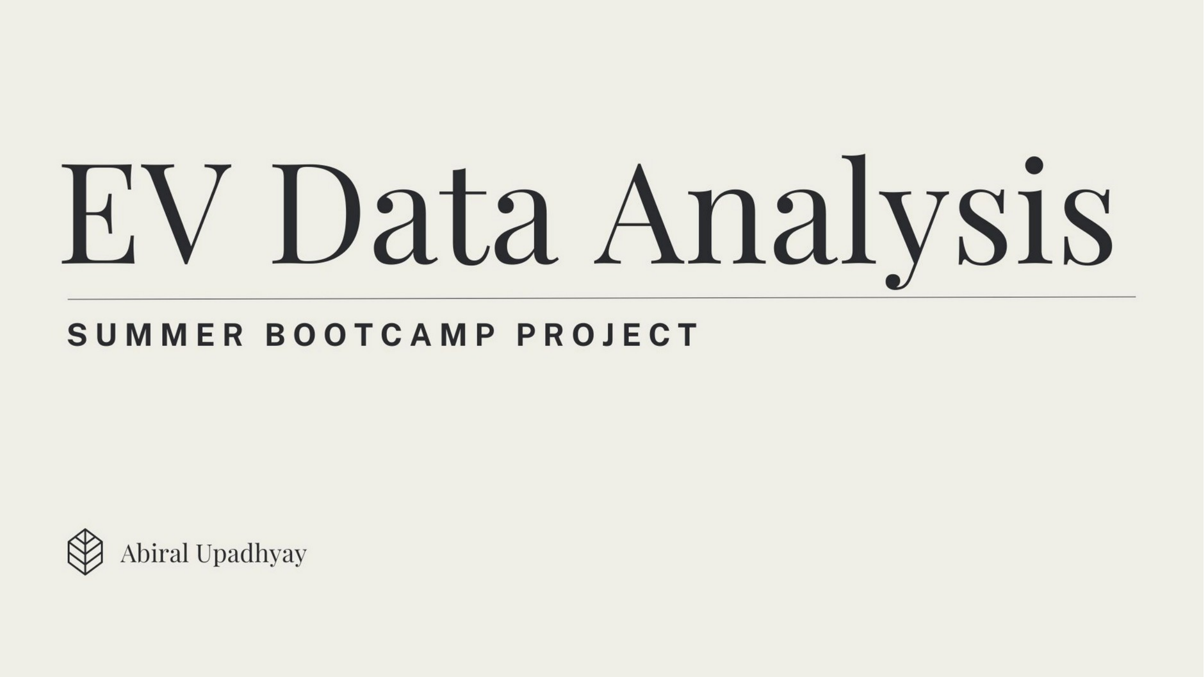


EV Data Analysis

SUMMER BOOTCAMP PROJECT

Abiral Upadhyay

#List of Tables

Importing Neccesarry Libraries

Libraries used:

- Numpy
- Pandas
- Matplotlib
- Seaborn

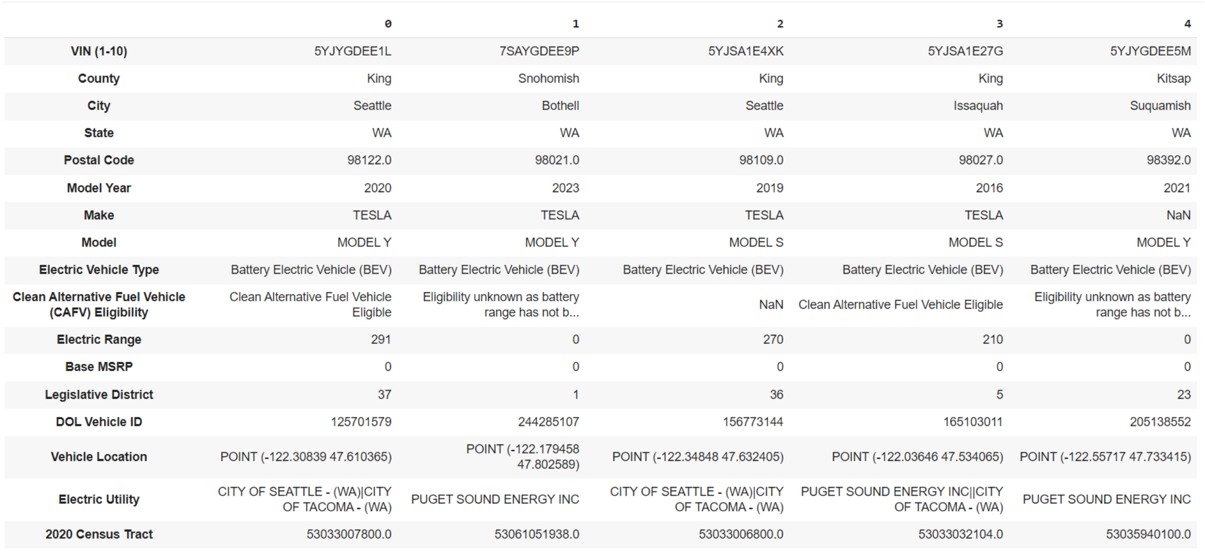
Importing the dataset

Importing the Electric Vehicle Population Dataset

Basic Exploration

1.) Head :-

First 5 entries of the dataset

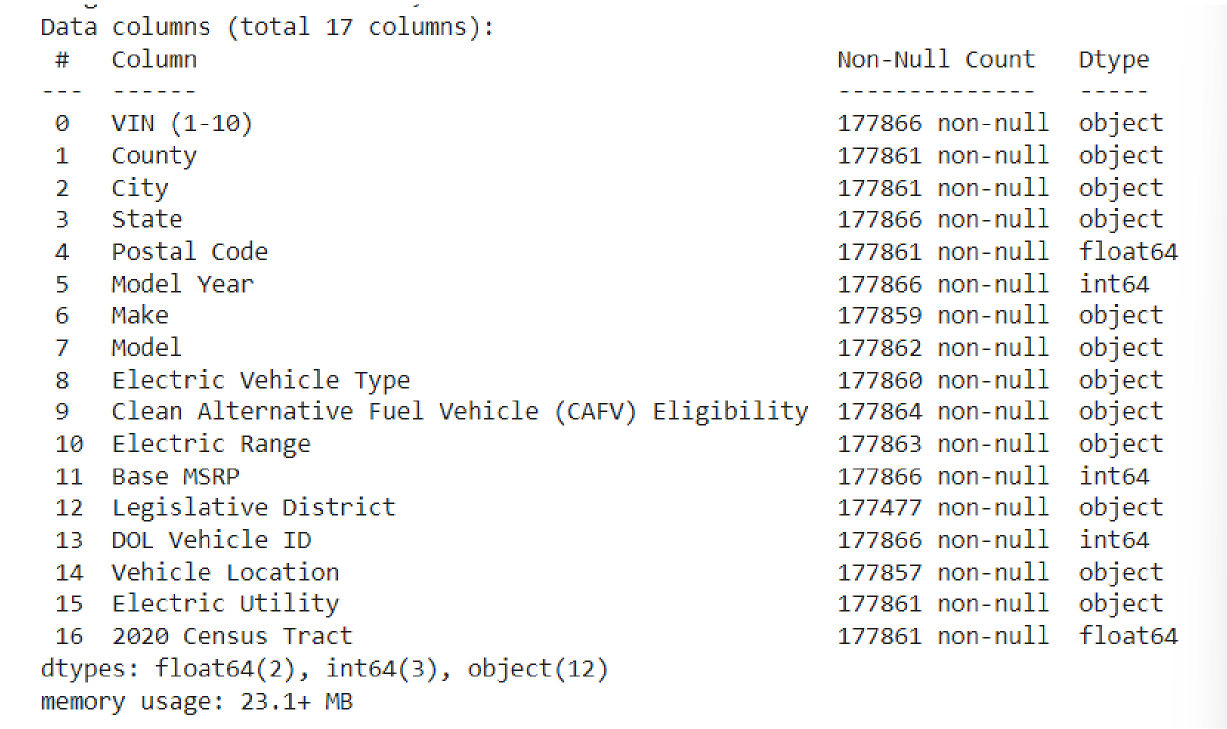


	0	1	2	3	4
VIN (1-10)	5YJYGDEE1L	7SAYGDEE9P	5YJSA1E4XK	5YJSA1E27G	5YJYGDEE5M
County	King	Snohomish	King	King	Kitsap
City	Seattle	Bothell	Seattle	Issaquah	Suquamish
State	WA	WA	WA	WA	WA
Postal Code	98122.0	98021.0	98109.0	98027.0	98392.0
Model Year	2020	2023	2019	2016	2021
Make	TESLA	TESLA	TESLA	TESLA	NaN
Model	MODEL Y	MODEL Y	MODEL S	MODEL S	MODEL Y
Electric Vehicle Type	Battery Electric Vehicle (BEV)	Battery Electric Vehicle (BEV)	Battery Electric Vehicle (BEV)	Battery Electric Vehicle (BEV)	Battery Electric Vehicle (BEV)
Clean Alternative Fuel Vehicle (CAFV) Eligibility	Clean Alternative Fuel Vehicle Eligible	Eligibility unknown as battery range has not b...	NaN	Clean Alternative Fuel Vehicle Eligible	Eligibility unknown as battery range has not b...
Electric Range	291	0	270	210	0
Base MSRP	0	0	0	0	0
Legislative District	37	1	36	5	23
DOL Vehicle ID	125701579	244285107	156773144	165103011	205138552
Vehicle Location	POINT (-122.30839 47.610365)	POINT (-122.179458 47.802589)	POINT (-122.34848 47.632405)	POINT (-122.03646 47.534065)	POINT (-122.55717 47.733415)
Electric Utility	CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)	PUGET SOUND ENERGY INC	CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)	PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)	PUGET SOUND ENERGY INC
2020 Census Tract	53033007800.0	53061051938.0	53033006800.0	53033032104.0	53035940100.0

- Base MSRP :- is 0 for all of the cars, which needs to be checked.
- Electric Range :- of some cars is 0 , the validation for this data is required
- Legislative District :- There exist a wrong entry '?' . This wrong entry needs to be treated.
- Null Values :- The dataset contains multiple null values.

Types of the data

info :- dislplays the type of the data present

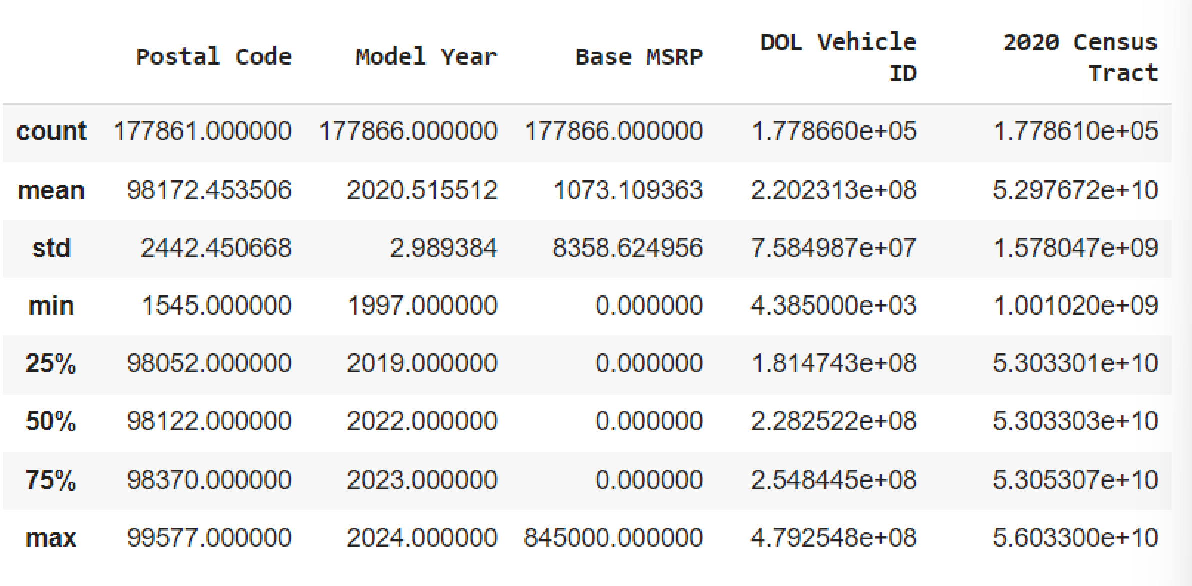


```
Data columns (total 17 columns):
 #   Column                                             Non-Null Count   Dtype  
---  ------                                             --------------   -----  
 0   VIN (1-10)                                         177866 non-null  object 
 1   County                                             177861 non-null  object 
 2   City                                               177861 non-null  object 
 3   State                                              177866 non-null  object 
 4   Postal Code                                        177861 non-null  float64
 5   Model Year                                         177866 non-null  int64  
 6   Make                                               177859 non-null  object 
 7   Model                                              177862 non-null  object 
 8   Electric Vehicle Type                              177860 non-null  object 
 9   Clean Alternative Fuel Vehicle (CAFV) Eligibility  177864 non-null  object 
 10  Electric Range                                     177863 non-null  object 
 11  Base MSRP                                          177866 non-null  int64  
 12  Legislative District                               177477 non-null  object 
 13  DOL Vehicle ID                                     177866 non-null  int64  
 14  Vehicle Location                                   177857 non-null  object 
 15  Electric Utility                                   177861 non-null  object 
 16  2020 Census Tract                                  177861 non-null  float64
dtypes: float64(2), int64(3), object(12)
memory usage: 23.1+ MB
```

- Postal Code :- the data type of this attribute should be object
- Electric Range :- The data type of this attribute should be integer or float

Statistical Summary of the data

Describe() :- Gives the statistical summary about the data

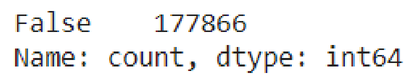


	Postal Code	Model Year	Base MSRP	DOL Vehicle ID	2020 Census Tract
count	177861.000000	177866.000000	177866.000000	1.778660e+05	1.778610e+05
mean	98172.453506	2020.515512	1073.109363	2.202313e+08	5.297672e+10
std	2442.450668	2.989384	8358.624956	7.584987e+07	1.578047e+09
min	1545.000000	1997.000000	0.000000	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

- Minimum Base MSRP is 0 , validity of the data is required to be checked. Outlier needed to be checked

Checking for Duplicated values

Duplicated() :- Returns True if the data is duplicated or else it returns false.



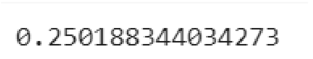
```
False    177866
Name: count, dtype: int64
```

The dataset contains NO duplicated values / row / entries

Checking for NULL Values

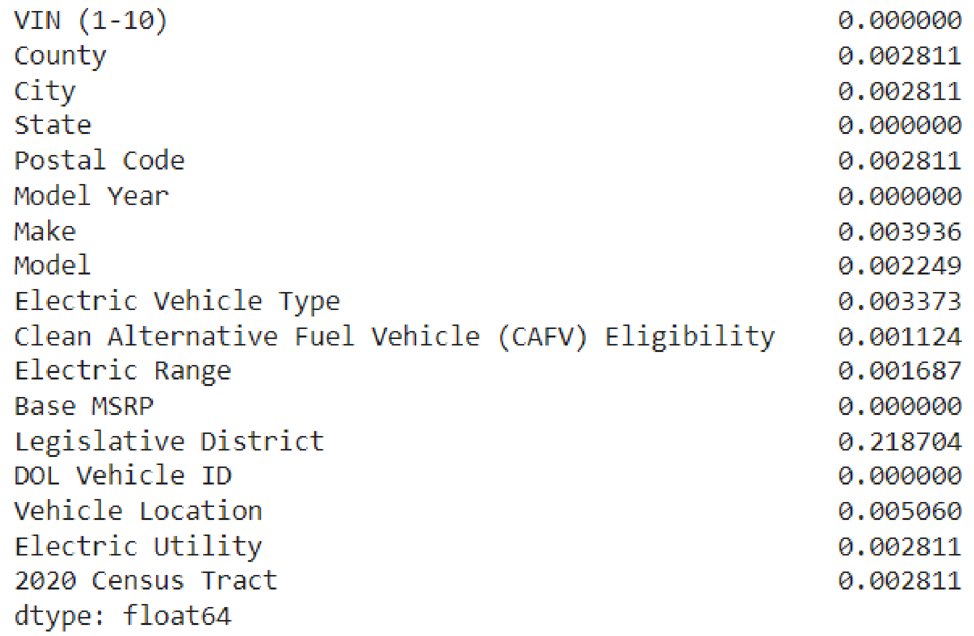
NULL() :- Returns True if the data is null or else returns false.

- The data set contains about 0.25% of null values



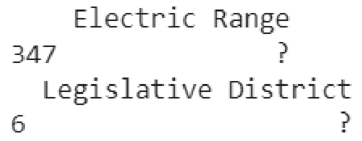
```
0.250188344034273
```

- Columnwise null value in per cent



```
VIN (1-10)                                            0.000000
County                                                0.002811
City                                                  0.002811
State                                                 0.000000
Postal Code                                           0.002811
Model Year                                            0.000000
Make                                                  0.003936
Model                                                 0.002249
Electric Vehicle Type                                 0.003373
Clean Alternative Fuel Vehicle (CAFV) Eligibility     0.001124
Electric Range                                        0.001687
Base MSRP                                             0.000000
Legislative District                                  0.218704
DOL Vehicle ID                                        0.000000
Vehicle Location                                      0.005060
Electric Utility                                      0.002811
2020 Census Tract                                     0.002811
dtype: float64
```

Checking for wrong entries



```
    Electric Range
347              ?
  Legislative District
6                    ?
```

The data set contains two worng entries '?'

- In Row: 347 and column : Electric Range
- In Row: 7 and Column : Legislative District

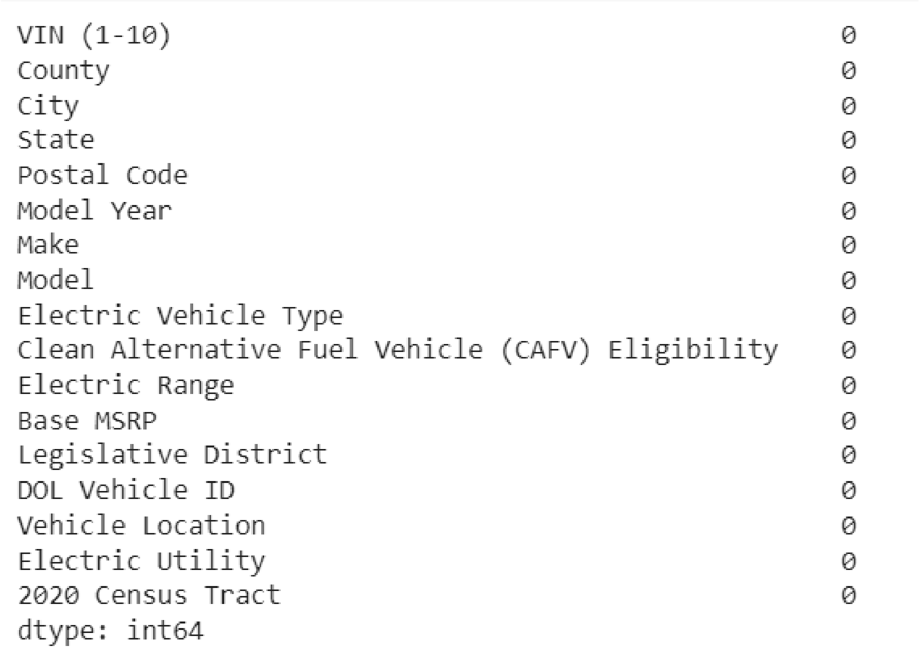
#Data Cleaning

## Treating Wrong Entries

Replacing '?' in Electric Range and Legislative District with null values

## Dropping the Null values

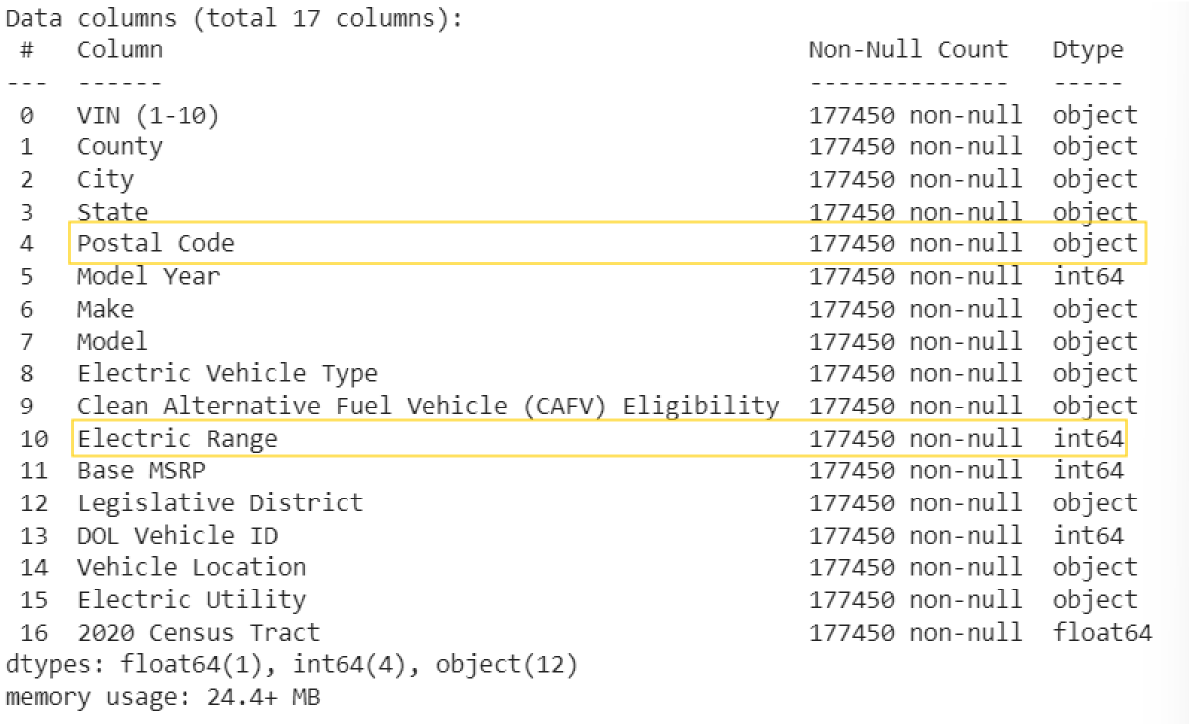
Dropping Null values in the dataset, as the total number of null values in the dataset is about 0.25% .



```
VIN (1-10)                                           0
County                                               0
City                                                 0
State                                                0
Postal Code                                          0
Model Year                                           0
Make                                                 0
Model                                                0
Electric Vehicle Type                                0
Clean Alternative Fuel Vehicle (CAFV) Eligibility    0
Electric Range                                       0
Base MSRP                                            0
Legislative District                                 0
DOL Vehicle ID                                       0
Vehicle Location                                     0
Electric Utility                                     0
2020 Census Tract                                    0
dtype: int64
```

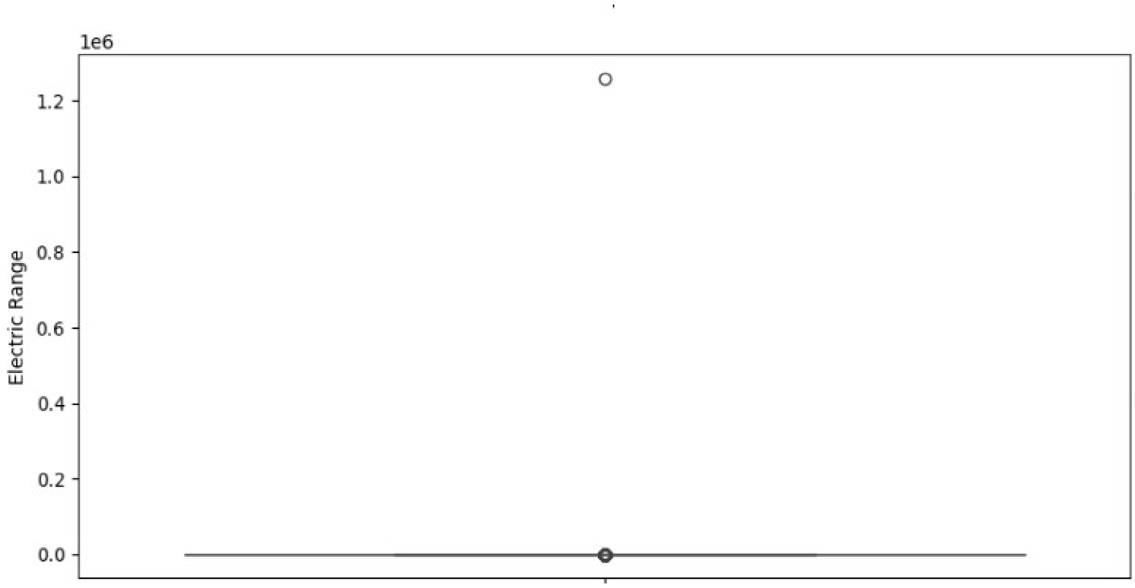
##Changing the data

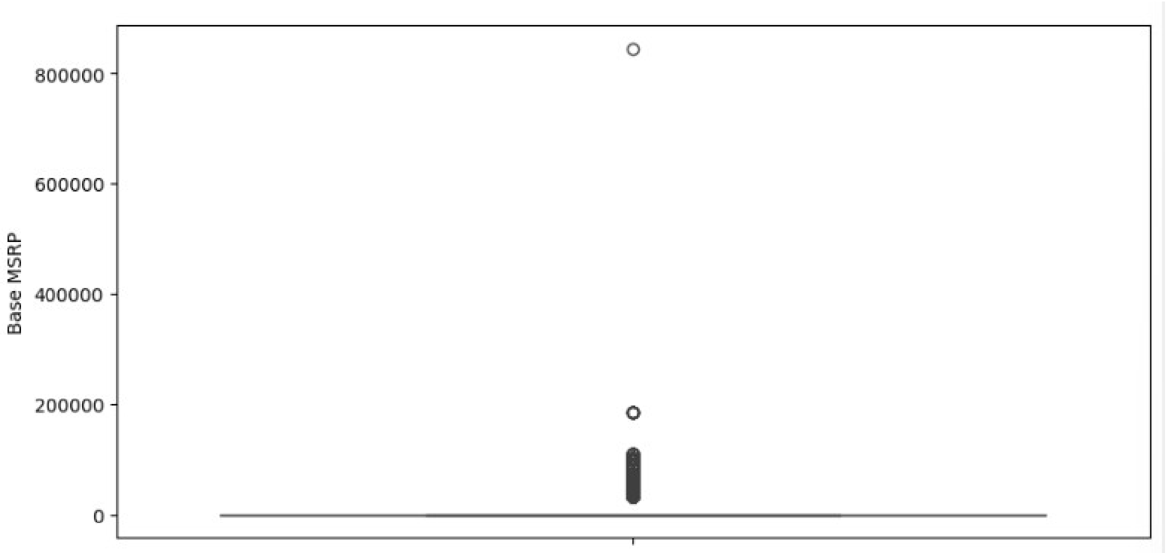
- Changing the datatype of "Postal Code" to "Object"
- Changing the datatype of "Electric Range" to "Integer"



```
Data columns (total 17 columns):
 #   Column                                             Non-Null Count   Dtype  
---  ------                                             --------------   -----  
 0   VIN (1-10)                                         177450 non-null  object 
 1   County                                             177450 non-null  object 
 2   City                                               177450 non-null  object 
 3   State                                              177450 non-null  object 
 4   Postal Code                                        177450 non-null  object 
 5   Model Year                                         177450 non-null  int64  
 6   Make                                               177450 non-null  object 
 7   Model                                              177450 non-null  object 
 8   Electric Vehicle Type                              177450 non-null  object 
 9   Clean Alternative Fuel Vehicle (CAFV) Eligibility  177450 non-null  object 
 10  Electric Range                                     177450 non-null  int64  
 11  Base MSRP                                          177450 non-null  int64  
 12  Legislative District                               177450 non-null  object 
 13  DOL Vehicle ID                                     177450 non-null  int64  
 14  Vehicle Location                                   177450 non-null  object 
 15  Electric Utility                                   177450 non-null  object 
 16  2020 Census Tract                                  177450 non-null  float64
dtypes: float64(1), int64(4), object(12)
memory usage: 24.4+ MB
```

Checking Outlier

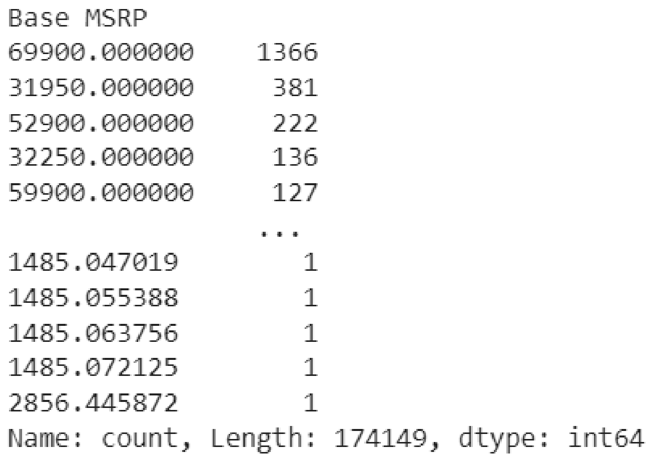




From the above plots we can infer that :-

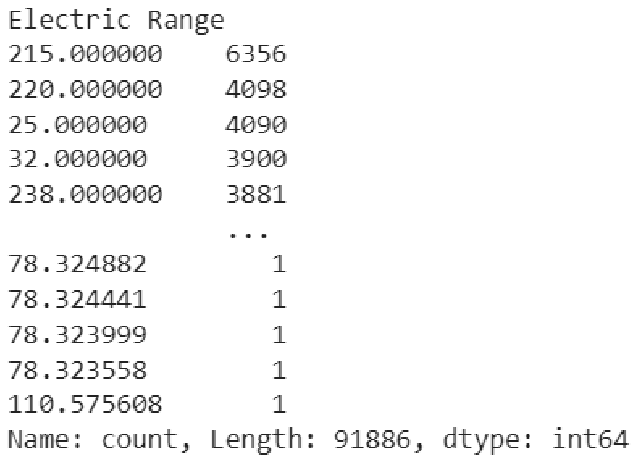
- Both 'Electric Range' and 'Base MSRP' contains outliers
- Both "Electric Range' and 'Base MSRP' contains more number of 0's than real data.

Replacing the 0's Values in Base MSRP with Mean:-



```
Base MSRP
69900.000000    1366
31950.000000     381
52900.000000     222
32250.000000     136
59900.000000     127
                ...
1485.047019        1
1485.055388        1
1485.063756        1
1485.072125        1
2856.445872        1
Name: count, Length: 174149, dtype: int64
```

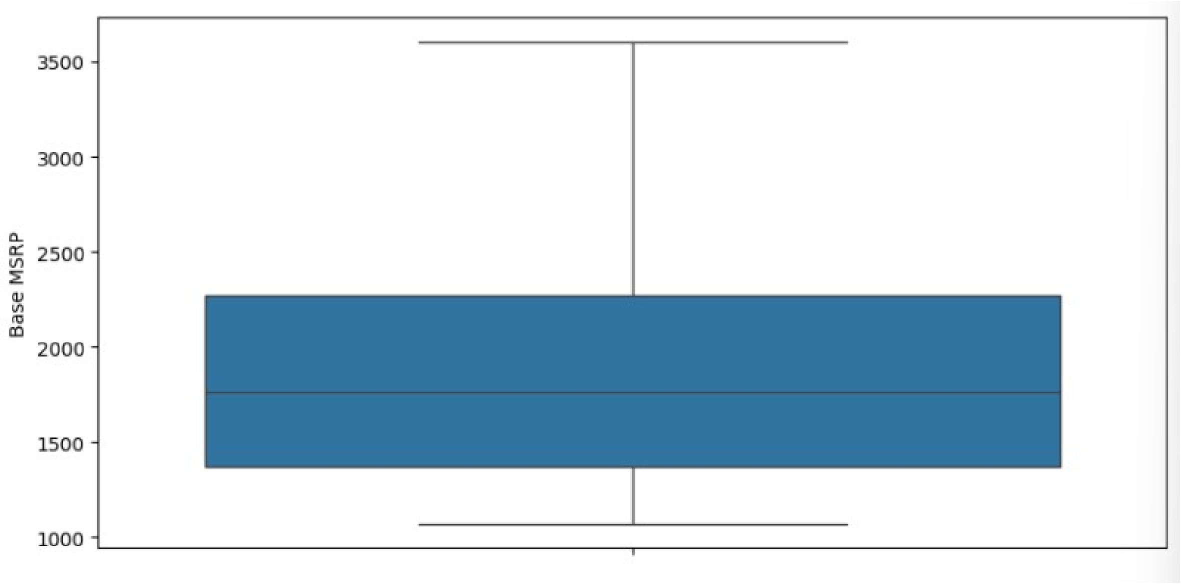
Replacing the 0's Values in Electric Range with Mean:-



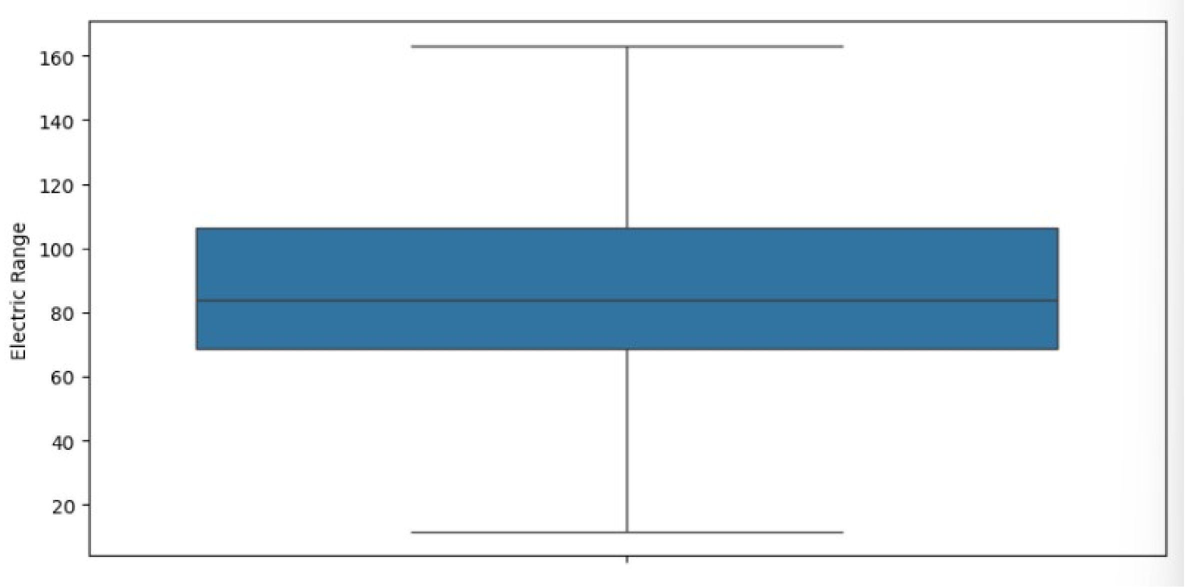
```
Electric Range
215.000000    6356
220.000000    4098
25.000000     4090
32.000000     3900
238.000000    3881
              ...
78.324882        1
78.324441        1
78.323999        1
78.323558        1
110.575608       1
Name: count, Length: 91886, dtype: int64
```

#Removing Outlier

Base MSRP

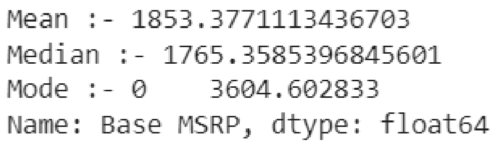


Electric Range



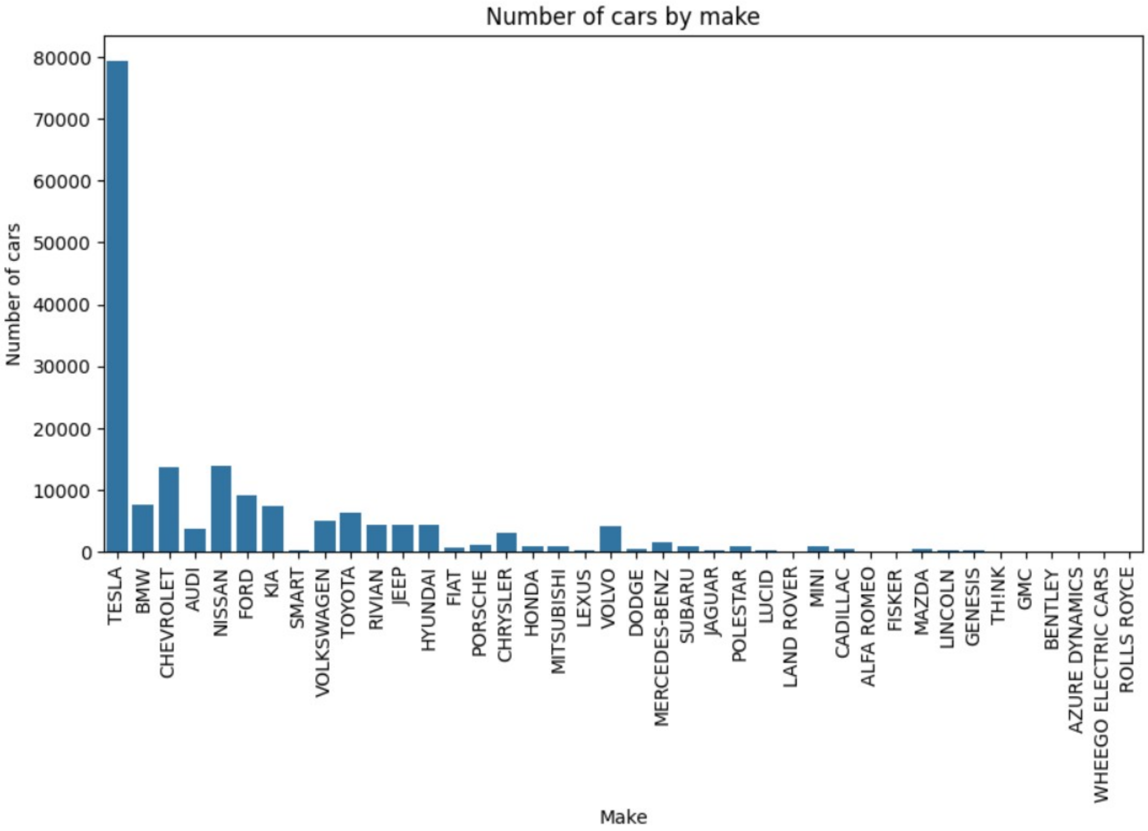
#Exploratory Data Analysis

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



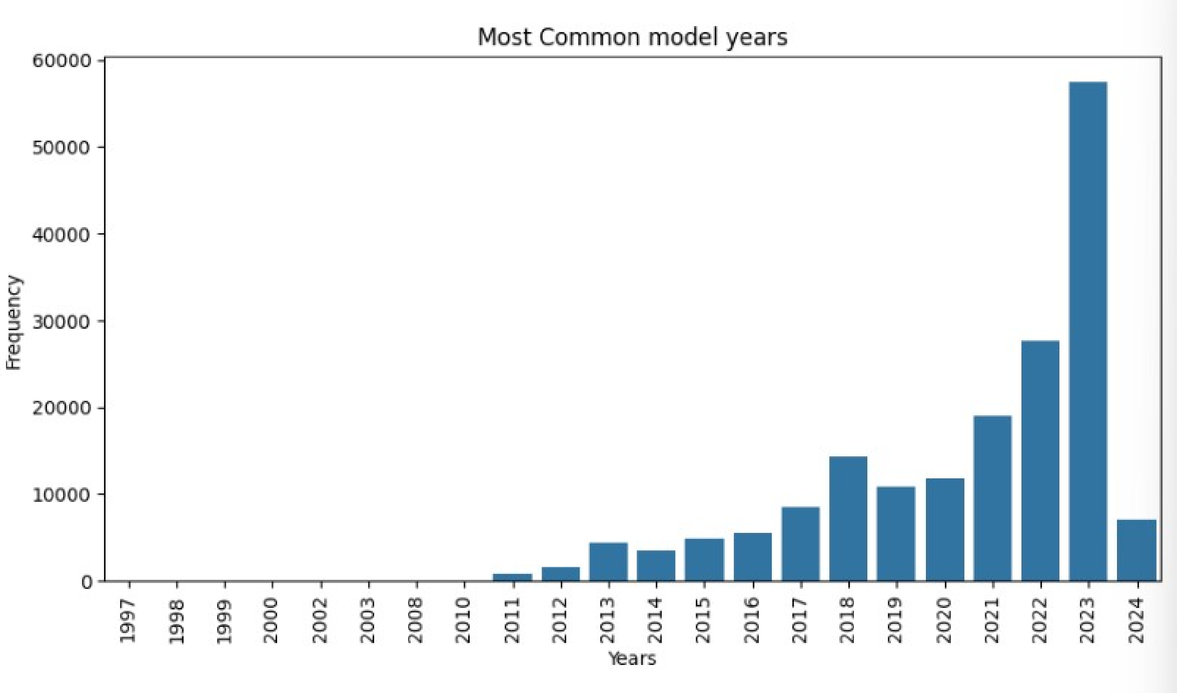
```
Mean :- 1853.3771113436703
Median :- 1765.3585396845601
Mode :- 0    3604.602833
Name: Base MSRP, dtype: float64
```

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



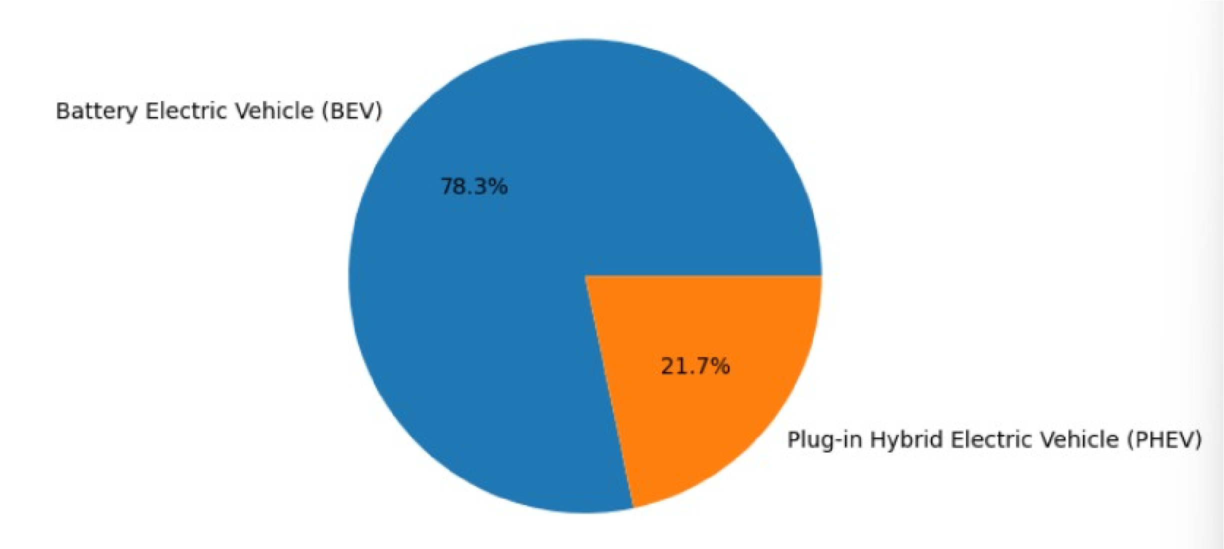
Tesla make most number of Electric Vehicle

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



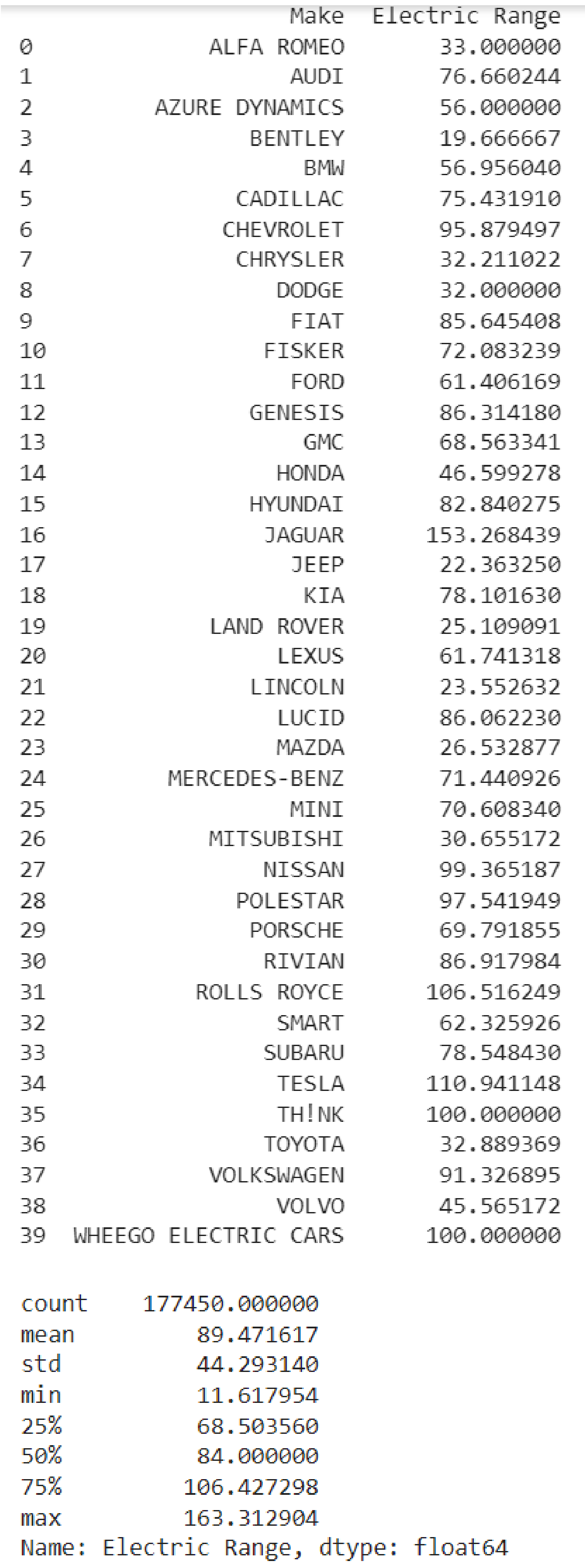
Most common year for Electric vehicle was 2023

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



The proportion of distribution of Battery Electric Vehicle (BEV) and Plug-in Hybrid Electric Vehicle (PHEV) is 78.3 : 21.7

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



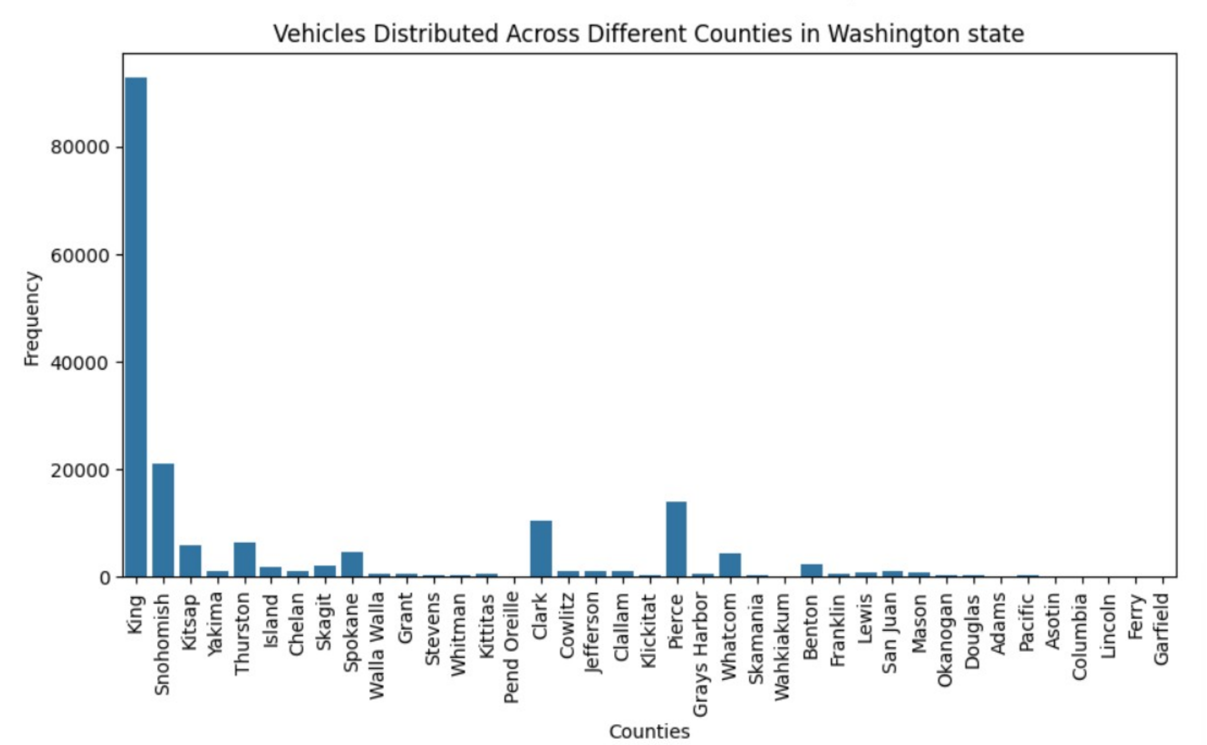
```
                   Make  Electric Range
0            ALFA ROMEO       33.000000
1                  AUDI       76.660244
2        AZURE DYNAMICS       56.000000
3               BENTLEY       19.666667
4                   BMW       56.956040
5              CADILLAC       75.431910
6             CHEVROLET       95.879497
7              CHRYSLER       32.211022
8                 DODGE       32.000000
9                  FIAT       85.645408
10               FISKER       72.083239
11                 FORD       61.406169
12              GENESIS       86.314180
13                  GMC       68.563341
14                HONDA       46.599278
15              HYUNDAI       82.840275
16               JAGUAR      153.268439
17                 JEEP       22.363250
18                  KIA       78.101630
19           LAND ROVER       25.109091
20                LEXUS       61.741318
21              LINCOLN       23.552632
22                LUCID       86.062230
23                MAZDA       26.532877
24        MERCEDES-BENZ       71.440926
25                 MINI       70.608340
26           MITSUBISHI       30.655172
27               NISSAN       99.365187
28             POLESTAR       97.541949
29              PORSCHE       69.791855
30               RIVIAN       86.917984
31          ROLLS ROYCE      106.516249
32                SMART       62.325926
33               SUBARU       78.548430
34                TESLA      110.941148
35                TH!NK      100.000000
36               TOYOTA       32.889369
37           VOLKSWAGEN       91.326895
38                VOLVO       45.565172
39  WHEEGO ELECTRIC CARS      100.000000

count    177450.000000
mean         89.471617
std          44.293140
min          11.617954
25%          68.503560
50%          84.000000
75%         106.427298
max         163.312904
Name: Electric Range, dtype: float64
```

Jaguar is having the maximum range : 153.26 Miles, whereas Bentley is having the minimum range : 19.67 Miles

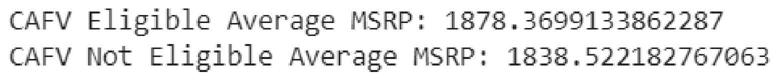
---

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



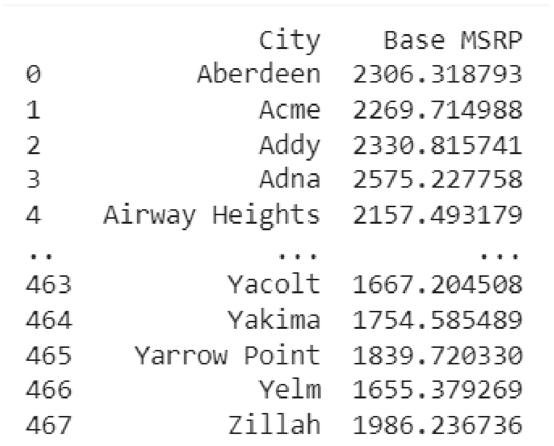
King county have the maximum number of Electric Vehicles registered, then at the Second we have Snohomish county .

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



```
CAFV Eligible Average MSRP: 1878.3699133862287
CAFV Not Eligible Average MSRP: 1838.522182767063
```

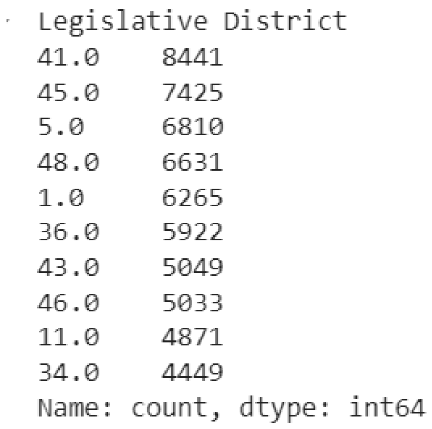
• How does the base MSRP vary across different cities in Washington state?



```
                City    Base MSRP
0           Aberdeen  2306.318793
1               Acme  2269.714988
2               Addy  2330.815741
3               Adna  2575.227758
4     Airway Heights  2157.493179
..               ...          ...
463           Yacolt  1667.204508
464           Yakima  1754.585489
465     Yarrow Point  1839.720330
466             Yelm  1655.379269
467           Zillah  1986.236736
```

We can observe that , Aberdeen have the highest Base MSRP for Electric Vehicle, then at the second postion we have Acme and then we havev Addy. The are the top 3 counties for having highest Base MSRP

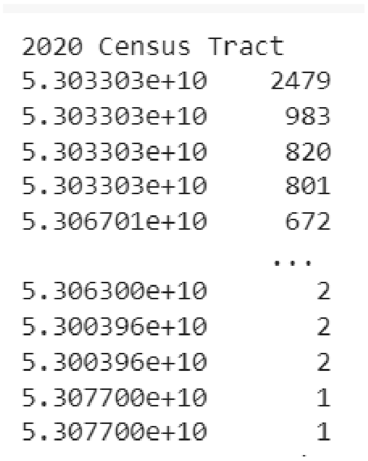
• Which legislative districts have the highest number of registered electric vehicles? Provide a ranked list.



```
Legislative District
41.0    8441
45.0    7425
5.0     6810
48.0    6631
1.0     6265
36.0    5922
43.0    5049
46.0    5033
11.0    4871
34.0    4449
Name: count, dtype: int64
```

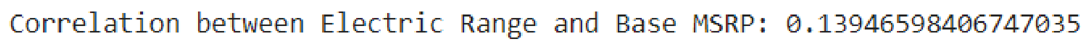
Legislative District 41 have the highest number of Electric Vehicle registered.

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



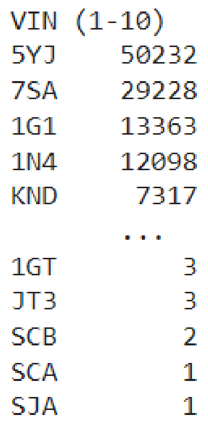
```
2020 Census Tract
5.303303e+10    2479
5.303303e+10     983
5.303303e+10     820
5.303303e+10     801
5.306701e+10     672
                ...
5.306300e+10       2
5.300396e+10       2
5.300396e+10       2
5.307700e+10       1
5.307700e+10       1
```

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



```
Correlation between Electric Range and Base MSRP: 0.13946598406747035
```

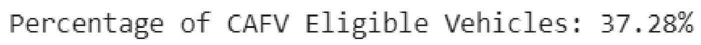
• Identify any patterns or commonalities in the VIN (1-10) for the vehicles. Are there any frequent prefixes or suffixes



```
VIN (1-10)
5YJ    50232
7SA    29228
1G1    13363
1N4    12098
KND     7317
       ...
1GT        3
JT3        3
SCB        2
SCA        1
SJA        1
```

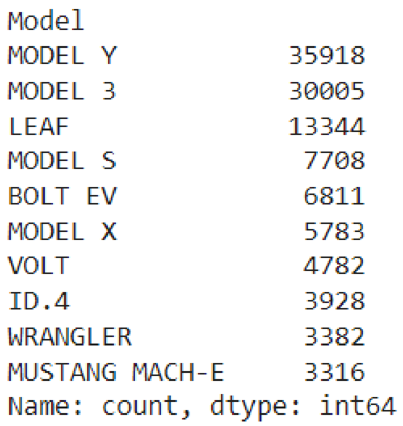
As we can observe from the above table, the prefix 5YJ is frequently used in the Electriv Vehicle registration.

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



```
Percentage of CAFV Eligible Vehicles: 37.28%
```

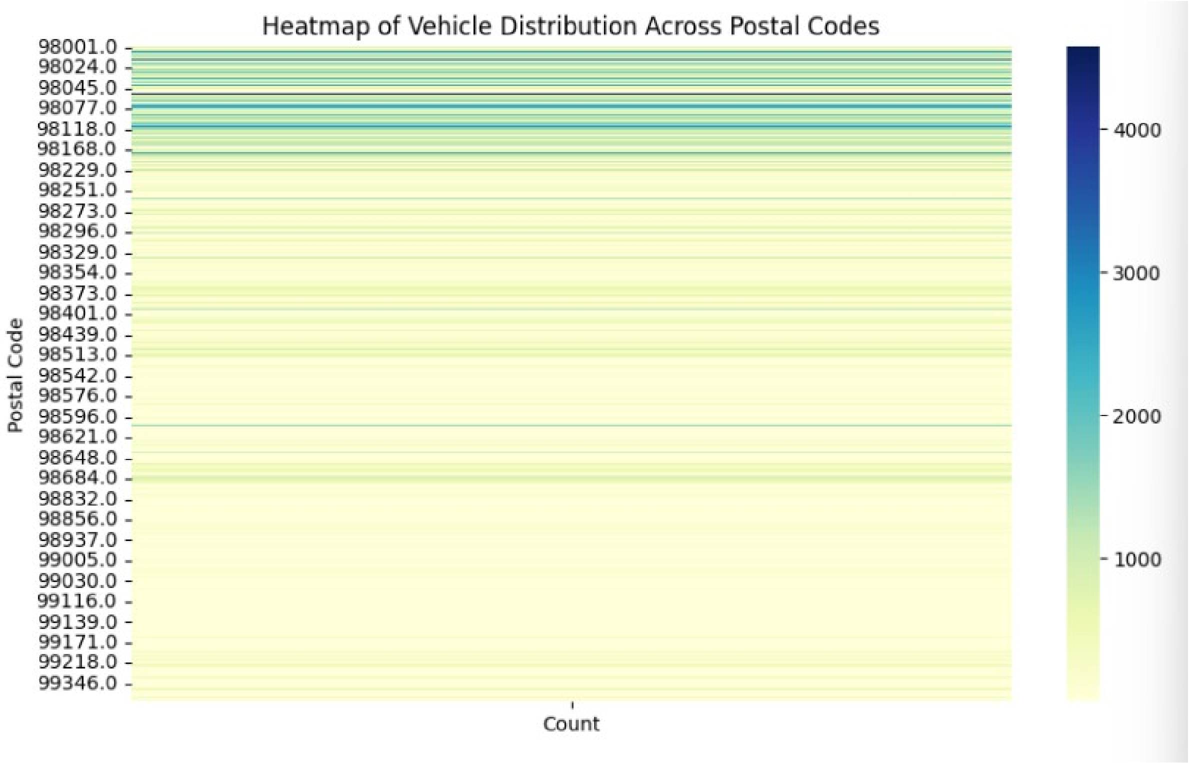
• Which vehicle models are the most popular in the dataset? Provide a frequency table of the top 10 models.



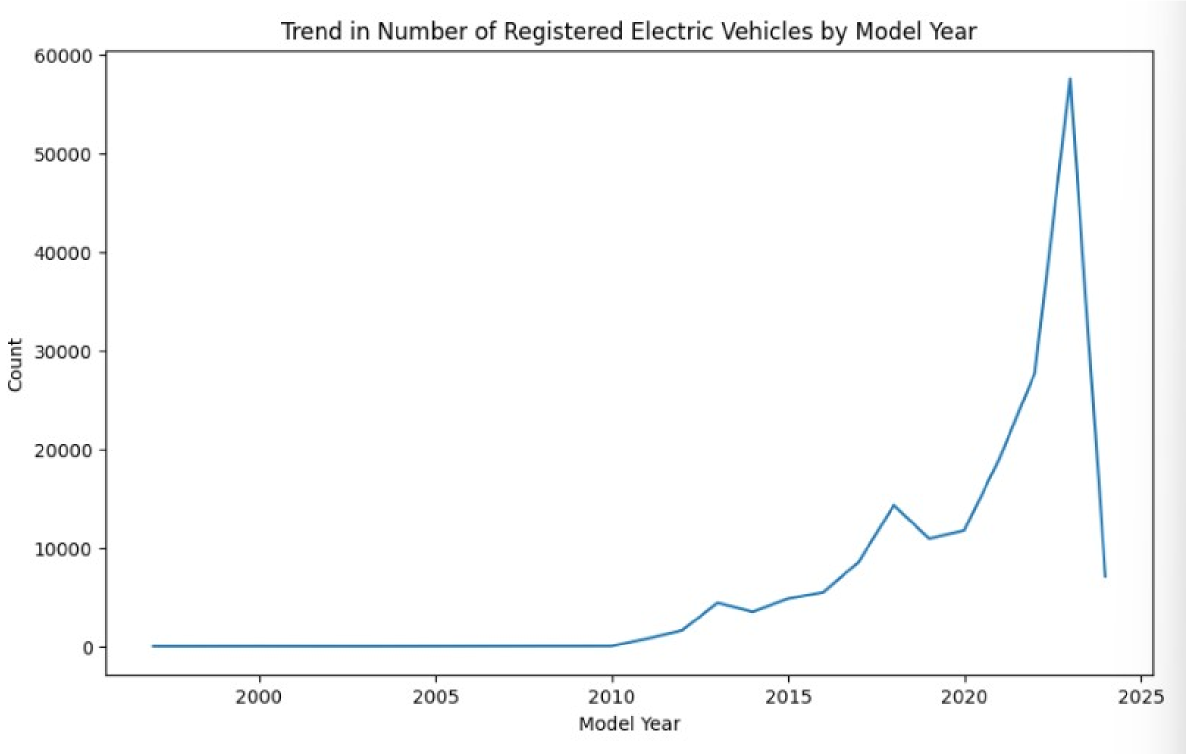
```
Model
MODEL Y           35918
MODEL 3           30005
LEAF              13344
MODEL S            7708
BOLT EV            6811
MODEL X            5783
VOLT               4782
ID.4               3928
WRANGLER           3382
MUSTANG MACH-E     3316
Name: count, dtype: int64
```

The most common model is Model Y from Tesla

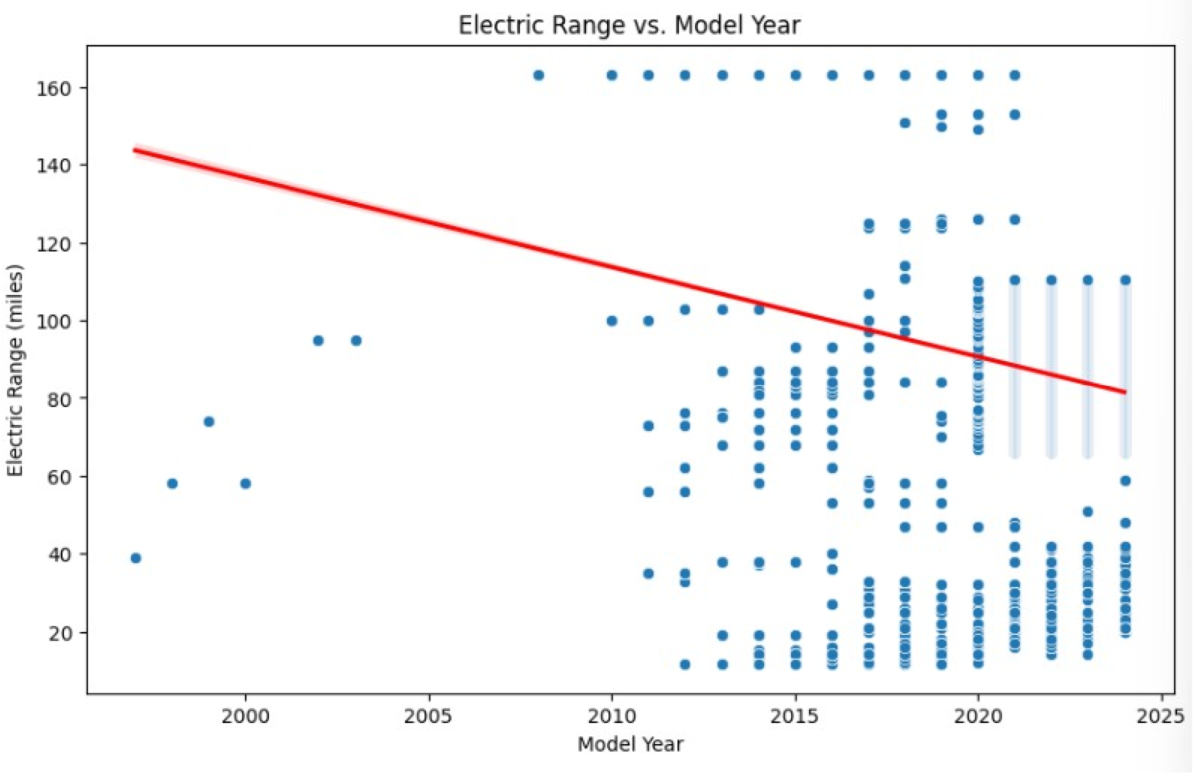
• How are vehicles distributed across different postal codes? Provide a heatmap or density plot.



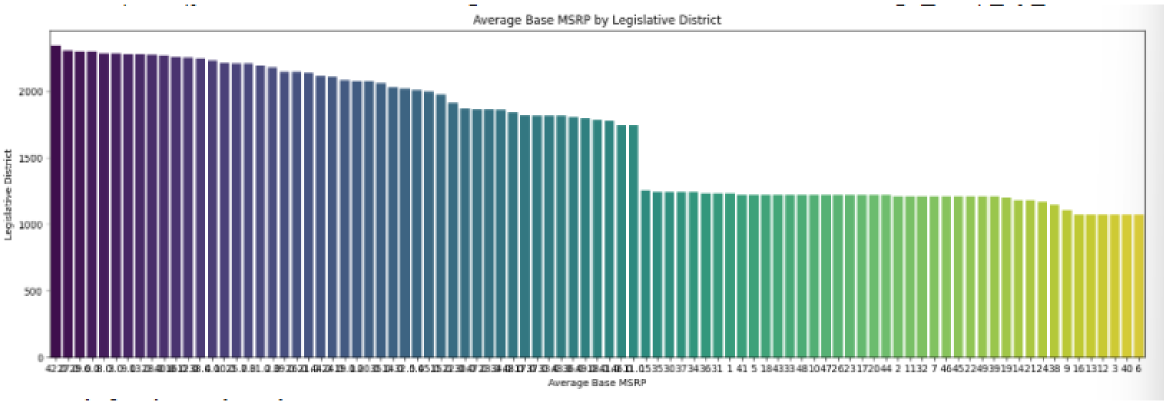
• Analyze the trend in the number of registered electric vehicles by model year. Provide a line chart to show any increase or decrease over the years.



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



- How does the average base MSRP vary across different legislative districts?



```
   Legislative District    Base MSRP
0                  42.0  2341.090058
1                  27.0  2301.736042
2                  29.0  2299.231995
3                   6.0  2298.479664
4                   8.0  2281.788409
..                  ...          ...
86                   13  1071.430688
87                   12  1071.413780
88                    3  1071.370299
89                   40  1071.325026
90                    6  1071.140892
```

# ***The End***