DAY-23

OCTOBER-09

PANDAS:

- Shape and dtypes are doing some action but doesnot considered as attributes.
- To see all the details about the table in only one function we can use the info() function.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 32 entries, 0 to 31
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   manufacturer  32 non-null     object 
 1   mpg           32 non-null     float64
 2   cyl           32 non-null     int64  
 3   disp          32 non-null     float64
 4   hp            32 non-null     int64  
 5   drat          32 non-null     float64
 6   wt            32 non-null     float64
 7   qsec          32 non-null     float64
 8   vs            32 non-null     int64  
 9   am            32 non-null     int64  
 10  gear          32 non-null     int64  
 11  carb          32 non-null     int64  
dtypes: float64(5), int64(6), object(1)
memory usage: 3.1+ KB
```

- To extract only one column we need to use variable followed by the column name.

```
0               Mazda RX4
1           Mazda RX4 Wag
2              Datsun 710
3          Hornet 4 Drive
4       Hornet Sportabout
5                 Valiant
6              Duster 360
7               Merc 240D
8                Merc 230
9                Merc 280
10              Merc 280C
11             Merc 450SE
12             Merc 450SL
13            Merc 450SLC
14     Cadillac Fleetwood
15    Lincoln Continental
16      Chrysler Imperial
17               Fiat 128
18            Honda Civic
19         Toyota Corolla
20          Toyota Corona
```

- The data will be extracted in series format to extract it in data frame format we need to use position format,use two square brackets.

	manufacturer
0	Mazda RX4
1	Mazda RX4 Wag
2	Datsun 710
3	Hornet 4 Drive
4	Hornet Sportabout
5	Valiant
6	Duster 360
7	Merc 240D
8	Merc 230
9	Merc 280
10	Merc 280C
11	Merc 450SE
12	Merc 450SL
13	Merc 450SLC

- To extract multiple columns in data frame format separate the column names by comma.

	manufacturer	mpg	cyl	gear
0	Mazda RX4	21.0	6	4
1	Mazda RX4 Wag	21.0	6	4
2	Datsun 710	22.8	4	4
3	Hornet 4 Drive	21.4	6	3
4	Hornet Sportabout	18.7	8	3
5	Valiant	18.1	6	3
6	Duster 360	14.3	8	3
7	Merc 240D	24.4	4	4
8	Merc 230	22.8	4	4
9	Merc 280	19.2	6	4
10	Merc 280C	17.8	6	4
11	Merc 450SE	16.4	8	3
12	Merc 450SL	17.3	8	3
13	Merc 450SLC	15.2	8	3
14	Cadillac Fleetwood	10.4	8	3
15	Lincoln Continental	10.4	8	3

- To use the same data multiple times better assign it to a variable.
  sdata = d2[['manufacturer','mpg','cyl','gear']]

- To create series data and data frame data manually:
  Series data - pd.Series([1,2,3,4,5])

```
0    1
1    2
2    3
3    4
4    5
dtype: int64
```

  Data frame data –
  Create a dictionary and change it to the data frame data, the dict should compulsorily have all the columns of same length.
  d = {'name':['a','b','c','d'],'id': [1,2,3,4],'sal':['10k','20k','30k','15k']}
  data = pd.DataFrame(d)

- To extract the d2 data from manufacturer till gear there are multiple ways:
  1. d2[['manufacturer','mpg','cyl','disp','hp','drat','wt','qsec','vs','am','gear']]

	manufacturer	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear
0	Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4
1	Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4
2	Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4
3	Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3
4	Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3
5	Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3

  2. d2.drop('carb',axis=1) (axis=0 – row,axis = 1 – column)

	manufacturer	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear
0	Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4
1	Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4
2	Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4
3	Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3
4	Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3
5	Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3

  3. d2.iloc[:,:-1]

	manufacturer	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear
0	Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4
1	Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4
2	Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4
3	Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3
4	Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3
5	Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3

4. d2.loc[:,'manufacturer':'gear']

	manufacturer	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear
0	Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4
1	Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4
2	Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4
3	Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3
4	Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3
5	Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3

- To delete the column permanently: d2 = d2.drop('carb',axis=1)

	manufacturer	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear
0	Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4
1	Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4
2	Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4
3	Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3
4	Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3
5	Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3

- To delete multiple columns permanently: d2 = d2.drop(['hp','wt'],axis = 1)

	manufacturer	mpg	cyl	disp	drat	qsec	vs	am	gear
0	Mazda RX4	21.0	6	160.0	3.90	16.46	0	1	4
1	Mazda RX4 Wag	21.0	6	160.0	3.90	17.02	0	1	4
2	Datsun 710	22.8	4	108.0	3.85	18.61	1	1	4
3	Hornet 4 Drive	21.4	6	258.0	3.08	19.44	1	0	3
4	Hornet Sportabout	18.7	8	360.0	3.15	17.02	0	0	3
5	Valiant	18.1	6	225.0	2.76	20.22	1	0	3

- To delete a row : d2 = d2.drop([1,3]) (no need to mention the axis as the default axis is row)

	manufacturer	mpg	cyl	disp	drat	qsec	vs	am	gear
0	Mazda RX4	21.0	6	160.0	3.90	16.46	0	1	4
2	Datsun 710	22.8	4	108.0	3.85	18.61	1	1	4
4	Hornet Sportabout	18.7	8	360.0	3.15	17.02	0	0	3
5	Valiant	18.1	6	225.0	2.76	20.22	1	0	3
6	Duster 360	14.3	8	360.0	3.21	15.84	0	0	3
7	Merc 240D	24.4	4	146.7	3.69	20.00	1	0	4

iloc and loc dataframe slicing:

var.iloc[row position ,column position]
var.iloc[rowstartvalue:rowstopvalue:rowincrement/decrement,columnstartvalue:columnstopvalue:columnincrement/decrement]

- To print all rows of first 5 columns: d2.iloc[:,:5]

	manufacturer	mpg	cyl	disp	drat
0	Mazda RX4	21.0	6	160.0	3.90
2	Datsun 710	22.8	4	108.0	3.85
4	Hornet Sportabout	18.7	8	360.0	3.15
5	Valiant	18.1	6	225.0	2.76
6	Duster 360	14.3	8	360.0	3.21
7	Merc 240D	24.4	4	146.7	3.69
8	Merc 230	22.8	4	140.8	3.92
9	Merc 280	19.2	6	167.6	3.92

- To print all the columns and rows: d2.iloc[:,:]
- Odd indexed rows with even indexed columns: d2.iloc[1::2,::2]

	manufacturer	cyl	drat	vs	gear
2	Datsun 710	4	3.85	1	4
5	Valiant	6	2.76	1	3
7	Merc 240D	4	3.69	1	4
9	Merc 280	6	3.92	1	4
11	Merc 450SE	8	3.07	0	3
13	Merc 450SLC	8	3.07	0	3
15	Lincoln Continental	8	3.00	0	3
17	Fiat 128	4	4.08	1	4
19	Toyota Corolla	4	4.22	1	4
21	Dodge Challenger	8	2.76	0	3
23	Camaro Z28	8	3.73	0	3
25	Fiat X1-9	4	4.08	1	4