PROJECT DOCUMENTATION

Indian Urban Air Quality and Health Impact 2019-2024 – Comprehensive data analysis Using Python

1. PROJECT OVERVIEW

The project analyzes and explores the air quality data across 24 Indian cities, to evaluate the health impact on the population. This dataset presents compiled air quality data and health impact indicators from major Indian cities between 2019–2024. It reflects pollutant measurements (PM2.5, PM10, NO_2, CO, SO_2, O_3), meteorological factors, and urban metrics such as traffic and industrial activity.

Domain: Healthcare

2. TOOLS USED
   - Pandas
   - Matplotlib
   - Seaborn

3. DATASET
   - Source: Indian Urban Air Quality and Health Impact 2019-2024.csv
   - Data contains: City, AQI, PM2.5, PM10, NO_2, SO_2, O_3, CO, Temperature, Humidity, Windspeed, Rainfall, Pressure, Vehicle count, Industrial Activity Index, Health Impact Score

4. STEPS FOLLOWED:

(i) Data Loading and Initial Overview:
   - Imported dataset using Pandas and conducted initial overview of the data
   - Number of rows and columns, data types of each column and initial observations were checked (head, info, describe)

(ii) Data pre-processing:
   - Checked for missing values and duplicates
   - Created derived column titled 'Levels of Health Concern' to indicate the health impact level based on Air Quality Index (AQI)
   - Filter and group by options were utilized to segregate the cities based on levels of health concerns

(iii) Exploratory Data Analysis (EDA)
   - Conducted univariate, bivariate and multivariate analysis
   - Explored Groupby, pivot tables and correlation analysis options.

(iv) Visualizations

- Generated bar plots, scatter plots, line charts, pie charts, histograms, box plots, reg plot, heatmaps
- Explanatory Markdown cells are added throughout the Jupyter notebook, to signify the interpretations

5. <u>KEY INSIGHTS</u>

- Air Quality Index - Interpretation of health concern:

Based on the AQI value, a derived column 'Levels of Health concern' is created. This helps to analyze the health impact on the Indian urban population upon exposure.

AQI 0-50 Good

AQI 51-100 Moderate

AQI 101-150 Unhealthy for sensitive groups

AQI 151-200 Unhealthy

AQI 201-300 Very Unhealthy

AQI 301-500 Hazardous

- All the major Indian cities, on average, have an Unhealthy air exposure with an AQI ~ 270
- Most of the AQI readings taken in the Indian cities fall under the Hazardous range. This indicates that the entire population is more likely to be affected, may also cause respiratory illness to the people on prolonged exposure.
- PM2.5 is the scale of Fine Particulate Matter which are particles in the air that measure less than 2.5 micrometers (µm) in diameter, typically composed of a mixture of solid or liquid particles suspended in air. These fine particles can travel deep into the respiratory tract, reaching the lungs and entering the blood stream. PM2.5 readings are taken considering a 24-hour exposure period (Unit: $\mu g/m^3$). PM2.5 at or below 12 $\mu g/m^3$ is indicative of Healthy with no/little risk scenario.
- As the average PM2.5 values fall in the range of 55.5 to 150.4 $\mu g/m^3$, this indicates an Unhealthy population of Indian cities with increased chances for respiratory effects.
- PM10 refers to particulate matter with a diameter of 10 microns or less, capable of entering the respiratory tract. It includes dust, pollen, mold, and emissions from vehicles, industries, and construction activities.

- PM10 readings are taken considering a 24-hour exposure period (Unit: µg/m³). Impacts of higher PM10 levels could be respiratory issues, cardiovascular damage and impaired lung growth in children.
- As the average PM10 values fall in the range of 155 to 254 µg/m³, this indicates an Unhealthy to sensitive group scenario where sensitive individuals in these cities are at higher risk of experiencing irritation and respiratory problems. These cities could be under the threat of chronic air pollution.
- The boxplot of NO_2, SO_2, and O_3 Concentrations shows that O_3 has the highest variability in concentration. O_3 has been recorded to have the highest outlier concentration (~100 µg/m³) found at Thane city, indicative of pollution events.
- In case of SO_2 concentrations, there is an outlier at ~ 50 µg/m³ which is recorded at cities such as Thane, Agra, Surat and Chennai. The corresponding Industrial Activity Indices of those cities are relatively high. The outlier correlates with the possible SO_2 emissions due to the prominent industrial activity in those locations.
- Highest average vehicle count is in Nagpur, and the same city has the highest PM2.5 and PM10, indicating the air pollution impact from vehicles. Though this exact trend is not replicated in Industrial Activity Index (IAI), the city Surat with highest average IAI has significantly high PM2.5 and PM10. Though not together correlated, IAI and Vehicle count are individually strongly correlated with the particulate matter in air (PM2.5 and PM10).
- The consistent standard deviation values across the cities signify that the spread of data/ amount of fluctuation in PM2.5 and PM10 levels is uniform. A similar trend is observed in the maximum and minimum values as well.
- The frequency of different value ranges of AQI, PM2.5, PM10 and Industrial Activity Index is visualized to check for central tendency (if any). As the histograms represent similar heights, unlike a normal distribution (with a peak), the dataset has no strong central tendency. Data is evenly spread across the range and every interval is almost equally likely.
- As humidity increases, PM10 levels at Rajkot also tend to increase slightly, but not strongly (~0.1 correlation factor). At high humidity, PM10 (larger particles like dust, ash, pollen) can absorb water, swell, and stay suspended longer instead of settling. Humid air is often denser and associated with low wind speeds, reducing pollutant dispersion, so particles accumulate. The relationship is weak, so humidity alone does not fully explain PM10 variations—other factors (traffic, industry, wind, rainfall) likely dominate.
- In humid conditions (>50% in Delhi), NO_2 reacts with hydroxyl radicals (OH·) and water vapor to form nitric acid (HNO_3), reducing its levels. CO, on the other hand, is relatively stable and oxidizes more slowly (via OH· radicals), as Delhi is an area of chronic air

pollution. Higher wind speeds dilute local NO_2, dispersing it over a larger area, resulting in a weak negative correlation.

- High humidity often leads to water condensation on particles, forming larger droplets. These droplets settle faster, reducing PM2.5 concentration. This dispersion is enhanced at warmer temperatures. This is signified by the weak negative correlation of PM2.5 with temperature during the high humid conditions in Delhi, i.e. when the relative humidity is above 50%.
- The cities (Eg: Hyderabad, Mumbai) which are predominantly affected by the PM10 levels due to Industrial activity, are indicated by the steep falls in PM10 levels (red line- at low IAI). The cities (Eg: Ahmedabad, Bhopal) which are predominantly affected by the PM2.5 levels due to Industrial activity, are indicated by the steep falls in PM2.5 levels (green line- at low IAI).
- The cities (Eg: Mumbai, Patna, Vadodara, Chennai etc.) which are predominantly affected by the NO2 emission from Industrial activity are indicated by the steep falls in NO2 levels (red line- at low IAI). The cities (Eg: Hyderabad, Patna) which are predominantly affected by the SO2 emission from Industrial activity are indicated by the steep falls in SO2 levels (green line- at low IAI).
- The cities (Eg: Bhopal, Jaipur, Ludhiana, Meerut, Rajkot etc.) which are predominantly affected by the CO emission from vehicles are indicated by the steep falls in CO levels (orange line- at low vehicle counts).
- Corrective measures:
    - Pollution standards to regulate the high-emission from vehicles can help improve air quality.
    - The industrial activity management needs to be systematically implemented.

6. FILES INCLUDED
    - Indian Urban Air Quality and Health Impact 2019-2024- Data Preprocessed.xlsx
    - Indian Air Quality_Final Project_17-08-2025.ipynb
    - Project Documentation_Python.pdf

7. HOW TO USE
    - Open Indian Urban Air Quality and Health Impact 2019-2024- Data Preprocessed.ipynb in Jupyter notebook to view the cleaned data.
    - Open Indian Air Quality_Final Project_17-08-2025.ipynb in Jupyter notebook to explore the visuals and interpretations.
    - Project Documentation_Python.pdf – For steps followed and insights arrived