Public Policy Data Analysis – Air Quality & Health in Pakistan

OBJECTIVE

This project investigates the interconnection between air quality levels and public health outcomes across major cities of Pakistan. Using data from 2019–2020, the analysis focuses on PM2.5 concentrations and their relationship with respiratory cases, assessing the impact of environmental policies and identifying key urban areas facing heightened air pollution risks.

The purpose of this project is to provide data-driven insights to support public policy decisions aimed at improving environmental health and mitigating pollution-related diseases.

TOOLS & METHODOLOGY

- **Data Analysis:** Conducted in RStudio using *tidyverse, dplyr, ggplot2,* and *lubridate* for cleaning, wrangling, and exploratory analysis.
- **Data Visualization:** Built in Power BI, featuring dynamic visuals that link air quality indicators with health metrics.
- **Data Sources:**
    - OpenAQ Platform – Historical PM2.5 data for major cities
    - WHO Air Quality Database
    - Pakistan Bureau of Statistics – Health and demographic indicators
    - Provincial Health Departments – Hospital admission records

---

DASHBOARD OVERVIEW

The Power BI dashboard visualizes the spatial and temporal dynamics of PM2.5 pollution and its correlation with respiratory illnesses.
It includes:

- **KPI Cards:**
    - **478 µg/m³** – Maximum PM2.5 recorded
    - **136.67 µg/m³** – National average PM2.5
- **Charts & Visuals:**

1. **PM2.5 by Year:** Displays average pollution trends from 2019–2020, showing gradual reduction in some regions.

2. **City and Average PM2.5 Map:** Highlights pollution hotspots, with Lahore and Karachi appearing most affected.

3. **Average Annual PM2.5 by City:** Compares pollutant levels, revealing Lahore as the city with the highest mean PM2.5 (~320 µg/m³).

4. **Effect of Air Quality Policies Over Time:** Shows pre- and post-policy PM2.5 levels, suggesting limited immediate impact.

5. **PM2.5 Trend Over Time by City:** Seasonal peaks, particularly during winter months, coincide with smog episodes.

6. **Average PM2.5 vs Respiratory Cases:** Demonstrates a positive correlation—cities with higher PM2.5 exhibit increased respiratory admissions.

DATA MODELLING & INSIGHTS

- **Correlation Analysis:**
  Using RStudio, Pearson correlation coefficients were calculated between PM2.5 averages and reported respiratory cases. The preliminary model indicates a strong positive correlation ($r \approx 0.78$), confirming that poor air quality significantly affects respiratory health outcomes.
- **Policy Evaluation:**
  Analysis of "Pre-policy" vs. "Post-policy" periods shows only a marginal decline in PM2.5, suggesting that policy enforcement and urban monitoring require strengthening.
- **City-Level Observations:**
  - **Lahore:** Exhibited the highest PM2.5 (≈ 320 µg/m³) and respiratory case rates.
  - **Peshawar:** Showed moderate pollution with seasonal fluctuations.
  - **Karachi:** Stable yet above WHO thresholds, indicating persistent air quality concerns.
  - **Islamabad:** Lowest PM2.5 among the cities analyzed, though still exceeding safe limits.

KEY TAKEAWAYS

1. Urban air quality is deteriorating, especially in Lahore and Karachi.
2. Respiratory cases rise proportionally with PM2.5 concentration, confirming the health-environment nexus.
3. Existing policy measures show limited impact, calling for integrated monitoring systems and public awareness initiatives.
4. Power BI visualization effectively communicates complex environmental-health data for policymaking and stakeholder engagement.

CONCLUSION

This project demonstrates how data analytics can inform public policy. By integrating open environmental and health datasets, performing R-based correlation modelling, and visualizing insights in Power BI, the study underscores the urgent need for sustainable air quality management in Pakistan.

It serves as a model for evidence-based decision-making, empowering policymakers to act on scientific insights and create measurable environmental reforms.