



## 1. 주제 및 선정 배경



# 1. 주제 및 선정 배경

<38대 서울특별시시장 공약이행현황(2021년 12월 말 기준)>

☐ 사업기간: '21 **③ 빈병 무인회수기 설치**

☐ 연차별 추진 ○ 빈병무인회수기 설치 확대 추진 : 시민들의 빈병 반환 편의성 제고

- 시민중심 제로웨이스트 및 재활용 문화 확산
- ('22년) 자원순환보증금관리센터와 **간밀히 협력하여 5개소 설치**
- ('23~'26년) **미설치된 자치구 위주로 40개소 추가설치**

단위사업	연기
	'21
포장 폐기물저감 (참여 매장 수)	200
제로웨이스트 인증제 (매장 수)	10
빈병무인 회수기 설치	-
편 폴리카펫 (개최 횟수)	-
자원순환 교육 (운영 횟수)	1

5	20	65	65	65	65	65	65	-	-
2	3	3	4	5	5	7	7	9	159

	'25 상반기	'25 하반기	'26 6월	임기 후
제로웨이스트 인증제 (매장 수)	10			
빈병무인 회수기 설치	-			
편 폴리카펫 (개최 횟수)	182			
자원순환 교육 (운영 횟수)				

그러나 서울시의 제로웨이스트 프로젝트 중 “빈병 무인회수기 설치” 공약의 구체적인 실행 계획이 부족함을 발견함



# 1. 주제 및 선정 배경

## 분석의 방향성



### 첫번째 목표

제로웨이스트 서울을 위한 최적의 “빈병 무인회수기” 입지 선정



### 추가적인 고려사항

빈병 무인회수기 입지 선정 과정에 “취약 계층의 접근성”을 변수로 고려하고자 함



# 1. 주제 및 선정 배경

## 분석의 방향성



### 두번째 목표

학습을 통한 빈병 무인회수기 바코드 인식 정교화 시도

바코드 인식기가 병의 바코드를 정확하게 인식할 수 있도록  
훼손된 바코드의 이미지를 복원하는 이미지 처리 학습을 진행





## 2. 전처리 및 시각화

1. 무인회수가능 공간
2. 음식점 수 & 주점 수
3. 생활인구/거주인구
4. 구별 쓰레기 배출량
5. 환경 인식 지수



## 2. 전처리 및 시각화

### 인구

#### 최종 데이터셋

시군구	행정동	행정동코드	전체거주인구	전체세대	1인세대	그외세대	...	생활인구
종로구	청운효자동	11110515	12177	5241	2000	3241	...	16571.55
종로구	사직동	11110530	9636	4696	2356	2340	...	22968.57
종로구	삼청동	11110540	2739	1229	625	604	...	5550.663
...	...	...	...	...	...	...	...	...

427X 13

시군구와 행정동별 각종 인구 데이터 완성



## 2. 전처리 및 시각화

### 구별 쓰레기 배출량

#### 최종 데이터셋

공병 재활용률 관련 파생변수!

시군구	재활용 배출량	혼합 배출량	합계	재활용 비율	재활용_유리병 _비율	서울_유리병 _비율
종로구	26384.8	79444.8	92175.3	0.286	0.07	0.024
강남구	69426.0	172032.2	236784.5	0.293	0.08	0.065
광진구	46604.7	75922.2	113683.8	0.410	0.05	0.028



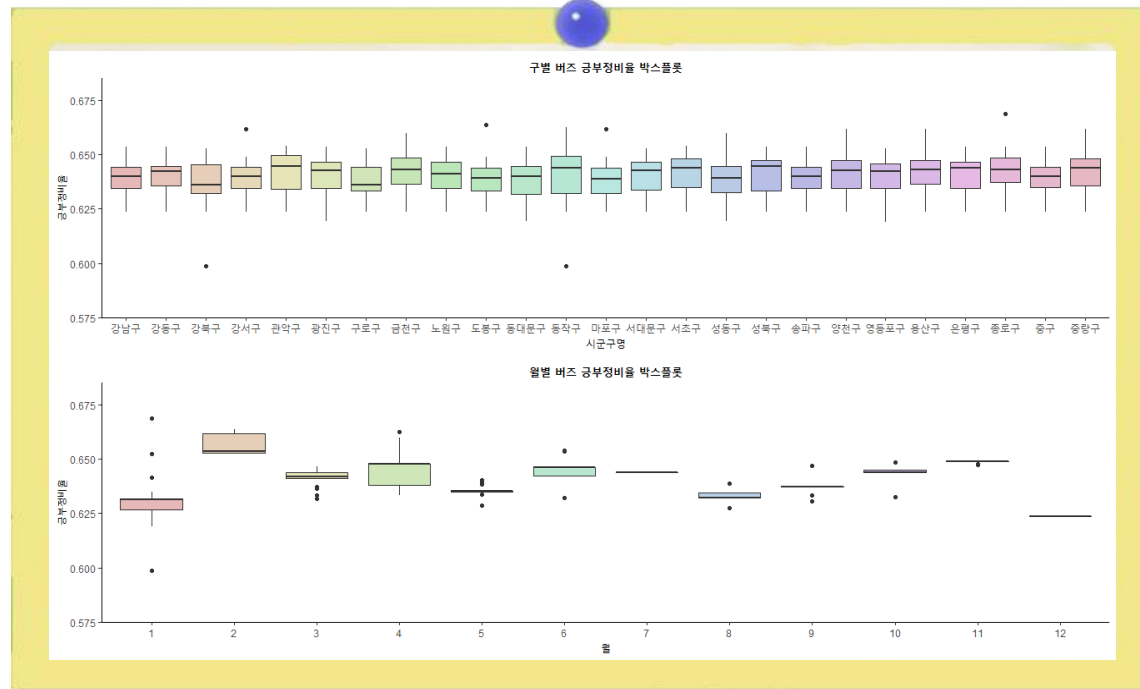


## 2. 전처리 및 시각화

### 환경 인식 지수

#### 환경 소셜 버즈 - 시각화를 통한 긍부정 평균 분석

월별\_긍부정\_총환경소셜버즈.csv



데이터를 연간데이터로 합치기 전에  
구별, 월별 긍부정 비율 박스플롯을 그려 확인

대체적으로 월에 대해서는 차이가 있어 보이지만,  
구에 대해서는 차이가 없는 것으로 보임.

→ ANOVA를 진행하여 확인



### 환경 인식 지수

#### 환경 소셜 버즈 - 금부정비율에 대한 ANOVA

월별\_금부정\_총환경소셜버즈.csv

RCBD 모델

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

$\mu = \text{grand mean}$

$\tau_i = i^{\text{th}} \text{ treatment effect (구)}$

$\beta_j = j^{\text{th}} \text{ block effect (월)}$

$$\sum_{i=1}^{25} \tau_i = \sum_{j=1}^{12} \beta_j = 0, \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

구를 treatment effect로 잡고,

월을 block effect로 잡아

RCBD 모델을 사용함

모든 구와 월에 대한 조사이기 때문에

Random Effect는 없다고 가정

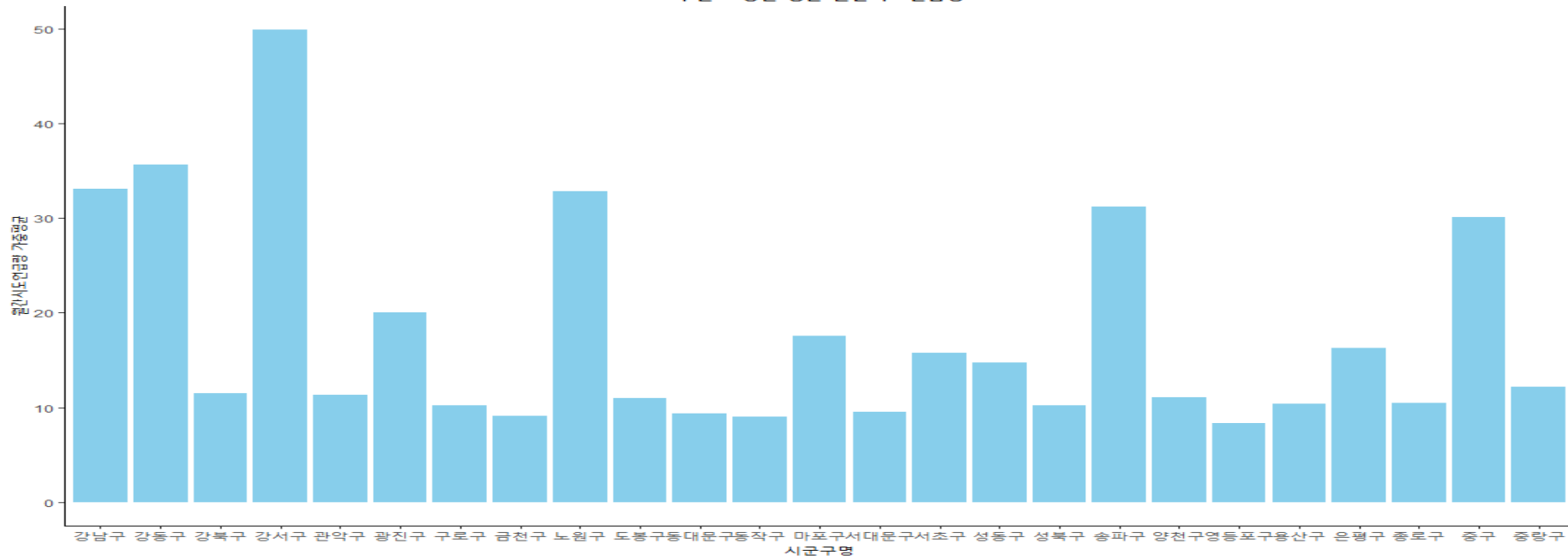


## 2. 전처리 및 시각화

### 환경 인식 지수

#### 금부정 비율로 조정한 월간 시도 언급량

구별 조정된 평균 월간시도언급량



구별로 조정된 평균 월간시도언급량을 확인하면

다른 구보다 확연하게 많이 언급된 구가 있다는 것을 확인 할 수 있음.



## 4. 회귀 모델링



## 4. 회귀모델링

### 회귀 모델링 흐름

#### 회귀모델링 흐름 정리

Train data(x) 강서구의 지역적 특성을 담은 변수

시군구	행정동	음식점 수	...	생활인구 수
강서구	가양1동	605		
강서구	가양2동	27		
강서구	가양3동	25		
	...	...		

X variables



강서구 행정동별  
쓰레기 배출량을  
예측하는 다중선형  
회귀모델 생성

Train data(y) 강서구의 쓰레기 배출량

행정동	쓰레기 배출량
가양1동	925,380
가양2동	31,240
가양3동	181,270
...	...

Y variable



### 회귀 모델 성능

#### 회귀모델링 시작

MODEL 1: 상업지수(인구), 주점 수, 주택 수, 음식점 수

변수 스케일링 진행!

- MODEL 2: 상업지수(인구), 주점 수, 주택 수, 음식점 수 → standard scaling

- MODEL 3: 상업지수(인구), 주점 수, 주택 수, 음식점 수 → minmax scaling

성능	MODEL1	MODEL2	MODEL3
F-stats	6.701	6.701	6.701
R-square	0.516	0.516	0.516

세 모델 모두 F-stats와 R-square값이 비슷하고 유의하다는 결과 도출





### 회귀 모델 처방

#### 회귀 모델링 - 회귀 모델 진단과 처방

정리하자면,

- ✓ 학습데이터의 수가 너무 적어 모델의 성능이 떨어짐
- ✓ 학습 데이터 수에 비해 너무 설명 변수가 많음
- ✓ 설명 변수를 최소한으로 학습 한 경우에는 F-test와 T-test를 통과하지만, 등분산성을 만족하지 못하는 경우가 발생함.





## 0. 전처리 마무리

1. 환경인식지수
2. 공시지가
3. 취약계층



## 0. 전처리 마무리

### 환경인식지수

#### 최종 데이터셋 - 환경인식지수

자치구	MM시도언급량	MM탄소배출량	MM재활용률	환경인식지수
강남구	0.595709	0.614535	0.385465	0.398322
강동구	0.657864	0.838101	0.482759	0.434174
...	...	...	...	...
중랑구	0.092328	0.284325	0.610345	0.472783

#### 환경인식지수

$$\frac{b + (1 - c) + r}{3}$$

$b$  = Min-Max 스케일링한 시도 언급량

$c$  = Min-Max 스케일링한 탄소 배출량

$r$  = Min-Max 재활용률

Min-Max 스케일링을 진행한 변수로 환경인식지수 완성  
이때, 탄소배출량은 작을수록 좋으므로 1-(스케일링된 값) 사용



# 1. 클러스터링

1. 무인 공병 회수기

2. 취약계층





## 1. 클러스터링

# 무인 공병 회수기

[illegible]

변수 조합별, 스케일링 종류별,  
클러스터링 종류별 모두 실행

변수 종류	스케일링 종류	클러스터링 종류
쓰레기 배출량 비율	None	K-means
거주 인구 비율	Minmax	K-medoids
인 가능 공간 개수 비율'에서 변경 ... 무인 공간 개수 비율	Standard	GMM
		Hierarchical

1주차 '무인 가능 공간 개수 비율'에서 변경 ...



가능한 모든 경우의 수로 클러스터링 진행 → 36가지의 경우의 수 비교





## 1. 클러스터링

### 무인 공병 회수기

#### 최적의 클러스터

시군구	행정동	쓰레기 배출량 비율	거주 인구 비율	무인 공간 개수 비율
중구	청구동	198.10	0.534	0
용산구	용문동	206.53	0.570	0
성동구	행당 2동	275.99	0.532	1
성북구	돈암 2동	234.63	0.610	2
성북구	길음 1동	205.17	0.501	2

무인공병회수기 클러스터링을 통해 최종적으로 추출된 16개 행정동의 일부 데이터



# 1. 클러스터링

## 취약계층

### <Preview>

변수 종류	스케일링 종류	클러스터링 종류
공시지가 & 고령자 비율 & 수급자 비율	Minmax / Standard	K-means
공시지가 & 고령자 비율	Minmax / Standard	K-means
공시지가 & 수급자 비율	Minmax / Standard	K-means
고령자 비율 & 수급자 비율	Minmax / Standard	K-means



더불어 공시지가의 경우 Scale이 매우 크기 때문에 Scaling을 고려하지 않은 경우는 분석에서 제외!



# 1. 클러스터링

## 취약계층 클러스터링

시군구	행정동
종로구	창신1동
용산구	남영동
용산구	보광동
성동구	송정동
...	...
강동구	강일동
강동구	상일2동



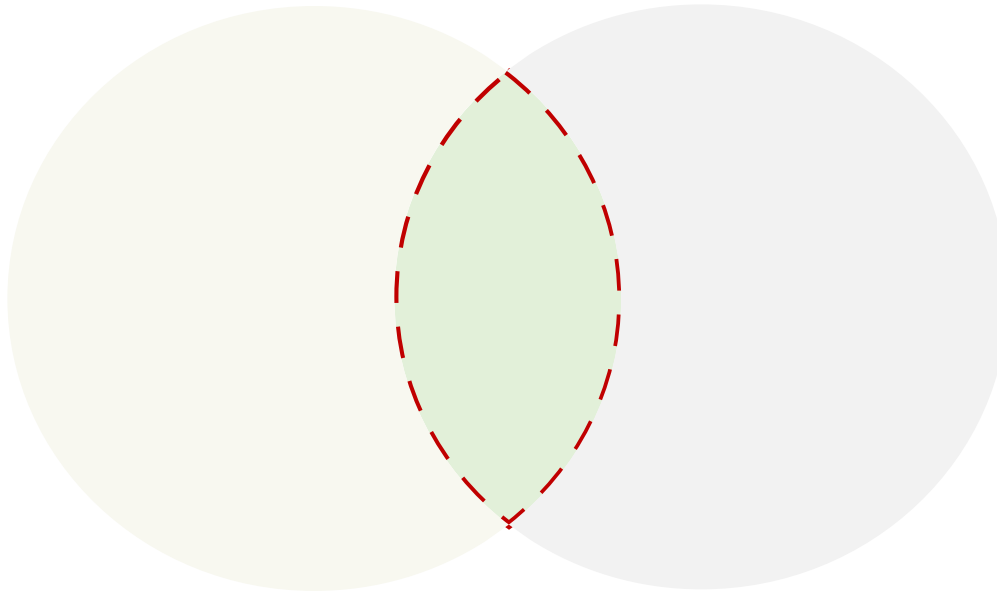
취약계층 클러스터링을 통해 최종적으로 추출된 61개의 행정동



## 최종 클러스터링

### 최종 클러스터링

무인공병회수기  
클러스터링  
16개 행정동



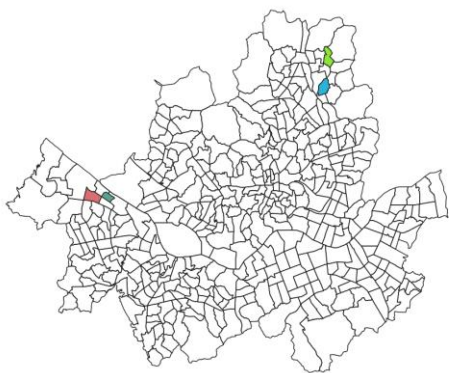
취약계층  
클러스터링  
61개 행정동

무인 공병 회수기를 설치할 최종 행정동은 무인 가능 공간 클러스터링과  
취약계층 클러스터링에 모두 포함된 행정동을 선정할 예정.



# 1. 클러스터링

## 최종 클러스터링



### 최종 4개 행정동 선정

노원구 중계 2.3동, 노원구 상계5동,  
강서구 등촌3동, 강서구 가양3동

## 최종 클러스터링

서울시 전체 425개 행정동 평균

	고령자비율	전체거주인구	공시지가	수급자비율
mean	0.168	22854	4528835	0.041

최종 4개 행정동 평균

	고령자비율	전체거주인구	공시지가	수급자비율
mean	0.223	26063	3348531	0.140

무인공병회수기 클러스터링과 취약계층 클러스터링 타겟에  
모두 속하는 행정동 추출 → 최종적으로 4개의 행정동이 선정



## 2. 입지최적화





## 2. 입지최적화

### 무인공병회수기 설치 계획을 반영한 P-median 입지최적화 모델

$$\text{minimize } \sum_i \sum_j h_i d_{ij} y_{ij}$$

$$\sum_j y_{ij} = 1, \sum_j x_j = p, y_{ij} \leq x_j, y_{ij} \in (0,1), x_j \in (0,1) \text{ for all } i, j$$

where  $i$  = 행정동 내 수요지(주거건물),  $j$  = 행정동 내 시설 입지 후보(편의점),

$h_i$  = 주거건물의 연면적(가중치),  $d_{ij}$  = 주거건물과 편의점 간 거리,  $p$  = 편의점의 수

$x_j$  = 편의점  $j$ 의 무인공병회수기 설치 여부,  $y_{ij}$  = 무인공병회수기의 총 수요 충족 여부

이때 수요량으로 주거건물 별 세대수를 사용할 예정이었으나,  
데이터를 구할 수 없어 이를 대체할 수 있는 연면적을 수요량으로 하여 모델의 가중치로 사용



## 2. 입지최적화

### 입지최적화 ① - P-median을 통한 노원구 입지 최적화

상계 5동 p-median pmp

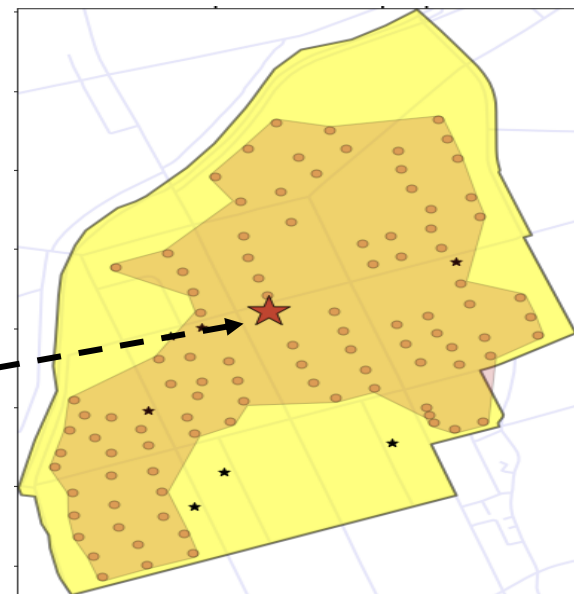


GS25 상계대박점

GS25 상계역점

GS25 중계무지개점

중계 2,3동 p-median pmp



P-median 모델을 통해 무인공병회수기를 설치할 3개의 최적의 편의점을 선정



### 3. 토이 프로젝트



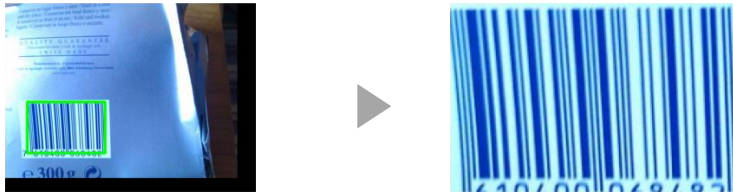
### 3. 토이 프로젝트

## 토이프로젝트 계획

무인공병회수기의 다른 문제점으로 바코드 인식률 문제가 존재. 공병 바코드가 훼손되거나 변질된 경우에는 회수기가 일련번호를 인식하지 못해 공병을 회수하지 못하는 문제가 발생. 이를 해결하고자 훼손된 바코드를 원본 바코드로 복원하는 프로젝트 진행.

#### 1. 사물 이미지에서 바코드 객체를 detection

정확도를 위해 Object detection 모델 중 two-stage 모델에 해당되는 FastRCNN 사용



epoch=3, train\_set: 300개 image, valid\_set: 65개 이미지

Train Loss: 0.5343

Valid Loss: 0.7890

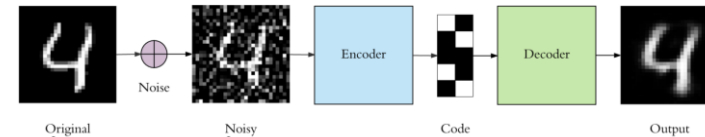
#### 2. crop된 바코드 이미지에 노이즈를 부여하여 이를 원본 이미지로 복원

##### 1) Noise method

Gaussian Method

Salt and Pepper Method

##### 2) DAE(denoising auto encoder)



Final model: conv - pooling - conv - pooling - conv - pooling

