11/18/2022
Anvinh Truong
Jia Cong Lin
Daniel Aguilar-Rodriguez

Air Quality in NYC

Motivation

Describe the problem you are trying to solve and/or the question you are trying to answer.

Regression Question: What will be the air quality in the future of New York City if traffic were to increase, decrease, or stay the same?

The problem we are trying to solve is to inform people of the impact that traffic has on air quality. This is an important issue to solve because it will show communities an indicator of how healthy their surroundings are and what it can be in the future. This will allow people to have an idea of how and when they should act to maintain a healthy community because bad air quality may limit the ability to be physically active and people with preexisting medical conditions such as asthma, emphysema or COPD are at greater risk from poorer air quality.

Data

Describe the data that you are going to use. Where did this data come from? How are you going to access it? Why did you choose this dataset? Also include a data dictionary for the raw data.

We are going to be using air quality and traffic in New York City data. The data regarding air quality and traffic were found through NYC Open Data and more datasets will be researched as we progress through the project if necessary. The data will be accessed by exporting them as csv files and those files will be used and manipulated. These datasets were chosen because they had relevance in the question we are trying to solve and they are targeted in New York City as well. The data dictionary for the raw data is described in the links below:

- Air Quality:
    - https://data.cityofnewyork.us/Environment/Air-Quality/c3uy-2p5r
- Car Traffic:
    - https://data.cityofnewyork.us/Transportation/Automated-Traffic-Volume-Counts/7ym2-wayt
- Mobility:

- https://www.google.com/covid19/mobility/

Plan

Describe at least one approach you will try to answer your question.

We are trying to show a correlation between traffic and air quality and the impact that traffic has on air quality so we are answering a regression question and we can create a linear regression model to help with our findings. These findings will also help predict air quality when there is a fluctuation of traffic in different parts of New York City. We will also try to answer our question by organizing it to be more user friendly followed by constructing graphs to outline the change in both traffic and air quality. It will determine if there is any correlation between traffic and air quality and if true, we can answer our question regarding traffic and the impact it has on air quality.

Challenges

Describe any challenges you anticipate while working on the project.

Some challenges that can occur while working on the project is to understand the raw data itself and see what is useful and what is irrelevant to our question. There can also be possible similarities to compare and merge between the datasets as well. This will also be our first time manipulating data using code to answer a particular question so that in itself will possibly bring challenges regarding efficient ways to scrape the data and clean it up.

Ethical Considerations

Who is being centered? Who is being marginalized?
What harm might this project cause?

The center of this project is going to be focused around New York City, and with traffic contributing to air pollution, there are concerns to the health of the citizens living in New York City. We will be showing the possibility of how cars are having a negative impact on the environment and how NYC as a whole is being a huge contributor because of the amount of traffic that's going on.