In [4]:
```
!pip install pandas numpy matplotlib seaborn scikit-learn
```

```
Requirement already satisfied: pandas in c:\users\user\appdata\local\programs\python\python312\lib\site-packages (2.2.2)
Requirement already satisfied: numpy in c:\users\user\appdata\local\programs\python\python312\lib\site-packages (2.0.2)
Requirement already satisfied: matplotlib in c:\users\user\appdata\local\programs\python\python312\lib\site-packages (3.9.
0)
Requirement already satisfied: seaborn in c:\users\user\appdata\local\programs\python\python312\lib\site-packages (0.13.2)
Requirement already satisfied: scikit-learn in c:\users\user\appdata\local\programs\python\python312\lib\site-packages (1.
5.1)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\user\appdata\local\programs\python\python312\lib\site-pac
kages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\user\appdata\local\programs\python\python312\lib\site-packages (fro
m pandas) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in c:\users\user\appdata\local\programs\python\python312\lib\site-packages (f
rom pandas) (2024.1)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\user\appdata\local\programs\python\python312\lib\site-packages
(from matplotlib) (1.2.1)
Requirement already satisfied: cycler>=0.10 in c:\users\user\appdata\local\programs\python\python312\lib\site-packages (fro
m matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\user\appdata\local\programs\python\python312\lib\site-packages
(from matplotlib) (4.53.1)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\user\appdata\local\programs\python\python312\lib\site-packages
(from matplotlib) (1.4.5)
Requirement already satisfied: packaging>=20.0 in c:\users\user\appdata\local\programs\python\python312\lib\site-packages
(from matplotlib) (24.1)
Requirement already satisfied: pillow>=8 in c:\users\user\appdata\local\programs\python\python312\lib\site-packages (from m
atplotlib) (10.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\user\appdata\local\programs\python\python312\lib\site-packages
(from matplotlib) (3.1.2)
Requirement already satisfied: scipy>=1.6.0 in c:\users\user\appdata\local\programs\python\python312\lib\site-packages (fro
m scikit-learn) (1.13.1)
Requirement already satisfied: joblib>=1.2.0 in c:\users\user\appdata\local\programs\python\python312\lib\site-packages (fr
om scikit-learn) (1.4.2)
Requirement already satisfied: threadpoolctl>=3.1.0 in c:\users\user\appdata\local\programs\python\python312\lib\site-packa
ges (from scikit-learn) (3.5.0)
Requirement already satisfied: six>=1.5 in c:\users\user\appdata\local\programs\python\python312\lib\site-packages (from py
thon-dateutil>=2.8.2->pandas) (1.16.0)
```

In [3]:
```
!python -m pip install --upgrade pip
```

```
Requirement already satisfied: pip in c:\users\user\appdata\local\programs\python\python312\lib\site-packages (24.3.1)
Collecting pip
  Downloading pip-25.1.1-py3-none-any.whl.metadata (3.6 kB)
Downloading pip-25.1.1-py3-none-any.whl (1.8 MB)
   ---------------------------------------- 0.0/1.8 MB ? eta -:--:--
   ----------- ---------------------------- 0.5/1.8 MB 4.2 MB/s eta 0:00:01
   ---------------------------- ----------- 1.3/1.8 MB 4.2 MB/s eta 0:00:01
   ---------------------------------- ----- 1.6/1.8 MB 4.0 MB/s eta 0:00:01
   ---------------------------------------- 1.8/1.8 MB 2.4 MB/s eta 0:00:00
Installing collected packages: pip
  Attempting uninstall: pip
    Found existing installation: pip 24.3.1
    Uninstalling pip-24.3.1:
      Successfully uninstalled pip-24.3.1
Successfully installed pip-25.1.1
```

In [5]:
```
import joblib
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.preprocessing import LabelEncoder
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
from sklearn.model_selection import RandomizedSearchCV
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
```

In [10]:
```
import pandas as pd

df = pd.read_csv("Electric_Vehicle_Population_By_County.csv", on_bad_lines='skip', engine='python')
```

In [11]:
```
# Load data
df = pd.read_csv("Electric_Vehicle_Population_By_County.csv")
```

In [12]:
```
df.head()
```

Out[12]:

	Date	County	State	Vehicle Primary Use	Battery EVs (BEVs)	Plug-In Hybrid EVs (PHEVs)	EV Total	Non-EV Total	Total Vehicles	Percent EV
0	September 30 2022	Riverside	CA	Passenger	7	0	7	460	467	1.50
1	December 31 2022	Prince William	VA	Passenger	1	2	3	188	191	1.57
2	January 31 2020	Dakota	MN	Passenger	0	1	1	32	33	3.03
3	June 30 2022	Ferry	WA	Truck	0	0	0	3,575	3,575	0.00
4	July 31 2021	Douglas	CO	Passenger	0	1	1	83	84	1.19

In [13]:
```
# no of rows and cols
df.shape
```

Out[13]:
```
(20819, 10)
```

In [14]:
```
# Data Types, class and memory alloc
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20819 entries, 0 to 20818
Data columns (total 10 columns):
 #   Column                      Non-Null Count  Dtype  
---  ------                      --------------  -----  
 0   Date                        20819 non-null  object 
 1   County                      20733 non-null  object 
 2   State                       20733 non-null  object 
 3   Vehicle Primary Use         20819 non-null  object 
 4   Battery EVs (BEVs)          20819 non-null  object 
 5   Plug-In Hybrid EVs (PHEVs)  20819 non-null  object 
 6   EV Total                    20819 non-null  object 
 7   Non-EV Total                20819 non-null  object 
 8   Total Vehicles              20819 non-null  object 
 9   Percent EV                  20819 non-null  float64
dtypes: float64(1), object(9)
memory usage: 1.6+ MB
```

In [15]:
```
df.isnull().sum()
```

Out[15]:
```
Date                          0
County                       86
State                        86
Vehicle Primary Use           0
Battery EVs (BEVs)            0
Plug-In Hybrid EVs (PHEVs)    0
EV Total                      0
Non-EV Total                  0
Total Vehicles                0
Percent EV                    0
dtype: int64
```

In [17]:
```
# Compute Q1 and Q3
Q1 = df['Percent EV'].quantile(0.25)
Q3 = df['Percent EV'].quantile(0.75)
IQR = Q3 - Q1

# Define outlier boundaries
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR
print('lower_bound:', lower_bound)
print('upper_bound:', upper_bound)

# Identify outliers
outliers = df[(df['Percent EV'] < lower_bound) | (df['Percent EV'] > upper_bound)]
print("Number of outliers in 'Percent EV':", outliers.shape[0])
```

```
lower_bound: -3.5174999999999996
upper_bound: 6.9025
Number of outliers in 'Percent EV': 2476
```

In [18]:
```
# Converts the "Date" column to actual datetime objects
df['Date'] = pd.to_datetime(df['Date'], errors='coerce')

# Removes rows where "Date" conversion failed
df = df[df['Date'].notnull()]

# Removes rows where the target (EV Total) is missing
df = df[df['EV Total'].notnull()]

# Fill missing values
df['County'] = df['County'].fillna('Unknown')
```

```
df['State'] = df['State'].fillna('Unknown')

# Confirm remaining nulls
print("Missing after fill:")
print(df[['County', 'State']].isnull().sum())

df.head()
```

```
Missing after fill:
County    0
State     0
dtype: int64
```

Out[18]:

	Date	County	State	Vehicle Primary Use	Battery EVs (BEVs)	Plug-In Hybrid EVs (PHEVs)	EV Total	Non-EV Total	Total Vehicles	Percent EV
0	2022-09-30	Riverside	CA	Passenger	7	0	7	460	467	1.50
1	2022-12-31	Prince William	VA	Passenger	1	2	3	188	191	1.57
2	2020-01-31	Dakota	MN	Passenger	0	1	1	32	33	3.03
3	2022-06-30	Ferry	WA	Truck	0	0	0	3,575	3,575	0.00
4	2021-07-31	Douglas	CO	Passenger	0	1	1	83	84	1.19

In [20]:

```
# Cap the outliers - it keeps all the data while reducing the skew from extreme values.

df['Percent EV'] = np.where(df['Percent EV'] > upper_bound, upper_bound,
                            np.where(df['Percent EV'] < lower_bound, lower_bound, df['Percent EV']))

# Identify outliers
outliers = df[(df['Percent EV'] < lower_bound) | (df['Percent EV'] > upper_bound)]
print("Number of outliers in 'Percent EV':", outliers.shape[0])
```

```
Number of outliers in 'Percent EV': 0
```

In [ ]: