**PROJECT TITLE: AIR POLLUTION AND RESPIRATORY DISEASE ANALYTICS IN GROWING AFRICAN CITIES**

**Case Study: Examining the Air Quality in Lagos, Nigeria**

**Project Description:**

African cities like Lagos (Nigeria), Nairobi (Kenya), Accra (Ghana), and Kinshasa (DR Congo) are witnessing rapid urbanization. However, this growth comes with worsening air quality due to vehicle emissions, industrial activities, and poor waste management.  
This project aims to analyze satellite-recorded air pollution data alongside hospital-reported respiratory disease data to understand the relationship between air pollution and health outcomes.  
Ultimately, it seeks to push for urban clean air initiatives across Africa.

**Background:**  
Lagos is Africa’s most populous city (>20 million people) and faces severe air pollution challenges due to high traffic, diesel generators, industrial zones, and waste burning.

**Global Evidence:**  
WHO reports Lagos PM2.5 levels are 5x higher than recommended limits.  
Hospitals like LUTH (Lagos University Teaching Hospital) have reported increasing respiratory cases — particularly during Harmattan (December-February) when dust intensifies.

**Why Lagos for Pilot:**

* Availability of satellite pollution data (e.g., Sentinel-5P, MODIS).
* Active hospital reporting systems.
* Urgent urban policy needs.

**Project Objectives:**

* Derive the pollution index (a composite feature comprises of all pollutants.
* Monitor trends in air pollution levels across various air pollutants over time.
* Analyze the relationship between pollution spikes and hospital respiratory cases.
* Predict respiratory disease surges using pollution data.
* Identify high-risk cities, periods, and pollutants.
* Recommend public health and urban environmental policies.

**Data Dictionary:**

| **Feature Name** | **Description** |
| --- | --- |
| City | |  | | --- |  | Local area in Lagos (e.g., Ikeja, Yaba, Ajah, Surulere, Lekki) | | --- | |
| Date | Date of observation (YYYY-MM-DD) |
| pm2\_5 | Fine particulate matter concentration (µg/m³) |
| pm10 | Coarse particulate matter concentration (µg/m³) |
| no2 | Nitrogen dioxide level (µg/m³) |
| so2 | Sulphur dioxide level (µg/m³) |
| o3 | Ozone concentration (µg/m³) |
| hospital\_id | Unique hospital identifier |
| respiratory\_cases | Number of new respiratory-related hospital cases reported |
| avg\_age\_of\_patients | Average age of patients reported |
| weather\_temperature | Average daily temperature (°C) |
| weather\_humidity | Average daily humidity (%) |
| wind\_speed | Average daily wind speed (m/s) |
| rainfall\_mm | Daily rainfall (mm) |
| population\_density | People per square kilometer in the city |
| industrial\_activity\_index | Proxy score (0–100) showing industrial pollution activity |

**Project Workflow**

**Exploratory Data Analysis (EDA)**

* Visualize pollution trends over time for each city.
* Analyze seasonal patterns (e.g., worse during dry season?).
* Explore spikes in hospital cases and identify which zones and months have higher pollution.

**Time-Series Trend Modeling**

* Detect seasonal variations in pollution and respiratory cases.

**Correlation Analysis**

* Correlate PM2.5, PM10, NO2 levels with total respiratory cases.
* Identify the strongest predictors of health burdens.
* What pollutants are most associated with health issues?

**Predictive Modeling**

* Build regression models (Linear Regression, Random Forests) to predict respiratory cases from pollution data.
* Additionally, try Time Series forecasting (Prophet, ARIMA).

**Insights and Recommendations**

* Identify high-risk periods (e.g., dry seasons) and cities.
* Recommend early warnings and public health interventions.

**Hypotheses:**

| **Hypothesis** | **Status** |
| --- | --- |
| H1: Higher PM2.5 levels will correlate with more respiratory hospital cases. | To be tested |
| H2: Cities with higher industrial indices have worse air quality. | To be tested |
| H3: Harmattan season will show spikes in PM10 and respiratory cases in West Africa. | To be tested |
| H4: Weather conditions (low humidity, high temperatures) worsen pollution impact. | To be tested |

**Research Questions:**

**1. Trend Analysis:**

* How have air pollution levels changed over time in these cities?
* Are there seasonal patterns in pollution levels?

**2. Health Impact:**

* Is there a correlation between air pollution spikes and hospital respiratory cases?
* Which pollutants have the strongest relationships with diseases?

**3. Predictive Modeling:**

* From the various model trained, which are the most performing features?

**4. Vulnerability and Risk:**

* Which cities/months are at highest risk?

**5. Policy and Intervention:**

* What policies could significantly reduce respiratory health issues based on findings?

**Deliverables:**

* Full EDA report (visuals + insights)
* Predictive models with evaluations
* Policy recommendations
* Presentation slides for stakeholders