



# 뉴노멀 시대 준비를 위한 서비스

---

# 분석 방법

빅데이터 분석 방법론 활용

데이터  
수집

제공된 데이터 중 4종류 활용  
추가적인 데이터 4종류 수집

데이터  
전처리

Data Preprocessing  
Feature Engineering

데이터  
EDA

Data Visualization  
Correlation Analysis

Insight  
도출

Insight 도출  
Pain Point 파악

서비스  
기획

서비스 아이디어 기획  
서비스 구체화  
프로토타입 제안

# 목차

---

## 제1부. 테이터

## 제2부. 서비스

# 제1부. 데이터

---

서비스 제안을 위한 데이터 탐색 / 분석

# 제1부. 데이터

## 진행 순서

### 제공된 데이터

유동인구데이터(SK텔레콤)  
카드매출데이터(신한카드)  
유통데이터(GS리테일)  
물류데이터(CJ올리브네트웍스)

### 코로나 집계

코로나집계데이터(Kaggle)  
코로나키워드 검색률데이터(네이버 데이터랩)

### 재난문자 집계

재난문자 건수데이터(행정안전부)  
재난문자키워드 검색률데이터(네이버 데이터랩)

# 제1부. 데이터

탐색 / 분석 순서

전처리

지역 단위 설정  
날짜 코드 변환  
날짜 이상치 처리

Feature  
Engineering

통합 테이블 정의  
Scaled 증가량 계산

EDA

Visualization  
Insight

# 제1부. 데이터

가설1. 제공된 데이터 간 유의미한 상관관계가 있을 것이다

# 제1부. 데이터

## 제공된 데이터

유동인구데이터(SK텔레콤)  
카드매출데이터(신한카드)  
유통데이터(GS리테일)  
물류데이터(CJ올리브네트웍스)

## 코로나 집계

코로나집계데이터(Kaggle)  
코로나키워드 검색률데이터(네이버 데이터랩)

## 재난문자 집계

재난문자 건수데이터(행정안전부)  
재난문자키워드 검색률데이터(네이버 데이터랩)

## 전처리

지역 단위 설정  
날짜 코드 변환  
날짜 이상치 처리

## Feature Engineering

통합 테이블 정의  
Scaled 증가량 계산

## EDA

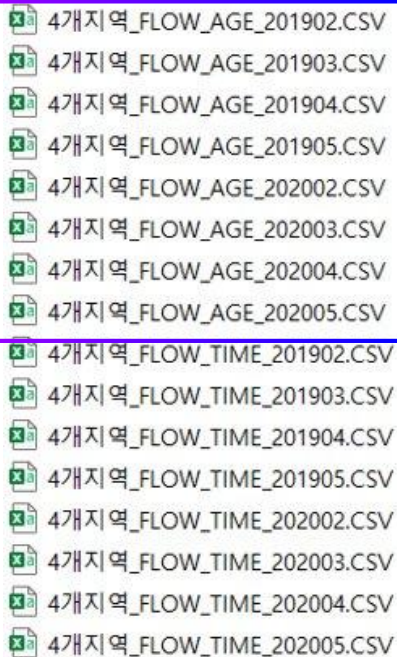
Visualization  
Insight



# 제1부. 데이터

제공된 데이터 - 유동인구 > 전처리

## ▶ 4개지역\_FLOW\_AGE 데이터를 이용했다



4개지역\_FLOW\_AGE\_201902.CSV  
4개지역\_FLOW\_AGE\_201903.CSV  
4개지역\_FLOW\_AGE\_201904.CSV  
4개지역\_FLOW\_AGE\_201905.CSV  
4개지역\_FLOW\_AGE\_202002.CSV  
4개지역\_FLOW\_AGE\_202003.CSV  
4개지역\_FLOW\_AGE\_202004.CSV  
4개지역\_FLOW\_AGE\_202005.CSV  
4개지역\_FLOW\_TIME\_201902.CSV  
4개지역\_FLOW\_TIME\_201903.CSV  
4개지역\_FLOW\_TIME\_201904.CSV  
4개지역\_FLOW\_TIME\_201905.CSV  
4개지역\_FLOW\_TIME\_202002.CSV  
4개지역\_FLOW\_TIME\_202003.CSV  
4개지역\_FLOW\_TIME\_202004.CSV  
4개지역\_FLOW\_TIME\_202005.CSV

# 제1부. 데이터

제공된 데이터 - 유동인구 > 전처리

▶ 코로나 감염증 확산으로 유동인구가 전반적으로 감소했을 것이다

" 우한 코로나'로 서울시 공무원임용 필기시험 또 6월 연기 " - 조선비즈, 2020.03.31

" [신종코로나] 확산 우려에 국내 축제 줄줄이 취소 " - 아주경제, 2020.02.01

" 한국철도, 코로나19로 광명역 도심공항터미널 운영 3개월 중단 " - 비즈니스포스트, 2020.03.31

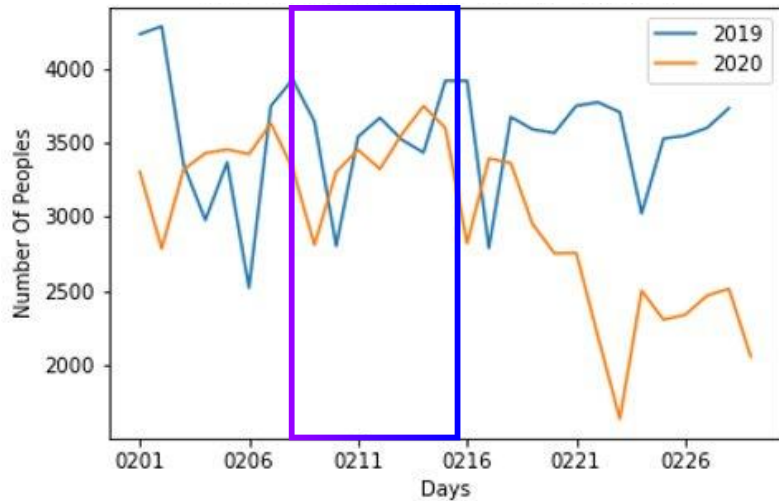
" 뮤지컬 '올 아이즈 온 미', 5월1일로 개막 연기 ...코로나19 여파 " - MK스포츠, 2020.03.31

# 제1부. 데이터

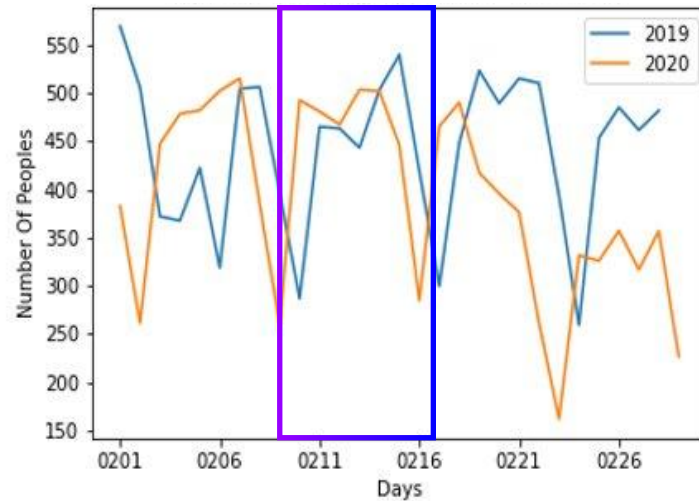
제공된 데이터 - 유동인구 > 전처리

- ▶ 행정동 별로 전년도 비교 시, 모든 경우에서 감소세를 보이는 것은 아니다

대봉1동 남성 40~44세 유동인구



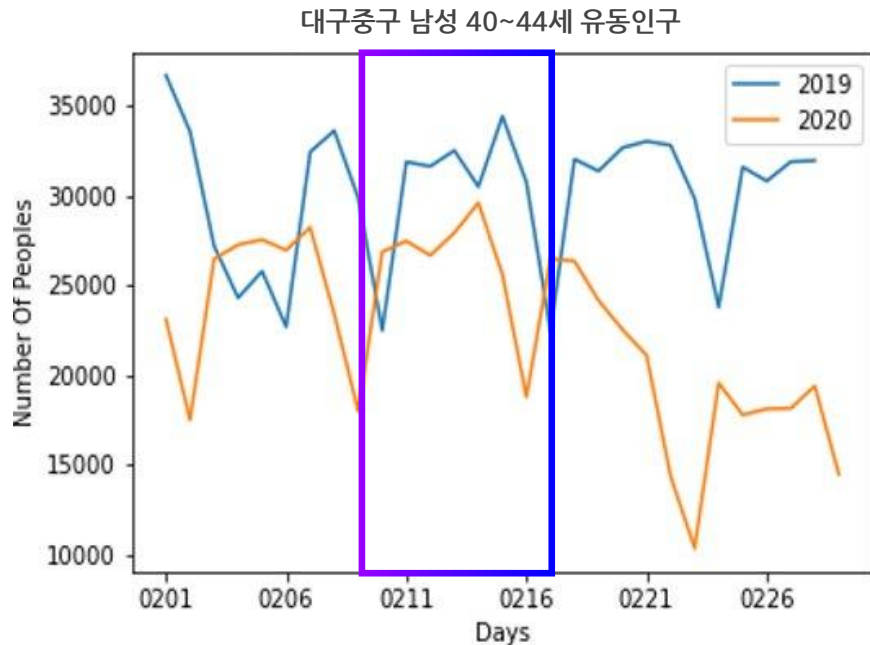
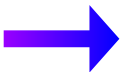
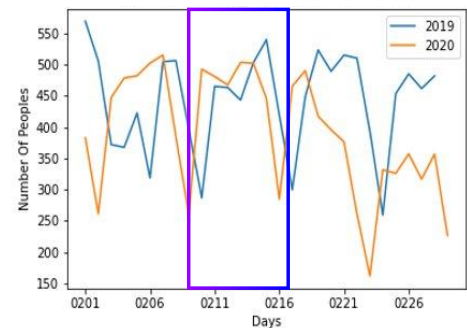
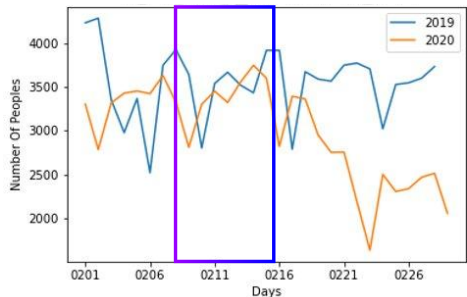
대봉2동 남성 40~44세 유동인구



# 제1부. 데이터

제공된 데이터 - 유동인구 > 전처리

▶ 구별로 전년도와 비교 시, 확실한 감소세를 보인다



# 제1부. 데이터

---

제공된 데이터 - 유동인구 > 전처리

## 지역 단위 설정

---

지역 단위를 행정동이 아닌 구 단위로 설정한다

# 제1부. 데이터

제공된 데이터 - 유동인구 > 전처리

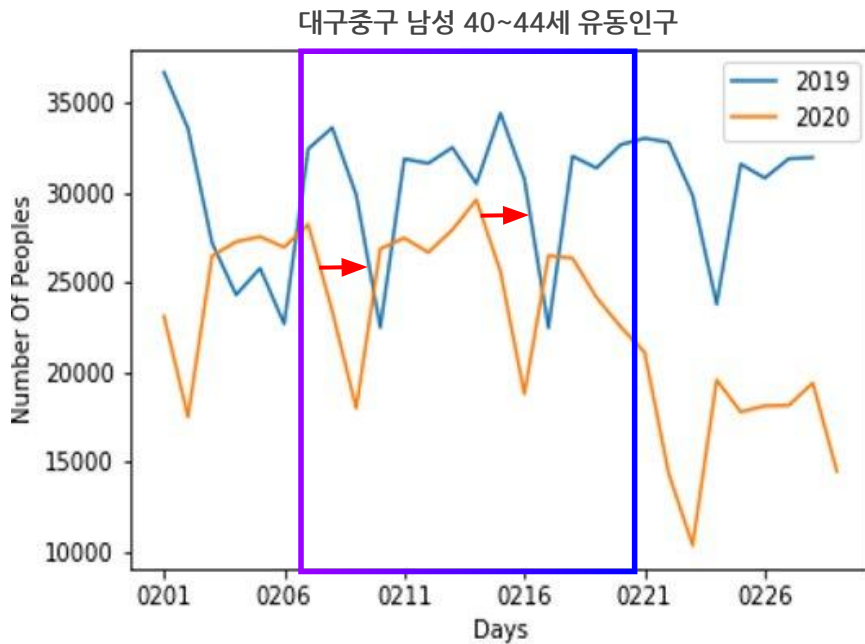
▶ 전년도 비교 시, 우선 동월동일 비교를 시도하였다



# 제1부. 데이터

제공된 데이터 - 유동인구 > 전처리

- ▶ 그래프를 평행이동 시킨 것과 같은 현상이 일어난다



# 제1부. 데이터

제공된 데이터 - 유동인구 > 전처리

▶ 동월동일이 아닌 동주차 동요일 비교를 하였다

2019년 3월

< >

일	월	화	수	목	금	토
24	25	26	27	28	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31	1	2	3	4	5	6

5주차

6주차

7주차

8주차

9주차

10주차

2020년 3월

< >

일	월	화	수	목	금	토
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	1	2	3	4
5	6	7	8	9	10	11

5주차

6주차

7주차

8주차

9주차

10주차

↔

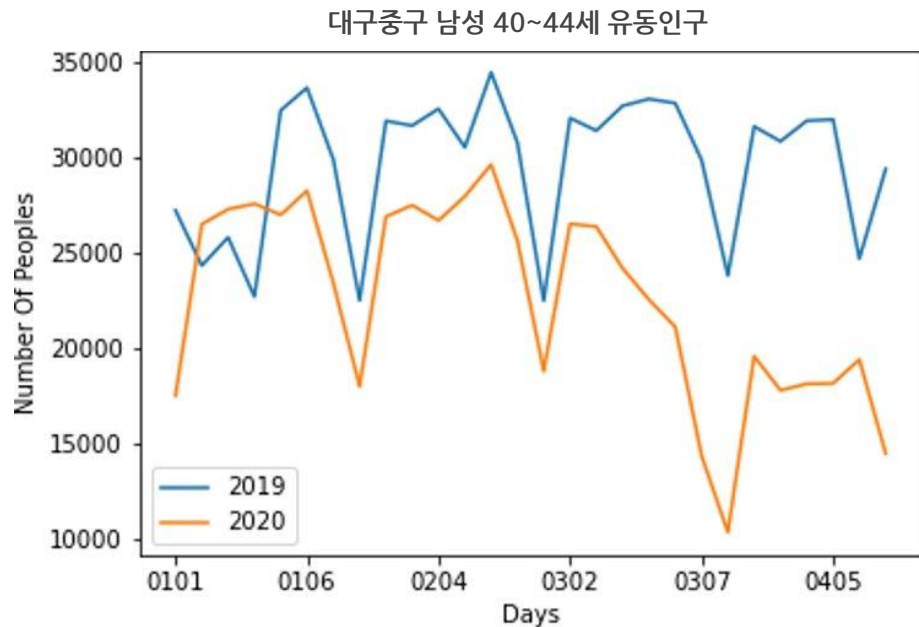
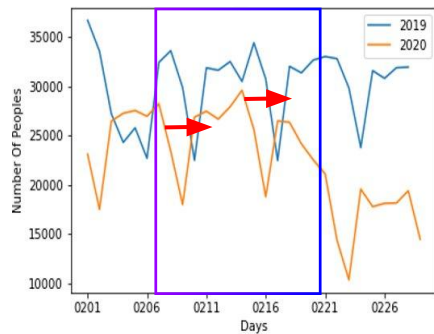
6주차 일요일



# 제1부. 데이터

제공된 데이터 - 유동인구 > 전처리

▶ 동주차 별 비교를 통해 전년 대비 감소세가 더 확실하게 확인된다



# 제1부. 데이터

제공된 데이터 - 유동인구 > 전처리

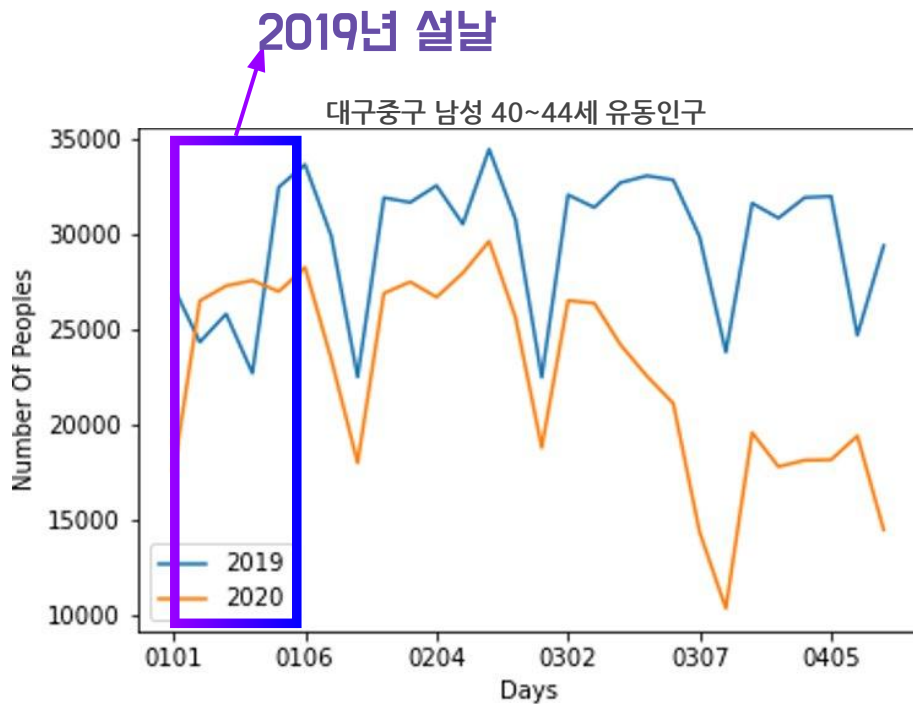
## 날짜 코드 변환

동주차 동요일로 데이터를 비교하기 위해 날짜 코드를 변환한다

# 제1부. 데이터

제공된 데이터 - 유동인구 > 전처리

- ▶ 공휴일로 인해 데이터 간 정확한 비교가 어렵다

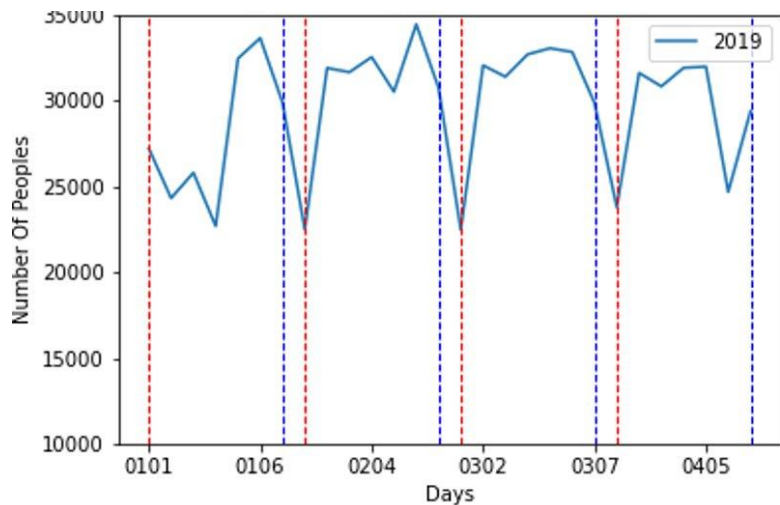


# 제1부. 데이터

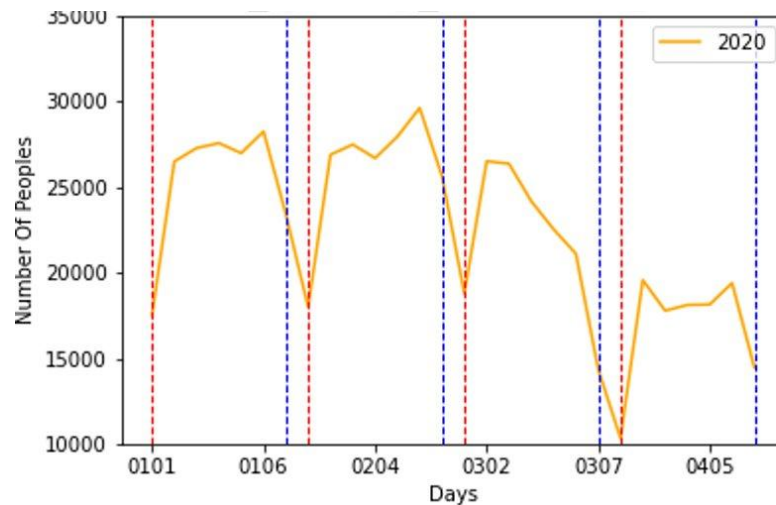
제공된 데이터 - 유동인구 > 전처리

## ▶ 주말과 평일 간 데이터 경향성이 달라 비교가 어렵다

대구중구 남성 40~44세 유동인구 - 2019년



대구중구 남성 40~44세 유동인구 - 2020년



----- 토요일

----- 일요일

# 제1부. 데이터

제공된 데이터 - 유동인구 > 전처리

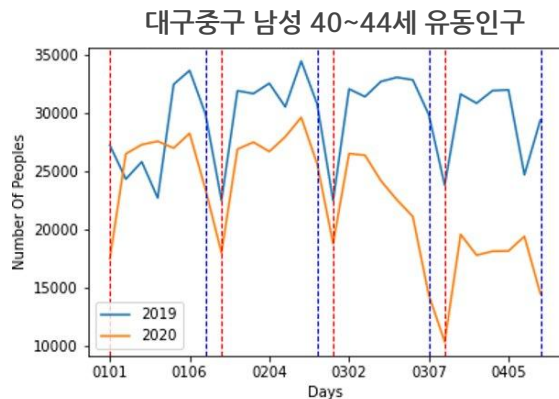
▶ 각 년도의 공휴일을 제거하고, 평일과 주말 데이터를 분리해야한다

	A	B	C	D	E	F
1	원래년도 ≡	원래날짜 ≡	주차인덱스 ≡	명칭 ≡	제거할 주차(2019) ≡	제거할 주차(2020) ≡
2	2019	02-04	0102	설날 연휴	190102	200102
3	2019	02-05	0103	설날 연휴	190103	200103
4	2019	02-06	0104	설날 연휴	190104	200104
5	2019	03-01	0406	3.1절	190406	200406
6	2019	05-01	1302	어린이날 대체	191302	201302
7	2020	02-14	0206	발렌타인데이	190206	200206
8	2020	03-01	0501	3.1절	190501	200501
9	2020	03-14	0607	화이트데이	190607	200607
10	2020	04-15	1104	총선	191104	201104
11	2020	04-30	1305	석가탄신일	191305	201305
12	2020	05-05	1403	어린이날	191403	201403

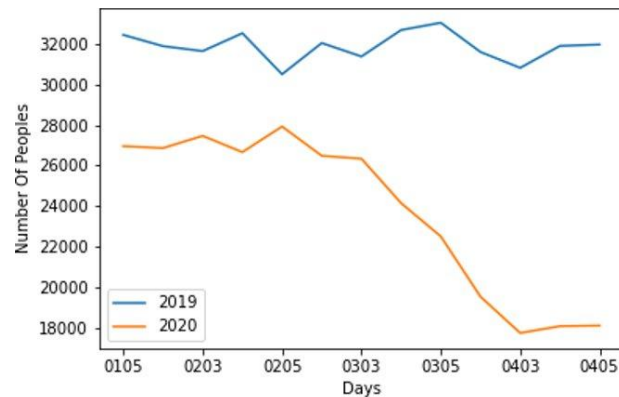
# 제1부. 데이터

제공된 데이터 - 유동인구 > 전처리

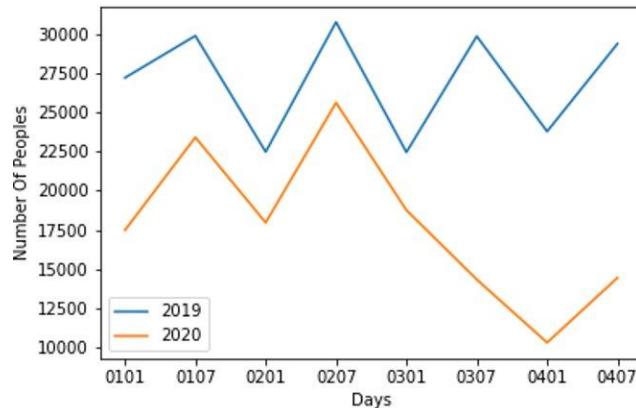
- ▶ 공휴일을 제거하고 평일/주말로 분리한 데이터의 패턴이 더 명확했다



공휴일제거-평일 / 대구중구 남성 40~44세 유동인구



공휴일제거-주말 / 대구중구 남성 40~44세 유동인구



# 제1부. 데이터

---

제공된 데이터 - 유동인구 > 전처리

## 날짜별 이상치 처리

---

공휴일은 제거하고 평일/주말을 분리하여 비교한다

# 제1부. 데이터

## 제공된 데이터

유동인구데이터(SK텔레콤)  
카드매출데이터(신한카드)  
유통데이터(GS리테일)  
물류데이터(CJ올리브네트웍스)

## 코로나 집계

코로나집계데이터(Kaggle)  
코로나키워드 검색률데이터(네이버 데이터랩)

## 재난문자 집계

재난문자 건수데이터(행정안전부)  
재난문자키워드 검색률데이터(네이버 데이터랩)

## 전처리

지역 단위 설정  
날짜 코드 변환  
날짜 이상치 처리

## Feature Engineering

통합 테이블 정의  
Scaled 증가량 계산

## EDA

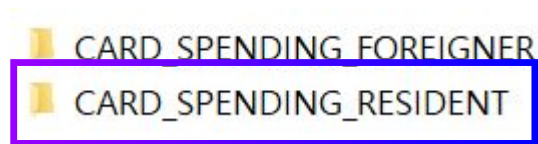
Visualization  
Insight



# 제1부. 데이터

제공된 데이터 - 카드 > 전처리

- ▶ CARD\_SPENDING\_RESIDENT 데이터에 대해 전과 동일하게 전처리를 진행했다



전처리

- 지역 단위 설정
- 날짜 코드 변환
- 날짜 이상치 처리

# 제1부. 데이터

## 제공된 데이터

유동인구데이터(SK텔레콤)  
카드매출데이터(신한카드)  
유통데이터(GS리테일)  
물류데이터(CJ올리브네트웍스)

## 코로나 집계

코로나집계데이터(Kaggle)  
코로나키워드 검색률데이터(네이버 데이터랩)

## 재난문자 집계

재난문자 건수데이터(행정안전부)  
재난문자키워드 검색률데이터(네이버 데이터랩)

## 전처리

지역 단위 설정  
날짜 코드 변환  
날짜 이상치 처리

## Feature Engineering

통합 테이블 정의  
Scaled 증가량 계산

## EDA

Visualization  
Insight

# 제1부. 데이터

제공된 데이터 - 유통 > 전처리

▶ 종합테이블 데이터를 사용했다

OPER_DT	PVN	BOR	ADMD	AMT_IND	LCLS_10_P	LCLS_20_P	LCLS_30_P	LCLS_40_P	LCLS_50_P	LCLS_60_P	LCLS_70_P	LCLS_80_P
20190201	대구광역시	수성구	고산1동	1.06	45.1%	17.2%	29.5%	2.3%	4.5%	0.5%	0.5%	0.4%
20190201	대구광역시	수성구	고산2동	1.08	45.8%	19.6%	29.6%	0.6%	3.4%	0.1%	0.9%	0.1%
20190201	대구광역시	수성구	고산3동	0.86	36.0%	19.7%	37.7%	1.3%	3.8%	0.9%	0.9%	0.6%
◀ ▶ 테이블정의서   참고)구_행정동코드   참고)분석용상품대분류코드   동별매출지수   카테고리별 매출비중   종합테이블(TB_DD_STR_PURCH_CLS) (+)												

동별매출지수 + 카테고리별 매출비중  
= 종합테이블

# 제1부. 데이터

제공된 데이터 - 유통 > 전처리

▶ 매출지수의 카테고리 별 비중을 고려하여 카테고리별 매출 지수로 변환했다

전처리

카테고리별 매출지수 변환

지역 단위 설정

날짜 코드 변환

날짜 이상치 처리

# 제1부. 데이터

제공된 데이터 - 유통 > 전처리

▶ 카테고리별 매출 지수 = 카테고리 별 비중 \* 매출지수

AMT_IN	매출지수
--------	------

LCLS_10_P	
-----------	--

LCLS_20_P	
-----------	--

LCLS_30_P	
-----------	--

LCLS_40_P	
-----------	--

LCLS_50_P	
-----------	--

LCLS_60_P	
-----------	--

LCLS_70_P	
-----------	--

LCLS_80_P	
-----------	--

LCLS_90_P	
-----------	--

카테고리 별 비중



카테고리 별 매출 지수

매출지수	AMT_IN
------	--------

LCLS_10_P * AMT_IN	
--------------------	--

LCLS_20_P * AMT_IN	
--------------------	--

LCLS_30_P * AMT_IN	
--------------------	--

LCLS_40_P * AMT_IN	
--------------------	--

LCLS_50_P * AMT_IN	
--------------------	--

LCLS_60_P * AMT_IN	
--------------------	--

LCLS_70_P * AMT_IN	
--------------------	--

LCLS_80_P * AMT_IN	
--------------------	--

LCLS_90_P * AMT_IN	
--------------------	--

# 제1부. 데이터

제공된 데이터 - 유통 > 전처리

▶ 이전과 동일하게 지역별 통합, 주차별 통합, 날짜별 이상치 분리, 총 3가지 전처리를 진행했다

## 전처리

카테고리별 매출지수 변환

지역 단위 설정

날짜 코드 변환

날짜 이상치 처리

# 제1부. 데이터

## 제공된 데이터

유동인구데이터(SK텔레콤)  
카드매출데이터(신한카드)  
유통데이터(GS리테일)  
물류데이터(CJ올리브네트웍스)

## 코로나 집계

코로나집계데이터(Kaggle)  
코로나키워드 검색률데이터(네이버 데이터랩)

## 재난문자 집계

재난문자 건수데이터(행정안전부)  
재난문자키워드 검색률데이터(네이버 데이터랩)

## 전처리

지역 단위 설정  
날짜 코드 변환  
날짜 이상치 처리

## Feature Engineering

통합 테이블 정의  
Scaled 증가량 계산

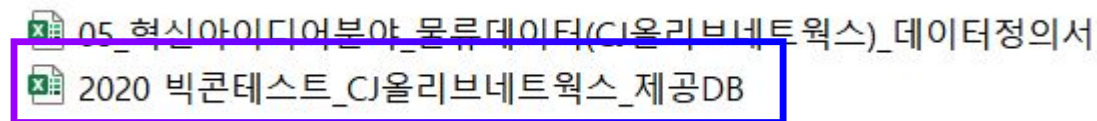
## EDA

Visualization  
Insight

# 제1부. 데이터

제공된 데이터 - 물류 > 전처리

- ▶ 이전과 동일하게 지역 단위 설정, 날짜 코드 변화, 날짜 이상치 처리 총 3가지 전처리를 진행했다



전처리

지역 단위 설정  
날짜 코드 변환  
날짜 이상치 처리



# 제1부. 데이터

## 제공된 데이터

유동인구데이터(SK텔레콤)  
카드매출데이터(신한카드)  
유통데이터(GS리테일)  
물류데이터(CJ올리브네트웍스)

## 코로나 집계

코로나집계데이터(Kaggle)  
코로나키워드 검색률데이터(네이버 데이터랩)

## 재난문자 집계

재난문자 건수데이터(행정안전부)  
재난문자키워드 검색률데이터(네이버 데이터랩)

## 전처리

지역 단위 설정  
날짜 코드 변환  
날짜 이상치 처리

## Feature Engineering

통합 테이블 정의  
Scaled 증가량 계산

## EDA

Visualization  
Insight

▶ 데이터 간 비교를 위해 SUM과 AVG를 포함한 통합 데이터 테이블을 정의했다

테이블 정의서					
테이블 개요		데이터 통합을 위한 전처리 통합 테이블			
순서	컬럼명	물리데이터 타입	설명	값 예시	Not Null
1	date	str	연주일 (주차별) 코드	190101 (19년 1주차 1번째 요일)	예
2	region	str	시-구 명	서울_중구, 서울_노원구, 대구_중구, 대구_수성구	아니오
3	type	str	성별/연령/업종/금액 등	20M, 20Mcnt, 레저용품cnt, 스포츠_레저	예
4	total_value	float	값의 총 합		예
5	avg_value	float	total_value / 지역구별 동 개수		아니오

# 제1부. 데이터

Feature Engineering - 새로운 컬럼 정의

▶ 전년대비 증감률을 계산하였으나, 2019년도 데이터 손실이 크다

	일요일	월요일	화요일	수요일	목요일	금요일	토요일
1주차							
2주차							
3주차							
4주차							
5주차							
6주차							
7주차							
8주차							
9주차							
10주차							
11주차							
12주차							
13주차							
14주차							
15주차							
16주차							
17주차							

증감률

=

2020데이터 - 2019데이터

2019데이터

1187개의 날짜코드 중 677개에 NaN값이 존재

카드데이터 2019년 NaN값

# 제1부. 데이터

Feature Engineering - 새로운 컬럼 정의

- ▶ 증감률이 아닌 증감량을 계산하였으나, 데이터별 범위가 달라 Scaling의 필요성을 느꼈다

Min-Max Scaling

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

$$x = 2020\text{데이터} - 2019\text{데이터}$$

# 제1부. 데이터

---

Feature Engineering - 새로운 컬럼 정의

## Feature Engineering

---

통합 테이블을 정의하고, Scaled 증감량을 계산한다

# 제1부. 데이터

## 제공된 데이터

유동인구데이터(SK텔레콤)  
카드매출데이터(신한카드)  
유통데이터(GS리테일)  
물류데이터(CJ올리브네트웍스)

## 코로나 집계

코로나집계데이터(Kaggle)  
코로나키워드 검색률데이터(네이버 데이터랩)

## 재난문자 집계

재난문자 건수데이터(행정안전부)  
재난문자키워드 검색률데이터(네이버 데이터랩)

## 전처리

지역 단위 설정  
날짜 코드 변환  
날짜 이상치 처리

## Feature Engineering

통합 테이블 정의  
Scaled 증가량 계산

## EDA

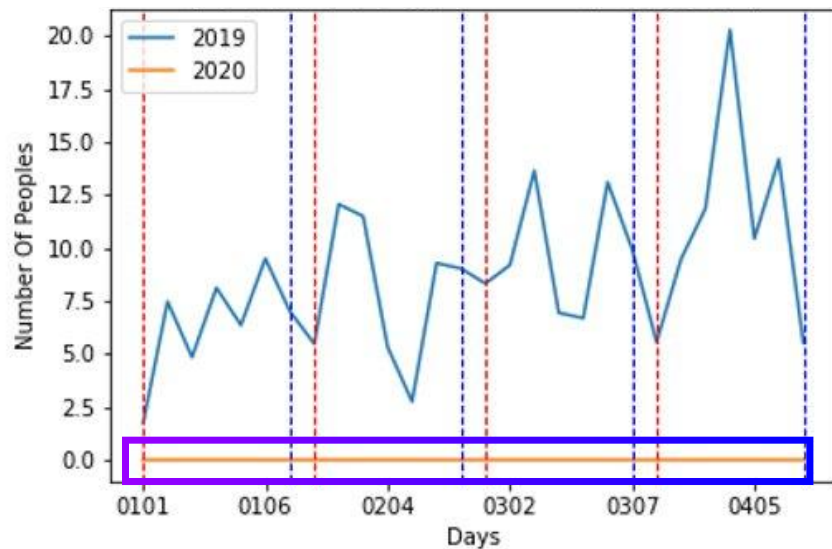
Visualization  
Insight

# 제1부. 데이터

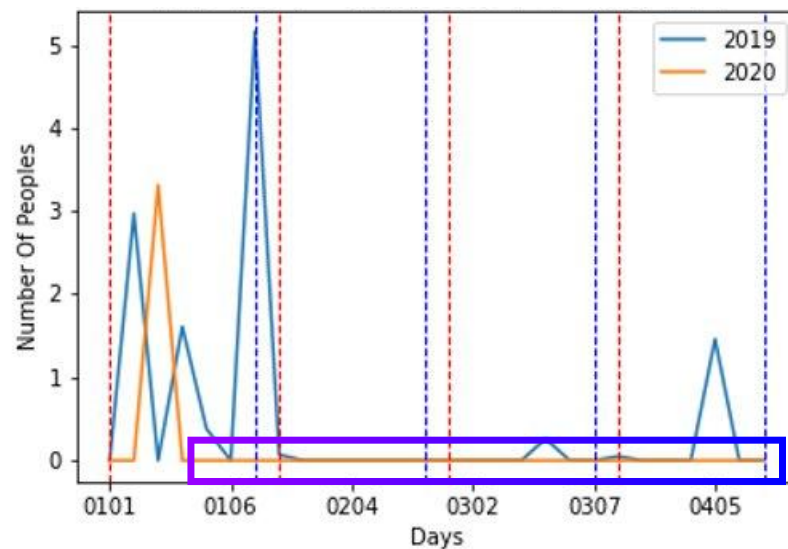
EDA - 유동인구

▶ 0 - 4세 인구의 경우 거의 이동하지 않는다

대구중구 남성 0~4세 유동인구



대구중구 여성 0~4세 유동인구

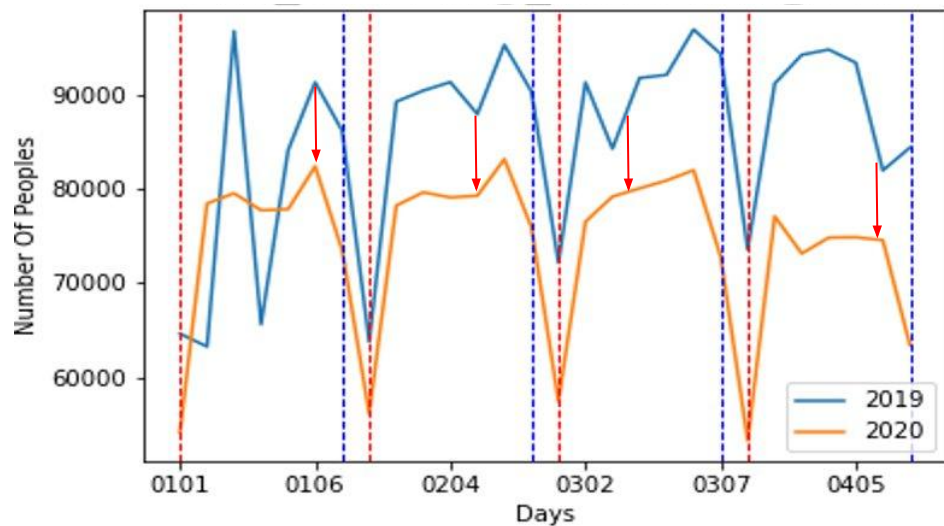


# 제1부. 데이터

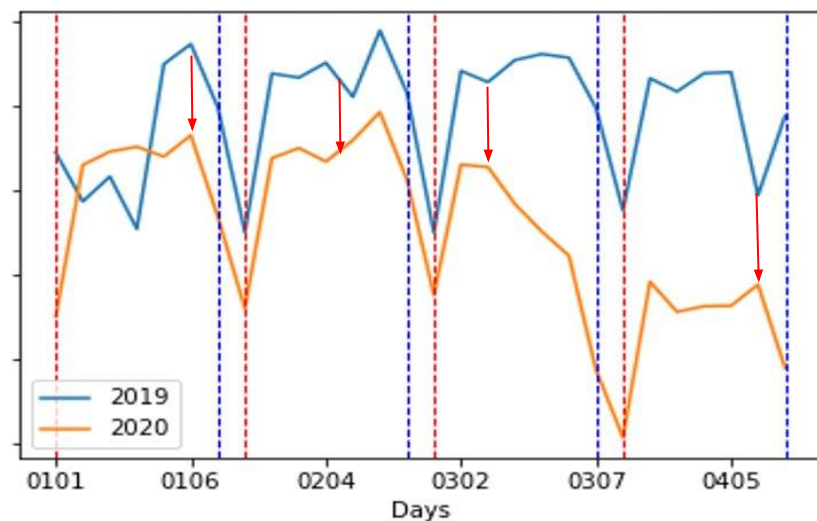
EDA - 유동인구

▶ 유동인구는 전년대비 평균적으로 \_\_ % 감소했다

서울노원구 남성 40~44세 유동인구



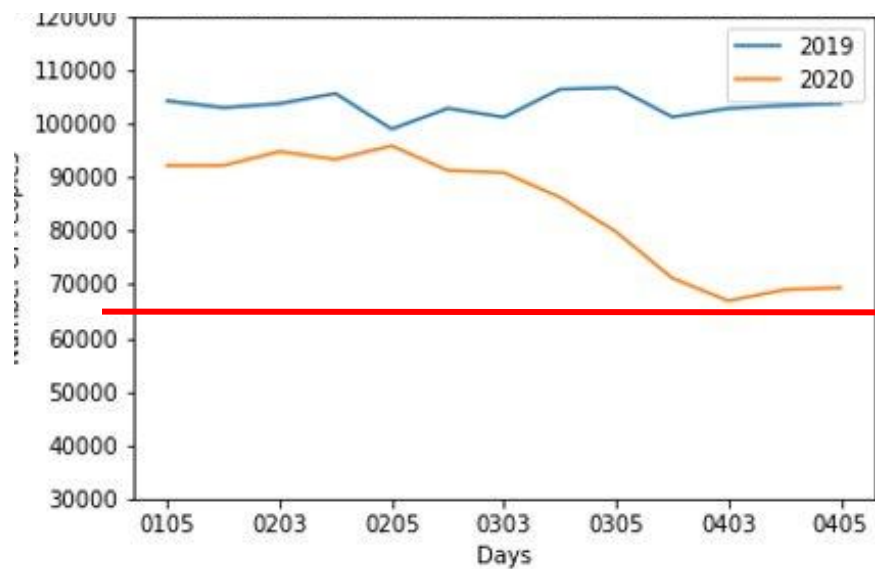
대구중구 남성 40~44세 유동인구



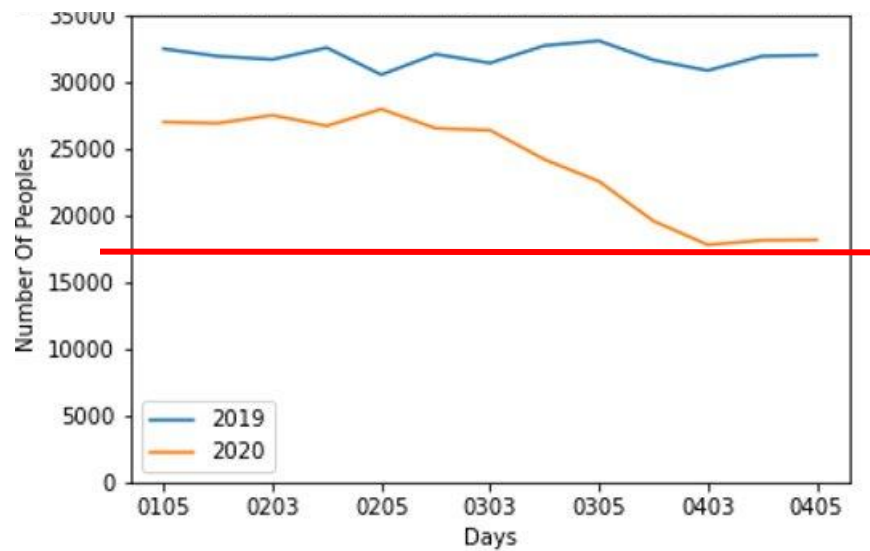


▶ 그러나 코로나 확산 후에도 일정이상의 유동인구는 지속적으로 존재했다

공휴일제거 평일 대구수성구 남성 40~44세 유동인구



공휴일제거 평일 대구중구 남성 40~44세 유동인구

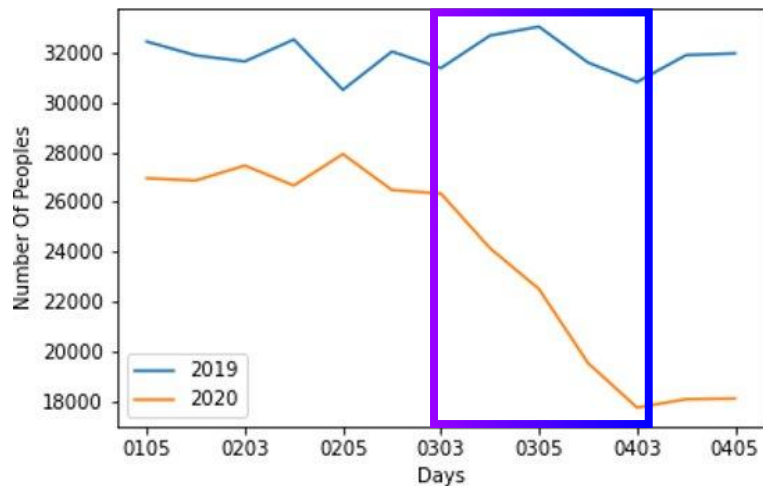


# 제1부. 데이터

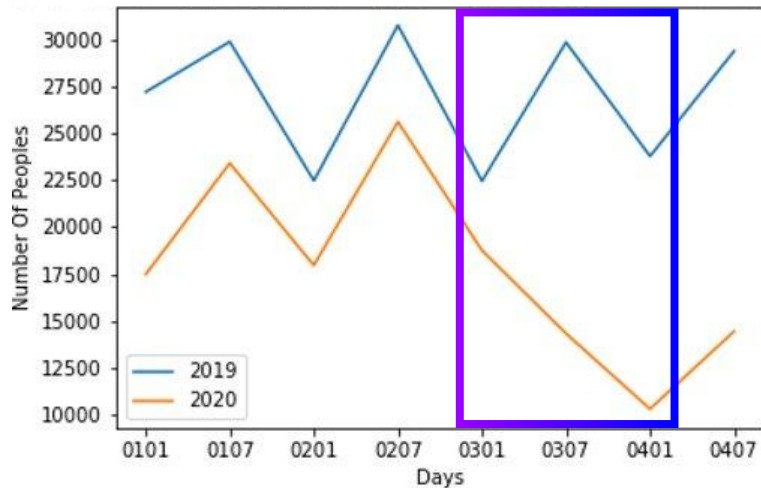
EDA - 유동인구

## ▶ 유동인구가 전년대비 급감하는 시기가 있다

공휴일제거 평일 대구중구 남성 40~44세 유동인구



공휴일제거 주말 대구중구 남성 40~44세 유동인구

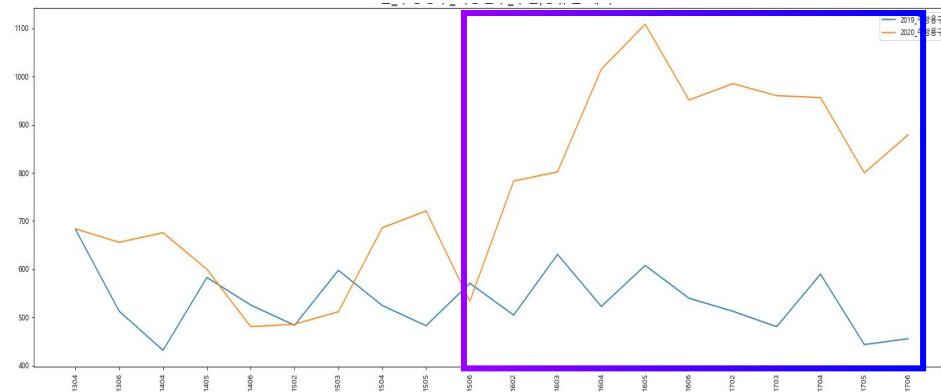


# 제1부. 데이터

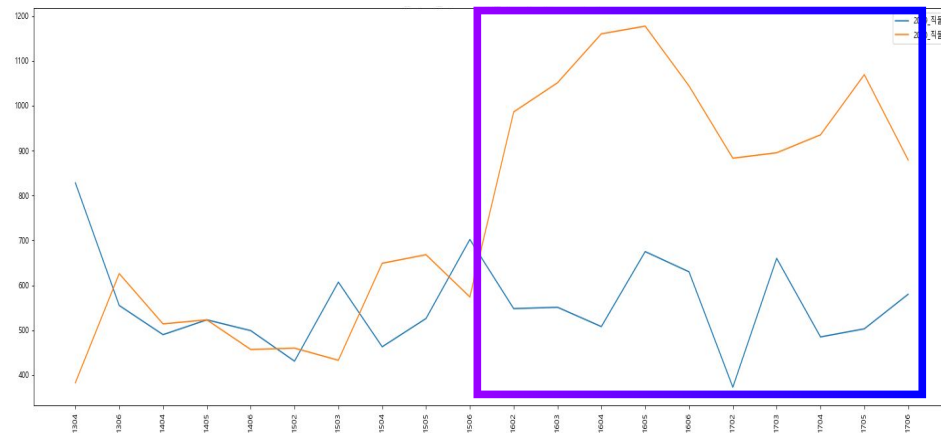
EDA - 카드

- ▶ 주방용구, 식물의 이용건수가 5월달에 급격히 증가했다.

공휴일제거 평일 주방용구 이용건수



공휴일제거 평일 식물 이용건수

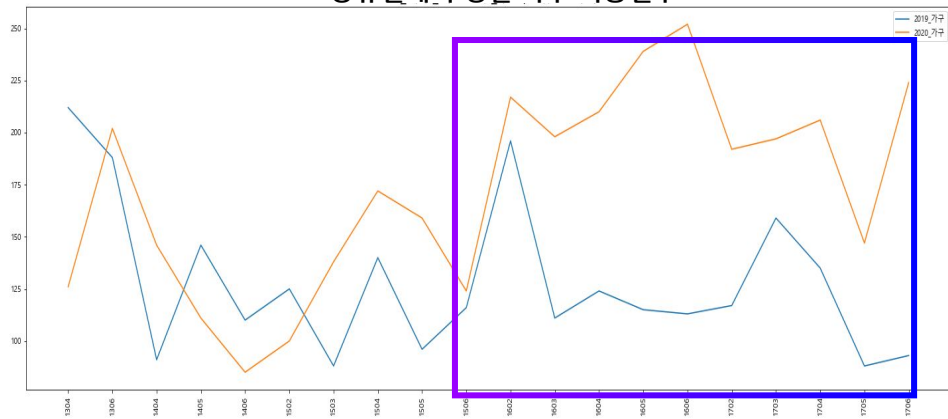


# 제1부. 데이터

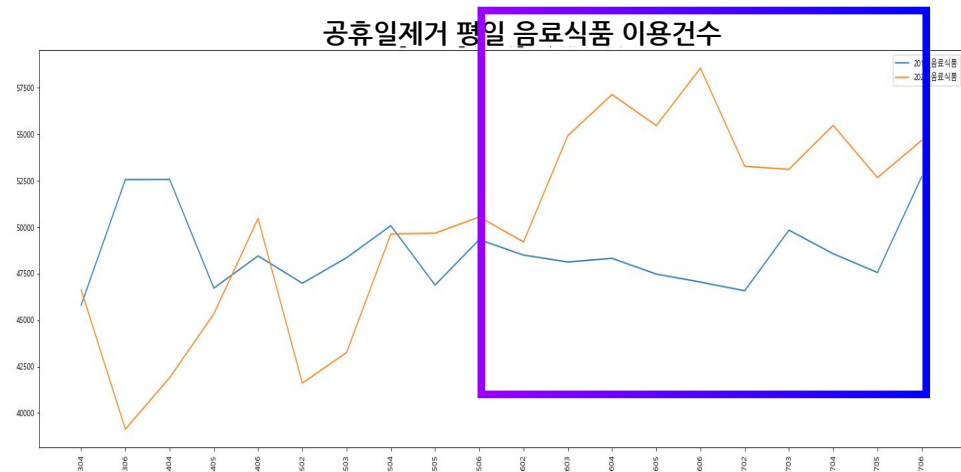
EDA - 카드

- ▶ 가구, 음료 식품의 이용건수가 5월달에 급격히 증가했다.

공휴일제거 평일 가구 이용건수

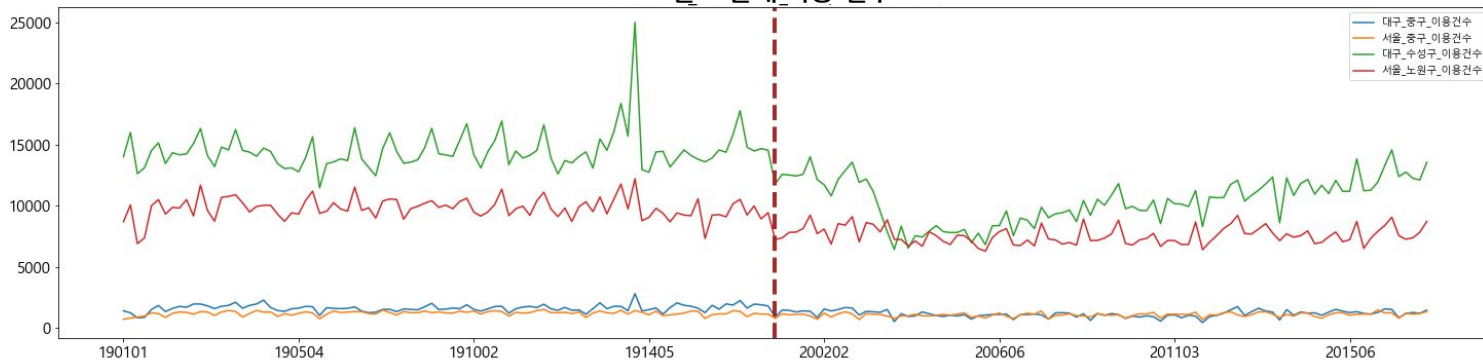


공휴일제거 평일 음료식품 이용건수



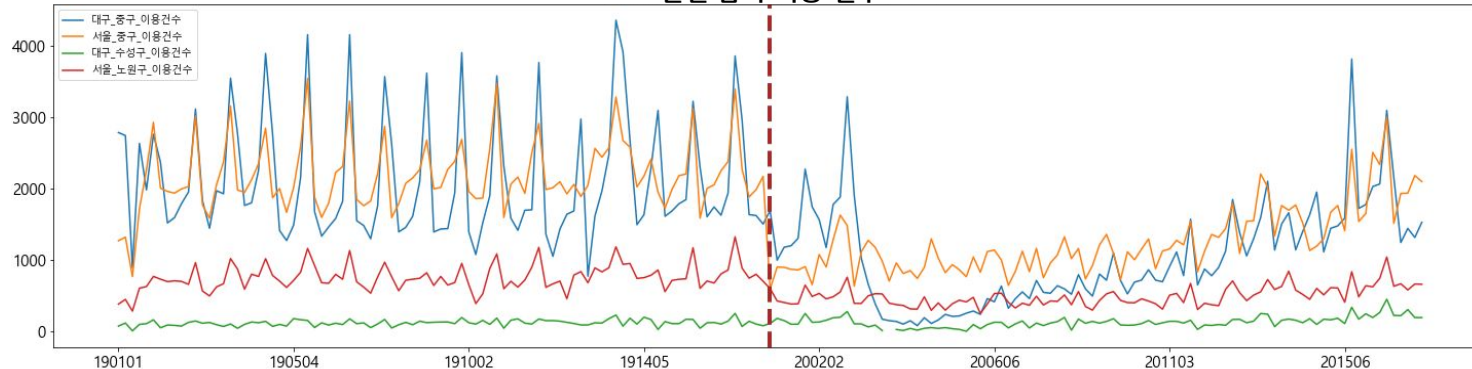
## ▶ 상업지구와 주거지구 간 차이를 보이는 업종이 있음

연료 판매 이용 건수

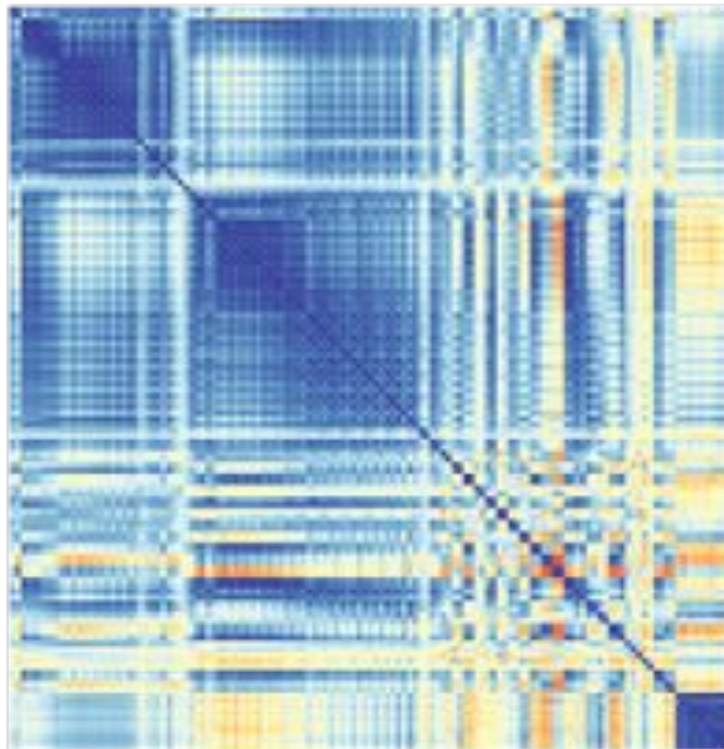


주거지구  
상업지구

신변 잡화 이용 건수



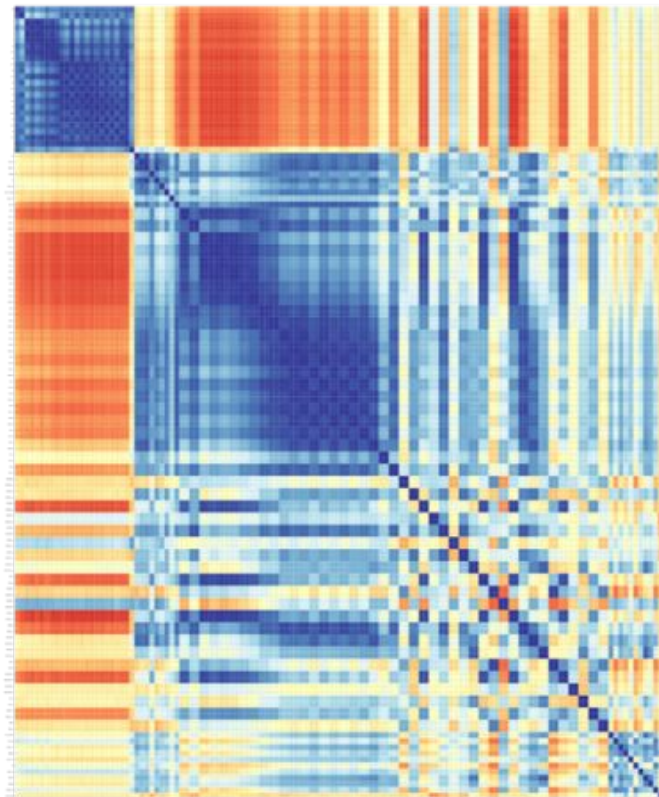
- ▶ 모든 데이터 간 상관관계를 파악하기 위해 Correlation Heat Map을 그렸다



# 제1부. 데이터

## EDA - Correlation

- ▶ 주말만 사용한 데이터가 Correlation이 더 뚜렷하게 나왔다.



# 제1부. 데이터

## EDA - Correlation

- ▶ 유동인구와 비교하였을 때, 대부분의 데이터는 양의 상관 관계를 가지고 있다

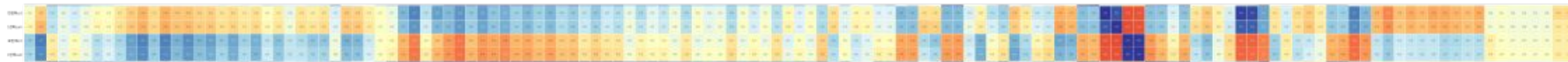




# 제1부. 데이터

## EDA - Correlation

- ▶ 유동인구와 비교하였을 때, 연료 판매와 신변잡화는 약한 음의 상관 관계를 보이고 있다



# 제1부. 데이터

가설1. 제공된 데이터 간 유의미한 상관관계가 있을 것이다

# 제1부. 데이터

가설1. 제공된 데이터 간 유의미한 상관관계가 있을 것이다

데이터 간 상관관계가 있다

— 그러나 서비스를 제안하는 단계에 도달하지 못했다

# 제1부. 데이터

가설2. 제공된 데이터와 코로나 확진자 수, 검색률 등 간에  
유의미한 상관관계가 있을 것이다

# 제1부. 데이터

## 가설2. 제공된 데이터와 코로나 확진자 수, 검색률 등 간에 유의미한 상관관계가 있을 것이다

근거

특정 시점에 유동인구가 크게 감소함  
제공된 데이터는 결국 코로나와 관련된 데이터일 것임

추가 데이터  
수집

코로나집계데이터(Kaggle)  
코로나키워드 검색률데이터(네이버 데이터랩)

검증 방법

데이터 간 상관관계 분석

# 제1부. 데이터

## 제공된 데이터

유동인구데이터(SK텔레콤)  
카드매출데이터(신한카드)  
유통데이터(GS리테일)  
물류데이터(CJ올리브네트웍스)

## 코로나 집계

코로나집계데이터(Kaggle)  
코로나키워드 검색률데이터(네이버 데이터랩)

## 재난문자 집계

재난문자 건수데이터(행정안전부)  
재난문자키워드 검색률데이터(네이버 데이터랩)

## 전처리

지역 단위 설정  
날짜 코드 변환  
날짜 이상치 처리

## Feature Engineering

통합 테이블 정의  
Scaled 증가량 계산

## EDA

Visualization  
Insight

# 제1부. 데이터

코로나 집계 - 코로나 집계

- ▶ 일별 코로나 검사자, 확진자, 완치자, 사망자 수에 대한 데이터를 수집하고, 이전과 동일한 방법으로 전처리를 진행했다



## Data Science for COVID-19 (DS4C)

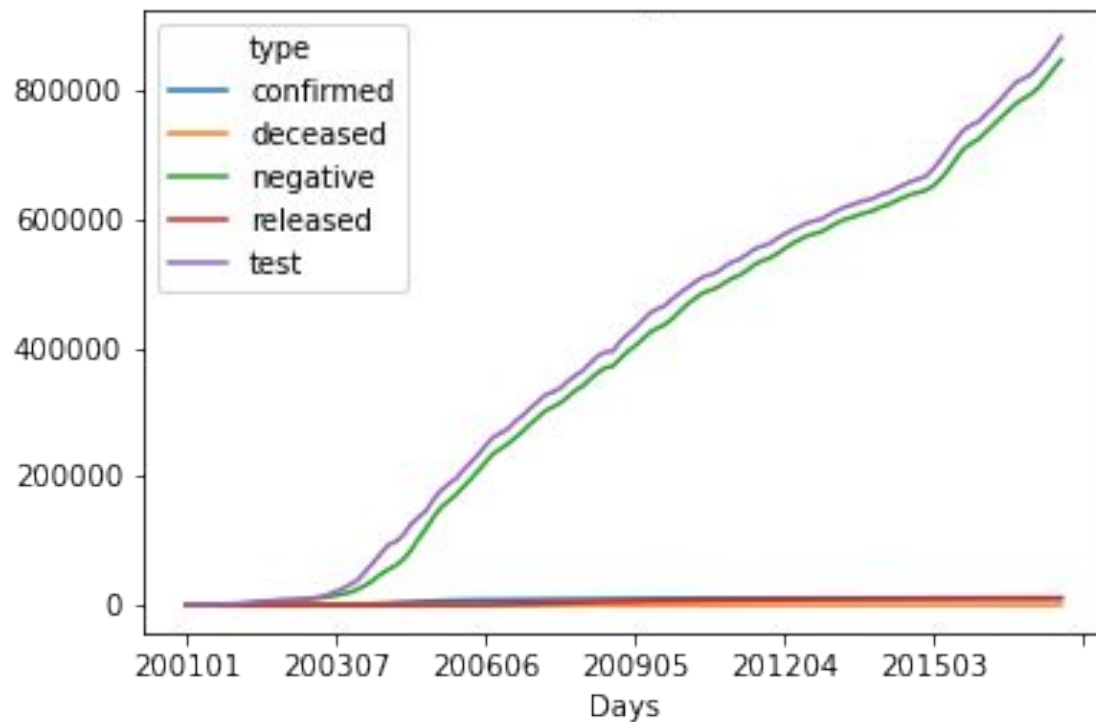
전처리

주차별 통합 - 동일자가 아닌 동주차로 데이터 비교  
날짜별 이상치(공휴일, 평일) 분리  
통합 테이블로 정의

# 제1부. 데이터

코로나 집계 - 코로나 집계

## ▶ 코로나 확진자, 사망자, 결과음성, 검사자 누계 그래프





# 제1부. 데이터

## 제공된 데이터

유동인구데이터(SK텔레콤)  
카드매출데이터(신한카드)  
유통데이터(GS리테일)  
물류데이터(CJ올리브네트웍스)

## 코로나 집계

코로나집계데이터(Kaggle)  
코로나키워드 검색률데이터(네이버 데이터랩)

## 재난문자 집계

재난문자 건수데이터(행정안전부)  
재난문자키워드 검색률데이터(네이버 데이터랩)

## 전처리

지역 단위 설정  
날짜 코드 변환  
날짜 이상치 처리

## Feature Engineering

통합 테이블 정의  
Scaled 증가량 계산

## EDA

Visualization  
Insight

# 제1부. 데이터

코로나 집계 - 코로나키워드 검색률

- ▶ 일별 코로나 관련 키워드의 검색률에 대한 데이터를 수집하고, 이전과 동일한 방법으로 전처리를 진행했다



키워드 : 코로나, corona, 우한 폐렴, Covid-19

전처리

주차별 통합 - 동일자가 아닌 동주차로 데이터 비교  
날짜별 이상치(공휴일, 평일) 분리  
통합 테이블로 정의

# 제1부. 데이터

코로나 집계 - 코로나키워드 검색률

▶ 2020년에 5월로 갈 수록 코로나 키워드 검색률이 줄어들었다.



유동인구, 카드 소비(5월로 갈수록 회복세)



코로나 키워드 검색률(5월로 갈수록 감소세)

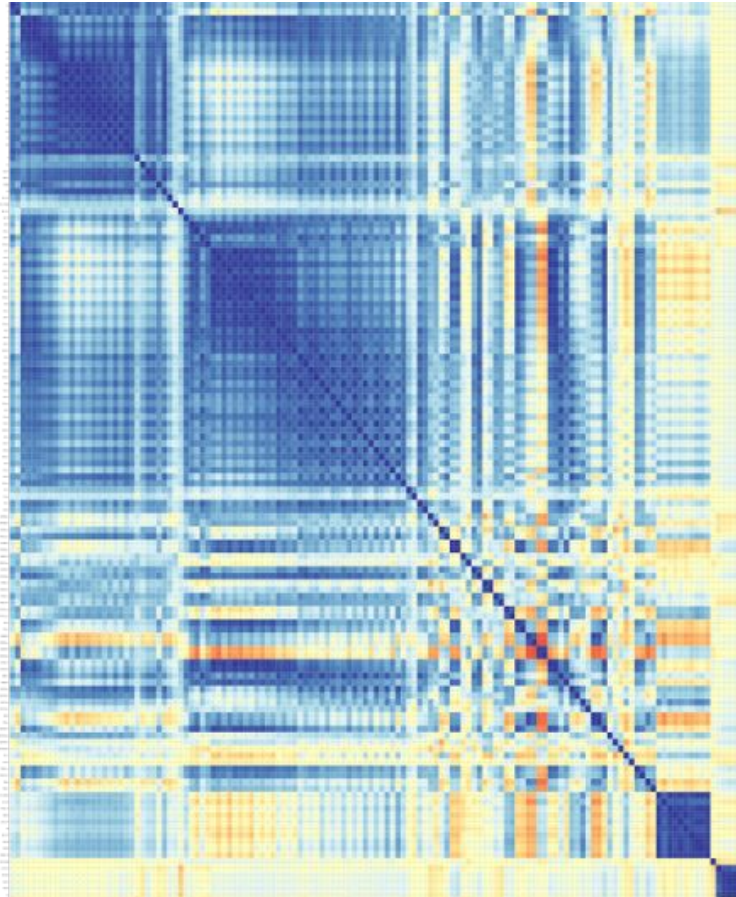


**대중의 경각심 하락**

# 제1부. 데이터

코로나 집계 - Correlation

## ▶ Correlation



# 제1부. 데이터

## 코로나 집계 - Correlation

- ▶ 코로나 데이터와의 상관계수가  $-0.2 \sim 0.2$ , 유의미한 상관관계를 찾을 수 없었다.



# 제1부. 데이터

가설2. 제공된 데이터와 코로나 확진자 수, 검색률 등 간에  
유의미한 상관관계가 있을 것이다

# 제1부. 데이터

가설2. 제공된 데이터와 코로나 확진자 수, 검색률 등 간에  
유의미한 상관관계가 있을 것이다

서비스를 제안할 수 있을 정도의  
\_\_\_\_\_ 유의미한 상관관계를 찾을 수 없다

## 제1부. 테이터

가설3. 사람들은 안전안내문자 수신으로 불편함을 느끼고 있을 것이다



# 제1부. 데이터

## 가설3. 사람들은 안전안내문자 수신으로 불편함을 느끼고 있을 것이다

이유

데이터 자체의 Insight가 아닌 실제 불편함을 고려해봐야할 필요성  
안전안내 문자로 인한 피로감을 느낀 적이 있음

추가 데이터  
수집

안전안내 문자 건수데이터(행정안전부)  
안전안내 문자키워드 검색률데이터(네이버 데이터랩)

검증 방법

안전안내 문자 건수 및 안전안내 문자 차단 관련 검색률 증감량 파악

# 제1부. 데이터

## 제공된 데이터

유동인구데이터(SK텔레콤)  
카드매출데이터(신한카드)  
유통데이터(GS리테일)  
물류데이터(CJ올리브네트웍스)

## 코로나 집계

코로나집계데이터(Kaggle)  
코로나키워드 검색률데이터(네이버 데이터랩)

## 재난문자 집계

안전안내문자 건수데이터(행정안전부)  
안전안내문자키워드 검색률데이터(네이버 데이터랩)

## 전처리

지역 단위 설정  
날짜 코드 변환  
날짜 이상치 처리

## Feature Engineering

통합 테이블 정의  
Scaled 증가량 계산

## EDA

Visualization  
Insight

# 제1부. 데이터

재난문자 집계 - 재난문자 건수

- ▶ OpenAPI를 활용하여 행정안전부의 일자별 안전안내문자 발송 건수 데이터를 수집하고, 이전과 동일한 방법으로 전처리를 진행했다

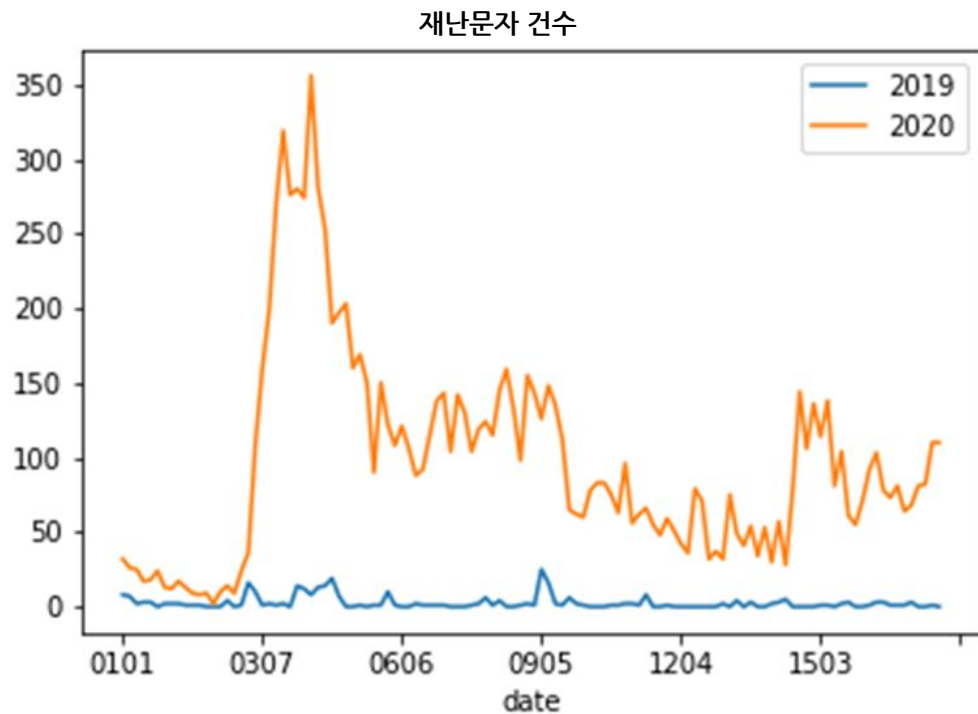


## 안전안내문자방송 발령현황

전처리

주차별 통합 - 동일자가 아닌 동주차로 데이터 비교  
날짜별 이상치(공휴일, 평일) 분리

- ▶ 재난문자 건수가 전년대비 37% 증가하였다



# 제1부. 데이터

## 제공된 데이터

유동인구데이터(SK텔레콤)  
카드매출데이터(신한카드)  
유통데이터(GS리테일)  
물류데이터(CJ올리브네트웍스)

## 코로나 집계

코로나집계데이터(Kaggle)  
코로나키워드 검색률데이터(네이버 데이터랩)

## 재난문자 집계

안전안내문자 건수데이터(행정안전부)  
안전안내문자키워드 검색률데이터(네이버 데이터랩)

## 전처리

지역 단위 설정  
날짜 코드 변환  
날짜 이상치 처리

## Feature Engineering

통합 테이블 정의  
Scaled 증가량 계산

## EDA

Visualization  
Insight

# 제1부. 데이터

재난문자 집계 - 재난문자 키워드 검색률

- ▶ 일별 안전안내문자 관련 키워드의 검색률에 대한 데이터를 수집하고, 이전과 동일한 방법으로 전처리를 진행했다

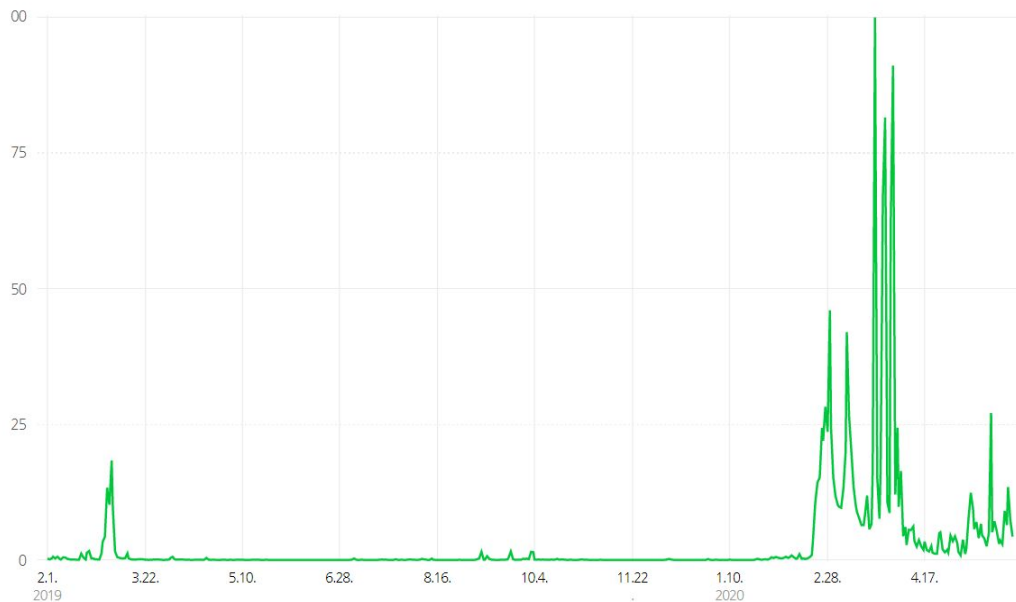


키워드 : 재난문자차단, 안전안내문자차단,  
안전문자차단, 재난안전문자차단, 재난문자안보기

전처리

주차별 통합 - 동일자가 아닌 동주차로 데이터 비교  
날짜별 이상치(공휴일, 평일) 분리

## ▶ 안전안내문자에 대한 불편함을 느꼈다



- 2월~5월 사이 에 검색률은 2019년 대비 2020년에 증가

- 전년대비 다수의 사람이 검색한 일수가 증가

## 제1부. 테이터

가설3. 사람들은 안전안내문자 수신으로 불편함을 느끼고 있을 것이다



# 제1부. 테이터

가설3. 사람들은 안전안내문자 수신으로 불편함을 느끼고 있을 것이다

— 실제로 불편함을 느끼고 있다

# 제1부. 테이터

안전안내문자의 불편함을 해소할 수 있는  
서비스를 개발할 필요성이 있다

# 제2부. 서비스

---

서비스 제안

---

우리의 서비스는?

뉴노멀시대를 대비한 마이데이터 활용 **개인 방역 강화** 서비스

안전재난 문자의 불편함 해소 위한 **동선 겹침 알림** 서비스

# 제2부. 서비스

서비스의 필요성

기존 서비스 분석

새로운 서비스 제안

## 제2부. 서비스

### 서비스의 필요성

- ▶ 감염병은 주기적으로 유행했고, 따라서 코로나 종식 이후에도 새로운 감염병 유행이 예상된다

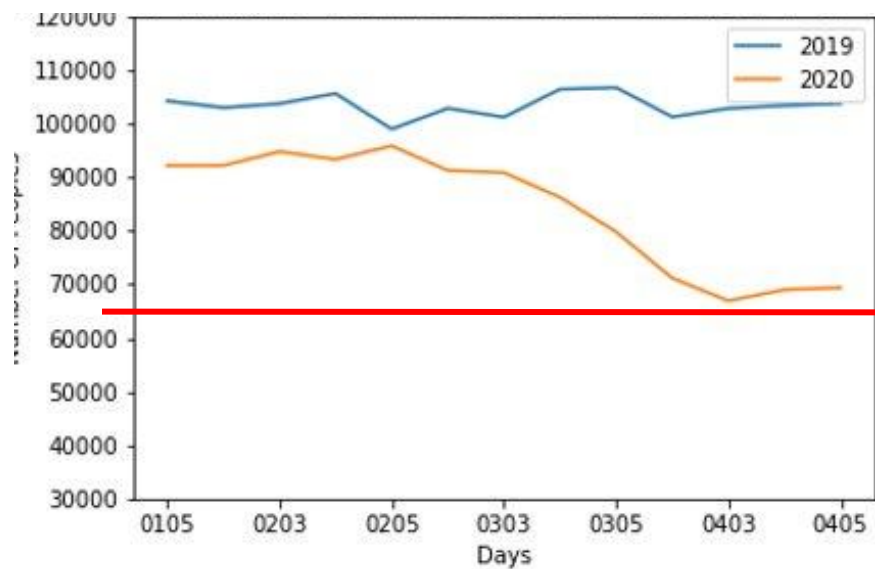


## 제2부. 서비스

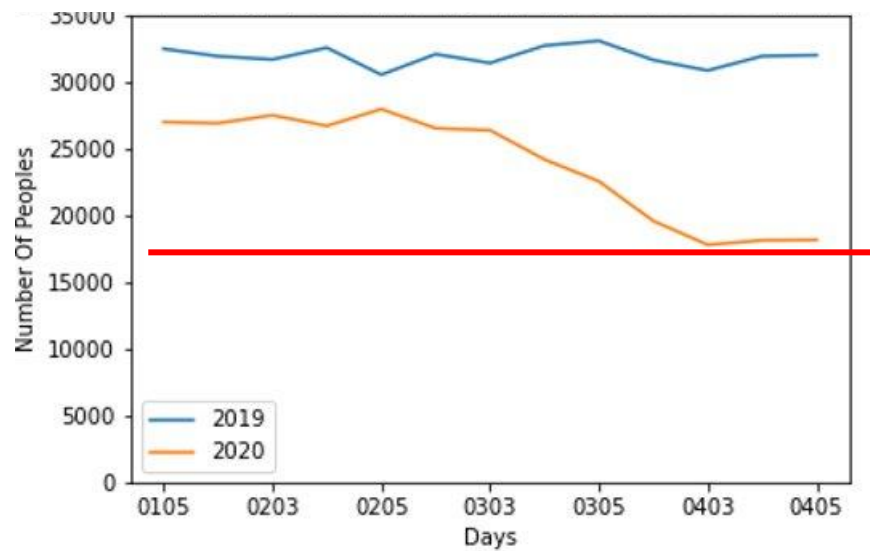
### 서비스의 필요성

- ▶ 감염병 유행 상황에서도 꼭 나가야 하는 사람들을 위한 개인 맞춤형 방역 방안의 필요성이 있다

공휴일제거 평일 대구수성구 남성 40~44세 유동인구



공휴일제거 평일 대구중구 남성 40~44세 유동인구



## 제2부. 서비스

### 서비스의 필요성

- ▶ 가장 활발히 이용되고 있는 방역 서비스는 안전 안내 문자 서비스이나, 여러가지 불편함이 있다

## PAIN POINT

안전 안내 문자 수신량 증가로 인한 불편함

텍스트만으로 이루어진 확진자 정보 확인에 대한 불편함

사용자의 동선을 사용자가 오롯이 기억하고 있어야 하는 불편함

지역 이동 시, 과거에 있었던 지역의 정보를 수신할 수 없음



- ▶ 가장 활발히 이용되고 있는 방역 서비스는 안전 안내 문자 서비스이나, 여러가지 불편함이 있다

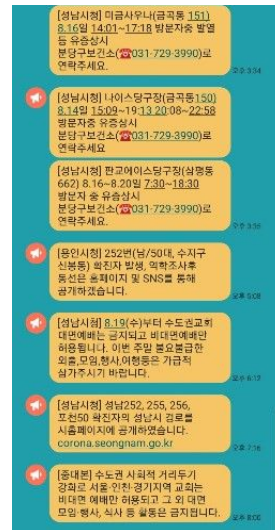
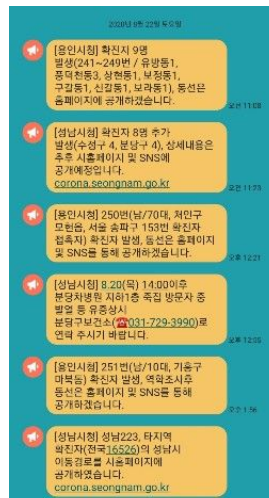
## PAIN POINT

안전 안내 문자 수신량 증가로 인한 불편함

텍스트만으로 이루어진 확진자 정보 확인에 대한 불편함

사용자의 동선을 사용자가 오롯이 기억하고 있어야 하는 불편함

지역 이동 시, 과거에 있었던 지역의 정보를 수신할 수 없음



2020년 하루에 수신한 안전안내문자 목록

▶ 가장 활발히 이용되고 있는 방역 서비스는 안전 안내 문자 서비스이나, 여러가지 불편함이 있다

PAIN POINT

안전 안내 문자 수신량 증가로 인한 불편함

텍스트만으로 이루어진 확진자 정보 확인에 대한 불편함

사용자의 동선을 사용자가 오롯이 기억하고 있어야 하는 불편함

지역 이동 시, 과거에 있었던 지역의 정보를 수신할 수 없음

확진환자 이동경로

장소유형	상호명	주소	노출일시	접촉자	소속여부
체육시설	시범한양아파트 단지 내 테니스장	분당구 중앙공원로 17 단지 내 테니스장	9.11(금) 10:17~12:30 9.15(화) 10:10~13:00	7명	소속완료
역국	메디컬은누리약국	분당구 동판교로 61	9.14(월) 12:20~12:22	2명	소속완료
오토랜드	넥센타이어오토랜드	분당구 분당로311번길 12	9.14(월) 16:24~16:25	1명	소속완료
병원	세봉이비인후과의원	분당구 운중로 239	9.15(화) 09:39~09:44	1명	소속완료
피부관리	서울피부원을 판교점	분당구 운중로 237	9.15(화) 15:00~16:00	1명	소속완료
병원	백문순 치과의원	중원구 여수동 205 신아람푸르지오시티1 2층	9.10(목) 14:37~15:32 9.17(목) 14:44~15:15	4명	소속완료

※ 9. 11(금) ~ 9.15(화) 「서현동 한양아파트(분당구 중앙공원로 17) 단지내 테니스장, 병문자(백문순, 동호회 활동)는 백문순입내에 가까운 보건소를 방문하여 코로나19 진단검사를 받으시기 바랍니다.

※ 9. 24(목) 21:00 ~ 9. 25(금) 06:00 스타트렉 오학실(수정구 산성대로255번길 8-4) 코인노래방 중원부스 이용객은 수정구보건소(☎031-729-3880)로 연락주시기 바랍니다.

※ 분당 아람동성남(분당구 매화로 35) 9. 13(월), 20(월), 11시 미사, 9. 24(목) 07시 장례미사 참여자 중 발열 및 기침 등 의심증상 발현시 가까운 보건소에서 코로나19 진단검사를 받으시기 바랍니다.

- \* 이동경로 및 방문장소는 모두 방역소독 완료하였으니 안심하고 이용하셔도 됩니다.
- \* 우러시는 확진자의 이동경로 등 정보공개 안내 제3판(2020.7.1. 시행)에 따라 확진환자의 개인을 특정하는 정보를 공개하지 않음
- \* 이동경로 공개는 증상 발현 또는 증상이 확인되지 않은 경우 감제 재취 2일전부터 격리일까지 공개(코로나19 대응지침(지자체용) 7-4 단 개정)
- \* 확진자가 마지막 접촉자와 접촉한 날로부터 14일 경과 시, 이동경로에 대한 부분은 공개되지 않음(확진환자의 이동경로 등 정보공개 안내제3판)
- \* 발생 상황 정보를 신속·투명의 알리기 위해 위 명단을 공개하여, 추가 역학조사 결과 등에 따라 수정 및 보완될 수 있음

텍스트로만 표현된 확진자 동선 정보

## 제2부. 서비스

### 서비스의 필요성

- ▶ 가장 활발히 이용되고 있는 방역 서비스는 안전 안내 문자 서비스이나, 여러가지 불편함이 있다

## PAIN POINT

안전 안내 문자 수신량 증가로 인한 불편함

텍스트만으로 이루어진 확진자 정보 확인에 대한 불편함

사용자의 동선을 사용자가 오롯이 기억하고 있어야 하는 불편함

지역 이동 시, 과거에 있었던 지역의 정보를 수신할 수 없음

- ▶ 가장 활발히 이용되고 있는 방역 서비스는 안전 안내 문자 서비스이나, 여러가지 불편함이 있다

### PAIN POINT

안전 안내 문자 수신량 증가로 인한 불편함

텍스트만으로 이루어진 확진자 정보 확인에 대한 불편함

사용자의 동선을 사용자가 오롯이 기억하고 있어야 하는 불편함

지역 이동 시, 과거에 있었던 지역의 정보를 수신할 수 없음

## 제2부. 서비스

---

### 서비스의 필요성

#### 서비스 필요성

---

뉴 노멀 시대에 어울리는 새로운 개인화 방역 서비스가 필요하다

# 제2부. 서비스

서비스의 필요성

기존 서비스 분석

새로운 서비스 제안

## 제2부. 서비스

### 기존 서비스 분석

- ▶ 코로나 맵과 안전디딤돌 서비스에는 장점도 있지만, 한계가 존재한다



안전디딤돌



코로나맵

#### 장점

사용자가 안전안내문자를 수신받을 지역(지자체)를 설정할 수 있다

지도를 이용하여 어느 지역에서 언제 확진이 발생했는지 확인할 수 있다

#### 한계

사용자 위치가 아닌 사용자가 설정한 지역의 문자만 받을 수 있다

서비스에서 동선비교 기능을 제공하지 않아 사용자가 직접확인해야한다

- ▶ 마이데이터와 기존데이터를 결합하여 새로운 서비스를 제공해 이러한 한계를 극복할 수 있다

### 1. 기존 서비스의 장점 응용

지도 데이터에 GPS를 이용하여, 효과적으로 시각화  
원하는 범위를 선택 가능하게 하여 서비스의 질 향상

### 2. 기존 서비스의 한계 극복

마이데이터와 기존 데이터들을 결합하여 새로운 서비스 제공  
확진자와의 동선 겹침 확인 등 사용자가 직접 수행해야했던 일을 자동화하여 편리함을 추구한다



# 제2부. 서비스

서비스의 필요성

기존 서비스 분석

새로운 서비스 제안

# 오늘의 방역 알리미

재난문자의 불편함 해소 및 개인 방역 강화를 위한 동선 겹침 알림

# 제2부. 서비스

새로운 서비스 제안

- ▶ 공공데이터와 마이데이터를 결합하여 사용자와 확진자의 동선을 파악한다

## 사용 데이터

공공데이터 | 각 지자체에서 공개하는 확진자 동선 데이터 > 확진자 동선 파악

마이데이터 | GPS 데이터  
카드 사용 이력 데이터 > 사용자 동선 파악  
QR 체크인 데이터

▶ 공공데이터와 마이데이터를 결합하여 사용자와 확진자의 동선을 파악한다

## 사용 데이터

공공데이터 | 각 지자체에서 공개하는 확진자 동선 데이터 > 예시 : 성남시청 공개 데이터

장소유형	상호명	주소	노출일자
체육시설	시범한양아파트 단지 내 테니스장	분당구 중앙공원로 17 단지 내 테니스장	9.11(금) 10:17~12:30 9.15(화) 10:10~13:00
병원	새봄이비인후과의원	분당구 운중로 239	9.15(화) 09:39~09:44
피부관리	서윤피부맑음 판교점	분당구 운중로 237	9.15(화) 15:00~16:00
병원	바른손 치과의원	중원구 여수동 205 신야탑푸르지오시티1 2층	9.10(목) 14:37~15:32 9.17(목) 14:44~15:15

- ▶ 공공데이터와 마이데이터를 결합하여 사용자와 확진자의 동선을 파악한다

## 사용 데이터

공공데이터 | 각 지자체에서 공개하는 확진자 동선 데이터

마이데이터

GPS 데이터

> SK Telecom 등 통신사

카드 사용 이력 데이터

> 신한카드 등 카드사

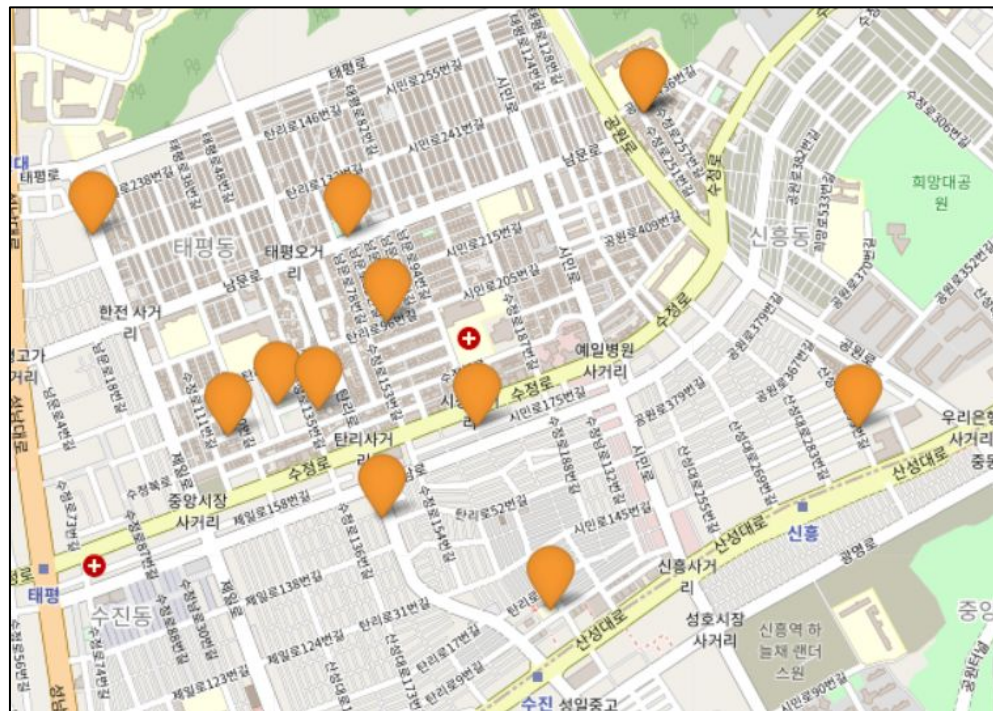
QR 체크인 데이터

> 카카오, 네이버

## 제2부. 서비스

새로운 서비스 제안

### ▶ 확진자 동선 정보 수집 후 시각화



확진자

## 제2부. 서비스

### 새로운 서비스 제안

#### ▶ 주요 기능 1 > 사용자와 확진자의 동선 겹침을 확인할 수 있다



나의 동선



내 동선과 겹친 확진자



확진자 동선정보(위치명, 시간)



### ▶ 주요 기능 2 > 사용자에게 필요한 정보 업데이트 시 푸시 알림을 통해 알려준다

내 동선과 겹쳤던 확진자 정보 업데이트 시 알림

내 오프라인 소비패턴 관련 방역 방침 알림

Example>

[확진자와 동선겹침]

2020/09/12 12:00~13:27 성남중앙지하도상가





# 오늘의 방역 알리미

재난문자의 불편함 해소 및 개인 방역 강화를 위한 동선 겹침 알림

---

## 감사합니다