기초 데이터 탐색 + 연구 목표 및 방향

1. 광양시\_공동주택\_분리수거장현황.csv

- complex\_nm 아파트단지명

- address 주소

- lon 경도

- lat 위도

2. 광양시\_재활용품분리수거\_장려금지급내역.csv

- date: 장려금 지급일자

- emd\_nm : 읍면동명

- complex\_nm : 아파트 단지명

- road\_nm : 도로명

- bldg\_no : 건물번호

- glass : 유리병

- pet : 페트병

- plastic : 플라스틱류

- lamp : 폐형광등

- vinyl : 필름류포장재비닐

- battery : 폐건전지류

- carton : 종이팩(종이컵)

- smallappliance : 소형가전

- hp : 휴대폰

- subsidy : 장려금 교부액

=> 결측치 약간 존재(3개) => 그냥 삭제해도 될듯.(1% 미만의 데이터.)

3. 광양시\_금호동\_집적장별\_재활용품통계.csv

- apt\_nm : 아파트 명

- can(kg)

- plastic(kg)

- glass(kg)

- paper(ton)

- scrapmetal : 고철(kg)

4. 광양시\_건축물\_표제부.csv

- land\_location : 주소(대지위치)(도/시/동/번지)

- sgg\_cd : 시군구 코드

- emd\_cd : 법정동 코드

- land\_gbn : 대지구분코드 <0밖에 없으므로 필요 없을듯>

- bun

- ji

- bldg\_nm : 건물명

* **사용 X**

- dong\_nm : 동명칭 (굉장히 많음. 숫자+문자 섞여 있음.)

* **사용 X**

- main\_annex\_gbn : 주부속구분코드명 (주건축물/부속건축물/Null)

* **‘부속건축물’ drop 코드**: df = df[df['main\_annex\_gbn'] == '주건축물']
* **‘etc\_purpose\_nm’ drop 값** : “창고시설” / “우산각(정자)” / “관리소”, “부대시설(전기실,펌프실)” / “창고시설” / “계단실, 가스저장실, 경비실, 관리실, 발전기실, 오수정화시설, 주차장, 펌프실”

- purpose\_cd : 주용도코드

- purpose\_nm : 주용도코드명

- etc\_purpose\_nm : 기타코드(주용도코드를 더 자세하게 설명.)

- household\_cnt : 세대수(세대)

- family\_cnt : 가구수(가구)

- ho\_cnt : 호수(호)=> 아파트에 호수 0이 왜 있지? 220세대가 사는데 호수가 39개인 경우는 또 뭐고? => 주차장 등 부속건물들이 따로 잡혀 있는데, 이름은 똑같이 “00아파트”로 들어가 있어서 그럼.

* ho\_cnt : 호수는 중요하지 않으므로 해당 열 삭제.

- lon : 경도

- lat : 위도

=> 결측치 꽤 있음. 147개 결측치. 없애기에 애매한 정도임.

=> 건물명(bldg\_name), 동 이름에 결측치가 꽤 있음.

5. 광양시\_건축물\_전유부.csv

- land\_location: 대지위치(지번주소)

- sgg\_cd : 시군구코드

- emd\_cd : 법정동코드

- land\_gbn : 대지구분코드

- bun

- ji

- bldg\_nm : 건물명

- dong\_nm : 동명칭

- ho\_nm : 호명칭

- floor\_gbn : 층구분

- floor\_num : 층번호

=> 결측치 많음. 데이터 자체도 많지만... 2203 개의 결측치 존재.

=> 동 이름 결측치가 왜 이렇게 많지?

**=> 5번파일 사용안하기로**

==================geojson 파일===================

6. 광양시\_법정경계(읍면동)

- emd\_cd : 읍면동 코드

- emd\_nm : 읍면동 명

7. 광양시\_도로명주소(건물)

- BDTYP\_CD 건물용도코드

- BULD\_NM 건축물대장 건물명

- BULD\_NM\_DC 상세건물명

- BULD\_SE\_CD 건물 구분 코드

- BUL\_MAN\_NO 건물 일련번호

- EMD\_CD 읍면동 코드

- GRO\_FLO\_CO 지상층수

- LNBR\_MNNM 지번본번

- LNBR\_SLNO 지번부번

- UND\_FLO\_CO 지하층수

8. 광양시\_도로명주소(건물출입구)

- BUL\_MAN\_NO 건물 일련 번호

- ENTRC\_SE 출입구 구분(RC:주차장출입구/RM:주 출입구/RS:보조출입구)

- ENT\_MAN\_NO 출입구일련번호

9. 광양시\_지적도

- pnu : 필지고유번호

- jibun : 지번

10. 광양시\_인구정보(총인구)

- gid : 격자고유번호

- vas : 격자내 인구수

================================================

11. 광양시\_동별인구수\_세대수

- emd\_nm 읍면동 명

- total\_pop\_2017 총인구수(2017)

- hh\_cnt\_2017 세대수(2017)

- hh\_pop\_cnt\_2017 세대당 인구(2017)

- male\_pop\_2017 남자 인구수(2017)

- female\_pop\_2017 여자 인구수(2017)

- sex\_ratio\_2017 남여 비율(2017)

- total\_pop\_2018 총인구수(2018)

- hh\_cnt\_2018 세대수(2018)

=> 이건 12 sample밖에 없어서.. 결측치 그럴것도 없네 시각화만 하면 끝날듯

12. 광양시\_장래인구\_추계치

- year 연도

- total\_pop 총인구(명)

- component\_ratio 구성비(%)

- male\_pop 남자인구(명)

- female\_pop 여자인구(명)

- sex\_ratio 성비(여자 1백명당)

- pop\_growth\_rate 인구성장율(%)

- nature\_inc\_rate 자연증가율(%)

- tot\_med\_age 중위연령전체(세)

- male\_med\_age 중위연령남자(세)

- female\_med\_age 중위연령여자(세)

- youth\_pop 유소년 인구(0~14세)

- youth\_comp\_ratio 유소년인구 구성비(%)

- prod\_pop 생산가능인구(15~64세)

- prod\_comp\_ratio 생산가능인구 구성비(%)

- old\_pop 고령인구(65세 이상)

- old\_comp\_ratio 고령인구 구성비(%)

=> 결측치가 3개씩 있는데... 이거는 인구 예측해 보라고 이렇게 넣어놓은듯. 알뜰한 광양시같으니라구^^ 뭘갖다 예측하라는거야 데이터도없는데 ^^;

=========================================================

gis에서 재활용수거차량이 다닐수 있는 도로 기반에서 소거대상(설치불가능위치) 넣고 인구, 세대수, 유동인구(있다면) 놓고 돌리는 방법을 한번 고민해 보세요

### **갖고 있는 데이터로 해야 할 목표(일단 csv파일로만.)**

1. 데이터 결측치 처리.
2. EDA와 시각화(장려금 어디가 많이 받았는지, 인구는 어디가 많은지, 인구 비율은 어떻게 되는지, 발생하는 폐품은 어디서 어떤 게 많이 발생하는지.)
3. 쓰레기 배출량 예측(대체어떻게!?)