In [1]:

```
# import lib
import numpy as np
import pandas as pd
```

In [ ]:

```
# Merging Fux

# Pandas Merge Connects Columns or Indexes in DataFrame based on keys (one or more then one
```

In [2]:

```
data1 = pd.DataFrame({
                        'Sno':[101,102,103,104,105,106,107],
                        'Name':['Harish','Suresh','Shoeb','Venkat','Wahed','Yogesh','Micky']
                       })
data1
```

Out[2]:

	Sno	Name
0	101	Harish
1	102	Suresh
2	103	Shoeb
3	104	Venkat
4	105	Wahed
5	106	Yogesh
6	107	Micky

In [3]:

```
data2 = pd.DataFrame({
                        'Sno':[101,102,103,104,105,106,107],
                        'City':['Hyd','Delhi','Puna','Ambala','Secand','Mumbai','Noida']
                       })
data2
```

Out[3]:

	Sno	City
0	101	Hyd
1	102	Delhi
2	103	Puna
3	104	Ambala
4	105	Secand
5	106	Mumbai
6	107	Noida

In [ ]:

```
# merge two dataset
'''
pd.merge(
    left,
    right,
    how='inner',
    on=None,
    left_on=None,
    right_on=None,
    left_index=False,
    right_index=False,
    sort=False,
    suffixes=('_x', '_y'),
    copy=True,
    indicator=False,
    validate=None,
)
'''
```

In [4]:

```
pd.merge(data1,data2, on='Sno')   # by default how='inner'
```

Out[4]:

	Sno	Name	City
0	101	Harish	Hyd
1	102	Suresh	Delhi
2	103	Shoeb	Puna
3	104	Venkat	Ambala
4	105	Wahed	Secand
5	106	Yogesh	Mumbai
6	107	Micky	Noida

In [5]:

```
pd.merge(data2,data1, on='Sno')
```

Out[5]:

	Sno	City	Name
0	101	Hyd	Harish
1	102	Delhi	Suresh
2	103	Puna	Shoeb
3	104	Ambala	Venkat
4	105	Secand	Wahed
5	106	Mumbai	Yogesh
6	107	Noida	Micky

In [ ]:

```

```

In [ ]:

```

```

In [6]:

```
data3 = pd.DataFrame({
                        'Sno':[101,102,103,104,105,108,109],
                        'Name':['Harish','Suresh','Shoeb','Venkat','Wahed','Yogesh','Micky']
                       })
data3
```

Out[6]:

	Sno	Name
0	101	Harish
1	102	Suresh
2	103	Shoeb
3	104	Venkat
4	105	Wahed
5	108	Yogesh
6	109	Micky

In [7]:

```
data4 = pd.DataFrame({
                        'Sno':[101,102,103,104,105,106,107],
                        'City':['Hyd','Delhi','Puna','Ambala','Secand','Mumbai','Noida']
                       })
data4
```

Out[7]:

	Sno	City
0	101	Hyd
1	102	Delhi
2	103	Puna
3	104	Ambala
4	105	Secand
5	106	Mumbai
6	107	Noida

In [ ]:

```

```

In [10]:

```
# Inner => both datset common value based on SNO
# how : {'left', 'right', 'outer', 'inner'}, default 'inner'
pd.merge(data3,data4, on='Sno',how='inner')
```

Out[10]:

	Sno	Name	City
0	101	Harish	Hyd
1	102	Suresh	Delhi
2	103	Shoeb	Puna
3	104	Venkat	Ambala
4	105	Wahed	Secand

In [ ]:

```

```

In [11]:

```
# how : {'left'}
pd.merge(data3,data4, on='Sno',how='left')
```

Out[11]:

	Sno	Name	City
0	101	Harish	Hyd
1	102	Suresh	Delhi
2	103	Shoeb	Puna
3	104	Venkat	Ambala
4	105	Wahed	Secand
5	108	Yogesh	NaN
6	109	Micky	NaN

In [ ]:

```

```

In [12]:

```
# how : {'right'}
pd.merge(data3,data4, on='Sno',how='right')
```

Out[12]:

	Sno	Name	City
0	101	Harish	Hyd
1	102	Suresh	Delhi
2	103	Shoeb	Puna
3	104	Venkat	Ambala
4	105	Wahed	Secand
5	106	NaN	Mumbai
6	107	NaN	Noida

In [ ]:

```

```

In [13]:

```
# how : {'outer'}
pd.merge(data3,data4, on='Sno',how='outer')
```

Out[13]:

	Sno	Name	City
0	101	Harish	Hyd
1	102	Suresh	Delhi
2	103	Shoeb	Puna
3	104	Venkat	Ambala
4	105	Wahed	Secand
5	108	Yogesh	NaN
6	109	Micky	NaN
7	106	NaN	Mumbai
8	107	NaN	Noida

In [ ]:

```

```

In [14]:

```
# indicator=False,(indicates values present in which dataset)
pd.merge(data3,data4, on='Sno',how='outer',indicator=True)
```

Out[14]:

	Sno	Name	City	_merge
0	101	Harish	Hyd	both
1	102	Suresh	Delhi	both
2	103	Shoeb	Puna	both
3	104	Venkat	Ambala	both
4	105	Wahed	Secand	both
5	108	Yogesh	NaN	left_only
6	109	Micky	NaN	left_only
7	106	NaN	Mumbai	right_only
8	107	NaN	Noida	right_only

In [ ]:

```

```

In [ ]:

```

```

In [21]:

```
data5 = pd.DataFrame({
                        'Sno':[101,102,103,104,105,106,107],
                        'Name':['Harish','Suresh','Shoeb','Venkat','Wahed','Yogesh','Micky']
                       })
data6 = pd.DataFrame({
                        'Sno':[111,112,113,114,115,116,117],
                        'Name':['Hyd','Delhi','Puna','Ambala','Secand','Mumbai','Noida']
                       })
data5
```

Out[21]:

	Sno	Name
0	101	Harish
1	102	Suresh
2	103	Shoeb
3	104	Venkat
4	105	Wahed
5	106	Yogesh
6	107	Micky

In [22]:

```
data6
```

Out[22]:

	Sno	Name
0	111	Hyd
1	112	Delhi
2	113	Puna
3	114	Ambala
4	115	Secand
5	116	Mumbai
6	117	Noida

In [20]:

```
pd.merge(data5,data6,on='Sno')
```

Out[20]:

	Sno	Name_x	Name_y

In [ ]:

```

```

In [26]:

```
# left_index=False,
#   right_index=False,
'''
left_index : bool, default False
    Use the index from the left DataFrame as the join key(s). If it is a
    MultiIndex, the number of keys in the other DataFrame (either the index
    or a number of columns) must match the number of levels.
right_index : bool, default False
    Use the index from the right DataFrame as the join key. Same caveats as
    left_index.
'''
pd.merge(data5,data6,left_index=True)
```

...

In [27]:

```
pd.merge(data5,data6,right_index=True)
```

...

In [28]:

```
data5
```

Out[28]:

	Sno	Name
0	101	Harish
1	102	Suresh
2	103	Shoeb
3	104	Venkat
4	105	Wahed
5	106	Yogesh
6	107	Micky

In [29]:

```
data6
```

Out[29]:

	Sno	Name
0	111	Hyd
1	112	Delhi
2	113	Puna
3	114	Ambala
4	115	Secand
5	116	Mumbai
6	117	Noida

In [25]:

```
pd.merge(data5,data6,left_index=True,right_index=True)
```

Out[25]:

	Sno_x	Name_x	Sno_y	Name_y
0	101	Harish	111	Hyd
1	102	Suresh	112	Delhi
2	103	Shoeb	113	Puna
3	104	Venkat	114	Ambala
4	105	Wahed	115	Secand
5	106	Yogesh	116	Mumbai
6	107	Micky	117	Noida

In [ ]:

```

```

In [ ]:

```

```

In [30]:

```
# IF BOTH DATA SET HAS SAME COLUMN NAME AND VALUES THEN ?
data7 = pd.DataFrame({
                        'Sno':[101,102,103,104,105,106,107],
                        'Name':['Harish','Suresh','Shoeb','Venkat','Wahed','Yogesh','Micky']
                       })
data8 = pd.DataFrame({
                        'Sno':[101,102,103,104,105,106,107],
                        'Name':['Harish','Suresh','Shoeb','Venkat','Wahed','Yogesh','Micky']
                       })
```

In [31]:

```
pd.merge(data7,data8,on='Sno')
```

Out[31]:

	Sno	Name_x	Name_y
0	101	Harish	Harish
1	102	Suresh	Suresh
2	103	Shoeb	Shoeb
3	104	Venkat	Venkat
4	105	Wahed	Wahed
5	106	Yogesh	Yogesh
6	107	Micky	Micky

In [32]:

```
# suffixes=('_x', '_y'),
pd.merge(data7,data8,on='Sno',suffixes=('_21-Oct', '_22-Oct'))
```

Out[32]:

	Sno	Name_21-Oct	Name_22-Oct
0	101	Harish	Harish
1	102	Suresh	Suresh
2	103	Shoeb	Shoeb
3	104	Venkat	Venkat
4	105	Wahed	Wahed
5	106	Yogesh	Yogesh
6	107	Micky	Micky

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

```

```

In [ ]:

```

```

In [ ]:

```

```