```
[84] # Importing libraries
     import pandas as pd
     import matplotlib.pyplot as plt
     import numpy as np
     import seaborn as sns
```

```
#Importing dataset
carData=pd.read_csv('/content/drive/My Drive/data.csv')
carData.head()
```

```
[86] #Data types
     carData.dtypes
```

```
[87] #Statistical summary
     carData.describe()
```

```
[88] #Shape
     carData.shape
```

```
[89] #NULL values
     d=carData.isnull().sum()
     d
```

77.68 GB available

From the above output we can clearly see that there are maximum number of null value ools Help All changes saved

Comment

+ Code + Text

RAM Disk

From the above output we can clearly see that there are maximum number of null values in the 'Market Category' column. Thus, for the ease of things we opt to remove it from the labelled columns. Also, we see that various more parameters do not have much significance in determining the prices of the cars. Thus dropping those columns won't make much of a difference in the processing of the model.

```
[90] #Eliminating the insignificant columns
     carData = carData.drop(['Engine Fuel Type','Number of Doors','Market Category'], axis = 1)
     carData.head()
```

```
[91] carData.shape
```

```
[92] carData.rename(columns = { "Engine HP": "HP", "Engine Cylinders": "Cylinders",
                                "Transmission Type": "Transmission", "Driven_Wheels": "Drive Mode","highway MPG": "MPG-H",
                                "city mpg": "MPG-C", "MSRP": "Price"}, inplace = True)
```

```
[93] carData.drop_duplicates()
```

```
[94] #NULL values
     print(carData.isnull().sum())
```

```
[95] carData=carData.dropna()
     carData.count()
```

```
carData.isnull().sum()
```

Activate Windows

ENG

+ Code + Text

```
[97] #Plotting graphs of Data (Columns)
     sns.boxplot(x=carData['HP'])

[98] sns.boxplot(x=carData['Cylinders'])

[99] sns.boxplot(x=carData['MPG-H'])

[100] sns.boxplot(x=carData['MPG-C'])

[101] sns.boxplot(x=carData['Popularity'])

[102] sns.boxplot(x=carData['Price'])

[103] q1=carData.quantile(0.25)
      q3=carData.quantile(0.75)
      iqr=q3-q1
      iqr

[104] carData=carData[~((carData<(q1-1.5*iqr))|(carData>(q3+1.5*iqr))).any(axis=1)]

[105] carData.shape
```

77.68 GB available

arch.google.com/drive/1dOYzkvnzt806ypkZerLZPp8t_s_2xNmG#scrollTo=Y7G6hKx9C6sh

Water Level Indicat... Ultrasonic Sensor B... FREE Marathi Typin... Download subtitles... Install Office 2019 L...

Runtime Tools Help All changes saved

+ Code + Text

[108]
```
#Percentage of car per brand
counts=carData['Make'].value_counts()*100/sum(carData['Make'].value_counts())
#Top 10 popular brands
popularCars=counts.index[:10]
#Plotting the bar plot
plt.figure(figsize=(10,5))
plt.bar(popularCars,height=counts[:10])
plt.title('Top 10 car Brands')
plt.show()
```

[109]
```
prices=carData[['Make','Price']].loc[(carData['Make']=='Chevrolet')|
                                     (carData['Make']=='Volkswagen')|
                                     (carData['Make']=='Toyota')|
                                     (carData['Make']=='Nissan')|
                                     (carData['Make']=='GMC')|
                                     (carData['Make']=='Dodge')|
                                     (carData['Make']=='Mazda')|
                                     (carData['Make']=='Honda')|
                                     (carData['Make']=='Suzuki')|
                                     (carData['Make']=='Infiniti')].groupby('Make').mean()
prices
```

```
#Correlation Matrix
corrMatrix=carData.corr()
sns.heatmap(corrMatrix,annot=True)
```

GB available

From the heatmap plotted above, it can be concluded that:

ıb ☆

sert Runtime Tools Help All changes saved

+ Code + Text

From the heatmap plotted above, it can be concluded that:

>>> Price is positively dependent on features and Horse Power(HP) and Year

>>> The features HP and Cylinders are positively dependent on each other

i.e. if number of Cylinders are increased, the HP also increases.

>>>MPG-H and MPG-C have strong negative correlation with Cylinders.

i.e. if number of cylinders are increased, MPG-H and MPG-C decreases.

```
[115] sns.barplot(carData['Year'],carData['Price'])

sns.barplot(carData['HP'],carData['Price'])

sns.barplot(carData['Cylinders'],carData['Price'])

sns.barplot(carData['MPG-H'],carData['Price'])
```

77.68 GB available

Tools Help All changes saved

+ Code + Text

```
sns.barplot(carData['Cylinders'],carData['Price'])
```

```
sns.barplot(carData['MPG-H'],carData['Price'])
```

```
[118] sns.barplot(carData['MPG-C'],carData['Price'])
```

```
sns.barplot(carData['Popularity'],carData['Price'])
```