# Comprehensive Data Analysis on Lahore’s Air Quality (2019–2023)

## 1. Introduction

Lahore has faced increasing air quality challenges, with PM2.5 (a key component of smog) frequently exceeding safe thresholds. This report analyzes air quality data from 2019 to 2023, exploring trends, drivers, and actionable solutions to mitigate smog in the region.

## 2. Data Overview

### Datasets:

1. Air Quality Index (AQI): Daily PM2.5 measurements alongside meteorological data:  
 - Temperature (max, avg, min in °F)  
 - Humidity (max, avg, min as a percentage)  
 - Wind Speed (max, avg, min in mph)  
 - Atmospheric Pressure (inHg)  
2. Population and Vehicle Data: Annual records for Lahore's population and registered vehicles.

### Data Cleaning and Preprocessing:

• Missing PM2.5 values (13.4% of data) were filled using the median.  
• Outliers in AQI and temperature were capped using interquartile range (IQR).  
• Population and vehicle data, originally in string format, were converted to numeric values.

### Data Size:

• AQI Dataset: 1644 records  
• Demographics Dataset: 5 records spanning 2019 to 2023.

## 3. Analysis of Air Quality Trends

### 3.1 Yearly PM2.5 Trends

• 2019: PM2.5 average at 183.9 µg/m³.  
• 2023: Improved slightly to 165.7 µg/m³.  
• Fluctuations suggest external interventions, weather variability, and local policies impact air quality.

### 3.2 Seasonal AQI Patterns

1. Winter (December–February):  
 - AQI peaks at 257 µg/m³ in January, largely due to:  
 - Temperature inversions trapping pollutants.  
 - Emissions from industrial and household heating.  
 - Poor wind activity compounds pollutant concentration.  
2. Spring (March–May):  
 - Gradual AQI decline to around 140 µg/m³ in April as temperatures rise.  
 - Improved dispersion and reduced heating emissions contribute to lower levels.  
3. Summer (June–August):  
 - AQI dips to ~100 µg/m³, driven by:  
 - Increased wind speeds aiding pollutant dispersion.  
 - Rainfall washing out airborne particulates.  
4. Autumn (September–November):  
 - PM2.5 surges again to ~230 µg/m³ in November due to:  
 - Crop residue burning in surrounding areas.  
 - Onset of cooler temperatures reducing atmospheric mixing.

## 4. Drivers of Air Quality Changes

### 4.1 Meteorological Drivers

1. Temperature:  
 - Higher temperatures reduce AQI by facilitating vertical air mixing.  
 - Correlation: Moderate negative (-0.45).  
2. Humidity:  
 - Increased humidity clusters particulate matter, reducing AQI.  
 - Correlation: Negative (-0.48).  
3. Wind Speed:  
 - Critical in dispersing pollutants, showing the strongest impact.  
 - Correlation: Strong negative (-0.67).  
 - Regression coefficient: -9.52 (for each 1 mph increase in wind speed, AQI reduces by ~9.52 units).

### 4.2 Non-Meteorological Drivers

1. Population:  
 - Growth of 8.1% (2019–2023) from 12.03M to 13.01M.  
 - Weak negative correlation (-0.12) with AQI suggests that population increase alone does not strongly drive smog.  
2. Vehicles:  
 - Registered vehicles increased by 17.6%, from 5.35M to 6.29M.  
 - Correlation (-0.35) indicates that vehicle emissions are partially offset by better fuel and technology standards.

## 5. Statistical Findings

### 5.1 Correlation Analysis

• Population and Vehicles: Strong correlation (+0.92), reflecting urbanization.  
• AQI and Weather: Strong dependence on wind speed and temperature; less sensitivity to population and vehicles.

### 5.2 Regression Analysis

• Wind Speed: Coefficient -9.52 (largest impact on AQI).  
• Temperature: Coefficient -2.72 (moderate effect).  
• Humidity: Coefficient -0.82 (minor effect).  
• Population: Coefficient +0.000045 (minimal impact).  
• Vehicles: Coefficient -0.000038 (minimal impact).

### 5.3 Seasonal Comparison: Winter vs. Other Seasons

1. Winter Temperature:  
 - Average: 56°F (vs. 82°F in other seasons).  
 - Colder temperatures trap pollutants, elevating AQI.  
2. Winter Wind Speed:  
 - Average: 10.4 mph (vs. 17.5 mph in other seasons).  
 - Lower wind speeds reduce pollutant dispersion.

## 6. Visualizations and Key Insights

• Boxplots and Histograms: Highlighted AQI outliers during winter.  
• Seasonal Plots: Demonstrated recurring winter AQI spikes.  
• Correlation Heatmap: Showed significant negative relationships between AQI and weather variables.  
• Regression Coefficients: Visualized wind speed as the most significant factor.

## 7. Recommendations

### 7.1 Policy Recommendations

1. Regulate industrial emissions, especially during smog-prone months.  
2. Ban crop residue burning with incentives for alternative practices.  
3. Enforce stricter vehicle emissions standards.

### 7.2 Urban and Infrastructure Planning

1. Develop urban corridors to facilitate wind flow.  
2. Increase vegetation to act as natural air filters.  
3. Expand public transportation to reduce vehicle dependency.

### 7.3 Public Awareness and Engagement

1. Educate citizens on reducing household emissions.  
2. Promote carpooling and use of cleaner technologies.

### 7.4 Research and Monitoring

1. Invest in high-frequency AQI monitoring stations.  
2. Study long-term impacts of meteorological changes on smog.

## 8. Conclusion

This comprehensive analysis confirms that Lahore's smog crisis is driven by a complex interplay of meteorological and anthropogenic factors. While weather conditions are the dominant seasonal drivers, targeted interventions in emissions control, urban planning, and public behavior can significantly reduce PM2.5 levels, improving the city's air quality and public health.