Team Mini Project: Alternative Fueling Locations

Ashley Ridley, Ty Johnson

DS450-01

Data Science Senior Capstone

There are various tools available for data visualization which can help in presenting graphics in a simple manner. Tableau is one of the fastest growing data visualization tools that provides the most efficient way to manipulate or transform raw data into an easily understandable format. Since this tool allows analyzation of data at a quicker rate, we decided to use Tableau to generate dashboards and workbooks to show insights. Since this project asks for various graphs, maps and color-coded fuel types based on code, we used Tableaus dashboards to separate, color code and label each map to make it easier to visualize and understand.

To build the maps, we used latitude and longitude coordinates to display a map of the US. From there, if we were asked to use a map of Alaska that displays a dot for each AFL, we would drag states to the filter section and select Alaska, making Alaska the only state shown. Next, we dragged the fuel type code to filters and color in the marks section. This allowed us to display a different color to each AFL making it more distinctive and select which AFL we want to display on the map. To display the dots depending on what was asked, we used state, street address, and measure values in the details section to get a more accurate and in-depth picture of every AFL in the United States.

When cleaning and preparing the data, we used the Python script and pandas library to manipulate the data. The dataset included many columns that were irrelevant to the project, so we dropped them from the dataset to improve the ease of working with it and performance of Tableau. Since we would need a column for regions corresponding to each state, we created a function to create the column. It accepted the state as input, and referenced lists of states that were in each region. We saw this as a more efficient way than individual mapping each state to a corresponding region. There were a few entries that did not have a listed address, so we went ahead and filled this with “unknown”. In a more polished project, this would create a problem when trying to find the precise location of a fueling station and would require additional processing to fix this, but in our case this would suffice.

For the parts-to-whole graphs, bar charts were the main thing that we went with. Combining these with stacked columns using the color feature of Tableau, we were able to easily visualize the proportion that each fuel type held

relative to the category we were looking at, whether that be individual states, region, etc. We found that packed bubbles were another type of graph that was extremely easy to visualize. Whenever we were looking into the regional AFLs, we found that it useful that it showed the concentration of each AFL type in each state all put together. For example, we could see that ELEC was the biggest in most states and easily identify the smaller types per state such as LPG.

There were a few problems that we encountered when creating the visualizations. First, not all the data included accurate latitude and longitude or featured an incorrect city name. This especially became an issue when creating maps for the AFL location in each region. The way that we fixed this issue was individually excluding the data points that were incorrect It was manageable since there were only about 1-7 incorrectly plotted dots per region, but this could become an issue when dealing with even larger datasets that have these issues. One potential fix if we came across this in a larger scale project could be to cross-reference the address with a geolocation API. One issue that we came across, that could be easily fixed, is the disproportionate numbers of EV fueling stations in comparison to all other ones. When creating graphs that examine all the AFL types together, it makes it difficult to see the other ones. We typically created two versions of each graph, one excluding the EV fuel type to get a better idea of the other types. When looking into EV Networks, there are many different types, some of them being subcategories of each other (ex. Tesla and Tesla Destination). This presented a problem, especially for the legend on the side of the graphs, it was unpleasant to look at. We dealt with this issue by creating some graphs that looked at only the top X amount of EV networks in the areas of interest. Doing further research and combining similar ones together into related categories could help with a situation like this in the future.