

DSC680 -Project Milestone 4 - sidbhaumik

April 6, 2024

0.0.1 Dataset: US_Electric_fuel_vehicles.csv

This csv dataset contains a list of all Electric Vehicles (EVs) and Alternative Fuel Vehicles (AFVs) available

in the US, as of July 2022

```
[2]: # Import required libraries
import pandas as pd
import numpy as np
```

```
[3]: # Read csv Dataset: US_Electric_fuel_vehicles.csv
# Importing a csv file into DataFrame format. This format is easy to read and
    ↳ query the data.
vehicle_df = pd.read_csv("US_Electric_fuel_vehicles.csv")
```

```
[4]: # Reading first 5 rows
vehicle_df.head()
```

```
[4]:
```

	Vehicle ID	Fuel ID	Fuel Configuration ID	Manufacturer ID	Category ID	\
0	13044	45	9.0	365	27	
1	12854	45	9.0	377	27	
2	12842	45	9.0	377	27	
3	12783	45	9.0	377	27	
4	12782	45	9.0	377	27	

	Model	Model Year	Alternative Fuel Economy City	\
0	NSX	2022	NaN	
1	A3	2022	NaN	
2	A3 quattro	2022	NaN	
3	A4 allroad quattro	2022	NaN	
4	A4 quattro	2022	NaN	

	Alternative Fuel Economy Highway	Alternative Fuel Economy Combined	...	\
0	NaN	NaN	...	
1	NaN	NaN	...	
2	NaN	NaN	...	
3	NaN	NaN	...	
4	NaN	NaN	...	

	Manufacturer URL	Category	Fuel Code	Fuel	\
0	http://www.acura.com/	Sedan/Wagon	HYBR	Hybrid Electric	
1	http://progress.audiusa.com/	Sedan/Wagon	HYBR	Hybrid Electric	
2	http://progress.audiusa.com/	Sedan/Wagon	HYBR	Hybrid Electric	
3	http://progress.audiusa.com/	Sedan/Wagon	HYBR	Hybrid Electric	
4	http://progress.audiusa.com/	Sedan/Wagon	HYBR	Hybrid Electric	

	Fuel Configuration Name	Electric-Only Range	PHEV Total Range	PHEV Type	\
0	Hybrid Electric	NaN	NaN	NaN	
1	Hybrid Electric	NaN	NaN	NaN	
2	Hybrid Electric	NaN	NaN	NaN	
3	Hybrid Electric	NaN	NaN	NaN	
4	Hybrid Electric	NaN	NaN	NaN	

	Notes	Drivetrain
0	NaN	AWD
1	NaN	FWD
2	NaN	AWD
3	NaN	AWD
4	NaN	AWD

[5 rows x 29 columns]

```
[5]: # Total number of rows and columns in the vehicle dataframe
vehicle_df.shape
```

```
[5]: (3008, 29)
```

```
[7]: # Creating new Dataframe with useful columns only.
vehicle_df2 = vehicle_df.drop(vehicle_df.columns[[2,7, 8, 10, 11, 14, 16, 17, 19, 23, 26, 27]], axis=1)
```

```
[8]: vehicle_df2.head()
```

	Vehicle ID	Fuel ID	Manufacturer ID	Category ID	Model	\
0	13044	45	365	27	NSX	
1	12854	45	377	27	A3	
2	12842	45	377	27	A3 quattro	
3	12783	45	377	27	A4 allroad quattro	
4	12782	45	377	27	A4 quattro	

	Model Year	Alternative Fuel Economy Combined	\
0	2022	NaN	
1	2022	NaN	
2	2022	NaN	
3	2022	NaN	

4	2022	NaN
---	------	-----

	Conventional Fuel Economy Combined	Transmission Type	Engine Size \
0	21.0	Auto	3.5L
1	32.0	Auto	2.0L
2	31.0	Auto	2.0L
3	26.0	Auto	2.0L
4	29.0	Auto	2.0L

	Manufacturer	Category	Fuel Code	Fuel	Electric-Only Range \
0	Acura	Sedan/Wagon	HYBR	Hybrid Electric	NaN
1	Audi	Sedan/Wagon	HYBR	Hybrid Electric	NaN
2	Audi	Sedan/Wagon	HYBR	Hybrid Electric	NaN
3	Audi	Sedan/Wagon	HYBR	Hybrid Electric	NaN
4	Audi	Sedan/Wagon	HYBR	Hybrid Electric	NaN

	PHEV Total Range	Drivetrain
0	NaN	AWD
1	NaN	FWD
2	NaN	AWD
3	NaN	AWD
4	NaN	AWD

```
[52]: # Renaming column names to ease dataframe operations
vehicle_df2.rename({'Vehicle ID': 'Vehicle_ID', 'Fuel ID': 'Fuel_ID', 'Fuel Code':
↳ 'Fuel_code', 'Model Year': 'Model_yr', 'Manufacturer ID':
↳ 'Manufacturer_ID', 'Transmission Type': 'Transmission_typ', 'Engine Size':
↳ 'Engine_Size', 'Electric-Only Range': 'Electric_Range', 'PHEV Total Range':
↳ 'PHEV_range', 'Alternative Fuel Economy Combined':
↳ 'Alt_fuel_Eco_Combd', 'Conventional Fuel Economy Combined':
↳ 'Conv_Fuel_Eco_Combd'}, axis=1, inplace=True)
```

```
[54]: vehicle_df2.head()
```

	Vehicle_ID	Fuel_ID	Manufacturer_ID	Category ID	Model \
0	13044	45	365	27	NSX
1	12854	45	377	27	A3
2	12842	45	377	27	A3 quattro
3	12783	45	377	27	A4 allroad quattro
4	12782	45	377	27	A4 quattro

	Model_yr	Alt_fuel_Eco_Combd	Conv_Fuel_Eco_Combd	Transmission_typ \
0	2022	NaN	21.0	auto
1	2022	NaN	32.0	auto
2	2022	NaN	31.0	auto
3	2022	NaN	26.0	auto
4	2022	NaN	29.0	auto

	Engine_Size	Manufacturer	Category	Fuel_code	Fuel	\
0	3.5L	Acura	Sedan/Wagon	HYBR	Hybrid Electric	
1	2.0L	Audi	Sedan/Wagon	HYBR	Hybrid Electric	
2	2.0L	Audi	Sedan/Wagon	HYBR	Hybrid Electric	
3	2.0L	Audi	Sedan/Wagon	HYBR	Hybrid Electric	
4	2.0L	Audi	Sedan/Wagon	HYBR	Hybrid Electric	

	Electric_Range	PHEV_range	Drivetrain
0	NaN	NaN	AWD
1	NaN	NaN	FWD
2	NaN	NaN	AWD
3	NaN	NaN	AWD
4	NaN	NaN	AWD

```
[55]: # Changing the column order to keep the most useful columns in the beginning
      ↪ and grouped together
      vehicle_df2.iloc[:, [0,1,2,3,10,4,5,12,13,11,14,15,16,6,7,8,9]]
```

```
[55]:
```

	Vehicle_ID	Fuel_ID	Manufacturer_ID	Category	ID	Manufacturer	\
0	13044	45	365		27	Acura	
1	12854	45	377		27	Audi	
2	12842	45	377		27	Audi	
3	12783	45	377		27	Audi	
4	12782	45	377		27	Audi	
...		
3003	656	3	225		9	Dodge	
3004	660	49	225		27	Dodge	
3005	658	49	219		9	Ford	
3006	650	49	215		27	Chevrolet	
3007	649	49	219		27	Ford	

	Model	Model_yr	Fuel_code	\
0	NSX	2022	HYBR	
1	A3	2022	HYBR	
2	A3 quattro	2022	HYBR	
3	A4 allroad quattro	2022	HYBR	
4	A4 quattro	2022	HYBR	
...	
3003	Ram Van B250	1992	CNG	
3004	Spirit FFV	1992	M85_GSLN	
3005	Econoline-Fleet Demo FFV	1992	M85_GSLN	
3006	Lumina FFV	1991	M85_GSLN	
3007	Taurus FFV	1991	M85_GSLN	

	Fuel	Category	Electric_Range	PHEV_range	\
0	Hybrid Electric	Sedan/Wagon	NaN	NaN	

1		Hybrid Electric	Sedan/Wagon		NaN	NaN
2		Hybrid Electric	Sedan/Wagon		NaN	NaN
3		Hybrid Electric	Sedan/Wagon		NaN	NaN
4		Hybrid Electric	Sedan/Wagon		NaN	NaN
...	
3003	CNG - Compressed Natural Gas		Van		NaN	NaN
3004		Methanol	Sedan/Wagon		NaN	NaN
3005		Methanol	Van		NaN	NaN
3006		Methanol	Sedan/Wagon		NaN	NaN
3007		Methanol	Sedan/Wagon		NaN	NaN

	Drivetrain	Alt_fuel_Eco_Combd	Conv_Fuel_Eco_Combd	Transmission_typ \
0	AWD	NaN	21.0	auto
1	FWD	NaN	32.0	auto
2	AWD	NaN	31.0	auto
3	AWD	NaN	26.0	auto
4	AWD	NaN	29.0	auto
...
3003	NaN	NaN	NaN	NaN
3004	NaN	NaN	NaN	NaN
3005	NaN	NaN	NaN	NaN
3006	NaN	NaN	NaN	NaN
3007	NaN	NaN	NaN	automatic

	Engine_Size
0	3.5L
1	2.0L
2	2.0L
3	2.0L
4	2.0L
...	...
3003	3.3 L
3004	NaN
3005	NaN
3006	NaN
3007	3.0 L

[3008 rows x 17 columns]

```
[56]: # Check for duplicates in key column
print("The column Vehicle_ID has duplicate - {}".format(any(vehicle_df2.
↪Vehicle_ID.duplicated())))
```

The column Vehicle_ID has duplicate - False

```
[57]: # Check for duplicates in the entire dataframe
```

```
print("Vehicle_df2 dataframe has duplicate - {}".format(any(vehicle_df2.
↳ duplicated())))
```

Vehicle_df2 dataframe has duplicate - False

```
[58]: # Checking outliers
print('Null values in PHEV_range column: ',vehicle_df2['PHEV_range'].isnull().
↳ sum())
```

Null values in PHEV_range column: 2965

```
[59]: # Checking for Unique values in Transmission_typ column
vehicle_df2['Transmission_typ'].unique()
```

```
[59]: array(['auto', nan, 'manual', 'ecvt', 'auto / man', 'automatic / 6 speed',
'semi-automatic', 'automatic / 4 speed', 'automatic',
'automatic / 1 speed', 'automatic / 5 speed', '6 speed automatic',
'cvt', 'manual / 6 speed', '3 mode / 1 speed',
'automatic / 7 speed', 'ecvt / 8 speed', 'automatic / variable',
'automatic / 8 speed', 'manual / 5 speed', 'none',
'5-speed automatic', 'v6, hybrid, 3.5l', 'automatic cvt',
'continuously variable', '4-speed automatic', '6-speed automatic',
'variable automatic', 'automatic evt', 'sohc', 'direct drive',
'7-speed automatic', 'multispeed automatic', '6-speed manual',
'4 speed automatic', '5 speed automatic', '4-speed automatic,',
'manual 5-speed or cvt', '5-speed auto with o/d',
'automatic (cvt)', 'electronic, 4-speed automatic',
'electronic 4-speed automatic', '4-speed automatic, overdrive',
'single-speed automatic', '4-speed auto or 5 speed manual',
'a/t 5 speed overdrive', 'transaxle w/2-stage gear set',
'single speed', '4-speed automatic/5 spd manual', 'single-speed',
'automatic overdrive', 'type single-speed automatic',
'single speed w/dual reduction', 'manual 5 or 6 speed/automatic',
'single-speed, fixed ratio', '3-speed automatic',
'4-speed automatic w/overdrive', 'cd4e, 4-speed automatic',
'single-speed, reduction gear'], dtype=object)
```

```
[60]: # Transmission_typ can be an useful column. So, need to fix the casing to all_
↳ lowercase
vehicle_df2['Transmission_typ'] = vehicle_df2['Transmission_typ'].str.lower()
```

```
[61]: # Checking the vehicle_df2 dataframe for Transmission_typ column casing changes
vehicle_df2.head()
```

```
[61]:   Vehicle_ID  Fuel_ID  Manufacturer_ID  Category ID  Model \
0      13044      45          365          27      NSX
1      12854      45          377          27      A3
```

2	12842	45	377	27	A3 quattro
3	12783	45	377	27	A4 allroad quattro
4	12782	45	377	27	A4 quattro

	Model_yr	Alt_fuel_Eco_Combd	Conv_Fuel_Eco_Combd	Transmission_typ	\
0	2022	NaN	21.0	auto	
1	2022	NaN	32.0	auto	
2	2022	NaN	31.0	auto	
3	2022	NaN	26.0	auto	
4	2022	NaN	29.0	auto	

	Engine_Size	Manufacturer	Category	Fuel_code	Fuel	\
0	3.5L	Acura	Sedan/Wagon	HYBR	Hybrid Electric	
1	2.0L	Audi	Sedan/Wagon	HYBR	Hybrid Electric	
2	2.0L	Audi	Sedan/Wagon	HYBR	Hybrid Electric	
3	2.0L	Audi	Sedan/Wagon	HYBR	Hybrid Electric	
4	2.0L	Audi	Sedan/Wagon	HYBR	Hybrid Electric	

	Electric_Range	PHEV_range	Drivetrain
0	NaN	NaN	AWD
1	NaN	NaN	FWD
2	NaN	NaN	AWD
3	NaN	NaN	AWD
4	NaN	NaN	AWD

```
[62]: # Checking for Unique values in Fuel_code column
vehicle_df2['Fuel_code'].unique()
```

```
[62]: array(['HYBR', 'PHEV', 'ELEC', 'BD', 'E85_GSLN', 'H2', 'LPG_GSLN',
        'CNG_GSLN', 'CNG', 'LPG', 'M85_GSLN'], dtype=object)
```

```
[63]: # I am only interested in Electric or Hybrid vehicles. So, filtering the rest
        ↳from my dataframe.
df1 = vehicle_df2[((vehicle_df2.Fuel_code=='HYBR') | (vehicle_df2.
        ↳Fuel_code=='PHEV') | (vehicle_df2.Fuel_code=='ELEC') | (vehicle_df2.
        ↳Fuel_code=='H2'))]

df1.head()
```

```
[63]:   Vehicle_ID  Fuel_ID  Manufacturer_ID  Category  ID  Model  \
0      13044      45      365      27      NSX
1      12854      45      377      27      A3
2      12842      45      377      27      A3 quattro
3      12783      45      377      27  A4 allroad quattro
4      12782      45      377      27      A4 quattro
```

	Model_yr	Alt_fuel_Eco_Combd	Conv_Fuel_Eco_Combd	Transmission_typ	\
--	----------	--------------------	---------------------	------------------	---

0	2022	NaN	21.0	auto
1	2022	NaN	32.0	auto
2	2022	NaN	31.0	auto
3	2022	NaN	26.0	auto
4	2022	NaN	29.0	auto

	Engine_Size	Manufacturer	Category	Fuel_code	Fuel \
0	3.5L	Acura	Sedan/Wagon	HYBR	Hybrid Electric
1	2.0L	Audi	Sedan/Wagon	HYBR	Hybrid Electric
2	2.0L	Audi	Sedan/Wagon	HYBR	Hybrid Electric
3	2.0L	Audi	Sedan/Wagon	HYBR	Hybrid Electric
4	2.0L	Audi	Sedan/Wagon	HYBR	Hybrid Electric

	Electric_Range	PHEV_range	Drivetrain
0	NaN	NaN	AWD
1	NaN	NaN	FWD
2	NaN	NaN	AWD
3	NaN	NaN	AWD
4	NaN	NaN	AWD

So, now I have a cleaner dataset with only useful columns and rows which I need to analyze further on growing Electric vehicle market and its future in US.

```
[64]: # Descriptive statistics and data visualization
print(df1.describe())
```

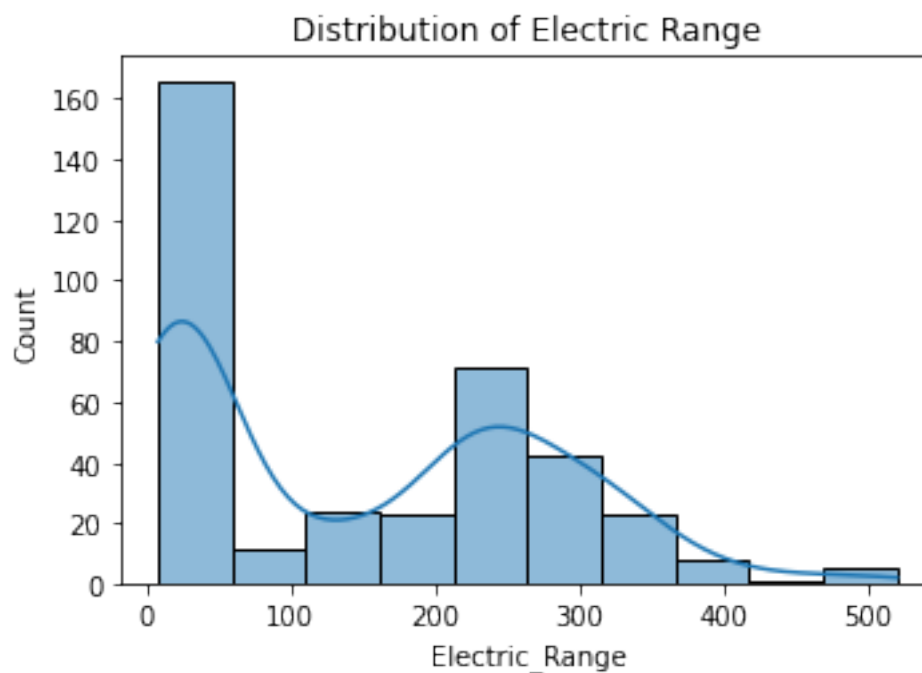
	Vehicle_ID	Fuel_ID	Manufacturer_ID	Category ID	Model_yr \
count	1506.000000	1506.000000	1506.000000	1506.000000	1506.000000
mean	10825.573041	45.382470	278.893758	27.099602	2017.067729
std	3700.268695	7.864792	73.758063	2.841208	5.508018
min	1.000000	9.000000	209.000000	3.000000	1995.000000
25%	11342.250000	41.000000	219.000000	27.000000	2015.000000
50%	12154.500000	45.000000	239.000000	27.000000	2019.000000
75%	12661.750000	45.000000	361.000000	29.000000	2021.000000
max	13105.000000	57.000000	470.000000	29.000000	2022.000000

	Alt_fuel_Eco_Combd	Conv_Fuel_Eco_Combd	Electric_Range	PHEV_range
count	376.000000	588.000000	374.000000	43.000000
mean	86.218085	29.705782	148.251337	450.465116
std	25.643823	10.745793	128.708353	93.042578
min	42.000000	15.000000	8.000000	290.000000
25%	66.000000	22.000000	21.000000	380.000000
50%	84.000000	26.000000	126.000000	460.000000
75%	108.000000	37.000000	254.750000	520.000000
max	142.000000	59.000000	520.000000	640.000000


```
[65]: # Importing visualization libraries
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LinearRegression
from sklearn.cluster import KMeans

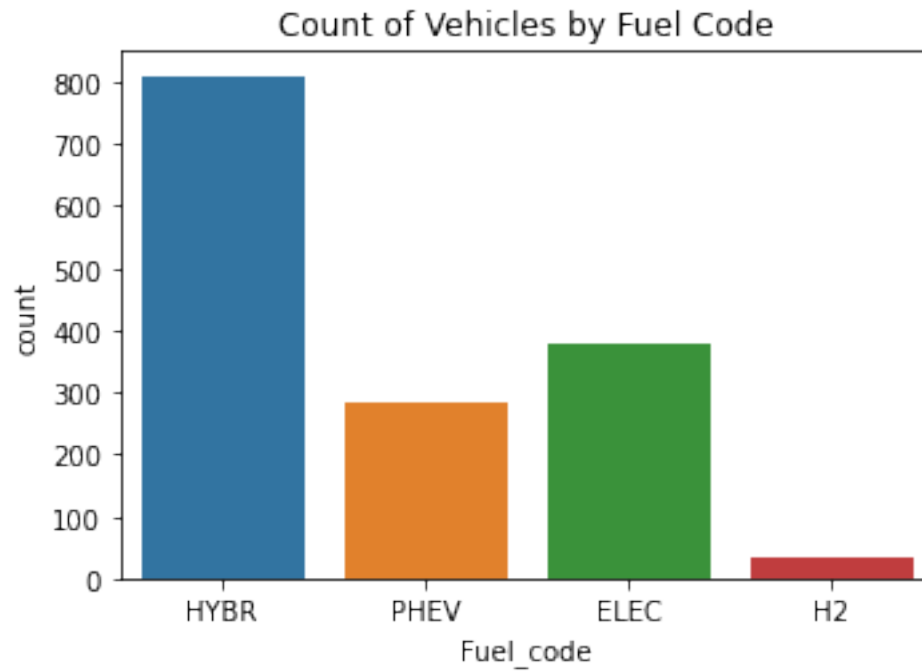
# Visualize the distribution of different variables
plt.figure(figsize=(12, 8))
plt.subplot(2, 2, 1)
sns.histplot(df1['Electric_Range'], kde=True)
plt.title('Distribution of Electric Range')
```

```
[65]: Text(0.5, 1.0, 'Distribution of Electric Range')
```



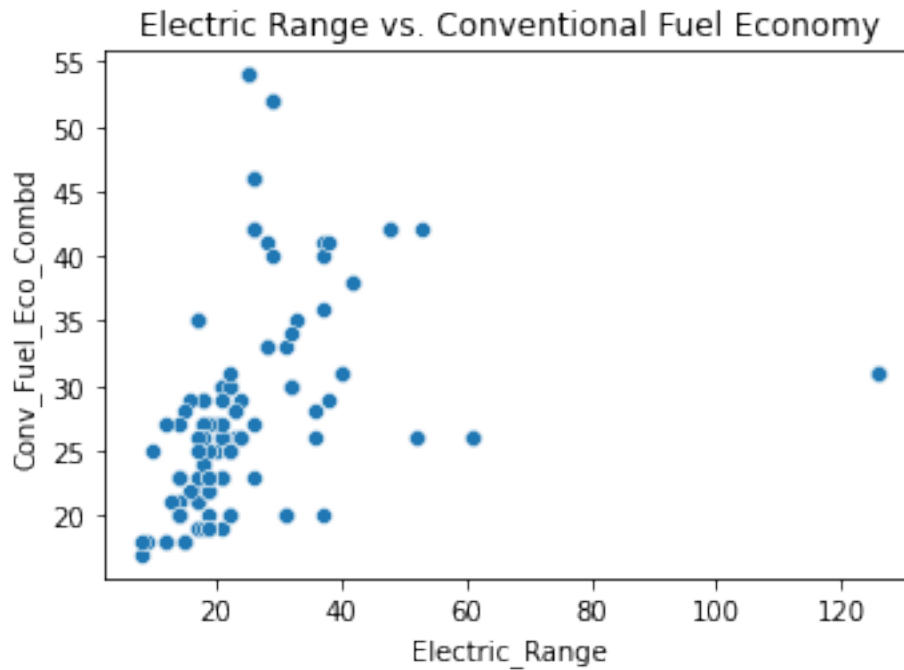
```
[66]: plt.figure(figsize=(12, 8))
plt.subplot(2, 2, 2)
sns.countplot(x='Fuel_code', data=df1)
plt.title('Count of Vehicles by Fuel Code')
```

```
[66]: Text(0.5, 1.0, 'Count of Vehicles by Fuel Code')
```



```
[69]: plt.figure(figsize=(12, 8))  
plt.subplot(2, 2, 4)  
sns.scatterplot(x='Electric_Range', y='Conv_Fuel_Eco_Combd', data=df1)  
plt.title('Electric Range vs. Conventional Fuel Economy')
```

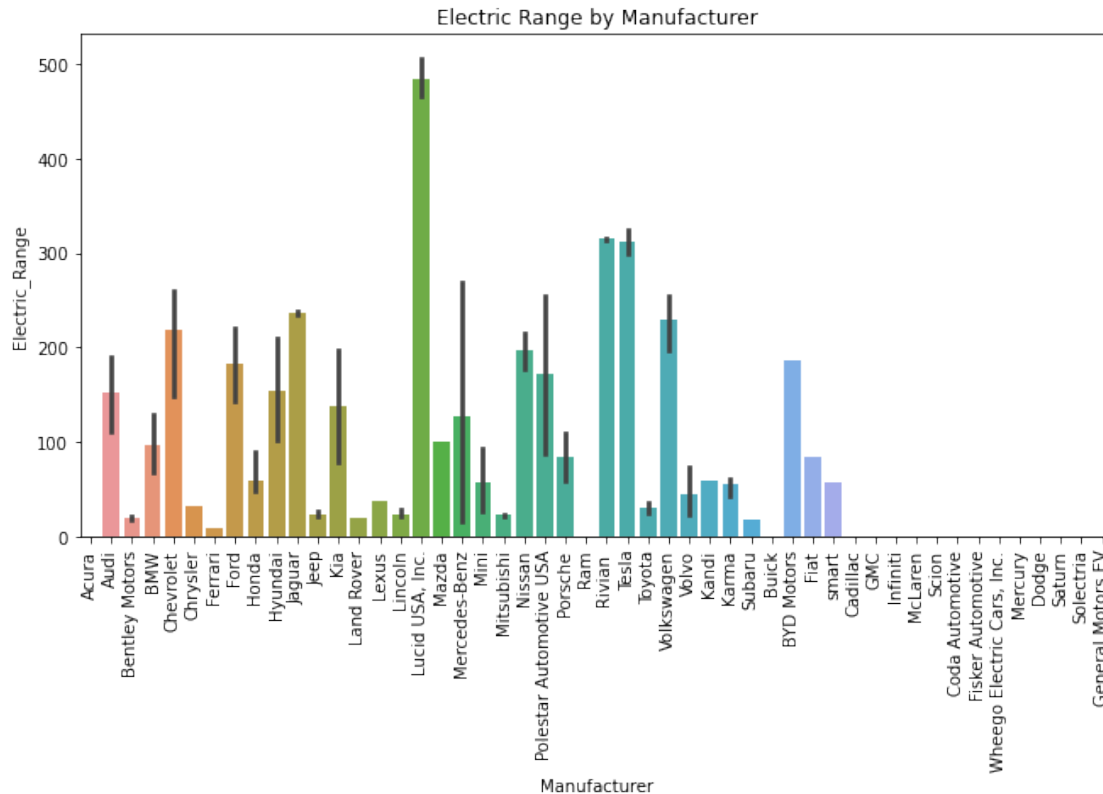
```
[69]: Text(0.5, 1.0, 'Electric Range vs. Conventional Fuel Economy')
```



```
[70]: import warnings
# Suppress all warnings
warnings.filterwarnings("ignore")

plt.figure(figsize=(24, 12))
plt.subplot(2, 2, 4)
sns.barplot(x='Manufacturer', y='Electric_Range', data=df1)
plt.xticks(rotation=90)
plt.title('Electric Range by Manufacturer')
```

```
[70]: Text(0.5, 1.0, 'Electric Range by Manufacturer')
```



```
[76]: # Check the column names
print(df1.columns)
```

```
Index(['Vehicle_ID', 'Fuel_ID', 'Manufacturer_ID', 'Category ID', 'Model',
      'Model_yr', 'Alt_fuel_Eco_Combd', 'Conv_Fuel_Eco_Combd',
      'Transmission_typ', 'Engine_Size', 'Manufacturer', 'Category',
      'Fuel_code', 'Fuel', 'Electric_Range', 'PHEV_range', 'Drivetrain'],
      dtype='object')
```

0.0.2 Regression analysis

```
[77]: X = df1[['Electric_Range', 'PHEV_range', 'Conv_Fuel_Eco_Combd']]
      y = df1['Alt_fuel_Eco_Combd']

      # Fill missing values with 0 for regression
      X = X.fillna(0)
      y = y.fillna(0) # Fill NaN values in the target variable

      model = LinearRegression()
      model.fit(X, y)

      print("Coefficient of determination (R-squared):", model.score(X, y))
```

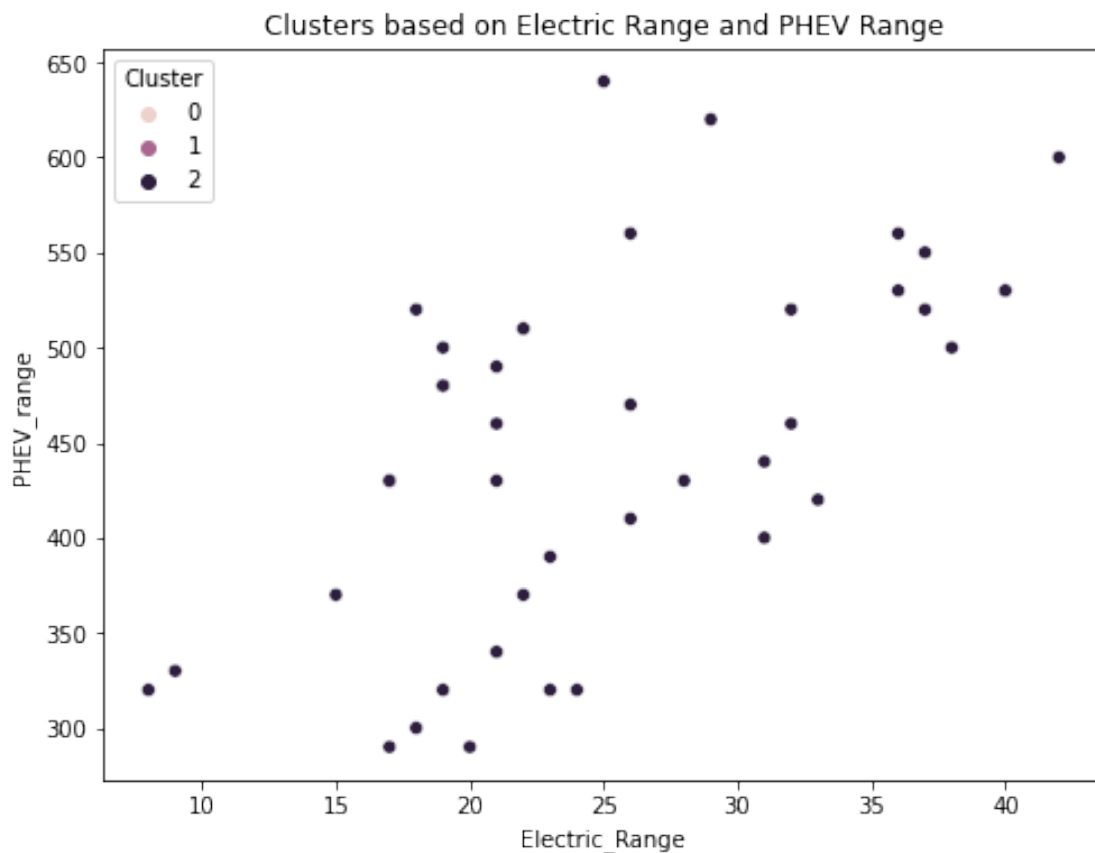
```
print("Coefficients:", model.coef_)
```

Coefficient of determination (R-squared): 0.7041390414337791

Coefficients: [0.36327423 0.1016857 0.34842643]

0.0.3 Cluster analysis

```
[78]: X = df1[['Electric_Range', 'PHEV_range', 'Conv_Fuel_Eco_Combd',  
            ↪ 'Alt_fuel_Eco_Combd']]  
X = X.fillna(0) # Fill missing values with 0 for clustering  
  
kmeans = KMeans(n_clusters=3, random_state=0)  
labels = kmeans.fit_predict(X)  
  
df1['Cluster'] = labels  
  
plt.figure(figsize=(8, 6))  
sns.scatterplot(x='Electric_Range', y='PHEV_range', hue='Cluster', data=df1)  
plt.title('Clusters based on Electric Range and PHEV Range')  
plt.show()  
  
print(df1.groupby('Cluster').mean())
```



	Vehicle_ID	Fuel_ID	Manufacturer_ID	Category ID	Model_yr \
Cluster					
0	12692.051813	41.497409	292.709845	27.683938	2020.927461
1	10469.418110	45.579528	275.151181	27.001575	2016.314173
2	12967.116279	57.000000	327.418605	27.372093	2022.000000

	Alt_fuel_Eco_Combd	Conv_Fuel_Eco_Combd	Electric_Range	PHEV_range
Cluster				
0	100.829016	31.000000	260.409326	NaN
1	71.664286	29.808905	29.768116	NaN
2	68.023256	28.232558	25.093023	450.465116

0.0.4 Dataset 2 : https://en.wikipedia.org/wiki/Electric_car

0.0.5 Description: Overall wiki information of currently available Electric vehicles and the top selling one's across the globe.

```
[80]: # Import required libraries
# pip install requests
import requests
from lxml import etree as et
from bs4 import BeautifulSoup # library to parse HTML documents

# Making a GET request
r = requests.get('https://en.wikipedia.org/wiki/Electric_car')

# check status code for response received
# success code - 200
print(r)
```

<Response [200]>

```
[81]: # parse data from the html into a beautifulsoup object
soup = BeautifulSoup(r.text, 'html.parser')
elec_table=soup.find('table',{'class':"wikitable"})

print(elec_table)
```

```
<table class="wikitable sortable mw-collapsible plainrowheaders"
id="Top_BEV_models">
<caption>All-time top-selling highway-capable<sup>(1)</sup> all-electric
passenger car nameplates
</caption>
<tbody><tr>
<th scope="col">Company
</th>
```

```

<th scope="col">Model
</th>
<th class="unsortable" scope="col">Image
</th>
<th scope="col">Market launch
</th>
<th scope="col">Lifetime global sales
</th>
<th scope="col">Total sales through
</th>
<th scope="col">Annual global sales
</th>
<th>Status
</th>
<th class="unsortable" scope="col"><abbr title="Reference(s)">Ref</abbr>
</th></tr>
<tr align="center">
<td><span typeof="mw:File"><a class="mw-file-description"
href="/wiki/File:Tesla_T_symbol.svg"></a></span><br/> <a
href="/wiki/Tesla,_Inc." title="Tesla, Inc.">Tesla, Inc.</a></td>
<td><a href="/wiki/Tesla_Model_Y" title="Tesla Model Y">Tesla Model Y</a></td>
<td align="left"><span typeof="mw:File"><a class="mw-file-description"
href="/wiki/File:Tesla_Model_Y_1X7A6211.jpg"></a></span></td>
<td>2020-03</td>
<td data-sort-value="1840,000">~2.49 million</td>
<td>2023-12</td>
<td>1,211,601 (2023)
</td>
<td style="background: #D5F5E3">In production
</td>
<td align="left"><sup class="reference" id="cite_ref-GlobalTopEVs2022_175-1"><a
href="#cite_note-GlobalTopEVs2022-175">[175]</a></sup><sup class="reference"
id="cite_ref-GlobalTopEVs2021_176-2"><a href="#cite_note-
GlobalTopEVs2021-176">[176]</a></sup><sup class="reference" id="cite_ref-
GlobalTopEVs062023_199-0"><a href="#cite_note-
GlobalTopEVs062023-199">[199]</a></sup><sup class="reference"

```

^[cite_ref-200][200]</sup>^{[201]}^{[202]}

<div>mw:File</div> <div>#cite_note-200</div> <div>Tesla T symbol.svg</div> <div>254x255px</div> <div>upload.wikimedia.org/wikipedia/commons/thumb/b/bb/Tesla_T_symbol.svg/35px-Tesla_T_symbol.svg.png</div> <div>srcset="//upload.wikimedia.org/wikipedia/commons/thumb/b/bb/Tesla_T_symbol.svg/53px-Tesla_T_symbol.svg.png 1.5x, //upload.wikimedia.org/wikipedia/commons/thumb/b/bb/Tesla_T_symbol.svg/70px-Tesla_T_symbol.svg.png 2x" width="35"/></div> <div>/wiki/Tesla,_Inc.</div> <div>Tesla, Inc.</div>	<div>Tesla Model 3</div> <div>Tesla Model 3</div>
<div>mw:File</div> <div>2019 Tesla Model 3 Performance AWD Front.jpg</div> <div>2143x4685px</div> <div>upload.wikimedia.org/wikipedia/commons/thumb/9/91/2019_Tesla_Model_3_Performance_AWD_Front.jpg/130px-2019_Tesla_Model_3_Performance_AWD_Front.jpg</div> <div>srcset="//upload.wikimedia.org/wikipedia/commons/thumb/9/91/2019_Tesla_Model_3_Performance_AWD_Front.jpg/195px-2019_Tesla_Model_3_Performance_AWD_Front.jpg 1.5x, //upload.wikimedia.org/wikipedia/commons/thumb/9/91/2019_Tesla_Model_3_Performance_AWD_Front.jpg/260px-2019_Tesla_Model_3_Performance_AWD_Front.jpg 2x" width="130"/></div>	<div>2017-07</div> <div>~2.06 million</div> <div>2023-06</div> <div>529,287 (2023)</div>
<div>In production</div>	<div>GlobalTopEVs2022-175-2</div> <div>#cite_note-GlobalTopEVs2022-175</div> <div>[175]</div> <div>GlobalTopEVs2021-176-3</div> <div>#cite_note-GlobalTopEVs2021-176</div> <div>[176]</div> <div>GlobalTopEVs2023-199-1</div> <div>#cite_note-GlobalTopEVs2023-199</div> <div>[199]</div> <div>GlobalTopEVs2022-175-2</div> <div>#cite_note-GlobalTopEVs2022-175</div> <div>[175]</div> <div>GlobalTopEVs2021-176-3</div> <div>#cite_note-GlobalTopEVs2021-176</div> <div>[176]</div> <div>GlobalTopEVs2023-199-1</div> <div>#cite_note-GlobalTopEVs2023-199</div> <div>[199]</div> <div>GlobalTopEVs2022-175-2</div> <div>#cite_note-GlobalTopEVs2022-175</div> <div>[175]</div> <div>GlobalTopEVs2021-176-3</div> <div>#cite_note-GlobalTopEVs2021-176</div> <div>[176]</div> <div>GlobalTopEVs2023-199-1</div> <div>#cite_note-GlobalTopEVs2023-199</div> <div>[199]</div> <div>GlobalTopEVs2022-175-2</div> <div>#cite_note-GlobalTopEVs2022-175</div> <div>[175]</div> <div>GlobalTopEVs2021-176-3</div> <div>#cite_note-GlobalTopEVs2021-176</div> <div>[176]</div> <div>GlobalTopEVs2023-199-1</div> <div>#cite_note-GlobalTopEVs2023-199</div> <div>[199]</div> <div>GlobalTopEVs2022-175-2</div> <div>#cite_note-GlobalTopEVs2022-175</div> <div>[175]</div> <div>GlobalTopEVs2021-176-3</div> <div>#cite_note-GlobalTopEVs2021-176</div> <div>[176]</div> <div>GlobalTopEVs2023-199-1</div> <div>#cite_note-GlobalTopEVs2023-199</div> <div>[199]</div> <div>GlobalTopEVs2022-175-2</div> <div>#cite_note-GlobalTopEVs2022-175</div> <div>[175]</div> <div>GlobalTopEVs2021-176-3</div> <div>#cite_note-GlobalTopEVs2021-176</div> <div>[176]</div> <div>GlobalTopEVs2023-199-1</div> <div>#cite_note-GlobalTopEVs2023-199</div> <div>[199]</div> <div>GlobalTopEVs2022-175-2</div> <div>#cite_note-GlobalTopEVs2022-175</div> <div>[175]</div> <div>GlobalTopEVs2021-176-3</div> <div>#cite_note-GlobalTopEVs2021-176</div> <div>[176]</div> <div>GlobalTopEVs2023-199-1</div> <div>#cite_note-GlobalTopEVs2023-199</div> <div>[199]</div> <div>GlobalTopEVs2022-175-2</div> <div>#cite_note-GlobalTopEVs2022-175</div> <div>[175]</div> <div>GlobalTopEVs2021-176-3</div> <div>#cite_note-GlobalTopEVs2021-176</div> <div>[176]</div> <div>GlobalTopEVs2023-199-1</div> <div>#cite_note-GlobalTopEVs2023-199</div> <div>[199]</div> <div>GlobalTopEVs2022-175-2</div> <div>#cite_note-GlobalTopEVs2022-175</div> <div>[175]</div> <div>GlobalTopEVs2021-176-3</div> <div>#cite_note-GlobalTopEVs2021-176</div> <div>[176]</div> <div>GlobalTopEVs2023-199-1</div> <div>#cite_note-GlobalTopEVs2023-199</div> <div>[199]</div> <div>GlobalTopEVs2022-175-2</div> <div>#cite_note-GlobalTopEVs2022-175</div> <div>[175]</div> <div>GlobalTopEVs2021-176-3</div> <div>#cite_note-GlobalTopEVs2021-176</div> <div>[176]</div> <div>GlobalTopEVs2023-199-1</div> <div>#cite_note-GlobalTopEVs2023-199</div> <div>[199]</div> <div>GlobalTopEVs2022-175-2</div> <div>#cite_note-GlobalTopEVs2022-175</div> <div>[175]</div> <div>GlobalTopEVs2021-176-3</div> <div>#cite_note-GlobalTopEVs2021-176</div> <div>[176]</div> <div>GlobalTopEVs2023-199-1</div> <div>#cite_note-GlobalTopEVs2023-199</div> <div>[199]</div> <div>GlobalTopEVs2022-175-2</div> <div>#cite_note-GlobalTopEVs2022-175</div> <div>[175]</div> <div>GlobalTopEVs2021-176-3</div> <div>#cite_note-GlobalTopEVs2021-176</div> <div>[176]</div> <div>GlobalTopEVs2023-199-1</div> <div>#cite_note-GlobalTopEVs2023-199</div> <div>[199]</div> <div>GlobalTopEVs2022-175-2</div> <div>#cite_note-GlobalTopEVs2022-175</div> <div>[175]</div> <div>GlobalTopEVs2021-176-3</div> <div>#cite_note-GlobalTopEVs2021-176</div> <div>[176]</div> <div>GlobalTopEVs2023-199-1</div> <div>#cite_note-GlobalTopEVs2023-199</div> <div>[199]</div> <div>GlobalTopEVs2022-175-2</div> <div>#cite_note-GlobalTopEVs2022-175</div> <div>[175]</div> <div>GlobalTopEVs2021-176-3</div> <div>#cite_note-GlobalTopEVs2021-176</div> <div>[176]</div> <div>GlobalTopEVs2023-199-1</div> <div>#cite_note-GlobalTopEVs2023-199</div> <div>[199]</div> <div>GlobalTopEVs2022-175-2</div> <div>#cite_note-GlobalTopEVs2022-175</div> <div>[175]</div> <div>GlobalTopEV</div>

srcset="//upload.wikimedia.org/wikipedia/commons/thumb/1/16/Wuling-logo.svg/90px-Wuling-logo.svg.png 1.5x, //upload.wikimedia.org/wikipedia/commons/thumb/1/16/Wuling-logo.svg/120px-Wuling-logo.svg.png 2x" width="60"/>
SAIC-GM-Wuling</td>

<td>Wuling Hongguang Mini EV</td>

<td align="left"></td>

<td>2020-07</td>

<td data-sort-value="1097,872">1,218,640</td>

<td>2023-12</td>

<td>118,834 (2023)</td>

<td style="background: #D5F5E3">In production</td>

<td align="left">^{[175]}^{[176]}^{[199]}^{[204]}^{[205]}</td></tr>

<tr align="center">

<td>
<a href="/wiki/Nissan"

title="Nissan">Nissan</td>

<td>Nissan Leaf

</td>

<td align="left"></td>

<td>2010-12</td>

<td data-sort-value="650,000">~650,000</td>

<td>2023-07</td>

<td>64,201 (2021)

</td>

<td style="background: #D5F5E3">In production

</td>

<td align="left">^{[176]}^{[191]}

</td></tr>

<tr align="center">

<td>
BYD

</td>

<td>BYD Yuan Plus / Atto 3

</td>

<td align="left">

</td>

<td>2022-02

</td>
 <td>614,260
 </td>
 <td>2023-12
 </td>
 <td>412,202 (2023)
 </td>
 <td style="background: #D5F5E3">In production
 </td>
 <td align="left">^{[205]}^{[206]}
 </td></tr>
 <tr align="center">
 <td>
BYD
 </td>
 <td>BYD Dolphin
 </td>
 <td align="left">
 </td>
 <td>2021-08
 </td>
 <td>602,434
 </td>
 <td>2023-12
 </td>
 <td>367,419 (2023)
 </td>
 <td style="background: #D5F5E3">In production
 </td>
 <td align="left">^{[175]}<sup class="reference"

id="cite_ref-GlobalTopEVs062023_199-3">[199]</sup>^{[207]}^{[208]}</td></tr>

| <td>
 GAC Group</td> <td>Aion S</td> <td align="left"></td> <td>2019-05</td> <td>485,369</td> <td>2023-12</td> <td>222,227 (2023)</td> <td style="background: #D5F5E3">In production</td> <td align="left">^{[175]}^{[176]}<sup class="reference" id="cite_ref-GlobalTopEVs062023_199-4"><a href="#cite_note- |

[illegible]

href="/wiki/Volkswagen" title="Volkswagen">Volkswagen

</td>

<td>Volkswagen ID.4

</td>

<td>

</td>

<td>2020-09

</td>

<td>493,219

</td>

<td>2023-12

</td>

<td>192,686 (2023)

</td>

<td style="background: #D5F5E3">In production

</td>

<td align="left">^{[175]}^{[176]}^{[217]}^{[202]}

</td></tr>

<tr align="center">

<td>
 GAC Group

```

</td>
<td><a href="/wiki/Aion_Y" title="Aion Y">Aion Y</a>
</td>
<td align="left"><span typeof="mw:File"><a class="mw-file-description"
href="/wiki/File:2021_GAC_Aion_Y_(front).jpg"></a></span>
</td>
<td>2021-04
</td>
<td>383,350
</td>
<td>2023-12
</td>
<td>229,555 (2023)
</td>
<td style="background: #D5F5E3">In production
</td>
<td align="left"><sup class="reference" id="cite_ref-:7_206-3"><a
href="#cite_note-:7-206">[206]</a></sup><sup class="reference"
id="cite_ref-:8_205-3"><a href="#cite_note-:8-205">[205]</a></sup><sup
class="reference" id="cite_ref-:9_211-1"><a
href="#cite_note-:9-211">[211]</a></sup>
</td></tr>
<tr align="center">
<td><span typeof="mw:File"><a class="mw-file-description"
href="/wiki/File:BYD_Auto_2022_logo.svg"></a></span> <br/><a
href="/wiki/BYD_Auto" title="BYD Auto">BYD</a>
</td>
<td><a href="/wiki/BYD_Han" title="BYD Han">BYD Han EV</a>
</td>
<td align="left"><span typeof="mw:File"><a class="mw-file-description"
href="/wiki/File:2020_BYD_Han_EV_front.jpg"><img class="mw-file-element" data-
file-height="3336" data-file-width="6385" decoding="async" height="68" src="//up
load.wikimedia.org/wikipedia/commons/thumb/b/bd/2020_BYD_Han_EV_front.jpg/130px-
2020_BYD_Han_EV_front.jpg" srcset="//upload.wikimedia.org/wikipedia/commons/thum
b/b/bd/2020_BYD_Han_EV_front.jpg/195px-2020_BYD_Han_EV_front.jpg 1.5x, //upload.

```


wikimedia.org/wikipedia/commons/thumb/b/bd/2020_BYD_Han_EV_front.jpg/260px-2020_BYD_Han_EV_front.jpg 2x" width="130"/>

</td>

<td>2020-03

</td>

<td>367,129

</td>

<td>2023-12

</td>

<td>106,502 (2023)

</td>

<td style="background: #D5F5E3">In production

</td>

<td align="left">^{[206]}^{[205]}^{[211]}^{[212]}

</td></tr>

<tr align="center">

<td>
 Tesla, Inc.</td>

<td>Tesla Model S

</td>

<td align="left"></td>

<td>2012-06</td>

<td data-sort-value="363,900">~363,900</td>

<td>2022-12</td>

<td data-sort-value="35,000">~35,000 (2022)

```

</td>
<td style="background: #D5F5E3">In production
</td>
<td align="left"><sup class="reference" id="cite_ref-TopModelZSW2022_218-0"><a
href="#cite_note-TopModelZSW2022-218">[218]</a></sup>
</td></tr>
<tr align="center">
<td><a href="/wiki/Chery" title="Chery">Chery</a>
</td>
<td><a href="/wiki/Chery_eQ1" title="Chery eQ1">Chery eQ1</a>
</td>
<td align="left"><span typeof="mw:File"><a class="mw-file-description"
href="/wiki/File:2022_Chery_eQ1_(facelift,_front).jpg"></a></span>
</td>
<td>2017-03
</td>
<td>338,051
</td>
<td>2023-12
</td>
<td>29,744 (2023)
</td>
<td style="background: #D5F5E3">In production
</td>
<td align="left"><sup class="reference" id="cite_ref-219"><a
href="#cite_note-219">[219]</a></sup><sup class="reference" id="cite_ref-220"><a
href="#cite_note-220">[220]</a></sup><sup class="reference" id="cite_ref-221"><a
href="#cite_note-221">[221]</a></sup>
</td></tr>
<tr align="center">
<td><span typeof="mw:File"><a class="mw-file-description"
href="/wiki/File:Hyundai_Motor_Company_logo.svg"></a></span> <br><a
href="/wiki/Hyundai_Motor_Company" title="Hyundai Motor Company">Hyundai</a>

```

```

</td>
<td><a class="mw-redirect" href="/wiki/Hyundai_Kona_Electric" title="Hyundai
Kona Electric">Hyundai Kona Electric</a>
</td>
<td align="left"><span typeof="mw:File"><a class="mw-file-description"
href="/wiki/File:Hyundai_Kona_Electric_(SX2)_1X7A1554.jpg"></a></span>
</td>
<td>2018-05
</td>
<td>329,643
</td>
<td>2023-12
</td>
<td>70,871 (2023)
</td>
<td style="background: #D5F5E3">In production
</td>
<td align="left"><sup class="reference" id="cite_ref-:23_222-0"><a
href="#cite_note-:23-222">[222]</a></sup>
</td></tr>
<tr align="center">
<td><span typeof="mw:File"><a class="mw-file-description"
href="/wiki/File:Volkswagen_logo_2019.svg"></a></span><br/> <a
href="/wiki/Volkswagen" title="Volkswagen">Volkswagen</a>
</td>
<td><a href="/wiki/Volkswagen_ID.3" title="Volkswagen ID.3">Volkswagen ID.3</a>
</td>
<td align="left"><span typeof="mw:File"><a class="mw-file-description"
href="/wiki/File:2020_Volkswagen_ID.3_1st_Front.jpg"></a></span>
</td>
<td>2019-11
</td>
<td>325,770
</td>
<td>2023-12
</td>
<td>139,268 (2023)
</td>
<td style="background: #D5F5E3">In production
</td>
<td align="left"><sup class="reference" id="cite_ref-223"><a
href="#cite_note-223">[223]</a></sup><sup class="reference" id="cite_ref-224"><a
href="#cite_note-224">[224]</a></sup><sup class="reference" id="cite_ref-225"><a
href="#cite_note-225">[225]</a></sup><sup class="reference"
id="cite_ref-:11_202-3"><a href="#cite_note-:11-202">[202]</a></sup>
</td></tr>
<tr align="center">
<td><span typeof="mw:File"><a class="mw-file-description"
href="/wiki/File:Hyundai_Motor_Company_logo.svg"></a></span> <br><a
href="/wiki/Hyundai_Motor_Company" title="Hyundai Motor Company">Hyundai</a>
</td>
<td><a href="/wiki/Hyundai_Ioniq_5" title="Hyundai Ioniq 5">Hyundai Ioniq 5</a>
</td>
<td><span typeof="mw:File"><a class="mw-file-description"
href="/wiki/File:Hyundai_Ioniq_5.jpg"></a></span>
</td>
<td>2021-03
</td>
<td>280,430
</td>
<td>2023-12
</td>

```

[illegible]

```
[82]: # Convert Wikipedia Table into a Python Dataframe
df=pd.read_html(str(elec_table))
# convert list to dataframe
df=pd.DataFrame(df[0])

df.head()
```

```
[82]:
```

	Company	Model	Image	Market launch \
0	Tesla, Inc.	Tesla Model Y	NaN	2020-03
1	Tesla, Inc.	Tesla Model 3	NaN	2017-07
2	SAIC-GM-Wuling	Wuling Hongguang Mini EV	NaN	2020-07
3	Nissan	Nissan Leaf	NaN	2010-12
4	BYD	BYD Yuan Plus / Atto 3	NaN	2022-02

	Lifetime global sales	Total sales through	Annual global sales \
0	~2.49 million	2023-12	1,211,601 (2023)
1	~2.06 million	2023-06	529,287 (2023)
2	1218640	2023-12	118,834 (2023)
3	~650,000	2023-07	64,201 (2021)
4	614260	2023-12	412,202 (2023)

	Status	Ref
0	In production	[175] [176] [199] [200] [201] [202]
1	In production	[175] [176] [203] [199] [202]
2	In production	[175] [176] [199] [204] [205]
3	In production	[176] [191]
4	In production	[205] [206]

```
[85]: # Renaming column names to ease dataframe operations
df.rename({'Market launch': 'Market_launch_dt','Lifetime global sales':
↳ 'Lifetime_global_sales','Annual global sales': 'Annual_global_sales','Total_
↳ global sales': 'Total_global_sales', 'Total sales through':
↳ 'Total_sales_through'}, axis=1 , inplace=True)
```

```
[86]: # Display top 5 rows after renaming header
df.head()
```

```
[86]:
```

	Company	Model	Image	Market_launch_dt \
0	Tesla, Inc.	Tesla Model Y	NaN	2020-03
1	Tesla, Inc.	Tesla Model 3	NaN	2017-07
2	SAIC-GM-Wuling	Wuling Hongguang Mini EV	NaN	2020-07
3	Nissan	Nissan Leaf	NaN	2010-12
4	BYD	BYD Yuan Plus / Atto 3	NaN	2022-02

	Lifetime_global_sales	Total_sales_through	Annual_global_sales \
0	~2.49 million	2023-12	1,211,601 (2023)
1	~2.06 million	2023-06	529,287 (2023)

2	1218640	2023-12	118,834 (2023)
3	~650,000	2023-07	64,201 (2021)
4	614260	2023-12	412,202 (2023)

	Status	Ref
0	In production	[175] [176] [199] [200] [201] [202]
1	In production	[175] [176] [203] [199] [202]
2	In production	[175] [176] [199] [204] [205]
3	In production	[176] [191]
4	In production	[205] [206]

```
[87]: # Creating new Dataframe with useful columns only.
vehicle_df = df.drop(['Ref'], axis=1)
```

```
[88]: # Display top 5 rows after dropping REF column
vehicle_df.head()
```

```
[88]:
```

	Company	Model	Image	Market_launch_dt	\
0	Tesla, Inc.	Tesla Model Y	NaN	2020-03	
1	Tesla, Inc.	Tesla Model 3	NaN	2017-07	
2	SAIC-GM-Wuling	Wuling Hongguang Mini EV	NaN	2020-07	
3	Nissan	Nissan Leaf	NaN	2010-12	
4	BYD	BYD Yuan Plus / Atto 3	NaN	2022-02	

	Lifetime_global_sales	Total_sales_through	Annual_global_sales	Status
0	~2.49 million	2023-12	1,211,601 (2023)	In production
1	~2.06 million	2023-06	529,287 (2023)	In production
2	1218640	2023-12	118,834 (2023)	In production
3	~650,000	2023-07	64,201 (2021)	In production
4	614260	2023-12	412,202 (2023)	In production

```
[89]: # check number of missing values in each column.
vehicle_df.isna().sum()
```

```
[89]: Company      0
      Model      0
      Image     18
      Market_launch_dt  0
      Lifetime_global_sales  0
      Total_sales_through  0
      Annual_global_sales  0
      Status      0
      dtype: int64
```

```
[90]: # Image column has majority missing values. So, dropping this column from the
      ↪ dataframe
vehicle_df = vehicle_df.drop(['Image'], axis=1)
```

```
[91]: vehicle_df.head()
```

```
[91]:
```

	Company	Model	Market_launch_dt	\
0	Tesla, Inc.	Tesla Model Y	2020-03	
1	Tesla, Inc.	Tesla Model 3	2017-07	
2	SAIC-GM-Wuling	Wuling Hongguang Mini EV	2020-07	
3	Nissan	Nissan Leaf	2010-12	
4	BYD	BYD Yuan Plus / Atto 3	2022-02	

	Lifetime_global_sales	Total_sales_through	Annual_global_sales	Status
0	~2.49 million	2023-12	1,211,601 (2023)	In production
1	~2.06 million	2023-06	529,287 (2023)	In production
2	1218640	2023-12	118,834 (2023)	In production
3	~650,000	2023-07	64,201 (2021)	In production
4	614260	2023-12	412,202 (2023)	In production

```
[92]: # Model and Company are useful columns. So, need to fix the casing to all_
      ↪uppercase
vehicle_df['Company'] = vehicle_df['Company'].str.upper()
vehicle_df['Model'] = vehicle_df['Model'].str.upper()
```

```
[93]: vehicle_df.head()
```

```
[93]:
```

	Company	Model	Market_launch_dt	\
0	TESLA, INC.	TESLA MODEL Y	2020-03	
1	TESLA, INC.	TESLA MODEL 3	2017-07	
2	SAIC-GM-WULING	WULING HONGGUANG MINI EV	2020-07	
3	NISSAN	NISSAN LEAF	2010-12	
4	BYD	BYD YUAN PLUS / ATTO 3	2022-02	

	Lifetime_global_sales	Total_sales_through	Annual_global_sales	Status
0	~2.49 million	2023-12	1,211,601 (2023)	In production
1	~2.06 million	2023-06	529,287 (2023)	In production
2	1218640	2023-12	118,834 (2023)	In production
3	~650,000	2023-07	64,201 (2021)	In production
4	614260	2023-12	412,202 (2023)	In production

0.0.6 Dataset 1

```
[99]: # Model and Manufacturer are useful columns. So, need to fix the casing to all_
      ↪uppercase
df1['Manufacturer'] = df1['Manufacturer'].str.upper()
df1['Model'] = df1['Model'].str.upper()
```

```
[100]: df1.head()
```



```
[100]:
```

	Vehicle_ID	Fuel_ID	Manufacturer_ID	Category	ID	Model	\
0	13044	45	365	27		NSX	
1	12854	45	377	27		A3	
2	12842	45	377	27		A3 QUATTRO	
3	12783	45	377	27	A4	ALLROAD QUATTRO	
4	12782	45	377	27		A4 QUATTRO	

	Model_yr	Alt_fuel_Eco_Combd	Conv_Fuel_Eco_Combd	Transmission_tpy	\
0	2022	NaN	21.0	auto	
1	2022	NaN	32.0	auto	
2	2022	NaN	31.0	auto	
3	2022	NaN	26.0	auto	
4	2022	NaN	29.0	auto	

	Engine_Size	Manufacturer	Category	Fuel_code	Fuel	\
0	3.5L	ACURA	Sedan/Wagon	HYBR	Hybrid Electric	
1	2.0L	AUDI	Sedan/Wagon	HYBR	Hybrid Electric	
2	2.0L	AUDI	Sedan/Wagon	HYBR	Hybrid Electric	
3	2.0L	AUDI	Sedan/Wagon	HYBR	Hybrid Electric	
4	2.0L	AUDI	Sedan/Wagon	HYBR	Hybrid Electric	

	Electric_Range	PHEV_range	Drivetrain	Cluster
0	NaN	NaN	AWD	1
1	NaN	NaN	FWD	1
2	NaN	NaN	AWD	1
3	NaN	NaN	AWD	1
4	NaN	NaN	AWD	1

```
[106]: # Get distinct Manufacturer names from the "Manufacturer" column
distinct_manuf = df1['Manufacturer'].unique()

# Display the distinct Manufacturer names
print(distinct_manuf)
```

```
['ACURA' 'AUDI' 'BENTLEY' 'BMW' 'CHEVROLET' 'CHRYSLER' 'FERRARI' 'FORD'
 'HONDA' 'HYUNDAI' 'JAGUAR' 'JEEP' 'KIA' 'LAND ROVER' 'LEXUS' 'LINCOLN'
 'LUCID' 'MAZDA' 'MERCEDES-BENZ' 'MINI' 'MITSUBISHI' 'NISSAN' 'POLESTAR'
 'PORSCHE' 'RAM' 'RIVIAN' 'TESLA' 'TOYOTA' 'VOLKSWAGEN' 'VOLVO' 'KANDI'
 'KARMA' 'SUBARU' 'BUICK' 'BYD' 'FIAT' 'SMART' 'CADILLAC' 'GMC' 'INFINITI'
 'MCLAREN' 'SCION' 'CODA AUTOMOTIVE' 'FISKER' 'WHEEGO ELECTRIC CARS, INC.'
 'MERCURY' 'DODGE' 'SATURN' 'SOLECTRIA' 'GENERAL MOTORS EV']
```

```
[104]: # Replace values in the "Company" column
df1['Manufacturer'] = df1['Manufacturer'].replace('BYD MOTORS', 'BYD')
df1['Manufacturer'] = df1['Manufacturer'].replace('BENTLEY MOTORS', 'BENTLEY')
df1['Manufacturer'] = df1['Manufacturer'].replace('LUCID USA, INC.', 'LUCID')
```

```
df1['Manufacturer'] = df1['Manufacturer'].replace('POLESTAR AUTOMOTIVE USA', 'POLESTAR')
df1['Manufacturer'] = df1['Manufacturer'].replace('FISKER AUTOMOTIVE', 'FISKER')
```

0.0.7 Dataset 2

```
[105]: # Get distinct Company names from the "Company" column
distinct_comp = vehicle_df['Company'].unique()

# Display the distinct model names
print(distinct_comp)
```

```
['TESLA' 'SAIC-GM-WULING' 'NISSAN' 'BYD' 'GAC GROUP' 'RENAULT'
 'VOLKSWAGEN' 'CHERY' 'HYUNDAI'
 'NOTES: (1) VEHICLES ARE CONSIDERED HIGHWAY-CAPABLE IF ABLE TO ACHIEVE AT LEAST
 A TOP SPEED OF 100\xa0KM/H (62\xa0MPH).']
```

```
[102]: # Replace values in the "Company" column
vehicle_df['Company'] = vehicle_df['Company'].replace('TESLA, INC.', 'TESLA')
```

0.0.8 Joining Dataset 1 and Dataset 2

```
[110]: # Merge the two DataFrames on the "Manufacturer" and "Company" columns
combined_df = pd.merge(df1, vehicle_df, left_on='Manufacturer',
    right_on='Company', how='inner')

# Display the combined DataFrame
combined_df.head()
```

```
[110]:
```

	Vehicle_ID	Fuel_ID	Manufacturer_ID	Category ID	Model_x \
0	12808	45	351	27	ELANTRA HYBRID
1	12808	45	351	27	ELANTRA HYBRID
2	12809	45	351	27	ELANTRA HYBRID BLUE
3	12809	45	351	27	ELANTRA HYBRID BLUE
4	12810	45	351	27	IONIQ

	Model_yr	Alt_fuel_Eco_Combd	Conv_Fuel_Eco_Combd	Transmission_typ \
0	2022	NaN	50.0	auto
1	2022	NaN	50.0	auto
2	2022	NaN	54.0	auto
3	2022	NaN	54.0	auto
4	2022	NaN	55.0	auto

	Engine_Size	...	PHEV_range	Drivetrain	Cluster	Company \
0	1.6L	...	NaN	FWD	1	HYUNDAI
1	1.6L	...	NaN	FWD	1	HYUNDAI

2	1.6L	...	NaN	FWD	1	HYUNDAI
3	1.6L	...	NaN	FWD	1	HYUNDAI
4	1.6L	...	NaN	FWD	1	HYUNDAI

	Model_y	Market_launch_dt	Lifetime_global_sales	\
0	HYUNDAI KONA ELECTRIC	2018-05	329643	
1	HYUNDAI IONIQ 5	2021-03	280430	
2	HYUNDAI KONA ELECTRIC	2018-05	329643	
3	HYUNDAI IONIQ 5	2021-03	280430	
4	HYUNDAI KONA ELECTRIC	2018-05	329643	

	Total_sales_through	Annual_global_sales	Status
0	2023-12	70,871 (2023)	In production
1	2023-12	114,988 (2023)	In production
2	2023-12	70,871 (2023)	In production
3	2023-12	114,988 (2023)	In production
4	2023-12	70,871 (2023)	In production

[5 rows x 25 columns]

```
[116]: #dropping this column from the dataframe
combined_df = combined_df.
↳drop(['Vehicle_ID', 'Fuel_ID', 'Manufacturer_ID', 'Category_
↳ID', 'Model_x', 'Company'], axis=1)
```

```
[117]: # Display the combined DataFrame
combined_df.head()
```

```
[117]: Model_yr  Alt_fuel_Eco_Combd  Conv_Fuel_Eco_Combd  Transmission_typ  \
0      2022                NaN                50.0            auto
1      2022                NaN                50.0            auto
2      2022                NaN                54.0            auto
3      2022                NaN                54.0            auto
4      2022                NaN                55.0            auto
```

	Engine_Size	Manufacturer	Category	Fuel_code	Fuel	\
0	1.6L	HYUNDAI	Sedan/Wagon	HYBR	Hybrid Electric	
1	1.6L	HYUNDAI	Sedan/Wagon	HYBR	Hybrid Electric	
2	1.6L	HYUNDAI	Sedan/Wagon	HYBR	Hybrid Electric	
3	1.6L	HYUNDAI	Sedan/Wagon	HYBR	Hybrid Electric	
4	1.6L	HYUNDAI	Sedan/Wagon	HYBR	Hybrid Electric	

	Electric_Range	PHEV_range	Drivetrain	Cluster	Model_y	\
0	NaN	NaN	FWD	1	HYUNDAI KONA ELECTRIC	
1	NaN	NaN	FWD	1	HYUNDAI IONIQ 5	
2	NaN	NaN	FWD	1	HYUNDAI KONA ELECTRIC	
3	NaN	NaN	FWD	1	HYUNDAI IONIQ 5	

4	NaN	NaN	FWD	1	HYUNDAI KONA ELECTRIC
---	-----	-----	-----	---	-----------------------

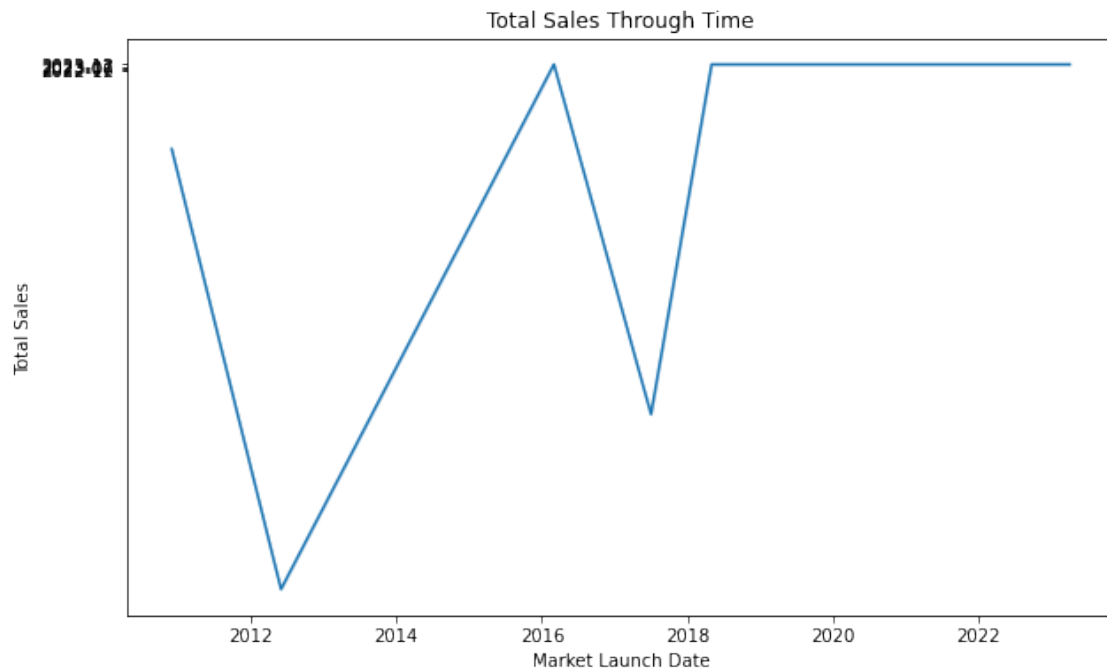
	Market_launch_dt	Lifetime_global_sales	Total_sales_through	\
0	2018-05	329643	2023-12	
1	2021-03	280430	2023-12	
2	2018-05	329643	2023-12	
3	2021-03	280430	2023-12	
4	2018-05	329643	2023-12	

	Annual_global_sales	Status
0	70,871 (2023)	In production
1	114,988 (2023)	In production
2	70,871 (2023)	In production
3	114,988 (2023)	In production
4	70,871 (2023)	In production

0.0.9 Time Series Analysis

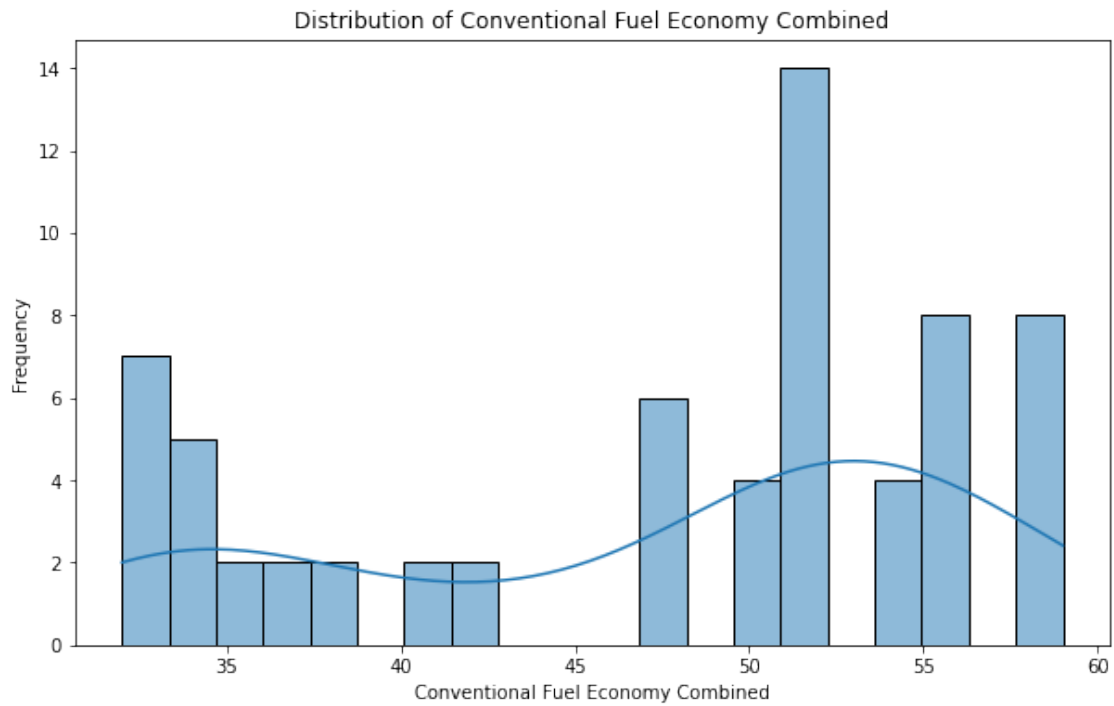
```
[125]: # Convert 'Market_launch_dt' column to datetime
combined_df['Market_launch_dt'] = pd.
    ↳to_datetime(combined_df['Market_launch_dt'])

# Time Series Analysis
plt.figure(figsize=(10, 6))
sns.lineplot(x='Market_launch_dt', y='Total_sales_through', data=combined_df,
    ↳estimator='sum')
plt.title('Total Sales Through Time')
plt.xlabel('Market Launch Date')
plt.ylabel('Total Sales')
plt.show()
```



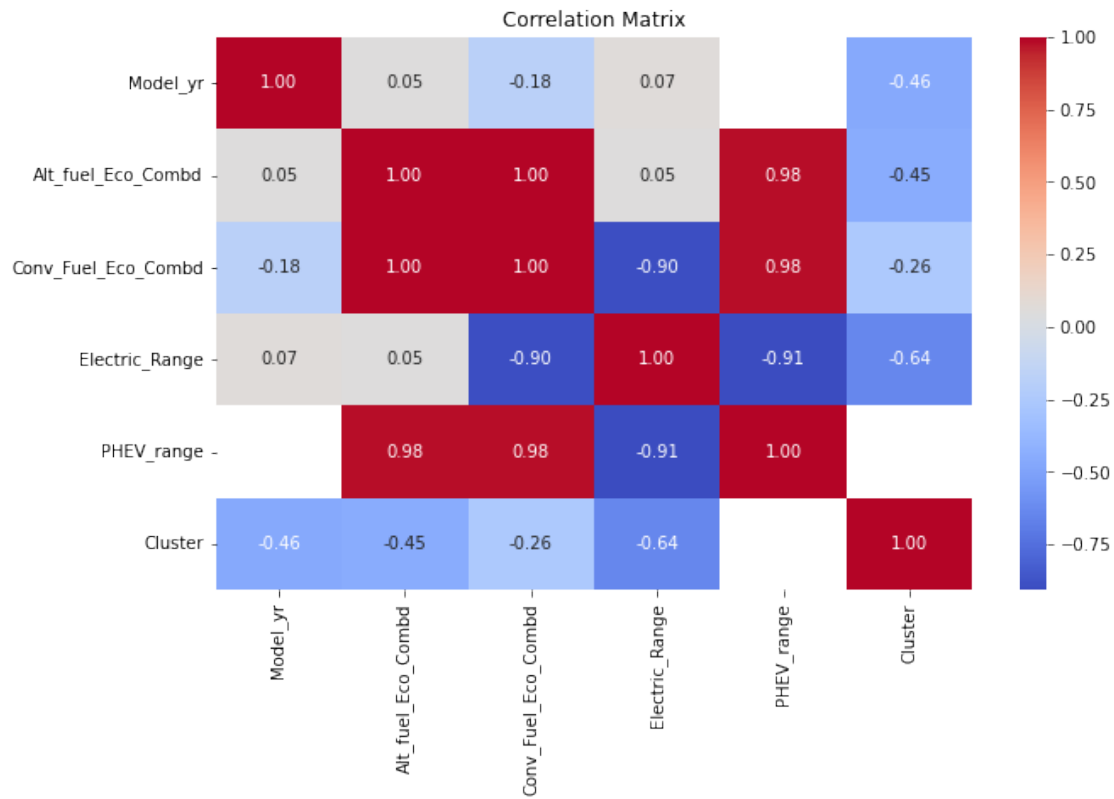
0.0.10 Distribution Analysis

```
[126]: plt.figure(figsize=(10, 6))
sns.histplot(combined_df['Conv_Fuel_Eco_Combd'].dropna(), bins=20, kde=True)
plt.title('Distribution of Conventional Fuel Economy Combined')
plt.xlabel('Conventional Fuel Economy Combined')
plt.ylabel('Frequency')
plt.show()
```



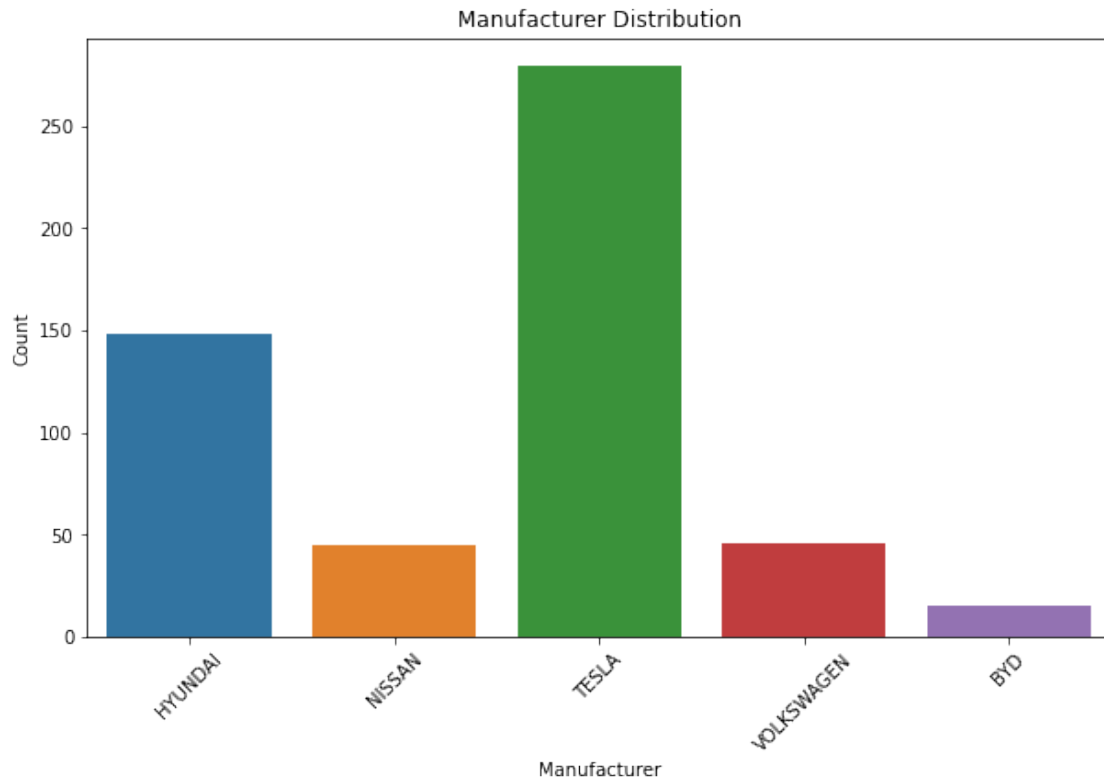
0.0.11 Correlation Analysis

```
[128]: plt.figure(figsize=(10, 6))
sns.heatmap(combined_df.corr(), annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Matrix')
plt.show()
```



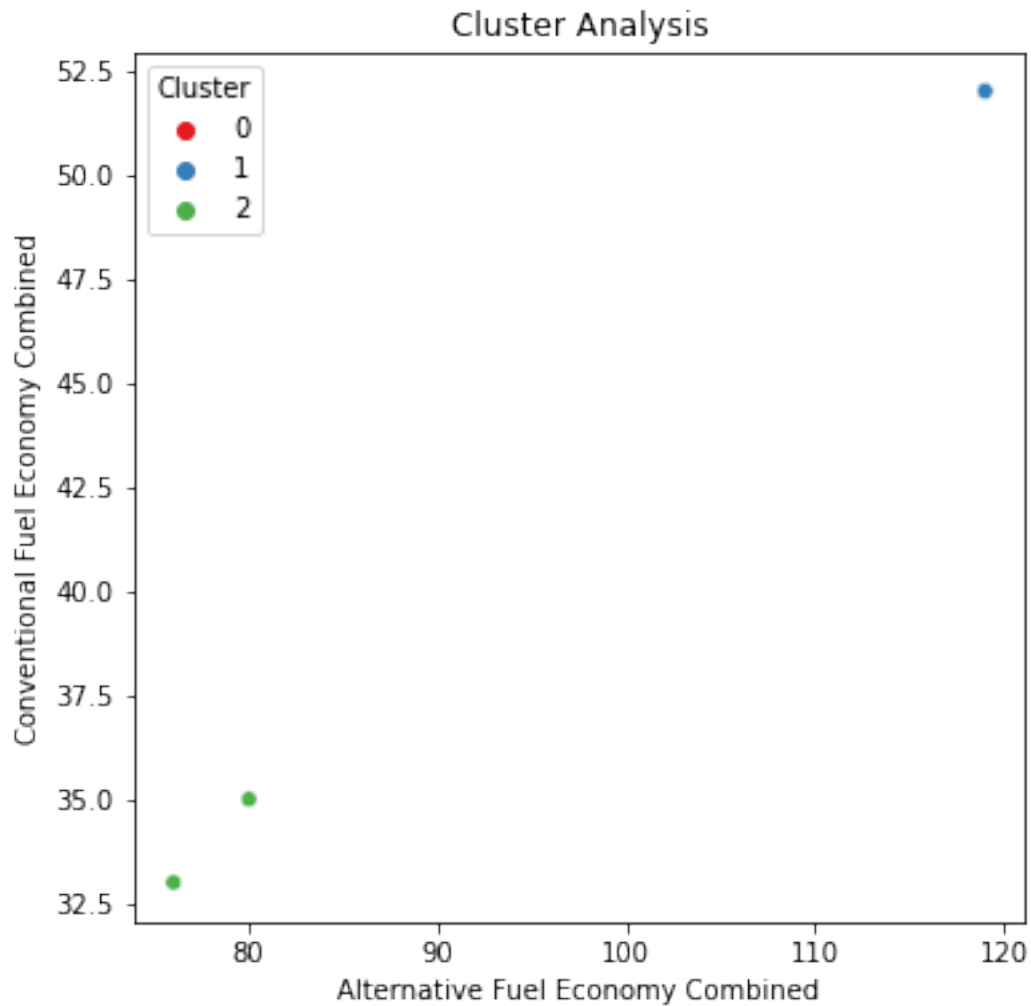
0.0.12 Categorical Analysis

```
[129]: plt.figure(figsize=(10, 6))
sns.countplot(x='Manufacturer', data=combined_df)
plt.title('Manufacturer Distribution')
plt.xlabel('Manufacturer')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.show()
```



0.0.13 Cluster Analysis

```
[131]: plt.figure(figsize=(6, 6))
sns.scatterplot(x='Alt_fuel_Eco_Combd', y='Conv_Fuel_Eco_Combd', hue='Cluster',
               data=combined_df, palette='Set1')
plt.title('Cluster Analysis')
plt.xlabel('Alternative Fuel Economy Combined')
plt.ylabel('Conventional Fuel Economy Combined')
plt.show()
```

0.0.14 Regression analysis

```
[132]: # Regression analysis on combined data
X = combined_df[['Electric_Range', 'PHEV_range', 'Conv_Fuel_Eco_Combd']]
y = combined_df['Alt_fuel_Eco_Combd']

# Fill missing values with 0 for regression
X = X.fillna(0)
y = y.fillna(0) # Fill NaN values in the target variable

model = LinearRegression()
model.fit(X, y)

print("Coefficient of determination (R-squared):", model.score(X, y))
print("Coefficients:", model.coef_)
```

Coefficient of determination (R-squared): 0.827892996525533
Coefficients: [0.34289337 0.14395417 0.07725874]

0.0.15 Cluster analysis

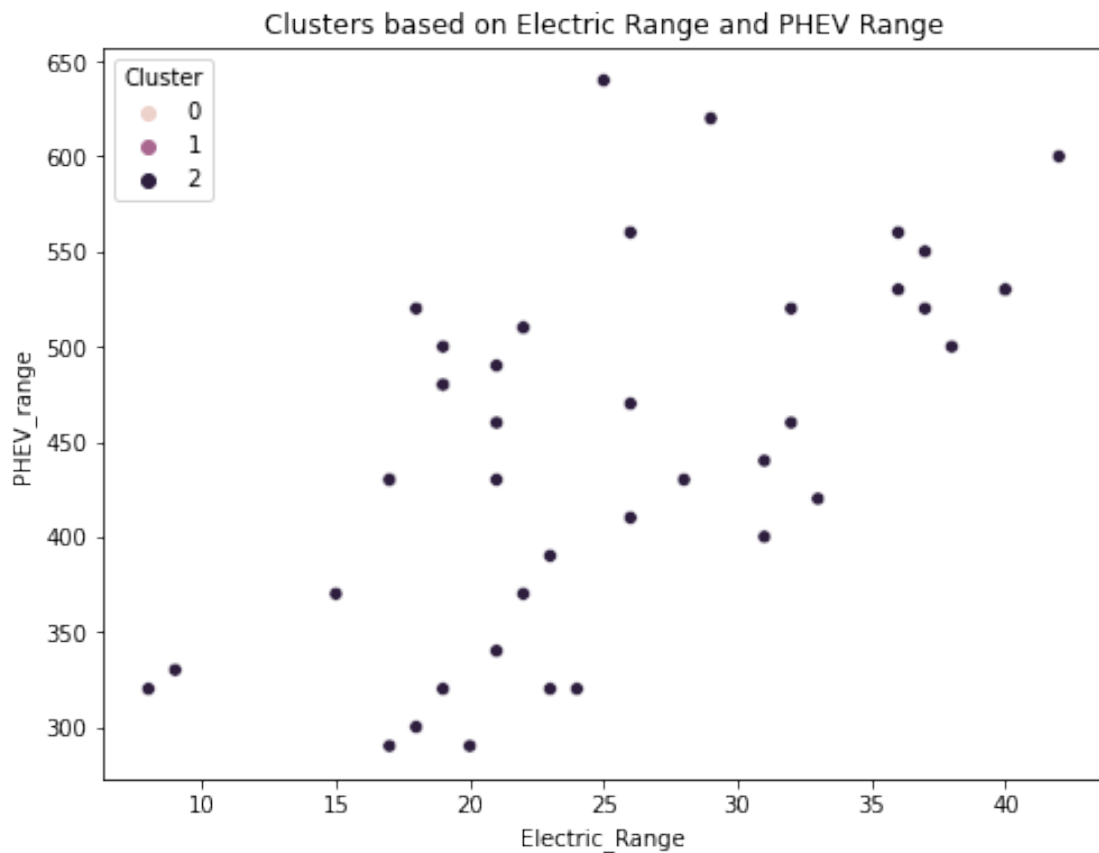
```
[133]: # Cluster analysis on combined data
X = combined_df[['Electric_Range', 'PHEV_range', 'Conv_Fuel_Eco_Combd', 'Alt_fuel_Eco_Combd']]
X = X.fillna(0) # Fill missing values with 0 for clustering

kmeans = KMeans(n_clusters=3, random_state=0)
labels = kmeans.fit_predict(X)

combined_df['Cluster'] = labels

plt.figure(figsize=(8, 6))
sns.scatterplot(x='Electric_Range', y='PHEV_range', hue='Cluster', data=df1)
plt.title('Clusters based on Electric Range and PHEV Range')
plt.show()

print(df1.groupby('Cluster').mean())
```



	Vehicle_ID	Fuel_ID	Manufacturer_ID	Category ID	Model_yr \
Cluster					
0	12692.051813	41.497409	292.709845	27.683938	2020.927461
1	10469.418110	45.579528	275.151181	27.001575	2016.314173
2	12967.116279	57.000000	327.418605	27.372093	2022.000000

	Alt_fuel_Eco_Combd	Conv_Fuel_Eco_Combd	Electric_Range	PHEV_range
Cluster				
0	100.829016	31.000000	260.409326	NaN
1	71.664286	29.808905	29.768116	NaN
2	68.023256	28.232558	25.093023	450.465116

[]: