Electric Vehicles Market Size Analysis using Python

In [2]:
```
import pandas as pd

data = pd.read_csv('Electric_Vehicle_Population_Data.csv')
```

Data Exploration

In [3]:
```
data.head()
```

Out[3]:

	VIN (1-10)	County	City	State	Postal Code	Model Year	Make	Model	Electric Vehicle Type	Clean Alternative Fuel Vehicle (CAFV) Eligibility	E
0	5YJYGDEE1L	King	Seattle	WA	98122.0	2020	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	
1	7SAYGDEE9P	Snohomish	Bothell	WA	98021.0	2023	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	
2	5YJSA1E4XK	King	Seattle	WA	98109.0	2019	TESLA	MODEL S	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	
3	5YJSA1E27G	King	Issaquah	WA	98027.0	2016	TESLA	MODEL S	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	
4	5YJYGDEE5M	Kitsap	Suquamish	WA	98392.0	2021	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	

In [4]:
```
data.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
```

In [5]:
```
data.isnull().sum()
```

Out[5]:
```
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
```

In [6]:
```
data = data.dropna()
```

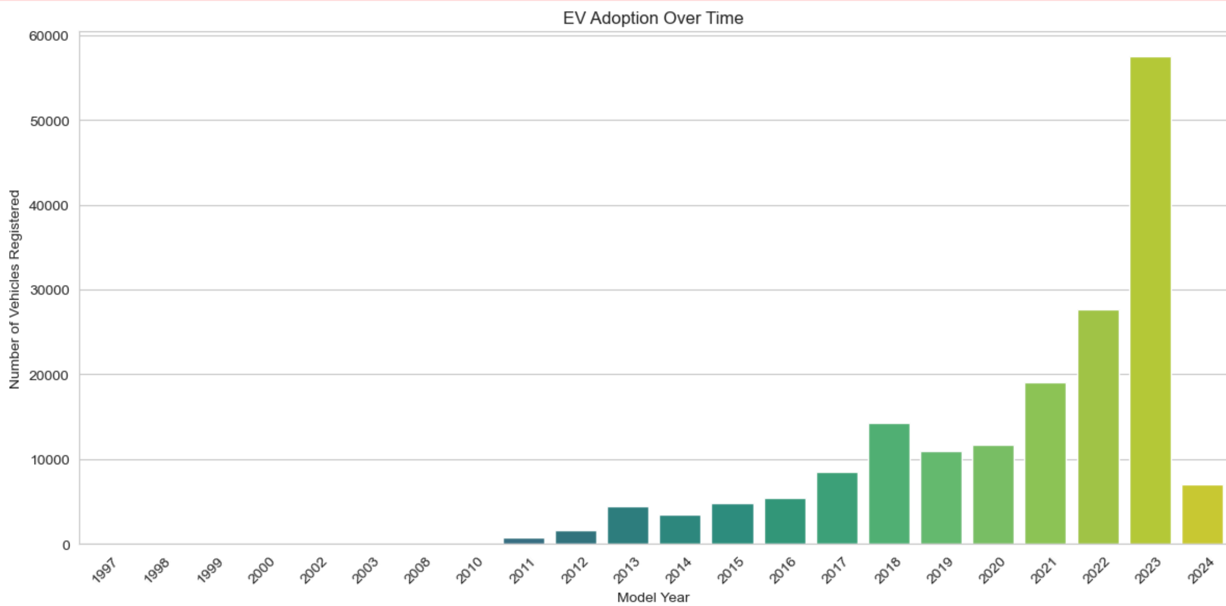
Analysis

In [7]:
```
import matplotlib.pyplot as plt
import seaborn as sns
sns.set_style("whitegrid")
```

EV Adoption Over Time

In [8]:
```
plt.figure(figsize=(12, 6))
ev_adoption_by_year = data['Model Year'].value_counts().sort_index()
sns.barplot(x=ev_adoption_by_year.index, y=ev_adoption_by_year.values, palette="viridi
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()
```

```
C:\Users\Bryan\AppData\Local\Temp\ipykernel_14868\1085525010.py:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.
0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

  sns.barplot(x=ev_adoption_by_year.index, y=ev_adoption_by_year.values, palette="vir
idis")
```

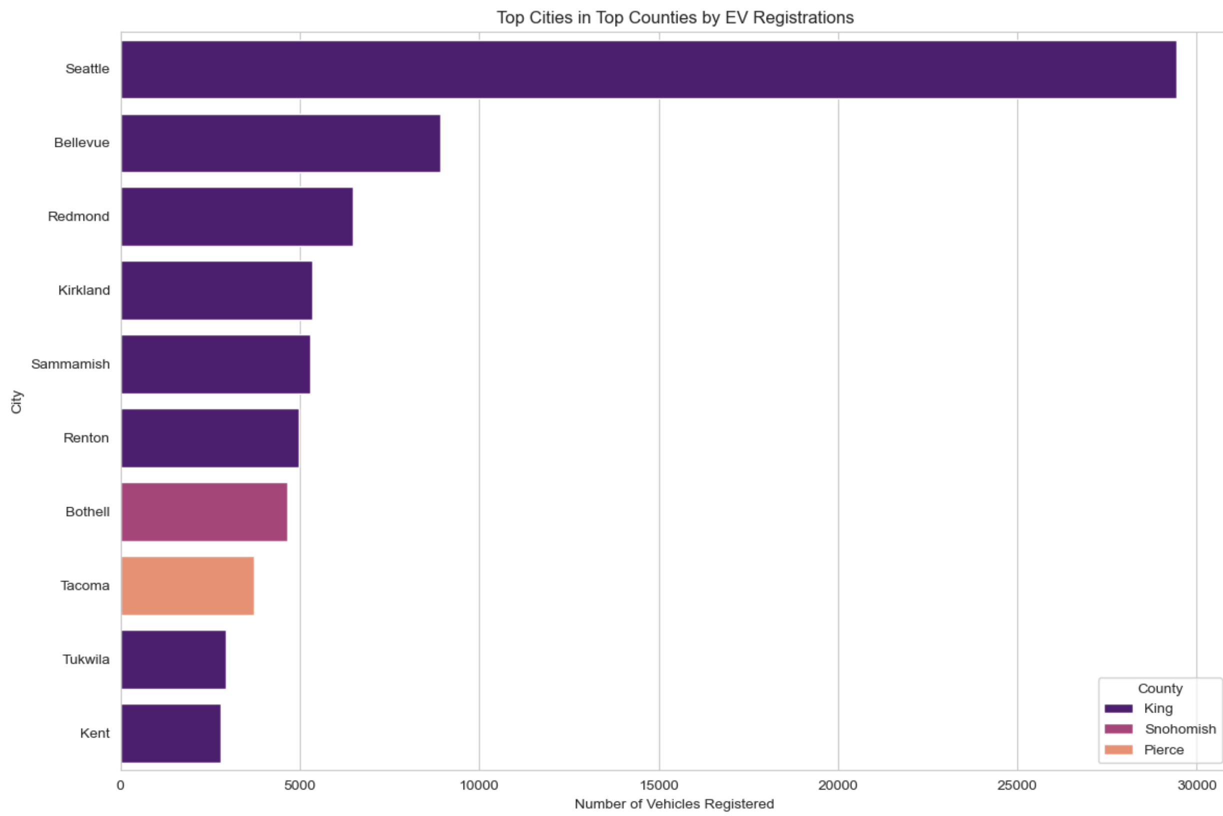


EV adoption has been increasing over time, especially noting a significant upward trend starting around 2016. The number of vehicles registered grows modestly up until that point and then begins to rise more rapidly from 2017 onwards.

Geographical distribution at county level

In [9]:
```
ev_county_distribution = data['County'].value_counts()
top_counties = ev_county_distribution.head(3).index

# filtering the dataset for these top counties
top_counties_data = data[data['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

# 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='
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()
```



- 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

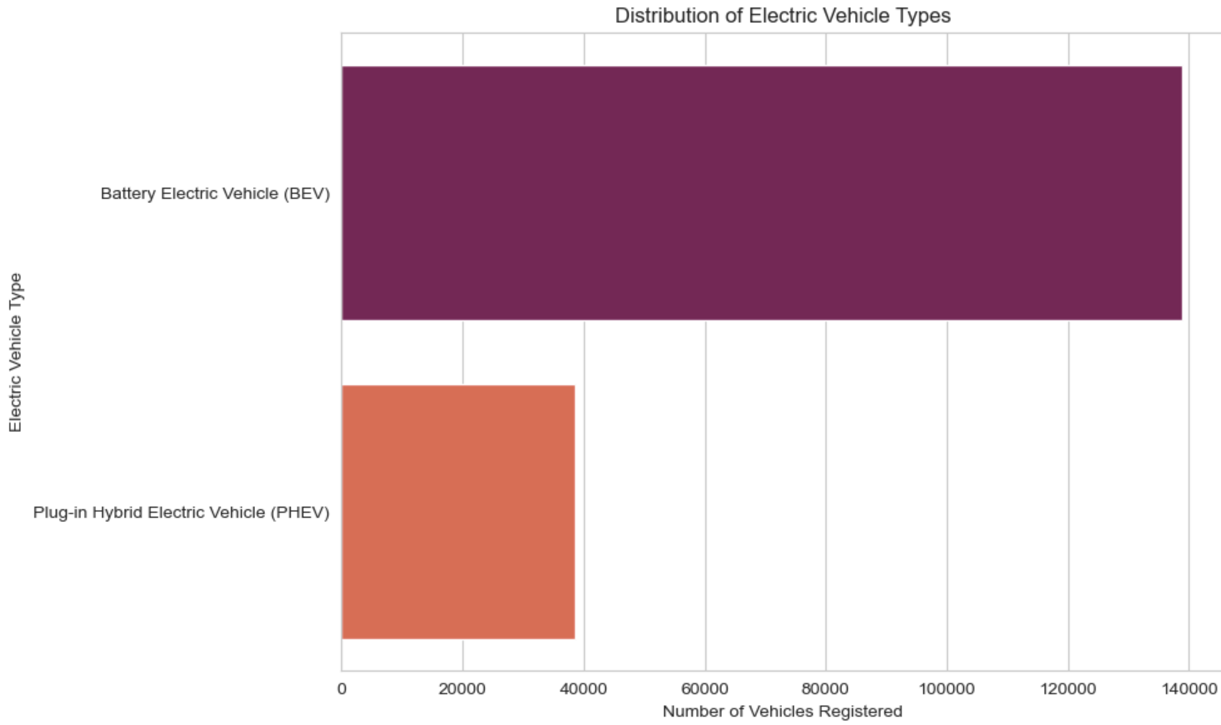
Distribution of electric vehicle Types

In [10]:

```
ev_type_distribution = data['Electric Vehicle Type'].value_counts()

plt.figure(figsize=(10, 6))
sns.barplot(x=ev_type_distribution.values, y=ev_type_distribution.index, palette="rock
```

```
plt.title('Distribution of Electric Vehicle Types')
plt.xlabel('Number of Vehicles Registered')
plt.ylabel('Electric Vehicle Type')
plt.tight_layout()
plt.show()
```

```
C:\Users\Bryan\AppData\Local\Temp\ipykernel_14868\2054861885.py:4: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.
0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

  sns.barplot(x=ev_type_distribution.values, y=ev_type_distribution.index, palette="r
ocket")
```

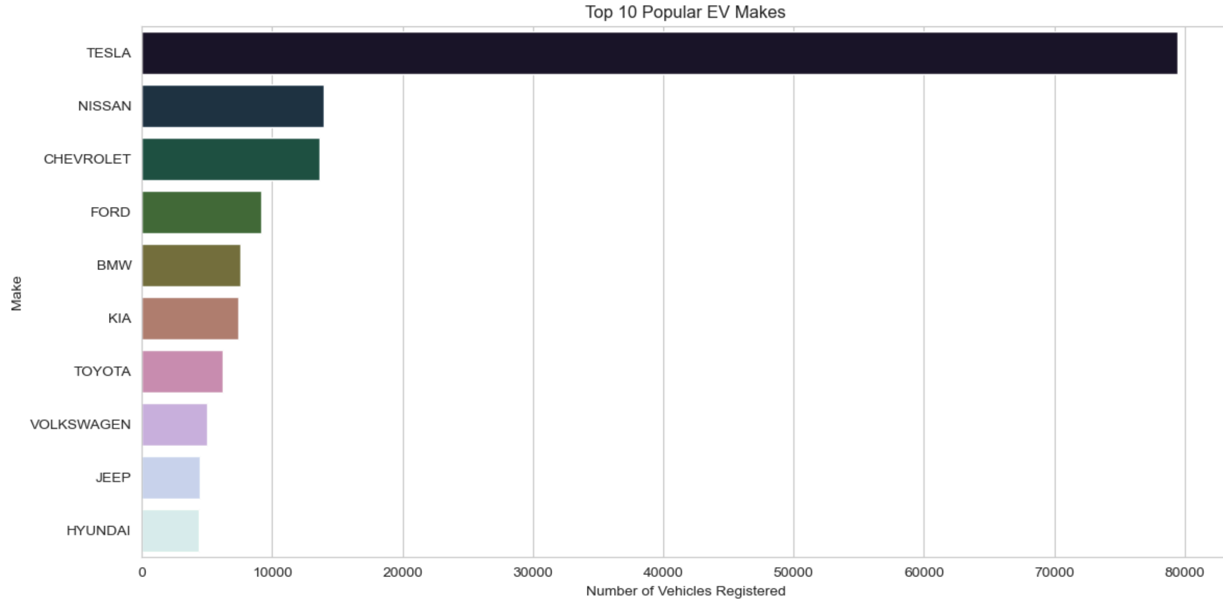


Popularity of EV manufacturers

In [11]:

```
ev_make_distribution = data['Make'].value_counts().head(10)  # Limiting to top 10 for

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

```
C:\Users\Bryan\AppData\Local\Temp\ipykernel_14868\3101160592.py:4: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.
0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

  sns.barplot(x=ev_make_distribution.values, y=ev_make_distribution.index, palette="c
ubehelix")
```

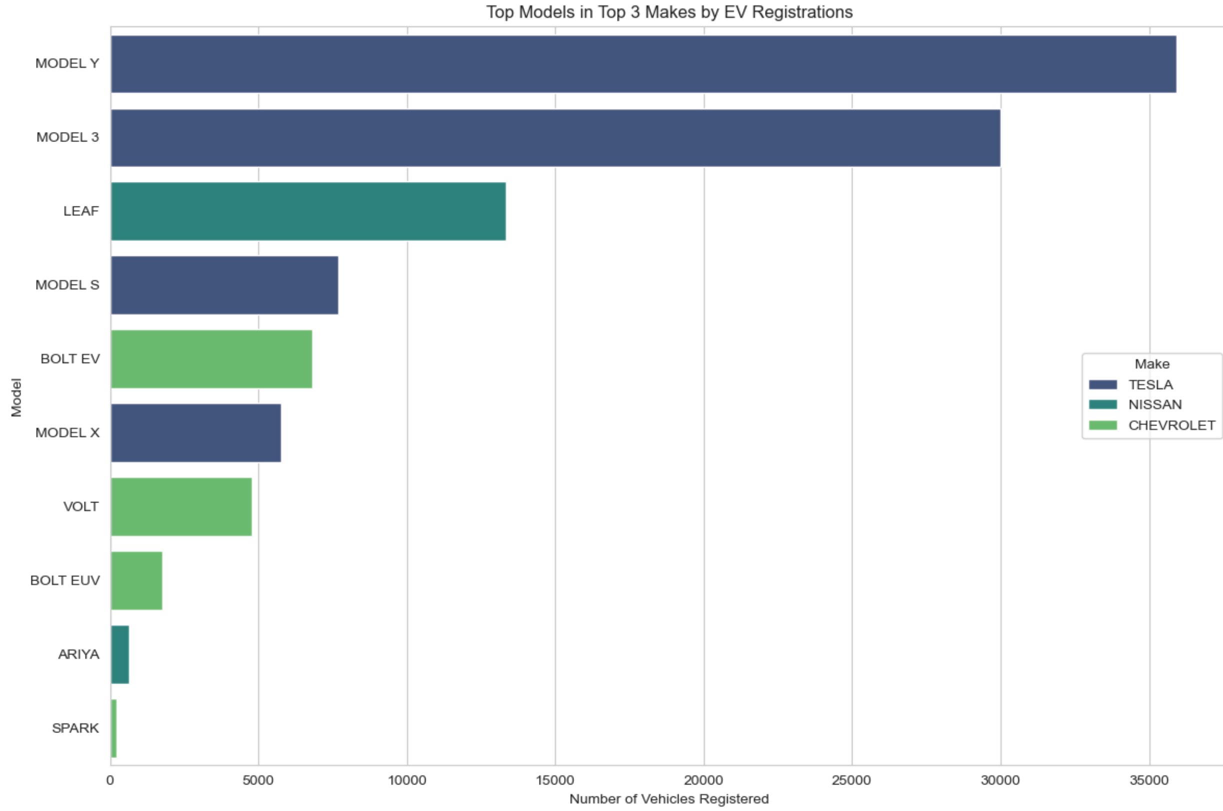


- TESLA leads by a substantial margin with the highest number of vehicles registered.
- NISSAN is the second most popular manufacturer, followed by CHEVROLET, though both have significantly fewer registrations than TESLA.
- FORD, BMW, KIA, TOYOTA, VOLKSWAGEN, JEEP, and HYUNDAI follow in decreasing order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

Top 3 manufacturers based on the number of vehicles registered

In [12]:

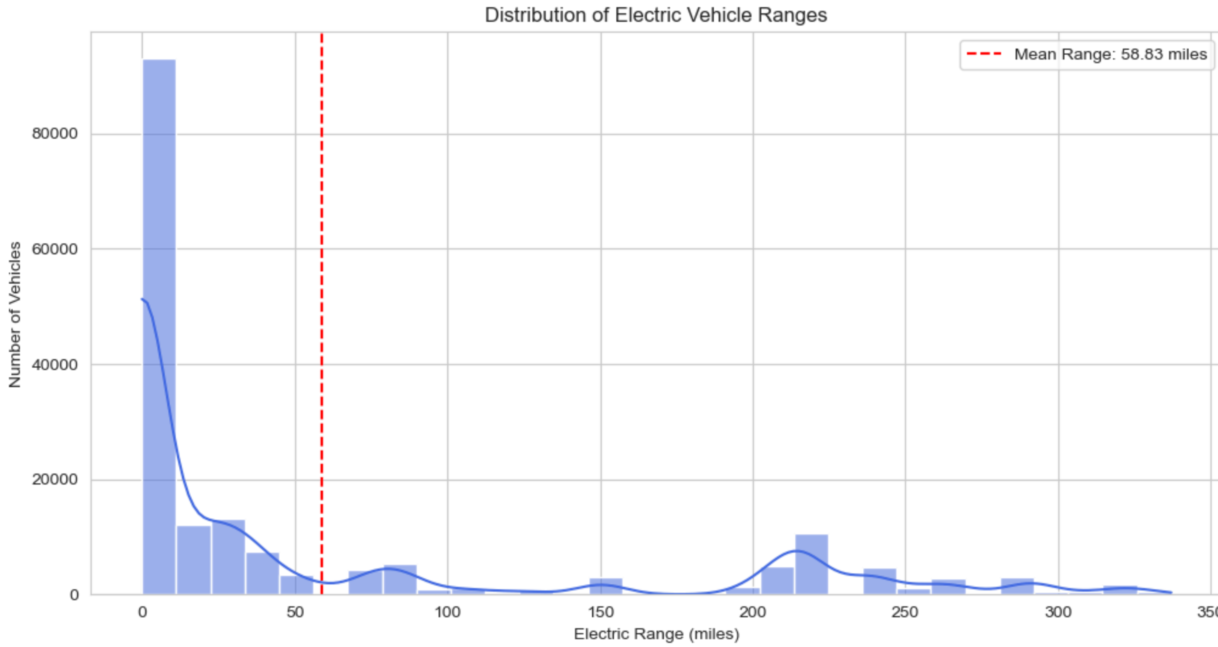
```
top_3_makes = ev_make_distribution.head(3).index

# filtering the dataset for these top manufacturers
top_makes_data = data[data['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().sor

# 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="v
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()
```



- 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.

Distribution of electric range

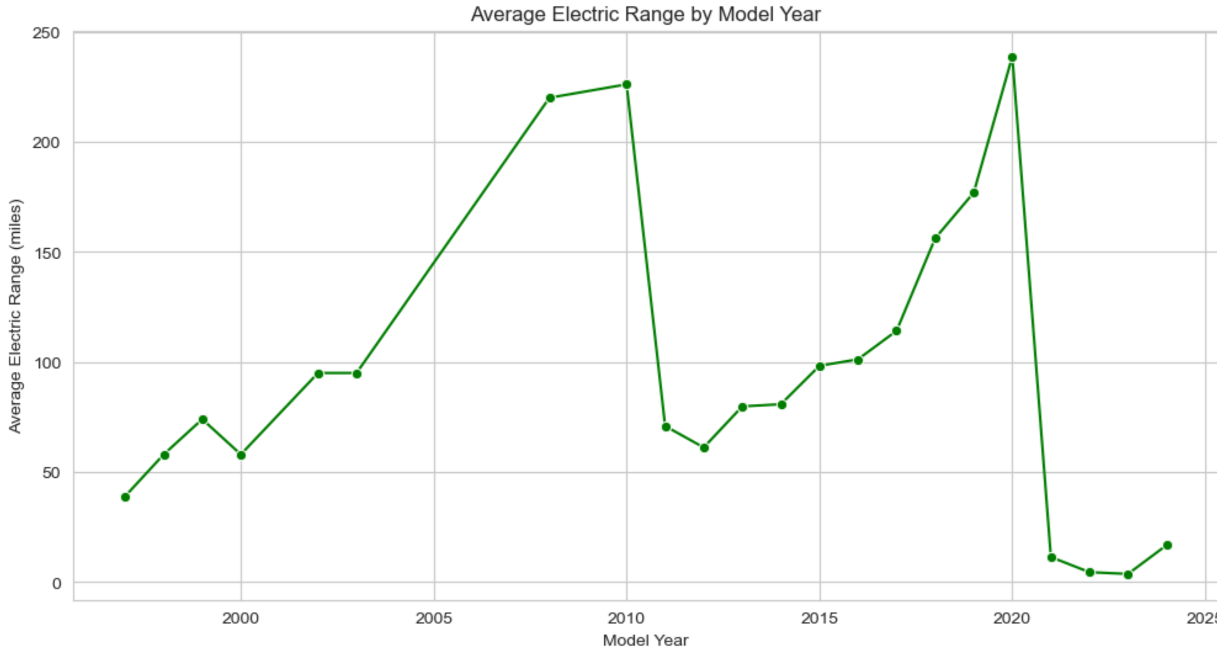
In [13]:
```
plt.figure(figsize=(12, 6))
sns.histplot(data['Electric Range'], bins=30, kde=True, color='royalblue')
plt.title('Distribution of Electric Vehicle Ranges')
plt.xlabel('Electric Range (miles)')
plt.ylabel('Number of Vehicles')
plt.axvline(data['Electric Range'].mean(), color='red', linestyle='--', label=f'Mean R
plt.legend()
plt.show()
```



- 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.

Average electric range by model year

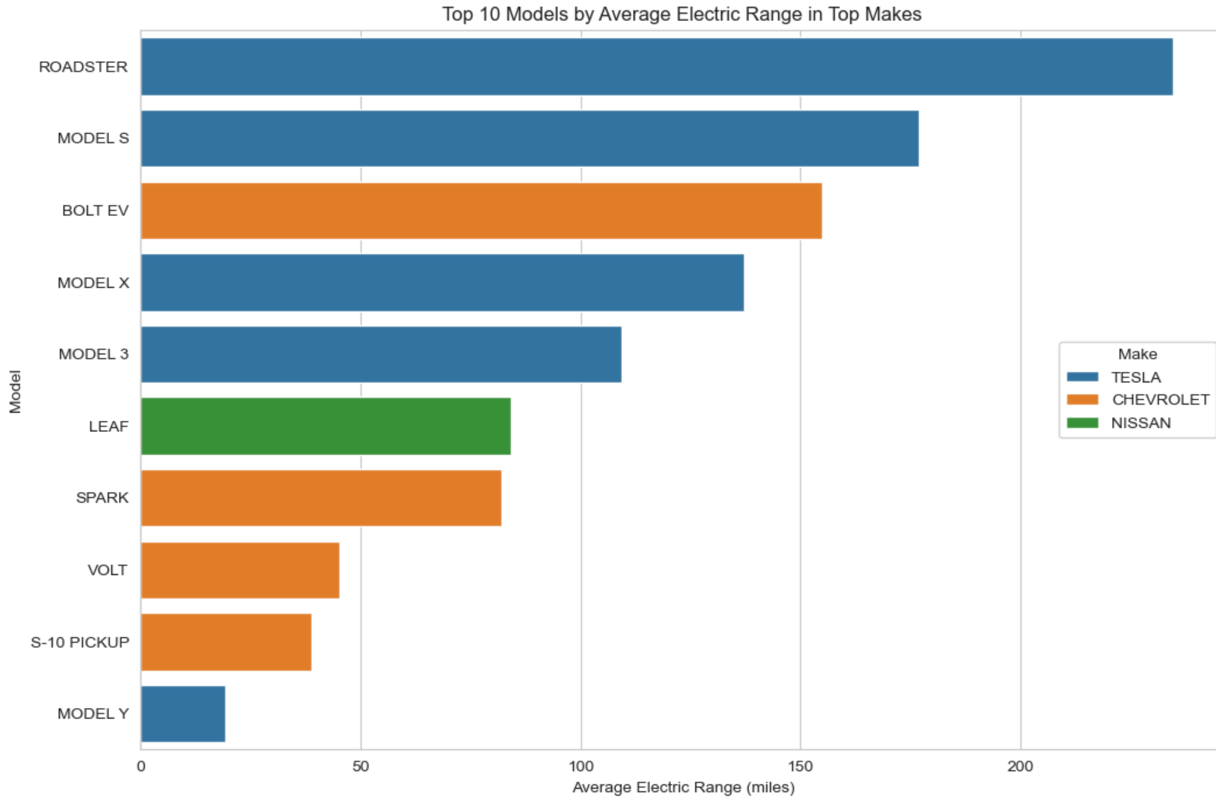
In [14]:
```
average_range_by_year = data.groupby('Model Year')['Electric Range'].mean().reset_inde

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



- 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.

Top 10 Models by Average Electric Range in Top Makes

In [16]:
```
average_range_by_model = top_makes_data.groupby(['Make', 'Model'])['Electric Range'].m
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
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.

Estimated Market Size Analysis of Electric Vehicles in the United States

In [19]:
```
# calculate the number of EVs registered each year
ev_registration_counts = data['Model Year'].value_counts().sort_index()
ev_registration_counts
```

Out[19]:
```
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
```

- 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.

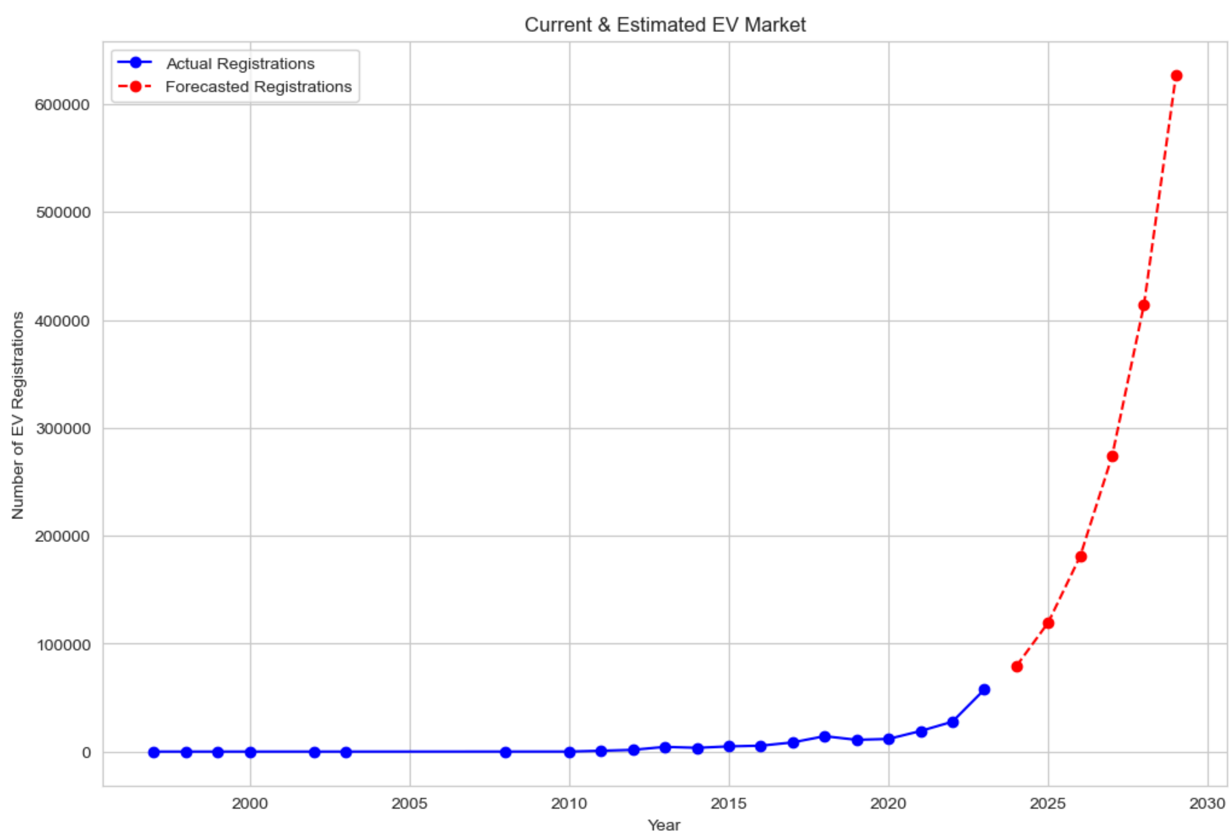
In [21]:
```python
from scipy.optimize import curve_fit
import numpy as np
```

In [22]:
```python
# filter the dataset to include years with complete data, assuming 2023 is the last co
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 yea
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_valu

print(forecasted_evs)
```

```
{2024: 79079.20808938889, 2025: 119653.96274428742, 2026: 181047.22020265696, 2027: 2
73940.74706208805, 2028: 414497.01805382164, 2029: 627171.3128407666}
```

Estimated market size data

In [23]:
```python
years = np.arange(filtered_years.index.min(), 2029 + 1)
actual_years = filtered_years.index
forecast_years_full = np.arange(2024, 2029 + 1)

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 Regist

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.

Conclusion

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.