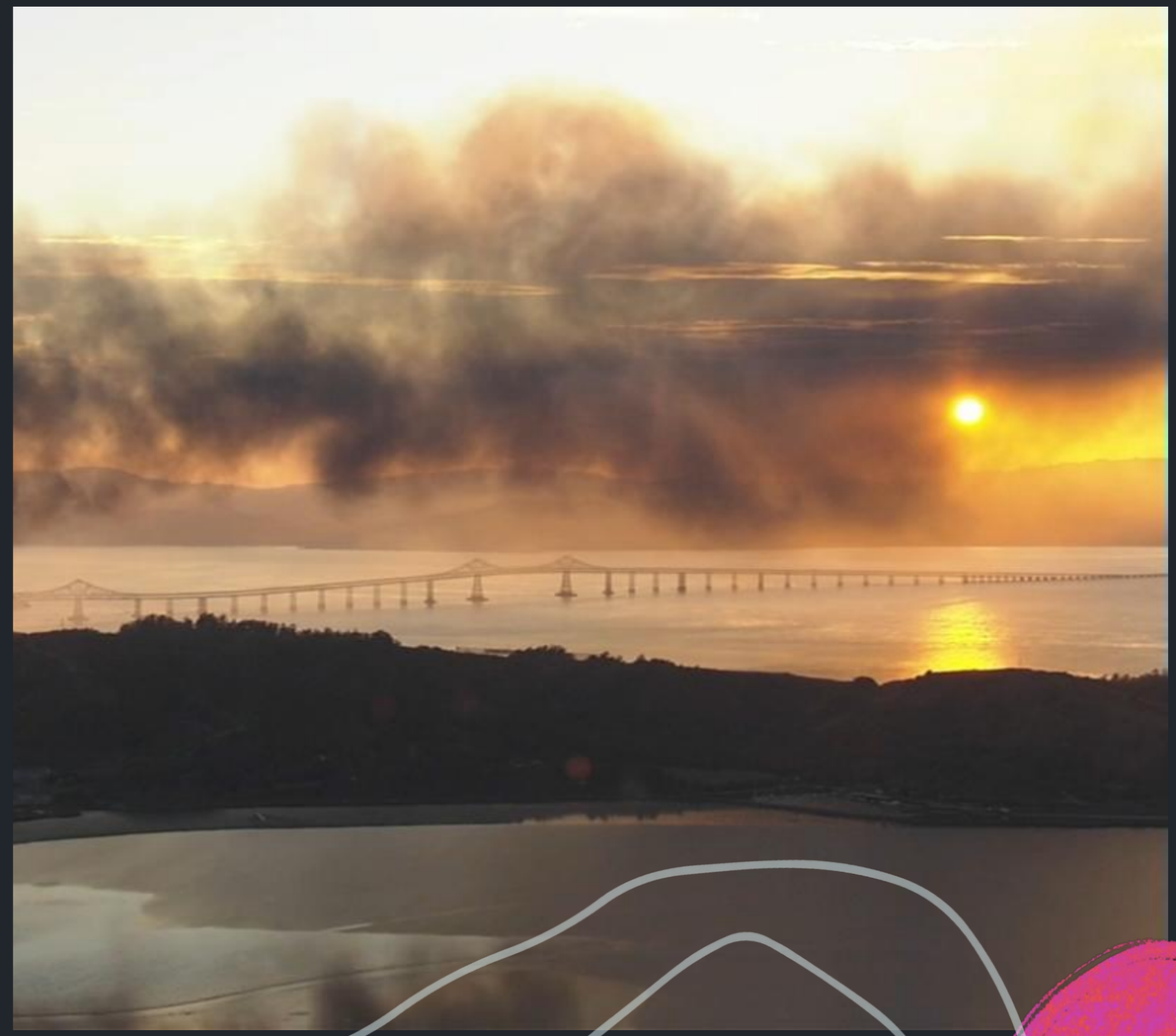
Air Pollution in the SF Bay Area



Group 3: Celina Kamler, Chris Kilkes, Divya Govil,
Jonathan Gonzalez, Meredith Jones, & Palash Raval

Overview

Topic:
2020 AQI & Income in Bay Area Cities

Intended Audience:
Residents of the SF Bay Area, policymakers, and leaders at large Greenhouse Gas polluters

Data Sources:
Environmental Protection Agency (EPA)
Air Quality System API
ACS Census Data from 2020

Topic background:
Environmental pollutants tend to impact communities of color and low-income neighborhoods disproportionately; asthma rates are nearly double in these areas

Disparate Health Impacts

OAKLAND HILLS

Childhood Asthma Cases

up to **20%** new cases from air pollution

Population

LESS THAN 30% people of color

WEST/DOWNTOWN OAKLAND

Childhood Asthma Cases

up to **50%** new cases from air pollution

Population

MORE THAN 70% people of color

source: https://www.edf.org/airqualitymaps/oakland/health-disparities

Our Process

01

ETL in Python

- Retrieved air quality data for various pollutants from the EPA's Air Quality System API for the Bay Area in 2020
- Merged the data with income data
- Aggregated the data into the max pollutant level for each city and date

02

Plotly Visualizations

- We utilized Plotly to create interactive visualizations that provide insights into the relationship between pollution levels and median income across different cities.
- Created line charts of the various pollutants across the different cities over the twelve months

03

PostgreSQL

- Generated master pollutant dataframes for future reference
- Created PostgreSQL database using robust table structure
- Exported SQL database & table files with ERD, enabling project dashboard

04

Leaflet

- Created a map of the Bay Area with layers for median annual income and major pollution sources

05

Flask & Javascript

- Created an app using Flask through the app route
- The Flask application responds to requests at 2 main endpoints and 5 data endpoints
- The endpoint fetches data from the SQL database and does Plotly and dropdowns for the pollutant and month visualization across cities

Database Design

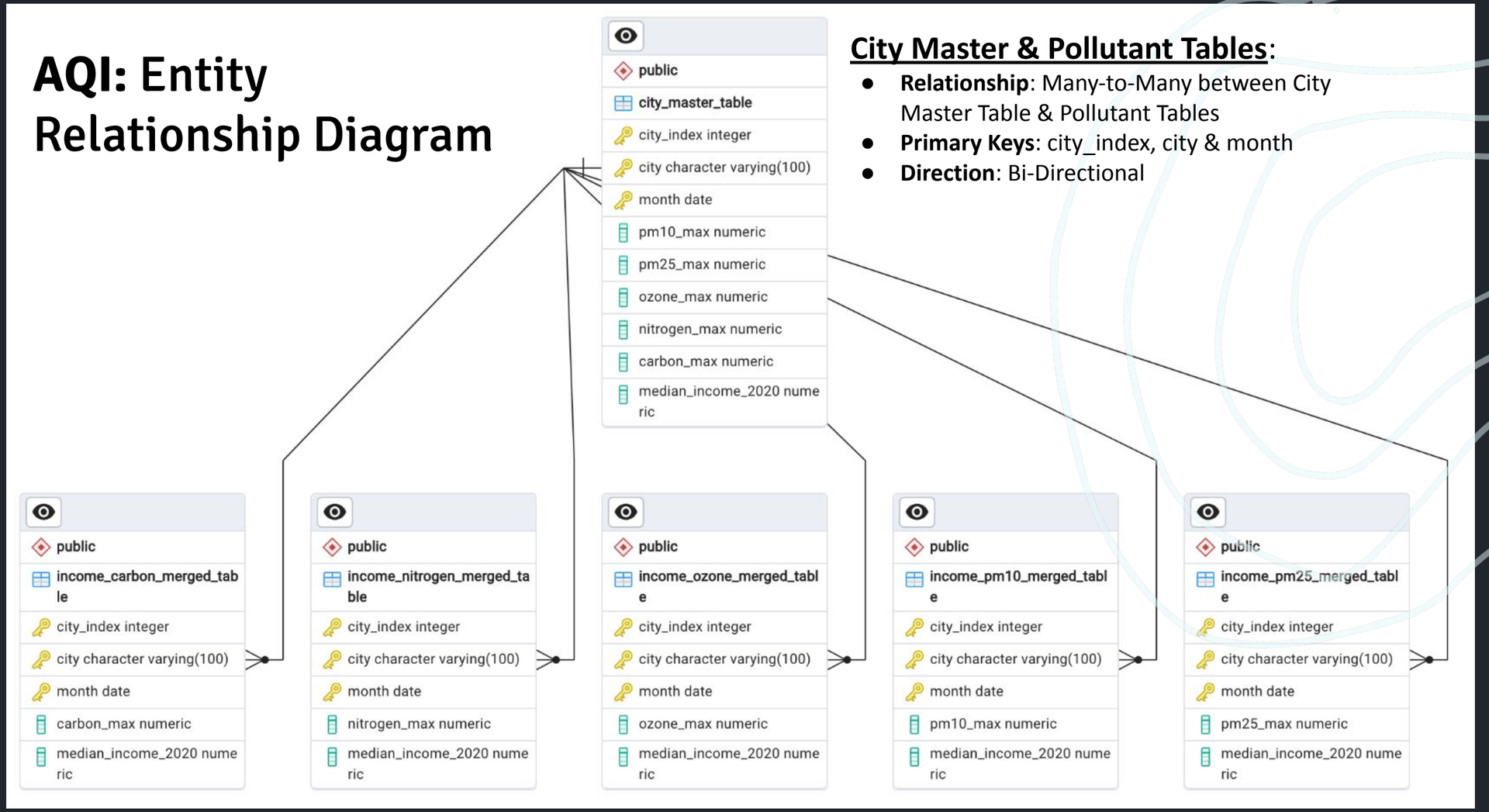
Purpose:

- Ingest, store and enable accurate pollutant and income data in SQL database for further analysis and use by Team #3 for dynamic visualization in the Air Quality & Income dashboard

Overall database design process:

- Clean pollutant data used post-ETL process using Pandas creating consolidated Bay Area pollutant data frame for storage and future reference
- Python used to create new consolidated .csv file providing a central referential SQL table for the new Air Quality & Income PostgreSQL database - maximizing integrity and utility
- pgAdmin used to create the new Air Quality & Income PostgreSQL database tying together the 1x new central City Master Table - holding the Primary Keys - with the 5x city & pollutant tables

AQI: Entity Relationship Diagram



City Master & Pollutant Tables:

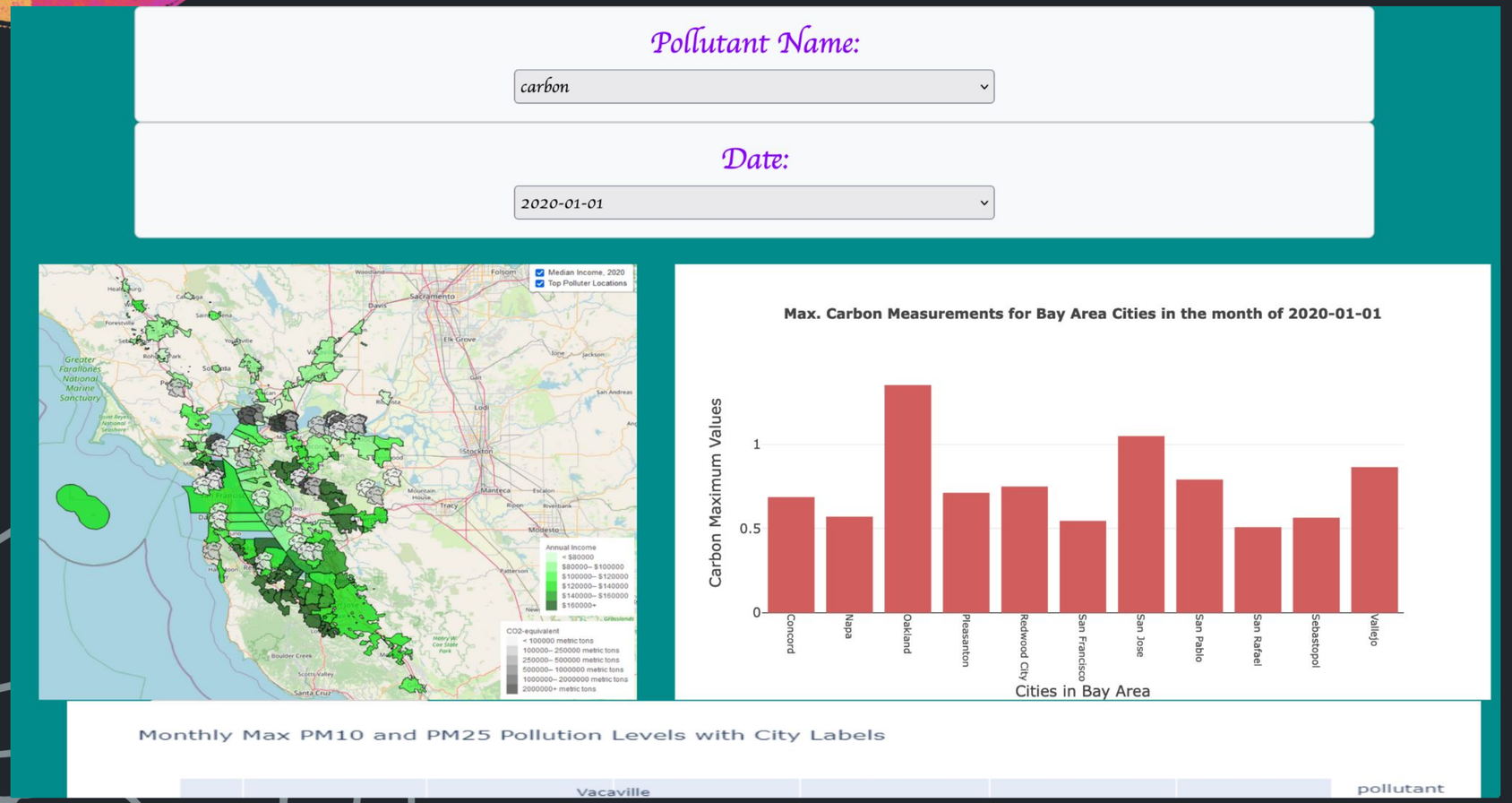
- **Relationship**: Many-to-Many between City Master Table & Pollutant Tables
- **Primary Keys**: city_index, city & month
- **Direction**: Bi-Directional

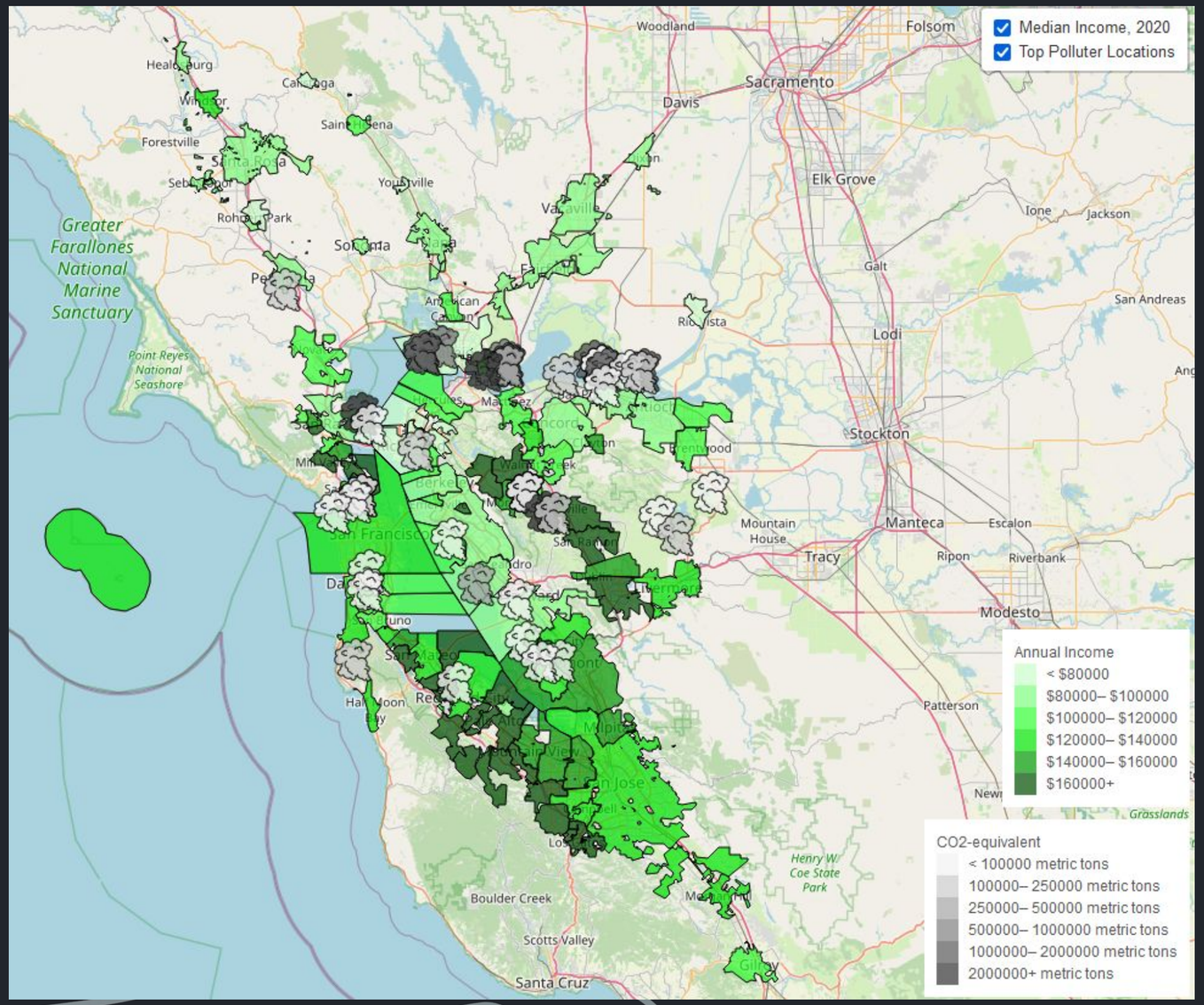
Overall data flow:

- City Master Table is built to serve as referential centerpiece to the Air Quality & Income Database enabling maximum data integrity and usability for long-term analytics use
- Each of the other five tables contain pollutant data for Bay Area cities

- “city_index”, “city”, and “month” fields serve as unique **Primary Keys** in the City Master Table, serving as connecting **Foreign Keys** in the other pollutant tables

Webpage Demo





Leaflet Map Demo

Data Wrangling/ Challenges

- Finding usable API data by city was challenging
- Flask: jsonify data for each pollutant to use d3.json for in JavaScript
- Leaflet: converted shapefile to geojson; converted geojson coordinates into a format Leaflet could use
- Getting dropdowns to work

Next Steps

- Expand geographic area to all of California
- Examine concentrated poverty (% of pop living below the federal poverty level)
- Get more complete AQI data for all neighborhoods and all cities

Art by Mark Fiore for KQED