```
import pandas as pd
df=pd.read_csv("files/dataset (1).csv")
```

Model Year Make: Model: Electric Vehicle Type: Electric Range: Base MSRP: County, City, State, Postal Code: Clean Alternative Fuel Vehicle (CAFV) Eligibility: Legislative District:

```
df.head()

   VIN (1-10)     County      City State  Postal Code  Model Year
Make  \
0  JTMEB3FV6N     Monroe  Key West    FL        33040        2022
TOYOTA
1  1G1RD6E45D      Clark  Laughlin    NV        89029        2013
CHEVROLET
2  JN1AZ0CP8B     Yakima    Yakima    WA        98901        2011
NISSAN
3  1G1FW6S08H     Skagit  Concrete    WA        98237        2017
CHEVROLET
4  3FA6P0SU1K  Snohomish   Everett    WA        98201        2019
FORD

        Model                   Electric Vehicle Type  \
0  RAV4 PRIME  Plug-in Hybrid Electric Vehicle (PHEV)
1        VOLT  Plug-in Hybrid Electric Vehicle (PHEV)
2        LEAF          Battery Electric Vehicle (BEV)
3     BOLT EV          Battery Electric Vehicle (BEV)
4      FUSION  Plug-in Hybrid Electric Vehicle (PHEV)

  Clean Alternative Fuel Vehicle (CAFV) Eligibility  Electric Range  \
0           Clean Alternative Fuel Vehicle Eligible              42
1           Clean Alternative Fuel Vehicle Eligible              38
2           Clean Alternative Fuel Vehicle Eligible              73
3           Clean Alternative Fuel Vehicle Eligible             238
4             Not eligible due to low battery range              26

   Base MSRP  Legislative District  DOL Vehicle ID  \
0          0                   NaN       198968248
1          0                   NaN         5204412
2          0                  15.0       218972519
3          0                  39.0       186750406
4          0                  38.0         2006714

              Vehicle Location        Electric Utility  2020 Census
Tract
0    POINT (-81.80023 24.5545)                     NaN
12087972100
1  POINT (-114.57245 35.16815)                     NaN
32003005702
2  POINT (-120.50721 46.60448)              PACIFICORP
53077001602
```

```
3   POINT (-121.7515 48.53892)  PUGET SOUND ENERGY INC
53057951101
4  POINT (-122.20596 47.97659)  PUGET SOUND ENERGY INC
53061041500

df.info

<bound method DataFrame.info of         VIN (1-10)     County
City State  Postal Code  Model Year  \
0       JTMEB3FV6N     Monroe       Key West    FL        33040
2022
1       1G1RD6E45D      Clark       Laughlin    NV        89029
2013
2       JN1AZ0CP8B     Yakima         Yakima    WA        98901
2011
3       1G1FW6S08H     Skagit       Concrete    WA        98237
2017
4       3FA6P0SU1K  Snohomish        Everett    WA        98201
2019
...            ...        ...            ...   ...          ...
...
112629  7SAYGDEF2N       King         Duvall    WA        98019
2022
112630  1N4BZ1CP7K   San Juan  Friday Harbor    WA        98250
2019
112631  1FMCU0KZ4N       King         Vashon    WA        98070
2022
112632  KNDCD3LD4J       King      Covington    WA        98042
2018
112633  YV4BR0CL8N       King      Covington    WA        98042
2022

             Make       Model                   Electric Vehicle Type
\
0          TOYOTA  RAV4 PRIME  Plug-in Hybrid Electric Vehicle (PHEV)

1       CHEVROLET        VOLT  Plug-in Hybrid Electric Vehicle (PHEV)

2          NISSAN        LEAF          Battery Electric Vehicle (BEV)

3       CHEVROLET     BOLT EV          Battery Electric Vehicle (BEV)

4            FORD      FUSION  Plug-in Hybrid Electric Vehicle (PHEV)

...           ...         ...                                     ...

112629      TESLA     MODEL Y          Battery Electric Vehicle (BEV)

112630     NISSAN        LEAF          Battery Electric Vehicle (BEV)

112631       FORD      ESCAPE  Plug-in Hybrid Electric Vehicle (PHEV)
```

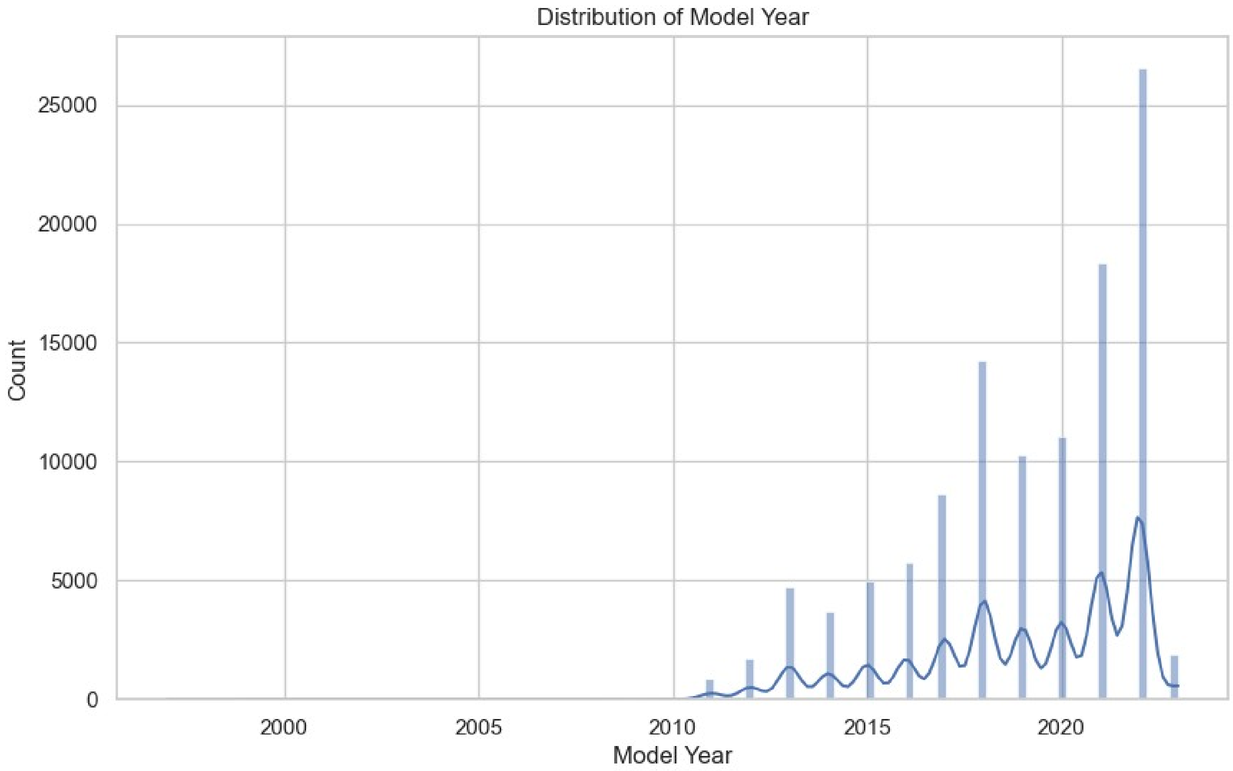
```
112632        KIA        NIRO  Plug-in Hybrid Electric Vehicle (PHEV)
112633      VOLVO        XC90  Plug-in Hybrid Electric Vehicle (PHEV)

                                         CAFV Eligibility  Electric
Range  \
0                 Clean Alternative Fuel Vehicle Eligible
42
1                 Clean Alternative Fuel Vehicle Eligible
38
2                 Clean Alternative Fuel Vehicle Eligible
73
3                 Clean Alternative Fuel Vehicle Eligible
238
4                   Not eligible due to low battery range
26
...                                                   ...
...
112629  Eligibility unknown as battery range has not b...
0
112630            Clean Alternative Fuel Vehicle Eligible
150
112631            Clean Alternative Fuel Vehicle Eligible
38
112632              Not eligible due to low battery range
26
112633              Not eligible due to low battery range
18

        Base MSRP  Legislative District  DOL Vehicle ID  \
0               0                   NaN       198968248
1               0                   NaN         5204412
2               0                  15.0       218972519
3               0                  39.0       186750406
4               0                  38.0         2006714
...           ...                   ...             ...
112629          0                  45.0       217955265
112630          0                  40.0       103663227
112631          0                  34.0       193878387
112632          0                  47.0       125039043
112633          0                  47.0       194673692

                   Vehicle Location  \
0         POINT (-81.80023 24.5545)
1       POINT (-114.57245 35.16815)
2       POINT (-120.50721 46.60448)
3        POINT (-121.7515 48.53892)
4       POINT (-122.20596 47.97659)
```

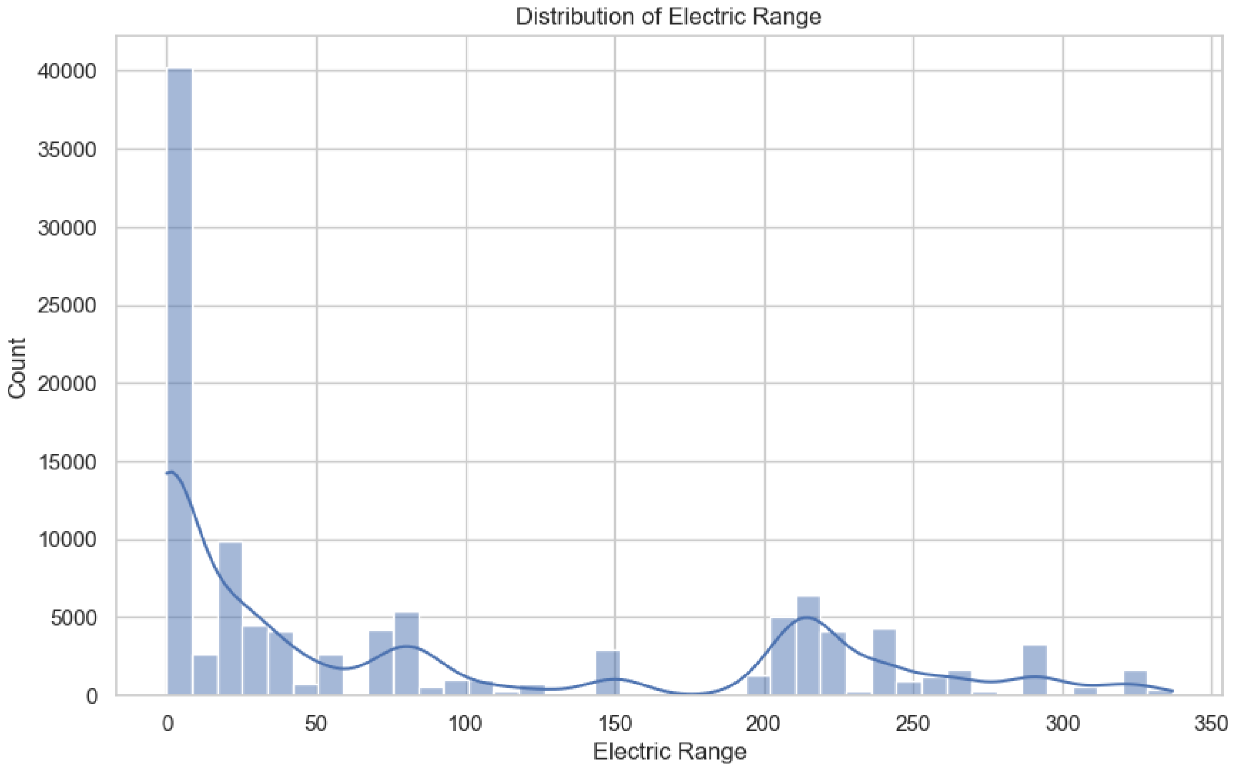
```
...                                ...
112629  POINT (-121.98609 47.74068)
112630  POINT (-123.01648 48.53448)
112631   POINT (-122.4573 47.44929)
112632  POINT (-122.09124 47.33778)
112633  POINT (-122.09124 47.33778)

                                         Electric Utility  2020 Census
Tract
0                                                     NaN
12087972100
1                                                     NaN
32003005702
2                                              PACIFICORP
53077001602
3                                    PUGET SOUND ENERGY INC
53057951101
4                                    PUGET SOUND ENERGY INC
53061041500
...                                                   ...
...
112629      PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)
53033032401
112630  BONNEVILLE POWER ADMINISTRATION||ORCAS POWER &...
53055960301
112631      PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)
53033027702
112632      PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)
53033032007
112633      PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)
53033032005

[112634 rows x 17 columns]>
```

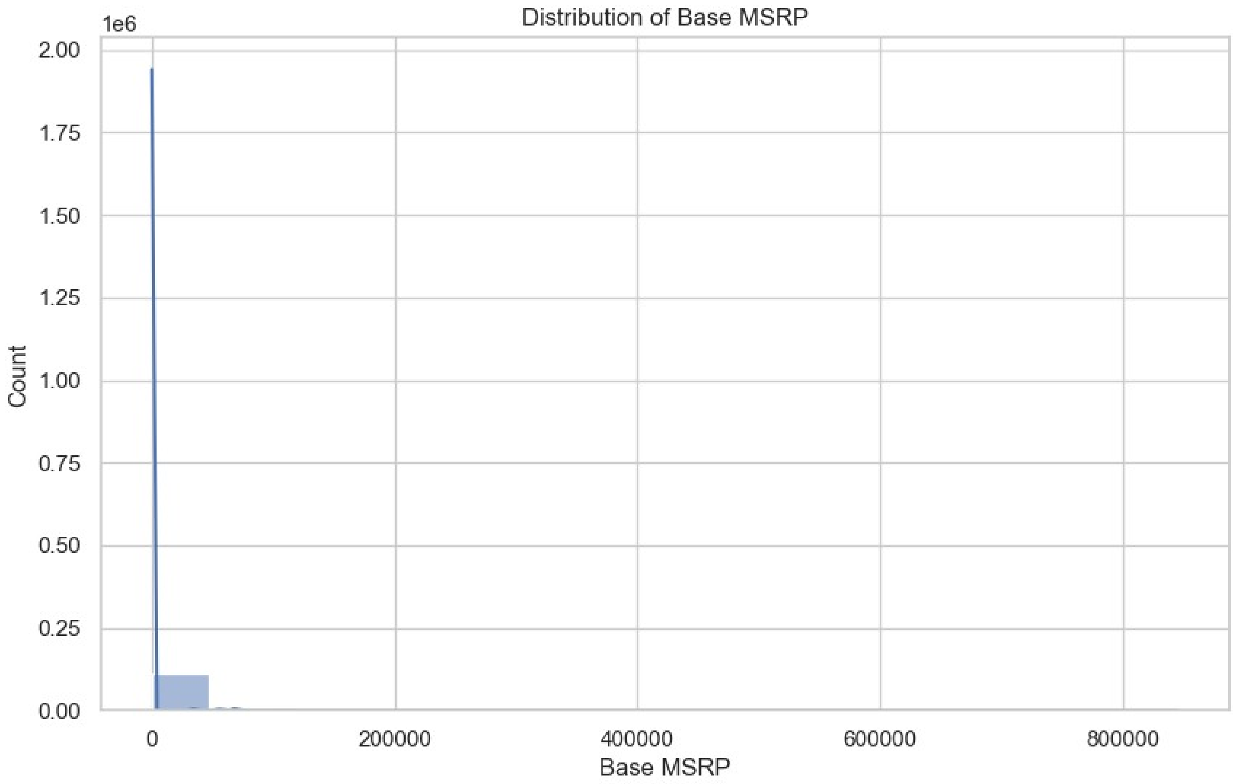
```python
# Assuming df is your DataFrame
df.rename(columns={'Clean Alternative Fuel Vehicle (CAFV)
Eligibility': 'CAFV Eligibility'}, inplace=True)

# Verify the renaming
print(df.columns)
```

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

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


# Clean the data
df.drop_duplicates(inplace=True)
df.dropna(subset=['Electric Range', 'Base MSRP'], inplace=True)
df['Electric Range'] = pd.to_numeric(df['Electric Range'],
errors='coerce')
df['Base MSRP'] = pd.to_numeric(df['Base MSRP'], errors='coerce')

# Optionally handle outliers
df = df[(df['Electric Range'] >= 0) & (df['Base MSRP'] >= 0)]

# Univariate analysis
sns.set(style='whitegrid')

plt.figure(figsize=(10, 6))
sns.histplot(df['Model Year'], kde=True)
plt.title('Distribution of Model Year')
plt.show()

plt.figure(figsize=(10, 6))
sns.histplot(df['Electric Range'], kde=True)
plt.title('Distribution of Electric Range')
plt.show()

plt.figure(figsize=(10, 6))
sns.histplot(df['Base MSRP'], kde=True)
plt.title('Distribution of Base MSRP')
plt.show()
```


Distribution of Model Year
Count
25000
20000
15000
10000
5000
0
2000
2005
2010
2015
2020
Model Year


Distribution of Electric Range
Count
40000
35000
30000
25000
20000
15000
10000
5000
0
0
50
100
150
200
250
300
350
Electric Range



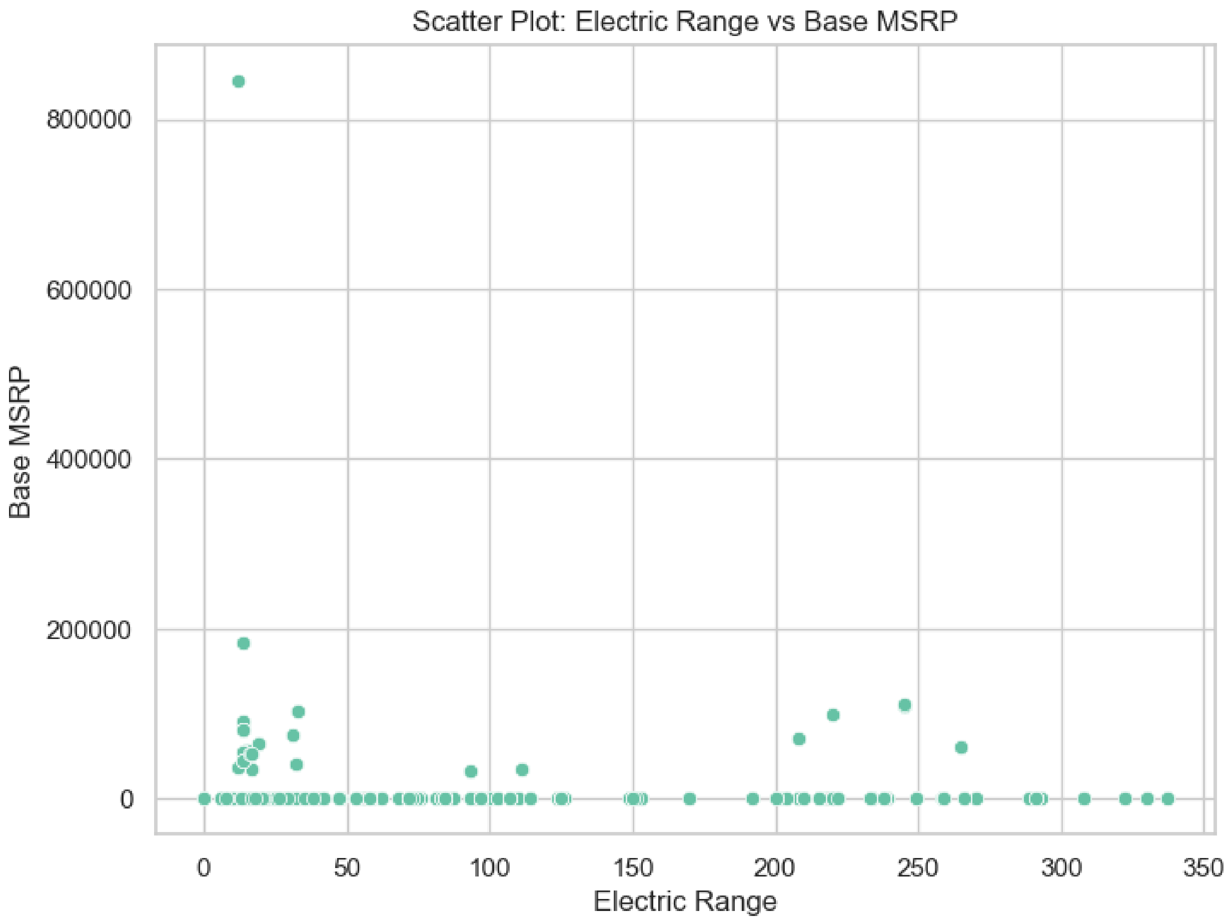
bivariate analysis

```
df_cleaned = df.drop(columns=['DOL Vehicle ID', 'Vehicle Location',
'2020 Census Tract'])
# df_cleaned['Electric Utility'].fillna('Unknown', inplace=True)
df_cleaned['VIN (1-10)'] = df_cleaned['VIN (1-10)'].astype(str)


make_state_crosstab = pd.crosstab(df_cleaned['Make'],
df_cleaned['State'])

print("Cross-tabulation between 'Make' and 'State':")
print(make_state_crosstab)





correlation = df_cleaned[['Electric Range', 'Base MSRP']].corr()
print("\nCorrelation between 'Electric Range' and 'Base MSRP':")
print(correlation)

plt.figure(figsize=(8, 6))
```

```
sns.scatterplot(x='Electric Range', y='Base MSRP', data=df_cleaned)
plt.title('Scatter Plot: Electric Range vs Base MSRP')
plt.show()

mean_electric_range_by_make = df_cleaned.groupby('Make')['Electric
Range'].mean()
print("\nAverage Electric Range by Make:")
print(mean_electric_range_by_make)

plt.figure(figsize=(10, 6))
sns.boxplot(x='Make', y='Electric Range', data=df_cleaned)
plt.title('Boxplot: Electric Range by Make')
plt.xticks(rotation=90)
plt.show()

Cross-tabulation between 'Make' and 'State':
State           AK  AL  AR  AZ  CA  CO  CT  DC  DE  FL  ...  RI  SC
SD  TN  \
Make                                                    ...

AUDI             0   0   0   0   2   0   0   0   0   0  ...   0   0
0   0
AZURE DYNAMICS   0   0   0   0   0   0   0   0   0   0  ...   0   0
0   0
BENTLEY          0   0   0   0   0   0   0   0   0   0  ...   0   0
0   0
BMW              0   0   1   0   1   1   0   1   0   0  ...   0   0
0   0
CADILLAC         0   0   0   0   0   0   0   0   0   0  ...   0   0
0   0
CHEVROLET        0   0   0   0   3   2   2   0   0   0  ...   0   0
0   0
CHRYSLER         0   0   0   0   2   1   1   0   0   0  ...   0   0
0   0
FIAT             0   0   0   0   0   0   0   0   0   0  ...   0   0
0   0
FISKER           0   0   0   0   0   0   0   0   0   0  ...   0   0
0   0
FORD             1   0   0   0   7   0   1   1   0   1  ...   0   1
0   0
GENESIS          0   0   0   0   0   0   0   0   0   0  ...   0   0
0   0
HONDA            0   0   0   0   1   0   0   0   0   0  ...   0   0
0   0
HYUNDAI          0   0   0   0   1   0   0   0   0   0  ...   0   0
0   0
JAGUAR           0   0   0   0   1   0   0   0   0   0  ...   0   0
0   0
JEEP             0   0   0   0   1   0   0   0   0   0  ...   0   0
0   0
```

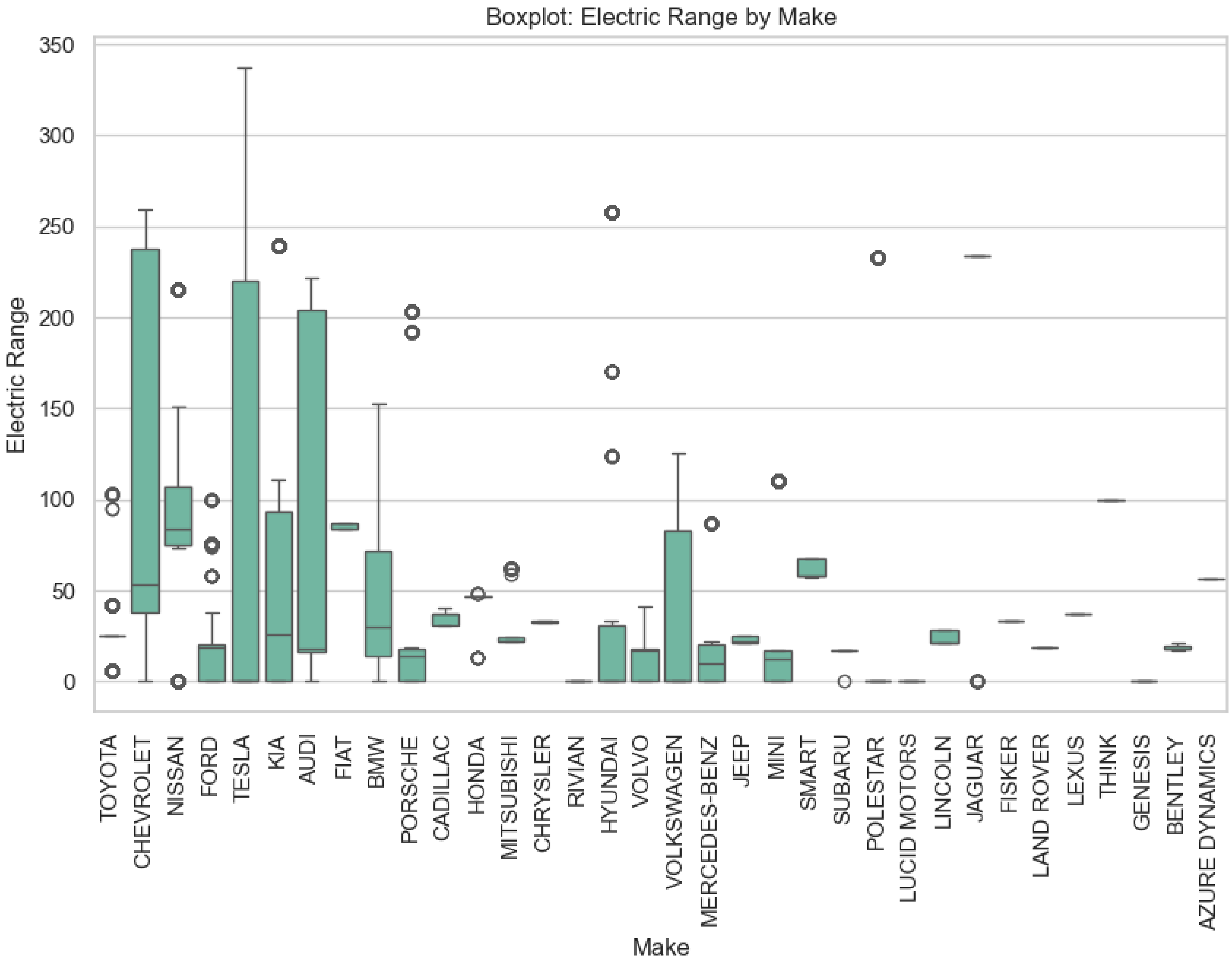
```
KIA               0   0   0   0   1   0   0   0   0   0  ...   0   0
0   0
LAND ROVER        0   0   0   0   0   0   0   0   0   0  ...   0   0
0   0
LEXUS             0   0   0   0   0   0   0   0   0   0  ...   0   0
0   0
LINCOLN           0   0   0   0   0   0   0   0   0   0  ...   0   0
0   0
LUCID MOTORS      0   0   0   0   0   0   0   0   0   0  ...   0   0
0   0
MERCEDES-BENZ     0   0   0   0   0   0   0   0   0   0  ...   0   0
0   0
MINI              0   0   0   0   0   0   0   0   0   0  ...   0   0
0   0
MITSUBISHI        0   0   0   0   1   0   0   0   0   0  ...   0   0
0   0
NISSAN            0   0   1   1   2   0   0   1   0   0  ...   1   0
0   0
POLESTAR          0   0   0   0   0   1   0   0   0   0  ...   0   0
0   0
PORSCHE           0   0   0   0   1   0   0   0   0   0  ...   0   0
0   0
RIVIAN            0   0   0   0   0   0   0   0   0   0  ...   0   0
0   0
SMART             0   0   0   0   0   0   0   0   0   0  ...   0   0
0   0
SUBARU            0   0   0   0   0   0   0   0   0   0  ...   0   0
0   0
TESLA             0   0   2   3  40   4   1   3   1   3  ...   0   3
1   3
TH!NK             0   0   0   0   0   0   0   0   0   0  ...   0   0
0   0
TOYOTA            0   1   0   1   6   0   1   0   0   2  ...   0   1
0   0
VOLKSWAGEN        0   0   0   0   2   0   0   0   0   0  ...   0   0
0   0
VOLVO             0   0   0   1   4   0   0   0   0   0  ...   0   0
0   0

State            TX  UT  VA     WA  WI  WY
Make
AUDI              0   0   0   2327   0   0
AZURE DYNAMICS    0   0   0      7   0   0
BENTLEY           0   0   0      3   0   0
BMW               0   0   3   4665   0   1
CADILLAC          0   0   0    108   0   0
CHEVROLET         0   0   4  10162   0   0
CHRYSLER          0   0   1   1781   0   0
FIAT              0   0   0    821   0   0
```

```
FISKER           0   0   0      19   0   0
FORD             4   0   2    5795   0   0
GENESIS          0   0   0      18   0   0
HONDA            0   0   0     790   0   0
HYUNDAI          0   0   0    1409   0   0
JAGUAR           0   0   0     218   0   0
JEEP             0   0   1    1146   0   0
KIA              0   1   1    4476   0   0
LAND ROVER       0   0   0      38   0   0
LEXUS            0   0   0      33   0   0
LINCOLN          0   0   0     168   0   0
LUCID MOTORS     0   0   0      65   0   0
MERCEDES-BENZ    0   0   1     504   0   0
MINI             0   0   0     632   0   0
MITSUBISHI       0   0   0     586   0   0
NISSAN           0   0   2   12866   0   0
POLESTAR         0   0   0     557   0   0
PORSCHE          0   0   0     817   0   0
RIVIAN           0   0   0     884   0   0
SMART            0   0   0     273   0   0
SUBARU           0   0   0      59   0   0
TESLA            9   1  17   51944   1   1
TH!NK            0   0   0       3   0   0
TOYOTA           1   1   3    4384   0   0
VOLKSWAGEN       0   1   1    2509   0   0
VOLVO            0   0   0    2281   0   0

[34 rows x 45 columns]

Correlation between 'Electric Range' and 'Base MSRP':
                Electric Range  Base MSRP
Electric Range        1.000000   0.085025
Base MSRP             0.085025   1.000000
```



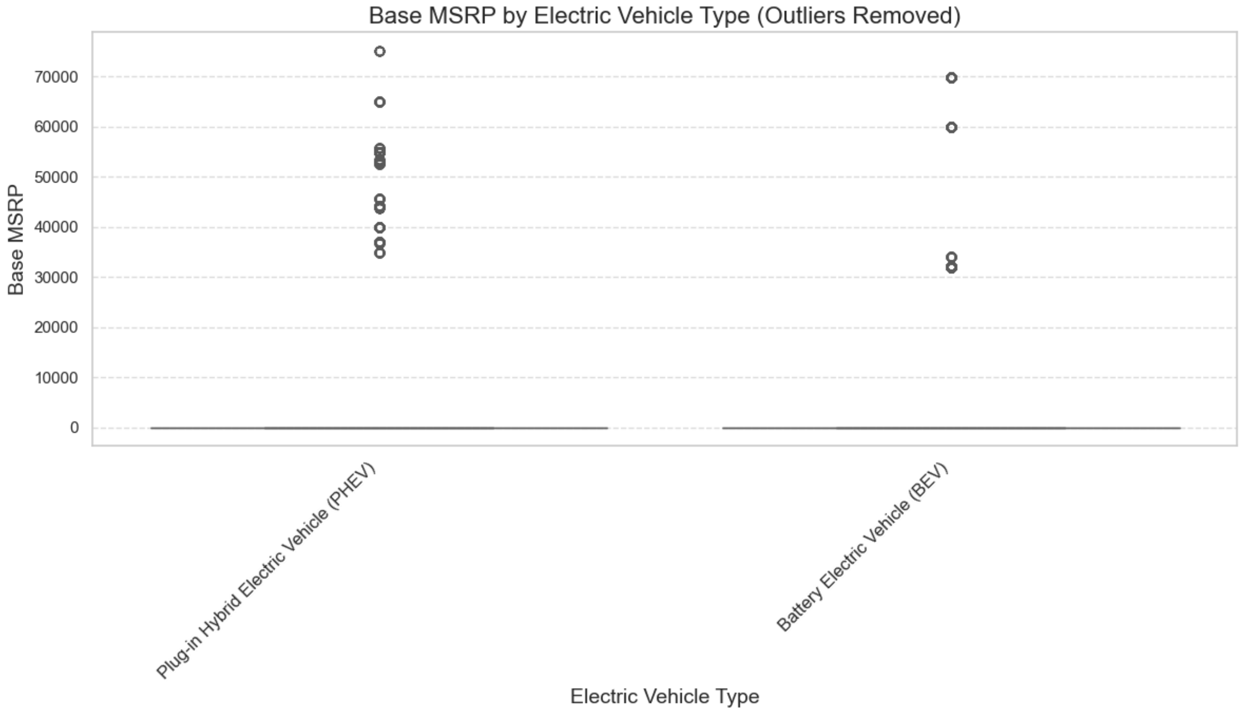
```
Average Electric Range by Make:
Make
AUDI                  62.876930
AZURE DYNAMICS        56.000000
BENTLEY               18.666667
BMW                   46.657479
CADILLAC              35.537037
CHEVROLET            109.766549
CHRYSLER              32.361204
FIAT                  85.624088
FISKER                33.000000
FORD                  16.848084
GENESIS                0.000000
HONDA                 46.618687
HYUNDAI               48.228754
JAGUAR               207.287671
JEEP                  22.707465
KIA                   67.631943
LAND ROVER            19.000000
LEXUS                 37.000000
```

```
LINCOLN            23.083333
LUCID MOTORS        0.000000
MERCEDES-BENZ      22.055336
MINI               26.604430
MITSUBISHI         26.746599
NISSAN             89.326941
POLESTAR           40.921147
PORSCHE            54.090465
RIVIAN              0.000000
SMART              62.282051
SUBARU             16.711864
TESLA             118.162756
TH!NK             100.000000
TOYOTA             26.044268
VOLKSWAGEN         43.762530
VOLVO              14.448864
Name: Electric Range, dtype: float64
```

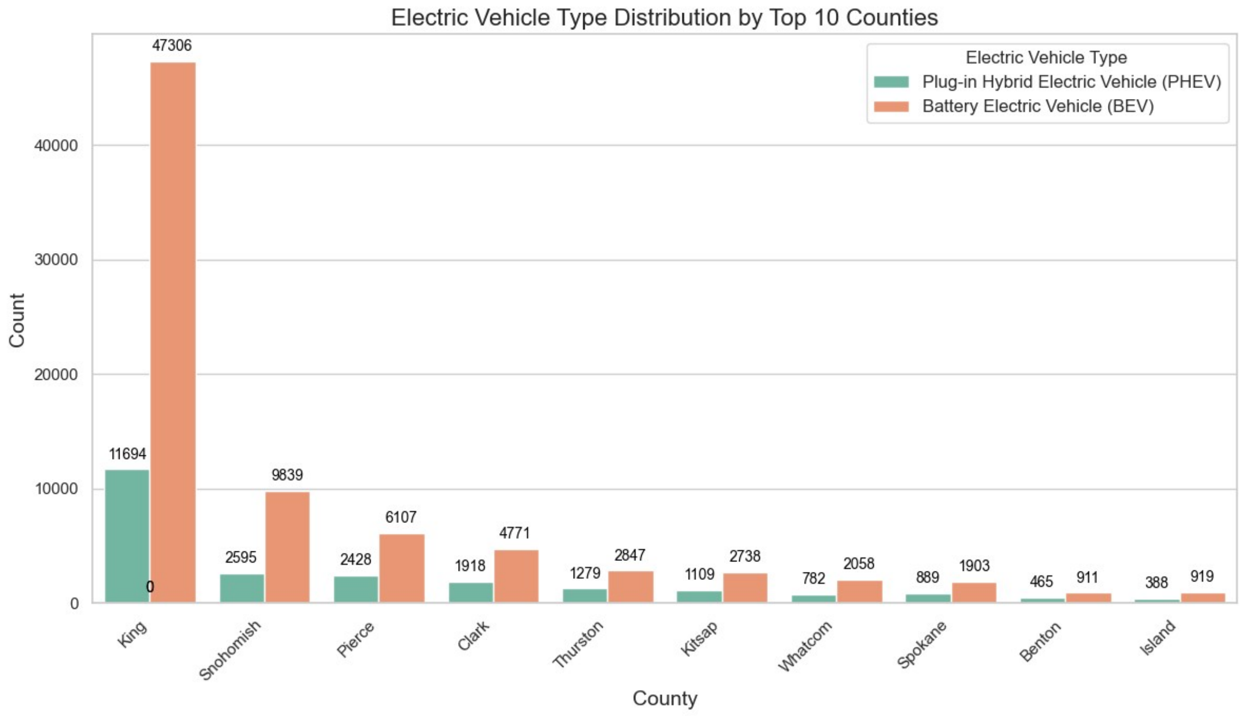


```
df_filtered = df[df['Base MSRP'] <= 80000]

plt.figure(figsize=(12, 7))
sns.set(style='whitegrid')
sns.boxplot(data=df_filtered, x='Electric Vehicle Type', y='Base
MSRP', palette='Set2')

plt.title('Base MSRP by Electric Vehicle Type (Outliers Removed)',
fontsize=16)
plt.xlabel('Electric Vehicle Type', fontsize=14)
plt.ylabel('Base MSRP', fontsize=14)

plt.xticks(rotation=45, ha='right', fontsize=12)
plt.grid(True, axis='y', linestyle='--', alpha=0.7)

plt.tight_layout()
plt.show()
```

```
C:\Users\sande\AppData\Local\Temp\ipykernel_11100\2523930659.py:5:
FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be
removed in v0.14.0. Assign the `x` variable to `hue` and set
`legend=False` for the same effect.

  sns.boxplot(data=df_filtered, x='Electric Vehicle Type', y='Base
MSRP', palette='Set2')
```

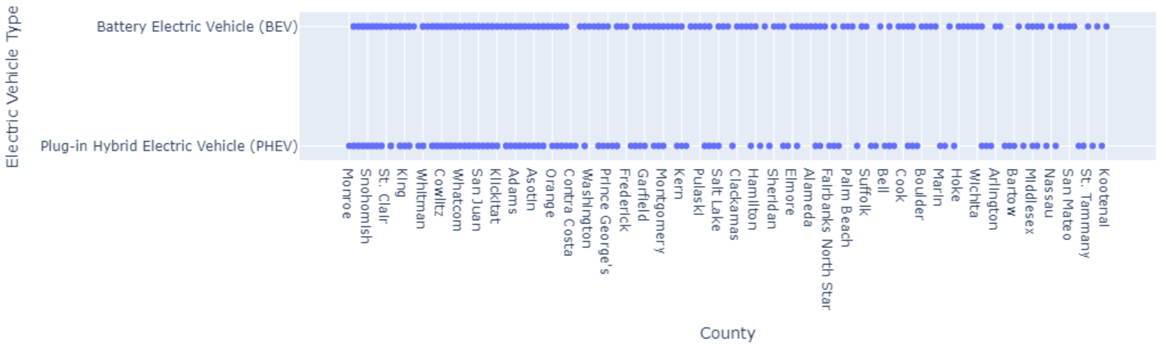


```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Assuming your dataset is already loaded into df
# First, aggregate the data to get the top counties by count
top_counties = df['County'].value_counts().head(10).index  # Limit to
top 10 counties

# Filter the dataset to include only the top 10 counties
df_filtered = df[df['County'].isin(top_counties)]

# Sort the counties by the total number of electric vehicles for
better visual representation
county_order = df_filtered['County'].value_counts().index

# Create the plot
plt.figure(figsize=(12, 7))  # Increase the figure size for better
readability
sns.set_palette('Set2')  # Use a more appealing color palette
sns.countplot(data=df_filtered, x='County', hue='Electric Vehicle
Type', order=county_order)

# Add title and labels with larger font size
plt.title('Electric Vehicle Type Distribution by Top 10 Counties',
fontsize=16)
plt.xlabel('County', fontsize=14)
plt.ylabel('Count', fontsize=14)

# Rotate x-axis labels for better readability
plt.xticks(rotation=45, ha='right')

# Add a legend with a more descriptive title
plt.legend(title='Electric Vehicle Type', fontsize=12)

# Add counts on top of the bars
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}', (p.get_x() +
p.get_width() / 2., p.get_height()),
                       ha='center', va='center', fontsize=10,
color='black', xytext=(0, 10),
                       textcoords='offset points')

plt.tight_layout()  # Adjust layout to prevent overlap
plt.show()
```

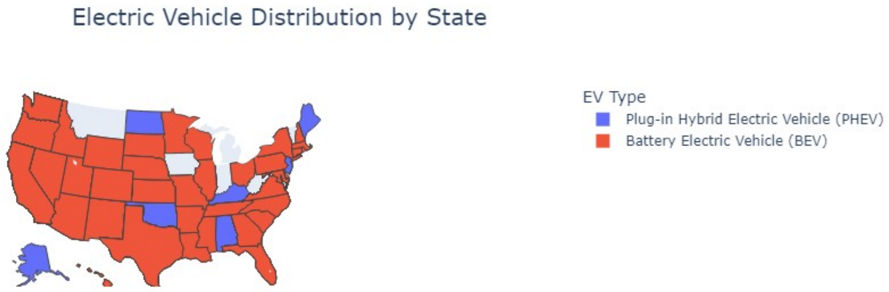


plotly

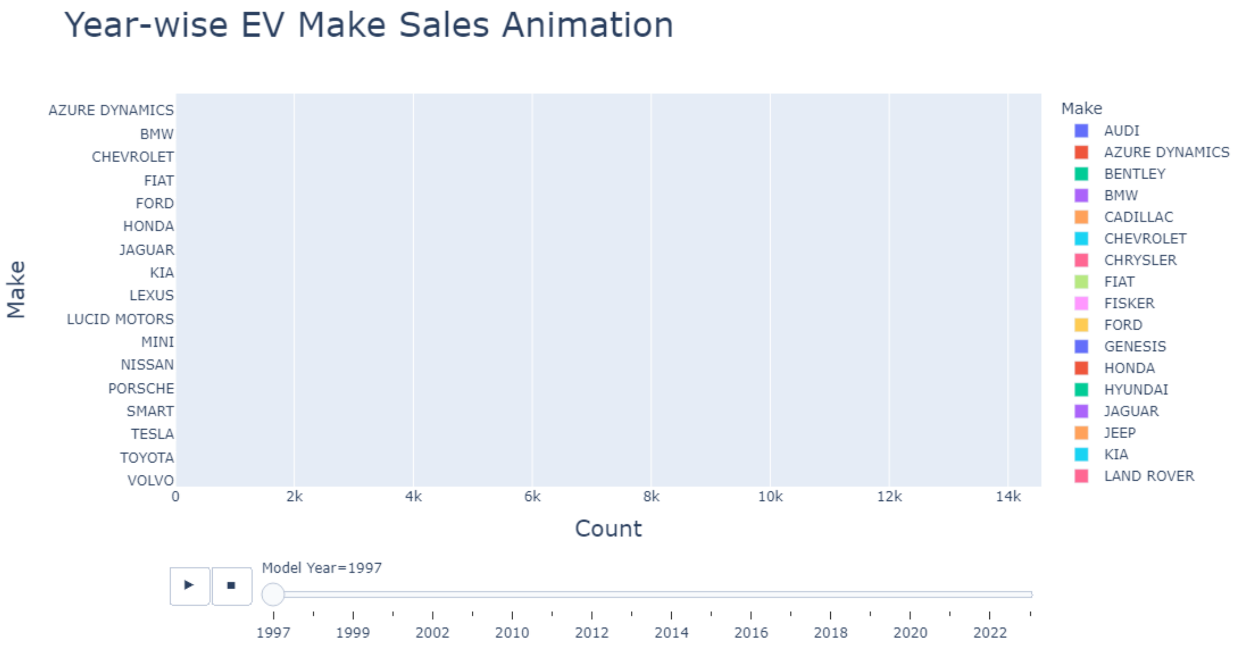
```
import plotly.express as px

px.scatter(df,x='County',y='Electric Vehicle Type')
```



```
fig = px.choropleth(df,
                    locations="State",
                    locationmode="USA-states",
                    color="Electric Vehicle Type",
                    hover_name="County",
                    hover_data=["Make", "Electric Range", "Base
MSRP"],
                    scope="usa",
                    color_continuous_scale="Viridis",
                    labels={'Electric Vehicle Type':'EV Type'},
                    title="Electric Vehicle Distribution by State")

fig.update_layout(geo=dict(showlakes=True, lakecolor='rgb(255, 255,
255)'),
                  title_font_size=20,
                  title_x=0.5,
                  legend_title="EV Type",
                  font=dict(size=12))

fig.show()
```


Electric Vehicle Distribution by State
EV Type
🟦 Plug-in Hybrid Electric Vehicle (PHEV)
🟥 Battery Electric Vehicle (BEV)

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

# Create a pivot table with counts of vehicles by 'Make' and 'Model
Year'
pivot_data = df.pivot_table(index="Model Year", columns="Make",
aggfunc="size", fill_value=0)

# Reset index to make 'Model Year' a column
pivot_data.reset_index(inplace=True)
melted_data = pivot_data.melt(id_vars=["Model Year"], var_name="Make",
value_name="Count")

# Create an animated bar plot
fig = px.bar(melted_data,
             x='Count',
             y='Make',
             color='Make',
             animation_frame='Model Year',
             range_x=[0, melted_data['Count'].max() + 10],  # Adjust
range for better visualization
             title='Year-wise EV Make Sales Animation',
             orientation='h')


fig.update_layout(
    title_font=dict(size=30),
    xaxis_title_font=dict(size=20),
    yaxis_title_font=dict(size=20),
    width=1000,
    height=600,
    bargap=0.1,
)

fig.show()
```


Year-wise EV Make Sales Animation
Make
AZURE DYNAMICS
BMW
CHEVROLET
FIAT
FORD
HONDA
JAGUAR
KIA
LEXUS
LUCID MOTORS
MINI
NISSAN
PORSCHE
SMART
TESLA
TOYOTA
VOLVO
2k
4k
6k
8k
10k
12k
14k
Count
Make
AUDI
AZURE DYNAMICS
BENTLEY
BMW
CADILLAC
CHEVROLET
CHRYSLER
FIAT
FISKER
FORD
GENESIS
HONDA
HYUNDAI
JAGUAR
JEEP
KIA
LAND ROVER
▶
■
Model Year=1997