실시간 수요 예측을 통한 공공자전거 제도 효율성 강화

Hackathon in ADW IoT Data Alchemist Hackathon

강원대 1팀 U-BIKE 타슈

INDEX



01 주제 선정 배경02 주제 및 분석 목표 설정

Analysis

01 분석 개요

02 데이터 전처리

03 1차 분석/ 2차 분석 /3차 분석

Conclusion

01 결론

02 활용 방안

03 제언

주제 선정 배경

전국의 지방자치단체에서 최근 화제가 되고 있는 공유경제와 스마트 모빌리티의 특징을 갖고 있는 공공자전거 제도를 시행을 통해 시민들의 높은 만족도와 교통 기본권을 확대하고 있다.



공공자전거 누구나(사용자), 언제든(24시간), 어디서든(출발지), 어디로든(목적지) 자전거를 이용할 수 있는 시스템

공공자전거의 현황

- A. 현재 전 세계 약 600여개의 도시에서 운영 중
- B. 단거리 자동차 이용 수요를 대체 전망



▶ 2016 기준

공공자전거의 효과

- A. 교통 측면
 - a. 접근성 향상
 - b. 교통 기본권 확대
- B. 사회적 측면
 - a. 자전거 보관 시설 비용 절감
 - b. 공간 자원의 효율적 활용
- C. 환경적 측면
 - a. 도시 이미지 및 미관 개선
 - b. 환경 오염 예방

부산광역시는 2015년 해운대 신도시 일대 지역에 공공자전거를 운행하였으나, 당시 부산광역시의 상황으로 인해 제대로 시행되지 못하였으나, 부산시 전 지역으로의 확장으로써 시민의 높은 만족도를 얻을 수 있을 것이다.



부산시 공공자전거 'U Bike'에 대하여

시행 당시 'U Bike'의 현황

- ❖ 부산시 면적에서 극히 일부 지역(해운대 신도시)에만 공공자전거가 설치됨
- ❖ 공공자전거 스테이션의 위치가 대부분 해안가에 자리 잡아 시민들의 접근성이 떨어짐



▶ 부산 지역 내 공공자전거가 설치된 지역

부산광역시는 대한민국 제 2의 수도로써 공공자전거 제도 시행 시 **높은 만족도**를 얻을 수 있을 것 공공자전거의 이용에 있어 제기된 불편 사항을 알아 본다. 다양한 민원 접수 중 스테이션에 자전거가 부족하여 이용하지 못 하는 의견을 확인할 수 있다.



제기된 불편 사항 공공자전거 이용자가 스테이션에 도착했을 때 이용할 수 없는 문제



▶ 공공자전거의 시민의견 게시물

불편 사항을 해결할 수 있는 방안은?

앞서 제기한 불편 사항에 대한 관리자의 답변은 다음과 같다. 이에 대한 원인은 자전거 쏠림 현상에 대한 선재적 대응 부족이라고 할 수 있으며, 이를 해결하기 위한 방안을 제시한다.



제기된 불만에 대한 답변

현재 서울공공자전거 '따름이'의 분배를 위해 서울시 전역에 24개 분배팀(분배직원 100여명, 분배차량 36대)이 자전거의 회수 및 거치를 지속적으로 분배하고 있으나, 지역별 특성에 따른 자전거의 쏠림현상이 과도하게 나타나 대응에 어려움을 겪고 있는 실정입니다.

▶ 공공자전거의 시민의견 게시물의 관리자 답변



문제 해결 방안

자전거 쏠림 현상을 해결하기 위해 스테이션별 수요에 맞는 **적절한 자전거 이동**이 필요함 적절한 자전거 이동을 위해 고려해야 할 것은 많은 사람들이 이용할 수 있도록 하는 것이다. 따라서 각 스테이션의 수요에 맞게 자전거를 배치 및 이동 시키는 것이 중요하며 이를 추정하기 위하여 센싱 데이터를 이용하도록 한다.



스테이션별 적절한 자전거 이동이 필요

사람들이 많이 이용할 수 있는 곳에 설치하는 것이 가장 중요한 만큼 스테이션 설치하는 데 가장 중요한 것은 **유동인구**



하지만 유동인구 데이터의 특수성으로 인해 유동인구를 추정할 데이터가 필요



유동인구 추정을 위하여 IOT 센싱데이터 이용

주제 및 분석 목표 설정

주제와 관련하여 유동인구 및 통행량이 영향을 끼치는 대기 지표를 활용하고자 한다. 이전 연구에 의해 CO, NO₂, VOCs, PM10, SO₂에 유동인구 및 통행량은 큰 영향을 끼친다고 알 수 있다.



유동인구 및 통행량이 영향을 주는 대기 지표

[Ref. 1] 〈표 1〉 자동차로 인한 오염물질별 배출 기여도

구분	∞	NO ₂	VOCs	PM10 ^{a)}	SO ₂
전국 ^{b)}	78.9%	40.9%	16.5%	43.3%	1.3%
경 기 ⁽⁾	90,3%	68,1%	18,8%	76.4%	10.1%
서울	90.1%	60.6%	31.9%	73.3%	11.0%

주: a) PM10의 경우 도로 이동 차량에 의한 날림먼지 양은 제외

b) 2003년 기준, 환경부(2006), 환경백서, p. 429-430.

c) 2003년 기준, 경기도(2006), 2006년도 경기도환경백서, p. 239.

[Ref. 2]

〈표 4〉NO₂ 측정값과 도시특성요소간의 상관분석

변수	상관계수		
유동인구수	Р	0.482	***
*** 유의수준 1%			

[Ref. 3]

되는 경유는 황 함량이 430ppm이었으며, 2005년 기준으로 전체 자동차 중 경유 차량이 1/3 가량 차지하는 구조가 도시 내 차량통행이 빈번한 도로를 중심으로 이산화황의 농도를 집중시키는 요인으로 작용하였다고 볼 수 있다. 2003년 서울시 자료를 이용한 연구에서는 VKT와 이산화황의 집중이 1%의 유의수준에서 양의 관계를 보이고 있다(Kim and Guldmann, 2011). 2004년 10월부터 황 함량이 30ppm

- ❖ [Ref.1] 연구결과를 활용: CO, NO₂, VOCs, PM10은 자동차 통행량과 상관성이 있음.
- ❖ [Ref.2] 연구결과를 활용: NO2와 유동인구가 양의 상관관계를 보임
- ❖ [Ref.3] 연구결과를 활용: 도시 내 차량 통행으로 인해 SO₂의 농도가 집중되며 양의 상관관계를 보임
- ▶ 김영국(2017.6), "GIS 응용 교통 및 토지이용자료를 기반으로 한 도시 대기오염 분석", 교통연구24(2), 67-81
- ▶ 홍재선,김학열,이승주(2007,9),"도시특성요소가 서울시 대기 중 이산화질소 오염에 미치는 영향", 서울도시연구 제8권 제3호, 117-130

주제 및 분석 목표 설정

실제 공공자전거의 운영은 자전거의 스테이션별 수요 예측을 하고 있다고 한다. 하지만 수요 예측이 이용자들에게 제공되는 정보로는 체계적이라고 말할 수 없다. 따라서 설명 가능한 수요 예측 방법으로써 이용자들의 불만을 해소할 필요가 있다.



수요 예측 방법

◆ 사용 데이터



SENSING DATA



AMOUT OF USE

- ◆ 데이터 이용 방법
 - A. 센싱 데이터 이용
 - ❖ 센싱 데이터의 **패턴** 확인
 - ❖ 공간 패턴 분석을 통한 실제 자전거 이용량과의 상관성
 - B. 자전거 이용량 데이터 이용
 - ❖ 자전거 이용량의 **시계열** 분석으로써 기존의 패턴 확인

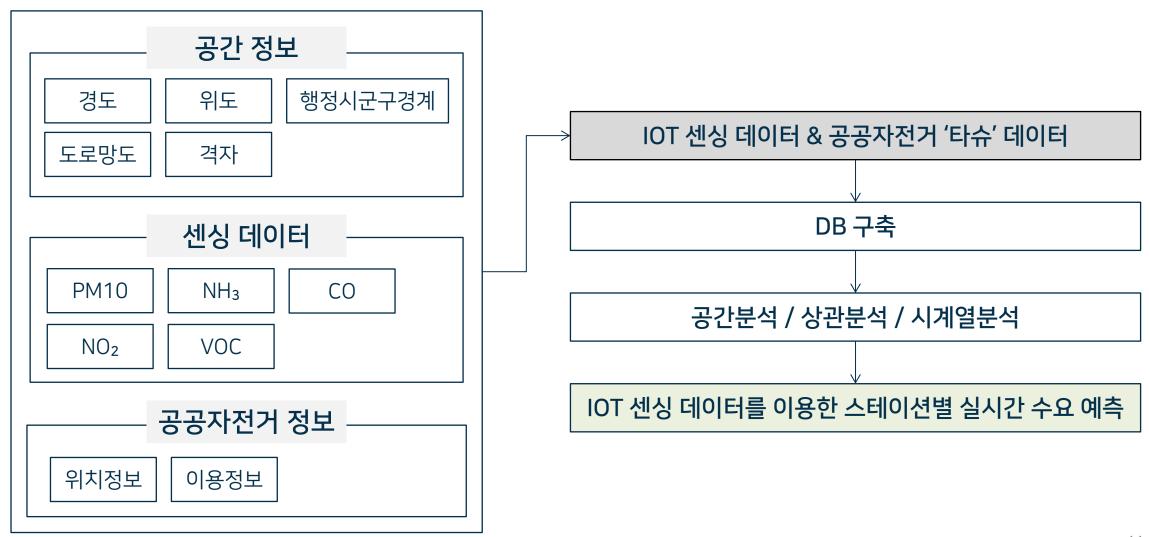


스테이션별 실시간 수요 예측이 가능





분석 프로세스





데이터 전처리 과정 1

대전광역시 공공자전거 '타슈'

- A. 대전, 대구 실시간 데이터의 일자별 필요 변수들만 추출
 - a. 연/월/일/시/분/초 추출
- B. '타슈'의 스테이션별 총 이용량과 일자별(시간별) 이용량 데이터 생성 -> 실시간 수요예측을 위하여 파생
 - a. 각 스테이션의 일별, 시별 총 이용량 데이터 생성
- C. '타슈' 자전거 스테이션의 데이터 조인
 - a. 타슈 자전거 현황 분석을 위해 위도, 경도를 추가하여 지도에 매핑
- D. 시계열 데이터 생성을 위해 '타슈' 데이터와 센싱 데이터와의 조인 및 정제
 - a. 센싱 데이터의 주말 부재로 인해 '타슈' 데이터를 주중으로만 구성
 - b. 주기를 맞추기 위해 이용량이 없는 시간에 대하여 이용량을 0으로 처리하여 모든 시간대를 생성
 - c. 05시~23시의 시간으로만 구성하며 이상치 00시 제거

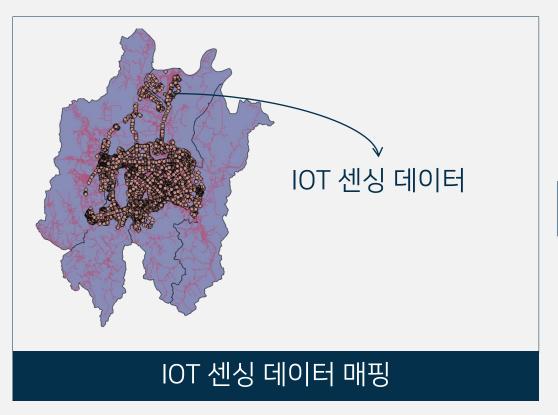


데이터 전처리 과정 2

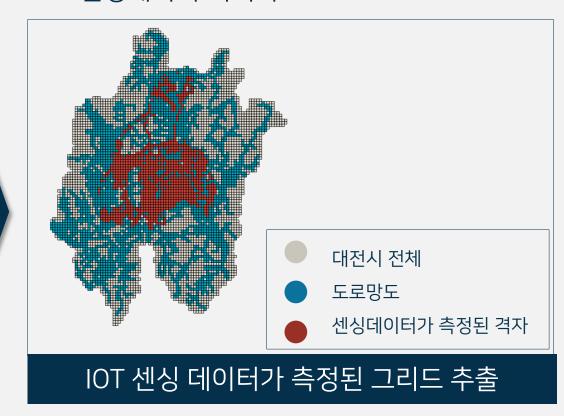
IOT 센싱 데이터

A. 공간 분석

a. 센싱데이터 처리 전



b. 센싱데이터 처리 후

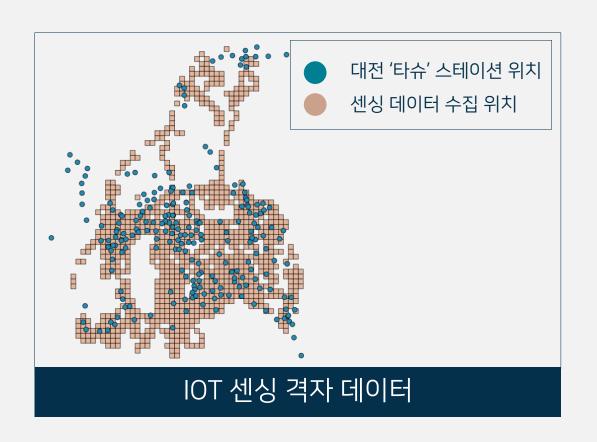


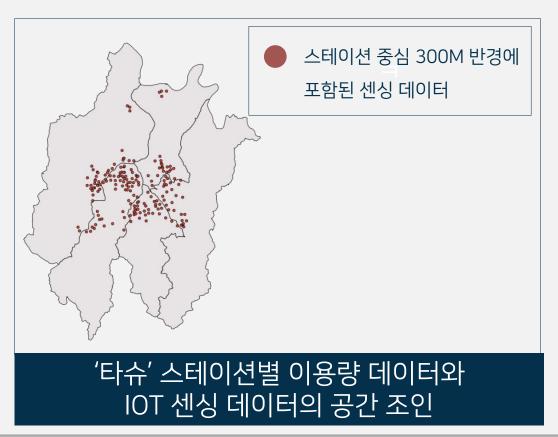


데이터 전처리 과정 2

IOT 센싱 데이터

B. 타슈와 센싱 데이터 공간데이터 조인





스테이션별 실시간 수요 예측을 위한 분석은 대전광역시의 IOT 센싱 데이터, 공공자전거 '타슈'의 이용량을 이용하며 분석의 단계는 다음과 같다.



▶ 대전 광역시

1차 분석: IOT 센싱데이터 핫스팟

가설: 시간/ 지역구별로 센싱 데이터 수치의 차이가 있다.

2차 분석: IOT 센싱데이터와 이용량의 상관분석

가설: 센싱 데이터와 이용량 간에 상관성이 있다.

3차 분석 : 시계열 분석

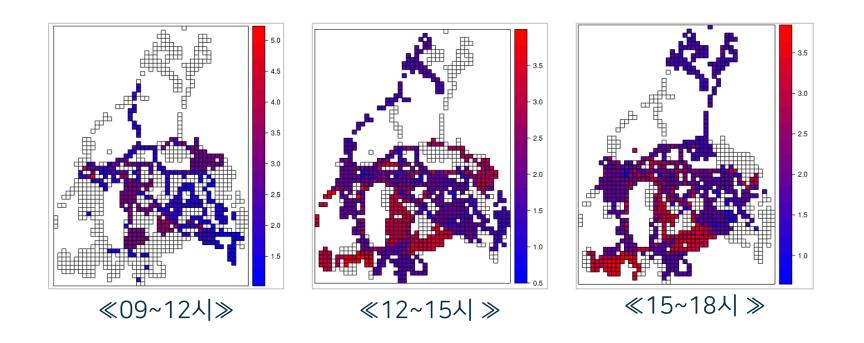
가설: 스테이션별 이용량에 시간대별 특정 패턴이 있다.



1차 분석 : IOT 센싱데이터 핫스팟

※ 대전광역시

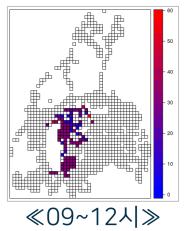
CO

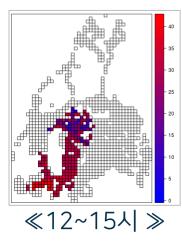


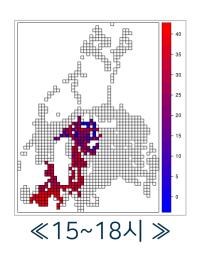
시간 지역구별로 센싱 데이터의 패턴 차이가 있음

% VOCs

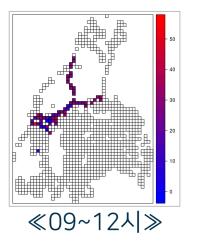
서구

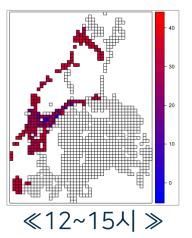


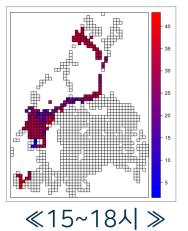




유성구





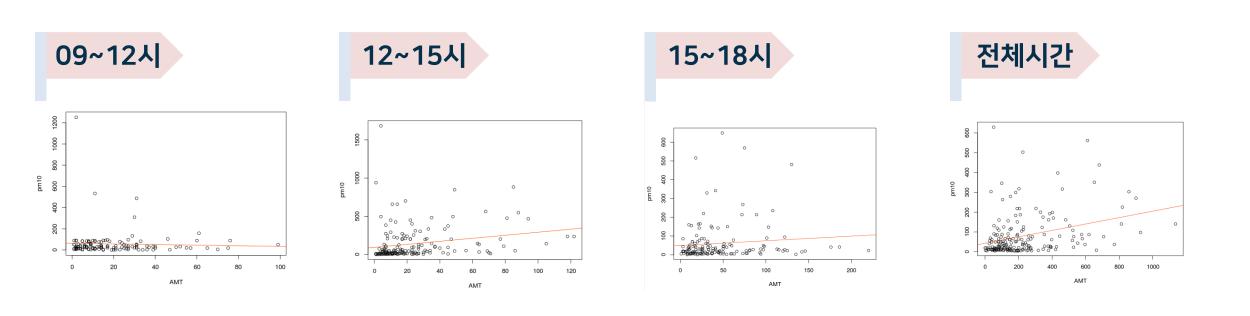


지역구에 따라 센싱 데이터의 패턴 차이가 있음



2차 분석 : IOT 센싱 데이터와 이용량의 상관분석

※ PM10



센싱데이터와 이용량의 패턴을 통해 이용량을 예측할 때 활용할 수 있음

3차 분석



3차 분석: 시계열 분석 '타슈' 데이터의 주기성을 확보 후, 시계열 분석 진행

시계열 데이터 구성

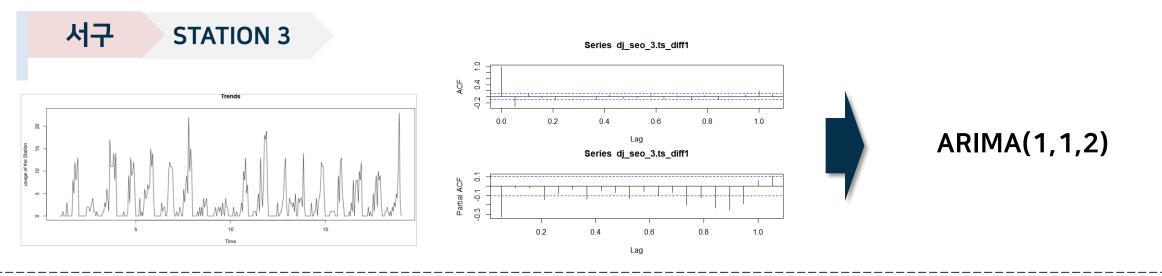
- A. '타슈' 이용량의 시간대를 05~23시로 고정하여 **주기 19**인 시계열 데이터 생성
- B. A 과정에서 서구와 유성구를 추출
- C. B 과정 후 해당 구 안에서 이용량이 가장 활발했던 서구의 스테이션 3, 유성구의 스테이션 31을 분석대상으로 삼음

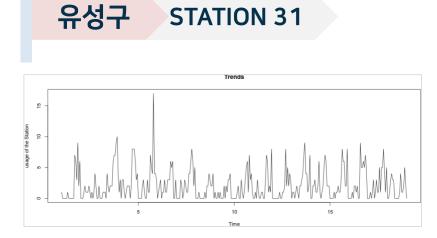
➤ 시계열 분석 R CODE(서구 스테이션 3)

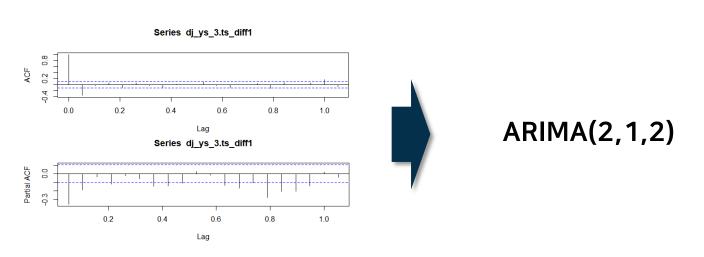
```
dj_seo_3 <- read.csv(file.choose(), header=T) # 데이터 입력 acf(dj_seo_3.ts_diff1, lag.max=20) # MA 2 pacf(dj_seo_3.ts_diff1, lag.max=20) # AR 1 library(TTR) # ARIMA(1,1,2) 적용 dj_seo_3.ts <- ts(dj_seo_3[6], st=1, end=19, fr=19) # 시계열 데이터 정의 dj_seo_3.ts_arima <- arima(dj_seo_3.ts, order=c(1,1,2)) # ARIMA 모형 적용 dj_seo_3.ts_arima dj_seo_3.ts_arima dj_seo_3.ts_arima dj_seo_3.ts_arima dj_seo_3.ts_arima dj_seo_3.ts_arima dj_seo_3.ts_arima dj_seo_3.ts_arima dj_seo_3.ts_arima,h=5) # 예측 dj_seo_3.ts_diff1 <- diff(dj_seo_3.ts_diff1) # 차분 1만 해도 어느 정도 stationary
```



3차 분석: 시계열 분석 '타슈' 데이터의 주기성을 확보 후, 시계열 분석 진행



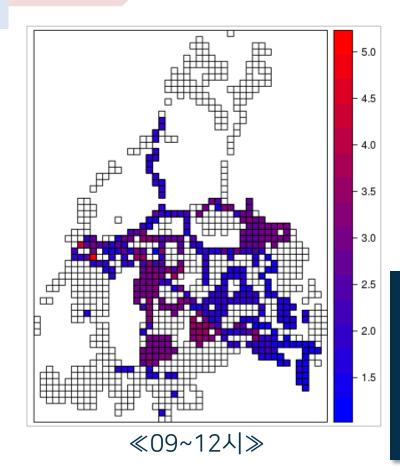






결론

CO

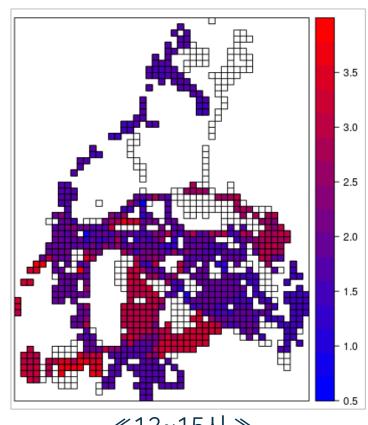






- ❖ 안골네거리 지역은 서대전역 부근의 기차역과 도마2동의 학교, 회사 밀집 지역으로의 이동에 의한 Hot spot
- ❖ Cold spot 지역이 Hot Spot 지역보다 CO 농도가 낮은 이유는 유동 인구 밀도가 비교적 낮고 구역을 넘어가는 크고 작은 도로가 많아 교통량 분산이 잘된다고 추정할 수 있음

CO



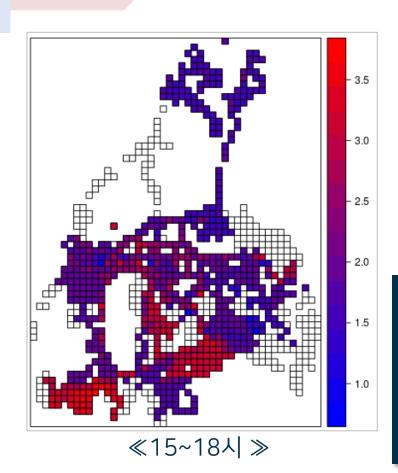




HOT SPOT

❖ 대전 남부지역에 3 구역의 hotSpot을 확인할 수 있는데, 건양대병원 부근은 외부지역에서 대전 중부지역으로 진입과 유입의 지역이며, 구농도원네거리, 문화초교네거리 지역의 부근은 연결되는 hotspot을 띄어 두 구역간의 교통 량 및 유동인구의 순환이 이루어진다고 추정할 수 있음

CO





부근의 hotspot이 다른 구역의 hotspot보다 농도가 옅음

❖ 12~15시와 비슷한 분포의 통행량으로 추정되지만 차이점은 구농도원네거리

HOT SPOT

❖ 그 이유는 화살표 방향의 다른 지역으로 통행량이 분산된다고 보여지는데 실제로 분산되는 부분의 CO의 농도가 12~15시와 비교했을 때 더 높음



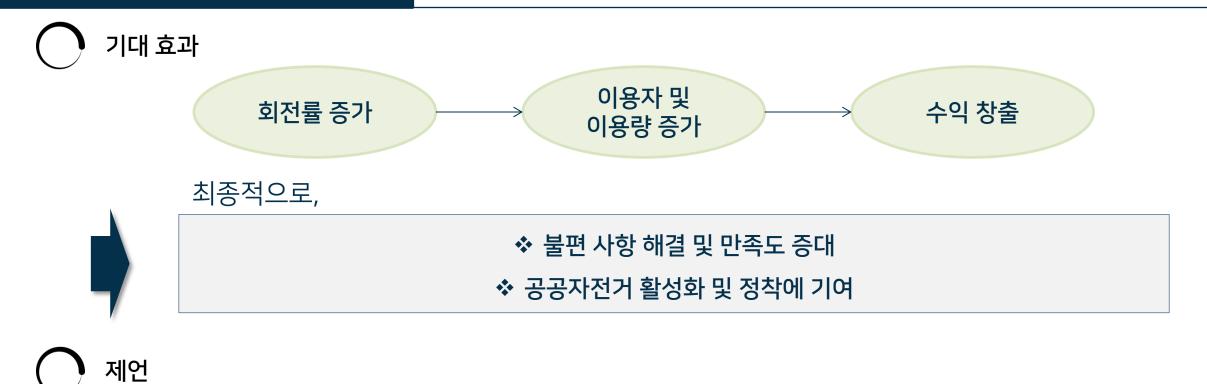
분석 결과

IOT 센싱 데이터 - 공간 분석

실시간 센싱 데이터는 유동인구 및 통행량 추정에 이용할 수 있음 스테이션의 이용량과 양의 상관관계를 가짐 따라서 실시간 유동인구 이슈에 활용이 가능

'타슈' 이용 이력 – 시계열 분석

'타슈' 이용 이력은 일정한 주기의 패턴을 갖고 있음 따라서 고정적 수요 예측이 가능





감사합니다.