Team-MetricMinds

Data Mining Project

GLOBAL AIR QUALITY & POLLUTION TRENDS

Presented By:

Lakshith, Sai Praneeth Sharma & Geethika

Why Air Quality Matters?

- Air pollution is a major global health risk → ~7 million deaths annually (WHO).
- Impacts climate, respiratory diseases, and overall life expectancy.
- Developing vs developed countries show major disparities.

Project Objectives and Goals

This project aims to analyze global air quality patterns and investigate how pollution varies across regions, time periods, economic contexts, urbanization levels, and seasons. We also explore correlations among major pollutants.

Methodology (Data Mining Steps)

Data Collection
- Extracted pollution data from OpenAQ.
- Retrieved socioeconomic indicators from the World Bank database.

Data Cleaning
- Handled missing data, standardized units, and removed outliers.

Data Processing & Mining
- Aggregation by country, city, month, and year.
- Trend analysis and correlation computations.
- Statistical and comparative analyses (e.g., urban vs. rural).

Visualization
- Time-series graphs, bar charts, scatterplots, and heatmaps.

Exploratory Question 1

Which countries/cities have the highest and lowest average pollution levels?

Relevance

- Helps identify global pollution hotspots
- Reveals inequality in air quality
- Important for environment & health-focused policy making

Understanding geographical disparities in air quality helps identify regions needing urgent policy intervention. It also highlights global inequity in exposure to harmful pollutants and supports data-driven prioritization for environmental health efforts.

Analysis Approach

- Calculated average AQI ($PM_{2.5}$/PM_{10} composite) for each country and city.
- Ranked all regions from highest to lowest average AQI.

Exploratory Question 1

WHICH COUNTRIES/CITIES HAVE THE HIGHEST AND LOWEST AVERAGE POLLUTION LEVELS?

Insights

- Country with maximum avg AQI value across all of its cities - Republic of Korea(421.0)
- Country with minimum avg AQi value across all of its cities - Palau (16.0)
- AQI Rank – India: 9th

	Country	Average_AQI
1	Republic of Korea	421.000000
2	Bahrain	188.000000
3	Mauritania	179.000000
4	Pakistan	178.788274
5	United Arab Emirates	163.666667
...	...	...
171	Bolivia (Plurinational State of)	23.787879
172	Iceland	23.000000
173	Maldives	19.000000
174	Solomon Islands	18.000000
175	Palau	16.000000

175 rows × 2 columns

Exploratory Question 2

How do pollutants (PM2.5, PM10, NO2, CO, O3) correlate with each other?

Relevance

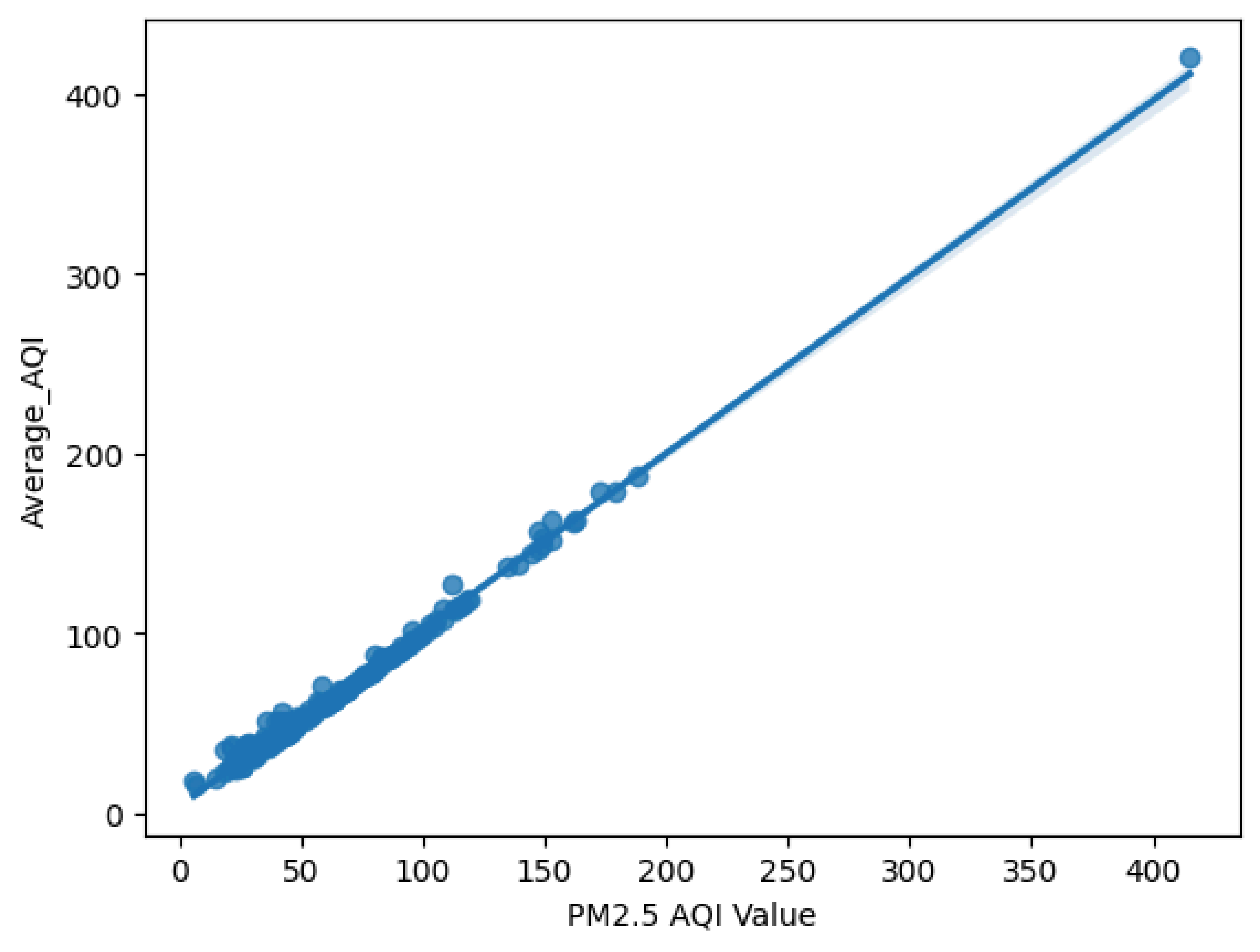
- Understanding how pollutants move together reveals:
- Whether they originate from common sources (e.g., vehicles, industry, biomass burning).
- How atmospheric chemistry links one pollutant to another.
- Which pollutants dominate overall Air Quality Index (AQI) calculations.
- This helps in designing targeted pollution-control strategies and developing predictive models.

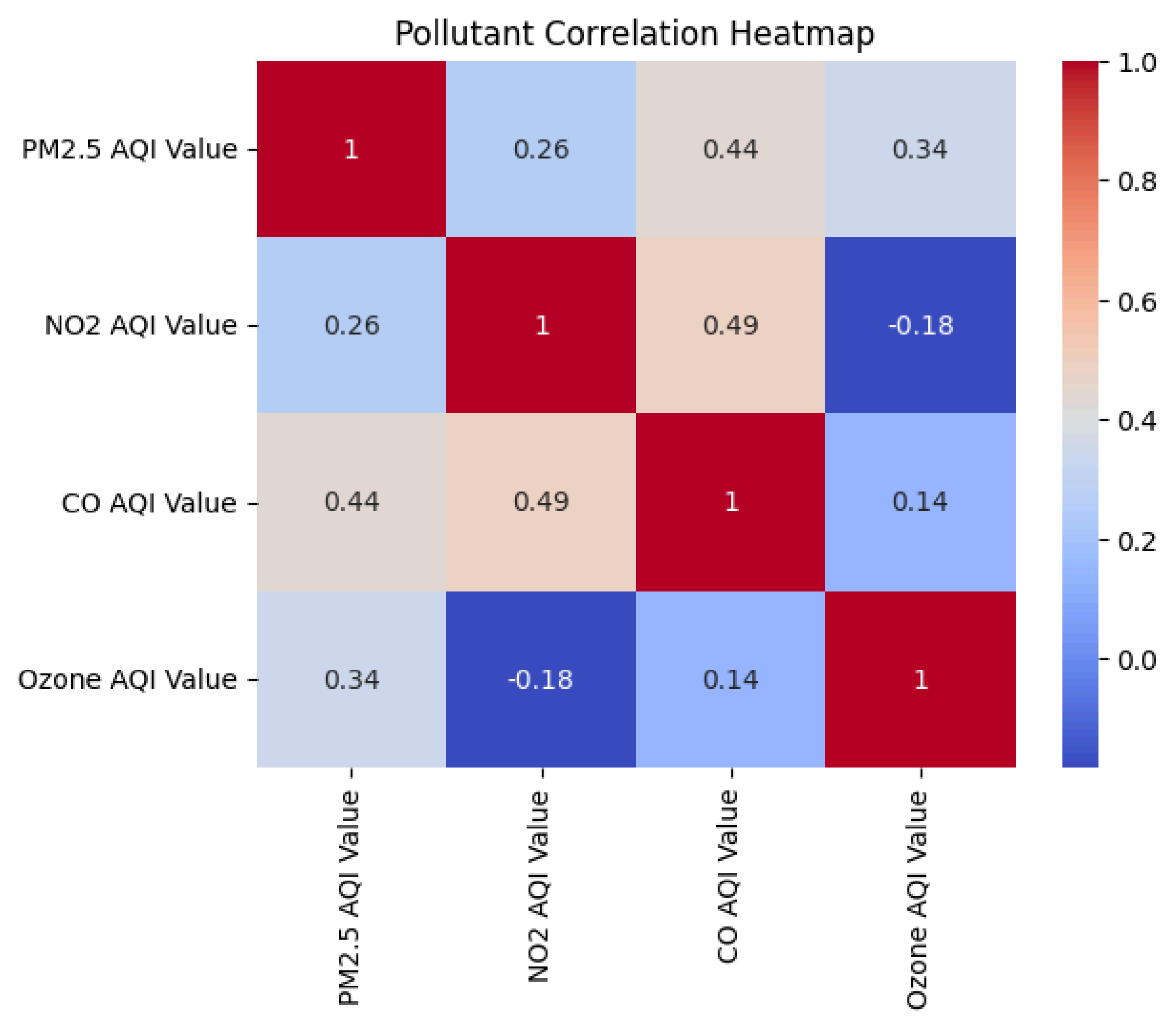
Analysis Approach

- Computed pairwise correlations between PM2.5, NO_2, CO, and Ozone.
- Visualized the correlation matrix using a heatmap.
- Used regression plots (e.g., PM2.5 vs. Average AQI) to confirm pollutant influence.

Exploratory Question 2

HOW DO POLLUTANTS (PM2.5, PM10, NO2, CO, O3) CORRELATE WITH EACH OTHER?





Insights

- PM2.5, NO_2, CO → positively correlated.
- Ozone → negative/weak correlations.
- The above graph shows a linear relationship between PM2.5 and Average AQI. This indicates that PM2.5 is a strong determinant of air quality (for all countries)

Exploratory Question 3

What is the relationship between GDP per capita and air pollution?

Relevance

- Economic status often correlates with environmental standards
- Helps understand if rich countries are cleaner

Exploring how economic growth relates to pollution levels helps validate environmental-economic theories such as the Environmental Kuznets Curve (EKC). It also guides policy by showing whether higher income enables cleaner technologies.

Analysis Approach

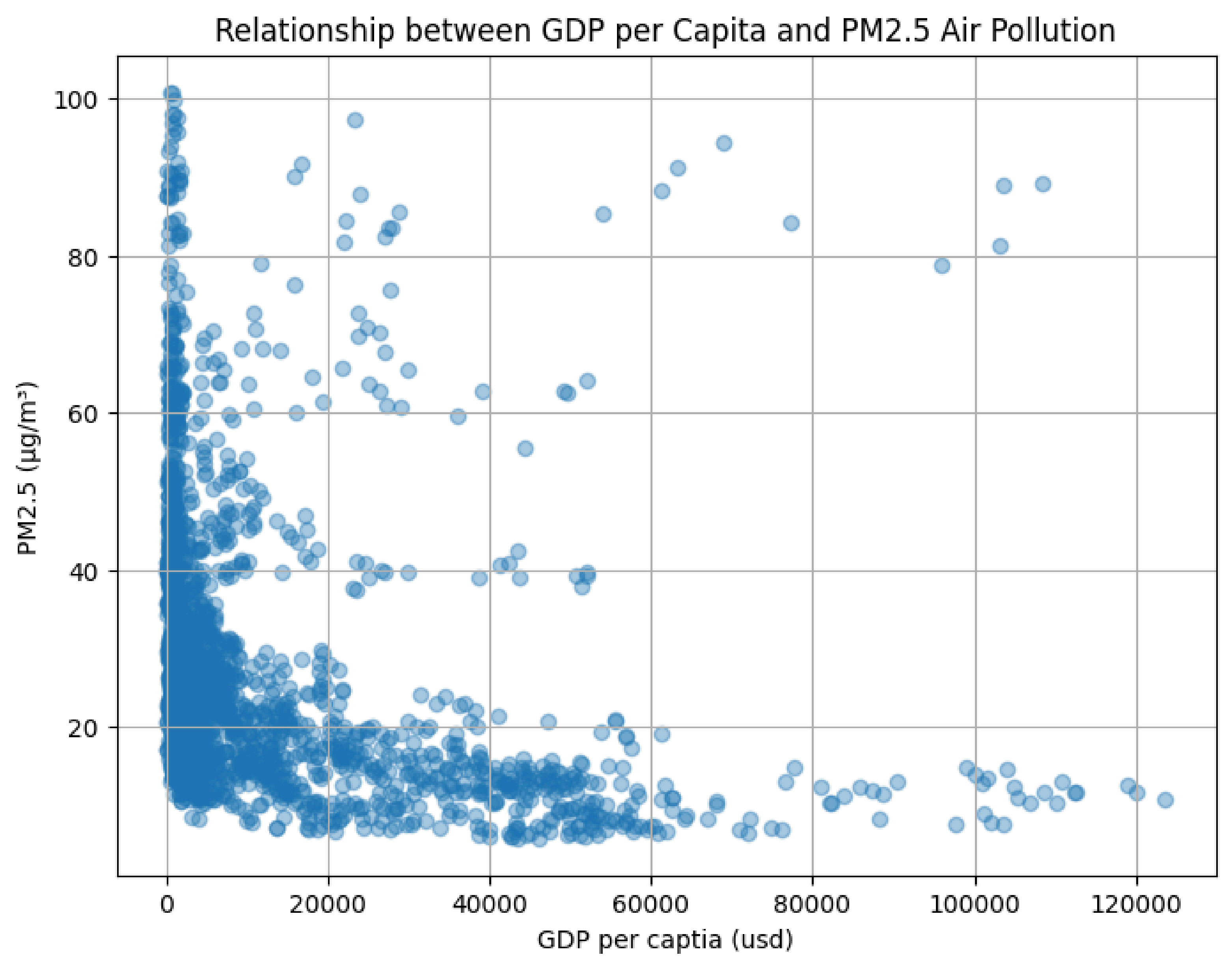
- Merged GDP per capita data with PM2.5 concentration for all countries and years.
- Conducted linear regression to quantify the relationship.
- Visualized the results using a scatterplot with a regression line.

Exploratory Question 3

WHAT IS THE RELATIONSHIP BETWEEN GDP PER CAPITA AND AIR POLLUTION?

Insights

According to the above analysis using Linear Regression we can say that, the slope of the linear line formed when we relate "GDP_per_captia" and "PM2.5" is negative which describes us that "Higher GDP => Lower Air Pollution"



Exploratory Question 3

WHAT IS THE RELATIONSHIP BETWEEN GDP PER CAPITA AND AIR POLLUTION?

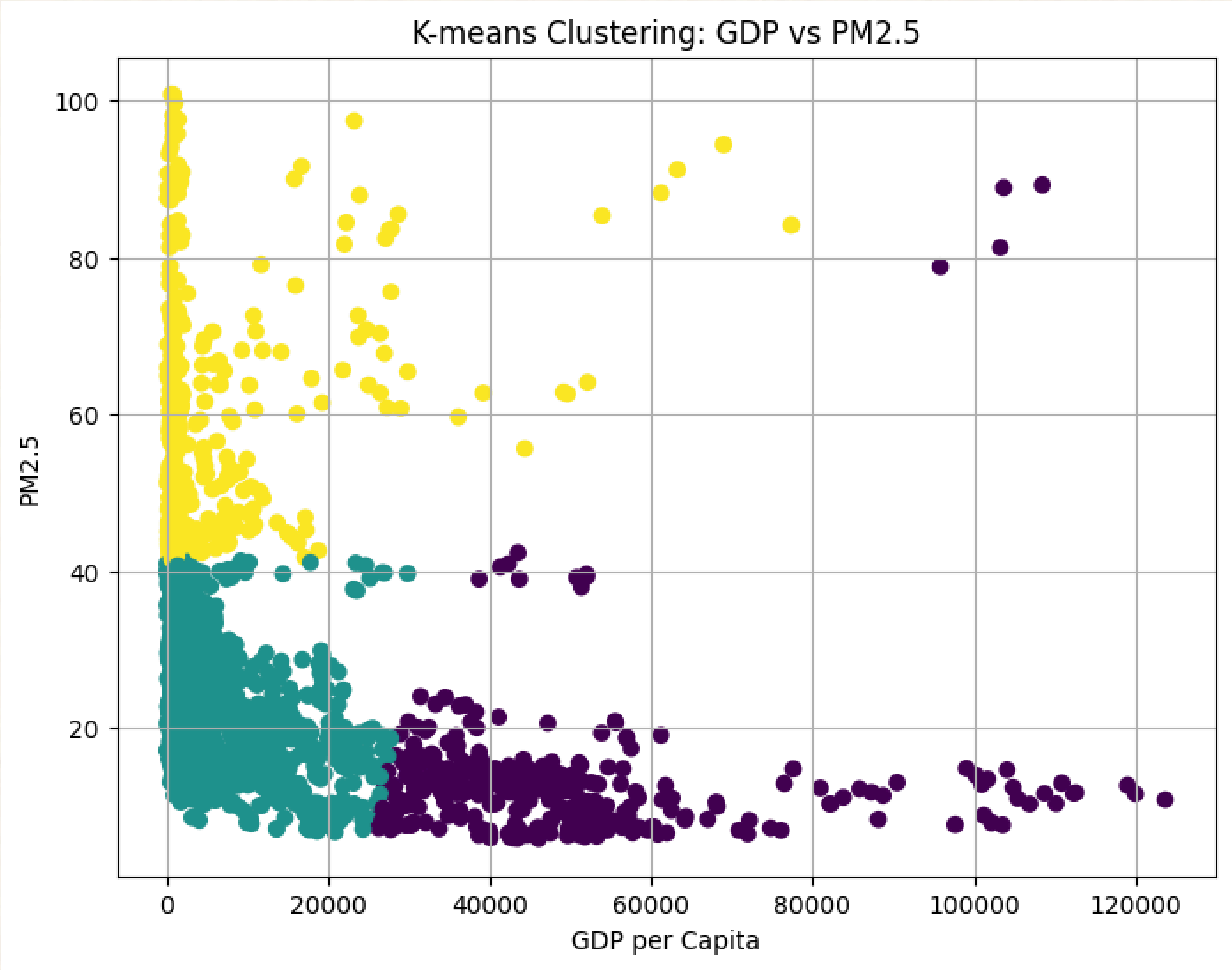
Insights

Clustering helps us group countries based on their PM2.5 levels and GDP. This reveals meaningful patterns such as:

1)Low pollution → Low GDP

2)High pollution → Low GDP

3)Low pollution → High GDP



Exploratory Question 4

Are there seasonal variations in pollution (winter vs summer)?

Relevance

- Helps understand environmental cycles
- Useful for public health warnings
- Pollution spikes usually due to winter inversion layers

Seasonal patterns influence air quality as weather conditions, temperature, wind flow, and emission cycles change across the year.

Analysis Approach

- Extracted month from timestamped pollution records.
- Computed monthly averages of PM2.5.
- Visualized seasonal trends using line charts.

Exploratory Question 4

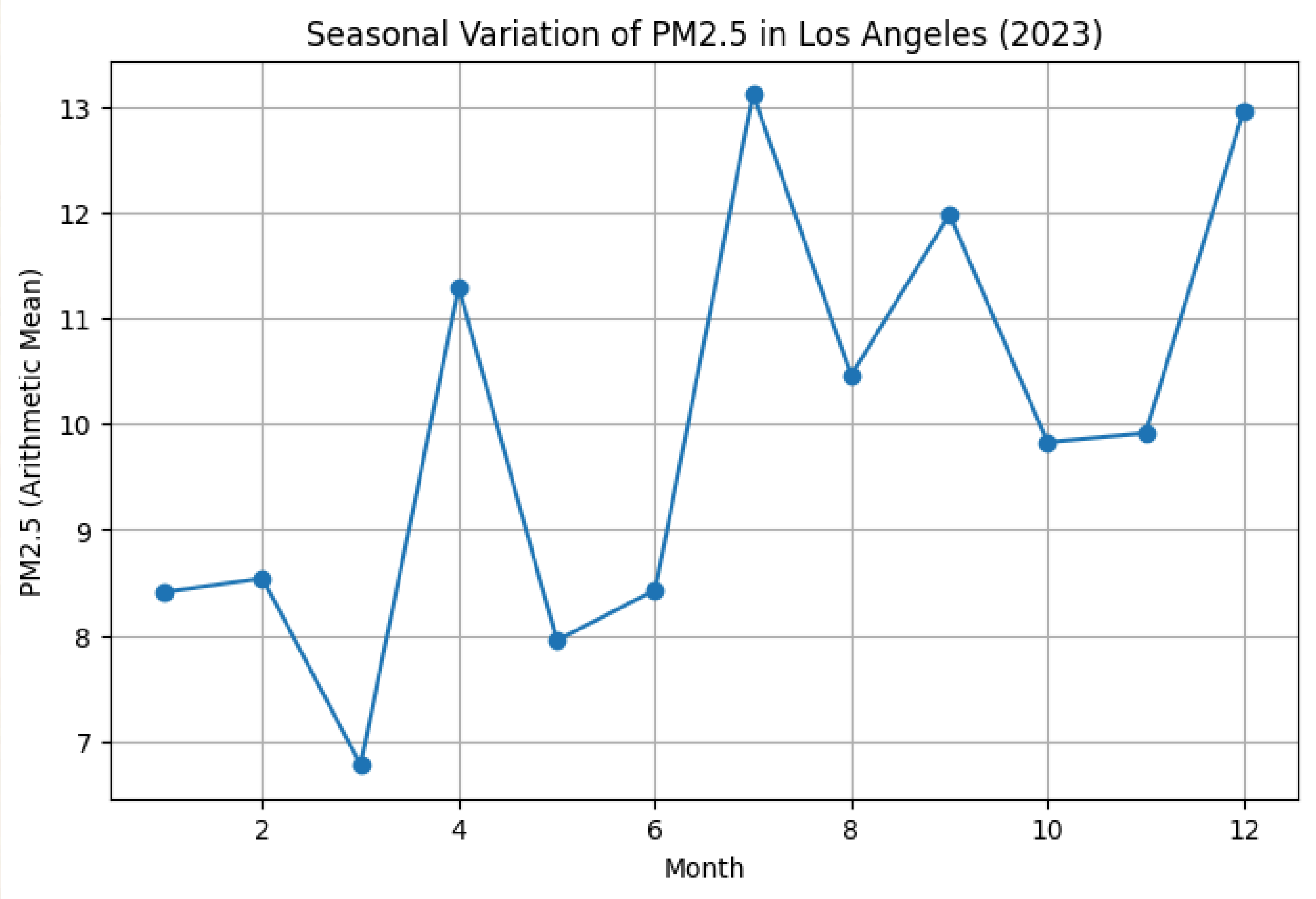
ARE THERE SEASONAL VARIATIONS IN POLLUTION (WINTER VS SUMMER)?

Insights

The seasonal analysis of PM2.5 levels in Los Angeles (2023) shows a clear and consistent pattern.

- PM2.5 concentrations are highest during the winter months (November–January), primarily due to temperature inversions, weak wind flow, and increased heating emissions.
- In contrast, PM2.5 levels drop significantly during the summer months (June–August), when stronger sunlight, warmer temperatures, and onshore sea breeze promote better atmospheric dispersion.

This demonstrates a strong inverse relationship between seasonal temperature and particulate pollution—colder seasons trap pollution, while warmer seasons help disperse it. Therefore, Los Angeles experiences poorer air quality in winter and much cleaner air in summer.



Exploratory Question 5

How has air quality changed over the last 20 years globally & regionally?

Relevance

Long-term pollution trends help assess whether air-quality policies are effective and identify regions where conditions are improving or deteriorating.

Analysis Approach

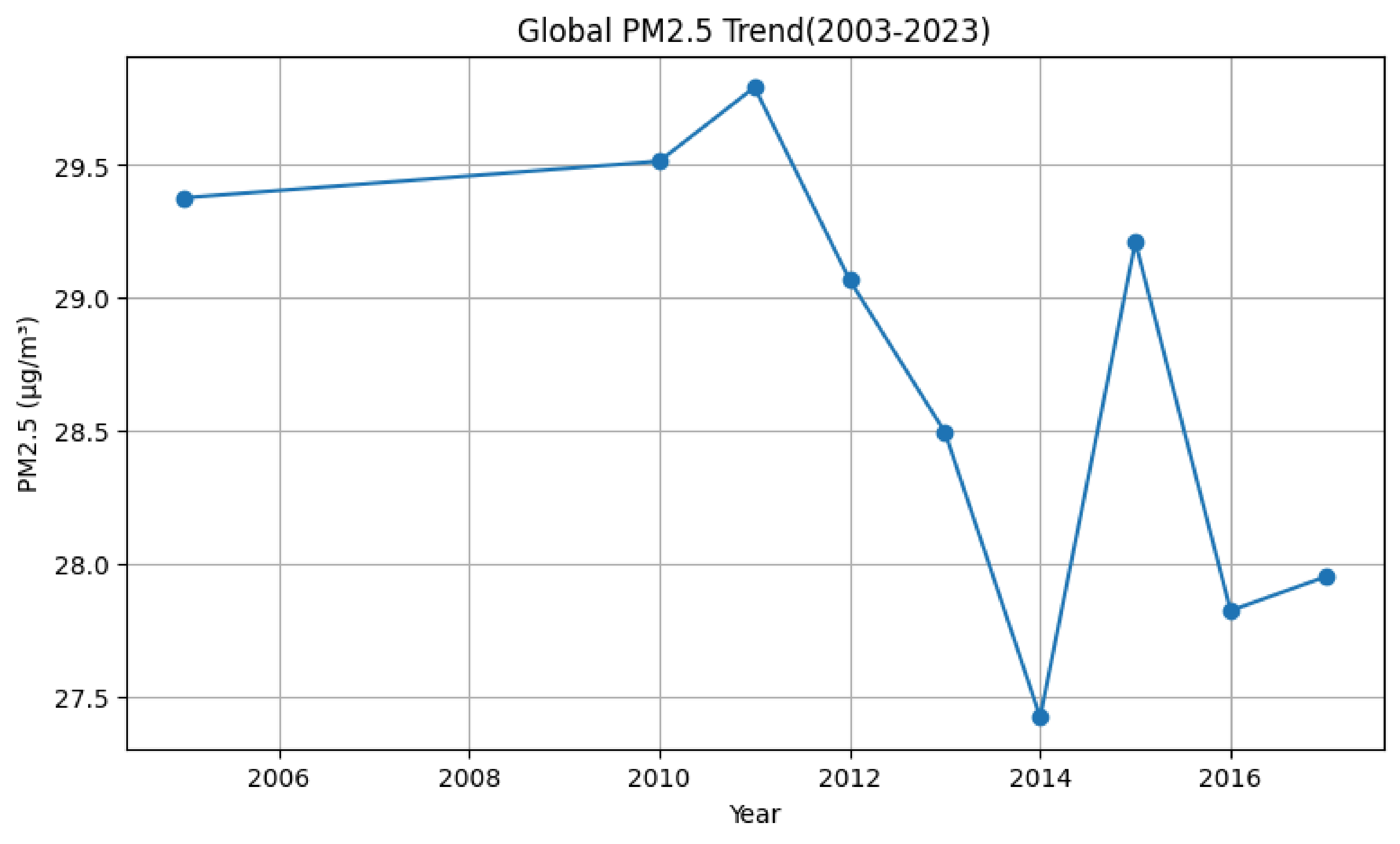
- Calculated annual mean $PM_{2.5}$ levels for each country (2000–2020).
- Aggregated results into regional averages (Asia, Europe, Americas, Africa).

Exploratory Question 5

HOW HAS AIR QUALITY CHANGED OVER THE LAST 20 YEARS GLOBALLY & REGIONALLY?

Insights

- The global PM2.5 trend over the last two decades shows a significant shift.
- Between 2003 and 2011, global pollution sharply increased due to industrial growth in developing regions. This was followed by a period of stability and then a gradual decline after 2015, driven mainly by aggressive clean-air policies in China and improvements in Europe and North America.
- While the downward trend is encouraging, global PM2.5 levels still remain higher than safe health guidelines, indicating that air pollution continues to be a major environmental and public health concern worldwide.



Exploratory Question 6

When Pollution Spikes: A Volatility Analysis of AQI

Relevance
Average AQI doesn't capture pollution spikes, which are crucial for public health. Volatility analysis helps identify regions with unpredictable air quality, guiding monitoring and emergency planning, especially for sensitive populations.
Analysis Approach

Analysis Process

- Aggregation: Grouped by country to compute AQI statistics.
- Volatility: Calculated standard deviation (SD) and coefficient of variation (CV) to measure absolute and relative fluctuations.
- Ranking & Visualization: Identified and charted the top 10 most volatile countries.

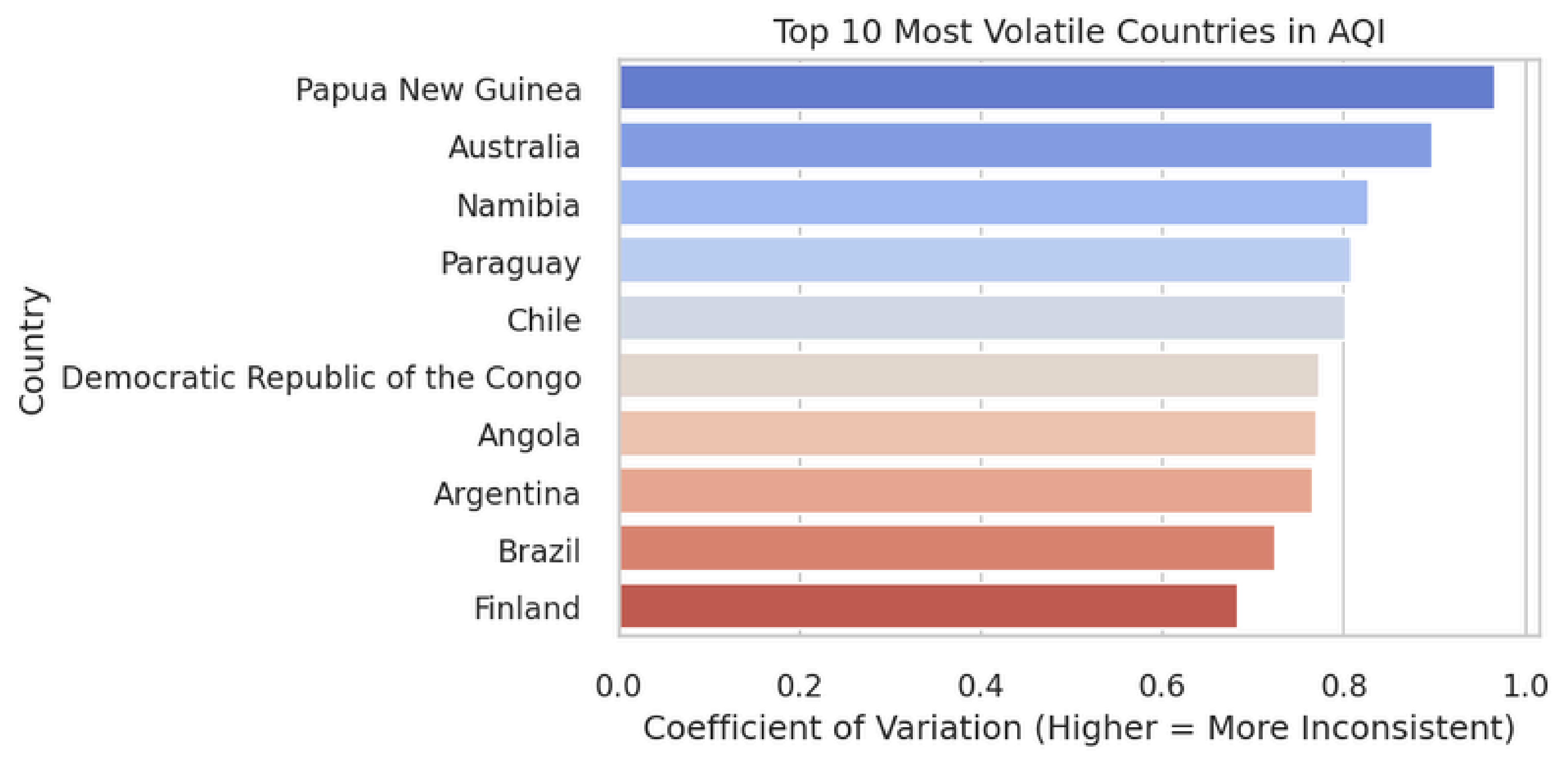
.

Exploratory Question 6

WHEN POLLUTION SPIKES: A VOLATILITY ANALYSIS OF AQI

Insights

- Papua New Guinea has the highest AQI volatility (CV ≈ 0.97).
- Australia, Namibia, Paraguay, and Chile also show large fluctuations despite moderate averages.
- Even countries with moderate mean AQI, like Finland and Brazil, experience notable swings.
- High volatility often signals episodic pollution events (wildfires, dust storms, industrial spikes).



Takeaway: Policymakers should consider AQI volatility, not just averages, to better protect public health.

Overall Findings

Conclusion – Key Insights

- Average AQI alone is not enough; volatility reveals dangerous pollution spikes.
- Countries like Papua New Guinea, Australia, and Chile show high AQI fluctuations despite moderate averages.
- Winter months have the worst air quality, while summer months are cleaner due to weather and wind patterns.
- Urbanization often increases pollution, but rural areas can also experience spikes from natural events.
- Pollution trends vary globally over time, influenced by policy, industrial activity, and climate.
- Pollutants such as PM2.5, PM10, NO2, CO, and O3 often correlate, highlighting combined pollution events.
- Socioeconomic factors like GDP per capita do not always predict cleaner air—management and regulations matter.

Team-MetricMinds