

# Problem Statement / Objective

Market Size Analysis is the process of estimating the potential sales for a product or service within a particular market segment. In the context of electric vehicles (EVs), it involves assessing the total volume of EV registrations to understand the growth of the market, forecast future trends, and help stakeholders make informed decisions regarding production, infrastructure development, and policy-making

The provided dataset contains the following columns, each representing different aspects of the electric vehicle (EV) population in the United States:

• VIN (1-10): Partial Vehicle Identification Number.

• County: The county in which the vehicle is registered.

• City: The city in which the vehicle is registered.

• State: The state in which the vehicle is registered. It appears that this dataset may be focused on Washington (WA) state.

• Postal Code: The postal code where the vehicle is registered.

• Model Year: The year of the vehicle model

• Make: The manufacturer of the vehicle.

• Model: The model of the vehicle.

• Electric Vehicle Type: The type of electric vehicle, e.g., Battery Electric Vehicle (BEV).

• Clean Alternative Fuel Vehicle (CAFV) Eligibility: Eligibility status for clean alternative fuel vehicle programs.

• Electric Range: The maximum range of the vehicle on a single charge (in miles).

• Base MSRP: The Manufacturer’s Suggested Retail Price.

• Legislative District: The legislative district where the vehicle is registered.

• DOL Vehicle ID: Department of Licensing Vehicle Identification.

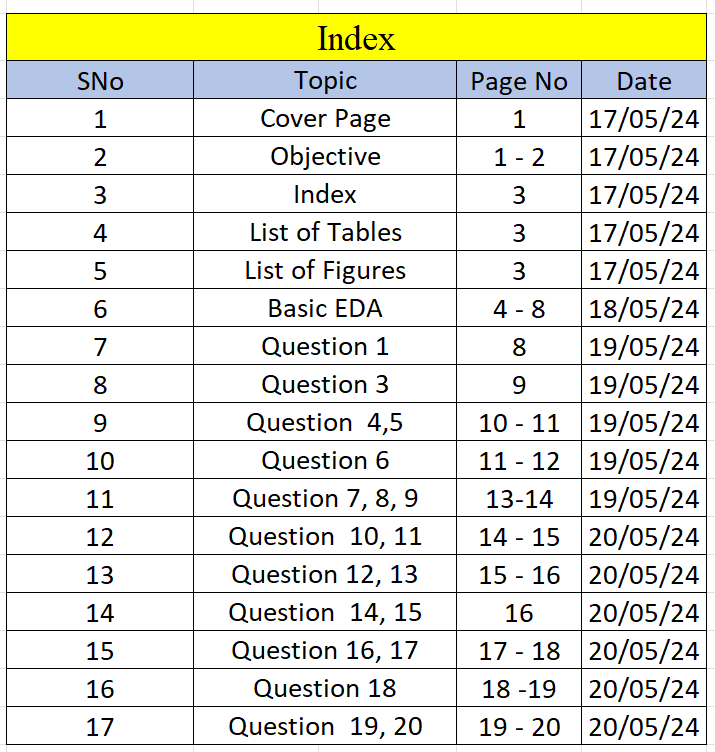
• Vehicle Location: Geographic coordinates of the vehicle location.

• Electric Utility: The electric utility service provider for the vehicle’s location.

• 2020 Census Tract: The census tract for the vehicle’s location.

The primary objective of this analysis is to leverage historical EV registration data to understand the current market penetration of EVs, predict future market growth, and identify key trends and factors driving market expansion.

# Index



# List of Tables

* Top 5 rows of dataset
* Bottom 5 rows of dataset
* Most common Model in a Year
* Summary Table for vehicles of different Make and Electric Range
* City vs Base MSRP
* Ranked list of most registered electric in Legislative District
* Most common VIN prefixes and suffix
* Frequency Table of top 10 Models
* Legislative District vs Base MSRP

# List of Figures

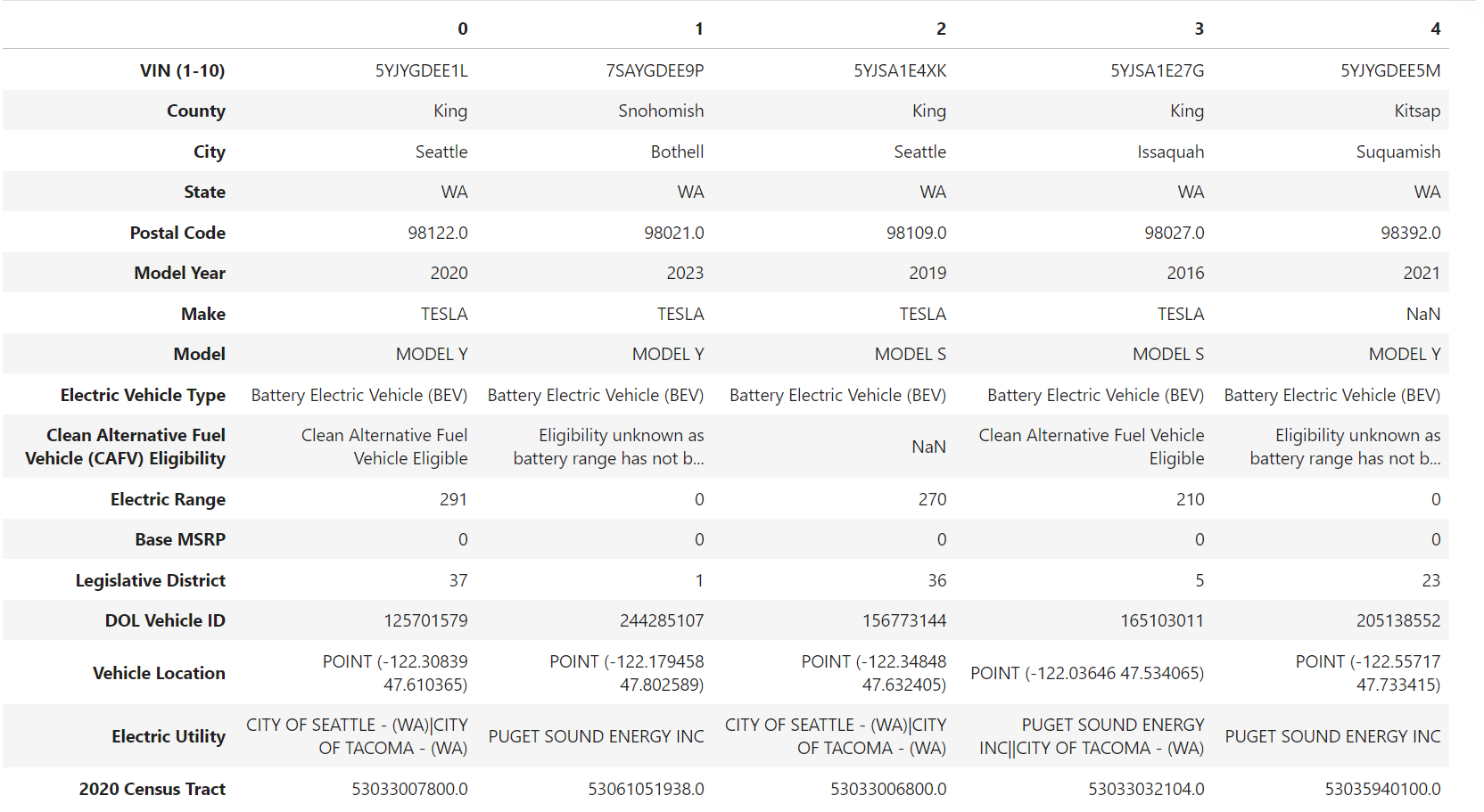
* Statistical data types
* Numerical datat types

# Data Dictionary

* 6-Electric\_Vehicle\_Population\_Data New.csv

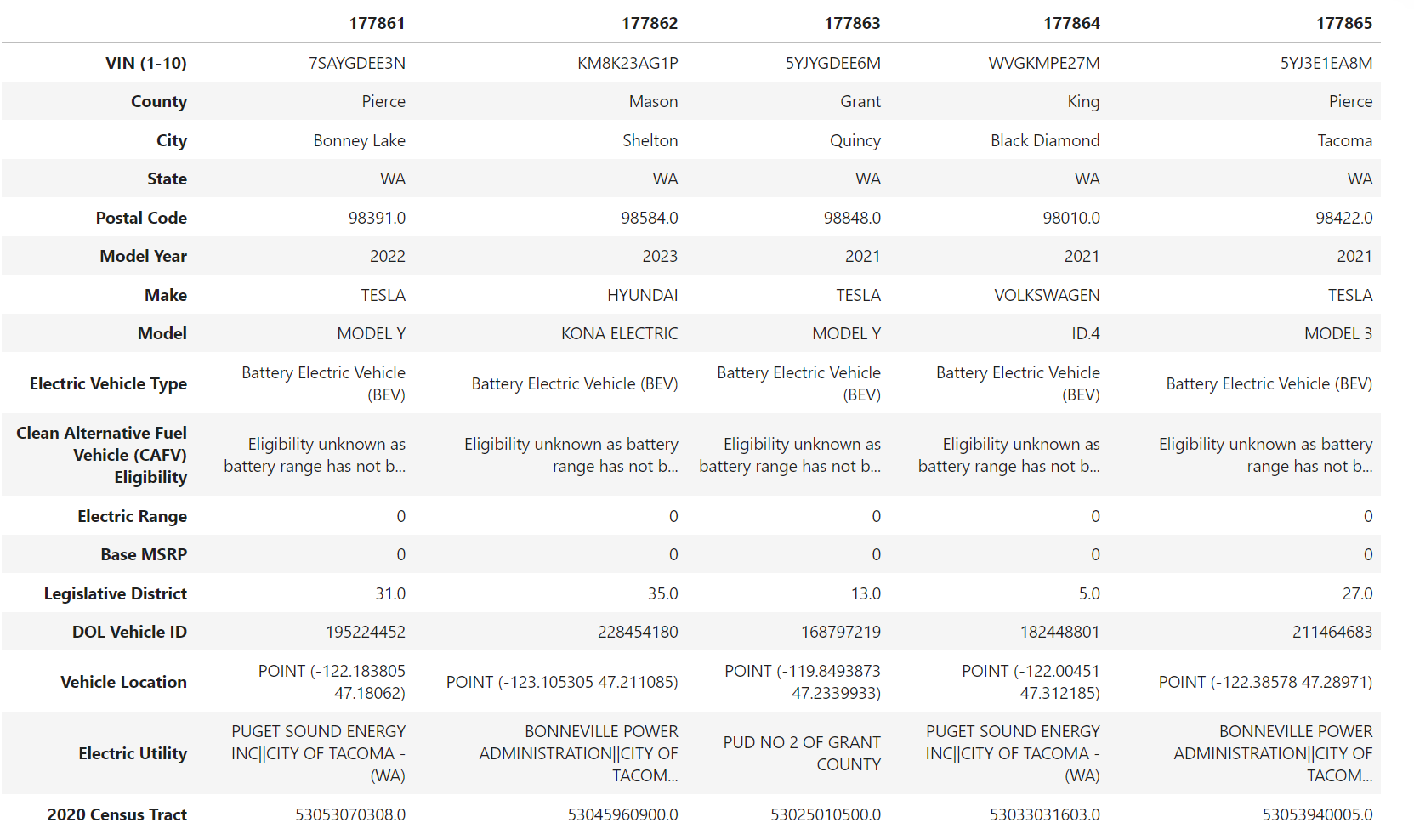
### The top 5 rows

## Observation



### The last 5 rows

## Observation



### Shape of Dataset

## Observation

* The dataset is having 17766 rows and 17 columns

### Datatypes of each feature

### Observation

* Postal Code is numerical type but should be Object
* Electric Range is Object type but should be Numerical
* DOL Vehicle ID is numerical type but should be Object
* 2020 Census Tract is numerical type but should be Object
* All these columns need to be checked

### Statistical Summary

## Observation

* Data set contains numerical datatypes of columns Postal code, DOL Vehicle ID and 2020 Census Tract which should be statistical types

### Null values

## Observation

Columns –

* County [5],
* City [5],
* Postal Code [5],
* Make [7],
* Model [4],
* Electric Vehicle Type [6],
* Clean Alternative Fuel Vehicle (CAFV) [2],
* Electric Range [3],
* Legislative District [389],
* Vehicle Location [9],
* Electric Utility [5],
* 2020 Census Tract [5]

contains null values as shown respectively

### Duplicate values

## Observation

* Data set contains no duplicate values

### Anomalies or wrong entry

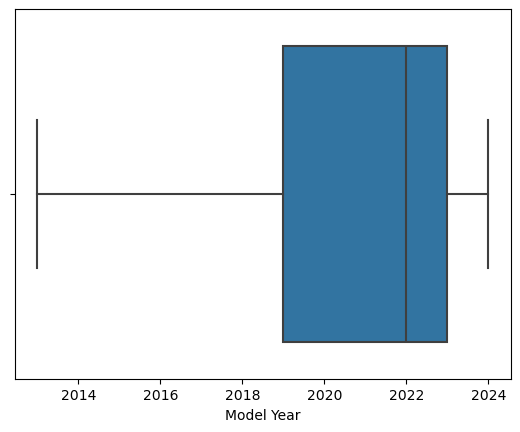
## Observation

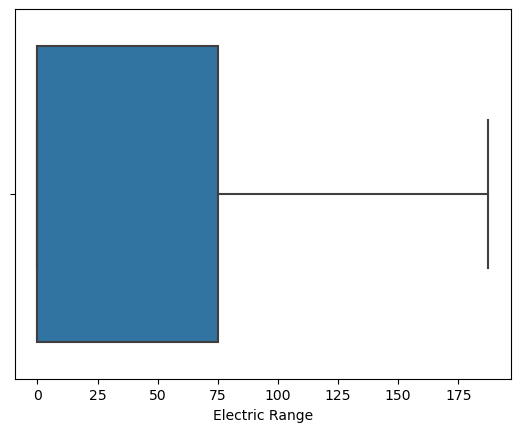
* Electric Range contains '?'
* Legislative District contains '?'
* For row 6, Country and City are both the same name, Yakima. Further checking is to be done.

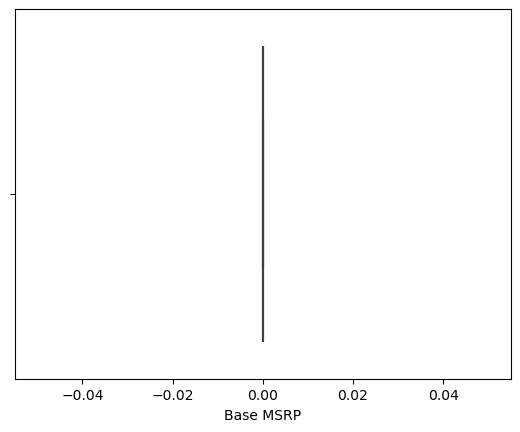
## Check the outliers and their authenticity

### Observation

* After removing anomalies and converting datatypes of each column with wrong entries, the outliers representation through box plot are







## Cleaning Data

## Observation

* Updated '?' to null value in columns- Electric Range and Legislative District
* Null values of columns Electric Range and Base MSRP are replaced by median
* Columns –
* [County, City, Postal Code, Make, Model, Electric Vehicle Type, Clean Alternative Fuel Vehicle (CAFV) Eligibility, Electric Range, Legislative District, Vehicle Location, Electric Utility, 2020 Census Tract]
* contains null values, all has been replaced by their respective Mode as they are of statistical datatype
* Outliers of every column has been removed

### 1. Descriptive Statistics:

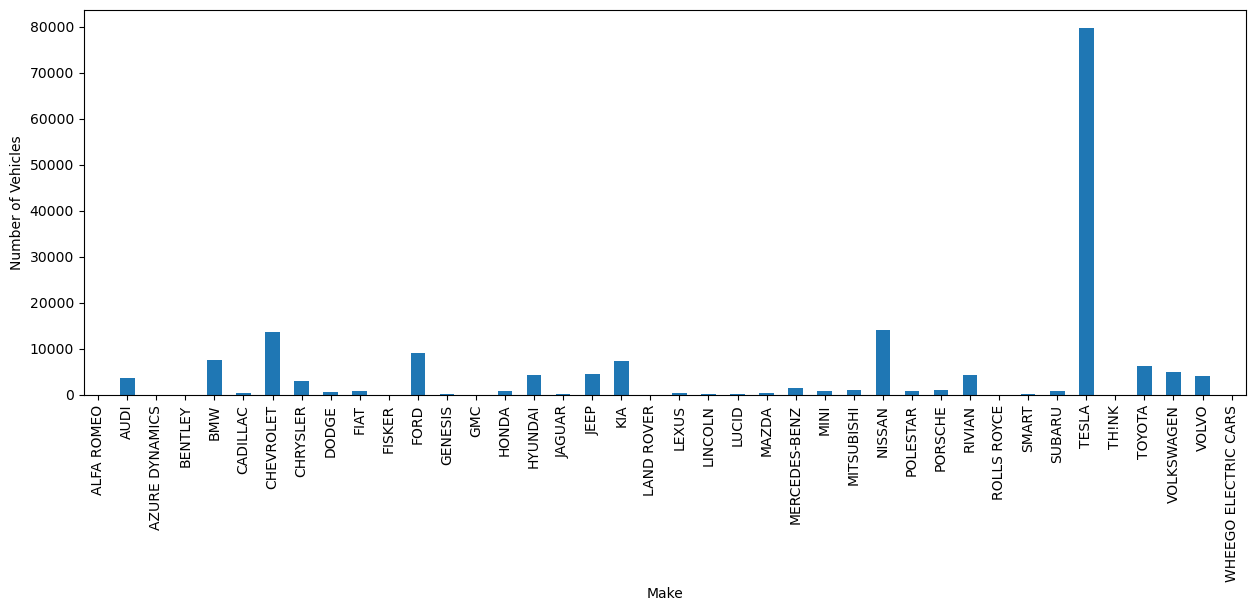
• What are the mean, median, and standard deviation of the base MSRP for the vehicles in the dataset?

## Observation

* The Mean of Base MSRP is 0.00
* The Median of Base MSRP is 0.0
* The Standard Deviation of Base MSRP is 0.00

### 2. Data Distribution:

• What is the distribution of vehicle makes in the dataset? Represent it using a bar chart.



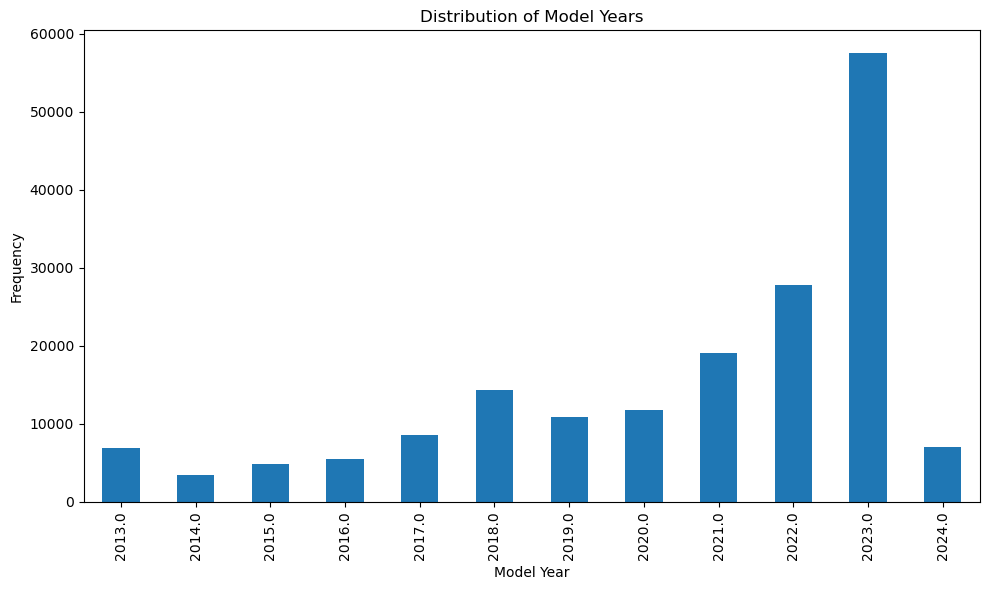
### 3. Model Year Analysis:

• What are the most common model years in the dataset? Provide

1. A frequency table
2. histogram.

## Observation

* Model Year  
  2013.0 6862  
  2014.0 3509  
  2015.0 4844  
  2016.0 5483  
  2017.0 8562  
  2018.0 14323  
  2019.0 10940  
  2020.0 11768  
  2021.0 19132  
  2022.0 27776  
  2023.0 57587  
  2024.0 7080



## Observation

* The year between 2023 had seen the most common modles in a year.

### 4. Electric Vehicle Type:

• What is the proportion of Battery Electric Vehicles (BEV) versus other types of electric vehicles?

## Observation

* The raito of Battery Electric Vehicles (BEV) to other types of electric vehicles is 13921 : 177866.
* Or Battery Electric Vehicles (BEV) is 78.26% more compared to other electric vehicles.

### 5. Electric Range Analysis:

• What is the average electric range for vehicles of different makes? Provide a summary tabe.

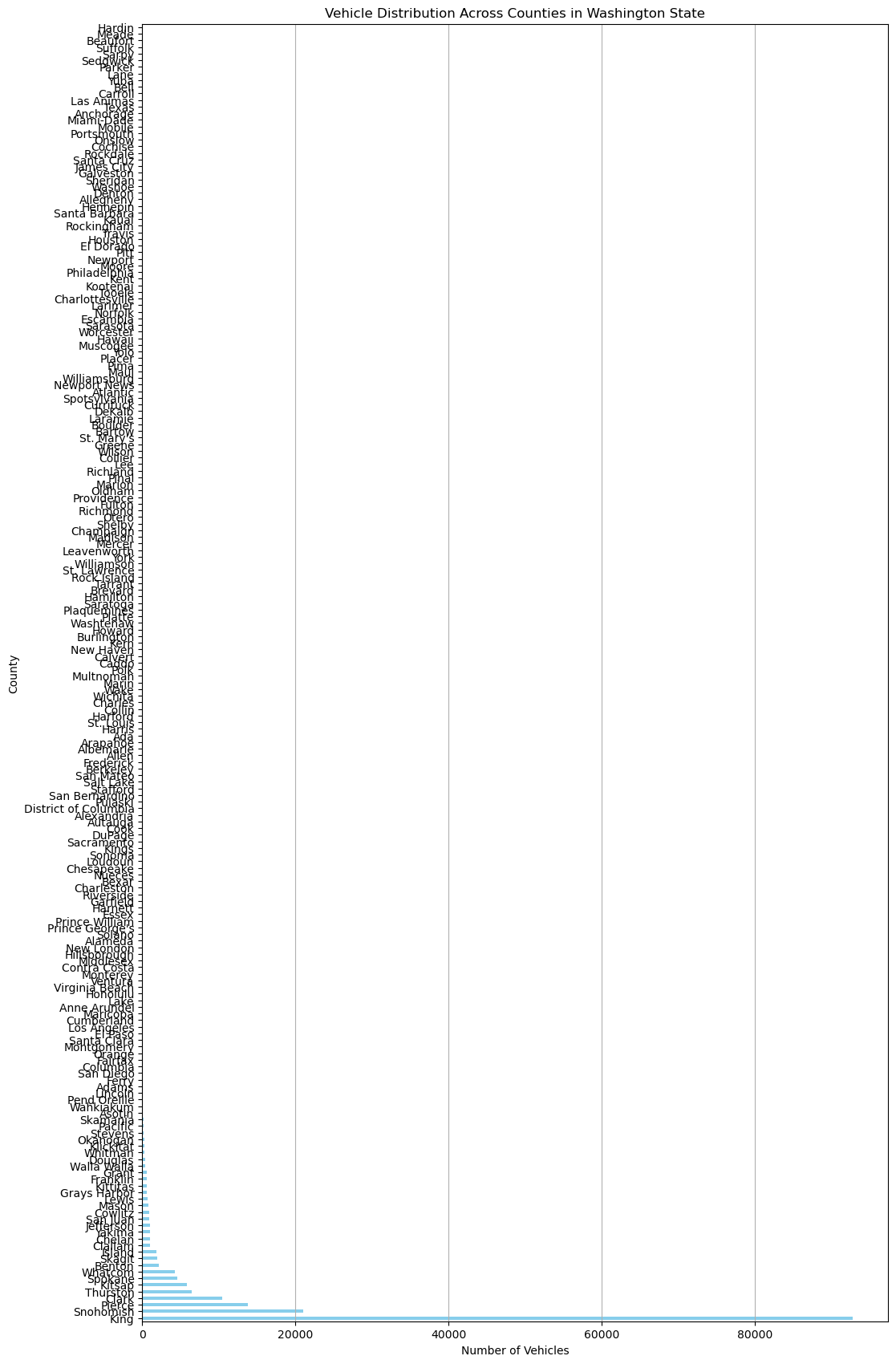
## Observation

* Summary Table for vehicles of different Make and Electric Range:

Make Electric Range  
0 ALFA ROMEO 33.000000  
1 AUDI 46.454745  
2 AZURE DYNAMICS 56.000000  
3 BENTLEY 19.666667  
4 BMW 34.711427  
5 CADILLAC 8.798429  
6 CHEVROLET 76.570332  
7 CHRYSLER 32.212162  
8 DODGE 32.000000  
9 FIAT 85.645408  
10 FISKER 8.755102  
11 FORD 10.812914  
12 GENESIS 0.000000  
13 GMC 0.000000  
14 HONDA 46.600240  
15 HYUNDAI 15.858715  
16 JAGUAR 163.254310  
17 JEEP 22.365402  
18 KIA 38.850175  
19 LAND ROVER 25.000000  
20 LEXUS 18.800000  
21 LINCOLN 23.543071  
22 LUCID 0.000000  
23 MAZDA 25.781513  
24 MERCEDES-BENZ 9.346130  
25 MINI 18.016704  
26 MITSUBISHI 30.646138  
27 NISSAN 79.839501  
28 POLESTAR 30.399660  
29 PORSCHE 42.693152  
30 RIVIAN 0.000000  
31 ROLLS ROYCE 0.000000  
32 SMART 62.325926  
33 SUBARU 1.350181  
34 TESLA 60.214775  
35 TH!NK 100.000000  
36 TOYOTA 28.095102  
37 VOLKSWAGEN 22.927058  
38 VOLVO 16.135737  
39 WHEEGO ELECTRIC CARS 100.000000

### 6. County Distribution:

• How are vehicles distributed across different counties in Washington state? Represent the distribution using a barh chart.



## Observation

* King has the greatest number of vehicles in the Washington state around 53%

### 7. Price Analysis:

• Compare the average base MSRP of vehicles eligible for the Clean Alternative Fuel Vehicle (CAFV) program versus those that are not

## Observation

* Average Base MSRP for CAFV Eligible: $0.00, Not Eligible: $0.00

### 8. Geographical Analysis:

• How does the base MSRP vary across different cities in Washington state?

## Observation

City Base MSRP  
0 Aberdeen 0.0  
1 Aberdeen Proving Ground 0.0  
2 Acme 0.0  
3 Adairsville 0.0  
4 Addy 0.0  
5 Adna 0.0  
6 Airway Heights 0.0  
7 Alameda 0.0  
8 Alderdale 0.0  
9 Alderwood Manor 0.0

### 9. Legislative Districts:

• Which legislative districts have the highest number of registered electric vehicles? Provide a ranked list

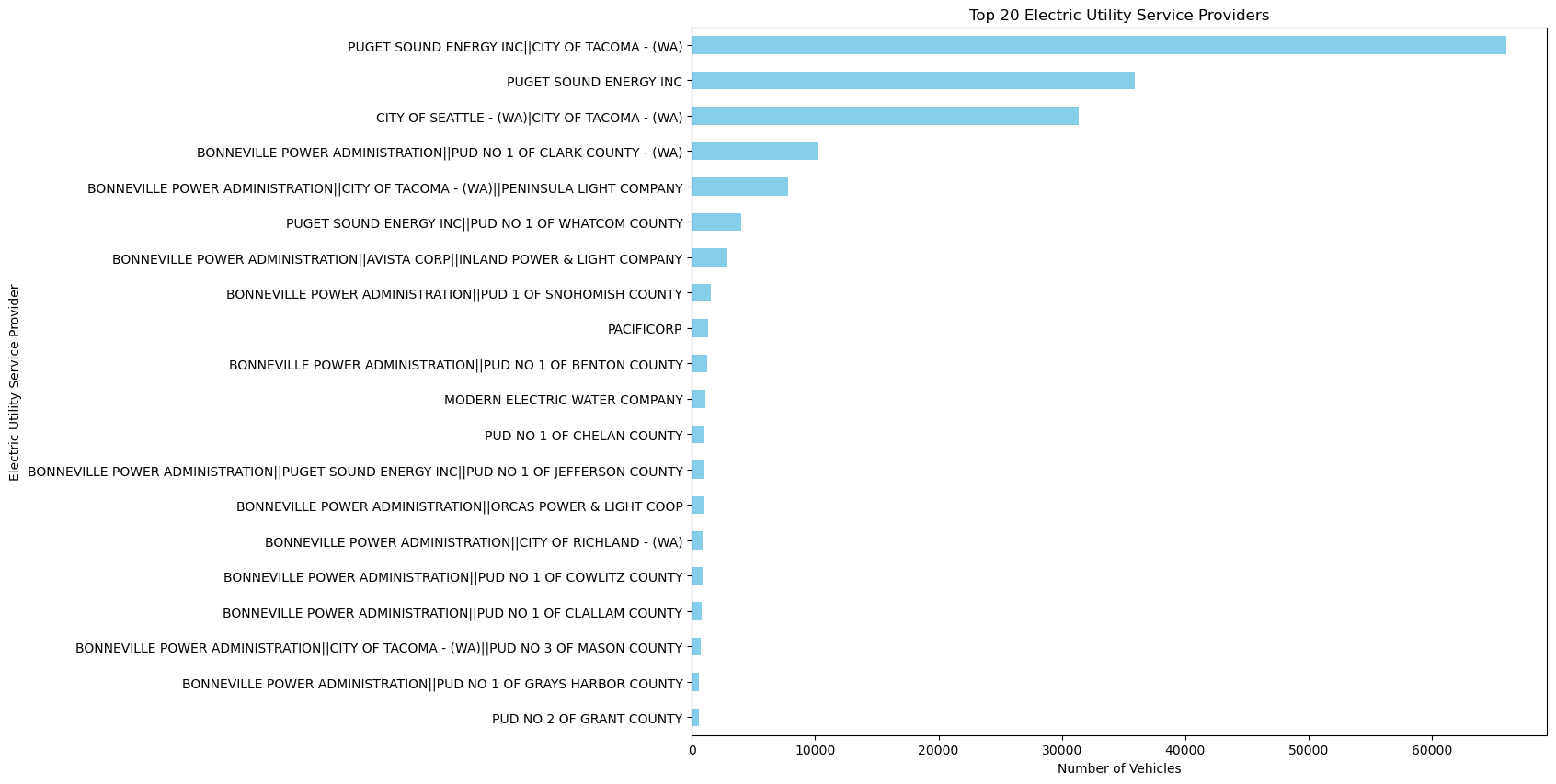
## Observation

* 41 legislative districts have the highest number of registered electric vehicles
* Ranked List:
* Legislative District  
  41.0 8831  
  45.0 7425  
  5.0 6810  
  48.0 6631  
  1.0 6265  
  Name: count, dtype: int64

Legislative District  
9 5  
16 4  
3 2  
40 2  
6 1  
Name: count, dtype: int64

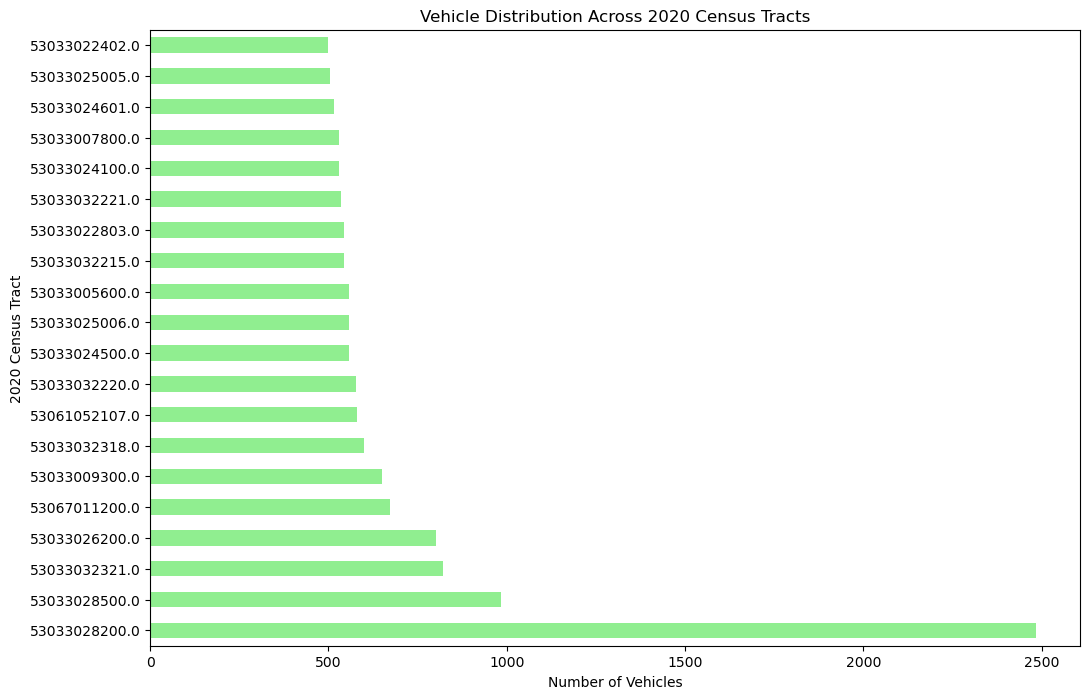
### 10. Electric Utility Providers:

• What is the distribution of electric utility service providers for the vehicles in the dataset?



### 11. Census Tract Analysis:

• How are vehicles distributed across different 2020 Census Tracts? Provide insights based on vehicle counts per tract.



### 12. Electric Range Correlation:

• Is there a correlation between the electric range and the base MSRP of the vehicles? Provide the correlation coefficient and interpret the result.

## Observation

* Correlation Coefficient between Electric Range and Base MSRP: nan
* Correlation Coefficient (r):

1: Perfect positive correlation.

0.7 to 0.9: Strong positive correlation.

0.4 to 0.6: Moderate positive correlation.

0.1 to 0.3: Weak positive correlation.

0: No correlation.

-0.1 to -0.3: Weak negative correlation.

-0.4 to -0.6: Moderate negative correlation.

-0.7 to -0.9: Strong negative correlation.

-1: Perfect negative correlation

* There cannot be a Correlation between Electric Range and Base MSRP

### 13. VIN Analysis:

• Identify any patterns or commonalities in the VIN (1-10) for the vehicles. Are there any frequent prefixes or suffixes

* Most common VIN prefixes (first 10 characters):  
  VIN (1-10)  
  7SAYGDEE6P 1239  
  7SAYGDEE7P 1235  
  7SAYGDEE8P 1197  
  7SAYGDEEXP 1191  
  7SAYGDEE5P 1177  
  Name: count, dtype: int64
* Most common VIN suffixes (last 10 characters):

VIN (1-10)  
7SAYGDEE6P 1239  
7SAYGDEE7P 1235  
7SAYGDEE8P 1197  
7SAYGDEEXP 1191  
7SAYGDEE5P 1177  
Name: count, dtype: int64

### 14. Eligibility Status:

• What percentage of vehicles are eligible for the Clean Alternative Fuel Vehicle (CAFV) program

## Observation

* Percentage of vehicles eligible for CAFV: 37.29%

### 15. Model Popularity:

• Which vehicle models are the most popular in the dataset? Provide a frequency table of the top 10 models

## Observation

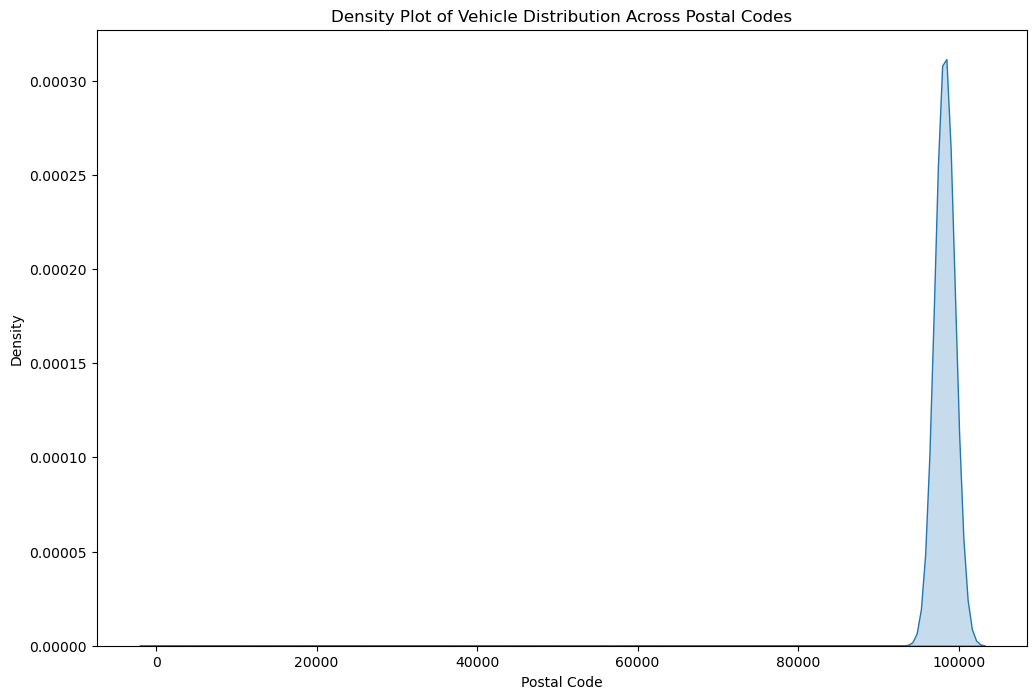
* Frequency Table of top 10 Models:

Model  
MODEL Y 35993  
MODEL 3 30091  
LEAF 13365  
MODEL S 7734  
BOLT EV 6821  
MODEL X 5796  
VOLT 4796  
ID.4 3937  
WRANGLER 3392  
MUSTANG MACH-E 3322  
Name: count, dtype: int64

* MODEL Y is most popular in the dataset

### 16. Postal Code Distribution:

• How are vehicles distributed across different postal codes? Provide a heatmap or density plot

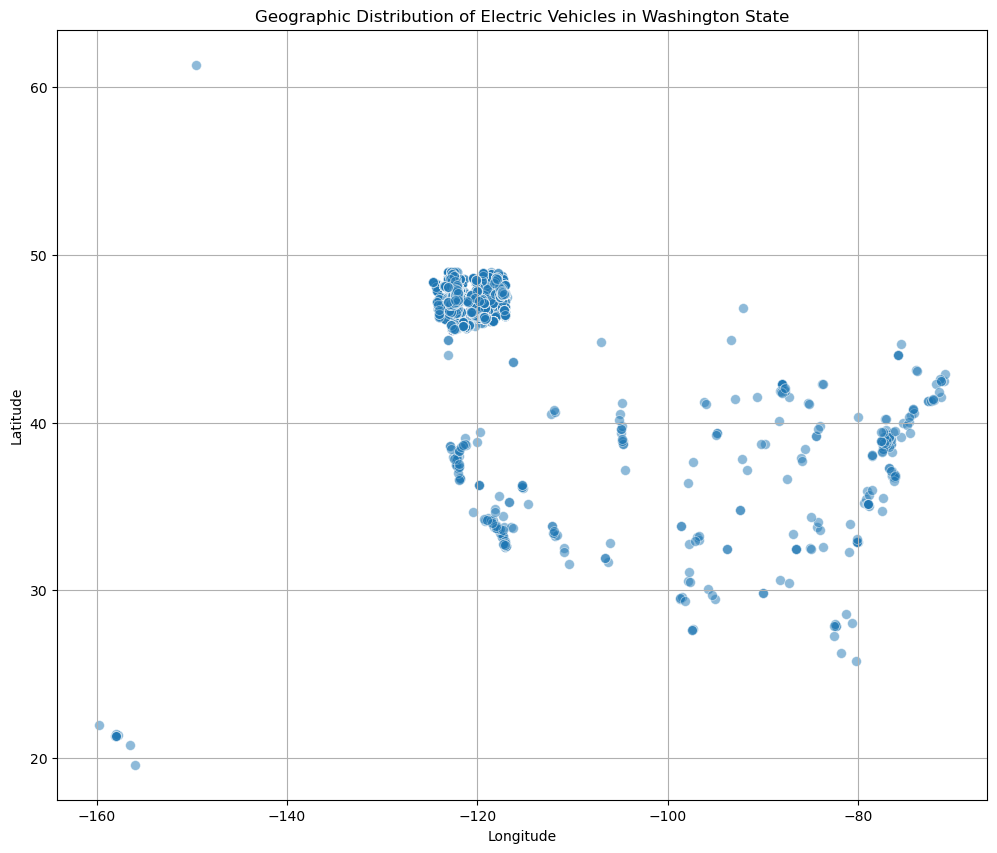


## Observation

* Density Plot of Vehicle Distribution Across Postal Codes shows postal code at 100000 is maximum

### 17. Vehicle Location Analysis:

• Analyze the geographic coordinates to determine any clusters of electric vehicles in certain areas of Washington state.

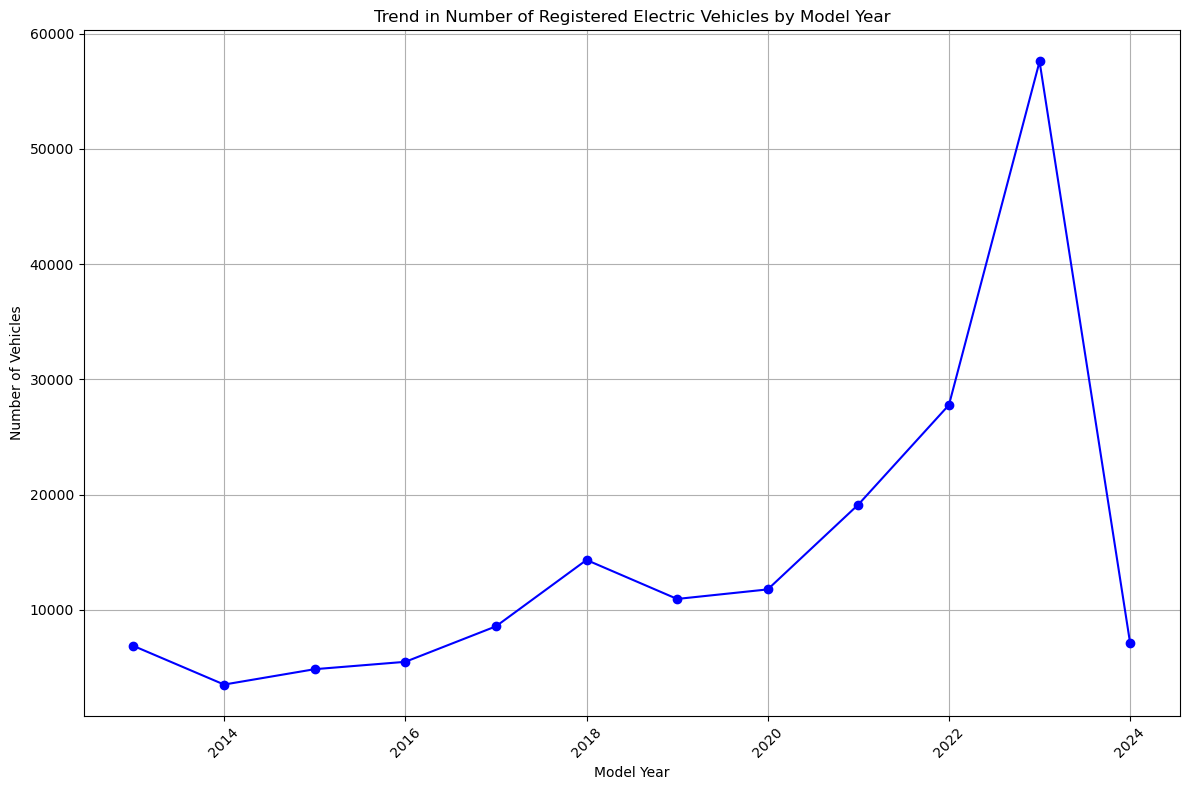


## Observation

* At longitude 49 and Longitude -120 the distribution is maximum

### 18. Model Year Trend:

• Analyze the trend in the number of registered electric vehicles by model year. Provide a line chart to show any increase or decrease over the years.

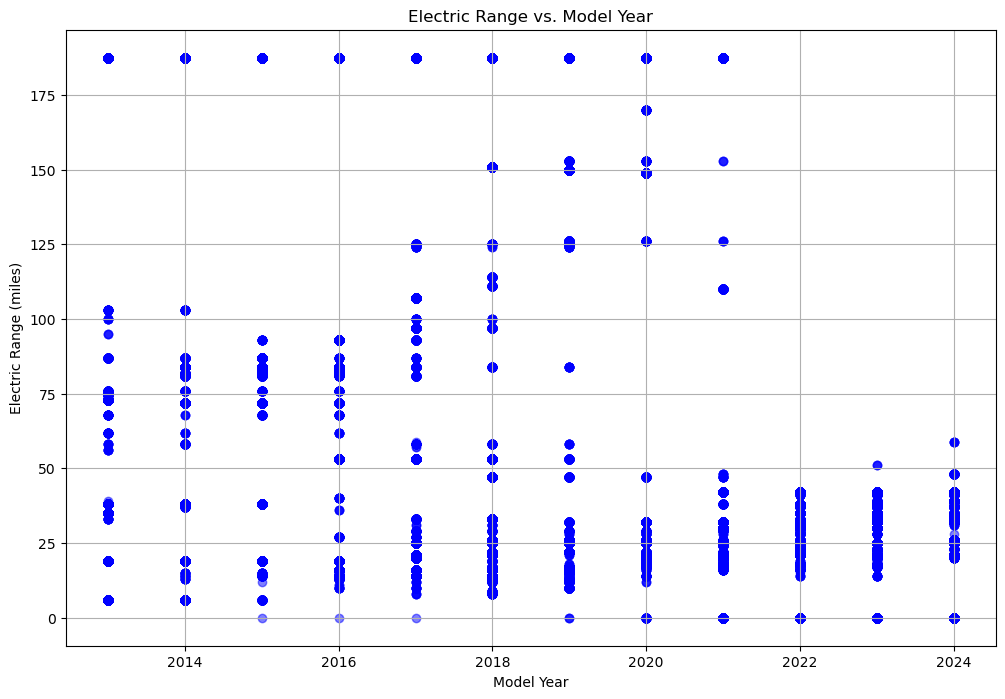


## Observation

* The was a steady growth in between year 2014 and 2020
* 2020 to 2023 saw a skyroketting growth in the number of vehicle registered
* After mid of 2023 there was a drastic fall in number of registrations in 2024

### 19. Range vs. Year:

• Is there a trend between the model year and the electric range of the vehicles? Provide a scatter plot and analyze the trend.



## Observation

* The correlation coefficient is -0.55
* This negative correlation indicates that newer model years generally have lower electric ranges compared to older model years.

### 20. Legislative District and MSRP:

• How does the average base MSRP vary across different legislative districts

## Observation

* Legislative District  
  1.0 0.0  
  24 0.0  
  23 0.0  
  22 0.0  
  21 0.0  
  Name: Base MSRP, dtype: float64
* The Base MSRP remains the same for every Legislative District
* There seems to be no connection between Legislative District and Base MSRP