**Analysis and Conclusions:**

Questions: What counties in NJ show higher levels of air pollution? Why do these areas have the high rates of pollution that they do?

Are there counties that have higher mean AQIs and max AQIs than the other across a 5 year period? What are those counties? Work and Visualization Done: Calculated mean and max/min AQI scores for each county in NJ between 2018-2023 to determine which counties had scores higher than the others. Selected 7 counties to further analyze. Visualized with bar charts (Mean AQIs for each county across 5 year period; max AQI for each county across 5 year period ). Answer: The 7 counties that seemed to have higher AQI scores were the following: Bergen, Essex, Camden, Gloucester, Hunterdon, Middlesex, Mercer.

Do these counties show a notable change in their AQI day categorization over time? Are there any perceivable trends? Work and Visualization Done: Calculated count of each type of day (ex. Good, moderate, etc.) for each of 7 counties across a 5 year period. Visualization: line plots across 5 year period for 7 selected counties to track count of each AQI score category for each day that year. Answer: Even the 7 selected concerning counties have by far more good and moderate days that the other more concerning category days. The trends for these 7 counties also seems similar to the overall trend for the state of NJ. Multiple of the 7 counties did not have high level concern (ex. Very unhealthy and hazardous) days. Most of the 7 counties (aside from Essex and Gloucester) had uptick in Good days from 2021 to 2022. It should be noted that some counties like Essex did not have data for 2023 and others did not have full 365 days for 2023.

Is there a relationship between AQI scores and the seasons/weather? Work and Visualization Done: We created calplot heat map across 5 year period for the high AQI counties to see if there were trends seasonally for the high AQI scores. Answer: We noted that high AQI scored days tended to be around the summer months, which would make sense as ozone levels also tend to be higher in high temperature days.

Are specific parameters contributing more to the AQI scores than others? (ANOVA test for stats significance) Work and Visualization Done: To visualize this, we created a boxplot of all of the AQI’s by defining parameters with a text of the p-value score. We also displayed a bar chart for the count of the parameters used per county to show which parameters were seen in the data more than others. Answer: We did an ANOVA test on the defining parameters on AQI. We grouped the parameters throughout all of the years and found a p-value that was significantly small enough to prove that the means of all of the AQI for each parameter is not the same. We also found that the counts of each defining parameter for each county was vastly different. The parameters PM2.5 and ozone were used the most by a large amount. Despite this, we found that a lot of the smaller AQI values produced were from parameters that had smaller counts like CO and NO2. From this, we can conclude that parameters such as PM2.5 and Ozone contribute moreso to higher AQI scores than other parameters, specifically PM2.5.

Are there any defining pollutants that are more dominant in specific counties? Work and Visualization Done: Created bar chart to see distribution of each defining pollutant per county for each year Answer: Ozone and PM2.5 seemed to be more dominant defining parameters used for high AQI counties

What are the parameters contributing to unhealthy air quality days in NJ? Work and Visualization Done: For the 6-years NJ AQI data we analyzed we counted the defining parameters i.e., count that contributed to the unhealthy, unhealthy for sensitive groups and very unhealthy AQI values. Answer: We noticed that ozone used to be the primary contributor for the unhealthy days before 2019, but noticing that there is a rise in PM2.5 values since the recent years. It appears that the trend is shifting.

Are there any parameters that impact the minimum AQI values more than the other pollutants? What are they? Work and Visualization Done: Created a bar chart of the count of the parameters that produced the minimum AQI values in a year. Answer: Despite being used the least, CO produced a lot of minimum values along with NO2. PM2.5 and Ozone had instances of producing small AQI values, but given that they were responsible for also producing the maximum AQI values we can conclude that they do not have an impact on minimum AQI values. We can conclude that CO contributed the most towards minimum AQI values and is a strong indicator of healthier days.

Are there any potential contributing factors to AQI scores we are observing? (ex. Nearby factory emissions) Work and Visualization Done: We used dataset with data on greenhouse gas (GHG) emissions (ex. CO2 and its equivalents) from large factories to determine what counties in NJ seemed to contribute more to GHGs in the state and how many factories the top 3 emitting counties had. Answer: The top GHG emitting factory locations seem to coinicde with our air pollution list, aside form Union. Despite Union have the highest score for GHG emissions but having less factories than the other top 2 GHG emitting counties, we can infer that factory count alone is not the only contributor to GHG emission score. We also want to note that GHG emissions are part of the picture in terms of air quality, but are not necessarily the top pollutants affecting AQI (ozone, PM2.5) in our EPA dataset.

What could be one area that air pollution impacts NJ? (ex. Agricultural emissions creating more cost) Work and Visualization Done: Looked at cost per emission for agricultural industry sectors in NJ, shown with pie chart or the most and least affected cost-wise. Answer: Air pollution does not only affect us in seemingly more obvious ways, but there are other figurative and literal costs.

Final Observations: Amongst 16 NJ counties in dataset, 7 counties had mean and max AQIs above the others. Overall Good and Moderate and relatively less unhealthy days. There seems to be relationship between Defining Parameters and AQI scores Ozone and PM2.5 highly contributed to defining parameter count; PM2.5 seemingly more relevant over time than Ozone as PM2.5 contributed more to unhealthy category. Most of the healthy days were during summertime; potential to dig deeper (ex. Temp, human activities, fires, etc.) Some counties with top GHG emissions did not appear in top AQI score list, likely due to type of emission Though not cause for major immediate concern now, still keep an eye on air quality as climate change is still a concern

Advancements: Potential health impacts of air pollution in NJ (ex. Lung-related illness rates) More potential influencing factors to AQI score in NJ (ex. Transportation, population density, migration within NJ, proximity to airports) Explore other areas in USA with concerning air quality and potential contributing factors. Predictive modeling eventually (machine learning)

Sources: <https://aqs.epa.gov/aqsweb/airdata/download_files.html#AQI> <https://aqicn.org/api/> <https://aqicn.org/json-api/doc/> Greenhouse Gas Emissions from Large Facilities (up to 2022) - <https://ghgdata.epa.gov/ghgp/main.do#/facility/?q=Find%20a%20Facility%20or%20Location&st=&bs=&et=&fid=&sf=11001100&lowE=-20000&highE=23000000&g1=1&g2=1&g3=1&g4=1&g5=1&g6=0&g7=1&g8=1&g9=1&g10=1&g11=1&g12=1&s1=1&s2=1&s3=1&s4=1&s5=1&s6=1&s7=1&s8=1&s9=1&s10=1&s201=1&s202=1&s203=1&s204=1&s301=1&s302=1&s303=1&s304=1&s305=1&s306=1&s307=1&s401=1&s402=1&s403=1&s404=1&s405=1&s601=1&s602=1&s701=1&s702=1&s703=1&s704=1&s705=1&s706=1&s707=1&s708=1&s709=1&s710=1&s711=1&s801=1&s802=1&s803=1&s804=1&s805=1&s806=1&s807=1&s808=1&s809=1&s810=1&s901=1&s902=1&s903=1&s904=1&s905=1&s906=1&s907=1&s908=1&s909=1&s910=1&s911=1&si=&ss=&so=0&ds=E&yr=2022&tr=current&cyr=2022&ol=0&sl=0&rs=ALL> Supply Chain Greenhouse Gas Emission Factors v1.2 by NAICS-6 - Catalog (data.gov) - <https://catalog.data.gov/dataset/supply-chain-greenhouse-gas-emission-factors-v1-2-by-naics-6>