

# 국내 코로나19 상황이 공공자전거 사용에 미친 영향 분석

(코로나19 발생 현황 및 서울시 공공자전거 이용 데이터 기반)

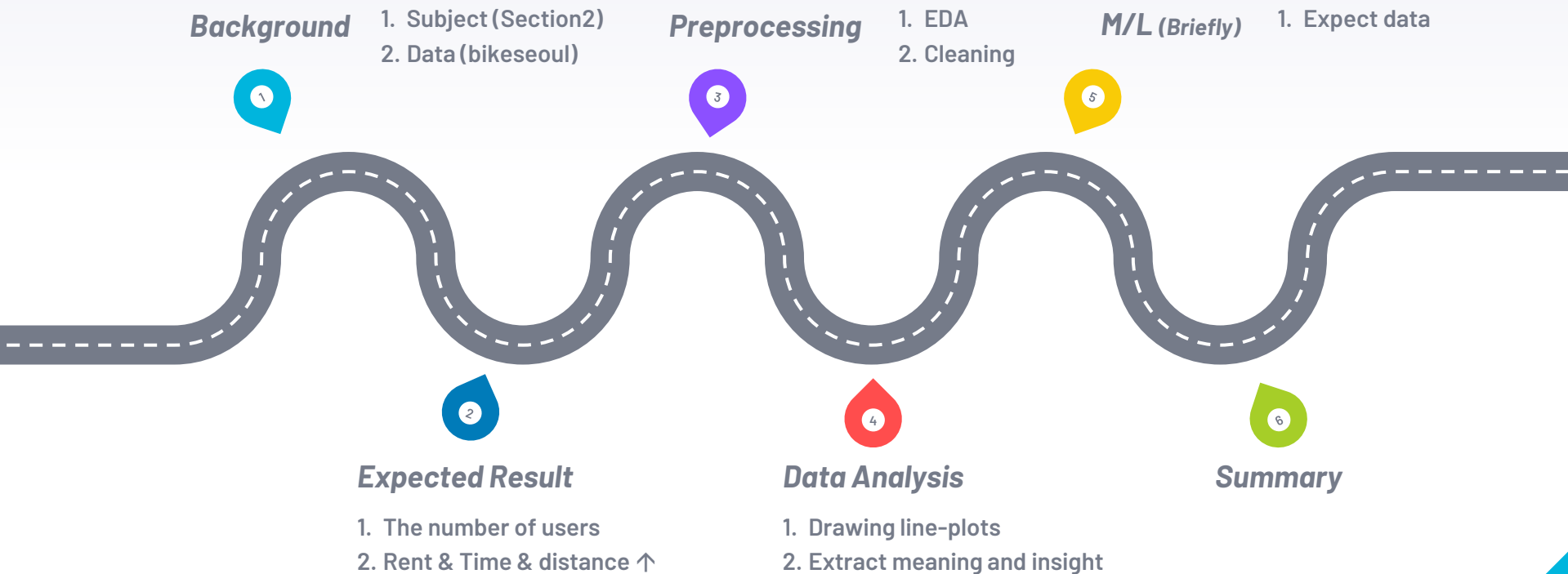
CODESTATES

AI-02

김병섭



# Contents



1,2

# *Background & Expected results*

---

1. Subject (Section2) 주제
2. Data (bikeseoul) 서울시 공공자전거 따릉이
3. Expected results 예상 결과





# Purpose

## Personal mobility is changing our lives.

(ex. Sharing economy, First-mile, Last-mile)

→ How to deal with the business in this **PANDEMIC?**



**Previous Project** (section 2, Lyft, San Francisco)

< Project = 선진국의 공유자전거 이용 현황 분석 및 이용자 예측 >

### • 진행 배경

1. 모빌리티 사업 내 공유 경제/구독 서비스의 사업 확대 증가
2. 모빌리티 다양화 및 시장 선점을 위한 사업 확대 (자전거, 스쿠터)
  - ※ 목적 : 변화하는 모빌리티 개념 및 수요 맞춤형 서비스 제공을 통한 고객 정보 확보  
(관련 용어 : 마이크로 모빌리티, 퍼스널 모빌리티, 라스트 마일)
  - ※ 대표 업체 : Lyft, Uber, Bolt, Lime (이상 해외), Elecle, HMC 등
3. COVID-19로 인한 모빌리티 사업 영향 : 공유 경제 수익성 악화 및 변화 필요



**Result = ★★★★★**

1. Limited information
2. Classify the membership



not good..



# Subject & Data

- ▶ Subject = Bikeseoul & COVID-19
- ▶ Data source = Seoul Open Data Plaza *서울시 열린데이터광장*
- ▶ Data information *사용한 데이터 정보*

1. Time = Jan. ~ June in 2019~2021 *지난 3년간 1월~6월*

2. File = New sub. Information (month) *공공자전거 신규 가입자 현황*

*(date, age code, sex, No. of people)*

**Bikeseoul usage info.(daily)** *공공자전거 이용 현황*

*(date, rental code, age code, No. of rental, distance, hours etc.)*

**Number of the Coronic in seoul** *서울특별시 코로나 확진자 현황*

*(date, No. of the Coronic)*



# EXPECTED RESULTS

Because of COVID-19,  
untact /personal(individual) life are becoming normal. So..



## 1. New subscribers, Users would increase (vs. 2020)

*코로나19가 장기화되면서 공공자전거 신규 가입자와 이용자가 2020년보다 늘어났을 것이다.*



## 2. No. of the rental, distance, hours would increase (vs. 2020)

*공공자전거/대여소 추가, 대중 교통 대체 수단으로 대여 건수, 이동 거리, 이용 시간 등 모두 증가했을 것이다.*

## 3,4

# *Preprocessing & Data Analysis*

---

1. EDA, Cleaning, etc.
2. Draw line plots & Extract meaning, insight



# Preprocessing (EDA, Data cleaning etc.)

- ▶ Unify the data file form to the latest version 양식 통일
- ▶ Delete NaN, '\n', 'BIL\_021' 결측치, 이상치, 오류 등 제거
- ▶ Delete data having zero in columns (distance, hours) 값인 0인 데이터 삭제
- ▶ Feature engineering (Speed, Day of week) 이동 속도, 평일/주말 구분
- ▶ Filtering the data : maximum speed, hours/rental 이상치 제거



## < Result >

Jan.~Jun. in 2021 = **2,739,751** (122,586 + 331,142 + 475,625 + 588,872 + 607,728 + 613,798)

Jan.~Jun. in 2020 = **1,190,594** (82,994 + 189,312 + 277,548 + 263,544 + 196,576 + 180,620)

Jan.~Jun. in 2019 = **1,524,670** (174,505 + 152,064 + 239,215 + 293,139 + 339,751 + 325,996)

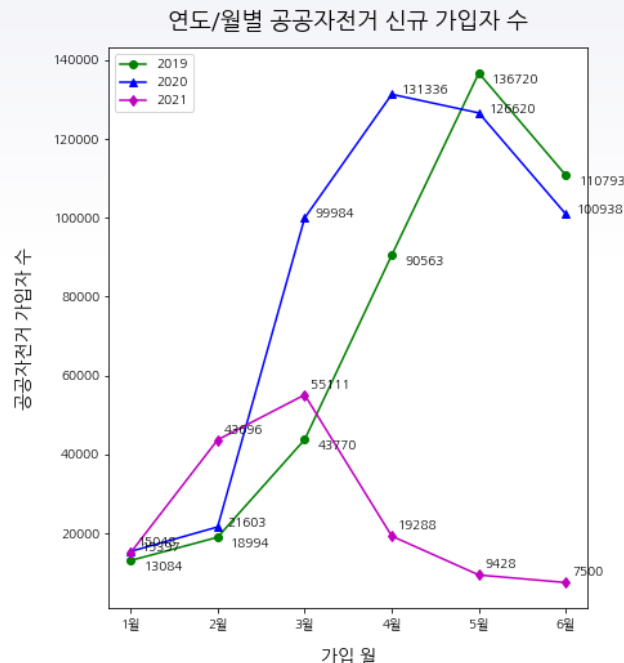
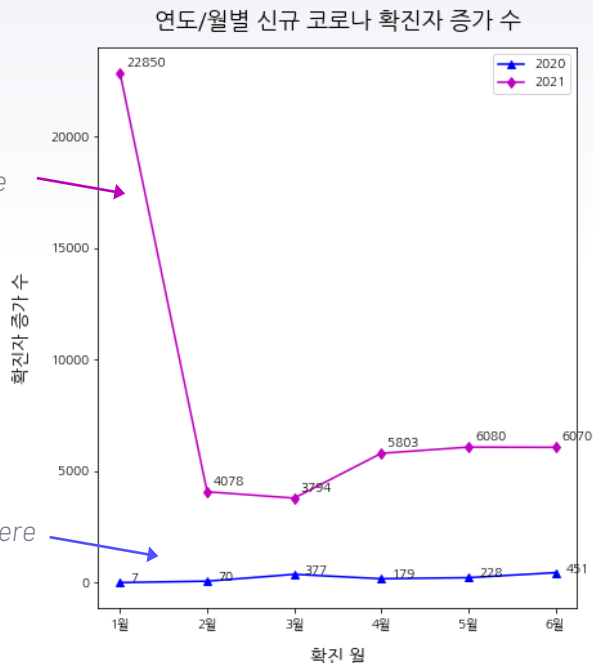




# Data Analysis\_COVID-19 & bicycles sub./month (by year)

[Status: 🔥]  
COVID-19 was severe

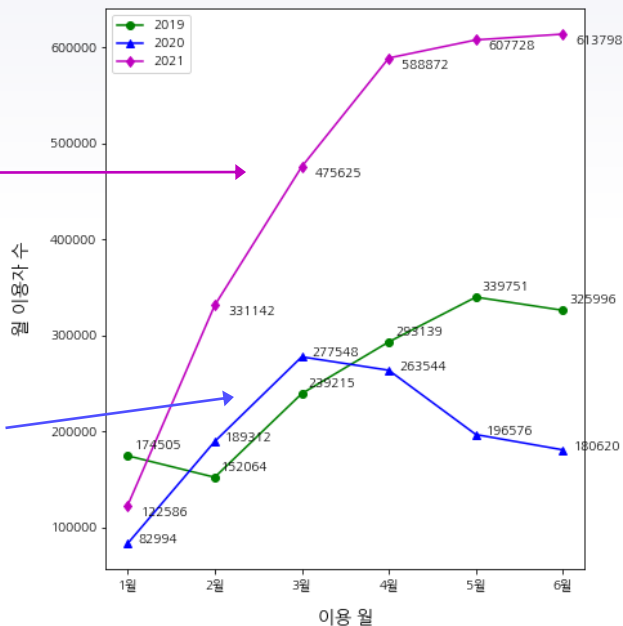
[Status: ☁️]  
COVID-19 wasn't severe



The increase of the Coronic decreases new subscribers. What about bicycles rental?

# Data Analysis\_The no. of bicycles users/month (by year)

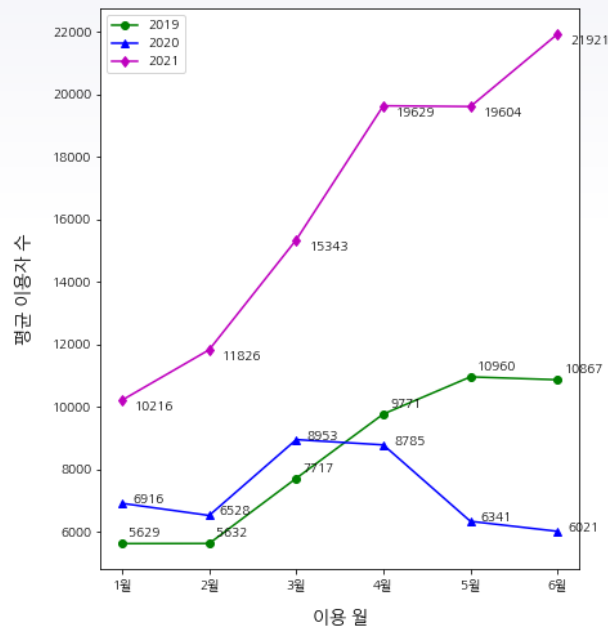
연도/월별 공공자전거 이용자 수



[Status: ]  
No. is increasing

[Status: ]  
No. is low

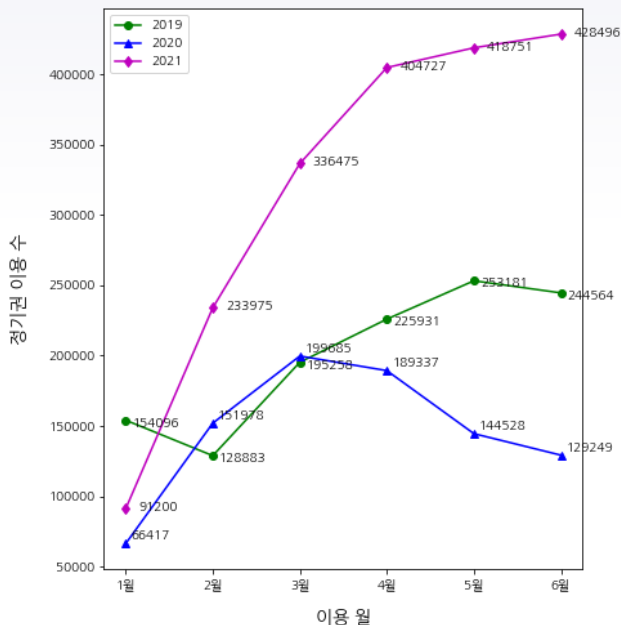
연도/월별 공공자전거 평균 이용자 수



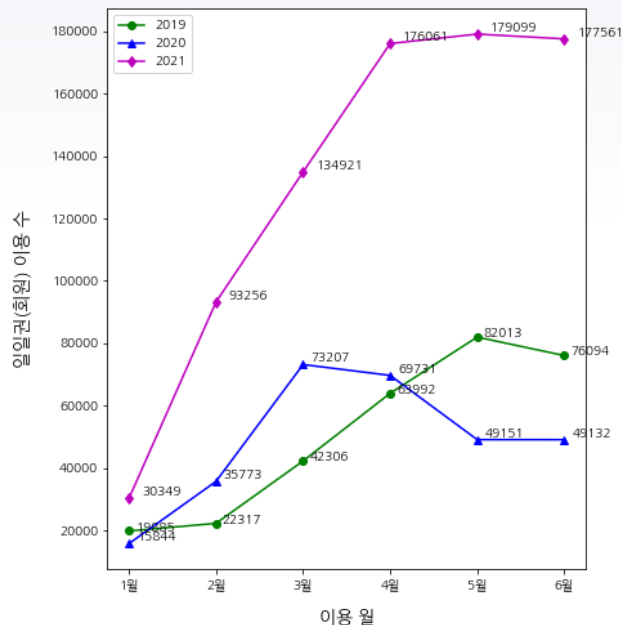
The no. of users (sum, mean) are continuously increasing, are they member?

# Data Analysis\_Rental code/month (by year)

연도/월별 대여 코드(정기)

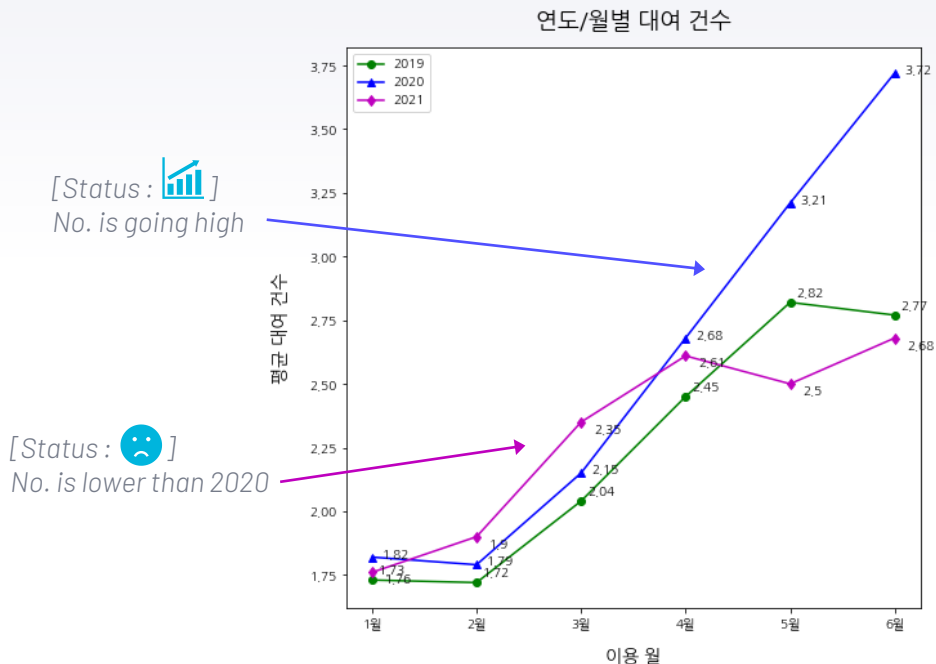


연도/월별 대여 코드(일일(회원))



The regular & daily ticket are taking the most in total (around 98%), **however..**

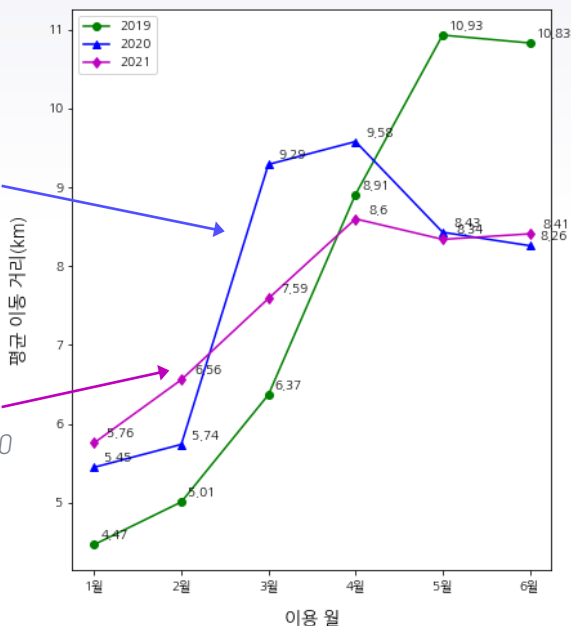
# Data Analysis\_The no. of rental/month (by year)



However, people don't use bicycles as many times as expected. What about distance & hours?

# Data Analysis\_Distance/month (by year)

연도/월별 평균 이동 거리



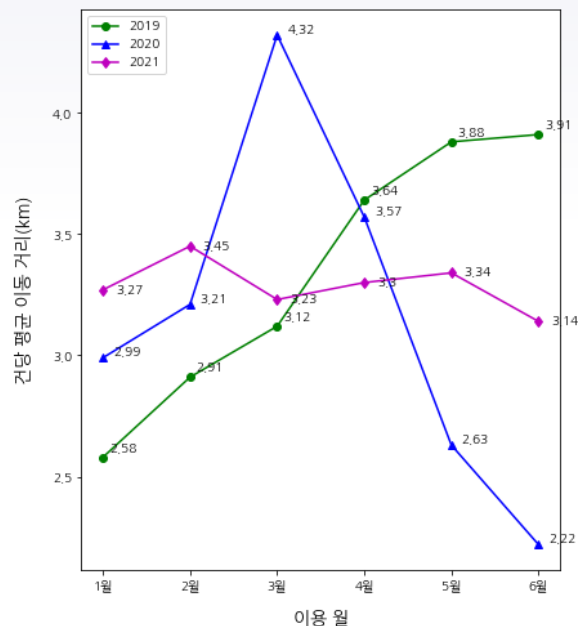
[Status: ]

No. is going high

[Status: ]

No. is lower than 2020

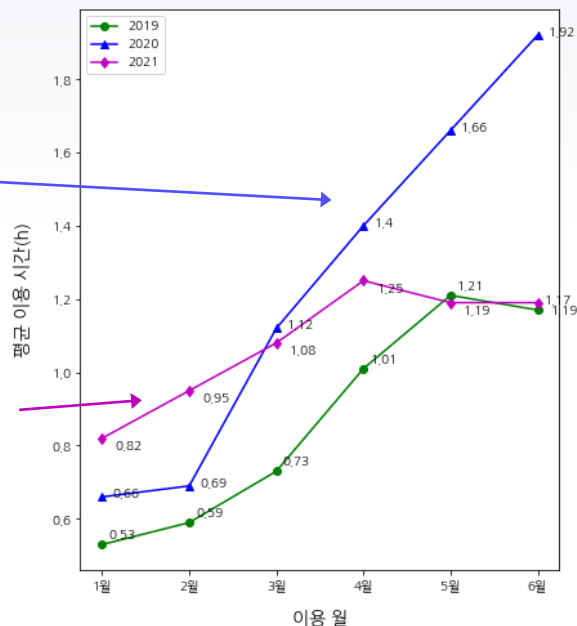
연도/월별 건당 평균 이동 거리



Distance (sum, mean) line plot shows people are using bicycles to go nearby place.

# Data Analysis\_Hours for riding bicycles/month(by year)

연도/월별 평균 이용 시간



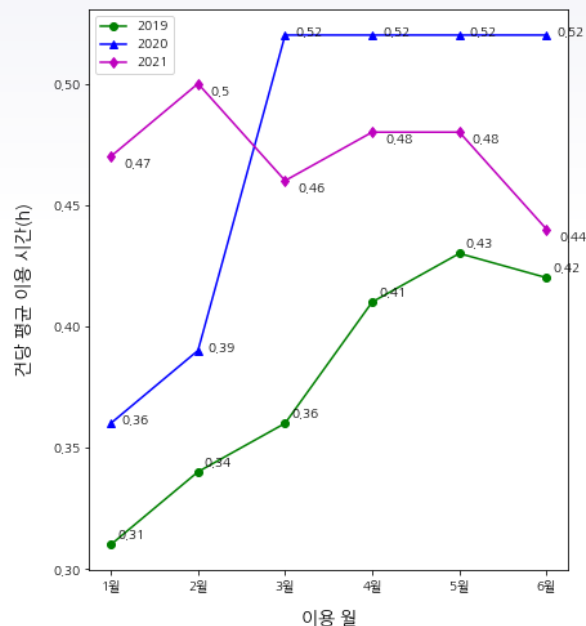
[Status: ]

No. is going high

[Status: ]

No. is lower than 2020

연도/월별 건당 평균 이용 시간

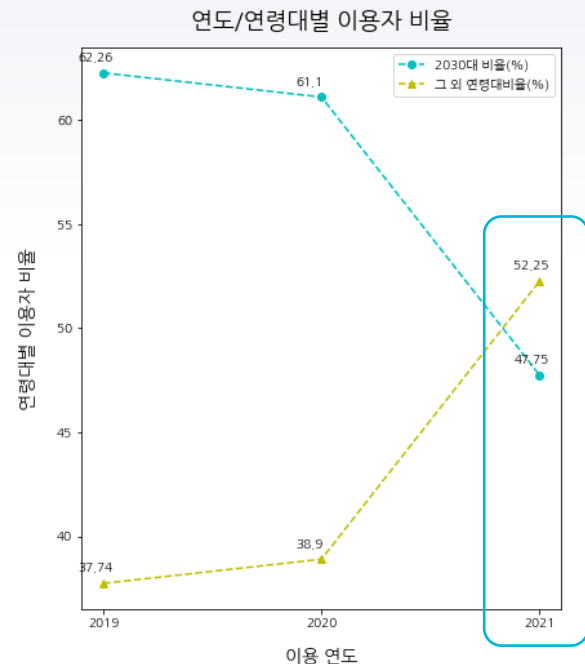
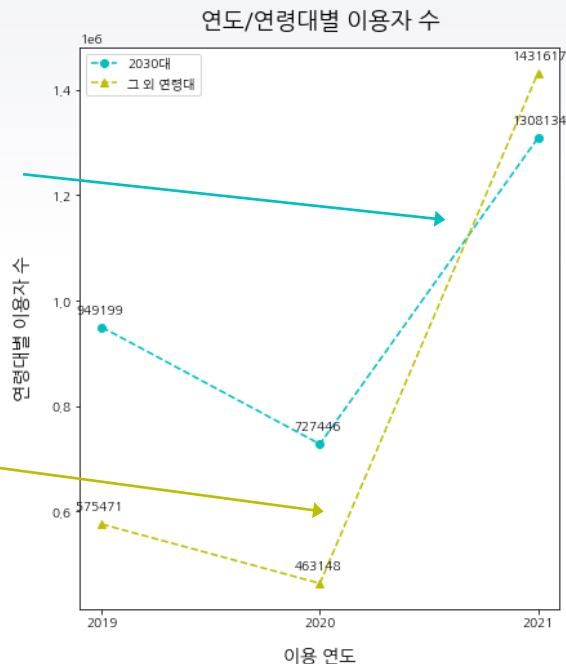


Hour in 2021 is also not bigger than hour in 2020. Why?

# Data Analysis\_2030 generation & others/month(by year)

[Status:   
No. is going high

[Status:   
No. is exploded!



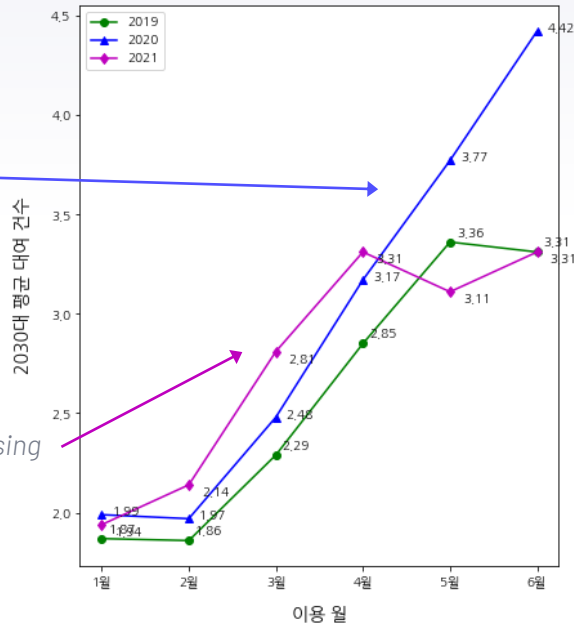
Overturn!



The No. of users is going high, but the group except 2030 gen. is increasing faster.

# Data Analysis\_2030 generation & others/month(by year)

연도/월별 2030대 대여 건수



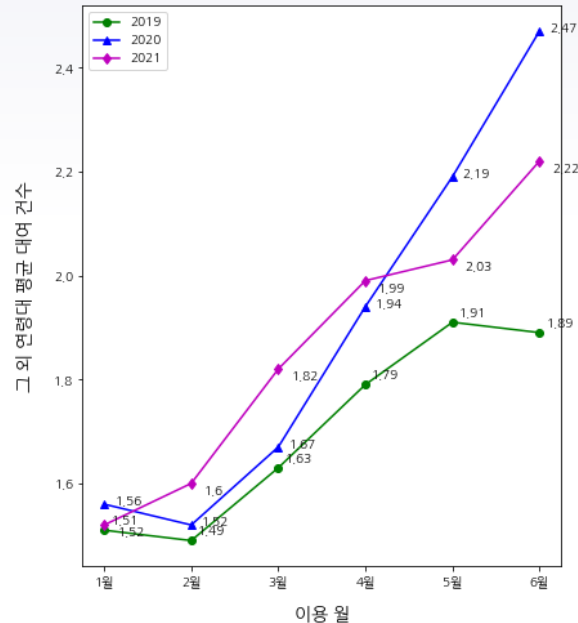
[Status: 

No. is going high

[Status: 

No. is generally increasing

연도/월별 그 외 연령대 대여 건수

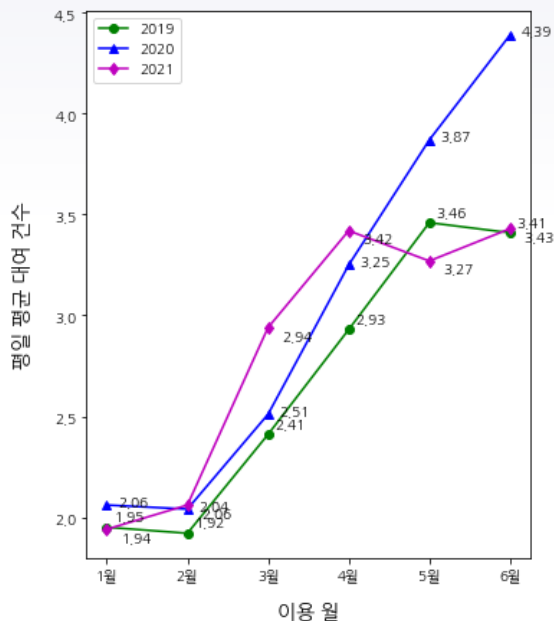


No. of the rental in '20~'21 is increasing, but the average value in '20 is bigger.

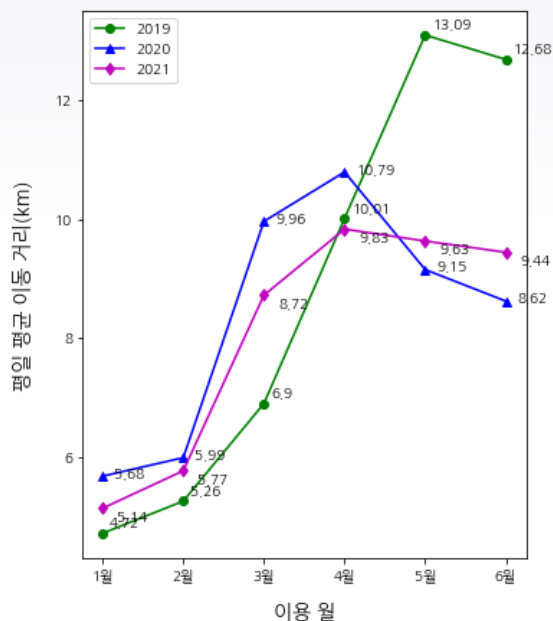


# Data Analysis\_2030 gen. weekday info./month(by year)

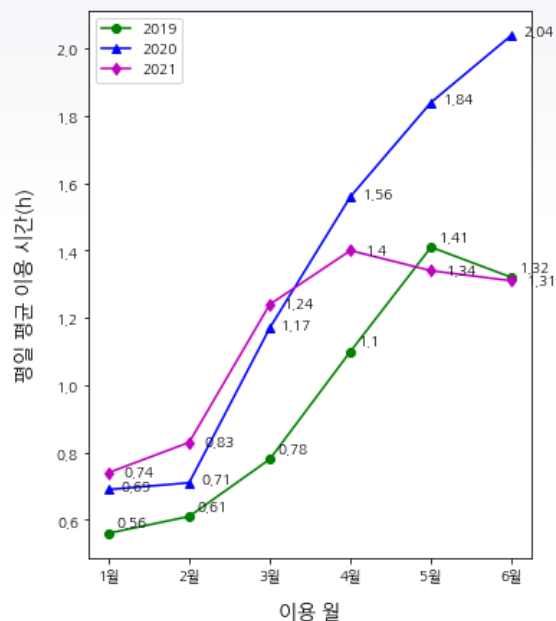
연도/월별 2030대 평일 평균 대여 건수



연도/월별 2030대 평일 평균 이동 거리



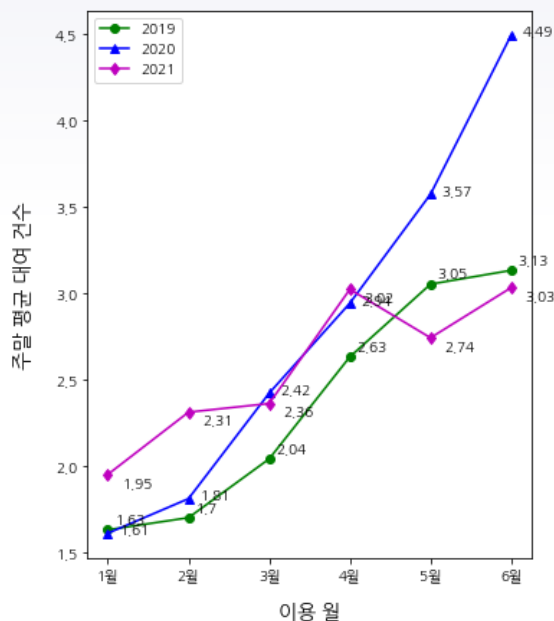
연도/월별 2030대 평일 평균 이용 시간



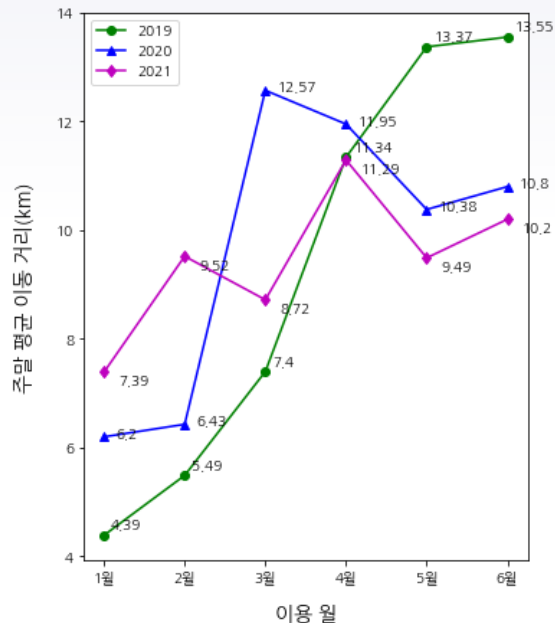
The results of 2030 gen. in 2021 is not bigger than the one in 2020.

# Data Analysis\_2030 gen. weekend info./month(by year)

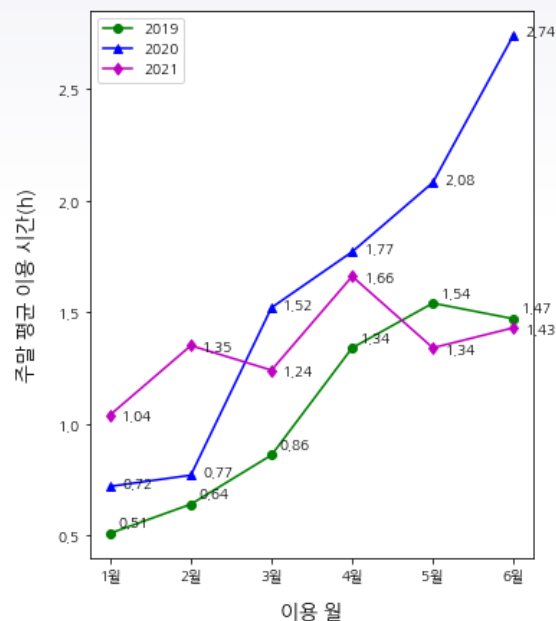
연도/월별 2030대 주말 평균 대여 건수



연도/월별 2030대 주말 평균 이동 거리



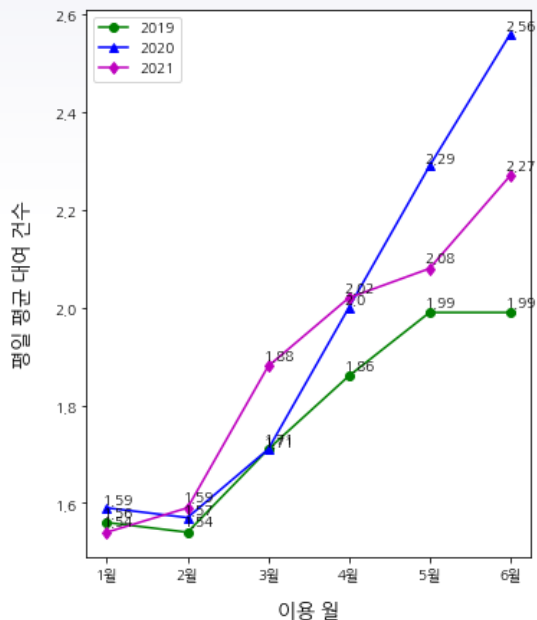
연도/월별 2030대 주말 평균 이용 시간



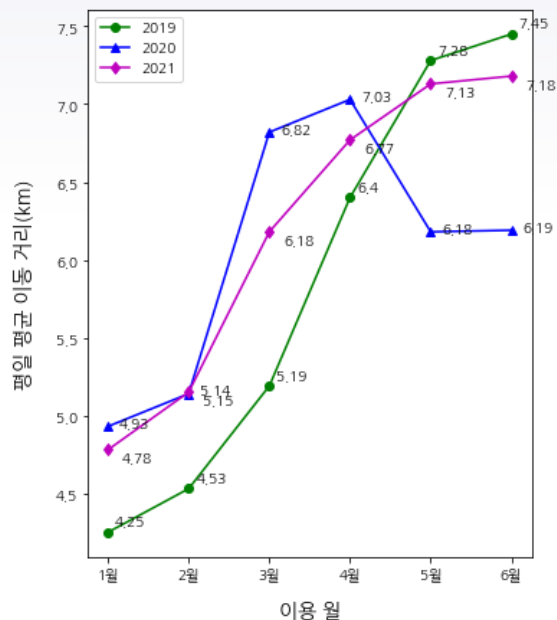
Similar results with weekdays. They are less using bicycles in 2021

# Data Analysis\_others gen. weekdays info./month(by year)

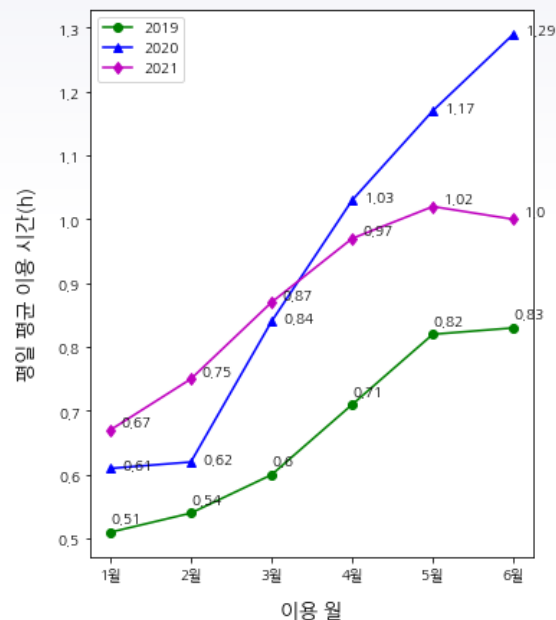
연도/월별 그 외 연령대 평일 평균 대여 건수



연도/월별 그 외 연령대 평일 평균 이동 거리



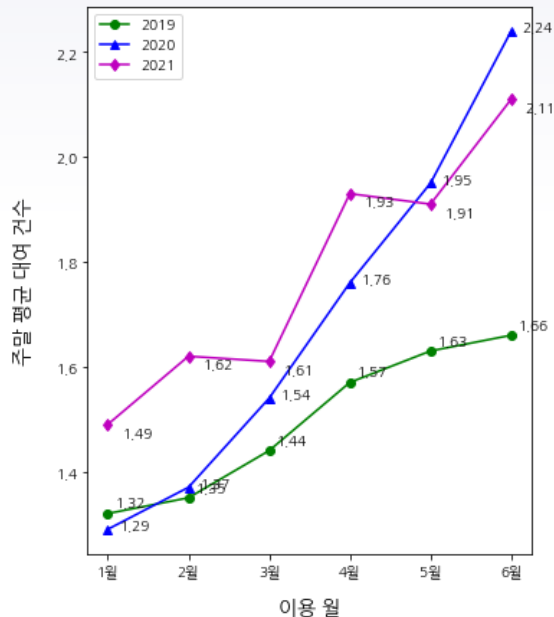
연도/월별 그 외 연령대 평일 평균 이용 시간



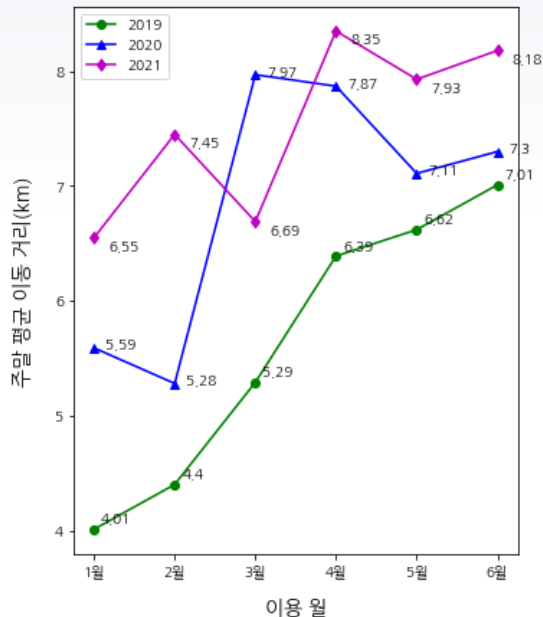
Unlike the large increase in new subscriber, they are not riding bicycles that much

# Data Analysis\_others gen. weekend info./month(by year)

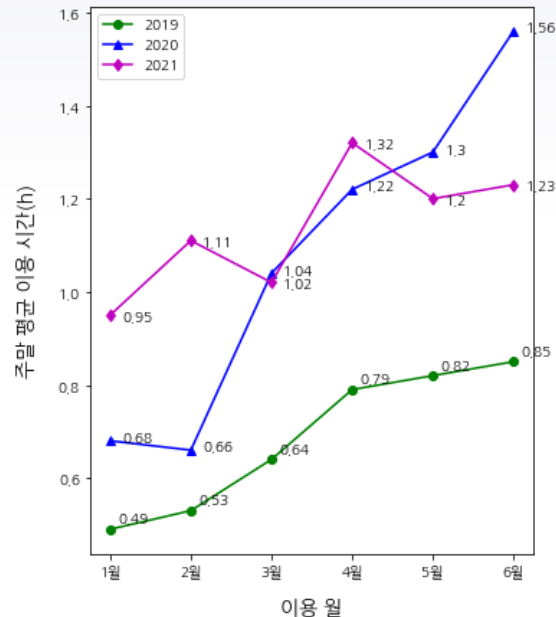
연도/월별 그 외 연령대 주말 평균 대여 건수



연도/월별 그 외 연령대 주말 평균 이동 거리



연도/월별 그 외 연령대 주말 평균 이용 시간



Similar patterns with 2030 gen., but They are using bicycles more!




# Summarize the data.

*By analyzing the Bikeseoul data, I think..*

  
match with  
expected  
result

**01** People are willing to use sharing bicycles in daily life in **COVID-19** pandemic.

  
not match  
with expected  
result

**02** Overall, the average values in almost aspects are less in 2021, but **scale is getting bigger.**

**03** Not only 20~30 gen., the other gen. are riding bicycles in untact life.



# 5

## *Machine Learning*

---

1. Expect data (briefly 🤖)  
→ Target = distance



# Why distance? & How to expect?

Main feature  
in data.



The most  
reliable info.



The base info  
for business.



The figure depends on  
personality.  
(Sex, Age, Rental code)

The least error in data  
comparing to the others

Bicycles, rental station  
management

Machine Learning,  
(Regression Model)

- RandomForestRegressor  
- XGBRegressor  
(OneHot & OrdinalEncoder,  
Hyperparameters)



Before (mae)

6.45

Baseline (mean)

After (mae)

2.75(81%)

Train = Jan~May.2021  
Test = Jun.2021



# Project Summary

---

- ▶ What if survey?
- ▶ Data analysis is still DIFFICULT.
- ▶ Machine Learning, It's GONE in my head...

**Any feedback or questions?**

You can find me at github: @Sammy308

