



# 따름이뽑기실패경험 ······<mark>따름이기피현상</mark>으로

따름이고장관리가원활하지않아서울시민의세금을투입하여적자운영중인따름이사업이용자수가점차줄어드는중

서울시 공공자전거 '따릉이' 이용 건수가 올해 상반기 처음으로 감소세로 돌아섰다.

Q

### 朝鮮日報

9만9728건으로 약 16.9%(2만357건) 감소했다.

따릉이 노후화와 방치에 따른 고장이 늘어나면 이용객이 줄어드는 악순환이 이어질 가능성도 있다. 실제 따릉이 대여 건수는 작년 상반기 2165만7097건에서 올 상반기 2141만 3204건으로 약 1.1%(24만3893건) 감소했다. 신규 유입도 마찬가지로 줄었다. 따릉이앱 신규 가입 건수는 작년 상반기 31만5724건에서 올 상반기 31만62건으로 약 1.8%(5662건) 감소했다. 따릉이의 주 이용 연령층인 20대의 앱 신규가입 건수는 12만85건에서

전문가들은 고장률이 높은 게 20대의 등을 돌리게 하는 요인이라고 지적한다. 자전거 온라인 커뮤니티엔 '따릉이 고르는 법' '따릉이 뽑기 실패' 등의 게시글을 흔히 볼 수 있다. 고장이 워낙 잦아 어떤 자전거를 타는지가 중요하기 때문에 고유 번호, 안장 주변부 생김 새 등을 따져 잘 나가는 따릉이를 구별해야 한다는 식이다. 김영성 서울시 보행자전거과 주무관은 "따릉이의 지속적 점검을 통해 안전한 이용이 되게끔 할 것"이라고 했다.

### 공유킥보드에 밀려…한풀 꺾인 '따릉이' 인기

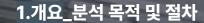
공공자전거인 따름이의 이용률이 줄어든 데는 민간 PM 확장이 영향을 미쳤단 분석이다. 카카오T바이크, 쏘카일레클, 지쿠 등 주요 민간 PM업체는 전기자전거와 전동킥보드의 편의성을 앞세워 고속 성장했다. 한국PM산업협회에 따르면 2020년 7만 대이던 국내 공유 PM은 지난해 29만 대로 네 배 이상으로 급증했다.

사업 건수는 12만85건에서 따름이는 1시간 이용권이 1000원으로 저렴한 비용이 장점이다. 민간 PM인 킥고잉은 따름이는 고장이 많다는 인식이 이용 감소에 영향을 준다는 의견 카일레클은 9500원에 달한다. 그럼에도 PM을

이용하는 소비자는 늘어나는 추세다. 전기 동력을 사용하는 PM은 속도가 빠르고 체력 소모가 적으며 대여 및 주차가 쉽다. 따름이는 대여소에서만 대여와 반납이 이뤄져

따릉이 관리 인력 확충이 없어 고장관리 운영의 효율화 필요

잦은 고장도 문제다. 서울시설공단에 따르면 고장 등 이유로 지난해 4500대의 따름이가 폐기됐고 13만7238대가 정비를 받았다. 지난 4년 동안 따름이 대수는 52.5% 늘었지만 운영 예산은 그대로라 관리인력 확충도 어려운 상황이다.●



# 문제생활 따름이관리인력부족 + 따름이고장에의한이용자유입감소

# 분석목전 따름이고장예방및관리효율화

따름이고장데이터분석으로

따름이고장과상관관계가있는

내부/외부요인확인

따름이고장을보다 효과적/효율적으로

관리할수있는방안제안

분석절차

**분석설계** (과제범위설정,데이터선정,) 데이터 탐색 (BDA등,) **기설설정** (시각화및인사이트도출) 기설검정 (통계검정법적용) 결론 및 제안 (검정결과를비탕으로)

# 1) 따릉이 별 고장내역과 연도별 날씨 데이터를 통합하여 날씨-고장데이터셋생성

따릉이고장내역을연도별로나눠서날씨데이터와통합하기위한전처리진행

#### 사용데이터

데이터전처리과정

서울시공공자전거고장신고내역 (2021.01~2023.12)

-출처: 서울열린데이터광장
-모든따릉이고장내역이월별로업데이트되는데이터
-연도별데이터로취합하여활용

\*4계절별 분석 위해, 1~6월만 있는 2024년 데이터는 제외

종합기상관측(ASOS)

 $(2021.01 \sim 2023.12)$ 

-출처: 기상청기상자료개방포털

-일별날씨데이터

-고장신고건수, 부품별고장건수데이터생성에사용

고장신고 데이터 연도별 데이터셋 병합 및 분리

고장건수 컬럼 생성, 각 일시별 고장구분 카운팅 고장신고이력 '월별', '일별' 부품별 고장 컬럼 생성

일시에 따른 날씨 컬럼 결측치 제거 (평균기온,회저기온,회고기온, 일강수량,평균상대습도', 일최심신적설,평균지면온도') 전처리된 고장신고내역, 종합기상관측(ASOS) 데이터셋병합처리

# 1) 자전거 별 고장내역과 연도별 날씨 데이터를 통합하여 날씨-고장데이터셋생성

자전거고장 내역을 연도별로 나눠서 날씨 데이터와 통합하기 위한 전처리 진행

Ⅰ 날씨별 고장 데이터셋 컬럼 소개

자전거 번호

개별따름이각각에게부여되는따름이인식번호

평균기온

하루동안관측하여측정한기온의평균수치형데이터

등록일시

개별따름이의등록된시간을알려주는날짜형데이터

평균최고기온

일사량이가장높은일간최고기온의평균 수치형데이터

고장구분

자전거의 고장을 분류하는 범주형 데이터

최저기온

**특정시간대에나타나는가장낮은기온** 수치형 데이터

평균지면온도

**하루동안지표면의평균온도** 수치형 데이터

# 2) 따름이 별대여이력과고장신고내역데이터를 통합하여 따름이 추적데이터셋생성

자전거 사용 이력을 추적하고 고장원인을 분석하기 위한 데이터 전처리 진행

#### 사용데이터

### 서울시공공자전거대여이력 (2020.01~2024.06)

-출처: 서울열린데이터광장

- -모든따름이일별대여내역이월단위로업데이트되는데이터
  - -연도별데이터로취합하여활용
- '평균이용시간', '평균이용거리', '총대여건수' 데이터생성에사

픙

## 서울시 공공자전거 고장신고 내역 (2020.01~2024.06)

-출처: 서울열린데이터 광장

-연도별따름이고장신고내역데이터

-고장신고횟수,고장신고건수데이터생성에사용

#### 데이터전처리과정

### 대여이력에서 불필요한 컬럼 삭제 대여대여소번호', 반납대여소번호', 대여일시'',반납일시'

이용거리와이용시간이 음수이거나 0인값은 결측치처리하여 삭제

따름이모델스펙상불가능한 시속36km이상이용내역삭제 (출처링크)

### 고장신고건수컬럼생성, 각연도별고장신고횟수키운팅

자전거번호기준으로 고장내역데이터와통합, 고장여부컬럼생성 고장신고가1회이상있을시1, 한번도없을시0으로표기

자전거번호컬럼기준으로 Groupby한데이터셋생성 이용거리, 이용시간 평균값과 총대여건수 컬럼생성

# 2) 자전거별대여이력과고장신고내역데이터를통합하여자전거추적데이터셋생성

신규데이터셋생성

I 자전거 추적데이터셋컬럼소개

자전거번호

개별따릏이각각에게부여되는따릏이인식번호

총\_대여건수

개별따름이의연도별대여건수합계수치형데이터

평균\_이용시간

개별따름이의연도별이용시간평균값수치형데이터

총\_고장신고건수

개별따름이의연도별고정신고건수합계수치형데이터

평균이용거리

개별따름이의 연도별이용거리 평균값수치형데이터

고장여부

고장신고기록이있을시1,없을시0으로표기된 범주형데이터

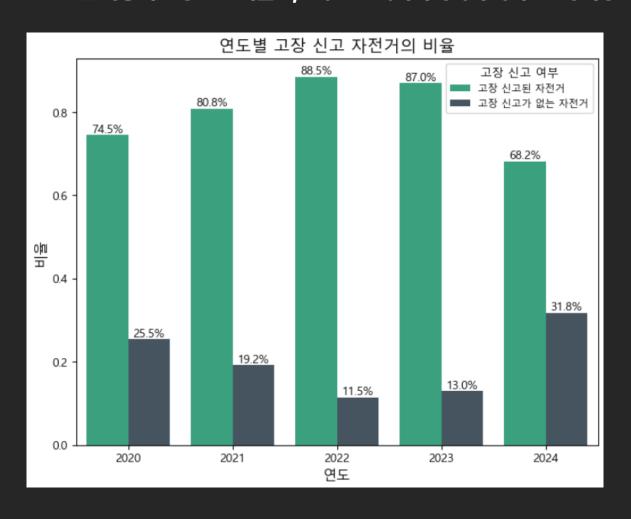
연도

해당행의데이터연도



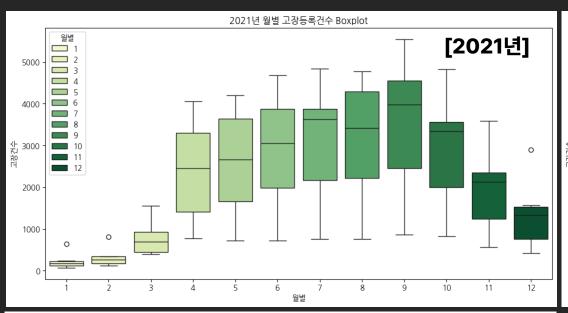
## 대부분의따름이는1년동안1회이상의고장신고를받음

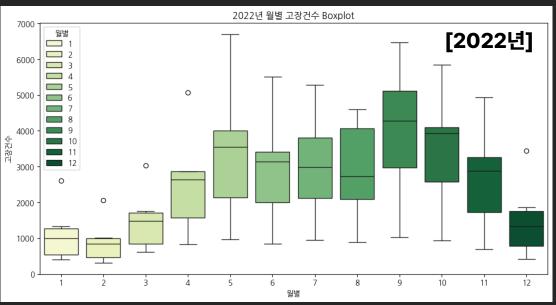
연도별따름이고장신고비율로, 1개연도대여이력에서적어도 1회 이상고장신고가존재하는따름이의비율확인가능

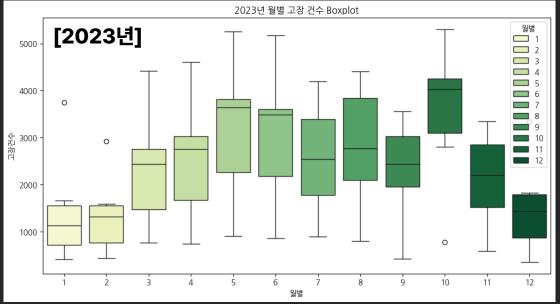


- 고장신고없이 1년 동안 운영되는 따름이는 전체의 11~19% 내외이다.
- 따름이는 1년 주기로 점검을 진행하기에,
   1년 주기의 대여기록 안에서 집계하였다.
- 20년 데이터는 QR따름이 교체 사업으로 데이터 수집 에 결측값이 많았다.
- 24년 데이터는최신화되지않아, 1~6월 데이터만존재하고, 이에따른고장피로누적이 낮아서적은고장신고 율을보이는것으로추정된다.

## 2.현황\_고장신고 연도별-월별 현황 ① 따름이 고장 신고내역 연도별 - 월 별 시각화





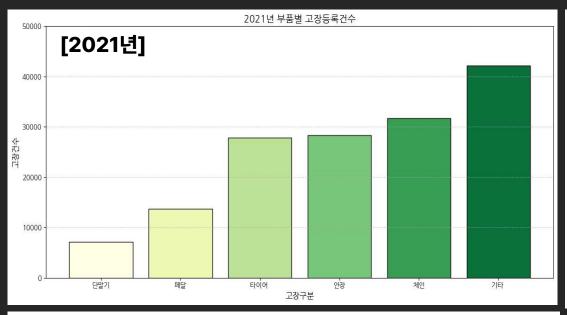


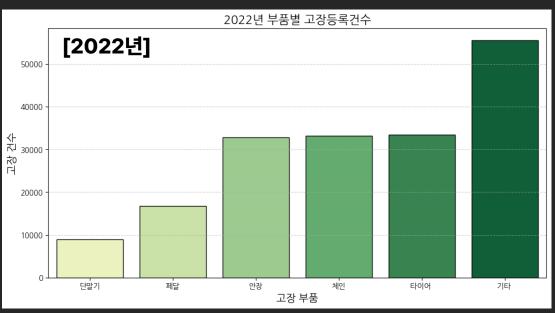
## 계절별 패턴 분석

- ✔ 여름철 고장 증가
- : 2021, 2022 년도에는 9월, 2023 년도에는 10월에 고장 건수 최고치를 기록
- ✔ 겨울철 고장 감소
- : 2021, 2022, 2023 년도 모두 1월, 2월, 12월에 고장 건수가 현저히 감소

#### 2.현황\_부품별 고장현황

## ② 따름이 부품별고장 등록 건수







## 주요 고장 부품 순위

: 기타 > 타이어 > 체인 > 안장 > 페달 > 단말기 순서로 고장 등록 건수가 많음

> 서울특별시 공공데이터 공개 사이트를 통해 확인해 본 결과, 기타 항목에는 전조등, 후미등, 프레임 등과 같은 세부적인 부품들을 일괄적으로 포함

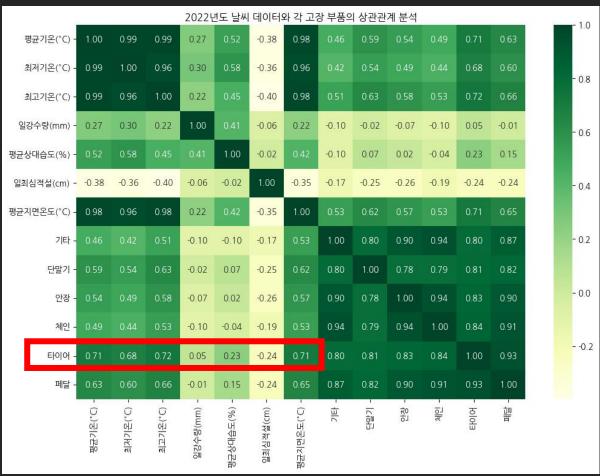
#### 2.현황\_날씨-고장 연도별 현황

## ③ 날씨데이터와 각고장 구분의 상관관계

## [2021년]

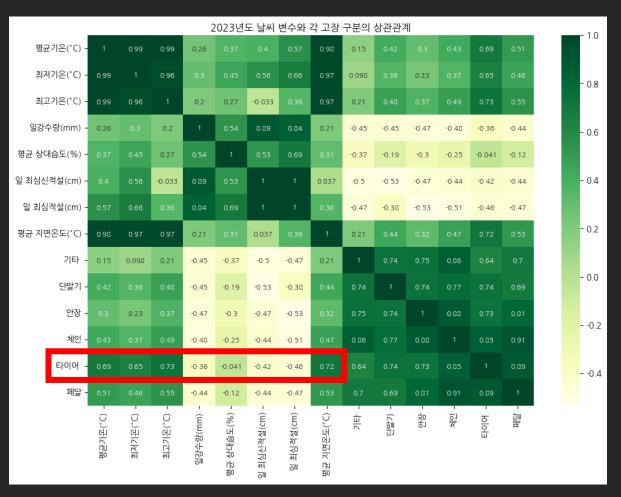


## [2022년]



## 2.현황\_날씨-고장 연도별 현황 ③ 날씨 데이터와 각 고장 구분의 상관관계

## [2023년]



### 날씨데이터 – 고장 분석

✔ 기온 관련

다른 날씨 변수들에 비해 고장 구분과 높은 상관관계 ✔ 습도 관련

평균 상대습도는 고장 구분과 미미한 상관관계를 보임

✔ 강수 및 적설 관련

강수량과 고장 구분 간의 전반적으로 낮은 상관관계 적설량은 대부분의 고장 구분과 음의 상관관계를 보임

## ③ 날씨데이터와 각고장 구분의 상관관계 (분석 방향성 설정)

### 분석 방향성에 따른 고장 구분 재분석

### ✔ 고장 빈도 측면

- 타이어의 경우, 단말기 & 안장 & 페달 & 체인 등 다른 고장 구분과 비교했을 때, 기타를 제외한 고장 구분 중 가장 높은 고장 등록 빈도를 나타냄

### ✓ 기온과의 상관관계

- 2021 년 ~ 2023 년의 3개년 데이터를 분석하였을 때, 타이어는 기온관련 변수들과 일관된 높은 상관관계를 나타냄 ※ 높은 신뢰성
  - 평균기온, 최고기온, 최저기온, 평균지면온도와의 상관계수가 타 고장 구분 대비 높음

## 날씨 데이터와 고장 구분 분석 방향

- ✔ 기온 타이어 관계성 분석
- 기온 변화에 따른 타이어 고장 패턴 심층 분석 - 월별 타이어 고장 특성 파악

- ✔ 예방적 관리 전략 수립
- 기온 변화에 따른 타이어 관리 방안 도출

## ③ 날씨데이터와 각고장 구분의 상관관계(강수 관련 데이터 제외 사유)

### 강수 관련 데이터 제외 사유 ① 높은 결측치 비율

일별 연 단위 데이터셋 ☞ 365 행으로 구성 이 데이터 중 상당수가 결측치로 확인

	일강수량 량	일최심신적설 설	일최심적설
결측치	215	356	349

- 일강수량: 215 일 결측 → 약 58 % 의 데이터 손실

- 일최심신적설 : 356 일 결측 → 약 **97** % 의 데이터 손실

- 일최심적설: 349 일 결측 → 약 96 % 의 데이터 손실

## 강수 관련 데이터 제외 사유 ② 데이터 처리 한계

미세 강수량의 경우, 결측치가 아닌 "0" 으로 기록 때 데이터 보정에 한계 존재

일반적인 결측치 처리 방법의 적용이 부적절 EX) fillna()

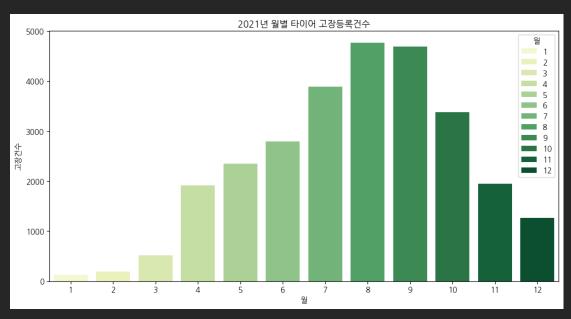
#### # 최심신적설

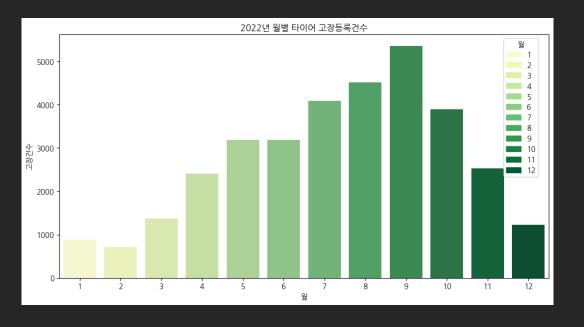
: 고려하고 있는 기간 동안 새롭게 쌓인 눈이 가장 두껍게 쌓여 있을 때의 눈의 두께

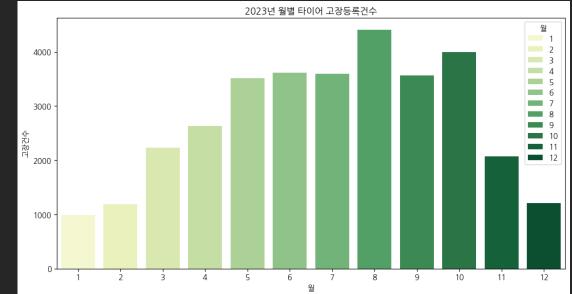
#### #최심적설

: 고려하고 있는 기간 동안, 전부터 내려 녹지 않고 쌓여 있을 눈을 포함하여 가장 두껍게 쌓여 있을 때의 눈의 두께

## 2.현황\_ 날씨-타이어 고장 현황 ④ 타이어 고장 데이터 심층 분석







## 연도별 일관된 패턴

- 2021년 ~ 2023년간 타이어 고장 건수가 매우 유사한 패턴
- 8월 ~ 10월에 최고점 기록 & 1월 ~ 2월에 최저점 기록

## 따릉이 외부적 요인과 고장 신고 간의 관계를 살펴본 결과

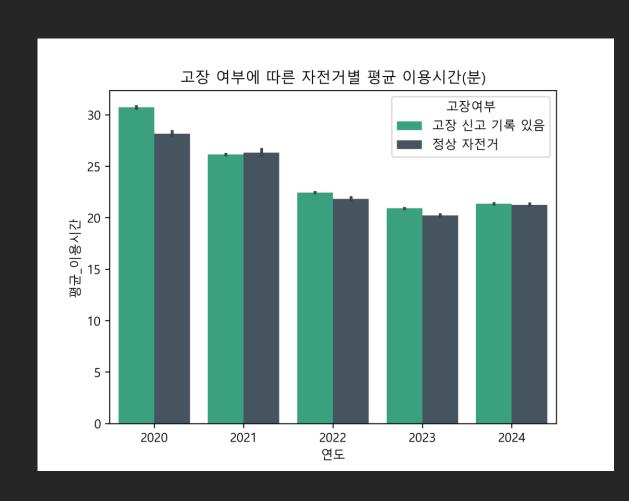
Insight ▶ 날씨요인중기온이 고장유형중 타이어고장과의 상관관계가있는 경향

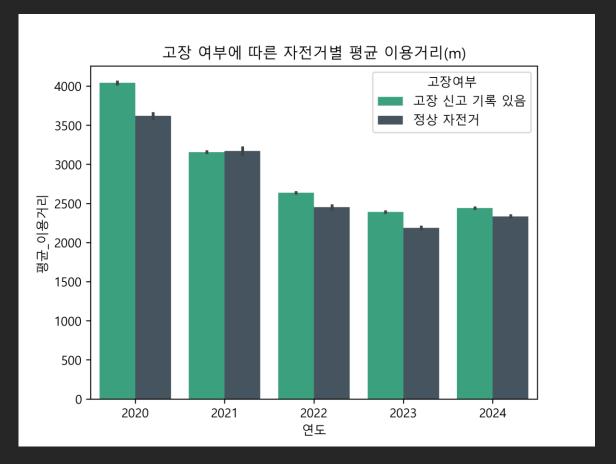
Next step ▶

실제로기온이따름이타이어고장과 통계적으로유의미한관계를 갖는지 검정

# 고장신고여부에따라각따름이의평균이용시간및평균이용거리는 차이존재

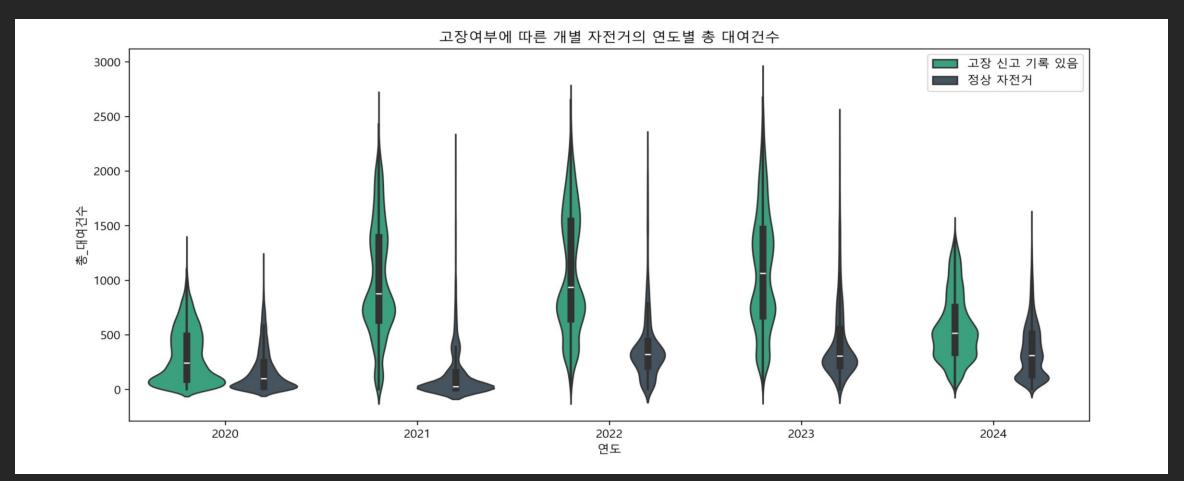
대체로 고장신고가 있는 따름이의 평균 이용시간/이용거리가 길지만 차이가 크지 않으므로 세부적인 분석 요망



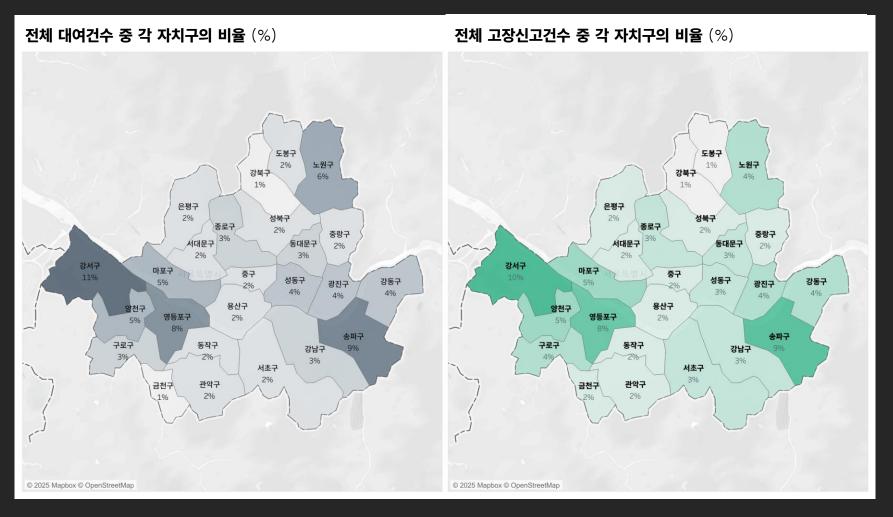


# 고장신고를 받은 따름이의 총대여건수가 높은 경향확인

고장 신고가 있는 자전거의 총 대여건수 평균이 더 높은 곳에 분포되어 있으며, max값도 더 높은 경향을 확인할 수 있다

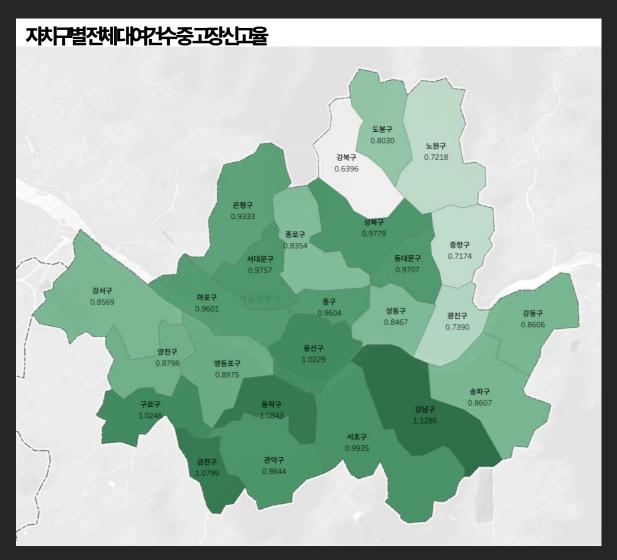


# 따름이 대여건수가 많은 자치구에서 고장신고도 높게 나타난다



대여건수 비율이 높은
 자치구의 분포와
 고장신고 건수 비율이 높은
 자치구의 분포가
 유사하게 나타나는 것을 확인

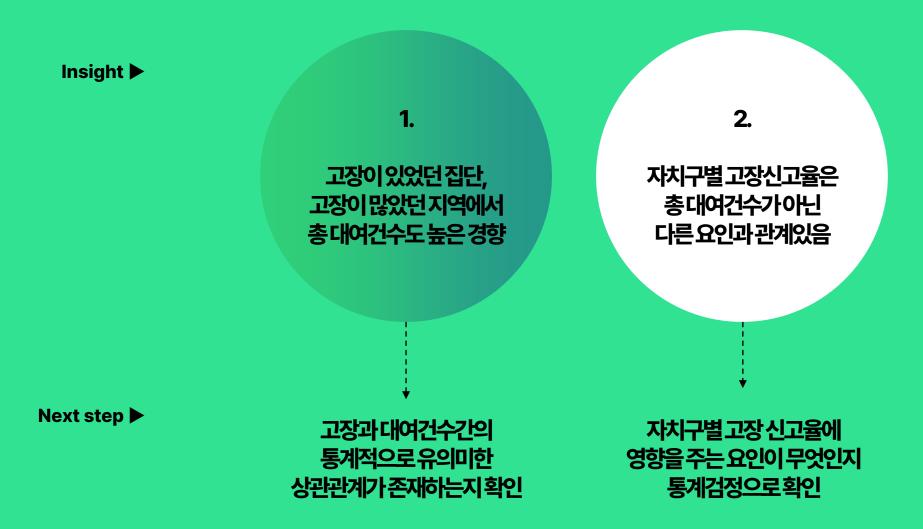
## 자치구별따름이고장신고율은대여건수와다른패턴



- 대여건수 높은 지역과 고장 신고율 높은 지역은 양상이 다름
- 자치구별로 고장신고율은 차이가 있을지
   ANOVA 검정
  - **검정 통계량**: 4.1343
  - P-값: 0.0
- 자치구별 고장 신고율은 유의미한 차이 있음
- 고장 신고율이 높은 상위 5개 지역은 강남구, 동작구, 금천구, 구로구, 용산구

### 2.현황\_따름이 내부요인 인사이트

## 따릉이 내부적 요인과 고장 신고 간의 관계를 살펴본 결과







## "날씨(기온)는 따름이 고장에 유의미한 영향을 미칠 것이다"

### 가설 설정

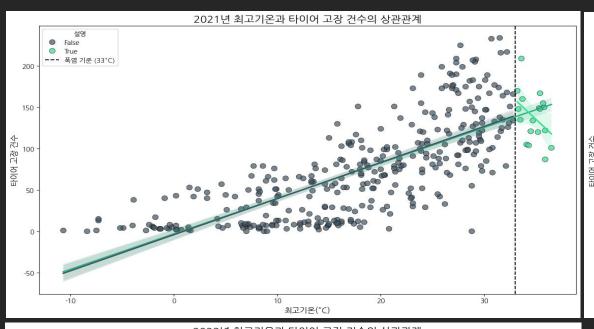
✓ 귀무가설 (HO)

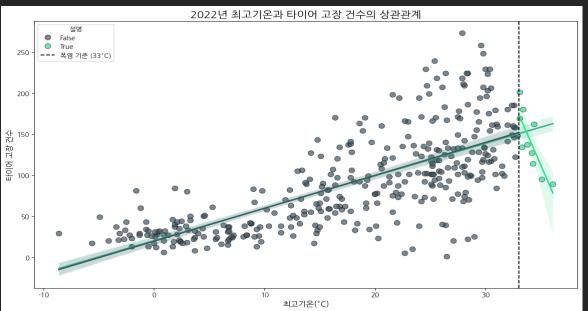
" 날씨(온도)는 따릉이 타이어 고장 건수에 유의미한 영향을 미치지 않는다. "

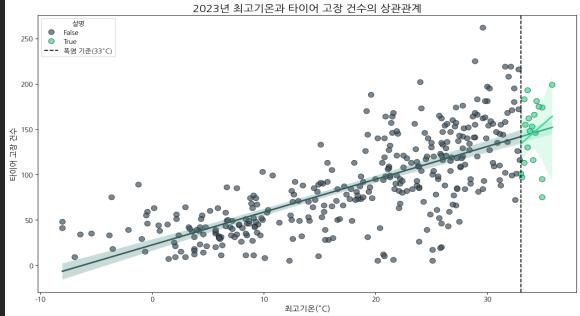
✓ 대립가설 (H1)

" 날씨(온도)는 따릉이 타이어 고장 건수에 유의미한 영향을 미친다. "

# "날씨(기온)는 따름이 고장에 유의미한 영향을 미칠 것이다"







### 시사점

- 전반적으로 기온이 상승할수록 타이어 고장 등록 건수가 증가하는 경향
- · 21~22년은 폭염에 따른 음의 상관관계, 23년은 양의 상관관계

이는 폭염일에 따른 표본의 크기에 의해 나타나는 현상으로 판단 유의한 상관계수가 아니라고 판단됨

## Q1.날씨와따름이타이어고장여부와상관이있을까?

# 다중화귀분석결과:날씨외 타이어의 상관관계는 통계적으로 유의미

### 통계검정설계

분석목적 타이어와 날씨 변수간 상관관계 파악

기설설정 귀무기설\_외부적(날씨) 요인에 의해 따름이 고

장이 발생할 수 없다.

대립가설\_외부적(날씨) 요인에 의해 따름이 고

장이 발생할 수 있다.

독립변인 평균기온, 최고기온, 평균지면온도

<u>종속변인</u> 타이어

### 통계검정결과

R 계수:

2021년 : 0.43, 2022년 : 0.52, 2023년 : 0.

89

최고기온 p-value가 0.00

**상관계수 결과 3개 년 평균** 0.61**이며** 

**독립변수의** p < 0.05 **이므로** 

기온과 타이어 간의 관계는 유의하다고

판단됨.

## "날씨(기온)는 따름이 고장에 유의미한 영향을 미칠 것이다"

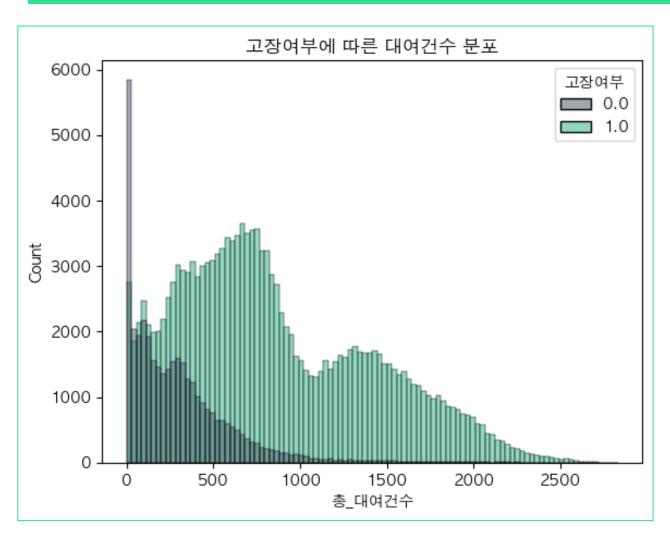
결론

기온 (특히 최고기온)은 타이어 고장 건수에 긍정적인 상관관계를 보임
BUT, 다중공선성 문제로 인해 명확한 회귀분석 결과 도출하지 못함
시각적 분석을 통해 기온이 증가함에 따라 타이어 고장이 증가하는 경향 확인



## Q1.서울시공공자전거대여건수는자전가의고장여부와상만이있을까?

# 고장닌 적 있는 따름이기 비교적 대여건수 많은 쪽에 분포하는 경항



- 고장난적없는집단은대여건수가적은쪽에밀집, 대여건수가많이질수록감소하는경향
- · 반면에고장난적있는집단은 비교적대여건수가많은쪽에분포되어있음

### Q1.서울시공공자전거대여건수는자전가의고장여부와상만이있을까?

# 로지스틱 화귀분석 결과:대여건수외 고장여부의 상관관계는 통계적으로 유의미

### 통계검정설계

분석목적 종속변수인고장여부가 더미변수로, 독립변수인 대여건수의 수치형 데이터와의 관계를 검증하기 위해 "로지스틱 회귀분석" 사용

가설설정 귀무가설

:따릉이 별총대여건수는 자전거의 고장여부와 상관이 없다

대립기설

:따릉이별총대여건수는지전거의고장여부와상관이있다

독립변인 따름이별1년내총대여건수

종속변인 따름이의고장여부

### 통계검정 결과

회귀식:고장여부~-0.3272+0.0034\*총\_대여건수

R<sup>2</sup>:0.202

총대여건수의 p-value가 0

p-value**는 0으로 통계적으로 유의미** 

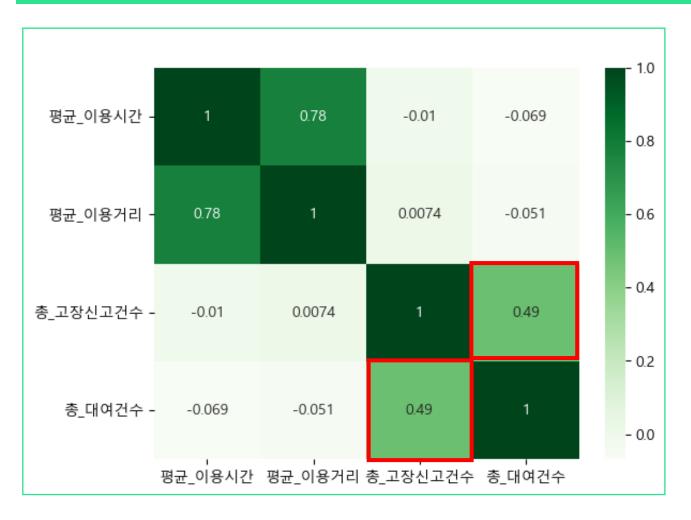
회귀계수는 0.0034으로 나타남

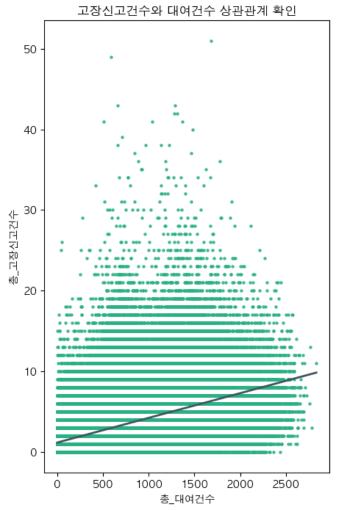
그러나 회귀계수가 0에 가까운 값이기 때문에 설명력이 부족하여 대여 건수가 고장여부의 영향을 주는 요인으로 결론 내리기에 부족

### 3.통계 검정<mark>\_내부 요인</mark>

## Q2서울시공공자전거대여건수는자전가의고장신고건수와상관이있을까?

# 총 고장신고 건수외 총 대여건수의 상괸 관계를 확인힐 수 있다





### 3.통계 검정<mark>\_내부 요인</mark>

Q2 서울시 공공자전거 대여 건수는 자전거의 고장신고건수와 상관이 있을까?

# 피어슨 상관분석 결과:대여건수외 고장신고건수는 통계적으로 뚜렷힌 양적 상관관계

### 통계검정설계

따름이별총대여건수와고장신고건수이두수치형변수들간의 

가설설정 귀무가설

:따릉이 별총대여건수는 지전거의 고장신고건수와 상관이 없다

대립기설

:따릉이별총대여건수는자전거의고장신고건수와상관이있다

독립변인 따름이별1년내총대여건수

종속변인 따름이의 고장신고건수

### 통계검정 결과

coef: 0.4918695754701043

p value: 0.0

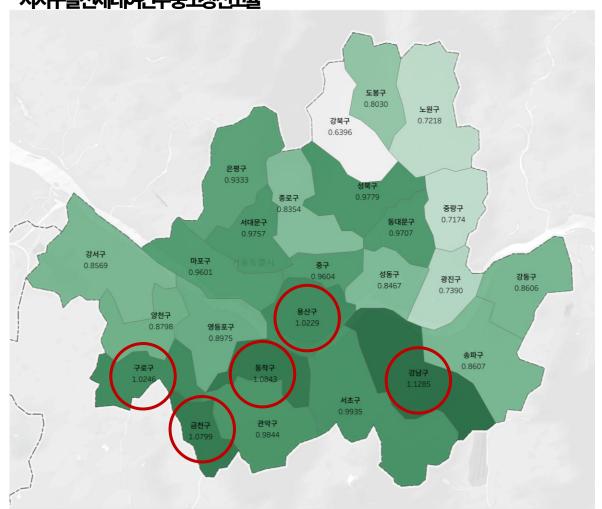
### p-value는 0으로 통계적으로 유의미

상관계수는 약 0.49186으로 두 변수는 양의 상관관계가 있습니다.

대여건수가 증가하면 고장신고건수도 증가함을 알 수 있습니다.

**Q3**이용거리와 시간은 자전거의 고장 신고율과 상관이 있을까?

# 대여건수 비율괴 무관하게 고장신고율이 높은 자치구기 존재한다



- 자치구별고장신고율은총대여건수가아닌 다른요인괴도관계가있을수도있다
- 고장신고율이높은집단과 [강남구,동작구,급천구,구로구,용산귀] 그외지역구로분리하여검증해본다

- \*고장신고율:(기준별)고장신고간수/대여간수
- \*평균이용시간&가리:
- 개별자전가임대여건당평균0용가임외평균0용사간을의미

### 3.통계 검정<mark>\_내부 요인</mark>

Q3 이용거리와 시간은 자전거의 고장 신고율과 상관이 있을까?

# T-Test분석결과: 고장신고율이 높은 자치구와 아닌 자치구 간

# 평균이용거리외이용시간에 통계적 차이기 없다.

### 통계검정설계

### 분석목적

Shapiro 정규성 검정 -> Bartlett/Levene 등분산검정 각각의 결과에 맞추어 T-Test 진행한다. 고장신고율이 높은 자 치구의평균이용량과그외자치구의평균이용량이통계적으로 유의한차이가있는지확인

### 가설설정 귀무가설

:고장신고율이높은자치구집단의개별따름이평균이용량과그 외자치구의평균이용량에 통계적차이가없다.

#### 대립기설

:고장신고율이높은자치구집단의개별따름이평균이용량과그 외자치구의 평균 이용량에 통계적 차이가 있다.

### 통계검정 결과

Shapiro : 정규성 만족

Bartlett : 등분산성 만족

T-Test 진행

P-value			
평균 이용거리	0.1029		
평균 이용시간	0.0977		

### 두 변수의 p-value가 유의수준 0.05보다 높다.

따라서. 귀무가설을 기각하지 못하며 두 변수 모두 고장 신고율로 나눈 두 자치구 그룹의 평균에 통계적인 차이가 없다.



# 결론 및 분석의 한계점\_외부 요인

## 결론\_날씨(기온)은 따름이 타이어 고장과 양의 상관관계

타이어고장은기온(평균기온,최고기온,최저기온,평균지면온도)과양의상관관계를나타낸다는사실을통해 타이어는기온이상승할수록높은고장률을보인다는것을확인할수있습니다.

이용자의 안전한 여름철 안전한 라이딩을 위해 여름, 특히 폭염 시기에는 타이어의 교체 및 관리 주기를 짧게 시행할 것을 제안 드립니다.

### 분석의 한계점

- 고장건수와 날씨데이터만 이용한 분석으로 얻을 수 있는 **인사이트의 한계**가 분명함
- 타이어교체주기 및 운영방침 미공개로 인해 기존 운영방침과 비교된 제안 시항을 도출하기에는 어려움이 있음
- 자전거방치상태및노후상태를고려하지않은분석으로, 신뢰하기어려움

# 결론 및 분석의 한계점 내부 요인

### 결론\_ 따름이 대여건수와 고장신고 건수는 양의 상관관계

따름이를자주사용하는것이고장을유발하는확률을높일수있기때문에

일정대여건수를넘긴따름이에대해서는특별정비를진행하거나

대여량이 많은 대여소의 정비 주기를 조정하는 것이 고장을 효율적으로 관리하는데 도움이 될 수 있습니다

또한대여량이 많은정비소의 따름이와 그렇지 않은 대여소의 따름이를 정기적으로 섞는 것도 방안이 될 수 있을 것입니다

### 분석의 한계점

- 고장'신고'데이터이기때문에고장인지 아닌지 구분할기준 없음
- 고장신고이후에정비진행여부 및 수리 완료여부를확인할 수 없음
- 운행이 불가한 고장난 자전거가 계속해서 대여되는 문제로, 이용거리가 정확하게 수집되지 않음
- 고장여부에서고장이 있었던 자전거의 데이터 크기가고장나지 않은 자전거에 비해 클 수 밖에 없음
- 이용거리, 이용시간, 대여건수 등분석하려는 변수의 단위가 달라서 정규화가 필요하였으나, 정규화 작업 진행이 안됨





¿ 서울시 공공자전거 이용량은 자전거의 고장 신고율과 상관이 있을까?

# 내부적 요인 관련 추가 검정

실제 따름이의 대여 기록 대비, 고장신고 양은 0.5%

실제 따름이 대여와 이용기록에 있어서, 고장난 상태보다 고장나지 않은 상태에서 운용된 기록 이 많을 수 밖에 없다. 전체의 평균값으로 계산 하면서 확인하고자 하는 '고장난 자전거'의 이 용패턴을 확인하기 어려움. 어떻게 이 한계를 극복할 수 있을까?

고장신고와 고장신고 사이의 간격이 짧은 데이터들은 고장에 대한 조치가 이루어지지 않은 상태에서 운용되었을 가능성이 높을 것이다.

그렇다면, 고장 신고 간격이 짧은 자전거에 대해서는 다른 이용패턴을 확인할 수 있지 않을까?

¿ 서울시 공공자전거 이용량은 자전거의 고장 신고율과 상관이 있을까?

# 개별따름이의 Consecutive 특성별일평균이용거리 분포

'Consecutive' : 연속 신고 앞선 고장신고와 다음 고장신고 사이의 간격이 n일 이내인 자전거

P-value Table	Consecutive 1 vs	Consecutive 1 vs	Consecutive 0 vs Normal	
기준일(n)	Consecutive 0	Normal		
31	0.0992	0.0052	0.2181	
14	0.0646	0.0089	0.1150	
7	0.0832	0.0211	0.0484	
5	0.0192	0.0067	0.1214	

연속 신고 기준이 5일인 경우

샘플링 된 데이터에서 연속 신고 집단과 정상 집단의 평균 차이가 통계적으로 유의 함과 동시에, 연속 신고 집단과 불연속 신고 집단의 평균 차이도 통계적으로 유의하게 달라진다.

Sapiro -> 정규성 불만족 Levene -> 등분산 불만족 Welch\_T-Test 진행

¿ 서울시 공공자전거 이용량은 자전거의 고장 신고율과 상관이 있을까?

# 개별따름이의 Consecutive 특성별일평균이용거리 분포

'Consecutive' : 연속 신고 앞선 고장신고와 다음 고장신고 사이의 간격이 n일 이내인 자전거

P-value Table	Consecutive 1 vs	Consecutive 1 vs	Consecutive 0 vs Normal	
기준일(n)	Consecutive 0	Normal		
31	0.0992	0.0052	0.2181	
14	0.0646	0.0089	0.1150	
7	0.0832	0.0211	0.0484	
5	0.0192	0.0067	0.1214	

연속 신고 기준이 5일인 경우

샘플링 된 데이터에서 연속 신고 집단과 정상 집단의 평균 차이가 통계적으로 유의 함과 동시에, 연속 신고 집단과 불연속 신고 집단의 평균 차이도 통계적으로 유의하게 달라진다.

Sapiro -> 정규성 불만족 Levene -> 등분산 불만족 Welch\_T-Test 진행

¿ 서울시 공공자전거 이용량은 자전거의 고장 신고율과 상관이 있을까?

# 개별따름이의 Consecutive 특성별일평균이용거리 분포

기준일(5일) 단위(m)	평균	중앙값(50%)	최대값(max)	표준편차
Consecutive 1	14578	10869	73336	14722
Consecutive 0	10612	9655	33809	6463
Normal	2421	1742	32445	2353

연속으로 신고된 자전거의 평균 일평균이용거리가 14578m로 가장 길고, 최대값 역시 가장 긴 73336m를 기록했다.

고장 신고 간격이 짧았다는 것이 고장난 자전거 자체를 의미하는 것은 아니므로, 고장 신고 간격이 짧은 자전거들의 이용 량이 평균적으로 많다.

고장난 자전거로 많은 이용을 했다는 해석보다는, 이용량이 많은 자전거가 고장이 난다. 더불어, 고장난 자전거에 대해서 적절한 조치가 이루어지고 있 다. 따라서 고장신고율이 높은 집단이 통계적으로 유의미한 결과를 보이고 있지 않았다고 짐작된다.

¿ 서울시 공공자전거 이용량은 자전거의 고장 신고율과 상관이 있을까?

# 개별따름이의 Consecutive 특성별일평균이용거리 분포

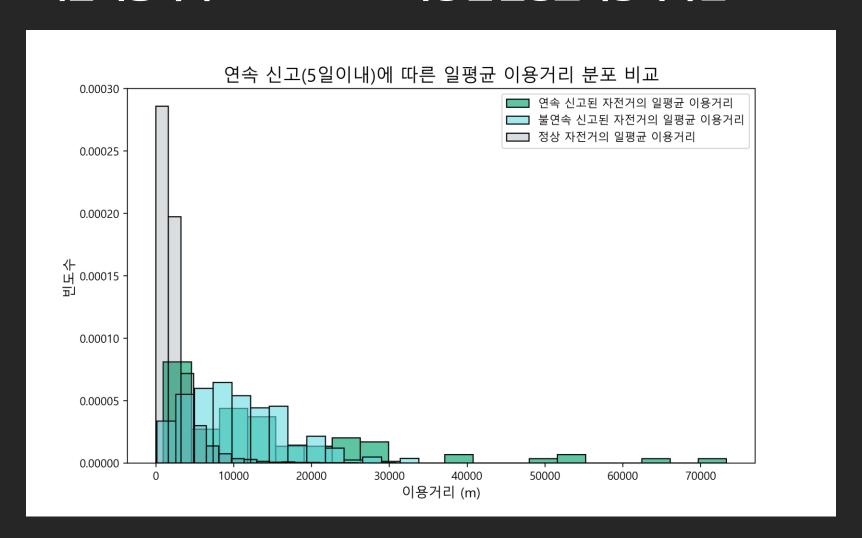
기준일(5일) 단위(m)	평균	중앙값(50%)	최대값(max)	표준편차
Consecutive 1	14578	10869	73336	14722
Consecutive 0	10612	9655	33809	6463
Normal	2421	1742	32445	2353

연속으로 신고된 자전거의 평균 일평균이용거리가 14578m로 가장 길고, 최대값 역시 가장 긴 73336m를 기록했다.

고장 신고 간격이 짧았다는 것이 고장난 자전거 자체를 의미하는 것은 아니므로, 고장 신고 간격이 짧은 자전거들의 이용 량이 평균적으로 많다는 것이다.

¿ 서울시 공공자전거 이용량은 자전거의 고장 신고율과 상관이 있을까?

# 개별따름이의 Consecutive 특성별일평균이용거리분포



개별 자전거를 기준으로 연속 신고 집단, 불연속 신고된 집단, 정상 집단에서 각각 100개 의 자전거를 샘플링.

개별 자전거의 한 해의(2022) 일 평균 이용거리의 분포를 나타내 었다.

모든 데이터가 Right Skewed하 게 나타났으며, Consecutive 속 성을 가지는 집단의 꼬리가 가장 길게 뻗어 있음을 확인할 수 있 다.