

데이터 가공 실전(아파트 매매 실거래 상세 조회)

pandas 가져오기

```
import pandas as pd
```

xml → dict로 만드는 xmltodict 설치하기

```
# xml을 dict로 변환하는 라이브러리 설치하기
```

```
!pip install xmltodict
```

```
Collecting xmltodict
  Downloading xmltodict-0.13.0-py2.py3-none-any.whl
(10.0 kB) Installing collected packages: xmltodict
Successfully installed xmltodict-0.13.0
```

아파트 매매 실거래가 자료 받아와서 xml을 dict로 변환하기

```
# requests 라이브러리 가져오기
```

```
import requests
```

```
url =
```

```
'http://openapi.molit.go.kr/OpenAPI_ToolInstallPackage/service/rest/RTMSOBSvc/getRTMSDataSvcAp
tTradeDev?serviceKey=Qm%2FSxTC4k4vYtw5LKeqYtg%2F params ={'pageNo' : '1', 'numOfRows' : '30',
'LAWD_CD' : '11650', 'DEAL_YMD' : '202401' }
```

```
response = requests.get(url, params=params)
```

```
# www.code.go.kr 에서 법정동 코드 10자리 중 앞 5자리
```

```
# 응답받은 xml형식의 데이터를 파이썬의 dict로 변경하기.
```

```
import xmltodict
```

```
data=xmltodict.parse(response.content)
```

pprint로 데이터 구조 파악하기

```
import pprint
```

```
pprint.pprint(data)
```

```
{'response': {'body': {'items': {'item': [{'거래금액': '260,000',
'거래유형': '중개거래',
'건축년도': '2001',
'년': '2024',
'도로명': '방배로',
'도로명건물본번호코드':
'00249',
'도로명건물부번호코드': '00000',
'도로명시군구코드': '11650',
'도로명일련번호코드': '01',
'도로명지상지하코드': '0',
'도로명코드': '3121010',
'동': None,
'등기일자':
None, '매도자':
'개인',
'매수자': '개인',
'법정동': '방배동',
'법정동본번호코드': '0772',
'법정동부번호코드': '0013',
'법정동시군구코드': '11650',
'법정동읍면동코드': '10100',
'법정동지번호코드': '1',
'아파트': '현대엠피스',
'월': '1',
```

```
'일': '2',
'일련번호': '11650-373',
'전용면적': '158.82',
'중개사소재지': '서울 서초구',
'지번': '772-13',
'지역코드': '11650',
'층': '11',
'해제사유발생일':
None, '해제여부':
None},
{'거래금액': '29,800',
'거래유형': '중개거래',
'건축년도': '2005',
'년': '2024',
'도로명': '서초대로',
'도로명건물본번호코드':
'00017',
'도로명건물부번호코드': '00000',
'도로명시군구코드': '11650',
'도로명일련번호코드': '01',
'도로명지상지하코드': '0',
'도로명코드': '2121002',
'동': None,
'등기일자': '24.02.01',
'매도자': '개인',
'매수자': '개인',
'법정동': '방배동',
'법정동본번호코드': '0866',
'법정동부번호코드': '0020',
'법정동시군구코드': '11650',
'법정동읍면동코드': '10100',
'법정동지번코드': '1',
'아파트': '방배대우디오빌',
'월': '1',
'일': '4',
'일련번호': '11650-3191',
'전용면적': '29.6',
```

전체 항목으로 데이터프레임 만들기

```
df=pd.DataFrame(data['response']['body']['items']['item'])
```

데이터프레임 정보 확인

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>RangeIndex: 30 entries, 0 to 29
Data columns (total 32 columns):
```

```
#   Column      Non-Null Count  Dtype
---

```

```
0   거래금액      30 non-null    object
1   거래유형      30 non-null    object
2   건축년도      30 non-null    object
3   년            30 non-null    object
4   도로명         30 non-null    object
5   도로명건물본번호코드  30 non-null    object
6   도로명건물부번호코드  30 non-null    object
7   도로명시군구코드  30 non-null    object
8   도로명일련번호코드  30 non-null    object
9   도로명지상지하코드  30 non-null    object
10  도로명코드     30 non-null    object
11  동             2 non-null     object
12  등기일자       3 non-null     object
13  매도자         30 non-null    object
14  매수자         30 non-null    object
15  법정동         30 non-null    object
16  법정동본번호코드  30 non-null    object
17  법정동부번호코드  30 non-null    object
18  법정동시군구코드  30 non-null    object
19  법정동읍면동코드  30 non-null    object
20  법정동지번코드  30 non-null    object
21  아파트         30 non-null    object
22  월             30 non-null    object
23  일             30 non-null    object
24  일련번호       30 non-null    object
25  전용면적       30 non-null    object
26  중개사소재지   29 non-null    object
27  지번           30 non-null    object
28  지역코드       30 non-null    object
29  층             30 non-null    object
30  해제사유발생일  0 non-null     object
31  해제여부       0 non-null     objectdtypes: object(32)
```

```
memory usage: 7.6+ KB
```

필요한 항목만으로 데이터프레임 만들기

필요한 필드(열) 선택해서 데이터프레임 만들기

```
df=df[['년', '월', '아파트', '건축년도', '거래금액', '전용면적', '층', '도로명', '법정동', '거래유형', '매도자', '매수자']]
```

```
df.head()
```

	년	월	아파트	건축년도	거래금액	전용면적	층	도로명	법정동	거래유형	매도자	매수자
0	2024	1	현대매퍼스	2001	260,000	158.82	11	방배로	방배동	중개거래	개인	개인
1	2024	1	방배대우디오빌	2005	29,800	29.6	9	서초대로	방배동	중개거래	개인	개인
2	2024	1	대림e-편한세상1	2003	255,000	164.25	14	서초대로34길	방배동	중개거래	개인	개인
3	2024	1	신동아	1981	235,000	106.42	2	효령로	방배동	중개거래	개인	개인
4	2024	1	방배아트자이	2018	160,000	59.9775	4	명달로9길	방배동	중개거래	개인	개인

건축년도, 거래금액, 전용면적을 숫자형으로 변경하기

astype()으로 데이터 형식 변경하기

```
df['건축년도'] = pd.to_numeric(df['건축년도']).astype(float)
df['거래금액'] = df['거래금액'].str.replace(',','',').astype(int)
df['전용면적'] = pd.to_numeric(df['전용면적']).astype(float)
df.head()
```

	년	월	아파트	건축년도	거래금액	전용면적	층	도로명	법정동	거래유형	매도자	매수자
0	2024	1	현대매퍼스	2001.0	260000	158.8200	11	방배로	방배동	중개거래	개인	개인
1	2024	1	방배대우디오빌	2005.0	29800	29.6000	9	서초대로	방배동	중개거래	개인	개인
2	2024	1	대림e-편한세상1	2003.0	255000	164.2500	14	서초대로34길	방배동	중개거래	개인	개인
3	2024	1	신동아	1981.0	235000	106.4200	2	효령로	방배동	중개거래	개인	개인
4	2024	1	방배아트자이	2018.0	160000	59.9775	4	명달로9길	방배동	중개거래	개인	개인

건축년도로 내림차순, 전용면적으로 내림차순하기

특정 열로 데이터프레임 정렬하기

```
sorted_df = df.sort_values(by=['건축년도', '전용면적'], ascending=[False, False])
sorted_df.head()
```

	년	월	아파트	건축년도	거래금액	전용면적	층	도로명	법정동	거래유형	매도자	매수자
4	2024	1	방배아트자이	2018.0	160000	59.9775	4	명달로9길	방배동	중개거래	개인	개인
16	2024	1	엘에이치서초5단지(토지임대부아파트)	2013.0	115000	84.9500	10	양재대로2길	우면동	중개거래	개인	개인
6	2024	1	롯데캐슬아르떼	2013.0	200000	84.9145	4	방배천로18길	방배동	중개거래	개인	개인
19	2024	1	엘에이치서초5단지(토지임대부아파트)	2013.0	90000	59.9300	6	양재대로2길	우면동	중개거래	개인	개인
20	2024	1	서초힐스	2012.0	134000	84.9600	19	양재대로2길	우면동	중개거래	개인	개인

데이터프레임을 엑셀 파일로 저장하기

```
sorted_df.to_excel('아파트매매가.xlsx')
```