

파이썬으로 엑셀 다루기

학습 목표

- 파이썬 환경에서 엑셀 파일을 다루는 기본 내용을 익힌다.

학습 내용

- 엑셀 파일 읽기(openpyxl, pandas)
- 내용 변경 후, 엑셀 추가

In [15]:

```
import openpyxl as oxl

print("openpyxl Ver : ", oxl.__version__)
```

openpyxl Ver : 3.0.5

01. 엑셀 파일 읽기

In [16]:

```
from openpyxl import Workbook, load_workbook
```

In [17]:

```
wb2 = load_workbook("주소록.xlsx")

print(wb2.sheetnames)
```

['Sheet', '두번째시트', 'Sheet1']

02. 시트의 내용 읽기

In [18]:

```
ws = wb2[ '두번째시트' ]

for row in ws.iter_rows():
    for cell in row:
        print(cell, cell.value)
```

```
<Cell '두번째시트'.A1> 이름
<Cell '두번째시트'.B1> 전화번호
<Cell '두번째시트'.A2> 홍길동
<Cell '두번째시트'.B2> 02-302-3333
<Cell '두번째시트'.A3> 김길동
<Cell '두번째시트'.B3> 010-222-3334
<Cell '두번째시트'.A4> toto
<Cell '두번째시트'.B4> 010-222-3334
<Cell '두번째시트'.A5> None
<Cell '두번째시트'.B5> 02-302-3333
<Cell '두번째시트'.A6> 홍길동
<Cell '두번째시트'.B6> 7777777
```

03. Pandas를 이용한 내용 읽기

In [19]:

```
import pandas as pd

print(pd.__version__)
```

1.1.3

In [20]:

```
# 제일 처음 내용 불러온다.
dat = pd.read_excel("주소록.xlsx")
dat
```

Out[20]:

—

In [22]:

```
# 시트를 지정해서 읽을 수 있음.
dat = pd.read_excel("주소록.xlsx", "두번째시트")
dat
```

Out[22]:

	이름	전화번호
0	홍길동	02-302-3333
1	김길동	010-222-3334
2	toto	010-222-3334
3	NaN	02-302-3333
4	홍길동	7777777

행단위로 읽어와 출력해 보기

In [23]:

```
for idx, row in dat.iterrows():
    print(idx, row[0], row[1])
```

```
0 홍길동 02-302-3333
1 김길동 010-222-3334
2 toto 010-222-3334
3 nan 02-302-3333
4 홍길동 7777777
```

In [24]:

```
### 원본 내용 백업
dat_ori = dat.copy()
```

이름 결측치 있는지 체크

In [25]:

```
dat.loc[ dat['이름'].isnull() , "이름체크" ] = "NOK"
dat.loc[ dat['이름'].notnull() , "이름체크" ] = "OK"
dat
```

Out[25]:

	이름	전화번호	이름체크
0	홍길동	02-302-3333	OK
1	김길동	010-222-3334	OK
2	toto	010-222-3334	OK
3	NaN	02-302-3333	NOK
4	홍길동	7777777	OK

전화번호 문제 없는지 확인

In [26]:

```
# 2,3글자 숫자-3,4글자 숫자-4글자 숫자
dat['전화번호'].str.contains('\d{2,3}-\d{3,4}-\d{4}')
```

Out[26]:

```
0      True
1      True
2      True
3      True
4     False
Name: 전화번호, dtype: bool
```

In [27]:

```
# 전화번호 확인
ch_tel = dat['전화번호'].str.contains('\d{2,3}-\d{3,4}-\d{4}')
```

In [28]:

```
dat.loc[ ch_tel , "전화번호체크" ] = "OK"
dat.loc[ -ch_tel , "전화번호체크" ] = "NOK"
dat
```

Out[28]:

	이름	전화번호	이름체크	전화번호체크
0	홍길동	02-302-3333	OK	OK
1	김길동	010-222-3334	OK	OK
2	toto	010-222-3334	OK	OK
3	NaN	02-302-3333	NOK	OK
4	홍길동	7777777	OK	NOK

Pandas 이용 파일로 만들기

- Pandas에서 ExcelWriter를 이용하면 기존의 엑셀 파일에 새로운 시트로 추가 가능.
- to_excel()로 하게 되면 동일 파일 있을 경우, 기존 파일 삭제 후, 새 파일이 만들어짐.

In [29]:

```
writer = pd.ExcelWriter("주소록_chk.xlsx", engine='openpyxl')
dat_ori.to_excel(writer, sheet_name="sheet_ori")
dat.to_excel(writer, sheet_name="sheet_chk")
writer.save()
```