

Python Basic data Cleaning

In [1]:

```
import pandas as pd
import numpy as np
data = {'A': [12.0, 63.0, np.nan, 24.0, 12.0],
        'B': [32.0, np.nan, 23.0, 41.0, 32.0]}
df = pd.DataFrame(data)
df
```

Out[1]:

	A	B
0	12.0	32.0
1	63.0	NaN
2	NaN	23.0
3	24.0	41.0
4	12.0	32.0

Remove duplicate data

In [2]:

```
df1=df.drop_duplicates()
df1
```

Out[2]:

	A	B
0	12.0	32.0
1	63.0	NaN
2	NaN	23.0
3	24.0	41.0

In [3]:

```
df1.isnull()
```

Out[3]:

	A	B
0	False	False
1	False	True
2	True	False
3	False	False

For NaN value, the isnull() is returning ture if it found nan value

In [6]:

```
df1.dropna()
```

Out[6]:

	A	B
--	---	---

	A	B
0	12.0	32.0
3	24.0	41.0

Using `dropna()`, the null value is removed

In [8]:

```
df.isnull().sum()
```

Out[8]:

```
A      1
B      1
dtype: int64
```

We got number of null value is 1 for each column

In []: