# 위치기반따름이 이용자 맞춤 광고 추천

팀 퍼스트

## 목차

1. 서론

01. 분석 배경

2. EDA

01. 데이터 이해02. 데이터 전처리03. 데이터 시각화

3. 모델링

01. 1차 모델링02. 2차 모델링

4. 결론

01. 활용방안

02. 기대효과



0

### 서론

01. 분석 배경





### 분석 배경

따름이 총 대여건수 2020년 24% 증가 2021년 35% 증가

잇는 '퍼스트-라스트 마일'

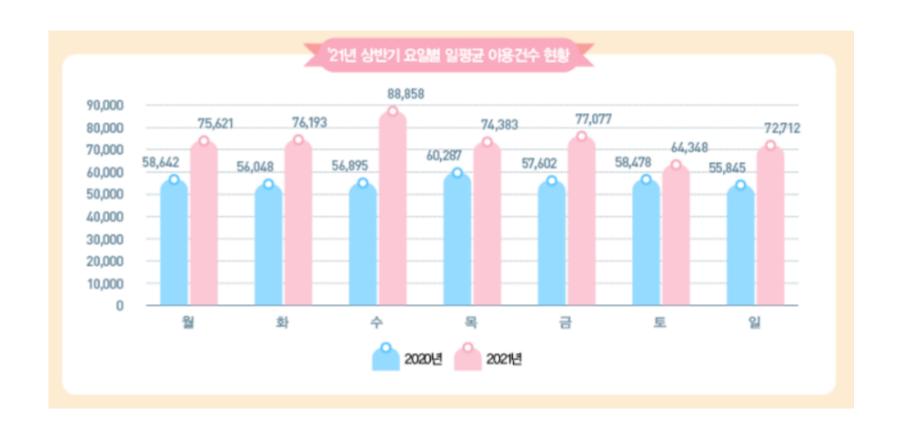
역할 톡톡

따름이 이용자

서울시민 3명 중 1명 꼴로 따름이 사용 경험 있음

증가 대중교통 이용 전후 틈새를

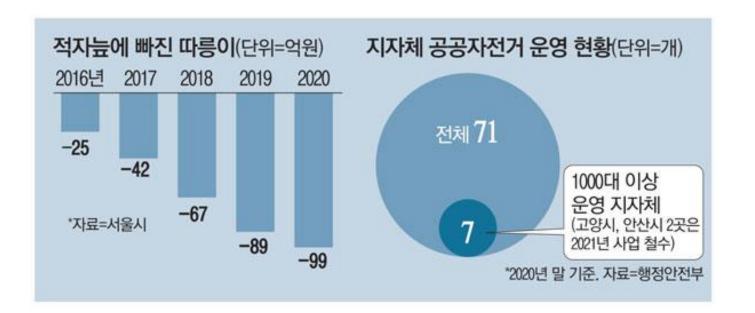
따름이 대여소 연말 3000개 증가 총 40500개



자전거, 대여소 노후화로 유지보수 비용 증가

따름이 적자 상황

적자폭 매년 증가 2021년도 103억 돌파



### [단독] 적자 100억 따름이 "당분간 안 늘린다...정책 전환"

2021. 6. 27. — [앵커] 서울시 공공자전거로 자리 잡은 '따름이'. 하지만 지난해 **적자**가 100억이 넘 을 정도로 운영: 지난해만 '100억 적자' 따름이...계속 달릴 수 있을까? - KBS뉴스 2022. 7. 24. — [앵커] 서울시 공공자전거를 '**따름이**'라고 하죠. 이용료가 저렴하고 접근성도 좋아

한해 100억 적자 '따릉이' 어쩌나...지방선 공공자전거 사업 철수

### 분석 배경

기업광고 달고 달리는 따름이

서울시는 적자가 누적될 경우 서비스 개선의 걸림돌로 작용할 수 있다고 판단, 수익구조를 다변화하는 차원, 기업광고를 유치하기로 함

#### 100억 적자 공공자전거 '따릉이' 연말에 기업광고 달린다

2022. 7. 5. — 지난해 100억원 이상에 달하는 **적자**를 내며 본격적인 수익구조 개선에 나선 것이 다. **따릉이**는 2015년 이후 이용자 수가 꾸준히 증가해 지난 4월 누적 1 ...

백호 서울시 도시교통실장은 "이용요금에만 국한됐던 따름이 수익구조를 다변화하고 운영효율을 높여나가겠다"며 "탄소 저감과 환경보호같이 따름이가 지향하는 가치에 부합하고 관심이 있는 기업들의 참여를 기대한다"고 말했다.

문제인식

따름이 이용자 증가 BUT 심각한 따름이 적자

서비스 제안

따릉이 대여소 및 프레임에 기업광고 도입

#### 서비스 관련 모델 제시

#### 활용 정보

1. 이용자 정보

- ① 성별
- ② L<sub>1</sub>01
- ③ 운동량

2. 대여소 위치 정보

해당 대여소 위치의 상권 특성

#### 광고 위치

- 1. 따름이 대여소 마다 스크린 설치하여 광고 노출
- 2. 따름이 반납 시 핸드폰에 광고 노출

02

### EDA

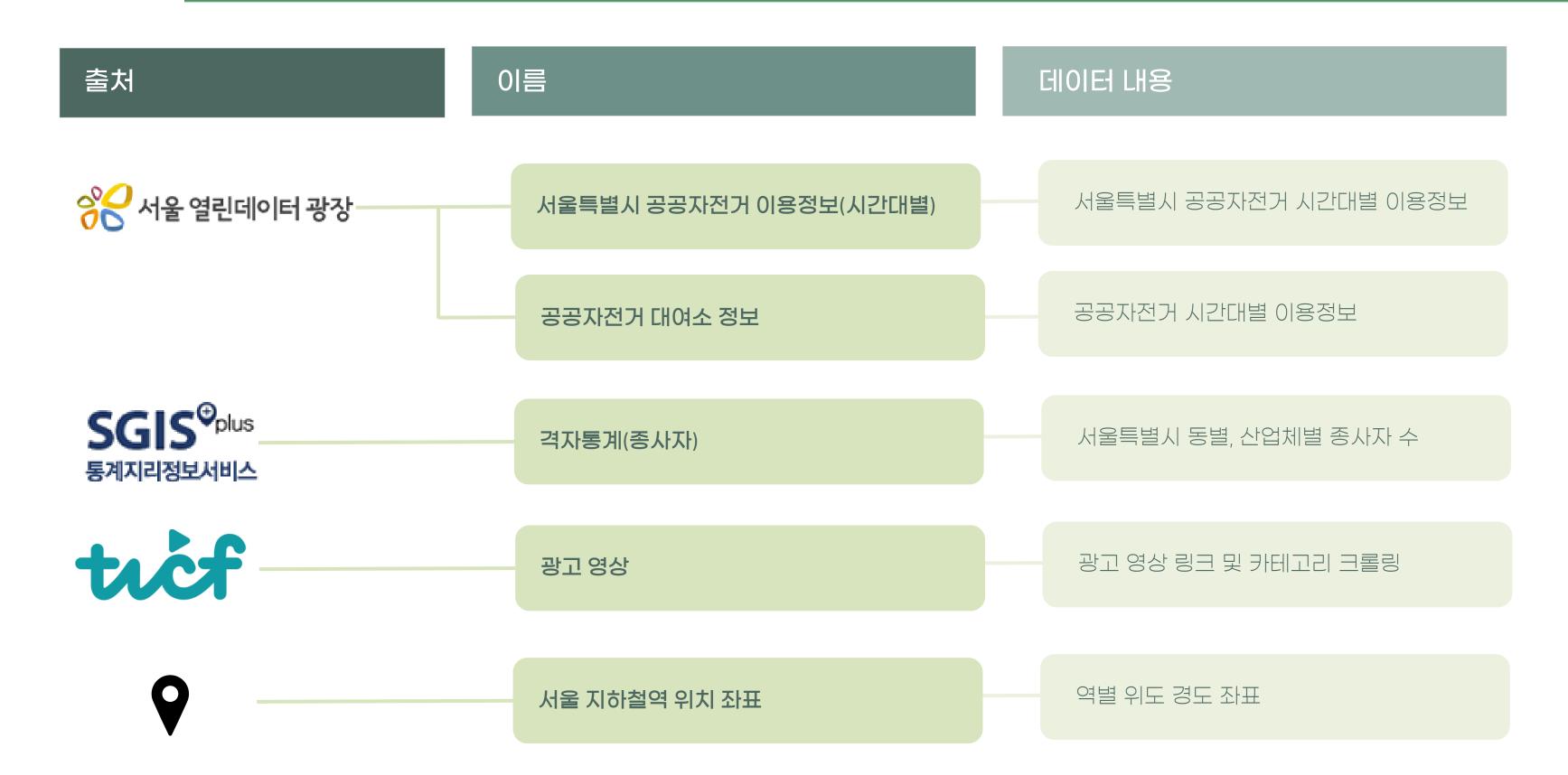
이. 데이터 이해

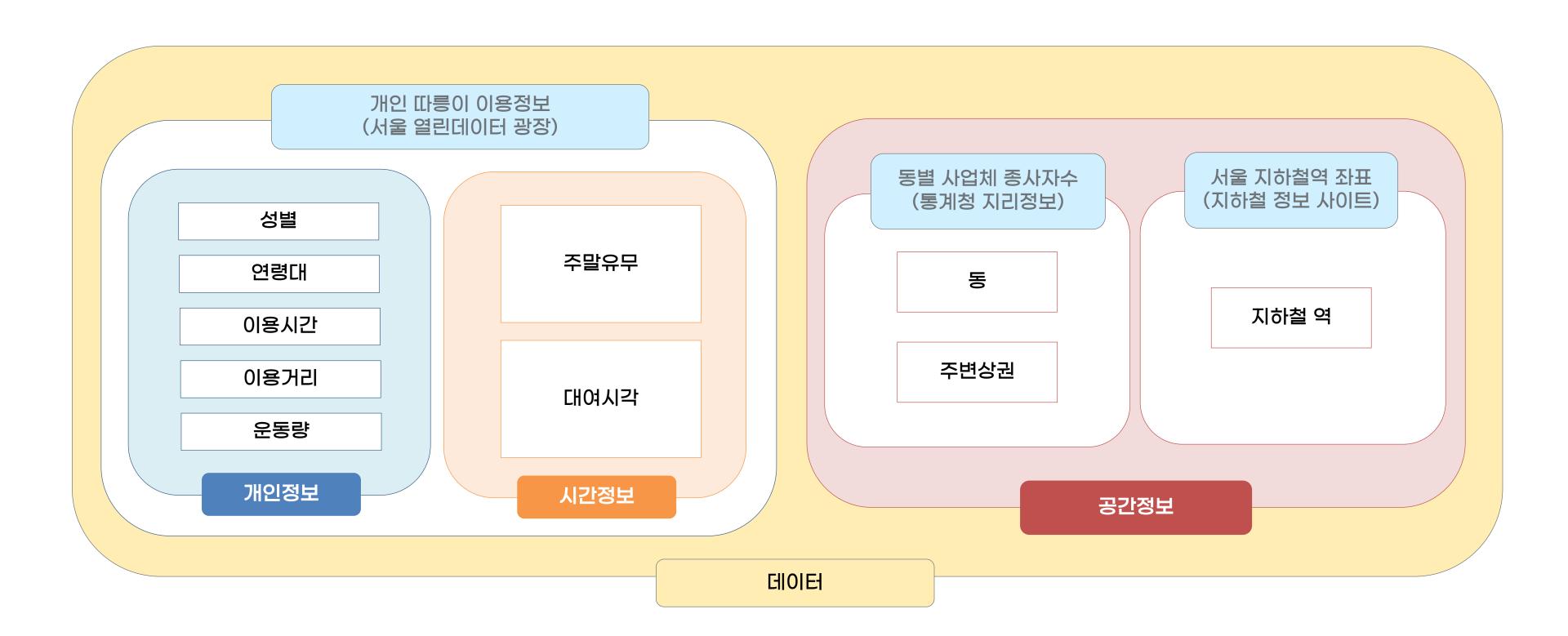
02. 데이터 전처리

03. 데이터 시각화









### 2. 돈 저 데이터 전처리

### 데이터

### 개인 따름이 이용 정보

	대여일자	대여시간	대여소번호	대여소명	대여구분코드	성별	연령대코드	이용건수	운동량	탄소량	이동거리(M)	이용시간(분)
0	2022-06-01	0	742	742. 등촌역 5번 출구 뒤	정기권	NaN	~10대	1	48.30	0.46	1999.70	10
1	2022-06-01	0	2059	2059. 보라매공원 정문	정기권	NaN	~10대	1	27.47	0.23	991.00	9
2	2022-06-01	0	1349	1349. 윌계2교 버스정류장 앞	정기권	NaN	~10대	1	20.63	0.20	868.17	3
3	2022-06-01	0	750	750. 연의근린공원 건너편	정기권	NaN	~10대	1	13.97	0.13	560.00	4
4	2022-06-01	0	755	755. 목동1단지아파트 상가 앞	정기권	NaN	~10┖∦	1	12.15	0.14	590.00	3

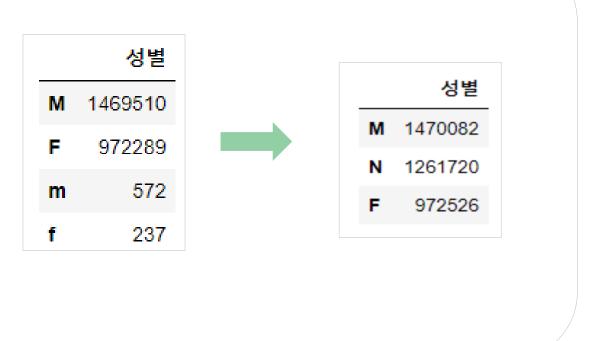
출처 : 서울 열린데이터 광장

#### 1

### 결측치 및 데이터 형식 처리

#### 결측치 처리 대여일자 대여시간 대여소번호 대여소명 대여구분코드 모든 결측치 'N'으로 대체 1261720 성별 연령대코드 이용건수 0 운동량 탄소량 이동거리(M) 이용시간(분) 0

### 데이터 형식 통일



### 2. 돈 저 데이터 전처리

이상치 제거 제거 전 제거후 이용시간 Boxplot 이동거리 & 대여소번호 Boxplot 이용시간 Boxplot 이동거리 & 대여소번호: 이동거리 & 대여소번호 Boxplot 200000 1750 -Q3 + IQR \* 1.5 보다 큰 값 1500 1250 이용시간: 대부분의 데이터가 10분대로 250 -180분 이상을 기준으로 제거 ... 이동거리(M) 대여소번호 이용시간(분) 이용시간(분) 이동거리(M) 대여소번호 3 필요한 컬럼 추가

### 대여 일자 → 요일

요일 금 578699 수 575342 화 548068 토 545882 월 501433 목 479471 일 475433

### 대여 일자 → 주말

주말 평일 2683013 주말 1021315

### 2. 돈 지 데이터 전처리

4 데이터 병합

### 1) 지하철 위.경도 데이터

	line	name	code	lat	Ing
0	01호선	녹양	1908.0	37.759380	127.042292
1	01호선	남영	1002.0	37.541021	126.971300
2	01호선	용산	1003.0	37.529849	126.964561

지하철역 위.경도 데이터를 이용하여 대여소와 가까운 지하철 역을 Mapping

### 2) 동별 사업체 종사자 수

	집계구번호	산업분류	종사자수
0	1114070010026	건설업	0.0
1	1115057010032	부동산업 및 임대업	0.0
2	1124055010013	전문, 과학 및 기술 서비스업	0.0

각 대여소가 위치한 동과 해당 동의 종사자 수가 최다인 산업분류와 해당 사업체 종사자 수를 Mapping

### 3) 따름이 대여소 정보

	대여소\n번호	보관소(대여소)명	자치구	상세주소	위도	경도
0	102.0	망원역 1번출구 앞	마포구	서울특별시 마포구 월드컵로 72	37.555649	126.910629
1	103.0	망원역 2번출구 앞	마포구	서울특별시 마포구 월드컵로 79	37.554951	126.910835
2	104.0	합정역 1번출구 앞	마포구	서울특별시 마포구 양화로 59	37.550629	126.914986

위도	경도	설치\n시기	거치\n대수	거치\n대수.1	운영\n방식
37.555649	126.910629	2022-05-03	NaN	15.0	QR
37.554951	126.910835	2022-05-03	NaN	14.0	QR
37.550629	126.914986	2022-05-12	NaN	13.0	QR

### 데이터

### 1~4 완료 후 데이터

대여역	일자	대여시간	대여소번호	대여소명	대여구분코드	성별	연령대코드	이용건수	운동량	탄소량	이동거리(M)	이용시간(분) 주말	요일	대여소근처역	소재지 위되	경도	동	산업종류
0 2022-06	6-01	0	742	742. 등촌역 5번 출구 뒤	정기권	N	10대	1	48.30	0.46	1999.70	10 평일	수	등촌	양천구 37.55073	126.864578	목3동	사업시설관리 및 사업지원 서비스업
1 2022-06	6-01	0	2059	2059. 보라매공원 정문	정기권	N	10대	1	27.47	0.23	991.00	9 평일	수	보라매	동작구 37.49707	126.917793	신대방2동	보건업 및 사회복지 서비스업
<b>2</b> 2022-06	6-01	0	1349	1349. 월계2교 버스정류장 앞	정기권	N	10대	1	20.63	0.20	868.17	3 평일	수	광운대	성북구 37.62382	127.050201	번동	금융 및 보험업
3 2022-06	6-01	0	750	750. 연의근린공원 건너편	정기권	N	10대	1	13.97	0.13	560.00	4 평일	수	신정네거리	양천구 37.51215	7 126.835625	신정3동	운수업
4 2022-06	6-01	0	755	755. 목동1단지아파트 상가 앞	정기권	N	10대	1	12.15	0.14	590.00	3 평일	수	신목동	양천구 37.53786	126.881409	목5동	보건업 및 사회복지 서비스업

5 수치형 데이터 카테고리화

### 대여 시각

#### 대여시각 분포를 기준으로

✓ 0: 밤(20시 ~ 06시), 새벽(00 ~ 06시))✓ 출근: 7시~9시

✓ 1 : 출근, 퇴근
 ✓ 2 : 낮(10시 ~ 16시)

✓ 불근, 가시~9시
✓ 퇴근: 17시~19시



### 이용시간, 이동거리, 운동량

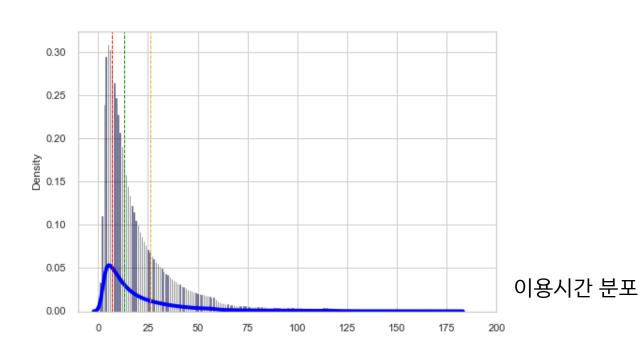
#### 각변수의 4분위를 기준으로

√ 0:0~25%

√ 1 : 26~50%

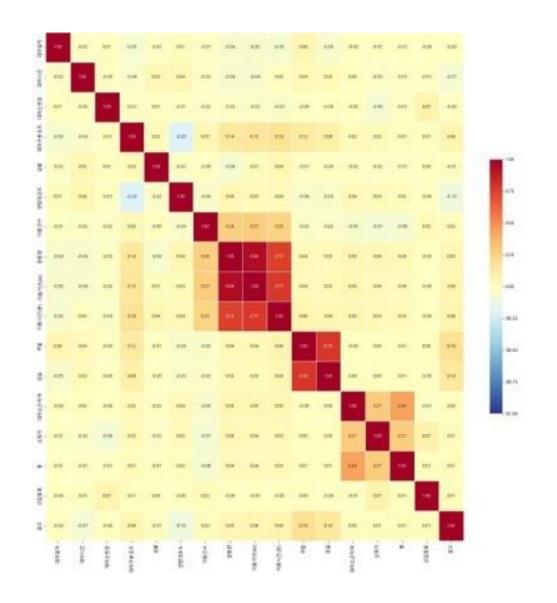
✓ 2:51~75%

✓ 3:76%~100%

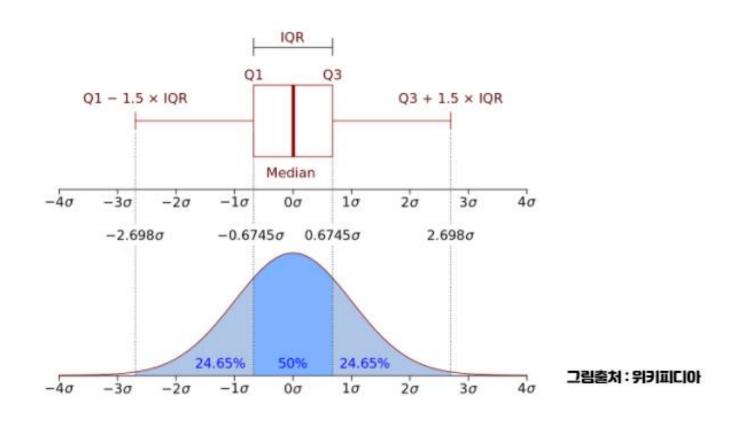


6 최종 데이터의 변수 중요도

### Feature Importance



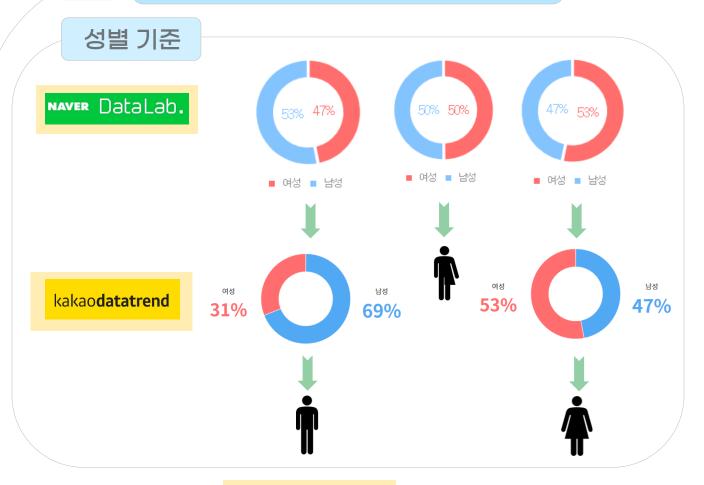
### Robust Scaler



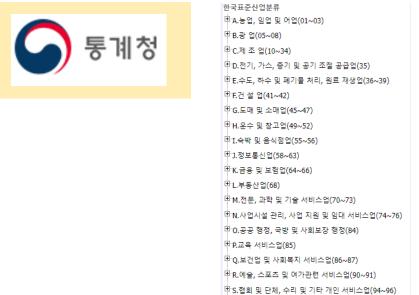
$$x' = \frac{x - median(x'')}{(Q_3 - Q_1)}$$

### 2. 돈 저 데이터 전처리

### 7 최종 데이터에 광고 라벨링

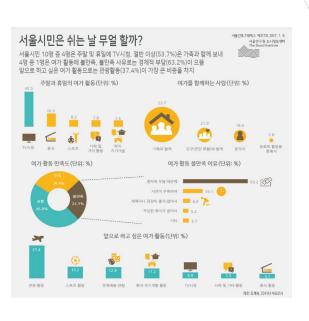


### 산업종류 기준



#### 평일, 주말 기준







### 대여시각 기준

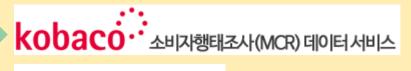


□ U.국제 및 외국기관(99)

田T.가구 내 고용활동 및 달리 분류되지 않은 자가 소비 생신

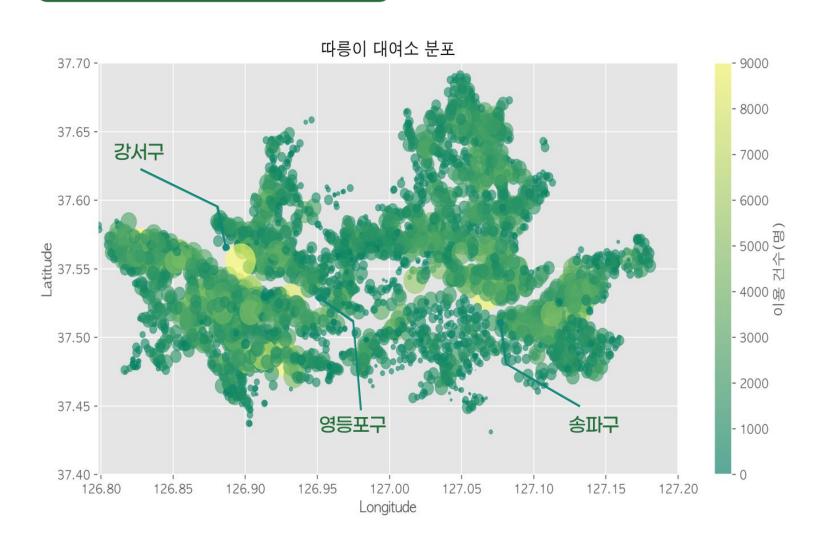


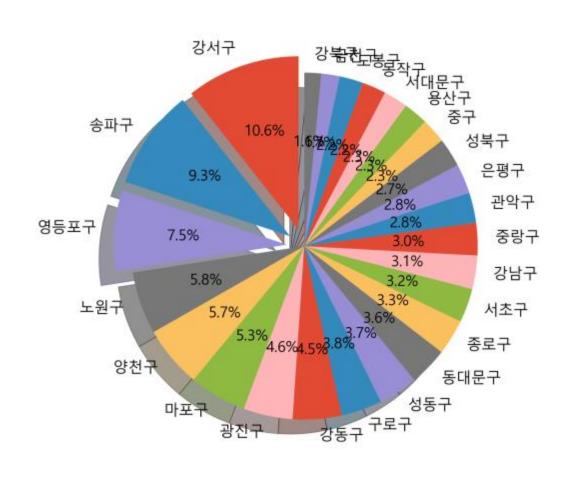
### **DMCREPORT**





### 따릉이 대여소 위치







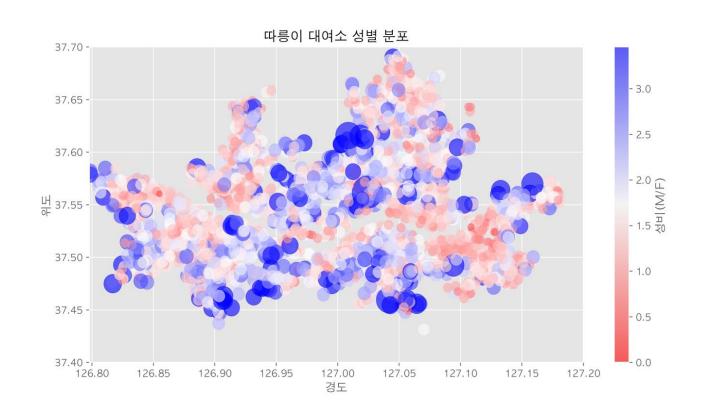
강서, 송파구에 대여소 가장 밀집되고 영등포구, 노원구가 뒤를 이음



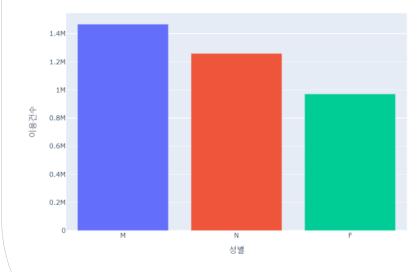
도봉, 강북, 성북, 중랑, 동작, 관악, 금천구 상대적으로 밀집도 낮음

### 2 두 수 데이터시각화-개인정보

### 1. 성별



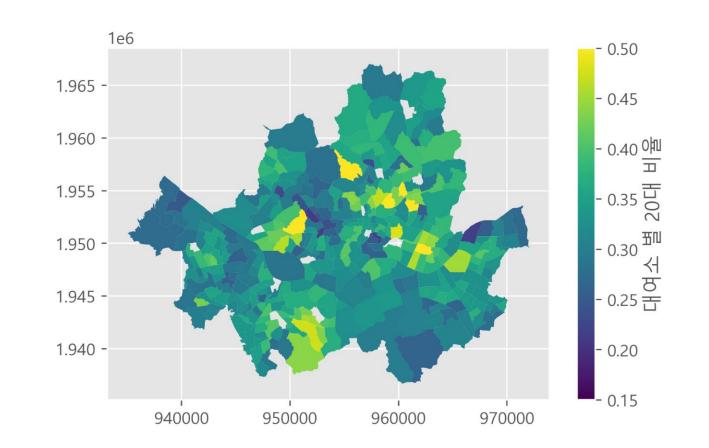
#### 성별에 따른 이용 건수

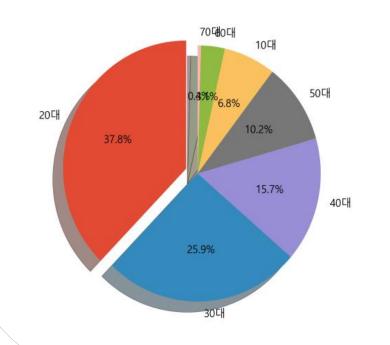




여성보다 남성 따름이 이용자가 많음

### 2. 연령대



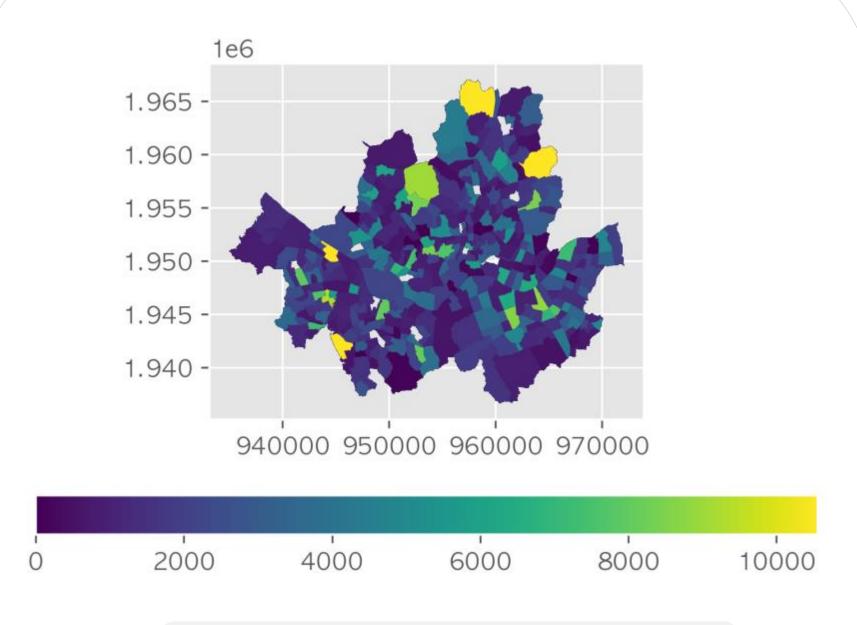




따름이의 가장 큰 이용자인 2030 세대 비중이 73.7%

### 2. 든 D A 데이터 시각화-개인정보

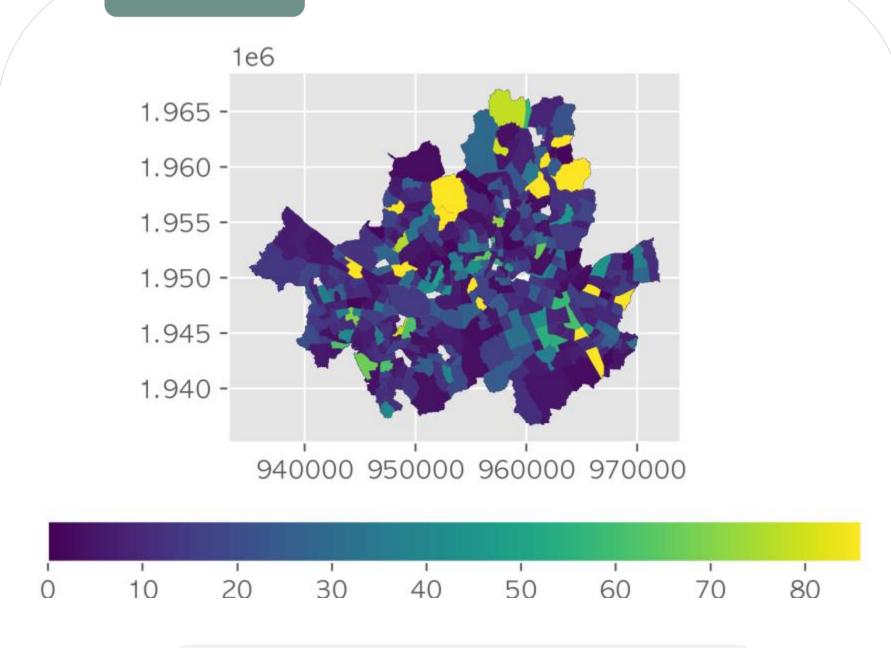






옥수동, 압구정동에서 이동거리가 가장 길며 이후 동작동, 망원동 대여소가 뒤를 이음

### 4. 이용시간



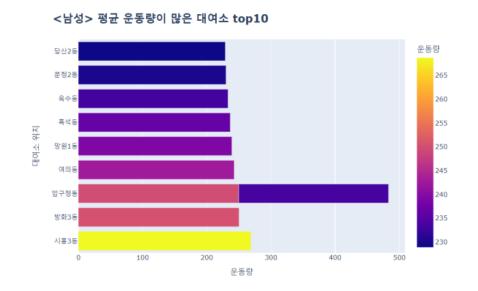


망원동, 동작동에서의 이용시간이 가장 길며 이후 압구정동, 옥수동의 대여소가 뒤를 이음

### 

### 5. 운동량 [남]

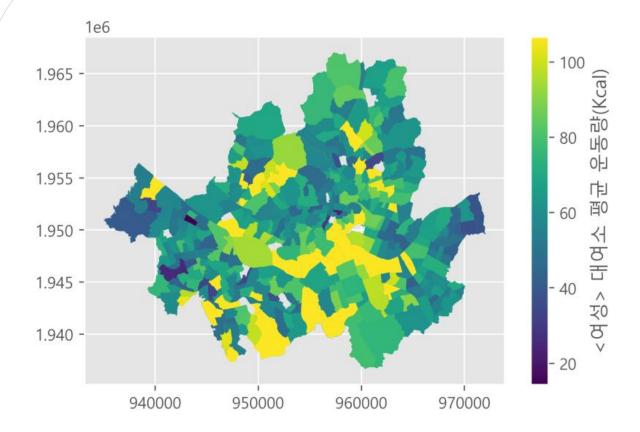




### 

<남성 대여자>의 운동량이 많은 대여소는 당산동, 망원동 등 한강 주변 지역이 높은 순위

### 5. 운동량 [여]

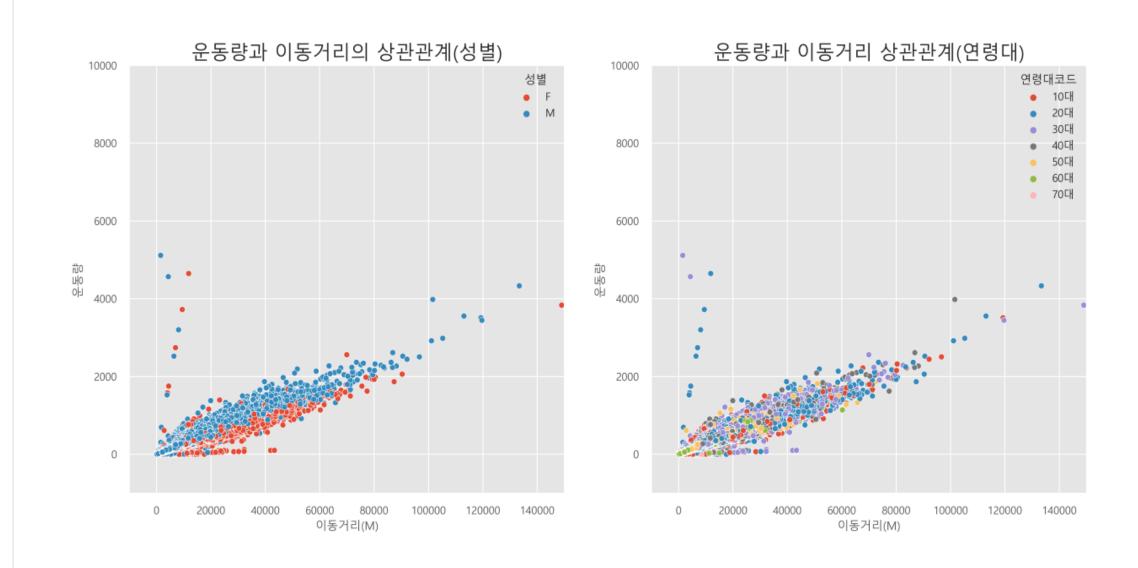






<여성 대여자>의 운동량이 많은 대여소는 고덕동, 시흥동 등 높은 순위

### 6. 상관관계- 운동량, 이동거리



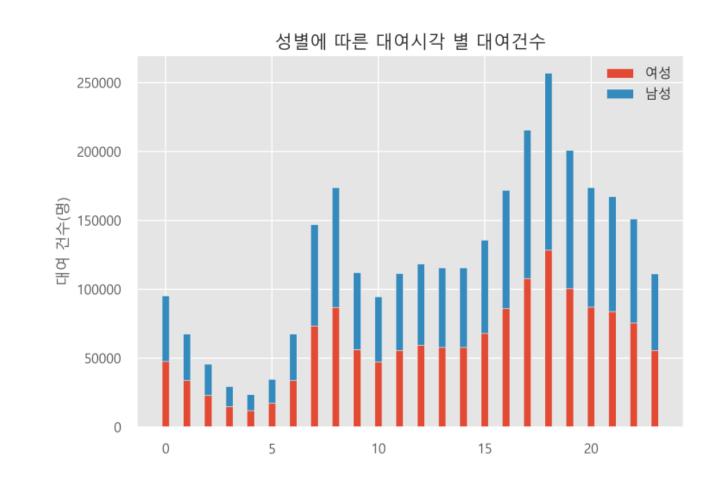


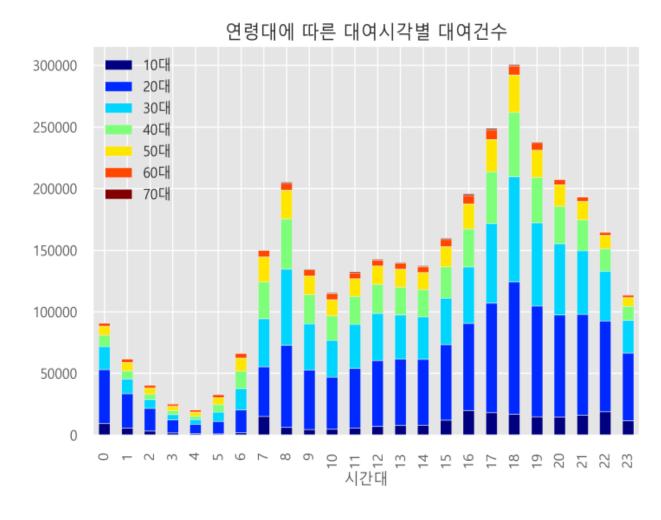
2030 세대가 운동량과 이동거리가 높은 성향을 보임



6070대는 주로 짧은 거리를 이용함

### 7. 상관관계-성별, 연령대





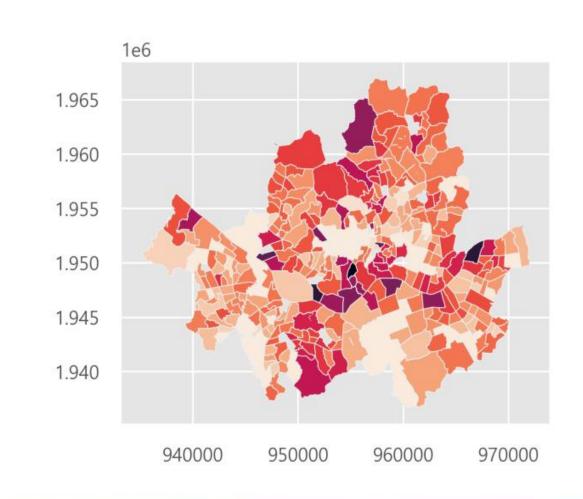


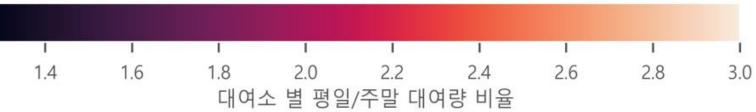
2030 세대가 운동량과 이동거리가 높은 성향을 보임

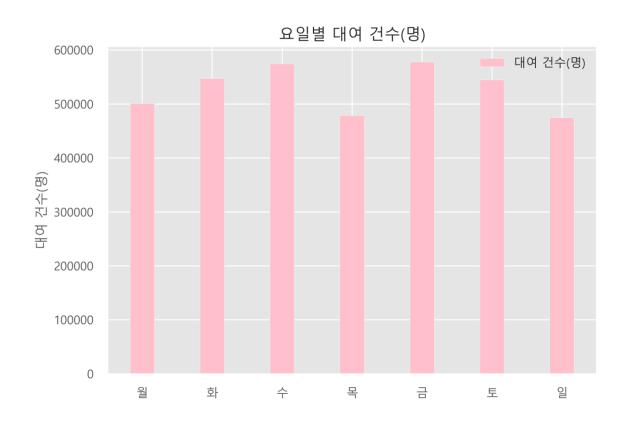


6070대는 주로 짧은 거리를 이용함

### 8. 요일







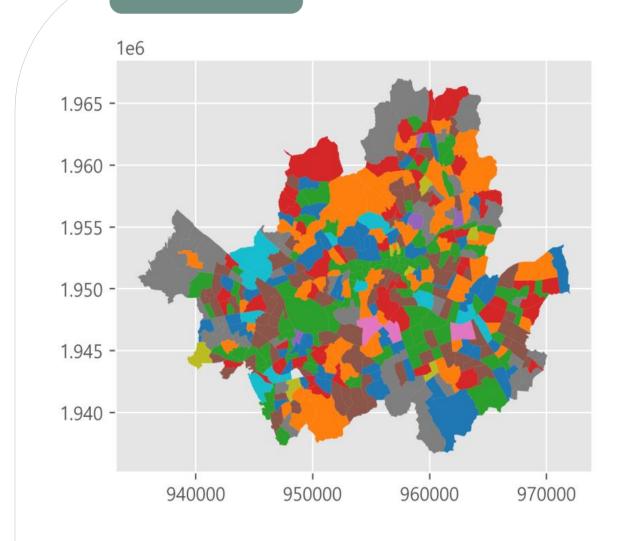


주말과 평일의 이용 평균이 큰 차이가 없으나 지역적으로 차이가 보임



금요일에 가장 많이 이용하며, 목요일에 가장 적게 이용됨

#### 9. 공간정보



- 건설업
- 공공행정, 국방 및 사회보장 행정
- 교육 서비스업
- 금융 및 보험업
- 도매 및 소매업
- 보건업 및 사회복지 서비스업
- 부동산업 및 임대업
- 사업시설관리 및 사업지원 서비스업
- 숙박 및 음식점업
- 예술, 스포츠 및 여가관련 서비스업
- 운수업
- 전문, 과학 및 기술 서비스업
- 제조업
- 출판, 영상, 방송통신 및 정보서비스업
- 하수 · 폐기물 처리, 원료재생 및 환경복원업

#### 읍면동코드 종사자수

		산업종류	동이름	구이름
5.0	1123068	건설업	개포1동	강남구
3 15.0	1123068	공공행정, 국방 및 사회보장 행정		
66.0	1123068	교육 서비스업		
6.0	1123068	금융 및 보험업		
37.0	1123068	도매 및 소매업		



강북 지역에 교육 서비스업이 밀집되어 있으며 강남 지역에 금융,보험,도매업이 밀집되어 있으며 서울 외각 지역에 운수업이 밀집되어 있는 경향을 보임

### 10. 대여소 주변역 정보

소재지	대여소근처역	대여소명	
강서구	가양	1158. 가양역 8번출구	6079
		3798. 그랜드아이파크아파트 앞	4523
		3780. 등촌서 <mark>광아파</mark> 트	3911
		2704.홈플러스(강서점)	3509
		1173. 강서구청사거리(SH타워)	3484
		2708.강서한강자이타워 앞	3448
		1157. 강서구청	3351
		1182. KBS 스포츠윌드	3039
		1120. 대림경동아파트앞	2338
		1185. 등촌9단지	1760
		2742.강서구민올림픽체육센터	1711
		1174. 강서구청사거리(부민병원)	1677
		2709.강변아파트 310동 앞	1437
		5065. 허준박물관삼거리	1211
		5080. 가양역 1번출구	591
		5082. 가양역 7번출구	4

#### 대여소명

소재지	대여소근처역	
강서구	가양	1120. 대림경동아파트앞
	개화	1200. 개화광역환승센터
	개화산	5053. 신마곡벽산블루밍메트로오피스텔앞
	공항시장	1109. 공항시 <u>장역 4</u> 번출구
	까치산	1146. 곰달래사거리
	등촌	1170. 강서구보건소
	마곡	2739.수명산파크1단지교차로
	마곡나루	2727.서울식물원 잔디마당 앞
	목동	1140. 목동사거리 버스정류장
	발산	1153. 발산역 1번, 9번 인근 대여소
	방화	1103. 방화역 4번출구앞
	송정	1150. 송정역 1번출구
	신방화	1163. 방화동강서기동대앞
	신정	5059. 화곡2동주민센터
	양천향교	1160. 양천향교역 7번출구앞
	염창	5067. 주원 리젠트 앞
	우장산	1126. 우장산역 1번출구옆(우장산아이파크105동앞)
	증미	1122. 황금내근린공원
	화곡	2711.가로공원공영주차장 1번 출구

03

### 모델링

01. 1차 모델링

02. 2차 모델링





### 3.모델링 차멜링

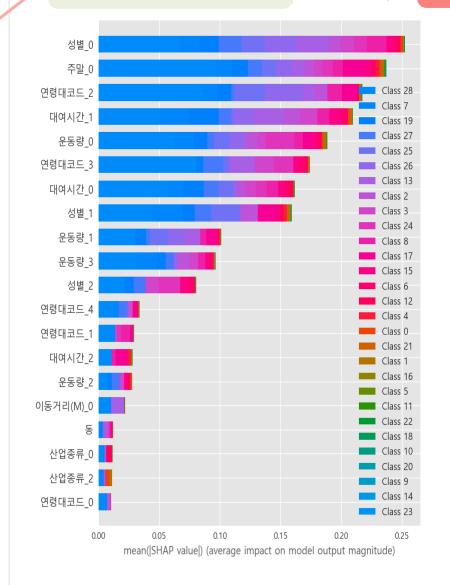
### AutoML

■ Pycaret 의 compare\_models()로 각각의 머신러닝 모델에 적용

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	II (Sec)
dt	Decision Tree Classifier	0.9550	0.0977	0.7846	0.9555	0.9551	0.9177	0.9177	0.1440
et	Extra Trees Classifier	0.9544	0.0996	0./192	0.9539	0.9540	0.9165	0.9165	2.1050
rf	Random Forest Classifier	0.9541	0.0999	0.6622	0.9532	0.9532	0.9158	0.9158	1.7560
nb	Naive Bayes	0.9537	0.0998	0.8672	0.9629	0.9539	0.9150	0.9154	0.1780
gbc	Gradient Boosting Classifier	0.9485	0.0992	0.5861	0.9451	0.9461	0.9055	0.9056	104.4000
lda	Linear Discriminant Analysis	0.8795	0.0975	0.8069	0.9258	0.8975	0.7960	0.8016	0.9100
ridge	Ridge Classifier	0.7090	0.0000	0.0890	0.5763	0.6105	0.2069	0.3246	0.1000
lr	Logistic Regression	0.6648	0.0492	0.0470	0.4419	0.5309	0.0000	0.0000	3.8850
dummy	Dummy Classifier	0.6648	0.0500	0.0470	0.4419	0.5309	0.0000	0.0000	0.0380
ada	Ada Boost Classifier	0.6628	0.0658	0.0449	0.4415	0.5300	-0.0020	-0.0136	1.9450
knn	K Neighbors Classifier	0.6442	0.0692	0.1737	0.5794	0.6004	0.2241	0.2349	0.6340
lightgbm	Light Gradient Boosting Machine	0.5553	0.0521	0.0939	0.5365	0.5283	0.1097	0.1142	3.9590
svm	SVM - Linear Kernel	0.4739	0.0000	0.0459	0.3461	0.3729	-0.0001	-0.0001	34.4260

- Decision Tree Classifier 가 가장 높은 성능 지표를 가짐
  - 하나의 트리 모델을 만들어 예측할 경우, 분기가 명확하여 좋은 설명력을 갖기 때문에 해당 모델을 사용함. 동시에 해당 모델로 Feature Importance를 확인함.

### **Decision Tree**



#### 과적합 위험

◀ Interaction Plot : 각 특성 간의 관계 파악함

$$\Phi_{i,j} = \sum_{S \subseteq N \setminus \{i,j\}} \frac{|S|!(M-|S|-2)!}{2(M-1)!} \nabla_{ij}(S),$$

when  $i \neq j$ , and

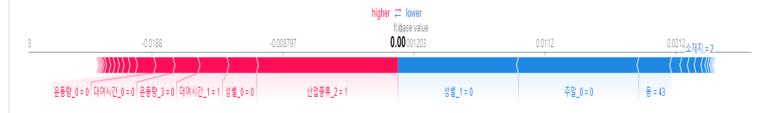
$$\nabla_{ij}(S) = f_x(S \cup \{i, j\}) - f_x(S \cup \{i\}) - f_x(S \cup \{j\}) + f_x(S)$$
$$= f_x(S \cup \{i, j\}) - f_x(S \cup \{j\}) - [f_x(S \cup \{i\}) - f_x(S)].$$

$$\Phi_{i,i} = \phi_i - \sum_{j 
eq i} \Phi_{i,j}$$

#### ☑ Shapely Interation Value 라?

Shapely value는 모두 동일한 단위(모델 예측 단위) 를 가지고 있어 클러스터링 기법으로 사용하기 적절함.

따라서, 모든 관측치를 '영향력(기여도)' 에 따라 클러스터링을 진행함.



•

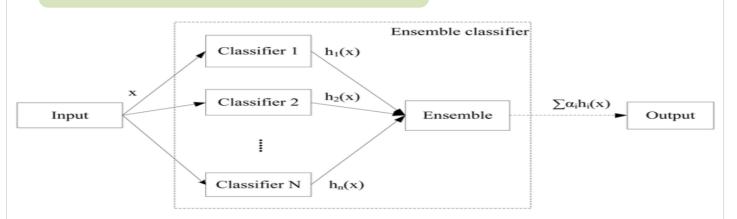
긍정적인 영향 : 대여시간, 운동량,

성별, 산업종류

부정적인 영향: 성별, 주말, 동

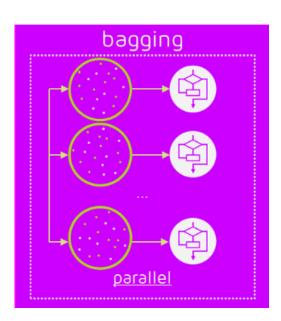
### 3. 모델링 2차모델링-앙상블, (bagging) Random Forest

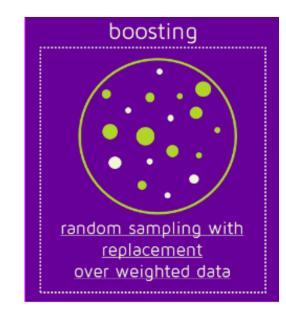
#### 앙상블 기법 Ensemble



#### ☑ Ensemble 기법 이란?

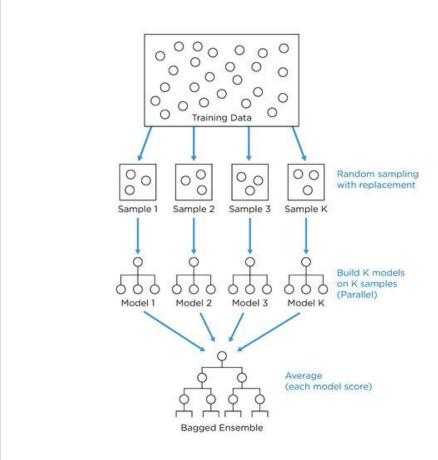
여러 개의 결정 트리를 경합하여 하나의 트리 보다 나은 성능을 내며 과적합의 문제점을 보완해주는 학습 방법





- 배강 (bagging) : 샘플을 여러 번 뽑아 각 모델을 학습, 각 모델은 서로 독립적
- 부스팅(boosting)
  - : 첫 모델값의 예측 결과에 따라 가중치 부여, 다음 모델에 영향을 미침

#### **Random Forest**



#### ☑ Random Forest 기법 이란?

무작위 샘플링을 사용하는 Random Forest 기법으로

무작위로 뽑은 데이터로 여러 개의 작은 트리를 만들고, 그 트리를 결합하여 과적합 문제를 방지함

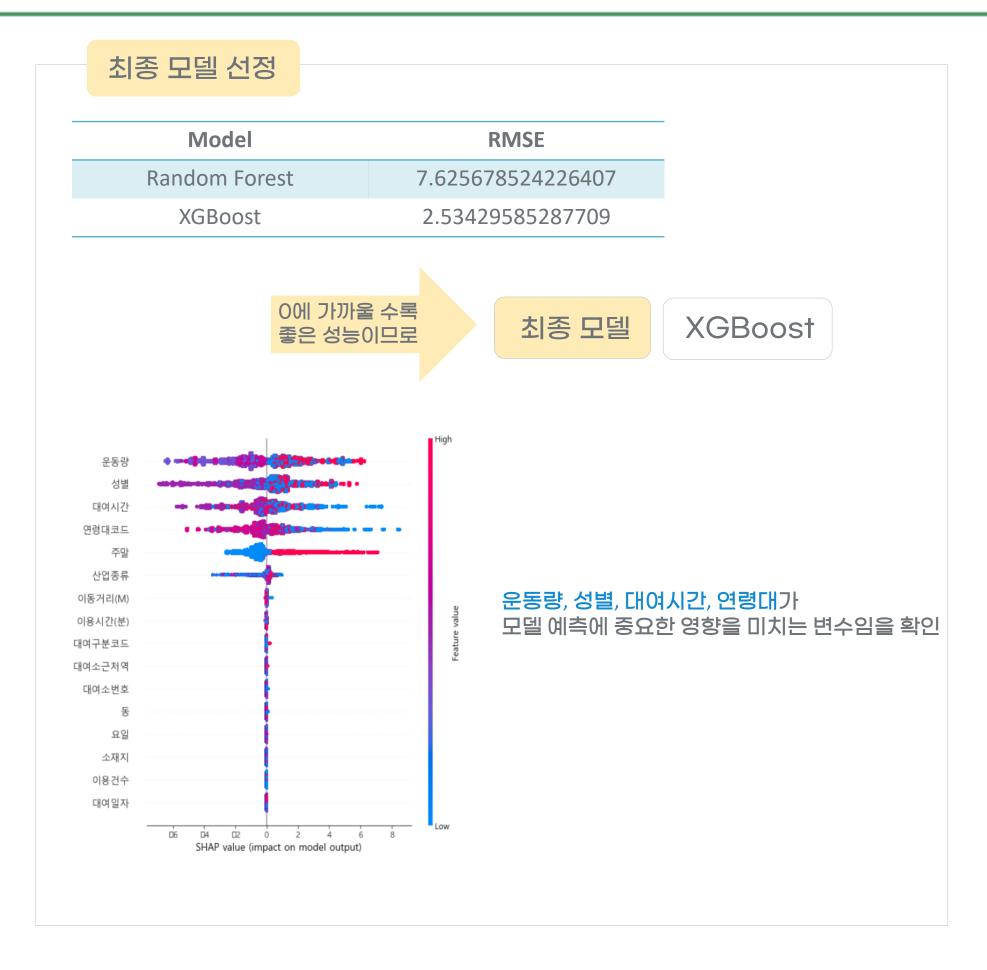
#### Random Forest 결과

	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
Fold							
0	0.9717	0.0000	0.6031	0.9692	0.9702	0.9272	0.9272
1	0.9687	0.0000	0.5374	0.9662	0.9667	0.9192	0.9193
2	0.9704	0.0000	0.5872	0.9681	0.9688	0.9237	0.9238
3	0.9708	0.0000	0.5965	0.9690	0.9695	0.9249	0.9250
4	0.9669	0.9649	0.5423	0.9649	0.9654	0.9148	0.9148
5	0.9696	0.0000	0.5512	0.9688	0.9685	0.9218	0.9219
6	0.9696	0.0000	0.5518	0.9686	0.9682	0.9217	0.9218
7	0.9716	0.0000	0.5677	0.9699	0.9701	0.9268	0.9269
8	0.9714	0.0000	0.5681	0.9684	0.9696	0.9264	0.9265
9	0.9674	0.0000	0.6286	0.9657	0.9661	0.9160	0.9160
Mean	0.9698	0.0965	0.5734	0.9679	0.9683	0.9223	0.9223
Std	0.0016	0.2895	0.0282	0.0016	0.0016	0.0042	0.0042

- ✓ Precision Recall Trade-Off 관계
  - Accuarcy, Precision 은 높게
  - Recall 은 낮게
- ✓ MCC 값 1에 근접
  - 무작위 추측이 아닌 잘 학습되었음을 판단 가능
- ✓ AUC
  - 분류 모델의 구분 성능이 낮다고 판단

### 3. 모델링 2차모델링-(boosting) XGBoost, 최종모델선정

#### **XGBoost** ✓ XGBoost 기법 이란? Extreme Gradient Boosting 의 약자로 기존의 Gradient Boosting 모델 보다 빠르고, 과적합 이슈를 방지할 수 있으며, Customizing에 용이하여 널리 사용되는 기법 XGBoost 결과 정확도: 0.9697 운동량 정밀도: 0.9690 재현율: 0.9697 연령대코드 F1: 0.9646 산업종류 긍정적인 영향: 대여시간, 운동량, 연령대 이동거리(M) 부정적인 영향: 주말 이용시간(분) 대여구분코드 대여소근처역 대여소번호 요일 소재지 이용건수 대여일자 025 0.50 0.75 1.00 1.25 1.50 1.75 mean(|SHAP value|) (average impact on model output magnitude) 25.01 26.01 27.01 28.01 29**29.31** 30.01 31.01 대여시간 = 0 운동량 = 3 연령대코드 = 0 주말 = 0



04

### 결론

01. 활용방안

02. 기대효과



### 4. 결론 활용방안

### 서비스 실현 계획

1. 따름이 어플에 광고 추천



- © 따름이 어플의 개인정보와 대여 유형을 이용하여 사용자 맞춤 추천 광고가 가능
  - □ 마릉이 앱에 등록된 이용자는 332만명으로 큰 광고 효과 기대

### 2. 따름이 대여소에 설치



♀ 대여소 별 분석한 내용을 토대로 대여소 유형(이용자 특성, 상권)에 알맞는 맞춤 광고 가능

> □ 대여소가 주로 역 근처에 많이 위치하므로 자연스레 이용자 외 사람에게도 노출

◎ 현재 도입예정인 따름이 광고의 모델을 위와 같이 제안하고자 함

### 4. 결론 활용방안

### 서비스 실현 예시

### 1. 따릉이 어플에 광고 추천

성별	연령대	대여 시각	운동량	이동 거리	이용 시간	주말	산업 종류
여성	20EH	20:00	145.9	5670	34	평일	도매





### 2. 따릉이 대여소에 설치

성별	연령대	대여 시각	운동량	이동 거리	이용 시간	주말	산업 종류
여성	30EH	10:00	325.4	8000	60	명인	금융





### 4. 결론 기대효과



# 감사합니다

