

[3-2. 엑셀파일 로딩 및 전처리]

1. 분개장 엑셀파일 로딩

1-1. 분개장 파일 확인

	A	B	C	D	E	F	G	H	I	J	K
1	분개장										
2											
3	회사명 : 패캠호텔										
4	일자	전표번호	계정코드	계정과목	적요	차변	대변	구분	거래처명		
5	01/02	00001	00001	자본금	자본금 입금	2000000000	2000000000	차변	하나은행		
6	01/03	00001	10300	보통예금	자본금 입금	330000000		대변	하나은행		
7	01/03	00001	10300	보통예금	객실 매출	330000000		차변	호텔스닷컴		
8	01/03	00002	80100	고객용품비	고객용품비 지출	200000000		대변	ABC용품사		
9	01/03	00003	80200	기타영업비용	영업비용A	150000000		차변	A사		
10	01/04	00001	10300	보통예금	객실 매출	250000000		대변	하나은행		
11	01/05	00001	40100	객실수입	객실 매출	300000000		차변	호텔스닷컴		
12	01/06	00001	10300	보통예금	객실 매출	250000000		대변	하나은행		
13	01/06	00001	40100	객실수입	객실 매출	250000000		차변	호텔스닷컴		

- 4번째 행 : 컬럼명
- 5번째 행 아래로 데이터 존재



	A	B	C	D	E	F	G	H	I	J	K
530			40100	보통예금	객실 매출		38000000	대변	하나은행		
539	05/28	00001	10300	보통예금	객실 매출	38000000		차변	하나은행		
540			40100	객실수입	객실 매출		38000000	대변	호텔스넷		
541	05/29	00001	10300	보통예금	객실 매출	25000000		차변	하나은행		
542			40100	객실수입	객실 매출		25000000	대변	호텔스넷		
543	05/29	00002	80300	판매수수료	객실판매수수료	3600000		차변	호텔스넷		
544			10300	보통예금	객실판매수수료		3600000	대변	하나은행		
545	05/29	00003	80300	판매수수료	객실판매수수료	3200000		차변	호텔스넷		
546			10300	보통예금	객실판매수수료		3200000	대변	하나은행		
547	05/29	00004	80100	고객용품비	고객용품비 지출	16100000		차변	ABC용품사		
548			10300	보통예금	고객용품비 지출		16100000	대변	하나은행		
549	05/29	00005	80200	기타영업비용	영업비용E	15000000		차변	C사		
550			10300	보통예금	영업비용E		15000000	대변	하나은행		
551	05/30	00001	10300	보통예금	객실 매출	29000000		차변	하나은행		
552			40100	객실수입	객실 매출		29000000	대변	호텔스넷		
553	합 계					10140136300	10140136300				
554											
555											
556											

- 마지막 행 : "합 계" 단어로 데이터 종료

2. 데이터 로딩 및 전처리

2-1. 라이브러리 импорт

```
In [ ]: import pandas as pd
from pandas import DataFrame, Series
from datetime import datetime, date
from openpyxl import load_workbook
```

2-2. 기본 키값 설정

```
In [ ]: # 컬럼명 위치 행번호
colname_row_number = 4

# 데이터 시작 행번호
start_row_number = 5

# 데이터 종료 구분값
last_row_value = "합 계"
```

2-3. 엑셀파일 로딩 및 전처리

```
In [ ]: filename = "분개장_샘플.xlsx"

wb = load_workbook(filename)
ws = wb.active
```

```
In [ ]: colnames = [x.value for x in ws[colname_row_number]]
print(colnames)

['일자', '전표번호', '계정코드', '계정과목', '적요', '차변', '대변', '구분', '거래처명']
```

```
In [ ]: datalist = []
for row in ws.iter_rows(
    min_row=start_row_number,
    max_col=len(colnames),
    values_only=True
):
    if row[0] == last_row_value:
        break
    row = list(row)

    # "일자", "전표번호" 컬럼에 값이 있는 경우 해당 값을 일자, 전표번호 변수에 저장
    if row[0] is not None:
        일자 = row[0].split('/')
        일자 = date(2024, int(일자[0]), int(일자[1]))
        전표번호 = row[1]

    # row 리스트의 첫번째 값은 "일자" 변수에 저장된 값으로 대체
    row[0] = 일자

    # row 리스트의 두번째 값은 "전표번호" 변수에 저장된 값으로 대체
    row[1] = 전표번호

    # 차변 컬럼의 값이 None인 경우 해당 값을 "0"으로 대체
    if pd.isna(row[5]):
        row[5] = 0

    # 대변 컬럼의 값이 None인 경우 해당 값을 "0"으로 대체
    if pd.isna(row[6]):
        row[6] = 0

    # 전처리 완료된 리스트를 datalist에 추가
    datalist.append(row)

# 전처리 완료된 데이터를 데이터프레임으로 전환
journal = DataFrame(datalist, columns=colnames)
```

```
In [ ]: journal
```

Out[]:

	일자	전표번호	계정코드	계정과목	적요	차변	대변	구분	거래처명
0	2024-01-02	00001	00001	자본금	자본금 입금	0	20000000000	차변	None
1	2024-01-02	00001	10300	보통예금	자본금 입금	20000000000	0	대변	하나은행
2	2024-01-03	00001	10300	보통예금	객실 매출	330000000	0	차변	하나은행
3	2024-01-03	00001	40100	객실수입	객실 매출	0	330000000	대변	호텔스닷컴
4	2024-01-03	00002	80100	고객용품비	고객용품비 지출	200000000	0	차변	ABC용품사
...
543	2024-05-29	00004	10300	보통예금	고객용품비 지출	0	161000000	대변	하나은행
544	2024-05-29	00005	80200	기타영업비용	영업비용E	150000000	0	차변	C사
545	2024-05-29	00005	10300	보통예금	영업비용E	0	150000000	대변	하나은행
546	2024-05-30	00001	10300	보통예금	객실 매출	290000000	0	차변	하나은행
547	2024-05-30	00001	40100	객실수입	객실 매출	0	290000000	대변	호텔스닷컴

548 rows × 9 columns

2-4. 분류값 추가

- 계정과목에 대하여 상위 분류값 추가

```
In [ ]: # 분류값 테이블 생성
category_table = pd.DataFrame([
    ['자본금', '자본', '자본금', '자본금'],
    ['보통예금', '유동자산', '현금및현금성자산', '보통예금'],
    ['객실수입', '매출', '매출', '객실수입'],
    ['고객용품비', '판매비와관리비', '고객용품비', '고객용품비'],
    ['기타영업비용', '판매비와관리비', '판매비용', '기타영업비용'],
    ['판매수수료', '판매비와관리비', '판매비용', '판매수수료'],
    ['세금과공과', '판매비와관리비', '기타판관비', '세금과공과'],
    ['수도광열비', '판매비와관리비', '기타판관비', '수도광열비'],
    ['급여', '판매비와관리비', '인건비', '급여'],
    ['노무용역비', '판매비와관리비', '인건비', '노무용역비'],
    ['이자수입', '영업외수익', '영업외수익', '이자수입']
],
    columns=['구분', '대분류', '중분류', '소분류']
)
category_table.set_index('구분', inplace=True)
```

```
In [ ]: category_table
```

```
Out[ ]:
```

	대분류	중분류	소분류
구분			
자본금	자본	자본금	자본금
보통예금	유동자산	현금및현금성자산	보통예금
객실수입	매출	매출	객실수입
고객용품비	판매비와관리비	고객용품비	고객용품비
기타영업비용	판매비와관리비	판매비용	기타영업비용
판매수수료	판매비와관리비	판매비용	판매수수료
세금과공과	판매비와관리비	기타판관비	세금과공과
수도광열비	판매비와관리비	기타판관비	수도광열비
급여	판매비와관리비	인건비	급여
노무용역비	판매비와관리비	인건비	노무용역비
이자수입	영업외수익	영업외수익	이자수입

```
In [ ]: # journal 데이터프레임에 "대분류", "중분류", "소분류" 컬럼 생성
journal['대분류'] = None
journal['중분류'] = None
journal['소분류'] = None

for idx in journal.index:
    # journal 데이터프레임의 "계정과목" 컬럼 값 추출
    key = journal.loc[idx, '계정과목']

    # category_table에서 추출된 키값에 해당하는 "대분류", "중분류", "소분류"
    # 값을 journal에 대체
    journal.loc[idx, '대분류'] = category_table.loc[key, '대분류']
    journal.loc[idx, '중분류'] = category_table.loc[key, '중분류']
    journal.loc[idx, '소분류'] = category_table.loc[key, '소분류']
```

```
In [ ]: journal
```

Out []:

	일자	전표번호	계정코드	계정과목	적요	차변	대변	구분	거래처명	대분류	중분류	소분류
0	2024-01-02	00001	00001	자본금	자본금입금	0 20000000000		차변	None	자본	자본금	자본금
1	2024-01-02	00001	10300	보통예금	자본금입금	20000000000	0	대변	하나은행	유동자산	현금및현금성자산	보통예금
2	2024-01-03	00001	10300	보통예금	객실매출	330000000	0	차변	하나은행	유동자산	현금및현금성자산	보통예금
3	2024-01-03	00001	40100	객실수입	객실매출	0 330000000		대변	호텔스닷컴	매출	매출	객실수입
4	2024-01-03	00002	80100	고객용품비	고객용품비지출	200000000	0	차변	ABC용품사	판매비와관리비	고객용품비	고객용품비
...

현

543	2024-05-29	00004	10300	보통예금	고객용품지출	0	16100000	대변	하나은행	유동자산	금및현금성자산	보통예금
544	2024-05-29	00005	80200	기타영업비용	영업비용E	15000000	0	차변	C사	판매비와관리비	판매비용	기타영업비용
545	2024-05-29	00005	10300	보통예금	영업비용E	0	15000000	대변	하나은행	유동자산	현금및현금성자산	보통예금
546	2024-05-30	00001	10300	보통예금	객실매출	29000000	0	차변	하나은행	유동자산	현금및현금성자산	보통예금
547	2024-05-30	00001	40100	객실수입	객실매출	0	29000000	대변	호텔스닷컴	매출	매출	객실수입

548 rows × 12 columns

2-5. 컬럼 순서 정리

```
In [ ]: journal = journal[
    [
        '일자', '전표번호', '대분류', '중분류', '소분류', '계정코드',
        '계정과목', '적요', '차변', '대변', '구분', '거래처명'
    ]
]
```

```
In [ ]: journal
```

Out []:

일자	전표번호	대분류	중분류	소분류	계정코드	계정과목	적요	차변	대변	구분	거래처명
----	------	-----	-----	-----	------	------	----	----	----	----	------

0	2024-01-02	00001	자본	자본금	자본금	00001	자본금	자본금입금	0	20000000000	차변	None
1	2024-01-02	00001	유동자산	현금및현금성자산	보통예금	10300	보통예금	자본금입금	20000000000	0	대변	하나은행
2	2024-01-03	00001	유동자산	현금및현금성자산	보통예금	10300	보통예금	객실매출	33000000	0	차변	하나은행
3	2024-01-03	00001	매출	매출	객실수입	40100	객실수입	객실매출	0	33000000	대변	호텔스닷컴
4	2024-01-03	00002	판매비와관리비	고객용품비	고객용품비	80100	고객용품비	고객용품비지출	20000000	0	차변	ABC용품사
...
543	2024-05-29	00004	유동자산	현금및현금성자산	보통예금	10300	보통예금	고객용품비지출	0	16100000	대변	하나은행
544	2024-05-29	00005	판매비와관리비	판매비용	기타영업비용	80200	기타영업비용	영업비용E	15000000	0	차변	C사
545	2024-05-29	00005	유동자산	현금및현금성	보통예금	10300	보통예금	영업비용	0	15000000	대변	하나은행

자 산 E											
546	2024-05-30	00001	유 동 자 산	현 금 및 현 금 성 자 산	보 통 예 금	10300	보 통 예 금	객 실 매 출	290000000	0	차 변 하 나 은 행
547	2024-05-30	00001	매 출	매 출	객 실 수 입	40100	객 실 수 입	객 실 매 출	0	290000000	대 변 호 텔 스 닷 컴

548 rows × 12 columns

In []: