





광양시 재활용품 자동수거기 최적 설치위치 선정 과제

골든플래닛 실습조 | 골드문 팀

김기영 박소연 송현정 이동렬



STEP 01

분석 개요

STEP **02**

EDA 및 전처리

EDA 데이터 통합



분석 내용

분리수거 발생량 예측 최적회귀모델 도출 점수화



분석 결과

분리수거기 최종 입지선정 결론 및 제언





• 광양시의 산업은 포스코 광양제철소가 위치한 금호동이 중심지 역할을 수행하며,

• 광양시의 인구는 광양읍과 중마동에 밀집되어 있음

• 행정동명 : 광양읍

● 인 구 : 48,284 (2021.09 기준)

● 특 이 사 항

- 광양시의 중심

- 호남 내 최다 인구를 보유한 읍

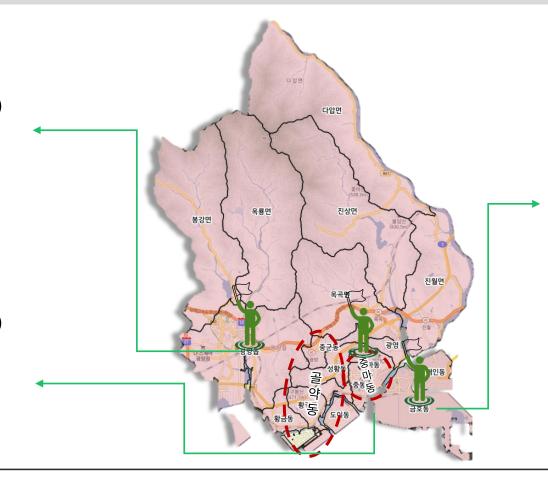
• 행정동명 : 중마동

● 인 구 : 57,492 (2021.09 기준)

● 특 이 사 항

- 광양시청이 위치함

- 다량의 주택지구가 들어서는 중임

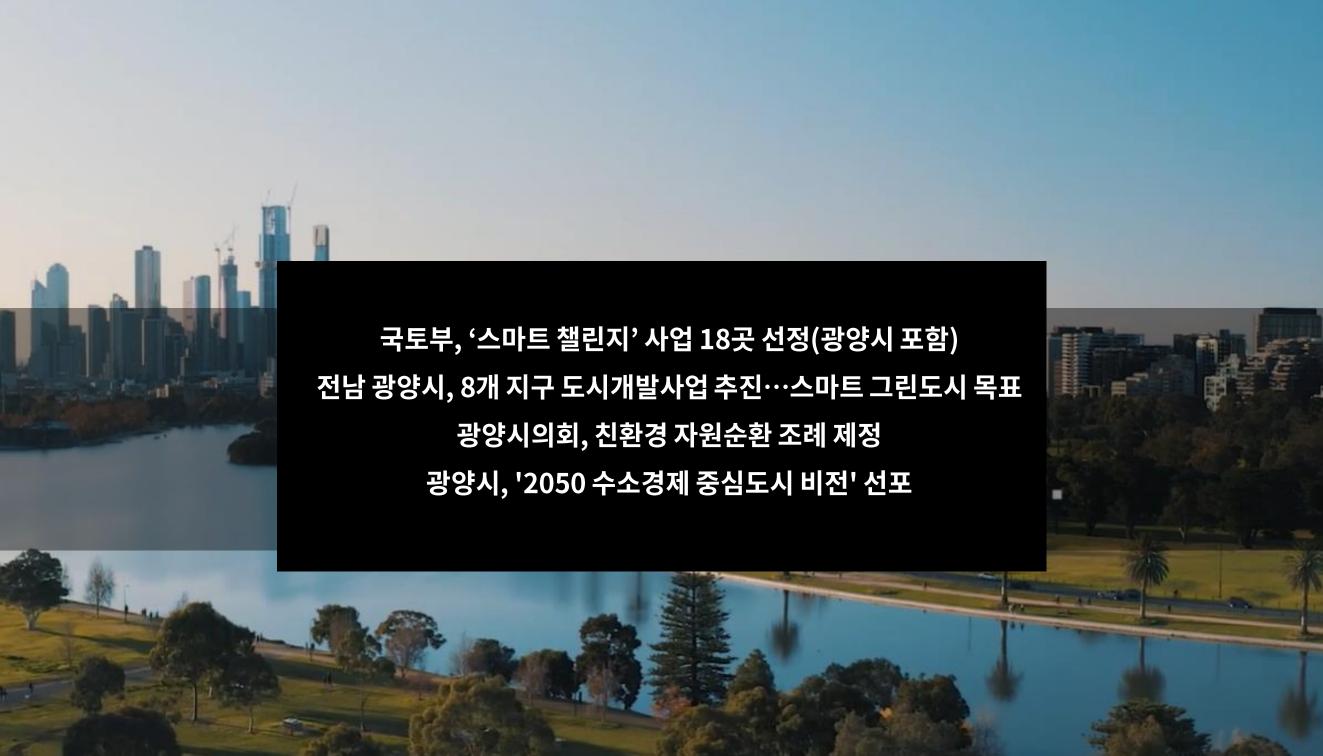


● 행정동명 : 금호동

• 인 구 : 11,691 (2021.09 기준)

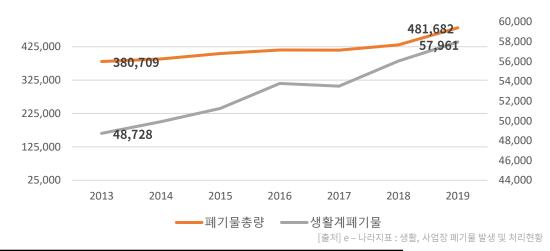
● 특 이 사 항

- 포스코 광양제철소를 중심으로 대규모 산업단지가 위치함



- 생활수준 향상 및 코로나 19이후 비대면 소비 증가로 폐기물 발생량 증가
- 재활용품 단가 하락으로 민간수거가 줄어들면서 공공수거 수요가 급증
- 광양시는 매립지 2037년 사용 종료 예정으로, 직매립 제로화와 자원순환 선진화 정책 수립 등 재활용 극대화를 위한 계획 필요.

대한민국 폐기물 배출 총량 및 생활 폐기물 배출량





[쓰레기 대란]① "쓰레기 버릴 곳 없어"…10년 내 '쓰레기 대란' 온다

송고시간 | 2021-07-23 08:00

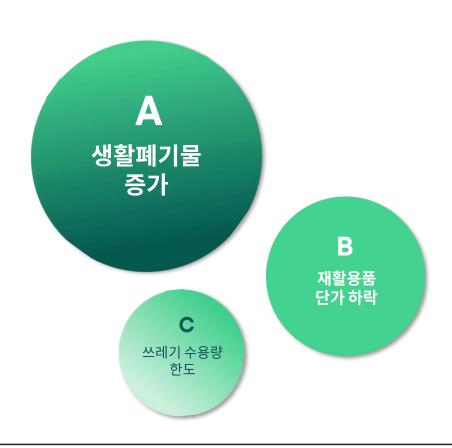
[출처] 연합뉴스 https://www.yna.co.kr/view/AKR2021071914540050

신아일보

홍천군, 코로나19 생활폐기물 증가 따른 소각능력 저하로 적치

🖰 조덕경 기자 📗 ② 승인 2021.09.23 15:41

출처] 신아일보 https://www.shinailbo.co.kr/news/articleView.html?idxno=1463269



- 올바른 방법으로 재활용품을 수거하여 재활용 가능한 쓰레기의 매립·소각을 줄임
- 수거되는 재활용품의 활용률 증가 및 생활폐기물 감소 효과
- 생활쓰레기 배출에 대한 주민들의 인식 개선 효과



교육 효과 시민들이 올바른 분리수거 방법을 자연스럽게 학습할 수 있음.



수거효율증대 올바른 방식으로 수거함으로써 재활용 가능한 자원의 질을 높일 수 있음.



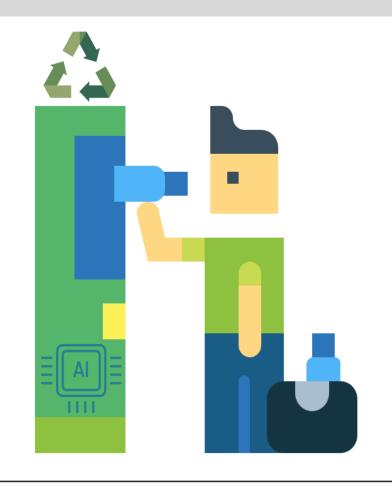
인력 감소 재활용품 분류에 필요한 인력이 감소.



환경 보전 자원 재활용률을 높여 소각으로 인한 대기 오염 예방.



경제적효율성 폐기물처리비용감소



- 적절한 위치에 자동수거기를 설치하여 수거기 운영 효율 증대
- 광양시 요구사항에 따라 생활 폐기물의 수거를 위하여 아파트 단지 내에 설치







02. EDA 및 전처리



데이터 탐색

- 1-1) 품목별, 동별 폐기물 발생현황
- 1-2) 아파트별 폐기물 발생현황
- 1-3) 금호동 폐기물 발생현황
- 1-4) 주거용도 건축물 위치확인

02

전처리

- 2-1) 사용 데이터 선정 및 필터링
- 2-2) 단위 변환 및 데이터 병합
- 2-3) 인구수 추정
- 2-4) 재활용품 품목 그룹화

03

분석 사용 컬럼

bldg_nm : 건물명

land locaiton: 주소

dong_nm: 동 이름

bun: 번

ji : 지

main_anex_gbn: 건물용도명

purpose_nm : 주용도명

etc_purpose_nm: 세부용도명

추정거주인구: 추정거주인구수

lon : 위도

lat : 경도

household_cnt : 세대수

pop_pred : 추정인구수

susidy : 총 장려금 교부액

A_recycle : 주요 배출 항목

B_recycle : 기타 배출 항목

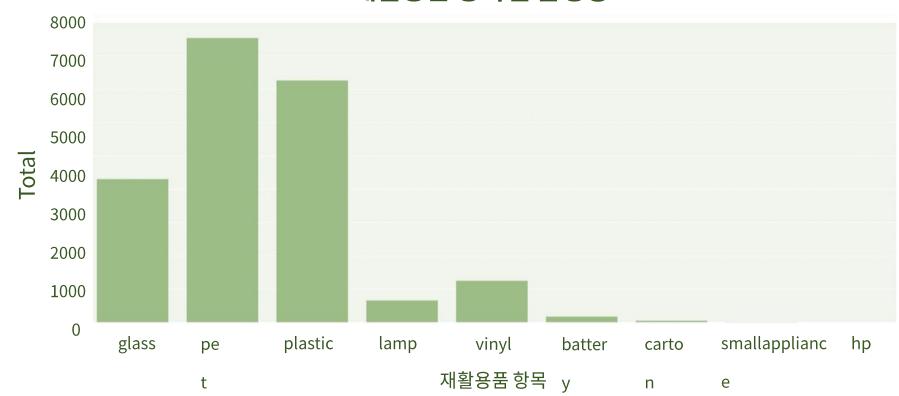
smallappliance, hp)

sum_recycle : 배출 kg수 총량

최종 17개 선정

- glass, pet, plastic 가장 많이 배출
- 세 가지 주요 배출항목을 고려한 입지 선정





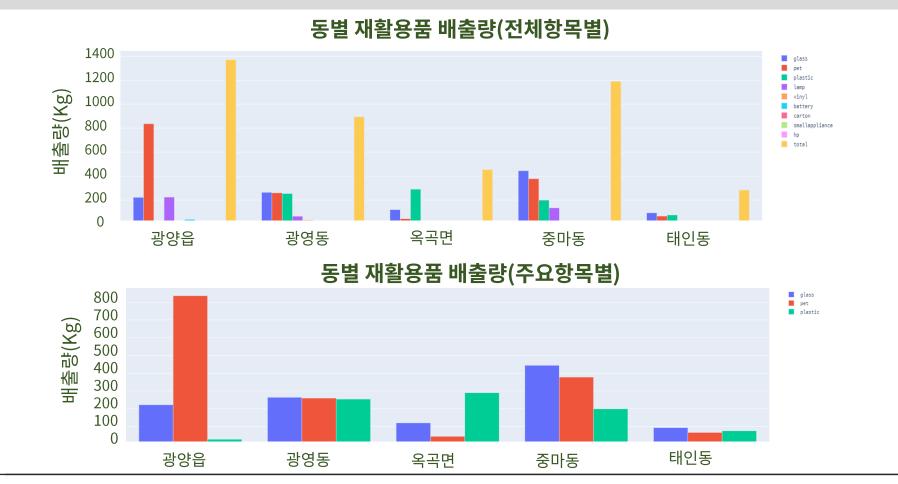
주요 탐색 대상

품목별 폐기물 발생량

사용 데이터

2.광양시_재활용품분리 수거_장려금지급내역.csv

- 데이터 특성상 데이터가 크게 튀는 금호동, 봉강면은 제외
- 장려금 데이터(원)를 무게(kg)으로 환산하여 시각화



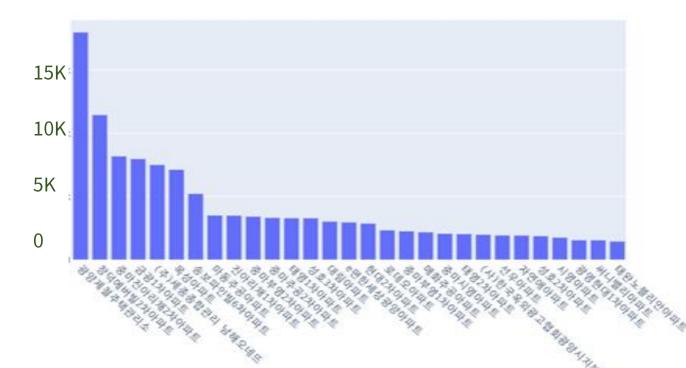
주요 탐색 대상

동별 폐기물 발생량

사용 데이터

2.광양시_재활용품분리 수거_장려금지급내역.csv • 대체로 고른 분포를 보여 아파트별로 회귀모델을 적용하는데에 적합하다고 판단

아파트별 재활용품 배출량(총합)



주요 탐색 대상

아파트별 폐기물 발생량

사용 데이터

2.광양시_재활용품분리 수거_장려금지급내역.csv

- 금호동 내 '사랑아파트'에서 가장 많은 폐기물 발생량 확인
- 사랑아파트에 재활용품 수거기 추가 설치 검토

금호동 아파트별 폐기물발생량(항목별)



주요 탐색 대상

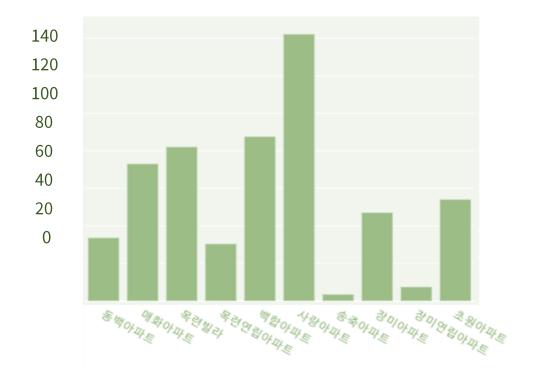
금호동 폐기물 발생량

사용 데이터

3.광양시_금호동_집적 장별_재활용품통계.csv

- 금호동 내 '사랑아파트'에서 가장 많은 폐기물 발생량 확인
- 사랑아파트에 재활용품 수거기 추가 설치 검토

금호동 아파트별 폐기물발생량(총량)



주요 탐색 대상

금호동 폐기물 발생량

사용 데이터

3.광양시_금호동_집적 장별_재활용품통계.csv

2. 단위 변환 및 데이터 병합

3. 인구수 추정

4. 재활용 품목 그룹화

- 데이터 탐색과정 이후, 아래 3개의 데이터를 분석 데이터로 선정 후 하나의 DataFrame으로 통합, 최적입지 선정에 활용

2.광양시_재활용품분리수거_ 장려금지급내역.csv

- 3년 간의 광양시 재활용품 분리수거
 장려금 지급내역 데이터
- 아파트별/품목별 재활용품 배출량 정보 활용

4.광양시_건축물_표제부.csv

- 21년 6월 광양시 건축물 표제부 데이터
- 건물의 좌표(주소의 center값), 용도 구분 및 세대수 정보 활용

11.광양시_동별인구수_세대수.csv

- 광양시 동별인구수/세대수 데이터.
- 세대당 인구 데이터를 활용하여 아파트별인구 추산

| 분석 시 고려사항

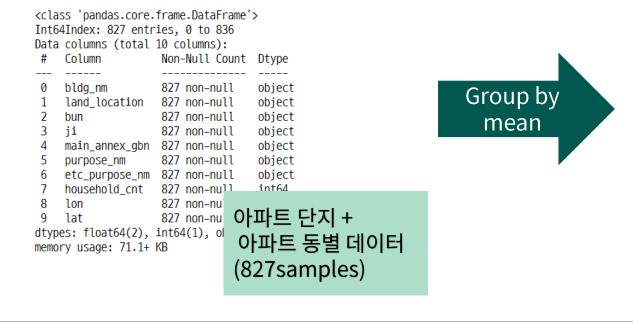
- 재활용품 자동수거기 설치 대상지역은 <mark>공동주택 단지 내에 설치되어야 하며</mark> 주민들의 쓰레기 배출동선 등을 고려하여 접근성 및 편리성이 확보되어야 합니다.
 - 해당 조건에 따라, 설치 지역은 "아파트 단지" 내에 설치되어야 한다고 정의하였음
 - 따라서 표제부에서 건물용도는 "주건축물", 주용도는 "공동주택"으로 필터링하였고, 세부용도는 모델 학습데이터 크기 확보를 위해 "아파트", "빌라", "다세대주택", "오피스텔" 을 필터링 하여 사용

2. 단위 변환 및 데이터 병합

3. 인구수 추정

4. 재활용 품목 그룹화

- 표제부 데이터 그룹화 요건 확인
 - 1. 배출량 데이터가 아파트 동별로 나와 있지 않고 아파트 단지별로 하나의 장려금 수령 주체만 존재
 - 2. 동일 단지에 거주하는 경우 아파트 주민들의 쓰레기 배출 동선이 "단지 "까지는 확장될 수 있다고 보았음
 - 아파트의 위경도 데이터가 단지별로 동일하였음. 아파트 단지의 center 값을 모든 동에 일괄 적용하였으므로,
 아파트 동별로 나와 있는 정보를 단지별로 합치는 데 무리가 없다고 판단하였음
- 위 분석 요건에 따라, 아파트 단지별로 세대수를 합하여 그룹화 하였음



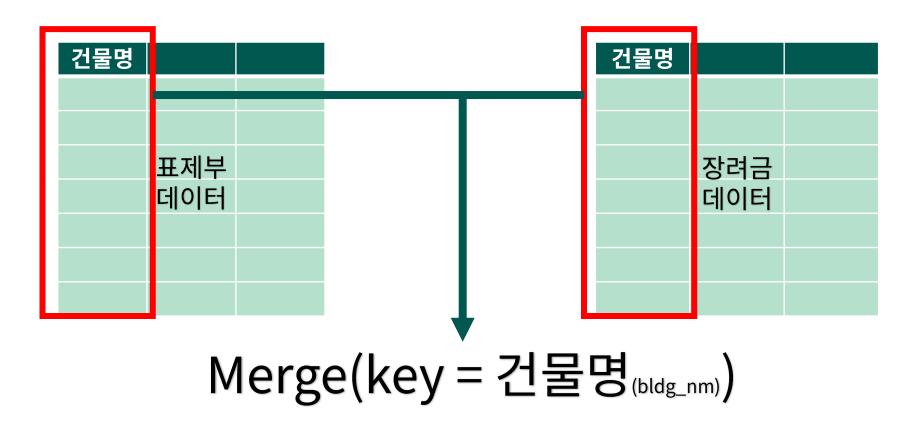
Int6	ss 'pandas.core. 4Index: 196 entr columns (total Column	10 columns):	
0 1 2 3 4 5	bldg_nm land_location bun ji main_annex_gbn	196 non-null 196 non-null 196 non-null 196 non-null 196 non-null 196 non-null	object object object object object object object
6		196 non-null	object float64
	lon lat household_cnt des: float64(2), ary usage: 16.8+	int64(1), KB CHC	아트 단지별

2. 단위 변환 및 데이터 병합

3. 인구수 추정

4. 재활용 품목 그룹화

- '2. 장려금 데이터', '4. 표제부 데이터'의 건물명을 대조하여 동일하게 통일
- 건물명을 기준으로 두 데이터를 병합



2. 단위 변환 및 데이터 병합

3. 인구수 추정

4. 재활용 품목 그룹화



관리자 2021-08-02 10:20:49

장려금 지급내역의 단위는 유리병(kg당 70원), 페트병(kg당 40원), 플라스틱류(kg당 30원), 폐형광등(kg당 150원), 필름류포장재비닐(kg당 200원), 폐건전지류(kg당 600원), 종이팩(kg당 200원), 소형가전(kg당 300원), 휴대폰(kg당 1000원)이며 금호동 재활용품 배출 통계의 단위는 캔(kg), 플라스틱(kg), 병(kg), 폐지(톤), 고철(kg) 입니다.

- 위 Q&A 답변내용을 바탕으로, 배출항목별 "장려금 교부액" 으로 되어 있던 데이터를 "재활용품 배출량(kg)"으로 변경해 줌
- 재활용품 배출 품목별로 단위당 장려금 금액이 다르기 때문에, 장려금을 기준으로 하면 배출항목별로 가중치가 들어가는 것이므로, 배출되는 "양"을 기준으로 삼기 위해 단위를 환산하였음

2. 단위 변환 및 데이터 병합

3. 인구수 추정

4. 재활용 품목 그룹화

- 분석 단위를 "아파트 단지"로 잡았으므로, 아파트 별로 데이터를 맵핑해야 함
- 아파트 단지별 인구 정보가 없으므로, 인구 추정을 다음의 두 가지 방법으로 시도하였음

*방법 1	건물 연면적 데이터 및 동별 인구수를 활용하여 동별 인구 1인당 주거건물 연면적을 구하고, 이를 통해서 아파트 단지별 인구 추산 A 아파트 거주인구 = A아파트 단지 연면적 / 행정동별 1인당 주거건물 연면적 문제 1. 연면적 데이터가 모든 거주용 건물들의 정보를 포함하고 있지 않음 문제 2. 표제부에 있는 데이터와 연면적 데이터에 존재하는 데이터가 다름	기각
방법 2	표제부 데이터의 세대수와 동별인구수_세대수 데이터의 동별 세대당 인구를 활용하여 인구 추산 A 아파트 거주인구 = A 아파트 세대수 * 행정동별 세대당 인구	<mark>채택</mark>

- 방법 1의 경우, 위 표에 기재되어 있는 것과 같이 문제가 발생하여 사용할 수 없었음
- 따라서 최종적으로 동별 세대당인구수 데이터를 활용하여 인구 추정을 진행함

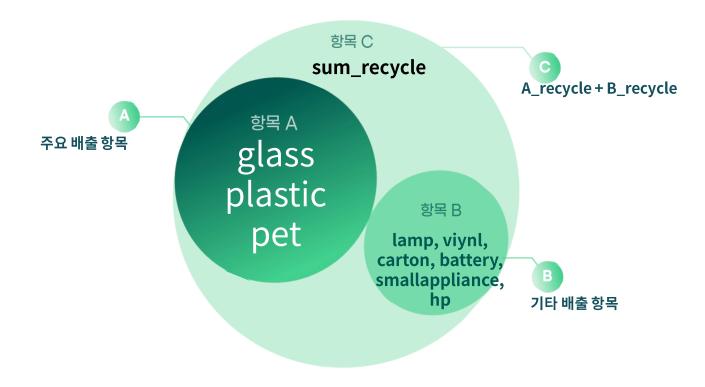
*compas 에 있는 광양시 건물분포도 데이터 활용

2. 단위 변환 및 데이터 병합

3. 인구수 추정

4. 재활용 품목 그룹화

- 가장 많이 발생하는 품목들을 "A_recycle", 그 외 배출항목들을 "B_recycle"로 분류

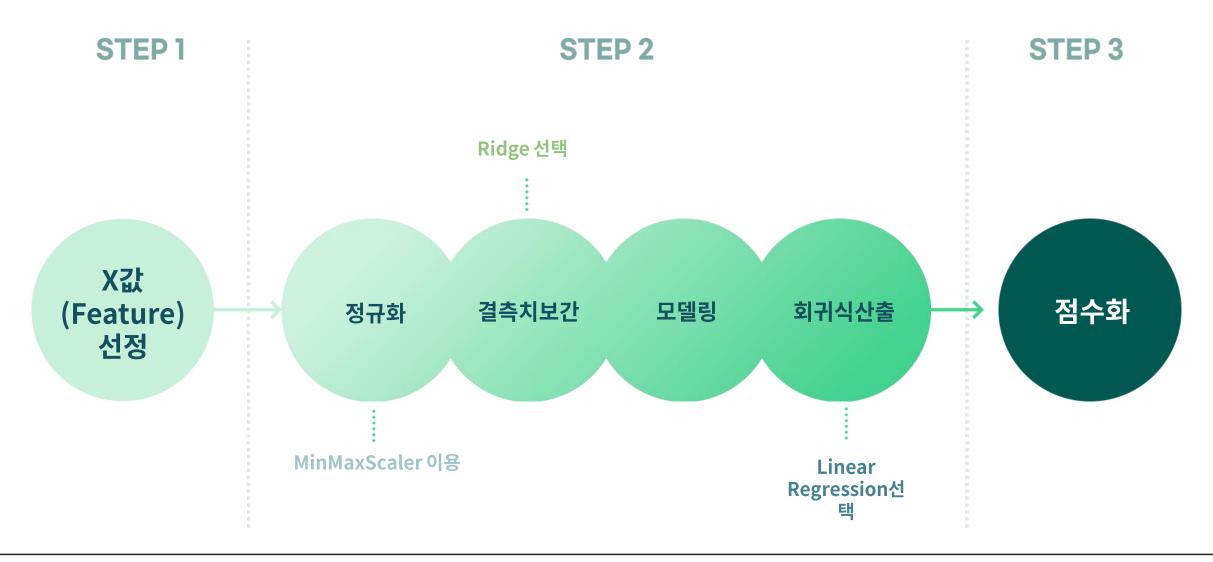


- 이유 1. 수거 효율성을 생각해 보았을 때, 재활용이 가장 활발하게 이루어지고, 대부분의 주거 단지에서 가장 많이 발생하는 재활용 종류를 그렇지 않은 것들과 구분해야 한다고 판단하였다
- 이유 2. 일부 품목들은 결측치가 과도하게 발생하므로, 그룹화 하는 것이 모델링에 유리하다고 판단하였다

02. EDA 및 전처리

- 전처리를 통해 최종 분석용 데이터 셋을 만듦

Columns명	설명
bldg_nm	건물명
land_location	주소
dong_nm	동 이름
bun	번
ji	지
main_annex_gbn	건물용도명
purpose_nm	주용도명
etc_purpose_nm	세부용도명
lon	위도
lat	경도
household_cnt	건물별 세대수
subsidy	총 장려금 교부액
A_recycle	주요 배출 항목(glass, plastic, pet)
B_recycle	기타 배출 항목 (lamp, vinyl, carton, battery, smallappliance, hp)
sum_recycle	재활용품 배출 총량(kg)
추정거주인구	추정거주인구수



'4.광양시_건축물_표제부.csv' 데이터를 활용하여, 장려금총액(subsidy), 주요품목배출량(A_recycle), 기타품목배출량(B_recycle) 3개의 값을 예측

- 예측에 필요한 X값(Feature) 선정 : 세대수(household_cnt), 추정인구수, 행정동
 - * '행정동'의 dtype이 object이므로 dummy처리 진행

	household_cnt	추정거주 인구	land_location_전라남도 광양시 광양읍	land_location_전라남도 광양시 광영동	land_location_전라남도 광양시 옥곡면	land_location_전라남도 광양시 중마동	land_location_전라남도 광양시 태인동
2	570	1442.10	0	0	0	1	0
6	132	333.96	0	0	0	1	0
10	295	666.70	0	1	0	0	0
19	270	683.10	0	0	0	1	0
29	140	316.40	0	1	0	0	0

- 예측을 위한 데이터 셋 분리

Train: 위 세 가지 특성값이 이미 존재해 예측값 도출을 위한 식을 생성해 줄 수 있는 데이터셋

Test: Target이 될 수 있는 값을 모아 놓은 데이터셋

- 더 높은 성능을 위해 MinMaxScale 적용

- 배출량 예측 시도한 모델 리스트

Ensemble: RandomForestRegressor, DecisionTreeRegressor, ExtraTreesRegressor

Linear: LinearRegression, Ridge, SGDRegressor, BayesianRidge, LogisticRegression, ElasticNet

- 최종 예측값을 도출하기위한 모델 선정 기준
- a. Overfitting되지 않아야한다
- b. Train 데이터셋에 분포되어 있는 Target값에 크게 벗어나지 않아야 한다
- c. 분산이 너무 크지 않아야 한다
- d. 이상치(마이너스값)가 다수 존재하지 않아야 한다



위의 조건들을 반영하여 회귀 예측 결과를 검토한 결과,

최종 채택 모델은 <u>"Ridge"</u>

- 모델 선정 이후, 스코어 값을 높이기위해 파라미터값 조정

A_recycle: [Ridge(solver = 'sparse_cg', alpha = 0.0001)]

B_recycle: [Ridge(solver = 'sparse_cg', alpha = 0.0001)]

subsidy: [Ridge(solver = 'sparse_cg', alpha = 0.0001)]

위와 같은 파라미터 값을 가지는 Ridge 회귀식을 사용하여 주요 배출품목 배출량(A_recycle), 기타 배출품목 배출량(B_recycle), 재활용 장려금 총액(subsidy) 예측값 도출

A_recycle	B_recycle	subsidy
0.149028	0.291266	0.158524
0.069416	0.257459	0.092285
0.246413	0.769934	0.334465
0.186072	1.000000	0.382337
0.037815	0.040179	0.014199
0.137666	0.073693	0.101309
0.116704	0.091197	0.092455
0.119738	0.096129	0.096072
0.120398	0.098628	0.097221
0.117627	0.093055	0.093646

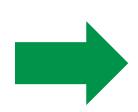
점수화에 필요한 최적회귀식 도출

X 값 : [land_location, household_cnt, 추정거주인구, B_recycle, subsidy]

Target 값 : [A_recycle]

Ridge, LinearRegression 모델을 비교해본 결과,

약 20% 높은 LinearRegression 모델 채택



```
# Weight
print("LR Weight: {}".format(model_lr.coef_))
# Bias
print("LR Bias: {}".format(model_lr.intercept_))

LR Weight: [[-3.71384403e-02 6.09567656e-02 -3.13094906e-01 1.22341251e+00 1.08072202e-03 6.88080633e-03 -6.45044614e-03 8.25544333e-04 -1.58436594e-03 -7.52260600e-04]]
LR Bias: [0.03020493]
```

- 점수화를 위해 각 Feature별 도출된 Weight값을 곱해 'Points' column 산출

	land_location	bldg_nm	lon	lat	household_cnt	추정거주인구	A_recycle	B_recycle	subsidy	Points
0	전라남도 광양시 광양읍	엘리시온	127.576751	34.979495	0.004915	0.004483	0.118047	0.093899	0.094188	0.085922
1	전라남도 광양시 광양읍	포인트빌	127.568893	34.967983	0.004915	0.004483	0.118047	0.093899	0.094188	0.085922
2	전라남도 광양시 중마동	스튜디오24원룸	127.688443	34.945190	0.004915	0.005020	0.137933	0.117623	0.123172	0.113986
3	전라남도 광양시 광양읍	엘도라도	127.577520	34.981129	0.004669	0.004260	0.117963	0.093730	0.094079	0.085838
4	전라남도 광양시 광양읍	베네스트A	127.583440	34.980059	0.004178	0.003813	0.117795	0.093393	0.093863	0.085670
181	전라남도 광양시 광양읍	시엘2차	127.590375	34.981895	0.002949	0.002696	0.137666	0.073693	0.101309	0.100925
182	전라남도 광양시 광양읍	자연애6차	127.576998	34.978288	0.000983	0.000909	0.116704	0.091197	0.092455	0.084576
183	전라남도 광양시 중마동	노블레스시티	127.697007	34.941416	0.013517	0.013620	0.119738	0.096129	0.096072	0.087766
184	전라남도 광양시 광양읍	광양스카이뷰	127.568414	34.967404	0.011797	0.010739	0.120398	0.098628	0.097221	0.088278
185	전라남도 광양시 광양읍	安愛家꿈꾸는집	127.582151	34.977269	0.003686	0.003366	0.117627	0.093055	0.093646	0.085501

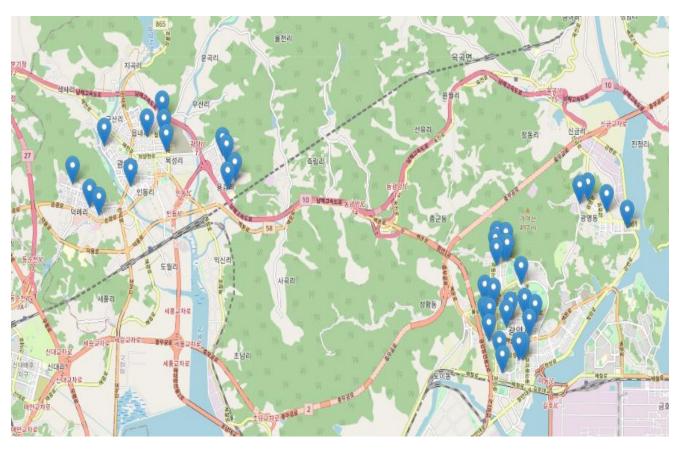
┃분석 시 고려사항

● 재활용품 자동수거기 설치 대상지역은 공동주택 단지 내에 설치되어야 하며 주민들의 쓰레기 배출동선 등을 고려하여 접근성 및 편리성이 확보되어야 합니다.

- 50개소 중 금호동을 제외한 38개소 수거기 설치

	bldg_nm	lon	lat	Points
0	창덕에버빌2차아파트	127.610219	34.973143	0.977567
1	금광1차아파트	127.690114	34.934728	0.806247
2	진아리채2차아파트	127.687123	34.938236	0.709583
3	목성아파트	127.591784	34.978266	0.614362
4	남해오네뜨	127.610419	34.967952	0.601955
5	중마2주공아파트	127.685890	34.945706	0.293818
6	마동주공아파트	127.689441	34.956343	0.265810
7	e편한세상광양아파트	127.698898	34.939212	0.259830
8	태영1차아파트	127.697087	34.934619	0.246376
9	진아리채1차아파트	127.696507	34.949411	0.239307
10	광영현대아파트	127.716059	34.965748	0.196658
11	광양송보파인빌7차아파트	127.612524	34.969653	0.185367
12	중마부영2차아파트	127.693434	34.942507	0.176692
13	중마부영1차아파트	127.700684	34.941885	0.174810
14	로데오아파트	127.713729	34.965156	0.170290
15	창덕에버빌1차아파트	127.609007	34.974636	0.163574
16	대림아파트	127.565044	34.969006	0.158766
17	광양광영사랑으로부영2차	127.727554	34.960347	0.155934
18	광양마동자연애아파트	127.688975	34.954945	0.151021

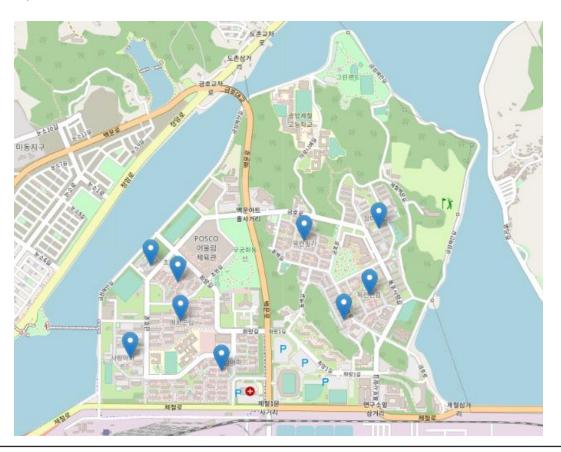
19	광양송보파인빌6차아파트	127.691960	34.956070	0.146729
20	광양송보파인빌5차아파트	127.689322	34.953125	0.146532
21	덕진광양의봄아파트	127.574215	34.976475	0.143476
22	중마시영아파트	127.685927	34.941713	0.137229
23	영신 그린빌	127.572741	34.962984	0.135722
24	선우아파트	127.692279	34.941017	0.130632
25	태영2차아파트	127.695730	34.934616	0.126603
26	성호2차아파트	127.690977	34.932170	0.126568
27	송보 파인빌 뷰아파트	127.691279	34.953878	0.126428
28	효성아파트	127.570006	34.964480	0.123858
29	광양 송보타워 아파트	127.692384	34.953971	0.118701
30	광동아파트	127.592489	34.975321	0.116622
31	신광맨션고층아파트	127.586823	34.978286	0.116622
32	동원아파트	127.591340	34.981753	0.115780
33	동원스타네옴	127.697559	34.943202	0.115689
34	스튜디오24원룸	127.688443	34.945190	0.113986
35	서울원룸	127.687275	34.941437	0.113528
36	대성빌라	127.581932	34.968605	0.113424
37	우주빌라	127.721520	34.963552	0.113376



| 분석 시 고려사항

- 재활용품 자동수거기 50개소 중 12개소는 금호동* 내 설치되어야 합니다.
 - * 금호동은 동 전체가 공동주택 단지로 조성된 지역으로 해당 지역의 특성 고려 필요
- <u>"50개소 중 12개소는 금호동내 설치되어야 합니다"</u>라는 분석 요건에 따라, 금호동 내의 아파트 단지들의 특성들을 위와 동일한 방법으로 점수화
- 금호동 내의 아파트 단지 수가 12개보다 적기에 모든 아파트 단지에 1개씩 설치하고, 점수 상위 3개의 아파트 단지에는 수거기를 1개씩 추가로 설치

		bldg_nm	lon	lat	Points	
	0	사랑아파트	127.717120	34.934343	0.095995	
	1	백합아파트	127.723044	34.933698	0.088955	
Ų	2	장미연립아파트	127.733266	34.941267	0.086332	
	3	목련연립아파트	127.732675	34.937696	0.057756	
	4	매화아파트	127.720397	34.936325	0.050301	
	5	목련빌라	127.728429	34.940589	0.049611	
	6	동백아파트	127.731015	34.936385	0.019655	
	7	초원아파트	127.720248	34.938398	0.017722	
	8	송죽아파트	127.718450	34.939268	0.000328	



분석의 기대 효과

- 본 데이터 분석을 통하여 입지 추천을 함으로써, 수거기가 효율적으로 운영될 것이라고 기대한다
 이를 통해 앞서 분석 배경에서 설명했던 인공지능 수거기의
 *기대 효과(교육 효과, 수거효율 증대, 인력 감소, 환경 보전, 경제적 효율성)을 극대화 할 수 있을 것이다
- 재활용품 배출량, 특히 주요 배출품목인 pet, plastic, glass 배출량이 높은 곳에 설치하였기 때문에 자동수거기가 광양시의 생활폐기물 처리 부담을 경감시켜 줄 것이다

제언

단지 내 모든 동을 고루 커버하기 위해, 최대한 단지의 중심에 설치하되,
 주민들의 배출 용이성 및 수거 차량의 수거 용이성, 재활용품 수거 장소에서 발생할 수 있는 냄새 등을 고려하여,
 주민들의 주요 이동 동선(출입구 근처 등)과 겹치지 않는 공터에 설치할 것을 제언한다