EV Market Analysis - Washington State Report

Roshen Abraham Sunny (2219833)
Shrinidhi Sudhir (2216988)
Gayathri Ravichandran Geetha (2220497)

INTRODUCTION

Lumon is an American electric vehicle and clean energy company. The company's success can be attributed to its innovative technology, stylish designs, and commitment to sustainability.

The company has a showroom and service center in Georgia and Texas. It has planned to start a new showroom in Washington as it has been a leader in promoting the adoption of electric vehicles. The state offers several incentives for EV buyers, including a sales tax exemption for qualifying vehicles and a reduced vehicle license fee. In addition, the state has implemented policies to support the deployment of EV charging infrastructure, including grants for public and private charging stations and requirements for new residential and commercial buildings to be EV-ready. The state has 281 cities and 39 counties. Lumon wants to know about the proportion of users owning electric vehicles in Washington in each county, to decide which county to be targeted to expand their business.

The company has decided to perform market analyses on both battery and plug-in hybrid electric vehicles that are currently registered in the Washington state department of licensing(DOL). The company wants to know which county has the maximum number of electric vehicles, which make, model, and model year are generally preferred, and which models are mostly preferred in plug-in hybrid and battery electric vehicles.

The company also wants to analyze the demographic data of each county to examine the correlation with the electric vehicle dataset and identify potential markets to expand their business.

DATASET

Dataset 1: Electric Vehicle Population Data

Data source: The data shows battery electric vehicles and Plug-in hybrid electric vehicles that are currently registered through the Washington state department of licensing (DOL). The dataset was published on Feb 17, 2023.
Source:https://catalog.data.gov/dataset/electric-vehicle-population-data

Attributes: It contains 17 attributes such as;
VIN(1-10) - A VIN consists of 17 characters, which are used to provide information about a vehicle's make, model, and other key details.
1st character: Indicates the country where the vehicle was manufactured.
2nd and 3rd character: Identifies the vehicle's manufacturer.
4th to 8th character: Gives information about the vehicle's model, engine size, and other details.
9th character: Used to Verify the validity of the VIN.
10th character: Indicates the year the vehicle was manufactured.
11th character: Identifies the manufacturing plant.
12th to 17th characters: Used to indicate the production sequence number
The obtained dataset shows only the first 10 characters of the VIN since the data is more focused on the vehicle's model, manufacturer, and distribution of vehicles across the state.
County - It shows 39 counties of Washington
City - It represents 281 cities in the state of Washington.
State - It shows the state in which the vehicle was registered.
Postal code - It shows various postal codes of Washington.
Model Year - The dataset shows the year in which the vehicle was released, starting from 1997 to 2023.
Make - It shows the manufacturer or brand of the vehicle. In this dataset, we have about 34 brand names.
Model - A vehicle's model refers to a specific version of the vehicle. In this dataset, we have around 120 models.
Electric Vehicle type - This dataset contains two types of Electric vehicles: Battery Electric vehicles (BEVs) and Plug-in Electric Vehicles (PHEVs).
Clean Alternative Fuel Vehicle (CAFV) Eligibility - It shows whether the particular vehicle is eligible for CAFV.
Electric Range - It shows the distance that the vehicle can travel on a single charge of its electric battery, without the need for any other power source.
Base MSRP - It refers to the base price at which a manufacturer suggests that a vehicle should be sold to customers. Here the price ranges from 31,950 USD to 8,45,000 USD.
Legislative District - The state of Washington has 49 legislative districts. In the dataset, the districts are numbered from 1- 49, and it represents at which district the vehicle was registered.
DOL Vehicle ID - In this dataset, it is the identification number assigned to a vehicle by the Department of Licensing in Washington.

Vehicle location - In this dataset, it shows the longitude and latitude of the respective registered vehicle.
Electric Utility - Here it shows the company or organization that generates, distributes, and sells electric power to the respective manufacturer.
Census Tract - Demographic and social characteristics within a specific geographic area are analyzed using the census tract.

Observations: The dataset contains around 118669 observations.

Dataset 2: Population density of Washington

Data source: The data shows the land area (square mile) and population density (people per square mile) of each county in the state of Washington in various years. From this dataset, we took land area in the year 2020 and population density in the year 2022.
Source Link:
https://ofm.wa.gov/washington-data-research/population-demographics/population-estimates/population-density

Attributes:
Land area (2020) - It shows the land area (Square mile) by each county in the state of Washington.
Population density (2022) - It represents people per square mile of each county in Washington, based on 2022 population estimates.

Observations: The dataset contains 39 observations as there are 39 counties in the state of Washington.

Dataset 3: US Counties

Data Source: The data shows the population spread (i.e.) census demographic data across each county in the USA. From this dataset, we took the Income per capita and Voting age of each county in the state of Washington, for this project.
Source link:
https://www.kaggle.com/datasets/muonneutrino/us-census-demographic-data?select=acs2017_county_data.csv

Attributes:
The income per cap - The data shows the average income earned per person across each county in the state of Washington. It ranges from 18415 USD to 46316 USD.
Voting age citizen - It represents the total number of people who are or above the age of 18 across each county in the state of Washington who has eligibility to vote.

Observations: The dataset contains about 3221 observations.

METHODOLOGY

What steps did you take to download, clean, import, and otherwise transform the data?

The three major datasets that we used include the details of Electric Vehicles registration in Washington state, the population density of Washington state for the year 2022, and the detailed infrastructure picture in various US counties.

From the electric vehicles dataset, we cleaned out data and removed all the null values. Though the data was primarily focused on Washington state, the dataset also has data on a few vehicles that were registered in other states of the US including California, Texas, etc. As this proportion was very negligible (<1%) of the total dataset we dropped these records and filtered out only those records that are required for our analysis.

The dataset population density of Washington state had multiple attributes including the land area and population density from the years 1990 to 2022. We selected only the latest attributes (i.e.) the population density of 2022 for our analysis. We also cleaned the data and dropped the records with NA values for easier analysis.

From the US County dataset, we took the major attributes which included income per capita ratio, to understand the income range of the population in various counties and the population who are eligible to vote. From the voting age population details, we can also understand the ratio of the population who are eligible to get a driver's license and a rough estimate of the population who drive in the county.

What DBMS(s) did you use?

After careful analyses of the dataset, considering the volume of the dataset, we decided to use PostgreSQL for our analyses.

PostgreSQL is considered to be one of the most advanced open-source relational database management systems because of various reasons. It is highly scalable and can handle large-volume datasets and high-traffic loads. In this project, we are using datasets that contain about 118669 observations, which makes PostgreSQL a better choice among other DBMS. Along with that, it supports clustering and replication, which makes it easier to scale. It offers a wide range of advanced features that are not available in other database management systems, such as full-text search, GIS, and JSON data types, which makes it a highly versatile DBMS that can be used for a wide range of applications. It also provides many built-in aggregation functions that are easy to use.

What tools did you use?

For this project, we used multiple tools to alter and update the datasets.

- PostgreSQL - Data Analysis using the SQL queries including aggregations, joins, etc. was performed

- Tableau - We created maps and charts which were used for the visualization
- Microsoft Excel - We used basic excel functions, formulas, and pivot tables to understand the data
- Google suite - Google spreadsheet and Google documents were used to create visualization and generate the report

What queries, functions, extensions, etc. did you use for the analysis?

The various queries used:

- Creating and importing the tables and importing that to PostgreSQL:

> Create table ev_population_wa

```
CREATE TABLE EV_Population_WA (
"DOL Vehicle ID"    NUMERIC PRIMARY KEY,
"VIN(1-10)"         VARCHAR(255),
"County"  VARCHAR(255),
"City"    VARCHAR(255),
"Make"    VARCHAR(255),
"Model"   VARCHAR(255),
"Electric Vehicle Type"         VARCHAR(255),
"Clean Alternative Fuel Vehicle (CAFV) Eligibility"  VARCHAR(255),
"Vehicle Location"  VARCHAR(255),
"Electric Utility"  VARCHAR(255),
"Postal Code"       NUMERIC,
"Model Year"        NUMERIC,
"Electric Range"    NUMERIC,
"Base MSRP"         NUMERIC,
"Legislative District" NUMERIC,
"2020 Census Tract"           NUMERIC);
```

> Create table POPDEN_WA

```
CREATE TABLE POPDEN_WA(
"County Name"       VARCHAR(255) PRIMARY KEY,
"Land Area (sq. mi.) 1990"    NUMERIC,
"Land Area (sq. mi.) 2000"    NUMERIC,
"Land Area (sq. mi.) 2010"    NUMERIC,
"Land Area (sq. mi.) 2020"    NUMERIC,
"Population Density 1900"     NUMERIC,
"Population Density 1910"     NUMERIC,
"Population Density 1920"     NUMERIC,
"Population Density 1930"     NUMERIC,
"Population Density 1940"     NUMERIC,
"Population Density 1950"     NUMERIC,
"Population Density 1960"     NUMERIC,
"Population Density 1970"     NUMERIC,
"Population Density 1980"     NUMERIC,
"Population Density 1990"     NUMERIC,
"Population Density 2000"     NUMERIC,
"Population Density 2010"     NUMERIC,
"Population Density 2020"     NUMERIC,
"Population Density 2021"     NUMERIC,
"Population Density 2022"     NUMERIC);
```

- Creating a new table by combining the 3 different tables (ev_population_wa, popden_wa,uscounties) using the join function - The tables are joined using the county name

```
CREATE TABLE EV_WA_COMB AS
(
SELECT "County","EV Count","Total Population","Population Density 2022",ROUND(("EV Count"/"Total Population"),8) AS "EV Per Capita",totalpop,incomepercap,"Eligible Driver Pop Percent"
FROM (
(
Select "County",Count("DOL Vehicle ID") as "EV Count" from ev_population_wa group by "County"
) AS subquery1
JOIN
(
Select "County Name", "Land Area (sq. mi.) 2020","Population Density 2022",ROUND("Land Area (sq. mi.) 2020" * "Population Density 2022") as "Total Population" from popden_wa
) AS subquery2
ON subquery1."County" = subquery2."County Name"
JOIN
(
SELECT LTRIM(RTRIM(REPLACE(county, 'County', ''))) AS "County_WA",totalpop,incomepercap,ROUND((votingagecitizen/totalpop)*100) as "Eligible Driver Pop Percent"
from uscounties where state = 'Washington'
) AS subquery3
ON subquery2."County Name" = subquery3."County_WA"
)
)
```

- Query to understand the various popular manufacturers of Electric Vehicles

```
SELECT "Make" , Count("Make") AS "Count of makes" FROM EV_Population_WA GROUP BY "Make" ORDER BY Count("Make") DESC
```

- The following query gives the most sold models of Electric Vehicles

```
SELECT "Model" , Count("County") AS "Count of model" FROM EV_Population_WA GROUP BY "Model" ORDER BY Count("Model") DESC
```

- This query gives us more information about each model that was released including their release year, the make, the place they were released, etc. We used the VIN to understand the models

```
SELECT "VIN(1-10)",COUNT("VIN(1-10)") AS "Count of a particular Make and model in a release year","County","Make","Model","Model Year" FROM EV_Population_WA GROUP BY "VIN(1-10)","Make","County","Model","Model Year" ORDER BY Count("VIN(1-10)") DESC
```

- We calculated the Electric Vehicles per capita using the formula Electric Vehicles/Total population. The following query gives us Electric Vehicles per capita for each county

```
SELECT "County","EV Count","TOTALPOP", ROUND(("EV Count"/"TOTALPOP"),8) AS "EV Per Capita"
FROM (
Select "County",Count("DOL Vehicle ID") as "EV Count" from ev_population_wa group by "County"
) AS subquery1
JOIN (
Select "County Name", "Land Area (sq. mi.) 2020","Population Density 2022",ROUND("Land Area (sq. mi.) 2020" * "Population Density 2022") as "TOTALPOP" from popden_wa
) AS subquery2
ON subquery1."County" = subquery2."County Name" ORDER BY "EV Per Capita" DESC;
```

- Electric Vehicles in each county in the state of Washington

SELECT "County",COUNT("Make") FROM EV_Population_WA GROUP BY "County" ORDER BY Count("Make")

- Population Density in each county in the state of Washington

SELECT "County Name", "Population Density 2022" FROM POPDEN_WA ORDER BY "Population Density 2022" DESC

- The preferred vehicle types among the two types of Electric Vehicles (BEV and PHEV)

SELECT "Electric Vehicle Type" , COUNT("Make") FROM EV_Population_WA GROUP BY "Electric Vehicle Type" ORDER BY Count("Make") DESC

- The popular vehicle type among each manufacturer

SELECT "Make","Electric Vehicle Type" , COUNT("Make") FROM EV_Population_WA GROUP BY "Make","Electric Vehicle Type" ORDER BY Count("Make") DESC

- To find the most sold manufacturers and models that has Plug-in Hybrid model

SELECT "Make",COUNT("Make") FROM EV_Population_WA WHERE "Electric Vehicle Type" = 'Plug-in Hybrid Electric Vehicle (PHEV)' GROUP BY "Make" ORDER BY Count("Make") DESC

SELECT "Model",COUNT("Model") FROM EV_Population_WA WHERE "Electric Vehicle Type" = 'Plug-in Hybrid Electric Vehicle (PHEV)' GROUP BY "Model" ORDER BY Count("Model") DESC

- To find the most sold manufacturers and models that has Battery Electric model

SELECT "Make",COUNT("Make") FROM EV_Population_WA WHERE "Electric Vehicle Type" = 'Battery Electric Vehicle (BEV)' GROUP BY "Make" ORDER BY Count("Make") DESC

SELECT "Model",COUNT("Model") FROM EV_Population_WA WHERE "Electric Vehicle Type" = 'Plug-in Hybrid Electric Vehicle (PHEV)' GROUP BY "Model" ORDER BY Count("Model") DESC

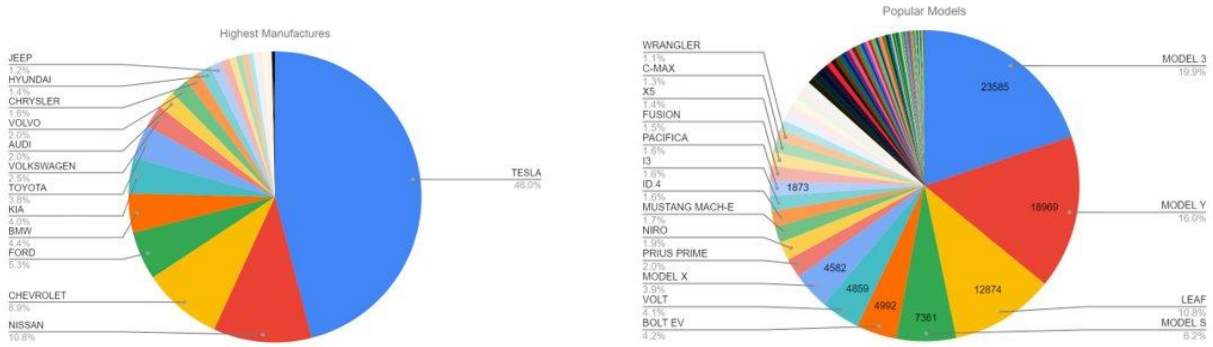
- Query to understand the relationship between the Electric Range for various make and models in PHEV

SELECT "Make", "Model", MAX("Electric Range"),MIN("Electric Range") FROM EV_Population_WA WHERE "Electric Vehicle Type" = 'Plug-in Hybrid Electric Vehicle (PHEV)' GROUP BY "Model","Make" ORDER BY COUNT("Model") DESC

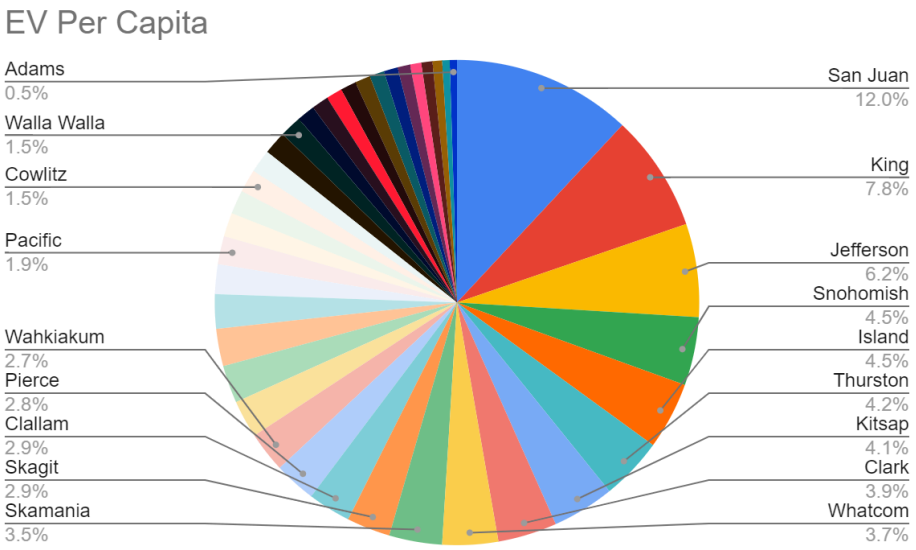
What problems did you encounter and how did you solve them?

Problem:
One of the major problems that we faced was that our primary database did not have the attributes which described the population's income per capita or the population density which were one of the major attributes that contributed to the key analysis.
Solution:
We got the required attributes from three different datasets (Electric Vehicles Population Data, US Counties data, and Population Density of US Counties) which we later joined and created a new table that had all the required attributes for our analysis

ANALYSIS AND FINDINGS



The market leader as of today for electric vehicles is Tesla which contributes around 46% of the total vehicles that have been registered in Washington and the most sold-out model is Model 3, Model Y, and Model S . Following Tesla, Nissan Leaf and Chevrolet's Bolt EV and Volt take the next lead.

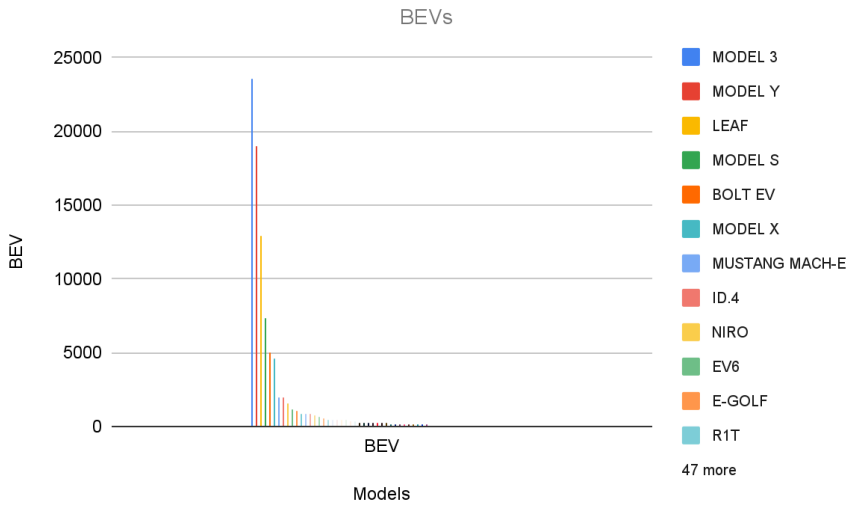


EV per capita denotes the number of electric vehicles by the total population. It can be noted that San Juan has the highest EV per capita ratio which is followed by Kings County, Jefferson County, Snohomish county, and Island county.

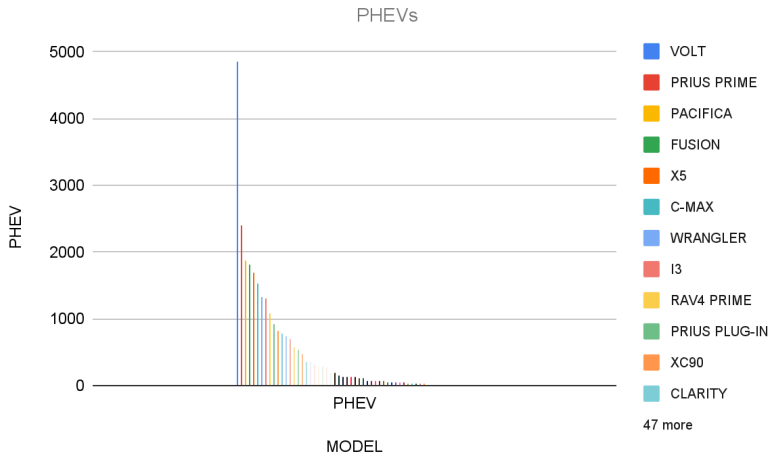
It is also noted that the majority of the population who bought electric vehicles prefer Battery Electric Vehicles when compared to Plug-in Hybrid models because of their high range.

Battery Electric Vehicles Vs Plug-In Hybrid Models

Tesla being the market leader, most of their vehicle types are battery electric vehicles. Tesla takes a huge lead of around twice the second largest manufacturer Nissan. The top three sold models in Battery Electric Vehicle type are Model 3, Model Y, and Volt.



Chevrolet's Volt takes a huge lead in the Plug-in Hybrid model type which is then followed by the Prius-Prime model of Toyota. One major reason for this is the Electric Range of the Volt is very high (i.e. minimum range of 25 mpg and maximum range of 53 mpg) when compared to any other models that have Plug-In Hybrid models.

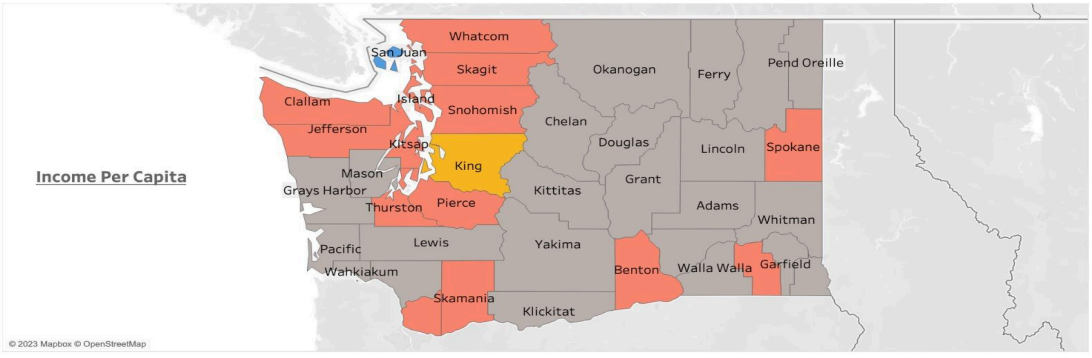


Result 1 | Messages

Make	Model	max	min
CHEVROLET	VOLT	53	35
TOYOTA	PRIUS PRIME	25	25
CHRYSLER	PACIFICA	33	32
FORD	FUSION	26	19
BMW	X5	30	13
FORD	C-MAX	20	19
JEEP	WRANGLER	25	21
BMW	I3	153	72
TOYOTA	RAV4 PRIME	42	42
TOYOTA	PRIUS PLUG-IN	6	6
VOLVO	XC90	35	13
HONDA	CLARITY	48	47
KIA	NIRO	26	26
VOLVO	XC60	35	17
AUDI	A3	16	16
MITSUBISHI	OUTLANDER	24	22
AUDI	Q5 E	20	18

Correlation between EV per capita and Income per capita:

It can be observed that there is a similar trend in the income distribution and the number of electric vehicles in each county within Washington state. San Juan has the highest EV per capita even though King county leads with the highest Income per capita, the reason being San Juan has around 83% of the eligible population who can drive whereas only 63% of the total population drive in King County.



CONCLUSION

Based on our analyses, we observed that Spokane and Benton counties are potential counties to start a showroom and service center. As Spokane county has an Income per capita of 28325 USD, EV per capita of 0.00531, and population density of 312.15, and also Benton county has an Income per capita of 30511 USD, EV per capita of 0.00679, and population density of 124.87. These data show that despite having a higher population of people who are eligible to drive and higher income per capita, we observed that EV per capita is less.

county character varying (255)	drive numeric	carpool numeric	transit numeric	walk numeric	othertransp numeric	workathome numeric	Eligible Driver Pop Percent, numeric
Benton County	81.2	10.7	1	1.7	1.1	4.2	67
Spokane County	78.7	9.2	2.5	2.8	1.7	5.2	75

Along with that as shown in the figure above, we can see that around 81.2 % of people in Benton county and 78.7 % of people in Spokane county prefer to drive. Hence, the proper marketing of new EV firms in these counties can fetch customers who will potentially buy EVs hence increasing the profit of the firm.

REFERENCES

1. :https://catalog.data.gov/dataset/electric-vehicle-population-data
2. https://ofm.wa.gov/washington-data-research/population-demographics/population-estimates/population-density
3. https://www.kaggle.com/datasets/muonneutrino/us-census-demographic-data?select=acs2017_county_data.csv