Cars Dataset

In [71]:
```
import pandas as pd
df = pd.read_csv(r'H:\DATA ANALYST AND SCIENCE SOFTWARE\PYTHON\Python Projects\Project 5 - Car Selling Analysis\2. Cars Data1.csv')
df.head()
```

Out[71]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepower
0	Acura	MDX	SUV	Asia	All	$36,945	$33,337	3.5	6.0	265
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820	$21,761	2.0	4.0	200
2	Acura	TSX 4dr	Sedan	Asia	Front	$26,990	$24,647	2.4	4.0	200
3	Acura	TL 4dr	Sedan	Asia	Front	$33,195	$30,299	3.2	6.0	270
4	Acura	3.5 RL 4dr	Sedan	Asia	Front	$43,755	$39,014	3.5	6.0	225

In [72]:
```
df.shape
```

Out[72]:
```
(428, 15)
```

In [73]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 428 entries, 0 to 427
Data columns (total 15 columns):
Make           428 non-null object
Model          428 non-null object
```

```
Type            428 non-null object
Origin          428 non-null object
DriveTrain      428 non-null object
MSRP            428 non-null object
Invoice         428 non-null object
EngineSize      428 non-null float64
Cylinders       426 non-null float64
Horsepower      428 non-null int64
MPG_City        428 non-null int64
MPG_Highway     428 non-null int64
Weight          428 non-null int64
Wheelbase       428 non-null int64
Length          428 non-null int64
dtypes: float64(2), int64(6), object(7)
memory usage: 50.3+ KB
```

Instruction 1 For Data Cleaning

1 . Check all null values in the data set.

In [74]:
```
df.isnull()
```

Out[74]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepower
0	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...
423	False	False	False	False	False	False	False	False	False	False
424	False	False	False	False	False	False	False	False	False	False

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepower
425	False	False	False	False	False	False	False	False	False	False
426	False	False	False	False	False	False	False	False	False	False
427	False	False	False	False	False	False	False	False	False	False

428 rows × 15 columns

In [75]:
```
df.isnull().sum()
```

Out[75]:
```
Make           0
Model          0
Type           0
Origin         0
DriveTrain     0
MSRP           0
Invoice        0
EngineSize     0
Cylinders      2
Horsepower     0
MPG_City       0
MPG_Highway    0
Weight         0
Wheelbase      0
Length         0
dtype: int64
```

In [76]:
```
df['Cylinders'].fillna(df['Cylinders'].mean(), inplace = True)
```

In [77]:
```
df.head(2)
```

Out[77]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepower
0	Acura	MDX	SUV	Asia	All	$36,945	$33,337	3.5	6.0	265

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepower
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820	$21,761	2.0	4.0	200

In [78]:
```
df.isnull().sum()
```

Out[78]:
```
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
```

Types Of Make and Count of Make Value

In [79]:
```
df['Make'].value_counts()
```

Out[79]:
```
Toyota           28
Chevrolet        27
Mercedes-Benz    26
Ford             23
BMW              20
Audi             19
Honda            17
Nissan           17
```

```
Volkswagen       15
Chrysler         15
Mitsubishi       13
Dodge            13
Hyundai          12
Volvo            12
Jaguar           12
Mazda            11
Kia              11
Subaru           11
Pontiac          11
Lexus            11
Buick             9
Lincoln           9
Mercury           9
Suzuki            8
Infiniti          8
Cadillac          8
GMC               8
Saturn            8
Acura             7
Porsche           7
Saab              7
Oldsmobile        3
Jeep              3
Land Rover        3
Scion             2
Isuzu             2
MINI              2
Hummer            1
Name: Make, dtype: int64
```

filtering method

Show All the REcords where origin is Asia or Europe

In [80]:
```
df.head(2)
```

Out[80]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepower
0	Acura	MDX	SUV	Asia	All	$36,945	$33,337	3.5	6.0	265
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820	$21,761	2.0	4.0	200

In [81]:
```
df[df['Origin'].isin(['Asia','Europe'])]
```

Out[81]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Hors
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	
...	...	...	...	...	...	...	...	...	...	
423	Volvo	C70 LPT convertible 2dr	Sedan	Europe	Front	$40,565	$38,203	2.4	5.0	
424	Volvo	C70 HPT convertible 2dr	Sedan	Europe	Front	$42,565	$40,083	2.3	5.0	
425	Volvo	S80 T6 4dr	Sedan	Europe	Front	$45,210	$42,573	2.9	6.0	
426	Volvo	V40	Wagon	Europe	Front	$26,135	$24,641	1.9	4.0	
427	Volvo	XC70	Wagon	Europe	All	$35,145	$33,112	2.5	5.0	

281 rows × 15 columns

remove unwanted records.

remove where weight is above 4000

In [82]:
```
df.head(2)
```

Out[82]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepower
0	Acura	MDX	SUV	Asia	All	$36,945	$33,337	3.5	6.0	265
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820	$21,761	2.0	4.0	200

In [83]:
```
df[~(df['Weight'] > 4000)]
```

Out[83]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Ho
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	
...	...	...	...	...	...	...	...	...	...	...
423	Volvo	C70 LPT convertible 2dr	Sedan	Europe	Front	$40,565	$38,203	2.4	5.0	
424	Volvo	C70 HPT convertible 2dr	Sedan	Europe	Front	$42,565	$40,083	2.3	5.0	
425	Volvo	S80 T6 4dr	Sedan	Europe	Front	$45,210	$42,573	2.9	6.0	

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Ho
426	Volvo	V40	Wagon	Europe	Front	$26,135	$24,641	1.9	4.0	
427	Volvo	XC70	Wagon	Europe	All	$35,145	$33,112	2.5	5.0	

325 rows × 15 columns

Apply Function on a column

Increase all the value of 'MPG_City' by 3

In [87]:
```
df.head(2)
```

Out[87]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepower
0	Acura	MDX	SUV	Asia	All	$36,945	$33,337	3.5	6.0	265
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820	$21,761	2.0	4.0	200

In [88]:
```
df['MPG_City'] = df['MPG_City'].apply(lambda x:x+3)
df.head(2)
```

Out[88]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepower
0	Acura	MDX	SUV	Asia	All	$36,945	$33,337	3.5	6.0	265
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820	$21,761	2.0	4.0	200

In [ ]:
```
```

In [ ]: