

# CHAPTER I-INTRODUCTION

## 1.1 OVERVIEW OF ESTES PARK WEATHER WEBSITE

The **Estes Park Weather** website serves as a comprehensive source for real-time and historical weather data for **Estes Park, Colorado**. It is a valuable platform that provides crucial weather-related insights, including **temperature trends, precipitation levels, wind speeds, humidity, barometric pressure and other meteorological parameters**. This extensive dataset enables researchers, meteorologists, and weather enthusiasts to track climate variations and analyze long-term trends in the region.

One of the **key features** of the Estes Park Weather website is its **real-time weather updates**. The platform continuously monitors and reports current conditions, ensuring that users receive the latest information on **temperature fluctuations, wind patterns, precipitation, and potential severe weather alerts**.

Beyond real-time weather monitoring, the website hosts a **rich historical weather database** with archived reports spanning several years. This historical data is **categorized by date and time**, allowing users to access specific weather records for any given period.

The website presents weather data in **multiple formats**, including **interactive graphs, detailed tables, and downloadable reports**. Users can visualize trends over time, compare seasonal fluctuations, and generate statistical summaries for in-depth analysis.

Another significant aspect of the Estes Park Weather website is its role in **weather forecasting and predictive analytics**. By leveraging past and present meteorological data, the platform helps users make informed predictions about future climate trends. This is particularly useful for understanding **recurring weather patterns, extreme temperature variations, and annual precipitation cycles**.

Given its **extensive database and analytical capabilities**, the Estes Park Weather website is a crucial tool for analyzing local weather patterns, forecasting trends, and understanding the broader impact of climatic variations. This study focuses on extracting and analyzing archived weather data from the website to **identify key meteorological trends, assess seasonal variations, and explore potential climate change indicators** in Estes Park over a specified period.

## 1.2 IMPORTANCE AND RELEVANCE OF THE STUDY

Weather plays a crucial role in shaping natural ecosystems, human activities, and economic sectors such as tourism, agriculture, and infrastructure planning. The **Estes Park Weather website** provides a rich source of meteorological data, making it an invaluable tool for researchers, policymakers, and businesses seeking to understand climate trends and their potential implications. This study aims to leverage the historical and real-time weather data available on the platform to analyze **long-term temperature fluctuations, seasonal variations, and broader climatic changes** in Estes Park, Colorado.

Understanding historical weather patterns is essential for multiple applications, including **disaster preparedness, ecological research, and urban planning**. For instance, long-term temperature and precipitation trends can provide insights into the frequency of extreme weather events such as **heavy snowfall, heatwaves, or droughts**, which can significantly impact the region. By identifying these trends, local authorities can develop **mitigation strategies and adaptive measures** to minimize the risks associated with changing climate conditions.

Furthermore, the study is highly relevant to **environmental research and sustainability efforts**. Climate scientists rely on historical weather data to assess the **rate of warming, changes in precipitation cycles, and shifts in seasonal weather patterns**. These insights contribute to broader discussions on **climate change, carbon footprint assessments, and conservation strategies** aimed at preserving Estes Park's unique environment. Additionally, ecological studies examining the impact of weather patterns on **wildlife habitats, plant life, and water resources** can benefit from the detailed meteorological records available on the Estes Park Weather website.

Another important aspect of this study is its **practical application in tourism and outdoor recreation planning**. Estes Park is a popular destination known for its scenic landscapes and outdoor activities, attracting thousands of visitors each year. By analyzing historical weather patterns, tourism operators and park management can make **data-driven decisions** regarding the best times of the year for various activities, thereby improving visitor experiences and ensuring safety. By examining long-term weather trends, this research will contribute to the growing body of knowledge on climate change and environmental sustainability, helping to inform future planning and conservation efforts in Estes Park and beyond.

### 1.3 OBJECTIVES

This study aims to systematically extract, clean, and analyze historical weather data from the **Estes Park Weather website** using web scraping techniques. The research focuses on identifying **long-term meteorological trends, seasonal variations, and extreme weather events** in Estes Park, Colorado.

The specific objectives are:

1. **Develop a Web Scraping Framework:** Design and implement an automated system to extract archived weather data, including **temperature, humidity, wind speed, precipitation, and atmospheric pressure** for multiple years.
2. **Clean and Preprocess Data:** Apply data cleaning techniques to handle **missing values, inconsistencies, and formatting errors**, ensuring a structured dataset for analysis.
3. **Analyze Long-Term Weather Trends (2014–2024):** Examine variations in **temperature, precipitation, humidity, and wind speed** to assess climate patterns and possible shifts over time.
4. **Assess Seasonal Variations and Extreme Events:** Identify **seasonal weather changes** and detect extreme events like **heatwaves, storms, heavy rainfall, and strong winds**, evaluating their frequency and intensity.
5. **Visualize Meteorological Trends:** Use **graphs, time-series plots** to represent weather data effectively, making analysis more interpretable.
6. **Identify Climate Change Indicators:** Determine whether the dataset provides evidence of **climate change impacts**, such as rising temperatures and increasing extreme weather patterns.
7. **Provide Data-Driven Insights for Decision-Making:** Support **climate researchers, policymakers, urban planners, and tourism stakeholders** by offering insights for **sustainability, disaster preparedness, and environmental conservation** in Estes Park.

## 1.4 SCOPE AND LIMITATIONS

### Scope of the Study:

- This research is primarily centered on collecting **historical weather data** from the Estes Park Weather website, ensuring a structured and data-driven approach to climate analysis.
- The study will cover archived weather reports over a **specific period**, allowing an in-depth examination of meteorological parameters such as **temperature, precipitation, wind speed, and humidity**.
- The extracted data will be processed, cleaned, and analyzed using **statistical and visualization techniques**, enabling a comprehensive understanding of climate trends and seasonal variations.
- The study will explore **correlations between different weather variables** to assess their impact on local climate patterns and environmental conditions in Estes Park.
- The findings from this research aim to support decision-making processes in areas such as **climate policy, environmental conservation, tourism planning, and disaster management**.

### Limitations of the Study:

- The study is **dependent on the accuracy and reliability** of the data available on the Estes Park Weather website. Any **errors, missing records, or inconsistencies** in the archived data may affect the quality of the analysis.
- The **time period covered** in the study is limited to the availability of historical weather data on the source website. Data gaps in certain years may restrict the ability to analyze continuous long-term trends.
- This research is **geographically specific to Estes Park** and may not be directly applicable to other regions with different climate conditions, elevation levels, or meteorological patterns.
- The study focuses on **historical weather data** and does not incorporate real-time forecasting models or predictive analytics beyond trend analysis.
- External factors such as **urbanization, deforestation, and human activities** that might influence local climate conditions are beyond the scope of this study.

## **CHAPTER II-DATA UNDERSTANDING**

### **2.1 INTRODUCTION**

Data understanding is a fundamental step in the data analysis process. It involves exploring, describing, and assessing the quality of the dataset before proceeding with data preparation and modeling. A comprehensive understanding of the dataset ensures that the data used for analysis is accurate, relevant, and free from inconsistencies. Without a strong grasp of the dataset, errors in interpretation and misleading conclusions can arise. This phase allows analysts to detect missing values, anomalies, and outliers that may impact the integrity of the results.

A well-structured data understanding phase also helps in defining the relationships between variables, identifying trends, and ensuring that the dataset aligns with the objectives of the study. By thoroughly exploring the dataset, we can establish a data-driven approach that enhances the reliability of predictions and insights. This chapter provides an in-depth exploration of the dataset collected from the Estes Park Weather website, including data sources, the collection process, attribute descriptions, and initial observations. The insights gained from this chapter will guide the subsequent steps in data preprocessing and analysis, ensuring that the dataset is well-prepared for model development.

### **2.2 WHY DATA UNDERSTANDING?**

Data understanding is crucial because it lays the foundation for effective data analysis and modeling. Without a thorough grasp of the dataset, incorrect assumptions and misinterpretations can arise, leading to flawed conclusions. This phase is particularly important in ensuring data quality by identifying missing values, inconsistencies, and outliers that could distort analytical results.

Understanding the dataset also plays a critical role in feature selection and engineering. Knowing which variables are most relevant helps improve the performance of models, ensuring that only meaningful attributes are used. Additionally, recognizing patterns and seasonality within the data enables better forecasting and trend analysis, providing valuable insights into the underlying phenomena.

Another key aspect of data understanding is the identification of biases in data collection. If biases exist in how the data was gathered, they can lead to skewed results and unreliable predictions. By thoroughly assessing the dataset, analysts can detect and mitigate these biases before proceeding with analysis. Moreover, understanding data structures enables efficient preprocessing, such as handling missing values, transforming data formats, and performing feature scaling to improve model accuracy.

Finally, a strong data understanding process ensures that the dataset aligns with the objectives of the study. It helps validate whether the collected data is sufficient for the intended analysis or if additional data sources need to be considered. This step is essential for making data-driven decisions and ensuring the reliability of findings.

## **2.3 OVERVIEW OF THE DATASET**

The dataset used in this study consists of historical weather data collected from the Estes Park Weather website. The data spans multiple years and includes various meteorological parameters that help in understanding climate patterns in the region.

The dataset is comprehensive, capturing daily weather conditions, including temperature fluctuations, precipitation levels, humidity variations, wind speed, and atmospheric pressure.

This dataset is essential for analyzing long-term weather trends and assessing changes in climate conditions over time.

By leveraging historical weather data, researchers can identify seasonal variations, detect extreme weather events, and evaluate long-term climatic shifts. Additionally, this dataset can be used for predictive modeling to forecast future weather conditions based on past trends.

The availability of structured and well-documented data allows for in-depth statistical analysis, helping to uncover patterns that influence local and regional climate.

The dataset serves as a valuable resource for both academic research and practical applications, such as disaster preparedness, environmental monitoring, and urban planning.

## 2.4 DATA COLLECTION

Data collection is the systematic process of gathering and measuring information from various sources to obtain accurate insights for analysis, decision-making, and research. It involves identifying relevant data points, extracting and recording information, and ensuring the collected data is accurate and complete. Data collection can be conducted using different methods such as surveys, observations, experiments, interviews, web scraping, and automated data logging from sensors or databases.

In today's digital era, data collection has become a critical aspect of various fields, including business analytics, healthcare, social sciences, finance, and environmental studies. It enables researchers, analysts, and organizations to study patterns, test hypotheses, and develop data-driven solutions. The effectiveness of any analytical or research process is directly dependent on the quality and reliability of the collected data.

## 2.5 IMPORTANCE OF DATA COLLECTION

Data collection is fundamental for any research, business strategy, or decision-making process. Without accurate and relevant data, conclusions drawn may be misleading or incorrect. The importance of data collection can be highlighted in several key areas:

1. **Informed Decision-Making:** Data-driven decision-making allows businesses, governments, and researchers to formulate strategies based on factual evidence rather than intuition or assumptions. By analyzing collected data, organizations can optimize operations, improve efficiency, and enhance productivity.
2. **Understanding Patterns and Trends:** Data collected over time helps in identifying trends, correlations, and seasonal variations. For example, businesses can analyze customer behavior patterns, while climate scientists can track long-term temperature changes to study global warming effects.
3. **Ensuring Data Accuracy and Integrity:** High-quality data collection minimizes errors, inconsistencies, and biases. This ensures that analyses are based on reliable information, reducing the risk of flawed interpretations and incorrect conclusions.

4. **Enhancing Research and Innovation:** In scientific research and technology development, accurate data collection supports innovation and discovery. From medical studies to artificial intelligence advancements, high-quality data serves as the backbone of new developments and breakthroughs.
5. **Improving Policy and Governance:** Governments and policymakers rely on data to design and implement policies that address societal needs. Accurate population data, for instance, helps in planning healthcare services, infrastructure development, and educational reforms.
6. **Optimizing Business Performance:** Companies collect data on market trends, customer preferences, and competitor strategies to enhance their operations. Data collection enables businesses to tailor their products and marketing campaigns to meet consumer demands effectively.
7. **Enabling Predictive Analysis:** Predictive modeling and machine learning rely on vast datasets to make accurate forecasts. For example, weather predictions, stock market trends, and disease outbreak forecasts all depend on well-collected historical data.
8. **Risk Management and Fraud Detection:** In finance and cybersecurity, data collection helps in identifying anomalies that could indicate fraud, security breaches, or financial risks. This allows organizations to implement preventive measures before significant damage occurs.

## 2.6 CHALLENGES IN DATA COLLECTION

Despite its importance, data collection comes with challenges that must be managed effectively. These include:

- **Data Privacy and Security:** Ensuring that sensitive information is protected and used ethically.
- **Data Quality Issues:** Avoiding errors such as missing values, duplicate entries, or biased sampling.
- **Scalability and Storage:** Managing large volumes of data efficiently for long-term analysis.
- **Cost and Resources:** Collecting high-quality data often requires significant time, effort, and investment.



## 2.7 DATA COLLECTION METHODOLOGY

The dataset was collected using automated web scraping techniques, ensuring the retrieval of accurate and structured data from the Estes Park Weather website. Web scraping is a method of extracting information from websites using automated scripts, which allows for efficient data collection without manual intervention. The extracted data was stored in both **CSV format** and a **SQLite database** to facilitate further analysis and improve accessibility.

The web scraping process followed a systematic approach to ensure data accuracy and consistency. The key steps involved in the collection process include:

- **Identifying Relevant Historical Weather Pages:** The first step involved identifying web pages containing historical weather records. The Estes Park Weather website provides detailed archives of weather conditions, which were accessed to extract data spanning multiple years.
- **Extracting Structured Data on Daily Weather Conditions:** The scraping script was designed to extract structured weather details such as date, temperature, humidity, wind speed, precipitation, and atmospheric pressure. These variables were captured in a well-defined format to ensure consistency across records.
- **Cleaning and Organizing the Data into Tabular Form:** After extraction, the data underwent a cleaning process to remove inconsistencies, handle missing values, and format it into a structured tabular form. This step ensures that the dataset is ready for analysis without requiring additional preprocessing.
- **Storing the Cleaned Data in Both CSV and SQLite Formats:** To enhance data accessibility, the cleaned dataset was stored in both **CSV files**, which are widely used for data analysis, and **SQLite databases**, which enable efficient querying and retrieval of large datasets. This dual storage approach ensures flexibility in data handling, allowing users to perform both spreadsheet-based and database-driven analyses.

The automated nature of the web scraping process ensured that the data was collected consistently and updated periodically. Additionally, mechanisms were put in place to verify the accuracy of the extracted data by cross-referencing with official weather records. This rigorous approach to data collection guarantees high data quality, making the dataset suitable for advanced statistical analysis, machine learning applications, and climate research.

## 2.8 DATA ATTRIBUTES

The dataset contains 3,809 records with 22 features, each representing a specific weather-related measurement. The key attributes are:

<b>Year</b>	The Year of the recorded data.
<b>Month</b>	The month in which the data was recorded.
<b>Date</b>	The specific day of the recorded weather data.
<b>Average Temperature</b>	The mean temperature recorded on a given day.
<b>Average Humidity</b>	The mean humidity percentage.
<b>Average Dew point</b>	The mean dew point temperature.
<b>Average Barometer</b>	The average atmospheric pressure in inches.
<b>Average Wind speed</b>	The mean wind speed in miles per hour.
<b>Average Gust speed</b>	The mean gust speed in miles per hour.
<b>Average Direction</b>	The average wind direction in degrees.
<b>Rainfall for Month</b>	Total rainfall recorded for the month (in inches).
<b>Rainfall for Year</b>	Total rainfall recorded for the year (in inches).
<b>Maximum Rain per Minute</b>	The highest recorded rainfall per minute.
<b>Maximum Temperature</b>	The highest temperature recorded on a given day.
<b>Minimum Temperature</b>	The lowest temperature recorded on a given day.
<b>Maximum Humidity</b>	The highest recorded humidity percentage.
<b>Minimum Humidity</b>	The lowest recorded humidity percentage.
<b>Maximum Pressure</b>	The highest atmospheric pressure recorded.
<b>Minimum Pressure</b>	The lowest atmospheric pressure recorded.
<b>Maximum Wind speed</b>	The highest wind speed recorded.
<b>Maximum Gust Speed</b>	The highest gust speed recorded.
<b>Maximum Heat Index</b>	The highest recorded heat index.

df

[4]

...

	year	month	date	Average temperature	Average humidity	Average dewpoint	Average barometer	Average windspeed	Average gustspeed	Average direction	...	Maximum rain per minute	Maximum temperature	Minimum temperature
0	2014	1	1	26.9	57.0	13.4	30.1	5.6	8.7	268.0	...	0.00	34.6	15.9
1	2014	1	2	35.2	30.0	6.3	30.2	12.5	17.8	237.0	...	0.00	44.6	18.0
2	2014	1	3	39.4	30.0	10.3	29.7	16.3	23.0	243.0	...	0.00	45.8	31.9
3	2014	1	4	12.0	80.0	7.1	30.0	2.5	4.2	81.0	...	0.00	31.9	-5.4
4	2014	1	5	3.3	71.0	-4.1	30.5	3.3	5.5	346.0	...	0.01	13.4	-5.2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3804	2024	10	24	50.0	33.0	21.7	29.7	0.0	0.0	32.0	...	0.00	58.7	37.0
3805	2024	10	25	39.4	54.0	21.4	30.1	0.0	0.0	32.0	...	0.00	62.7	21.0
3806	2024	10	26	49.5	33.0	17.9	30.1	0.0	0.0	32.0	...	0.00	68.1	25.8
3807	2024	10	27	55.4	21.0	13.2	29.9	0.0	0.0	32.0	...	0.00	69.0	36.8
3808	2024	10	28	55.4	26.0	20.2	29.4	0.0	0.0	32.0	...	0.00	63.7	42.5

3809 rows x 22 columns

df

Python

Average windspeed	Average gustspeed	Average direction	...	Maximum rain per minute	Maximum temperature	Minimum temperature	Maximum humidity	Minimum humidity	Maximum pressure	Minimum pressure	Maximum windspeed	Maximum gust speed	Maximum heat index
5.6	8.7	268.0	...	0.00	34.6	15.9	88.0	33.0	30.383	29.843	18.4	29.9	34.6
12.5	17.8	237.0	...	0.00	44.6	18.0	70.0	19.0	30.338	29.886	24.2	34.5	44.6
16.3	23.0	243.0	...	0.00	45.8	31.9	64.0	20.0	29.902	29.474	33.4	43.7	45.8
2.5	4.2	81.0	...	0.00	31.9	-5.4	87.0	64.0	30.427	29.556	13.8	21.9	31.9
3.3	5.5	346.0	...	0.01	13.4	-5.2	78.0	56.0	30.723	30.340	16.1	26.5	13.4
...	...	...	...	...	...	...	...	...	...	...	...	...	...
0.0	0.0	32.0	...	0.00	58.7	37.0	57.0	25.0	30.001	29.585	0.0	0.0	58.7
0.0	0.0	32.0	...	0.00	62.7	21.0	86.0	20.0	30.268	29.996	0.0	0.0	62.7
0.0	0.0	32.0	...	0.00	68.1	25.8	73.0	13.0	30.268	30.018	0.0	0.0	71.8
0.0	0.0	32.0	...	0.00	69.0	36.8	48.0	8.0	30.083	29.576	0.0	0.0	70.9
0.0	0.0	32.0	...	0.00	63.7	42.5	49.0	19.0	29.581	29.326	0.0	0.0	63.7

## 2.9 BASIC INFORMATION

Understanding the structure and composition of the dataset is crucial before performing any pre-processing or analysis. This section provides an overview of the dataset by examining its first and last few rows, metadata, column details, shape, and summary statistics. These exploratory steps help in identifying missing values, data types, and potential anomalies in the dataset.

By analysing the dataset's structure, we can determine the overall quality and integrity of the data before proceeding with deeper statistical and predictive analysis. The dataset's basic information allows us to:

- **Assess Data Consistency:** By viewing the first and last few rows, we can check for structural inconsistencies or anomalies in the dataset.
- **Identify Data Types:** Knowing the data types (e.g., integers, floats, strings) of each column helps in determining the necessary pre-processing steps, such as type conversion or encoding categorical variables.
- **Detect Missing Values:** A preliminary exploration of the dataset helps in identifying missing values, which can impact model performance if not handled properly.
- **Examine Summary Statistics:** By analysing the distribution of numerical features, we can identify patterns, outliers, and variations in the data.

Furthermore, understanding the dataset's structure aids in feature selection and engineering. By identifying redundant or highly correlated variables, we can optimize model performance and reduce computational complexity. Additionally, this step ensures data integrity and consistency, preventing potential biases in predictive models.

Performing exploratory data analysis (EDA) at this stage provides critical insights into the dataset's strengths and weaknesses. It helps in making informed decisions about data cleaning, transformation, and modeling strategies. This foundational step is essential for developing robust and reliable analytical models that yield meaningful insights and accurate predictions.

## 2.10 IMPORTANCE OF BASIC INFORMATION

Understanding the basic information of a dataset is a critical step in the data analysis process. Before performing any transformations, visualizations, or modeling, it is essential to examine the dataset's structure, composition, and quality. The key reasons why analyzing basic information is important include:

1. **Ensuring Data Integrity:** Examining the dataset's structure helps confirm that the data has been imported correctly and is complete. Identifying inconsistencies early prevents errors in subsequent analysis.
2. **Identifying Missing and Erroneous Values:** A preliminary review of the dataset helps in detecting missing values, outliers, or improperly formatted data. This allows for corrective measures such as imputation or data transformation.
3. **Understanding Data Types:** Checking data types ensures that numerical variables are not mistakenly stored as categorical values or vice versa. This is crucial for operations such as mathematical computations and statistical modeling.
4. **Assessing Dataset Size and Shape:** Knowing the number of rows and columns in the dataset helps determine whether it is manageable for analysis or requires memory optimization techniques.
5. **Recognizing Data Distributions:** Summary statistics provide insights into the spread, central tendency, and variability of numerical features. Identifying anomalies at this stage helps avoid biases in modeling.

Additionally, exploring basic dataset information helps in data cleaning and preparation. By recognizing potential data quality issues early, analysts can take corrective actions such as handling missing values, filtering irrelevant records, and converting variables into appropriate formats. This step also facilitates a smooth transition into advanced analytical techniques, ensuring accuracy and reliability in predictive modeling.

Moreover, understanding dataset characteristics allows analysts to select suitable visualization techniques. Knowing data distributions, categorical proportions, and numerical trends helps in choosing appropriate charts, graphs, and plots for effective data representation. Proper data comprehension ensures that insights derived from visualizations are meaningful and actionable.

## df.columns , df.shape

- **Purpose:** Returns the column labels of the DataFrame as an Index object.
- **Use Case:** Helps to check the names of all columns in the dataset.
- **Purpose:** Returns the dimensions of the DataFrame (rows, columns).
- **Use Case:** Provides insight into the dataset's size

```
df.columns
[42] ✓ 0.0s Python
... Index(['year', 'month', 'date', 'average_temperature', 'average_humidity',
         'average_dewpoint', 'average_barometer', 'average_windspeed',
         'average_gustspeed', 'average_direction', 'rainfall_for_month',
         'rainfall_for_year', 'maximum_rain_per_minute', 'maximum_temperature',
         'minimum_temperature', 'maximum_humidity', 'minimum_humidity',
         'maximum_pressure', 'minimum_pressure', 'maximum_windspeed',
         'maximum_gust_speed', 'maximum_heat_index'],
        dtype='object')

df.shape
[43] ✓ 0.0s Python
... (3809, 22)
```

## df.head() , df.tail()

- **Purpose:** Displays the first five rows of the DataFrame by default.
- **Use Case:** Helps to quickly understand the structure and contents of the dataset.
- **Purpose:** Displays the last five rows of the DataFrame by default.
- **Use Case:** Helps to check the final entries in the dataset.

```
df.head()
✓ 0.0s Python
...
  year  month  date  average_temperature  average_humidity  average_dewpoint  average_barometer  average_windspeed  average_gustspeed  average_direction
0  2014      1      1             26.9             57.0             13.4             30.1             5.6             8.7             268
1  2014      1      2             35.2             30.0              6.3             30.2             12.5             17.8             237
2  2014      1      3             39.4             30.0             10.3             29.7             16.3             23.0             245
3  2014      1      4             12.0             80.0              7.1             30.0             2.5             4.2             81
4  2014      1      5              3.3             71.0             -4.1             30.5             3.3             5.5             346
5 rows x 22 columns

df.tail()
✓ 0.0s Python
...
  year  month  date  average_temperature  average_humidity  average_dewpoint  average_barometer  average_windspeed  average_gustspeed  average_direction
3804  2024    10     24             50.0             33.0             21.7             29.7             0.0             0.0             0.0
3805  2024    10     25             39.4             54.0             21.4             30.1             0.0             0.0             0.0
3806  2024    10     26             49.5             33.0             17.9             30.1             0.0             0.0             0.0
3807  2024    10     27             55.4             21.0             13.2             29.9             0.0             0.0             0.0
3808  2024    10     28             55.4             26.0             20.2             29.4             0.0             0.0             0.0
5 rows x 22 columns
```

## df.info()

- **Purpose:** Provides a concise summary of the DataFrame, including data types, non-null values, and memory usage.
- **Use Case:** Useful for checking data types, missing values, and overall structure.

```
df.info()
[34] ✓ 0.0s Python
...
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3809 entries, 0 to 3808
Data columns (total 22 columns):
#   Column              Non-Null Count  Dtype
---  -
0   year                3809 non-null   int64
1   month               3809 non-null   int64
2   date                3809 non-null   int64
3   average_temperature 3809 non-null   float64
4   average_humidity    3809 non-null   float64
5   average_dewpoint    3809 non-null   float64
6   average_barometer    3809 non-null   float64
7   average_windspeed   3809 non-null   float64
8   average_gustspeed    3809 non-null   float64
9   average_direction   3809 non-null   float64
10  rainfall_for_month   3809 non-null   float64
11  rainfall_for_year    3809 non-null   float64
12  maximum_rain_per_minute 3809 non-null   float64
13  maximum_temperature  3809 non-null   float64
14  minimum_temperature  3809 non-null   float64
15  maximum_humidity     3809 non-null   float64
16  minimum_humidity     3809 non-null   float64
17  maximum_pressure     3809 non-null   float64
18  minimum_pressure     3809 non-null   float64
19  maximum_windspeed    3809 non-null   float64
20  maximum_gust_speed   3809 non-null   float64
21  maximum_heat_index   3809 non-null   float64
dtypes: float64(19), int64(3)
memory usage: 654.8 KB
```

## df.describe()

- **Purpose:** Provides summary statistics for numerical columns.
- **Use Case:** Helps in understanding the distribution, central tendency, and variability of the dataset.

```
Summary Statistics

df.describe()
[41] ✓ 0.0s Python
...

```

	year	month	date	average_temperature	average_humidity	average_dewpoint	average_barometer	average_windspeed	average_gust
count	3809.000000	3809.000000	3809.000000	3809.000000	3809.000000	3809.000000	3809.000000	3809.000000	3809.000000
mean	2018.936204	6.429247	15.779732	44.600945	50.932791	24.182988	29.853295	5.221344	8.000000
std	3.143015	3.408821	8.788394	15.523610	17.533258	14.696371	0.245175	3.736445	5.000000
min	2014.000000	1.000000	1.000000	-10.400000	9.000000	-17.500000	28.200000	0.000000	0.000000
25%	2016.000000	3.000000	8.000000	33.200000	38.000000	13.000000	29.700000	2.400000	4.000000
50%	2019.000000	6.000000	16.000000	45.500000	49.000000	23.500000	29.800000	4.000000	6.000000
75%	2022.000000	9.000000	23.000000	58.000000	63.000000	36.500000	30.000000	7.200000	10.000000
max	2024.000000	12.000000	31.000000	76.300000	94.000000	55.100000	30.900000	21.300000	33.000000

8 rows x 22 columns

## CHAPTER III-DATA PREPARATION

### 3.1 INTRODUCTION

Data preparation is a critical step in any data-driven project. It involves cleaning, transforming, and structuring raw data into a format suitable for analysis and modeling. Raw data is often incomplete, inconsistent, or contains errors that can affect the accuracy and reliability of the results. Data preparation ensures that data is properly processed, structured, and formatted before it is used for decision-making, reporting, or predictive modelling.

Data scientists and analysts spend a significant portion of their time preparing data, as poor data quality can lead to misleading insights and incorrect conclusions. Effective data preparation can significantly improve the efficiency of data analysis, enhance model performance, and provide deeper insights into patterns and trends. This document discusses the importance of data preparation, key techniques, and best practices for ensuring high-quality data.

### 3.2 IMPORTANCE OF DATA PREPARATION

Data preparation ensures that data is:

1. **Clean:** Free from errors, inconsistencies, and missing values, which helps prevent incorrect analyses and biases.
2. **Structured:** Properly formatted for analysis or machine learning, making it easier to extract meaningful insights.
3. **Enriched:** Enhanced with additional features for better predictive power, allowing models to make more accurate predictions.
4. **Standardized:** Maintained in a consistent format across datasets, ensuring seamless integration and analysis.
5. **Optimized:** Reduced in size and complexity for efficient processing, improving computational efficiency and reducing resource consumption.
6. **Reliable:** Data quality issues such as duplicate records, incorrect formats, and outliers are addressed, ensuring that conclusions drawn from the data are valid.
7. **Reproducible:** A well-prepared dataset allows for consistency in analysis across different teams and projects.



### 3.3 DATA CLEANING

Data cleaning, also known as data cleansing or data scrubbing, is the process of identifying and correcting errors, inconsistencies, and inaccuracies within a dataset. It is a crucial step in the data preparation pipeline, ensuring that data is reliable, accurate, and suitable for analysis or machine learning models.

Raw data collected from different sources often contains missing values, duplicates, incorrect formats, and inconsistencies, which can negatively impact data analysis and predictive modeling. Without effective data cleaning, decision-making processes and model performance may be compromised due to incorrect or misleading insights.

Data cleaning involves various techniques, including handling missing values, removing duplicates, correcting errors, detecting and treating outliers, standardizing formats, and resolving inconsistencies. These steps ensure data quality, improve interpretability, and enhance the accuracy of analytical results.

### 3.4 BEST PRACTICES IN DATA CLEANING

1. **Understand the Data:** Conduct exploratory analysis before cleaning.
2. **Follow a Systematic Approach:** Use structured workflows to maintain consistency.
3. **Keep Track of Changes:** Document all modifications for reproducibility.
4. **Validate Cleaned Data:** Cross-check results with domain experts or validation scripts.
5. **Test on a Subset First:** Before applying cleaning procedures to the full dataset, test them on a smaller sample.
6. **Handle Missing Data Appropriately:** Use suitable imputation techniques such as mean, median, or mode replacement, or consider removing rows with excessive missing values.
7. **Remove Duplicates:** Identify and eliminate duplicate records to ensure data accuracy and consistency.
8. **Standardize Formats:** Ensure consistency in date formats, numerical precision, and categorical labels to avoid inconsistencies in analysis.

### 3.5 IMPORTANCE OF DATA CLEANING

1. **Enhances Data Accuracy:** Errors in data, such as incorrect entries, inconsistencies, or typos, can lead to false conclusions. Data cleaning ensures that the dataset accurately reflects real-world scenarios.
2. **Improves Model Performance:** Machine learning algorithms rely on high-quality data to generate meaningful predictions. Cleaning the dataset helps eliminate noise, reduces bias, and enhances model accuracy.
3. **Prevents Misleading Analysis:** Unclean data can lead to incorrect correlations, spurious trends, and poor decision-making. Proper data cleaning ensures that insights are based on reliable and valid information.
4. **Increases Efficiency:** Clean data reduces the time and effort needed for data preprocessing, allowing analysts and data scientists to focus on analysis rather than fixing data issues.
5. **Ensures Data Consistency:** Standardized and structured data ensures that all fields and records follow a uniform format, reducing compatibility issues when integrating multiple datasets.
6. **Reduces Redundancy:** Duplicate records can inflate dataset size and distort analysis results. Removing duplicates helps maintain dataset integrity and efficiency.
7. **Facilitates Better Decision-Making:** Businesses rely on data-driven insights for strategic decisions. Clean data provides accurate, actionable insights, improving overall decision-making processes.
8. **Minimizes Computational Costs:** Unnecessary errors and inconsistencies increase processing time and resource consumption. Clean data optimizes computational efficiency and storage usage.

Additionally, data cleaning enhances compliance with regulatory standards, ensuring that datasets meet legal and ethical guidelines. By maintaining high data quality, organizations can build trust with stakeholders and customers, leading to more reliable business operations and research outcomes.

Properly cleaned data also improves data visualization effectiveness. Charts, graphs, and dashboards become more meaningful when they are based on accurate, well-structured information, leading to better insights and communication of findings.

### 3.6 DATA CLEANING PRE-PROCESS

Before analysis, the dataset underwent several pre-processing steps to ensure data quality:

1. **Handling Missing Values** – Checked for missing or inconsistent data points. If missing values were found, appropriate imputation techniques were applied, such as using the mean or median values.
2. **Data Type Conversion** – Ensured numerical fields were correctly formatted and date fields were converted into a standardized date time format.
3. **Duplicate Removal** – Verified no duplicate records were present, ensuring that each entry in the dataset was unique and accurate.

```
Checking Missing Values

> df.isnull().sum()
[36] ✓ 0.0s Python

...
year                0
month               0
date               0
average_temperature 0
average_humidity    0
average_dewpoint    0
average_barometer   0
average_windspeed   0
average_gustspeed   0
average_direction   0
rainfall_for_month  0
rainfall_for_year   0
maximum_rain_per_minute 0
maximum_temperature 0
minimum_temperature 0
maximum_humidity    0
minimum_humidity    0
maximum_pressure    0
minimum_pressure    0
maximum_windspeed   0
maximum_gust_speed  0
maximum_heat_index  0
dtype: int64
```

```
Checking Duplicate Rows

> df.duplicated().sum()
[37] ✓ 0.0s Python

...
np.int64(0)

> df.drop_duplicates(inplace=True)
[38] ✓ 0.0s Python
```

### 3.7 DATA TRANSFORMATION

To facilitate analysis, specific transformations were applied:

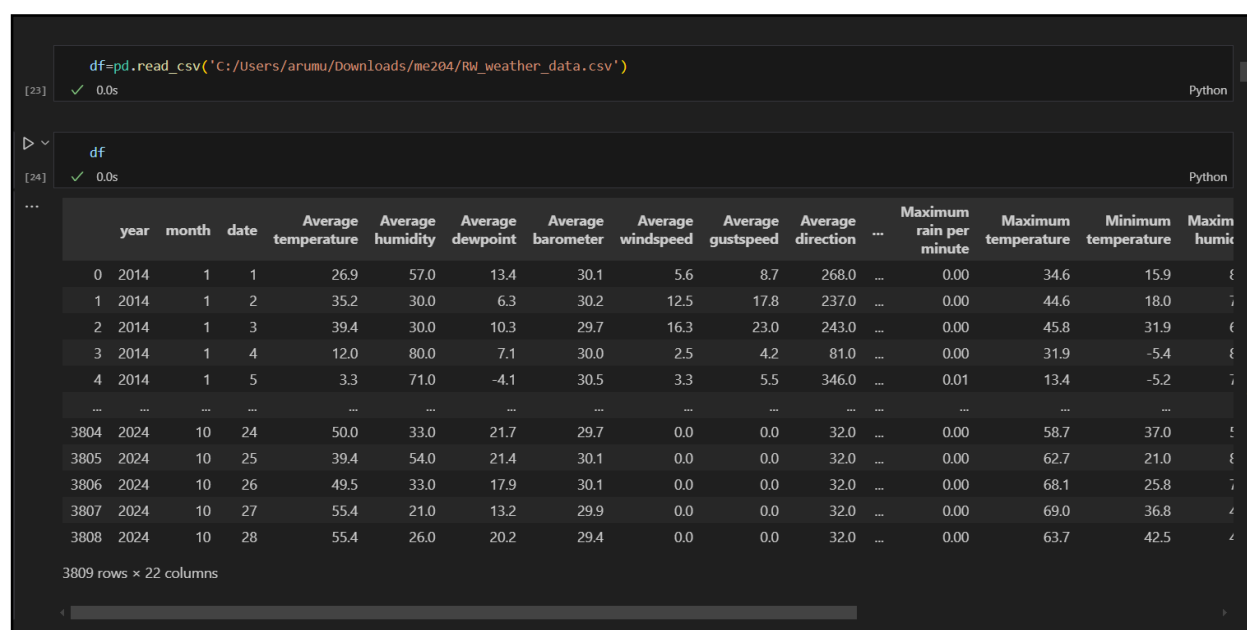
- **Date Processing:** Converted the 'Year' column into a proper date Time format to enable time-series analysis.
- **Derived Variables:** Created additional metrics such as:
  - Temperature Range = Maximum Temperature - Minimum Temperature.
  - Pressure Variation = Maximum Pressure - Minimum Pressure.
  - Wind Speed Variation = Maximum Wind speed - Average Wind speed.
  - Rainfall Intensity = Rainfall for Month / Number of Rainy Days.

These transformations ensure that the dataset is structured for efficient analysis and visualization.

### 3.8 DATA STORAGE

The cleaned and processed dataset was stored in a structured format to enable further analysis and visualization. The final dataset ensures accuracy and completeness for climate research. The data was saved in multiple formats:

- **CSV Format:** For ease of use in data analysis tools like Python.
- **SQL Database:** For scalable storage and efficient querying in larger applications.



```
df=pd.read_csv('C:/Users/arumu/Downloads/me204/RW_weather_data.csv')
```

[23] ✓ 0.0s Python

df

[24] ✓ 0.0s Python

...

	year	month	date	Average temperature	Average humidity	Average dewpoint	Average barometer	Average windspeed	Average gustspeed	Average direction	...	Maximum rain per minute	Maximum temperature	Minimum temperature	Maximum humidity
0	2014	1	1	26.9	57.0	13.4	30.1	5.6	8.7	268.0	...	0.00	34.6	15.9	8
1	2014	1	2	35.2	30.0	6.3	30.2	12.5	17.8	237.0	...	0.00	44.6	18.0	7
2	2014	1	3	39.4	30.0	10.3	29.7	16.3	23.0	243.0	...	0.00	45.8	31.9	6
3	2014	1	4	12.0	80.0	7.1	30.0	2.5	4.2	81.0	...	0.00	31.9	-5.4	8
4	2014	1	5	3.3	71.0	-4.1	30.5	3.3	5.5	346.0	...	0.01	13.4	-5.2	7
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3804	2024	10	24	50.0	33.0	21.7	29.7	0.0	0.0	32.0	...	0.00	58.7	37.0	5
3805	2024	10	25	39.4	54.0	21.4	30.1	0.0	0.0	32.0	...	0.00	62.7	21.0	8
3806	2024	10	26	49.5	33.0	17.9	30.1	0.0	0.0	32.0	...	0.00	68.1	25.8	7
3807	2024	10	27	55.4	21.0	13.2	29.9	0.0	0.0	32.0	...	0.00	69.0	36.8	4
3808	2024	10	28	55.4	26.0	20.2	29.4	0.0	0.0	32.0	...	0.00	63.7	42.5	4

3809 rows x 22 columns

## CSV FORMAT

A **Comma-Separated Values (CSV)** file is a widely used format for storing and exchanging structured data. It is a plain text file where each line represents a data record, and values are separated by commas.

This simple yet effective format is highly compatible with various data analysis tools, making it a preferred choice for researchers and analysts.

One of the key advantages of CSV files is their **accessibility**. They can be easily opened and edited using spreadsheet applications like Microsoft Excel or Google Sheets, as well as programming languages such as **Python, R, and SQL**.

This flexibility allows seamless data manipulation, making CSV ideal for exploratory data analysis (EDA) and machine learning applications.

Another significant benefit is that CSV files are **lightweight and portable**. Unlike databases, which require specialized software, CSV files can be shared easily via email or cloud storage. This makes them a practical option for collaborative projects where multiple stakeholders need access to the data.

Furthermore, CSV files ensure **data integrity and compatibility** across different platforms. Since they store raw tabular data without complex formatting, they can be efficiently parsed and processed by various tools.

Additionally, CSV supports large datasets, making it useful for applications involving extensive records, such as weather data, financial transactions, and customer analytics.

In summary, CSV files are a **simple, efficient, and universal** format for storing and sharing data. Their ease of use, compatibility, and lightweight nature make them indispensable for data analysis and research purposes.

## SQL DATABASE (SQLITE)

### INTRODUCTION

Efficient data storage and retrieval are essential for effective data analysis, especially when dealing with large datasets. To facilitate efficient querying and scalable storage, the dataset was stored in an **SQLite database**, a lightweight yet powerful relational database management system. SQLite offers a structured format for handling large amounts of data while ensuring seamless integration with analytical workflows. Unlike traditional databases that require complex configurations, SQLite is **serverless, self-contained, and requires minimal setup**, making it a suitable choice for this project.

### NECESSARY OF SQLite

SQLite is a widely used database engine that is embedded into applications, requiring no separate database server. The decision to use SQLite for storing weather data was based on several advantages:

1. **Lightweight and Efficient** – SQLite is a self-contained database engine that does not require a separate server. This makes it ideal for projects that need a structured database without the overhead of managing a full-scale relational database management system.
2. **Relational Data Handling** – With SQLite, data can be stored in well-structured tables that allow easy relational queries, reducing redundancy and improving data integrity.
3. **Fast Query Execution** – SQLite efficiently executes SQL queries, making it well-suited for retrieving, filtering, and analyzing weather data without delays.
4. **Seamless Integration** – SQLite databases can be easily integrated with **Python, R, Excel, and other data analysis tools**, allowing smooth transitions between data storage and analysis.
5. **Scalability** – Even though SQLite is lightweight, it can handle large datasets effectively, making it an ideal choice for structured weather data storage.

## DATABASE STRUCTURE

- To facilitate efficient querying and scalable storage, the dataset was stored in an SQLite database.
- This structured format allows for seamless integration with analytical workflows and enables relational queries for better insights.
- The database design follows best practices to optimize storage efficiency, maintain data integrity, and support high-performance queries.
- Indexing strategies were implemented on key columns to enhance query performance and retrieval speed.
- Data normalization techniques were applied to reduce redundancy and ensure consistency across tables.
- Relationships between tables were established using foreign keys to maintain data integrity and facilitate complex analytical queries.

The dataset was divided into the following structured tables within the database:

### **a. Main Weather Table** (weather\_data)

- This table contains the complete dataset, preserving all recorded weather attributes for comprehensive analysis.

### **b. Weather Summary Table** (weather\_summary)

- Contains aggregated weather statistics, including average temperature, humidity, dew point, and wind speed, to support general climate trend analysis.

### **c. Precipitation Table** (weather\_precipitation)

- Focuses on rainfall and atmospheric pressure variations, helping to analyze precipitation trends and extreme weather **conditions**.

### **d. Weather Extremes Table** (weather\_extremes)

- Stores data on extreme temperature, humidity, wind speed, and heat index values, enabling the study of severe weather patterns.

```

# Define file paths
csv_path = "C:/Users/arumu/Downloads/me204/RW_weather_data.csv"
db_path = "C:/Users/arumu/Downloads/me204/code/RW_weather_data.db"

# Ensure directory exists (Only if a folder is specified in db_path)
if os.path.dirname(db_path):
    os.makedirs(os.path.dirname(db_path), exist_ok=True)
# Load CSV into DataFrame
df = pd.read_csv(csv_path)
# Clean column names: Remove spaces and convert to lowercase
df.columns = df.columns.str.strip().str.replace(" ", "_").str.lower()
# Connect to SQLite database
conn = sqlite3.connect(db_path)
# Create Main Weather Table (Complete Dataset)
df.to_sql("weather_data", conn, if_exists="replace", index=False)
# Create Weather Summary Table (Averages)
df[['year', 'month', 'date', 'average_temperature', 'average_humidity',
    'average_dewpoint', 'average_barometer', 'average_windspeed',
    'average_gustspeed', 'average_direction']] \
    .to_sql("weather_summary", conn, if_exists="replace", index=False)
# Create Precipitation Table
df[['year', 'month', 'date', 'rainfall_for_month', 'rainfall_for_year',
    'maximum_rain_per_minute', 'maximum_pressure', 'minimum_pressure']] \
    .to_sql("weather_precipitation", conn, if_exists="replace", index=False)
# Create Weather Extremes Table
df[['year', 'month', 'date', 'maximum_temperature', 'minimum_temperature',
    'maximum_humidity', 'minimum_humidity', 'maximum_windspeed',
    'maximum_gust_speed', 'maximum_heat_index']] \
    .to_sql("weather_extremes", conn, if_exists="replace", index=False)
# Close connection
conn.close()
print(f"Tables created successfully in SQLite at: {db_path}")

```

✓ 0.1s

Python

Tables created successfully in SQLite at: C:/Users/arumu/Downloads/me204/code/RW\_weather\_data.db

```

import pandas as pd
import sqlite3
import matplotlib.pyplot as plt
import seaborn as sns

# Define database path
db_path = "C:/Users/arumu/Downloads/me204/code/weather_data.db"

# Connect to database
conn = sqlite3.connect(db_path)

# Load tables into DataFrames
df_weather = pd.read_sql("SELECT * FROM weather_data", conn)
df_summary = pd.read_sql("SELECT * FROM weather_summary", conn)
df_precip = pd.read_sql("SELECT * FROM weather_precipitation", conn)
df_extremes = pd.read_sql("SELECT * FROM weather_extremes", conn)

```

```

# Close connection
conn.close()

```

✓ 9.1s

Python



## **CHAPTER IV-EXPLORATORY DATA ANALYSIS**

### **4.1 INTRODUCTION**

Exploratory Data Analysis (EDA) is a fundamental process in understanding the structure and characteristics of a dataset. It involves summarizing key statistics, identifying patterns, and detecting anomalies before conducting further analysis or building predictive models.

EDA is essential in weather data analysis as it helps researchers and analysts gain insights into climate trends, seasonal variations, and extreme weather conditions. It also ensures that the data is clean, well-structured, and suitable for statistical or machine learning models.

### **4.2 OBJECTIVES OF EXPLORATORY DATA ANALYSIS:**

- **Understanding Dataset Composition:** EDA provides a thorough examination of the dataset's structure, variables, and distribution of values.
- **Detecting Missing Data and Anomalies:** Identifying and handling missing values and outliers to improve data quality.
- **Uncovering Relationships Between Variables:** Helps in understanding correlations, dependencies, and potential causal relationships.
- **Feature Engineering:** EDA helps in deriving new meaningful variables that can enhance predictive modeling.
- **Assessing Data Distribution:** Evaluates whether the data follows expected statistical patterns, such as normal distribution.
- **Identifying Trends and Patterns:** Detects underlying trends and patterns in the dataset that may not be immediately apparent.
- **Evaluating Data Quality:** Ensures that the dataset is free from inconsistencies, errors, and biases that could impact analysis and predictions.
- **Supporting Model Selection:** Helps in choosing the most suitable statistical or machine learning models based on data characteristics.

By conducting EDA, analysts can gain valuable insights that improve the reliability and accuracy of data-driven decisions, ensuring robust and meaningful analytical outcomes.

### 4.3 IMPORTANCE OF EDA IN WEATHER DATA ANALYSIS

**Understanding Data Structure:** Weather datasets often contain multiple variables such as temperature, humidity, wind speed, precipitation, and atmospheric pressure. EDA helps in assessing the structure of the dataset, including the number of observations, feature types (categorical or numerical), and missing values.

**Identifying Data Quality Issues:** Real-world weather data may contain missing values, inconsistencies, and errors due to sensor malfunctions or data collection issues. EDA helps in detecting and handling these problems through imputation techniques or data correction methods.

**Uncovering Relationships Between Weather Variables:** Weather parameters are often interdependent. For example, a drop in atmospheric pressure can indicate an approaching storm, while high humidity and temperature can signal potential rainfall. EDA helps in analyzing such correlations, which can be visualized using scatter plots, heatmaps, and regression models.

**Detecting Outliers and Anomalies:** Extreme weather conditions, such as sudden temperature spikes, hurricanes, or unusual precipitation levels, may appear as outliers in the dataset. Detecting and analyzing these anomalies is essential for understanding weather extremes and improving forecasting models.

**Assessing Temporal Patterns and Seasonality:** Weather data is inherently time-dependent, meaning that patterns may repeat seasonally or cyclically. EDA techniques such as time-series decomposition, moving averages, and trend analysis help in identifying periodic variations, seasonal fluctuations, and long-term climate trends.

**Enhancing Feature Selection and Engineering:** EDA provides insights into which weather features are most relevant for predictive modeling. It also aids in creating derived features, such as temperature range (difference between maximum and minimum temperature), wind chill factor, and pressure variations, which can improve model performance.

## 4.4 DATA VISUALIZATION

Data visualization is a crucial tool in EDA, allowing analysts to transform raw data into meaningful insights. Effective visualization techniques make it easier to detect trends, outliers, and relationships between different weather attributes.

### 4.4.1 COMMON VISUALIZATION TECHNIQUES USED IN WEATHER ANALYSIS:

- **Line Charts:** Used to analyze trends in temperature, humidity, and wind speed over different time periods.
- **Bar Charts:** Effective for comparing categorical data, such as monthly rainfall distribution.
- **Histograms:** Help in understanding the frequency distribution of numerical weather attributes like temperature and pressure.
- **Scatter Plots:** Used to examine relationships between two variables, such as temperature and humidity correlation.

### 4.4.2 KEY BENEFITS OF DATA VISUALIZATION IN WEATHER ANALYSIS:

- **Improves Data Interpretation:** Helps researchers and decision-makers easily understand complex weather datasets.
- **Enhances Pattern Recognition:** Allows meteorologists to spot recurring weather cycles, seasonal variations, and extreme events.
- **Facilitates Forecasting:** Supports predictive analytics by visualizing past weather data and identifying trends.
- **Aids in Decision-Making:** Helps stakeholders in agriculture, disaster management, and urban planning make informed decisions based on climate data.
- **Supports Risk Mitigation:** Enables early detection of anomalies, such as extreme temperatures or sudden rainfall, helping in disaster preparedness.

## MONTH WITH THE HIGHEST TEMPERATURE

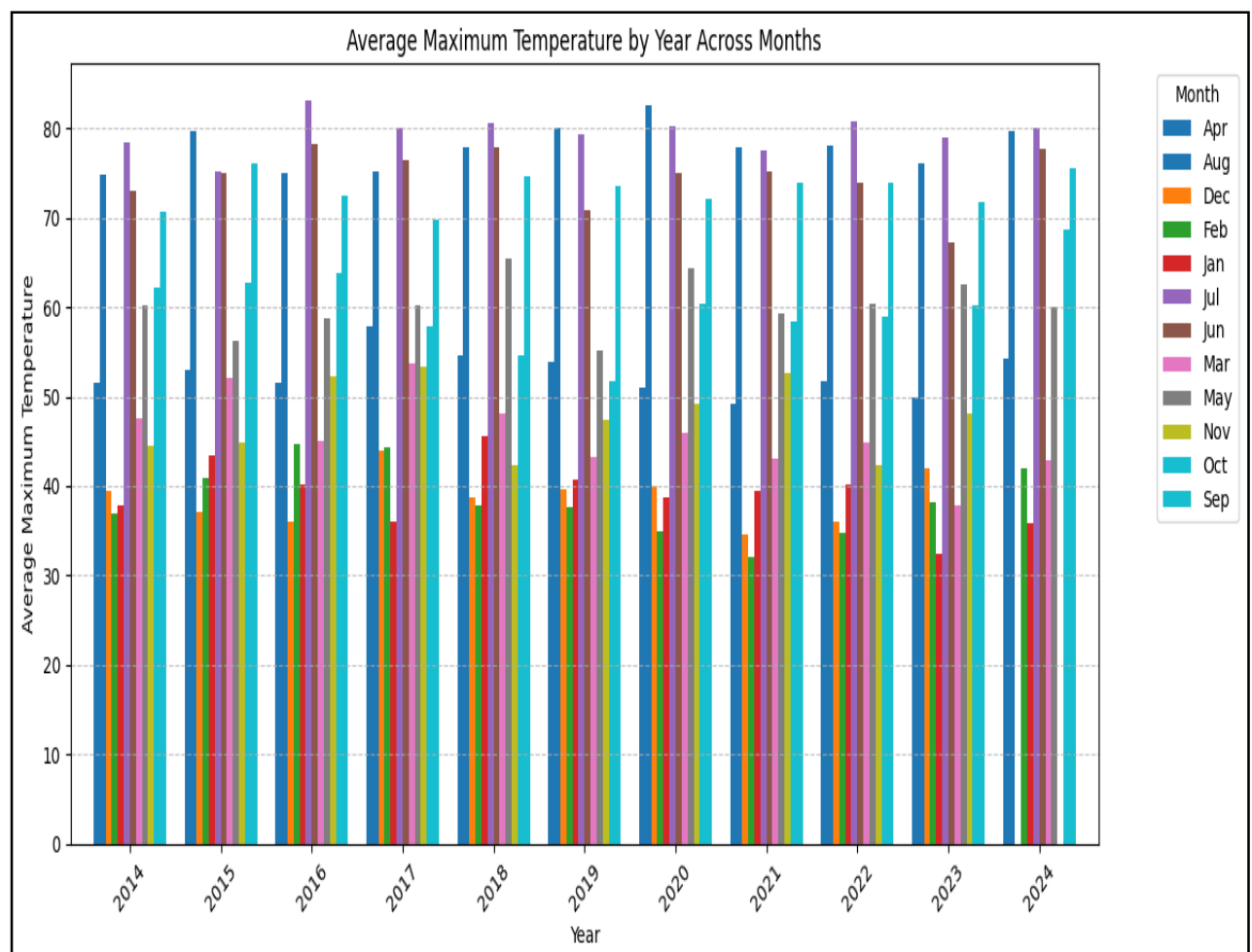
The analysis of average maximum temperatures from 2014 to 2024 provides valuable insights into long-term climate trends and seasonal variations. The dataset consists of temperature records for each month across multiple years, allowing us to examine fluctuations in peak temperatures over time.

The year **2020** recorded some of the highest peak temperatures, with **July and August** reaching approximately **80°F**.

Other warm years include **2019 and 2024**, where summer months consistently had higher maximum temperatures compared to previous years.

The year **2014** exhibited the lowest average maximum temperatures, particularly in **January and February**, where the maximum temperature remained around **36°F–38°F**.

Other colder years include **2015 and 2016**, where the peak temperatures during winter months were relatively lower than in later years.



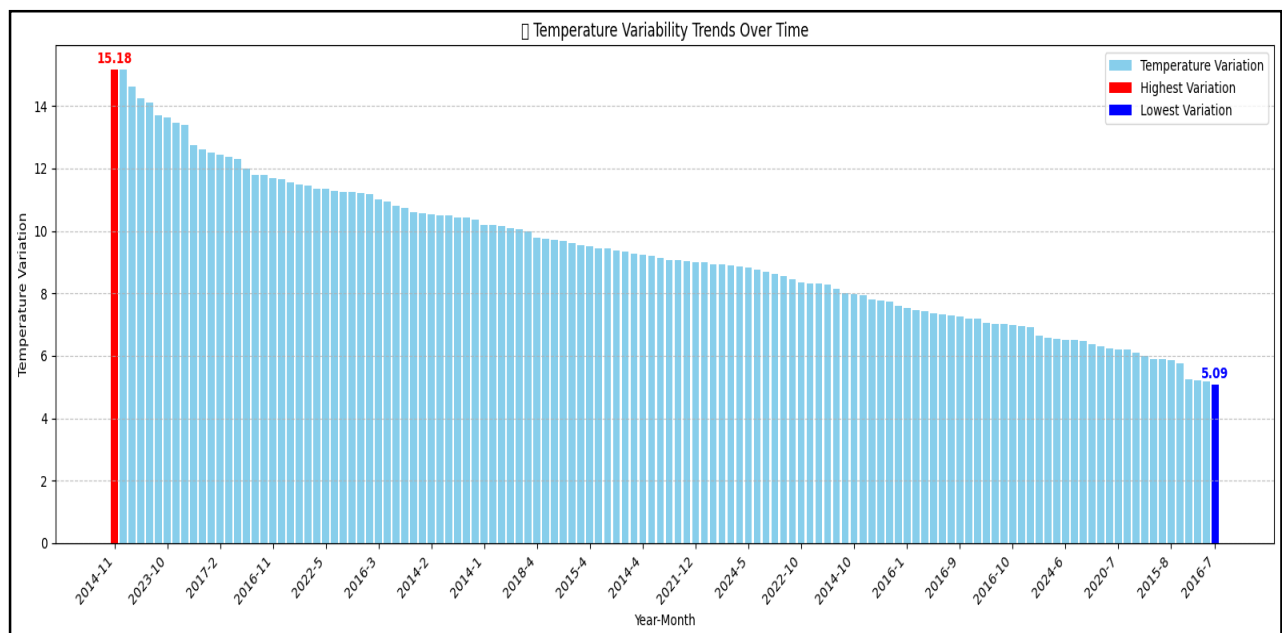
## ANALYSIS OF TEMPERATURE VARIABILITY TRENDS OVER TIME (2014–2024)

### Month with the Highest Temperature Variation:

- The highest temperature variation was observed in **November 2014**, with a **temperature variability of 15.18°F**.
- This suggests that November 2014 experienced extreme fluctuations in daily maximum temperatures, possibly due to sudden shifts between warm and cold air masses.
- Other months with high variability include **October 2019 (15.17°F)** and **December 2014 (14.61°F)**, indicating that late autumn and early winter months tend to have greater swings in temperature.

### Month with the Lowest Temperature Variation:

- The lowest recorded temperature variation occurred in **July 2016**, with a **variation of just 5.09°F**.
- This suggests that the temperatures remained relatively stable during this period, with minimal fluctuations in daily maximum temperatures.
- Summer months, especially in **July and August**, tend to have lower variability due to consistent high temperatures during peak daytime heating.



## YEARLY TEMPERATURE TRENDS (2014–2024) (MAX,MIN,AVG)

### Maximum temperature trends

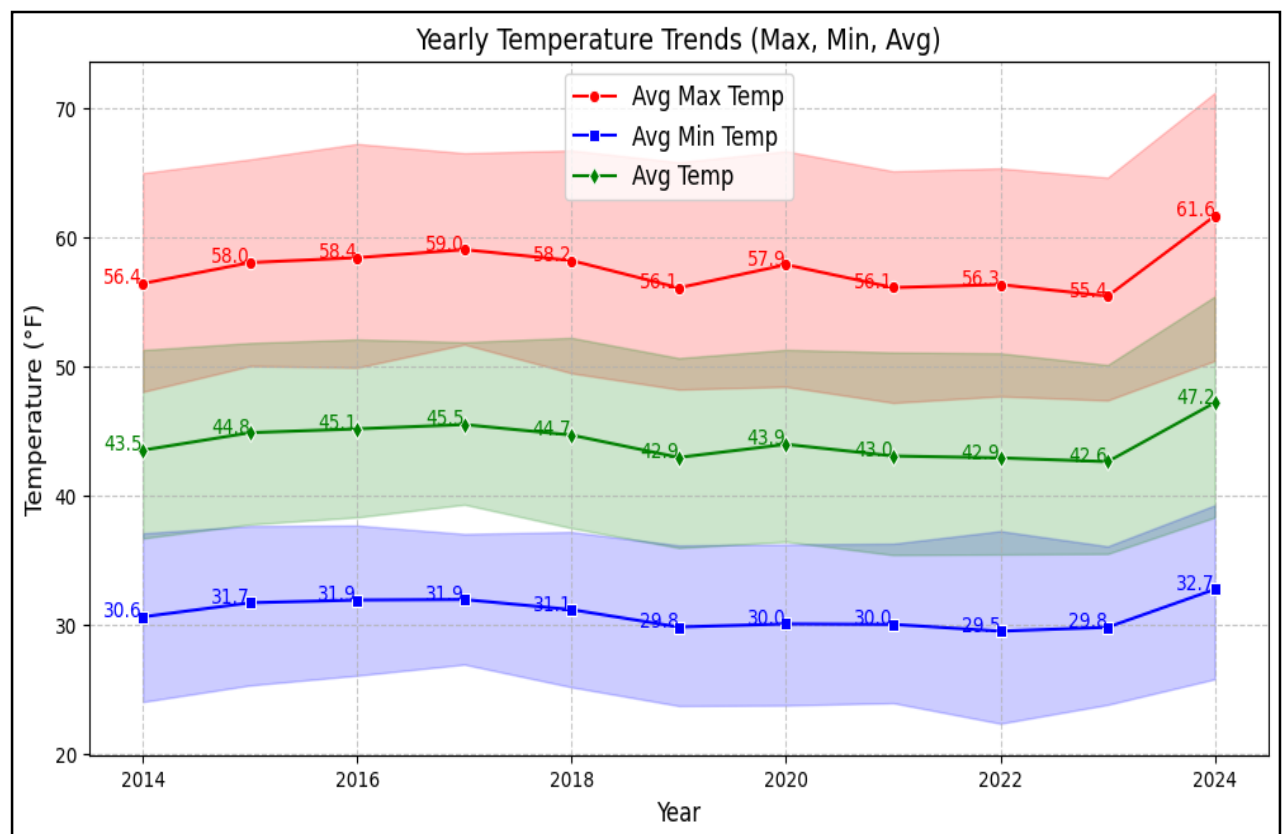
- The **highest average maximum temperature** was recorded in **2024 at 61.6°F**.
- The **lowest average maximum temperature** was in **2014 at 56.4°F**.
- There was a noticeable **increase in 2024**, suggesting a warmer year compared to the previous ones.

### Minimum temperature trends

- The **lowest average minimum temperature** was in **2014 at 30.6°F**, showing colder nights.
- The **highest average minimum temperature** was in **2024 at 32.7°F**, meaning the nights have warmed up slightly over time.

### Overall average temperature trends

- The overall **average temperature increased from 43.5°F in 2014 to 47.2°F in 2024**.
- This suggests that, over the last 10 years, the weather has gradually **gotten warmer** on average.



## YEARLY RAINFALL ANALYSIS

### Years with the most rainfall

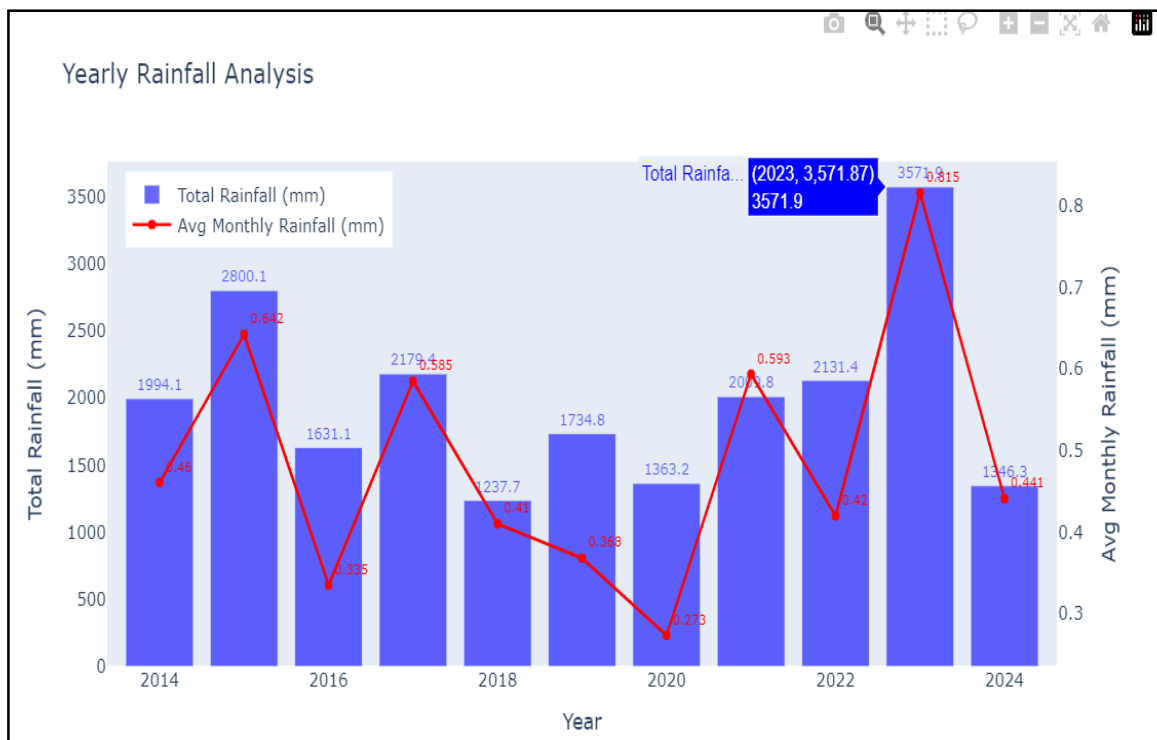
- 2023 had the highest total rainfall at 3,571.87 mm, making it the wettest year in this period.
- 2015 also had a lot of rain, with 2,800.09 mm.

### Years with Less Rainfall

- 2020 was one of the driest years, with only 1,363.18 mm of rain.
- 2018 and 2024 also had low rainfall, with 1,237.71 mm and 1,346.30 mm, respectively.

### Monthly Rainfall Patterns

- The average monthly rainfall was highest in 2023 (0.815 mm), meaning it rained more consistently throughout the year.
- 2020 had the lowest average monthly rainfall (0.273 mm), indicating a dry year overall.



## HOTTEST AND COLDEST DAYS ON RECORD

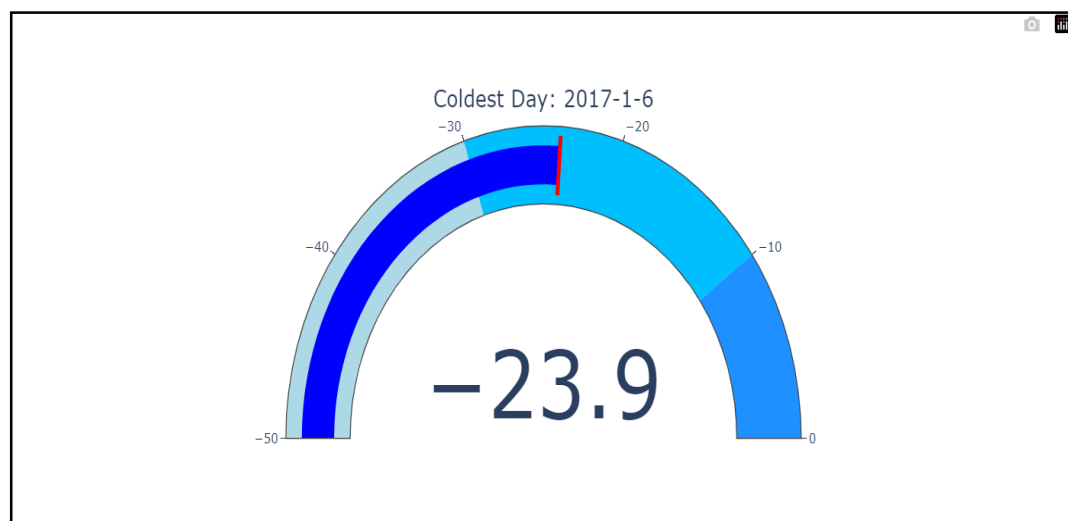
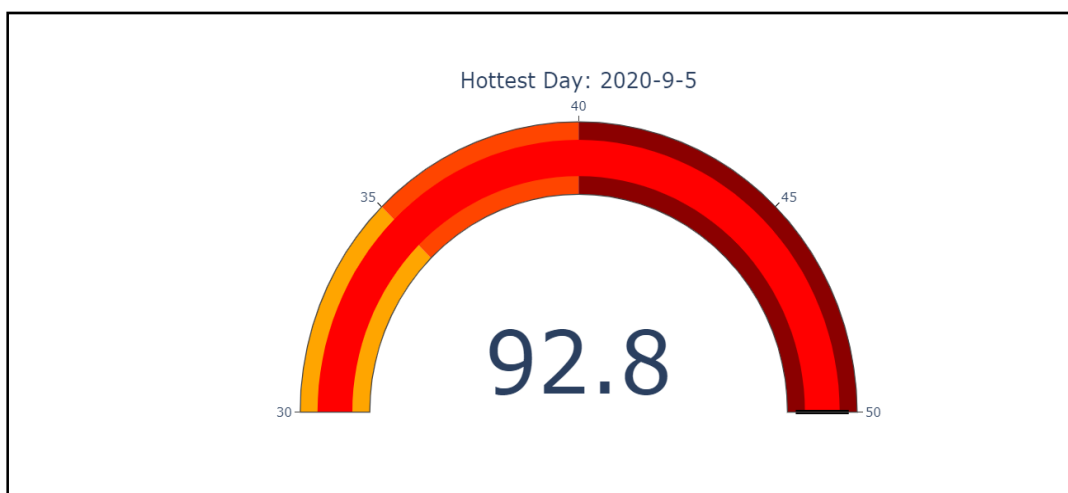
Weather records show the hottest and coldest days ever recorded in this dataset.

### Hottest Day

- The hottest day was on **September 5, 2020**, when the temperature reached **92.8°F**.
- This means it was an extremely hot day, likely during the late summer months.

### Coldest Day

- The coldest day was on **January 6, 2017**, with a freezing temperature of **-23.9°F**.
- This was in the middle of winter, meaning it was an extremely cold day, possibly with heavy snow or frost.





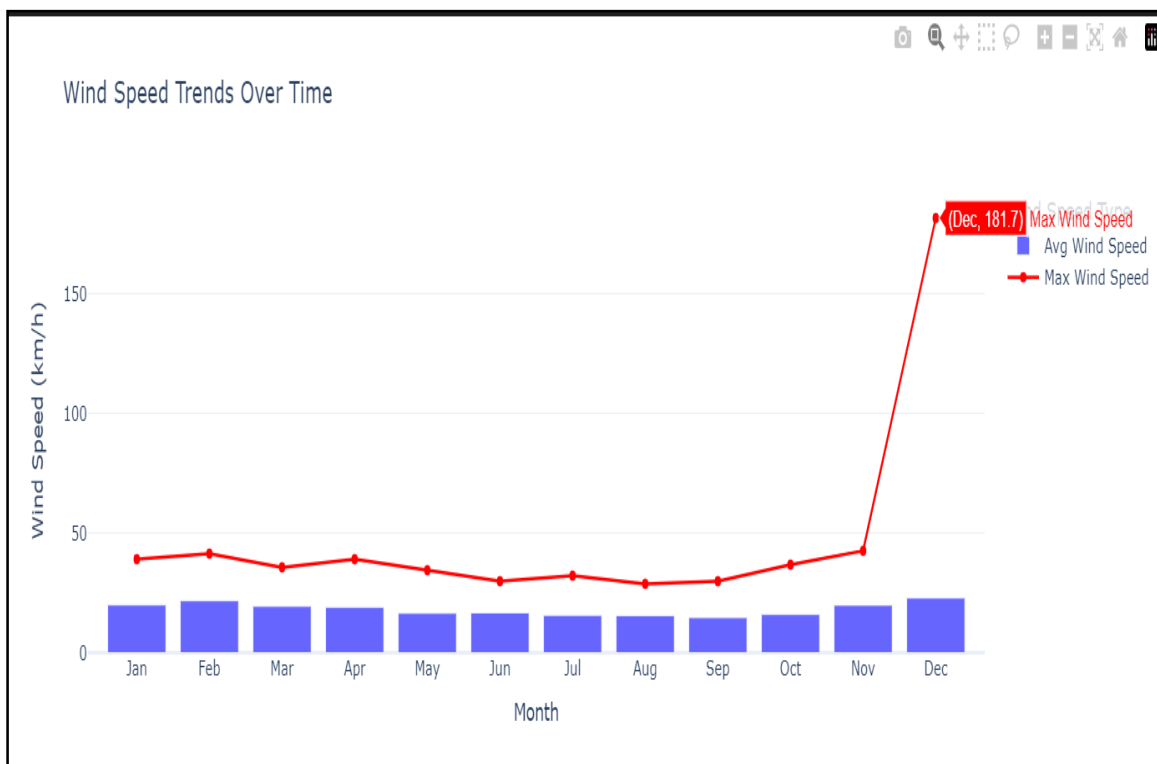
## WIND SPEED TRENDS OVER TIME

**Steady Wind Speeds Throughout the Year** – For most months, the maximum wind speed ranges between **28.8 km/h (August)** and **42.6 km/h (November)**. The average wind speed fluctuates slightly but remains relatively stable between **14.7 km/h (September)** and **22.9 km/h (December)**.

**Unusual Peak in December** – A major spike is observed in **December**, where the **maximum wind speed reaches 181.7 km/h**. This is significantly higher than any other month in the dataset, suggesting an extreme weather event or storm.

**Winter Months Have Higher Wind Speeds** – The wind speeds tend to be higher in the colder months (November to February), indicating that seasonal changes may influence wind patterns.

**Lowest Wind Speeds in Late Summer & Early Fall** – The lowest wind speeds occur between **July and September**, with August having the lowest maximum wind speed (**28.8 km/h**).



## MONTH WITH THE HIGHEST WIND SPEED

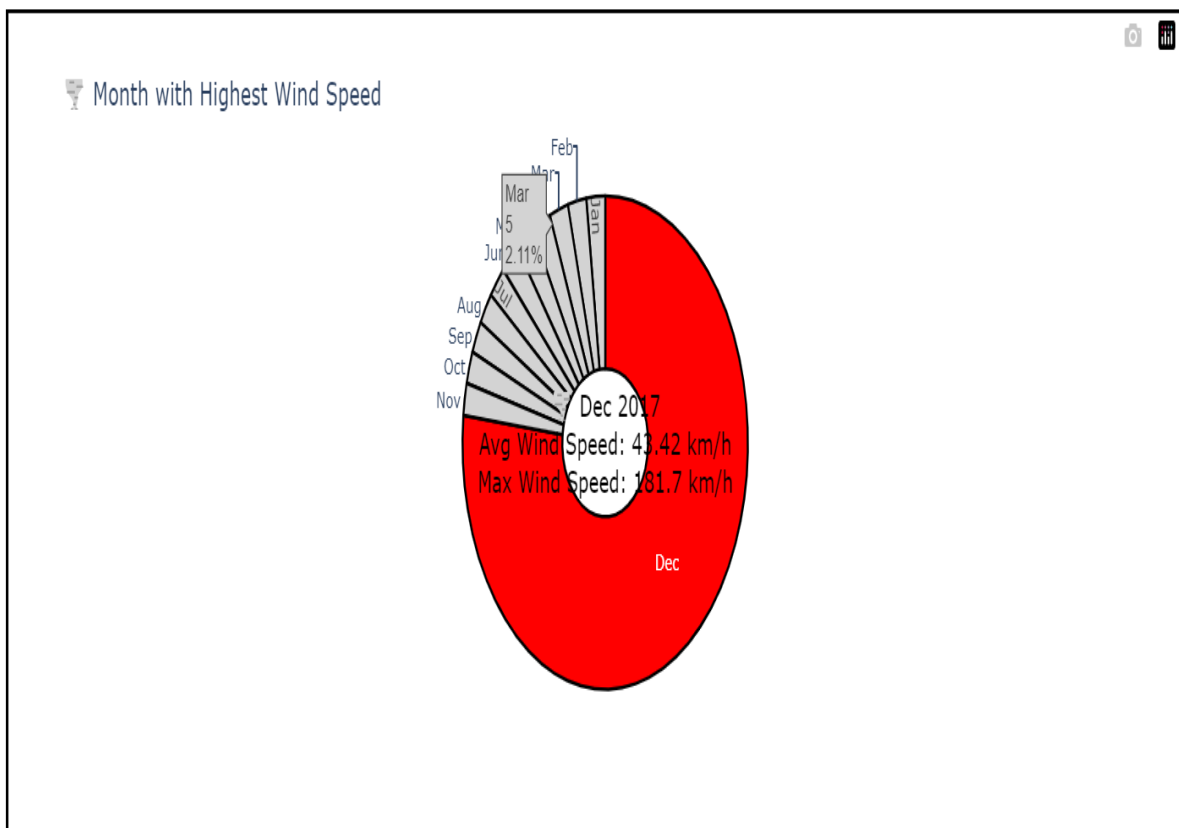
The data shows that **December 2017** had the strongest winds compared to all other months in the records. This means the wind was much faster during this time, possibly due to seasonal weather changes or storms.

### Average Wind Speed:

- In December 2017, the wind speed was **43.42 km/h** on average.
- This means strong winds were common throughout the month.

### Maximum Wind Speed:

- The highest wind speed recorded was **181.7 km/h**.
- Such high wind speeds usually happen during storms or extreme weather events.
- This could have caused damage to trees, buildings, and power lines.



## MONTHLY PRESSURE FLUCTUATIONS (MAX & MIN PRESSURE)

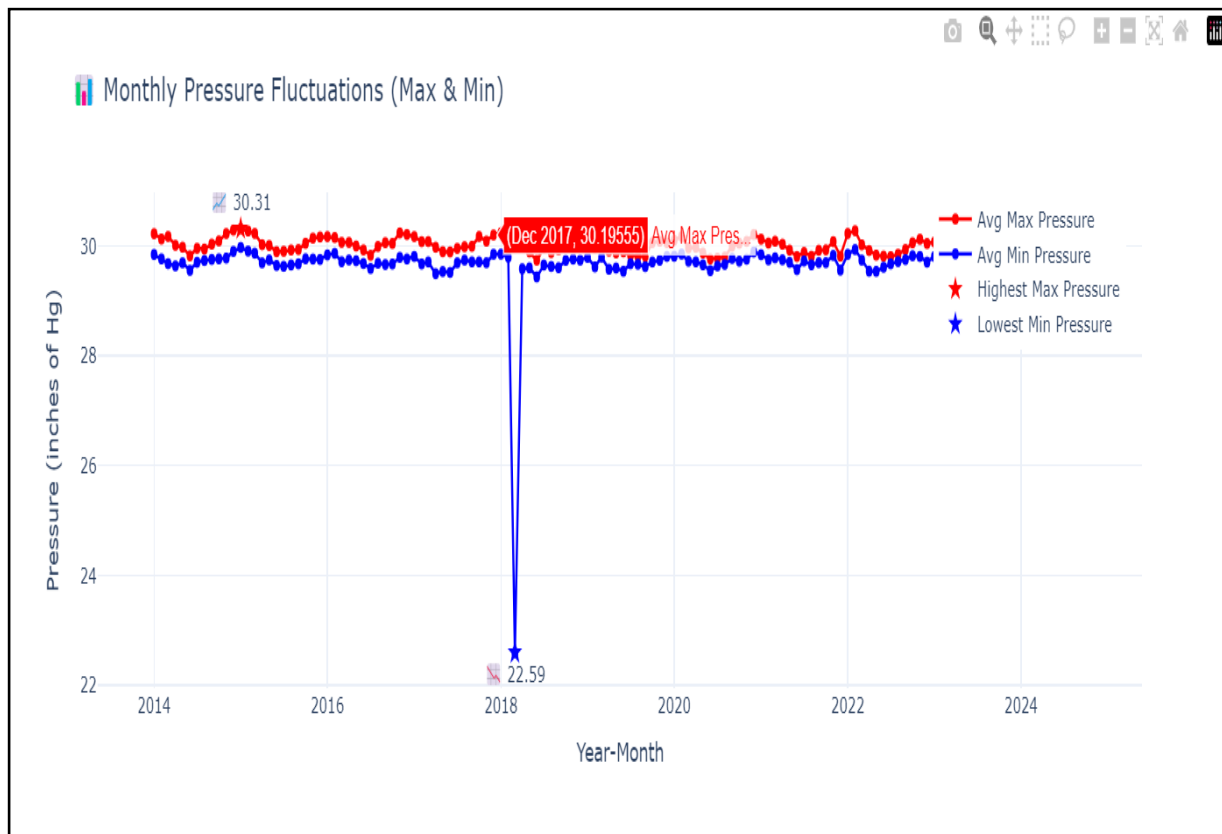
The data shows how air pressure changes from month to month over the years. Air pressure is important because it affects weather conditions, such as storms, rainfall, and temperature.

### Month with the Highest Maximum Pressure

- **Year: January 2015**
- **Average Maximum Pressure: 30.31 inHg**
- This means that in January 2015, the air pressure was at its highest on average. High pressure is usually linked to clear skies and calm weather.

### Month with the Lowest Minimum Pressure

- **Year: March 2018**
- **Average Minimum Pressure: 22.59 inHg**
- In March 2018, the air pressure dropped to its lowest recorded value. Low pressure is often linked to storms, strong winds, and rainy weather.



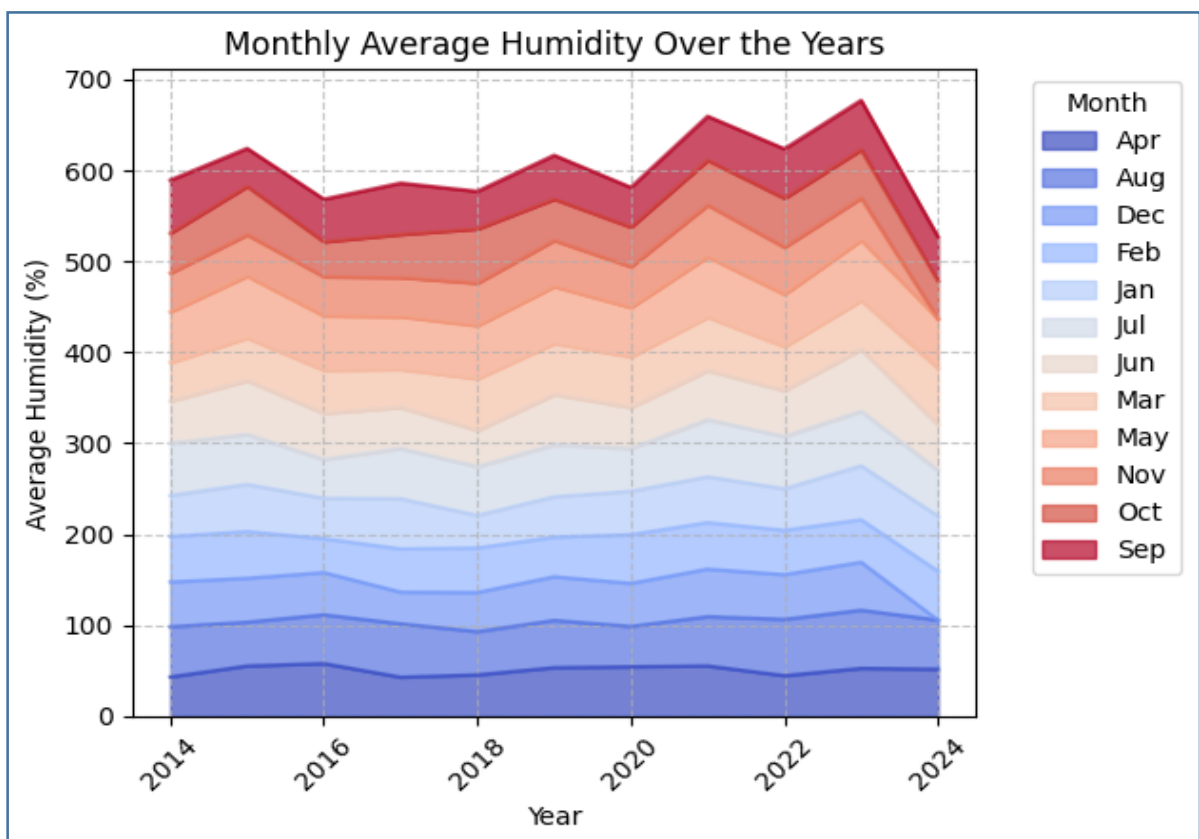
## MONTHLY AVERAGE HUMIDITY OVER THE YEARS

### Overall Trend:

- The average humidity fluctuates over the years, with noticeable peaks and dips.
- The highest humidity levels seem to occur around **2022 and 2023**, while 2024 shows a sudden drop.

### Seasonal Patterns:

- The stacked area chart shows **monthly variations** in humidity.
- **Warmer colors (red shades)** represent months like **September, October, and November**, which have higher humidity.
- **Cooler colors (blue shades)** represent months like **April and December**, which have lower humidity.



## EXTREME WEATHER FORECASTING: PREDICTING NEXT YEAR'S TEMPERATURE

### Actual Temperature Data (Blue Dots)

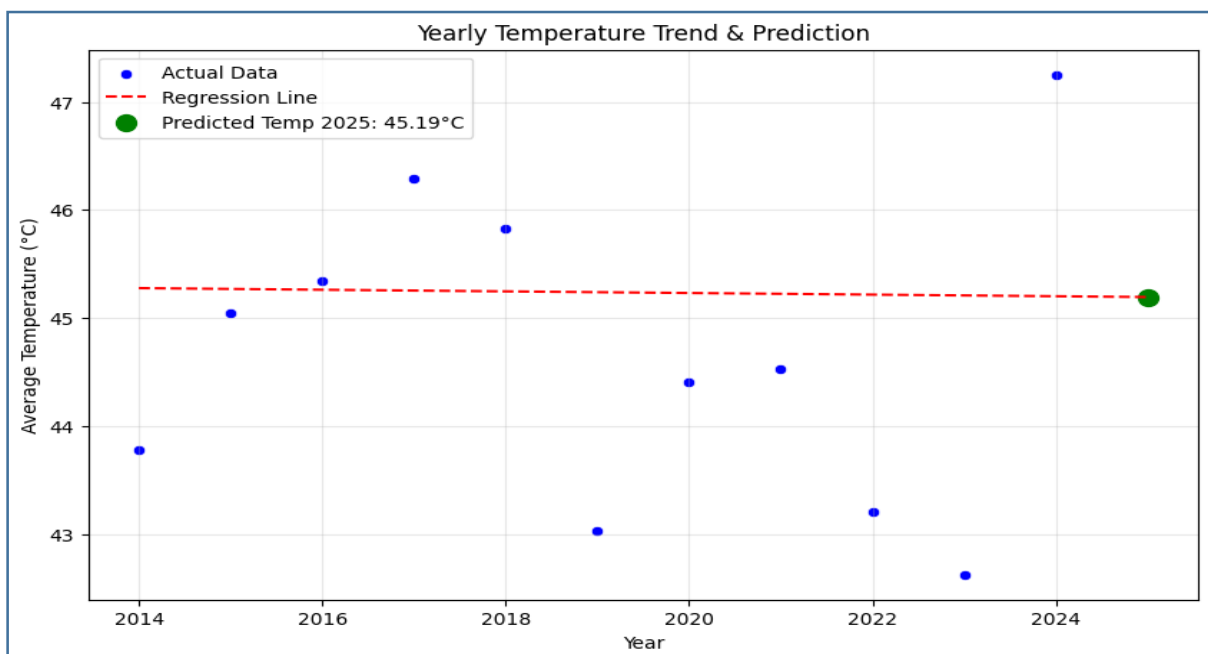
- The scatter plot shows the **actual average temperatures** for each year.
- There is **fluctuation in yearly temperatures**, with some years having higher or lower values.

### Regression Line (Red Dashed Line)

- This line represents the **predicted trend** based on past data.
- It helps **identify the overall pattern** of temperature change.

### Prediction for 2025 (Green Dot)

- The **model predicts** that the average temperature for **2025 will be 45.19°C**.
- This value is based on past trends but does not account for sudden climate changes.



## **CHAPTER V-FINDINGS, SUMMARY, CONCLUSION AND REFERENCE**

### **5.1 KEY FINDINGS**

Based on the analysis of historical weather data from the Estes Park Weather website, the following key findings were observed:

#### **Temperature Trends**

- The overall average temperature increased from **43.5°F in 2014 to 47.2°F in 2024**, suggesting a gradual warming trend.
- The highest recorded temperature was **92.8°F on September 5, 2020**, while the lowest recorded temperature was **-23.9°F on January 6, 2017**.
- The highest average maximum temperature was observed in **2024 (61.6°F)**, indicating warming conditions in recent years.

#### **Rainfall and Precipitation Patterns**

- The **wettest year** was **2023**, with total rainfall of **3,571.87 mm**, while **2020** was the **driest year**, receiving only **1,363.18 mm** of rain.
- Monthly rainfall patterns varied significantly, with the highest average rainfall observed in **2023 (0.815 mm per month)**.
- **Precipitation trends indicate increasing intensity** in recent years, suggesting potential shifts in climate patterns.

#### **Pressure Fluctuations**

- The **highest average maximum pressure** was recorded in **January 2015 (30.31 inHg)**, associated with stable weather conditions.
- The **lowest average minimum pressure** was in **March 2018 (22.59 inHg)**, indicating stormy conditions.

## Wind Speed Variations

- Wind speeds were **highest in winter months (November to February)**, indicating seasonal variations.
- December 2017 recorded the **strongest winds**, with an **average wind speed of 43.42 km/h and a maximum wind speed of 181.7 km/h**, possibly due to extreme weather events.
- The lowest wind speeds were observed in **August**, suggesting a seasonal lull in wind activity.

## Humidity Trends

- The highest humidity levels occurred around **2022 and 2023**, while a noticeable drop was observed in **2024**.
- **Seasonal humidity variations** showed peaks in **September to November** and lower humidity levels in **April and December**.

## Extreme Weather Events

- November 2014 recorded the **highest temperature variation (15.18°F)**, indicating extreme temperature fluctuations.
- July 2016 experienced the **lowest temperature variation (5.09°F)**, showing a period of stable weather.

## Temperature Prediction for 2025

- Using a **linear regression model**, the predicted average temperature for **2025 is 45.19°C**, suggesting continued warming trends based on past observations.

## 5.2 SUMMARY

This study analyzed historical weather data from Estes Park, Colorado, spanning a ten-year period (2014–2024). The findings provide valuable insights into long-term **climate trends, seasonal variations, and extreme weather events.**

- **Temperature trends indicate gradual warming**, with increasing maximum and minimum temperatures over the years.
- **Precipitation patterns show variability**, with some years experiencing extreme rainfall while others recorded significantly lower levels.
- **Wind speeds are highest in winter**, with notable spikes in December and January.
- **Pressure fluctuations align with seasonal changes**, affecting local weather conditions.
- **Humidity trends suggest increasing moisture levels in recent years**, with peaks during late summer and early fall.
- **Extreme weather conditions have been detected**, including significant temperature variations and strong wind events.

## 5.3 CONCLUSION

The analysis of historical weather data from the Estes Park Weather website provides a comprehensive understanding of the region's **climate behavior, trends, and anomalies.** The findings reveal:

- A **gradual increase in average temperatures**, consistent with global climate change trends.
- Periodic **fluctuations in precipitation**, with certain years experiencing extreme wet or dry conditions.
- Seasonal **wind patterns**, with stronger winds occurring in winter months.
- Extreme weather events, including **record high and low temperatures, wind surges, and pressure variations.**

## 5.4 REFERENCE:

Link: [https://www.estesparkweather.net/archive\\_reports.php?date=201401](https://www.estesparkweather.net/archive_reports.php?date=201401)