

파이썬으로 엑셀 다루기

학습 목표

- 주어진 엑셀의 파일의 내용이 정확한지 파일의 내용을 읽어와 체크를 해본다.

학습 내용

- openpyxl를 이용하여 엑셀 파일을 읽어오는 것을 알아본다.
- pandas를 이용하여 엑셀 파일 읽기를 알아본다.
- 엑셀 파일의 내용을 체크하여 결과를 추가해 본다.
- 최종 확인한 결과를 파일로 저장하여 보자.

목차

01. 엑셀 파일 읽기

02. Pandas를 이용한 내용 읽기

```
In [ ]: import openpyxl as oxl

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

openpyxl Ver : 3.0.5
```

01. 엑셀 파일 읽기

목차로 이동하기

```
In [ ]: from openpyxl import Workbook, load_workbook
```

파일 읽어오기

```
In [ ]: wb2 = load_workbook("주소록.xlsx")
print(wb2.sheetnames)

ws = wb2['두번째시트']
print(type(ws))

['Sheet', '두번째시트', 'Sheet1']
<class 'openpyxl.worksheet.worksheet.Worksheet'>
```

시트의 내용 읽기

- iter_rows()를 사용하여 행 단위로 내용에 접근 가능
- iter_cols()를 사용하여 열 단위로 내용에 접근 가능

```
In [ ]: 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

```

02. Pandas를 이용한 내용 읽기

목차로 이동하기

```

In [ ]: import pandas as pd

print(pd.__version__)

1.1.3

```

```

In [ ]: # 제일 처음 시트의 내용 불러온다.
dat = pd.read_excel("주소록.xlsx")
dat

```

Out[]: 시트명 Sheet

- 두번째 시트의 내용을 불러오고 싶을 때,

```

In [ ]: # 시트를 지정해서 읽을 수 있음.
dat = pd.read_excel("주소록.xlsx", sheet_name="두번째시트")
dat

```

Out[]:

	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 [ ]: 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 [ ]: ### 원본 내용 백업
dat_ori = dat.copy()

```

이름 결측치 있는지 체크

```

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

```

```
Out[ ]:
```

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

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

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

```
In [ ]: dat.loc[ch_tel, "전화번호체크"] = "OK"
dat.loc[~ch_tel, "전화번호체크"] = "NOK"
dat
```

```
Out[ ]:
```

	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

- ExcelWriter를 사용하지 않고, to_excel()로만 파일 생성 시, 기존 파일 삭제 후, 새 파일이 만들어집니다.
- 만약 기존의 파일에 시트로 추가하고 싶을 때는 ExcelWriter를 이용하면 새로운 시트로 추가가 가능합니다.

```
In [ ]: # engine : 쓰기를 위해 사용하는 엔진.
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()
```

REFERENCE

목차로 이동하기

- openpyxl 전체 메뉴얼 : <https://openpyxl.readthedocs.io/en/stable/>
- openpyxl Tutorial : <https://openpyxl.readthedocs.io/en/latest/tutorial.html>
- ExcelWriter Class :
<https://pandas.pydata.org/docs/reference/api/pandas.ExcelWriter.html>