```python
import pandas as pd
```

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

```python
data.head()
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New_Price
0	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
1	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	13 km/kg	1199 CC	88.7 bhp	5.0	8.61 Lakh
2	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77 kmpl	1248 CC	88.76 bhp	7.0	NaN
3	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
4	6	Nissan Micra Diesel XV	Jaipur	2013	86999	Diesel	Manual	First	23.08 kmpl	1461 CC	63.1 bhp	5.0	NaN

Next steps: [Generate code with `data`]   [🔘 View recommended plots]

```python
#Checking for NULL values
```

```python
data.isna().sum()
```

```
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
```

```python
d1= pd.DataFrame(data)
```

B)Removing the units from Mileage, Engine, Power and New_Price and converting them into float

```python
d1['Mileage'] = d1['Mileage'].str.extract('(\d+\.\d+)').astype(float)
```

```python
d1['Engine'] = d1['Engine'].str.replace(' CC', '').astype(float)
```

```python
d1['Power'] = d1['Power'].str.extract('(\d+\.\d+)').astype(float)
```

```python
d1['New_Price'] = d1['New_Price'].str.extract('(\d+\.\d+)').astype(float)
```

```python
d1.head()
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Own
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	

Next steps: **Generate code with** d1   **View recommended plots**

A)Filling missing values in Mileage, Engine, Power with mean values,Seats with mode values

```
d1['Mileage'].fillna(d1['Mileage'].mean(),inplace=True)
d1['Engine'].fillna(d1['Engine'].mean(),inplace=True)
d1['Power'].fillna(d1['Power'].mean(),inplace=True)

d1['Seats'].fillna(d1['Seats'].mode()[0],inplace=True)
d1=d1.dropna()

d1.head()
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owne
1	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	
8	10	Maruti Ciaz Zeta	Kochi	2018	25692	Petrol	Manual	
13	15	Mitsubishi Pajero Sport 4X4	Delhi	2014	110000	Diesel	Manual	
18	20	BMW 3 Series 320d	Kochi	2014	32982	Diesel	Automatic	
26	28	Honda WRV i-VTEC VX	Kochi	2018	37430	Petrol	Manual	

Next steps: **Generate code with** d1   **View recommended plots**

```
output_csv_file = 'preprocessed_data.csv'
d1.to_csv(output_csv_file, index=False)


d1.head()
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owne
1	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	
8	10	Maruti Ciaz Zeta	Kochi	2018	25692	Petrol	Manual	
13	15	Mitsubishi Pajero Sport 4X4	Delhi	2014	110000	Diesel	Manual	
18	20	BMW 3 Series 320d	Kochi	2014	32982	Diesel	Automatic	
26	28	Honda WRV i-VTEC VX	Kochi	2018	37430	Petrol	Manual	

```
d2=pd.read_csv("/content/preprocessed_data.csv")
```

```
d2.isna().sum()
```

```
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
New_Price            0
Price                0
Current_Age          0
dtype: int64
```

Take a look at the unique values for some of our categorical variables: "Fuel_Type," "Transmission," and "Owner_Type."

```
d2['Fuel_Type'].unique()
```

```
array([1, 0, 2])
```

```
d2['Transmission'].unique()
```

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

```
d2['Owner_Type'].unique()
```

```
array(['First', 'Second', 'Third'], dtype=object)
```

C) To work with our categorical variables, we are converting them into numerical data using one-hot encoding.

```
d2['Fuel_Type'].replace({'Diesel': 0, 'Petrol': 1, 'Electric': 2}, inplace=True)
d2['Transmission'].replace({'Manual': 0, 'Automatic': 1}, inplace=True)
d2['Owner_Type'].replace({'First': 1, 'Second': 2, 'Third': 3, 'Fourth & Above': 4}, inplace=True)
```

```
d2.head()
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owne
1	2	Honda Jazz V	Chennai	2011	46000	1	0	
8	10	Maruti Ciaz Zeta	Kochi	2018	25692	1	0	
13	15	Mitsubishi Pajero Sport 4X4	Delhi	2014	110000	0	0	
18	20	BMW 3 Series 320d	Kochi	2014	32982	0	1	
26	28	Honda WRV i-VTEC VX	Kochi	2018	37430	1	0	

```
#Save the encoded data in separate csv file

output_csv_file = 'encoded_data.csv'
d2.to_csv(output_csv_file, index=False)
```

D)We are adding a new column called "CURRENT AGE OF THE CAR" to our dataset, which will help us understand how old each car is.

```
from datetime import datetime

current_year = datetime.now().year

d2['Current_Age'] = current_year - d1['Year']


#Save the encoded data in separate csv file
output_csv_file = 'newColumnData.csv'
d2.to_csv(output_csv_file, index=False)


d3=pd.read_csv("/content/newColumnData.csv")
d3.head()
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner
0	2	Honda Jazz V	Chennai	2011	46000	1	0	
1	10	Maruti Ciaz Zeta	Kochi	2018	25692	1	0	
2	15	Mitsubishi Pajero Sport 4X4	Delhi	2014	110000	0	0	
3	20	BMW 3 Series 320d	Kochi	2014	32982	0	1	
4	28	Honda WRV i-VTEC VX	Kochi	2018	37430	1	0	

Next steps: **Generate code with** d3   🔘 **View recommended plots**

e)Performing Select,Filter,Rename,Mutate,,Arrange and Summarize

SELECT

```
selected_columns = d3[['Seats', 'Current_Age', 'Year', 'Mileage', 'Price']]
selected_columns
```

	Seats	Current_Age	Year	Mileage	Price
0	5.0	13	2011	18.159379	4.50
1	5.0	6	2018	21.560000	9.95
2	7.0	10	2014	13.500000	15.00
3	5.0	10	2014	22.690000	18.55
4	5.0	6	2018	17.500000	9.90
...	...	...	...	...	...
796	5.0	8	2016	17.570000	4.00
797	5.0	13	2011	16.090000	3.25
798	5.0	8	2016	24.300000	7.43
799	5.0	11	2013	18.500000	3.20
800	5.0	10	2014	28.400000	4.75

801 rows × 5 columns

Next steps: Generate code with `selected_columns`   View recommended plots

FILTER

```
d3['Mileage'] = pd.to_numeric(d3['Mileage'], errors='coerce')
filtered_data = d3[d3['Mileage'] > 20]
filtered_data
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	
1	10	Maruti Ciaz Zeta	Kochi	2018	25692	1	0	
3	20	BMW 3 Series 320d	Kochi	2014	32982	0	1	
6	36	Maruti Vitara Brezza ZDi Plus	Kochi	2018	50075	0	0	
11	69	Hyundai Creta 1.4 E Plus Diesel	Mumbai	2018	13000	0	0	
17	113	BMW 3 Series 320d Luxury Line	Mumbai	2015	56087	0	1	
...	...	...	...	...	...	...	...	
791	5971	Volkswagen Polo 1.5 TDI Highline	Hyderabad	2013	65000	0	0	
793	5981	Maruti Baleno Alpha	Mumbai	2017	6000	1	0	
794	5987	Tata Tiago 1.2 Revotron XT	Kochi	2017	15386	1	0	
798	6005	Maruti Vitara Brezza VDi	Pune	2016	37208	0	0	
800	6014	Maruti Swift VDI	Delhi	2014	27365	0	0	

369 rows × 15 columns

Next steps: **Generate code with** `filtered_data`   ⬤ **View recommended plots**

RENAME

```
from os import rename

d3.rename(columns={'Year': 'Origin_Year'}, inplace=True)
d3
```

Name	Location	Origin_Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mi
a Jazz V	Chennai	2011	46000	1	0	1	18.1
i Ciaz Zeta	Kochi	2018	25692	1	0	1	21.5
ubishi Pajero rt 4X4	Delhi	2014	110000	0	0	1	13.5
MW 3 320d	Kochi	2014	32982	0	1	1	22.6
Honda VRV i- EC VX	Kochi	2018	37430	1	0	1	17.5
...	...	...	...	...	...	...	
a Bolt votron XT	Chennai	2016	10000	1	0	1	17.5
vagen to 1.6 ghline	Mumbai	2011	38000	1	0	1	16.0
Maruti Vitara a VDi	Pune	2016	37208	0	0	1	24.3
a Brio 'X MT	Delhi	2013	33746	1	0	1	18.5
i Swift VDI	Delhi	2014	27365	0	0	1	28.4

Next steps: Generate code with d3   View recommended plots

MUTATE

```
current_year = 2024
d3['Age'] = current_year - d3['Origin_Year']
d3
```

	Unnamed: 0	Name	Location	Origin_Year	Kilometers_Driven	Fuel_Type	Transmis
0	2	Honda Jazz V	Chennai	2011	46000	1	
1	10	Maruti Ciaz Zeta	Kochi	2018	25692	1	
2	15	Mitsubishi Pajero Sport 4X4	Delhi	2014	110000	0	
3	20	BMW 3 Series 320d	Kochi	2014	32982	0	
4	25	WRV F	Kochi	2018	37456	1	

Next steps: Generate code with d3   ● View recommended plots

ARRANGE

```
sorted_data = d3.sort_values(by='Price', ascending=False)
sorted_data
```

	Unnamed: 0	Name	Location	Origin_Year	Kilometers_Driven	Fuel_Type	Transmis
531	4079	Land Rover Range Rover 3.0 Diesel LWB Vogue	Hyderabad	2017	25000	0	
197	1505	Land Rover Range Rover Sport SE	Kochi	2019	26013	0	
277	2095	Mercedes-Benz SLC 43 AMG	Coimbatore	2019	2526	1	
243	1885	Mercedes-Benz GLS 350d Grand Edition	Delhi	2018	6000	0	
		Mercedes-					