Team-MetricMinds

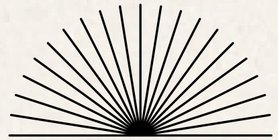
Data Mining Project

GLOBAL AIR QUALITY & POLLUTION TRENDS

{Evaluation - 2}

Presented By:

Lakshith, Sai Praneeth Sharma & Geethika



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.

Dataset Description (Exploratory Question 1)

	Country	City	AQI Value	AQI Category	CO AQI Value	CO AQI Category	Ozone AQI Value	Ozone AQI Category	NO2 AQI Value	NO2 AQI Category	PM2.5 AQI Value	PM2.5 AQI Category
0	Russian Federation	Praskoveya	51	Moderate	1	Good	36	Good	0	Good	51	Moderate
1	Brazil	Presidente Dutra	41	Good	1	Good	5	Good	1	Good	41	Good
2	Italy	Priolo Gargallo	66	Moderate	1	Good	39	Good	2	Good	66	Moderate
3	Poland	Przasnysz	34	Good	1	Good	34	Good	0	Good	20	Good
4	France	Punaauia	22	Good	0	Good	22	Good	0	Good	6	Good
...	...	...	...	...	...	...	...	...	...	...	...	...
23458	India	Gursahaiganj	184	Unhealthy	3	Good	154	Unhealthy	2	Good	184	Unhealthy
23459	France	Sceaux	50	Good	1	Good	20	Good	5	Good	50	Good
23460	India	Mormugao	50	Good	1	Good	22	Good	1	Good	50	Good
23461	United States of America	Westerville	71	Moderate	1	Good	44	Good	2	Good	71	Moderate
23462	Malaysia	Marang	70	Moderate	1	Good	38	Good	0	Good	70	Moderate

23035 rows × 12 columns

Global Air Pollution Dataset

- Contains measurements of PM2.5, PM10, CO, NO2, SO2, O3
- Geographical coverage: >100 countries
- Several measurement locations per city

Dataset Description (Exploratory Question 1)

	Country Name	Country Code	Indicator Name	Indicator Code	1960	1961	1962	1963	1964	1965	...	2016	2017	2018
0	Aruba	ABW	GDP per capita (current US$)	NY.GDP.PCAP.CD	NaN	NaN	NaN	NaN	NaN	NaN	...	27441.529662	28440.051964	30082.127645
1	Africa Eastern and Southern	AFE	GDP per capita (current US$)	NY.GDP.PCAP.CD	186.121835	186.941781	197.402402	225.440494	208.999748	226.876513	...	1329.807285	1520.212231	1538.901679
2	Afghanistan	AFG	GDP per capita (current US$)	NY.GDP.PCAP.CD	NaN	NaN	NaN	NaN	NaN	NaN	...	522.082216	525.469771	491.337221
3	Africa Western and Central	AFW	GDP per capita (current US$)	NY.GDP.PCAP.CD	121.939925	127.454189	133.827044	139.008291	148.549379	155.565216	...	1630.039447	1574.230560	1720.140280
4	Angola	AGO	GDP per capita (current US$)	NY.GDP.PCAP.CD	NaN	NaN	NaN	NaN	NaN	NaN	...	1807.952941	2437.259712	2538.591391

5 rows × 70 columns

World Bank Dataset

- GDP per capita
- Urban population percentage
- Other socioeconomic indicators

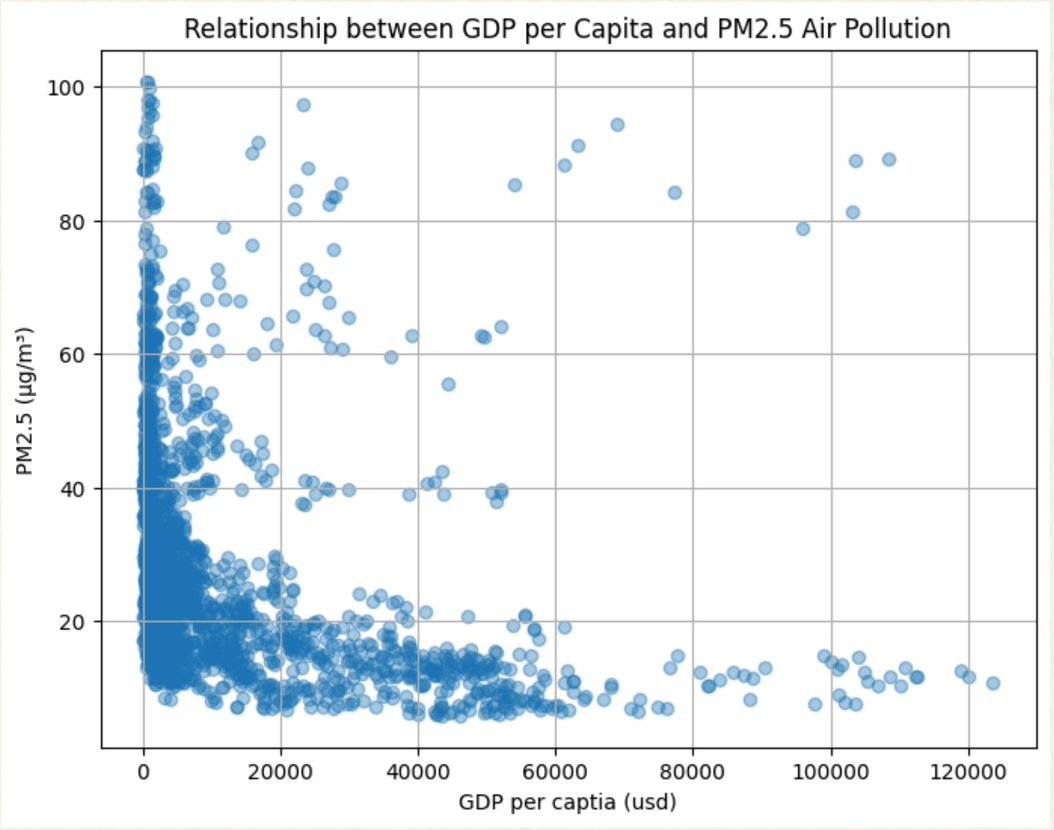
Together, these datasets offer a comprehensive view of both environmental and socioeconomic factors.

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"



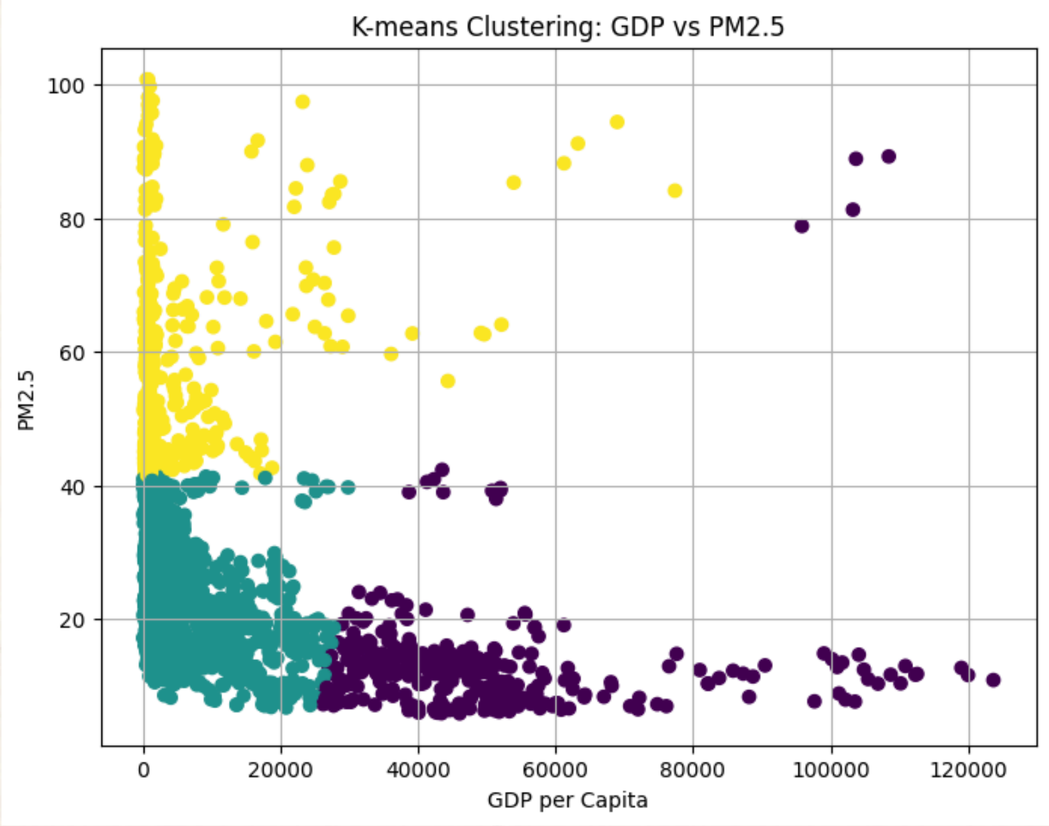
Exploratory Question 3

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.

Dataset Description (Exploratory Question 4)

	State Code	County Code	Site Num	Parameter Code	POC	Latitude	Longitude	Datum	Parameter Name	Sample Duration	...	AQI	Method Code	Method Name	Local Site Name	Address	State Name	County Name	City Name	CBSA Name	Date of Last Change
0	1	3	10	88101	3	30.497478	-87.880258	NAD83	PM2.5 - Local Conditions	1 HOUR	...	NaN	209	Met One BAM-1022 Mass Monitor w/ VSCC or TE-PM...	FAIRHOPE, Alabama	FAIRHOPE HIGH SCHOOL, 1 PIRATE DRIVE, FAIRHOPE...	Alabama	Baldwin	Fairhope	Daphne-Fairhope-Foley, AL	2024-08-06
1	1	3	10	88101	3	30.497478	-87.880258	NAD83	PM2.5 - Local Conditions	1 HOUR	...	NaN	209	Met One BAM-1022 Mass Monitor w/ VSCC or TE-PM...	FAIRHOPE, Alabama	FAIRHOPE HIGH SCHOOL, 1 PIRATE DRIVE, FAIRHOPE...	Alabama	Baldwin	Fairhope	Daphne-Fairhope-Foley, AL	2024-08-06
2	1	3	10	88101	3	30.497478	-87.880258	NAD83	PM2.5 - Local Conditions	1 HOUR	...	NaN	209	Met One BAM-1022 Mass Monitor w/ VSCC or TE-PM...	FAIRHOPE, Alabama	FAIRHOPE HIGH SCHOOL, 1 PIRATE DRIVE, FAIRHOPE...	Alabama	Baldwin	Fairhope	Daphne-Fairhope-Foley, AL	2024-08-06
3	1	3	10	88101	3	30.497478	-87.880258	NAD83	PM2.5 - Local Conditions	1 HOUR	...	NaN	209	Met One BAM-1022 Mass Monitor w/ VSCC or TE-PM...	FAIRHOPE, Alabama	FAIRHOPE HIGH SCHOOL, 1 PIRATE DRIVE, FAIRHOPE...	Alabama	Baldwin	Fairhope	Daphne-Fairhope-Foley, AL	2024-08-06
4	1	3	10	88101	3	30.497478	-87.880258	NAD83	PM2.5 - Local Conditions	1 HOUR	...	NaN	209	Met One BAM-1022 Mass Monitor w/ VSCC or TE-PM...	FAIRHOPE, Alabama	FAIRHOPE HIGH SCHOOL, 1 PIRATE DRIVE, FAIRHOPE...	Alabama	Baldwin	Fairhope	Daphne-Fairhope-Foley, AL	2024-08-06

5 rows × 29 columns

US EPA PM2.5 Monitoring Data

- Contains hourly PM2.5 measurements from U.S. EPA air quality monitoring stations.
- Includes detailed location and site metadata such as state, county, latitude, longitude, and local site name.
- Provides pollutant information, monitoring method details, and associated air quality indicators (e.g., AQI).

Exploratory Question 4

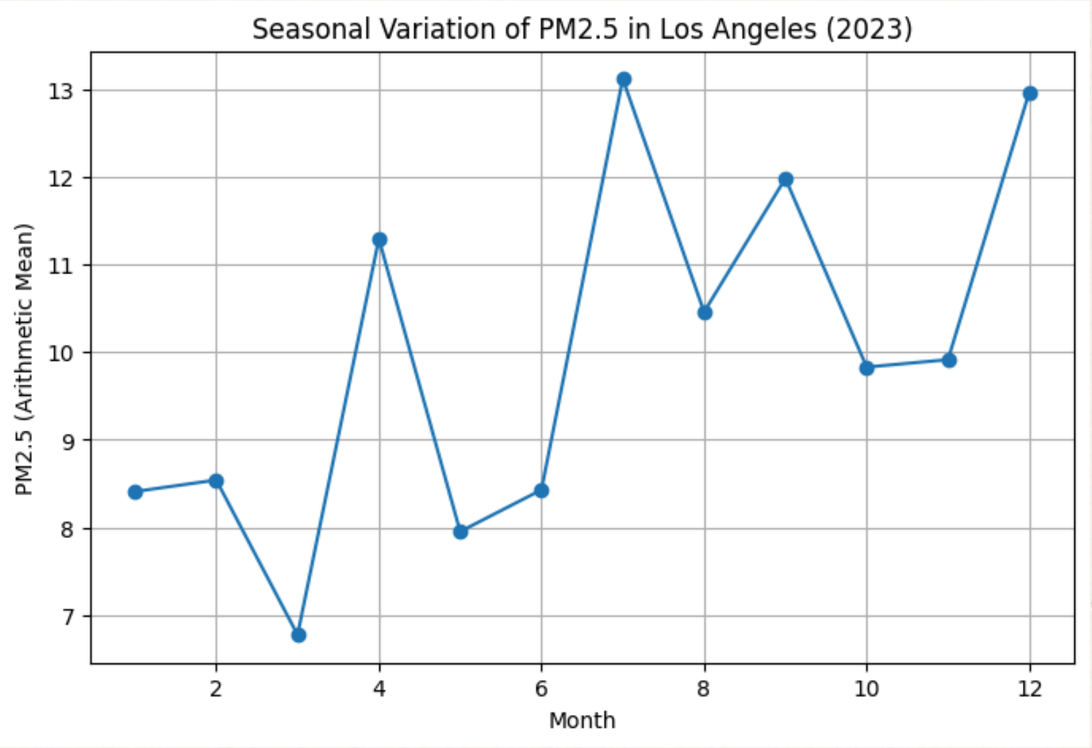
ARE THERE SEASONAL VARIATIONS IN POLLUTION (WINTER VS SUMMER)?

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).

Dataset Description (Exploratory Question 5)

Entity	Year	pm25_exposure
Afghanistan	1990	65.48679352
Afghanistan	1995	65.17951202
Afghanistan	2000	64.92246246
Afghanistan	2005	64.52268219
Afghanistan	2010	65.24559021
Afghanistan	2011	66.83572388
Afghanistan	2012	66.02317047
Afghanistan	2013	61.366745
Afghanistan	2014	59.0103302
Afghanistan	2015	61.25265503
Afghanistan	2016	56.28704834
Afghanistan	2017	56.91080856
Albania	1990	22.71866417
Albania	1995	22.02921867
Albania	2000	21.99824905

PM25- Exposure-Worldbank Dataset

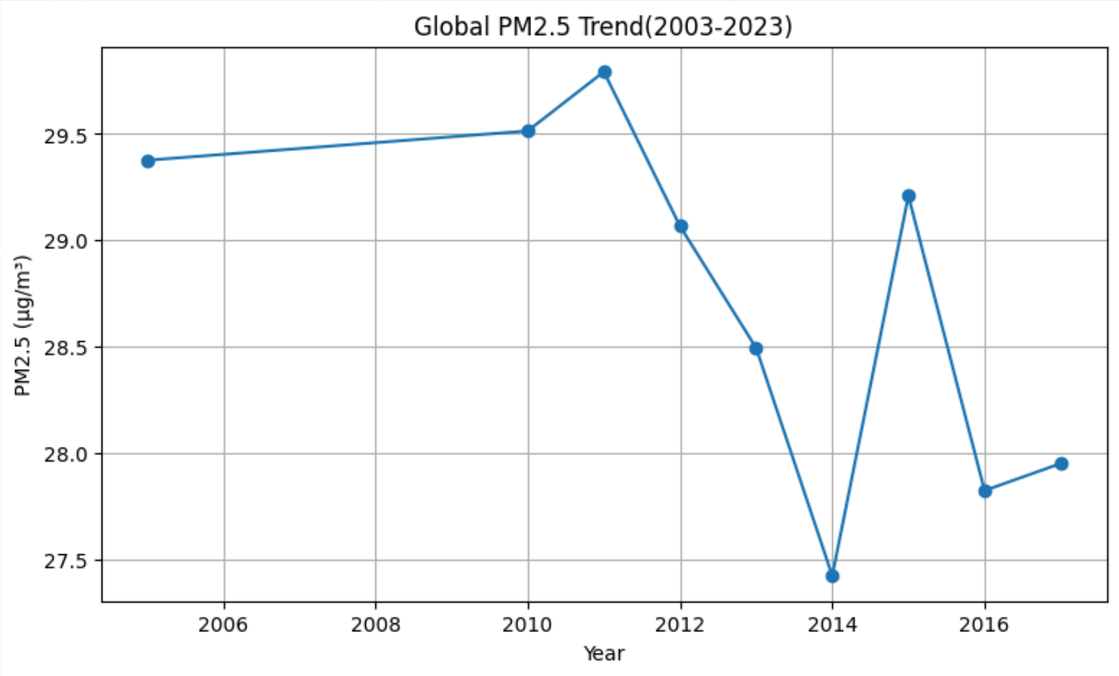
- Contains country-level PM2.5 exposure data from the World Bank
- Includes annual values for each country
- PM2.5 exposure measured as population-weighted average concentration ($\mu g/m^3$)
- Helps analyze air pollution trends over time
- Useful for cross-country comparisons of air quality
- Supports research on public health and environmental policies

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.



Key Insights

01 Winter months have the worst air quality, while summer months are cleaner due to weather and wind patterns.

02 Pollution trends vary globally over time, influenced by policy, industrial activity, and climate.

03 Socioeconomic factors like GDP per capita do not always predict cleaner air—management and regulations matter.

Team - MetricMinds