidity, wind speed, pressure, visibility, and UV index.

This includes:

- Mean, median, and standard deviation
- Minimum and maximum values

The statistical summary provides an initial understanding of central tendency, dispersion, and range of weather attributes.

---

5. Categorical Variable Exploration

5.1 Weather Types

- Unique weather types present in the dataset are identified.
- Frequency analysis highlights the most and least common weather conditions.

5.2 Seasons

- Distinct seasons included in the dataset are explored.
- The distribution confirms that multiple seasons are represented, enabling seasonal comparisons.

---

6. Univariate Analysis

Univariate analysis is conducted to examine the distribution and characteristics of individual numerical variables.

6.1 Temperature

- Histogram and KDE plots reveal the distribution of temperature values.
- Skewness analysis indicates a **right-skewed distribution**, suggesting the presence of higher extreme temperature values.
- Boxplots identify a few high-end outliers, which represent realistic extreme weather conditions.

Conclusion: Temperature values are mostly concentrated around moderate levels, with occasional extreme highs.

6.2 Humidity

- Distribution analysis shows how atmospheric moisture varies across observations.
- Outlier detection highlights extreme humidity levels, often associated with rainfall or monsoon conditions.

Conclusion: Humidity exhibits variability with noticeable seasonal influence.

6.3 Wind Speed

- Wind speed distribution indicates moderate values with several extreme outliers.
- These outliers represent strong wind or storm events and are retained for analysis.

Conclusion: Wind speed contains valid extreme values, making both mean and median relevant depending on analytical goals.

6.4 Atmospheric Pressure

- Pressure values display relatively low variability.
- Boxplots confirm the absence of extreme fluctuations.

Conclusion: Atmospheric pressure remains fairly stable across observations.

---

7. Bivariate Analysis

Bivariate analysis is performed to study relationships between two variables, particularly between weather attributes and seasons or locations.

7.1 Season-wise Analysis

- **Season-wise Average Temperature:** Bar plots reveal clear seasonal variation, with higher temperatures in summer and lower temperatures in winter.
- **Season-wise Wind Speed:** Certain seasons experience higher average wind speeds, reflecting seasonal wind patterns.
- **Season-wise Humidity:** Humidity peaks during the monsoon season due to increased moisture levels.
- **Season-wise Atmospheric Pressure:** Pressure shows minor seasonal fluctuations, indicating overall stability.
- **Season-wise Visibility:** Reduced visibility is observed during moisture-heavy seasons such as monsoon.

7.2 Location-wise Analysis

- **Location-wise Weather Type:** Count plots show how weather conditions vary geographically.

- **Location-wise Temperature:** Boxplots highlight differences in median temperature and variability across locations.
- **Location-wise Wind Speed:** Regional wind patterns and extreme wind events are clearly visible.
- **Location-wise UV Index:** Some locations experience significantly higher UV exposure, indicating potential health risk zones.

---

8. Key Insights

- Temperature distribution is positively skewed with realistic extreme values.
- Humidity and visibility are strongly influenced by seasonal changes.
- Wind speed shows regional and seasonal variability with meaningful outliers.
- Atmospheric pressure remains relatively stable throughout seasons.
- Geographic location plays a significant role in determining weather patterns and UV exposure.

---

9. Conclusion

This EDA provides a thorough understanding of the weather dataset through structured data inspection, univariate, and bivariate analyses. The findings highlight clear seasonal and geographical patterns in key meteorological variables. These insights can support further predictive modeling, weather classification, or decision-making applications related to climate analysis.

Project Status: *EDA Completed Successfully*
Thank You