question

Problem Statement: Data Analysis The dataset contains more of 10, 000 rows and more than 10 columns which contains features of the car and its (MSRP) manufacturer's suggested retail price. Clean the data and analyse it making it ready for modelling.

solution

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

In [4]: dt=pd.read_csv('Documents/github/JeetPython_tops/assignment/data analisis with pyth
In [5]: dt
```

	Make	Model	Year	Engine Fuel Type	Engine HP	Engine Cylinders	Transmission Type	Driven_Wheels	Nu
	0 BMW	1 Series M	2011	premium unleaded (required)	335.0	6.0	MANUAL	rear wheel drive	
	1 BMW	1 Series	2011	premium unleaded (required)	300.0	6.0	MANUAL	rear wheel drive	
	2 BMW	1 Series	2011	premium unleaded (required)	300.0	6.0	MANUAL	rear wheel drive	
	3 BMW	1 Series	2011	premium unleaded (required)	230.0	6.0	MANUAL	rear wheel drive	
	4 BMW	1 Series	2011	premium unleaded (required)	230.0	6.0	MANUAL	rear wheel drive	
119	09 Acura	ZDX	2012	premium unleaded (required)	300.0	6.0	AUTOMATIC	all wheel drive	
119	10 Acura	ZDX	2012	premium unleaded (required)	300.0	6.0	AUTOMATIC	all wheel drive	
119	11 Acura	ZDX	2012	premium unleaded (required)	300.0	6.0	AUTOMATIC	all wheel drive	
119	12 Acura	ZDX	2013	premium unleaded (recommended)	300.0	6.0	AUTOMATIC	all wheel drive	
119	13 Lincoln	Zephyr	2006	regular unleaded	221.0	6.0	AUTOMATIC	front wheel drive	

11914 rows × 16 columns

In [6]: dt.head(5)

Out[6]:		Make	Model	Year	Engine Fuel Type	Engine HP	_	Transmission Type	Driven_Wheels	Number of Doors	Ma
	0	BMW	1 Series M	2011	premium unleaded (required)	335.0	6.0	MANUAL	rear wheel drive	2.0	Tun
	1	BMW	1 Series	2011	premium unleaded (required)	300.0	6.0	MANUAL	rear wheel drive	2.0	Luxu
	2	BMW	1 Series	2011	premium unleaded (required)	300.0	6.0	MANUAL	rear wheel drive	2.0	
	3	BMW	1 Series	2011	premium unleaded (required)	230.0	6.0	MANUAL	rear wheel drive	2.0	Luxu
	4	BMW	1 Series	2011	premium unleaded (required)	230.0	6.0	MANUAL	rear wheel drive	2.0	
											>

In [7]: dt.tail(5)

Out[7]:

	Make	Model	Year	Engine Fuel Type	Engine HP	Engine Cylinders	Transmission Type	Driven_Wheels	Nu I
11909	Acura	ZDX	2012	premium unleaded (required)	300.0	6.0	AUTOMATIC	all wheel drive	
11910	Acura	ZDX	2012	premium unleaded (required)	300.0	6.0	AUTOMATIC	all wheel drive	
11911	Acura	ZDX	2012	premium unleaded (required)	300.0	6.0	AUTOMATIC	all wheel drive	
11912	Acura	ZDX	2013	premium unleaded (recommended)	300.0	6.0	AUTOMATIC	all wheel drive	
11913	Lincoln	Zephyr	2006	regular unleaded	221.0	6.0	AUTOMATIC	front wheel drive	
									•

In [8]: dt.dtypes

```
object
            Make
  Out[8]:
            Model
                                     object
                                       int64
            Year
            Engine Fuel Type
                                     object
            Engine HP
                                    float64
            Engine Cylinders
                                    float64
            Transmission Type
                                     object
            Driven_Wheels
                                     object
            Number of Doors
                                    float64
            Market Category
                                     object
            Vehicle Size
                                     object
            Vehicle Style
                                     object
            highway MPG
                                       int64
            city mpg
                                       int64
            Popularity
                                       int64
            MSRP
                                       int64
            dtype: object
            dt = dt.drop(['Engine Fuel Type', 'Market Category', 'Vehicle Style', 'Popularity'
  In [9]:
            dt.head(5)
  Out[9]:
                                    Engine
                                              Engine Transmission
                                                                                   highway
                                                                                             city
                              Year
                                                                                                   MSRP
                Make Model
                                                                    Driven_Wheels
                                       HP
                                                                                       MPG
                                            Cylinders
                                                                                             mpg
                                                              Type
                           1
                                                                         rear wheel
                BMW
                              2011
                                      335.0
                                                  6.0
                                                          MANUAL
                                                                                                   46135
            0
                       Series
                                                                                         26
                                                                                               19
                                                                             drive
                          M
                                                                         rear wheel
                BMW
                              2011
                                      300.0
                                                  6.0
                                                          MANUAL
                                                                                         28
                                                                                               19
                                                                                                   40650
                       Series
                                                                             drive
                                                                         rear wheel
                BMW
                              2011
                                      300.0
                                                  6.0
                                                          MANUAL
                                                                                         28
                                                                                                   36350
                       Series
                                                                             drive
                                                                         rear wheel
                BMW
                              2011
                                      230.0
                                                  6.0
                                                          MANUAL
                                                                                         28
                                                                                               18
                                                                                                   29450
                       Series
                                                                             drive
                                                                         rear wheel
                BMW
                              2011
                                      230.0
                                                  6.0
                                                          MANUAL
                                                                                         28
                                                                                               18
                                                                                                   34500
                       Series
                                                                             drive
4
            dt = dt.rename(columns={"Engine HP": "HP", "Engine Cylinders": "Cylinders", "Transi
 In [10]:
            dt.head(5)
                                                                                  MPG-
                                                                                          MPG-
 Out[10]:
                Make
                                        HP Cylinders Transmission
                                                                     Drive Mode
                                                                                                  Price
                        Model
                                Year
                                                                                      Н
                                                                                              C
                                                                       rear wheel
                        1 Series
            0
                BMW
                                2011
                                      335.0
                                                  6.0
                                                           MANUAL
                                                                                                46135
                                                                                     26
                                                                                             19
                            M
                                                                           drive
                                                                       rear wheel
                                2011
                                      300.0
                                                           MANUAL
                                                                                                40650
                BMW
                        1 Series
                                                  6.0
                                                                                     28
                                                                                              19
                                                                           drive
                                                                       rear wheel
            2
                BMW
                        1 Series
                                2011
                                      300.0
                                                  6.0
                                                           MANUAL
                                                                                      28
                                                                                                 36350
                                                                           drive
                                                                       rear wheel
            3
                BMW
                                2011
                                      230.0
                                                  6.0
                                                           MANUAL
                                                                                      28
                                                                                                29450
                        1 Series
                                                                                             18
                                                                           drive
                                                                       rear wheel
                BMW
                        1 Series 2011 230.0
                                                  6.0
                                                           MANUAL
                                                                                     28
                                                                                             18
                                                                                                34500
                                                                           drive
            duplicate_rows_dt = dt[dt.duplicated()]
 In [13]:
```

print("number of duplicate rows: ", duplicate_rows_dt.shape)

number of duplicate rows: (989, 10) In [15]: dt.count() 11914 Make Out[15]: Model 11914 Year 11914 HP 11845 Cylinders 11884 Transmission 11914 Drive Mode 11914 MPG-H 11914 MPG-C 11914 Price 11914 dtype: int64 In [16]: dt = dt.drop_duplicates() dt.head(5) MPG-MPG-Out[16]: Make Model Year **HP Cylinders Transmission Drive Mode Price** Н C rear wheel 1 Series 0 BMW 2011 335.0 6.0 MANUAL 26 19 46135 drive rear wheel 1 Series 2011 300.0 BMW 6.0 MANUAL 28 19 40650 drive rear wheel MANUAL 28 **BMW** 1 Series 2011 300.0 6.0 20 36350 drive rear wheel MANUAL 3 BMW 1 Series 2011 230.0 6.0 28 18 29450 drive rear wheel 4 BMW 1 Series 2011 230.0 6.0 MANUAL 28 18 34500 drive In [17]: dt.count() 10925 Make Out[17]: Model 10925 Year 10925 HP 10856 Cylinders 10895 Transmission 10925 Drive Mode 10925 MPG-H 10925 MPG-C 10925 Price 10925 dtype: int64 In [18]: print(dt.isnull().sum()) Make 0 Model 0 0 Year HP 69 Cylinders 30 0 Transmission Drive Mode 0 MPG-H 0

MPG-C

Price

dtype: int64

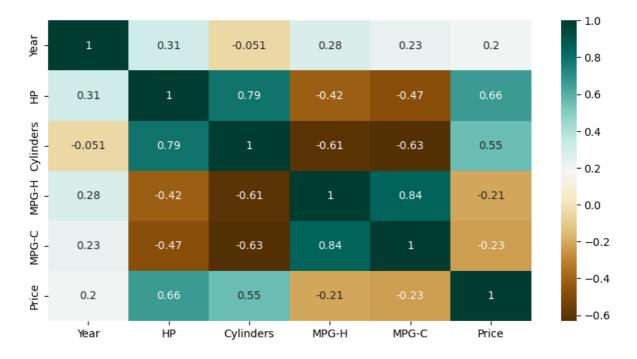
0 0

```
dt = dt.dropna()
In [20]:
               dt.count()
               Make
                                         10827
Out[20]:
               Model
                                         10827
               Year
                                         10827
               HP
                                         10827
               Cylinders
                                         10827
               Transmission
                                         10827
               Drive Mode
                                        10827
               MPG-H
                                         10827
               MPG-C
                                         10827
               Price
                                         10827
               dtype: int64
In [22]: print(dt.isnull().sum())
               Make
                                         0
               Model
                                         0
               Year
                                         0
               ΗP
                                         0
               Cylinders
                                         0
               Transmission
                                         0
               Drive Mode
                                         0
               MPG-H
                                         0
               MPG-C
                                         0
               Price
                                         0
               dtype: int64
In [40]: plt.subplots(figsize=(15,10))
               plt.scatter(dt['Price'],dt['Make'])
               plt.xlabel('Price')
               plt.ylabel('Make')
               plt.show()
                   Bugatti
HUMMER
Genesis
Scion
Infiniti
Lotus
Land Rover
Aston Martin
                   Rolls-Royce
Acura
Buick
Spyker
Volkswagen
Subaru
Lincoln
Lamborghini
Dodge
Chevrolet
Bentley
Kaila
                    Kia
Cadillac
Ford
Suzuki
Oldsmobile
Honda
Plymouth
                     Maybach
McLaren
                   McLaren
Toyota
Alfa Romeo
Ferrari
Mitsubishi
Mazda
Volvo
Nissan
Chrysler
rcedes-Benz
FIAT
Audi
BMW
                                                           0.5
                                                                                     1.0
Price
                                                                                                               1.5
                                 0.0
                                                                                                                                        2.0
               plt.figure(figsize=(10,5))
In [38]:
               c= dt.corr()
               sns.heatmap(c,cmap="BrBG",annot=True)
```

C:\Users\Sony Vaio\AppData\Local\Temp\ipykernel_7976\2442648811.py:2: FutureWarnin
g: The default value of numeric_only in DataFrame.corr is deprecated. In a future
version, it will default to False. Select only valid columns or specify the value
of numeric_only to silence this warning.
 c= dt.corr()

Out[38]:

	Year	НР	Cylinders	MPG-H	MPG-C	Price
Year	1.000000	0.314971	-0.050598	0.284237	0.234135	0.196789
НР	0.314971	1.000000	0.788007	-0.420281	-0.473551	0.659835
Cylinders	-0.050598	0.788007	1.000000	-0.611576	-0.632407	0.554740
MPG-H	0.284237	-0.420281	-0.611576	1.000000	0.841229	-0.209150
MPG-C	0.234135	-0.473551	-0.632407	0.841229	1.000000	-0.234050
Price	0.196789	0.659835	0.554740	-0.209150	-0.234050	1.000000



In []: