Locations for Tesla Supercharger Installation

Document Title	Project Report
Name	Priyanka Gupta

Table of Content

1. Introduction:

In an era when technological innovation and environmental sustainability are coming together, the electric vehicle (EV) market is expanding at a rate never seen before. As the leader of this revolution, Tesla is about to take a big step forward in building up its infrastructure to meet the growing demand for electric vehicles.

The number of electric vehicles in Washington state is described by a large dataset that has undergone a thorough analysis for this project. In order to assist Tesla in determining the best locations for these charging stations, we want to extract relevant data. In order to encourage more people to convert to electric vehicles, we are working to make EV charging more simple and easily accessible.

2. Business Problem:

By late 2024, Tesla would open 3,500 new Superchargers along highway corridors to Tesla and non-Tesla customers across United States. This project uses sample of electric vehicle population dataset for Washington State to predict which cities Tesla should strategically deploy its Supercharger stations across Washington. Our findings will empower Tesla to enhance its distribution networks, fine-tune marketing strategies, and align inventory with the unique needs of each location, thus fostering the growth and adoption of electric vehicles across Washington.

Our data-driven Supercharger installation suggestions will enable Tesla to expand its charging network in Washington, thereby promoting greater adoption of electric vehicles as the company takes the forefront in advancing sustainable transportation.

3. Objective:

The primary objective of our consultancy project is to provide Tesla with exact, data-driven suggestions for the best locations for Supercharger stations throughout the state of Washington. Our objective is to evaluate and pinpoint crucial sites that would optimize the usability and effectiveness of Tesla's charging network for both present and prospective EV owners by utilising an extensive dataset on the population of electric vehicles in the area. In addition to improving convenience for both Tesla and non-Tesla EV users, this strategic deployment intends to strengthen Tesla's market position and further its larger purpose of supporting sustainable mobility.

4. Problem Statement:

As a leading consultancy firm, our objective is to provide recommendations for best locations for Tesla Supercharging Station installations in the area of Washington. It's difficult to determine which parts of Washington are most suited for Supercharger installations. Analysing a number of variables is necessary, such as traffic patterns, accessibility for both Tesla and non-Tesla EVs, proximity to key highways, and existing and forecast EV ownership.

STEPS TO BE FOLLOWED TO SOLVE THIS PROBLEM:

Pre-processing Data

This stage includes the loading, exploring, and cleaning of the data. To comprehend the many sorts of variables and variable handling strategies, the data is examined. The detection of missing values for imputation or omission is another step included in this process.

Dimensionality Reduction

Variables that are not necessary are removed during this phase. The dataset's dimensionality is decreased at this stage by using Principal Component Analysis (PCA). This aids in figuring out any interdependencies between variables in the dataset as well as the impact of different columns on the target variable.

Data Mining Task Specification

This stage involves assessment of the dataset and identification of target variables and outcome to determine the data mining tasks or methods to perform. According to the selected dataset, clustering and decision tree shall be applied to compare and finalize the better method for prediction of the target variable.

Data Partition

To build linear regression models, compare models, and evaluating the dataset, the dataset is divided into many categories such as training data and test data. The partitioned dataset's distribution of the data is random.

Assessment of results

The above selected methods are compared to identify the best model among them to use for predicting the target variable. The business value of the problem is considered to choose the best model.


Charged

Tesla to open U.S. charging network to rivals in $7.5 bln federal program

By **Hyunjoo Jin** and **Jarrett Renshaw**

February 15, 2023 6:22 PM CST · Updated 8 months ago

5. Source of the dataset:

Electric Vehicle Population Data: https://www.kaggle.com/datasets/usamabuttar/electric-vehicle-population-data-washington-us
Charging Station Location Data: https://afdc.energy.gov/data_download
Subset of Dataset : Updated Dataset using Tableau Tables

5.1. Description of Data

We are using 2 datasets Electric Vehicle population Dataset and Electric Charging station dataset for Washington State Government. The first dataset for Electric Vehicle population was extracted from Kaggle and includes 121654 rows and 13 columns (Vin(1-10), County, City, State, Postal Code, Model Year, Make, Model, Electric Vehicle Type, Electric range, DOL Vehicle ID, Vehicle Location and Electric Utility). The second dataset for Charging Station in Washington State is taken from US Department of Energy and it includes 1984 rows and 6 columns (Station Name-Charging Station, Street Address-Charging Station, City-Charging Station, Zip-Charging Station, Latitude-Charging Station and Longitude-Charging Station).

The Electric Vehicle and Electric Charging Station datasets were combined to create the subset of our dataset. This procedure includes utilising Tableau's table function to retrieve important information such as the number of Electric Charging Stations, the average and median Electric Range, and the number of vehicles for any given zip code in Washington. This subset is then used to build the required prediction models.

Vehicle Dataset Description:

Vin (1-10)	Motor vehicles can be uniquely identified by their Vehicle Identification Number (VIN). It is used to identify certain automobiles and is normally up of 10 characters
County	Refers to the county in Washington
City	Refer to the city in Washington
State	Washington
Postal Code	Different zip codes in Washington
Model Year	the year that the car model was produced.
Make	Refers to the vehicle's brand or make (e.g., Tesla, Audi, Nissan).
Model	Gives the vehicle's exact model name or number (Model 3, E-TRON, LEAF, etc.)
Electric Vehicle Type	Explains the type of electric car (plug-in hybrid or battery electric vehicle).
Electric Range	Usually expressed in miles, this is the maximum distance an electric car can go on a full charge before needing to be recharged.
DOL Vehicle ID	Most likely a special number that the Department of Licencing (DOL) gives to every car.
Vehicle Location	Identifies the location of the vehicle's latitude and longitude.
Electric Utility	Shows the name of the electric utility company that supplies power to the vehicle's location.

Electric Charging Station Dataset Description:

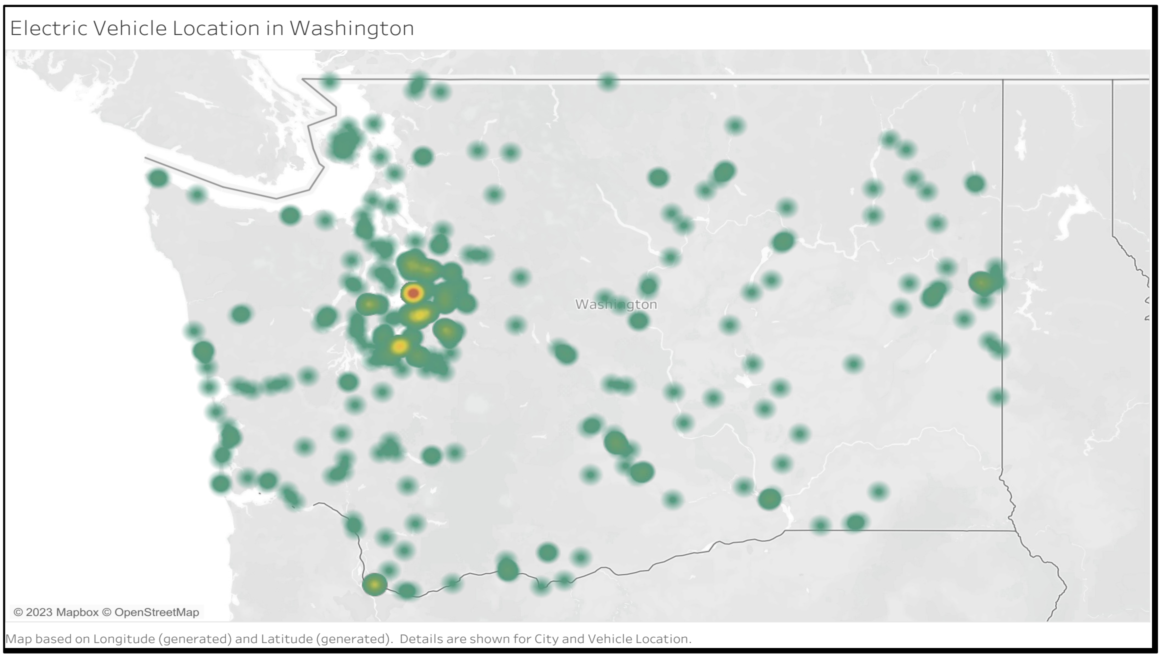
Station name - Charging Station	Station name of that Charging station
Station Address - Charging Station	Station Address of that Charging station
City – Charging Station	City in Washington
Zip – Charging Station	Zip Code of that location
Latitude – Charging Station	Latitude of Charging Station location
Longitude – Charging Station	Longitude of Charging Station location

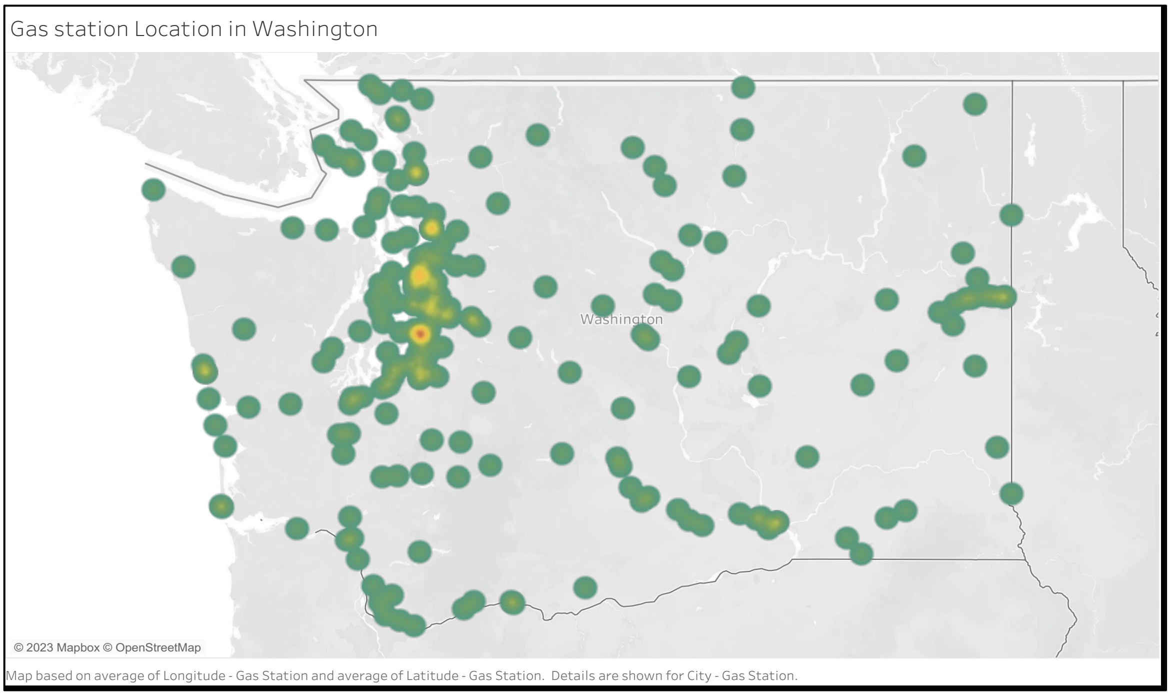
6. Data Visualization

We used 3 Visualisation techniques (Density plot, Heat maps and Tables) using Tableau to understand and analyse both Electric Vehicle and Electric Charging station dataset.

6.1. Density Plot

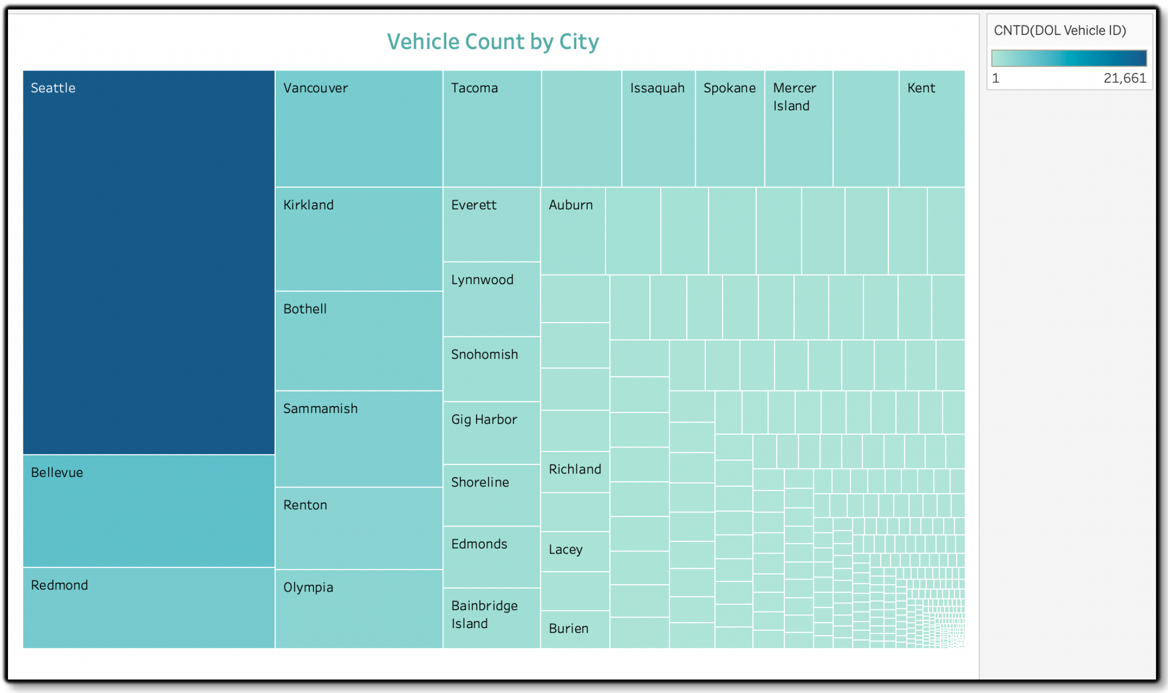
By analyzing density plots for vehicle locations using Electric Vehicle dataset and Electric Charging stations using Electric Charging station dataset, a clear trend emerges, as we observe that areas with a high density of old and existing Charging stations correspond to locations with a greater number of vehicles, and vice versa. To maximize its customer base, Tesla should strategically position superchargers in areas with the highest density. Through the data visualized in the two heatmaps, we can see that these areas of higher density are in the urban areas in the north-western part of the state in cities such as Seattle and its surrounding areas rather than the rural areas throughout the eastern part of the state. In this project, we will employ data mining algorithms, such as clustering and Linear Regression, to identify these high-density locations with a substantial number of electric vehicle users. This strategic approach will enable Tesla to maximize its profits from these strategically placed superchargers.

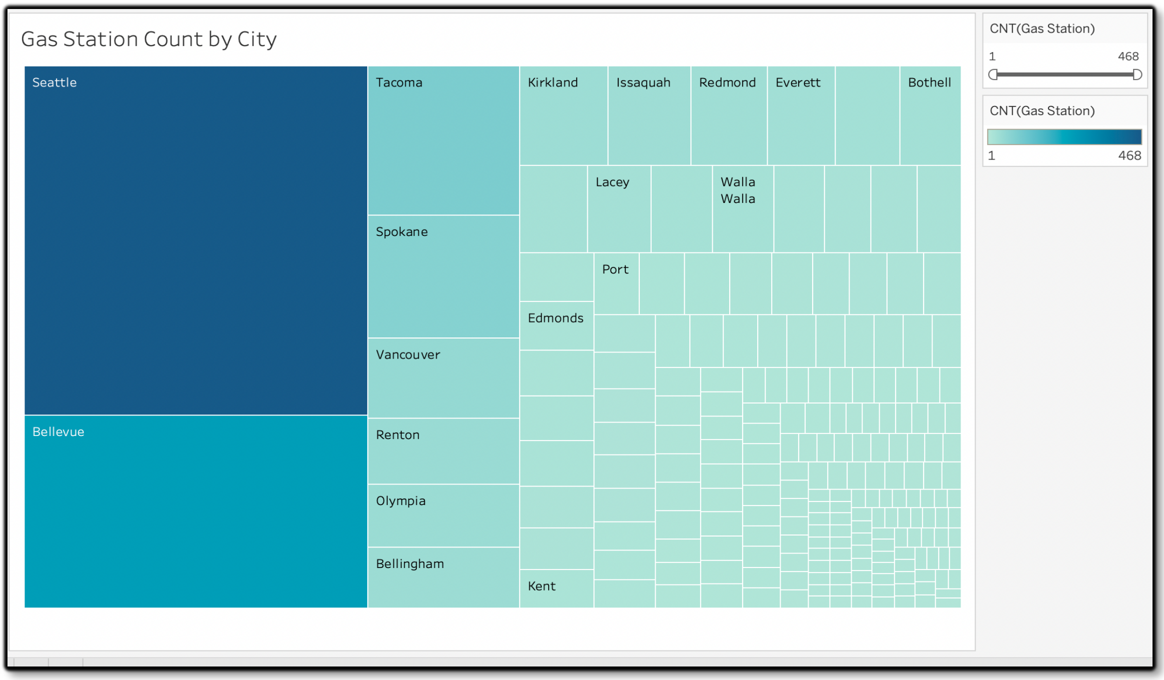




6.2. Heat Maps

We performed Data Visualisation using Heat-maps where we figured out the relation between count of Electric Vehicles and Charging Station in each city. The below heatmap give the clear relation between the two where most of places which has maximum Charging Stations has a greater number of electric cars. Tesla should target the highways in these cities to gain maximum profit. On the other hand, there is relation where cities which has only 1 or 2 Charging stations has only few Electric vehicles so Tesla should dig more into this market to see if there is any possibility to get increased car sales in this market by installing any Charging station in that city.





6.3. Tables

By examining both datasets, we created table attached in the subset of dataset in Data Description. If we analyze the table, we can say that, as the number of electric vehicles grows, so does the need for greater charging infrastructure in any zip code. Analyzing this relationship can provide insight into an area's suitability for electric vehicles.

Upon thorough examination of the dataset, we notice that certain zip codes, such as 98052 and 98072, exhibit both high vehicle counts and a substantial number of electric charging stations. These locations could serve as prime candidates for installing new superchargers. However, there are other areas like 98109 and 98125 with similarly high vehicle counts but limited charging stations. Investigating these areas could uncover opportunities to establish new charging stations, potentially expanding market coverage and reaching a broader customer base.

Through a comprehensive zip code analysis for each area, Tesla can determine the optimal cities to focus on for the launch of its new electric vehicle charging stations. To perform this analysis, we will be using Data mining and prediction techniques to suggest good locations for these Superchargers.

1	Postal Code	Number of Electric Charging Station	Number of Vehicles	Avg Electric Range	Median Electric Range
2	98052	45	3,192	81.97117795	19
3	98072	43	1,787	94.39899273	33
4	98012	36	2,073	75.53738543	12
5	98115	36	2,008	81.29731076	32.5
6	98004	33	2,117	94.48795465	18
7	98033	32	2,217	97.62426703	26
8	98034	31	1,713	85.49854057	25
9	98040	31	1,782	98.73681257	30
10	98021	27	1,288	80.01785714	17
11	98006	26	2,002	90.25924076	25
12	98074	25	1,761	89.84951732	19
13	98110	23	1,321	81.56320969	30
14	98117	23	1,275	78.37882353	32
15	98026	21	826	80.54600484	30
16	98027	20	1,140	91.43684211	25
17	98053	20	1,269	95.74310481	30
18	98059	20	1,311	84.05797101	19
19	98258	20	850	75.38	19
20	98391	20	876	79.03424658	21
21	98075	19	1,438	93.13630042	21
22	98103	19	1,620	83.45	32
23	98290	18	668	90.85179641	38
24	98607	18	1,093	82.2516011	26
25	98112	17	1,185	92.69957806	30
26	98177	17	845	82.97633136	32
27	98105	16	1,091	84.34738772	25
28	98106	16	496	75.43145161	29
29	98133	16	930	77.94086022	32
30	98272	16	498	71.85943775	19
31	98370	16	507	77.16568047	33
32	98501	16	1,293	84.63959783	26
33	98029	15	1,286	84.05365474	9
34	98028	14	738	88.89159892	32.5
35	98038	14	910	73.28901099	20.5
36	98221	14	505	82.45346535	30
37	98225	14	714	80.95658263	47
38	98296	14	766	81.22845953	23.5
39	98502	14	678	87.24631268	32
40	98005	13	927	90.03128371	21
41	98036	13	725	77.42896552	25
42	98087	13	608	69.37664474	0
43	98109	13	1,137	102.6473175	32
44	98125	13	1,024	80.61132813	38

Zip Code Dataset Project

Query	Explanation
How many observations in the dataset?	For Electric Vehicle - 121654 observations. For Charging Station – 1984 Observations.
How many binary/categorical variables?	For Electric Vehicle Dataset: ‘7’ County (Categorical), City(Categorical), State(Categorical), Make(Categorical), Model(Categorical), Electric Vehicle Type(binary), Electric Utility(Categorical) For Charging Station: ‘3’ City – Charging Station(Categorical), Station name – Charging Station(Categorical), Station Address – Charging Station(Categorical)

How many continuous variables?	For Electric Vehicle Dataset: '6'. Postal Code, Model Year, Electric Range, DOL Vehicle ID, Vehicle Location, Vin(1-10) Charging Station Dataset: '3' Zip, Latitude-Charging Station, Longitude- Charging Station
What is the outcome/target variable?	NA
If binary or categorical: What percentage of the variables belong to each class?	NA
If continuous: What is the mean value of the target variable?	NA
Before doing any further processing, what would your prediction of the target variable be?	Optimal locations for Tesla Supercharger stations in Washington.

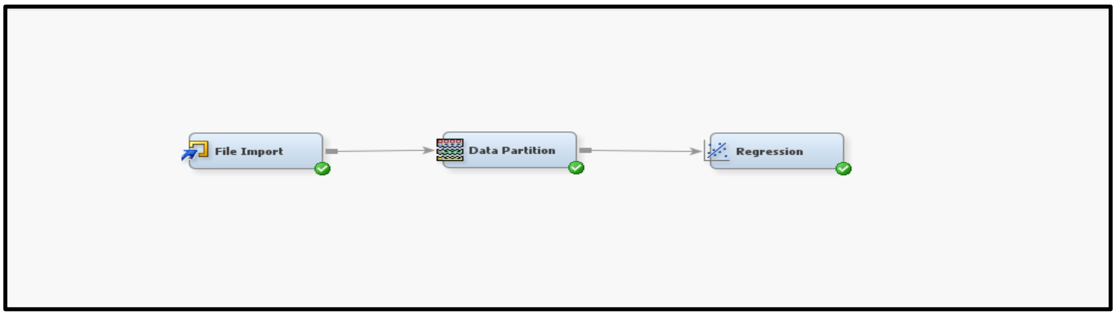
7. Data Prediction Techniques:

We examined two models: Linear Regression and Clustering. Opting for the Linear Regression model was to find the correlation between the number of Electric Vehicles and Electric Charging stations. On the other hand, we employed clustering to identify good locations for Superchargers. Through clustering, we aimed to cluster zip codes with high vehicle counts that would benefit from additional Electric Charging stations, thereby enhancing our business and help us to make strategic decisions to capture EV market.

7.1. Linear Regression

Linear regression is a statistical technique used to examine the correlation between number of Electric Vehicles and number of Electric charging stations. By utilizing this method, it becomes possible to predict how many electric charging stations need to be present for a specific number of vehicles.

To predict the number of electric charging stations availability based on the number of vehicles the dataset was partitioned into training and testing data sets.



Training data set: Comprising 80% of the total data, this subset was used to train the linear regression model. The model learned the relation between the number of vehicles and electric stations availability from this dataset.

Testing data set: Consisting of the remaining 20% of the data, this subset was utilized to assess the model's performance on unseen data. The testing dataset provided an independent evaluation to gauge the model's ability to generalize to new observations.

We used SAS in order to perform the linear regression.

Model Fit Statistics

R-Square	0.9020	Adj R-Sq	0.9025
AIC	638.3111	BIC	640.3305
SBC	646.3629	C(p)	2.0000

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.1443	0.1239	1.17	0.2446
Count_of_Postal_Code	1	0.0157	0.000254	61.85	<.0001

The linear regression equation used for this analysis is given by:

*Count of Electric charging stations (Y) = 0.1443 + 0.0157*Count of vehicles(x) + ε*

Where Y is the Electric station availability, which is dependent variable, X is the number of vehicles which is independent variable, $\beta 0$ is the intercept and $\beta 1$ is the coefficient for the number of vehicles.

After performing the linear regression using SAS, we got the following results:

- $\beta 0 = 0.1443$
- $\beta 1 = 0.0157$
- R-square = 0.9020
- ε is the error term

The positive coefficient $\beta 1 = 0.0157$ indicates a positive relationship between the number of vehicles and electric station availability. As the number of vehicles increases, electric station availability tends to increase according to our model.

The high R-square value (90.2%) suggests that it is a better fit and this model successfully explains a sizeable portion of the variability in electric station availability based on the number of vehicles. This means that when the number of vehicles changes, we can expect a corresponding change in electric station availability. This finding is crucial for making informed decisions and planning, as it provides valuable insights into how these two factors are linked.

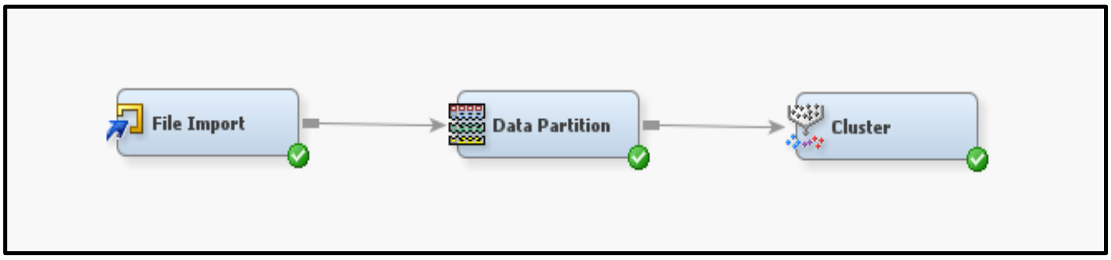
Stakeholders, such as planners or decision-makers, can trust that our model performs well not just on the data it was trained on but is likely to work effectively in similar situations. The model's ability to explain and predict these relationships is robust, giving confidence in its reliability for practical applications and decision support.

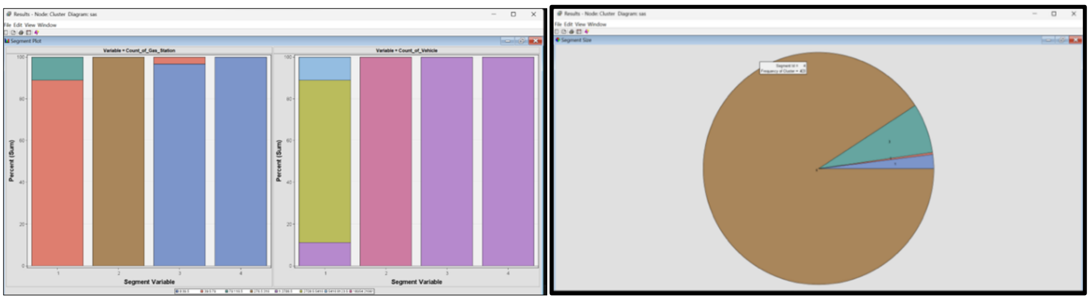
In summary, the high R-square value of 90.2% tells us that our linear regression model is good at explaining why electric station availability changes. This makes the model a valuable tool for figuring out and predicting how the number of vehicles on the road affects how easy it is to find an electric station. It is like having a reliable guide to understand and anticipate the impact of changes in the number of vehicles on electric station accessibility.

7.2. Clustering

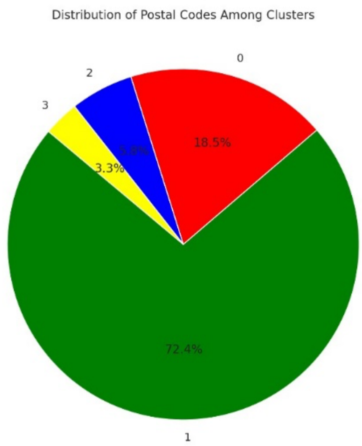
We performed clustering analysis on our dataset, to predict the postal codes to install the superchargers. The below results we gathered shows what we discovered using SAS Enterprise Miner in our project. Firstly, it tells us which details carry the most weight and statistical significance in creating our various groups (or clusters) in our data. The larger the importance, the more these details affect the clusters. In these clusters we considered the count of gas stations and the count of electrical vehicles as our dependent variables, from which four major clusters were formed. Through the analysis using SAS Enterprise Miner and using several combinations used from dataset, we found that the count of electric charging stations and electrical vehicles were the most significant variables to group the specific zip codes that can be used to help us find specific groups in our data.

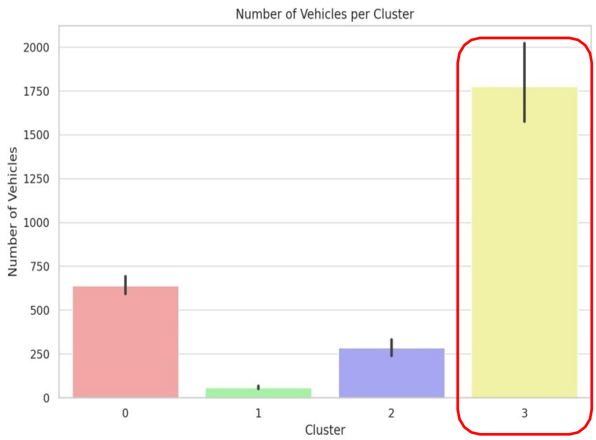
For our report, these cluster results would help in finding out the postal codes where we can suggest the installation of superchargers. This will also help us to identify any outliers in the data that could point to cities that might require more in-depth investigation before beginning installation of new charging station.

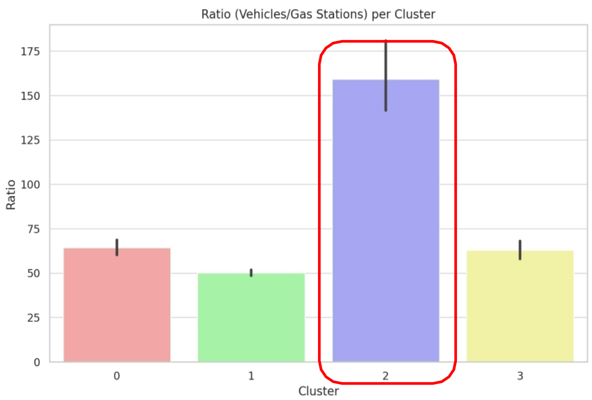


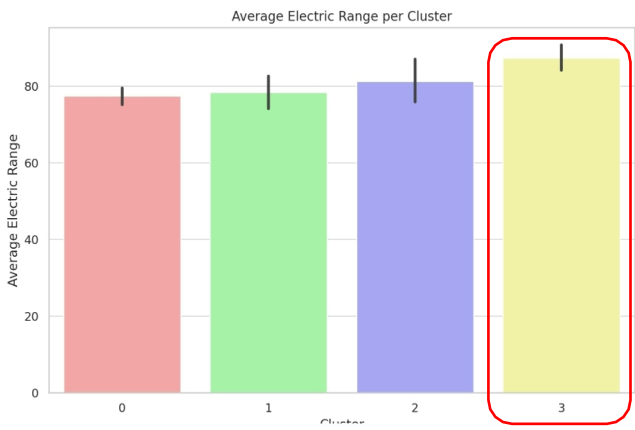


Analysing the outputs, it is evident that Cluster 3 exhibits a higher average electric range compared to other clusters and also the number of vehicles in are more in cluster 3. Our observations reveal that cluster 2 exhibits a higher ratio of vehicles to gas stations when compared to cluster 3.









To enhance the precision of our findings, we also used k-means clustering using Python, aligning with the outcomes obtained from SAS. The clusters were created based on variables such as the count of gas stations and electric vehicles, creating four unique clusters with the following characteristics:

1. Cluster 0: This Cluster indicates a well-balanced Market with Moderate Demand. that the ratio of vehicles to gas stations is 64.25, implying an average of approximately 64 vehicles for every one gas station.

- Average Number of Gas Stations: 10.51
- Average Number of Vehicles: 638.92
- Average Electric Range: 77.49
- Average Ratio (Vehicles/Gas station): 64.25

2. Cluster 1: This cluster signifies an Emerging Market with Lower Demand, as evidenced by a ratio of vehicles to gas station at 50, suggesting an average of 50 vehicles per gas station.

- Average Number of Gas Stations: 1.14
- Average Number of Vehicles: 57.59
- Average Electric Range: 78.39
- Average Ratio: 50.18

3. Cluster 2: This cluster suggests a high potential for EV Charging stations as we observe highest no of vehicles in this cluster with a lesser ratio of gas station. The results suggest that there are 159 vehicles per gas station. In comparison with the other clusters the ratio of vehicles per gas station is more in this cluster with an average of 160 vehicles per gas station.

- Average Number of Gas Stations: 1.93
- Average Number of Vehicles: 284.90
- Average Electric Range: 81.27
- Average Ratio: 159.24

4. Cluster 3: This cluster represents a High-Demand Urban Areas. This cluster implies that there is an average of 62 vehicles per gas station which is like cluster 0.

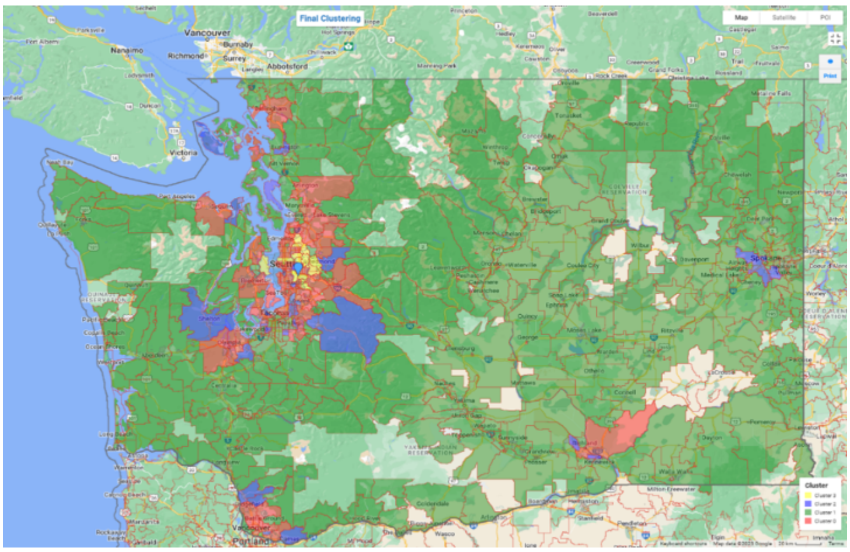
- Average Number of Gas Stations: 28.76
- Average Number of Vehicles: 1774.94
- Average Electric Range: 87.41
- Average Ratio: 62.94

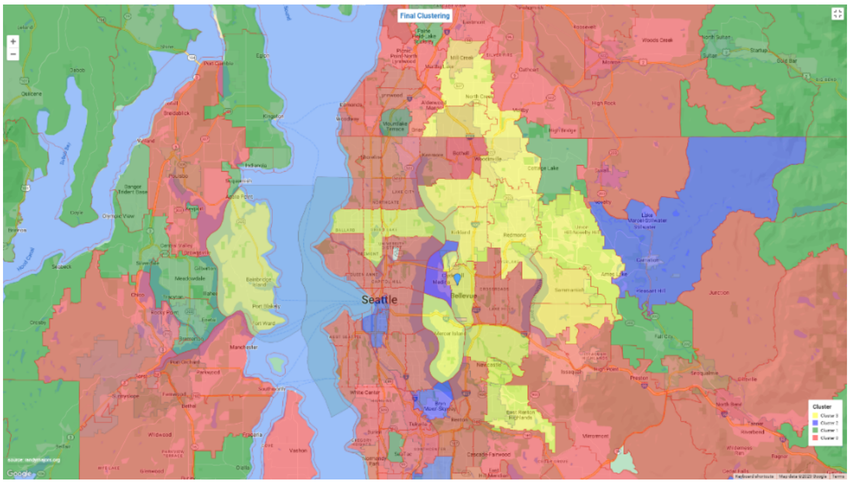
8. Conclusion:

With the various data visualization techniques we employed, we found there is a strong relation between the number of vehicles and number of Charging Station for a city. Using linear regression, we were able to identify and gain statistical evidence that there is a strong correlation between the number of vehicles and the ratio of number of vehicles to charging stations in predicting the amount of charging stations. We also performed K-means clustering and were able to generate 4 clusters, each with postal codes with their unique and distinct set of characteristics. We used the data gathered in each cluster to gather insights into the condition of the electric vehicle market in each area. From the Average Number of Vehicles in each postal code we were able to observe the electric vehicle density in each area and get insight into the amount of electric vehicle traffic that each area experiences, which is a key market metric when deciding where to place a new charging station. Additionally, we used the ratio variable which we decided to create after performing our data visualizations to as well to give us a sense of the amount of demand of charging stations in each postal code. Lastly, we used the variable "Average Electric Range" per postal code as an indication to the maturity of the electric vehicle adoption in a particular area. We drew this insight by analyzing our data.

Through our clustering analysis, we were able to see that Cluster 3 was the area with the highest average number of electric vehicles with 1774.94 vehicles per postal code. This indicates that in the areas within Cluster 3 there is a significantly higher electric vehicle density and potential vehicular traffic compared to the remaining three clusters. Another important characteristic of this cluster was the average electric vehicle range of 87.41 miles per vehicle, which was also significantly higher compared to the remaining three clusters. This suggests that this cluster is a more mature electric vehicle market where there is an increased adoption and receptance to electric vehicle technology. Although, Cluster 3's ratio did not stand out among the three clusters, through its other key features, specifically its significantly higher average number of vehicles per postal code, we decided to present this cluster as an area of recommendation when considering where to place a new supercharging station in the state of Washington.

In addition, our team also wanted to present Cluster 2 and the postal codes contained therein as an additional area of recommendation to Tesla. This Cluster had an average number of vehicles of 284.90 (which is far lower than 1774.94 recorded in Cluster 3) indicating that there is decreased electric vehicle density and probable electric vehicular traffic in this area. However, one key characteristic of this Cluster was its average ratio of electric vehicles per charging station was 159.24, which was significantly higher than the remaining three clusters. This suggests that in the postal codes located within this cluster there is a large unmet demand for electric vehicle charging stations.





The images above show two geographical maps of Washington state and the various Clusters that we generated with their specific postal codes displayed. In these maps, we can see that Cluster 1 (shown by the green areas on the map) takes up the majority of the geographical area throughout the state. We can also see that Cluster 3 (shown by the yellow areas on the map) primarily contains the postal codes located Northeast to Seattle, Washington's largest city. Additionally, we see that Cluster 2 (shown by the blue areas on the map) also contains some postal codes located in the cities on the outskirts of the Seattle area, and also notably the city of Spokane (Washington's second largest city)

8.1. Managerial Implications:

Through our data analysis, we were able to present primary areas within the state of Washington for Tesla to consider launching a new supercharging station. However, we recommend that managers proceed to perform extensive research into each specific postal code by looking into factors such as the area's electric grid capacity, nearby highways, when considering where to plant a new charging station.

We were able to successfully recommend the areas in Cluster 2 and Cluster 3 as the areas Tesla should consider deploying their supercharging station. Out of the entire state of Washington, these areas account for less than 9.1% of the postal codes in the state of Washington.

Tesla managers can decide between these two Clusters that we presented based on their specific business goals and objectives in the placement of a new supercharging station.

If Tesla managers desire to target a potentially underserved area for electric vehicle charging stations, then they should consider researching Cluster 2 for future development and growth. However, if the goal is to target a more established electric vehicle market, then the managers should consider Cluster 3 as their primary area of interest, due to its high electric vehicle traffic and existing demand.