Daniel Vogler

DATA 512

Collaboration Reflection Statement

In the first phase of this project, I collaborated with others in the course to make important decisions about the most efficient tools to work with the data, how to filter it and pre-process it, designing my smoke intensity metric, and how to approach predictive modeling.

Collaboration helped me select the right tools for different tasks required to complete this assignment. For example, Apoorva Sheera and Manya Chadha advised me to use geopandas because this package is optimized for dealing with geoJSONs with geometry objects. Taking this advice allowed me to efficiently work with dataframes, which simplified the work required to merge datasets across different tables.

I also worked with others to resolve perceived ambiguities in the assignment specification and make judgements about the most effective ways to pre-process and filter the data. For example, I discussed with Ed Seryozhenkov and Sid Gurajala whether to filter down the dataset to only wildfires or to include prescribed burns.

Furthermore, Ed, Sid and I collaborated to develop out smoke intensity metrics. We discussed whether to use centroid or edge distances; models that assumed a Gaussian distribution of smoke around the point source; and justifications for settling on an inverse-square law. This was particularly helpful. Even as we each implemented out metrics in slightly different ways, the act of discussing them forced us to clarify our thoughts, express them in clear equations, and justify one modeling choice over another.

Finally, it was helpful to consult collaborators for the predictive modeling aspect of this assignment. Sid Gurajala and Alex Netzley shared with me that traditional, non-seasonal ARIMA models had yielded poor results in their trials. This led me to scrutinize trends in my smoke intensity score and examine whether a *seasonal* model might be more appropriate. This in turn led me to choose a SARIMA model.

In conclusion, the collaborative efforts in this project significantly enhanced the depth and quality of our analyses, and forced me to clearly communicate my reasoning for certain analytical decisions. By engaging in discussions with my peers, I was able to navigate the complexities of data wrangling and predictive modeling more effectively. The insights shared by Apoorva, Manya, Ed, Sid, and Alex not only influenced the tools I selected but also helped me develop a clearer and better-justified smoke intensity metric and predictive model.