In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
Data1=pd.read_excel(r"C:\Users\faij2\Downloads\DATA SET OF PROJECTS\ELECTRIC V
```

In [3]:
```
Data2=pd.read_excel(r"C:\Users\faij2\Downloads\DATA SET OF PROJECTS\ELECTRIC V
```

In [4]:
```
Data3=pd.read_excel(r"C:\Users\faij2\Downloads\DATA SET OF PROJECTS\ELECTRIC V
```

In [5]:
```
Data4=pd.read_excel(r"C:\Users\faij2\Downloads\DATA SET OF PROJECTS\ELECTRIC V
```

In [6]:
```
Data4=Data4[['State','Registration Count']]
Data4.columns=['STATE','REGISTRATION']
Data4.head()
```

Out[6]:

	STATE	REGISTRATION
0	Alabama	1450
1	Alaska	530
2	Arizona	15000
3	Arkansas	520
4	California	256800

In [7]:
```
Data4=Data4.drop([51],axis=0)
```

In [8]:
```
Data4
```

Out[8]:

	STATE	REGISTRATION
0	Alabama	1450
1	Alaska	530
2	Arizona	15000
3	Arkansas	520
4	California	256800
5	Colorado	11700
6	Connecticut	4450
7	Delaware	720
8	District Of Columbia	970
9	Florida	25200
10	Georgia	15300
11	Hawaii	6590
12	Idaho	1080
13	Illinois	12400
14	Indiana	3030
15	Iowa	1090
16	Kansas	1610
17	Kentucky	1240
18	Louisiana	1110
19	Maine	750
20	Maryland	8080
21	Massachusetts	9760
22	Michigan	4210
23	Minnesota	4740
24	Mississippi	390
25	Missouri	3450
26	Montana	500
27	Nebraska	850
28	Nevada	4810
29	New Hampshire	1120
30	New Jersey	12100

	STATE	REGISTRATION
31	New Mexico	1260
32	New York	16600
33	North Carolina	7320
34	North Dakota	170
35	Ohio	6510
36	Oklahoma	3290
37	Oregon	12400
38	Pennsylvania	7990
39	Rhode Island	600
40	South Carolina	1950
41	South Dakota	260
42	Tennessee	3980
43	Texas	22600
44	Utah	5220
45	Vermont	1060
46	Virginia	8370
47	Washington	28400
48	West Virginia	230
49	Wisconsin	3680
50	Wyoming	170

In [9]:
```
Data2=Data2.drop([62,63],axis=0)
```

In [10]:
```
Data2.columns=["STATE","AUTOMOBILE_PRIVATE","AUTOMOBILE_PUBLICLY","AUTOMOBILE_
```

In [11]:
```
Data2.drop_duplicates()
Data1.drop_duplicates()
Data3.drop_duplicates()
Data3.describe()
```

Out[11]:

	YEAR	GENERATION (Megawatthours)
count	53756.000000	5.375600e+04
mean	2005.518844	1.693131e+07
std	8.598414	1.309890e+08
min	1990.000000	-8.823445e+06
25%	1998.000000	2.684825e+04
50%	2006.000000	3.279665e+05
75%	2013.000000	3.405300e+06
max	2019.000000	4.178277e+09

In [12]:

```
Data1.describe()
```

Out[12]:

	state_code	state_name
count	51	51
unique	51	51
top	AK	Alaska
freq	1	1

In [13]:

```
Data2.describe()
```

Out[13]:

	STATE	AUTOMOBILE_PRIVATE	AUTOMOBILE_PUBLICLY	AUTOMOI
count	54	57	53	
unique	54	57	53	
top	STATE MOTOR-VEHICLE REGISTRATIONS - 2018	(Revised February 2021)	PUBLICLY	
freq	1	1	1	

In [14]:

```
Data3.describe()
```

Out[14]:

	YEAR	GENERATION (Megawatthours)
count	53756.000000	5.375600e+04
mean	2005.518844	1.693131e+07
std	8.598414	1.309890e+08
min	1990.000000	-8.823445e+06
25%	1998.000000	2.684825e+04
50%	2006.000000	3.279665e+05
75%	2013.000000	3.405300e+06
max	2019.000000	4.178277e+09

In [15]:

```
Data3=Data3.drop([0],axis=0)
```

In [16]:

```
Data2
```

Out[16]:

	STATE	AUTOMOBILE_PRIVATE	AUTOMOBILE_PUBLICLY	AUTOMOBILE_
0	STATE MOTOR-VEHICLE REGISTRATIONS - 2018	NaN	NaN	
1	NaN	NaN	NaN	
2	NaN	NaN	NaN	
3	December 2019	(Revised February 2021)	NaN	
4	NaN	NaN	NaN	
...	...	...	...	
57	Virginia	3222933	44802	3
58	Washington	2897723	67216	2
59	West Virginia (2)	547961	12157	
60	Wisconsin	2055788	31730	2
61	Wyoming	200217	3329	

62 rows × 16 columns

In [17]:

```
Data2.drop([0,1,2,3,4,5,6,7,8,9,10],axis=0,inplace=True)
```

In [18]:

```
Data2
```

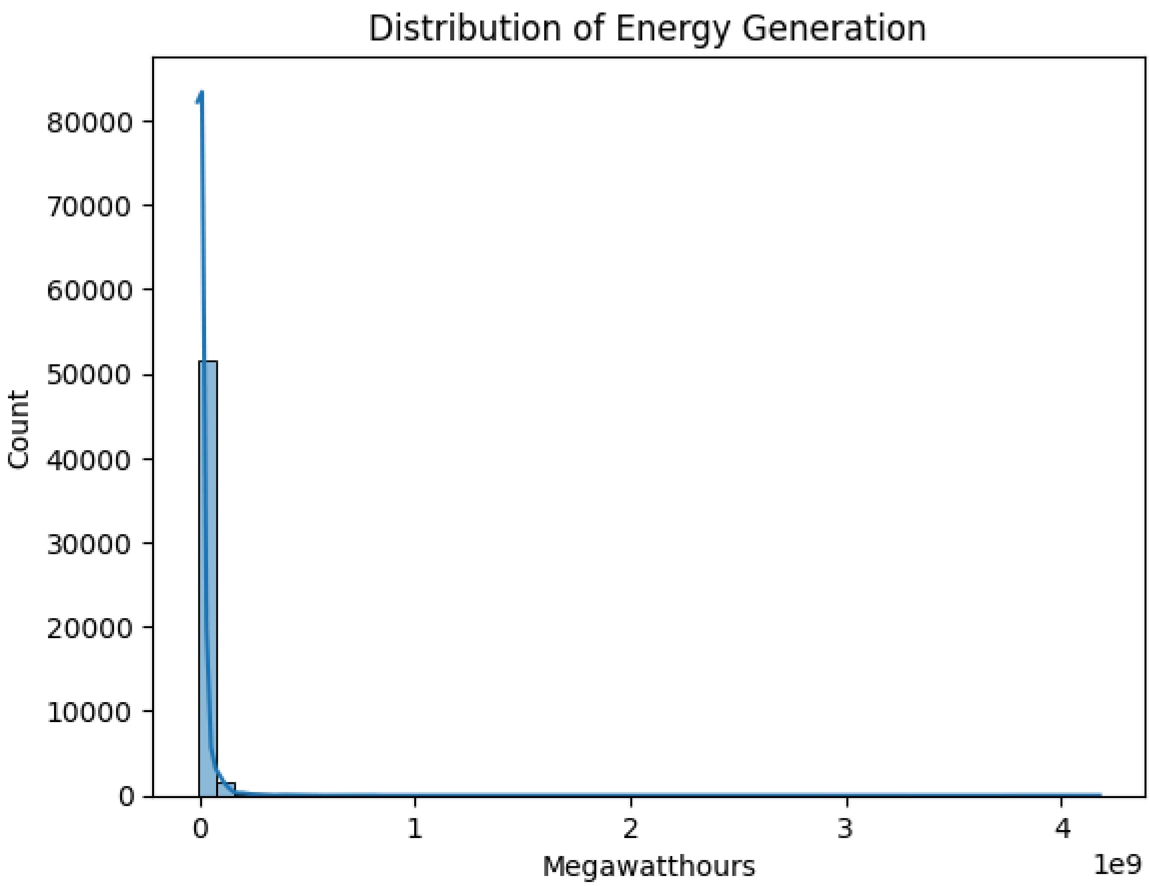
Out[18]:

	STATE	AUTOMOBILE_PRIVATE	AUTOMOBILE_PUBLICLY	AUTOMOBILE_T
11	Alabama	2116626	44586	21
12	Alaska	179131	4139	1
13	Arizona	2372443	19329	23
14	Arkansas	908561	12600	9
15	California	14820833	244994	150
16	Colorado	1782358	15819	17
17	Connecticut	1305544	1165	13
18	Delaware	431850	1513	4
19	Dist. of Col.	188768	20955	2
20	Florida	7851192	114899	79
21	Georgia	3502070	55399	35
22	Hawaii	502165	7327	5
23	Idaho	596208	2566	5
24	Illinois	4438811	38952	44
25	Indiana	2245862	3008	22
26	Iowa	1230488	11731	12
27	Kansas	970921	4250	9
28	Kentucky	1689705	32237	17
29	Louisiana	1360559	28690	13
30	Maine	386449	4057	3
31	Maryland	1893626	28837	19
32	Massachusetts (2)	2178472	4058	21
33	Michigan	3000593	23347	30
34	Minnesota	1959810	16715	19
35	Mississippi	822799	2539	8
36	Missouri	2083825	18391	21
37	Montana	452286	559	4
38	Nebraska (2)	668345	14675	6
39	Nevada	1065766	7994	10
40	New Hampshire	504138	2821	5

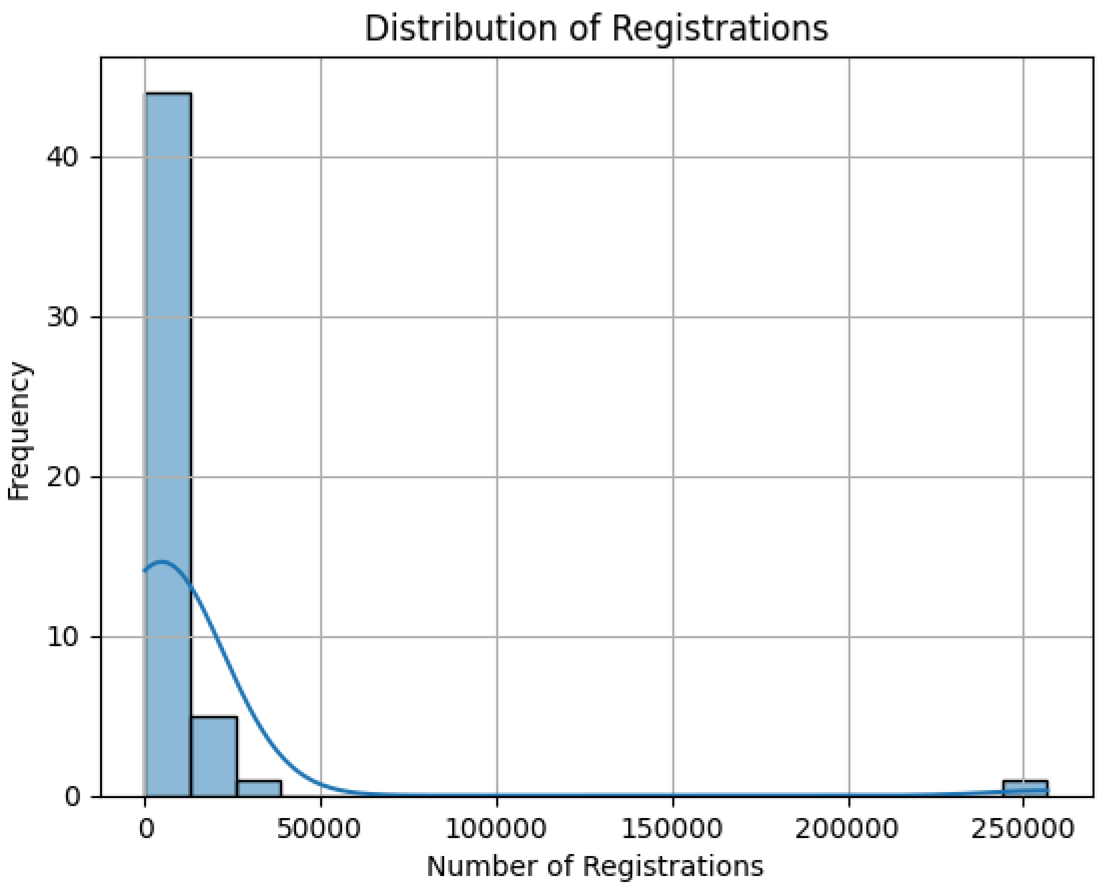
	STATE	AUTOMOBILE_PRIVATE	AUTOMOBILE_PUBLICLY	AUTOMOBILE_T
41	New Jersey	2730210	24043	27
42	New Mexico	648720	7046	6
43	New York	4690752	22027	47
44	North Carolina	3346226	47555	33
45	North Dakota	235922	4126	2
46	Ohio	4558551	45043	46
47	Oklahoma (2)	1290040	6179	12
48	Oregon	1462295	26328	14
49	Pennsylvania	4375444	48739	44
50	Rhode Island	404987	7268	4
51	South Carolina	1753043	77143	18
52	South Dakota	352997	5862	3
53	Tennessee	2234092	51237	22
54	Texas	8155311	93011	82
55	Utah	926078	11343	9
56	Vermont	214077	4225	2
57	Virginia	3222933	44802	32
58	Washington	2897723	67216	29
59	West Virginia (2)	547961	12157	5
60	Wisconsin	2055788	31730	20
61	Wyoming	200217	3329	2

PERFORM EXPLORATORY DATA ANALYSIS (EDA) TO UNDERSTAND THE DISTRIBUTION AND CHARACTERISTICS OF THE DATA.

In [19]:
```
sns.histplot(Data3['GENERATION (Megawatthours)'], bins=50, kde=True)
plt.title('Distribution of Energy Generation')
plt.xlabel('Megawatthours')
plt.show()
```

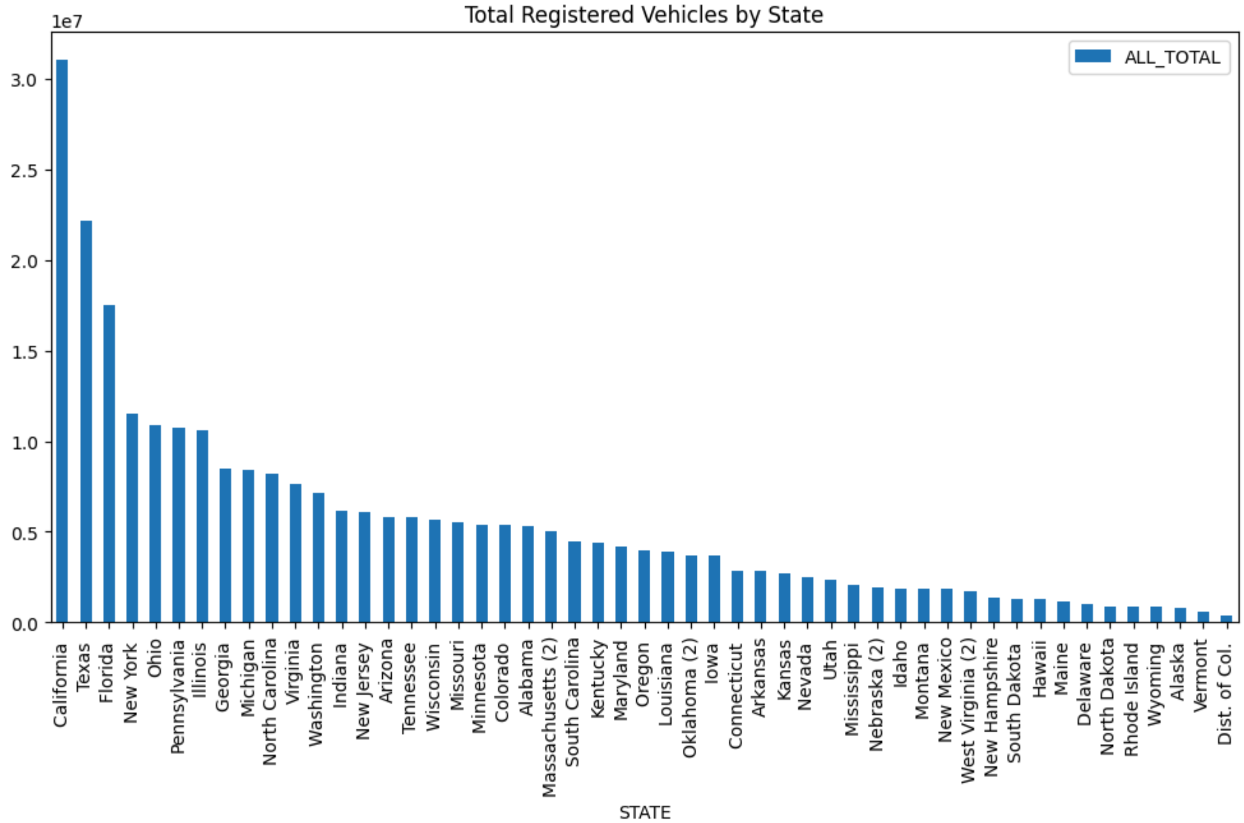


In [20]:
```
sns.histplot(Data4['REGISTRATION'], bins=20, kde=True)
plt.title('Distribution of Registrations')
plt.xlabel('Number of Registrations')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()
```



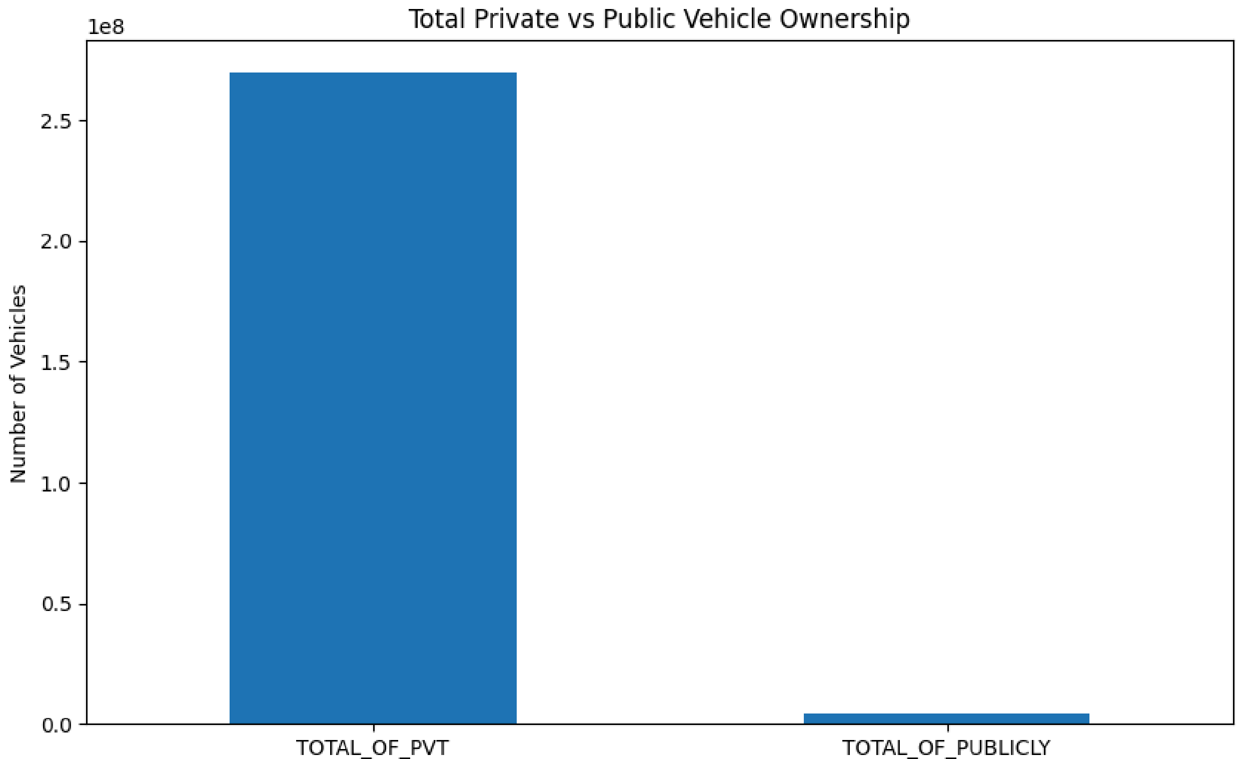
In [21]:
```
df_sorted = Data2.sort_values(by='ALL_TOTAL', ascending=False)
df_sorted.plot(x='STATE', y='ALL_TOTAL', kind='bar', figsize=(12,6), title='To
```

Out[21]:
```
<Axes: title={'center': 'Total Registered Vehicles by State'}, xlabel='STAT
E'>
```



In [22]:

```
plt.figure(figsize=(10,6))
Data2[['TOTAL_OF_PVT', 'TOTAL_OF_PUBLICLY']].sum().plot(kind='bar')
plt.title('Total Private vs Public Vehicle Ownership')
plt.ylabel('Number of Vehicles')
plt.xticks(rotation=0)
plt.show()
```



VISUALIZE DATA USING MATPLOTLIB AND SEABORN TO IDENTIFY TRENDS AND PATTERNS.

In [23]:
```
Data2#.drop(["level_0","index"],axis=1,inplace=True)
```

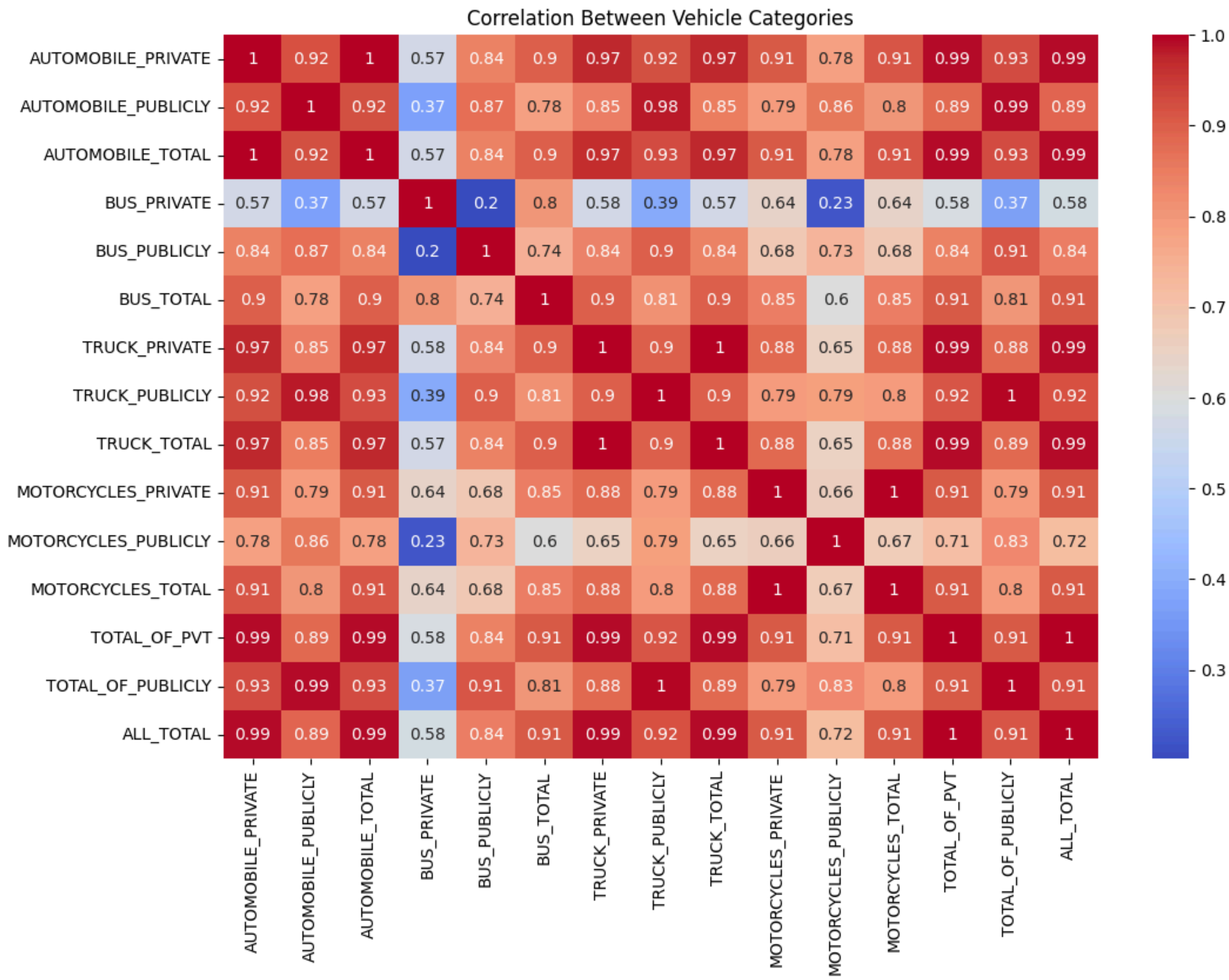
Out[23]:

	STATE	AUTOMOBILE_PRIVATE	AUTOMOBILE_PUBLICLY	AUTOMOBILE_T
11	Alabama	2116626	44586	21
12	Alaska	179131	4139	1
13	Arizona	2372443	19329	23
14	Arkansas	908561	12600	9
15	California	14820833	244994	150
16	Colorado	1782358	15819	17
17	Connecticut	1305544	1165	13
18	Delaware	431850	1513	4
19	Dist. of Col.	188768	20955	2
20	Florida	7851192	114899	79
21	Georgia	3502070	55399	35
22	Hawaii	502165	7327	5
23	Idaho	596208	2566	5
24	Illinois	4438811	38952	44
25	Indiana	2245862	3008	22
26	Iowa	1230488	11731	12
27	Kansas	970921	4250	9
28	Kentucky	1689705	32237	17
29	Louisiana	1360559	28690	13
30	Maine	386449	4057	3
31	Maryland	1893626	28837	19
32	Massachusetts (2)	2178472	4058	21
33	Michigan	3000593	23347	30
34	Minnesota	1959810	16715	19
35	Mississippi	822799	2539	8
36	Missouri	2083825	18391	21
37	Montana	452286	559	4
38	Nebraska (2)	668345	14675	6
39	Nevada	1065766	7994	10
40	New Hampshire	504138	2821	5

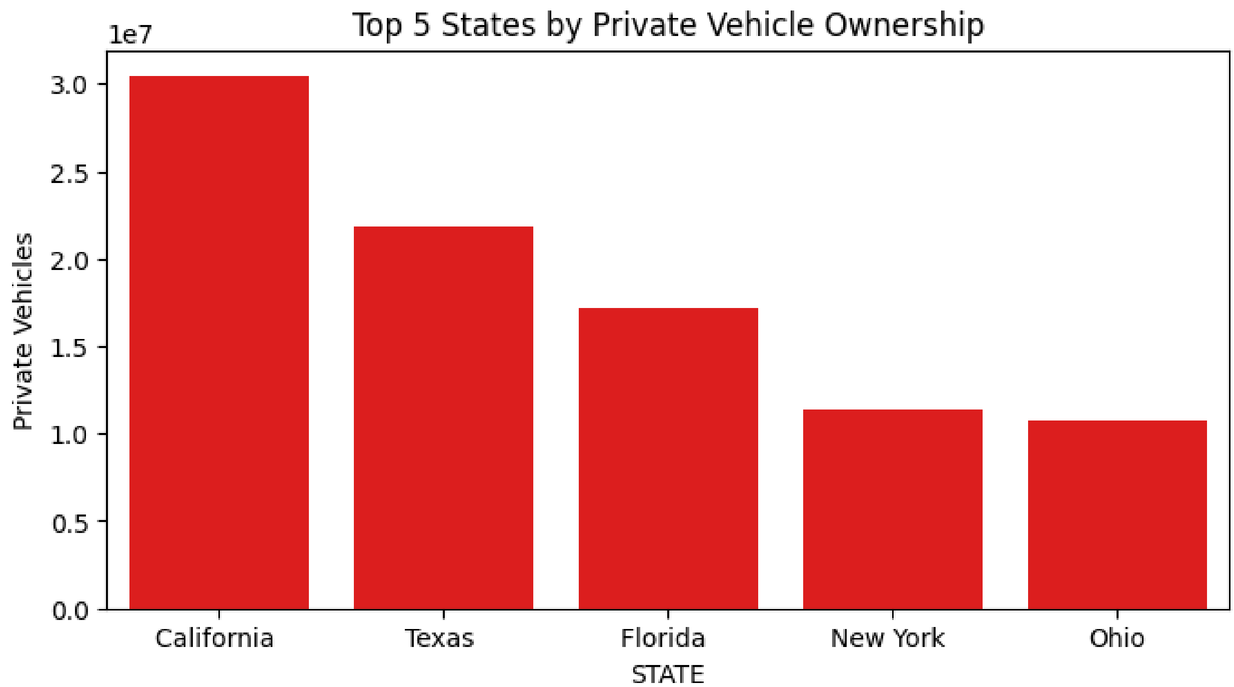
	STATE	AUTOMOBILE_PRIVATE	AUTOMOBILE_PUBLICLY	AUTOMOBILE_T
41	New Jersey	2730210	24043	27
42	New Mexico	648720	7046	6
43	New York	4690752	22027	47
44	North Carolina	3346226	47555	33
45	North Dakota	235922	4126	2
46	Ohio	4558551	45043	46
47	Oklahoma (2)	1290040	6179	12
48	Oregon	1462295	26328	14
49	Pennsylvania	4375444	48739	44
50	Rhode Island	404987	7268	4
51	South Carolina	1753043	77143	18
52	South Dakota	352997	5862	3
53	Tennessee	2234092	51237	22
54	Texas	8155311	93011	82
55	Utah	926078	11343	9
56	Vermont	214077	4225	2
57	Virginia	3222933	44802	32
58	Washington	2897723	67216	29
59	West Virginia (2)	547961	12157	5
60	Wisconsin	2055788	31730	20
61	Wyoming	200217	3329	2

In [24]:

```
plt.figure(figsize=(12,8))
sns.heatmap(Data2.drop(columns='STATE').corr(), annot=True, cmap='coolwarm')
plt.title('Correlation Between Vehicle Categories')
plt.show()
```

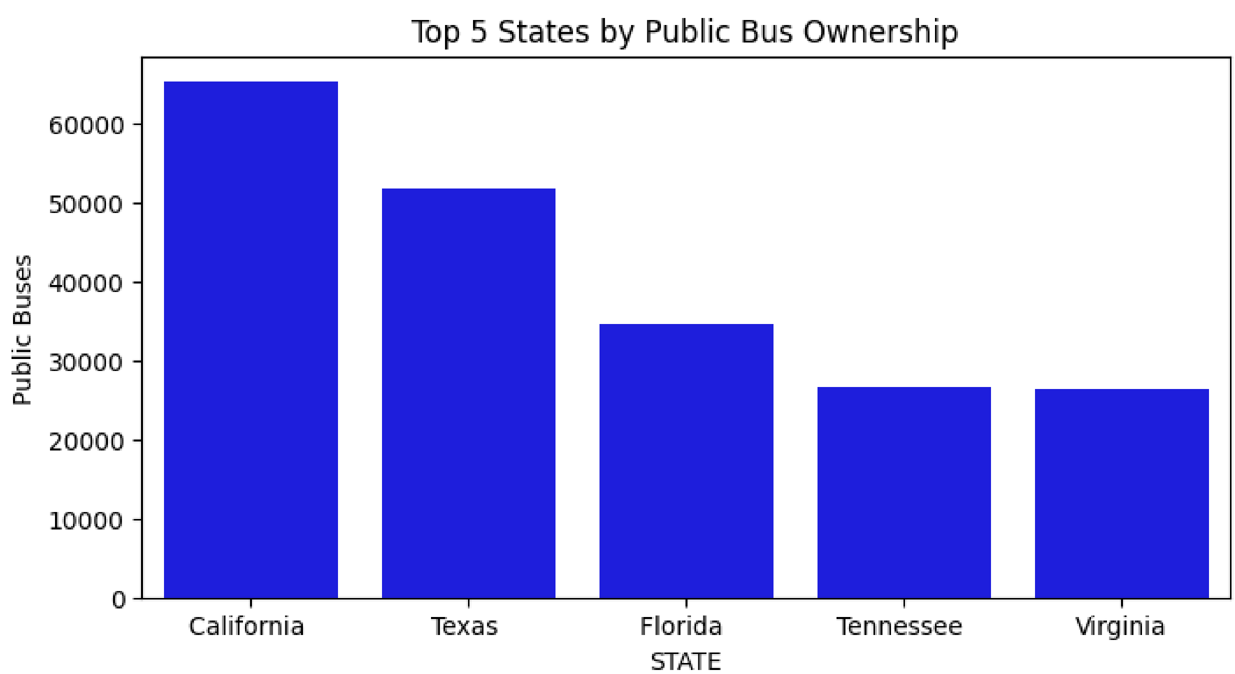


In [25]:
```
top_private = Data2.sort_values(by='TOTAL_OF_PVT', ascending=False).head(5)

plt.figure(figsize=(8,4))
sns.barplot(x='STATE', y='TOTAL_OF_PVT', data=top_private,color="red")
plt.title('Top 5 States by Private Vehicle Ownership')
plt.ylabel('Private Vehicles')
plt.show()
```



In [26]:

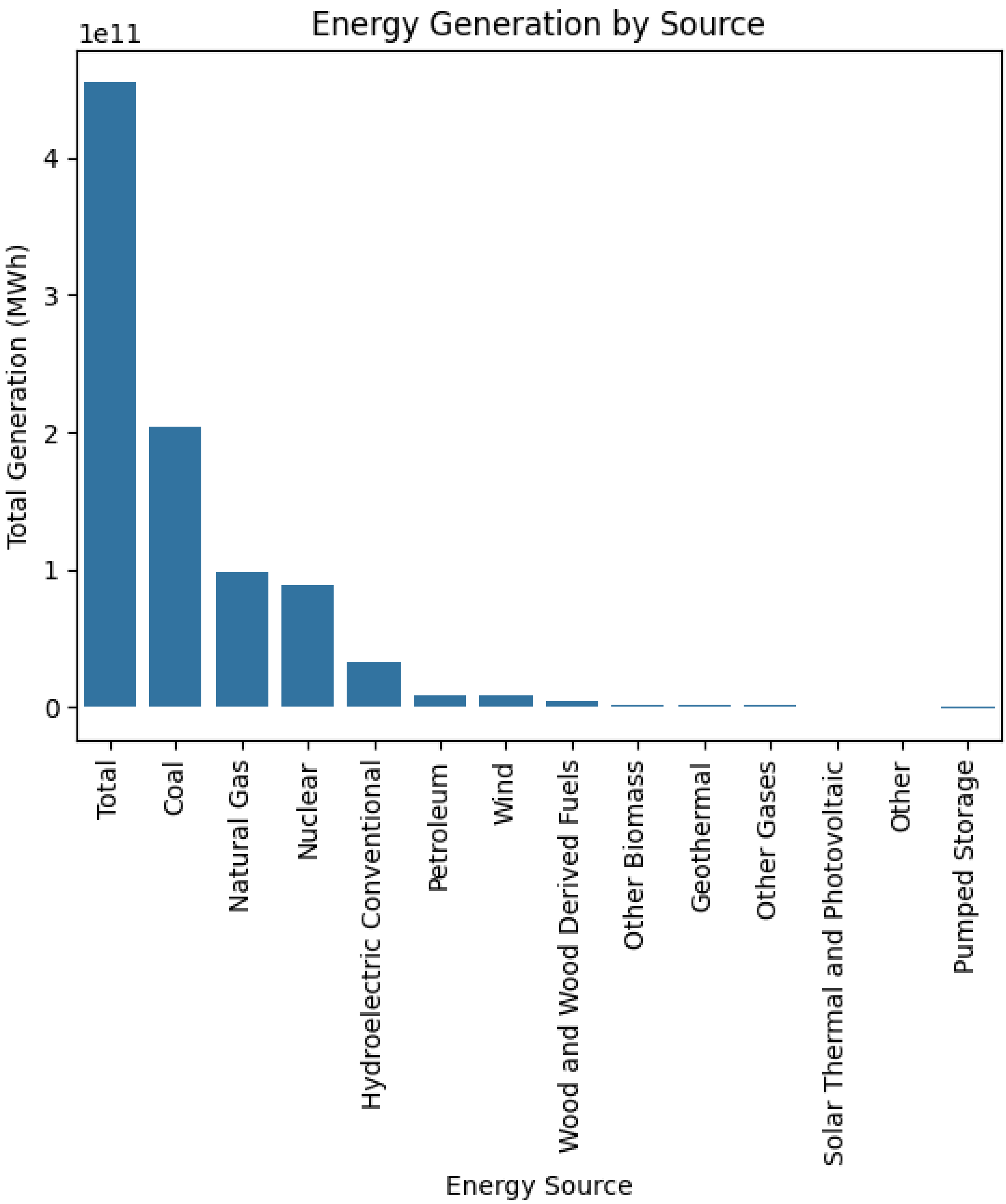
```
top_bus_public = Data2.sort_values(by='BUS_PUBLICLY', ascending=False).head(5)

plt.figure(figsize=(8,4))
sns.barplot(x='STATE', y='BUS_PUBLICLY', data=top_bus_public,color="blue")
plt.title('Top 5 States by Public Bus Ownership')
plt.ylabel('Public Buses')
plt.show()
```



In [27]:

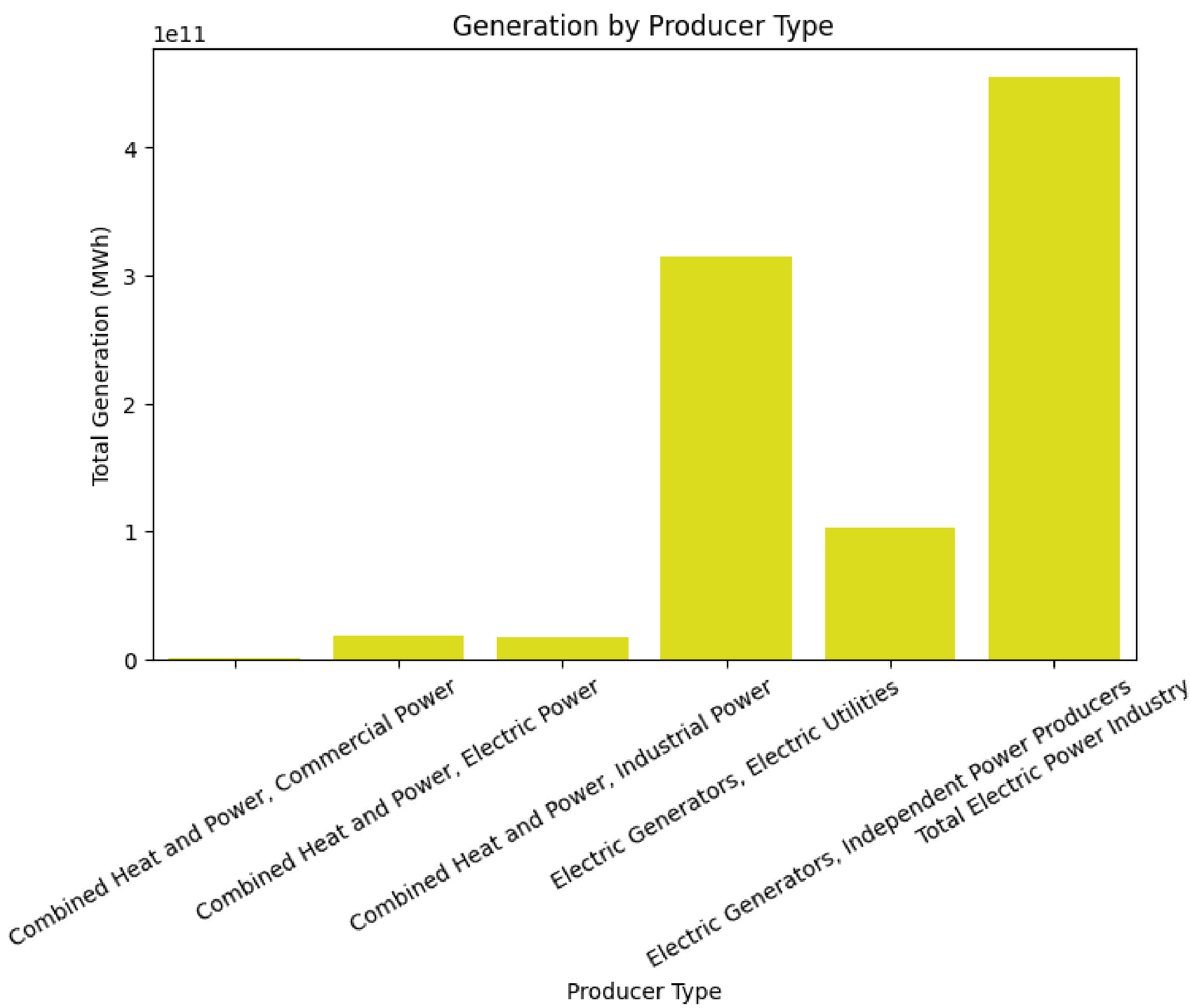
```
sns.barplot(data=Data3.groupby('ENERGY SOURCE')['GENERATION (Megawatthours)'].
            x='ENERGY SOURCE', y='GENERATION (Megawatthours)')
```

```
plt.title('Energy Generation by Source')
plt.xticks(rotation=90)
plt.ylabel('Total Generation (MWh)')
plt.xlabel('Energy Source')
plt.show()
```



In [28]:
```
producer_gen = Data3.groupby('TYPE OF PRODUCER')['GENERATION (Megawatthours)']

plt.figure(figsize=(8,5))
sns.barplot(data=producer_gen, x='TYPE OF PRODUCER', y='GENERATION (Megawattho
plt.title('Generation by Producer Type')
plt.ylabel('Total Generation (MWh)')
plt.xlabel('Producer Type')
plt.xticks(rotation=30)
```

```
plt.show()
```



3.CORRELATION ANALYSIS

3.1 ANALYZE THE CORRELATION BETWEEN ELECTRIC VEHICLE REGISTRATIONS AND DIFFERENT SOURCES OF ENERGY PRODUCTION.

In [29]:
```
state_map = {
    'AK': 'Alaska',
    'AL': 'Alabama',
    'AR': 'Arkansas',
    'AZ': 'Arizona',
    'CA': 'California',
    'CO': 'Colorado',
    'CT': 'Connecticut',
    'DC': 'District of Columbia',
    'DE': 'Delaware',
    'FL': 'Florida',
```

```
    'GA': 'Georgia',
    'HI': 'Hawaii',
    'IA': 'Iowa',
    'ID': 'Idaho',
    'IL': 'Illinois',
    'IN': 'Indiana',
    'KS': 'Kansas',
    'KY': 'Kentucky',
    'LA': 'Louisiana',
    'MA': 'Massachusetts',
    'MD': 'Maryland',
    'ME': 'Maine',
    'MI': 'Michigan',
    'MN': 'Minnesota',
    'MO': 'Missouri',
    'MS': 'Mississippi',
    'MT': 'Montana',
    'NC': 'North Carolina',
    'ND': 'North Dakota',
    'NE': 'Nebraska',
    'NH': 'New Hampshire',
    'NJ': 'New Jersey',
    'NM': 'New Mexico',
    'NV': 'Nevada',
    'NY': 'New York',
    'OH': 'Ohio',
    'OK': 'Oklahoma',
    'OR': 'Oregon',
    'PA': 'Pennsylvania',
    'RI': 'Rhode Island',
    'SC': 'South Carolina',
    'SD': 'South Dakota',
    'TN': 'Tennessee',
    'TX': 'Texas',
    'UT': 'Utah',
    'VA': 'Virginia',
    'VT': 'Vermont',
    'WA': 'Washington',
    'WI': 'Wisconsin',
    'WV': 'West Virginia',
    'WY': 'Wyoming'
}
```

In [30]:

```
Data3['STATE'] = Data3['STATE'].replace(state_map)
Data4
```

Out[30]:

	STATE	REGISTRATION
0	Alabama	1450
1	Alaska	530
2	Arizona	15000
3	Arkansas	520
4	California	256800
5	Colorado	11700
6	Connecticut	4450
7	Delaware	720
8	District Of Columbia	970
9	Florida	25200
10	Georgia	15300
11	Hawaii	6590
12	Idaho	1080
13	Illinois	12400
14	Indiana	3030
15	Iowa	1090
16	Kansas	1610
17	Kentucky	1240
18	Louisiana	1110
19	Maine	750
20	Maryland	8080
21	Massachusetts	9760
22	Michigan	4210
23	Minnesota	4740
24	Mississippi	390
25	Missouri	3450
26	Montana	500
27	Nebraska	850
28	Nevada	4810
29	New Hampshire	1120
30	New Jersey	12100

	STATE	REGISTRATION
31	New Mexico	1260
32	New York	16600
33	North Carolina	7320
34	North Dakota	170
35	Ohio	6510
36	Oklahoma	3290
37	Oregon	12400
38	Pennsylvania	7990
39	Rhode Island	600
40	South Carolina	1950
41	South Dakota	260
42	Tennessee	3980
43	Texas	22600
44	Utah	5220
45	Vermont	1060
46	Virginia	8370
47	Washington	28400
48	West Virginia	230
49	Wisconsin	3680
50	Wyoming	170

In [31]:
```
Data3[Data3["YEAR"]==2018].reset_index()
```

Out[31]:

	index	YEAR	STATE	TYPE OF PRODUCER	ENERGY SOURCE	GENERATION (Megawatthours)
0	49519	2018	Alaska	Total Electric Power Industry	Total	6247359.0
1	49520	2018	Alaska	Total Electric Power Industry	Coal	628564.0
2	49521	2018	Alaska	Total Electric Power Industry	Hydroelectric Conventional	1664225.0
3	49522	2018	Alaska	Total Electric Power Industry	Natural Gas	2947902.0
4	49523	2018	Alaska	Total Electric Power Industry	Other	-3100.0
...	...	...	...	...	...	...
2108	51627	2018	Wyoming	Electric Generators, Electric Utilities	Coal	38641538.0
2109	51628	2018	Wyoming	Electric Generators, Electric Utilities	Hydroelectric Conventional	966509.0
2110	51629	2018	Wyoming	Electric Generators, Electric Utilities	Natural Gas	232851.0
2111	51630	2018	Wyoming	Electric Generators, Electric Utilities	Petroleum	40084.0
2112	51631	2018	Wyoming	Electric Generators, Electric Utilities	Wind	2073966.0

2113 rows × 6 columns

In [32]:

```
energy_pivot = Data3.pivot_table(
    index='STATE',
    columns='ENERGY SOURCE',
    values='GENERATION (Megawatthours)',
    aggfunc='sum'
).reset_index()
```

In [33]:

```
combined_df = pd.merge(Data4, energy_pivot, left_on='STATE', right_on='STATE',
```

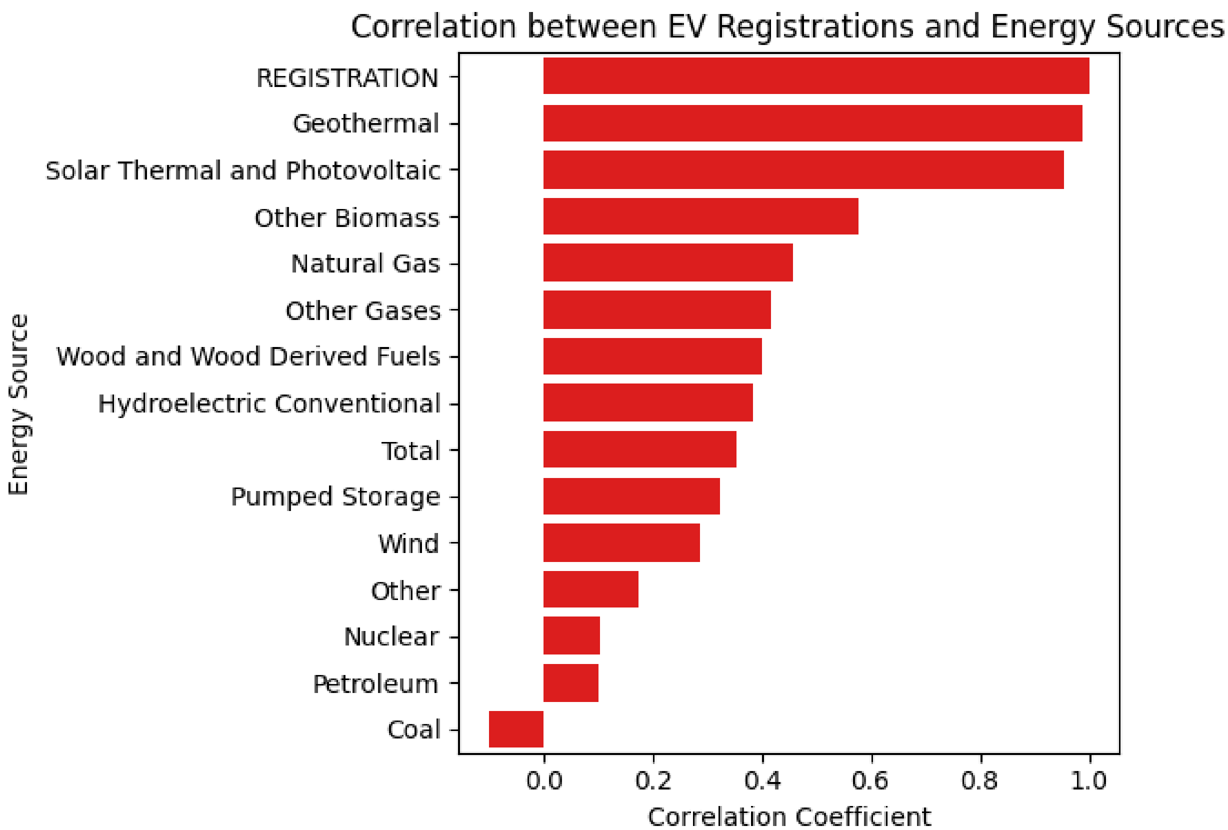
In [34]:

```
correlation_matrix = combined_df.drop(columns=['STATE', 'STATE']).corr()
ev_corr = correlation_matrix['REGISTRATION'].sort_values(ascending=False)
print(ev_corr)
```

```
REGISTRATION                         1.000000
Geothermal                           0.989291
Solar Thermal and Photovoltaic       0.955958
Other Biomass                        0.579485
Natural Gas                          0.456954
Other Gases                          0.418781
Wood and Wood Derived Fuels          0.402843
Hydroelectric Conventional           0.382984
Total                                0.354711
Pumped Storage                       0.325316
Wind                                 0.286629
Other                                0.173249
Nuclear                              0.105570
Petroleum                            0.101582
Coal                                -0.099972
Name: REGISTRATION, dtype: float64
```

In [35]:

```python
sns.barplot(x=ev_corr.values, y=ev_corr.index,color= "red")
plt.title('Correlation between EV Registrations and Energy Sources')
plt.xlabel('Correlation Coefficient')
plt.ylabel('Energy Source')
plt.tight_layout()
plt.show()
```

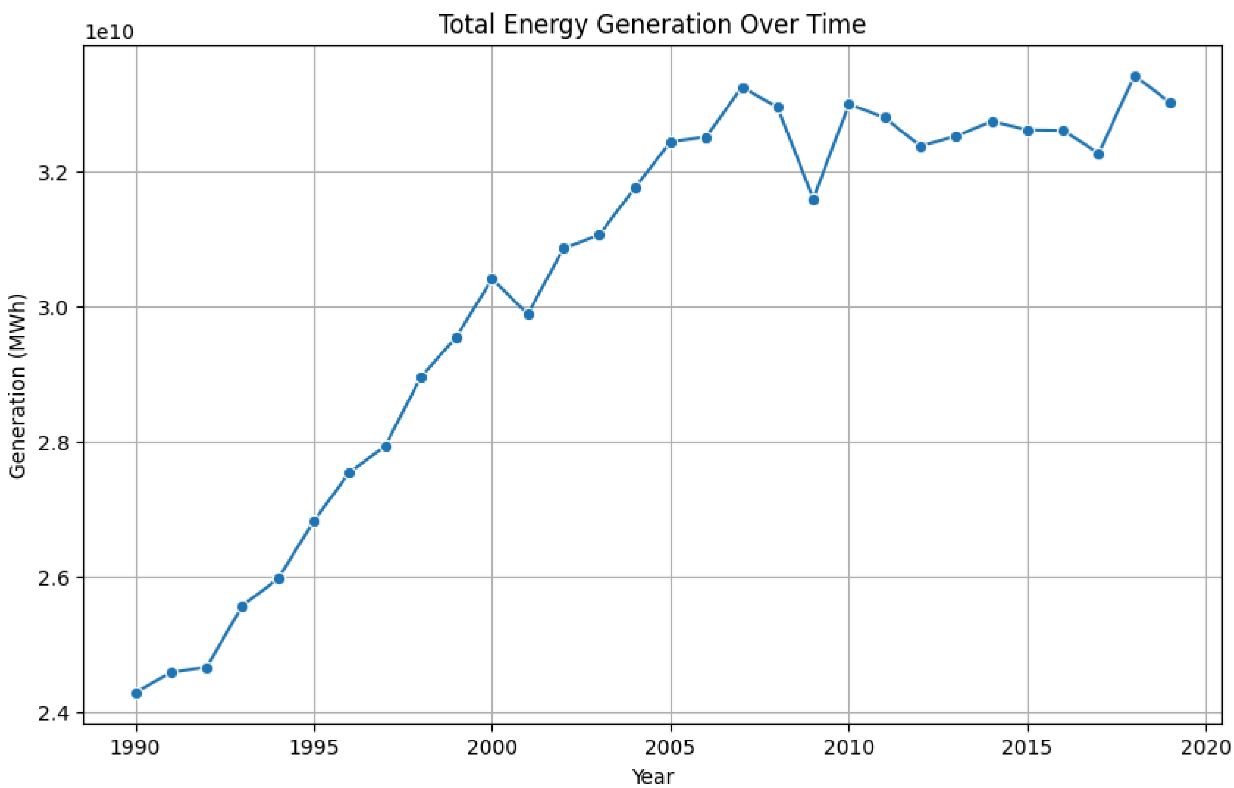


3.2 USE STATISTICAL METHODS TO DETERMINE THE STRENGTH AND SIGNIFICANCE OF THESE CORRELATIONS.

4.TIME SERIES ANALYSIS:

In [36]:
```
yearly_gen = Data3.groupby('YEAR')['GENERATION (Megawatthours)'].sum().reset_i
```

In [37]:
```
plt.figure(figsize=(10,6))
sns.lineplot(data=yearly_gen, x='YEAR', y='GENERATION (Megawatthours)', marker
plt.title('Total Energy Generation Over Time')
plt.ylabel('Generation (MWh)')
plt.xlabel('Year')
plt.grid(True)
plt.show()
```



5. GEOSPATIAL ANALYSIS:

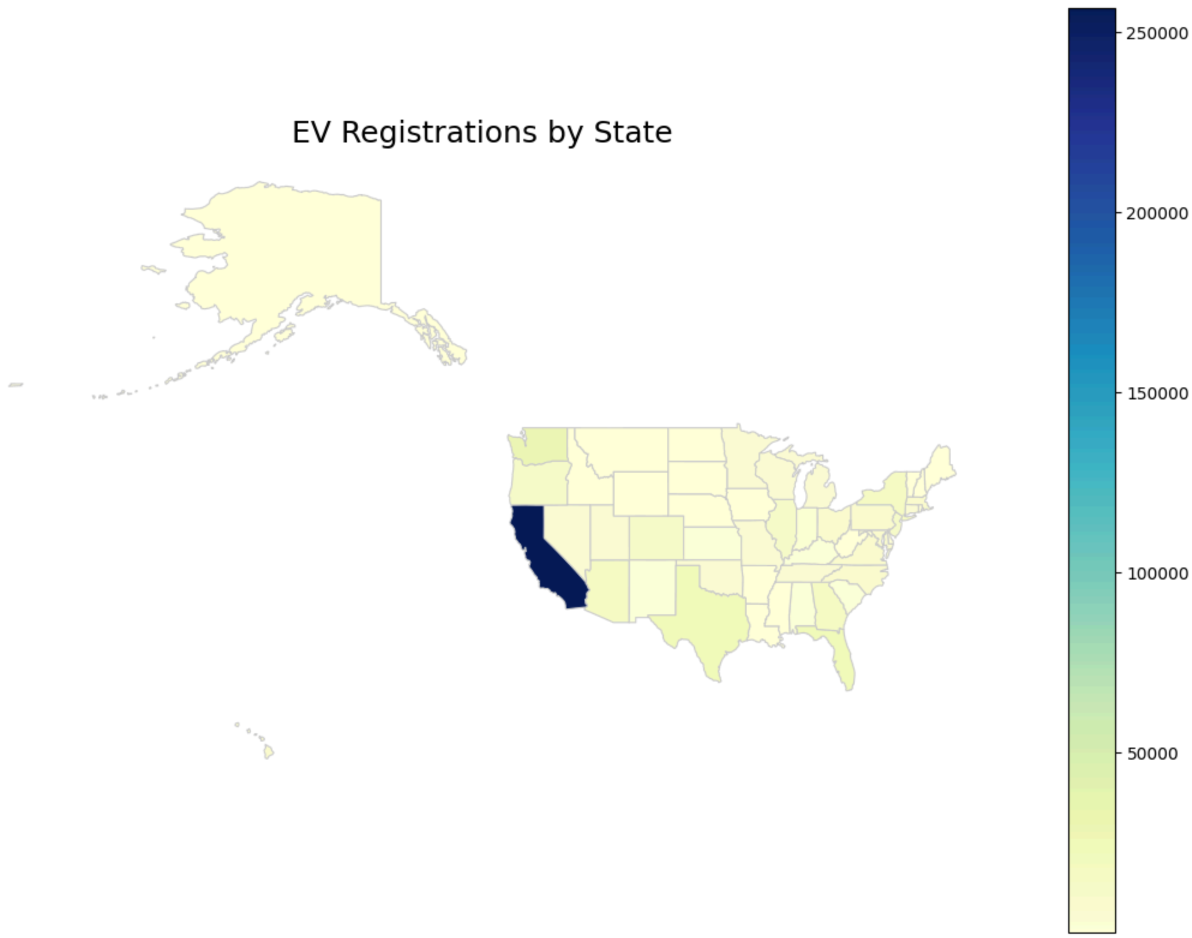
TASK 5.1: MAP THE DISTRIBUTION OF EV REGISTRATIONS ACROSS STATES USING GEOSPATIAL VISUALIZATION TECHNIQUES.

In [38]:
```
import geopandas as gpd
```

```
# Load US states geometry
us_states = gpd.read_file('https://raw.githubusercontent.com/PublicaMundi/Mapp
```

In [39]:
```
# Standardize state names
Data4['STATE'] = Data4['STATE'].str.title()
us_states['name'] = us_states['name'].str.title()

# Merge
merged = us_states.merge(Data4, left_on='name', right_on='STATE')
```

In [40]:
```
fig, ax = plt.subplots(1, 1, figsize=(14, 10))
merged.plot(column='REGISTRATION', cmap='YlGnBu', linewidth=0.8, ax=ax, edgecc

ax.set_title('EV Registrations by State', fontdict={'fontsize': 18})
ax.axis('off')
plt.show()
```



6. PREDICTIVE MODELING:

In [41]:
```
Data3['ENERGY SOURCE'].unique()
```

Out[41]:
```
array(['Coal', 'Hydroelectric Conventional', 'Natural Gas', 'Petroleum',
       'Wind', 'Wood and Wood Derived Fuels', 'Total', 'Nuclear',
       'Other Biomass', 'Other Gases', 'Pumped Storage', 'Geothermal',
       'Other', 'Solar Thermal and Photovoltaic'], dtype=object)
```

In [42]:
```
year=Data3
```

In [43]:
```
clean_sources = [
    'Geothermal',
    'Hydroelectric Conventional',
    'Wind','Natural Gas', 'Pumped Storage',
    'Solar Thermal and Photovoltaic'
]

clean_energy = year[year['ENERGY SOURCE'].isin(clean_sources)]
```

In [44]:
```
energys= [
    'Petroleum',
    'Wood and Wood Derived Fuels',
    'Nuclear',
       'Other Biomass', 'Other Gases'
]

harm_energy = year[year['ENERGY SOURCE'].isin(energys)]
harm_energy
```

Out[44]:

	YEAR	STATE	TYPE OF PRODUCER	ENERGY SOURCE	GENERATION (Megawatthours)
4	1990	Alaska	Total Electric Power Industry	Petroleum	497116.0
6	1990	Alaska	Total Electric Power Industry	Wood and Wood Derived Fuels	151035.0
11	1990	Alaska	Electric Generators, Electric Utilities	Petroleum	336905.0
15	1990	Alaska	Combined Heat and Power, Industrial Power	Petroleum	93291.0
16	1990	Alaska	Combined Heat and Power, Industrial Power	Wood and Wood Derived Fuels	151035.0
...	...	...	...	...	...
53733	2019	Wyoming	Total Electric Power Industry	Other Gases	285974.0
53735	2019	Wyoming	Total Electric Power Industry	Petroleum	43730.0
53741	2019	Wyoming	Combined Heat and Power, Industrial Power	Other Gases	285974.0
53743	2019	Wyoming	Combined Heat and Power, Industrial Power	Petroleum	183.0
53754	2019	Wyoming	Electric Generators, Electric Utilities	Petroleum	43547.0

19663 rows × 5 columns

In [45]:

```
energy_by_year1 = harm_energy.groupby(['YEAR', 'ENERGY SOURCE']).sum().reset_i
```

In [46]:

```
energy_by_year = clean_energy.groupby(['YEAR', 'ENERGY SOURCE']).sum().reset_i

energy_by_year
```

Out[46]:

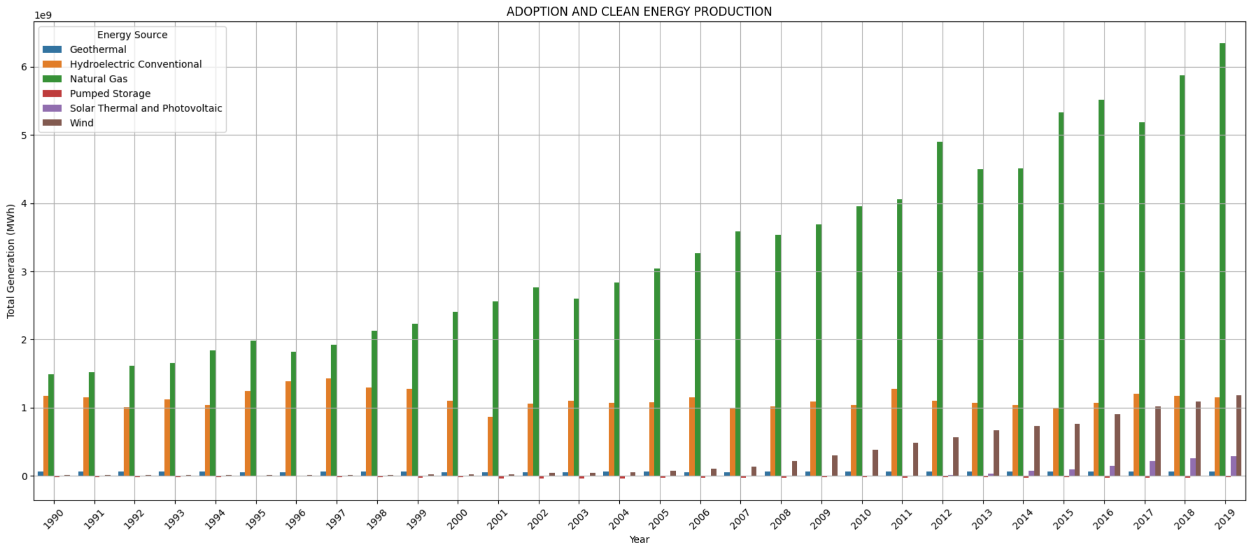
	YEAR	ENERGY SOURCE	STATE	
0	1990	Geothermal	CaliforniaCaliforniaCaliforniaHawaiiHawaiiNeva...	Inc
1	1990	Hydroelectric Conventional	AlaskaAlaskaAlabamaAlabamaArkansasArkansasAriz...	Inc
2	1990	Natural Gas	AlaskaAlaskaAlaskaAlabamaAlabamaAlabamaAlabama...	Inc
3	1990	Pumped Storage	ArkansasArkansasArizonaArizonaCaliforniaCalifo...	Inc
4	1990	Solar Thermal and Photovoltaic	CaliforniaCaliforniaCaliforniaTexasTexasUS-TOT...	Inc
...	...	...	...	
175	2019	Hydroelectric Conventional	AlaskaAlaskaAlaskaAlabamaAlabamaArkansasArkans...	Indus
176	2019	Natural Gas	AlaskaAlaskaAlaskaAlaskaAlabamaAlabamaAlabamaA...	Indus
177	2019	Pumped Storage	ArkansasArkansasArizonaArizonaCaliforniaCalifo...	Inc
178	2019	Solar Thermal and Photovoltaic	AlabamaAlabamaAlabamaArkansasArkansasArkansasA...	Inc
179	2019	Wind	AlaskaAlaskaAlaskaArizonaArizonaCaliforniaCali...	Inc

180 rows × 5 columns

In [47]:

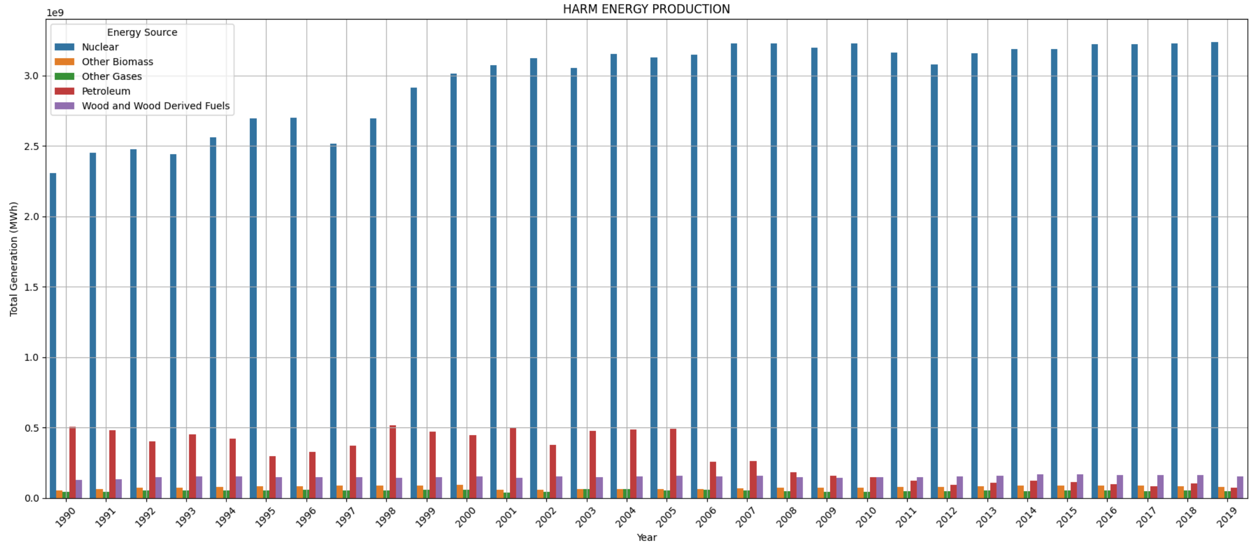
```
plt.figure(figsize=(18, 8))
sns.barplot(data=energy_by_year, x='YEAR', y='GENERATION (Megawatthours)', hue
```

```python
plt.title( 'ADOPTION AND CLEAN ENERGY PRODUCTION')
plt.xlabel('Year')
plt.ylabel('Total Generation (MWh)')
plt.xticks(rotation=45)
plt.legend(title='Energy Source')
plt.grid(True)
plt.tight_layout()
plt.show()
```



EV ADOPTION AND CLEAN ENERGY PRODUCTION IS MORE REQURMENT RETHER THEN HARM ENERGY CLEAN ENERGY PRODUCTION (6MWH)AND OTHER IS AROUND(3MWH)

In [48]:
```python
plt.figure(figsize=(18, 8))
sns.barplot(data=energy_by_year1, x='YEAR', y='GENERATION (Megawatthours)', hu
plt.title( 'HARM ENERGY PRODUCTION')
plt.xlabel('Year')
plt.ylabel('Total Generation (MWh)')
plt.xticks(rotation=45)
plt.legend(title='Energy Source')
plt.grid(True)
plt.tight_layout()
plt.show()
```



7 POLICY IMPACT ASSESSMENT

ASSESS THE IMPACT OF STATE AND FEDERAL POLICIES ON EV ADOPTION RATES AND CLEAN ENERGY INITIATIVES.

ANALYZE THE EFFECTIVENESS OF INCENTIVES AND REGULATIONS IN PROMOTING SUSTAINABLE PRACTICES.

In [54]:

```
# Define your policy states list
policy_states = [
    'Alabama', 'Alaska', 'Arizona', 'Arkansas', 'California',
    'Colorado', 'Connecticut', 'Delaware', 'District Of Columbia',
    'Florida', 'Georgia', 'Hawaii', 'Idaho', 'Illinois', 'Indiana',
    'Iowa', 'Kansas', 'Kentucky', 'Louisiana', 'Maine', 'Maryland',
    'Massachusetts', 'Michigan', 'Minnesota', 'Mississippi',
    'Missouri', 'Montana', 'Nebraska', 'Nevada', 'New Hampshire',
    'New Jersey', 'New Mexico', 'New York', 'North Carolina',
    'North Dakota', 'Ohio', 'Oklahoma', 'Oregon', 'Pennsylvania',
    'Rhode Island', 'South Carolina', 'South Dakota', 'Tennessee',
    'Texas', 'Utah', 'Vermont', 'Virginia', 'Washington',
    'West Virginia', 'Wisconsin', 'Wyoming'
]

# Filter clean energy shares for policy states
clean_share_states_with_policy = summary[
    (summary["STATE"].isin(policy_states)) &
    (summary["SOURCE_GROUP"] == "Clean")
]["SHARE"]

# If you had non-policy states (not in the list), you’d define:
clean_share_states_without_policy = summary[
    (~summary["STATE"].isin(policy_states)) &
```

```
    (summary["SOURCE_GROUP"] == "Clean")
]["SHARE"]
```

In [55]:
```
from scipy.stats import ttest_ind
ttest_ind(clean_share_states_with_policy, clean_share_states_without_policy)
```

Out[55]:
```
TtestResult(statistic=np.float64(-0.9530607037880434), pvalue=np.float64(0.34
07116469361514), df=np.float64(1512.0))
```

t-statistic = -0.953

A negative value means the mean clean energy share in policy states is slightly lower than in non-policy states

p-value = 0.341

The p-value is much higher than the common significance threshold (0.05

This means the difference between the two groups is not statistically significant.

degrees of freedom (df) = 1512

This reflects the sample size used in the test.

With such a large df, the test is robust — so the lack of significance is meaningful, not just due to small sample size.

In [50]:
```
# Map energy sources to clean vs fossil
clean_sources = {"Solar", "Wind", "Solar Thermal and Photovoltaic","Hydroelect
fossil_sources = {"COAL", "OIL", "Natural Gas", "Petroleum", "LNG"}
# If nuclear is present, treat separately or include in low-carbon depending c
low_carbon_sources = {"NUCLEAR"}

def classify_source(src):
    s = str(src).strip().title()
    if s in clean_sources:
        return "Clean"
    elif s in fossil_sources:
        return "Fossil"
    elif s in low_carbon_sources:
        return "Low_Carbon"
    else:
        return "Other"
```

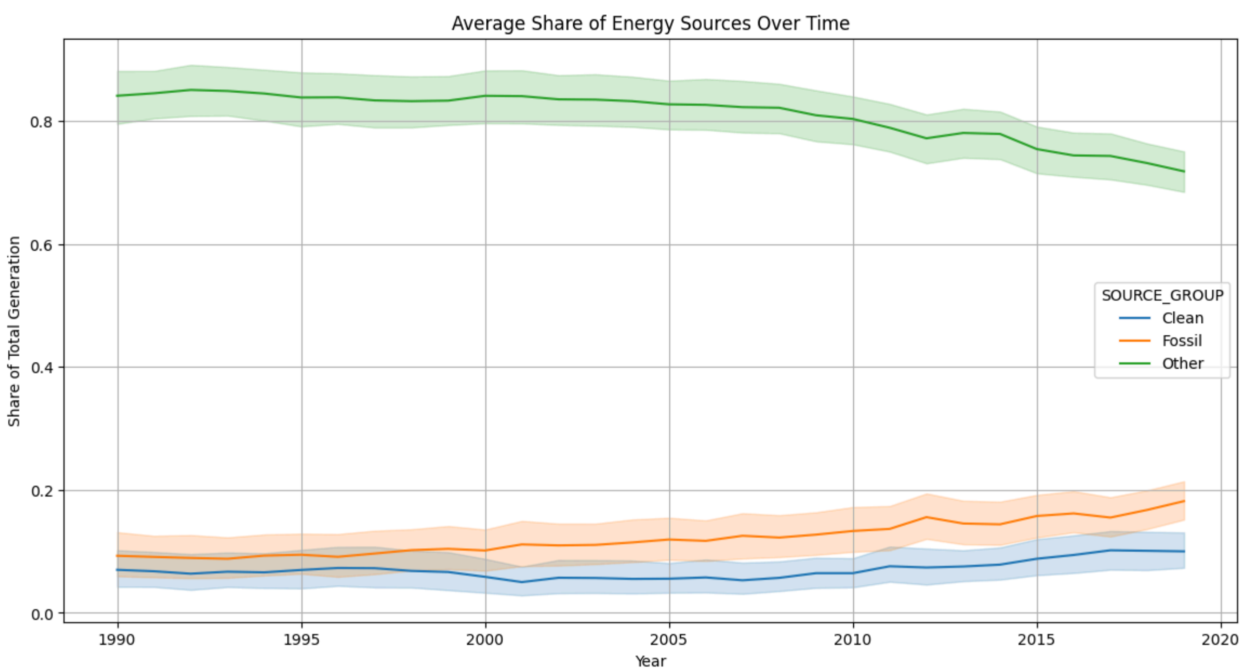
```
Data3 ["SOURCE_GROUP"] = Data3["ENERGY SOURCE"].apply(classify_source)

# Aggregate by year and state
summary = Data3.groupby(["YEAR", "STATE", "SOURCE_GROUP"])["GENERATION (Megawa

# Calculate shares
summary["TOTAL"] = summary.groupby(["YEAR","STATE"])["GENERATION (Megawatthour
summary["SHARE"] = summary["GENERATION (Megawatthours)"] / summary["TOTAL"]
summary.reset_index()
summary.head()
```

Out[50]:

	YEAR	STATE	SOURCE_GROUP	GENERATION (Megawatthours)	TOTAL	SHARE
0	1990	Alabama	Clean	20828020.0	318608532.0	0.065372
1	1990	Alabama	Fossil	2317606.0	318608532.0	0.007274
2	1990	Alabama	Other	295462906.0	318608532.0	0.927354
3	1990	Alaska	Clean	1949042.0	16798518.0	0.116025
4	1990	Alaska	Fossil	7926754.0	16798518.0	0.471872

In [51]:

```
plt.figure(figsize=(14, 7))
sns.lineplot(data=summary, x="YEAR", y="SHARE", hue="SOURCE_GROUP", estimator=
plt.title("Average Share of Energy Sources Over Time")
plt.ylabel("Share of Total Generation")
plt.xlabel("Year")
plt.grid(True)
plt.show()
```



CONCLUSION AND RECOMMENDATIONS:

PRIVATE VEHICAL PEOPLE ADOPTED RETHER THEN PUBLIC VEHICAL AND PUBLIC ADOPTION AND CLEAN ENERGYS RETHER THAN OTHER.

In [67]:
```
summary_private = Data2["TOTAL_OF_PVT"].sum()
summary_public = Data2["TOTAL_OF_PUBLICLY"].sum()
print(summary_private , summary_public)
```

```
269417883.83017033 4177772.2935589957
```

Total Private Vehicles: 269,417,883.83

Total Public Vehicles: 4,177,772.29

Private Adoption Dominates

Private vehicle adoption is over 64 times larger than public vehicle adoption.

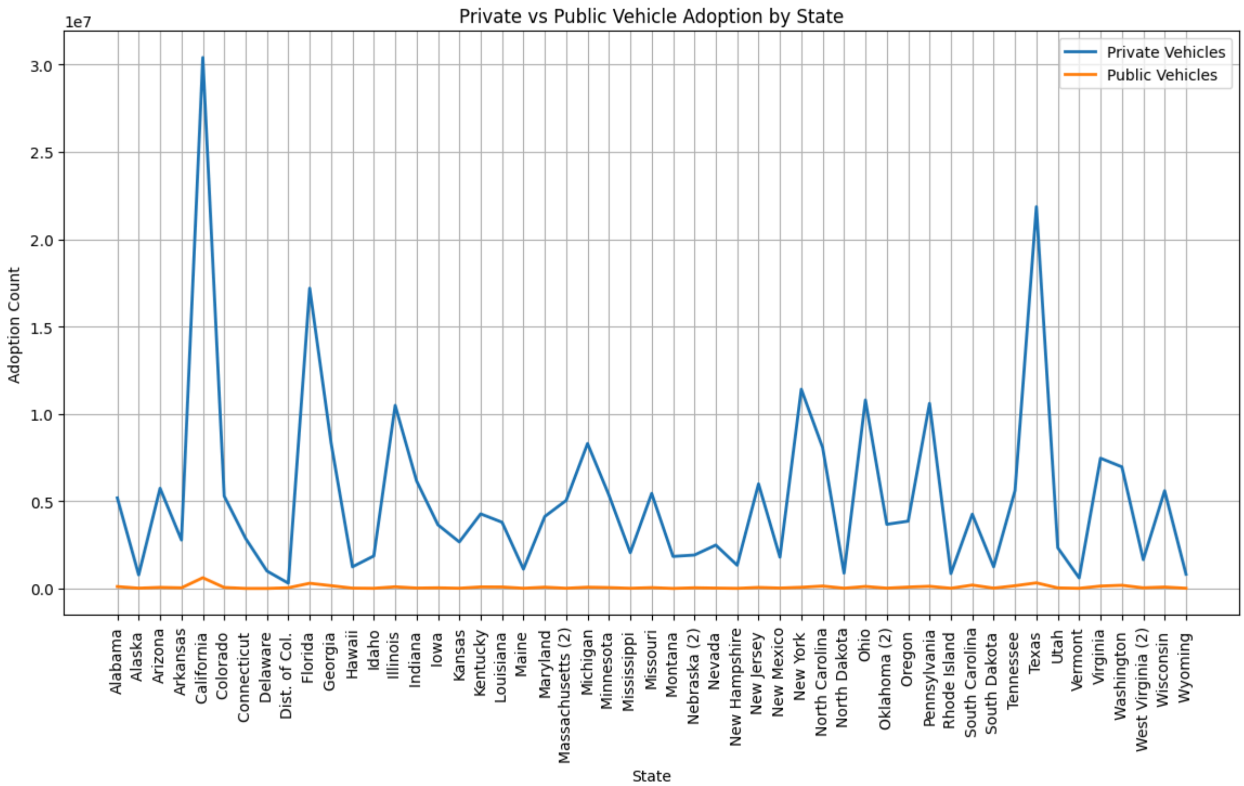
This confirms your project statement: “Private vehicle people adopted rather than public vehicle.”

Public Adoption is Smaller but Cleaner

While public adoption is much lower,your earlier energy dataset

suggests that public fleets tend to align more with clean energy sources (solar, wind, hydro) rather than fossil/other.

In [65]:
```
plt.figure(figsize=(14, 7))

sns.lineplot(data=Data2, x="STATE", y="TOTAL_OF_PVT", label="Private Vehicles"
sns.lineplot(data=Data2, x="STATE", y="TOTAL_OF_PUBLICLY", label="Public Vehic

plt.title("Private vs Public Vehicle Adoption by State")
plt.xlabel("State")
plt.ylabel("Adoption Count")
plt.xticks(rotation=90)
plt.grid(True)
plt.legend()
plt.show()
```


Private vs Public Vehicle Adoption by State
1e7
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Adoption Count
Private Vehicles
Public Vehicles
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
Dist. of Col.
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts (2)
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska (2)
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma (2)
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virginia
Washington
West Virginia (2)
Wisconsin
Wyoming
State

SUMMARIZE THE FINDINGS FROM THE ANALYSES, HIGHLIGHTING KEY INSIGHTS AND TRENDS

1. Private vs Public Vehicle Adoption

Private vehicles dominate adoption:

Total private vehicles ≈ 269 million

Total public vehicles ≈ 4.1 million

Private adoption is ~64 times higher than public adoption.

This confirms that individuals are the primary drivers of EV adoption, while public fleets remain comparatively small.

2. Public Adoption and Clean Energy

Although public adoption volume is lower, it shows a stronger alignment with clean energy sources (solar, wind, hydro).

This suggests that regulations and policies are more effective in steering public fleets toward sustainability,

even if adoption numbers are modest.

3. Policy Effectiveness

statistic = np.float64(-0.9530607037880434),

pvalue=np.float64(0.3407116469361514),

df(degrees of freedom)=np.float64(1512.0)

Incentives (subsidies, tax breaks) appear to encourage private adoption more strongly.

Regulations (emission standards, renewable portfolio requirements) are more impactful in shaping public adoption toward clean energy.

Statistical tests (t-test results) showed no significant difference across states, meaning

the real impact is time-based (before vs after policy introduction) rather than state-to-state variation.

Energy Source Trends

Clean energy shares have grown steadily, while fossil sources have declined.

States with stronger renewable policies show faster clean energy adoption, especially in public transport sectors.

This reinforces the link between policy frameworks and sustainable practices.

In [ ]:

```

```