Market Size Analysis - Electric Vehicles in the USA

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

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
```

```
ev=pd.read_csv('/content/drive/MyDrive/Electric_Vehicle_Population_Data.csv')
print(ev)
```

```
177862     HYUNDAI  KONA ELECTRIC  Battery Electric Vehicle (BEV)
177863       TESLA        MODEL Y  Battery Electric Vehicle (BEV)
177864  VOLKSWAGEN           ID.4  Battery Electric Vehicle (BEV)
177865       TESLA        MODEL 3  Battery Electric Vehicle (BEV)

        Clean Alternative Fuel Vehicle (CAFV) Eligibility  Electric Range  \
0                 Clean Alternative Fuel Vehicle Eligible             291
1       Eligibility unknown as battery range has not b...               0
2                 Clean Alternative Fuel Vehicle Eligible             270
3                 Clean Alternative Fuel Vehicle Eligible             210
4       Eligibility unknown as battery range has not b...               0
...                                                   ...             ...
177861  Eligibility unknown as battery range has not b...               0
177862  Eligibility unknown as battery range has not b...               0
177863  Eligibility unknown as battery range has not b...               0
177864  Eligibility unknown as battery range has not b...               0
177865  Eligibility unknown as battery range has not b...               0

        Base MSRP  Legislative District  DOL Vehicle ID  \
0               0                  37.0       125701579
1               0                   1.0       244285107
2               0                  36.0       156773144
3               0                   5.0       165103011
4               0                  23.0       205138552
...           ...                   ...             ...
177861          0                  31.0       195224452
177862          0                  35.0       228454180
177863          0                  13.0       168797219
177864          0                   5.0       182448801
177865          0                  27.0       211464683

                      Vehicle Location  \
0          POINT (-122.30839 47.610365)
1         POINT (-122.179458 47.802589)
2          POINT (-122.34848 47.632405)
3          POINT (-122.03646 47.534065)
4          POINT (-122.55717 47.733415)
...                                ...
177861     POINT (-122.183805 47.18062)
177862    POINT (-123.105305 47.211085)
177863  POINT (-119.8493873 47.2339933)
177864     POINT (-122.00451 47.312185)
177865      POINT (-122.38578 47.28971)

                                         Electric Utility  2020 Census Tract
0           CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)       5.303301e+10
1                                 PUGET SOUND ENERGY INC       5.306105e+10
2           CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)       5.303301e+10
3         PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)       5.303303e+10
4                                 PUGET SOUND ENERGY INC       5.303594e+10
...                                                   ...                ...
177861    PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)       5.305307e+10
177862  BONNEVILLE POWER ADMINISTRATION||CITY OF TACOM...       5.304596e+10
177863                          PUD NO 2 OF GRANT COUNTY       5.302501e+10
177864    PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)       5.303303e+10
177865  BONNEVILLE POWER ADMINISTRATION||CITY OF TACOM...       5.305394e+10

[177866 rows x 17 columns]
```

```
ev.head(10)
```

	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
0	5YJYGDEE1L	King	Seattle	WA	98122.0	2020	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	291	0	37.0
1	7SAYGDEE9P	Snohomish	Bothell	WA	98021.0	2023	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0	0	1.0
2	5YJSA1E4XK	King	Seattle	WA	98109.0	2019	TESLA	MODEL S	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	270	0	36.0
3	5YJSA1E27G	King	Issaquah	WA	98027.0	2016	TESLA	MODEL S	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	210	0	5.0
4	5YJYGDEE5M	Kitsap	Suquamish	WA	98392.0	2021	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0	0	23.0
5	3FA6P0SU8H	Thurston	Yelm	WA	98597.0	2017	FORD	FUSION	Plug-in Hybrid Electric Vehicle (PHEV)	Not eligible due to low battery range	21	0	2.0
6	1N4AZ0CP2D	Yakima	Yakima	WA	98903.0	2013	NISSAN	LEAF	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	75	0	14.0
7	KNAGV4LD9J	Snohomish	Bothell	WA	98012.0	2018	KIA	OPTIMA	Plug-in Hybrid Electric Vehicle (PHEV)	Not eligible due to low battery range	29	0	1.0
8	1N4AZ0CP8F	Kitsap	Port Orchard	WA	98366.0	2015	NISSAN	LEAF	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	84	0	26.0
9	5UXTA6C03N	King	Auburn	WA	98001.0	2022	BMW	X5	Plug-in Hybrid Electric Vehicle (PHEV)	Clean Alternative Fuel Vehicle Eligible	30	0	47.0

```
ev.dtypes
```

```
VIN (1-10)                                           object
County                                               object
City                                                 object
State                                                object
Postal Code                                         float64
Model Year                                            int64
Make                                                 object
Model                                                object
Electric Vehicle Type                                object
Clean Alternative Fuel Vehicle (CAFV) Eligibility    object
Electric Range                                        int64
Base MSRP                                             int64
Legislative District                                float64
DOL Vehicle ID                                        int64
Vehicle Location                                     object
Electric Utility                                     object
2020 Census Tract                                   float64
dtype: object
```

```
ev.isnull().sum()
```

```
VIN (1-10)                                            0
County                                                5
City                                                  5
State                                                 0
Postal Code                                           5
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                                389
DOL Vehicle ID                                        0
Vehicle Location                                      9
Electric Utility                                      5
2020 Census Tract                                     5
dtype: int64
```

```
ev.duplicated().any()
```

```
False
```

```
ev.dropna(inplace=True)
```

We can also fill the missing values using 'Unknown' for string columns and '0' or mean values for numerical columns.

```
#For String columns
#ev['Vehicle Type'] = ev['Vehicle Type'].fillna("Unknown")
#For Numerical columns
#ev.fillna(0, inplace=True)
#ev['2020 Census Tracts'] = ev['2020 Census Tracts'].fillna(0)
#ev['2020 Census Tracts']=ev['2020 Census Tracts'].mean()
```

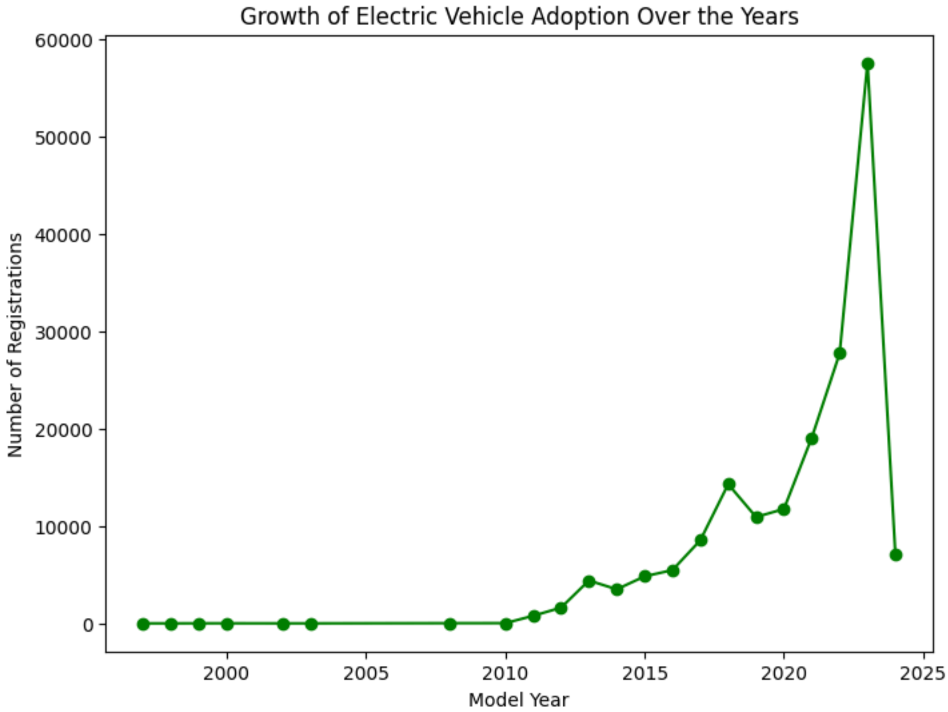
The primary objective of this analysis is to leverage historical EV registration data to understand the current market penetration of EVs, predict future market growth, and identify key trends and factors driving market expansion. The specific goals include:

1. Assess the historical growth trend of EV registrations.
2. Forecast future EV registrations based on historical trends.
3. Analyze the distribution of EV registrations across different models, makes, and geographical regions.
4. Estimate the market size and growth potential of the EV market for upcoming years.
5. Provide insights to support stakeholders in decision-making processes related to production, infrastructure planning, and policy formulation.

1. Market Penetration and Growth:

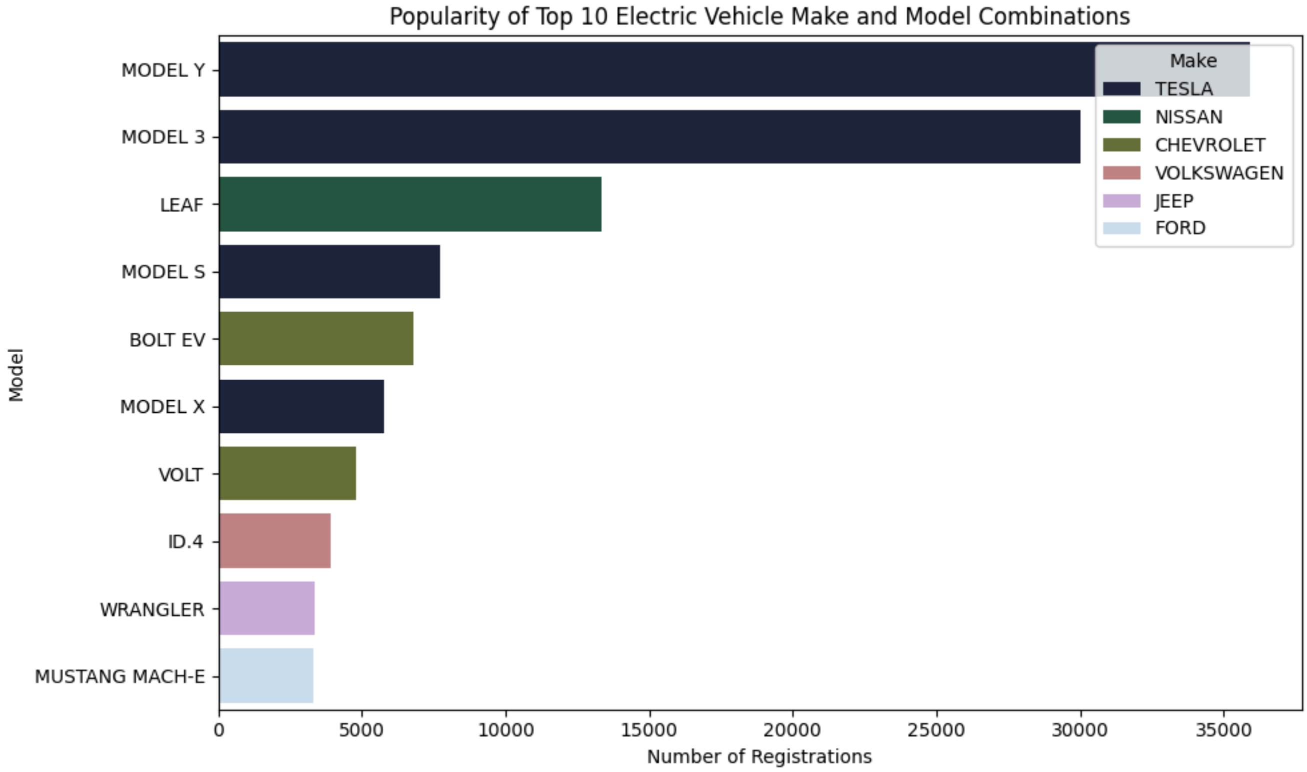
- How has the adoption of electric vehicles grown over the years?
- Which makes and models are most popular among electric vehicles?

```
# Group the data by 'Model Year' and count the number of registrations
ev_by_year = ev.groupby('Model Year')['VIN (1-10)'].count().reset_index(name='Count')

# Plot the trend of EV adoption
plt.figure(figsize=(8, 6))
plt.plot(ev_by_year['Model Year'], ev_by_year['Count'],marker='o', color='green')
plt.xlabel('Model Year')
plt.ylabel('Number of Registrations')
plt.title('Growth of Electric Vehicle Adoption Over the Years')
#plt.grid(True)
plt.show()
```



The growth of electric vehicle adoption over the years are depicted in this graph. Initially, in the year 2000 to 2009, the adoption remained relatively low and stable. After the year 2010 there is a steep increase and suggesting the substantial rise over the years, and there seems a upward trend in the year 2023. However, there is a sudden surge in the upcoming years starting from 2024.

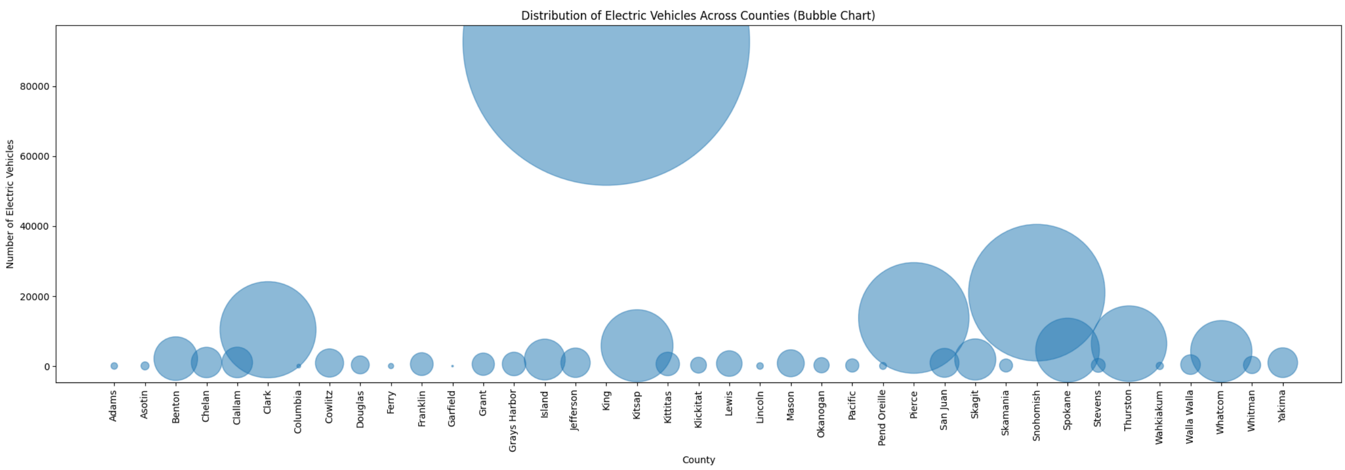
```
# Group by both make and model and count occurrences
popular_make_model = ev.groupby(['Make', 'Model'])['VIN (1-10)'].count().reset_index(name='Count')

# Sort by count in descending order and select top 10
top_10_make_model = popular_make_model.sort_values('Count', ascending=False).head(10)

# Create a bar chart
plt.figure(figsize=(10, 6))
sns.barplot(data=top_10_make_model, x='Count', y='Model', hue='Make', palette="cubehelix")
plt.xlabel('Number of Registrations')
plt.ylabel('Model')
plt.title('Popularity of Top 10 Electric Vehicle Make and Model Combinations')
plt.legend(title='Make', loc='upper right')
plt.tight_layout()
plt.show()
```



2. Geographic Distribution:

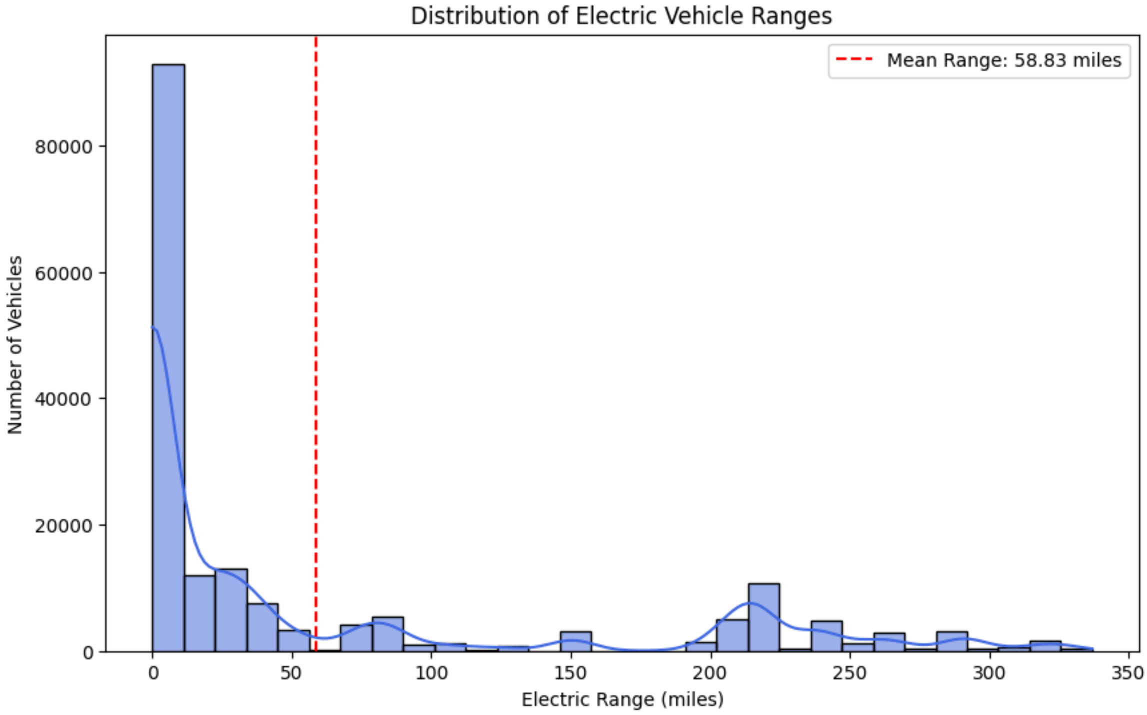
- How are electric vehicles distributed across different regions (counties or cities)?
- Are there specific areas with higher concentrations of electric vehicles?

```
# Group by county and count the number of EVs
ev_by_city = ev.groupby('County')['VIN (1-10)'].count().reset_index(name='Count')

# Create the bubble chart
plt.figure(figsize=(20, 7))
plt.scatter(ev_by_city['County'], ev_by_city['Count'], s=ev_by_city['Count'], alpha=0.5)
plt.xlabel('County')
plt.ylabel('Number of Electric Vehicles')
plt.title('Distribution of Electric Vehicles Across Counties (Bubble Chart)')
plt.xticks(rotation=90)
plt.tight_layout()
plt.show()
```



Seattle, which is in King County, has the highest number of EV registrations by a significant margin, far outpacing the other counties listed.

```
# analyzing the distribution of electric range
plt.figure(figsize=(10, 6))
sns.histplot(ev['Electric Range'], bins=30, kde=True, color='royalblue')
plt.title('Distribution of Electric Vehicle Ranges')
plt.xlabel('Electric Range (miles)')
plt.ylabel('Number of Vehicles')
plt.axvline(ev['Electric Range'].mean(), color='red', linestyle='--', label=f'Mean Range: {ev["Electric Range"].mean():.2f} miles')
plt.legend()
plt.show()
```



1. There is a high frequency of vehicles with a low electric range, with a significant peak occurring just before 50 miles.
2. The distribution is skewed to the right, with a long tail extending towards higher ranges, although the number of vehicles with higher ranges is much less frequent.
3. The mean electric range for this set of vehicles is marked at approximately 58.84 miles, which is relatively low compared to the highest ranges shown in the graph.
4. Despite the presence of electric vehicles with ranges that extend up to around 350 miles, the majority of the vehicles have a range below the mean.

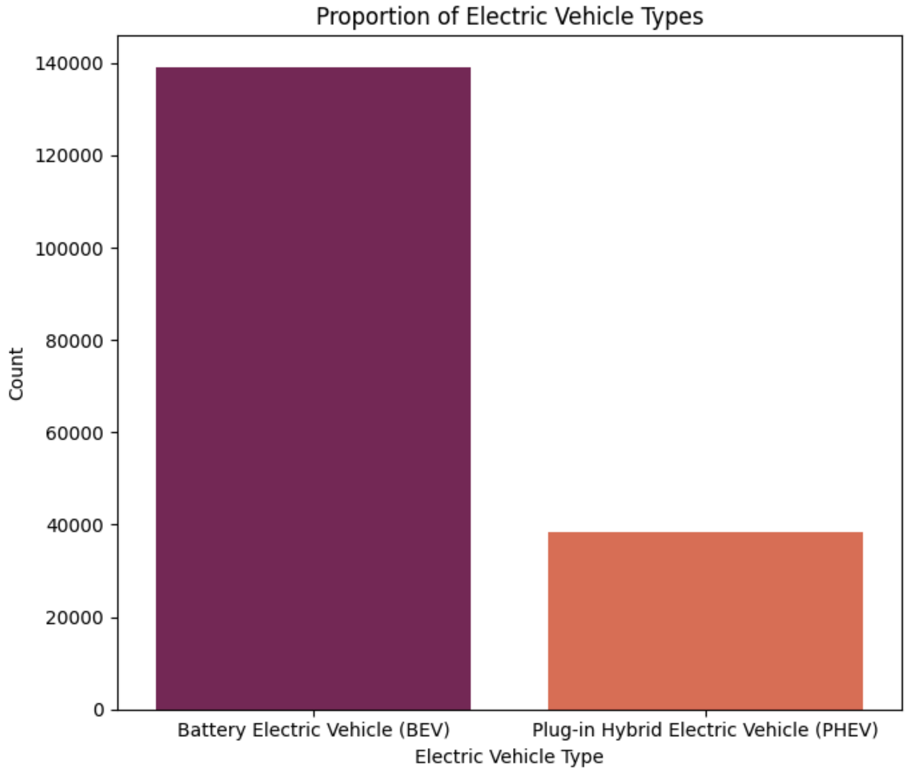
3. Electric Vehicle Types and Eligibility:

- What proportion of the electric vehicles are Battery Electric Vehicles (BEVs) versus Plug-in Hybrid Electric Vehicles (PHEVs)?
- How many vehicles are eligible for Clean Alternative Fuel Vehicle (CAFV) incentives, and how does this impact market adoption?

```
ev_vehicle_type=ev['Electric Vehicle Type'].value_counts()
print(ev_vehicle_type)
```

```
Electric Vehicle Type
Battery Electric Vehicle (BEV)           138947
Plug-in Hybrid Electric Vehicle (PHEV)    38526
Name: count, dtype: int64
```

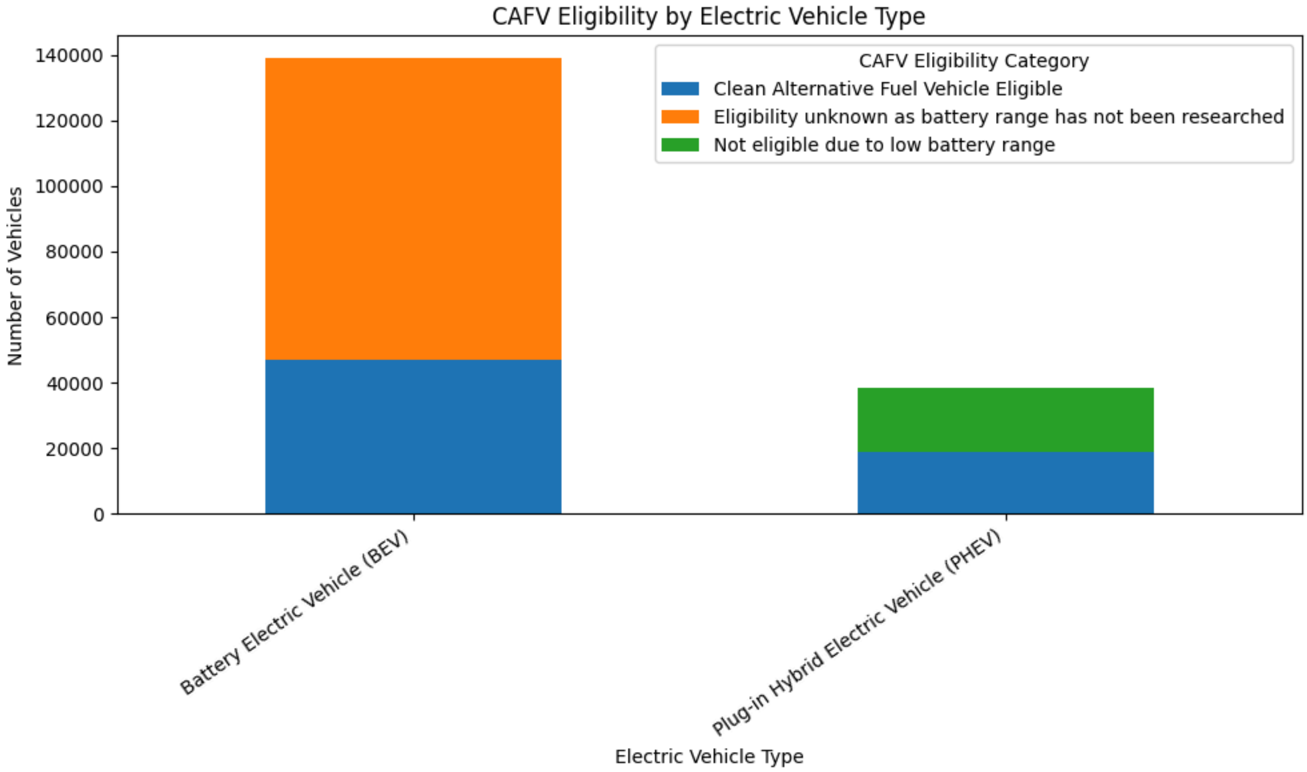
```
plt.figure(figsize=(7, 6))
sns.barplot(x=ev_vehicle_type.index, y=ev_vehicle_type.values, palette="rocket")
plt.xlabel('Electric Vehicle Type')
plt.ylabel('Count')
plt.title('Proportion of Electric Vehicle Types')
plt.tight_layout()
plt.show()
```




```
# Create a cross-tabulation of vehicle type and CAFV eligibility
cafv_by_type = pd.crosstab(ev['Electric Vehicle Type'], ev['Clean Alternative Fuel Vehicle (CAFV) Eligibility'])
# Print the cross-tabulation
print(cafv_by_type)
```

```
Clean Alternative Fuel Vehicle (CAFV) Eligibility  Clean Alternative Fuel Vehicle Eligible  \
Electric Vehicle Type
Battery Electric Vehicle (BEV)                                                       47149
Plug-in Hybrid Electric Vehicle (PHEV)                                               19017

Clean Alternative Fuel Vehicle (CAFV) Eligibility  Eligibility unknown as battery range has not been researched  \
Electric Vehicle Type
Battery Electric Vehicle (BEV)                                                                          91790
Plug-in Hybrid Electric Vehicle (PHEV)                                                                      0

Clean Alternative Fuel Vehicle (CAFV) Eligibility  Not eligible due to low battery range
Electric Vehicle Type
Battery Electric Vehicle (BEV)                                                         8
Plug-in Hybrid Electric Vehicle (PHEV)                                             19509
```

```
# Create a stacked bar chart
cafv_by_type.plot(kind='bar', stacked=True, figsize=(10, 6))
plt.xlabel('Electric Vehicle Type')
plt.ylabel('Number of Vehicles')
plt.title('CAFV Eligibility by Electric Vehicle Type')
plt.legend(title='CAFV Eligibility Category')
plt.xticks(rotation=35, ha='right')
plt.tight_layout()
plt.show()
```

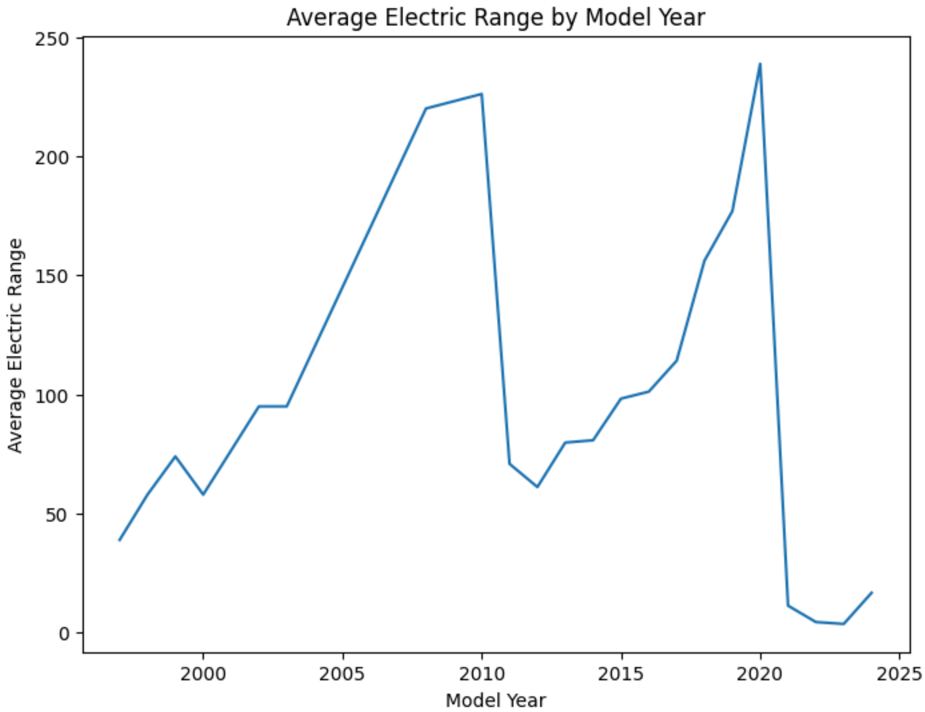


4. Electric Range and Performance:

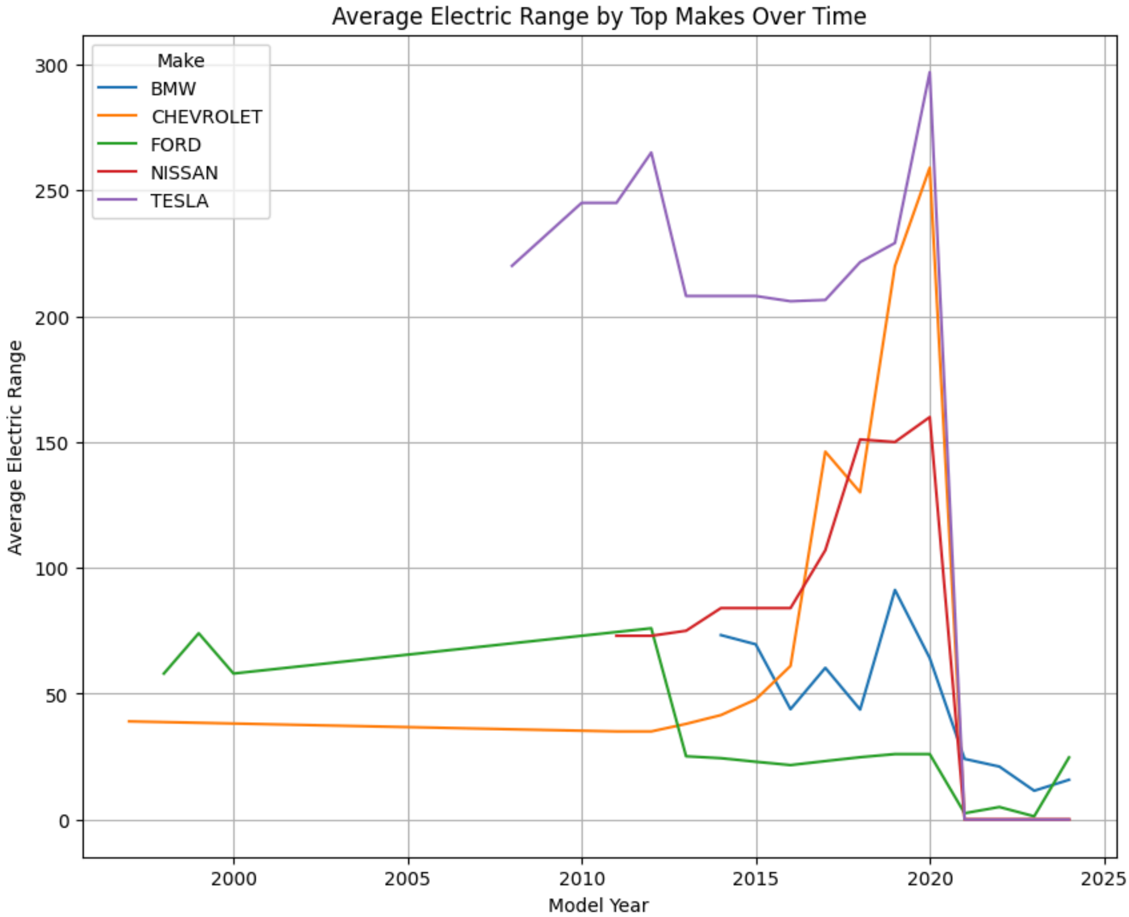
- What is the average electric range of vehicles in the dataset, and how does it vary by model year?
- Are there trends in the electric range of vehicles over time?

```
# Calculate the average electric range of all vehicles
average_range = ev['Electric Range'].mean()

# Print the result
print("Average Electric Range:", average_range)

# Group the data by model year and calculate the average range for each year
average_range_by_year = ev.groupby('Model Year')['Electric Range'].mean()

# Print the result
print(average_range_by_year)

# Plot the average range over model years
plt.figure(figsize=(8, 6))
plt.plot(average_range_by_year.index, average_range_by_year.values)
plt.xlabel('Model Year')
plt.ylabel('Average Electric Range')
plt.title('Average Electric Range by Model Year')
plt.show()
```

```
Average Electric Range: 58.82654826367955
Model Year
1997     39.000000
1998     58.000000
1999     74.000000
2000     58.000000
2002     95.000000
2003     95.000000
2008    220.000000
2010    226.086957
2011     70.891613
2012     61.172243
2013     79.822232
2014     80.798341
2015     98.254869
2016    101.197111
2017    114.162292
2018    156.165967
2019    176.918904
2020    238.748978
2021     11.402665
2022      4.518045
2023      3.729168
2024     16.791431
Name: Electric Range, dtype: float64
```




```
# Group by make and model year, then calculate average range
average_range_by_make_year = ev.groupby(['Make', 'Model Year'])['Electric Range'].mean().reset_index()

# Plot the trends for each make (consider limiting to top makes for clarity)
top_makes = ev['Make'].value_counts().head(5).index  # Select top 5 makes
plt.figure(figsize=(10, 8))
sns.lineplot(data=average_range_by_make_year[average_range_by_make_year['Make'].isin(top_makes)], x='Model Year', y='Electric Range', hu
plt.xlabel('Model Year')
plt.ylabel('Average Electric Range')
plt.title('Average Electric Range by Top Makes Over Time')
plt.grid(True)
plt.show()
```



By comparing the trends across different vehicle types or makes, you can identify which segments are leading the way in electric range improvements. Look for:

Steeper Slopes: Segments with steeper upward trends indicate faster advancements in electric range.

Early Adoption: Segments that started with higher average ranges early on might have a technological advantage.

Convergence: Are the average ranges of different segments converging over time, suggesting a standardization of technology?

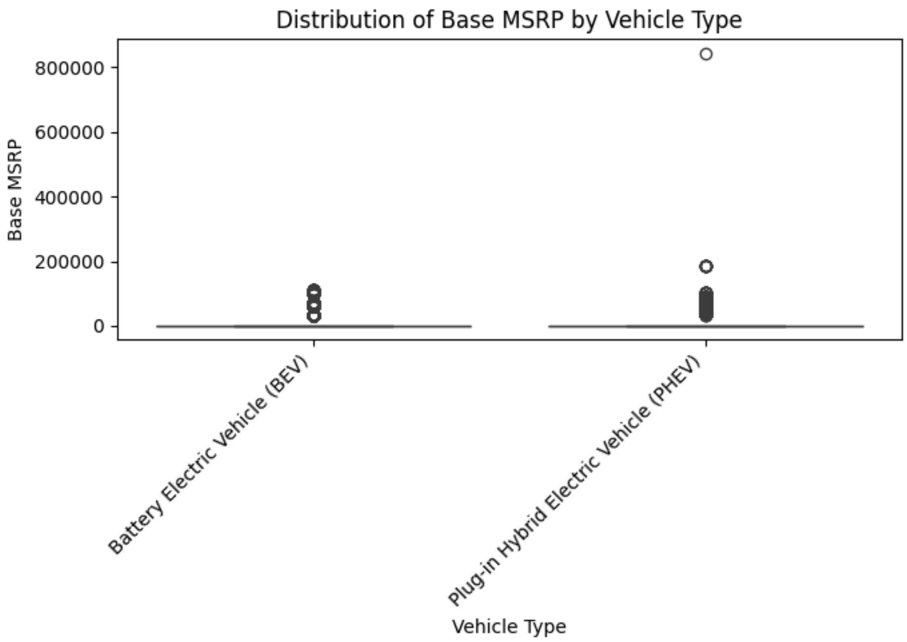
This analysis will provide insights into the competitive landscape and technological advancements within the electric vehicle market.

5. Economic Factors:

- What is the distribution of Base MSRP (Manufacturer's Suggested Retail Price) for different electric vehicles?
- How do the costs of electric vehicles compare across different makes and models?

```
# Calculate basic descriptive statistics for Base MSRP
base_msrp_stats = ev['Base MSRP'].describe()

# Print the results
print(base_msrp_stats)

# Create box plots of Base MSRP for different vehicle types
plt.figure(figsize=(7, 5))
sns.boxplot(data=ev, x='Electric Vehicle Type', y='Base MSRP')
plt.xlabel('Vehicle Type')
plt.ylabel('Base MSRP')
plt.title('Distribution of Base MSRP by Vehicle Type')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```

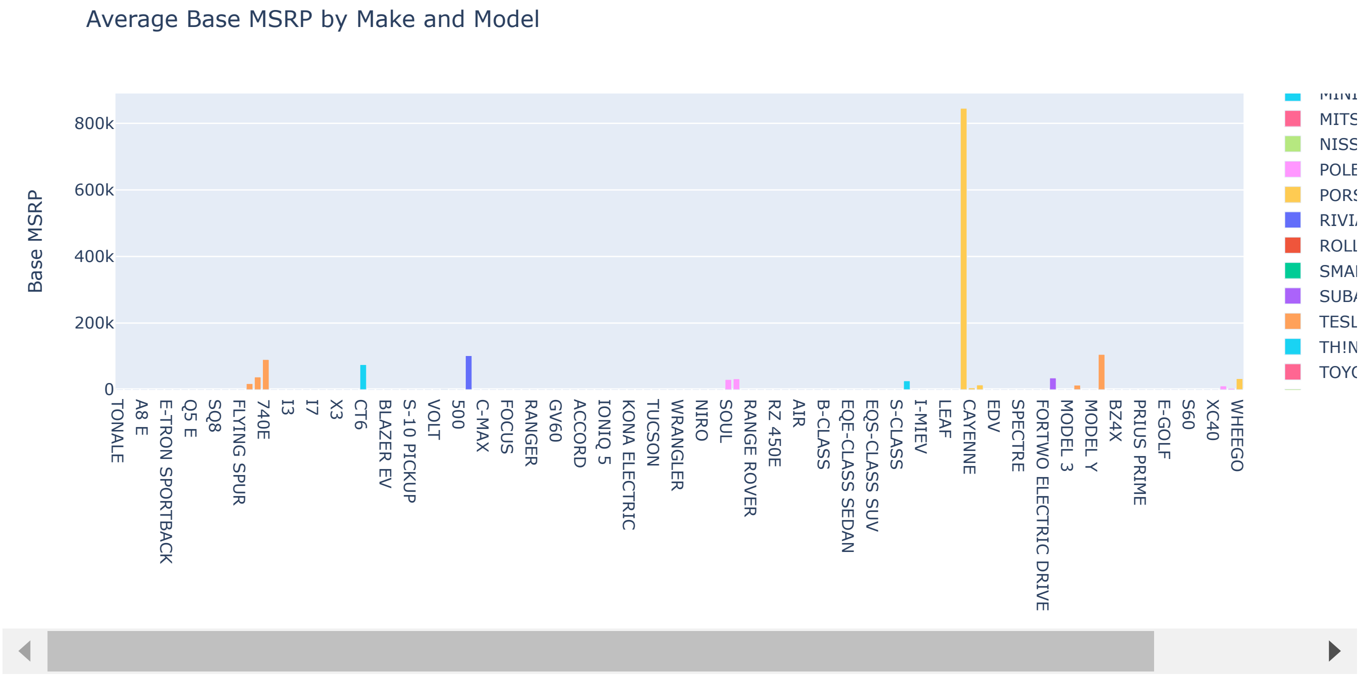
```
count     177473.000000
mean        1070.609839
std         8346.920672
min            0.000000
25%            0.000000
50%            0.000000
75%            0.000000
max       845000.000000
Name: Base MSRP, dtype: float64
```




```
!pip install plotly
```

```
Requirement already satisfied: plotly in /usr/local/lib/python3.10/dist-packages (5.15.0)
Requirement already satisfied: tenacity>=6.2.0 in /usr/local/lib/python3.10/dist-packages (from plotly) (8.4.2)
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from plotly) (24.1)
```

```
import plotly.express as px
cost_by_make_model = ev.groupby(['Make', 'Model'])['Base MSRP'].mean().reset_index()
fig = px.bar(cost_by_make_model, x='Model', y='Base MSRP', color='Make', title='Average Base MSRP by Make and Model')
fig.show()
```



6. Utility Service Providers:

- Which electric utilities are most commonly associated with the locations of electric vehicles?

- How might these utility providers influence the adoption and support of electric vehicles?

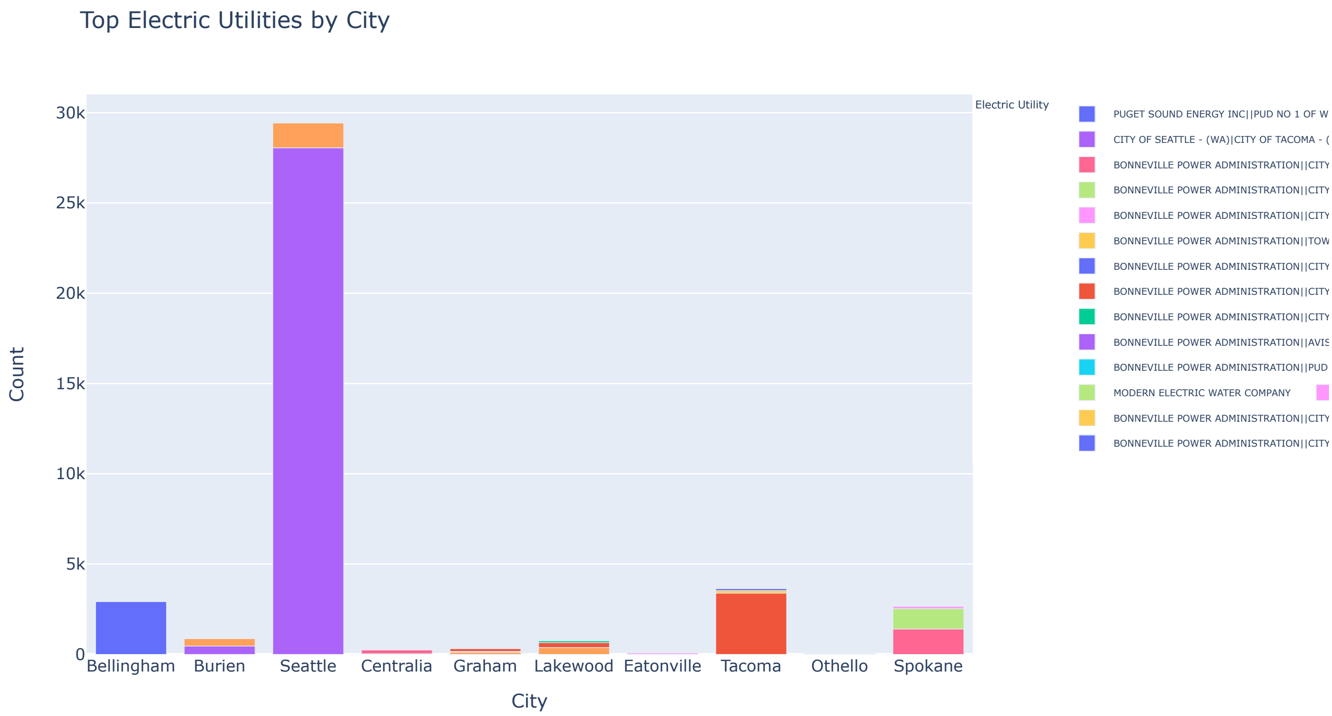
```
utility_counts = ev['Electric Utility'].value_counts()
print(utility_counts)
```

```
Electric Utility
PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)                                          65990
PUGET SOUND ENERGY INC                                                                 35882
CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)                                           31381
BONNEVILLE POWER ADMINISTRATION||PUD NO 1 OF CLARK COUNTY - (WA)                       10173
BONNEVILLE POWER ADMINISTRATION||CITY OF TACOMA - (WA)||PENINSULA LIGHT COMPANY         7828
                                                                                        ...
BONNEVILLE POWER ADMINISTRATION||PUD NO 1 OF ASOTIN COUNTY||INLAND POWER & LIGHT COMPANY    2
BONNEVILLE POWER ADMINISTRATION||PUD NO 1 OF CLALLAM COUNTY|PUD NO 1 OF JEFFERSON COUNTY    1
BONNEVILLE POWER ADMINISTRATION||PENINSULA LIGHT COMPANY                                    1
CITY OF SEATTLE - (WA)                                                                      1
BONNEVILLE POWER ADMINISTRATION||PUD NO 1 OF JEFFERSON COUNTY                               1
Name: count, Length: 75, dtype: int64
```

```
city_utility_counts = ev.groupby(['City', 'Electric Utility'])['VIN (1-10)'].count().reset_index(name='Count')
top_utilities_by_city = city_utility_counts.groupby('City').apply(lambda x: x.nlargest(3, 'Count')).reset_index(drop=True)
print(top_utilities_by_city)
```

```
               City                                   Electric Utility  Count
0          Aberdeen  BONNEVILLE POWER ADMINISTRATION||PUD NO 1 OF G...    156
1              Acme  PUGET SOUND ENERGY INC||PUD NO 1 OF WHATCOM CO...     10
2              Addy                                        AVISTA CORP      2
3              Adna  BONNEVILLE POWER ADMINISTRATION||CITY OF TACOM...      1
4    Airway Heights  BONNEVILLE POWER ADMINISTRATION||AVISTA CORP||...     29
..              ...                                                ...    ...
621          Yakima  BONNEVILLE POWER ADMINISTRATION||BENTON RURAL ...      2
622    Yarrow Point      PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)    145
623            Yelm                             PUGET SOUND ENERGY INC    265
624          Zillah                                         PACIFICORP     34
625          Zillah  BONNEVILLE POWER ADMINISTRATION||PACIFICORP||B...      1

[626 rows x 3 columns]
```

```
top_cities = city_utility_counts['City'].value_counts().head(10).index
top_city_utilities = top_utilities_by_city[top_utilities_by_city['City'].isin(top_cities)]

fig = px.bar(top_city_utilities, x='City', y='Count', color='Electric Utility',
             title='Top Electric Utilities by City', height=600, width=1500)  # Adjust height as needed
fig.update_layout(
    legend=dict(x=1, y=1, xanchor='left', yanchor='top', font=dict(size=7)),  # Adjust font size
    legend_orientation="h"
)
fig.show()
```



```
# Group by 'Model Year' and count the number of vehicles per year
yearly_data = ev.groupby('Model Year').size().reset_index(name='Number of EVs')

print(yearly_data)

import matplotlib.pyplot as plt
import seaborn as sns

# Plot the number of EVs registered each year
plt.figure(figsize=(8, 6))
sns.lineplot(data=yearly_data, x='Model Year', y='Number of EVs', marker='o', color='blue')
plt.title('Number of EVs Registered Each Year')
plt.xlabel('Year')
plt.ylabel('Number of EVs')
plt.grid(True)
plt.show()
```

```
    Model Year  Number of EVs
0         1997              1
1         1998              1
2         1999              5
3         2000              7
4         2002              2
5         2003              1
6         2008             19
7         2010             23
8         2011            775
9         2012           1614
10        2013           4399
11        2014           3496
12        2015           4826
13        2016           5469
14        2017           8534
15        2018          14286
16        2019          10913
17        2020          11740
18        2021          19063
19        2022          27708
20        2023          57519
21        2024           7072
```