

따릉이 데이터 분석 및 전략 제언

데이터 분석 및 경영 - E조
강병찬 권다인 김성필
박세웅 이태희 이현우

서울자전거
SEOUL





CONTENTS

///

01

**주제 및
선정 배경**

02

변수 설정

03

**회귀분석 및
clustering 해석**

04

한계점





1. 주제 및 선정 배경



서울시 자전거 따릉이



“ 서울시의 **교통체증, 대기오염, 고유가 문제**를 해결하고
건강한 사회 및 시민들의 삶의 질을 높이고자 마련됨.

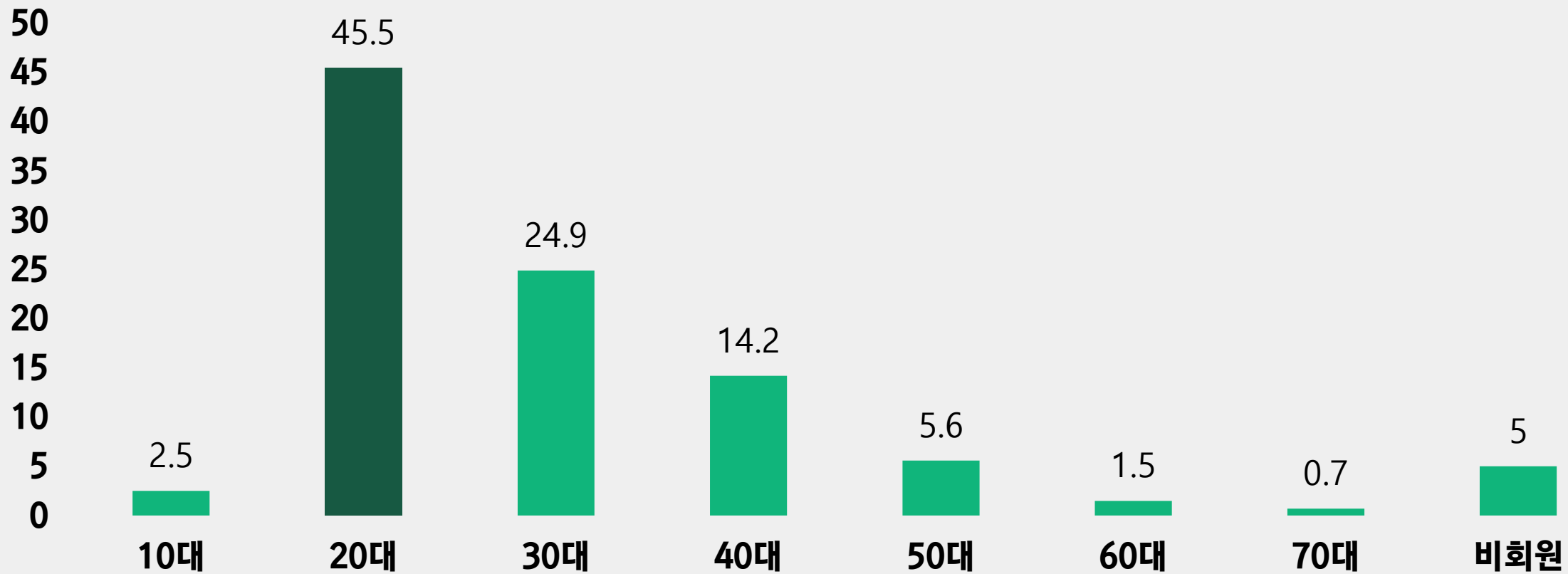
녹색 성장
탄소 배출 저하

자전거로 지키는
건강한 도시

편리한
무인대여 시스템



파름이 연령대별 이용 비율



단위 - 퍼센트(%)

출처 - 서울시설공단



매년 늘고 있는 파릉이 이용건수

2016년 161만 건(하루 4,403건)에서 2017년 503만 건(하루 13,784건)으로 1년 만에 **3배 이상 증가**



이용자 절반은 20대, 여성보다 남성이 많고 회원이용자가 94.8%

2017년 이용자 연령대는 **20대가 절반가량**(45.5%), 이어 30대(24.9%), 40대(14.2%) 순이고
남성(59.4%), 여성(36.8%)



평일 퇴근시간대(18시~21시), 휴일 오후시간대(15~18시)이용자 많아

요일별로 이용자가 금요일(15,162건)이 가장 많고 일요일(12,681건)이 최소

스마트 모빌리티의 대두

스마트 모빌리티

첨단 기술이 융합되고 보다 지능화된 교통 서비스,
1~2인승 개념의 소형 개인 이동 수단에 집중되어 있으며
공영화된 사례는 "서울시 따릉이"가 대표적

미래형 교통수단 '스마트 모빌리티' 시대

한국경제매거진 2017/08

전기자전거 '보급형으로' 다시 예 열

헤럴드경제 2017/06

공공자전거 따릉이 2만대로 늘어난다...서울 전역서

연합뉴스 2017/07

차세대 이동수단 '스마트 모빌리티' / YTN 사이언스

YTN 2015/08

누구나 편하게 이동하는 스마트 모빌리티 시티

건설경제 2016/09

일상에 자리 잡은 스마트 모빌리티

스마트PC 2017/06

서울 시민의 발 따릉이, 이용 간편해진다

자전거뉴스 2017/07



주제 제시

“ **따름이 이용자 분석을 통한 따름이의
지속적인 성장 방안 제시** ”



데이터 전처리

RULE1

5개 구의 행정동 별로 구분하여 전처리.

RULE2

**출근 지역(회사 지역) – 종로구, 중구, 영등포구/퇴근 지역(비회사 지역)–서대문구, 마포구
한강 지역–마포구, 영등포구/비한강 지역–종로구, 중구, 서대문구**



테이터 전처리

RULE1

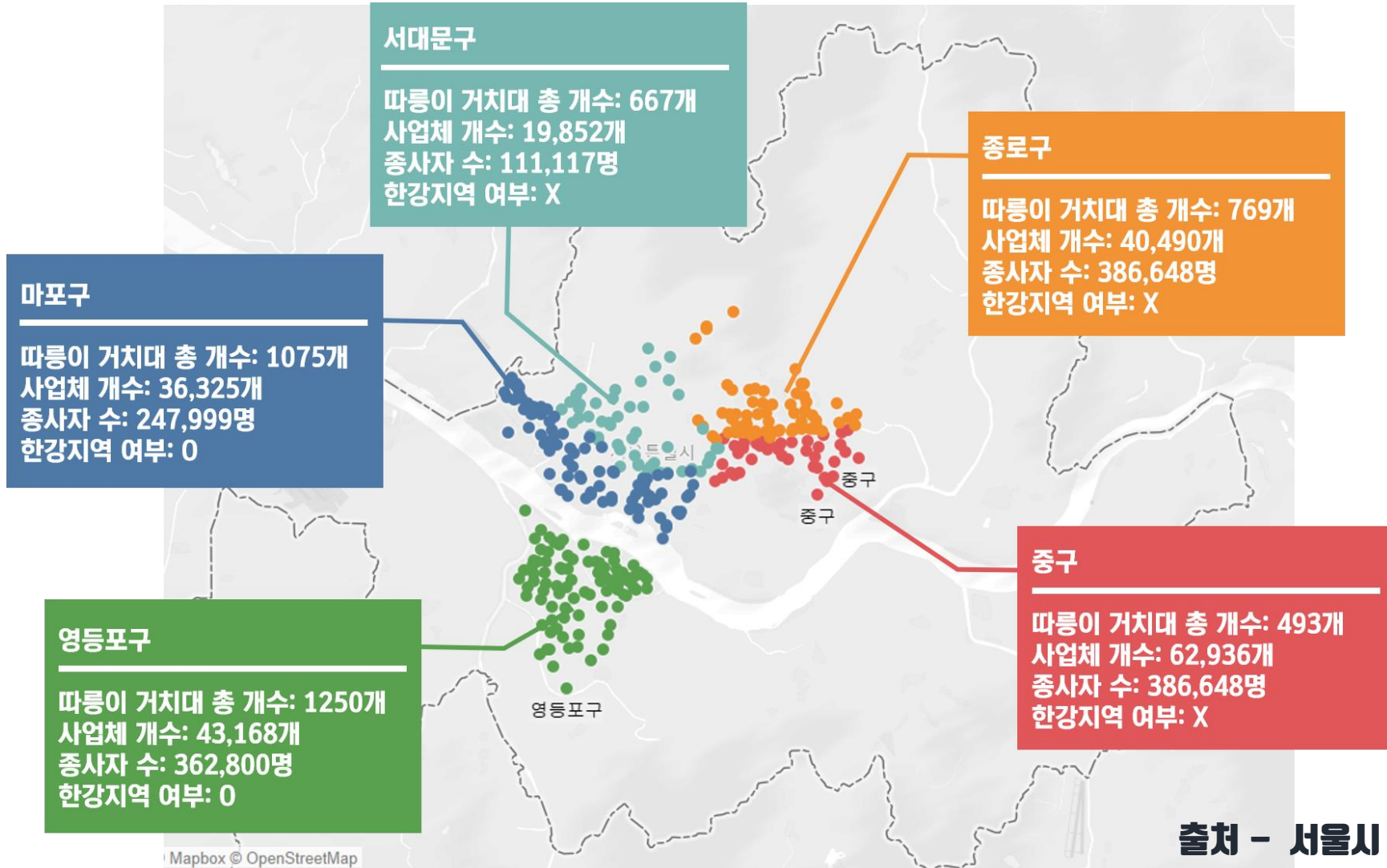
5개 구의 행정동 별로 구분하여 전처리.

RULE2

출근 지역(회사 지역) – 종로구, 중구, 영등포구/퇴근 지역(비회사 지역)–서대문구, 마포구
한강 지역–마포구, 영등포구/비한강 지역–종로구, 중구, 서대문구



구별 정보(따릉이 거치대 수, 사업체 수, 종사자 수, 한강지역 여부)





2. 변수 설정





변수 설정



종속 변수

따릉이 이용자 건 수

독립 변수

- | | |
|-------------|------------------------|
| 1인 가구 수 | - 1인가구의 따릉이 열람 |
| 한강공원 유무 | - 따릉이의 주요 주행 장소 |
| 버스 이용자 건 수 | - 대중교통으로서의 따릉이 |
| 따릉이 거치대 수 | - 거치대 수와 이용자의 관계 |
| 지하철 이용자 건 수 | - 대중교통으로서의 따릉이 |
| 주차장 개수 | - 자가용 이용과 따릉이의 관계 |
| 가로수 현존 개수 | - 가로수 길에서 따릉이 남만을 |
| 자전거 전용도로 길이 | - 따릉이로 이동할 수 있는 길 (KM) |

주택

아파트

연립주택

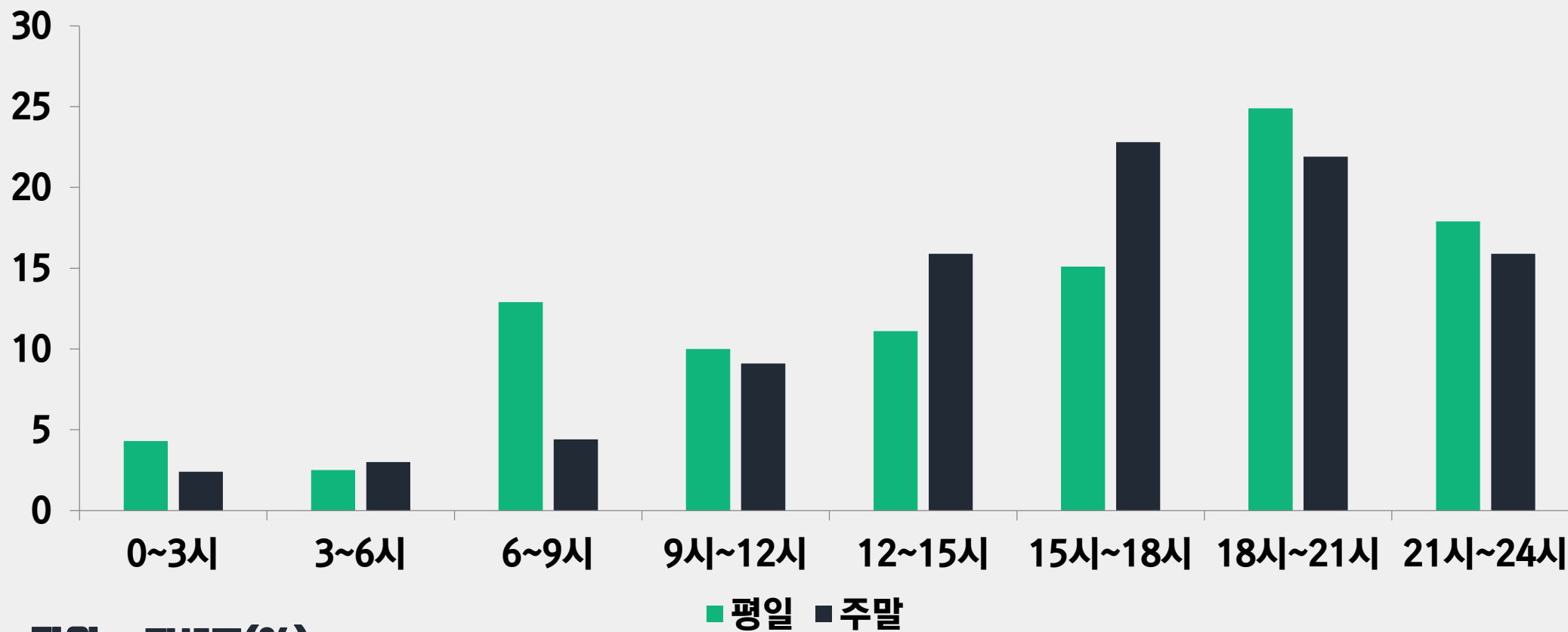
다세대 주택

비거주용 건물내 주택

- 주택 타입과 따릉이 이용 간의 상관관계 관찰



/// 파릉이 시간대별 이용 비율



단위 - 퍼센트(%)

출처 - 서울시설공단

3. 회귀분석 및 clustering 해석

서울자전거
SEOUL BIKE

LLL



3.1 회귀분석 결과 : 출퇴근 시간

회사 지역

```
Call:
lm(formula = 따릉이이용자 ~ x1인가구 + 버스 + 주택 + 거주공간 +
  가로수 + 자전거전용도로 + 따릉이거치대수 + 지하철이용자,
  data = 평일출근_df_n_a)

Residuals:
    Min       1Q   Median       3Q      Max
-0.087238 -0.027331  0.001149  0.024465  0.079745

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.17581    0.04656   3.776  0.00183 **
x1인가구     -0.12408    0.03257  -3.809  0.00171 **
버스          -0.35585    0.10277  -3.462  0.00348 **
주택         -0.05400    0.03759  -1.437  0.17135
거주공간     -0.06990    0.04148  -1.685  0.11267
가로수        0.18622    0.10418   1.788  0.09406 .
자전거전용도로 0.04991    0.03722   1.341  0.19986
따릉이거치대수 0.97613    0.14282   6.835 5.65e-06 ***
지하철이용자  -0.07670    0.06820  -1.125  0.27845
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0463 on 15 degrees of freedom
Multiple R-squared:  0.9647,    Adjusted R-squared:  0.9459
F-statistic: 51.25 on 8 and 15 DF,  p-value: 1.648e-09
```

1% 유의수준에서
통계적으로 유의함.

따라서 따릉이 이용자와 출근
시간 버스 이용자는 유의 상관
관계가 있음

출퇴근시간 회사지역에서는
정류장 및 지하철역이
회사 건물과 비교적 가깝게
위치하고 있기 때문에
따릉이를 이용할
필요가 없음

3.1 회귀분석 결과 : 출퇴근 시간

비회사 지역

```
Call:
lm(formula = 따릉이이용자 ~ x1인가구 + 한강공원 + 버스 + 주차장 +
    주택 + 아파트 + 연립주택 + 다세대주택 + 가로수 + 따릉이거치대수 +
    지하철이용자, data = 평일출근_df_n_b)
```

Residuals:

충현동	천연동	신촌동	홍제제2동	홍제제3동	남가좌제1동	아현동
-0.0036191	0.0009003	-0.0102565	-0.0064341	0.0136875	0.0057219	-0.0138989
공덕동	도화동	용강동	대흥동	서강동	서교동	합정동
0.0022671	-0.0078665	0.0086990	0.0096175	0.0049316	0.0027644	-0.0070336
성산제2동						
0.0005194						

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.0273487	0.0310615	-0.880	0.44343
x1인가구	-0.1508861	0.0578278	-2.609	0.07974 .
한강공원	0.0007924	0.0129860	0.061	0.95518
버스	0.0459409	0.0375879	1.222	0.30889
주차장	0.1061702	0.0397899	2.668	0.07580 .
주택	0.0477332	0.0357497	1.335	0.27408
아파트	-0.1430310	0.0340253	-4.204	0.02458 *
연립주택	-0.0686475	0.0507666	-1.352	0.26921
다세대주택	0.0884390	0.0428147	2.066	0.13079
가로수	0.2284058	0.1098155	2.080	0.12901
따릉이거치대수	0.9467935	0.1532822	6.177	0.00854 **
지하철이용자	0.2348230	0.0698551	3.362	0.04368 *

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

Residual standard error: 0.01723 on 3 degrees of freedom
Multiple R-squared: 0.993, Adjusted R-squared: 0.9673
F-statistic: 38.7 on 11 and 3 DF, p-value: 0.005944

1% 유의수준에서
통계적으로 유의함.
따라서 따릉이 이용자와
출근시간 지하철 이용자는
양의 상관관계가 있음

출퇴근시간 비회사지역에서는 따릉이를
타고 대중교통을 이용하는 곳까지
이동하는 경우가 있으므로 따릉이 이용자
와 지하철이용자 간에 양의 상관관계를
보이는 것으로 판단됨

Implication

따라서 주택가가 많고
지하철역(또는 버스 정류장)과
거리가 먼 곳에 따릉이를
설치하는 것이 필요함

3.1 회귀분석 결과 : 평일

```
Call:
lm(formula = 따릉이이용자 ~ ., data = 평일_df_nnew)

Residuals:
    Min       1Q   Median       3Q      Max
-0.122670 -0.022429 -0.003062  0.025593  0.095056

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.065130   0.041079   1.585  0.12495
x1인가구    -0.110217   0.034719  -3.175  0.00384 **
한강공원     -0.052116   0.064902  -0.803  0.42926
버스          -0.054652   0.054096  -1.010  0.32167
주차장       0.039445   0.050662   0.779  0.44325
주택        -0.007033   0.056322  -0.125  0.90158
아파트      -0.036360   0.038206  -0.952  0.35002
연립주택     0.081996   0.061618   1.331  0.19484
다세대주택  -0.035749   0.070093  -0.510  0.61434
가로수       0.179111   0.096975   1.847  0.07616 .
자전거전용도로 0.052365   0.082197   0.637  0.52966
따릉이거치대수 0.782674   0.125472   6.238 1.34e-06 ***
지하철이용자  0.003361   0.067243   0.050  0.96051
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0564 on 26 degrees of freedom
Multiple R-squared:  0.9171,    Adjusted R-squared:  0.8788
F-statistic: 23.96 on 12 and 26 DF,  p-value: 5.033e-11
```

5% 유의수준에서
통계적으로 유의함.
따라서 따릉이 이용자와
가로수는 양의 상관관계
가 있음

평일 따릉이 이용자와
가로수 개수 및 따릉이 이용
자와 따릉이 거치대수는
각각 유의수준 5% 및
1%에서 통계적으로
유의하므로 양의 상관관계가
있음

3.1 회귀분석 결과 : 평일

	A	F	G	H
1	행정동	따릉이 거치대수	가로수	가로수/거치대수
2	공덕동	70	664	9.4857142857
3	광희동	8	393	49.125
4	남가좌제1동	35	1148	32.8
5	당산제1동	50	780	15.6
6	당산제2동	138	1198	8.6811594203
7	대림제3동	26	564	21.692307692
8	대흥동	77	582	7.5584415584
9	도화동	7	270	38.571428571
10	명동	25	1153	46.12
11	문래동	58	1897	32.706896552
12	사직동	35	2042	58.342857143
13	서강동	40	993	24.825
14	서교동	132	2499	18.931818182
15	성산제2동	107	813	7.5981308411
16	소공동	8	1689	211.125
17	송인제1동	20	72	3.6
18	신길제1동	10	652	65.2
19	신길제6동	30	564	18.8
20	신길제7동	37	211	5.7027027027
21	신촌동	138	1228	8.8985507246
22	아현동	11	332	30.181818182
23	양평제1동	65	670	10.307692308
24	양평제2동	40	778	19.45
25	여의동	504	6223	12.347222222
26	영등포동	121	2118	17.504132231
27	영등포본동	50	11	0.22
28	용강동	11	968	88
29	을지로동	18	1300	72.222222222
30	이화동	35	24	0.68571428571
31	장충동	28	420	15
32	종로1,2,3,4	116	2915	25.129310345
33	창신제1동	8	53	6.625
34	천연동	15	135	9
35	충현동	43	610	14.186046512
36	필동	20	672	33.6
37	합정동	15	2364	157.6
38	홍제제2동	8	378	47.25
39	홍제제3동	8	450	56.25
40	회현동	5	292	58.4
41		평균		34.854465794

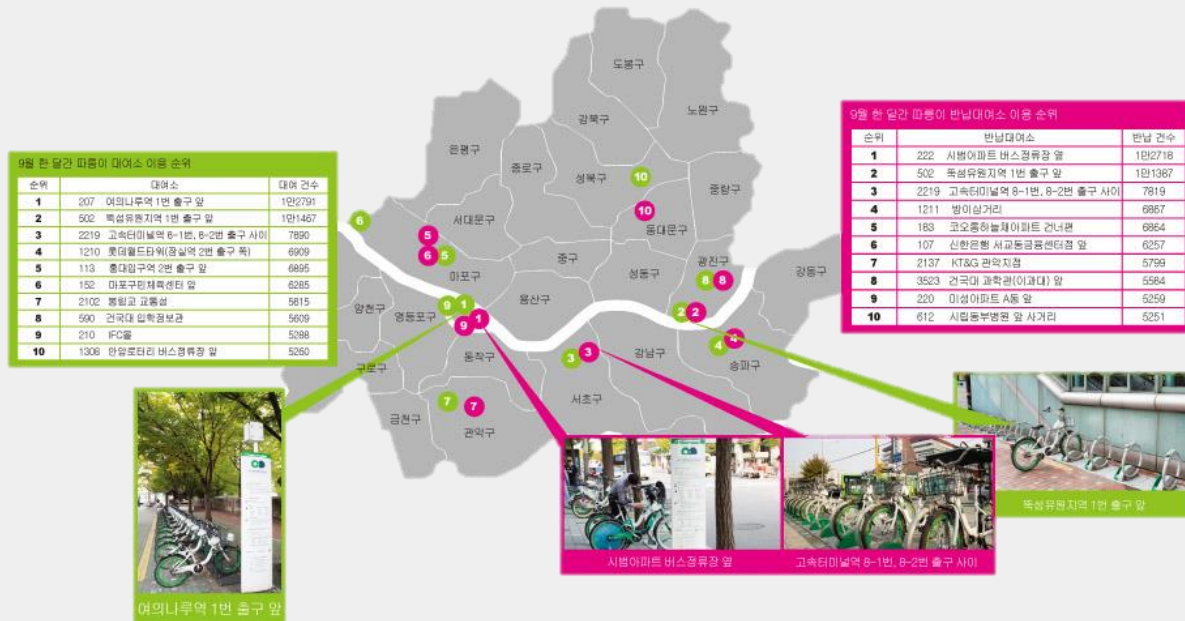
▲따릉이 거치대수 및 가로수 개수

- 가로수가 많다는 것은 인근 지역의 도로가 잘 형성되어 있다는 것으로 판단할 수 있으며, 자전거 전용도로 또한 설치 가능할 것으로 예상됨
- 지역별로 따릉이 거치대 개수가 적절히 설치되었는지 알아보기 위해 서울시 공공 데이터를 활용하였음

Implication

- 가로수 개수에 비해 따릉이 거치대 수가 적은 지역에 따릉이 거치대 설치
- 자전거 전용로 확충

3.1 회귀분석 결과 : 주말(한강공원)



- 파종이 이용 빈도가 가장 높은 대여소는 한강공원 인근 지역(여의도)으로 집계됨(서울시, 2018)
- 2018년 9월 한 달 동안 서울 시민들이 파종이를 많이 빌려간 대여소 1위는 서울지하철 5호선 ‘여의나루역 1번 출구 앞’으로 1만2791건
- 서울 시민들이 파종이를 많이 반납한 대여소 1위는 여의동 ‘시범아파트 버스정류장 옆’으로 1만2718건으로 집계됨
- 마포대교와 원효대교 사이에 자리해 파종이로 다리를 건너려는 이용자 수도 많은 것으로 집계됨
- 이용 빈도가 높은 대여소는 여의도 한강공원과 이어져 있었음
- 여의도 지역은 자전거 거치대만 50여 개로 다른 대여소에 비해 압도적으로 많았음
- 또한, 여의도 지역은 도로가 모두 평지인 데다 대부분 자전거도로가 마련돼 있어 대중교통보다 접근성이 더 좋은 지역임

3.1 회귀분석 결과 : 주말(한강공원)

```
> summary(modelb)

Call:
lm(formula = 따름이이용자 ~ ., data = 주말_df_n_b)

Residuals:
    Min       1Q   Median       3Q      Max
-0.080171 -0.026412  0.007545  0.025980  0.090976

Coefficients: (1 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.76356    0.41979   1.819  0.10642
x1언가구      -0.24139    0.10806  -2.234  0.05597 .
한강공원       NA         NA      NA      NA
버스          -0.33031    0.13036  -2.534  0.03504 *
주차장        0.37773    0.11350   3.328  0.01042 *
주택         -0.01939    0.07388  -0.262  0.79963
아파트       -0.31746    0.07881  -4.028  0.00380 **
연립주택     0.23510    0.08703   2.701  0.02702 *
다세대주택   0.06979    0.09421   0.741  0.47998
비거주용건물내주택 0.28522    0.08766   3.254  0.01164 *
가로수      -0.28750    0.21608  -1.330  0.22003
자전거전용도로 -0.78697    0.37408  -2.104  0.06854 .
따름이거처대수 1.54786    0.29580   5.233  0.00079 ***
지하철이용자  0.04849    0.16304   0.297  0.77372
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07291 on 8 degrees of freedom
Multiple R-squared:  0.9702,    Adjusted R-squared:  0.9256
F-statistic: 21.73 on 12 and 8 DF,  p-value: 8.769e-05
```

```
> summary(modela)

Call:
lm(formula = 따름이이용자 ~ ., data = 주말_df_n_a)

Residuals:
    광희동    남가좌제1동    명동    사직동    소공동    송인제1동    신촌동    을지로동    이화동
-0.066329  -0.032249  -0.029363  0.047411  0.034292  0.062642  0.024319  -0.052674  0.025860
    장충동    종로1.2.3.4가동    창신제1동    천연동    충현동    필동    홍제제2동    홍제제3동    회현동
-0.082345  -0.029687  -0.086994  0.025642  -0.035517  0.171538  -0.013891  -0.008212  0.045556

Coefficients: (1 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.217209    0.165489   1.313  0.2464
x1언가구      -0.111113    0.093330  -1.191  0.2873
한강공원       NA         NA      NA      NA
버스          -0.271942    0.249107  -1.092  0.3248
주차장        -0.123417    0.358300  -0.344  0.7445
주택         0.003554    0.103106   0.034  0.9738
아파트       -0.076125    0.146134  -0.521  0.6247
연립주택     -0.039812    0.118061  -0.337  0.7497
다세대주택   -0.104452    0.182324  -0.573  0.5915
비거주용건물내주택 0.144367    0.122047   1.183  0.2900
가로수       0.050581    0.426619   0.119  0.9102
자전거전용도로 -0.090933    0.220587  -0.412  0.6972
따름이거처대수 1.551560    0.750769   2.067  0.0936 .
지하철이용자  -0.135876    0.186543  -0.728  0.4990
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1154 on 5 degrees of freedom
Multiple R-squared:  0.7368,    Adjusted R-squared:  0.105
F-statistic: 1.166 on 12 and 5 DF,  p-value: 0.4638
```

- 코딩 시 한강공원이 있는 구(영등포구/마포구)와 없는 구(서대문구/종로구/중구)에 각각 1 및 0을 넣어 분석함
- 1은 자전거를 타고 갈만한 거리에 한강공원이 있다는 것을 의미함

3.1 회귀분석 결과 : 주말(한강공원)

```
> summary(modelb)

Call:
lm(formula = 따릉이이용자 ~ ., data = 주말_df_n_b)

Residuals:
    Min       1Q   Median       3Q      Max
-0.080171 -0.026412  0.007545  0.025980  0.090976

Coefficients: (1 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.76356    0.41979   1.819  0.10642
x1인가구     -0.24139    0.10806  -2.234  0.05597 .
한강공원      NA         NA      NA      NA
버스          -0.33031    0.13036  -2.534  0.03504 *
주차장        0.37773    0.11350   3.328  0.01042 *
주택         -0.01939    0.07388  -0.262  0.79963
아파트       -0.31746    0.07881  -4.028  0.00380 **
연립주택     0.23510    0.08703   2.701  0.02702 *
다세대주택   0.06979    0.09421   0.741  0.47998
비거주용건물내주택 0.28522    0.08766   3.254  0.01164 *
가로수       -0.28750    0.21608  -1.330  0.22003
자전거전용도로 -0.78697    0.37408  -2.104  0.06854 .
따릉이거치대수 1.54786    0.29580   5.233  0.00079 ***
지하철이용자  0.04849    0.16304   0.297  0.77372
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07291 on 8 degrees of freedom
Multiple R-squared:  0.9702,    Adjusted R-squared:  0.9256
F-statistic: 21.73 on 12 and 8 DF, p-value: 8.769e-05
```

주말 주차장과 지하철 이용자는
따릉이 이용자 수와 양의 상관관계를 가짐

3.1 회귀분석 결과 : 주말(지하철)



- 서울시에 따르면, 따봉이 이용 상위 지역은 지하철역 인근 지역으로 조사됨(여의도, 독섬 등)
- 독섬유원지역 1번 출구 앞은 평일 낮인데도 따봉이 대여소에 비치된 자전거가 1대밖에 없을 정도로 인기가 높은 곳임
- 한강공원 내 지하철역 인근 지역은 이용자 수 및 이용 빈도가 꾸준히 증가하고 있음

Implication

한강공원 내에서 지하철역 출구에 따봉이 거치대 수를 늘린다면
한강공원 이용객수가 많은 휴일에도 시민들의 따봉이 수요를
충족시킬 수 있을 것으로 판단됨

3.1 회귀분석 결과 : 주말(주차장)



- 서울시에 따르면, 여가를 즐기기 위한 목적으로 파퓰어를 이용하는 시민이 많은 것으로 집계되었음
- 파퓰어 대여 4위를 기록한 ‘롯데월드타워(1210)’, 반남 4위인 ‘방이삼거리(1211)’는 모두 석촌호수 인근임
- 한강공원 이용 목적으로만 파퓰어를 대여하는 것은 아닌 것으로 조사되었으며, 한강을 둘러싼 서울 지역은 교통체증이 많이 발생하는 지역으로 시민들이 주차장에 차를 정차해 놓고 파퓰어를 이용하여 약속 장소까지 이동하는 경우가 많은 것으로 파악되었음



Implication

한강 공원 내 혹은 근처 주차장에 파퓰어 거치대 추가를 통한 파퓰어 이용자 수 증가

3.1 회귀분석 결과 : 다중공선성 문제 - VIF

평일 출퇴근 (회사)

```
> vif(modela)
x1인가구      버스      주택      거주공간      가로수
1.131251      3.927895      1.153390      1.402172      5.345372
자전거전용도로  따릉이거치대수  지하철이용자
3.004174      8.994998      2.503161
```

평일 출퇴근 (비회사)

```
> vif(modelb)
x1인가구      한강공원      버스      주차장      주택
11.076473      2.044847      4.739906      5.137798      6.652875
아파트      연립주택      다세대주택      가로수  따릉이거치대수
4.723679      13.264640      10.973485      7.274017      9.540006
지하철이용자
5.681892
```

평일

```
> vif(model)
x1인가구      한강공원      버스      주차장      주택
1.296465      12.835723      1.816717      1.626794      2.830578
아파트      연립주택      다세대주택      가로수  자전거전용도로
1.419460      2.528519      3.267397      3.684483      13.210697
따릉이거치대수  지하철이용자
5.305680      2.185576
```

주말(한강)

```
> vif(modelb)
x1인가구      버스      주차장      주택
2.341142      3.116164      3.337656      1.660575
아파트      연립주택      다세대주택  비거주용건물내주택
1.891524      2.284235      3.291591      2.321161
가로수      자전거전용도로  따릉이거치대수  지하철이용자
8.256068      3.084547      14.873180      4.954761
```

VIF가 대부분 10 이하의 값을 가짐 -> 다중공선성의 문제 가능성 미미

3.2 Clustering 결과

대여소번호	대여소명	대여구분코드	성별	연령대코드	이용건수	운동량	탄소량	이동거리	이동시간
300	정동사거리	정기권	F	~10대	1	37.32	0.34	1450	23
505	자양사거리 광진아크로텔 앞	정기권	F	20대	1	54.83	0.49	2130	15
1531	미아사거리 1번 출구	정기권	F	20대	1	78.34	0.73	3140	17
516	광진메디칼 앞	일반권	M	20대	1	29.95	0.29	1240	5
3508	화양사거리	정기권	F	20대	1	46.09	0.45	1940	43
315	신한은행 안국역지점 옆	정기권	F	20대	1	141.09	1.18	5090	36
320	을지로입구역 4번출구 앞	정기권	F	20대	1	77.34	1.01	4340	30

*운동량= 이동거리 * 사용자 체중(미입력시 65kg로 계산) ex) 5.94 / 15

*탄소량(kg)= 이동거리 * 0.232
탄소량의 뜻은, 해당 자전거로 대여한 이동거리를 자동차로 이동했을 시에 발생하는 이산화탄소의 량(탄소절감량)

*이동거리 단위=미터(m) / 이동시간 단위=분

*연령대코드 -> 10대=1, 20대=2, 30대=3, 40대=4, 50대=5, 60대=6, 70~=7로 변환

연령대코드, 이용건수, 운동량, 탄소량, 이동거리 변수 Normalize 후 K=2로 Clustering 진행

*K=2~7일 때 모두 진행하였으나 K=2일 때 고객 군의 특징 설명

3.2 Clustering 결과

각 cluster의 변수에 대한 평균

```
> aggregate(연령대코드 ~ cluster, 따릉이_clustering, mean)
cluster 연령대코드
1      1      5.392519
2      2      2.785110
> aggregate(이용건수 ~ cluster, 따릉이_clustering, mean)
cluster 이용건수
1      1 1.024874
2      2 1.082443
> aggregate(운동량 ~ cluster, 따릉이_clustering, mean)
cluster 운동량
1      1 102.22767
2      2  96.45872
> aggregate(이동거리 ~ cluster, 따릉이_clustering, mean)
cluster 이동거리
1      1 3786.496
2      2 3504.516
> aggregate(탄소량 ~ cluster, 따릉이_clustering, mean)
cluster 탄소량
1      1 0.8784710
2      2 0.8130381
```



평균	Cluster1	Cluster2
연령대	4~60대	2~30대
이용건수	1.024874	1.082443
운동량	약 102.2	약 96.6
이동거리	약 3786m	약 3505m
탄소절감량	약 0.878	약 0.813

각 cluster의 변수에 대한 합

```
> aggregate(이용건수 ~ cluster, 따릉이_clustering, sum)
cluster 이용건수
1      1 33045
2      2 232603
> aggregate(탄소량 ~ cluster, 따릉이_clustering, sum)
cluster 탄소량
1      1 28324.54
2      2 174711.31
```



합	Cluster1	Clustert2
이용건수	33045건	232603건
탄소절감량	약 28325	약 174711

3.2 Clustering 결과

각 cluster의 변수에 대한 평균

평균	Cluster1	Cluster2
연령대	4~60대	2~30대
이용건수	1.024874	1.082443
운동량	약 102.2	약 96.6
이동거리	약 3786m	약 3505m
탄소절감량	약 0.878	약 0.813

각 cluster의 변수에 대한 합

합	Cluster1	Clustert2
이용건수	33045건	232603건
탄소절감량	약 28325	약 174711

Cluster1 특징:

- 연령대 평균이 50대
- 운동량 및 이동거리가 cluster2에 비해 높음

Cluster2 특징:

- 연령대 평균이 20~30대
- 이용 건수 및 탄소절감량의 합이 cluster1에 비해 8배 가량 됨

3.2 Clustering 결과

압도적인 이용 건수에 의해 clsuter2를 타겟으로 한 정기권 제도가 이미 도입된 상태

상품안내

구분	정기권			일일권(회원)		일일권(비회원)	
	종별	일반권 (60분)	프리미엄권 (120분)	일반권 (60분)	프리미엄권 (120분)	일반권 (60분)	프리미엄권 (120분)
상품	7일권	3,000원	4,000원	1,000원	2,000원	1,000원	2,000원
	30일권	5,000원	7,000원				
	180일권	15,000원	20,000원				
	365일권	30,000원	40,000원				
결제	휴대폰, 신용카드, 모바일 티머니			휴대폰, 신용카드, 모바일 티머니		휴대폰, 신용카드	
	· 일반권						

정기권을 구매하여 사용할 경우 이용 요금을 대폭 감소시킬 수 있음.
타릉이를 정기적으로 이용하는 고객들에게 유용

타릉이가 2030 틈새 교통수단이
라는 점에서, cluster2는 (실제)
기존 타겟층
그러나 이미 cluster2의 타릉이
이용률이 전체의 70%에 달하고,
누적 인원이 계속해서 증가하고
있는 상태이기 때문에
이제는 새로운 타겟층 유입을 위
한 고객 확대 전략 필요

3.2 Clustering 결과

새로운 전략 제안

운동량 및 이동거리, 탄소절감량이 높은 cluster1에 초점을 맞춘 마케팅 전략

For example)

1시간 단위로 전체 평균 이동거리를 기준으로 측정한 거리인 3550m 이상 이동 시 다음 이용 할인권이나 일일 1시간 이용권을 지급

예상 결과

- 쿠폰 지급을 통해서 고객들의 추가적인 파름이 이용을 유도 가능
- 탄소절감량을 더욱 낮추게 함으로서 친환경적인 캠페인 가능
- 회원가입 고객 유치 가능



4. 한계점





4. 한계점

✓ **유의미하지 않은 P-value 값**

✓ **상관관계 분석과 실제 모델**

- 1인 가구, 주차장



1. 최대한 고려했으나, 데이터와 변수 개수 부족

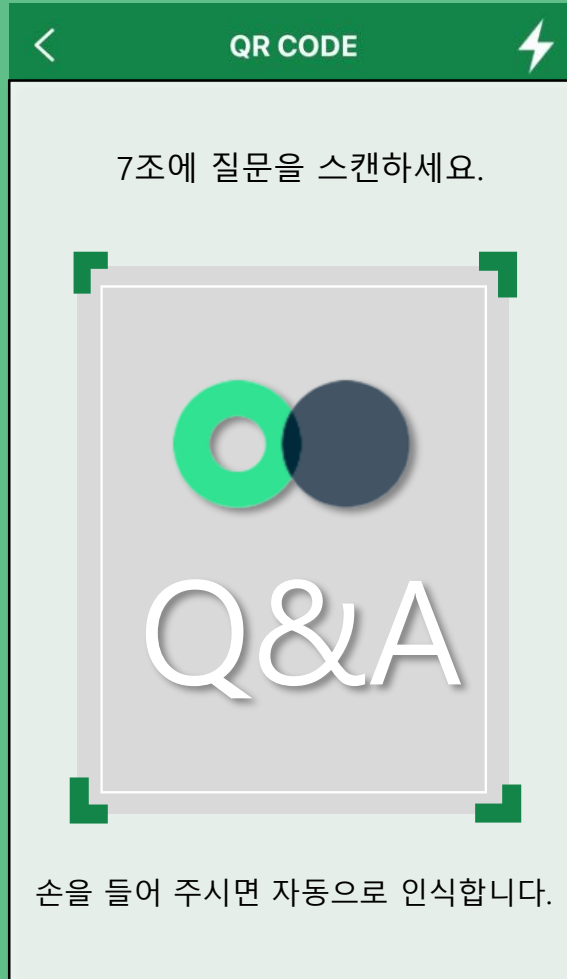
2. 상관 관계

데이터 전처리에서 나온 상관관계 결과와 실제 모델에서 차이가 있었음.

1인 가구) 상관관계 양으로 높을 것이라 예상,
But 실제 모델에서는 음의 상관 관계
: 가구의 구성원 수보다는 연령대와 더 관계가 있을 것

주차장) 상관관계가 낮을 것이라 예상,
but 실제 모델에서는 유의미 했음
: '한강공원'과 '주차장'의 관계 존재







THANK YOU