

Working on real project with python

(A part of big data analysis)

Cars Dataset

Here,the data of different cars is given their specifications. This data is available as a CSV file. we are going to analyze the data set using the pandas dataframe.

```
In [2]: import pandas as pd
car=pd.read_csv(r"C:\Users\hp\Downloads\2. Cars Data1.csv")

In [3]: car

Out[3]:
```

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepower	MPG_City	MPG_Highway	Weight	Wheelbase	Length
0	Acura	MDX	SUV	Asia	All	\$36,945	\$33,337	3.5	6.0	265.0	17.0	23.0	4451.0	106.0	189.0
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	\$23,820	\$21,761	2.0	4.0	200.0	24.0	31.0	2778.0	101.0	172.0
2	Acura	TSX 4dr	Sedan	Asia	Front	\$26,990	\$24,647	2.4	4.0	200.0	22.0	29.0	3230.0	105.0	183.0
3	Acura	TL 4dr	Sedan	Asia	Front	\$33,195	\$30,299	3.2	6.0	270.0	20.0	28.0	3575.0	108.0	186.0
4	Acura	3.5 RL 4dr	Sedan	Asia	Front	\$43,755	\$39,014	3.5	6.0	225.0	18.0	24.0	3880.0	115.0	197.0
...
427	Volvo	C70 LPT convertible 2dr	Sedan	Europe	Front	\$40,565	\$38,203	2.4	5.0	197.0	21.0	28.0	3450.0	105.0	186.0
428	Volvo	C70 HPT convertible 2dr	Sedan	Europe	Front	\$42,565	\$40,083	2.3	5.0	242.0	20.0	26.0	3450.0	105.0	186.0
429	Volvo	S80 T6 4dr	Sedan	Europe	Front	\$45,210	\$42,573	2.9	6.0	268.0	19.0	26.0	3653.0	110.0	190.0
430	Volvo	V40	Wagon	Europe	Front	\$26,135	\$24,641	1.9	4.0	170.0	22.0	29.0	2822.0	101.0	180.0
431	Volvo	XC70	Wagon	Europe	All	\$35,145	\$33,112	2.5	5.0	208.0	20.0	27.0	3823.0	109.0	186.0

432 rows × 15 columns

```
In [4]: car.head()

Out[4]:
```

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepower	MPG_City	MPG_Highway	Weight	Wheelbase	Length
0	Acura	MDX	SUV	Asia	All	\$36,945	\$33,337	3.5	6.0	265.0	17.0	23.0	4451.0	106.0	189.0
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	\$23,820	\$21,761	2.0	4.0	200.0	24.0	31.0	2778.0	101.0	172.0
2	Acura	TSX 4dr	Sedan	Asia	Front	\$26,990	\$24,647	2.4	4.0	200.0	22.0	29.0	3230.0	105.0	183.0
3	Acura	TL 4dr	Sedan	Asia	Front	\$33,195	\$30,299	3.2	6.0	270.0	20.0	28.0	3575.0	108.0	186.0
4	Acura	3.5 RL 4dr	Sedan	Asia	Front	\$43,755	\$39,014	3.5	6.0	225.0	18.0	24.0	3880.0	115.0	197.0

```
In [5]: car.shape

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

1)instruction (for data cleaning)

- find all null value in the dataset.if there is any value in any column,then fill it with the mean of that column.

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

Out[9]:
Make          4
Model         4
Type          4
Origin        4
DriveTrain    4
MSRP          4
Invoice       4
EngineSize    4
Cylinders     6
Horsepower    4
MPG_City      4
MPG_Highway   4
Weight        4
Wheelbase     4
Length        4
dtype: int64

In [13]: car['Cylinders'].fillna(car['Cylinders'].mean(),inplace= True)

In [15]: car.isnull().sum()

Out[15]:
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
```

2) instruction (filtering)

- 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 [16]: car.head(2)

Out[16]:
```

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepower	MPG_City	MPG_Highway	Weight	Wheelbase	Length
0	Acura	MDX	SUV	Asia	All	\$36,945	\$33,337	3.5	6.0	265.0	17.0	23.0	4451.0	106.0	189.0
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	\$23,820	\$21,761	2.0	4.0	200.0	24.0	31.0	2778.0	101.0	172.0

```
In [17]: car['Make'].value_counts()

Out[17]:
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: Make, dtype: int64
```

3) instruction (filtering)

- show all the records where origin is asia or europe.

```
In [18]: car.head(2)

Out[18]:
```

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepower	MPG_City	MPG_Highway	Weight	Wheelbase	Length
0	Acura	MDX	SUV	Asia	All	\$36,945	\$33,337	3.5	6.0	265.0	17.0	23.0	4451.0	106.0	189.0
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	\$23,820	\$21,761	2.0	4.0	200.0	24.0	31.0	2778.0	101.0	172.0

```
In [20]: car[car['Origin'].isin(['Asia','Europe'])]

Out[20]:
```

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepower	MPG_City	MPG_Highway	Weight	Wheelbase	Length
0	Acura	MDX	SUV	Asia	All	\$36,945	\$33,337	3.5	6.0	265.0	17.0	23.0	4451.0	106.0	189.0
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	\$23,820	\$21,761	2.0	4.0	200.0	24.0	31.0	2778.0	101.0	172.0
2	Acura	TSX 4dr	Sedan	Asia	Front	\$26,990	\$24,647	2.4	4.0	200.0	22.0	29.0	3230.0	105.0	183.0
3	Acura	TL 4dr	Sedan	Asia	Front	\$33,195	\$30,299	3.2	6.0	270.0	20.0	28.0	3575.0	108.0	186.0
4	Acura	3.5 RL 4dr	Sedan	Asia	Front	\$43,755	\$39,014	3.5	6.0	225.0	18.0	24.0	3880.0	115.0	197.0
...
427	Volvo	C70 LPT convertible 2dr	Sedan	Europe	Front	\$40,565	\$38,203	2.4	5.0	197.0	21.0	28.0	3450.0	105.0	186.0
428	Volvo	C70 HPT convertible 2dr	Sedan	Europe	Front	\$42,565	\$40,083	2.3	5.0	242.0	20.0	26.0	3450.0	105.0	186.0
429	Volvo	S80 T6 4dr	Sedan	Europe	Front	\$45,210	\$42,573	2.9	6.0	268.0	19.0	26.0	3653.0	110.0	190.0
430	Volvo	V40	Wagon	Europe	Front	\$26,135	\$24,641	1.9	4.0	170.0	22.0	29.0	2822.0	101.0	180.0
431	Volvo	XC70	Wagon	Europe	All	\$35,145	\$33,112	2.5	5.0	208.0	20.0	27.0	3823.0	109.0	186.0

281 rows × 15 columns

4)instruction (removing unwanted records)

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

```
In [22]: car.head(2)

Out[22]:
```

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepower	MPG_City	MPG_Highway	Weight	Wheelbase	Length
0	Acura	MDX	SUV	Asia	All	\$36,945	\$33,337	3.5	6.0	265.0	17.0	23.0	4451.0	106.0	189.0
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	\$23,820	\$21,761	2.0	4.0	200.0	24.0	31.0	2778.0	101.0	172.0

```
In [25]: car[car['Weight'] > 4000]

Out[25]:
```

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepower	MPG_City	MPG_Highway	Weight	Wheelbase	Length
0	Acura	MDX	SUV	Asia	All	\$36,945	\$33,337	3.5	6.0	265.0	17.0	23.0	4451.0	106.0	189.0
15	Audi	A4 3.0 Quattro convertible 2dr	Sedan	Europe	All	\$44,240	\$40,075	3.0	6.0	220.0	18.0	25.0	4013.0	105.0	180.0
17	Audi	A6 4.2 Quattro 4dr	Sedan	Europe	All	\$49,690	\$44,936	4.2	8.0	300.0	17.0	24.0	4024.0	109.0	193.0
18	Audi	A8 L Quattro 4dr	Sedan	Europe	All	\$69,190	\$64,740	4.2	8.0	330.0	17.0	24.0	4399.0	121.0	204.0
20	Audi	RS 6 4dr	Sports	Europe	Front	\$84,600	\$76,417	4.2	8.0	450.0	15.0	22.0	4024.0	109.0	191.0
...
405	Volkswagen	Touareg V6	SUV	Europe	All	\$35,515	\$32,243	3.2	6.0	220.0	15.0	20.0	5086.0	112.0	187.0
415	Volkswagen	Phaeton 4dr	Sedan	Europe	Front	\$65,000	\$59,912	4.2	8.0	335.0	16.0	22.0	5194.0	118.0	204.0
416	Volkswagen	Phaeton W12 4dr	Sedan	Europe	Front	\$75,000	\$69,130	6.0	12.0	420.0	12.0	19.0	5399.0	118.0	204.0
419	Volkswagen	Passat W8	Wagon	Europe	Front	\$40,235	\$36,956	4.0	8.0	270.0	18.0	25.0	4067.0	106.0	184.0
420	Volvo	XC90 T6	SUV	Europe	All	\$41,250	\$38,851	2.9	6.0	268.0	15.0	20.0	4638.0	113.0	189.0

103 rows × 15 columns

5)instruction (applying function on a column)

- increase all the value of 'MPG_Highway' column by 3.

```
In [26]: car.head(2)

Out[26]:
```

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepower	MPG_City	MPG_Highway	Weight	Wheelbase	Length
0	Acura	MDX	SUV	Asia	All	\$36,945	\$33,337	3.5	6.0	265.0	17.0	23.0	4451.0	106.0	189.0
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	\$23,820	\$21,761	2.0	4.0	200.0	24.0	31.0	2778.0	101.0	172.0

```
In [32]: car['MPG_Highway']= car['MPG_Highway'].apply(lambda x:x+3)

In [33]: car

Out[33]:
```

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepower	MPG_City	MPG_Highway	Weight	Wheelbase	Length
0	Acura	MDX	SUV	Asia	All	\$36,945	\$33,337	3.5	6.0	265.0	25.0	26.0	4451.0	106.0	189.0
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	\$23,820	\$21,761	2.0	4.0	200.0	32.0	34.0	2778.0	101.0	172.0
2	Acura	TSX 4dr	Sedan	Asia	Front	\$26,990	\$24,647	2.4	4.0	200.0	30.0	32.0	3230.0	105.0	183.0
3	Acura	TL 4dr	Sedan	Asia	Front	\$33,195	\$30,299	3.2	6.0	270.0	28.0	31.0	3575.0	108.0	186.0
4	Acura	3.5 RL 4dr	Sedan	Asia	Front	\$43,755	\$39,014	3.5	6.0	225.0	26.0	27.0	3880.0	115.0	197.0
...
427	Volvo	C70 LPT convertible 2dr	Sedan	Europe	Front	\$40,565	\$38,203	2.4	5.0	197.0	29.0	31.0	3450.0	105.0	186.0
428	Volvo	C70 HPT convertible 2dr	Sedan	Europe	Front	\$42,565	\$40,083	2.3	5.0	242.0	28.0	29.0	3450.0	105.0	186.0
429	Volvo	S80 T6 4dr	Sedan	Europe	Front	\$45,210	\$42,573	2.9	6.0	268.0	27.0	29.0	3653.0	110.0	190.0
430	Volvo	V40	Wagon	Europe	Front	\$26,135	\$24,641	1.9	4.0	170.0	30.0	32.0	2822.0	101.0	180.0
431	Volvo	XC70	Wagon	Europe	All	\$35,145	\$33,112	2.5	5.0	208.0	28.0	30.0	3823.0	109.0	186.0

432 rows × 15 columns

```
In [ ]:
```