

```
In [ ]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_csv('/content/train-data.csv')
df
```

		Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price
0	0	Maruti Wagon R LXI CNG	Mumbai	2010		72000	CNG	Manual	First	26.6 km/kg	998 CC	58.16 bhp	5.0	NaN	1.75
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015		41000	Diesel	Manual	First	19.67 kmpl	1582 CC	126.2 bhp	5.0	NaN	12.50
2	2	Honda Jazz V	Chennai	2011		46000	Petrol	Manual	First	18.2 kmpl	1199 CC	88.7 bhp	5.0	8.61 Lakh	4.50
3	3	Maruti Ertiga VDI	Chennai	2012		87000	Diesel	Manual	First	20.77 kmpl	1248 CC	88.76 bhp	7.0	NaN	6.00
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013		40670	Diesel	Automatic	Second	15.2 kmpl	1968 CC	140.8 bhp	5.0	NaN	17.74
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
6014	6014	Maruti Swift VDI	Delhi	2014		27365	Diesel	Manual	First	28.4 kmpl	1248 CC	74 bhp	5.0	7.88 Lakh	4.75
6015	6015	Hyundai Xcent 1.1 CRDi S	Jaipur	2015		100000	Diesel	Manual	First	24.4 kmpl	1120 CC	71 bhp	5.0	NaN	4.00
6016	6016	Mahindra Xylo D4 BSIV	Jaipur	2012		55000	Diesel	Manual	Second	14.0 kmpl	2498 CC	112 bhp	8.0	NaN	2.90
6017	6017	Maruti Wagon R VXI	Kolkata	2013		46000	Petrol	Manual	First	18.9 kmpl	998 CC	67.1 bhp	5.0	NaN	2.65
6018	6018	Chevrolet Beat Diesel	Hyderabad	2011		47000	Diesel	Manual	First	25.44 kmpl	936 CC	57.6 bhp	5.0	NaN	2.50

6019 rows × 14 columns

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

		Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price
0	0	Maruti Wagon R LXI CNG	Mumbai	2010		72000	CNG	Manual	First	26.6 km/kg	998 CC	58.16 bhp	5.0	NaN	1.75
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015		41000	Diesel	Manual	First	19.67 kmpl	1582 CC	126.2 bhp	5.0	NaN	12.50
2	2	Honda Jazz V	Chennai	2011		46000	Petrol	Manual	First	18.2 kmpl	1199 CC	88.7 bhp	5.0	8.61 Lakh	4.50
3	3	Maruti Ertiga VDI	Chennai	2012		87000	Diesel	Manual	First	20.77 kmpl	1248 CC	88.76 bhp	7.0	NaN	6.00
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013		40670	Diesel	Automatic	Second	15.2 kmpl	1968 CC	140.8 bhp	5.0	NaN	17.74

```
In [ ]: df.tail()
```

		Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price
6014	6014	Maruti Swift VDI	Delhi	2014		27365	Diesel	Manual	First	28.4 kmpl	1248 CC	74 bhp	5.0	7.88 Lakh	4.75
6015	6015	Hyundai Xcent 1.1 CRDi S	Jaipur	2015		100000	Diesel	Manual	First	24.4 kmpl	1120 CC	71 bhp	5.0	NaN	4.00
6016	6016	Mahindra Xylo D4 BSIV	Jaipur	2012		55000	Diesel	Manual	Second	14.0 kmpl	2498 CC	112 bhp	8.0	NaN	2.90
6017	6017	Maruti Wagon R VXI	Kolkata	2013		46000	Petrol	Manual	First	18.9 kmpl	998 CC	67.1 bhp	5.0	NaN	2.65
6018	6018	Chevrolet Beat Diesel	Hyderabad	2011		47000	Diesel	Manual	First	25.44 kmpl	936 CC	57.6 bhp	5.0	NaN	2.50

```
In [ ]: df.isna().sum()
```

```
Out[ ]: Unnamed: 0          0
Name            0
Location        0
Year            0
Kilometers_Driven  0
Fuel_Type        0
Transmission     0
Owner_Type       0
Mileage          2
Engine           36
Power            36
Seats            42
New_Price         5195
Price             0
dtype: int64
```

```
In [ ]: df.dtypes
```

```
Out[ ]: Unnamed: 0      int64
Name          object
Location      object
Year          int64
Kilometers_Driven    int64
Fuel_Type     object
Transmission   object
Owner_Type    object
Mileage        object
Engine         object
Power          object
Seats          float64
New_Price     object
Price          float64
dtype: object
```

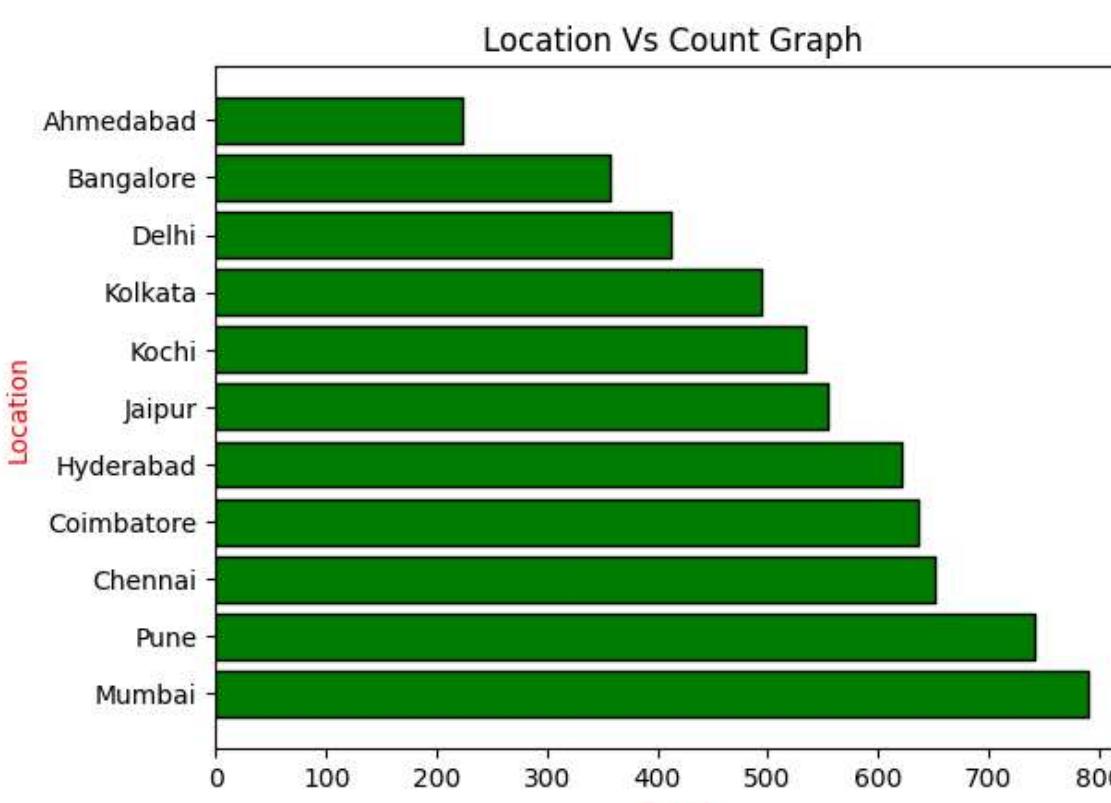
```
In [ ]: lst=['Location','Fuel_Type','Transmission','Owner_Type']
for i in lst:
    x=df[i].unique()
    print(i,':',x)
```

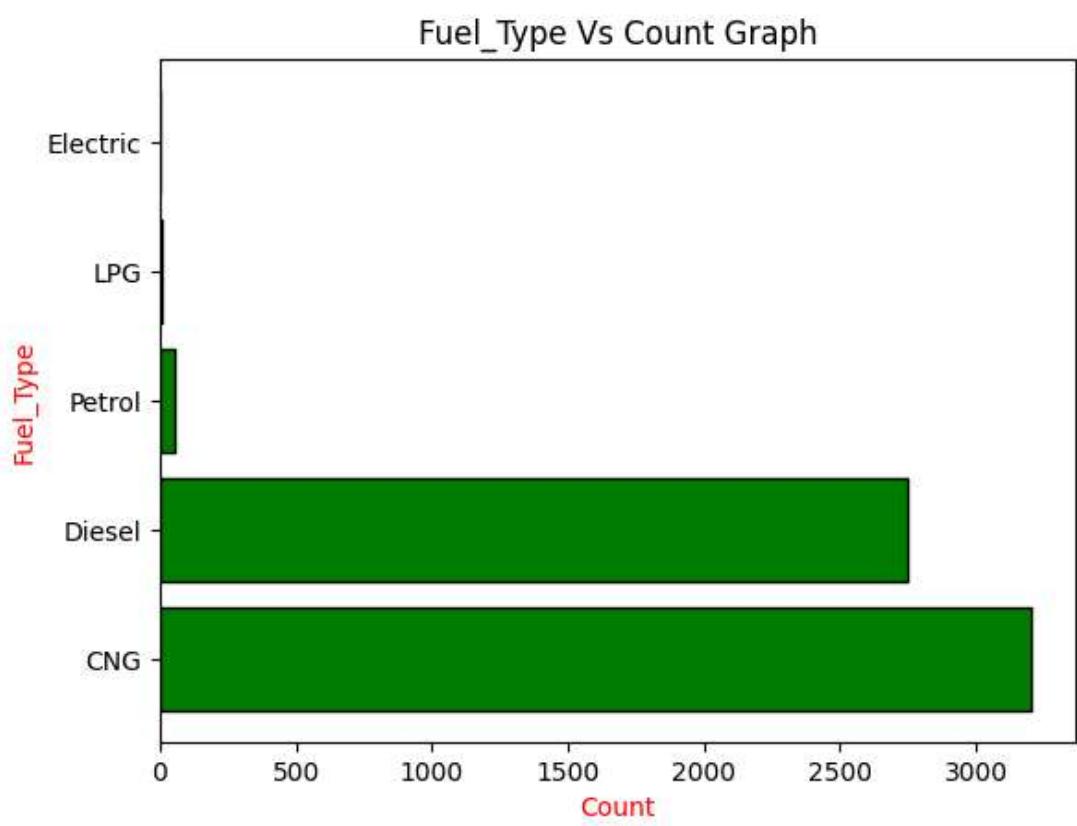
```
Location : ['Mumbai' 'Pune' 'Chennai' 'Coimbatore' 'Hyderabad' 'Jaipur' 'Kochi'
 'Kolkata' 'Delhi' 'Bangalore' 'Ahmedabad']
Fuel_Type : ['CNG' 'Diesel' 'Petrol' 'LPG' 'Electric']
Transmission : ['Manual' 'Automatic']
Owner_Type : ['First' 'Second' 'Fourth & Above' 'Third']
```

```
In [ ]: for i in lst:
    y=df[i].value_counts()
    print(y)
    print('*'*100)
```

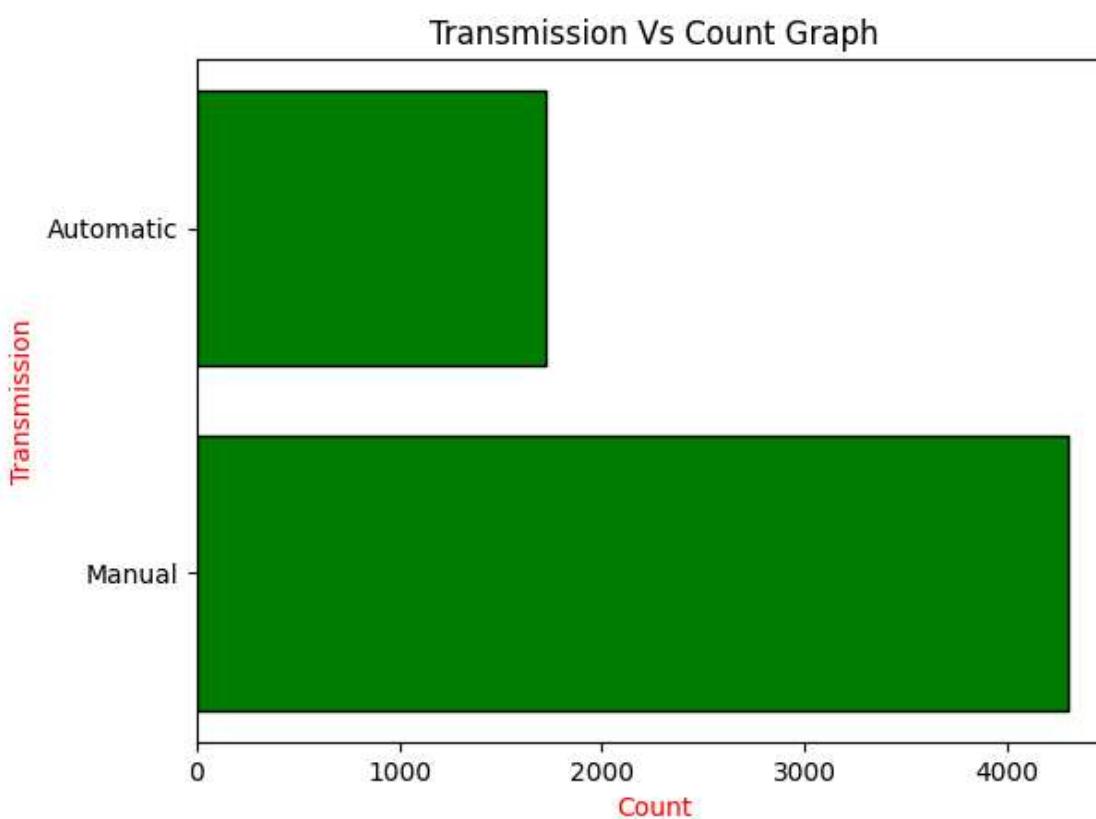
```
Location
Mumbai      790
Hyderabad   742
Kochi       651
Coimbatore  636
Pune        622
Delhi        554
Kolkata     535
Chennai      494
Jaipur       413
Bangalore    358
Ahmedabad   224
Name: count, dtype: int64
*****
Fuel_Type
Diesel      3205
Petrol      2746
CNG         56
LPG         10
Electric     2
Name: count, dtype: int64
*****
Transmission
Manual      4299
Automatic   1720
Name: count, dtype: int64
*****
Owner_Type
First       4929
Second      968
Third       113
Fourth & Above  9
Name: count, dtype: int64
*****
```

```
In [ ]: for i in lst:
    plt.barh(df[i].unique(),df[i].value_counts(), color='green',edgecolor='black')
    plt.xlabel('Count', color='red')
    plt.ylabel(i, color='red')
    plt.title(f'{i} Vs Count Graph')
    plt.show()
    print('*'*100)
```

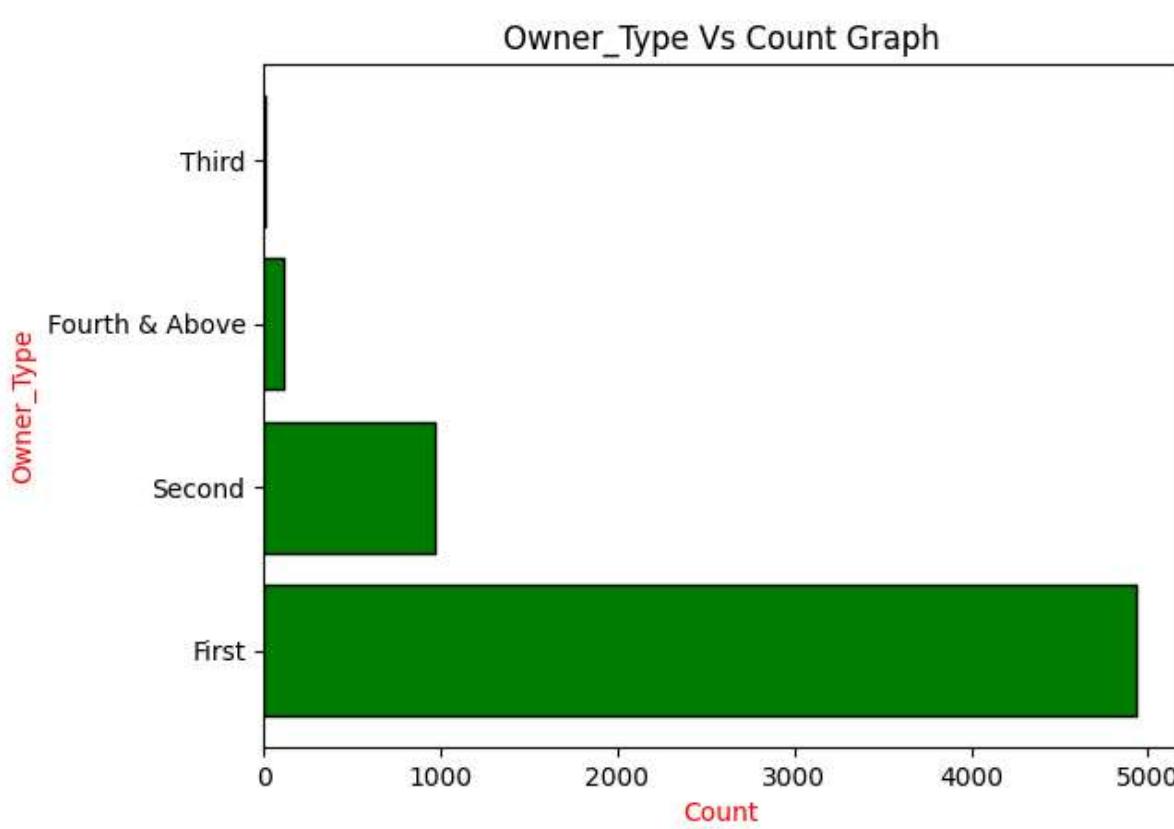




\*\*\*\*\*



\*\*\*\*\*



\*\*\*\*\*

In [ ]: `df['Name'].unique()`

Out[ ]: `array(['Maruti Wagon R LXI CNG', 'Hyundai Creta 1.6 CRDi SX Option', 'Honda Jazz V', ..., 'Tata Bolt Revotron XT', 'Honda', 'Mahindra Xylo D4 BSIV'], dtype=object)`

In [ ]: `df['Name'].value_counts()`

```
Out[ ]: Name
Mahindra XUV500 W8 2WD          49
Maruti Swift VDI                 45
Honda City 1.5 S MT              34
Maruti Swift Dzire VDI           34
Maruti Swift VDI BSIV             31
..
Ford Fiesta Titanium 1.5 TDCi    1
Mahindra Scorpio S10 AT 4WD       1
Hyundai i20 1.2 Era               1
Toyota Camry W4 (AT)              1
Mahindra Xylo D4 BSIV             1
Name: count, Length: 1878, dtype: int64
```

```
In [ ]: df['Brand']=df['Name'].apply(lambda x:x.split()[0])
df
```

		Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price	Price	Brand
0	0	Maruti Wagon R LXI CNG	Mumbai	2010		72000	CNG	Manual	First	26.6 km/kg	998 CC	58.16 bhp	5.0	NaN	1.75	Maruti
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015		41000	Diesel	Manual	First	19.67 kmpl	1582 CC	126.2 bhp	5.0	NaN	12.50	Hyundai
2	2	Honda Jazz V	Chennai	2011		46000	Petrol	Manual	First	18.2 kmpl	1199 CC	88.7 bhp	5.0	8.61 Lakh	4.50	Honda
3	3	Maruti Ertiga VDI	Chennai	2012		87000	Diesel	Manual	First	20.77 kmpl	1248 CC	88.76 bhp	7.0	NaN	6.00	Maruti
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013		40670	Diesel	Automatic	Second	15.2 kmpl	1968 CC	140.8 bhp	5.0	NaN	17.74	Audi
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
6014	6014	Maruti Swift VDI	Delhi	2014		27365	Diesel	Manual	First	28.4 kmpl	1248 CC	74 bhp	5.0	7.88 Lakh	4.75	Maruti
6015	6015	Hyundai Xcent 1.1 CRDi S	Jaipur	2015		100000	Diesel	Manual	First	24.4 kmpl	1120 CC	71 bhp	5.0	NaN	4.00	Hyundai
6016	6016	Mahindra Xylo D4 BSIV	Jaipur	2012		55000	Diesel	Manual	Second	14.0 kmpl	2498 CC	112 bhp	8.0	NaN	2.90	Mahindra
6017	6017	Maruti Wagon R VXI	Kolkata	2013		46000	Petrol	Manual	First	18.9 kmpl	998 CC	67.1 bhp	5.0	NaN	2.65	Maruti
6018	6018	Chevrolet Beat Diesel	Hyderabad	2011		47000	Diesel	Manual	First	25.44 kmpl	936 CC	57.6 bhp	5.0	NaN	2.50	Chevrolet

6019 rows x 15 columns

```
In [ ]: df['Brand'].value_counts()
```

```
Out[ ]: Brand
Maruti      1211
Hyundai     1107
Honda       608
Toyota      411
Mercedes-Benz 318
Volkswagen   315
Ford         300
Mahindra     272
BMW          267
Audi          236
Tata          186
Skoda         173
Renault       145
Chevrolet     121
Nissan        91
Land          60
Jaguar        40
Fiat          28
Mitsubishi    27
Mini          26
Volvo          21
Porsche        18
Jeep          15
Datsun        13
Force          3
ISUZU          2
Smart          1
Ambassador    1
Isuzu          1
Bentley        1
Lamborghini    1
Name: count, dtype: int64
```

```
In [ ]:
```

```
In [ ]: df1=pd.get_dummies(df[['Location','Fuel_Type','Transmission','Owner_Type','Brand']],dtype=int,drop_first=True)
df1
```

Out[ ]:	Location_Bangalore	Location_Chennai	Location_Coimbatore	Location_Delhi	Location_Hyderabad	Location_Jaipur	Location_Kochi	Location_Kolkata	Location_Mumbai
0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0
3	0	1	0	0	0	0	0	0	0
4	0	0	1	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...
6014	0	0	0	1	0	0	0	0	0
6015	0	0	0	0	0	1	0	0	0
6016	0	0	0	0	0	1	0	0	0
6017	0	0	0	0	0	0	0	0	1
6018	0	0	0	0	1	0	0	0	0

6019 rows × 48 columns

```
In [ ]: dfe=pd.concat([df,df1],axis=1)
dfe
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	...	Brand_Mitsubishi	Brand_Nissan	Brand_Po
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	First	26.6 km/kg	998 CC	...	0	0	
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	1582 CC	...	0	0	
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	18.2 kmpl	1199 CC	...	0	0	
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77 kmpl	1248 CC	...	0	0	
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.2 kmpl	1968 CC	...	0	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
6014	6014	Maruti Swift VDI	Delhi	2014	27365	Diesel	Manual	First	28.4 kmpl	1248 CC	...	0	0	
6015	6015	Hyundai Xcent 1.1 CRDi S	Jaipur	2015	100000	Diesel	Manual	First	24.4 kmpl	1120 CC	...	0	0	
6016	6016	Mahindra Xylo D4 BSIV	Jaipur	2012	55000	Diesel	Manual	Second	14.0 kmpl	2498 CC	...	0	0	
6017	6017	Maruti Wagon R VXI	Kolkata	2013	46000	Petrol	Manual	First	18.9 kmpl	998 CC	...	0	0	
6018	6018	Chevrolet Beat Diesel	Hyderabad	2011	47000	Diesel	Manual	First	25.44 kmpl	936 CC	...	0	0	

6019 rows × 63 columns

```
In [ ]: dfe.drop(['Unnamed: 0','Name','Location','Fuel_Type','Transmission','Owner_Type','New_Price','Brand'],axis=1,inplace=True)  
dfe
```

Out[ ]:

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Price	Location_Bangalore	Location_Chennai	Location_Coimbatore	...	Brand_Mitsubishi	Brand_Nissan	B
0	2010	72000	26.6 km/kg	998 CC	58.16 bhp	5.0	1.75	0	0	0	...	0	0	0
1	2015	41000	19.67 kmpl	1582 CC	126.2 bhp	5.0	12.50	0	0	0	...	0	0	0
2	2011	46000	18.2 kmpl	1199 CC	88.7 bhp	5.0	4.50	0	1	0	...	0	0	0
3	2012	87000	20.77 kmpl	1248 CC	88.76 bhp	7.0	6.00	0	1	0	...	0	0	0
4	2013	40670	15.2 kmpl	1968 CC	140.8 bhp	5.0	17.74	0	0	1	...	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
6014	2014	27365	28.4 kmpl	1248 CC	74 bhp	5.0	4.75	0	0	0	...	0	0	0
6015	2015	100000	24.4 kmpl	1120 CC	71 bhp	5.0	4.00	0	0	0	...	0	0	0
6016	2012	55000	14.0 kmpl	2498 CC	112 bhp	8.0	2.90	0	0	0	...	0	0	0
6017	2013	46000	18.9 kmpl	998 CC	67.1 bhp	5.0	2.65	0	0	0	...	0	0	0
6018	2011	47000	25.44 kmpl	936 CC	57.6 bhp	5.0	2.50	0	0	0	...	0	0	0

6019 rows × 15 columns

In [ ]:

```
dfe['Mileage']=dfe['Mileage'].str.replace('km/kg',' ')
dfe['Mileage']=dfe['Mileage'].str.replace('kmpl',' ')
dfe['Engine']=dfe['Engine'].str.replace('CC',' ')
dfe['Power']=dfe['Power'].str.replace('bhp',' ')
dfe
```

Out[ ]:

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Price	Location_Bangalore	Location_Chennai	Location_Coimbatore	...	Brand_Mitsubishi	Brand_Nissan	B
0	2010	72000	26.6	998	58.16	5.0	1.75	0	0	0	...	0	0	0
1	2015	41000	19.67	1582	126.2	5.0	12.50	0	0	0	...	0	0	0
2	2011	46000	18.2	1199	88.7	5.0	4.50	0	1	0	...	0	0	0
3	2012	87000	20.77	1248	88.76	7.0	6.00	0	1	0	...	0	0	0
4	2013	40670	15.2	1968	140.8	5.0	17.74	0	0	1	...	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
6014	2014	27365	28.4	1248	74	5.0	4.75	0	0	0	...	0	0	0
6015	2015	100000	24.4	1120	71	5.0	4.00	0	0	0	...	0	0	0
6016	2012	55000	14.0	2498	112	8.0	2.90	0	0	0	...	0	0	0
6017	2013	46000	18.9	998	67.1	5.0	2.65	0	0	0	...	0	0	0
6018	2011	47000	25.44	936 CC	57.6	5.0	2.50	0	0	0	...	0	0	0

6019 rows × 15 columns

In [ ]:

```
#null==>0
dfe['Mileage']=dfe['Mileage'].str.replace('null','0')
dfe['Power']=dfe['Power'].str.replace('null','0')
dfe['Engine']=dfe['Engine'].str.replace('null','0')
```

In [ ]:

```
#datatype conversion
dfe['Mileage']=dfe['Mileage'].astype(float)
dfe['Power']=dfe['Power'].astype(float)
dfe['Engine']=dfe['Engine'].astype(float)
dfe.dtypes
```

```
Out[ ]: Year           int64
Kilometers_Driven      int64
Mileage                 float64
Engine                  float64
Power                   float64
Seats                   float64
Price                   float64
Location_Bangalore     int64
Location_Chennai        int64
Location_Coimbatore    int64
Location_Delhi          int64
Location_Hyderabad      int64
Location_Jaipur         int64
Location_Kochi          int64
Location_Kolkata        int64
Location_Mumbai         int64
Location_Pune           int64
Fuel_Type_Diesel        int64
Fuel_Type_Electric      int64
Fuel_Type_LPG            int64
Fuel_Type_Petrol         int64
Transmission_Manual    int64
Owner_Type_Fourth & Above int64
Owner_Type_Second       int64
Owner_Type_Third         int64
Brand_Audi               int64
Brand_BMW                int64
Brand_Bentley             int64
Brand_Chevrolet          int64
Brand_Datsun              int64
Brand_Fiat                int64
Brand_Force               int64
Brand_Ford                int64
Brand_Honda               int64
Brand_Hyundai              int64
Brand_ISUZU               int64
Brand_Isuzu               int64
Brand_Jaguar               int64
Brand_Jeep                 int64
Brand_Lamborghini          int64
Brand_Land                 int64
Brand_Mahindra             int64
Brand_Maruti               int64
Brand_Mercedes-Benz        int64
Brand_Mini                 int64
Brand_Mitsubishi            int64
Brand_Nissan               int64
Brand_Porsche              int64
Brand_Renault              int64
Brand_Skoda                 int64
Brand_Smart                 int64
Brand_Tata                 int64
Brand_Toyota               int64
Brand_Volkswagen            int64
Brand_Volvo                 int64
dtype: object
```

```
In [ ]: #convert 0 to a missing value
dfe.loc[dfe.Engine==0,'Engine']=np.Nan
dfe.loc[dfe.Mileage==0,'Mileage']=np.Nan
dfe.loc[dfe.Power==0,'Power']=np.Nan
```

```
In [ ]: dfe.isna().sum()
```

```
Out[ ]: Year          0  
Kilometers_Driven      0  
Mileage            70  
Engine             36  
Power             143  
Seats              42  
Price              0  
Location_Bangalore    0  
Location_Chennai       0  
Location_Coimbatore     0  
Location_Delhi         0  
Location_Hyderabad      0  
Location_Jaipur        0  
Location_Kochi         0  
Location_Kolkata        0  
Location_Mumbai        0  
Location_Pune          0  
Fuel_Type_Diesel       0  
Fuel_Type_Electric      0  
Fuel_Type_LPG          0  
Fuel_Type_Petrol        0  
Transmission_Manual     0  
Owner_Type_Fourth & Above 0  
Owner_Type_Second       0  
Owner_Type_Third        0  
Brand_Audi            0  
Brand_BMW             0  
Brand_Bentley          0  
Brand_Chevrolet        0  
Brand_Datsun           0  
Brand_Fiat             0  
Brand_Force            0  
Brand_Ford             0  
Brand_Honda            0  
Brand_Hyundai          0  
Brand_ISUZU            0  
Brand_Isuzu             0  
Brand_Jaguar            0  
Brand_Jeep              0  
Brand_Lamborghini       0  
Brand_Land              0  
Brand_Mahindra          0  
Brand_Maruti            0  
Brand_Mercedes-Benz     0  
Brand_Mini              0  
Brand_Mitsubishi        0  
Brand_Nissan            0  
Brand_Porsche           0  
Brand_Renault           0  
Brand_Skoda              0  
Brand_Smart              0  
Brand_Tata              0  
Brand_Toyota             0  
Brand_Volkswagen        0  
Brand_Volvo              0  
dtype: int64
```

```
In [ ]: dfe['Mileage']=dfe['Mileage'].fillna(dfe['Mileage'].mean())  
dfe['Engine']=dfe['Engine'].fillna(dfe['Engine'].mean())  
dfe['Power']=dfe['Power'].fillna(dfe['Power'].mean())  
dfe['Seats']=dfe['Seats'].fillna(dfe['Seats'].mode()[0])
```

```
In [ ]: dfe.isna().sum()
```

```
Out[ ]: Year          0  
Kilometers_Driven      0  
Mileage          0  
Engine            0  
Power             0  
Seats             0  
Price             0  
Location_Bangalore    0  
Location_Chennai       0  
Location_Coimbatore     0  
Location_Delhi         0  
Location_Hyderabad      0  
Location_Jaipur        0  
Location_Kochi         0  
Location_Kolkata        0  
Location_Mumbai         0  
Location_Pune          0  
Fuel_Type_Diesel        0  
Fuel_Type_Electric      0  
Fuel_Type_LPG           0  
Fuel_Type_Petrol         0  
Transmission_Manual     0  
Owner_Type_Fourth & Above 0  
Owner_Type_Second        0  
Owner_Type_Third         0  
Brand_Audi             0  
Brand_BMW              0  
Brand_Bentley          0  
Brand_Chevrolet         0  
Brand_Datsun           0  
Brand_Fiat             0  
Brand_Force            0  
Brand_Ford              0  
Brand_Honda             0  
Brand_Hyundai           0  
Brand_ISUZU             0  
Brand_Isuzu             0  
Brand_Jaguar            0  
Brand_Jeep              0  
Brand_Lamborghini        0  
Brand_Land              0  
Brand_Mahindra          0  
Brand_Maruti             0  
Brand_Mercedes-Benz      0  
Brand_Mini              0  
Brand_Mitsubishi         0  
Brand_Nissan             0  
Brand_Porsche            0  
Brand_Renault            0  
Brand_Skoda              0  
Brand_Smart              0  
Brand_Tata              0  
Brand_Toyota             0  
Brand_Volkswagen          0  
Brand_Volvo              0  
dtype: int64
```

```
In [ ]: dfe.drop(['Brand_Bentley','Brand_Chevrolet','Brand_ISUZU','Brand_Isuzu','Brand_Force','Brand_Lamborghini','Fuel_Type_Electric'],axis=1,inplace=True)
```

```
In [ ]: x=dfe.drop(['Price'],axis=1)  
y=dfe['Price']  
x.isna().sum()
```

```
Out[ ]: Year          0  
Kilometers_Driven 0  
Mileage          0  
Engine            0  
Power             0  
Seats             0  
Location_Bangalore 0  
Location_Chennai 0  
Location_Coimbatore 0  
Location_Delhi   0  
Location_Hyderabad 0  
Location_Jaipur  0  
Location_Kochi   0  
Location_Kolkata 0  
Location_Mumbai  0  
Location_Pune    0  
Fuel_Type_Diesel 0  
Fuel_Type_LPG   0  
Fuel_Type_Petrol 0  
Transmission_Manual 0  
Owner_Type_Fourth & Above 0  
Owner_Type_Second 0  
Owner_Type_Third  0  
Brand_Audi        0  
Brand_BMW         0  
Brand_Datsun     0  
Brand_Fiat        0  
Brand_Ford        0  
Brand_Honda       0  
Brand_Hyundai    0  
Brand_Jaguar      0  
Brand_Jeep         0  
Brand_Land        0  
Brand_Mahindra   0  
Brand_Maruti      0  
Brand_Mercedes-Benz 0  
Brand_Mini         0  
Brand_Mitsubishi 0  
Brand_Nissan      0  
Brand_Porsche     0  
Brand_Renault     0  
Brand_Skoda        0  
Brand_Smart        0  
Brand_Tata         0  
Brand_Toyota       0  
Brand_Volkswagen 0  
Brand_Volvo        0  
dtype: int64
```

```
In [ ]: df2=pd.read_csv('/content/test-data.csv')  
df2
```

		Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price
0	0	Maruti Alto K10 LXI CNG		Delhi	2014	40929	CNG	Manual	First	32.26 km/kg	998 CC	58.2 bhp	4.0	NaN
1	1	Maruti Alto 800 2016-2019 LXI		Coimbatore	2013	54493	Petrol	Manual	Second	24.7 kmpl	796 CC	47.3 bhp	5.0	NaN
2	2	Toyota Innova Crysta Touring Sport 2.4 MT		Mumbai	2017	34000	Diesel	Manual	First	13.68 kmpl	2393 CC	147.8 bhp	7.0	25.27 Lakh
3	3	Toyota Etios Liva GD	Hyderabad		2012	139000	Diesel	Manual	First	23.59 kmpl	1364 CC	null bhp	5.0	NaN
4	4	Hyundai i20 Magna		Mumbai	2014	29000	Petrol	Manual	First	18.5 kmpl	1197 CC	82.85 bhp	5.0	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1229	1229	Volkswagen Vento Diesel Trendline		Hyderabad	2011	89411	Diesel	Manual	First	20.54 kmpl	1598 CC	103.6 bhp	5.0	NaN
1230	1230	Volkswagen Polo GT TSI		Mumbai	2015	59000	Petrol	Automatic	First	17.21 kmpl	1197 CC	103.6 bhp	5.0	NaN
1231	1231	Nissan Micra Diesel XV		Kolkata	2012	28000	Diesel	Manual	First	23.08 kmpl	1461 CC	63.1 bhp	5.0	NaN
1232	1232	Volkswagen Polo GT TSI		Pune	2013	52262	Petrol	Automatic	Third	17.2 kmpl	1197 CC	103.6 bhp	5.0	NaN
1233	1233	Mercedes-Benz E-Class 2009-2013 E 220 CDI Avantgar...		Kochi	2014	72443	Diesel	Automatic	First	10.0 kmpl	2148 CC	170 bhp	5.0	NaN

1234 rows × 13 columns

```
In [ ]: df2.isna().sum()
```

```
Out[ ]: Unnamed: 0          0  
Name            0  
Location        0  
Year            0  
Kilometers_Driven 0  
Fuel_Type       0  
Transmission    0  
Owner_Type      0  
Mileage          0  
Engine           10  
Power            10  
Seats            11  
New_Price        1052  
dtype: int64
```

```
In [ ]: df2.dtypes
```

```
Out[ ]: Unnamed: 0          int64
Name            object
Location        object
Year             int64
Kilometers_Driven    int64
Fuel_Type        object
Transmission     object
Owner_Type       object
Mileage           object
Engine            object
Power             object
Seats              float64
New_Price         object
dtype: object
```

```
In [ ]: lst1=['Location','Fuel_Type','Transmission','Owner_Type']
for i in lst1:
    print(i,':',df2[i].unique())
```

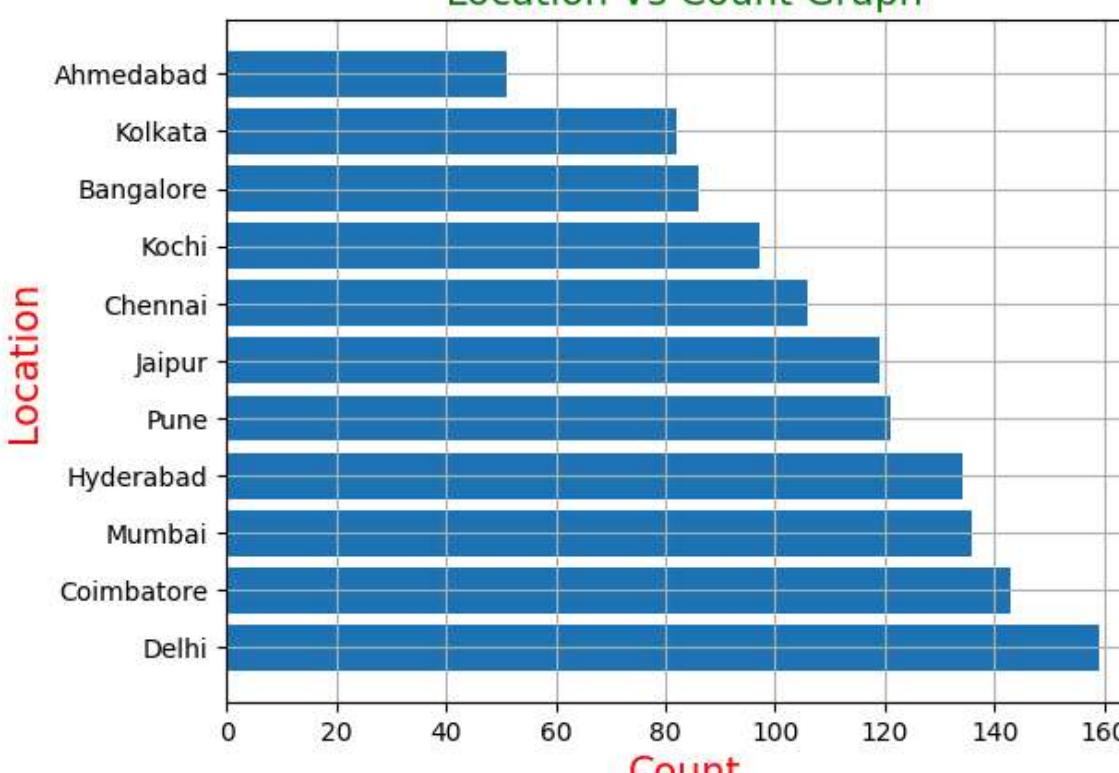
```
Location : ['Delhi' 'Coimbatore' 'Mumbai' 'Hyderabad' 'Pune' 'Jaipur' 'Chennai'
 'Kochi' 'Bangalore' 'Kolkata' 'Ahmedabad']
Fuel_Type : ['CNG' 'Petrol' 'Diesel' 'LPG']
Transmission : ['Manual' 'Automatic']
Owner_Type : ['First' 'Second' 'Third' 'Fourth & Above']
```

```
In [ ]: for i in lst1:
    print(i,':',df2[i].value_counts())
    print('*'*100)
```

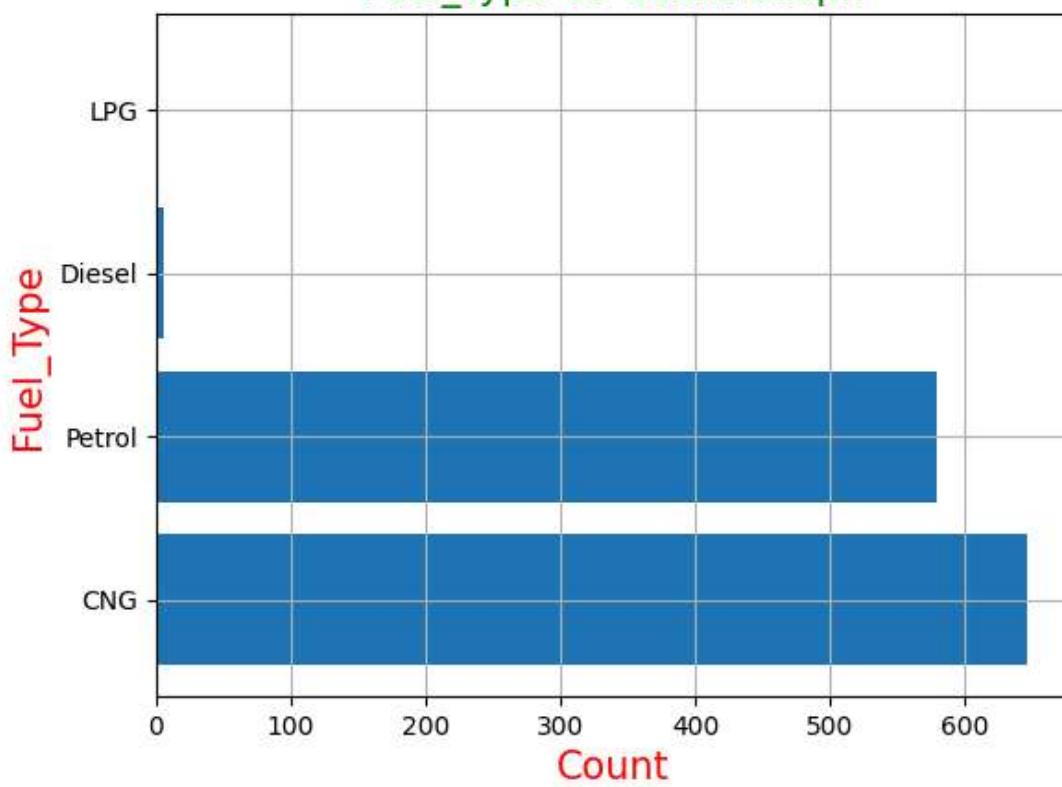
```
Location : Location
Mumbai      790
Hyderabad   742
Kochi       651
Coimbatore  636
Pune        622
Delhi        554
Kolkata     535
Chennai     494
Jaipur      413
Bangalore   358
Ahmedabad   224
Name: count, dtype: int64
*****
Fuel_Type : Fuel_Type
Diesel      3205
Petrol      2746
CNG         56
LPG         10
Electric     2
Name: count, dtype: int64
*****
Transmission : Transmission
Manual      4299
Automatic   1720
Name: count, dtype: int64
*****
Owner_Type : Owner_Type
First       4929
Second      968
Third       113
Fourth & Above  9
Name: count, dtype: int64
*****
```

```
In [ ]: for i in lst1:
    plt.barh(df2[i].unique(),df2[i].value_counts())
    plt.title(f'{i} Vs Count Graph',fontsize=15,color='g')
    plt.xlabel('Count',fontsize=15,color='r')
    plt.ylabel(i,fontsize=15,color='r')
    plt.grid(True)
    print('*'*100)
    plt.show()
```

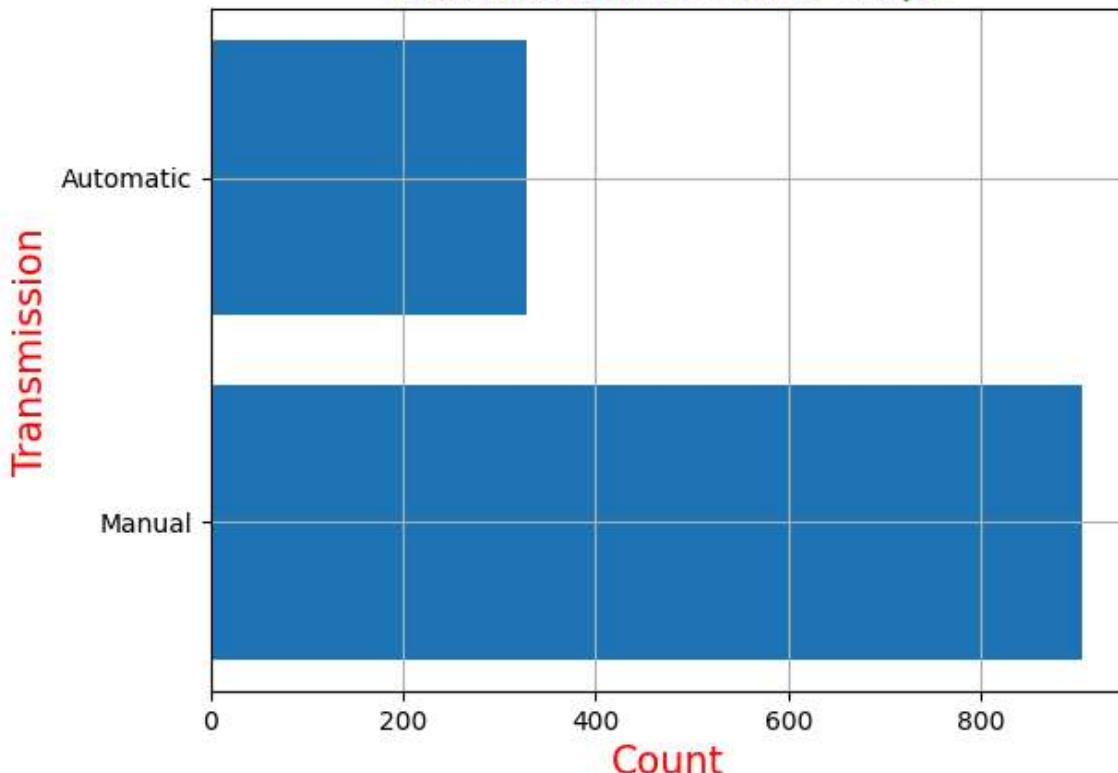
Location Vs Count Graph



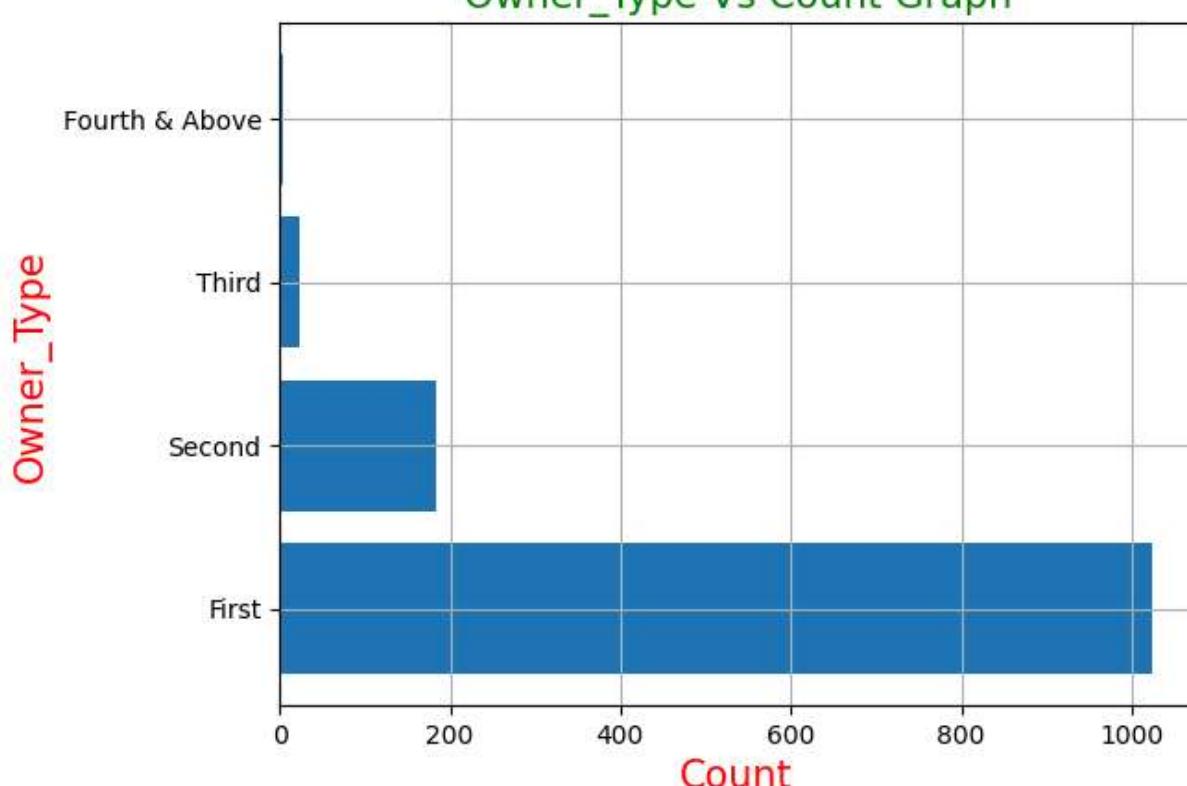
### Fuel\_Type Vs Count Graph



### Transmission Vs Count Graph



### Owner\_Type Vs Count Graph



```
In [ ]: df2['Brand']=df['Name'].apply(lambda x:x.split()[0])
df2['Brand']
```

```
Out[ ]: 0      Maruti
1      Hyundai
2      Honda
3      Maruti
4      Audi
...
1229    Chevrolet
1230      Toyota
1231    Mahindra
1232      Hyundai
1233  Mercedes-Benz
Name: Brand, Length: 1234, dtype: object
```

```
In [ ]: df3=pd.get_dummies(df2[['Location','Fuel_Type','Transmission','Owner_Type','Brand']],dtype=int,drop_first=True)
```

```
In [ ]: dft=pd.concat([df2,df3],axis=1)
```

```
dft
```

		Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	...	Brand_Mitsubishi	Brand_Nissan	Brand_Pc
0	0	Maruti Alto K10 LXI CNG	Delhi	2014		40929	CNG	Manual	First	32.26 km/kg	998 CC	...	0	0	0
1	1	Maruti Alto 800 2016-2019 LXI	Coimbatore	2013		54493	Petrol	Manual	Second	24.7 kmpl	796 CC	...	0	0	0
2	2	Toyota Innova Crysta Touring Sport 2.4 MT	Mumbai	2017		34000	Diesel	Manual	First	13.68 kmpl	2393 CC	...	0	0	0
3	3	Toyota Etios Liva GD	Hyderabad	2012		139000	Diesel	Manual	First	23.59 kmpl	1364 CC	...	0	0	0
4	4	Hyundai i20 Magna	Mumbai	2014		29000	Petrol	Manual	First	18.5 kmpl	1197 CC	...	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1229	1229	Volkswagen Vento Diesel Trendline	Hyderabad	2011		89411	Diesel	Manual	First	20.54 kmpl	1598 CC	...	0	0	0
1230	1230	Volkswagen Polo GT TSI	Mumbai	2015		59000	Petrol	Automatic	First	17.21 kmpl	1197 CC	...	0	0	0
1231	1231	Nissan Micra Diesel XV	Kolkata	2012		28000	Diesel	Manual	First	23.08 kmpl	1461 CC	...	0	0	0
1232	1232	Volkswagen Polo GT TSI	Pune	2013		52262	Petrol	Automatic	Third	17.2 kmpl	1197 CC	...	0	0	0
1233	1233	Mercedes-Benz E-Class 2009-2013 E 220 CDI Avan...	Kochi	2014		72443	Diesel	Automatic	First	10.0 kmpl	2148 CC	...	0	0	0

1234 rows × 56 columns

```
In [ ]: dft.drop(['Unnamed: 0','Name','Location','Fuel_Type','Transmission','Owner_Type','New_Price','Brand'],axis=1,inplace=True)
```

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Location_Bangalore	Location_Chennai	Location_Coimbatore	Location_Delhi	...	Brand_Mitsubishi	Brand_Nissan	Brand_Pc
0	2014	40929	32.26 km/kg	998 CC	58.2 bhp	4.0	0	0	0	0	1	...	0	0
1	2013	54493	24.7 kmpl	796 CC	47.3 bhp	5.0	0	0	0	1	0	...	0	0
2	2017	34000	13.68 kmpl	2393 CC	147.8 bhp	7.0	0	0	0	0	0	...	0	0
3	2012	139000	23.59 kmpl	1364 CC	null bhp	5.0	0	0	0	0	0	...	0	0
4	2014	29000	18.5 kmpl	1197 CC	82.85 bhp	5.0	0	0	0	0	0	...	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1229	2011	89411	20.54 kmpl	1598 CC	103.6 bhp	5.0	0	0	0	0	0	...	0	0
1230	2015	59000	17.21 kmpl	1197 CC	103.6 bhp	5.0	0	0	0	0	0	...	0	0
1231	2012	28000	23.08 kmpl	1461 CC	63.1 bhp	5.0	0	0	0	0	0	...	0	0
1232	2013	52262	17.2 kmpl	1197 CC	103.6 bhp	5.0	0	0	0	0	0	...	0	0
1233	2014	72443	10.0 kmpl	2148 CC	170 bhp	5.0	0	0	0	0	0	...	0	0

1234 rows × 48 columns

```
In [ ]: dft['Mileage']=dft['Mileage'].str.replace('km/kg',' ')
dft['Mileage']=dft['Mileage'].str.replace('kmpl',' ')
dft['Engine']=dft['Engine'].str.replace('CC',' ')
dft['Power']=dft['Power'].str.replace('bhp',' ')
dft['Mileage']=dft['Mileage'].str.replace('null','0')
dft['Engine']=dft['Engine'].str.replace('null','0')
dft['Power']=dft['Power'].str.replace('null','0')
```

```
In [ ]: dft['Mileage']=dft['Mileage'].astype(float)
dft['Power']=dft['Power'].astype(float)
dft['Engine']=dft['Engine'].astype(float)
dft.dtypes
```

```
Out[ ]: Year           int64
Kilometers_Driven      int64
Mileage                 float64
Engine                  float64
Power                   float64
Seats                   float64
Location_Bangalore     int64
Location_Chennai        int64
Location_Coimbatore    int64
Location_Delhi          int64
Location_Hyderabad     int64
Location_Jaipur         int64
Location_Kochi          int64
Location_Kolkata        int64
Location_Mumbai         int64
Location_Pune           int64
Fuel_Type_Diesel        int64
Fuel_Type_LPG            int64
Fuel_Type_Petrol         int64
Transmission_Manual    int64
Owner_Type_Fourth & Above int64
Owner_Type_Second       int64
Owner_Type_Third         int64
Brand_Audi               int64
Brand_BMW                int64
Brand_Chevrolet          int64
Brand_Datsun              int64
Brand_Fiat                int64
Brand_Ford                int64
Brand_Honda                int64
Brand_Hyundai              int64
Brand_Jaguar                int64
Brand_Jeep                  int64
Brand_Land                  int64
Brand_Mahindra              int64
Brand_Maruti                int64
Brand_Mercedes-Benz        int64
Brand_Mini                  int64
Brand_Mitsubishi             int64
Brand_Nissan                int64
Brand_Porsche                int64
Brand_Renault                int64
Brand_Skoda                  int64
Brand_Smart                  int64
Brand_Tata                  int64
Brand_Toyota                  int64
Brand_Volkswagen             int64
Brand_Volvo                  int64
dtype: object
```

```
In [ ]: dft.loc[dft.Engine==0, 'Engine']=np.NaN
dft.loc[dft.Mileage==0, 'Mileage']=np.NaN
dft.loc[dft.Power==0, 'Power']=np.NaN
```

```
In [ ]:
```

```
In [ ]: dft['Engine']=dft['Engine'].fillna(dft['Engine'].mean())
dft['Power']=dft['Power'].fillna(dft['Power'].mean())
dft['Mileage']=dft['Mileage'].fillna(dft['Mileage'].mean())
dft['Seats']=dft['Seats'].fillna(dft['Seats'].mode()[0])
```

```
In [ ]: dft.isna().sum()
```

```
Out[ ]: Year          0  
Kilometers_Driven      0  
Mileage          0  
Engine            0  
Power             0  
Seats             0  
Location_Bangalore    0  
Location_Chennai       0  
Location_Coimbatore    0  
Location_Delhi         0  
Location_Hyderabad     0  
Location_Jaipur        0  
Location_Kochi          0  
Location_Kolkata        0  
Location_Mumbai         0  
Location_Pune           0  
Fuel_Type_Diesel        0  
Fuel_Type_LPG           0  
Fuel_Type_Petrol         0  
Transmission_Manual     0  
Owner_Type_Fourth & Above 0  
Owner_Type_Second        0  
Owner_Type_Third          0  
Brand_Audi              0  
Brand_BMW               0  
Brand_Chevrolet         0  
Brand_Datsun            0  
Brand_Fiat              0  
Brand_Ford              0  
Brand_Honda             0  
Brand_Hyundai            0  
Brand_Jaguar             0  
Brand_Jeep               0  
Brand_Land              0  
Brand_Mahindra           0  
Brand_Maruti              0  
Brand_Mercedes-Benz      0  
Brand_Mini               0  
Brand_Mitsubishi          0  
Brand_Nissan              0  
Brand_Porsche             0  
Brand_Renault             0  
Brand_Skoda              0  
Brand_Smart               0  
Brand_Tata               0  
Brand_Toyota              0  
Brand_Volkswagen          0  
Brand_Volvo               0  
dtype: int64
```

```
In [ ]: dft.drop(['Brand_Chevrolet'],axis=1,inplace=True)
```

```
In [ ]: #model creation  
from sklearn.linear_model import LinearRegression  
model=LinearRegression()  
model.fit(x,y)  
y_pred=model.predict(dft)  
y_pred
```

```
Out[ ]: array([ 2.50786059, -0.04472804, 14.66981066, ..., -1.96269802,  
   6.51223448, 24.66480752])
```