

# Covid vs. Pollutants in New York State

Vrishti Jain  
 Priyanshu Tripathi  
 Harsh Sugandh  
 Abhishek Gupta  
 Michael Morrison

## Introduction

Our goal is to verify if there is correlation between Covid-19 data and the air quality Index of an area and analyze the trend if the correlation exists, specifically in some counties of New York State.

## Data Description

To capture the Covid-19 data, We collected the data from the New York State Govt NY: <https://github.com/owid/covid-19-data/blob/master/public/data/README.md>.

For air pollutant data we used the Air Quality Index Daily Values Report from the the United States Environmental Protection Agency: <https://www.epa.gov/outdoor-air-quality-data/air-quality-index-daily-values-report>.

The reason for choosing these datasets is that it is a comprehensive collection in both cases, with varied columns about covid-19 testing and results as well as each pollutant has a specific data sheet with various columns of AQI, latitude, longitude of the county and so on.

Also, this data is as authentic for New York state as possible, since both these belong to Government agencies.

Since, we wanted to get these authentic datasets, we searched upon various Github repositories and blogs to locate and capture them. The dataset files are downloaded in .csv and stored on Google Drive for a shared medium as well on Github for version control, analysis and project development.

## Hypothesis

**Hypothesis A:** There is some type of correlation between pollutants in the air before and after the lockdown took place in NYS.

**Hypothesis B:** The concentration of pollutants changed this year compared to previous years.

### References:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7314691/>  
<https://www.nature.com/articles/s41598-020-73197-8>  
<https://plotly.com/python/time-series/>

### Resources:

Covid Data: <https://github.com/owid/covid-19-data/blob/master/public/data/README.md>  
 Pollutant Data: <https://www.epa.gov/outdoor-air-quality-data/air-quality-index-daily-values-report>

## Model Development

We started off by exploring our covid dataset for new positives in various counties in NYS. We looked into the general distribution of the data to start off. We saw that in general, there was an increasing trend in the number of new positives all throughout the year of 2020. We then performed our exploratory analysis on the pollutants dataset where we explored the data for all the pollutants namely CO, NO2, Ozone, PM2, and SO2. After exploring each of the two datasets individually, we merged these two datasets using the basis of counties and pollutants present in that county. We joined our two datasets using an inner join to get a dataframe.

After merging and performing initial preprocessing on our data, we looked into the daily concentration of various pollutants corresponding to various counties, using box-plots. We then selected a particular pollutant, namely NO2, for our further analysis. We calculated the daily AQI value and average NO2 concentration for all the months from January to February for various counties.

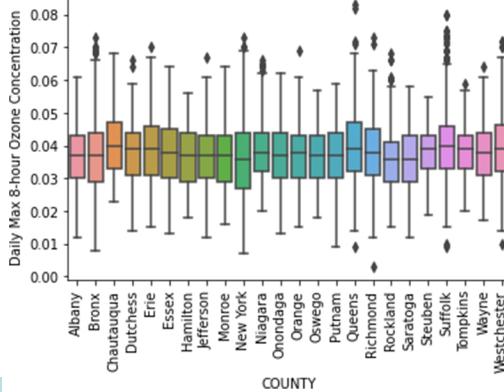


Fig 1: Daily Ozone Concentration

### For Albany AVG by month

Month	AVG DAILY_AQI_VALUE	Avg Ozone
January	26.344828	0.028448
February	31.896552	0.034517
March	35.516129	0.038452
April	39.300000	0.042400
May	37.032258	0.039968
June	38.933333	0.041200
July	35.612903	0.038484
August	33.064516	0.035742
September	31.466667	0.033967
October	26.838710	0.029065
November	27.566667	0.029800
December	25.333333	0.027333

Fig 2: AQI Value and Average NO2 - Albany

## Results

We found that there was a decreasing trend in NO2 concentration from January to December.

To further validate our hypothesis, we looked into the NO2 concentration for previous years and found out that there was a decrease in the amount of pollutants from previous years. This helped us conclude that lockdown coming into effect from around March in NYS helped reduce the concentration of various pollutants in the air. We were also able to verify that lockdown had a greater impact on larger and more populated counties. The smaller counties did not see a significant change in the pollutants concentration throughout the year.

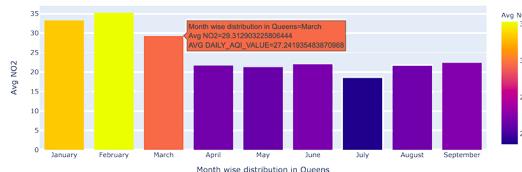


Fig 3: Monthly Concentration of NO2 - Queens

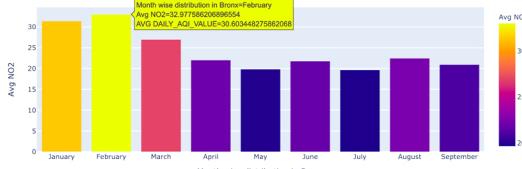


Fig 4: Monthly Concentration of NO2 - Bronx

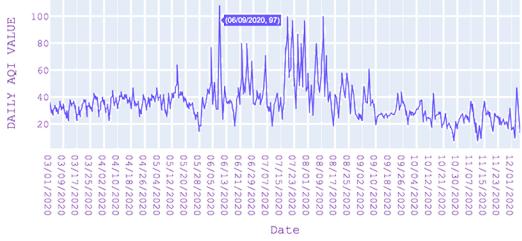


Fig 5: Daily AQI Change - Bronx

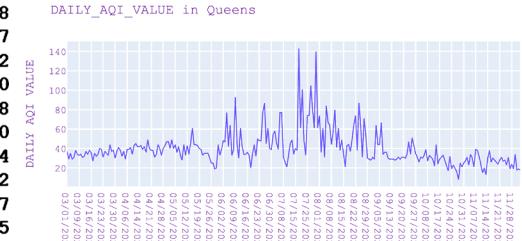


Fig 6: Daily AQI Change - Queens