

```
In [ ]: ▶ import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

df=pd.read_csv("/content/train-data.csv")
df
```

Out[1]:

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic
...	...	...	...	...	...	...	...
6014	6014	Maruti Swift VDI	Delhi	2014	27365	Diesel	Manual
6015	6015	Hyundai Xcent 1.1 CRDi S	Jaipur	2015	100000	Diesel	Manual
6016	6016	Mahindra Xylo D4 BSIV	Jaipur	2012	55000	Diesel	Manual
6017	6017	Maruti Wagon R VXI	Kolkata	2013	46000	Petrol	Manual
6018	6018	Chevrolet Beat Diesel	Hyderabad	2011	47000	Diesel	Manual

6019 rows × 14 columns



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

```
Out[2]:
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	Individual
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	Individual
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	Individual
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	Individual
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Individual

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

```
Out[3]:
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type
6014	6014	Maruti Swift VDI	Delhi	2014	27365	Diesel	Manual	Individual
6015	6015	Hyundai Xcent 1.1 CRDi S	Jaipur	2015	100000	Diesel	Manual	Individual
6016	6016	Mahindra Xylo D4 BSIV	Jaipur	2012	55000	Diesel	Manual	Individual
6017	6017	Maruti Wagon R VXI	Kolkata	2013	46000	Petrol	Manual	Individual
6018	6018	Chevrolet Beat Diesel	Hyderabad	2011	47000	Diesel	Manual	Individual

```
In [ ]: df.columns
```

```
Out[4]: Index(['Unnamed: 0', 'Name', 'Location', 'Year', 'Kilometers_Driven', 'Fuel_Type', 'Transmission', 'Owner_Type', 'Mileage', 'Engine', 'Power', 'Seats', 'New_Price', 'Price'], dtype='object')
```

In [ ]: ▶

In [ ]: ▶ `df.isna().sum()`

```
Out[5]: 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
```

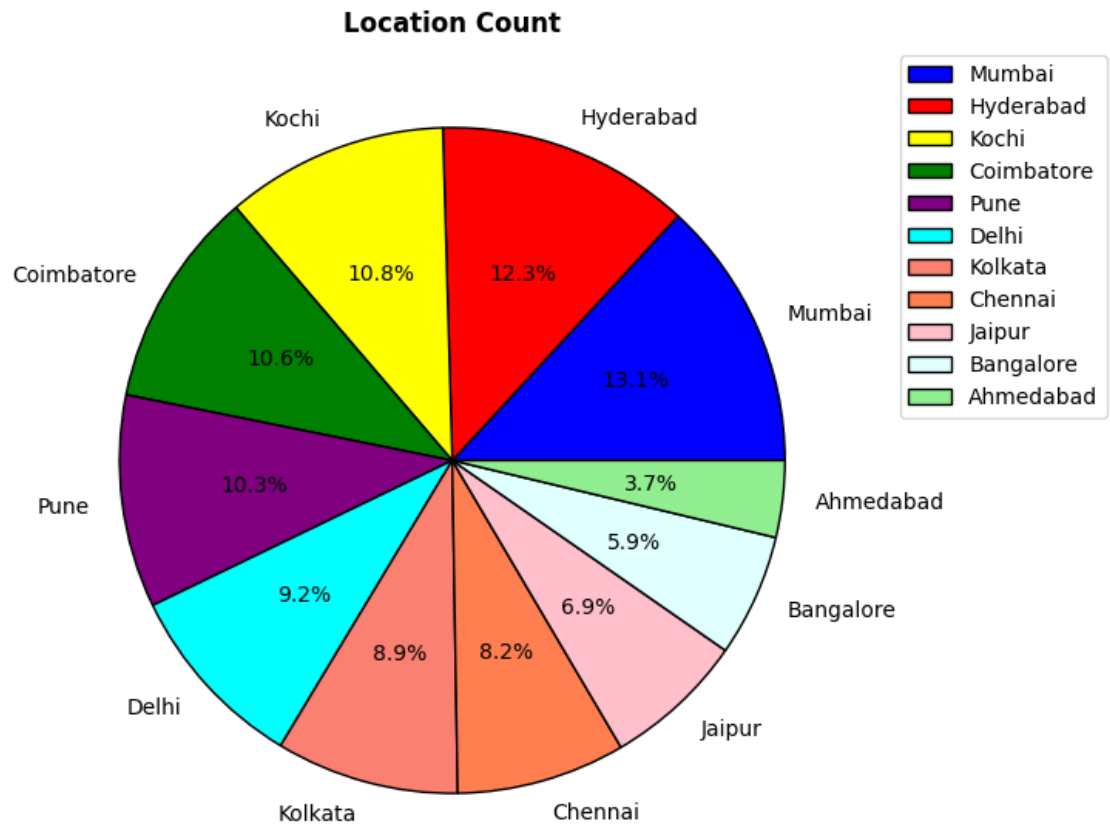
In [ ]: ▶ `df.dtypes`

```
Out[6]: 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 [ ]: ▶ `df['Location'].value_counts()`

```
Out[7]: Mumbai      790
        Hyderabad    742
        Kochi        651
        Coimbatore    636
        Pune         622
        Delhi        554
        Kolkata      535
        Chennai      494
        Jaipur       413
        Bangalore    358
        Ahmedabad    224
        Name: Location, dtype: int64
```

```
In [ ]: ▶ count_lo = df['Location'].value_counts()
plt.figure(figsize=(7,7))
plt.pie(count_lo, labels=count_lo.index, autopct='%1.1f%%',
        colors=['blue', 'red', 'yellow', 'green', 'purple', 'cyan', 'salmon', 'lightblue', 'pink', 'lightgreen'],
        wedgeprops=dict(linewidth=1, edgecolor='black'))
plt.legend(loc='upper right', bbox_to_anchor=(1.3, 1))
plt.title("Location Count", fontweight="bold")
plt.show()
```

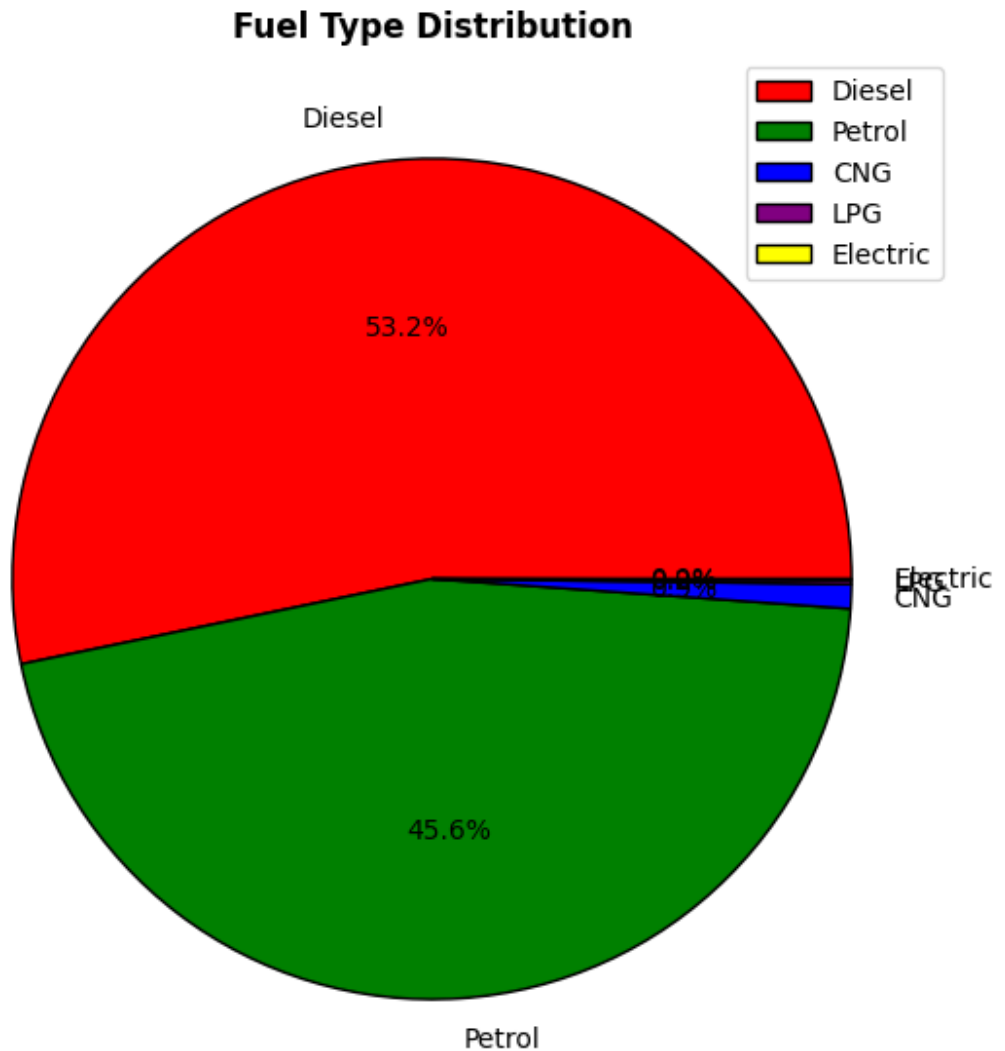


```
In [ ]: ▶ df['Fuel_Type'].value_counts()
```

```
Out[9]: Diesel      3205
        Petrol      2746
        CNG         56
        LPG         10
        Electric     2
        Name: Fuel_Type, dtype: int64
```

```
In [ ]: fuel_type_counts = df['Fuel_Type'].value_counts()

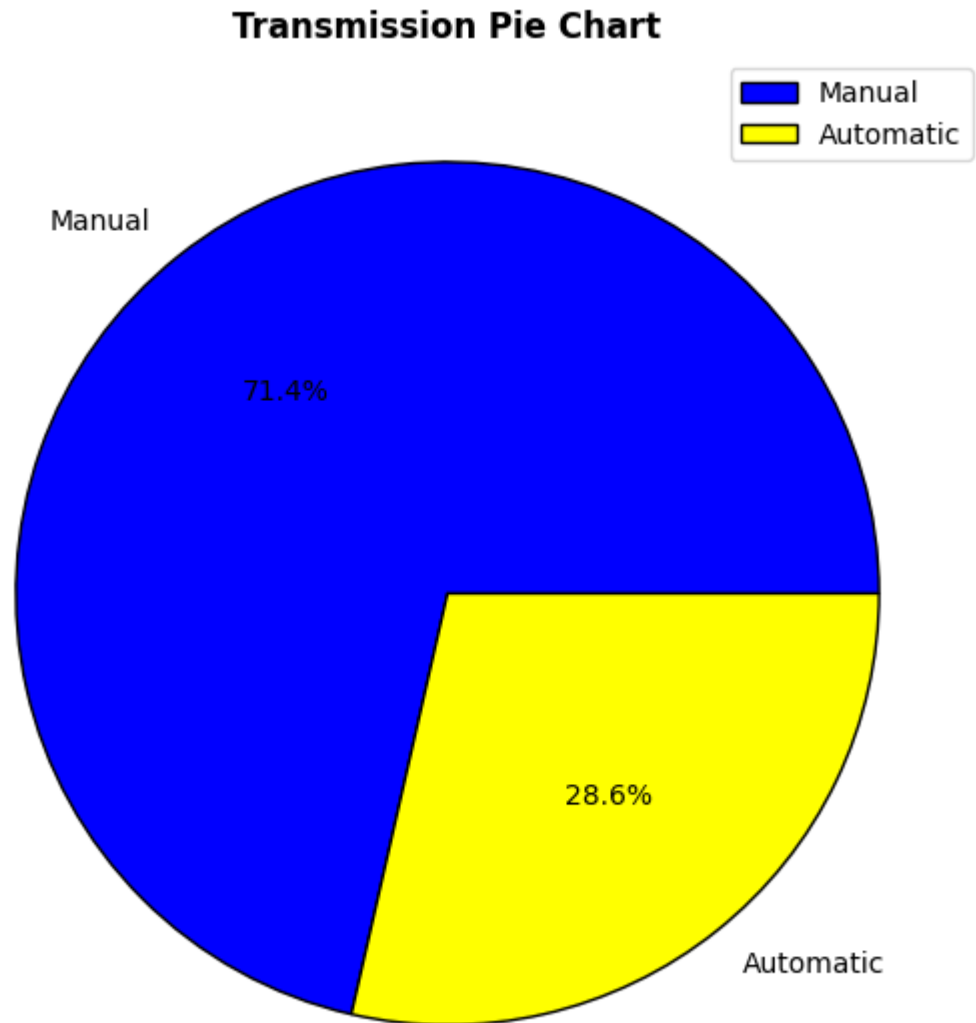
plt.figure(figsize=(7, 7))
plt.pie(fuel_type_counts, labels=fuel_type_counts.index, autopct='%1.1f%%',
        wedgeprops=dict(linewidth=1, edgecolor='black'))
plt.legend(loc='upper right', bbox_to_anchor=(1, 1))
plt.title("Fuel Type Distribution", fontweight='bold')
plt.show()
```



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

```
Out[11]: Manual      4299
Automatic    1720
Name: Transmission, dtype: int64
```

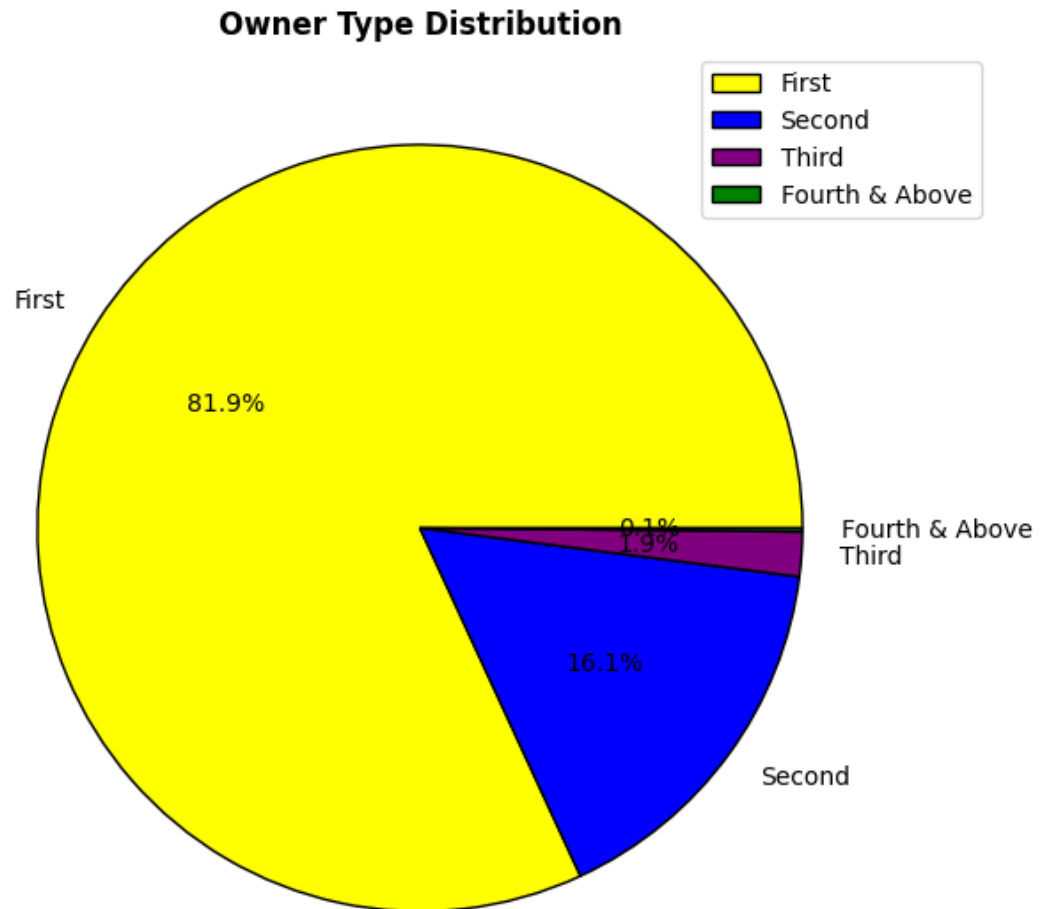
```
In [ ]: count_trans=df['Transmission'].value_counts()
plt.figure(figsize=(7,7))
plt.pie(count_trans,labels=count_trans.index,autopct='%1.1f%%',colors=['b','y'],
        wedgeprops=dict(linewidth=1,edgecolor='black'))
plt.legend(loc='upper right',bbox_to_anchor=(1,1))
plt.title('Transmission Pie Chart',fontweight='bold')
plt.show()
```



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

```
Out[13]: First          4929
Second         968
Third          113
Fourth & Above    9
Name: Owner_Type, dtype: int64
```

```
In [ ]: ▶ count_own=df['Owner_Type'].value_counts()
plt.figure(figsize=(7,7))
plt.pie(count_own,labels=count_own.index,autopct='%1.1f%%',colors=['yellow',
plt.title('Owner Type Distribution',fontweight='bold')
plt.legend(loc='upper right',bbox_to_anchor=(1.1,1))
plt.show()
```



```
In [ ]: ▶ #get_dummies encoding ..not in machine learning so import from pandas
df1=pd.get_dummies(df[['Location','Fuel_Type','Transmission','Owner_Type']]
df1
```

Out[15]:

	Location_Bangalore	Location_Chennai	Location_Coimbatore	Location_Delhi	Location
0	0	0	0	0	
1	0	0	0	0	
2	0	1	0	0	
3	0	1	0	0	
4	0	0	1	0	
...	...	...	...	...	...
6014	0	0	0	1	
6015	0	0	0	0	
6016	0	0	0	0	
6017	0	0	0	0	
6018	0	0	0	0	

6019 rows × 18 columns





In [ ]: `#need to join df +df1 concatenate`

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

Out[16]:

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic
...	...	...	...	...	...	...	...
6014	6014	Maruti Swift VDI	Delhi	2014	27365	Diesel	Manual
6015	6015	Hyundai Xcent 1.1 CRDi S	Jaipur	2015	100000	Diesel	Manual
6016	6016	Mahindra Xylo D4 BSIV	Jaipur	2012	55000	Diesel	Manual
6017	6017	Maruti Wagon R VXI	Kolkata	2013	46000	Petrol	Manual
6018	6018	Chevrolet Beat Diesel	Hyderabad	2011	47000	Diesel	Manual

6019 rows × 32 columns



In [ ]: `df.columns`

Out[17]: Index(['Unnamed: 0', 'Name', 'Location', 'Year', 'Kilometers\_Driven', 'Fuel\_Type', 'Transmission', 'Owner\_Type', 'Mileage', 'Engine', 'Power', 'Seats', 'New\_Price', 'Price'], dtype='object')

```
In [ ]: dfe.drop(['Unnamed: 0', 'Name', 'Location', 'Fuel_Type', 'Transmission', 'Owner', 'dfe'])
```

Out[18]:

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

6019 rows × 24 columns

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

```
Out[19]: Year                int64
Kilometers_Driven          int64
Mileage                    object
Engine                    object
Power                     object
Seats                    float64
Price                    float64
Location_Bangalore         uint8
Location_Chennai           uint8
Location_Coimbatore        uint8
Location_Delhi             uint8
Location_Hyderabad         uint8
Location_Jaipur            uint8
Location_Kochi             uint8
Location_Kolkata           uint8
Location_Mumbai            uint8
Location_Pune              uint8
Fuel_Type_Diesel           uint8
Fuel_Type_LPG              uint8
Fuel_Type_Petrol           uint8
Transmission_Manual        uint8
Owner_Type_Fourth & Above  uint8
Owner_Type_Second          uint8
Owner_Type_Third           uint8
dtype: object
```

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

```
Out[20]: Year                int64
Kilometers_Driven          int64
Mileage                    float64
Engine                    float64
Power                      float64
Seats                      float64
Price                      float64
Location_Bangalore         uint8
Location_Chennai           uint8
Location_Coimbatore        uint8
Location_Delhi             uint8
Location_Hyderabad         uint8
Location_Jaipur            uint8
Location_Kochi             uint8
Location_Kolkata           uint8
Location_Mumbai            uint8
Location_Pune              uint8
Fuel_Type_Diesel           uint8
Fuel_Type_LPG              uint8
Fuel_Type_Petrol           uint8
Transmission_Manual        uint8
Owner_Type_Fourth & Above  uint8
Owner_Type_Second          uint8
Owner_Type_Third           uint8
dtype: object
```

```
In [ ]: ▶ #replace 0 to a numerical 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[22]: 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_LPG         0
Fuel_Type_Petrol       0
Transmission_Manual   0
Owner_Type_Fourth & Above 0
Owner_Type_Second     0
Owner_Type_Third      0
dtype: int64
```

```
In [ ]: ▶ #MISSING VALUE
```

```
dfe['Mileage']=dfe['Mileage'].fillna(dfe['Mileage'].mean()) #mean bcoz of
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[24]: 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_LPG     0
Fuel_Type_Petrol  0
Transmission_Manual 0
Owner_Type_Fourth & Above 0
Owner_Type_Second 0
Owner_Type_Third  0
dtype: int64
```

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

```
Out[25]:
```

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Location_Bangalore	Location_Chennai	Location_Coimbatore	Location_Delhi	Location_Hyderabad	Location_Jaipur	Location_Kochi	Location_Kolkata	Location_Mumbai	Location_Pune	Fuel_Type_Diesel	Fuel_Type_LPG	Fuel_Type_Petrol	Transmission_Manual	Owner_Type_Fourth & Above	Owner_Type_Second	Owner_Type_Third
0	2010	72000	26.60	998.0	58.16	5.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	2015	41000	19.67	1582.0	126.20	5.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2011	46000	18.20	1199.0	88.70	5.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	2012	87000	20.77	1248.0	88.76	7.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	2013	40670	15.20	1968.0	140.80	5.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
6014	2014	27365	28.40	1248.0	74.00	5.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6015	2015	100000	24.40	1120.0	71.00	5.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6016	2012	55000	14.00	2498.0	112.00	8.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6017	2013	46000	18.90	998.0	67.10	5.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6018	2011	47000	25.44	936.0	57.60	5.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

6019 rows × 23 columns



```
In [ ]: ▶ y=dfe['Price']  
y
```

```
Out[26]: 0      1.75  
1     12.50  
2      4.50  
3      6.00  
4     17.74  
      ...  
6014    4.75  
6015    4.00  
6016    2.90  
6017    2.65  
6018    2.50  
Name: Price, Length: 6019, dtype: float64
```

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

Out[27]:

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission
0	0	Maruti Alto K10 LXI CNG	Delhi	2014	40929	CNG	Manual
1	1	Maruti Alto 800 2016-2019 LXI	Coimbatore	2013	54493	Petrol	Manual
2	2	Toyota Innova Crysta Touring Sport 2.4 MT	Mumbai	2017	34000	Diesel	Manual
3	3	Toyota Etios Liva GD	Hyderabad	2012	139000	Diesel	Manual
4	4	Hyundai i20 Magna	Mumbai	2014	29000	Petrol	Manual
...	...	...	...	...	...	...	...
1229	1229	Volkswagen Vento Diesel Trendline	Hyderabad	2011	89411	Diesel	Manual
1230	1230	Volkswagen Polo GT TSI	Mumbai	2015	59000	Petrol	Automatic
1231	1231	Nissan Micra Diesel XV	Kolkata	2012	28000	Diesel	Manual
1232	1232	Volkswagen Polo GT TSI	Pune	2013	52262	Petrol	Automatic
1233	1233	Mercedes-Benz E-Class 2009-2013 E 220 CDI Avan...	Kochi	2014	72443	Diesel	Automatic

1234 rows × 13 columns





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

Out[28]:

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Own
0	0	Maruti Alto K10 LXI CNG	Delhi	2014	40929	CNG	Manual	
1	1	Maruti Alto 800 2016-2019 LXI	Coimbatore	2013	54493	Petrol	Manual	
2	2	Toyota Innova Crysta Touring Sport 2.4 MT	Mumbai	2017	34000	Diesel	Manual	
3	3	Toyota Etios Liva GD	Hyderabad	2012	139000	Diesel	Manual	
4	4	Hyundai i20 Magna	Mumbai	2014	29000	Petrol	Manual	

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

Out[29]:

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Own
1229	1229	Volkswagen Vento Diesel Trendline	Hyderabad	2011	89411	Diesel	Manual	
1230	1230	Volkswagen Polo GT TSI	Mumbai	2015	59000	Petrol	Automatic	
1231	1231	Nissan Micra Diesel XV	Kolkata	2012	28000	Diesel	Manual	
1232	1232	Volkswagen Polo GT TSI	Pune	2013	52262	Petrol	Automatic	
1233	1233	Mercedes-Benz E-Class 2009-2013 E 220 CDI Avan...	Kochi	2014	72443	Diesel	Automatic	

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

```
Out[30]: 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
```

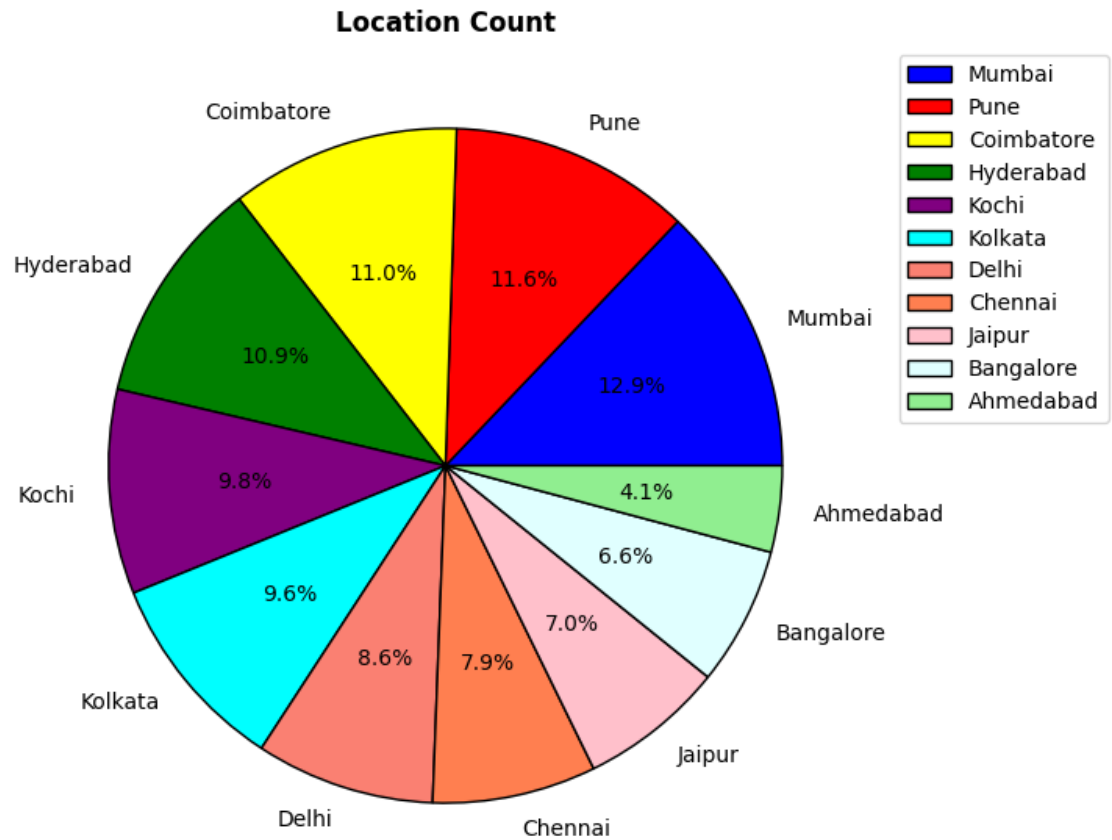
```
In [ ]: df.isna().sum()
```

```
Out[31]: 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 [ ]: loc=df['Location'].value_counts()  
loc
```

```
Out[32]: Mumbai      159  
Pune      143  
Coimbatore  136  
Hyderabad  134  
Kochi      121  
Kolkata    119  
Delhi      106  
Chennai    97  
Jaipur     86  
Bangalore  82  
Ahmedabad  51  
Name: Location, dtype: int64
```

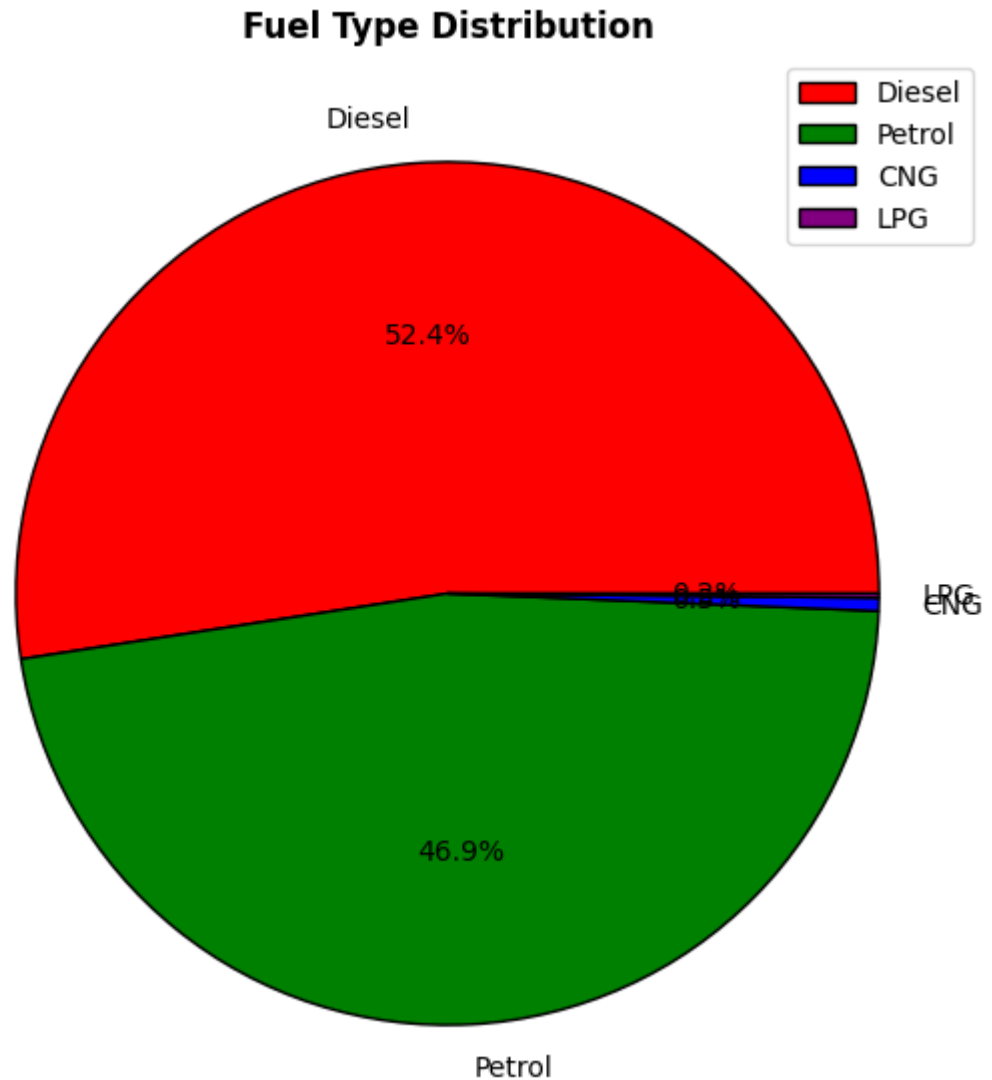
```
In [ ]: ▶ plt.figure(figsize=(7,7))
plt.pie(loc,labels=loc.index,autopct='%1.1f%%', colors=['blue', 'red', 'yellow', 'green', 'purple', 'cyan', 'red', 'orange', 'pink', 'lightblue', 'lightgreen'],
        wedgeprops=dict(linewidth=1, edgecolor='black'))
plt.legend(loc='upper right',bbox_to_anchor=(1.3,1))
plt.title("Location Count",fontweight="bold")
plt.show()
```



```
In [ ]: ▶ fuel=df['Fuel_Type'].value_counts()
fuel
```

```
Out[34]: Diesel      647
         Petrol      579
         CNG          6
         LPG          2
         Name: Fuel_Type, dtype: int64
```

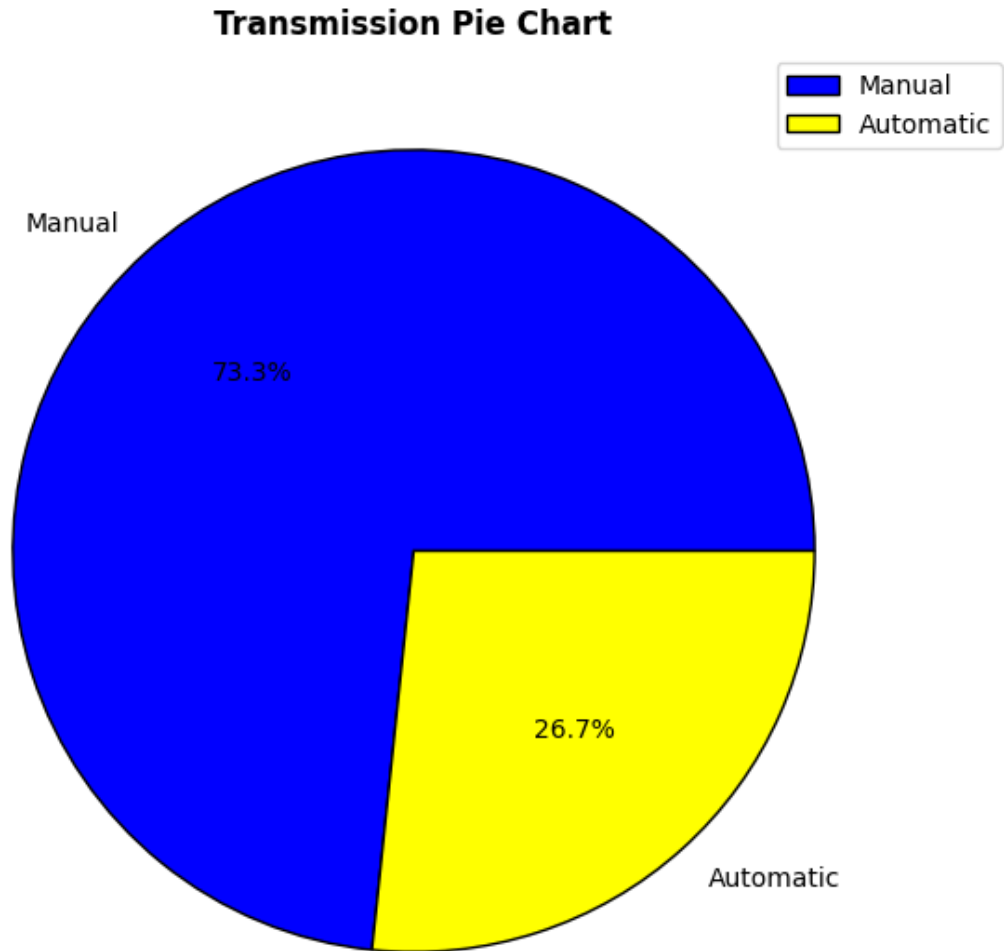
```
In [ ]: ▶ plt.figure(figsize=(7,7))
plt.pie(fuel,labels=fuel.index,autopct='%1.1f%%',colors=['red', 'green', 'blue', 'purple'],
        wedgeprops=dict(linewidth=1, edgecolor='black'))
plt.legend(loc='upper right', bbox_to_anchor=(1, 1))
plt.title("Fuel Type Distribution",fontweight='bold')
plt.show()
```



```
In [ ]: ▶ trans=df['Transmission'].value_counts()
trans
```

```
Out[36]: Manual      905
Automatic    329
Name: Transmission, dtype: int64
```

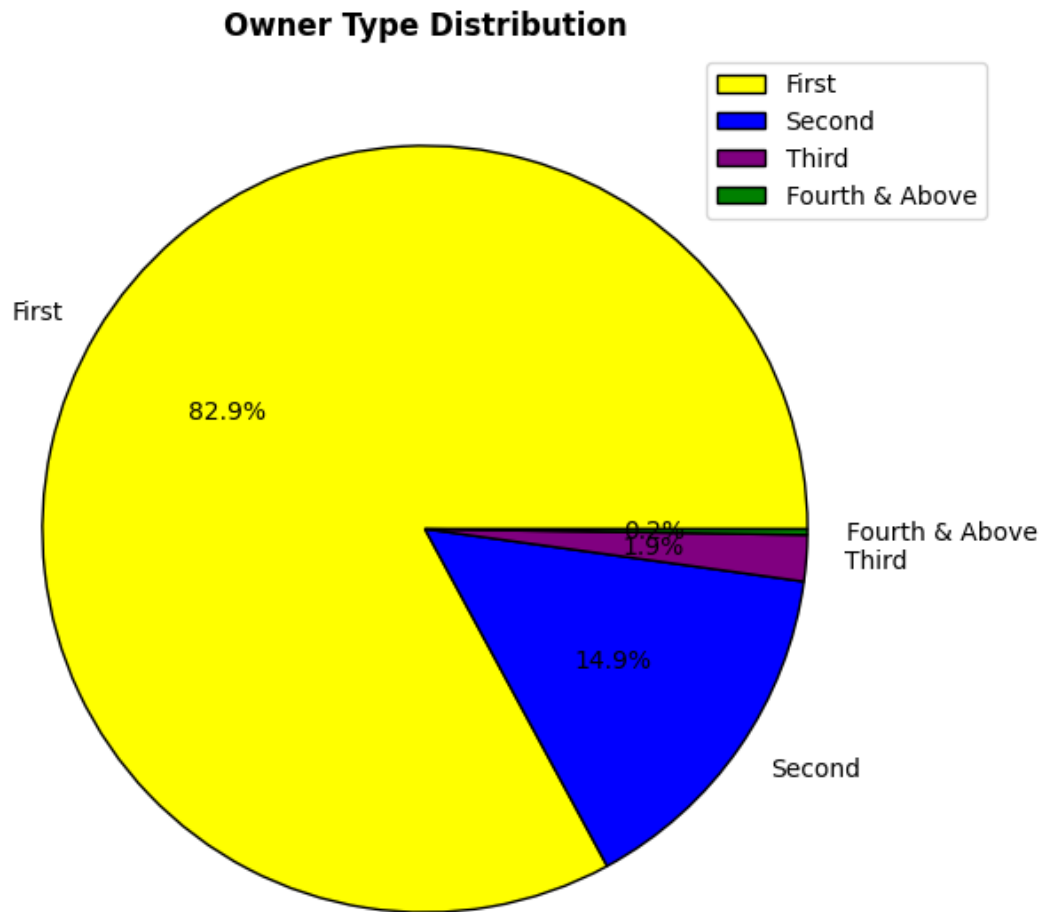
```
In [ ]: ▶ plt.figure(figsize=(7,7))
plt.pie(trans,labels=trans.index,autopct='%1.1f%',colors=['blue','yellow']
        wedgeprops=dict(linewidth=1,edgecolor='black'))
plt.legend(loc='upper right',bbox_to_anchor=(1.1,1))
plt.title('Transmission Pie Chart',fontweight='bold')
plt.show()
```



```
In [ ]: ▶ own=df['Owner_Type'].value_counts()
own
```

```
Out[38]: First          1023
Second         184
Third           24
Fourth & Above    3
Name: Owner_Type, dtype: int64
```

```
In [ ]: ▶ plt.figure(figsize=(7,7))
plt.pie(own,labels=own.index,autopct='%1.1f%%',colors=['yellow','blue','pu
plt.title('Owner Type Distribution',fontweight='bold')
plt.legend(loc='upper right',bbox_to_anchor=(1.1,1))
plt.show()
```



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

```
Out[40]: Maruti Alto LXi 9
Honda City 1.5 V MT 8
Maruti Swift Dzire VDI 8
Volkswagen Polo 1.2 MPI Highline 8
Hyundai i10 Magna 7
..
Hyundai Santro GLS I - Euro II 1
Honda City i DTEC VX Option BL 1
Land Rover Discovery 4 SDV6 SE 1
Hyundai Verna CRDi 1.6 SX Option 1
Mercedes-Benz E-Class 2009-2013 E 220 CDI Avantgarde 1
Name: Name, Length: 769, dtype: int64
```

```
In [ ]: ▶ #Encoding
#Get_dummies
df1=pd.get_dummies(df[['Location','Fuel_Type','Transmission','Owner_Type']]
df1
```

Out[41]:

	Location_Bangalore	Location_Chennai	Location_Coimbatore	Location_Delhi	Location_Mumbai
0	0	0	0	1	0
1	0	0	1	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
...	...	...	...	...	...
1229	0	0	0	0	0
1230	0	0	0	0	0
1231	0	0	0	0	0
1232	0	0	0	0	0
1233	0	0	0	0	0

1234 rows × 17 columns



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

Out[42]:

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission
0	0	Maruti Alto K10 LXI CNG	Delhi	2014	40929	CNG	Manual
1	1	Maruti Alto 800 2016- 2019 LXI	Coimbatore	2013	54493	Petrol	Manual
2	2	Toyota Innova Crysta Touring Sport 2.4 MT	Mumbai	2017	34000	Diesel	Manual
3	3	Toyota Etios Liva GD	Hyderabad	2012	139000	Diesel	Manual
4	4	Hyundai i20 Magna	Mumbai	2014	29000	Petrol	Manual
...	...	...	...	...	...	...	...
1229	1229	Volkswagen Vento Diesel Trendline	Hyderabad	2011	89411	Diesel	Manual
1230	1230	Volkswagen Polo GT TSI	Mumbai	2015	59000	Petrol	Automatic
1231	1231	Nissan Micra Diesel XV	Kolkata	2012	28000	Diesel	Manual
1232	1232	Volkswagen Polo GT TSI	Pune	2013	52262	Petrol	Automatic
1233	1233	Mercedes- Benz E- Class 2009- 2013 E 220 CDI Avan...	Kochi	2014	72443	Diesel	Automatic

1234 rows × 30 columns





```
In [ ]: dfe.drop(['Unnamed: 0', 'Name', 'Location', 'Fuel_Type', 'Transmission', 'Owner', 'dfe
```

Out[43]:

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

1234 rows × 23 columns



```
In [ ]: ▶ dfe.dtypes
```

```
Out[44]: Year                int64
Kilometers_Driven          int64
Mileage                    object
Engine                    object
Power                     object
Seats                    float64
Location_Bangalore         uint8
Location_Chennai           uint8
Location_Coimbatore        uint8
Location_Delhi             uint8
Location_Hyderabad         uint8
Location_Jaipur            uint8
Location_Kochi             uint8
Location_Kolkata           uint8
Location_Mumbai            uint8
Location_Pune              uint8
Fuel_Type_Diesel           uint8
Fuel_Type_LPG              uint8
Fuel_Type_Petrol           uint8
Transmission_Manual        uint8
Owner_Type_Fourth & Above  uint8
Owner_Type_Second          uint8
Owner_Type_Third           uint8
dtype: object
```

```
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[45]:

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

1234 rows × 23 columns



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

```
In [ ]: ▶ dfe.dtypes
```

```
Out[47]: Year                int64
Kilometers_Driven          int64
Mileage                    object
Engine                    object
Power                     object
Seats                    float64
Location_Bangalore         uint8
Location_Chennai           uint8
Location_Coimbatore        uint8
Location_Delhi             uint8
Location_Hyderabad         uint8
Location_Jaipur            uint8
Location_Kochi             uint8
Location_Kolkata           uint8
Location_Mumbai            uint8
Location_Pune              uint8
Fuel_Type_Diesel           uint8
Fuel_Type_LPG              uint8
Fuel_Type_Petrol           uint8
Transmission_Manual        uint8
Owner_Type_Fourth & Above  uint8
Owner_Type_Second          uint8
Owner_Type_Third           uint8
dtype: object
```

```
In [ ]: ▶ #data type conversion
dfe['Mileage']=dfe['Mileage'].astype(float)
dfe['Engine']=dfe['Engine'].astype(float)
dfe['Power']=dfe['Power'].astype(float)
dfe.dtypes
```

```
Out[48]: Year                int64
Kilometers_Driven          int64
Mileage                    float64
Engine                    float64
Power                     float64
Seats                     float64
Location_Bangalore         uint8
Location_Chennai           uint8
Location_Coimbatore        uint8
Location_Delhi             uint8
Location_Hyderabad         uint8
Location_Jaipur            uint8
Location_Kochi             uint8
Location_Kolkata           uint8
Location_Mumbai            uint8
Location_Pune              uint8
Fuel_Type_Diesel           uint8
Fuel_Type_LPG              uint8
Fuel_Type_Petrol           uint8
Transmission_Manual        uint8
Owner_Type_Fourth & Above  uint8
Owner_Type_Second          uint8
Owner_Type_Third           uint8
dtype: object
```

```
In [ ]: ▶ #replace numerical 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['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['Power'].mode()[0])
```

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

```
Out[51]: 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
dtype: int64
```

```
In [ ]: dfe
```

```
Out[52]:
```

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Location_Bangalore	Loc
0	2014	40929	32.26	998.0	58.20000	4.0	0	
1	2013	54493	24.70	796.0	47.30000	5.0	0	
2	2017	34000	13.68	2393.0	147.80000	7.0	0	
3	2012	139000	23.59	1364.0	110.38042	5.0	0	
4	2014	29000	18.50	1197.0	82.85000	5.0	0	
...	...	...	...	...	...	...	...	...
1229	2011	89411	20.54	1598.0	103.60000	5.0	0	
1230	2015	59000	17.21	1197.0	103.60000	5.0	0	
1231	2012	28000	23.08	1461.0	63.10000	5.0	0	
1232	2013	52262	17.20	1197.0	103.60000	5.0	0	
1233	2014	72443	10.00	2148.0	170.00000	5.0	0	

1234 rows × 23 columns



```
In [ ]: dfe.dtypes
```

```
Out[53]: Year                int64
Kilometers_Driven          int64
Mileage                    float64
Engine                    float64
Power                     float64
Seats                     float64
Location_Bangalore         uint8
Location_Chennai           uint8
Location_Coimbatore        uint8
Location_Delhi             uint8
Location_Hyderabad         uint8
Location_Jaipur            uint8
Location_Kochi             uint8
Location_Kolkata           uint8
Location_Mumbai            uint8
Location_Pune              uint8
Fuel_Type_Diesel           uint8
Fuel_Type_LPG              uint8
Fuel_Type_Petrol           uint8
Transmission_Manual        uint8
Owner_Type_Fourth & Above  uint8
Owner_Type_Second          uint8
Owner_Type_Third           uint8
dtype: object
```

```
In [ ]: z=dfe
z
```

```
Out[54]:
```

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Location_Bangalore	Loc
0	2014	40929	32.26	998.0	58.20000	4.0	0	
1	2013	54493	24.70	796.0	47.30000	5.0	0	
2	2017	34000	13.68	2393.0	147.80000	7.0	0	
3	2012	139000	23.59	1364.0	110.38042	5.0	0	
4	2014	29000	18.50	1197.0	82.85000	5.0	0	
...	...	...	...	...	...	...	...	...
1229	2011	89411	20.54	1598.0	103.60000	5.0	0	
1230	2015	59000	17.21	1197.0	103.60000	5.0	0	
1231	2012	28000	23.08	1461.0	63.10000	5.0	0	
1232	2013	52262	17.20	1197.0	103.60000	5.0	0	
1233	2014	72443	10.00	2148.0	170.00000	5.0	0	

1234 rows × 23 columns



In [ ]: x

Out[55]:

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Location_Bangalore	Location_Chennai	Location_Coimbatore	Location_Delhi	Location_Hyderabad	Location_Jaipur	Location_Kochi	Location_Kolkata	Location_Mumbai	Location_Pune	Fuel_Type_Diesel	Fuel_Type_LPG	Fuel_Type_Petrol	Transmission_Manual	Owner_Type_Fourth & Above	Owner_Type_Second	Owner_Type_Third
0	2010	72000	26.60	998.0	58.16	5.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	2015	41000	19.67	1582.0	126.20	5.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2011	46000	18.20	1199.0	88.70	5.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	2012	87000	20.77	1248.0	88.76	7.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	2013	40670	15.20	1968.0	140.80	5.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
6014	2014	27365	28.40	1248.0	74.00	5.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6015	2015	100000	24.40	1120.0	71.00	5.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6016	2012	55000	14.00	2498.0	112.00	8.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6017	2013	46000	18.90	998.0	67.10	5.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6018	2011	47000	25.44	936.0	57.60	5.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

6019 rows × 23 columns

In [ ]: x.dtypes

Out[56]:

Year	int64
Kilometers_Driven	int64
Mileage	float64
Engine	float64
Power	float64
Seats	float64
Location_Bangalore	uint8
Location_Chennai	uint8
Location_Coimbatore	uint8
Location_Delhi	uint8
Location_Hyderabad	uint8
Location_Jaipur	uint8
Location_Kochi	uint8
Location_Kolkata	uint8
Location_Mumbai	uint8
Location_Pune	uint8
Fuel_Type_Diesel	uint8
Fuel_Type_LPG	uint8
Fuel_Type_Petrol	uint8
Transmission_Manual	uint8
Owner_Type_Fourth & Above	uint8
Owner_Type_Second	uint8
Owner_Type_Third	uint8
dtype:	object



In [ ]: ▶ y

```
Out[57]: 0      1.75
          1     12.50
          2      4.50
          3      6.00
          4     17.74
          ...
        6014    4.75
        6015    4.00
        6016    2.90
        6017    2.65
        6018    2.50
        Name: Price, Length: 6019, dtype: float64
```

In [ ]: ▶ z

```
Out[58]:
```

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Location_Bangalore	Loc
0	2014	40929	32.26	998.0	58.20000	4.0		0
1	2013	54493	24.70	796.0	47.30000	5.0		0
2	2017	34000	13.68	2393.0	147.80000	7.0		0
3	2012	139000	23.59	1364.0	110.38042	5.0		0
4	2014	29000	18.50	1197.0	82.85000	5.0		0
...	...	...	...	...	...	...		...
1229	2011	89411	20.54	1598.0	103.60000	5.0		0
1230	2015	59000	17.21	1197.0	103.60000	5.0		0
1231	2012	28000	23.08	1461.0	63.10000	5.0		0
1232	2013	52262	17.20	1197.0	103.60000	5.0		0
1233	2014	72443	10.00	2148.0	170.00000	5.0		0

1234 rows × 23 columns



```
In [ ]: ▶ z.dtypes
```

```
Out[59]: Year                int64
Kilometers_Driven          int64
Mileage                    float64
Engine                    float64
Power                     float64
Seats                     float64
Location_Bangalore         uint8
Location_Chennai           uint8
Location_Coimbatore        uint8
Location_Delhi             uint8
Location_Hyderabad         uint8
Location_Jaipur            uint8
Location_Kochi             uint8
Location_Kolkata           uint8
Location_Mumbai            uint8
Location_Pune              uint8
Fuel_Type_Diesel           uint8
Fuel_Type_LPG              uint8
Fuel_Type_Petrol           uint8
Transmission_Manual        uint8
Owner_Type_Fourth & Above  uint8
Owner_Type_Second          uint8
Owner_Type_Third           uint8
dtype: object
```

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

```
Out[60]: array([ 2.87588492, -1.29344912, 16.1069494 , ...,  0.1378514 ,
                9.27293255, 21.48043251])
```

```
In [ ]: ▶ print(model.predict)
```

```
<bound method LinearModel.predict of LinearRegression(>
```