No Urban Air Quality and Citizen Health: An Exploratory Analysis

Briefing for the Mayor of London Attractive Cities Network -Thematic Lead Analysis

Candidate No: TKJL0

Executive Summary

The report carries a detailed exploratory analysis of air quality and health data from various global cities from 1994 to 2023. Cities like Beijing in Asia, have consistently shown the highest PM2.5 levels. Despite similar pollution levels, cities with high GDP tend to show better health outcomes than those with low GDP. Therefore, the Mayors should be cautious about data quality issues, including missing values, outliers, and the fact that city averages hide within-city inequalities.

Part 1: Exploratory Data Analysis

1a. Key Trends, Patterns and Anomalies

H1: Cities with higher PM2.5 concentrations are associated with poorer population health outcomes, including lower life expectancy, higher child mortality, and lower Health Index scores.

H2: Economic development, proxied by GDP per capita, strongly conditions the relationship between air quality and health, such that higher-GDP cities achieve better health outcomes at similar levels of pollution.

H3: Regional context explains variation in health outcomes beyond air quality alone, reflecting differences in healthcare access, governance capacity, and broader socioeconomic conditions.

Which cities are consistently most polluted? Beijing, Mumbai, Mexico, and Los Angeles are the most consistently polluted cities, with mean PM2.5 concentrations of 30.0 µg/m³, 16.1 µg/m³, 15.3 µg/m³, 14.5 µg/m³, and 10.6 µg/m³, respectively.

PM2.5 remained stable until 2009 peaked around 2011 and declined again by 2023.

The above declared hypotheses are tested and explored through descriptive statistics, data visualization, correlation methods, and exploratory regression.

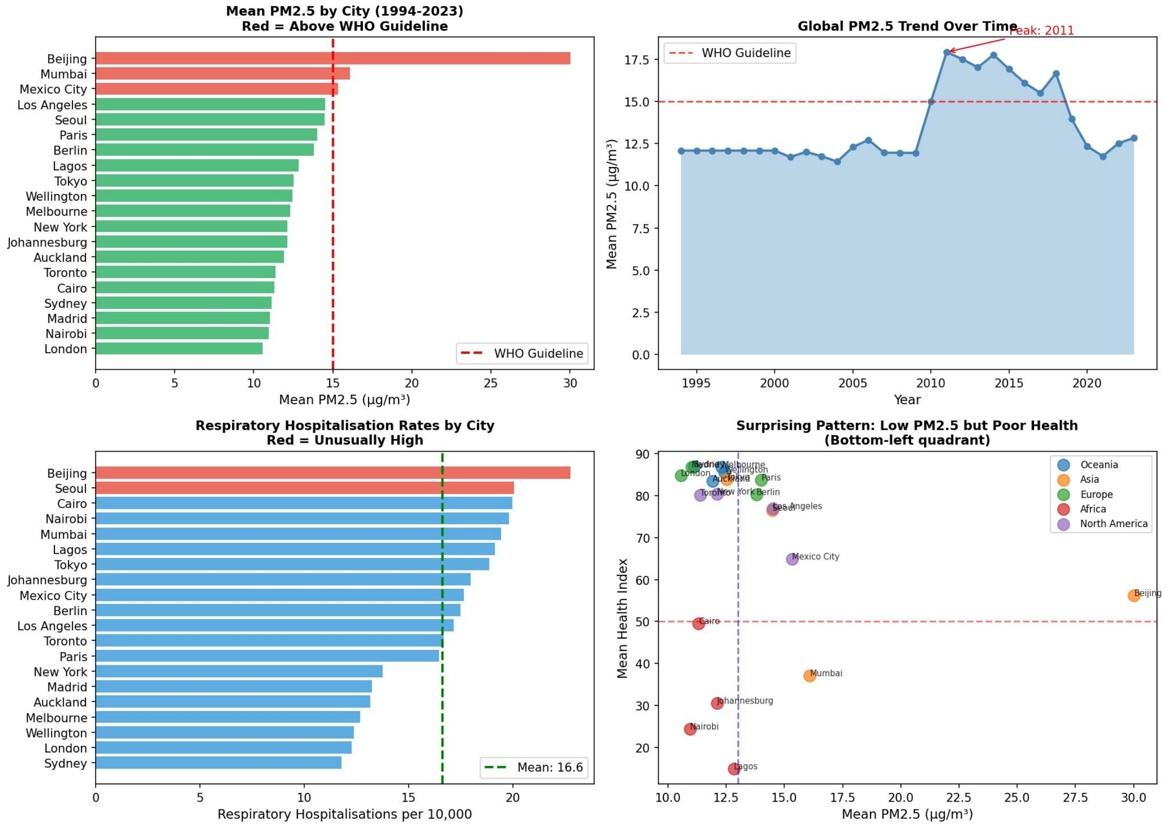
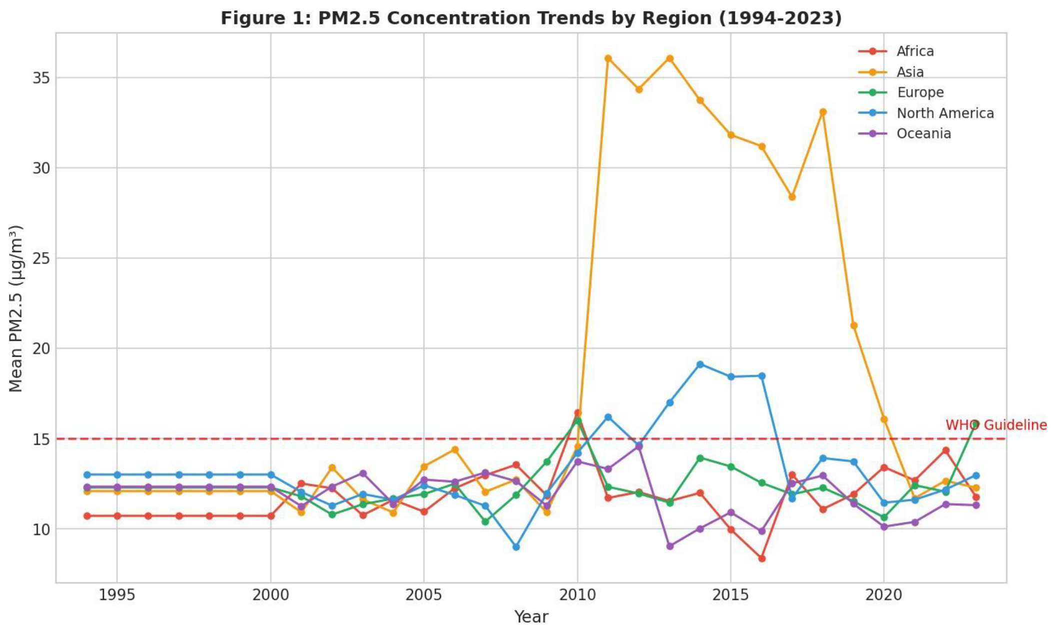


Figure 1: PM2.5 patterns, trends, respiratory rates, and health-pollution relationships

Surprising patterns: African cities: Johannesburg, Lagos, and Nairobi, show poor health outcomes despite moderate PM2.5 levels. Whereas high respiratory hospitalisation rates are recorded in Beijing and Seoul (22.7 and 20.1 per 10,000, respectively. With Beijing reaching 101.6 µg/m³ in 2013, which is seven times the WHO guideline.



Coverage: 600 observations across 20 cities.

Regional patterns: Asia showed the highest and most volatile PM2.5 levels, while Europe and Oceania showed the lowest.

Health trends: Health index increased by 21%, and life expectancy increased from 72.5 years to 77 years.

Anomalies: 39 PM2.5 outliers (6.5%) were concentrated in Los Angeles and Mexico because of pollution or measurement issues.

1b. Interrelationships Between Air Quality and Health

PM2.5 and Life Expectancy: Low life expectancy is experienced in cities with higher PM2.5. $r = -0.06$ showing a negative association.

PM2.5 and Child Mortality: Cities with higher PM2.5 tend to have higher child mortality. $r = 0.08$ showing a positive association.

PM2.5 and Health Index: Strongest association with $r = -0.16$ and $p < 0.001$

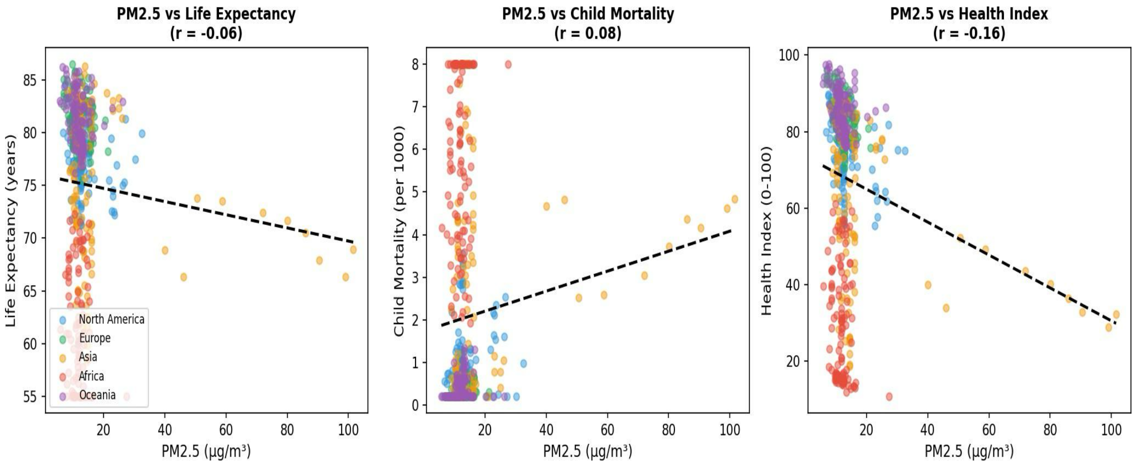
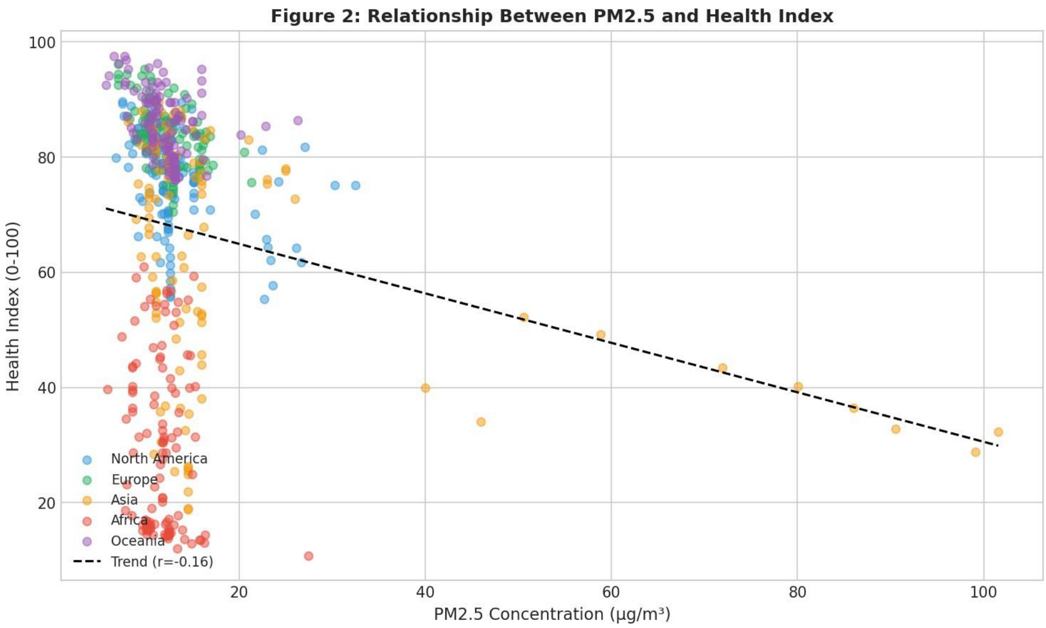


Figure 2: Associations between PM2.5 and health outcomes

Important: These are associations, not causal relationships, controlling for GDP, R^2 increases from 0.004 to 0.608, indicating economic development explains most health variation.



Air-quality health relationships

Unadjusted relationship: Weak and insignificant as PM2.5 vs life expectancy r = -0.06 and Health Index r = - 0.16).

Adjusted model: Controlling for GDP, PM2.5 is statistically associated with : → −0.11 years of life expectancy ($p < 0.001$).

Explanatory power: suggesting GDP dominates health outcomes.

1c. Differences by City Characteristics

High-GDP vs Low-GDP Cities: Cities with high-GDP have a Health Index of 81.8, and cities with low-GDP have it for 54.5. The life expectancy in high-GDP cities is 80 years, and in low-GDP cities, it is 70.2 years. Lastly, child mortality is six times higher in cities with low-GDP.

Key finding: At similar PM2.5 levels, high-GDP cities have a Health Index of 81.6 while low-GDP cities have 57.5, suggesting that economic development matters more than pollution.

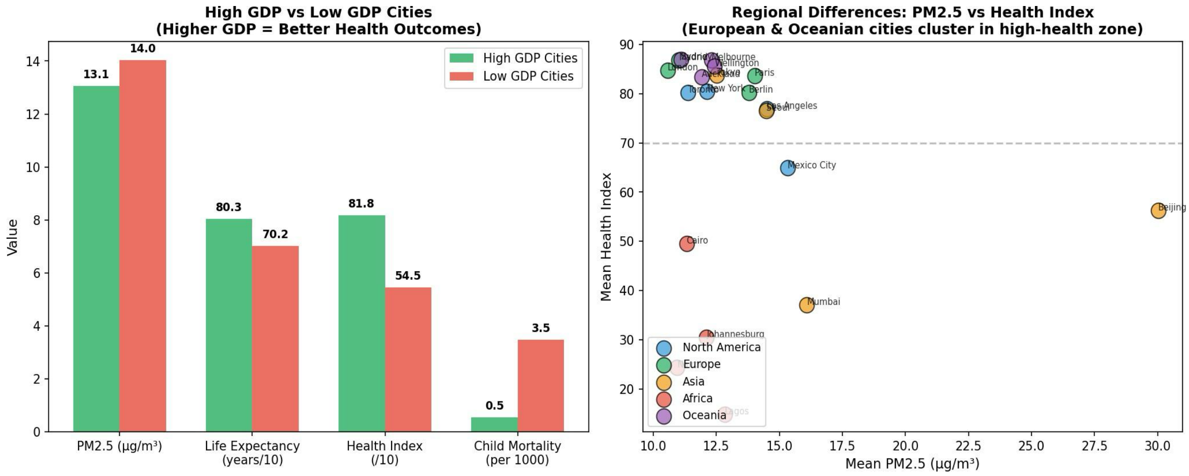
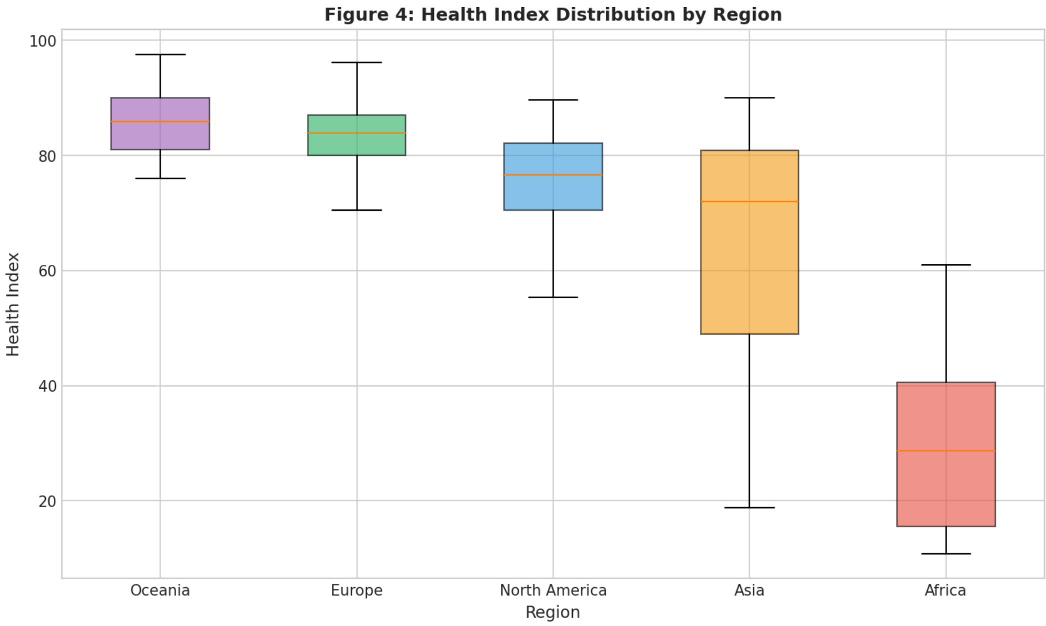


Figure 3: Health outcomes by GDP and regional clustering

Regional patterns: Europe and Oceania fall within the high-health zone, with a Health Index between 84 and 86. African cities have a poor Health Index of 29.8 despite a low PM2.5 level of 11.8 μg/m³.

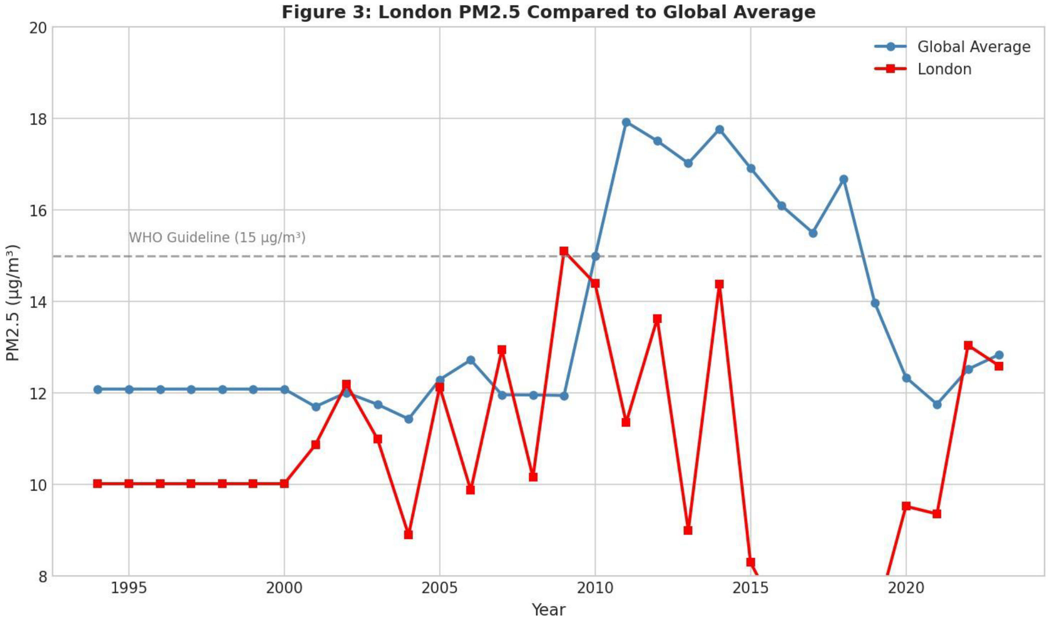


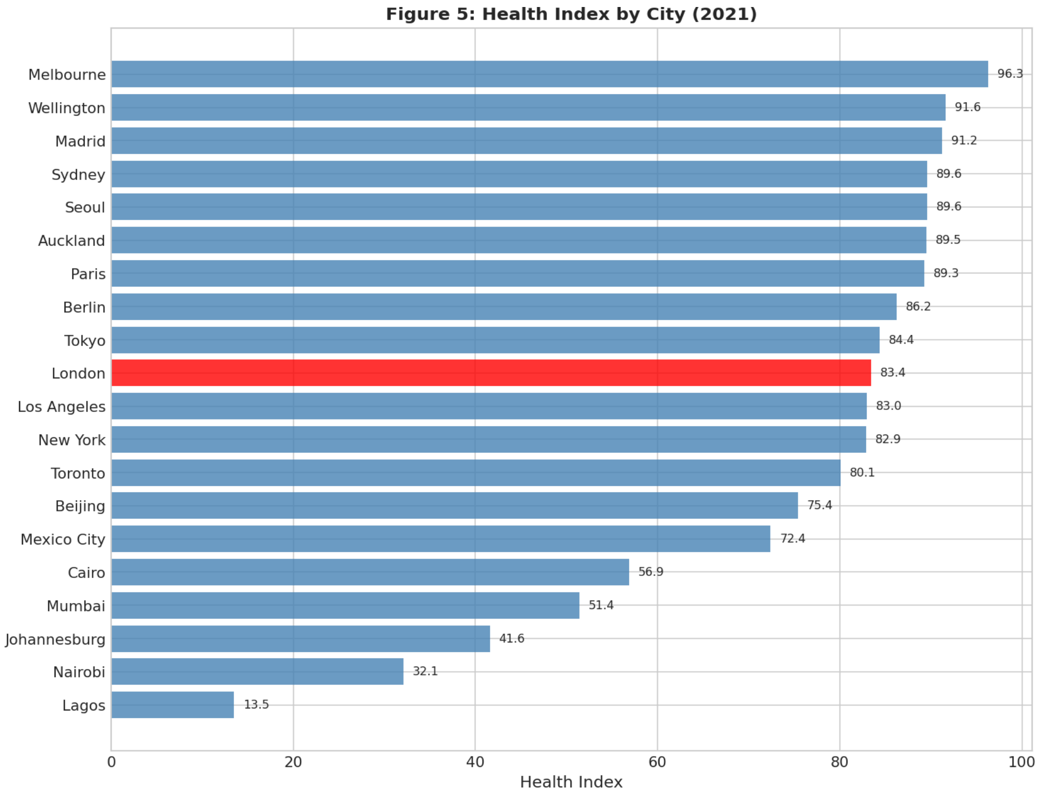
Regional disparities: Health Index is highest in Oceania and Europe at 85.7 and 83.9, respectively, and lowest in Africa at 29.8.

Key insight: Africa has low PM2.5 but poor health, suggesting that air quality is not the sole factor.

1d. London's Position

London had a PM2.5 of 9.4 µg/m³ in 2021





Air quality: Consistently low PM2.5 of 9.4 µg/m³ in 2021, below the global average

Health ranking: Health Index of 83.4, below Madrid and Paris, suggesting room for improvement.

This supports H2 as London's strong air quality has not delivered top health outcomes and H3 as its health performance reflects European regional patters rather than pollution levels.

1e. Data Problems and Uncertainty

Data Gaps: 20% of non-motorised modal split and Health Index and Life Expectancy data for 2022-2023 are missing, limiting assessment of recent trends. Claims about health status are based on 2021 data.

Outliers and Extreme Values: I identified 39 PM2.5 outliers concentrated in Beijing and Mexico using the IQR method. Beijing's 2013 value 101.6 µg/m³, is exceptionally extreme.

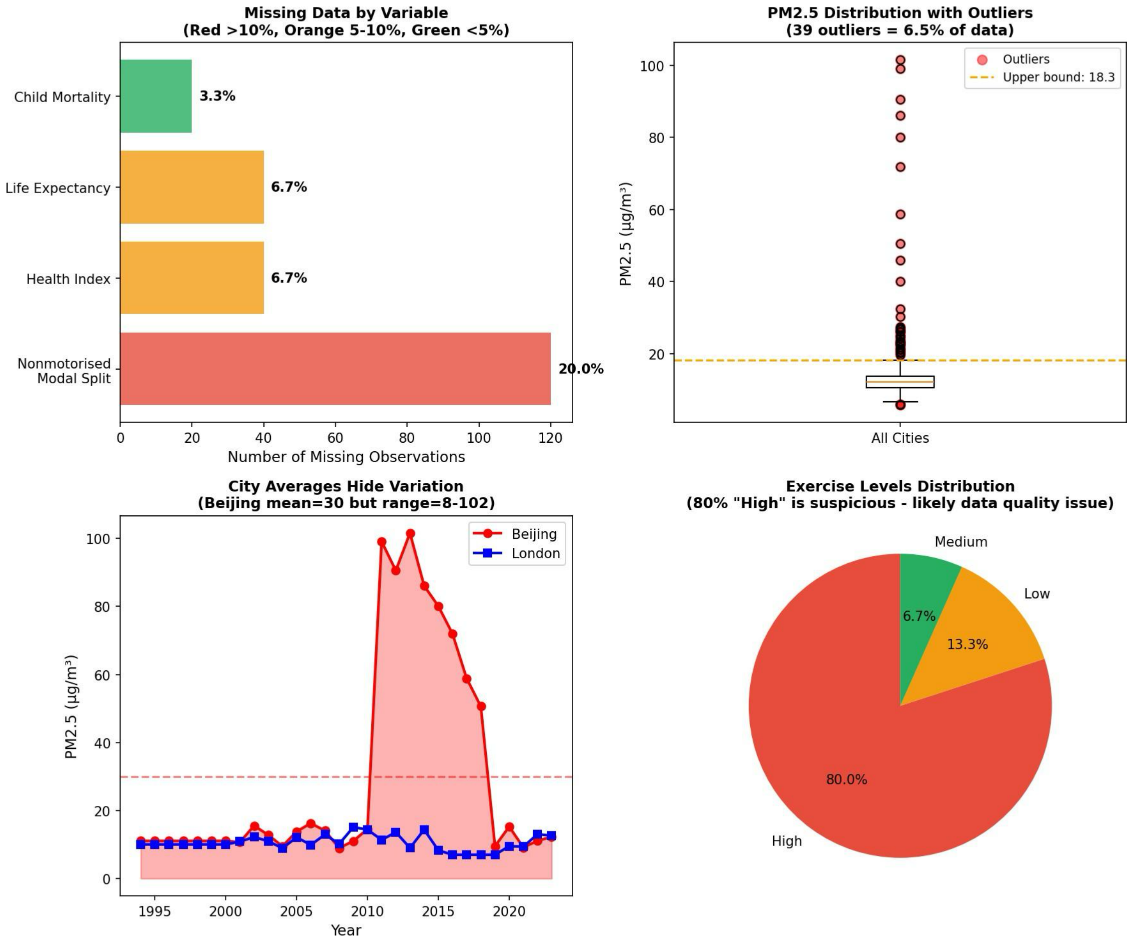


Figure 6: Data quality issues—missing values, outliers, variation hidden by averages, and suspicious distributions

City Averages Hide Inequalities: The data indicates average levels city-by-city. Still, pollution varies widely within cities as Beijing's mean PM2.5 is 30 µg/m³, but values range from 9 to 102 µg/m³. Poor neighborhoods are more vulnerable to higher levels of pollution due to their proximity to roads and industrial sites, and the totalizing view of statistics cannot reveal this micro picture.

Measurement Inconsistencies: Different cities use different monitoring and measuring protocols and reporting standards, making comparisons uncertain.

Data is Simplified/Fictional: The data simplifies complex realities and excludes health determinants.

What the Mayor should be cautious about: The major should not that results are associative, recent trends are uncertain due to missing data, rankings are sensitive to measurement differences and city averages hide inequalities.

Key Recommendations

I observed a weak relationship between PM2.5 and Health (H1), while GDP and regional context explain most variation.

Observed London's good air hasn't led to top health records, supporting H2 (GDP is more important than PM2.5 alone) and it also supports H3 (health aligns with Europe, not air quality).Based on this analysis, I recommend that the mayor:

(1) Policy development as approaches must account for different baseline conditions across cities including local diversities.

(2) Address development factors, as high-GDP cities tend to have better health regardless of pollution.

(3) Commission improved data collection with standardized measurement protocols across cities to conduct more robust comparative analysis.

(4) Commission primary data collection with standardized protocols.

(5) Require intra-city data to identify vulnerable populations.

(6) Consider findings as exploratory, as real policy needs more substantial evidence.

Confidence Assessments: Regional patterns have high confidence; London's comparative position has medium confidence; casual claims and exercise-level relationships have low confidence due to data limitations.

Part 2: Informing Policy Decisions

Q2. Interpreting Probabilistic Analysis

(a) Relative Uncertainty: Cost modelling, through Monte Carlo simulation, reveals a key difference in unpredictability. Two options are available: Option A) (Monitoring Network) has a narrow cost distribution (£48.1 M – £137M). Option B (Clean air Zone) has a much higher cost (£30.7M to £186.1M), indicating costs are less certain and carries higher risk.

(b) Recommendations: I recommend pursuing option A due to lower risk rates. My part 1 analysis showed that London already has excellent air quality (PM2.5 = 9.4 µg/m³, the lowest globally), so the marginal benefit of aggressive emissions may be limited. Choice A's monitoring infrastructure would help the network understand city-specific outcomes. While Choice B has a lower mean cost (£60.2M vs £75M), costs exceed £120M.

(c) Caveats: Caveats: Here are a few caveats for mayors to understand that predictions, which left spaces for uncertainty. Part 1 showed data gaps and inconsistent city-level measurement. Cost-benefit ignores health outcomes. My analysis found that African cities have poor health despite moderate air quality, monitoring alone may miss critical health determinants; (iii) implementation risks and behavioral responses are challenging to model; and (iv) the distributions assume independence between cost factors, which may not hold in practice.

Q3. Additional Analysis

(a) Before final judgement, I would suggest that the Mayor commission a cost-effectiveness analysis that considers health outcomes per pound spent. My Part 1 revealed that the relationship between air quality and health is weak ($r = -0.06$ to -0.16) and heavily confounded by GDP. The Monte Carlo analysis addresses cost uncertainty but not effectiveness. My analysis showed that higher city GDP correlates with a 27-point higher Health Index than low-GDP cities, even at similar pollution levels. Interventions targeting broader determinants of health may be more cost-effective than air quality measures alone. Considering the cost per QALY (quality-adjusted life year) gained would enable a more informed picture, given in my finding that London's Health Index (83.4) lags behind Madrid (91.2) and Paris (89.3) despite excellent air quality.

Q4. Multi-Criteria Analysis Advice

(a) Should she use MCA? Yes, MCA is the most methodologically sound choice. This means disturbing £5M across six competing priorities with different beneficiaries, evidence bases, and time zones. My foundational (part 1) analysis addresses relationship complexity: air quality alone explains little health variation, regional contexts differ dramatically (Health Index ranges from 29.8 in Africa to 85.7 in Oceania), and city averages hide within-city inequalities. MCA's strengths include: making these trade-offs explicit, enabling structured comparison of incommensurable outcomes, and providing transparency for stakeholders.

(b) Three Critical Design Choices: The Mayor must create a mutually inclusive team in scoring and weighting. My Part 1 analysis revealed that the city-by-city expertise requirement differs. African cities face different health challenges than European cities, despite similar air quality. I recommend including: public health officials, environmental scientists, transport planners, community representatives from high-exposure areas, and economists. This balanced approach will mitigate the dual risk: technical bias and unwieldy process

Criteria Weighting Process: Process inclusion for weighted fundamentally alters outcomes. My analysis shows exclusive air quality metrics would miss variable, as African cities have poor health despite low pollution. Thus, the Mayor should use a deliberative process where stakeholders are free to discuss and dissent on weights. So these weights can account for distributional impact, which I elaborated in part 1 of my analysis.

Stakeholder Inclusion: The Mayor must decide which London communities participate. As aggregate data in my Part 1 finding hides within-city inequalities, it is directly relevant here. Economically weaker neighborhoods typically experience higher pollution, so, the process must ensure the inclusion of these communities in the MCA process with a proactive inclusion of historically underrepresented groups.

Q5. Critical Reflexivity

(a) What my choices privileged and obscured: Part 1 analysis centred around quantitative, aggregate relationships, focusing on correlations, R-squared values, and city-level incidents that structure this report. This methodological choice effectively highlighted broad trends but inherently obscured critical macro- and micro-scale variations.
My regression approach assumed linear relationships. The choice to control for GDP treated economic development as a confounder to be 'removed' rather than as deeply intertwined with pollution governance and health behaviours. Community-level knowledge and lived experiences of pollution were excluded from my quantitative framework.

(b) Whose knowledge matters: More reliance on city-level data obscured several dimensions, like the fact that African cities have poor health despite moderate air quality. The factors like this suggested I couldn't study the air quality of inhabited spaces and residential spaces, which can be a major concern than PM2.5. For a diverse process, I suggest compilation of qualitative case studies, participatory mapping exercises, ethnographic research, and holding deliberative workshops.

(c) My recommendation is shaped by my role as an analyst, where I focus on cost certainty and what the data can show with confidence. This makes me cautious about options with higher uncertainty. Acknowledging this matters, as the analysis reflects my perspective and should sit alongside political judgement and local experience.

Q6. Scenarios for Deep Uncertainty

This scenarios analysis builds on Part 1 findings: H1 identifies a weak PM2.5- health relationship, H2 shows that GDP explains most health variation and H3 highlights strong regional differences. These scenarios below explore how these relationships may evolve under future uncertainty.

(A)What are scenarios?

Scenarios are structured, plausible narratives of alternative futures under different conditions, used to explore deep uncertainty where probabilistic forecasts are not meaningful. We cannot meaningfully say climate change has a ‘60% chance’ of causing specific air quality changes. Scenarios enable decision-makers to think through the implications of fundamentally alternative realities rather than optimizing for a single predicted outcome.

(B) How scenarios could help:

For air monitoring network, I would recommend developing 3-4 scenarios dealing with alternative uncertainties:(i) Green Transition Accelerates the rapid EV adoption and clean energy reduce transport emissions faster than expected, but new pollutant can emerge; (ii) Climate Disruption: extreme weather events can spike pollution levels, overwhelming monitoring systems; (iii) Divergent Pathways: some cities relatively achieve improvements while others face declining capacity, creating network tensions; (iv) Technological Disruption: new mobility patterns fundamentally alter urban emissions rates.

A Mayor could assess whether proposed policies remain valuable across alternative futures using scenarios and develop a contingency blueprint for scenario-specific challenges

(C) Limitations:

Scenarios carry significant methodological constraints. They are tilted towards dominant worldviews assuming cognitive bias of designers. They require significant investment in facilitation and stakeholder engagement to be useful. In the network context specifically, the challenge is developing scenarios that are meaningful across diverse cities while avoiding homogeneous logic like Western-centric assumptions about progress.

References

Dataset: aqhealthcities.csv (provided for assignment)

WHO (2021). WHO global air quality guidelines. World Health Organization.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist’s companion*. Princeton University Press. https://press.princeton.edu/books/hardcover/9780691120355/mostly-harmless-econometrics

Diez Roux, A. V. (2001). Investigating neighborhood and area effects on health. *American Journal of Public Health, 91*(11), 1783–1789. https://doi.org/10.2105/AJPH.91.11.1783

Marmot, M. (2005). Social determinants of health inequalities. *The Lancet, 365*(9464), 1099–1104. https://doi.org/10.1016/S0140-6736(05)71146-6

Morgan, M. G., & Henrion, M. (1990). *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge University Press. https://www.cambridge.org/core/books/uncertainty

Pope, C. A., & Dockery, D. W. (2006). Health effects of fine particulate air pollution: Lines that connect. *Journal of the Air & Waste Management Association, 56*(6), 709–742. https://doi.org/10.1080/10473289.2006.10464485

Rydin, Y., Bleahu, A., Davies, M., Dávila, J. D., Friel, S., De Grandis, G., Groce, N., Hallal, P. C., Hamilton, I., Howden-Chapman, P., Lai, K.-M., Lim, C.-J., Martins, J., Osrin, D., Ridley, I., Scott, I., Taylor, M., Wilkinson, P., & Wilson, J. (2012). Shaping cities for health: Complexity and the planning of urban environments. *The Lancet, 379*(9831), 2079–2108. https://doi.org/10.1016/S0140-6736(12)60435-8