electric-vehicle-task

October 12, 2024

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

[2]:
```
df = pd.read_csv(r'C:\Users\user\Downloads\dataset.csv')
```

[3]:
```
df
```

[3]:

	VIN (1-10)	County	City	State	Postal Code	Model Year	Make	Model	Electric Vehicle Type	Clean Alternative Fuel Vehicle (CAFV) Eligibility	Electric Range
0	JTMEB3FV6N	Monroe	Key West	FL	33040	2022	TOYOTA	RAV4 PRIME	Plug-in Hybrid Electric Vehicle (PHEV)	Clean Alternative Fuel Vehicle Eligible	42
1	1G1RD6E45D	Clark	Laughlin	NV	89029	2013	CHEVROLET	VOLT	Plug-in Hybrid Electric Vehicle (PHEV)		
2	JN1AZ0CP8B	Yakima	Yakima	WA	98901	2011	NISSAN	LEAF	Battery Electric Vehicle (BEV)		
3	1G1FW6S08H	Skagit	Concrete	WA	98237	2017	CHEVROLET	BOLT EV	Battery Electric Vehicle (BEV)		
4	3FA6P0SU1K	Snohomish	Everett	WA	98201	2019	FORD	FUSION	Plug-in Hybrid Electric Vehicle (PHEV)		
…	…	…	…	…	…	…	…	…	…	…	…
112629	7SAYGDEF2N	King	Duvall	WA	98019	2022	TESLA	MODEL Y	Battery Electric Vehicle (BEV)		
112630	1N4BZ1CP7K	San Juan	Friday Harbor	WA	98250	2019	NISSAN	LEAF	Battery Electric Vehicle (BEV)		
112631	1FMCU0KZ4N	King	Vashon	WA	98070	2022	FORD	ESCAPE	Plug-in Hybrid Electric Vehicle (PHEV)		
112632	KNDCD3LD4J	King	Covington	WA	98042	2018	KIA	NIRO	Plug-in Hybrid Electric Vehicle (PHEV)		
112633	YV4BR0CL8N	King	Covington	WA	98042	2022	VOLVO	XC90	Plug-in Hybrid Electric Vehicle (PHEV)		

```
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]
```

[4]:
```
df.head()
```

[4]:
```
    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
```

[5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 112634 entries, 0 to 112633
Data columns (total 17 columns):
```

```
 #   Column                                             Non-Null Count   Dtype
---  ------                                             --------------   -----
 0   VIN (1-10)                                         112634 non-null  object
 1   County                                             112634 non-null  object
 2   City                                               112634 non-null  object
 3   State                                              112634 non-null  object
 4   Postal Code                                        112634 non-null  int64
 5   Model Year                                         112634 non-null  int64
 6   Make                                               112634 non-null  object
 7   Model                                              112614 non-null  object
 8   Electric Vehicle Type                              112634 non-null  object
 9   Clean Alternative Fuel Vehicle (CAFV) Eligibility  112634 non-null  object
 10  Electric Range                                     112634 non-null  int64
 11  Base MSRP                                          112634 non-null  int64
 12  Legislative District                               112348 non-null  float64
 13  DOL Vehicle ID                                     112634 non-null  int64
 14  Vehicle Location                                   112610 non-null  object
 15  Electric Utility                                   112191 non-null  object
 16  2020 Census Tract                                  112634 non-null  int64
dtypes: float64(1), int64(6), object(10)
memory usage: 14.6+ MB
```

[6]:
```
df.isnull().sum()
```

[6]:
```
VIN (1-10)                                           0
County                                               0
City                                                 0
State                                                0
Postal Code                                          0
Model Year                                           0
Make                                                 0
Model                                               20
Electric Vehicle Type                                0
Clean Alternative Fuel Vehicle (CAFV) Eligibility    0
Electric Range                                       0
Base MSRP                                            0
Legislative District                               286
DOL Vehicle ID                                       0
Vehicle Location                                    24
Electric Utility                                   443
2020 Census Tract                                    0
dtype: int64
```

[7]:
```
df.describe()
```

[7]:
```
          Postal Code     Model Year  Electric Range      Base MSRP  \
count  112634.000000  112634.000000   112634.000000  112634.000000
```

```
mean    98156.226850    2019.003365       87.812987    1793.439681
std      2648.733064       2.892364      102.334216   10783.753486
min      1730.000000    1997.000000        0.000000       0.000000
25%     98052.000000    2017.000000        0.000000       0.000000
50%     98119.000000    2020.000000       32.000000       0.000000
75%     98370.000000    2022.000000      208.000000       0.000000
max     99701.000000    2023.000000      337.000000  845000.000000

       Legislative District  DOL Vehicle ID  2020 Census Tract
count         112348.000000    1.126340e+05       1.126340e+05
mean              29.805604    1.994567e+08       5.296650e+10
std               14.700545    9.398427e+07       1.699104e+09
min                1.000000    4.777000e+03       1.101001e+09
25%               18.000000    1.484142e+08       5.303301e+10
50%               34.000000    1.923896e+08       5.303303e+10
75%               43.000000    2.191899e+08       5.305307e+10
max               49.000000    4.792548e+08       5.603300e+10
```

[9]:
```
df.shape
```

[9]:
```
(112634, 17)
```

0.1 EDA_Exploratary Data Analysis

[10]:
```
df.duplicated().sum()
```

[10]:
```
0
```

[11]:
```
# df["Model"]=df["Model"].fillna(df["Model"].mode()[0])
```

[12]:
```
# df["Legislative District"]=df["Legislative District"].fillna(df["Legislative
↪District"].mean())
```

[13]:
```
# df["Vehicle Location"]=df["Vehicle Location"].fillna(df["Vehicle Location"].
↪mode()[0])
```

[15]:
```
# df["Electric Utility"]=df["Electric Utility"].fillna(df["Electric Utility"].
↪mode()[0])
```

[16]:
```
# df.isna().sum()
```

[17]:
```
# df.to_csv("Analysis on Electric Vehicles")
```

[18]:
```
# df.shape
```

[19]:
```
# df.info()
```

0.2 Univariate Analysis

0.2.1 Import required library - plotly.express

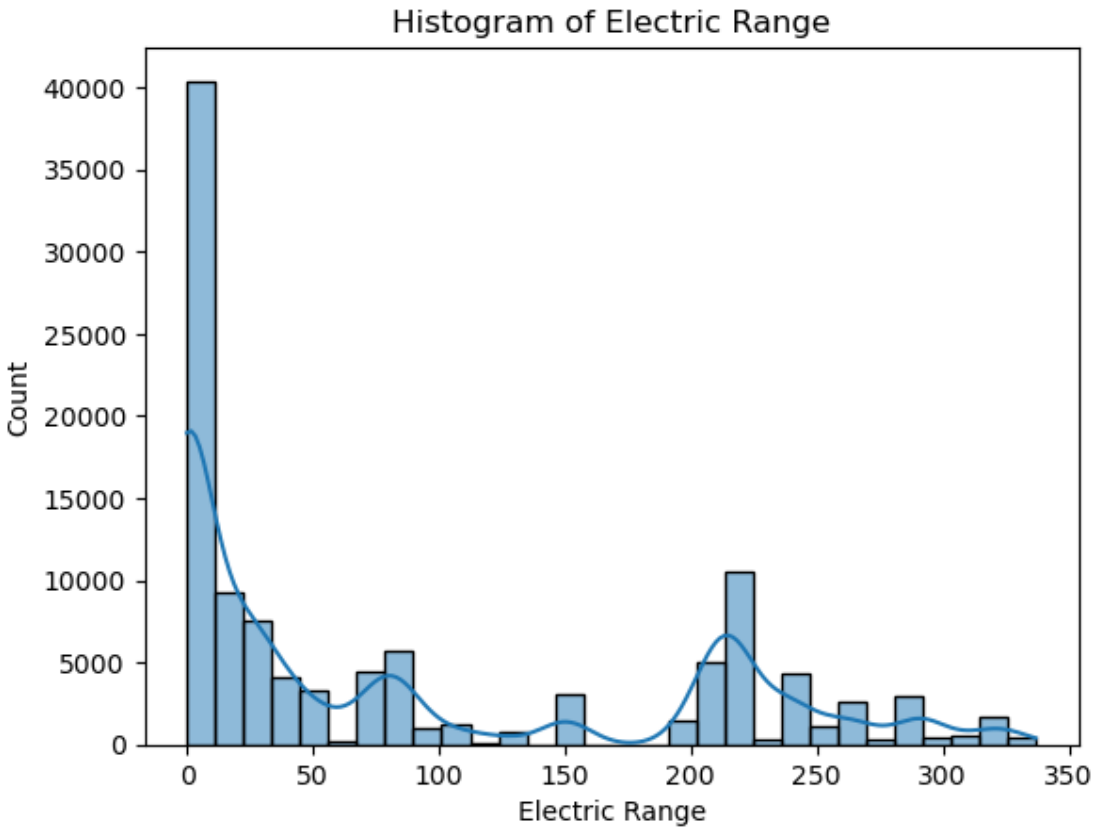
[20]:
```
import plotly.express as px
```

[21]:
```
fig = px.box(df, y='Electric Range',
             title="Box Plot of Electric Range",
             labels={'Electric Range': 'Electric Range'},
             color_discrete_sequence=["#FF5733"])  # Custom color (e.g., orange)

# Customize layout
fig.update_layout(yaxis_title="Electric Range", width=800, height=600)

# Show plot
fig.show()
```

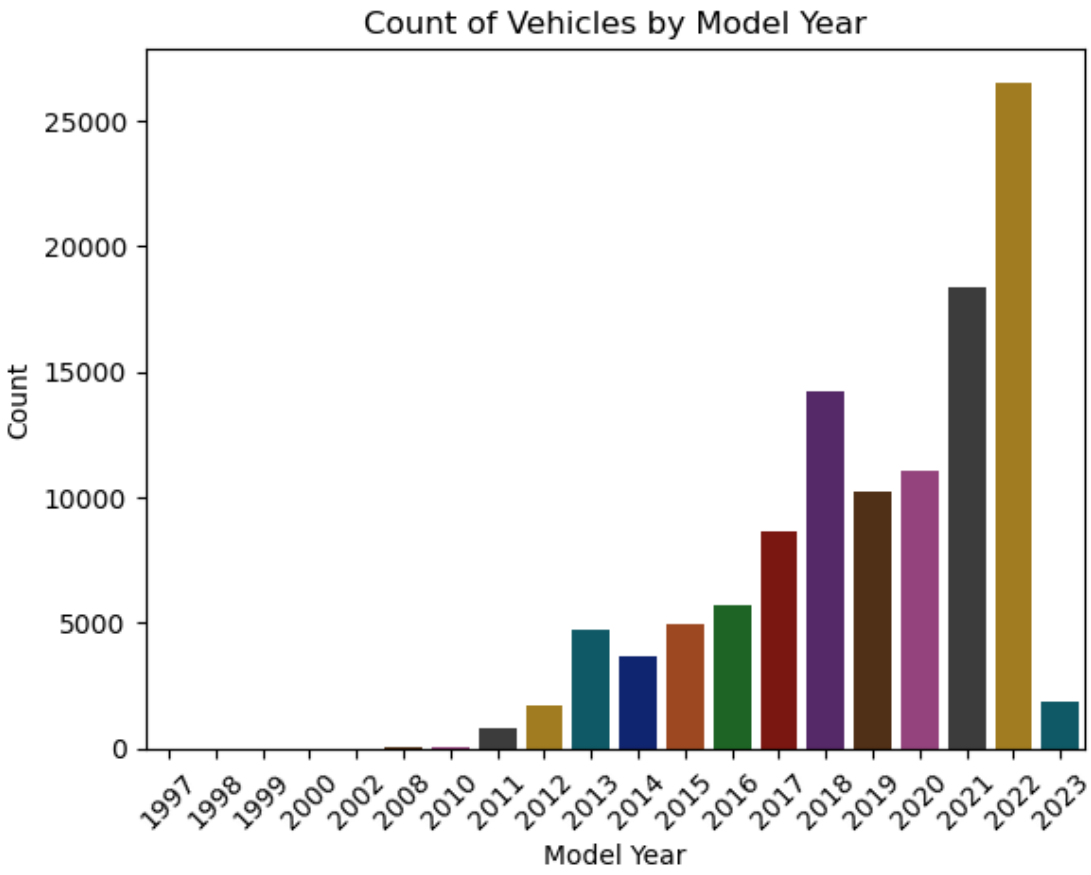
[22]:
```
px.histogram(df, x='Base MSRP', title="Histogram of Base MSRP", nbins=30,
color_discrete_sequence=['#EF553B']).update_layout(
xaxis_title="Base MSRP", yaxis_title="Count", width=800, height=600).show()
```

0.2.2 Histograms for numerical features

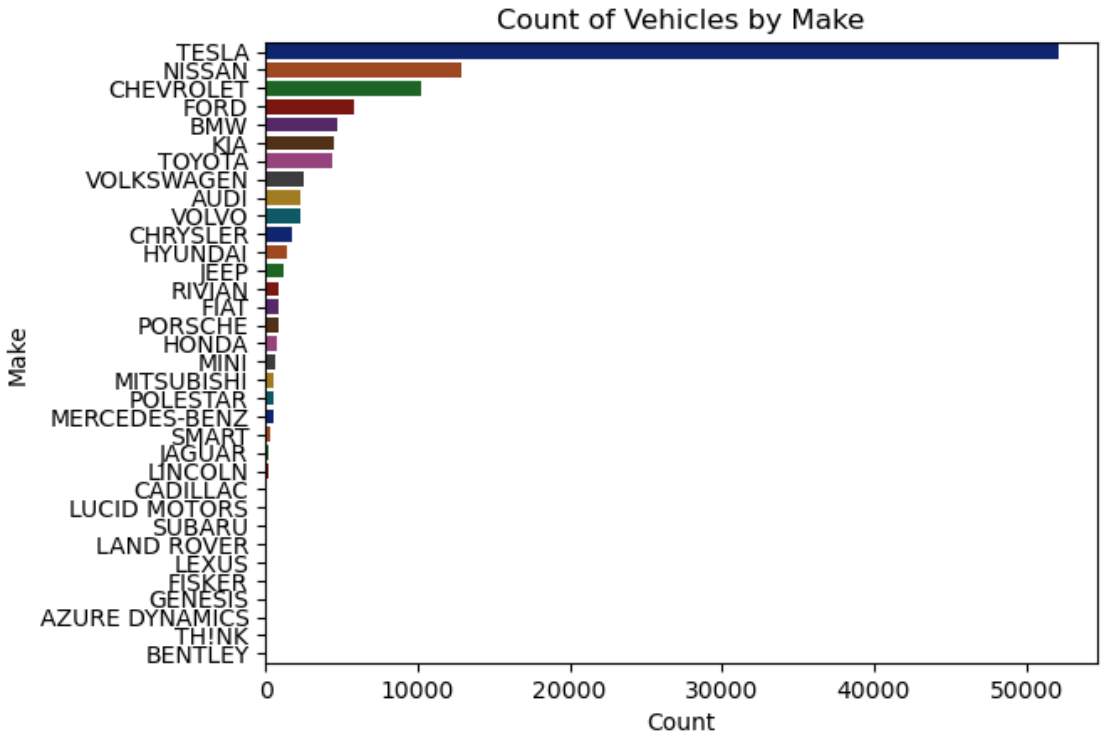
[23]:
```
# Replace 'Electric Range' with the actual column name if it contains spaces.
sns.histplot(df['Electric Range'],
bins=30, kde=True).set_title('Histogram of Electric Range')
plt.show()
```



[24]:
```
# Use a darker color palette
sns.countplot(x='Model Year', data=df, palette='dark')
plt.title('Count of Vehicles by Model Year')
plt.xticks(rotation=45)
plt.xlabel('Model Year')
plt.ylabel('Count')
plt.show()
```



[25]:
```
# Use a dark color palette
sns.countplot(y='Make', data=df, order=df['Make'].value_counts().index,␣
 ↪palette='dark')
plt.title('Count of Vehicles by Make')
plt.ylabel('Make')
plt.xlabel('Count')
plt.show()
```



0.3 Bivariate Analysis

0.3.1 Scatter plot using plotly.express Numerical vs Numerical A(Electric Range vs Make)

0.3.2 Task- This is an open ended problem.apply exploratory data analysis (Univariate and Bivariate) on the dataset available above.

[26]:
```
px.scatter(df,x = "Make",y ="Electric Range")
```

0.4 Box plot using plotly.Express

[27]:
```
px.box(df, x = "Electric Vehicle Type", y = "Electric Range")
```

0.4.1 pie chart plot using plotly.Express

[28]:
```
px.pie(df,names = "Make", values = "2020 Census Tract")
```

[29]:
```
df["State"].unique()
```

[29]:
```
array(['FL', 'NV', 'WA', 'IL', 'NY', 'VA', 'OK', 'KS', 'CA', 'NE', 'MD',
       'CO', 'DC', 'TN', 'SC', 'CT', 'OR', 'TX', 'SD', 'HI', 'GA', 'MS',
       'AR', 'NC', 'MO', 'UT', 'PA', 'DE', 'OH', 'WY', 'AL', 'ID', 'AZ',
```

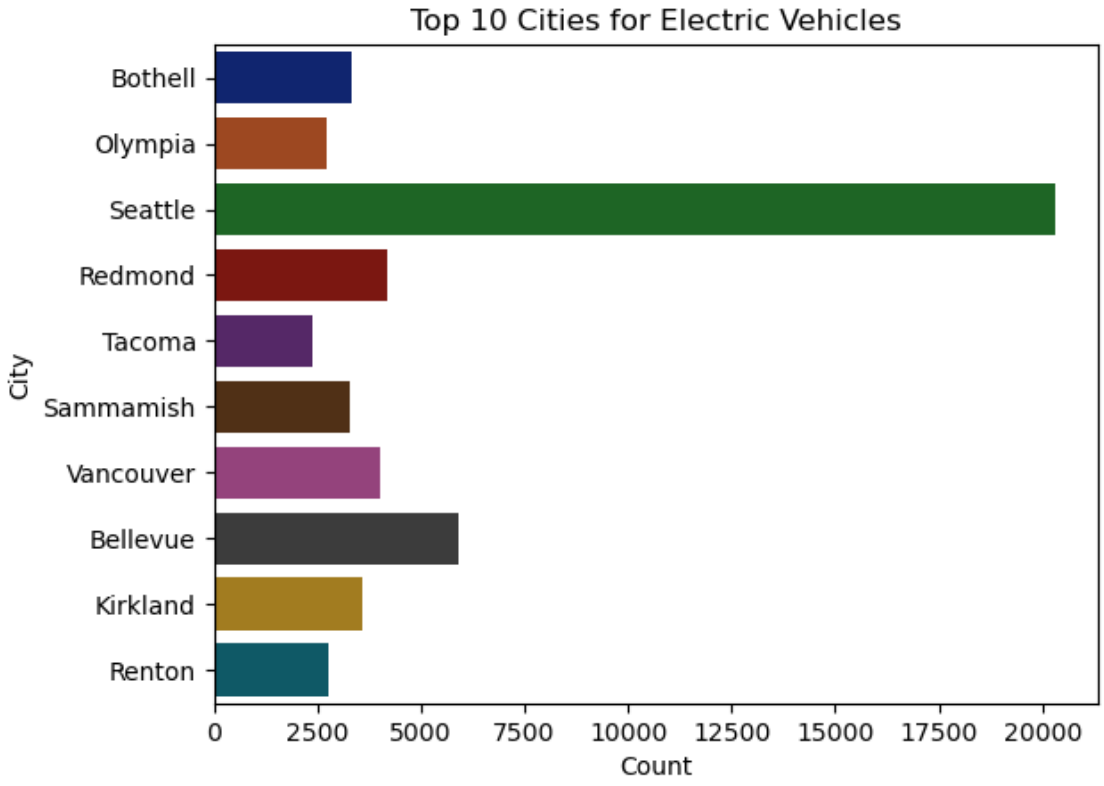
```
'AK', 'LA', 'NM', 'WI', 'KY', 'NJ', 'MN', 'MA', 'ME', 'RI', 'NH',
'ND'], dtype=object)
```

[30]:
```
grouped_df = df.groupby("State").agg({"Electric Range":"mean"})
```

[31]:
```
df.shape
```

[31]:
```
(112634, 17)
```

[32]:
```
# Use a dark color palette
sns.set_palette("dark")

top_cities = df['City'].value_counts().nlargest(10).index  # Top 10 cities
sns.countplot(y='City', data=df[df['City'].isin(top_cities)])
plt.title('Top 10 Cities for Electric Vehicles')
plt.xlabel('Count')
plt.ylabel('City')
plt.show()
```



0.4.2 Heatmap of Correlation (for numeric variables)

[40]:
```
# Select only numeric columns
df_numeric = df.select_dtypes(include=[np.number])

# Calculate the correlation matrix
correlation_matrix = df_numeric.corr()

# Create heatmap for correlation matrix
import plotly.express as px
fig = px.imshow(correlation_matrix, title="Correlation Heatmap", text_auto=True)
fig.show()
```

[34]:
```
# Pie chart for Electric Vehicle Type
fig = px.pie(df, names='Electric Vehicle Type', title="Distribution of Electric␣
 ↪Vehicle Types")
fig.show()
```

[36]:
```
import pandas as pd, plotly.express as px

df = pd.read_csv(r"C:\Users\user\Downloads\dataset.csv")
state_nyc = df.groupby(['Postal Code', 'Model Year']).size().
 ↪reset_index(name='Number_of_Vehicles')

fig = px.choropleth_mapbox(state_nyc, geojson='https://raw.githubusercontent.
 ↪com/python-visualization/folium/master/examples/data/us-states.json',
    locations='Postal Code', color='Number_of_Vehicles',␣
 ↪featureidkey="properties.ZCTA5CE10", mapbox_style="carto-positron",
    zoom=5, center={"lat": 47.7511, "lon": -120.7401}, animation_frame='Model␣
 ↪Year')

fig.update_layout(margin={"r": 0, "t": 0, "l": 0, "b": 0}).show()
```

[38]:
```
# pip install bar_chart_race
```

[39]:
```
import pandas as pd, plotly.express as px

df = pd.read_csv(r"C:\Users\user\Downloads\dataset.csv")
d = df.groupby(['Make', 'Model Year']).size().
 ↪reset_index(name='Number_of_Vehicles')

px.bar(d, x='Number_of_Vehicles', y='Make', color='Make',␣
 ↪animation_frame='Model Year', orientation='h',
       title='EV Makes and Their Count Over the Years', range_x=[0, 3000]).
 ↪update_traces(
    texttemplate='%{x}', textposition='outside').update_layout(
```

```
    yaxis_title='EV Makes', xaxis_title='Number of EV Vehicles', title_x=0.5,␣
↪width=800, height=600).show()
```

[ ]: