Reviewer Response, Gillian Grasso and Jackie Katsirubas

Given the reviews of the final project notebook provided by Devon and Giovanni, one improvement to the notebook that Giovanni suggested after our presentation and that Devon also touched on in her peer review was to incorporate code for both the filtered Boston and Denver datasets that does not isolate the 6AM time. By including data from all times of day for the filtered datasets, this increases the amount of data points and thus produces more statistically significant cluster plots. We included code that accomplishes this directly after the set of cluster plots specific to the 6AM filtered Denver data. This recommendation produced cluster plots with much more data and helped to further define the clusters identified in the 6AM filtered data. However, an interesting consequence of including all times of day in the cluster plots is the unintentional vertical lines that appear in the Denver plots; we are unsure of why this occurs.

A second adjustment to the notebook offered by Devon in her peer review is the addition of code that produces cluster plots with 3 clusters instead of 4. This is included beneath the code that generates cluster plots for all times within the filtered Boston and Denver datasets and uses these particular dataframes to produce the new cluster plots. Although 4 clusters seem to fit both datasets, it is interesting to visually compare the statistical distribution of datapoints when 3 clusters are used.

Partner Collaboration

Throughout this project, we collaborated to determine the most useful code to include in our project notebook from previous Application Labs as well as to write our own code that would efficiently filter both datasets. Jackie's background in statistical analysis was helpful in understanding the significance behind the clustering analysis and Gillian was able to provide background in Python coding to work through many of the coding issues we ran into while filtering the data.