

R 프로그래밍을 통한 세미 프로젝트

# 긱은 날씨에는 누가 백화점에 올까?

H A P P Y   B E A N

2019.12.20

# 목차

## 기상 조건에 따른 백화점 유인효과 분석

01

### INTRO

- 비즈니스 이해
- 프로젝트 정의

02

### METHODS

- 집계 구역
- 생활 인구
- 기상조건

03

### RESULTS

- 선형 회귀
- 의사결정나무

04

### DISCUSSION

- 결과 해석
- 추후 목표

# INTRO

비 즈 니 스 이 해  
프 로 젝 트 정 의

## // 비즈니스 이해 //

미세먼지 피해서 백화점·쇼핑몰로...방문객 '북적'

2019.03.06 15:29

: 미세먼지 공습을 피해 실내공간 방문객이 늘면서 유통업계가 특수를 누리고 있다. 롯데백화점의 경우, 지난 3월 미세먼지 비상저감 조치가 발령된 5일간 전국지점 총 58개 점포 매출이 지난해 같은 기간보다 9.1% 상승했다. 구매고객은 이보다 많은 18.8%의 상승률을 보여주었다.

## // 프로젝트 정의 //

1. 날씨가 좋지 않은 날, 백화점에 고객이 더 많이 올까?
2. 얼마나 더 / 덜 올까?
3. 각 매장별 차이는 없을까?

**기상 조건 별 방문객 예측 및 세그먼트  
백화점 날씨 마케팅 전략 제안**

# METHODS

집	계	구	역
생	활	인	구
기	상	조	건

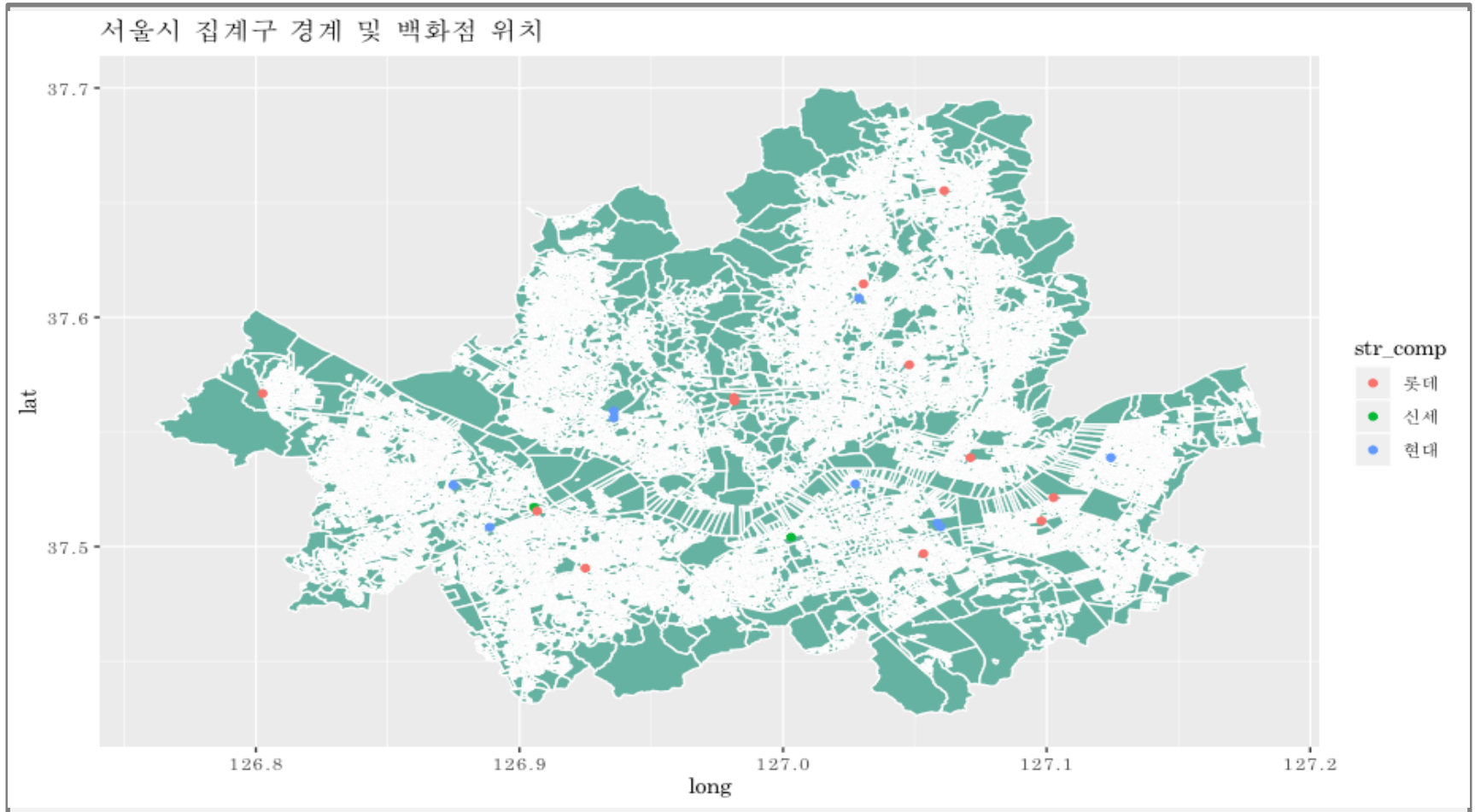
## // 집계 구역 //

통계청에서 제공하는 2016 집계 구역 데이터셋

상대적으로 작고 구체적으로 원하는 지역 선정 가능

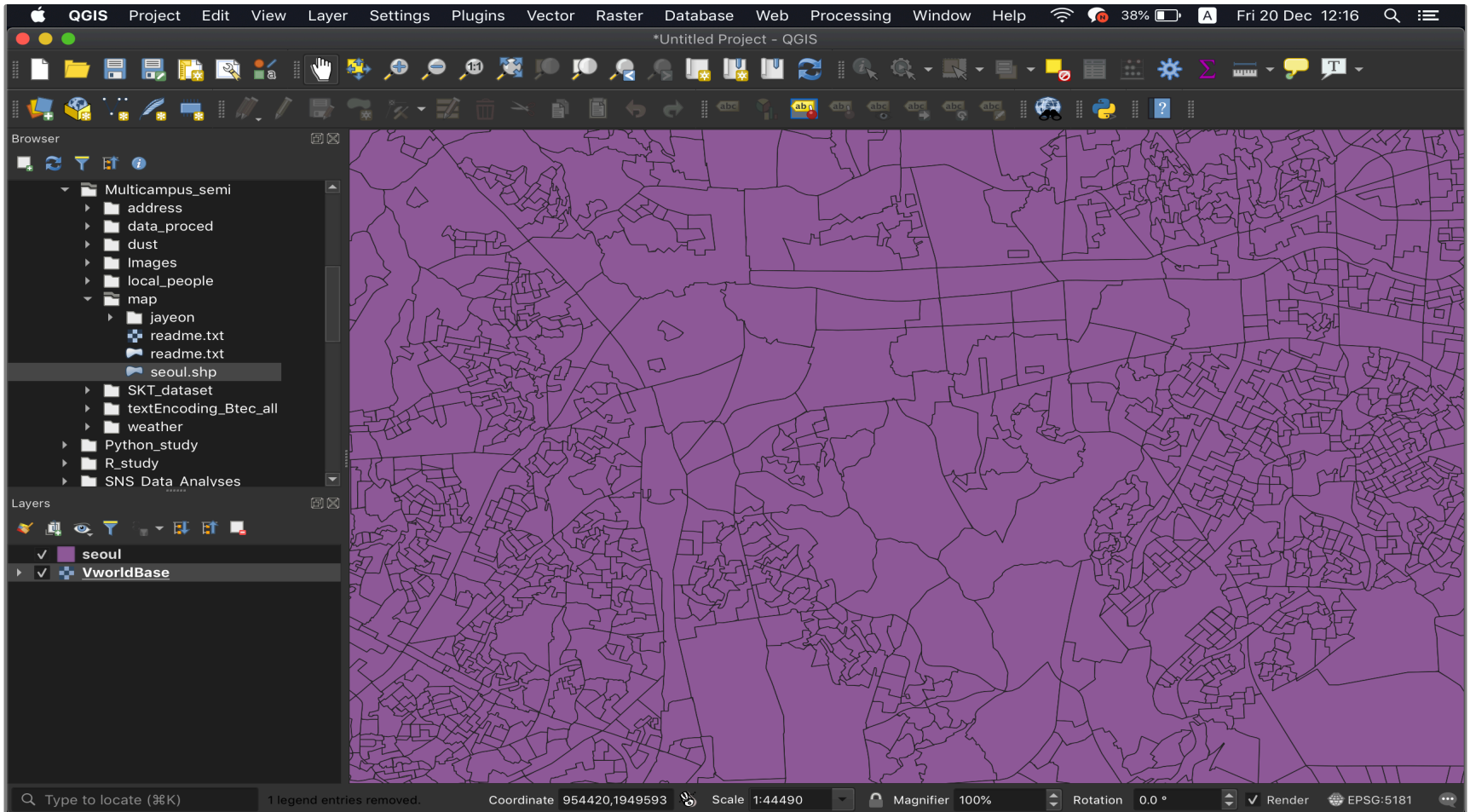
구: 25 -> 동: 424 -> 집계 구역: 19,153

## // 집계 구역 //





## // 집계 구역 //



## // 집계 구역 //

서울시 소재 백화점 3사 (롯데, 현대, 신세계) 총 25개소

- Google geocoding API, 백화점 주소지 위도, 경도 변환
- 유클리디안 거리 기준, 최근접 집계 구역 식별
- 이후 육안으로 사후 검정 실시

## // 집계 구역 //

서울시 소재 백화점 3사 (롯데, 현대, 신세계) 총 25개소

- 공항, 기차역, 버스터미널 등에 위치한 백화점 제외
- 동일 집계 구역 내 위치한 백화점은 하나로 계산

최종 16개 집계 구역 식별

// 

## 생활 인구

 //

KT LTE 통신 가입자 기반의 서울 생활 인구 '추정' 데이터

제공단위:

일(2017.01.01 ~ 2019.11.30) \* 시각(24) \* 집계구역(19,153)

제공속성

성(2) \* 연령(10세 미만, 10~79세 (5세 단위), 80세 이상)

## // 생활 인구 //

그림 2-3 전체인구로 확장을 위한 보정계수 적용 방법

KTMS보정계수  
적용

- KT의 휴대폰 시장점유율 역수  
(참고) MS(Market Share) : KT 시장점유율

LTE가입률 보정계수  
적용

- KT 내 휴대폰 가입자 중 LTE가입자 비중
- 행정동\*성\*연령 단위로 적용

휴대폰 On비율  
보정계수 적용

- KT 휴대폰 On비율  
(휴대폰OFF, 기기결함, 기지국 통신결함 등 포함)

전국 성\*연령 비교  
보정

- 1~3단계 보정을 적용한 전국 KT고객수와 전국 주민등록인구와 비교

//

## 날씨

//

우선적 관심 변수로 미세먼지와 강수를 선택

미세먼지

WHO 8단계 기준, 미세먼지 | 초미세 5단계 이상인 경우

강수

일 최대 시간당 강수량 5mm 이상인 경우

# RESULTS

선형회귀  
의사결정나무

## // 선형회귀 //

총 3가지 유형 식별

1. 미세먼지 영향으로 생활 인구 증가한 경우
2. 미세먼지 영향으로 생활 인구 감소한 경우
3. 미세먼지 영향이 없는 경우



## // 선형 회귀 //

### 1. 미세먼지 영향으로 생활 인구 증가한 경우

롯데 본점 (명동), 롯데 잠실점, 롯데 에비뉴엘점,  
현대 무역센터점, 현대 압구정 본점

## // 선형 회귀 //

## 1. 미세먼지 영향으로 생활 인구 증가한 경우

```
[1] 롯데.잠실점
16 Levels: 롯데.강남점 롯데.관악점 롯데.노원점 ... 현대.천호점

Call:
lm(formula = temp$mean_visit_pop ~ IsDustyDay + IsRainy, data = temp)

Residuals:
    Min       1Q   Median       3Q      Max
-20596.6  -3221.6   -38.5    2550.3   10241.8

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  20701.8      319.4    64.822 < 0.00000000000000002 ***
IsDustyDay1   1483.3       495.1     2.996   0.00299 **
IsRainy1     -799.2       919.6    -0.869   0.38563

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

Residual standard error: 3951 on 265 degrees of freedom
Multiple R-squared:  0.0365,    Adjusted R-squared:  0.02923
F-statistic: 5.019 on 2 and 265 DF,  p-value: 0.007251
```

## // 선형 회귀 //

### 2. 미세먼지 영향으로 생활 인구 감소한 경우

롯데 관악점

## // 선형 회귀 //

## 2. 미세먼지 영향으로 생활 인구 감소한 경우

```
[1] 롯데.관악점
16 Levels: 롯데.강남점 롯데.관악점 롯데.노원점 ... 현대.천호점

Call:
lm(formula = temp$mean_visit_pop ~ IsDustyDay + IsRainy, data = temp)

Residuals:
    Min       1Q   Median       3Q      Max
-971.0 -329.8 -196.7  142.8 2255.1

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1046.99     51.19   20.452 < 0.0000000000000002 ***
IsDustyDay1   -291.28     79.36   -3.670   0.000293 ***
IsRainy1       -13.49    147.42   -0.091   0.927173
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 633.3 on 265 degrees of freedom
Multiple R-squared:  0.04842,    Adjusted R-squared:  0.04123
F-statistic: 6.742 on 2 and 265 DF,  p-value: 0.001394
```

## // 선형 회귀 //

연령대 별 효과?

-> 지점마다 특징적인 패턴 가능성

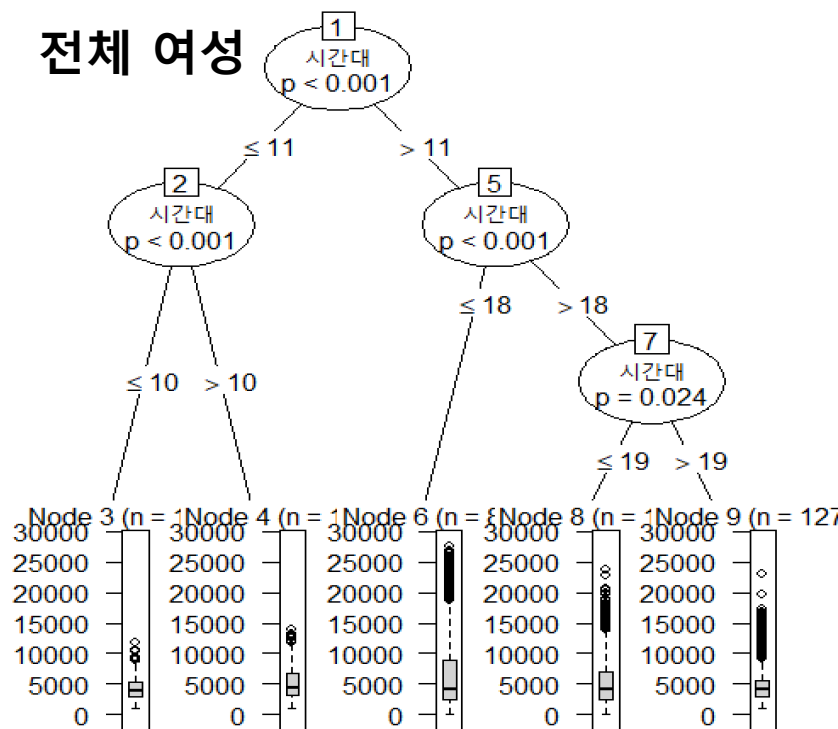
Ex)목동점

남/여 40, 45 미세먼지의 영향으로 방문객 증가

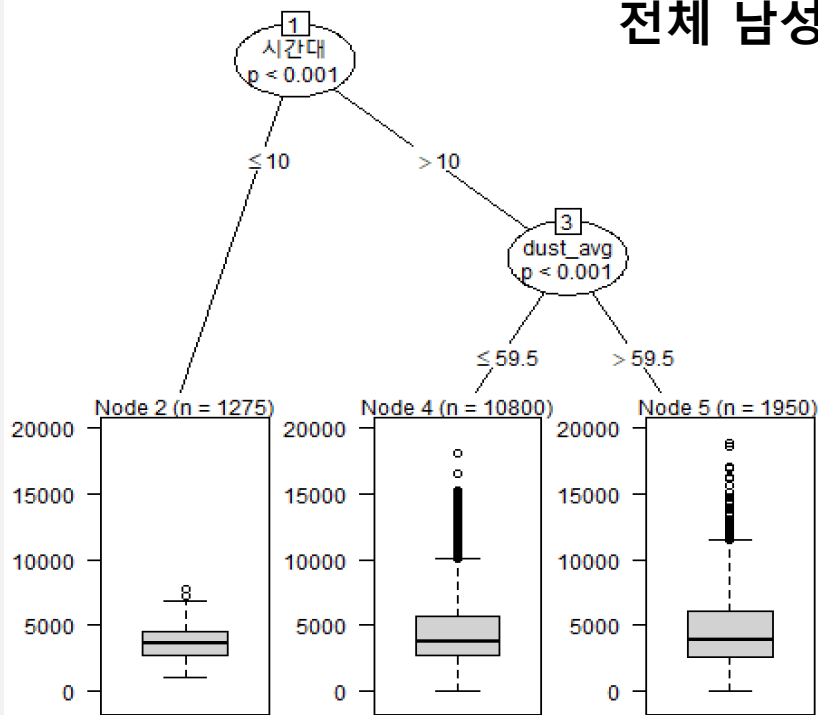
## // 의사결정 나무 //

## 의사결정 나무

전체 여성



전체 남성



# DISCUSSION

진행계획

## // 논의 //

지점별 특성을 고려한 모델 수립 가능성 제시

-> 주변 상권의 특징, 가족 구성의 형태, 영/유아의 수 등

전체 인구 수에서의 차이는 없으나 세부 연령별 특징 패턴 확인

-> 세부 연령별 기상 조건 영향 분석 가능

생활 인구를 유동 인구로 가정할 수 있을까?



## // 논의 //

생활 인구를 유동 인구로 가정할 수 있을까?

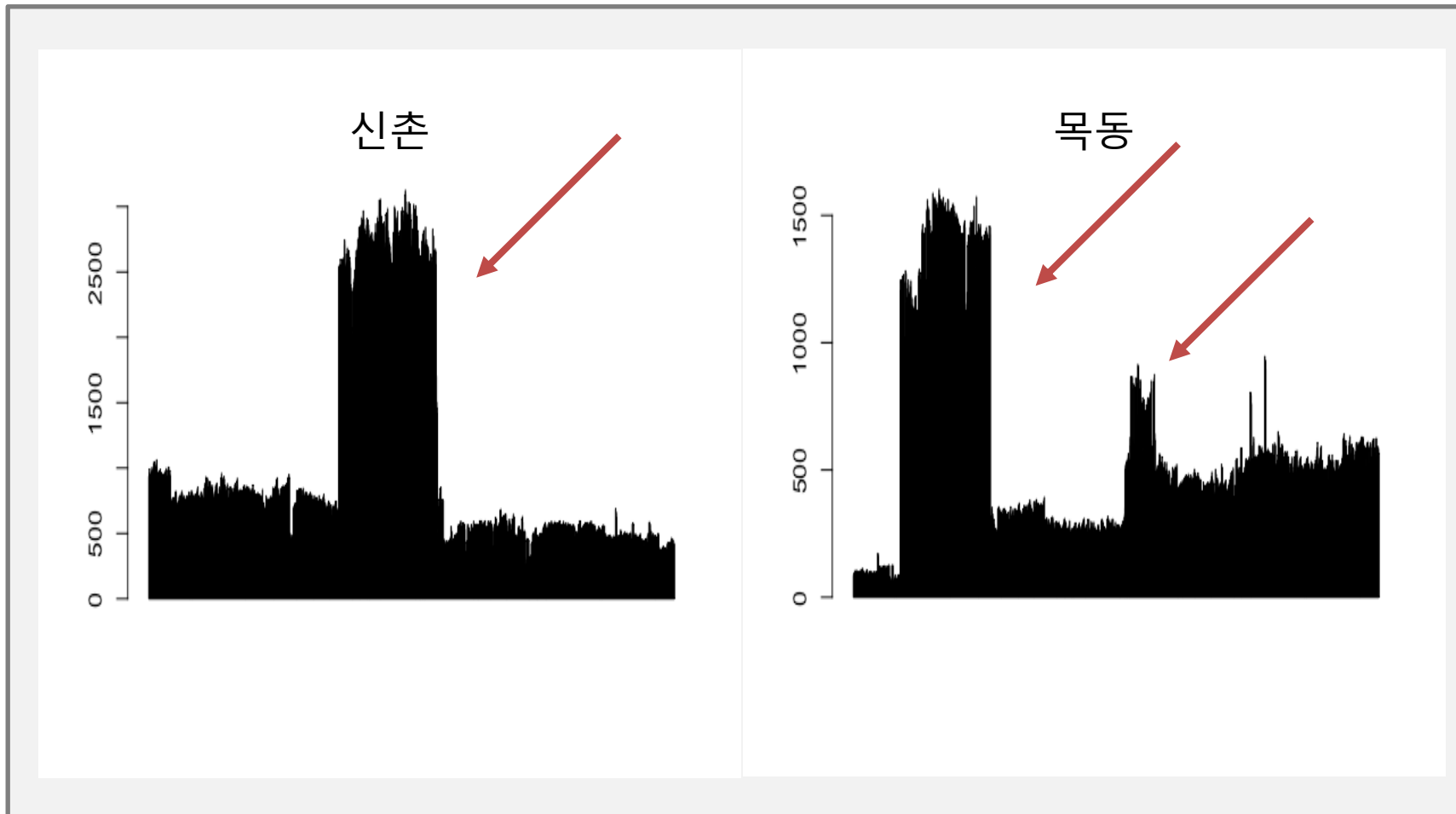
거주인구

화, 수, 목 / 02~05시 집계 구역 1년의 생활 인구 평균

유동인구

토, 일 / 10~20시 생활 인구 일 평균 - 당해 연도 거주인구

## // 논의 //



## // 추후 목표 //

SKT 데이터 허브에 공개된 구 단위 유동 인구 데이터와의 비교

-> KT 생활 인구 데이터에 대한 유효성 검증

임의로 설정한 기상 조건을 실제 값 혹은 다른 기준으로 변환하여 적용

의사결정 나무를 활용하여 고객 방문 예측 변수 파악

기상 조건 반응에 대해 유사 패턴 갖는 백화점 클러스터링

THANK  
YOU

H A P P Y   B E A N