General Assembly

Data Analytics Immersive Bootcamp Fellow

ELECTRIC VEHICLES ANALYSIS

Capstone Project

by

Maryam Mohamed Ali Abdulkarim Mohamed

September 29 - October 20, 2025

Acknowledgements

I would like to thank my instructors, Christopher Harvell, Husain Amer, Mohamad Hindi and my associate instructors, Mohamed Ahmed and Siddiqa Shubbar, for their constant guidance and support throughout this Bootcamp. Their help was essential at every stage of my work. I dedicate this thesis to my late grandmother, who would have been proud of my achievements. My sincere thanks also go to my family and friends for their endless encouragement and support through every challenge. Finally, I thank everyone who inspired or helped me, this work would not have been possible without you.

Abstract

In this report, we study the global market for electric vehicles (EVs) by looking at its size, growth, between major cities and regional differences. Using past sales data, we analyze how fast the EV market is expanding, which regions and cities lead in growth, and what factors drive it, such as government policies, better technology, and charging infrastructure. The results show a strong rise in EV adoption and changes in top regions, along with challenges like limited charging stations and high costs in developing areas. These findings can help governments, manufacturers, and investors focus on the right regions, improve their plans, and support sustainable growth in the EV industry.

Contents

1 Introduction

In recent years, the demand for electric vehicles (EVs) has been growing slowly because they produce less pollution. Transport emissions have been rising faster than in most other sectors. In this study, EV data is examined for insights into trends, performance, and the main reasons for adoption such as vehicle type, range, and location. Governments support EVs through funding, new technology, and better infrastructure. However, the number of charging points is limited, and the price of EVs is substantial for many countries. These issues must be addressed before EVs can become widely accepted. The results can help improve EV design, reduce costs, expand infrastructure, and support the global shift to cleaner, greener transportation.

2 Problem Statement

Between 1990 and 2022, transport emissions grew by about 1.7% each year, faster than most other sectors. The use of electric vehicles (EVs) is the cleanest alternative, which can contribute to the reduction of emissions and thus better air quality, among other environmental benefits. Unfortunately, the majority of the public and even the legislators have not yet grasped the whole picture regarding the EV adoption and its driving forces. High EV prices, limited charging stations. This project analyzes EV market data from 2011–2024 to find the main drivers of sales, efficiency, and adoption, and to provide insights for reducing global emissions.

3 Objectives and Target Audience

The underlying purpose is to contribute to the development of cleaner transport and the improvement of air quality. Specifically, the project will:

- Analyze the EV data to determine the main sales and performance trends.
- Investigate the influence of EV attributes such as range on adoption.
- Present the findings through visualizations and summaries that are simple to comprehend.
- Provide insights that will make it easier for people to switch to EVs.

The report is useful for:

- **Automakers and dealers:** to plan EV strategies.
- **Government:** to formulate EV regulations and plan the location of charging stations.
- **NGOs and academics:** to track EV adoption and its environmental impact.
- **Public:** to learn about the benefits of EVs and enjoy cleaner air.

4 Dataset

Electric Vehicle Population Dataset Columns (150483 x 17)
Data Link: https://www.kaggle.com/datasets/willianoliveiragibin/electric-vehicle-population/data

Column Name	Description
VIN (1-10)	First 10 characters of Vehicle ID
County	Registration county
City	Registration city
State	Registration state
Postal Code	ZIP code
Model Year	Vehicle manufacturing year
Make	Manufacturer (e.g., Tesla)
Model	Vehicle model
Electric Vehicle Type	Battery Electric or Plug-in Hybrid
Clean Alternative Fuel Vehicle (CAFV) Eligibility	CAFV Eligibility
Electric Range	Miles per charge
Base MSRP	Manufacturer's suggested price
Legislative District	Numbered district of registration
DOL Vehicle ID	Unique ID from licensing
Vehicle Location	Latitude & longitude
Electric Utility	Charging power provider
2020 Census Tract	Census tract code

5 Data Handling

I dropped the County column because the information was incorrect. I also removed Base MSRP because it was unnecessary and contained errors. I checked for duplicates and found none. Then, I examined null values three rows contained nulls in several columns, including the most important one, City, so I dropped those rows. There were 341 nulls in Legislative District and 7 in Vehicle Location, but I kept them since those columns were not used in my analysis. I added a new column, State Name, using the VLOOKUP function. I also created two columns: Electric_Vehicle_Type_Code (for the code) and Electric_Vehicle_Type (for the name). Additionally, I added Clean_Alternative_Fuel_Vehicle_Eligibility, where I classified entries as Eligible, Not Eligible, or Unknown.

6 Analysis and Findings

- **EV Adoption Over Time (2011–2024):** EV registrations have grown steadily. In 2011, 796 cars were registered, rising to 37,079 in 2023, showing a gradual shift toward cleaner transport.

- **EV Type Distribution:** Battery electric vehicles (BEVs) dominate with 116,804 cars. Plug-in hybrids (PHEVs) are fewer, totaling 33,675, making BEVs the most preferred type.

- **Vehicle Eligibility for Clean Alternative Fuel:** 69,698 vehicles are "Unknown," 62,948 are "Eligible," and 17,833 are "Not Eligible," showing a need for better classification and awareness.

- **Top Cities by EV Registrations:** Seattle leads EV registrations, followed by Bellevue and Redmond. Other high-registration cities include Vancouver, Bothell, Kirkland, and Sammamish.

- **Top Manufacturers:** Tesla leads with 68,981 vehicles, followed by Nissan (13,496) and Chevrolet (12,026). Other notable brands are Ford, BMW, KIA, and Toyota.

- **Most Popular Models:** Tesla Model Y (28,502) and Model 3 (27,708) are the most common. Nissan Leaf and Tesla Model S follow, showing a preference for long-range, efficient EVs.

- **Average EV Range Over Time (2011–2024):** EV ranges have improved greatly. In 2011, the average range was about 71 miles, rising to over 240 miles by 2020, showing advances in battery technology.

- **EV Range by Model:** Tesla models have the longest ranges. For example, Tesla Roadster reaches 235 miles, while Chevrolet Spark only goes 82 miles.

- **EV Market Growth in the U.S.:** Actual registrations from 2011–2023 and forecasts to 2029 show continuous growth, expected to exceed 175,000 registrations by 2029.

In [1]:

```
import pandas as pd
```

In [2]:

```
import numpy as np
```

In [3]:

```
import matplotlib.pyplot as plt
```

In [44]:

```
import matplotlib.ticker as mticker
```

In [5]:

```
import seaborn as sns
```

In [80]:

```
EV = pd.read_excel("ALLDATA.xlsx", sheet_name="Electric_Vehicle_Population_c")
```

In [7]:

```
EV.tail(1)
```

Out[7]:

	VIN (1-10)	City	State_Name	State	Postal Code	Model Year	Make	Model	Electric_Vehicle_Type_Code	Electri
150479	0	0	NaN	0	0	0	0	0	NaN	

In [8]:

```
EV = EV.iloc[:-1] # Removing the last row as i added it to count the null while cleaning
```

In [9]:

```
EV.shape
```

Out[9]:

```
(150479, 20)
```

In [10]:

```
EV.dtypes
```

Out[10]:

```
VIN (1-10)                                             object
City                                                   object
State_Name                                             object
State                                                  object
Postal Code                                             int64
Model Year                                              int64
Make                                                   object
Model                                                  object
Electric_Vehicle_Type_Code                             object
Electric_Vehicle_Type                                  object
Electric Vehicle Type                                  object
Clean_Alternative_Fuel_Vehicle_Eligibility             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                                         int64
dtype: object
```

In [11]:

```
EV.columns
```

Out[11]:

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

In [71]:

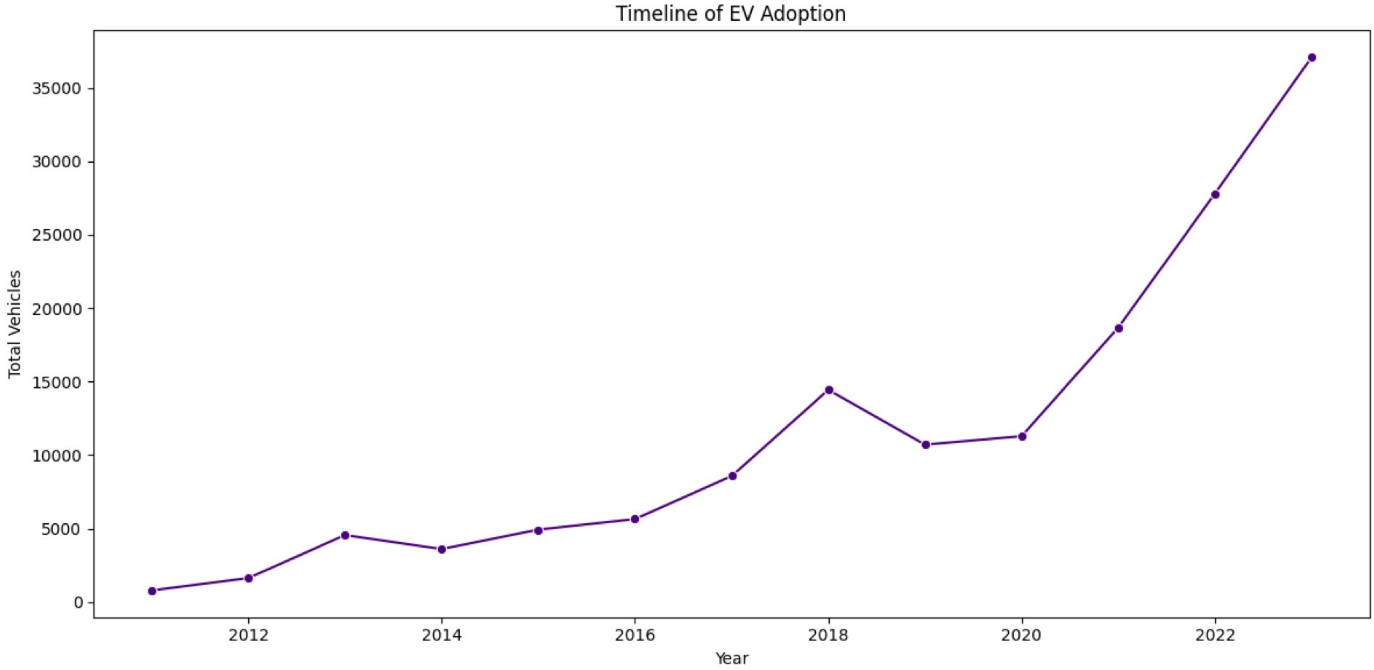
```
EV[EV["Model Year"] != 2024]["Model Year"].value_counts()
```

Out[71]:

```
Model Year
2023    37079
2022    27799
2021    18684
2018    14441
2020    11294
2019    10716
2017     8574
2016     5650
2015     4934
2013     4565
2014     3613
2012     1633
2011      796
2010       24
2008       18
2000        8
1999        4
2002        2
1998        1
2003        1
1997        1
Name: count, dtype: int64
```

In [74]:

```
plt.figure(figsize=(12,6))
EV_year = EV[(EV["Model Year"] > 2010) & (EV["Model Year"] < 2024)]["Model Year"].value_
sns.lineplot(x=EV_year.index, y=EV_year.values, marker='o', color='#4B0082')
plt.title('Timeline of EV Adoption')
plt.xlabel('Year')
plt.ylabel("Total Vehicles")
plt.tight_layout()
plt.show()
```



In [75]:

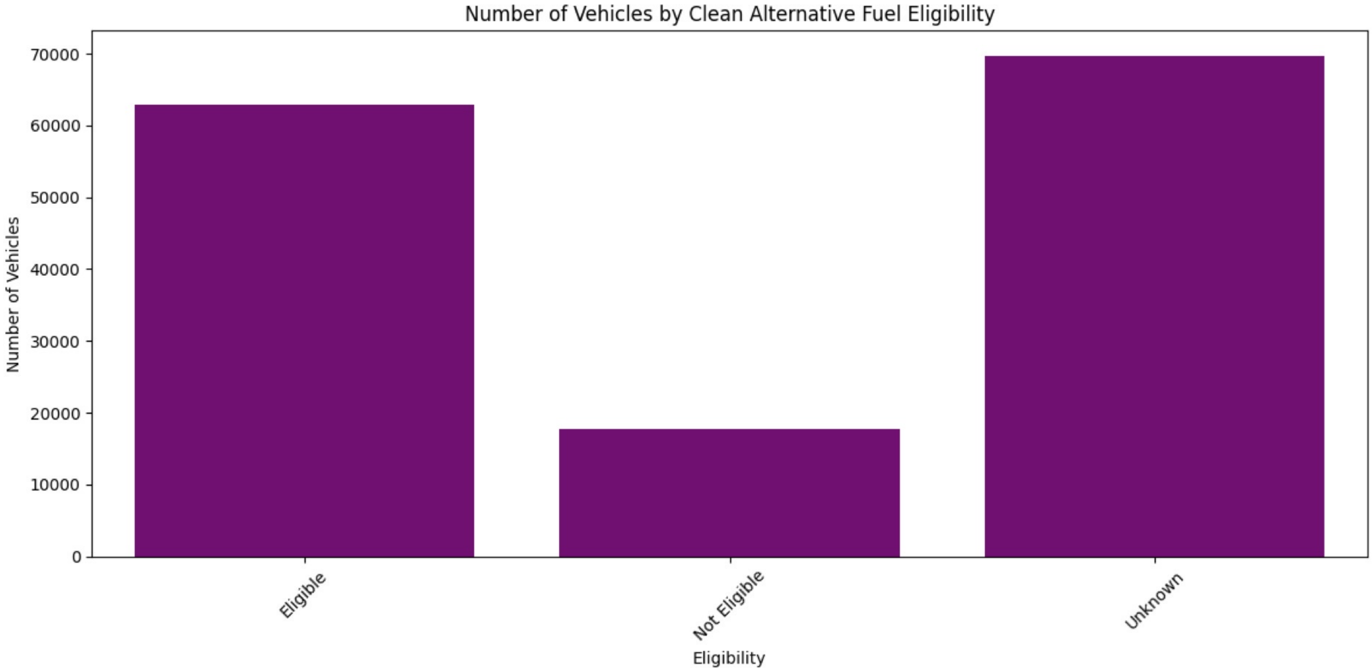
```
EV_type = EV["Electric_Vehicle_Type_Code"].value_counts().reset_index()
EV_type.columns = ['Electric Vehicle Type', 'Total Vehicles']
EV_type
```

Out[75]:

	Electric Vehicle Type	Total Vehicles
0	BEV	116804
1	PHEV	33675

In [76]:

```
EV_type = EV["Electric_Vehicle_Type_Code"].value_counts()
plt.figure(figsize=(12,6))
sns.barplot(
    x=EV_type.values,
    y=EV_type.index,
    hue=EV_type.index,
    palette=["#4B0082", "#E6E6FA"],
)
plt.title('Total Electric Vehicles by Type')
plt.xlabel('Total Vehicles')
plt.ylabel('Electric Vehicle Type')
plt.tight_layout()
plt.show()
```



In [18]:

```
ev_city = (EV.groupby('City').size().sort_values(ascending=False)
    .reset_index(name='Total Vehicles'))
top_city = ev_city.head(10)
top_city
```

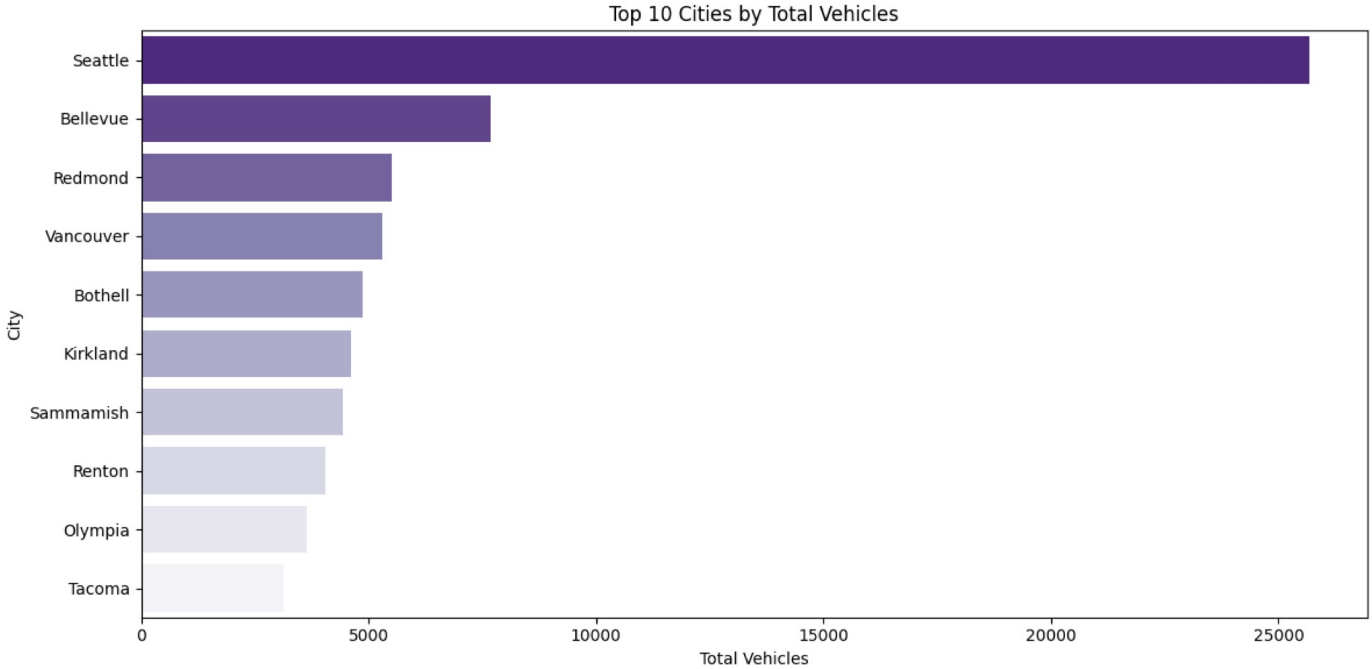
Out[18]:

	City	Total Vehicles
0	Seattle	25675
1	Bellevue	7691
2	Redmond	5502
3	Vancouver	5310
4	Bothell	4861
5	Kirkland	4622
6	Sammamish	4436
7	Renton	4043
8	Olympia	3634
9	Tacoma	3121

In [20]:

```
plt.figure(figsize=(12,6))
sns.barplot(
    x='Total Vehicles',
    y='City',
    hue='City',
    data=top_city,
    palette=sns.color_palette("Purples_r",n_colors=10),
    dodge=False,
    legend=False
)
```

```
plt.title('Top 10 Cities by Total Vehicles')
plt.xlabel('Total Vehicles')
plt.ylabel('City')
plt.tight_layout()
plt.show()
```



In [21]:

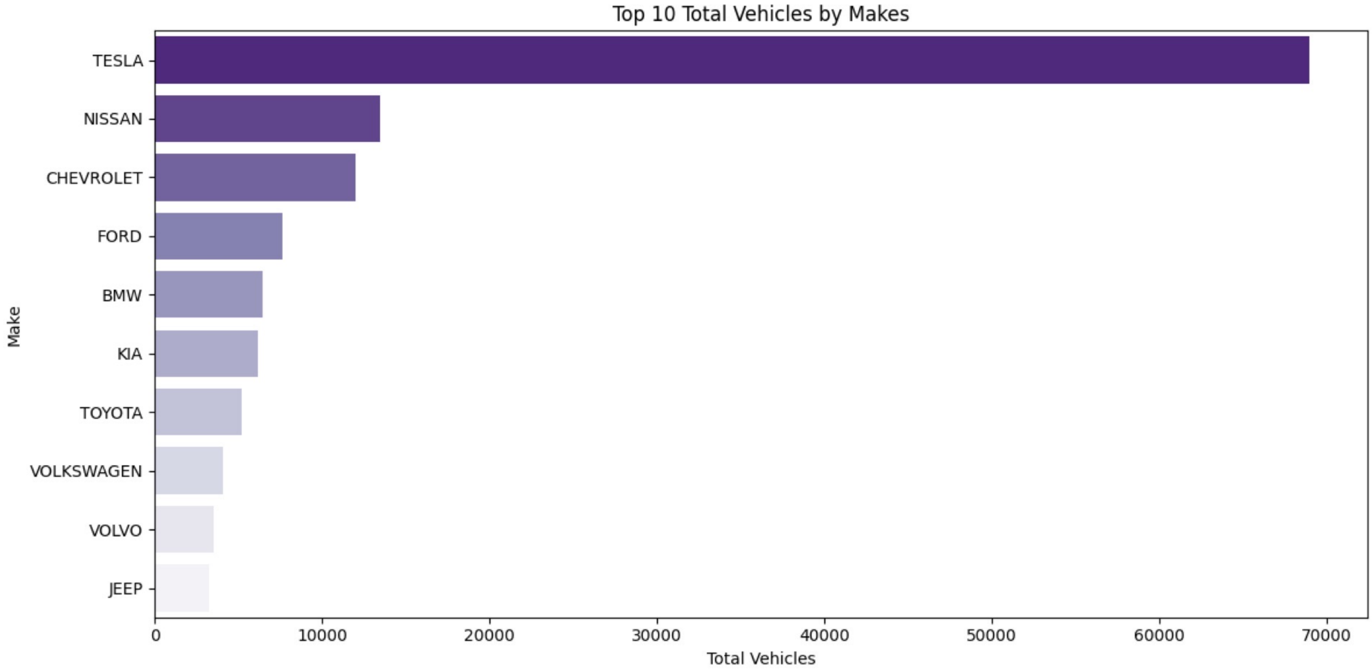
```
EV_make = EV["Make"].value_counts().head(10)
EV_make
```

Out[21]:

```
Make
TESLA         68981
NISSAN        13496
CHEVROLET     12026
FORD           7614
BMW            6439
KIA            6198
TOYOTA         5223
VOLKSWAGEN     4074
VOLVO          3536
JEEP           3292
Name: count, dtype: int64
```

In [22]:

```
plt.figure(figsize=(12,6))
sns.barplot(
    x=EV_make.values,
    y=EV_make.index,
    hue=EV_make.index,
    palette=sns.color_palette("Purples", n_colors=10)[::-1]
)
plt.title('Top 10 Total Vehicles by Makes')
plt.xlabel("Total Vehicles")
plt.ylabel('Make')
plt.tight_layout()
plt.show()
```



In [23]:

```
top_3_make = EV_make.head(3).index
top_make = EV[EV['Make'].isin(top_3_make)]

ev_model_top_make=(top_make.groupby(['Make', 'Model']).size().sort_values(ascending=Fals
    .head(10).reset_index(name='Count'))
ev_model_top_make
```

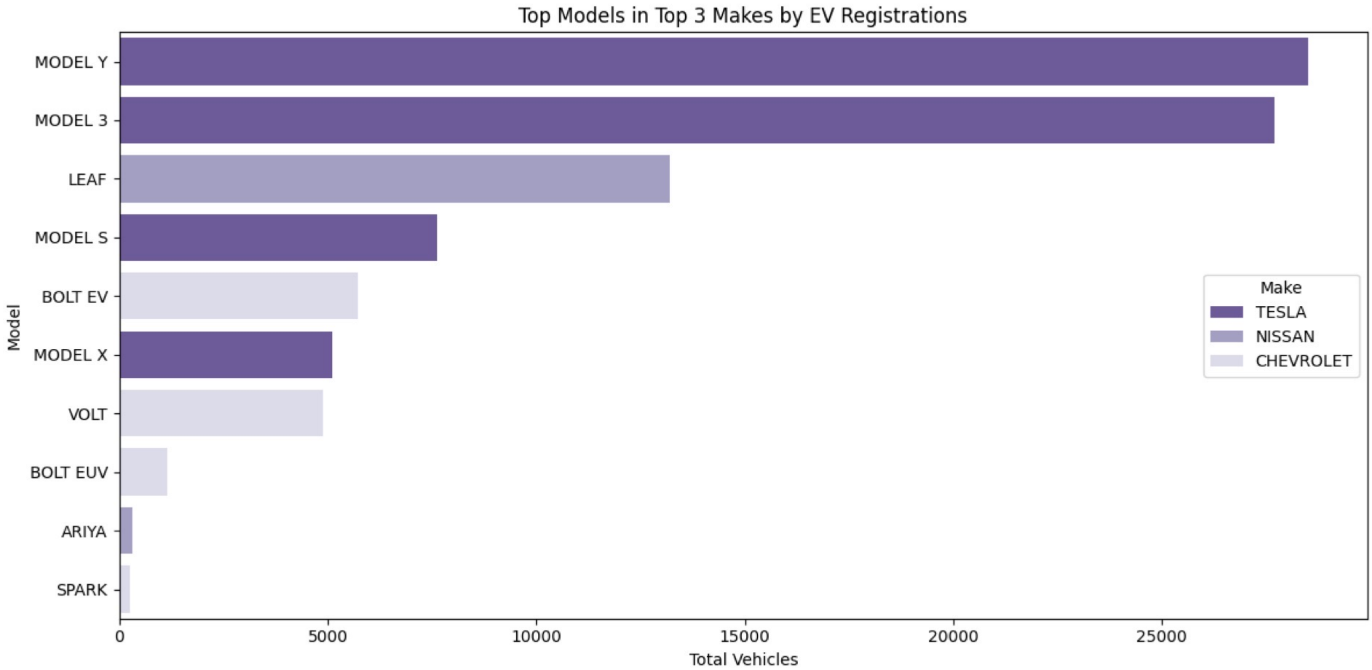
Out[23]:

	Make	Model	Count
0	TESLA	MODEL Y	28502
1	TESLA	MODEL 3	27708
2	NISSAN	LEAF	13186
3	TESLA	MODEL S	7611
4	CHEVROLET	BOLT EV	5733
5	TESLA	MODEL X	5114
6	CHEVROLET	VOLT	4890
7	CHEVROLET	BOLT EUV	1153
8	NISSAN	ARIYA	310
9	CHEVROLET	SPARK	249

In [24]:

```
plt.figure(figsize=(12, 6))
sns.barplot(
    data=ev_model_top_make,
    x='Count',
    y='Model',
    hue='Make',
    palette=sns.color_palette("Purples", n_colors=len(top_3_make))[::-1],
    dodge=False
```

```
)
plt.title('Top Models in Top 3 Makes by EV Registrations')
plt.xlabel('Total Vehicles')
plt.ylabel('Model')
plt.legend(title='Make', loc='center right')
plt.tight_layout()
plt.show()
```



In [25]:

```
EV['Electric Range'].mean()
```

Out[25]:

67.8757700410024

In [26]:

```
counts, bin_edges = np.histogram(EV['Electric Range'], bins=30)
hist_data = pd.DataFrame({
    'Bin Start': bin_edges[:-1],
    'Bin End': bin_edges[1:],
    'Vehicle Count': counts
})
hist_data
```

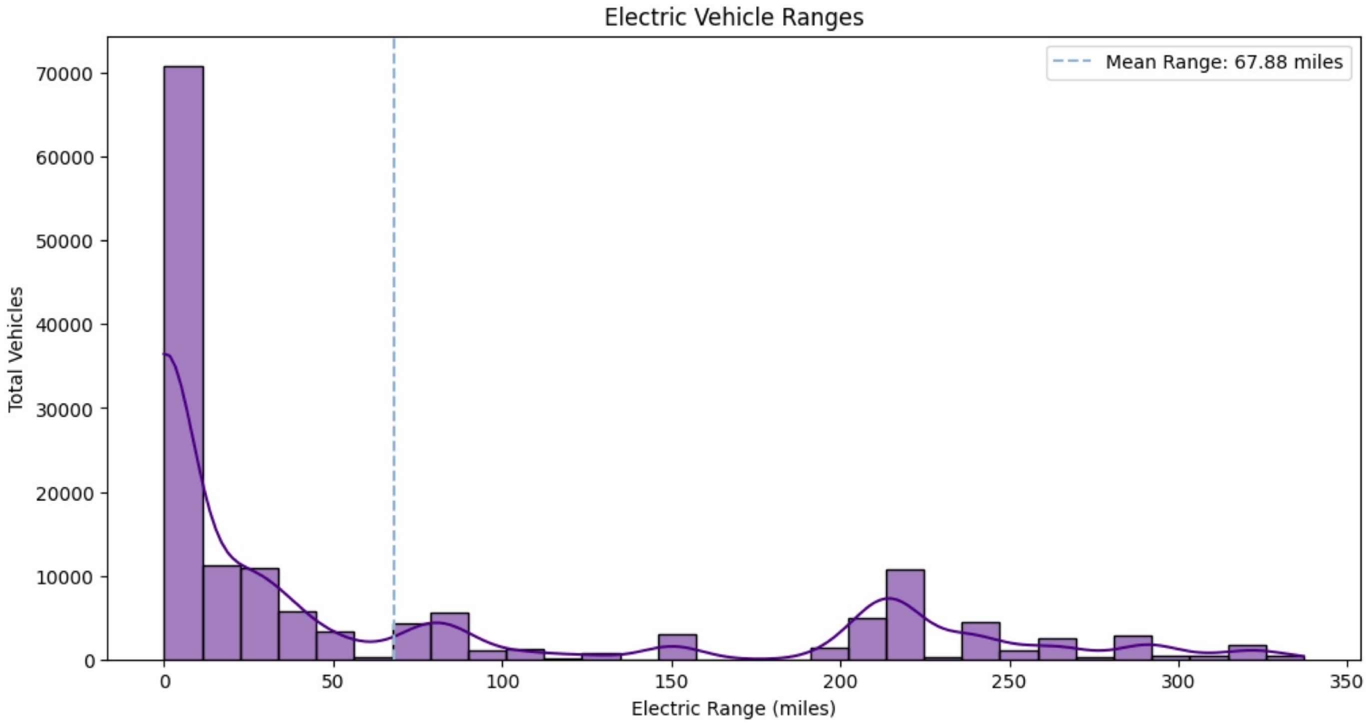
Out[26]:

	Bin Start	Bin End	Vehicle Count
0	0.000000	11.233333	70861
1	11.233333	22.466667	11263
2	22.466667	33.700000	10829
3	33.700000	44.933333	5788
4	44.933333	56.166667	3323
5	56.166667	67.400000	222
6	67.400000	78.633333	4271
7	78.633333	89.866667	5608

	Bin Start	Bin End	Vehicle Count
8	89.866667	101.100000	1034
9	101.100000	112.333333	1173
10	112.333333	123.566667	103
11	123.566667	134.800000	822
12	134.800000	146.033333	0
13	146.033333	157.266667	3081
14	157.266667	168.500000	0
15	168.500000	179.733333	23
16	179.733333	190.966667	0
17	190.966667	202.200000	1411
18	202.200000	213.433333	4997
19	213.433333	224.666667	10820
20	224.666667	235.900000	318
21	235.900000	247.133333	4487
22	247.133333	258.366667	1132
23	258.366667	269.600000	2609
24	269.600000	280.833333	269
25	280.833333	292.066667	2926
26	292.066667	303.300000	450
27	303.300000	314.533333	513
28	314.533333	325.766667	1747
29	325.766667	337.000000	399

In [27]:

```
plt.figure(figsize=(12,6))
sns.histplot(EV['Electric Range'],
             bins=30,
             kde=True,
             color='#4B0082')
plt.title('Electric Vehicle Ranges')
plt.xlabel('Electric Range (miles)')
plt.ylabel('Total Vehicles')
plt.axvline(EV['Electric Range'].mean(),
            color='#8eb5d4',
            linestyle='--',
            label=f'Mean Range: {EV["Electric Range"].mean():.2f} miles')

plt.legend()
plt.show()
```



In [28]:

```
avg_range_year = EV.groupby('Model Year')['Electric Range'].mean().reset_index()
avg_range_year = avg_range_year[avg_range_year['Model Year'] >2010]
avg_range_year
```

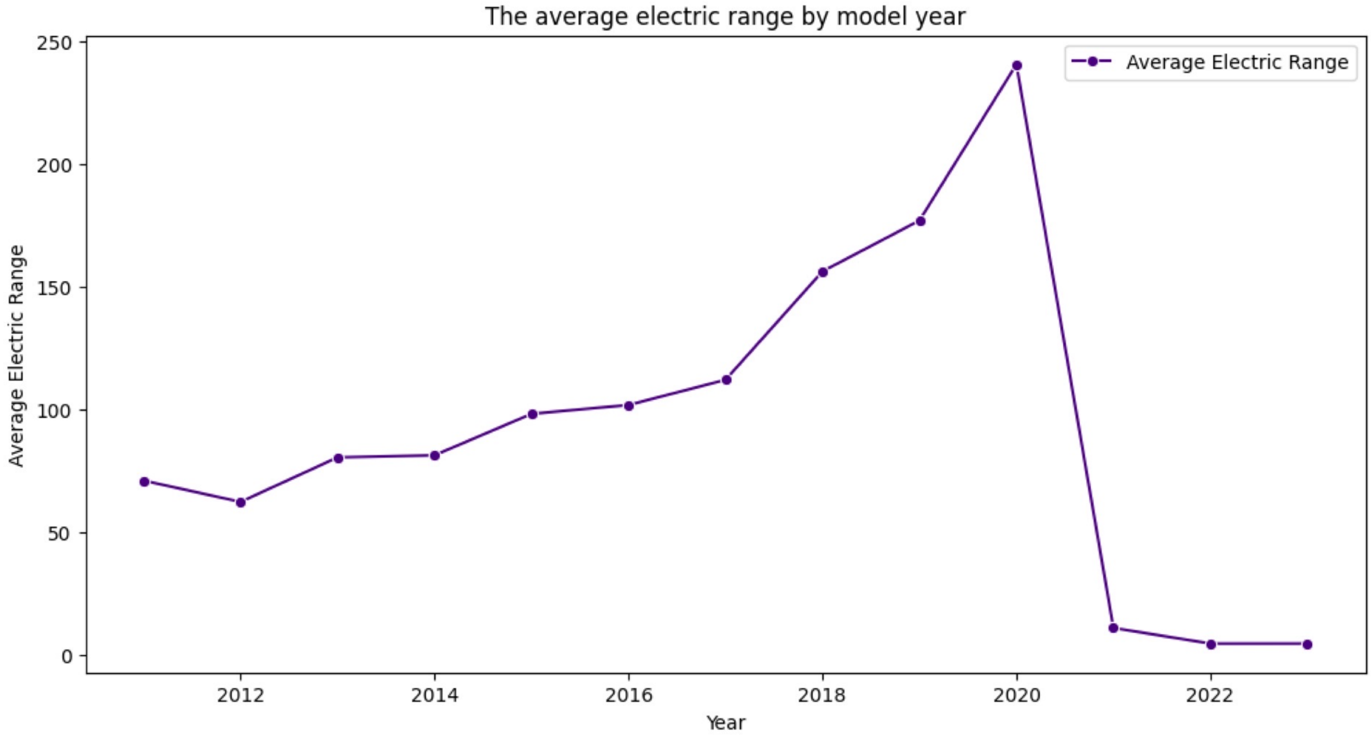
Out[28]:

	Model Year	Electric Range
8	2011	70.851759
9	2012	62.186160
10	2013	80.319387
11	2014	81.172710
12	2015	98.060195
13	2016	101.651858
14	2017	111.916608
15	2018	156.087875
16	2019	176.810284
17	2020	240.410926
18	2021	10.812942
19	2022	4.470449
20	2023	4.479409
21	2024	17.683801

In [77]:

```
plt.figure(figsize=(12,6))
sns.lineplot(
    x='Model Year',
```

```
    y='Electric Range',
    data=avg_range_year[avg_range_year['Model Year'] != 2024],
    marker='o',
    color='#4B0082',
    label='Average Electric Range'
)

plt.title('The average electric range by model year')
plt.xlabel('Year')
plt.ylabel('Average Electric Range')

plt.legend()
plt.show()
```



In [30]:

```
average_range_model = (top_make.groupby(['Make', 'Model'])['Electric Range'].mean()
    .sort_values(ascending=False).reset_index())

average_range_model.head(10)
```

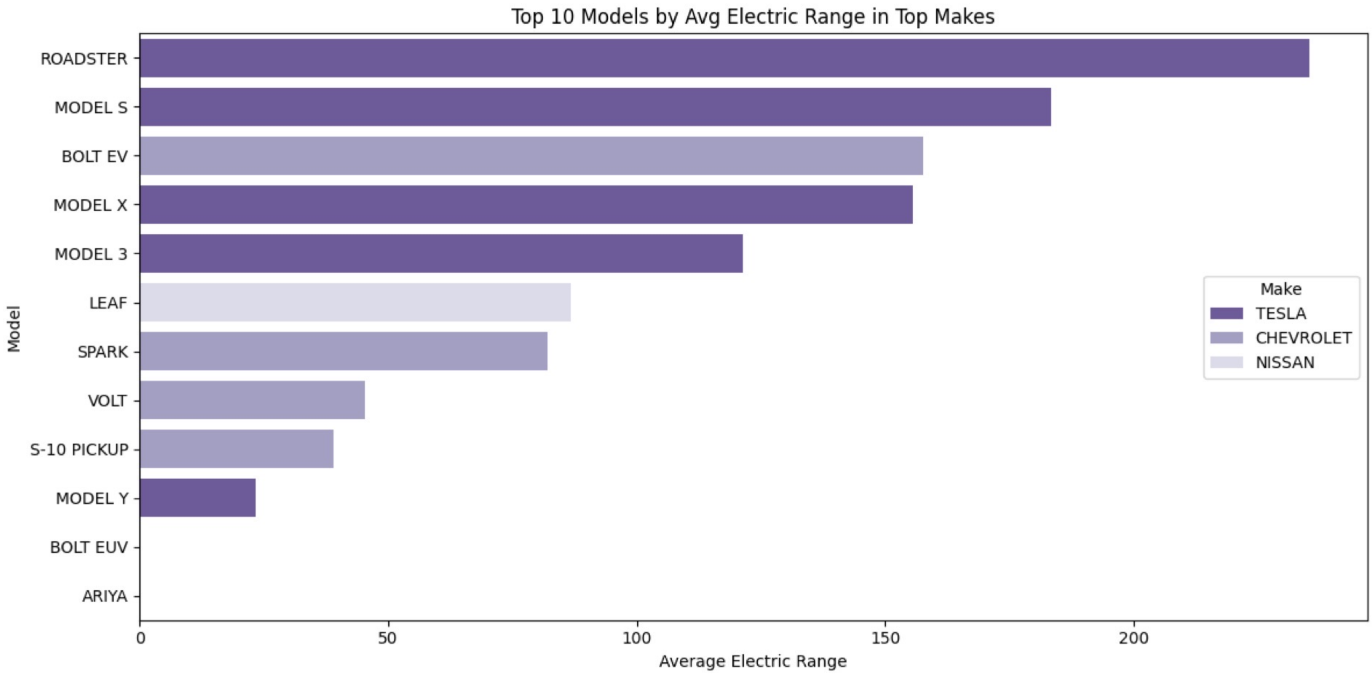
Out[30]:

	Make	Model	Electric Range
0	TESLA	ROADSTER	235.217391
1	TESLA	MODEL S	183.456182
2	CHEVROLET	BOLT EV	157.583639
3	TESLA	MODEL X	155.539304
4	TESLA	MODEL 3	121.375884
5	NISSAN	LEAF	86.717503
6	CHEVROLET	SPARK	82.000000
7	CHEVROLET	VOLT	45.358896

	Make	Model	Electric Range
8	CHEVROLET	S-10 PICKUP	39.000000
9	TESLA	MODEL Y	23.257947

In [31]:

```
plt.figure(figsize=(12,6))
sns.barplot(
    x='Electric Range',
    y='Model',
    hue='Make',
    data=average_range_model,
    palette=sns.color_palette("Purples", n_colors=len(average_range_model['Make'].unique
    dodge=False
)
plt.title("Top 10 Models by Avg Electric Range in Top Makes")
plt.xlabel("Average Electric Range")
plt.ylabel("Model")
plt.legend(title='Make', loc='center right')
plt.tight_layout()
plt.show()
```



In [78]:

```
ev_registration_counts = EV['Model Year'].value_counts().sort_index()
```

In [33]:

```
from scipy.optimize import curve_fit
import matplotlib.ticker as mticker

def exponential_growth(x, a, b): return a * np.exp(b * x)

the_filter = ev_registration_counts[(ev_registration_counts.index > 2010) & (ev_registra

x_vals=the_filter.index-the_filter.index.min()
y_vals=the_filter.values

fit_params, fit_cov = curve_fit(exponential_growth, x_vals, y_vals)
```

```
future_years = np.arange(2024, 2030)-the_filter.index.min()
forecasted_vals = exponential_growth(future_years, *fit_params)

ev_forecast_dict = dict(zip(future_years + the_filter.index.min(), forecasted_vals))
ev_forecast_dict
```

Out[33]:

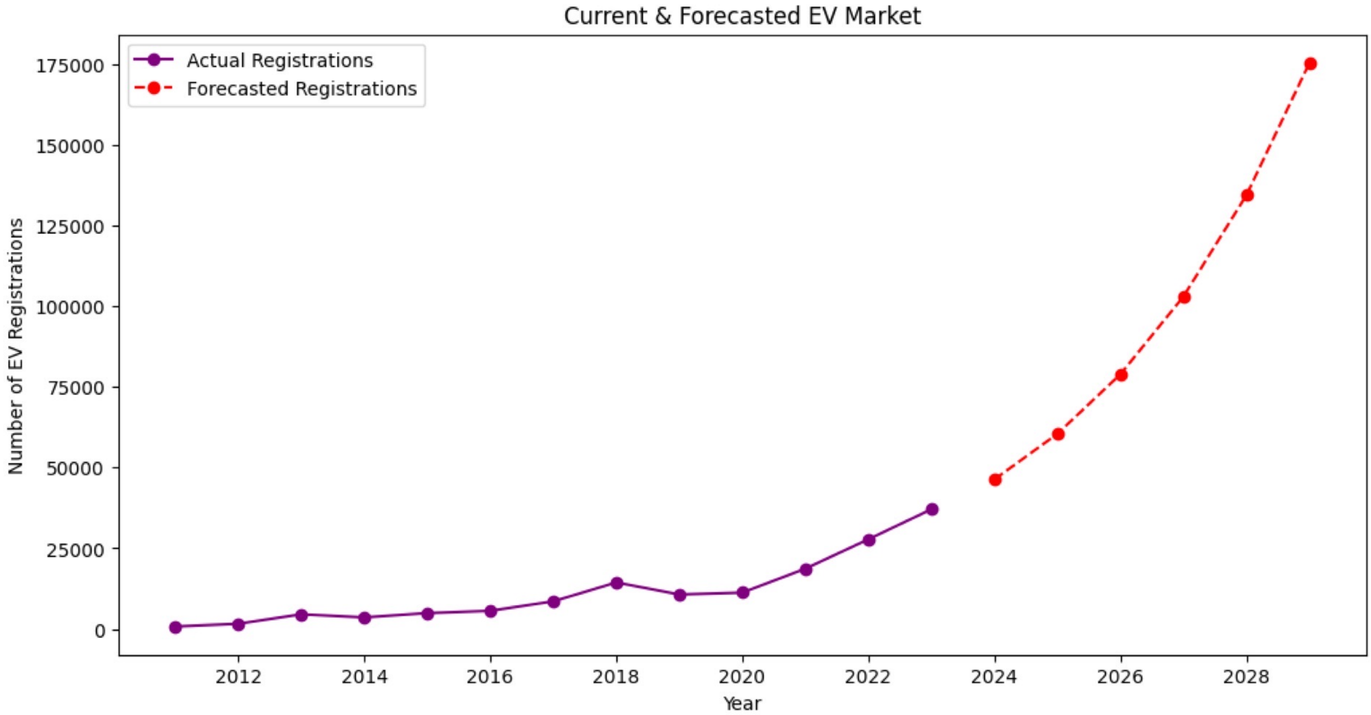
```
{2024: 46330.89713171541,
 2025: 60463.79298372942,
 2026: 78907.82368374843,
 2027: 102978.06887803155,
 2028: 134390.76348563237,
 2029: 175385.667133093}
```

In [45]:

```
all_years = np.arange(the_filter.index.min(), 2030)
actual_years = the_filter.index
forecast_years_plot = np.arange(2024, 2030)

actual_values = the_filter.values
forecast_values_plot = [ev_forecast_dict[year] for year in forecast_years_plot]
```

In [79]:

```
plt.figure(figsize=(12, 6))

plt.plot(actual_years, actual_values, color='#800080', marker='o', linestyle='-', label=
plt.plot(forecast_years_plot, forecast_values_plot, color='red', marker='o', linestyle='

plt.title('Current & Forecasted EV Market')
plt.xlabel('Year')
plt.ylabel('Number of EV Registrations')
plt.legend()
plt.grid(False)
plt.gca().xaxis.set_major_locator(mticker.MaxNLocator(integer=True))
plt.show()
```

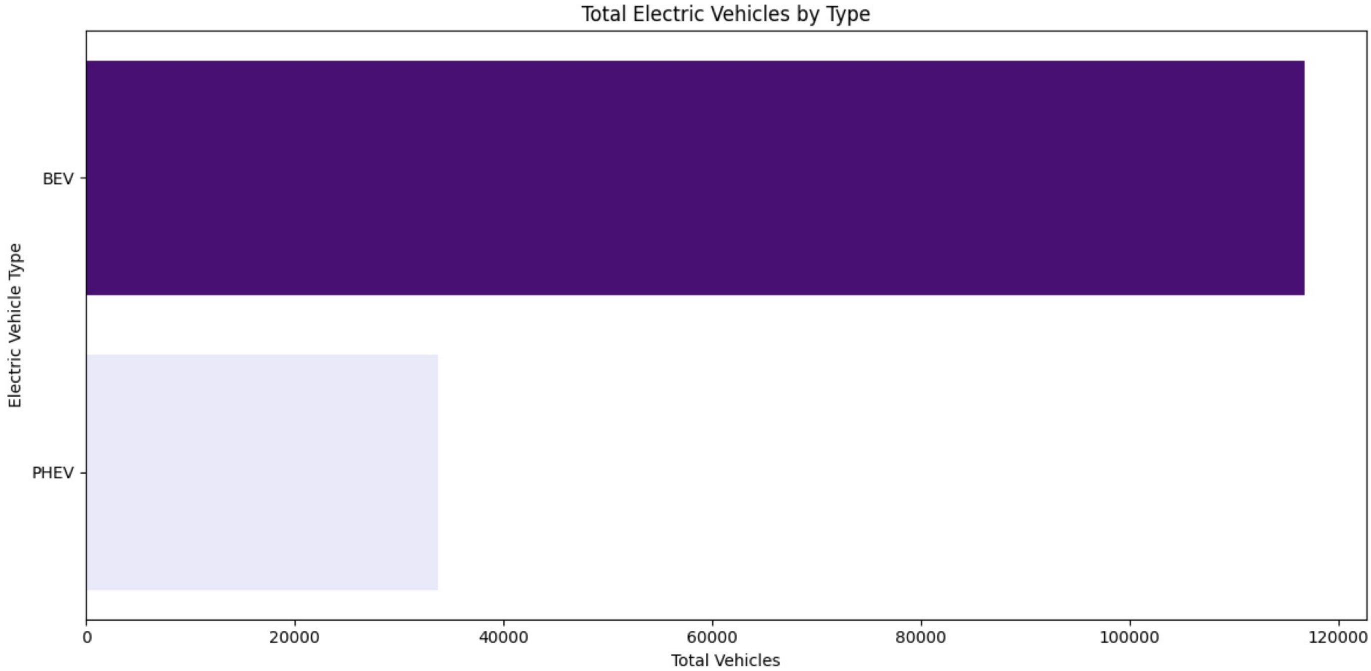


In [42]:

```python
forecast_years = np.arange(2024, 2030)
forecast_values = [ev_forecast_dict[year] for year in forecast_years]

actual_years=the_filter.index
actual_values= the_filter.values

actual_df= pd.DataFrame({
    'Year': actual_years,
    'EV_Registrations': actual_values})

forecast_df= pd.DataFrame({
    'Year': forecast_years,
    'Forecasted_EV_Registrations': forecast_values})

combined_df =pd.merge(actual_df, forecast_df, on='Year', how='outer')
combined_df = combined_df.sort_values('Year').reset_index(drop=True)

combined_df
```

	Year	EV_Registrations	Forecasted_EV_Registrations
0	2011	796.0	NaN
1	2012	1633.0	NaN
2	2013	4565.0	NaN
3	2014	3613.0	NaN
4	2015	4934.0	NaN
5	2016	5650.0	NaN
6	2017	8574.0	NaN
7	2018	14441.0	NaN
8	2019	10716.0	NaN
9	2020	11294.0	NaN
10	2021	18684.0	NaN
11	2022	27799.0	NaN
12	2023	37079.0	NaN
13	2024	NaN	46330.897132
14	2025	NaN	60463.792984
15	2026	NaN	78907.823684
16	2027	NaN	102978.068878
17	2028	NaN	134390.763486
18	2029	NaN	175385.667133

In [ ]:



In [16]:

```
vehicle_counts = EV.groupby('Clean_Alternative_Fuel_Vehicle_Eligibility').size().reset_i
vehicle_counts
```

Out[16]:

	Clean_Alternative_Fuel_Vehicle_Eligibility	Number_of_Vehicles
0	Eligible	62948
1	Not Eligible	17833
2	Unknown	69698

In [17]:

```
plt.figure(figsize=(12,6))
sns.barplot(
    data=vehicle_counts,
    x='Clean_Alternative_Fuel_Vehicle_Eligibility',
    y='Number_of_Vehicles',
    color='purple'
)
plt.xticks(rotation=45)
plt.title('Number of Vehicles by Clean Alternative Fuel Eligibility')
plt.ylabel('Number of Vehicles')
plt.xlabel('Eligibility')
plt.tight_layout()
plt.show()
```

7 Limitations and Assumptions

The focus is on the U.S. and major markets, and only commonly registered EV types (BEVs and PHEVs) were considered.

8 Recommendations

- Introduce support and reduce taxes for eco-friendly cars.
- Run EV awareness campaigns in cities with few EVs, like Tacoma, Olympia, and Renton.
- Make people excited about EVs and clean energy.
- Organize test-drive events and use influencers to promote EVs.
- Build more EV charging stations.
- Focus on areas where EV adoption can grow fast.

9 The Features

- Electricity is unlimited, and in the future we might not need gas at all.
- EVs are cheaper to fix and maintain, they don't need oil changes like normal cars.
- Gas prices change with global events, but electricity mainly depends on how much we use.
- EVs are safer in accidents because gas cars can catch fire easily.
- EVs are quiet, smooth, and comfortable to drive.
- Home charging is convenient.

10 Conclusion

This study focused on electric vehicles. The EVs are viewed as the prominent solution to tackle transportation emissions and improve the air quality. Their sales have increased in a number of countries, with battery electric vehicles (BEVs) being the main. The market has witnessed growth despite limited charging stations, and variable adoption rates. With the usage of EVs, establishment of more charging stations, public awareness, policy support, manufacturer assistance, and progress monitoring, the transition to electric cars can be accelerated for everyone. These steps will contribute to the generation of cleaner, greener transport and wider accessibility of EVs in the global market.

References

[1] IEA, 2023. *Transport.* [online] Available at: https://www.iea.org/energy-system/transport [Accessed 2 October 2025].

[2] IEA, 2024. *Electric Vehicles.* [online] Available at: https://www.iea.org/energy-system/transport/electric-vehicles [Accessed 20 October 2025].

[3] U.S. Department of Transportation, 2023. *EV Challenges and Evolving Solutions for Rural Communities.* [online] Available at: https://www.transportation.gov/rural/ev/toolkit/ev-benefits-and-challenges/challenges-and-evolving-solutions [Accessed 3 October 2025].

[4] World Health Organization, 2024. *Air Pollution.* [online] Available at: https://www.who.int/health-topics/air-pollution#tab=tab_1 [Accessed 5 October 2025].

[5] November, C.E. and 2022, 2022. *Growing demand for electric cars: Are you ready to make the switch?.* [online] Available at: https://www.currencytransfer.com/blog/expert-analysis/growing-demand-electric-cars-2022 [Accessed 9 October 2025].