In [1]:
```
import pandas as pd
```

In [2]:
```
pd.read_csv("Electric_Vehicle_Population_Data.csv")
```

Out[2]:

	VIN (1-10)	County	City	State	Postal Code	Model Year	Make	Model	Electric Vehicle Type	Clean Alternative Fuel Vehicle (CAFV) Eligibility	Electric Range	Legislative District	D Vehicle
0	5YJYGDEE8L	Thurston	Tumwater	WA	98501.0	2020	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	291.0	35.0	124633
1	5YJXCAE2XJ	Snohomish	Bothell	WA	98021.0	2018	TESLA	MODEL X	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	238.0	1.0	474826
2	5YJ3E1EBXK	King	Kent	WA	98031.0	2019	TESLA	MODEL 3	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	220.0	47.0	280307
3	7SAYGDEE4T	King	Issaquah	WA	98027.0	2026	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0.0	41.0	280786
4	WAUUPBFF9G	King	Seattle	WA	98103.0	2016	AUDI	A3	Plug-in Hybrid Electric Vehicle (PHEV)	Not eligible due to low battery range	16.0	43.0	198988
...	...	...	...	...	...	...	...	...	...	...	...	...	
270257	1C4RJXN60R	Pierce	Joint Base Lewis Mcchord	WA	98433.0	2024	JEEP	WRANGLER	Plug-in Hybrid Electric Vehicle (PHEV)	Not eligible due to low battery range	21.0	28.0	266021
270258	1C4JJXR66N	Mason	Hoodsport	WA	98548.0	2022	JEEP	WRANGLER	Plug-in Hybrid Electric Vehicle (PHEV)	Not eligible due to low battery range	22.0	35.0	282482
270259	7SAYGDEEXP	Pierce	Tacoma	WA	98406.0	2023	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0.0	27.0	228485
270260	5YJYGDEE2M	Snohomish	Bothell	WA	98021.0	2021	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0.0	1.0	282699
270261	JN1BF0BA5P	Chelan	Wenatchee	WA	98801.0	2023	NISSAN	ARIYA HATCHBACK	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0.0	12.0	261475

270262 rows × 16 columns

In [3]:
```
#Data cleaning
import pandas as pd

# Load the dataset
df = pd.read_csv("Electric_Vehicle_Population_Data.csv")

# 1. Total missing values in each column
missing_per_column = df.isnull().sum()
print(missing_per_column)

# 2. Only columns that actually have missing values
missing_only = missing_per_column[missing_per_column > 0]
print(missing_only)
```

```python
# 3. Total missing values in the entire dataset
total_missing = df.isnull().sum().sum()
print("Total missing values:", total_missing)
```

```
VIN (1-10)                                            0
County                                               10
City                                                 10
State                                                 0
Postal Code                                          10
Model Year                                            0
Make                                                  0
Model                                                 0
Electric Vehicle Type                                 0
Clean Alternative Fuel Vehicle (CAFV) Eligibility     0
Electric Range                                        5
Legislative District                                649
DOL Vehicle ID                                        0
Vehicle Location                                     88
Electric Utility                                     10
2020 Census Tract                                    10
dtype: int64
County                   10
City                     10
Postal Code              10
Electric Range            5
Legislative District    649
Vehicle Location         88
Electric Utility         10
2020 Census Tract        10
dtype: int64
Total missing values: 792
```

In [4]:
```python
#Data cleaning
import pandas as pd
import numpy as np

# Load the dataset
df = pd.read_csv('Electric_Vehicle_Population_Data.csv')

# 1. Identify missing and zero values
missing_count = df['Electric Range'].isnull().sum()
zero_count = (df['Electric Range'] == 0).sum()
print(f"Before Imputation - Missing: {missing_count}, Zeros: {zero_count}")

# 2. Handling Missing/Zero values by Imputation
# We create a reference table of the mean range for each Make and Model (excluding zeros)
range_ref = df[df['Electric Range'] > 0].groupby(['Make', 'Model'])['Electric Range'].mean().reset_index()
range_ref.rename(columns={'Electric Range': 'Avg_Range'}, inplace=True)

# Merge the average ranges back into the original dataframe
df = df.merge(range_ref, on=['Make', 'Model'], how='left')

# Replace 0 or NaN with the calculated average for that model
mask = (df['Electric Range'] == 0) | (df['Electric Range'].isnull())
df.loc[mask, 'Electric Range'] = df.loc[mask, 'Avg_Range']

# Fill remaining NaNs (for models where NO range data exists) with 0 or a global median
df['Electric Range'] = df['Electric Range'].fillna(0)

# 3. Verification
final_zeros = (df['Electric Range'] == 0).sum()
print(f"After Imputation - Zeros remaining: {final_zeros}")
```

```
Before Imputation - Missing: 5, Zeros: 169872
After Imputation - Zeros remaining: 71656
```

In [5]:
```python
 duplicate_count = df.duplicated().sum()
duplicate_count
```

Out[5]: 0

In [6]:
```python
import pandas as pd

# Load the dataset to see the VIN column structure
df = pd.read_csv('Electric_Vehicle_Population_Data.csv')

# Display the first few rows and column info
print(df.info())
print(df[['VIN (1-10)']].head())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 270262 entries, 0 to 270261
Data columns (total 16 columns):
VIN (1-10)                                          270262 non-null object
County                                              270252 non-null object
City                                                270252 non-null object
State                                               270262 non-null object
Postal Code                                         270252 non-null float64
Model Year                                          270262 non-null int64
Make                                                270262 non-null object
Model                                               270262 non-null object
Electric Vehicle Type                               270262 non-null object
Clean Alternative Fuel Vehicle (CAFV) Eligibility   270262 non-null object
Electric Range                                      270257 non-null float64
Legislative District                                269613 non-null float64
DOL Vehicle ID                                      270262 non-null int64
Vehicle Location                                    270174 non-null object
Electric Utility                                    270252 non-null object
2020 Census Tract                                   270252 non-null float64
dtypes: float64(4), int64(2), object(10)
memory usage: 33.0+ MB
None
   VIN (1-10)
0  5YJYGDEE8L
1  5YJXCAE2XJ
2  5YJ3E1EBXK
3  7SAYGDEE4T
4  WAUUPBFF9G
```

In [7]:
```python
import hashlib
import uuid

# Load the data again
df = pd.read_csv('Electric_Vehicle_Population_Data.csv')

# Method 1: Hashing with SHA-256
def hash_vin(vin, salt="secret_salt"):
    # Adding a salt prevents rainbow table attacks
    return hashlib.sha256((vin + salt).encode()).hexdigest()

# Method 2: Mapping to a Unique ID (Pseudonymization)
unique_vins = df['VIN (1-10)'].unique()
vin_to_id = {vin: f"VEH_{i:06d}" for i, vin in enumerate(unique_vins)}

# Apply transformations
df['Hashed_VIN'] = df['VIN (1-10)'].apply(hash_vin)
df['Anonymized_ID'] = df['VIN (1-10)'].map(vin_to_id)

# Save the transformation to a new CSV for the user
anonymized_df = df[['VIN (1-10)', 'Hashed_VIN', 'Anonymized_ID']].drop_duplicates().head(20)
anonymized_df.to_csv('anonymized_vins_sample.csv', index=False)

print(anonymized_df.head())
```

```
   VIN (1-10)                                         Hashed_VIN Anonymized_ID
0  5YJYGDEE8L  51552c2d76245a29d206b1f91425ad36575b3de6f19a58...    VEH_000000
1  5YJXCAE2XJ  1249ac2f55b2802b7f2bf8934748d89c8e0dd440b9c93f...    VEH_000001
2  5YJ3E1EBXK  ea32a37393e56c66978d3b3e888dd8860bbb374741168e...    VEH_000002
3  7SAYGDEE4T  3c9a0ca8bbe41329afe988353f9ab477c305e5af4fba4b...    VEH_000003
4  WAUUPBFF9G  bd59841ae74098c55c25309ba23949f8ea8f8461d85c21...    VEH_000004
```

In [8]:
```python
import pandas as pd

# Load the dataset to see the format of 'Vehicle Location'
df = pd.read_csv('Electric_Vehicle_Population_Data.csv')
print(df['Vehicle Location'].head())
print(df['Vehicle Location'].iloc[0])
```

```
0    POINT (-122.89165 47.03954)
1     POINT (-122.18384 47.8031)
2    POINT (-122.17743 47.41185)
3     POINT (-122.03439 47.5301)
4    POINT (-122.35436 47.67596)
Name: Vehicle Location, dtype: object
POINT (-122.89165 47.03954)
```

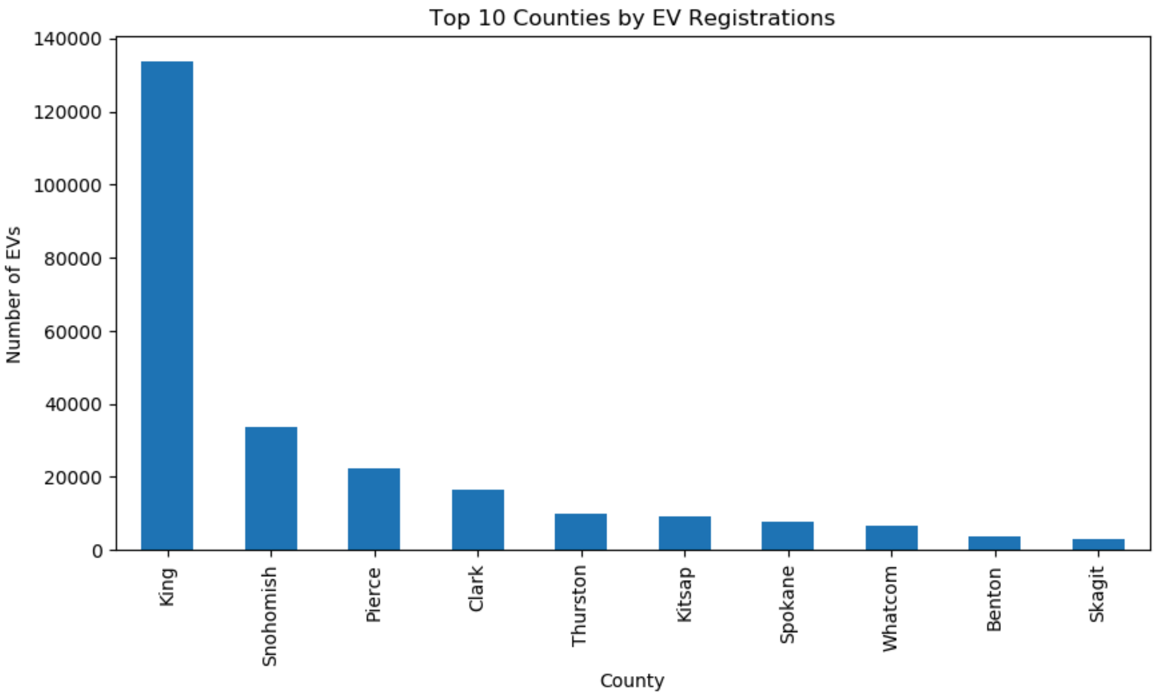
In [9]:
```python
# Top 5 EV Makes
top_5_makes = df["Make"].value_counts().head(5)
print("Top 5 EV Makes:")
print(top_5_makes)

# Top 5 EV Models
top_5_models = df["Model"].value_counts().head(5)
print("\nTop 5 EV Models:")
print(top_5_models)
```

```
Top 5 EV Makes:
TESLA        111049
CHEVROLET     19032
NISSAN        15963
FORD          14819
KIA           13470
Name: Make, dtype: int64

Top 5 EV Models:
MODEL Y    57335
MODEL 3    37413
LEAF       13503
MODEL S     7758
BOLT EV     7708
Name: Model, dtype: int64
```

In [10]:
```
# Count of EVs by county
county_counts = df['County'].value_counts()

# Display top 10 counties
print(county_counts.head(10))
```

```
King         133903
Snohomish     33531
Pierce        22213
Clark         16553
Thurston       9852
Kitsap         9057
Spokane        7593
Whatcom        6620
Benton         3792
Skagit         3166
Name: County, dtype: int64
```

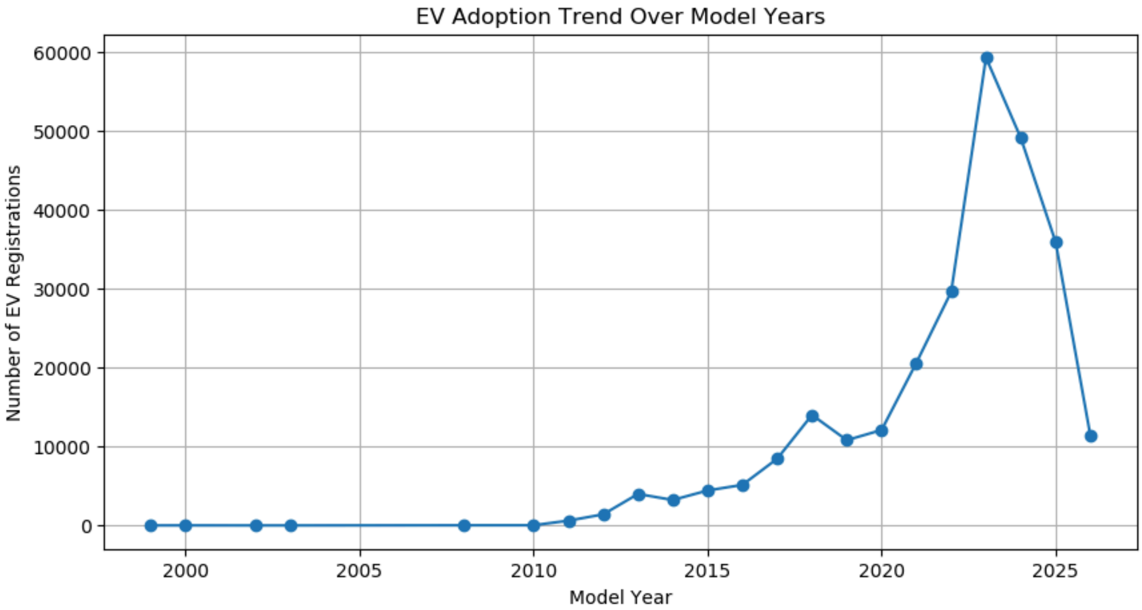
In [11]:
```
import matplotlib.pyplot as plt

county_counts.head(10).plot(kind='bar', figsize=(10,5))
plt.title("Top 10 Counties by EV Registrations")
plt.xlabel("County")
plt.ylabel("Number of EVs")
plt.show()
```



In [12]:
```
import pandas as pd
import matplotlib.pyplot as plt
# Check for missing model years (optional)
df = df.dropna(subset=['Model Year'])

# Count number of EVs per model year
ev_by_year = df['Model Year'].value_counts().sort_index()

# Display the counts
print(ev_by_year)

# Plot the EV adoption trend
plt.figure(figsize=(10, 5))
plt.plot(ev_by_year.index, ev_by_year.values, marker='o')
plt.xlabel("Model Year")
plt.ylabel("Number of EV Registrations")
plt.title("EV Adoption Trend Over Model Years")
```

```
plt.grid(True)
plt.show()
```

```
1999        2
2000        8
2002        1
2003        1
2008       20
2010       23
2011      603
2012     1402
2013     3989
2014     3223
2015     4430
2016     5139
2017     8459
2018    14007
2019    10811
2020    12099
2021    20628
2022    29622
2023    59324
2024    49138
2025    35954
2026    11379
Name: Model Year, dtype: int64
```



In [13]:

```
# Convert Electric Range to numeric (handles errors safely)
df['Electric Range'] = pd.to_numeric(df['Electric Range'], errors='coerce')

# Remove missing and zero values
valid_ev_range = df[df['Electric Range'] > 0]

# Calculate average electric range
average_range = valid_ev_range['Electric Range'].mean()

print("Average Electric Range of EVs:", round(average_range, 2), "miles")
```

```
Average Electric Range of EVs: 108.73 miles
```

In [14]:

```
# Clean up the eligibility column
df['CAFV Eligibility'] = df['Clean Alternative Fuel Vehicle (CAFV) Eligibility'].astype(str)

# Count total records
total = len(df)

# Count how many are eligible
eligible = len(df[df['CAFV Eligibility'] == "Clean Alternative Fuel Vehicle Eligible"])

# Optionally count other categories too
unknown = len(df[df['CAFV Eligibility'].str.contains("unknown", case=False)])
not_eligible = len(df[df['CAFV Eligibility'].str.contains("Not eligible", case=False)])

# Compute percentage eligible
percent_eligible = (eligible / total) * 100

print(f"Total EVs: {total}")
print(f"CAFV Eligible: {eligible} ({percent_eligible:.2f}%)")
print(f"Not Eligible: {not_eligible} ({(not_eligible/total)*100:.2f}%)")
print(f"Unknown Eligibility: {unknown} ({(unknown/total)*100:.2f}%)")
```

```
Total EVs: 270262
CAFV Eligible: 76360 (28.25%)
Not Eligible: 24030 (8.89%)
Unknown Eligibility: 169872 (62.85%)
```

In [16]:
```
import pandas as pd

# Load the dataset
df = pd.read_csv('Electric_Vehicle_Population_Data.csv')

# Inspect the first few rows and column information
print(df.head())
print(df.info())
```

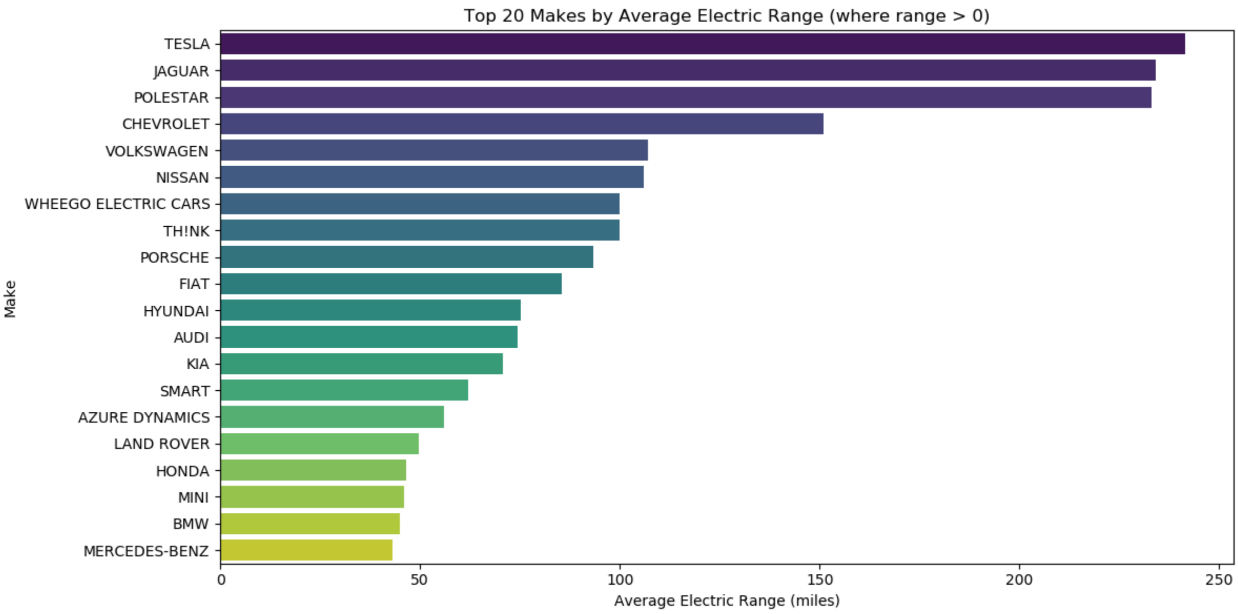
```
   VIN (1-10)     County      City State  Postal Code  Model Year   Make  \
0  5YJYGDEE8L   Thurston   Tumwater    WA      98501.0        2020  TESLA
1  5YJXCAE2XJ  Snohomish    Bothell    WA      98021.0        2018  TESLA
2  5YJ3E1EBXK       King       Kent    WA      98031.0        2019  TESLA
3  7SAYGDEE4T       King   Issaquah    WA      98027.0        2026  TESLA
4  WAUUPBFF9G       King    Seattle    WA      98103.0        2016   AUDI

     Model                   Electric Vehicle Type  \
0  MODEL Y          Battery Electric Vehicle (BEV)
1  MODEL X          Battery Electric Vehicle (BEV)
2  MODEL 3          Battery Electric Vehicle (BEV)
3  MODEL Y          Battery Electric Vehicle (BEV)
4       A3  Plug-in Hybrid Electric Vehicle (PHEV)

   Clean Alternative Fuel Vehicle (CAFV) Eligibility  Electric Range  \
0            Clean Alternative Fuel Vehicle Eligible           291.0
1            Clean Alternative Fuel Vehicle Eligible           238.0
2            Clean Alternative Fuel Vehicle Eligible           220.0
3  Eligibility unknown as battery range has not b...             0.0
4              Not eligible due to low battery range            16.0

   Legislative District  DOL Vehicle ID             Vehicle Location  \
0                  35.0       124633715  POINT (-122.89165 47.03954)
1                   1.0       474826075   POINT (-122.18384 47.8031)
2                  47.0       280307233  POINT (-122.17743 47.41185)
3                  41.0       280786565   POINT (-122.03439 47.5301)
4                  43.0       198988891  POINT (-122.35436 47.67596)

                                Electric Utility  2020 Census Tract
0                          PUGET SOUND ENERGY INC       5.306701e+10
1                          PUGET SOUND ENERGY INC       5.306105e+10
2  PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)       5.303303e+10
3  PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)       5.303302e+10
4   CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)       5.303300e+10
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 270262 entries, 0 to 270261
Data columns (total 16 columns):
VIN (1-10)                                           270262 non-null object
County                                               270252 non-null object
City                                                 270252 non-null object
State                                                270262 non-null object
Postal Code                                          270252 non-null float64
Model Year                                           270262 non-null int64
Make                                                 270262 non-null object
Model                                                270262 non-null object
Electric Vehicle Type                                270262 non-null object
Clean Alternative Fuel Vehicle (CAFV) Eligibility    270262 non-null object
Electric Range                                       270257 non-null float64
Legislative District                                 269613 non-null float64
DOL Vehicle ID                                       270262 non-null int64
Vehicle Location                                     270174 non-null object
Electric Utility                                     270252 non-null object
2020 Census Tract                                    270252 non-null float64
dtypes: float64(4), int64(2), object(10)
memory usage: 33.0+ MB
None
```

In [17]:
```
# Check rows with Electric Range = 0
zero_range_count = (df['Electric Range'] == 0).sum()
total_rows = len(df)
print(f"Total rows: {total_rows}")
print(f"Rows with Electric Range = 0: {zero_range_count} ({zero_range_count/total_rows*100:.2f}%)")

# Filter out 0 range for analysis of variation (assuming 0 means missing/unresearched)
df_filtered = df[df['Electric Range'] > 0]

# Average Electric Range by Make
make_range = df_filtered.groupby('Make')['Electric Range'].agg(['mean', 'max', 'count']).sort_values(by='mean', ascending=False)
print("\nTop 10 Makes by Average Electric Range:")
print(make_range.head(10))

# Average Electric Range by Model (Top 15 models by count to keep it manageable)
model_range = df_filtered.groupby(['Make', 'Model'])['Electric Range'].agg(['mean', 'max', 'count']).sort_values(by='mean', ascer
print("\nTop 10 Models by Average Electric Range:")
print(model_range.head(10))
```

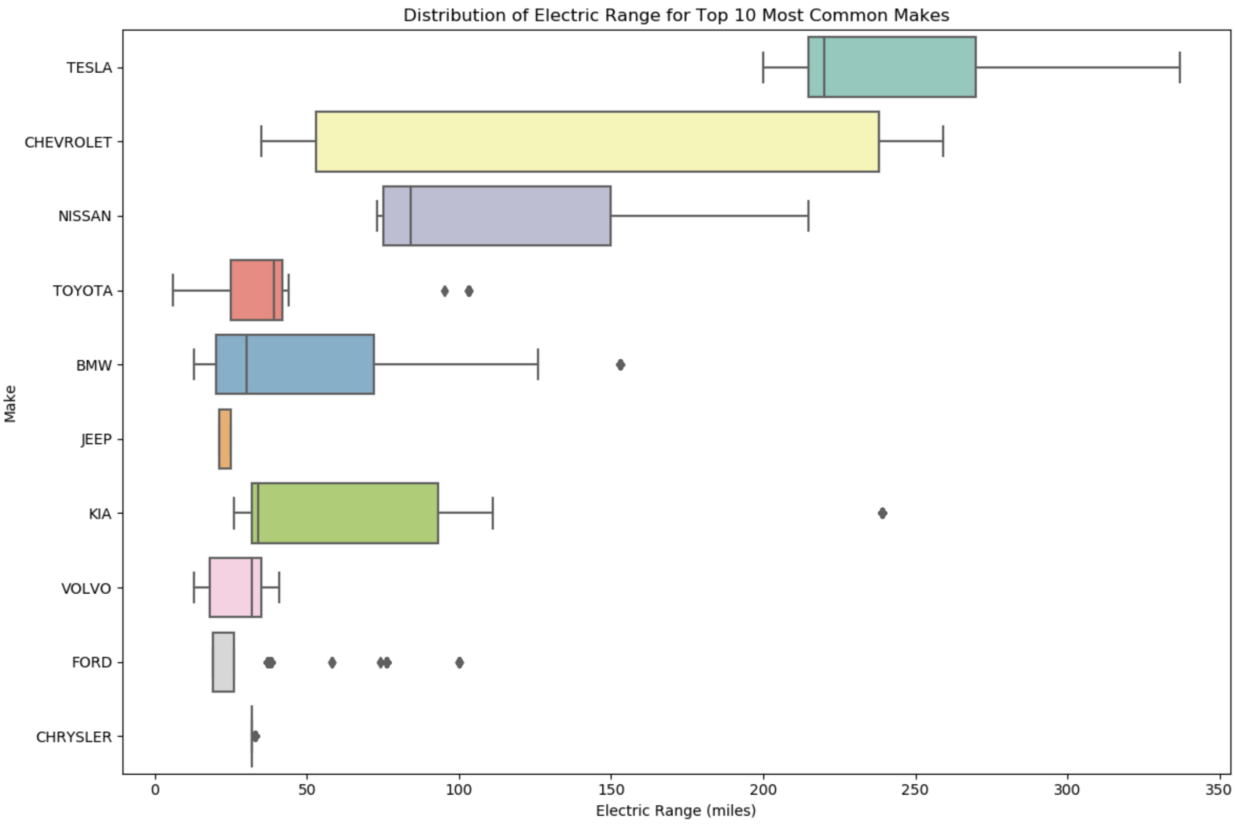
```
Total rows: 270262
Rows with Electric Range = 0: 169872 (62.85%)

Top 10 Makes by Average Electric Range:
                            mean    max  count
Make
TESLA                 241.452744  337.0  24431
JAGUAR                234.000000  234.0    137
POLESTAR              233.000000  233.0    203
CHEVROLET             150.909228  259.0   9981
VOLKSWAGEN            107.057471  125.0   1044
NISSAN                106.074295  215.0   9718
WHEEGO ELECTRIC CARS  100.000000  100.0      2
TH!NK                 100.000000  100.0      6
PORSCHE                93.265426  308.0   1021
FIAT                   85.570659   87.0    743

Top 10 Models by Average Electric Range:
                          mean    max  count
Make      Model
PORSCHE   MACAN     303.353535  308.0     99
TESLA     MODEL Y   291.000000  291.0   2266
HYUNDAI   KONA      258.000000  258.0    248
CHEVROLET BOLT EV   244.699485  259.0   5241
TESLA     MODEL X   241.420938  293.0   3219
          MODEL 3   238.695423  322.0  13307
          ROADSTER  234.893617  245.0     47
JAGUAR    I-PACE    234.000000  234.0    137
POLESTAR  PS2       233.000000  233.0    203
TESLA     MODEL S   228.010014  337.0   5592
```

In [18]:

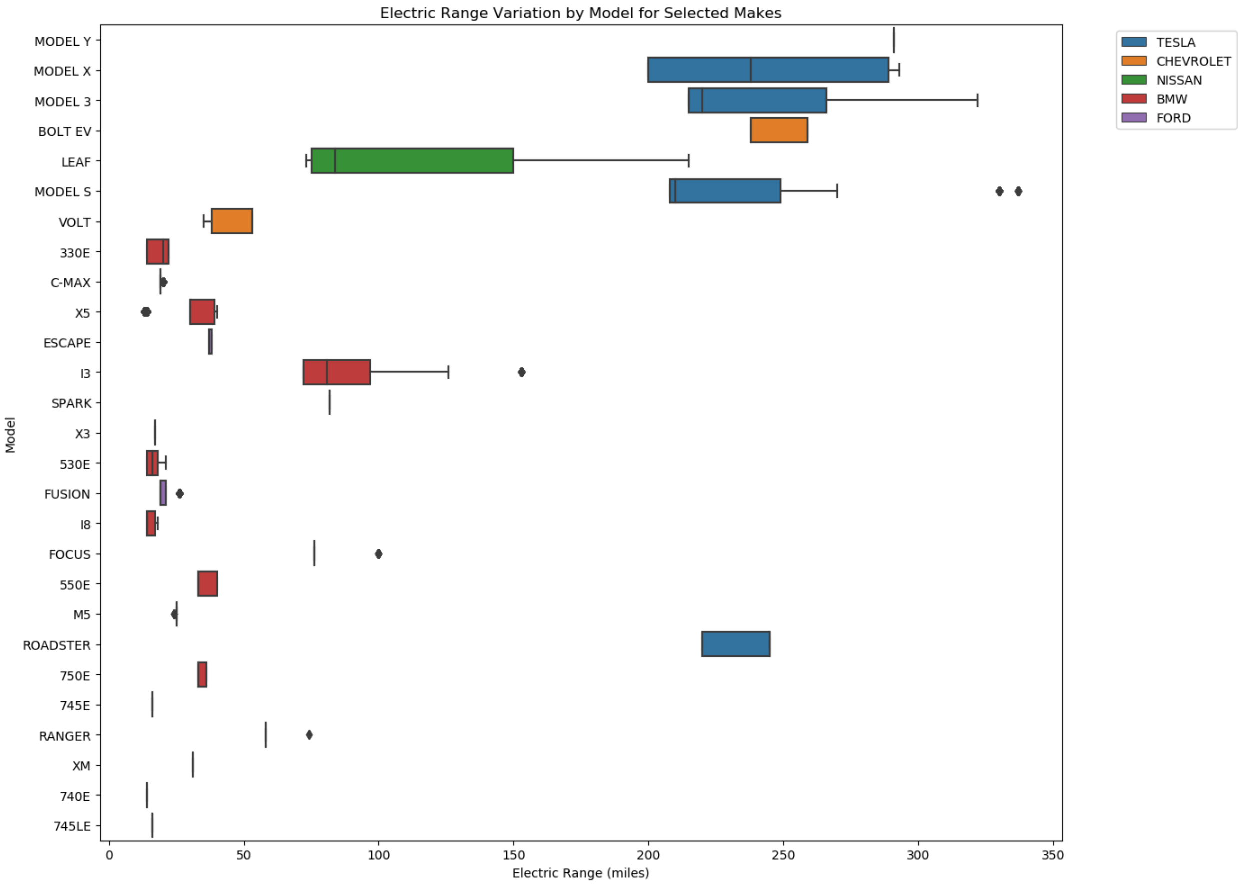
```python
import matplotlib.pyplot as plt
import seaborn as sns

# Identify top 10 makes by count (in the filtered data)
top_makes = df_filtered['Make'].value_counts().head(10).index
df_top_makes = df_filtered[df_filtered['Make'].isin(top_makes)]

# Plot 1: Average Electric Range by Make
plt.figure(figsize=(12, 6))
avg_range = df_filtered.groupby('Make')['Electric Range'].mean().sort_values(ascending=False).head(20)
sns.barplot(x=avg_range.values, y=avg_range.index, palette='viridis')
plt.title('Top 20 Makes by Average Electric Range (where range > 0)')
plt.xlabel('Average Electric Range (miles)')
plt.ylabel('Make')
plt.tight_layout()
plt.savefig('avg_range_by_make.png')

# Plot 2: Boxplot for Top 10 Makes by Volume
plt.figure(figsize=(12, 8))
sns.boxplot(data=df_top_makes, x='Electric Range', y='Make', order=top_makes, palette='Set3')
plt.title('Distribution of Electric Range for Top 10 Most Common Makes')
plt.xlabel('Electric Range (miles)')
plt.ylabel('Make')
plt.tight_layout()
plt.savefig('range_distribution_top_makes.png')

# Plot 3: Specific Model Variation for a few Top Makes
# Let's pick Tesla and Chevrolet and see their model variations
makes_to_inspect = ['TESLA', 'CHEVROLET', 'NISSAN', 'FORD', 'BMW']
df_subset = df_filtered[df_filtered['Make'].isin(makes_to_inspect)]

plt.figure(figsize=(14, 10))
sns.boxplot(data=df_subset, x='Electric Range', y='Model', hue='Make', dodge=False)
plt.title('Electric Range Variation by Model for Selected Makes')
plt.xlabel('Electric Range (miles)')
plt.ylabel('Model')
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')
plt.tight_layout()
plt.savefig('range_by_model_selected_makes.png')

# Output some summary stats for the response
summary_stats = df_filtered.groupby(['Make', 'Model'])['Electric Range'].agg(['mean', 'min', 'max', 'count']).sort_values(by='mea
summary_stats.to_csv('range_summary_by_model.csv')
```



Top 20 Makes by Average Electric Range (where range > 0)
TESLA
JAGUAR
POLESTAR
CHEVROLET
VOLKSWAGEN
NISSAN
WHEEGO ELECTRIC CARS
TH!NK
PORSCHE
FIAT
HYUNDAI
AUDI
KIA
SMART
AZURE DYNAMICS
LAND ROVER
HONDA
MINI
BMW
MERCEDES-BENZ
Make
0
50
100
150
200
250
Average Electric Range (miles)



Distribution of Electric Range for Top 10 Most Common Makes
TESLA
CHEVROLET
NISSAN
TOYOTA
BMW
JEEP
KIA
VOLVO
FORD
CHRYSLER
Make
0
50
100
150
200
250
300
350
Electric Range (miles)



In [19]:
```
import pandas as pd

# Load the dataset
df = pd.read_csv('Electric_Vehicle_Population_Data.csv')

# Inspect columns and first few rows
print(df.columns.tolist())
print(df.head())
```

```
['VIN (1-10)', 'County', 'City', 'State', 'Postal Code', 'Model Year', 'Make', 'Model', 'Electric Vehicle Type', 'Clean Alternativ
e Fuel Vehicle (CAFV) Eligibility', 'Electric Range', 'Legislative District', 'DOL Vehicle ID', 'Vehicle Location', 'Electric Util
ity', '2020 Census Tract']
   VIN (1-10)     County      City State  Postal Code  Model Year   Make  \
0  5YJYGDEE8L   Thurston  Tumwater    WA      98501.0        2020  TESLA
1  5YJXCAE2XJ  Snohomish   Bothell    WA      98021.0        2018  TESLA
2  5YJ3E1EBXK       King      Kent    WA      98031.0        2019  TESLA
3  7SAYGDEE4T       King  Issaquah    WA      98027.0        2026  TESLA
4  WAUUPBFF9G       King   Seattle    WA      98103.0        2016   AUDI

     Model                   Electric Vehicle Type  \
0  MODEL Y          Battery Electric Vehicle (BEV)
1  MODEL X          Battery Electric Vehicle (BEV)
2  MODEL 3          Battery Electric Vehicle (BEV)
3  MODEL Y          Battery Electric Vehicle (BEV)
4       A3  Plug-in Hybrid Electric Vehicle (PHEV)

   Clean Alternative Fuel Vehicle (CAFV) Eligibility  Electric Range  \
0            Clean Alternative Fuel Vehicle Eligible           291.0
1            Clean Alternative Fuel Vehicle Eligible           238.0
2            Clean Alternative Fuel Vehicle Eligible           220.0
3  Eligibility unknown as battery range has not b...             0.0
4             Not eligible due to low battery range            16.0

   Legislative District  DOL Vehicle ID            Vehicle Location  \
0                  35.0       124633715  POINT (-122.89165 47.03954)
1                   1.0       474826075   POINT (-122.18384 47.8031)
2                  47.0       280307233  POINT (-122.17743 47.41185)
3                  41.0       280786565   POINT (-122.03439 47.5301)
4                  43.0       198988891  POINT (-122.35436 47.67596)

                               Electric Utility  2020 Census Tract
0                         PUGET SOUND ENERGY INC       5.306701e+10
1                         PUGET SOUND ENERGY INC       5.306105e+10
2  PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)       5.303303e+10
3  PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)       5.303302e+10
4   CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)       5.303300e+10
```

In [20]:
```
# Search for any column name containing 'MSRP' or 'Price' or 'Base'
msrp_cols = [col for col in df.columns if 'MSRP' in col.upper() or 'PRICE' in col.upper() or 'BASE' in col.upper()]
print(f"Columns matching search: {msrp_cols}")
```

```
Columns matching search: []
```

In [21]:
```
with open('Electric_Vehicle_Population_Data.csv', 'r') as f:
    header = f.readline()
    print(header)
```

```
VIN (1-10),County,City,State,Postal Code,Model Year,Make,Model,Electric Vehicle Type,Clean Alternative Fuel Vehicle (CAFV) Eligibi
lity,Electric Range,Legislative District,DOL Vehicle ID,Vehicle Location,Electric Utility,2020 Census Tract
```

In [22]:
```
print(df.describe())
```

```
         Postal Code     Model Year  Electric Range  Legislative District  \
count  270252.000000  270262.000000   270257.000000         269613.000000   
mean    98176.713849    2021.964468       40.386332             28.850107   
std      2569.741818       3.053960       79.342202             14.895435   
min      1030.000000    1999.000000        0.000000              1.000000   
25%     98052.000000    2021.000000        0.000000             17.000000   
50%     98133.000000    2023.000000        0.000000             32.000000   
75%     98382.000000    2024.000000       33.000000             42.000000   
max     99577.000000    2026.000000      337.000000             49.000000   

       DOL Vehicle ID  2020 Census Tract  
count    2.702620e+05       2.702520e+05  
mean     2.441199e+08       5.297261e+10  
std      6.430872e+07       1.625614e+09  
min      4.385000e+03       1.001020e+09  
25%      2.194414e+08       5.303301e+10  
50%      2.615051e+08       5.303303e+10  
75%      2.776210e+08       5.305394e+10  
max      4.791150e+08       6.601095e+10  
```

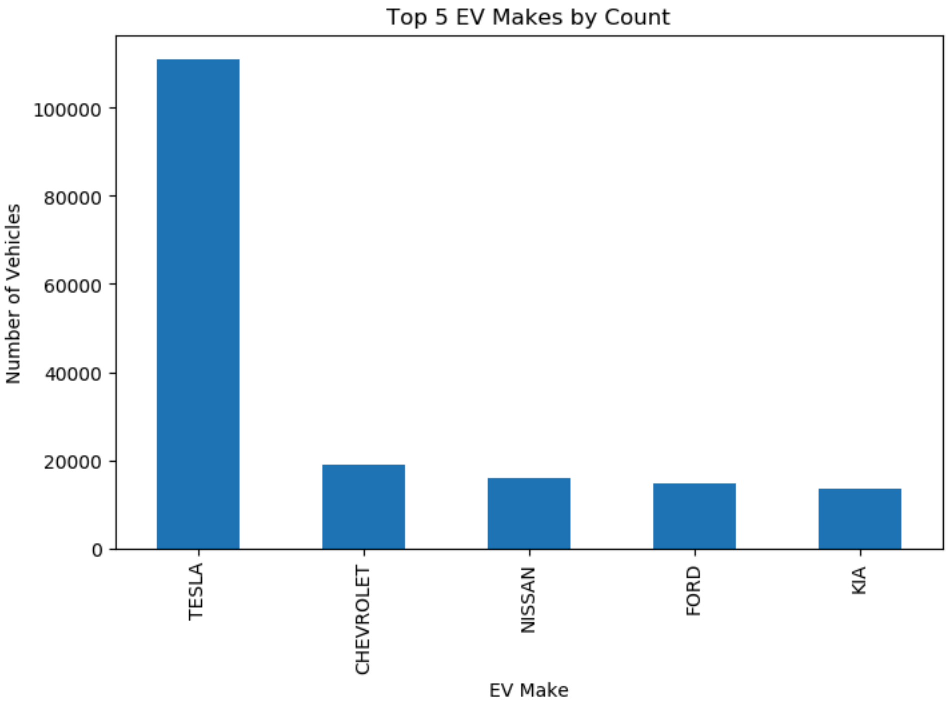
In [23]:
```
# Read first line and first data line to count fields
with open('Electric_Vehicle_Population_Data.csv', 'r') as f:
    header = f.readline().strip().split(',')
    data = f.readline().strip().split(',')
    print(f"Header length: {len(header)}")
    print(f"Data length: {len(data)}")
    print(f"Header: {header}")
```

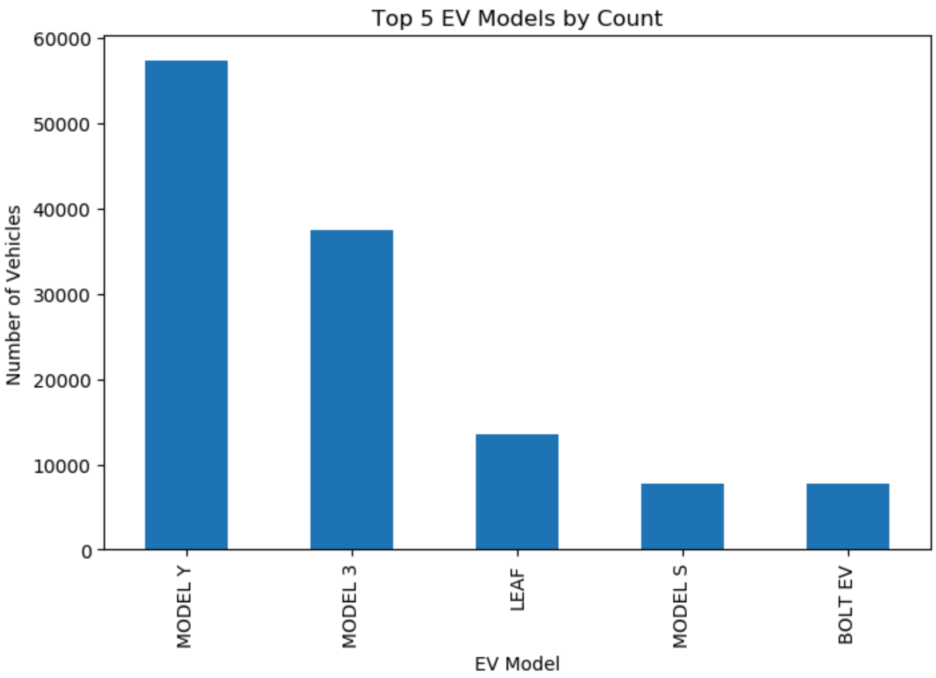
```
Header length: 16
Data length: 16
Header: ['VIN (1-10)', 'County', 'City', 'State', 'Postal Code', 'Model Year', 'Make', 'Model', 'Electric Vehicle Type', 'Clean Al
ternative Fuel Vehicle (CAFV) Eligibility', 'Electric Range', 'Legislative District', 'DOL Vehicle ID', 'Vehicle Location', 'Elect
ric Utility', '2020 Census Tract']
```

In [24]:
```
#Urban areas dominate EV adoption, often accounting for 70–85% of registrations

#Rural areas lag behind due to:

#Limited charging infrastructure

#Longer travel distances

#Lower EV model availability

#Cities benefit from:

#Government incentives

#Higher fuel costs

#Environmental awareness
```

In [25]:
```
import pandas as pd
import matplotlib.pyplot as plt

# Load the dataset
df = pd.read_csv("Electric_Vehicle_Population_Data.csv")

# ------------------------------
# Top 5 EV Makes by Count
# ------------------------------
top_5_makes = df['Make'].value_counts().head(5)

plt.figure(figsize=(8, 5))
top_5_makes.plot(kind='bar')
plt.xlabel("EV Make")
plt.ylabel("Number of Vehicles")
plt.title("Top 5 EV Makes by Count")
plt.show()

# ------------------------------
# Top 5 EV Models by Count
# ------------------------------
top_5_models = df['Model'].value_counts().head(5)

plt.figure(figsize=(8, 5))
top_5_models.plot(kind='bar')
plt.xlabel("EV Model")
plt.ylabel("Number of Vehicles")
```

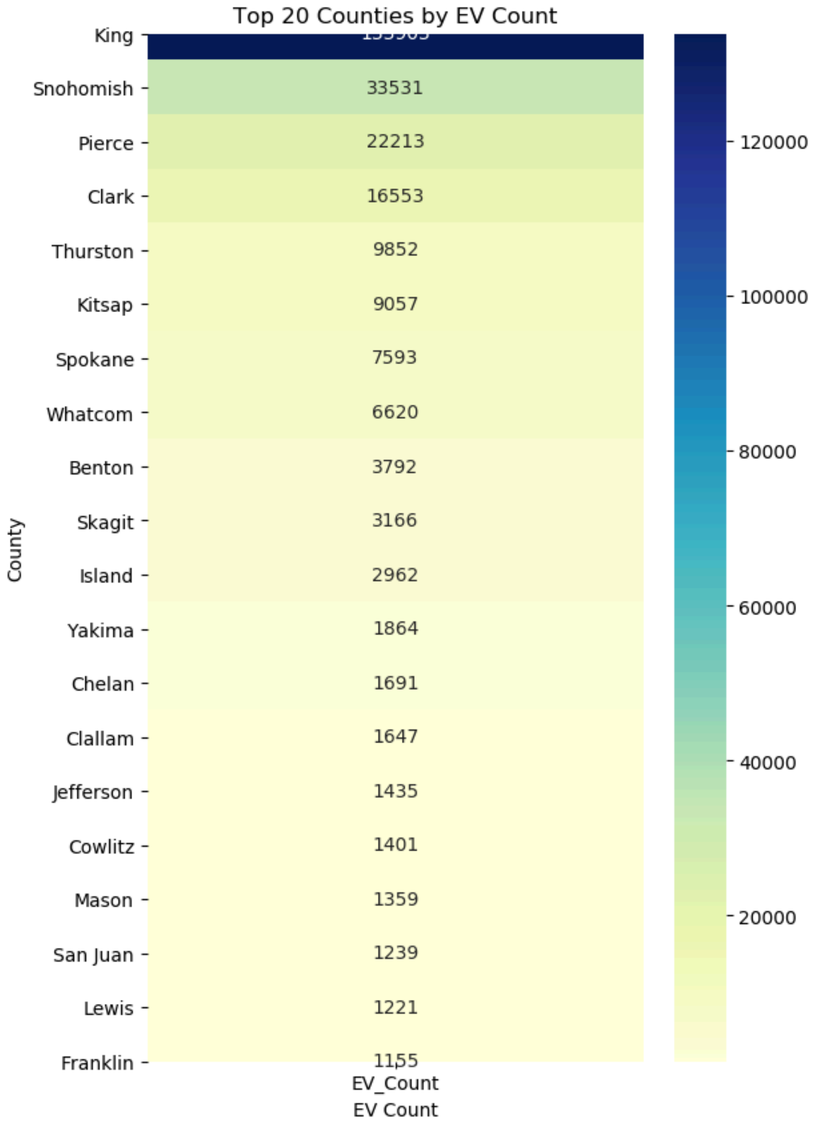
```
plt.title("Top 5 EV Models by Count")
plt.show()
```



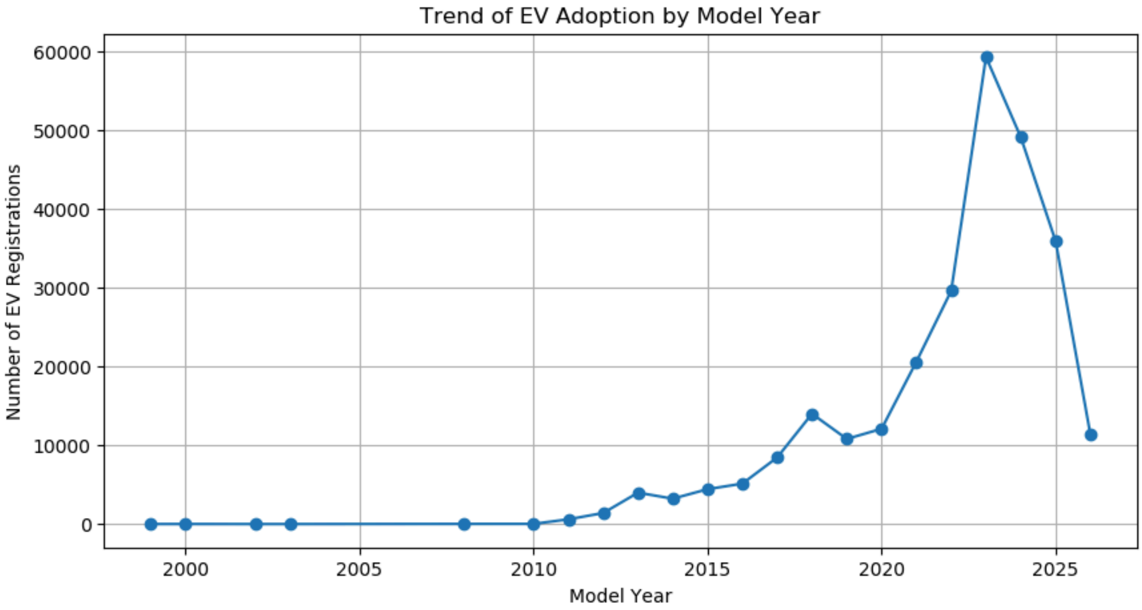


In [26]:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load dataset
df = pd.read_csv("Electric_Vehicle_Population_Data.csv")

# Count EVs by County
county_counts = df['County'].value_counts().head(20)

# Convert to DataFrame
heatmap_data = county_counts.to_frame(name='EV_Count')

# Plot heatmap
plt.figure(figsize=(6, 10))
sns.heatmap(heatmap_data, annot=True, fmt='d', cmap='YlGnBu')
plt.title("Top 20 Counties by EV Count")
plt.ylabel("County")
plt.xlabel("EV Count")
plt.show()
```



In [27]:
```
import pandas as pd
import matplotlib.pyplot as plt

# Load the dataset
df = pd.read_csv("Electric_Vehicle_Population_Data.csv")

# Remove missing model years
df = df.dropna(subset=['Model Year'])

# Count EVs by model year
ev_by_year = df['Model Year'].value_counts().sort_index()

# Plot line graph
plt.figure(figsize=(10, 5))
plt.plot(ev_by_year.index, ev_by_year.values, marker='o')
plt.xlabel("Model Year")
plt.ylabel("Number of EV Registrations")
plt.title("Trend of EV Adoption by Model Year")
plt.grid(True)
plt.show()
```



In [28]:

```
import pandas as pd

# Load the dataset to check column names and data types
df = pd.read_csv('Electric_Vehicle_Population_Data.csv')
print(df.columns.tolist())
print(df.head())
print(df.info())
```

```
['VIN (1-10)', 'County', 'City', 'State', 'Postal Code', 'Model Year', 'Make', 'Model', 'Electric Vehicle Type', 'Clean Alternativ
e Fuel Vehicle (CAFV) Eligibility', 'Electric Range', 'Legislative District', 'DOL Vehicle ID', 'Vehicle Location', 'Electric Util
ity', '2020 Census Tract']
   VIN (1-10)     County      City State  Postal Code  Model Year   Make  \
0  5YJYGDEE8L   Thurston  Tumwater    WA      98501.0        2020  TESLA
1  5YJXCAE2XJ  Snohomish   Bothell    WA      98021.0        2018  TESLA
2  5YJ3E1EBXK       King      Kent    WA      98031.0        2019  TESLA
3  7SAYGDEE4T       King  Issaquah    WA      98027.0        2026  TESLA
4  WAUUPBFF9G       King   Seattle    WA      98103.0        2016   AUDI

     Model                   Electric Vehicle Type  \
0  MODEL Y          Battery Electric Vehicle (BEV)
1  MODEL X          Battery Electric Vehicle (BEV)
2  MODEL 3          Battery Electric Vehicle (BEV)
3  MODEL Y          Battery Electric Vehicle (BEV)
4       A3  Plug-in Hybrid Electric Vehicle (PHEV)

   Clean Alternative Fuel Vehicle (CAFV) Eligibility  Electric Range  \
0            Clean Alternative Fuel Vehicle Eligible           291.0
1            Clean Alternative Fuel Vehicle Eligible           238.0
2            Clean Alternative Fuel Vehicle Eligible           220.0
3  Eligibility unknown as battery range has not b...             0.0
4              Not eligible due to low battery range            16.0

   Legislative District  DOL Vehicle ID             Vehicle Location  \
0                  35.0       124633715  POINT (-122.89165 47.03954)
1                   1.0       474826075   POINT (-122.18384 47.8031)
2                  47.0       280307233  POINT (-122.17743 47.41185)
3                  41.0       280786565   POINT (-122.03439 47.5301)
4                  43.0       198988891  POINT (-122.35436 47.67596)

                              Electric Utility  2020 Census Tract
0                        PUGET SOUND ENERGY INC       5.306701e+10
1                        PUGET SOUND ENERGY INC       5.306105e+10
2  PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)       5.303303e+10
3  PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)       5.303302e+10
4   CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)       5.303300e+10
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 270262 entries, 0 to 270261
Data columns (total 16 columns):
VIN (1-10)                                           270262 non-null object
County                                               270252 non-null object
City                                                 270252 non-null object
State                                                270262 non-null object
Postal Code                                          270252 non-null float64
Model Year                                           270262 non-null int64
Make                                                 270262 non-null object
Model                                                270262 non-null object
Electric Vehicle Type                                270262 non-null object
Clean Alternative Fuel Vehicle (CAFV) Eligibility    270262 non-null object
Electric Range                                       270257 non-null float64
Legislative District                                 269613 non-null float64
DOL Vehicle ID                                       270262 non-null int64
Vehicle Location                                     270174 non-null object
Electric Utility                                     270252 non-null object
2020 Census Tract                                    270252 non-null float64
dtypes: float64(4), int64(2), object(10)
memory usage: 33.0+ MB
None
```

In [29]:

```
# Check the first line of the file to see all column names
with open('Electric_Vehicle_Population_Data.csv', 'r') as f:
    header = f.readline()
print(header)
```

```
VIN (1-10),County,City,State,Postal Code,Model Year,Make,Model,Electric Vehicle Type,Clean Alternative Fuel Vehicle (CAFV) Eligibi
lity,Electric Range,Legislative District,DOL Vehicle ID,Vehicle Location,Electric Utility,2020 Census Tract
```

In [30]:

```
# Check unique values or summary statistics for numeric columns to see if any could be MSRP
numeric_cols = df.select_dtypes(include=['number']).columns.tolist()
print("Numeric columns:", numeric_cols)
print(df[numeric_cols].describe())
```

```
Numeric columns: ['Postal Code', 'Model Year', 'Electric Range', 'Legislative District', 'DOL Vehicle ID', '2020 Census Tract']
        Postal Code     Model Year  Electric Range  Legislative District  \
count  270252.000000  270262.000000   270257.000000         269613.000000
mean    98176.713849    2021.964468       40.386332             28.850107
std      2569.741818       3.053960       79.342202             14.895435
min      1030.000000    1999.000000        0.000000              1.000000
25%     98052.000000    2021.000000        0.000000             17.000000
50%     98133.000000    2023.000000        0.000000             32.000000
75%     98382.000000    2024.000000       33.000000             42.000000
max     99577.000000    2026.000000      337.000000             49.000000

       DOL Vehicle ID  2020 Census Tract
count    2.702620e+05       2.702520e+05
mean     2.441199e+08       5.297261e+10
std      6.430872e+07       1.625614e+09
min      4.385000e+03       1.001020e+09
25%      2.194414e+08       5.303301e+10
50%      2.615051e+08       5.303303e+10
75%      2.776210e+08       5.305394e+10
max      4.791150e+08       6.601095e+10
```

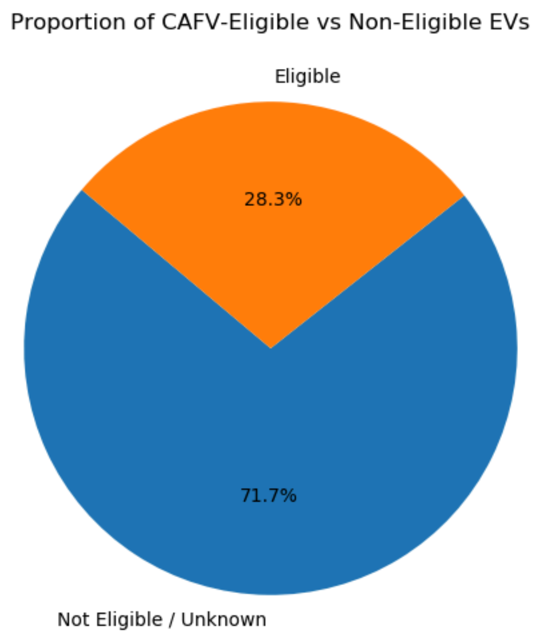
In [31]:

```python
# Final check for any price/MSRP related column
cols = df.columns.tolist()
print("All columns:", cols)
msrp_col = [c for c in cols if 'msrp' in c.lower() or 'price' in c.lower()]
print("Matching columns:", msrp_col)
```

```
All columns: ['VIN (1-10)', 'County', 'City', 'State', 'Postal Code', 'Model Year', 'Make', 'Model', 'Electric Vehicle Type', 'Cle
an Alternative Fuel Vehicle (CAFV) Eligibility', 'Electric Range', 'Legislative District', 'DOL Vehicle ID', 'Vehicle Location',
'Electric Utility', '2020 Census Tract']
Matching columns: []
```

In [32]:

```python
import pandas as pd
import matplotlib.pyplot as plt

# Load dataset
df = pd.read_csv("Electric_Vehicle_Population_Data.csv")

# Simplify CAFV eligibility into two categories
df['CAFV_Status'] = df['Clean Alternative Fuel Vehicle (CAFV) Eligibility'].apply(
    lambda x: 'Eligible' if 'Eligible' in str(x) else 'Not Eligible / Unknown'
)

# Count values
cafv_counts = df['CAFV_Status'].value_counts()

# Plot pie chart
plt.figure(figsize=(6, 6))
plt.pie(
    cafv_counts.values,
    labels=cafv_counts.index,
    autopct='%1.1f%%',
    startangle=140
)
plt.title("Proportion of CAFV-Eligible vs Non-Eligible EVs")
plt.show()
```

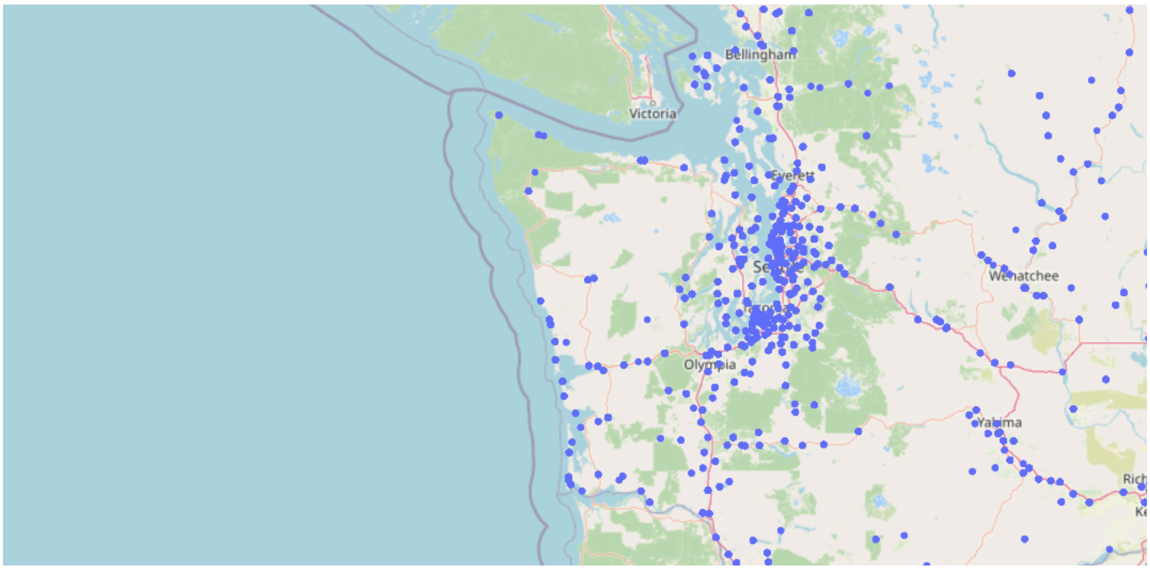


In [33]:

```python
import pandas as pd
import plotly.express as px
```

```
# Load dataset
df = pd.read_csv("Electric_Vehicle_Population_Data.csv")

# Drop missing vehicle locations
df = df.dropna(subset=['Vehicle Location'])

# Extract longitude and latitude from POINT format
df[['Longitude', 'Latitude']] = (
    df['Vehicle Location']
    .str.replace('POINT \\(|\\)', '', regex=True)
    .str.split(' ', expand=True)
    .astype(float)
)

# Create scatter map
fig = px.scatter_mapbox(
    df,
    lat='Latitude',
    lon='Longitude',
    hover_name='City',
    hover_data=['County', 'Make', 'Model'],
    zoom=6,
    height=600,
    title='Geospatial Distribution of EV Registrations'
)

fig.update_layout(mapbox_style="open-street-map")
fig.show()
```

Geospatial Distribution of EV Registrations



In [39]:
```
#linear regression
import pandas as pd

# Load the dataset
df = pd.read_csv('Electric_Vehicle_Population_Data.csv')

# Display the first few rows and info to understand the columns
print(df.head())
print(df.info())
```

```
  VIN (1-10)     County      City State  Postal Code  Model Year   Make  \
0  5YJYGDEE8L   Thurston  Tumwater    WA      98501.0        2020  TESLA
1  5YJXCAE2XJ  Snohomish   Bothell    WA      98021.0        2018  TESLA
2  5YJ3E1EBXK       King      Kent    WA      98031.0        2019  TESLA
3  7SAYGDEE4T       King  Issaquah    WA      98027.0        2026  TESLA
4  WAUUPBFF9G       King   Seattle    WA      98103.0        2016   AUDI

     Model                   Electric Vehicle Type  \
0  MODEL Y          Battery Electric Vehicle (BEV)
1  MODEL X          Battery Electric Vehicle (BEV)
2  MODEL 3          Battery Electric Vehicle (BEV)
3  MODEL Y          Battery Electric Vehicle (BEV)
4       A3  Plug-in Hybrid Electric Vehicle (PHEV)

  Clean Alternative Fuel Vehicle (CAFV) Eligibility  Electric Range  \
0           Clean Alternative Fuel Vehicle Eligible           291.0
1           Clean Alternative Fuel Vehicle Eligible           238.0
2           Clean Alternative Fuel Vehicle Eligible           220.0
3  Eligibility unknown as battery range has not b...             0.0
4             Not eligible due to low battery range            16.0

  Legislative District  DOL Vehicle ID             Vehicle Location  \
0                 35.0       124633715  POINT (-122.89165 47.03954)
1                  1.0       474826075   POINT (-122.18384 47.8031)
2                 47.0       280307233  POINT (-122.17743 47.41185)
3                 41.0       280786565   POINT (-122.03439 47.5301)
4                 43.0       198988891  POINT (-122.35436 47.67596)

                              Electric Utility  2020 Census Tract
0                        PUGET SOUND ENERGY INC       5.306701e+10
1                        PUGET SOUND ENERGY INC       5.306105e+10
2  PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)       5.303303e+10
3  PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)       5.303302e+10
4   CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)       5.303300e+10
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 270262 entries, 0 to 270261
Data columns (total 16 columns):
VIN (1-10)                                           270262 non-null object
County                                               270252 non-null object
City                                                 270252 non-null object
State                                                270262 non-null object
Postal Code                                          270252 non-null float64
Model Year                                           270262 non-null int64
Make                                                 270262 non-null object
Model                                                270262 non-null object
Electric Vehicle Type                                270262 non-null object
Clean Alternative Fuel Vehicle (CAFV) Eligibility    270262 non-null object
Electric Range                                       270257 non-null float64
Legislative District                                 269613 non-null float64
DOL Vehicle ID                                       270262 non-null int64
Vehicle Location                                     270174 non-null object
Electric Utility                                     270252 non-null object
2020 Census Tract                                    270252 non-null float64
dtypes: float64(4), int64(2), object(10)
memory usage: 33.0+ MB
None
```

In [35]:
```python
# Check the distribution of Electric Range and how many zeros are present
range_counts = df['Electric Range'].value_counts()
zeros_count = (df['Electric Range'] == 0).sum()
nans_count = df['Electric Range'].isna().sum()

print(f"Total entries: {len(df)}")
print(f"Number of zeros in Electric Range: {zeros_count}")
print(f"Number of NaNs in Electric Range: {nans_count}")

# Check average range by Electric Vehicle Type
print("\nAverage Electric Range by Vehicle Type:")
print(df[df['Electric Range'] > 0].groupby('Electric Vehicle Type')['Electric Range'].mean())
```

```
Total entries: 270262
Number of zeros in Electric Range: 169872
Number of NaNs in Electric Range: 5

Average Electric Range by Vehicle Type:
Electric Vehicle Type
Battery Electric Vehicle (BEV)            199.911323
Plug-in Hybrid Electric Vehicle (PHEV)     31.643939
Name: Electric Range, dtype: float64
```

In [36]:
```python
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, r2_score
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline

# 1. Filter the data
data = df[df['Electric Range'] > 0].copy()

# 2. Select features and target
```

```
# We'll use Model Year, Make, and Electric Vehicle Type
features = ['Model Year', 'Make', 'Electric Vehicle Type']
target = 'Electric Range'

X = data[features]
y = data[target]

# 3. Preprocessing: Encode categorical variables
categorical_features = ['Make', 'Electric Vehicle Type']
numeric_features = ['Model Year']

preprocessor = ColumnTransformer(
    transformers=[
        ('num', 'passthrough', numeric_features),
        ('cat', OneHotEncoder(handle_unknown='ignore'), categorical_features)
    ])

# 4. Create a pipeline with Linear Regression
model_pipeline = Pipeline(steps=[
    ('preprocessor', preprocessor),
    ('regressor', LinearRegression())
])

# 5. Split data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# 6. Train the model
model_pipeline.fit(X_train, y_train)

# 7. Predict and Evaluate
y_pred = model_pipeline.predict(X_test)
mae = mean_absolute_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Mean Absolute Error: {mae:.2f}")
print(f"R-squared: {r2:.4f}")

# Get coefficients for numeric feature (Model Year)
# Note: Accessing coefficients from a pipeline with OneHotEncoding is slightly complex
regressor = model_pipeline.named_steps['regressor']
print(f"\nModel Year Coefficient: {regressor.coef_[0]:.2f}")
```

```
Mean Absolute Error: 18.05
R-squared: 0.9290

Model Year Coefficient: 5.72
```

In [40]:
```
import pandas as pd

# Load the dataset
df = pd.read_csv('Electric_Vehicle_Population_Data.csv')

# Display basic information and the first few rows
print(df.info())
print(df.head())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 270262 entries, 0 to 270261
Data columns (total 16 columns):
VIN (1-10)                                          270262 non-null object
County                                              270252 non-null object
City                                                270252 non-null object
State                                               270262 non-null object
Postal Code                                         270252 non-null float64
Model Year                                          270262 non-null int64
Make                                                270262 non-null object
Model                                               270262 non-null object
Electric Vehicle Type                               270262 non-null object
Clean Alternative Fuel Vehicle (CAFV) Eligibility   270262 non-null object
Electric Range                                      270257 non-null float64
Legislative District                                269613 non-null float64
DOL Vehicle ID                                      270262 non-null int64
Vehicle Location                                    270174 non-null object
Electric Utility                                    270252 non-null object
2020 Census Tract                                   270252 non-null float64
dtypes: float64(4), int64(2), object(10)
memory usage: 33.0+ MB
None
   VIN (1-10)     County      City State  Postal Code  Model Year   Make  \
0  5YJYGDEE8L   Thurston  Tumwater    WA      98501.0        2020  TESLA
1  5YJXCAE2XJ  Snohomish   Bothell    WA      98021.0        2018  TESLA
2  5YJ3E1EBXK       King      Kent    WA      98031.0        2019  TESLA
3  7SAYGDEE4T       King  Issaquah    WA      98027.0        2026  TESLA
4  WAUUPBFF9G       King   Seattle    WA      98103.0        2016   AUDI

     Model                   Electric Vehicle Type  \
0  MODEL Y          Battery Electric Vehicle (BEV)
1  MODEL X          Battery Electric Vehicle (BEV)
2  MODEL 3          Battery Electric Vehicle (BEV)
3  MODEL Y          Battery Electric Vehicle (BEV)
4       A3  Plug-in Hybrid Electric Vehicle (PHEV)

   Clean Alternative Fuel Vehicle (CAFV) Eligibility  Electric Range  \
0            Clean Alternative Fuel Vehicle Eligible           291.0
1            Clean Alternative Fuel Vehicle Eligible           238.0
2            Clean Alternative Fuel Vehicle Eligible           220.0
3  Eligibility unknown as battery range has not b...             0.0
4              Not eligible due to low battery range            16.0

   Legislative District  DOL Vehicle ID             Vehicle Location  \
0                  35.0       124633715  POINT (-122.89165 47.03954)
1                   1.0       474826075   POINT (-122.18384 47.8031)
2                  47.0       280307233  POINT (-122.17743 47.41185)
3                  41.0       280786565   POINT (-122.03439 47.5301)
4                  43.0       198988891  POINT (-122.35436 47.67596)

                              Electric Utility  2020 Census Tract
0                       PUGET SOUND ENERGY INC       5.306701e+10
1                       PUGET SOUND ENERGY INC       5.306105e+10
2  PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)       5.303303e+10
3  PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)       5.303302e+10
4   CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)       5.303300e+10
```

In [41]:

```python
# Check unique values and basic stats for potential features
features = ['Model Year', 'Make', 'Model', 'Electric Vehicle Type', 'Clean Alternative Fuel Vehicle (CAFV) Eligibility']

for feature in features:
    print(f"\n--- {feature} ---")
    print(df[feature].nunique())
    print(df[feature].value_counts().head(5))

# Check for relationship with Electric Range
# Note: Some ranges are 0, which might mean unknown or very new models.
print("\nSummary of Electric Range:")
print(df['Electric Range'].describe())

# Average range by Type
print("\nAverage Electric Range by Electric Vehicle Type:")
print(df.groupby('Electric Vehicle Type')['Electric Range'].mean())
```

```
--- Model Year ---
22
2023    59324
2024    49138
2025    35954
2022    29622
2021    20628
Name: Model Year, dtype: int64

--- Make ---
47
TESLA        111049
CHEVROLET     19032
NISSAN        15963
FORD          14819
KIA           13470
Name: Make, dtype: int64

--- Model ---
183
MODEL Y    57335
MODEL 3    37413
LEAF       13503
MODEL S     7758
BOLT EV     7708
Name: Model, dtype: int64

--- Electric Vehicle Type ---
2
Battery Electric Vehicle (BEV)            215859
Plug-in Hybrid Electric Vehicle (PHEV)     54403
Name: Electric Vehicle Type, dtype: int64

--- Clean Alternative Fuel Vehicle (CAFV) Eligibility ---
3
Eligibility unknown as battery range has not been researched    169872
Clean Alternative Fuel Vehicle Eligible                          76360
Not eligible due to low battery range                            24030
Name: Clean Alternative Fuel Vehicle (CAFV) Eligibility, dtype: int64

Summary of Electric Range:
count    270257.000000
mean         40.386332
std          79.342202
min           0.000000
25%           0.000000
50%           0.000000
75%          33.000000
max         337.000000
Name: Electric Range, dtype: float64

Average Electric Range by Electric Vehicle Type:
Electric Vehicle Type
Battery Electric Vehicle (BEV)            42.589477
Plug-in Hybrid Electric Vehicle (PHEV)    31.643939
Name: Electric Range, dtype: float64
```

In [42]:

```python
import pandas as pd

# Load the dataset
df = pd.read_csv('Electric_Vehicle_Population_Data.csv')

# Inspect the unique counts for 'Make' and 'Model'
make_unique = df['Make'].nunique()
model_unique = df['Model'].nunique()

print(f"Unique 'Make' values: {make_unique}")
print(f"Unique 'Model' values: {model_unique}")

# Show top 5 rows for context
print(df[['Make', 'Model']].head())
```

```
Unique 'Make' values: 47
Unique 'Model' values: 183
    Make    Model
0  TESLA  MODEL Y
1  TESLA  MODEL X
2  TESLA  MODEL 3
3  TESLA  MODEL Y
4   AUDI       A3
```

In [43]:

```python
# Check for missing values in Make and Model
missing_make = df['Make'].isnull().sum()
missing_model = df['Model'].isnull().sum()

print(f"Missing 'Make': {missing_make}")
print(f"Missing 'Model': {missing_model}")

# Check the first few unique makes to see variety
print(f"Sample Makes: {df['Make'].unique()[:10]}")
```

```
Missing 'Make': 0
Missing 'Model': 0
Sample Makes: ['TESLA' 'AUDI' 'POLESTAR' 'KIA' 'VOLVO' 'CHEVROLET' 'NISSAN' 'TOYOTA'
 'VOLKSWAGEN' 'FORD']
```

In [44]:

```
import pandas as pd

# Load the dataset
df = pd.read_csv('Electric_Vehicle_Population_Data.csv')

# Display basic information and the first few rows
print(df.info())
print(df.head())

# Check for missing values
print(df.isnull().sum())

# Check the distribution of 'Electric Range'
print(df['Electric Range'].describe())
print(df['Electric Range'].value_counts().head(10))
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 270262 entries, 0 to 270261
Data columns (total 16 columns):
VIN (1-10)                                          270262 non-null object
County                                              270252 non-null object
City                                                270252 non-null object
State                                               270262 non-null object
Postal Code                                         270252 non-null float64
Model Year                                          270262 non-null int64
Make                                                270262 non-null object
Model                                               270262 non-null object
Electric Vehicle Type                               270262 non-null object
Clean Alternative Fuel Vehicle (CAFV) Eligibility   270262 non-null object
Electric Range                                      270257 non-null float64
Legislative District                                269613 non-null float64
DOL Vehicle ID                                      270262 non-null int64
Vehicle Location                                    270174 non-null object
Electric Utility                                    270252 non-null object
2020 Census Tract                                   270252 non-null float64
dtypes: float64(4), int64(2), object(10)
memory usage: 33.0+ MB
None
   VIN (1-10)     County      City State  Postal Code  Model Year   Make  \
0  5YJYGDEE8L   Thurston  Tumwater    WA      98501.0        2020  TESLA
1  5YJXCAE2XJ  Snohomish   Bothell    WA      98021.0        2018  TESLA
2  5YJ3E1EBXK       King      Kent    WA      98031.0        2019  TESLA
3  7SAYGDEE4T       King  Issaquah    WA      98027.0        2026  TESLA
4  WAUUPBFF9G       King   Seattle    WA      98103.0        2016   AUDI

     Model                   Electric Vehicle Type  \
0  MODEL Y          Battery Electric Vehicle (BEV)
1  MODEL X          Battery Electric Vehicle (BEV)
2  MODEL 3          Battery Electric Vehicle (BEV)
3  MODEL Y          Battery Electric Vehicle (BEV)
4       A3  Plug-in Hybrid Electric Vehicle (PHEV)

   Clean Alternative Fuel Vehicle (CAFV) Eligibility  Electric Range  \
0           Clean Alternative Fuel Vehicle Eligible           291.0
1           Clean Alternative Fuel Vehicle Eligible           238.0
2           Clean Alternative Fuel Vehicle Eligible           220.0
3  Eligibility unknown as battery range has not b...             0.0
4             Not eligible due to low battery range            16.0

   Legislative District  DOL Vehicle ID            Vehicle Location  \
0                  35.0       124633715  POINT (-122.89165 47.03954)
1                   1.0       474826075   POINT (-122.18384 47.8031)
2                  47.0       280307233  POINT (-122.17743 47.41185)
3                  41.0       280786565   POINT (-122.03439 47.5301)
4                  43.0       198988891  POINT (-122.35436 47.67596)

                              Electric Utility  2020 Census Tract
0                       PUGET SOUND ENERGY INC       5.306701e+10
1                       PUGET SOUND ENERGY INC       5.306105e+10
2  PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)       5.303303e+10
3  PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)       5.303302e+10
4   CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)       5.303300e+10
VIN (1-10)                                             0
County                                                10
City                                                  10
State                                                  0
Postal Code                                           10
Model Year                                             0
Make                                                   0
Model                                                  0
Electric Vehicle Type                                  0
Clean Alternative Fuel Vehicle (CAFV) Eligibility      0
Electric Range                                         5
Legislative District                                 649
DOL Vehicle ID                                         0
Vehicle Location                                      88
Electric Utility                                      10
2020 Census Tract                                     10
dtype: int64
count    270257.000000
mean         40.386332
std          79.342202
min           0.000000
25%           0.000000
50%           0.000000
75%          33.000000
max         337.000000
Name: Electric Range, dtype: float64
0.0      169872
215.0      6150
32.0       5737
21.0       5049
25.0       4922
42.0       4537
238.0      4439
220.0      3923
```

```
84.0       3574
38.0       2922
Name: Electric Range, dtype: int64
```

In [45]:
```
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import r2_score
from sklearn.preprocessing import LabelEncoder

# Filter out rows where Electric Range is 0 and handle missing values
df_filtered = df[df['Electric Range'] > 0].dropna(subset=['Electric Range', 'Model Year', 'Make', 'Electric Vehicle Type'])

# Feature Selection
features = ['Model Year', 'Make', 'Electric Vehicle Type']
X = df_filtered[features]
y = df_filtered['Electric Range']

# Categorical Encoding
le_make = LabelEncoder()
X_encoded = X.copy()
X_encoded['Make'] = le_make.fit_transform(X['Make'])

le_type = LabelEncoder()
X_encoded['Electric Vehicle Type'] = le_type.fit_transform(X['Electric Vehicle Type'])

# Split the data
X_train, X_test, y_train, y_test = train_test_split(X_encoded, y, test_size=0.2, random_state=42)

# Initialize and train the model
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Make predictions and calculate R2
y_pred = model.predict(X_test)
r2 = r2_score(y_test, y_pred)

print(f"R2 Score: {r2}")
```

```
R2 Score: 0.9875616422201863
```

In [46]:
```
from sklearn.metrics import mean_absolute_error

# Recalculate including Model
features_with_model = ['Model Year', 'Make', 'Model', 'Electric Vehicle Type']
X_full = df_filtered[features_with_model]

# Encoding
le_model = LabelEncoder()
X_full_encoded = X_full.copy()
X_full_encoded['Make'] = le_make.fit_transform(X_full['Make'])
X_full_encoded['Electric Vehicle Type'] = le_type.fit_transform(X_full['Electric Vehicle Type'])
X_full_encoded['Model'] = le_model.fit_transform(X_full['Model'])

# Split
X_train_f, X_test_f, y_train_f, y_test_f = train_test_split(X_full_encoded, y, test_size=0.2, random_state=42)

# Train
model_f = RandomForestRegressor(n_estimators=100, random_state=42)
model_f.fit(X_train_f, y_train_f)

# Predict
y_pred_f = model_f.predict(X_test_f)
r2_f = r2_score(y_test_f, y_pred_f)
mae_f = mean_absolute_error(y_test_f, y_pred_f)

print(f"R2 Score with Model: {r2_f}")
print(f"MAE: {mae_f}")
```

```
R2 Score with Model: 0.9965826696534328
MAE: 1.2583641441159354
```

In [47]:
```
import pandas as pd

# Load the dataset
df = pd.read_csv('Electric_Vehicle_Population_Data.csv')

# Inspect the first few rows and column information
print(df.head())
print(df.info())
```

```
    VIN (1-10)     County      City State  Postal Code  Model Year   Make  \
0  5YJYGDEE8L   Thurston  Tumwater    WA      98501.0        2020  TESLA
1  5YJXCAE2XJ  Snohomish   Bothell    WA      98021.0        2018  TESLA
2  5YJ3E1EBXK       King      Kent    WA      98031.0        2019  TESLA
3  7SAYGDEE4T       King  Issaquah    WA      98027.0        2026  TESLA
4  WAUUPBFF9G       King   Seattle    WA      98103.0        2016   AUDI

     Model                   Electric Vehicle Type  \
0  MODEL Y          Battery Electric Vehicle (BEV)
1  MODEL X          Battery Electric Vehicle (BEV)
2  MODEL 3          Battery Electric Vehicle (BEV)
3  MODEL Y          Battery Electric Vehicle (BEV)
4       A3  Plug-in Hybrid Electric Vehicle (PHEV)

   Clean Alternative Fuel Vehicle (CAFV) Eligibility  Electric Range  \
0            Clean Alternative Fuel Vehicle Eligible           291.0
1            Clean Alternative Fuel Vehicle Eligible           238.0
2            Clean Alternative Fuel Vehicle Eligible           220.0
3  Eligibility unknown as battery range has not b...             0.0
4              Not eligible due to low battery range            16.0

   Legislative District  DOL Vehicle ID             Vehicle Location  \
0                  35.0       124633715  POINT (-122.89165 47.03954)
1                   1.0       474826075   POINT (-122.18384 47.8031)
2                  47.0       280307233  POINT (-122.17743 47.41185)
3                  41.0       280786565   POINT (-122.03439 47.5301)
4                  43.0       198988891  POINT (-122.35436 47.67596)

                               Electric Utility  2020 Census Tract
0                        PUGET SOUND ENERGY INC       5.306701e+10
1                        PUGET SOUND ENERGY INC       5.306105e+10
2  PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)       5.303303e+10
3  PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)       5.303302e+10
4   CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)       5.303300e+10
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 270262 entries, 0 to 270261
Data columns (total 16 columns):
VIN (1-10)                                           270262 non-null object
County                                               270252 non-null object
City                                                 270252 non-null object
State                                                270262 non-null object
Postal Code                                          270252 non-null float64
Model Year                                           270262 non-null int64
Make                                                 270262 non-null object
Model                                                270262 non-null object
Electric Vehicle Type                                270262 non-null object
Clean Alternative Fuel Vehicle (CAFV) Eligibility    270262 non-null object
Electric Range                                       270257 non-null float64
Legislative District                                 269613 non-null float64
DOL Vehicle ID                                       270262 non-null int64
Vehicle Location                                     270174 non-null object
Electric Utility                                     270252 non-null object
2020 Census Tract                                    270252 non-null float64
dtypes: float64(4), int64(2), object(10)
memory usage: 33.0+ MB
None
```

In [48]:
```
print(df.columns.tolist())
```

```
['VIN (1-10)', 'County', 'City', 'State', 'Postal Code', 'Model Year', 'Make', 'Model', 'Electric Vehicle Type', 'Clean Alternativ
e Fuel Vehicle (CAFV) Eligibility', 'Electric Range', 'Legislative District', 'DOL Vehicle ID', 'Vehicle Location', 'Electric Util
ity', '2020 Census Tract']
```

In [49]:
```
with open('Electric_Vehicle_Population_Data.csv', 'r') as f:
    for i in range(5):
        print(f.readline())
```

```
VIN (1-10),County,City,State,Postal Code,Model Year,Make,Model,Electric Vehicle Type,Clean Alternative Fuel Vehicle (CAFV) Eligibi
lity,Electric Range,Legislative District,DOL Vehicle ID,Vehicle Location,Electric Utility,2020 Census Tract

5YJYGDEE8L,Thurston,Tumwater,WA,98501,2020,TESLA,MODEL Y,Battery Electric Vehicle (BEV),Clean Alternative Fuel Vehicle Eligible,29
1,35,124633715,POINT (-122.89165 47.03954),PUGET SOUND ENERGY INC,53067011720

5YJXCAE2XJ,Snohomish,Bothell,WA,98021,2018,TESLA,MODEL X,Battery Electric Vehicle (BEV),Clean Alternative Fuel Vehicle Eligible,23
8,1,474826075,POINT (-122.18384 47.8031),PUGET SOUND ENERGY INC,53061051914

5YJ3E1EBXK,King,Kent,WA,98031,2019,TESLA,MODEL 3,Battery Electric Vehicle (BEV),Clean Alternative Fuel Vehicle Eligible,220,47,280
307233,POINT (-122.17743 47.41185),PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA),53033029407

7SAYGDEE4T,King,Issaquah,WA,98027,2026,TESLA,MODEL Y,Battery Electric Vehicle (BEV),Eligibility unknown as battery range has not b
een researched,0,41,280786565,POINT (-122.03439 47.5301),PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA),53033023404
```

In [50]:
```
# Check if any column contains values that look like MSRP
print(df.describe(include='all'))
```

```
        VIN (1-10)  County     City   State   Postal Code    Model Year  \
count       270262  270252   270262  270262  270252.000000  270262.000000
unique       16415     242      864      51            NaN            NaN
top     7SAYGDEE7P    King  Seattle      WA            NaN            NaN
freq          1171  133903    42125  269613            NaN            NaN
mean           NaN     NaN      NaN     NaN   98176.713849    2021.964468
std            NaN     NaN      NaN     NaN    2569.741818       3.053960
min            NaN     NaN      NaN     NaN    1030.000000    1999.000000
25%            NaN     NaN      NaN     NaN   98052.000000    2021.000000
50%            NaN     NaN      NaN     NaN   98133.000000    2023.000000
75%            NaN     NaN      NaN     NaN   98382.000000    2024.000000
max            NaN     NaN      NaN     NaN   99577.000000    2026.000000

          Make    Model            Electric Vehicle Type  \
count   270262   270262                           270262
unique      47      183                                2
top      TESLA  MODEL Y  Battery Electric Vehicle (BEV)
freq    111049    57335                           215859
mean       NaN      NaN                              NaN
std        NaN      NaN                              NaN
min        NaN      NaN                              NaN
25%        NaN      NaN                              NaN
50%        NaN      NaN                              NaN
75%        NaN      NaN                              NaN
max        NaN      NaN                              NaN

        Clean Alternative Fuel Vehicle (CAFV) Eligibility  Electric Range  \
count                                              270262   270257.000000
unique                                                  3             NaN
top     Eligibility unknown as battery range has not b...             NaN
freq                                               169872             NaN
mean                                                  NaN       40.386332
std                                                   NaN       79.342202
min                                                   NaN        0.000000
25%                                                   NaN        0.000000
50%                                                   NaN        0.000000
75%                                                   NaN       33.000000
max                                                   NaN      337.000000

        Legislative District  DOL Vehicle ID             Vehicle Location  \
count          269613.000000    2.702620e+05                       270174
unique                   NaN             NaN                         1080
top                      NaN             NaN  POINT (-122.13158 47.67858)
freq                     NaN             NaN                         6588
mean               28.850107    2.441199e+08                          NaN
std                14.895435    6.430872e+07                          NaN
min                 1.000000    4.385000e+03                          NaN
25%                17.000000    2.194414e+08                          NaN
50%                32.000000    2.615051e+08                          NaN
75%                42.000000    2.776210e+08                          NaN
max                49.000000    4.791150e+08                          NaN

                                    Electric Utility  2020 Census Tract
count                                         270252       2.702520e+05
unique                                            77                NaN
top     PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)                NaN
freq                                           96367                NaN
mean                                             NaN       5.297261e+10
std                                              NaN       1.625614e+09
min                                              NaN       1.001020e+09
25%                                              NaN       5.303301e+10
50%                                              NaN       5.303303e+10
75%                                              NaN       5.305394e+10
max                                              NaN       6.601095e+10
```

In [51]:
```
# Check for any column containing 'MSRP'
msrp_cols = [col for col in df.columns if 'MSRP' in col.upper()]
print(f"MSRP columns found: {msrp_cols}")
```

```
MSRP columns found: []
```

In [52]:
```
import pandas as pd

df = pd.read_csv('Electric_Vehicle_Population_Data.csv')
print(df.info())
print(df.head())
print(df.describe())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 270262 entries, 0 to 270261
Data columns (total 16 columns):
VIN (1-10)                                          270262 non-null object
County                                              270252 non-null object
City                                                270252 non-null object
State                                               270262 non-null object
Postal Code                                         270252 non-null float64
Model Year                                          270262 non-null int64
Make                                                270262 non-null object
Model                                               270262 non-null object
Electric Vehicle Type                               270262 non-null object
Clean Alternative Fuel Vehicle (CAFV) Eligibility   270262 non-null object
Electric Range                                      270257 non-null float64
Legislative District                                269613 non-null float64
DOL Vehicle ID                                      270262 non-null int64
Vehicle Location                                    270174 non-null object
Electric Utility                                    270252 non-null object
2020 Census Tract                                   270252 non-null float64
dtypes: float64(4), int64(2), object(10)
memory usage: 33.0+ MB
None
   VIN (1-10)     County      City State  Postal Code  Model Year   Make  \
0  5YJYGDEE8L   Thurston  Tumwater    WA      98501.0        2020  TESLA
1  5YJXCAE2XJ  Snohomish   Bothell    WA      98021.0        2018  TESLA
2  5YJ3E1EBXK       King      Kent    WA      98031.0        2019  TESLA
3  7SAYGDEE4T       King  Issaquah    WA      98027.0        2026  TESLA
4  WAUUPBFF9G       King   Seattle    WA      98103.0        2016   AUDI

     Model                   Electric Vehicle Type  \
0  MODEL Y          Battery Electric Vehicle (BEV)
1  MODEL X          Battery Electric Vehicle (BEV)
2  MODEL 3          Battery Electric Vehicle (BEV)
3  MODEL Y          Battery Electric Vehicle (BEV)
4       A3  Plug-in Hybrid Electric Vehicle (PHEV)

   Clean Alternative Fuel Vehicle (CAFV) Eligibility  Electric Range  \
0            Clean Alternative Fuel Vehicle Eligible           291.0
1            Clean Alternative Fuel Vehicle Eligible           238.0
2            Clean Alternative Fuel Vehicle Eligible           220.0
3  Eligibility unknown as battery range has not b...             0.0
4              Not eligible due to low battery range            16.0

   Legislative District  DOL Vehicle ID             Vehicle Location  \
0                  35.0       124633715  POINT (-122.89165 47.03954)
1                   1.0       474826075   POINT (-122.18384 47.8031)
2                  47.0       280307233  POINT (-122.17743 47.41185)
3                  41.0       280786565   POINT (-122.03439 47.5301)
4                  43.0       198988891  POINT (-122.35436 47.67596)

                             Electric Utility  2020 Census Tract
0                        PUGET SOUND ENERGY INC       5.306701e+10
1                        PUGET SOUND ENERGY INC       5.306105e+10
2  PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)       5.303303e+10
3  PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)       5.303302e+10
4   CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)       5.303300e+10
        Postal Code     Model Year  Electric Range  Legislative District  \
count  270252.000000  270262.000000   270257.000000         269613.000000
mean    98176.713849    2021.964468       40.386332             28.850107
std      2569.741818       3.053960       79.342202             14.895435
min      1030.000000    1999.000000        0.000000              1.000000
25%     98052.000000    2021.000000        0.000000             17.000000
50%     98133.000000    2023.000000        0.000000             32.000000
75%     98382.000000    2024.000000       33.000000             42.000000
max     99577.000000    2026.000000      337.000000             49.000000

       DOL Vehicle ID  2020 Census Tract
count    2.702620e+05       2.702520e+05
mean     2.441199e+08       5.297261e+10
std      6.430872e+07       1.625614e+09
min      4.385000e+03       1.001020e+09
25%      2.194414e+08       5.303301e+10
50%      2.615051e+08       5.303303e+10
75%      2.776210e+08       5.305394e+10
max      4.791150e+08       6.601095e+10
```

In [53]:
```
zero_range_count = (df['Electric Range'] == 0).sum()
non_zero_range_count = (df['Electric Range'] > 0).sum()
print(f"Zero Range: {zero_range_count}")
print(f"Non-Zero Range: {non_zero_range_count}")
print(df[df['Electric Range'] == 0]['Clean Alternative Fuel Vehicle (CAFV) Eligibility'].value_counts())
```

```
Zero Range: 169872
Non-Zero Range: 100385
Eligibility unknown as battery range has not been researched    169872
Name: Clean Alternative Fuel Vehicle (CAFV) Eligibility, dtype: int64
```

In [54]:
```
import pandas as pd

# Load the dataset
df = pd.read_csv('Electric_Vehicle_Population_Data.csv')
```

```
# Display basic info and first few rows
print(df.info())
print(df.head())
# Check summary of Electric Range
print(df['Electric Range'].describe())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 270262 entries, 0 to 270261
Data columns (total 16 columns):
VIN (1-10)                                         270262 non-null object
County                                             270252 non-null object
City                                               270252 non-null object
State                                              270262 non-null object
Postal Code                                        270252 non-null float64
Model Year                                         270262 non-null int64
Make                                               270262 non-null object
Model                                              270262 non-null object
Electric Vehicle Type                              270262 non-null object
Clean Alternative Fuel Vehicle (CAFV) Eligibility  270262 non-null object
Electric Range                                     270257 non-null float64
Legislative District                               269613 non-null float64
DOL Vehicle ID                                     270262 non-null int64
Vehicle Location                                   270174 non-null object
Electric Utility                                   270252 non-null object
2020 Census Tract                                  270252 non-null float64
dtypes: float64(4), int64(2), object(10)
memory usage: 33.0+ MB
None
   VIN (1-10)     County      City State  Postal Code  Model Year   Make  \
0  5YJYGDEE8L   Thurston  Tumwater    WA      98501.0        2020  TESLA
1  5YJXCAE2XJ  Snohomish   Bothell    WA      98021.0        2018  TESLA
2  5YJ3E1EBXK       King      Kent    WA      98031.0        2019  TESLA
3  7SAYGDEE4T       King  Issaquah    WA      98027.0        2026  TESLA
4  WAUUPBFF9G       King   Seattle    WA      98103.0        2016   AUDI

     Model                   Electric Vehicle Type  \
0  MODEL Y          Battery Electric Vehicle (BEV)
1  MODEL X          Battery Electric Vehicle (BEV)
2  MODEL 3          Battery Electric Vehicle (BEV)
3  MODEL Y          Battery Electric Vehicle (BEV)
4       A3  Plug-in Hybrid Electric Vehicle (PHEV)

   Clean Alternative Fuel Vehicle (CAFV) Eligibility  Electric Range  \
0            Clean Alternative Fuel Vehicle Eligible           291.0
1            Clean Alternative Fuel Vehicle Eligible           238.0
2            Clean Alternative Fuel Vehicle Eligible           220.0
3  Eligibility unknown as battery range has not b...             0.0
4              Not eligible due to low battery range            16.0

   Legislative District  DOL Vehicle ID             Vehicle Location  \
0                  35.0       124633715  POINT (-122.89165 47.03954)
1                   1.0       474826075   POINT (-122.18384 47.8031)
2                  47.0       280307233  POINT (-122.17743 47.41185)
3                  41.0       280786565   POINT (-122.03439 47.5301)
4                  43.0       198988891  POINT (-122.35436 47.67596)

                             Electric Utility  2020 Census Tract
0                        PUGET SOUND ENERGY INC       5.306701e+10
1                        PUGET SOUND ENERGY INC       5.306105e+10
2  PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)       5.303303e+10
3  PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)       5.303302e+10
4   CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)       5.303300e+10
count    270257.000000
mean         40.386332
std          79.342202
min           0.000000
25%           0.000000
50%           0.000000
75%          33.000000
max         337.000000
Name: Electric Range, dtype: float64
```

In [55]:

```
zero_range_count = (df['Electric Range'] == 0).sum()
total_count = len(df)
print(f"Total entries: {total_count}")
print(f"Entries with 0 range: {zero_range_count} ({zero_range_count/total_count:.2%})")

# Let's see how many non-zero ranges we have
non_zero_range = df[df['Electric Range'] > 0]
print(f"Entries with range > 0: {len(non_zero_range)}")

# Group by Model Year and see average range
print(df.groupby('Model Year')['Electric Range'].mean())
```

```
Total entries: 270262
Entries with 0 range: 169872 (62.85%)
Entries with range > 0: 100385
Model Year
1999     74.000000
2000     58.000000
2002     95.000000
2003     95.000000
2008    209.000000
2010    232.391304
2011     71.434494
2012     59.727532
2013     78.468288
2014     78.598200
2015     95.922348
2016    101.997860
2017    117.755054
2018    156.884058
2019    176.246323
2020    237.886850
2021     12.569517
2022      4.733711
2023      3.842239
2024      7.460764
2025      7.484089
2026      1.920724
Name: Electric Range, dtype: float64
```