

가. Python openpyxl로 Excel 파일 읽고 쓰기

앞서 에서 Python의 Pandas를 이용하여 Excel 파일을 읽고 쓰는 것은 간단히 해봤습니다. 언제나 그렇듯이 제 개인적인 사용범위에서 말씀드리면, 데이터에만 관심이 있다 보니 pandas로 읽고 쓰는 정도로도 충분한데, 아주 가끔 엑셀 파일로 보고서 등을 작성할 때 아쉬울 때가 있습니다.

업무를 하다보면 아주 정형화된 보고서를 작성하는 일이 반복적으로 이루어질 때가 있습니다. 예를 들어, 학생의 성적표나 개인 건강검진표, 관리비 고지서, 견적서, 거래명세서 등과 같이 특정 셀에 값만 바뀔 뿐 전체적인 형태가 바뀌지 않을 만한 문서를 만들 때, openpyxl 패키지를 이용한 엑셀 수정이 유용할 수 있습니다.

openpyxl은 엑셀 파일을 읽는 것 뿐만 아니라 엑셀파일을 작성할때, 셀을 병합하고 폰트, 셀 등의 스타일을 변경하는 등의 많은 기능을 제공합니다. 여기서 제 개인적으로는 셀의 스타일을 변경하면서 아무것도 없는 상태에서 엑셀파일을 생성하는 것보다는 미리 작성된 엑셀 파일을 최소로 수정하여 새로운 이름으로 파일을 저장하는 방법을 권해드리며, 간단한 예제를 보이도록 하겠습니다.

1) openpyxl로 excel 파일 읽기

2020년 2학기 중간고사 성적											
	A	B	C	D	E	F	G	H	I	J	K
1	2020년 2학기 중간고사 성적										
2	이름	과목							평균	등수	총 인원
3		국어	영어	수학	과학	체육	음악	미술			
4	하윤	88	56	25	73	59	16	17	47.7	12	20
5	서윤	13	36	76	43	59	18	53	42.6	18	20
6	서연	71	12	37	24	88	50	86	52.6	8	20
7	하은	42	59	66	38	71	67	39	54.6	5	20
8	지유	63	51	23	66	14	13	71	43.0	17	20
9	지우	20	58	29	26	27	89	53	43.1	16	20
10	하린	79	27	36	60	81	20	70	53.3	6	20
11	수아	15	42	75	26	57	58	63	48.0	11	20
12	지아	40	47	74	47	57	15	91	53.0	7	20
13	서아	96	16	39	72	41	36	16	45.1	13	20
14	도윤	30	27	74	77	59	59	17	49.0	10	20
15	하준	27	27	20	55	16	23	79	35.3	20	20
16	서준	58	63	80	46	84	57	59	63.9	2	20
17	시우	46	84	78	93	79	78	39	71.0	1	20
18	민준	81	26	18	17	80	62	23	43.9	15	20
19	예준	76	14	39	34	72	50	23	44.0	14	20
20	주원	22	54	98	42	80	82	20	56.9	4	20
21	유준	71	74	13	56	81	43	26	52.0	9	20
22	지호	60	16	30	47	76	95	95	59.9	3	20
23	준우	53	53	36	62	12	32	41	41.3	19	20
24											
25											

위의 그림은 학생 20명의 각 과목의 성적, 평균, 등수 등이 기재되어 있는 엑셀파일입니다.

위의 코드의 출력 결과는 다음과 같습니다.

시트에서 'A' 컬럼에 있는 값들을 출력해보고, 'A'컬럼에 몇개의 로우(row)가 있는지 출력해보겠습니다.

위에 엑셀 파일을 보시면 4번째 row에서부터 실질적인 학생 이름과 성적, 등수 등이 있습니다. 이를 tmp라는 리스트에 담아 보도록 하겠습니다.

위의 그림은 학생 20명의 각 과목의 성적, 평균, 등수 등이 기재되어 있는 엑셀파일입니다.

```
from openpyxl import load_workbook
```

```
wb = load_workbook(filename = "2 학기중간고사통합.xlsx", data_only=True)
```

```
print(wb.sheetnames)
```

```
ws = wb[wb.sheetnames[0]]
```

```
print(ws)
```

위의 코드의 출력 결과는 다음과 같습니다.

```
['2 학기중간고사']
```

```
<Worksheet "2 학기중간고사">
```

시트에서 'A' 컬럼에 있는 값들을 출력해보고, 'A'컬럼에 몇개의 로우(row)가 있는지 출력해보겠습니다.

```
for cell in ws['A']:
```

```
    print(cell.value)
```

```
print(len(ws['A']))
```

```
print(ws.cell(row = 3, column = 1).value)
```

```
2020 년 2 학기 중간고사 성적
```

```
이름
```

```
None
```

```
하윤
```

```
서윤
```

```
...
```

```
지호
```

```
준우
```

```
23
```

```
None
```

위에 엑셀 파일을 보시면 4 번째 row 에서부터 실질적인 학생 이름과 성적, 등수 등이 있습니다. 이를 tmp 라는 리스트에 담아 보도록 하겠습니다.

```
tmp = []

for i in range(4,ws.max_row+1):

    tmp.append([])

    for cell in ws[i]:

        tmp[-1].append(cell.value)

print(tmp)

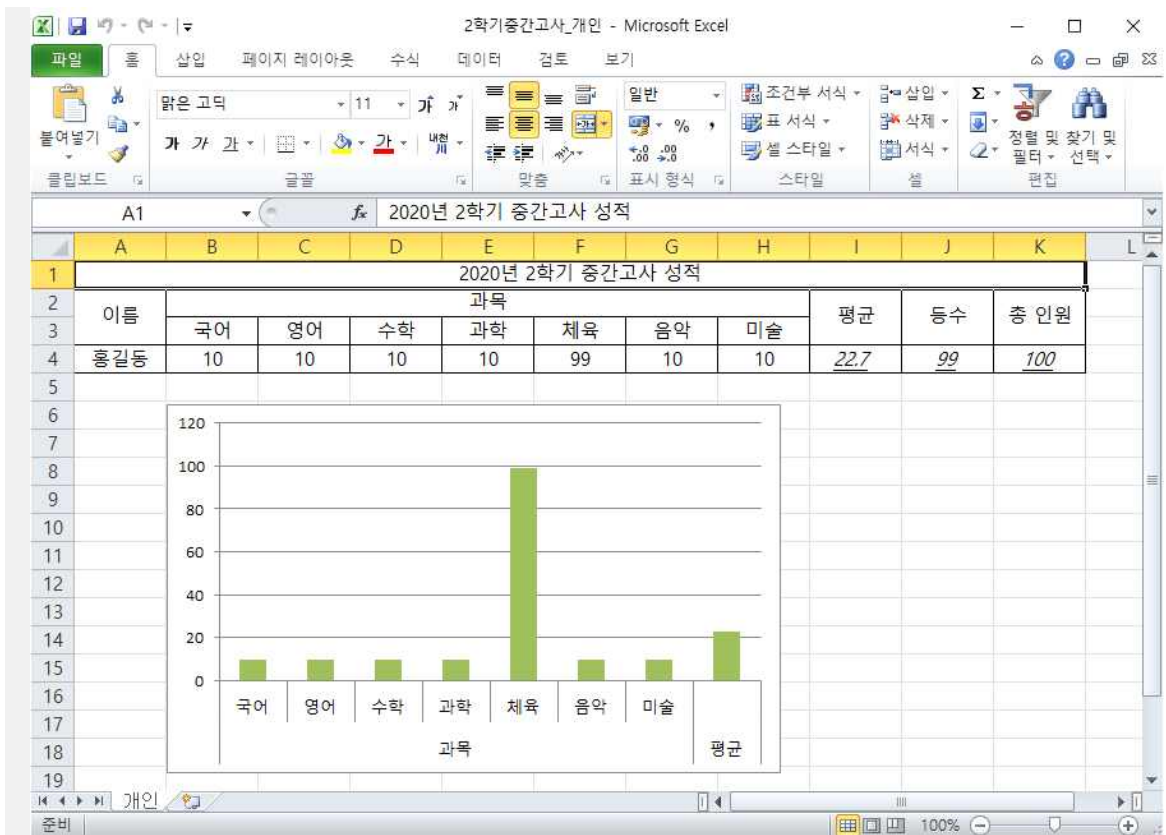
[['하윤', 88, 56, 25, 73, 59, 16, 17, 47.714285714285715, 12, 20],
 ['서윤', 13, 36, 76, 43, 59, 18, 53, 42.57142857142857, 18, 20],
 ...
 ['지호', 60, 16, 30, 47, 76, 95, 95, 59.857142857142854, 3, 20],
 ['준우', 53, 53, 36, 62, 12, 32, 41, 41.285714285714285, 19, 20]]
```

출처: <https://besixdouze.net/39> [Bésixdouze]

2) openpyxl로 excel 파일 쓰기




















앞서 잠깐 언급하였듯 openpyxl로 엑셀 파일을 작성할 때 스타일 편집 등이 다양하게 제공되고 있지만, 개인적으로 원하는 것은 이미 엑셀 서식을 만들어놓고 일부 데이터만 수정하는 것이 여러모로 편하긴 합니다.

위에서 엑셀파일을 openpyxl로 읽었으니, 개인별 성적표를 만들어보겠습니다. 우선 아래와 같이 개인 성적만 기재되어 있고, 간단한 차트 하나 그려진 서식을 만들어봤습니다.

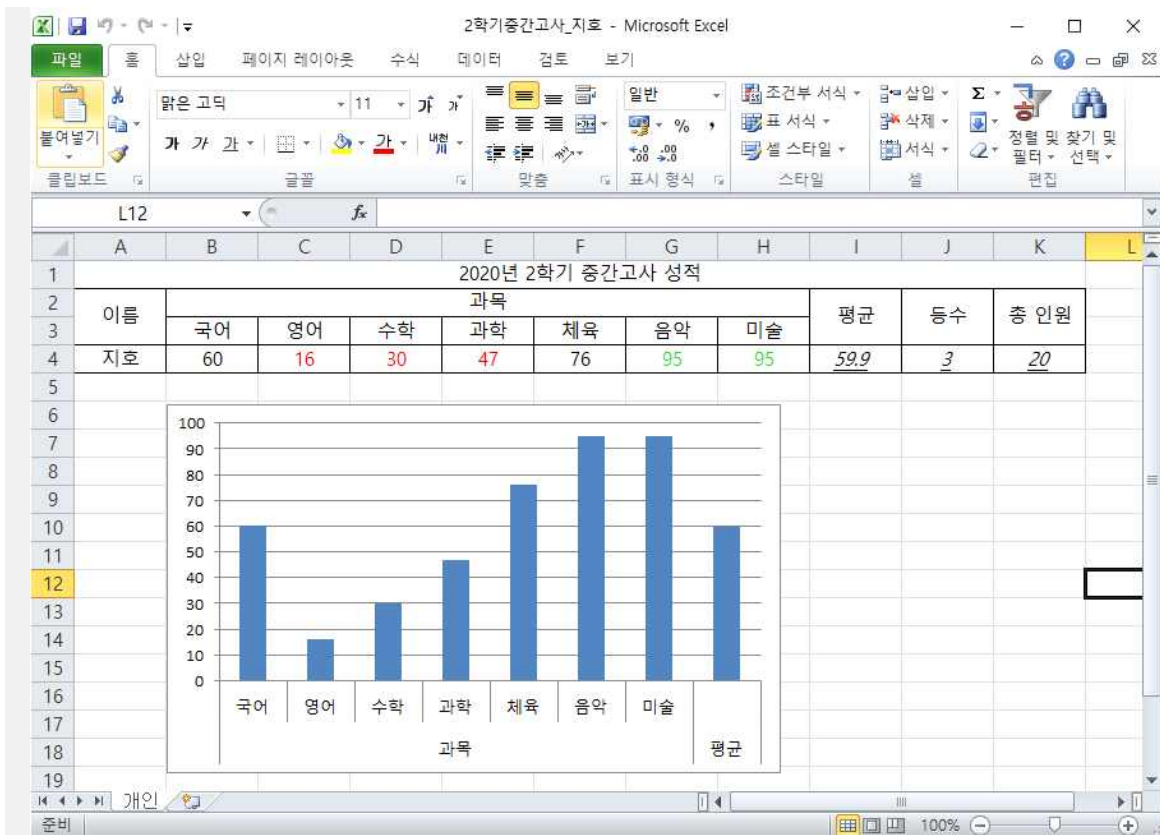


앞서 20명의 학생이 있었으니, 20개의 성적표 파일("2학기중간고사_이름.xlsx")을 만들겠습니다. VBA를 잘하시는 분들이야 엑셀에서 모든 것을 해결하시겠지만, 저는 엑셀을 못 하기에 파이썬을 합니다.

위의 코드를 실행했더니 서식파일을 기반으로 각 학생들의 성적표가 각각의 파일로 생성되었습니다.

이름	수정한
 2학기중간고사_민준	2020-0
 2학기중간고사_서아	2020-0
 2학기중간고사_서연	2020-0
 2학기중간고사_서윤	2020-0
 2학기중간고사_서준	2020-0
 2학기중간고사_수아	2020-0
 2학기중간고사_시우	2020-0
 2학기중간고사_예준	2020-0
 2학기중간고사_유준	2020-0
 2학기중간고사_주원	2020-0
 2학기중간고사_준우	2020-0
 2학기중간고사_지아	2020-0
 2학기중간고사_지우	2020-0
 2학기중간고사_지유	2020-0
 2학기중간고사_지호	2020-0
 2학기중간고사_하린	2020-0
 2학기중간고사_하윤	2020-0
 2학기중간고사_하은	2020-0
 2학기중간고사_하준	2020-0

이제 성적표의 점수에 따라 폰트 색깔이 바뀌었는지 확인해보면 아래와 같습니다.



반복적으로 정형화된 문서를 작성해야 할 때 시간을 내어 코드를 하나 만들어두면, 업무 자동화가 되어 시간을 상당히 단축시킬 수 있습니다.

openpyxl에 대한 예제는 다음 링크에 가면 더 많이 확인 가능합니다.

2. openpyxl 에 대한 예제는 다음 링크에 가면 더 많이 확인 가능합니다.

3. openpyxl.readthedocs.io/en/stable/tutorial.html

4. 출처: <https://besixdouze.net/39> [Bésixdouze]

5. 1. Excel 에 데이터 쓰기

```
from openpyxl import Workbook

# 엑셀파일 쓰기

write_wb = Workbook()

# 이름이 있는 시트를 생성

write_ws = write_wb.create_sheet('생성시트')

# Sheet1 에다 입력

write_ws = write_wb.active

write_ws['A1'] = '숫자'

#행 단위로 추가

write_ws.append([1,2,3])
```



```
#셀 단위로 추가

write_ws.cell(5, 5, '5 행 5 열')

write_wb.save("C:/Users/Administrator/Desktop/기준/프로그래밍/과제대행/주식데이터크롤링/숫자.xlsx")
```

6. 2. Excel 데이터 읽기

```
from openpyxl import load_workbook

# data_only=True 로 해줘야 수식이 아닌 값으로 받아온다.

load_wb = load_workbook("C:/Users/Administrator/Desktop/기준/프로그래밍/과제대행/주식데이터크롤링/output.xlsx",
data_only=True)

# 시트 이름으로 불러오기

load_ws = load_wb['Sheet1']
```

```
# 셀 주소로 값 출력
```

```
print(load_ws['B2'].value)
```

```
# 셀 좌표로 값 출력
```

```
print(load_ws.cell(3, 2).value)
```

```
# 지정한 셀의 값 출력
```

```
get_cells = load_ws['B3' : 'B6']
```

```
for row in get_cells:
```

```
    for cell in row:
```

```
        print(cell.value)
```

```
# 모든 행 단위로 출력
```

```
for row in load_ws.rows:
```

```
    print(row)
```

```
# 모든 열 단위로 출력
```

```
for column in load_ws.columns:
```

```
    print(column)
```

```
# 모든 행과 열 출력
```

```
all_values = []
```

```
for row in load_ws.rows:
```

```
    row_value = []
```

```
    for cell in row:
```

```
        row_value.append(cell.value)
```

```
    all_values.append(row_value)
```

```
print(all_values)
```

```
load_ws.cell(3, 3, 51470)
```

```
load_ws.cell(4, 3, 21470)
```

```
load_ws.cell(5, 3, 1470)
```

```
load_ws.cell(6, 3, 6470)
```

```
load wb.save("C:/Users/Administrator/Desktop/기준/프로그래밍/과제대행/주식데이터크롤링/output.xlsx")
```

7.