# 수원시 스마트 버스 정류장 우선 설치위치 선정

[Team B3A1]

# CONTENTS


## 문제 정의
배경

## 목표 설정
고려요인
분석방법 설정

## DATA
데이터 Filter
데이터 전처리

## MODEL
SPCA
K-means Clustering
PLS Regression

## 결론
결과 및 해석
결론

# 1

# 문제 정의

# 배경

## 1. 수원시 현황

### 1) **스마트 챌린지 솔루션 확산 사업 대상지** 선정

수원시는 "시민 삶에 플러스가 되는 스마트 시티 솔루션 플러스
(**스마트 버스 정류장**, 스마트 횡단보도, 공유주차)" 를 주제로 응모해 선정

스마트버스정류장

스마트공유주차

스마트횡단보도

**스마트 버스 정류장**

**AI 음성인식/안내**
기상특보, 재난재해, 음성호출

**힐링쉼터**
냉난방, 공기질 측정, 체온 측정

**편의시설**
휴대폰 충전
무료 와이파이

스마트 챌린지 솔루션 확산 사업은 오는 4월에 시작해 **올해 안에 마무리할 계획**
"**스마트 도시**" 로 자리매김하는 발판을 마련

# 배경

## 1. 수원시 현황

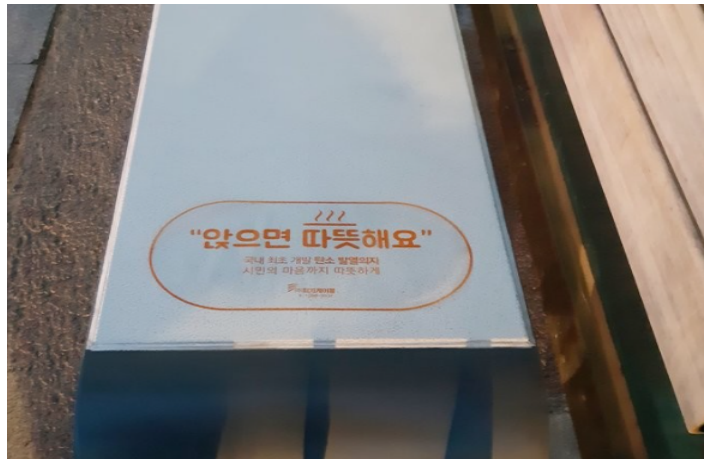
### 2) **교통 서비스의 품질을 높이기** 위한 지속적인 노력

**"수원시, 광역버스 졸음운전 탐지해 시민 안전 챙긴다"**

"**2021년 디지털 공공서비스 혁신프로젝트"에 선정**되어 국비 확보

졸음운전 탐지 및 안전 서비스 구축을 통해 버스 운전기사를 관리하고
운수종사자, 수원 시민, 경기도민의 안전을 확보하는 것이 목표

**최강 한파 속 수원시 버스정류장 온열의자 인기**



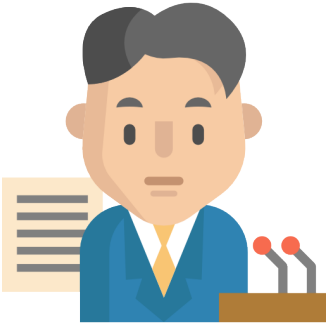

시민들이 버스를 기다릴 때 이용할 수 있도록
2015년부터 **온열의자**를 설치하여 운영 중

고효율 탄소 소재 발열체가 내장되어
소량의 전기로 최대 40도까지 발열

# 배경

## 1. 수원시 현황

### 3) 미세먼지 감축을 위한 지속적인 노력과 투자



"지속가능한 도시생태계를 조성하겠습니다."

"그린 인프라 구축과 도시 숲 조성으로 **미세먼지 저감**을 위한 도시환경 문제에 **적극 대응**해 갈 것입니다."

-2021년 수원시장 신년사 中

**미세먼지 감축을 위한 수원시의 주요 사업과 정책**

- 미세먼지 **계절 관리제** 시행
- 저공해 차량 보급 및 충전 인프라 확대
- 도시숲 확대 및 생태공원 조성
- 대기오염 정보 공개 (**수원시 대기질 알리미**)
- 노후 경유차 조기폐차 및 LPG 화물차 신차 구입에 관한 차량 관련 보조금 지급
- 2024년까지 2016년 대비 **초미세먼지 연평균농도 25% 이상 저감** 목표
- 친환경 전기자동차 보급 및 보조금 지원

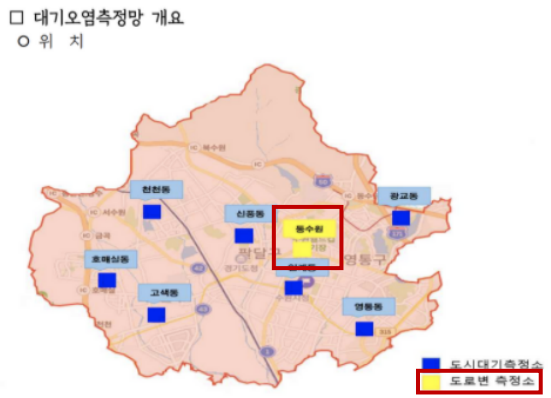
# 배경

## 2. 문제 제기

1) 수원시 미세먼지 저감정책의 해결책, **스마트버스 정류장**

**Q1.** 현재 수원시 내 **미세먼지 측정소는 8곳 뿐**이며, 특히 미세먼지가 심한 **도로변의 측정소는 1곳 뿐**임.

**A1.** 스마트 버스 정류장 설치를 통한 **도로변 미세먼지 측정소 확대**




**Q2.** 버스정류장 중앙차로의 이산화질소 농도는 도심지역의 2.5배~ 5.3배에 달해
**버스 이용자가 승차대기 중 받게 되는 대기오염 피해는 상당히 높음**

**A2.** 스마트 버스 정류장을 통해 주변 오염물질 유입을 방지하고, 공기정화장치 등을 통해 공기질 개선
실제로, 고양향동지구 내 설치한 스마트 버스 정류장에서 측정한 초미세먼지 농도 조사에 따르면
**10% ~ 33%의 초미세먼지 저감효과**를 보임

# 배경

## 2. 문제 제기

### 스마트 버스 정류장이란?

LH 스마트클린버스쉘터

미세먼지, 차량배기 가스 등
정류장 주변 오염물질 유입 방지

공기정화장치 등을 통해
버스정류장 내 공기질을 개선하는 시설

Wi-Fi 및 범죄 예방을 위한 안심벨 설치

**"미래형 교통 인프라의 초석"**

# 2
# 목표 설정

# 고려 요인 설정

## 1. 중심 요인

### 1) 중심 요인 설정 배경

"서울시는 최첨단 ICT기술, 신재생에너지, 공기청정시설등 다양한 기능이 집약된 **미래형 버스정류소를 설치**한다고 밝혔다."

"**후보지는 이용현황을 바탕**으로 합정역, 홍대입구역, 왕십리광장등 10개소가 선정됐다."

**서울시 스마트버스쉘터 담당자 전화 인터뷰** (서울시 정류소관리팀 김**님)

서울시 스마트 쉘터를 설치하는데 **주로 고려하신 사항**은 무엇인가요?

중앙버스정류소 위주로 설치하였고 **사람이 많은 곳, 지역의 상징성**을 고려하여 선정하였습니다.

주요기능 중 하나인 **미세먼지 차단과 정보수집** 측면에서 무엇을 우선적으로 고려하셨나요?

**심각한 미세먼지 오염으로부터의 대피**를 통한 시민 건강 증진에 초점을 두고 사업을 진행하였습니다.

정류장별 **광고 수익**을 위해 무엇을 고려하셨나요?

사람이 많이 모이는 노선들을 중점으로 **유동인구**를 주요하게 고려하였습니다.

감사합니다!

# 고려 요인 설정

## 1. 중심 요인

2) 중심 요인 설정

| | |
|---|---|
| 유동인구 | 서울시 스마트 버스쉘터를 선례로 참고하여,<br>잠재적 스마트 버스 정류장 이용자 혹은 광고 타겟층인<br>**유동인구**를 주요변수로 분석 |
| 초승+환승 수 | 최대한 많은 시민이 해당 버스정류장을 이용하도록<br>**실질적인 사용량**을 고려하여 분석 |
| 전체 교통량 | 교통량이 많을수록 미세먼지 배출량이 높고<br>대기시간동안 노출도도 커지므로<br>미세먼지 노출도를 **전체 교통량**과 연관시켜 분석 |

# 고려 요인 설정

## 2. 기타 고려 요인

1) **취약 계층**을 고려한 위치 선정으로 **공공의 이익** 실현

| 국내 정책 현황 |
| --- |
| - 생물학적요인 (영유아, 어린이, 65세 이상 고령자)을 대상으로 하는 주요 정책은 이미 수립되어 있음 |
| **국내 정책 한계점** |
| - 스스로 보호하기 어려운 **사회적 민감군(저소득계층)에 대한 지원이 미약**한 실정 |

출처: 수원시 대기오염 민감군 피해저감 방안 연구 (수원시정연구원 연구책임자 강은하)

| 해결방안 |
| --- |
| 미세먼지 취약계층이자 미세먼지 정보를 상대적으로 접하기 어려운 **고령인구와 저소득계층**도 함께 고려한 분석 진행 |
| **"소외 계층 없는 공공의 이익 실현"** |

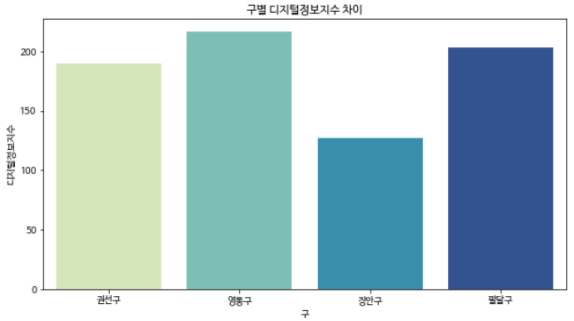
# 2 분석 방법 설정

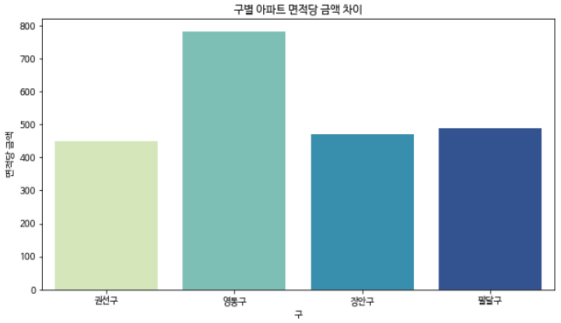
목표설정

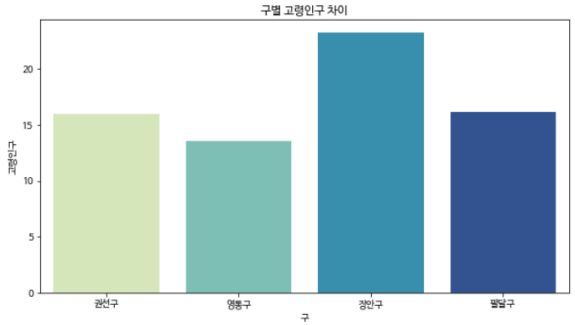
## 1. 구별로 나누어 분석 진행

- 각 구별 특성이 다르기 때문에 **차이점을 반영**하여 분석하고자, 구별로 나누어 분석
- **[공공성] 최대한 많은 시민들**이 스마트 버스 정류장을 경험할 수 있도록 **골고루 배치**

| 디지털정보지수 | | 아파트 면적당 금액 | | 고령인구 | |
|---|---|---|---|---|---|
| 1등 | 영통구 | 1등 | 영통구 | 1등 | 장안구 |
| 4등 | 장안구 | 3등 | 장안구 | 4등 | 영통구 |



# 분석 방법 설정

## 2. 스마트 버스 정류장 광고 타겟팅

### 광고 타겟팅 방법

스마트 버스 정류장 광고를 **성별, 연령대별로** 타겟팅하기 위해
**성별, 연령대별 유동인구**를 사용해서 분석

성별에 따라 광고에 영향 받는 정도가 다르다고 생각해
**(성별에 따른 가중치) X (유동인구 값)**
으로 새로운 **광고 타겟 유동인구 지수** 생성

관여도와 성별에 따른 구매의도의 평균

| | | 성별 | | |
|---|---|---|---|---|
| | | 남성 | 여성 | 전체 |
| 관여도 | 고관여 | 1.84 | 2.92 | 2.38 |
| | 저관여 | 3.97 | 5.62 | 4.79 |
| | 전체 | 2.94 | 4.27 | 3.59 |

**관여도**: 제품이나 이슈, 상황 등에 대해 개인이 부여하는 중요도의 정도 또는 관심의 정도
(**고관여**: 관심이 많은 경우, **저관여**: 관심이 적은 경우)

목표와 관계없는 변수이기때문에
관여도 항목 중 '전체' 항목의 값들 이용

여성

(해당 나이 대 여성 유동인구 수)
X
4.27
= (광고 타겟 여성 유동인구)

남성

(해당 나이 대 남성 유동인구 수)
X
2.94
= (광고 타겟 남성 유동인구)

출처: 성별과 관여도에 따른 옥외광고의 커뮤니케이션 효과(한광석, 2005)

# 3

# DATA

# MODEL FLOW

## Data Preprocessing

데이터 Filter

- 인도 폭
- BIS 설치 유무

데이터 전처리

- 내부 데이터 병합
- 외부 데이터 병합
- 파생변수 생성

버스정류장 데이터

## Modeling

SPCA

K-means Clustering

Target 클러스터

PLS Regression

- 추정치를 이용한 스마트지수 생성

## 최종 설치 입지 선정

### 설치위치 선정

스마트 지수

구별 순위

Constraint

- 구별 정류장 비율
- 정류장 간 거리

### 최종 결과

통합 순위

Constraint

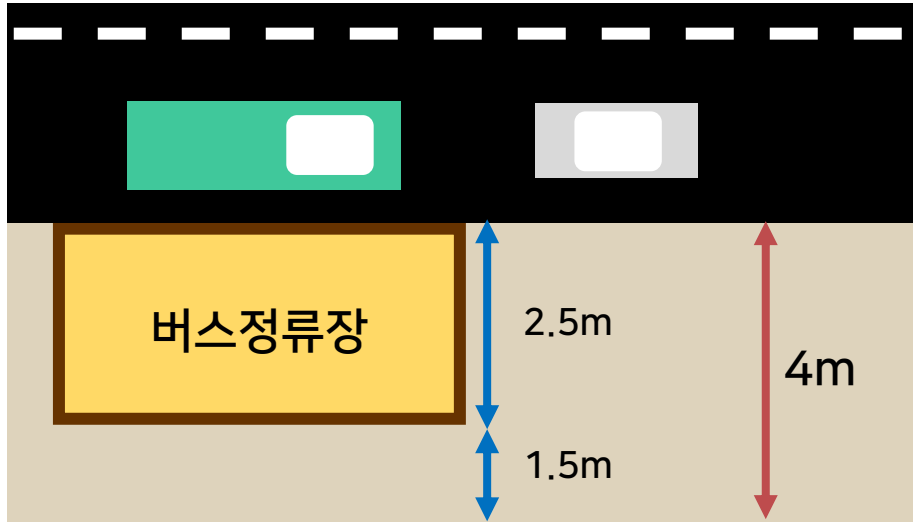
광고 Targeting

- 성별 가중치 부여

# Data Filtering

## 1. 인도 폭 이용해 후보 버스정류장 Filtering

- 버스정류장의 크기는 5.5m X 2.5m로 규정
- 인도 폭 1.5m 이상을 확보하기 위해 인도 폭이 4m 이상인 곳의 버스정류장만을 선택




| 버스정류장 폭 | | 선택 인도 폭 |
|---|---|---|
| 2.5m 이상 | + | 4m 이상 |
| 인도 확보 폭 | | |
| 1.5m 이상 | | |

3

DATA

# Data Filtering

## 1. 인도 폭 이용해 후보 버스정류장 Filtering

### CASE 1. 인도폭이 4m 이상인 경우

#### STEP1. 폭이 4m 이상인 인도 filter

| 인도 ID | 인도 폭 | geometry |
|---|---|---|
| 154931 | 4 | multilinestring |
| 154932 | 4.5 | multilinestring |
| 154947 | 6 | multilinestring |
| 154993 | 9.4 | multilinestring |

26.수원시_인도(2017).geojson
데이터를 인도폭 변수(WIDT)를 기준으로
4 이상인 인도만 filter

#### STEP2. 폭을 이용해 buffer 생성

QGIS에서
filter된 인도 데이터에 대해
미터 단위의 인도폭으로 버퍼 생성

#### STEP3. 버스정류장 filter

| 정류소 ID | 정류장명 |
|---|---|
| 200000006 | 광교공원 경기대<br>수원캠퍼스 입구 |
| 200000008 | 문암골 |
| 200000036 | 풍림아파트 입구 |
| 200000047 | 조원뉴타운 |

QGIS에서
인도 버퍼 내에 있는
버스정류장만 filter

# Data Filtering

## 1. 인도 폭 이용해 후보 버스정류장 Filtering

**CASE 2. 인도폭이 0m인 경우**

인도폭이 0m인 경우 측정되지 않은 값이라고 생각하여, 분석 데이터에 포함

### STEP1. 폭이 0m인 인도 filter

| 인도 ID | 인도 폭 | geometry |
|---|---|---|
| 154767 | 0 | multilinestring |
| 154768 | 0 | multilinestring |
| 154769 | 0 | multilinestring |
| 154929 | 0 | multilinestring |

26.수원시_인도(2017).geojson 데이터를 인도폭 변수(WIDT)를 기준으로 0인 인도만 filter

### STEP2. 폭을 설정해 buffer 생성

QGIS에서 filter된 인도 데이터에 대해 주어진 데이터내의 최대 인도폭으로 버퍼 생성

### STEP3. 버스정류장 filter

| 정류소 ID | 정류장명 |
|---|---|
| 203000439 | 광교카페거리입구 |
| 203000357 | 호반베르디움 |
| 203000344 | 광교중학교 |
| 203000317 | 광교대광로제비앙 |

QGIS에서 인도 버퍼 내에 있는 버스정류장만 filter

# Data Filtering

## 2. BIS 설치 유무 이용해 후보 버스정류장 Filtering

- BIS 시스템 설치 정류장은 전기 및 통신공사가 완료
- 전기와 통신의 추가 공사는 불가하기 때문에 **이미 BIS 시스템이 설치된 정류소 선택**

### STEP1. 인도 filter 데이터 merge

| 정류장 ID | BIS 설치여부 | geometry |
|---|---|---|
| 201000344 | 0 | multilinestring |
| 201000268 | 1 | multilinestring |
| 201000055 | 0 | multilinestring |
| 202000092 | 1 | multilinestring |

인도폭 기준으로 필터링 된
두 case의 버스정류장 데이터를
row 기준으로 merge

### STEP2. BIS 변수 1인 데이터 filter

| 정류장 ID | BIS 설치여부 | geometry |
|---|---|---|
| 201000345 | 1 | multilinestring |
| 201000268 | 1 | multilinestring |
| 202000090 | 1 | multilinestring |
| 202000092 | 1 | multilinestring |

QGIS에서 filter된 버스정류장 데이터에
대해 BIS 설치여부 변수가 1인
버스정류장만 filter

### STEP3. 데이터 저장

bus_filter_final.geojson

QGIS에서
filter된 버스정류장 데이터를
bus_filter_final.geojson으로 내보내기

3

DATA

# 버스 정류장 데이터

## 사용 데이터

- 버스 정류장 입지 선정을 위해, 버스 정류장을 기준으로 데이터 병합
- **버스 정류장 고유 특성 / 버스 노선의 특성 / 일자별 특성**을 모두 반영하기 위하여 데이터를 세분화하여 전처리

**버스 정류장 (516)**

- 수원시 버스정류장
- 수원시 버스정류장별 노선현황
- 수원시 지하철역 위치정보 + 이용현황
- 수원시 시간대별/성연령별/요일별 유동인구
- 수원시 인구정보 격자
- 수원시 교통량/혼잡시간강도/혼잡빈도강도
- 수원시 부동산 실거래가 (외부데이터) : 국도교통부 수원시 구별 실거래가
- 경기도 상권 데이터 (외부데이터) : 공공데이터포털 소상공인시상진흥공단 상가(상권) 데이터
- 수원시 관광명소 데이터 (크롤링) : 수원시 홈페이지 > 도시사용설명서

**버스 노선 (3926)**

- 경기도 버스 노선 순서 (외부데이터) : 경기버스정보 > 버스 노선 정보
- 경기도 버스 노선 정보 (외부데이터) : 경기버스정보 > 노선 경유 정보

**일자별 특성 (1326103)**

- 수원시 버스정류장별 승하차이력 (주중/주말/공휴일 데이터 세분화)
- 수원시 일자별 기상 정보 (외부데이터) : 기상자료개방포털 > 종관기상관측(ASOS)
- 수원시 일자별 미세먼지 정보 (크롤링) : 에어코리아 > 시도별 대기정보 (PM2.5 & PM10)

### 버스 정류장 (516, )

| 정류소ID | 정류장명 |
|---|---|
| 200000006 | 광교공원.경기대수원캠퍼스입구.연무시장 |
| 200000008 | 문암골 |
| 200000036 | 풍림아파트입구 |

### 버스 노선 (3926, )

| 정류소ID | 정류장명 | 노선ID | 노선번호 |
|---|---|---|---|
| 200000006 | 광교공원.경기대수원캠퍼스입구.연무시장 | 200000037 | 13 |
| | | 200000099 | 37 |
| | | 200000045 | 7_1 |

### 일자별 특성 (1326103, )

| 정류소ID | 정류장명 | 일자 | 노선ID | 노선번호 |
|---|---|---|---|---|
| 200000006 | 광교공원.경기대수원캠퍼스입구.연무시장 | 2020-01-01 | 200000037 | 13 |
| | | | 200000099 | 37 |
| | | | 200000045 | 7_1 |

# 3 외부 데이터 병합

DATA

## 경기 버스 정보 (외부데이터 : 버스 노선 정보 & 노선 경유 정보)

- 노선 순서, 배차 간격, 첫차·막차 정보 등 **다양한 버스 노선 정보** 반영

| 정류소ID | 정류장명 | 노선번호 |
|---|---|---|
| 200000006 | 광교공원.경기대수원캠퍼스입구.연무시장 | 13 |
| | | 37 |
| | | 7_1 |
| 200000008 | 문암골 | 13 |
| 200000036 | 풍림아파트입구 | 112 |
| 200000047 | 조원뉴타운 | 4_1 |
| | | 88 |
| 200000066 | 수성중사거리 | 112 |
| | | 19 |
| | | 37 |

수원시_버스정류장별_승하차이력 (1) ~ (5).csv
(1364613, 25)

정류소ID, 노선번호 기준 merge

| 정류소ID | 노선번호 | 노선ID | 노선순서 |
|---|---|---|---|
| 200000006 | 13 | 200000037 | 13 |
| | 37 | 200000099 | 1 |
| | 7_1 | 200000045 | 55 |
| 200000008 | 13 | 200000037 | 11 |
| 200000036 | 112 | 200000049 | 29 |
| 200000047 | 4_1 | 200000101 | 46 |
| | 88 | 200000107 | 56 |
| 200000066 | 112 | 200000049 | 24 |
| | 19 | 200000266 | 67 |
| | 37 | 200000099 | 9 |

GGD_RouteStationInfo_M.xls
버스 노선 정보
(192708, 5)

노선ID, 노선번호 기준 merge

| 정류소ID | 노선ID | 노선번호 | 주중/주말배차간격 | 상/하행 첫차 | 상/하행 막차 |
|---|---|---|---|---|---|
| 200000006 | 200000037 | 13 | 5분~7분 / 5분~7분 | 06:00 / 05:40 | 22:00 / 23:10 |
| | 200000099 | 37 | 10분~13분 / 15분~20분 | 04:50 / 06:00 | 21:40 / 23:10 |
| | 200000045 | 7_1 | 7분~9분 / 9분~12분 | 05:00 / 06:00 | 22:40 / 23:45 |
| 200000008 | 200000037 | 13 | 5분~7분 / 5분~7분 | 06:00 / 05:40 | 22:00 / 23:10 |
| 200000036 | 200000049 | 112 | 7분~9분 / 9분~12분 | 05:00 / 06:05 | 22:50 / 23:55 |
| 200000047 | 200000101 | 4_1 | 30분~35분 / 30분~40분 | 05:00 / 05:00 | 22:30 / 23:20 |
| | 200000107 | 88 | 12분~14분 / 20분~28분 | 04:40 / 05:20 | 22:50 / 23:30 |
| 200000066 | 200000049 | 112 | 7분~9분 / 9분~12분 | 05:00 / 06:05 | 22:00 / 23:55 |
| | 200000266 | 19 | 15분~20분 / 18분~20분 | 05:00 / 06:00 | 22:00 / 23:00 |
| | 200000099 | 37 | 10분~13분 / 15분~20분 | 04:50 / 06:00 | 21:40 / 23:10 |

GGD_RouteInfo_M.xls
노선 경유 정보
(3480, 15)

# 3 DATA

# 파생 변수 생성

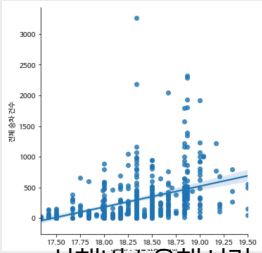
## 1. [운행시간, 배차횟수]

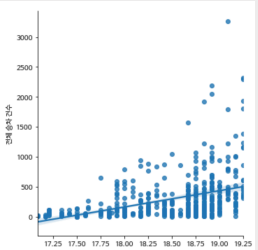
### 운행시간

정류소 운행시간 = (버스 정류장에 가장 늦게 오는 버스의 막차 시간) – (가장 빨리 오는 버스의 첫차 시간)

| 정류소ID | 노선ID | 노선번호 | 상행 첫차 | 상행 막차 |
|---|---|---|---|---|
| 200000006 | 200000037 | 13 | 06:00 | 22:00 |
| | 200000099 | 37 | 04:50 | 21:40 |
| | 200000045 | 7_1 | 05:00 | 22:40 |

**정류소 운행 시간**
22:40 – 04:50
= 17 시간 50분
(= 17.8333 시간)


상행버스운행시간


하행버스운행시간

**운행 시간**이 증가할수록
**전체 승차 건수**가
증가하는 경향을 보임

### 배차횟수

정류소 배차횟수 = ∑ (각 버스 노선의 운행 시간 / 배차 간격)

| 정류소ID | 노선ID | 노선번호 | 운행 시간 | 배차 간격 | 배차 횟수 | 배차간격_15분 | 배차간격_60분 |
|---|---|---|---|---|---|---|---|
| 200000006 | 200000037 | 13 | 16시간 | 6분 | 160 | 1 | 0 |
| | 200000099 | 37 | 16시간 50분 | 11.5분 | 87.82 | 1 | 0 |
| | 200000045 | 7_1 | 17시간 40분 | 8분 | 132.5 | 1 | 0 |

10분~13분 형식으로 되어 있는 경우,
(10+13)/2 = 11.5분으로 전처리

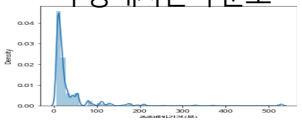
**정류소 배차 횟수**
160 + 87.82
+ 132.5
= 380.32

주중 배차간격과
주말 배차간격이
상이함을 고려하여,

주중 / 주말 / 공휴일
세 가지 데이터로
세분화하여
파생변수 생성

**주중배차횟수**
**주말배차횟수**
**공휴일배차횟수**

주중배차간격 분포



주중 배차 간격이 15분 이하인 버스 : 47개
주중 배차 간격이 60분 이상인 버스 : 27개

주말 배차 간격이 15분 이하인 버스 : 28개
주말 배차 간격이 60분 이상인 버스 : 39개

버스의 특성에 따라 배차 간격이 다름
**15분, 60분을 기준**으로 버스 배차 간격의 길고 짧음을 판단

**배차 간격 15분 이하 버스 개수** = 1+1+1 = 3
**배차 간격 60분 이상 버스 개수** = 0+0+0 = 0

# 3 DATA

# 파생 변수 생성

## 2. [승차·환승·하차 비율]

### 승차·환승·하차 비율

승차·환승·하차 비율 = (승차·환승·하차 건수) / (승차 건수 + 환승 건수 + 하차 건수)

| 일자 | 정류소ID | 노선ID | 노선번호 | 초승 건수 | 환승 건수 | 하차 건수 | 승차 비율 | 환승 비율 | 하차 비율 |
|---|---|---|---|---|---|---|---|---|---|
| 2020-01-01 | 200000006 | 200000037 | 13 | 160 | 6 | 67 | 160/233 | 6/233 | 67/233 |
| | | 200000099 | 37 | 33 | 3 | 0 | 33/36 | 3/36 | 0 |
| | | 200000045 | 7_1 | 162 | 25 | 190 | 162/377 | 25/377 | 190/377 |

**정류소 승차 비율**
= 승차 비율의 평균
(0.712 + 0.916 + 0.430) / 3
= 0.686

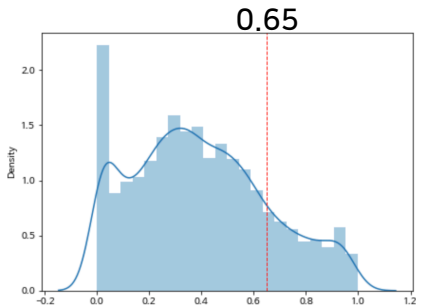
주중/ 주말 / 공휴일
승하차 분포가 다름을 고려하여

주중 / 주말 / 공휴일
세 가지 데이터로 세분화하여
파생변수 생성

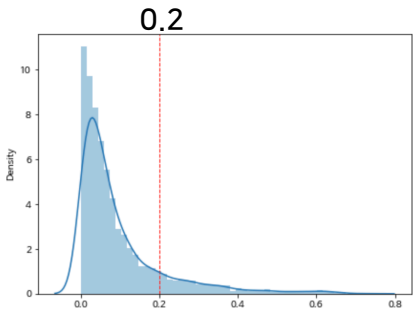
**주중 승차·환승·하차 비율**
**주말 승차·환승·하차 비율**
**공휴일 승차·환승·하차 비율**

### 승차·환승·하차 많은 버스

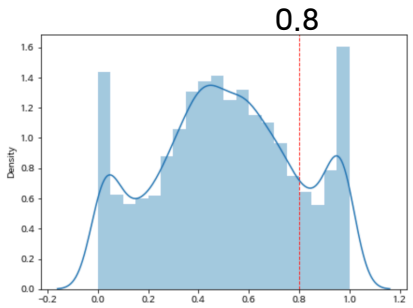
승차·환승·하차 비율 > threshold




주중승차비율 분포




주중환승비율 분포




주중하차비율 분포

**정류소 승차 많은 버스**
= 승차 비율 > 0.65 인 버스의 개수
= 2

주중 승차 비율 > 0.65 인 버스 정류장의 비율 : 0.1735
주중 환승 비율 > 0.20 인 버스 정류장의 비율 : 0.0960
주중 하차 비율 > 0.80 인 버스 정류장의 비율 : 0.1793

버스의 특성에 따라
승차·환승·하차 주요 역할이 달라짐

기준치를 넘으면 특징이 있는 버스라 판단

# 파생 변수 생성

## 3. [대기시간노출도]

- 정류장의 편리성 및 미세먼지 노출도를 반영하기 위해, **대기시간노출도** 변수 생성
- 평균 대기시간 x 정류장 이용객 수와 의미가 유사한, **배차 간격 x 승차 건수**를 통해 값 산출

### 대기시간 노출도

정류장 대기시간 노출도 = 배차 간격 x 승차 건수

| 정류소ID | 노선번호 | 주중배차간격 | 전체 승차 건수 | 평균주중배차간격 | 주중대기시간노출도 |
|---|---|---|---|---|---|
| 200000006 | 13 | 6 | 205.23 | max(6,11,11.5,8) = 11.5 | 6 x 205.23 = 1231.38 |
| | 16 | 210 | 2.21 | | 11.5 x 2.21 = 25.42 |
| | 16_1 | 195 | 2.33 | | 11.5 x 2.33 = 26.8 |
| | 16_2 | 135 | 5.13 | | 11.5 x 5.13 = 59 |
| | 32_5 | 210 | 2.05 | | 11.5 x 2.05 = 23.58 |
| | 35 | 11 | 83.17 | | 11 x 83.17 = 914.87 |
| | 37 | 11.5 | 47.84 | | 11.5 x 47.84 = 550.16 |
| | 7_1 | 8 | 226 | | 8 x 226 = 1808 |

버스 배차 간격의 경우,
평균값을 취할 시 **이상치**에 크게 영향을 받게 됨

따라서, **배차 간격 < 60분**인
버스 노선에서의 **배차 간격 최댓값**을
평균 주중 배차 간격으로 간주하고,

**1. 배차 간격 > 60분 버스는**
**평균 주중 배차 간격 x 승차 건수,**

**2. 배차 간격 < 60분 버스는**
**배차 간격 x 승차 건수**

연산을 통해 주중 대기시간 노출도 값 산출

# 3 파생 변수 생성

DATA

## 4. [강수량상관관계, 온도상관관계, 미세먼지상관관계]

- 정류장 승·하차 건수와 일자 별 기상정보 사이의 상관관계 반영

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 전체승하차건수   R-squared (uncentered):              0.855
Model:                            OLS   Adj. R-squared (uncentered):              0.855
Method:                 Least Squares   F-statistic:                         3.634e+05
Date:                Sat, 27 Mar 2021   Prob (F-statistic):                       0.00
Time:                        16:59:13   Log-Likelihood:                    -4.0955e+05
No. Observations:              185554   AIC:                                 8.191e+05
Df Residuals:                  185551   BIC:                                 8.191e+05
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
최고강수량          0.0414      0.001     43.482      0.000       0.039       0.043
최고체감온도          0.1204      0.000    320.820      0.000       0.120       0.121
미세먼지           0.0716      0.000    381.933      0.000       0.071       0.072
==============================================================================
Omnibus:                     1006.572   Durbin-Watson:                   0.470
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1123.364
Skew:                          -0.139   Prob(JB):                    1.16e-244
Kurtosis:                       3.260   Cond. No.                         8.52
==============================================================================
```

[가정] 날씨에 따라서 정류소 승하차 건수가 달라질 것이다

- Y : log(전체승하차건수) ~ X : 일별 최고강수량 + 최고체감온도 + 가장 가까운 측정소의 미세먼지
- 초미세먼지 변수의 경우, p=0.206, corr(미세먼지, 초미세먼지) = 0.846 이므로 제외
- R-squared =0.855 이므로 가정이 합리적이고, P < 0.05 이므로 모든 변수가 유의미하다고 판단

| | 일자 | 정류소ID | 전체승하차건수 | 최고기온 | 최저기온 | 최고강수량 | 최고체감온도 | 최저체감온도 | 미세먼지 | 초미세먼지 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 2020-01-01 | 200000006 | 646 | 0.200000 | -6.300000 | 0.000000 | 2.332181 | -10.377334 | 34 | 22 |
| 1 | 2020-01-01 | 200000008 | 173 | 0.200000 | -6.300000 | 0.000000 | 2.332181 | -10.377334 | 34 | 22 |
| 2 | 2020-01-01 | 200000036 | 112 | 0.200000 | -6.300000 | 0.000000 | 2.332181 | -10.377334 | 34 | 22 |
| 3 | 2020-01-01 | 200000047 | 187 | 0.200000 | -6.300000 | 0.000000 | 2.332181 | -10.377334 | 34 | 22 |
| 4 | 2020-01-01 | 200000066 | 330 | 0.200000 | -6.300000 | 0.000000 | 2.332181 | -10.377334 | 34 | 22 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 185549 | 2020-12-31 | 203000385 | 94 | -5.000000 | -12.600000 | 0.000000 | -3.146994 | -17.938886 | 28 | 12 |
| 185550 | 2020-12-31 | 203000390 | 57 | -5.000000 | -12.600000 | 0.000000 | -3.146994 | -17.938886 | 27 | 8 |
| 185551 | 2020-12-31 | 203000399 | 1209 | -5.000000 | -12.600000 | 0.000000 | -3.146994 | -17.938886 | 27 | 8 |
| 185552 | 2020-12-31 | 203000439 | 1 | -5.000000 | -12.600000 | 0.000000 | -3.146994 | -17.938886 | 27 | 8 |
| 185553 | 2020-12-31 | 203000450 | 552 | -5.000000 | -12.600000 | 0.000000 | -3.146994 | -17.938886 | 28 | 12 |

정류소ID (516) + 2020 일자별 기상정보 : (185554, 10)

체감온도 : 13.12 + 0.6215T - 11.37V^0.16 + 0.3965V^0.16T (T : 기온(℃), V : 풍속(km/h))

| | 전체승하차건수 | 최고강수량 | 최고체감온도 | 미세먼지 |
|---|---|---|---|---|
| 전체승하차건수 | 1.000000 | -0.194810 | 0.402356 | 0.023169 |
| 최고강수량 | -0.194810 | 1.000000 | 0.177766 | -0.245471 |
| 최고체감온도 | 0.402356 | 0.177766 | 1.000000 | -0.208063 |
| 미세먼지 | 0.023169 | -0.245471 | -0.208063 | 1.000000 |

정류소ID : 20000006
2020.01.01 ~ 2020.12.31 **일자**의
**승하차건수**와 **날씨**와의 상관관계

| | 정류소ID | 강수량상관관계 | 온도상관관계 | 미세먼지상관관계 |
|---|---|---|---|---|
| 0 | 200000006 | -0.194810 | 0.402356 | 0.023169 |
| 1 | 200000008 | -0.206834 | 0.225646 | 0.040863 |
| 2 | 200000036 | -0.013023 | 0.087172 | 0.058419 |
| 3 | 200000047 | -0.034233 | 0.050548 | 0.078566 |
| 4 | 200000066 | -0.006900 | 0.172847 | 0.013248 |
| 5 | 200000068 | -0.023610 | -0.010020 | 0.150519 |
| 6 | 200000070 | 0.073989 | 0.190359 | -0.029807 |
| 7 | 200000071 | 0.024425 | 0.026264 | 0.051416 |
| 8 | 200000073 | -0.041819 | -0.000303 | 0.086151 |
| 9 | 200000075 | -0.044161 | -0.013004 | 0.067826 |

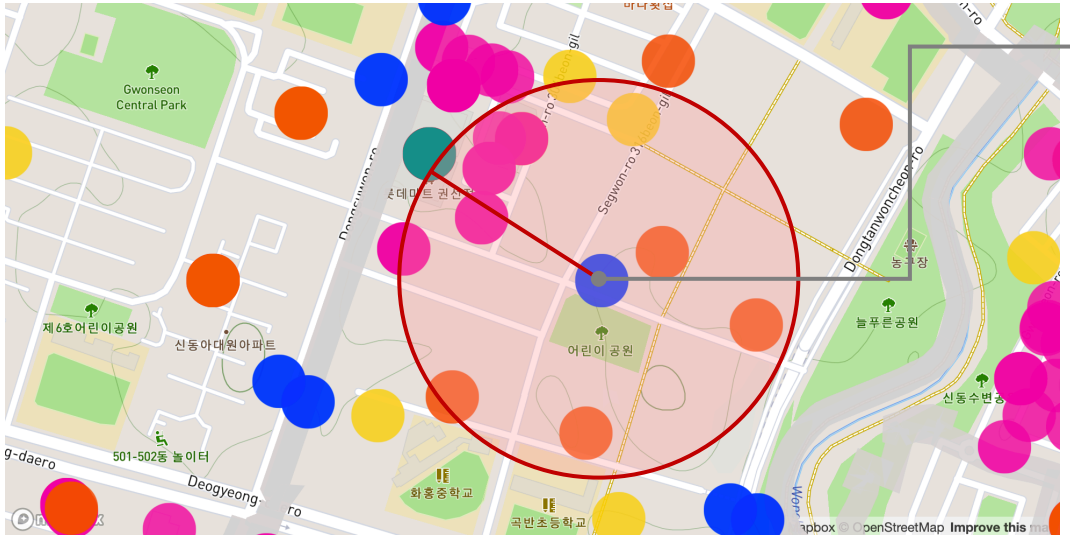
다중공선성을 고려하여,
상관관계 높은 설명변수들 제외

# 3 DATA

# 파생 변수 생성

## 5. [500m 내 상권 개수]

- 공공데이터포털, **소상공인시장진흥공단 경기도 상권정보** 데이터 사용
- 버스 정류장 500m 내의 상권 정보 반영




**버스정류장 500m 반경 내 상권 개수**

생활 편의 시설 : 마트, 편의점
쇼핑 시설 · 번화가 : 백화점, 카페
주거 지역 : 학원, 세탁소

- 버스정류장
- 마트
- 편의점
- 카페
- 세탁소

주변 상권 정보를 통해 정류소 입지 주변 정보 반영

### 수원시 상권 정보

**생활 편의 시설**
- 마트 : 이마트 6 + 홈플러스 5 + 롯데마트 4 + 하나로마트 9 = 24개
- 편의점 : GS25 + 세븐일레븐 + CU + 이마트 24 = 839 개

**쇼핑 시설 · 번화가**
- 백화점 : AKplaza + NC백화점 + 갤러리아백화점 + 롯데몰 = 4개
- 커피전문점/카페 : 2156 개

**주거 지역**
- 학원 : 1657 개
- 세탁소 : 727 개

| 정류소ID | 정류장명 | 500미터내마트개수 | 500미터내백화점개수 | 500미터내편의점개수 | 500미터내카페개수 | 500미터내학원개수 | 500미터세탁소개수 |
|---|---|---|---|---|---|---|---|
| 201000236 | 곡반중학교.수원아이파크 | 2 | 0 | 0 | 8 | 0 | 16 |
| 202000010 | 수원시청역8번출구.씨네파크.한혈의집 | 0 | 1 | 8 | 46 | 0 | 13 |
| 202000020 | 수원시청역9번출구.국민연금공단 | 0 | 1 | 8 | 48 | 8 | 12 |
| 203000160 | 삼성2차아파트 | 1 | 0 | 7 | 6 | 33 | 12 |
| 202000027 | 수원시청역1번출구.국민연금공단.경인일보 | 0 | 1 | 11 | 41 | 0 | 11 |
| 202000073 | 아주대입구.우리은행 | 0 | 0 | 10 | 16 | 29 | 11 |
| 201000001 | 한마음외과 | 0 | 1 | 3 | 15 | 4 | 11 |

# 파생 변수 생성

## 6. [NLP Cluster]

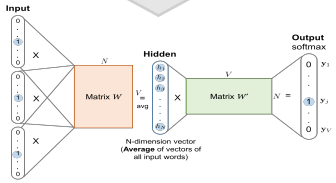
- 버스 정류장명에서 추출할 수 있는 단어 고유의 특성 반영

### 버스 정류장 embedding

수원시 버스정류장 + 수원시 관광명소
= 중복 제거 후 **807개** 데이터 학습

| 정류장명 | 불용어 제거 | Tokenizing |
|---|---|---|
| 곡반중학교.수원아이파크 | 곡반중학교수원아이파크 | [곡반, 중학교, 수원, 아이파크] |
| 고색역.고색초교.태산아파트 | 고색역고색초교태산아파트 | [고색역, 고색, 초교, 태산, 아파트] |
| 영통역8번출구 | 영통역번출구 | [영통역, 번, 출구] |

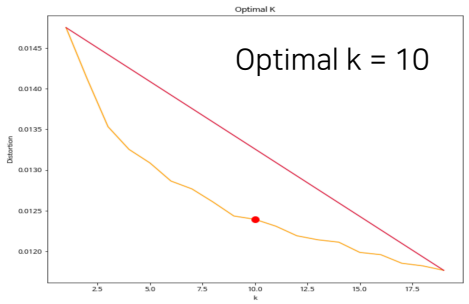
Data Cleansing + Tokenizing (Twitter)




GloVe Embedding

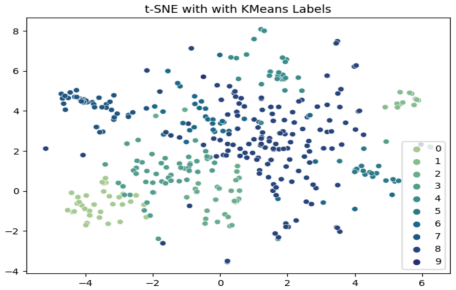
단어 유사도와 동시등장정보를 모두 반영할 수 있는 임베딩 모델
각 token 을 128 차원의 벡터로 embedding 하여,
정류장마다 고유의 embedding vector 생성

### Clustering




K-means Clustering

embedding vector 를 Clustering




Clustering 결과 (t-SNE)

비슷한 특성을 가진 정류장명끼리
잘 뭉쳐있는 양상 확인 가능

**Cluster 0 : 아파트**
신명아파트
풍림아파트
우방아파트.한화아파트.SK아파트
청구1차아파트
...
우만동선경아파트

**Cluster 1 : 사거리**
수성중사거리
정자초교사거리
수원시청사거리
세곡사거리
...
구법원사거리

**Cluster 5 : 역**
수원시청역9번출구.국민연금공단
영통역7번출구.영덕고등학교
매교역6번출구.문화맨션
청명역4번출구
...
망포역7번출구.벽적골8단지

**Cluster 7 : 학교 & 공원**
효원초등학교
잠원중학교
숙지중고교.서호공원
광교호수공원.제1주차장
...
수원월드컵경기장.동성중학교

**10 개의 Cluster 로 분류**

Cluster 마다 고유의 특성 존재

# 3 DATA 내부 데이터 병합

## 1. 미세먼지

### 1) 미세먼지 (11. 대기오염도_측정현황)

- 각 버스 정류장마다 가장 가까운 측정소의 월별 대기 환경 수치를 평균 내어 대기오염 변수 생성
  - 미세먼지, 초미세먼지,오존,이산화질소,아황산가스,일산화탄소

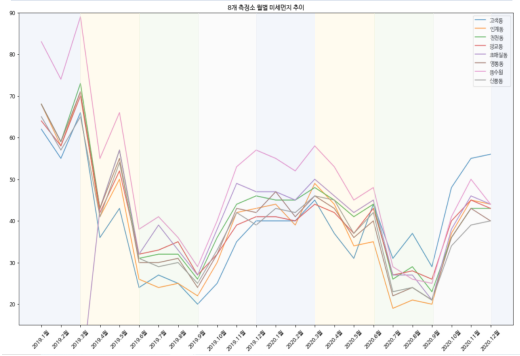
ex) 권선구청 정류장에서
가장 가까운 고색동 측정소의
수치를 이용

미세먼지 측정소
버스 정류장( 권선구청)

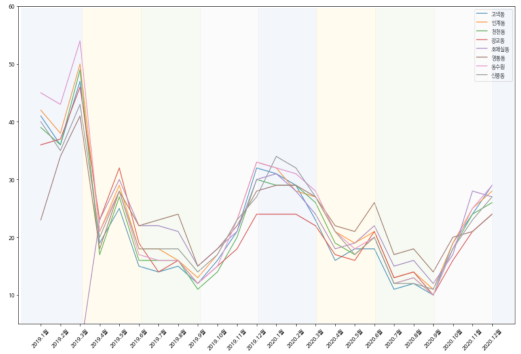
### 2) 미세먼지- 계절적 특성 반영 (11. 대기오염도_측정현황)

- 8개의 측정소에서 측정한 미세먼지와 초미세먼지의 계절별 양상이 다르기 때문에 이를 변수에 반영
  - 미세먼지_봄, 미세먼지_여름, 미세먼지_가을, 미세먼지_겨울,
    초미세먼지_봄, 초미세먼지_여름, 초미세먼지_가을, 초미세먼지_겨울

미세먼지 월별 수치



초미세먼지 월별 수치

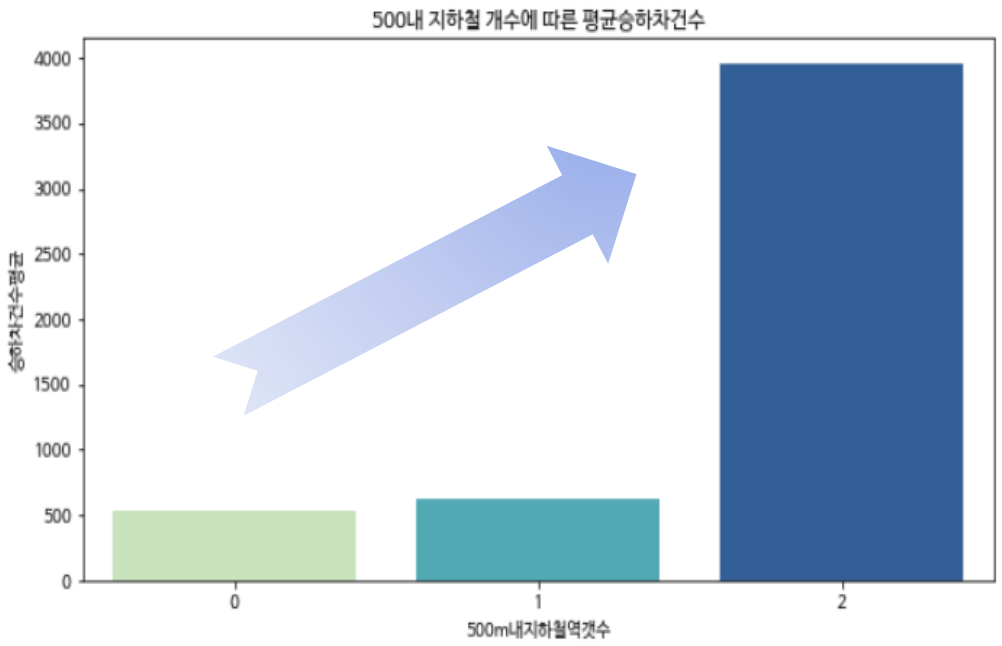


봄 여름
가을 겨울

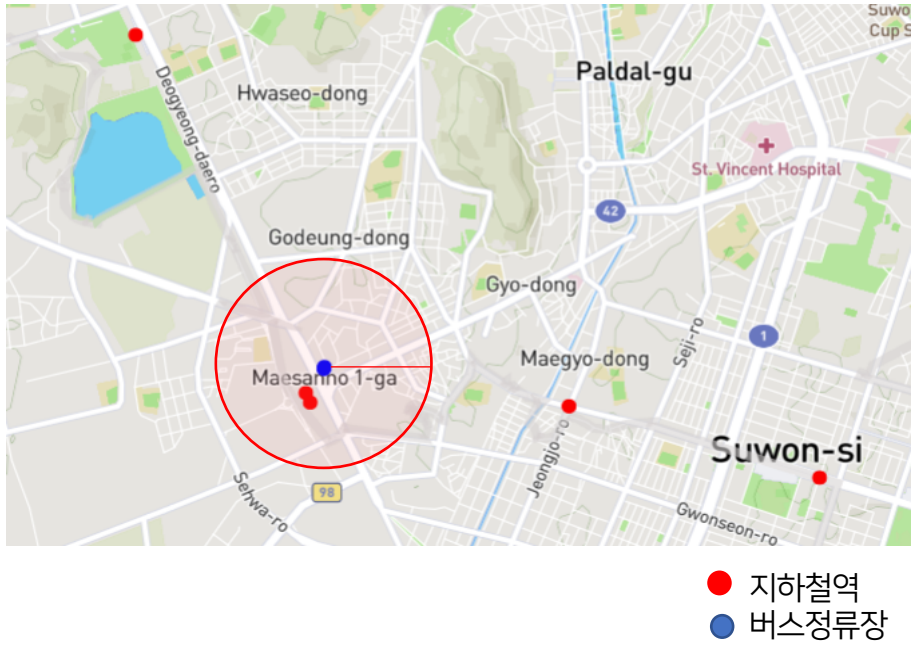
# 3 DATA

# 내부 데이터 병합

## 2. 지하철역 (08.지하철역 위치정보, 09. 지하철역별 이용현황)




500m내 지하철역의 개수가 커짐에 따라 승·하차 평균이 커짐
→ 500m내 지하철역의 개수가 유의미한 변수라 판단




- 정류장 별 반경 500m내 **지하철역**의 개수 변수 생성
  500m내 지하철역개수
- 각 지하철역의 **이용객 수**를 더해 변수 생성
  지하철이용객

# 내부 데이터 병합

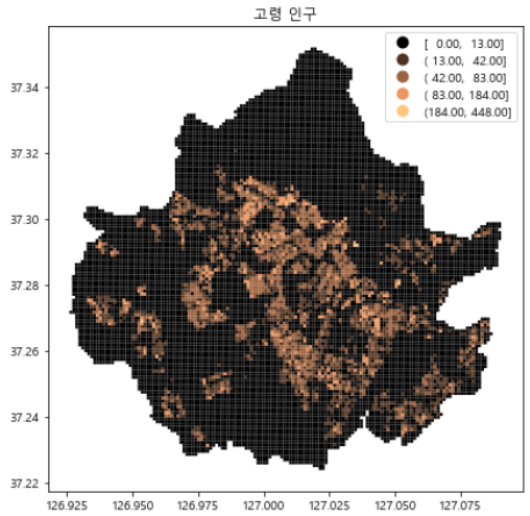
## 3. 고정인구 (17.고령인구, 18.생산가능인구, 19.유소년인구)

- 버스정류장데이터의 경도,위도를 point형태로 변환하여, **고령인구, 유소년인구, 생산가능인구의 격자 안에 들어가면** 해당 인구 변수 생성
- 각 연령층별 인구 비율변수 생성 : 각 인구를 전체 인구로 나눈 변수 생성

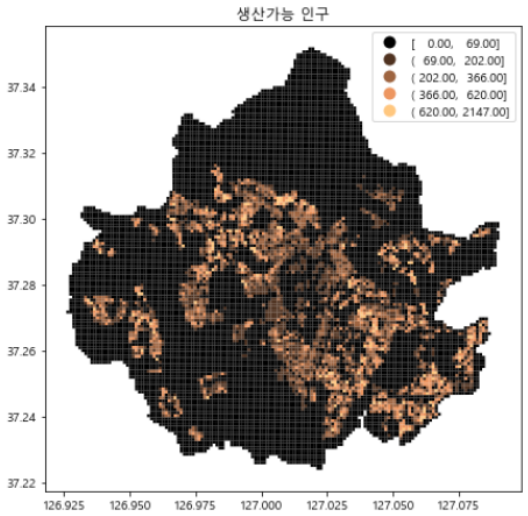
고령인구, 유소년인구, 생산가능인구

고령_전체, 생산_전체, 유소년_전체

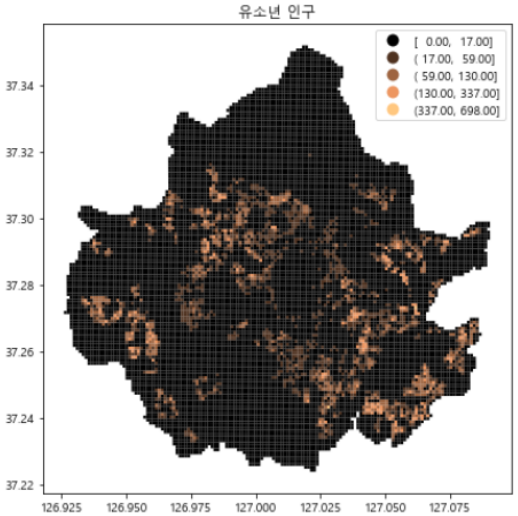
**고령 인구**




**생산가능 인구**




**유소년 인구**




- 고령인구와 생산가능인구는 골고루 인구분포가 많음
- 유소년인구는 중심보다는 주변에 인구가 많이 분포

인구 적음 인구 많음

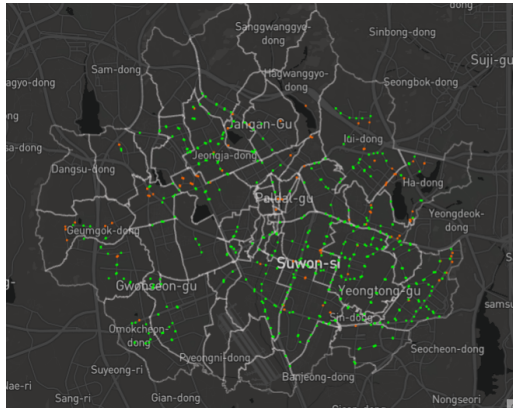
# 3 DATA 내부 데이터 병합

## 4. 유동인구

### 1) 평일, 주말 유동인구 (16. 요일별_유동인구)

- 각 정류장과 가장 가까운 유동인구 격자의 중심점을 기준으로 요일별 유동인구를 평일과 주말로 나누어 변수 생성

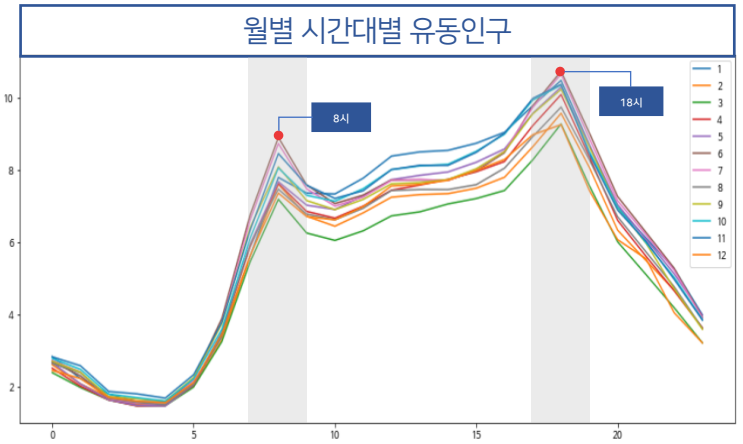
평일유동인구, 주말유동인구




● 주말에 유동인구가 많은 정류장
● 평일에 유동인구가 많은 정류장

- 대부분의 버스 정류장에서 주말 유동인구보다 평일 유동인구가 많음
- 공원이나 박물관,관광지 주변의 일부 버스정류장에서는 주말 유동인구가 많음

### 2) 출퇴근 유동인구 (14. 수원시_시간대별_유동인구)




- 일별, 시간대별 데이터의 평균을 이용하여 월별, 시간대별 데이터로 변환
- 월별 경향성은 비슷하고 시간대별로는 다른 양상을 보임
- 그래프 peak를 기준으로 **앞, 뒤 한 시간씩을 포함하여** 출퇴근시간으로 판단하여 변수 생성

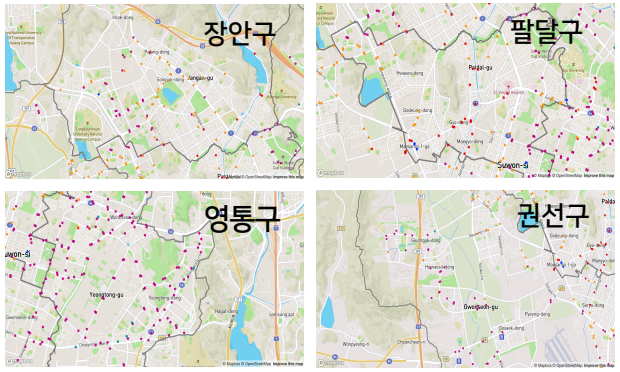
출근시간: 7~9시,
퇴근시간: 17~19시

유동인구_출근, 유동인구_퇴근,
유동인구_출퇴근, 유동인구_전체

### 3) 성연령별 유동인구 (15. 수원시_성연령별_유동인구)

- 각 정류장과 가장 가까운 유동인구 격자의 중심점을 기준으로 성·연령별 유동인구로 나누어 변수 생성

10대남자유동인구, 20대남자유동인구, … , 60대남자유동인구
10대여자유동인구, 20대여자유동인구, … , 60대여자유동인구

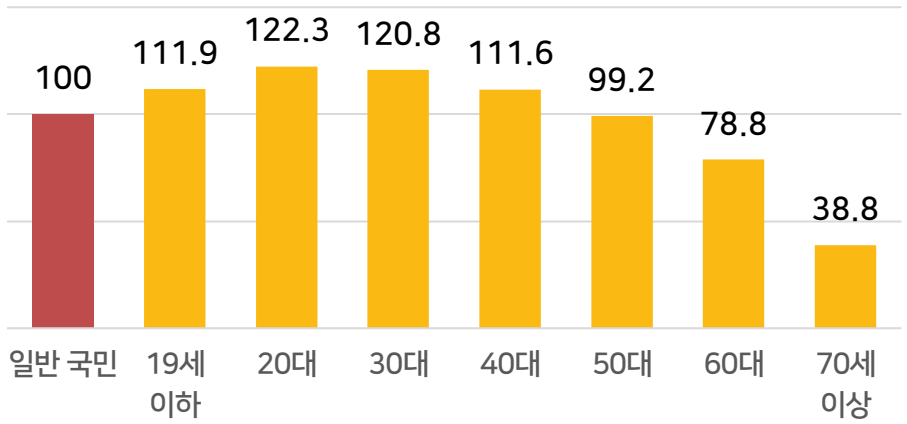



● 60대 남자 유동인구 비율이 많은 버스정류장
● 50대 남자 유동인구 비율이 많은 버스정류장
● 40대 남자 유동인구 비율이 많은 버스정류장
● 30대 남자 유동인구 비율이 많은 버스정류장
● 20대 남자 유동인구 비율이 많은 버스정류장

# 3 DATA 파생 변수 생성

## 5. [디지털정보지수] (15. 수원시_성연령별_유동인구)

- 사회적 민감군 (취약계층) 에 대한, **정보화 격차**를 고려하기 위한 변수
- 정류장에 가장 가까운 유동인구 격자를 기준으로, 디지털정보지수 값 산출




**연령별 디지털정보화 역량수준**

일반국민의 디지털정보화 수준 (컴퓨터·모바일 기기 기본 이용 능력) 을 100으로 할 때, 일반국민 대비 연령별 디지털정보화 역량수준

출처 : 과학기술정보통신부, 2020 디지털정보격차 실태조사 보고서

| 정류소ID | 남자10대유동인구 | 남자20대유동인구 | 남자30대유동인구 | 남자40대유동인구 | 남자50대유동인구 | 남자60대유동인구 | 여자10대유동인구 | 여자20대유동인구 | 여자30대유동인구 | 여자40대유동인구 | 여자50대유동인구 | 여자60대유동인구 | 디지털정보지수 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 200000006 | 2.365833 | 7.491667 | 9.765833 | 13.972500 | 20.946667 | 25.994167 | 1.820833 | 6.397500 | 6.496667 | 10.760000 | 17.704167 | 16.250833 | 59.483040 |
| 200000008 | 1.715833 | 5.220833 | 9.853333 | 15.887500 | 22.269167 | 30.544167 | 1.487500 | 4.343333 | 5.940833 | 10.950000 | 20.195000 | 19.827500 | 63.382027 |
| 200000036 | 2.625000 | 4.275000 | 6.124167 | 8.574167 | 8.933333 | 7.345833 | 2.758333 | 3.272500 | 3.898333 | 7.241667 | 6.602500 | 4.957500 | 32.321112 |
| 200000047 | 4.225833 | 4.836667 | 6.171667 | 10.119167 | 11.035833 | 9.412500 | 4.180000 | 4.864167 | 5.212500 | 8.629167 | 8.676667 | 6.777500 | 41.475248 |
| 200000066 | 2.580833 | 8.575000 | 13.912500 | 17.730000 | 20.731667 | 19.568333 | 2.568333 | 7.763333 | 8.851667 | 12.262500 | 13.874167 | 12.340000 | 62.581964 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 203000316 | 0.366667 | 1.070833 | 2.985000 | 4.309167 | 3.944167 | 2.564167 | 0.349167 | 0.776667 | 1.482500 | 1.880000 | 1.520000 | 1.101667 | 10.723160 |
| 203000317 | 0.760000 | 2.993333 | 6.454167 | 8.130833 | 7.505833 | 5.322500 | 0.774167 | 2.575833 | 3.470000 | 3.752500 | 3.745833 | 2.519167 | 22.190258 |
| 203000344 | 6.990833 | 16.029167 | 29.757500 | 33.933333 | 27.127500 | 21.744167 | 8.079167 | 18.980833 | 23.552500 | 27.526667 | 22.737500 | 14.344167 | 118.942575 |
| 203000357 | 15.593333 | 35.610833 | 77.781667 | 111.862500 | 101.076667 | 69.138333 | 17.014167 | 27.270833 | 48.920000 | 63.266667 | 52.160000 | 34.044167 | 317.226568 |
| 203000439 | 1.758333 | 4.480833 | 9.762500 | 12.303333 | 10.271667 | 6.612500 | 2.107500 | 4.222500 | 6.437500 | 7.803333 | 5.705000 | 3.732500 | 36.341893 |

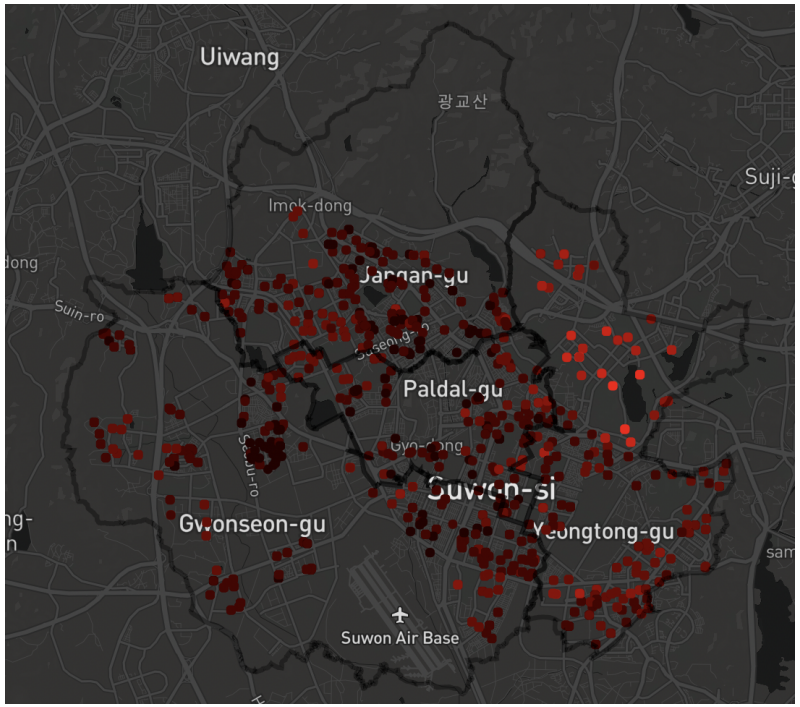
**디지털정보지수**

= (10대 유동인구 x 1.119) + (20대 유동인구 x 1.223) + (30대 유동인구 x 1.208) + (40대 유동인구 x 1.116) + (50대 유동인구 x 0.992) + 60대 유동인구 x {(0.778 + 0.388)/2}

# 외부 데이터 병합

## 6. 실거래가 (외부데이터 : 수원시 실거래가)

- 국가교통부 실거래가 공개시스템에서 2020년 1월부터 2020년 12월까지의 **아파트 실거래가를 각 구별로 병합**
- Geocoder-Xr을 이용하여 **시군구+도로명 기준으로 좌표 생성**




면적당 금액 낮음 ● ● ● 면적당 금액 높음

면적당 금액= 거래금액/전용면적

- 전용면적, 층, 거래금액, 면적당 금액 변수 생성
- 색이 밝을수록 주변 아파트의 실거래가 높은 버스 정류장
- 다른 구에 비해 영통구의 면적당 금액이 높은 편임을 알 수 있음
- 아파트 면적당 금액 변수는 **소득 수준을 반영하는 수치**로서 취약계층을 고려하여 분석하는데 필요한 변수라고 판단

# 4

# MODEL

# MODEL FLOW

## Data Preprocessing

데이터 Filter

- 인도 폭
- BIS 설치 유무

데이터 전처리

- 내부 데이터 병합
- 외부 데이터 병합
- 파생변수 생성

버스정류장 데이터

## Modeling

SPCA

K-means Clustering

Target 클러스터

PLS Regression

- 추정치를 이용한 스마트지수 생성

## 최종 설치 입지 선정

### 설치위치 선정

스마트 지수

구별 순위

Constraint

- 구별 정류장 비율
- 정류장 간 거리

### 최종 결과

통합 순위

Constraint

광고 Targeting

- 성별 가중치 부여

# SparsePCA & Clustering

## 차원 축소 기법

### *Sparse PCA* : *Sparse Principal Component Analysis*

다변량 데이터에서 많이 쓰는 차원축소 방법
고차원 데이터의 최대 분산 방향을 찾아 새로운 공간에서 저차원으로 투영하는 PCA 에, **추가적인 규제** (Ridge, Lasso)를 주는 방식
모든 변수의 linear 조합으로 만들어지는 PCA와 달리, **몇 개의 변수 조합만을 이용하기 때문에** 해석에 용이하다는 장점이 있음

## 클러스터링 기법

### *K-means Clustering*

데이터를 k개의 cluster로 묶는 클러스터링 알고리즘
동일한 군집에 속하는 데이터는 어떠한 **중심**을 기준으로 분포할 것이라는 가정에 기반하여,
각 클러스터 별 중심점과 클러스터 내의 점들의 거리 차이 **분산을 최소화**하는 방식으로 작동

# SparsePCA & Clustering

## 차원축소 & 클러스터링 순서도

### 구별로 데이터 나누기

1) 수원시의 구별 특징 반영
2) 스마트버스쉘터의 공공성

**구별로 데이터 분할**

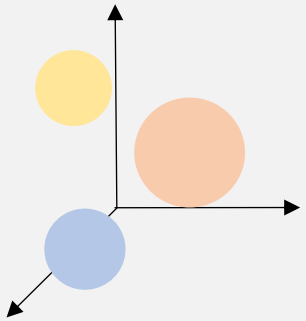
### Sparse PCA 차원 축소

145개의 변수로 이루어진
데이터의 해석을 용이하게
하기 위해 Sparse PCA 적용

| 1 | 2 | … | 145 |
|---|---|---|---|
| | | | |
| | | | |

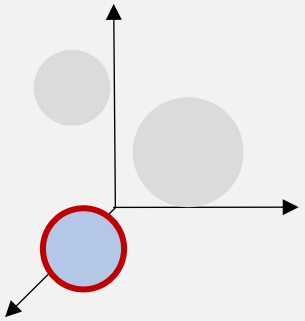
| 1 | 2 | 3 |
|---|---|---|
| | | |
| | | |

### K-means clustering



Sparse PCA 변수를 축으로
**구별 클러스터링**을 통해
비슷한 특징 갖는 버스정류장 파악

### Target 그룹 선택

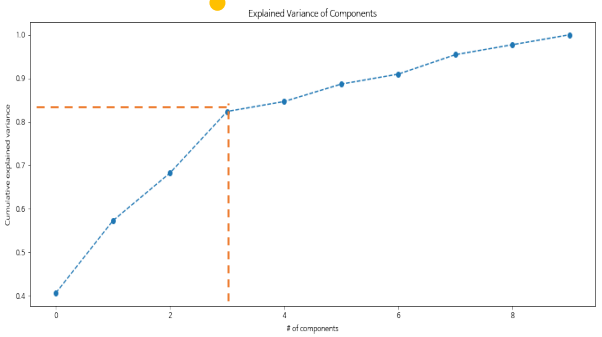


각 그룹의 Sparse PCA로 해석된
변수들의 **평균을 비교**해
target 그룹 선택

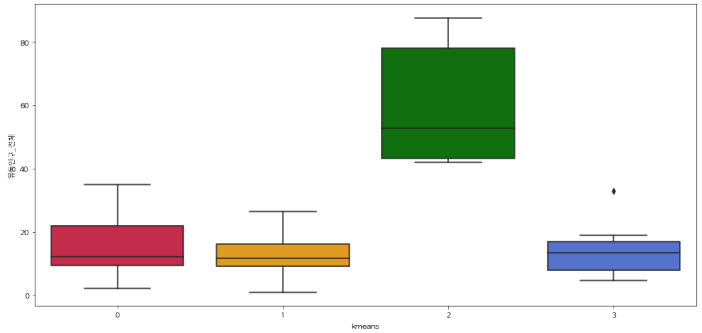
# SparsePCA & Clustering

## 차원축소 & 클러스터링 상세 과정

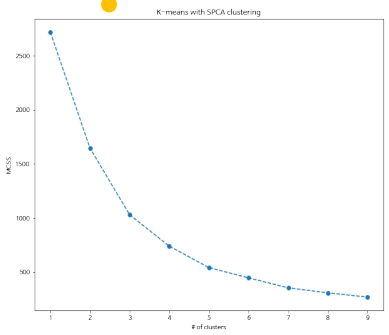
1. Sparse PCA 분석 과정에 필요 없는 변수 제거
2. **Explained variance plot**을 이용해 적절한 component 수 설정
3. Sparse PCA로 만들어진 축을 기준으로 **elbow method**를 이용해 K-means clustering
4. Cluster내의 변수들의 평균 비교 (Anova test & **box plot**)




팔달구 Explained Variance Plot




팔달구 cluster별 Box Plot




팔달구 K-means Elbow Method Plot

# 구 별 SPCA & Clustering 결과

## 1. 팔달구 Target

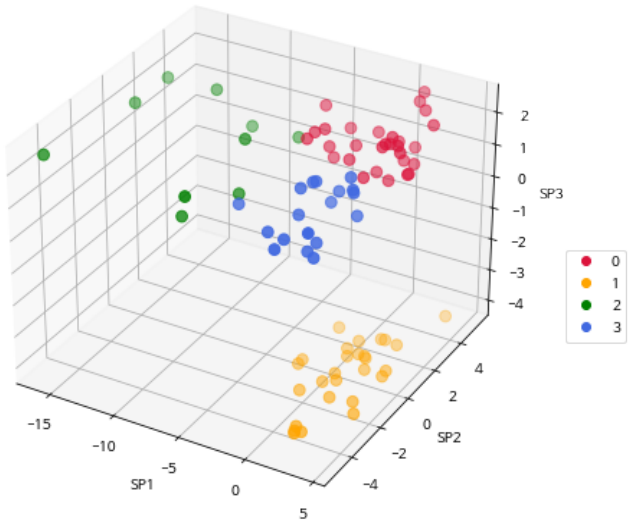
### 차원축소 & 클러스터링 결과

### 팔달구

**SPCA별 유의미한 변수**

| SPCA | 변수명 |
|---|---|
| SP1 | 유동인구_전체 |
| SP2 | 주말승차비율 |
| SP3 | 이산화질소(ppm) |

K-means
Clustering



# 구 별 SPCA & Clustering 결과

## Group별 SPCA 변수 평균 비교 결과 (37개)

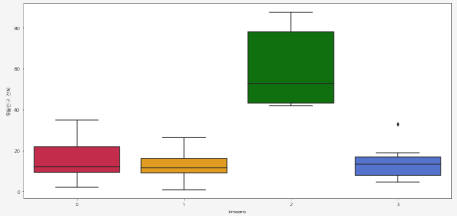
| group # | 유동인구_전체 | pop_주말 | 디지털 정보지수 | 미세_겨울 | 이산화질소 (ppm) |
|---|---|---|---|---|---|
| group0 | 15.382 | 337.006 | 165.702 | 49.103 | 0.0258 |
| group1 | 12.410 | 258.108 | 135.975 | 60.833 | 0.034 |
| group2 | 59.422 | 1352.14 | 597.524 | 48.333 | 0.0259 |
| group3 | 13.178 | 278.643 | 146.537 | 49.241 | 0.0259 |

**group1**
디지털정보지수 낮음, 미세먼지 높음

**group2**
유동인구 높음

## ANOVA test & Box plot

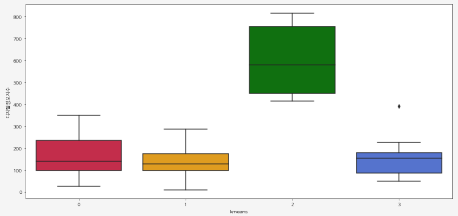
### 유동인구_전체



| | df | sum_sq | mean_sq | F | PR(>F) |
|---|---|---|---|---|---|
| C(kmeans) | 3 | 18498.094 | 6166.031 | 64.142 | 0.00 |
| Residual | 80 | 7690.49 | 96.131 | nan | nan |

→ **유동인구_전체** 변수가 그룹별로 유의미한 차이가 있음

### 디지털정보지수



| | df | sum_sq | mean_sq | F | PR(>F) |
|---|---|---|---|---|---|
| C(kmeans) | 3 | 1775499.3 | 591833.1 | 62.648 | 0.00 |
| Residual | 80 | 755751.58 | 9446.895 | nan | nan |

→ **디지털정보지수** 변수가 그룹별로 유의미한 차이가 있음

# 구 별 SPCA & Clustering 결과

## 2. 장안구 Target

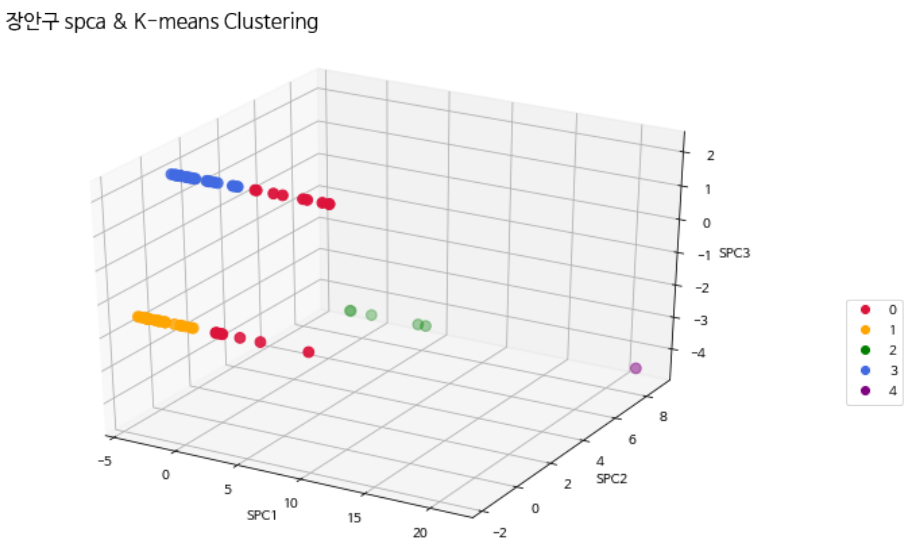
### 차원축소 & 클러스터링 결과

### 장안구

**SPCA별 유의미한 변수**

| SPCA | 변수명 |
|---|---|
| SP1 | 유동인구_전체 |
| SP2 | 미세_봄 |
| SP3 | 초미세먼지(μg/㎥) |

K-means Clustering



# 구 별 SPCA & Clustering 결과

**Group별 SPCA 변수 평균 비교 결과 (29개)**

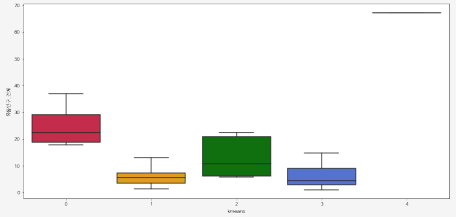
| group # | 유동인구_전체 | pop_평일 | 디지털 정보지수 | 미세_봄 | 초미세먼지 (μg/m³) |
|---|---|---|---|---|---|
| group0 | 24.356 | 604.087 | 278.325 | 49.583 | 22.688 |
| group1 | 5.909 | 149.692 | 66.094 | 48 | 23.125 |
| group2 | 13.200 | 316.165 | 141.769 | 61 | 24.167 |
| group3 | 6.084 | 153.571 | 70.771 | 51.167 | 22.25 |
| group4 | 67.300 | 1662.53 | 712.880 | 61 | 24.167 |

**group1**
디지털정보지수 낮음

**group4**
유동인구 높음, 미세먼지 높음

**ANOVA test & Box plot**

유동인구_전체



| | df | sum_sq | mean_sq | F | PR(>F) |
|---|---|---|---|---|---|
| C(kmeans) | 4 | 7825.150 | 1956.287 | 83.464 | 0.00 |
| Residual | 76 | 1781.347 | 23.439 | nan | nan |

→ **유동인구_전체** 변수가 그룹별로 유의미한 차이가 있음

# 구 별 SPCA & Clustering 결과

## 3. 권선구 Target

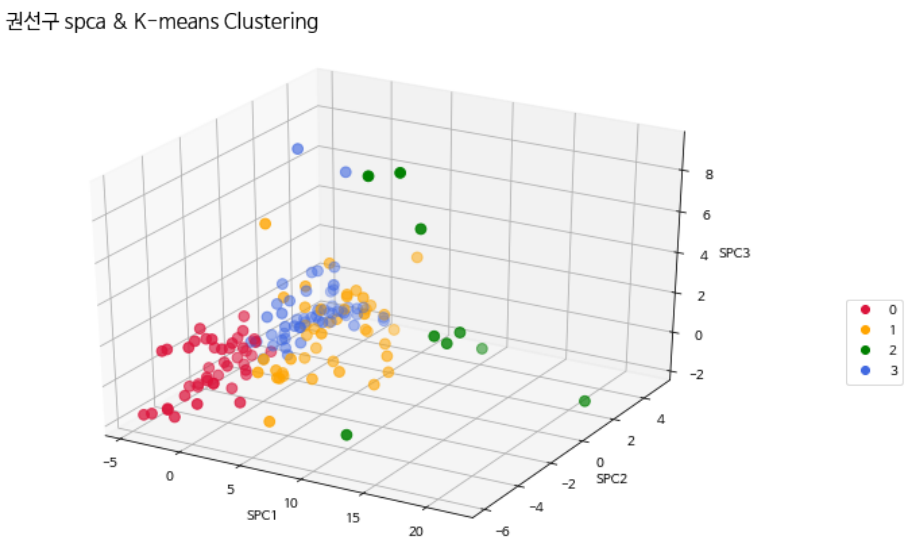
### 차원축소 & 클러스터링 결과

### 권선구

**SPCA별 유의미한 변수**

| SPCA | 변수명 |
|---|---|
| SP1 | 유동인구_전체 |
| SP2 | 하차비율 |
| SP3 | 주말배차간격_60분 |

K-means Clustering



# 구 별 SPCA & Clustering 결과

## Group별 SPCA 변수 평균 비교 결과 (51개)

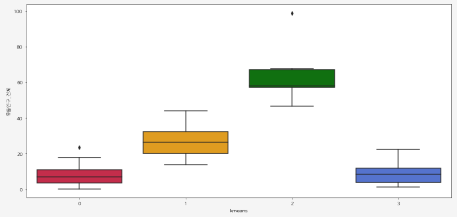
| group # | 유동인구_전체 | pop_평일 | 하차비율 | 주말배차간격_60분 | 운행노선수 |
|---|---|---|---|---|---|
| group0 | 7.369 | 186.467 | 0.271 | 1.250 | 6.955 |
| group1 | 27.186 | 673.394 | 0.525 | 2.409 | 9.886 |
| group2 | 63.663 | 1563.56 | 0.496 | 7.111 | 15.778 |
| group3 | 8.715 | 219.524 | 0.757 | 1.340 | 6.660 |

**group1**
유동인구 높음, 하차비율 높음

**group2**
유동인구 높음, 배차간격_60분 높음, 운행노선수 높음

## ANOVA test & Box plot

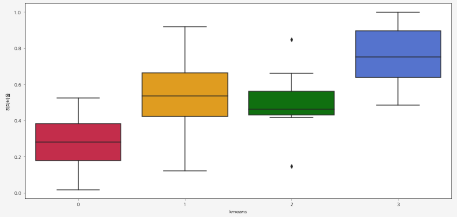
### 유동인구_전체



| | df | sum_sq | mean_sq | F | PR(>F) |
|---|---|---|---|---|---|
| C(kmeans) | 3 | 31881.34 | 10627.11 | 216.789 | 0.00 |
| Residual | 146 | 7156.995 | 49.021 | nan | nan |

→ **유동인구_전체** 변수가 그룹별로 유의미한 차이가 있음

### 하차비율



| | df | sum_sq | mean_sq | F | PR(>F) |
|---|---|---|---|---|---|
| C(kmeans) | 3 | 5.692 | 1.897 | 74.909 | 0.00 |
| Residual | 146 | 3.698 | 0.025 | nan | nan |

→ **하차비율** 변수가 그룹별로 유의미한 차이가 있음

# 구 별 SPCA & Clustering 결과

## 4. 영통구 Target

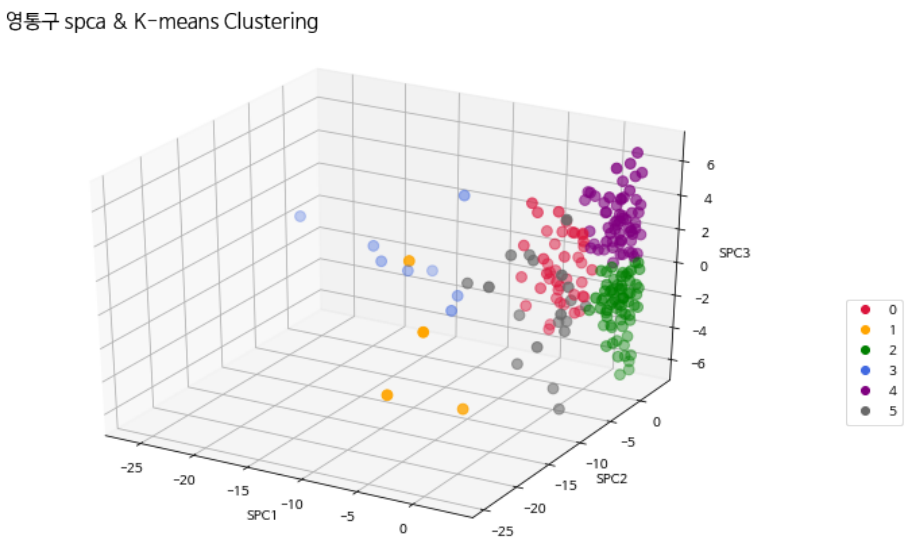
### 차원축소 & 클러스터링 결과

### 영통구

**SPCA별 유의미한 변수**

| SPCA | 변수명 |
|---|---|
| SP1 | 주중전체승하차건수 |
| SP2 | 유동인구_전체 |
| SP3 | 주말승차비율 |

K-means Clustering



# 구 별 SPCA & Clustering 결과

## Group별 SPCA 변수 평균 비교 결과 (31개)

| group # | 주중전체 승하차건수 | 운행노선수 | 유동인구 _전체 | pop_평일 |
|---|---|---|---|---|
| group0 | 1081.439 | 15.095 | 17.268 | 429.499 |
| group1 | 1032.083 | 18.750 | 133.496 | 3193.767 |
| group2 | 193.458 | 3.985 | 8.443 | 217.020 |
| group3 | 4102.100 | 26.250 | 27.512 | 791.537 |
| group4 | 249.322 | 4.984 | 8.063 | 203.663 |
| group5 | 582.233 | 9.632 | 60.564 | 1518.980 |

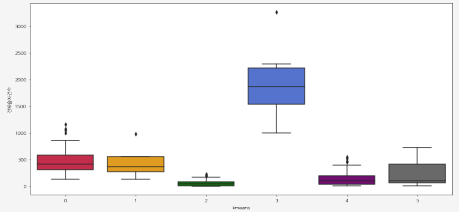
**group1**
유동인구 높음

**group5**
유동인구 높음

**group3**
전체승하차건수 높음,
운행노선수 높음

## ANOVA test & Box plot

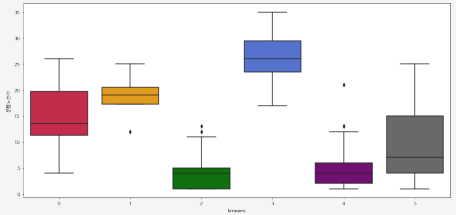
### 전체승차건수



| | df | F | PR(>F) |
|---|---|---|---|
| C(kmeans) | 5 | 124.626 | 0.00 |
| Residual | 195 | nan | nan |

→ **전체승차건수** 변수가 그룹별로
유의미한 차이가 있음

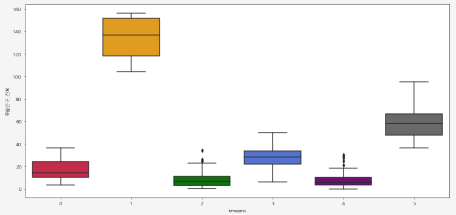
### 운행노선수



| | df | F | PR(>F) |
|---|---|---|---|
| C(kmeans) | 5 | 63.188 | 0.00 |
| Residual | 195 | nan | nan |

→ **운행노선수** 변수가 그룹별로
유의미한 차이가 있음

### 유동인구_전체



| | df | F | PR(>F) |
|---|---|---|---|
| C(kmeans) | 5 | 207.067 | 0.00 |
| Residual | 195 | nan | nan |

→ **유동인구_전체** 변수가 그룹별로
유의미한 차이가 있음

# Multivariate Regression & PLSR

## 회귀 모델

### *Multivariate Regression*: 다변량 회귀분석

$$\begin{pmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ \vdots & \vdots & \vdots \\ y_{n1} & y_{n2} & y_{n3} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1\,79} \\ 1 & x_{21} & \cdots & x_{2\,79} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{n\,79} \end{pmatrix} \begin{pmatrix} \beta_{01} & \beta_{02} & \beta_{03} \\ \beta_{11} & \beta_{12} & \beta_{13} \\ \vdots & \vdots & \vdots \\ \beta_{79\,1} & \beta_{79\,2} & \beta_{79\,3} \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} & \varepsilon_{13} \\ \varepsilon_{21} & \varepsilon_{22} & \varepsilon_{23} \\ \vdots & \vdots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \varepsilon_{n3} \end{pmatrix}$$

$n$: 구별 데이터 개수

$$cov(\varepsilon_{ij}, \varepsilon_{kl}) = 0, \quad \mathrm{i} \neq k$$
$$cov(\varepsilon_{ij}, \varepsilon_{il}) = \Sigma_{jl}$$

- X 사이의 상관성 뿐만 아니라 Y 사이의 상관성도 함께 고려할 수 있음
- 다만, X변수 간의 **다중공선성 위험이 있어** PLSR에서 multivariate regression 사용
- 주요 3가지 변수 **Y1: 교통량 ,Y2: 유동인구 , Y3: 전체 승차건수**를 모두 종속변수로 설정해 다변량 회귀분석 진행

# Multivariate Regression & PLSR

## 회귀 모델

### *PLSR: Partial Least Square Regression*

- **Y와의 공분산이 높은 k개의 선형조합 변수를 추출**하여 적합시킨 회귀모형
- X 선형결합의 분산을 최대화할 뿐만 아니라 **Y의 상관관계도 최대화하는 축을 찾아 차원축소**
- 추출된 변수가 **Y와의 상관관계를 반영**

### *Why PLSR?*

- **설명변수** (79개)에 비해 **관측치** (구 별-37개,29개,53개,31개)가 **많지 않기 때문에** 사용
- **다변량** (상관관계를 갖는 다수의 Y)에 대한 회귀분석 가능
- 종속 변수와의 상관성을 고려한 잠재변수 도출을 통해 데이터 차원을 축소할 수 있음
- X변수들 간 **다중공선성의 위험성 감소**
- 추출된 소수의 변수로 효율적인 모델 구축 가능

# Multivariate Regression & PLSR

## PLSR Flow

- PLSR은 각 **구별**로 선정된 클러스터로 진행 (모든 구의 전체적인 흐름 동일)
- **Y1: 교통량 ,Y2: 유동인구 , Y3: 전체 승차건수**를 **동시에 고려**하는 다변량 분석기법 사용
- 예측한 Y1, Y2, Y3를 더해 **스마트 지수 생성** → 지수가 클수록 스마트버스정류장으로 교체 우선순위에 해당

**X변수 삭제**
Y와의 상관성이 높은 X변수는 삭제하여 136개에서 79개로 축소

→

**Y의 값 분포 확인**
기본값이 skewed 되어 있다면, log scaling을 진행하여 정규분포 형태로 변환

→

**최적 차원 개수 찾기**
MSE 최소, RPD 최대, $R^2$ 최대가 되는 **최적의 차원 개수** 찾기

→

**PLSR**
최적의 차원 개수로 **회귀분석 진행**

→

**Y 예측값**
PLSR을 통해 얻은 예측값 확인

→

**exp(Y 예측값)**
회귀 전 log scaling 했다면, 다시 exp 취함

→

**스케일링**
$\widehat{y_1}$, $\widehat{y_2}$ , $\widehat{y_3}$ 의 범위가 다르기 때문에 minmaxscaling 진행

→

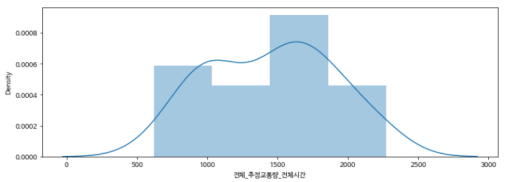
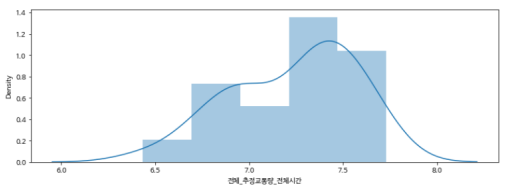
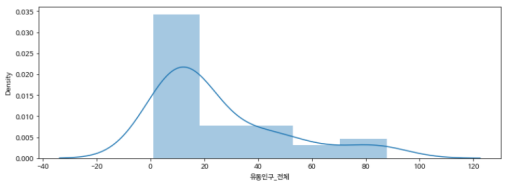
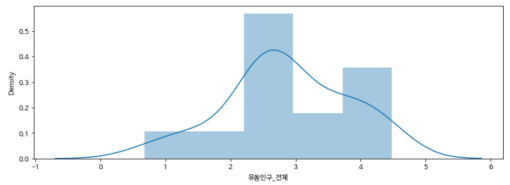
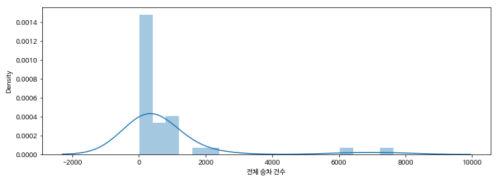
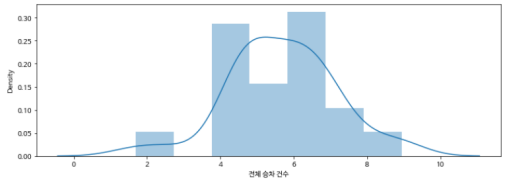
**스마트지수 생성**
정류장별로 scaled $\widehat{y_1}$,scaled $\widehat{y_2}$ , scaled $\widehat{y_3}$ 를 모두 더하여 **스마트지수 생성**

# 구 별 PLSR 결과

## PLSR 결과

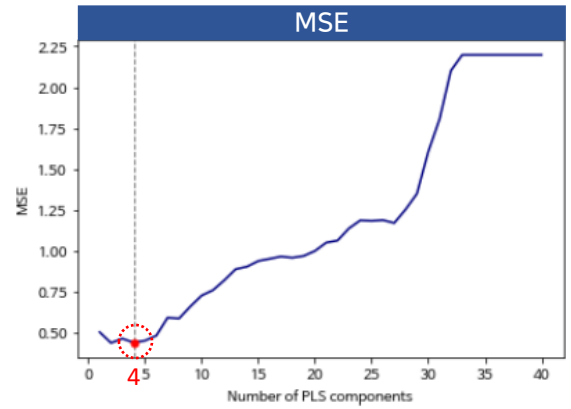
### 팔달구

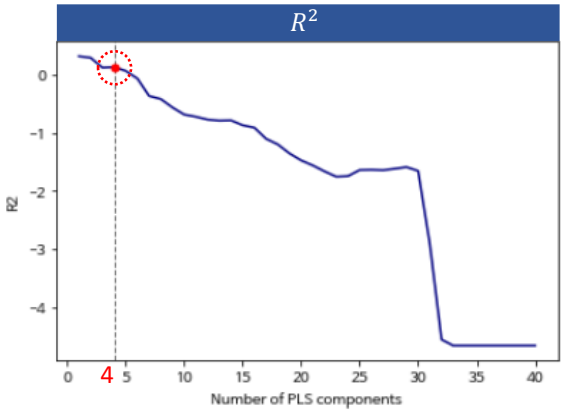
- 팔달구에서 선택된 2개의 클러스터로 진행 (10+27=총 37개 버스 정류장)
- Y의 값이 다소 skewed된 형태이기 때문에 log scaling 후 진행

| | scaling 전 | scaling 후 |
|---|---|---|
| 교통량 |  |  |
| 유동인구 |  |  |
| 승차건수 |  |  |

# 구 별 PLSR 결과

## 최적 차원 개수 찾기

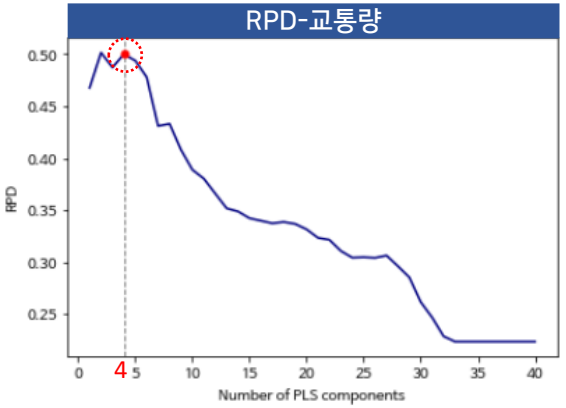


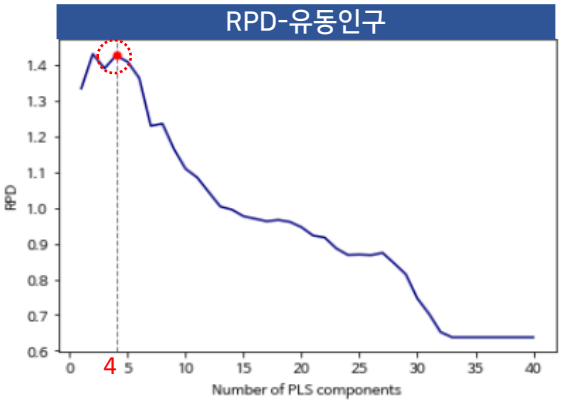



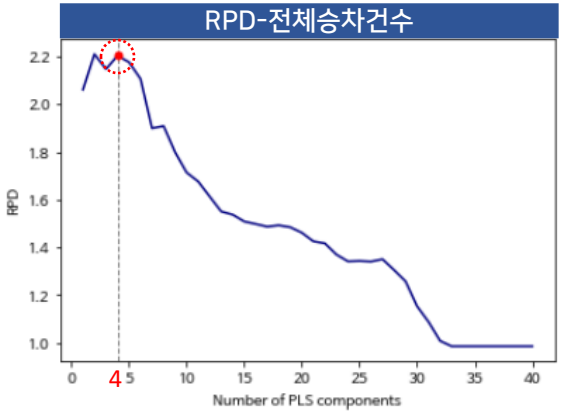

$$\text{MSE}=\frac{1}{df_E}\sum(Y_i-\hat{Y}_i)^2$$

$$R^2=\frac{\sum(\hat{Y}_i-\bar{Y}_i))^2}{\sum(Y_i-\bar{Y}_i)^2}$$

$$\text{RPD}=\frac{\sqrt{\frac{\sum(Y_i-Y_i)^2}{n-1}}}{\sqrt{MSE}}$$



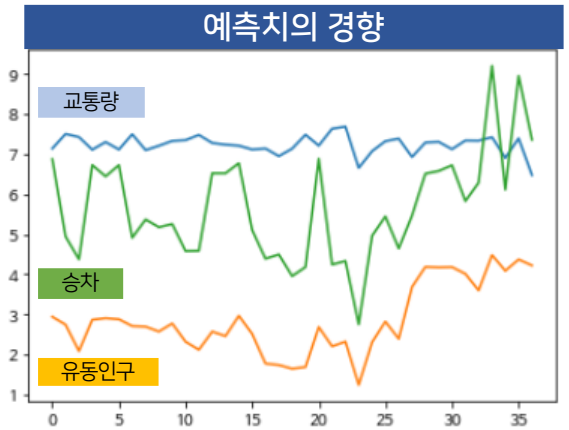



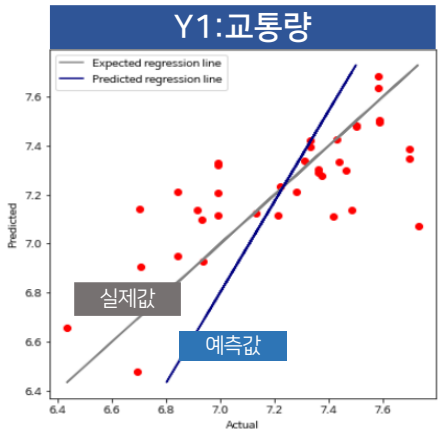



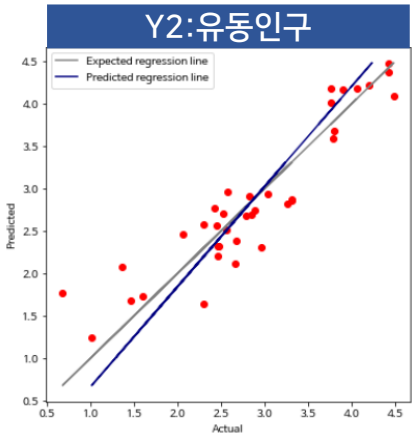

→ MSE 최소, RPD최대, $R^2$최대가 되는 차원개수 : 4

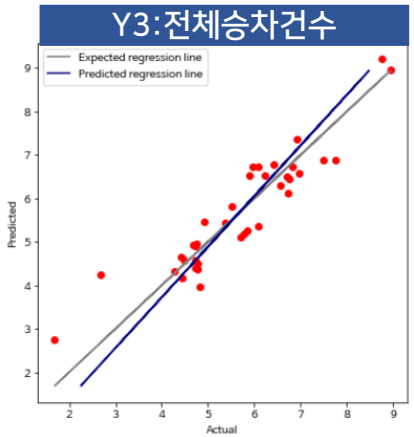
# 구 별 PLSR 결과

최적 차원 개수 4개로 PLSR 진행 ($R^2$: 0.7483 , MSE:187865.3472)




- 전체승차수와 유동인구 예측치가 매우 비슷한 양상
- 팔달구에서 전체 승차가 많은 버스 정류장은 유동인구가 많은 곳이라 할 수 있음

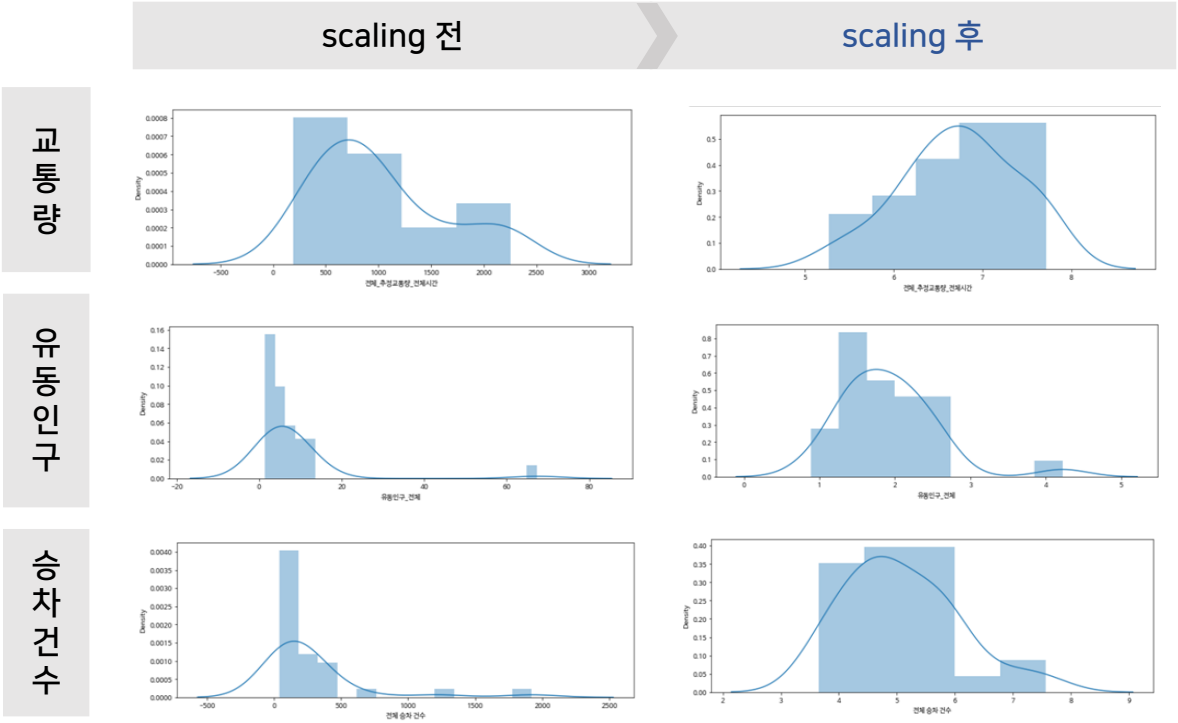
- 유동인구와 전체승차건수가 모형에 매우 잘 적합했음을 볼 수 있음
- 전체 $R^2$ = 0.75로 모형이 전체적으로 적합하다고 판단
- 위의 회귀식을 이용하여 $\widehat{y_1}$, $\widehat{y_2}$ , $\widehat{y_3}$ 을 각각 구함
- 최종 스마트 지수 생성 : smart= scaled $\widehat{y_1}$+scaled $\widehat{y_2}$ + scaled $\widehat{y_3}$

# 구 별 PLSR 결과

## PLSR 결과

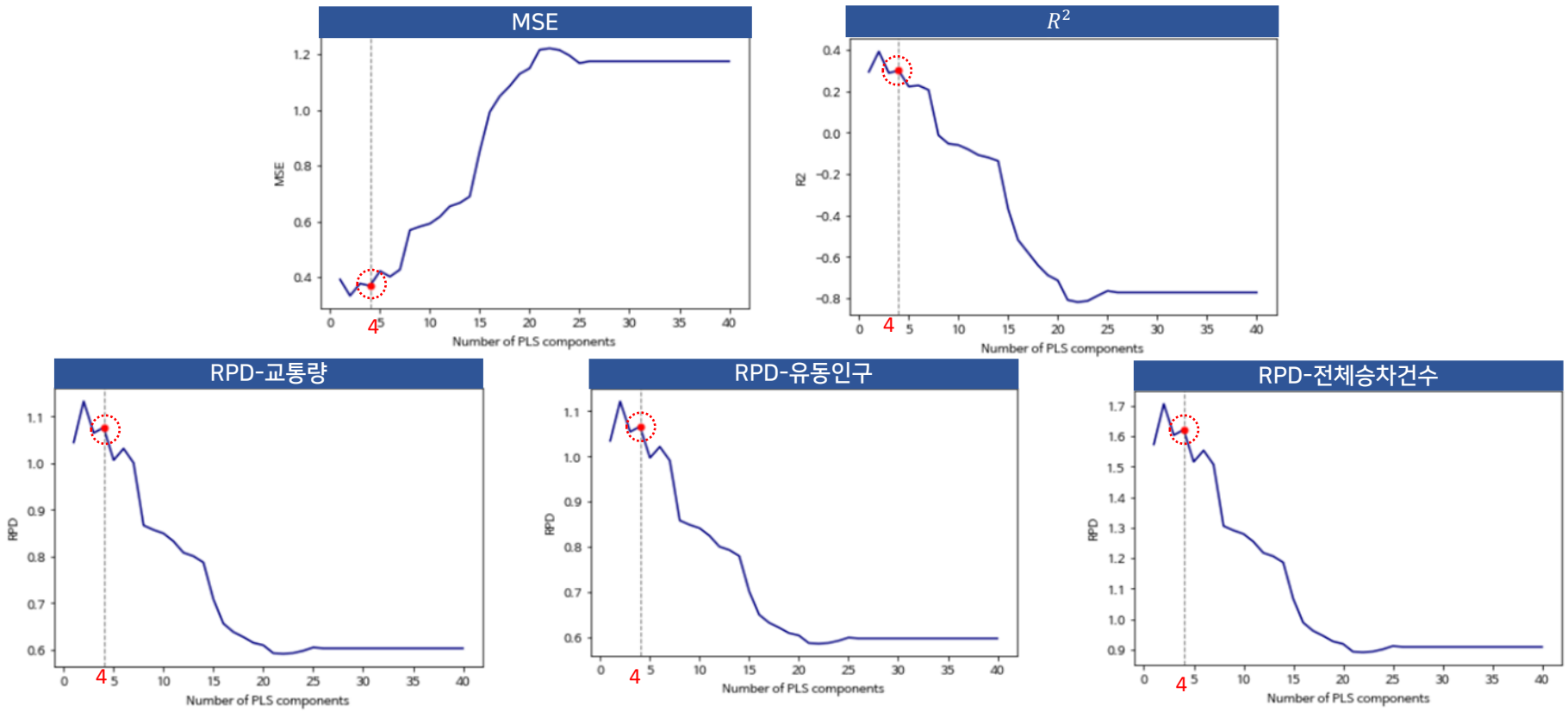
### 장안구

- 장안구에서 선택된 2개의 클러스터로 진행 (1+28=총 29개 버스 정류장)
- Y의 값이 다소 skewed된 형태이기 때문에 log scaling 후 진행



# 구 별 PLSR 결과

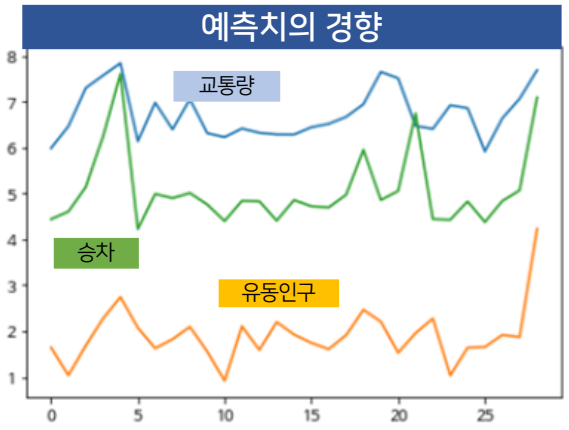
## 최적 차원 개수 찾기

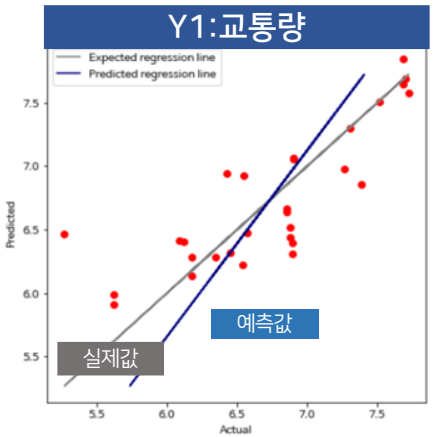


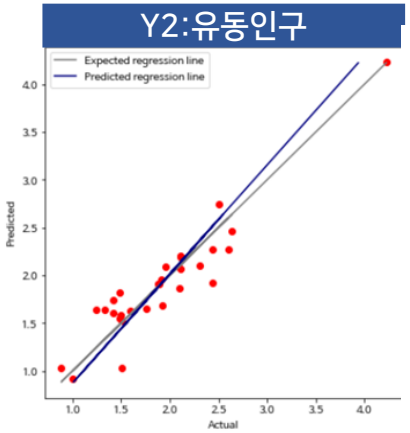

→ MSE 최소, RPD최대, $R^2$최대가 되는 차원개수 : 4

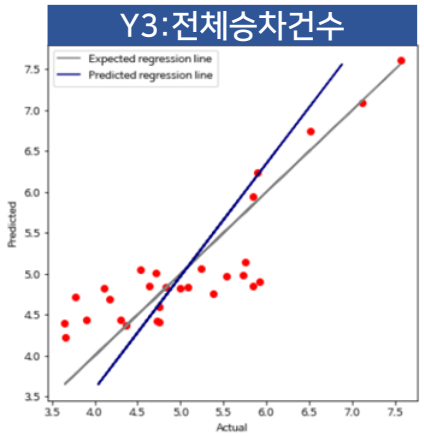
# 구 별 PLSR 결과

최적 차원 개수 4개로 PLSR 진행 ($R^2$: 0.9106 , MSE: 26553.7553)




- 전체승차수, 유동인구, 교통량 예측치가 비슷한 양상
- 장안구에서 전체 승차가 많은 버스 정류장은 유동인구가 많고 교통량이 많은 곳이라 생각됨

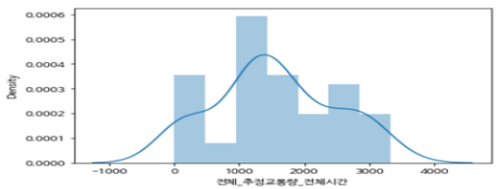
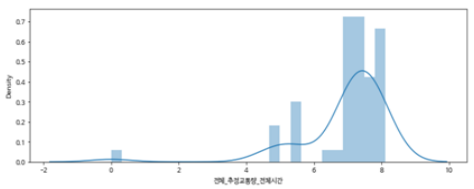
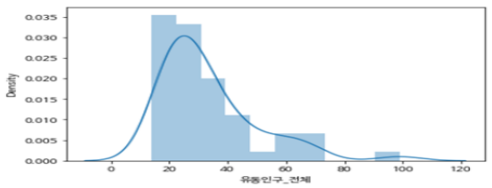
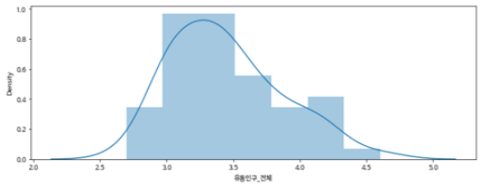
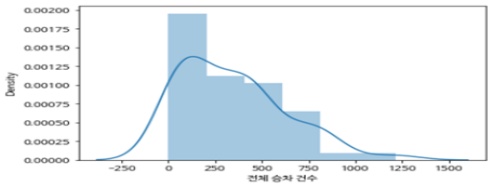
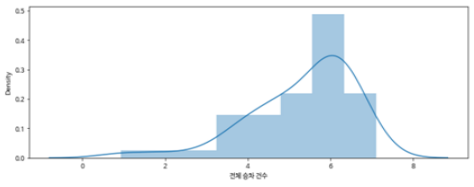
- 유동인구와 전체승차건수가 모형에 매우 잘 적합했음을 볼 수 있음
- 전체 $R^2$ = 0.91로 모형이 전체적으로 적합하다고 판단
- 위의 회귀식을 이용하여 $\widehat{y_1}$, $\widehat{y_2}$ , $\widehat{y_3}$ 을 각각 구함
- 최종 스마트 지수 생성 : smart= scaled $\widehat{y_1}$+scaled $\widehat{y_2}$ + scaled $\widehat{y_3}$

# 구 별 PLSR 결과

## PLSR 결과

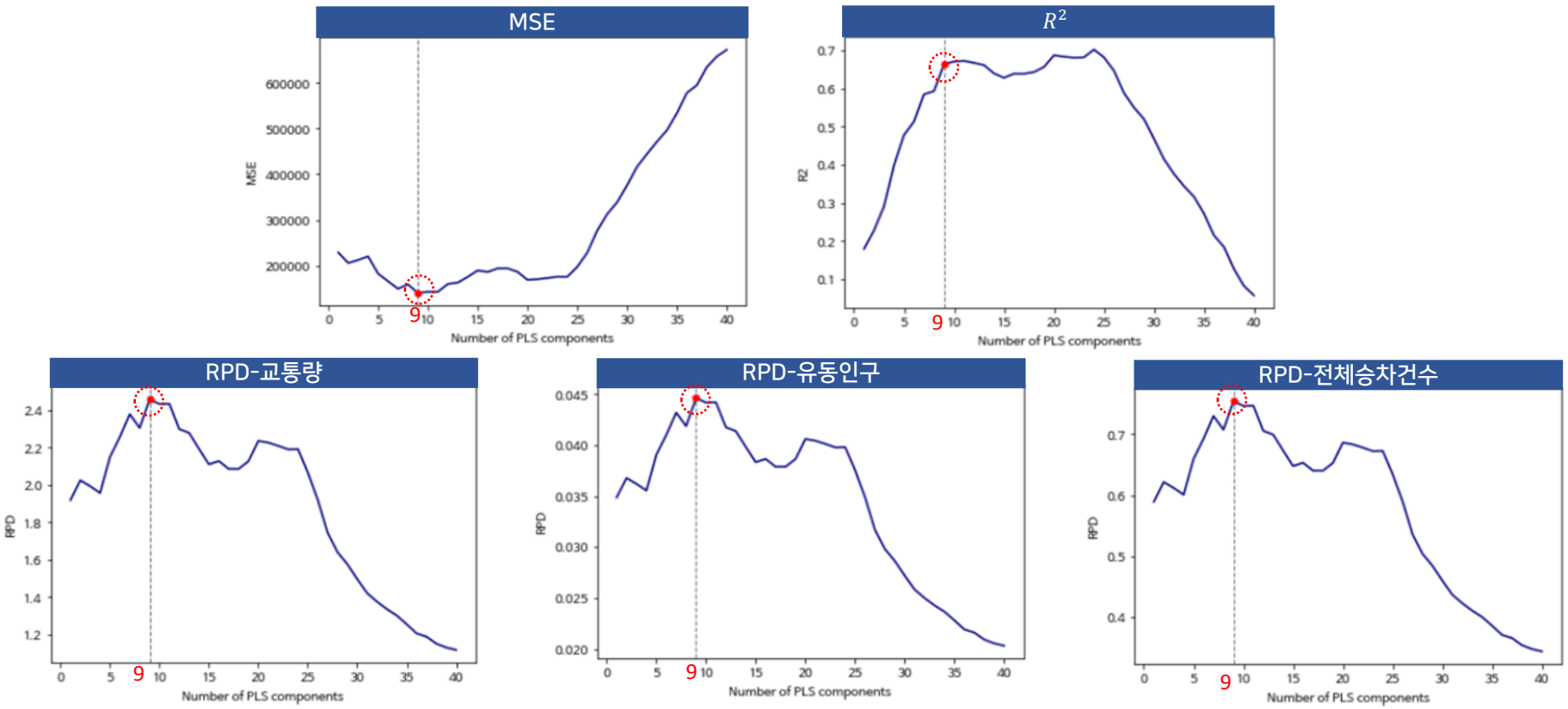
### 권선구

- 권선구에서 선택된 2개의 클러스터로 진행 (9+44=총 53개 버스 정류장)
- Y의 분포의 왜곡 정도가 심하지 않고 스케일링시 왜곡이 더 심해지므로 스케일링 없이 진행

| | scaling 전 | scaling 후 |
|---|---|---|
| 교통량 |  |  |
| 유동인구 |  |  |
| 승차건수 |  |  |

# 구 별 PLSR 결과

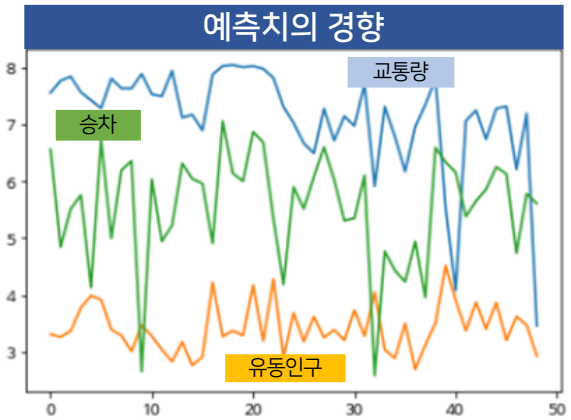
## 최적 차원 개수 찾기

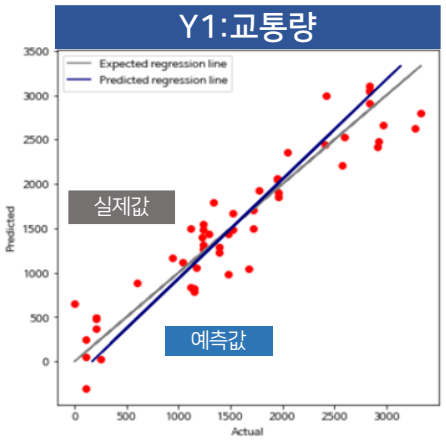


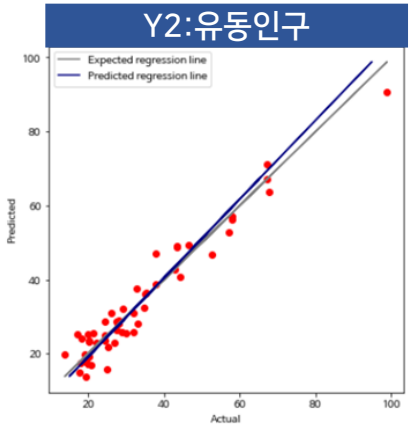

→ MSE 최소, RPD최대, $R^2$최대가 되는 차원개수 : 9

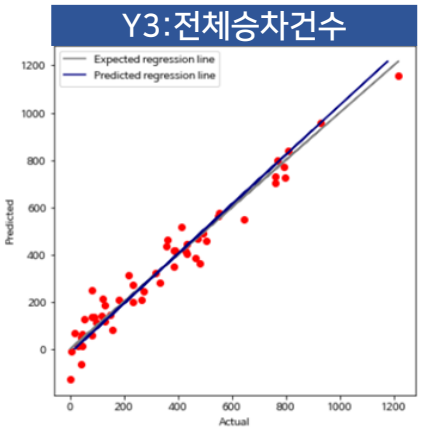
# 구 별 PLSR 결과

최적 차원 개수 4개로 PLSR 진행 ($R^2$: 0.9282, MSE: 31074.2998)




- 권선구는 다른 구에 비해 변수 간 경향성이 뚜렷하게 비슷하지는 않으며, 정도의 차이는 있지만 등락의 경향이 비슷한 양상을 보임

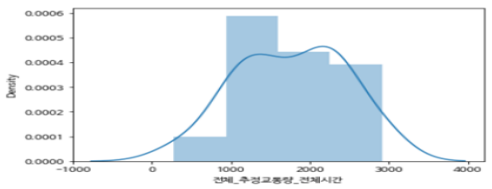
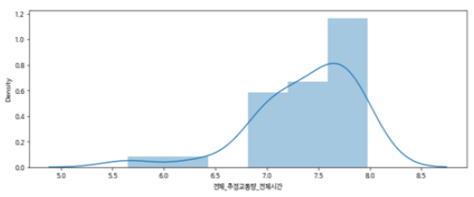
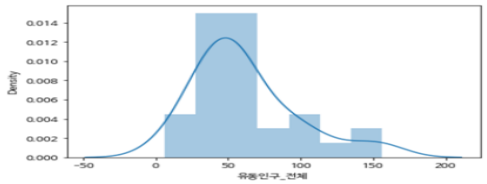
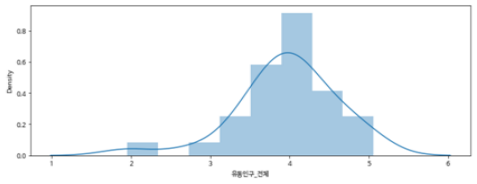
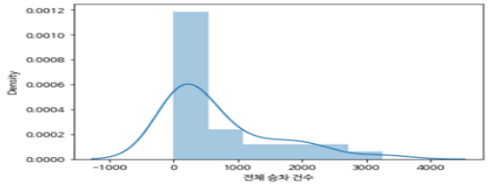
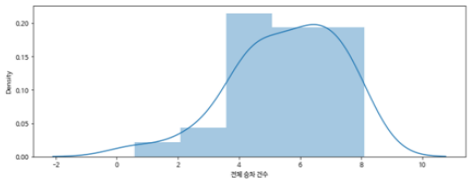
- 교통량, 유동인구와 전체승차건수가 모형에 매우 잘 적합했음을 볼 수 있음
- 전체 $R^2$ = 0.93로 모형이 전체적으로 적합하다고 판단
- 위의 회귀식을 이용하여 $\widehat{y_1}$, $\widehat{y_2}$ , $\widehat{y_3}$ 을 각각 구함
- 최종 스마트 지수 생성 : smart= scaled $\widehat{y_1}$+scaled $\widehat{y_2}$ + scaled $\widehat{y_3}$

# 구 별 PLSR 결과

## PLSR 결과

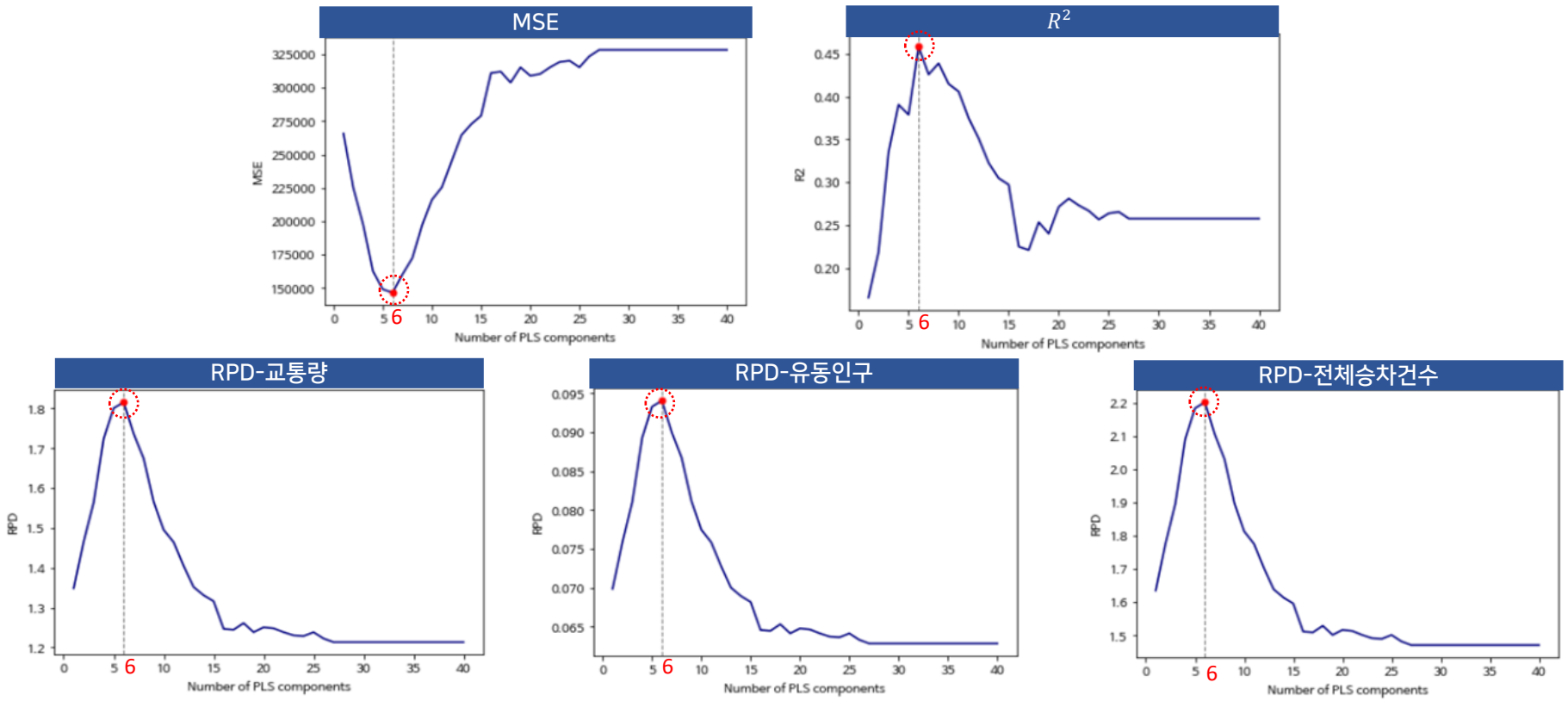
### 영통구

- 영통구에서 선택된 3개의 클러스터로 진행 (8+4+19=총 31개 버스 정류장)
- Y의 분포의 왜곡 정도가 심하지 않고 스케일링시 왜곡이 더 심해지므로 스케일링 없이 진행

| | scaling 전 | scaling 후 |
|---|---|---|
| 교통량 |  |  |
| 유동인구 |  |  |
| 승차건수 |  |  |

# 구 별 PLSR 결과

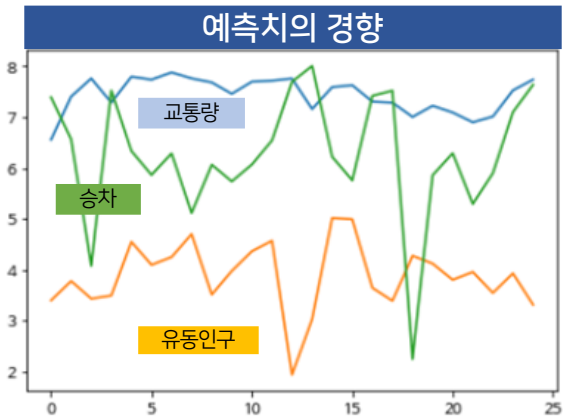
## 최적 차원 개수 찾기

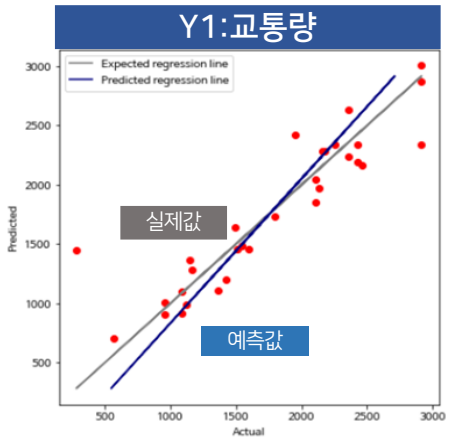


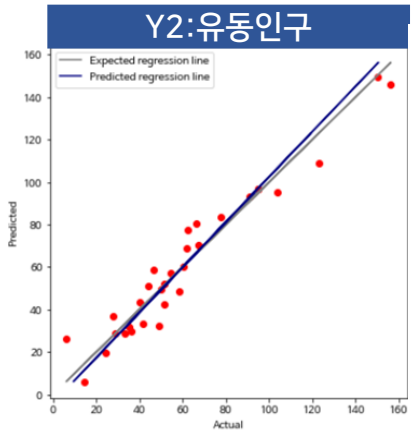

→ MSE 최소, RPD최대, $R^2$최대가 되는 차원개수 : 6

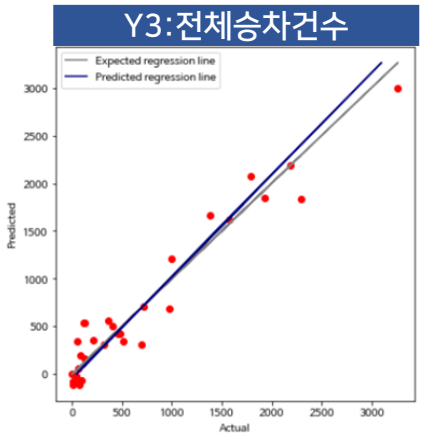
# 구 별 PLSR 결과

최적 차원 개수 4개로 PLSR 진행 ($R^2$: 0.8992, MSE: 42478.2176)




- 영통구는 다른 구에 비해 변수간 경향성이 비슷하지 않음

- 교통량, 유동인구와 전체승차건수가 모형에 매우 잘 적합했음을 볼 수 있음
- 전체 $R^2$ = 0.899로 모형이 전체적으로 적합하다고 판단
- 위의 회귀식을 이용하여 $\widehat{y_1}$, $\widehat{y_2}$ , $\widehat{y_3}$ 을 각각 구함
- 최종 스마트 지수 생성 : smart= scaled $\widehat{y_1}$+scaled $\widehat{y_2}$ + scaled $\widehat{y_3}$

# 최적 입지 선정

## 1. 스마트 지수 기준 Ranking

### *스마트 지수*

- 스마트 버스정류장 수요를 수치화한 값
- *PLS Regression*으로 추정한 추정치를 **스케일링**하여 더한 값
- (스마트 지수) = $\widehat{y_1} + \widehat{y_2} + \widehat{y_3}$

$\widehat{y_1}$ : 전체 교통량 추정치
$\widehat{y_2}$ : 유동인구 추정치
$\widehat{y_3}$ : 승차건수 추정치

스마트 지수가 높을수록 스마트 버스정류장에 대한 수요가 높은 것을 의미하므로
높은 우선순위 부여

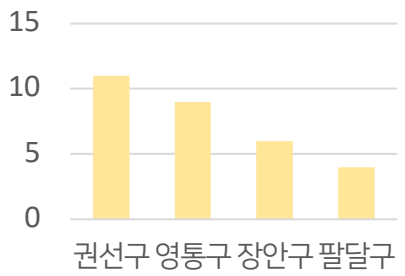
# 최적 입지 선정

## 2. Constraint

: 최대한 많은 사람들의 스마트 버스정류장 이용을 위한 **제한조건**

### 1. 구별 버스정류장 비율

기존의 구별 버스정류장 비율을 유지하여
고르게 구별 설치 개수 설정




| | 기존 | 비율 | 설치개수 |
|---|---|---|---|
| 권선구 | 429 | 0.36 | 11 |
| 영통구 | 345 | 0.29 | 9 |
| 장안구 | 253 | 0.21 | 6 |
| 팔달구 | 152 | 0.13 | 4 |

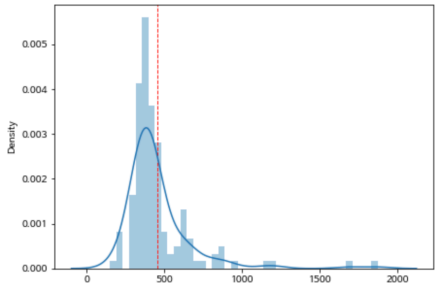
### 2. 버스정류장 간 평균 거리 계산

근접 거리에 스마트 버스정류장이 여러 개 설치되는 것 방지

PROCESS

1. 수원시_버스정류장 + 경기도 버스 노선 순서 & 정보 데이터를 이용하여, 버스 정류장에 대한 노선 순서 데이터 생성
2. 버스 노선 순서로 정렬 후, 이전 정류장과의 거리 계산 (연속된 순서의 정류장끼리만 계산)
3. **노선마다 정류장 간 평균 거리 계산**
4. 전체 노선에 대해 평균 거리 값 구함

| 노선번호 | 정류장명 | lon | lat | next_lon | next_lat | 거리(m) | 노선 순서 |
|---|---|---|---|---|---|---|---|
| 16 | 화서역 | 126.990107 | 37.283895 | 126.991358 | 37.281791 | 197.940 | 15 |
| | 숙지중고교.서호공원 | 126.991358 | 37.281791 | 126.995919 | 37.273909 | 731.636 | 16 |
| | 쌍우물택시쉼터 | 126.995919 | 37.273909 | . | . | | 17 |
| | 수원역환승센터(2승강장) | 126.999279 | 37.265404 | 126.999248 | 37.262758 | 177.117 | 21 |
| | 평동동남아파트 | 126.999248 | 37.262758 | 126.995538 | 37.260691 | 435.015 | 22 |




전체 노선 평균 거리
Mean = 457.691
→ Constraint 458(m) 설정

4

MODEL

# Optimization Process

## 최적 입지 선정 과정

### 1. 구내

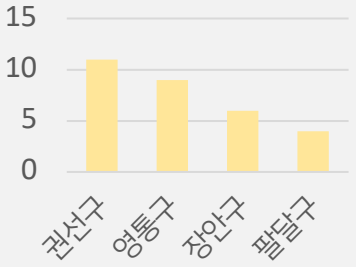
*Step1.*

**스마트 지수 기준**으로
구내 버스정류장 우선순위 부여

*Step2.*

[constraint]
일정거리보다
가까운 버스정류장 중
**낮은 우선순위의**
**버스정류장 제거**

*Step3.*




x 2

우선순위의
버스정류장
구별 설치 개수의
약 **2배수** 선택

### 2. 통합

*Step4.*

구별로 선택된 버스정류장
**통합** 후 **우선순위 재부여**

*Step5.*

[constraint]
구 간 근접하게 선택된
버스정류장을 고려하여
**일정거리 내에 있는**
**낮은 순위의 버스정류장 제거**

*Step6.*

구별 버스정류장 **비율 유지**하여
**우선순위**의 버스정류장 선택

*Step7.*

유동인구 기준으로
**광고 Target층** 선정

4

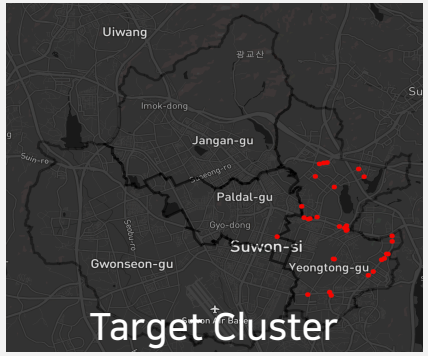
MODEL

# 구내 Constraint

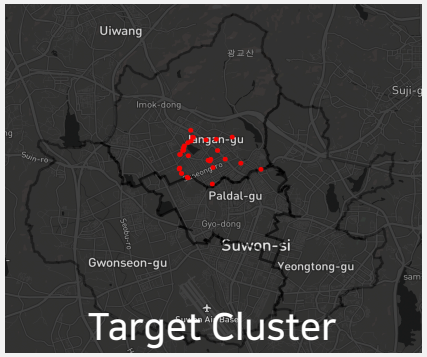
1. **스마트 지수 기준**으로 구내 버스정류장 우선순위 부여
2. 일정거리보다 가까운 버스정류장 중 **낮은 우선순위의 버스정류장 제거**
3. 이후 구 경계에 있는 버스정류장까지 고려하기 위해 **높은 순위의 버스정류장 구별 설치 개수의 약 2배수 선택**

## 팔달구

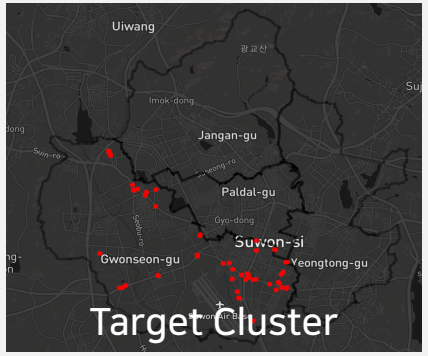
## 영통구

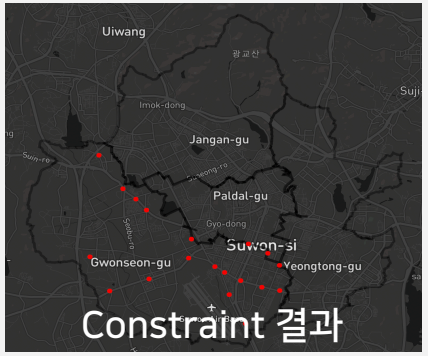


## 장안구



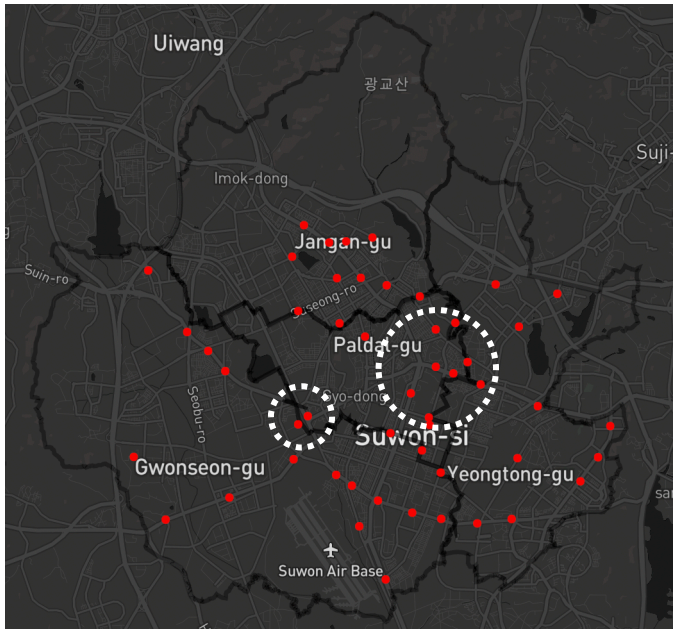
## 권선구



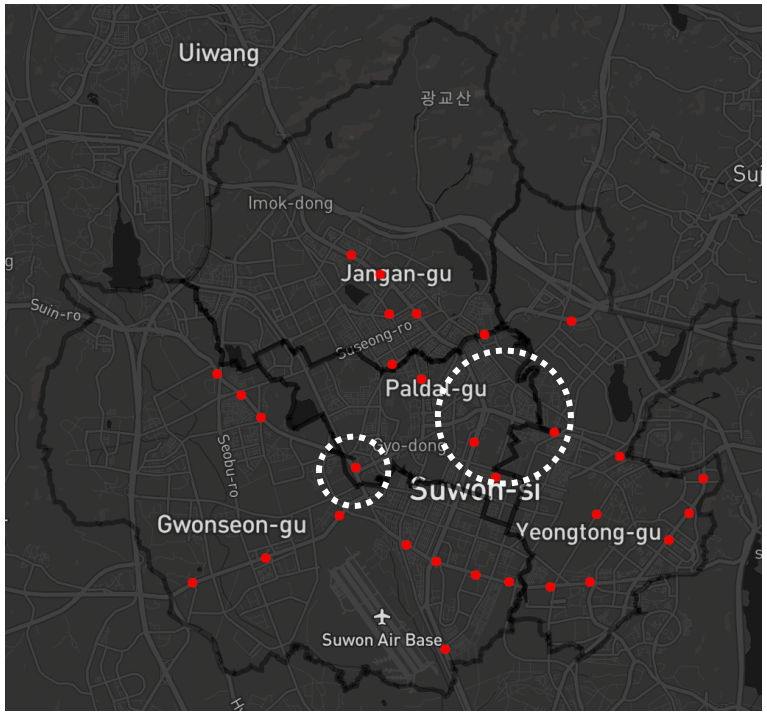



# 통합 Constraint

4. 구별로 선택된 **버스정류장 통합 후 우선순위 재부여**
5. 구 경계에 있는 버스정류장 고려하여 **일정거리보다 가까운 버스정류장 제거**
6. 구별 버스정류장 **비율 유지**하여 **높은 순위**의 버스정류장 선택




구내 Constraint 적용 결과




통합 Constraint 적용 결과

4 MODEL

# 최종 입지 선정 결과

## 스마트 버스정류장 우선 설치위치 30개소

| 순위 | 정류장명 |
|---|---|
| 1 | 수원역.AK플라자 |
| 2 | 수원버스터미널 |
| 3 | 우만동4단지 |
| 4 | 한일타운.경기일보.홈플러스 |
| 5 | 곡반초등학교 |
| 6 | 영통아이파크캐슬1단지.래미안영통마크원2단지 |
| 7 | 고색역.고색초교.태산아파트 |
| 8 | 수원아이파크시티.선일초교 |
| 9 | 원천레이크파크아파트 |
| 10 | 수원광교박물관.광교카페거리 |

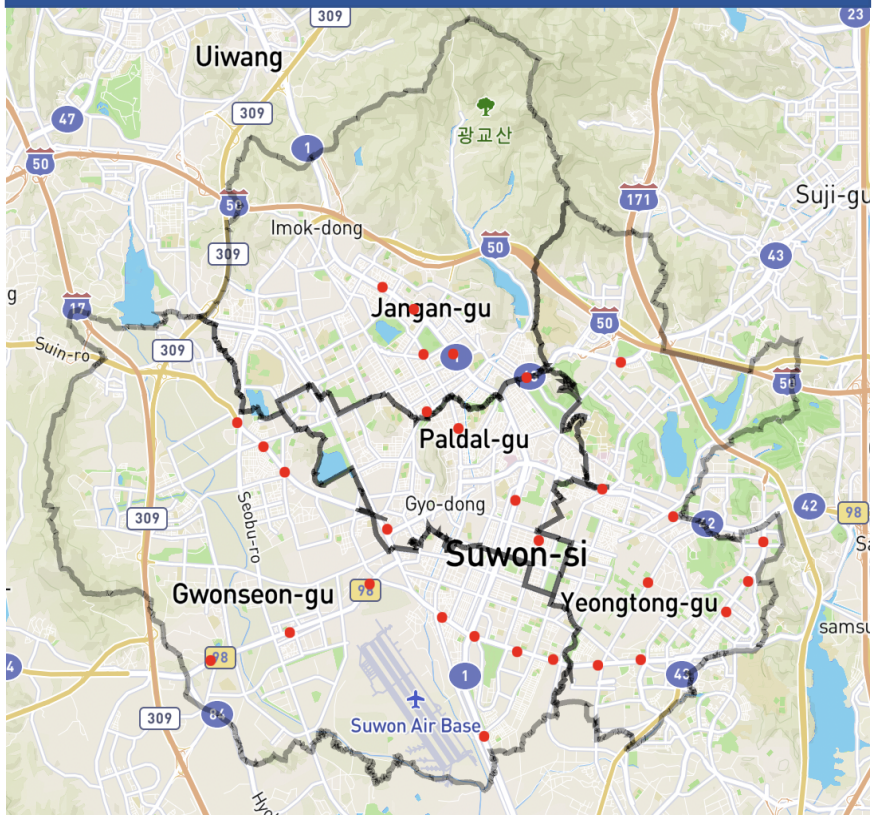
| 순위 | 정류장명 |
|---|---|
| 11 | 청명역3번출구 |
| 12 | 세류2동주민센터 |
| 13 | 망포역4번출구.영통2동주민센터 |
| 14 | 영통역 |
| 15 | 오목천동태산아파트.서희스타힐스 |
| 16 | 뉴코아아울렛 |
| 17 | 웃거리 |
| 18 | 구운오거리.경기중소기업성장지원센터 |
| 19 | 곡반정동입구 |
| 20 | 공구유통타운.중앙자동차매매단지 |

| 순위 | 정류장명 |
|---|---|
| 21 | 구운사거리.농수산물유통센터 |
| 22 | 구법원사거리 |
| 23 | 수일중학교 |
| 24 | 황골주공.벽산아파트 |
| 25 | 영화초등학교.한국건강관리협회 |
| 26 | 팔달구청.화성행궁.수원성지 |
| 27 | KBS수원센터정문.한화생명 |
| 28 | 삼성DigitalCity홍보관 |
| 29 | 수원종합운동장,로얄팰리스,아너스빌위즈 |
| 30 | 한국은행 |

# 최종 입지 선정 결과

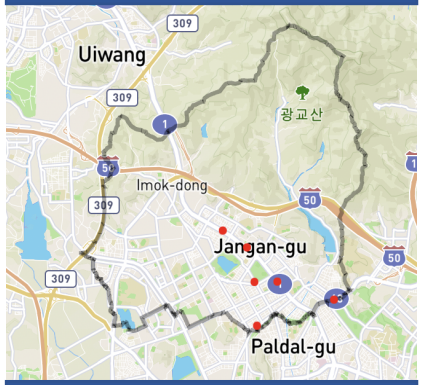
## 최종 결과 확인

최종 선정 입지 30개소



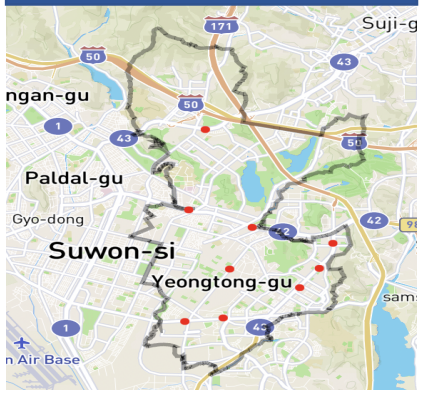

권선구(11개)

장안구(6개)




팔달구(4개)

영통구(9개)

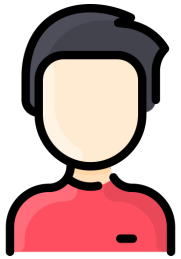


# 광고 Targeting

## 성별 가중치 부여

- 성별에 따른 광고 타겟팅 효과를 반영하기 위한 **가중치 (구매의도지수)** 부여

(여자 유동인구 수)
x (가중치 = 4.27)



(남자 유동인구 수)
x (가중치 = 2.94)

광고 Target층
: 유동인구가 가장 많은 (연령대 + 성별)

| 성별 | 여성 | 남성 |
|---|---|---|
| 구매 의도 | 4.27 | 2.94 |

옥외광고에서, 성별에 따른 광고효과의 차이 존재

출처: 성별과 관여도에 따른 옥외광고의 커뮤니케이션 효과(한광석,2005)

| 순위 | 정류장명 | 광고 Target층 |
|---|---|---|
| 1 | 수원역.AK플라자 | 20대 여성 |
| 2 | 수원버스터미널 | 40대 남성 |
| 3 | 우만동4단지 | 40대 남성 |
| 4 | 한일타운.경기일보.홈플러스 | 50대 여성 |
| 5 | 곡반초등학교 | 50대 여성 |
| 6 | 영통아이파크캐슬1단지.래미안영통마크원2단지 | 30대 여성 |
| 7 | 고색역.고색초교.태산아파트 | 50대 남성 |
| 8 | 수원아이파크시티.선일초교 | 40대 여성 |
| 9 | 원천레이크파크아파트 | 40대 여성 |
| 10 | 수원광교박물관.광교카페거리 | 40대 여성 |
| … | … | … |
| 30 | 한국은행 | 60대 남성 |

# 5

# 결론

# 5 결론

# 결과 해석

## 1. 유동인구 & 교통량 많은 버스정류장

### 우선순위1. 수원역 .AK 플라자

- 근거리에 수원역 분당선과 1호선이 있고 AK 플라자, 롯데백화점, 수원메쎄, 아파트 단지들이 많이 있어 스마트버스정류장의 **수요가 높은 곳**이라 예상할 수 있음.
- 또한, **중앙차로**에 위치하고 있어 이용객의 **대기먼지 노출도**가 심하기 때문에 최우선적으로 고려되어야 할 곳
- 광고 타겟층: 20대 여성

### 우선순위2. 수원버스터미널

- NC 백화점 앞에 위치하고 있고 근거리에 수원버스터미널, 이마트, 초등학교, 중학교 및 아파트 단지들이 있어 **이용인원**이 많을 것이라 판단 가능
- 또한, 옆에 **사거리**를 두고 있어 교통량 또한 많아, 광고 접근성이 좋을것이라 예상됨
- 광고 타겟층: 40대 남성

### 우선순위3. 우만동4단지

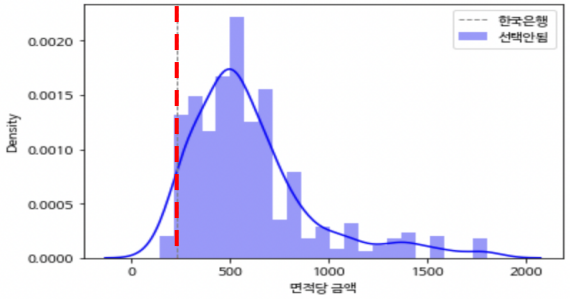
- 아파트단지 앞에 위치하고 있고 주변에 초등학교와 중학교, 월드컵 경기장 입구가 있어 **이용인원**이 많을 것이라 판단 가능
- 또한, 인근에 공영주차장과 사거리가 있고 **8차선 도로**에 위치하고 있어 교통량 또한 많아 광고 접근성이 좋을 것이라 예상됨
- 광고 타겟층: 40대 남성

# 결과 해석

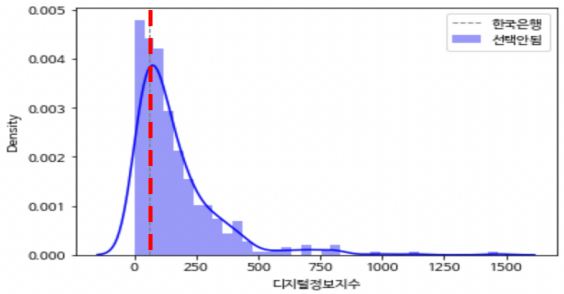
## 2. 취약계층을 위한 버스정류장

### 우선순위30. 한국은행

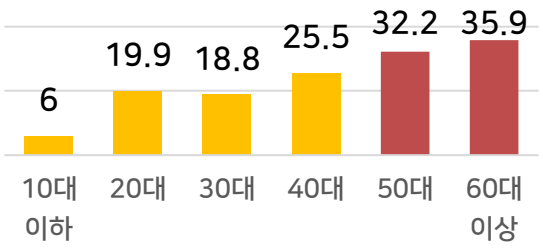
- 주변에 경로당, 요양병원, 한의원 등이 위치하고 있는 것으로 보아 **고령인구**가 많이 사용하는 지역임을 알 수 있음
- 광고 타겟층: 60대 남성




면적당 금액이 다른 지역에 비해 낮은 것으로 보아 상대적으로 **저소득층** 지역임을 알 수 있음




디지털정보화지수가 다른 지역에 비해 낮은 것으로 보아 상대적으로 **정보 소외계층**임을 알 수 있음




**연령별 유동인구 수**

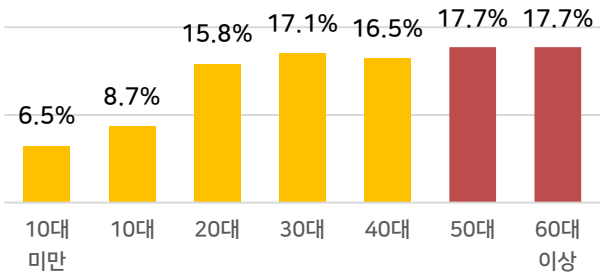
상대적으로 50대, 60대 이상의 유동인구 수가 많은 것으로 보아 고령인구가 많이 이용하는 지역임을 알 수 있음

# 결과 해석

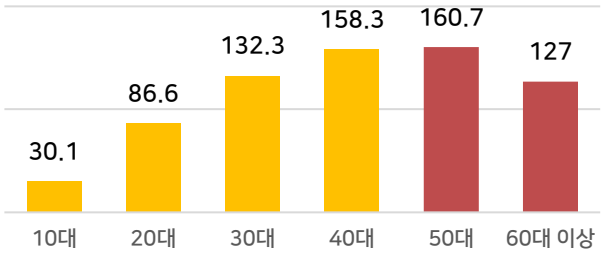
## 2. 취약계층을 위한 버스정류장

### 우선순위12. 세류2동 주민센터

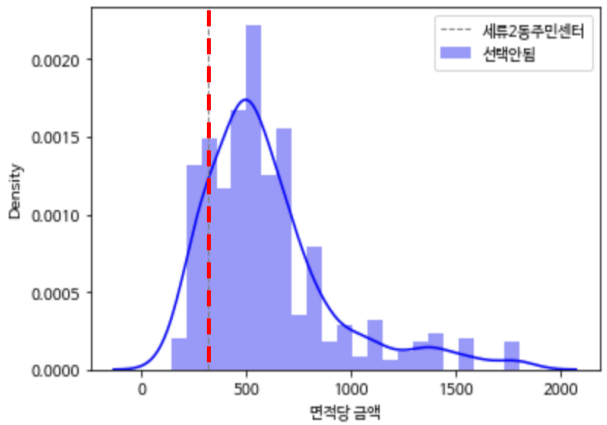
- 정류소 주변에 행정복지센터가 위치하고 있어 **시민 편익 증진** 및 주민 자치 기능에 기여하므로 공공성 실현
- 광고 타겟층: 50대 남성




연령별 거주 인구




연령별 유동인구



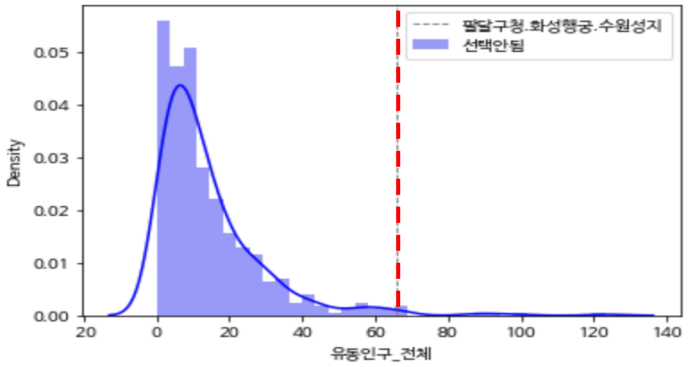

세류2동 주민센터 정류장 주변 아파트의
**면적당 금액**은 평균에 비해 낮으므로
상대적 **저소득층** 지역이라고 할 수 있음
따라서 기존 취약계층을 고려한
분석 목표에 부합하는 입지라 판단됨

# 결과 해석

## 3. 지역의 상징성 고려한 버스정류장

### 우선순위26. 팔달구청, 화성행궁, 수원성지

- 정류장 주변에는 박물관,전시관, 문화관 등의 **관광지 및 명소**가 많아 **수원시의 상징성을 보여주는 지역**이라 할 수 있음
- 광고 타겟층: 50대 남성




타 정류장에 비해 **유동인구**가
월등히 많은 양상을 띠는 것으로 보아
**관광지**의 특성이 드러남

상징적이고 관광객이 많은 위치에
스마트 버스정류장을 설치함으로써
수원시 이미지 제고

# 결론

## 의의 및 한계점

| | |
|---|---|
| 의의 | • 취약계층을 고려한 분석<br>- 디지털정보지수, 면적 당 금액 변수를 사용한 분석<br>• 최대한 많은 수원시민에게 스마트 버스정류장을 경험할 수 있도록 배치<br>- 구별로 각자 특성을 반영하여 고르게 배치<br>• 실제 사례를 참고하여 목표 설정하였기 때문에 현실성 있고 구체적인 결과 도출 |
| 한계점 | • 데이터 제약<br>- 격자 데이터를 가장 가까운 버스정류장에 병합했기 때문에 다소 부정확한 수 있음<br>- 수원시 미세먼지 측정소가 8개뿐이어서 정류장별 차이를 정확하게 파악할 수 없었음 |

# Reference

## 1. 출처

[1] ‘LH, 미세먼지 잡는 버스정류장 선보인다’, 중앙일보, 2019.04.28
[2] ‘수원시, '스마트챌린지 솔루션 확산사업' 대상지 선정’, 에듀인뉴스, 2021.03.08
[3] ‘수원시, 과학기술정보통신부 공모 선정으로 국비 27억 7000만원 확보’, 한국분양정보, 2021.02.22
[4] ‘수원시, 버스정류장 14곳에 온열의자 추가 설치’, 연합뉴스, 2018.02.13
[5] 2021년 신년사, 수원시장, 2021.01.01
[6] 수원시 대기오염 민감군 피해저감 방안 연구, 수원시정연구원 강은하, 2019
[7] 다변량 통계분석 방법을 이용한 Online Monitoring Methods , 포항공과대학교 화학공학과
[8] Variable and subset selection in PLS regression, technical university of Denmark, 2000.11.06
[9] Multivariate Linear Models, Washington University, 2010.11.08
[10] 성별과 관여도에 따른 옥외광고의 커뮤니케이션 효과, 한광석, 2005
[11] http://www.gisdeveloper.co.kr/?p=4784

## 2. 분석 툴

python™ QGIS

# 감사합니다