EV Chargers: Modelling for number of EV Chargers and Recommendations in Ontario

1.0 Introduction

1.1 Problem Statement

EV Station is a start-up company providing multi-type EV Chargers for all types of EVs on the market. The company has successfully installed over 10K EV Chargers over 10 states in the United States, and are looking to enter the Canadian EV charging market. EV Station data science team was consulted by the management team for estimating the Canadian market size. The management would like to investigate the estimated number of additional EV chargers needed currently in Canada, revenue for the company if they fill the gap, and how it will be needed in the next few years.

Photo 1. Public EV chargers in Ontario (source: https://www.autotrader.ca/editorial/20211201/level-3-ev-chargers-to-be-installed-along-ontario-s-hwy-400-and-401/)

1.2 Goal

The aim of this project is to study if there are any opportunities for the EV Station company to enter the Canadian EV Charging Market, especially for Ontario, through installing new Public EV Chargers, as the emerging trend of using Electrical Vehicles. The addition of EV chargers in Ontario will be estimated for the management team to study the feasibility of entering the Canadian Market.

2.0 Datasets

2.1 Charging Stations Dataset

The main dataset used in this study was retrieved and downloaded from a public data source: [Electric Charging and Alternative Fuelling Stations Locator]() from the Government of Canada. The website provides a visualization tool for users to locate public chargers in both the US and Canada. It also enables embedding and downloading for the original datasets for public uses.

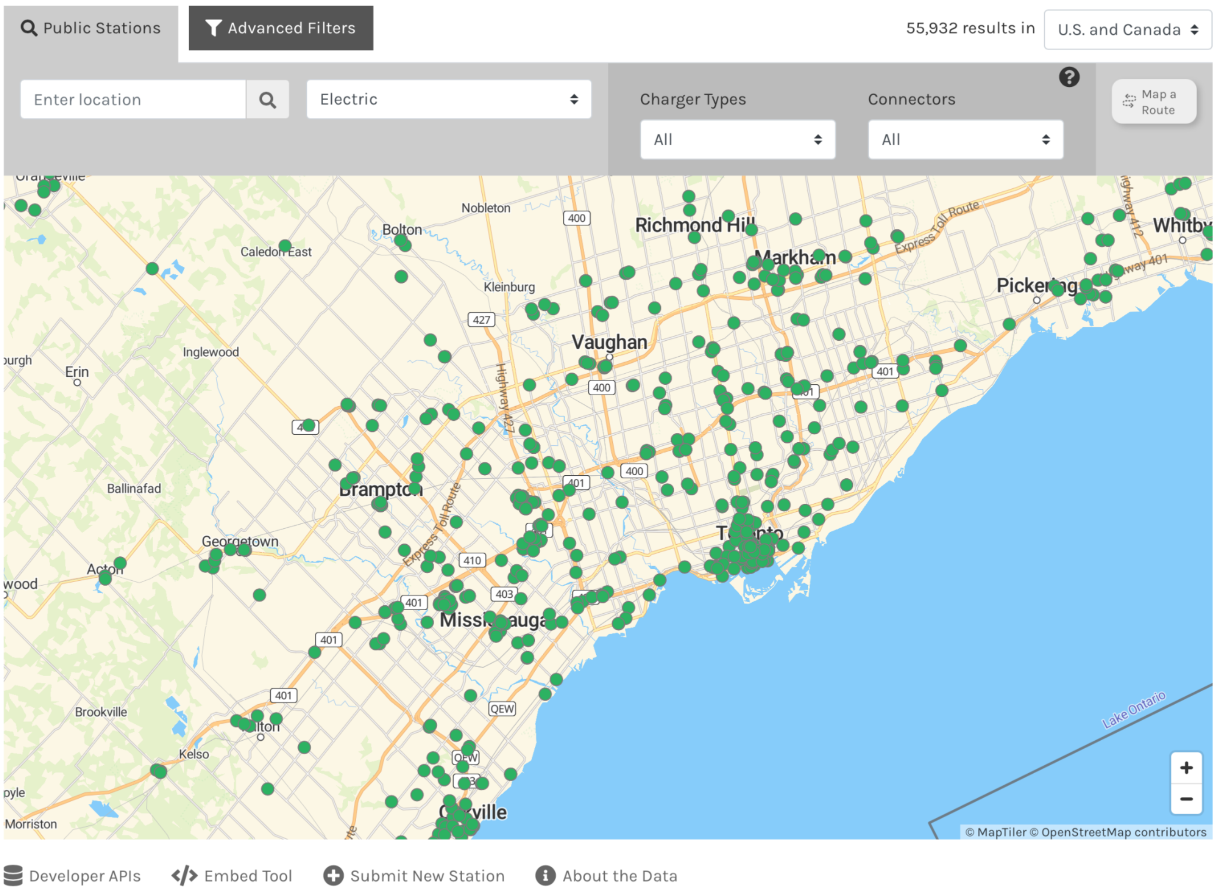


Photo 1. Electric Charging and Alternative Fuelling Stations Locator website from the Government of Canada, showing electrical EV Chargers in the Greater Toronto Area (source: https://www.nrcan.gc.ca/energy-efficiency/transportation-alternative-fuels/electric-charging-alternative-fuelling-stationslocator-map/20487#/find/nearest?fuel=ELEC&ev_levels=all)

This database provided locations for 53,292 electrical charging stations in the United States and Canada. The basic information of the charging stations also included, type of chargers, coordinates, number of chargers, charger speed, types of ports, etc.

2.2 Other Stations Dataset

Other public datasets used in this study included:

1) GDP: https://www.bea.gov/data/gdp/gdp-stat ; https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3610040202

2) Population: https://www.census.gov/popclock/; https://www150.statcan.gc.ca/n1/pub/71-607-x/71-607-x2018005-eng.htm
3) EV and Car Registration: https://afdc.energy.gov/data/; https://www.statcan.gc.ca/en/topics-start/automotive

These public datasets were aggregated and summarized in a single dataset.

3.0 Data Cleaning & Data Wrangling

The datasets used in this study were relatively very cleaned, since the data is used as tools for public uses. Some of the data cleaning and wrangling approaches were used to ensure the modeling processes will be smooth.

3.1 Data Cleaning

First of all, NAN columns or columns containing large amounts of NAN values were removed from the dataset, as they did not provide any useful information. In the meantime, some of the columns out of the scope of this study were dropped from the main dataset, including ZIP code, Federal Agency ID, EV Pricing, EV On-site Renewable Sources etc. Although these columns may provide some related information, I did not expect they can contribute to the success of the study, and they may increase the complexity of a model.
For some of the useful features in the dataset, a lot of values that are NAN was filled with 0. The data can either be missing or not collected, and it is reasonable to assume that they are zero.

3.2 Data Wrangling

The raw data of the charging stations is probably not recognizable by the models, and it is necessary to conduct some data wrangling.
For instance, To understand the column, related to charging connector types, it is necessary to know the differences. As we may notice, each charging station may comprise multiple Some of the charging cables can be convertible to the other, while some of them may not. Existing dataset provided the charging connect types in a single column, which will definitely be text columns that won't be able to work in the modeling. As a result, the column was splitted by these charging port types and it was binary coded by existing (1) or not existing (0) at each charging station.

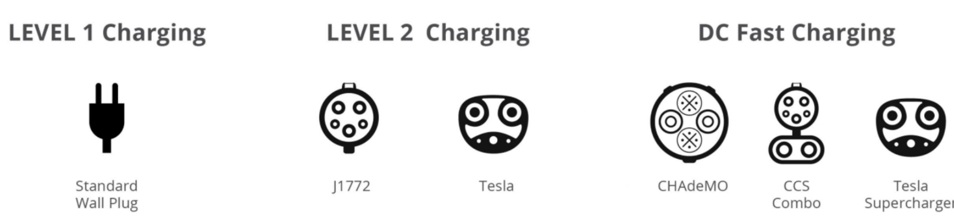
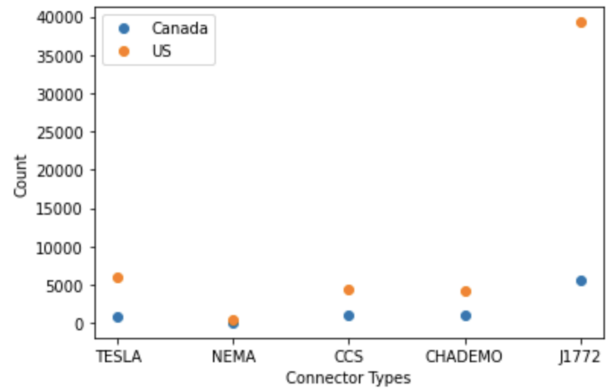


Photo 3. Charging port types and Level of charging speed (source: https://blinkcharging.com/understanding-ev-charging-plugs/?locale=en)

After the data was transformed, the following figure was generated to explore each type of EV charger connector type. The J1772 is the most used connector in both Canada and the US. The US definitely has more chargers than Canada. J1772 in the US is about 8 times more than Tesla and other types of charge, which is very interesting.

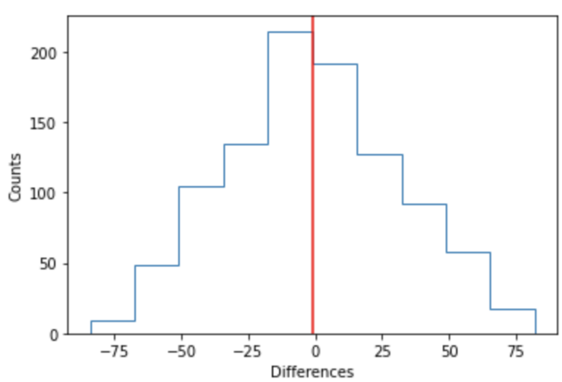


4.0 Exploratory Data Analysis

By doing the exploratory data analysis, a few interesting conclusions can be made through data conversion, feature engineering, or merging between datasets.

4.1 CCS & TESLA Chargers are identical between provinces

The mean and standard deviation for Tesla and CCS are very close to each other. They are both around 100 per states/provinces and the standard deviation is 170. But the average number of Tesla is greater than CCS. We assume that the EV Charger company can install all types of chargers, and they want to see two charging types TESLA and CCS, as they can provide Level 3 Fast Charging and more EV Manufactures are increasingly using them.



As a result, it is meaningful to find some relationship between these two types of ports in all states/provinces, which may provide some hints of whether the EV Charger Company should evenly install these two types of chargers or more Tesla.

A permutation test was completed by creating a null hypothesis that there is no difference between the average number of CCS and TESLA. Differences were calculated and resampled. The differences were plotted, and the p-value was calculated to 0.239 < 0.5, which rejected the hypothesis. Our conclusion for doing this test was there was no obvious evidence for the differences between the average number of CCS and TESLA chargers per state/province. It is recommended that these chargers may be evenly installed in the future for level 3 fast charging.

4.2 Linear Relationships between features

A fast correlation tests are completed for between features, a few linear relationships stand out (Correlation Coefficient > 0.8), including:

1) Total Ports and Total Stations
   The total ports and total stations in each state/province have a very strong linear relationship. It is reasonable, as each station may consist of 2-10 stations, depending on the size and locations.

2) 2019 Total Registered Cars and Population
   The total registered cars also had a relatively strong (Coefficient = 0.91) relationship with population.

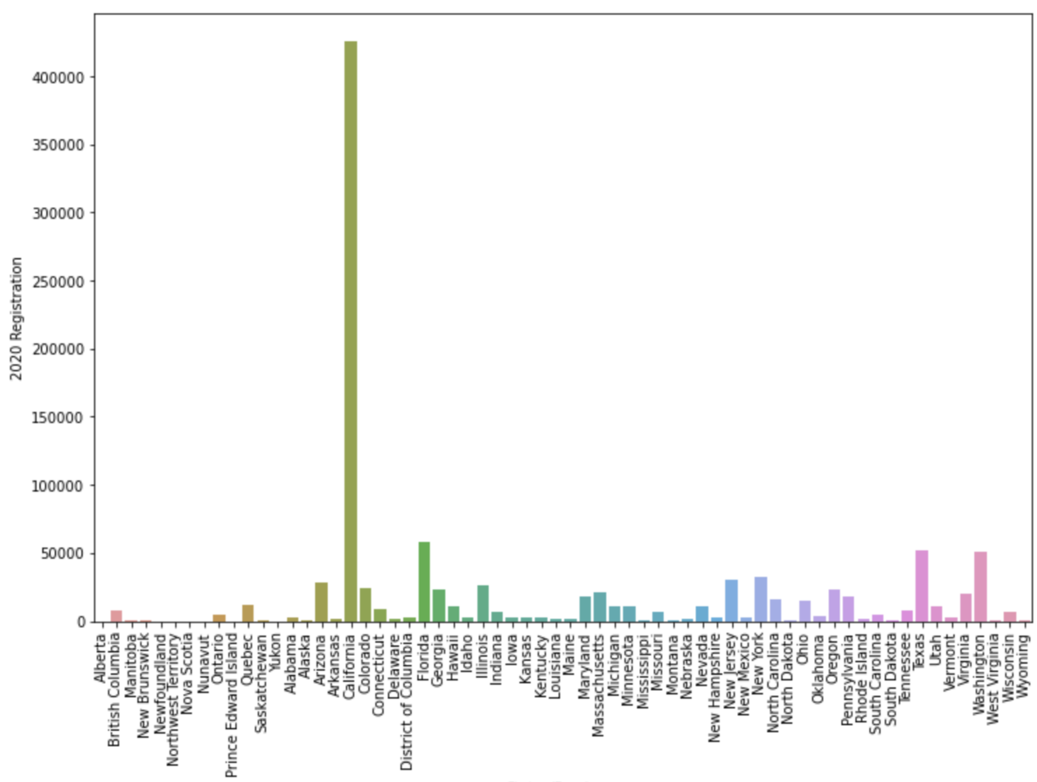
3) CCS and CHADEMO
   The number of CCS and CHADEMO are strongly related. It suggested that these two ports are probably often installed together in every station.

4.3 Dimension Reduction

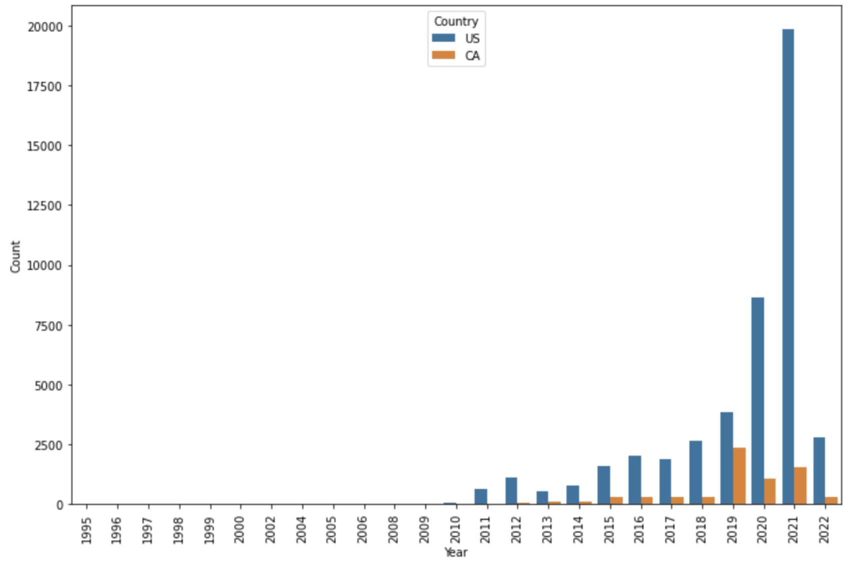
A PCA Test was used to explore the relationship between features and targets. It was determined that four features can explain 90% of the variance. However, after the dimension is reduced, we did not see an obvious pattern or groups.

4.4 Other interesting findings

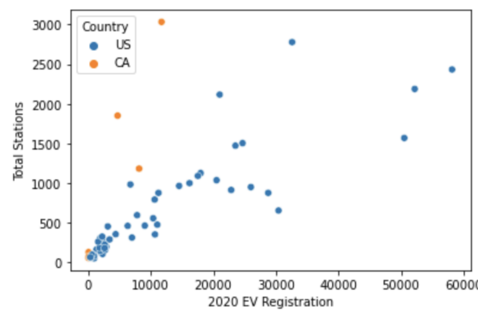
1) EV registration in California is 8 or more times than the Florida, Taxes, and Washington, which are 2nd highest EV registration states.



2) EV Charger installation number in 2021 was almost twice of chargers installed in 2020 in the US. Similar trends for 2020 and 2019. It suggested that chargers are emerging in the recent a few years in the US.
   In Canada, the trend of installation seems to be different. In 2019, above 2500 chargers were installed but twice less in 2020. More chargers were installed in 2021 but less than 2019.



3) EV registration number in 2020 seem to have a linear relationship with the total number of charging stations, as shown. Canada seems to have deeper slope than the US, which means in the US, the same number of 2020 EV registration will have less total stations



5.0 Modeling and Findings

5.1 Feature Preparation

After the exploratory data analysis, training and testing datasets were prepared to train machine learning models. The data set included the following features:

1) **Total ports**: total ports that were installed in the states/provinces regardless of the port types.
2) **TESLA, NEMA, CCS, CHADEMO, J1772**: total number of respective charging connectors in the states/provinces.
3) **2020 EV Registration**: EV registered in 2020.
4) **M_Pop**: Million of population in the most recent census.
5) **B_GDP**: Billion of GDP in 2021
6) **M_Reg_Cars**: Million of cars registered, regardless of the types of cars (EV, PHEV, Fuel).
7) **M_Area, M_Land, M_Water**: Million square meters for total area, land area, and water area.

8) **Country_CA (Binary)**: 1 is in Canada, 0 is in the US.

The **Total Stations** column is the target for modeling and prediction.

5.2 Modeling

In the modeling process, a few approaches were tried to find the best fit for the model, including training-test split, cross validation, grid search and random search hyperparameter tuning.
The following models were fitted and validated:

1) Linear Regression Model
2) Random Forest
3) XGBoost
4) RANSAC Linear

Each model's performance was evaluated through mean absolute errors (MAE) and R2 scores. The RANSAC Linear model performs the best as it has the most R2 score and the lowest MAE. After a grid search cross validation process, the best-fit hyperparameters were chosen as the best model.
The model was then fitted to the entire dataset, excluding Ontario (Target Province). Then the model was used to predict the EV Chargers in Ontario for the current scenario. The prediction model has provided a fairly accurate prediction. The actual charger number is 1850, and the predicted EV charger number is 1894. The MAE indicated that the number of chargers are plus or minus 6, which is very small compared to the actual number.
The modeling results of this scenario might provided an idea to the company that entering Canada to install more public chargers might not be profitable, as the existing chargers have already sufficient to supply most of the EVs.

5.3 Scenarios

EV Registration has increased a lot during these two years. There are many reasons for the increase of EV purchasing. One of the most important reason is the increase of the gas prices, and the demonstrated capacity of EV batteries. According to Stats Canada, there are about 14,748 battery EVs registered in 2021, which was more than 2 times than 2020. The increasing rate is also predicted to be more than 100% in the future years.
As a result, the following scenarios are modelled, and the following results were predicted

1. The EV Registration number increase to 15K, the predicted number of chargers is 1901
2. The EV Registration number increase to 30K, the predicted number of chargers is 1912
3. The EV Registration number increase to 60K, the predicted number of chargers is 1933

It is very strange to see that, even though EVs are increasing more than 10 times, the public charging stations are still not very needed. Why does this happen? What has caused this problem? Can we find some other opportunities in considering these questions? What other analysis and modeling can we do?

5.4 Discussion

The following things can be considered as the causes of the results:
1. Most of the people charge their EV at home, as a result, there is no such needs to use public chargers. In Canada, a lot of people live in houses with individual garages. These garages

enable the owners to install chargers or even plug in electricity directly. The public chargers are most likely to be used by people who live in an apartment with public garages. To explore this, we can investigate and model urban EV chargers and number of EVs only. Another thing that we can do is to find the data for home chargers, although it is very difficult.

2. People are not using their EVs to travel. People may choose a gas vehicle for long-range travels, as there might be a shortage of electrical chargers in rural areas. Even if there are some destination chargers for travellers, a lot of them are actually private properties, i.e. hotels, resorts, airbnbs. These chargers may not be included in this dataset.

6.0 Future Works

1. We will explore not only provinces, but also cities in the dataset. There are two benefits for doing this: 1) Increase the number of data; 2) focus on places where public stations are used more. People living in a house, usually do not use a public stations.

2. If possible, looking for data about home chargers. Home chargers may be the major charging approaches for EV owners. Home charger is still critical to EV, due to the charging speed, type of chargers, and available chargers. As an owner of a Tesla, I found that we rarely need a public charger, as I can charge the EV every night, rather than wait for 20-30 min to charge at a charging station. If there is data for home chargers, we can make some modeling about the following questions: How many home chargers are needed? Are there any opportunities for the company to help installing home chargers, instead of commercial/public chargers? What the revenue would be? Any possibilities of making shared private chargers?

3. We can also explore EV charger types, based on the EV types. Up to this point, we only focused on the total station numbers. But currently, there are a few charging port types, and a lot of them are not transferable. What the most needed charging ports can be a great question to be answered through modeling. Also, is it more profitable to install uniform charging ports, or focus on one charging port?