Pandas

1. Create DataFrame

In [3]:

```
import pandas as pd

path = 'mhs.txt'
df = pd.read_csv(path, delimiter=',') # read file

print(df)
```

```
        nim    nama   umur       asal
0  10519001    Adit     19    Bandung
1  10519002    Dede     20      Garut
2  10519003    Bayu     20    Cirebon
3  10519004    Cici     19   Surabaya
```

In [4]:

```
path = 'mhs.csv'
df = pd.read_csv(path, delimiter=',') # read file

print(df)
```

```
        nim  nama  umur      asal
0  10519001  Adit    19   Bandung
1  10519002  Dede    20     Garut
2  10519003  Bayu    20   Cirebon
3  10519004  Cici    19  Surabaya
```

In [5]:

```
path = 'mhs.xlsx'
df = pd.read_excel(path, delimiter=',') # read file

print(df)
```

```
        nim  nama  umur      asal
0  10519001  Adit    19   Bandung
1  10519002  Dede    20     Garut
2  10519003  Bayu    20   Cirebon
3  10519004  Cici    19  Surabaya
```

In [32]:

```
newDf = pd.DataFrame([['10519005', 'Firman', 20, 'Bandung'],['10519006', 'Adit', 20, 'Ciamis']], co
lumns=('nim', 'nama', 'umur', 'asal'))
print(newDf)
```

```
        nim    nama  umur     asal
0  10519005  Firman    20  Bandung
1  10519006    Adit    20   Ciamis
```

In [33]:

```
concatDf = pd.concat([df, newDf])
print(concatDf)
concatDf = pd.concat([df, newDf], ignore_index=True)
print(concatDf)
```

```
        nim    nama  umur      asal
0  10519001    Adit    19   Bandung
```

```
1  10519002    Dede    20     Garut
2  10519003    Bayu    20   Cirebon
3  10519004    Cici    19  Surabaya
0  10519005  Firman    20   Bandung
1  10519006    Adit    20    Ciamis
        nim    nama  umur      asal
0  10519001    Adit    19   Bandung
1  10519002    Dede    20     Garut
2  10519003    Bayu    20   Cirebon
3  10519004    Cici    19  Surabaya
4  10519005  Firman    20   Bandung
5  10519006    Adit    20    Ciamis
```

In [53]:

```
copy = concatDf.copy()
jk = ['L','L','L','P','L','L']
copy['jk'] = jk
print(copy)
```

```
        nim    nama  umur      asal jk
0  10519001    Adit    19   Bandung  L
1  10519002    Dede    20     Garut  L
2  10519003    Bayu    20   Cirebon  L
3  10519004    Cici    19  Surabaya  P
4  10519005  Firman    20   Bandung  L
5  10519006    Adit    20    Ciamis  L
```

2. Access Items

In [55]:

```
print('Index DataFrame : ',df.index)
print('Rows : ',len(df))
```

```
Index DataFrame :  RangeIndex(start=0, stop=4, step=1)
Rows :  4
```

In [57]:

```
print('Data Baris Pertama : \n', df.iloc[0,:])
print('Data Kolom Nama : \n', df.iloc[:,1])
print('Data Kolom NIM : \n', df['nim'])
print('Data Kolom NIM & Nama : \n', df[['nim','nama']])
```

```
Data Baris Pertama :
 nim     10519001
nama        Adit
umur          19
asal     Bandung
Name: 0, dtype: object
Data Kolom Nama :
 0    Adit
1    Dede
2    Bayu
3    Cici
Name: nama, dtype: object
Data Kolom NIM :
 0    10519001
1    10519002
2    10519003
3    10519004
Name: nim, dtype: int64
Data Kolom NIM & Nama :
         nim  nama
0  10519001  Adit
1  10519002  Dede
2  10519003  Bayu
3  10519004  Cici
```

In [60]:

```
print('Data Mahasiswa Umur 20 : \n')
print(df.loc[df['umur'] == 20])
```

```
Data Mahasiswa Umur = 20 :

        nim  nama  umur     asal
1  10519002  Dede    20    Garut
2  10519003  Bayu    20  Cirebon
```

In [15]:

```
print('Hapus Data Asal : \n', df.drop('asal', axis=1))
print('Hapus Data Asal : \n', df.drop(0, axis=0))
```

```
Hapus Data Asal :
         nim  nama  umur
0  10519001  Adit    19
1  10519002  Dede    20
2  10519003  Bayu    20
3  10519004  Cici    19
Hapus Data Asal :
         nim  nama  umur      asal
1  10519002  Dede    20     Garut
2  10519003  Bayu    20   Cirebon
3  10519004  Cici    19  Surabaya
```

3. Preprocessing

In [36]:

```
noDup = concatDf.drop_duplicates('nama')
print(noDup)
```

```
        nim    nama  umur      asal
0  10519001    Adit    19   Bandung
1  10519002    Dede    20     Garut
2  10519003    Bayu    20   Cirebon
3  10519004    Cici    19  Surabaya
4  10519005  Firman    20   Bandung
```

In [52]:

```
# Normalisasi Jangkauan [0,1]
min = noDup['umur'].min()
max = noDup['umur'].max()
# noDup['umur2'] = (noDup['umur']-min)/(max-min)
# print(noDup)
```

In [51]:

```
copyDf = noDup.copy()
copyDf['umur2'] = (noDup['umur']-min)/(max-min)
print(copyDf)
```

```
        nim    nama  umur      asal  umur2
0  10519001    Adit    19   Bandung    0.0
1  10519002    Dede    20     Garut    1.0
2  10519003    Bayu    20   Cirebon    1.0
3  10519004    Cici    19  Surabaya    0.0
4  10519005  Firman    20   Bandung    1.0
```

In [65]:

```
# Agregasi
agreDf = copyDf.groupby(['asal'])['umur'].mean()
```

```
print(agreDf)
```

```
asal
Bandung     19.5
Cirebon     20.0
Garut       20.0
Surabaya    19.0
Name: umur, dtype: float64
```