Rohan Chauhan

Project Topic :-- Electric Vehicle Population Data

```
#Importing the necessaray libraries

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Reading a File
#Loading  data into data frame
df = pd.read_csv("Electric Vehicle Population  Data.csv")
```

This file conatins a 16 columns and 138779 rows. Here the columns are :- 1.VIN:- Vechicle Identification Number 2.County 3.City 4.State 5.Postal Code 6.Model Year 7.Make 8.Model 9.Electric Vehicle Type 10.Electric Range 11.Base MSRP 12.Legislative District 13.DOL Vehicle ID 14.Vehicle Location 15.Electric Utility 16.2020 Census Tract

```
# To display the top  5 rows
print("first five Data are here:-")
df.head(5)
```

```
first five Data are here:-

   VIN (1-10)     County          City State  Postal Code  Model Year
Make  \
0  1N4AZ0CP5D     Kitsap     Bremerton    WA      98310.0        2013
NISSAN
1  1N4AZ1CP8K     Kitsap  Port Orchard    WA      98366.0        2019
NISSAN
2  5YJXCAE28L       King       Seattle    WA      98199.0        2020
TESLA
3  SADHC2S1XK   Thurston       Olympia    WA      98503.0        2019
JAGUAR
4  JN1AZ0CP9B  Snohomish       Everett    WA      98204.0        2011
NISSAN

     Model           Electric Vehicle Type  Electric Range  Base MSRP
\
0     LEAF  Battery Electric Vehicle (BEV)            75.0        0.0

1     LEAF  Battery Electric Vehicle (BEV)           150.0        0.0

2  MODEL X  Battery Electric Vehicle (BEV)           293.0        0.0

3   I-PACE  Battery Electric Vehicle (BEV)           234.0        0.0

4     LEAF  Battery Electric Vehicle (BEV)            73.0        0.0
```

```
   Legislative District  DOL Vehicle ID  \
0                  23.0     214384901.0
1                  26.0     271008636.0
2                  36.0       8781552.0
3                   2.0       8308492.0
4                  21.0     245524527.0

                                Vehicle Location  \
0  POINT (-122.61136499999998 47.575195000000065)
1   POINT (-122.63926499999997 47.53730000000007)
2            POINT (-122.394185 47.63919500000003)
3                      POINT (-122.8285 47.03646)
4   POINT (-122.24128499999995 47.91088000000008)

                             Electric Utility  2020 Census Tract
0                      PUGET SOUND ENERGY INC       5.303508e+10
1                      PUGET SOUND ENERGY INC       5.303509e+10
2  CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)       5.303301e+10
3                      PUGET SOUND ENERGY INC       5.306701e+10
4                      PUGET SOUND ENERGY INC       5.306104e+10
```

```
# To  display the bottom  5 rows
df.tail()
```

```
       VIN (1-10)    County      City State  Postal Code  Model
Year  \
50588  5YJSA1E49K  Kittitas  Cle Elum    WA      98922.0        2019

50589  7SAYGDEE3P     Clark     Camas    WA      98607.0        2023

50590  1G1FX6S06H      King   Seattle    WA      98103.0        2017

50591  5YJSA1E43H    Pierce       Roy    WA      98580.0        2017

50592  1G1FW6S08H  Kittitas  Cle Elum    WA      98922.0        2017


            Make    Model           Electric Vehicle Type  Electric
Range  \
50588      TESLA  MODEL S  Battery Electric Vehicle (BEV)
270.0
50589      TESLA  MODEL Y  Battery Electric Vehicle (BEV)
0.0
50590  CHEVROLET  BOLT EV  Battery Electric Vehicle (BEV)
238.0
50591      TESLA  MODEL S  Battery Electric Vehicle (BEV)
210.0
50592  CHEVROLET   BOLT E                              NaN
NaN
```

```
       Base MSRP  Legislative District  DOL Vehicle ID  \
50588        0.0                  13.0     447161816.0
50589        0.0                  18.0     240763603.0
50590        0.0                  43.0     156970089.0
50591        0.0                   2.0     148554024.0
50592        NaN                   NaN             NaN

                                    Vehicle Location  \
50588   POINT (-120.93829499999998 47.19538000000006)
50589   POINT (-122.40556499999997 45.59009000000003)
50590  POINT (-122.34300999999999 47.659185000000036)
50591   POINT (-122.52298499999995 46.98760000000004)
50592                                             NaN

                                    Electric Utility  2020 Census
Tract
50588                              PUGET SOUND ENERGY INC
5.303798e+10
50589  BONNEVILLE POWER ADMINISTRATION||PUD NO 1 OF C...
5.301104e+10
50590       CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)
5.303300e+10
50591  BONNEVILLE POWER ADMINISTRATION||CITY OF TACOM...
5.305307e+10
50592                                                NaN
NaN

# To display random  10 rows
df.sample(10)

       VIN (1-10)     County               City State  Postal Code  \
17660  KNDC4DLC4P       King            Seattle    WA      98109.0
17042  5YJSA1E26H       King            Seattle    WA      98101.0
29325  1N4AZ0CP6F       King            Seattle    WA      98107.0
45256  5YJ3E1EC8N       King             Auburn    WA      98092.0
49457  1N4AZ0CP0F     Kitsap  Bainbridge Island    WA      98110.0
20162  5YJYGDEE7M     Pierce             Tacoma    WA      98405.0
21892  KM8JFDA29N      Clark          Vancouver    WA      98685.0
39485  5YJ3E1EBXN    Douglas     East Wenatchee    WA      98802.0
5178   1N4AZ0CP4E  Snohomish            Everett    WA      98204.0
6      5YJ3E1EB7K    Spokane            Spokane    WA      99203.0

       Model Year     Make    Model                     Electric Vehicle
Type  \
17660        2023      KIA      EV6          Battery Electric Vehicle
(BEV)
17042        2017    TESLA  MODEL S          Battery Electric Vehicle
(BEV)
29325        2015   NISSAN     LEAF          Battery Electric Vehicle
```

```
(BEV)
45256        2022    TESLA  MODEL 3          Battery Electric Vehicle
(BEV)
49457        2015   NISSAN     LEAF          Battery Electric Vehicle
(BEV)
20162        2021    TESLA  MODEL Y          Battery Electric Vehicle
(BEV)
21892        2022  HYUNDAI   TUCSON  Plug-in Hybrid Electric Vehicle
(PHEV)
39485        2022    TESLA  MODEL 3          Battery Electric Vehicle
(BEV)
5178         2014   NISSAN     LEAF          Battery Electric Vehicle
(BEV)
6            2019    TESLA  MODEL 3          Battery Electric Vehicle
(BEV)

       Electric Range  Base MSRP  Legislative District  DOL Vehicle ID
\
17660             0.0        0.0                  36.0     239704685.0

17042           210.0        0.0                  43.0     220620732.0

29325            84.0        0.0                  36.0     148012406.0

45256             0.0        0.0                  31.0     185616196.0

49457            84.0        0.0                  23.0     160022418.0

20162             0.0        0.0                  27.0     193568056.0

21892            33.0        0.0                  18.0     196100712.0

39485             0.0        0.0                  12.0     212311466.0

5178             84.0        0.0                  21.0     349945837.0

6               220.0        0.0                   6.0     241573384.0


                                      Vehicle Location  \
17660            POINT (-122.34848 47.63240500000006)
17042  POINT (-122.33534499999996 47.610790000000065)
29325                   POINT (-122.378895 47.66905)
45256                POINT (-122.1820969 47.3198995)
49457  POINT (-122.52357809999995 47.629332300000044)
20162                   POINT (-122.45153 47.251135)
21892   POINT (-122.70301999999998 45.70370600000007)
39485   POINT (-120.28673999999995 47.41760000000005)
5178    POINT (-122.24128499999995 47.91088000000008)
6      POINT (-117.42526499999997 47.635365000000036)
```

```
                                          Electric Utility  2020 Census
Tract
17660       CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)
5.303301e+10
17042       CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)
5.303301e+10
29325       CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)
5.303300e+10
45256      PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)
5.303303e+10
49457                              PUGET SOUND ENERGY INC
5.303509e+10
20162  BONNEVILLE POWER ADMINISTRATION||CITY OF TACOM...
5.305306e+10
21892  BONNEVILLE POWER ADMINISTRATION||PUD NO 1 OF C...
5.301104e+10
39485                            PUD NO 1 OF DOUGLAS COUNTY
5.301795e+10
5178                               PUGET SOUND ENERGY INC
5.306104e+10
6                            MODERN ELECTRIC WATER COMPANY
5.306300e+10
```

```
# df.dtypes check  the datatypes of columns
# returns a series with the data type of each column
df.dtypes
```

```
VIN (1-10)                 object
County                     object
City                       object
State                      object
Postal Code               float64
Model Year                  int64
Make                       object
Model                      object
Electric Vehicle Type      object
Electric Range            float64
Base MSRP                 float64
Legislative District      float64
DOL Vehicle ID            float64
Vehicle Location           object
Electric Utility           object
2020 Census Tract         float64
dtype: object
```

```
#df.info() shows the information  about the  data frame
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50593 entries, 0 to 50592
Data columns (total 16 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   VIN (1-10)             50593 non-null  object
 1   County                 50590 non-null  object
 2   City                   50590 non-null  object
 3   State                  50593 non-null  object
 4   Postal Code            50590 non-null  float64
 5   Model Year             50593 non-null  int64
 6   Make                   50593 non-null  object
 7   Model                  50494 non-null  object
 8   Electric Vehicle Type  50592 non-null  object
 9   Electric Range         50592 non-null  float64
 10  Base MSRP              50592 non-null  float64
 11  Legislative District   50472 non-null  float64
 12  DOL Vehicle ID         50592 non-null  float64
 13  Vehicle Location       50588 non-null  object
 14  Electric Utility       50589 non-null  object
 15  2020 Census Tract      50589 non-null  float64
dtypes: float64(6), int64(1), object(9)
memory usage: 6.2+ MB
```

The above code df.info() shows the information about data frame . it can shows the class as pandas. 2.RangeIndex(rows)-138779 and columns-16 3.displaying the name of columns and their Dtypes. in above data the float datatypes repeated 3 times,int datatypes repeated 4 times . and others 9 are objects.

```
#Dropping Irrelevant Columns
df=df.drop(['County',
            'DOL Vehicle ID',
            'Legislative District',
            'State',
            'Base MSRP',
            '2020 Census Tract',
            'Vehicle Location',
            'Electric Utility'],axis=1)
df.head(5)
```

	VIN (1-10)	City	Postal Code	Model Year	Make	Model
0	1N4AZ0CP5D	Bremerton	98310.0	2013	NISSAN	LEAF
1	1N4AZ1CP8K	Port Orchard	98366.0	2019	NISSAN	LEAF
2	5YJXCAE28L	Seattle	98199.0	2020	TESLA	MODEL X
3	SADHC2S1XK	Olympia	98503.0	2019	JAGUAR	I-PACE

```
4  JN1AZ0CP9B       Everett      98204.0        2011  NISSAN     LEAF


            Electric Vehicle Type  Electric Range
0  Battery Electric Vehicle (BEV)            75.0
1  Battery Electric Vehicle (BEV)           150.0
2  Battery Electric Vehicle (BEV)           293.0
3  Battery Electric Vehicle (BEV)           234.0
4  Battery Electric Vehicle (BEV)            73.0
```

```
#Renaming The columns
df=df.rename(columns={"VIN (1-10)":"Vechicle ID",
                      "Make":"EcarBrand",
                       "Model":"Car Model"})
df.head(5)
```

```
  Vechicle ID          City  Postal Code  Model Year EcarBrand Car
Model  \
0  1N4AZ0CP5D     Bremerton      98310.0        2013    NISSAN
LEAF
1  1N4AZ1CP8K  Port Orchard      98366.0        2019    NISSAN
LEAF
2  5YJXCAE28L       Seattle      98199.0        2020     TESLA   MODEL
X
3  SADHC2S1XK       Olympia      98503.0        2019    JAGUAR    I-
PACE
4  JN1AZ0CP9B       Everett      98204.0        2011    NISSAN
LEAF
```

```
            Electric Vehicle Type  Electric Range
0  Battery Electric Vehicle (BEV)            75.0
1  Battery Electric Vehicle (BEV)           150.0
2  Battery Electric Vehicle (BEV)           293.0
3  Battery Electric Vehicle (BEV)           234.0
4  Battery Electric Vehicle (BEV)            73.0
```

```
#df.shape display the total number of entries(rows) and total number
of columns
df.shape
```

```
(50593, 8)
```

```
#it is use to print the duplicates
df[df.duplicated()]
```

```
      Vechicle ID          City  Postal Code  Model Year EcarBrand
Car Model  \
171    5YJ3E1EA6J       Bothell      98012.0        2018     TESLA
MODEL 3
330    1N4AZ1CP7J       Kenmore      98028.0        2018    NISSAN
```

```
LEAF
547    2C4RC1L74P        Tukwila      98188.0        2023  CHRYSLER
PACIFICA
619    7SAYGDEF8P       Bellevue      98006.0        2023     TESLA
MODEL Y
620    5YJ3E1EA6N         Seatac      98148.0        2022     TESLA
MODEL 3
...           ...            ...          ...         ...       ...
...
50575  5YJYGDEE2M      Sammamish      98074.0        2021     TESLA
MODEL Y
50578  5YJ3E1EBXJ  Mercer Island      98040.0        2018     TESLA
MODEL 3
50580  7SAYGDEE0P       Bellevue      98008.0        2023     TESLA
MODEL Y
50587  7SAYGDEE5N   Maple Valley      98038.0        2022     TESLA
MODEL Y
50589  7SAYGDEE3P          Camas      98607.0        2023     TESLA
MODEL Y
```

	Electric Vehicle Type	Electric Range
171	Battery Electric Vehicle (BEV)	215.0
330	Battery Electric Vehicle (BEV)	151.0
547	Plug-in Hybrid Electric Vehicle (PHEV)	32.0
619	Battery Electric Vehicle (BEV)	0.0
620	Battery Electric Vehicle (BEV)	0.0
...	...	...
50575	Battery Electric Vehicle (BEV)	0.0
50578	Battery Electric Vehicle (BEV)	215.0
50580	Battery Electric Vehicle (BEV)	0.0
50587	Battery Electric Vehicle (BEV)	0.0
50589	Battery Electric Vehicle (BEV)	0.0

```
[6798 rows x 8 columns]
```

```
#it prints the total number of duplicate  entries and total number of
columns
duplicate_rows_df=df[df.duplicated()]
print("NUmber of Duplicate Rows : ",duplicate_rows_df.shape)
```

```
NUmber of Duplicate Rows :  (6798, 8)
```

```
#use to count the number of rows
df.count()
```

```
Vechicle ID              50593
City                     50590
Postal Code              50590
Model Year               50593
EcarBrand                50593
```

```
Car Model                 50494
Electric Vehicle Type     50592
Electric Range            50592
dtype: int64
```

```
#Use to  drop the duplicates from the data .
df=df.drop_duplicates()
df.head(5)
```

```
  Vechicle ID          City  Postal Code  Model Year EcarBrand Car
Model  \
0  1N4AZ0CP5D     Bremerton      98310.0        2013    NISSAN
LEAF
1  1N4AZ1CP8K  Port Orchard      98366.0        2019    NISSAN
LEAF
2  5YJXCAE28L       Seattle      98199.0        2020     TESLA   MODEL
X
3  SADHC2S1XK       Olympia      98503.0        2019    JAGUAR    I-
PACE
4  JN1AZ0CP9B       Everett      98204.0        2011    NISSAN
LEAF
```

```
            Electric Vehicle Type  Electric Range
0  Battery Electric Vehicle (BEV)            75.0
1  Battery Electric Vehicle (BEV)           150.0
2  Battery Electric Vehicle (BEV)           293.0
3  Battery Electric Vehicle (BEV)           234.0
4  Battery Electric Vehicle (BEV)            73.0
```

```
# Use to counts number of roes after dropping
df.count()
```

```
Vechicle ID               43795
City                      43792
Postal Code               43792
Model Year                43795
EcarBrand                 43795
Car Model                 43697
Electric Vehicle Type     43794
Electric Range            43794
dtype: int64
```

```
#print(df.isnull().sum()) returns the number of missing values  from
the data
print(df.isnull().sum())
```

```
Vechicle ID               0
City                      3
Postal Code               3
Model Year                0
EcarBrand                 0
```

```
Car Model                  98
Electric Vehicle Type       1
Electric Range              1
dtype: int64
```

```
#Dropping the missing values
df=df.dropna()
df.count()
```

```
Vechicle ID               43693
City                      43693
Postal Code               43693
Model Year                43693
EcarBrand                 43693
Car Model                 43693
Electric Vehicle Type     43693
Electric Range            43693
dtype: int64
```

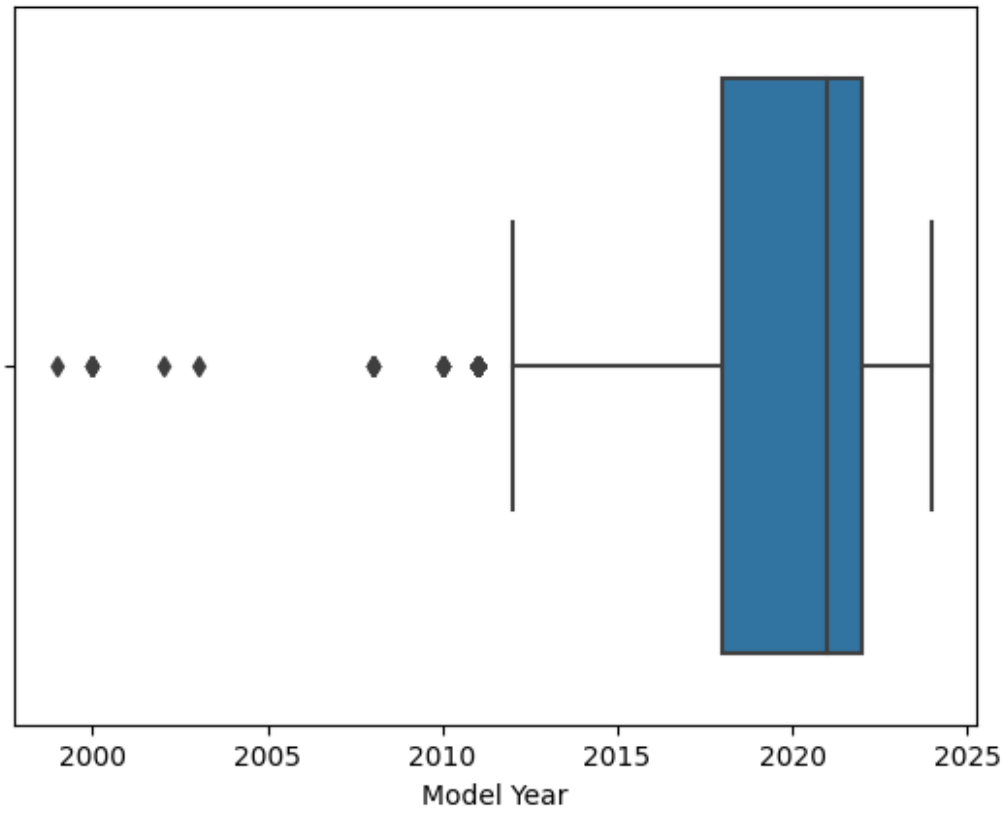
```
# After dropping values
print(df.isnull().sum())
```

```
Vechicle ID               0
City                      0
Postal Code               0
Model Year                0
EcarBrand                 0
Car Model                 0
Electric Vehicle Type     0
Electric Range            0
dtype: int64
```

Outliers

```
sns.boxplot(x=df['Model Year'])
```

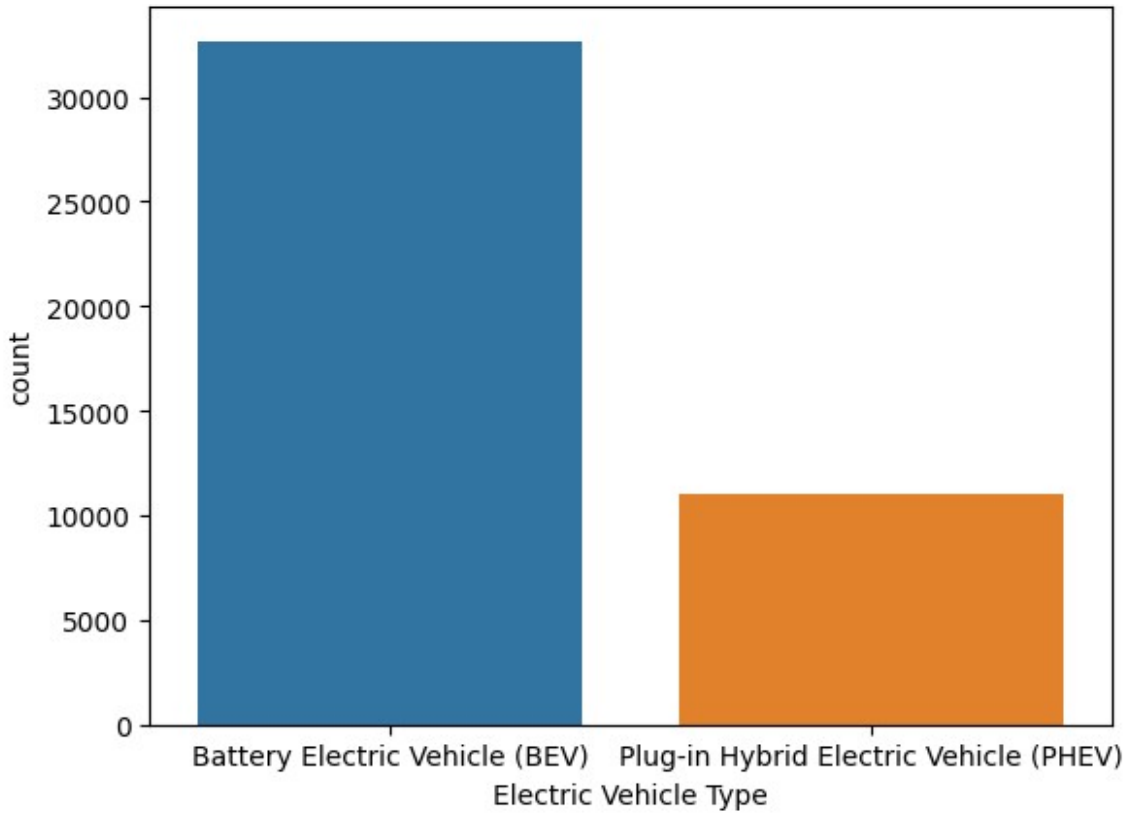
```
<Axes: xlabel='Model Year'>
```



Above we have plotted the boxplot with the help of seborn library. In boxplot xlabel is Model Year.

```
sns.countplot(x="Electric Vehicle Type" ,data=df)
```

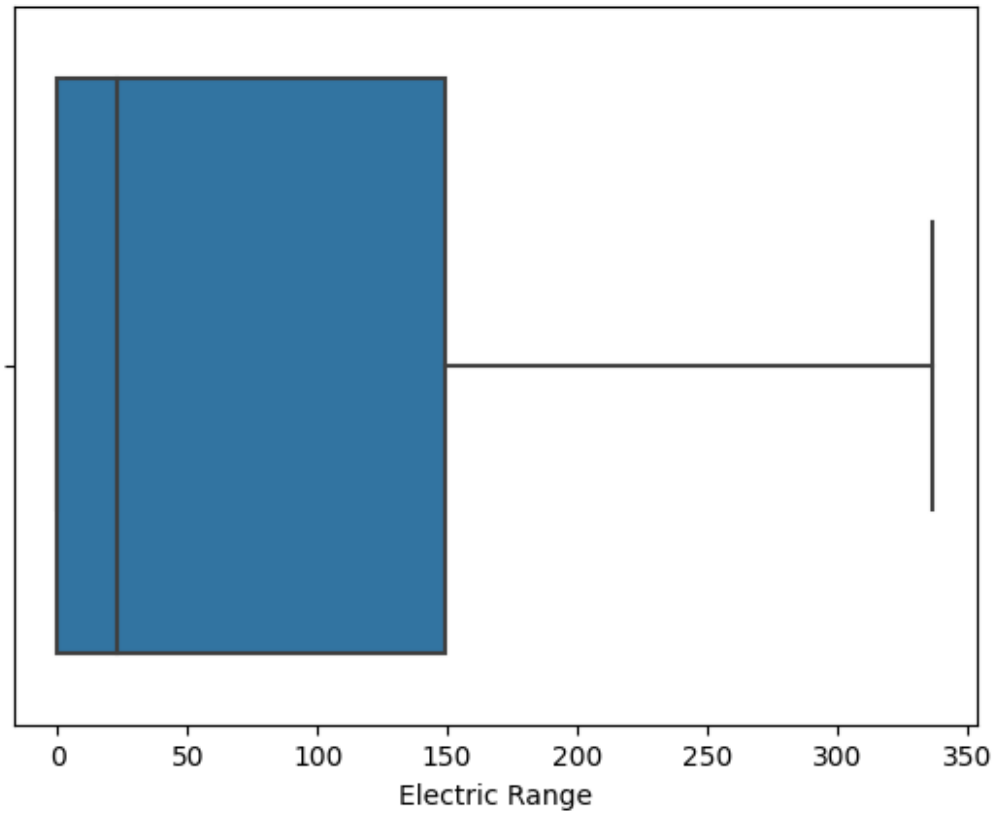
```
<Axes: xlabel='Electric Vehicle Type', ylabel='count'>
```



Above we have plotted the countplot with the help of seborn library. In boxplot xlabel is Electric Vehicle Type and ylabel is count . The count of Battery Electric Vehicle(BEV) is more than 30000. The Count of plug-in Hybrid Electric Vehicle(PHEV) is 10000.

```
sns.boxplot(x=df['Electric Range'])
```

```
<Axes: xlabel='Electric Range'>
```



Above we have plotted the boxplot with the help of seborn library. In boxplot xlabel is Electric Range.

```
#Handling outliers
R1=df.quantile(0.25)
R2=df.quantile(0.75)
IQR=R2-R1
print(IQR)
```

```
Postal Code       334.0
Model Year          4.0
Electric Range    149.0
dtype: float64
```

```
<ipython-input-42-3fe758c2fc3d>:2: FutureWarning: The default value of
numeric_only in DataFrame.quantile is deprecated. In a future version,
it will default to False. Select only valid columns or specify the
value of numeric_only to silence this warning.
  R1=df.quantile(0.25)
<ipython-input-42-3fe758c2fc3d>:3: FutureWarning: The default value of
numeric_only in DataFrame.quantile is deprecated. In a future version,
it will default to False. Select only valid columns or specify the
value of numeric_only to silence this warning.
  R2=df.quantile(0.75)
```

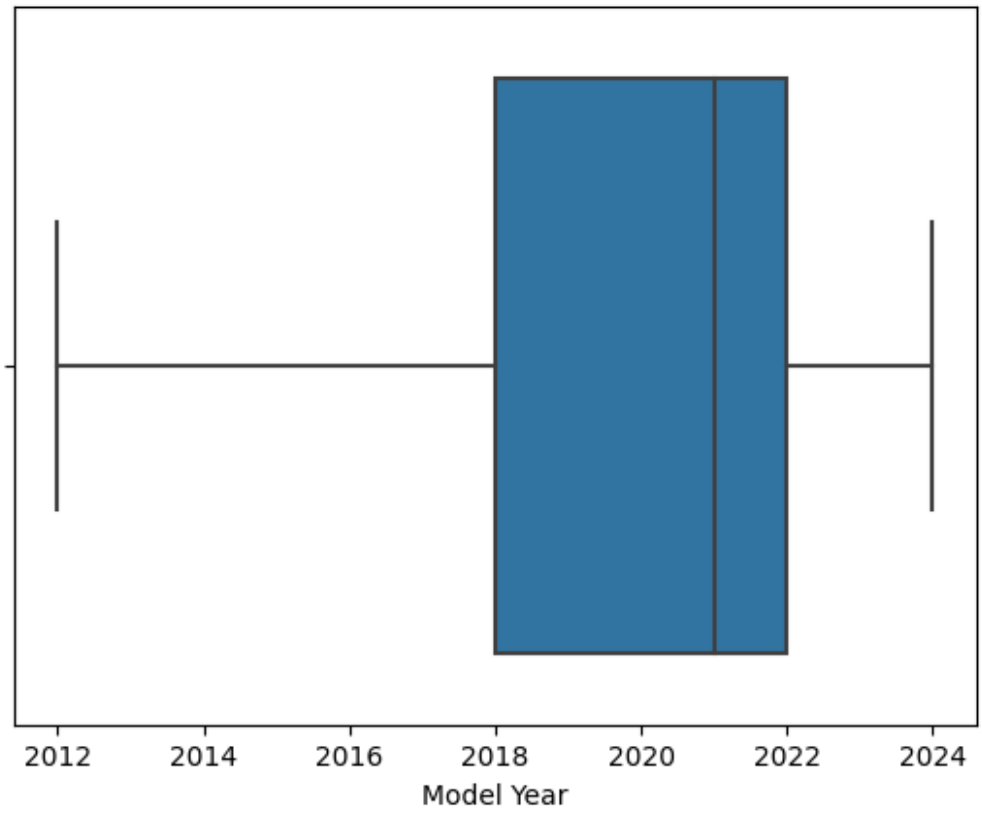
```
#Formula for Handling outliers
df=df[~((df<(R1-1.5*IQR))|(df>(R2+1.5*IQR))).any(axis=1)]
df.shape
```

```
<ipython-input-43-54eca9e89891>:2: FutureWarning: Automatic reindexing
on DataFrame vs Series comparisons is deprecated and will raise
ValueError in a future version. Do `left, right = left.align(right,
axis=1, copy=False)` before e.g. `left == right`
  df=df[~((df<(R1-1.5*IQR))|(df>(R2+1.5*IQR))).any(axis=1)]
```

```
(40539, 8)
```

```
sns.boxplot(x=df['Model Year'])
```

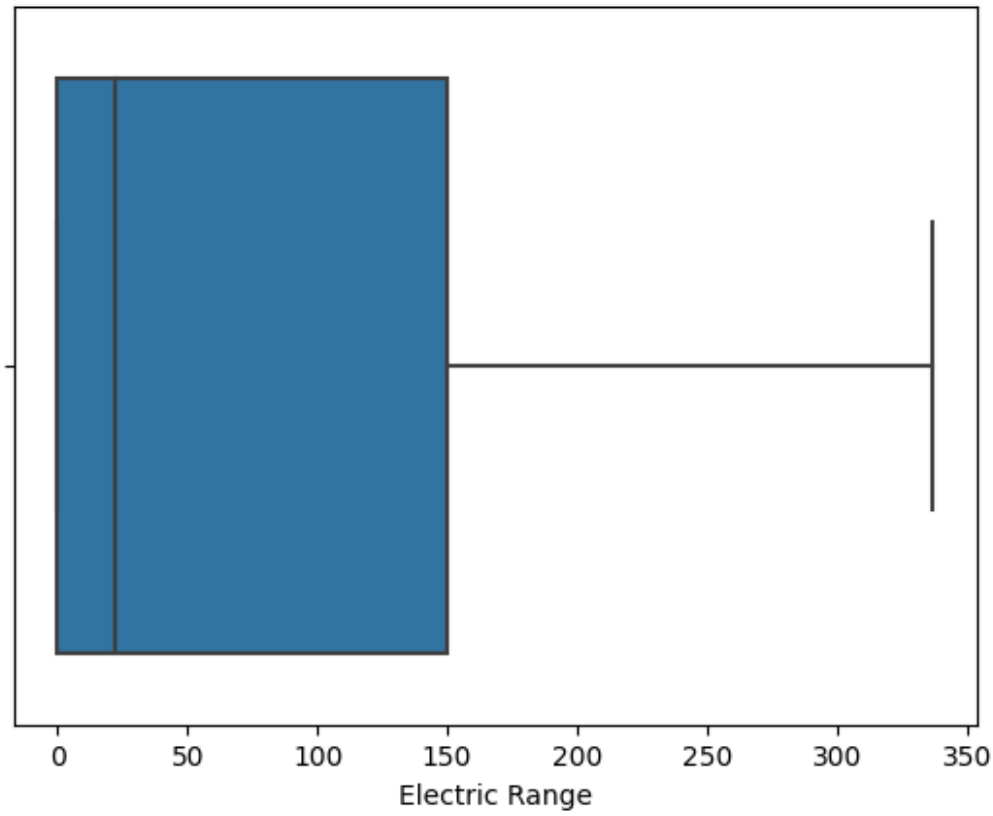
```
<Axes: xlabel='Model Year'>
```



After Handling Outliers . Above we have plotted the boxplot with the help of seborn library. In boxplot xlabel is Model Year.

```
sns.boxplot(x=df['Electric Range'])
```

```
<Axes: xlabel='Electric Range'>
```

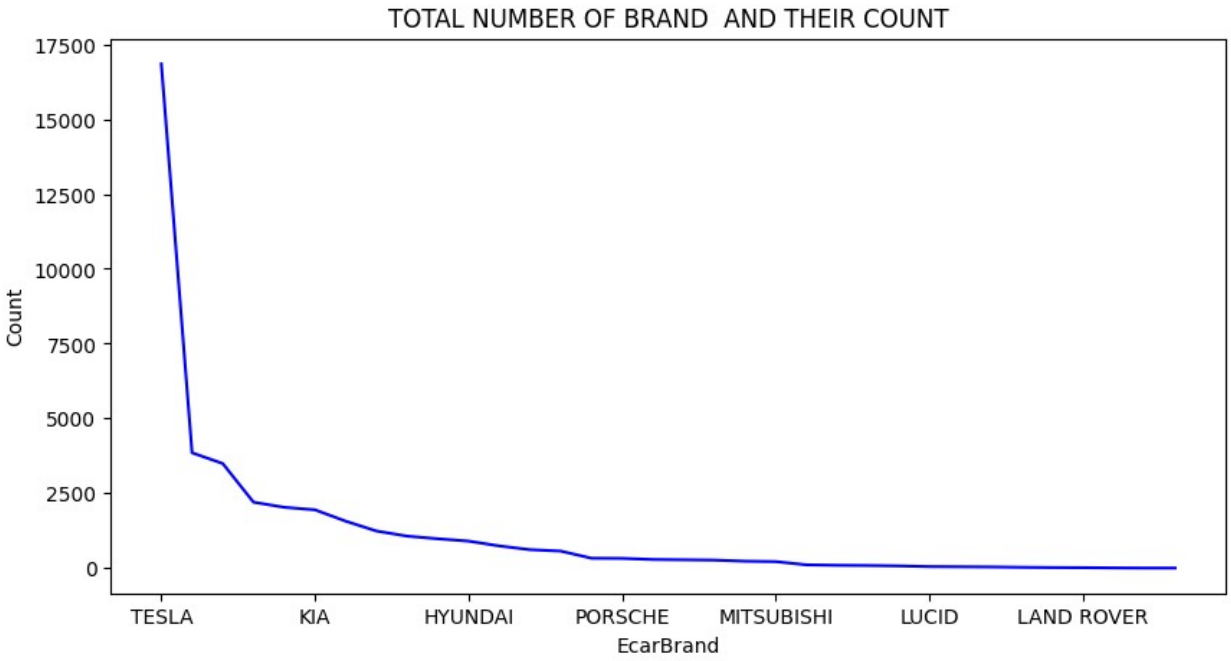



```
#Printing a top 5 rows
df.head()
```

```
  Vechicle ID          City  Postal Code  Model Year EcarBrand Car
Model  \
0  1N4AZ0CP5D     Bremerton      98310.0        2013    NISSAN
LEAF
1  1N4AZ1CP8K  Port Orchard      98366.0        2019    NISSAN
LEAF
2  5YJXCAE28L       Seattle      98199.0        2020     TESLA   MODEL
X
3  SADHC2S1XK       Olympia      98503.0        2019    JAGUAR    I-
PACE
7  3FA6P0SU5E      Tumwater      98501.0        2014      FORD
FUSION

                    Electric Vehicle Type  Electric Range
0          Battery Electric Vehicle (BEV)            75.0
1          Battery Electric Vehicle (BEV)           150.0
2          Battery Electric Vehicle (BEV)           293.0
3          Battery Electric Vehicle (BEV)           234.0
7  Plug-in Hybrid Electric Vehicle (PHEV)            19.0
```

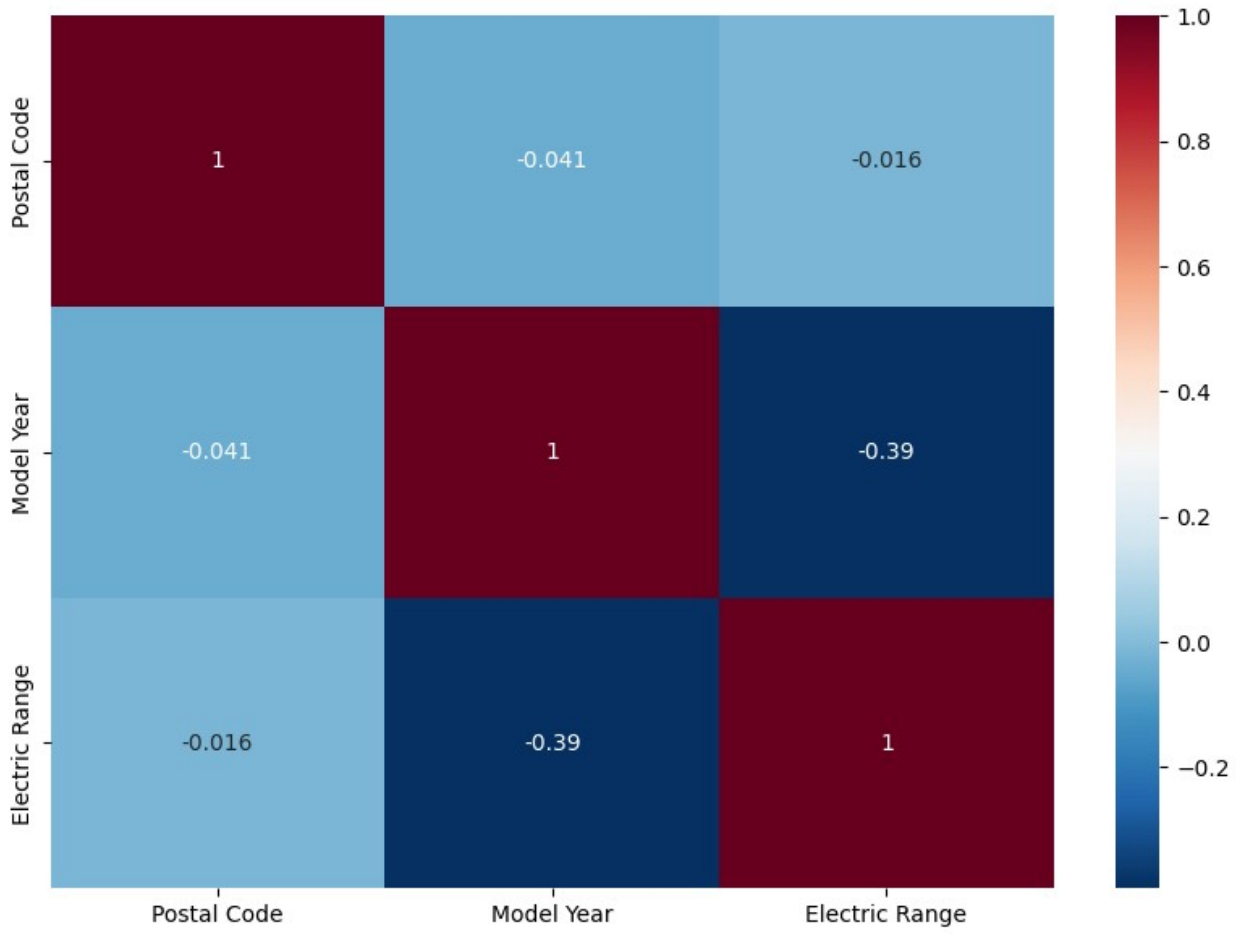
```
df.EcarBrand.value_counts().plot(kind='line',figsize=(10,5),color='Blu
e')
```

```
plt.title("TOTAL NUMBER OF BRAND  AND THEIR COUNT")
plt.ylabel('Count')
plt.xlabel('EcarBrand')

Text(0.5, 0, 'EcarBrand')
```



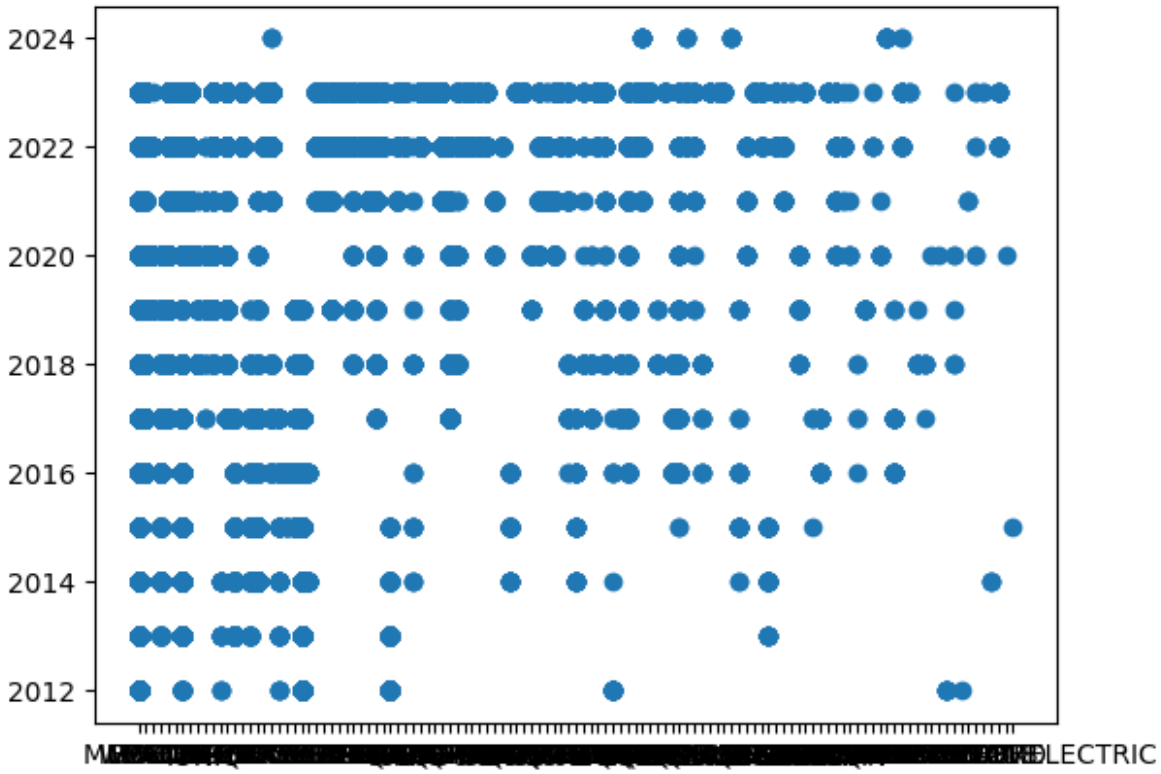
Above we have plotted the linegraph with the help of Matplotlib library. in lineplot the xlabel is EcarBrand and ylabel is count and Title is TOTAL NUMBER OF BRAND AND THEIR COUNT. here in above the EcarBrands are :-- TESLA,KIA,AUDI,MINI,MITSUBISHI,LUCID. In above the TESLA has most number of count i.e more than 2000 plus.

```
#plotting a heatmap
plt.figure(figsize=(10,7))
c=df.corr()
sns.heatmap(c,cmap="RdBu_r",annot=True)
c

<ipython-input-48-a003024a8e30>:3: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it
will default to False. Select only valid columns or specify the value
of numeric_only to silence this warning.
  c=df.corr()

                Postal Code  Model Year  Electric Range
Postal Code        1.000000   -0.041164       -0.016392
Model Year        -0.041164    1.000000       -0.394309
Electric Range    -0.016392   -0.394309        1.000000
```




```
plt.scatter(df['Car Model'],df['Model Year'])
```

```
<matplotlib.collections.PathCollection at 0x79e33619be80>
```




```
from sklearn.preprocessing import LabelEncoder
```

ENCODING

```
x=df.iloc[:,:-1]
y=df.iloc[:,-1]

Re=LabelEncoder()
x["Vechicle ID"]=Re.fit_transform(x["Vechicle ID"])
x["City"]=Re.fit_transform(x["City"])
x["EcarBrand"]=Re.fit_transform(x["EcarBrand"])
x["Car Model"]=Re.fit_transform(x["Car Model"])
x["Electric Vehicle Type"]=Re.fit_transform(x["Electric Vehicle
Type"])
x
```

```
       Vechicle ID  City  Postal Code  Model Year  EcarBrand  Car
Model  \
0             1021    26      98310.0        2013         24
67
1             1113   204      98366.0        2019         24
67
2             2918   229      98199.0        2020         30
71
3             4905   183      98503.0        2019         13
```

```
55
7             1710   274      98501.0        2014          9
46
...            ...   ...         ...         ...        ...        ..
.
50583         1164    32      98168.0        2017         24
67
50584         1089    24      98011.0        2019         24
67
50585         6694   109      98027.0        2021         33
117
50590          483   229      98103.0        2017          5
16
50591         2673   222      98580.0        2017         30
70

       Electric Vehicle Type
0                          0
1                          0
2                          0
3                          0
7                          1
...                      ...
50583                      0
50584                      0
50585                      1
50590                      0
50591                      0

[40539 rows x 7 columns]

df.head()

  Vechicle ID          City  Postal Code  Model Year ECarBrand Car
Model  \
0  1N4AZ0CP5D     Bremerton      98310.0        2013    NISSAN
LEAF
1  1N4AZ1CP8K  Port Orchard      98366.0        2019    NISSAN
LEAF
2  5YJXCAE28L       Seattle      98199.0        2020     TESLA   MODEL
X
3  SADHC2S1XK       Olympia      98503.0        2019    JAGUAR    I-
PACE
7  3FA6P0SU5E      Tumwater      98501.0        2014      FORD
FUSION

                     Electric Vehicle Type  Electric Range
0          Battery Electric Vehicle (BEV)            75.0
1          Battery Electric Vehicle (BEV)           150.0
2          Battery Electric Vehicle (BEV)           293.0
```

```
3          Battery Electric Vehicle (BEV)           234.0
7  Plug-in Hybrid Electric Vehicle (PHEV)            19.0
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 40539 entries, 0 to 50591
Data columns (total 8 columns):
 #   Column                 Non-Null Count  Dtype  
---  ------                 --------------  -----  
 0   Vechicle ID            40539 non-null  object 
 1   City                   40539 non-null  object 
 2   Postal Code            40539 non-null  float64
 3   Model Year             40539 non-null  int64  
 4   EcarBrand              40539 non-null  object 
 5   Car Model              40539 non-null  object 
 6   Electric Vehicle Type  40539 non-null  object 
 7   Electric Range         40539 non-null  float64
dtypes: float64(2), int64(1), object(5)
memory usage: 2.8+ MB
```

```
x.head()
```

```
   Vechicle ID  City  Postal Code  Model Year  EcarBrand  Car Model  \
0         1021    26      98310.0        2013         24         67   
1         1113   204      98366.0        2019         24         67   
2         2918   229      98199.0        2020         30         71   
3         4905   183      98503.0        2019         13         55   
7         1710   274      98501.0        2014          9         46   

   Electric Vehicle Type  
0                      0  
1                      0  
2                      0  
3                      0  
7                      1  
```

```
# splitting data
from sklearn.model_selection import train_test_split
xtrain,xtest,ytrain,ytest=train_test_split(x,y,test_size=0.3,random_st
ate=42)

#importing the model
from sklearn.linear_model import LinearRegression

#step2 -: initalize the model
model=LinearRegression()
model
model.fit(xtrain,ytrain)
ypred=model.predict(xtest)
```

```
from sklearn.metrics import
mean_absolute_error,mean_squared_error,r2_score

from sklearn.metrics import r2_score
print(f"Accuracy : {r2_score(ytest, ypred)}")
```

```
Accuracy : 0.2802319899129866
```

```
print(mean_absolute_error(ytest,ypred))
```

```
69.12002954899253
```

```
print(mean_squared_error(ytest,ypred))
```

```
6928.211220710704
```

```
print(r2_score(ytest,ypred))
```

```
0.2802319899129866
```

```
print(model.score(xtrain,ytrain))
```

```
0.27918006748886837
```

```
print(model.score(xtest,ytest))
```

```
0.2802319899129866
```