Project Report: Weather Insights in Forbes Top 100 Cities

Autor: Maria Jose Teran

1. Introduction

- ***Objective:***

Analyze the weather data of Forbes' Top 100 Cities from 2020 to 2024, and uncover potential relationships between weather conditions and social factors.

- ***Dataset Overview:***

Dataset name: **Forbes Top 100 Cities Weather Data (2020 – YTD)**

Source: Kaggle

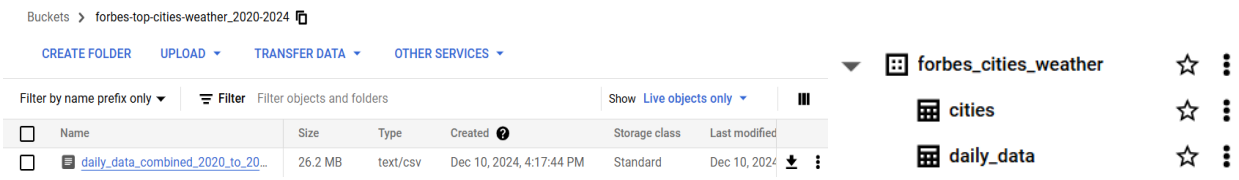
The dataset provides a detailed look at the climate for some of the world's most influential cities. Contains weather attributes (e.g., temperature, precipitation, humidity) for leading global cities.

This dataset tracks the weather of the principal urban centers, which affects Travel and Tourism, Urban Living and Economic Impacts. Analyzing this data unlocks opportunities to create more sustainable and efficient cities, smarter business strategies, and enhance the quality of life for people around the world.

For this project it was chosen the daily temperature records over the hourly records.

- ***Dataset Creation:***

The dataset was downloaded from Kaggle and charged into Google Cloud, using Buckets. Then, importing into the dataset created.

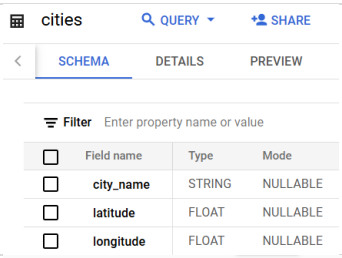


2. Data Exploration

Initial data exploration:

1. Table: cities

Schema



Basic Overview

```
-- 1. explore the datasets

--CITIES
SELECT *
FROM `ds203-finalproject.forbes_cities_weather.cities`
ORDER BY city_name;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUT

Row	city_name	latitude	longitude
1	Abu Dhabi	24.466669	54.366669
2	Amsterdam	52.374031	4.88969
3	Athens	37.97945	23.716221
4	Atlanta	33.749001	-84.387978
5	Auckland	-36.85582	174.743042
6	Austin	34.998421	-91.983757
7	Bangkok	13.87719	100.71991
8	Barcelona	41.38879	2.15899

Total rows

```
SELECT COUNT(DISTINCT city_name) AS total_cities
FROM `ds203-finalproject.forbes_cities_weather.cities`;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON

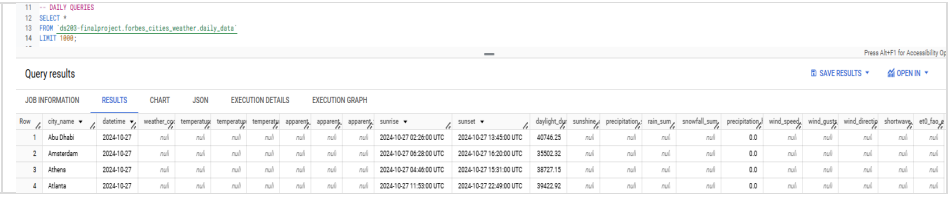
Row	total_cities
1	100

Highlights:

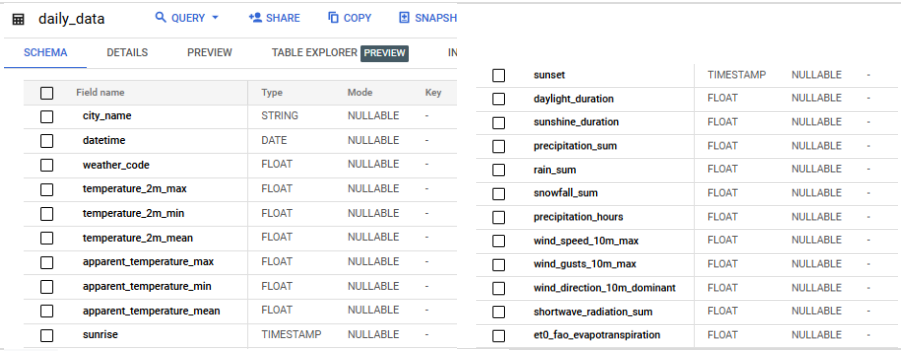
- This table provides information about 100 cities, including: City names, their latitude (north-south position) and longitude (east-west position).
- It helps identify the location of each city on a map.

2. Table: daily_data

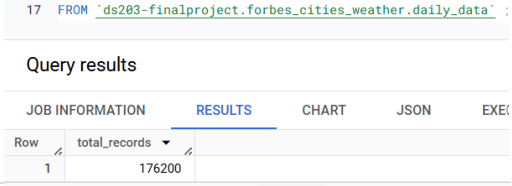
Basic Overview



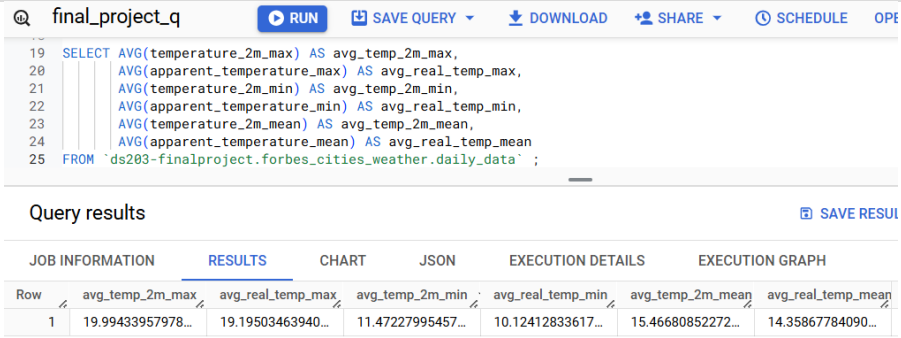
Schema



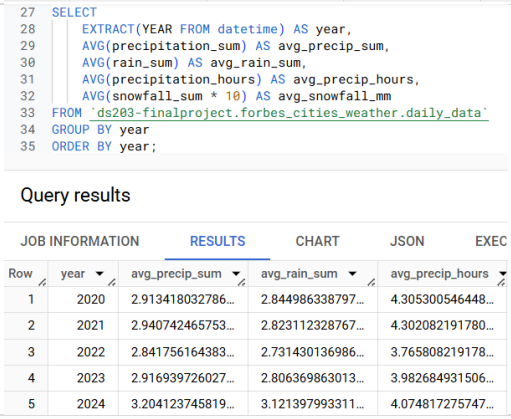
Total rows



Comparing the columns: temperature_2m_*, apparent_temperature_*



Comparing the columns: *precipitation_sum*, *snowfall_sum*, *rain_sum*, *precipitation_hours*



Key columns:

Column	Description
city_name	Name of the city.
datetime	Date of the data record.
weather_code	Code describing the weather condition (e.g., sunny, rainy, cloudy).
temperature_2m_max	Maximum temperature at 2 meters above ground (°C).
temperature_2m_min	Minimum temperature at 2 meters above ground (°C).
temperature_2m_mean	Average temperature at 2 meters above ground.
apparent_temperature_max	Maximum apparent temperature (°C).

apparent_temperature_min	Minimum apparent temperature (°C).
apparent_temperature_mean	Average apparent temperature (°C).
daylight_duration	Duration of daylight in hours.
precipitation_sum	Total precipitation (mm) (including rain, showers, and snowfall).
rain_sum	Total rainfall (mm).
snowfall_sum	Total snowfall (cm)
precipitation_hours	Total hours with precipitation.
wind_speed_10m_max	Maximum wind speed at 10 meters above ground (m/s).
shortwave_radiation_sum	Total shortwave solar radiation (kWh/m²).

Highlights:

- This table provides a lot of information about the weather in each city across the years.
- It contains nulls values, which should be removed.

3. Data Cleaning

Actions taken:

- Search rows without data in the columns “city_name” or “datetime”, without results. So, no entire rows were dropped.

```
SELECT *
FROM
    `ds203-finalproject.forbes_cities_weather.daily_data`
WHERE city_name IS NULL OR datetime IS NULL;
```

ery results

INFORMATION | RESULTS | CHART | JSON | EX

There is no data to display.

- Create a new table, dropping some columns non relevant for the analysis.

Column	Description	Note
sunrise	Exact time of sunrise.	**DROP:** Data contemplated in daylight_duration.
sunset	Exact time of sunset.	**DROP:** Data contemplated in daylight_duration.
sunshine_duration	Duration of sunshine in hours.	**DROP:** Not relevant for analysis.
wind_gusts_10m_max	Maximum wind gust speed at 10 meters above ground (m/s).	**DROP:** The wind gust data is not relevant for the analysis.
wind_direction_10m_dominant	Dominant wind direction at 10 meters (degrees).	**DROP:** The wind direction is not relevant for the analysis.
et0_fao_evapotranspiration	Total evapotranspiration following FAO standards (mm).	**DROP:** Parameter specific to agriculture.

- Count the null values.

```sql
--Count the null values from the new table
SELECT COUNT(*) AS null_value_count
FROM `ds203-finalproject.forbes_cities_weather.daily_data_cleaned`
WHERE
    city_name IS NULL OR
    datetime IS NULL OR
    temperature_max IS NULL OR
    temperature_min IS NULL OR
    temperature_mean IS NULL OR
    temp_feels_like_max IS NULL OR
    temp_feels_like_min IS NULL OR
    temp_feels_like_mean IS NULL OR
    daylight_duration IS NULL OR
    precipitation_sum IS NULL OR
    rain_sum IS NULL OR
    snowfall_sum IS NULL OR
    precipitation_hours IS NULL OR
    wind_speed_10m_max IS NULL OR
    shortwave_radiation_sum IS NULL;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DE

Row	null_value_count
1	200

- Fill the null values with the average value of each column, and replace values in the table cleaned.

```sql
CREATE OR REPLACE TABLE `ds203-finalproject.forbes_cities_weather.daily_data_cleaned` AS
SELECT
    city_name,
    datetime,
    IFNULL(temperature_max, AVG(temperature_max) OVER ()) AS temperature_max,
    IFNULL(temperature_min, AVG(temperature_min) OVER ()) AS temperature_min,
    IFNULL(temperature_mean, AVG(temperature_mean) OVER ()) AS temperature_mean,
    IFNULL(temp_feels_like_max, AVG(temp_feels_like_max) OVER ()) AS temp_feels_like_max,
    IFNULL(temp_feels_like_min, AVG(temp_feels_like_min) OVER ()) AS temp_feels_like_min,
    IFNULL(temp_feels_like_mean, AVG(temp_feels_like_mean) OVER ()) AS temp_feels_like_mean,
    IFNULL(daylight_duration, AVG(daylight_duration) OVER ()) AS daylight_duration,
    IFNULL(precipitation_sum, AVG(precipitation_sum) OVER ()) AS precipitation_sum,
    IFNULL(rain_sum, AVG(rain_sum) OVER ()) AS rain_sum,
    IFNULL(snowfall_sum, AVG(snowfall_sum) OVER ()) AS snowfall_sum,
    IFNULL(precipitation_hours, AVG(precipitation_hours) OVER ()) AS precipitation_hours,
    IFNULL(wind_speed_10m_max, AVG(wind_speed_10m_max) OVER ()) AS wind_speed_10m_max,
    IFNULL(shortwave_radiation_sum, AVG(shortwave_radiation_sum) OVER ()) AS shortwave_radiation_sum
FROM
    `ds203-finalproject.forbes_cities_weather.daily_data_cleaned`;
```

ery results

NFORMATION | RESULTS | EXECUTION DETAILS | EXECUTION GRAPH

This statement replaced the table named daily_data_cleaned.

And finally, the data does not contain any NULL values.

```sql
SELECT COUNT(*) AS null_value_count
FROM `ds203-finalproject.forbes_cities_weather.daily_data_cleaned`
WHERE
    city_name IS NULL OR
    datetime IS NULL OR
    temperature_max IS NULL OR
    temperature_min IS NULL OR
    temperature_mean IS NULL OR
    temp_feels_like_max IS NULL OR
    temp_feels_like_min IS NULL OR
    temp_feels_like_mean IS NULL OR
    daylight_duration IS NULL OR
    precipitation_sum IS NULL OR
    rain_sum IS NULL OR
    snowfall_sum IS NULL OR
    precipitation_hours IS NULL OR
    wind_speed_10m_max IS NULL OR
    shortwave_radiation_sum IS NULL;
```

ery results

INFORMATION | RESULTS | CHART | JSON | EXECUTION

null_value_count
0

- Find inconsistent values:

```
SELECT *
FROM `ds203-finalproject.forbes_cities_weather.daily_data_cleaned`
WHERE
    (temperature_max < temperature_min OR
    temp_feels_like_max < temp_feels_like_min );
```

ery results

INFORMATION | RESULTS | CHART | JSON | EXECUTION I

There is no data to display.

- Find duplicate values:

```
SELECT
    city_name,
    datetime,
    COUNT(*) AS count
FROM `ds203-finalproject.forbes_cities_weather.daily_data_cleaned`
GROUP BY city_name, datetime
HAVING COUNT(*) > 1;
```

ery results

INFORMATION | RESULTS | CHART | JSON | EXECUTION

There is no data to display.

4. Data Analysis

Research Questions:

Q1 - Which cities are most suitable for solar and wind energy projects across geographic regions?

- Purpose: Identify cities with high solar and wind speed potential for energy projects.
- Parameters Used: City name, Radiation sum, Wind speed, geographical location.

Findings:

TOP 10 cities with Solar potential for Energy projects (across all the data).

```
SELECT
    d.city_name,
    c.longitude,
    c.latitude,
    ROUND(AVG(d.shortwave_radiation_sum),2) AS avg_shortwave_radiation_sum,
    ROUND(AVG(d.wind_speed_10m_max),2) AS avg_wind_speed
FROM `ds203-finalproject.forbes_cities_weather.daily_data_cleaned` AS d
INNER JOIN `ds203-finalproject.forbes_cities_weather.cities` AS c
ON c.city_name = d.city_name
GROUP BY d.city_name, c.longitude, c.latitude
ORDER BY avg_shortwave_radiation_sum DESC
LIMIT 10;
```

Row	city_name	longitude	latitude	avg_shortwave_radia	avg_wind_speed
1	Mexico City	-99.127663	19.428471	22.32	19.34
2	Riyadh	46.721851	24.687731	21.17	22.32
3	Muscat	58.592201	23.613871	21.12	19.01
4	Abu Dhabi	54.366669	24.466669	21.05	23.75
5	Dubai	55.304722	25.258169	21.0	23.37
6	Doha	51.533333	25.286667	20.92	24.18
7	Las Vegas	-105.2239	35.593929	20.33	23.76
8	San Jose	-121.894958	37.33939	19.71	19.9
9	Bogota	-74.063644	4.624335	19.71	13.03
10	Kuwait City	47.978329	29.36972	19.67	25.44

Radiation: 0 – 22+ MJ/m² (where 0 indicates no solar radiation, and the maximum values indicates exceptionally sunny, clear skies, or high altitude).

TOP 10 cities with Wind potential for Energy projects (across all the data).

```
SELECT
    d.city_name,
    c.longitude,
    c.latitude,
    ROUND(AVG(d.shortwave_radiation_sum),2) AS avg_shortwave_radiation_sum,
    ROUND(AVG(d.wind_speed_10m_max),2) AS avg_wind_speed
FROM `ds203-finalproject.forbes_cities_weather.daily_data_cleaned` AS d
INNER JOIN `ds203-finalproject.forbes_cities_weather.cities` AS c
ON c.city_name = d.city_name
GROUP BY d.city_name, c.longitude, c.latitude
ORDER BY avg_wind_speed DESC
LIMIT 10;
```

Row	city_name	longitude	latitude	avg_shortwave_radia	avg_wind_speed
1	Helsinki	24.93545	60.16952	10.9	29.04
2	San Francisco	-122.419418	37.774929	18.73	28.09
3	Tokyo	139.691711	35.689499	14.6	26.67
4	Auckland	174.743042	-36.85582	15.63	25.9
5	Copenhagen	12.56553	55.675941	11.43	25.76
6	Chicago	-87.650047	41.850029	15.15	25.64
7	Kuwait City	47.978329	29.36972	19.67	25.44
8	Busan	129.050003	35.133331	15.42	24.19

Wind speed: 1 – 29+km/h (where 1 Indicates calm conditions, and the maximum value represents high wind speeds).

Q2 - How have temperature and precipitation trends changed over the years across cities?

- Purpose: Detect long-term climate changes by exploring how temperature and precipitation patterns over the years.
- Parameters Used: City, Year, Temperature, Precipitation sum.

```
SELECT
    city_name,
    EXTRACT(YEAR FROM datetime) AS year,
    AVG(temperature_mean) AS avg_temperature_mean,
    AVG(precipitation_sum) AS Avg_precipitation_sum
FROM `ds203-finalproject.forbes_cities_weather.daily_data_cleaned`
GROUP BY city_name, year
ORDER BY city_name, year ASC;
```

Row	city_name	year	avg_temperature_me	Avg_precipitation_su
1	Abu Dhabi	2020	27.91857923497…	0.223770491803…
2	Abu Dhabi	2021	28.52136986301…	0.017260273972…
3	Abu Dhabi	2022	28.17808219178…	0.120273972602…
4	Abu Dhabi	2023	28.26602739726…	0.119726027397…
5	Abu Dhabi	2024	29.02801866128…	0.847204919208…
6	Amsterdam	2020	11.34426229508…	2.522131147540…
7	Amsterdam	2021	10.19095890410…	2.586027397260…
8	Amsterdam	2022	11.22904109589…	2.197534246575…
9	Amsterdam	2023	11.41534246575…	3.402191780821…
10	Amsterdam	2024	12.41904856161…	3.243550434158…
11	Athens	2020	18.33989071038…	1.583606557377…

Average value across the years:

```
SELECT
    EXTRACT(YEAR FROM datetime) AS year,
    AVG(temperature_mean) AS avg_temperature_mean,
    AVG(precipitation_sum) AS Avg_precipitation_sum,
    CORR(temperature_mean, precipitation_sum) AS Temp_Precip_Corr
FROM `ds203-finalproject.forbes_cities_weather.daily_data_cleaned`
GROUP BY year
ORDER BY year ASC;
```

Row	year	avg_temperature_me	Avg_precipitation_su	Temp_Precip_Corr
1	2020	15.31733606557...	2.913418032786...	0.041013587155...
2	2021	14.90412328767...	2.940742465753...	0.054457254654...
3	2022	15.14577260273...	2.841756164383...	0.052873006511...
4	2023	15.52204109589...	2.916939726027...	0.020531668223...
5	2024	16.65320470779...	3.202464055421...	0.022366072832...

Findings:

- The average temperature has been increasing, going from around 15.3°C in 2020 to about 16.7°C in 2024.
- Precipitation has also been showing an upward trend, from 2.9 mm in 2020 to 3.2 mm in 2024, with variations.
- The correlation between temperature and precipitation is weak. Precipitation changes aren't directly ruled by temperature trends.

Q3 - Which cities have the best comfortable temperatures and precipitation levels?

- Purpose: Identify cities offering favorable conditions for outdoor activities for living and tourism.
- Parameters Used: City, "Feels Like" Temperature, Precipitation level and hours.

Comfortable weather conditions set	Comfortable temperature: 10 – 28 ºC Comfortable precipitation level: until 10 mm (light to moderate rain) Comfortable precipitation hours: until 4 h (light to moderate rain)

```
SELECT
    city_name,
    AVG(temp_feels_like_mean) AS avg_temp_feels_like,
    AVG(temp_feels_like_min) AS avg_temp_feels_like_min,
    AVG(temp_feels_like_max) AS avg_temp_feels_like_max,
    AVG(precipitation_sum) AS Avg_precipitation_sum,
    AVG(precipitation_hours) AS Avg_precipitation_hours
FROM `ds203-finalproject.forbes_cities_weather.daily_data_cleaned`
GROUP BY city_name
HAVING (AVG(temp_feels_like_mean) >= 10 AND AVG(temp_feels_like_mean) <= 28
    AND AVG(temp_feels_like_min) >= 10 AND AVG(temp_feels_like_min) <= 28
    AND AVG(temp_feels_like_max) >= 10 AND AVG(temp_feels_like_max) <= 28
    AND AVG(precipitation_sum) <= 10 AND AVG(precipitation_hours) <= 4);
```

Number of cities with comfortable weather:

Count the number of cities with subquery.

```
SELECT COUNT(city_name) AS cities_confortable_weather
FROM (
        SELECT
            city_name,
            AVG(temp_feels_like_mean) AS avg_temp_feels_like,
            AVG(temp_feels_like_min) AS avg_temp_feels_like_min,
            AVG(temp_feels_like_max) AS avg_temp_feels_like_max,
            AVG(precipitation_sum) AS Avg_precipitation_sum,
            AVG(precipitation_hours) AS Avg_precipitation_hours
        FROM `ds203-finalproject.forbes_cities_weather.daily_data_cleaned`
        GROUP BY city_name
        HAVING (AVG(temp_feels_like_mean) >= 10 AND AVG(temp_feels_like_mean) <= 28
                AND AVG(temp_feels_like_min) >= 10 AND AVG(temp_feels_like_min) <= 28
                AND AVG(temp_feels_like_max) >= 10 AND AVG(temp_feels_like_max) <= 28
                AND AVG(precipitation_sum) <= 10 AND AVG(precipitation_hours) <= 4) );
```

Row	cities_confortable_w
1	18

Top 10 Cities by Temperature and Precipitation:

```
SELECT
    city_name,
    AVG(temp_feels_like_mean) AS avg_temp_feels_like,
    AVG(temp_feels_like_min) AS avg_temp_feels_like_min,
    AVG(temp_feels_like_max) AS avg_temp_feels_like_max,
    AVG(precipitation_sum) AS Avg_precipitation_sum,
    AVG(precipitation_hours) AS Avg_precipitation_hours
FROM `ds203-finalproject.forbes_cities_weather.daily_data_cleaned`
GROUP BY city_name
HAVING (AVG(temp_feels_like_mean) >= 10 AND AVG(temp_feels_like_mean) <= 28
        AND AVG(temp_feels_like_min) >= 10 AND AVG(temp_feels_like_min) <= 28
        AND AVG(temp_feels_like_max) >= 10 AND AVG(temp_feels_like_max) <= 28
        AND AVG(precipitation_sum) <= 10 AND AVG(precipitation_hours) <= 4)
ORDER BY avg_temp_feels_like DESC, Avg_precipitation_sum
LIMIT 10;
```

Row	city_name	avg_temp_feels_like	avg_temp_feels_like	avg_temp_feels_like	Avg_precipitation_su	Avg_precipitation_ho
1	Tel Aviv	21.88843209743...	17.57708520336...	26.21475314111...	1.472649648514...	2.237230419977...
2	Boston	21.09910179096...	16.20478100359...	27.44091659173...	3.543364744995...	3.495459704880...
3	Philadelphia	18.84620735282...	13.56437237703...	25.59579740898...	4.005964063951...	3.609534619750...
4	Nashville	18.38826183636...	13.14598418180...	25.10992907754...	3.860788127515...	3.438706015891...
5	Athens	18.14927205203...	13.45795921017...	22.55493475291...	1.291775641703...	1.826901248581...
6	Valencia	17.74944231309...	13.26823162788...	22.49386778356...	1.330311396527...	1.846197502837...
7	Washington	17.62146274442...	12.41874241108...	24.08001988345...	3.272763155892...	3.395005675368...
8	Naples	17.37106546860...	14.87220438611...	19.72502555881...	3.183659864178...	3.528944381384...
9	Austin	17.30699055373...	12.27816352345...	23.71412885053...	3.653523655324...	3.270147559591...
10	Los Angeles	16.72140599073...	11.61760733730...	24.23206301625...	1.374579273939...	1.439841089670...

Findings:

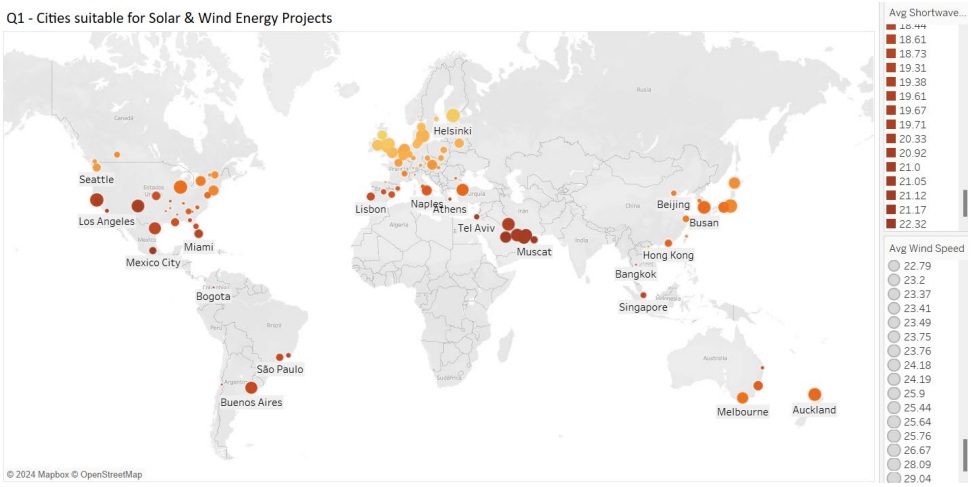
- There are 18 cities that meet the conditions set for comfortable weather.
- Tel Aviv (21.8°C) and Boston (21°C) lead with the highest average "feels like" temperatures among the listed cities and few precipitation levels.

5. Data Visualization

Q1 - Which cities are most suitable for solar and wind energy projects across geographic regions?

- Purpose: Identify cities with high solar and wind speed potential for energy projects.

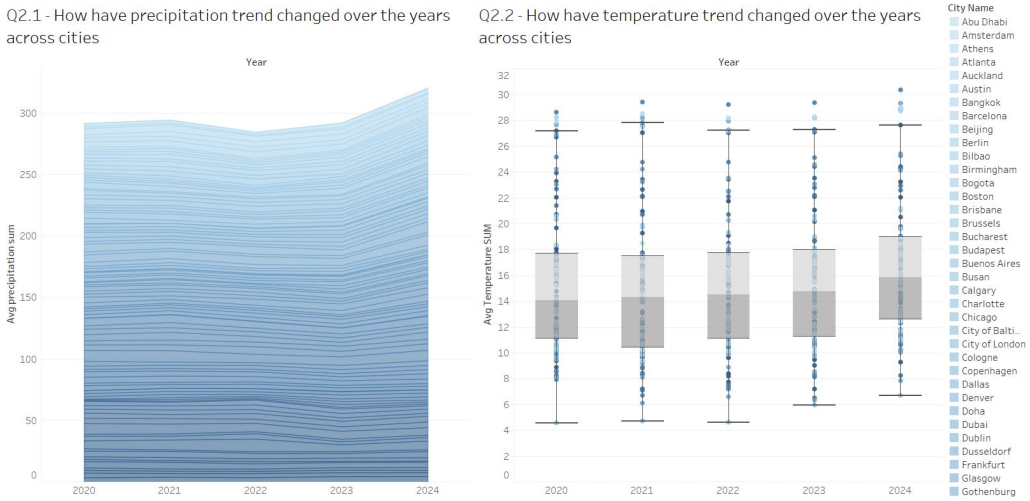
Visualization:



- Solar Energy Potential: Cities like Muscat, Tel Aviv, and Los Angeles have high solar radiation, making them ideal for solar energy projects.
- Wind Energy Potential: Helsinki, Tokio, and San Francisco have strong wind speeds, perfect for wind energy developments.
- Hybrid Opportunities: Locations like San Francisco, Doha and Kuwait show a mix of high solar radiation and wind speed, making them suitable for combined solar and wind energy initiatives.

Q2 - How have temperature and precipitation trends changed over the years across cities?

- Purpose: Detect long-term climate changes by exploring how temperature and precipitation patterns over the years.

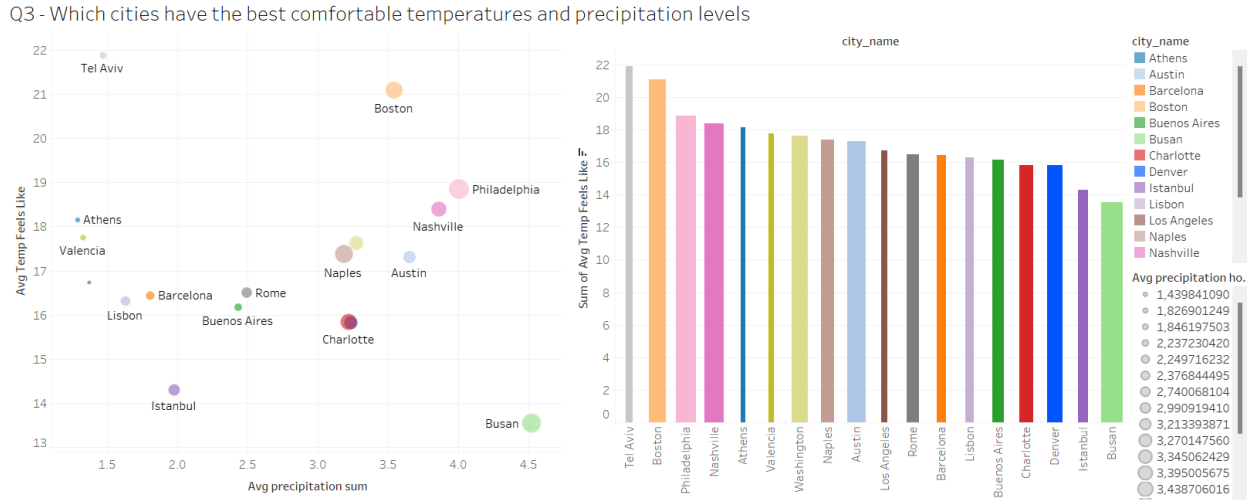


Findings:

- The average temperature has been increasing, going from around from 2020 to 2024.
- Precipitation has also been showing an upward trend.
- This suggests a clear warming trend, which aligns with broader concerns about global climate change.

Q3 - Which cities have the best comfortable temperatures and precipitation levels?

- Purpose: Identify cities offering favorable conditions for outdoor activities for living and tourism.



Note:

Comfortable weather conditions set	*Comfortable temperature: 10 – 28 ºC* *Comfortable precipitation level: until 10 mm (light to moderate rain)* *Comfortable precipitation hours: until 4 h (light to moderate rain)*

Findings:

- From the 18 cities that meet the conditions set for comfortable weather.
- From the top 5, Tel Aviv and Athens have the highest average "feels like" temperatures with few precipitation levels.

6. Conclusion

Findings:

Climate Trends:

- Average global temperatures and precipitation are rising across the years due to global warming.
- Temperature and precipitation trends are weakly correlated, suggesting independent drivers.

Energy Potential:

- High solar potential cities: Muscat, Tel Aviv, Los Angeles.
- High wind speed cities: Helsinki, Tokyo, San Francisco.
- Hybrid opportunities: San Francisco, Doha, Kuwait (suitable for combined solar and wind energy projects).

Comfortable Weather:

- 18 cities meet favorable conditions for outdoor activities.
- Top cities: Tel Aviv and Athens stand out with warm, consistent temperatures and low precipitation.

Recommendations:

- Focus solar energy developments in cities with high-radiation levels (e.g., Muscat).
- Invest in wind energy initiatives in areas with strong wind speeds (e.g., Helsinki).
- Explore hybrid energy projects in cities with mixed potentials (e.g., San Francisco).
- Prioritize tracking of temperature and precipitation to adapt urban and environmental strategies.
- Highlight cities with comfortable weather for tourism promotion (e.g., Tel Aviv).
- Explore across the findings to identify cities best suited for outdoor activities.

7. Appendix

References:

- Hourly and Daily Weather Dataset of Forbes Top 100 Best Cities To Live, Work And Visit from: https://www.kaggle.com/datasets/bwandowando/forbes-top-100-cities-weather-data2020-ytd?select=cities.csv
- https://open-meteo.com/en/docs/historical-weather-api#start_date=2020-01-01&end_date=2024-10-27