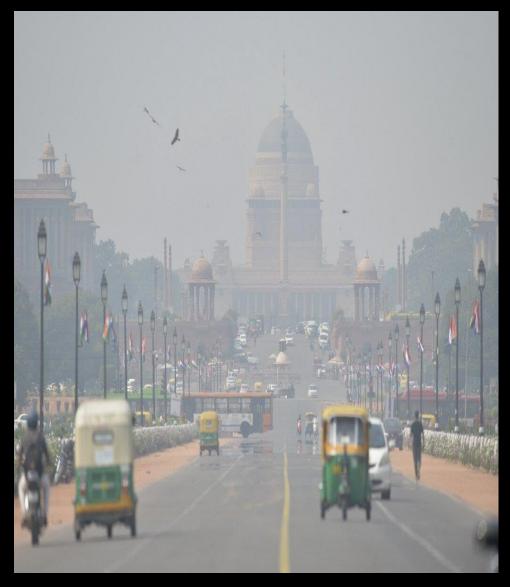# Air Quality Data Analysis

❑ Project Overview :

Air quality data analysis involves the systematic examination of data related to pollutants like PM2.5, NO2, and O3 to assess environmental health. The process begins with collecting data from monitoring stations, satellites, or crowdsourced platforms, followed by preprocessing steps to clean and prepare the data. Statistical and machine learning techniques are then applied to identify trends, patterns, and sources of pollution. Visualization tools, such as maps and graphs, help in presenting the data clearly, making it easier to understand and communicate findings. The analysis provides crucial insights into pollution levels, identifies hotspots, and suggests actions to improve air quality. These insights are vital for public health, informing policy decisions, and guiding urban planning efforts to mitigate pollution impacts.

❑ **Project Goal :**

The goal of air quality data analysis is to understand the levels, patterns, and sources of air pollution to protect public health and the environment. By analysing trends over time, the analysis aims to identify pollution hotspots and determine the effectiveness of regulatory measures. Predictive models help forecast future air quality, enabling timely interventions. Additionally, the analysis seeks to uncover correlations between pollution levels and health outcomes, supporting evidence-based policy-making. Ultimately, the goal is to provide actionable insights that can guide efforts to reduce pollution, improve air quality, and enhance the quality of life in affected areas.

❑ Tasks :

1. What is the air quality data recorded in the city of "New York"?
2. What was the air quality in "Los Angeles" on "2024-08- 01"?
3. How many entries exist for each city in the dataset?
4. What is the maximum PM2.5 level recorded for each city?
5. Are there any records with missing PM10 values?
6. What are the average PM2.5, PM10, NO2, and SO2 levels for each city?
7. On which dates was the PM2.5 level above 100 in "Mumbai"?
8. What are the average PM2.5 and PM10 levels by country, sorted in descending order?
9. What is the correlation between temperature and PM2.5 levels by city?
10. Which city had the highest average NO2 levels in July 2023?
11. What is the trend in PM2.5 levels over time in "Beijing"? Show monthly averages for the past year.
12. Which are the top 5 cities with the highest average PM2.5 levels over the last 6 months?
13. How does wind speed affect PM10 levels in different cities? Compare days with wind speeds above and below 15 km/h
14. Are there any days in "Dubai" where PM2.5 and PM10 levels spiked by more than 50% compared to the previous day?

Visualization :

🤔 🤔

- **Question :**

➢ What is the air quality data recorded in the city of

"New York"?

- **Query :**

➢ select * from air_quality_data where city="New York";

- **Result :**

City	Country	Date	PM2.5	PM10	NO2	SO2	CO	O3	Temperature	Humidity	Wind Speed
New York	USA	2023-11-23	72.52	23.43	92.01	15.26	4.57	106.5	32.14	76.57	9.42
New York	USA	2023-06-26	82.4	80.55	91.9	46.52	0.27	146.03	30.84	49.96	17.45
New York	USA	2023-10-12	62.34	38.48	57.36	32.1	2.58	154.98	37.45	48.65	17.61
New York	USA	2023-03-11	71.19	152.28	59.61	38.15	8.53	95.3	33.59	43.37	13.96
New York	USA	2023-06-06	21.8	71.51	74.7	11.27	1.75	76.6	39.05	34.01	11.81
New York	USA	2023-06-16	134.05	182.71	19.94	38.66	1.06	83.26	34.37	64.41	16.65

- **Explanation :**

➢ This question seeks to extract all air quality records for New York City, providing a complete view of the pollutant levels recorded over time. It helps in understanding the overall air quality trends and identifying specific pollution events in the city.

Question :

➤ What was the air quality in "Los Angeles" on "2024-08-01"?

Query :

➤ select * from air_quality_data where city="Los Angeles" and date= '2023-08-01';

Result :

City	Country	Date	PM2.5	PM10	NO2	SO2	CO	O3	Temperature	Humidity	Wind Speed
Los Angeles	USA	2023-08-01	23.91	97.33	5.26	42.39	5.99	127.49	25.26	27.83	7.56
Los Angeles	USA	2023-08-01	14.45	15.16	60.45	24.23	8.62	181.65	27.59	43.17	12.58
Los Angeles	USA	2023-08-01	51.33	181.54	5.68	35.14	0.75	174.93	22.87	85.14	11.17
Los Angeles	USA	2023-08-01	55.23	75.74	6.19	24.98	3.91	152.42	13.12	14.08	12.13

Explanation :

➤ The goal is to retrieve air quality data for Los Angeles on a particular date, which allows for the examination of environmental conditions on that specific day. This is useful for assessing the impact of particular events or weather conditions on air quality.

- **Question :**

  ➢ How many entries exist for each city in the dataset?

- **Query :**

  ➢ SELECT City, COUNT(*) AS EntryCount FROM Air_Quality_data GROUP BY City;

- **Result :**

City	EntryCount
▶ Bangkok	499
Istanbul	492
Rio de Janeiro	491
Mumbai	540
Paris	489
Los Angeles	497
Tokyo	491
New York	465
London	489
Seoul	522
Cairo	510
Mexico City	463

- **Explanation :**

  ➢ This question aims to determine the volume of data available for each city in the dataset, which is important for assessing data coverage and reliability when conducting comparative analysis across different locations.

- **Question :**

➤ What is the maximum PM2.5 level recorded for each city?

- **Query :**

➤ SELECT City, MAX(`PM2.5`) AS `Max_PM2.5` FROM Air_Quality_data GROUP BY City;

- **Result :**

	City	Max_PM2.5
▶	Bangkok	149.86
	Istanbul	149.91
	Rio de Janeiro	149.71
	Mumbai	149.59
	Paris	149.38
	Los Angeles	149.95
	Tokyo	148.4
	New York	149.69
	London	149.92
	Seoul	149.71

- **Explanation :**

➤ The focus here is on identifying the highest PM2.5 levels recorded in each city, which is crucial for pinpointing extreme pollution events and understanding the severity of air quality issues across different regions.

- **Question :**

➢ Are there any records with missing PM10 values?

- **Query :**

➢ SELECT *FROM Air_Quality_dataWHERE PM10 IS NULL;

- **Result :**

City	Country	Date	PM2.5	PM10	NO2	SO2	CO	O3	Temperature	Humidity	Wind Speed

- **Explanation :**

➢ This question checks for any instances where the PM10 values are missing, which is essential for data quality assurance. Missing data can impact the accuracy of analysis and need to be addressed appropriately . No missing value was there .

- **Question :**

  ➤ What are the average PM2.5, PM10, NO2, and SO2 levels for each city?

- **Query :**

  ➤ SELECT city, ROUND(AVG(`PM2.5`), 2) AS `avg_pm2.5`, ROUND(AVG(PM10), 2) AS avg_pm10, ROUND(AVG(NO2), 2) AS avg_no2, ROUND(AVG(SO2), 2) AS avg_so2 FROM air_quality_data GROUP BY city;

  ➤ **Result :**

city	avg_pm2.5	avg_pm10	avg_no2	avg_so2
Bangkok	77.46	103.93	51.51	25.58
Istanbul	77.71	103.14	52.02	25.77
Rio de Janeiro	75.67	105.44	52.1	24.58
Mumbai	78.9	105.84	53.34	25.03
Paris	74.69	104.64	53.82	24.57
Los Angeles	76.42	105.54	53.52	26.15
Tokyo	78.87	105.04	52.67	27.98
New York	77.86	102.95	51	25.17
London	77.61	103.84	52.33	25.5
Seoul	74.8	105	51.88	25.01

- **Explanation :**

  ➤ This question calculates the average levels of key pollutants (PM2.5, PM10, NO2, SO2) for each city, providing a summary of typical air quality conditions. It helps in comparing pollution levels across cities and identifying areas with consistently high pollution.

- **Question :**

➢ On which dates was the PM2.5 level above 100 in "Mumbai"?

- **Query :**

➢ SELECT date FROM air_quality_data WHERE city = 'Mumbai' AND `PM2.5` > 100;

- **Result :**

date
► 2023-03-16
2023-06-26
2023-06-10
2023-03-13
2023-04-02
2023-01-23
2023-04-28
2023-04-09
2023-06-17
2023-03-13
2023-04-07
2023-12-28

- **Explanation :**

➢ The goal is to identify specific dates in Mumbai when the PM2.5 levels exceeded a critical threshold (e.g., 100), which can highlight days with particularly poor air quality and potential health risks.

- **Question :**

➢ What are the average PM2.5 and PM10 levels by country, sorted in descending order?

- **Query :**

➢ SELECT city, round(AVG(`PM2.5`),2) as `avg_pm2.5`, round(AVG(pm10),2) as avg_pm10FROM air_quality_datagroup by city order by `avg_pm2.5`desc,avg_pm10 desc;

- **Result :**

city	avg_pm2.5	avg_pm10
▶ Dubai	80.01	103.89
Sydney	78.93	103.37
Mumbai	78.9	105.84
Tokyo	78.87	105.04
Mexico City	78.86	104.53
Beijing	78.63	103.19
Moscow	77.88	105.94

- **Explanation :**

➢ This question seeks to compare average PM2.5 and PM10 levels across different countries, ranked in descending order. This allows for an assessment of how different nations are performing in terms of air quality and pollution control.

- **Question :**

  ➢ What is the correlation between temperature and PM2.5 levels by city?

- **Query :**

  ➢ SELECT City,

    round((SUM(Temperature * `PM2.5`) - SUM(Temperature) * SUM(`PM2.5`) / COUNT(*)) /

     (SQRT((SUM(Temperature * Temperature) - SUM(Temperature) * SUM(Temperature) / COUNT(*)) *

      (SUM(`PM2.5` * `PM2.5`) - SUM(`PM2.5`) * SUM(`PM2.5`) / COUNT(*)))),2) AS Temperature_PM25_Correlation

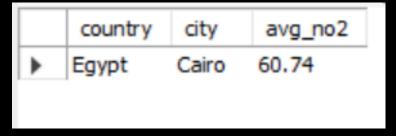   FROM Air_Quality_data

    GROUP BY City;

- **Result :**

City	Temperature_PM25_Correlation
▶ Bangkok	-0.01
Istanbul	0
Rio de Janeiro	-0.01
Mumbai	-0.07
Paris	-0.05
Los Angeles	-0.05
Tokyo	-0.03

- **Explanation :**

➢ The objective is to explore the relationship between temperature and PM2.5 levels in each city, which can reveal how weather conditions influence air pollution. Understanding this correlation can help in predicting pollution levels based on temperature forecasts.

- **Question :**

➢ Which city had the highest average NO2 levels in July 2023?

- **Query :**

➢ select country,city, round(avg(NO2),2) as `avg_no2` from air_quality_data where month(date)=7 and year(date)=2023 group by country, city order by `avg_no2` desc limit 1;

➢ **Result :**

country	city	avg_no2
▶ Egypt	Cairo	60.74

- **Explanation :**

➢ This question aims to find the city with the highest average NO2 levels during July 2023. It identifies locations with significant NO2 pollution, which can be critical for targeted air quality interventions during that period.

- **Question :**

  ➢ What is the trend in PM2.5 levels over time in "Beijing"?

- **Query :**

  ➢ SELECT
     DATE_FORMAT(Date, '%Y-%m') AS Month,
     ROUND(AVG(`PM2.5`), 2) AS `Avg_PM2.5`
     FROM air_quality_data
     WHERE City = 'Beijing'
     GROUP BY Month
     ORDER BY Month ASC;

- **Question :**

  ➢ Show monthly averages for the past year.

- **Query :**

➢ SELECT
     DATE_FORMAT(Date, '%Y-%m') AS Month,
     ROUND(AVG(`PM2.5`), 2) AS `Avg_PM2.5`
    FROM air_quality_data
    WHERE City = 'Beijing'
    AND Date >= DATE_SUB(CURDATE(), INTERVAL 1 YEAR)
    GROUP BY Month
    ORDER BY Month ASC

Result :

Month	Avg_PM2.5
▶ 2023-08	77.03
2023-09	75.36
2023-10	80.35
2023-11	83.65
2023-12	88.22

Month	Avg_PM2.5
▶ 2023-01	67.25
2023-02	76.34
2023-03	70.31
2023-04	82.02
2023-05	69.82
2023-06	86.96
2023-07	84

Explanation :

➢ The focus is on analysing the trend of PM2.5 levels in Beijing over the past year, with monthly averages providing a clear view of how pollution levels have changed over time. This helps in identifying seasonal patterns or long-term improvements or deteriorations in air quality.

▪ Question :

➤ Which are the top 5 cities with the highest average PM2.5 levels over the last 6 months?

▪ Query :

➤ SELECT City, ROUND(AVG(`PM2.5`), 2) AS Avg_PM25FROM air_quality_dataWHERE Date BETWEEN '2023-07-01' AND '2023-12-31'GROUP BY CityORDER BY Avg_PM25 DESCLIMIT 5;

▪ Result :

City	Avg_PM25
▶ Dubai	84.39
Beijing	82.25
Moscow	81.26
Mexico City	79.05
Mumbai	78.97

▪ Explanation :

➤ This question identifies the top 5 cities with the highest average PM2.5 levels over the last 6 months, highlighting regions with severe air pollution. This information is useful for prioritizing pollution control efforts in the most affected cities.

- **Question :**

➤ How does wind speed affect PM10 levels in different cities? Compare days with wind speeds above and below 15 km/h.

- **Query :**

➤ SELECT
    City,
    CASE
    WHEN `Wind Speed` < 5 THEN '< 5 m/s'
    WHEN `Wind Speed` BETWEEN 5 AND 10 THEN '5-10 m/s'
    WHEN `Wind Speed` BETWEEN 10 AND 15 THEN '10-15 m/s'
    WHEN `Wind Speed` BETWEEN 15 AND 20 THEN '15-20 m/s'
    END AS WindSpeedRange,
    ROUND(AVG(PM10), 2) AS Avg_PM10
  FROM air_quality_data GROUP BY City, WindSpeedRange
  ORDER BY City, WindSpeedRange;

- **Question :**

➤ Compare days with wind speeds above and

below 15 km/h.

- **Query :**

➤ SELECT
    City,
   CASE
   WHEN `Wind Speed` >= 15 THEN 'Above 15 km/h'
   ELSE 'Below 15 km/h'
   END AS WindSpeedCategory,
   ROUND(AVG(PM10), 2) AS Avg_PM10
   FROM air_quality_data
   GROUP BY City, WindSpeedCategory
   ORDER BY City, WindSpeedCategory;

- **Result :**

	City	WindSpeedRange	Avg_PM10
▶	Bangkok	< 5 m/s	101.36
	Bangkok	10-15 m/s	107
	Bangkok	15-20 m/s	97.96
	Bangkok	5-10 m/s	109.64
	Beijing	< 5 m/s	95.79

	City	WindSpeedCategory	Avg_PM10
▶	Bangkok	Above 15 km/h	97.96
	Bangkok	Below 15 km/h	106.13
	Beijing	Above 15 km/h	103.99
	Beijing	Below 15 km/h	102.85
	Berlin	Above 15 km/h	102.5

- **Explanation :**

➢ The goal is to compare PM10 levels on days with high versus low wind speeds across different cities, revealing how wind conditions influence the dispersion or concentration of particulate matter. This can inform strategies for mitigating pollution during low wind conditions.

- **Question :**

➤ Which are the top 5 cities with the highest average PM2.5 levels over the last 6 months?

- **Query :**

➤ WITH PreviousDayData AS (
    SELECT
    City,
    Date,
    PM2.5`,
    PM10,
     LAG(`PM2.5`) OVER (PARTITION BY City ORDER BY Date) AS `Prev_PM2.5`,
        LAG(PM10) OVER (PARTITION BY City ORDER BY Date) AS Prev_PM10
    FROM air_quality_data
    WHERE City = 'Dubai'
)

```sql
SELECT
 Date,
 `PM2.5`,
 PM10,
 `Prev_PM2.5`,
 Prev_PM10,
 ROUND((`PM2.5` - `Prev_PM2.5`) / `Prev_PM2.5` * 100, 2) AS `PM2.5_Spike_Percentage`,
ROUND((PM10 - Prev_PM10) / Prev_PM10 * 100, 2) AS PM10_Spike_Percentage
FROM PreviousDayData
WHERE
  ((`PM2.5` - `Prev_PM2.5`) / `Prev_PM2.5` * 100 > 50  OR (PM10 - Prev_PM10) /
Prev_PM10 * 100 > 50)    AND `Prev_PM2.5` IS NOT NULL
  AND Prev_PM10 IS NOT NULL;
```

➤ Result :

Date	PM2.5	PM10	Prev_PM2.5	Prev_PM10	PM2.5_Spike_Percentage	PM10_Spike_Percenta
2023-01-04	38.88	58.26	5.37	38.54	624.02	51.17
2023-01-05	138.61	31.71	38.88	58.26	256.51	-45.57
2023-01-06	123.67	186.24	138.61	31.71	-10.78	487.32
2023-01-06	40.27	68.38	6.4	26.34	529.22	159.61
2023-01-06	44.38	117.92	40.27	68.38	10.21	72.45
2023-01-07	82.67	76.38	19.93	92.69	314.8	-17.6
2023-01-07	141.39	128.93	13.19	47.5	971.95	171.43
2023-01-10	126	23.14	5.03	114.87	2404.97	-79.86
2023-01-10	91.01	134.79	126	23.14	-27.77	482.5

■ Explanation :

➤ This question looks for days in Dubai where both PM2.5 and PM10 levels spiked significantly compared to the previous day, indicating sudden pollution events. Identifying such spikes is crucial for understanding and addressing the causes of these sharp increases in pollution.

Thank You