INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

Electric Vehicle

By – SARTHAK HIRPURKAR

About me

- **Background – Pursuing a Master of Computer Applications (MCA), and you already hold a Bachelor of Computer Applications (BCA) degree**
- **Why I want to learn Data Science? – I have a passion for using the data to solve real-world complex problems. Data is everything, it can solve the problem and Data Science provides a way to do so.**
- **Linkedin - https://www.linkedin.com/in/sarthakhirpurkar/**
- **Github - https://github.com/sarthak2863**

- **Business Problem:**
  - As the Electric Vehicle market grows exponentially, it becomes important for Manufacturers to analyze its growth and future potential in order to maximize profit and meet market needs.

- **Objective of the Project:**
  - This project aims to analyze the past data available and extract insights from it which includes customer preferences, market, geographic locations, type, in the USA.

Column	Description
VIN (1-10)	VIN (Vehicle Identification Number). It provides information about the Manufacturer, Model, Model Year. It's like a car's fingerprints.
County	A county is a subdivision of a state. It is a form of local government. A county is usually a group of towns or villages.
City	A city is a subdivision of county. A city government is responsible for providing public services, maintaining infrastructure, and enforcing local laws.
State	A State is a subdivision of the entire country.
Postal Code	This is also called ZIP Code (Zone Improvement Plan). This helps in breaking down a state further more.
Model Year	The Year when the Model of the Vehicle was manufactured.
Make	The Name of the Company/ Manufacturer.
Model	The Model of the Electric Vehicle.
Electric Vehicle Type	Types of Electric Vehicle: * Battery Electric Vehicle (BEV): Fully powered by electricity. * Plug-in Hybrid Electric Vehicle (PHEV): Uses both an internal combustion engine and a battery charged from an external socket (they have a plug).
Clean Alternative Fuel Vehicle (CAFV) Eligibility	The Column name is itself explainatory. It consists of 3 values: • Clean Alternative Fuel Vehicle Eligible • Not eligible due to low battery range • Eligibility unknown as battery range has not been researched
Electric Range	The amount of electricity the battery can store before its battery needs to be recharged.
Base MSRP	Base MSRP stands for Manufacturer's Suggested Retail Price.
Legislative District	District where EV is registered.

|	
DOL Vehicle ID	It is a unique value given by Department of Licensing to each vehicle
Vehicle Location	This Provides the geographical coordinates for corresponding ZIP Code.
Electric Utility	Electric Utility is a company that generates, transmits, and distributes electricity to customers in a specific geographic area.
Census Tract	It is a further subdivision of the Postal Code, allowing for analysis based on census demographics.

Data Dictionary

Exploratory
Data Analysis

Understanding Missing Data

- The entire data consists of 1,12,634 observations and 17 columns.

- There are 20 Missing Values in the Model, 286 in the Legislative District, 24 in Vehicle Location, and 443 in Electric Utility.

- There are 3 types of Missing Values:

- MCAR

- MAR

- MNAR

- Heatmap helps evaluate the type of missing value in the dataset.

- Observation:

- Wherever there is a missing value in Legislative District Column there is also a missing value in the Electric Utility column. But not vice versa. This suggests that there is a one-way dependency in the column and is a case of MNAR.

- Further analyzing the data, I observed that 99% of the data belongs to Washington State and only for this State Legislative District values exist whereas for others it is a Null.

INNOMATICS
RESEARCH LABS

- The missing values in the Model are dependent on the Make, Electric Vehicle type, and Model Year (MAR) which is Volvo, Battery Electric Vehicle, and the year is 2023. By analyzing the data I can conclude that the Missing values belong to either or both of them which is "XC40" or "C40" model.
- The missing values in the column Vehicle Location is a case of MCAR. After carefully analyzing the data I couldn't find the relationship with other columns in the missing data.

Univariate Analysis: Describing the Column

Electric Range
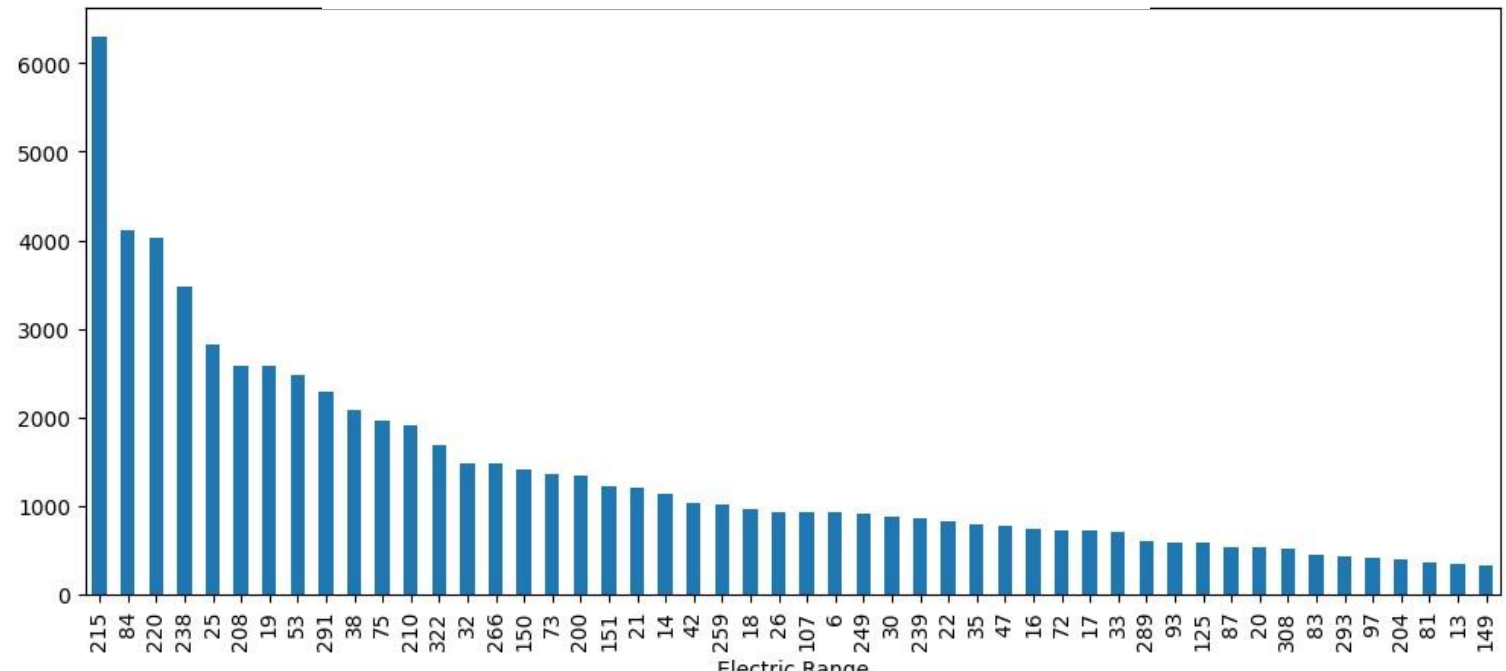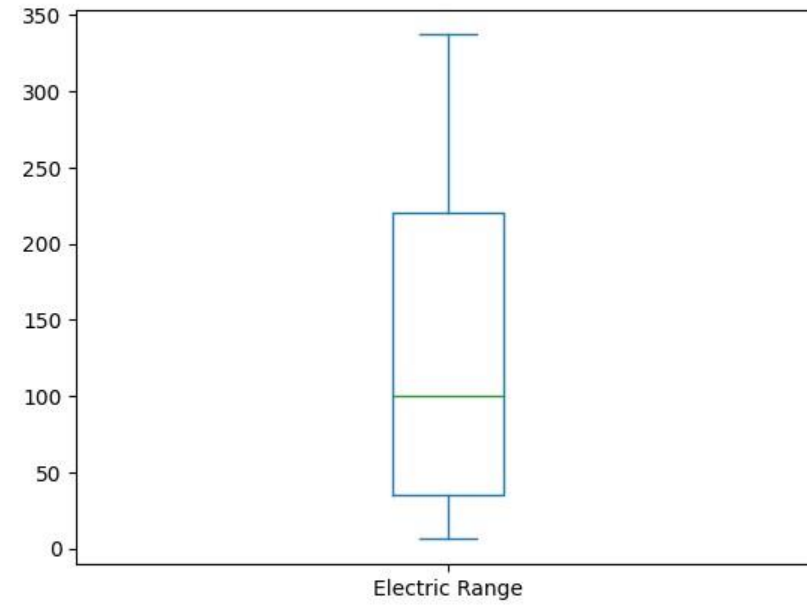
- There are in total of 101 unique rows in this column.

- There are 35% of rows in which Electric Range is set to 0.

- Without considering the values which is 0. I made both box plot and count plot because the values are continuous in nature but since there are only 101 unique values I also made count plot.

- The box plot shows the spread of Electric Range values. The median electric range is around 100 (green line inside the box). The interquartile range (IQR), which covers the middle 50% of the data, spans from approximately 50 to 200. The whiskers extend from about 0 to 350, showing the overall spread of the data without any extreme outliers.

- The bar plot provides a frequency distribution of Electric Range values. The distribution indicates that certain electric ranges (like 215, 284, and 204 miles) are much more common in the dataset than others.

Base MSRP

- The column doesn't give any importance in the EDA only because 96% of the data is imputed to 0.

- Analyzing the data after not considering the 0 value and the outliers.

- The box plot shows the spread of Base MSRP values. The median MSRP is around 60000(green line inside the box). The interquartile range (IQR), which covers the middle 50% of the data, spans from approximately 40000 to 70000. The whiskers extend from about 32000 to 110000, showing the overall spread of the data without any extreme outliers.

- Note: This is highly unreliable.
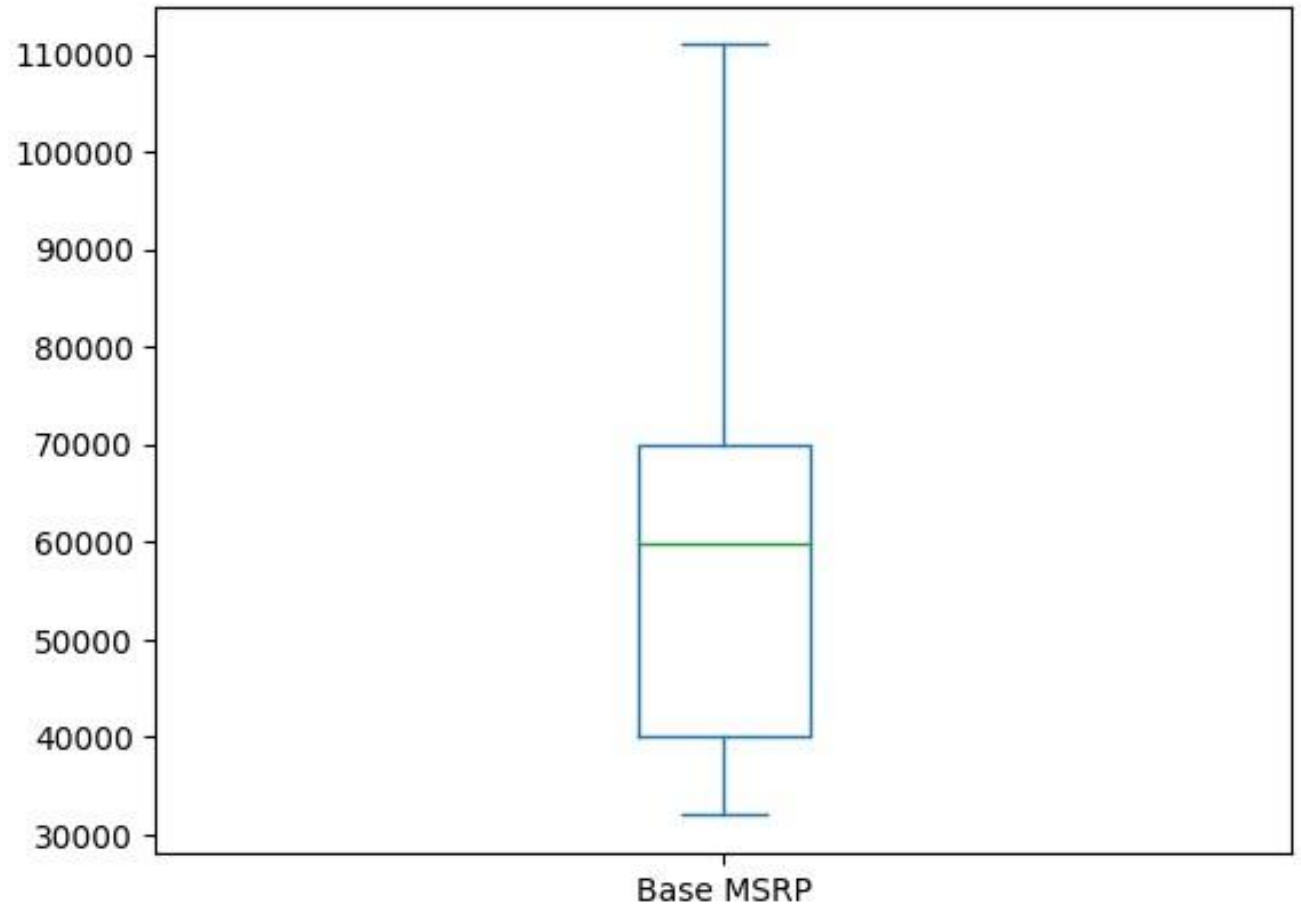
Make

- This is count plot of Make and it shows that Tesla is currently leading in this dataset by covering 46% of the dataset.

- Nissan only covers 11.4% of sales followed by Chevrolet which covers 9%.
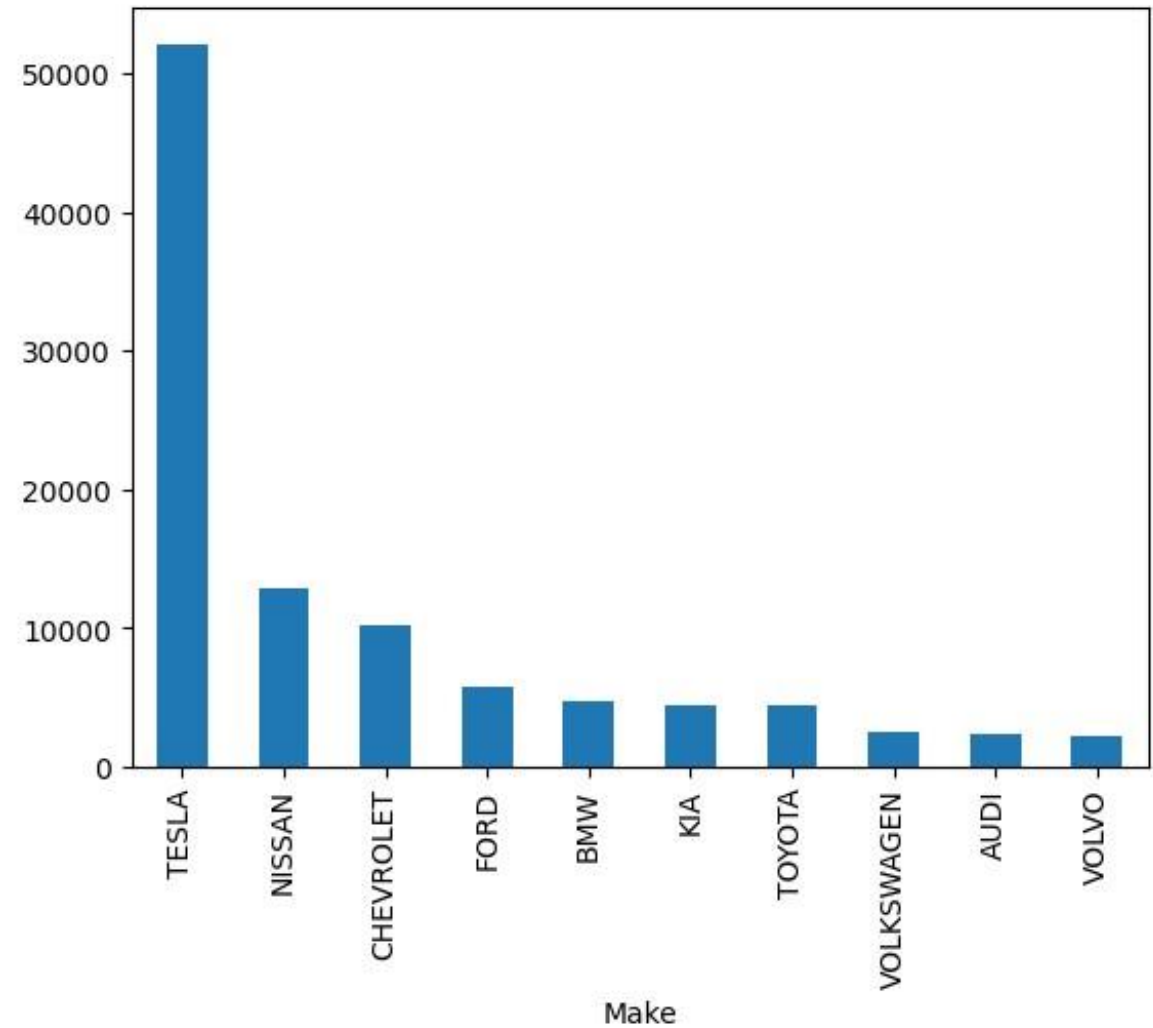
Model
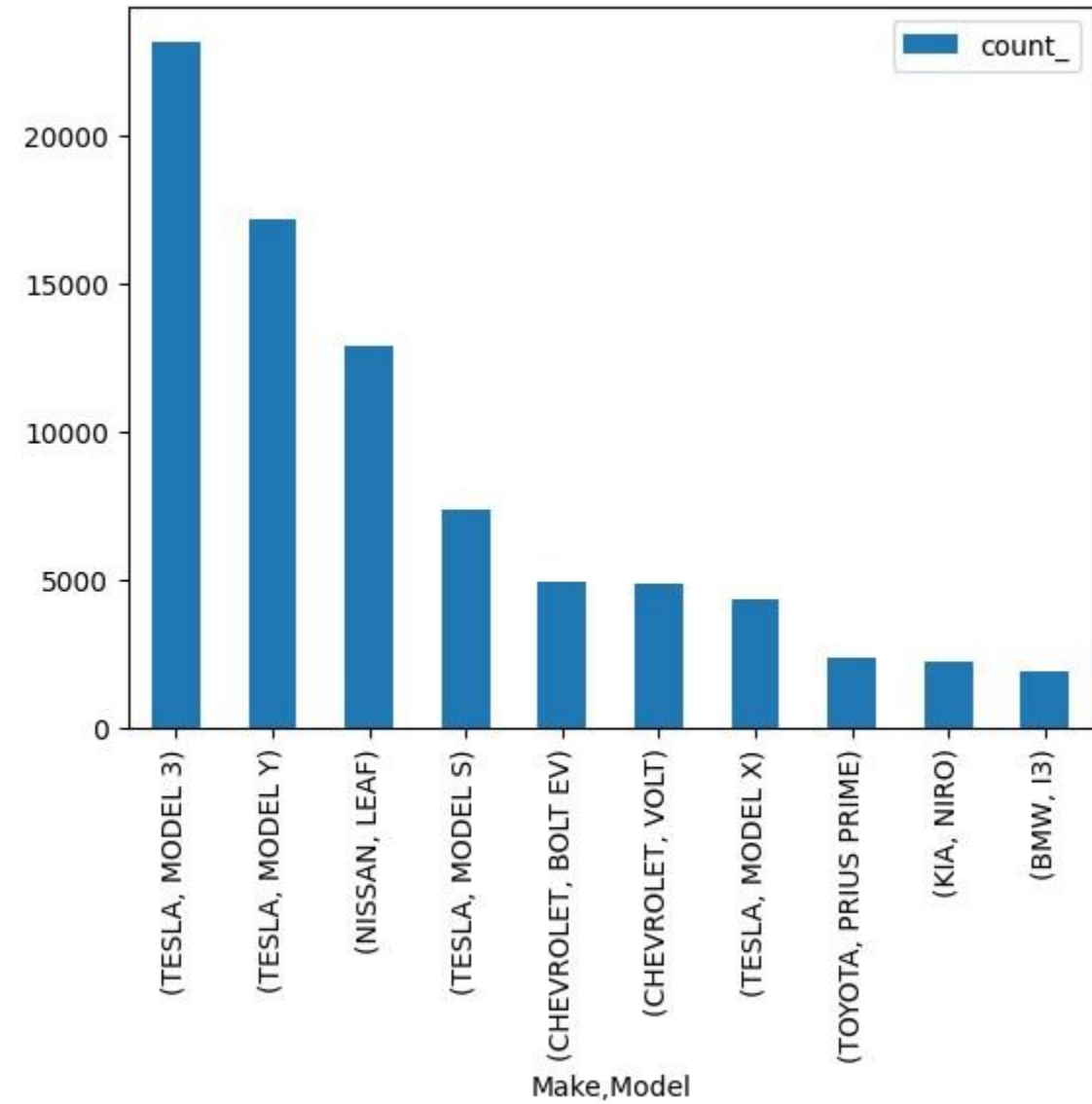
- This is a count plot of Model Column. It clearly shows that Model 3 by Tesla is currently leading by covering 20% of dataset followed by Model Y of Tesla which covers 15% of dataset, followed by Leaf by Nissan which covers 11% of dataset.

Model Year

- This is count plot on the column Model Year

- This shows the sales of EV in USA. The sales Started from the Year 1997 and significantly increased in 2011 and then on kept increasing.
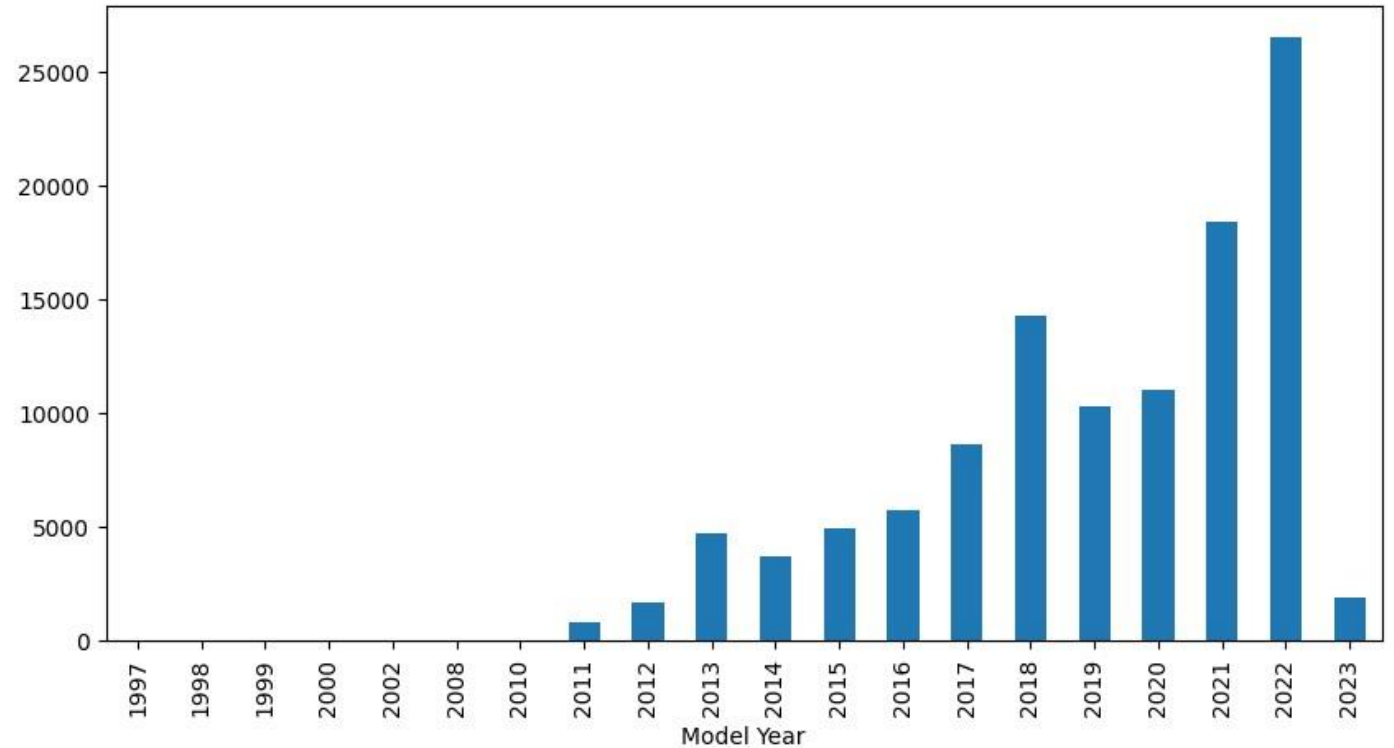
State

- There were found to be 45 states in the entire dataset. The highest sales were recorder in the dataset for the state Washington (WA) with over 99.7% (1,12,348) of the data. Whereas for others it was significantly low (<100).

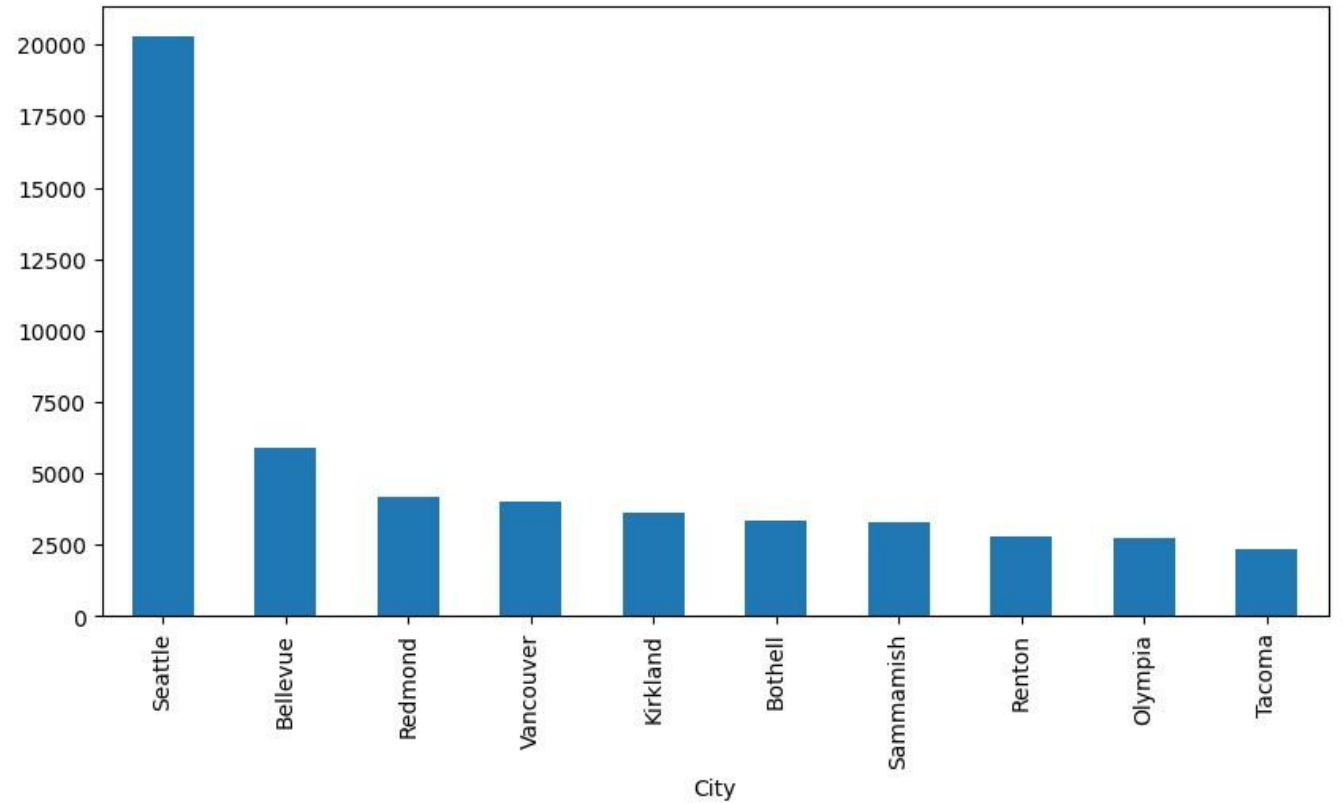- The count plot shows the top 5 States with having higher sales in this dataset.
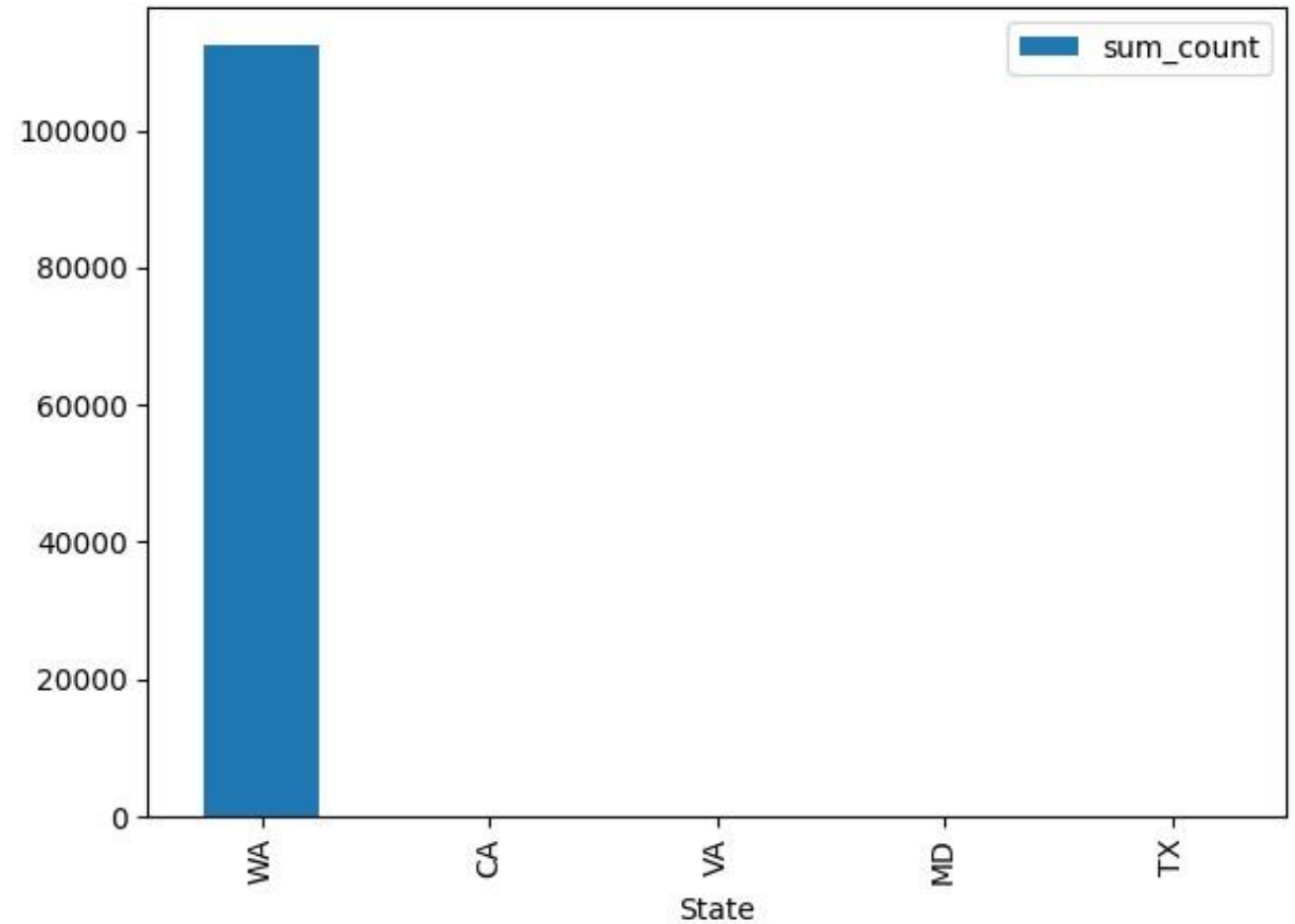
County

- This Count plot tells the sales in each county

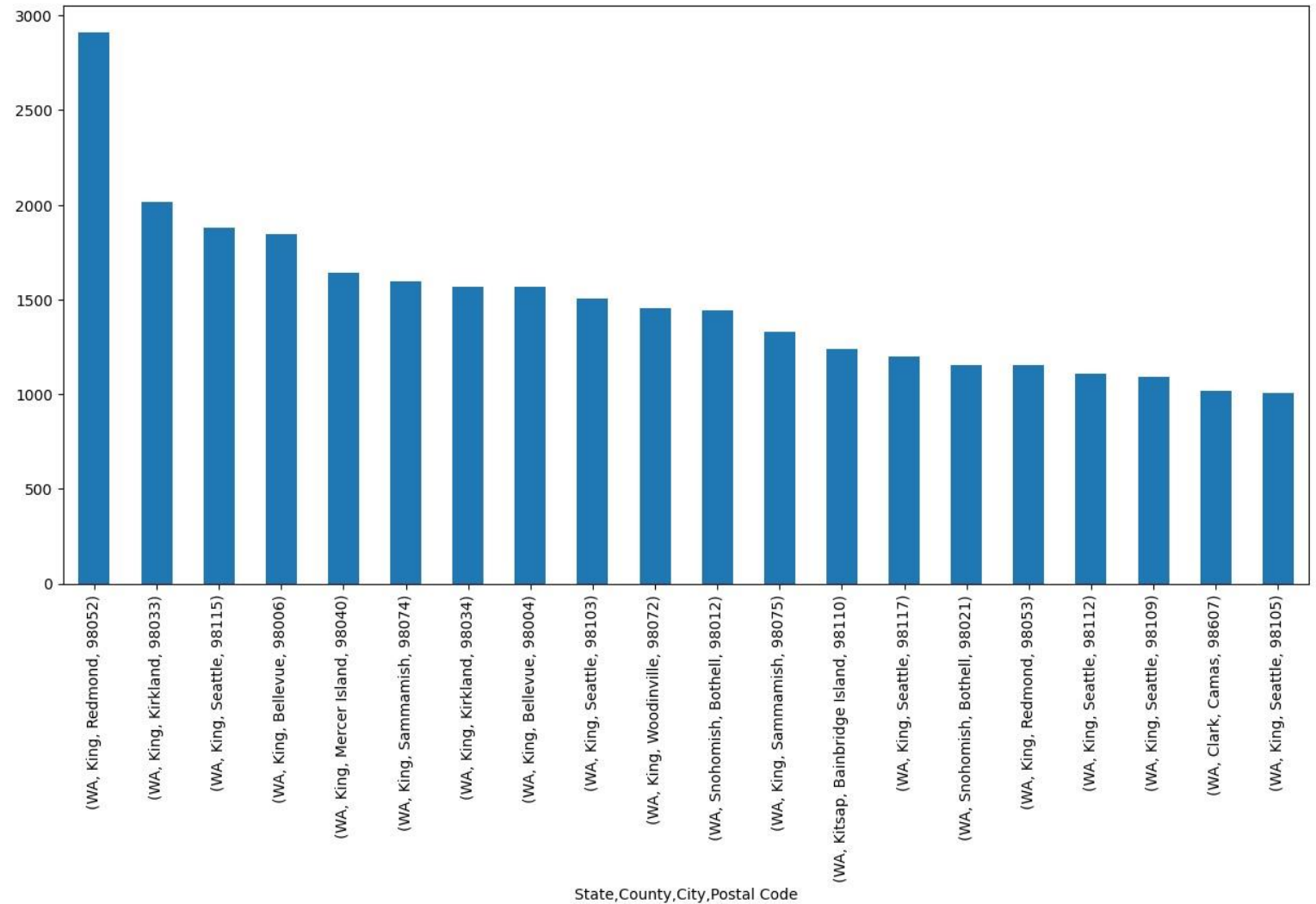- It evident that highest sales is for the King County which is in WA.
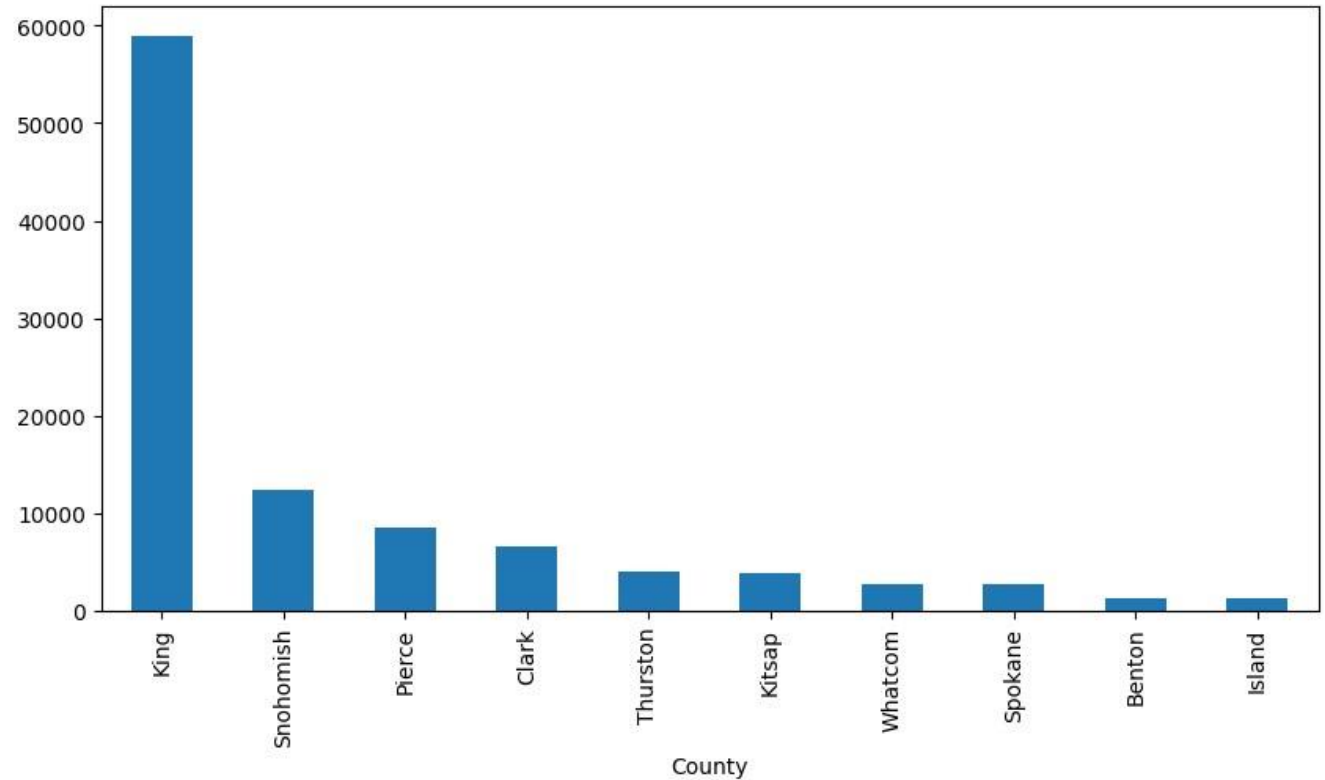
City

- There was highest Sales recorded in the City called Seattle of about 20000.

Postal Code

- This is a count plot of the most occurred Postal code in the dataset.

- It is clear that the postal code 98052 records the highest EV compared to other postal codes.
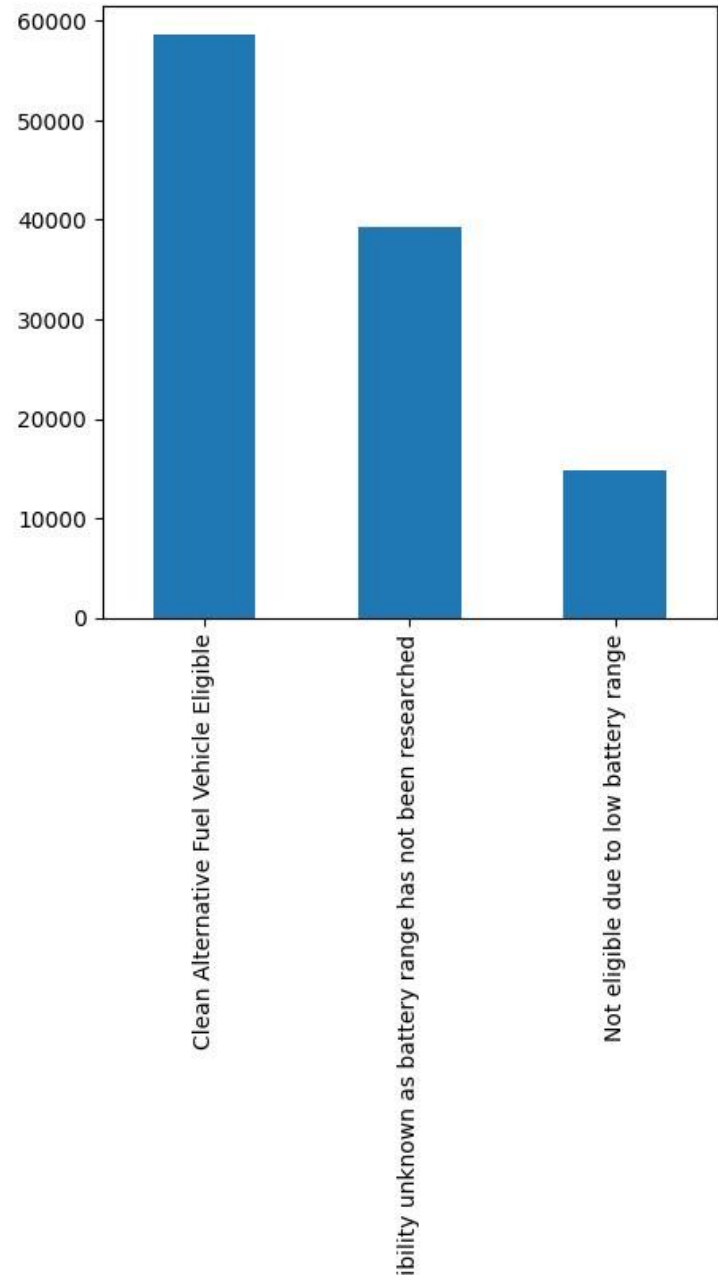

State,County,City,Postal Code

Electric Vehicle Type

- This column has 2 unique values. Battery Electric Vehicle and Plug in Hybrid Vehicle.

- The BEV accounts to 76% of the dataset. Whereas the PHEV accounts to only 23%.
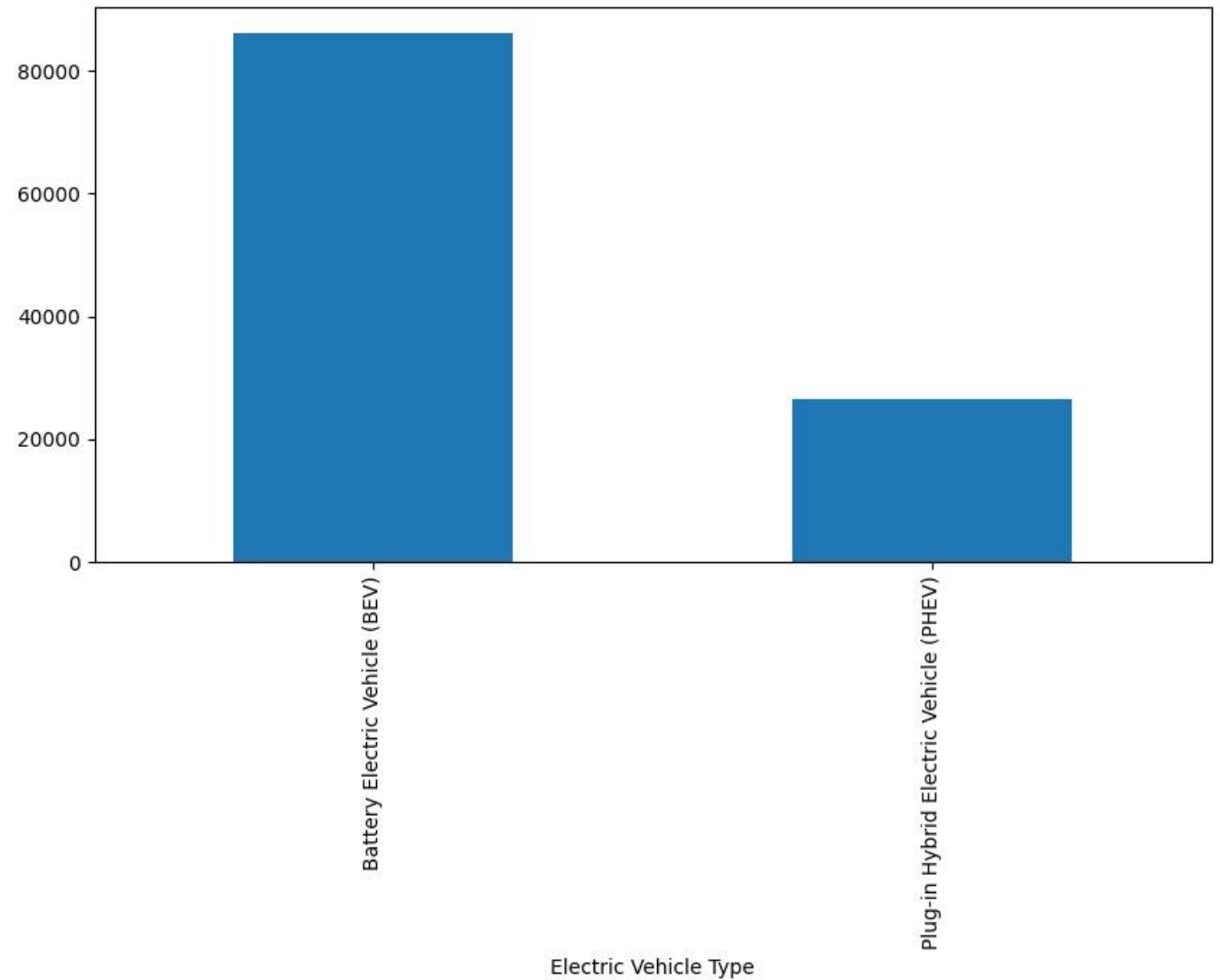
Clean Alternative Fuel

Vehicle (CAFV) Eligibility

- By Analyzing this column it was observed that 35% EV's Clean Alternative Fuel Vehicle (CAFV) Eligibility is unknown because the battery range has not been researched.

- 52% of the data is Eligible and only 13% of data is not eligible due to low battery range.

Bivariate Analysis:

Relationship B/w 2 Variables

Relationship B/w Electric Vehicle Type, Clean Alternative Fuel Vehicle (CAFV) Eligibility, Electric Range

- If the car's battery range is unknown, the electric range is set to 0, meaning the car's eligibility status hasn't been determined because the battery range hasn't been researched and all of these are Battery Electric Vehicle (BEV)

- When the electric range is less than 30, the vehicle is marked as "Not eligible due to low battery range."

- For cars with an electric range of 30 or more, the vehicle qualifies as "Clean Alternative Fuel Vehicle Eligible."

- Battery Electric Vehicle tends to have high electric range (337) whereas the maximum for Plug-in Hybrid Electric
 Vehicle (PHEV) was recorded till 153.

Electric Range Comparison: BEV vs PHEV
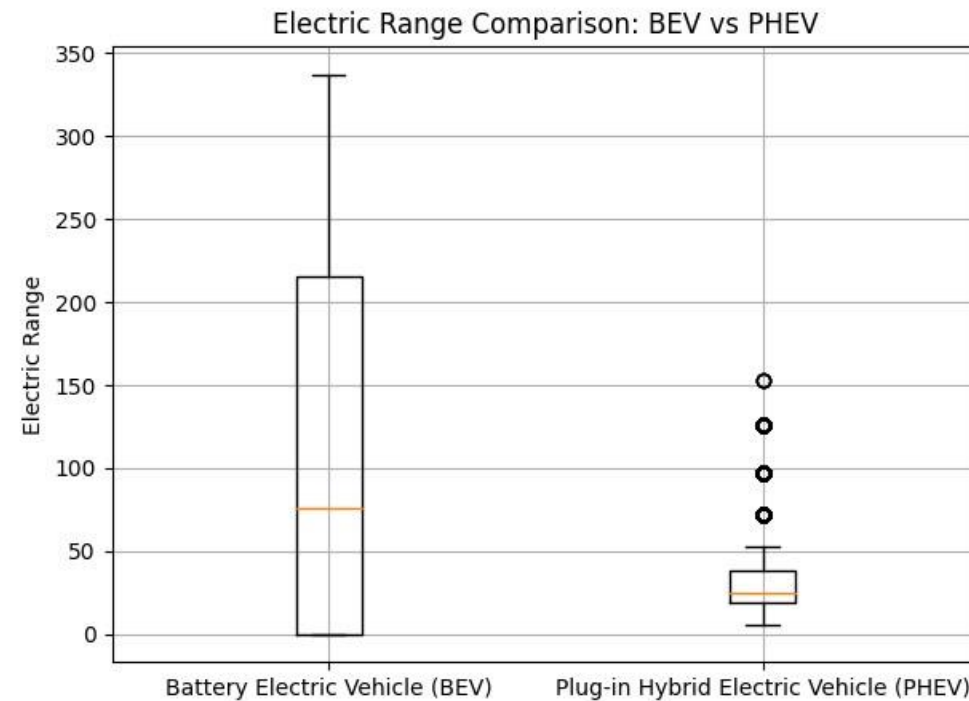
Electric Range Comparison: Not eligible due to low battery range vs Clean Alternative Fuel Vehicle Eligible
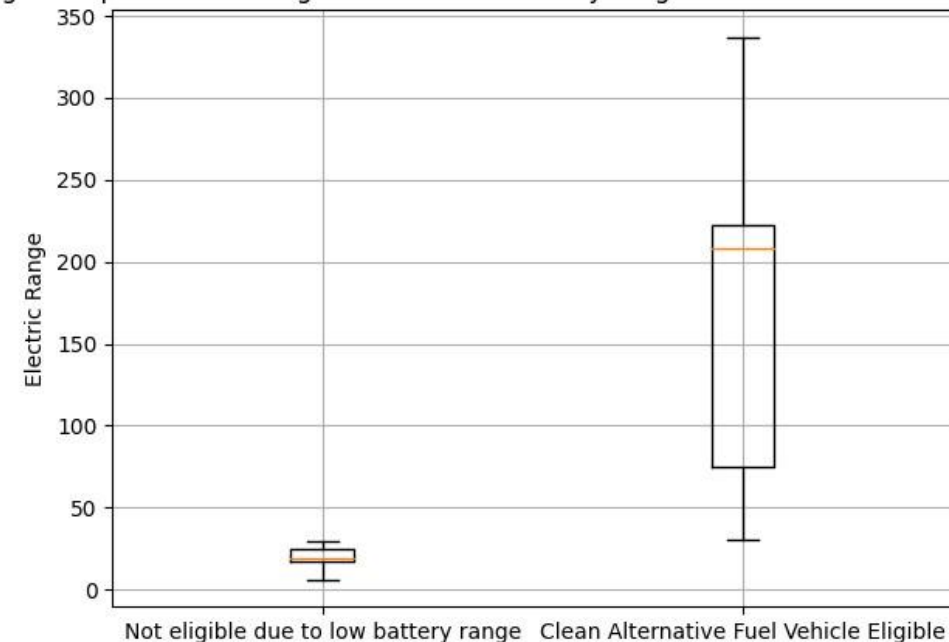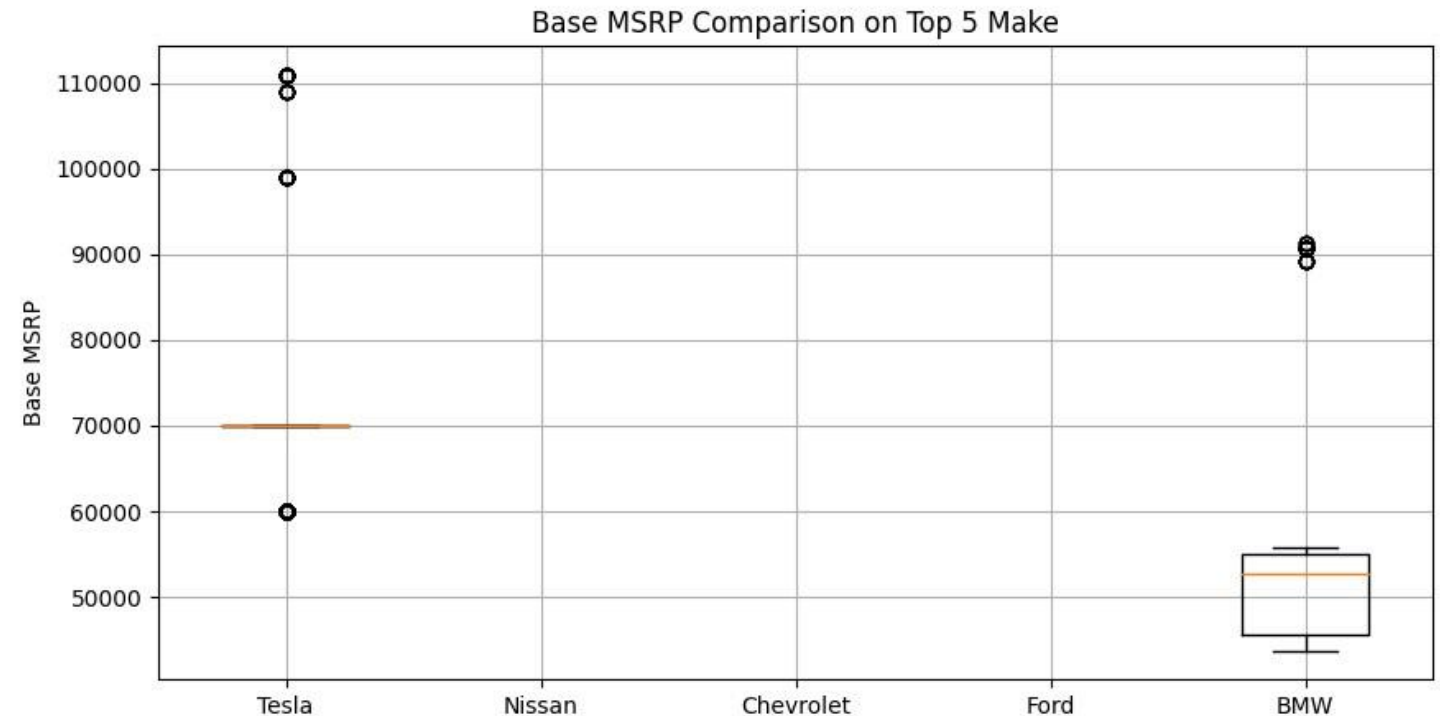
Electric Range Vs Top 5 Make

- This box plot was analyzed without considering the 0 Value present in the Electric Range.

- Form this plot it is clear that Tesla offers the best Electric Range which starts from 200. It has the highest median electric range, with a box that is positioned higher than the others.

- Chevrolet though offers wide range of electric range, but it is still lower than Tesla.

- Ford has the lowest median electric range among the five makes and its maximum is at 100.

Electric Range Comparison on Top 5 Make
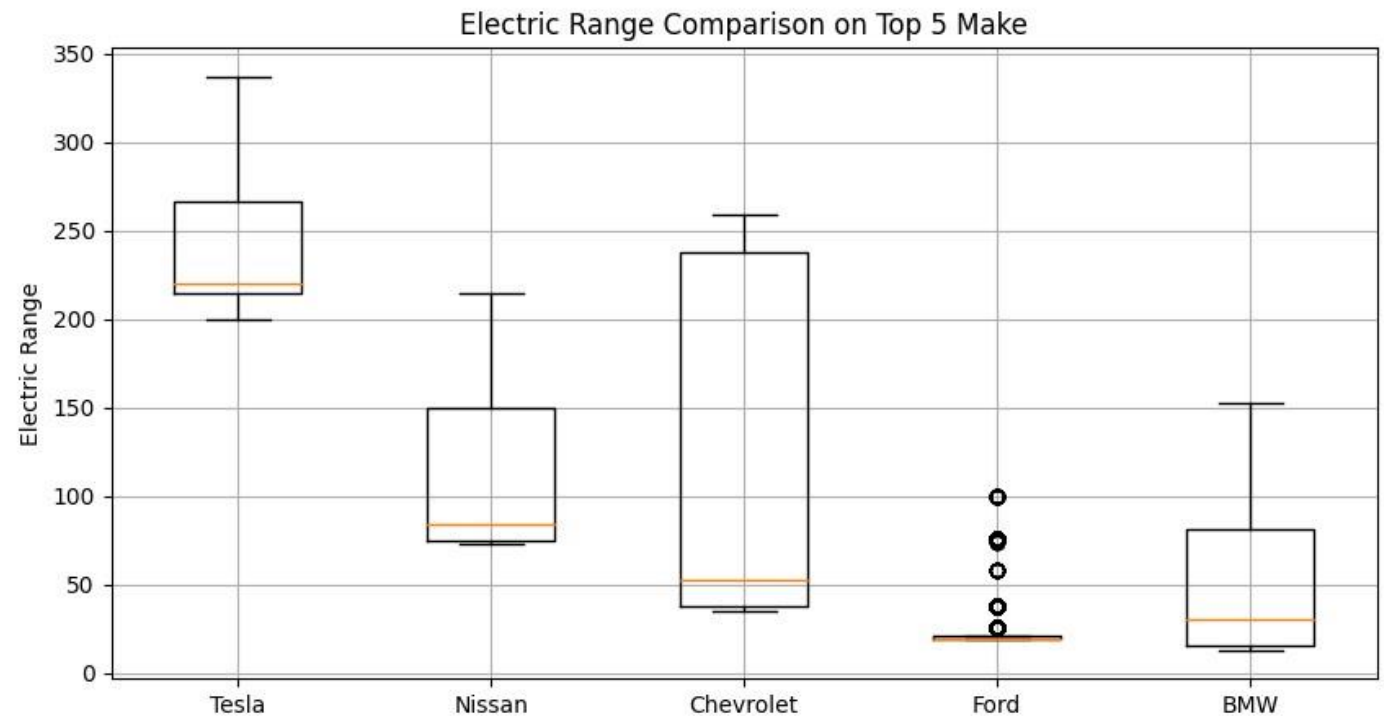
Base MSRP vs Top 5 Make

- The plot is a box plot that compares the base MSRP (manufacturer's suggested retail price) of five different car makes: Tesla, Nissan, Chevrolet, Ford, and BMW. This plot was made without considering the 0 values in Base MSRP.

- Nissan, Chevrolet and Ford didn't provide Base MSRP besides the 0 Value.

- BMW has only 10% of data for Base MSRP. It median is approximately 52,600.

- Tesla has only 3% of data for Base MSRP. Its median is at 69900.


Base MSRP Comparison on Top 5 Make

Electric Range vs top 5 County

- The plot is a box plot that compares the electric range of electric vehicles in the top 5 counties in the United States. This plot was made without considering the 0 value in electric Range.

- King County has the highest median electric range, with a box that is positioned slightly higher than the others. It has the median of 150 which is higher than other county followed by Snohomish at approximately 110.

- Pierce, Clark, and Thurston Counties have similar median electric ranges, which are slightly lower than King County.

- The whiskers of the box plots show the minimum and maximum values for each county. All five counties have similar lengths of whiskers, indicating a similar range of electric range values among electric vehicles in each county.

Electric Range Comparison on Top 5 County

INNOMATICS
RESEARCH LABS
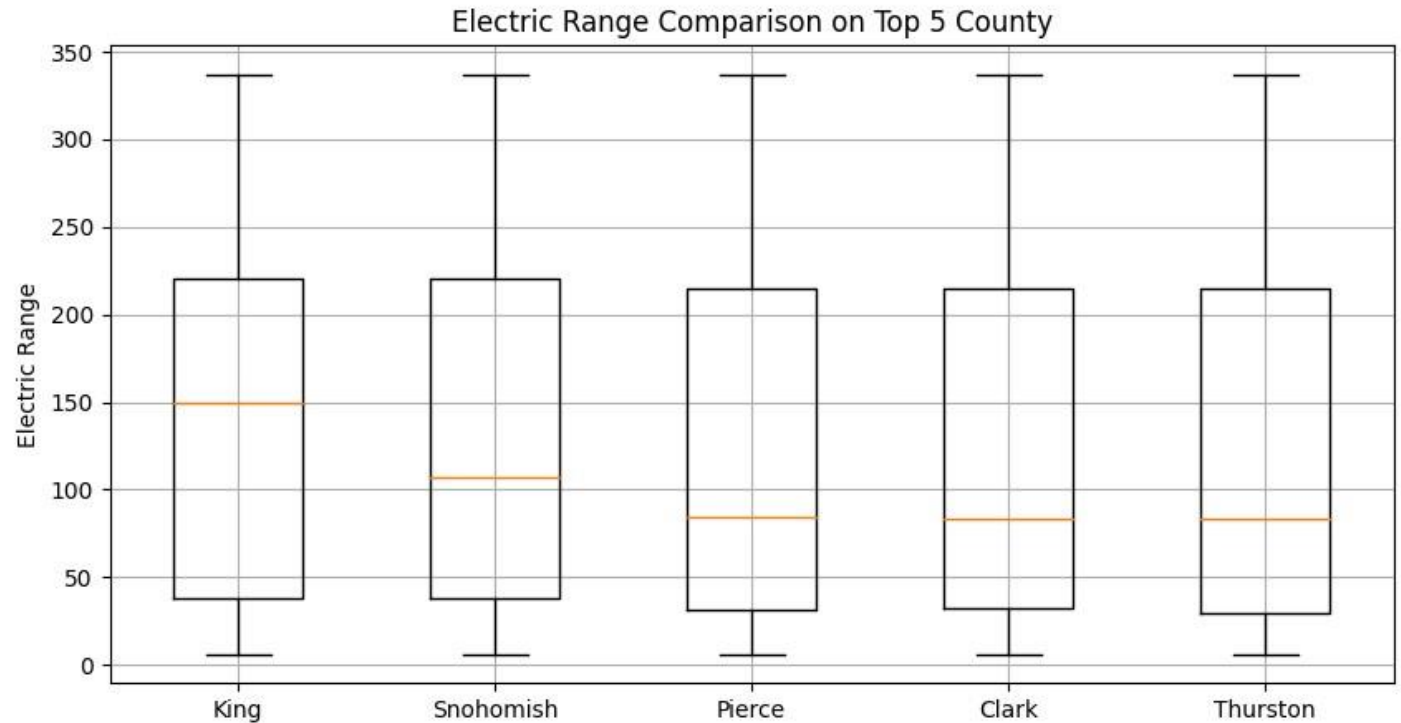
Electric Range vs top 5 City

- The plot is a box plot that compares the electric range of electric vehicles in the top 5 cities in the United States.

- Seattle which has the highest count recorded has low electric range around 110.

- Bellevue, Redmond and Kirkland has have similar median electric ranges, with the boxes being positioned at same height. Their Median is at 210.

- Vancouver has slightly lower Median range compared to Seattle and is the lowest in the top 5 city. Its median is approximately 70.

Electric Range Comparison on Top 5 City
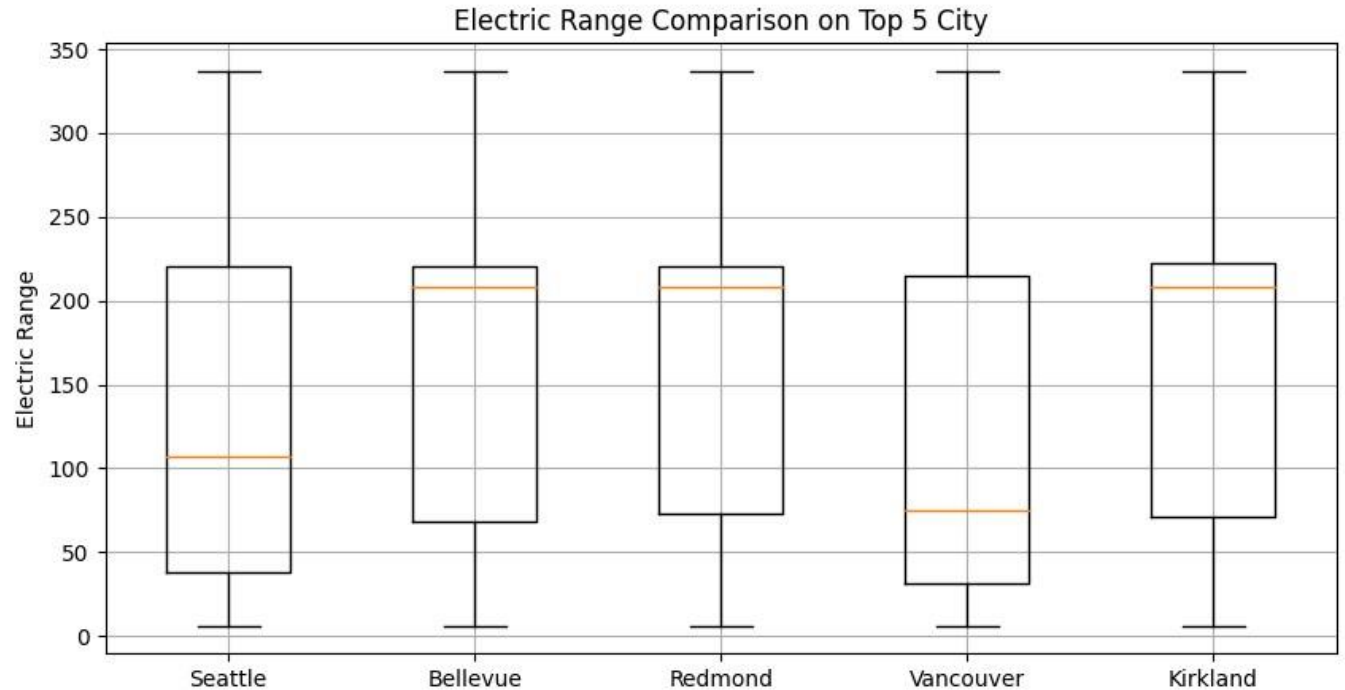
Electric Range vs Top 6 Years

- The plot is a box plot that compares the electric range of electric vehicles across six years: 2023, 2022, 2021, 2020, 2019, and 2018. This Analysis was conducted by not considering Electric Range value at 0.

- It was observed that from the year 2018 to 2020 there was an increase in Electric Range.

- But in the year 2021 and afterwards there is decline in electric range. It is due to the fact that the Manufacturers were not providing the Electric range therefore it was set to 0. And only the vehicle having electric range below 50 was provided

Electric Range Comparison on Top 6 Years
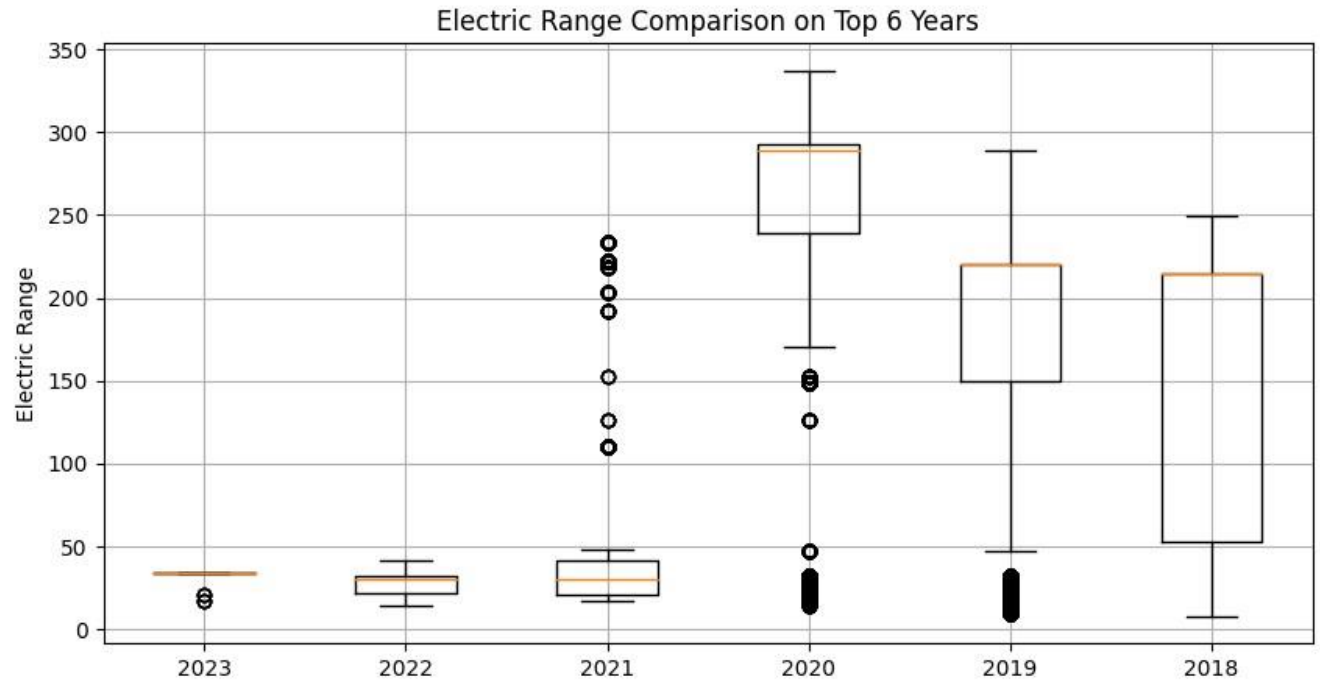
# Electric Range vs top 5 Electric Utility providers

- This is a box plot comparing the electric range of vehicles across the top 5 electric utilities.

- All 5 Electric Utilities seem to have a similar range distribution, which means that the electric range is fairly similar across these providers.

- The Median of the 1st provider is comparatively higher than other (i.e 200). Followed by 3rd provider at 110.

- The median of the other 3 providers is around 90.
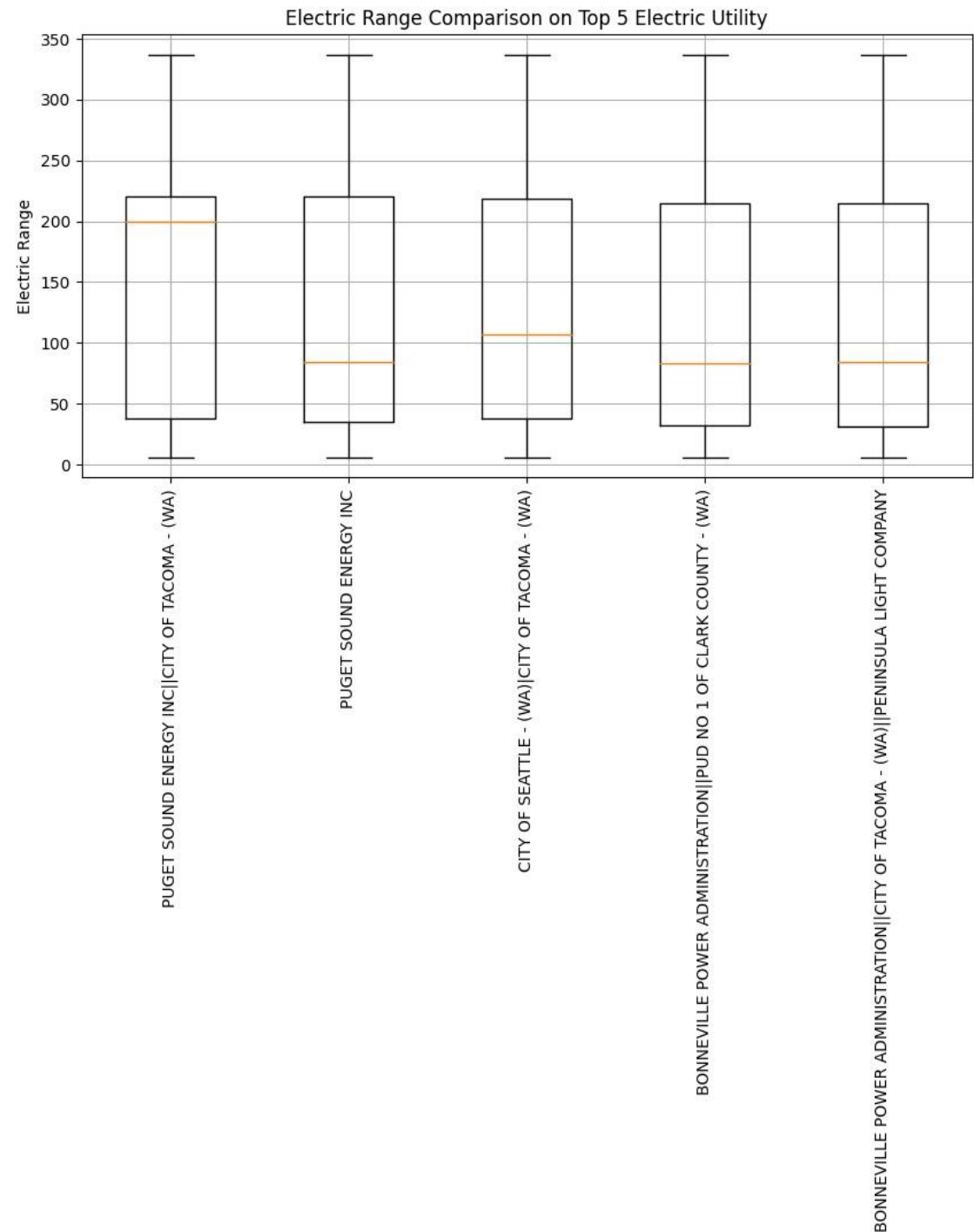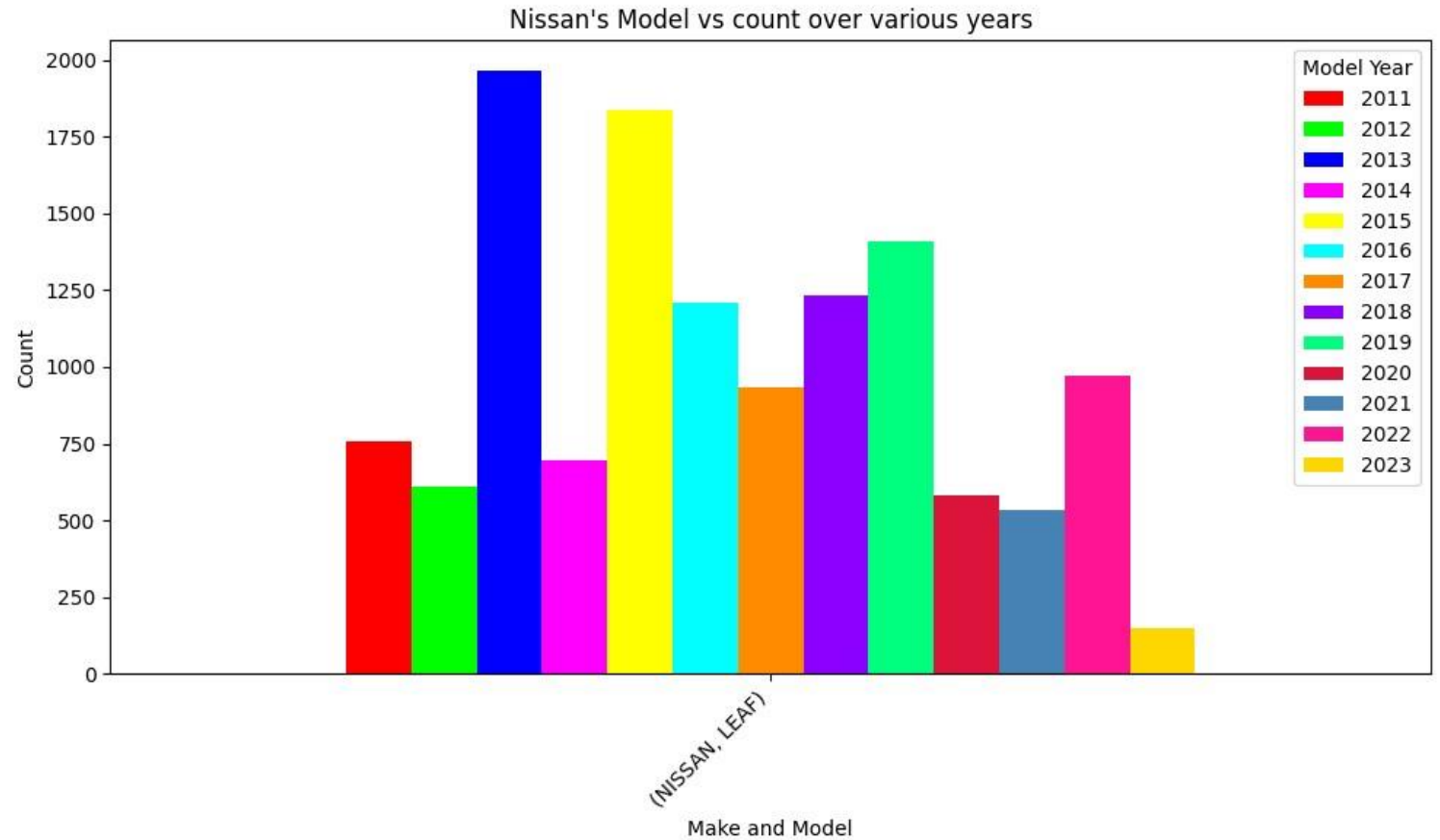
Tesla's Model vs count over various years

- The plot is a grouped bar plot. On Xaxis we have Make and Model whereas on Y-axis it is the count.

Electric Range Comparison on Top 5 Electric Utility

- The dataset has 5 unique Model for the Make Tesla

- Tesla's first model is Roadster which hasn't recorded significant purchases only a few in 2008, 2010, 2011.

- Model S was released in year 2012. It has been recorded continuous increase in purchase upto 2015. After which the purchases dropped down.

- Model 3 was released in year 2017. It recorded high purchase in the Year 2018. and 2022.

- Model Y was released in the year 2020 (Most recent) and recorded the highest sales upto now in the year 2020, 2021, 2022, 2023.
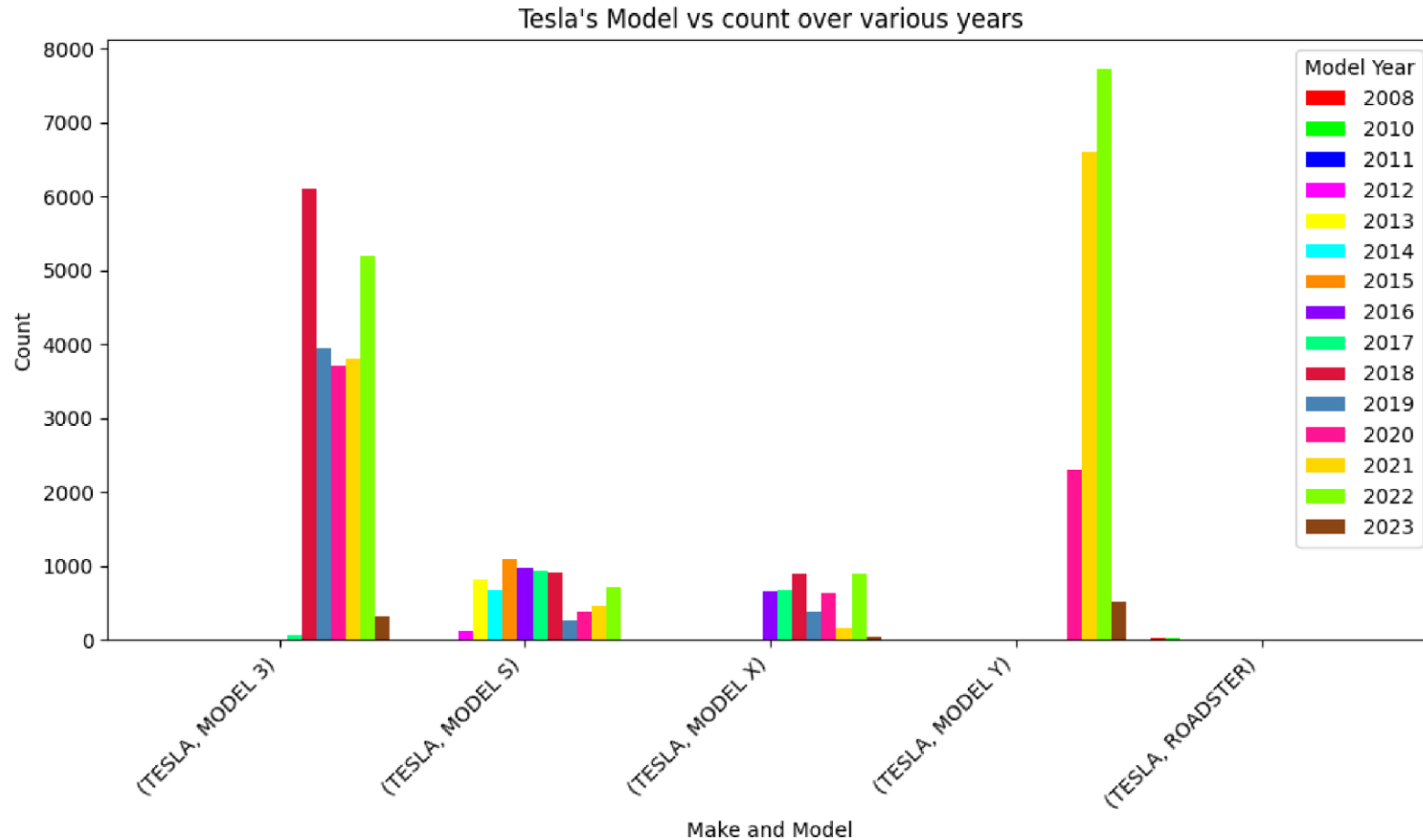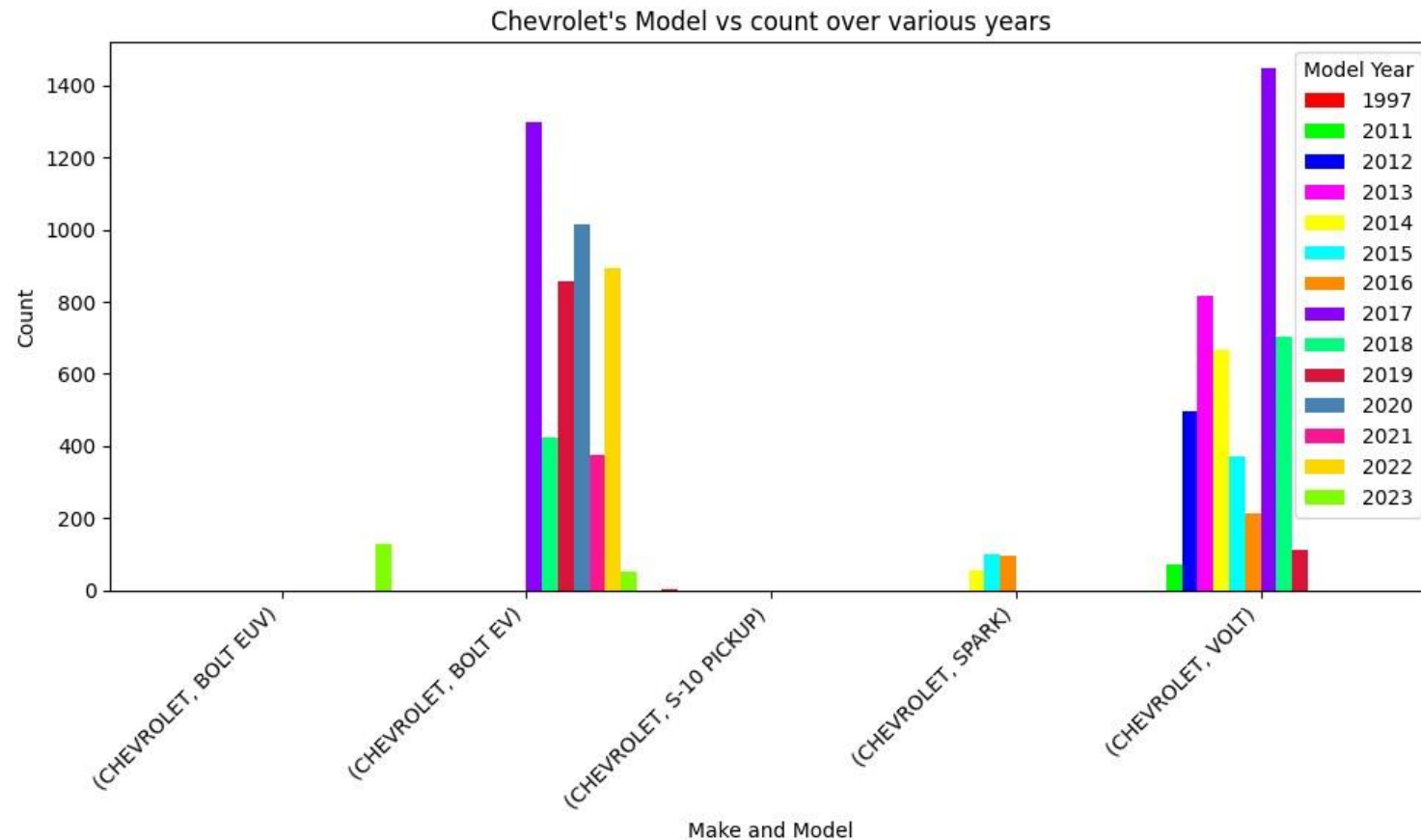

Tesla's Model vs count over various years

Nissan's Model vs count over various years

- The plot is a grouped bar plot. On Xaxis we have Make and Model whereas on Y-axis it is the count.

- In this dataset, It only has one Model which is Leaf. This was launched in the year 2011.

- It has been recorded continuous increase in sales till year 2019. 2013 and 2015 has recorded the highest sales.

- From the year 2020, it seems that there was a decrease in sales.

- The highest sales recorded were approximately 1900.

Nissan's Model vs count over various years
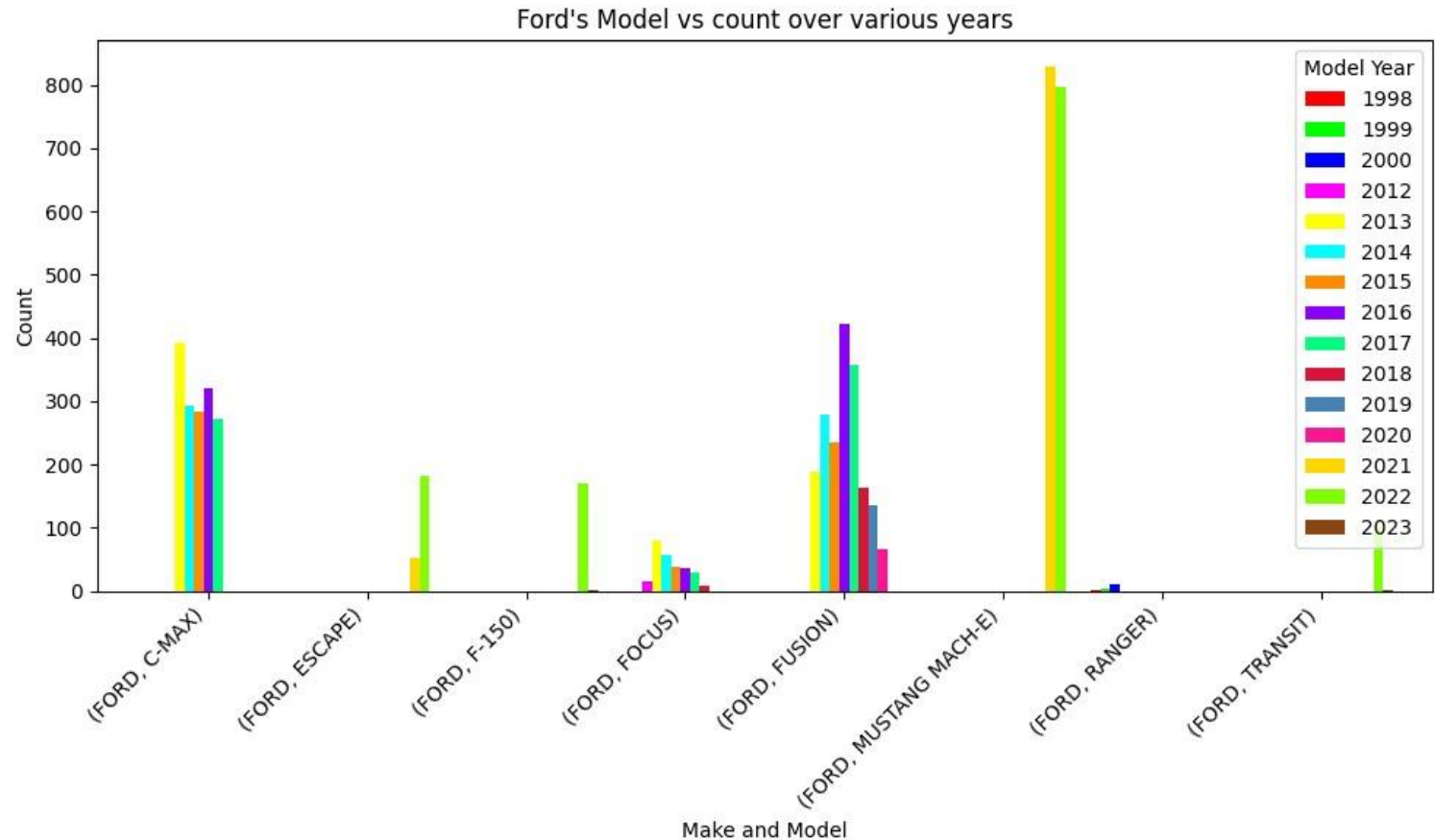
Chevrolet's Model vs count over various years

- The plot is a grouped bar plot. On Xaxis we have Make and Model whereas on Y-axis it is the count.

- It has released upto 6 Model till now.

- S-10 Pickup was the first EV model by Chevrolet which recorded only 1 sale in the year 1997.

- Volt was the second model which was released in the year 2011. It has recorded highest sales in 2017.

- Spark was the next model which was released in the year 2014. It didn't perform very well (in terms of sales) and lasted till 2016.

- Bolt EV is the next Model which was released in the year 2017 and its sales was high in the same year.

Chevrolet's Model vs count over various years

- Bolt EUV is the most recent model which was released in year 2023. Since the dataset is limited till year 2023 its sales was unrecorded.

Ford's Model vs count over various years

- The plot is a grouped bar plot. On Xaxis we have Make and Model whereas on Y-axis it is the count.

- It has released upto 8 Model till now.

- C-MAX, Fusion, and Mustang Mach-E has recorded the highest sales compared to other Models.

- C-Max was released in year 2013 and recorded highest sales in the same year. Its distribution seems similar for the rest of the years.

- Fusion was also released in year 2013 and recorded highes sales in the year 2016. After which its sales started declining.

- Mustang Mach-E was released in year 2021 and recorded highest sales compared to the other Model

- Transit was released in 2022 but it didn't perform very well in the market.

BMW's Model vs count over various years

- The plot is a grouped bar plot. On Xaxis we have Make and Model whereas on the Y-axis it is the count.

- It has released upto 11 Models till now.

- Model I3 and X5 seems to have performed exceptionally well interms of other models.

- I3 was released in year 2014, Its sales was also the highest in the same year. It sales has been declining after the year 2018.

BMW's Model vs count over various years
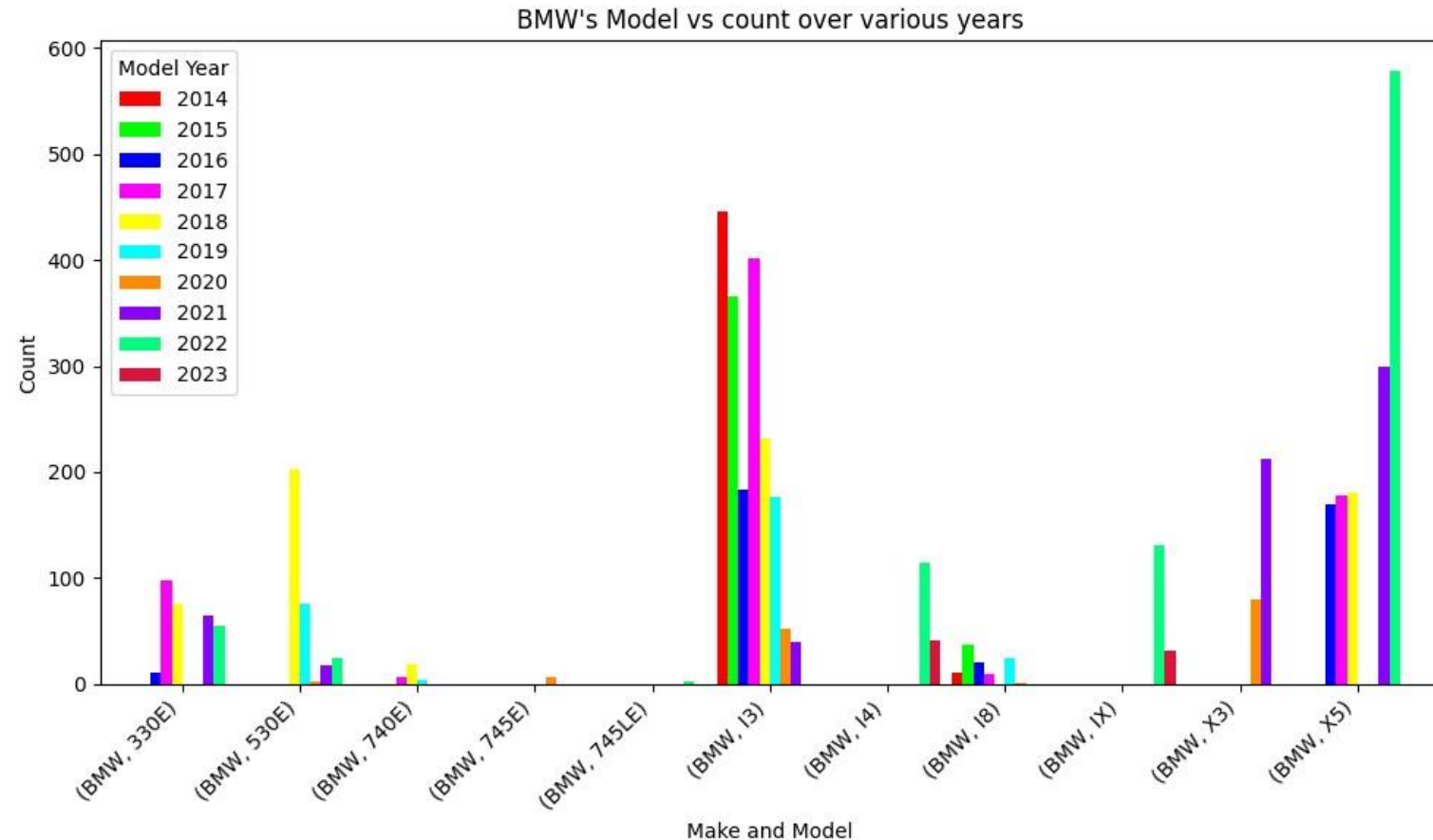
- X5 was released in the year 2016. Its sales has been continuously increasing from the year of its release. It was also noted that the sales in year 2019 and 2023 is completely missing in the dataset and needs to be attended.

Visualizing the Geographical region of the highest purchase

- This was made using choropleth map available in plotly express.

- To get the geographical coordinates I downloaded geojson data from the github repository.

- This helps the Sales Executive identify the key market locations to promote Evs.

Count of purchase in Corresponding Zip Code for Washington State

Zip Code=98052
count=2916

count
2500
2000
1500
1000
500
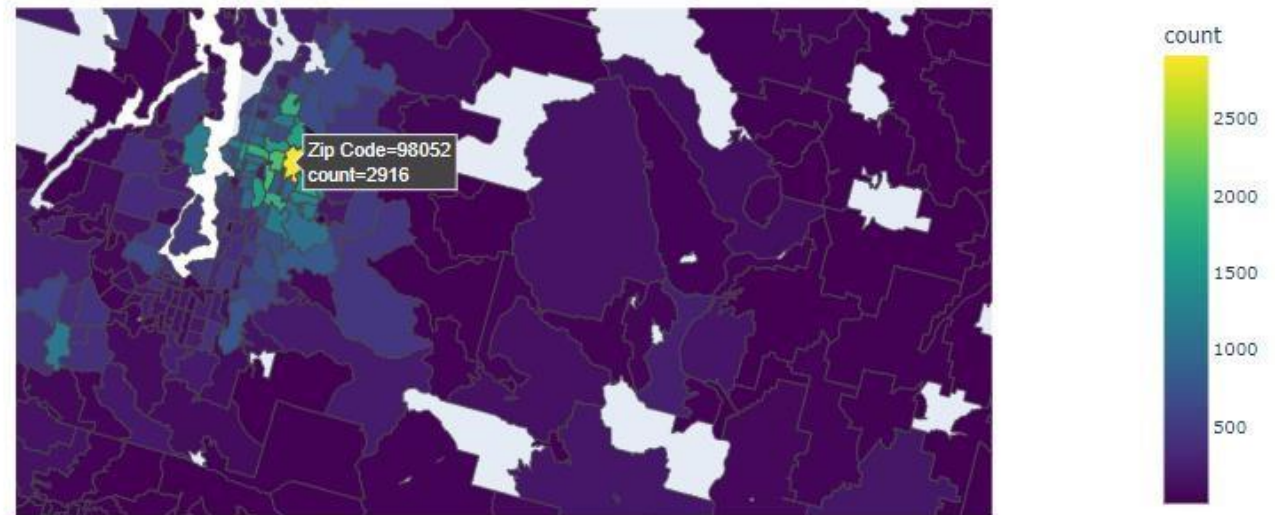
Racing Bar Plot of Car Makes Over Years

- bar-chart-race is a library in python that allows one to create video animation of bar graph for varying column.

- This helps in evaluating the top leaders for every corresponding varying column.

- Here I plotted count plot of Car Make with respect to each year.

Conclusion

The Analysis is extremely useful for the Sales Exectives in making key decisions such as:

Racing Bar Plot of Car Makes Over Years
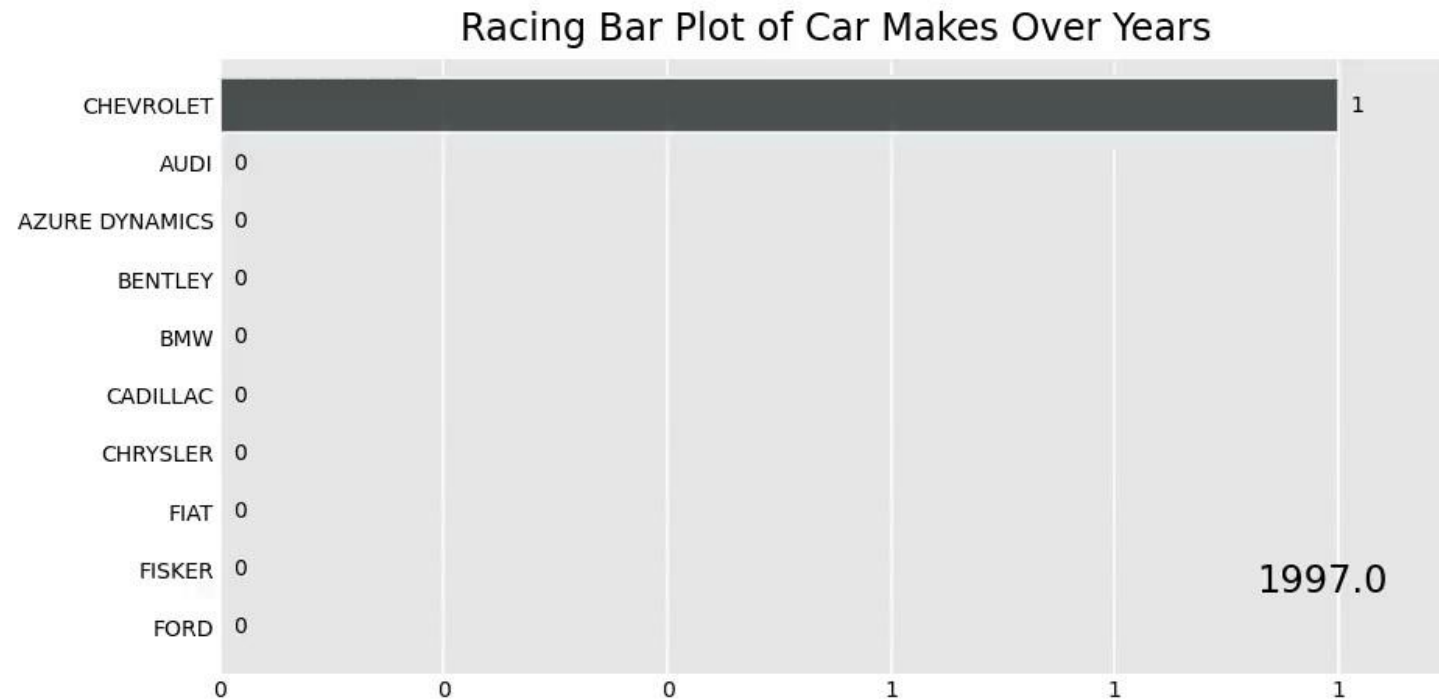
- Electric Range across various different fields
- Addressing the reasoning of the existence of the Missing values.
- Regional Sales Trends in USA
- Customer Preferences over Make and Model.
- Growth of EV across various years.

These insights enables sales executives to strategize, optimize, and enhance their approach to driving electric vehicle sales.

THANK YOU