April 4, 2024

**Air Quality in NYC in Relation to ENJ Communities Before After and During 2023 Canada Wildfires**

**Clio Bate**

**OBJECTIVES**

I will assess air quality in New York City using GEE’s Aerosol Index and Carbon Monoxide (CO) Concentrations from the GEE Catalog. I would like to compare the air quality during the period of Canadian Fires last Summer to the same period one year prior. According to [Yale School of Public Health,](https://ysph.yale.edu/news-article/canadian-wildfire-smoke-associated-with-increased-asthma-cases-in-nyc/#:~:text=Looking%20at%20data%20from%20June,June%206%20to%20June%208.) the worst period for NYC was between June 6th and June 8th. Therefore, I will focus my analysis during that time frame. I will use Cloud SQL with Google Cloud Console to perform these spatial queries: I also want to use NYC census blocks to identify who was affected the most severely (in terms of race/ethnicity, income, language, etc.). I will also compare these areas (or maybe the “recovery” also) to that of vegetation (NDVI also from GEE and parks boundaries).

I know this is a big undertaking, but these are skills I would really like to strengthen, I am very interested in this topic, and am willing to devote a lot of time to this project!

**ASSIGNMENTS**

Complete the following assignments. Deliverables will include pushing to your public GitHub and updating the README at every step.

**Project Proposal**

1. **Fill in the highlighted portions of this prompt. Utilize the Group Prompts for inspiration.**
   * Propose at least 5 vector datasets and 2 raster images that will address your topic.
   * What relationships will you analyze? Propose at least 3 spatial queries.

**2. Create a new Final Project repository and invite Jon & Kunal to collaborate on GitHub.**

**Due Friday, April 5 @ 5 pm (10 Points)**

**Assignment 1 – Data Acquisition, Processing, & Database Setup**

1. **Find and Process Geospatial Data**
   * Acquire data for at least 5 vector layers & 2 rasters:
     + **Vector Data**
       1. City boundaries
       2. Parks
       3. US census data
       4. [Neighborhood data](https://data.cityofnewyork.us/City-Government/2010-Neighborhood-Tabulation-Areas-NTAs-/cpf4-rkhq)
       5. [Borough boundaries](https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm)
     + **Raster Data**
       1. [UV Aerosol Index (GEE)](https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S5P_NRTI_L3_AER_AI)
       2. [CO Concentrations(GEE)](https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S5P_OFFL_L3_CO)
       3. NDVI- or other vegetation index of nyc (to get street trees(?))- I have scripts from a previous class I could adapt.
   * Be sure to provide sources, descriptions, and visualizations in your README.
2. **Set Up Database Schema**
   * Create schema for your chosen topic.
   * What attributes should you be mindful of?
3. **Pre-process the Data**
   * Process the data to align different datasets temporally and spatially.
   * Be sure to capture the details in your README.

**Due Friday, April 12 @ 5 pm (10 Points)**

**Assignment 2 – Import Spatial Data & Normalize Tables**

1. **Import your data into PostgreSQL tables/schema created in Assignment 1.**
2. **Normalize your tables (1NF up to possibly 4NF, depending on your data) and explain the logic in your README.**
   * Even if normalization is not required, explain why in your README.

**Due Friday, April 19 @ 5 pm (20 Points)**

**Assignment 3 - Spatial Queries & Presentation**

Perform spatial analyses to determine:

* SELECT the CO and Aerosol Index over the parks/vegetated areas and calculate the average concentration/ intensity over these areas. Do the same for Each neighborhood – see how those areas compare to pre/during/post values
* Use SQL to identify areas of income below a certain level and above a certain level
* Identify neighborhoods that have with languages other than English spoken at home over 25% (this number may change)
* See if there is a correlation between race, income, ethnicity, languages and concentrations of CO or aerosol

**Spatial Analysis & Presentation are Due Thursday, April 25 @ 10:15 am (40 Points)**

**Final GitHub Repo & README are Due Friday, May 3 @ 5 pm (30 Points)**

**Total: 110 Points**

**NO LATE SUBMISSIONS ACCEPTED AFTER MAY 3 -- Plan accordingly.**