# ev-plotyexpress

October 10, 2024

### 0.0.1 Quick Description of the Project

**Project Title**: **Data Analysis and Visualization of Electric Vehicle Trends**

**Description**: This project focuses on analyzing a dataset related to electric vehicles (EVs) through various data analysis and visualization techniques. The aim is to uncover insights into the adoption and distribution of electric vehicles globally, using tools like Python and Plotly Express for visual storytelling.

**Objectives**: 1. **Exploratory Data Analysis (EDA)**: Perform univariate and bivariate analysis to explore key patterns in the EV dataset. - Understand the distribution of key variables like the number of EVs by year, country, and make. - Analyze relationships between variables to identify trends over time.

2. **Choropleth Map Creation**: Visualize the distribution of electric vehicles across different locations using Plotly Express.
   - Generate a Choropleth map that displays EV adoption based on geographic regions.
3. **Racing Bar Plot**: Animate the evolution of electric vehicle makes over time.
   - Implement a racing bar plot animation to dynamically show how different EV manufacturers' market shares have changed yearly.
4. **LinkedIn Post**: Present your work and key insights on LinkedIn to engage with the data science community, showcasing the project's real-world impact and visualization capabilities.

**Tools Used**: - **Pandas** for data manipulation and analysis. - **Plotly Express** for creating interactive maps and animations. - **Bar-Chart-Race** for generating animated visualizations of trends over time.

**Deliverables**: - **Choropleth Map** visualizing EV distribution across the world. - **Racing Bar Plot** showing the rise of various EV makes across years. - **LinkedIn Post** summarizing project insights.

**Collaboration**: This project is done in collaboration with Innomatics Research Labs, reflecting a real-world application of data analysis skill.

### 0.0.2 Project Summary:

This project focuses on analyzing a dataset related to electric vehicles (EVs) to uncover insights into their distribution, types, and trends over time. Through exploratory data analysis (EDA), the project reveals patterns of EV adoption, highlighting key features such as vehicle make, model, range, and geographic location.

Key findings include a growing trend in EV adoption, especially in states like California and Washington, with manufacturers like Tesla and Nissan leading the market. The project also emphasizes the importance of vehicle range and clean fuel eligibility in driving consumer choices. Visualizations such as Choropleth maps and Racing Bar Plots help illustrate the geographic distribution and evolving market dynamics over the years.

The project showcases advanced data analysis and visualization techniques, providing insights into the electric vehicle landscape and supporting the push toward sustainable transportation solutions.

### 0.0.3 Libraries

[90]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
```

[91]:

```
warnings.filterwarnings("ignore")
```

[12]:

```
!pip install plotly
```

```
Requirement already satisfied: plotly in
c:\users\lenovo\anaconda3\envs\notebook\lib\site-packages (5.23.0)
Requirement already satisfied: tenacity>=6.2.0 in
c:\users\lenovo\anaconda3\envs\notebook\lib\site-packages (from plotly) (8.5.0)
Requirement already satisfied: packaging in
c:\users\lenovo\anaconda3\envs\notebook\lib\site-packages (from plotly) (23.1)
```

[208]:

```
!pip install bar-chart-race
```

```
Collecting bar-chart-race
  Downloading bar_chart_race-0.1.0-py3-none-any.whl.metadata (4.2 kB)
Requirement already satisfied: pandas>=0.24 in
c:\users\lenovo\anaconda3\envs\notebook\lib\site-packages (from bar-chart-race)
(2.2.1)
Requirement already satisfied: matplotlib>=3.1 in
c:\users\lenovo\anaconda3\envs\notebook\lib\site-packages (from bar-chart-race)
(3.8.4)
Requirement already satisfied: contourpy>=1.0.1 in
c:\users\lenovo\anaconda3\envs\notebook\lib\site-packages (from
matplotlib>=3.1->bar-chart-race) (1.2.1)
Requirement already satisfied: cycler>=0.10 in
c:\users\lenovo\anaconda3\envs\notebook\lib\site-packages (from
matplotlib>=3.1->bar-chart-race) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
c:\users\lenovo\anaconda3\envs\notebook\lib\site-packages (from
matplotlib>=3.1->bar-chart-race) (4.51.0)
Requirement already satisfied: kiwisolver>=1.3.1 in
```

```
c:\users\lenovo\anaconda3\envs\notebook\lib\site-packages (from
matplotlib>=3.1->bar-chart-race) (1.4.5)
Requirement already satisfied: numpy>=1.21 in
c:\users\lenovo\anaconda3\envs\notebook\lib\site-packages (from
matplotlib>=3.1->bar-chart-race) (1.26.4)
Requirement already satisfied: packaging>=20.0 in
c:\users\lenovo\anaconda3\envs\notebook\lib\site-packages (from
matplotlib>=3.1->bar-chart-race) (23.1)
Requirement already satisfied: pillow>=8 in
c:\users\lenovo\anaconda3\envs\notebook\lib\site-packages (from
matplotlib>=3.1->bar-chart-race) (10.3.0)
Requirement already satisfied: pyparsing>=2.3.1 in
c:\users\lenovo\anaconda3\envs\notebook\lib\site-packages (from
matplotlib>=3.1->bar-chart-race) (3.1.2)
Requirement already satisfied: python-dateutil>=2.7 in
c:\users\lenovo\anaconda3\envs\notebook\lib\site-packages (from
matplotlib>=3.1->bar-chart-race) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
c:\users\lenovo\anaconda3\envs\notebook\lib\site-packages (from
pandas>=0.24->bar-chart-race) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.7 in
c:\users\lenovo\anaconda3\envs\notebook\lib\site-packages (from
pandas>=0.24->bar-chart-race) (2024.1)
Requirement already satisfied: six>=1.5 in
c:\users\lenovo\anaconda3\envs\notebook\lib\site-packages (from python-
dateutil>=2.7->matplotlib>=3.1->bar-chart-race) (1.16.0)
Downloading bar_chart_race-0.1.0-py3-none-any.whl (156 kB)
   ---------------------------------------- 0.0/156.8 kB ? eta -:--:--
   ---------- ----------------------------- 41.0/156.8 kB 667.8 kB/s eta 0:00:01
   ---------------------------------------  153.6/156.8 kB 1.8 MB/s eta 0:00:01
   ---------------------------------------- 156.8/156.8 kB 1.6 MB/s eta 0:00:00
Installing collected packages: bar-chart-race
Successfully installed bar-chart-race-0.1.0
```

[92]:

```
import plotly.express as px
import bar_chart_race as bcr
```

### 0.0.4 EDA

[93]:

```
df = pd.read_csv(r"C:\Users\Lenovo\Assignment\Plotly Express_Electric␣
 ↪Vehicles\dataset.csv")
```

[95]:

```
df.head()
```

[95]:

```
   VIN (1-10)     County      City State  Postal Code  Model Year       Make  \
0  JTMEB3FV6N     Monroe  Key West    FL        33040        2022     TOYOTA
1  1G1RD6E45D      Clark  Laughlin    NV        89029        2013  CHEVROLET
```

```
2  JN1AZ0CP8B     Yakima    Yakima    WA        98901        2011      NISSAN
3  1G1FW6S08H     Skagit  Concrete    WA        98237        2017   CHEVROLET
4  3FA6P0SU1K  Snohomish   Everett    WA        98201        2019        FORD

        Model                   Electric Vehicle Type  \
0  RAV4 PRIME  Plug-in Hybrid Electric Vehicle (PHEV)
1        VOLT  Plug-in Hybrid Electric Vehicle (PHEV)
2        LEAF          Battery Electric Vehicle (BEV)
3     BOLT EV          Battery Electric Vehicle (BEV)
4      FUSION  Plug-in Hybrid Electric Vehicle (PHEV)

  Clean Alternative Fuel Vehicle (CAFV) Eligibility  Electric Range  \
0           Clean Alternative Fuel Vehicle Eligible              42
1           Clean Alternative Fuel Vehicle Eligible              38
2           Clean Alternative Fuel Vehicle Eligible              73
3           Clean Alternative Fuel Vehicle Eligible             238
4             Not eligible due to low battery range              26

   Base MSRP  Legislative District  DOL Vehicle ID  \
0          0                   NaN       198968248
1          0                   NaN         5204412
2          0                  15.0       218972519
3          0                  39.0       186750406
4          0                  38.0         2006714

              Vehicle Location       Electric Utility  2020 Census Tract
0    POINT (-81.80023 24.5545)                    NaN        12087972100
1  POINT (-114.57245 35.16815)                    NaN        32003005702
2  POINT (-120.50721 46.60448)             PACIFICORP        53077001602
3   POINT (-121.7515 48.53892)  PUGET SOUND ENERGY INC        53057951101
4  POINT (-122.20596 47.97659)  PUGET SOUND ENERGY INC        53061041500
```

[96]:

```
df.shape
```

[96]: (112634, 17)

[97]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 112634 entries, 0 to 112633
Data columns (total 17 columns):
 #   Column                                             Non-Null Count   Dtype
---  ------                                             --------------   -----
 0   VIN (1-10)                                         112634 non-null  object
 1   County                                             112634 non-null  object
 2   City                                               112634 non-null  object
 3   State                                              112634 non-null  object
 4   Postal Code                                        112634 non-null  int64
 5   Model Year                                         112634 non-null  int64
 6   Make                                               112634 non-null  object
 7   Model                                              112614 non-null  object
 8   Electric Vehicle Type                              112634 non-null  object
 9   Clean Alternative Fuel Vehicle (CAFV) Eligibility  112634 non-null  object
 10  Electric Range                                     112634 non-null  int64
 11  Base MSRP                                          112634 non-null  int64
 12  Legislative District                               112348 non-null  float64
 13  DOL Vehicle ID                                     112634 non-null  int64
 14  Vehicle Location                                   112610 non-null  object
 15  Electric Utility                                   112191 non-null  object
 16  2020 Census Tract                                  112634 non-null  int64
dtypes: float64(1), int64(6), object(10)
memory usage: 14.6+ MB
```

```
[98]: df.describe()
```

```
[98]:          Postal Code     Model Year  Electric Range      Base MSRP  \
       count  112634.000000  112634.000000   112634.000000  112634.000000
       mean    98156.226850    2019.003365       87.812987    1793.439681
       std      2648.733064       2.892364      102.334216   10783.753486
       min      1730.000000    1997.000000        0.000000       0.000000
       25%     98052.000000    2017.000000        0.000000       0.000000
       50%     98119.000000    2020.000000       32.000000       0.000000
       75%     98370.000000    2022.000000      208.000000       0.000000
       max     99701.000000    2023.000000      337.000000  845000.000000

              Legislative District  DOL Vehicle ID  2020 Census Tract
       count         112348.000000    1.126340e+05       1.126340e+05
       mean              29.805604    1.994567e+08       5.296650e+10
       std               14.700545    9.398427e+07       1.699104e+09
       min                1.000000    4.777000e+03       1.101001e+09
       25%               18.000000    1.484142e+08       5.303301e+10
       50%               34.000000    1.923896e+08       5.303303e+10
       75%               43.000000    2.191899e+08       5.305307e+10
       max               49.000000    4.792548e+08       5.603300e+10
```

```
[99]: df.isna().sum()
```

```
[99]: VIN (1-10)                                              0
      County                                                  0
      City                                                    0
      State                                                   0
      Postal Code                                             0
      Model Year                                              0
      Make                                                    0
```

```
Model                                                 20
Electric Vehicle Type                                  0
Clean Alternative Fuel Vehicle (CAFV) Eligibility      0
Electric Range                                         0
Base MSRP                                              0
Legislative District                                 286
DOL Vehicle ID                                         0
Vehicle Location                                      24
Electric Utility                                     443
2020 Census Tract                                      0
dtype: int64
```

```
[100]: df["Make"].value_counts()
```

```
[100]: Make
       TESLA                   52078
       NISSAN                  12880
       CHEVROLET               10182
       FORD                     5819
       BMW                      4680
       KIA                      4483
       TOYOTA                   4405
       VOLKSWAGEN               2514
       AUDI                     2332
       VOLVO                    2288
       CHRYSLER                 1794
       HYUNDAI                  1412
       JEEP                     1152
       RIVIAN                    885
       FIAT                      822
       PORSCHE                   818
       HONDA                     792
       MINI                      632
       MITSUBISHI                588
       POLESTAR                  558
       MERCEDES-BENZ             506
       SMART                     273
       JAGUAR                    219
       LINCOLN                   168
       CADILLAC                  108
       LUCID MOTORS               65
       SUBARU                     59
       LAND ROVER                 38
       LEXUS                      33
       FISKER                     20
       GENESIS                    18
       AZURE DYNAMICS              7
```

```
TH!NK                    3
BENTLEY                  3
Name: count, dtype: int64
```

[101]:

```
missing_models = df[df['Model'].isna()]
missing_models
```

[101]:

```
        VIN (1-10)    County          City State  Postal Code  Model Year  \
13874   YV4ED3GM2P      King       Seattle    WA        98115        2023
30517   YV4ED3UL3P      King       Seattle    WA        98115        2023
31936   YV4ED3GM4P   Clallam        Sequim    WA        98382        2023
37517   YV4ED3UW2P  Snohomish      Edmonds    WA        98026        2023
58071   YV4ED3UM4P      King        Renton    WA        98058        2023
61626   YV4ED3GM5P    Pierce        Tacoma    WA        98465        2023
63240   YV4ED3GMXP      King       Redmond    WA        98052        2023
63380   YV4ED3GM7P      King       Seattle    WA        98122        2023
63462   YV4ED3UW4P      King     Newcastle    WA        98059        2023
78472   YV4ED3UM1P      King     Fall City    WA        98024        2023
81302   YV4ED3UM5P      King       Redmond    WA        98052        2023
84142   YV4ED3UM2P      King    North Bend    WA        98045        2023
86960   YV4ED3UM9P      King     Sammamish    WA        98075        2023
88687   YV4ED3GM5P      King  Maple Valley    WA        98038        2023
89882   YV4ED3UM5P      King      Bellevue    WA        98006        2023
93197   YV4ED3GM8P  Snohomish      Bothell    WA        98021        2023
103099  YV4ED3UW6P    Pierce        Milton    WA        98354        2023
103394  YV4ED3GM5P      King       Seattle    WA        98133        2023
108116  YV4ED3GL1P      King       Seattle    WA        98104        2023
112622  YV4ED3GM0P      King     Covington    WA        98042        2023

         Make Model           Electric Vehicle Type  \
13874   VOLVO   NaN  Battery Electric Vehicle (BEV)
30517   VOLVO   NaN  Battery Electric Vehicle (BEV)
31936   VOLVO   NaN  Battery Electric Vehicle (BEV)
37517   VOLVO   NaN  Battery Electric Vehicle (BEV)
58071   VOLVO   NaN  Battery Electric Vehicle (BEV)
61626   VOLVO   NaN  Battery Electric Vehicle (BEV)
63240   VOLVO   NaN  Battery Electric Vehicle (BEV)
63380   VOLVO   NaN  Battery Electric Vehicle (BEV)
63462   VOLVO   NaN  Battery Electric Vehicle (BEV)
78472   VOLVO   NaN  Battery Electric Vehicle (BEV)
81302   VOLVO   NaN  Battery Electric Vehicle (BEV)
84142   VOLVO   NaN  Battery Electric Vehicle (BEV)
86960   VOLVO   NaN  Battery Electric Vehicle (BEV)
88687   VOLVO   NaN  Battery Electric Vehicle (BEV)
89882   VOLVO   NaN  Battery Electric Vehicle (BEV)
93197   VOLVO   NaN  Battery Electric Vehicle (BEV)
103099  VOLVO   NaN  Battery Electric Vehicle (BEV)
```

```
103394  VOLVO   NaN  Battery Electric Vehicle (BEV)
108116  VOLVO   NaN  Battery Electric Vehicle (BEV)
112622  VOLVO   NaN  Battery Electric Vehicle (BEV)

        Clean Alternative Fuel Vehicle (CAFV) Eligibility  Electric Range  \
13874   Eligibility unknown as battery range has not b…                 0
30517   Eligibility unknown as battery range has not b…                 0
31936   Eligibility unknown as battery range has not b…                 0
37517   Eligibility unknown as battery range has not b…                 0
58071   Eligibility unknown as battery range has not b…                 0
61626   Eligibility unknown as battery range has not b…                 0
63240   Eligibility unknown as battery range has not b…                 0
63380   Eligibility unknown as battery range has not b…                 0
63462   Eligibility unknown as battery range has not b…                 0
78472   Eligibility unknown as battery range has not b…                 0
81302   Eligibility unknown as battery range has not b…                 0
84142   Eligibility unknown as battery range has not b…                 0
86960   Eligibility unknown as battery range has not b…                 0
88687   Eligibility unknown as battery range has not b…                 0
89882   Eligibility unknown as battery range has not b…                 0
93197   Eligibility unknown as battery range has not b…                 0
103099  Eligibility unknown as battery range has not b…                 0
103394  Eligibility unknown as battery range has not b…                 0
108116  Eligibility unknown as battery range has not b…                 0
112622  Eligibility unknown as battery range has not b…                 0

        Base MSRP  Legislative District  DOL Vehicle ID  \
13874           0                  46.0       221526476
30517           0                  43.0       223881556
31936           0                  24.0       219769000
37517           0                  32.0       218357779
58071           0                  11.0       224511766
61626           0                  28.0       224496702
63240           0                  48.0       221295224
63380           0                  37.0       224280472
63462           0                  41.0       218912410
78472           0                   5.0       224631494
81302           0                  48.0       220511791
84142           0                   5.0       223998148
86960           0                  41.0       214714706
88687           0                   5.0       224709726
89882           0                  41.0       214731254
93197           0                   1.0       220532063
103099          0                  30.0       213335454
103394          0                  46.0       220589967
108116          0                  37.0       219268451
112622          0                  47.0       224307996
```

```
                   Vehicle Location  \
13874   POINT (-122.31765 47.70013)
30517   POINT (-122.31765 47.70013)
31936   POINT (-123.10367 48.07965)
37517   POINT (-122.31768 47.87166)
58071    POINT (-122.08747 47.4466)
61626   POINT (-122.52886 47.24977)
63240   POINT (-122.13158 47.67858)
63380   POINT (-122.31009 47.60803)
63462   POINT (-122.15771 47.50549)
78472   POINT (-121.89086 47.56812)
81302   POINT (-122.13158 47.67858)
84142    POINT (-121.7831 47.49348)
86960   POINT (-122.03539 47.61344)
88687   POINT (-122.04526 47.39394)
89882   POINT (-122.12096 47.55584)
93197    POINT (-122.18384 47.8031)
103099  POINT (-122.32172 47.24898)
103394   POINT (-122.3503 47.71868)
108116  POINT (-122.32945 47.60357)
112622  POINT (-122.09124 47.33778)

                                         Electric Utility  2020 Census Tract
13874       PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)        53033002200
30517        CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)        53033003601
31936   BONNEVILLE POWER ADMINISTRATION||PUD NO 1 OF C…        53009002301
37517                              PUGET SOUND ENERGY INC        53061050700
58071       PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)        53033031911
61626   BONNEVILLE POWER ADMINISTRATION||CITY OF TACOM…        53053061001
63240       PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)        53033032324
63380        CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)        53033007800
63462       PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)        53033025005
78472       PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)        53033032221
81302       PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)        53033022902
84142       PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)        53033032704
86960       PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)        53033032213
88687       PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)        53033031604
89882       PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)        53033025007
93197                              PUGET SOUND ENERGY INC        53061051926
103099  BONNEVILLE POWER ADMINISTRATION||CITY OF MILTO…        53053070703
103394       CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)        53033000601
108116       CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)        53033009300
112622      PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)        53033031709
```

### 0.0.5 Data Description

The dataset consists of 112,634 entries and 17 columns, which provide detailed information about electric vehicles (EVs). Here is a description of the key features:

1. **VIN (1-10)**: A truncated version of the Vehicle Identification Number, unique to each vehicle.
2. **County**: The county where the vehicle is registered.
3. **City**: The city where the vehicle is located.
4. **State**: The U.S. state where the vehicle is registered.
5. **Postal Code**: The postal code of the vehicle's location.
6. **Model Year**: The year in which the vehicle was manufactured.
7. **Make**: The manufacturer of the vehicle (e.g., Toyota, Chevrolet).
8. **Model**: The specific model of the electric vehicle.
9. **Electric Vehicle Type**: Indicates whether the vehicle is a battery electric vehicle (BEV) or a plug-in hybrid electric vehicle (PHEV).
10. **Clean Alternative Fuel Vehicle (CAFV) Eligibility**: Whether the vehicle is eligible for clean alternative fuel incentives.
11. **Electric Range**: The estimated range of the vehicle on electric power alone, in miles.
12. **Base MSRP**: The manufacturer's suggested retail price for the vehicle.
13. **Legislative District**: The legislative district where the vehicle is registered.
14. **DOL Vehicle ID**: The Department of Licensing vehicle identification number.
15. **Vehicle Location**: The geographic coordinates (latitude and longitude) of the vehicle's registration.
16. **Electric Utility**: The utility provider for the vehicle's location.
17. **2020 Census Tract**: The census tract where the vehicle is registered, used for demographic analysis.

These features provide extensive information about EV distribution, types, and associated attributes such as range, price, and eligibility for alternative fuel programs.

[102]:

```
missing_value_percentage = (df.isna().sum()/len(df))*100
missing_value_percentage
```

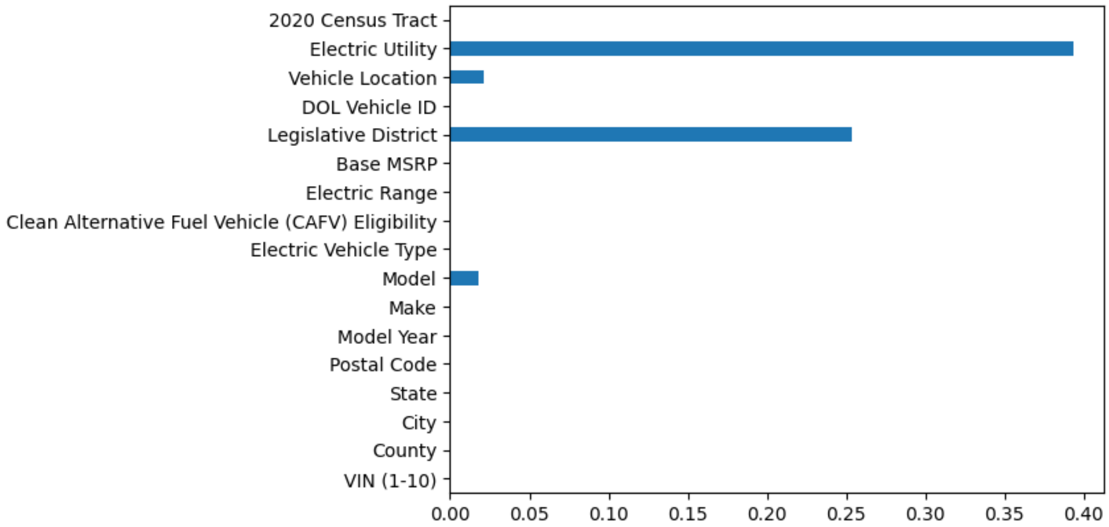
[102]:

```
VIN (1-10)                                           0.000000
County                                               0.000000
City                                                 0.000000
State                                                0.000000
Postal Code                                          0.000000
Model Year                                           0.000000
Make                                                 0.000000
Model                                                0.017757
Electric Vehicle Type                                0.000000
Clean Alternative Fuel Vehicle (CAFV) Eligibility    0.000000
Electric Range                                       0.000000
Base MSRP                                            0.000000
Legislative District                                 0.253920
DOL Vehicle ID                                       0.000000
Vehicle Location                                     0.021308
```

```
Electric Utility                                     0.393309
2020 Census Tract                                    0.000000
dtype: float64
```

```
[103]: missing_value_percentage.plot(kind="barh")
```

```
[103]: <Axes: >
```



```
[104]: makes_with_missing_models = missing_models.groupby('Make').size().
        ↪reset_index(name='missing_model_count')
       makes_with_missing_models
```

```
[104]:     Make  missing_model_count
       0  VOLVO                   20
```

```
[105]: model_mode = df["Model"].mode()[0]
       model_mode
```

```
[105]: 'MODEL 3'
```

```
[106]: df["Model"].fillna(model_mode, inplace=True)
```

```
[107]: df.isna().sum()
```

```
[107]: 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                               286
DOL Vehicle ID                                       0
Vehicle Location                                    24
Electric Utility                                   443
2020 Census Tract                                    0
dtype: int64
```

[108]:

```
location_mode = df["Vehicle Location"].mode()[0]
location_mode
```

[108]:

```
'POINT (-122.13158 47.67858)'
```

[109]:

```
df["Vehicle Location"].fillna(location_mode,inplace=True)
```

[110]:

```
df.isna().sum()
```

[110]:

```
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                               286
DOL Vehicle ID                                       0
Vehicle Location                                     0
Electric Utility                                   443
2020 Census Tract                                    0
dtype: int64
```

[111]:

```
df["Legislative District"].value_counts()
```

[111]:

```
Legislative District
41.0    7605
45.0    7112
```

```
48.0    6462
36.0    5251
46.0    4723
1.0     4715
5.0     4694
43.0    4621
37.0    3556
34.0    3478
18.0    3024
22.0    2782
32.0    2709
11.0    2707
44.0    2670
40.0    2633
23.0    2626
21.0    2617
26.0    2267
33.0    2112
10.0    2061
31.0    1912
17.0    1907
47.0    1876
24.0    1664
27.0    1654
42.0    1626
35.0    1620
39.0    1574
49.0    1573
28.0    1448
30.0    1268
2.0     1226
8.0     1157
38.0    1079
25.0    1049
6.0     1041
12.0    1004
20.0     973
4.0      845
13.0     748
14.0     720
29.0     692
19.0     672
16.0     611
9.0      606
3.0      557
7.0      544
15.0     277
```

```
Name: count, dtype: int64
```

[112]:

```
df[df["Legislative District"].isna()]
```

[112]:

```
        VIN (1-10)        County              City State  Postal Code  \
0       JTMEB3FV6N        Monroe           Key West    FL        33040
1       1G1RD6E45D         Clark           Laughlin    NV        89029
12      3C3CFFGE3G     St. Clair          Mascoutah    IL        62258
19      5YJXCAE28G      Saratoga  Greenfield Center    NY        12833
41      5YJSA1E22G  Newport News       Newport News    VA        23602
…            …           …                …   …          …
112153  1G1RB6E46F       Carroll       North Conway    NH         3860
112301  5YJ3E1EB9M    Dorchester        Summerville    SC        29483
112394  5UXKT0C36H   Leavenworth            Lansing    KS        66043
112541  JA4J24A50J      Williams          Williston    ND        58802
112603  7FCTGAAL7N      Kootenai             Worley    ID        83876

        Model Year        Make       Model  \
0             2022      TOYOTA  RAV4 PRIME
1             2013   CHEVROLET        VOLT
12            2016        FIAT         500
19            2016       TESLA     MODEL X
41            2016       TESLA     MODEL S
…            …         …         …
112153        2015   CHEVROLET        VOLT
112301        2021       TESLA     MODEL 3
112394        2017         BMW          X5
112541        2018  MITSUBISHI   OUTLANDER
112603        2022      RIVIAN         R1T

                         Electric Vehicle Type  \
0       Plug-in Hybrid Electric Vehicle (PHEV)
1       Plug-in Hybrid Electric Vehicle (PHEV)
12              Battery Electric Vehicle (BEV)
19              Battery Electric Vehicle (BEV)
41              Battery Electric Vehicle (BEV)
…                                        …
112153  Plug-in Hybrid Electric Vehicle (PHEV)
112301          Battery Electric Vehicle (BEV)
112394  Plug-in Hybrid Electric Vehicle (PHEV)
112541  Plug-in Hybrid Electric Vehicle (PHEV)
112603          Battery Electric Vehicle (BEV)

        Clean Alternative Fuel Vehicle (CAFV) Eligibility  Electric Range  \
0                 Clean Alternative Fuel Vehicle Eligible              42
1                 Clean Alternative Fuel Vehicle Eligible              38
12                Clean Alternative Fuel Vehicle Eligible              84
```

```
19                Clean Alternative Fuel Vehicle Eligible             200
41                Clean Alternative Fuel Vehicle Eligible             210
…                                                   …             …
112153            Clean Alternative Fuel Vehicle Eligible              38
112301  Eligibility unknown as battery range has not b…             0
112394              Not eligible due to low battery range              14
112541              Not eligible due to low battery range              22
112603  Eligibility unknown as battery range has not b…             0

        Base MSRP  Legislative District  DOL Vehicle ID  \
0               0                   NaN       198968248
1               0                   NaN         5204412
12              0                   NaN       153786167
19              0                   NaN       218050878
41              0                   NaN       111593331
…           …                   …             …
112153          0                   NaN       177816061
112301          0                   NaN       179604183
112394          0                   NaN       122897484
112541          0                   NaN         2592005
112603          0                   NaN       211894693

                   Vehicle Location Electric Utility  2020 Census Tract
0         POINT (-81.80023 24.5545)              NaN        12087972100
1       POINT (-114.57245 35.16815)              NaN        32003005702
12       POINT (-89.79939 38.49028)              NaN        17163504356
19        POINT (-73.84643 43.1284)              NaN        36091060601
41       POINT (-76.53585 37.10499)              NaN        51700032131
…                               …              …                …
112153   POINT (-71.12513 44.04945)              NaN        33003955302
112301   POINT (-80.17601 33.01897)              NaN        45035010506
112394   POINT (-94.89874 39.23762)              NaN        20103071104
112541     POINT (7.86484 51.32975)              NaN        38105954100
112603  POINT (-116.91895 47.40077)              NaN        16055940000

[286 rows x 17 columns]
```

**as Legislative District code is a number based on different location and it doesn't make any sense to fill NaN values with using mean or median we will be filling the missing value with "0"**

[113]:
```
df["Legislative District"].fillna(0, inplace=True)
```

[114]:
```
df.isna().sum()
```

[114]:
```
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                                     443
2020 Census Tract                                      0
dtype: int64
```

[115]:

```
df[df["Electric Utility"].isna()]
```

[115]:

```
        VIN (1-10)       County              City State  Postal Code  \
0       JTMEB3FV6N       Monroe          Key West    FL        33040
1       1G1RD6E45D        Clark          Laughlin    NV        89029
12      3C3CFFGE3G    St. Clair         Mascoutah    IL        62258
19      5YJXCAE28G     Saratoga  Greenfield Center    NY        12833
21      1G1RD6S55H      Stevens   Nine Mile Falls    WA        99026
…              …            …                 …     …          …
112301  5YJ3E1EB9M   Dorchester       Summerville    SC        29483
112321  1N4BZ1CP7K     Thurston           Olympia    WA        98502
112394  5UXKT0C36H  Leavenworth           Lansing    KS        66043
112541  JA4J24A50J     Williams         Williston    ND        58802
112603  7FCTGAAL7N     Kootenai            Worley    ID        83876

        Model Year        Make        Model  \
0             2022      TOYOTA   RAV4 PRIME
1             2013   CHEVROLET         VOLT
12            2016        FIAT          500
19            2016       TESLA      MODEL X
21            2017   CHEVROLET         VOLT
…              …           …            …
112301        2021       TESLA      MODEL 3
112321        2019      NISSAN         LEAF
112394        2017         BMW           X5
112541        2018  MITSUBISHI   OUTLANDER
112603        2022      RIVIAN          R1T

                         Electric Vehicle Type  \
0       Plug-in Hybrid Electric Vehicle (PHEV)
```

```
1       Plug-in Hybrid Electric Vehicle (PHEV)
12              Battery Electric Vehicle (BEV)
19              Battery Electric Vehicle (BEV)
21      Plug-in Hybrid Electric Vehicle (PHEV)
…                                         …
112301          Battery Electric Vehicle (BEV)
112321          Battery Electric Vehicle (BEV)
112394  Plug-in Hybrid Electric Vehicle (PHEV)
112541  Plug-in Hybrid Electric Vehicle (PHEV)
112603          Battery Electric Vehicle (BEV)

        Clean Alternative Fuel Vehicle (CAFV) Eligibility  Electric Range  \
0                  Clean Alternative Fuel Vehicle Eligible              42
1                  Clean Alternative Fuel Vehicle Eligible              38
12                 Clean Alternative Fuel Vehicle Eligible              84
19                 Clean Alternative Fuel Vehicle Eligible             200
21                 Clean Alternative Fuel Vehicle Eligible              53
…                                                    …             …
112301  Eligibility unknown as battery range has not b…                0
112321             Clean Alternative Fuel Vehicle Eligible             150
112394               Not eligible due to low battery range              14
112541               Not eligible due to low battery range              22
112603  Eligibility unknown as battery range has not b…                0

        Base MSRP  Legislative District  DOL Vehicle ID  \
0               0                   0.0       198968248
1               0                   0.0         5204412
12              0                   0.0       153786167
19              0                   0.0       218050878
21              0                   7.0       141964049
…           …                    …             …
112301          0                   0.0       179604183
112321          0                  22.0       142814556
112394          0                   0.0       122897484
112541          0                   0.0         2592005
112603          0                   0.0       211894693

                   Vehicle Location Electric Utility  2020 Census Tract
0         POINT (-81.80023 24.5545)              NaN        12087972100
1       POINT (-114.57245 35.16815)              NaN        32003005702
12       POINT (-89.79939 38.49028)              NaN        17163504356
19        POINT (-73.84643 43.1284)              NaN        36091060601
21      POINT (-117.54392 47.77676)              NaN        53065951402
…                              …              …                 …
112301   POINT (-80.17601 33.01897)              NaN        45035010506
112321  POINT (-122.92333 47.03779)              NaN        53067012002
112394   POINT (-94.89874 39.23762)              NaN        20103071104
```

```
112541     POINT (7.86484 51.32975)            NaN        38105954100
112603  POINT (-116.91895 47.40077)            NaN        16055940000

[443 rows x 17 columns]
```

[116]:

```
df.groupby("State")["Electric Utility"].unique().reset_index()
```

[116]:

```
   State                        Electric Utility
0     AK                                   [nan]
1     AL                                   [nan]
2     AR                                   [nan]
3     AZ                                   [nan]
4     CA                                   [nan]
5     CO                                   [nan]
6     CT                                   [nan]
7     DC                                   [nan]
8     DE                                   [nan]
9     FL                                   [nan]
10    GA                                   [nan]
11    HI                                   [nan]
12    ID                                   [nan]
13    IL                                   [nan]
14    KS                                   [nan]
15    KY                                   [nan]
16    LA                                   [nan]
17    MA                                   [nan]
18    MD                                   [nan]
19    ME                                   [nan]
20    MN                                   [nan]
21    MO                                   [nan]
22    MS                                   [nan]
23    NC                                   [nan]
24    ND                                   [nan]
25    NE                                   [nan]
26    NH                                   [nan]
27    NJ                                   [nan]
28    NM                                   [nan]
29    NV                                   [nan]
30    NY                                   [nan]
31    OH                                   [nan]
32    OK                                   [nan]
33    OR                                   [nan]
34    PA                                   [nan]
35    RI                                   [nan]
36    SC                                   [nan]
37    SD                                   [nan]
38    TN                                   [nan]
```

```
39    TX                                                    [nan]
40    UT                                                    [nan]
41    VA                                                    [nan]
42    WA  [PACIFICORP, PUGET SOUND ENERGY INC, PUD NO 2 …
43    WI                                                    [nan]
44    WY                                                    [nan]
```

[117]:

```
df["Electric Utility"].fillna("unknown", inplace=True)
```

[118]:

```
df.isna().sum()
```

[118]:

```
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
```

[119]:

```
df.duplicated().sum()
```

[119]:

```
0
```

### 0.0.6 Unique value in each feature

[120]:

```
for i in df.columns:
    print(f"{i}:{df[i].nunique()} are the total unique values out of {df[i].
 ↪count()}")
```

```
VIN (1-10):7548 are the total unique values out of 112634
County:165 are the total unique values out of 112634
City:629 are the total unique values out of 112634
State:45 are the total unique values out of 112634
Postal Code:773 are the total unique values out of 112634
Model Year:20 are the total unique values out of 112634
Make:34 are the total unique values out of 112634
```

```
Model:114 are the total unique values out of 112634
Electric Vehicle Type:2 are the total unique values out of 112634
Clean Alternative Fuel Vehicle (CAFV) Eligibility:3 are the total unique values
out of 112634
Electric Range:101 are the total unique values out of 112634
Base MSRP:30 are the total unique values out of 112634
Legislative District:50 are the total unique values out of 112634
DOL Vehicle ID:112634 are the total unique values out of 112634
Vehicle Location:758 are the total unique values out of 112634
Electric Utility:74 are the total unique values out of 112634
2020 Census Tract:2026 are the total unique values out of 112634
```

[121]:

```
df["Make_Model"] = df["Make"]+ "-" +df["Model"]
```

[122]:

```
df["Make_Model"].value_counts()
```

[122]:

```
Make_Model
TESLA-MODEL 3           23135
TESLA-MODEL Y           17142
NISSAN-LEAF             12880
TESLA-MODEL S            7377
CHEVROLET-BOLT EV        4910
                         …
BMW-745LE                   2
CHEVROLET-S-10 PICKUP       1
SUBARU-SOLTERRA             1
PORSCHE-918                 1
BENTLEY-FLYING SPUR         1
Name: count, Length: 115, dtype: int64
```

[123]:

```
df.columns
```

[123]:

```
Index(['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',
       'Make_Model'],
      dtype='object')
```

[124]:

```
df.head()
```

[124]:

```
   VIN (1-10)     County       City State  Postal Code  Model Year       Make  \
0  JTMEB3FV6N     Monroe   Key West    FL        33040        2022     TOYOTA
1  1G1RD6E45D      Clark   Laughlin    NV        89029        2013  CHEVROLET
2  JN1AZ0CP8B     Yakima     Yakima    WA        98901        2011     NISSAN
3  1G1FW6S08H     Skagit   Concrete    WA        98237        2017  CHEVROLET
```

```
4  3FA6P0SU1K  Snohomish   Everett    WA        98201        2019       FORD

        Model                   Electric Vehicle Type  \
0  RAV4 PRIME  Plug-in Hybrid Electric Vehicle (PHEV)
1        VOLT  Plug-in Hybrid Electric Vehicle (PHEV)
2        LEAF          Battery Electric Vehicle (BEV)
3     BOLT EV          Battery Electric Vehicle (BEV)
4      FUSION  Plug-in Hybrid Electric Vehicle (PHEV)

  Clean Alternative Fuel Vehicle (CAFV) Eligibility  Electric Range  \
0           Clean Alternative Fuel Vehicle Eligible              42
1           Clean Alternative Fuel Vehicle Eligible              38
2           Clean Alternative Fuel Vehicle Eligible              73
3           Clean Alternative Fuel Vehicle Eligible             238
4             Not eligible due to low battery range              26

   Base MSRP  Legislative District  DOL Vehicle ID  \
0          0                   0.0       198968248
1          0                   0.0         5204412
2          0                  15.0       218972519
3          0                  39.0       186750406
4          0                  38.0         2006714

              Vehicle Location        Electric Utility  2020 Census Tract  \
0    POINT (-81.80023 24.5545)                 unknown        12087972100
1  POINT (-114.57245 35.16815)                 unknown        32003005702
2  POINT (-120.50721 46.60448)              PACIFICORP        53077001602
3   POINT (-121.7515 48.53892)  PUGET SOUND ENERGY INC        53057951101
4  POINT (-122.20596 47.97659)  PUGET SOUND ENERGY INC        53061041500

          Make_Model
0  TOYOTA-RAV4 PRIME
1     CHEVROLET-VOLT
2        NISSAN-LEAF
3  CHEVROLET-BOLT EV
4        FORD-FUSION
```

### 0.0.7 Relationship between numeric features

[125]:

```
df_numeric = df.select_dtypes(include=[np.number])
df_numeric
```

[125]:

```
        Postal Code  Model Year  Electric Range  Base MSRP  \
0             33040        2022              42          0
1             89029        2013              38          0
2             98901        2011              73          0
3             98237        2017             238          0
```

```
4             98201         2019              26           0
…             …          …             …         …
112629        98019         2022               0           0
112630        98250         2019             150           0
112631        98070         2022              38           0
112632        98042         2018              26           0
112633        98042         2022              18           0

        Legislative District  DOL Vehicle ID  2020 Census Tract
0                        0.0       198968248        12087972100
1                        0.0         5204412        32003005702
2                       15.0       218972519        53077001602
3                       39.0       186750406        53057951101
4                       38.0         2006714        53061041500
…                        …             …                 …
112629                  45.0       217955265        53033032401
112630                  40.0       103663227        53055960301
112631                  34.0       193878387        53033027702
112632                  47.0       125039043        53033032007
112633                  47.0       194673692        53033032005

[112634 rows x 7 columns]
```

```
[126]: corr_matrix = df_numeric.corr()
       styled_corr = corr_matrix.style.background_gradient(cmap='cividis')
       styled_corr
```

```
[126]: <pandas.io.formats.style.Styler at 0x214bed58250>
```

```
[127]: df.columns
```

```
[127]: Index(['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',
              'Make_Model'],
             dtype='object')
```

### 1. Which counties have the highest number of electric vehicles (EVs)?

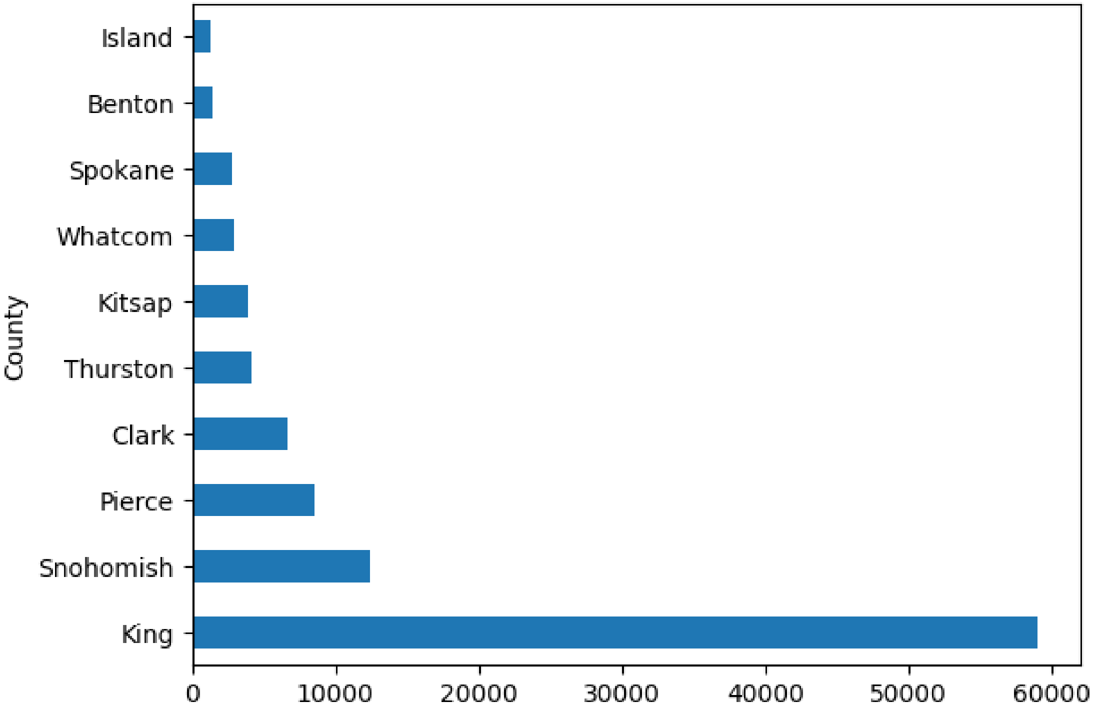
```
[128]: df['County'].value_counts().head(10)
```

```
[128]: County
       King         59000
       Snohomish    12434
       Pierce        8535
```

```
Clark          6689
Thurston       4126
Kitsap         3847
Whatcom        2840
Spokane        2792
Benton         1376
Island         1307
Name: count, dtype: int64
```

[129]:
```
df['County'].value_counts().head(10).plot(kind="barh")
```

[129]:
```
<Axes: ylabel='County'>
```



**2. Which cities have the highest number of electric vehicles (EVs)?**

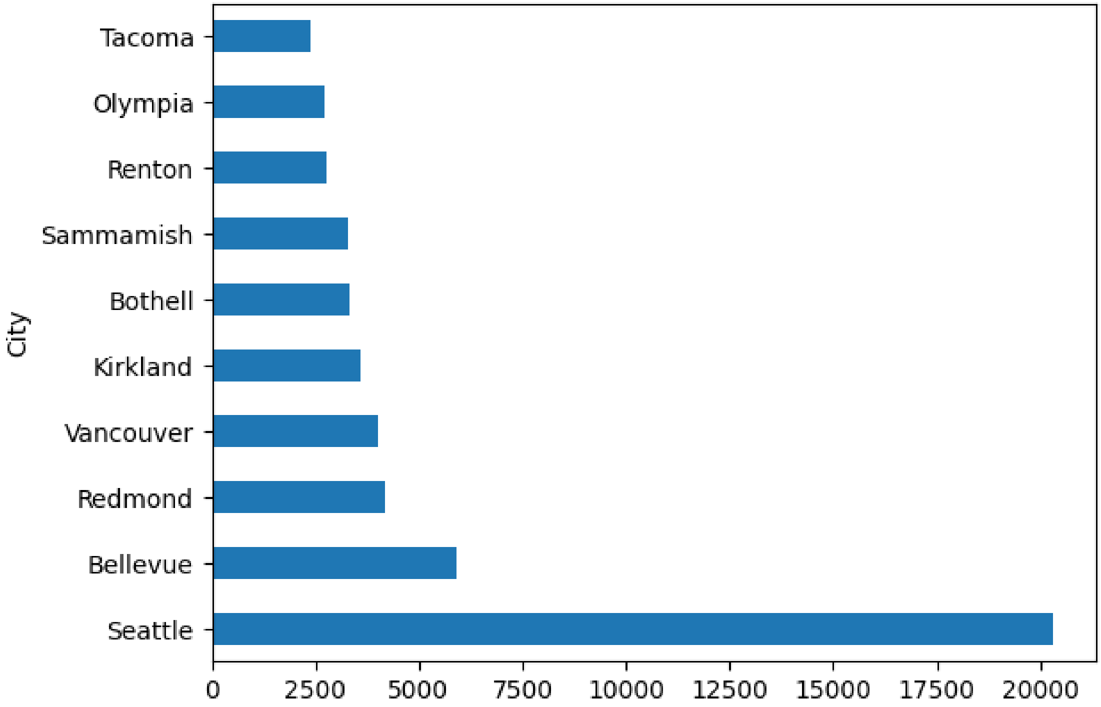
[130]:
```
df['City'].value_counts().head(10)
```

[130]:
```
City
Seattle       20305
Bellevue       5921
Redmond        4201
Vancouver      4013
Kirkland       3598
Bothell        3335
```

```
Sammamish     3292
Renton        2778
Olympia       2745
Tacoma        2379
Name: count, dtype: int64
```

[131]:
```
df['City'].value_counts().head(10).plot(kind="barh")
```

[131]:
```
<Axes: ylabel='City'>
```



### 3. Who are the top 10 EV manufacturers?

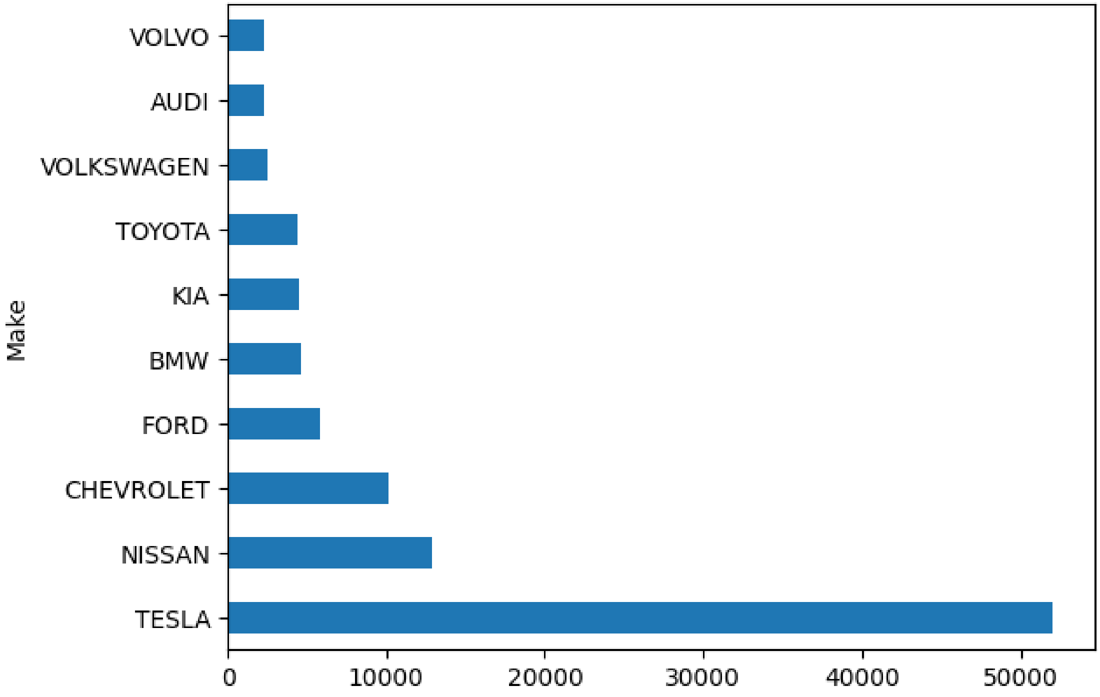
[132]:
```
df['Make'].value_counts().head(10)
```

[132]:
```
Make
TESLA         52078
NISSAN        12880
CHEVROLET     10182
FORD           5819
BMW            4680
KIA            4483
TOYOTA         4405
VOLKSWAGEN     2514
AUDI           2332
```

```
VOLVO           2288
Name: count, dtype: int64
```

```
[133]: df['Make'].value_counts().head(10).plot(kind="barh")
```

```
[133]: <Axes: ylabel='Make'>
```



### 4. What are the top 10 EV models?

```
[134]: df['Model'].value_counts().head(10)
```

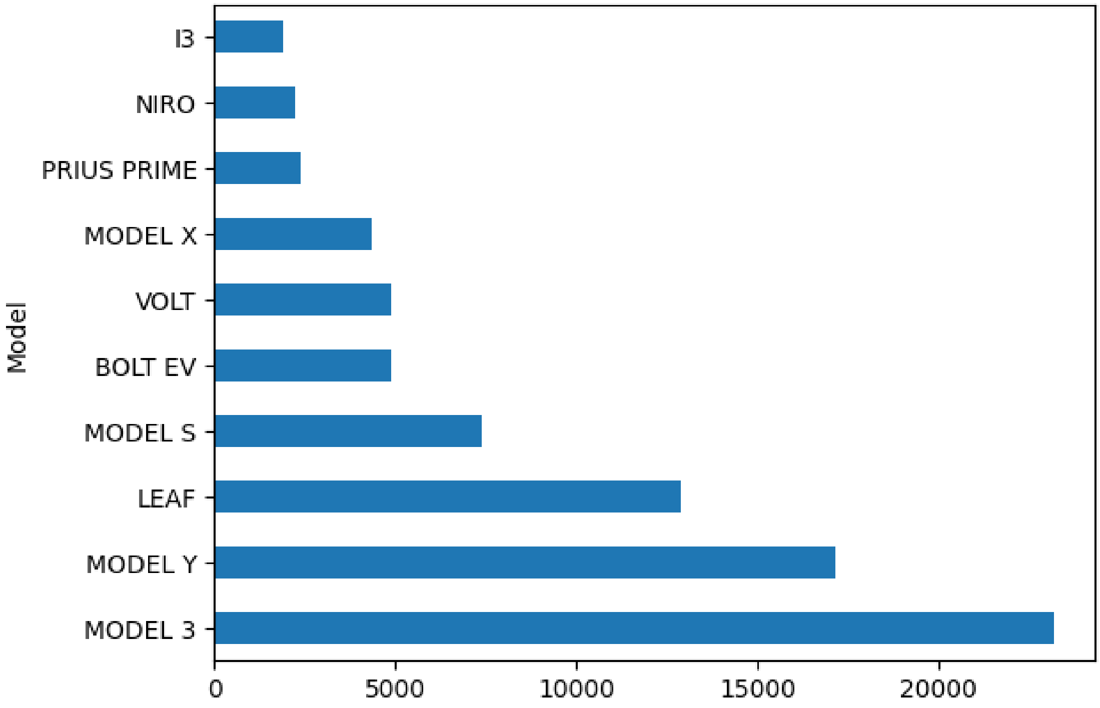
```
[134]: Model
       MODEL 3         23155
       MODEL Y         17142
       LEAF            12880
       MODEL S          7377
       BOLT EV          4910
       VOLT             4896
       MODEL X          4370
       PRIUS PRIME      2380
       NIRO             2260
       I3               1896
       Name: count, dtype: int64
```

[135]:

```
df['Model'].value_counts().head(10).plot(kind="barh")
```

[135]:

```
<Axes: ylabel='Model'>
```



### 5. What is the count and percentage distribution of Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs)?

[136]:

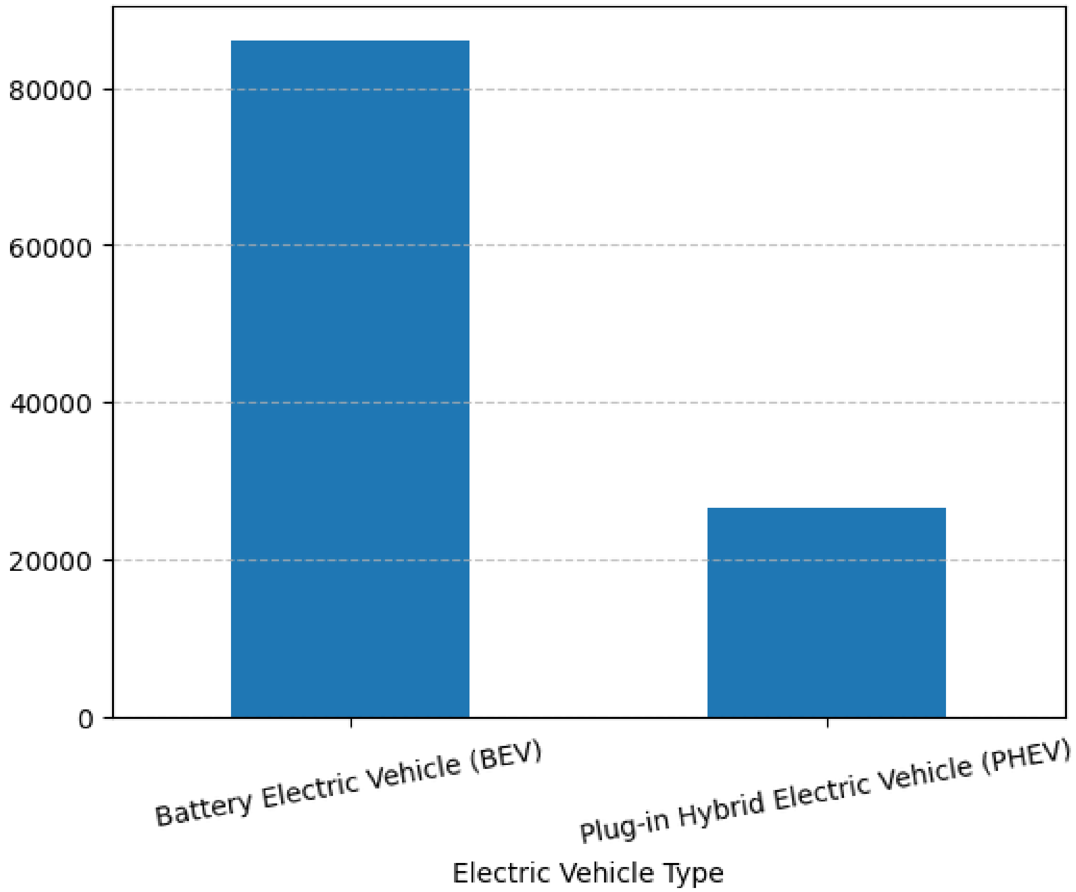
```
df['Electric Vehicle Type'].value_counts()
```

[136]:

```
Electric Vehicle Type
Battery Electric Vehicle (BEV)            86044
Plug-in Hybrid Electric Vehicle (PHEV)    26590
Name: count, dtype: int64
```

[137]:

```
df['Electric Vehicle Type'].value_counts().plot(kind="bar")
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.xticks(rotation=10)
plt.show()
```

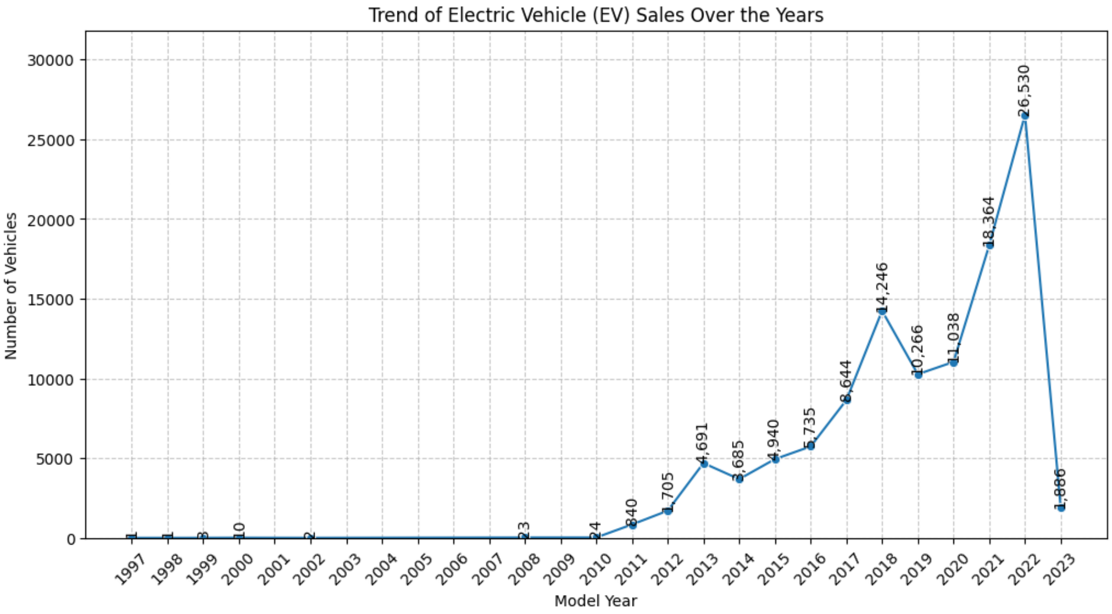


**6. What is the trend of EV sales over the years?**

[138]:
```
year_wise = df['Model Year'].value_counts()
```

[139]:
```
plt.figure(figsize= (12,6))
sns.lineplot( x = year_wise.index , y =year_wise.values  ,marker= 'o')
plt.title("Trend of Electric Vehicle (EV) Sales Over the Years")
plt.xlabel('Model Year')
plt.ylabel('Number of Vehicles')
plt.grid('True' ,linestyle ='--', alpha=0.7 )
plt.xticks(rotation = 45)

#incrase the number of interval on y axis
max_value = df['Model Year'].value_counts().max()
plt.ylim(0, max_value *1.2)

# Increase the number of intervals on the x-axis
plt.xticks(ticks=np.arange(year_wise.index.min(), year_wise.index.max() + 1, 1))
```

```
# Annotate each point with the exact count
for index, value in enumerate(year_wise.values):
    plt.text(year_wise.index[index], value, f'{value:,}', ha='center',␣
 ↪va='bottom' ,rotation=90)

# Show the plot
plt.show()
# 'top', 'bottom', 'center', 'baseline', 'center_baseline'
```



### 7. What is the count and percentage distribution of Clean Alternative Fuel Vehicles (CAFVs)?

[140]:
```
df.columns
```

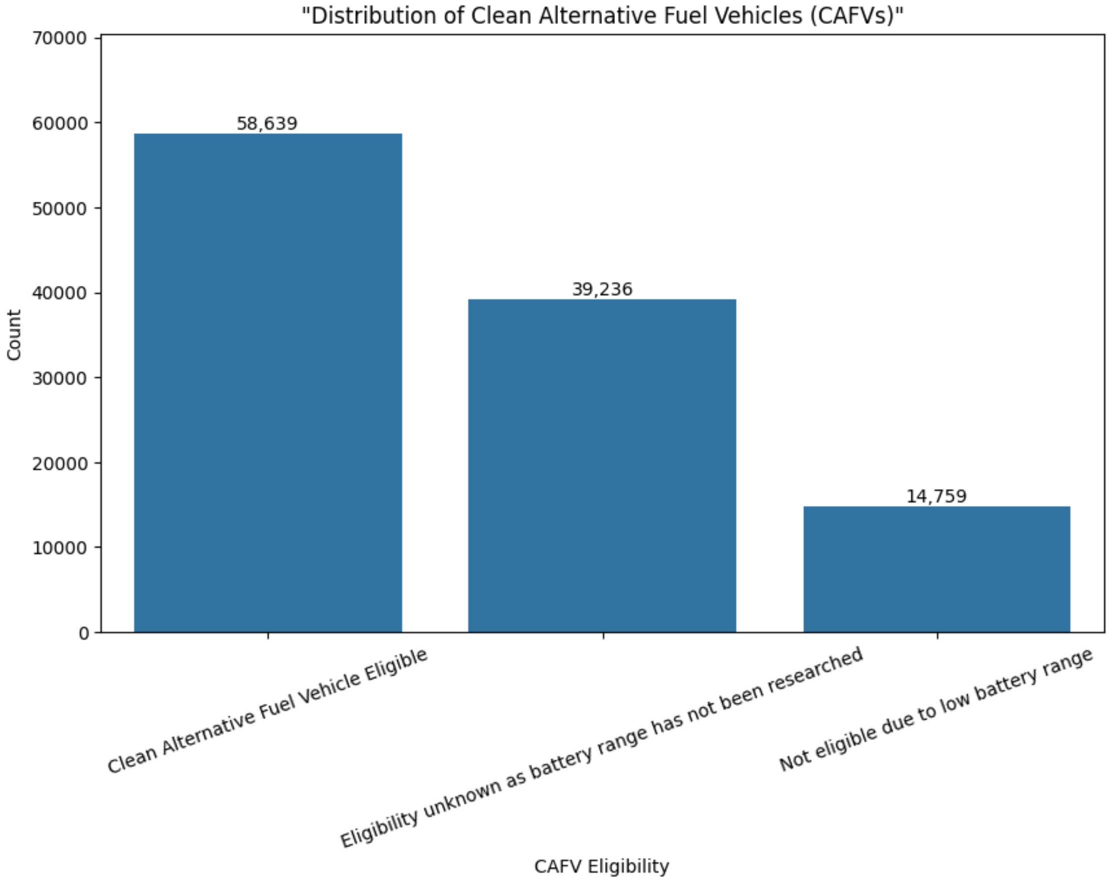
[140]:
```
Index(['VIN (1-10)', 'County', 'City', 'State', 'Postal Code', 'Model Year',
       'Make', 'Model', 'Electric Vehicle Type',
       'Clean Alternative Fuel Vehicle (CAFV) Eligibility', 'Electric Range',
       'Base MSRP', 'Legislative District', 'DOL Vehicle ID',
       'Vehicle Location', 'Electric Utility', '2020 Census Tract',
       'Make_Model'],
      dtype='object')
```

[141]:
```
percentage_distribution_CAFV = df['Clean Alternative Fuel Vehicle (CAFV)␣
 ↪Eligibility'].value_counts()
```

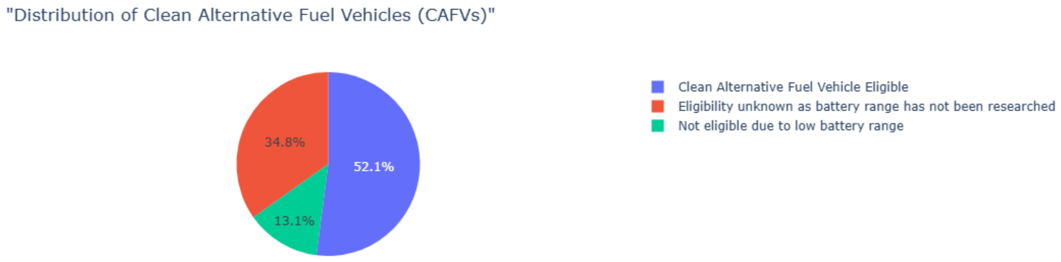
[142]:
```
percentage_distribution_CAFV
```

```
[142]: Clean Alternative Fuel Vehicle (CAFV) Eligibility
       Clean Alternative Fuel Vehicle Eligible                          58639
       Eligibility unknown as battery range has not been researched    39236
       Not eligible due to low battery range                            14759
       Name: count, dtype: int64
```

```
[143]: plt.figure(figsize=(10,6))
       sns.barplot(x=percentage_distribution_CAFV.index,␣
         ↪y=percentage_distribution_CAFV.values)
       plt.title('"Distribution of Clean Alternative Fuel Vehicles (CAFVs)"')
       plt.xlabel('CAFV Eligibility')
       plt.ylabel('Count')
       plt.xticks(rotation =20)
       # Display the count above each bar
       for index, values in enumerate(percentage_distribution_CAFV.values):
           plt.text(index, values, f'{values:,}' , ha='center' ,va='bottom')
       # Increase the range of the y-axis to add some space above the highest bar
       plt.ylim(0, percentage_distribution_CAFV.max() * 1.2)
       plt.show()
```



[144]:
```
fig1 = px.pie(values = percentage_distribution_CAFV.values ,␣
  ↪names=percentage_distribution_CAFV.index, title='"Distribution of Clean␣
  ↪Alternative Fuel Vehicles (CAFVs)"' )
fig1.show()
```



### 8. What is the average range of different makes?

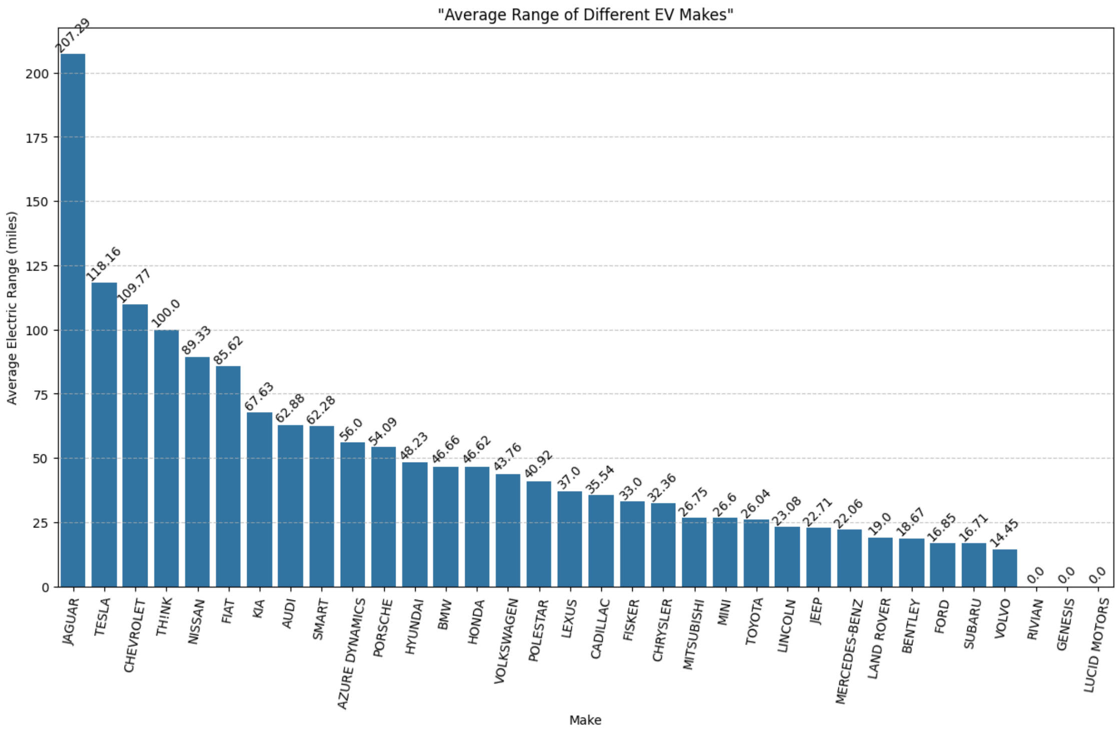
[145]:
```
makewise_range = df.groupby("Make")["Electric Range"].mean().
  ↪sort_values(ascending=False)
makewise_range
```

[145]:
```
Make
JAGUAR            207.287671
TESLA             118.162756
CHEVROLET         109.766549
TH!NK             100.000000
NISSAN             89.326941
FIAT               85.624088
KIA                67.631943
AUDI               62.876930
SMART              62.282051
AZURE DYNAMICS     56.000000
PORSCHE            54.090465
HYUNDAI            48.228754
BMW                46.657479
HONDA              46.618687
VOLKSWAGEN         43.762530
POLESTAR           40.921147
LEXUS              37.000000
CADILLAC           35.537037
FISKER             33.000000
CHRYSLER           32.361204
MITSUBISHI         26.746599
MINI               26.604430
```

```
TOYOTA             26.044268
LINCOLN            23.083333
JEEP               22.707465
MERCEDES-BENZ      22.055336
LAND ROVER         19.000000
BENTLEY            18.666667
FORD               16.848084
SUBARU             16.711864
VOLVO              14.448864
RIVIAN              0.000000
GENESIS             0.000000
LUCID MOTORS        0.000000
Name: Electric Range, dtype: float64
```

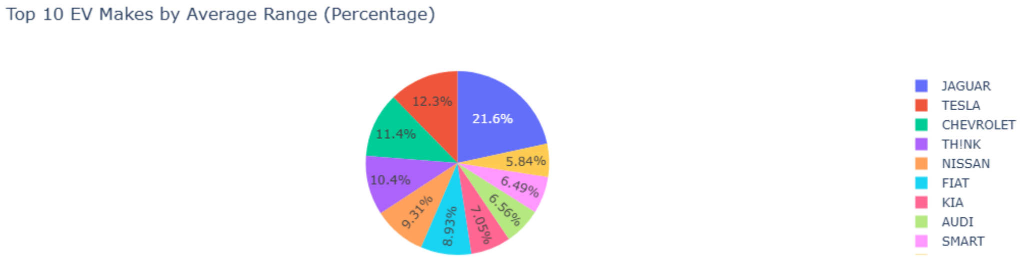
[146]:

```
plt.figure(figsize=(15,8))
sns.barplot(x= makewise_range.index , y=makewise_range.values )
plt.title('"Average Range of Different EV Makes"')
plt.xlabel('Make')
plt.ylabel('Average Electric Range (miles)')
plt.xticks(rotation =80)
plt.grid(axis ='y' , linestyle ='--'  , alpha=0.7)
# Increase the range of the y-axis to add some space above the highest bar
max_average_range_by_make = makewise_range.max()
# Display the count above each line
for index, value in enumerate(np.round(makewise_range.values,2)):
    plt.text(index , value, f'{value:,}', ha='center', va='bottom' ,rotation =␣
 ↪45)
plt.show()
```



[147]:

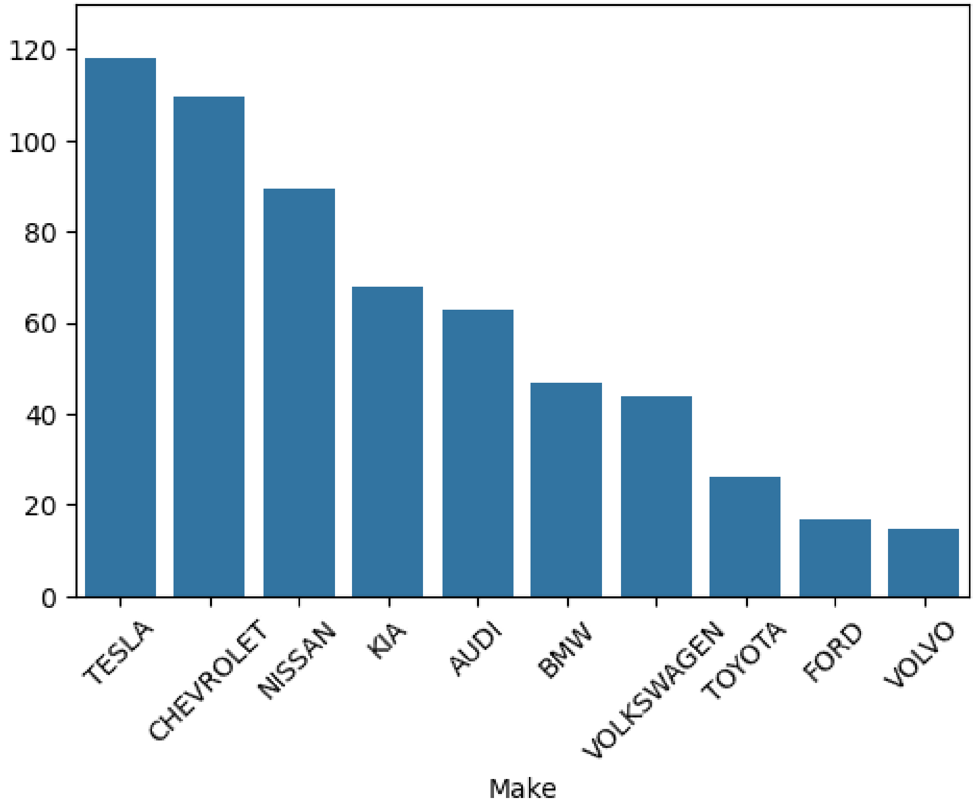
```
average_range_by_make =  df.groupby('Make')['Electric Range'].mean().
 ↪nlargest(10)
fig1 = px.pie(values = average_range_by_make.values ,␣
 ↪names=average_range_by_make.index, title="Top 10 EV Makes by Average Range␣
 ↪(Percentage)" )
fig1.show()
```



[148]:

```
# Step 1: Get the top 10 makes by their count
top_10_make = df["Make"].value_counts().nlargest(10)
```

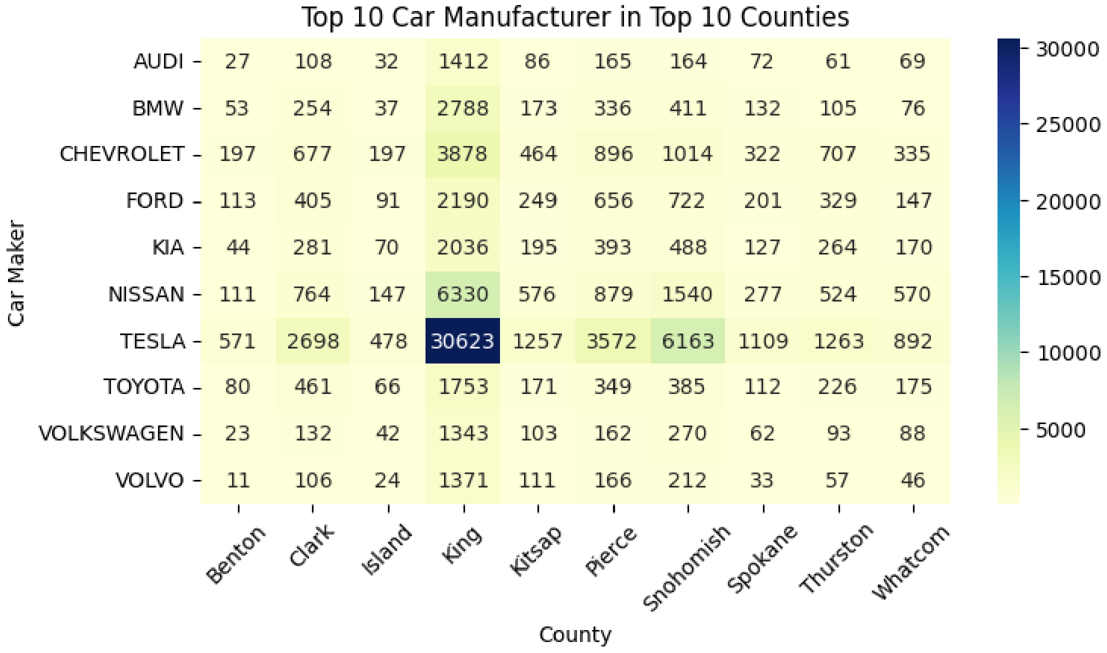
```
# Step 2: Filter the DataFrame to include only the top 10 makes
top_10_make_df = df[df["Make"].isin(top_10_make.index)]

# Step 3: Calculate the average electric range for each of the top 10 makes and␣
 ↪sort in descending order
average_range_by_top_10_make = top_10_make_df.groupby("Make")["Electric Range"].
 ↪mean().sort_values(ascending= False)

# Display the result
plt.figure(figsize=(6,4))
sns.barplot(x=average_range_by_top_10_make.index,␣
 ↪y=average_range_by_top_10_make.values)
# Increase the range of the y-axis to add some space above the highest bar
max1 = average_range_by_top_10_make.max()
plt.ylim(0, max1 *1.1)
plt.xticks(rotation=45)

plt.show()
```



**11. Which are the top 10 makes with the highest number of vehicles in the top 10 counties?**

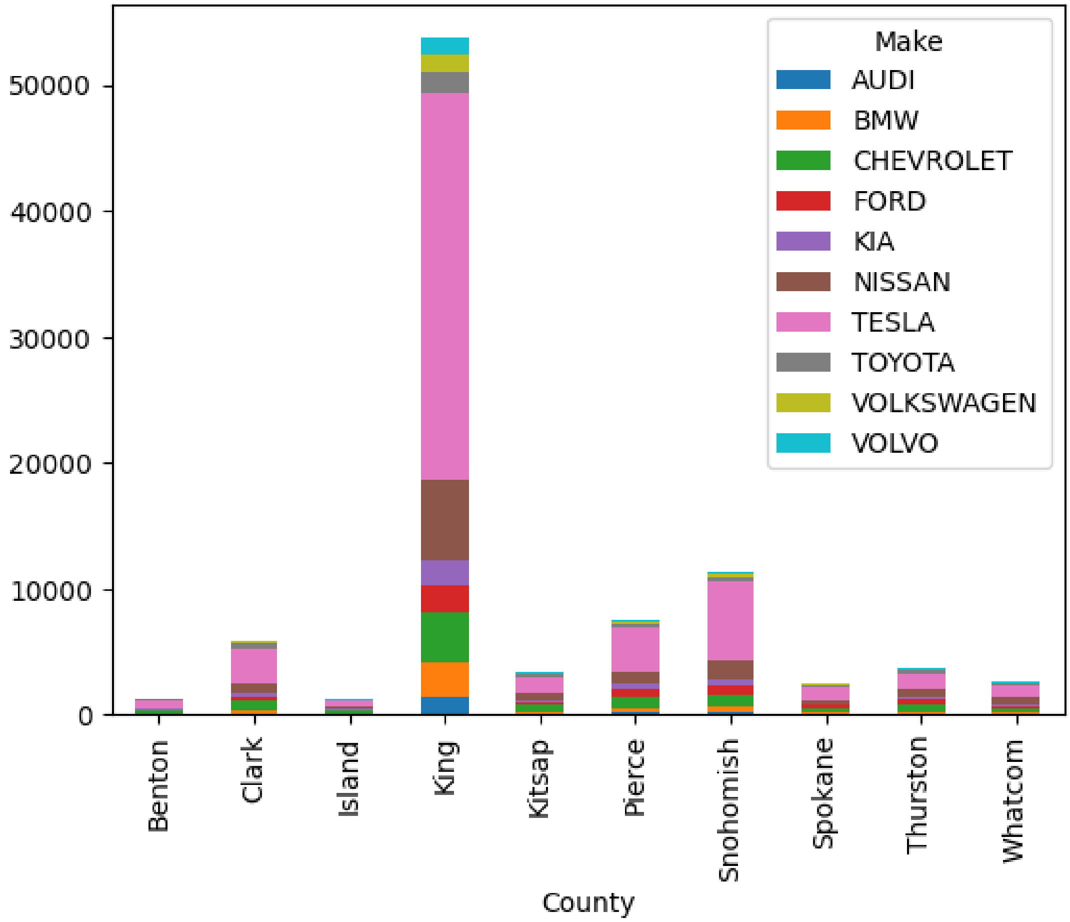
```
[149]: # 1. Get the top 10 counties by occurrence
       Top_10_county = df['County'].value_counts().nlargest(10)
       # 2. Filter the DataFrame to include only the rows where County is in the top 10
       top_10_county_df = df[df['County'].isin(Top_10_county.index)]
       # 3. Get the top 10 Model by occurrence
       top_10_make_df  = top_10_county_df['Make'].value_counts().nlargest(10)
       # 4. Filter the DataFrame to include only the rows where Model is in the top 10
       top_10_county_df = top_10_county_df[top_10_county_df['Make'].
        ↪isin(top_10_make_df.index)]
       top_10_county_df['count1'] =1

       pivot_table = top_10_county_df.pivot_table(
           index='Make',
           columns='County',
           aggfunc='size'
       )
```

```
[150]: plt.figure(figsize=(8, 4))
       sns.heatmap(pivot_table, annot=True , fmt='d' ,cmap='YlGnBu')
       plt.title('Top 10 Car Manufacturer in Top 10 Counties')
       plt.xlabel('County')
       plt.ylabel('Car Maker')
       plt.xticks(rotation=45)  # Rotate x-axis labels for better readability
       plt.yticks(rotation=0)
       plt.show()
```



```
[151]: pivot_table = top_10_county_df.pivot_table(
           index='County',
           columns='Make',
           aggfunc='size'
       )
       pivot_table
       pivot_table.plot(kind='bar', stacked=True)
```

```
[151]: <Axes: xlabel='County'>
```



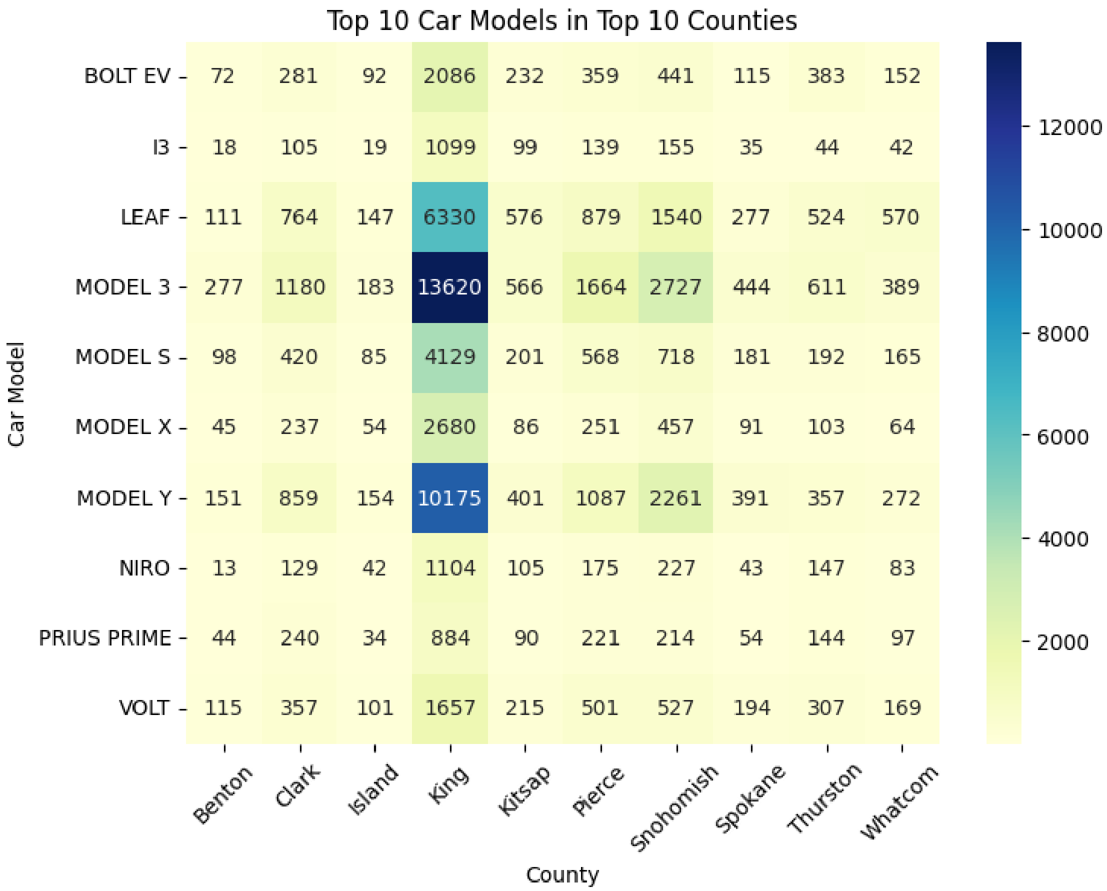
**12. Which are the top 10 models with the highest number of vehicles in the top 10 counties?**

```
[152]: # 1. Get the top 10 counties by occurrence
       Top_10_county = df['County'].value_counts().nlargest(10)
       # 2. Filter the DataFrame to include only the rows where County is in the top 10
```

```
top_10_county_df = df[df['County'].isin(Top_10_county.index)]
# 3. Get the top 10 Model by occurrence
top_10_model_df  = top_10_county_df['Model'].value_counts().nlargest(10)
# 4. Filter the DataFrame to include only the rows where Model is in the top 10
top_10_county_df = top_10_county_df[top_10_county_df['Model'].
 ↪isin(top_10_model_df.index)]
top_10_county_df['count1'] =1

pivot_table = top_10_county_df.pivot_table(
    index='Model',
    columns='County',
    aggfunc='size'
)
```

[153]:

```
plt.figure(figsize=(8, 6))
sns.heatmap(pivot_table, annot=True , fmt='d' ,cmap='YlGnBu')
plt.title('Top 10 Car Models in Top 10 Counties')
plt.xlabel('County')
plt.ylabel('Car Model')
plt.xticks(rotation=45)  # Rotate x-axis labels for better readability
plt.yticks(rotation=0)
plt.show()
```



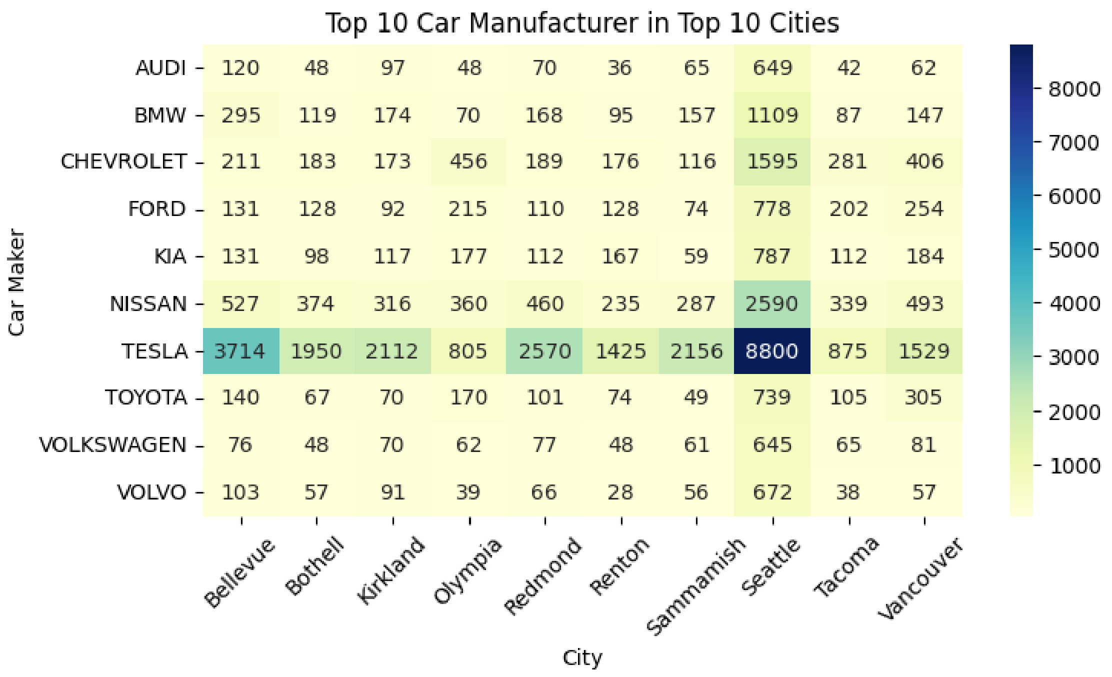
### 13. Which are the top 10 makes with the highest number of vehicles in the top 10 cities?

[154]:

```
# 1. Get the top 10 counties by occurrence
Top_10_city = df['City'].value_counts().nlargest(10)
# 2. Filter the DataFrame to include only the rows where County is in the top 10
top_10_city_df = df[df['City'].isin(Top_10_city.index)]
# 3. Get the top 10 Model by occurrence
top_10_make_df  = top_10_city_df['Make'].value_counts().nlargest(10)
# 4. Filter the DataFrame to include only the rows where Model is in the top 10
top_10_city_df = top_10_city_df[top_10_city_df['Make'].isin(top_10_make_df.
 ↪index)]
top_10_city_df['count1'] =1

pivot_table = top_10_city_df.pivot_table(
    index='Make',
    columns='City',
    aggfunc='size'
)
```

[155]:

```
plt.figure(figsize=(8, 4))
sns.heatmap(pivot_table, annot=True , fmt='d' ,cmap='YlGnBu')
plt.title('Top 10 Car Manufacturer in Top 10 Cities')
plt.xlabel('City')
plt.ylabel('Car Maker')
plt.xticks(rotation=45)  # Rotate x-axis labels for better readability
plt.yticks(rotation=0)
plt.show()
```



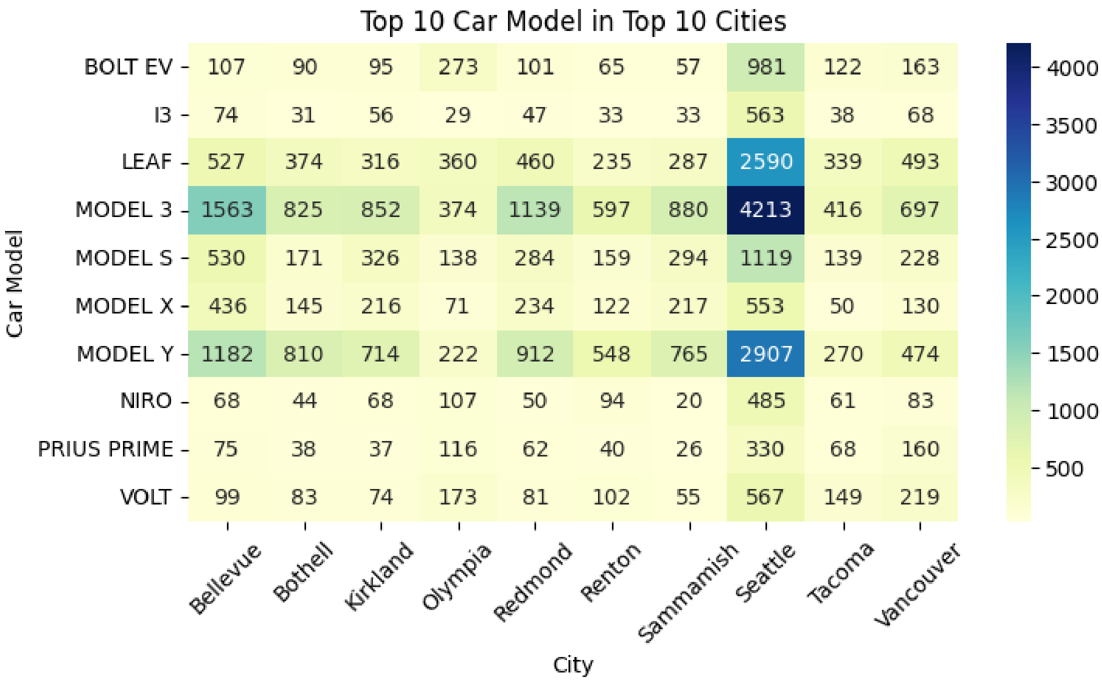
**14. Which are the top 10 models with the highest number of vehicles in the top 10 cities?**

[156]:

```
# 1. Get the top 10 counties by occurrence
Top_10_city = df['City'].value_counts().nlargest(10)
# 2. Filter the DataFrame to include only the rows where County is in the top 10
top_10_city_df = df[df['City'].isin(Top_10_city.index)]
# 3. Get the top 10 Model by occurrence
top_10_make_df  = top_10_city_df['Model'].value_counts().nlargest(10)
# 4. Filter the DataFrame to include only the rows where Model is in the top 10
top_10_city_df = top_10_city_df[top_10_city_df['Model'].isin(top_10_make_df.
 ↪index)]
top_10_city_df['count1'] =1

pivot_table = top_10_city_df.pivot_table(
    index='Model',
    columns='City',
    aggfunc='size'
)
pivot_table
```

[156]:
```
City          Bellevue  Bothell  Kirkland  Olympia  Redmond  Renton  Sammamish  \
Model
BOLT EV            107       90        95      273      101      65         57
I3                  74       31        56       29       47      33         33
LEAF               527      374       316      360      460     235        287
MODEL 3           1563      825       852      374     1139     597        880
MODEL S            530      171       326      138      284     159        294
MODEL X            436      145       216       71      234     122        217
MODEL Y           1182      810       714      222      912     548        765
NIRO                68       44        68      107       50      94         20
PRIUS PRIME         75       38        37      116       62      40         26
VOLT                99       83        74      173       81     102         55

City          Seattle  Tacoma  Vancouver
Model
BOLT EV           981     122        163
I3                563      38         68
LEAF             2590     339        493
MODEL 3          4213     416        697
MODEL S          1119     139        228
MODEL X           553      50        130
MODEL Y          2907     270        474
NIRO              485      61         83
PRIUS PRIME       330      68        160
VOLT              567     149        219
```

[157]:
```
plt.figure(figsize=(8, 4))
sns.heatmap(pivot_table, annot=True , fmt='d' ,cmap='YlGnBu')
plt.title('Top 10 Car Model in Top 10 Cities')
plt.xlabel('City')
plt.ylabel('Car Model')
plt.xticks(rotation=45)  # Rotate x-axis labels for better readability
plt.yticks(rotation=0)
plt.show()
```



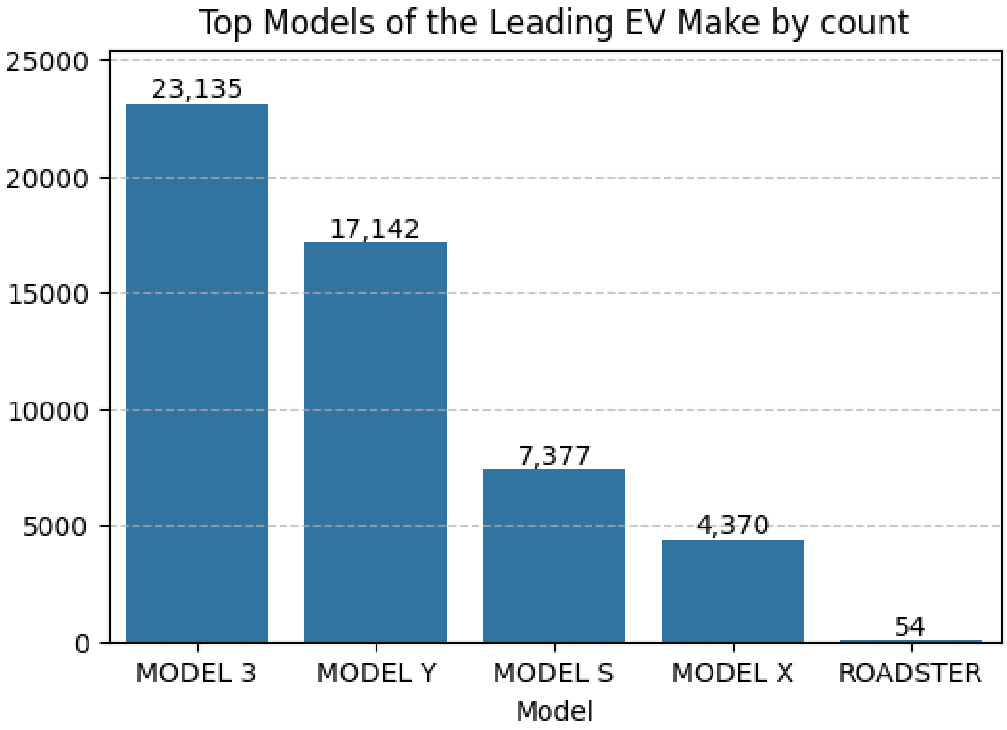
Top 10 Car Model in Top 10 Cities

| Car Model | Bellevue | Bothell | Kirkland | Olympia | Redmond | Renton | Sammamish | Seattle | Tacoma | Vancouver |
|---|---|---|---|---|---|---|---|---|---|---|
| BOLT EV | 107 | 90 | 95 | 273 | 101 | 65 | 57 | 981 | 122 | 163 |
| I3 | 74 | 31 | 56 | 29 | 47 | 33 | 33 | 563 | 38 | 68 |
| LEAF | 527 | 374 | 316 | 360 | 460 | 235 | 287 | 2590 | 339 | 493 |
| MODEL 3 | 1563 | 825 | 852 | 374 | 1139 | 597 | 880 | 4213 | 416 | 697 |
| MODEL S | 530 | 171 | 326 | 138 | 284 | 159 | 294 | 1119 | 139 | 228 |
| MODEL X | 436 | 145 | 216 | 71 | 234 | 122 | 217 | 553 | 50 | 130 |
| MODEL Y | 1182 | 810 | 714 | 222 | 912 | 548 | 765 | 2907 | 270 | 474 |
| NIRO | 68 | 44 | 68 | 107 | 50 | 94 | 20 | 485 | 61 | 83 |
| PRIUS PRIME | 75 | 38 | 37 | 116 | 62 | 40 | 26 | 330 | 68 | 160 |
| VOLT | 99 | 83 | 74 | 173 | 81 | 102 | 55 | 567 | 149 | 219 |

### 15. What are the top models of the top 1 make by count?

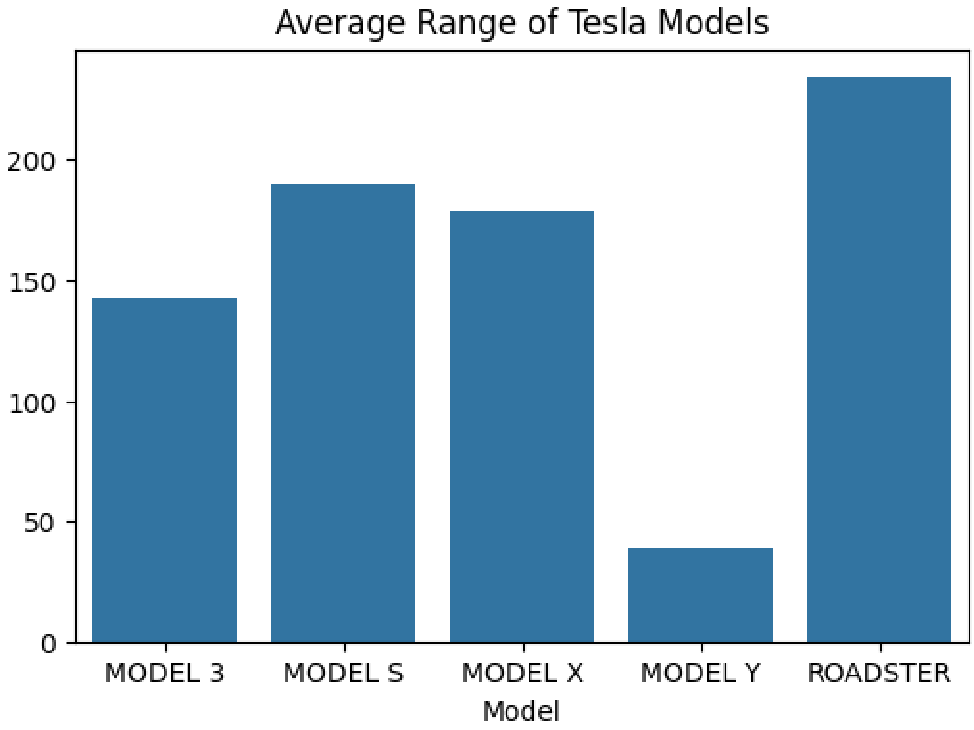
[158]:

```
Tesla = df['Make'].value_counts().nlargest(1)
Tesla = df[ df['Make'].isin(Tesla.index) ]
Tesla_models = Tesla['Model'].value_counts()
# Increase the range of the y-axis to add some space above the highest bar
max1= Tesla_models.max()
plt.figure(figsize=(6,4))
sns.barplot(x=Tesla_models.index , y=Tesla_models.values)
plt.title('Top Models of the Leading EV Make by count')
# plt.grid(True)
plt.grid(axis ='y' , linestyle='--', alpha=0.7)
plt.ylim(0, max1 * 1.1 )
# Display the count above each bar
for index, value in enumerate(Tesla_models.values):
    plt.text(index, value, f'{value:,}' , ha='center' , va='bottom')
```



**16. Tesla Models Electric Range**

```
[159]: plt.figure(figsize=(6,4))
average_range_tesla_model = Tesla.groupby('Model')['Electric Range'].mean()
sns.barplot(x=average_range_tesla_model.index , y=average_range_tesla_model.
  ↪values)
plt.title('Average Range of Tesla Models')
plt.show()
```



**BiVariate Analysis**

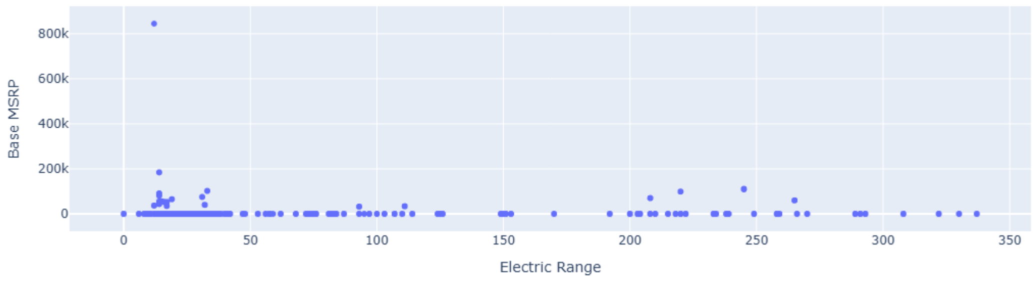
[160]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 112634 entries, 0 to 112633
Data columns (total 18 columns):
 #   Column                                             Non-Null Count   Dtype
---  ------                                             --------------   -----
 0   VIN (1-10)                                         112634 non-null  object
 1   County                                             112634 non-null  object
 2   City                                               112634 non-null  object
 3   State                                              112634 non-null  object
 4   Postal Code                                        112634 non-null  int64
 5   Model Year                                         112634 non-null  int64
 6   Make                                               112634 non-null  object
 7   Model                                              112634 non-null  object
 8   Electric Vehicle Type                              112634 non-null  object
 9   Clean Alternative Fuel Vehicle (CAFV) Eligibility  112634 non-null  object
 10  Electric Range                                     112634 non-null  int64
 11  Base MSRP                                          112634 non-null  int64
 12  Legislative District                               112634 non-null  float64
 13  DOL Vehicle ID                                     112634 non-null  int64
 14  Vehicle Location                                   112634 non-null  object
 15  Electric Utility                                    112634 non-null  object
 16  2020 Census Tract                                   112634 non-null  int64
 17  Make_Model                                          112634 non-null  object
dtypes: float64(1), int64(6), object(11)
memory usage: 15.5+ MB
```

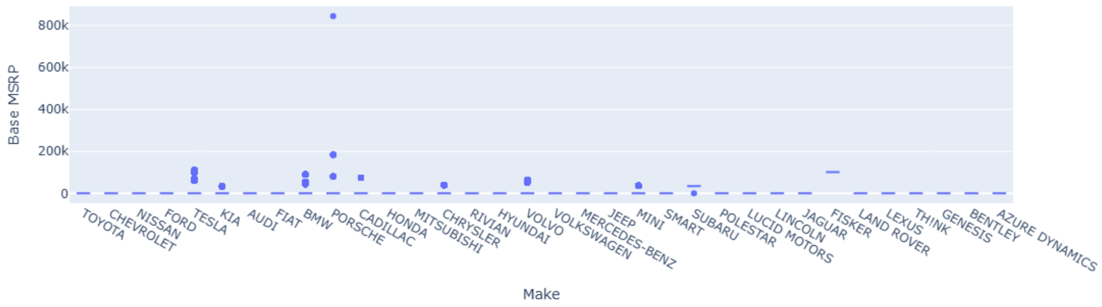
[161]:
```
df.columns
```

[161]:
```
Index(['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',
       'Make_Model'],
      dtype='object')
```

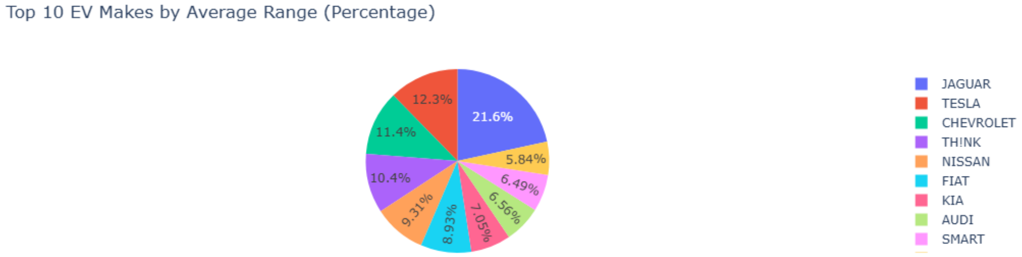
[162]:
```
px.scatter(df,x="Electric Range",y="Base MSRP")
```



[163]:
```
px.box(df,x="Make",y="Base MSRP")
```

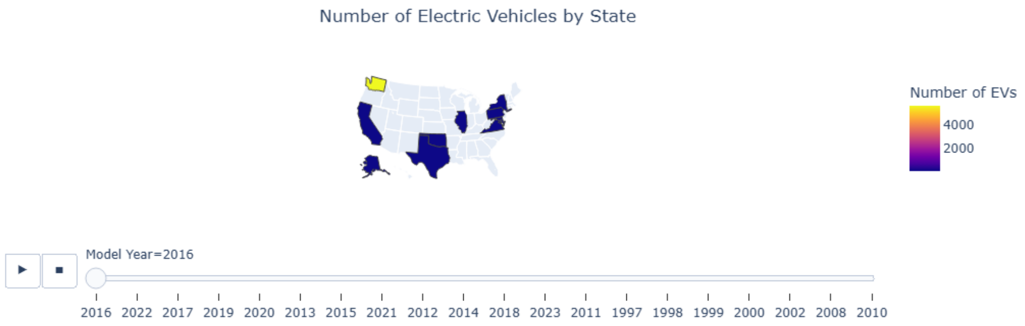


```
[164]: average_range_by_make =  df.groupby('Make')['Electric Range'].mean().
        ↪nlargest(10)
       fig1 = px.pie(values = average_range_by_make.values ,␣
        ↪names=average_range_by_make.index, title="Top 10 EV Makes by Average Range␣
        ↪(Percentage)" )
       fig1.show()
```



```
[165]: ev_count_by_state = df.groupby(['State', 'Model Year']).size().
        ↪reset_index(name='EV_Count')
```

```
[172]: fig = px.choropleth(
           ev_count_by_state,
           locations='State',
           locationmode='USA-states',
           color='EV_Count',
           scope='usa',
           labels={'EV_Count': 'Number of EVs'},
           title='Number of Electric Vehicles by State',
           animation_frame = "Model Year"
       )

       fig.update_layout(
           title_x=0.5,
       )

       fig.show()
```



[167]:
```
final = df.groupby(['Model Year', 'Make']).size().unstack(fill_value=0)
```

[168]:
```
final
```

[168]:
```
Make        AUDI  AZURE DYNAMICS  BENTLEY  BMW  CADILLAC  CHEVROLET  CHRYSLER  \
Model Year
1997           0               0        0    0         0          1         0
1998           0               0        0    0         0          0         0
1999           0               0        0    0         0          0         0
2000           0               0        0    0         0          0         0
2002           0               0        0    0         0          0         0
2008           0               0        0    0         0          0         0
2010           0               0        0    0         0          0         0
2011           0               4        0    0         0         71         0
2012           0               3        0    0         0        496         0
2013           0               0        0    0         0        818         0
2014           0               0        0  457        58        724         0
2015           0               0        0  403         0        467         0
2016         214               0        0  383        18        309         0
2017         187               0        0  692        15       2744        94
2018         174               0        0  710        17       1126       554
2019         392               0        0  279         0        966       119
2020         224               0        1  143         0       1014       286
2021         544               0        1  635         0        377       504
2022         585               0        1  905         0        892       237
2023          12               0        0   73         0        177         0

Make        FIAT  FISKER  FORD  …  POLESTAR  PORSCHE  RIVIAN  SMART  SUBARU  \
Model Year                      …
1997           0       0     0  …         0        0       0      0       0
1998           0       0     1  …         0        0       0      0       0
1999           0       0     3  …         0        0       0      0       0
2000           0       0    10  …         0        0       0      0       0
2002           0       0     0  …         0        0       0      0       0
```

```
2008           0       0     0  ...       0        0       0      0       0
2010           0       0     0  ...       0        0       0      0       0
2011           0       0     0  ...       0        0       0      0       0
2012           0      20    15  ...       0        0       0      0       0
2013         106       0   662  ...       0        0       0     29       0
2014          97       0   628  ...       0        8       0     71       0
2015         242       0   556  ...       0       33       0     52       0
2016         148       0   778  ...       0       78       0     31       0
2017         167       0   659  ...       0       71       0     34       0
2018          53       0   170  ...       0       78       0     47       0
2019           9       0   136  ...       0       62       0      9      58
2020           0       0    65  ...       0      143       0      0       0
2021           0       0   882  ...      98      145       0      0       0
2022           0       0  1250  ...     372      200     885      0       0
2023           0       0     4  ...      88        0       0      0       1

Make        TESLA  TH!NK  TOYOTA  VOLKSWAGEN  VOLVO
Model Year
1997            0      0       0           0      0
1998            0      0       0           0      0
1999            0      0       0           0      0
2000            0      0       0           0      0
2002            0      0       2           0      0
2008           23      0       0           0      0
2010           24      0       0           0      0
2011            7      3       0           0      0
2012          134      0     385           0      0
2013          814      0     296           0      0
2014          683      0     215           0      0
2015         1089      0      89         124      0
2016         1639      0       0         319    112
2017         1679      0     899          52    115
2018         7919      0     659          39    226
2019         4583      0     190         495    190
2020         7018      0     311           0    162
2021        11028      0     935         988    580
2022        14548      0     424         428    882
2023          890      0       0          69     21

[20 rows x 34 columns]
```

[169]:

```
bcr.bar_chart_race(
    df=final,
    title='EV Make and its Count Each Year',
    orientation='h',
    sort='desc',
    n_bars=10,
    steps_per_period=40,
    period_length=2000,
    bar_size=0.95,
    title_size=24,
    period_label={'x': .95, 'y': .25, 'fontsize': 12},
    perpendicular_bar_func='median',
    period_summary_func=lambda v, r: {'x': .2, 'y': .9, 's': f'Total EVs: {v.
 ↪sum():,.0f}', 'ha': 'center', 'size': 12},
    figsize=(6, 4),
    dpi=144,
    cmap='tab20'
)
```

```
[169]: <IPython.core.display.HTML object>
```

### 0.0.8 Conclusion of the Complete EDA

Based on the exploratory data analysis (EDA) of the electric vehicle (EV) dataset, here are some key conclusions:

1. **Growth in EV Adoption**: There has been a steady increase in the number of electric vehicles registered each year, particularly for **Battery Electric Vehicles (BEVs)**. This indicates a growing acceptance of EVs as a sustainable transportation solution.

2. **Geographic Distribution**: Certain states and cities, such as those in **California, Washington, and Florida**, have a significantly higher number of EV registrations. This could be attributed to supportive local policies, greater availability of charging infrastructure, and environmental awareness.

3. **Popular EV Makes and Models**: Major automakers like **Tesla, Nissan, and Chevrolet** are dominant in the electric vehicle market, with models such as the **Tesla Model 3** and **Nissan Leaf** showing widespread adoption. Plug-in hybrids, such as the **Chevrolet Volt**, are also prevalent.

4. **Electric Range Variations**: The **electric range** of vehicles varies significantly across models. BEVs typically offer higher ranges (e.g., **Tesla Model S**), while plug-in hybrids have a lower range, often below 50 miles. This impacts the eligibility for clean alternative fuel incentives, where longer-range vehicles are more likely to qualify.

5. **Price Insights**: The **Base MSRP** (Manufacturer's Suggested Retail Price) varies widely, with luxury brands like Tesla having a higher base price compared to other manufacturers like Nissan and Chevrolet, which offer more affordable EV models.

6. **CAFV Eligibility**: A significant portion of the vehicles are **Clean Alternative Fuel Vehicle (CAFV) eligible**, which suggests that many of these EVs are contributing to environmental goals by utilizing cleaner energy sources.

7. **Utility Providers**: The dataset provides insight into how different **electric utility companies** support EV infrastructure in different areas, with utilities in EV-heavy states likely being more involved in supporting charging networks.

Overall, the EDA reveals a positive trend in EV adoption, with certain regions and manufacturers leading the market. The data also highlights the importance of vehicle range and pricing in shaping EV adoption patterns. It will overall give a better idea to the industries with respect to future aspects.

ions.

[ ]:

```
```