4/30 건축물 데이터 전처리 상태

진행도: 90% (6 / 7)

1. 건축물대장 - 완료

2. 표제부 - 완료

3. 총괄표제부 - 완료

4. 층별개요 – 완료

5. 지역지구구역 – 진행중(80%)

6. 전유부 - 완료

7. 전유공유면적표 – 완료

\*\*통합 – pk로 매칭이 안되는 부분이 많아서 건축물 데이터 담당자에게 질의 예정

<세부사항>

**층별 개요**

* 주용도코드 치환하는 과정에서 1000번대별로 대분류가 존재함을 확인.
* 대분류를 통해 상가가 들어올 수 있는 근린생활시설임을 알 수 있었음.
* 기타 분류 column을 통해 해당 건물의 해당 층이 무슨 용도로 사용되는지 알 수 있음
* 주용도코드를 1000단위로 치환함.
* 1000단위로 치환한 용도코드를 코드표를 이용해 한글로 치환함
* 현재 주.부속 건축물에 대한 고려 않는 것으로 보고 처리 안했음.

**지역 지구 구역**

* 제일 골때리는게 얘는 한글로 치환된 자료랑 코드로 남아있는 자료가 공존하는 형태.
* 근데 코드 정보가 없다..?
* 영어로 시작하는 코드는 국가표준코드에서 찾을 수 있었지만 숫자로만 표현된 코드는 어디서도 정보를 얻을 수 없었음.
* 해당 문제의 경우 통합에 관련한 시청 질문 시 같이 물어볼 예정
* 이게 해결되든 말든 해야 초과용적률 등 관련된 자료에 대한 고민 해볼 듯

**전유 공유 면적표**

* 전유\_공용\_구분\_코드에서 1만 남기고 처리한 결과 엄청나게 데이터가 줄어들어서 일단은 조금 더 볼 예정
* 층별 개요와 마찬가지로 주용도코드 치환함
* 지난 주에 통합 어떻게 하는 건지 모르겠다고 했는데 이것도 물어볼 예정

**통합**

* 일단 이거 건축물 대장 전체를 풀 테이블로 만드는 것은 불가능함.
* 건물마다 나오는 건축물 대장이 다르기 때문에 NaN값이 엄청 많아짐
* 지금 생각은 묶을 수 있는 자료들끼리 묶어서 최소한의 큰 테이블을 만드는거?
* 아니면 이 데이터들을 다 DB에 넣어서 건물 관련된 정보 보여주거나 하는 작업은 View 띄우고, 자료를 분석에 사용할 때는 SQL 쿼리 찍어서 가져오는 방법? 기술 스택을 생각한다면 Hive 활용해서 하는게 가능할 듯..