**Fall 2022 DSI Capstone Project**

**Progress Report 1**

**EV Charging stations in Washington State**

Oct 22, 2022

**Student Team Members**

**Anqi Lin** al4215

**Clarissa Tai** rt2822

**Mengchen Xu** mx2257

**Yue Zhang** yz4155

**Yu-Chieh Chen** yc4015

**Industry Mentors**

**Alexander Voet**, Data Scientist, Associate at KPMG Lighthouse

**Arvind Sathi**, Director at KPMG Lighthouse

**Chengwei Wang**, D&A Data Scientist, Associate at KPMG Lighthouse

**Jeffrey Lee**, Lighthouse D&A Modeler, Associate at KPMG

**John Heath**, Director at KPMG

**Jordan Jalving**, Sr Associate at KPMG

**Laura Uguccioni**, D&A Data Scientist, Manager at KPMG Lighthouse

**Michelle Zee**, D&A Consultant, Sr. Associate at KPMG Lighthouse

**Sydney (Bolim) Son**, D&A Data Modeler, Associate at KPMG Lighthouse

## 

## 

**Table of Contents**

[**1. Project Background and Problem Definition**](#_aahzs218ut95) **3**

[**2. Exploratory Data Analysis**](#_fl0rvsjcskca) **3**

[**2.1 Why Washington State?**](#_xz5y4byr7xon) **3**

[**2.2 Natural Risk Index**](#_v0jay4quhfms) **5**

[**2.3 Traffic**](#_lsdb6uzamngp) **6**

[**2.4 Crime rate**](#_6gpvfmngdr7u) **7**

[**2.5 Tourists Attraction**](#_27c3af5jxou7) **8**

[**2.6 E-grid**](#_ggwqdaip0eqx) **9**

[**2.7 Gas station distribution**](#_4xhu2b7s3py) **9**

[**3. Data Preprocessing**](#_fhpjqd9dxkol) **10**

[**3.1 Route Distance**](#_49d68gth20ln) **10**

[**3.2 Highway**](#_r8hiqnvzes1q) **10**

[**3.3 City categorization**](#_d74lgzyz2qe2) **11**

[**3.4 Missing value**](#_5mwthvnjlfa4) **11**

[**3.5 Generalization specific data processing**](#_9mellh7w7l4) **12**

[**3.6 Dataset schema**](#_xyq8wx797u4s) **12**

[**4. Baseline Model: Classification**](#_vvwc1b6gcphb) **13**

[**5. Contributions**](#_lys71hg9tp38) **14**

[**6. Appendix**](#_mku8nzyjed5e) **15**

## 

## **1. Project Background and Problem Definition**

## In November 2021, President Biden officially signed the bipartisan infrastructure bill, which included a $5 billion investment in state-administered grants for nationwide electric vehicle (EV) charging stations. These funds will help states develop charging networks across communities that include rural, disadvantaged, and hard-to-reach areas, propelling the administration’s decarbonization goals through state-level transportation projects.[[1]](#footnote-0) Considering rural, disadvantaged, and hard-to-reach areas, people living in rural areas are deterred from owning an EV car because of the lack of EV chargers, especially when they plan to travel far. The Chicken-and-Egg Conundrum exists here is that drivers choose not to own an EV being afraid of not enough power, while lack of enough EV owners will not support the idea of establishing an EV charging station.[[2]](#footnote-1)

## In this project, we will tackle the challenge of how to best electrify our roads and communities to encourage balanced growth in our communities. Therefore, our defined research question is: what are the most important factors to consider when choosing EV charger locations in Washington state and where are potential optimal locations.

## **2. Exploratory Data Analysis**

### **2.1 Why Washington State?**

To deliver a meaningful result in the end, we need to choose a specific state or region to focus our efforts on. After doing some basic research, we chose Washington state as our target for a couple of reasons as follows:

1. To deliver a meaningful result in the end, we need to choose a specific Looking into the EV registrations and public EV charging stations statistics, we found that Washington has comparatively more EV registered and not severely lacking EV stations for modeling purposes. WA is ranked middle in all three statistics we calculated. WA has a good portion of EV cars among all vehicles from the top two graphs in Figure 2.1.1. In addition, WA still has the potential to build more EV stations compared to other large EV-used states in the current state from the last graph in Figure 2.1.1.
2. WA has an EV Deployment Goal that all light-duty vehicles sold, purchased, or registered in WA must all be EVs by model year 2030[[3]](#footnote-2) even though current EVs only account for 0.8% of all registered vehicles in WA. This indicates a large potential market for EVs along with its supporting infrastructure.
3. In Figure 2.1.2 and 2.1.3, we observe that EV registration density is not necessarily the same as EV charging station density. There are counties with relatively high EV registration density but low EV charging station density, which are the counties we should focus more on. 
4. WA has supported policies and grants for EVs and EV infrastructures. The federal government establishes a National Electric Vehicle Infrastructure Formula Program (“NEVI Formula”) to provide funding to States to deploy EV charging infrastructure and to establish an interconnected network to facilitate data collection, access, and reliability. WA is allocated about $71 million from this program distributed in 5 years as shown in Figure 2.1.4. In addition to the federal program, WA also has about $198 million state funding devoted to this area in the 2021-2023 biennium, among which $140 million are specifically for developing EV Charging Infrastructures.

### 

### **2.2 Natural Risk Index**

Risk of natural disaster is an important aspect when designing EV station locations. Exposure to natural disasters will directly lead to possible electric shocks and outages which further lead to severe damage to the infrastructure. Moreover, sustainability and environmental vulnerability of the charging stations take a big part when achieving business goals. The National Risk Index is an application from FEMA.gov that identifies communities most at risk to 18 natural hazards. According to the data, Washington's most common natural disasters include volcanic activity, avalanches, earthquakes, etc. Natural risk index is calculated by three components, which are annual economic loss, social vulnerability, and community resilience. Overall, coastal suburban areas have higher risk scores, due to lower social vulnerability and community resilience to recover from the disaster. In this project, we will only use the overall risk score which includes all 18 hazards. The missing values are from lacking historic records.



### **2.3 Traffic**

In order to make sure the EV stations are placed around traffic accessible areas, traffic count becomes an important factor to consider when determining the location. A total of 188 highway routes exist in WA in 2021 and major traffic is happening along interstate highways and state routes in the Seattle area as shown in Figure 2.3.1. When overlaying the EV station locations with the traffic, we realized that WA is lacking EV chargers in rural areas within comparatively large traffic access in 3 major areas (Figure 2.3.2). According to the deep dive analysis on the three areas, most areas can be covered by at least one EV Level 2 charger if the radius of EV charging stations is at least 50 miles. In addition, some traffic near Colville National Forest will only be covered by one EV charging station with only two Level 2 chargers, which shows an apparent lack of EV charging stations supply (squared area in yellow in Figure 2.3.3). In conclusion, traffic count data provides us baseline areas where we may need to build EV stations.

### **2.4 Crime rate**

Crime rate in one specific area is also an important aspect to be taken into account when deciding EV station locations. Out of safety concerns, EV car owners may be unwilling to leave their car for a couple of hours to charge in dangerous areas (especially high motor theft, robbery, etc.), leaving the already built EV stations useless. Therefore, to take the crime situation into account, we collected crime count data from the FBI website. To eliminate the population effect on the arbitrary crime count, we take the crime rate (total crimes divided population), which indicates the dangerousness of the neighborhood in each city. From the visualization, most current EV charging stations are not located in high crime rate areas, except some crowded charging stations in Spokane (Figure 2.4). In addition, assumed ideal places for EV chargers discussed in previous slides are mostly located in low crime areas.

### **2.5 Tourists Attraction**

Tourists account for a huge amount of EV station users. In order to determine the optimal locations for EV stations, we need to consider the places where tourists prefer to visit, and these places may provide helpful insights on potential areas that may need EV stations. From the visualization (Figure 2.5), we can see that the blue circled area includes some tourism locations without current EV charging stations along major routes (mainly I-5) with higher traffic identified previously. Therefore, we can focus more on these areas, which may be potential baseline areas that need to build EV stations.



### **2.6 E-grid**

To build new EV charging stations, we need to assure that electricity supply is available in the surroundings. According to the Clean Energy Transformation Act (SB 51116, 2019), Washington state is going to be an electricity supply free of greenhouse gas emission by 2045. Therefore, electricity from renewable energy plants are preferred and more analysis on these plants is needed. Currently, about 29% of plants are from renewable sources, offering about 12% of electricity capacity in Washington state. From figure 2.6, it shows that more renewable energy plants are proposed compared to nonrenewable ones. The distribution of plants are scattered around the whole state except the upper area. Most proposed energy plants are along the road, which implies future electricity accessibility among major routes. However, we are unable to find detailed distribution and usage information for each plant so we cannot come to the conclusion on which location will be accessible from E-grids.



### **2.7 Gas station distribution**

From the above section, it shows the power plants’ locations but there is no public information on where electricity is delivered from each power plant. We are unable to determine the availability of electricity. Since gas chargers and convenience stores in gas stations are powered by electricity, the locations of gas stations guarantee the electricity accessibility. Therefore, an analysis of gas stations is crucial. From the figure 2.7, we can see most current EV charging stations overlap with gas stations in cities but not in rural areas. Some rural areas still lack charging stations. When proposing new EV charging stations, we can assume that the presence of gas stations indicates the electricity can be delivered.



## 

## **3. Data Preprocessing**

### **3.1 Route Distance**

To prepare the features needed for modeling, we consider setting gas station locations in WA as our base references to determine the optimal locations for EV stations and EV chargers. We are curious about the number of EV stations, highway exits, crime counts, traffic counts, tourist attraction places, and the risk of natural disaster around gas stations, which might be useful features to help us determine the optimal EV station locations. However, it is inexplicable to find these numbers in a circle area with gas station locations as centers and direct distance as radius.

People who need to find EV stations are drivers who can only access gas stations through existing routes. Therefore, route distance is more reasonable to be considered in this situation.

The method we come up with to figure out the route distances between locations is called MapQuest API. MapQuest API can provide route distance between two locations, and it also can perform radius search by setting center locations and route distances as radius.

### **3.2 Highway**

Since the main federal funding source, the NEVI Formula Program, requires that any EV charging infrastructure projects acquired or installed with NEVI Formula funding must be located along a designated alternative fuel corridor[[4]](#footnote-3), we consider the distance to the highway might be an important factor to our model. Instead of using the direct distance from location to highway, we chose to use the route distance from a certain location to its nearest highway exit, because this gives us the best estimate of how much a driver needs to drive from highway in order to access the EV charging station. We used MapQuest's Corridor Search API to filter the gas stations within a 5 mile driving distance from the target highway. Then we used MapQuest’s Direction API to obtain the route distance from each filtered gas station to its nearest highway exit.

### **3.3 City Categorization**

In this project, our focus is on rural areas. Hence, it is important to define the boundaries of urban, suburban, and rural areas to prevent models from skewing toward urban areas where most EV charging stations and population are located. Recall the fact that the census tract is planned according to population density, by an average of 4000 inhabitants in each area. It is reasonable to classify the census tracts by its area. The graph below shows the distribution of urban (purple), suburban (blue) and rural (light blue) census tracts. In the total of 1780 valid census tracts, we labeled 1332 as urban, 226 as suburban, and 222 as rural.



### **3.4 Missing Value**

Features related to crime and NRI contain missing values due to lack of record. To fill the missing values in NRI, we take the average since it is impossible to have no risk of natural disaster and NRI are preprocessed also using mean values. There is no crime recorded in some cities on the FBI website, so for gas stations in those cities, we have missing crime data. Considering that the missing crime data may be related to the lack of large population in some cities so small amounts of crimes weren’t recorded and apparently it does not indicate no crimes happening in that city, we decided to fill in missing values in crime data with minimum count.

### **3.5 Dataset Schema**

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| gas\_key | Numerical | Gas station key |
| gas\_name | Categorical | Gas station name |
| gas\_lat | Numerical | Gas station location latitude |
| gas\_long | Numerical | Gas station location longitude |
| attr\_cnt\_1mile | Numerical | Number of tourism attractions within 1 mile distance |
| attr\_cnt\_5mile | Numerical | Number of tourism attractions within 5 mile distance |
| distance\_to\_nearest\_attr | Numerical | Distance to the nearest tourists attraction |
| crime\_coord | Numerical | Crime data location coordinate: (longitude, latitude) |
| crime\_county | Categorical | Crime happened county label |
| total\_crime | Numerical | Number of summed crimes (theft, burglary, murder, etc.) |
| highway | Categorical | Name of Highway, for example “I5” |
| distance\_to\_nearest\_exit | Numerical | Distance to the nearest highway exit |
| num\_EV\_in\_2\_miles\_of\_gas | Numerical | Number of EV stations within 2 mile distance |
| num\_EV\_in\_20\_miles\_of\_gas | Numerical | Number of EV charging stations within 20 mile distance |
| distance\_to\_closest\_ev\_station | Numerical | Distance to the nearest EV charging station |
| nri\_geoid | Categorical | Geographic identifiers that uniquely identify all administrative geographic area |
| nri\_county | Categorical | County names that corresponds to the Geographic identifiers in the previous column, for example “King” as “King County” |
| nri\_risk\_score | Numerical | Relative risk of natural hazards at a location, higher value means higher risk, a number in range [0, 100] |
| nri\_risk\_rating | Categorical | Relative rating of communities at the same level, categories include “Very High”, “Relatively High”, “Relatively Moderate”,“Very Low” and “Relatively Low” |
| traff\_cnt\_5m\_max | Numerical | Maximum traffic count within 5 miles |
| traff\_cnt\_10m\_max | Numerical | Maximum traffic count within 10 miles |

## **4. Baseline Model: Classification**

To make good use of the existing gas station distribution, we want to use a supervised ML model to give a rough baseline for EV charging station locations based on gas station locations. Based on our current gas station and EV station data, in our selected focus area (I-5 south), we can classify which gas stations can be chosen as locations to place EV chargers. According to our modeling goal, we can create a binary classification model.

The target variable is defined as follows: within 20m precision, for each gas station location, if there is a nearby EV station (indicates that the location is possible to place EV station) then 1 else 0. Since we want to deploy the dataset on the focused I-5 south area, we separate those gas stations as our deployment data (without labeling the target). The training and validation data are the gas stations except ones along the I-5 south area. When exploring the training features, we realized Imbalanced target with 63% unmatched gas stations (0) and 37% matched gas stations (1). In this case, we used SMOTE to balance targets. In addition, we removed highly correlated and unuseful features (including id, name, lat/long, etc) to reach our final dataset for modeling.

The evaluation metrics chosen are recall, average precision and AUROC. Our cost of false negatives (predict actual matched location as unmatched) is high so we choose recall. We also want two metrics (average precision and AUROC) to evaluate the general performance.

Among all classification models, we chose logistic regression for three reasons: a) we have a small dataset so we want to build a more simple model; b) we have a binary target; and c) logistic regression is easy to interpret.

After performing 5-fold cross validation with logistic regression on our data, the model performance is as follows: Recall: 0.79, Average Precision: 0.79, AUC: 0.81. The top 4 important features are burglary rate, motor vehicle theft rate, natural risk rating and maximum traffic within 5 miles as shown in Figure 4.1. This result corresponds to our expectation that crime rate, natural risk index and traffic will affect the decision on making EV charging stations.

In Figure 4.2, the confusion matrix shows that the model predicts 0 (gas stations without nearby EV stations) more correctly but the model is not accurate on predicting 1 (gas station with EV stations). However, we don’t care whether the model predicted unmatched gas stations as EV stations because our goal is to have more EV stations. Thus, the model performed as expected. 

After deployment, as shown in Figure 4.3, many of our model predicted EV charging locations are outside of existing EV charging locations, which provides us several potential locations to do optimization models on.



## **5. Appendix - Contributions**

* Anqi Lin: Team captain. Gathered, cleaned and preprocessed traffic and crime data, and responsible for exploratory data analysis on those two dataset. Main contributor on classification modeling.
* Clarissa Tai: Gathered, cleaned, preprocessed, and visualized natural risk index data. Responsible for exploratory data analysis on census tract level data, and aggregate feature and dataset for generalization.
* Mengchen Xu: Gathered, cleaned and preprocessed commute and attraction data. Responsible for exploratory data analysis on commute and attraction data. Main contributor on MapQuest API for route distance.
* Yue Zhang: Gathered, cleaned and preprocessed government and exit data. Responsible for exploratory data analysis on government funding data. Main contributor on MapQuest search API and optimization approach.
* Yu-Chieh Chen: Gathered, cleaned and preprocessed gas station and e-grid data. Responsible for exploratory data analysis on gas station and e-grid data, as well as maintaining merged main dataset.

## 

## **6. Appendix - Data Source**

* Washington State Plan for Electric Vehicle Infrastructure Deployment, July 2022 <https://wsdot.wa.gov/construction-planning/statewide-plans/washington-state-plan-electric-vehicle-infrastructure-deployment>
* U.S. Energy Information Administration. “Electricity.” <https://www.eia.gov/electricity/>. Accessed 5 October 2022.
* Energy Efficiency & Renewable Energy. “Alternative Fuel Data Center.” <https://afdc.energy.gov/fuels/electricity_locations.html#/analyze?fuel=ELEC>. Accessed 5 October 2022.
* National Risk Index <https://hazards.fema.gov/nri/data-resources>. Accessed 5 October 2022.
* Annual Average Traffic Count <https://gisdata-wsdot.opendata.arcgis.com/datasets/WSDOT::wsdot-traffic-counts-aadt-1/explore?location=47.271387%2C-119.745108%2C6.90> Accessed 11 October 2022.
* Crime <https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019/tables/table-8/table-8-state-cuts/washington.xls> Accessed 5 October 2022.
* MAPQUEST <https://developer.mapquest.com/documentation> Accessed 15 October 2022.
* Washington Geospatial Open Data Portal <https://geo.wa.gov/datasets/WSDOT::wsdot-interstate-exit-numbers-1/api> Accessed 12 October 2022.
* Tourist Attraction <https://mygeodata.cloud/data/download/osm/tourist-attractions/united-states-of-america--washington> Accessed 5 October 2022.

1. [EU Support Grows for Russia Oil Ban Over Ukraine War - WSJ](https://www.wsj.com/articles/eu-support-grows-for-russia-oil-ban-for-ukraine-war-11647883376) [↑](#footnote-ref-0)
2. [Biden Has a $5 Billion Plan to Eliminate America's EV Charging Deserts](https://www.bloomberg.com/news/features/2022-07-29/biden-has-a-5-billion-plan-to-eliminate-america-s-ev-charging-deserts) [↑](#footnote-ref-1)
3. [Electric Vehicle (EV) Deployment Goal](https://afdc.energy.gov/laws/12873#:~:text=All%20light%2Dduty%20vehicles%20sold,goal%20by%20December%2031%2C%202022) [↑](#footnote-ref-2)
4. [National Electric Vehicle Infrastructure Formula Program](https://www.fhwa.dot.gov/bipartisan-infrastructure-law/nevi_formula_program.cfm) [↑](#footnote-ref-3)