CLEANING THE DATA SET

BEFORE CLEANING:

	Car_Name	Year	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission	Owner	
0	ritz	2014	3.35	5.59	27000	Petrol	Dealer	Manual	0	
1	sx4	2013	4.75	9.54	43000	Diesel	Dealer	Manual	0	
2	ciaz	2017	7.25	9.85	6900	Petrol	Dealer	Manual	0	
3	wagon r	2011	2.85	4.15	5200	Petrol	Dealer	Manual	0	
4	swift	2014	4.60	6.87	42450	Diesel	Dealer	Manual	0	

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

df=pd.read_csv("car data.csv")
df.head()

df.shape

print(df["Seller_Type"].unique())
print(df["Fuel_Type"].unique())
print(df["Transmission"].unique())
print(df["Owner"].unique())
```

```
df.isnull().sum()

df.describe()

fd=df[["Year","Selling_Price","Present_Price","Kms_Driven","Fuel_Type","Seller_Type","Transmission","Owner"]]

fd.head()

fd["current_year"]=2020

fd.head()

fd["Years_Old"]=fd["current_year"]-fd["Year"]

fd.head()

fd=pd.get_dummies(fd,drop_first=True)

fd.head()

fd.drop(['Year'],axis=1,inplace=True)

fd.drop(['current_year'],axis=1,inplace=True)

fd.head()

fd=pd.get_dummies(fd,drop_first=True)

fd.head()
```

AFTER CLEANING:

	Selling_Price	Present_Price	Kms_Driven	Owner	Years_Old	Fuel_Type_Diesel	Fuel_Type_Petrol	Seller_Type_Individual	Transmission_Manual	
0	3.35	5.59	27000	0	6	0	1	0	1	
1	4.75	9.54	43000	0	7	1	0	0	1	

2	7.25	9.85	6900	0	3	0	1	0	1
3	2.85	4.15	5200	0	9	0	1	0	1
4	4.60	6.87	42450	0	6	1	0	0	