

안전하고 편리하게, 따릉이 이용 최적화



<빅데이터 처리와 시각화>
- 재무금융전공 20202817 신수옥
- AI빅데이터융합경영학과 20212544 김승혁
- AI빅데이터융합경영학과 20212548 김지은

contents

프로젝트 배경

데이터 전처리
및 시각화

앱 개발 소개

최적화 방안
및 기대효과

프로젝트 배경



따릉이 대여건수 매년 증가

이에 따른
'따릉이의 배치의 효율성 증대,
따릉이 사고 감소에 일조'

데이터 소개

'따릉이'의 검색 결과 16건을 찾았습니다. 정확도순 조회



[교통]

서울시 따릉이 대여소별 대여/반납 승객수 정보

서울시 1일 1시간 단위로, 따릉이 대여소별 출발지 도착지 승객수를 확인할 수 있습니다. 따릉이 대...
수정일자: 2022-06-12 제공기관: 서울특별시 제공부서: 도시교통실 보행친화기획관 교통정...

FILE

공공 데이터



[교통]

서울시 따릉이대여소 마스터 정보

서울시 따릉이대여소 마스터 정보입니다. 서울시 따릉이대여소에 대한 대여소 ID, 역 주소, 좌표 정...
수정일자: 2022-04-13 제공기관: 서울특별시 제공부서: 도시교통실 보행친화기획관 교통정...

SHEET OpenAPI

공공 데이터



[교통]

서울시 공공자전거 대여소 정보

서울특별시 공공자전거 대여소(따릉이) 현황정보입니다. 대여소의 이름, 관리번호, 위치정보와, 거치대...
수정일자: 2022-04-13 제공기관: 서울특별시 제공부서: 도시교통실 보행친화기획관 자전거...

FILE

공공 데이터

1. 서울시 따릉이 대여소별 합계
2. 서울시 따릉이 시간대별 대여 합계
3. 서울시 따릉이 사고발생 합계

데이터 전처리

메인 데이터 : 시간대별 데이터

대여일자	대여시간	대여소번호	대여소명	대여구분	성별	연령대	코드	이용건수	운동량	탄소량	이동거리	사용시간
2021-01-01	0	1952	1952. 천왕	정기권	F	~10대		1	22.15	0.2	860.6	6
2021-01-01	0	828	828. 숙대	정기권	F	20대		1	0	0	0	41
2021-01-01	0	1047	1047. 건강	정기권	F	20대		1	27.13	0.32	1370	12
2021-01-01	0	112	112. 극동	정기권	F	20대		1	0	0	0	70
2021-01-01	0	2620	2620. 송파	정기권	F	20대		1	0	0	0	20
2021-01-01	0	3411	3411.종묘	정기권	F	20대		1	27.18	0.24	1055.93	8

A. 대여일자 -> string

B. 대여소명 코드 중복

C. 대여시간 / 사용시간 string

D. 데이터 제공형태 : 월별 데이터

데이터 전처리

```
def absd(df):  
    df = pd.read_csv(df)  
    df = df.query("1300 <= 대여소번호 < 1400 or 4403 <= 대여소번호 <= 4415")  
    df['대여일자'] = pd.to_datetime(df['대여일자'], format="%Y %m %d")  
    df['year'] = df['대여일자'].dt.year  
    df['month'] = df['대여일자'].dt.month  
    df['day'] = df['대여일자'].dt.day  
    df['성별'].replace('m', 'M', inplace = True)  
    df['성별'].replace('f', 'F', inplace = True)
```

```
def(month, latlon):  
    month = pd.read_csv(month)  
    time19 = pd.concat([time19, time20], axis = 0)  
    time19 = pd.concat([time19, time21], axis = 0)  
    latlon = pd.read_csv(latlon, encoding = 'cp949')
```

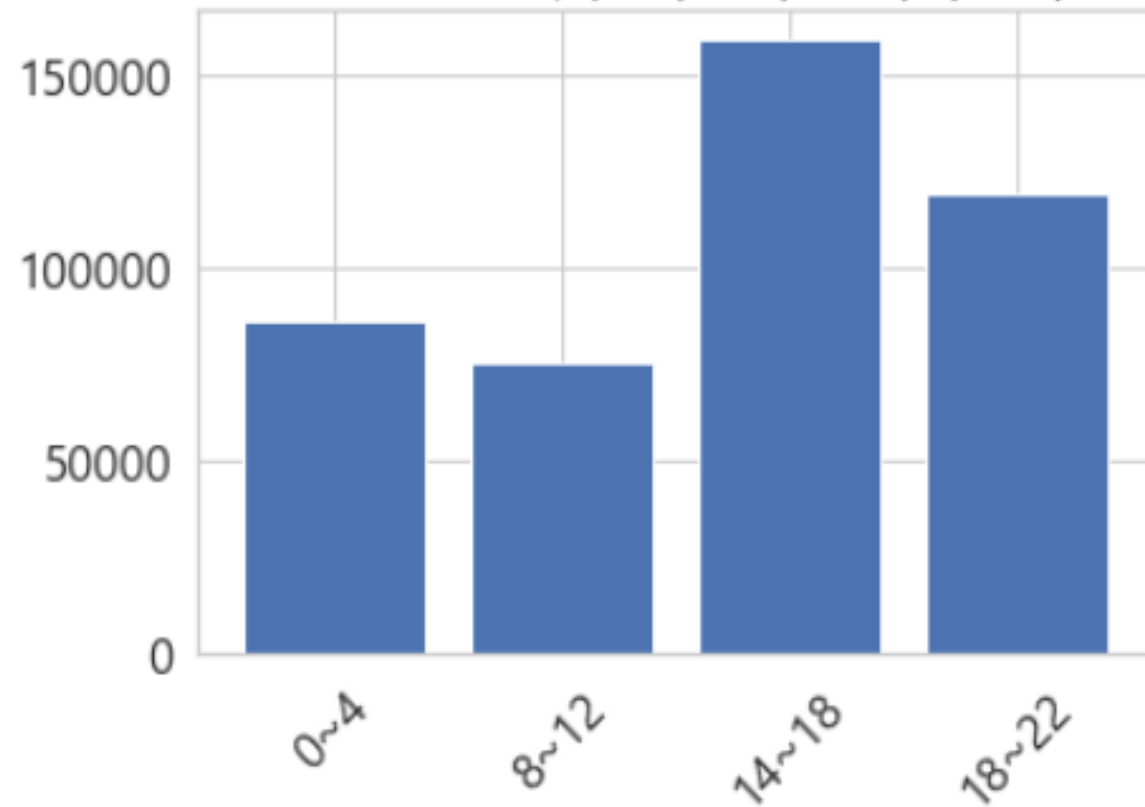
각 데이터 전처리
: 매번 코드를 다시 설정해야 함



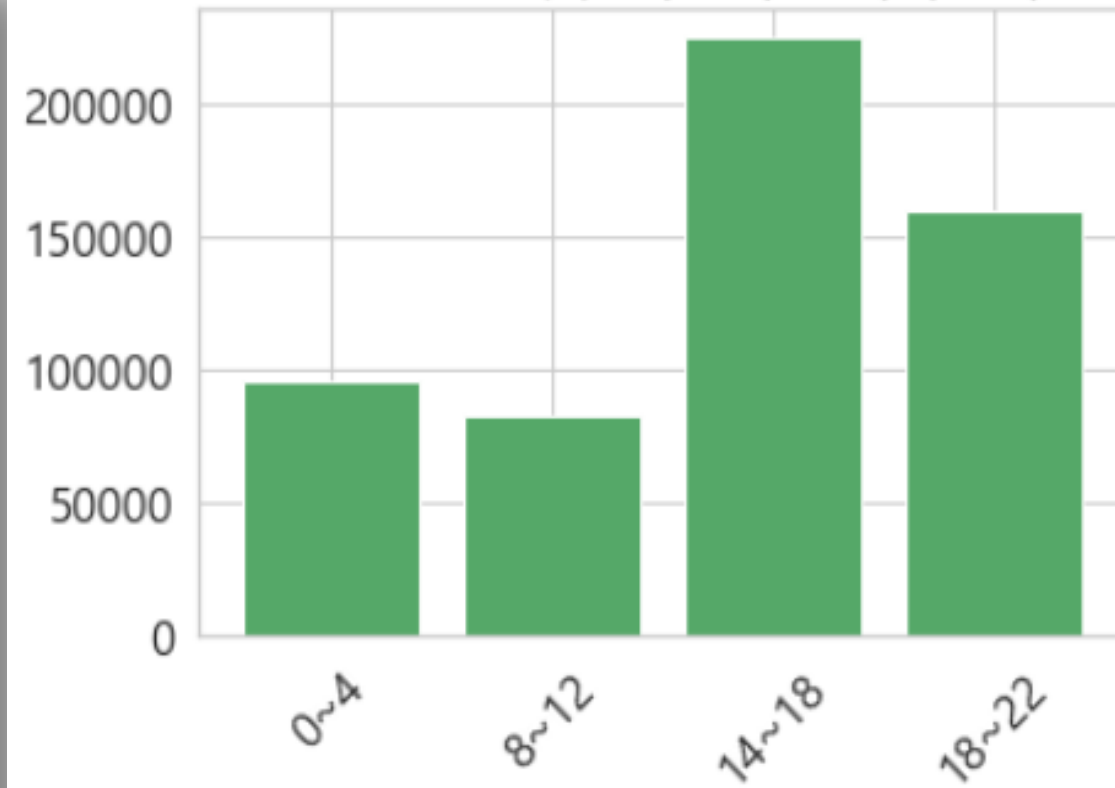
함수 작성으로 데이터 전처리의 자동화

데이터 전처리 및 시각화

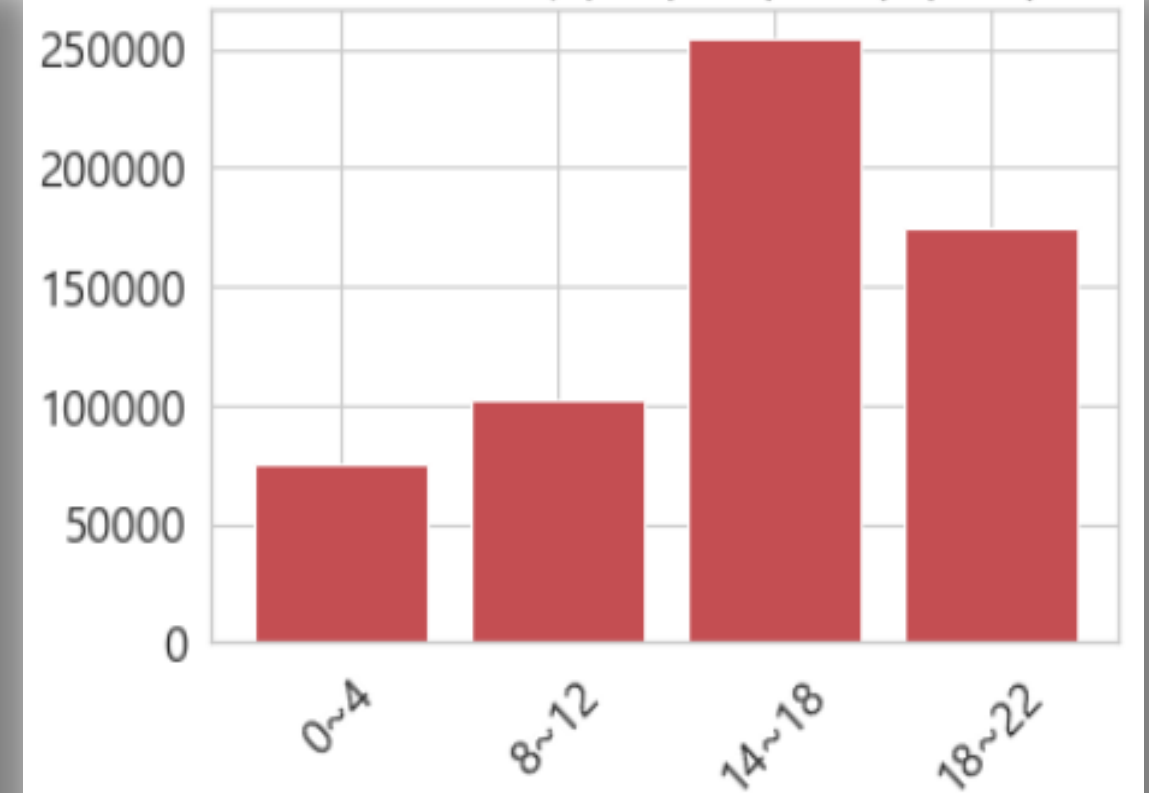
2019년 성북구 시간대별 대여합계



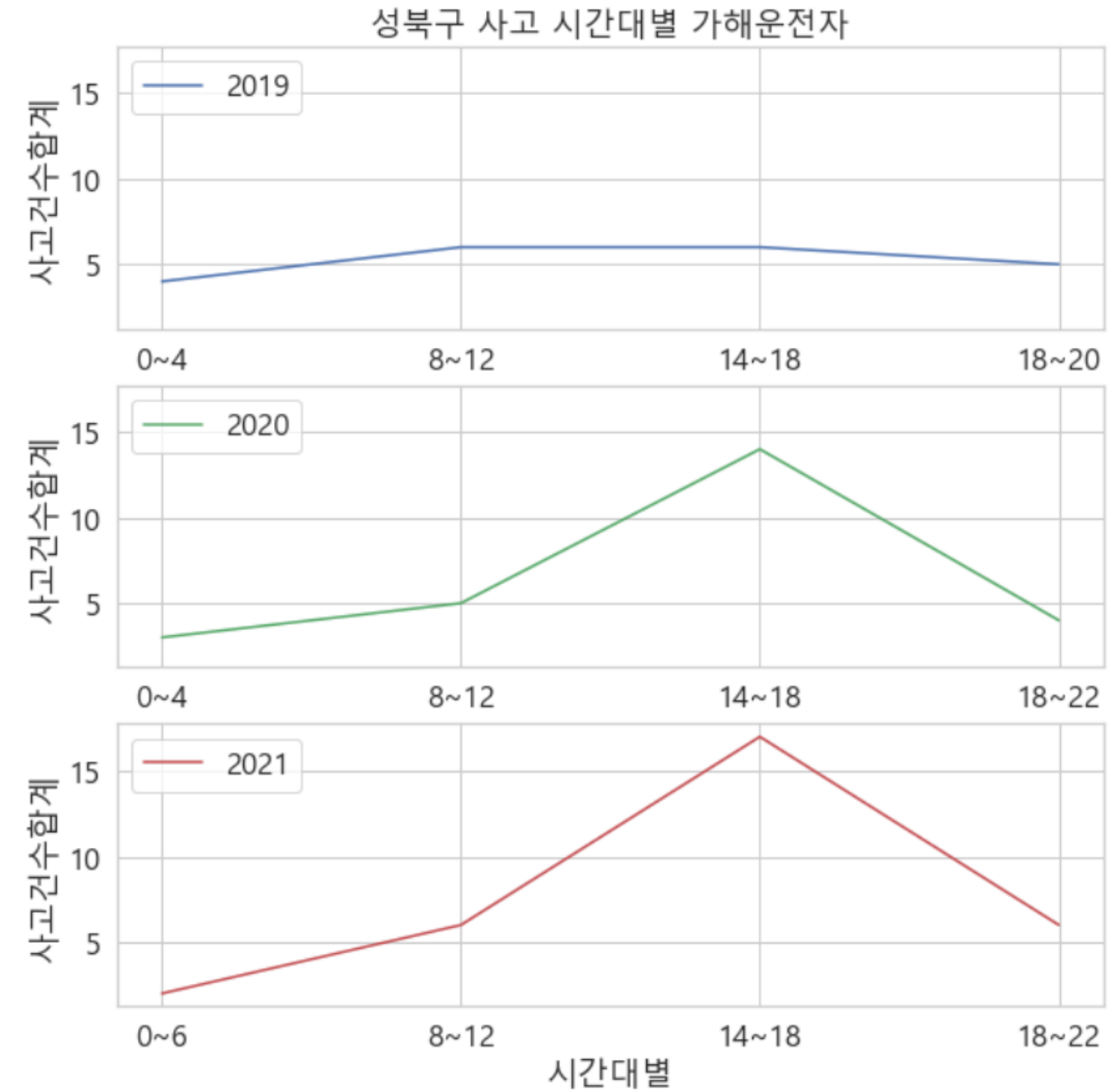
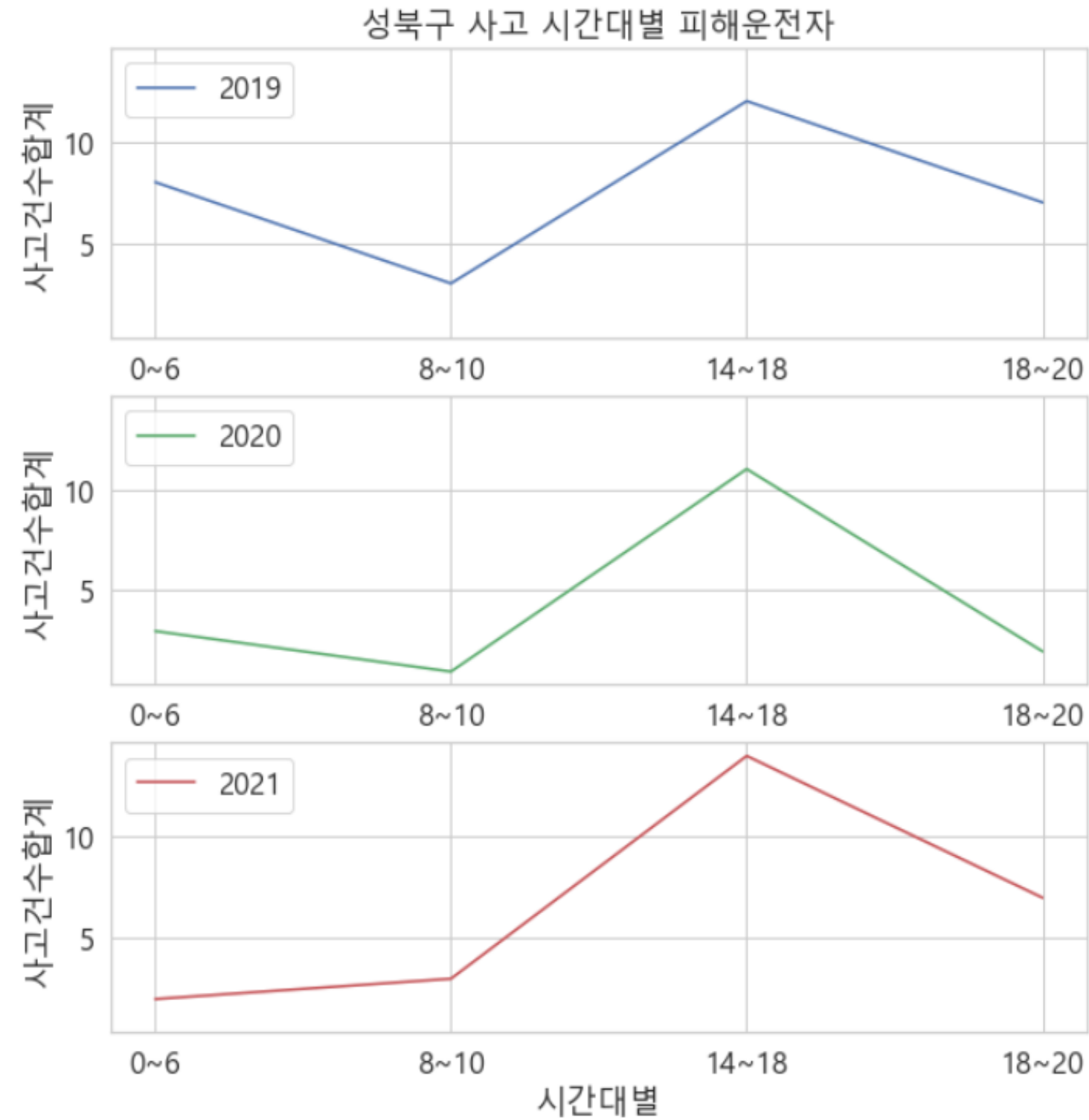
2020년 성북구 시간대별 대여합계



2021년 성북구 시간대별 대여합계



데이터 전처리 및 시각화



Streamlit data app

성북구 따릉이 Dashboard

새로고침

0. Overview

평균이동거리

4.041

↑ 0.0

최다 이동거리

13204.0

평균사용시간

28.684

↑ 0.0

최다 사용시간

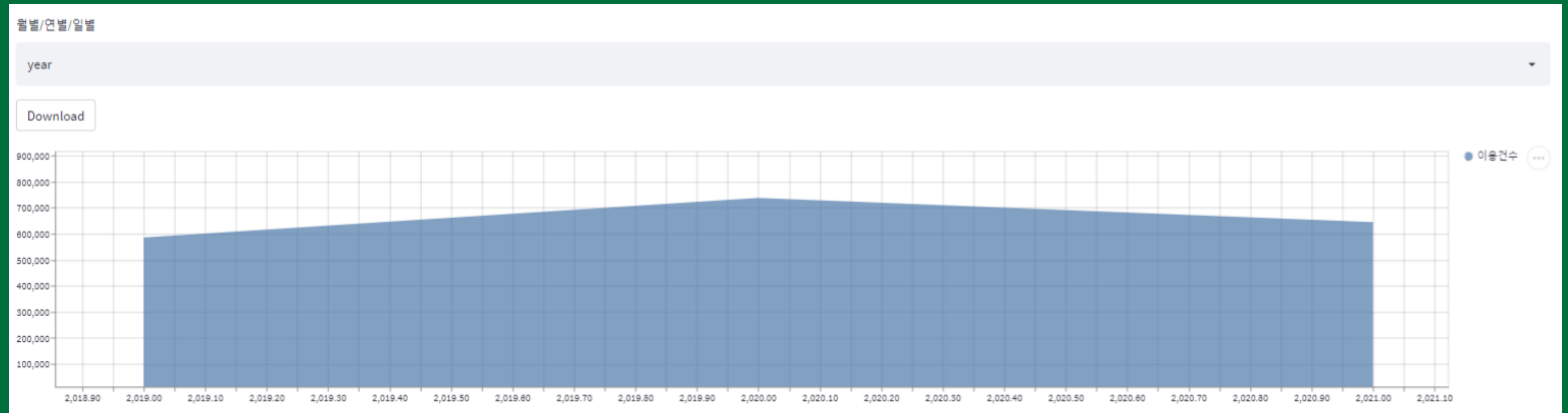
7727

19



Streamlit data app

year



Streamlit data app

month

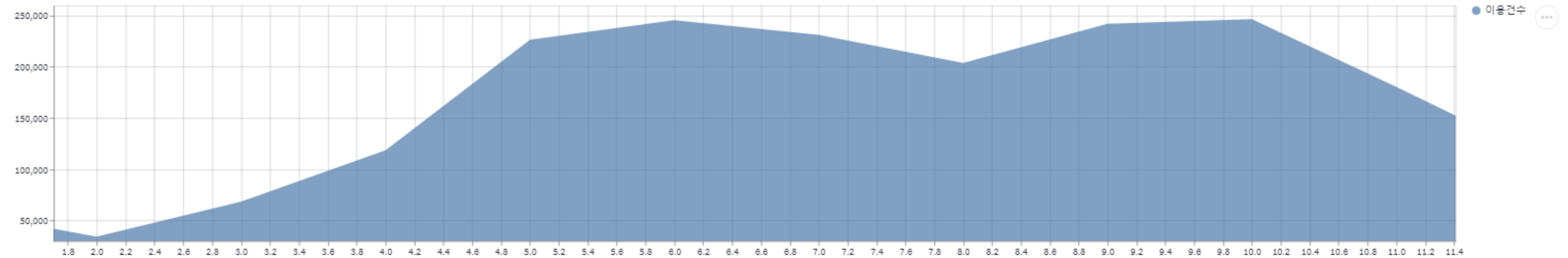
1. 이용건수 분석

전체

월별/연별/일별

month

Download

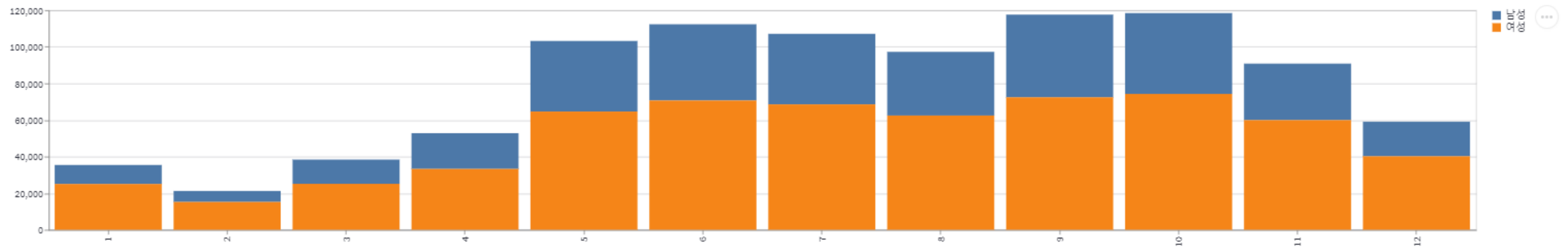


Streamlit data app

성별 대여건수

성별 현황

성별 사용건수

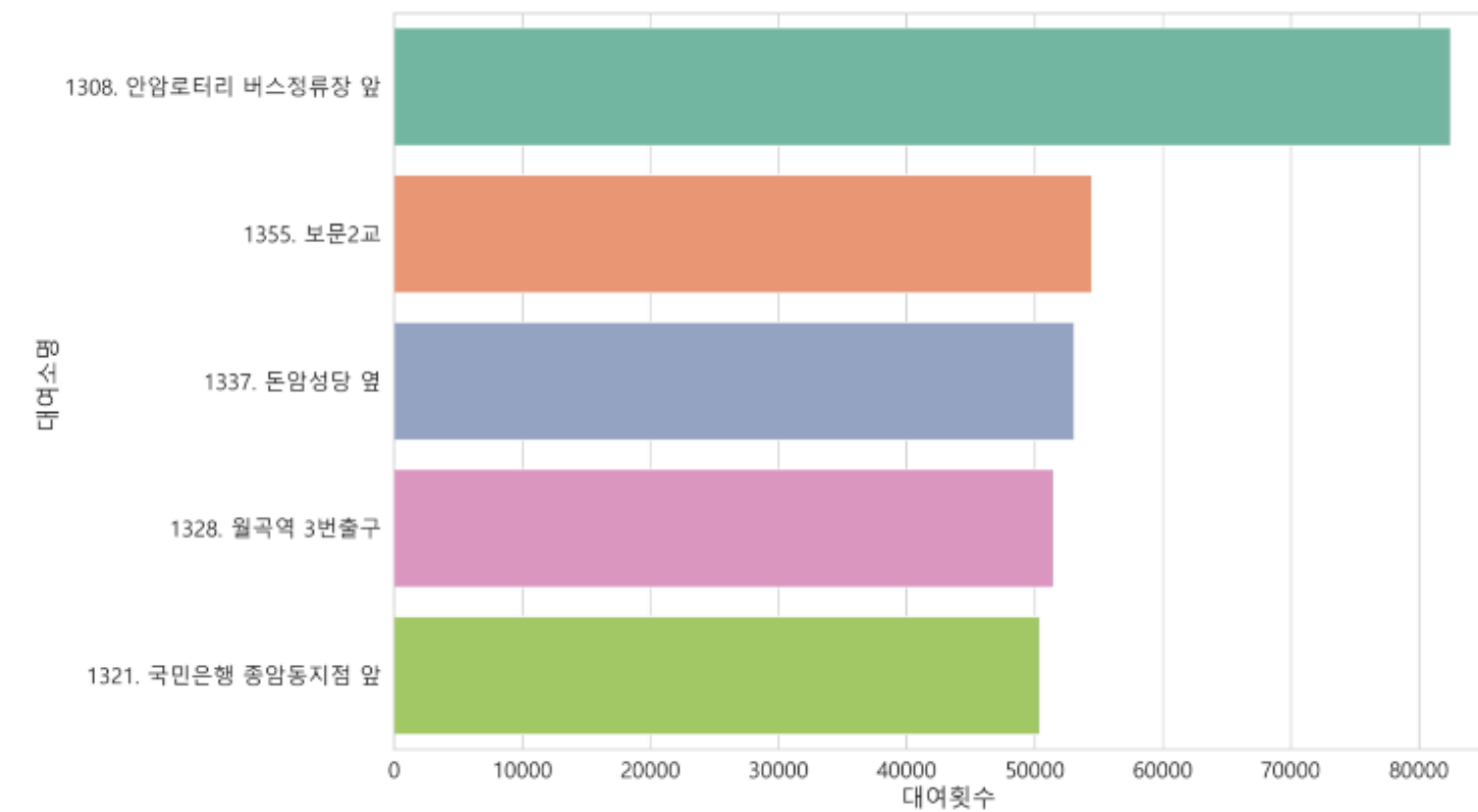


Streamlit data app

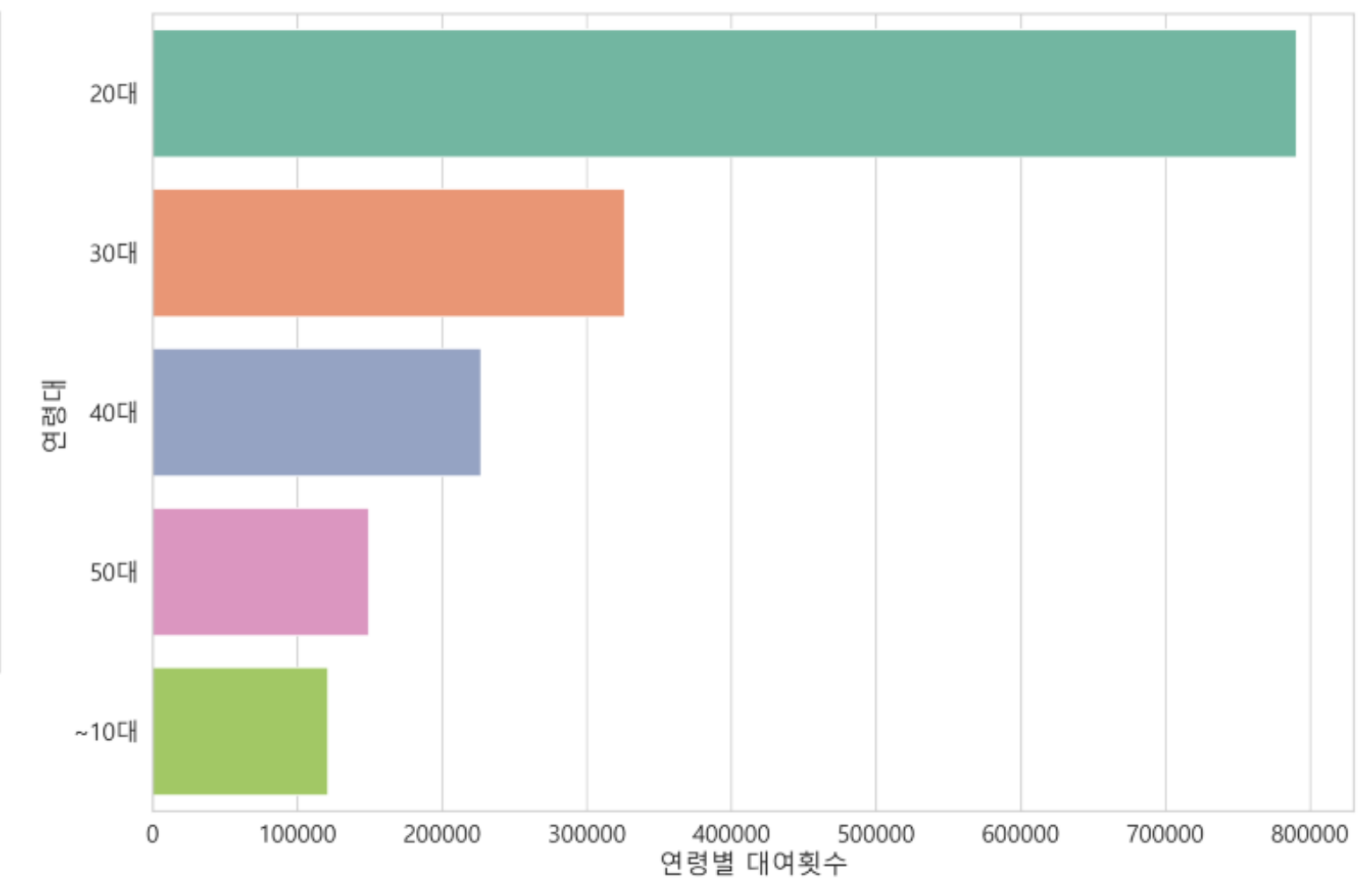
Top5 대여소

2. Top5 비교

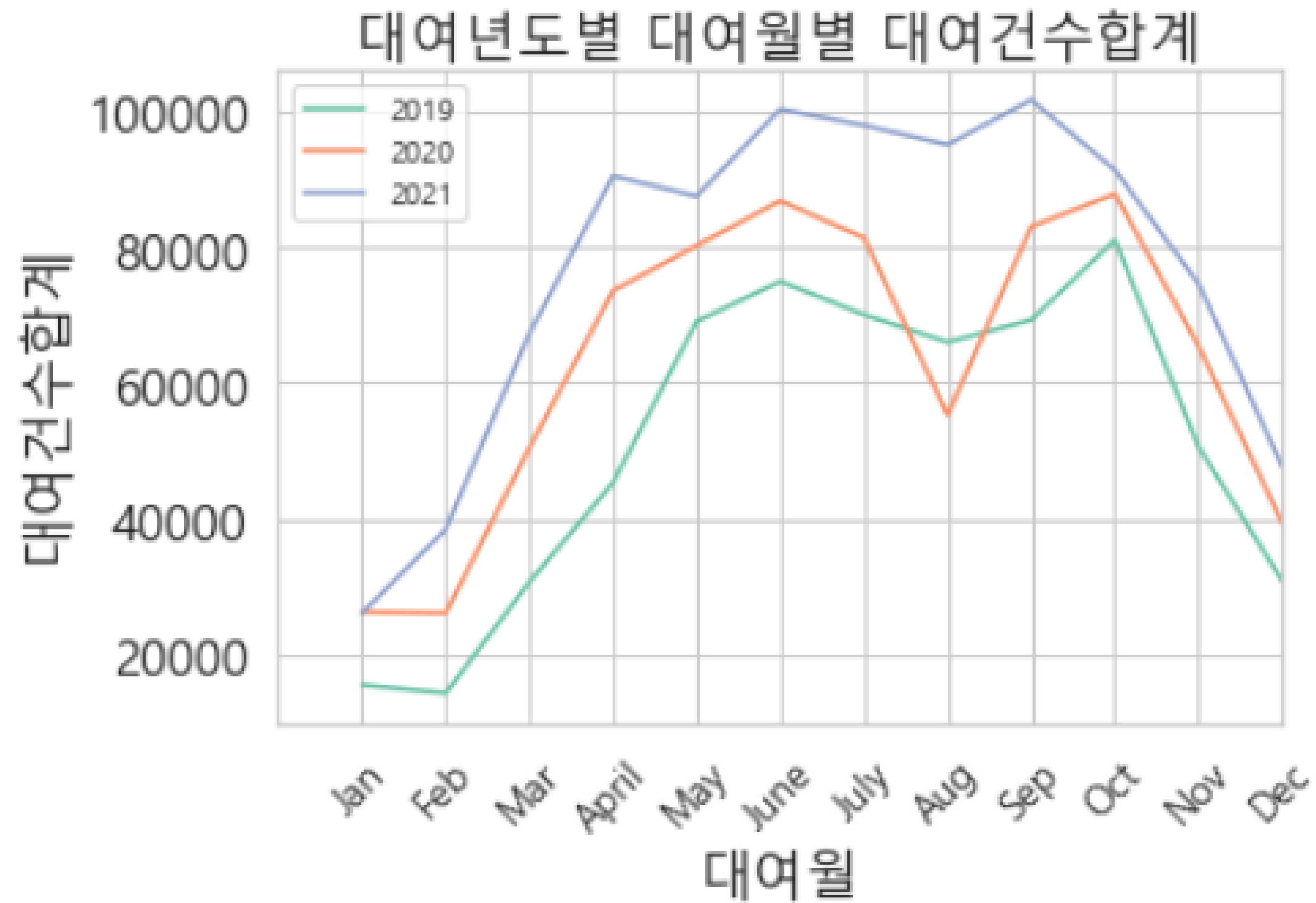
Top5 대여소



Top5 연령대



Streamlit data app



Insight 도출

**A. 퇴근시간보다
오후 시간이 많은 이유**

- 1. 주말 데이터 나들이**
- 2. 퇴근 시간 데이터의 포함**

Insight 도출

**B. 새벽시간이 출근시간
보다 많은 이유**

**막차가 끊긴 후 퇴근 용도로
사용하는 경우**

Insight 도출

C. 사고 시간대별 데이터

사고 피해 운전자와
가해 운전자는 거의 동일
-> 자전거 끼리의 사고

출근시간에
총 6건의 추가 피해사고 발견

Insight 도출

D. 평균이동거리 <<< 평균 사용시간

1. 평균 사용시간 다수의 이상치 발견
2. 최다 사용시간 실제 문의

-> 실제로 해당 데이터 시점
기준으로 미반납

Insight 도출

E. 5월 / 9월 / 10월 최다 대여

1. 날씨가 온후한 봄 / 가을에 증가
2. 방학 시즌인 봄에는 더욱 증가

Insight 도출

F. 여성 데이터의 우위

1. 자가용 소유 여부
2. 여성 중 20대 이하 데이터 다수 발견

Insight 도출

**G. 2019 / 2020 데이터에 비해
2021년 대여 데이터 감소**

-> 코로나19 상황

Main Insight A

대여소 1~5위
데이터 확인 및 문의

자전거 보충 필요 없음

Main Insight B

연령대 20 -> 30 -> 40 -> 50 대 순서

젊은 층

대여량 압도적 다수 :
2030세대 홍보 효과

한계점

1. 주말 데이터 분류의 부재
2. 위경도별 데이터 의미 부재
3. Insight의 간단함



Thank you

