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

df=pd.read_csv("Cars_data.csv")

df.shape
```

```
(11914, 16)
```

```python
print("No. of rows:",df.shape[0])
print("No. of columns:",df.shape[1])
```

```
No. of rows: 11914
No. of columns: 16
```

```python
#checking empty(null) values
df.isnull().sum()
```

```
Make                    0
Model                   0
Year                    0
Engine Fuel Type        3
Engine HP              69
Engine Cylinders       30
Transmission Type       0
Driven_Wheels           0
Number of Doors         6
Market Category      3742
Vehicle Size            0
Vehicle Style           0
highway MPG             0
city mpg                0
Popularity              0
MSRP                    0
dtype: int64
```

```python
df=df.fillna(0)
#if mean has to be filled
#df["Engine HP"].fillna(df["Engine HP"].mean(),inplace = True)

#rechecking null values
df.isnull().sum()
```

```
Make                 0
Model                0
Year                 0
Engine Fuel Type     0
Engine HP            0
Engine Cylinders     0
```

```
Transmission Type    0
Driven_Wheels        0
Number of Doors      0
Market Category      0
Vehicle Size         0
Vehicle Style        0
highway MPG          0
city mpg             0
Popularity           0
MSRP                 0
dtype: int64

#checking duplicate values
df_dup=df.duplicated().any()
print(df_dup)

True

#dropping duplicate values
df=df.drop_duplicates()

df.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 11199 entries, 0 to 11913
Data columns (total 16 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Make               11199 non-null  object
 1   Model              11199 non-null  object
 2   Year               11199 non-null  int64
 3   Engine Fuel Type   11199 non-null  object
 4   Engine HP          11199 non-null  float64
 5   Engine Cylinders   11199 non-null  float64
 6   Transmission Type  11199 non-null  object
 7   Driven_Wheels      11199 non-null  object
 8   Number of Doors    11199 non-null  float64
 9   Market Category    11199 non-null  object
 10  Vehicle Size       11199 non-null  object
 11  Vehicle Style      11199 non-null  object
 12  highway MPG        11199 non-null  int64
 13  city mpg           11199 non-null  int64
 14  Popularity         11199 non-null  int64
 15  MSRP               11199 non-null  int64
dtypes: float64(3), int64(5), object(8)
memory usage: 1.5+ MB

df.describe()

               Year    Engine HP  Engine Cylinders  Number of Doors  \
count  11914.000000  11845.00000      11884.000000     11908.000000
```

```
mean    2010.384338    249.38607          5.628829          3.436093
std        7.579740    109.19187          1.780559          0.881315
min     1990.000000     55.00000          0.000000          2.000000
25%     2007.000000    170.00000          4.000000          2.000000
50%     2015.000000    227.00000          6.000000          4.000000
75%     2016.000000    300.00000          6.000000          4.000000
max     2017.000000   1001.00000         16.000000          4.000000

        highway MPG      city mpg    Popularity          MSRP
count  11914.000000  11914.000000  11914.000000  1.191400e+04
mean      26.637485     19.733255   1554.911197  4.059474e+04
std        8.863001      8.987798   1441.855347  6.010910e+04
min       12.000000      7.000000      2.000000  2.000000e+03
25%       22.000000     16.000000    549.000000  2.100000e+04
50%       26.000000     18.000000   1385.000000  2.999500e+04
75%       30.000000     22.000000   2009.000000  4.223125e+04
max      354.000000    137.000000   5657.000000  2.065902e+06
```

```
df.head()
```

```
  Make       Model  Year             Engine Fuel Type  Engine HP  \
0  BMW  1 Series M  2011  premium unleaded (required)      335.0
1  BMW    1 Series  2011  premium unleaded (required)      300.0
2  BMW    1 Series  2011  premium unleaded (required)      300.0
3  BMW    1 Series  2011  premium unleaded (required)      230.0
4  BMW    1 Series  2011  premium unleaded (required)      230.0

   Engine Cylinders Transmission Type     Driven_Wheels  Number of
Doors  \
0               6.0            MANUAL  rear wheel drive
2.0
1               6.0            MANUAL  rear wheel drive
2.0
2               6.0            MANUAL  rear wheel drive
2.0
3               6.0            MANUAL  rear wheel drive
2.0
4               6.0            MANUAL  rear wheel drive
2.0

                        Market Category Vehicle Size Vehicle Style  \
0  Factory Tuner,Luxury,High-Performance      Compact         Coupe
1                     Luxury,Performance      Compact   Convertible
2                Luxury,High-Performance      Compact         Coupe
3                     Luxury,Performance      Compact         Coupe
4                                 Luxury      Compact   Convertible

   highway MPG  city mpg  Popularity   MSRP
0           26        19        3916  46135
1           28        19        3916  40650
```

```
2          28         20         3916  36350
3          28         18         3916  29450
4          28         18         3916  34500
```

```
#what are the different types of make in the dataset and what is the count of each make in the data?
df["Make"].value_counts()
```

```
Chevrolet        1123
Ford              881
Volkswagen        809
Toyota            746
Dodge             626
Nissan            558
GMC               515
Honda             449
Mazda             423
Cadillac          397
Mercedes-Benz     353
Suzuki            351
BMW               334
Infiniti          330
Audi              328
Hyundai           303
Volvo             281
Subaru            256
Acura             252
Kia               231
Mitsubishi        213
Lexus             202
Buick             196
Chrysler          187
Pontiac           186
Lincoln           164
Oldsmobile        150
Land Rover        143
Porsche           136
Saab              111
Aston Martin       93
Plymouth           82
Bentley            74
Ferrari            69
FIAT               62
Scion              60
Maserati           58
Lamborghini        52
Rolls-Royce        31
Lotus              29
Tesla              18
HUMMER             17
```

```
Maybach            16
Alfa Romeo          5
McLaren             5
Spyker              3
Genesis             3
Bugatti             3
Name: Make, dtype: int64
```

```python
#show all the records where vehicle style is coupe or sedan.
df[df["Vehicle Style"].isin(["Coupe","Sedan"])]
```

```
          Make       Model  Year             Engine Fuel Type  Engine
HP  \
0          BMW  1 Series M  2011  premium unleaded (required)
335.0
2          BMW    1 Series  2011  premium unleaded (required)
300.0
3          BMW    1 Series  2011  premium unleaded (required)
230.0
5          BMW    1 Series  2012  premium unleaded (required)
230.0
7          BMW    1 Series  2012  premium unleaded (required)
300.0
...        ...         ...   ...                          ...          .
..
11809   Toyota    Yaris iA  2017             regular unleaded
106.0
11810   Toyota    Yaris iA  2017             regular unleaded
106.0
11890      BMW        Z4 M  2007  premium unleaded (required)
330.0
11892      BMW        Z4 M  2008  premium unleaded (required)
330.0
11913  Lincoln      Zephyr  2006             regular unleaded
221.0

       Engine Cylinders Transmission Type      Driven_Wheels  Number
of Doors  \
0                   6.0            MANUAL   rear wheel drive
2.0
2                   6.0            MANUAL   rear wheel drive
2.0
3                   6.0            MANUAL   rear wheel drive
2.0
5                   6.0            MANUAL   rear wheel drive
2.0
7                   6.0            MANUAL   rear wheel drive
2.0
...                 ...               ...                ...
...
```

```
11809                4.0            MANUAL  front wheel drive
4.0
11810                4.0         AUTOMATIC  front wheel drive
4.0
11890                6.0            MANUAL   rear wheel drive
2.0
11892                6.0            MANUAL   rear wheel drive
2.0
11913                6.0         AUTOMATIC  front wheel drive
4.0

                             Market Category Vehicle Size Vehicle
Style  \
0      Factory Tuner,Luxury,High-Performance      Compact
Coupe
2                     Luxury,High-Performance      Compact
Coupe
3                          Luxury,Performance      Compact
Coupe
5                          Luxury,Performance      Compact
Coupe
7                     Luxury,High-Performance      Compact
Coupe
...                                      ...          ...             ..
.
11809                                    NaN      Compact
Sedan
11810                                    NaN      Compact
Sedan
11890  Factory Tuner,Luxury,High-Performance      Compact
Coupe
11892  Factory Tuner,Luxury,High-Performance      Compact
Coupe
11913                                 Luxury      Midsize
Sedan

       highway MPG  city mpg  Popularity   MSRP
0               26        19        3916  46135
2               28        20        3916  36350
3               28        18        3916  29450
5               28        18        3916  31200
7               28        20        3916  39300
...            ...       ...         ...    ...
11809           39        30        2031  15950
11810           40        32        2031  17050
11890           22        15        3916  50100
11892           23        15        3916  50400
11913           26        17          61  28995

[4259 rows x 16 columns]
```

```
#remove all the records where popularity is above 5000
df = df[~(df["Popularity"] > 5000)]

#increase all the values of city mpg by 3
df["city mpg"]=df["city mpg"].apply(lambda x:x+3)

C:\Users\hp\AppData\Local\Temp\ipykernel_10264\1111787861.py:2:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  df["city mpg"]=df["city mpg"].apply(lambda x:x+3)

df

          Make       Model  Year                Engine Fuel Type
Engine HP  \
0          BMW  1 Series M  2011     premium unleaded (required)
335.0
1          BMW    1 Series  2011     premium unleaded (required)
300.0
2          BMW    1 Series  2011     premium unleaded (required)
300.0
3          BMW    1 Series  2011     premium unleaded (required)
230.0
4          BMW    1 Series  2011     premium unleaded (required)
230.0
...        ...         ...   ...                             ...
...
11909    Acura         ZDX  2012     premium unleaded (required)
300.0
11910    Acura         ZDX  2012     premium unleaded (required)
300.0
11911    Acura         ZDX  2012     premium unleaded (required)
300.0
11912    Acura         ZDX  2013  premium unleaded (recommended)
300.0
11913  Lincoln      Zephyr  2006                regular unleaded
221.0

       Engine Cylinders Transmission Type      Driven_Wheels  Number
of Doors  \
0                   6.0            MANUAL   rear wheel drive
2.0
1                   6.0            MANUAL   rear wheel drive
2.0
2                   6.0            MANUAL   rear wheel drive
```

```
2.0
3                    6.0            MANUAL   rear wheel drive
2.0
4                    6.0            MANUAL   rear wheel drive
2.0
...                  ...               ...                ...
...
11909                6.0         AUTOMATIC    all wheel drive
4.0
11910                6.0         AUTOMATIC    all wheel drive
4.0
11911                6.0         AUTOMATIC    all wheel drive
4.0
11912                6.0         AUTOMATIC    all wheel drive
4.0
11913                6.0         AUTOMATIC  front wheel drive
4.0

                             Market Category Vehicle Size  Vehicle
Style  \
0      Factory Tuner,Luxury,High-Performance      Compact
Coupe
1                         Luxury,Performance      Compact
Convertible
2                    Luxury,High-Performance      Compact
Coupe
3                         Luxury,Performance      Compact
Coupe
4                                     Luxury      Compact
Convertible
...                                      ...          ...            .
..
11909             Crossover,Hatchback,Luxury      Midsize  4dr
Hatchback
11910             Crossover,Hatchback,Luxury      Midsize  4dr
Hatchback
11911             Crossover,Hatchback,Luxury      Midsize  4dr
Hatchback
11912             Crossover,Hatchback,Luxury      Midsize  4dr
Hatchback
11913                                 Luxury      Midsize
Sedan

       highway MPG  city mpg  Popularity   MSRP
0               26        22        3916  46135
1               28        22        3916  40650
2               28        23        3916  36350
3               28        21        3916  29450
4               28        21        3916  34500
```

```
...                ...        ...          ...    ...
11909               23         19          204  46120
11910               23         19          204  56670
11911               23         19          204  50620
11912               23         19          204  50920
11913               26         20           61  28995

[11033 rows x 16 columns]
```