

Final Progress Report

Date: November 28, 2021

Accomplishments:

- Reworked several of the visualizations in PM2.5, PM2.5 part 2, Ozone, AQI, and Asthma studies and uploaded the modified Jupyter notebook files to the GitHub project repository at https://github.com/mjlaw/FIT_capstone.
- Completed the analyses of air quality and asthma at the U.S. state level for 2012. Air quality features were mean PM2.5, mean ozone, median AQI, max AQI, days CO, days NO2, days Ozone, days SO2, days PM2.5, days PM10, unhealthy for sensitive group days, unhealthy days, very unhealthy days, and hazardous days. Created a correlation matrix with a heatmap and scatterplots to visualize the relationship between air quality and asthma. Uploaded the Air Quality and Asthma Jupyter notebook to the GitHub project repository.
- Completed the analyses of air quality and asthma at the county level for two states with a high prevalence of asthma in 2012—Vermont and Kentucky. These states report asthma in two- or three-year periods. The asthma datasets I used were for 2011-2012. Therefore, I cleaned daily PM2.5 datasets, daily ozone datasets, and annual AQI by county datasets for 2011 to add to the previously cleaned and saved 2012 datasets. Air quality features for Kentucky were mean PM2.5, mean ozone, good days, unhealthy for sensitive group days, unhealthy days, max AQI, median AQI, days NO2, days Ozone, and Days SO2. (Other features were not included because the values were all 0 or a max of 1.) Created a correlation matrix with a heatmap and a scatterplot. Air quality features for Vermont were good days, unhealthy for sensitive group days, max AQI, median AQI, days CO, days NO2, days Ozone, and days SO2. Uploaded the Air Quality and Asthma part2 Jupyter notebook to the GitHub project repository.
- Uploaded zip files for annual AQI by county for years 2011, 2017, and 2018 and for daily PM2.5 and daily Ozone for years 2011, 2017, and 2018 to the GitHub project repository Original datasets folder.
- Created a new Jupyter notebook for Air Quality and Asthma part3, which will examine the relationship between air quality and asthma in California. Read in daily PM2.5, daily ozone, and annual AQI files for years 2017 and 2018. Limited the datasets to California. Cleaned the PM2.5 and ozone datasets. Uploaded the Air Quality and Asthma part3 Jupyter notebook to the GitHub project repository.
- Uploaded the Final Progress Report to the GitHub project repository.

Current Activities: I am currently working on a new study on air quality and asthma in California and finalizing visualizations.

Challenges: Working with limited data is challenging. Vermont had three counties with PM2.5 data, two counties with ozone data, and four counties with AQI data. I did not use PM2.5 data or ozone data in the analysis because there weren't enough observations. The correlation matrices and heatmaps created thus far show very little correlation between any of the air quality features and asthma prevalence. Although I had initially intended to limit the scope of the

study to specific years—1999, 2012, and 2020 , I added the 2011 datasets because states report asthma in two-year periods. I had also planned to look at asthma prevalence in a state or states having hazardous days. I could not find asthma reports for those states in my target years. After further research into states with the largest numbers of counties reporting air quality data, I decided to examine the relationship between air quality and asthma in California. The most recent reports for asthma in California are for 2017-2018, necessitating the download and cleaning of additional datasets. The map visualizations are taking the largest amount of time and are the most challenging for me. I'm working on adding layers to maps using geopandas and modifying the color/shape/size of color bars.

Work to be Completed: All remaining visualizations, the California air quality and asthma study, and the final project report.