Can You Breathe Here??

Air Quality & Public Health Risk Analysis

Project Overview:

Air pollution affects millions of people every day, yet its impact is often discussed using technical numbers that are hard to understand. Air Quality Index (AQI) values and health statistics exist, but they are rarely connected in a way that clearly answers a simple question:

Can people breathe safely where they live?

This project brings air quality and public health data together to create a clear, easy-to-understand picture of environmental health risk across countries and years. Instead of focusing only on charts or raw numbers, the project introduces a new, interpretable measure called the **Breathability Score**, designed to make complex data meaningful for everyone.

Objectives:

This project aims to :

- Understand how air pollution varies across countries and over time
- Explore how pollution relates to respiratory health outcomes
- Combine datasets that were originally reported at different levels
- Communicate risk in a way that non-technical audiences can understand
- Present insights responsibly and transparently

Data Sources:

- Air Quality Data from Kaggle
- Public health data from Kaggle

Cleaning and Preparing the Data

The raw data required preparation before analysis:

- Country names were extracted from city names so all data could be aligned
- Only relevant variables were kept to avoid unnecessary complexity
- Data types and missing values were checked to ensure reliability

This step ensured that all later analysis was based on clean, trustworthy data.

Solving a Real-World Data Challenge: City vs Country Data

A key challenge was that:

- Air quality data was reported by city
- Health data was reported by country

Because these levels didn’t match, the datasets could not be compared directly.

The solution:

All city-level AQI values were combined into a single country-level value for each year

For each country and year, the project calculated:

- The average AQI
- The median AQI

- The number of cities contributing data

Preparing the Health Data

The health dataset already summarized information at the country level, but still required refinement:

- Only indicators related to respiratory health were selected
- Column names were simplified for clarity
- Duplicate country-year records were combined to ensure consistency

This produced one clean health profile per country per year.

Combining the Datasets

Once both datasets were prepared:

- were merged using **country and year**
- Only years where both air quality and health data were available were kept

The result was a single, aligned dataset that allowed pollution and health outcomes to be examined together.

Exploring the Data: What the Patterns Show

After combining the data, several patterns became clear:

- Countries with higher air pollution often show worse respiratory health outcomes
- Urbanization and pollution tend to increase together
- Strong healthcare systems reduce risk but cannot fully offset poor air quality

These patterns were explored using:

- Scatter plots to compare pollution and health outcomes
- Correlation heatmaps to show relationships between multiple factors
- Country-level comparisons to highlight global differences

The "Breathability Score" and why create it?

Raw numbers like pollution levels or mortality rates can be confusing, especially for people without a technical background.

To solve this, the project introduced a **Breathability Score**, a single number that summarizes how safe the air is to breathe in a given country and year.

How the Score Works

Two things were considered:

- **How polluted the air is**
- **How serious are respiratory health outcomes are**

Both values were scaled to the same range and given equal importance.

Breathability Score = $100 - (50 \times \text{AQI}_{normalized} + 50 \times \text{Mortality}_{normalized})$

Then they were combined into one score from **0 to 100**:

- Higher score → healthier air
- Lower score → higher health risk

How to Read the Score

- 80–100: Clean air and low health risk

- 50–80: Moderate air quality and some risk
- Below 50: Poor air quality and high health risk

Results & Insights

Countries with Better Air to Breathe

These countries generally had:

- Lower pollution levels
- Strong healthcare systems
- Well-managed urban development

Countries with Higher Risk

These countries were often affected by:

- Persistent air pollution
- Rapid or unplanned urban growth
- Higher rates of respiratory illness

Key Takeaways

- Poor air quality is closely linked to respiratory health problems.
- Urban growth increases pollution risks when infrastructure does not keep up.
- Looking at pollution across entire countries reveals long-term exposure patterns.
- The Breathability Score makes complex environmental health risks easy to understand and communicate.

Limitations

Important limitations include:

- Country-level data can hide differences between cities.
- The analysis shows **relationships**, not direct cause-and-effect.
- Some countries have missing or incomplete data.

Tools & Technologies

- **Python**
- **Pandas & NumPy** — data processing
- **Matplotlib & Seaborn** — visualization
- **Scikit-learn** — normalization
- **VS Code** — developing the project