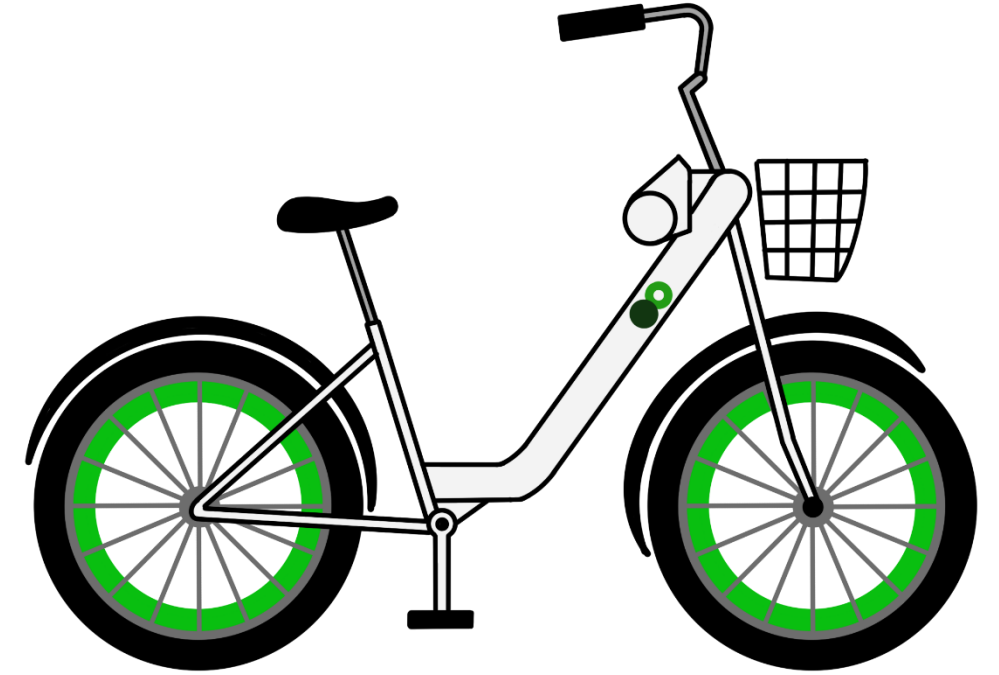


빅데이터분석(가) 최종발표

# 따릉이, 여기까지 타봤다!

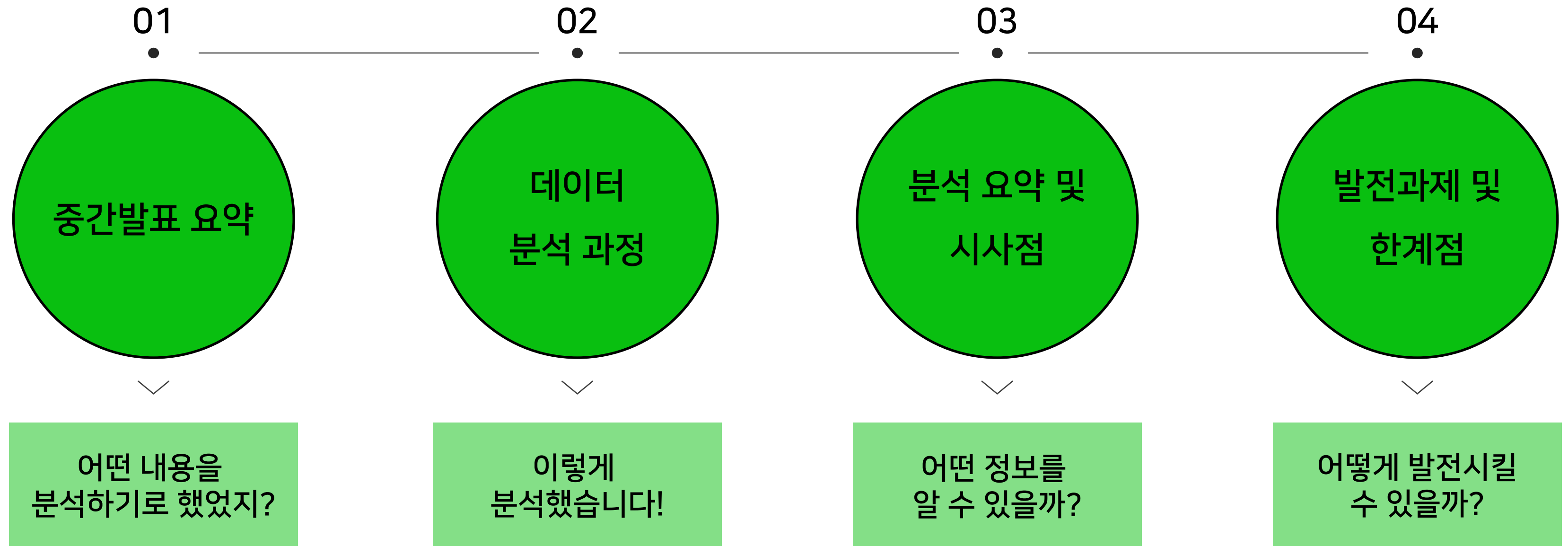
- 서울시 공용자전거 사용 데이터 분석을 통한 인사이트 도출 -



글로벌통상학과 20183076 이상훈

2024.06.11

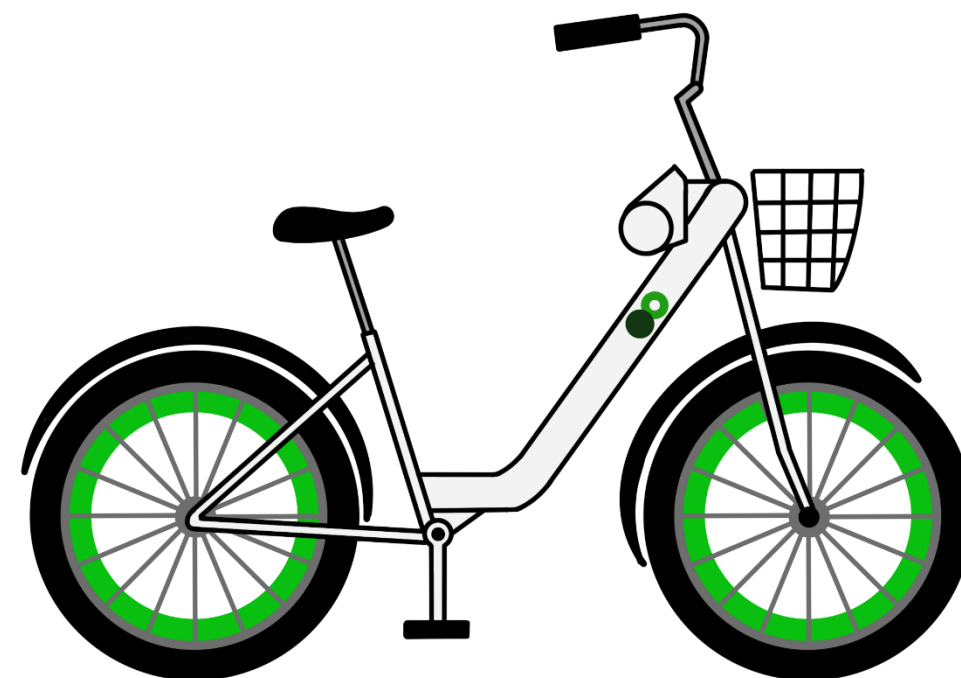
## 0. 목차



# 01

## 중간발표 요약

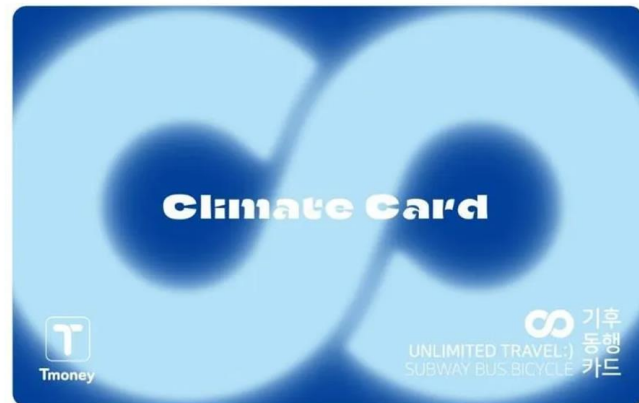
어떤 내용을 분석하기로 했었지?



# 기후동행카드 사업이 **따릉이 사용량**에 어떠한 영향을 미쳤는지 궁금해져 시작된 이번 텀 프로젝트

## 1. 주제 선정 배경

서울시에서 시범적으로 운영하고 있는 대중교통 통합 정기권,  
**기후동행카드**



[그림 1] 기후동행카드의 실물 이미지 (출처: 서울시)

- 사업 기간 2024년 1월 27일(토) ~ 6월 30일(일)
- 이용 범위 서울지역 지하철 + 김포골드라인, 서울시 면허 시내·마을버스, **따릉이**
- 이용 제외 신분당선, 서울지역 외 지하철, 광역/공항버스, 타 지역 면허버스
- 카드 가격 ① 62,000원권(지하철 + 버스)  
② 65,000원권(지하철 + 버스 + **따릉이**)

4/17

## 2. 분석 목표

**3개**의 핵심 목표를 중심으로 한 분석 진행 예정

1

기후동행카드 사업 시행 전후의  
따릉이 사용량 변화

2

지역구별 대중교통 정류장 개수와  
따릉이 사용량의 상관관계

3

지역구별 특성에 따른  
따릉이 배치 분석

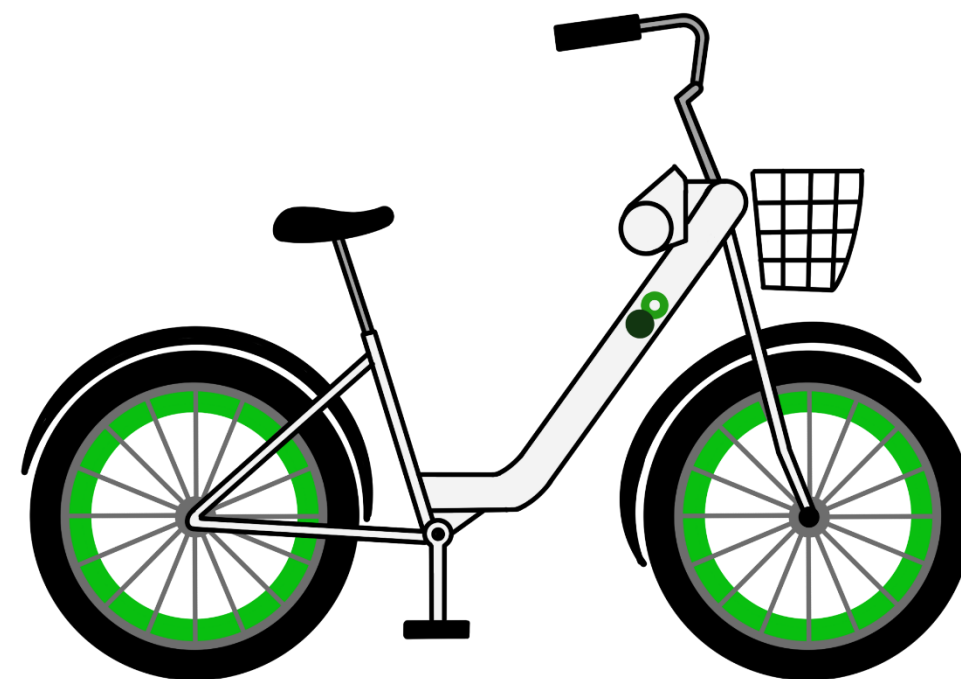
8/17

서울시, 통계청 등 공공기관의 공공데이터를 활용하여  
기후동행카드 사업 시행 전후의 변화 분석을 목표

# 02

## 데이터 분석 과정

이렇게 분석했습니다!



## 공공기관의 공공 데이터와 GitHub의 오픈소스 데이터를 활용



통계청

국가통계포털

행정구역(시군구)별/1세별 주민등록인구  
(2023년 4월, 2024년 4월)



서울시 열린데이터 광장

서울시 버스노선별 정류장별 승하차 인원 정보  
서울시 버스정류소 위치정보  
서울시 지하철호선별 역별 승하차 인원 정보  
서울시 역사마스터 정보  
서울시 따릉이 대여소별 대여/반납 승객수 정보  
서울시 따릉이대여소 마스터 정보



서울\_자치구\_경계\_2017.geojson\*

## 2. 데이터 분석 과정\_ ② 데이터 전처리

# 따릉이 대여소 데이터: 주소 없음 데이터 추가

```
분석(가)_템프로젝트
View Insert Cell Kernel Navigate Widgets Help
2581 ST-164 NaN 37.484215 126.996292
In [75]: 1 sbike_spot[sbike_spot["주소1"].isna()]["대여소_ID"].values
array(['ST-3191', 'ST-3190', 'ST-2426', 'ST-1168', 'ST-1165',
       'ST-1164', 'ST-1162', 'ST-1161', 'ST-1160', 'ST-1158', 'ST-1157',
       'ST-1154', 'ST-1144', 'ST-1143', 'ST-1142', 'ST-1137', 'ST-1123',
       'ST-1110', 'ST-1109', 'ST-1103', 'ST-1102', 'ST-1099', 'ST-1098',
       'ST-1096', 'ST-1095', 'ST-1093', 'ST-1057', 'ST-1042', 'ST-1039',
       'ST-1038', 'ST-1037', 'ST-1035', 'ST-1034', 'ST-1033', 'ST-1032',
       'ST-1029', 'ST-1028', 'ST-1027', 'ST-1025', 'ST-1024'],
      dtype=object)
```

	A	B	C	D	E
1	대여소_	주소1	주소2	위도	경도
889	ST-3164	서울특별시 강남구 영동대로 646 위레벤646		37.51974	127.0577
2000	ST-2164	서울특별시 마포구 상암동 1727		37.5793	126.8873
2573	ST-1649	서울특별시 영등포구 양산로 213	아크로타워	37.52252	126.9072
2574	ST-1648	서울특별시 금천구 시흥대로 39-7	석수역1번	37.43727	126.9027
2575	ST-1647	서울특별시 강서구 방화대로5길 7	송정중학교	37.55473	126.8161
2576	ST-1646	서울특별시 강서구 화곡로63가길 92	등촌9단지	37.5629	126.8495
2577	ST-1645	서울특별시 강서구 방화대로7길 6	마곡13단지	37.55645	126.816
2578	ST-1644	서울특별시 서대문구 연희로 136 서울연희초등학교		37.56963	126.9333
2579	ST-1643	서울특별시 서대문구 홍제천로 111	홍남교 두	37.57491	126.9291
2580	ST-1642	서울특별시 노원구 노원로34길 43	노원정보도	37.66071	127.0653
2581	ST-1641	서울특별시 노원구 동일로 1449	도봉운전민	37.65773	127.0593
2582	ST-1640	서울특별시 노원구 동일로 지하 1520-1	마들연 3번	37.66525	127.0579
2583	ST-1640	주소 없음		37.48422	126.9963
3107	ST-1164	서울특별시 서초동 1748-33	교내입구	37.48731	127.0106

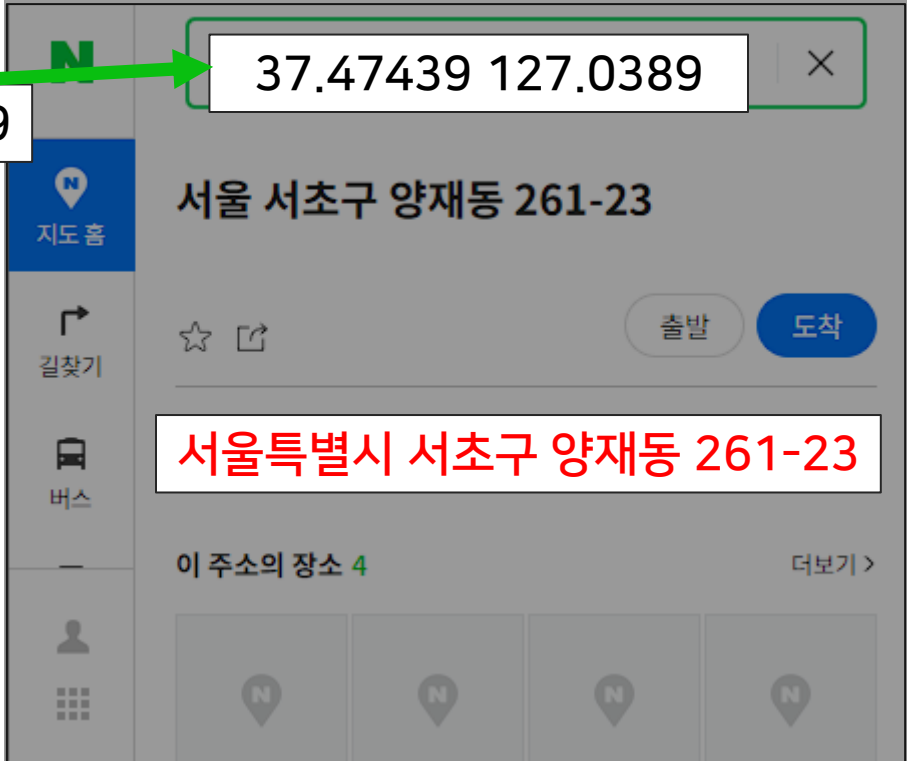


# 따릉이 대여소 데이터: 도로명 주소 누락 데이터 수정

```
분석(가)_템프로젝트
View Insert Cell Kernel Navigate Widgets Help
2581 ST-164 NaN 37.484215 126.996292
In [75]: 1 sbike_spot[sbike_spot["주소1"].isna()]["대여소_ID"].values
array(['ST-3191', 'ST-3190', 'ST-2426', 'ST-164', 'ST-1165',
      'ST-1164', 'ST-1162', 'ST-1161', 'ST-1160', 'ST-1158', 'ST-1157',
      'ST-1154', 'ST-1144', 'ST-1143', 'ST-1142', 'ST-1137', 'ST-1123',
      'ST-1110', 'ST-1109', 'ST-1103', 'ST-1102', 'ST-1099', 'ST-1098',
      'ST-1096', 'ST-1095', 'ST-1093', 'ST-1057', 'ST-1042', 'ST-1039',
      'ST-1038', 'ST-1037', 'ST-1035', 'ST-1034', 'ST-1033', 'ST-1032',
      'ST-1029', 'ST-1028', 'ST-1027', 'ST-1025', 'ST-1024'],
      dtype=object)
```

ST-1168

	A	B	C	D	E
	대여소 ID	주소	위도	경도	
3288	ST-1168	서울특별시 양재동 261-23	37.47439	127.0389	
3289					
3290					
3291					
3292					
3293					
3294					
3295					
3296					
3297					
3298					
3299					
3300					
3301					





# 기후동행카드 사용불가 노선 필터링

버스(13개)

연번	노선번호	승·하차
1	6633	X
2	9401	X
3	9401-1	X
4	9404	X
5	9408	X
6	9409	X
7	9701	X
8	9707	X
9	9711	X
10	9714	X
11	서울01	X
12	서울03	X
13	서울06	X

지하철(262개)

연번	역명	노선	승차	하차	판매	충전	사용정지 처리
1	가능	1호선	X	X	X	X	X
2	간석	1호선	X	X	X	X	X
3	관악	1호선	X	X	X	X	X
4	광명	1호선	X	X	X	X	X
5	군포	1호선	X	X	X	X	X
6	금정	1호선	X	X	X	X	X
7	녹양	1호선	X	X	X	X	X
8	당정	1호선	X	X	X	X	X
9	덕계	1호선	X	X	X	X	X
10	덕정	1호선	X	X	X	X	X
11	도원	1호선	X	X	X	X	X
12	도화	1호선	X	X	X	X	X
13	등두천	1호선	X	X	X	X	X

```
1 # 이용불가 노선 13개
2 bus_unavailable = ["6633", "9401", "9401-01", "9404", "9408", "9409", "9701",
                    "9707", "9711", "9714", "서울01", "서울03", "서울06"]
```

```
1 # 이용불가 역 262개
2 subway_unavailable = ["가능", "간석", "관악", "광명", "군포", "금정", "녹양",
                        "당정", "덕계", "도원", "도화", "등두천", "등두천중앙", "동암", "동인천", "두
                        정", "망월사", "명학", "배방", "백운", "병점", "보산", "봉명", "부개", "부천",
```

# 지하철역, 버스정류소 행정구 라벨링

지하철(771개)

역한글명칭	호선명칭	환승역X좌표	환승역Y좌표
삼성중앙	9호선(연장)	127.053282	37.513011
사평	9호선	127.015259	37.504206
구반포	9호선	126.987332	37.501364
흑석(중앙대입구)	9호선	126.963708	37.508770
샛강	9호선	126.928422	37.517274
...	...	...	...
관악산(서울대)	신림선	126.945064	37.469102
신논현	신분당선(연장2)	127.025060	37.504598
당정	장항선	127.084850	36.788660
신사	신분당선(연장2)	127.020114	37.516334
논현	신분당선(연장2)	127.021415	37.511093



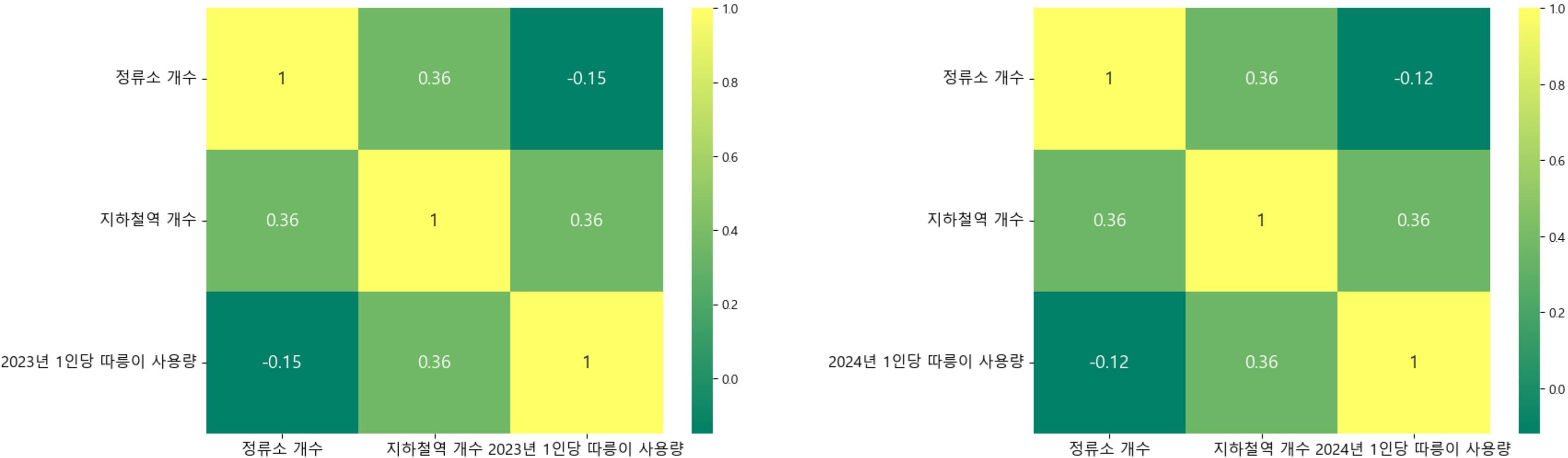
SIG_KOR_NM	geometry
종로구	POLYGON ((127.00864 37.58047, 127.00871 37.580...
중구	POLYGON ((127.02314 37.57196, 127.02336 37.571...
용산구	POLYGON ((126.96918 37.55566, 126.96917 37.554...
성동구	POLYGON ((127.04341 37.57234, 127.04524 37.571...
광진구	POLYGON ((127.10166 37.57240, 127.10224 37.572...
동대문구	POLYGON ((127.07108 37.60732, 127.07117 37.606...
중랑구	POLYGON ((127.11131 37.62069, 127.11214 37.620...
성북구	POLYGON ((127.01059 37.58025, 127.01030 37.580...
강북구	POLYGON ((127.00457 37.68508, 127.00553 37.684...
도봉구	POLYGON ((127.01509 37.64929, 127.01477 37.649...
노원구	POLYGON ((127.11085 37.63841, 127.11090 37.638...



버스(11261개)

정류소명	X좌표	Y좌표
종로2가사거리	126.987752	37.569806
창경궁.서울대학교병원	126.996521	37.579433
명륜3가.성대입구	126.998251	37.582580
종로2가.삼일교	126.987613	37.568579
혜화동로터리.여운형활동터	127.001744	37.586243
...	...	...
우성아파트	127.139339	37.550386
우성아파트	127.140046	37.550643
조밀약국	127.123596	37.533630
성내시장	127.125497	37.536155
천효우체국.로데오거리	127.127337	37.540343

# 버스정류소 개수, 지하철역 개수 - 행정구별 1인당 따릉이 사용량의 선형적인 상관관계 확인



버스정류소 개수와 지하철역 개수 모두 약한 상관관계를 보임

## 버스정류소 개수, 지하철역 개수 - 행정구별 1인당 따릉이 사용량의 선형회귀 · 중선형회귀 진행

### 귀무가설

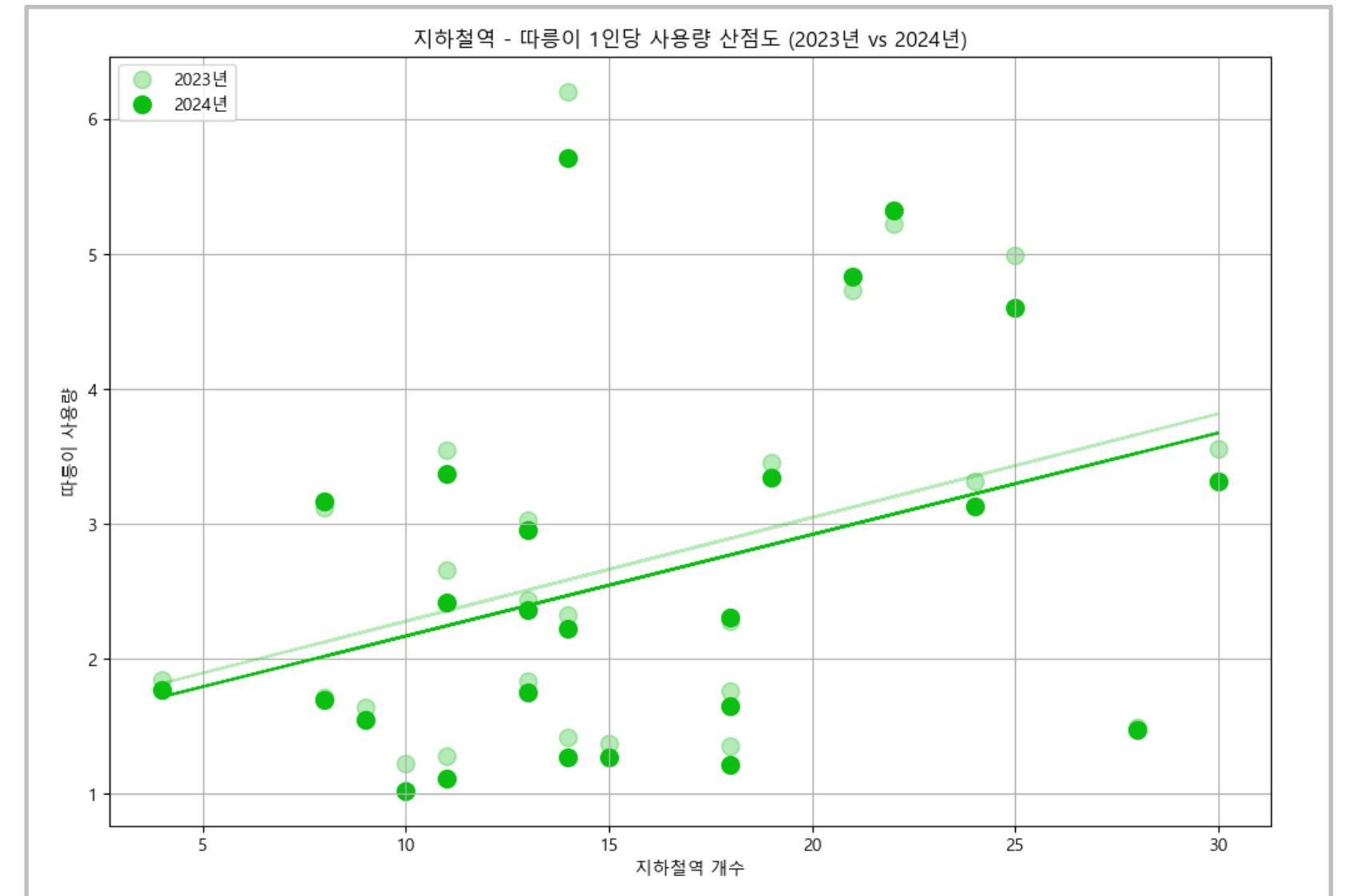
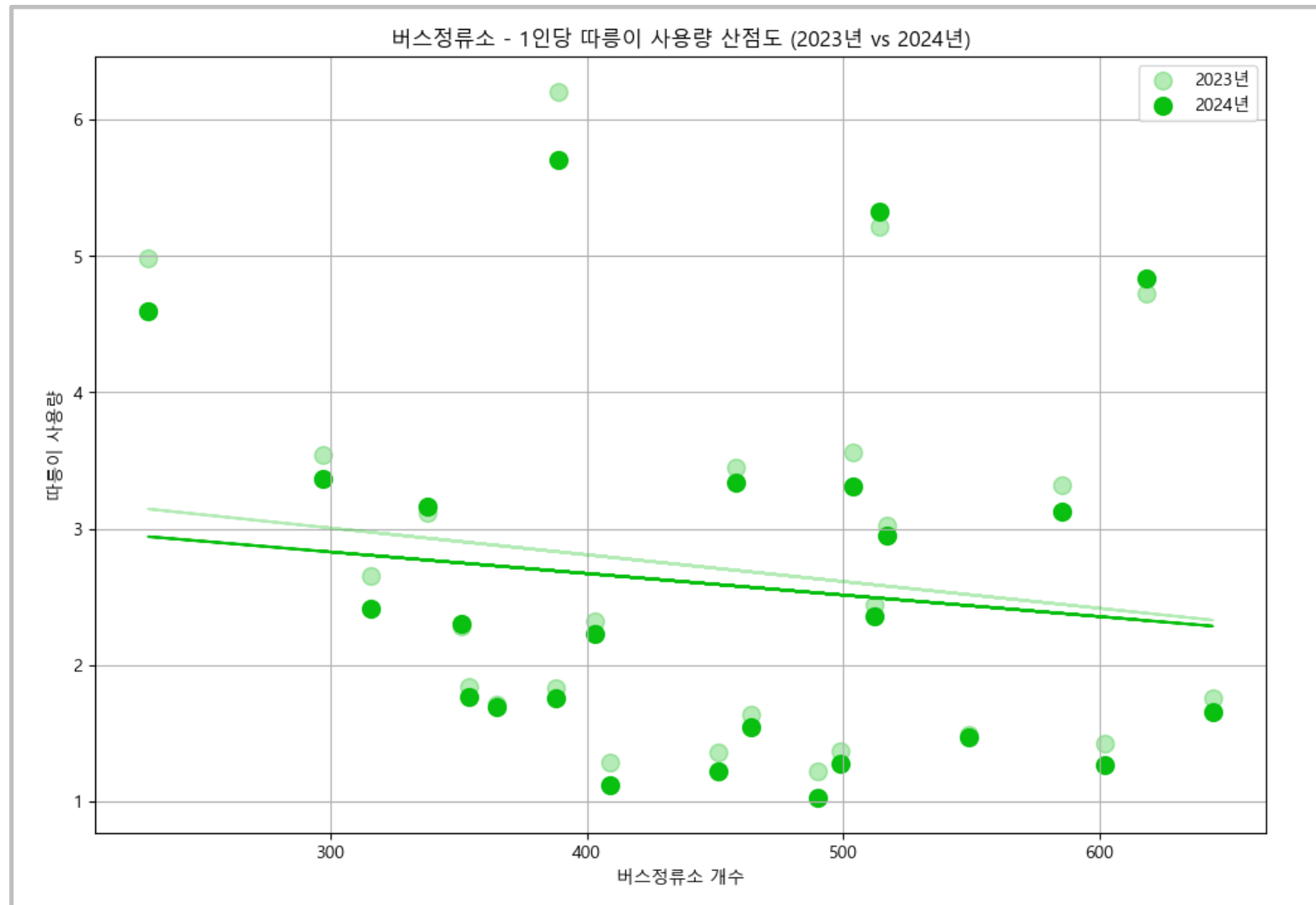
버스정류소 개수, 지하철역 개수는 따릉이 사용량과 관계가 없을 것이다

### 대립가설

버스정류소 개수, 지하철역 개수가 적으면 따릉이 사용량이 증가할 것이다

선형적인 상관관계 및 선형회귀분석을 통한 인과관계를  
확인하여 가설 검증을 진행

# 버스정류소 개수, 지하철역 개수 - 행정구별 1인당 따릉이 사용량의 선형회귀 · 중선형회귀 진행



산점도와 회귀선을 이용한 단일선형회귀 시각화

# 버스정류소 개수, 지하철역 개수 - 행정구별 1인당 따릉이 사용량의 선형회귀 · 중선형회귀 진행

OLS Regression Results

Dep. Variable:	per_person_2023	R-squared:	0.023			
Model:	OLS	Adj. R-squared:	-0.019			
Method:	Least Squares	F-statistic:	0.5426			
Date:	Tue, 04 Jun 2024	Prob (F-statistic):	0.469			
Time:	09:03:55	Log-Likelihood:	-42.928			
No. Observations:	25	AIC:	89.86			
Df Residuals:	23	BIC:	92.29			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	3.5941	1.233	2.915	0.008	1.043	6.145
bus	-0.0020	0.003	-0.737	0.469	-0.007	0.004
Omnibus:	5.042	Durbin-Watson:	1.468			
Prob(Omnibus):	0.080	Jarque-Bera (JB):	4.025			
Skew:	0.983	Prob(JB):	0.134			
Kurtosis:	2.973	Cond. No.	2.03e+03			

2023년	버스 정류소 개수
R <sup>2</sup> (결정계수)	0.023
P-value (유의확률)	0.469
Coef (회귀계수)	-0.002

OLS Regression Results

Dep. Variable:	per_person_2024	R-squared:	0.016			
Model:	OLS	Adj. R-squared:	-0.027			
Method:	Least Squares	F-statistic:	0.3647			
Date:	Tue, 04 Jun 2024	Prob (F-statistic):	0.552			
Time:	09:03:55	Log-Likelihood:	-42.471			
No. Observations:	25	AIC:	88.94			
Df Residuals:	23	BIC:	91.38			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	3.3032	1.211	2.728	0.012	0.799	5.808
bus	-0.0016	0.003	-0.604	0.552	-0.007	0.004
Omnibus:	4.547	Durbin-Watson:	1.437			
Prob(Omnibus):	0.103	Jarque-Bera (JB):	3.678			
Skew:	0.937	Prob(JB):	0.159			
Kurtosis:	2.868	Cond. No.	2.03e+03			

2024년	버스 정류소 개수
R <sup>2</sup> (결정계수)	0.016
P-value (유의확률)	0.552
Coef (회귀계수)	-0.001

유의확률이 0.005 이상이므로 귀무가설을 채택

# 버스정류소 개수, 지하철역 개수 - 행정구별 1인당 따릉이 사용량의 선형회귀 · 중선형회귀 진행

OLS Regression Results

Dep. Variable:	per_person_2023	R-squared:	0.132			
Model:	OLS	Adj. R-squared:	0.094			
Method:	Least Squares	F-statistic:	3.488			
Date:	Tue, 04 Jun 2024	Prob (F-statistic):	0.0746			
Time:	09:03:55	Log-Likelihood:	-41.455			
No. Observations:	25	AIC:	86.91			
Df Residuals:	23	BIC:	89.35			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	1.5078	0.696	2.166	0.041	0.068	2.948
subway	0.0769	0.041	1.868	0.075	-0.008	0.162
Omnibus:	5.551	Durbin-Watson:	2.024			
Prob(Omnibus):	0.062	Jarque-Bera (JB):	3.650			
Skew:	0.865	Prob(JB):	0.161			
Kurtosis:	3.715	Cond. No.	44.6			

2023년	지하철역 개수
R2 (결정계수)	0.132
P-value (유의확률)	0.075
Coef (회귀계수)	0.076

OLS Regression Results

Dep. Variable:	per_person_2024	R-squared:	0.132			
Model:	OLS	Adj. R-squared:	0.094			
Method:	Least Squares	F-statistic:	3.494			
Date:	Tue, 04 Jun 2024	Prob (F-statistic):	0.0744			
Time:	09:03:55	Log-Likelihood:	-40.900			
No. Observations:	25	AIC:	85.80			
Df Residuals:	23	BIC:	88.24			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	1.4147	0.681	2.078	0.049	0.007	2.823
subway	0.0752	0.040	1.869	0.074	-0.008	0.158
Omnibus:	3.551	Durbin-Watson:	1.989			
Prob(Omnibus):	0.169	Jarque-Bera (JB):	2.390			
Skew:	0.754	Prob(JB):	0.303			
Kurtosis:	3.150	Cond. No.	44.6			

2024년	지하철역 개수
R2 (결정계수)	0.132
P-value (유의확률)	0.074
Coef (회귀계수)	0.075

유의확률이 0.074 이상이므로 귀무가설을 기각하고 대립가설을 채택하지만,  
설명력이 0.132로 낮음

# 버스정류소 개수, 지하철역 개수 - 행정구별 1인당 따릉이 사용량의 선형회귀 · 중선형회귀 진행

OLS Regression Results

Dep. Variable:	per_person_2023	R-squared:	0.225			
Model:	OLS	Adj. R-squared:	0.154			
Method:	Least Squares	F-statistic:	3.186			
Date:	Tue, 04 Jun 2024	Prob (F-statistic):	0.0609			
Time:	09:03:55	Log-Likelihood:	-40.041			
No. Observations:	25	AIC:	86.08			
Df Residuals:	22	BIC:	89.74			
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	3.0195	1.149	2.629	0.015	0.637	5.402
bus	-0.0042	0.003	-1.623	0.119	-0.010	0.001
subway	0.1021	0.043	2.391	0.026	0.014	0.191
Omnibus:	7.289	Durbin-Watson:	1.953			
Prob(Omnibus):	0.026	Jarque-Bera (JB):	5.185			
Skew:	0.989	Prob(JB):	0.0748			
Kurtosis:	4.032	Cond. No.	2.07e+03			

2023년	버스 정류소 개수	지하철역 개수
R2 (결정계수)	0.225	
P-value (유의확률)	0.119	0.026
Coef (회귀계수)	-0.004	0.102

OLS Regression Results

Dep. Variable:	per_person_2024	R-squared:	0.208			
Model:	OLS	Adj. R-squared:	0.136			
Method:	Least Squares	F-statistic:	2.890			
Date:	Tue, 04 Jun 2024	Prob (F-statistic):	0.0768			
Time:	09:03:56	Log-Likelihood:	-39.752			
No. Observations:	25	AIC:	85.50			
Df Residuals:	22	BIC:	89.16			
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	2.7539	1.135	2.425	0.024	0.399	5.109
bus	-0.0038	0.003	-1.455	0.160	-0.009	0.002
subway	0.0976	0.042	2.312	0.031	0.010	0.185
Omnibus:	5.763	Durbin-Watson:	1.896			
Prob(Omnibus):	0.056	Jarque-Bera (JB):	3.996			
Skew:	0.941	Prob(JB):	0.136			
Kurtosis:	3.544	Cond. No.	2.07e+03			

2024년	버스 정류소 개수	지하철역 개수
R2 (결정계수)	0.208	
P-value (유의확률)	0.160	0.031
Coef (회귀계수)	-0.003	0.097

중선형회귀는 단일선형회귀보다 결정계수는 낮고, 유의확률은 높아졌음



## 버스정류소 개수, 지하철역 개수 - 행정구별 1인당 따릉이 사용량의 선형회귀 · 중선형회귀 진행

### 귀무가설

버스정류소 개수, 지하철역 개수는 따릉이 사용량과 관계가 없을 것이다

### 대립가설

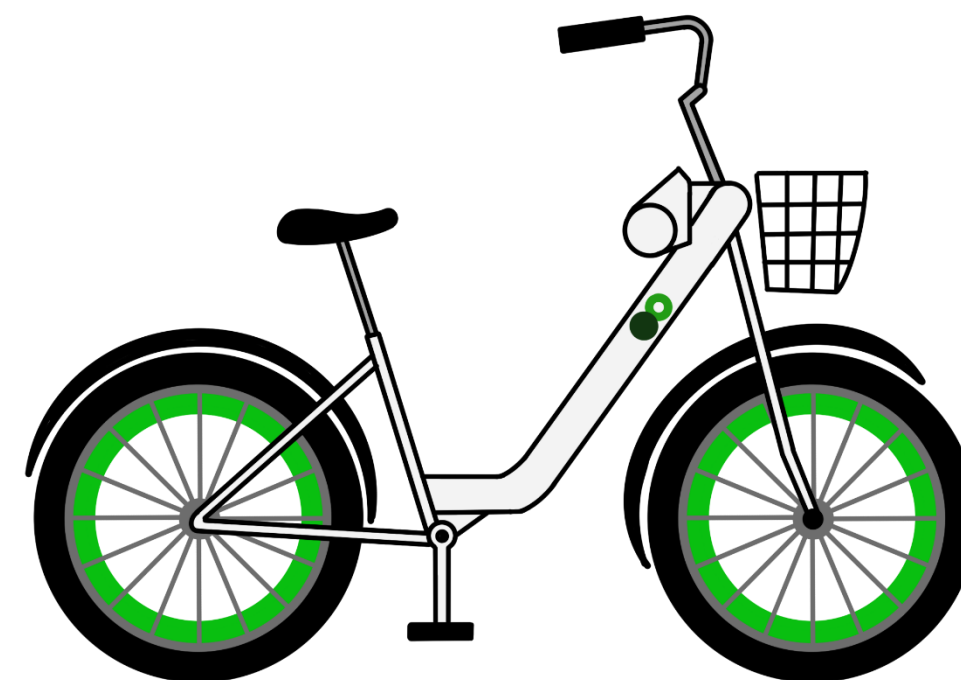
~~버스정류소 개수, 지하철역 개수가 적으면 따릉이 사용량이 증가할 것이다~~

초기의 분석 목표였던 대중교통 접근성과 따릉이 사용량 사이의 관계는  
관련이 없는 것으로 밝혀짐

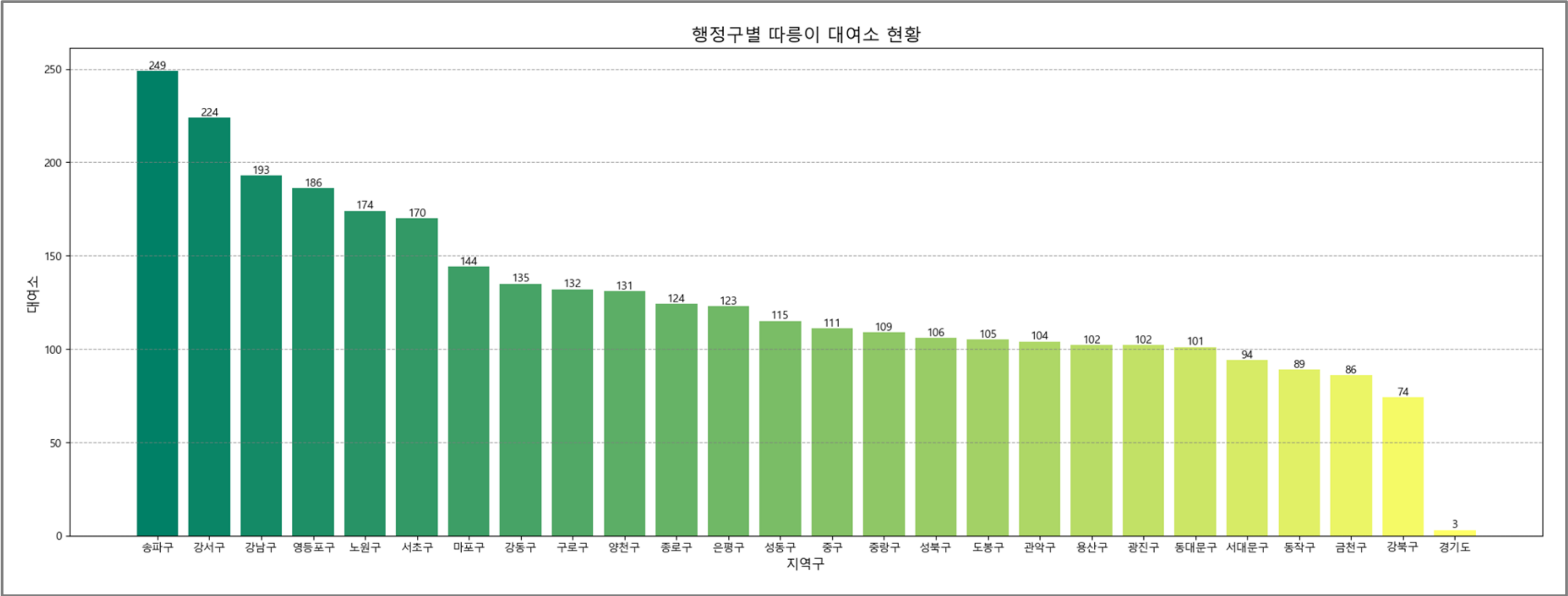
# 03

## 분석 요약 및 시사점

어떤 정보를 알 수 있을까?

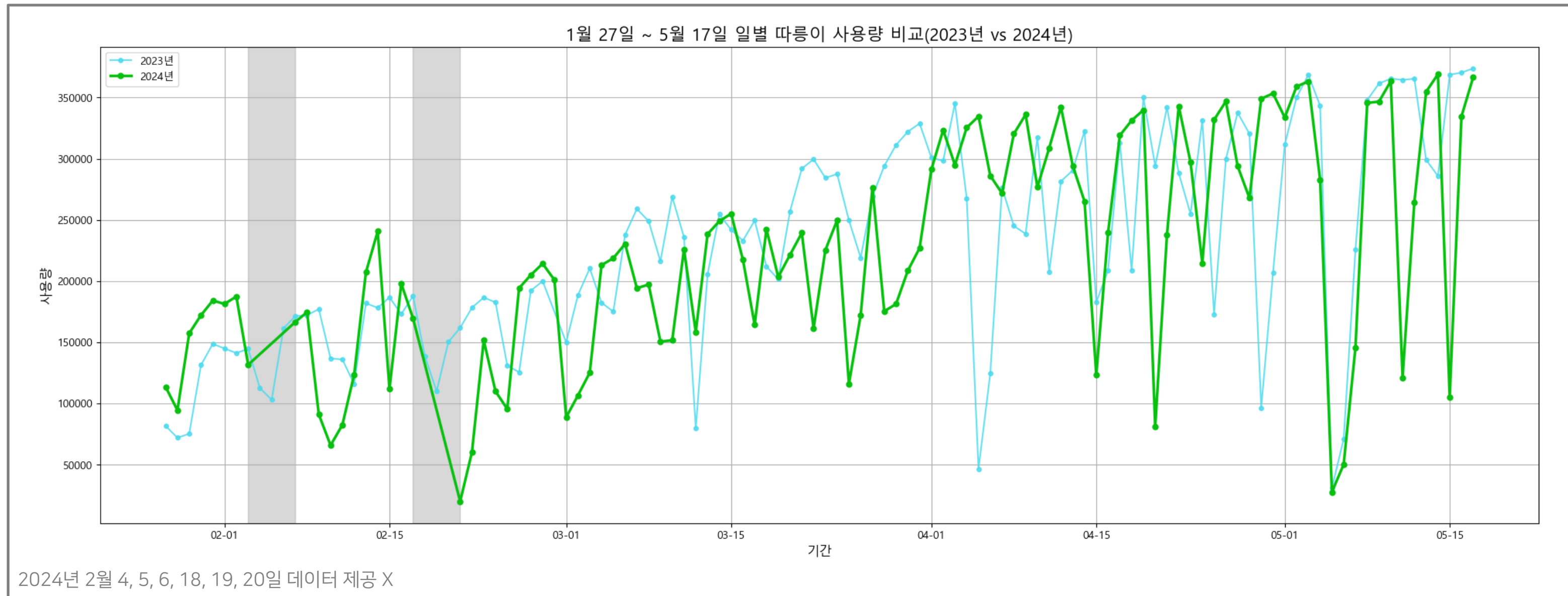


① 행정구별 따릉이 대여소의 개수는 송파구가 제일 많음



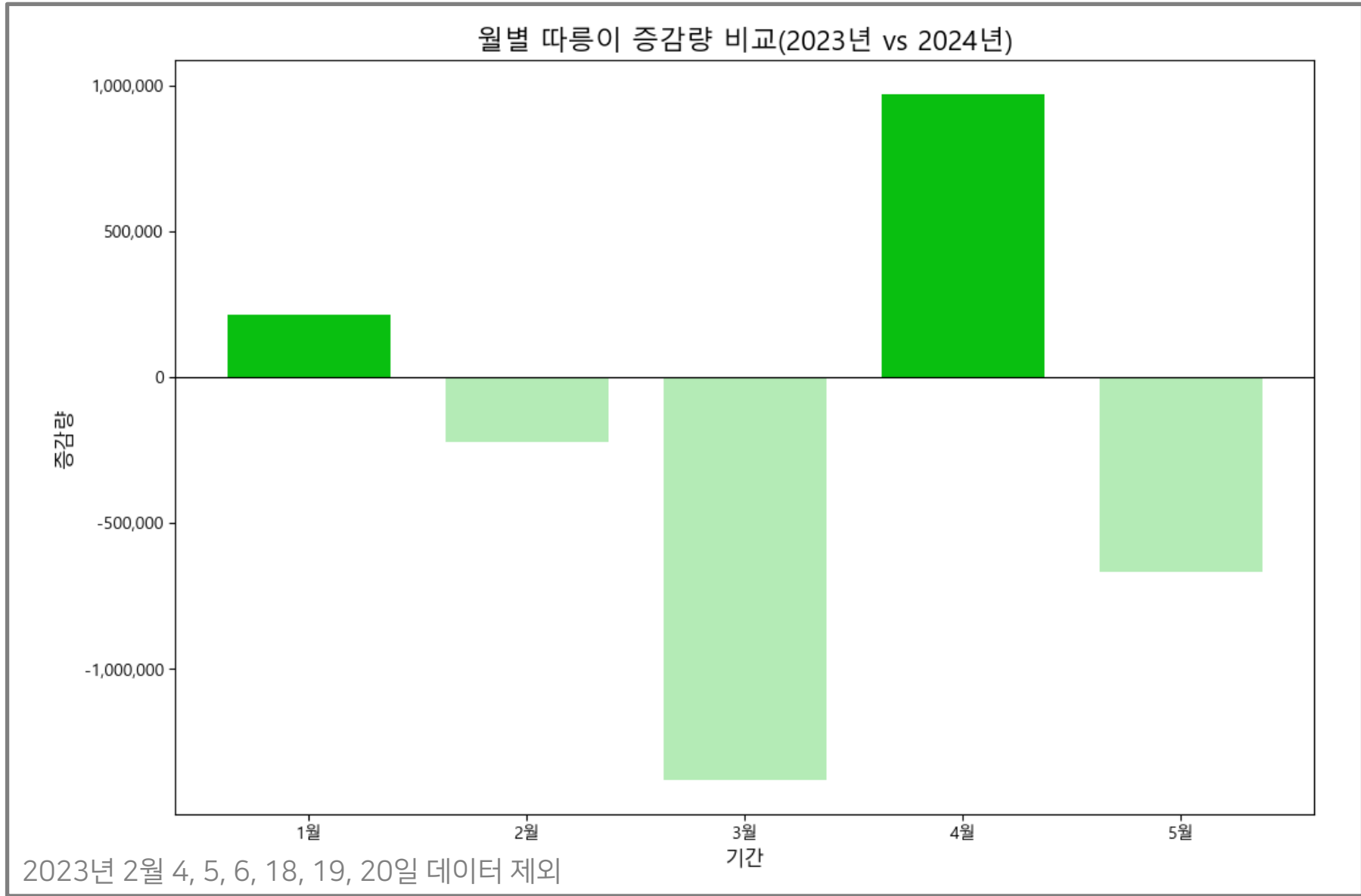
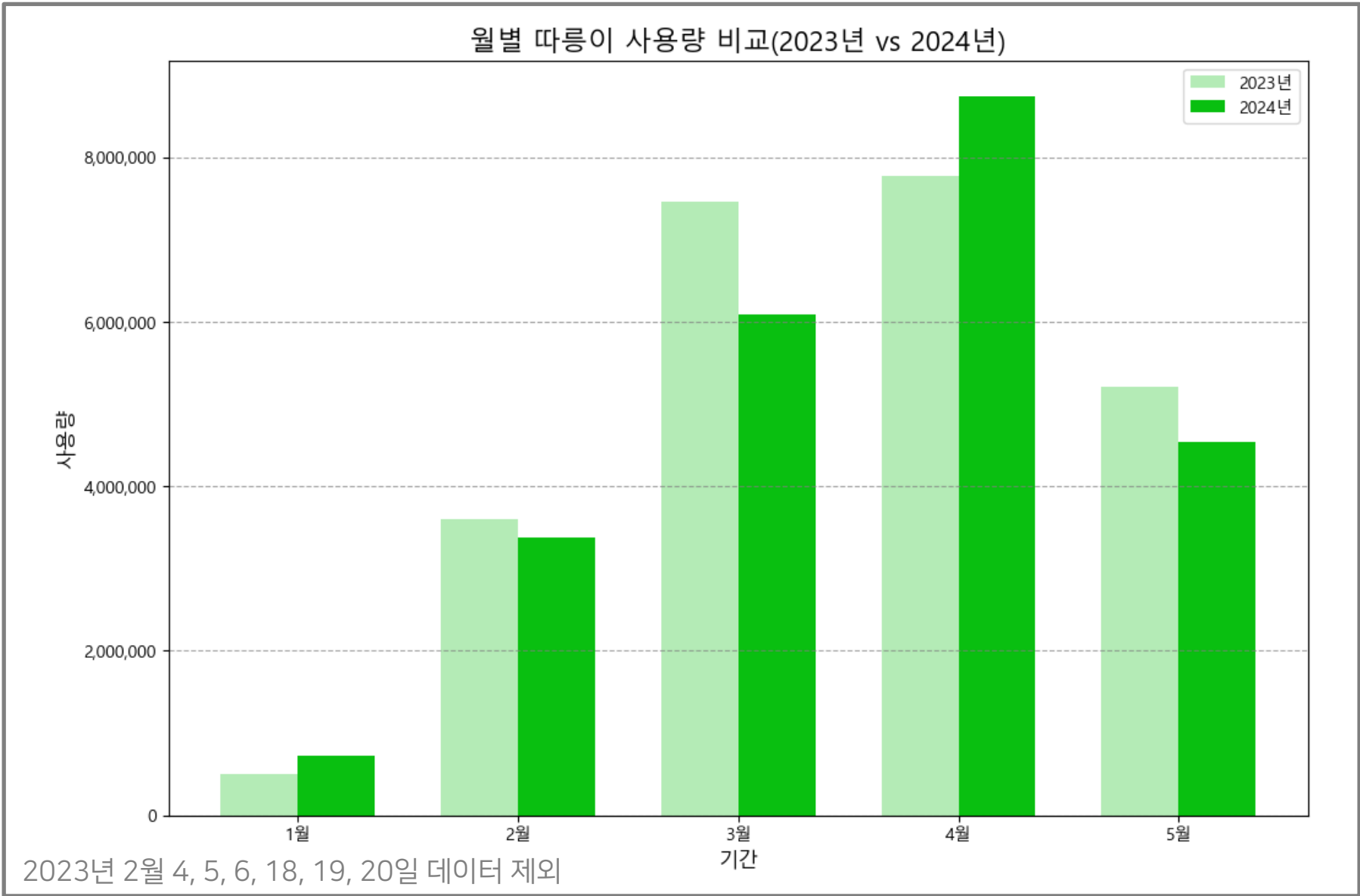
1위는 송파구로 249개, 그 뒤로 강서구, 강남구, 영등포구,  
그리고 노원구 순으로 사용량이 많았음

## ② 일별 따릉이 사용량은 전반적으로 감소한 추세를 보임



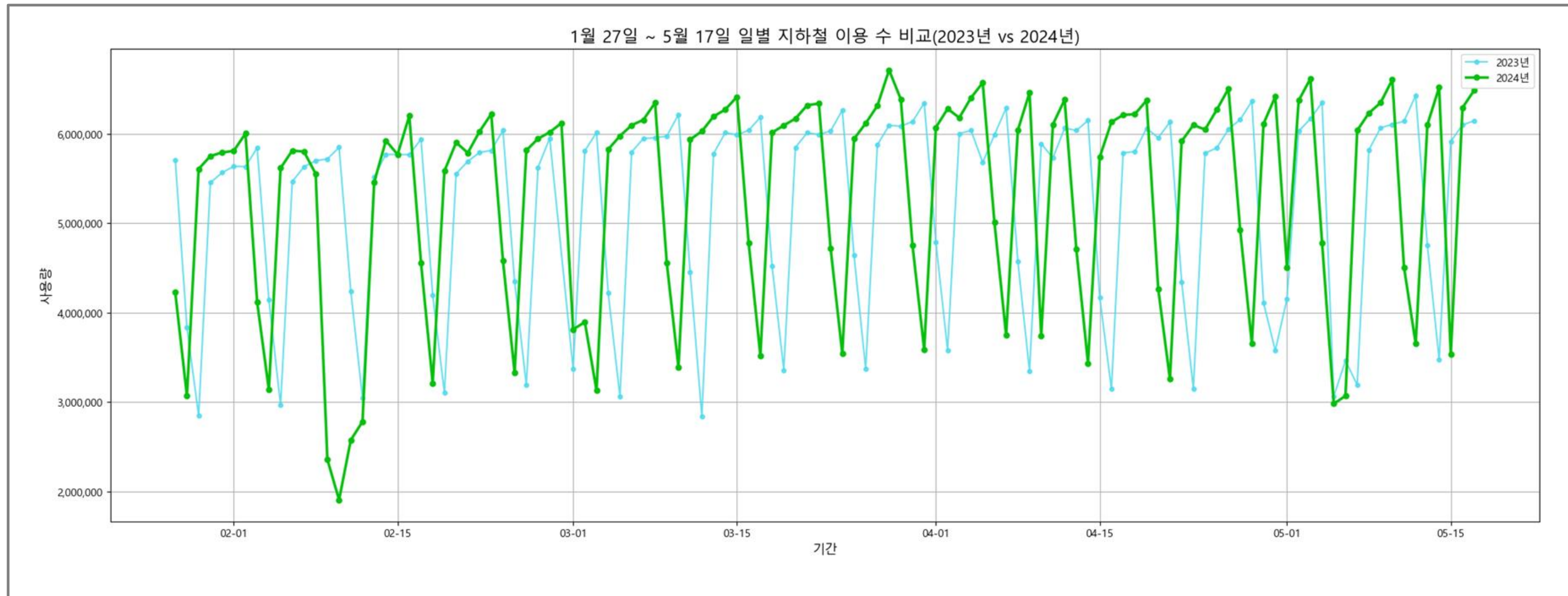
시간이 흐름에 따라 사용량이 꾸준히 증가하는 2023년에 비해  
2024년은 3월까지 사용량이 일정하다가 4월 이후 증가하는 추세를 보임

③ 월별 따릉이 사용량은 3월에 큰 폭으로 감소, 4월에 소폭 증가



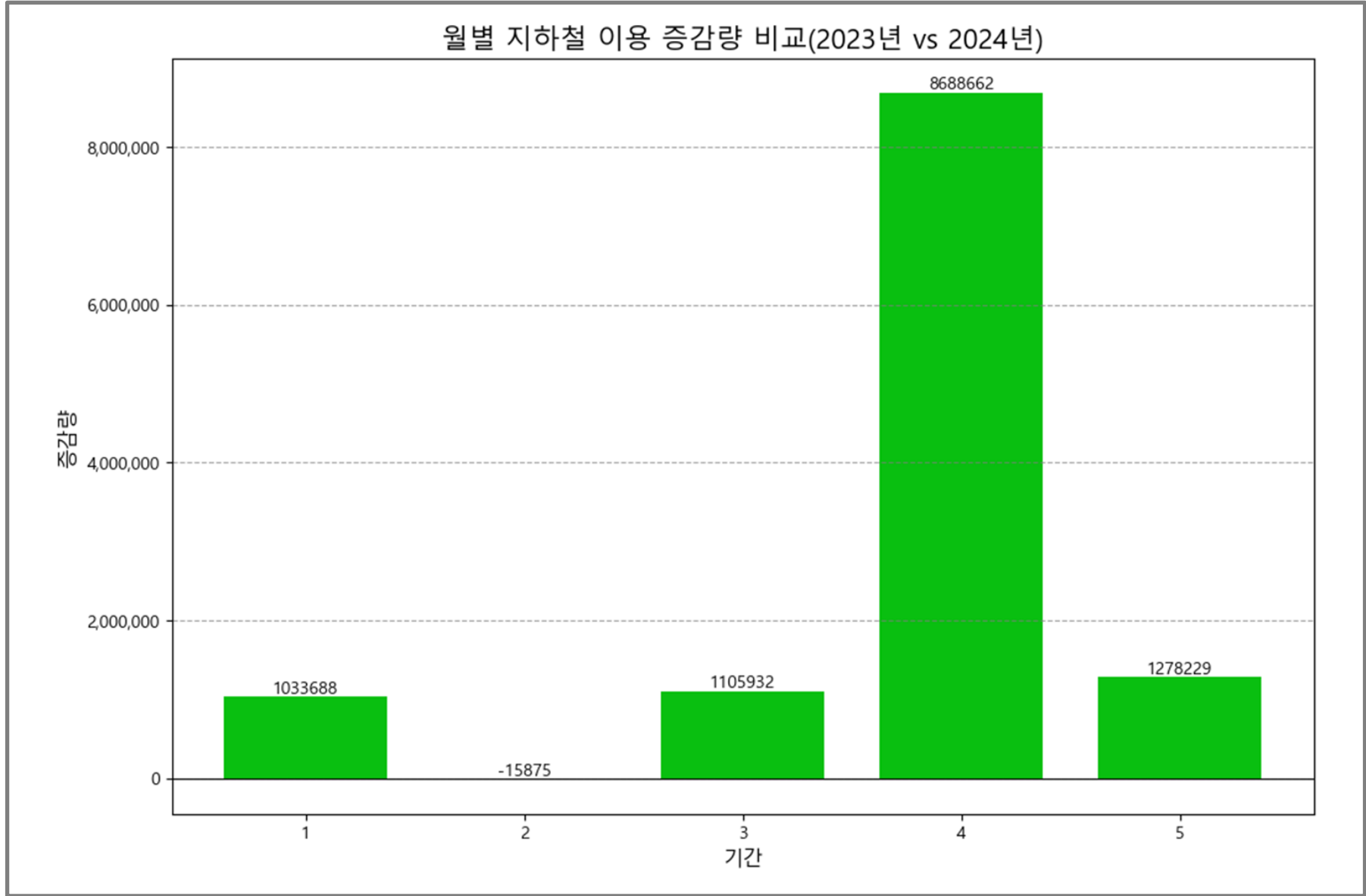
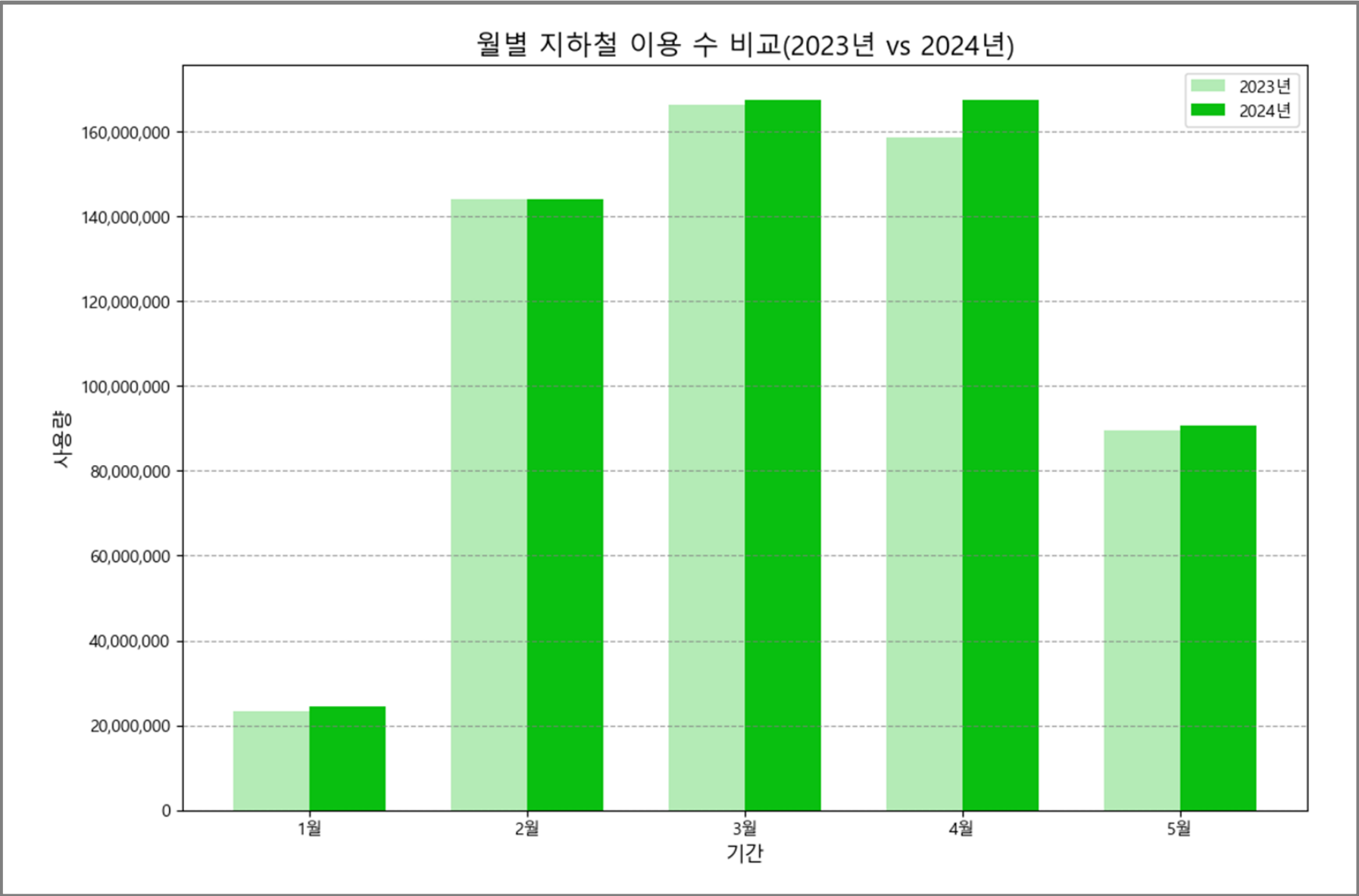
3월에 1,381,099건 감소, 4월에 970,209건 증가하여  
작년 동기 대비 총 1,090,957건 감소함

## ④ 일별 지하철 이용 수는 전반적으로 증가한 추세를 보임



평일에 높은 이용량, 주말에 낮은 이용량을 보이는 계절성을 보이며  
작년 동기 대비 소폭의 상승세를 보임

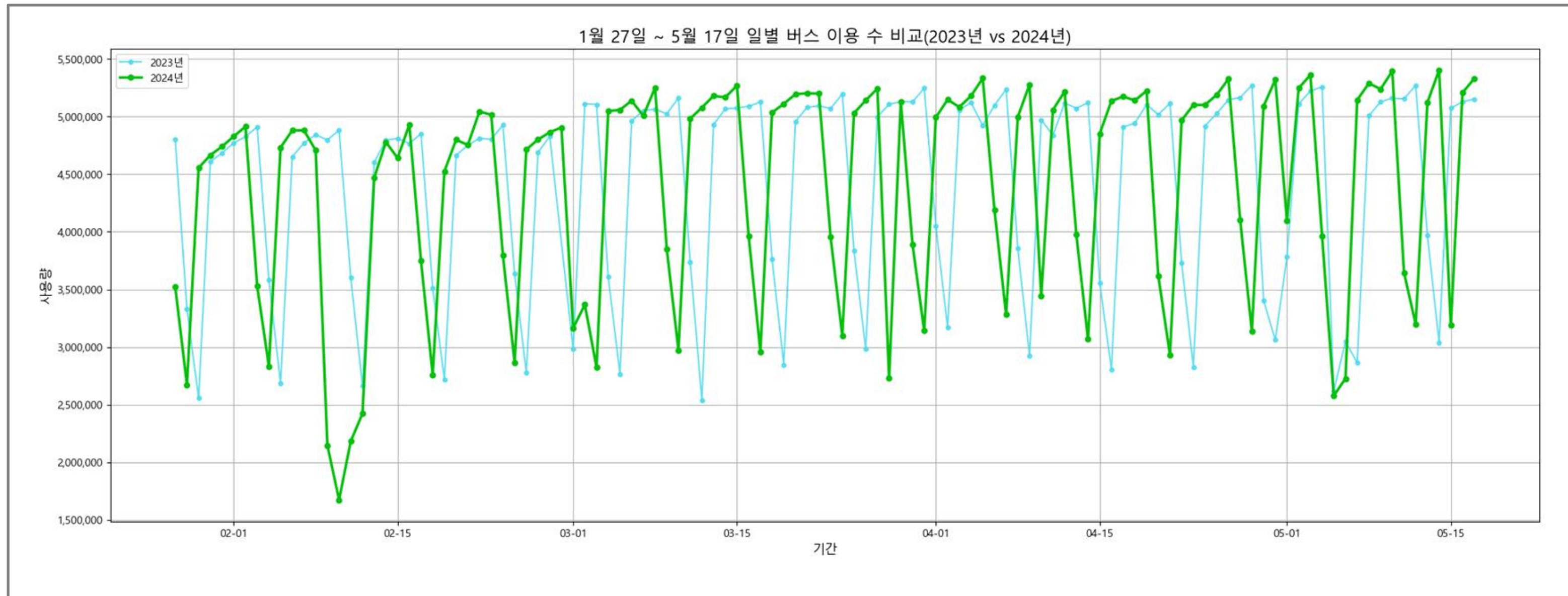
## ⑤ 월별 지하철 이용 수는 4월에 큰 폭으로 증가함



2월에 15,875건 감소를 제외하고 전 기간 증가하여  
작년 동기 대비 총 12,090,636건 증가함



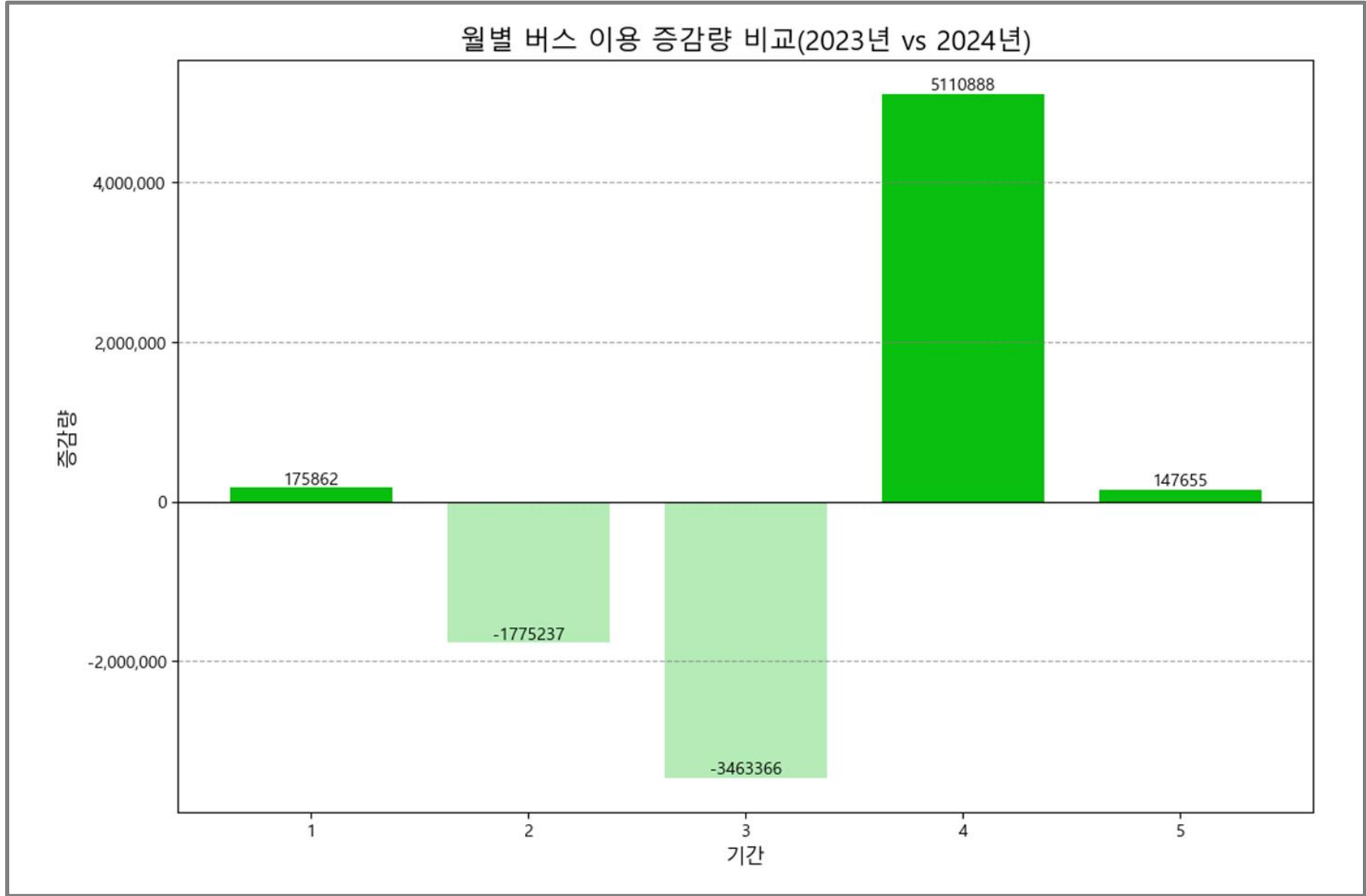
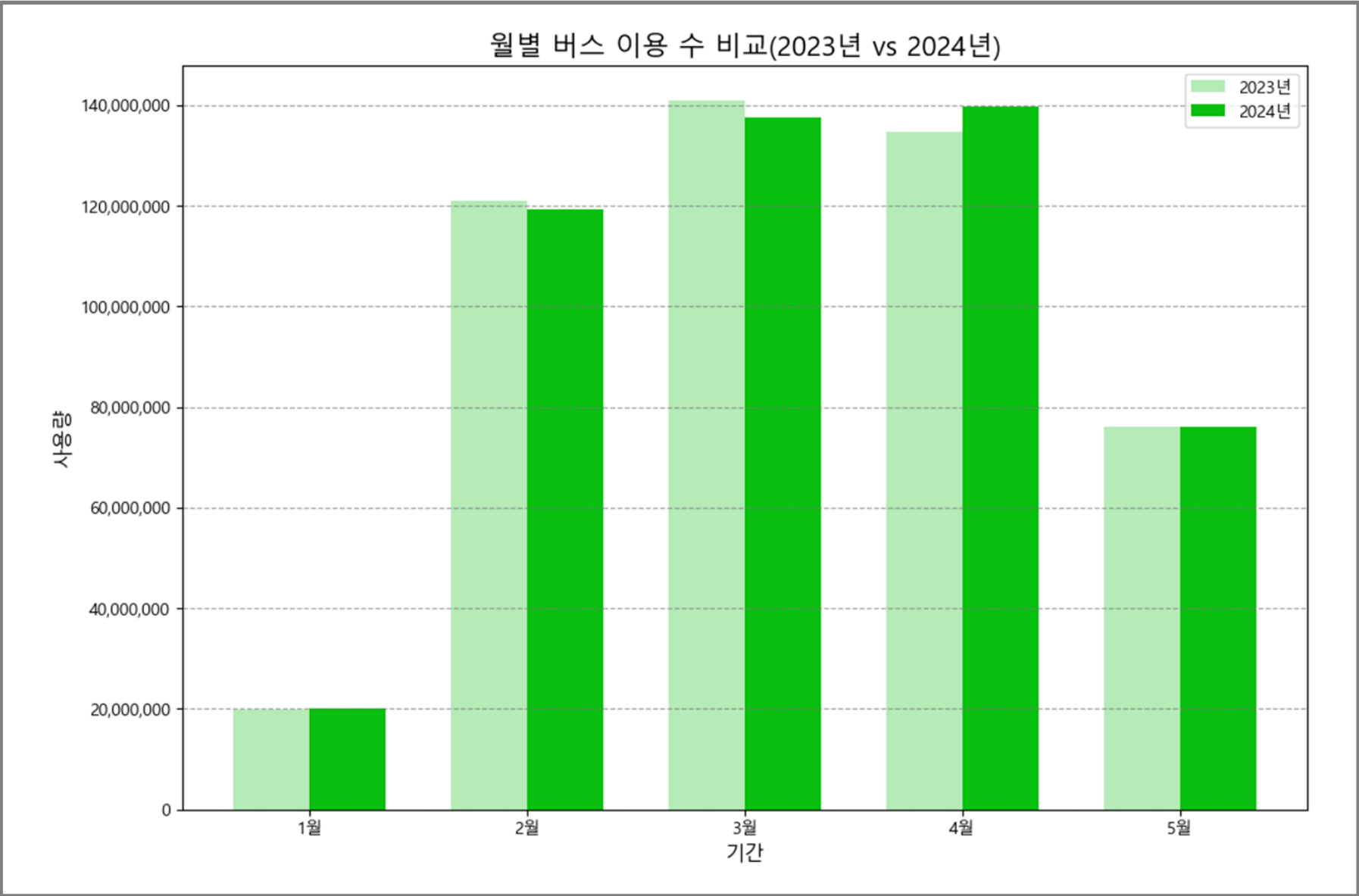
## ⑥ 일별 버스 이용 수는 전반적으로 증가한 추세를 보임



평일에 높은 이용량, 주말에 낮은 이용량을 보이는 계절성을 보이며  
작년 동기 대비 소폭의 상승세를 보임



# ⑦ 월별 버스 이용 수는 4월에 큰 폭으로 증가함



2월 1,775,237건, 3월 3,463,366건 감소를 제외하고 전 기간 증가하여  
작년 동기 대비 총 195,802건 증가함

## ⑧ 2024년 행정구별 따릉이 사용량은 강서구가 제일 많음



1위는 강서구로 2,709,792건, 그 뒤로 송파구, 영등포구, 노원구,  
그리고 양천구 순으로 사용량이 많았음

## ⑨ 2024년 행정구별 1인당 따릉이 사용량은 종로구가 제일 많음



1위는 종로구로 5.71건, 그 뒤로 영등포구, 강서구, 중구,  
그리고 광진구 순으로 사용량이 많았음

② ~ ⑦

기후동행카드는 따릉이의 사용량을 감소,  
대중교통의 이용을 증가시킴

	2023년	2024년	증감량	증감률
따릉이	24,548,453*	23,457,496*	-1,090,957	-4.44%
지하철	581,675,165	593,765,801	12,090,636	2.07%
버스	492,311,553	492,507,355	195,802	0.04%

기후동행카드로 인해 따릉이 사용이 활성화되기보단,  
기존에 따릉이를 사용하던 사람들이 따릉이 대신 대중교통을 이용한 것으로 추정

\* 2월 4, 5, 6, 18, 19, 20일 데이터 제외

⑧, ⑨  
행정구별 따릉이 사용량과 1인당 따릉이 사용량  
상위권 행정구들의 순위가 일치하지 않음

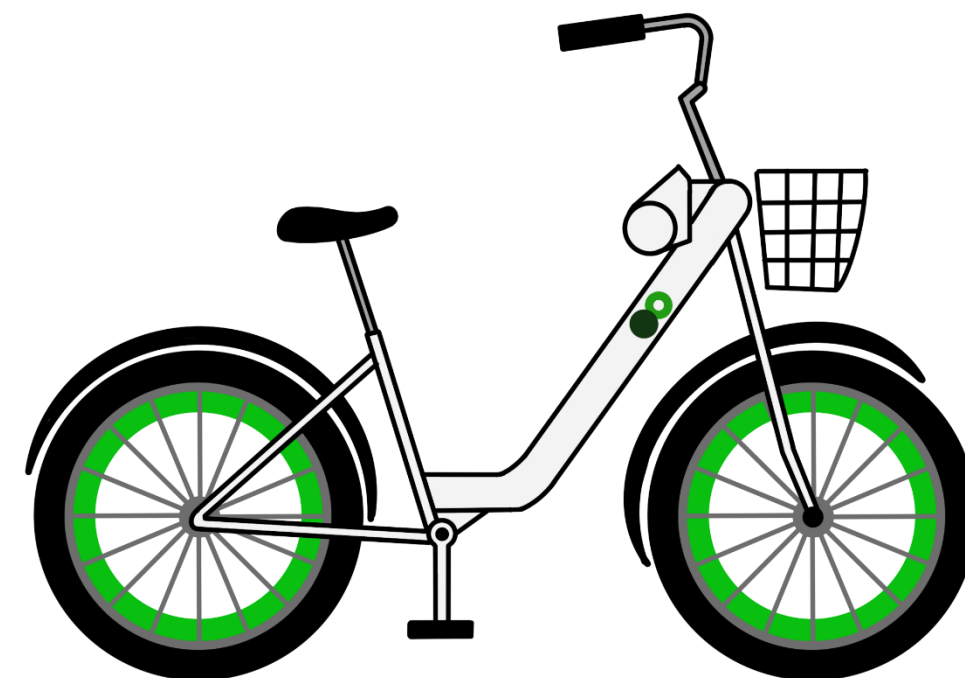
	전체 사용량	1인당 사용량
1위	강서구	종로구
2위	송파구	영등포구
3위	영등포구	강서구
4위	노원구	중구
5위	양천구	광진구

주민등록인구 수에 비례하여 따릉이 사용량이 많을 것으로 예상하였지만,  
따릉이 사용량에 크게 관계가 없는 것으로 확인됨

# 04

## 발전과제 및 한계점

어떻게 발전시킬 수 있을까?





## 따릉이 사용량에 영향을 미치는 변수 찾기

### 한계점

- 1** 성별, 연령별 분포 - 따릉이 사용량 관계 분석 X
- 2** 행정구별 대여소 개수 활용 X
- 3** 따릉이 사용 영향 외부 요인 고려 X (날씨 등)

### 발전과제

- 1** 행정구별 특성 반영하여 따릉이 사용량에 영향을 미치는 변수 찾기
- 2** 머신러닝, 딥러닝을 이용하여 시간대별, 정류소별 따릉이 사용량 예측
- 3** 외부 요인 고려한 따릉이 사용량의 증감량 분석



A row of white bicycles with green accents is parked on a sidewalk. The bicycles are parked in a line, with their front wheels facing towards the left. The background shows a path lined with trees, and a signpost is visible in the distance. The overall scene is dimly lit, suggesting dusk or dawn.

경청해주셔서 감사합니다!