Data Analysis Project

Data Analyst: Shivangi Gupta

Client/Sponsor: Public Health & Environment Monitoring Authority

Purpose:

Write a brief description of why this project is happening below. Why is this project happening? What are the goals?

The project aims to analyze the air quality data to understand the impact of Fine Particles PM 2.5, (NO_2) Nitrogen Dioxide, and (O_3) Ozone levels on asthma in New York City. The dataset contains air quality readings, possibly from different stations, timeframes, or pollutants. First, we'll inspect the structure. By analyzing this dataset, we can identify:

1. ***Air Quality Trends**: How have pollutant levels (PM2.5, NO_2, O_3) changed over time and by season?*
2. ***Geographic Disparities:** Which neighborhoods experience the worst air quality and health outcomes?*
3. ***Health Outcomes:** Are emergency visits, hospitalizations, and deaths increasing or decreasing?*
4. ***Cause Attribution:** Which pollutants are most associated with asthma, cardiovascular, or respiratory conditions?*

Scope / Major Project Activities:

What are the major parts of this project? List out the high-level steps, activities, or stages of the project, and give a brief description for each.

Activity	Description
Data Cleaning	Remove duplicates, handle missing values, convert units (µg/m³), fix encoding issues, and normalize measures for consistency.

Exploratory Data Analysis (EDA)	Analyze distributions, detect trends in pollutant levels, visualize seasonal and annual behaviors, and assess pollutant concentration across boroughs.
Air Quality Trends	Explore how the pollutant levels (PM2.5, NO_2, O_3) changed over time and by season.
Geographic Disparities	Identify neighborhoods with the highest exposure to PM2.5, NO_2, and O_3; visualize via maps and ranked bar charts.
Health Outcomes Analysis	Explore the relationship between air quality and health outcomes including asthma-related visits, hospitalizations, and deaths.
Correlation Analysis	Examine how specific pollutants affect different causes (e.g., asthma, cardiovascular disease) using bivariate analysis.

This project does not include:

Specify the things that this project isn't responsible for doing (out of scope). For instance, "this project does not involve a summation of 2019 data analysis"

1. *Predictive Modeling or machine learning analysis.*
2. *Primary data collection or consumer surveys*
3. *Marketing campaigns.*

Deliverables:

A specific list of things that your project will deliver.

Deliverable	Description/ Details
Business Task Statement	Clear articulation of project goals: analyzing pollution trends, health effects, and geographic inequalities.

Cleaned Dataset	Standardized dataset with cleaned columns, date-formatted fields, seasonally and spatially tagged records
Data Visualizations	Python plots showing time-series trends, pollutant distributions, seasonal variations, and health impact breakdowns.
Power BI Dashboard	Multi-page interactive dashboard displaying KPIs, geographic hotspots, seasonal insights, and health outcome patterns.

Schedule Overview / Major Milestones:

The expected schedule for the project. This can be defined by milestones (e.g. "all data is cleaned and processed"), periods of time ("Week 1 / Week 2"), or other ways based on the needs of the project.

Milestone	Expected Completion Date	Description/Details
Data Understanding	*Week 1*	*Review dataset structure, requirements, and define EDA plan.*
Data Cleaning Completed	*Week 1*	*Dataset ready for analysis: formatted, de-duplicated, and unit-consistent.*
EDA & Visualization	*Week 1*	*Perform trend, seasonality, analyses; generate key visualizations.*
Health Outcome Linkage	*Week 2*	*Associate pollutants with ED visits, hospitalizations, and deaths.*
Power BI Dashboard Finalization	*Week 3*	*Interactive dashboard pages and presentation-ready insights compiled.*

*Estimated date for completion:

Four weeks from the start date.