Data Analytics/Visualization Bootcamp
Project 2 Technical Report
Team 5: Kristin Scholten, Jordan Morales, Hope Youngblood

The objective of the study using our extracted, transformed, and loaded (ETL) data was to examine the relationship between wildfires and air quality in California. We accessed the 1980-2021 Air Quality Index (AQI) from the EPA and the California WildFires (2013-2020) datasets, both of which were retrieved from kaggle.com. The data was extracted using Pandas in Jupyter Notebook.

The AQI datatset was transformed to remove all states except California and dates from 01-01-2013 to 12-31-20 because this was the date parameters of the California wildfire data. The California WildFires dataset was kept intact so that the analytics team could use as many of the variables (e.g., acreage, number of crews involved, amount of water used, length of burning time) as possible to determine the role each may have played in affecting the AQI.

The data was then converted into datasets and loaded into SQL. The analytics team has the option of using a number of corresponding variables between the datasets, including longitudinal and latitudinal coordinates and county name to determine where the wildfires were located and how it affected the AQI in the surrounding area. Some of the possible analyses that could be performed include health effects on crews that are exposed to wildfires and the resultant poor air quality, determine how rapidly the area surrounding the wildfires returned to baseline after it was extinguished, comparison between acres burned and AQI, and topographical effects on AQI.