

---

# 광양시 재활용품 자동수거기 설치 최적위치 선정 과제

COMPAS | 골드문 팀

2021.08.31

---

# Contents



STEP  
01

## 분석 배경

재활용품 발생 현황

분석 목적

STEP  
02

## 데이터 탐색 및 전처리

EDA

데이터 통합 및 단지별 인구수 추정

STEP  
03

## 분석 과정

폐기물 발생량 예측을 이용한 최적 모델 선정

cluster별 지수화로 우선순위 산출

STEP  
04

## 분석 결과

분리수거기 최종 입지선정

결론 및 제언

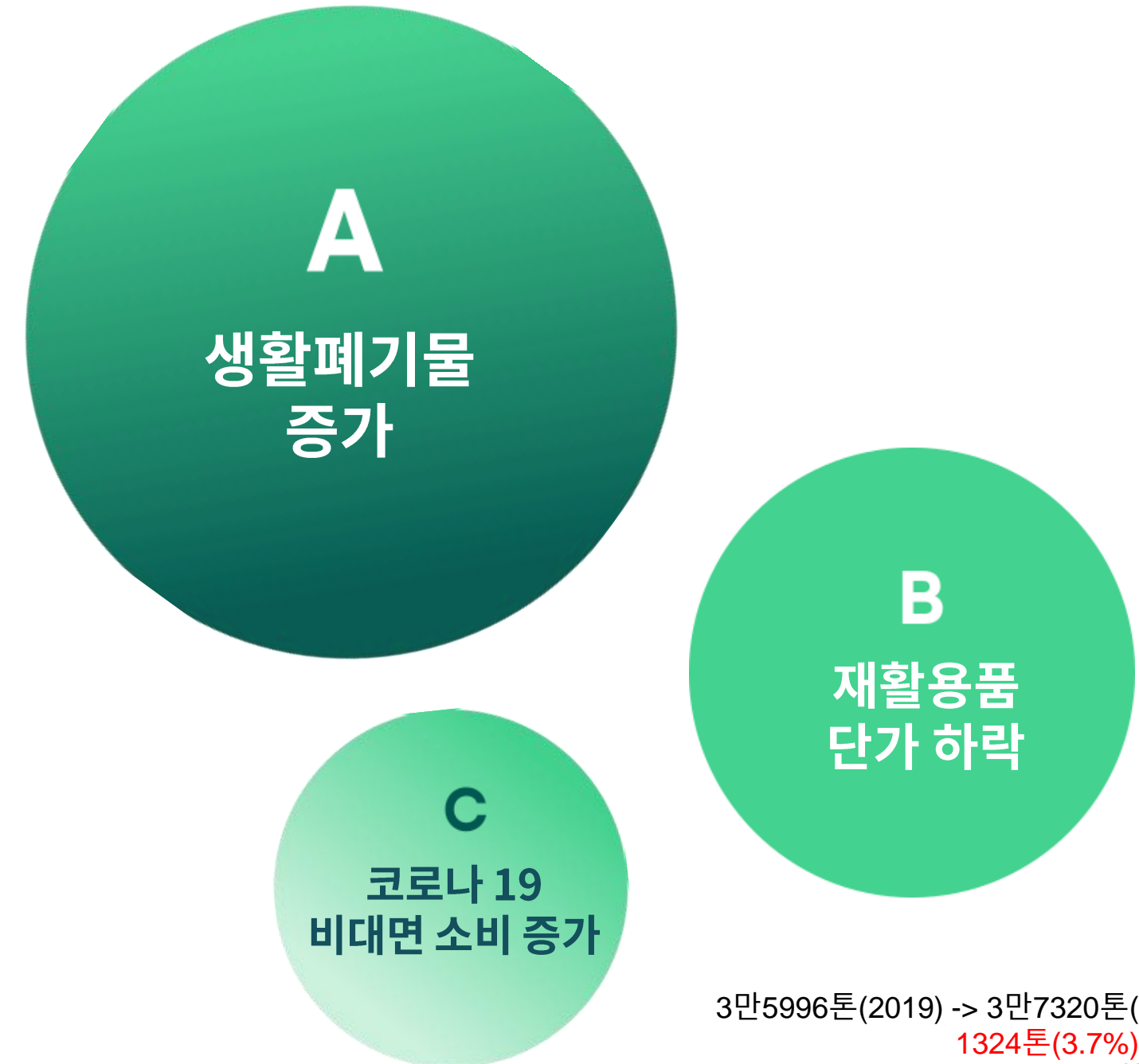
한계 및 보완점

# 1. 분석 배경

## <광양시>

광양제철소를 비롯한 순환재활용 기업이 가장 많이 밀집한 도시.

1. 코로나 19 이전: 민간수집상이 공동주택 재활용품을 수거  
코로나 19 이후: 판매단가 하락으로 인해  
민간업체 수거 거부 및 공공수거로 전환
2. 광양시는 인구 증가와 코로나 19로 인한 1회용품  
사용 증가로 생활폐기물 발생량이 증가
3. 또한 2037년 매립지 사용 종료를 앞두고  
직매립 제로화와 자원순환 선진화 정책을 수립하는 등  
재활용 극대화를 위한 계획 필요



3만5996톤(2019) -> 3만7320톤(2020)  
1324톤(3.7%) 증가.  
-2021.08.26 환경부-

생활폐기물 발생억제 감량화 + 발생된 쓰레기는 재활용 및 에너지화 추진

➡ “재활용품 수거의 효율성 증대” 필요성이 대두됨.

# 1. 분석 배경

## 자동 수거기의 기대 효과

교육 효과	시민들이 올바른 분리수거 방법을 자연스럽게 학습할 수 있음.
수거효율 증대	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.
경제적 효율성	폐기물 처리비용 감소

자동수거기를 적절한 위치에 설치하여 수거기 운영 효율을 높이는 것 중요.

따라서, 데이터 분석을 통해  
광양시 공동주택 내 재활용품 자동수거기 50개소 최적위치를 도출하고자 함.



## 2. 데이터 탐색 및 전처리

01

### 데이터 탐색

1-1) 항목별, 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

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 16개 선정

## 2. 데이터 탐색 및 전처리

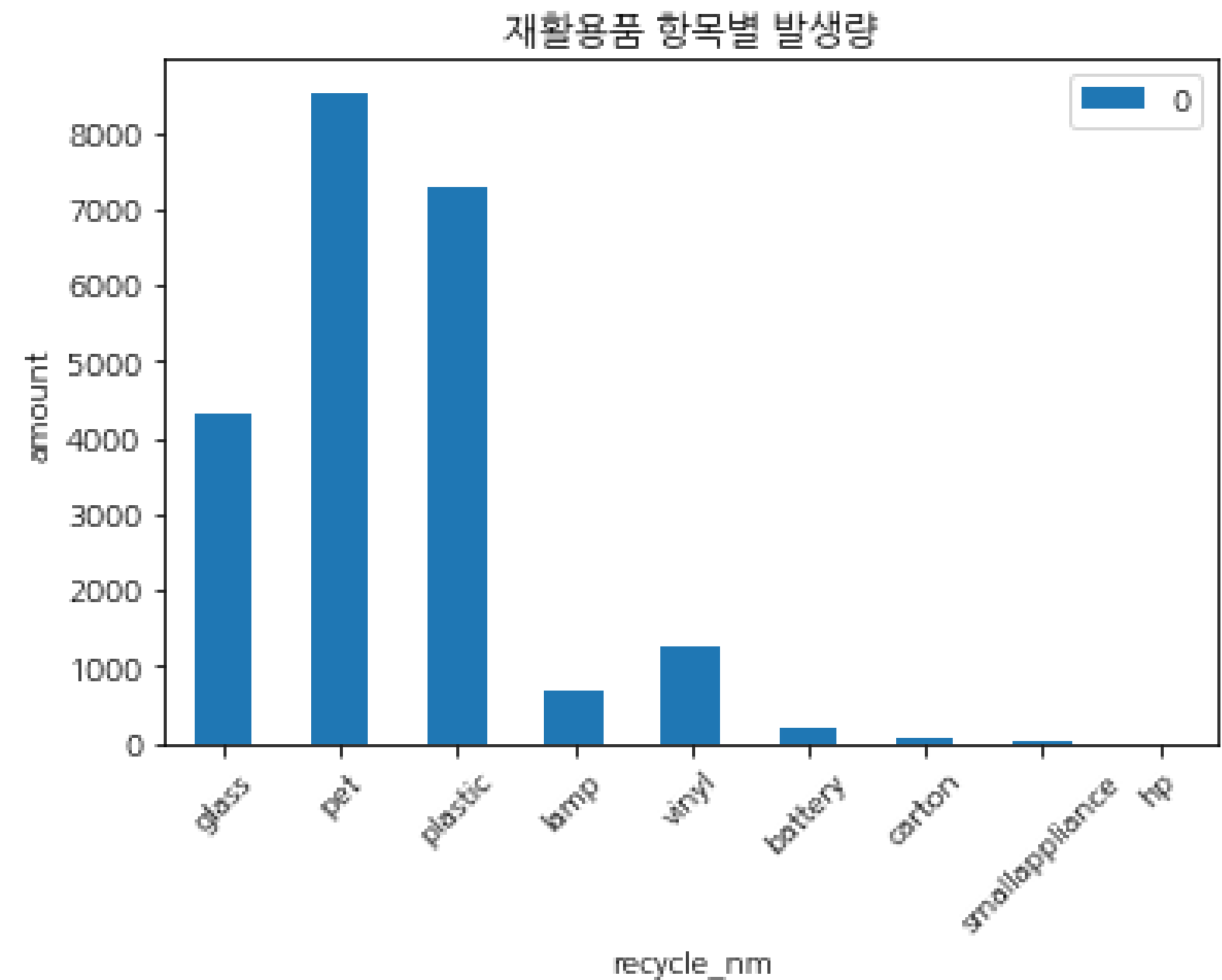
주요 탐색 대상 : 항목별 폐기물 발생량

사용 데이터 명 : 2.광양시\_재활용품분리수거\_장려금지급내역(수정).csv

탐색 결과

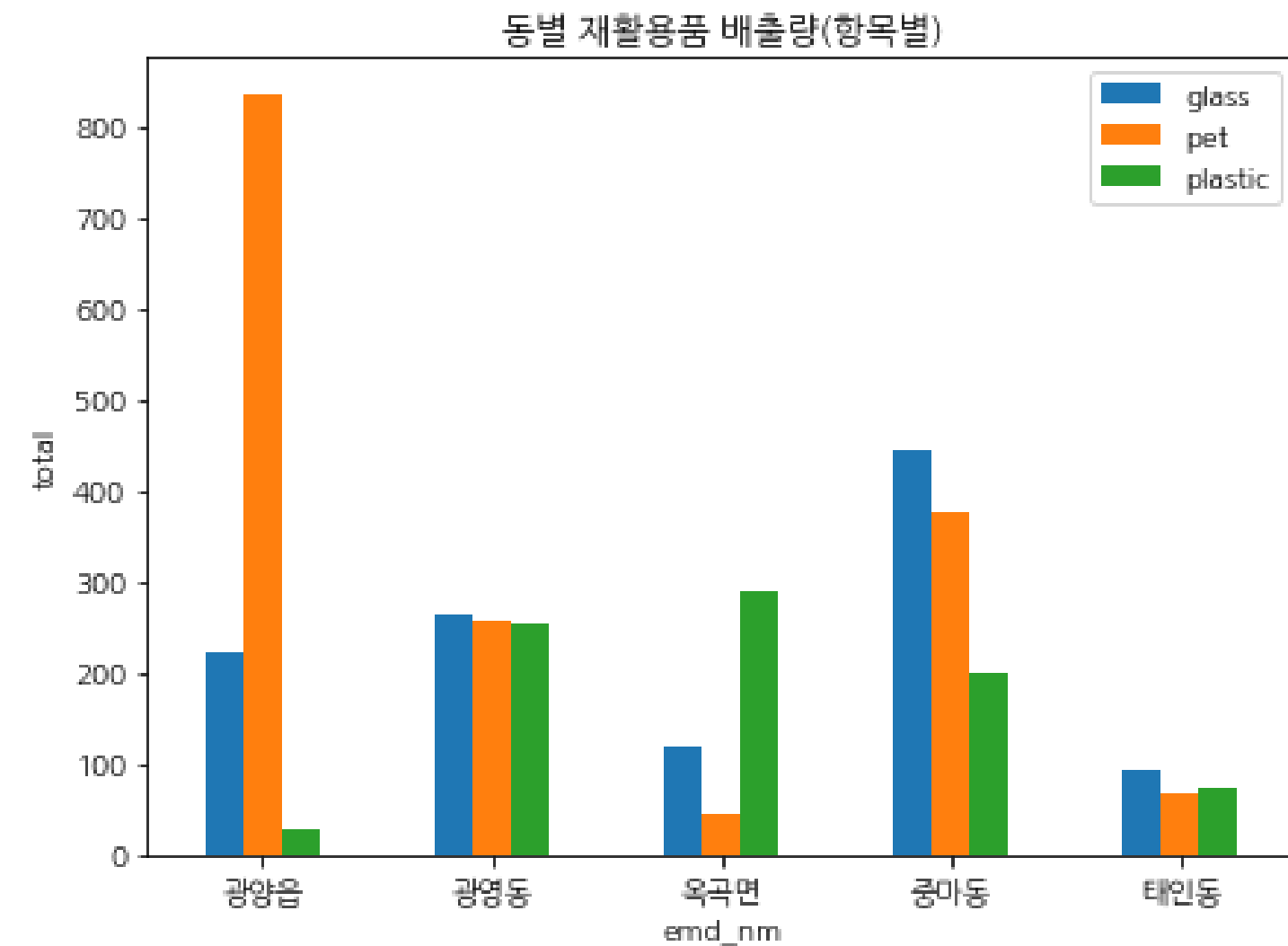
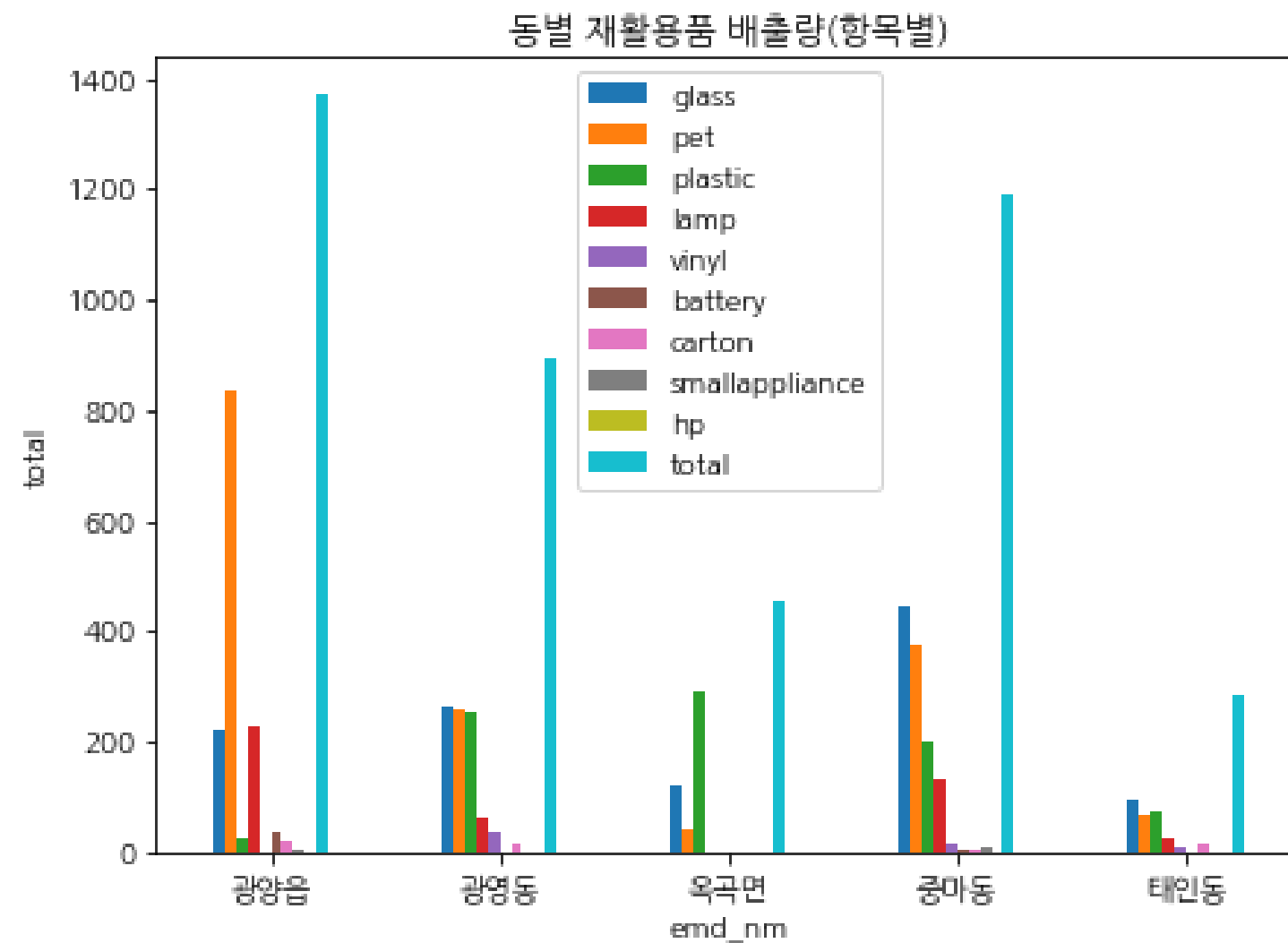
glass, pet, plastic 가장 많이 배출,  
3가지 주요 배출항목을 고려한 입지 선정.

램프, 비닐, 배터리, 폐종이류는 적게 배출,  
소형 가전이나 휴대폰은 거의 없음.



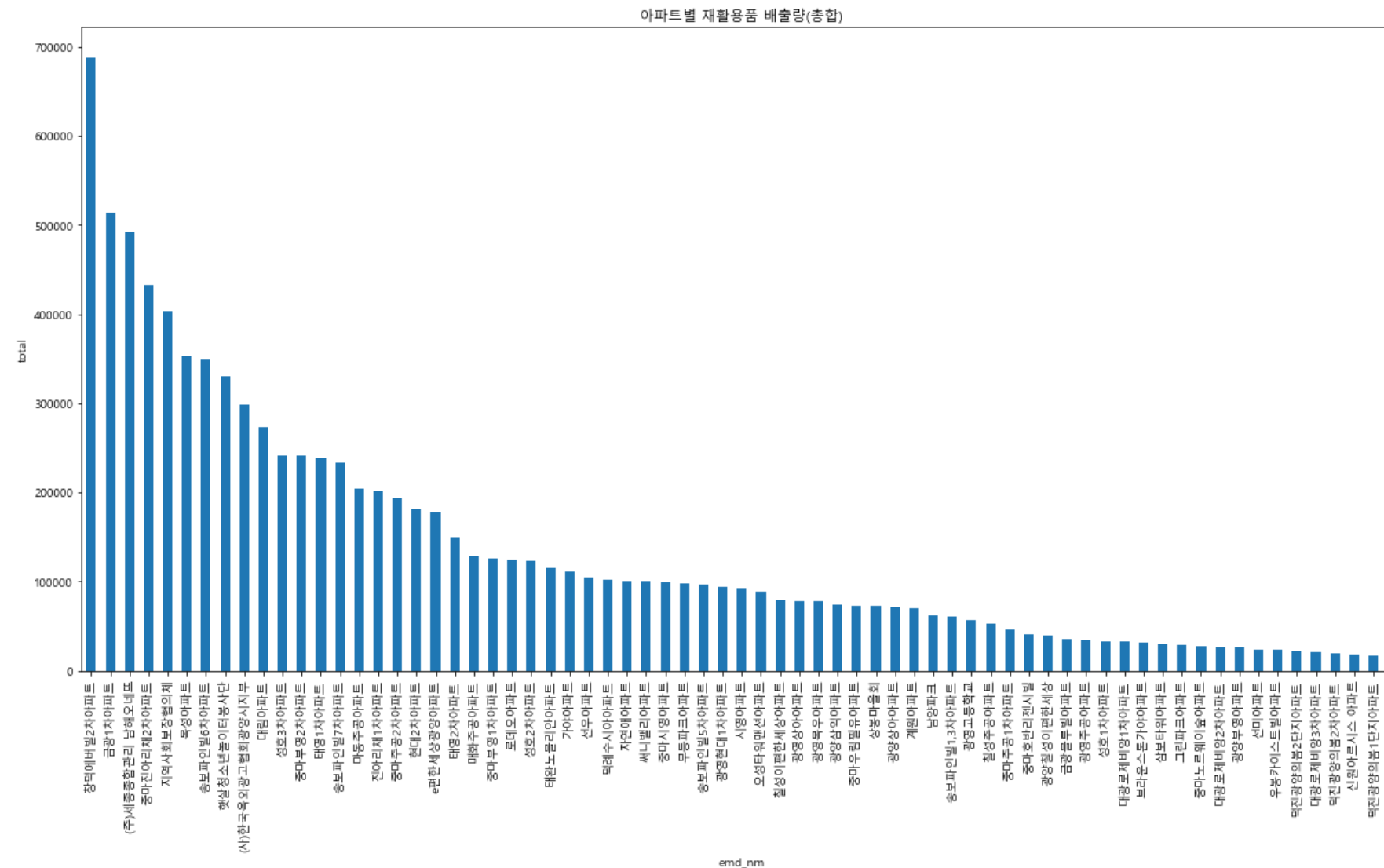
## 2. 데이터 탐색 및 전처리

- 주요 탐색 대상 : 동별 폐기물 발생 현황
- 사용 데이터 명 : 2.광양시\_재활용품분리수거\_장려금지급내역(수정).csv
- 시각화 방법: 데이터가 크게 튀는 금호동, 봉강면을 제외한 뒤, 장려금 단위를 kg으로 바꾸어 시각화.



## 2. 데이터 탐색 및 전처리

- 주요 탐색 대상 : 아파트별 배출량 확인
- 사용 데이터 명 : 2.광양시\_재활용품분리수거\_장려금지급내역(수정).csv

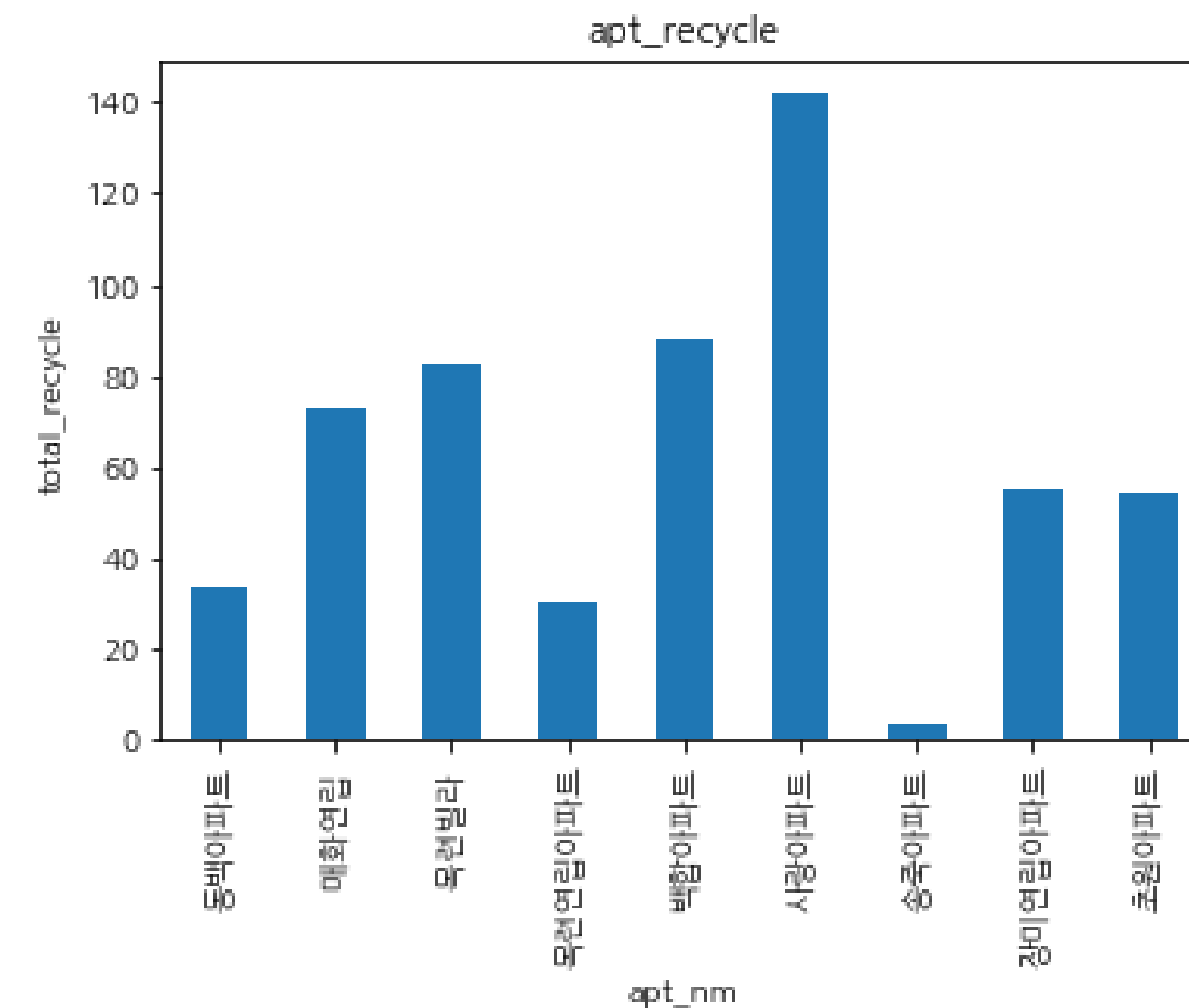
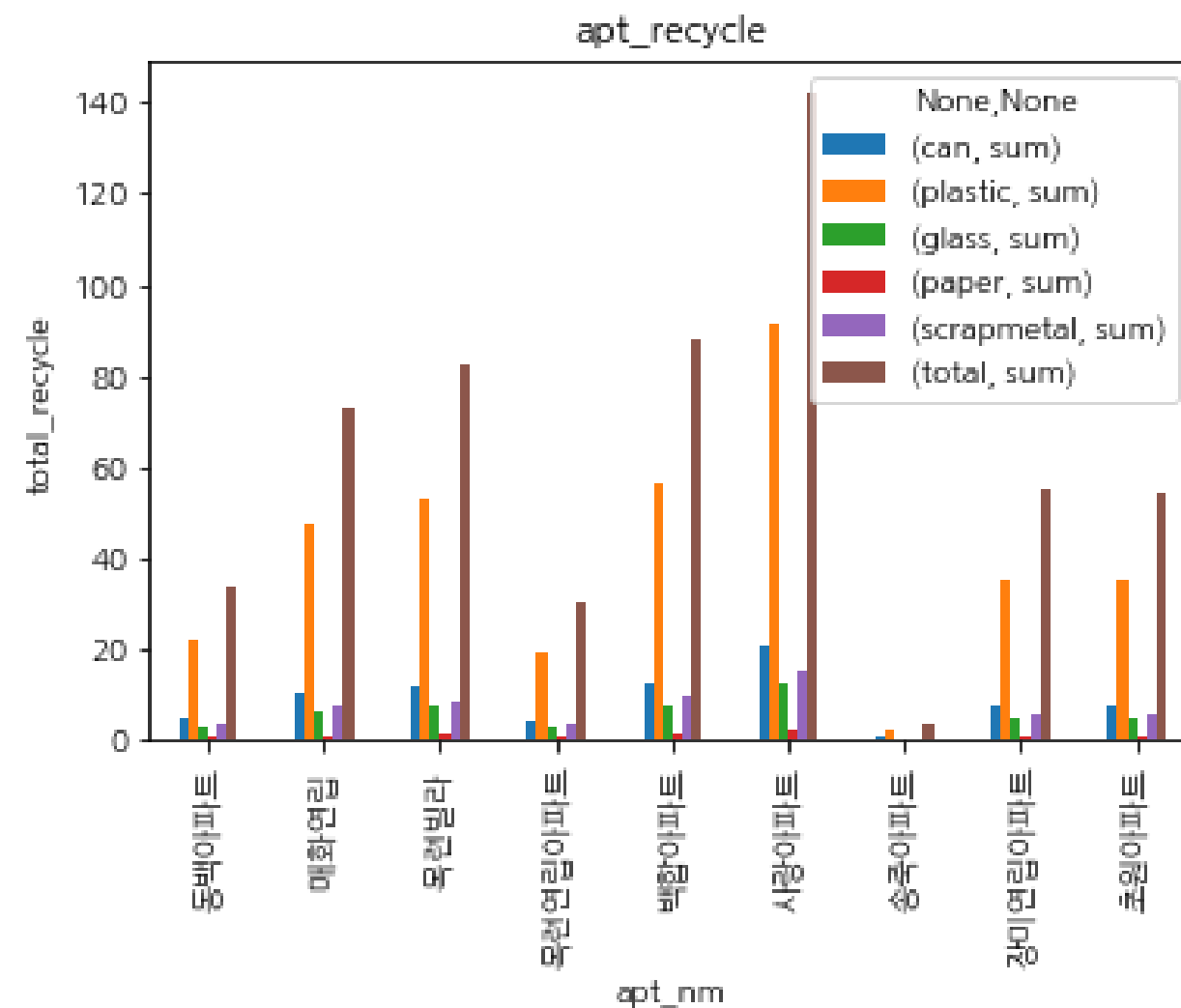


- 탐색 결과 : 대부분 고른 분포를 보임. 따라서 아파트별로 회귀모델을 적용하기에 적합하다고 판단.



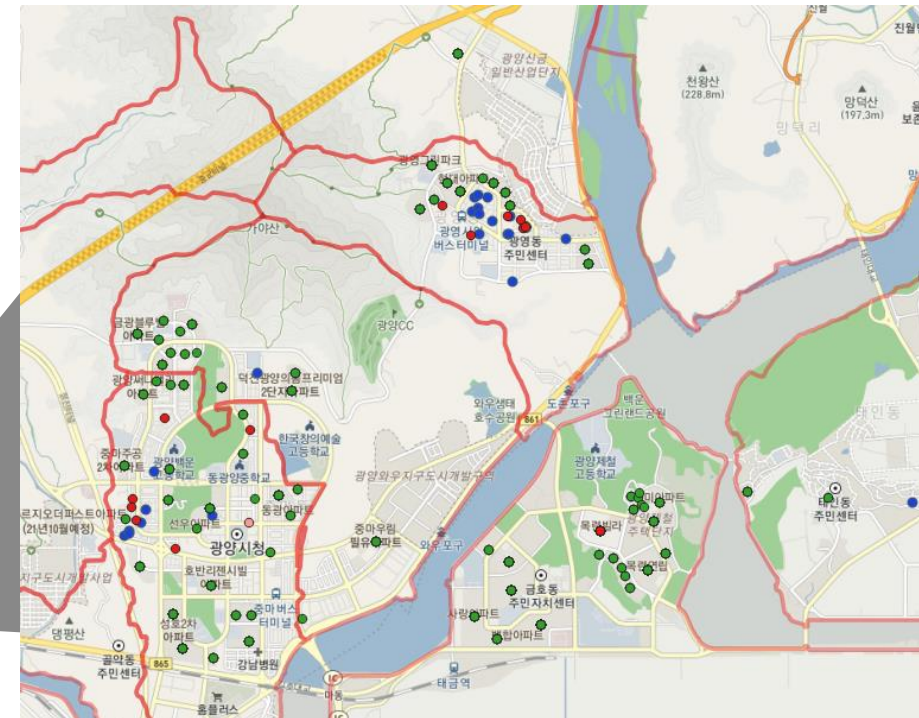
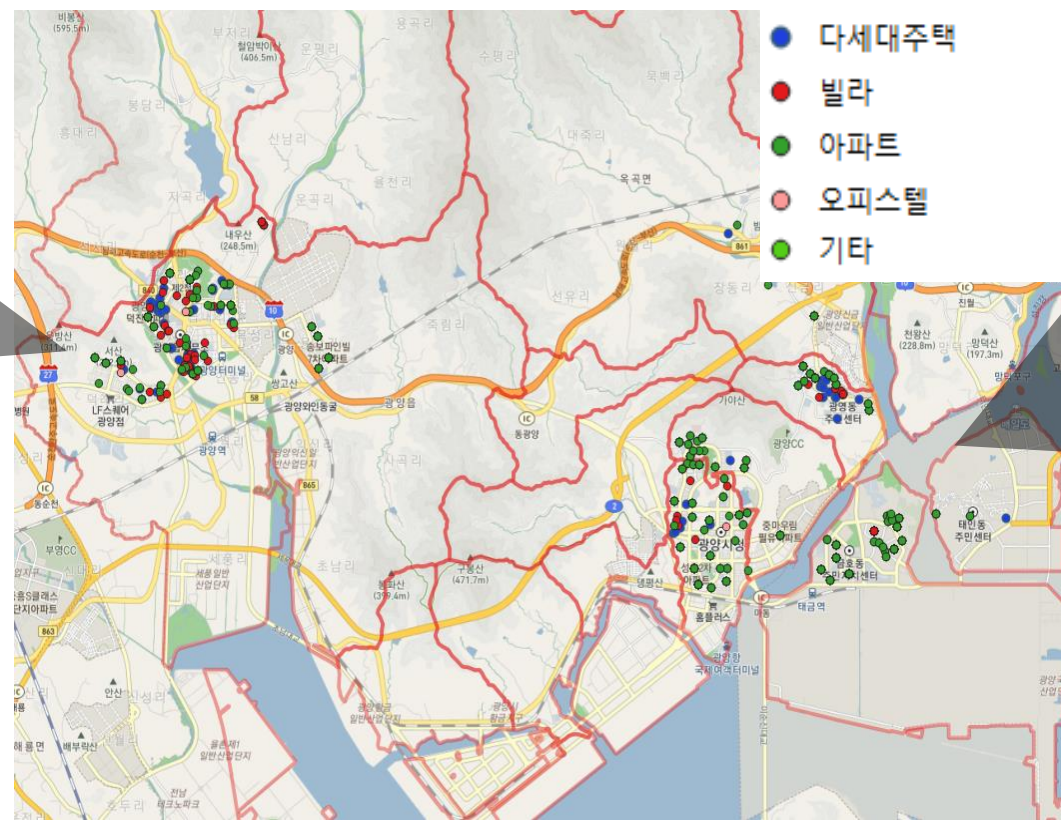
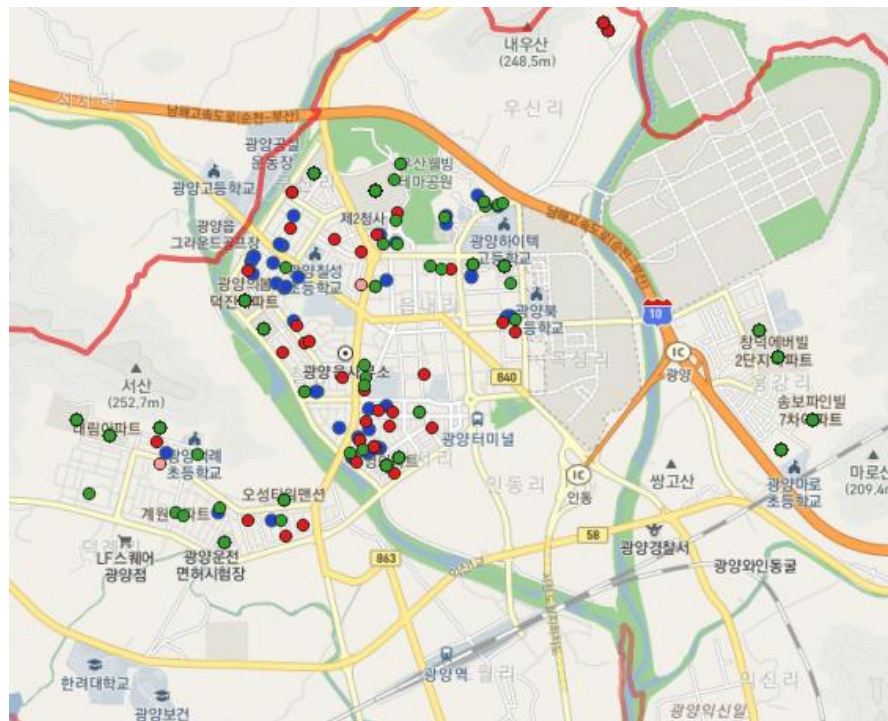
## 2. 데이터 탐색 및 전처리

- 주요 탐색 대상 : 금호동의 폐기물 발생량
- 사용 데이터 명 : 3.광양시\_금호동\_집적장별\_재활용품통계.csv
- 탐색 결과
  - > 금호동 내 '사랑아파트'에서 가장 많은 폐기물 발생량 확인.
  - > 사랑아파트에 재활용품 수거기 추가 설치 검토.



## 2. 데이터 탐색 및 전처리

- 건축물 위치 확인
- 사용데이터 : 4.광양시\_건축물\_표제부(수정).csv
  - >광양시청, 광양시청 2청사, 광영동 주변 다수의 주택단지가 분포.
  - >분리수거기 위치 선정시 주택단지 밀집지역에 가중치 주는 것 고려 가능.



## 2. 데이터 탐색 및 전처리

1. 사용 데이터  
선정 및 필터링

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 아파트별 인구 추산.

### 분석 시 고려사항

- 재활용품 자동수거기 설치 대상지역은 **공동주택 단지 내에 설치되어야 하며** 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 사용하였음.



## 2. 데이터 탐색 및 전처리

1. 사용 데이터  
선정 및 필터링

2. 단위 변환  
및 데이터 병합

3. 인구수 추정

4. 재이용 품목  
그룹화

- 표제부 데이터 그룹화 요건 확인

1. 배출량 데이터가 아파트 동별로 나와 있지 않고 아파트 단지별로 하나의 장려금 수령 주체만 존재함.
2. 동일 단지에 거주하는 경우 아파트 주민들의 쓰레기 배출 동선이 "단지" 까지는 확장될 수 있다고 보았음.
3. 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.

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 827 entries, 0 to 836
Data columns (total 10 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   bldg_nm         827 non-null    object
1   land_location   827 non-null    object
2   bun             827 non-null    object
3   ji              827 non-null    object
4   main_annex_gbn  827 non-null    object
5   purpose_nm      827 non-null    object
6   etc_purpose_nm    827 non-null    object
7   household_cnt   827 non-null    int64
8   lon             827 non-null    float64
9   lat             827 non-null    float64
dtypes: float64(2), int64(1), object(7)
memory usage: 71.1+ KB
```

아파트 단지 +  
아파트 동별 데이터  
(827samples)

Group by  
sum

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 196 entries, 0 to 195
Data columns (total 10 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   bldg_nm         196 non-null    object
1   land_location   196 non-null    object
2   bun             196 non-null    object
3   ji              196 non-null    object
4   main_annex_gbn  196 non-null    object
5   purpose_nm      196 non-null    object
6   etc_purpose_nm    196 non-null    object
7   lon             196 non-null    float64
8   lat             196 non-null    float64
9   household_cnt   196 non-null    int64
dtypes: float64(2), int64(1), object(7)
memory usage: 16.8+ KB
```

아파트 단지별  
데이터  
(196 samples)

## 2. 데이터 탐색 및 전처리

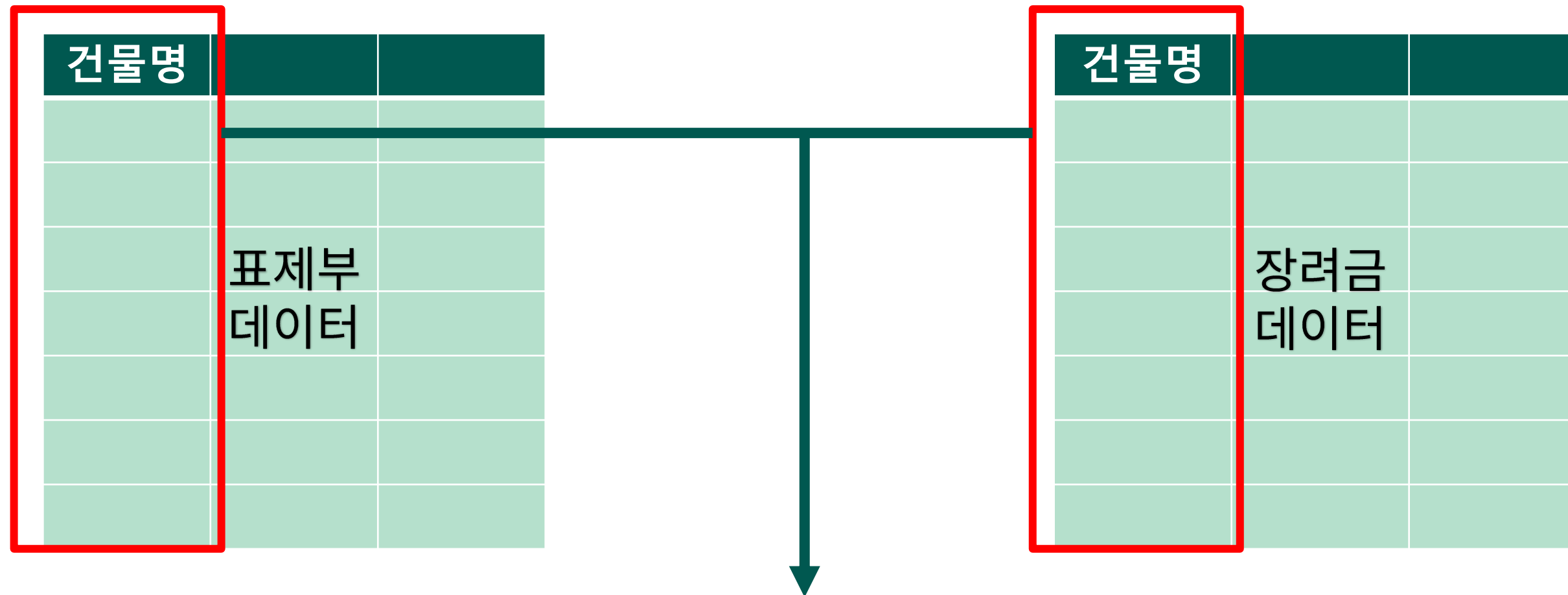
1. 사용 데이터  
선정 및 필터링

2. 단위 변환  
및 데이터 병합

3. 인구수 추정

4. 재이용 품목  
그룹화

- 2. 장려금 데이터, 4. 표제부 데이터의 건물명을 대조하여 동일하게 통일.
- 건물명을 기준으로 두 데이터를 병합함.



Merge(key = 건물명<sub>(bldg\_nm)</sub>)

## 2. 데이터 탐색 및 전처리

1. 사용 데이터  
선정 및 필터링

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. 데이터 탐색 및 전처리

1. 사용 데이터  
선정 및 필터링

2. 단위 변환  
및 데이터 병합

3. 인구수 추정

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

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

## 2. 데이터 탐색 및 전처리

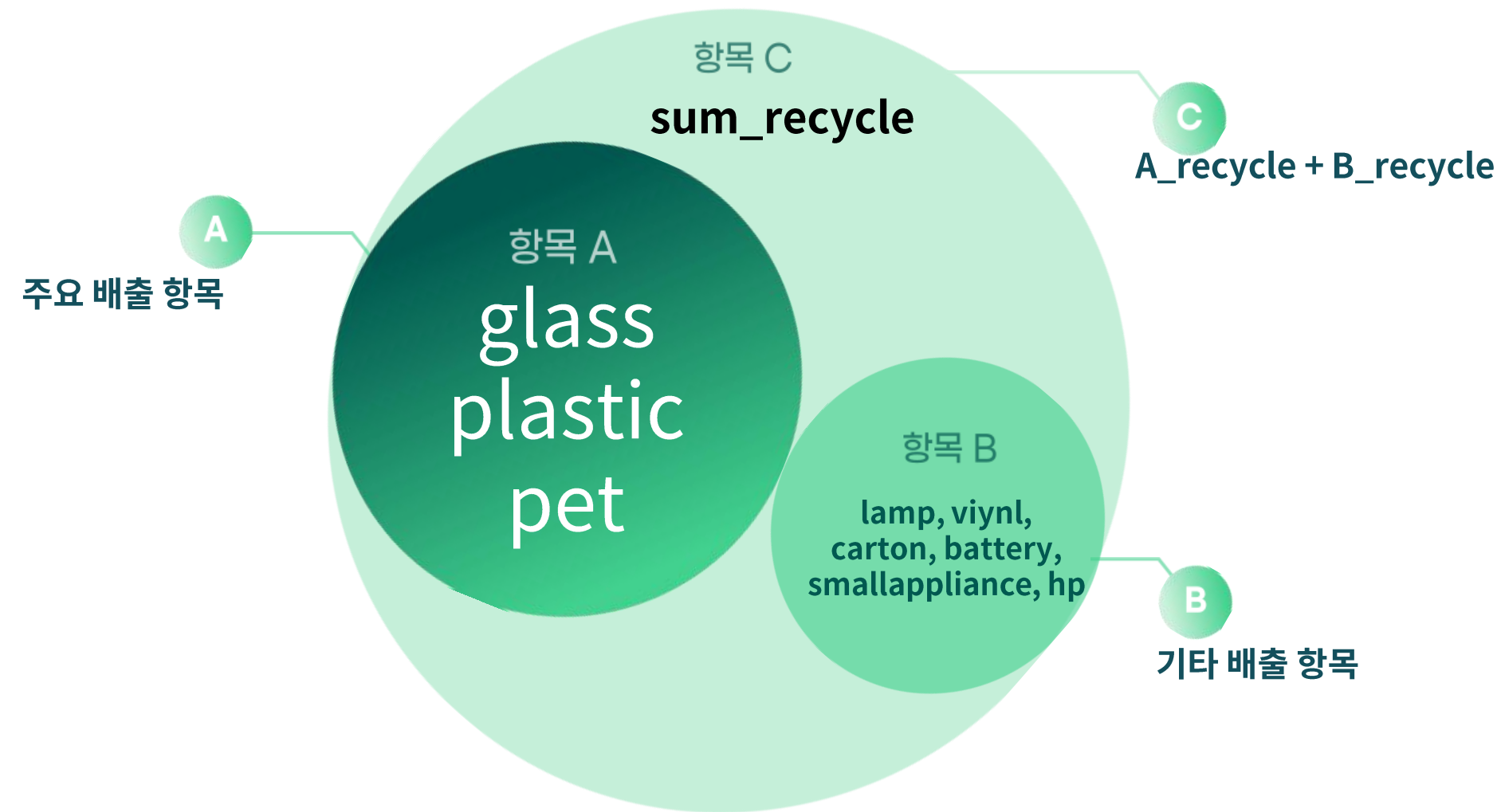
1. 사용 데이터  
선정 및 필터링

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.

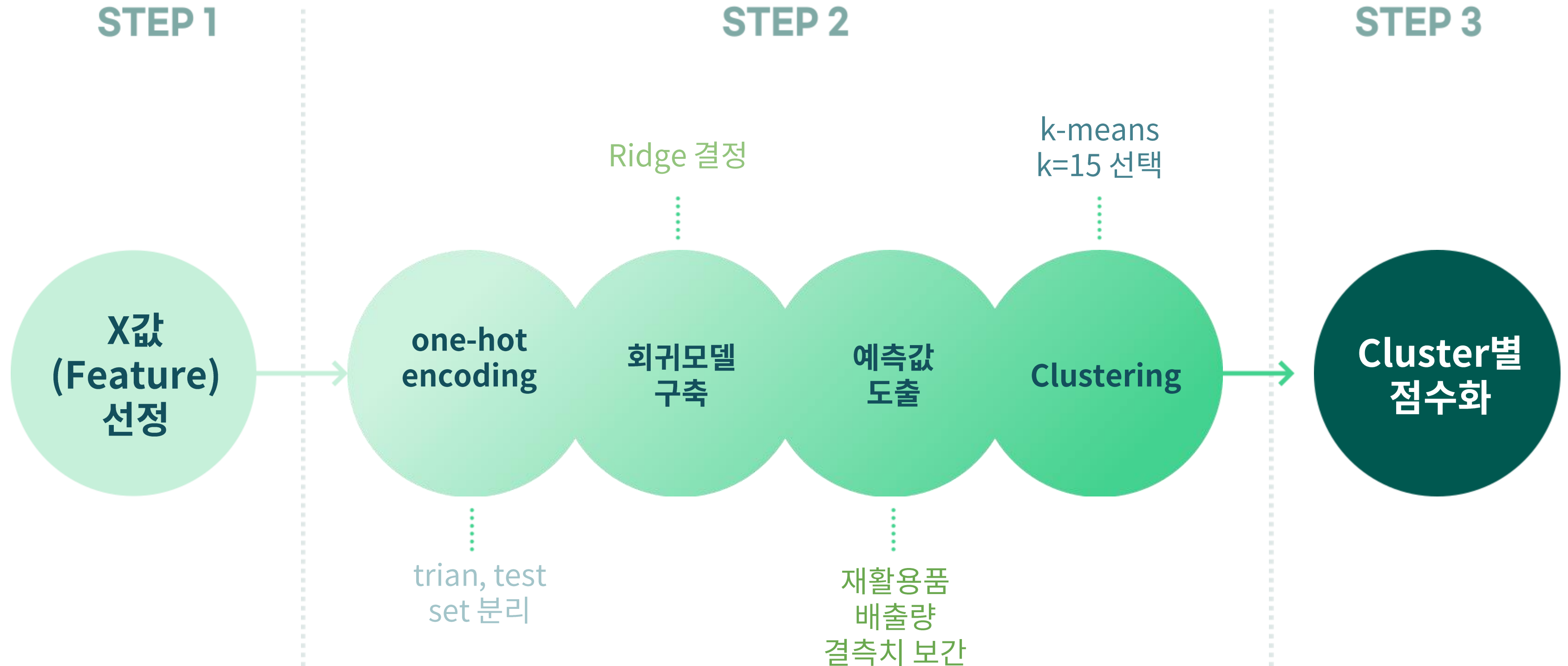


## 2. 데이터 탐색 및 전처리

- 전처리를 통해 최종 분석용 데이터 셋을 만듦.(final)

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	추정인구수
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)

### 3. 분석 과정



# 3. 분석 과정

1. Feature  
선정 및  
훈련/검증 분할

2. 회귀모델  
구축 및  
예측값 도출

3. Clustering

4. 점수화

- 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

- '4.광양시\_건축물\_표제부.csv' 데이터를 활용하여,  
장려금총액(subsidy) , 주요품목배출량(A\_recycle), 기타품목배출량(B\_recycle) 3개의 값을 예측함

-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

# 3. 분석 과정

1. Feature  
선정 및  
훈련/검증 분할

2. 회귀모델  
구축 및  
예측값 도출

3. Clustering

4. 점수화

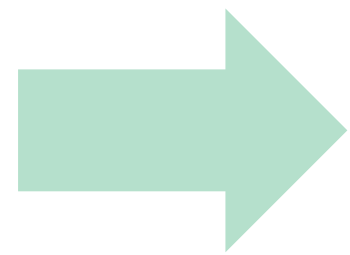
- 배출량 예측 시도한 모델 리스트

Ensemble : RandomForestRegressor, DecisionTreeRegressor, ExtraTreesRegressor

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

최종 예측값을 도출하기위한 모델 선정 기준

- a. Train 데이터셋에 분포되어 있는 Target값에 크게 벗어나지 않아야 한다.
- b. 분산이 너무 크지 않아야 한다.
- c. 이상치(마이너스값)가 다수 존재하지 않아야 한다.



위의 조건들을 반영하여 회귀 예측 결과를 검토한 결과,

최종 채택 모델은 **“Ridge”**

# 3. 분석 과정

1. Feature  
선정 및  
훈련/검증 분할

2. 회귀모델  
구축 및  
예측값 도출

3. Clustering

4. 점수화

- 모델 선정 이후, 더 실제에 가까운 예측값을 도출해내기 위해 파라미터값을 조정함.

A\_recycle : [Ridge(solver = 'svd', alpha = 1)]

B\_recycle : [Ridge(solver = 'svd', alpha = 0.1)]

subsidy: [Ridge(solver = 'svd', alpha = 0.1)]

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
30000.000000	240.000000	2.335540e+06
9561.000000	1737.000000	1.019790e+06
10443.000000	120.000000	9.642500e+05
13800.000000	1452.000000	1.912450e+06
16026.000000	216.000000	1.040500e+06
39821.000000	555.000000	2.483570e+06
280.000000	780.000000	2.091000e+05
12010.000000	150.000000	8.280300e+05
17591.000000	6299.000000	3.217270e+06
1964.000000	0.000000	1.337800e+05
21190.000000	350.000000	1.375900e+06
26390.000000	2290.000000	1.970200e+06

# 3. 분석 과정

1. Feature  
선정 및  
훈련/검증 분할

2. 회귀모델  
구축 및  
예측값 도출

3. Clustering

4. 점수화

- 이상치 보정 ) 클러스터링 전 마이너스값으로 예측된 값을 각 컬럼별 median값으로 대체하였음.

In [102]: final\_final.loc[[43]]

Out[102]:

cation	bun	ji	main_annex_gbn	purpose_nm	etc_purpose_nm	lon	lat	household_cnt	추정 거주 인구	A_recycle	B_recycle	subsidy
도 광양 다압면	"0407"	"0004"	주건축물	공동주택	다세대주택	127.737085	35.058258	33	58.08	-4020.064185	-1443.339266	-1.849272e+06

In [104]: final\_final.loc[[43]]

Out[104]:

land_location	bun	ji	main_annex_gbn	purpose_nm	etc_purpose_nm	lon	lat	household_cnt	추정 거주 인구	A_recycle	B_recycle	subsidy
전라남도 광양 시 다압면	"0407"	"0004"	주건축물	공동주택	다세대주택	127.737085	35.058258	33	58.08	11394.770791	168.812292	670965.0

# 3. 분석 과정

1. Feature  
선정 및  
훈련/검증 분할

2. 회귀모델  
구축 및  
예측값 도출

3. Clustering

4. 점수화

- 세대수, 추정거주인구, 장려금총액, 주요품목배출량, 기타품목배출량] 컬럼을 사용하여 클러스터링 진행.
- K-means 클러스터링을 진행하였고, 파라미터 값은 기본으로 적용하였음.

```
In [45]: cluster_features = final_final[['household_cnt', '추정거주인구', 'subsidy', 'A_recycle', 'B_recycle']]
cluster_features
```

Out[45]:

	household_cnt	추정거주인구	subsidy	A_recycle	B_recycle
2	570	1442.10	9.328100e+05	11904.000000	256.000000
6	132	333.96	1.818590e+06	34486.000000	30.000000
10	295	666.70	7.166400e+05	11885.000000	731.000000
19	270	683.10	2.045000e+05	3252.000000	134.000000
29	140	316.40	3.733300e+05	6766.000000	164.000000
...	...	...	...	...	...
189	9	20.70	6.650446e+05	11367.848689	160.681367
190	16	36.80	6.719524e+05	11430.666928	164.475799
191	24	55.20	6.798471e+05	11502.459202	168.812292
193	52	119.60	7.074783e+05	11753.732159	183.990019
194	19	43.70	6.749129e+05	11457.589031	166.101984

### 3. 분석 과정

1. Feature  
선정 및  
훈련/검증 분할

2. 회귀모델  
구축 및  
예측값 도출

3. Clustering

4. 점수화

	K = 11	K = 12	K = 13	K = 14	K = 15
Cluster_0	95	95	93	93	93
Cluster_1	6	5	5	5	5
Cluster_2	20	20	20	20	20
Cluster_3	2	2	2	2	5
Cluster_4	12	12	11	2	1
Cluster_5	5	5	5	9	2
Cluster_6	5	4	4	2	2
Cluster_7	20	20	20	20	6
Cluster_8	6	6	5	5	20
Cluster_9	10	10	7	4	2
Cluster_10	5	5	5	7	5
Cluster_11	-	2	2	5	5
Cluster_12	-	-	7	7	7
Cluster_13	-	-	-	5	10
Cluster_14	-	-	-	-	3

최적 군집 도출을 위하여 클러스터 개수를 늘려 가면서, 각 클러스터별로 데이터가 어떻게 묶이는지 직접 확인 후, 비슷한 특성을 가진 데이터끼리 잘 묶이면서, 다른 특성을 가진 것들과 적당히 구분되는 클러스터 개수는 15개로 결론 내림.



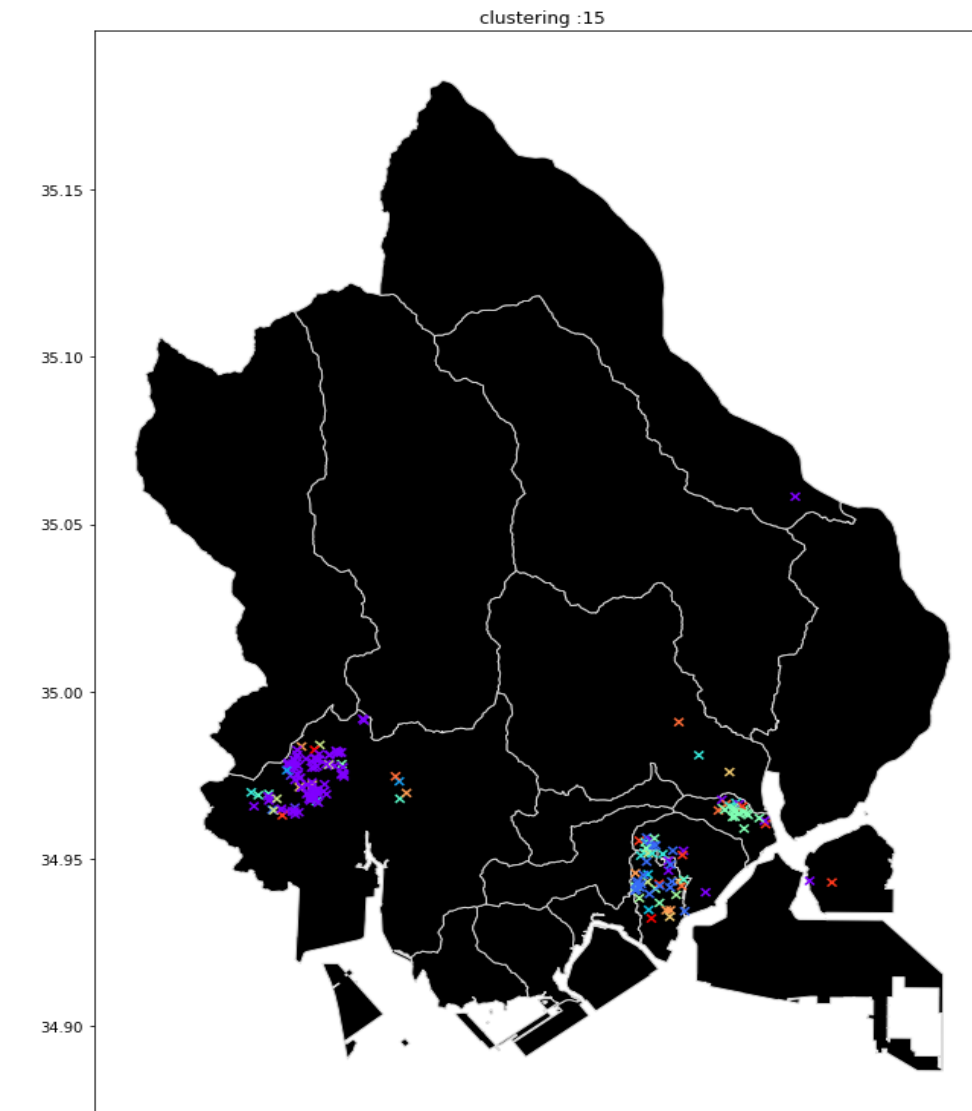
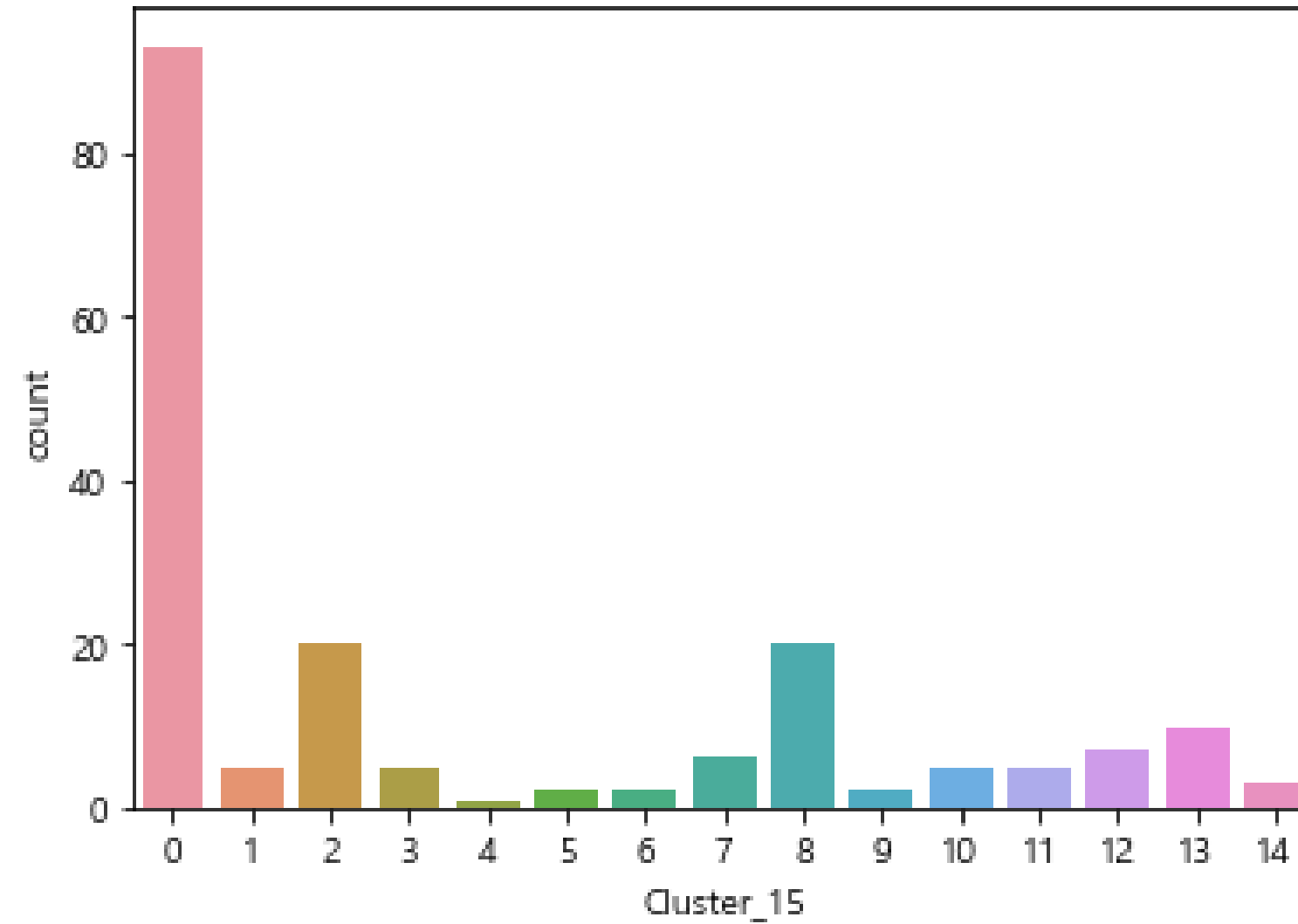
# 3. 분석 과정

1. Feature  
선정 및  
훈련/검증 분할

2. 회귀모델  
구축 및  
예측값 도출

3. Clustering

4. 점수화



최적 군집 도출을 위하여 클러스터 개수를 늘려 가면서, 각 클러스터별로 데이터가 어떻게 묶이는지 직접 확인 후, 비슷한 특성을 가진 데이터끼리 잘 묶이면서, 다른 특성을 가진 것들과 적당히 구분되는 클러스터 개수는 15개로 결론 내림.

# 3. 분석 과정

1. Feature  
선정 및  
훈련/검증 분할

2. 회귀모델  
구축 및  
예측값 도출

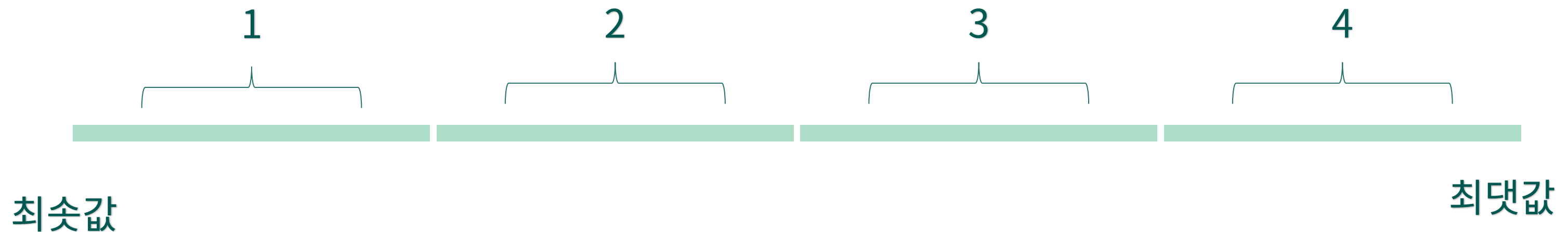
3. Clustering

4. 점수화

각 특성을 구간분할하여, 모든 아파트 단지별 특성에 점수를 부여함.

- 1.내추럴 브레이크
- 2.분위수 활용
- 3. 등간격 분할

점수화는 위와 같은 데이터 분할 방법을 고민하였는데,  
입지 선정에 있어 각 특성의 값 크기가 중요하고, 극단적으로 크거나 작은 값이 없다고 판단되어 등간격으로 값을 분할하는 방법을 사용함. 1~4의 값으로 분할하여 점수 부여.



# 3. 분석 과정

1. Feature  
선정 및  
훈련/검증 분할

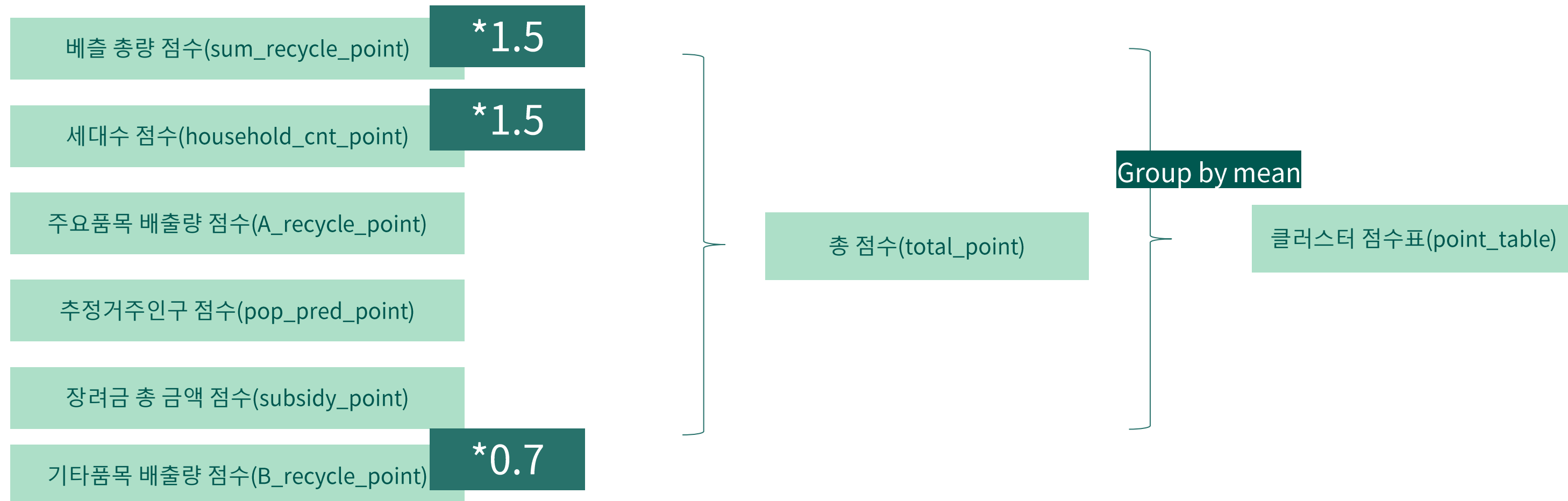
2. 회귀모델  
구축 및  
예측값 도출

3. Clustering

4. 점수화

점수화 시킨 값에 특성별로 가중치를 부여함.

- 점수화 과정에서, 배출 항목을 2개로 나누되, 재활용품 배출 총량이 우선 요인이라고 생각하여 주요항목 배출량(A\_recycle), 기타항목 배출량(B\_recycle)을 합쳐 재활용품 배출 총량(sum\_recycle) 데이터를 만듦.
- 가중치가 부여된 점수를 모두 더한 total\_point 컬럼을 만들고, 클러스터별로 평균값으로 그룹화하여 클러스터 점수 데이터 프레임을 생성.



## 4. 분석 결과

1. Feature  
선정 및  
훈련/검증 분할

2. 회귀모델  
구축 및  
예측값 도출

3. Clustering

4. 점수화

최종 입지 선정 결과

앞에서 도출한 클러스터 점수표(point\_table)에서, 클러스터별 평균점수의 총합인 total\_point를 기준으로 정렬.  
38개소를 구하기 위한 상위 클러스터들의 개체(아파트 단지)를 추출.

클러스터 분석 결과를 우선고려하되, total\_point도 함께 고려하여 주기 위하여,  
클러스터 점수를 내림차순으로 부여하고 여기에 \*3의 가중치를 주어 최종 입지 선정 결과를 도출함.

*“recommend\_point = (클러스터 점수 \* 3 + total\_point)”*

위 기준에 따라 상위 38개의 아파트 단지를 추출.

```
In [74]: recommend.index
```

```
Out[74]: Index(['성호2차아파트', 'e편한세상광양아파트', '칠성주공아파트', '무등파크맨션아파트', '금광1차아파트', '태영1차아파트',  
               '광양마동자연애아파트', '오성타워맨션아파트', '칠성e-편한세상아파트', '대림아파트', '남해오네뜨', '덕례수시아아파트',  
               '태완노블리안아파트', '태영2차아파트', '매화마을주공아파트', '중마2주공아파트', '광양송보파인빌7차아파트',  
               '창덕에버빌2차아파트', '중마부영1차아파트', '가야아파트', '창덕에버빌1차아파트', '덕진광양의봄아파트',  
               '광영상아아파트', '대광로제비앙2차아파트', '광영현대아파트', '중마노르웨이숲아파트', '브라운스톤가야아파트',  
               '대광로제비앙1차아파트', '흥한에르가', '화진그린맨션', '중마부영2차아파트', '덕진광양의봄2단지아파트',  
               '광양써니밸리아파트', '선미태인하이츠아파트', '금광블루빌아파트', '광양부영아파트', '광양광영사랑으로부영2차',  
               '영신 그린빌'],  
              dtype='object', name='bldg_nm')
```

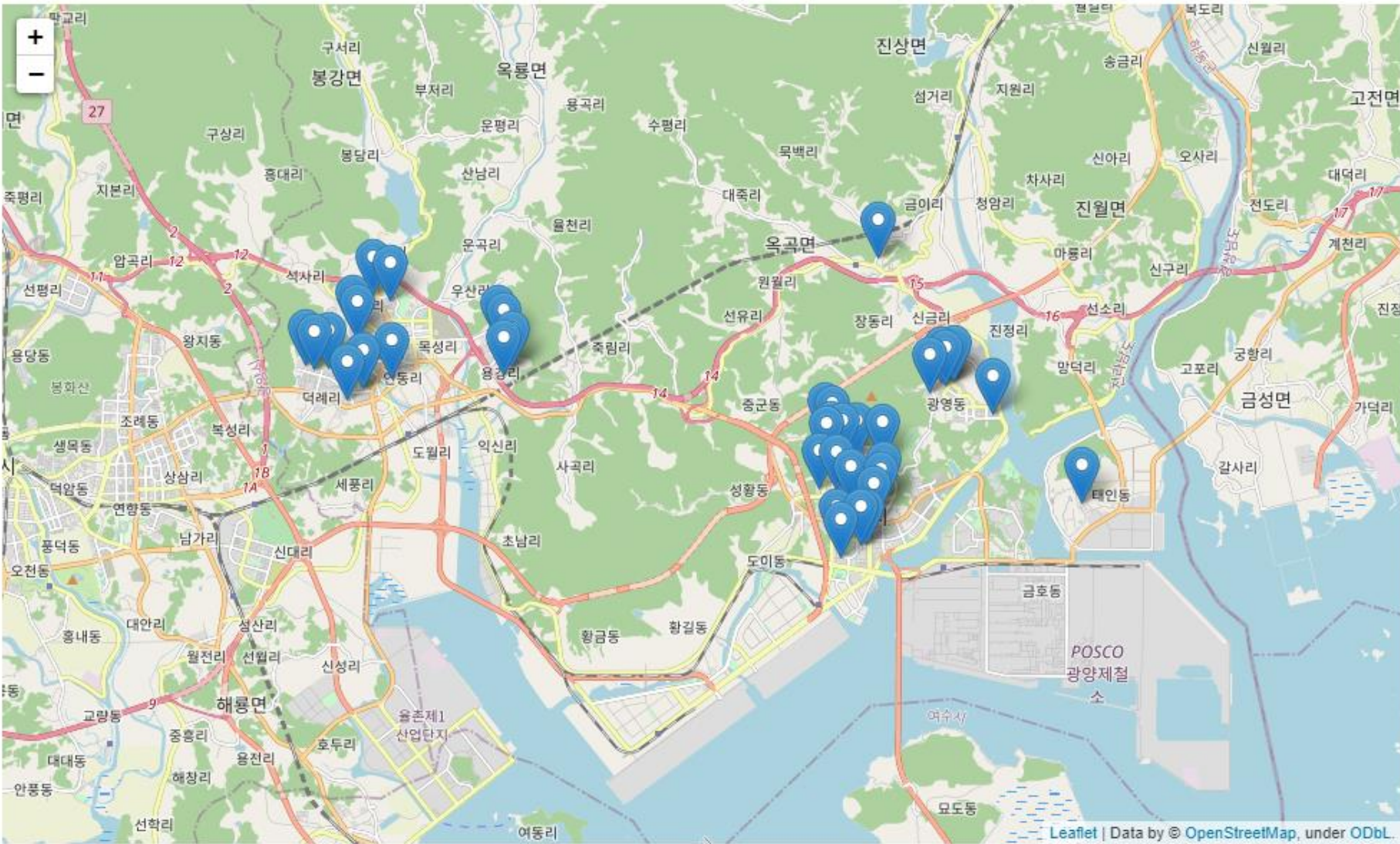


# 4. 분석 결과

- 최종 선정된 입지를 지도상에 시각화 하면, 다음과 같은 분포를 보임을 알 수 있음.

	lon	lat	Cluster_15	recommend_point
bldg_nm				
성호2차아파트	127.690977	34.932170	5	45.0
e편한세상광양아파트	127.698898	34.939212	5	42.0
철성주공아파트	127.583055	34.982528	5	43.0
무등파크맨션아파트	127.689866	34.945355	4	39.0
금광1차아파트	127.690114	34.934728	4	41.0
태영1차아파트	127.697087	34.934619	14	32.0
광양마동자연애아파트	127.688975	34.954945	14	35.0
오성타워맨션아파트	127.576791	34.965288	12	30.0
철성e-편한세상아파트	127.575439	34.974893	12	30.0
대림아파트	127.565044	34.969006	1	25.0
남해오네프	127.610419	34.967952	1	28.0
덕레수시아아파트	127.568404	34.969387	1	25.0
태완노블리안아파트	127.701356	34.943798	1	26.0
태영2차아파트	127.695730	34.934616	8	22.0
매화마을주공아파트	127.578910	34.983559	8	21.0
중마2주공아파트	127.685890	34.945706	8	23.0
광양송보파인빌7차아파트	127.612524	34.969653	8	21.0
창덕에버빌2차아파트	127.610219	34.973143	3	20.0
중마부영1차아파트	127.700684	34.941885	3	18.0
가야아파트	127.714951	34.966349	3	19.0
창덕에버빌1차아파트	127.609007	34.974636	3	18.0

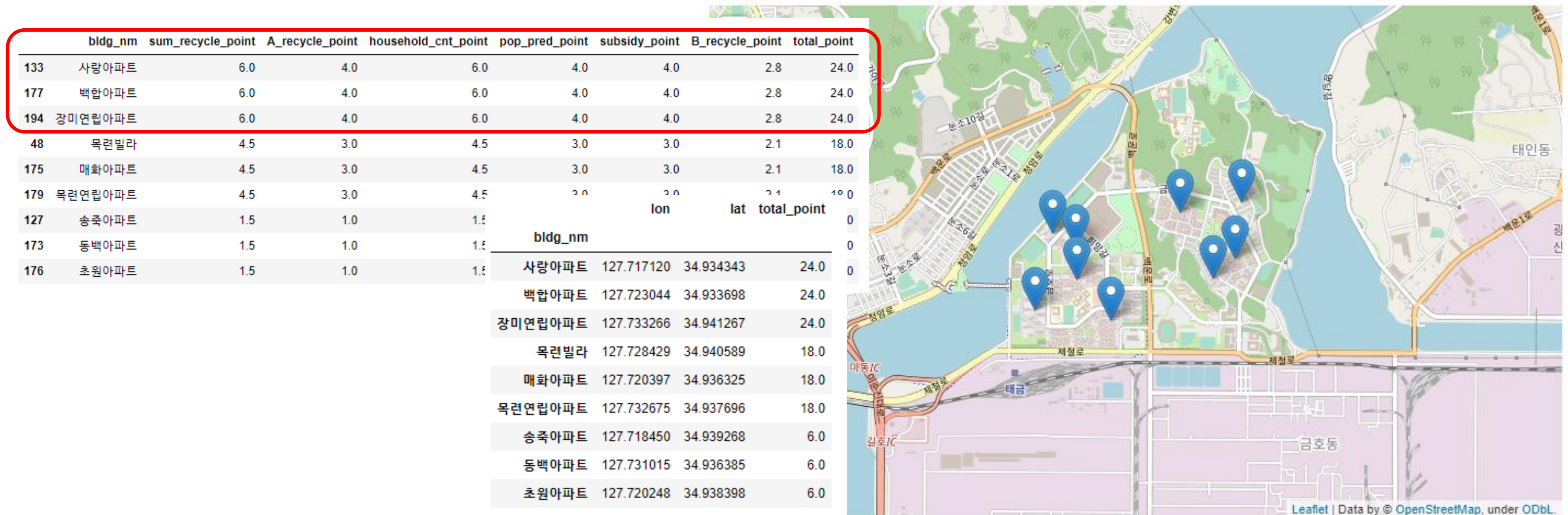
덕진광양의봄아파트	127.574215	34.976475	3	18.0
광영상아아파트	127.718082	34.966717	10	14.0
대광로제비양2차아파트	127.688735	34.951843	10	15.0
광영현대아파트	127.716059	34.965748	10	15.0
중마노르웨이숲아파트	127.694638	34.951398	10	15.0
브라운스톤가야아파트	127.712450	34.964441	10	15.0
대광로제비양1차아파트	127.691211	34.951606	10	15.0
흥한에르가	127.562805	34.969854	10	15.0
화진그린맨션	127.699772	34.990859	10	15.0
중마부영2차아파트	127.693434	34.942507	6	12.0
덕진광양의봄2단지아파트	127.700913	34.951116	6	12.0
광양씨니벨리아파트	127.687589	34.950930	6	12.0
선미태인하이츠아파트	127.748806	34.942922	6	10.0
금광블루빌아파트	127.687112	34.955400	6	12.0
광양부영아파트	127.583680	34.967214	6	12.0
광양광영사랑으로부영2차	127.727554	34.960347	6	12.0
영신 그린빌	127.572741	34.962984	6	12.0





## 4. 분석 결과

- “금호동에는 12개의 수거기를 별도 설치한다” 라는 분석 요건에 따라, 금호동 내의 아파트 단지들의 특성들을 위와 동일한 방법으로 점수화하였습니다. 금호동 내의 아파트 단지 수가 12개보다 적기에, 모든 아파트 단지에 1개씩 설치하고, 점수 상위 3개의 아파트 단지에는 수거기를 1개씩 추가로 설치하는 것을 제안합니다.



## 4. 분석 결과 – 결론 및 제언

### 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.

## 4. 분석 결과 - 한계 및 보완점

### 1) 회귀식 산정시, feature 수 부족

데이터 탐색 기간을 늘려 외부 데이터 수집을 하여 회귀모델에 반영되는 특성을 늘렸다면 좀 더 정교한 회귀예측이 가능할 것이다.

### 2) 모델의 예측 성능을 판별하는 score가 낮았던 점.

딥러닝을 시도하여 배출량 예측의 정확도를 높일 수 있을 것이다.

부족한 학습데이터를 보완하기 위한 방법을 시도해 본다면 더 나은 결과물이 나오지 않을까 기대한다.  
(oversampling 등)

### 3) 지적도나 아파트출입구, 도로 데이터 등을 활용하지 못한 점.

분석단위를 아파트 단지로 잡지 않고, 격자로 잡아 다른 지리데이터를 맵핑/또는 레이어를 활용하여 분석했다면, 더 많은 데이터를 고려할 수 있었을 것이다.



---

COMPAS | 골드문 팀

『광양시 재활용품 자동수거기 설치를 위한 최적위치 선정 과제』 마침.

---