**Capstone Project-**

**Weather Analysis**

|  |
| --- |
| Data Analytics |
| Kapil Verma |

# OVERVIEW

The dataset contains weather-related information for various cities, including attributes such as humidity, pressure, temperature, weather descriptions, wind direction, and wind speed. It includes data on a city's geographical details, making it suitable for location-based analysis. The dataset's individual tables provide hourly data for these attributes, enabling the analysis of weather trends, seasonal variations, and correlations among weather attributes. The dataset can be leveraged for gaining insights into weather patterns and developing a Power BI dashboard for weather monitoring and historical weather analysis.

# THE PROCESS

**1. Retrieving data from a GitHub repository:**

**Acquire the necessary dataset from a designated GitHub repository containing comprehensive meteorological data. This dataset includes information on temperature, humidity, wind patterns, and geographical attributes, spanning various cities and regions.**

**2. Data Transformation and Enhancement:**

**Execute data preprocessing steps, ensuring data quality and consistency. This may involve tasks like data cleaning, handling missing values, and standardizing formats. Additionally, consider enriching the dataset with new problem statements or insights to enhance the depth of analysis.**

**3. Connecting with Analytical Tools:**

**Establish connections between the dataset and analytical tools like Power BI, Excel, and SQL for efficient data integration and processing. This step ensures that I can effectively analysis and visualize the data.**

**4. MECE Breakdown Analysis in Excel:**

**Create a structured analysis plan by breaking down the project into manageable components. This ensures a clear and organized approach to solving specific weather-related problems.**

**5. Problem Statement Solution in Power BI :**

**Implement Power BI to visualize and analysis the data comprehensively. Leverage its features for creating interactive dashboards and deriving solutions to the weather-related problem statements.**

**6. Exploratory Data Analysis (EDA):**

**Perform exploratory data analysis using either Excel or SQL Workbench, depending on the complexity of the analysis. Extract meaningful patterns, relationships, and trends from the data to inform subsequent decision-making.**

**7. Creation of Visual and Insightful PowerPoint:**

**Develop a comprehensive PowerPoint presentation that encapsulates the project's objectives, methodologies, problem statement solutions, and key visualizations. Each problem statement should be accompanied by a dedicated section with pertinent conclusions and insights.**

**8. Detailed Documentation:**

**Compile a detailed report that meticulously documents the entire project lifecycle. Include sections on data collection, transformation, problem statement formulation, tools integration, Power BI solutions, EDA insights, and PowerPoint visualizations.**

# OBJECTIVE

**The objective of this project is to analysis weather trends, seasonal variations, and correlations between weather attributes using a comprehensive dataset that includes information on city attributes, humidity, pressure, temperature, weather descriptions, wind direction, and wind speed. This analysis will be complemented by the development of a Power BI dashboard for weather monitoring and historical weather analysis. The project aims to provide valuable insights into weather patterns and conditions for various cities, aiding decision-making and planning in various domains, including agriculture, energy consumption, and disaster management.**

**Analyzing weather trends and conditions is crucial for a variety of industries and decision-making processes. However, there are challenges such as variations in weather data sources, seasonal fluctuations, correlations between weather attributes, and the impact of these factors on various applications.**

**The project will involve the following tasks:**

* **This involves exploring variations in weather data, seasonal patterns, and correlations between weather attributes.**
* **The project aims to draw meaningful conclusions from the analysis and identify actionable insights for various industries.**
* **Creating a user-friendly Power BI dashboard.**
* **Compiling the results, insights, and dashboard for stakeholders.**

**The success of this project will be measured by:**

* **The quality and depth of the weather analysis.**
* **The relevance and applicability of the derived insights.**
* **The utility of the Power BI dashboard for decision-making.**

**The impact of the recommendations:**

**This project is significant because it has the potential to enhance decision-making across multiple sectors by leveraging historical weather data to understand trends and patterns. This understanding can lead to more informed resource allocation, improved efficiency, and better preparedness for weather-related challenges.**

# SIGNIFICANCE

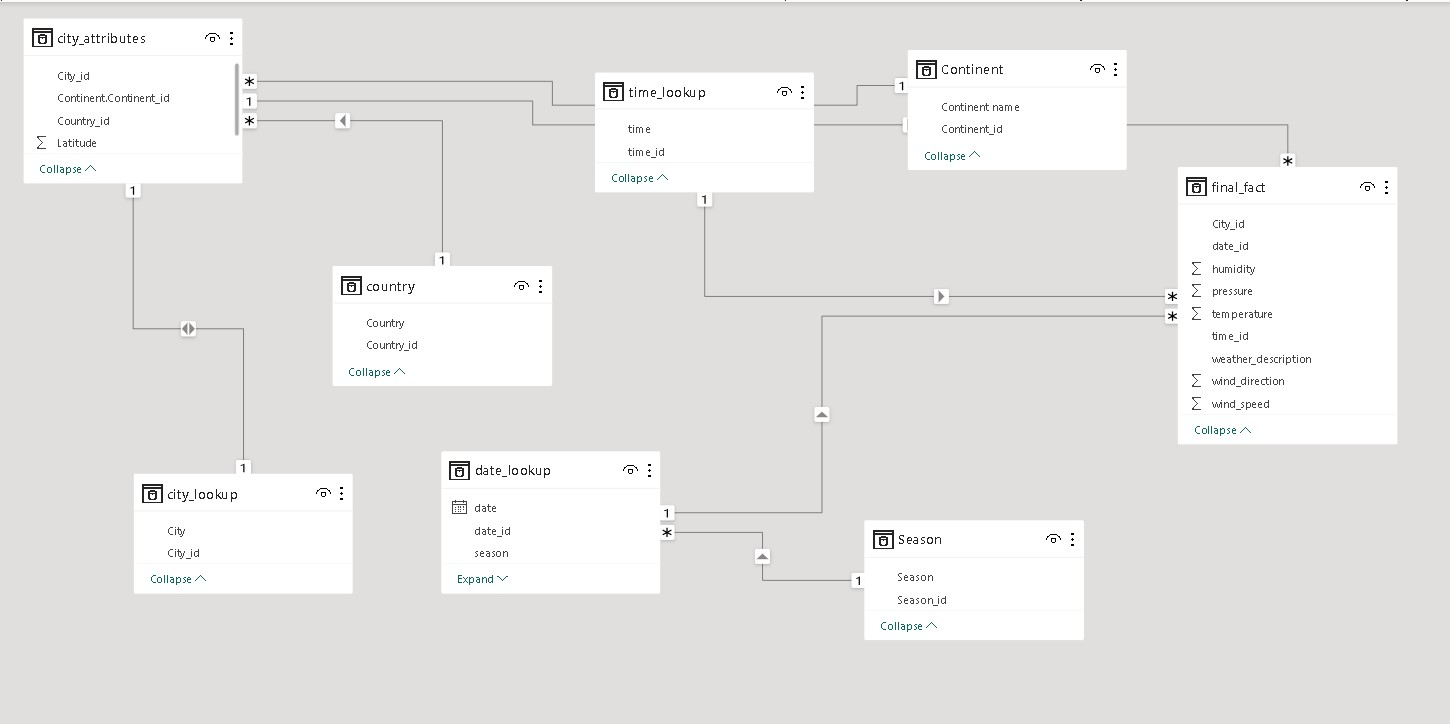
**Weather analysis serves as a critical component in a wide range of industries and sectors, influencing decision-making processes and resource allocation. By delving into historical weather data and trends, individuals and organizations can gain valuable insights, impacting multiple aspects of their operations.**

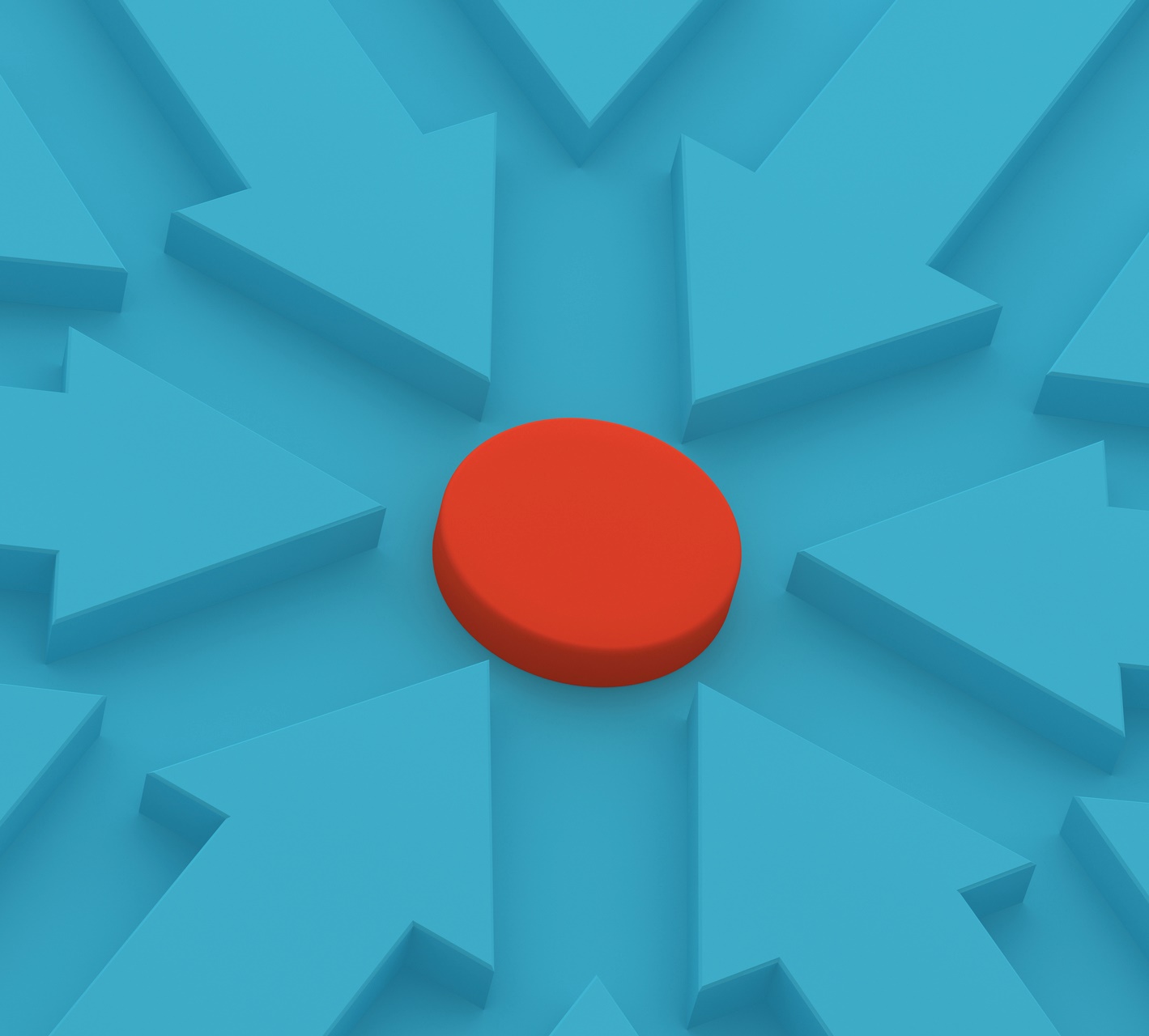
**For various sectors such as agriculture, energy, and disaster management, understanding weather patterns and their implications is essential. By analyzing historical weather data, stakeholders can make informed decisions about resource allocation, energy consumption, agricultural planning, and disaster preparedness. This knowledge empowers them to adapt to seasonal variations, optimize energy usage, improve agricultural yield, and enhance preparedness for weather-related emergencies.**

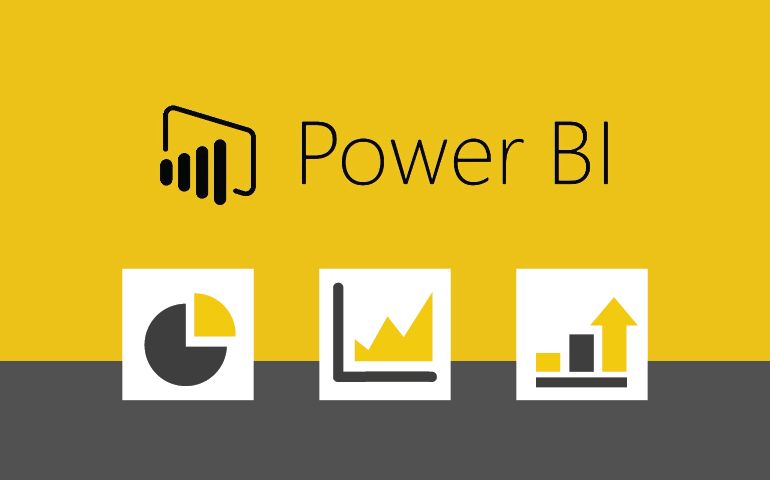
**Furthermore, the development of a user-friendly Power BI dashboard for weather monitoring and historical analysis brings real-time data access and visualization to the forefront. This dashboard is invaluable for businesses, government agencies, and researchers, providing up-to-date weather information and in-depth analysis. Users can customize the dashboard to track specific weather attributes and locations, enabling timely decision-making.**

**In conclusion, the weather analysis project's significance lies in its potential to impact decision-making across various sectors. It equips stakeholders with knowledge and tools to make more informed choices, optimize resource allocation, and prepare for weather-related challenges. By embracing the insights gained from weather analysis, stakeholders can collectively work towards more resilient and efficient operations, ultimately leading to better decision-making and enhanced resource management.**

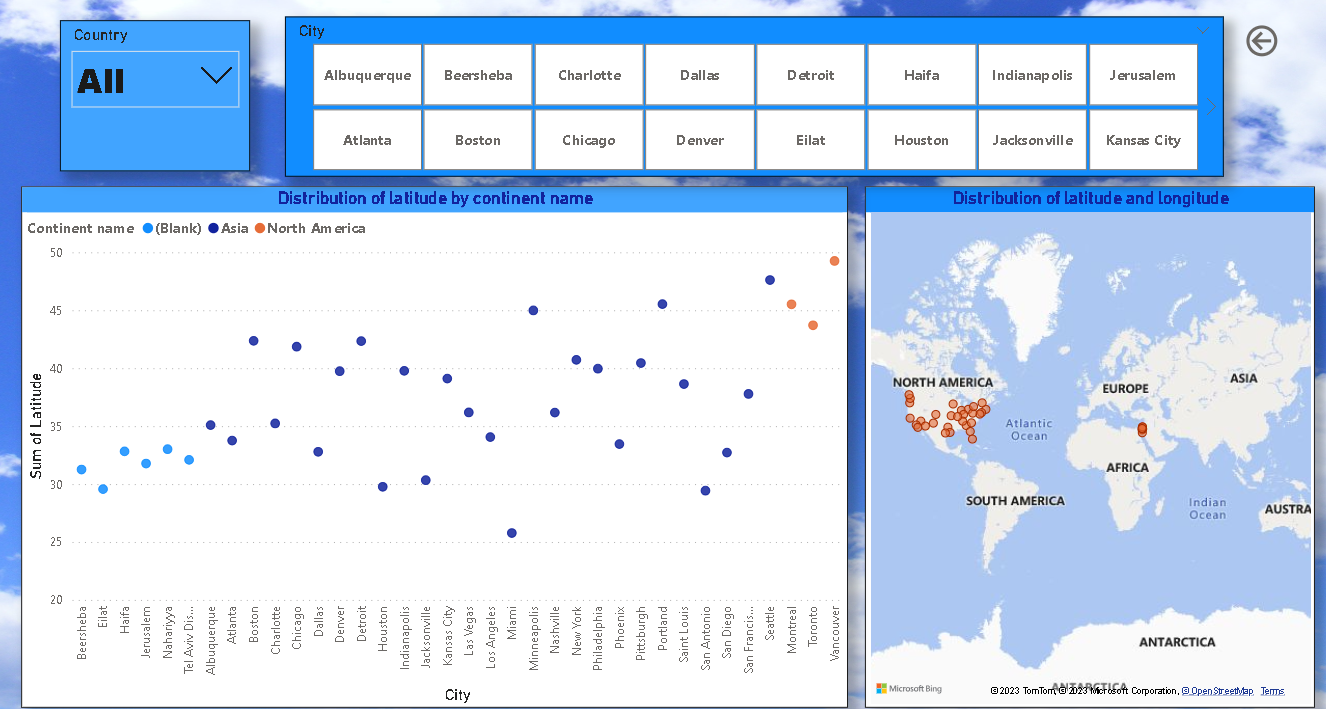
# ER DIAGRAM





**POWER BI OF PROBLEM STATEMENT**

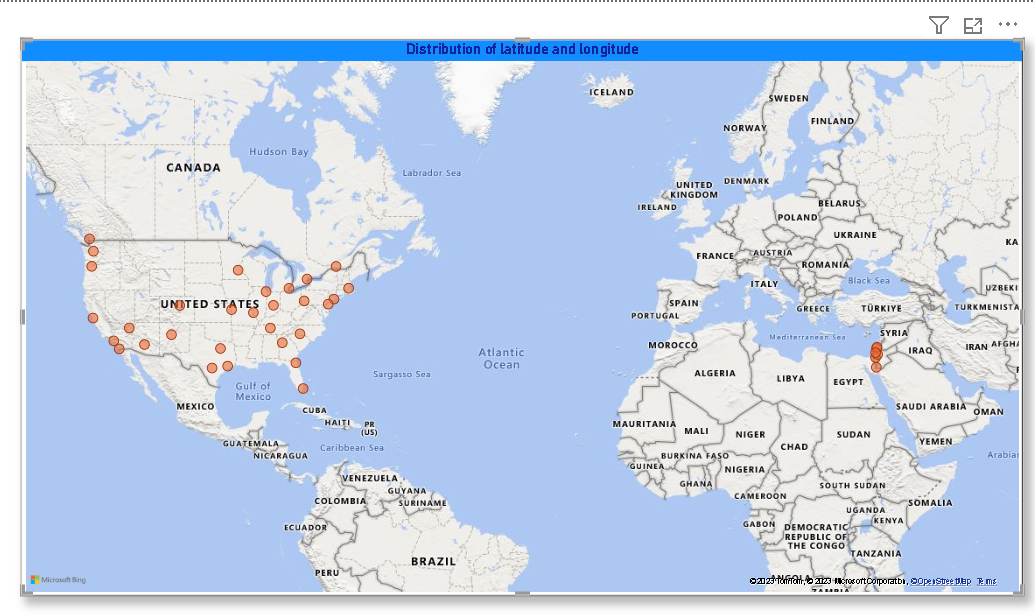
**Geographical Analysis**



1.Can you create a geographical map in Power BI showing the distribution of cities in the dataset based on their latitude and longitude?

* I created a geographical map visualization to showcase the distribution of cities in the dataset based on their latitude and longitude. The map's location parameter was set to "City," and the latitude and longitude were aggregated using the sum function. This visualization provides a clear overview of the spatial distribution of cities across the dataset, allowing us to see their geographic spread and analyze potential patterns or clusters in the data.

The map will show markers or points for each city, with their positions determined by the latitude and longitude values. This allows me to visually understand where these cities are located in the world. You can observe clusters of cities in specific regions, identify proximity to geographical features like coastlines or mountains, and gain insights into how the distribution of cities may impact localized weather patterns or climate variations.

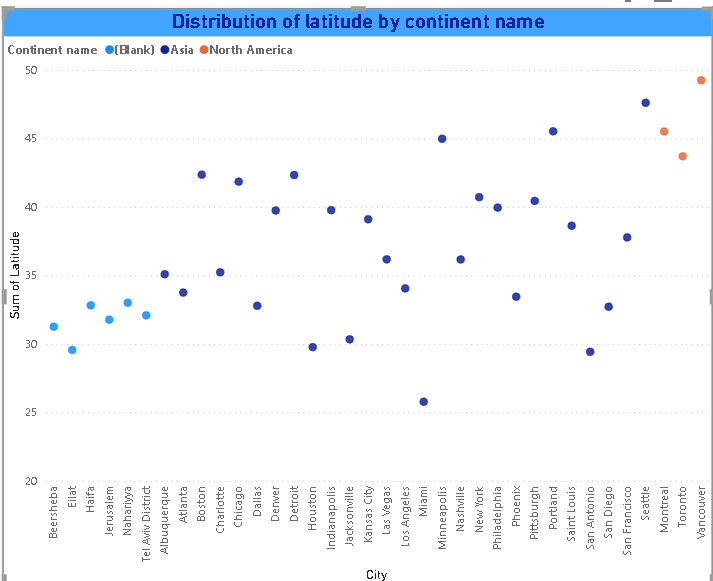


2. In Power BI, can you create a bar chart representing the top 10 countries with the highest number of cities in the dataset?

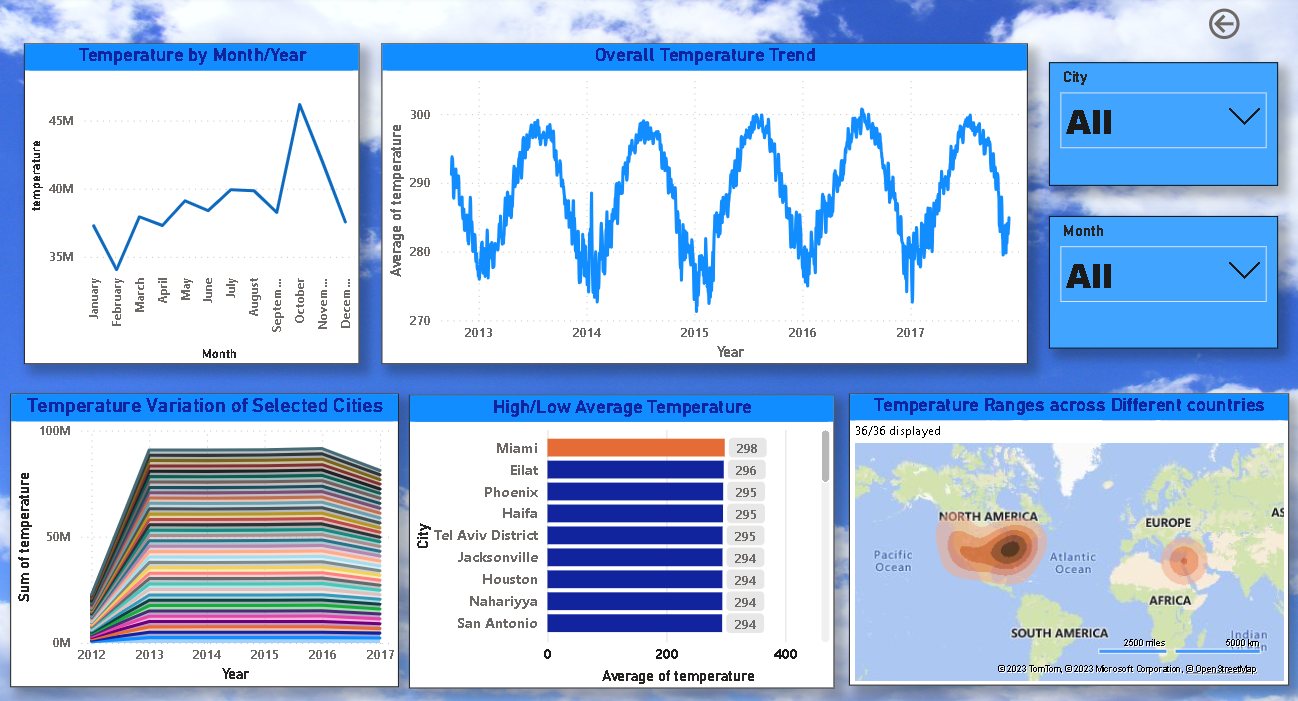
* After analyzing the data, I have determined that there is no relevant data to create a bar chart representing the top 10 countries with the highest number of cities in the dataset, it's a valid and important observation.

3.How does the distribution of cities in terms of latitude vary across different continents? Create a scatter plot in Power BI to illustrate this.

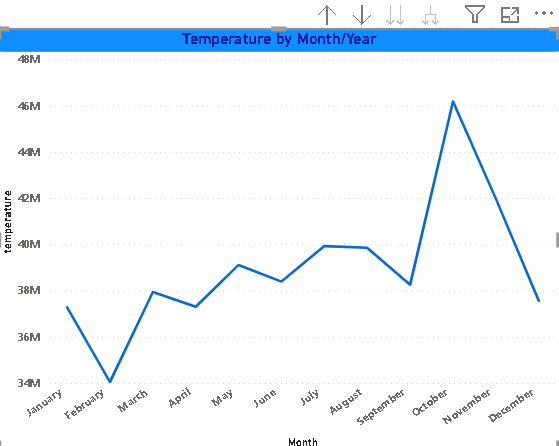
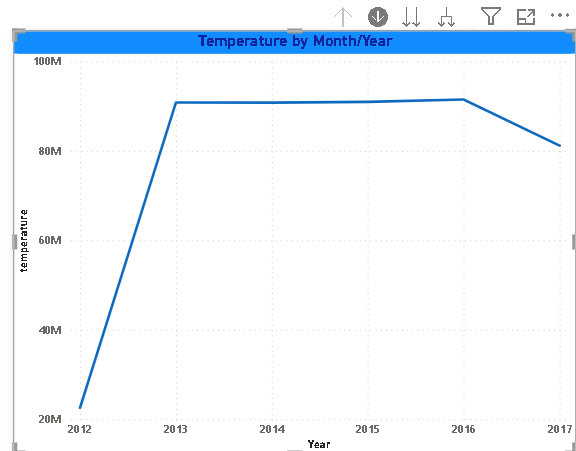
* The scatter plot illustrates how cities are distributed across various latitudes and allows for a clear differentiation by continent name. It is evident that cities are not uniformly distributed but cluster around specific latitudinal ranges. This variation in city distribution signifies the impact of geographical and climatic factors on urban development. For example, cities located closer to the equator tend to have different climatic conditions compared to those in higher latitudes.



**Temperature Analysis**



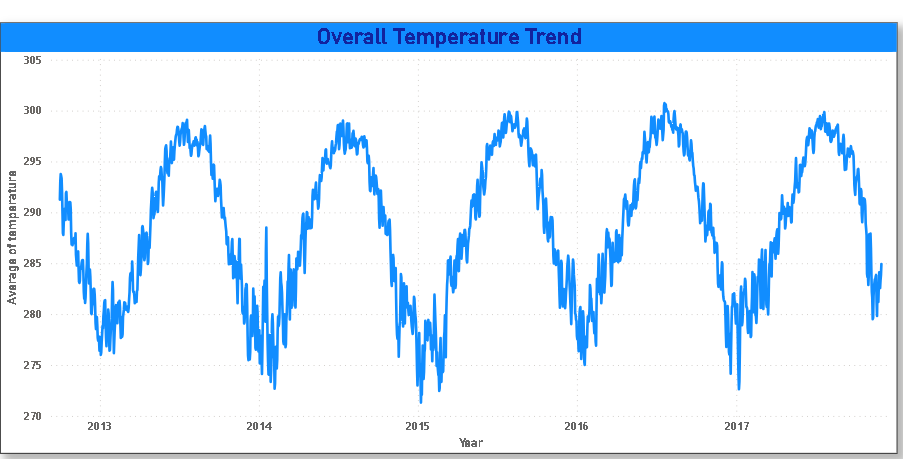
Create a line chart in Power BI to display the temperature trends over time for a selected city. Highlight extreme temperature events

In the provided line chart within Power BI, temperature trends over time for a specific city are displayed. The x-axis represents the date, organized by month and year, while the y-axis showcases temperature values. This chart allows for the observation of temperature fluctuations over a continuous timeline for the selected city. It provides a visual representation of temperature patterns, facilitating the identification of seasonal variations, long-term trends, and noteworthy temperature events.  
Here I showed Two chart 1st one is showing temperature by month and 2nd one showing temperature by Year by using month or year slicer I changed this line for clearly visualization .

Additionally, the chart's potential to highlight extreme temperature events offers a crucial insight. By incorporating visual cues or annotations to mark extreme temperature points, users can quickly identify periods of unusual weather conditions, which could be significant for various applications, such as assessing heatwaves, cold spells, or temperature records. The line chart thus serves as a valuable tool for monitoring temperature trends over time and pinpointing extreme temperature occurrences, enabling more informed decision-making and analysis.

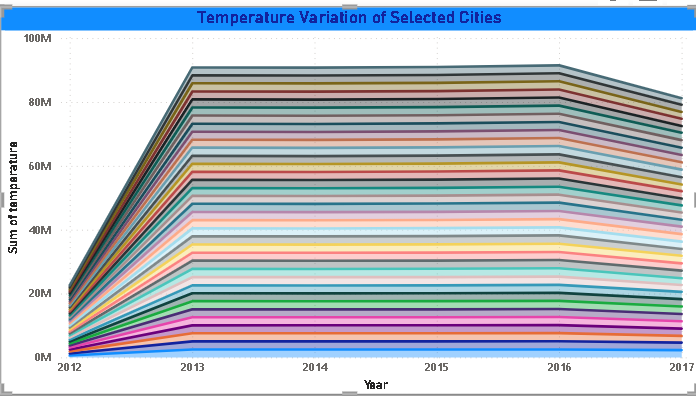
Create a time-series line chart in Power BI to show the overall temperature trends over the entire dataset.



The time-series line chart in Power BI is a powerful visualization tool that offers a complete overview of temperature trends across the entire dataset. The chart has two axes: the x-axis displays the date, providing flexibility to view data at yearly, monthly, or daily intervals, and the y-axis represents temperature values.

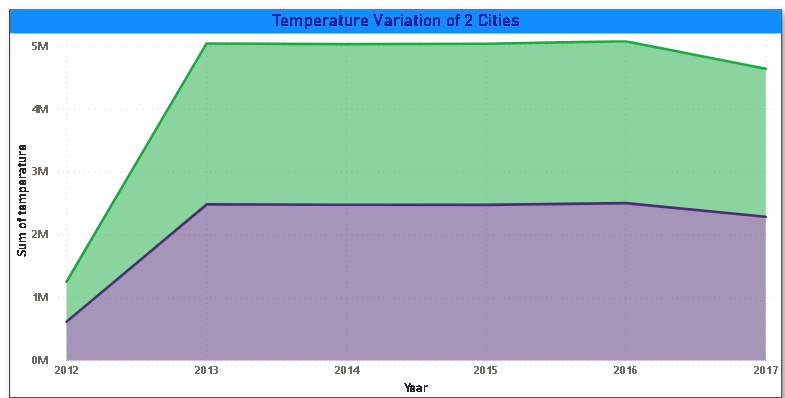
This chart serves to impact by presenting a visual representation of how temperature has changed over time. It helps in identifying long-term trends, seasonal fluctuations, and temperature variations. By observing the line's trajectory, you can discern temperature patterns, which could be crucial for understanding climate changes, pinpointing extreme temperature events, and informing decision-making across various sectors like agriculture, energy management, and meteorology.

Create a Power BI chart comparing the temperature variations between two selected cities over a specific timeframe.

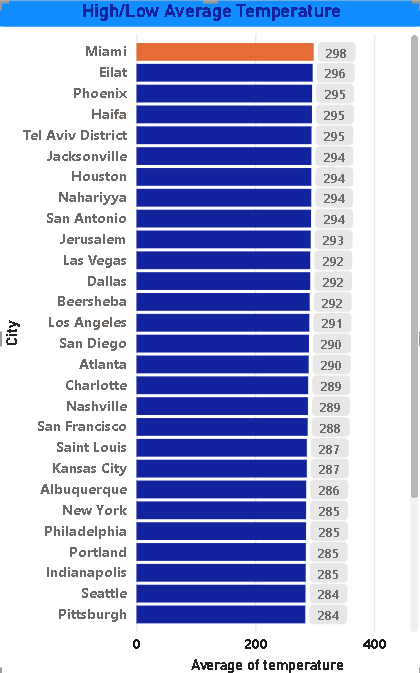


According to my analysis, I have employed a stacked area chart in Power BI to effectively compare temperature variations between two selected cities over a specific timeframe. This chart represents a valuable approach to visualize and understand temperature trends. The x-axis is dedicated to time, offering the flexibility to observe data at different intervals, such as by date or year. The y-axis showcases the sum of temperature values, enabling a clear comparison of temperature variations. Meanwhile, the legend corresponds to the city, making it easy to distinguish between the two selected cities.

Filtering the chart to include only two cities out of the available 36 simplifies the visualization, focusing precisely on the comparison between these specific locations. The stacked area chart's primary impact is to provide a side-by-side representation of temperature variations, facilitating a direct and efficient comparison between the chosen cities. This approach can reveal temperature trends, similarities, differences, and any significant events that may have occurred within the specified timeframe.



Create a bar chart in Power BI to highlight cities with the highest and lowest average temperatures in the dataset.

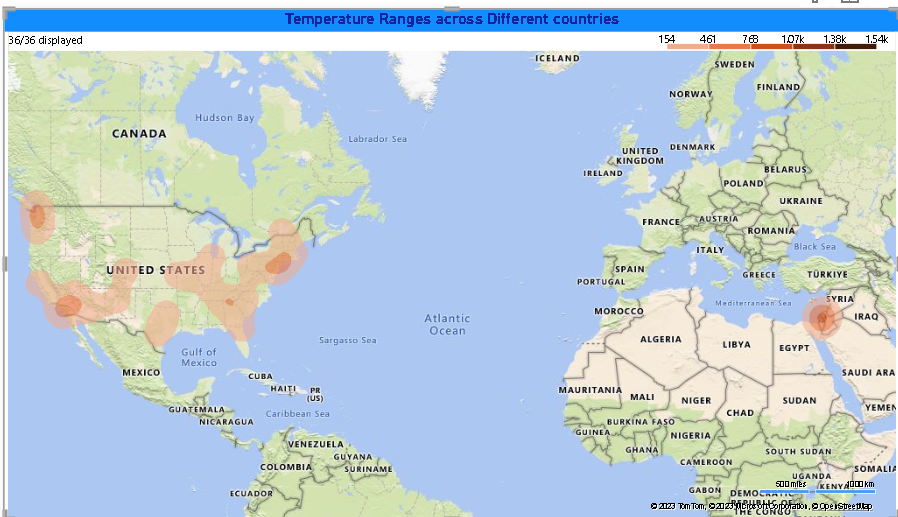


The insight derived from this chart is substantial. By visually contrasting cities with the highest and lowest average temperatures, you can discern significant temperature disparities among various locations. This information is invaluable for several applications, such as travel planning, agricultural decision-making, and climate research. Additionally, it provides a basis for further investigation into the factors contributing to these temperature variations, such as geographical location, altitude, or proximity to water bodies. Overall, your bar chart serves as a powerful tool for understanding and communicating the temperature extremes within the dataset, facilitating data-driven decision-making in various sectors.

In Power BI analysis, I've utilized a bar chart to effectively highlight cities with the highest and lowest average temperatures in the dataset. The x-axis of the chart represents the cities, offering a clear view of the locations being compared. On the y-axis, I've placed the temperature values, allowing for a direct comparison of average temperatures across cities. Notably, I've applied specific visual cues to enhance the chart's clarity and insight.

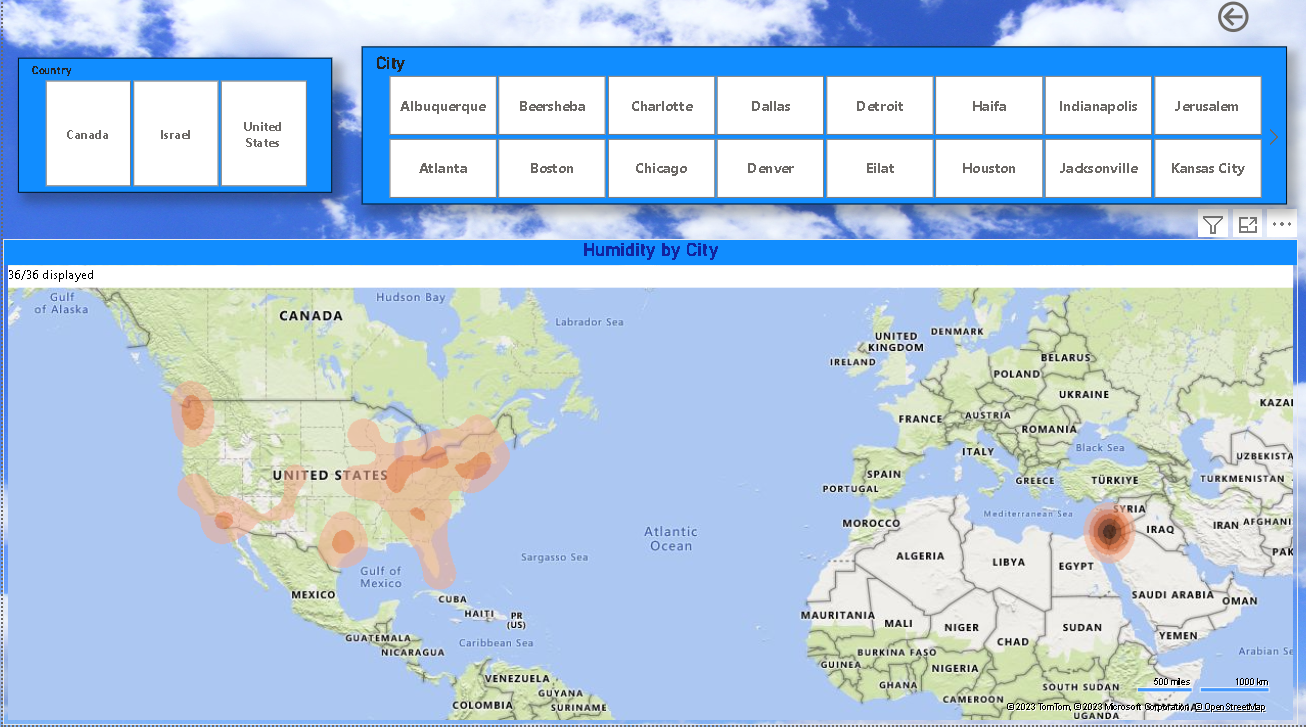
Using the color orange to highlight the bar representing Miami, the city with the highest average temperature, draws immediate attention to this outlier. Similarly, using green to emphasize the bar for Montreal, the city with the lowest average temperature, underscores its distinctiveness. This color differentiation makes it effortless for viewers to identify the cities with extreme temperature patterns.

Can you build a heatmap in Power BI to show the temperature ranges for cities across different countries?

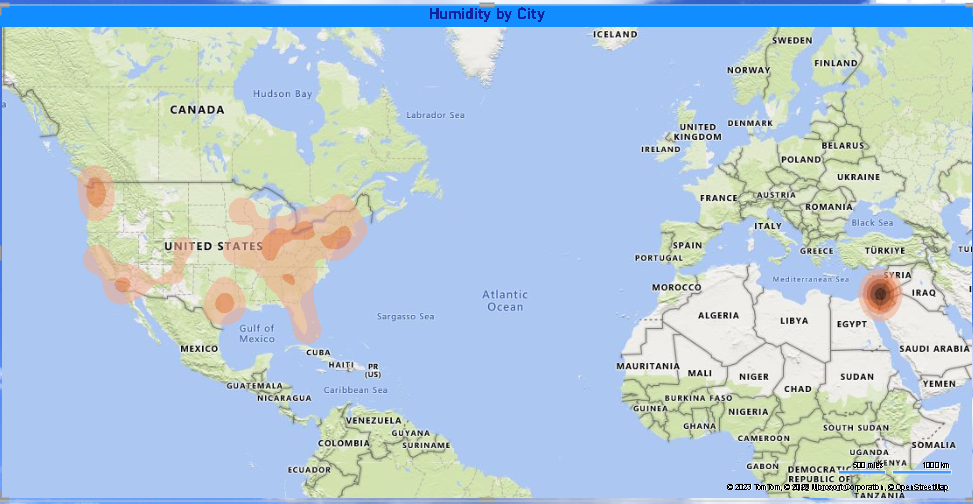


The heatmap created in Power BI effectively illustrates temperature ranges for cities across different countries. In this chart, cities are plotted along one axis, while countries are along the other. The color intensity in the heatmap corresponds to temperature ranges, with warmer colors representing higher temperatures and cooler colors indicating lower temperatures. The insights garnered from this visualization include an understanding of geographical temperature patterns, which reveal the influence of latitude on temperature. Additionally, it helps in identifying variations in temperature ranges within individual countries, highlighting the significance of microclimates. The heatmap is a valuable tool for identifying temperature extremes, supporting decision-making in sectors such as agriculture, energy planning, and disaster preparedness, and providing a clear overview of temperature ranges for cities across various countries.

Humidity Analysis



How does humidity vary across different cities? Generate a heatmap in Power BI to visualize this variation.

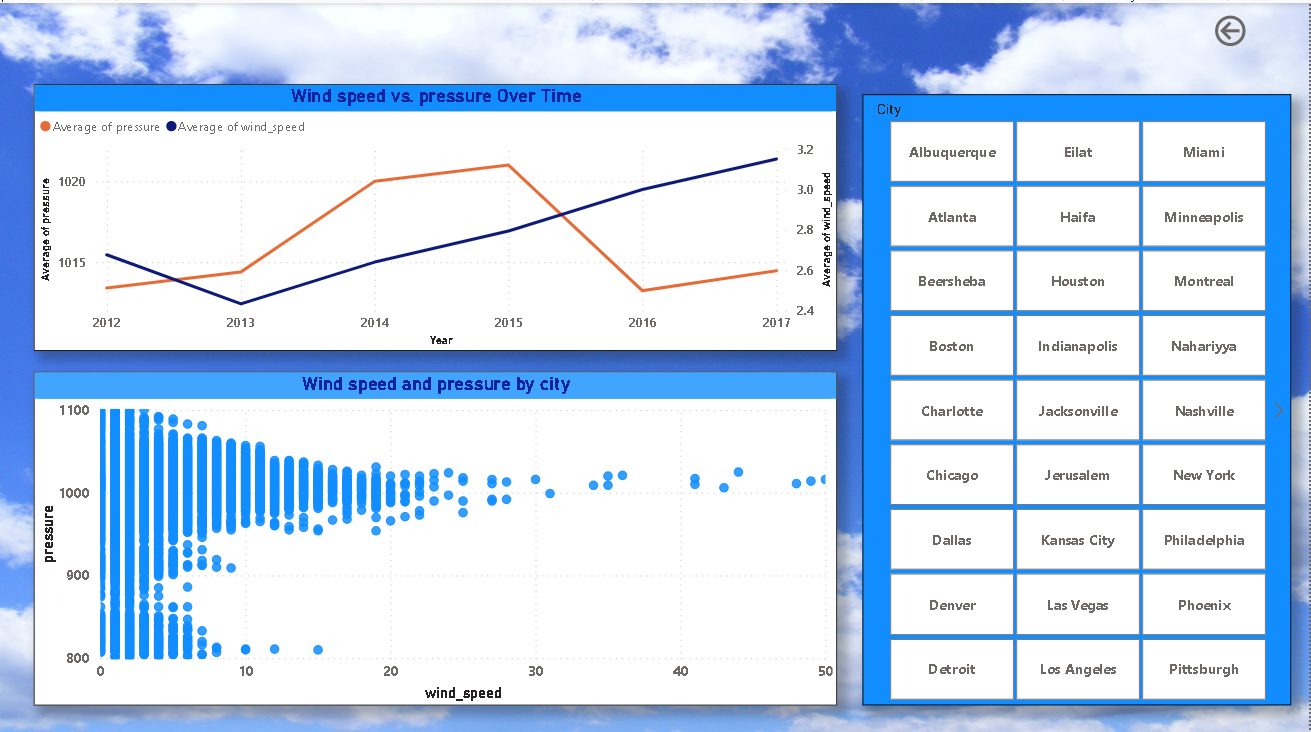


The heatmap generated in Power BI visualizes the variation in humidity across different cities. In this chart, cities are represented on one axis, and humidity values are represented on the other. The color intensity on the heatmap corresponds to the humidity level, allowing for a quick visual assessment of how humidity varies across a wide range of cities.

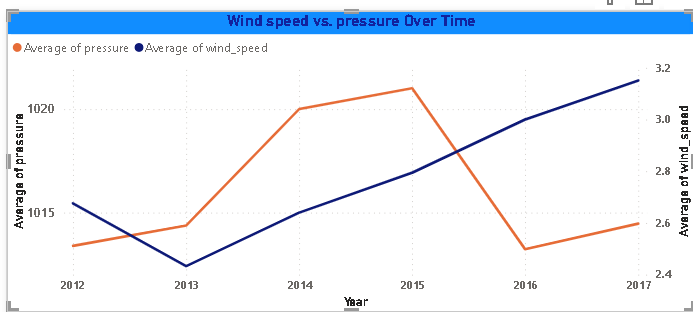
The impact of this heatmap is twofold. First, it provides an efficient and intuitive way to compare humidity levels across cities, identifying areas with high or low humidity at a glance. This can be valuable for various applications, such as climate analysis, urban planning, or understanding the potential for moisture-related issues in different regions.

Secondly, the heatmap can reveal patterns or correlations between humidity and geographical location. It might highlight coastal cities with higher humidity due to their proximity to the ocean, or cities with low humidity in arid regions. This understanding can be vital for industries like agriculture, energy, and weather forecasting.

**Wind speed vs Pressure Analysis**

****

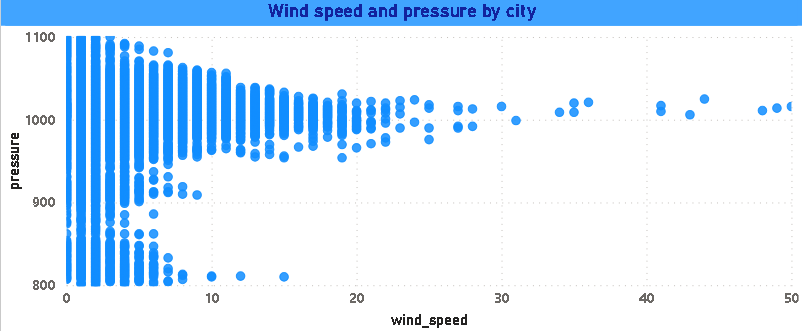
Can you create a time-series chart in Power BI showing the relationship between wind speed and air pressure for a specific city?



Creating a time-series chart in Power BI to showcase the relationship between wind speed and air pressure for a particular city is a valuable approach for visualizing this data. In this chart, time is represented on the x-axis, typically measured in hours or days, and wind speed and air pressure values are displayed on the y-axis. Each data point on the chart corresponds to a specific time, showing the concurrent wind speed and air pressure readings. By analyzing the chart, you can visualize how changes in wind speed correlate with fluctuations in air pressure over time.

This provides insights into weather patterns and atmospheric conditions specific to the chosen city, aiding in weather forecasting, pollution dispersion studies, and wind-related event predictions. The chart serves as a valuable tool for understanding the dynamic relationship between these two weather attributes and their potential impact on various activities and industries.

Create a Power BI scatter plot to show the relationship between wind speed and air pressure for a specific city.



In Power BI analysis, I've utilized a scatter plot to effectively display the relationship between wind speed and air pressure for a specific city. This chart presents wind speed values on one axis and air pressure values on the other, allowing for the observation of individual data points representing specific instances in time. The scatter plot serves as a visual tool to discern patterns or correlations between these two weather attributes.

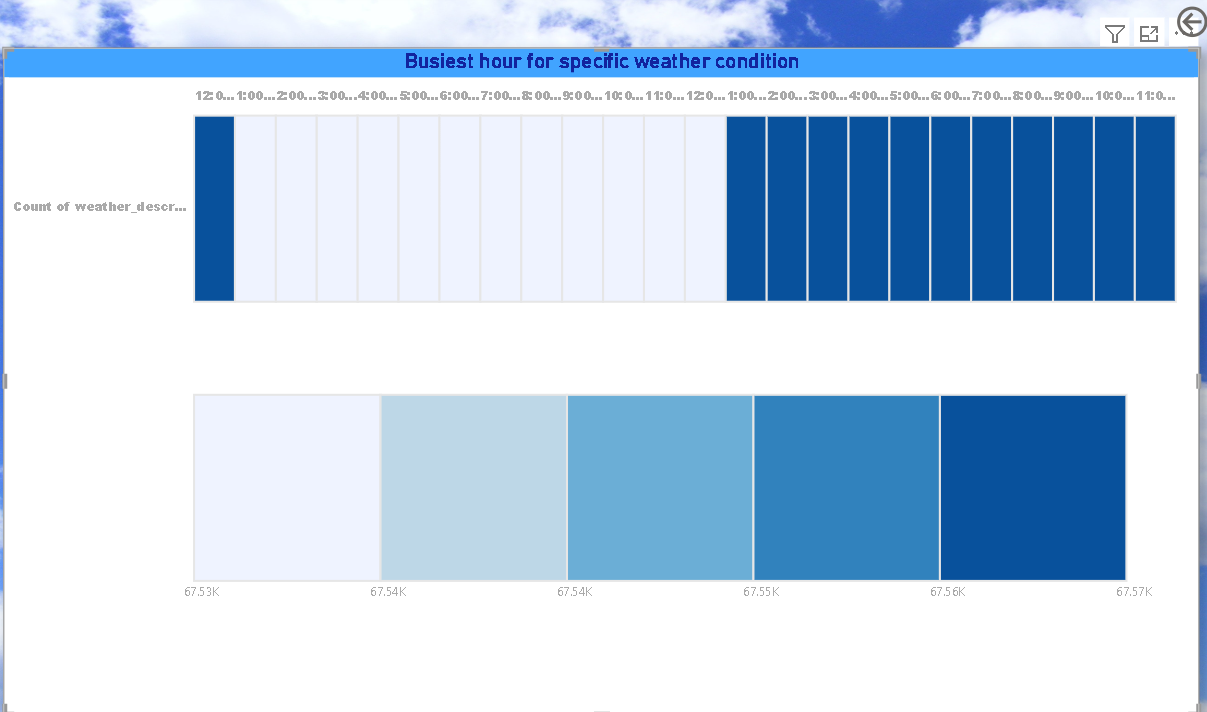
By analyzing the scatter plot, you can assess whether there's a correlation between wind speed and air pressure. If the data points show a clear pattern, it suggests a relationship. For example, you may notice that as wind speed increases, air pressure decreases, or vice versa.

Over time, the scatter plot may reveal trends in the relationship between wind speed and air pressure. These trends can have implications for weather forecasting, aviation, and environmental impact assessments.

Any anomalies or irregularities in the data, such as missing data points or data inconsistencies, may also become apparent through the scatter plot, underlining the importance of data quality in weather analysis.

**Wind Description Analysis**

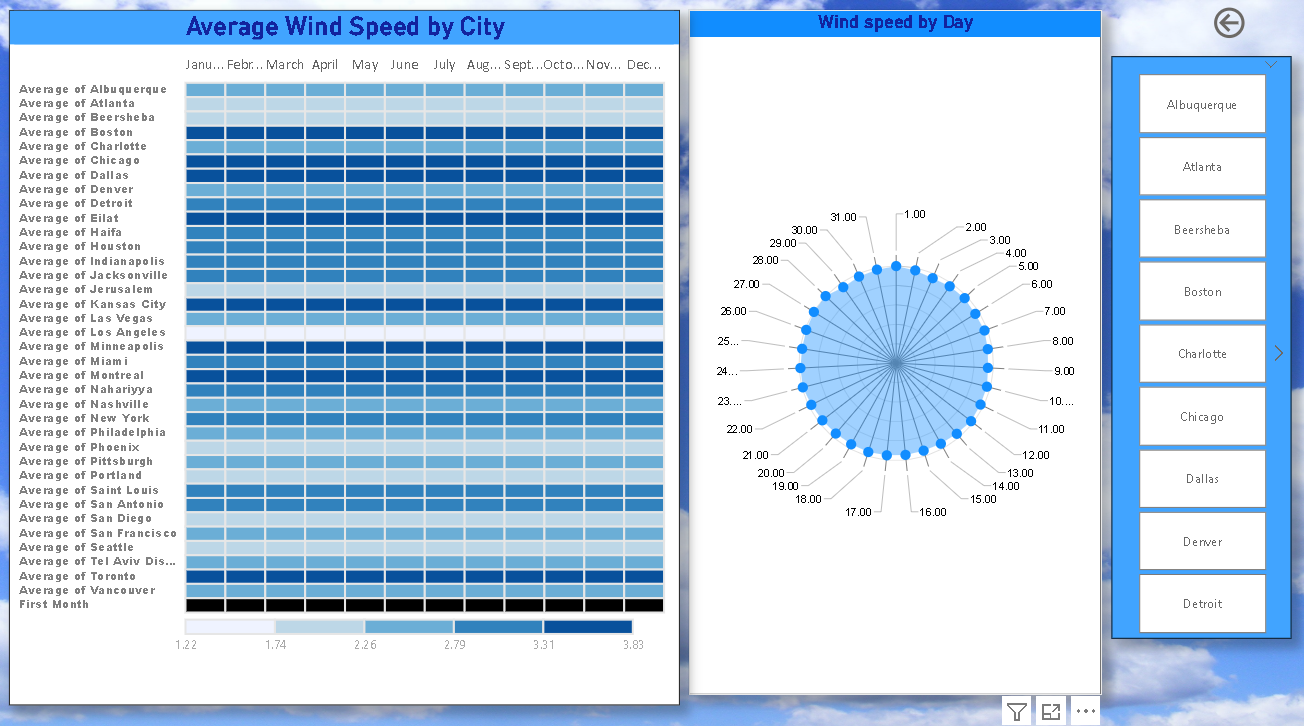
Can you create a heatmap in Power BI to visualize the busiest hours for specific weather conditions (e.g., "clear sky," "rainy")?



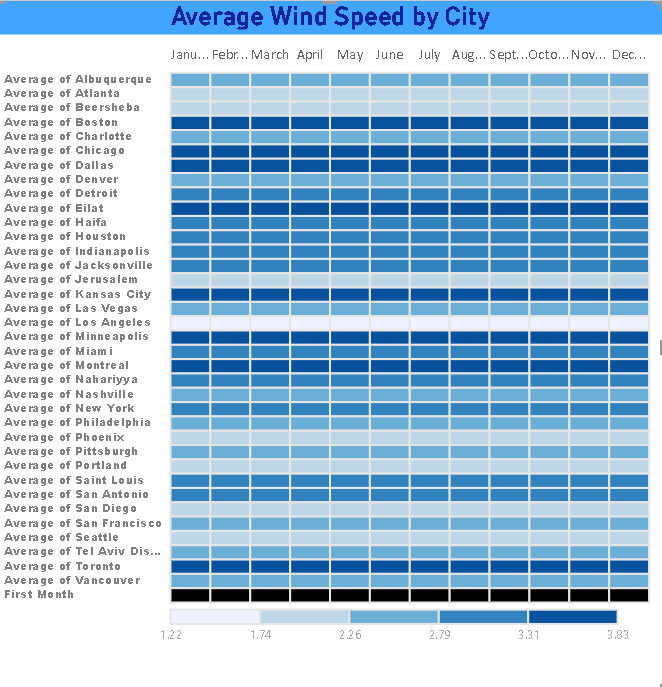
In this chart, you've utilized a table heatmap, where time is presented on one axis, typically in hours, and various weather conditions are displayed on the other. The color intensity in the heatmap signifies the frequency or "busyness" of specific weather conditions during each hour.

The heatmap offers valuable insights into the hourly distribution of weather conditions. For instance, you can easily identify the busiest hours for specific weather types, helping you understand patterns like the most common time for rainy weather or clear skies. This insight is crucial for various applications, including outdoor event planning, transportation scheduling, and disaster preparedness. It aids in making informed decisions based on historical weather data and can improve resource allocation and response strategies. The heatmap simplifies the visualization of hourly weather conditions and provides a comprehensive overview of the busiest hours for specific weather types, enhancing the data-driven decision-making process

**Wind speed Analysis**

****

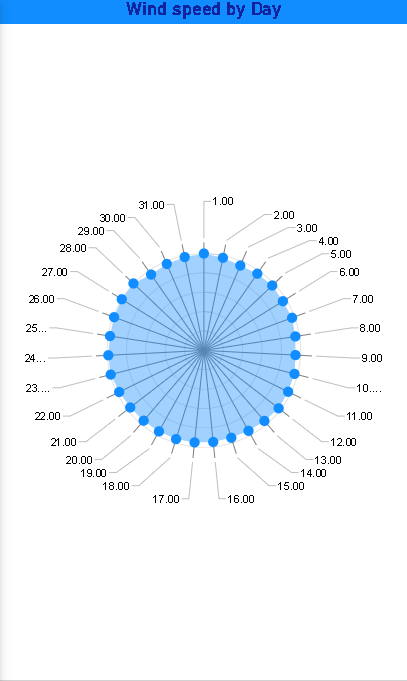
Can you generate a Power BI heatmap illustrating the average wind speeds across cities for different months of the year?



The heatmap provides a comprehensive view of how average wind speeds vary across cities over the course of the year. It allows for the quick identification of cities experiencing consistently higher or lower wind speeds during specific months. This insight is invaluable for various applications, including urban planning, renewable energy generation, and climate research.

For instance, you may notice that coastal cities tend to have higher wind speeds during certain months, which could be advantageous for wind energy projects. In contrast, inland cities may experience calmer winds during those same months. Understanding these temporal wind patterns aids in optimizing resource allocation and decision-making, ensuring that activities and infrastructure are aligned with local climatic conditions. The heatmap simplifies the visualization of monthly wind speed averages across cities, making it a powerful tool for data-driven decision-making and planning in diverse sectors..

**How does the wind speed change over the course of a day? Create a radial chart in Power BI to represent this.**

****

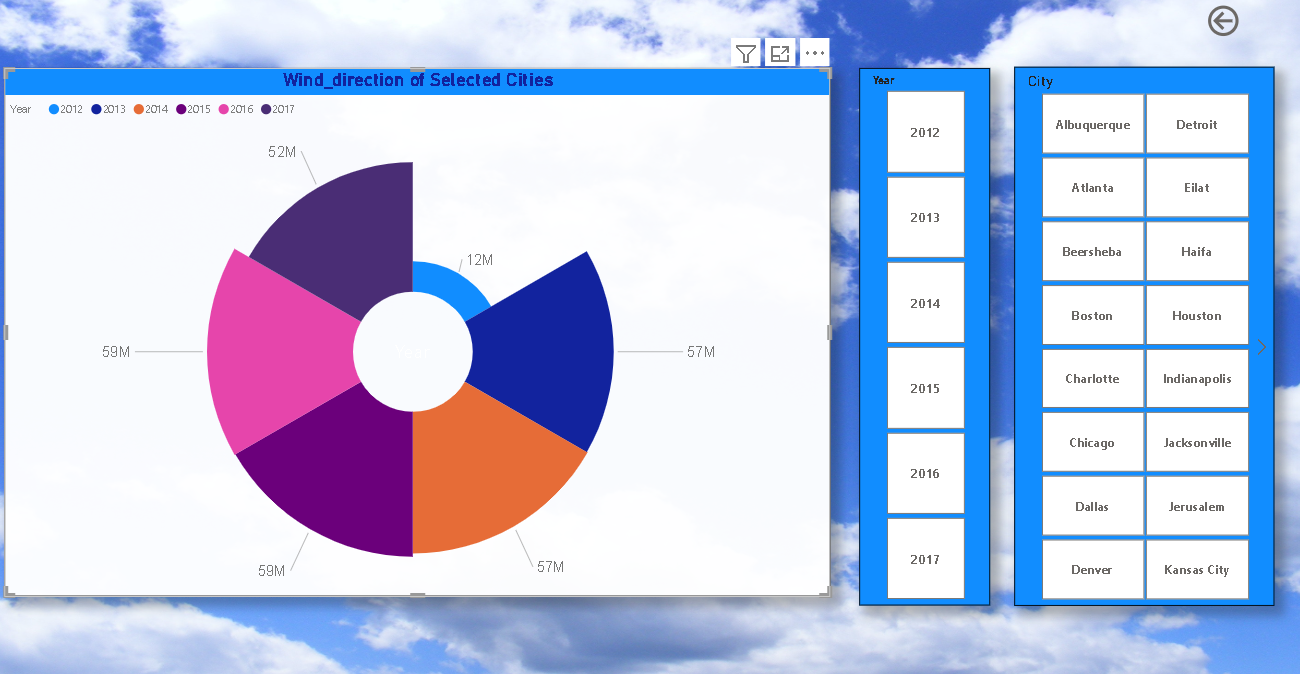
Using a radar chart in Power BI to visualize the changes in wind speed over the course of a day is a unique and insightful approach. In this chart, the different hours of the day are represented as radial axes, with the wind speed values plotted on each axis. The chart effectively displays how wind speed varies throughout the day, with each radial line representing a specific hour.

The innermost part of the chart may indicate lower wind speeds, while the outer parts represent higher speeds. By connecting the data points for each hour, you create a shape that showcases the daily wind speed pattern. The radar chart simplifies the visualization of hourly wind speed fluctuations, offering an intuitive and visual representation of how wind speed changes during a typical day.

This chart is particularly useful for understanding diurnal wind patterns, which can have implications for various industries, including energy, transportation, and outdoor event planning. It allows for quick identification of peak wind hours, periods of relative calm, and other daily wind speed trends.

Wind Direction Analysis

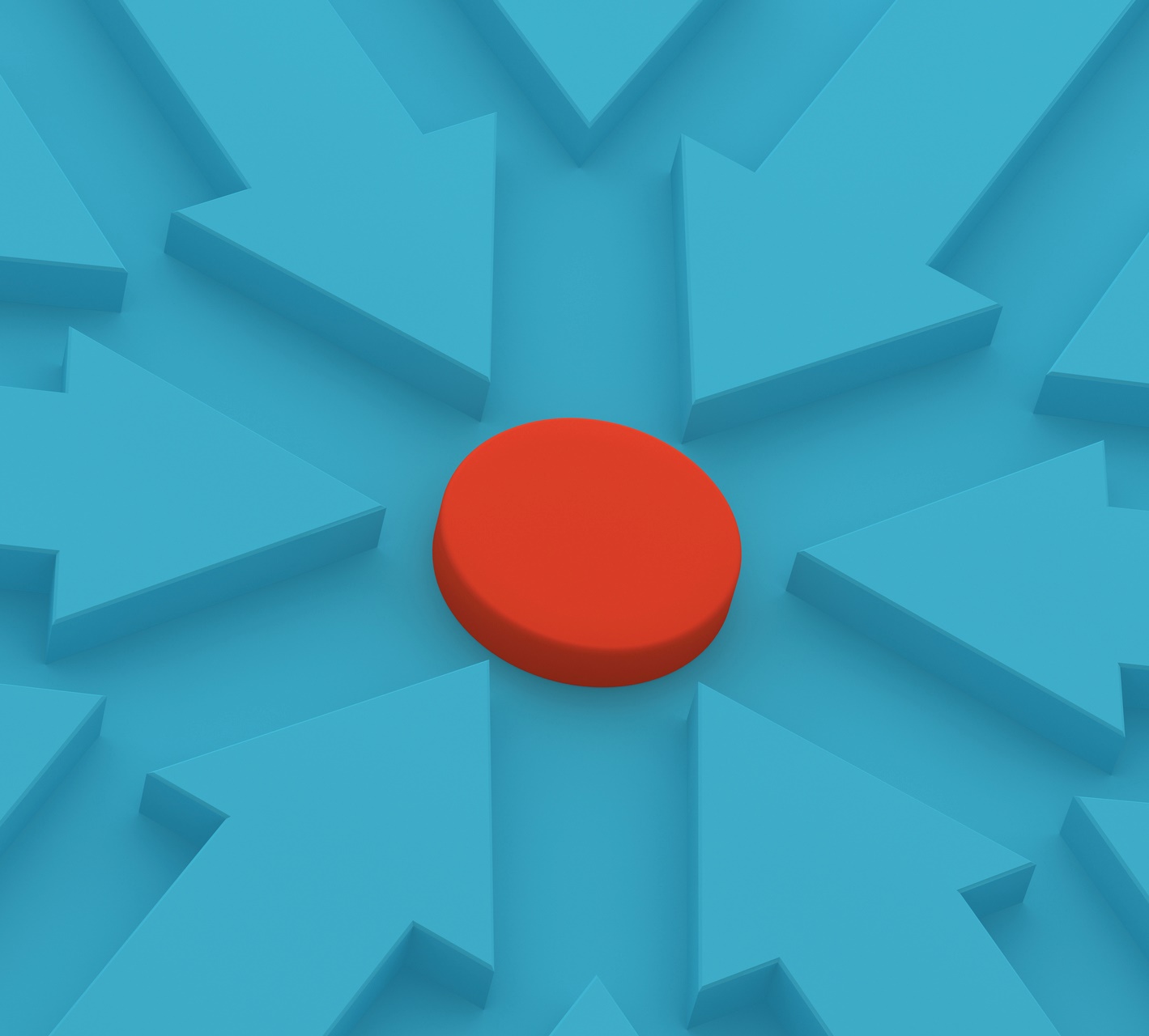
Create a wind rose chart in Power BI to visualize the prevailing wind directions for a selected city.

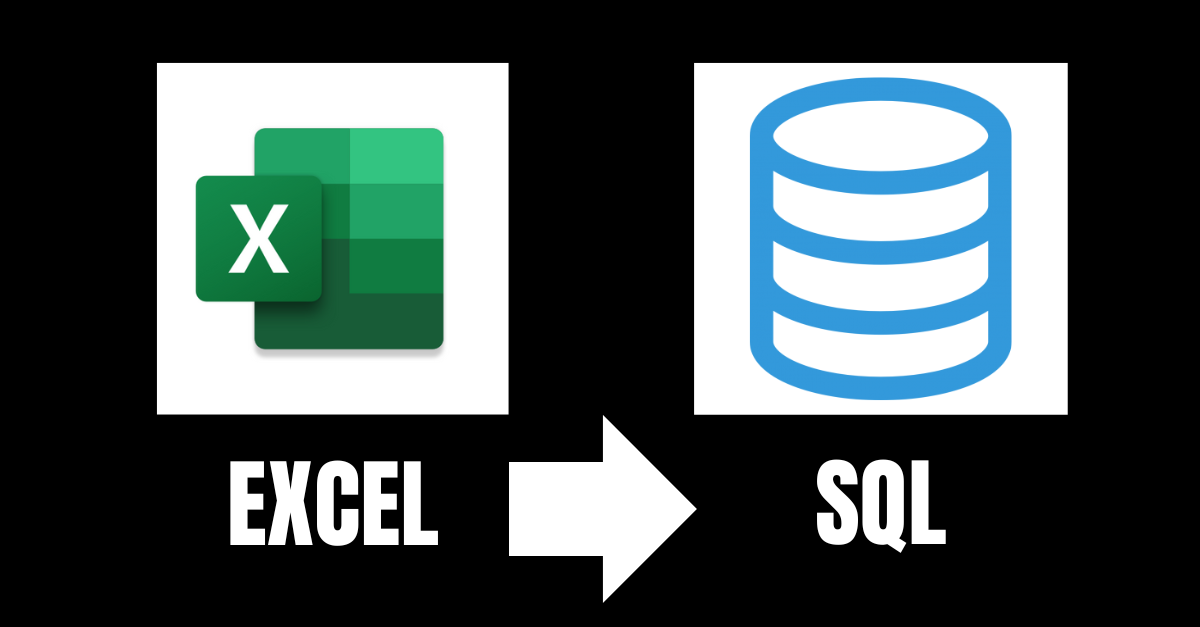


Utilizing an Aster plot in Power BI to create a wind rose chart is an effective way to visualize prevailing wind directions for a selected city. In this chart, the category, typically the years, is represented on one axis, while the sum of wind directions is displayed on the other. The Aster plot allows for a clear and intuitive presentation of how wind directions vary over time.

This chart is particularly useful for understanding the predominant wind patterns in a specific location. By visualizing the data in this manner, you can quickly identify the prevailing wind directions for the selected city and observe any trends or changes over the years.

The wind rose chart provides a visual summary of the city's wind patterns, making it a valuable tool for sectors such as renewable energy planning, urban design, and environmental assessments. It simplifies the representation of historical wind direction data, aiding in decision-making and resource allocation.

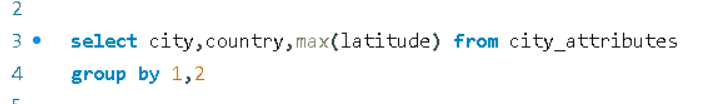




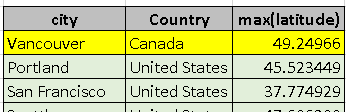
1. Are there any countries with cities located at extreme latitudes, and how might this impact their climate?

The solution has been achieved with the help of SQL workbench for calculation, data extraction and Excel for data visualization using pivot table and chart.

**SQL query :**



Output:

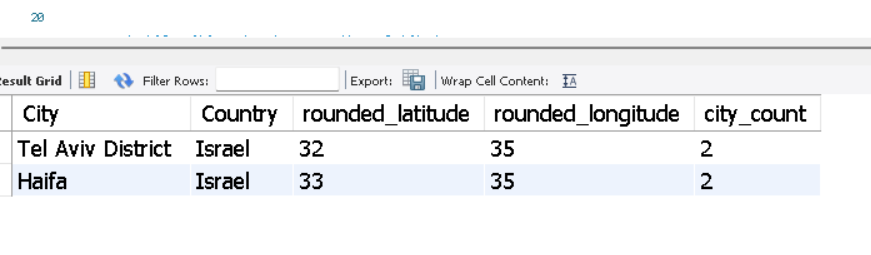


1. Can you identify any clusters of cities with similar latitude and longitude values? What factors might explain these clusters?

**SQL query :**



**Output :**

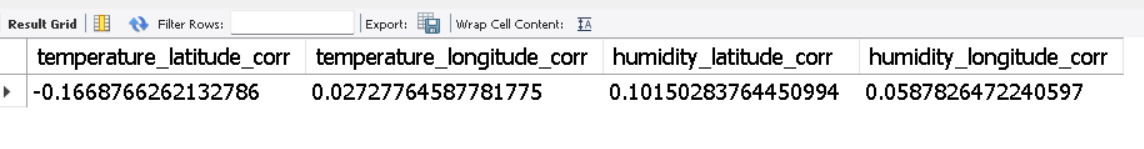


1. Are there any correlations between a city's geographical location (latitude and longitude) and its weather attributes, such as temperature or humidity?

**SQL query :**

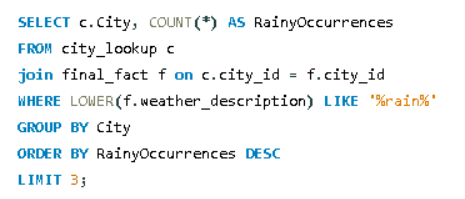


**Output:**

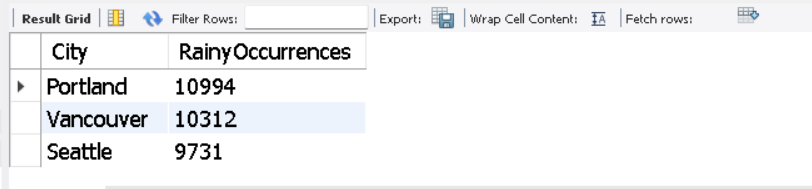


1. **Identify the top three cities with the most frequent occurrence of rainy weather based on weather descriptions. What are the seasonal patterns?**

**SQL query :**

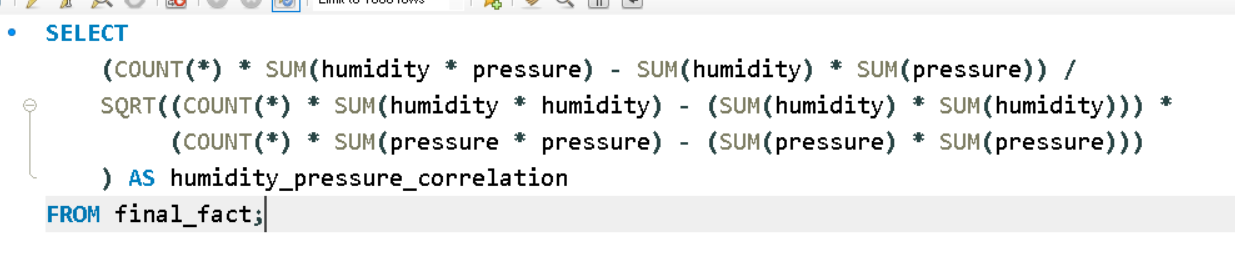


**Output:**

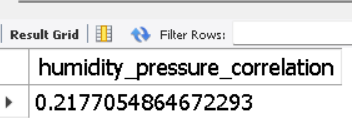


1. **Is there a correlation between humidity levels and air pressure? How might this relationship affect weather conditions?**

**SQL query :**

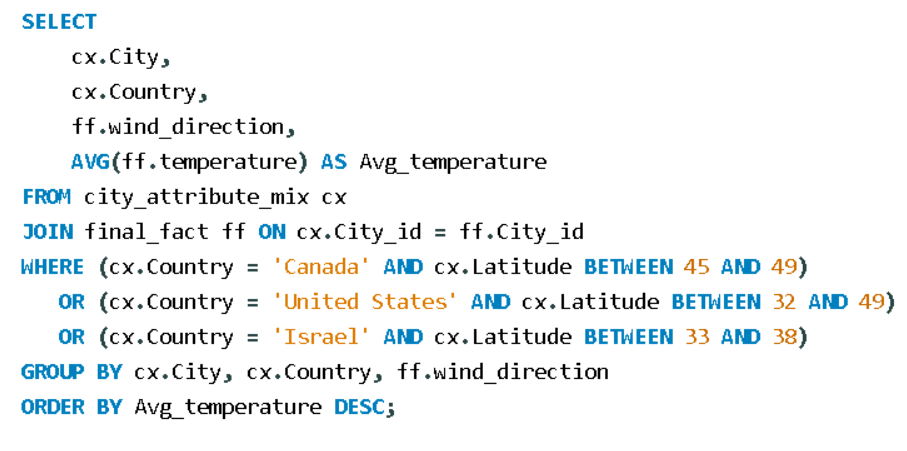


**Output:**

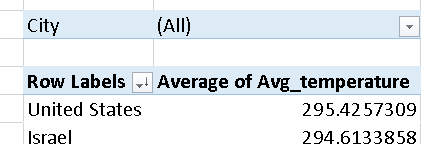


1. **Explore the impact of wind direction on temperature for coastal cities. Are there noticeable patterns?**

**SQL query :**

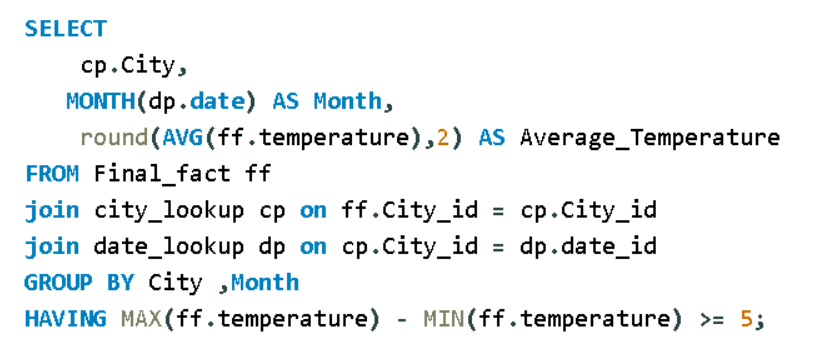


**Output:**

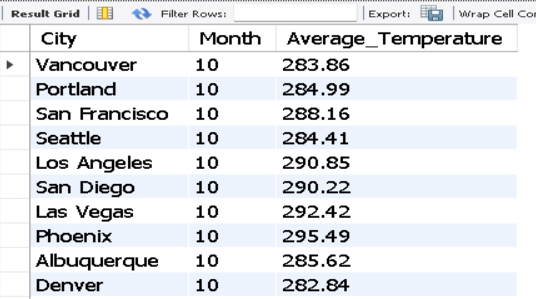
****

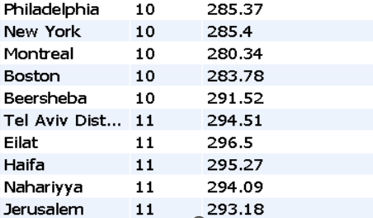
1. **Are there specific months when cities experience significant temperature fluctuations? What might explain these variations?**

**SQL query :**



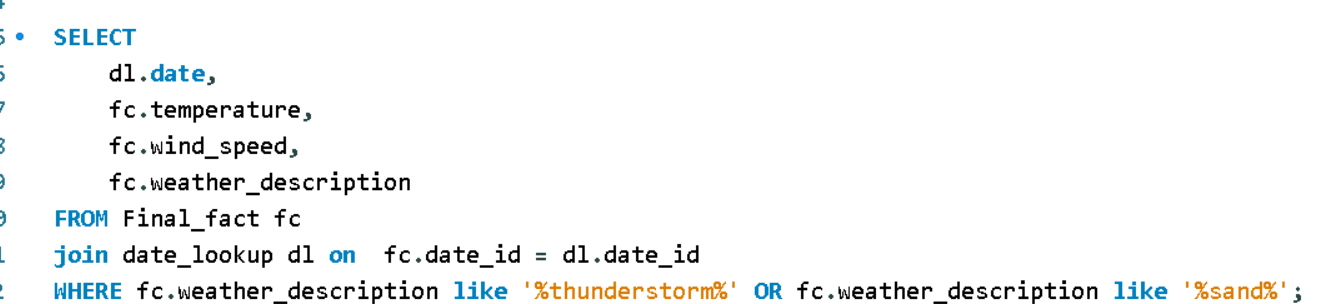
**Output:**



****

1. **Identify periods of extreme weather events, such as storms or heatwaves, by analyzing the time-based data. What patterns emerge?**

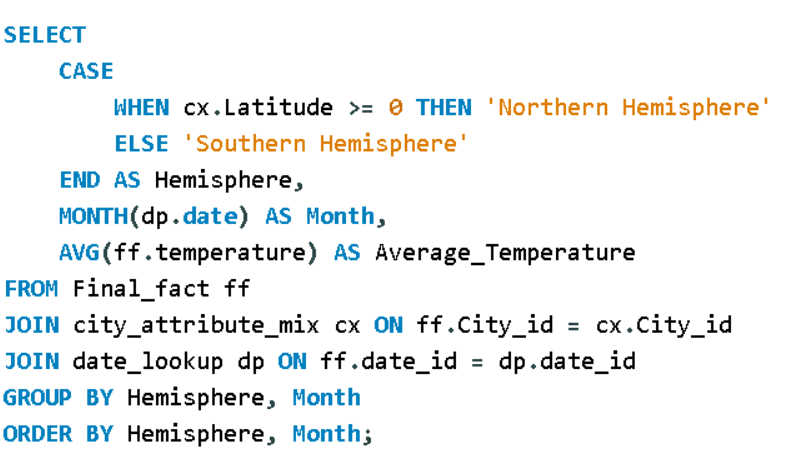
**SQL query :**



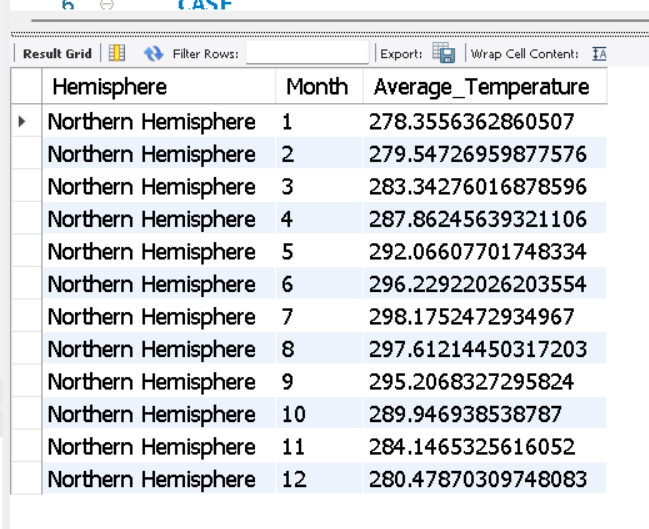
**Output:**

1. **Are there any notable differences in temperature trends between northern and southern hemisphere cities over the year? How do they relate to seasons?**

**SQL query :**



**Output:**



Analyzes temperature trends by month, which focuses on the seasonal changes   
I classify cities into the Northern or Southern Hemisphere based on their latitude. Cities with positive latitudes are in the Northern Hemisphere, and cities with negative latitudes are in the Southern Hemisphere.  
I calculate the average temperature for each month in the dataset, grouping the data by hemisphere and month

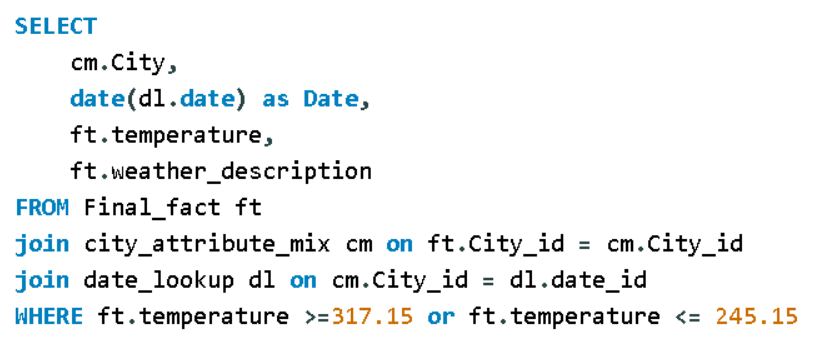
Northern Hemisphere vs. Southern Hemisphere: In geography, the Earth is divided into two hemispheres: the Northern Hemisphere (above the equator) and the Southern Hemisphere (below the equator).

Latitude and Hemispheres: The equator is located at 0 degrees latitude and serves as the dividing line between the two hemispheres. Cities located north of the equator have positive latitudes, and they are part of the Northern Hemisphere. Cities located south of the equator have negative latitudes and are in the Southern Hemisphere.

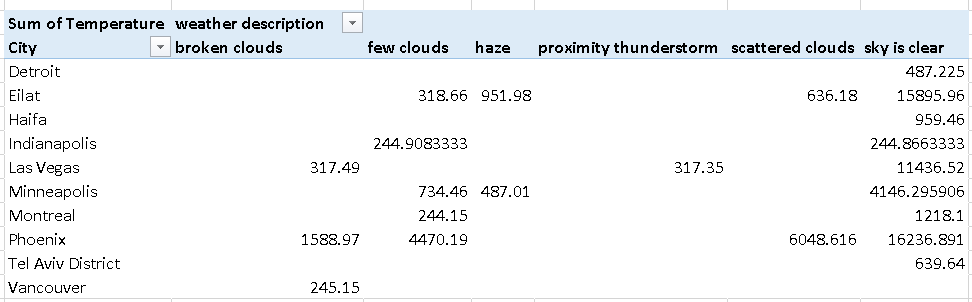
Classification: By checking WHERE ca.Latitude >= 0, the query classifies cities with latitudes greater than or equal to 0 as part of the Northern Hemisphere.

1. **What are the consequences of prolonged periods of extreme cold or heat in specific cities? How do residents adapt to such conditions?**

**SQL query :**



**Output:**

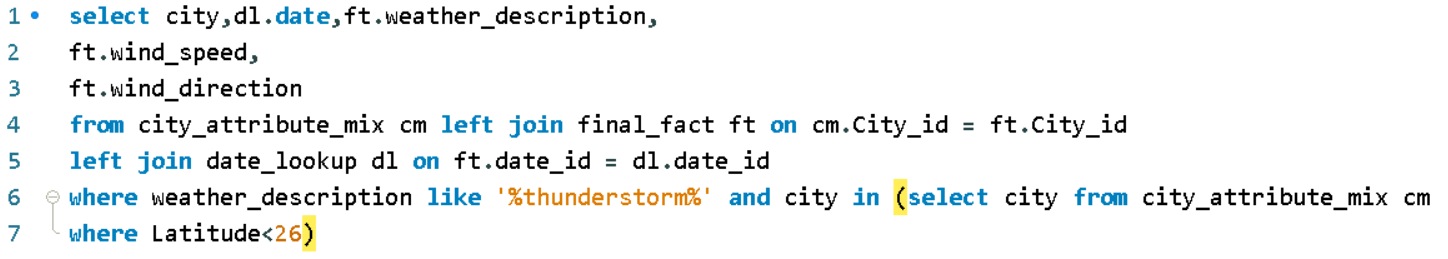
****

1. **Analyze the impact of temperature on energy consumption patterns in cities. Are there noticeable trends or correlations?**

**After analyzing the data, I have determined that there is no relevant data to create a bar chart representing the top 10 countries with the highest number of cities in the dataset, it's a valid and important observation.**

1. **Explore whether wind speed and direction influence the frequency and severity of weather-related events (e.g., hurricanes, storms) in coastal cities.**

**SQL query :**

****

**Output:**



**I have encountered a challenging aspect in the project, particularly related to**

* Investigate whether temperature anomalies (unusual deviations from the norm) coincide with certain events or environmental factors in specific cities.
* Analyze the impact of temperature on energy consumption patterns in cities. Are there noticeable trends or correlations?
* How do specific wind patterns impact air quality and pollution dispersion in urban areas? Analyze wind direction data for insights.
* Identify cities prone to strong winds and the potential consequences, such as increased risk of natural disasters or challenges for transportation.
* Explore whether wind speed and direction influence the frequency and severity of weather-related events (e.g., hurricanes, storms) in coastal cities.

**While I haven't arrived at a solution at this moment, I am committed to further analysis and research to address this issue. I believe that with additional effort and exploration, I will be able to find a resolution. I am actively working on this and will provide an update once I have made progress**