evmarketsizeanalysis

July 20, 2024

[51]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

[52]:
```
df = pd.read_csv('/content/Electric_Vehicle_Population_Data.csv')
```

[53]:
```
df.head()
```

[53]:
```
   VIN (1-10)     County       City State  Postal Code  Model Year   Make  \
0  5YJYGDEE1L       King    Seattle    WA      98122.0        2020  TESLA
1  7SAYGDEE9P  Snohomish    Bothell    WA      98021.0        2023  TESLA
2  5YJSA1E4XK       King    Seattle    WA      98109.0        2019  TESLA
3  5YJSA1E27G       King   Issaquah    WA      98027.0        2016  TESLA
4  5YJYGDEE5M     Kitsap  Suquamish    WA      98392.0        2021  TESLA

     Model           Electric Vehicle Type  \
0  MODEL Y  Battery Electric Vehicle (BEV)
1  MODEL Y  Battery Electric Vehicle (BEV)
2  MODEL S  Battery Electric Vehicle (BEV)
3  MODEL S  Battery Electric Vehicle (BEV)
4  MODEL Y  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

   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

                  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)

                                 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
```

[54]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 177866 entries, 0 to 177865
Data columns (total 17 columns):
 #   Column                                             Non-Null Count   Dtype
---  ------                                             --------------   -----
 0   VIN (1-10)                                         177866 non-null  object
 1   County                                             177861 non-null  object
 2   City                                               177861 non-null  object
 3   State                                              177866 non-null  object
 4   Postal Code                                        177861 non-null  float64
 5   Model Year                                         177866 non-null  int64
 6   Make                                               177866 non-null  object
 7   Model                                              177866 non-null  object
 8   Electric Vehicle Type                              177866 non-null  object
 9   Clean Alternative Fuel Vehicle (CAFV) Eligibility  177866 non-null  object
 10  Electric Range                                     177866 non-null  int64
 11  Base MSRP                                          177866 non-null  int64
 12  Legislative District                               177477 non-null  float64
 13  DOL Vehicle ID                                     177866 non-null  int64
 14  Vehicle Location                                   177857 non-null  object
 15  Electric Utility                                   177861 non-null  object
 16  2020 Census Tract                                  177861 non-null  float64
dtypes: float64(3), int64(4), object(10)
memory usage: 23.1+ MB
```

[55]:
```
df.isnull().sum()
```

[55]:
```
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
```

[56]:
```
df = df.dropna()
```

[57]:
```
df.isnull().sum()
```

[57]:
```
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
```

[58]:
```
df['Model Year'].unique()
```

[58]:
```
array([2020, 2023, 2019, 2016, 2021, 2017, 2013, 2018, 2015, 2022, 2014,
       2012, 2024, 2011, 2000, 2008, 2010, 2002, 1998, 1999, 1997, 2003])
```

- We have data from 1997-2024(half)

[59]:
```
df['City'].nunique()
```

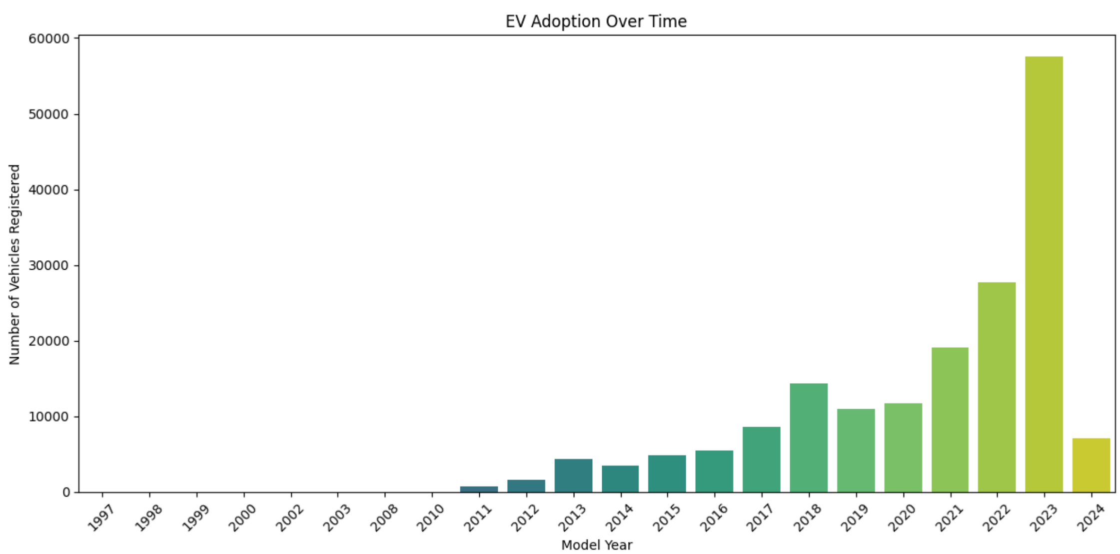
[59]: 468

[60]:
```
df['Electric Vehicle Type'].unique()
```

[60]:
```
array(['Battery Electric Vehicle (BEV)',
       'Plug-in Hybrid Electric Vehicle (PHEV)'], dtype=object)
```

0.0.1 EV Adoption over the years

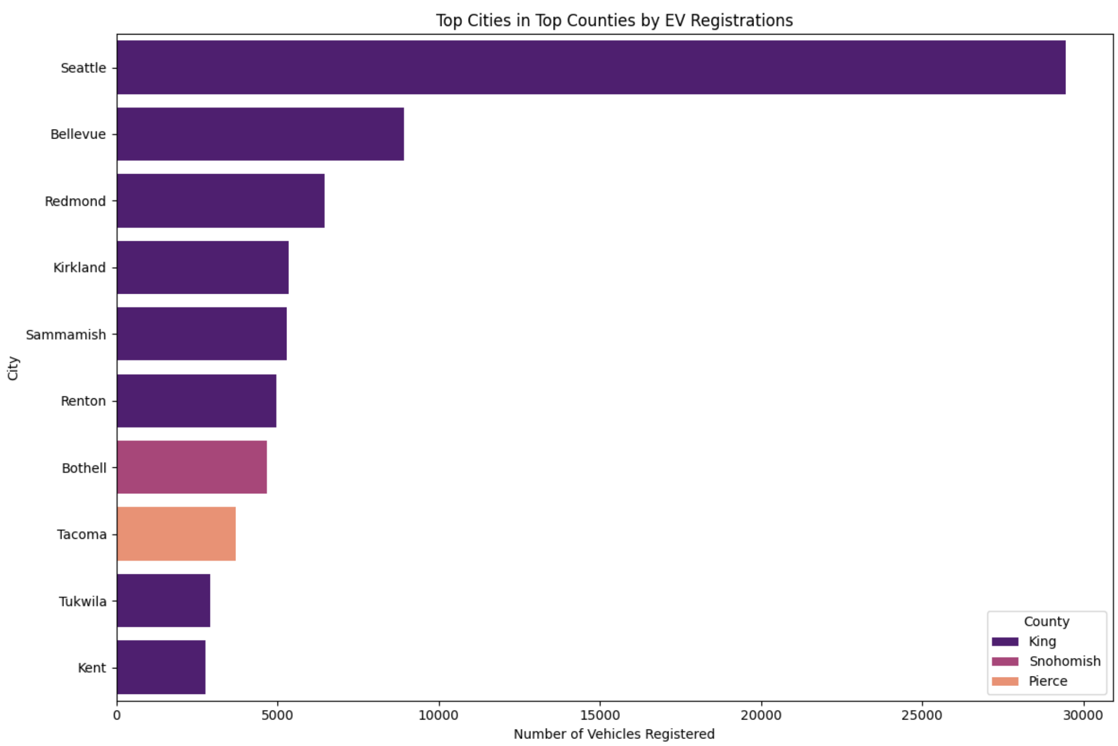
[61]:
```
plt.figure(figsize=(12, 6))
ev_adoption_by_year = df['Model Year'].value_counts().sort_index()
sns.barplot(x=ev_adoption_by_year.index, y=ev_adoption_by_year.values,␣
 ↪palette="viridis")
plt.title('EV Adoption Over Time')
plt.xlabel('Model Year')
plt.ylabel('Number of Vehicles Registered')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



From the above bar chart, it's clear that EV adoption has been increasing over time, especially noting a significant upward trend starting around 2016. The number of vehicles registered grows modestly up until that point and then begins to rise more rapidly from 2017 onwards. The year 2023 shows a particularly sharp increase in the number of registered EVs, with the bar for 2023 being the highest on the graph, indicating a peak in EV adoption.

[63]:

```
# geographical distribution at county level
ev_county_distribution = df['County'].value_counts()
top_counties = ev_county_distribution.head(3).index

# filtering the dataset for these top counties
top_counties_data = df[df['County'].isin(top_counties)]

# analyzing the distribution of EVs within the cities of these top counties
ev_city_distribution_top_counties = top_counties_data.groupby(['County',␣
↪'City']).size().sort_values(ascending=False).reset_index(name='Number of␣
↪Vehicles')

# visualize the top 10 cities across these counties
top_cities = ev_city_distribution_top_counties.head(10)

plt.figure(figsize=(12, 8))
sns.barplot(x='Number of Vehicles', y='City', hue='County', data=top_cities,␣
↪palette="magma")
plt.title('Top Cities in Top Counties by EV Registrations')
plt.xlabel('Number of Vehicles Registered')
plt.ylabel('City')
plt.legend(title='County')
plt.tight_layout()
plt.show()
```

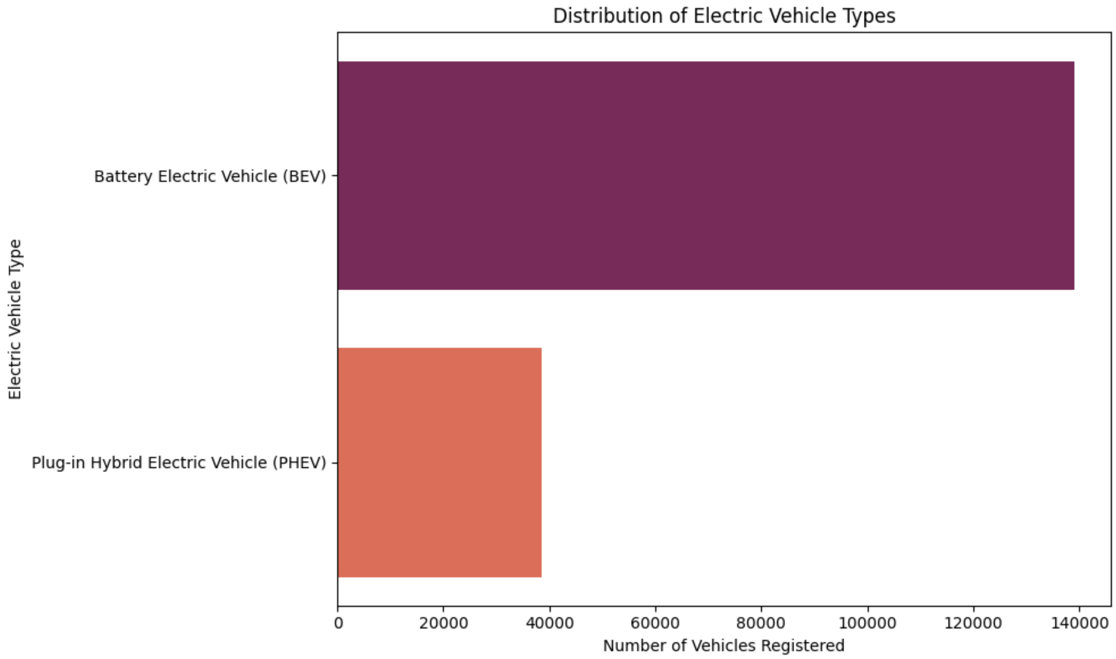


The above graph compares the number of electric vehicles registered in various cities within three counties: King, Snohomish, and Pierce. The horizontal bars represent cities, and their length corresponds to the number of vehicles registered, colour-coded by county. Here are the key findings from the above graph:

- Seattle, which is in King County, has the highest number of EV registrations by a significant margin, far outpacing the other cities listed.
- Bellevue and Redmond, also in King County, follow Seattle with the next highest registrations, though these are considerably less than Seattle's.
- Cities in Snohomish County, such as Kirkland and Sammamish, show moderate EV registrations.
- Tacoma and Tukwila, representing Pierce County, have the fewest EV registrations among the cities listed, with Tacoma slightly ahead of Tukwila.
- The majority of cities shown are from King County, which seems to dominate EV registrations among the three counties.

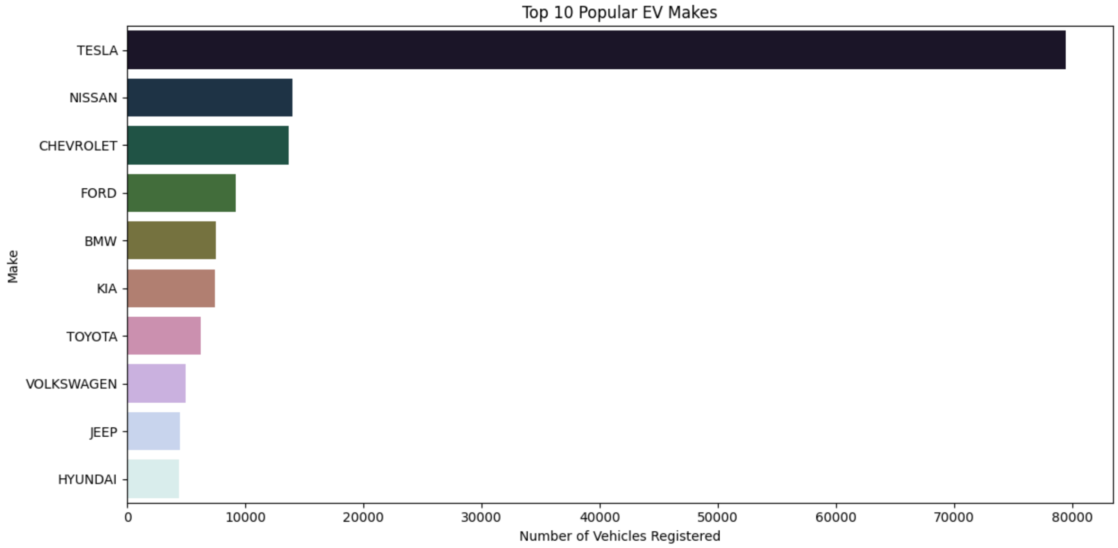
Overall, the graph indicates that EV adoption is not uniform across the cities and is more concentrated in certain areas, particularly in King County.

[64]:
```
# analyzing the distribution of electric vehicle Types
ev_type_distribution = df['Electric Vehicle Type'].value_counts()

plt.figure(figsize=(10, 6))
sns.barplot(x=ev_type_distribution.values, y=ev_type_distribution.index,␣
 ↪palette="rocket")
plt.title('Distribution of Electric Vehicle Types')
plt.xlabel('Number of Vehicles Registered')
plt.ylabel('Electric Vehicle Type')
plt.tight_layout()
plt.show()
```



- The above graph shows that BEVs are more popular or preferred over PHEVs among the electric vehicles registered in the United States.

[65]:
```
# analyzing the popularity of EV manufacturers
ev_make_distribution = df['Make'].value_counts().head(10)  # Limiting to top 10␣
↪for clarity

plt.figure(figsize=(12, 6))
sns.barplot(x=ev_make_distribution.values, y=ev_make_distribution.index,␣
↪palette="cubehelix")
plt.title('Top 10 Popular EV Makes')
plt.xlabel('Number of Vehicles Registered')
plt.ylabel('Make')
plt.tight_layout()
plt.show()
```

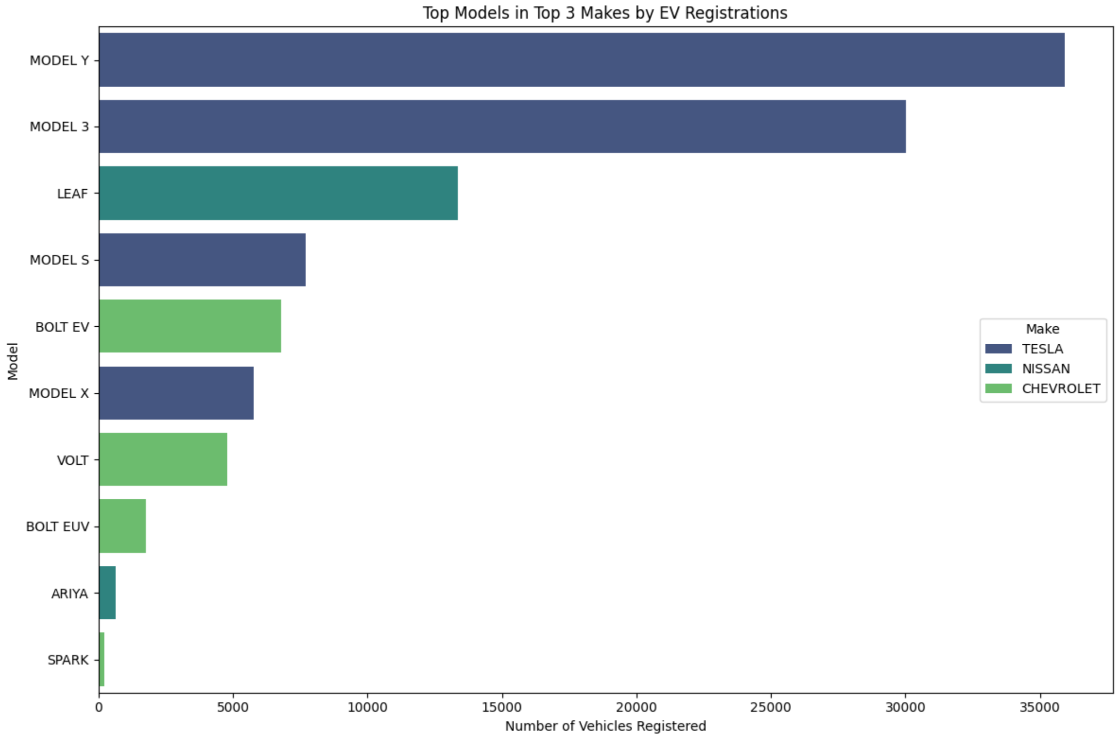


The above chart shows that:

- TESLA leads by a substantial margin with the highest number of vehicles registered.
- NISSAN is the second most popular manufacturer, followed by CHEVROLET, though both have significantly fewer registrations than TESLA.
- FORD, BMW, KIA, TOYOTA, VOLKSWAGEN, JEEP, and HYUNDAI follow in decreasing order of the number of registered vehicles.

[66]:
```
# selecting the top 3 manufacturers based on the number of vehicles registered
top_3_makes = ev_make_distribution.head(3).index

# filtering the dataset for these top manufacturers
top_makes_data = df[df['Make'].isin(top_3_makes)]

# analyzing the popularity of EV models within these top manufacturers
ev_model_distribution_top_makes = top_makes_data.groupby(['Make', 'Model']).
 ↪size().sort_values(ascending=False).reset_index(name='Number of Vehicles')

# visualizing the top 10 models across these manufacturers for clarity
top_models = ev_model_distribution_top_makes.head(10)

plt.figure(figsize=(12, 8))
sns.barplot(x='Number of Vehicles', y='Model', hue='Make', data=top_models,␣
 ↪palette="viridis")
plt.title('Top Models in Top 3 Makes by EV Registrations')
plt.xlabel('Number of Vehicles Registered')
plt.ylabel('Model')
plt.legend(title='Make', loc='center right')
plt.tight_layout()
```

```
plt.show()
```

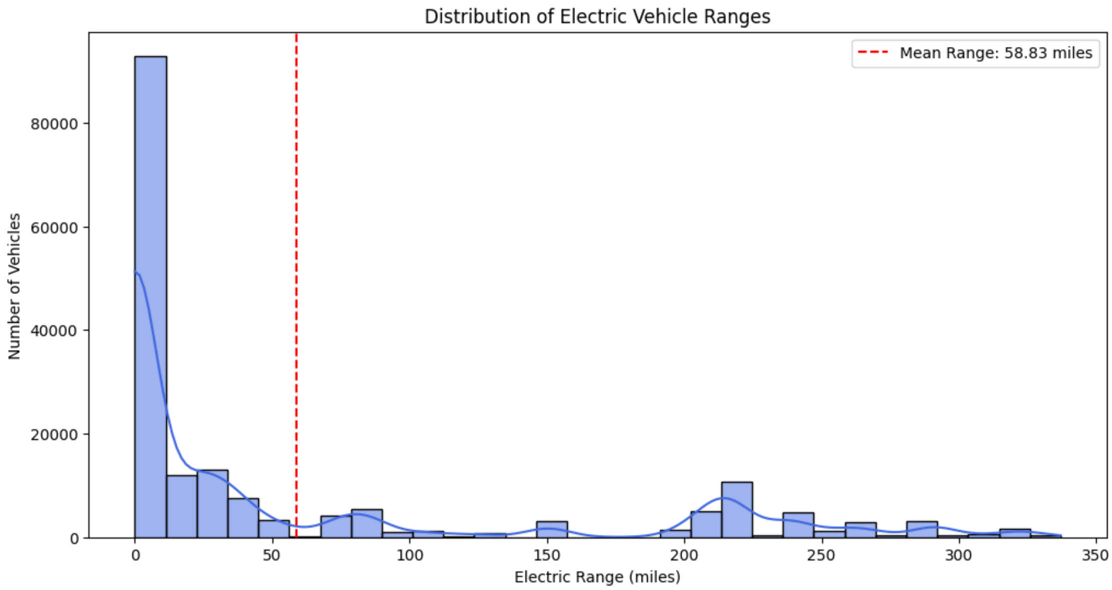


The above graph shows the distribution of electric vehicle registrations among different models from the top three manufacturers: TESLA, NISSAN, and CHEVROLET. Here are the findings:

- TESLA's MODEL Y and MODEL 3 are the most registered vehicles, with MODEL Y having the highest number of registrations.
- NISSAN's LEAF is the third most registered model and the most registered non-TESLA vehicle.
- TESLA's MODEL S and MODEL X also have a significant number of registrations.
- CHEVROLET's BOLT EV and VOLT are the next in the ranking with considerable registrations, followed by BOLT EUV.
- NISSAN's ARIYA and CHEVROLET's SPARK have the least number of registrations among the models shown.

[67]:
```
# analyzing the distribution of electric range
plt.figure(figsize=(12, 6))
sns.histplot(df['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(df['Electric Range'].mean(), color='red', linestyle='--',␣
 ↪label=f'Mean Range: {df["Electric Range"].mean():.2f} miles')
plt.legend()
```

```
plt.show()
```

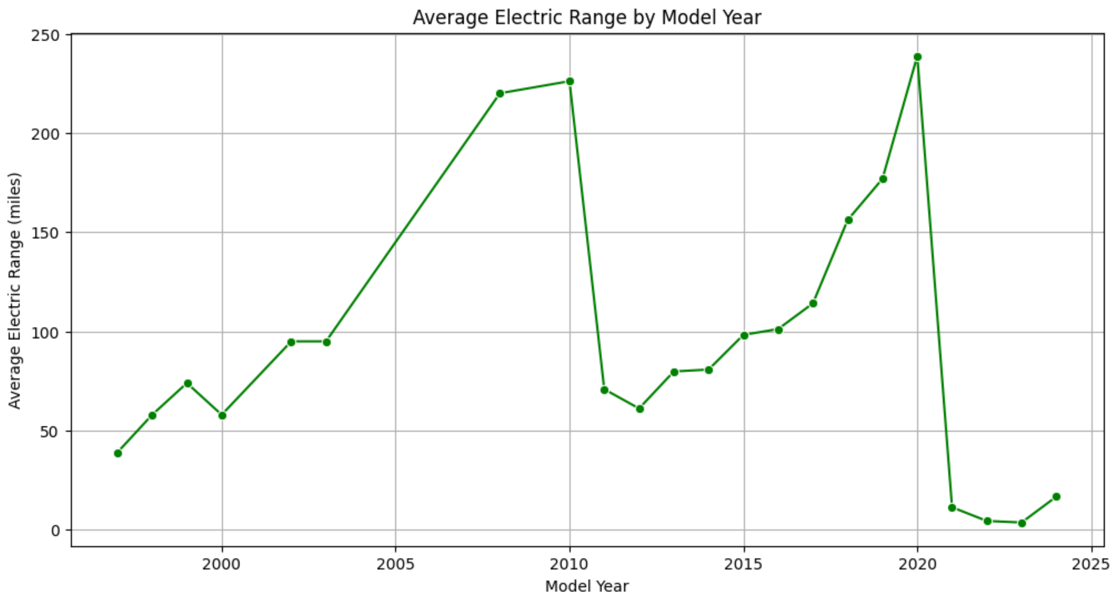


The above graph shows the mean electric range. Key observations from the graph include:

- There is a high frequency of vehicles with a low electric range, with a significant peak occurring just before 50 miles.
- 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.
- 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.
- 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.

It suggests that while there are EVs available with high electric ranges, the average range is skewed lower due to a substantial number of vehicles with shorter ranges

[68]:
```
# calculating the average electric range by model year
average_range_by_year = df.groupby('Model Year')['Electric Range'].mean().
 ↪reset_index()

plt.figure(figsize=(12, 6))
sns.lineplot(x='Model Year', y='Electric Range', data=average_range_by_year,␣
 ↪marker='o', color='green')
plt.title('Average Electric Range by Model Year')
plt.xlabel('Model Year')
plt.ylabel('Average Electric Range (miles)')
plt.grid(True)
plt.show()
```



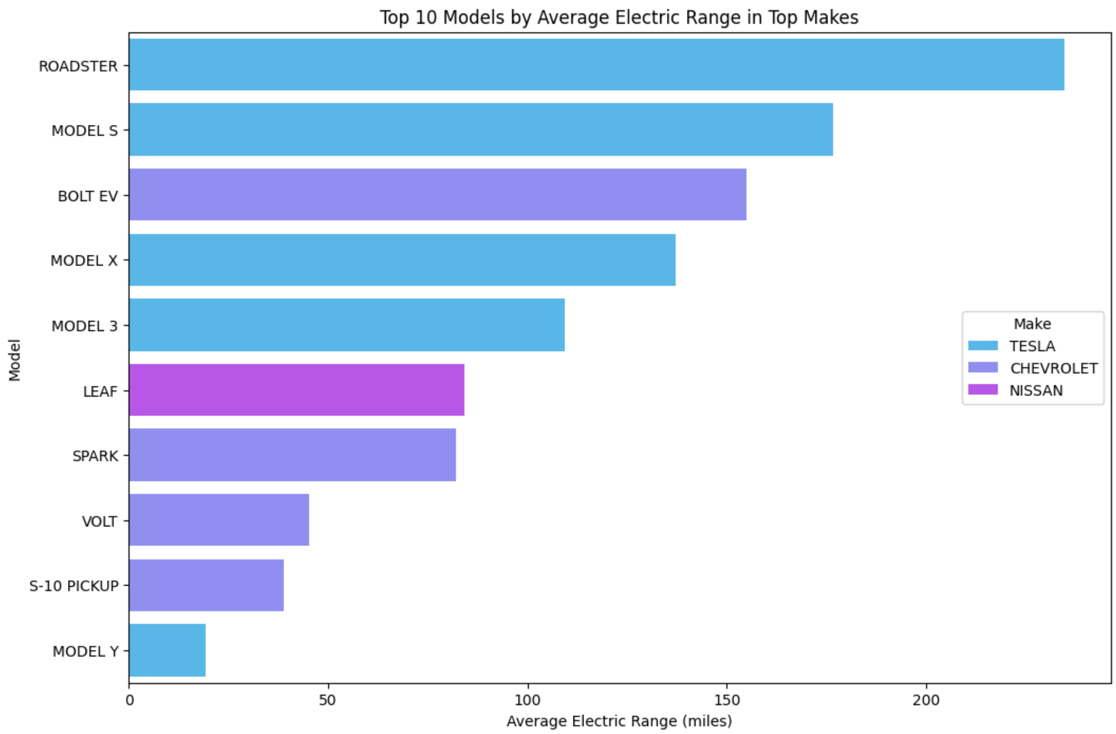
The above graph shows the progression of the average electric range of vehicles from around the year 2000 to 2024. Key findings from the graph:

- There is a general upward trend in the average electric range of EVs over the years, indicating improvements in technology and battery efficiency.
- There is a noticeable peak around the year 2020 when the average range reaches its highest point.
- Following 2020, there's a significant drop in the average range, which could indicate that data for the following years might be incomplete or reflect the introduction of several lower-range models.
- After the sharp decline, there is a slight recovery in the average range in the most recent year shown on the graph.

The data suggest that while there have been fluctuations, the overall trend over the last two decades has been toward increasing the electric range of EVs.

[69]:
```
average_range_by_model = top_makes_data.groupby(['Make', 'Model'])['Electric␣
↪Range'].mean().sort_values(ascending=False).reset_index()

# the top 10 models with the highest average electric range
top_range_models = average_range_by_model.head(10)

plt.figure(figsize=(12, 8))
barplot = sns.barplot(x='Electric Range', y='Model', hue='Make',␣
↪data=top_range_models, palette="cool")
plt.title('Top 10 Models by Average Electric Range in Top Makes')
plt.xlabel('Average Electric Range (miles)')
plt.ylabel('Model')
```

```
plt.legend(title='Make', loc='center right')
plt.show()
```



The TESLA ROADSTER has the highest average electric range among the models listed. TESLA's models (ROADSTER, MODEL S, MODEL X, and MODEL 3) occupy the majority of the top positions, indicating that on average, TESLA's vehicles have higher electric ranges. The CHEVROLET BOLT EV is an outlier among the CHEVROLET models, having a substantially higher range than the VOLT and S-10 PICKUP from the same maker. NISSAN's LEAF and CHEVROLET's SPARK are in the lower half of the chart, suggesting more modest average ranges.

[70]:
```
# calculate the number of EVs registered each year
ev_registration_counts = df['Model Year'].value_counts().sort_index()
ev_registration_counts
```

[70]:
```
Model Year
1997        1
1998        1
1999        5
2000        7
2002        2
2003        1
2008       19
2010       23
2011      775
```

```
2012     1614
2013     4399
2014     3496
2015     4826
2016     5469
2017     8534
2018    14286
2019    10913
2020    11740
2021    19063
2022    27708
2023    57519
2024     7072
Name: count, dtype: int64
```

The dataset provides the number of electric vehicles registered each year from 1997 through 2024. However, the data for 2024 is incomplete as it only contains the data till March. Here's a summary of EV registrations for recent years:

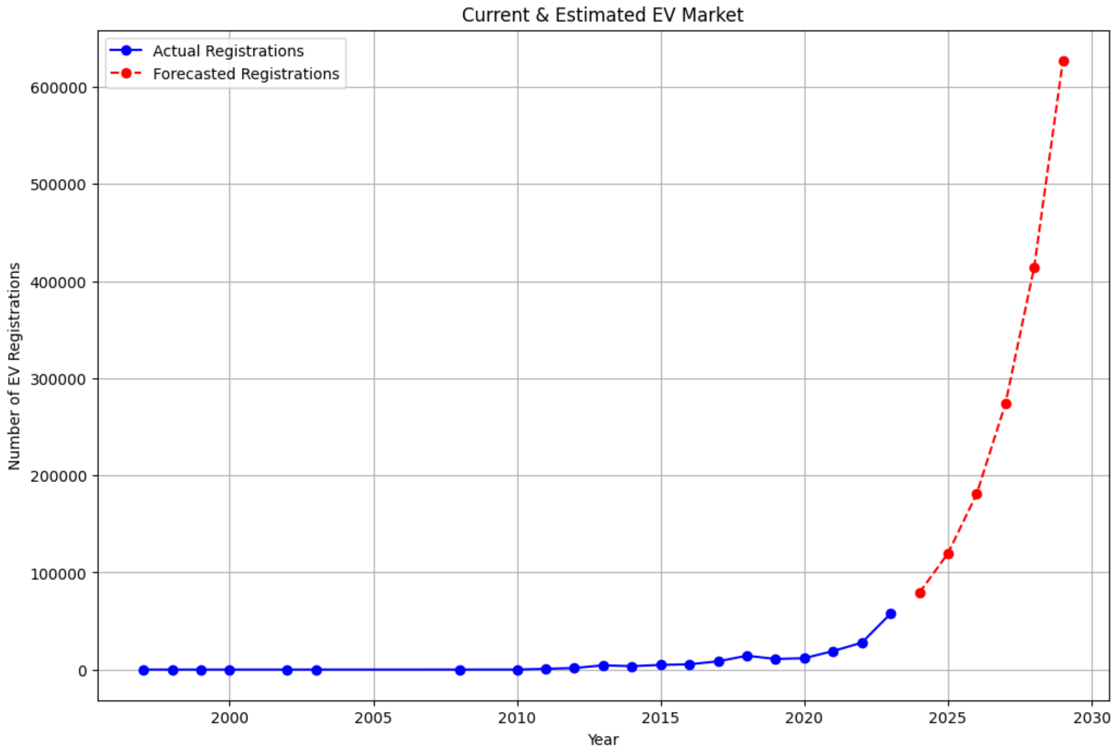
- In 2021, there were 19,063 EVs registered.
- In 2022, the number increased to 27708 EVs.
- In 2023, a significant jump to 57,519 EVs was observed.
- For 2024, currently, 7,072 EVs are registered, which suggests partial data.

[71]:
```
from scipy.optimize import curve_fit
import numpy as np

# filter the dataset to include years with complete data, assuming 2023 is the
 ↪last complete year
filtered_years = ev_registration_counts[ev_registration_counts.index <= 2023]

# define a function for exponential growth to fit the data
def exp_growth(x, a, b):
    return a * np.exp(b * x)

# prepare the data for curve fitting
x_data = filtered_years.index - filtered_years.index.min()
y_data = filtered_years.values

# fit the data to the exponential growth function
params, covariance = curve_fit(exp_growth, x_data, y_data)

# use the fitted function to forecast the number of EVs for 2024 and the next
 ↪five years
forecast_years = np.arange(2024, 2024 + 6) - filtered_years.index.min()
forecasted_values = exp_growth(forecast_years, *params)

# create a dictionary to display the forecasted values for easier interpretation
```

```
forecasted_evs = dict(zip(forecast_years + filtered_years.index.min(),␣
↪forecasted_values))

print(forecasted_evs)
```

```
{2024: 79079.2066611501, 2025: 119653.95934090775, 2026: 181047.21317328632,
2027: 273940.7335817853, 2028: 414496.9933533305, 2029: 627171.2689549965}
```

[72]:
```
# prepare data for plotting
years = np.arange(filtered_years.index.min(), 2029 + 1)
actual_years = filtered_years.index
forecast_years_full = np.arange(2024, 2029 + 1)

# actual and forecasted values
actual_values = filtered_years.values
forecasted_values_full = [forecasted_evs[year] for year in forecast_years_full]

plt.figure(figsize=(12, 8))
plt.plot(actual_years, actual_values, 'bo-', label='Actual Registrations')
plt.plot(forecast_years_full, forecasted_values_full, 'ro--', label='Forecasted␣
↪Registrations')

plt.title('Current & Estimated EV Market')
plt.xlabel('Year')
plt.ylabel('Number of EV Registrations')
plt.legend()
plt.grid(True)

plt.show()
```



From the above graph, we can see:

- The number of actual EV registrations remained relatively low and stable until around 2010, after which there was a consistent and steep upward trend, suggesting a significant increase in EV adoption.
- The forecasted EV registrations predict an even more dramatic increase in the near future, with the number of registrations expected to rise sharply in the coming years.
- Given the growing trend in actual EV registrations and the projected acceleration as per the forecast data, we can conclude that the EV market size is expected to expand considerably. The steep increase in forecasted registrations suggests that consumer adoption of EVs is on the rise, and this trend is likely to continue. Overall, the data point towards a promising future for the EV industry, indicating a significant shift in consumer preferences and a potential increase in related investment and business opportunities.

1 Summary

So, market size analysis is a crucial aspect of market research that determines the potential sales volume within a given market. It helps businesses understand the magnitude of demand, assess market saturation levels, and identify growth opportunities. From our market size analysis of electric vehicles, we found a promising future for the EV industry, indicating a significant shift in consumer preferences and a potential increase in related investment and business opportunities.

[72]: