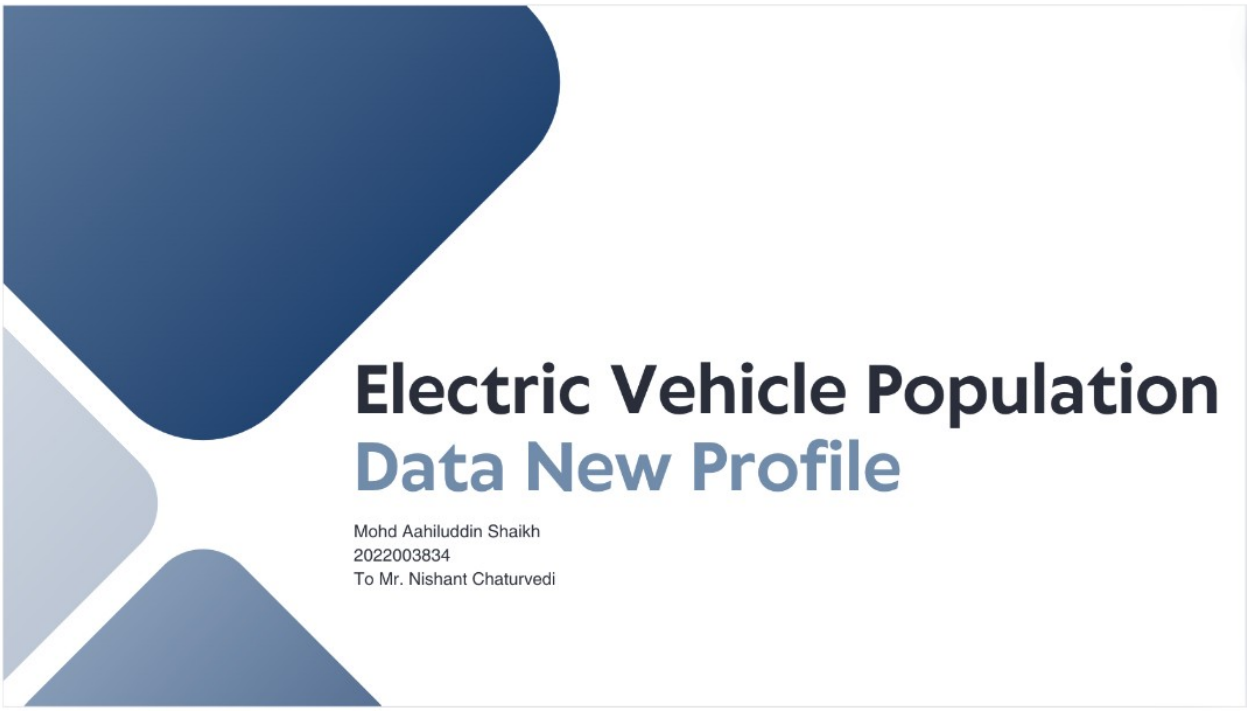


Data

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 registerd. It appears that this dataset may be focused on Washington (WA) stat* • Postal Code: The postal code where the vehicle is register* . • Model Year: The year of the vehicle mo* l • Make: The manufacturer of the veh* le. • Model: The model of the ve* cle. • 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 p* grams. • Electric Range: The maximum range of the vehicle on a single charge (i* miles). • Base MSRP: The Manufacturer's Suggested Ret* l Price. • Legislative District: The legislative district where the vehicle is * gistered. • DOL Vehicle ID: Department of Licensing Vehicle Ide* ification. • Vehicle Location: Geographic coordinates of the vehi* e location. • Electric Utility: The electric utility service provider for the vehic* 's location. • 2020 Census Tract: The census tract for the vehicle's location.

1- Display the top 5 rows.

[4]:

	VIN (1-10)	County	City	State	Postal Code	Model Year	Make	Model	Electric Vehicle Type	Clean Alternative Fuel Vehicle (CAFV) Eligibility	Electric Range	Base MSRP	Legislative District	DOL Vehicle ID	Vehicle Location	Electric Utility
0	5YJYGDEE1L	King	Seattle	WA	98122.0	2020	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	291	0	37	125701579	POINT (-122.30839 47.610365)	CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)
1	7SAYGDEE9P	Snohomish	Bothell	WA	98021.0	2023	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0	0	1	244285107	POINT (-122.179458 47.802589)	PUGET SOUND ENERGY INC
2	5YJSA1E4XK	King	Seattle	WA	98109.0	2019	TESLA	MODEL S	Battery Electric Vehicle (BEV)	NaN	270	0	36	156773144	POINT (-122.34848 47.632405)	CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)
3	5YJSA1E27G	King	Issaquah	WA	98027.0	2016	TESLA	MODEL S	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	210	0	5	165103011	POINT (-122.03646 47.534065)	PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)
4	5YJYGDEE5M	Kitsap	Suquamish	WA	98392.0	2021	NaN	MODEL Y	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0	0	23	205138552	POINT (-122.55717 47.733415)	PUGET SOUND ENERGY INC

2- Display the last 5 rows

[4]:

	VIN (1-10)	County	City	State	Postal Code	Model Year	Make	Model	Electric Vehicle Type	Clean Alternative Fuel Vehicle (CAFV) Eligibility	Electric Range	Base MSRP	Legislative District	DOL Vehicle ID	Vehicle Location
177861	7SAYGDEE3N	Pierce	Bonney Lake	WA	98391.0	2022	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0	0	31.0	195224452	POINT (-122.183805 47.18062)
177862	KM8K23AG1P	Mason	Shelton	WA	98584.0	2023	HYUNDAI	KONA ELECTRIC	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0	0	35.0	228454180	POINT (-123.105305 47.211085)
177863	5YJYGDEE6M	Grant	Quincy	WA	98848.0	2021	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0	0	13.0	168797219	POINT (-119.8493873 47.2339933)
177864	WVGKMPE27M	King	Black Diamond	WA	98010.0	2021	VOLKSWAGEN	ID.4	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0	0	5.0	182448801	POINT (-122.00451 47.312185)
177865	5YJ3E1EA8M	Pierce	Tacoma	WA	98422.0	2021	TESLA	MODEL 3	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has	0	0	27.0	211464683	POINT (-122.38578 47.28971)

3- Check the shape of dataset.

```
[5]: (177866, 17)
```

4- Check the datatypes of each feature.

[6]:

	Postal Code	Model Year	Base MSRP	DOL Vehicle ID	2020 Census Tract
count	177861.000000	177866.000000	177866.000000	1.778660e+05	1.778610e+05
mean	98172.453506	2020.515512	1073.109363	2.202313e+08	5.297672e+10
std	2442.450668	2.989384	8358.624956	7.584987e+07	1.578047e+09
min	1545.000000	1997.000000	0.000000	4.385000e+03	1.001020e+09
25%	98052.000000	2019.000000	0.000000	1.814743e+08	5.303301e+10
50%	98122.000000	2022.000000	0.000000	2.282522e+08	5.303303e+10
75%	98370.000000	2023.000000	0.000000	2.548445e+08	5.305307e+10
max	99577.000000	2024.000000	845000.000000	4.792548e+08	5.603300e+10

5- Check the Statistical summary

[7]:

	VIN (1-10)	County	City	State	Postal Code	Model Year	Make	Model	Electric Vehicle Type	Clean Alternative Fuel Vehicle (CAFV) Eligibility	Electric Range	Base MSRP	Legislative District	DOL Vehicle ID	Vehicle Location	Electric Utility	2020 Census Tract
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	True	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	True	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
177861	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
177862	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
177863	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
177864	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
177865	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False

177866 rows × 17 columns

6- Check the null values

```
[8]: VIN (1-10)                                          0
     County                                              5
     City                                                5
     State                                               0
     Postal Code                                         5
     Model Year                                          0
     Make                                                7
     Model                                               4
     Electric Vehicle Type                               6
     Clean Alternative Fuel Vehicle (CAFV) Eligibility   2
     Electric Range                                      3
     Base MSRP                                           0
     Legislative District                              389
     DOL Vehicle ID                                      0
     Vehicle Location                                    9
     Electric Utility                                    5
     2020 Census Tract                                   5
     dtype: int64
```

7- Check the duplicate values

```
[9]: 445
```

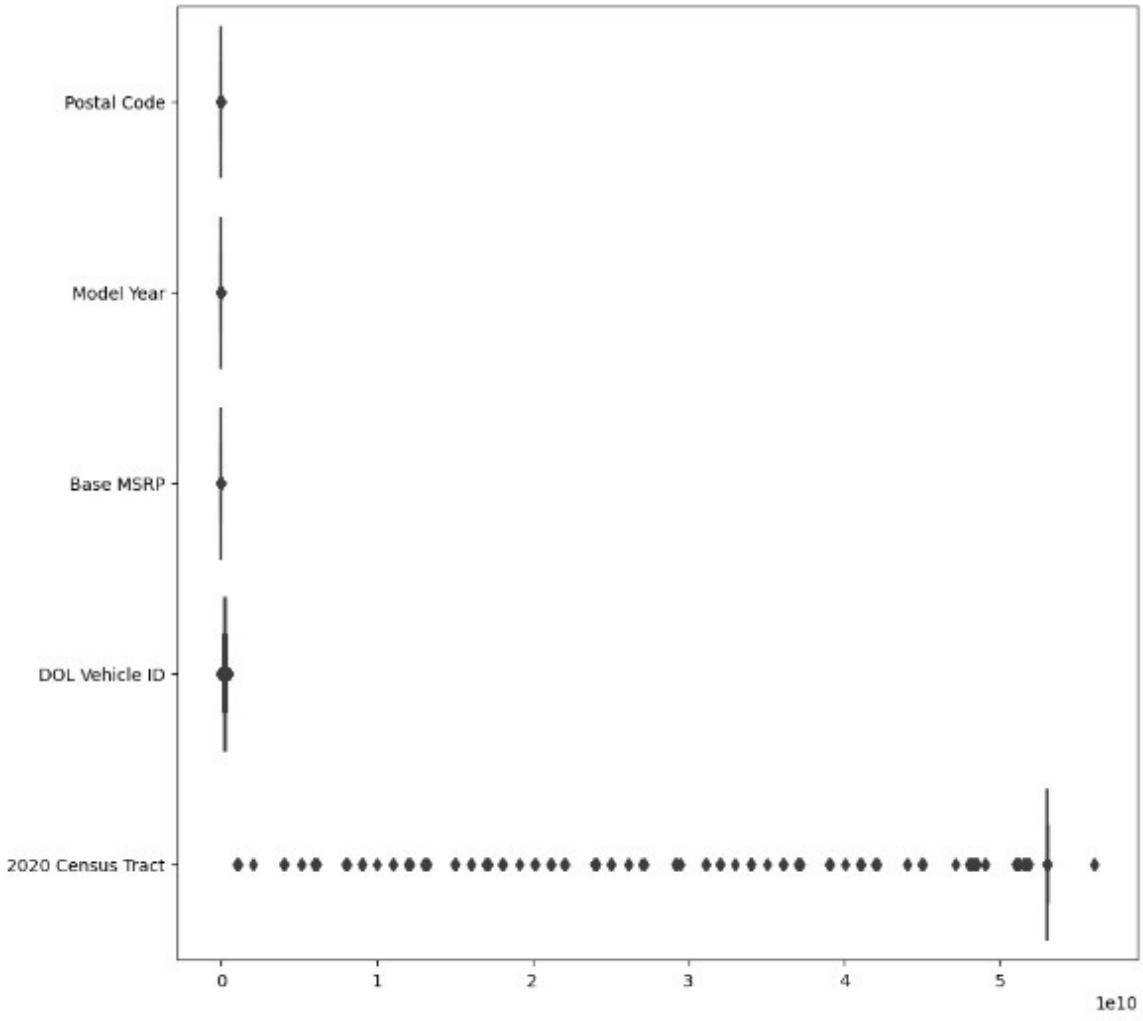
8- Check the anomalies or wrong entries.

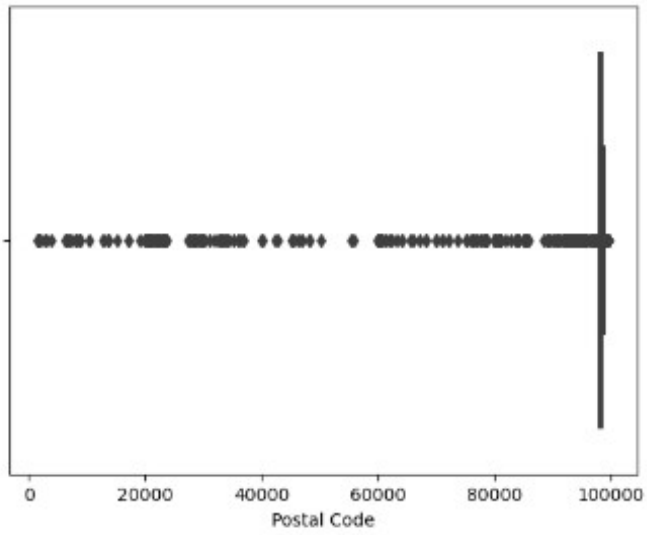
```
[10]: 0         False
      1         False
      2         False
      3         False
      4         False
                ...
      177861    False
      177862    False
      177863    False
      177864    False
      177865    False
      Length: 177866, dtype: bool
```

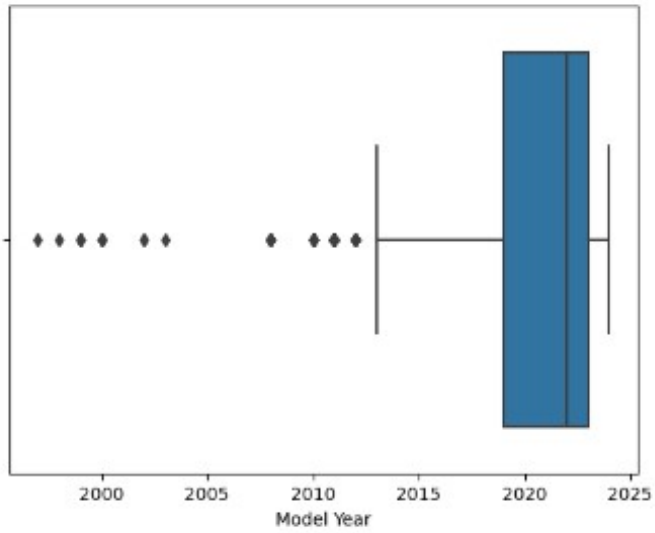
9- Check the outliers and their authenticity.

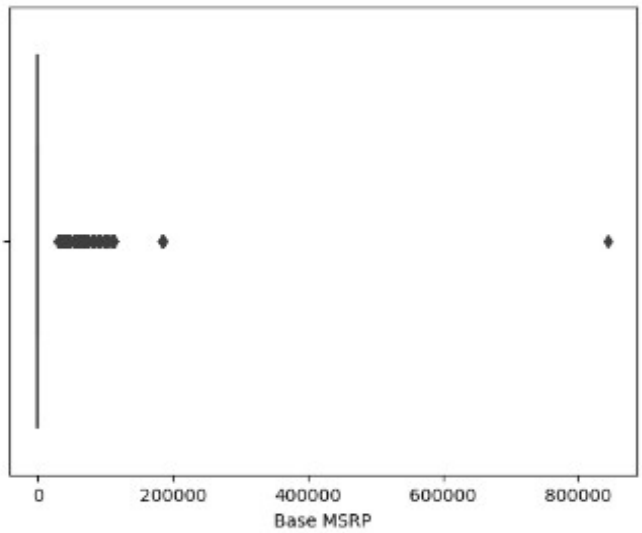
0

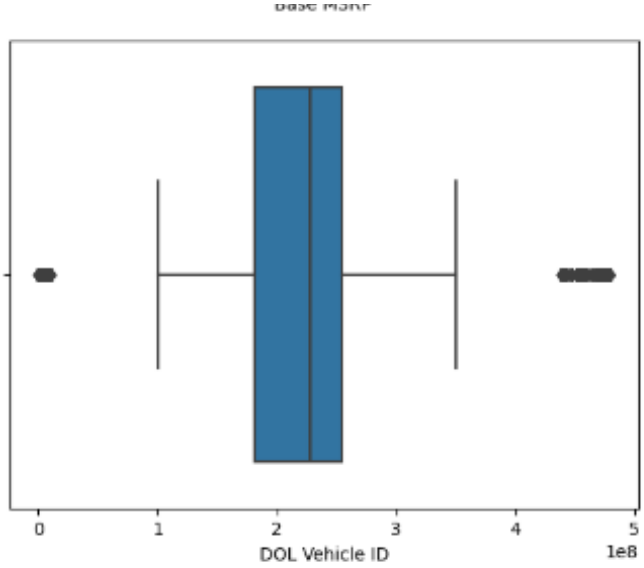
10- Do the necessary data cleaning steps like dropping duplicates, unnecessary columns, null value imputation, outliers treatment etc.

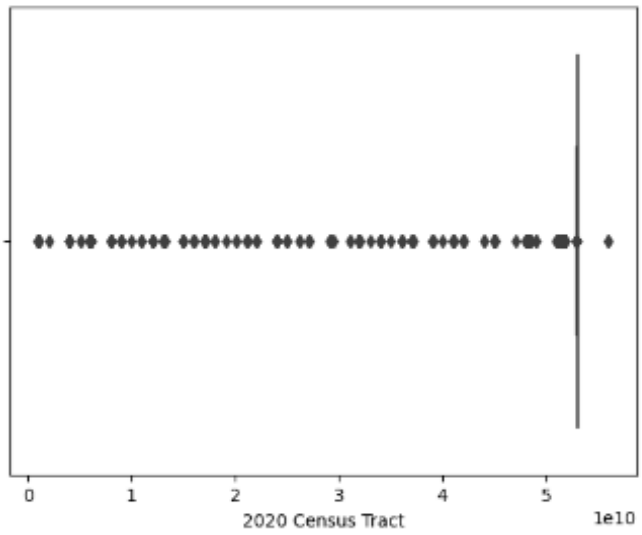

Postal Code
Model Year
Base MSRP
DOL Vehicle ID
2020 Census Tract
0
1
2
3
4
5
1e10


0
20000
40000
60000
80000
100000
Postal Code


2000
2005
2010
2015
2020
2025
Model Year


0
200000
400000
600000
800000
Base MSRP


Base MSRP
0
1
2
3
4
5
DOL Vehicle ID
1e8


0
1
2
3
4
5
2020 Census Tract
1e10