4/24 건축물 데이터 전처리 상태

진행도: 60% (4 / 8)

1. 건축물대장 - 완료

2. 표제부 - 완료

3. 총괄표제부 - 완료

4. 층별개요 – 진행중(60%)

5. 지역지구구역 – 진행중(80%)

6. 전유부 - 완료

7. 전유공유면적표 – 진행중(60%)

8. 통합 - 준비중

<세부사항>

**층별 개요**

* 주용도코드에 대해 코드 형태의 데이터를 우리가 읽을 수 있는 형태로 치환 예정
* 부속건축물에 대한 데이터도 6685개가 존재함을 확인. 대부분의 상가는 주건축물에 존재하지만 부속 건축물에 상가를 차린다고 한다면 어떻게 해야 할지 고민해볼 문제.
* 주용도코드 치환 후 어느정도 내용을 담고 있다면 기타용도 칼럼 지울 예정.

**지역 지구 구역**

* 노트북에 주석으로 남겨놓은 자료인데, 지역지구구역별 용적률의 최고치가 존재하는데 그를 바탕으로 초과 용적률과 미활용 용적률이라는 변수를 생성할 수 있다. 근데 용적률이라는게 높을 수록 좋은거라 두 변수 모두 중소형 빌딩의 매매시에 통계적으로 유의미한 영향을 준다는 논문이 있다.
* [신보연(Shin Bo Yeon),and 유선종(Yoo Seon Jong). "용적률 활용이 중소형 빌딩 가격에 미치는 영향에 관한 연구." 주거환경(한국주거환경학회논문집) 16.3 (2018): 195-213.](https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artiId=ART002391941)
* 여기서 208쪽
* 어떻게 활용할까에 대한 생각은 조금 더 필요할 듯
* 제일 골때리는게 얘는 한글로 치환된 자료랑 코드로 남아있는 자료가 공존하는 형태.
* 담주중에 다 해결할 예정. 시험만 끝나라..

**전유 공유 면적표**

* 전유공용구분 코드에서 1이 전유임을 나타내는 값이다. 공용으로 사용하는 부분은 제외하기 위해 1만 남기고 삭제할 예정
* 이것도 주용도코드 바꿔보고 주용도코드가 정보를 충분히 담고있다면 기타용도 삭제 예정
* 이 데이터의 가장 큰 문제점은 pk를 어떻게 연결하는지 모르겠다는 것이다.
* 이 부분은 통합 에서 설명

**통합**

* 이 데이터들 무지하게 불친절하다.
* 합치기 엄청나게 어렵게 해놨다…….
* 정 안되면 주택정책실 주택정책지원센터에서 제공하는 자료니까 전화해보던지 가보던지 하겠슴,,,
* 서울 열린 데이터 광장에서 데이터 받은 곳 가면 댓글 남기는 기능도 있는데 그것도 활용할 예정.
* 시험기간이라 많이 못해서 죄송합니다ㅠㅠㅠㅠㅠㅠㅠ
* 담주에 미친사람처럼 할께여,,,,,
* 남은 시험 홧팅ㅎㅎㅎ