



# 서울시 출퇴근 인구 분석을 통한 공유 우산 대여 시스템 구축

서울우산 토닥이 팀

양창석  
박현종  
유인호  
유준호

# CONTENTS



**01**

분석배경 및  
목적

**02**

분석 데이터  
및 방법

**03**

분석

**04**

최종결과

**05**

활용 방안  
및 문제점

**06**

참고 문헌  
및 데이터  
출처

# 갑자기 비가 온다면?

## 01 분석 배경 및 목적 Umbrella Sharing System





## 편의점, 폭우에 우산 판매량 '급증'...

입력 2017-07-26 15:12    진범용     

지난 6월부터 이어진 장마의 영향으로 편의점 우산이 불티나게 팔린 것으로 나타났다. 이는 올해 장마가 게릴라성 집중호우로 단시간 많은 양의 비를 동반해 미리 우산을 준비하지 못한 고객들이 편의점으로 몰렸기 때문이라는 분석이다.

26일 관련 업계에 따르면 이번 장마는 지역별로도 비가 내리는 양이 다르고 예고 없이 비가 쏟아지는 경우가 많았다. 이에 따라 편의점 우산 판매량이 두 자릿수 이상 신장했다.

실제로 이달 1일부터 23일 기준으로 세븐일레븐에서 판매하는 카카오프렌즈 우산은 전년대비 156.7% 판매량이 급증했다. 일회용 우산은 48.2% 신장했다. 카카오프렌즈 우산의 가격은 1만6000원으로 일회용 우산(4500원)보다 가격이 3배 이상 높다.

출처 : 뉴데일리경제(2017.07.26)



예고 없이 비가  
쏟아 지는 경우

일회용 우산 구매  
증가



## "앗, 비오네"...'썩는데 100년' 비닐우산 또 샀다

머니투데이 | 남형도 기자

VIEW 11,784 | 2017.08.01 06:25

#7월31일 오후 12시 광화문 일대. 아침에 우산을 미처 챙기지 못한 직장인 김모씨(31)가 한 편의 점으로 향했다. 그가 집어든 것은 5000원짜리 일회용 비닐우산. 김씨는 "돈이 아깝긴 하지만 오늘 비가 많이 온다고 해서 어쩔 수 없이 샀다"며 "비가 오락가락해 이달에 비닐우산만 3개를 샀다"고 말했다.

쓰고 버리는 일회용 우산 등 우산 쓰레기도 많다. 매년 2000만~3000만 개의 우산이 우리 손으로 들어온 뒤 다시 쓰레기통으로 들어가는 것으로 나타났다. 비가 오면 일회용 우산을 손쉽게 산 뒤 대중교통 등에 두고 잃어버리고, 조금만 망가져도 버리기 때문이다.

이로 인한 환경 오염 피해도 심각하다. 우산 비닐커버의 원료는 고밀도 폴리에틸렌(HDPE)이기 때문에 땅 속에 매립할 경우 썩는 데 100년이 걸리는 것으로 나타났다. 소각하면 다이옥신 등 유해 성분을 배출하고 온실가스 메탄 등도 배출하는 것으로 알려져 있다. 환경오염의 주범이 되는 것이다.

출처 : 머니투데이(2017.07.26)

<개인적 관점>

갑작스러운 비에 의한  
**불필요한 우산 구매**

<환경적 관점>

일회용 비닐 우산 썩는 데  
걸리는 시간 100년

**환경오염의 주범**

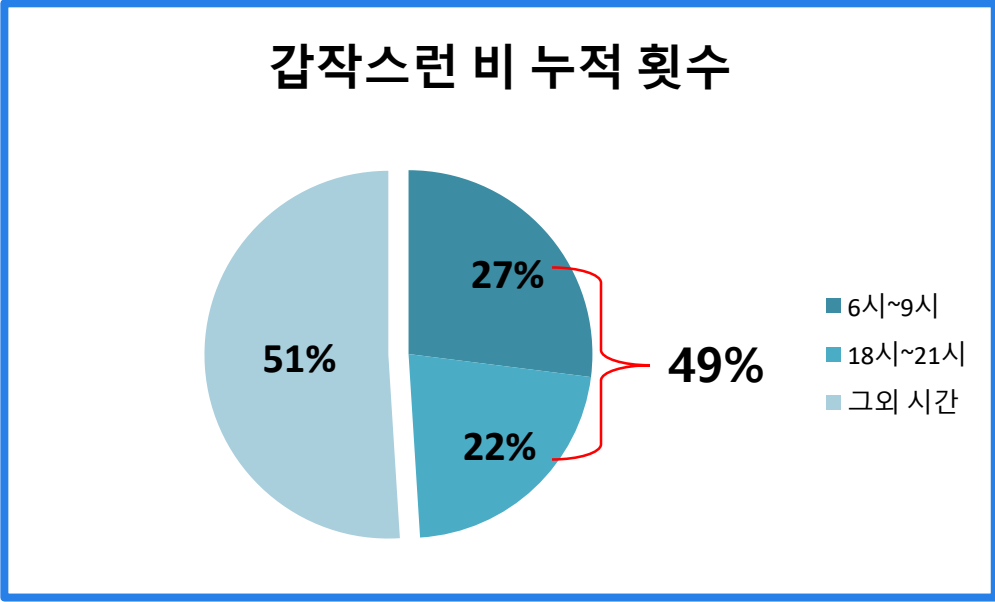
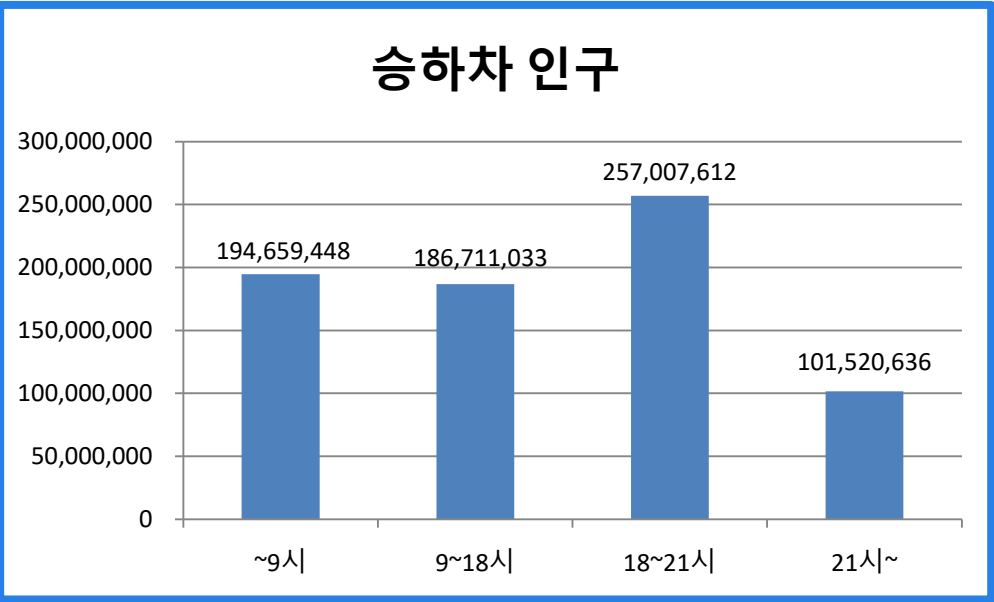
# 왜 출퇴근 시간 인가?



- 1. 출퇴근 시간 때 지하철 이용객 수가 **전체의 61%** 차지
- 2. 출퇴근 시간 때 갑작스런 비 횡수가 **전체의 49.3%** 차지
- > 출퇴근 직장 인구에게 새로운 시민 편의 제공
- 3 . 작년 **서울시**에서 시행한 **정책 방향성** 참고

서울 출퇴근 대중교통 내일 또 무료...미세먼지 저감 조치

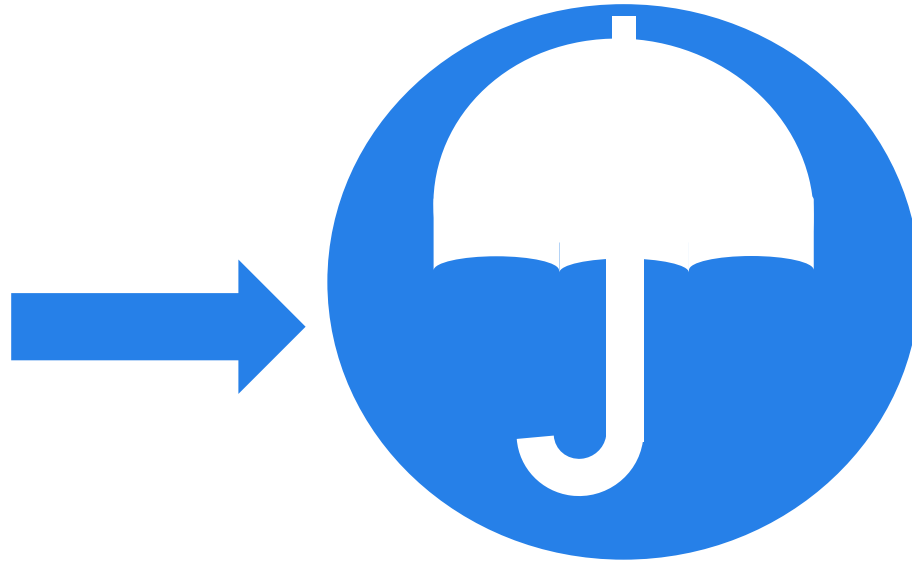
장세만 기자  
입력 : 2018.01.17 17:45 | 수정 : 2018.01.17 17:45







서울자전거  
SEOUL BIKE 따릉이



서울우산 토닥이

## 토닥토닥 (순우리말)

1. 빗물이 우산에 부딪힐 때 나는 소리
2. 출퇴근 시간 직장인의 지친 마음을 “토닥토닥” 위로해준다

원하는 곳에 대여하고 반납이 가능한 “**공유경제**”를 이용해 집에서 우산을 준비하지 못한 시민에게 **필요할 때 대여하고 반납이 가능**하도록 하자



## 서울시 날씨 정보 데이터



# 데이터 주요 정보(출처: 기상청)

1. 서울시 시간대별 기상관측 데이터  
(기온, 강수량, 습도, 적설 등)

## 서울시 지하철 이용통계 데이터



# 데이터 주요 정보(출처: 서울시 빅데이터캠퍼스)

1. 서울 지하철역 266개의 운행 일자 정보
2. 역별 시간대 승/하차 총 승객 수

## 서울시 거주인구/ 직장인구 데이터



# 데이터 주요 정보(출처: 서울시 빅데이터 캠퍼스)

### 1. 거주 인구

: 서울시 행정동 단위 거주 인구 자료

### 2. 직장 인구

: 서울시 사업체 정보 및 해당 종사 인구 자료



강수량에 따른 승차, 하차 인구 변화 패턴 분석

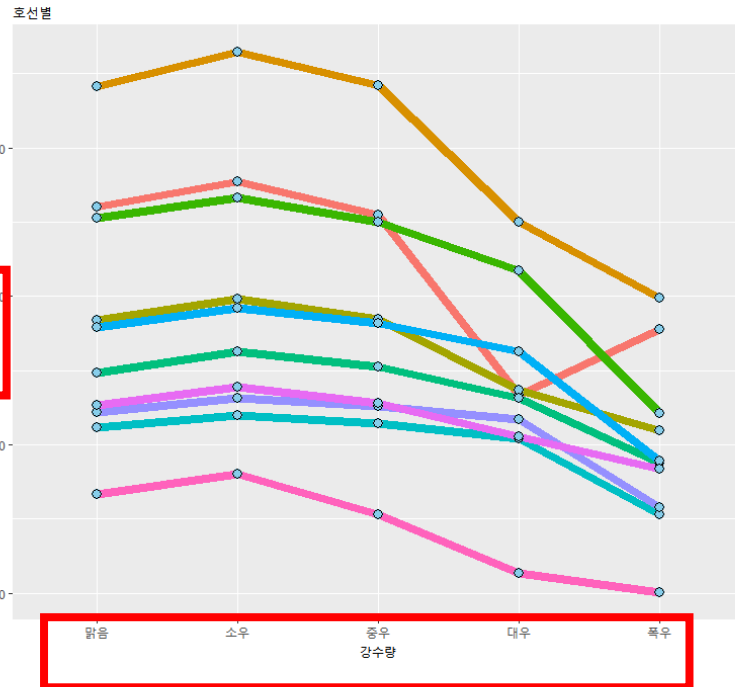


실제 거주/직장 인구와 상관 분석

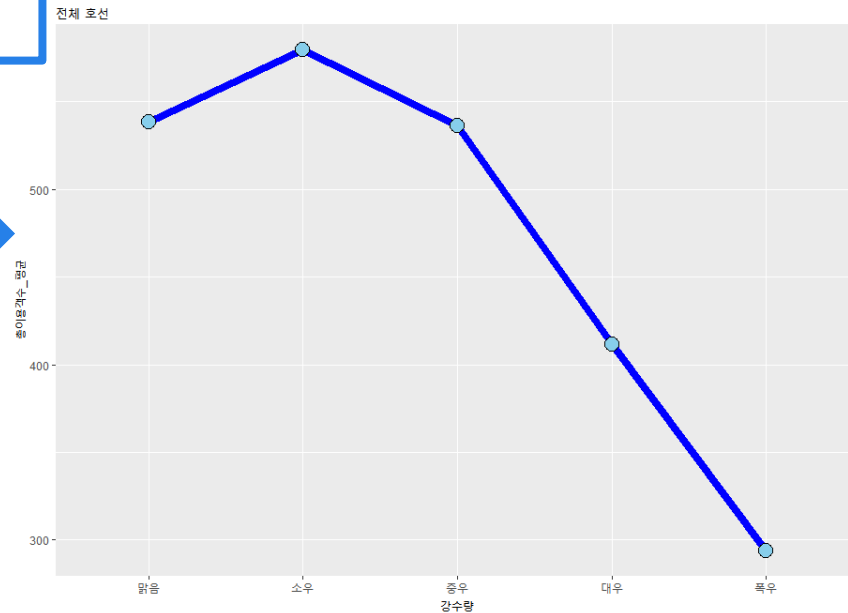




- 비의 세기에 따른 지하철 이용객 추이



전체호선으로  
확장



<총 이용객 수>

mean(승차 인원 + 하차 인원)

<강수량>

맑음(0mm)

소우(0.1mm~0.9mm)

중우(1mm~4mm),

대우(5mm~10.5mm)

폭우(12mm~20.5mm)

(Natural Break 로 범위 산정)

-> 데이터 값의 자연스러운 배열을

찾아 최적화 하여 묶는 법

소우(0.1mm~0.9mm)일 때 지하철 이용객 수가 가장 많다.

그 이후는 대부분 호선이 **비가 많이 올 수록 줄어드는 추세**를 보인다.

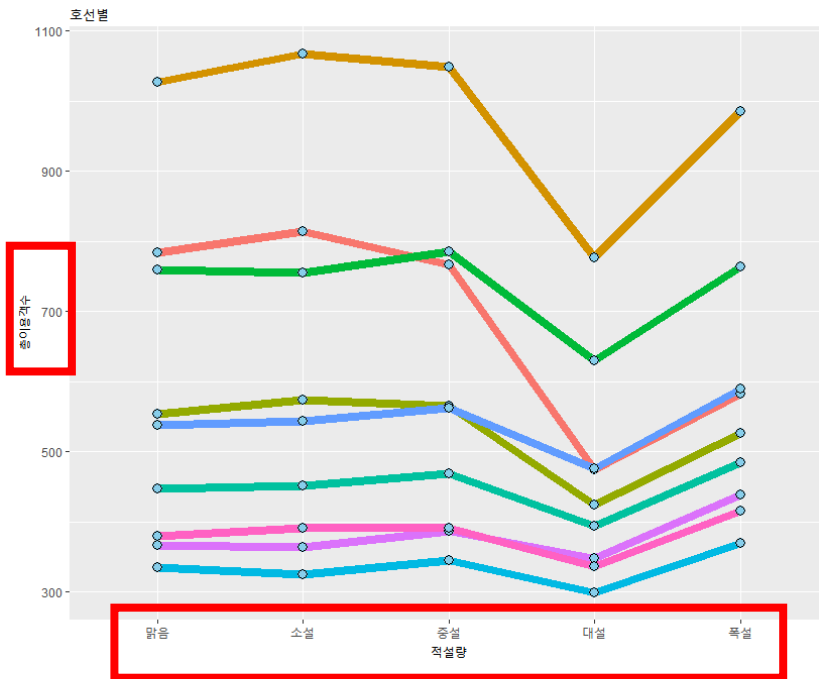
-> 폭우시 지하철 보다 다른 교통 수단 (자가용, 택시) 을 선호하는 것으로 보임

# 적설량에 따른 지하철 출퇴근 인구 변화

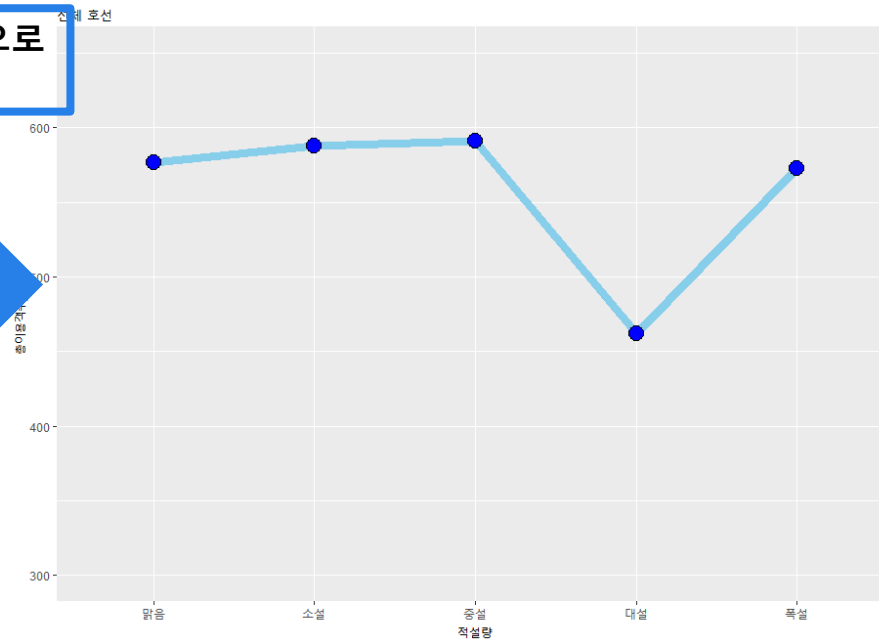
02 분석 데이터 및 방법  
Umbrella Sharing System



- 눈의 세기에 따른 지하철 이용객 추이



전체호선으로  
확장



<총 이용객 수>

mean(승차 인원 + 하차 인원)

<적설량>

맑음(0mm)

소설(0.1mm~0.9mm)

중설(1mm~2.4mm),

대설(2.5mm~4mm)

폭설(4.1mm~6.1mm)

(Natural Break 로 범위 산정)

-> 데이터 값의 자연스러운 배열을

찾아 최적화 하여 묶는 법

대설(2.5~4mm)일 경우 지하철 이용객 수 감소, 폭설(4.1~6.1mm)에서 증가 추세

-> 눈이 어중간하게 오면 다른 교통 수단 (자가용, 택시)를 선호하나

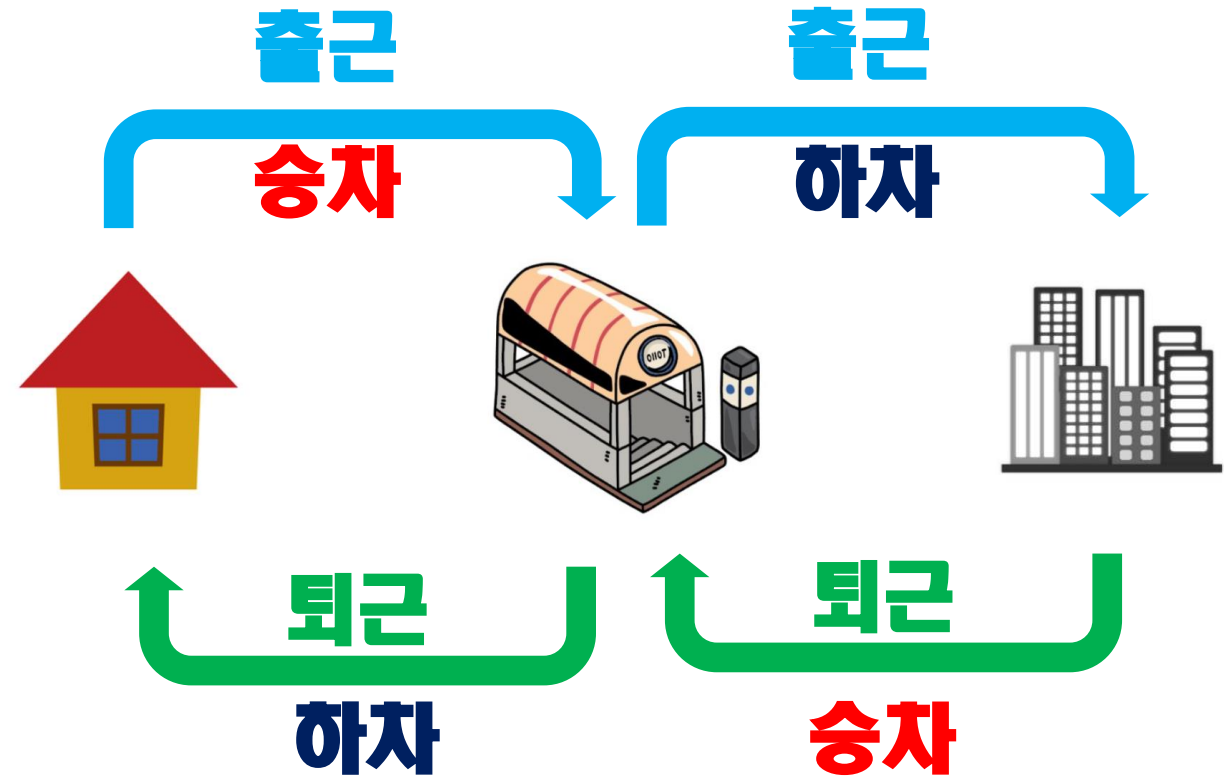
폭설일 경우 안전한 지하철을 선호하는 경향을 보임.

# 승하차 인구에 따른 집, 직장 예상

02 분석 데이터 및 방법  
Umbrella Sharing System



출근/퇴근 시간대	계산	예상
출근	승차 승객 수 - 하차 승객 수 > 0	집 거주인구 多
	승차 승객 수 - 하차 승객 수 < 0	직장 직장인구 多
퇴근	승차 승객 수 - 하차 승객 수 > 0	직장 직장인구 多
	승차 승객 수 - 하차 승객 수 < 0	집 거주인구 多

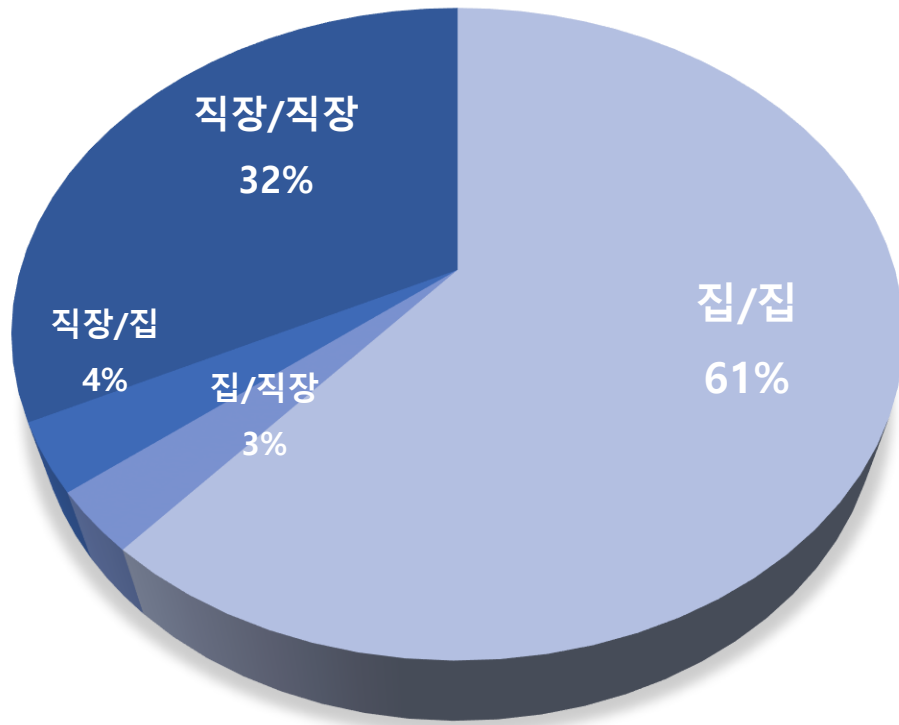


유출입인구 수식

(출근 승차 - 출근 하차) - (퇴근 승차 - 퇴근 하차)



예측



■ 집/집 ■ 집/직장 ■ 직장/집 ■ 직장/직장

(출근 시간 대 예측 / 퇴근 시간 대 예측)

집/집 : 61.28 %

직장/직장 : 32.33 %

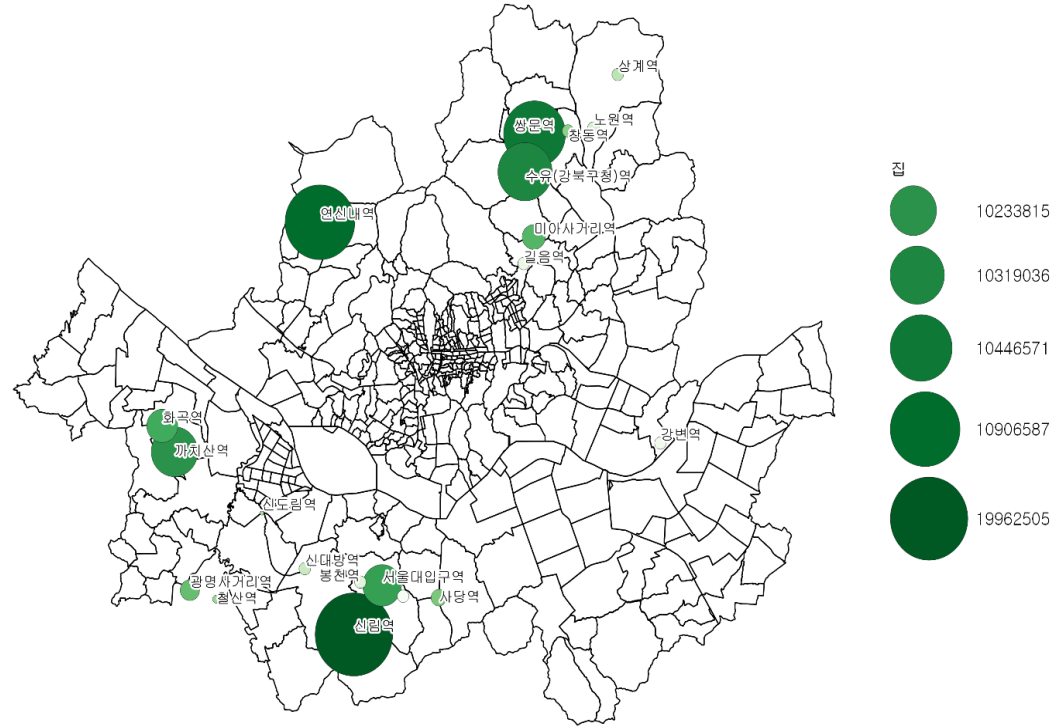
=> 집, 직장 밀집 지역으로 예측되는 수치가 전체의 **93.61%** 로 **역 마다 집과 직장을 구분**해 사용해도 될 것이라 판단됨

# 유출입 인구에 따른 집, 직장 예상

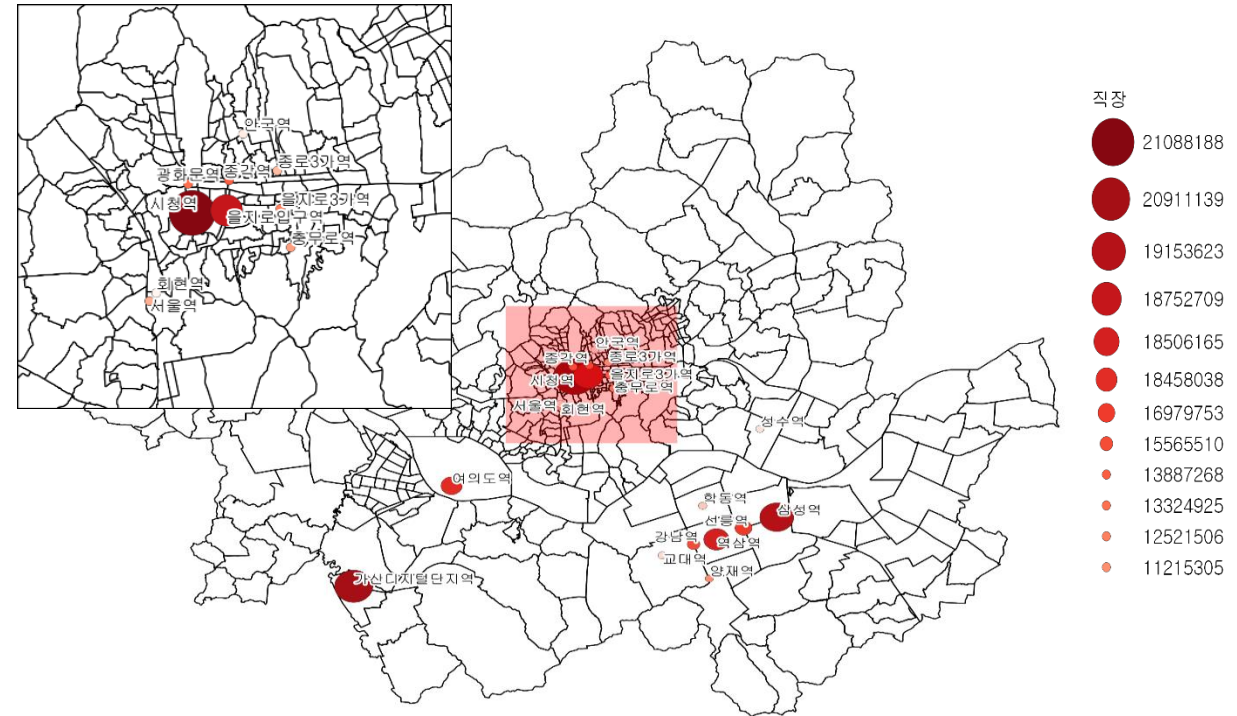
02 분석 데이터 및 방법  
Umbrella Sharing System



<거주 인구 밀집 지역>



<직장 인구 밀집 지역>



거주 밀집 지역 Top5		
순위	역	지하철 총 이용객 수
1	신림역	19,962,505
2	연신내역	10,906,587
3	쌍문역	10,446,571
4	수유역	10,319,036
5	까치산역	10,233,815

직장 밀집 지역 Top5		
순위	역	지하철 총 이용객 수
1	시청역	21,088,188
2	가산디지털단지역	20,911,139
3	삼성역	19,153,623
4	을지로입구역	18,752,709
5	역삼역	18,506,165



## 분석 방법

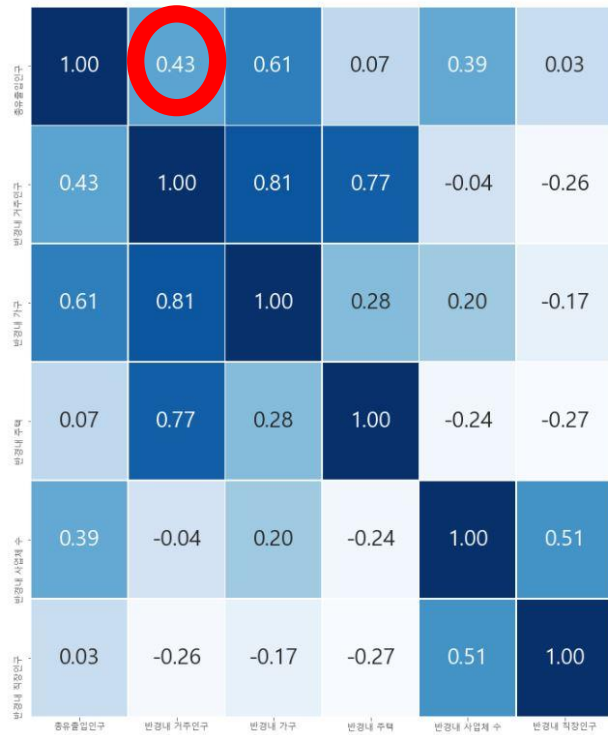
### 1. 지하철 역마다 반경 500m (역세권) 에 있는 거주인구, 직장인구 계산

거주, 직장 인구 데이터( 서울시 빅데이터 캠퍼스 제공) 의 위 경도 값을 계산, 이를 토대로 역 기준 클러스터링함

(생성변수 :지하철 역, 유출입 인구, 반경내 거주인구, 반경내 가구, 반경내 주택, 반경내 사업체 수, 반경내 직장인구)

### 2. Pearson 상관분석을 통해 변수간의 상관관계를 파악함 (Heatmap 시각화)

<거주지역>



<직장 지역>



### <Pearson 상관 분석 >

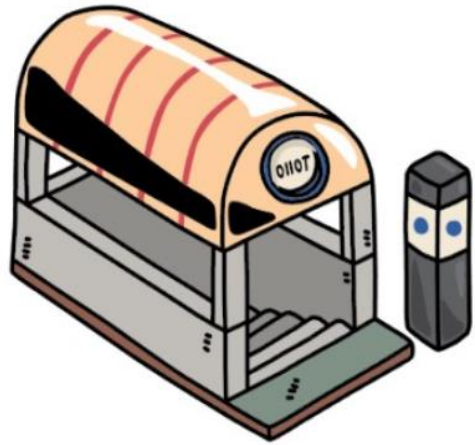
1. 거주가 많다고 추정된 지하철역의 유출입인구와 실제 거주인구와의 Pearson R값 : 0.43
2. 직장이 많다고 추정된 지하철역의 유출입인구와 실제 직장인구와의 Pearson R값 : 0.60

-> 거주, 직장 분석 결과 모두 Pearson R값이 0.3~0.7 이므로 뚜렷한 상관관계가 있음을 확인함

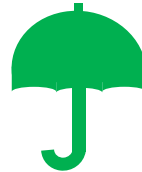


# 승하차 인구에 따른 대여 반납 예상

02 분석 데이터 및 방법  
Umbrella Sharing System



대여 = 하차역 기준

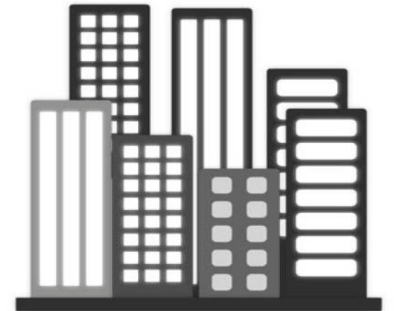


반납 = 승차역 기준

(출퇴근 관계 없이 특정 역 기준)



OR



하차 승객 수 - **승차** 승객 수 > 0

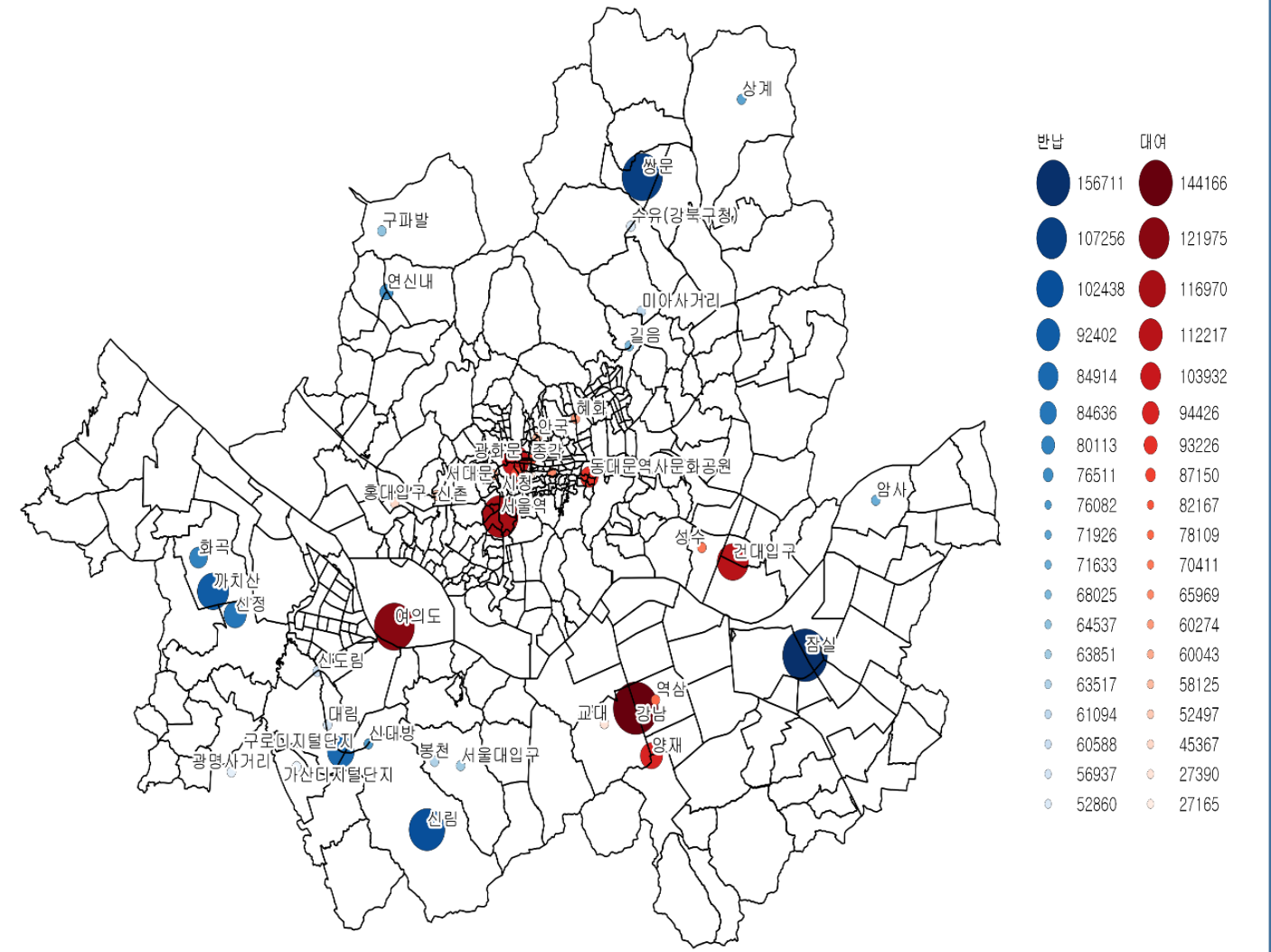
대여

하차 승객 수 - **승차** 승객 수 < 0

반납



<우산 대여 및 반납 수량 예측>



$$X = (\text{하차인구}) - (\text{승차인구}) \begin{cases} X > 0 : \text{대여} \\ X < 0 : \text{반납} \end{cases}$$

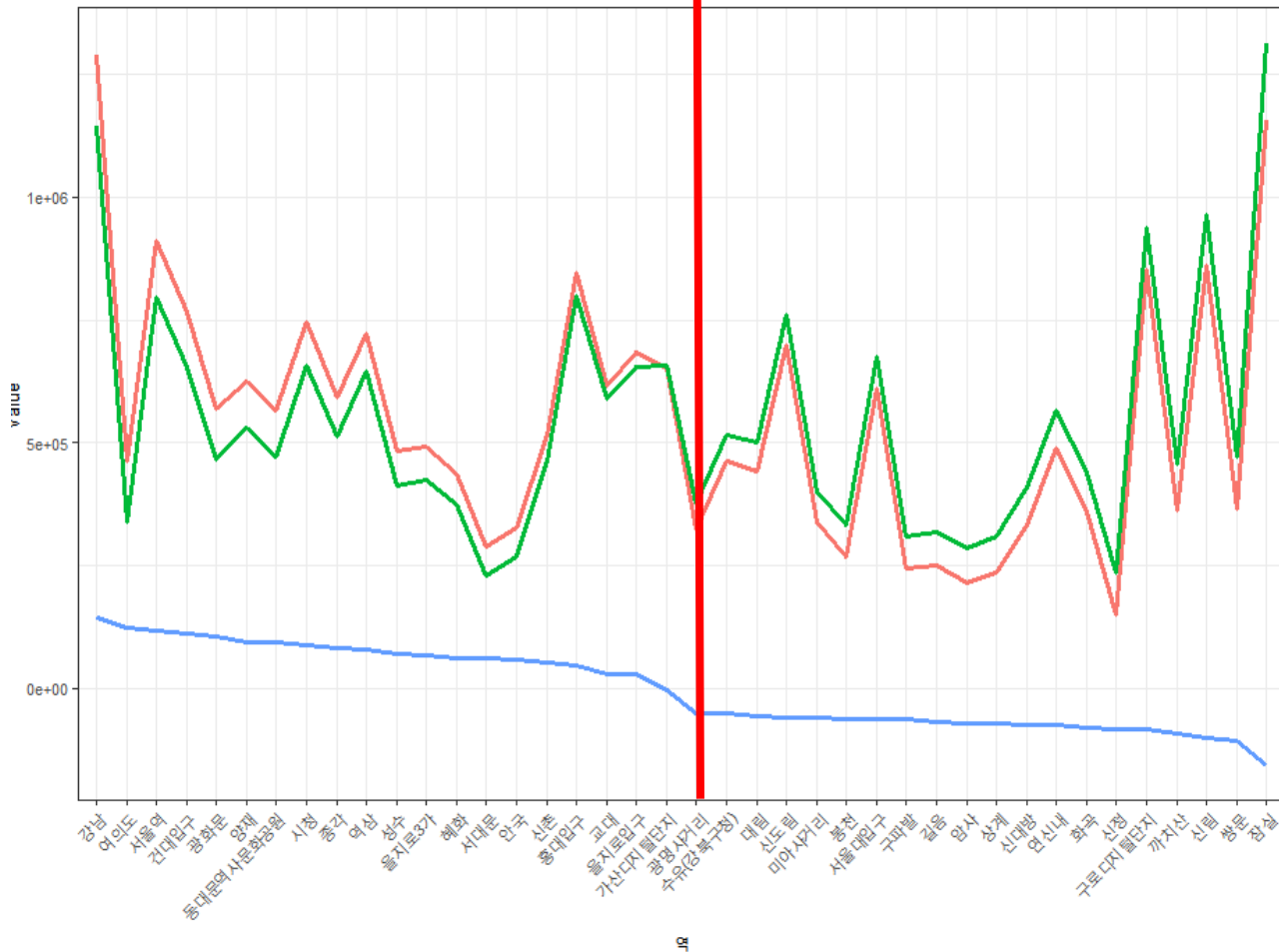
대여 Top5		
순위	역	추정치( X )
1	강남	144,166
2	여의도	121,975
3	서울역	116,970
4	건대입구	112,217
5	광화문	103,932

반납 Top5		
순위	역	추정치( X )
1	잠실	-156,711
2	쌍문	-107,256
3	신림	-102,438
4	까치산	-92,402
5	구로디지털단지	-84,914



<대여가 더 많은 역>

<반납이 더 많은 역>



우산 **대여** 수요 가 높은 **상위 20개 지하철 역**

우산 **반납** 수요 가 높은 **상위 20개 지하철 역**

**“ 40개 ”** 역을 **대표역**으로 선정

지하철 (266개역)	승 하차인구 비중
10 + 10 역 (7%)	18 %
20 + 20 역 (15%)	14 %
30 + 30 역 (22%)	6 %

- 20+ 20개역 선정 이유

-> 가장 필요한 역을 기준으로 **“시범 사업”** 시행

# Sampling

## 03 분석

Umbrella Sharing System

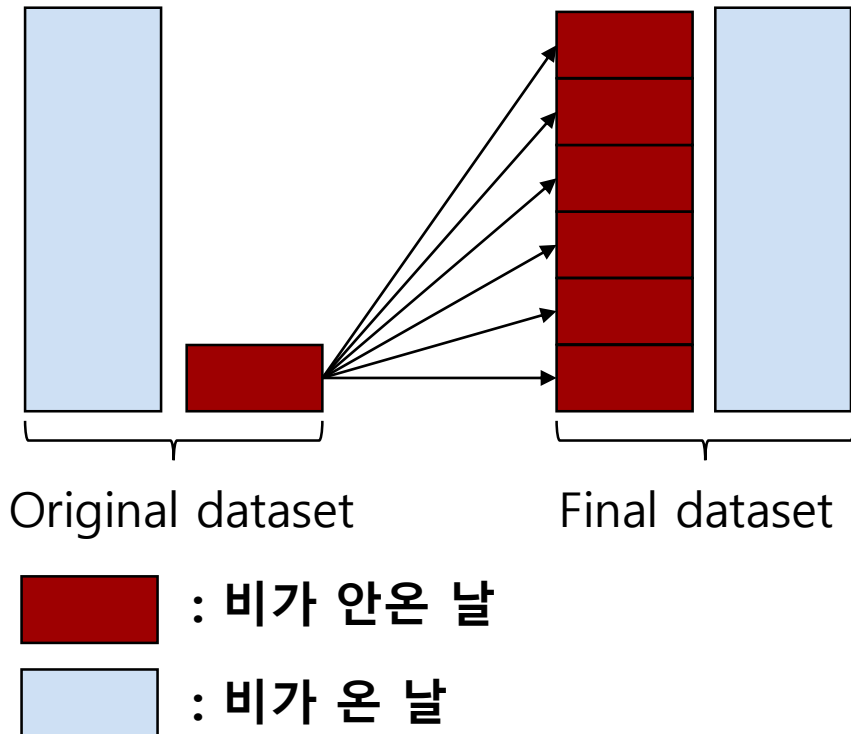


비가 온 날의 데이터 수 : 98,088

비가 안온 날의 데이터 수 : 1,970,777

**약 20배의 차이**

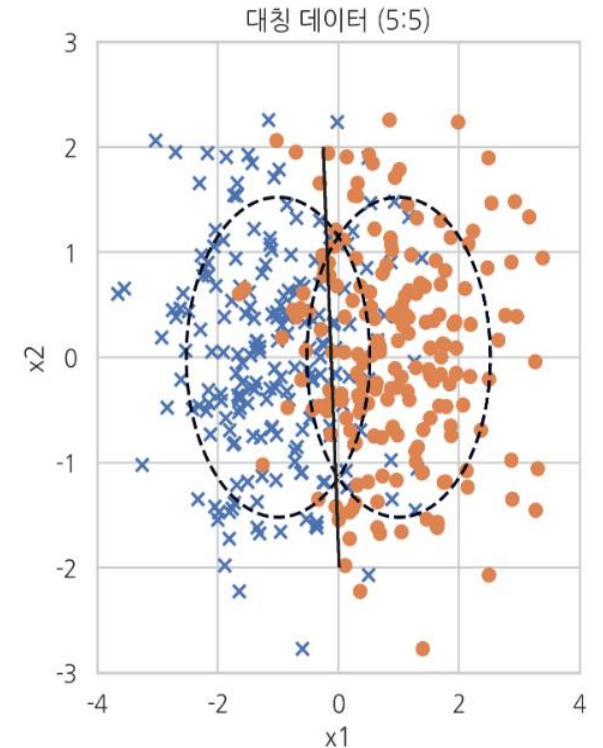
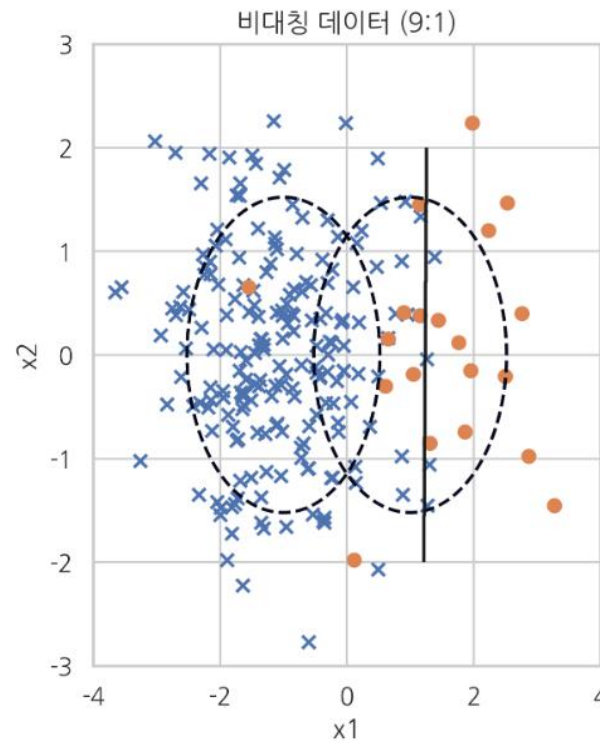
Oversampling



## SMOTE 를 이용한 샘플링 장점

-> 오버샘플링으로 인한 과적합 문제를 인스턴스 복제 보다는 합성 예제 생성으로 완화 시킬 수 있고 유용한 데이터의 손실이 없다.

불균형 -> 균형





강우량 정도에 따라 달라지는

역별 승/하차 인구를 예측한다

모델의 성능 평가 척도

1. R-square(결정계수)

-> 추정된 선형 모형이 주어진 자료에 적합한 정도의 척도이다.  
반응 변수의 변동량 중에서 적용한 모형으로 **설명가능 한 부분의 비율**을 가리킨다.

2. RMSE(평균 제곱근 오차)

-> 추정 값 또는 모델이 예측한 값과 **실제 환경에서 관찰되는 값과의 차이**를 다룰 때 흔히 사용하는 척도이다.

좋은 모형 기준

=> **R-square ↑ RMSE ↓**

모델 비교			
순위	모델	R-square	RMSE
1	Random Forest	0.63	947.0419
2	MLP	0.60	968.3885
3	Gradient Boost	0.48	1039.5894
4	Ada Boost	0.19	1294.4379

=> **Random Forest 선택**





승하차 인구 모두가 우산이 필요할까?

- 집에서 우산을 가지고 온 사람을 어떻게 배제할까?  
-> 편의점 우산 판매량을 실제 우산 수요라 생각하자!!





## “서울역” 편의점 Storyway의 2018년 7월 우산판매량

편의점 시간대별 우산 판매량						
시간대별 매출 조회(일반조회)						
조회일자 : 20180701~20180731						
				(단위 : 수량)		
본부명	업종명	매장코드	매장명	06시~07시	07시~08시	08시~09시
서울	편의점	205570	서울중앙직영편의점	5	18	25
				18시~19시	19시~20시	20시~21시
				10	21	11

출근시간

퇴근시간

“서울역” - 유일한 우산판매량 데이터 확보

비온 횟수, 하차 인구를 고려  
다른 역의 우산 판매량을 추정

## [우산 판매량 추측 Process]

<기준 : 서울역>

$x_j$  = 우산판매량  $K = 100m$  내 편의점 수  
 $n$  = 18년 7월 비온 횟수

$$\text{총 우산 판매량} = \frac{K}{n} \times \sum_{i=j}^{i+2} x_j, j = \begin{cases} 6 & (\text{출근}) \\ 18 & (\text{퇴근}) \end{cases}$$

<추정 : 다른 역>

하차인구 비율 = (다른 역) 하차인구 / (서울역) 하차인구

비온 횟수 비율 = (다른 역) 비온 횟수 / (서울역) 비온 횟수

**(다른 역)우산 판매량 산정 공식**

=> 총 우산 판매량 X 하차인구 비율 X 비온 횟수 비율



## < 대여 수요 >

$$(\text{모델링을 통해 나온 역별 하차 인구}) \times \frac{\text{역별 추정 우산 판매량 (18년 7월)}}{\text{18년 7월 비온 시간대의 역별 하차 인구}}$$

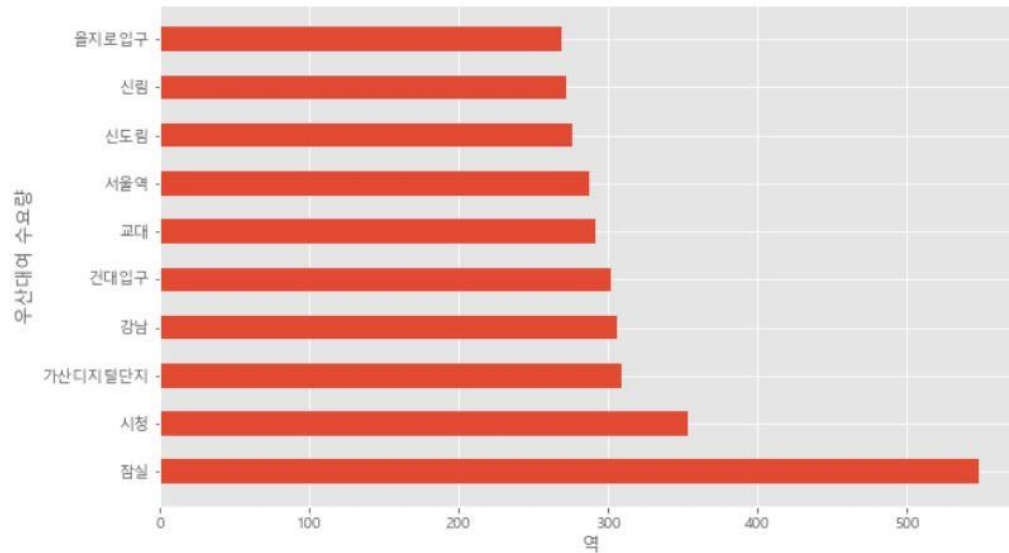
## < 반납 수요 >

전제조건 : 우산의 대여량과 우산의 반납량은 항상 같다

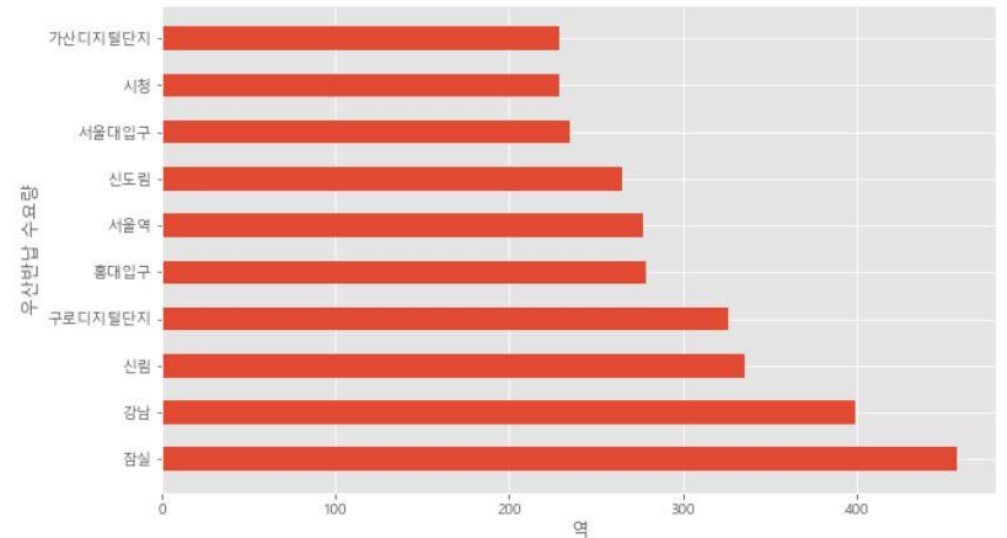
$$(\text{대표역 우산 판매량들의 총 합}) \times \frac{\text{18년도 특정 역의 승차인구}}{\text{18년도 대표역들의 승차인구 총합}}$$



<우산 대여 수요 상위 10개 역>



<우산 반납 수요 상위 10개역>



<분석 결과>

1. 잠실역 이 대여 수요량 (548개), 반납수요량(458개)로 대여/반납 모두 가장 많은 역
2. 8개 역이 대여 수요 상위 10 과 반납 수요 상위 10을 동시에 차지함

=>대여가 많은 지역에 대체로 반납도 많이 이루어진다는 결론 도출

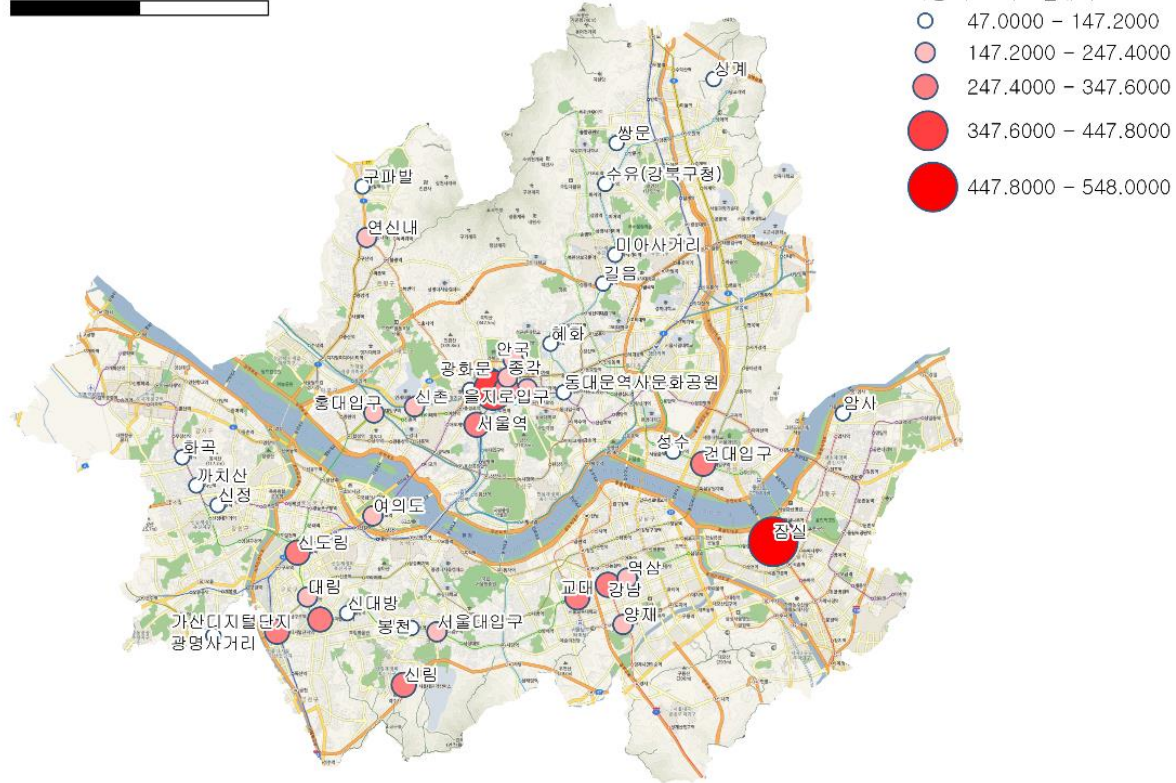
# 최종 역별 우산 수요량

## 04 결과 Umbrella Sharing System



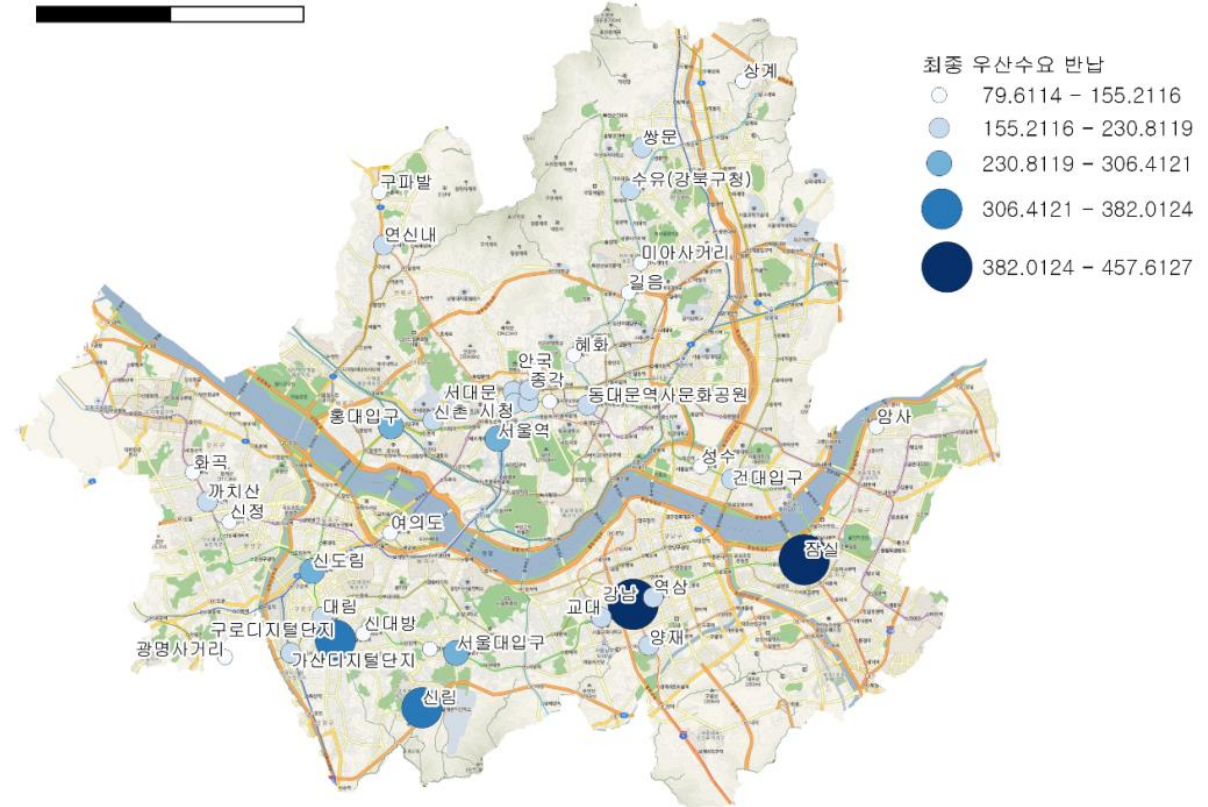
< 대표역별 우산 대여 수요 >

0 5 10 km



< 대표역별 우산 반납 수요 >

0 5 10 km



**"7,541"** 개의 **우산** 배치가 필요하다는 결론 도출



## < 지하철 출퇴근 인구 분석 >

1. 비/눈 에 따른 지하철 승하차 인구 변화는 강우/적설 정도에 따라 다양한 변화를 보였음
2. 지하철 유출입 인구를 통한 수학적 추정 방법은 실제 거주/ 인구 직장인구와 상관 분석 결과 상당한 일치를 보였음

## < 서울시 공유 우산 대여 시스템 기획 >

“출퇴근인구”를 대상으로 우산 대여/반납이 많은 40개역을 선정하여 진행

모델링을 통해 강우량 정도에 따른 **역별 승/하차 인구**를 예측하고

우산 판매량 데이터를 활용하여 **역별 우산 대여 / 반납 수요량**도 추정 계산함

그리고 원활한 시스템 도입을 위한 **“총 우산 배치 필요량”**도 계산함



## 활용 방안

### 1. 타 기상상황에 따른 휴대 용품 공유

-> 폭염, 황사 등 다양한 기상 환경에 따른 필요한 물품(손난로, 손풍기 등)을 공유 활용 할 수 있다.

### 2. 시범 사업을 거쳐 서울시 전역 서비스 확대(공유자전거 따릉이)

-> 시범 사업을 거쳐 대여 관련 데이터가 많이 축적되면 따릉이처럼 서울시 전역으로 서비스 확대할 수 있다.

## 한계점

### 1. 정확한 우산 판매량 데이터

-> 영업 기밀로 인해 우산 판매량 데이터를 확보하지 못하였다. 실제 데이터가 주어졌다면 수요 관련 더 정밀한 예측이 가능할 것이다.

### 2. 날씨 예보 데이터 필요

-> 과거 날씨 예보에 쓰였던 데이터가 있다면 집에서 자가우산 비율을 추정할 수 있어 더 정확한 예측이 가능하다.





따릉이

시범 사업 자전거 수 : 약 2,000 대  
자전거 1대 가격 : 890,000 원  
-> Total : 1,780,000,000 원



서울우산 토닥이

시범 사업 필요 우산 수 : 7,541 개  
우산 소비자 가격 : 5,000원  
-> Total : 37,705,000 원



## 참고 문헌

- 「효과적 수요 예측 방법과 사례」 - SERI 이슈페이퍼. 박성배 수석연구원(2012)
- 「기상정보를 활용한 의류제품 판매 예측 시스템 연구」 - 한국의류산업학회지. 오재선, 오희선(2017)
- 대중교통 승하차 수요분석을 통한 서울시 역세권 유형화 및 토지이용 특성 연구 (이정우, 2015)



## 데이터 출처

데이터	형식	출처	기준연도
서울시 지하철 데이터	csv	서울시 빅데이터 캠퍼스	2014~2015
서울시 거주인구	csv	서울시 빅데이터캠퍼스	2014~2015
서울시 직장인구	shp	서울시 빅데이터캠퍼스	2014~2015
날씨 데이터	csv	기상청	2014~2015
서울시 지하철 데이터	csv	서울교통공사	2018
편의점 우산판매량 (서울역)	xlsx	코레일 유통 본사 판매팀	2018

## <활용 분석 도구>





감사합니다.