

따릉이 이용 요인분석 및 대여-반납건수 예측

범주형 자료분석팀

강지원 이정현 박서현 최재영 이동욱

목차 INDEX

I. 공모 배경

가) 주제 선정 배경

나) 분석 방향

II. 활용 데이터 정의

가) 데이터 수집 및 전처리

① 대여소 장소 선정 및 기간 선정

나) 데이터 수집

① 따릉이 이용자 데이터

② 따릉이 대여소 데이터

③ 자전거 도로 데이터

④ 축제 데이터

⑤ 날씨 데이터

- 강수량 변수

- 평균풍속 변수

- 상대습도 변수

- 일교차 변수

⑥ 미세먼지 데이터

⑦ 장소와 날짜에 대한 파생변수

III. 데이터 처리 방안 및 활용 분석 기법

가) PCA를 활용한 대여소 지수 생성

① 한량지수(여가 지수)

② 노예지수(출퇴근 지수)

나) 네트워크 분석을 통한 대여소 클러스터링

① 네트워크 분석과 시각화

② 네트워크 분석과 중심성 산출

다) 최종 데이터

IV. 분석 결과

가) 평가 지표 선정

나) 해석 모델

① Month 변수

② Day 변수

③ Index 변수

④ Bike road 변수

⑤ 기온/강수량/습도 변수

⑥ 미세먼지 변수

다) 예측 모델

① Prophet

② Random Forest

③ XGboost

④ Entity Embedding

라) 최종 평가

V. 서비스 활용 방안

가) 날씨 정보로 업데이트 되는 따릉이 대여소 정보

나) 따릉이 대여소 추가설치에 대한 시민 요구 수렴

다) 따릉이 이용으로 인한 환경적 비용 시각화

VI. 서비스 기대효과

가) 따릉이의 효율적인 배치와 사회적 파급 효과

나) 공유경제의 환경보호 역할 가시화

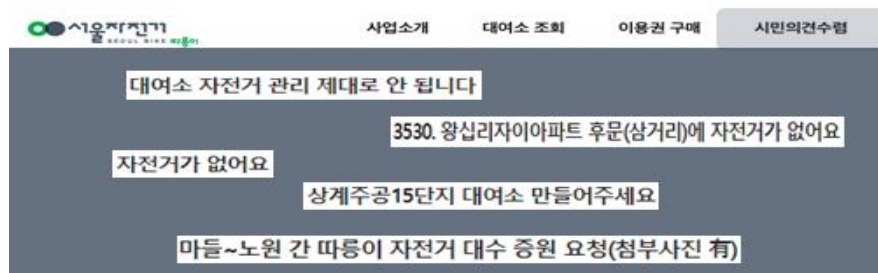
I. 공모 배경

가) 주제 선정 배경

자전거 수가 인구 수 보다 많은 네덜란드는 공영자전거 정책과 운용에 있어서 가장 선진화 되어 있는 국가 중 하나이다. “에코 시티”라는 이름에 걸맞게 네덜란드 수도 암스테르담의 경우, 자전거의 도심교통 분담률은 60%를 웃도는 수준이다. 이는 2015 년 기준 네덜란드의 이산화탄소 배출량이 아시아, 유럽 전역과 비교해 보았을 때 0.46%로 매우 낮을 수밖에 없었던 가장 큰 이유라고 할 수 있다. 실제 자전거 이용만으로도 1 킬로미터 당 NOx 의 배출량을 65 퍼센트 가량 줄일 수 있고 1 킬로미터 당 자전거 이용에 따르는 사회적 비용은 0.68 유로로, 버스와 자가용의 경우 각각 29 유로, 37 유로인 것을 고려했을 때 자전거의 그것 비해 최대 50 배까지 차이가 난다.

한국도 마찬가지로 2008 년 창원시의 ‘누비자’로 시작되어, 대전광역시의 ‘타슈’, 순천시의 ‘온누리’ 그리고 서울시의 ‘따릉이’와 광주광역시의 “타랑께”까지 국내 많은 도시들이 자전거 공공 대여 서비스를 제공하기 시작했다. 특히 ‘따릉이’의 경우 엄청난 성공을 거두었는데, 등장한지 불과 2 년 만에 인프라는 10 배, 회원은 20 배가 늘었다. 정확한 수치로는 2015 년 150 개의 대여소와 2000 대의 따릉이로 시작하여, 2019 년 현재 1290 개의 대여소와 20000 대의 따릉이를 보유 중이고 회원은 2015 년 말 3 만 4 천명에서 시작해 현재 60 만 명의 회원이 따릉이를 사용 중이다. 이에 서울시는 멈추지 않고 따릉이를 3 만대로 확장하겠다고 발표했다. 이는 서울시가 공영 자전거 대여 시스템에 큰 의미를 부여하고 있으며, 그에 따른 성공적인 결과를 기대하고 있다고 볼 수 있다.

그러나, 과연 실제로 서울특별시의 따릉이는 잘 운영되고 있을까?



▲ 그림1. 서울자전거 공식홈페이지, 시민의견수렴 페이지

현재 따릉이의 가장 큰 문제점은 따릉이의 수요가 공급보다 훨씬 많아, 원활하고 효율적인 따릉이 배치 및 운영이 이루어지지 않고 있다는 것이었다. 즉, 단순히 자전거의 전체 개수가 아닌 각 대여소별 적정 대여 건수에 따른 따릉이의 순환이 원활하지 않다는 것이다. 따릉이는 대여한 곳이 아닌 곳에서도 반납이 가능한 만큼, 대여 수요가 많은 인기 대여소의 경우 따릉이가 다시 공급되기까지는 많은 시간이 걸리는 것이다. 예를 들어, 여가를 즐길 수 있는 한강 공원 주변의 대여소인 “여의나루” 혹은 “뚝섬한강공원”의 대여소 혹은 회사 밀집 지역인 “광화문”, “을지로” 혹은 “여의도”의 대여소를 보면 확실한 문제점이 드러난다. 실제로 여의도 부근에 많은 대여소가 있음에도 불구하고, 퇴근시간에 따릉이를 이용하고자 하는 사람들은 많지만 부족한 자전거로 인해

다들 다른 교통수단으로 발걸음을 옮긴다.



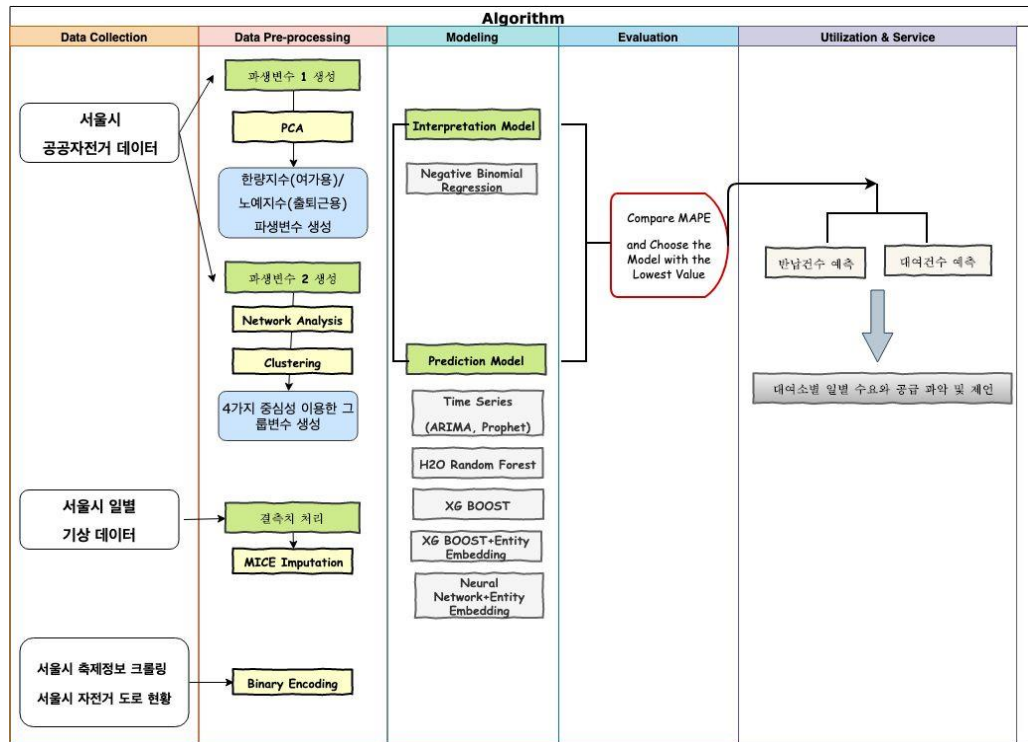
▲ 그림2. 따릉이 어플, 시간과 공간에 따른 잔여 자전거 수 편차

나) 분석 방향

따라서 본 팀은 효율적인 자전거 배치제고와 추가 대여소 설치를 개선하기 위해 우선적으로 현재 시행되고 있는 시스템의 현위치를 파악하고자 한다. 이용자들이 자전거를 이용하고 싶을 때 이용할 수 있도록 적재적소에 따릉이가 배치되었는지 궁금하였다. 우리는 서울시의 따릉이 대여소별로 일별 수요를 예측하여 실제 대여건수와 비교한 후, 해당 대여소에 자전거의 배치가 더 필요한 곳인지 혹은 과한 공급인지 파악하고자 한다.

날짜 + 대여소	날짜 INFO	대여소 INFO	대여건수
2019-01-01-101			14
2019-01-02-101			45
.	.	.	.
2019-03-30-993			3
2019-03-31-993			14

▲ 그림3. 데이터 구성표



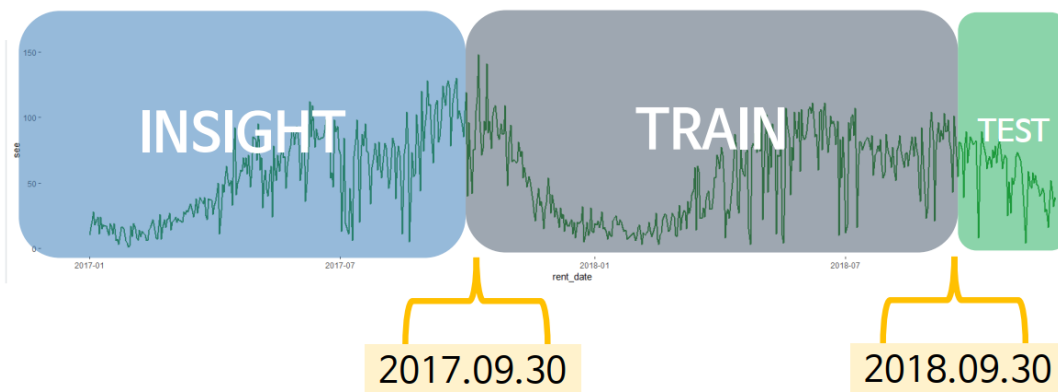
▲ 그림4. 분석 Flow Chart

분석 과정에서 자전거 대여에 영향을 주는 요소 중 날씨(온도, 습도, 풍속, 미세먼지 등)가 자전거 이용에 요인분석 중 가장 유의미하게 나왔다는 연구를 참고하여 날씨 변수를 가장 중점적으로 사용하되, 각 대여소의 위치와 날씨의 특성까지 고려하여 변수간의 관계와 최종 대여건수 예측까지의 분석을 시행할 것이다. 최종적으로 어느 조건에 어느 대여소가 해당 날짜에 자전거의 수요-공급 현황을 확인하여 서울시 에코 시티로 발돋움 할 수 있도록, 따릉이의 운영 개선방안을 제시할 것이다.

II. 활용 데이터 정의

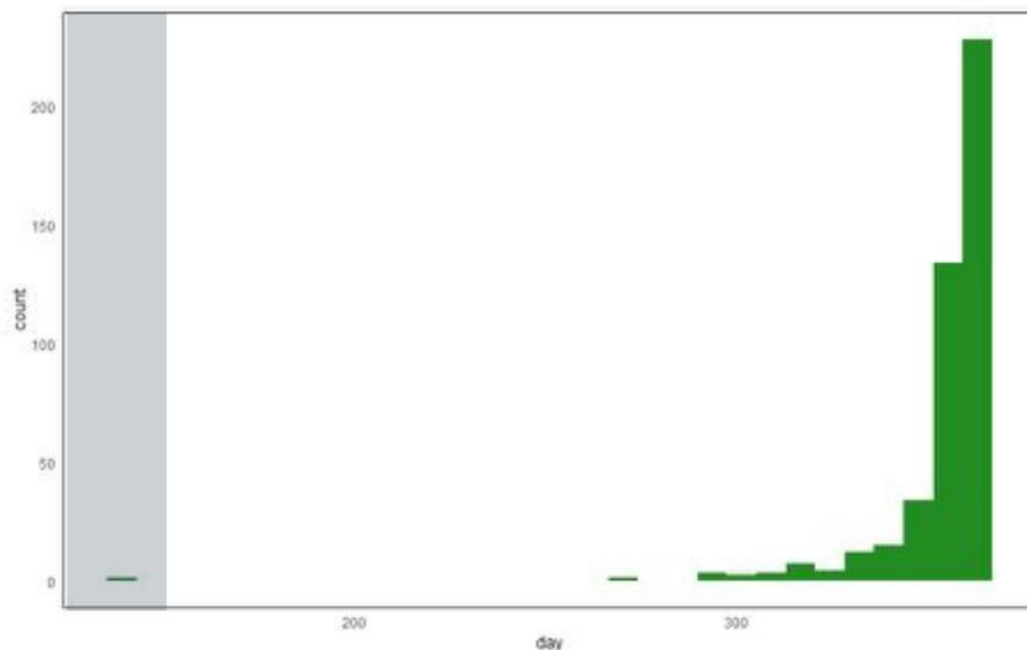
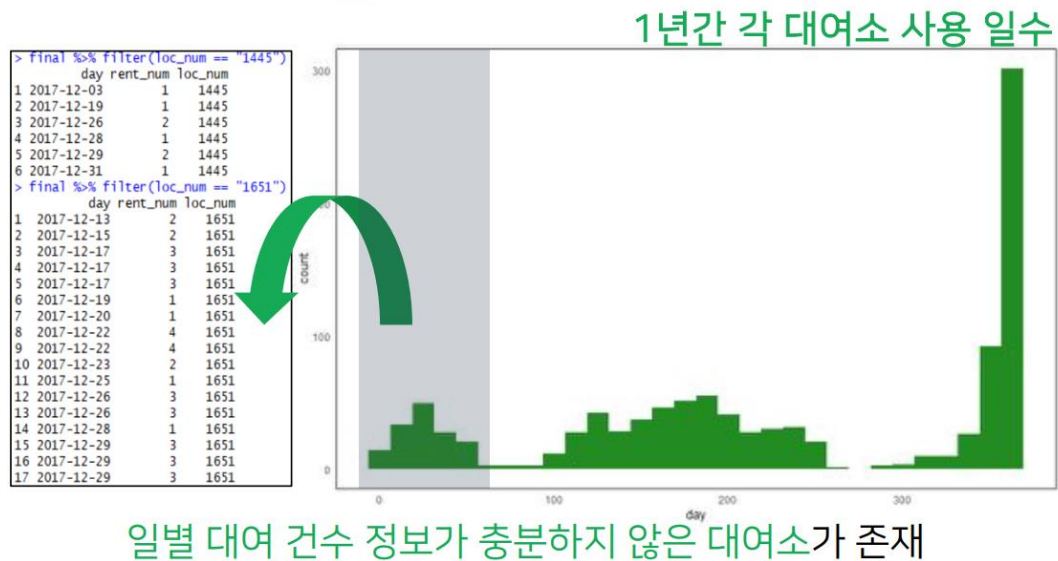
가) 데이터 수집 및 전처리

① 대여소 장소 선정 및 기간 선정



▲ 그림 5. 데이터 사용기간 정의

먼저 전체 데이터 중, 2017/01/01~2017/09/30기간의 데이터는 이미 나와있는 대여-반납 건수와 그 특징을 사용하여 대여소에 대한 탐색을 하는 데에 사용할 것이다. 이후 실제 모델링을 위해 2017/10/01~2018/09/30기간을 모델 학습에 사용한다. 이는 계절과 날씨를 고려해 주기 위한 결정으로 총 1년의 시간이다. 이후 2018/10/01~2018/11/30기간을 모델 테스트에 사용한다. 가장 최근의 데이터를 모델 테스트에 사용함으로써 모델의 정확성을 높이며 최신 현상을 반영하는 모델링이 될 것이다.



▲ 그림 6. 데이터 사용공간 정의 (데이터가 충분하지 않은 대여소 삭제)

전체 대여소에 대해서 데이터가 충분히 존재하는 가에 대한 탐색을 해본 결과, 2017년도 중 12월에 생긴 대여소나 사용 중 사라진 대여소 등 모델링에 적합하지 않은 표본들이 발견되었다. 본

팀은 정밀한 모델링을 위해 다음과 같이 2017년의 첫 주에 사용 정보가 있는 총 443개의 대여소들만을 모델에 사용하기로 결정하였다.

나) 데이터 수집

따릉이의 수요를 정확히 예측하기 위해 우선 '서울시 열린데이터광장' (<https://data.seoul.go.kr/dataList/datasetView.do?infId=OA-15493&srvType=A&serviceKind=1¤tPageNo=1>) 에서 따릉이 자전거과 대여소, 그리고 자전거 도로에 대한 데이터를 수집했다. 또한 날씨와 미세먼지를 고려하기 위해 '기상청' (www.kma.go.kr/info_open/public_data/request.jsp) 의 날씨와 미세먼지 데이터를 수집하였고 '대한민국 구석구석' (https://korean.visitkorea.or.kr/list/fes_list.do#0^All^All^1^All^All^1^10^1^)에서 분석 기간 동안 서울시에서 열린 축제 데이터들을 수집했다.

① 따릉이 이용자 데이터

대여일자	대여시간	대여소번호	대여소명	대여구분코드	성별	연령대코드	이용건수
2018-08-01	0	1361	홍익중고 입구	정기권	F	~10대	1
2018-08-01	0	165	중앙근린공원	정기권	F	~10대	1
2018-08-01	0	704	남부법원검찰청 교차로	정기권	F	~10대	1
2018-08-01	0	1139	용문사 버스정류장	정기권	F	~10대	1
2018-08-01	0	1503	이디야 커피 수유역점 앞	정기권	F	~10대	1
2018-08-01	0	1920	서울미래초등학교 사거리	정기권	F	~10대	1
2018-08-01	0	503	덕삼스타시티 C동 앞	정기권	F	20대	3
2018-08-01	0	935	역촌역 1번 출구	정기권	F	20대	1

▲ 표1. 이용자 데이터

따릉이 이용자들의 성별, 나이, 대여구분코드에 대한 정보가 담겨있어 이를 차후 PCA를 통한 지수생성에 사용하기 위해 팩터처리 되어있는 변수들을 long데이터에서 wide데이터로 전처리 해주었다. 이후 필요한 PCA에 쓰이는 변수들만 골라내어 남겨두었다.

morning_0612	night_1800	female_30	female_40	female_50	male_30	male_40	male_50	commutation	commutation_2hour	weekday
938	1279	464	102	79	997	105	8	2779	188	2712
2519	6678	1914	486	337	3068	908	238	10948	338	10670
1572	4686	1051	452	151	1934	576	342	7519	335	7456
1239	5442	1345	264	99	2234	832	287	7794	429	7933
1172	2855	820	163	93	1063	410	124	4737	204	4894
1830	8643	1885	404	143	2756	1298	484	11175	647	11500
2368	4745	1335	504	142	2267	1676	337	10089	380	9285
607	3794	647	145	26	1365	508	106	5873	234	5840

afternoon_1218	female_20	male_20	Unknown	group	oneday_2hour	oneday_no_id	oneday_id	weekend
1026	838	935	64	41	31	108	503	938
3915	4067	3128	357	99	275	660	2518	4168
3000	2461	3060	250	93	240	465	2056	3252
3687	2895	2792	248	92	255	508	2205	3350
2210	1875	2133	160	47	196	333	1498	2121
4736	4355	5181	398	154	608	842	3996	5922
3407	2277	3107	162	30	92	278	1320	2904
2501	2083	2955	193	57	126	389	1582	2421

▲ 표2. 성별/나이/대여구분코드를 펼친 wide data

② 따릉이 대여소 데이터

	A	B	C	D	E	F	G	H	I	J	K
1	자전거번호	대여일시	대여소번호	대여소명	거치대번호	반납일시	반납대여소번호	반납대여소명	반납 거치대	이용시간(분)	이용거리(m)
2	SPB-00230	#####	419	홍플러스 앞	5	#####	914	새절역 2번출구	18	20	3340
3	SPB-02803	#####	825	서빙고동 주민센터 앞	2	#####	822	이촌1동 마을공	7	14	3380
4	SPB-03956	#####	906	연신내역 5번출구15	1	#####	931	역촌파출소	6	11	1610
5	SPB-02864	#####	906	연신내역 5번출구15	9	#####	912	응암오거리	1	24	3620
6	SPB-01388	#####	330	청계천 한빛광장	6	#####	171	임광빌딩 앞	9	17	2280
7	SPB-04437	#####	347	동대문역사문화공원	16	#####	344	성균관대 E하우	4	11	1380

▲ 표3. 대여한 날짜/장소와 반납한 날짜/장소가 나와있는 데이터

‘누가’, ‘어디서’, ‘언제’ 따릉이를 대여하고 반납했는지에 대한 데이터이다. 차후 네트워크 분석에서의 링크와 노드 데이터로 사용하기 위해 대여한 장소와 반납한 장소가 같은 경우의 수를 합쳐 다음과 같이 전처리 해주었다.

loc_num	tot_usage	rent_loc_num	return_loc_num	tot_use
101	3650	101	101	611
102	14838	101	102	79
103	10708	101	103	120
104	11283	101	104	369
105	7015	101	105	60
		101	106	446

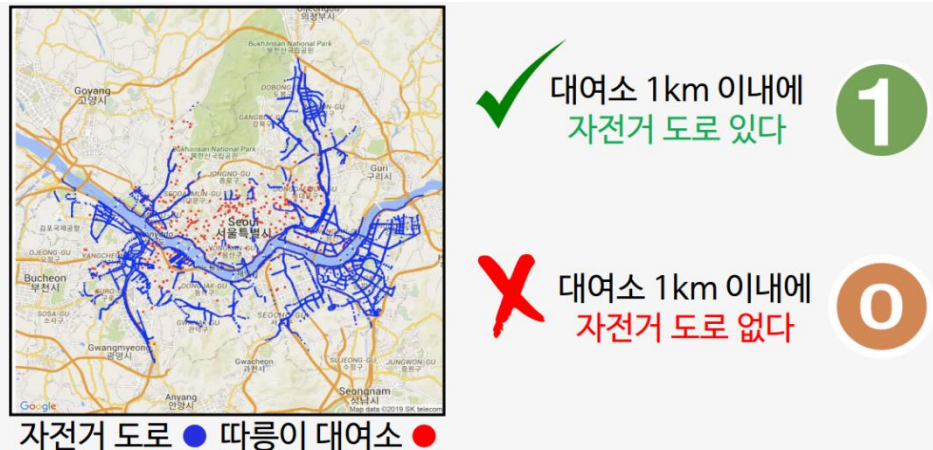
▲ 표4. 노드데이터와 링크데이터

③ 자전거 도로 데이터

도로명	자치구명	도로길이	도로폭	자전거도로종류	분리시설	도로포장종류	도로포장색상	설치장소
도봉로	도봉구	31.14177485	2.0	자전거보행자겸...		보도블럭	회색	도로변
노해로	도봉구	87.44974597	2.0	자전거보행자겸...	연석	아스콘	회색	도로변
	노원구	5.741990713	0.0	자전거보행자겸...				
노원로	노원구	239.5600977	1.2	자전거보행자겸...	경제석	아스콘	적색	도로변
덕릉로	노원구	175.8778607	2.1	자전거보행자겸...	경제석	아스콘,고압블록	적색	도로변

▲ 표5. 자전거 도로 데이터

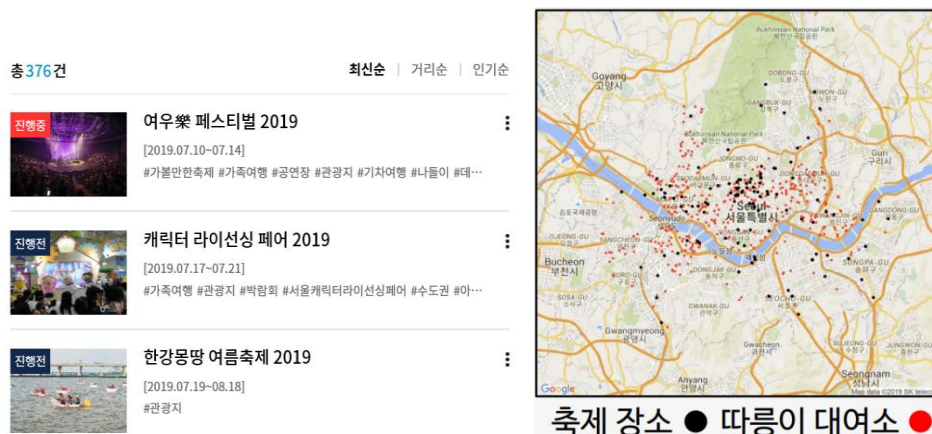
자전거 도로의 위치정보, 도로 길이와 폭 등에 대한 데이터이다. 이 데이터를 바탕으로 따릉이 대여소를 기준으로 일정 반경 내에 자전거도로가 있으면 1, 없으면 0으로 나타내는 파생 변수를 만들었다.



▲ 그림 7. 지도에 나타낸 자전거 도로

④ 축제 데이터

‘대한민국 구석구석’ 사이트의 축제 섹션에서 분석 기간 동안 열린 축제의 장소와 기간을 크롤링하여 수집하였다. 이 데이터를 바탕으로 자전거 대여소의 파생변수를 만드는 방식과 동일한 방식으로 따릉이 대여소 일정 반경 내에 축제가 열리면 1, 그렇지 않으면 0으로 나타내는 파생변수를 생성하였다.



▲ 그림 8. 지도에 나타낸 축제 장소

축제 장소를 고려함과 동시에 축제 기간도 고려해주기 위해 본 팀은 Matrix Computation을 사용하여 변수들을 파생하였다.

Festival에 대한 날짜 정보

fest1
fest2

fest203

2017-10-01
0
1

2017-10-02
0
1

2017-10-03
0
1

2017-10-04
1
1

2018-09-28
1
1

2018-09-29
1
0

2018-09-30
1
0

...

1
0

Festival에 대한 가까운 대여소 정보

101
102
103
993

fest1
0
1
0
1

fest2
0
1
0
0

fest3
1
1
0
0

fest203
1
1
0
0

X

=

최종 Festival 변수

101
102
993

2017-10-01
0
0
0

2017-10-02
0
0
0

2017-10-03
0
0
1

2017-10-04
3
2
0

2018-09-28
0
1
0

2018-09-29
2
0
0

2018-09-30
1
0
0

...

▲ 그림 9. 축제 정보 파생변수

⑤ 날씨 데이터

평균기온(°)	최저기온(°)	최저기온 /	최고기온(°)	최고기온 /	강수 계속/10분	최다 10분	최다 10분	최다 1시간	최다 1시간	최다 일강수량(mm)	최대 순간
2.7	-1.6	540	6.9	1419							4.5
5	1.8	2355	9.2	1355	1			0.3	600	0.3	7
2	-2.3	731	7.7	1444							7.2
3.9	1	1	8.9	1519							6.5
3.8	-0.1	316	7.3	1444							9.4
5.4	2.5	804	11.4	1548							8.7
4.6	0	815	10.5	1327							5.3
6.5	4	458	10.9	1350	0						8
1.5	-3.1	2357	4.3	1454	0						9.5

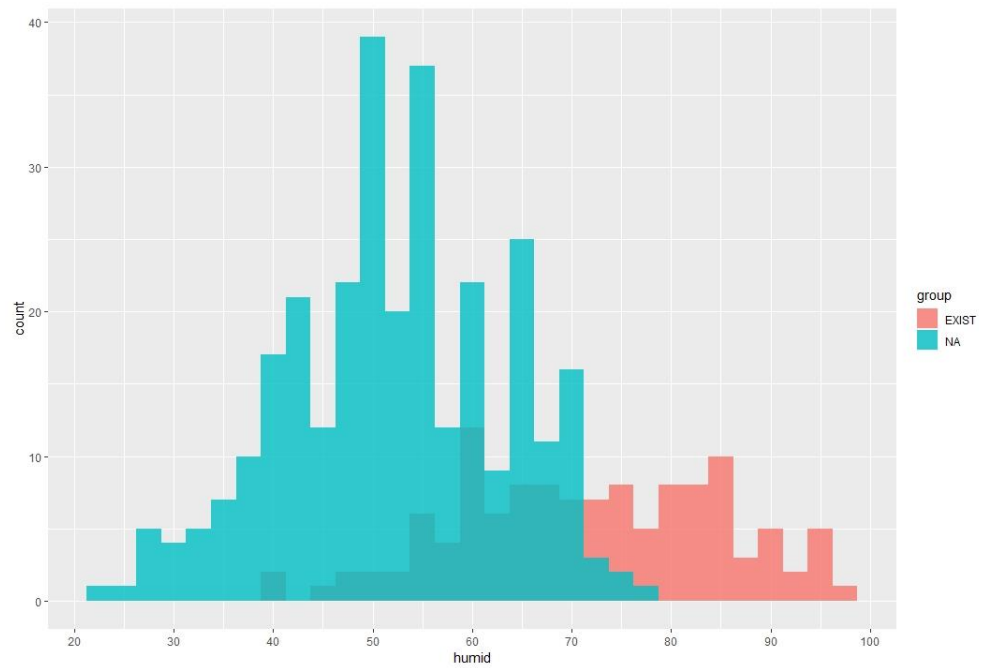
▲ 표6. 날씨 데이터

날씨에 관련한 변수가 세부적인 변수까지 포함해 약 50개였으나 자전거 이용에 크게 영향을 줄 수 있는 강수량, 기온, 습도 등의 변수들을 선택하였다. 하지만 변수를 선택하기에 앞서 발견된 결측치들을 MICE패키지를 사용하여 imputation해 주었다.

MICE(Multiple Imputation via Chained Equation) 기법은 하나의 변수를 단일 결측치 처리하는 것이 아닌 주변 변수들과의 관계를 이용하여 결측치를 처리해주는 방법으로 변수들간의 상관관계가 중요하다. 따라서 각 변수들의 경우를 하나하나 살펴보았다.

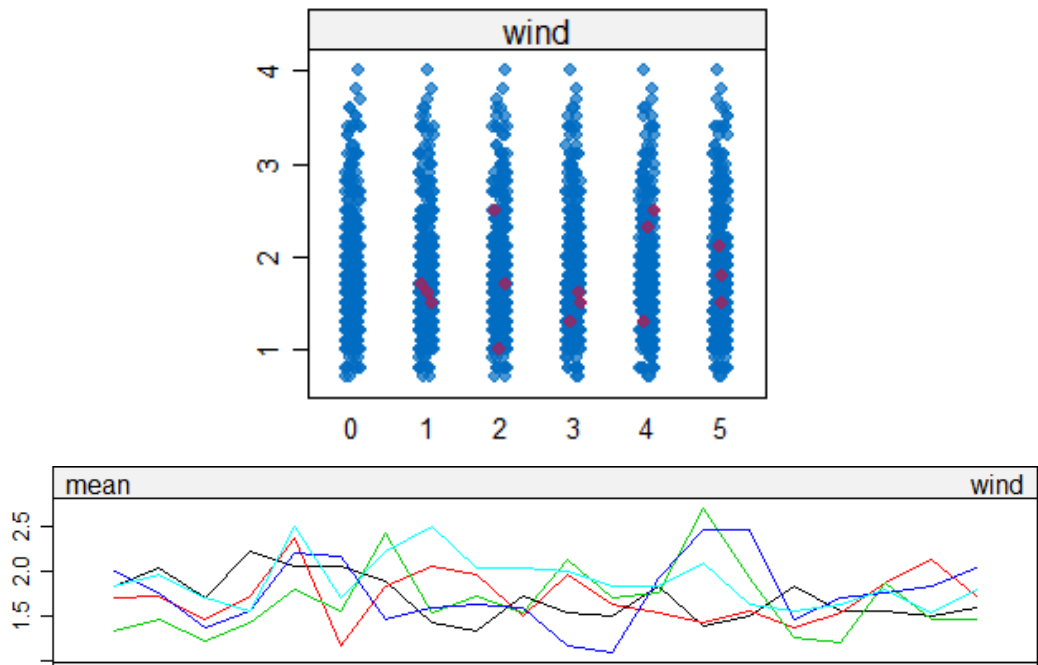
- 강수량 변수

강수량 결측치들의 습도 분포를 살펴본 결과, NA들의 습도가 모두 상대적으로 낮게 분포해있다는 것을 알 수 있다. 즉, NA들은 missing at random의 결측치들이 아닌, 강수량이 0인, 비가 내리지 않았을 때를 나타낸다. 따라서 본 팀은 강수량 변수의 NA는 모두 0으로 처리해 주었다.



▲ 그림 10. 강수량이 NA인 값들의 습도 분포

- 평균풍속 변수

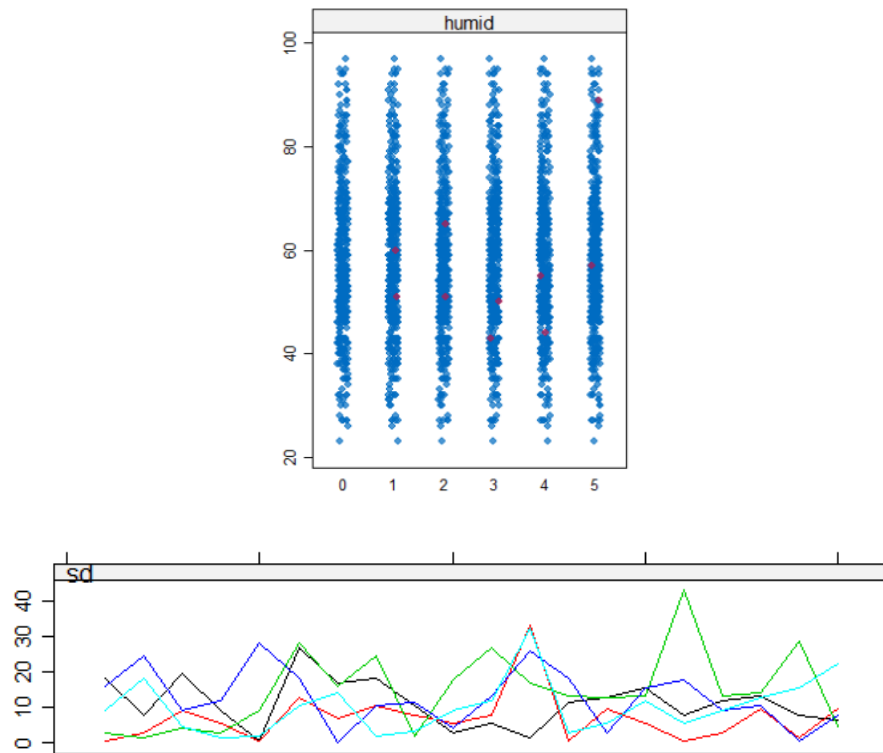


▲ 그림 11. MICE를 이용한 풍속 결측치 처리

풍속들과 관련된 변수(일시, 최대순간 풍속, 최대순간 풍속풍향, 최대순간 풍속시각, 최대 풍속, 최대풍속 풍향, 최대풍속 시각, 평균풍속, 풍정합, 최다풍향) 들을 사용하여 imputation한 결과 결측 처리된 값들에 이상치가 없고 결측처리를 한 값들의 평균이 수

렴한다는 것을 알 수 있다. (MICE 파라미터는 모두 $m = 5$, $maxit = 20$, $method = 'cart'$ 를 사용하였다.)

- 상대습도 변수



▲ 그림 12. MICE를 이용한 상대 습도 결측치 처리

습도와 관련된 변수(일시, 평균 이슬점온도, 최소 상대습도, 최소 상대습도 시각, 평균 상대습도) 들을 사용하여 imputation한 결과 결측 처리된 값들에 이상치가 없고 결측처리를 한 값들의 표준편차가 수렴한다는 것을 알 수 있다.

- 일교차 변수

마지막으로 최고기온과 최저기온의 차를 이용한 일교차 변수를 생성했다.

$$\text{diff_temp} = \text{max_temp} - \text{min_temp}$$

▲ 식1. 일교차 공식

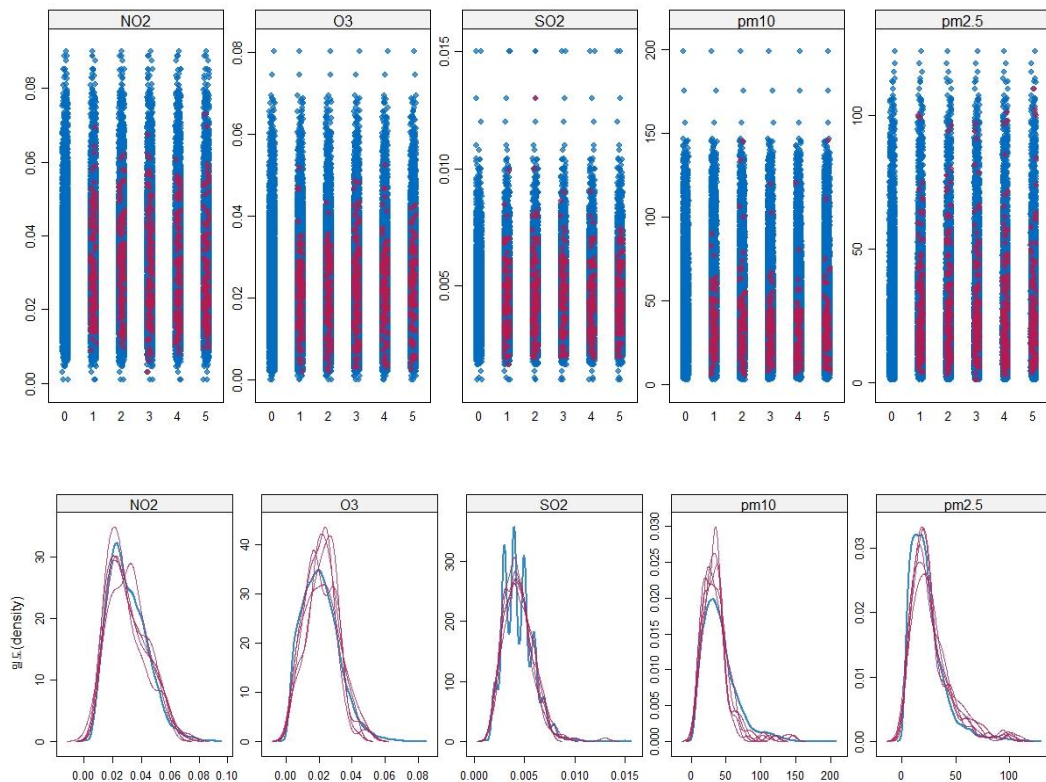
최종 기온/습도/강수량 데이터는 다음과 같다.

date	mean_temp	precip	wind	humid	diff_temp
2017-10-01	18.2	4.0	2.40	75.0	5.7
2017-10-02	22.0	1.5	2.40	62.0	13.8
2017-10-03	17.6	0.0	2.30	50.0	10.2
2017-10-04	16.7	0.0	1.80	49.0	13.6
2017-10-05	18.7	0.0	2.10	56.0	9.5
2017-10-06	18.9	0.5	2.40	66.0	7.1
2017-10-07	21.9	0.0	1.10	68.0	11.9
2017-10-08	23.0	0.0	1.50	67.0	9.4

▲ 표7. 최종 기온/습도/강수량 데이터

⑥ 미세먼지 데이터

전 날짜에 대한 미세먼지 데이터를 구하였고, 미세먼지 또한 결측치가 존재하였다. 하지만 미세먼지는 변수들간의 상관관계가 뚜렷이 드러났으므로 MICE를 바로 사용해 주었다.



▲ 그림 13. MICE를 이용한 미세먼지 결측치 처리

미세먼지 변수들을 imputation한 결과 결측 처리된 값들에 이상치가 없고 각 변수의 특수한 분포를 잘 고려해주어 결측치 처리 되었음을 확인할 수 있다.

최종 미세먼지 데이터는 다음과 같다.

when	where	NO2	O3	SO2	pm10	pm2.5
20171001	강남구	0.026	0.014	0.005	35.00000	24
20171001	강동구	0.025	0.017	0.004	32.00000	17
20171001	강북구	0.022	0.017	0.003	29.00000	20
20171001	강서구	0.036	0.017	0.006	36.00000	22
20171001	관악구	0.025	0.015	0.005	29.00000	17
20171001	광진구	0.032	0.027	0.003	28.00000	20
20171001	구로구	0.024	0.014	0.006	33.00000	21

▲ 표8. 최종 미세먼지 데이터

⑦ 장소와 날짜에 대한 파생변수

rent_date	loc_num	gu	heuil	month	day
2017-10-01	270	영등포구	1	10	1
2017-10-01	271	영등포구	1	10	1
2017-10-01	272	영등포구	1	10	1
2017-10-01	300	종로구	1	10	1
2017-10-01	301	종로구	1	10	1
2017-10-01	302	종로구	1	10	1

▲ 표9. 장소와 날짜 데이터

대여소가 속해있는 구 정보와 날짜의 월, 요일, 공휴일 정보를 추가하였다.

III. 데이터 처리 방안 및 활용 분석 기법

가) PCA를 활용한 대여소 지수 생성

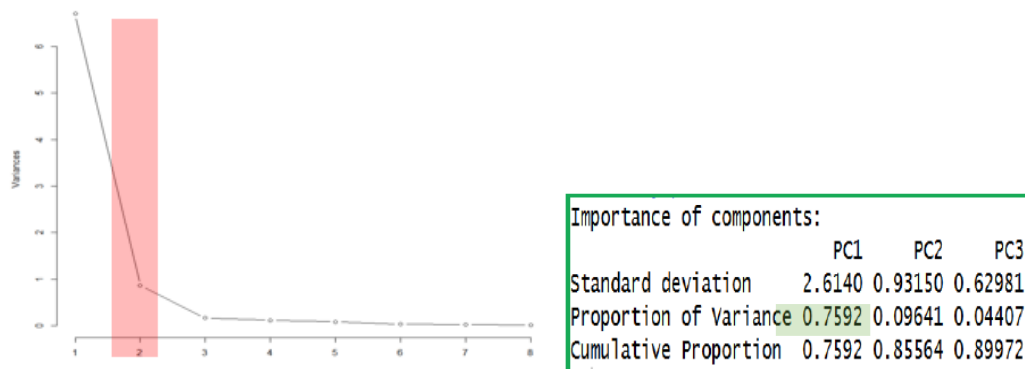
① 한량지수(여가 지수)

위에서 정제한 데이터 중 따릉이를 이용한 사람들의 나이, 성별, 대여구분코드에 대한 데이터를 사용하여 지수를 파생하고자 했다.

따릉이의 이용 목적이 다른 목적들보단 '여가'에 가깝다고 판단한 변수들을 선정하여 PCA를 진행하였다. 이 때 쓰인 변수들은 다음과 같다.



▲ 그림 14. 한량지수 파생에 쓰인 변수들



▲ 그림 15. Scree plot과 PV

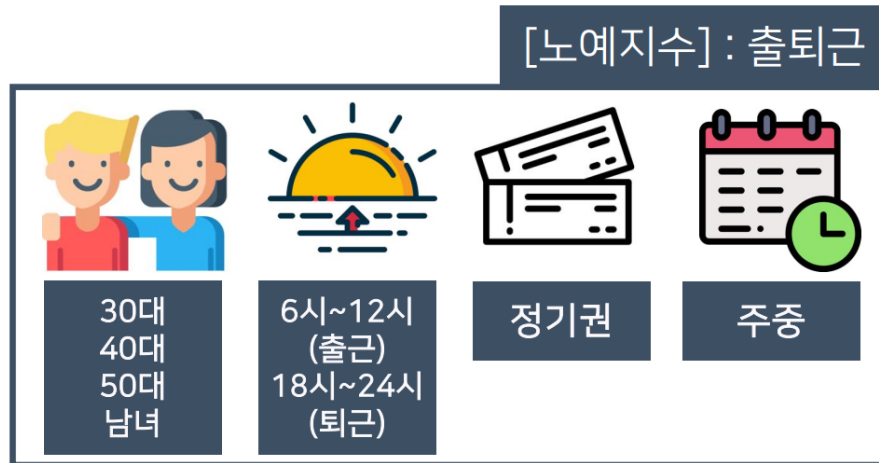
	PC1	PC2	PC3
afternoon_1218	0.11079810	0.097475435	0.23960041
female_20	0.11337636	0.099539744	0.05767349
male_20	0.09051286	0.260497932	0.12998089
Unknown	0.11417668	0.103564043	0.15123296
group	0.10900298	0.137460742	0.20326451
oneday_2hour	0.11264037	0.126522961	0.07356278
oneday_no_id	0.11452713	0.110907803	0.11763629
oneday_id	0.11920380	0.005217022	0.02447658
weekend	0.11706312	0.063862347	0.09861097

▲ 그림 16. 각 변수의 loading값

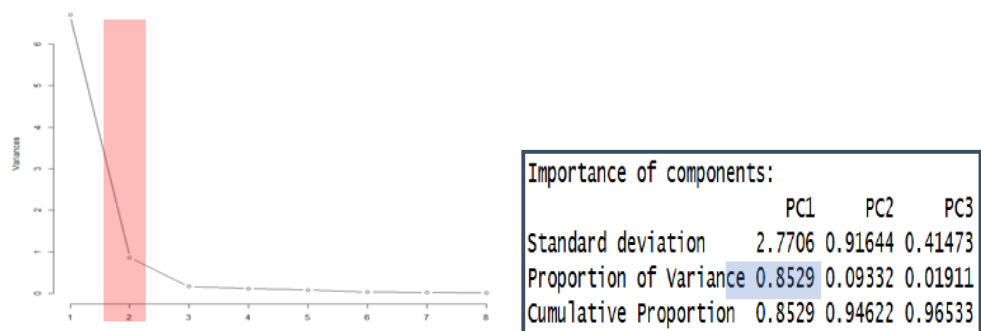
Scree plot과 Proportion of Variation을 살펴본 결과, PC1만으로 데이터의 75%이상을 설명하므로 PC1이 데이터를 충분히 설명한다고 볼 수 있었다. 또한 각 변수의 loading도 PC1에서 가장 고르게 분포되어 있었으므로 각 변수들이 PC1에 사용된 비중만큼 가중 평균을 내어 지수를 산출하였다. 대여소가 여가용으로 사용되는 정도를 나타내는 '한량 지수'를 파생하였다.

② 노예지수(출퇴근 지수)

위의 한량 지수 방식과 동일하게 따릉이의 이용 목적이 다른 목적들보단 '출퇴근용'에 가깝다고 판단한 변수들을 선정하여 PCA를 진행하였다. 이 때 쓰인 변수들은 다음과 같다.



▲ 그림 17. 노예지수 파생에 쓰인 변수들



▲ 그림 18. Scree plot과 PV

	PC1	PC2	PC3
morning_0612	0.08717924	0.01612582	0.1467631250
night_1800	0.09829478	0.09699387	0.0901765979
female_30	0.09731568	0.01998574	0.0897378637
female_40	0.08278857	0.14273362	0.0833125454
female_50	0.06863468	0.22476258	0.1171128869
male_30	0.09778900	0.07854089	0.0720209503
male_40	0.09179930	0.04050215	0.1073595128
male_50	0.07933425	0.13711406	0.1382860369
commutation	0.10213949	0.07822882	0.0304905076
commutation_2hour	0.09012707	0.08047884	0.1244901173
weekday	0.10459795	0.08453362	0.0002498563

▲ 그림 19. 각 변수의 loading값

Scree plot과 Proportion of Variation을 살펴본 결과, PC1만으로 데이터의 85%이상을 설명하므로 PC1이 데이터를 충분히 설명한다고 볼 수 있었다. 또한 각 변수의 loading도 PC1에서 가장 고르게 분포되어 있었으므로 각 변수들이 PC1에 사용된 비중만큼 가중 평균을 내어 지수를 산출하였다. 대여소가 여가용으로 사용되는 정도를 나타내는 '노예 지수'를 파생하였다.



▲ 그림 20. 최종 대여소별 지수 index 예시1

최종 산출된 지수들을 예시를 통해 살펴본 결과, 유의미한 결과가 도출 되었다. '서울숲 남문 버스정류장 옆' 대여소는 공원 옆으로, 여가 생활과 관련이 있다. 한량 지수는 2.7, 노예 지수는 -1로 유의미한 차이가 났다.



▲ 그림 21. 최종 대여소별 지수 index 예시2

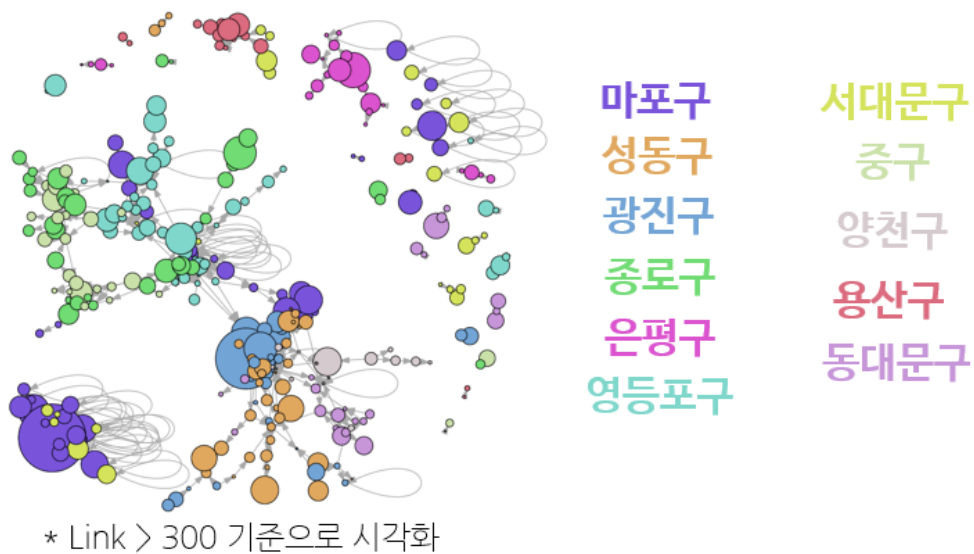
이와 반대로 '정동 사거리' 대여소는 회사 주변에 존재하며 한량 지수는 -0.42, 노예 지수는 3으로 유의미한 차이가 났다.

나) 네트워크 분석을 통한 대여소 클러스터링

① 네트워크 분석과 시각화

위에서 정제한 데이터 중 노드와 링크 데이터를 사용하여 네트워크 분석을 진행하였다. 이를

구 별로 시각화한 결과는 다음과 같다. (링크의 개수는 보다 나은 시각화를 위해 300개를 기준으로 정제되었다.)



▲ 그림 22. 네트워크 분석 그래프

위의 결과에서 우리는 대여소들이 구 별로 가깝게 모이는 것을 확인할 수 있다. 즉, '구' 또한 유의미한 변수가 될 것이며, 대여소별 특징이 확연히 존재한다는 것을 알 수 있다.

② 네트워크 분석과 중심성 산출

네트워크 분석을 이용하여 우리는 노드들이 얼마나 연결되어있고, 가까운지를 기준으로 총 4개의 중심성을 산출할 수 있다. 가장 빈번하게 사용되는 중심성 3개를 본 팀의 분석상황과 연결지어 나타내면 다음과 같다.

매개 중심성(Betweenness centrality): 다른 대여소를 얼마나 잘 연결시켜주는가?

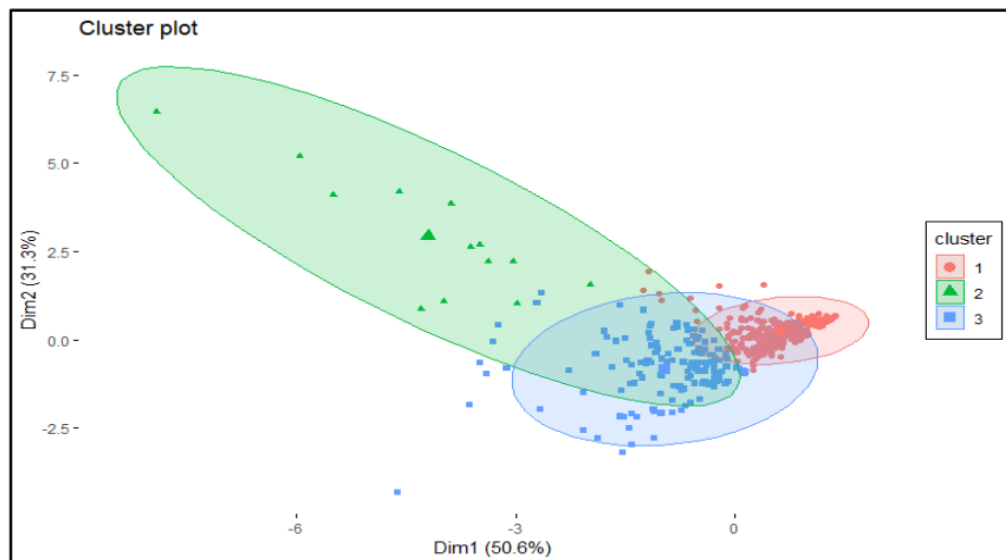
연결정도 중심성(Degree centrality): 몇 개의 대여소와 연결되어 있는가?

위세 중심성(Eigenvector centrality): 중요한 대여소와 연결되어 있는가?

이러한 중심성 변수를 그대로 사용할 수도 있지만, 중심성 변수를 이용하여 본 팀은 대여소들의 특성을 반영한 비지도 학습인 클러스터링을 진행했다.

btw	degree	eigen
110.210152	115	0.33637099
10.635965	19	0.04755840
9.281161	19	0.04990792
78.884304	146	0.37186261
0.750000	16	0.04979088
175.939734	141	0.39729732
278.963849	120	0.32125661
276.895491	189	0.50894613
1942.525995	315	0.83627831
1305.444845	253	0.66723780

▲ 표10. 중심성 데이터



```
> summary(group1)
```

degree	btw	eigen
Min. : 1318	Min. : 343.6	Min. : 0.2060
1st Qu.: 5918	1st Qu.: 625.8	1st Qu.: 0.3671
Median : 6580	Median : 729.3	Median : 0.4660
Mean : 7196	Mean : 813.1	Mean : 0.5184
3rd Qu.: 9191	3rd Qu.: 834.8	3rd Qu.: 0.6582
Max. : 12865	Max. : 1649.5	Max. : 1.0000

```
> summary(group2)
```

degree	btw	eigen
Min. : 498	Min. : 169.7	Min. : 4.346e-05
1st Qu.: 1888	1st Qu.: 791.3	1st Qu.: 2.747e-03
Median : 3474	Median : 1103.6	Median : 7.605e-03
Mean : 4133	Mean : 1151.6	Mean : 2.028e-02
3rd Qu.: 5413	3rd Qu.: 1359.0	3rd Qu.: 2.276e-02
Max. : 13606	Max. : 3536.0	Max. : 2.318e-01

```
> summary(group3)
```

degree	btw	eigen
Min. : 1	Min. : 0.0	Min. : 0.0000000
1st Qu.: 370	1st Qu.: 151.0	1st Qu.: 0.0005522
Median : 1056	Median : 337.2	Median : 0.0024886
Mean : 1229	Mean : 350.9	Mean : 0.0145554
3rd Qu.: 1791	3rd Qu.: 528.5	3rd Qu.: 0.0111369
Max. : 4682	Max. : 1070.8	Max. : 0.2556124

▲ 그림23. 클러스터링 결과

클러스터링 결과, 대여소는 총 3개의 그룹으로 나뉘었다. 각 그룹별 특징을 살펴보니, 그룹1은 연결되어 있는 대여소가 많으며, 그 중에서도 특히 중요한 대여소가 많이 연결이 되어 있었다. 그룹2는 다른 대여소들을 가장 많이 연결시켜 주었으나, 중요한 대여소들과는 비교적 적게 연결되어 있었다. 그룹3는 연결되어 있는 대여소도 적으며, 다른 대여소들을 연결시켜주는 역할도 가장 작았다.

다) 최종 데이터

지수 변수와 클러스터링 변수까지 총합하여 최종 데이터셋이 만들어 졌다.

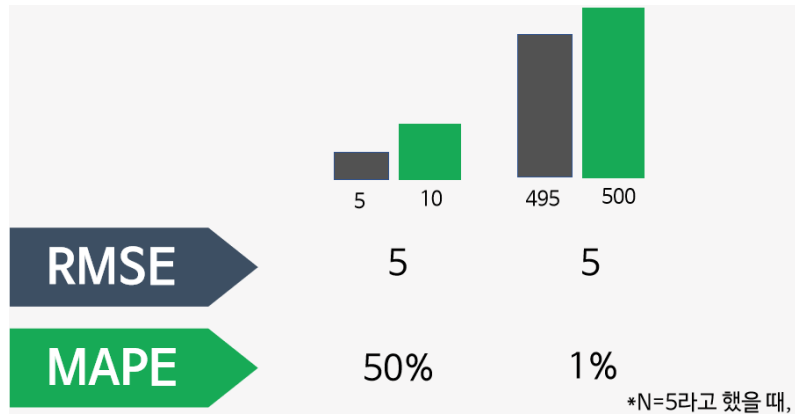
변수명	설명
Y	따릉이 대여건수
rent_date	TRAIN DATA : 2017-10-01~2018-09-30 TEST DATA : 2018-10-01 ~ 2018-11-30
loc_num	101-933번 (총 443개)
gu	광진구 동대문구 마포구 서대문구 성동구 양천구 영등포구 용산구 은평구 종로구 중구
rent_month	1월-12월
day	월요일-일요일
redday	공휴일: 1 공휴일 X: 0
festival	대여소 100m 반경 축제 진행 : 1
bikeroad	대여소 100m 반경 이내 자전거 도로 존재 : 1
networkgroup	중심성에 따라 1,2,3으로 군집화
hanlyang_index	여가용으로 사용되는 대여소
slave_index	출퇴근용으로 사용되는 대여소
mean_temp	평균기온 (°C)
precip	강수량 (mm)
wind	풍속 (m/s)
humid	상대습도 (%)
diff_temp	일교차 (°C)
NO2	이산화질소 (ppm)
O3	오존 (ppm)
SO2	아황산가스/이산화황(ppm)
pm10	미세먼지 ($\mu\text{g}/\text{m}^3$)
pm2.5	초미세먼지 ($\mu\text{g}/\text{m}^3$)

▲ 표11. 최종 데이터셋

IV. 분석 결과

가) 평가 지표 선정

모델의 예측 성능을 확인하기 위해선 그 예측의 오차가 적어야 할 것이다. 모델의 예측력 비교를 할 때, 가장 많이 쓰이는 지표로는 MSE, MAE, RMSE, MAPE 정도가 있다. 이 중 MSE는 모델의 실제값과 예측값의 차를 제곱을 하기 때문에 outlier(특이치)에 민감하다는 단점이 있다. 이에 우리의 데이터가 특이치에 민감할 수 있는 데이터임을 감안하여 MSE 보다 특이치에 덜 민감하게 반응하는 MAE를 선택하는 것이 합리적일 것이라고 판단했다. 더 나아가, 아래 <참조>에서 볼 수 있듯, 오차의 값이 같은 각각의 데이터 또한 데이터의 크기에 따라 오차의 비율은 달라질 수 있는 점을 고려하여, 이번 분석에서의 모델 예측 성능 비교 지표는 'MAPE'로 결정하였다.



▲ 그림24. RMSE와 MAPE의 비교

앞서 선정한 모델들을 최종 test set으로 검증하기 전에 다섯 모델 간, train set을 5 fold cross validation(교차검증) 했을 때 얻는 MAPE값을 통해 비교를 먼저 할 것이다. MAPE는 결국 오차(error)이므로 MAPE가 가장 낮은 모델이 선택될 것이다.

나) 해석 모델

날씨를 포함해 날짜와 대여소에 대한 정보를 포함한 변수들이 대여-반납 건수에 대해 유의미한 변수들일지 알아보기 위해 회귀 분석을 실시하였다. 그러나 우리에게 주어진 반응 변수는 도수 자료로, 일반적인 연속형 자료와는 다른 특성을 가지고 있다.

따라서 본 팀은 반응 변수가 도수 자료로 주어졌을 때 효과적이며 포아송 분포의 과산포 문제가 해결된 음이항 회귀모형을 사용하였다.

	Variables	VIF
1	mean_temp	15.199252
2	precip	1.573400
3	wind	1.886915
4	humid	3.124304
5	diff_temp	2.231899
6	gu	NA
7	NO2	4.650682
8	O3	3.246368
9	SO2	2.727317
10	pm10	7.135787
11	pm2.5	7.142322
12	redday	NA
13	festival	NA
14	hanlyang_index	2.876908
15	slave_index	2.537135
16	networkgroup	NA
17	bikeroad	NA
18	rent_month	NA
19	day	NA

▲ 사진25. VIF값 계산 결과

회귀모형은 다중공선성이 존재한다면 유의미하지 않은 변수도 유의미하다는 결과가 나오거나

변수해석의 결과가 달라질 수 있으므로, VIF값이 10보다 큰 mean_temp변수를 제거해주고 진행해주었다. 회귀모형에 적합한 결과, 대부분의 변수가 유의미한 결과를 보였으며 해석은 다음과 같다. 대여 건수와 반납 건수는 정비례하는 관계를 보였으므로, 대여 건수를 중심으로 해석한다.

① Month 변수

rent_month2	8.421e-02	7.438e-03	11.321	< 2e-16	***
rent_month3	8.532e-01	7.554e-03	112.943	< 2e-16	***
rent_month4	1.181e+00	8.027e-03	147.147	< 2e-16	***
rent_month5	1.427e+00	8.591e-03	166.137	< 2e-16	***
rent_month6	1.644e+00	8.256e-03	199.084	< 2e-16	***
rent_month7	1.424e+00	7.832e-03	181.815	< 2e-16	***
rent_month8	1.367e+00	8.181e-03	167.047	< 2e-16	***
rent_month9	1.509e+00	8.476e-03	177.975	< 2e-16	***
rent_month10	1.578e+00	7.736e-03	204.024	< 2e-16	***
rent_month11	1.066e+00	7.117e-03	149.735	< 2e-16	***
rent_month12	1.996e-01	7.260e-03	27.493	< 2e-16	***

▲ 사진26. 음이항 회귀모형 해석 예시1

1월을 base-line으로 잡은 회귀모형 결과이다. 모든 계수에는 exponential을 취하여 계산한다. 3월 대여건수는 1월 대여건수의 2.35배이며, 6월 대여건수는 1월 대여건수의 5.18배이다. 이는 즉, 날씨가 풀리고 해가 길어지는 달에 따릉이 대여-반납 건수가 증가한다는 것을 알 수 있다. 평균기온 변수는 삭제되었지만, 이를 통해 평균기온이 증가하면 따릉이 대여-반납 건수가 증가할 것이라고 유추할 수 있다.

② Day 변수

dayFriday	1.965e-01	5.692e-03	34.527	< 2e-16	***
dayMonday	1.301e-01	5.505e-03	23.635	< 2e-16	***
daySaturday	1.276e-01	4.899e-03	26.046	< 2e-16	***
dayThursday	1.700e-01	5.669e-03	29.979	< 2e-16	***
dayTuesday	2.029e-01	5.636e-03	35.992	< 2e-16	***
dayWednesday	2.128e-01	5.546e-03	38.369	< 2e-16	***

▲ 사진27. 음이항 회귀모형 해석 예시2

Sunday를 base-line으로 잡은 회귀모형 결과이다. 수요일의 대여건수는 일요일 대여건수의 1.24배이다. 즉, 평일보다 휴일에 사람들이 따릉이를 많이 이용한다는 것을 알 수 있다.

③ Index 변수

hanlyang_index	-0.0052937	0.0006227	-8.501	< 2e-16	***
slave_index	0.1776590	0.0006838	259.821	< 2e-16	***

▲ 사진28. 음이항 회귀모형 해석 예시3

한량지수가 1 unit 증가할 때, 대여건수는 0.99배씩 증가한다. 즉, 여가를 목적으로 따릉이를 이용하는 대여소에서는 다른 모든 변수가 고정되었을 때, 오히려 대여건수가 줄어듦을 알 수 있다. 즉, 여가가 목적인 지역에서는 따릉이 외에도 다른 여가 수단이 존재함으로 오히려 많은 따릉이의 재배치가 요구되지는 않는다는 것을 알 수 있다.

이와 반대로 노예지수가 1 unit 증가할 때, 대여건수는 1.2배씩 증가한다. 출퇴근을 목적으로 따릉이를 이용하는 대여소에서는 다른 모든 변수가 고정되었을 때, 대여건수가 증가한다. 이는 따릉이가 대중교통으로써 자리잡고 있음을 반영한다.

④ Bike road 변수

bikeroad1	5.863e-02	2.897e-03	20.236	< 2e-16 ***
-----------	-----------	-----------	--------	-------------

▲ 사진29. 음이항 회귀모형 해석 예시4

자전거 도로 변수 또한 유의미한 결과가 나왔는데, 자전거 도로가 있을 때의 대여건수가 없을 때의 대여건수의 1.07배라는 계수해석이 가능하다. 즉, 따릉이의 대여-반납 건수가 증가하는 데에는 자전거 도로 변수도 유의미한 관계가 존재한다는 것을 알 수 있다.

⑤ 기온/강수량/습도 변수

precip	-0.0200946	0.0001438	-139.734	< 2e-16 ***
wind	-0.0383030	0.0023864	-16.051	< 2e-16 ***
humid	-0.0005583	0.0001272	-4.391	1.13e-05 ***
diff_temp	0.0262635	0.0005479	47.933	< 2e-16 ***

▲ 사진30. 음이항 회귀모형 해석 예시5

강수량, 평균풍속, 상대습도, 일교차는 모두 유의미한 변수로 각각의 변수는 1 unit이 증가할 때마다, 대여건수는 0.98배, 0.96배, 0.99배, 1.02배로 증가하였다. 즉, 비가 오거나 바람이 많이 불 때, 습할 때는 모두 대여건수가 감소한다는 뜻이다. 더불어 일교차가 커질수록 대여건수는 증가한다. 계절의 변화가 있는 날들에 수요가 늘어나는 것으로 보인다. 이는 미세하지만 선형적인 관계를 가정했을 때 날씨와 기온의 수치에 따라 대여-반납 건수가 예측이 가능하다는 것을 증명한다.

⑥ 미세먼지 변수

NO2	2.0073773	0.1700419	11.805	< 2e-16	***
O3	3.5305034	0.1736232	20.334	< 2e-16	***
SO2	29.0803328	1.1960125	24.314	< 2e-16	***
pm10	-0.0011650	0.0001078	-10.803	< 2e-16	***
pm2.5	0.0010021	0.0001773	5.653	1.58e-08	***

▲ 사진31. 음이항 회귀모형 해석 예시6

이산화질소, 오존 모두 유의미한 변수로 각각의 변수는 1 unit이 증가할 때마다, 대여건수는 7.44배, 21.76배로 증가한다. 특히나 이산화황의 경우 $4.260181e+12$ 라는 큰 계수를 가지게 되는데, 이는 이산화황의 단위가 매우 작아 나타나는 현상으로 보여진다. 반대로 미세먼지와 초미세먼지 변수는 각각 1 unit이 증가할 때마다, 0.99배, 1.001배 등으로 변화를 보이는데, 이는 미세먼지가 사회적 이슈가 되고 경각심이 생기기 시작한 2019년도 이전의 데이터를 사용함으로써 아직 충분한 정보가 쌓이지 않은 것으로 보인다.

다) 예측 모델

대여-반납 건수만 존재했다면 시계열 분석까지만 시도할 수 있었다면, 본 팀의 데이터로는 다양한 정보를 포함한 변수가 존재함으로 트리 기반 모델과 딥러닝 모델까지 보다 더 예측력을 높이기 위한 시도를 할 수 있었다.

① Prophet

가장 먼저 자료가 시계열 자료라는 점을 고려하여 Facebook AI에서 개발한 가장 최근의 시계열 모델인 프로페트를 적용하여 보았다. (*Grid Search를 통해 데이터에 가장 최적화된 yearly.seasonality=F, weekly.seasonality=T, daily.seasonality=F, change_point_prior_scale=0.01, seasonality_prior_scale=100, holidays_prior_scale=10의 파라미터 사용*) 자체 내장되어있는 cv함수를 이용하여 rolling window 기법을 이용하여 MAPE값을 산출하였다. (*initial = 186, period=12, horizon = 61 사용*)

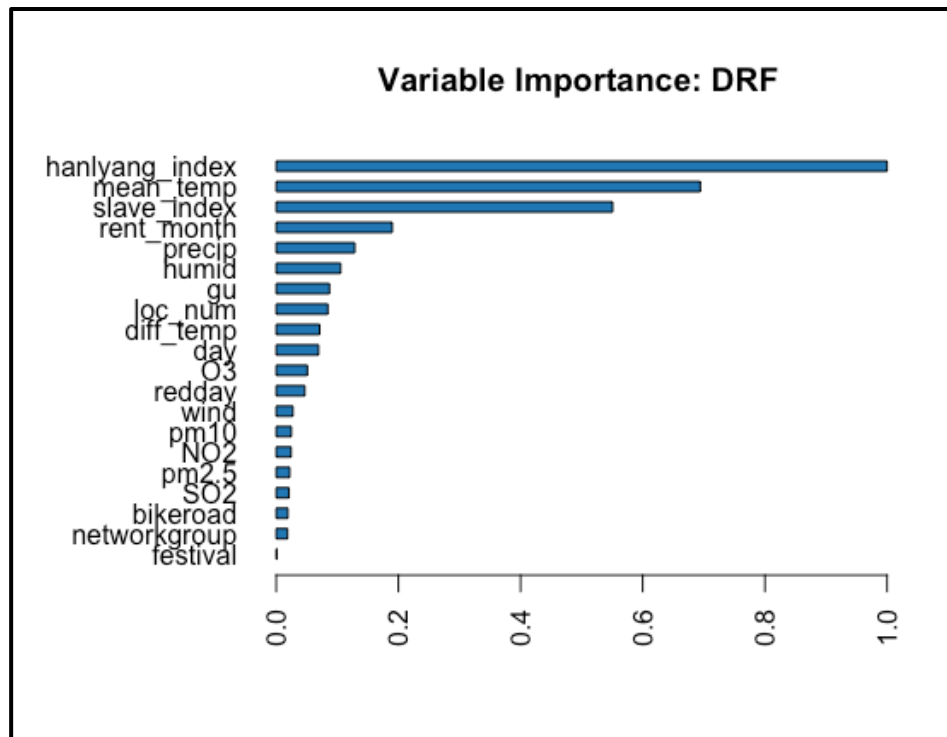
하지만 각 대여소마다 다른 시계열 흐름을 보여주므로 다른 파라미터를 적용해야하나 443개의 대여소에 대해 일괄적으로 파라미터 조정을 할 수 밖에 없었고 평균 MAPE는 0.5이상이라는 결과를 가져왔다. 가장 작은 MAPE의 값 또한 0.436이었다.

② Random Forest

트리 기반 모형 중, 비선형 관계를 설명해주는 Decision Tree모형의 variance문제가 해결된 랜덤 포레스트 모형을 사용해보았다. 컴퓨팅 시간이 빠른 h2o기반 랜덤 포레스트 모델을 적용한 결과 MAPE값은 0.2806이 나왔다. (*Grid Search를 통해 데이터에 가장 최적화된 ntree=100, mtries=9의 파라미터 사용*)

추가 해석을 위해 랜덤 포레스트 모형이 데이터를 사용할 때, error값을 가장 효율적으로 줄

여주는 변수들을 importance plot으로 나열해 보았다.



▲ 사진32. Importance plot

기온/습도/미세먼지 등 날씨와 관련된 변수들이 높은 순위를 차지하는 것을 확인할 수 있었다. 또한 장소와 날짜에 대한 변수도 상위권에 속하는 것을 보아 대여소별 특징을 파악하고자 한 본 팀의 의도가 유의미한 것으로 나타났다.

③ XGboost

랜덤 추출된 변수로 Decision tree를 만들고 합친 랜덤포레스트 모델과 다르게 에러값에 weak 모델을 적합시키는 Boosting모델을 사용해보기로 했다. 그 중에서도 높은 예측력을 자랑하는 XGBoost모델을 사용하였다. 하지만 MAPE값이 0.37로 높은 error값이 나왔다. (*Grid Search를 통해 데이터에 가장 최적화된 max_depth=14, min_child_weight=5, subsample=0.8, colsample_bytree=0.8의 파라미터 사용*)

에러 값이 높게 나온 이유는 대여소는 범주형 변수로 level이 443개나 되어 변수의 차원이 너무 커지는 것을 막기 위해 변수 제거를 해주었기 때문이다. 따라서 본 팀은 범주형 자료를 차원 축소를 통해 모델에 사용할 수 있도록 딥러닝을 사용하여 entity embedding을 사용하였다.

④ Entity Embedding

Embedding이란, 개체를 다른 위상(차원)에 끼워 넣는다는 의미로, 기존에 one-hot encoding으로 표현해야만 했던 범주형 변수들을 축소된 차원으로 표현하고자 작은 차원에 학습시키는 방법이다. 이렇게 축소된 차원은 다른 모델에도 응용가능하며, 축소된 범주형 변수와 차원은 다음과 같다.

Model	기존 차원	축소 차원
loc_num	443	10
gu	11	3
month	12	4
day	7	2
date	31	5

▲ 표12. 축소된 범주형 변수와 차원

축소된 차원의 변수를 이용하여 XGboost모델에 다시 적용시키니, 0.3262라는 MAPE 결과값이 나왔다. 이는 대여소 변수를 고려하지 않았을 때보다 고려하였을 때 훨씬 낮은 error값이 나온을 말해준다.

더불어 Neural Network을 통해 축소된 차원의 변수를 바로 사용하여 대여 건수를 예측한 결과 0.32이라는 매우 작은 MAPE 결과값이 나왔다.

라) 최종 평가

다음은 우리가 위에서 진행한 모델의 MAPE를 비교한 표이다.

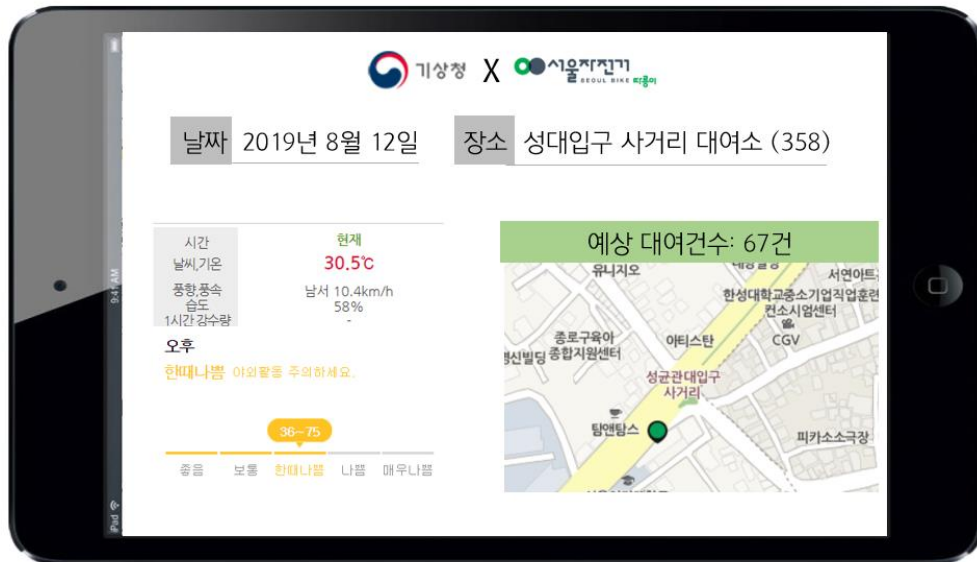
Model	5 fold train CV MAPE
프로펫 시계열 모형	0.4360
음이항 회귀 모형	0.3923
H2O Random Forest	0.2806
XG BOOST	0.37
XG BOOST + Entity Embedding	0.3262
Neural Network + Entity Embedding	0.22

▲ 표13. 평가지표 비교

주어진 train set에 대해 가장 예측 성공률이 높은 모델은 MAPE 값이 0.22로 가장 낮은 Neural Network + Entity Embedding 모델임을 확인했다. 따라서 이 모델을 통해 test set에 적용시켰을 때 최종 CV error는 0.3276으로, 과적합 되지 않은 순수한 예측 성능을 보였다.

V. 서비스 활용 방안

가) 날씨 정보로 업데이트 되는 따릉이 대여소 정보



“오늘 성대앞 대여소에는 따릉이가 충분하네! 따릉이로 하교할 수 있겠어”

▲ 사진33. 따릉이 소비자용 어플 예시

현재 사용되고 있는 대여소 현황 알리는 현재 남아있는 자전거의 개수를 알려주는 사후 알림의 역할밖에 하지 못한다. 본 팀이 개발해 낸 대여건수와 반납건수 예측모델을 사용한다면 이미 0대가 남아있기 전에, 오늘의 수요와 반납의 상태를 미리 예측하여 사전에 하루 대중교통 이용경로와 일과를 미리 계획할 수 있다.

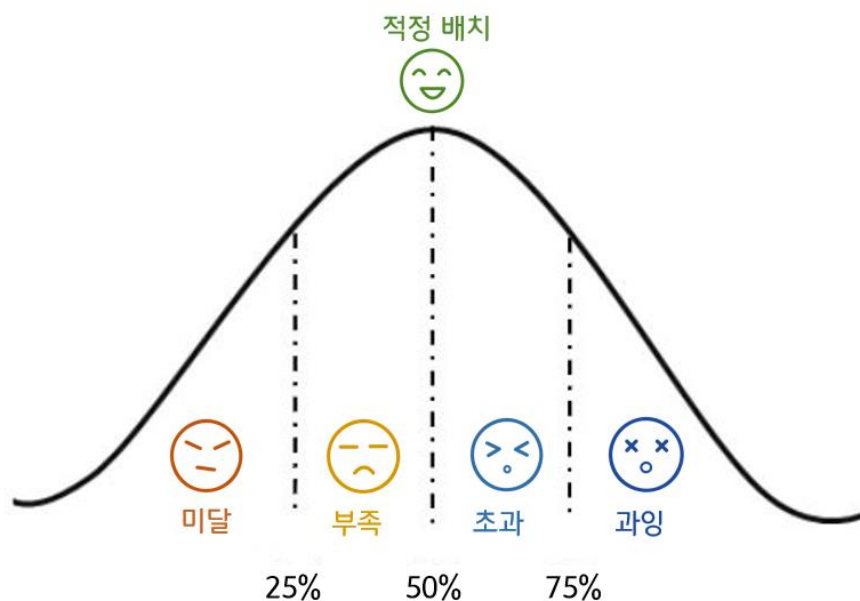


“오늘 종각역 근처 대여소는 평소보다 따릉이가 부족하겠어”

▲ 사진34. 따릉이 공급자용 어플 예시

이를 이용하여 소비자뿐만 아니라 공급자 또한 미리 24시간 전에 파악하여 대여-반납 건수의 비율이 맞아 떨어지는 중심 지점으로 이동시키기 위해 필요한 자전거의 건수를 파악할 수 있다. 대여-반납 건수의 차 (반납-대여)를 이용하여 음수가 나타난다면 추가 배치를, 양수가 나타난다면 자전거 배치를 줄일 수 있을 것이다.

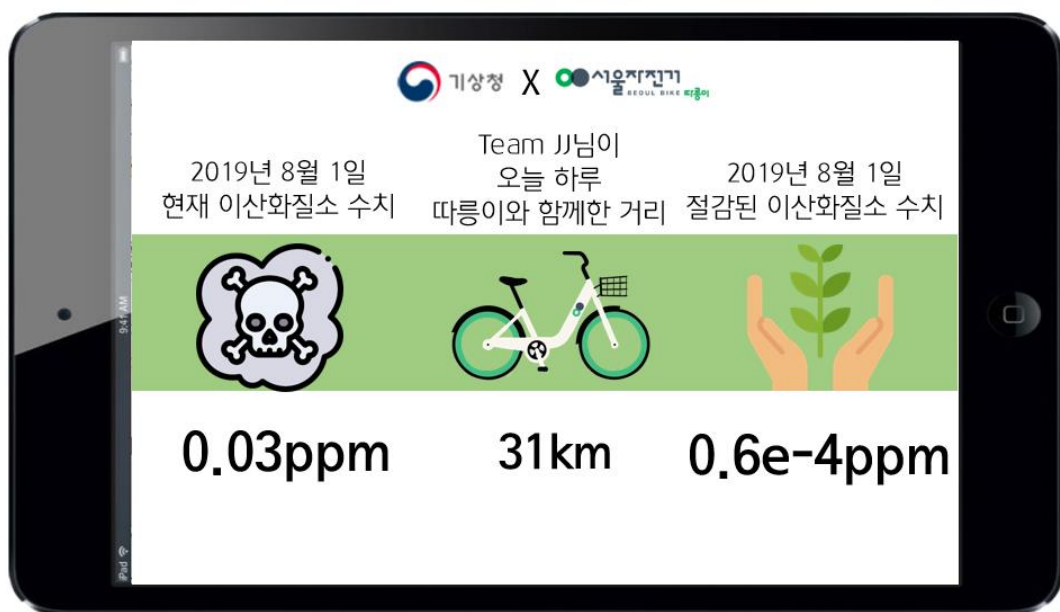
나) 따릉이 대여소 추가설치에 대한 시민 요구 수렴



▲ 사진35. 따릉이 부족경보 예시

또한 따릉이 대여소 별 수요예측을 통해 따릉이 수요가 가장 많은 지역과 버스와 지하철이 인접해있는 지역을 선별한 뒤, 순차적으로 대여소 추가 설치가 가능하다. 선별 방식으로는 대여소 별로 만들어진 대여-반납 건수의 차의 분포를 이용하여 전체 대여소 분포의 25%, 50%, 75%, 100% 분위수를 사용하여 '따릉이 부족경보'를 만들어 낼 수 있다. 따릉이 부족경보가 미달일 시, 추가 대여소 설치가 필요하다는 주장이 가능하다. 추가 설치 시, 예상되는 비용 대비 실제로 많이 사용될 것인가에 대한 조사가 필요한데, 이 때 우리의 대여-반납 건수의 예측을 이용하여 다른 사람들을 설득할 수 있다. 부족한 반납 건수의 자전거 개수만큼의 이산화탄소와 환경적 비용의 절감이 이루어진다면 대여소를 설치할 시 드는 비용을 충당할 수 있다는 근거를 내세울 수 있다.

다) 따릉이 이용으로 인한 환경적 비용 시각화



▲ 사진36. 날씨와 따릉이 상호영향 시각화 예시

날씨 일기예보와 대여소예보와 더불어 따릉이 대여-반납 건수의 예측으로 파급되는 환경비용 및 생태계의 비용을 시각화하여 따릉이 어플 및 홈페이지 혹은 기상청 홈페이지에 매일 업데이트 되어 기재된다면 공영자전거의 활발한 수요 및 이용이 증가할 것이다. 일반적으로 자동차의 평균 연비는 약 10km이며 최근 휘발유 평균 가격은 1L에 약 1200원. 1일 출퇴근으로 평균 20km를 주행한다고 할 시, 1년만 꾸준히 자전거 이용을 한다면 약 100만 원이 절약된다. 이렇게 줄어드는 경제적 비용과 더불어 1km 자전거 이용시 감축되는 이산화탄소는 0.4kg으로 낙엽송 0.5그루와 같은 효과를 가진다. 이와 같은 정보들을 자전거 예상 대여건수와 연계하여 데이터 정보를 주고 받을 수 있다면, 실제 이용자들이 따릉이를 이용함으로써

얼마나 환경에 기여하고 있는지 시각화된 정보를 통해 알게 할 수 있다.

VI. 서비스 기대효과

가) 따릉이의 효율적인 배치와 사회적 파급 효과

공공자전거 ‘따릉이’ 옮기는 ‘우렁각시’를 아세요?

출근 시간엔 지하철역 주변 자전거 부족
퇴근 시간엔 직장근처 대여소에 부족해
분배트럭으로 일일이 옮겨줘야
분배팀 “거치율 70% 유지가 목표”



▲ 사진37. 인기대여소 집중관리

따릉이 대여소 관리와 관련하여 어르신일자리 사업에 선한 영향력을 줄 수 있다. 무인대여시스템인 따릉이는 인기대여소의 경우 자전거 점검 및 불법 부착물 점검 등 대여소 환경관리에 일손이 필요한 상황이다. 이에 동작구는 어르신 일자리 사업을 따릉이로 확장하여 만 65세 이상 기초연금 수급자 12명을 대상으로 따릉이 환경 관리 사업에 들어갔다. 이 사례를 통해, 따릉이가 완전한 교통연계수단으로 자리잡기 전까지 생기는 문제들에 대해 따릉이 대여소 별 수요 예측을 통해 어느 대여소에 일손이 필요한지 파악한 뒤, 어르신일자리 사업과 연계하여 노인 복지 및 일자리 창출에 도움을 줄 수 있음을 알 수 있다.

나) 공유경제의 환경보호 역할 가시화

따릉이는 그저 서울시민의 편의만을 위한 공영 자전거 시스템으로 남기에는 도시환경 및 계획분야에 있어서 아주 큰 잠재력을 가지고 있다. OECD ITF(국제교통포럼)에 따르면, 교통수단을 공유하는 사회에 대한 결과로 현재 자동차 대수의 3%만으로도 현재의 교통 수요를 처리

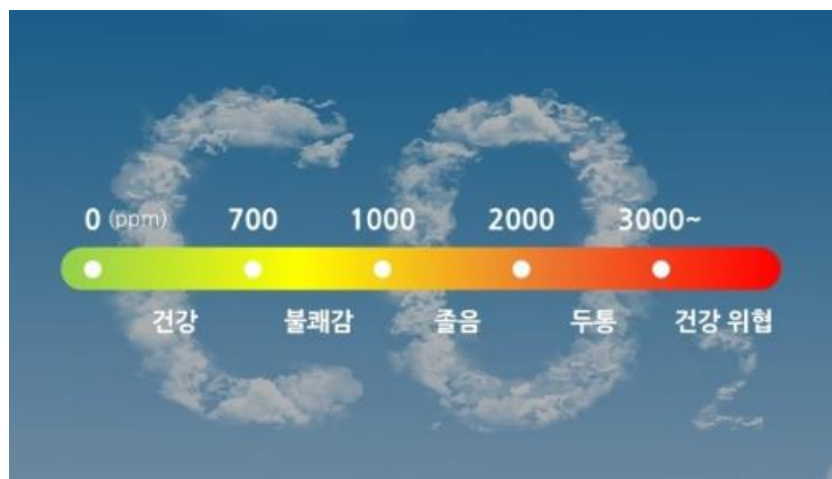
할 수 있다고 한다. 그리고 이 경우, 이동 비용의 50%가 감소하고 이산화탄소배출량은 37% 감소하며 공공주차공간의 95%가 불필요해진다. 즉, 자가용보다는 대중교통의 이용이 도시 환경과 경제에 미치는 환경 및 경제적 이익이 더 크다는 것이다. 이에 더 나아가, 따릉이가 환승 가능한 대중교통수단이 되었을 때, 도시 환경 및 경제에 미치는 영향력은 절대 무시할 수치가 아닐 것이다.

대분류	분류	도보	자전거	자전거+기타교통수단	오토바이	버스(버스간환승포함)	지하철(철도)	버스+지하철	택시	승용차
	종로구	15.5	2.7	-	1.1	37.2	6.1	16.9	0.3	18.6
	중구	20.1	1.5	-	0.5	24.5	17.3	25.9	-	7.9
	용산구	15.8	2.8	0	0.4	37.9	9.6	14.2	0.1	18.2
	성동구	21.7	3.2	0.2	1.1	18.8	14.2	15.4	0.1	22.3
	광진구	20.4	4.2	0	0.4	18.9	13.6	25.6	0.8	15
	동대문구	14.7	2.7	-	0.3	26.5	16.6	18.8	-	20
	중랑구	13.6	3	-	0.3	28.8	11.2	21.2	0.1	21
	성북구	15.2	4.6	-	0.4	28.6	8.4	25	0	13.7
	강북구	18.4	5.1	-	1.4	23.9	8	20.4	-	22.2
	도봉구	17.9	1.9	-	2.2	28.5	11	17.9	-	19.7
	노원구	16	3.2	0.1	0.3	19	18.4	16.5	0	24

▲ 표14. 서울시 통근/통학시 이용하는 교통수단 통계

하지만 따릉이 수요 증가추세와는 다르게 2018년 기준 실제로 따릉이의 도심교통 분담률이 최소 1.5%로 다른 출퇴근 교통수단과 비해 잘 쓰이지 않음을 알 수 있다. 뿐만 아니라, 확실히 다른 교통수단과 연계가 되지 않는 모습을 보인다. 버스와 지하철은 서로 유기적으로 사용되고 있지만, 따릉이의 경우 타 교통수단과의 연계성까지 처참한 상황이다.

만약 따릉이 이용이 얼마나 도시 환경에 기여하는지 따릉이 이용자가 직접 눈으로 그 수치를 확인할 수 있다면 도시 환경에 미치는 공영 자전거의 중요성을 다시금 깨닫게 하며, 이용자 본인의 환경 개선에 대한 기여, 만족도를 충족시켜 따릉이 이용의 선순환적인 구조를 만들 수 있을 것이다. 예를 들면, '오토에코'라는 앱도 존재하는데, 이 앱은 기후변화 문제에 대처하기 위해 만들어졌으며, 사용자의 생활습관 중 어느 영역이 지구 온난화에 영향을 미치는지 그래프로 보여준다.



▲ 사진38. 환경적 비용 시각화

이에 따라, 대여소별 수요예측을 통해 따릉이의 원활한 수급 및 이용의 예상수치를 활용하여 '오토에코'의 방식처럼, 사용자가 따릉이를 이용함으로써 얼마나 환경에 기여하고 있는지 시각화된 정보를 통해 알게 되는 것이다. 가령, 그리고 그 데이터를 SNS를 통해 다른 사람들과 공유하고, 현재 교통수단으로 인한 환경오염이 어느 정도인지 정확한 수치로 인지하고, 공유자전거의 이용의 중요성을 대중들에게 자연스럽게 고취시킨다면 친환경적 문화를 만드는 동시에 환경에 대한 관심 재고 및 해결 가능성에 이바지할 것이다.