2. Data Dictionary

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Variable Name** | **Measurement Unit** | **Allowed Values** | **Description** |
| City Name | city | String | Letters | Name of the city where data was recorded |
| Month of the Year | year\_month | YYYY-MM | 2020-2025/01-12 | Month of observation (YYYY-MM-DD) |
| Air Quality Index | aqi | Numeric | 0 - 500 | Air Quality Index value |
| Raw Pollutant Matter | pm2\_5 | Numeric | 0 - 100 | PM2.5 concentration (µg/m³) |
| Raw Pollutant Matter | pm10 | Numeric | 0 - 100 | PM10 concentration (µg/m³) |
| Nitrogen Dioxide | no2 | Numeric | 0 - 100 | Nitrogen dioxide concentration (ppb) |
| Ozone | o3 | Numeric | 0 - 100 | Ozone concentration (ppb) |
| Temperature | temperature | Numeric | –8 to 45 | Daily average temperature (°C) |
| Levels of Humidity | humidity | Numeric | 0 - 100 | Daily average relative humidity (%) |
| Number of Hospital Admissions | hospital\_admissions | Numeric | 0 - 500 | Number of daily hospital admissions for respiratory issues |
| Available Beds | hospital\_capacity | Numeric | 0 - 2500 | Number of beds available in local hospitals |
| Types of Population | population\_density | String | Urban, Suburban, or Rural | Urban, Suburban, or Rural classification |

3. Data Collection Methodology. State how the data is collected, by whom, how often, etc.

The “Global Air Quality and Respiratory Health Outcomes” dataset combines information from air quality monitoring stations (operated by government or environmental agencies) with public health and hospital records on respiratory-related hospitalizations. It encompasses multiple cities worldwide. Air quality indicators are originally measured on daily basis and then aggregated into monthly averages to align with the health outcome data, which are reported over corresponding timeframes. Since the dataset spans diverse cities and healthcare systems, raw pollutant concentration measurements are used as a common basis to match with hospitalization outcomes.

4. Why does this data set intrigue you?

This dataset is particularly valuable because it offers a global, multi-city perspective that spans diverse climates and standards of living. What makes it especially compelling is its linkage of air quality metrics with respiratory-related hospital admissions. Such data can support public health analyses aimed at identifying which air quality factors most strongly affect health outcomes in these cities, thereby enabling more informed decision-making for public health officials and healthcare systems. Moreover, it provides a strong foundation for exploring the relationship between urban air pollution and health outcomes through advanced machine learning applications.