Cars Dataset

Here, The Data of different cars is given with their specifications.

This data is available as a CSV file. We are going to analyze this data set using the Pandas DataFrame.

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

In [2]:
```
car= pd.read_csv("file.csv")
```

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

Out[3]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepo
0	Acura	MDX	SUV	Asia	All	$36,945	$33,337	3.5	6.0	2
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820	$21,761	2.0	4.0	2
2	Acura	TSX 4dr	Sedan	Asia	Front	$26,990	$24,647	2.4	4.0	2
3	Acura	TL 4dr	Sedan	Asia	Front	$33,195	$30,299	3.2	6.0	2
4	Acura	3.5 RL 4dr	Sedan	Asia	Front	$43,755	$39,014	3.5	6.0	2

In [5]:
```
car.shape
```

Out[5]:
```
(432, 15)
```

1. Instruction (For Data Cleaning)

Find all Null Value in the Dataset. If there is any null value in any column, then fill it with the mean of that column.

In [9]:
```
car.isnull()
```

Out[9]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepo
0	False	False	False	False	False	False	False	False	False	F
1	False	False	False	False	False	False	False	False	False	F
2	False	False	False	False	False	False	False	False	False	F
3	False	False	False	False	False	False	False	False	False	F
4	False	False	False	False	False	False	False	False	False	F
...	...	...	...	...	...	...	...	...	...	
427	False	False	False	False	False	False	False	False	False	F
428	False	False	False	False	False	False	False	False	False	F
429	False	False	False	False	False	False	False	False	False	F
430	False	False	False	False	False	False	False	False	False	F
431	False	False	False	False	False	False	False	False	False	F

432 rows × 15 columns

In [20]:
```
car.isnull().sum()
```

Out[20]:
```
Make          4
Model         4
Type          4
Origin        4
DriveTrain    4
MSRP          4
Invoice       4
EngineSize    4
Cylinders     0
Horsepower    4
MPG_City      4
MPG_Highway   4
Weight        4
Wheelbase     4
Length        4
dtype: int64
```

In [15]:
```
print(car.columns)
```

```
Index(['Make', 'Model', 'Type', 'Origin', 'DriveTrain', 'MSRP', 'Invoice',
       'EngineSize', 'Cylinders', 'Horsepower', 'MPG_City', 'MPG_Highway',
       'Weight', 'Wheelbase', 'Length'],
      dtype='object')
```

In [16]:
```
print(car.head())  # Display the first few rows of the DataFrame
```

```
    Make           Model   Type Origin DriveTrain      MSRP   Invoice  \
0  Acura             MDX    SUV   Asia        All  $36,945   $33,337
1  Acura  RSX Type S 2dr  Sedan   Asia      Front  $23,820   $21,761
2  Acura         TSX 4dr  Sedan   Asia      Front  $26,990   $24,647
3  Acura          TL 4dr  Sedan   Asia      Front  $33,195   $30,299
4  Acura      3.5 RL 4dr  Sedan   Asia      Front  $43,755   $39,014

   EngineSize  Cylinders  Horsepower  MPG_City  MPG_Highway  Weight  \
0         3.5        6.0       265.0      17.0         23.0  4451.0
1         2.0        4.0       200.0      24.0         31.0  2778.0
2         2.4        4.0       200.0      22.0         29.0  3230.0
3         3.2        6.0       270.0      20.0         28.0  3575.0
4         3.5        6.0       225.0      18.0         24.0  3880.0

   Wheelbase  Length
0      106.0   189.0
1      101.0   172.0
2      105.0   183.0
3      108.0   186.0
4      115.0   197.0
```

In [18]:
```
car['Cylinders'].fillna(car['Cylinders'].mean(), inplace=True)
```

In [23]:
```
car.isnull().sum()
```

Out[23]:
```
Make          0
Model         0
Type          0
Origin        0
DriveTrain    0
MSRP          0
Invoice       0
EngineSize    0
Cylinders     0
Horsepower    0
MPG_City      0
MPG_Highway   0
Weight        0
Wheelbase     0
Length        0
dtype: int64
```

In [25]:
```
car.shape
```

Out[25]:
```
(428, 15)
```

2. Question (Based on value Counts)

Check what are the different types of make are there in our dataset, And, What is the count(occurrence) of each make in the data?

In [27]:
```
car.head(2)
```

Out[27]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepo
0	Acura	MDX	SUV	Asia	All	$36,945	$33,337	3.5	6.0	2
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820	$21,761	2.0	4.0	2

In [28]:
```
car['Make'].value_counts()
```

Out[28]:
```
Make
Toyota           28
Chevrolet        27
Mercedes-Benz    26
Ford             23
BMW              20
Audi             19
Honda            17
Nissan           17
Volkswagen       15
Chrysler         15
Dodge            13
Mitsubishi       13
Volvo            12
Jaguar           12
Hyundai          12
Subaru           11
Pontiac          11
Mazda            11
Lexus            11
Kia              11
Buick             9
Mercury           9
Lincoln           9
Saturn            8
Cadillac          8
Suzuki            8
Infiniti          8
GMC               8
Acura             7
Porsche           7
Saab              7
Land Rover        3
Oldsmobile        3
Jeep              3
Scion             2
Isuzu             2
MINI              2
Hummer            1
Name: count, dtype: int64
```

3. Instruction (Filtering)

Show all the records where Origin is Asia or europe

In [29]:
```
car['Origin'].value_counts()
```

Out[29]:
```
Origin
Asia      158
USA       147
Europe    123
Name: count, dtype: int64
```

In [31]:
```
car[car['Origin'].isin(['Asia','Europe'])]
```

Out[31]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	I
0	Acura	MDX	SUV	Asia	All	$36,945	$33,337	3.5	6.0	
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820	$21,761	2.0	4.0	
2	Acura	TSX 4dr	Sedan	Asia	Front	$26,990	$24,647	2.4	4.0	
3	Acura	TL 4dr	Sedan	Asia	Front	$33,195	$30,299	3.2	6.0	
4	Acura	3.5 RL 4dr	Sedan	Asia	Front	$43,755	$39,014	3.5	6.0	
...	...	...	...	...	...	...	...	...	...	
427	Volvo	C70 LPT convertible 2dr	Sedan	Europe	Front	$40,565	$38,203	2.4	5.0	
428	Volvo	C70 HPT convertible 2dr	Sedan	Europe	Front	$42,565	$40,083	2.3	5.0	
429	Volvo	S80 T6 4dr	Sedan	Europe	Front	$45,210	$42,573	2.9	6.0	
430	Volvo	V40	Wagon	Europe	Front	$26,135	$24,641	1.9	4.0	
431	Volvo	XC70	Wagon	Europe	All	$35,145	$33,112	2.5	5.0	

281 rows × 15 columns

4. Instruction (Removing unwanted records)

Remove all the records(rows) where weight is above 4000.

In [37]:
```
print(car.columns)
```

```
Index(['Make', 'Model', 'Type', 'Origin', 'DriveTrain', 'MSRP', 'Invoice',
       'EngineSize', 'Cylinders', 'Horsepower', 'MPG_City', 'MPG_Highway',
       'Weight', 'Wheelbase', 'Length'],
      dtype='object')
```

In [39]:
```
car[~(car['Weight'] > 4000)]
```

Out[39]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820	$21,761	2.0	4.0
2	Acura	TSX 4dr	Sedan	Asia	Front	$26,990	$24,647	2.4	4.0
3	Acura	TL 4dr	Sedan	Asia	Front	$33,195	$30,299	3.2	6.0
4	Acura	3.5 RL 4dr	Sedan	Asia	Front	$43,755	$39,014	3.5	6.0
5	Acura	3.5 RL w/Navigation 4dr	Sedan	Asia	Front	$46,100	$41,100	3.5	6.0
...	...	...	...	...	...	...	...	...	...
427	Volvo	C70 LPT convertible 2dr	Sedan	Europe	Front	$40,565	$38,203	2.4	5.0
428	Volvo	C70 HPT convertible 2dr	Sedan	Europe	Front	$42,565	$40,083	2.3	5.0
429	Volvo	S80 T6 4dr	Sedan	Europe	Front	$45,210	$42,573	2.9	6.0
430	Volvo	V40	Wagon	Europe	Front	$26,135	$24,641	1.9	4.0
431	Volvo	XC70	Wagon	Europe	All	$35,145	$33,112	2.5	5.0

325 rows × 15 columns

In [41]:
```
car.shape
```

Out[41]:
```
(428, 15)
```

In [43]:
```
428-103
```

Out[43]:
```
325
```

5. Instruction (Applying function on a column)

Increase all the values of 'MPG_City' column by 3.

In [44]:
```
car.head()
```

Out[44]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepo
0	Acura	MDX	SUV	Asia	All	$36,945	$33,337	3.5	6.0	2
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820	$21,761	2.0	4.0	2
2	Acura	TSX 4dr	Sedan	Asia	Front	$26,990	$24,647	2.4	4.0	2
3	Acura	TL 4dr	Sedan	Asia	Front	$33,195	$30,299	3.2	6.0	2
4	Acura	3.5 RL 4dr	Sedan	Asia	Front	$43,755	$39,014	3.5	6.0	2

In [45]:
```
car['MPG_City']= car['MPG_City'].apply(lambda x:x+3)
```

In [46]:
```
car
```

Out[46]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	
0	Acura	MDX	SUV	Asia	All	$36,945	$33,337	3.5	6.0	
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820	$21,761	2.0	4.0	
2	Acura	TSX 4dr	Sedan	Asia	Front	$26,990	$24,647	2.4	4.0	
3	Acura	TL 4dr	Sedan	Asia	Front	$33,195	$30,299	3.2	6.0	
4	Acura	3.5 RL 4dr	Sedan	Asia	Front	$43,755	$39,014	3.5	6.0	
...	...	...	...	...	...	...	...	...	...	...
427	Volvo	C70 LPT convertible 2dr	Sedan	Europe	Front	$40,565	$38,203	2.4	5.0	
428	Volvo	C70 HPT convertible 2dr	Sedan	Europe	Front	$42,565	$40,083	2.3	5.0	
429	Volvo	S80 T6 4dr	Sedan	Europe	Front	$45,210	$42,573	2.9	6.0	
430	Volvo	V40	Wagon	Europe	Front	$26,135	$24,641	1.9	4.0	
431	Volvo	XC70	Wagon	Europe	All	$35,145	$33,112	2.5	5.0	

428 rows × 15 columns

In [ ]:
```

```