

작성일자	2022-04-08
작성자	유호준

# 5조 프로젝트 제안서

따릉이 프로젝트

조장	박종원
조원	김형우
	류제범
	유호준
	임낙현
	하동호

---

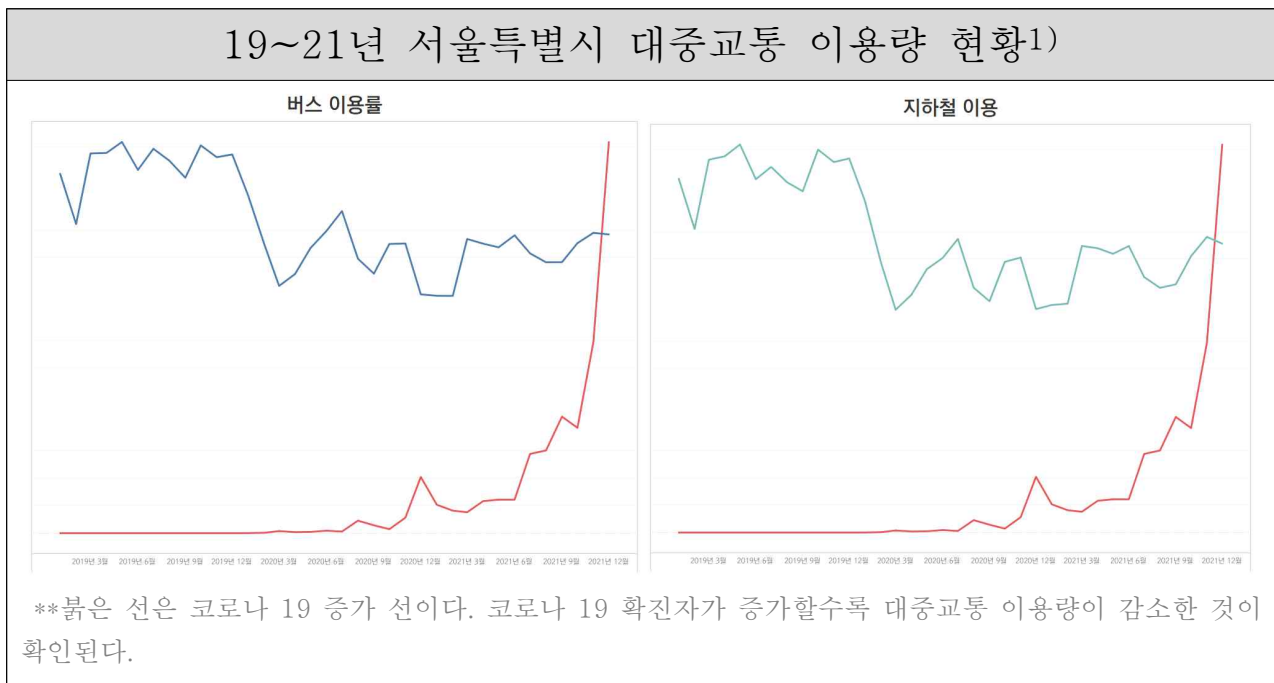
# 목차

---

1. 기획 의도 .....	3p
2. 데이터셋 .....	4~5p
3. 분석 .....	6~9p
4. 모델링 .....	10p
5. 결론 .....	11p
6. 팀원별 역할 .....	12p
7. 일정 .....	13p

# 1. 기획 의도

- 서울특별시에서는 코로나 19로 인해 대중교통 이외의 제3의 이동수단이 떠오르고 있다.
- 본 제안은 시민들의 새로운 니즈에 맞춰 적절한 제3의 교통수단을 개발하기 위해 현재 주로 이용하는 교통수단 중 적절한 수단을 선정하고, 더 발전시킬 수 있는 방안을 고안하는 것이다.
- 코로나 19는 우리 삶에 다양한 변화를 초래했는데 그중에서도 일상생활에 가장 많은 영향을 주었다. 아래의 그래프를 보면 실제로 19년 대비 20~21년 국내 대중교통 이용량은 급격하게 줄었음을 알 수 있다.



- 우리 조는 줄어든 대중교통 이용대신 시민들이 어떤 이동수단을 이용할까? 에 대한 의문점을 가지게 되었고 프로젝트 주제를 선정하게 되었다. 조사 대상으로 정한 제3의 교통수단은 아래와 같다.

제3의 교통수단 리스트				
전동 킥보드	사륜차	이륜차	렌트카	따릉이 (공공 자전거)

1) 티머니 월간 교통카드 통계 자료 <https://pay.tmoney.co.kr/>

## 2. 데이터셋

### ☐ 따릉이

- 서울특별시 19~21년 따릉이 대여 정보
- ~~ 컬럼 이용

따릉이 데이터셋 <sup>2)</sup>			
구분	컬럼	로우	발체형식
대여정보	4	72,642,362	CSV
고장정보	3	337,800	CSV

### ☐ 대중교통

- 서울특별시 19~21년 버스, 지하철 이용 정보
- 사용월, 승차승객수 컬럼 이용

대중교통 데이터셋 <sup>3)</sup>			
구분	컬럼	로우	발체형식
버스	9	1,403,217	CSV
지하철	7	21,654	CSV

### ☐ 사·이륜차

- 전국·서울 19~21년 사·이륜차 등록 정보
- 기간, 승용차 합계, 승합차 합계, 이륜차 합계 컬럼 이용

사·이륜차 데이터셋 <sup>4)</sup>			
구분	컬럼	로우	발체형식
수치	25	973	CSV

2) 서울시 공공자전거 대여이력 정보 <https://data.seoul.go.kr/dataList/OA-15182/F/1/datasetView.do>

서울시 공공자전거 고장신고 내역 <https://data.seoul.go.kr/dataList/OA-15644/F/1/datasetView.do>

3) 티머니 월간 교통카드 통계 자료 <https://pay.tmoney.co.kr/>

4) 서울시 자동차등록 (월별/구별) 통계 <https://data.seoul.go.kr/dataList/255/S/2/datasetView.do>

## ☐ 코로나 19

- 전국 19~21년 코로나 확진자 정보
- 년도, 확진자 수 컬럼 이용

코로나19 데이터셋 <sup>5)</sup>			
구분	컬럼	로우	발체형식
수치	2	37	CSV

## ☐ 렌트카·전동 킥보드

- 서울특별시 19~20년 렌트카, 전동 킥보드 대여 정보
- 날짜, 카테고리, 사용 카운트 컬럼 이용

렌트카·전동 킥보드 데이터셋 <sup>6)</sup>			
구분	컬럼	로우	발체형식
수치	4	39	CSV

5) 질병관리청 코로나19 <https://www.kdca.go.kr/index.es?sid=a2>

6) 금융데이터 거래소:

[https://www.findatamall.or.kr/fsec/dataProd/generalDataProdDetail.do?cmnx=44&goods\\_id=98851420-d86b-11ea-a506-cf31c4c94055](https://www.findatamall.or.kr/fsec/dataProd/generalDataProdDetail.do?cmnx=44&goods_id=98851420-d86b-11ea-a506-cf31c4c94055)

### 3. 데이터 분석

#### ① 코로나 19 증감에 따른 서울시 대중교통 이용량 분석

표1. 19~21년 서울특별시 대중교통 이용량 현황

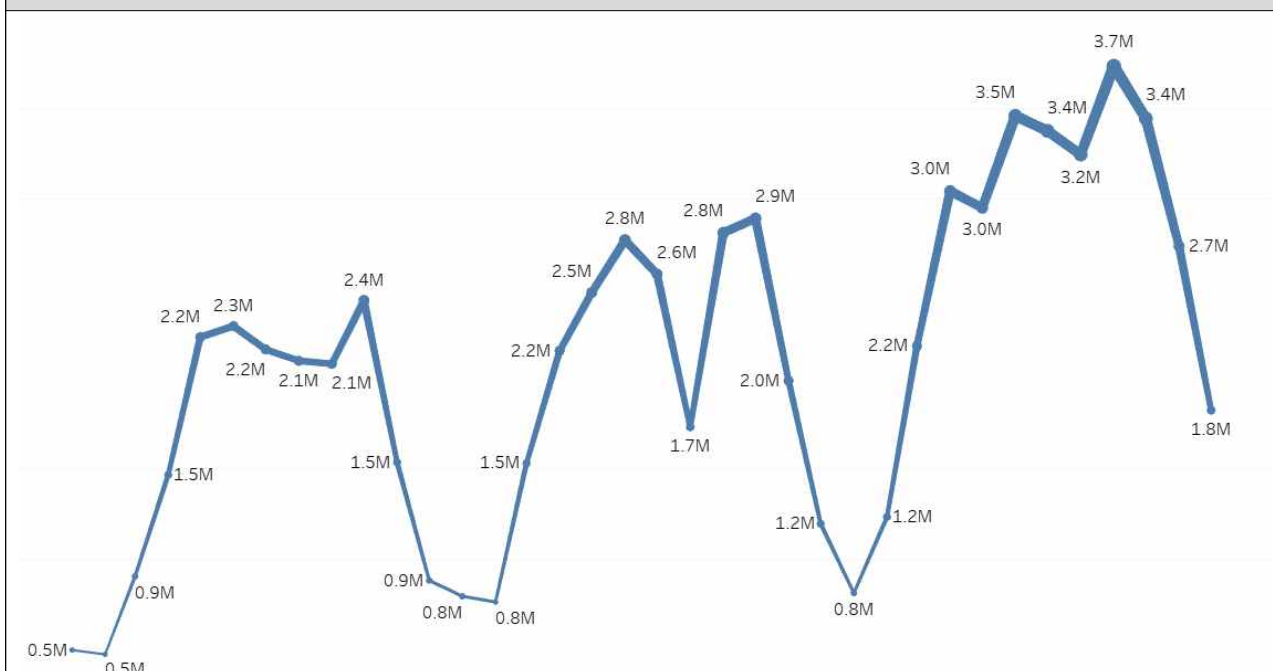


\*\*붉은 선은 코로나 19 증가 선이다. 코로나 19 확진자가 증가할수록 대중교통 이용량이 감소한 것이 확인된다.

#### ② 서울시 대중교통 이용량 증감에 따른 이동수단 이용량 분석

○ 따릉이의 경우 계절에 영향을 받지만 꾸준히 증가하는 추세

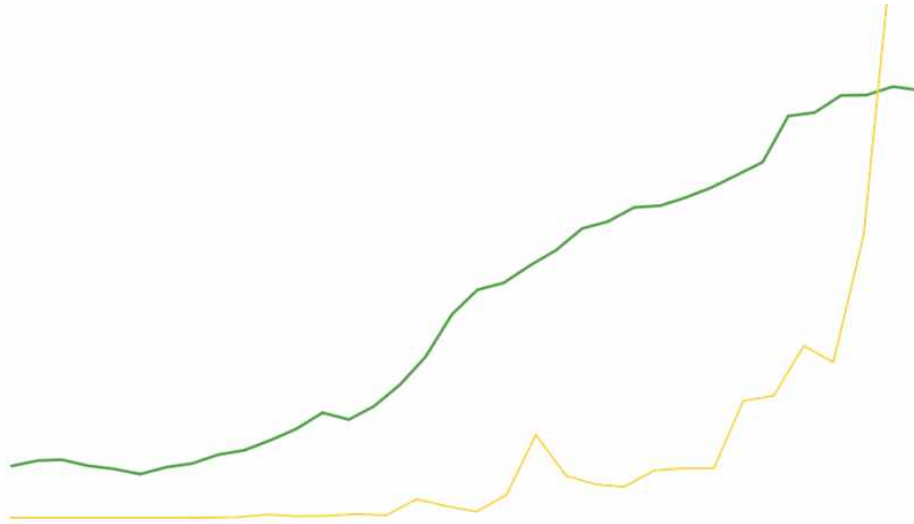
표2. 따릉이 이용 추세



○ 코로나 19가 점점 증가할 때마다 서울특별시 4륜차 등록 수가 올라가는 양상을 보임

표3. 코로나 19 영향에 따른 4륜차 등록 추세

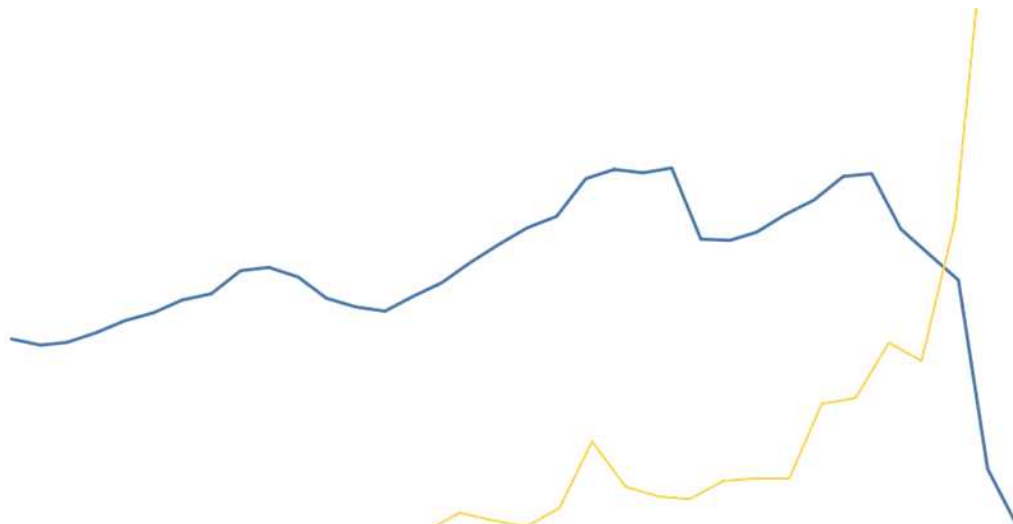
노란 그래프 : 코로나 증가율



○ 4륜차와 반대로 2륜차의 경우 코로나 19가 증가할수록 감소하는 양상을 보임

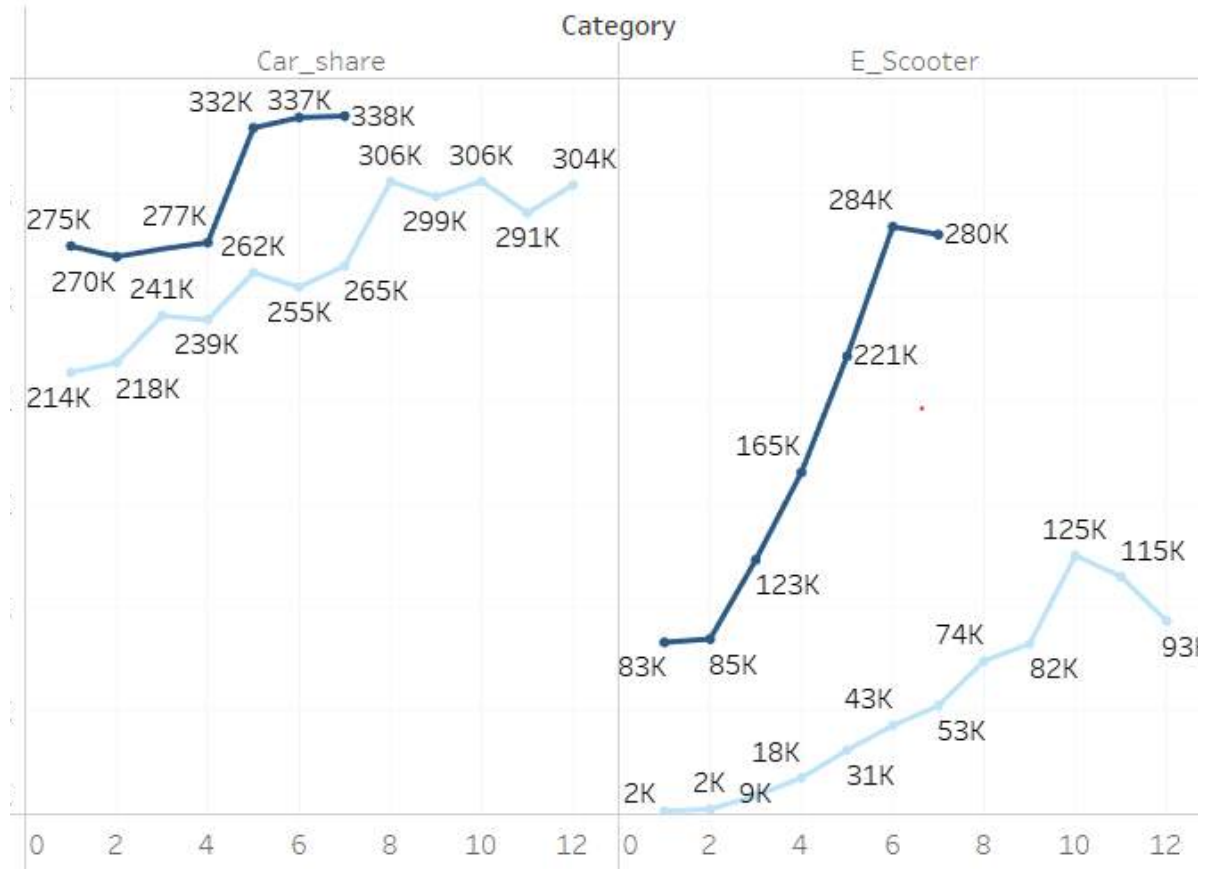
표4. 코로나 19 영향에 따른 2륜차 등록 추세

노란 그래프 : 코로나 증가율



○ 렌트카와 공유 킥보드 또한 코로나 이전보다 사용량이 큰 폭으로 증가함

표5. 공유 킥보드, 렌트카 이용 추세



\*\* 진한 그래프 : 2020년, 연한 그래프 : 2019년

## □ 타겟 교통수단 선정

○ 그래프 분석 결과 대중교통과 이륜 자동차를 제외하곤 여러 이동수단의 이용량이 큰 폭으로 증가하고 있었으며, 특이사항으로 따릉이의 경우 계절의 영향에 따라 수요의 변화가 있었다.

○ 우리 조는 개발 이동수단 대상으로 '따릉이'를 선정

○ 따릉이 선정 이유는 다음과 같다.

- 이용자 입장에서 비용적으로 저렴하다. 이용자에게 이륜, 사륜차의 구매와 렌트카의 경우 큰 비용을 요구한다.
- 개발할 점이 명확하고, 단순하다. 데이터셋 분석 결과 서울시 따릉이는 유지보수로 인한 적자 상황이었다. 유지보수 내용은 단순 관리 장소와 노후화 문제였다.
- 친환경적이다. 다른 이동수단과 다르게 따릉이는 환경 측면에서 공공의 이익이 된다.



## 4. 모델링

### 1. 고장 주기 예측

```
from sklearn.linear_model import HuberRegressor

from sklearn.model_selection import train_test_split

train_features, test_features, train_labels, test_labels = train_test_split(features, label)

from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
train_features = scaler.fit_transform(train_features)
test_features = scaler.transform(test_features)

model = HuberRegressor()
model.fit(train_features, train_labels)

print(model.score(train_features, train_labels))
print(model.score(test_features, test_labels))
```

0.23337673938012726  
0.327065294477877

```
from sklearn.ensemble import GradientBoostingRegressor

from sklearn.model_selection import train_test_split
train_features, test_features, train_labels, test_labels = train_test_split(features, label)

from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
train_features = scaler.fit_transform(train_features)
test_features = scaler.transform(test_features)

model = GradientBoostingRegressor()
model.fit(train_features, train_labels)

print(model.score(train_features, train_labels))
print(model.score(test_features, test_labels))
```

0.512214582265118  
0.29857811877250096

○ HuberRegressor의 경우가 test데이터의 수치가 가장 높았다.

### 따릉이의 세대 구분

종류	고유번호	비고
LCD형	<u>00001~10000</u>	1K-1K H 초기형
	<u>10001~20000</u>	2 후기형
	<u>20001~27000</u>	3 개선형 <sup>[4]</sup>
QR형 따릉이	<u>30001~</u>	4 뉴따릉이
	<u>40001~</u>	5 QR단말기(뉴따릉이)형 <sup>[비고]</sup>
	47000~60000	6 뉴따릉이형. 시트포스트가 검정이고 안장 바로 밑에서 높낮이 조절이 가능하다. 4만번대에서 지적되던 너무 낮은 안장 문제가 해결되었다. <sup>[6]</sup>
새싹따릉이	<u>80001~</u>	7 어린이용 따릉이

## 2. 세대별 고장 부위 예측

```
from sklearn.ensemble import GradientBoostingRegressor

from sklearn.model_selection import train_test_split
train_features, test_features, train_labels, test_labels = train_test_split(features, label)

from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
train_features = scaler.fit_transform(train_features)
test_features = scaler.transform(test_features)

model = GradientBoostingRegressor()
model.fit(train_features, train_labels)

print(model.score(train_features, train_labels))
print(model.score(test_features, test_labels))

0.12962081458685837
0.014885264216688121
```

○ GradientBoostingRegressor로 분석시 test데이터가 1%가 나왔다.

---

## 5. 결론

---

### ☐ 대중교통

- 코로나 19 이후 2020년 급격히 확진자가 증가하는 경우 대중교통의 이용자 또한 확연히 줄어들고 이후 다시 조금씩 늘어가는 추세를 확인하였다. 다만 21년 이후 확진자의 수가 급격히 증가하여도 이용자 수가 줄어드는 것은 20년보다 덜 한 것을 확인할 수 있었다. 하지만 코로나 이전과 비교하였을 경우 이용자 수가 많이 줄어들었음을 부인할 수 없다.
- 이에 따라 우리 조는 서울의 인구는 다수의 사람과 접촉하여 이동하는 대중교통이 코로나 확진의 원인이 될 수 있기에 이용하지 않았으리라 생각하였고 이를 대체할 수 있는 교통수단이 무엇인지 확인해보았다.

### ☐ 따릉이 선정

- 이후 서울에 존재하는 교통수단인 자차, 따릉이, 공유 킥보드, 렌트카에 대하여 조사를 해보았으며, 이 중 따릉이에 집중하여 보았다.
- 이유는 따릉이가 코로나 이후 가장 사용량이 많아진 교통수단이었으며 친환경적이고 이용자의 비용부담이 적었다. 자차의 경우 대중교통의 대체수단이라하기엔 무리라 할 수 있고, 공유 킥보드와 렌트카의 경우도 서울시 전체의 데이터를 확인하기엔 무리였다.(금전적인 이유와 데이터 부족)

### ☐ 그룹의 목표

- 따릉이는 사용량이 늘어난 만큼 공급량 또한 증가하였고 계속 증가시킬 예정이다.<sup>7)</sup> 하지만 그 사용량과는 별개로 적자의 폭이 늘어나고 있다.<sup>8)</sup> 이 적자를 줄이기 위해 고장주기분석과 고장부위의 분석, 예측을 통해 적절한 시기에 고장을 고칠 수 있다면 년마다 늘어나는 따릉이 완제품 공급량을 좀 더 조절할 수 있다고 생각하였다.

### ☐ 분석결과

- 고장주기와 부위를 예측한 결과 고장부위 예측은 힘들었으며, 이를 극복하기 위해서는 따릉이에 대한 더욱 다양하고 세부적인 데이터가 필요해보인다.  
1세대 생산품의 경우 이전고장 부위를 분석하였을 때, 이후에도 동일 부위의 신고가 계속되었다. 이에 2~7세대들의 고장재발률을 GradientBoostingRegressor으로 분석하였을 때 5~12%의 결과를 얻었다.

---

7) <https://mediahub.seoul.go.kr/archives/2003645> 따릉이 이용증가로 공급 증가

8) <https://www.hankyung.com/politics/article/202104291609i> 따릉이 적자

## 6. 팀원별 역할

박종원	유호준	김형우
		
조장, CTO	기획자	시각화 (그래프, ppt)
주제 선정 데이터셋 탐색 전처리 머신러닝	주제 선정 데이터셋 탐색 전처리 기획서 작성	주제 선정 데이터셋 탐색 전처리 ppt 제작

류제범	하동호	임낙현
		
CTO	시각화 (그래프)	CTO
주제 선정 데이터셋 탐색 전처리 머신러닝	주제 선정 데이터셋 탐색 전처리 시각화 담당	주제 선정 데이터셋 탐색 전처리

## 7. 일정

### ☐ 타임 테이블

1일차 (2022-04-05)		
	주제 선정	09:00~11:40
	점심	11:40~13:10
	주제 선정	13:10~18:00
2일차 (2022-04-06)		
	데이터셋 조사	09:00~11:40
	점심	11:40~13:10
	데이터셋 조사, 정제	13:10~18:00
3일차 (2022-04-07)		
	데이터 정제	09:00~11:40
	점심	11:40~13:10
	데이터 정제 데이터 시각화 기획서 작성 머신러닝	13:10~18:00
4일차 (2022-04-08)		
	머신러닝 데이터 시각화 기획서 작성	09:00~11:40
	점심	11:40~13:10
	발표 준비 및 발표	13:10~18:00