RSC Team 2 Deliverable Documentation

Project Summary & Technical Handoff

Objective

For our deliverable, we developed an interactive dashboard to visualize air pollution-related risks in South San Francisco and San Bruno. By integrating air quality data with public health vulnerability metrics, we generated a Composite Risk Score at the census tract level. We designed the dashboard as a tool to help community members and decision-makers identify neighborhoods at higher environmental and health risk.

Data

We collected air quality data from three sources: Clarity monitor data, provided by Rise South City, covers the period from October 30, 2024 to March 31, 2025 and is stored in the file `clean_clarity.csv`. PurpleAir API data, pulled by our team, extends coverage from March 30, 2024 to May 3, 2025 and is saved in `clean_api_purpleair.csv`. In addition, we merged historical PurpleAir data from December 27, 2018 to April 9, 2025 with API records to create `clean_purpleair.csv`. These datasets were cleaned, merged, and standardized using EPA's AQI formula, and final results were stored in both CSV and GeoJSON formats for visualization.

Health vulnerability data was processed into a Health Risk Index stored in `health_risk_index.csv`. This index, normalized between 0 and 1, draws from CalEnviroScreen 4.0 and San Mateo County's All Together Better health indicators. Metrics included respiratory disease prevalence, access to care, and structural inequities. Raw indicators were standardized and combined using the methodology documented in `health.ipynb` and `health_preproc.ipynb`.

Composite Risk Score

The dashboard's main feature is the Composite Risk Score, which combines the Health Risk Index (HRI) and the Air Quality Index (AQI) using a user-defined weighting parameter. The formula is:

$$\text{Risk} = \alpha \times \text{HRI} + (100 - \alpha) \times \text{AQI}$$

Where the user selects the weight α via an interactive slider. Users then can explore how different weightings of environmental versus health vulnerabilities affect overall risk across tracts.

Consistency and Predictability Scores

To assess sensor reliability, we calculated consistency and predictability scores. The consistency score measures how accurately a monitor predicts its own next-day PM2.5 levels using data from the previous seven days, modeled with a Random Forest. The predictability score measures how well a monitor's reading can be inferred from its five nearest neighbors on the same day. Both scores are based on model accuracy and provide a basis for evaluating confidence in the sensor data.

Additional Visualizations

The "`figures`" folder includes maps of AQI, HRI, and Composite Risk Scores by census tract, along with graphs of daily PM2.5 trends. These visualizations supported weekly meetings, presentations, and the final dashboard. Additional contextual datasets—such as airport traffic and census tract boundaries—are stored in the "miscellaneous" folder and facilitate further spatial analysis.

FAQs

1. How do I run the dashboard?
   To run the dashboard locally, run the following command from the root directory of the drive/repository:
   ```
   streamlit run code/streamlit_app.py
   ```

2. How do I deploy and maintain the dashboard?
   The dashboard is deployed via Streamlit, an open-source Python framework used to produce dynamic applications efficiently. To deploy the dashboard, it is hosted directly through **Streamlit Community Cloud** by connecting your **GitHub account** and respective **repository** and specifying the main app file (`code/streamlit_app.py`). However, if the app becomes inactive or goes to sleep due to inactivity, you can **simply visit the dashboard link to wake it up and restart the app.**

3. How do I update the dashboard?
   Because the dashboard is connected to your GitHub project, any time you make changes to the app or its files and save them on GitHub, the dashboard will automatically update with those changes. You don't need to manually re-upload anything — it stays in sync with your project.