

2020  
공공빅데이터  
청년인턴십

전주 1반 4조

# ● 공공형 택시의 최적 지역 분석 및 이용객 수 예측

데이터 기반 행정으로 국민의 삶의 질을 개선하라!

# 목차

Contents

- 01  
주제 선정

주제 선정 배경 및 목표

- 02  
데이터 분석

데이터 전처리 및 모델 구축

- 03  
결론

분석 내용 및 결과

# 01

Key word

## 주제 선정 배경 및 문제점

1. 공공형 택시 정의 및 목적
2. 전주 모심 택시 분석배경
3. 분석목표 및 절차

## 공공형 택시란?

버스 노선이 없거나 버스 정류장에서 거리가 멀어 대중교통 이용이 불편한 지역 주민들을 대상으로 택시비를 지원하는 사업. 이용자가 일정액을 부담하면 나머지 요금을 국비와 도비, 시·군비로 보조함



### 목 적

- ✓ 교통 취약 지역에 있는 계층에게 형평성 있는 교통편의를 제공하여 대중교통 공백 해소
- ✓ 100원 택시, 마중택시, 이음택시, 희망택시, 행복택시 등 지자체에서 다양한 이름으로 운영 중
- ✓ 전주시도 2017년부터 '모심 택시'라는 이름으로 공공형 택시 운영 중

### 다양한 공공형 택시



화순군 100원 택시



보성군 행복택시



전주시 모심택시

## 01 주제 선정 배경 및 문제점

### - 전주 모심 택시 분석배경

#### 문제점

##### 거리

마을에서 가장 가까운 승강장까지  
거리가 800m 이상인 마을

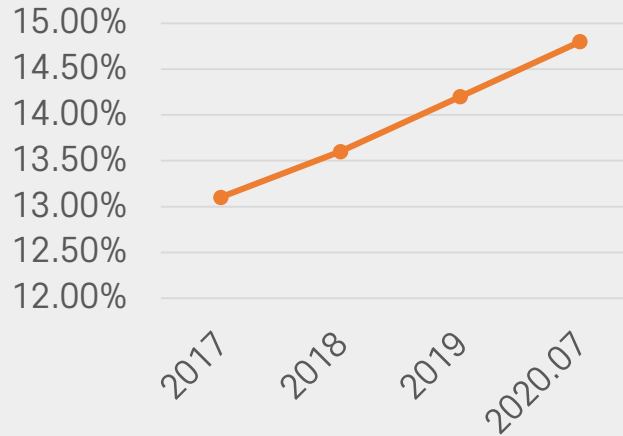
##### 배차시간

현재 시내버스가 운행되지 않거나,  
배차간격이 3시간 이상인 마을

모심 택시의 마을 선정 방식은  
거리나 버스의 배차시간만을  
고려한 정도로 **사각지대를  
해소하기에 너무 단순화** 됨

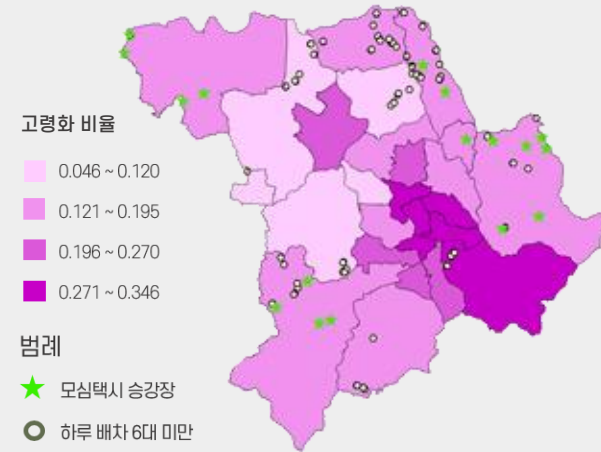
거리와 배차조건에 적합한데도  
선정되지 않은 마을 다수

##### 전주시 고령화 비율



주 이용객이 고령인 것을 고려할  
때, 전주시 고령화 비율은  
지속적으로 늘어났지만  
**모심택시는 시행 이후 증차된  
적이 없음**

##### 고령화 비율 및 모심택시 승강장 비교



전주시 동별 고령화 비율과  
모심택시 승강장 비교 결과  
**고령화 비율이 높은 곳에  
승강장이 없음**을 확인

최적화된 공공형 택시 배차기준과 위치를 확대 및 재조정 필요

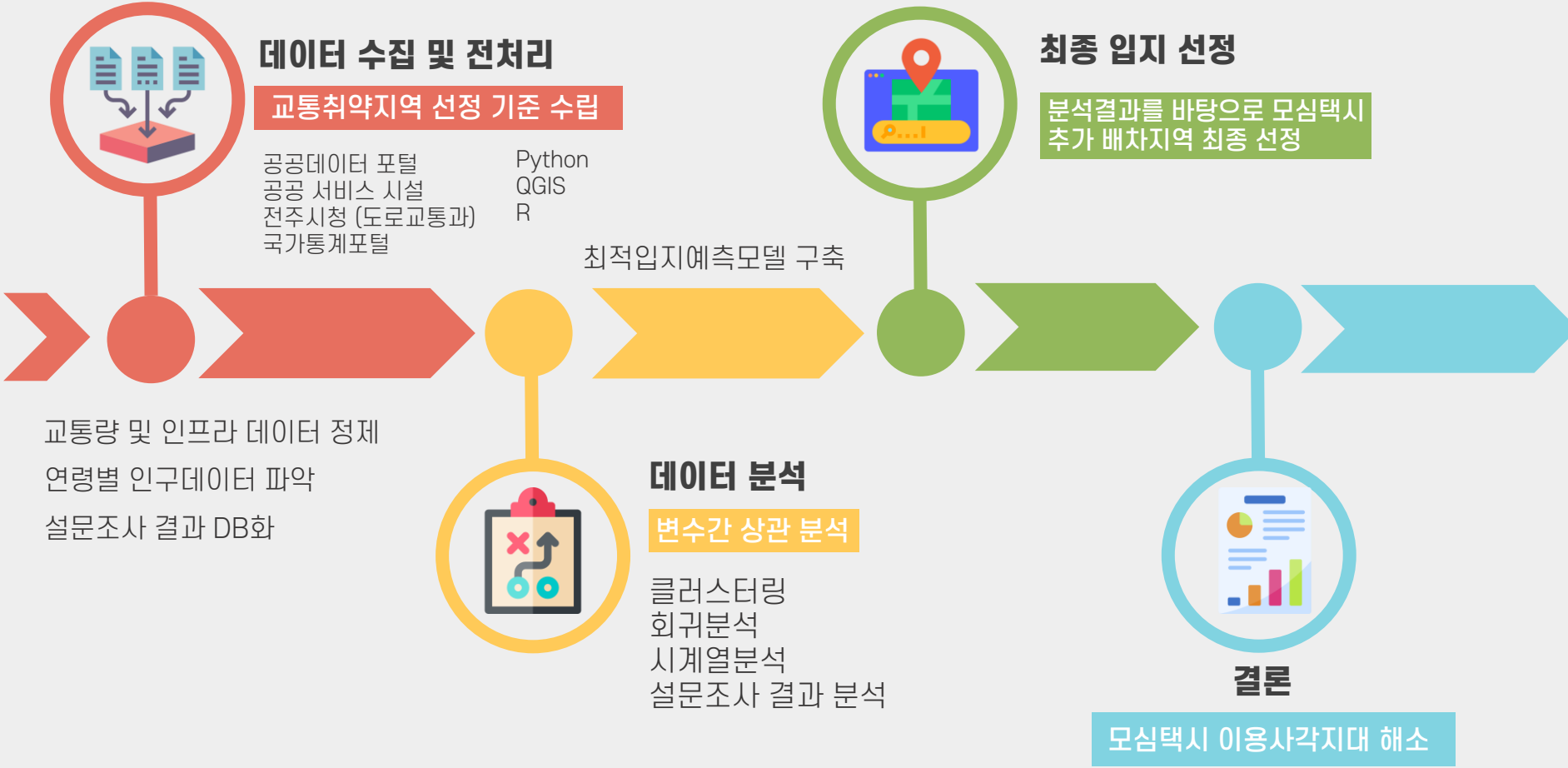
01 주제 선정 배경 및 문제점

- 분석 목표 및 절차

프로젝트 분석 목표

모심 택시 확대 및 재조정을 통해서 교통 취약 지역에게 형평성 있는 교통편의를 제공

프로젝트 분석 절차



# 02

## 데이터 분석

1. 현장 조사
2. 데이터 전처리
3. 데이터 분석 및 모델 구축

전주 시청 담당자 인터뷰

Monday, September 21, 2020



담당 공무원

대상의 제한은 없지만 60대 이상의 고령층이 주로 이용

병원 치료

토요일 오전까지 추가 운행 요망, 운행 시작 시간 변경 희망, 노선 별 운전자 고정 요망 등의 의견 존재

노선별로 출발지역과 종착지역이 고정되어 있으나, 운행자 분의 재량껏 노선 중간 지역에서도 하차 할 수 있도록 유동적으로 운행 중.

가능 하나 올해의 경우 코로나 사태로 인해 되도록 자제할 것으로 운행 지침 중

전주시 모심택시 주요 이용층

설문 내용에서 이용자들의 가장 빈도가 높은 이용 목적

이용자들의 건의사항

승차지역과 하차지역의 고정 여부

승차지역이 같은 경우, 합승 가능 여부



우아동의 3개 마을(장재, 원산정, 왜망실) 조사 및 시민인터뷰

Monday, September 21, 2020

모심택시 만족도 조사



장재 마을 주민(여성, 70대, 택시 주 1~2회 이용, 버스 X)

- 모심택시에 관해서 만족도가 **매우** 높음
- 모심택시 생긴 후 버스 보다는 모심택시를 더 많이 이용
- 모심택시 주 이용 목적 : **병원 진료** 및 **장**을 보러 시장을 나감
- 모심택시 이용 후 달라진 점 : 생활편의가 한없이 좋아져서 정말 좋음
- 장점 : 택시기사가 친절하며, 가격이 저렴해서 자주 이용



원산정 마을 주민(남성, 70대, 택시 주 3-4회 이용, 버스 주 1-2회 이용)

- 대체적으로 만족하나 배차시간이 더 많았으면 좋겠음.
- 모심 택시 주 이용 목적 : **병원 진료** 혹은 **모임**이나 **약속**
- 모심 택시 이용 후 달라진 점 : 생활 편의
- 모심 택시 장점 : 항상 제 시간에 오는 점이 좋음
- 희망 개선 사항 : **증차** 강력희망, **승강장과의 거리가 멀**



왜망실 마을 주민(여성, 60대, 택시 주 1~2회 이용, 주 1~2회 버스 이용)

- 모심택시 운영에 대체적으로 만족하나, 택시 배차시간 및 승강장까지의 거리에 있어 불편 호소
- 모심택시 주 이용 목적 : **장보기**
- 모심택시 이용 후 달라진 점 : 이용 후 생활편리
- 모심택시 장점 : 택시기사가 친절하며 가격도 저렴
- 희망 개선 사항 : **배차 시간이 더 가까웠음을 희망**

현장조사

전주시청 모심 택시 담당자 인터뷰



모심택시 운행 지역 주민 인터뷰



현장 조사 및 관련자 인터뷰를 통해 데이터 수집 및 데이터 분석 프로세스 구상

## 02 데이터 분석

### - 데이터 전처리

#### 분석 데이터 추가

##### 기존 모심택시 데이터

거리

배차  
시간

기존 모심택시 데이터는  
거리와 배차시간만 고려



##### 재조정 된 모심택시 데이터

거리



배차  
시간

취약인구비율

교통취약지역

의료서비스 취약지역

공공서비스취약지역

마을 회관 위치(경로당)

취약 지연을 보다 구체적으로  
알기 위해 조건&입지 추가

#### 데이터 목록

##### 대중 교통 데이터



##### 의료 시설 데이터



##### 인구 데이터



##### 인구 데이터



##### 공공 서비스 데이터



##### 상권 데이터



##### 전주시 모심 택시 데이터



## 02 데이터 분석

### - 데이터 전처리

#### 사용 언어

R



시계열 분석  
설문조사 분석  
회귀 분석

QGIS



지도 시각화

Python



데이터 전처리  
클러스터링  
상관관계 분석  
예측모델 구축

## 02 데이터 분석

### - 시계열 분석



#### 데이터 분석

R을 이용한 시계열 분석

Arima model을 이용한 월별 모심택시 이용객수 예측

### Arima model 이란?

시계열로 변환한 데이터에 대해 정상성을 만족하는 적절한 모델을 형성  
월별 행복택시 탑승 인원 데이터로 향후 행복택시 이용객을 예측하고자 사용

Arima 모형

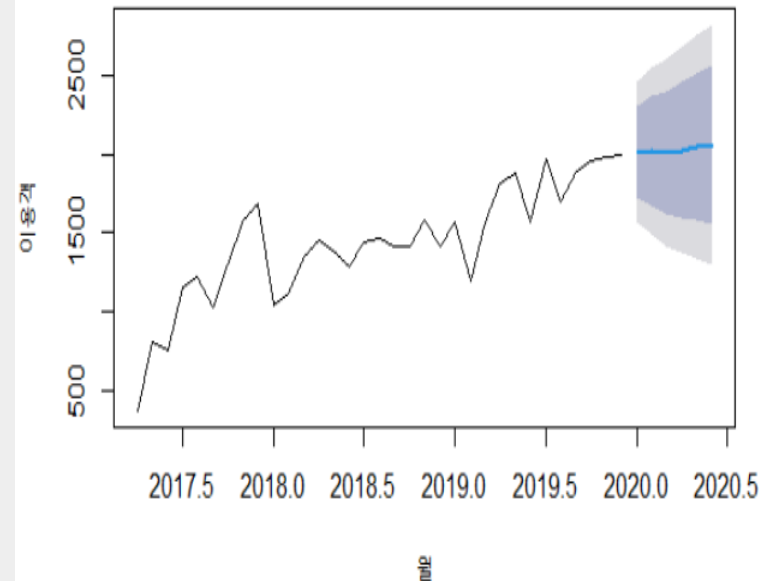
*ARIMA(0,1,1)*

표준오차를 통한 모델의 유의성 검정

$$|\hat{\phi}_1| = 0.5990 > 0.1996 \times 1.96 = 0.3912^*$$

$$|b| = 42.8017 > 16.4488 \times 1.96 = 32.2396^*$$

모심택시 월별 이용객 예측



## 02 데이터 분석

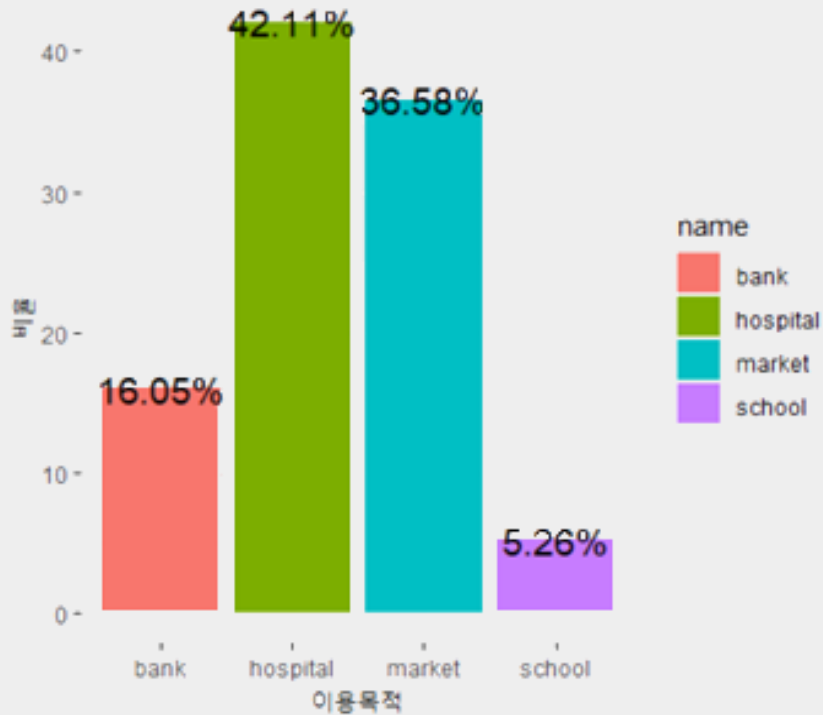
### - 상관 분석



#### 데이터 분석

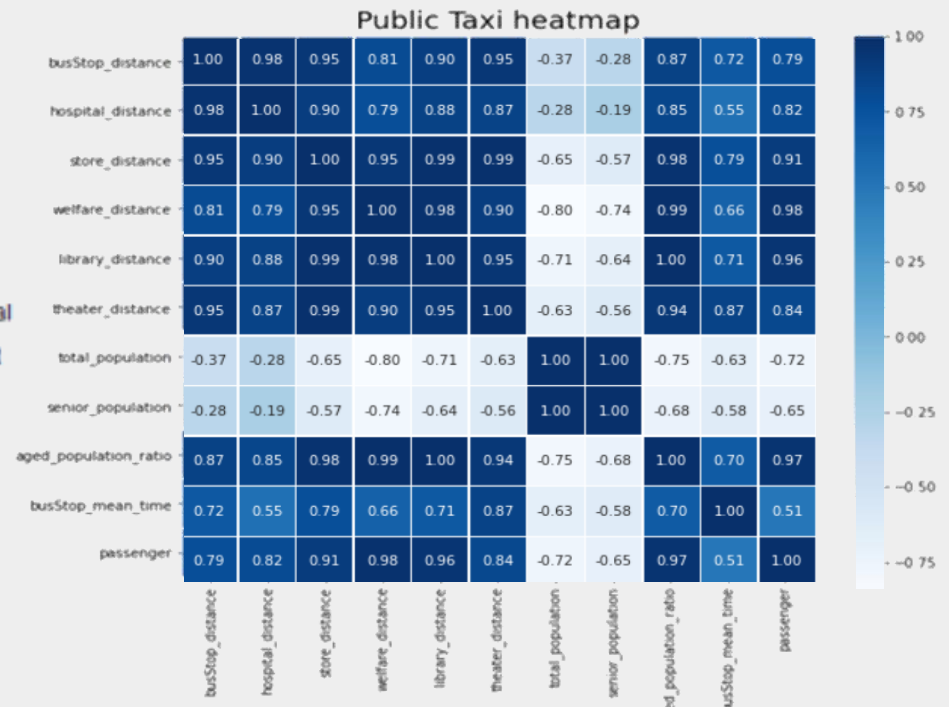
1단계 설문조사 분석 및 변수 간의 상관분석 실시

모심택시 이용목적 (%)



주 대상은 교통 취약 지역의 거주민이기 때문에 비교적 도심에 위치에 있는 병원, 은행·우체국, 시장, 학교, 직장으로 이동을 목적으로 이용

변수간의 상관분석



각 변수들 간의 관계를 파악하기 위해 상관분석 실시  
변수들 사이의 관계가 유의미함을 파악  
상관관계를 바탕으로 클러스터링에 필요한 가중치 선정

## 02 데이터 분석

### - 클러스터링



## 데이터 분석

### 2단계 클러스터링

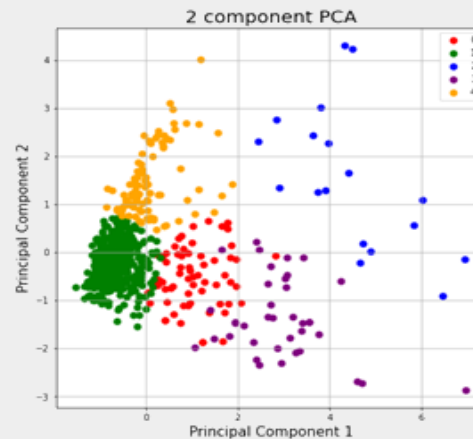
교통, 인프라, 노인인구 비율을 고려하여 군집화 실행

#### 변수 목록

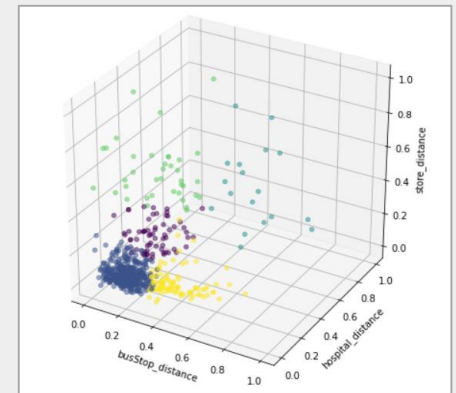
- ◇ 버스 정류장과 마을 회관의 거리
- ◇ 의료시설과 마을회관의 거리
- ◇ 상권과 마을회관의 거리
- ◇ 문화, 서비스 센터(1. 복지센터 2. 도서관 3. 영화관)와 마을회관의 거리
- ◇ 버스 배차 시간의 평균
- ◇ 노인인구 비율

#### K-Means Clustering

#### 2차원 클러스터링



#### 3차원 클러스터링



#### Clustering 시각화

“다양한 영향 변수를 고려한  
최적의 취약 지역을 추정 ”

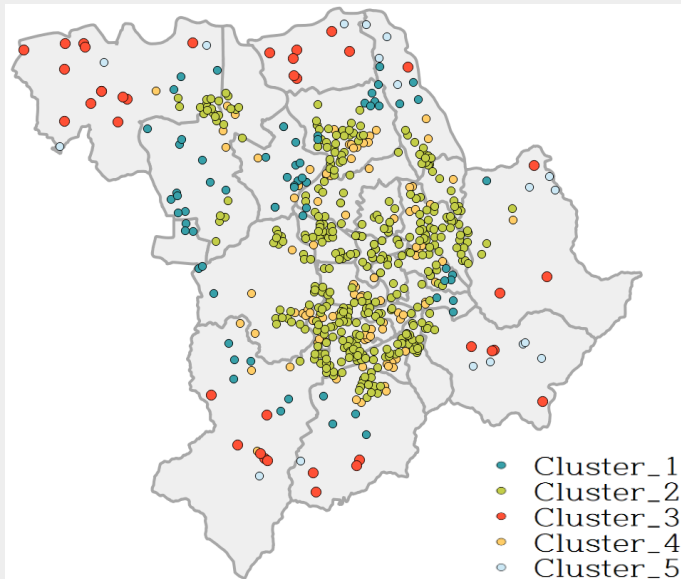
주어진 데이터의 중심으로 군집화 하는 알고리즘  
K=5 설정 후 변수를 기반으로 지역을 최종 ‘5개’로 군집화



데이터 분석

2단계 클러스터링

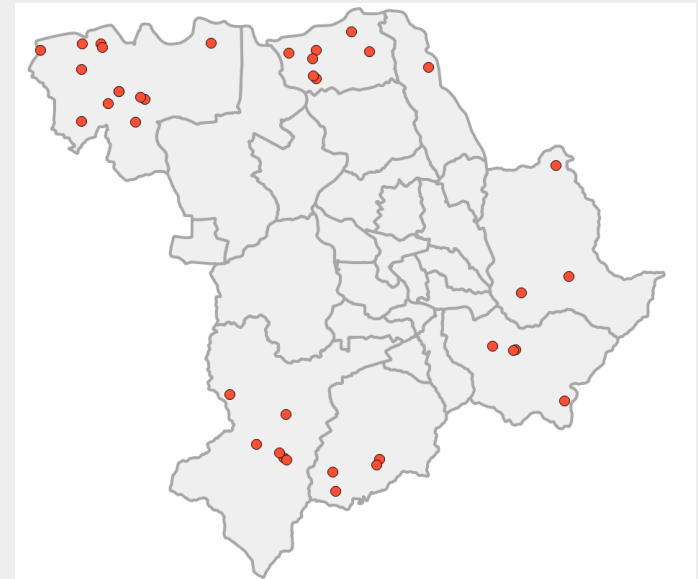
K- means 결과



“가중치를 이용해 나눈 취약 지역 클러스터 순위”



최종 취약 지역



“선정된 클러스터에서 상위 10개 지역 추출”

“다양한 영향 변수를 고려한 최적의 취약 지역을 추정”



## 02 데이터 분석

### - 예측 모델



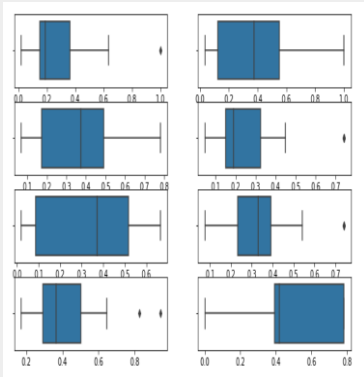
## 데이터 분석

## 3단계 최적지역 예측 모델

### 예측을 위한 데이터 전처리

- ◇ “MinMax Scaler 활용”
- ◇ 데이터를 0~1 사이 값으로 표준화

### 이상치 제거



변수별 이상치 확인 및 제거

### 상관계수 확인



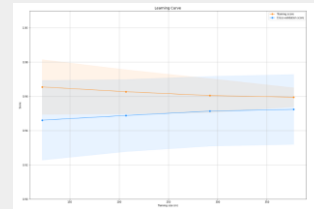
변수의 상관계수를 확인하고  
유의미한 변수 추출

### 인공지능 모델 학습

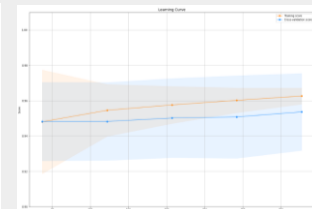
#### Learning Curve

어떤 특정한 대상을 학습하는 데 시간대비 학습 성취도 확인

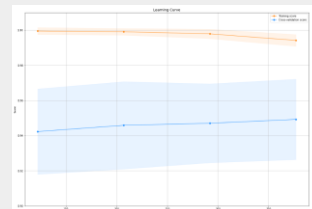
#### Logistic Regression



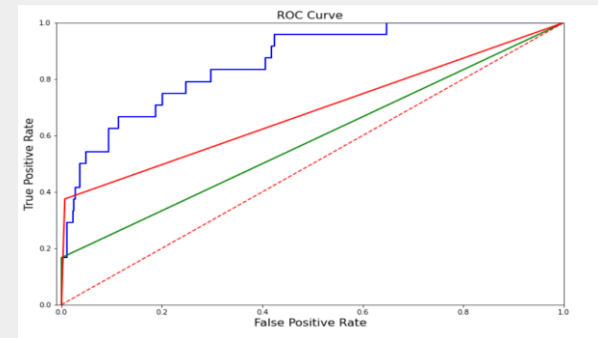
#### Random Forest



#### Gradient Boosting



### ROC Curve : 분류 모델의 성능 확인



#### Learning Curve, ROC 확인 결과

“Logistic Regression” 모델이 가장 성능이 좋음 확인

선정한 변수를 바탕으로 최적 지역을 예측하는 인공지능 모델을 구축

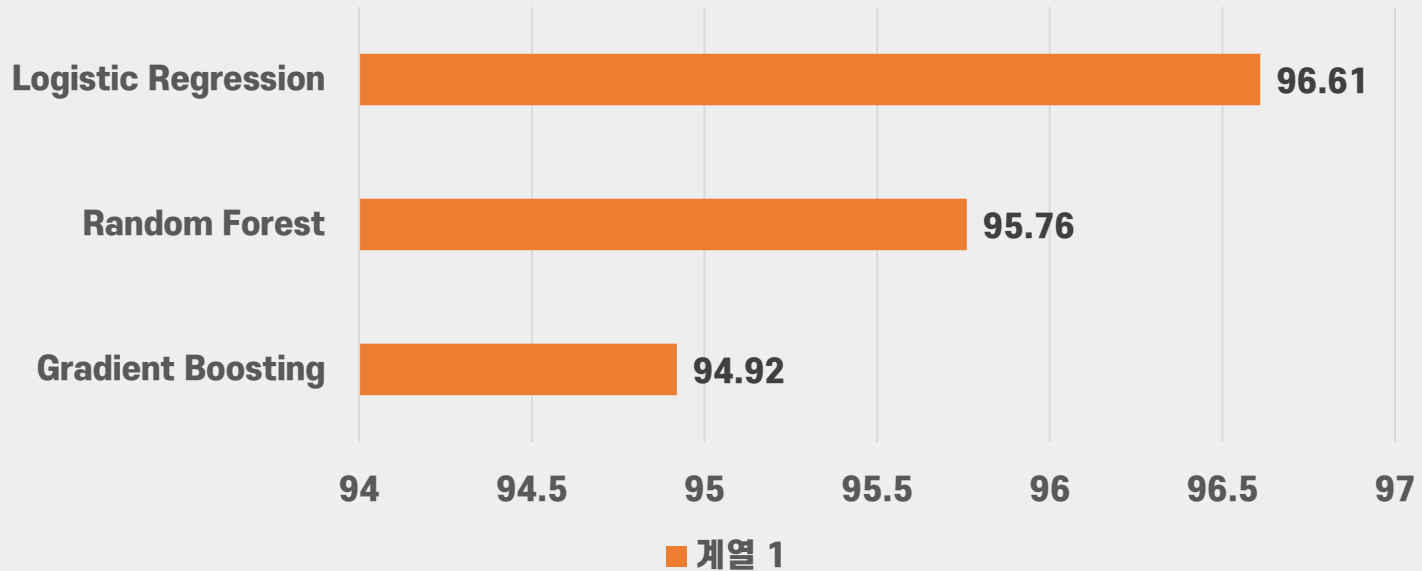


데이터 분석

3단계 최적지역 예측 모델

“정확도 96%”

## 정확도



정확도가 가장 높은 “Logistic Regression” 모델 선정  
예측 모델을 활용해 취약 지역을 예측 가능

## 02 데이터 분석

### - 회귀 분석



데이터 분석

회귀 분석

마을 별 인구 대비 연간 모신택시 탑승 인구에 대한 회귀분석 결과, 인구가 많은 마을에서 모신택시 이용 빈도가 높은 것으로 보임

회귀모델은  $p=0.03824$ 로 유의수준 0.05에서 유의하다고 판단되며 설명력은 **25.6%**로 추정.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

독립변수: 마을별 인구

종속변수: 마을별 연간 모신택시 탑승객 수

```
> summary(HT1)
```

```
Call:
lm(formula = passenger ~ population, data = HT)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-817.1  -284.3  -111.0   406.4   948.2
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  701.772    266.388   2.634  0.0188 *
population     3.603      1.586   2.272  0.0382 *
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 514.9 on 15 degrees of freedom
Multiple R-squared:  0.256,    Adjusted R-squared:  0.2064
F-statistic: 5.162 on 1 and 15 DF,  p-value: 0.03824
```

## 02 데이터 분석

### - 회귀 분석



## 데이터 분석

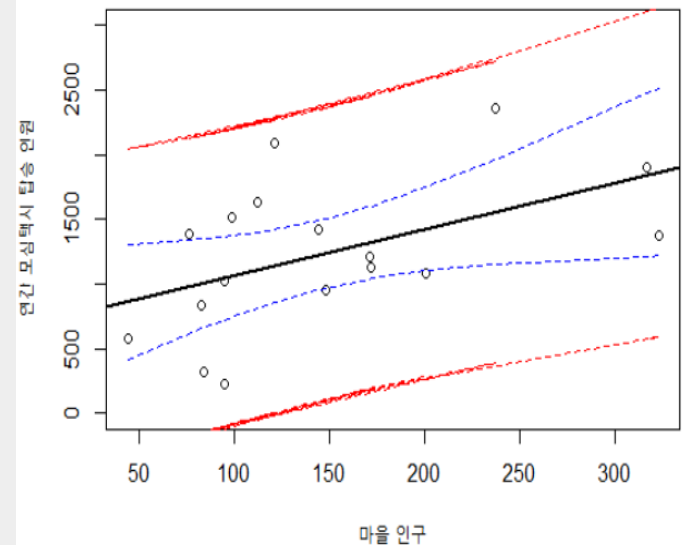
## 회귀 분석

형성된 회귀모델을 바탕으로 신뢰 · 예측구간의 그래프를 그린 후, 추가 배차마을의 인구를 고려하여 연간 탑승인원을 예측해 보았을 때 약 8,100여 명이 추가로 모신택시를 이용 예상

```
> pr <- predict(HT1,newdata=new,interval="confidence")
> pr
```

	fit	lwr	upr
1	1335.977	1053.8539	1618.101
2	1548.580	1152.7753	1944.385
3	1692.718	1188.5662	2196.869
4	1036.892	711.5077	1362.276
5	1285.529	1015.3691	1555.689
6	1278.322	1009.2344	1547.410

$$\hat{y} \pm 1.96\hat{\sigma}_e \sqrt{1 + \frac{1}{T} + \frac{(x - \bar{x})^2}{(T-1)s_x^2}},$$



# 03

## 결론 및 활용 방안

1. 모심택시 최종 배차 지도
2. 개선 방안
3. 활용 방안

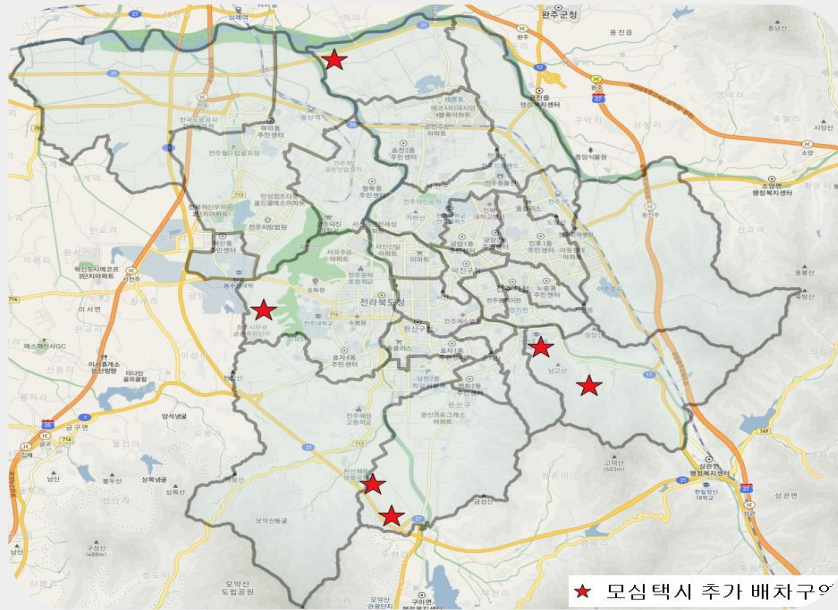
## 04 결론 및 활용 방안

### - 분석 결론

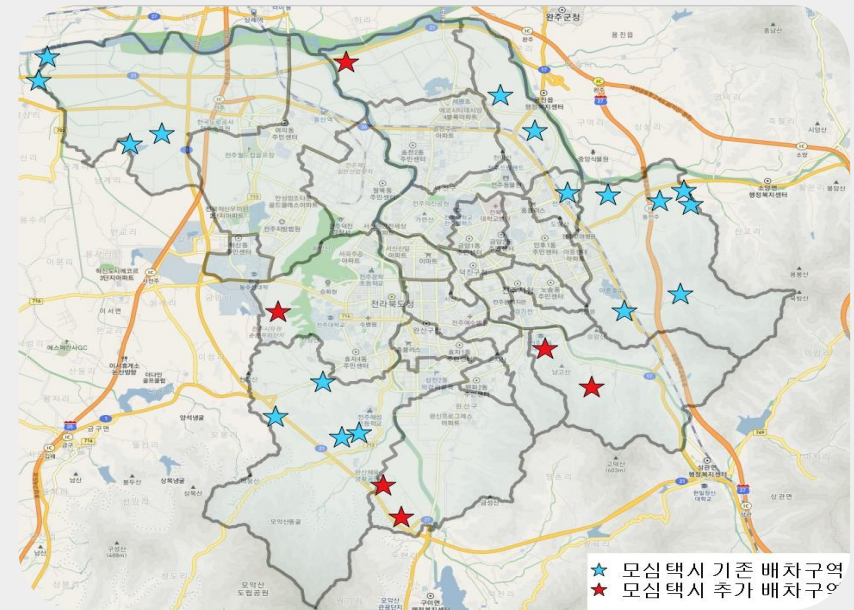
#### 결론



### 모심 택시 추가 배차 지도



### 모심 택시 최종 배차 지도



## 04 결론 및 활용 방안

### - 분석 결론

#### 개선 방안

#### 지자체 관계자 측면

노인인구 비율이 높은 지역임에도 불구하고 모심택시 운영을 하지 않아 교통서비스 이용에 불편함을 느끼는 교통취약계층을 고려한 모심택시 운영이 필요

#### 개선 방안

#### 사업시행 실무자 측면

향후 모심택시 추가 배치 고려시 최적입지 선정 결과를 참고 하여 확대 가능

#### 정책 활용 측면

공공시설과 마을회관거리 분석을 통해 더 가까운 시설을 쉽게 이용할 수 있도록 모심택시 노선 개선

모심택시의 사각지대를 분석하여 더 많은 소외지역 주민들의 생활편의 제공

## 04 결론 및 활용 방안

### - 분석 결론

#### 활용 방안

#### 교통 서비스 측면

모심택시의 기존의 운영방식보다는 개선된 운영방식으로, 기존의 승강장과 마을회관같은 거리를 줄이고, 배차시간 공공시설 및 병원등과의 거리를 측정한 최적화 된 입지에 승강장을 배치 및 운영

#### 활용 방안

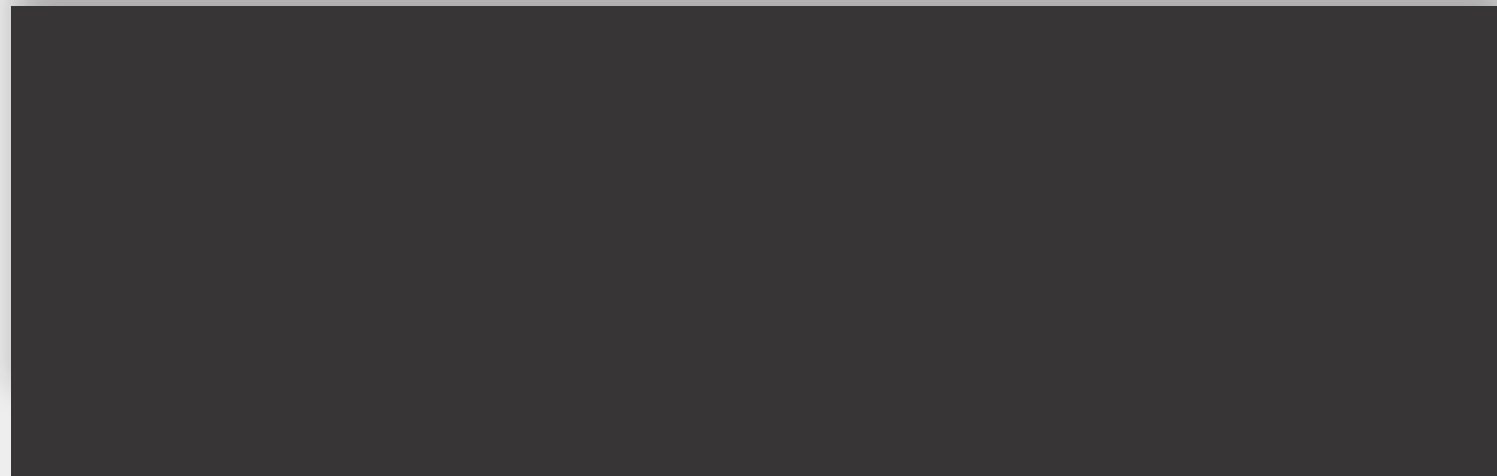
#### 도시 재생 측면

기존 버스 노선 개편에 활용  
개편된 공공형 택시의 효율성 있는 운행으로 수요 증가  
지역활성화 및 도시뉴딜 정책에 활용 및 기여

#### 정책 활용 측면

동일한 환경에서 문제를 겪고 있는 타 지역의  
공공형 택시 운영에 있어서 표준모델로 활용





••

발표를 들어주셔서

감사합니다 :) )