Capstone Project

Coursera - Google Data Analytics Professional Certificate
December 2021

Fernando Castellanos

CASE STUDY: Analysis to identify the two preferable cities in Washington State to use as pilot sites for the installation of ZAP EV PowerHood.

Scenario

As a junior data analyst working in a data consultancy company, I form part of our internal working group assembled to assist one of our clients, company ZAP. ZAP is a company working in the area of power electronics and artificial intelligence and it is ready to try out one of its new inventions: the EV PowerHood. The general idea of the EV PowerHood is to interconnect several EVs connected to their chargers and that are located in the same proximity area. EV PowerHood optimizes their combine operation to provide electrical power to the homes of the EV owners and EVs that need to charge. Combining several EVs allows for a large pool of energy that will reduce the energy cost of the EV owners.

The ZAP directors of marketing, finances and research are leading a project to analyse the market conditions in Washington State. The company has chosen this State because of its average income level and the general attitude of its population to be early adapters of electronic devices in general. As part of a pilot program ZAP company is interested in deploying its EV PowerHood systems to two cities. The cities should be on the top 5 cities in terms of EV numbers and should be close to a balanced and homogeneous distribution of EVs across the city. Additionally, they want to know which are the top electric vehicles makers in the chosen cities to analyse plugs and electronics compatibility with their system. Finally, they want to understand the development in time of the number EVs in the chosen cities and the evolution of the number of EVs in terms of makers.

1. ASK

I will be exploring and analyzing the EV population in Washington State and its characteristics such as location, make and time evolution. Using this I expect to identify the two cities in the State that are among the top five in terms of numbers of EVs and are close to a homogeneous distribution of EVs, as per ZAP requirements. I will use EV registration State public data as the source of information and some of the metrics that I will use are: selection of the 5 top cities in terms of number of EVs and for each of these cities calculation of the top 5 EV makers, measurements of uptake, distribution and clustering of EVs across the city. Additionally, an analysis of EV registration along several previous years to observe the rate of increase of EVs in the selected cities and the rate of increase of EVs per maker.

The two cities selected and a summary of the analysis will be presented to ZAP directors of marketing, finance and research as the proposed option for the launching of the pilot programme of the EV PowerHood. Additional consultations could be required with the State Department of Transportation and the local electrical utility.

Deliverable
Explore and analyze EV population data in Washington State to identify the two preferable cities in the State that are among the top five in terms of numbers of EVs and the closest to a uniform distribution of EVs, as per ZAP requirements. These cities will be selected to participate in the pilot programme of ZAP EV PowerHood.

2. PREPARE

Deliverable
The original data for the study is located at the Washington State repository of public open data under the category Transportation. (https://data.wa.gov/browse?sortBy=relevance&category=Transportation&undefined=).

Two sets of data were downloaded:
- *Electric Vehicle Population Data* (https://data.wa.gov/Transportation/Electric-Vehicle-Population-Data/f6w7-q2d2).
- *Electric Vehicle Title and Registration Activity* (https://data.wa.gov/Transportation/Electric-Vehicle-Title-and-Registration-Activity/rpr4-cgyd).

The Department of Licensing provided the two sets and they were last updated 9th Nov 2021. Once the two datasets were downloaded, copies of the two sets were made and stored on a backup disk.

The dataset *'Electric Vehicle Population Data'* is organized in 15 columns and 84476 rows. Each row corresponds to a Battery Electric Vehicle (BEV) or a Plug-in Hybrid Electric Vehicle (PHEVs) that is currently registered through Washington State Department of Licensing (DOL).
The dataset *'Electric Vehicle Title and Registration Activity'* is organized in 30 columns and 479357 rows. Each row corresponds to a title activity (transactions recording changes of ownership), or registration activity (transactions authorizing vehicles to be used on Washington public roads).
Lists of the columns and their descriptions for the two sets are shown in Appendix A.

The source of the datasets is directly the Department of Licensing of Washington State, therefore the datasets are:
- unbiased, credible and reliable as they are issued by a government agency, which follows State guidelines for the disclosure of public data.
- comprehensive as they provide a wide range of information related to the registration of EVs in the State and supply the information required for my analysis.
- the datasets are current, as their last update was 9th November 2021.
- the websites for the datasets indicated that they have been downloaded many times.

All these attributes make the datasets ROCCCCs.

The two datasets are located in Washington State repository of public open data, as such there is no need for a license. Privacy is of critical importance to the State and measures such as providing only a portion of the EV identification number and the EVs location resolution only unto the ZIP level, provide conditions to assure the privacy of the EVs owners.

For safety and security the two original downloaded datasets were stored in my computer under the “Capstone Project” folder, subfolder “Data”. Three copies of both datasets were made and distributed to a backup folder in the computer, an external backup disk and a cloud personal data repository.

The size of the two datasets is large, but they are easily managed by data tools such as Excel, SQL or R. I choose Excel as the tool to facilitate the sorting and filtering of the data. A summary of the changes applied to the datasets is shown below.

Dataset *‘Electric Vehicle Population Data’* - Initial size 15 columns by 84476 rows.
Deleted the following columns, as they are of not relevance to the project:
VIN(1-10),Model Year, Model, Electric Vehicle Type, Clean Alternative Fuel Vehicle (CAF) Eligibility, Electric Range, Base MSRP, Legislative District, DOL Vehicle ID and Vehicle Location.
After, filtered for residents of Washington State and removed rows belonging to EV owners that have residence outside Washington State. Filtered for rows with “blank” for the County column and removed them.The resultant dataset was checked for possible duplication of records and unusual values on any of the columns. No problems were detected and the final dataset had 5 columns and 84254 rows.

Dataset *‘Electric Vehicle Title and Registration Activity’* - Initial size 30 columns and 479357 rows.
Deleted the following columns, as they are of not relevance to the project:
Clean Alternative Fuel Vehicle Type, VIN (1-10), Model Year, Model, New or Used Vehicle, Sale price, Electric Vehicle Fee Paid, Electric Range, Base MSRP, 2015 HB 2778 Exemption Eligibility, Sale Date, Vehicle Primary Use, DOL Vehicle ID, Legislative District, 2019 HB 2042 Clean Alternative Fuel Vehicle (CAFV) Eligibility, Meets 2019 HB 2042 Electric Range Requirement, Meets 2019 HB 2042 Sale Date Requirement, Meets 2019 HB 2042 Sale Price/Value Requirement, Odometer Reading, Odometer Code, Transportation Electrification Fee Paid and Hybrid Vehicle Electrification Fee Paid. After, filtered for residents of Washington State and removed rows belonging to EV owners that have residence outside Washington State. Filtered for rows with “blank” for any of the columns and removed them. Filtered the column *“Transaction Type”* and kept only the rows with *“Original Registration”* as this identifies the date the EV was initially register. The resultant dataset was checked for duplicates and the final dataset had 5 columns and 208883 rows.

3. PROCESS

Deliverable

To study and analyse the data I decided to use Excel and used some of its features as filters, duplication detection and pivot tables. The integrity of the data was evaluated by running filters and counters along the columns to determine if there were any unusual or missing values. Some integrity analysis was carried out by checking that values in the *“Transaction Year”* column agreed with the values corresponding to the year in *“DOL Transaction Date”*. All these different inspections were recorded in a log file, together with the filtering and sorting from the previous section. After performing all this manipulation of the datasets, both were considered ready for analysis.

4. ANALYZE

I started my analysis with the set *'Electric Vehicle Population Data'.* Using a series of pivot tables I was able to calculate the following results:

Top 5 cities in terms of EVs

CITY	NUMBER EVs
Seattle	16008
Bellevue	4371
Redmond	3241
Vancouver	3011
Kirkland	2674
Total Washington State	**84254**

Top 5 EV makers in Seattle

MAKER	NUMBER EVs
Tesla	6728
Nissan	2426
Chevrolet	1551
BMW	924
Ford	658

Top 5 EV makers in Bellevue

MAKER	NUMBER EVs
Tesla	2625
Nissan	502
BMW	219
Chevrolet	213
Kia	113

Top 5 EV makers in Redmond

MAKER	NUMBER EVs
Tesla	1867
Nissan	454
Chevrolet	205
BMW	148
Toyota	88

Top 5 EV makers in Vancouver

MAKER	NUMBER EVs
Tesla	1032
Nissan	473
Chevrolet	369
Toyota	216
Ford	199

Top 5 EV makers in Kirkland

MAKER	NUMBER EVs
Tesla	1481
Nissan	303
Chevrolet	185
BMW	126
Kia	89

These initial results provide some insights: the city with most EVs is Seattle and it has almost four times the number than the second city. However, the number of EVs in Seattle corresponds only to almost 20% of the total number of EVs in Washington State. With the exception of Vancouver, the other cities are in very close proximity of Seattle and belong

to the King county. Even more, Bellevue is considered as part of the “Seattle Metropolitan Area”. These emphasise the concentration of EVs to Seattle and its vicinity.

From the tables showing the top five EVs makers for these five cities is possible to observe Tesla is number one in all of them. It surpasses the second maker, that is Nissan in the five cities by two or more times its amount of EVs. This represents a very strong position in the market. The third maker is Chevrolet across the five cities (in Bellevue BMW is third ahead of Chevrolet only by six EVs, so they can be interchangeable). The fourth maker is BMW with only the exception of Vancouver, where BMW does not appears in the top five makers. Finally, Ford, Kia and Toyota are the final makers. Therefore, it is possible to conclude the three more popular makers are Tesla, Nissan and Chevrolet in that order.

In order to evaluate the distribution of EVs in a city, the following methodology was used. Three key indexes were calculated for each of the five cities: level of uptake EVs, level of distribution EVs and level of clustering EVs. The first index is the level of uptake an it is defined as:

$$Level\ of\ uptake = \frac{Number\ of\ EVs}{Estimated\ number\ of\ cars}$$

where the estimated number of cars was calculated from the estimated number of households in the city (obtained from the United States Census Bureau) and the estimated average number of cars by household in the city (obtained from the Data USA). This index calculates the portion of the total number of cars that is EVs. The next index is the level of distribution and it is defined as:

$$Level\ of\ distribution = \frac{Number\ of\ ZIP\ code\ areas\ with\ EVs}{Number\ of\ ZIP\ areas}$$

where the number of ZIP code areas with EVs is calculated from the dataset ‘*Electric Vehicle Population Data’* and the total number of ZIP areas for the city is obtained from the US Postal Service. This index provides a rough estimated of the portion of the city area that have EVs. The last index is the level of clustering and it is defined as:

$$Level\ of\ clustering = Std.\ deviation\ of\ EVs\ across\ their\ ZIP\ code\ areas$$

where using the data from the dataset ‘*Electric Vehicle Population Data’* the standard deviation of the number of EVs on each ZIP code is calculated.

Using the combination of these three indexes it is then possible to provide an approximation of the level of homogeneity of the distribution of EVs on the city. This all round index is calculated as level of uptake times level of distribution divided by the level of clustering. The following table presents the results.

City/Index	Uptake	Distribution	Clustering	All Round Index
Seattle	0.032	0.79	357.73	0.007
Bellevue	0.047	1.00	490.59	0.0096
Redmond	0.075	0.50	1069.16	0.003
Vancouver	0.027	1.00	95.99	0.028
Kirkland	0.046	1.00	798.48	0.005

From the table is possible to observe the number of EVs, compare to the total number of cars in each city is very small, with the largest in Redmond with 7.5%. In terms of how the EVs are distributed around each city, three cities have EVs in all their local ZIP codes, while for example Redmond the EVs cover half of its ZIP codes. The level of clustering index clearly indicates that Redmond has a very poor distribution of its EVs around its ZIP codes. In terms of clustering Vancouver is the best as its standard deviation is small compared to the other cities.

Finally, the all round index provide us with an estimated metric of the EV distribution homogeneity on the five cities. The results show the two best cities to carry the pilot project for ZAP EV PowerHood: Vancouver and Bellevue, and if required Seattle could be a third candidate. From the table is possible to see that for example, Vancouver has a low uptake of EVs, but they are distributed across all the local ZIP codes in a close to homogeneous distribution.

The final part is the calculation of the historic data using the reduced and manipulated dataset *'Electric Vehicle Title and Registration Activity'.* The data clearly indicates the increase of EVs in the five cities. However, Seattle shows a steep increase in the rate of growth after 2015, that does not appear in the other cities. This increase on the rate is reflected in the EVs numbers at the State level. In terms of EV makers, Tesla is clearly ahead of the rest. Data shows Tesla was growing to a similar rate to the other EV makers. However, after 2017 Tesla started to grow at a high rate, which has continued until know. In comparison the other EV makers have maintained more or less a constant number of their EVs registered. For the cities of Vancouver and Bellevue the rate of growth for EVs is similar.

A summary of the findings is:

- Seattle and surrounding cities account for a large amount of the EVs in Washington State. The top five cities are: Seattle, Bellevue, Redmond, Vancouver and Kirkland. The number of EVs in Seattle has above four times the other four cities.
- In terms of EV makers the four top in the five cities are: Tesla, Nissan, Chevrolet and BMW. Ford, Kia and Toyota share the fifth place. Tesla has very strong position in the market at its numbers are two or more times the other cities.
- The percentage of EVs, compared to the total estimated number of cars is below 8% for all the cities. Redmond is the city with largest uptake of EVs of 7.5%.
- From the top five cities, three have EVs on each of their respective local ZIP codes: Bellevue, Vancouver and Kirkland.
- However, the distribution of EVs across the local ZIPs is very uneven, with the worst case Redmond.
- The calculation of the all round shows that the two cities to carry the pilot test are: Vancouver and Bellevue. If required a third city will be Seattle.

- EVs rate of growth in Seattle showed a sharp increase in 2015 and continues until today. The other four sities present a more close to constant lower rate since 2010.
- From 2017 Tesla increased its market share rapidly and now has by a large margin surpassed the other four makers. This group of four EV makers has maintained about a same number of EVs year after year, with some losing market share.

5. SHARE

To share the analysis and results with colleagues and our clients, ZAP, I decided a way to demonstrate the versatility and applications of visualisations was to use two different tools to create graphics. Both set of graphics support the analysis and insights discussed previously.

I decided that Tableau will be a good platform to use and collate several graphs on a dashboard. During the design I identify the graphs to include in the dashboard:

- A map of Washington State with the location of the five cities with a color “level” indicating the number of EVs in each city. Hovering of the mouse over the city will provide the EV related data.
- Some type of “matrix map” showing the relationships between the top five cities and the top five EV makers.
- Several bar plots for number of EVs, level of uptake EVs, level of distribution EVs and level of clustering EVs for each city.

The second set of graphs was for the historical evolution of the EVs in terms of numbers and EV makers. These graphs I decided to use Excel and try several things such as lines, bars and areas and see which combination could represent better the findings.

After some work, trial and error, I finished with the following:

- A Tableau dashboard EVs in Washington State - Top 5 performers.

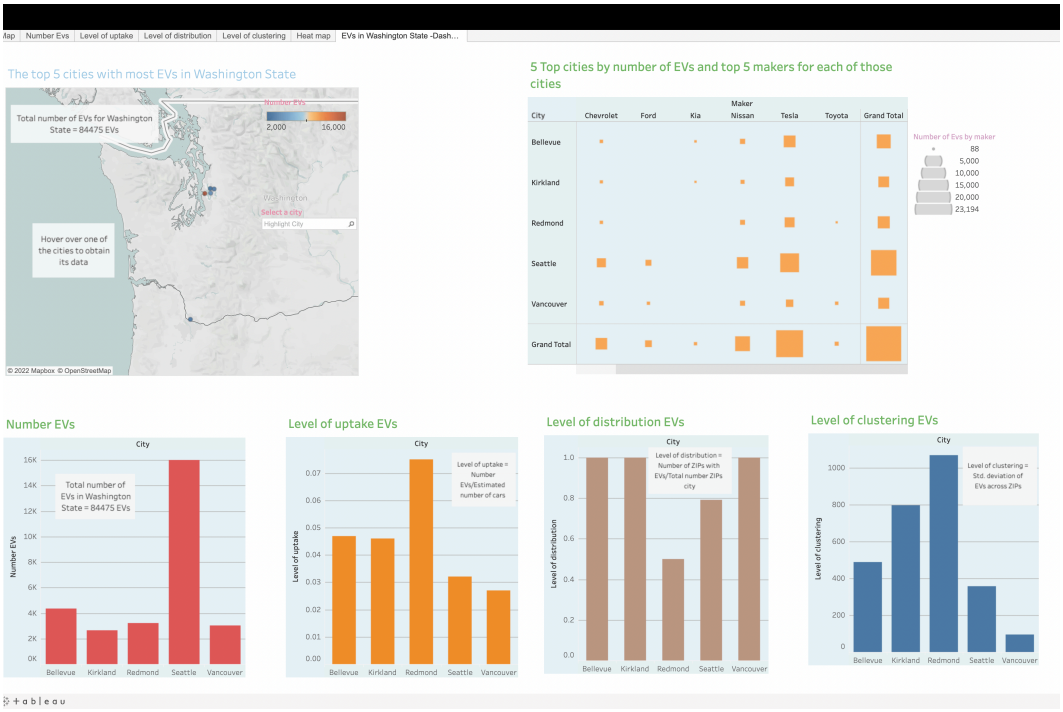
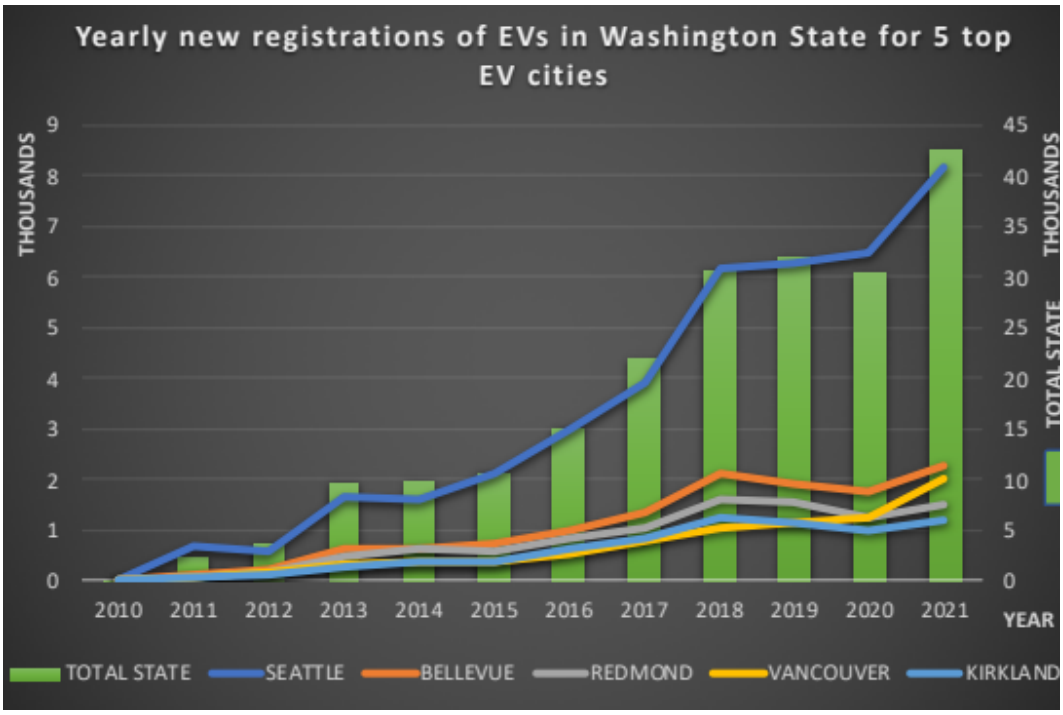


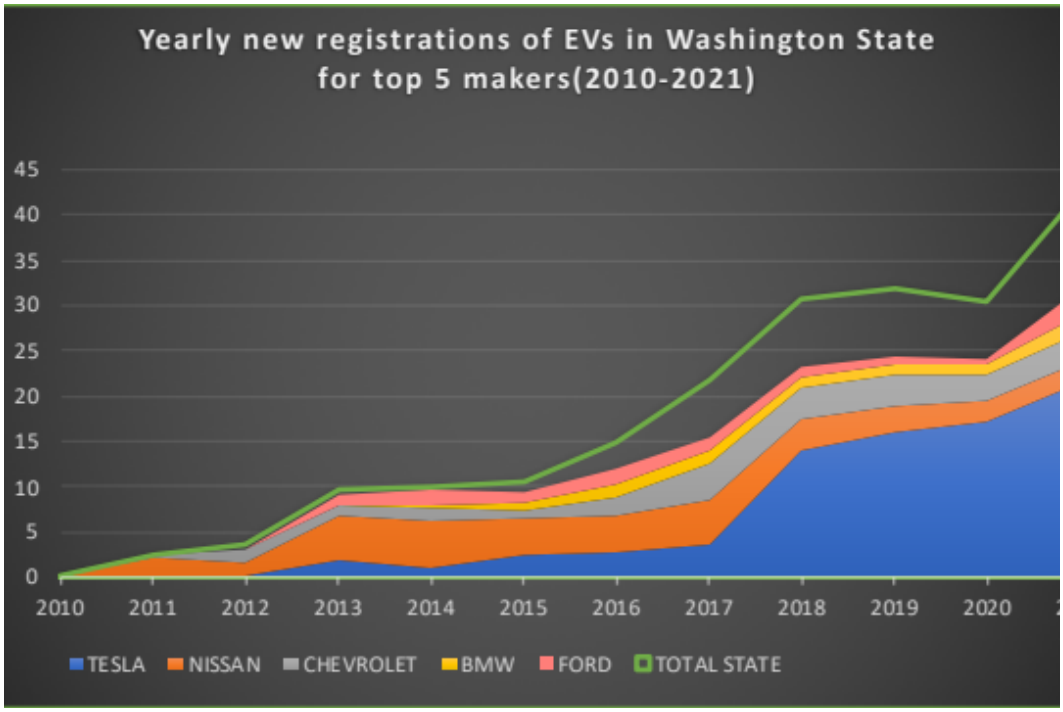
Tableau dashboard

The dashboard presents in a very concise way the main results of the analysis with a variety of graphs. I specially like the “heat map” summarizing the number of EVs and EV makers for each of the cities.

- The historical evolution of EVs in the 5 cities and the makers is presented by two Excel graphs shown below.



Historical evolution of EV numbers in the five cities and the State



Historical evolution of EV numbers for the top five EV makers

The first graph takes advantage of a secondary y axis to display on the same graph the values corresponding to the cities (represented with lines) and the state. The state value (represented with bars) is an order of magnitude larger than the values of the cities and both of them would not be able to display properly if only one scale was used. The use of bars for the state makes easy to identify it from the lines representing the cities.

The second graph is an area plot which represents very clear the evolution of each of the EV makers.

All the graphs were designed with bright colors an high contrast, to facilitate its use by visually impaired people and to attract attention.

6. ACT

The main findings are:

- The two cities recommended for the pilot programme are Vancouver and Bellevue. They are among the five top cities in term of EV numbers and are the closest cities to a homogenous distribution of EVs in their cities.
- A possible third city could be Seattle, but Seattle has shown from the historical and present records that its EV development is different from the other cities.
- The four main EV makers in the five top cities are Tesla, Nissan, Chevrolet and BMW.
- Even Tesla came to produce EVs later than the other makers and, had some slow initial years, its market presence from 2015 has been growing at a faster rate that any other EV maker. The present result is that Tesla EV numbers are twice or more than the other EV makers. Other EV makers seem to maintain almost invariable numbers for the last six years.
- There is a large concentration of EVs in Seattle and surrounding cities. It is important to note that one of the recommended cities is Bellevue, which is very close to Seattle. Therefore is expected that the pilot programme in Bellevue would be able to capture some of the dynamics around Seattle and surroundings.
- The penetration of EVs in the car market still not reach 10% in any of the five top cities. All the five cities, with exception of Seattle, have a similar historical growth.
- As a next step before the pilot programme, the suggestion is to try to identify other sources of data that could reduce the geographical area of study, the ZIP code area, to a smaller one and provide the information about the number of EVs in that area. This would help to refine the possible locations in the recommended cities that could participate in the pilot programme.

APPENDIX A

Set: ***Electric vehicle population data***
https://data.wa.gov/Transportation/Electric-Vehicle-Population-Data/f6w7-q2d2
Data provided by Department of Licensing of Washington State.
Last updated 9th November 2021
Size: 15 columns and 84476 rows
This dataset shows the Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs) that are currently registered through Washington State Department of Licensing (DOL). Each row corresponds to one registered vehicle.

Column Name	Description	Type
VIN (1-10)	The 1st 10 characters of each vehicle's Vehicle Identification Number (VIN).	Plain Text
County	The county in which the registered owner resides.	Plain Text
City	The city in which the registered owner resides.	Plain Text
State	The state in which the registered owner resides.	Plain Text
ZIP Code	The 5 digit zip code in which the registered owner resides.	Plain Text
Model Year	The model year of the vehicle, determined by decoding the Vehicle Identification Number (VIN).	Plain Text
Make	The manufacturer of the vehicle, determined by decoding the Vehicle Identification Number (VIN).	Plain Text
Model	The model of the vehicle, determined by decoding the Vehicle Identification Number (VIN).	Plain Text
Electric Vehicle Type	This distinguishes the vehicle as all electric or a plug-in hybrid.	Plain Text
Clean Alternative Fuel Vehicle (CAFV) Eligibility	This categorizes vehicle as Clean Alternative Fuel Vehicles (CAFVs) based on the fuel requirement and electric-only range requirement in House Bill 2042 as passed in the 2019 legislative session.	Plain Text
Electric Range	Describes how far a vehicle can travel purely on its electric charge.	Number
Base MSRP	This is the lowest Manufacturer's Suggested Retail Price (MSRP) for any trim level of the model in question.	Number
Legislative District	The specific section of Washington State that the vehicle's owner resides in, as represented in the state legislature.	Plain Text
DOL Vehicle ID	Unique number assigned to each vehicle by Department of Licensing for identification purposes.	Plain Text
Vehicle Location	The center of the ZIP Code for the registered vehicle.	Geo Point

Set: ***Electric vehicle title and registration activity***
https://data.wa.gov/Transportation/Electric-Vehicle-Title-and-Registration-Activity/rpr4-cgyd
Data provided by Department of Licensing of Washington State.
Last updated 9th November 2021
Size: 30 columns by 479357 rows.
This dataset contains the records of title activity (transactions recording changes of ownership), and registration activity (transactions authorizing vehicles to be used on Washington State public roads). Each row represents a transaction.

Column Name	Description	Type
Clean Alternative Fuel Vehicle Type	This distinguishes the vehicle as all electric, plug-in hybrid, or compressed natural gas.	Plain Text
VIN (1-10)	The 1st 10 characters of each vehicle's Vehicle Identification Number (VIN).	Plain Text
Model Year	The model year of the vehicle, determined by decoding the Vehicle Identification Number (VIN).	Number
Make	The manufacturer of the vehicle, determined by decoding the Vehicle Identification Number (VIN).	Plain Text
Model	The model of the vehicle, determined by decoding the Vehicle Identification Number (VIN).	Plain Text
New or Used Vehicle	If a vehicle is being sold by its manufacturer to its first owner, it it considered 'new'. Otherwise it is considered 'used'.	Plain Text
Sale Price	The amount that was reported to have been paid for a vehicle. A value of zero indicates that the sale price was not available.	Number
DOL Transaction Date	The day upon which a transaction was recorded into Department of Licensing's computer system.	Date & Time
Transaction Type	The category of activity that was performed.	Plain Text
Transaction Year	The year upon which a transaction was recorded into Department of Licensing's computer system.	Number
Electric Vehicle Fee Paid	The Electric Vehicle Fee is charged to some electric vehicles when they renew their registration. This indicates if it was collected during the transaction.	Plain Text
County	The County that the vehicle's owner is listed to reside within.	Plain Text
City	The City that the vehicle's owner is listed to reside within.	Plain Text
Zip	The 5 digit zip code that the vehicle's owner is listed to reside within.	Plain Text
Electric Range	This tells how far a vehicle can travel purely on its electric charge.	Number
Base MSRP	This is the lowest Manufacturer's Suggested Retail Price (MSRP) for any version of the model in question.	Number

2015 HB 2778 Exemption Eligibility	Shows if new vehicle title transactions were eligible for the sales tax exemption authorized by House Bill 2778 during the 2015 Legislative Session. If not eligible, reasons are provided.	Plain Text
Sale Date	The day on which the vehicle changed ownership.	Date & Time
Vehicle Primary Use	This describes the primary intended use of the vehicle.	Plain Text
State of Residence	This describes the residential or business location of the primary vehicle owner.	Plain Text
DOL Vehicle ID	A unique identification number for each vehicle present in Transactions dataset. Transactions done on the same vehicle will have the same DOL Vehicle ID.	Plain Text
Legislative District	The specific section of Washington State that the vehicle's owner resides in, as represented in the state legislature.	Plain Text
2019 HB 2042 Clean Alternative Fuel Vehicle (CAFV) Eligibility	Shows if vehicle title transactions were eligible for the sales tax exemption authorized by House Bill 2042 during the 2019 Legislative Session. If not eligible, reasons are provided.	Plain Text
Meets 2019 HB 2042 Electric Range Requirement	True = The vehicle model can travel 30 miles or more solely on electricity. Otherwise, False.	Checkbox
Meets 2019 HB 2042 Sale Date Requirement	True = The vehicle was sold on or after August 1, 2019. Otherwise, False.	Checkbox
Meets 2019 HB 2042 Sale Price/ Value Requirement	True = The vehicle sale price/value was $45,000 or less for new vehicles, or $30,000 or less for used vehicles. Otherwise, False.	Checkbox
Odometer Reading	The odometer value taken at the time of the Title Transaction.	Number
Odometer Code	The type of odometer value that was taken at the time of a Title Transaction.	Plain Text
Transportation Electrification Fee Paid	The Transportation Electrification Fee is charged to some electric vehicles when they renew their registration. This indicates if it was collected during the transaction.	Plain Text
Hybrid Vehicle Electrification Fee Paid	The Hybrid Vehicle Electrification Fee is charged to some electric vehicles when they renew their registration. This indicates if it was collected during the transaction.	Plain Text