Team Mini Project: Alternative Fueling Locations

Brandon Thomas, Kadin McWilliams

DS-450-01

Data Science Senior Capstone

Alternative Fueling Locations

**Toolset Choice**

Tableau was used for this analysis because it is a familiar tool that all group members have used in prior projects, ensuring an efficient workflow. It has a user-friendly interface and a particularly helpful drag-and-drop functionality that makes it easily accessible for beginning and experienced users alike. Tableau can also handle large datasets and visualize complex data through interactive dashboards and diverse chart options. It is especially helpful for its ability to handle geographic data well, as demonstrated by this project. Other programs like Python were not needed to clean and prepare the data. Tableau was able to handle any values that were illogical or bad data points, while performing our graphical exploration. Tableau can integrate with various data sources like Excel, SQL, and cloud services, streamlining the data importation process, which is often a challenge using other tools. Finally, Tableau is advantageous due to its emphasis on visualizations which are able to quickly convey pertinent information to stakeholders or business leaders who may not have extensive data analysis experience.

**Visualization Generation**

To generate the visualizations for the US\_AFL file, one method was used and then slightly modified to satisfy the requirements of each of the suggested visualizations. To start, the longitude and latitude data column dimensions were used as columns and rows respectively, generating a coordinate map of each combination of latitude and longitude that were automatically overlaid on a world map. However, only the continental US was requested so coordinates in Puerto Rico, Alaska, Hawaii, and any other erroneous data points that did not lie in the continental US were excluded by adding a filter on the state. Next, fuel type code was dragged into the marks section and differentiated by color so that each alternative fuel type has its own color. To generate the remaining visualizations for the file, the same process was followed, filtering by state for each individual state requested. For the generation of visualizations where each fuel type is displayed by itself, a very similar process was used—with one exception. To display all graphs on one dashboard, two calculated columns were created that manually specified the position of each fuel type location map. This created a table where the locations of each individual alternative fuel could be easily compared. This process was essentially the same for the regions, electric vehicles, and specific state data portions of the project. Filters were applied and relevant data was marked by color for differentiation.

**Part-to-a-whole Graphs**

The three main part-to-a-whole graphs that were used for this project were tree maps, bar graphs and bubble charts. The tree maps were used to show the distribution of fuel alternatives through each region of the United States as well as the distribution of electric vehicle distribution. The tree maps provided a clear visual of which fuel alternatives were most abundant in each region. The most common graph that was used were bar graphs. Bar graphs are commonly used and widely understood. These graphs were simple and effective in showing the number and distribution of fueling locations or electric vehicle network sites in individual states, regions, or the continental US. Bubble charts were added in pairing with bar graphs to have different visual representations other than bar graphs. These bubble charts made understanding the data more digestible. The combination of these three part-to-a-whole graphs were used to show distributions and quantities of fueling locations.

**Exploration Findings**

From exploring the data, we found that electric fuel was the most common and widely used alternative fuel source. While other sources did come close to competing with the electric fuel sources, electric fuel was the most common across all states and regions. The most common places for electric fuel sources to be located were in dense urban areas. A few trends that were found amount the different alternatives were that electric fuel, liquified petroleum gas and compressed natural gas were the most used fuel alternatives. Liquified petroleum gas and compressed natural gas were more commonly found in more rural states. In the Midwest however, E85 fuel is used the same amount as electric fuel. This is likely since the Midwest is a large producer of corn—which is used to generate E85 fuel—and is often referred to as the ‘Corn Belt.’ These two fuel sources account for roughly 70% of all fuel alternative in the Midwest. Hydrogen fuel was the least used fuel sources across the whole United States. Biodiesel and liquified natural gas were not as common across the United States. Hydrogen fuel was the least common fuel source by a significant margin with most regions of the country only having a handful of locations with most occurring in coastal states. In summary, electric fuel is the most used fuel alternative followed by E85, liquified petroleum gas and compressed natural gas. Other fuel alternatives such as hydrogen fuel, biodiesel and liquified natural gas were far less common across the whole country and largely insignificant.

**Problems Encountered**

The data used to generate the visualizations did not require much data cleaning or preparing. There were, however, some erroneous data points that showed latitude and longitude coordinates in Spain, the middle of the ocean, as well as in Puerto Rico, which was not the area of study for this project. Thus, these points were excluded from the analysis in Tableau. An additional problem faced getting the correct latitude and longitudinal coordinates for plotting the alternative fuel locations. There were already generated latitude and longitude points that were separate from the data set. This problem was fixed by changing the correct coordinate points to being dimensions and not calculated fields. The final problem encountered was that for three of the five regions, an outdated version of the data set was used. To resolve this problem, without having to redo all the work, a connection was made to the updated data set that replaced the fuel type code, latitude and longitude. This resolved the issue easily and additional work did not need to be done.