

성연

▼ 폐의약품

폐의약품_수집_오염.pdf

• 문제점

- 어디에 지을지는 이미 거의 확정..
- 14곳 중 9곳 확정.. (나머지 5곳은 주민 의견 수렴 안됨)

거침없는 기후대응댐...후보지 조선 14곳 중 9곳 확정
정부가 지난해 7월 발표된 기후대응댐 14개 후보지 조선 9곳을 후보지로 우선 확정했다. 지역 공감대가 형성된 9곳은 후보지로 후속 절차를 추진, 타당성 등을 구체적으로 검증해 나간다는 계획이다. 아직 주민 ..

<https://www.nocutnews.co.kr/news/6306890>



- 댐 건설 후 물 부족량 82% 해소 문제 등의 환경부 검토 결과들이 있음에도, 주민들 공감대 없이 강제로 진행하지 않는다는 원칙 있음
 - 주민 설득을 위해 오히려 데이터적으로(?) 분석한 결과가 환경부 입장에서는 필요할수도..?
- 문제는 어디 짓느냐는 아니라 14곳에 왜 지어야 하는지로 가야할듯 (그러기 위해서는 댐 건설 기준이 되는 feature 가 필요)
 - 어디에 짓느냐로 가면 있을 곳은 이미 있을듯..
- 기준이 되는 feature 가 있다고 해도, 그것만으로 왜 지어야 하는지를 결정하는게 맞을지 의문
 - 이미 댐 후보지로 환경부가 지정했기 때문에, 이것만으로 지어야 하는 이유가 되나는 반박이 생길 우려는 없을수도 있을것 같음
- 필요한 features 가 많기도 하고 원하는 데이터들이 있는지 (너무 어려운 문제가 될수도..)
- 14곳에 대한 데이터가 있는지도 의문 (정보나 위치등)
- 댐 건설 기준이 있음 (<https://m.blog.naver.com/pilestory/222286190049>)
 - 댐 형식 결정의 경우, 지형, 지질 및 기초 상태, 기상 등을 확인하는 듯
 - 그러나 댐 형식은 이미 환경부에서 정했고, 왜 14곳에 지어야 하는지 보여주려면, 해당 url의 "댐 위치의 결정" 부분을 보면 되는데, 아래와 같이 딱히 정해진 기준은 없는듯.. (추상적)
 - 2.4.1 댐 위치의 결정
- (1) 일반사항
 - 댐 위치는 댐의 입지에 대한 기술적 사항과 경제성 측면뿐만 아니라 댐 개발지역의 사회, 경제, 자연환경 및 정치적인 사항까지 고려하여 결정한다.
- (2) 댐 위치 결정에 고려한 사항
 - ① 댐 위치는 댐 개발 목적과의 적합성, 지형 및 지질 등의 자연조건과 공사상의 문제점, 지역경제와의 연관성과 기득 수리권과의 조정문제, 장래개발 가능성과의 연관성, 단일댐 개발과 댐군 개발의 비교, 자연환경과의 조화와 보전, 생태계 보전을 면밀히 검토하여 결정한다.
 - ② 댐의 예상 수율지역 내에 중요한 문화적, 역사적 유적이거나 수려한 자연경관이 위치하고 있을 경우에는 댐 위치의 변경 또는 규모축소 등도 고려한다.

◦ 방향성 및 요약

- 문제점 요약
 - 지어야 하는 기준(feature) 이 명확히 없음
 - 당연한 결과
 - 14곳의 정보가 있는지 (특징이 명확하지 않을 뿐더러 있다해도 그 특징들만으로 건설 여부를 따지는게 맞는지)
- 2024 대상작 재활용 문제
 - 우수 지역의 일부 feature 를 미흡지역에 적용한 새로운 데이터 생성 하는것도 좋은 방법임
 - 우수/미흡을 만들어내는 핵심 feature 찾는게 좋은 방향 같음 (feature 중요도, 인과 추정)
- 2023 대상작 대기질 오염에 영향을 주는 feature 를 찾고, 해결책(도시숲 조성) 제시 후 개선 결과를 보여줌 (이 경우, 관련 정책과 예산도 고려해서 입지를 따로 설정한게 point 인듯)
 - 이 경우 데이터끼리 합친듯
- feature 중요도, 인과추정 주제

1. 도시 열섬

- 녹지 감소 및 도시화, 인공열, 대기 오염
- 데이터
 - 열섬 현상 기준이 되는 feature(NDVI, NDBI, LST: 모두 수식 있음)
 - NDVI: 녹색도
 - NDBI: 도시화·인공 구조물(건축물) 영역을 강조하는 지표
 - LST: 위성 지표면 온도
 - 지역별 기상 정보
 - 지역별 인구, 교통
 - 지역별 공원
- 방향성 및 기법
 - 피처 중요도 추출 후 "NDVI가 0.1 증가할 때 LST가 평균 0.7 °C 감소" 식으로 정량화
 - 공원, 녹지 조성 지역과 비교 후 효과 추정

• 열섬 데이터 (데이터 많은 것 같음, 위도, 경도)

예시) 확인 필요

```
import geopandas as gpd
import pandas as pd
import rasterio
from rasterio.mask import mask
import numpy as np
import xgboost as xgb
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error

# --- 1. 공간격자 불러오기 (1km 그리드) ---
grid = gpd.read_file('grid_1km.shp') # 필드: grid_id, geometry

# --- 2. 래스터(zonal stats) 함수 정의 ---
def zonal_stat(raster_path, grid_gdf, stat='mean'):
    with rasterio.open(raster_path) as src:
        results = []
        for geom in grid_gdf.geometry:
            out_image, _ = mask(src, [geom], crop=True)
            arr = out_image.data[0].astype(float)
            arr[arr == src.nodata] = np.nan
            if stat == 'mean':
                results.append(np.nanmean(arr))
            elif stat == 'median':
                results.append(np.nanmedian(arr))
            elif stat == 'std':
                results.append(np.nanstd(arr))
        return results

# --- 3. LST, NDVI, NDBI 통계값 추출 ---
grid['LST_mean'] = zonal_stat('modis_lst_summer2023.tif', grid)
grid['NDVI_mean'] = zonal_stat('sentinel2_ndvi_summer2023.tif', grid)
grid['NDBI_mean']= zonal_stat('sentinel2_ndbi_summer2023.tif', grid)

# --- 4. 인구밀도 & 교통량 데이터 병합 ---
pop = pd.read_csv('pop_density_grid.csv') # 열: grid_id, pop_density
traffic = pd.read_csv('traffic_count_grid.csv') # 열: grid_id, traffic_count
grid = grid.merge(pop, on='grid_id').merge(traffic, on='grid_id')

# --- 5. 공원-녹지 비율 계산 ---
parks = gpd.read_file('parks.shp') # 필드: park_id, geometry
parks = parks.to_crs(grid.crs)
```

```

overlay = gpd.overlay(grid[['grid_id','geometry']], parks, how='intersection')
park_area = overlay.geometry.area.groupby(overlay.grid_id).sum()
grid['park_area'] = grid['grid_id'].map(park_area).fillna(0)
grid['park_ratio'] = grid['park_area'] / grid.geometry.area

# --- 6. 모델링을 위한 DataFrame 준비 ---
df = pd.DataFrame({
    'LST' : grid['LST_mean'],
    'NDVI' : grid['NDVI_mean'],
    'NDBI' : grid['NDBI_mean'],
    'pop_density' : grid['pop_density'],
    'traffic_count': grid['traffic_count'],
    'park_ratio' : grid['park_ratio']
}).dropna()

X = df[['NDVI','NDBI','pop_density','traffic_count','park_ratio']]
y = df['LST']

# --- 7. XGBoost 회귀모델 학습 & 평가 ---
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
model = xgb.XGBRegressor()
model.fit(X_train, y_train)
preds = model.predict(X_test)

rmse = mean_squared_error(y_test, preds, squared=False)
importances = pd.Series(model.feature_importances_, index=X.columns)

print(f'RMSE: {rmse:.3f}')
print('Feature Importances:\n', importances)

```

파일명	데이터 내용	출처 예시
grid_1km.shp	1km×1km 정사각형 격자 폴리곤 (<code>grid_id</code> , <code>geometry</code>)	직접 생성 또는 공공데이터포털/국토교통부 공간정보 오픈마켓에서 “격자(shapefile)” 검색 → 다운로드
modis_lst_summer2023.tif	2023년 6~8월 일별 MODIS LST_Day_1km 평균 지표면 온도 GeoTIFF	Google Earth Engine → MODIS/006/MOD11A 컬렉션 필터→평균→Export to Drive → GeoTIFF
sentinel2_ndvi_summer2023.tif	2023년 여름 Sentinel-2 (B8, B4) 밴드로 계산한 NDVI 평균 GeoTIFF	GEE에서 COPERNICUS/S2 컬렉션 → normalizedDifference(["B8","B4"]) (또는 equivalent) → 기간 필터 → 평균 → Export GeoTIFF
sentinel2_ndbi_summer2023.tif	2023년 여름 Sentinel-2 (SWIR, NIR) 밴드로 계산한 NDBI 평균 GeoTIFF	GEE에서 normalizedDifference(["B11","B8"]) (또는 equivalent) → 기간 필터 → 평균 → Export GeoTIFF
pop_density_grid.csv	각 격자(<code>grid_id</code>)별 인구밀도 (명/km ²)	통계청 KOSIS 행정동 인구 데이터 또는 WorldPop 격자 인구 GeoTIFF → 공간조인(Spatial Join)→ CSV로 저장
traffic_count_grid.csv	각 격자(<code>grid_id</code>)별 교통량(대수/일)	한국도로공사 VDS API나 서울 TOPIS API → 관측소별 교통량 JSON → 공간조인 → 격자별 합계 → CSV
parks.shp	도시 공원 폴리곤 (<code>park_id</code> , <code>geometry</code>)	공공데이터포털 “전국도시공원정보” 다운로드 → Shapefile (위도·경도, 면적, 공원명 등 필드 포함)

- feature 중요도, 인과추정 외의 방향성 및 주제
 - 쓰레기 무단 투기 행동 예측 (데이터 부족)
 - 시간대, 지역별 무단 투기 발생 가능성
 - 데이터
 - 생활 쓰레기 신고 현황
 - cctv 단속 이력
 - 인구, 상권 정보
 - 방향성 및 기법
 - 위치적으로 어디서 무단투기가 발생하는지 시각화
 - Gradient Boosting** 기반 예측모델로 피처 중요도 분석 (날씨나 축제등)
 - 투기 밀도가 높은 곳 지도에 표시
 - 기대 인사이트
 - 어느 시간대에 상업 밀집도 무단 투기 발생 확률 높프로
 - 단속 차량 순회 경로 최적화나 cctv 설치가 안된곳 등의 해결책
 - 소음 공해 지역 예측 (데이터 부족)
 - 소음 수준이 일정 기준 이상인 구역 집중 관리
 - 데이터
 - 위치, 시간별 소음(dB)
 - 역 주변 소음 유발 요인 (지하철 운행정보)
 - 상업 정보
 - 방향성 및 기법
 - 소음 군집화
 - 시간대, 교통량, 건물 밀도등의 변수 중요도
 - 위치별 소음도 시각화
 - 기대 인사이트
 - 방음벽 설치 우선 지역 설정

▼ EDA 분석 결과_1차 데이터

[1] df1 ([seoul_recycle.csv](#))

1-1. basic EDA

▼ [df.info\(\)](#)

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25 entries, 0 to 24
Data columns (total 55 columns):
Column Non-Null Count Dtype
--- 
0 city      25 non-null  object
1 gu        25 non-null  object
2 waste_generation  25 non-null  float64
3 recycled_amount   25 non-null  float64
4 recycling_rate    25 non-null  float64
5 population_total   25 non-null  float64
6 househole    25 non-null  float64
7 population_per_household 25 non-null  float64
8 man_population    25 non-null  float64
9 woman_population   25 non-null  float64
10 man_woman_ratio   25 non-null  float64
11 Oto9_pop     25 non-null  int64
12 10to19_pop    25 non-null  int64
13 20to29_pop    25 non-null  int64
14 30to39_pop    25 non-null  int64
15 40to49_pop    25 non-null  int64
16 50to59_pop    25 non-null  int64
17 60to69_pop    25 non-null  int64
18 70to79_pop    25 non-null  int64

```

```

19 80to89_pop      25 non-null int64
20 90to99_pop      25 non-null int64
21 100over_pop     25 non-null int64
22 0to9_man_pop    25 non-null int64
23 10to19_man_pop  25 non-null int64
24 20to29_man_pop  25 non-null int64
25 30to39_man_pop  25 non-null int64
26 40to49_man_pop  25 non-null int64
27 50to59_man_pop  25 non-null int64
28 60to69_man_pop  25 non-null int64
29 70to79_man_pop  25 non-null int64
30 80to89_man_pop  25 non-null int64
31 90to99_man_pop  25 non-null int64
32 100over_man_pop 25 non-null int64
33 0to9_woman_pop  25 non-null int64
34 10to19_woman_pop 25 non-null int64
35 20to29_woman_pop 25 non-null int64
36 30to39_woman_pop 25 non-null int64
37 40to49_woman_pop 25 non-null int64
38 50to59_woman_pop 25 non-null int64
39 60to69_woman_pop 25 non-null int64
40 70to79_woman_pop 25 non-null int64
41 80to89_woman_pop 25 non-null int64
42 90to99_woman_pop 25 non-null int64
43 100over_woman_pop 25 non-null int64
44 P1              25 non-null int64
45 P2              25 non-null int64
46 P3              25 non-null int64
47 P4              25 non-null int64
48 P5              25 non-null int64
49 P6              25 non-null int64
50 P7              25 non-null int64
51 P8              25 non-null int64
52 P9              25 non-null int64
53 zero_restaurant 25 non-null int64
54 zero_re_cnt     25 non-null int64

```

dtypes: float64(9), int64(44), object(2)
memory usage: 10.9+ KB

▼ df.describe()

	waste_generation	recycled_amount	recycling_rate	population_total	househole	population_per_household	man_population	woman_population	man_woman_ratio	Oto9_pop	...	P2	P3	P4
count	25.000000	25.000000	25.000000	25.000000	25.000000	25.000000	25.000000	25.000000	25.000000	25.000000	...	25.0	25.00	25.000000
mean	336.636000	174.160000	52.996000	379875.480000	178424.160000	2.115600	184297.600000	195577.880000	0.944800	22713.840000	...	0.4	0.04	0.120000
std	261.038445	117.856166	9.679476	127585.540947	56488.95869	0.169928	61449.69844	66281.839682	0.030155	9600.862616	...	0.5	0.20	0.331662
min	182.400000	87.400000	38.030000	122054.000000	63880.000000	1.730000	59368.000000	62686.000000	0.890000	6091.000000	...	0.0	0.00	0.000000
25%	218.800000	120.500000	46.830000	305615.000000	144866.000000	2.000000	145681.000000	159934.000000	0.920000	17276.000000	...	0.0	0.00	0.000000
50%	249.300000	135.600000	52.550000	385755.000000	181921.000000	2.090000	186455.000000	197161.000000	0.940000	22332.000000	...	0.0	0.00	0.000000
75%	336.500000	177.900000	57.400000	461417.000000	202489.000000	2.260000	226062.000000	235355.000000	0.960000	28528.000000	...	1.0	0.00	0.000000
max	1478.300000	625.000000	80.250000	662960.000000	285939.000000	2.450000	319461.000000	343499.000000	1.020000	48063.000000	...	1.0	1.00	1.000000

8 rows x 53 columns

▼ df.isnull().sum()

	0
city	0
gu	0
waste_generation	0
recycled_amount	0
recycling_rate	0
population_total	0
househole	0
population_per_household	0
man_population	0
woman_population	0
man_woman_ratio	0
Oto9_pop	0
10to19_pop	0
20to29_pop	0
30to39_pop	0
40to49_pop	0
50to59_pop	0
60to69_pop	0
70to79_pop	0
80to89_pop	0
90to99_pop	0
100over_pop	0
Oto9_man_pop	0
10to19_man_pop	0
20to29_man_pop	0

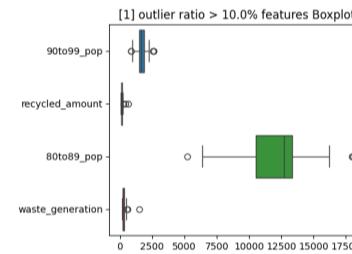
30to39_man_pop	0
40to49_man_pop	0
50to59_man_pop	0
60to69_man_pop	0
70to79_man_pop	0
80to89_man_pop	0
90to99_man_pop	0
100over_man_pop	0
Oto9_woman_pop	0
10to19_woman_pop	0
20to29_woman_pop	0
30to39_woman_pop	0
40to49_woman_pop	0
50to59_woman_pop	0
60to69_woman_pop	0
70to79_woman_pop	0
80to89_woman_pop	0
90to99_woman_pop	0
100over_woman_pop	0
P1	0
P2	0
P3	0
P4	0
P5	0
P6	0
P7	0
P8	0
P9	0
zero_restaurant	0
zero_re_cnt	0

dtype: int64

▼ df.columns

```
Index(['city', 'gu', 'waste_generation', 'recycled_amount', 'recycling_rate',
       'population_total', 'househole', 'population_per_household',
       'man_population', 'woman_population', 'man_woman_ratio', 'Oto9_pop',
       '10to19_pop', '20to29_pop', '30to39_pop', '40to49_pop', '50to59_pop',
       '60to69_pop', '70to79_pop', '80to89_pop', '90to99_pop', '100over_pop',
       'Oto9_man_pop', '10to19_man_pop', '20to29_man_pop', '30to39_man_pop',
       '40to49_man_pop', '50to59_man_pop', '60to69_man_pop', '70to79_man_pop',
       '80to89_man_pop', '90to99_man_pop', '100over_man_pop', 'Oto9_woman_pop',
       '10to19_woman_pop', '20to29_woman_pop', '30to39_woman_pop',
       '40to49_woman_pop', '50to59_woman_pop', '60to69_woman_pop',
       '70to79_woman_pop', '80to89_woman_pop', '90to99_woman_pop',
       '100over_woman_pop', 'P1', 'P2', 'P3', 'P4', 'P5', 'P6', 'P7', 'P8',
       'P9', 'zero_restaurant', 'zero_re_cnt'],
      dtype='object')
```

▼ outlier with boxplot (10% 이상의 이상치만 표시)



1-2. features 간의 corr

- value 가 unique 한 features: city, P7
- 두 feature 는 corr 에서 제외 (variance 가 0 이 되므로)

▼ continuous columns

```
Index(['Oto9_man_pop', 'Oto9_pop', 'Oto9_woman_pop', '100over_man_pop',
       '100over_pop', '100over_woman_pop', '10to19_man_pop', '10to19_pop',
       '10to19_woman_pop', '20to29_man_pop', '20to29_pop', '20to29_woman_pop',
       '30to39_man_pop', '30to39_pop', '30to39_woman_pop', '40to49_man_pop',
       '40to49_pop', '40to49_woman_pop', '50to59_pop', '50to59_man_pop',
       '50to59_woman_pop', '60to69_man_pop', '60to69_pop', '60to69_woman_pop',
       '70to79_man_pop', '70to79_pop', '70to79_woman_pop', '80to89_man_pop',
       '80to89_pop', '80to89_woman_pop', '90to99_man_pop', '90to99_pop',
       '90to99_woman_pop', 'househole', 'man_population', 'man_woman_ratio',
       'population_per_household', 'population_total', 'recycled_amount',
       'recycling_rate', 'waste_generation', 'woman_population',
       'zero_re_cnt'],
      dtype='object')
```

▼ binary columns

only 0/1: [P1, P2, P3, P4, P5, P6, P8, P9, 'zero_restaurant']

▼ categorical columns

```
Index(['gu'], dtype='object')
```

▼ 기본 corr (각 feature)

- abs threshold 가 0.5 이상인 것

```
[cont-cont]
→ waste_generation & ≥ 0.5:
recycled_amount 0.94
30to39_woman_pop 0.60
30to39_pop 0.60
30to39_man_pop 0.58
90to99_woman_pop 0.53
90to99_pop 0.52
househole 0.51

→ recycled_amount & ≥ 0.5:
waste_generation 0.94
30to39_pop 0.53
30to39_woman_pop 0.52
30to39_man_pop 0.52
```

[cont-bin]
→ zero_re_cnt & ≥ 0.4:
zero_restaurant 0.91

[cont-cate]
→ waste_generation & ≥ 0.4:
gu_gangseo_gu 0.91

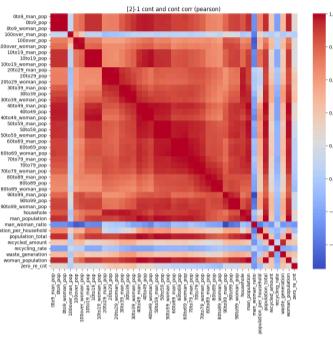
→ recycled_amount & ≥ 0.4:
gu_gangseo_gu 0.80

→ recycling_rate & ≥ 0.4:
gu_yongsan_gu 0.59

[bin-bin]
→ P1 & ≥ 0.5:
P2 0.84

[bin-cate]
없음

▼ cont & cont: pearson

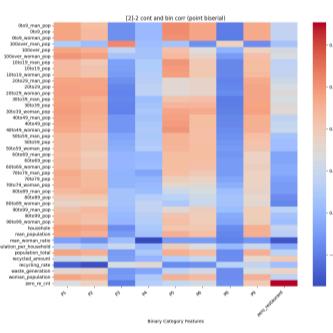


▼ waste_generation, recycled_amount, recycling_rate, zero_re_cmn 와의 corr 이 threshold 0.5 (abs) 이상인 continuous features

→ waste_generation & ≥ 0.5:
recycled_amount 0.94
30to39_woman_pop 0.60
30to39_pop 0.60
30to39_man_pop 0.58
90to99_woman_pop 0.53
90to99_pop 0.52
househole 0.51

→ recycled_amount & ≥ 0.5:
waste_generation 0.94
30to39_pop 0.53
30to39_woman_pop 0.52
30to39_man_pop 0.52

▼ cont & binary: point biserial (biserial, polyserial)

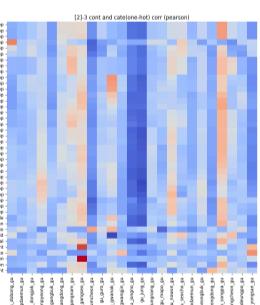


▼ waste_generation, recycled_amount, recycling_rate, zero_re_cmn 와의 corr 이 threshold 0.4 (abs) 이상인 binary features

→ recycled_amount & ≥ 0.4:
zero_restaurant 0.40

→ zero_re_cnt & ≥ 0.4:
zero_restaurant 0.91
P9 0.40

▼ cont & cate(one-hot): pearson



▼ waste_generation, recycled_amount, recycling_rate, zero_re_cmn 와의 corr 이 threshold 0.4 (abs) 이상인 catgrorical features

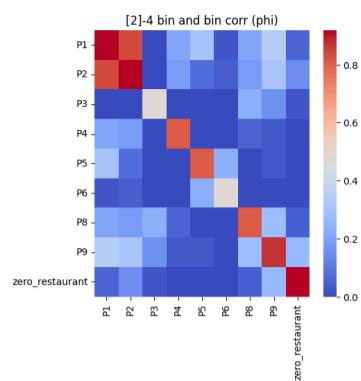
→ waste_generation & ≥ 0.4:
gu_gangseo_gu 0.91

→ recycled_amount & ≥ 0.4:
gu_gangseo_gu 0.80
gu_yongsan_gu 0.44

→ recycling_rate & ≥ 0.4:
gu_yongsan_gu 0.59
gu_gwangjin_gu 0.41

→ zero_re_cnt & ≥ 0.4:
gu_seongdong_gu 0.44

▼ binary & binary: cramers V (카이제곱, chi2_contingency)



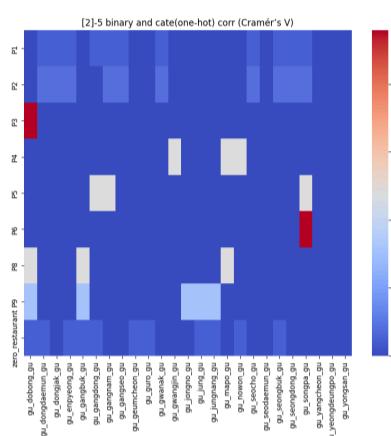
▼ `binary_cols` 의 각 corr이 threshold 0.5 (abs) 이상인 binary features

→ P1 & ≥ 0.5 :
P2 0.84

→ P2 & ≥ 0.5 :
P1 0.84

▼ cate & cate: 불가

▼ binary & cate(one-hot): cramér's V (카이제곱, `chi2_contingency`)



▼ `binary_cols` 와의 corr이 threshold 0.4 (abs) 이상인 categorical features

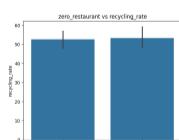
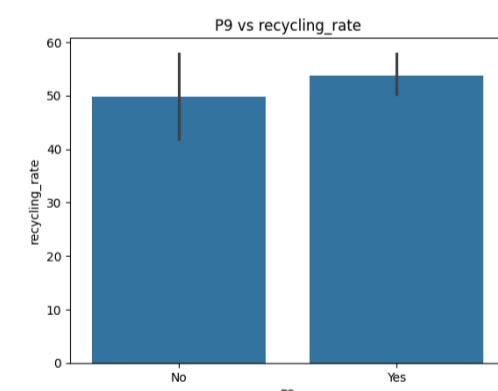
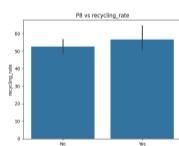
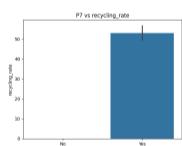
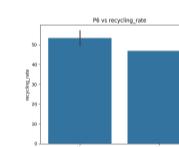
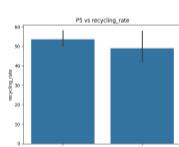
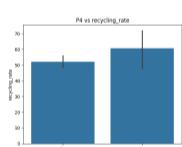
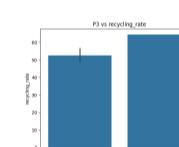
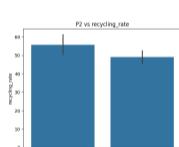
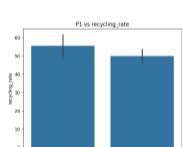
→ P3 & ≥ 0.4 :
gu_dobong_gu 0.48

→ P6 & ≥ 0.4 :
gu_songpa_gu 0.48

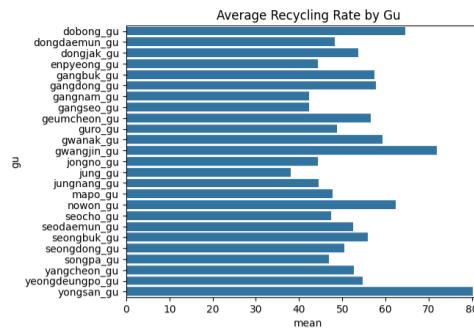
1-3. extra visualization

▼ barplot

▼ policy별 recycling rate



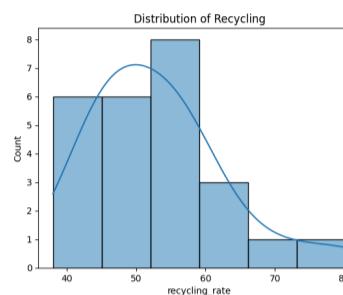
▼ gu 별



▼ (추가: pop,age등) 별

▼ histplot

▼ recycling rate 분포



1-4. Clustering

1. [Cluster] pop & recycling rate (population_density & recycling_rate)

• (waste_generation 추가 시 clustering 부적절)

▼ add pop density column using area_km2

```
df['pop_density'] = df['population_total'] / df['area_km2']
```

▼ Clustering

▼ cluster features: recycling_rate, pop_density

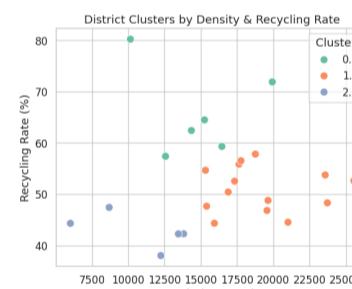
▼ cluster 기법: KMeans

▼ cluster 결과

Cluster 0 (초록)
- 대체로 중간~높은 인구밀도에
- 높은 재활용률을 보이는 구들

Cluster 1 (주황)
- 가장 높은 인구밀도에
- 재활용률은 중간 정도인 구들

Cluster 2 (파랑)
- 낮은 인구밀도에
- 재활용률도 낮은 구들

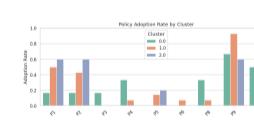
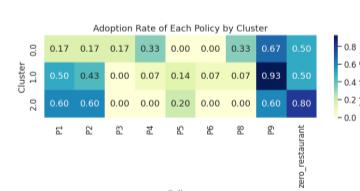


▼ cluster 결과 확인 지표 (적당한 cluster)

Silhouette Score: 0.367
Calinski-Harabasz Index: 17.4
Davies-Bouldin Index: 0.854

▼ Clustering insight

▼ policy (P7 제외)



▼ 최종 분석 결과_[insight_policy]

Cluster ↓ \ Policy →	P1	P2	P3	P4	P5	P6	P8	P9	zero_restaurant
0 (중밀집·고재활용)	0.17	0.17	0.17	0.33	0.00	0.33	0.67	0.50	
1 (고밀집·중재활용)	0.50	0.43	0.00	0.07	0.14	0.07	0.07	0.93	0.50
2 (저밀집·저재활용)	0.60	0.60	0.00	0.20	0.00	0.00	0.60	0.60	0.80

• P8 (특수품목분리수거)

- cluster0에서만 33% adoption → 고재활용에 높은 영향
- 1,2 cluster 의 중, 저재활용 -> 확대해야할 정책 후보

• P4 (봉제원단재활용)

- Cluster 0이 33%, Cluster 1:2 거의 0% 채택 → Cluster 0 재활용을 상승에 기여했을 가능성 큼
- 1,2 cluster 의 중, 저재활용 -> 확대해야할 정책 후보

• P1, P2 (재활용정거장·무인회수기)

- Cluster 2(저밀집·저재활용)에서 각각 60% 채택
- Cluster 0(고성과)에서는 17%밖에 안 채택 ⇒ 단독 도입만으로는 재활용을 개선 효과가 미미한 듯

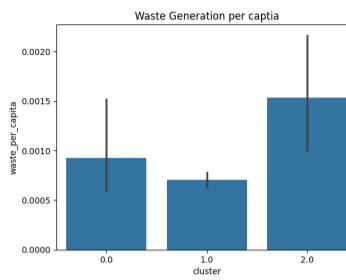
• P9 (회수보상제)

- 거의 전 클러스터(60~93%)에서 도입되었지만, - Cluster 1(93%)은 중간 재활용율(약 55%) → P9만으로는 한계가 있는 정책

• zero_restaurant

- Cluster 2에서 80%로 가장 높으나 재활용율은 제일 낮음 ⇒ "제로 식당" 정책 단독 도입만으로는 재활용율을 끌어올리지 못함

▼ waste_generation_captia



- anova 결과: ANOVA: F = 5.65, p = 0.010 (통계적 유의미)

▼ 최종 분석 결과_[insight_waste_generation_captia]

Cluster	waste_generation_captia
0 (중밀집·고재활용)	0.000929
1 (고밀집·중재활용)	0.000707
2 (저밀집·저재활용)	0.001533

- Cluster0: 중간 1인당 배출량 (0.000929)
- Cluster1: 가장 낮은 1인당 배출량 (0.000707)
 - 인구 밀집도가 높고, 재활용 시스템이 적당히 잘 갖춰진 환경에서 1인당 배출을 최소화하고 있음으로 예상
- Cluster2: 주택 면적이 넓거나 소비 패턴이 도심과 달라서 쓰레기 발생이 큰 것으로 추정 (거의 Cluster1의 두배)

▼ pop_distribution

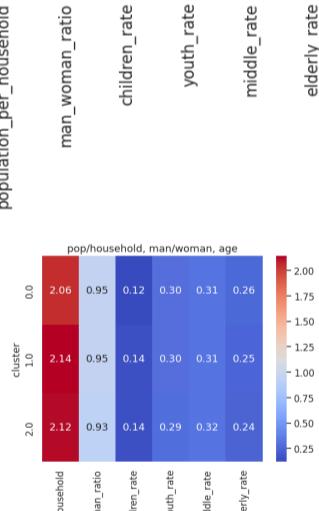
- age column

```
'10to19_pop', '20to29_pop', '30to39_pop', '40to49_pop', '50to59_pop',
'60to69_pop', '70to79_pop', '80to89_pop', '90to99_pop', '100over_pop'
```

- cluster 별 비교할 features

```
'population_per_household', 'man_woman_ratio',
'children_rate', 'youth_rate', 'middle_rate', 'elderly_rate'
```

	population_per_household	man_woman_ratio	children_rate	youth_rate	middle_rate	elderly_rate
cluster						
0.0	2.06	0.95	0.12	0.30	0.31	0.26
1.0	2.14	0.95	0.14	0.30	0.31	0.25
2.0	2.12	0.93	0.14	0.29	0.32	0.24



▼ 최종 분석 결과_[insight_waste_generation_captia]

지표	0 클러스터	1 클러스터	2 클러스터	최대-최소 차이
population_per_household	2.06	2.14	2.12	0.08
man_woman_ratio	0.95	0.95	0.93	0.02
children_rate	0.12	0.14	0.14	0.02
youth_rate	0.30	0.30	0.29	0.01
middle_rate	0.31	0.31	0.32	0.01
elderly_rate	0.26	0.25	0.24	0.02

- 별로 의미 없어보임 (분류 기준으로 보기 힘들듯)

▼ zero_restaurant_count, zero_re_density

cluster	area_km2	zero_re_cnt	zero_re_density
0.0	24.698333	22.333333	0.904245
1.0	21.088571	24.714286	1.171928
2.0	32.358000	36.800000	1.137277



▼ 최종 분석 결과_[insight_waste_generation_captia]

Cluster	area_km2	zero_re_cnt	zero_re_density
0 (중밀집·고재활용)	24.70	22.33	0.904245
1(고밀집·중재활용)	21.09	24.71	1.171928
2 (저밀집·저재활용)	32.36	36.80	1.137277

- Cluster 0(중밀집·고재활용)
 - 단위 면적당 식당이 상대적으로 적어, “제로 식당”이 미미하게 분포
- Cluster 1(고밀집·중재활용)
 - 면적 대비 식당 수가 특히 많아 밀도가 가장 높음
- Cluster 2(저밀집·저재활용)
 - 전체 수는 많지만 면적도 넓어, 밀도는 Cluster 1에 비해 약간 낮은 중간값

▼ 최종 분석 결과_policy별

정책 코드	정의	채택율 (Cluster 0 / 1 / 2)	관찰된 효과	권장 액션
P1	재활용정책	0.17 / 0.50 / 0.60	- Cluster 0: 저채택에도 재활용률 최고 - Cluster 2: 높은 채택에도 재활용률 최저 ⇒ 단독효과 제한적	다른 정책(예: P4-P8)과 묶어 보완적 확대
P2	무인회수기	0.17 / 0.43 / 0.60	P1과 유사 양상: 단독 도입만으로 재활용률 개선에 한계	P1과 병행 도입, 접근성 높은 지역 우선 확대
P3	폐현수막재활용	0.17 / 0.00 / 0.00	Cluster 0에서만 소량 채택, 효과 겸증 불충분	시범사업 형태로 Cluster 1-2에 파일럿 확대 검토
P4	봉제원단 재활용	0.33 / 0.00 / 0.00	Cluster 0에서만 채택 → 재활용률 상승 가능성 높음	Cluster 1-2 전면 확대 도입
P5	일회용컵보증금제	0.00 / 0.14 / 0.20	중·저재활용 clusters 만 일부 채택, 재활용률과 명확한 상관관계 미확인	추가 데이터 수집 후 효과 평가
P6	다회용기도입	0.00 / 0.07 / 0.00	채택 극소수, 영향 평가 불가	우선순위 낮음, 필요시 타 정책과 패키지화
P8	특수품목분리수거	0.33 / 0.07 / 0.00	Cluster 0에서만 채택 → 고재활용률과 연관성 시사	Cluster 1-2 확대 도입
P9	재활용품회수보상제	0.67 / 0.93 / 0.60	전 클러스터에서 광범위 도입 ⇒ Cluster 1-2 재활용률 증·저위치 ⇒ 단독효과 한계	인센티브 강화 방안 검토, 보상 수준 조정
zero_restaurant	‘제로 식당’(배출 저감 식당)	0.50 / 0.50 / 0.80	Cluster 2 채택 최다에도 재활용률 최저 ⇒ 단일 정책으로 효과 미흡	다른 분리수거·보상 정책과 연계 패키지화

▼ 최종 분석 결과_cluster 별 basic

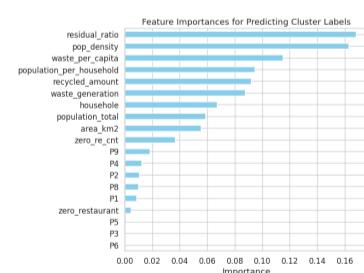
Cluster (특성)	District	Avg. Waste per capita	Zero Restaurant Density	주요 정책 확장 후 보	Pop. Distribution 차이 (의미 X)	District 특징 요약	확대·신규 도입할 정책	그 외 추천 프로그램·조치
0 (중밀집·고재활용)	[yongsan_gu, gwangjin_gu, gangbuk_gu, dobong_gu, nowon_gu, gwanak_gu]	0.000929	0.9042	—(현재 성과 유지)	미미 (최대-최소 0.08)	- 재활용률 최고 - 1인당 배출량 중간 (0.00093) - 제로 식당 집계 밀도 낮음 (0.90)	- P5 신규 검토 (채택률 0.00) - P6 신규 검토 (채택률 0.00)	- 제로 식당 추가 설치 - 포장재 감량·분리배출 교육
1(고밀집·중재활용)	[seongdong_gu, dongdaemun_gu, jungnang_gu, seongbuk_gu, enpyeong_gu, seodaemun_gu, mapo_gu, yangcheon_gu, guro_gu, geumcheon_gu, yeongdeungpo_gu, dongjak_gu, songpa_gu, gangdong_gu]	0.000707	1.1719	P4, P8	미미 (최대-최소 0.08)	- 재활용률 중간 - 1인당 배출량 최저 (0.00071) - 제로 식당 집계 밀도 최고 (1.17)	- P3 신규 도입 (채택률 0.00) - P4 확대 (채택률 0.07) - P6 신규 도입 (채택률 0.07) - P8 확대 (채택률 0.07)	- 스마트 분리배출 알림 - 재활용 캠페인 (청년·맞벌이)
2 (저밀집·저재활용)	[jongno_gu, jung_gu, gangseo_gu, seocho_gu, gangnam_gu]	0.001533	1.1373	P4, P8	미미 (최대-최소 0.08)	- 재활용률 최저 - 1인당 배출량 최고 (0.00153) - 제로 식당 집계 밀도 중간 (1.14)	- P3, P4, P6, P8 전부 신규 도입 (채택률 0.00)	- 감방형 요금제(RFID) (배출량 만큼만 수수료를 지불하는 음식물쓰레기 종량제 시스템) - 음식물 감량 캠페인 - 소규모 무인 수거함 설치

▼ 최종 분석 결과_cluster 별 policy

Cluster (특성)	District	P1	P2	P3	P4	P5	P6	P8	P9	zero_restaurant
0 (중밀집·고재활용)	yongsan_gu, gwangjin_gu, gangbuk_gu, dobong_gu, nowon_gu, gwanak_gu	17% → 현행 유지 (P4-P8 패키지 연계)	17% → 현행 유지 (접근성 보완)	0% → 우선순위 낮음(시범사업 불필요)	33% → 현행 유지	0% → 파일럿 검토	0% → 파일럿 검토	33% → 현행 유지	67% → 인센티브 구조화	50% → 추가 설치·교체
1(고밀집·중재활용)	seongdong_gu, dongdaemun_gu, jungnang_gu, seongbuk_gu, enpyeong_gu, seodaemun_gu, mapo_gu, yangcheon_gu, guro_gu, geumcheon_gu, yeongdeungpo_gu, dongjak_gu, songpa_gu, gangdong_gu	50% → 분리수거 패키지 병행	43% → 분리수거 패키지 병행	0% → 신규 도입 강력 권장(파일럿→전면)	7% → 즉시 확대 도입	14% → 데이터 추가 수집	7% → 신규 도입 검토	7% → 즉시 확대 도입	93% → 보상 수준 조정·정책 연계	50% → 패키지 포함
2 (저밀집·저재활용)	jongno_gu, jung_gu, gangseo_gu, seocho_gu, gangnam_gu	60% → 보완 필요	60% → 보완 필요	0% → 신규 도입 필수	0% → 신규 도입 필수	20% → 시범사업 후 확대 검토	0% → 신규 도입 검토	0% → 신규 도입 필수	60% → 인센티브 강화	80% → 분리수거 패키지 연계

2. other clustering

1-5. Feature Importance using Cluster



[2] df2 (zero_restaurant_gu_en.xlsx)

2-1. basic EDA

df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 664 entries, 0 to 663
Data columns (total 4 columns):
 #   Column   Non-Null Count Dtype  
--- 
 0   store_name    664 non-null   object  
 1   address       664 non-null   object  
 2   district      664 non-null   object  
 3   district_en   664 non-null   object  
dtypes: object(4)
memory usage: 20.9+ KB
```

df.describe()

	store_name	address	district	district_en
count	664	664	664	664
unique	662	627	14	14
top	금강에프엔비주식회사	서울 송파구 문정동 652	성동구	seongdong_gu
freq	2	6	82	82

▼ df.isnull().sum()

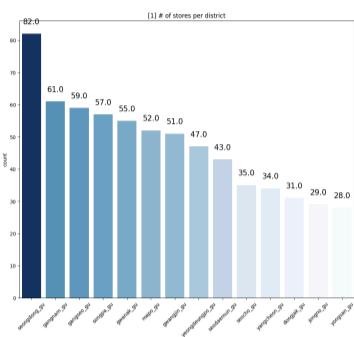
	0
store_name	0
address	0
district	0
district_en	0

dtype: int64

▼ df.columns

Index(['store_name', 'address', 'district', 'district_en'], dtype='object')

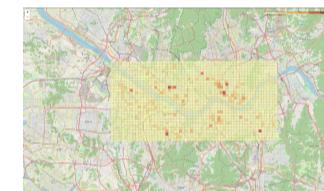
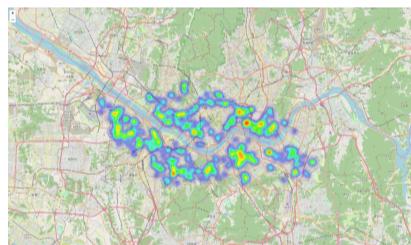
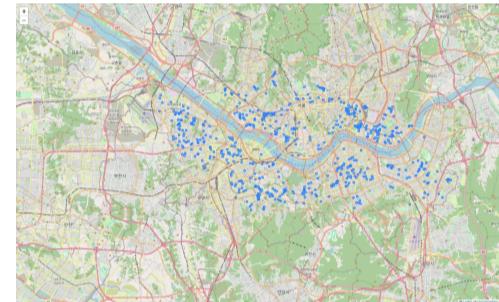
2-2. barplot (district & store_name)



2-3. map (with geocoding)

- geocoding (longitude, latitude 열 추가)

<https://velog.io/@elij9501/%ED%8C%8C%EC%9D%84%EC%8D%AC%EC%9C%BC%EB%A1%9C-%EC%9C%84%EB%8E%84%EA%B2%BD%EB%8E%84-%EC%80%BE%EA%B8%80@geocoder.geocoding-API%EA%B5%AC%EA%B8%80-%EC%8A%A4%ED%94%84%EB%A0%88%EB%93%9C%EC%BB%9C%ED%BA%B8>



2-4. text mining



▼ code

```

tokens = df['store_name'].\
    str.replace('[^ㄱ-ㅎㅏ-ㅣㅏ-ㅣ]', '') \
    .str.split() \
    .explode()
freq = Counter(tokens)

font_path = '/usr/share/fonts/truetype/nanum/NanumBarunGothic.ttf'

wc = WordCloud(
    font_path=font_path,
    width=800,
    height=400,
    background_color='white'
).generate_from_frequencies(freq)

plt.figure(figsize=(10,5))
plt.imshow(wc, interpolation='bilinear')
plt.axis('off')
plt.show()

• top 20

[('두점', 13), ('주식회사', 10), ('신촌점', 6), ('마라탕', 6), ('두마리찜닭', 6),
 ('한뿌리죽', 5), ('구의점', 5), ('송파점', 5), ('분점', 5), ('강남점', 4),
 ('떡볶이', 4), ('김치점', 4), ('진전복삼계탕', 4), ('왕십리점', 4),
 ('영등포점', 4), ('교대점', 3), ('키친', 3), ('신림점', 3), ('봉천점', 3),
 ('건대점', 3)]

```

[3] Merge (seoul_recycle_2.csv + zero_restaurant_gu_en_2.csv)

3-1. clustering with text mining

- 이 경우, 각 zero_rest 의 업종 column 필요할 듯
- 각 rest 의 업종 열을 추가한다면, text mining 필요없음 (각 rest 의 업종 열을 각 업종 개수를 값으로 가지는 열로 세분화한 후, 바로 clustering)
 - 예시

```
import re

def categorize(name):
    name = name.lower()
    if re.search('치킨', name):
        return 'chicken'
    if re.search('카페|coffee', name):
        return 'cafe'
    if re.search('분식|떡볶이|김밥', name):
        return 'snack'
    if re.search('초밥|초밥집|스시', name):
        return 'sushi'
    if re.search('한식|비빔밥|국수|찌개', name):
        return 'korean'
    if re.search('중식|중국집', name):
        return 'chinese'
    if re.search('양식|파스타|피자|스테이크', name):
        return 'western'
    # ...
    return 'other'

df['category'] = df['store_name'].apply(categorize)
```

▼ 예시 merge 데이터 형태

district_en	waste_generation	recycling_rate	pop_density	chinese	japanese	korean	western	snack	other
gangnam_gu	474.9	42.3	13461.87	12	8	15	5	3	18
seoho_gu	350.6	47.6	8689.23	4	5	7	2	1	9
seongdong_gu	209.8	50.4	16894.95	3	2	9	1	4	11
yeongdeungpo_gu	326.3	54.7	15327.49	6	3	8	4	2	12
jongno_gu	208.5	44.3	6016.31	2	1	12	0	2	7

▼ 각 업종별로 묶은뒤 avg_waste, avg_recycling_rate 이용해서 cluster

- 예상 인사이트

```
==== Category Summary with Clusters ====
   category count avg_waste avg_recycling cr_cluster
0 chicken 120  500.2    45.6       0
1 snack   150  300.1    50.8       1
2 cafe    200  320.4    55.3       1
3 sushi   80   280.7    60.2       1
4 korean  90   480.1    42.1       0
5 chinese 70   450.3    40.5       0
6 western  60   350.8    47.0       2
7 other   40   290.5    48.9       2
```

해석 예시: cluster1 이 저배출, 고재활용 이므로, snack, cafe, sushi 의 경우 정책효과 잘 나타남

▼ 각 구별로 feature 는 pop_density, recycling rate, chicken 등 업종 열

```
==== District-level Data with food_cluster ====
   district_en food_cluster pop_density recycling_rate chicken cafe snack sushi korean chinese western other
0 gangnam_gu      1  13461.87      42.3   12  20  5  3  8  10  2  4
1 seoho_gu        2  8689.23      47.6   4  8  3  2  5  4  1  3
2 seongdong_gu    0  16894.95      50.4   3  5  4  1  6  2  1  2
3 yeongdeungpo_gu 1  15327.49      54.7   6  12  2  1  7  3  2  5
4 jongno_gu       0  6016.31      44.3   2  4  3  0  10  1  0  2

food_cluster
0 16234.12
1 13461.87
2 8689.23
Name: pop_density, dtype: float64
```

해석 예시: food_cluster 0(한식·분식 중심)은 재활용률이 비교적 높으니, 이 유형의 업종을 다른 지역에도 확산시키면 효과적

3-2. cross tab

- 업종별 policy 채택률 확인
 - 가능 insight ex: 'zero rest 중 일식은 여전히 P1 이 잘 이루어지지 않는다'

▼ EDA 분석 결과_2차 데이터

[1] df1 (seoul_recycle.csv)

1-1. basic EDA

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25 entries, 0 to 24
Data columns (total 79 columns):
 #   Column          Non-Null Count Dtype  
--- 
 0   city            25 non-null   object  
 1   gu              25 non-null   object  
 2   waste_generation 25 non-null   float64 
 3   recycled_amount  25 non-null   float64 
 4   recycling_rate   25 non-null   float64 
 5   population_total 25 non-null   int64   
 6   household        25 non-null   int64   
 7   population_per_household 25 non-null   float64 
 8   man_population   25 non-null   int64   
 9   woman_population 25 non-null   int64   
 10  man_woman_ratio  25 non-null   float64 
 11  Oto9_pop         25 non-null   int64   
 12  10to19_pop       25 non-null   int64   
 13  20to29_pop       25 non-null   int64   
 14  30to39_pop       25 non-null   int64   
 15  40to49_pop       25 non-null   int64   
 16  50to59_pop       25 non-null   int64   
 17  60to69_pop       25 non-null   int64   
 18  70to79_pop       25 non-null   int64   
 19  80to89_pop       25 non-null   int64   
 20  90to99_pop       25 non-null   int64   
 21  100over_pop      25 non-null   int64   
 22  Oto9_man_pop     25 non-null   int64   
 23  10to19_man_pop   25 non-null   int64   
 24  20to29_man_pop   25 non-null   int64   
 25  30to39_man_pop   25 non-null   int64
```

```

26 40to49_man_pop    25 non-null int64
27 50to59_man_pop    25 non-null int64
28 60to69_man_pop    25 non-null int64
29 70to79_man_pop    25 non-null int64
30 80to89_man_pop    25 non-null int64
31 90to99_man_pop    25 non-null int64
32 100over_man_pop   25 non-null int64
33 Oto9_woman_pop    25 non-null int64
34 10to19_woman_pop  25 non-null int64
35 20to29_woman_pop  25 non-null int64
36 30to39_woman_pop  25 non-null int64
37 40to49_woman_pop  25 non-null int64
38 50to59_woman_pop  25 non-null int64
39 60to69_woman_pop  25 non-null int64
40 70to79_woman_pop  25 non-null int64
41 80to89_woman_pop  25 non-null int64
42 90to99_woman_pop  25 non-null int64
43 100over_woman_pop 25 non-null int64
44 P1                 25 non-null int64
45 P2                 25 non-null int64
46 P3                 25 non-null int64
47 P4                 25 non-null int64
48 P5                 25 non-null int64
49 P6                 25 non-null int64
50 P7                 25 non-null int64
51 P8                 25 non-null int64
52 P9                 25 non-null int64
53 zero_restaurant   25 non-null int64
54 zero_re_cnt        25 non-null int64
55 area_km2           25 non-null float64
56 pop_density         25 non-null float64
57 cluster            25 non-null int64
58 waste_per_captia  25 non-null float64
59 children_rate       25 non-null float64
60 youth_rate          25 non-null float64
61 middle_rate          25 non-null float64
62 elderly_rate         25 non-null float64
63 zero_re_density    25 non-null float64
64 residual_waste     25 non-null float64
65 residual_ratio      25 non-null float64
66 Total_Boarding     25 non-null int64
67 Total_Alighting    25 non-null int64
68 num_businesses      25 non-null int64
69 num_employees        25 non-null int64
70 Beverage_Retail     25 non-null int64
71 Buffet              25 non-null int64
72 Chinese_Food        25 non-null int64
73 General_Lodging     25 non-null int64
74 Japanese_Food       25 non-null int64
75 Korean_Food          25 non-null int64
76 Other_Lodging        25 non-null int64
77 Pub                 25 non-null int64
78 Western_Food         25 non-null int64
dtypes: float64(15), int64(62), object(2)
memory usage: 15.6+ KB

```

▼ df.describe()

	waste_generation	recycled_amount	recycling_rate	population_total	househole	population_per_household	man_population	woman_population	man_woman_ratio	Oto9_pop	...	num_employees	Beverage_Retail	Buffet
count	25.000000	25.000000	25.000000	25.000000	25.000000	25.000000	25.000000	25.000000	25.000000	25.000000	...	25.000000	25.0000	25.000000
mean	336.636800	174.160000	52.996000	379875.480000	178424.160000	2.115600	184297.600000	195577.880000	0.944800	22713.840000	...	231817.000000	29.6400	44.040000
std	261.038445	117.856166	9.679476	127585.540947	56488.95869	0.169928	61449.69844	66281.839682	0.030155	9600.862616	...	167725.603353	25.0913	25.430428
min	182.400000	87.400000	38.030000	122054.000000	63880.000000	1.730000	59368.000000	62686.000000	0.890000	6091.000000	...	77618.000000	8.0000	15.000000
25%	218.800000	120.500000	46.830000	305615.000000	144866.000000	2.000000	145681.000000	159934.000000	0.920000	17276.000000	...	125727.000000	13.0000	28.000000
50%	249.300000	135.600000	52.550000	385755.000000	181921.000000	2.090000	186455.000000	197161.000000	0.940000	22332.000000	...	151851.000000	22.0000	36.000000
75%	336.500000	177.900000	57.400000	461417.000000	202489.000000	2.260000	226062.000000	235355.000000	0.960000	28528.000000	...	278403.000000	31.0000	58.000000
max	1478.300000	625.000000	80.250000	662960.000000	285939.000000	2.450000	319461.000000	343499.000000	1.020000	48063.000000	...	802908.000000	122.0000	125.000000

8 rows × 77 columns

▼ df.isnull().sum()

	0
city	0
gu	0
waste_generation	0
recycled_amount	0
recycling_rate	0
population_total	0
househole	0
population_per_household	0
man_population	0
woman_population	0
man_woman_ratio	0
Oto9_pop	0
10to19_pop	0
20to29_pop	0
30to39_pop	0
40to49_pop	0
50to59_pop	0
60to69_pop	0
70to79_pop	0
80to89_pop	0
90to99_pop	0
100over_pop	0
Oto9_man_pop	0
10to19_man_pop	0
20to29_man_pop	0
30to39_man_pop	0
40to49_man_pop	0
50to59_man_pop	0
60to69_man_pop	0
70to79_man_pop	0
80to89_man_pop	0
90to99_man_pop	0
100over_man_pop	0
Oto9_woman_pop	0
10to19_woman_pop	0
20to29_woman_pop	0
30to39_woman_pop	0
40to49_woman_pop	0
50to59_woman_pop	0
60to69_woman_pop	0

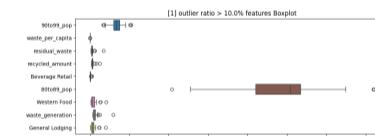
70to79_woman_pop	0
80to89_woman_pop	0
90to99_woman_pop	0
100over_woman_pop	0
P1	0
P2	0
P3	0
P4	0
P5	0
P6	0
P7	0
P8	0
P9	0
zero_restaurant	0
zero_re_cnt	0
area_km2	0
pop_density	0
cluster	0
waste_per_captia	0
children_rate	0
youth_rate	0
middle_rate	0
elderly_rate	0
zero_re_density	0
residual_waste	0
residual_ratio	0
Total_Boarding	0
Total_Alighting	0
num_businesses	0
num_employees	0
Beverage Retail	0
Buffet	0
Chinese Food	0
General Lodging	0
Japanese Food	0
Korean Food	0
Other Lodging	0
Pub	0
Western Food	0

dtype: int64

▼ df.columns

```
Index(['city', 'gu', 'waste_generation', 'recycled_amount', 'recycling_rate',
       'population_total', 'househole', 'population_per_household',
       'man_population', 'woman_population', 'man_woman_ratio', '0to9_pop',
       '10to19_pop', '20to29_pop', '30to39_pop', '40to49_pop', '50to59_pop',
       '60to69_pop', '70to79_pop', '80to89_pop', '90to99_pop', '100over_pop',
       '0to9_man_pop', '10to19_man_pop', '20to29_man_pop', '30to39_man_pop',
       '40to49_man_pop', '50to59_man_pop', '60to69_man_pop', '70to79_man_pop',
       '80to89_man_pop', '90to99_man_pop', '100over_man_pop', '0to9_woman_pop',
       '10to19_woman_pop', '20to29_woman_pop', '30to39_woman_pop',
       '40to49_woman_pop', '50to59_woman_pop', '60to69_woman_pop',
       '70to79_woman_pop', '80to89_woman_pop', '90to99_woman_pop',
       '100over_woman_pop', 'P1', 'P2', 'P3', 'P4', 'P5', 'P6', 'P7', 'P8',
       'P9', 'zero_restaurant', 'zero_re_cnt', 'area_km2', 'pop_density',
       'cluster', 'waste_per_captia', 'children_rate', 'youth_rate',
       'middle_rate', 'elderly_rate', 'zero_re_density', 'residual_waste',
       'residual_ratio', 'Total_Boarding', 'Total_Alighting', 'num_businesses',
       'num_employees', 'Beverage Retail', 'Buffet', 'Chinese Food',
       'General Lodging', 'Japanese Food', 'Korean Food', 'Other Lodging',
       'Pub', 'Western Food'],
      dtype='object')
```

▼ outlier with boxplot (10% 이상의 이상치만 표시)



1-2. features 간의 corr

- value 가 unique 한 features: city, P7
 - 두 feature 는 corr 에서 제외 (variance 가 0 이 되므로)

▼ continuous columns (cluster 을 제외)

```
Index(['0to9_man_pop', '0to9_pop', '0to9_woman_pop', '100over_man_pop',
       '100over_pop', '100over_woman_pop', '10to19_man_pop', '10to19_pop',
       '10to19_woman_pop', '20to29_man_pop', '20to29_pop', '20to29_woman_pop',
       '30to39_man_pop', '30to39_pop', '30to39_woman_pop', '40to49_man_pop',
       '40to49_pop', '40to49_woman_pop', '50to59_man_pop', '50to59_pop',
       '50to59_woman_pop', '60to69_man_pop', '60to69_pop', '60to69_woman_pop',
       '70to79_man_pop', '70to79_pop', '70to79_woman_pop', '80to89_man_pop',
       '80to89_pop', '80to89_woman_pop', '90to99_man_pop', '90to99_pop',
       '90to99_woman_pop', 'Beverage Retail', 'Buffet', 'Chinese Food',
       'General Lodging', 'Japanese Food', 'Korean Food', 'Other Lodging',
       'Pub', 'Total_Alighting', 'Total_Boarding', 'Western Food', 'area_km2',
       'children_rate', 'elderly_rate', 'househole', 'man_population',
       'man_woman_ratio', 'middle_rate', 'num_businesses', 'num_employees',
       'pop_density', 'population_per_household', 'population_total',
       'recycled_amount', 'recycling_rate', 'residual_ratio', 'residual_waste',
       'waste_generation', 'waste_per_captia', 'woman_population',
       'youth_rate', 'zero_re_cnt', 'zero_re_density'],
      dtype='object')
```

▼ binary columns

only 0/1: [P1, P2, P3, P4, P5, P6, P7, P8, P9, zero_restaurant]

▼ categorical columns

```
Index(['gu'], dtype='object')
```

▼ 기본 corr (추가된 열만)

▼ 추가된 열

```
['Total_Boarding',
 'Total_Alighting',
 'num_businesses',
```

```
'num_employees',
'Beverage Retail',
'Buffet',
'Chinese Food',
'General Lodging',
'Japanese Food',
'Korean Food',
'Other Lodging',
'Pub',
'Western Food']
```

▼ [cont-cont] (변경)

→ Total_Boarding & ≥ 0.5:

```
Total_Alighting 1.00
Korean Food 0.75
num_employees 0.73
num_businesses 0.72
Japanese Food 0.72
Beverage Retail 0.72
Buffet 0.65
Western Food 0.62
Pub 0.60
General Lodging 0.52
```

→ Total_Alighting & ≥ 0.5:

```
Total_Boarding 1.00
Korean Food 0.76
num_employees 0.74
Japanese Food 0.73
num_businesses 0.73
Beverage Retail 0.73
Buffet 0.66
Western Food 0.63
Pub 0.61
General Lodging 0.53
```

→ num_businesses & ≥ 0.5:

```
num_employees 0.97
Korean Food 0.92
Buffet 0.89
Beverage Retail 0.87
Japanese Food 0.81
Chinese Food 0.80
Total_Alighting 0.73
Total_Boarding 0.72
Pub 0.68
Western Food 0.62
recycling_rate -0.56
residual_ratio 0.56
elderly_rate -0.50
```

→ num_employees & ≥ 0.5:

```
num_businesses 0.97
Buffet 0.95
Beverage Retail 0.93
Korean Food 0.90
Japanese Food 0.84
Chinese Food 0.76
Total_Alighting 0.74
Total_Boarding 0.73
Western Food 0.72
Pub 0.65
elderly_rate -0.51
```

→ Beverage Retail & ≥ 0.5:

```
num_employees 0.93
num_businesses 0.87
Buffet 0.87
Korean Food 0.85
Japanese Food 0.84
Western Food 0.80
Total_Alighting 0.73
Total_Boarding 0.72
Chinese Food 0.61
Pub 0.60
elderly_rate -0.54
children_rate 0.53
area_km2 0.51
```

→ Buffet & ≥ 0.5:

```
num_employees 0.95
num_businesses 0.89
Beverage Retail 0.87
Korean Food 0.81
Japanese Food 0.77
Western Food 0.72
Chinese Food 0.69
Total_Alighting 0.66
Total_Boarding 0.65
Pub 0.60
pop_density -0.50
```

→ Chinese Food & ≥ 0.5:

```
num_businesses 0.80
Korean Food 0.78
num_employees 0.76
Pub 0.71
Japanese Food 0.69
Buffet 0.69
Beverage Retail 0.61
elderly_rate -0.54
```

→ General Lodging & ≥ 0.5:

```
Pub 0.64
Western Food 0.56
Japanese Food 0.55
Total_Alighting 0.53
Total_Boarding 0.52
60to69_man_pop -0.50
```

→ Japanese Food & ≥ 0.5:

```
Korean Food 0.91
Western Food 0.90
Pub 0.86
Beverage Retail 0.84
num_employees 0.84
num_businesses 0.81
Buffet 0.77
Total_Alighting 0.73
Total_Boarding 0.72
Chinese Food 0.69
elderly_rate -0.67
zero_re_cnt 0.58
```

General Lodging 0.55

→ Korean Food & ≥ 0.5:
num_businesses 0.92
Japanese Food 0.91
num_employees 0.90
Pub 0.88
Beverage Retail 0.85
Buffet 0.81
Chinese Food 0.78
Total_Alighting 0.76
Total_Boarding 0.75
Western Food 0.73
elderly_rate -0.53
area_km2 0.52
recycling_rate -0.51
residual_ratio 0.51

→ Other Lodging & ≥ 0.5:
20to29_woman_pop 0.53
20to29_pop 0.53
20to29_man_pop 0.52
youth_rate 0.51

→ Pub & ≥ 0.5:
Korean Food 0.88
Japanese Food 0.86
Western Food 0.72
Chinese Food 0.71
num_businesses 0.68
num_employees 0.65
General Lodging 0.64
Total_Alighting 0.61
Total_Boarding 0.60
Beverage Retail 0.60
Buffet 0.60
zero_re_cnt 0.51
elderly_rate -0.50

→ Western Food & ≥ 0.5:
Japanese Food 0.90
Beverage Retail 0.80
Korean Food 0.73
Buffet 0.72
num_employees 0.72
Pub 0.72
Total_Alighting 0.63
num_businesses 0.62
Total_Boarding 0.62
General Lodging 0.56
elderly_rate -0.54
zero_re_cnt 0.50

▼ [cont-bin] (변경)

→ num_businesses & ≥ 0.5:
P5 0.54

→ num_employees & ≥ 0.5:
P5 0.51

→ Beverage Retail & ≥ 0.5:
P5 0.59

→ Japanese Food & ≥ 0.5:
P5 0.52
zero_restaurant 0.51

→ Korean Food & ≥ 0.5:
P5 0.55

→ Other Lodging & ≥ 0.5:
P2 0.52
P1 0.52

→ Western Food & ≥ 0.5:
zero_restaurant 0.51

▼ [cont-cat] (변경)

→ num_businesses & ≥ 0.5:
gu_gangnam_gu 0.64

→ num_employees & ≥ 0.5:
gu_gangnam_gu 0.71

→ Beverage Retail & ≥ 0.5:
gu_gangnam_gu 0.77

→ Buffet & ≥ 0.5:
gu_gangnam_gu 0.66

→ Chinese Food & ≥ 0.5:
gu_yeongdeungpo_gu 0.64

→ General Lodging & ≥ 0.5:
gu_mapo_gu 0.75

→ Japanese Food & ≥ 0.5:
gu_gangnam_gu 0.72

→ Korean Food & ≥ 0.5:
gu_gangnam_gu 0.69

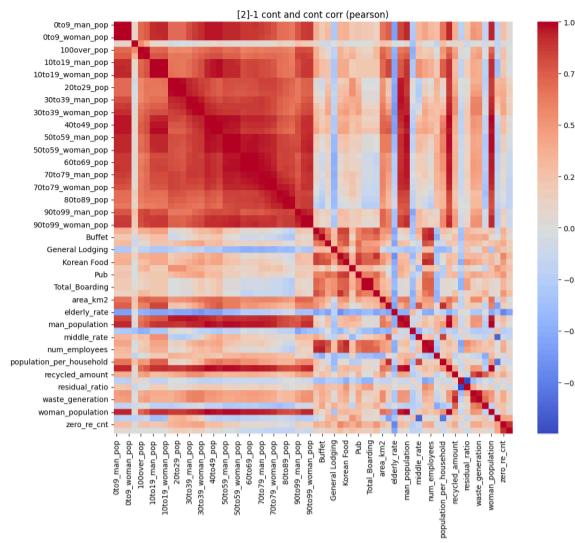
→ Other Lodging & ≥ 0.5:
gu_dongjak_gu 0.55
gu_gwanak_gu 0.55

→ Pub & ≥ 0.5:
gu_gangnam_gu 0.52

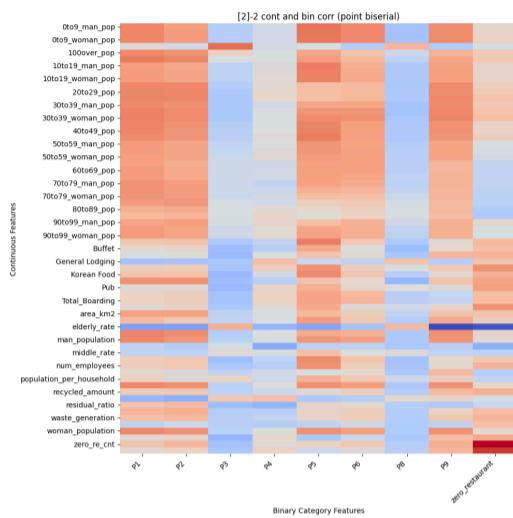
→ Western Food & ≥ 0.5:
gu_gangnam_gu 0.71

▼ corr visualization

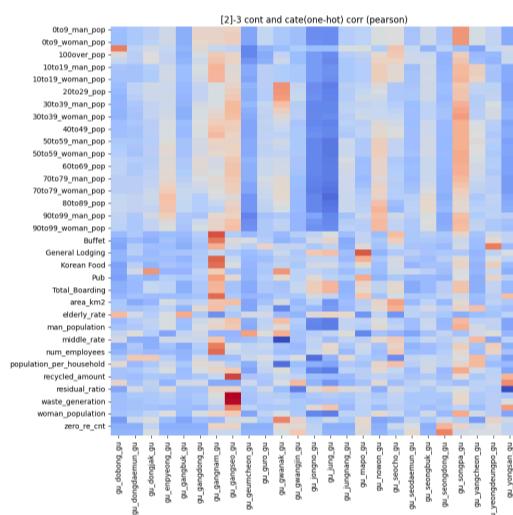
▼ cont & cont: pearson



▼ cont & binary: point biserial (biserial, polyserial)



▼ cont & cate(one-hot): pearson

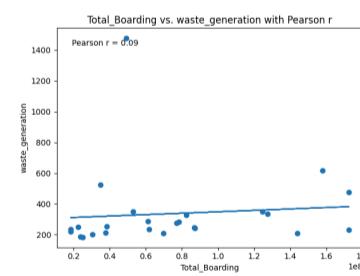


1-3. visualization

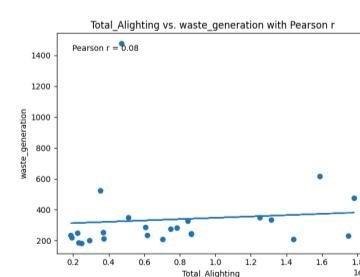
▼ 논의 사항

▼ 강서구, 용산구 쓰레기 많은데, 이유가 되는지 확인 (유동인구) (`Total_Boarding`, `Total_Aighting`)

▼ `Total_Boarding` & `waste_generation` (scatter + regplot) → `pearson = 0.09` (상관관계 없음)



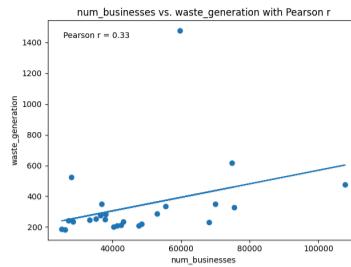
▼ `Total_Aighting` & `waste_generation` (scatter + regplot) → `pearson = 0.08` (상관관계 없음)



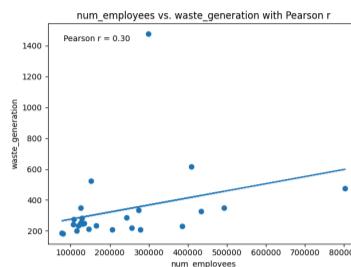
▼ ⇒ clustering 으로 시도

▼ `num_businesses`, `num_employees`: 쓰레기 배출량과 연관있는지 확인 (배달, 거주자 외 업무 중 생기는 쓰레기)

▼ `num_businesses` & `waste_generation` (scatter + regplot) → `pearson = 0.33` (상관관계 낮음)



▼ num_employees & waste_generation (scatter + regplot) → pearson = 0.30 (상관관계 낮음)



▼ ⇒ clustering 으로 시도

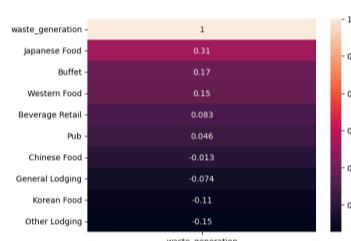
▼ zero rest

▼ extra visualization

▼ 상권 업종(retail_cols) & waste_generation corr (heatmap)

```
retail_cols = ['Beverage Retail', 'Buffet', 'Chinese Food',
               'General Lodging', 'Japanese Food', 'Korean Food', 'Other Lodging',
               'Pub', 'Western Food']
```

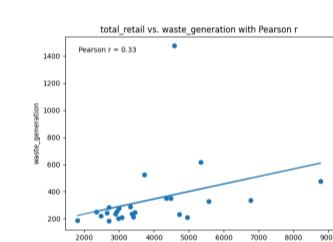
- 모든 업종이 상관계수가 높은건 아님 (업종 하나만으로 waste_generation 예측 X)



▼ 총 입종 (구별, total_retail) & waste_generation corr (regplot)

```
df['total_retail'] = df[retail_cols].sum(axis=1)
```

- corr 0.33 으로 낮음



1-4. Clustering

1. [Cluster1] pop & recycling rate (population_density & recycling_rate)

- ← waste_generation 추가 시 clustering 부정질
- ▼ add pop_density column using area_km2

```
df['pop_density'] = df['population_total'] / df['area_km2']
```

▼ Clustering

▼ cluster features: recycling_rate, pop_density

▼ cluster 기법: KMeans

▼ cluster 결과

Cluster 0 (초록)

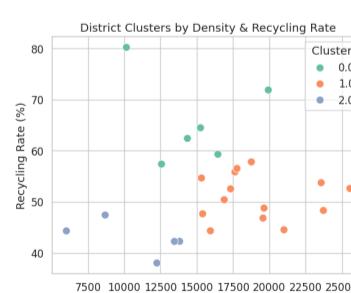
- 대체로 중간~높은 인구밀도에
- 높은 재활용률을 보이는 구들

Cluster 1 (주황)

- 가장 높은 인구밀도에
- 재활용률은 중간 정도인 구들

Cluster 2 (파랑)

- 낮은 인구밀도에
- 재활용률도 낮은 구들

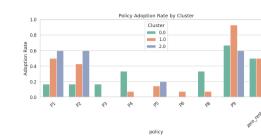
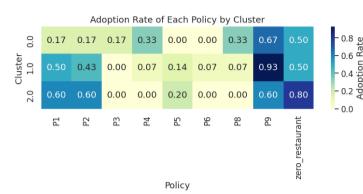


▼ cluster 결과 확인 지표 (적당한 cluster)

```
Silhouette Score: 0.367
Calinski-Harabasz Index: 17.8
Davies-Bouldin Index: 0.854
```

▼ Clustering insight

▼ policy (P7 제외)

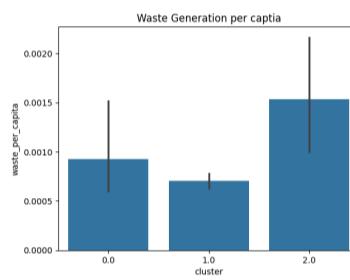


▼ 최종 분석 결과_[insight_policy]

Cluster ↓ \ Policy →	P1	P2	P3	P4	P5	P6	P8	P9	zero_restaurant
0 (중밀집·고재활용)	0.17	0.17	0.17	0.33	0.00	0.33	0.67	0.50	0.00
1(고밀집·중재활용)	0.50	0.43	0.00	0.07	0.14	0.07	0.07	0.93	0.50
2 (저밀집·저재활용)	0.60	0.60	0.00	0.00	0.20	0.00	0.00	0.60	0.80

- P8 (특수목분리수거)
 - cluster0 에서만 33% adoption → 고재활용에 높은 영향
 - 1,2 cluster 의 중, 저재활용 -> 확대해야할 정책 후보
- P4 (봉제원단재활용)
 - Cluster 0이 33%, Cluster 1:2 거의 0% 채택 → Cluster 0 재활용을 상승에 기여했을 가능성 큼
 - 1,2 cluster 의 중, 저재활용 -> 확대해야할 정책 후보
- P1, P2 (재활용정거장·무인화수거)
 - Cluster 2(저밀집·저재활용)에서 각각 60% 채택
 - Cluster 0(고성과)에서는 17%밖에 안 채택 ⇒ 단독 도입만으로는 재활용률 개선 효과가 미미한 듯
- P9 (회수보상제)
 - 거의 전 클러스터(60~93%)에서 도입되었지만, - Cluster 1(93%)은 중간 재활용률(약 55%) → P9만으로는 한계가 있는 정책
- zero_restaurant
 - Cluster 2에서 80%로 가장 높으나 재활용률은 제일 낮음 ⇒ "제로 식당" 정책 단독 도입만으로는 재활용률을 끌어올리지 못함

▼ waste_generation_captia



• anova 결과: ANOVA: F = 5.65, p = 0.010 (통계적 유의미)

▼ 최종 분석 결과_[insight_waste_generation_captia]

Cluster	waste_generation_captia
0 (중밀집·고재활용)	0.000929
1(고밀집·중재활용)	0.000707
2 (저밀집·저재활용)	0.001533

- Cluster0: 중간 1인당 배출량 (0.000929)
- Cluster1: 가장 낮은 1인당 배출량 (0.000707)
 - 인구 밀집도가 높고, 재활용 시스템이 적당히 잘 갖춰진 환경에서 1인당 배출을 최소화하고 있음으로 예상
- Cluster2: 주택 면적이 넓거나 소비 패턴이 도심과 달라서 쓰레기 발생이 큰 것으로 추정 (거의 Cluster1의 두배)

▼ pop_distribution

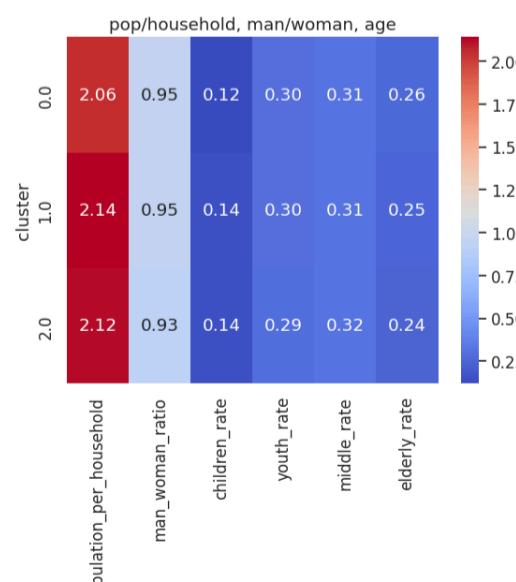
- age column

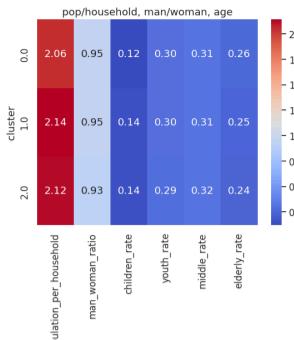
```
'10to19_pop', '20to29_pop', '30to39_pop', '40to49_pop', '50to59_pop',
'60to69_pop', '70to79_pop', '80to89_pop', '90to99_pop', '100over_pop'
```

- cluster 별 비교할 features

```
'population_per_household', 'man_woman_ratio',
'children_rate', 'youth_rate', 'middle_rate', 'elderly_rate'
```

cluster	population_per_household	man_woman_ratio	children_rate	youth_rate	middle_rate	elderly_rate
0.0	2.06	0.95	0.12	0.30	0.31	0.26
1.0	2.14	0.95	0.14	0.30	0.31	0.25
2.0	2.12	0.93	0.14	0.29	0.32	0.24





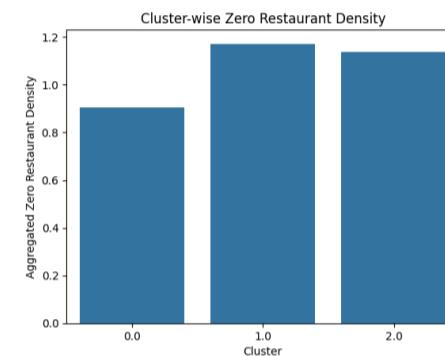
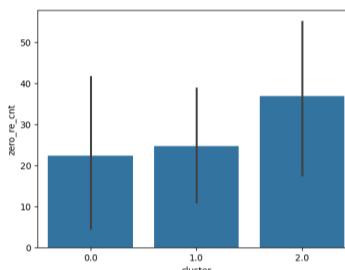
▼ 최종 분석 결과_[insight_waste_generation_captia]

지표	0 클러스터	1 클러스터	2 클러스터	최대-최소 차이
population_per_household	2.06	2.14	2.12	0.08
man_woman_ratio	0.95	0.95	0.93	0.02
children_rate	0.12	0.14	0.14	0.02
youth_rate	0.30	0.30	0.29	0.01
middle_rate	0.31	0.31	0.32	0.01
elderly_rate	0.26	0.25	0.24	0.02

- 별로 의미 없어보임 (분류 기준으로 보기 힘들듯)

▼ zero_restaurant_count, zero_re_density

cluster	area_km2	zero_re_cnt	zero_re_density
0.0	24.698333	22.333333	0.904245
1.0	21.088571	24.714286	1.171928
2.0	32.358000	36.800000	1.137277



▼ 최종 분석 결과_[insight_waste_generation_captia]

Cluster	area_km2	zero_re_cnt	zero_re_density
0 (중밀집·고재활용)	24.70	22.33	0.904245
1(고밀집·중재활용)	21.09	24.71	1.171928
2 (저밀집·저재활용)	32.36	36.80	1.137277

- Cluster 0(중밀집·고재활용)
 - 단위 면적당 식당이 상대적으로 적어, “제로 식당”이 미미하게 분포
- Cluster 1(고밀집·중재활용)
 - 면적 대비 식당 수가 특히 많아 밀도가 가장 높음
- Cluster 2(저밀집·저재활용)
 - 전체 수는 많지만 면적도 넓어, 밀도는 Cluster 1에 비해 약간 낮은 중간값

▼ 최종 분석 결과_policy별

정책 코드	정의	체택율 (Cluster 0 / 1 / 2)	관찰된 효과	권장 액션
P1	재활용정거장	0.17 / 0.50 / 0.60	- Cluster 0: 저체택률로 재활용률 최고 - Cluster 2: 높은 체택률로 재활용률 최저 → 단독효과 제한적	다른 정책(예: P4-P8)과 묶어 보완적 확대
P2	무인회수기	0.17 / 0.43 / 0.60	P1과 유사 양상: 단독 도입만으로 재활용률 개선에 한계	P1과 병행 도입, 접근성 높은 지역 우선 확대
P3	폐현수수막재활용	0.17 / 0.00 / 0.00	Cluster 0에서만 소량 체택, 효과 검증 불충분	시범사업 형태로 Cluster 1-2에 파일럿 확대 검토
P4	봉제원단 재활용	0.33 / 0.00 / 0.00	Cluster 0에서만 체택 → 재활용률을 증상 기여한 가능성 높음	Cluster 1-2 전면 확대 도입
P5	일회용컵보증금제	0.00 / 0.14 / 0.20	중·저재활용 clusters 만 일부 체택, 재활용률과 명확한 상관관계 미확인	추가 데이터 수집 후 효과 재평가
P6	디회용기도입	0.00 / 0.07 / 0.00	체택 극소량, 영향 평가 불가	우선순위 낮음, 필요시 타 정책과 패키지화
P8	특수품목분리수거	0.33 / 0.07 / 0.00	Cluster 0에서만 체택 → 고재활용률과 연관성 시사	Cluster 1-2 확대 도입
P9	재활용품회수보상제	0.67 / 0.93 / 0.60	전 클러스터에서 광범위 도입 → Cluster 1-2 재활용률 증·저위치 → 단독효과 한계	인센티브 강화 방안 검토, 보상 수준 조정
zero_restaurant	‘제로 식당’ (배출 저감 식당)	0.50 / 0.50 / 0.80	Cluster 2 체택 최다에도 재활용률 최저 → 단일 정책으로 효과 미흡	다른 분리수거·보상 정책과 연계 패키지화

▼ 최종 분석 결과_cluster 별 basic

Cluster (특성)	District	Avg. Waste per captia	Zero Restaurant Density	주요 정책 확장 후보	Pop. Distribution 차이 (의미 X)	District 특징 요약	획대·신규 도입할 정책	그 외 추천 프로그램·조치
0 (중밀집·고재활용)	[yongsan_gu, gwangjin_gu, gangbuk_gu, dobong_gu, nowon_gu, gwanak_gu]	0.000929	0.9042	—(현재 성과 유지)	미미 (최대-최소 0.08) (0.00093)	- 재활용률 최고 - 1인당 배출량 중간 - P5 신규 검토 (체택률 0.00) - P6 신규 검토 (체택률 0.00)	- P5 신규 검토 (체택률 0.00) - 제로 식당 추가 설치 - 포장재 감량·분리배출 교육	
1(고밀집·중재활용)	[seongdong_gu, dongdaemun_gu, jungnang_gu, seongbuk_gu, enpyeong_gu, seodaemun_gu, mapo_gu, yangcheon_gu, guro_gu, geumcheon_gu, yeongdeungpo_gu, dongjak_gu, songpa_gu, gangdong_gu]	0.000707	1.1719	P4, P8	미미 (최대-최소 0.08) (0.00071)	- 재활용률 중간 - 1인당 배출량 최저 - P4 확대 (체택률 0.07) - P6 신규 도입 (체택률 0.07) - P8 확대 (체택률 0.07)	- P3 신규 도입 (체택률 0.00) - 스마트 분리배출 알림 - 재활용 캠페인 (청년·맞벌이)	
2 (저밀집·저재활용)	[jongno_gu, jung_gu, gangseo_gu, seocho_gu, gangnam_gu]	0.001533	1.1373	P4, P8	미미 (최대-최소 0.08) (0.00153)	- 재활용률 최저 - 1인당 배출량 최고 - P3, P4, P6, P8 전면 확대 (체택률 0.00) - 신규 도입 (체택률 0.00)	- 감량형 요금제(RFID) (배출량 만큼만 수수료 지불하는 음식물쓰레기 종량제 시스템) - 음식물 감량 캠페인 - 소규모 무인 수거함 설치	

▼ 최종 분석 결과_cluster 별 policy

Cluster (특성)	District	P1	P2	P3	P4	P5	P6	P8	P9	zero_restaurant
0 (중밀집·고재활용)	yongsan_gu, gwangjin_gu, gangbuk_gu, dobong_gu, nowon_gu, gwanak_gu	17% → 현행 유지 (P4-P8 패키지 연계)	17% → 현행 유지 (접근성 보완)	0% → 우선순위 낮음(시범사업 불필요)	33% → 현행 유지	0% → 파일럿 검토	0% → 파일럿 검토	33% → 현행 유지	67% → 인센티브 구조 확장	50% → 추가 설치·교체
1(고밀집·중재활용)	seongdong_gu, dongdaemun_gu, jungnang_gu, seongbuk_gu, enpyeong_gu, seodaemun_gu, mapo_gu, yangcheon_gu, guro_gu, geumcheon_gu, yeongdeungpo_gu, dongjak_gu, songpa_gu, gangdong_gu	50% → 분리수거 페키지 병행	43% → 분리수거 페키지 병행	0% → 신규 도입 강력 권장(파일럿→전면)	7% → 즉시 확대 도입	14% → 데이터 추가 수집	7% → 신규 도입 검토	7% → 즉시 확대 도입	93% → 보상 수준 조정·정책 연계	50% → 패키지 포함
2 (저밀집·저재활용)	jongno_gu, jung_gu, gangseo_gu, seocho_gu, gangnam_gu	60% → 보완 필요	60% → 보완 필요	0% → 신규 도입 필수	0% → 신규 도입	20% → 시범사업 후 확대 검토	0% → 신규 도입 필수	0% → 신규 도입 검토	0% → 신규 도입 필수	80% → 분리수거 패키지 연계

2. [Cluster2] cluster_2 (Total_Boarding or Total_Alighting)

▶ Clustering

- cluster features: `Total_Boarding` or `Total_Alighting` (두 features 상관관계 1.00 이므로 둘 중 하나만)

- cluster 기법: KMeans

▼ cluster 결과

cluster	boarding_mean	alighting_mean	waste_mean	recycling_rate_mean	businesses_mean	employees_mean	waste_per_rider
0	7.30×10^7	7.27×10^7	273.33	55.94 %	42 736	182 987	1.88×10^{-6}
1	1.50×10^8	1.52×10^8	369.40	44.42 %	70 628	440 635	1.23×10^{-6}
2	2.98×10^7	2.95×10^7	373.95	55.49 %	37 163	150 473	6.31×10^{-6}

waste_per_rider = waste_mean ÷ (boarding_mean + alighting_mean)

▼ cluster 별 district (강서구, 용산구 cluster2)

gu
0 [dongjak_gu, enpyeong_gu, gangdong_gu, guro_gu, gwanak_gu, gwangjin_gu, nowon_gu, seongdong_gu, yeongdeungpo_gu]
1 [gangnam_gu, jongno_gu, jung_gu, mapo_gu, seocho_gu, songpa_gu]
2 [dobong_gu, dongdaemun_gu, gangbuk_gu, gangseo_gu, geumcheon_gu, jungnang_gu, seodaemun_gu, seongbuk_gu, yangcheon_gu, yongsan_gu]

▼ cluster 결과 확인 지표

Silhouette Score: 0.648
Calinski-Harabasz Index: 135.6
Davies-Bouldin Index: 0.407

▼ 통계적 검정

▼ insight_policy 채택률

군집	재활용정거장(P1)	무인회수기(P2)	폐현수막재활용(P3)	봉제원단재활용(P4)	일회용컵보증금제(P5)	다회용기도입(P6)	특수품목분리수거(P8)	재활용품회수보상제(P9)	제로식당 협업
0	56%	44%	0%	22%	11%	0%	0%	100%	56%
1	50%	50%	0%	17%	33%	17%	17%	67%	83%
2	30%	30%	10%	0%	0%	0%	20%	70%	40%

▼ 최종 결과

Cluster_2	특징	강점	개선 방안
0 (중간 이동량·중간 규모 허브)	- boarding, alighting: 약 73 M / 72 M (중간) - waste_generation: 가장 낮음 - recycling_rate: 가장 높음 - businesses, employees: 보통 - waste_per_rider: 중간	- 재활용품회수보상제(P9) 100% → 보상제 기반 홍보·참여율 우수 - 재활용정거장(P1) 56% - 무인회수기(P2) 44% → 거점수거장·회수기 어느 정도 구축	- 봉제원단재활용(P4) 22% - 일회용컵보증금제(P5) 11% → 업소 대상 설유·컵 제도 확대 시도- 다회용기도입(P6) 0% - 특수품목분리수거(P8) 0% → F&B 연관 소규모 프로그램 도입 검토 - 제로식당 협업 56% → 나머지 44% 제로식당 발굴 및 협업 강화
1(고이동량·대형 상업·환승 허브)	- boarding, alighting: 가장 높음 - waste_generation: 중간 이상 - recycling_rate: 가장 낮음 - business, employees: 가장 많음 - waste_per_rider: 가장 낮음 => 1인당 배출량은 낮은데, 절대적 배출량이 많음, 환승구역 곳곳에 재활용 전용 수거함 추가	- 제로식당 협업 83% - 재활용정거장(P1) 50% - 무인회수기(P2) 50% → 대형 허브 중심 정체 인프라 절반 이상 보유 - 일회용컵보증금제(P5) 33% - 다회용기도입(P6) 17% → 대규모 시설에서 시범 시행 중	- 봉제원단재활용(P4) 17% → 설유업소 집중 수거함 추가 - 특수품목분리수거(P8) 17% → 소형가전 등 회수 확대 - 재활용품회수보상제(P9) 67% → 나머지 33% 구역 보상제 매핑·확대
2 (저이동량·주거 밀집 소규모역)	- boarding, alighting: 가장 낮음 - waste_generation: 가장 높음 - recycling_rate: 높은편 - business, employees: 가장 낮음 - waste_per_rider: 가장 높음 => 지하철 이용객 수는 적지만 1인당 배출량이 매우 높음	- 폐현수막재활용(P3) 10% - 특수품목분리수거(P8) 20% → 일부 업사이클링·재활용품 관리 시작 - 재활용품회수보상제(P9) 70% → 주민 참여 보상제 활용 가능	- 재활용정거장(P1) 30%- 무인회수기(P2) 30% → 기본 분리수거 인프라 대폭 확충 필요 - 봉제원단재활용(P4) 0% - 일회용컵보증금제(P5) 0% - 다회용기도입(P6) 0% → 가정·소규모 대상 시범 프로그램 기획 - 제로식당 협업 40% → 주민 대상 제로식당 프로모션·안내 캠페인

3. [Cluster3] cluster_3 (num_businesses , num_employees , total_retail)

▼ Clustering

- cluster features: num_businesses, num_employees, total_retail
- cluster 기법: KMeans

▼ cluster 결과

cluster_3	num_businesses_mean	num_employees_mean	total_retail_mean	waste_generation_mean	recycling_rate_mean	zero_restaurant_rate
0.0	47,947.86	227,613.57	3,598.86	407.00	49.79%	43%
1.0	75,274.33	466,440.50	5,948.00	389.03	46.15%	83%
2.0	32,728.67	116,957.25	2,961.75	269.39	58.29%	50%

▼ cluster 별 district

gu
0 [dongdaemun_gu, gangdong_gu, gangseo_gu, geumcheon_gu, guro_gu, jongno_gu, seongdong_gu]
1 [gangnam_gu, jung_gu, mapo_gu, seocho_gu, songpa_gu, yeongdeungpo_gu]
2 [dobong_gu, dongjak_gu, enpyeong_gu, gangbuk_gu, gwanak_gu, gwangjin_gu, jungnang_gu, nowon_gu, seodaemun_gu, seongbuk_gu, yangcheon_gu, yongsan_gu]

▼ cluster 결과 확인 지표

Silhouette Score: 0.323
Calinski-Harabasz Index: 26.5
Davies-Bouldin Index: 0.910

▼ 통계적 검정

▼ insight_policy 채택률

cluster_3	P1	P2	P3	P4	P5	P6	P8	P9	제로식당 협업
0.0	0.57	0.43	0.00	0.00	0.14	0.00	0.00	0.86	0.43
1.0	0.50	0.50	0.00	0.17	0.33	0.17	0.17	0.83	0.83
2.0	0.33	0.33	0.08	0.17	0.00	0.00	0.17	0.75	0.50

▼ 최종 결과

Cluster_3	특징	강점	개선 방안
0.0(소상공인 밀집 상권)	- 회사 수·회사원 수·총 업종 수: 중간 - 폐기물량: 최고(407) - 재활용률: 중간(49.8%)	- 재활용품회수보상제(P9): 86% - 재활용정거장(P1): 56% - 무인회수기(P2): 44%	- 일소 대상 음식물 수거 스케줄 조정 - P1·P2 인프라 추가 보강 - P3·P4·P5·P6·P8 시범사업 기획 - 제로식당 협업 확대
1.0(상업·업무 중심지)	- 회사 수·회사원 수·총 업종 수: 최고 - 폐기물량: 2위(389) - 재활용률: 최저(46.2%)	- 제로식당 협업: 83% - 재활용정거장(P1): 50% - 무인회수기(P2): 50%	- P1(P2)(재활용정거장) 확대 설치 - P2(무인회수기)·P8(특수품목) 추가 배치 - 제로식당 협업 모델 확장 및 우수사례 전파
2.0(주택·주거 위주)	- 회사 수·회사원 수·총 업종 수: 최저 - 폐기물량: 최저(269) - 재활용률: 최고(58.3%)	- 재활용률: 58.3% - 재활용품회수보상제(P9): 75%	- P1·P2(기본 분리수거 인프라) 대폭 확충 - P3·P4·P5·P6 소규모 시범사업 기획 - 제로식당 협업 홍보·프로모션 강화

4. [Cluster4] Cluster_4 (total_boarding , total_alighting , num_businesses , num_employees)

▼ Clustering

- cluster features: total_boarding, total_alighting, num_businesses, num_employees
- cluster 기법: KMeans

▼ cluster 결과

cluster	boarding_mean	alighting_mean	businesses_mean	employees_mean	waste_mean	recycling_rate_mean
0	6.93×10^7	6.86×10^7	40 979.78	168 108.11	401.33	54.57 %
1	1.34×10^8	1.37×10^8	81 996.75	534 454.25	441.98	47.80 %
2	2.77×10^7	2.75×10^7	34 662.89	133 934.00	251.24	56.96 %
3	1.48×10^8	1.50×10^8	57 084.67	313 076.33	258.27	43.34 %

▼ cluster 별 district

gu
0 [dongjak_gu, enpyeong_gu, gangdong_gu, gangseo_gu, guro_gu, gwanak_gu, gwangjin_gu, nowon_gu, seongdong_gu]
1 [gangnam_gu, seocho_gu, songpa_gu, yeongdeungpo_gu]

2 [dobong_gu, dongdaemun_gu, gangbuk_gu, geumcheon_gu, jungnang_gu, seodaemun_gu, seongbuk_gu, yangcheon_gu, yongsan_gu]
 3 [jongno_gu, jung_gu, mapo_gu]

▼ cluster 결과 확인 지표

Silhouette Score: 0.339
 Calinski-Harabasz Index: 29.0
 Davies-Bouldin Index: 1.027

▼ 통계적 검정

▼ insight_policy 채택률

cluster	P1	P2	P3	P4	P5	P6	P8	P9	제로식당 협업
0	67 %	56 %	0 %	22 %	11 %	0 %	0 %	100 %	56 %
1	75 %	75 %	0 %	0 %	50 %	25 %	0 %	100 %	100 %
2	22 %	22 %	11 %	0 %	0 %	0 %	22 %	67 %	33 %
3	0 %	0 %	0 %	33 %	0 %	0 %	33 %	33 %	67 %

▼ 최종 결과

cluster_4	특징	강점	개선 방안
0	- 승·하차·업종 규모 모두 중간 - 폐기물량 높은(401) - 재활용률 높음(54.6 %)	- 재활용률 회수보상제(P9) 100 % - 재활용정거장(P1) 67 % - 무인회수기(P2) 56 % 설치	- P4-P8-P6 미도입 → 다회용기·특수품목 분리수거 시범 도입 - P5(11 %) 소량 → 요금제 시범 확대- 분리수거 인프라(P1/P2) 보강
1	- 승·하차·업종 규모 최고 - 폐기물량 최고(442) - 재활용률 낮음(47.8 %)	- P1-P2 양호(75 %)- P5(50 %)-P6(25 %) 시범 진행 중 - 제로식당 협업 100 %	- P4-P8 미도입 → 복제원단·특수품목 수거함 설치 - 재활용률 개선 위해 분리수거 가정 추가 - 성공사례 전파를 위한 우수 모델 공유
2	- 승·하차·업종 규모 최저 - 폐기물량 최저(251) - 재활용률 높음(57.0 %)	- 재활용률 56.96 % - P9 67 % 활용 → 주민 보상제 참여도 양호	- P1-P2 설치율 낮음(22 %) → 기본 분리수거 인프라 대폭 확충 - P5-P6 미시행 → 소규모 시범사업(감량형 요금제-다회용기) 추진 - 제로식당 홍보 확대
3	- 승·하차량 두 번째 최고(148 M) - 업종·직원 증간 수준 - 폐기물량 증간 이하(258) - 재활용률 최저(43.3 %)	- P4-P8-P9 각각 33 % 설치 → 일부 특수품목·수거보상 제도 적용 - 제로식당 협업 67 %	- P1-P2-P3 미도입 → 거점수거장·무인회수기·업사이클링 프로그램 설치 - 재활용률 저하 대응 위해 분리배출 안내 강화 - 스마트 빈 도입

5. [Cluster5] Cluster_5 (total_boarding , total_alighting , retail_cols)

▼ Clustering

- cluster features: total_boarding , total_alighting , retail_cols
- cluster 기법: KMeans

▼ cluster 결과

cluster	boarding_mean	alighting_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	zero_restaurant_rate	Beverage_Retail	Buffet	Chinese_Food	General_Lodging	Japanese_Food	Korean_Food	Other_Lodging	Pub
0.0	61,326,745.5	60,914,950.0	610.88	0.00	51.26	75%	24.50	49.00	342.50	219.50	287.75	2,333.25	156.50	864.25
1.0	149,632,975.5	150,553,900.0	351.25	0.00	44.15	75%	53.25	66.50	230.75	381.25	357.25	2,523.00	137.00	764.75
2.0	47,322,908.8	47,041,260.0	250.40	0.00	56.89	40%	18.20	31.13	156.53	138.40	180.73	1,567.27	94.80	536.73
3.0	149,815,188.5	154,810,200.0	405.70	0.00	44.96	100%	78.50	86.00	375.50	645.50	795.50	3,527.00	165.50	1,258.00

▼ cluster 별 district

gu
 0 [gangseo_gu, guro_gu, gwanak_gu, yeongdeungpo_gu]
 1 [jongno_gu, jung_gu, seocho_gu, songpa_gu]
 2 [dobong_gu, dongdaemun_gu, dongjak_gu, enpyeong_gu, gangbuk_gu, gangdong_gu, geumcheon_gu, gwangjin_gu, jungnang_gu, nowon_gu, seodaemun_gu, seongbuk_gu, seongdong_gu, yangcheon_gu, yongsan_gu]
 3 [gangnam_gu, mapo_gu]

▼ cluster 결과 확인 지표

Silhouette Score: 0.286
 Calinski-Harabasz Index: 12.6
 Davies-Bouldin Index: 1.203

▼ 통계적 검정

▼ insight_policy 채택률

cluster	P1	P2	P3	P4	P5	P6	P8	P9	제로식당 협업
0.0	50%	50%	0%	0%	0%	0%	0%	100%	75%
1.0	50%	50%	0%	0%	25%	25%	0%	50%	75%
2.0	40%	33%	7%	13%	7%	0%	13%	80%	40%
3.0	50%	50%	0%	50%	50%	0%	50%	100%	100%

▼ 최종 결과

cluster_5	특징	강점	개선 방안
0.0	- 승·하차·업종 규모 중간- 폐기물량 최고(610.9) - 재활용률 증간(51.3%)- 제로식당 협업 75% - 주력 업종: Buffet-Pub 등 F&B 비율 높음	- 제로식당 협업 75% 확보- P9(보상제) 100% - P1-P2(거점수거장·무인회수기) 50% 도입	- P3-P4-P6-P8(업사이클링·다회용기·특수품목) 시범 도입 - P5(일회용컵 보증금제) 시범 확대 - 분리수거 인프라(P1/P2) 추가 보강
1.0	- 승·하차·업종 규모 높음- 폐기물량 중간(351.3) - 재활용률 낮음(44.2%) - 제로식당 협업 75% - 주력 업종: Buffet-Korean Food	- P1-P2 50% 도입 - P5-P6(요금제-다회용기) 25% 시범 - 제로식당 협업 75%	- P3-P4-P8(업사이클링·봉제원단·특수품목) 프로그램 도입 - 재활용정거장 확대 설치 - 제로식당 협업 고도화
2.0	- 승·하차·업종 규모 최저 - 폐기물량 최저(250.4) - 재활용률 높음(56.9%) - 제로식당 협업 40% - 주거 밀집-소규모 상권 특징	- 높은 재활용률(56.9%) - P9(보상제) 80% 활용	- 분리수거 인프라(P1-P2) 대폭 확충 - 소규모 대상 P5-P6(감량형 요금제-다회용기) 시범 - 제로식당 협업 확대
3.0	- 승·하차·업종 규모 가장 높음- 폐기물량 높음(405.7) - 재활용률 가장 낮음(44.96%) - 숙박·관광 업종 물류(Lodging-F&B 비중↑) - 제로식당 협업 100%	- P9(보상제) 100% - P4-P8 50% 도입(봉제원단·특수품목) - 제로식당 협업 100%	- P1-P2-P3(거점수거장·무인회수기·업사이클링) 설치 - 분리배출 안내 강화 - 고빈도 환승구역용 스마트 빈 도입

6. [Cluster6] Cluster_6 (poulation_per_household , household , area_km2)

▼ Clustering

- cluster features: poulation_per_household , household , area_km2
- cluster 기법: KMeans

▼ cluster 결과

cluster_6	boarding_mean	alighting_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	zero_restaurant_rate	pop_per_household_mean	house_hold_mean	area_km2_mean	Beverage_Retail	Buffet	Chinese_Food	General_Lodging	Jap
0.0	6.3420×10^7	6.3771×10^7	254.70	0.00	54.44	0.50	2.09	154,480.72	19.37	23.50	38.33	201.39	250.28	232
1.0	1.2505×10^8	1.2710×10^8	369.73	0.00	50.71	0.67	2.33	207,810.00	40.64	70.33	80.00	285.00	164.67	498
2.0	8.4149×10^7	8.2988×10^7	680.52	0.00	48.19	0.75	2.08	264,130.25	33.65	26.75	42.75	228.00	192.75	302

▼ cluster 별 district

gu
 0 [dobong_gu, dongdaemun_gu, dongjak_gu, gangbuk_gu, gangdong_gu, geumcheon_gu, guro_gu, gwangjin_gu, jongno_gu, jung_gu, jungnang_gu, mapo_gu, seodaemun_gu, seongbuk_gu, seongdong_gu, yangcheon_gu, yongsan_gu]
 1 [gangnam_gu, nowon_gu, seocho_gu]
 2 [enpyeong_gu, gangseo_gu, gwanak_gu, songpa_gu]

▼ cluster 결과 확인 지표

Silhouette Score: 0.357
 Calinski-Harabasz Index: 11.6
 Davies-Bouldin Index: 1.080

▼ 통계적 검정

▼ insight_policy 채택률

cluster_6	P1	P2	P3	P4	P5	P6	P8	P9	zero_restaurant
0.0	0.28	0.22	0.06	0.11	0.06	0.00	0.17	0.72	0.50
1.0	0.67	0.67	0.00	0.33	0.33	0.00	0.00	1.00	0.67
2.0	1.00	1.00	0.00	0.00	0.25	0.25	0.00	1.00	0.75

▼ 최종 결과

cluster_6	특징	강점	개선 방안
0.0	- 낮은 승-하차량 (~63 M) - 최저 폐기율량 (254.7) - 최고 재활용률 (54.44%) - 중간 1가구당 인구(2.09) - 작은 가구 수·면적(154 K 가구, 19 km ²) - 제로식당 협업 50%	- 이미 재활용 참여도가 매우 높음 - 소규모 지역 관리 용이 - P9(보상제) 도입률 72%	- 제로식당 협업 확대(→75% 목표) - P1-P2(거점수거·무인회수기) 보강 - 소규모 주거지 대상 분리배출 홍보 강화
1.0	- 가장 높은 승-하차량 (-125 M) - 중간 폐기율량 (369.7) - 중간 재활용률 (50.71%) - 높은 1가구당 인구(2.33) - 넓은 면적(40.6 km ²) - 제로식당 협업 67%	- 유동인구 활용한 대규모 캠페인 가능 - P1-P2 채택률 67% - P9 100% 도입 완료	- P3-P4-P8(업사이클링·다회용기·특수품목) 시범 확대 - P5-P6 요금제·다회용기 보급 강화- 핵심 상권 스마트 빈 설치
2.0	- 중간 승-하차량 (-84 M) - 최고 폐기율량 (680.5) - 최저 재활용률 (48.19%) - 중간 1가구당 인구(2.08) - 넓은 가구 수(264 K) - 제로식당 협업 75%	- 활발한 제로식당 파트너십(75%) - P1-P2-P9 100% 도입 가능 - 광역 대상 집중 관리 용이	- 재활용률 제고 교육·인센티브 강화 - 폐기율 감량형 요금제(P5) 파일럿 도입 - 분리수거 인프라 집중 확충

7. [Cluster7] Cluster_7 (man_woman_ratio, children_rate, youth_rate, middle_rate, elderly_rate)

▼ Clustering

- cluster features: man_woman_ratio, children_rate, youth_rate, middle_rate, elderly_rate

- cluster 기법: KMeans

- cluster 결과

cluster_7	boarding_mean	alighting_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	zero_restaurant_rate	man_woman_ratio_mean	children_rate_mean	youth_rate_mean	middle_rate_mean	Beverage Retail	Buffet	Chinese Food	G L
0.0	65,461,086.78	65,332,597.67	222.77	0.00	49.64	0.11	0.96	0.12	0.29	0.31	20.56	37.78	178.89	2
1.0	93,662,417.43	94,480,264.86	351.77	0.00	52.18	0.57	0.94	0.16	0.27	0.33	48.43	56.29	228.43	1
2.0	67,614,021.11	67,973,252.44	438.73	0.00	56.98	1.00	0.94	0.13	0.33	0.30	24.11	40.78	242.56	3

- cluster 별 district

gu
0 [dobong_gu, dongdaemun_gu, empyeong_gu, gangbuk_gu, geumcheon_gu, guro_gu, jongno_gu, jung_gu, jungnang_gu]
1 [gangdong_gu, gangnam_gu, nowon_gu, seocho_gu, seongbuk_gu, songpa_gu, yangcheon_gu]
2 [dongjak_gu, gangseo_gu, gwanak_gu, gwangjin_gu, mapo_gu, seodaemun_gu, seongdong_gu, yeongdeungpo_gu, yongsan_gu]

- cluster 결과 확인 지표

Silhouette Score: 0.295
Calinski-Harabasz Index: 11.9
Davies-Bouldin Index: 1.08

- 통계적 검정

▼ insight_policy 채택률

cluster_7	P1	P2	P3	P4	P5	P6	P8	P9	zero_restaurant
0.0	0.22	0.22	0.11	0.00	0.00	0.22	0.44	0.11	
1.0	0.71	0.57	0.00	0.14	0.43	0.14	0.00	1.00	0.57
2.0	0.44	0.44	0.00	0.22	0.00	0.11	1.00	1.00	

▼ 최종 결과

cluster_7	특징	강점	개선 방안
0.0	- 인구 중장년층 비율↑ (middle_rate 31%) - 어린이 비율 낮음(12%), 청년 29% - 남녀비 균형(0.96) - 낮은 유동인구(boarding 65M) - 재활용률 낮음(49.64%)	- 소규모 주거지 특성으로 집중 관리 용이 - P9(보상제) 도입률 44% 시작 단계	- 재활용 참여 캠페인 강화(특히 중장년 대상 및 출현 홍보) - P1-P2 인프라 도입 확대 - 어린이·청년 대상 교육 프로그램 운영
1.0	- 청년층 비율 증간(27%), 중장년 33% - 어린이 비율 16% - 남녀비 0.94 (약간 여성 많음) - 매우 높은 유동인구(boarding 94M) - 재활용률 증간(52.18%)	- 유동인구 많은 상업·관광지로 캠페인 효과↑- P1(71%), P2(57%), P5(43%) 채택률 높음	- P3-P8(업사이클링·특수품목) 시범 확대 - 다회용기(P6) 보급 강화 - 핵심 상권 스마트 분리수거함 설치
2.0	- 청년 비율 높음(33%), 중장년 30% - 어린이 비율 13% - 남녀비 0.94 - 중간 유동인구(boarding 68M) - 높은 재활용률(56.98%) - 제로식당 100% 도입	- 모범적인 재활용률(56.98%) 달성 - 제로식당 협업 원천 구축	- 폐기율 감량형 요금제(P5) 파일럿 시행 - P1-P2 인프라 대폭 확충 - 청년층 대상 재활용 챌린지 프로그램 기획

▼ 최종 데이터 클러스터별 분석 결과 (cluster1 결과 변경)

▼ cluster1 (변경)

	cluster_1	boarding_mean	alighting_mean	businesses_mean	employees_mean	waste_mean	recycling_rate_mean	pop_density_mean	P1	P2	P3	P4	P5	P6	P8
0	0.0	5.481897e+07	5.476494e+07	42901.91	184165.91	268.99	51.64	20135.61	0.55	0.45	0.00	0.00	0.18	0.09	0.00
1	1.0	5.485885e+07	5.478078e+07	37568.71	161153.57	300.17	64.36	14866.86	0.14	0.14	0.14	0.29	0.00	0.00	0.29
2	2.0	1.237560e+08	1.250335e+08	63589.00	377360.71	479.40	43.76	12227.93	0.57	0.57	0.00	0.14	0.00	0.00	0.14

▼ cluster2

	cluster_2	boarding_mean	alighting_mean	waste_mean	recycling_rate_mean	businesses_mean	employees_mean	P1	P2	P3	P4	P5	P6	P8	P9
0	0.0	7.301615e+07	7.273185e+07	273.33	55.94	42736.44	1								

0	0.0	6.342009e+07	6.377076e+07	254.70	0.0	54.44	0.50	2.09		154480.72	19.37	23.50	38.33	201.39	251
1	1.0	1.250518e+08	1.270958e+08	369.73	0.0	50.71	0.67	2.33		207810.00	40.64	70.33	80.00	285.00	162
2	2.0	8.414898e+07	8.298838e+07	680.52	0.0	48.19	0.75	2.08		264130.25	33.65	26.75	42.75	228.00	192

▼ cluster7

	cluster_7	boarding_mean	alighting_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	zero_restaurant_rate	man_woman_ratio_mean	children_rate_mean	youth_rate_mean	middle_rate_mean	Beverage_Retail	Buffet
0	0.0	65461086.78	65332597.67	222.77	0.0	49.64	0.11	0.96	0.12	0.29	0.31	20.56	37.78
1	1.0	93662417.43	94480264.86	351.77	0.0	52.18	0.57	0.94	0.16	0.27	0.33	48.43	56.29
2	2.0	67614021.11	67973252.44	438.73	0.0	56.98	1.00	0.94	0.13	0.33	0.30	24.11	40.78

▼ 보고서

▼ 1. 과제 개요 (150자)

| 분석과제의 구성 및 제안 배경

| 수행한 분석과제의 목표, 활용 대상, 법령, 기대효과를 작성

- 본 과제는 서울시 25개 자치구의 폐기물 배출량, 유동인구(지하철 승·하차), 사업체·종사자 수, 상권 특성을 종합 분석하여, 구별 재활용률 제고 및 폐기물 저감 맞춤형 정책을 제안하는 것을 목표로 한다.

▼ 2. 활용 데이터 및 분석도구

| 분석 결과를 도출하기 위해 활용한 데이터 및 분석 도구의 정보

| (데이터) 보유 시스템(운영기관), 데이터 형태(형식), 내용, 규모, 특징, 수집 방안, 비용 소요 여부 등

| (분석 도구) 분석 도구(언어), 내용, 특징, 제작자(기업), 사용 버전(유료, 비전), 비용 소요 여부 등

- 데이터 (기존 데이터와 area_km2 데이터)
 - https://stat.esouls.go.kr/statHtml/statHtml.do?orgId=201&tblId=DT_201004_0010001&conn_path=l2&mode=tab&obj_var_id=&up_itm_id=

| 분석 도구

- 언어/플랫폼: Python (Colab)

- 주요 라이브러리: Pandas, Numpy, Scikit-learn, Seaborn, Matplotlib

▼ 3. 분석내용

| 데이터 전처리 및 탐색적 데이터 분석, 분석 모델링 선정 방안 등

▼ 데이터 전처리

- null 확인 및 outlier 탐지
- 파생 변수 생성: pop_density, waste_per_capita, children_rate, youth_rate, middle_rate, elderly_rate, zero_re_density, residual_waste, residual_ratio, total_retail, waste_per_rider

▼ EDA

- basic: info(), describe()

- feature importance

▼ features corr & visualization

- cont-cont: pearson
- cont-bin: point biserial
- cont-cate: one-hot & pearson
- bin-bin: phi(cramer's V)
- bin-cate: cramer's V

▼ visualization

- barplot (policy & recycling rate)
- barplot (gu & recycling rate)
- histplot(distribution of recycling rate)
- qcut & boxplot (Total_Boarding, Total_Alighting & waste_generation)

▼ 분석 모델링 선정 방안

- 알고리즘: K-Means
- 사용 클러스터링 (EDA2 참고)
- 클러스터링 별 조합
- 클러스터링 품질 평가 지표: Silhouette Score, Calinski-Harabasz Index, Davies-Bouldin Index
- 클러스터링 별 프로파일링: 군집별 주요 특징 평균 값 확인

▼ 4. 창의성

| 분석과제의 차이점 및 독창적인 특징

- 도메인별 클러스터링 비교: 단일 변수에 그치지 않고 여러 도메인으로 여러 클러스터링 생성 및 비교

| 규모 보정 지표 도입: pop_density, waste_per_capita, zero_re_density, waste_per_rider,

- df['pop_density'] = df['population_total'] / df['area_km2']
- df['waste_per_capita'] = df['waste_generation'] / df['population_total']
- df['zero_re_density'] = df['zero_re_cnt'] / df['area_km2']
- df['waste_per_rider'] = (df['waste_generation'] / (df['Total_Boarding'] + df['Total_Alighting']))

- 정책 채택률을 융합 분석: 각 클러스터별 별 정책 채택률을 바탕으로 정책 우선순위 도출

- 여러 데이터 병합

▼ 5. 적합성

| 분석 수행 시 활용한 데이터의 융합 및 활용성

| 분석 과정 및 사용 모델의 내용에 대한 구체적인 설명

| (분석모델의 구현 예시 시각화 결과 및 설명 작성)

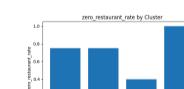
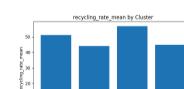
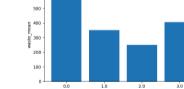
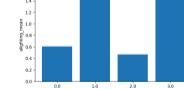
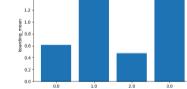
- 데이터 응합: 재활용률 · 폐기물 · 유동인구 · 사업체 · 상권 · 정책 데이터를 통합

▼ 모델 선택 근거

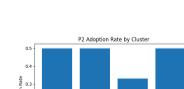
- 비지도 학습(K-Means)은 사전 레이블 없이 세그먼트 식별에 적합
- 각 지표로 객관적 군집 품질 평가 가능

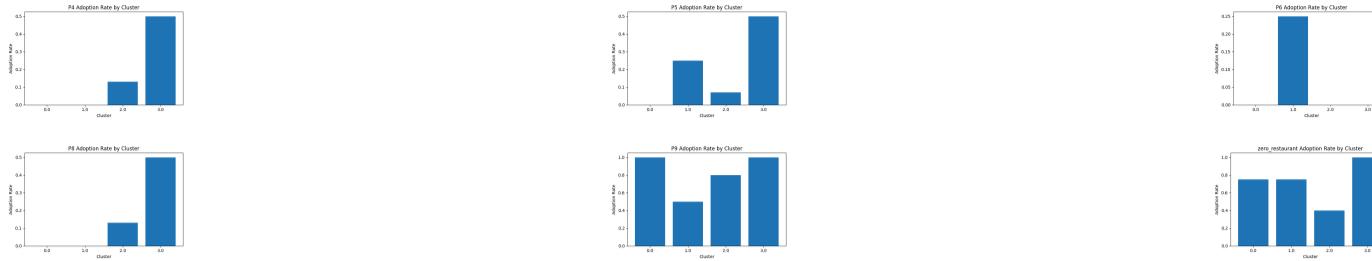
- 재현성: Colab 에 분석 워크플로우 기록, 누구나 재실행 가능

- 시각화 예시1 (cluster5의 클러스터별 주요 feature의 mean 값 barplot)



- 시각화 예시2 (cluster5 의 클러스터별 정책 채택률 barplot)





▼ 6. 활용성

- | 해당 분석모델을 업무수행에 어떻게 활용할 수 있는지
- | 구체적인 업무 활용 계획(활용 대상, 소요 예산, 절차 등)을 구체적으로 작성
- 업무 프로세스:
 1. 연 1회 데이터 갱신 → 자동화 스크립트 실행
 2. 신규 클러스터·프로파일 산출 → 보고서
 - 예산·절차:
 - 시스템: 무료 오픈소스 활용, 추가 비용 X
 - 활용 대상: 서울시 자원순환과, 각 구청 환경과

▼ 7. 정책 활용

- | 분석 결과가 어떻게 정책에 적용될 수 있는지
- | 관련된 법령·고사·지침 등을 구체적으로 기재

- 각 클러스터에 필요한 정책은 [⑥. 기대효과](#) 의 모델 결과로 직접 확인 (수치적 근거)

▼ 정책 관련 법령·고사·지침

법령·조항	정책 코드	URL	근거 설명
자원순환기본법 제5조(국가 및 지방자치단체의 책무)	전반(기획·홍보·교육)	https://faolex.fao.org/docs/pdf/kor232114.pdf	자원순환 전환을 위한 시책 수립·홍보·교육 프로그램 의무
자원순환기본법 제7조(국민의 책무)	전반(분리배출, 회용자재)	https://www.lawnb.com/Info/ContentView?sid=L0004EA923938C5F_7	1회용품 사용 자체·분리배출 의무 부과
자원순환기본법 제8조(자원순환사회로의 발전을 위한 문화 조성)	전반(문화조성)	https://faolex.fao.org/docs/pdf/kor232114.pdf	시민 대상 순환경체 문화 조성 사업 근거
폐기물관리법 제13조의2(재활용센터의 설치·운영 등)	P1(재활용정거장)	https://www.easylaw.go.kr/CSP/CnpClMain.laf?popMenu=ovk&cmSeq=1452&ccfNo=3&cclNo=1&cnpCisNo=1&search_put_=	지자체·사업자의 재활용정거장 설치·운영 근거
폐기물관리법 제14조(생활폐기물의 처리 등)	전반(수수료 부과)	https://law.go.kr/법령/폐기물관리법/제14조	생활폐기물 처리 수수료 부과 및 관리 근거 (부과 수수료 부과 근거)
폐기물관리법 제16조(제조업자 등의 재활용의무)	P4(봉제원단재활용)	https://elaw.kiri.re.kr/kor_service/lawViewTitle.do?hseq=33581	섬유류 제조업자에 재활용 계획 수립·이행 의무 부과
폐기물관리법 시행규칙 제15조의2(생활폐기물을 수집·운반·재활용)	P6(특수품목분리수거)	https://www.nhis.or.kr/lm/lmxsr/law/lawFullView.do?SEQ=913&SEQ_HISTORY=28086	특수·섬유류를 별도 수거·운반·재활용하도록 규정
자원의 절약과 재활용촉진에 관한 법률 제13조의2(재활용센터 설치·운영)	P1(재활용정거장)	https://www.law.go.kr/법령/자원의절약과재활용촉진에관한법률	민용기센터 등 설치·운영 근거
자원의 절약과 재활용촉진에 관한 법률 제15조의2(번용기·1회용컵 촉진)	P2(무인화수기)	https://www.law.go.kr/법령/자원의절약과재활용촉진에관한법률	무인화수기 설치·보증금 환급 제도 근거
환경정책기본법 제10조(자원 등의 절약 및 순환적 사용 촉진)	전반(자원절약·순환사용)	https://law.go.kr/법령/환경정책기본법	국가 자원절약·순환사용 정책 근거
환경정책기본법 제25조(환경보전에 관한 교육 등)	전반(환경교육)	https://law.go.kr/법령/환경정책기본법	환경보전 교육·홍보 프로그램 시행 근거
순환경체사회 전환 촉진법 제5조(국가 및 지방자치단체의 책무)	전반(순환경체 시책)	https://law.go.kr/LSW//lInfoP.do?lId=012567&ancYnChk=0#0000	순환경체 전환 촉진을 위한 정부·지자체 책무
서울특별시 쓰레기줄이기와 재활용촉진에 관한 조례 제3조제2항6호(재활용에 관한 교육 및 홍보)	전반(재활용 교육·홍보)	https://www.law.go.kr/LSW/ordinInfoP.do?ordinId=2000226	시 차원의 분리배출·재활용 교육·홍보 의무
서울특별시 쓰레기줄이기와 재활용촉진에 관한 조례 시행 규칙 제4조(보조금 등의 지급)	zero_restaurant 보조금	https://www.law.go.kr/LSW/ordinInfoP.do?ordinId=2000226	제로 식당 인프라 설치 시 보조금 지급 기준
서울특별시 쓰레기줄이기와 재활용촉진에 관한 조례 제8조(교육·홍보)	전반(교육·홍보)	https://www.law.go.kr/LSW/ordinInfoP.do?ordinId=2000226	시 차원의 분리배출·재활용 교육·홍보 의무
서울특별시 쓰레기줄이기와 재활용촉진에 관한 조례 제9조(자발적 협약의 체결)	zero_restaurant 협약	https://www.law.go.kr/LSW/ordinInfoP.do?ordinId=2000226	민간 업체와 자발적 협약 체결 근거
서울특별시 강남구 1회용품 사용 줄이기에 관한 조례 제3조(구청장의 책무)	P5(일회용컵보증금제)	https://www.law.go.kr/LSW/ordinInfoP.do?ordinSeq=1832841	구청장은 공공기관 및 주민의 1회용품 사용을 줄이기 위한 시책 수립·시행하고, 교육·홍보 등 필요한 조치를 해야 함
재활용지정사업자의 재활용 지침 제3조(재활용계획의 수립 등)	전반(사업자 계획 수립)	https://law.go.kr/LSW/admRulLInfoP.do?admRulId=32318&efYd=0	지정사업자의 연간 재활용계획 수립·보고 규정
재활용가능자원의 분리수거 등에 관한 지침 제4조(분리배출 유형)	P6(특수품목분리수거)	https://www.law.go.kr/행정규칙/재활용가능자원의분리수거등에관한지침	재활용품별 분리배출 유형·방법 구체화
파주시 천환경 소재 사용 촉진 및 재활용 활성화 조례 제 1조 (목적)	P3(폐현수막 재활용)	https://www.elis.go.kr/allar/selectArlBrdOne?aino=41510121265010&histNo=001&menuNm=main	지자체별로 조례가 다르므로, 각 시·군 조례를 확인·제정
대구광역시 종기 1회용품 사용 줄이기 및 다회용기 사용 활성화 지원에 관한 조례 제 1조 (목적)	P6(다회용기 도입)	https://www.law.go.kr/LSW/ordinInfoP.do?ordinSeq=2001597	광역·기초단체 조례로 활성화 지원 근거
자원의 절약과 재활용촉진에 관한 법률 시행규칙 제8조(플라스틱 제품·재료·용기의 폐기물부담금 면제를 위한 회수·재활용 방법)	P9(재활용품회수보상제)	https://www.law.go.kr/법령/자원의절약과재활용촉진에관한법률시행규칙	국·공유 폐기물부담금 면제 신청과 연계해 보상체계를 설계
재활용가능자원의 분리수거 등에 관한 지침 제6조(분리수거용기의 설치·관리 등) 4항	P9(재활용품회수보상제)	https://www.law.go.kr/행정규칙/재활용가능자원의분리수거등에관한지침	④ 시장·군수·구청장은 재활용가능자원의 분리배출 편의성 제고 및 수거량 확보 등을 위하여 재활용가능자원을 지역화폐 등으로 보상해주는 수거보상제도(무인화수기 포함한다)를 운영할 수 있다.

▼ 8. 기대효과

- | 분석 결과를 정책에 반영할 경우 기대되는 사회적 기여도, 경제적 효과, 국민편의 증진, 업무 효율성 등
- | 구체적인 수치와 근거를 제시

- OLS 회귀 모델 ($Y_i = \beta_1 X_{i1} + \dots + \beta_p X_{ip}$) 이용 (다른 모델?·마신러닝 모델?)
- 최소제곱법

[1] 각 클러스터별 가장 `recycling_rate` 높이는 `policy` 출력하는 코드

▼ code

```
import pandas as pd
import statsmodels.api as sm

cluster_cols = ['cluster', 'cluster_2', 'cluster_3', 'cluster_4', 'cluster_5', 'cluster_6', 'cluster_7']
policy_cols = ['P1','P2','P3','P4','P5','P6','P8','P9','zero_restaurant']
delta = 0.05 # 5%
```

```
def best_policy_per_cluster(df, cluster, policy_cols, delta):
    results = []
    for cluster_id in sorted(df[cluster].unique()):
        cdf = df[df[cluster]==cluster_id].dropna(subset=policy_cols + ['recycling_rate'])

        # OLS model
        X = sm.add_constant(cdf[policy_cols])
        y = cdf['recycling_rate']
        model = sm.OLS(y, X).fit()

        # β_policy * delta
        effects = model.params[policy_cols] * delta

        best_policy = effects.idxmax()
        best_effect = effects.max()

        results.append({
            'cluster': cluster_id,
            'best_policy': best_policy,
            'delta recycling_rate': best_effect
        })
```

```

return pd.DataFrame(results)

for cluster in cluster_cols:
    best_df = best_policy_per_cluster(df, cluster, policy_cols, delta)
    print(best_df)
    print("\n")

▼ result ex

cluster      best_policy  delta recycling_rate
0         0             P9        1.402333
1         1             P9        0.376700
2         2  zero_restaurant     0.314500

cluster_2      best_policy  delta recycling_rate
0         0.0           P9        2.367893
1         1.0  zero_restaurant     0.314500
2         2.0           P9        0.866118

cluster_3      best_policy  delta recycling_rate
0         0.0           P9        0.319
1         1.0           P9        0.416
2         2.0           P8        0.643

cluster_4      best_policy  delta recycling_rate
0         0.0           P9        2.43950
1         1.0           P9        1.36675
2         2.0           P8        0.64300
3         3.0  zero_restaurant     0.31450

cluster_5      best_policy  delta recycling_rate
0         0.0           P9        2.439500
1         1.0  zero_restaurant     0.314500
2         2.0           P8        0.643000
3         3.0           P9        0.711031

cluster_6      best_policy  delta recycling_rate
0         0.0           P3        0.958684
1         1.0           P9        1.676571
2         2.0           P1        0.739000

cluster_7      best_policy  delta recycling_rate
0         0             P8        0.80575
1         1             P9        3.05300
2         2             P9        1.56225

```

[2] 각 정책별 recycling rate 높이는 적정 delta 찾는 코드

▼ code

```

import pandas as pd
import statsmodels.api as sm

def find_policy_deltas_for_targets(
    cluster_df: pd.DataFrame,
    policies: list[str],
    target_rates: dict[int, float],
    cluster_col: str
) -> pd.DataFrame:

    coefs = {}
    for pol in policies:
        X = sm.add_constant(cluster_df[[pol]])
        y = cluster_df['recycling_rate_mean']
        model = sm.OLS(y, X).fit()
        coefs[pol] = model.params[pol]

    rows = []

    for cluster, vals in cluster_df.iterrows():
        current = vals['recycling_rate_mean']
        target = target_rates.get(cluster)

        needed = target - current
        for pol in policies:
            coef = coefs[pol]
            if coef > 0: # policy 종 recycling rate + 하는 것만
                req_delta = needed / coef
                rows.append({
                    'cluster_col': cluster,
                    'policy': pol,
                    'coef': coef,
                    'current_rate': current,
                    'target_rate': target,
                    'required_delta': req_delta
                })

    return pd.DataFrame(rows)

#print(cluster_7_df)
#print("\n")

# 클러스터별 목표 재활용률 지정 (예: 0→60%, 1→65%, 2→70%)
target_rates = {0: 60, 1: 65, 2: 70}

# 델타 계산
df_deltas = find_policy_deltas_for_targets(
    cluster_df = cluster_7_df,
    policies    = ['P1','P2','P3','P4','P5','P6','P8','P9','zero_restaurant'],
    target_rates = target_rates,
    cluster_col  = 'cluster_7'
)

print(df_deltas)

```

▼ result ex

	cluster_7	policy	coef	current_rate	target_rate
0	0	P1	4.325955	49.64	60
1	0	P2	10.009585	49.64	60
2	0	P4	31.645161	49.64	60
3	0	P9	8.821429	49.64	60
4	0	zero_restaurant	8.215548	49.64	60

```

5   1     P1  4.325955  52.18   65
6   1     P2  10.009585  52.18   65
7   1     P4  31.645161  52.18   65
8   1     P9  8.821429  52.18   65
9   1 zero_restaurant  8.215548  52.18   65
10  2     P1  4.325955  56.98   70
11  2     P2  10.009585  56.98   70
12  2     P4  31.645161  56.98   70
13  2     P9  8.821429  56.98   70
14  2 zero_restaurant  8.215548  56.98   70

required_delta
0  2.394847
1  1.035008
2  0.327380
3  1.174413
4  1.261024
5  2.963508
6  1.280772
7  0.405117
8  1.453279
9  1.560456
10 3.009740
11 1.300753
12 0.411437
13 1.475951
14 1.584800

```

[3] 동적으로 각 cluster, policy 적용해서 delta recycling rate 출력하는 코드

▼ code

```

import pandas as pd
import statsmodels.api as sm

policy_cols = ['P1','P2','P3','P4','P5','P6','P8','P9','zero_restaurant']

def predict_policy_effect(
    df: pd.DataFrame,
    clusters,
    cluster_id,
    policies,
    delta: float
) → pd.DataFrame:

    if isinstance(clusters, (int, float)):
        clusters = [int(clusters)]
    if isinstance(policies, str):
        policies = [policies]

    rows = []
    for cid in clusters:
        cdf = df[df[cluster_id] == cid].dropna(subset=policy_cols + ['recycling_rate'])
        if cdf.empty:
            continue

        # OLS model
        X = sm.add_constant(cdf[policy_cols])
        y = cdf['recycling_rate']
        model = sm.OLS(y, X).fit()

        for pol in policies:
            if pol not in model.params:
                continue
            effect = model.params[pol] * delta
            rows.append({
                'cluster_id': cid,
                'policy': pol,
                'delta_policy': delta,
                'delta_recycling_rate': effect
            })

    return pd.DataFrame(rows)

# cluster 0,1,2 모두에 대해 P1과 P3를 5%p 늘렸을 때
res = predict_policy_effect(df, clusters=[0,1,2], cluster_id='cluster_7', policies=['P1','P3'], delta=0.05)
print(res)

print("\n")

# cluster 10에 대해 zero_restaurant만 10%p 늘렸을 때
res2 = predict_policy_effect(df, clusters=1, cluster_id='cluster_5', policies='zero_restaurant', delta=0.10)
print(res2)

```

▼ result ex

```

cluster_7 policy delta_policy delta_recycling_rate
0     0     P1      0.05   -1.583750e-01
1     0     P3      0.05   3.555000e-01
2     1     P1      0.05   9.700000e-02
3     1     P3      0.05   -3.423206e-16
4     2     P1      0.05   -2.760625e-01
5     2     P3      0.05   -4.758390e-17

cluster_5      policy delta_policy delta_recycling_rate
0  1 zero_restaurant      0.1       0.629

```

▼ zero_re_cnt clustering

▼ cluster1

- recycling_rate, pop_density
- zero_re_cnt_mean : 22.45 / 25.86 / 33.71

cluster_1	boarding_mean	alighting_mean	waste_mean	recycling_rate_mean	businesses_mean	employees_mean	zero_re_cnt_mean
0.0	5.481897e+07	5.476494e+07	268.99	51.64	42901.91	184165.91	22.45
1.0	5.485885e+07	5.478078e+07	300.17	64.36	37568.71	161153.57	25.86
2.0	1.237560e+08	1.250335e+08	479.40	43.76	63589.00	377360.71	33.71

▼ district

```

gu
0 [dongdaemun_gu, dongjak_gu, gangdong_gu, geumcheon_gu, guro_gu, jungnang_gu, seodaemun_gu, seongbuk_gu, seongdong_gu, songpa_gu, yangcheon_gu]
1 [dobong_gu, gangbuk_gu, gwanak_gu, gwangjin_gu, nowon_gu, yeongdeungpo_gu, yongsan_gu]
2 [enpyeong_gu, gangnam_gu, gangseo_gu, jongno_gu, jung_gu, mapo_gu, seocho_gu]

```

▼ cluster2

- Total_Boarding, Total_alighting
- zero_re_cnt_mean : 29.56, 39.00, 16.40

cluster_2	boarding_mean	alighting_mean	waste_mean	recycling_rate_mean	businesses_mean	employees_mean	zero_re_cnt_mean

0.0	7.301615e+07	7.273185e+07	273.33	55.94	42736.44	182987.33	29.56
1.0	1.496937e+08	1.519726e+08	369.40	44.42	70628.50	440634.83	39.00
2.0	2.980053e+07	2.946921e+07	373.95	55.49	37162.60	150473.00	16.40

▼ district

gu
0 [dongjak_gu, enpyeong_gu, gangdong_gu, guro_gu, gwanak_gu, gwangjin_gu, nowon_gu, seongdong_gu, yeongdeungpo_gu]
1 [gangnam_gu, jongno_gu, jung_gu, mapo_gu, seocho_gu, songpa_gu]
2 [dobong_gu, dongdaemun_gu, gangbuk_gu, gangseo_gu, geumcheon_gu, jungnang_gu, seodaemun_gu, seongbuk_gu, yangcheon_gu, yongsan_gu]

▼ cluster3

- num_businesses , num_employees
- zero_re_cnt_mean : 24.29, 42.00, 20.17

cluster_3	num_businesses_mean	num_employees_mean	total_retail_mean	waste_generation_mean	recycling_rate_mean	zero_restaurant_rate	zero_re_cnt_mean
0.0	47947.86	227613.57	3598.86	407.00	49.79	0.43	24.29
1.0	75274.33	466440.50	5948.00	389.03	46.15	0.83	42.00
2.0	32728.67	116957.25	2961.75	269.39	58.29	0.50	20.17

▼ district

gu
0 [dongdaemun_gu, gangdong_gu, gangseo_gu, geumcheon_gu, guro_gu, jongno_gu, seongdong_gu]
1 [gangnam_gu, jung_gu, mapo_gu, seocho_gu, songpa_gu, yeongdeungpo_gu]
2 [dobong_gu, dongjak_gu, enpyeong_gu, gangbuk_gu, gwanak_gu, gwangjin_gu, jungnang_gu, nowon_gu, seodaemun_gu, seongbuk_gu, yangcheon_gu, yongsan_gu]

▼ cluster4_1 (k=3)

- total_boarding , total_allighting , num_businesses , num_employees
- zero_re_cnt_mean : 30.89, 40.14, 11.67

cluster_4(1)	boarding_mean	allighting_mean	businesses_mean	employees_mean	waste_mean	recycling_rate_mean	zero_re_cnt_mean
0.0	6.929765e+07	6.858616e+07	40979.78	168108.11	401.33	54.57	30.89
1.0	1.400963e+08	1.423407e+08	71320.14	439578.00	363.24	45.89	40.14
2.0	2.766225e+07	2.749488e+07	34662.89	133934.00	251.24	56.96	11.67

▼ district

gu
0 [dongjak_gu, enpyeong_gu, gangdong_gu, gangseo_gu, guro_gu, gwanak_gu, gwangjin_gu, nowon_gu, seongdong_gu]
1 [gangnam_gu, jongno_gu, jung_gu, mapo_gu, seocho_gu, songpa_gu, yeongdeungpo_gu]
2 [dobong_gu, dongdaemun_gu, gangbuk_gu, geumcheon_gu, jungnang_gu, seodaemun_gu, seongbuk_gu, yangcheon_gu, yongsan_gu]

▼ cluster4_2 (k=4)

- total_boarding , total_allighting , num_businesses , num_employees
- zero_re_cnt_mean : 30.89, 50.00, 11.67, 27.00

cluster_4(2)	boarding_mean	allighting_mean	businesses_mean	employees_mean	waste_mean	recycling_rate_mean	zero_re_cnt_mean
0.0	6.929765e+07	6.858616e+07	40979.78	168108.11	401.33	54.57	30.89
1.0	1.342946e+08	1.365818e+08	81996.75	534454.25	441.98	47.80	50.00
2.0	2.766225e+07	2.749488e+07	34662.89	133934.00	251.24	56.96	11.67
3.0	1.478318e+08	1.500193e+08	57084.67	313076.33	258.27	43.34	27.00

▼ district

gu
0 [dongjak_gu, enpyeong_gu, gangdong_gu, gangseo_gu, guro_gu, gwanak_gu, gwangjin_gu, nowon_gu, seongdong_gu]
1 [gangnam_gu, seocho_gu, songpa_gu, yeongdeungpo_gu]
2 [dobong_gu, dongdaemun_gu, gangbuk_gu, geumcheon_gu, jungnang_gu, seodaemun_gu, seongbuk_gu, yangcheon_gu, yongsan_gu]
3 [jongno_gu, jung_gu, mapo_gu]

▼ cluster5_1 (k=3)

- total_boarding , total_allighting , total_retail , retail_cols
- zero_re_cnt_mean : 40.25, 39.00, 17.93

cluster_5_(1)	boarding_mean	allighting_mean	businesses_mean	employees_mean	waste_mean	recycling_rate_mean	zero_restaurant_rate	Beverage_Retail	Buffet	Chinese_Food	General_Lodging	Japanese_Food	Korean_Food	Other_Lodging	Pub	1
0.0	6.132675e+07	60914949.5	610.88	0.0	51.26	0.75		24.50	49.00	342.50	219.50	287.75	2333.25	156.5	864.25	.
1.0	1.496937e+08	151972636.5	369.40	0.0	44.42	0.83		61.67	73.00	279.00	469.33	503.33	2857.67	146.5	929.17	:
2.0	4.732291e+07	47041259.2	250.40	0.0	56.89	0.40		18.20	31.13	156.53	138.40	180.73	1567.27	94.8	536.73	.

▼ district

gu
0 [gangseo_gu, guro_gu, gwanak_gu, yeongdeungpo_gu]
1 [gangnam_gu, jongno_gu, jung_gu, mapo_gu, seocho_gu, songpa_gu]
2 [dobong_gu, dongdaemun_gu, dongjak_gu, enpyeong_gu, gangbuk_gu, gangdong_gu, geumcheon_gu, gwangjin_gu, jungnang_gu, nowon_gu, seodaemun_gu, seongbuk_gu, seongdong_gu, yangcheon_gu, yongsan_gu]

▼ cluster5_2 (k=4)

- total_boarding , total_allighting , total_retail , retail_cols
- zero_re_cnt_mean : 40.25, 30.25, 17.93, 56.50

cluster_5_(2)	boarding_mean	allighting_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	zero_restaurant_rate	Beverage_Retail	Buffet	Chinese_Food	General_Lodging	Japanese_Food	Korean_Food	Other_Lodging	Pub	1
0.0	61326745.5	6.091495e+07	610.88	0.0	51.26	0.75	24.50	49.00	342.50	219.50	287.75	2333.25	156.5	864.25	.
1.0	149632975.5	1.505539e+08	351.25	0.0	44.15	0.75	53.25	66.50	230.75	381.25	357.25	2523.00	137.0	764.75	:
2.0	47322908.8	4.704126e+07	250.40	0.0	56.89	0.40	18.20	31.13	156.53	138.40	180.73	1567.27	94.8	536.73	.
3.0	149815188.5	1.548102e+08	405.70	0.0	44.96	1.00	78.50	86.00	375.50	645.50	795.50	3527.00	165.5	1258.00	1

▼ district

gu
0 [gangseo_gu, guro_gu, gwanak_gu, yeongdeungpo_gu]
1 [jongno_gu, jung_gu, seocho_gu, songpa_gu]
2 [dobong_gu, dongdaemun_gu, dongjak_gu, enpyeong_gu, gangbuk_gu, gangdong_gu, geumcheon_gu, gwangjin_gu, jungnang_gu, nowon_gu, seodaemun_gu, seongbuk_gu, seongdong_gu, yangcheon_gu, yongsan_gu]
3 [gangnam_gu, mapo_gu]

▼ cluster6

- population_per_household , house_hold , area_km2
- zero_re_cnt_mean : 22.67, 38.25, 57.00, 14.25

cluster_6	boarding_mean	allighting_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	zero_restaurant_rate	pop_per_household_mean	house_hold_mean	area_km2_mean	Beverage_Retail	Buffet	Chinese_Food	General_Lodging	Jai	Ja
0.0	5.661833e+07	5.664938e+07	245.37	0.											

cluster_7	boarding_mean	alighting_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	zero_restaurant_rate	man_woman_ratio_mean	children_rate_mean	youth_rate_mean	middle_rate_mean	Beverage_Retail	Buffet	Chinese Food
0.0	65461086.78	65332597.67	222.77	0.0	49.64	0.11	0.96	0.12	0.29	0.31	20.56	37.78	178.89
1.0	93662417.43	94480264.86	351.77	0.0	52.18	0.57	0.94	0.16	0.27	0.33	48.43	56.29	228.43
2.0	67614021.11	67973252.44	438.73	0.0	56.98	1.00	0.94	0.13	0.33	0.30	24.11	40.78	242.56

▼ district

```
gu
0 [dobong_gu, dongdaemun_gu, enpyeong_gu, gangbuk_gu, geumcheon_gu, guro_gu, jongno_gu, jung_gu, jungnang_gu]
1 [gangdong_gu, gangnam_gu, nowon_gu, seocho_gu, seongbuk_gu, songpa_gu, yangcheon_gu]
2 [dongjak_gu, gangseo_gu, gwanak_gu, gwangjin_gu, mapo_gu, seodaemun_gu, seongdong_gu, yeongdeungpo_gu, yongsan_gu]
```

▼ cluster_PCA1

- Total_Boarding, Total_Alighting, num_businesses, num_employees, population_per_household, household, area_km2, total_retail, retail_cols
- zero_re_cnt_mean : 21.28, 36.67, 61.00

cluster_PCA1	boarding_mean	alighting_mean	businesses_mean	employees_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	zero_restaurant_rate	man_woman_ratio_mean	children_rate_mean	youth_rate_mean	middle_rate_mean	population_
0.0	4.847995e+07	4.804052e+07	37821.33	151021.06	326.29	0.0	55.76	0.44	0.95	0.13	0.30	0.31	2.11
1.0	1.347223e+08	1.363830e+08	65239.50	379023.00	344.63	0.0	46.49	0.83	0.93	0.14	0.31	0.31	2.10
2.0	1.723401e+08	1.780871e+08	107804.00	802908.00	474.90	0.0	42.26	1.00	0.92	0.17	0.28	0.34	2.26

▼ district

```
gu
0 [dobong_gu, dongdaemun_gu, dongjak_gu, enpyeong_gu, gangbuk_gu, gangdong_gu, gangseo_gu, geumcheon_gu, guro_gu, gwanak_gu, gwangjin_gu, jungnang_gu, nowon_gu, seodaemun_gu, seongbuk_gu, seongdong_gu, yangcheon_gu, yongsan_gu]
1 [jungno_gu, jung_gu, mapo_gu, seocho_gu, songpa_gu, yeongdeungpo_gu]
2 [gangnam_gu]
```

▼ cluster_PCA2

- Total_Boarding, Total_Alighting, num_businesses, num_employees, man_woman_ratio, children_rate, youth_rate, middle_rate, elderly_rate, totaretail_cols
- zero_re_cnt_mean : 34.00, 40.14, 8.56

cluster_PCA2	boarding_mean	alighting_mean	businesses_mean	employees_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	zero_restaurant_rate	man_woman_ratio_mean	children_rate_mean	youth_rate_mean	middle_rate_mean	population_
0.0	5.523150e+07	5.492179e+07	41132.11	185032.33	418.91	0.0	56.84	0.67	0.96	0.12	0.33	0.30	2.01
1.0	1.400963e+08	1.423407e+08	71320.14	439578.00	363.24	0.0	45.89	0.86	0.93	0.14	0.30	0.32	2.13
2.0	4.172840e+07	4.115925e+07	34510.56	117009.78	233.67	0.0	54.68	0.22	0.94	0.14	0.27	0.32	2.21

▼ district

```
gu
0 [dongdaemun_gu, dongjak_gu, gangseo_gu, geumcheon_gu, guro_gu, gwanak_gu, gwangjin_gu, seongdong_gu, yongsan_gu]
1 [gangnam_gu, jongno_gu, jung_gu, mapo_gu, seocho_gu, songpa_gu, yeongdeungpo_gu]
2 [dobong_gu, enpyeong_gu, gangbuk_gu, gangdong_gu, jungnang_gu, nowon_gu, seodaemun_gu, seongbuk_gu, yangcheon_gu]
```

▼ cluster_PCA3

- Total_Boarding, Total_Alighting, num_businesses, num_employees, population_per_household, household, area_km2, man_woman_ratio, children_rate, youth_rate, middle_rate, elderly_rate, retail_cols
- zero_re_cnt_mean : 16.81, 41.75, 61.00

cluster_PCA3	boarding_mean	alighting_mean	businesses_mean	employees_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	zero_restaurant_rate	man_woman_ratio_mean	children_rate_mean	youth_rate_mean	middle_rate_mean	population_
0.0	4.817657e+07	4.790617e+07	36519.38	143291.25	252.78	0.0	56.38	0.38	0.95	0.13	0.29	0.31	2.14
1.0	1.137685e+08	1.145661e+08	60988.88	337482.12	487.08	0.0	47.57	0.88	0.94	0.13	0.32	0.31	2.05
2.0	1.723401e+08	1.780871e+08	107804.00	802908.00	474.90	0.0	42.26	1.00	0.92	0.17	0.28	0.34	2.26

▼ district

```
gu
0 [dobong_gu, dongdaemun_gu, dongjak_gu, enpyeong_gu, gangbuk_gu, gangdong_gu, geumcheon_gu, guro_gu, gwangjin_gu, jungnang_gu, nowon_gu, seodaemun_gu, seongbuk_gu, seongdong_gu, yangcheon_gu, yongsan_gu]
1 [gangseo_gu, gwanak_gu, jongno_gu, jung_gu, mapo_gu, seocho_gu, songpa_gu, yeongdeungpo_gu]
2 [gangnam_gu]
```

▼ cluster_PCA4

- Total_Boarding, Total_Alighting, population_per_household, household, area_km2, man_woman_ratio, children_rate, youth_rate, middle_rate, elderly_rate
- zero_re_cnt_mean : 14.38, 38.25, 40.50

cluster_PCA4	boarding_mean	alighting_mean	businesses_mean	employees_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	zero_restaurant_rate	man_woman_ratio_mean	children_rate_mean	youth_rate_mean	middle_rate_mean	population_
0.0	3.988729e+07	3.964065e+07	37372.85	148190.46	251.62	0.0	54.92	0.31	0.95	0.13	0.29	0.32	2.14
1.0	1.332510e+08	1.350171e+08	72633.75	458444.50	431.32	0.0	49.74	0.75	0.92	0.17	0.27	0.33	2.33
2.0	1.002218e+08	1.007147e+08	50455.38	254396.38	427.44	0.0	51.50	0.88	0.94	0.12	0.33	0.30	1.97

▼ district

```
gu
0 [dobong_gu, dongdaemun_gu, enpyeong_gu, gangbuk_gu, gangdong_gu, geumcheon_gu, guro_gu, gwangjin_gu, jungnang_gu, nowon_gu, seodaemun_gu, seongbuk_gu, seongdong_gu, yangcheon_gu, yongsan_gu]
1 [gangnam_gu, nowon_gu, seocho_gu, songpa_gu]
2 [dongjak_gu, gangseo_gu, gwanak_gu, gwangjin_gu, jongno_gu, jung_gu, mapo_gu, yeongdeungpo_gu]
```

▼ cluster_PCA5

- population_per_household, household, area_km2, man_woman_ratio, children_rate, youth_rate, middle_rate, elderly_rate
- zero_re_cnt_mean : 22.54, 37.40, 26.29

cluster_PCA5	boarding_mean	alighting_mean	businesses_mean	employees_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	zero_restaurant_rate	man_woman_ratio_mean	children_rate_mean	youth_rate_mean	middle_rate_mean	population_
0.0	5.848198e+07	5.837180e+07	39993.46	172528.54	355.05	0.0	53.14	0.46	0.94	0.13	0.29	0.31	2.12
1.0	1.10997e+08	1.125342e+08	65661.80	393518.60	394.92	0.0	50.32	0.80	0.93	0.17	0.27	0.33	2.35
2.0	7.679269e+07	7.708715e+07	47400.14	226423.00	260.81	0.0	54.65	0.57	0.97	0.12	0.34	0.30	1.95

▼ district

```
gu
0 [dobong_gu, enpyeong_gu, gangbuk_gu, gangdong_gu, gangseo_gu, guro_gu, gwangjin_gu, jungsang_gu, mapo_gu, seodaemun_gu, seongbuk_gu, seongdong_gu, yangcheon_gu, yongsan_gu]
1 [gangnam_gu, nowon_gu, seocho_gu, songpa_gu, yangcheon_gu]
2 [dongdaemun_gu, dongjak_gu, geumcheon_gu, gwanak_gu, gwangjin_gu, jung_gu, yeongdeungpo_gu]
```

▼ zero_re_ratio clustering

- 새로운 열 설명
- zero_re_ratio : zero_re_cnt / total_retail (total_retail 은 전체 식당 수)
- waste_per_rider : df['waste_generation'] / (df['Total_Boarding'] + df['Total_Alighting'])

▼ children_rate, youth_rate, middle_rate, elderly_rate

```
df['children_rate'] = (
    df['0to9_pop'] + df['10to19_pop']
) / df['population_total']

df['y
```

) / df['population_total']																																													
	df['elderly_rate'] = (df[['60to69_pop','70to79_pop','80to89_pop','90to99_pop','100over_pop']].sum(axis=1)) / df['population_total']																																													
▼ cluster1	<ul style="list-style-type: none"> recycling_rate, pop_density zero_re_ratio_mean : 0.006695 / 0.005580 																																													
cluster_1	<table border="1"> <thead> <tr><th>cluster_1</th><th>boarding_mean</th><th>alighting_mean</th><th>waste_mean</th><th>recycling_rate_mean</th><th>businesses_mean</th><th>employees_mean</th><th>zero_re_ratio_mean</th></tr> </thead> <tbody> <tr><td>0.0</td><td>7.760516e+07</td><td>7.799498e+07</td><td>341.614286</td><td>49.800</td><td>50259.142857</td><td>252706.714286</td><td>0.006695</td></tr> <tr><td>1.0</td><td>5.590111e+07</td><td>5.580500e+07</td><td>310.500000</td><td>69.775</td><td>31145.750000</td><td>122146.000000</td><td>0.005580</td></tr> </tbody> </table>	cluster_1	boarding_mean	alighting_mean	waste_mean	recycling_rate_mean	businesses_mean	employees_mean	zero_re_ratio_mean	0.0	7.760516e+07	7.799498e+07	341.614286	49.800	50259.142857	252706.714286	0.006695	1.0	5.590111e+07	5.580500e+07	310.500000	69.775	31145.750000	122146.000000	0.005580																					
cluster_1	boarding_mean	alighting_mean	waste_mean	recycling_rate_mean	businesses_mean	employees_mean	zero_re_ratio_mean																																							
0.0	7.760516e+07	7.799498e+07	341.614286	49.800	50259.142857	252706.714286	0.006695																																							
1.0	5.590111e+07	5.580500e+07	310.500000	69.775	31145.750000	122146.000000	0.005580																																							
▼ district	<p>gu</p> <p>0 [dongdaemun_gu, dongjak_gu, enpyeong_gu, gangbuk_gu, gangdong_gu, gangnam_gu, gangseo_gu, geumcheon_gu, guro_gu, gwanak_gu, jongno_gu, jung_gu, jungsang_gu, mapo_gu, seocho_gu, seodaemun_gu, seongbuk_gu, seongdong_gu, songpa_gu, yangcheon_gu, yeongdeungpo_gu]</p> <p>1 [dobong_gu, gwangjin_gu, nowon_gu, yongsan_gu]</p>																																													
▼ cluster2	<ul style="list-style-type: none"> Total_Boarding, Total_alighting zero_re_ratio_mean : 0.006525, 0.006489 																																													
cluster_2	<table border="1"> <thead> <tr><th>cluster_2</th><th>boarding_mean</th><th>alighting_mean</th><th>waste_mean</th><th>recycling_rate_mean</th><th>businesses_mean</th><th>employees_mean</th><th>zero_re_ratio_mean</th></tr> </thead> <tbody> <tr><td>0.0</td><td>5.027108e+07</td><td>4.996204e+07</td><td>326.289474</td><td>55.703158</td><td>39802.842105</td><td>165874.526316</td><td>0.006525</td></tr> <tr><td>1.0</td><td>1.496937e+08</td><td>1.519726e+08</td><td>369.400000</td><td>44.423333</td><td>70628.500000</td><td>440634.833333</td><td>0.006489</td></tr> </tbody> </table>	cluster_2	boarding_mean	alighting_mean	waste_mean	recycling_rate_mean	businesses_mean	employees_mean	zero_re_ratio_mean	0.0	5.027108e+07	4.996204e+07	326.289474	55.703158	39802.842105	165874.526316	0.006525	1.0	1.496937e+08	1.519726e+08	369.400000	44.423333	70628.500000	440634.833333	0.006489																					
cluster_2	boarding_mean	alighting_mean	waste_mean	recycling_rate_mean	businesses_mean	employees_mean	zero_re_ratio_mean																																							
0.0	5.027108e+07	4.996204e+07	326.289474	55.703158	39802.842105	165874.526316	0.006525																																							
1.0	1.496937e+08	1.519726e+08	369.400000	44.423333	70628.500000	440634.833333	0.006489																																							
▼ district	<p>gu</p> <p>0 [dobong_gu, dongdaemun_gu, dongjak_gu, enpyeong_gu, gangbuk_gu, gangdong_gu, gangnam_gu, gangseo_gu, geumcheon_gu, guro_gu, gwanak_gu, gwangjin_gu, jungsang_gu, nowon_gu, seodaemun_gu, seongbuk_gu, seongdong_gu, yangcheon_gu, yeongdeungpo_gu, yongsan_gu]</p> <p>1 [gangnam_gu, jongno_gu, jung_gu, mapo_gu, seocho_gu, songpa_gu]</p>																																													
▼ cluster3	<ul style="list-style-type: none"> num_businesses, num_employees zero_ratio_mean : 0.006029, 0.007768 																																													
cluster_3	<table border="1"> <thead> <tr><th>cluster_3</th><th>num_businesses_mean</th><th>num_employees_mean</th><th>total_retail_mean</th><th>waste_generation_mean</th><th>recycling_rate_mean</th><th>zero_re_ratio_mean</th></tr> </thead> <tbody> <tr><td>0.0</td><td>37151.055556</td><td>149858.777778</td><td>3119.166667</td><td>255.744444</td><td>55.873889</td><td>0.006029</td></tr> <tr><td>1.0</td><td>73043.714286</td><td>442566.714286</td><td>5753.714286</td><td>544.642857</td><td>45.595714</td><td>0.007768</td></tr> </tbody> </table>	cluster_3	num_businesses_mean	num_employees_mean	total_retail_mean	waste_generation_mean	recycling_rate_mean	zero_re_ratio_mean	0.0	37151.055556	149858.777778	3119.166667	255.744444	55.873889	0.006029	1.0	73043.714286	442566.714286	5753.714286	544.642857	45.595714	0.007768																								
cluster_3	num_businesses_mean	num_employees_mean	total_retail_mean	waste_generation_mean	recycling_rate_mean	zero_re_ratio_mean																																								
0.0	37151.055556	149858.777778	3119.166667	255.744444	55.873889	0.006029																																								
1.0	73043.714286	442566.714286	5753.714286	544.642857	45.595714	0.007768																																								
▼ district	<p>gu</p> <p>0 [dobong_gu, dongdaemun_gu, dongjak_gu, enpyeong_gu, gangbuk_gu, gangdong_gu, gangnam_gu, gangseo_gu, geumcheon_gu, guro_gu, gwanak_gu, gwangjin_gu, jongno_gu, jungsang_gu, nowon_gu, seodaemun_gu, seongbuk_gu, seongdong_gu, yangcheon_gu, yongsan_gu]</p> <p>1 [gangnam_gu, gangseo_gu, jung_gu, mapo_gu, seocho_gu, songpa_gu, yeongdeungpo_gu]</p>																																													
▼ cluster4_1	<ul style="list-style-type: none"> total_boarding, total_alighting, num_businesses, num_employees zero_re_ratio_mean : 0.006419, 0.006767 																																													
cluster_4(1)	<table border="1"> <thead> <tr><th>cluster_4(1)</th><th>boarding_mean</th><th>alighting_mean</th><th>businesses_mean</th><th>employees_mean</th><th>waste_mean</th><th>recycling_rate_mean</th><th>zero_re_ratio_mean</th></tr> </thead> <tbody> <tr><td>0.0</td><td>4.847995e+07</td><td>4.804052e+07</td><td>37821.333333</td><td>151021.055556</td><td>326.288889</td><td>55.760556</td><td>0.006419</td></tr> <tr><td>1.0</td><td>1.400963e+08</td><td>1.423407e+08</td><td>71320.142857</td><td>439578.000000</td><td>363.242857</td><td>45.887143</td><td>0.006767</td></tr> </tbody> </table>	cluster_4(1)	boarding_mean	alighting_mean	businesses_mean	employees_mean	waste_mean	recycling_rate_mean	zero_re_ratio_mean	0.0	4.847995e+07	4.804052e+07	37821.333333	151021.055556	326.288889	55.760556	0.006419	1.0	1.400963e+08	1.423407e+08	71320.142857	439578.000000	363.242857	45.887143	0.006767																					
cluster_4(1)	boarding_mean	alighting_mean	businesses_mean	employees_mean	waste_mean	recycling_rate_mean	zero_re_ratio_mean																																							
0.0	4.847995e+07	4.804052e+07	37821.333333	151021.055556	326.288889	55.760556	0.006419																																							
1.0	1.400963e+08	1.423407e+08	71320.142857	439578.000000	363.242857	45.887143	0.006767																																							
▼ district	<p>gu</p> <p>0 [dobong_gu, dongdaemun_gu, dongjak_gu, enpyeong_gu, gangbuk_gu, gangdong_gu, gangnam_gu, gangseo_gu, geumcheon_gu, guro_gu, gwanak_gu, gwangjin_gu, jongno_gu, jungsang_gu, nowon_gu, seodaemun_gu, seongbuk_gu, seongdong_gu, yangcheon_gu, yongsan_gu]</p> <p>1 [gangnam_gu, jongno_gu, jung_gu, mapo_gu, seocho_gu, songpa_gu, yeongdeungpo_gu]</p>																																													
▼ cluster5_1	<ul style="list-style-type: none"> total_boarding, total_alighting, total_retail, retail_cols zero_re_ratio_mean : 0.006419, 0.006767 																																													
cluster_5_(1)	<table border="1"> <thead> <tr><th>cluster_5_(1)</th><th>boarding_mean</th><th>alighting_mean</th><th>waste_mean</th><th>waste_per_rider_mean</th><th>recycling_rate_mean</th><th>zero_restaurant_mean</th><th>total_retail_mean</th><th>Beverage_Retail</th><th>Buffet</th><th>Chinese_Food</th><th>General_Lodging</th><th>Japanese_Food</th><th>Korean_Food</th><th>Other_Lodging</th></tr> </thead> <tbody> <tr><td>0.0</td><td>4.847995e+07</td><td>4.804052e+07</td><td>326.288889</td><td>0.000004</td><td>55.760556</td><td>0.444444</td><td>3098.722222</td><td>18.333333</td><td>32.888889</td><td>176.666667</td><td>150.055556</td><td>192.611111</td><td>1661.555556</td><td>106.111111</td></tr> <tr><td>1.0</td><td>1.400963e+08</td><td>1.423407e+08</td><td>363.242857</td><td>0.000001</td><td>45.887143</td><td>0.857143</td><td>5806.285714</td><td>58.714286</td><td>72.714286</td><td>316.000000</td><td>438.428571</td><td>487.857143</td><td>2868.571429</td><td>145.285714</td></tr> </tbody> </table>	cluster_5_(1)	boarding_mean	alighting_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	zero_restaurant_mean	total_retail_mean	Beverage_Retail	Buffet	Chinese_Food	General_Lodging	Japanese_Food	Korean_Food	Other_Lodging	0.0	4.847995e+07	4.804052e+07	326.288889	0.000004	55.760556	0.444444	3098.722222	18.333333	32.888889	176.666667	150.055556	192.611111	1661.555556	106.111111	1.0	1.400963e+08	1.423407e+08	363.242857	0.000001	45.887143	0.857143	5806.285714	58.714286	72.714286	316.000000	438.428571	487.857143	2868.571429	145.285714
cluster_5_(1)	boarding_mean	alighting_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	zero_restaurant_mean	total_retail_mean	Beverage_Retail	Buffet	Chinese_Food	General_Lodging	Japanese_Food	Korean_Food	Other_Lodging																																
0.0	4.847995e+07	4.804052e+07	326.288889	0.000004	55.760556	0.444444	3098.722222	18.333333	32.888889	176.666667	150.055556	192.611111	1661.555556	106.111111																																
1.0	1.400963e+08	1.423407e+08	363.242857	0.000001	45.887143	0.857143	5806.285714	58.714286	72.714286	316.000000	438.428571	487.857143	2868.571429	145.285714																																
▼ district	<p>gu</p> <p>0 [dobong_gu, dongdaemun_gu, dongjak_gu, enpyeong_gu, gangbuk_gu, gangdong_gu, gangnam_gu, gangseo_gu, geumcheon_gu, guro_gu, gwanak_gu, gwangjin_gu, jongno_gu, jungsang_gu, nowon_gu, seodaemun_gu, seongbuk_gu, seongdong_gu, yangcheon_gu, yongsan_gu]</p> <p>1 [gangnam_gu, jongno_gu, jung_gu, mapo_gu, seocho_gu, songpa_gu, yeongdeungpo_gu]</p>																																													
▼ cluster6	<ul style="list-style-type: none"> population_per_household, house_hold, area_km2 zero_re_ratio_mean : 0.006924, 0.005467 																																													
cluster_6	<table border="1"> <thead> <tr><th>cluster_6</th><th>boarding_mean</th><th>alighting_mean</th><th>waste_mean</th><th>waste_per_rider_mean</th><th>recycling_rate_mean</th><th>zero_restaurant_rate</th><th>pop_per_household_mean</th><th>house_hold_mean</th><th>area_km2_mean</th><th>Beverage_Retail</th><th>Buffet</th><th>Chinese_Food</th><th>General_Lodging</th><th>Jai</th></tr> </thead> <tbody> <tr><td>0.0</td><td>6.291169e+07</td><td>6.317898e+07</td><td>261.144444</td><td>0.000003</td><td>54.527778</td><td>0.555556</td><td>2.056667</td><td>158823.388889</td><td>19.648889</td><td>22.722222</td><td>37.444444</td><td>205.388889</td><td>257.611111</td><td>23:</td></tr> <tr><td>1.0</td><td>1.029861e+08</td><td>1.034133e+08</td><td>530.757143</td><td>0.000004</td><td>49.057143</td><td>0.571429</td><td>2.267143</td><td>228826.142857</td><td>35.934286</td><td>47.428571</td><td>61.000000</td><td>242.142857</td><td>161.857143</td><td>38:</td></tr> </tbody> </table>	cluster_6	boarding_mean	alighting_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	zero_restaurant_rate	pop_per_household_mean	house_hold_mean	area_km2_mean	Beverage_Retail	Buffet	Chinese_Food	General_Lodging	Jai	0.0	6.291169e+07	6.317898e+07	261.144444	0.000003	54.527778	0.555556	2.056667	158823.388889	19.648889	22.722222	37.444444	205.388889	257.611111	23:	1.0	1.029861e+08	1.034133e+08	530.757143	0.000004	49.057143	0.571429	2.267143	228826.142857	35.934286	47.428571	61.000000	242.142857	161.857143	38:
cluster_6	boarding_mean	alighting_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	zero_restaurant_rate	pop_per_household_mean	house_hold_mean	area_km2_mean	Beverage_Retail	Buffet	Chinese_Food	General_Lodging	Jai																																
0.0	6.291169e+07	6.317898e+07	261.144444	0.000003	54.527778	0.555556	2.056667	158823.388889	19.648889	22.722222	37.444444	205.388889	257.611111	23:																																
1.0	1.029861e+08	1.034133e+08	530.757143	0.000004	49.057143	0.571429	2.267143	228826.142857	35.934286	47.428571	61.000000	242.142857	161.857143	38:																																
▼ district	<p>gu</p> <p>0 [dobong_gu, dongdaemun_gu, dongjak_gu, enpyeong_gu, gangbuk_gu, geumcheon_gu, guro_gu, gwanak_gu, gwangjin_gu, jongno_gu, jung_gu, jungsang_gu, mapo_gu, seodaemun_gu, seongbuk_gu, seongdong_gu, yangcheon_gu, yeongdeungpo_gu, yongsan_gu]</p> <p>1 [enpyeong_gu, gangdong_gu, gangnam_gu, gangseo_gu, nowon_gu, seocho_gu, songpa_gu]</p>																																													
▼ cluster7	<ul style="list-style-type: none"> man_woman_ratio, children_rate, youth_rate, middle_rate, elderly_rate zero_re_ratio_mean : 0.006784, 0.005946 																																													
cluster_7	<table border="1"> <thead> <tr><th>cluster_7</th><th>boarding_mean</th><th>alighting_mean</th><th>waste_mean</th><th>waste_per_rider_mean</th><th>recycling_rate_mean</th><th>zero_restaurant_rate</th><th>man_woman_ratio_mean</th><th>children_rate_mean</th><th>youth_rate_mean</th><th>middle_rate_mean</th><th>Beverage_Retail</th><th>Buffet</th><th>Chinese_Food</th></tr> </thead> <tbody> <tr><td>0.0</td><td>6.296386e+07</td><td>6.283643e+07</td><td>330.411765</td><td>0.000004</td><td>53.644706</td><td>0.529412</td><td>0.951765</td><td>0.122672</td><td>0.309428</td><td>0.306750</td><td>21.588235</td><td>38.823529</td><td>205.117647</td></tr> <tr><td>1.0</td><td>9.786590e+07</td><td>9.911189e+07</td><td>349.862500</td><td>0.000002</td><td>51.617500</td><td>0.625000</td><td>0.930000</td><td>0.161055</td><td>0.281673</td><td>0.326017</td><td>46.750000</td><td>55.125000</td><td>238.125000</td></tr> </tbody> </table>	cluster_7	boarding_mean	alighting_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	zero_restaurant_rate	man_woman_ratio_mean	children_rate_mean	youth_rate_mean	middle_rate_mean	Beverage_Retail	Buffet	Chinese_Food	0.0	6.296386e+07	6.283643e+07	330.411765	0.000004	53.644706	0.529412	0.951765	0.122672	0.309428	0.306750	21.588235	38.823529	205.117647	1.0	9.786590e+07	9.911189e+07	349.862500	0.000002	51.617500	0.625000	0.930000	0.161055	0.281673	0.326017	46.750000	55.125000	238.125000			
cluster_7	boarding_mean	alighting_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	zero_restaurant_rate	man_woman_ratio_mean	children_rate_mean	youth_rate_mean	middle_rate_mean	Beverage_Retail	Buffet	Chinese_Food																																	
0.0	6.296386e+07	6.283643e+07	330.411765	0.000004	53.644706	0.529412	0.951765	0.122672	0.309428	0.306750	21.588235	38.823529	205.117647																																	
1.0	9.786590e+07	9.911189e+07	349.862500	0.000002	51.617500	0.625000	0.930000	0.161055	0.281673	0.326017	46.750000	55.125000	238.125000																																	
▼ district	<p>gu</p> <p>0 [dobong_gu, dongdaemun_gu, dongjak_gu, enpyeong_gu, gangbuk_gu, geumcheon_gu, guro_gu, gwanak_gu, gwangjin_gu, jongno_gu, jung_gu, jungsang_gu, mapo_gu, seodaemun_gu, seongbuk_gu, seongdong_gu, yeongdeungpo_gu, yongsan_gu]</p> <p>1 [gangdong_gu, gangnam_gu, mapo_gu, nowon_gu, seocho_gu, songpa_gu, yangcheon_gu]</p>																																													
▼ cluster_PCA1	<ul style="list-style-type: none"> Total_Boarding, Total_Alighting, num_businesses, num_employees, population_per_household, house_hold, area_km2, total_retail, retail_cols zero_re_ratio_mean : 0.006419, 0.006767 																																													

cluster_PCA1	boarding_mean	alighting_mean	businesses_mean	employees_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	zero_restaurant_rate	man_woman_ratio_mean	children_rate_mean	youth_rate_mean	middle_rate_mean	population_mean
0.0	4.847995e+07	4.804052e+07	37821.333333	151021.055556	326.288889	0.000004	55.760556	0.444444	0.950000	0.131738	0.299744	0.311269	2.111667
1.0	1.400963e+08	1.423407e+08	71320.142857	439578.000000	363.242857	0.000001	45.887143	0.857143	0.931429	0.143225	0.302609	0.317148	2.125714

▼ district

gu
0 [dobong_gu, dongdaemun_gu, dongjak_gu, enpyeong_gu, gangbuk_gu, gangdong_gu, gangseo_gu, geumcheon_gu, guro_gu, gwanak_gu, gwangjin_gu, jungnang_gu, nowon_gu, seodaemun_gu, seongbuk_gu, seongdong_gu, yangcheon_gu, yongsan_gu]
1 [gangnam_gu, jongno_gu, jung_gu, mapo_gu, seocho_gu, songpa_gu, yeongdeungpo_gu]

▼ cluster_PCA2

- Total_Boarding, Total_Alighting, num_businesses, num_employees, man_woman_ratio, children_rate, youth_rate, middle_rate, elderly_rate, total_retail, retail_cols
- zero_re_ratio_mean : 0.006419, 0.006767

cluster_PCA2	boarding_mean	alighting_mean	businesses_mean	employees_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	man_woman_ratio_mean	children_rate_mean	youth_rate_mean	middle_rate_mean	population_per_household_mean
0.0	4.847995e+07	4.804052e+07	37821.333333	151021.055556	326.288889	0.000004	55.760556	0.950000	0.131738	0.299744	0.311269	2.111667
1.0	1.400963e+08	1.423407e+08	71320.142857	439578.000000	363.242857	0.000001	45.887143	0.931429	0.143225	0.302609	0.317148	2.125714

▼ district

gu
0 [dobong_gu, dongdaemun_gu, dongjak_gu, enpyeong_gu, gangbuk_gu, gangdong_gu, gangseo_gu, geumcheon_gu, guro_gu, gwanak_gu, gwangjin_gu, jungnang_gu, nowon_gu, seodaemun_gu, seongbuk_gu, seongdong_gu, yangcheon_gu, yongsan_gu]
1 [gangnam_gu, jongno_gu, jung_gu, mapo_gu, seocho_gu, songpa_gu, yeongdeungpo_gu]

▼ cluster_PCA3

- Total_Boarding, Total_Alighting, num_businesses, num_employees, population_per_household, household, area_km2, man_woman_ratio, children_rate, youth_rate, middle_rate, elderly_rate, retail_cols
- zero_re_cnt_mean : 0.006419, 0.006767

cluster_PCA3	boarding_mean	alighting_mean	businesses_mean	employees_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	man_woman_ratio_mean	children_rate_mean	youth_rate_mean	middle_rate_mean	population_per_household_mean
0.0	4.847995e+07	4.804052e+07	37821.333333	151021.055556	326.288889	0.000004	55.760556	0.950000	0.131738	0.299744	0.311269	2.111667
1.0	1.400963e+08	1.423407e+08	71320.142857	439578.000000	363.242857	0.000001	45.887143	0.931429	0.143225	0.302609	0.317148	2.125714

▼ district

gu
0 [dobong_gu, dongdaemun_gu, dongjak_gu, enpyeong_gu, gangbuk_gu, gangdong_gu, gangseo_gu, geumcheon_gu, guro_gu, gwanak_gu, gwangjin_gu, jungnang_gu, nowon_gu, seodaemun_gu, seongbuk_gu, seongdong_gu, yangcheon_gu, yongsan_gu]
1 [gangnam_gu, jongno_gu, jung_gu, mapo_gu, seocho_gu, songpa_gu, yeongdeungpo_gu]

▼ cluster_PCA4

- Total_Boarding, Total_Alighting, population_per_household, household, area_km2, man_woman_ratio, children_rate, youth_rate, middle_rate, elderly_rate
- zero_re_ratio_mean : 0.006150, 0.007979

cluster_PCA4	boarding_mean	alighting_mean	businesses_mean	employees_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	man_woman_ratio_mean	children_rate_mean	youth_rate_mean	middle_rate_mean	population_per_household_mean
0.0	6.489073e+07	64922174.7	42585.8	191391.6	322.065	0.000004	53.666	0.948	0.126267	0.308285	0.308038	2.0565
1.0	1.110997e+08	112534202.2	65661.8	393518.6	394.920	0.000002	50.316	0.932	0.169704	0.269590	0.332422	2.3520

▼ district

gu
0 [dobong_gu, dongdaemun_gu, dongjak_gu, enpyeong_gu, gangbuk_gu, gangdong_gu, gangseo_gu, geumcheon_gu, guro_gu, gwanak_gu, gwangjin_gu, jongno_gu, jung_gu, jungnang_gu, mapo_gu, seodaemun_gu, seongbuk_gu, seongdong_gu, yongsan_gu]
1 [gangnam_gu, nowon_gu, seocho_gu, songpa_gu, yangcheon_gu]

▼ cluster_PCA5

- population_per_household, household, area_km2, man_woman_ratio, children_rate, youth_rate, middle_rate, elderly_rate
- zero_re_ratio_mean : 0.006832, 0.006043

cluster_PCA5	boarding_mean	alighting_mean	businesses_mean	employees_mean	waste_mean	waste_per_rider_mean	recycling_rate_mean	man_woman_ratio_mean	children_rate_mean	youth_rate_mean	middle_rate_mean	population_per_household_mean
0.0	6.296048e+07	6.306930e+07	41452.733333	190371.533333	257.433333	0.000003	55.022667	0.956667	0.120874	0.310913	0.306211	2.023333
1.0	9.089057e+07	9.150751e+07	55823.400000	293985.200000	455.440000	0.000004	49.956000	0.927000	0.156075	0.284995	0.322972	2.254000

▼ district

gu
0 [dobong_gu, dongdaemun_gu, dongjak_gu, enpyeong_gu, gangbuk_gu, gangdong_gu, gangseo_gu, geumcheon_gu, guro_gu, gwanak_gu, gwangjin_gu, jongno_gu, jung_gu, jungnang_gu, mapo_gu, seodaemun_gu, seongbuk_gu, seongdong_gu, yongsan_gu]
1 [enpyeong_gu, gangdong_gu, gangnam_gu, gangseo_gu, mapo_gu, nowon_gu, seocho_gu, seongbuk_gu, songpa_gu, yangcheon_gu]

▼ recycling_machine clustering

Cluster1

- 예상 인사이트: 기계 설치 밀도가 높을수록 잔여 쓰레기율이 줄고 재활용률이 높은가?
- 사용한 features: machine_per_km2, machine_per_capita, pop_density

▼ 새로운 features

```
df['machine_per_km2'] = df['recycling_machine'] / df['area_km2']
df['machine_per_capita'] = df['recycling_machine'] / df['population_total']
```

▼ 결과

지표 \ Cluster	0	1
인구밀도 평균(명/km ²)	15,831.58	18,904.73
면적 평균(km ²)	25.99	17.07
기계 밀도 평균(대/km ²)	0.25	1.63
1인당 기계 수 평균(대/명)	0.000015	0.000088
잔여 쓰레기율 평균(%)	47.25	46.05
재활용률 평균(%)	52.76	53.96
해석	• 인구밀도·기계 밀도 모두 낮고 잔여율 높음, 재활용률 낮음→ 추가 설치·접근성 개선 필요	• 인구밀도·기계 밀도 모두 높고 잔여율 낮음, 재활용률 높음→ 운영·정비 최적화로 성과 극대화 가능

▼ district

gu
0 [dongdaemun_gu, enpyeong_gu, gangbuk_gu, gangdong_gu, gangnam_gu, gangseo_gu, guro_gu, gwanak_gu, gwangjin_gu, jongno_gu, jung_gu, jungnang_gu, mapo_gu, nowon_gu, seocho_gu, seodaemun_gu, seongbuk_gu, songpa_gu, yangcheon_gu, yeongdeungpo_gu, yongsan_gu]
1 [dobong_gu, dongjak_gu, geumcheon_gu, jungnang_gu, seongdong_gu]

Cluster2

- 예상 인사이트: 설치만 했는지, 정책 채택만 했는지, 둘다 했는지 (정책이 성과로 이어졌는가? 채택 없이도 성과가 있는가?)
- 채택 했는데 성과 없으면 점검 / 채택 안했는데 성과 좋으면 자발 성공 모델 / 둘 다 안되면 지원
- 사용한 features: machine_per_km2, machine_per_capita, P2

▼ 결과

지표 \ Cluster	0.0	1.0	2.0
인구밀도 평균(명/km ²)	15,776.53	16,987.85	17,723.26
면적 평균(km ²)	21.00	30.69	18.66
기계 밀도 평균(대/km ²)	0.33	0.35	1.90
1인당 기계 수 평균(대/명)	0.000018	0.000020	0.000017
잔여 쓰레기율 평균(%)	44.23	51.07	46.84
재활용률 평균(%)	55.77	48.93	53.16

P2 채택률 평균	0.00	1.00	0.33
해석	- 정책 미채택에도 성과 우수(잔여율 44.2%, 재활용률 55.8%)→ 민·관 자발 운영 모델 벤치마킹	- 정책 채택했으나 설치 부족·성과 최저(잔여율 51.1%, 재활용률 48.9%)→ 예산·집행력 보강 필요	- 일부 채택에도 설치·성과 양호(잔여율 46.8%, 재활용률 53.2%)→ 민간 주도 성공 모델 확대 권장

▼ district

gu
0 [gangbuk_gu, gangdong_gu, geumcheon_gu, guro_gu, gwangjin_gu, jongno_gu, jung_gu, mapo_gu, nowon_gu, seodaemun_gu, yangcheon_gu, yeongdeungpo_gu, yongsan_gu]
1 [dongdaemun_gu, dongjak_gu, enpyeong_gu, gangnam_gu, gangseo_gu, gwanak_gu, seocho_gu, seongbuk_gu, songpa_gu]
2 [dobong_gu, jungnang_gu, seongdong_gu]

▼ Cluster3

▼ gap: 정책 - 기계 수 정규화 값

```
p2 = df['P2']
machines = df['machine_per_km2']
scaler = MinMaxScaler()
machines_norm = scaler.fit_transform(machines.values.reshape(-1,1)).flatten()
df['policy_machine_gap'] = p2 - machines_norm
```

- gap > 0: 정책 채택했는데, 설치 수는 적은 지역 → 성과 저조
- gap < 0: 정책 채택 없어도 설치가 다수 이루어진 지역 → 우수
- 사용한 features: 'policy_machine_gap', 'residual_ratio', 'pop_density'

▼ 결과

지표 \ 클러스터	0	1	2
pop_density_mean	20,302.09	14,060.65	14,927.15
area_km2_mean	18.49	28.65	23.76
machine_per_km2_mean	0.8189	0.2509	0.6795
machine_per_capita_mean	0.000040	0.000018	0.000043
residual_waste_ratio_mean	47.82%	51.99%	30.23%
recycling_rate_mean	52.18%	48.01%	69.78%
policy_machine_gap_mean	-0.1144	+0.5635	-0.2794
scaled_machine_mean	0.3367	0.1032	0.2794
p2_mean	22.22%	66.67%	0.00%
해석	<ul style="list-style-type: none"> - 매우 높은 인구밀도(20,302)와 설치밀도(0.82) - 정책채택률 낮음(22%) - gap<0 - 재활용·잔여율 중간 수준 <p>⇒ 주민 주도 설치 활발, 정책 채택 지원 필요</p>	<ul style="list-style-type: none"> - 낮은 인구밀도(14,061)-넓은 면적(28.65) - 정책채택률 높음(67%)-설치밀도 극히 저조 - gap>0 - 재활용률 최저-잔여율 최고 <p>⇒ 행정 실행력 강화, 인프라 조속 확충 필요</p>	<ul style="list-style-type: none"> - 중간 인구밀도-설치밀도(0.68) - 정책채택 전무(0%) - gap<0 - 재활용률 최고(69.8%) - 잔여율 최저 <p>⇒ 우수 사례(주민 자발 모델)로 벤치마킹·확산 필요</p>

▼ district

gu
0 [dongdaemun_gu, dongjak_gu, gangdong_gu, geumcheon_gu, guro_gu, jungnang_gu, seodaemun_gu, yangcheon_gu, yeongdeungpo_gu]
1 [enpyeong_gu, gangbuk_gu, gangnam_gu, gangseo_gu, gwanak_gu, jongno_gu, jung_gu, mapo_gu, seocho_gu, seongbuk_gu, seongdong_gu, songpa_gu]
2 [dobong_gu, gwangjin_gu, nowon_gu, yongsan_gu]

팀명: DRPR (District-level Recycling Policy Recommendation)

과제명: 머신러닝 기반 서울시 폐기물 클러스터링 솔루션: 유동인구·상권·재활용 정책 인프라 반영