

Mission Briefing: Operation Clean Air

Background: Biohacking entrepreneur walks out due to air pollution

In early February 2025, American entrepreneur Bryan Johnson visited India to participate in a podcast with Zerodha co-founder Nikhil Kamath. During the recording in Mumbai, Johnson experienced significant discomfort due to the city's poor air quality, reporting symptoms such as skin rashes and irritation in his eyes and throat. Despite wearing an N95 mask and using an air purifier, he found the indoor Air Quality Index (AQI) to be 130, with PM2.5 levels at $75 \mu\text{g}/\text{m}^3$ —equivalent to smoking 3.4 cigarettes over 24 hours. Consequently, Johnson decided to leave the podcast early.

Following this incident, Johnson took to social media to highlight the severe health risks associated with India's air pollution. He referenced studies indicating that prolonged exposure to PM2.5 pollutants can lead to liver inflammation, blood fat imbalances, and gene dysregulation linked to cancer. He emphasized that addressing air quality issues in India could significantly enhance public health, potentially more so than curing all cancers.

Nithin Kamath, co-founder of Zerodha, acknowledged the widespread nature of air quality problems in India, noting that cities beyond Delhi, such as Bengaluru, also experience high AQI levels. He suggested that property prices should reflect air quality levels and urged researchers to explore this topic further.

This incident underscores the pressing need to address air pollution in India, as it poses significant health risks to residents and visitors alike.

Read more here:

[X](#)

[Business News India](#)

[mint](#)

Mission Impossible

~~im~~possible

Your Mission, Should You Choose to Accept It:

You are doing the ES 114 course at IITGN. The country needs you! The Central Pollution Control Board (CPCB) tasks with a critical assignment. Recent events have highlighted the pervasive issue of air pollution across India. Notably, during a podcast recording in Mumbai's Bandra area, the Air Quality Index (AQI) was alarmingly high, exceeding 160. This incident debunked the myth that severe air quality problems are confined to Delhi during winter months. In light of these revelations, Nithin Kamath, co-founder of Zerodha, proposed linking property prices to AQI levels, suggesting that areas with poorer air quality should have lower real estate values.

Your objective is to analyze the provided datasets to uncover insights that could inform policies aimed at improving air quality. This mission is of utmost importance, as it could influence future strategies to combat air pollution across the nation.

With a **submission deadline of February 18th 11:00 PM** fast approaching, time is critical. Your task is clear: Work with your team, analyze the data, extract meaningful questions, and deliver the insights that could contribute to improving his work on air quality. The secret team has provided you with the below dataset.

Remember: Lives are at stake!

This operation is classified. Precision and efficiency are paramount.

Tools at Your Disposal:

To successfully complete this mission, you'll need to employ a suite of powerful tools:

- **NumPy**: Your go-to for numerical computations, enabling efficient array operations and mathematical functions.
- **Pandas**: The data manipulation expert, allowing you to handle and analyze structured data with ease.
- **Matplotlib & Seaborn**: Your visualization specialists, ready to help you create insightful plots and charts to communicate your findings effectively.
- **SciPy**: The scientific computing powerhouse, offering advanced statistical analysis capabilities.
- **Jupyter Notebook**: Your mission control center, providing an interactive environment to write and execute your code, document your process, and present your results.

These tools are essential for data analysis and visualization in Python, enabling you to process and interpret complex datasets efficiently.

Security Protocols:

- **Private GitHub Repository** – Store mission files securely. No leaks. No compromises.
- **Slack HQ** – Have questions? Drop them in the dedicated Slack channel. Help is always a ping away.

Assignment starts here

Understanding Dataset.

The folder Link provided below has 3 CSV files:

- **Data.csv:** This CSV file contains information about the Air Quality Station, state, pollutant measurement, and a few other things about station location
- **State_data.csv:** This CSV file contains states, their area, and population
- **NCAP_Funding.csv:** This CSV file contains city, state, and funds received by them for different years

Dataset link

[Dataset](#)

Questions

Spatial aggregation. [4 Marks]

1. Which state (consider all stations in that state) has the highest average PM2.5 concentration across all stations and across all years?
2. Which state (consider all stations in that state) had the most days with hazardous PM2.5 levels (above $300 \mu\text{g}/\text{m}^3$) for the year 2023?
3. Which state has the highest variability in PM2.5 levels across its monitoring stations in 2023? (Think how would you measure variability)
4. Which state has the lowest average PM2.5 levels during the Covid period (The year 2020-2021 both years inclusive)?

Temporal Aggregation [3 Marks]

1. Report the station that recorded the highest value of PM 2.5 for the month Aug of 2020.
2. For the year 2018, calculate the average PM2.5 concentration for each season (Seasons being Winter: December-February, Summer: March-May, and Monsoon: June-September) for stations **Lal Bahadur Shastri Nagar, Kalaburagi - KSPCB**. Identifying the season with the highest pollution.
[Thought Question] Can you suggest some potential factors on why certain seasons are generally more polluted than others?
3. Compare the average PM2.5 levels during weekends and weekdays for the station **Lal Bahadur Shastri Nagar, Kalaburagi - KSPCB** over the year 2021. Create a plot to show these trends Steps:
 - a. Calculate the **monthly average PM2.5** for weekdays and weekends separately
 - b. **X-axis:** Months (January to December).
 - c. **Y-axis:** Average PM2.5 levels.
 - d. Use **two lines**: one for the monthly average PM2.5 levels on **weekdays** and one for **weekends**.

Spatio-temporal aggregation [3 marks]

1. For each state, calculate the percentage increase or decrease in PM2.5 levels during monsoon months (June to September) compared to summer months (March to May). Which state has the most difference (increase or decrease) for the year 2022?
2. Plot the average PM2.5 levels for Delhi across each season (Winter: December-February, Summer: March-May, Monsoon: June-September) from 2017 to 2023. What seasonal variation, if any, can you observe in pollution levels?

3. Plot a time-series graph comparing Delhi's PM2.5 levels with Mumbai's from 2017 to 2023. Which city had more fluctuations? (Hint: Think what fluctuation means)

Population-Based [3 Marks]

1. Which state in India has the highest number of monitoring stations relative to its population?
2. Plot a bar chart showing the top 5 most polluted states in terms of per capita PM2.5 exposure in 2023.
3. Plot the relationship between population density and average PM2.5 concentration for each state. What conclusion can you draw about the correlation between population density and pollution?
 - a. **X-axis:** Population density (people per square kilometer).
 - b. **Y-axis:** Average PM2.5 concentration.
 - c. Use a **scatter plot** to show the relationship, where each point represents a state.

Area Based [3 Marks]

1. Which state has the highest PM2.5 concentration per square kilometer? Plot the pollution levels for all states relative to their area.
 - a. **X-axis:** States.
 - b. **Y-axis:** PM2.5 concentration per square kilometer.
 - c. Use a **bar plot** where each bar represents a state, and the height shows PM2.5 per sq km.
2. Compare the total number of monitoring stations per square kilometer in each state. Identify the state with the highest density of monitoring stations relative to its area. Show a bar plot where each bar represents a state, and the height shows the number of monitoring stations per square kilometer.
3. Compare the average PM2.5 levels between Maharashtra and Madhya Pradesh over the year 2021. Given that both states have similar area sizes but different population densities, analyze how population density impacts pollution levels in these two states.

Funding Based [3 Marks]

1. Compare the average PM2.5 levels between states that received NCAP funding and those that did not in 2021.
2. Create a time series plot showing PM2.5 levels with an increment in NCAP funding for Assam. Has an increment of the financing led to noticeable improvements in air quality?

3. Create a scatter plot to showcase the area versus funding received by a state. Mark every state with a different color and show the legend outside the plot.

Miscellaneous Questions [5 Marks]

1. For the most polluted city in 2023 (highest average pollution in 2023), calculate the percentage of days with PM2.5 levels above the hazardous threshold ($300 \mu\text{g}/\text{m}^3$)[**1 Mark**]
2. Analyze the PM2.5 levels in Delhi during Diwali (October-November) and New Year compared to regular months over the 3 years (2018-2020). Do you observe any significant pollution spikes during these festivals? If spikes are observed, Can you mark these spikes using arrows?[**1.5 Mark**]
3. Identify the state that received the highest NCAP funding in 2020 and determine if its air pollution levels improved significantly by 2023. Use a line graph to compare yearly average PM2.5 trends.[**1 Mark**]
4. Draw the map of India and plot the sensor locations on the map.[**1.5 Mark**]

Tell us something new [1 Mark]

1. Tell us something interesting from the dataset that we have not asked. Anything that intrigues you, anything that might be useful to the policymakers.

Assignment ends here

Acknowledgment:

Your efforts will significantly contribute to ongoing research and policy formulation, potentially leading to cleaner air and healthier living conditions for millions.

This message will self-destruct in five seconds.

.....

.....

.....

.....

.....

....

...

..

.

Well not really!

More reading

1. [Vartalaap: What Drives #AirQuality Discussions: Politics, Pollution or Pseudo-science?](#)

2. [Samachar: Print News Media on Air Pollution in India](#)
3. [VayuBuddy: an LLM-Powered Chatbot to Democratize Air Quality Insights](#)