Group Project 3: Air pollution for the top 10 most populous cities in 2022.

Team members: Alexandr Gorbulin, Abigail Janvier and Aida Ndiongue

Github Repository: <https://github.com/Gorbulin1989/Team-Project-3>

Purpose:

This app is looking into air pollution for 10 major cities (by population) across the world from 01-01-2022 to 01-01-2023 in order to assess how it might be affected by population, precipitation and wind speed. We will be focusing on the levels in the air of 8 common pollutants such as: carbon monoxide (CO), nitric oxide (NO), nitrogen dioxide (NO2), trioxygen (O3), sulfur dioxide(SO2), particulate matter 2.5, particulate matter 10 and ammonia (NH3) in each of the top 10 cities in the world, and the weather recorded in those cities at the time of pollution levels collection.

1. Discuss your dataset(s).

We decided to collect data from several different sources to maximize the amount of variables to include in our project. We used openweathermap API to generate data for the air pollutants and quality index by city using coordinate data found from a simple google search. This API yielded concentration values for CO, NO, NO2, O3, SO2, PM2\_5, PM10, and NH3 in μg/m3 per hour from 01-01-2022 to 01-01-2023. It also included the city’s coordinates and air quality index ranging from 1 (good) to 5 (very poor) for each hour as well. Our second dataset came from historical weather API to collect temperature, precipitation, cloud cover, and wind speed data hourly for the same time period as the previous dataset per city. We obtained data for population from a google search as well.

* 10 most populous cities: <https://en.wikipedia.org/wiki/List_of_largest_cities>
* Air quality index and concentration of major pollutants: <https://openweathermap.org/api/air-pollution>
* Coordinates for cities: generic Google search
* Weather data: <https://open-meteo.com/en/docs/historical-weather-api#api-documentation>

1. Discuss your variables and your expectations on the relationships between those variables.

Variables:

City, population, air quality index, precipitation, wind speed, cloud cover, CO, NO, NO2, O3, SO2, PM2\_5, PM10, and NH3 concentrations.

Relationships:

1. The concentrations of the common pollutants are directly related to the air quality index. The higher the concentration, the poorer the air quality and vice versa.
2. We expect wind speed to have a negative correlation with air quality index, meaning that the higher the average wind speed, the better the air quality (AQI of a low value) and the lower the concentration of pollutants.
3. We expect population to have a positive correlation to AQI meaning that the higher the population, the worse the air quality (AQI of a high value) and the greater the concentration of common pollutants.
4. We expect that as precipitation increases, the concentration of common pollutants will decrease thus improving the air quality index.
5. We expect that regions with greater cloud cover will present lower air quality indexes and higher concentrations of common pollutants.
6. Based on the expectations, what are the main questions you are asking?

Questions:

1. Does population size play a role in the air quality of a city?
2. Is there a significant relationship between precipitation, wind speed, cloud cover, and the air quality index of a given city and what is that relationship?
3. Will different time periods in the year show varying levels of pollutants and air quality indexes for the same city?
4. What are the most polluted cities and how does population size and weather contribute to their AQI?
5. Based on these questions, what sort of charts and visualizations are you planning to have? (You are expected to generate between five and ten different charts in the project and get at least three of them to have different view types (interactions)).

We are planning on having the following graphs to depict the relationships between;

1. population and pollutant levels: a side by side horizontal bar graph showing all cities on the y-axis and population and pollutant levels on the x-axis, with a drop down menu to select an index to look at
2. a set of line plots for all indices, precipitation and wind speed versus a timeline to see how they all directly correlate with each other. This graph will also have a drop down menu for a list of cities.
3. A bar graph summarizing all pollutants for a given city over the dataset timeframe.
4. 3 AQI speedometers/gauges for the top 3 most polluted cities
5. Finally a map of these 10 cities with circle markers.