



적자 '따릉이'광고 달고 달려야하는데...

기사입력 2018-03-29 11:29

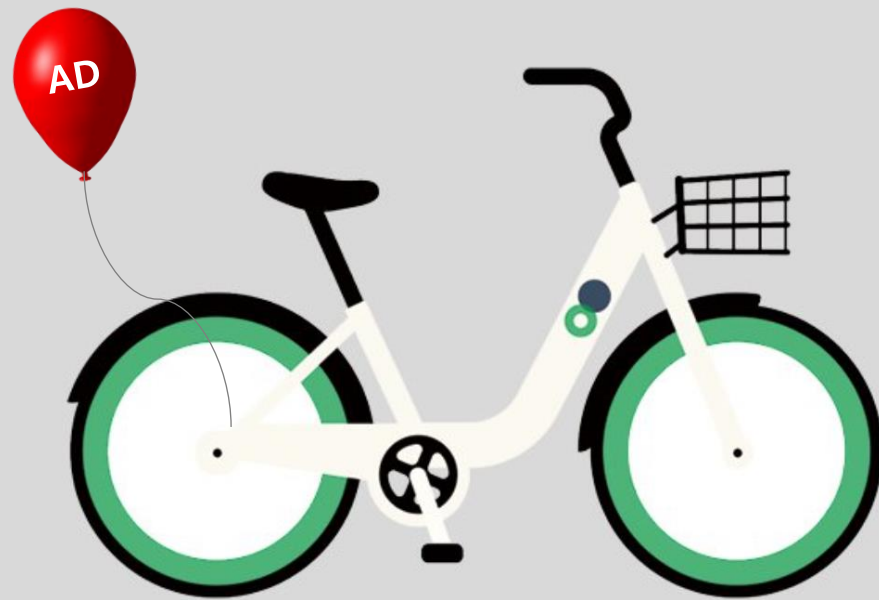
운영비용 급증...유치방안 고심

서울시가 공공자전거 '따릉이' 광고권을 파는 데 난항을 겪고 있다. 시는 적자 해소 방안으로 다음 달 중 광고 사업자를 뽑는 방침을 세웠지만, 기업들은 불확실성을 지적하며 이를 망설이는 것으로 알려졌다.

시가 따릉이 광고권을 판매하려는 것은 매년 운영적자가 눈덩이처럼 불고 있기 때문이다. 시의 '따릉이 적자 규모'를 보면 따릉이로 인한 적자는 2015년(9~12월) 약 3억7300만원, 2016년 약 23억4900만원, 지난해 약 38억8100만원(추정)에 이른다. 복지 차원에서 이용료를 최소화해 타는 시민이 늘수록 적자는 늘어나는 형식이다. 따릉이는 현재 대여료를 통해서만 수익을 내고 있다.



서울시가 공공자전거 따릉이 '광고권'을 파는 데 난항을 겪고 있다. 기업들은 불확실성 등을 이유로 광고유치를 망설이는 것으로 알려졌다.



따릉이는 광고를 심고~

Team : 4 Merge

오세민 | 표병수 | 한용희 | 한호재

목차

- 프로젝트 개요
- 데이터 수집 및 탐색
- 데이터 전처리
- 데이터 분석
- 요약

프로젝트 개요

연구배경

- 현재 따릉이 '이용 요금 수입' 대비 '운영 비용' 이 너무 큰 실정
- 15년 9월 이후 매년 '적자' 기록 (17년 약 38억)
- 현재 따릉이 광고 효과와 방식에 대한 의문

효율적인 광고를 위한 따릉이 빅데이터 분석



데이터 수집 및 탐색

데이터 수집 및 탐색

데이터셋 리스트

- 서울시 공공자전거 대여이력 정보 (2017.07~2018.06)
- 서울시 실시간 자치구별 대기환경 현황(API) (2017.06~2018.06)
- 종관기상관측 데이터 (2017.06~2018.06)
- 서울시 공공자전거 대여소 정보

데이터 수집 및 탐색

서울시 공공자전거 대여이력 정보 (2017.06~2018.06)

'자전거번호'	'대여일시'	'대여대여소번호'	'대여대여소'	'대여거치대'	'반납일시'	'반납대여소번호'	'반납대여소'	'반납거치대'	'이용시간(분)'	'이용거리(M)'
---------	--------	-----------	---------	---------	--------	-----------	---------	---------	-----------	-----------

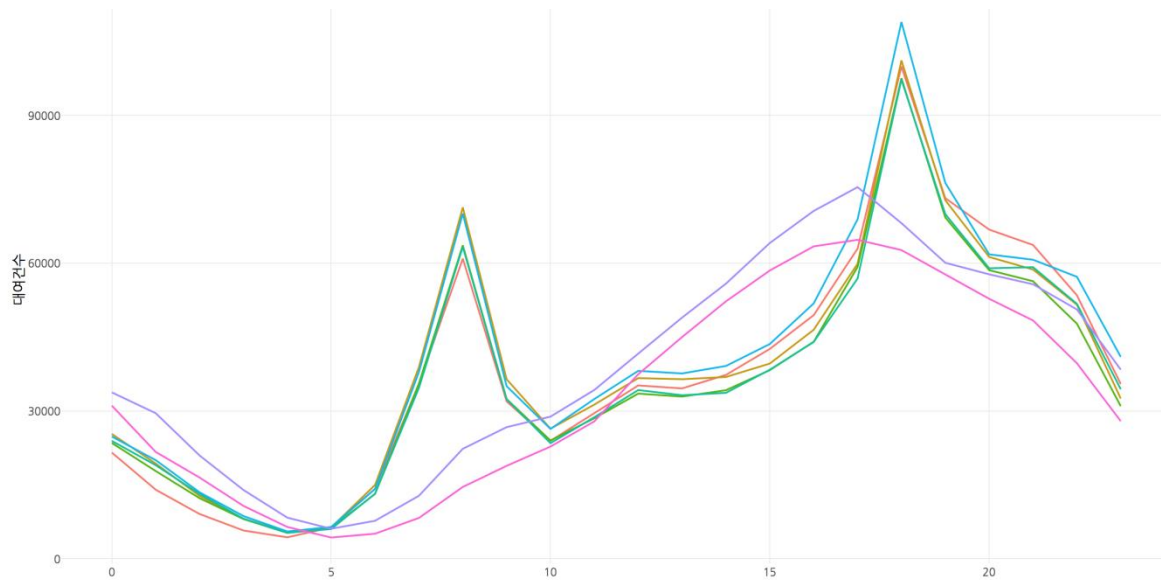
680만 rows

대여일시 | 대여대여소번호 | 반납일시 | 반납대여소번호

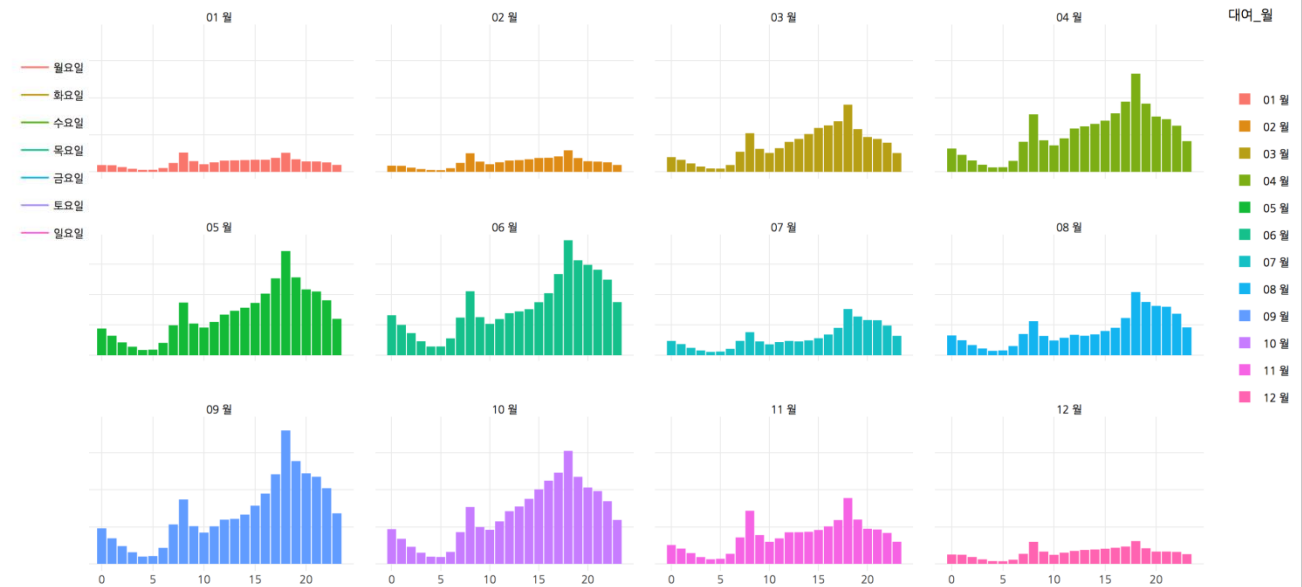
데이터 수집 및 탐색

서울시 공공자전거 대여이력 정보 (2017.06~2018.06)

■ 탐색적 데이터 분석



< 요일별 대여건수 >



< 월별 대여건수 >

데이터 셋의 요일별, 월별 등 데이터 패턴 파악

데이터 수집 및 탐색

서울시 공공자전거 대여이력 정보 (2017.06~2018.06)

■ 데이터 특성 파악에 따른 속성 세분화

'자전거번호'	'대여일시'	'대여대여소번호'	'대여대여소'	'대여거치대'	'반납일시'	'반납대여소번호'	'반납대여소'	'반납거치대'	'이용시간(분)'	'이용거리(M)'
---------	--------	-----------	---------	---------	--------	-----------	---------	---------	-----------	-----------

680만 rows

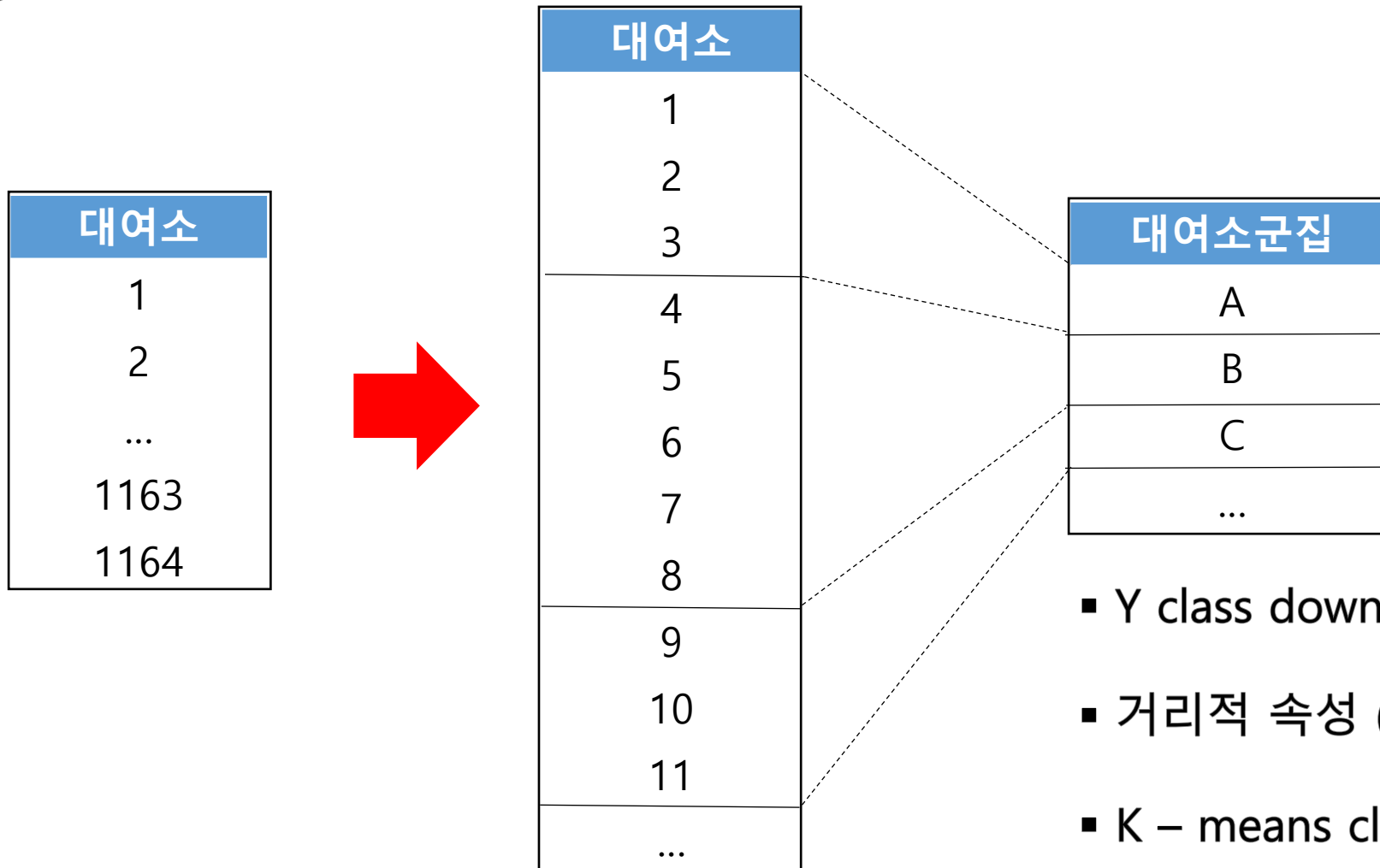
대여일시 | 대여대여소번호 | 반납일시 | 반납대여소번호

대여일시 : 주말 및 공휴일 | 시간대 (범주화) | 계절

데이터 전처리

데이터 전처리

Clustering



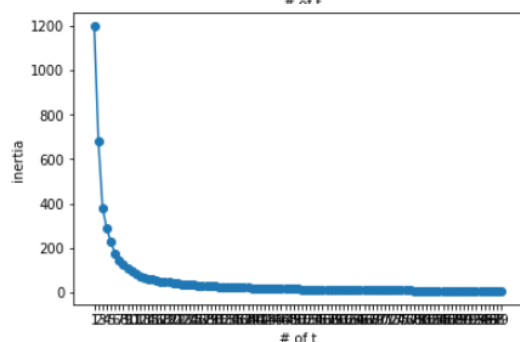
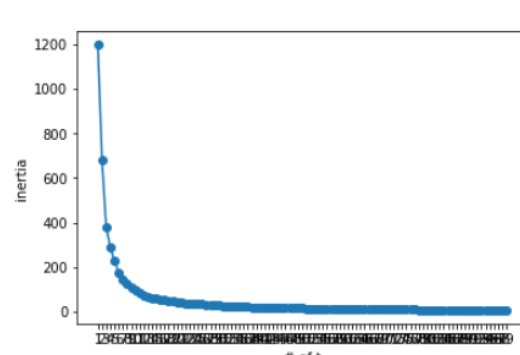
- Y class downsizing
- 거리적 속성 (위도 | 경도)
- K – means clustering

데이터 전처리

Clustering

- 한강을 포함하지 않도록 [강남 / 강북] 개별 클러스터링

- # of K : K inertias



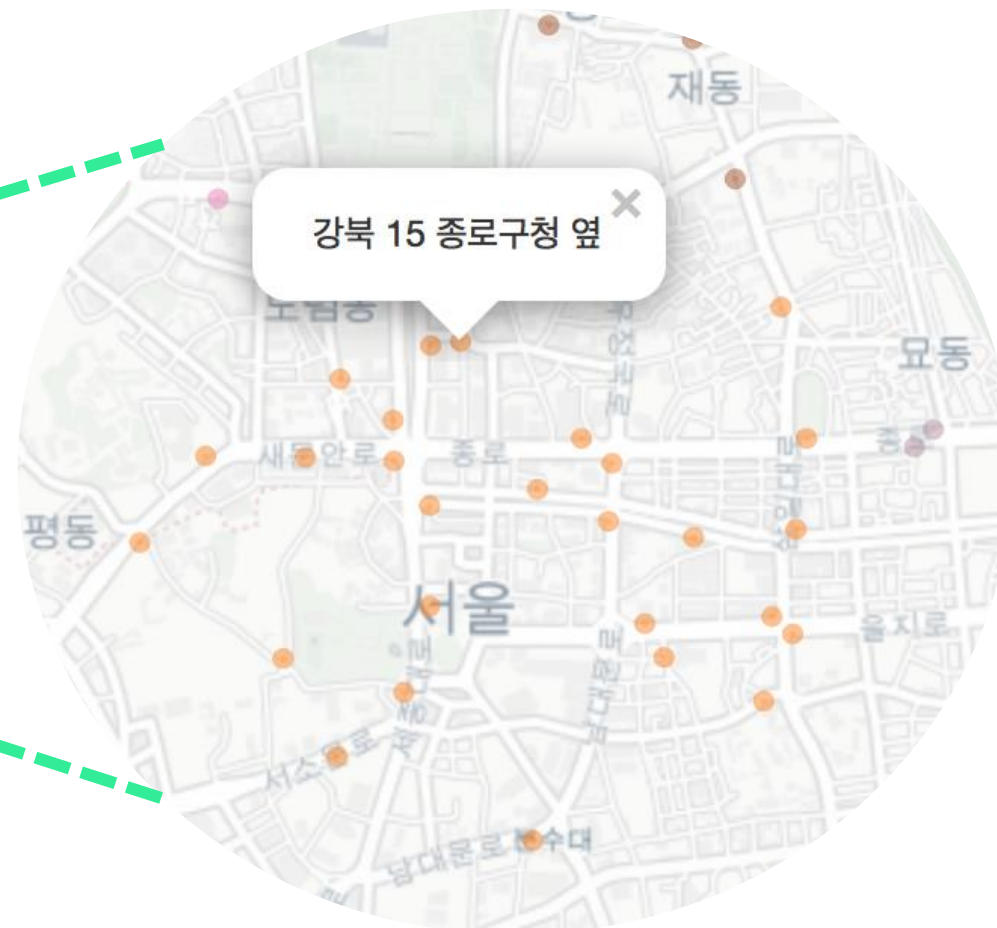
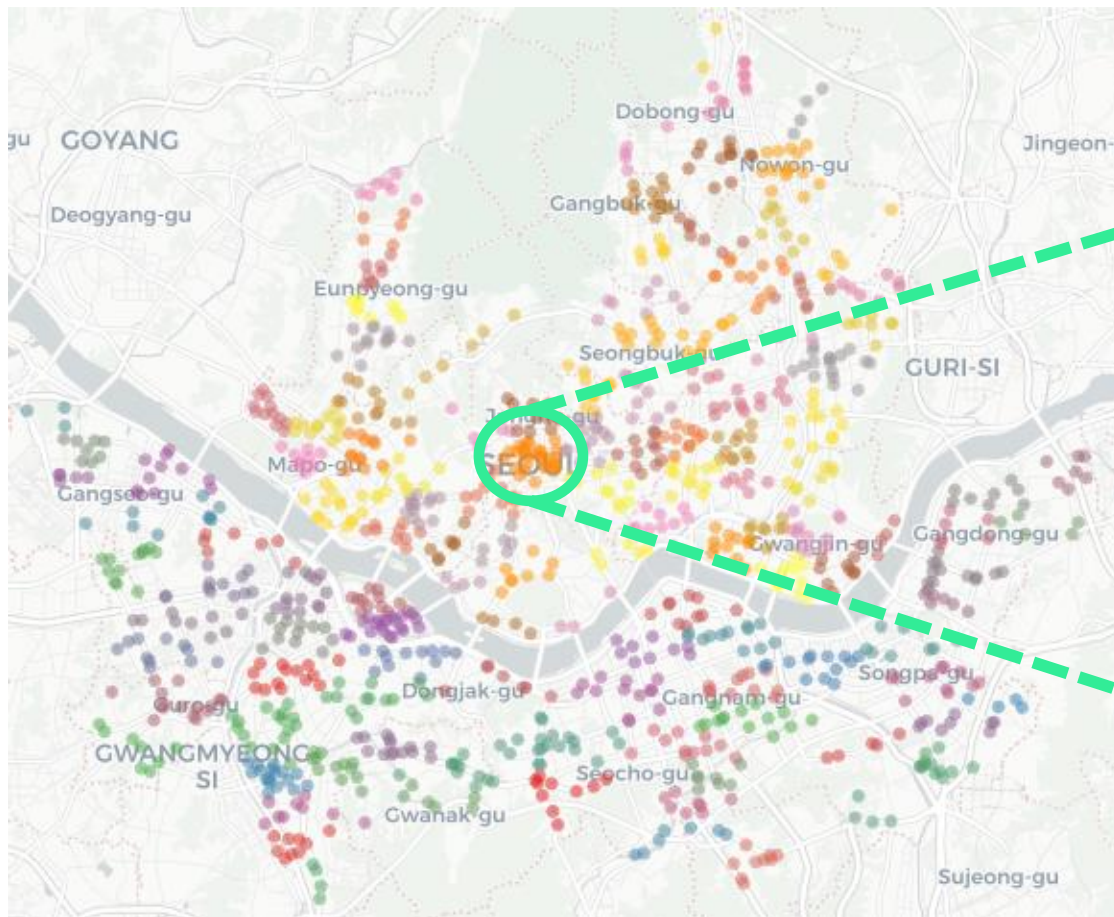
강남 (56)	강북 (68)
50 : 11.753786416039409	60 : 11.764441488738816
51 : 11.502021092798353	61 : 11.733674146177346
52 : 11.486817549576198	62 : 11.339631990577777
53 : 10.935358374696865	63 : 10.880722839156787
54 : 10.665048074611438	64 : 10.882493134821468
55 : 10.294963375064171	65 : 10.719609210691301
56 : 10.179419673631104	66 : 10.390857082249687
57 : 9.870218996436403	67 : 10.347897253063886
58 : 9.658744254618298	68 : 10.257112988663998
59 : 9.51955309212502	69 : 9.974066418231512
60 : 9.418198309569162	70 : 9.88694322118134
61 : 9.172987105905586	71 : 9.65752575808309
62 : 8.99757050370815	72 : 9.443444077885587
63 : 8.668025416648522	73 : 9.493332444083729
64 : 8.659338161697676	74 : 9.135014264566188
65 : 8.184573014148	75 : 8.892674510781779
66 : 8.192341251430133	76 : 8.918278192879043
67 : 8.25556065799842	77 : 8.60915463683103
68 : 7.944583842989262	78 : 8.610573310317166

<Select K>

- K inertias 값은 작을 수록 좋은
- 하지만, 너무 작게 군집화가 됨을 방지
- K inertias = 10 에서 군집개수 결정

데이터 전처리

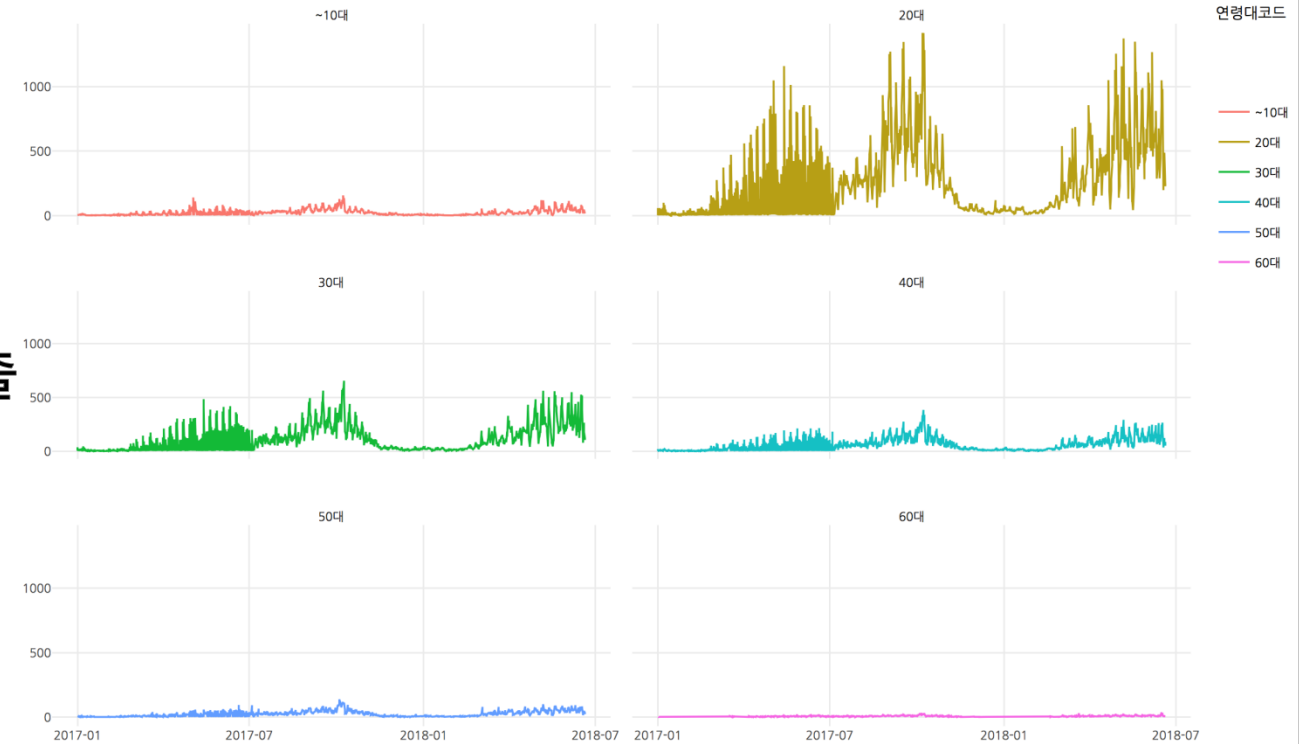
Clustering



데이터 전처리

데이터 생성

- 타겟 광고를 위한 데이터 생성
- 따릉이 어플 가입자의 성별 및 연령 통계 도출
- 통계의 가중치 바탕으로 개인정보 랜덤 생성



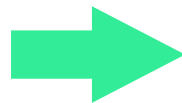
< 서울시 공공자전거 연령별 신규가입자수 시각화 >

데이터 전처리

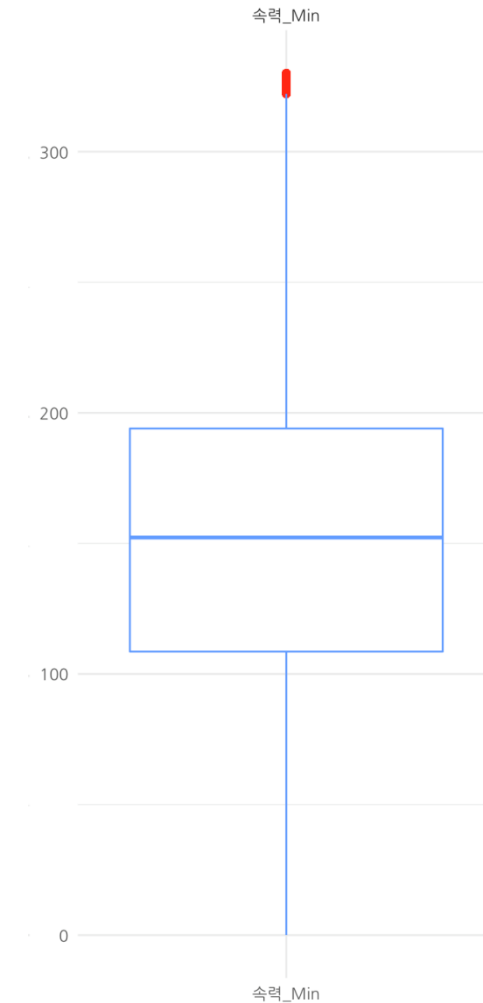
데이터 전처리 결과

- 결측치(Null) 제거
- 이상치 제거 (Box-plot Outlier)
- 군집간 이동 365회 미만 제거 (군집간 이동 1일 1회 미만)

680만 rows



530만 rows



< 속력데이터 Boxplot 예시 >

네트워크 분석을 통한 광고 활용 방안

네트워크 분석을 통한 광고 활용 방안

1) 각 군집별 내부 순환율 도출

- 군집간의 이동 [대여 군집 -> 반납 군집] 순서쌍 도출
- 각 순서쌍의 대여이력 횟수 count

대여 군집	반납 군집	count
A	A	1,156
A	B	967
B	B	3,200
C	A	500
C	D	2,307
D	C	1,893
...

* 내부 순환율

$$\frac{\text{(반납 군집이 대여군집인 횟수)}}{\text{(대여 군집의 총 대여 횟수)}}$$

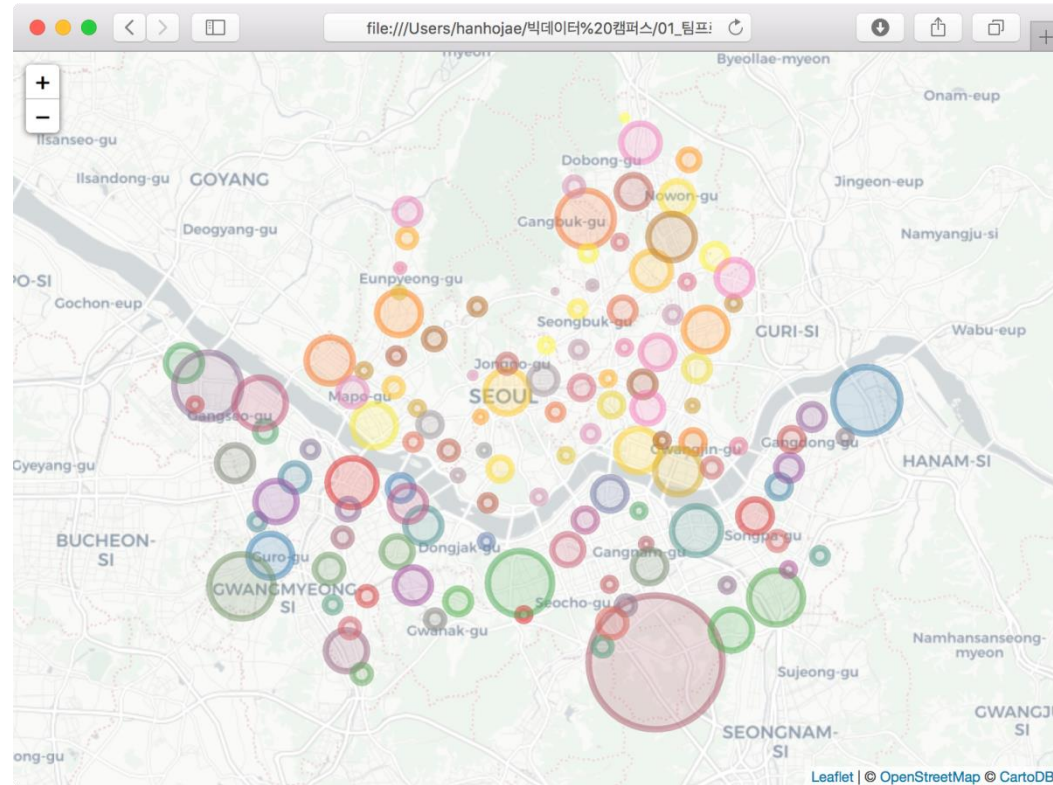
ex) A 군집의 내부순환율

$$\frac{1156}{(1156 + 967)} = 54.45(\%)$$

네트워크 분석을 통한 광고 활용 방안

2) 군집별 속성을 이용한 옥외 광고 예시

- 내부순환율이 높은 군집 옥외 광고



< 군집별 총이용건수 대비 내부이용율 Bubble plot >

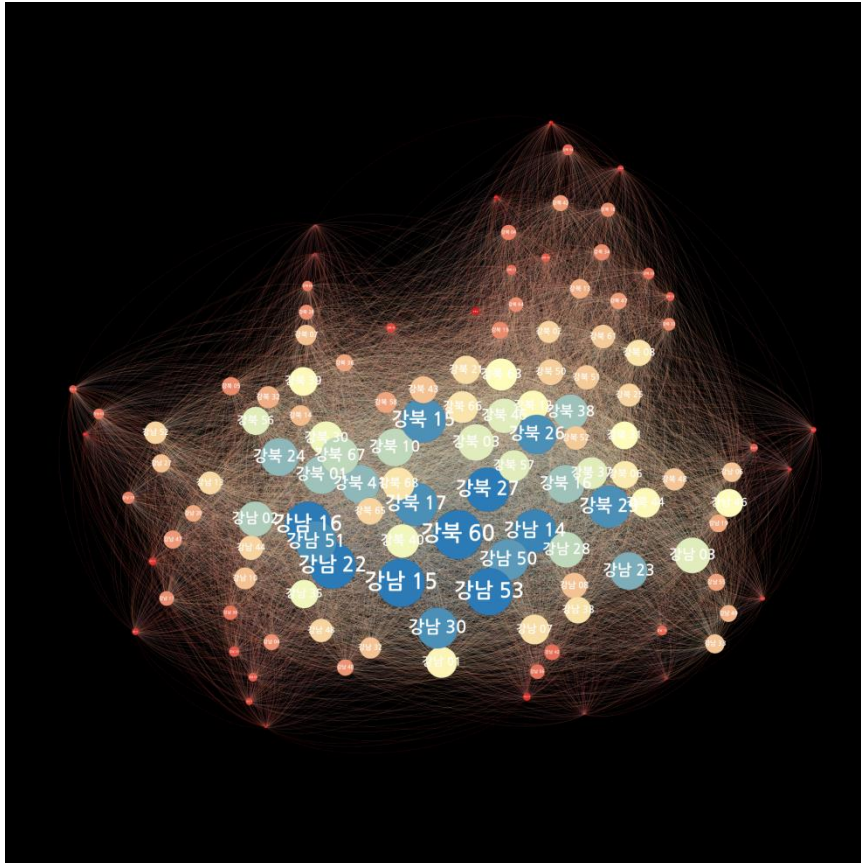
- 내부순환율이 높아 군집 이탈율이 낮음
- ex) 지역속성을 대표하는 광고 설치



네트워크 분석을 통한 광고 활용 방안

2) 군집별 속성을 이용한 옥외 광고 예시

- 내부순환율이 낮은 군집 옥외 광고



- 내부순환율이 낮아 군집 이탈율이 높음
- ex) 지역속성의 영향을 적게 받는 광고 설치

Coca-Cola



STARBUCKS®



주식회사 **오뚜기**

네트워크 분석을 통한 광고 활용 방안

4) 군집별 속성을 이용한 옥외 광고 설치 예시



Citibike(NY)



바이쿨

XGBoost를 활용한 타겟 광고 활용 방안

XGBoost를 활용한 타겟 광고 활용 방안

1) XGBoost

<통합 데이터 (530만 rows)>

Data Columns
반납균집
대여균집
주말 및 공휴일
시간대 (범주화)
기온
계절
미세먼지농도
초미세먼지농도
성별
연령대

▪ Target Variable (Y)

▪ Independent Variables (X)

<XGBoost 과정>

- 하드웨어 Computing power 한계
- 530만 -> 100만 rows 랜덤 샘플링
- 범주형 변수의 더미화

연령대	10대	...	60대
10대	1	..	0
..

▪ [Train : Test] = 7 : 3

XGBoost를 활용한 타겟 광고 활용 방안

2) XGBoost 결과

```
# 0825 17:15
xgb_4 = XGBClassifier(booster='gbtree',
                      n_estimators=30,
                      silent=0,
                      max_depth=7,
                      eta=0.2,
                      objective='multi:softprob', num_class=124,
                      eval_metric='error',
                      gamma=0,
                      subsample=0.8,
                      colsample_bytree=0.8,
                      n_jobs=-1, # 모든 CPU 사용
                      cv=4, # 병렬처리
                      verbose=1
)
```

```
# 평가
xgb_4.score(x_test, y_test)
```

```
/anaconda3/lib/python3.5/site-packages/sklearn/preproc
: The truth value of an empty array is ambiguous. Retu
sult in an error. Use `array.size > 0` to check that a
if diff:
```

0.395968

<Score 해석>

- 약 40%의 정확도
- <전문가 의견 (서울대학교 박사과정 김강진)>

“X data의 Column의 개수와 Y의 Class 개수를 고려하였을 때 (Prob = 1/124), 모델의 적합도 40%는 높은 수치라 생각합니다.”

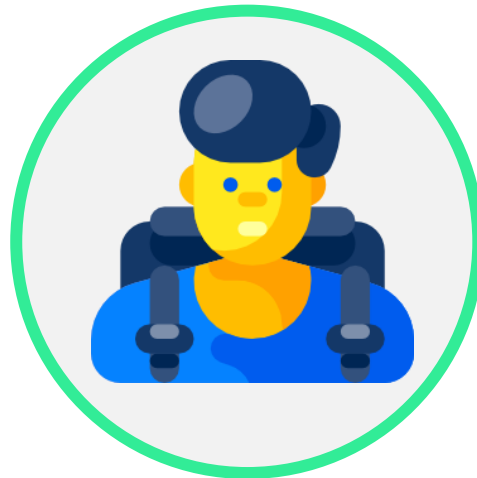
- 따라서, XGBoost를 사용하여 모델 구축

XGBoost를 활용한 타겟 광고 활용 방안

3) 모델을 활용한 앱 광고 실행 방안

<상황 설정>

20대 | 남성 | 가을 | 25.0 °C | 주말 | 퇴근시간대 | 사당역 5번출구 대여 |

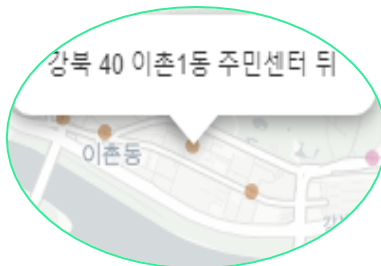
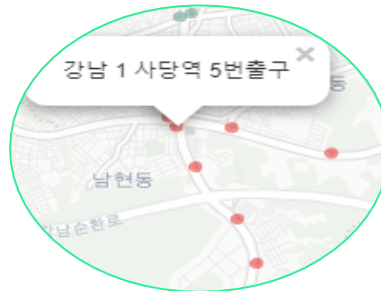
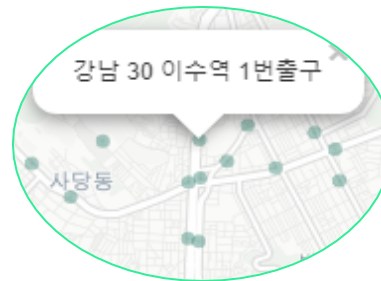
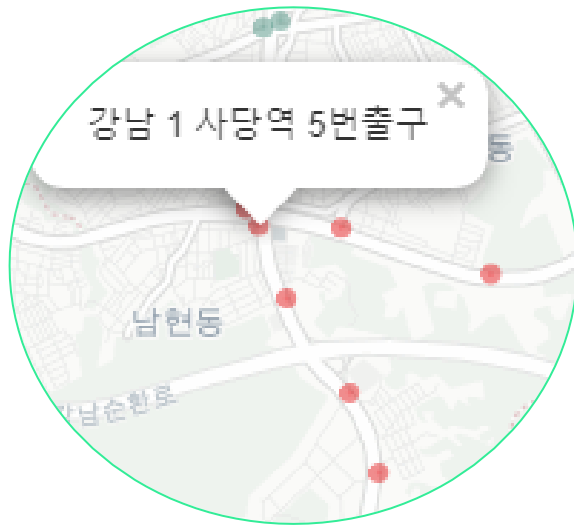


[사당역 5번출구 따릉이 대여소]

XGBoost를 활용한 타겟 광고 활용 방안

3) 모델을 활용한 앱 광고 실행 방안

<결과 예측> (by XGBoost.predict_proba)



1순위 강남 30 (0.201527)
ex) 미용실, 생활용품, 편의점...

2순위 강남 01 (0.141744)
ex) 직장인 점심맛집...

3순위 강북 40 (0.121398)
ex) 명품샵, 백화점 ..

XGBoost를 활용한 타겟 광고 활용 방안

3) 모델을 활용한 앱 광고 실행 예시



사용자 대여완료 순간 상위군집 타겟 광고 팝업

Development

Development



Thank You

Q & A

#. References



- 뉴스 기사 (적자, 따릉이 광고 실패 달려야하는데...) – http://news.heraldcorp.com/view.php?ud=20180329000358&md=20180401003321_BL
- 서울시 종관기상관측현황 데이터 – 기상자료개방포털
- 서울시 공공자전거 대여이력 정보 – 서울열린데이터광장
- 서울시 공공자전거 대여소 정보 – 서울열린데이터광장
- 서울시 자치구별 대기환경 현황 – 서울열린데이터광장