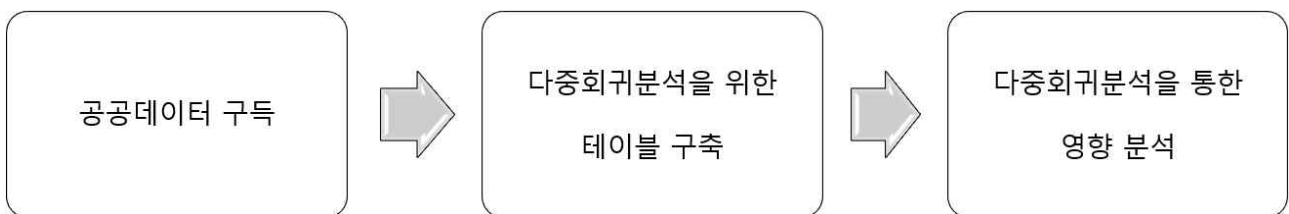


연구 제목	다중회귀분석을 활용한 공공자전거 이용에 날씨가 미치는 영향 분석
제 출 자	제태성, 정홍, 김상우, 이강신

1. 연구배경 및 목적

전 세계적으로 기후 변화 및 화석연료 고갈 등과 대도시 교통 혼잡 및 공기오염 등에 대해 해결책으로 제시되고 있는 것 중 하나인 자전거 이용이 주목받고 있음, 이미 전 세계 100개 이상의 도시에서 공공자전거 서비스를 운영하고 있으며, 이는 자전거 이용 활성화를 위한 효과적인 정책으로 평가받고 있음. 우리나라의 자전거 이용률은 이미 자전거 이용이 활성화되어 있는 국가들에 비해 낮은 수준이다. 자전거 활성화 정책이 효과적으로 수립되고 시행되려면 자전거 이용에 미치는 여러 요인들에 대한 분석이 선행되어야 한다. 이에 본 연구에서는 자전거 이용에 영향을 미치는 여러 요인들 중 날씨가 미치는 영향을 서울시를 대상으로 분석하고자 한다.

2. 연구방법

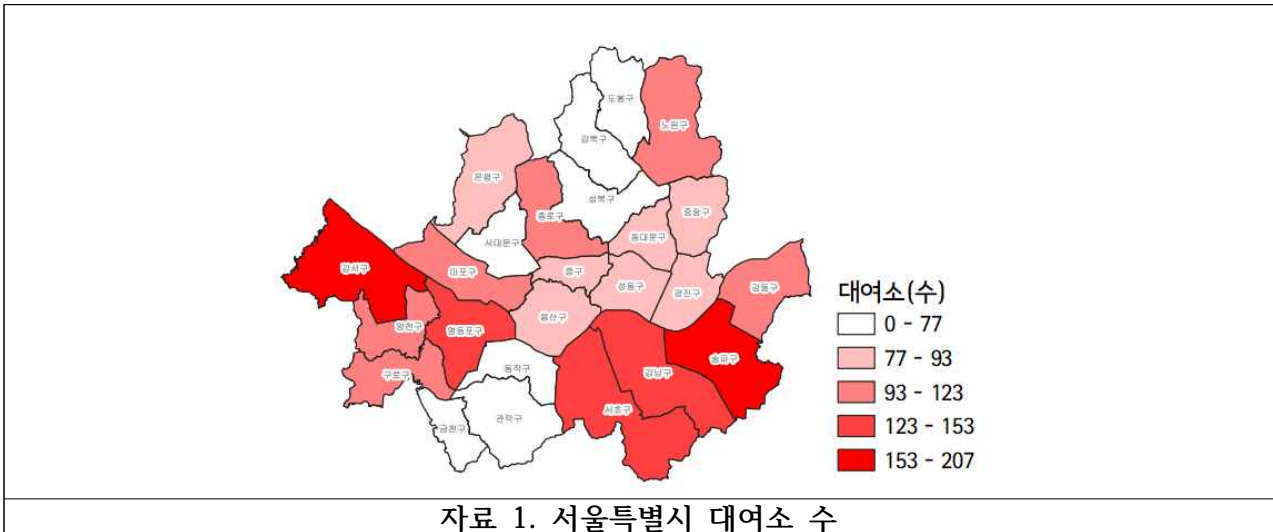


2-1. 분석 자료

2-1-1 공공자전거 이용정보 데이터

(서울 열린 데이터 광장 - 2021년 서울시 공공자전거 이용정보(일별))

대여 일자, 대여소 번호, 대여소명, 성별, 연령대 코드 등의 자료가 포함되어 있으며 본 연구에서는 대여 일자에 대한 이용량을 다중회귀분석의 종속변수(Y)로 사용되며, 종속변수에 log를 변환을 했다. log 변환을 하는 이유는 종속변수와 독립변수 간 왜도(skewness, 도수분포에서 평균값에 관한 좌우 비대칭성)와 첨도(kurtosis, 뾰족한 정도)를 줄여 정규성을 높이고 분석에서 정확한 값을 얻기 위함이다. 공공자전거 대여소 수는 22년 05 기준 서울특별시내 총 2586개(강남구:153, 강동구:117, 강북구:58, 강서구:181, 관악구:76, 광진구:86, 구로구:99, 금천구: 72, 노원구:123, 도봉구: 68, 동대문구: 82, 동작구: 70, 마포구: 117, 서대문구: 75, 서초구: 141, 성동구: 84, 성북구: 77, 송파구:207, 양천구: 103, 영등포구: 142, 용산구: 83, 은평구: 93, 종로구: 105, 중구: 84, 중랑구: 90)의 대여소가 있었으며 가장 많은 대여소가 있었던 지역은 송파구로 207개의 대여소가 있었고 가장 적게 대여소를 가지고 있었던 지역은 강북구로 58개의 대여소를 가지고 있었다. 분석은 2021년의 대여소를 기준으로 진행하며, 2021년 이후 신설된 대여소는 제외하고 분석을 진행하였다.



2-1-2 기상정보 데이터
(기상청 - 방재기상관측 데이터)
일시, 평균기온(° C), 최저기온(° C), 최저기온 시각(hhmi), 최고기온(° C), 최고기온 시각(hhmi), 일강수량(mm), 최대순간풍속(m/s), 최대순간풍속 시각(hhmi), 평균풍속(m/s), 최대순간풍속 풍향(deg) 자료가 포함되어 있으며 본 연구에서는 공공자전거 데이터의 대여 일자와 매칭하여 평균기온(° C), 최저기온(° C), 최고기온(° C), 일강수량(mm), 최대순간풍속(m/s), 평균풍속(m/s)이 독립변수(x)로 사용되며 계산된 지표들의 단위 등이 서로 상이하고 다양하므로 지표 측정의 단위를 표준화하여 그 값을 사용한다. 표준화된 지표값의 범위는 -2에서 2 사이로 변화하며 그 식은 식 1. 지표 표준화와 같다.

$$X = \frac{(\text{해당지표값} - \text{해당지표의 평균값})}{\text{해당지표의 표준편차}}$$

식 1. 지표 표준화

2-2. 다중 회귀 분석(multiple regression analysis)
회귀분석이란 종속변수와 독립변수 간의 함수적 관계를 파악하는 통계적 기법을 말하며, 회귀분석은 회귀 모형 구축에 의한 자료의 기술과 모수(parameter, 모집단의 특성을 나타내는 수치)추정, 적합한 모형을 통한 예측을 목적으로 한다. 회귀 모형은 종속변수를 설명하기 위한 독립변수의 수에 따라 단순회귀 모형과 다중회귀모형으로 분류되며, 단순회귀모형은 하나의 독립변수와 종속변수의 관계를 나타내고 다중회귀모형은 두 개 이상의 독립변수와 종속변수의 관계를 함수적으로 나타내는 것을 말한다. 본 연구에서는 6가지 기상 데이터에 따른 공공자전거의 이용량을 예측하는 것을 목적으로 하기 때문에 다중회귀 분석을 적용하였다. 식 2. 다중회귀모형은 자료의 개수가 n 개일 때, 독립변수 x1, x2, x3를 이용하여 종속변수 y를 추정하기 위한 다중 회귀모형을 나타낸 것이다.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (1)$$

식 2. 다중회귀모형(multiple regression analysis)

여기서 $\beta_0, \beta_1, \beta_2, \beta_3$ 은 회귀계수를 나타내며, ε_i 는 오차항을 나타낸다. 회귀계수를 추정하기 위해서 주로 최소제곱법(least square method)의 원리를 적용한다. 최소제곱법은 식 2를 회귀 모형으로 가정하였을 때, 회귀 모형에 의한 모든 독립변수와 종속변수 간의 관계에서 나타나는 오차항 ε_i 의 제곱의 합, 즉, 식 3의 값이 최소가 되도록 하는 회귀계수를 추정하는 방법이다.

$$\sum \varepsilon_i^2 = [y_i - (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i})]^2, i = 1, 2, \dots, n \quad (2)$$

식 3. 최소제곱법(least square method)

본 연구에서는 기상 데이터와 공공자전거 이용량 간의 상관관계를 예측하기 위하여 다중선형회귀분석을 적용하였다. 다중선형회귀분석은 종속변수와 각각의 독립변수를 설명하는 회귀계수가 선형을 이루고 있는 모형에 대한 분석을 말하며, 독립변수와 종속변수의 관계가 1차 식으로 구성된 다중선형회귀모형을 적용할 경우 각각의 회귀계수는 종속변수를 예측하기 위한 독립변수의 영향력을 의미한다.

3. 연구 범위

3-1. 연구의 공간적 범위

본 연구는 서울시의 공공자전거 따릉이의 이용량과 기상데이터와의 상관관계를 분석하는 연구이고 각 대여소별 대여량을 시군구 단위로 합산하여 분석을 진행하기 때문에 연구의 공간적 범위는 서울특별시의 자치구 단위로 한다.



자료 2. 서울특별시 자치구

3-2. 연구의 시간적 범위

본 연구의 시간적 범위는 서울특별시의 공공자전거인 따릉이 서비스가 안정적으로 도입되기 시작한 2021년을 시간적 범위로 한다.

4. 데이터 전처리

4-1. 공공자전거 데이터 분석의 흐름

4-1-1. 자료의 인코딩 변경

```
import pandas as pd
import numpy as np
df = pd.read_csv('C:/Users/bangw/Desktop/도시데이터코딩_자료/도시데이터_따릉이/공공자전거_이용정보(일일)_2102.csv', encoding='utf-8')
df
```

대여일자 대여소번호 대여소 대여구분코드 성별 연령대코드 윗층좌석번호 승차권 종류 윗층좌석번호(M) 윗층좌석번호(N)

* 최초 자료에서 한글 깨짐 현상이 생겨 Sublime text를 활용하여 파일의 인코딩을 cpc949로 변경

```
import pandas as pd
import numpy as np
df = pd.read_csv('C:/Users/sangw/Desktop/도시데이터코딩_자료/도시데이터_파블이/클라우드전거 이용정보(일일)_2102.csv', encoding='cp949')
df
```

	대여일자	대여소번호	대여소	대여구분코드	성별	연령대코드	이용건수	운동량	탄소량	이동거리(M)	이용시간(분)
0	2021-02-01	3	중랑센터	일일(회원)	M	AGE_003	1	23.01	0.19	830.00	8
1	2021-02-01	3	중랑센터	정기	M	AGE_003	1	112.38	0.56	2404.50	25
2	2021-02-01	10	위트클럽	정기	M	AGE_003	7	25.91	0.12	531.94	66
3	2021-02-01	101	101. (구)합정동 주민센터	정기	W	AGE_002	4	234.79	1.68	7284.77	183
4	2021-02-01	101	101. (구)합정동 주민센터	정기	W	AGE_003	2	245.79	1.67	7187.86	99
...
613802	2021-02-28	3588	3588.세종대학교(영실관)	정기	M	AGE_002	3	39.25	0.37	1605.64	13
613803	2021-02-28	3588	3588.세종대학교(영실관)	정기	M	AGE_003	1	20.02	0.2	842.60	13
613804	2021-02-28	9999	시스템관리팀	단체	F	AGE_002	3	37.22	0.37	1566.59	135
613805	2021-02-28	9999	시스템관리팀	일일(비회원)	W	AGE_008	1	5.15	0.05	200.00	5
613806	2021-02-28	9999	시스템관리팀	일일(회원)	F	AGE_004	1	171.43	1.55	6660.00	53

613807 rows x 11 columns

4-1-2. 대여소번호 & 대여일자별 이용건수 합계 산출

```
df2 = df.pivot_table('이용건수', '대여소번호', '대여일자', aggfunc=np.sum, margins=True)
```

대여일자	2021-12-01	2021-12-02	2021-12-03	2021-12-04	2021-12-05	2021-12-06	2021-12-07	2021-12-08	2021-12-09	2021-12-10	...	2021-12-23	2021-12-24	2021-12-25	2021-12-26	2021-12-27	2021-12-28	2
대여소번호																		1
3	1.0	NaN	NaN	NaN	1.0	NaN	NaN	NaN	NaN	1.0	...	NaN	NaN	NaN	NaN	NaN	NaN	
5	NaN	NaN	NaN	NaN	NaN	NaN	1.0	1.0	NaN	NaN	...	1.0	1.0	NaN	NaN	NaN	NaN	1.0
102	53.0	56.0	65.0	68.0	63.0	79.0	87.0	98.0	89.0	80.0	...	85.0	52.0	18.0	10.0	34.0	50.0	
103	44.0	40.0	43.0	42.0	56.0	67.0	78.0	76.0	70.0	52.0	...	54.0	43.0	17.0	13.0	29.0	31.0	
104	38.0	43.0	40.0	38.0	27.0	46.0	64.0	46.0	56.0	38.0	...	43.0	25.0	13.0	6.0	33.0	25.0	
...	
9998	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	
9999	1.0	NaN	5.0	6.0	1.0	3.0	8.0	3.0	4.0	5.0	...	3.0	4.0	2.0	1.0	3.0	2.0	
88888	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	
99999	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	1.0	
All	62785.0	65547.0	70087.0	59475.0	58055.0	84020.0	89379.0	90993.0	88633.0	73232.0	...	75058.0	55216.0	18859.0	16485.0	42521.0	51268.0	520

2587 rows x 32 columns

```
df3 = df2.fillna(0)
```

```
write_csv = df3.to_csv("C:/Users/sangw/Desktop/2112_자전거", csv")
```

* 대여소별로 부여된 번호를 서울시의 자치구로 매칭한 후 날짜별로 이용 건수를 합산(피벗테이블 기능 사용)

4-2. 기상정보 데이터 분석의 흐름

4-2-1. 최초 기상정보 데이터

지점	일시	평균기온(°C)	최저기온(°C)	최저기온 시작(hh)	최고기온(°C)	최고기온 시작(hh)	일강수량(mm)	최대 순간 풍속(m)	최대 순간 풍속 시	평균 풍속(m/s)	최대 순간 풍속 풍향(deg)
400	2021-01-01	-3.6	-9	538	2.2	1416	0	6.4	2250	1.1	329.1
400	2021-01-02	-3.8	-6.6	752	-0.3	1425	0	6.9	1544	1.7	334.7
400	2021-01-03	-4.4	-9.8	750	-0.7	1304	0	6.2	1210	1.4	250.3
400	2021-01-04	-2.8	-8.1	416	0.9	1325	0	4	2354	0.8	81.6
400	2021-01-05	-4.2	-7.9	2352	-1.4	408	0	7.7	1049	1.9	320.6
400	2021-01-06	-6.3	-10.7	511	-1.7	1455	6	7.4	2337	1.3	315
400	2021-01-07	-13.4	-15.7	755	-7.4	1	0	9.7	350	2.8	343.1

4-2-2. 분석에 필요한 마스터 파일

month	sgg	rental(Y)	평균기온(°C)	최저기온(°C)	최고기온(°C)	일강수량(mm)	최대순간풍속(m/s)	평균풍속(m/s)
2021-01-01	강남구	744	-3.6	-9	2.2	0	6.4	1.1
2021-01-02	강남구	708	-3.8	-6.6	-0.3	0	6.9	1.7
2021-01-03	강남구	631	-4.4	-8.8	-0.7	0	6.2	1.4
2021-01-04	강남구	1072	-2.8	-8.1	0.9	0	4	0.8
2021-01-05	강남구	946	-4.2	-7.9	-1.4	0	7.7	1.9
2021-01-06	강남구	679	-6.3	-10.7	-1.7	6	7.4	1.3
2021-01-07	강남구	110	-13.4	-15.7	-7.4	0	9.7	2.8
2021-01-08	강남구	154	-13.7	-17	-9.5	0	7.7	2.2
2021-01-09	강남구	158	-11.3	-15.8	-6.6	0	8.2	1.5

서울특별시의 기상관측 지점(2586개) 별 기상자료와 자치구 별 공공자전거 이용건수 합계를 일자 기준으로 매칭하였다. 서울특별시 25개 자치구의 365일 공공자전거 이용건수와 기상 데이터를 구축하였다.

4-3. 분석 절차

분석 결과의 설명력을 높이기 위하여 총 5가지 case를 통해 분석을 진행하였고, 설명력이 0.435로 가장 높았던 case 5(종속변수 log 치환, 독립변수 표준화)를 통한 분석 결과를 최종적으로 보고서에 기입하였다.

4-3-1. 최초 파일

	month	sgg	rental(Y)	평균기온(°C)	최저기온(°C)	최고기온(°C)	일강수량(mm)	최대순간풍속(m/s)	평균풍속(m/s)
0	2021-01-01	강남구	744	-3.6	-9.0	2.2	0.0	6.4	1.1
1	2021-01-02	강남구	708	-3.8	-6.6	-0.3	0.0	6.9	1.7
2	2021-01-03	강남구	631	-4.4	-8.8	-0.7	0.0	6.2	1.4
3	2021-01-04	강남구	1072	-2.8	-8.1	0.9	0.0	4.0	0.8
4	2021-01-05	강남구	946	-4.2	-7.9	-1.4	0.0	7.7	1.9
...
9120	2021-12-27	중랑구	1175	-6.8	-11.1	-3.7	0.0	4.9	1.3
9121	2021-12-28	중랑구	1350	-3.4	-7.4	-0.1	0.0	4.0	1.2
9122	2021-12-29	중랑구	1262	1.1	-2.7	5.9	0.0	7.5	1.7
9123	2021-12-30	중랑구	1253	-2.8	-5.0	0.2	0.0	10.1	2.8
9124	2021-12-31	중랑구	1073	-5.6	-7.7	-3.1	0.0	9.4	2.8

9125 rows × 9 columns

4-3-2. 결측치 제거

```
master4 = master3.dropna(axis=0)
master4
```

	month	sgg	rental(Y)	평균기온(°C)	최저기온(°C)	최고기온(°C)	일강수량(mm)	최대순간풍속(m/s)	평균풍속(m/s)
0	2021-01-01	강남구	744	-3.6	-9.0	2.2	0.0	6.4	1.1
1	2021-01-02	강남구	708	-3.8	-6.6	-0.3	0.0	6.9	1.7
2	2021-01-03	강남구	631	-4.4	-8.8	-0.7	0.0	6.2	1.4
3	2021-01-04	강남구	1072	-2.8	-8.1	0.9	0.0	4.0	0.8
4	2021-01-05	강남구	946	-4.2	-7.9	-1.4	0.0	7.7	1.9
...
9120	2021-12-27	중랑구	1175	-6.8	-11.1	-3.7	0.0	4.9	1.3
9121	2021-12-28	중랑구	1350	-3.4	-7.4	-0.1	0.0	4.0	1.2
9122	2021-12-29	중랑구	1262	1.1	-2.7	5.9	0.0	7.5	1.7
9123	2021-12-30	중랑구	1253	-2.8	-5.0	0.2	0.0	10.1	2.8
9124	2021-12-31	중랑구	1073	-5.6	-7.7	-3.1	0.0	9.4	2.8

8771 rows × 9 columns

최초 9125개의 일별 데이터를 구축하였지만, 결측치가 포함되어 있어 결측치가 포함되어있는 일별 데이터는 dropna를 사용하여 결측치를 제거하였고 최종적으로 8771개의 데이터가 남았다.

4-3-3. log값 추가

```
#log를 이용하여 변환
master4["log_rental(Y)"] = np.log1p(master4["rental(Y)"])
master4

C:\Users\sangw\AppData\Local\Temp\ipykernel_18480\3808100948.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-copy
master4["log_rental(Y)"] = np.log1p(master4["rental(Y)"])
```

	month	sgg	rental(Y)	평균기온(°C)	최저기온(°C)	최고기온(°C)	일강수량(mm)	최대순간풍속(m/s)	평균풍속(m/s)	log_rental(Y)
0	2021-01-01	강남구	744	-3.6	-9.0	2.2	0.0	6.4	1.1	6.613384
1	2021-01-02	강남구	708	-3.8	-8.6	-0.3	0.0	6.9	1.7	6.563856
2	2021-01-03	강남구	631	-4.4	-8.8	-0.7	0.0	6.2	1.4	6.448889
3	2021-01-04	강남구	1072	-2.8	-8.1	0.9	0.0	4.0	0.8	6.978214
4	2021-01-05	강남구	948	-4.2	-7.9	-1.4	0.0	7.7	1.9	6.853299

4-3-4. 분석에 불필요한 컬럼 제거

```
master4 = master4.drop(['month', 'sgg', 'rental(Y)'], axis=1)
master4
```

	평균기온(°C)	최저기온(°C)	최고기온(°C)	일강수량(mm)	최대순간풍속(m/s)	평균풍속(m/s)	log_rental(Y)
0	-3.6	-9.0	2.2	0.0	6.4	1.1	6.613384
1	-3.8	-8.6	-0.3	0.0	6.9	1.7	6.563856
2	-4.4	-8.8	-0.7	0.0	6.2	1.4	6.448889
3	-2.8	-8.1	0.9	0.0	4.0	0.8	6.978214
4	-4.2	-7.9	-1.4	0.0	7.7	1.9	6.853299
...
9120	-8.8	-11.1	-3.7	0.0	4.9	1.3	7.089874
9121	-3.4	-7.4	-0.1	0.0	4.0	1.2	7.208600
9122	1.1	-2.7	5.9	0.0	7.5	1.7	7.141245
9123	-2.8	-5.0	0.2	0.0	10.1	2.8	7.134094
9124	-5.6	-7.7	-3.1	0.0	9.4	2.8	6.979145

8771 rows x 7 columns

4-3-5. 지표 표준화

```
x1_mean, x1_std = master4['평균기온(°C)'].mean(), master4['평균기온(°C)'].std()
x2_mean, x2_std = master4['최저기온(°C)'].mean(), master4['최저기온(°C)'].std()
x3_mean, x3_std = master4['최고기온(°C)'].mean(), master4['최고기온(°C)'].std()
x4_mean, x4_std = master4['일강수량(mm)'].mean(), master4['일강수량(mm)'].std()
x5_mean, x5_std = master4['최대순간풍속(m/s)'].mean(), master4['최대순간풍속(m/s)'].std()
x6_mean, x6_std = master4['평균풍속(m/s)'].mean(), master4['평균풍속(m/s)'].std()
```

```
master4['평균기온'] = (master4['평균기온(°C)']-x1_mean)/x1_std
master4['최저기온'] = (master4['최저기온(°C)']-x2_mean)/x2_std
master4['최고기온'] = (master4['최고기온(°C)']-x3_mean)/x3_std
master4['일강수량'] = (master4['일강수량(mm)']-x4_mean)/x4_std
master4['최대순간풍속'] = (master4['최대순간풍속(m/s)']-x5_mean)/x5_std
master4['평균풍속'] = (master4['평균풍속(m/s)']-x6_mean)/x6_std
print(master4[['평균기온', '최저기온', '최고기온', '일강수량', '최대순간풍속', '평균풍속']])
```

	평균기온	최저기온	최고기온	일강수량	최대순간풍속	평균풍속
0	-1.666171	-1.710937	-1.570370	-0.305067	-0.268729	-0.707492
1	-1.685405	-1.488078	-1.807722	-0.305067	-0.077020	0.113945
2	-1.743106	-1.692365	-1.845699	-0.305067	-0.345413	-0.296773
3	-1.589237	-1.627365	-1.693793	-0.305067	-1.188932	-1.118210
4	-1.723872	-1.608793	-1.912157	-0.305067	0.229714	0.387758
...
9120	-1.973908	-1.905938	-2.130521	-0.305067	-0.843856	-0.433679
9121	-1.646938	-1.562364	-1.788734	-0.305067	-1.188932	-0.570586
9122	-1.214184	-1.125933	-1.219090	-0.305067	0.153031	0.113945
9123	-1.589237	-1.339506	-1.760252	-0.305067	1.149917	1.619914
9124	-1.858507	-1.590222	-2.073556	-0.305067	0.881525	1.619914

[8771 rows x 6 columns]

4-3-6. 다중 회귀 분석을 위한 종속변수

```

master4_data = master4.drop(['log_rental(Y)'],axis=1) # Feature에서 Target Data 분리
target = master4[['log_rental(Y)']] # Target 변수 저장

x_data = master4[['평균기온', '최저기온', '최고기온', '일강수량', '최대순간풍속', '평균풍속']]
# df[[col]] - DataFrame형태로 컬럼 추출 / df[col] - Series형태로 컬럼추출

x_data.head()

```

	평균기온	최저기온	최고기온	일강수량	최대순간풍속	평균풍속
0	-1.666171	-1.710937	-1.570370	-0.305067	-0.268729	-0.707492
1	-1.685406	-1.488078	-1.807722	-0.305067	-0.077020	0.113945
2	-1.743106	-1.692365	-1.845699	-0.305067	-0.345413	-0.296773
3	-1.589237	-1.627365	-1.693793	-0.305067	-1.188932	-1.118210
4	-1.723872	-1.608793	-1.912157	-0.305067	0.229714	0.387758

4-5-7. 다중 회귀 분석

```

# 회귀분석을 하기 위한 B.O. 상수항 추가
x_data1 = sm.add_constant(x_data, has_constant = "add")

# 회귀모델 적합
multi_model = sm.OLS(target, x_data1)
fitted_multi_model = multi_model.fit()

# summary결과를 통해 OLS 결과 출력
fitted_multi_model.summary()

```

C:\Users\hsangw\Anaconda3\lib\site-packages\statsmodels\tools\statstools.py:142:
ments of concat except for the argument 'objs' will be keyword-only
x = pd.concat(x[::-order], 1)

OLS Regression Results

Dep. Variable:	log_rental(Y)	R-squared:	0.435			
Model:	OLS	Adj. R-squared:	0.435			
Method:	Least Squares	F-statistic:	1126.			
Date:	Thu, 16 Jun 2022	Prob (F-statistic):	0.00			
Time:	17:34:45	Log-Likelihood:	-8795.7			
No. Observations:	8771	AIC:	1.761e+04			
Df Residuals:	8764	BIC:	1.766e+04			
Df Model:	6					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	7.8329	0.007	1111.707	0.000	7.819	7.847
평균기온	0.1932	0.121	1.595	0.111	-0.044	0.431
최저기온	-0.1943	0.067	-2.893	0.004	-0.326	-0.063
최고기온	0.4930	0.065	7.570	0.000	0.365	0.621
일강수량	-0.2880	0.007	-39.204	0.000	-0.302	-0.274
최대순간풍속	-0.1225	0.010	-11.856	0.000	-0.143	-0.102
평균풍속	0.0528	0.010	5.070	0.000	0.032	0.073
Omnibus:	4379.977	Durbin-Watson:	0.790			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	134939.660			
Skew:	-1.796	Prob(JB):	0.00			
Kurtosis:	21.877	Cond. No.	36.9			

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

5. 결과 분석

[그림1. 날씨(기후) 변수의 기초 통계량]

	평균기온(°C)	최저기온(°C)	최고기온(°C)	일강수량(mm)	최대순간풍속(m/s)	평균풍속(m/s)
count	8771.000000	8771.000000	8771.000000	8771.000000	8771.000000	8771.000000
mean	13.725698	9.425356	18.740531	3.071258	7.100878	1.616771
std	10.398509	10.769162	10.532885	10.067497	2.608120	0.730427
min	-17.100000	-22.600000	-12.700000	0.000000	1.900000	0.200000
25%	6.400000	1.700000	11.200000	0.000000	5.300000	1.100000
50%	14.000000	9.400000	19.400000	0.000000	6.600000	1.500000
75%	22.700000	19.000000	27.700000	0.500000	8.400000	2.000000
max	33.000000	29.500000	39.200000	104.500000	45.000000	6.800000

[그림2. 날씨(기후) 변수의 분석통계량]

OLS Regression Results

Dep. Variable:	log_rental(Y)	R-squared:	0.435			
Model:	OLS	Adj. R-squared:	0.435			
Method:	Least Squares	F-statistic:	1126.			
Date:	Thu, 09 Jun 2022	Prob (F-statistic):	0.00			
Time:	17:17:00	Log-Likelihood:	-8795.7			
No. Observations:	8771	AIC:	1.761e+04			
Df Residuals:	8764	BIC:	1.766e+04			
Df Model:	6					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	7.8329	0.007	1111.707	0.000	7.819	7.847
평균기온	0.1932	0.121	1.595	0.111	-0.044	0.431
최저기온	-0.1943	0.067	-2.893	0.004	-0.326	-0.063
최고기온	0.4930	0.065	7.570	0.000	0.365	0.621
일강수량	-0.2880	0.007	-39.204	0.000	-0.302	-0.274
최대순간풍속	-0.1225	0.010	-11.856	0.000	-0.143	-0.102
평균풍속	0.0528	0.010	5.070	0.000	0.032	0.073
Omnibus:	4379.977	Durbin-Watson:	0.790			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	134939.660			
Skew:	-1.796	Prob(JB):	0.00			
Kurtosis:	21.877	Cond. No.	36.9			

날씨가 전체 자전거 이용에 미치는 영향을 분석한 결과 온도 등의 날씨 변수가 자전거 이용에 영향을 미치는 것으로 나타났다. 모형의 수정된 R제곱 값은 0.435로, 분석한 변수들이 약 43.5%의 설명력을 가짐을 알 수 있었다.

채택된 변수는 6개 중 최저기온, 최고기온, 일 강수량, 최대순간풍속, 평균풍속이다. 기온의 경우 평균기온과 최저기온 최고기온으로 나누어 확인하였는데, 평균기온 변수는 p값이 0.111으로 채택되지 않았다. 반면, 최저기온과 최고기온은 매우 유의한 변수로 채택되었다. 최저기온의 경우 음의 값을 가지는데, 이

는 최저기온 내에서 비교적 기온이 덜 낮으면 자전거 이용이 증가하고, 온도가 그보다 더 낮으면 이용이 감소하는 것으로 해석할 수 있다. 최고기온은 날씨 변수 중 자전거 이용수에 가장 큰 영향을 미치는 변수임을 확인했다. 상관계수를 해석해 보면 최저기온과 반대로 비교적 온도가 낮아지면 자전거 이용수가 감소하며 온도가 높아지면 자전거 이용수가 증가하는 것으로 해석할 수 있었다.

강수량은 음의 값으로 나타났는데, 이는 비가 많이 오면 자전거 이용이 감소하고, 비가 적게 오면 자전거 이용이 증가함을 뜻한다. 풍속의 경우에는 평균풍속과 최대순간풍속의 상관계수가 다른 영향을 미칠 수 있다. 먼저 최대순간풍속은 음의 값이 나왔는데, 이것은 강한 돌풍 등의 바람이 불면 자전거 이용이 감소한다는 것으로 해석할 수 있다. 평균풍속은 양의 값으로 나타났으며 이것은 평균 바람의 세기가 강해질수록 자전거 이용 수가 증가하는 것으로 이해할 수 있었다.

6. 결론

본 연구에서 도출한 결과는 날씨가 자전거 이용에 영향을 미친다는 것을 보여주고 있다. 결과적으로 온도와 강수량, 그리고 바람(풍속)이 영향을 미치는 것으로 나타났다. 그러나 겨울기간에 대한 적설량 변수가 분석 대상에 포함되지 않아 향후 이를 포함한 분석도 수행되어야 공공자전거 이용 수와 날씨의 관계를 더 명확히 볼 수 있을 것이다. 본 연구결과는 앞으로 자전거 정책을 수립하고 시행하는데 기초적인 자료가 될 수 있을 것으로 판단된다.

7. 참고자료 및 문헌

- 1) 서울 열린데이터 광장 (<https://data.seoul.go.kr/dataList/OA-15246/F/1/datasetView.do>)
 - 서울시 공공자전거 이용정보(일별).
 - 공공자전거 대여소 정보.
- 2) 기상청 기상자료개방포털 (<https://data.kma.go.kr/data/rmt/rmtList.do?code=400&pgmNo=570>)
 - 평균기온, 최저기온, 최고기온, 일 강수량, 최대순간풍속, 평균풍속.
- 3) 이장호, 정경옥, 신희철 (2016), “기상조건과 입지특성이 공공자전거 이용에 미치는 영향분석”, 한국교통대학교 철도시설공학과, 한국교통연구원.
- 4) 김동준, 신희철, 박준식, 임형준 (2012), “날씨가 자전거 이용에 미치는 영향 분석 - 고양시 공공자전거를 대상으로”, 한국교통연구원.
- 5) “서울시, 따릉이 4년… 56% ‘출퇴근시간대’ 집중, 70% ‘단거리 이동’ 이용” (I·SEOUL·U 2019년 11월 4일 조간용 보도자료, 담당부서: 도시교통실 자전거정책과)