In [1]:
```
import pandas as pd
```

In [2]:
```
data = pd.read_csv('C:/Users/91830/Downloads/train.csv')
```

In [3]:
```
data.head()
```

Out[3]:

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission
0	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual
1	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual
2	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual
3	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic
4	6	Nissan Micra Diesel XV	Jaipur	2013	86999	Diesel	Manual

In [4]:
```
#Check for null values
```

In [5]:
```
data.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                38
New_Price          5032
Price                 0
dtype: int64
```

In [6]:
```
d2 = pd.DataFrame(data)
```

In [7]:
```
# (b) Removed the units from Mileage, Engine, Power and New_Price
```

In [8]:
```
d2['Mileage'] = d2['Mileage'].str.extract('(\d+\.\d+)').astype(float)

d2['Engine'] = d2['Engine'].str.replace(' CC', '').astype(float)

d2['Power'] = d2['Power'].str.extract('(\d+\.\d+)').astype(float)

d2['New_Price'] = d2['New_Price'].str.extract('(\d+\.\d+)').astype(floa
```

In [9]:
```
# (a)Now we are looking for missing values
#Replace the NA values of Mileage, Engine and Power with median
#Median is the best replacement technique because it preserves the cent
#Replace the NA values of Seat with Mode.
```

In [10]:
```
d2['Mileage'].fillna(d2['Mileage'].median(),inplace=True)
d2['Engine'].fillna(d2['Engine'].median(),inplace=True)
d2['Power'].fillna(d2['Power'].median(),inplace=True)
d2['Seats'].fillna(d2['Seats'].mode()[0],inplace=True)

d2.drop('New_Price', axis=1, inplace=True)#axis=1 indicates to drop a c
```

In [11]:
```
output_csv_file = ('C:/Users/91830/Downloads/clean.csv')
d2.to_csv(output_csv_file, index=False)
```

In [12]:
```
data1=pd.read_csv('C:/Users/91830/Downloads/clean.csv')
```

In [13]:
```
data1.isna().sum()
```

Out[13]:
```
Unnamed: 0           0
Name                 0
Location             0
Year                 0
Kilometers_Driven    0
Fuel_Type            0
Transmission         0
Owner_Type           0
Mileage              0
Engine               0
Power                0
Seats                0
Price                0
dtype: int64
```

In [14]:
```
#Check for the unique values of Categorical variables: Fuel_Type, Trans
```

In [15]:
```
data1['Fuel_Type'].unique()
```

Out[15]:
```
array(['Diesel', 'Petrol', 'Electric'], dtype=object)
```

In [16]:
```
data1['Transmission'].unique()
```

Out[16]:
```
array(['Manual', 'Automatic'], dtype=object)
```

In [17]:
```
data1['Owner_Type'].unique()
```

Out[17]:
```
array(['First', 'Second', 'Fourth & Above', 'Third'], dtype=object)
```

In [18]:
```
data1['Fuel_Type'].replace({'Diesel': 0, 'Petrol': 1, 'Electric': 2}, i
data1['Transmission'].replace({'Manual': 0, 'Automatic': 1}, inplace=Tr
```

In [19]:
```
data1.head()
```

Out[19]:

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission
0	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	0	0
1	2	Honda Jazz V	Chennai	2011	46000	1	0
2	3	Maruti Ertiga VDI	Chennai	2012	87000	0	0
3	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	0	1
4	6	Nissan Micra Diesel XV	Jaipur	2013	86999	0	0

In [20]:
```
#Save encoded data in csv file
```

In [21]:
```
output_csv_file = ('C:/Users/91830/Downloads/clean.csv')
data1.to_csv(output_csv_file, index=False)
```

In [22]:
```
# (d) Add the column 'CURRENT AGE' OF THE CAR
```

In [23]:
```
from datetime import datetime
```

In [24]:
```
current_year = datetime.now().year

data1['Current_Age'] = current_year - data1['Year']
```

In [25]:
```
#Save the file with newly added Current_Age colmun in separate file
output_csv_file = ('C:/Users/91830/Downloads/clean.csv')
data1.to_csv(output_csv_file, index=False)
```

In [26]:
```
data3=pd.read_csv("C:/Users/91830/Downloads/clean.csv")
data3.head()
```

Out[26]:

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission
0	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	0	0
1	2	Honda Jazz V	Chennai	2011	46000	1	0
2	3	Maruti Ertiga VDI	Chennai	2012	87000	0	0
3	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	0	1
4	6	Nissan Micra Diesel XV	Jaipur	2013	86999	0	0

In [30]:

```
# (e) Selecting specific columns
selected_columns = data[['Name', 'Year', 'Kilometers_Driven', 'Fuel_Typ
print("Selected columns:")
print(selected_columns.head())

# (e) Filtering the dataset based on a condition
filtered_data = data[data['Fuel_Type'] == 0]  # Filtering only Diesel c
print("\nFiltered data (Diesel cars only):")
print(filtered_data.head())

# (e) Renaming columns
renamed_data = data.rename(columns={'Name': 'Car_Name', 'Kilometers_Dri
print("\nRenamed columns:")
print(renamed_data.head())

# (e) Mutating/Creating new columns
data['Price_in_lakhs'] = data['Price'] / 100000  # Converting Price to
print("\nMutated data (Added Price_in_lakhs column):")
print(data.head())

# (e) Arranging/Sorting the dataset based on a column
sorted_data = data.sort_values(by='Price', ascending=False)  # Sorting
print("\nSorted data (Based on Price in descending order):")
print(sorted_data.head())

# (e) Summarizing data with group by operations
grouped_data = data.groupby('Fuel_Type').agg({'Price': 'mean', 'Year':
print("\nGrouped data (Mean Price and Max Year by Fuel_Type):")
print(grouped_data)
```

```
Selected columns:
                              Name  Year  Kilometers_Driven Fuel_Type
\
0  Hyundai Creta 1.6 CRDi SX Option  2015              41000    Diesel
1                      Honda Jazz V  2011              46000    Petrol
2                 Maruti Ertiga VDI  2012              87000    Diesel
3   Audi A4 New 2.0 TDI Multitronic  2013              40670    Diesel
4            Nissan Micra Diesel XV  2013              86999    Diesel

  Transmission  Price
0       Manual  12.50
1       Manual   4.50
2       Manual   6.00
3    Automatic  17.74
4       Manual   3.50

Filtered data (Diesel cars only):
Empty DataFrame
Columns: [Unnamed: 0, Name, Location, Year, Kilometers_Driven, Fuel_Ty
pe, Transmission, Owner_Type, Mileage, Engine, Power, Seats, New_Pric
e, Price]
Index: []

Renamed columns:
   Unnamed: 0                          Car_Name    Location  Year  Kms
_Driven  \
0           1  Hyundai Creta 1.6 CRDi SX Option        Pune  2015
41000
1           2                      Honda Jazz V     Chennai  2011
46000
2           3                 Maruti Ertiga VDI     Chennai  2012
87000
3           4   Audi A4 New 2.0 TDI Multitronic  Coimbatore  2013
40670
4           6            Nissan Micra Diesel XV      Jaipur  2013
86999

  Fuel_Type Transmission Owner_Type     Mileage   Engine      Power  S
eats  \
0    Diesel       Manual      First  19.67 kmpl  1582 CC  126.2 bhp
5.0
1    Petrol       Manual      First    13 km/kg  1199 CC   88.7 bhp
5.0
2    Diesel       Manual      First  20.77 kmpl  1248 CC  88.76 bhp
7.0
3    Diesel    Automatic     Second   15.2 kmpl  1968 CC  140.8 bhp
5.0
4    Diesel       Manual      First  23.08 kmpl  1461 CC   63.1 bhp
5.0

   New_Price  Price
0        NaN  12.50
1  8.61 Lakh   4.50
2        NaN   6.00
3        NaN  17.74
4        NaN   3.50

Mutated data (Added Price_in_lakhs column):
   Unnamed: 0                              Name    Location  Year  \
0           1  Hyundai Creta 1.6 CRDi SX Option        Pune  2015
1           2                      Honda Jazz V     Chennai  2011
```

```
2           3                Maruti Ertiga VDI     Chennai  2012
3           4   Audi A4 New 2.0 TDI Multitronic  Coimbatore  2013
4           6            Nissan Micra Diesel XV      Jaipur  2013

   Kilometers_Driven Fuel_Type Transmission Owner_Type     Mileage   E
ngine  \
0              41000    Diesel       Manual      First  19.67 kmpl  15
82 CC
1              46000    Petrol       Manual      First    13 km/kg  11
99 CC
2              87000    Diesel       Manual      First  20.77 kmpl  12
48 CC
3              40670    Diesel    Automatic     Second   15.2 kmpl  19
68 CC
4              86999    Diesel       Manual      First  23.08 kmpl  14
61 CC

       Power  Seats  New_Price  Price  Price_in_lakhs
0  126.2 bhp    5.0        NaN  12.50        0.000125
1   88.7 bhp    5.0  8.61 Lakh   4.50        0.000045
2  88.76 bhp    7.0        NaN   6.00        0.000060
3  140.8 bhp    5.0        NaN  17.74        0.000177
4   63.1 bhp    5.0        NaN   3.50        0.000035

Sorted data (Based on Price in descending order):
      Unnamed: 0                                          Name    Locat
ion  \
3952        4079  Land Rover Range Rover 3.0 Diesel LWB Vogue   Hydera
bad
5620        5781                   Lamborghini Gallardo Coupe       De
lhi
5752        5919                       Jaguar F Type 5.0 V8 S   Hydera
bad
1457        1505             Land Rover Range Rover Sport SE       Ko
chi
1917        1974                           BMW 7 Series 740Li  Coimbat
ore

      Year  Kilometers_Driven Fuel_Type Transmission Owner_Type     Mi
leage  \
3952  2017              25000    Diesel    Automatic      First  13.33
kmpl
5620  2011               6500    Petrol    Automatic      Third    6.4
kmpl
5752  2015               8000    Petrol    Automatic      First   12.5
kmpl
1457  2019              26013    Diesel    Automatic      First  12.65
kmpl
1917  2018              28060    Petrol    Automatic      First  12.05
kmpl

       Engine      Power  Seats New_Price   Price  Price_in_lakhs
3952  2993 CC    255 bhp    5.0    2.3 Cr  160.00        0.001600
5620  5204 CC    560 bhp    2.0       NaN  120.00        0.001200
5752  5000 CC  488.1 bhp    2.0       NaN  100.00        0.001000
1457  2993 CC    255 bhp    5.0   1.39 Cr   97.07        0.000971
1917  2979 CC    320 bhp    5.0       NaN   93.67        0.000937

Grouped data (Mean Price and Max Year by Fuel_Type):
               Price  Year
Fuel_Type
```

```
Diesel     12.960686  2019
Electric   12.875000  2016
Petrol      5.756688  2019
```

In [ ]: