DGR - Milestone 1

Beatrice Grassano | 370780
Lorenzo Drudi | 367980
Emanuele Rimoldi | 377013
DGR

1 Problematic

Air quality isn't just about fresh air—it's about health, productivity, and the future of our planet. Poor air quality has been linked to respiratory diseases, heart conditions, and even cognitive decline. In the past, air pollution was often considered a local issue, only drawing attention when it reached extreme levels. However, with growing climate awareness and advancements in monitoring technology, we now recognize it as a global concern. Wildfires, urban smog, and industrial emissions impact not only major cities but entire ecosystems. Achieving cleaner air isn't just an aspiration—it's essential for a sustainable and healthier future.

Project Overview

Our project aims to illustrate how air quality in the European Union has evolved over the past 20 years and project future trends based on current patterns.

We will analyze a wide range of pollutants and metrics, exploring their sources and impact. We are deeply committed to this issue and want our project to engage all of society—from experts to the general public—through interactive explanations.

Additionally, we will include a *Curiosities* section featuring engaging visuals that highlight intriguing aspects of air quality, such as the impact of electric vehicles and how air pollution influences the real estate market (National Bureau of Economic Research , 2019).

Research Questions

- How has air quality changed globally over the past 20 years, and what are the main factors driving these changes?
- What are the projected air quality trends for the next 10–20 years if current patterns continue?
- How do innovations like electric vehicles and urban green spaces influence air quality, and what unexpected effects do they have on the economy and society?

2 Dataset

- **OpenAQ** (OpenAQ , 2023): OpenAQ is a nonprofit platform that aggregates and harmonizes air quality data from 163 global providers into an open-access database. Data can be accessed via an API (with rate limits) or downloaded from **AWS S3** and queried using **AWS Athena**. Given the large volume of historical data (hundreds of gigabytes), we use OpenAQ primarily as a reference to validate our results rather than for direct analysis.
- **AirBase** (Environmental European Agency , 2025): Maintained by the European Environment Agency (EEA), AirBase is our primary data source, compiling air quality measurements from EU Member States, EEA countries, and partner nations. It provides a multi-year time series of pollutant levels, along with metadata on monitoring networks and stations. A user-friendly data downloader facilitates access. The dataset spans from 2002 to 2025 and is categorized as:
  - **Historical** (2002–2012): Archived data.
  - **Verified** (2013–2023): Reported annually by 30 September.
  - **Unverified** (2024–onward): Continuously transmitted.

To explore correlations between air quality and socio-economic factors, we integrate additional datasets:

- **House Prices**: Environmental factors influence real estate, yet no historical dataset covers geographic price trends.

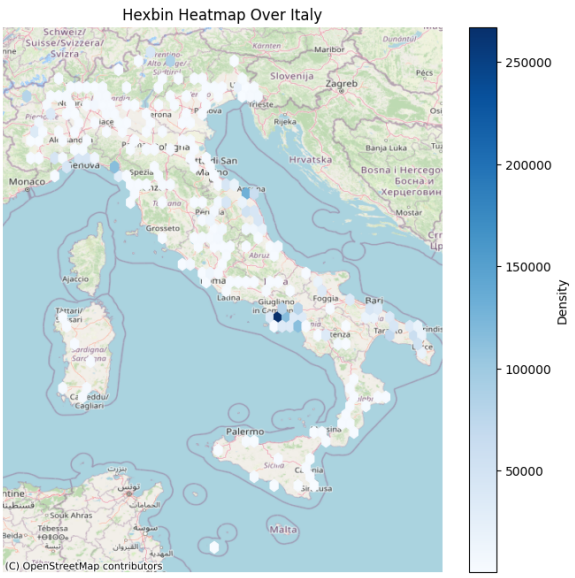

Figure 1: Hexbin Map over Italy showing locations distribution and measurement density.

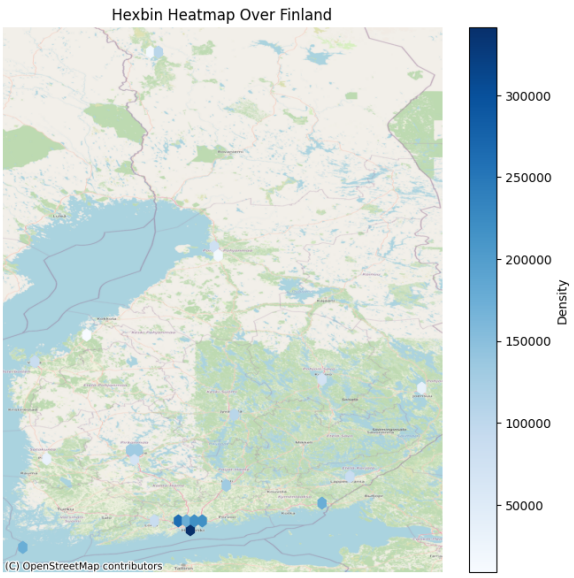

Figure 2: Hexbin Map over Finland showing locations distribution and measurement density.

To address this, we scraped data (see *scraper_immobiliare_it.py* in the data folder) from Immobiliare.it (Immobiliare.it , 2025) for 2019–2025. This analysis is included as a curiosity section on our website.

- **Electric Vehicle Usage**: The **Global EV Outlook** (Electric Vehicles Initiative , 2024) is an annual report on electric mobility trends worldwide. This dataset helps assess the potential link between air quality improvements and EV adoption.

3 Exploratory Data Analysis

We conducted an *exploratory data analysis* (eda) on a selected set of 10 air pollutants: *PM2.5*, *PM10*, *NO*, *NO2*, *NO3*, *O3*, *CO*, *Co*, *SO2*, *NH3*, and *Pb*. In particular, the notebook *eea_air_quality_data_eda* (located in the analysis folder) presents an in-depth analysis of *PM2.5*, which is considered one of the most harmful pollutants.

As part of our analysis, we examined air quality data from a subset of four countries—*Italy*, *Poland*, *Finland*, and *Spain*—carefully selected to ensure a balanced distribution of samples across regions with some of the cleanest air, as determined by *PM2.5* concentration per cubic meter (World Health Organization , 2025).

Our analysis revealed differences in data coverage across countries. Some countries, such as Italy (see Fig. 1), have a more comprehensive spatial distribution of measurement locations, whereas others, like Finland (see Fig. 2), are less uniformly covered. Specifically, Italy features numerous monitoring locations that collectively provide broad national coverage. In contrast, Finland's monitoring sites, while fewer, are well-distributed and maintain a consistent level of measurement density. Conversely, Italy exhibits a more uneven distribution, with certain locations collecting a disproportionately large amount of data while others have significantly fewer measurements.

Several techniques can be applied to address these issues:

- **Uneven Spatial Coverage**: Apply spatial normalization techniques (*e.g.*, dividing measurements by population density or land area), identify clusters of monitoring stations to assess potential over-representation, or use geostatistical interpolation (such as Kriging) to estimate values in areas with fewer measurements.
- **Imbalance in Measurement Density**: Use weighting based on measurement frequency to mitigate bias from overrepresented locations. Alternatively, reduce the number of data points in highly sampled areas or generate additional points in underrepresented regions using statistical methods.

4 Related Work

To leverage our chosen datasets in an innovative way, it is essential to examine how researchers and

platforms have already utilized them:

- The **AirBase** dataset has been used to analyze historical air quality trends and assess the impact of EU policies on pollution reduction. Spatial analysis has helped identify regions with chronic pollution and seasonal spikes in pollutants such as *PM10*, *NO2*, and *Ozone*.
- The **OpenAQ** dataset powers the OpenAQ Explorer, allowing users to map air quality and filter by pollutant type. However, it lacks key context, such as pollutant descriptions and effects, and omits time series and correlation analysis, limiting trend insights.

Beyond understanding the existing uses of these datasets, we must also consider the many platforms that already provide global air quality visualizations. For instance, *Google Maps* and weather applications like *3B Meteo* integrate air quality features. However, two specific websites have been key sources of inspiration for our project:

- Waqi.info: This website displays the *Air Quality Index* (AQI) on a map with color-coded labels and pollutant details. However, its usability is limited—navigation is unintuitive, and users cannot manually select pollutants to visualize but must wait for an automatic panel transition.
- AQI: More comprehensive than *Waqi.info*, this site offers historical pollution data over multiple time frames (24 hours, 7 days, 30 days), inspiring our project to extend this to daily, monthly, and yearly views. However, despite better navigation, it lacks an educational component.

The originality of our approach lies in creating a platform that is not only user-friendly and visually appealing but also fosters discussion and education. Through concise informational insights, we aim to raise awareness of how air quality is both influenced by human behavior (e.g., choosing low-emission vehicles) and affects daily life (e.g., real estate values). Our goal is to make air pollution more "visible," helping people recognize an often-overlooked problem simply because it is not immediately perceptible.

References

Electric Vehicles Initiative . 2024. Global EV Outlook . https://www.iea.org/data-and-statistics/data-tools/global-ev-data-explorer. Accessed: 2025-03-20.

Environmental European Agency . 2025. European Air Quality Dataset . https://eeadmz1-downloads-webapp.azurewebsites.net/. Accessed: 2025-03-20.

Immobiliare.it . 2025. Dati dei prezzi degli immobili in Italia . https://www.immobiliare.it/mercato-immobiliare/. Accessed: 2025-03-20.

National Bureau of Economic Research . 2019. The Effect of Air Pollution on Housing Prices . https://www.nber.org/digest/mar99/cleaner-air-results-higher-home-prices. Accessed: 2025-03-20.

OpenAQ . 2023. Retrieved from https://api.openaq.org . https://api.openaq.org. Accessed: 2025-03-20.

World Health Organization . 2025. Countries with the Cleanest Air . https://global-relocate.com/rankings/the-cleanest-air. Accessed: 2025-03-20.