Multivariate Cluster Analysis of Electric Vehicle Charging Station Locations in Pittsburgh

Alexander Wolfe

**1. Introduction**

Electric Vehicles have continued their rise in popularity as technological advancements have unlocked improved battery durations, resulting in longer drive times and less frequent recharging for electric vehicles. Concern over harmful emissions, the health of the environment, and dependence on oil have also contributed to their popularity. As the demand for electric vehicles increases, additional infrastructure, in the form of Electric Vehicle Charging Stations, is needed to keep electric vehicles on the road. Unfortunately, not all neighborhoods in Pittsburgh are ideal locations for charging stations. This lack of suitability may be a result of economic hardships that have fallen upon residents in a neighborhood which results in fewer residents owning electric vehicles and thus, less demand for public charging stations.

Due to the high initial cost of purchasing an Electric Vehicle, identifying neighborhoods containing residents with high economic standing is important as these residents may be more likely to purchase electric vehicles, and therefore seek the use of charging stations.

It is important to consider that residents of neighborhoods who own their own homes may have the opportunity to charge their EV batteries at their house while residents who rent may be forced to rely on public charging stations. Following the assumption that higher-than-average income residents are more likely to purchase EV’s, renters who pay higher-than-average rent may also serve as a useful indicator of where EV charging stations are typically built.

While accounting for various economic and demographic factors present in each neighborhood in Pittsburgh, the relationship between these variables and the presence of existing electric vehicle charging stations was explored through Multivariate Clustering Analysis.

**2. Materials**

**The Pittsburgh Neighborhood Project Data:**

“Pittsburgh Neighborhood Project” Data was obtained by downloading from the ArcGIS online portal within ArcGIS. It contains demographic and economic information on each neighborhood in Pittsburgh. The variables included in the dataset describe the median home value of each neighborhood, the population, median rent price, family poverty rate, and several additional demographic and economic indicators (Cotter, 2023).

The data was originally downloaded as a WFL layer which allowed data tables to be exported from ArcGIS to Microsoft Excel for cleaning and transformation. Three variables in the dataset were originally formatted as percentages so they were converted to proportions to allow for more accurate analysis. These three variables were the percentage of residents in each neighborhood without a bachelor's degree, the percentage without access to a vehicle, and the percentage without internet access (Cotter, 2023).

All numeric variables retained and included in analysis were normalized in order to provide for more reliable analysis. This was achieved through z-score normalization. The result of the normalization process is that each variable will have a mean of 0 and a standard deviation of 1(Esri, 2024).

**Equation 1**

**Z-score Normalization Formula**

The cleaned and transformed data was added to ArcGIS as a standalone table.

**Neighborhoods Data:**

“Neighborhoods” is a Shapefile that includes geospatial references to each neighborhood in Pittsburgh and their boundaries (Pgh.admin, 2024). The “Neighborhoods” file was downloaded from the ArcGIS open data portal from within ArcGIS. To prepare the file for use, data from The Pittsburgh Neighborhood Project was joined to Neighborhoods in ArcGIS to link numeric data to the location of each neighborhood in Pittsburgh. Both datasets contained a unique neighborhood identifier which served as a key during joining to ensure data was joined to the appropriate neighborhood (Pgh.admin, 2024); (Cotter, 2023).

**Electric Vehicle Charging Station Locations Data:**

The locations of Electric Vehicle charging stations dataset was download from the United States Department of Energy Alternative Fuels Data Center Website. It includes geospatial coordinates of the location of each Electric Vehicle charging station in the United States (United States Department of Energy, 2024). The data was downloaded as a table and filtered in Microsoft Excel to exclude charging stations that are not located in Pittsburgh or its surrounding areas. Once prepared, the data was loaded to ArcGIS as a point layer.

**3. Analysis**

**Multivariate Clustering:**

K-means clustering is an unsupervised machine learning algorithm that classifies observations of data into clusters based on their feature variable values. Observations that are classified into the same cluster can be thought of as being similar or alike based on their feature values.

K-means clustering was performed using the joined Neighborhoods-Pittsburgh Neighborhood Project layer as the input field. This included eight separate normalized variables that correspond to each neighborhood: total population, median home value, median rent price, family poverty rate, single mother rate, proportion of residents without a bachelor’s degree, proportion of residents without internet access, and proportion of residents without access to a vehicle. Four distinct clusters were defined, and “Optimized Seed Locations” was chosen as the Initialization Method.

**Model Diagnostics:**

Included as output for Multivariate Clustering analysis as **Figure 1**, is a graphic that shows the “Optimized Pseudo-F Statistic Chart” for the k-means model. The F-Statistic is used in Multivariate Clustering to select the optimal number of clusters for the model. The optimal number of clusters corresponds to the highest peak on the chart, which was three for the clustering model (Esri, 2024).

**Figure 1**

A graph with a line

Description automatically generated

The k-means model was respecified to generate three distinct clusters instead of four. The model included the same variables from the Neighborhoods-Pittsburgh Neighborhood Project layer and the model output is shown below in **Table 1**.

**Table 1**

**K-means Clustering Output, k = 3.**

A screenshot of a graph

Description automatically generated

Median Home Value

Proportion without Internet Access

Proportion without Bachelor’s degree

Family Poverty Rate

Population

Median Gross Rent

Proportion without Access to a Vehicle

Single Mother Rate

Model Summary output in **Table 1** details each variable included in the Multivariate Clustering model along with calculated descriptive statistics. Perhaps most importantly, the R2 statistic gives the proportion of variance in cluster labeling that is explained by each of the variables included in the model. Put simply, the R2 statistic details how useful each variable is at separating each neighborhood into one of the three clusters (Esri, 2024). A closer look at the output in **Table 1** shows that the most useful attributes in determining which cluster a neighborhood fell into were:

1. Proportion of Residents Without a Bachelor’s degree,

2. Median Home Value in each neighborhood, and

3. Proportion of Residents Without Internet Access.

The output also indicates that Neighborhood Population and Median Gross Rent were the least helpful.

**4. Discussion**

A map of a city

Description automatically generated

As shown by the map of clustered neighborhoods (**Map 1**), it is evident that many of the existing locations of EV Charging Stations fall into areas classified as Cluster 2 or border upon neighborhoods within this cluster. It is apparent by **Map 1** that there is a relationship between the locations of existing Electric Vehicle charging stations in Pittsburgh and the economic and demographic characteristics of neighborhoods where the stations are located. Furthermore, **Map 1** shows that the variables included in the Multivariate Clustering Analysis are accurate indicators of determining where EV Charging Stations are typically located in Pittsburgh.

By examining the clusters created by the Multivariate Clustering analysis, we can see that neighborhoods that are considered to be more affluent fall into cluster 2 and neighborhoods that are considered to be economically disadvantaged tend to fall into cluster 3. Neighborhoods in cluster 2 have the highest median home values and the lowest proportion of residents without a bachelor’s degree. A further description of neighborhood demographic and economic indicator values relative to cluster is shown in **Figure 2**.

**Figure 2**

**Multivariate Clustering Box-Plots**

A graph with lines and dots

Description automatically generated

To further examine the relationship between neighborhood indicators, existing electric vehicle charging stations, and the results of cluster classification, locations of existing Electric Vehicle charging stations were aggregated by the clusters produced by Multivariate Clustering Analysis. This allowed the exact number of charging stations contained within each cluster of neighborhoods to be determined. As can be seen in **Figure 3**, cluster 2 neighborhoods contain the majority of Electric Vehicle charging stations which supports the hypothesis of a relationship between neighborhood level economic indicators and the presence or absence of charging stations. Cluster 2 contains 141 of the 234 charging stations within Pittsburgh city limits, or roughly 60% of all charging stations.

**Figure 3**

A graph with a bar chart

Description automatically generated with medium confidence

The number of Electric Vehicle charging stations located within each neighborhood in Pittsburgh is detailed in **Map 2**.

A map of a city

Description automatically generated

**5. Conclusion:**

Based on the clusters produced by K-means clustering and the locations of existing electric vehicle charging stations in Pittsburgh, it is apparent that a relationship exists between specified demographic and economic conditions within neighborhoods and the presence of electric vehicle charging stations. Neighborhoods with the lowest proportion of residents without a bachelor’s degree, lowest proportion of residents without Internet access, and lowest single mother rate are typically associated with a higher number of electric vehicle charging stations. Similarly, neighborhoods with the highest median home value and highest median gross rent are also associated with the presence of electric vehicle charging stations. Neighborhoods that were calculated to have fulfilled these criteria were labeled as cluster 2 through the utilization of a k-means clustering algorithm.

It is possible that neighborhoods could be designated as suitable for the construction of new electric vehicle charging stations based on their clustering classification and the number of existing charging stations located within the neighborhood. Though, to accomplish this properly, historical neighborhood data may be needed in order to analyze changing demographics and economic conditions within each neighborhood. Furthermore, additional analysis is required to forecast how individual neighborhoods may be impacted by changing demographics or economic conditions to identify neighborhoods that may have a higher need for EV Charging Stations in the not-too-distant future. Based on the mostly exploratory nature of this analysis, any decision on the suitability of a particular neighborhood for the construction of new electric vehicle charging stations would rely much more heavily on judgement rather than statistical evidence

**Limitations:**

This project is subject to limitations. While the United States Department of Energy Alternative Fuels Data Center appears to be thorough and accurate, it is likely that there are more charging stations in existence across Pittsburgh than included in the dataset, particularly on the West side of the city (United States Department of Energy, 2024). It may be appropriate to repeat the research contained in this project with a focus on the eastern side of the city of Pittsburgh due to this uncertainty in data reliability.

**References**

Cotter, N. (2023, September 14). The Pittsburgh Neighborhood Project. Retrieved November 17, 2024. <https://www.arcgis.com/home/item.html?id=4dc69b9e08b64c4f9d5c3f472780adf8>

Esri. (2024). *How multivariate clustering works*. How Multivariate Clustering works-ArcGIS Pro | Documentation. <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/how-multivariate-clustering-works.htm#:~:text=The%20R2%20value%20reflects,at%20discriminating%20among%20your%20features>.

Gorr, W. L., & Kurland, K. S. (2023). *GIS tutorial for ArcGIS pro 3.1*. Esri Press.

Pgh.admin. (2024, July 30) Neighborhoods. Retrieved November 10, 2024. <https://www.arcgis.com/home/item.html?id=dbd133a206cc4a3aa915cb28baa60fd4>

United States Department of Energy. (2024). Alternative Fuels Data Center: Electric Vehicle Charging Station Locations. Retrieved November 14, 2024. <https://afdc.energy.gov/fuels/electricity-locations#/analyze?country=US&region=US-PA>