Project Title: Comprehensive Air Quality and Health Impact Analysis in New York City

Business Problem Statement:

You have been provided with a detailed dataset on New York City's air quality, which includes pollutant levels (NO_2, PM2.5, Ozone), emission sources (vehicle miles traveled, boiler emissions), population density, geographic locations, seasonal and yearly time frames, and associated health indicators (asthma, cardiovascular issues, respiratory hospitalizations, and death rates).

Your objective is to perform a comprehensive data analysis using Python and Excel to uncover trends, correlations, and regional patterns that influence air quality and public health across NYC. The goal is to support government agencies, environmental strategists, and public health professionals with actionable insights.

You will create an analytical report and an interactive dashboard that answer the following integrated business questions:

1. **Pollution Levels & Sources:** What are the average, minimum, and maximum concentrations of pollutants across different districts? How do vehicle emissions and boiler usage contribute to air pollution?

2. **Temporal Trends:** How do air quality and health outcomes vary by season and over the years? Are there any improvements or worsening trends?

3. **Geographic Insights:** Which areas (boroughs or districts) are most affected by pollution and health risks? How does population density relate to air quality metrics?

4. **Health Impact Assessment:** What is the relationship between pollution levels and hospitalization/emergency rates for asthma, respiratory, cardiovascular issues, and mortality?

5. **Demographic Patterns:** How do different age groups (All Ages, Under 18, 18+, 30+) experience health impacts in relation to air quality?

6. **Outlier Detection & Data Quality:** Are there any extreme values in pollutant or health data? What steps were taken to handle them for accurate modeling?

7. **Measure Standardization:** How were units and measurement standards (e.g., $\mu g/m^3$, per 100,000 population) harmonized for cross-comparison?

Deliverables:

- A cleaned and transformed dataset with standardized measures and derived insights

- Visual reports and summary dashboards built in Excel and Python (Matplotlib/Seaborn/Pandas)

- A brief insight document outlining major findings and actionable recommendations