

```
In [5]: import pandas as pd
import numpy as np
```

```
In [6]: ds=pd.DataFrame([[1,2,3,4,5,np.NaN,7],
                        [11,2,np.NaN,33,np.NaN,88,10],
                        [np.NaN,10,15,22,27,29,11],
                        [32,21,25,26,np.NaN,14,19],
                        [17,33,51,24,26,np.NaN,37],
                        [18,14,np.NaN,51,56,48,49],
                        [np.NaN,51,89,56,23,84,56],
                        [np.NaN,7,np.NaN,89,np.NaN,np.NaN,85],
                        [np.NaN,86,np.NaN,48,np.NaN,np.NaN,59],
                        [np.NaN,89,99,32,np.NaN,np.NaN,27]],
                        index="a b c d e f g h i j".split(),
                        columns="A B C D E F G".split())
```

```
In [7]: ds
```

```
Out[7]:
```

	A	B	C	D	E	F	G
a	1.0	2	3.0	4	5.0	NaN	7
b	11.0	2	NaN	33	NaN	88.0	10
c	NaN	10	15.0	22	27.0	29.0	11
d	32.0	21	25.0	26	NaN	14.0	19
e	17.0	33	51.0	24	26.0	NaN	37
f	18.0	14	NaN	51	56.0	48.0	49
g	NaN	51	89.0	56	23.0	84.0	56
h	NaN	7	NaN	89	NaN	NaN	85
i	NaN	86	NaN	48	NaN	NaN	59
j	NaN	89	99.0	32	NaN	NaN	27

```
In [8]: ds.isnull()
```

```
Out[8]:
```

	A	B	C	D	E	F	G
a	False	False	False	False	False	True	False
b	False	False	True	False	True	False	False
c	True	False	False	False	False	False	False
d	False	False	False	False	True	False	False
e	False	False	False	False	False	True	False
f	False	False	True	False	False	False	False
g	True	False	False	False	False	False	False
h	True	False	True	False	True	True	False
i	True	False	True	False	True	True	False
j	True	False	False	False	True	True	False

```
In [9]: ds.isnull().any()
```

```
Out[9]: A      True
        B      False
        C      True
        D      False
        E      True
        F      True
        G      False
dtype: bool
```

```
In [10]: ds.dropna(axis=1)
```

```
Out[10]:
```

	B	D	G
a	2	4	7
b	2	33	10
c	10	22	11
d	21	26	19
e	33	24	37
f	14	51	49
g	51	56	56
h	7	89	85
i	86	48	59
j	89	32	27

```
In [31]: ds.dropna(thresh=6)
```

```
Out[31]:
```

	A	B	C	D	E	F	G
a	1.0	2	3.0	4	5.0	NaN	7
c	NaN	10	15.0	22	27.0	29.0	11
d	32.0	21	25.0	26	NaN	14.0	19
e	17.0	33	51.0	24	26.0	NaN	37
f	18.0	14	NaN	51	56.0	48.0	49
g	NaN	51	89.0	56	23.0	84.0	56

```
In [77]: ds["new"]=[1,2,3,4,5,6,7,8,9,10]
```

In [78]:

```
ds
```

Out[78]:

	A	B	C	D	E	F	G	new
a	1.0	2	3.0	4	5.0	NaN	7	1
b	11.0	2	NaN	33	NaN	88.0	10	2
c	NaN	10	15.0	22	27.0	29.0	11	3
d	32.0	21	25.0	26	NaN	14.0	19	4
e	17.0	33	51.0	24	26.0	NaN	37	5
f	18.0	14	NaN	51	56.0	48.0	49	6
g	NaN	51	89.0	56	23.0	84.0	56	7
h	NaN	7	NaN	89	NaN	NaN	85	8
i	NaN	86	NaN	48	NaN	NaN	59	9
j	NaN	89	99.0	32	NaN	NaN	27	10

In [79]:

```
ds.set_index("new",inplace=True)
```

In [80]: ds

Out[80]:

	A	B	C	D	E	F	G
new							
1	1.0	2	3.0	4	5.0	NaN	7
2	11.0	2	NaN	33	NaN	88.0	10
3	NaN	10	15.0	22	27.0	29.0	11
4	32.0	21	25.0	26	NaN	14.0	19
5	17.0	33	51.0	24	26.0	NaN	37
6	18.0	14	NaN	51	56.0	48.0	49
7	NaN	51	89.0	56	23.0	84.0	56
8	NaN	7	NaN	89	NaN	NaN	85
9	NaN	86	NaN	48	NaN	NaN	59
10	NaN	89	99.0	32	NaN	NaN	27

In [81]: ds.reset_index(inplace=True)

In [82]: ds

Out[82]:

	new	A	B	C	D	E	F	G
0	1	1.0	2	3.0	4	5.0	NaN	7
1	2	11.0	2	NaN	33	NaN	88.0	10
2	3	NaN	10	15.0	22	27.0	29.0	11
3	4	32.0	21	25.0	26	NaN	14.0	19
4	5	17.0	33	51.0	24	26.0	NaN	37
5	6	18.0	14	NaN	51	56.0	48.0	49
6	7	NaN	51	89.0	56	23.0	84.0	56
7	8	NaN	7	NaN	89	NaN	NaN	85
8	9	NaN	86	NaN	48	NaN	NaN	59
9	10	NaN	89	99.0	32	NaN	NaN	27

In [83]: ds.drop("new",axis=1,inplace=True)

In [116]: ds

Out[116]:

	A	B	C	D	E	F	G
0	1.0	2	3.0	4	5.0	NaN	7
1	11.0	2	NaN	33	NaN	88.0	10
2	1.0	10	15.0	22	27.0	29.0	11
3	32.0	21	25.0	26	NaN	14.0	19
4	17.0	33	51.0	24	26.0	NaN	37
5	18.0	14	NaN	51	56.0	48.0	49
6	1.0	51	89.0	56	23.0	84.0	56
7	1.0	7	NaN	89	NaN	NaN	85
8	1.0	86	NaN	48	NaN	NaN	59
9	1.0	89	99.0	32	NaN	NaN	27

```
In [60]: ds.head(2)
```

```
Out[60]:
```

	A	B	C	D	E	F	G
0	1.0	2	3.0	4	5.0	NaN	7
1	11.0	2	NaN	33	NaN	88.0	10

```
In [61]: ds.tail(2)
```

```
Out[61]:
```

	A	B	C	D	E	F	G
8	NaN	86	NaN	48	NaN	NaN	59
9	NaN	89	99.0	32	NaN	NaN	27

```
In [63]: ds['B'].unique()
```

```
Out[63]: array([ 2, 10, 21, 33, 14, 51,  7, 86, 89], dtype=int64)
```

```
In [66]: ds['B'].value_counts()
```

```
Out[66]: 2      2
14      1
10      1
89      1
7       1
86      1
21      1
51      1
33      1
Name: B, dtype: int64
```

```
In [67]: ds.fillna(value=100)
```

```
Out[67]:
```

	A	B	C	D	E	F	G
0	1.0	2	3.0	4	5.0	100.0	7
1	11.0	2	100.0	33	100.0	88.0	10
2	100.0	10	15.0	22	27.0	29.0	11
3	32.0	21	25.0	26	100.0	14.0	19
4	17.0	33	51.0	24	26.0	100.0	37
5	18.0	14	100.0	51	56.0	48.0	49
6	100.0	51	89.0	56	23.0	84.0	56
7	100.0	7	100.0	89	100.0	100.0	85
8	100.0	86	100.0	48	100.0	100.0	59
9	100.0	89	99.0	32	100.0	100.0	27

```
In [120]: ds.fillna(ds.mean())
```

```
Out[120]:
```

	A	B	C	D	E	F	G
a	1.0	2	3.0	4	5.0	52.6	7
b	11.0	2	47.0	33	27.4	88.0	10
c	15.8	10	15.0	22	27.0	29.0	11
d	32.0	21	25.0	26	27.4	14.0	19
e	17.0	33	51.0	24	26.0	52.6	37
f	18.0	14	47.0	51	56.0	48.0	49
g	15.8	51	89.0	56	23.0	84.0	56
h	15.8	7	47.0	89	27.4	52.6	85
i	15.8	86	47.0	48	27.4	52.6	59
j	15.8	89	99.0	32	27.4	52.6	27


```
In [121]: ds.fillna(ds.median())
```

```
Out[121]:
```

	A	B	C	D	E	F	G
a	1.0	2	3.0	4	5.0	48.0	7
b	11.0	2	38.0	33	26.0	88.0	10
c	17.0	10	15.0	22	27.0	29.0	11
d	32.0	21	25.0	26	26.0	14.0	19
e	17.0	33	51.0	24	26.0	48.0	37
f	18.0	14	38.0	51	56.0	48.0	49
g	17.0	51	89.0	56	23.0	84.0	56
h	17.0	7	38.0	89	26.0	48.0	85
i	17.0	86	38.0	48	26.0	48.0	59
j	17.0	89	99.0	32	26.0	48.0	27

```
In [122]: ds["A"].mode()
```

```
Out[122]: 0    1.0
1    11.0
2    17.0
3    18.0
4    32.0
dtype: float64
```

```
In [128]: ds["A"].fillna(ds["A"].mode()[0])
```

```
Out[128]: a    1.0
b    11.0
c    1.0
d    32.0
e    17.0
f    18.0
g    1.0
h    1.0
i    1.0
j    1.0
Name: A, dtype: float64
```

In [129]: ds

Out[129]:

	A	B	C	D	E	F	G
a	1.0	2	3.0	4	5.0	NaN	7
b	11.0	2	NaN	33	NaN	88.0	10
c	1.0	10	15.0	22	27.0	29.0	11
d	32.0	21	25.0	26	NaN	14.0	19
e	17.0	33	51.0	24	26.0	NaN	37
f	18.0	14	NaN	51	56.0	48.0	49
g	1.0	51	89.0	56	23.0	84.0	56
h	1.0	7	NaN	89	NaN	NaN	85
i	1.0	86	NaN	48	NaN	NaN	59
j	1.0	89	99.0	32	NaN	NaN	27

In [132]: ds["B"].mode()

Out[132]: 0 2
dtype: int64

In [133]: ds["B"].fillna(ds["B"].mode()[0])

Out[133]: a 2
b 2
c 10
d 21
e 33
f 14
g 51
h 7
i 86
j 89
Name: B, dtype: int64

In [134]:

```
ds
```

Out[134]:

	A	B	C	D	E	F	G
a	1.0	2	3.0	4	5.0	NaN	7
b	11.0	2	NaN	33	NaN	88.0	10
c	1.0	10	15.0	22	27.0	29.0	11
d	32.0	21	25.0	26	NaN	14.0	19
e	17.0	33	51.0	24	26.0	NaN	37
f	18.0	14	NaN	51	56.0	48.0	49
g	1.0	51	89.0	56	23.0	84.0	56
h	1.0	7	NaN	89	NaN	NaN	85
i	1.0	86	NaN	48	NaN	NaN	59
j	1.0	89	99.0	32	NaN	NaN	27

In [137]:

```
dataset=pd.read_csv(r"D:\ML_Course\Works_on_python\tips.csv")
```

In [139]: dataset

Out[139]:

	Unnamed: 0	sepal_length	sepal_width	petal_length	petal_width	species
0	0	5.1	3.5	1.4	0.2	setosa
1	1	4.9	3.0	1.4	0.2	setosa
2	2	4.7	3.2	1.3	0.2	setosa
3	3	4.6	3.1	1.5	0.2	setosa
4	4	5.0	3.6	1.4	0.2	setosa
5	5	5.4	3.9	1.7	0.4	setosa
6	6	4.6	3.4	1.4	0.3	setosa
7	7	5.0	3.4	1.5	0.2	setosa
8	8	4.4	2.9	1.4	0.2	setosa
9	9	4.9	3.1	1.5	0.1	setosa
10	10	5.4	3.7	1.5	0.2	setosa
11	11	4.8	3.4	1.6	0.2	setosa
12	12	4.8	3.0	1.4	0.1	setosa
13	13	4.3	3.0	1.1	0.1	setosa
14	14	5.8	4.0	1.2	0.2	setosa
15	15	5.7	4.4	1.5	0.4	setosa
16	16	5.4	3.9	1.3	0.4	setosa
17	17	5.1	3.5	1.4	0.3	setosa
18	18	5.7	3.8	1.7	0.3	setosa
19	19	5.1	3.8	1.5	0.3	setosa
20	20	5.4	3.4	1.7	0.2	setosa
21	21	5.1	3.7	1.5	0.4	setosa
22	22	4.6	3.6	1.0	0.2	setosa
23	23	5.1	3.3	1.7	0.5	setosa
24	24	4.8	3.4	1.9	0.2	setosa

Unnamed: 0	sepal_length	sepal_width	petal_length	petal_width	species	
25	25	5.0	3.0	1.6	0.2	setosa
26	26	5.0	3.4	1.6	0.4	setosa
27	27	5.2	3.5	1.5	0.2	setosa
28	28	5.2	3.4	1.4	0.2	setosa
29	29	4.7	3.2	1.6	0.2	setosa
...
120	120	6.9	3.2	5.7	2.3	virginica
121	121	5.6	2.8	4.9	2.0	virginica
122	122	7.7	2.8	6.7	2.0	virginica
123	123	6.3	2.7	4.9	1.8	virginica
124	124	6.7	3.3	5.7	2.1	virginica
125	125	7.2	3.2	6.0	1.8	virginica
126	126	6.2	2.8	4.8	1.8	virginica
127	127	6.1	3.0	4.9	1.8	virginica
128	128	6.4	2.8	5.6	2.1	virginica
129	129	7.2	3.0	5.8	1.6	virginica
130	130	7.4	2.8	6.1	1.9	virginica
131	131	7.9	3.8	6.4	2.0	virginica
132	132	6.4	2.8	5.6	2.2	virginica
133	133	6.3	2.8	5.1	1.5	virginica
134	134	6.1	2.6	5.6	1.4	virginica
135	135	7.7	3.0	6.1	2.3	virginica
136	136	6.3	3.4	5.6	2.4	virginica
137	137	6.4	3.1	5.5	1.8	virginica
138	138	6.0	3.0	4.8	1.8	virginica
139	139	6.9	3.1	5.4	2.1	virginica
140	140	6.7	3.1	5.6	2.4	virginica
141	141	6.9	3.1	5.1	2.3	virginica

Unnamed: 0	sepal_length	sepal_width	petal_length	petal_width	species
142	142	5.8	2.7	5.1	1.9 virginica
143	143	6.8	3.2	5.9	2.3 virginica
144	144	6.7	3.3	5.7	2.5 virginica
145	145	6.7	3.0	5.2	2.3 virginica
146	146	6.3	2.5	5.0	1.9 virginica
147	147	6.5	3.0	5.2	2.0 virginica
148	148	6.2	3.4	5.4	2.3 virginica
149	149	5.9	3.0	5.1	1.8 virginica

150 rows × 6 columns

In []: