Working on Real Project with Python

(A part of Big Data Analysis)

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.

```python
import pandas as pd

car = pd.read_csv(r"C:\Users\91999\Desktop\IT SPARKS\8. Project\7. Cars Data.csv")

car.head()
```

```
    Make          Model     Type Origin DriveTrain  Unnamed: 5   MSRP  \
0  Acura           MDX      SUV   Asia          All        NaN  $36,945

1  Acura   RSX Type S 2dr  Sedan  Asia        Front        NaN  $23,820

2  Acura         TSX 4dr   Sedan  Asia        Front        NaN  $26,990

3  Acura          TL 4dr   Sedan  Asia        Front        NaN  $33,195

4  Acura      3.5 RL 4dr   Sedan  Asia        Front        NaN  $43,755


   Invoice  EngineSize  Cylinders  Horsepower  MPG_City  MPG_Highway  \
0  $33,337         3.5        6.0       265.0      17.0         23.0

1  $21,761         2.0        4.0       200.0      24.0         31.0

2  $24,647         2.4        4.0       200.0      22.0         29.0

3  $30,299         3.2        6.0       270.0      20.0         28.0

4  $39,014         3.5        6.0       225.0      18.0         24.0
```

```
    Unnamed: 13  Weight  Wheelbase  Length
0            NaN  4451.0      106.0   189.0
1            NaN  2778.0      101.0   172.0
2            NaN  3230.0      105.0   183.0
3            NaN  3575.0      108.0   186.0
4            NaN  3880.0      115.0   197.0
```

```
car.shape
```

```
(432, 17)
```

1) Instruction (For Data Cleaning)

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

```
car.isnull().sum()
```

```
Make            4
Model           4
Type            4
Origin          4
DriveTrain      4
Unnamed: 5    432
MSRP            4
Invoice         4
EngineSize      4
Cylinders       6
Horsepower      4
MPG_City        4
MPG_Highway     4
Unnamed: 13   432
Weight          4
Wheelbase       4
Length          4
dtype: int64
```

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

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 ?

```
car.head(2)
```

```
     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


   EngineSize  Cylinders  Horsepower  MPG_City  MPG_Highway  Weight  \
0         3.5        6.0         265        17           23    4451
1         2.0        4.0         200        24           31    2778

   Wheelbase  Length
0        106     189
1        101     172
```

```python
car['Make'].value_counts()
```

```
Toyota           28
Chevrolet        27
Mercedes-Benz    26
Ford             23
BMW              20
Audi             19
Honda            17
Nissan           17
Chrysler         15
Volkswagen       15
Mitsubishi       13
Dodge            13
Jaguar           12
Volvo            12
Hyundai          12
Kia              11
Pontiac          11
Subaru           11
Mazda            11
Lexus            11
Mercury           9
Buick             9
Lincoln           9
Saturn            8
Suzuki            8
Cadillac          8
GMC               8
Infiniti          8
Porsche           7
Saab              7
Acura             7
Jeep              3
Oldsmobile        3
```

```
Land Rover        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.

```
car.head(2)
```

	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

	EngineSize	Cylinders	Horsepower	MPG_City	MPG_Highway	Weight
0	3.5	6.0	265	17	23	4451
1	2.0	4.0	200	24	31	2778

	Wheelbase	Length
0	106	189
1	101	172

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

	Make	Model	Type	Origin	DriveTrain	MSRP
0	Acura	MDX	SUV	Asia	All	$36,945
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820
2	Acura	TSX 4dr	Sedan	Asia	Front	$26,990
3	Acura	TL 4dr	Sedan	Asia	Front	$33,195
4	Acura	3.5 RL 4dr	Sedan	Asia	Front	$43,755
..	...	...	...	...	...	..
423	Volvo	C70 LPT convertible 2dr	Sedan	Europe	Front	$40,565
424	Volvo	C70 HPT convertible 2dr	Sedan	Europe	Front	$42,565
425	Volvo	S80 T6 4dr	Sedan	Europe	Front	$45,210

426	Volvo		V40	Wagon	Europe	Front $26,135
427	Volvo		XC70	Wagon	Europe	All $35,145

	Invoice	EngineSize	Cylinders	Horsepower	MPG_City	MPG_Highway
0	$33,337	3.5	6.0	265	17	23
1	$21,761	2.0	4.0	200	24	31
2	$24,647	2.4	4.0	200	22	29
3	$30,299	3.2	6.0	270	20	28
4	$39,014	3.5	6.0	225	18	24
..	...	...	...	...	...	..
423	$38,203	2.4	5.0	197	21	28
424	$40,083	2.3	5.0	242	20	26
425	$42,573	2.9	6.0	268	19	26
426	$24,641	1.9	4.0	170	22	29
427	$33,112	2.5	5.0	208	20	27

	Weight	Wheelbase	Length
0	4451	106	189
1	2778	101	172
2	3230	105	183
3	3575	108	186
4	3880	115	197
..	...	...	...
423	3450	105	186
424	3450	105	186
425	3653	110	190
426	2822	101	180
427	3823	109	186

[281 rows x 15 columns]

```python
car['Origin'].value_counts()
```

```
Asia      158
USA       147
```

```
Europe      123
Name: Origin, dtype: int64
```

```
158 + 123
```

```
281
```

---

4) Instruction (Removing unwanted records)

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

```
car.head(2)
```

```
    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


    EngineSize   Cylinders   Horsepower   MPG_City   MPG_Highway   Weight  \
0          3.5         6.0          265         17            23     4451
1          2.0         4.0          200         24            31     2778

    Wheelbase   Length
0         106      189
1         101      172
```

```
car[~(car['Weight'] > 4000)]
```

```
        Make                     Model    Type   Origin DriveTrain
Unnamed: 5  \
1     Acura            RSX Type S 2dr   Sedan     Asia      Front
NaN
2     Acura                    TSX 4dr   Sedan     Asia      Front
NaN
3     Acura                     TL 4dr   Sedan     Asia      Front
NaN
4     Acura                 3.5 RL 4dr   Sedan     Asia      Front
NaN
5     Acura   3.5 RL w/Navigation 4dr   Sedan     Asia      Front
NaN
..      ...                        ...     ...      ...        ...
...
427   Volvo   C70 LPT convertible 2dr   Sedan   Europe      Front
NaN
428   Volvo   C70 HPT convertible 2dr   Sedan   Europe      Front
NaN
429   Volvo                 S80 T6 4dr   Sedan   Europe      Front
```

```
NaN
430   Volvo                        V40   Wagon   Europe        Front
NaN
431   Volvo                        XC70  Wagon   Europe          All
NaN

         MSRP     Invoice   EngineSize   Cylinders   Horsepower
MPG_City  \
1     $23,820    $21,761         2.0         4.0        200.0        24.0

2     $26,990    $24,647         2.4         4.0        200.0        22.0

3     $33,195    $30,299         3.2         6.0        270.0        20.0

4     $43,755    $39,014         3.5         6.0        225.0        18.0

5     $46,100    $41,100         3.5         6.0        225.0        18.0

..        ...        ...         ...         ...          ...         ...

427   $40,565    $38,203         2.4         5.0        197.0        21.0

428   $42,565    $40,083         2.3         5.0        242.0        20.0

429   $45,210    $42,573         2.9         6.0        268.0        19.0

430   $26,135    $24,641         1.9         4.0        170.0        22.0

431   $35,145    $33,112         2.5         5.0        208.0        20.0


      MPG_Highway   Unnamed: 13   Weight   Wheelbase   Length
1            31.0           NaN   2778.0       101.0    172.0
2            29.0           NaN   3230.0       105.0    183.0
3            28.0           NaN   3575.0       108.0    186.0
4            24.0           NaN   3880.0       115.0    197.0
5            24.0           NaN   3893.0       115.0    197.0
..            ...           ...      ...         ...      ...
427          28.0           NaN   3450.0       105.0    186.0
428          26.0           NaN   3450.0       105.0    186.0
429          26.0           NaN   3653.0       110.0    190.0
430          29.0           NaN   2822.0       101.0    180.0
431          27.0           NaN   3823.0       109.0    186.0

[329 rows x 17 columns]

428 - 103

325
```

5) Instruction (Applying function on a column)

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

```
car.head(2)
```

	Make	Model	Type	Origin	DriveTrain	Unnamed: 5	MSRP
0	Acura	MDX	SUV	Asia	All	NaN	$36,945
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	NaN	$23,820

	Invoice	EngineSize	Cylinders	Horsepower	MPG_City	MPG_Highway
0	$33,337	3.5	6.0	265.0	17.0	23.0
1	$21,761	2.0	4.0	200.0	24.0	31.0

	Unnamed: 13	Weight	Wheelbase	Length
0	NaN	4451.0	106.0	189.0
1	NaN	2778.0	101.0	172.0

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

```
car
```

	Make	Model	Type	Origin	DriveTrain	Unnamed: 5
0	Acura	MDX	SUV	Asia	All	NaN
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	NaN
2	Acura	TSX 4dr	Sedan	Asia	Front	NaN
3	Acura	TL 4dr	Sedan	Asia	Front	NaN
4	Acura	3.5 RL 4dr	Sedan	Asia	Front	NaN
..	...	...	...	...	...	...
427	Volvo	C70 LPT convertible 2dr	Sedan	Europe	Front	NaN
428	Volvo	C70 HPT convertible 2dr	Sedan	Europe	Front	NaN
429	Volvo	S80 T6 4dr	Sedan	Europe	Front	NaN
430	Volvo	V40	Wagon	Europe	Front	NaN
431	Volvo	XC70	Wagon	Europe	All	NaN

```
          MSRP    Invoice  EngineSize  Cylinders  Horsepower
MPG_City  \
0     $36,945    $33,337         3.5        6.0       265.0        20.0

1     $23,820    $21,761         2.0        4.0       200.0        27.0

2     $26,990    $24,647         2.4        4.0       200.0        25.0

3     $33,195    $30,299         3.2        6.0       270.0        23.0

4     $43,755    $39,014         3.5        6.0       225.0        21.0

..        ...        ...         ...        ...         ...         ...

427   $40,565    $38,203         2.4        5.0       197.0        24.0

428   $42,565    $40,083         2.3        5.0       242.0        23.0

429   $45,210    $42,573         2.9        6.0       268.0        22.0

430   $26,135    $24,641         1.9        4.0       170.0        25.0

431   $35,145    $33,112         2.5        5.0       208.0        23.0


      MPG_Highway  Unnamed: 13  Weight  Wheelbase  Length
0            23.0          NaN  4451.0      106.0   189.0
1            31.0          NaN  2778.0      101.0   172.0
2            29.0          NaN  3230.0      105.0   183.0
3            28.0          NaN  3575.0      108.0   186.0
4            24.0          NaN  3880.0      115.0   197.0
..            ...          ...     ...        ...     ...
427          28.0          NaN  3450.0      105.0   186.0
428          26.0          NaN  3450.0      105.0   186.0
429          26.0          NaN  3653.0      110.0   190.0
430          29.0          NaN  2822.0      101.0   180.0
431          27.0          NaN  3823.0      109.0   186.0

[432 rows x 17 columns]
```