Data sets for charts – the data will be merged as follows:

Each City will have all 6 pollutants merged into a CSV	There will be 3 city CSV files for 2016, 2018, 20202	The columns and rows will be dropped
Each Pollutant will have the monitoring data for the six cities	There will be 3 pollutant CSV files for 2016, 2018, 2020	The columns and rows will be dropped
The data files can be merged or segmented more based on what visualizations are being used		

Data sets for machine learning – the data will be merged as follows:

Air Quality predictions in a specific city over the years 2016, 2018, 2020 – initial training may only take one year into account before trying to run all three years together to see what the training results are.	Predictor Variables are: Location Date Lead reading Carbon Monoxide reading Ozone reading Nitrogen Dioxide reading Sulfur Dioxide reading Particulate Matter 10 reading [150 ug/m3] means [150 micrograms/cubic meter] [added to the set if time permits] Seasonal date Location [taken from weather api if historical data can be captured] Humidity Temperature	Prediction: Air Quality will change based on the number of pollutants and their monitored volumes. A subset of this prediction is that 2022 air quality should show more pollutants present after everyone goes back to work.