Exploratory Data Analysis of the Seoul Bike

2020310243
Data Science Convergence
Bohyun Kim

CONTENTS

Introduction

Data

1. About datasets
2. Data manipulation

EDA

1. Overall trend in using Seoul Bike
2. Hypothesis1: correlation between COVID-19 and bike rental degree
3. Hypothesis2: where most people rented bikes

Conclusions

Introduction

Public system to lend bikes, **Seoul Bike**

important transportation for Seoul citizens

Several problems

How to develop and maintain our public system

Data

1. About datasets

2	강남구	2302. 교보타워 버스정류장(신논현역 3번출구 후면)	201812	500
3	강남구	2303. 논현역 7번출구	201812	286
4	강남구	2304. 신영 ROYAL PALACE 앞	201812	149
5	강남구	2305. MCM 본사 직영점 앞	201812	145
6	강남구	2306. 압구정역 2번 출구 옆	201812	457
7	강남구	2307. 압구정 한양 3차 아파트	201812	279
8	강남구	2308. 압구정파출소 앞	201812	292
9	강남구	2309. 청담역(우리들병원 앞)	201812	152
10	강남구	2310. 청담동 맥도날드 옆(위치)	201812	214
11	강남구	2311. 학동로 래미안 아파트 앞	201812	154
12	강남구	2312. 청담역 13번 출구 앞	201812	151
13	강남구	2313. 금원빌딩 앞	201812	268
14	강남구	2314. 청담나들목입구	201812	187
15	강남구	2315. 봉은사역 5번출구 옆	201812	447
16	강남구	2317. 삼성도로공원	201812	130
17	강남구	2318. 삼성중앙역4번출구(문화센터더 리빌)	201812	195
18	강남구	2319. 포스코4거리 서측(수협 삼성동 지점)	201812	218
19	강남구	2320. 도곡역 대치지구대 방향	201812	212
20	강남구	2321. 학여울역 사거리	201812	256

Seoul Bike data

- from Seoul open data square
- January 2017 to January 2021

	일자	계(명)	국내발생(명)	해외유입(명)
2	누적(명)	138,311	129,336	8,975
3	2020-01-20	1	0	1
4	2020-01-21	0	0	0
5	2020-01-22	0	0	0
6	2020-01-23	0	0	0
7	2020-01-24	1	0	1
8	2020-01-25	0	0	0
9	2020-01-26	1	0	1
10	2020-01-27	1	0	1
11	2020-01-28	0	0	0
12	2020-01-29	0	0	0
13	2020-01-30	3	1	2
14	2020-01-31	4	3	1
15	2020-02-01	1	0	1
16	2020-02-02	3	1	2
17	2020-02-03	0	0	0

COVID-19 data

- from the Korean ministry of Health and welfare

Data

2. Data manipulation

```
#load bike dataset
bike_201701_201806 = pd.read_csv('공공자전거 대여소별 이용정보_201701_201806.csv', encoding='cp949',
                                 names=['date', 'spot_num', 'spot', 'rental'],  header=None)
bike_201807_201811 = pd.read_csv("공공자전거 대여소별 이용정보_201807_11.csv", encoding='cp949',
                                 names=['date', 'spot_num', 'spot', 'rental'],  header=None)
bike_201812_202101 = pd.read_csv('공공자전거 대여소별 이용정보_201812_202101.csv', encoding='cp949',
                                 names=['group', 'spot', 'date', 'rental'],  header=None)
```

load csv files

```
#NaN이 있는 행 삭제
bike_201812_202101 = bike_201812_202101.dropna()
```

remove NaN values

make a new data frame

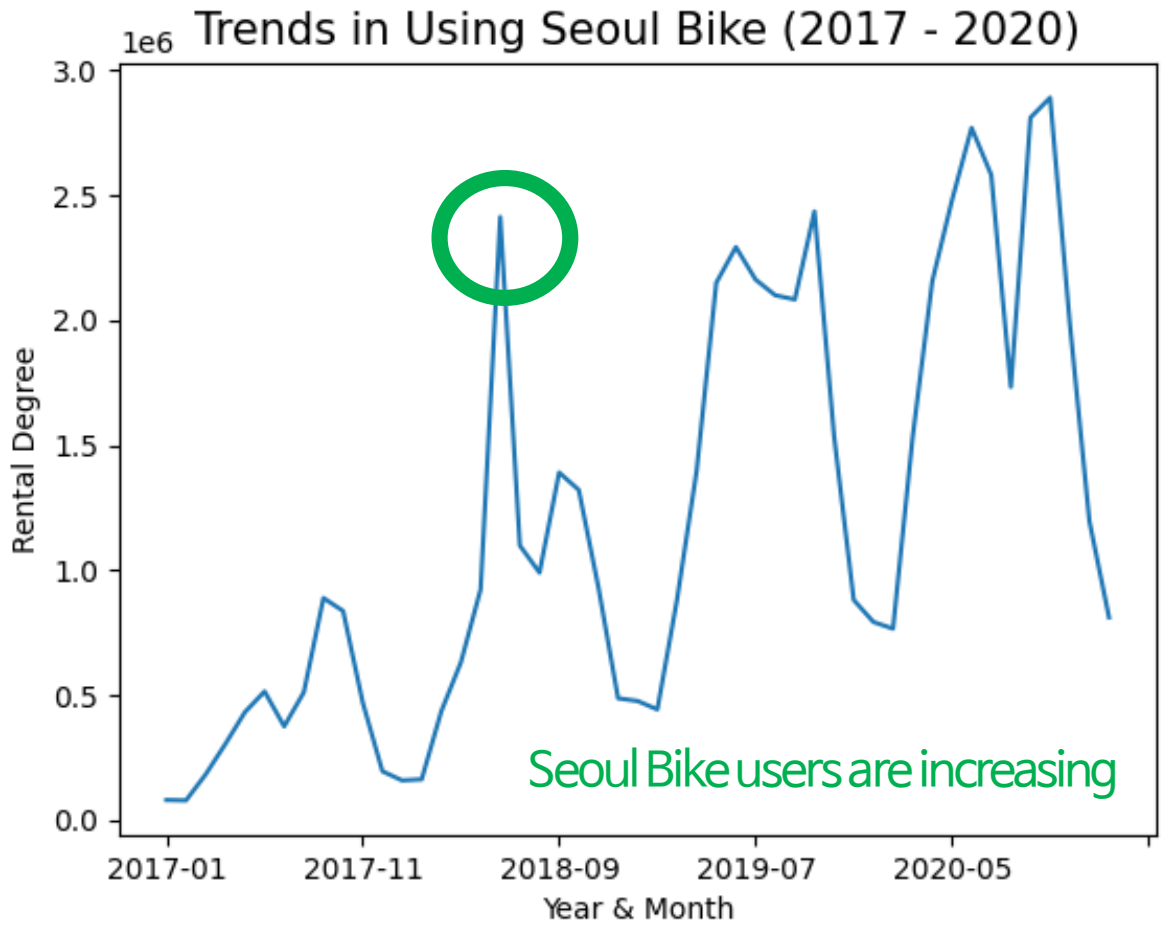
```
#df 만들기
total_rental = pd.concat([bike_201701_201806['rental'], bike_201807_201811['rental'],
                          bike_201812_202101['rental']], axis = 0, ignore_index=True)
date = pd.concat([bike_201701_201806['date'], bike_201807_201811['date'],
                  bike_201812_202101['date']], axis = 0, ignore_index=True)
bike = pd.concat([date, total_rental], axis = 1)

#date로 groupby
bike_bydate = bike.groupby('date').sum()
```

manipulate data

EDA (Exploratory Data Analysis)

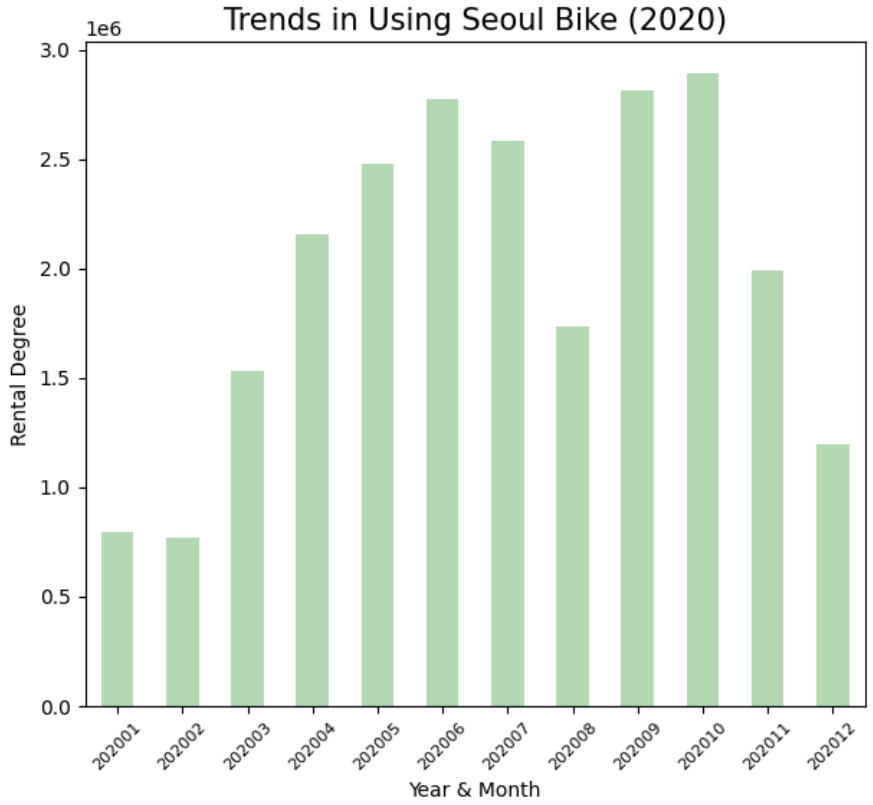
1. Overall trend in using Seoul Bike



EDA (Exploratory Data Analysis)

2. Hypothesis1 : correlation between COVID-19 and bike rental degree

- Hypothesis1: There is a specific correlation between COVID-19 and Seoul bike rental degree.



EDA (Exploratory Data Analysis)

2. Hypothesis1 : correlation between COVID-19 and bike rental degree

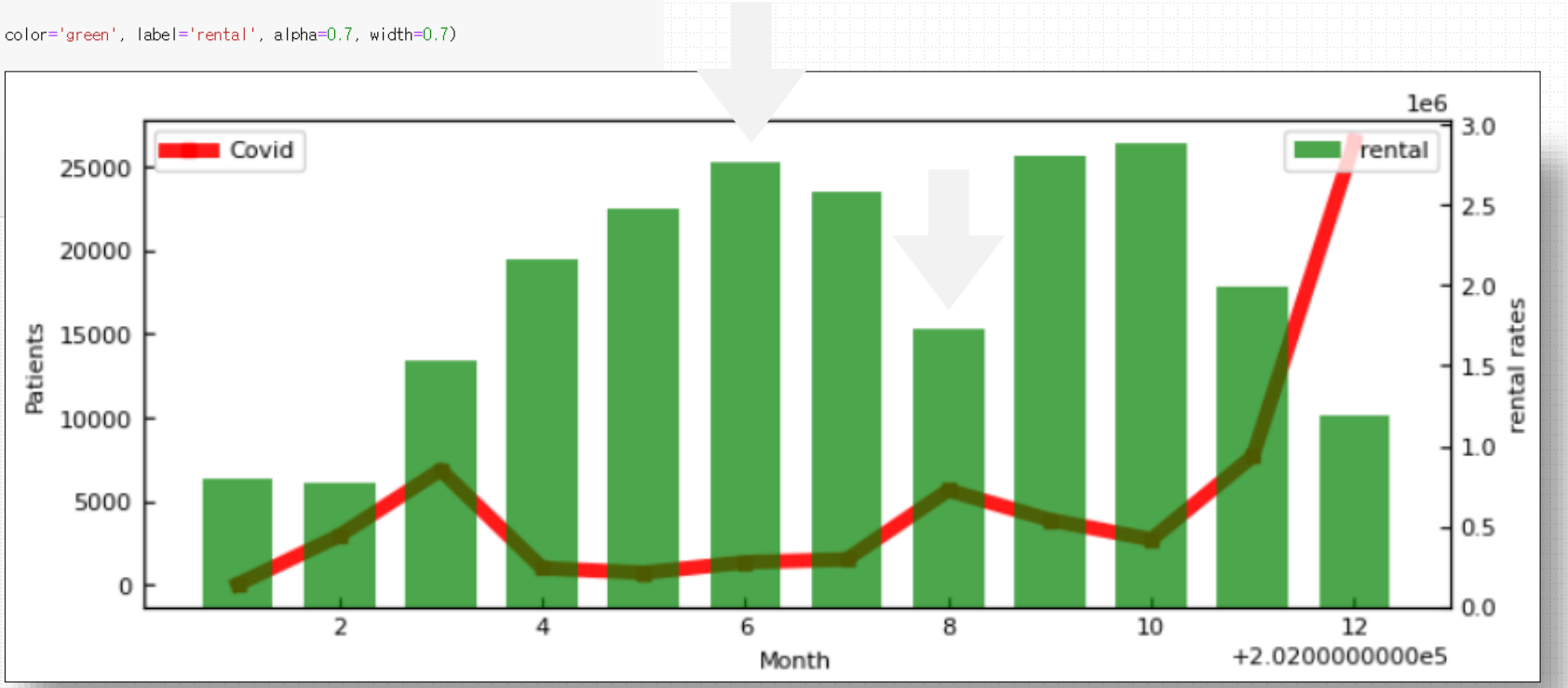
In [521]:
```
fig, ax1 = plt.subplots()

ax1.plot(covid_by2020['일자'], covid_by2020['계(명)'], '-s', color='red', markersize=4, linewidth=5, alpha=0.9, label='Covid')
ax1.set_ylim()
ax1.set_xlabel('Month')
ax1.set_ylabel('Patients')
ax1.tick_params(axis='both', direction='in')

ax2 = ax1.twinx()
ax2.bar(bike_by2020['date'],bike_by2020['rental'], color='green', label='rental', alpha=0.7, width=0.7)
ax2.set_ylim()
ax2.set_ylabel('rental rates')
ax2.tick_params(axis='y', direction='in')

ax1.legend(loc='upper left')
ax2.legend(loc='upper right')

plt.show()
```

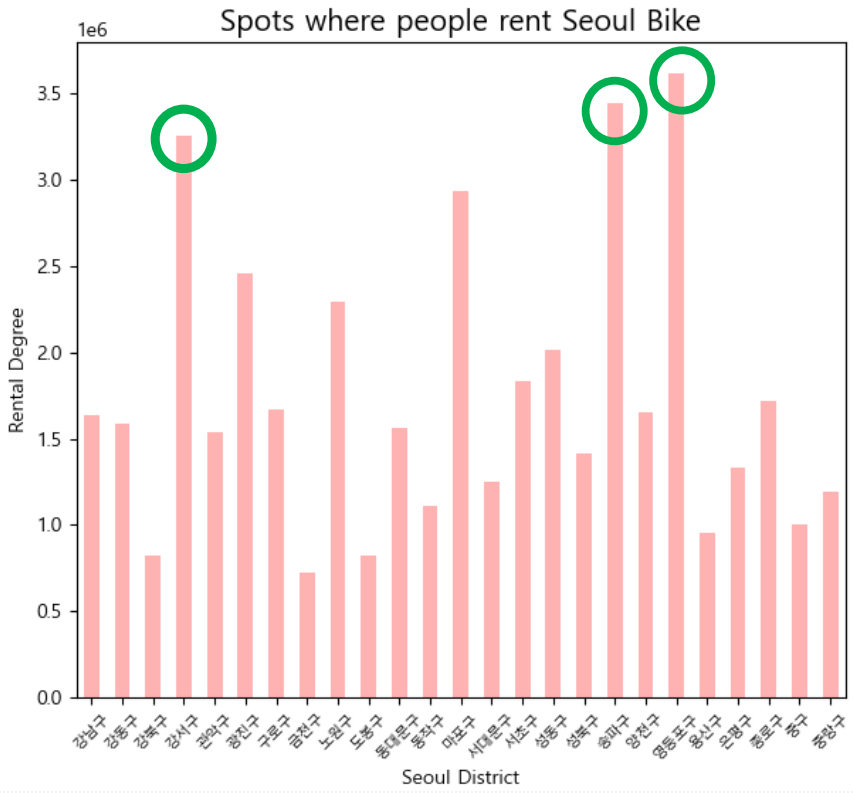


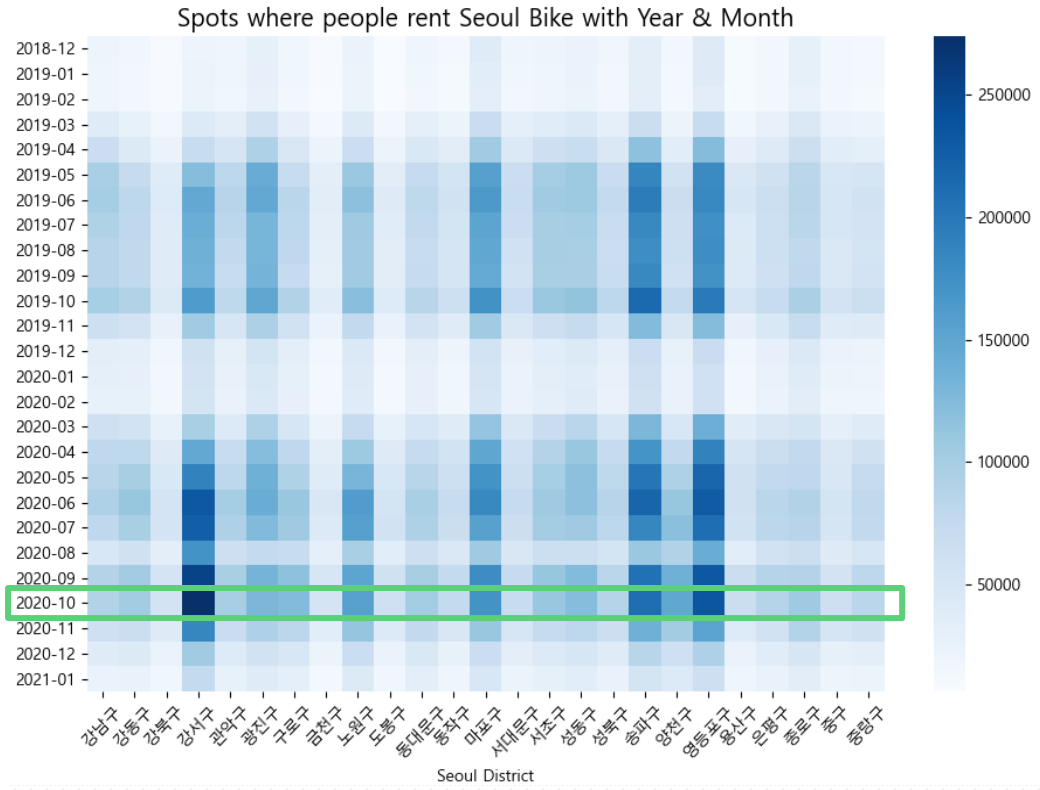
COVID-19 is one of the factors that related to Seoul Bike rental degree

EDA (Exploratory Data Analysis)

3. Hypothesis2: where most people rented bikes

- Hypothesis2: Bike rental degree will be the highest at spots near Han River Park.





EDA (Exploratory Data Analysis)

3. Hypothesis2: where most people rented bikes

In [365]:
```
import folium

geo_content = 'https://raw.githubusercontent.com/southkorea/seoul-maps/master/kostat/2013/json/seoul_municipalities_geo_simple.json'
```

In [366]:
```
map = folium.Map(location = [37.566345,126.977893], tiles = 'Stamen Terrain', zoom_start = 10)
```

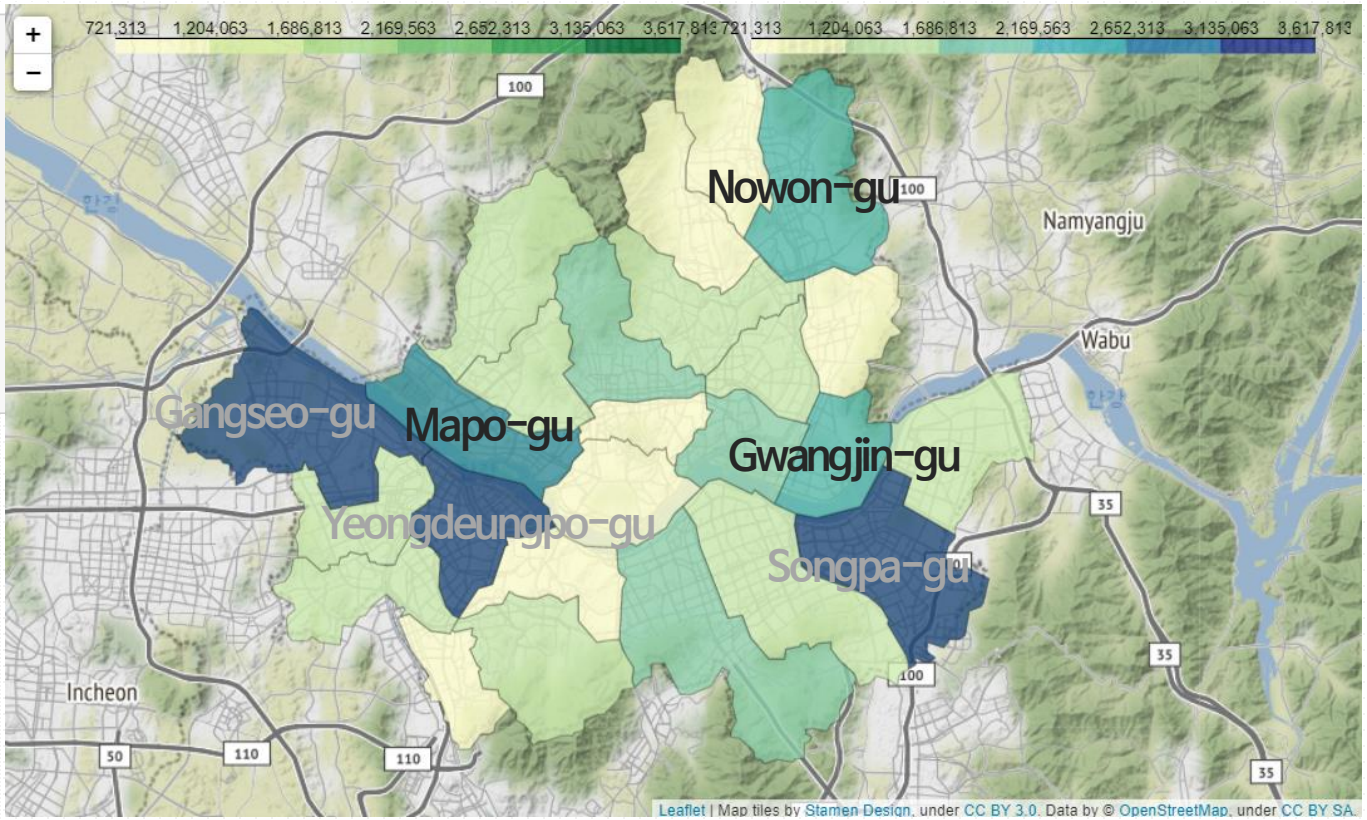
In [369]:
```
folium.Choropleth(
    geo_data = geo_content,
    name = 'chorpleth',
    data = newbike_bygroup,
    columns = ['group', 'rental'],
    key_on = 'feature.properties.name',
    fill_color = 'YlGnBu',
    fill_opacity = 0.5,
    line_opacity = 0.2,
).add_to(map)

map
```

Gangseo-gu
- Gangseo Han River Park

Songpa-gu
- Jamsil Han River Park

Yeongdeungpo-gu
- Yeouido Han River Park



EDA (Exploratory Data Analysis)

3. Hypothesis2: where most people rented bikes

```
#About top3 (강서, 송파, 영등포구)
gangseo = newbike[newbike.group == '강서구']
songpa = newbike[newbike.group == '송파구']
yeongdeungpo = newbike[newbike.group == '영등포구']

gangseo_by = gangseo.groupby('spot').sum()
songpa_by = songpa.groupby('spot').sum()
yeongdeungpo_by = yeongdeungpo.groupby('spot').sum()

print("================================")
print('<The most rented place in Gangseo-gu>', gangseo_by[gangseo_by['rental'] == max(gangseo_by['rental'])])
print('<The most rented place in Songpa-gu>', songpa_by[songpa_by['rental'] == max(songpa_by['rental'])])
print('<The most rented place in Yeongdeungpo-gu>', yeongdeungpo_by[yeongdeungpo_by['rental'] == max(yeongdeungpo_by['rental'])])
print("================================")
```

```
================================
<The most rented place in Gangseo-gu>                        rental
spot
2701. 마곡나루역 5번출구 뒤편  119753.0
<The most rented place in Songpa-gu>                           rental
spot
1210. 롯데월드타워(잠실역2번출구 쪽)  140370.0
<The most rented place in Yeongdeungpo-gu>                     rental
spot
207. 여의나루역 1번출구 앞  237293.0
================================
```

The city should put more bikes and rental spots especially in these places

Conclusions

1. The overall trend is increasing rental degree.

2. There is a correlation between COVID-19 and bike usage.

3. Gangseo-gu, Songpa-gu, Yeongdeungpo-gu near Han River Park are Seoul districts most citizens rented bikes

Thank you

By. Bohyun Kim