US Electric Car & Renewable Energy

In [1]:
```
# importing major libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px

# additional libraries
import warnings
warnings.filterwarnings('ignore')
```

In [30]:
```
# 1. State codes
df_state_codes = pd.read_excel("state_codes.xlsx")

# 2. All vehicle registrations (2018)
df_all_vehicle = pd.read_excel("States_All_Vehicle_Registrations_2018.xlsx")

# 3. Electric vehicle registrations (2018)
df_ev = pd.read_excel("States_Electric_Vehicle_Registrations_2018 (1).xlsx")

# 4. Annual energy generation (1990–2019)
df_energy = pd.read_excel("States_Annual_Energy_Generation_Sources_1990_2019.x
```

In [31]:
```
print("EV columns:\n", ev.columns)
print("\nEnergy columns:\n", energy.columns)
print("\nAll vehicle columns:\n", all_vehicle.columns)
print("\nState columns:\n", states.columns)
```

```
EV columns:
 Index(['unnamed:_0', 'unnamed:_1', 'unnamed:_2', 'unnamed:_3', 'unnamed:_4'],
dtype='object')

Energy columns:
 Index(['state_historical_tables_for_2019\nreleased:_september_2020\nrevised:_f
ebruary_2021\nnext_update:_november_2021',
       'unnamed:_1', 'unnamed:_2', 'unnamed:_3', 'unnamed:_4'],
      dtype='object')

All vehicle columns:
 Index([' ', 'Unnamed: 1', 'Unnamed: 2', 'Unnamed: 3', 'Unnamed: 4',
       'Unnamed: 5', 'Unnamed: 6', 'Unnamed: 7', 'Unnamed: 8', 'Unnamed: 9',
       'Unnamed: 10', 'Unnamed: 11', 'Unnamed: 12', 'Unnamed: 13',
       'Unnamed: 14', 'Unnamed: 15'],
      dtype='object')

State columns:
 Index(['state_code', 'state_name'], dtype='object')
```

In [32]:
```
ev_state = ev.iloc[:, 0]        # State name / code
```

```
ev_year = ev.iloc[:, 1]         # Year
ev_count = ev.iloc[:, 2]        # Electric vehicle registrations
```

In [33]:
```
energy_state = energy.iloc[:, 0]      # State
energy_year = energy.iloc[:, 1]       # Year
renewable_energy = energy.iloc[:, 2]  # Renewable energy generation
```

In [34]:
```
vehicle_state = all_vehicle.iloc[:, 0]   # State
vehicle_year = all_vehicle.iloc[:, 1]    # Year
total_vehicles = all_vehicle.iloc[:, 2]  # Total vehicle registrations
```

In [35]:
```
state_code = states.iloc[:, 0]   # State code
state_name = states.iloc[:, 1]   # State full name
```

In [36]:
```
ev_year = pd.to_numeric(ev_year, errors="coerce")
ev_count = pd.to_numeric(ev_count, errors="coerce")

energy_year = pd.to_numeric(energy_year, errors="coerce")
renewable_energy = pd.to_numeric(renewable_energy, errors="coerce")

vehicle_year = pd.to_numeric(vehicle_year, errors="coerce")
total_vehicles = pd.to_numeric(total_vehicles, errors="coerce")
```

In [37]:
```
ev_state = ev.iloc[:, 0]
ev_year = pd.to_numeric(ev.iloc[:, 1], errors="coerce")
ev_count = pd.to_numeric(ev.iloc[:, 2], errors="coerce")

energy_state = energy.iloc[:, 0]
energy_year = pd.to_numeric(energy.iloc[:, 1], errors="coerce")
renewable_energy = pd.to_numeric(energy.iloc[:, 2], errors="coerce")

vehicle_state = all_vehicle.iloc[:, 0]
vehicle_year = pd.to_numeric(all_vehicle.iloc[:, 1], errors="coerce")
total_vehicles = pd.to_numeric(all_vehicle.iloc[:, 2], errors="coerce")
```

In [38]:
```
print(ev.head())
print(energy.head())
print(states.head())
```

```
   unnamed:_0                                 unnamed:_1          unnamed:_2  \
0         0.0  Electric Vehicle Registrations by State                   0
1         0.0                                   State  Registration Count
2         0.0                                 Alabama                1450
3         0.0                                  Alaska                 530
4         0.0                                 Arizona               15000

   unnamed:_3 unnamed:_4
0         0.0          0
1         0.0          0
2         0.0          0
3         0.0          0
4         0.0          0
  state_historical_tables_for_2019\nreleased:_september_2020\nrevised:_februar
y_2021\nnext_update:_november_2021  \
0                                               YEAR
1                                               1990
2                                               1990
3                                               1990
4                                               1990

  unnamed:_1                    unnamed:_2                 unnamed:_3  \
0      STATE               TYPE OF PRODUCER              ENERGY SOURCE
1         AK  Total Electric Power Industry                      Total
2         AK  Total Electric Power Industry                       Coal
3         AK  Total Electric Power Industry  Hydroelectric Conventional
4         AK  Total Electric Power Industry                Natural Gas

                  unnamed:_4
0  GENERATION (Megawatthours)
1                     5599506
2                      510573
3                      974521
4                     3466261
  state_code  state_name
0         AK      Alaska
1         AL     Alabama
2         AR    Arkansas
3         AZ     Arizona
4         CA  California
```

In [39]:
```
ev.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 54 entries, 0 to 53
Data columns (total 5 columns):
 #   Column      Non-Null Count  Dtype  
---  ------      --------------  -----  
 0   unnamed:_0  54 non-null     float64
 1   unnamed:_1  54 non-null     object 
 2   unnamed:_2  54 non-null     object 
 3   unnamed:_3  54 non-null     float64
 4   unnamed:_4  54 non-null     object 
dtypes: float64(2), object(3)
memory usage: 2.2+ KB
```

In [40]:
```
energy.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 53757 entries, 0 to 53756
Data columns (total 5 columns):
 #   Column
Non-Null Count  Dtype
---  ------
--------------  -----
 0   state_historical_tables_for_2019
released:_september_2020
revised:_february_2021
next_update:_november_2021  53757 non-null  object
 1   unnamed:_1
53757 non-null  object
 2   unnamed:_2
53757 non-null  object
 3   unnamed:_3
53757 non-null  object
 4   unnamed:_4
53757 non-null  object
dtypes: object(5)
memory usage: 2.1+ MB
```

In [41]:
```
states.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51 entries, 0 to 50
Data columns (total 2 columns):
 #   Column      Non-Null Count  Dtype 
---  ------      --------------  ----- 
 0   state_code  51 non-null     object
 1   state_name  51 non-null     object
dtypes: object(2)
memory usage: 948.0+ bytes
```

In [42]:
```
ev.drop_duplicates(inplace=True)
energy.drop_duplicates(inplace=True)
states.drop_duplicates(inplace=True)
```

In [43]:
```
# Data Cleaning
```

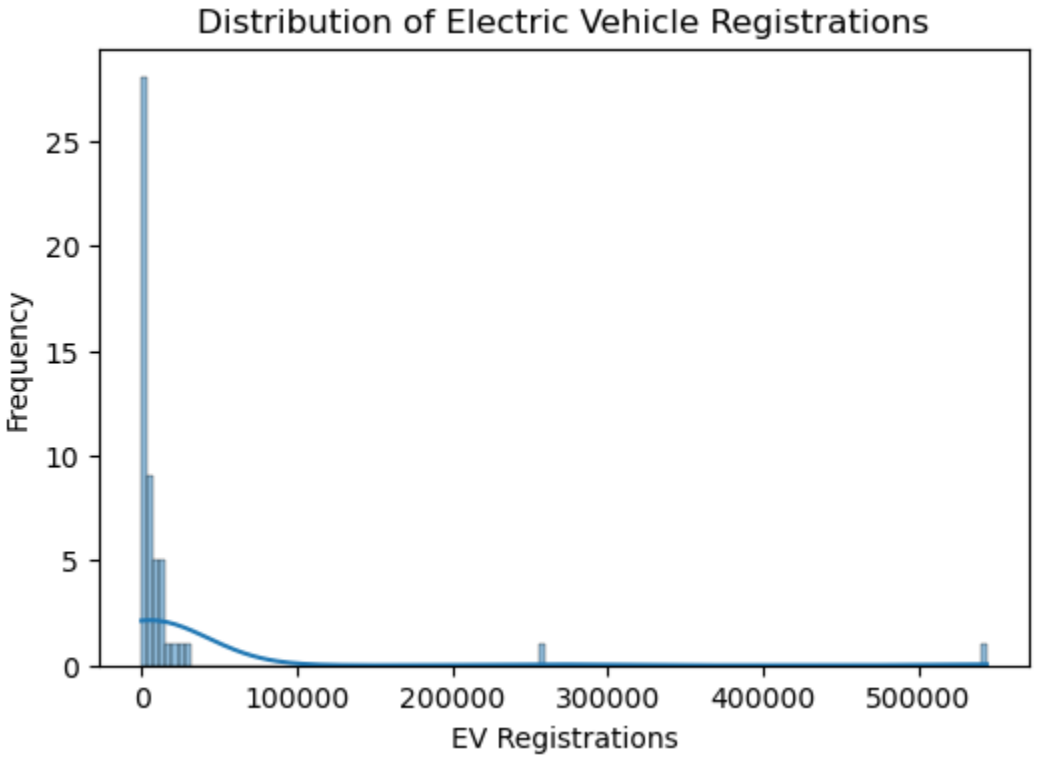
```
ev.drop_duplicates(inplace=True)
energy.drop_duplicates(inplace=True)
states.drop_duplicates(inplace=True)
```

In [44]:
```
#Missing Values Check
ev.isnull().sum()
energy.isnull().sum()
ev.fillna(0, inplace=True)
energy.fillna(0, inplace=True)
```

In [45]:
```
#Column Name Standardization
ev.columns = ev.columns.str.lower().str.replace(" ", "_")
energy.columns = energy.columns.str.lower().str.replace(" ", "_")
states.columns = states.columns.str.lower().str.replace(" ", "_")
```

Exploratory Data Analysis (EDA)

In [46]:
```
#EV Registrations Distribution
plt.figure(figsize=(6,4))
sns.histplot(ev_count.dropna(), kde=True)
plt.title("Distribution of Electric Vehicle Registrations")
plt.xlabel("EV Registrations")
plt.ylabel("Frequency")
plt.show()
```

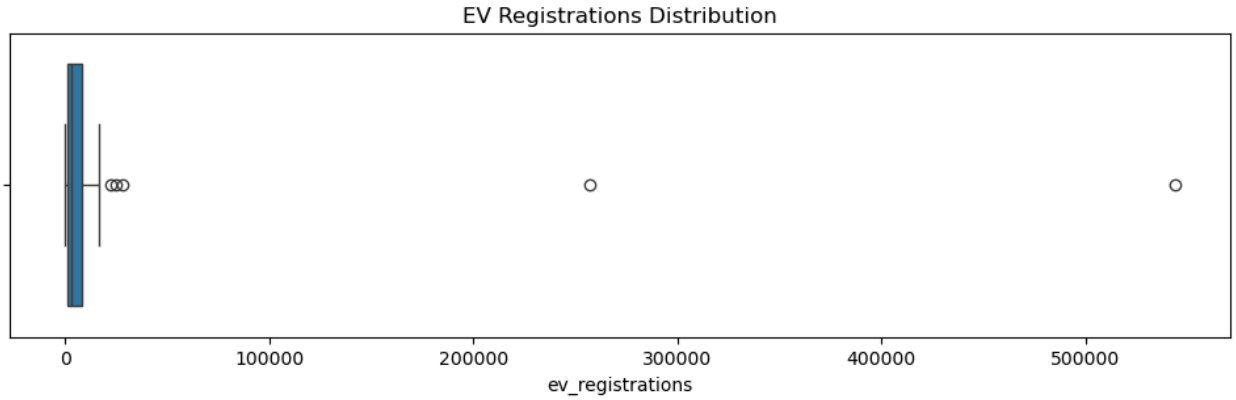


Insight

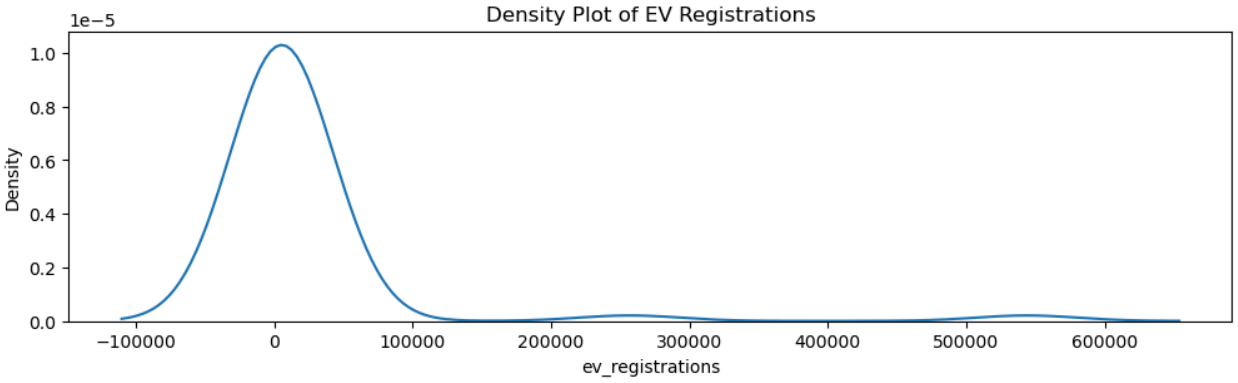
Most states have low EV adoption, while a few states dominate registrations.

In [47]:
```
df = pd.DataFrame()
df["ev_registrations"] = (
    ev.iloc[:, 2]
    .astype(str)
    .str.replace(",", "", regex=True))
df["ev_registrations"] = pd.to_numeric(
    df["ev_registrations"], errors="coerce")
# Remove empty values
df.dropna(inplace=True)
```

In [48]:
```
#BOX PLOT
plt.figure(figsize=(12,3))
sns.boxplot(x=df['ev_registrations'])
plt.title('EV Registrations Distribution')
plt.show()
```



In [49]:
```
#KDE PLOT
plt.figure(figsize=(12,3))
sns.kdeplot(x=df['ev_registrations'])
plt.title('Density Plot of EV Registrations')
plt.show()
```



In [50]:
```
df['ev_registrations'].skew()
```

Out[50]: np.float64(5.8210856473184)

Insights

EV registrations are right-skewed, indicating that a few states have very high EV adoption while most states have lower registrations.

Key Findings (Summary)

- EV adoption is concentrated in renewable-energy-rich states
- Solar and wind energy have strong relationships with EV growth
- Policy support likely plays a key role in high-adoption states

Conclusion & Recommendations

- States should invest in renewable infrastructure to accelerate EV adoption
- Policymakers must align EV incentives with clean energy goals
- Future studies can include predictive modeling and policy variables

End of Project

- This capstone project demonstrates the strong linkage between clean energy production and sustainable transportation adoption in the United States.

In [ ]: