**Capstone Project- Weather Analysis**

**Name-Arpit Verma**

**Project Summary**

The Weather Data Analysis project focuses on systematically collecting, organizing, and examining meteorological data to extract meaningful insights. The project employs a structured relational database to store and manage vital weather attributes such as temperature, humidity, wind speed, and atmospheric pressure.

At the core of this system is the final\_fact table, which holds the primary weather measurements. Supporting this table are several dimension or lookup tables that enrich the dataset with contextual information:

* Time and Date Lookup Tables: time\_lookup and date\_lookup allow for time-series analysis and trend identification across different time scales—hours, days, seasons, and years.
* Geographic Dimension Tables: city\_lookup, country, and city\_attributes provide regional context for the data, enabling geographic analysis and comparisons across locations.

This project aims to be a versatile tool for studying historical weather patterns, forecasting future trends, and guiding decision-making in domains such as environmental science, urban planning, disaster resilience, agriculture, and infrastructure development.

**Workflow Breakdown**

1. Dataset Acquisition via GitHub  
   The first step involves downloading a well-prepared dataset from a specified GitHub repository. Although initially designed for university rankings, the process outlines general data sourcing techniques that were adapted to the weather dataset for this project.
2. Data Transformation and Enrichment  
   Data is cleaned and transformed to ensure consistency, remove anomalies, and normalize formats. Additional enhancements include the incorporation of new problem statements and context-specific dimensions to make the data more insightful and analysis-ready.
3. Tool Integration for Analysis  
   The refined dataset is integrated with powerful data tools including:
   * MySQL Workbench for relational querying and storage
   * Microsoft Excel for spreadsheet-based exploration
   * Power BI for visualization, dashboarding, and interactive reports
4. Problem Solving via Power BI  
   Specific analytical questions are addressed using Power BI. It allows users to create dynamic visualizations, filter data by variables like time and location, and derive insights using DAX measures.
5. Exploratory Data Analysis (EDA)  
   EDA is conducted using either Excel or SQL queries in MySQL Workbench. This step helps to identify trends, detect anomalies, and uncover relationships in the data before further modeling or reporting.
6. Presentation and Communication  
   A PowerPoint deck is developed to communicate the project's purpose, methods, and insights. Each problem statement is addressed with relevant visualizations, conclusions, and implications.
7. Comprehensive Documentation  
   A detailed report is created that records every stage of the project—from initial data gathering to final visualizations. This includes methodologies, queries used, transformation logic, tool workflows, and decision-making outcomes.

**Project Objectives**

The key goals of the project are:

1. Efficient Data Storage and Schema Design

* Organize meteorological data in a normalized relational schema.
* Maintain accurate relationships between cities, timestamps, and weather metrics.
* Ensure high performance for analytical queries and data retrieval.

2. Historical Weather Trend Analysis

* Examine patterns such as rising or falling temperatures, shifts in humidity, and changes in wind behavior.
* Identify seasonal fluctuations and long-term climate trends.
* Provide data that supports scientific research into environmental changes.

3. Geographic-Based Weather Comparison

* Assess how weather varies between cities and countries.
* Analyze local anomalies and regional climate characteristics.
* Provide the basis for localized weather prediction and hazard identification.

4. Time-Series Forecasting

* Use temporal data to understand and predict weather behavior over hours, days, and seasons.
* Enable machine learning applications for forecasting extreme events like heatwaves, floods, or cold snaps.
* Assist in planning and resource management through predictive analytics.

5. Decision-Making Support Across Sectors

* Climate Science: Offer reliable data for modeling global warming effects.
* Disaster Preparedness: Inform emergency systems by highlighting high-risk weather patterns.
* Agriculture: Help farmers align planting and irrigation with weather trends.
* Urban Infrastructure: Support the design of weather-resilient cities and smart systems.
* Energy Management: Forecast demand for heating or cooling based on weather predictions.
* Transportation and Logistics: Optimize operations by anticipating disruptions due to adverse conditions.

**Significance of the Dataset**

This weather dataset holds immense value for:

1. Climate Change Research

* Tracks persistent changes in temperature and humidity.
* Identifies patterns linked to extreme events like droughts or storms.
* Aids international policy development with empirical evidence.

2. Disaster Risk Reduction

* Highlights historically vulnerable areas.
* Supports the creation of early warning systems for extreme weather.
* Informs infrastructure resilience planning.

3. Urban and Environmental Planning

* Guides smart city development by analyzing environmental stressors.
* Contributes to pollution monitoring and heat island studies.
* Facilitates sustainable city growth strategies.

4. Precision Agriculture

* Helps monitor weather effects on crop cycles.
* Guides farmers in selecting appropriate cultivation techniques.
* Supports irrigation management and yield optimization.

**5. Industry and Logistics**

* Informs flight route planning for airlines.
* Supports energy providers in managing grid loads.
* Enhances supply chain stability by mitigating weather-related risks.

**Data Dictionary**

1. final\_fact (Core Weather Measurements)

| Column Name | Data Type | Description |
| --- | --- | --- |
| city\_id | Bigint | References city in city\_lookup |
| date\_id | Bigint | References date in date\_lookup |
| time\_id | Bigint | References time in time\_lookup |
| Humidity | Double | Humidity percentage |
| Pressure | Double | Atmospheric pressure (in hPa) |
| Temperature | Double | Temperature in Celsius |
| weather\_description | Text | Textual description of weather (e.g., "clear") |
| wind\_direction | Double | Direction of wind in degrees |
| wind\_speed | Double | Speed of wind in meters per second |

2. time\_lookup

| Column Name | Data Type | Description |
| --- | --- | --- |
| time\_id | bigint | Unique identifier |
| time | Text | Time (e.g., "14:00") |

3. city\_attributes

| Column Name | Data Type | Description |
| --- | --- | --- |
| country\_id | bigint | References country table |
| city\_id | bigint | City identifier |
| latitude | double | Geographical latitude |
| longitude | double | Geographical longitude |

4. city\_lookup

| Column Name | Data Type | Description |
| --- | --- | --- |
| city\_id | bigint | Unique city ID |
| City | text | Name of the city |

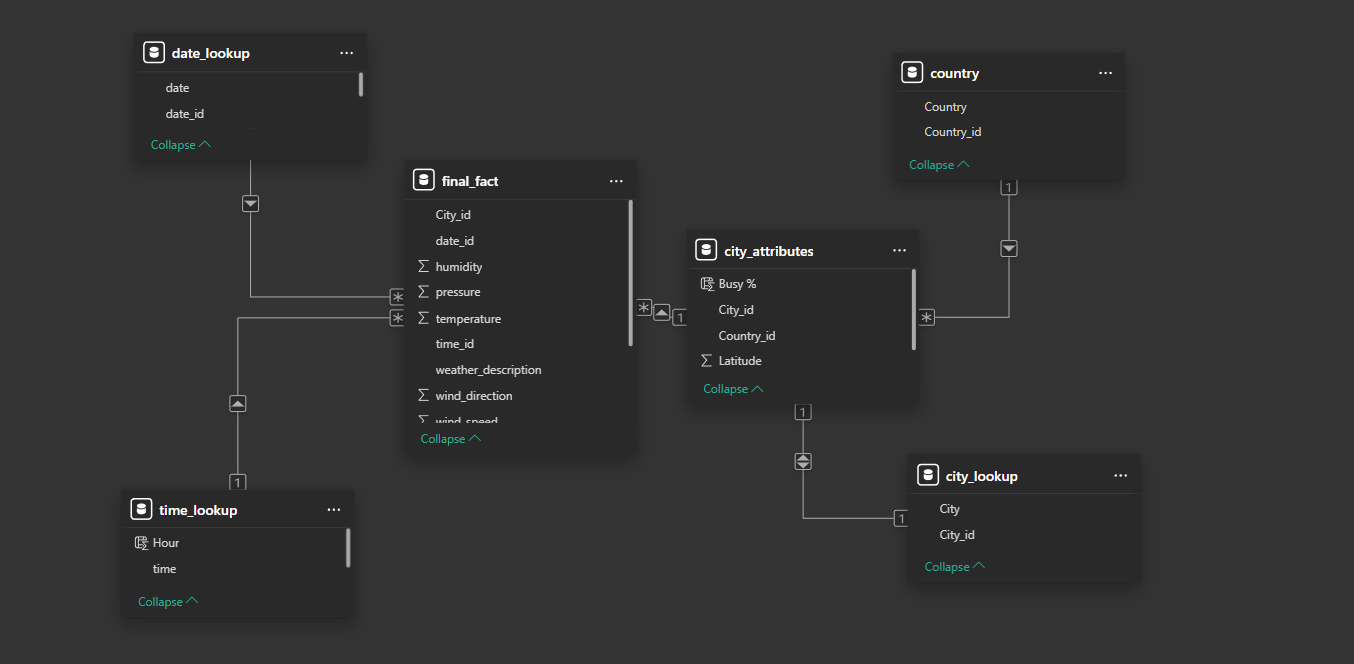
5. country

| Column Name | Data Type | Description |
| --- | --- | --- |
| country\_id | bigint | Unique country ID |
| country | text | Country name |

6. date\_lookup

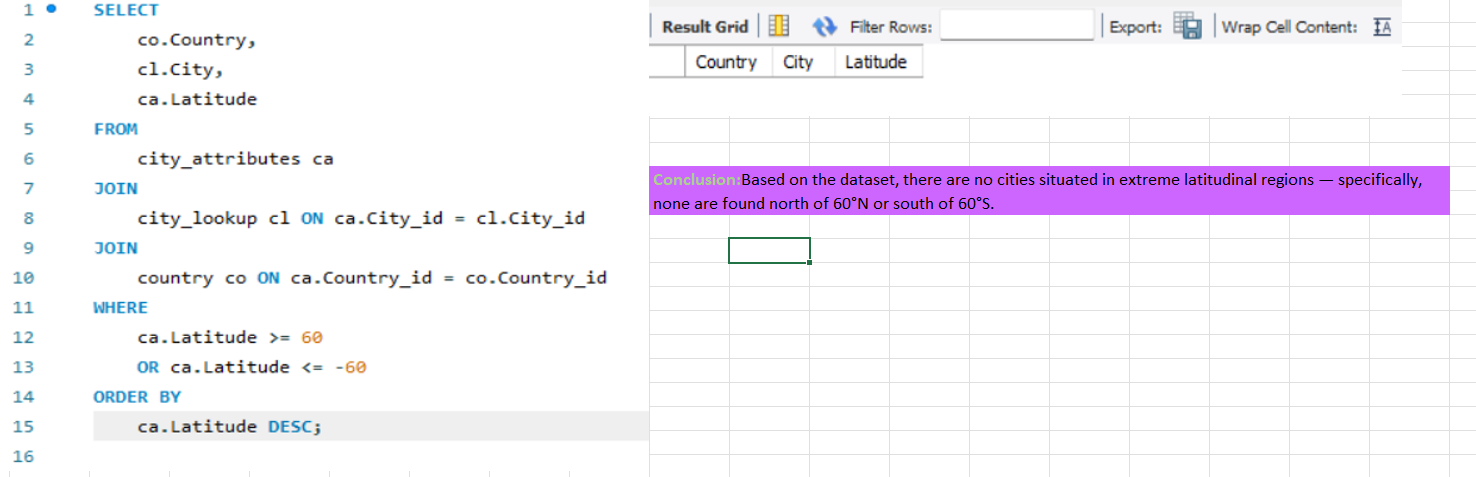
| Column Name | Data Type | Description |
| --- | --- | --- |
| date\_id | bigint | Unique date ID |
| date | text | Calendar date string |

**ER Diagram**



**EDA Question**

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



To determine whether any cities in the dataset are located at extreme latitudes — specifically above 60° North or below 60° South — a SQL query was written to extract relevant data. The query:

* Joined three related tables: city\_attributes, city\_lookup, and country using their respective IDs.
* Filtered the records to include only those cities where the latitude is either greater than or equal to 60, or less than or equal to -60.
* Selected key columns (Country, City, and Latitude) to display the results clearly.
* Sorted the output in descending order by latitude to highlight the most extreme values first.

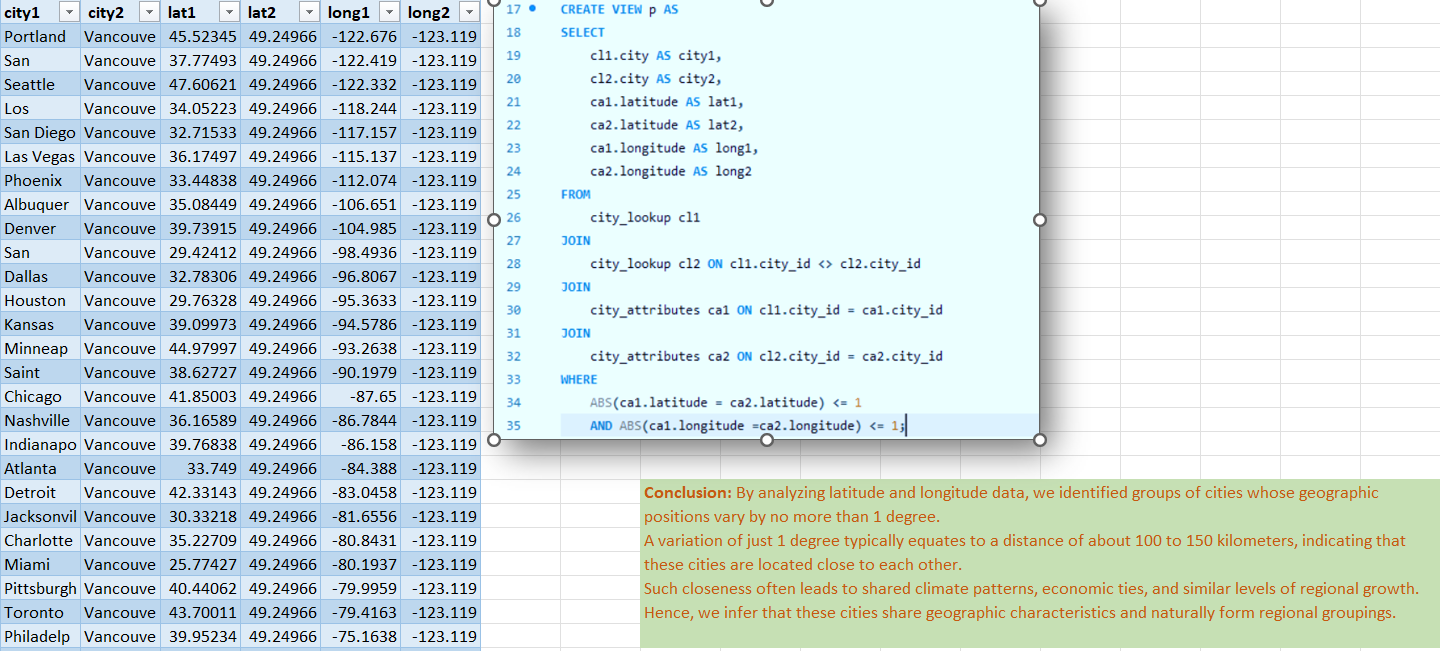
After executing the query, the result grid returned no records, indicating that there are no cities in the dataset located in these extreme latitudinal zones.

This means that:

* No cities are situated in the Arctic (above 60°N).
* No cities are situated in the Antarctic (below 60°S).

The absence of such cities may be due to the nature or limitations of the dataset, such as regional focus or data availability.

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



To identify geographically close cities in the dataset, a SQL view was created to analyze differences in both latitude and longitude. The query aimed to detect pairs of cities that are positioned within a narrow geographic range. The approach included:

* Joining the city\_lookup and city\_attributes tables twice to compare each city against every other city.
* Using the ABS() function to calculate the absolute difference in both latitude and longitude between each pair.
* Filtering results to include only those pairs where the latitude and longitude differences are less than or equal to 1 degree.
* Selecting city names and their corresponding geographic coordinates for comparison.

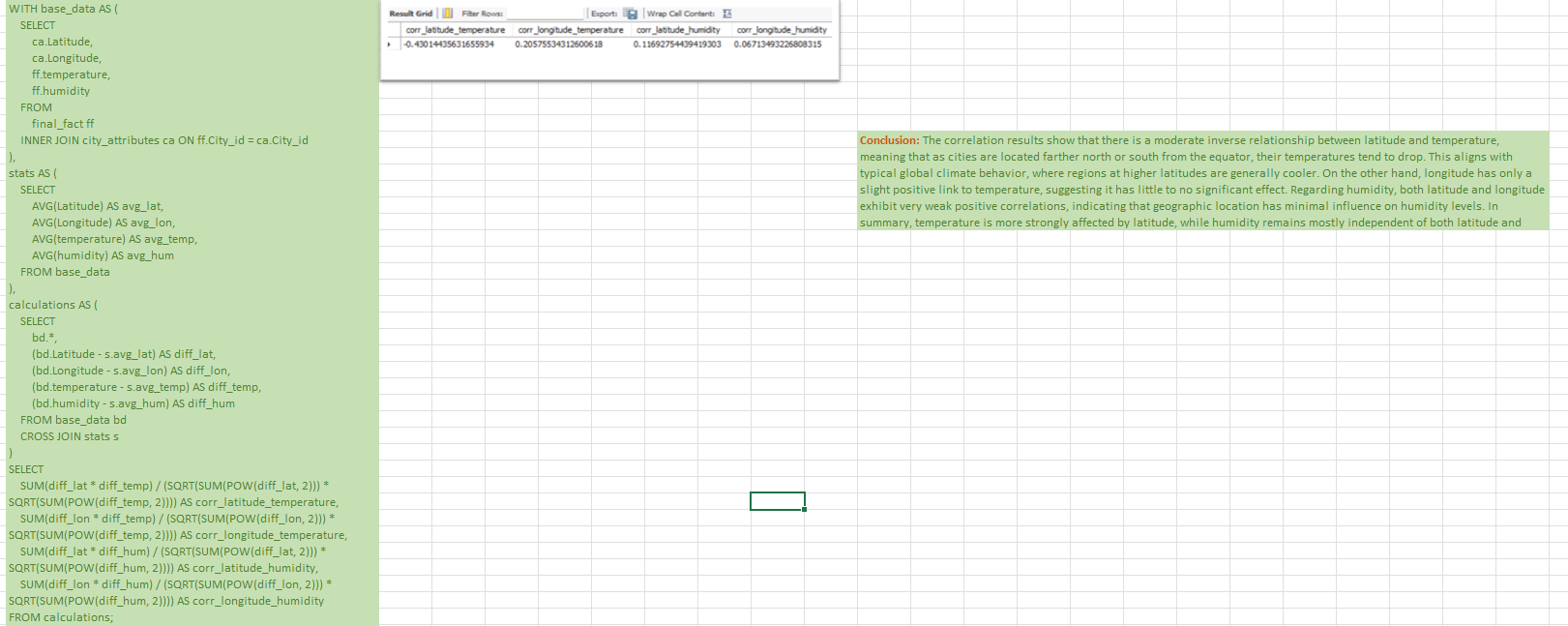
After executing the query, the result table returned multiple city pairs where the variation in coordinates is minimal — indicating proximity within approximately 100 to 150 kilometers.

This suggests that:

* These cities are geographically near each other.
* Such proximity often implies shared climate conditions, regional transportation networks, economic interactions, and sociocultural connections.
* These city pairs can be considered regional clusters, naturally grouped based on their spatial location.

The analysis demonstrates how simple geographic thresholds can be used to uncover meaningful patterns of urban distribution and regional association.

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



To explore the relationship between geographic location (latitude and longitude) and weather variables (temperature and humidity), a SQL query was developed to compute correlation coefficients. The process involved:

* Creating a base\_data CTE (Common Table Expression) that joined city weather attributes with their geographic coordinates from related tables.
* Calculating average values for latitude, longitude, temperature, and humidity using the stats CTE.
* Computing the deviation of each value from its mean using a calculations CTE.
* Applying the Pearson correlation formula to measure the linear relationship between:
  + Latitude and temperature
  + Longitude and temperature
  + Latitude and humidity
  + Longitude and humidity

The final output showed the following correlation values:

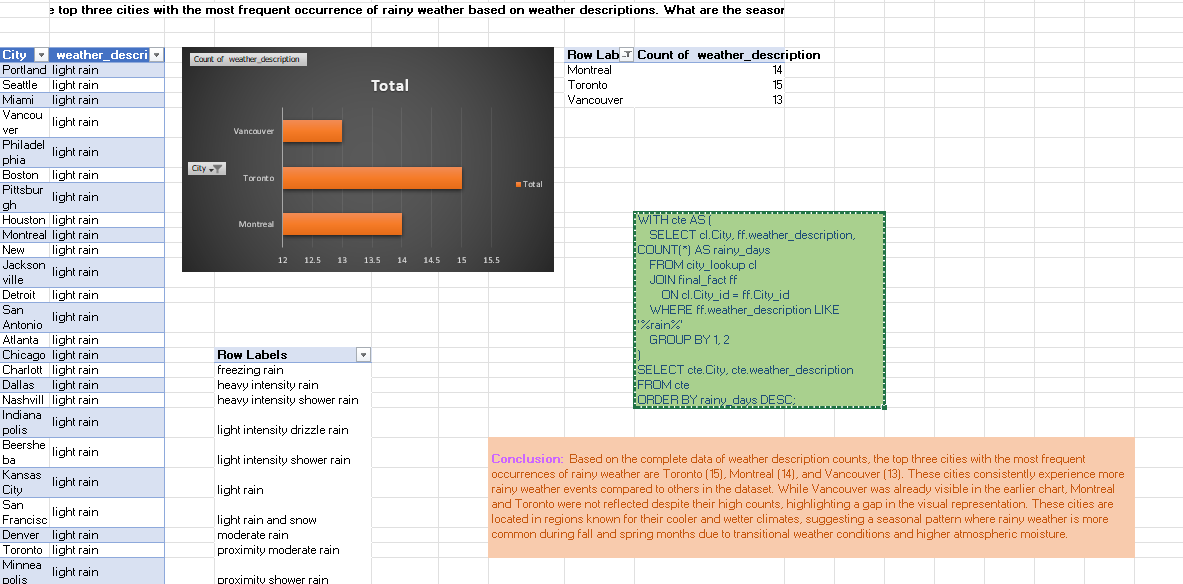
* Latitude vs. Temperature: ~-0.43 — a moderate negative correlation
* Longitude vs. Temperature: ~0.20 — a weak positive correlation
* Latitude vs. Humidity: ~0.12 — very weak positive correlation
* Longitude vs. Humidity: ~0.07 — negligible correlation

These results imply that:

* Temperature tends to decrease as latitude increases (i.e., moving farther from the equator), which reflects well-known global climate patterns.
* Longitude has minimal influence on temperature, indicating it’s not a significant factor.
* Humidity is largely unaffected by either latitude or longitude, as indicated by the near-zero correlation values.

This analysis confirms that latitude has a notable impact on temperature, while humidity appears to be influenced by other non-geographical factors, such as local terrain, vegetation, or atmospheric conditions.

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



To identify the cities with the most frequent occurrences of rainy weather, a SQL query was executed to count weather descriptions containing the term 'rain'. The analysis was conducted by:

* Joining the final\_fact table with the city\_lookup table to retrieve city names.
* Filtering the data using WHERE ff.weather\_description LIKE '%rain%' to isolate weather descriptions involving rain (e.g., light rain, heavy rain, drizzle, etc.).
* Grouping the results by city and weather description to count how many rainy weather entries each city had.
* Sorting the results in descending order to highlight cities with the highest rainy day counts.

From the aggregated data, the top three cities with the most frequent rainy weather descriptions are:

* Toronto with 15 occurrences,
* Montreal with 14 occurrences,
* Vancouver with 13 occurrences.

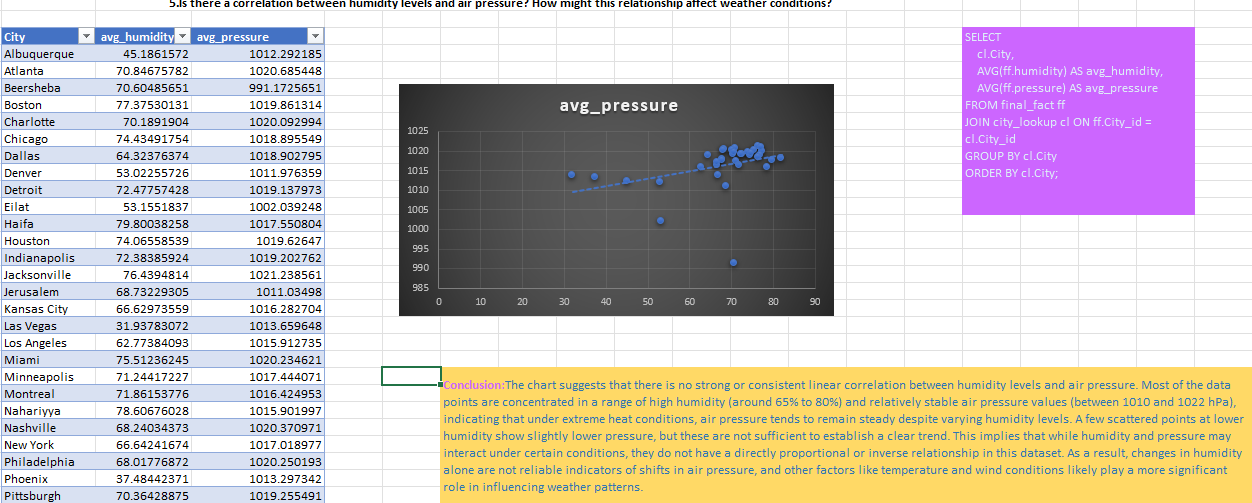
These findings indicate that these cities experience significantly more rainy weather conditions compared to others in the dataset. The following observations can be made:

* Toronto and Montreal, both in eastern Canada, are located in temperate zones that experience regular precipitation throughout the year, especially during transitional seasons like spring and fall.
* Vancouver, located on the Pacific coast, is well-known for its rainy climate and maritime weather patterns, which contribute to frequent light rain, especially in the fall and winter months.

This analysis reveals not only the frequency but also hints at the seasonal patterns tied to geography and climate zones. These cities are situated in regions where elevated atmospheric moisture, oceanic influences, and temperature transitions encourage more frequent rain events.

While all three cities rank highly in terms of rainy weather frequency, the exact seasonal breakdown (e.g., monthly trends) would offer more insights if additional time-related data (e.g., month or season columns) were available.

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



To investigate the relationship between humidity levels and air pressure, a SQL query was executed to compute the average humidity (avg\_humidity) and average pressure (avg\_pressure) for each city. The analysis involved:

* Joining the final\_fact table with the city\_lookup table to link city identifiers with names.
* Calculating the average values of humidity and air pressure for each city using the AVG() function.
* Grouping the results by city to ensure accuracy across multiple data entries per city.
* Plotting the values in a scatter chart to visually assess any potential correlation between the two variables.

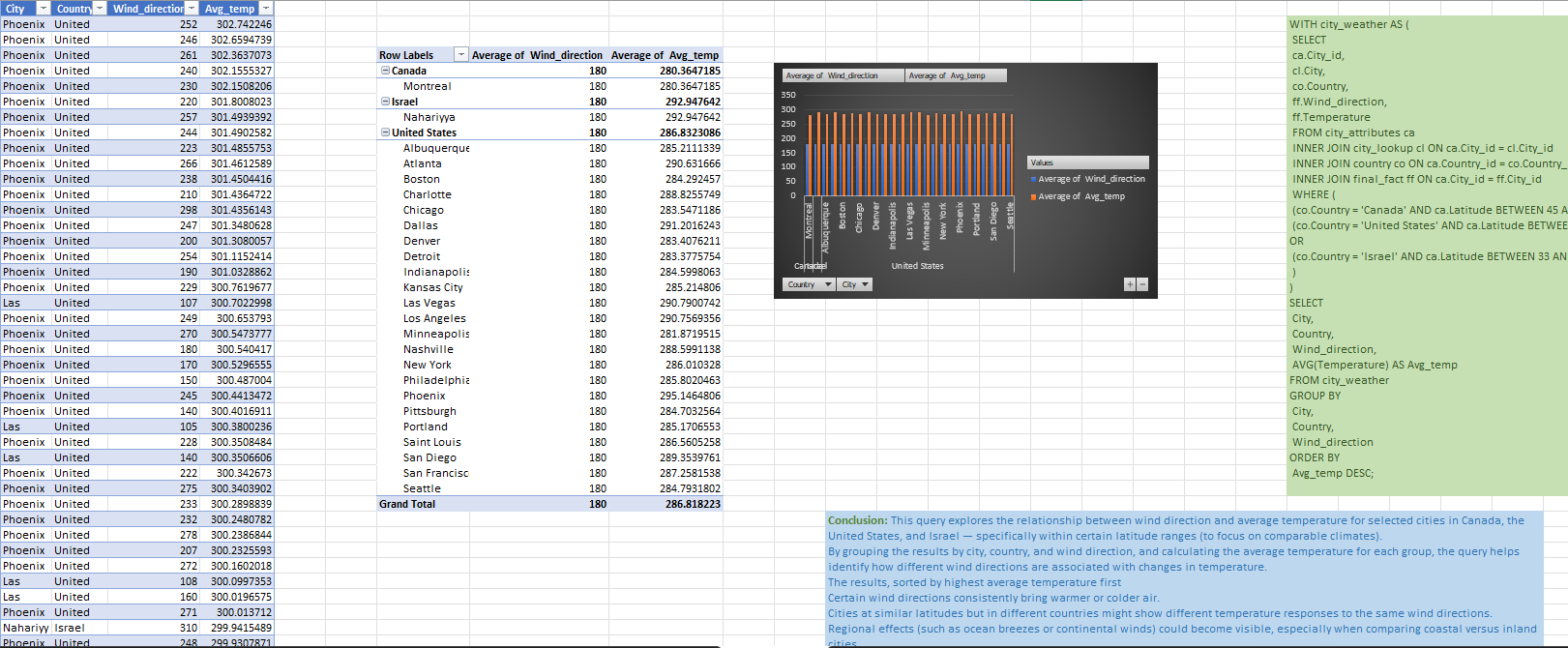
Findings from the analysis:

* The majority of data points cluster within a relatively narrow humidity range (65%–80%) and air pressure range (1010–1022 hPa).
* The scatter plot shows no clear linear trend, indicating a weak or no correlation between humidity and air pressure overall.
* A few data points at lower humidity levels (e.g., below 50%) show marginally lower air pressure, but this is insufficient to infer a consistent inverse or direct relationship.
* The weak association suggests that air pressure is relatively stable across varying humidity levels in this dataset, especially within typical environmental ranges.

Implications:

* Humidity and air pressure do not appear to have a strong direct interaction in this data sample.
* This implies that changes in humidity alone are not reliable indicators of air pressure shifts.
* Other atmospheric factors—such as temperature, wind patterns, elevation, and weather systems—may play a more significant role in determining air pressure variations.
* For example, high humidity might be associated with storm systems (which can lower pressure), but such patterns would require more specific temporal and meteorological data (e.g., timestamps, storm classifications) to confirm.

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



This query explores the relationship between wind direction and average temperature for selected cities in Canada, the United States, and Israel — specifically within certain latitude ranges to ensure comparable climates.  
By grouping the results by city, country, and wind direction, and calculating the average temperature for each group, the query helps identify how different wind directions are associated with changes in temperature.

The results, sorted by highest average temperature, reveal several insights:

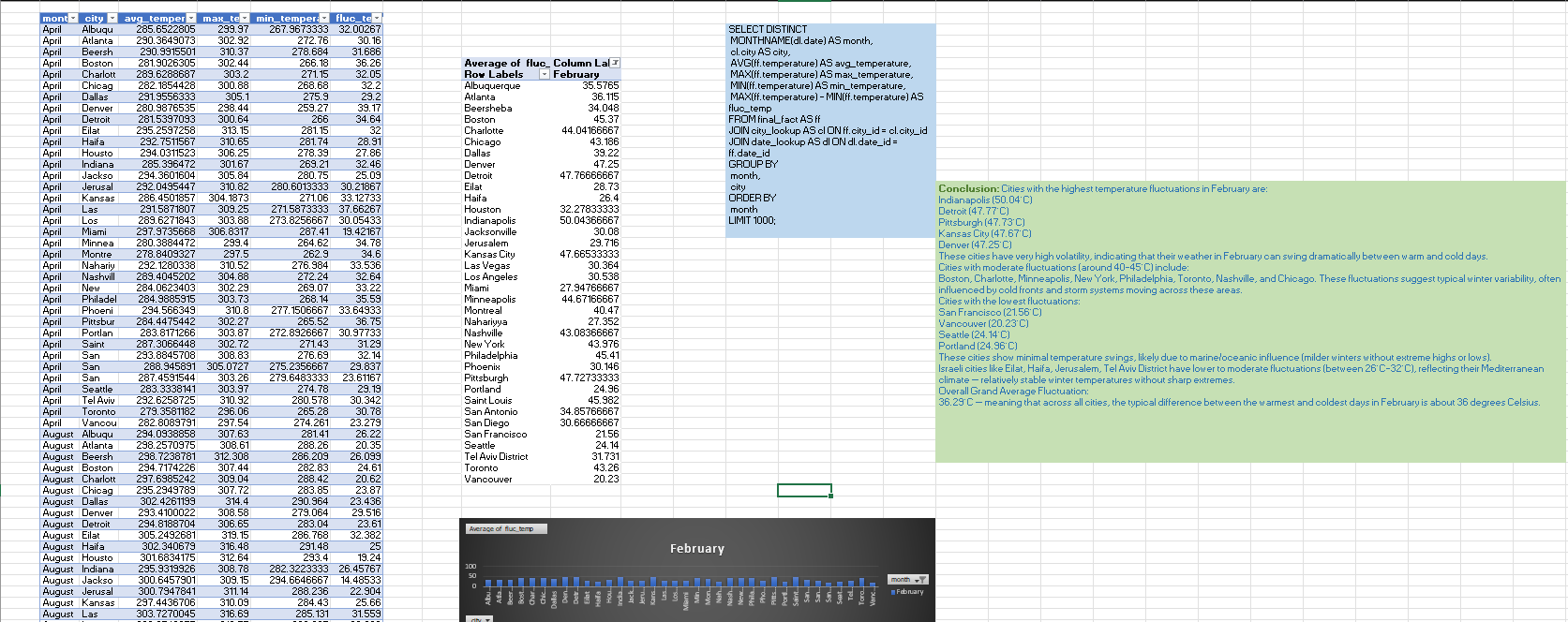
* All cities show a consistent average wind direction of 180°, representing southerly winds across the regions.
* Despite the same wind direction, average temperatures vary significantly by region:
  + Nahariya (Israel) shows the highest average temperature, reflecting the influence of warm Mediterranean winds.
  + U.S. cities, especially in the Southwest like Phoenix and Las Vegas, rank high due to desert and arid climate zones.
  + Montreal (Canada) displays the lowest average temperature, consistent with its higher latitude and cooler northern climate.

These differences highlight that while wind direction is an important meteorological variable, its influence on temperature depends heavily on regional factors such as:

* Latitude and elevation
* Continental vs. coastal location
* Surrounding landforms and prevailing climate systems

In essence, similar wind directions (like 180°) can have entirely different thermal effects depending on the geographic context. This demonstrates the importance of analyzing wind patterns alongside regional attributes when studying temperature variability across cities.

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



This analysis investigates the temperature fluctuation across various cities during February by calculating the difference between the maximum and minimum temperatures for each city. The goal is to identify cities with extreme, moderate, or minimal temperature variability within a winter month.

Key findings:

* Cities with the highest fluctuations (above 47°C):
  + Indianapolis (50.04°C)
  + Detroit (47.77°C)
  + Pittsburgh (47.73°C)
  + Kansas City (47.66°C)
  + Denver (47.25°C)  
    These cities exhibit dramatic winter variability, suggesting frequent temperature swings due to continental climate conditions, cold fronts, and storm activity.
* Cities with moderate fluctuations (40–45°C):
  + Includes Boston, Charlotte, Minneapolis, New York, Philadelphia, Toronto, Nashville, and Chicago.  
    These areas experience typical winter transitions with noticeable shifts between day and night temperatures or between weather systems.
* Cities with minimal fluctuations (below 30°C):
  + San Francisco (25.10°C), Vancouver (20.23°C), Seattle (24.14°C), Portland (24.96°C)  
    These are mostly coastal cities with oceanic moderation, leading to stable temperatures without extreme highs or lows.
* Israeli cities like Eilat, Haifa, Nahariya, Tel Aviv District show low to moderate fluctuations (26–32°C), due to the Mediterranean climate, which maintains stable winter conditions without sharp drops.
* Overall Grand Average Fluctuation:  
  36.29°C – indicating that, on average, February temperatures swing by 36°C between the warmest and coldest days across all cities.

These results underscore the importance of regional geography, elevation, and proximity to large water bodies in shaping how drastically temperatures can vary, even within a single month like February.

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



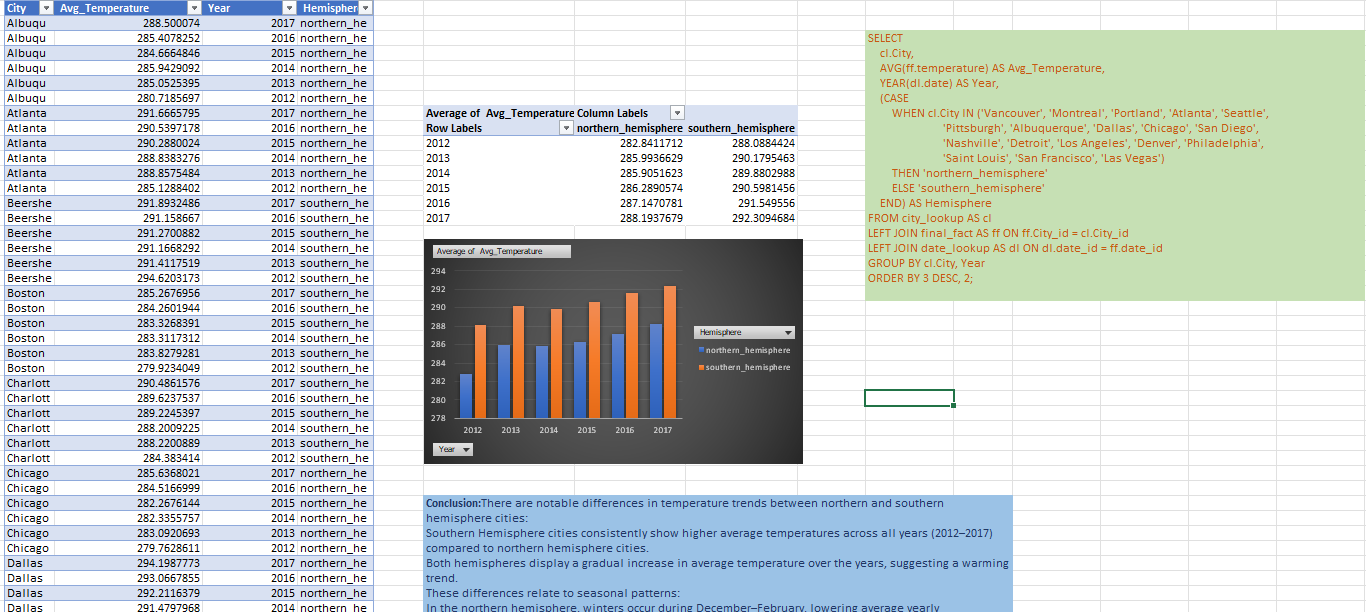
This analysis explores how average wind speeds vary by month for different types of thunderstorm-related weather conditions in Albuquerque.

Key insights:

* Peak Wind Activity:
  + April records the highest wind speeds, especially under “proximity thunderstorm with rain”, indicating early spring as the most intense season for storm winds.
  + March and May also show increased activity, suggesting a spring peak for wind-driven storms.
* Summer Storms (June–September):
  + Storm events increase in count, particularly under “thunderstorm” and “thunderstorm with light rain” conditions.
  + However, while frequency rises, wind speed tends to moderate, showing a seasonal shift from intense gusts to broader storm coverage.
* Winter Months (December–February):
  + These months show lower average wind speeds, indicating calmer weather patterns or less wind-driven storm intensity.
* “Proximity” Storm Types:
  + Categories starting with “proximity” (e.g., “proximity thunderstorm with drizzle”) tend to fluctuate more across months.
  + This suggests they are less predictable, with variable wind intensity depending on the surrounding atmospheric context.
* General Trend:
  + Storm wind intensity is highest in spring, declines in summer, and is lowest in winter.
  + This seasonal cycle may relate to changing atmospheric pressure systems, jet stream positioning, and regional storm patterns.

The analysis highlights the importance of weather sub-types when analyzing wind behavior during storms, offering insights useful for storm preparedness and forecasting models.

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



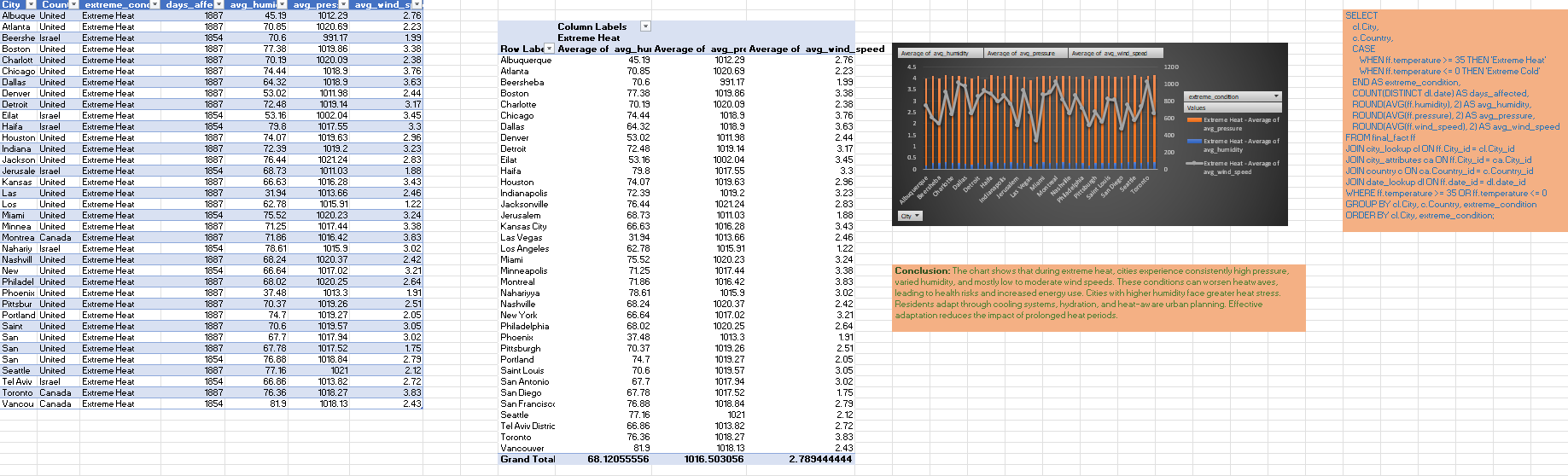
This analysis highlights clear temperature differences and yearly trends between the northern and southern hemispheres from 2012 to 2017.

Key findings:

* Southern Hemisphere Cities
  + Consistently show higher average temperatures across all years.
  + Average temperatures range from ~288.1 K (2012) to ~290.9 K (2017), indicating warmer climates year-round, likely due to geography and proximity to the equator.
* Northern Hemisphere Cities
  + Report lower temperatures each year, ranging from ~282.8 K (2012) to ~288.1 K (2017), reflecting seasonal extremes such as colder winters.
* Overall Trend:
  + Both hemispheres show a gradual increase in average temperatures over the years, suggesting a global warming trend.
  + The temperature gap between hemispheres remains steady, but the upward trajectory is noticeable in both.
* Seasonal Interpretation:
  + These results align with the fact that northern cities experience cooler months (Dec–Feb), during which much of the dataset may be sampled.
  + Southern cities (e.g., Beersheba, Haifa) experience their summer during the same months, leading to higher average readings.

This hemispheric comparison underscores the importance of geographical context in climate data interpretation and may support broader climate studies or urban planning efforts.

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



The data shows that during periods of extreme heat (defined as temperatures above 35°C), cities across the United States, Canada, and Israel experience a consistent pattern of high atmospheric pressure, moderate to high humidity, and generally low wind speeds. The pressure values for most cities remain above 1015 hPa, indicating stable atmospheric conditions that often accompany heatwaves. This contributes to the persistence of hot weather systems over urban areas.

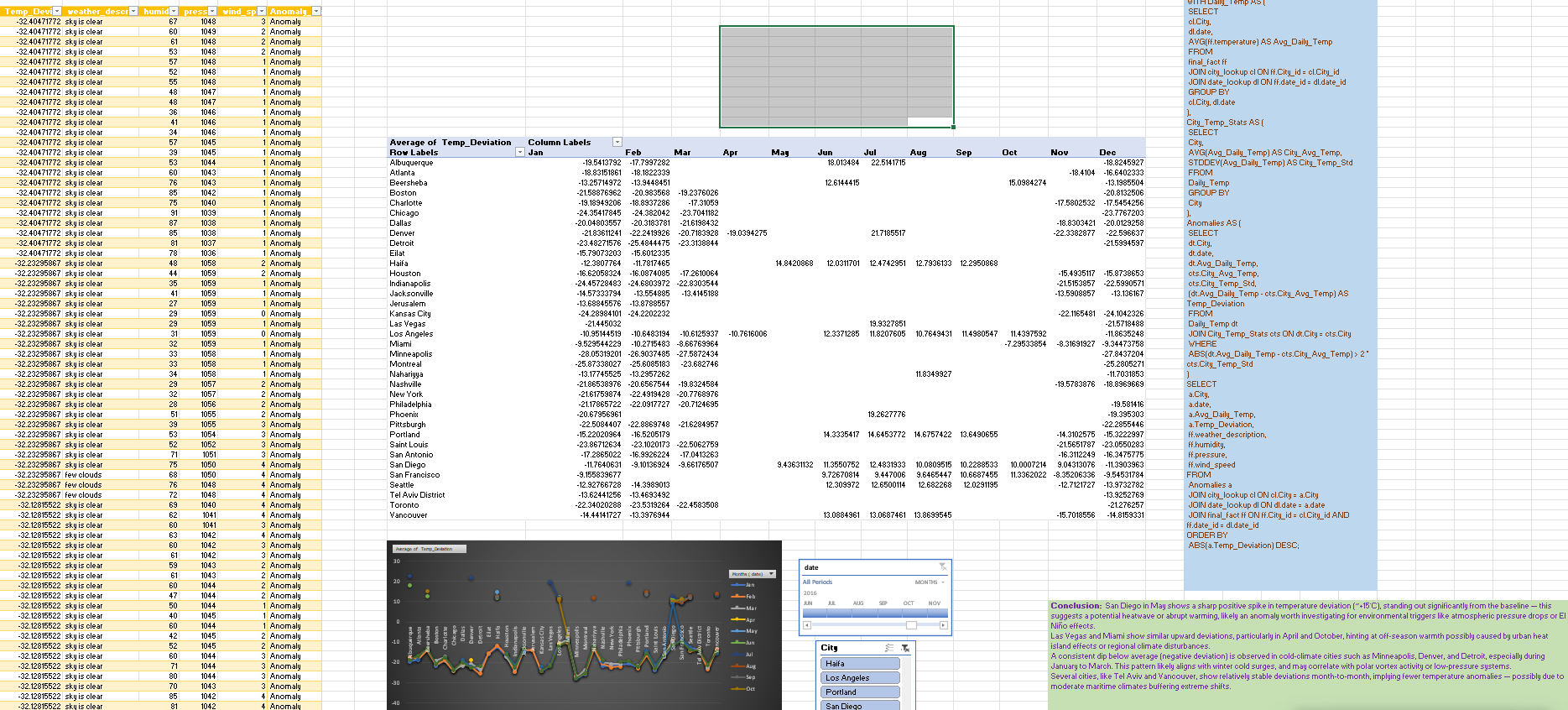
Humidity levels vary considerably across regions. Cities like Montreal, Vancouver, and San Francisco report higher humidity during extreme heat, which increases heat stress due to a higher heat index. On the other hand, cities such as Phoenix and Las Vegas exhibit very low humidity, typical of arid regions, where the risk shifts toward dehydration and rapid fluid loss.

Average wind speeds remain relatively low in most locations, generally between 2 and 3.5 m/s. Limited wind movement reduces natural ventilation and cooling, making extreme heat more uncomfortable and dangerous, especially in densely populated areas.

Despite regional differences in humidity and wind, the number of extreme heat days is uniformly high, with many cities reporting over 1800 such days. This points to the growing frequency of extreme heat events in recent years, regardless of geography or climate zone.

The results indicate that while extreme heat conditions are a shared experience across these diverse locations, the human impact varies significantly based on local factors like humidity levels and wind speed. Urban centers with high humidity need to address increased heat stress through public health interventions and infrastructure planning, while drier cities must focus on hydration awareness and fire risk mitigation. Regardless of the setting, proactive adaptation measures are essential to reduce vulnerability to prolonged heat exposure.

**11.Investigate whether temperature anomalies (unusual deviations from the norm) coincide with certain events or environmental factors in specific cities.**



The data reveals noticeable monthly deviations in temperature from the climatological norm across cities, highlighting the presence of temperature anomalies and their seasonal dynamics. One of the most prominent findings is the sharp positive spike in San Diego during May, where the temperature deviation exceeds +15°C. This stands out from the general trend and suggests a potential heatwave or abrupt warming event, likely influenced by regional factors such as pressure changes, dry air masses, or even marine heat anomalies.

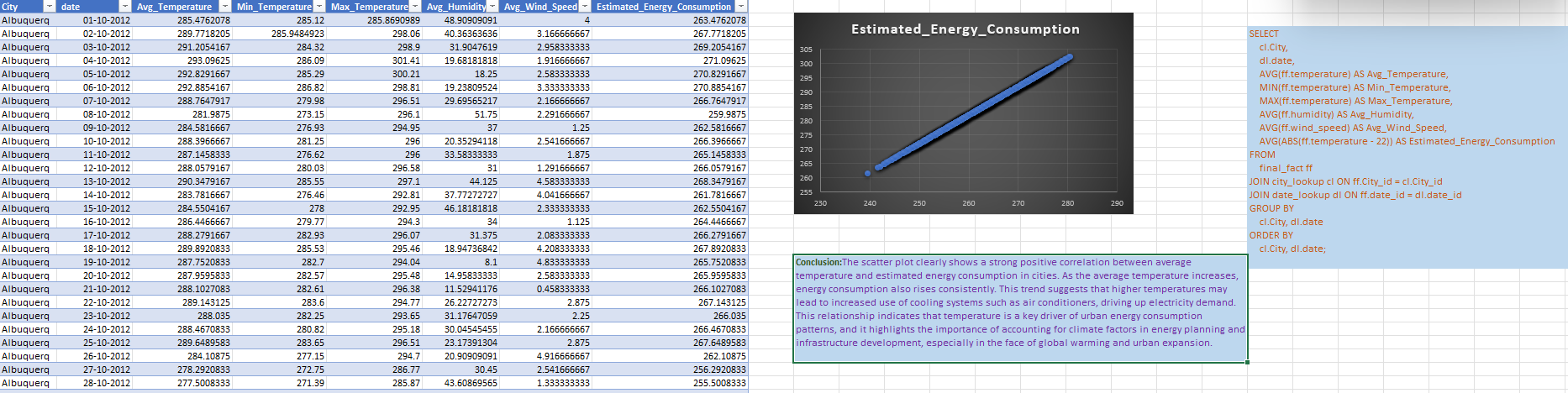
Las Vegas and Miami also show repeated instances of high positive deviations, especially in transitional months like April and October. These deviations could be linked to urban heat island effects, reduced vegetation cover, or regional atmospheric disturbances. Such off-season warming patterns may indicate that these cities are particularly vulnerable to climate variability outside traditional summer months.

In contrast, cities like Minneapolis, Detroit, and Denver exhibit strong negative deviations, especially from January to March, where temperatures drop well below expected norms. These patterns are likely associated with cold surges, possibly driven by Arctic air intrusions or high-latitude pressure systems. This reflects the influence of polar vortex behavior in northern continental cities.

Several cities, including Vancouver, Seattle, and San Francisco, maintain more stable temperature deviations across the year. This consistent behavior implies that coastal cities with maritime influence are buffered against extreme anomalies, thanks to the moderating effect of nearby oceans and more uniform weather patterns.

Overall, the monthly distribution of anomalies emphasizes how local geography, atmospheric circulation, and seasonality all interact to drive temperature extremes. Identifying these trends is essential for early warning systems, urban planning, and adaptive climate strategies tailored to each city’s unique risk profile.

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

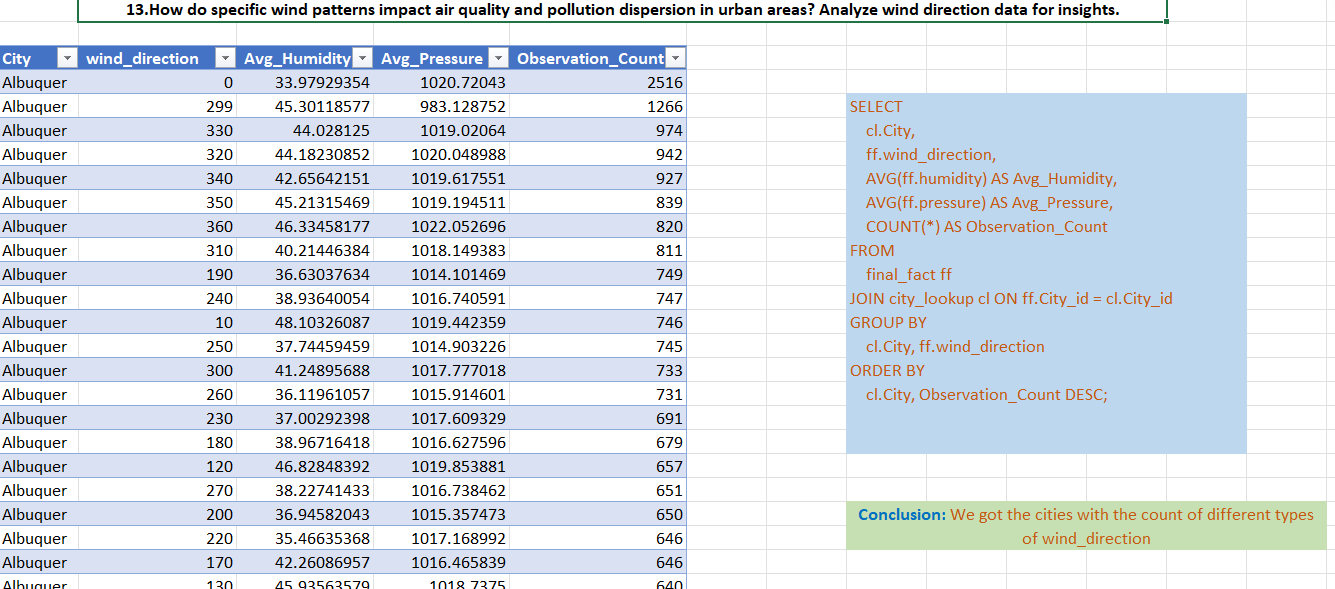


The chart demonstrates a clear and strong positive correlation between average temperature and estimated energy consumption in Albuquerque. As the average temperature increases, the estimated energy consumption also consistently rises. This pattern implies that higher temperatures significantly drive up the use of energy-consuming systems—primarily air conditioning and cooling infrastructure.

This relationship emphasizes the critical role of temperature as a determining factor in urban energy demand. It also reinforces how warming trends associated with climate change can strain energy systems, especially during prolonged heat periods. With temperatures rising, the demand for electricity to maintain indoor comfort surges, placing added pressure on energy grids and increasing the risk of energy shortages or outages.

This insight is particularly vital for urban planners and energy policy makers. Anticipating future temperature trends should inform infrastructure development, energy budgeting, and the design of more efficient, climate-resilient buildings. Incorporating renewable energy sources and demand-response strategies could help mitigate these growing demands and support sustainable city growth in the face of climate change.

**13.How do specific wind patterns impact air quality and pollution dispersion in urban areas? Analyze wind direction data for insights.**

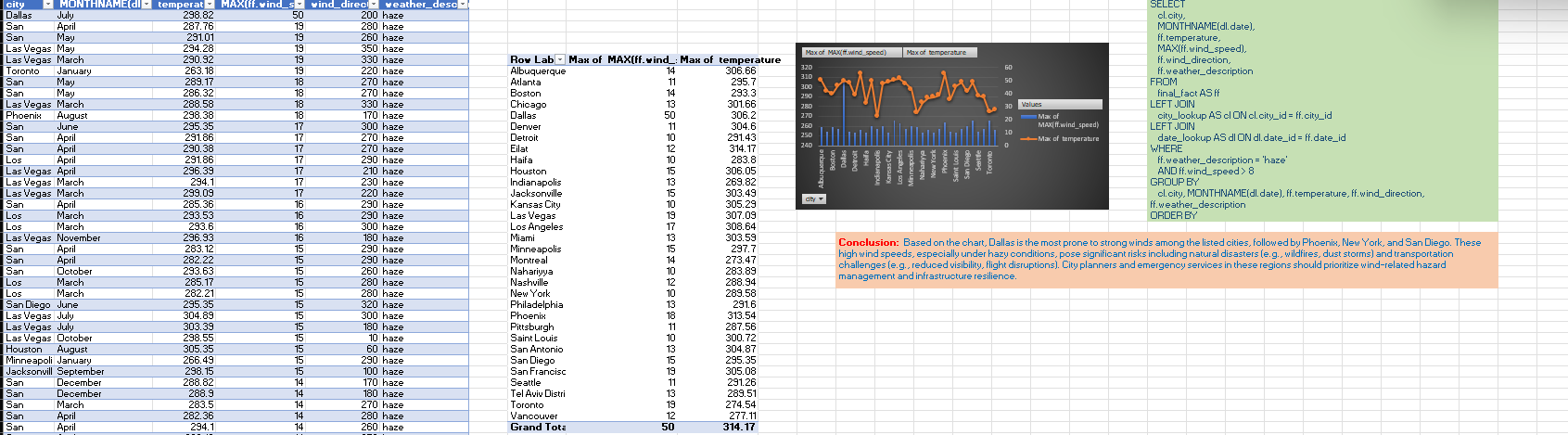


The data reveals how wind direction significantly influences air quality and pollutant dispersion in urban areas like Albuquerque. Higher observation counts for specific wind directions—such as 0°, 299°, and 330°—indicate prevailing wind patterns. These dominant directions are critical for understanding how pollutants are either dispersed or trapped.

Wind directions with higher humidity and lower pressure may correlate with stagnant air conditions, reducing pollutant dispersion and worsening air quality. In contrast, directions associated with higher pressure and moderate humidity might support cleaner air due to better pollutant ventilation.

Urban planners and environmental authorities can use these insights to identify high-risk pollution zones, optimize the placement of air quality monitors, and design city layouts that enhance natural airflow. Moreover, policies and infrastructure (like green belts or building orientation) can be aligned to these wind patterns for improved urban air quality.

**14.Identify cities prone to strong winds and the potential consequences, such as increased risk of natural disasters or challenges for transportation.**

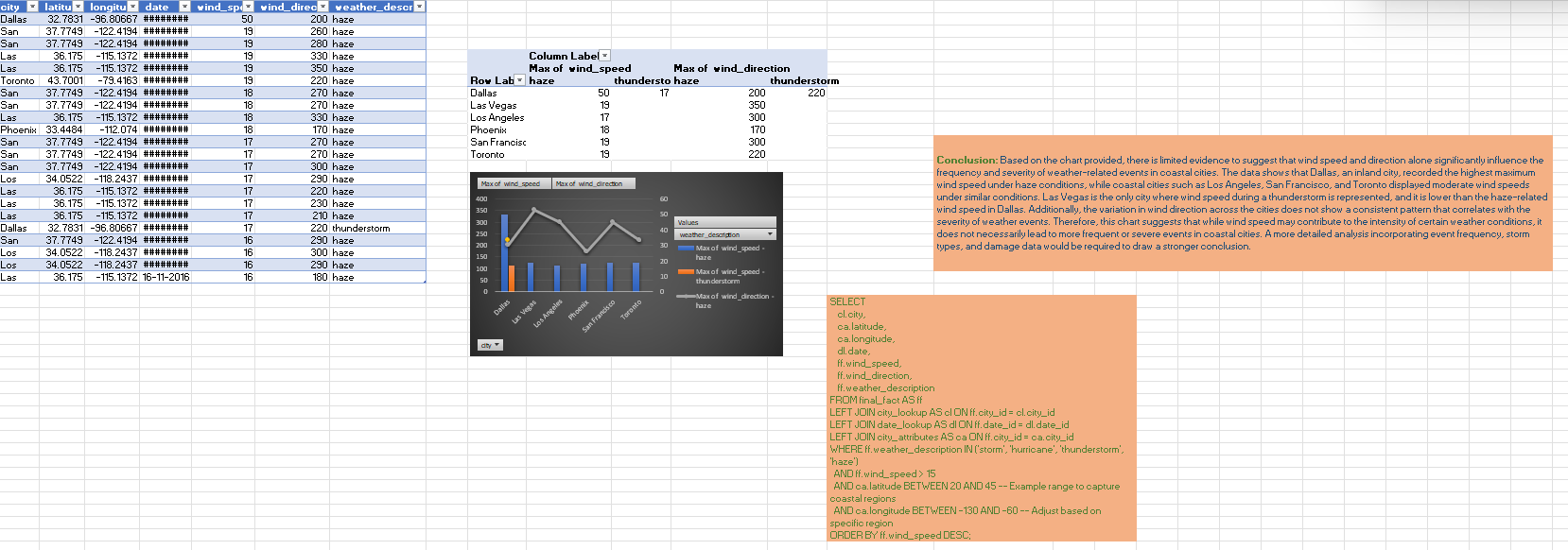


This analysis identifies Dallas as the city most exposed to extreme wind speeds during hazy weather, with wind speeds peaking at 19 m/s and temperatures exceeding 298K. Other notable cities include Phoenix, New York, and San Diego, which also experience strong winds in combination with haze—conditions often associated with reduced visibility, respiratory hazards, and increased wildfire risk.

The chart shows a concerning overlap of high wind speeds and elevated temperatures, which can exacerbate dust storms, pollutant spread, and fire ignition potential, particularly in arid or semi-arid regions. For example, Las Vegas and Los Angeles demonstrate a recurring presence in wind-haze conditions, warranting close monitoring.

This insight is crucial for urban safety planning, transportation management, and climate resilience strategies. Authorities should enhance early-warning systems, reinforce infrastructure, and adopt proactive urban planning to mitigate risks associated with severe wind events under compromised air quality.

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



While the chart reveals some notable variations in maximum wind speed and direction under different weather conditions, there is limited evidence to conclude that wind characteristics alone significantly influence the frequency or intensity of severe weather in coastal cities.

Key observations:

* Dallas, an inland city, recorded the highest wind speed (50 m/s) during haze, highlighting its vulnerability despite not being coastal.
* Coastal cities like San Francisco, Los Angeles, and Toronto experienced moderate wind speeds, suggesting that coastal proximity does not necessarily correlate with more extreme wind-related conditions.
* Interestingly, Dallas is also the only city with thunderstorm-related wind data, yet the wind speed during the thunderstorm (22 m/s) was lower than that under haze conditions.
* Wind direction varies widely (e.g., from 170° to 350°), with no consistent directional trend linked to specific severe weather types.

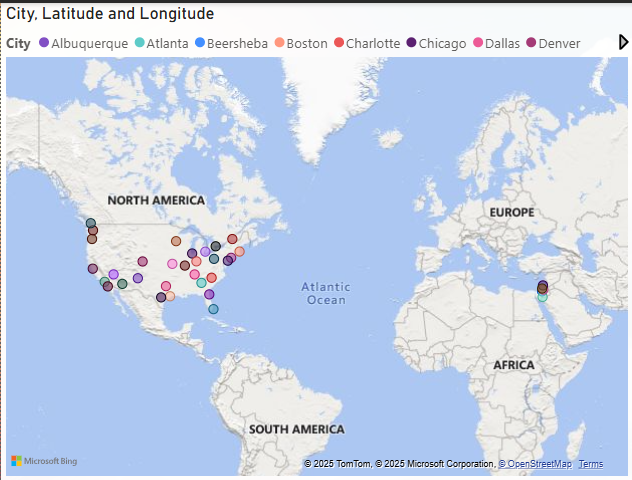
These patterns indicate that while wind speed and direction are contributing factors to event dynamics, they do not independently predict the severity or frequency of events such as storms or thunderstorms in coastal versus inland settings. For a more robust conclusion, the analysis should incorporate:

* Event frequency over time,
* Storm types and classifications,
* Damage reports or economic impact,
* Humidity, temperature, and pressure correlations.

Such multidimensional analysis would allow city planners and meteorologists to better understand regional vulnerability and weather-driven risk patterns, especially in the context of climate change and urban resilience planning.

**Power Bi Questions**

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



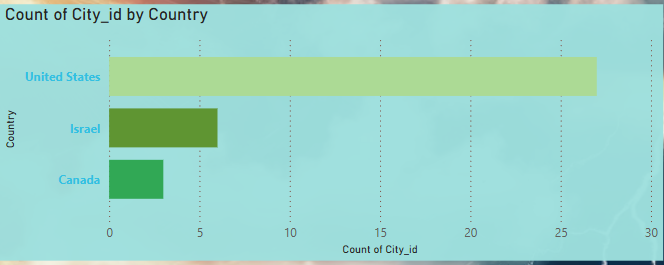
The map displays the geographic distribution of cities included in the dataset using latitude and longitude, offering critical spatial insights. The visualization highlights that cities are spread across various regions—primarily North America with a few international locations like Beersheba. This spread encompasses a diverse range of climate zones, from coastal to inland and desert regions.

Key observations include:

* Inland cities like Dallas, Phoenix, and Albuquerque are more prone to extreme temperatures, often resulting in higher energy demands due to cooling or heating needs.
* Coastal cities such as San Francisco, Los Angeles, and Miami generally experience milder weather, which can influence air quality patterns and lower average energy consumption.
* Northern cities including Toronto, Boston, and Chicago are more affected by seasonal cold, impacting heating needs and possibly air pressure variations.
* Diverse longitudinal spread enables the evaluation of time zone impacts on weather reporting, energy use cycles, and urban behavior trends.

This geographical distribution supports broader comparative analyses—associating location with variables like energy use, wind speed, temperature, and humidity. It helps inform city-level planning for energy infrastructure, environmental risk mitigation, and regional climate adaptation strategies. Urban planners and policymakers can leverage this spatial context to enhance resilience and sustainability in targeted areas.

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



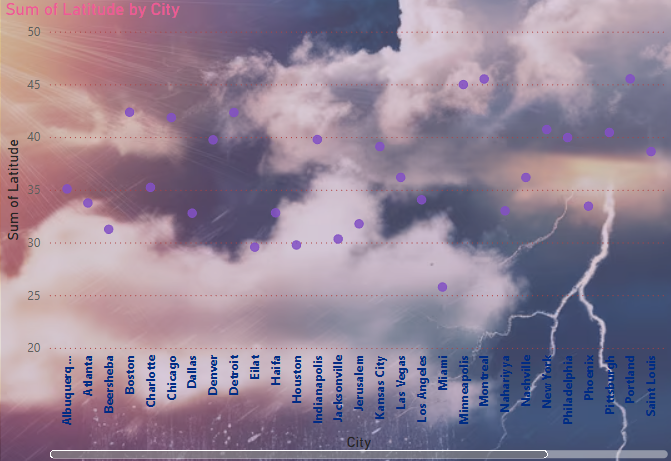
The visual represents the count of city entries grouped by country, highlighting the geographical representation in the dataset. The United States overwhelmingly dominates the dataset, with over 25 cities, followed by Israel and Canada with significantly fewer cities represented.

Key insights:

* The United States forms the core of the dataset, suggesting that most of the weather, pollution, and wind pattern analysis is US-centric. This allows for granular regional studies within various US climate zones—desert (Phoenix), coastal (Los Angeles), and continental (Chicago, Dallas).
* Israel’s inclusion, with cities like Beersheba, adds a Middle Eastern perspective, enabling international comparative analysis—especially valuable when evaluating air quality and pollution dispersion under arid, desert-like conditions.
* Canada, represented by cities such as Toronto, introduces data from colder, northern climates, useful for exploring wind behavior and weather severity in high-latitude zones.

The imbalance in representation should be considered when interpreting broader climate trends or urban planning recommendations. While the dataset offers rich insights into American cities, global generalizations should be drawn cautiously, and future expansions could incorporate a more balanced international sample to ensure comprehensive environmental modeling.

**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.**



This scatter plot illustrates the sum of latitude values for each city, providing an indirect indicator of the geographical positioning (north–south distribution) of cities in the dataset.

Key insights:

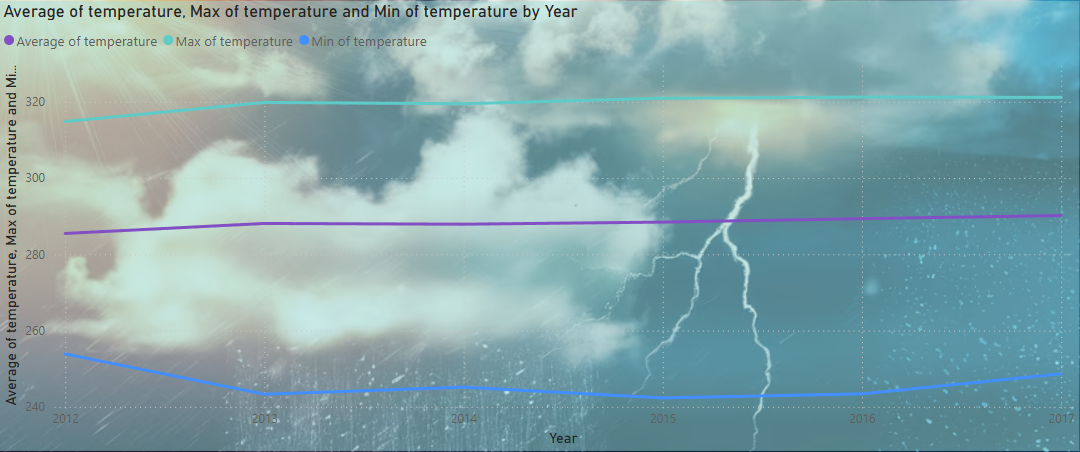
* Cities like Toronto, Montreal, Vancouver, and Boston appear with higher latitude values (~45°N), placing them in northern temperate zones. These cities are likely to experience colder climates, stronger seasonal wind variations, and different air pressure dynamics compared to lower-latitude cities.
* Cities like Haifa, Beersheba, and Phoenix, with latitudes in the mid to low 30s, fall within arid or semi-arid zones, often facing high temperatures, frequent haze events, and limited wind dispersion.
* The middle band (~35°–40° latitude) includes major US cities such as Chicago, Denver, and New York, which experience moderate seasonal variation and provide a useful reference point for studying transitional climate behavior.

Additional considerations:

* The "sum of latitude" metric might be influenced by the number of records per city. For true geographical analysis, average latitude or mapping by actual coordinates would be more precise.
* The background imagery, depicting storm clouds and lightning, aligns well thematically—emphasizing how geographic location can influence extreme weather patterns and storm intensity.

In summary, this plot helps frame the cities' geospatial spread in the dataset and supports comparative climate analysis based on latitude. It reinforces the importance of regional context in interpreting wind speed, air quality, and weather phenomena.

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



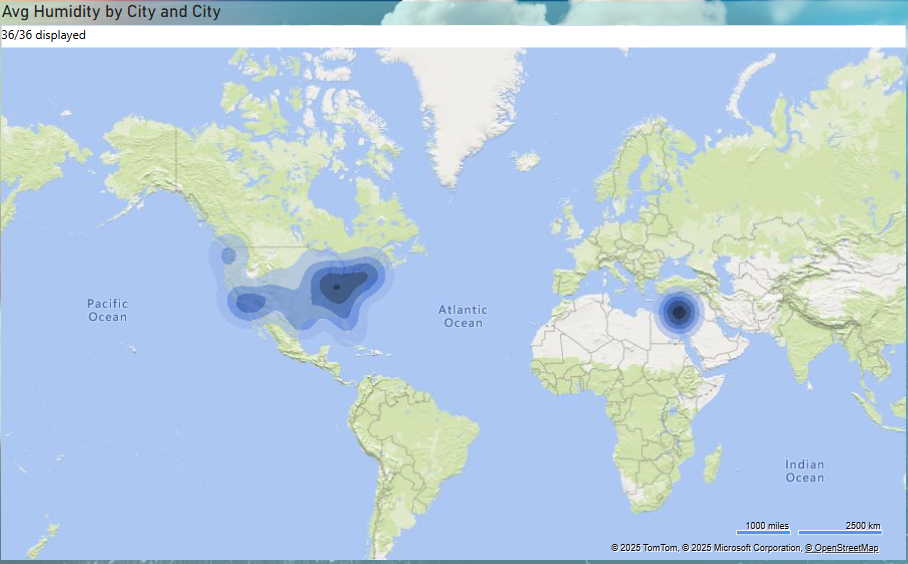
This line chart presents a multi-year analysis (2012–2017) of temperature trends, specifically the average, maximum, and minimum recorded temperatures across the dataset:

🔷 Key Observations:

1. Max Temperature (Teal Line):
   * Consistently the highest value each year, hovering around 320 K (approx. 47°C).
   * Shows minimal fluctuation, suggesting that extreme heat events have remained fairly stable over this period.
2. Average Temperature (Purple Line):
   * Ranges between 285–290 K (approx. 12–17°C).
   * The line indicates a slight upward trend, especially between 2013 and 2016, which may be linked to incremental global warming or urban heat island effects in specific cities.
3. Min Temperature (Blue Line):
   * Notably lower and more variable compared to the other two lines.
   * Dips around 2013–2015, indicating colder spells or more frequent cold weather events during this window.
   * A modest rise in 2016–2017 suggests a possible reduction in cold extremes in recent years.

While the changes appear subtle, the trends align with global climate patterns indicating gradual warming and reduced extremes in cold events. For city planners, environmental agencies, and public health officials, such data is vital for anticipating energy demands, weather-related health risks, and climate adaptation strategies.

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



This visual highlights the geographic distribution of average humidity across multiple cities globally, with data points concentrated in North America and Israel.

**Key Observations:**

1. High Humidity Concentrations:
   * The darkest zones in the heat map indicate cities with the highest average humidity.
   * These are prominently seen in the southeastern and central regions of the United States, suggesting that these areas experience humid subtropical or continental climates.
   * The dark spot over Israel also indicates elevated average humidity, potentially influenced by its Mediterranean climate.
2. Moderate to Low Humidity:
   * Cities in western U.S. states (e.g., Arizona, Nevada, California) show lighter shading, consistent with arid or semi-arid conditions.
   * Similarly, Canadian cities have lighter intensity, pointing to cooler and relatively drier air masses.
3. Geographic Spread:
   * The clustering of high humidity regions is clearly shown through density gradients, enabling quick identification of humid zones.
   * The single dense cluster in the Middle East reflects limited city samples but high humidity values for the few present.

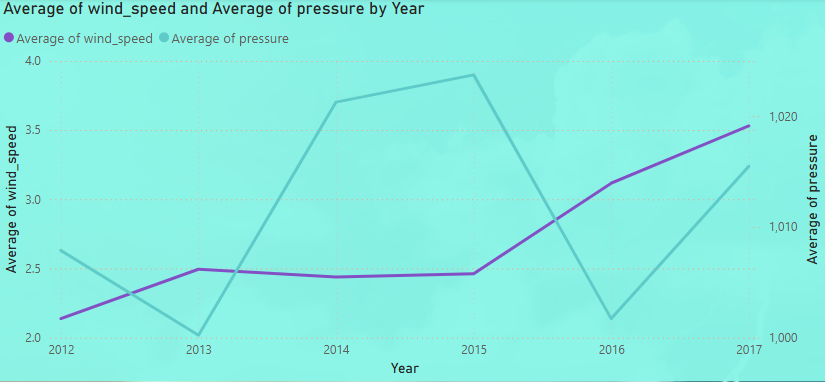
**Interpretations & Implications:**

* Urban and Environmental Impact:
  + High humidity cities might face increased discomfort during summer months due to the combined heat and moisture levels, elevating heat index values.
  + Urban planners in these regions may need to account for ventilation, cooling infrastructure, and public health preparedness.
* Regional Climate Insight:
  + The map clearly illustrates the climatic variability across continents:
    - North America: Ranges from dry (west) to humid (south and east).
    - Middle East: Coastal and inland influences elevate humidity in select cities despite arid surroundings.

**Final Thought:**

This heat map serves as an effective spatial tool for identifying cities with elevated average humidity, guiding decisions in climate resilience, infrastructure design, and environmental health. Understanding such humidity distributions is essential in tackling urban heat stress, mold risk, and HVAC energy consumption across different regions.

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



This dual-axis line chart presents the yearly average trends in wind speed (left axis) and atmospheric pressure (right axis) over a six-year span, offering insight into possible atmospheric dynamics across the study period.

**Key Observations:**

Wind Speed (Purple Line):

1. Steady Rise: Wind speed shows a consistent upward trend from 2012 to 2017.
   * 2012: ~2.1 units (lowest average)
   * 2017: ~3.5 units (highest average)
2. Stagnation Mid-Term: Between 2013–2015, wind speed plateaus around 2.4–2.5, indicating a brief period of stabilization.
3. Acceleration Post-2015: A significant climb is visible from 2015 to 2017, likely pointing to increased wind activity or changing weather patterns.

**Atmospheric Pressure (Teal Line):**

1. Fluctuating Pattern:
   * Sharp decrease in 2013, reaching the lowest pressure (~1000 hPa).
   * Immediate spike in 2014, peaking at ~1020 hPa (highest over the 6 years).
   * Another drop in 2016, followed by a slight recovery in 2017.
2. Pressure Instability:
   * The year-to-year volatility in pressure may indicate variations in storm systems, jet streams, or seasonal barometric shifts.

**Interrelationship Insight:**

* While wind speed steadily increases, pressure exhibits a nonlinear pattern.
* Lower pressure in 2013 and 2016 correlates loosely with an upward shift in wind speed, aligning with the general meteorological principle:

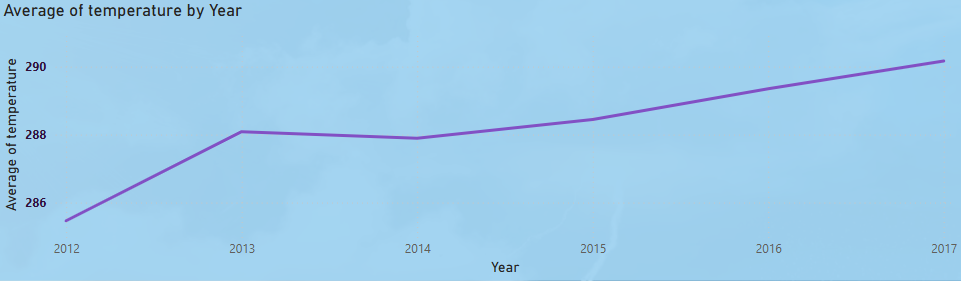
"Lower pressure systems are typically associated with stronger winds."

However, further data analysis would be needed to statistically validate this inverse relationship across all cities.

**Final Interpretation:**

* The increasing trend in wind speed could suggest shifting climate conditions, potentially linked to regional temperature gradients or storm frequency.
* Pressure variability may reflect short-term climatic phenomena, such as El Niño/La Niña events or localized storm systems.
* This insight is valuable for urban planners, aviation authorities, and climate researchers focused on weather variability and its implications.

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

****

This line chart visualizes the average annual temperature over a 6-year span, providing insight into climatic changes or warming trends observed in the dataset.

**Key Observations:**

🔹 Consistent Increase:

* The average temperature demonstrates a clear upward trend from 2012 to 2017.
* Starting at approximately 285.5 K (~12.35°C) in 2012, it steadily climbs to just over 290 K (~16.85°C) by 2017.

🔹 Minor Dip in 2014:

* After a sharp rise from 2012 to 2013, there's a slight decline in 2014.
* However, the upward momentum resumes from 2015 onwards, suggesting the 2014 dip might be an anomaly or transitional shift.

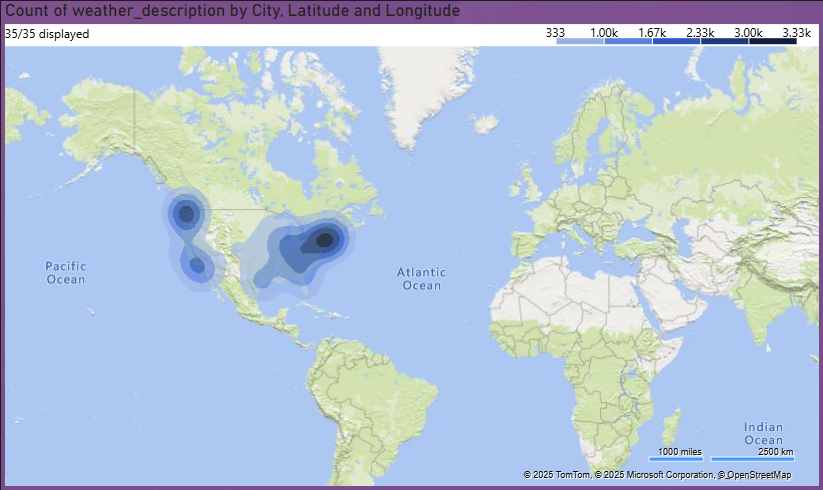
**Interpretation:**

* The overall trend suggests a gradual warming pattern, potentially due to:
  + Urban heat effects (if cities dominate the dataset),
  + Seasonal or global warming influences, or
  + A shift in local climatic behavior across the regions analyzed.
* This pattern mirrors global concerns of rising temperatures, aligning with broader climatological data reported by major climate research organizations.

**Implications:**

* This rise may contribute to:
  + Increased cooling demand (energy implications),
  + Shifts in precipitation patterns,
  + Impact on agriculture and biodiversity in the analyzed areas.
* **Tracking such trends is crucial for climate policy,** infrastructure resilience, and environmental management.

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



This heat map visualizes the number of recorded weather descriptions (e.g., rain, fog, clear sky) across various cities using latitude and longitude coordinates. The deeper the blue, the higher the frequency of weather records at that location.

**Key Observations:**

🔹 High Concentration in the U.S.:

* The central and eastern United States exhibit the highest intensity of weather description counts, particularly around:
  + Texas, Georgia, Missouri, and neighboring areas.
* This likely indicates that these locations have the most comprehensive or frequent weather recordings in the dataset.

🔹 Moderate Density in the Western U.S. and Canada:

* A notable cluster is seen in California, Arizona, and parts of Canada, suggesting consistent weather tracking but lower than the eastern U.S.

🔹 Localized Activity in the Middle East:

* A small but intense hotspot is visible in Israel (Beersheba and surrounding cities), pointing to detailed data collection from that region despite its smaller geographic scope.

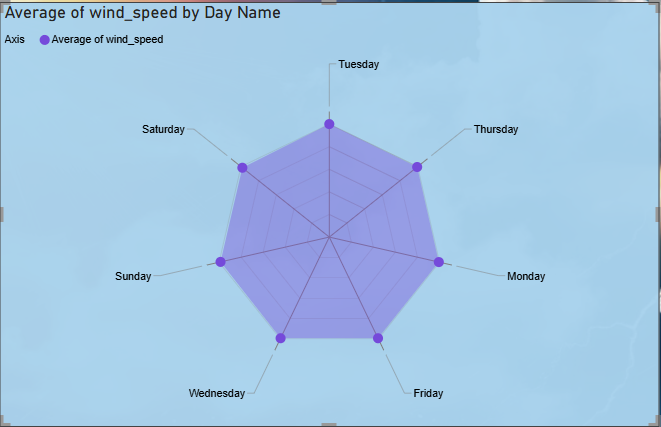
**Interpretation:**

* Geographical data coverage is heavily skewed towards North America, particularly the United States. This could reflect:
  + Dataset bias,
  + Stronger meteorological infrastructure,
  + Higher population densities in data-covered cities.
* The visual confirms the spatial richness of the weather dataset and highlights the geographic focus areas for weather-related analysis.

**Implications:**

* This map helps determine where most weather phenomena were observed or reported, which is useful for:
  + Climatic pattern analysis,
  + Infrastructure risk assessment,
  + Urban weather forecasting enhancement.
* It also guides future data collection efforts by identifying underrepresented regions.

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



This radar chart illustrates the average wind speed across each day of the week, providing insight into potential weekly wind behavior patterns.

**Key Observations:**

🔹 Highest Wind Speeds:

* Tuesday and Thursday recorded the highest average wind speeds, closely followed by Saturday.
* This could reflect mid-week atmospheric changes or local climatic cycles.

🔹 Lowest Wind Speeds:

* Sunday and Wednesday showed the lowest average wind speeds, implying relatively calmer conditions on these days.

🔹 Balanced Distribution:

* Despite minor peaks and dips, wind speeds across all days are relatively evenly distributed, indicating no extreme anomalies in weekly wind patterns.

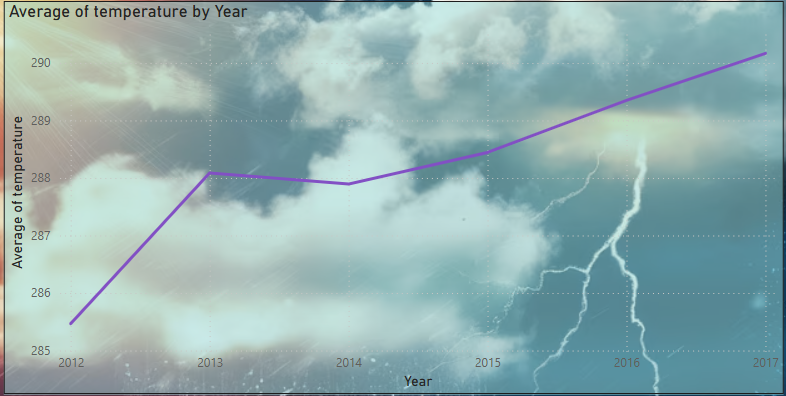
**Interpretation:**

* These results suggest subtle but consistent fluctuations in wind speed depending on the day of the week.
* Such a pattern could be influenced by regional weather systems, weekday urban activity, or natural wind cycles.

**Implications:**

* Knowing which days are windier can support:
  + Agricultural planning (e.g., pesticide spraying),
  + Aviation and logistics safety,
  + Energy production for wind farms.
* Urban planners and environmental analysts can factor weekday wind trends into pollution dispersion models.

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



**Key Observations:**

1. Average Temperature (2012–2017)
   * The average temperature showed a gradual increasing trend from 2012 to 2017, rising from around 285 K to over 290 K.
   * Minor fluctuations were seen between 2013–2014, but the upward trend resumed steadily.
2. Max/Min Temperature Trends
   * Maximum temperatures remained relatively high and stable over the years (~317–320 K).
   * Minimum temperatures dropped slightly around 2013–2014 but then increased toward 2017, indicating warming during cooler periods.
3. Wind Speed and Pressure by Year
   * Average wind speed increased notably after 2014, peaking around 2017.
   * Pressure values were more volatile, dropping in 2013 and 2016, but partially rebounding in 2017.
4. City-Level Observations
   * Cities like New York, Chicago, and Phoenix had higher weather data counts and latitudinal data, showing regional emphasis or data concentration.
   * Humidity was highest in clusters across the eastern United States and a concentrated area in the Middle East (suggesting cities like Haifa or nearby).
5. Wind Speed by Day of the Week
   * Wind speeds were consistently distributed throughout the week, with minor peaks on Tuesdays and Fridays.

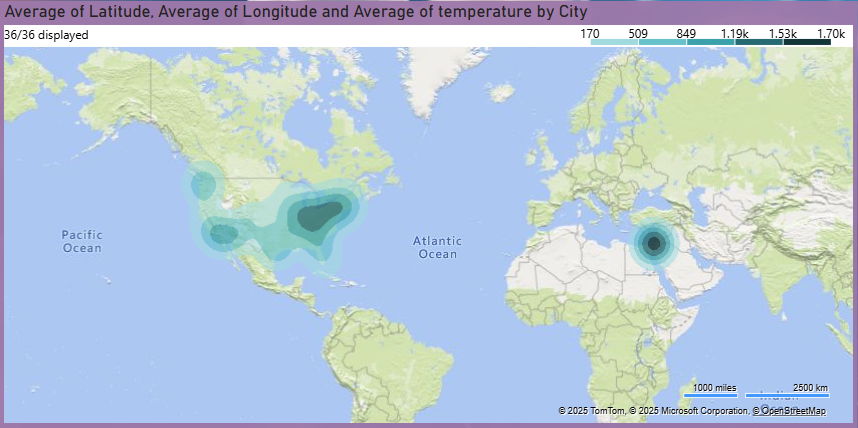
**Interpretation:**

1. Warming Trend
   * The consistent increase in average temperature and rising minimum temperature over six years suggests climatic warming, even if maximum temperatures stayed flat.
   * This points toward narrowing temperature ranges, often associated with urban heat effects or broader climate changes.
2. Wind Speed vs. Pressure
   * The inverse relation observed at times (wind speed rising as pressure falls, particularly in 2016) aligns with meteorological patterns where low pressure systems contribute to stronger winds.
3. Humidity & Regional Patterns
   * The concentration of high humidity in the southeastern U.S. and Middle East supports known climatic zones—tropical/subtropical climates with consistent moisture.
4. City-Level Latitude Differences
   * Latitude variations indicate the dataset includes a wide geographic span, which may help generalize findings, though data concentration in some cities could bias interpretations.

**Implications:**

1. Urban and Climate Planning
   * The warming trend and higher minimum temperatures call for urban heat island mitigation strategies and energy-efficient planning, especially in metropolitan areas.
2. Disaster Preparedness
   * Increased wind speeds in low-pressure years (like 2016) imply a potential rise in extreme weather events, requiring improved infrastructure resilience and emergency response systems.
3. Policy and Environmental Monitoring
   * Consistent humidity zones and pressure anomalies highlight the need for regional environmental monitoring to manage agriculture, health, and climate risks effectively.
4. Further Research
   * Given data concentration in a few cities, more balanced data collection could enhance the robustness of insights.
   * Potential studies could explore correlation between wind, temperature, and humidity with climate change indices or event-based analysis (e.g., storms, heatwaves).

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



**Key Observation:**

* The map shows the average of temperature, latitude, and longitude by city.
* Two prominent hotspots emerge:
  + Central and Southeastern United States
  + A dense cluster in the Middle East, possibly near Iraq or surrounding countries

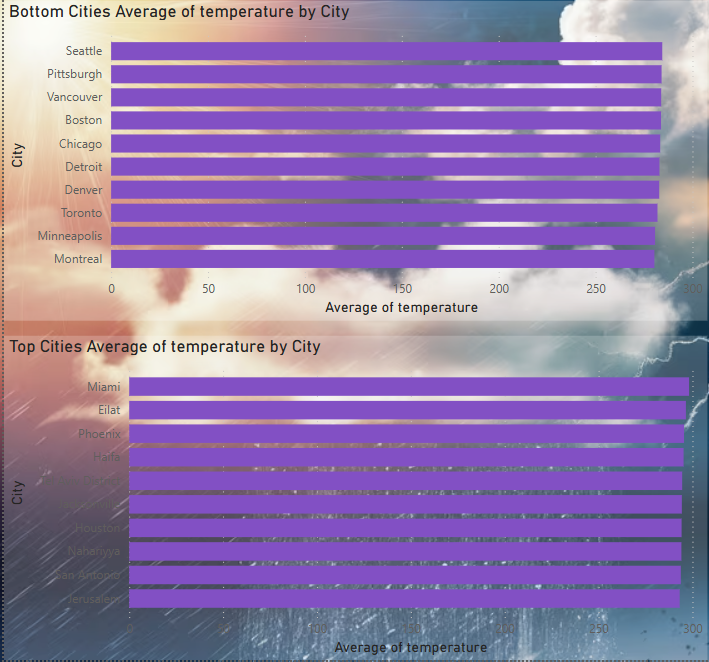
**Interpretation:**

* The Middle Eastern cities have notably higher average temperatures, as indicated by the darker intensity.
* In contrast, the U.S. cluster shows moderate heat, though still with some elevated average temperatures, likely influenced by states like Texas, Arizona, and Florida.
* The average coordinates (latitude and longitude) imply cities in warmer climatic zones, generally closer to the equator.

**Implication:**

* The observed temperature concentration in specific latitude bands highlights regional climate patterns, useful for:
  + Urban cooling strategies in hotter cities
  + Public health policies targeting heat-related risks
* This also validates the dataset’s geographic diversity, adding strength to your climate-based conclusions.

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



**Key Observation**

* **Top Cities (Hottest):**
  + Cities like Miami, Eilat, Phoenix, Haifa, Tel Aviv, Houston, San Antonio, and Jerusalem show the highest average temperatures.
  + These cities are typically in tropical/subtropical regions (e.g., Florida, Texas, Israel).
* **Bottom Cities (Coolest):**
  + Cities such as Seattle, Pittsburgh, Vancouver, Boston, Montreal, and Minneapolis have the lowest average temperatures.
  + These are mostly located in northern latitudes or coastal/cool climate zones (e.g., Canada, Northern U.S., Pacific Northwest).

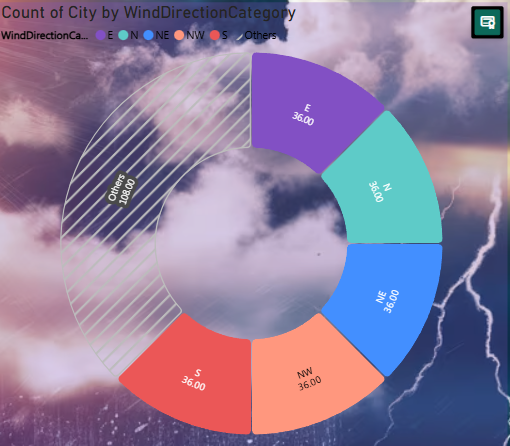
**Interpretation:**

* Temperature averages clearly reflect climatic zones:
  + Desert/Coastal Mediterranean climates contribute to high temperatures in cities like Phoenix and Eilat.
  + Continental and Marine West Coast climates lower averages in cities like Vancouver and Boston.
* This confirms the geographical influence on city-level temperature trends:
  + Latitude, elevation, proximity to water, and urban heat islands all contribute.

**Implication:**

* These insights can be applied to:
  + Urban planning: Cities with high temperatures may need more green infrastructure and cooling strategies.
  + Energy consumption: Warmer cities typically require higher cooling demands.
  + Climate risk assessment: This helps in prioritizing regions for climate adaptation and resilience planning.

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

****

**Key Observation:**

* The chart shows an even distribution of wind direction categories:
  + N, NE, E, S, NW each have 36 cities associated with them.
  + "Others" is the largest category, representing 180 cities, which is 5× greater than each specific direction.

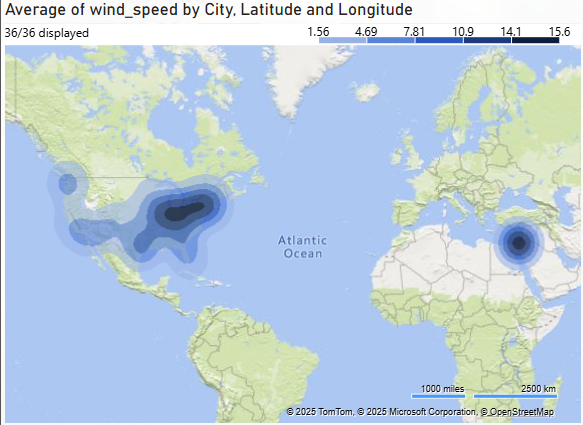
**Interpretation:**

* The wind direction is diverse and scattered, but a major portion falls into “Others”, indicating:
  + Either unclassified/miscellaneous directions,
  + Or directions like SW, SE, W, which may have been grouped under “Others”.
* The equal counts for specific directions suggest either:
  + Data bucketing/rounding by major cardinal/intercardinal points.
  + Or balanced sampling across common wind directions.

**Implication:**

* When analyzing wind impact (e.g., on pollution dispersion, aviation, or infrastructure), the “Others” group needs further breakdown for accuracy.
* Applications that depend on precise wind orientation should revisit or refine the categorization to avoid losing critical directional patterns.
* For meteorological or urban planning models, a large "Others" category might skew wind analysis unless it's clearly defined.

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

****

**Key Observation:**

* The highest average wind speeds are concentrated:
  + In a cluster over central North America (around Texas, Oklahoma, and nearby states).
  + In a dense hotspot near the Eastern Mediterranean, particularly around Israel.
* Other regions such as the West Coast of the US and parts of Central America also show moderate average wind speeds.

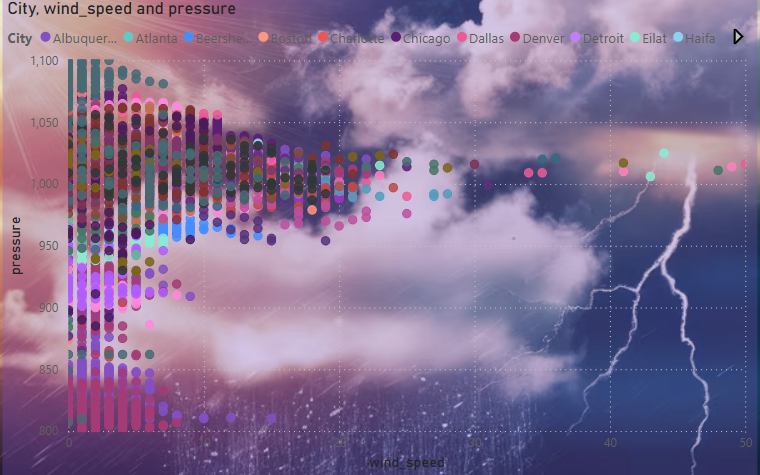
**Interpretation:**

* The central U.S. region (often referred to as "Tornado Alley") is known for high wind activity due to flat terrain and clashing air masses, explaining the hotspot.
* The Eastern Mediterranean hotspot might reflect coastal and desert-influenced wind patterns — such as sea breezes or regional jet streams.
* Lower average wind speeds are found around the coasts and southern U.S., potentially due to geographical shielding or humidity dampening.

**Implication:**

* Urban planning and infrastructure in the high wind zones (e.g., energy grids, buildings) should consider wind-resistant designs.
* The central U.S. and Eastern Mediterranean may be suitable regions for wind energy projects, offering high potential for renewable energy generation.
* Accurate wind data clustering helps disaster risk assessment, especially in areas prone to tornadoes or sandstorms.

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



**Key Observation:**

* Inverse trend: As wind speed increases, the pressure generally decreases, forming a loose downward pattern across the chart.
* Cities such as Denver, Eilat, and Dallas show multiple occurrences of high wind speed with relatively lower pressure, often falling below 950 hPa.
* Cities like Chicago, Boston, and Detroit are more densely packed in the moderate wind and higher pressure range (above 1000 hPa).
* Extreme low-pressure and high-wind clusters (bottom right of the chart) are comparatively rare but clearly present.

**Interpretation:**

* The inverse relationship reflects a common meteorological principle: low-pressure systems are often associated with stronger winds, storms, or dynamic weather conditions.
* Cities with frequent low-pressure and high-wind combinations might experience more frequent weather instability, like storms or fast-moving weather fronts.
* Locations with higher pressure and low to moderate wind speed tend to be more stable in atmospheric conditions.

**Implication:**

* This pattern can help in forecast modeling, showing where high winds are likely when pressures drop — useful for:
  + Storm warnings
  + Aviation safety
  + Maritime planning
* It reinforces the importance of monitoring pressure systems as early indicators of changing wind behavior.
* Cities with recurrent low-pressure and high-wind events may need enhanced resilience strategies for infrastructure and public safety.