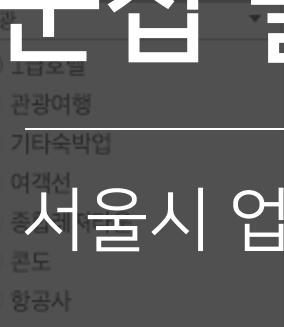


Corona Virus Disease 2019

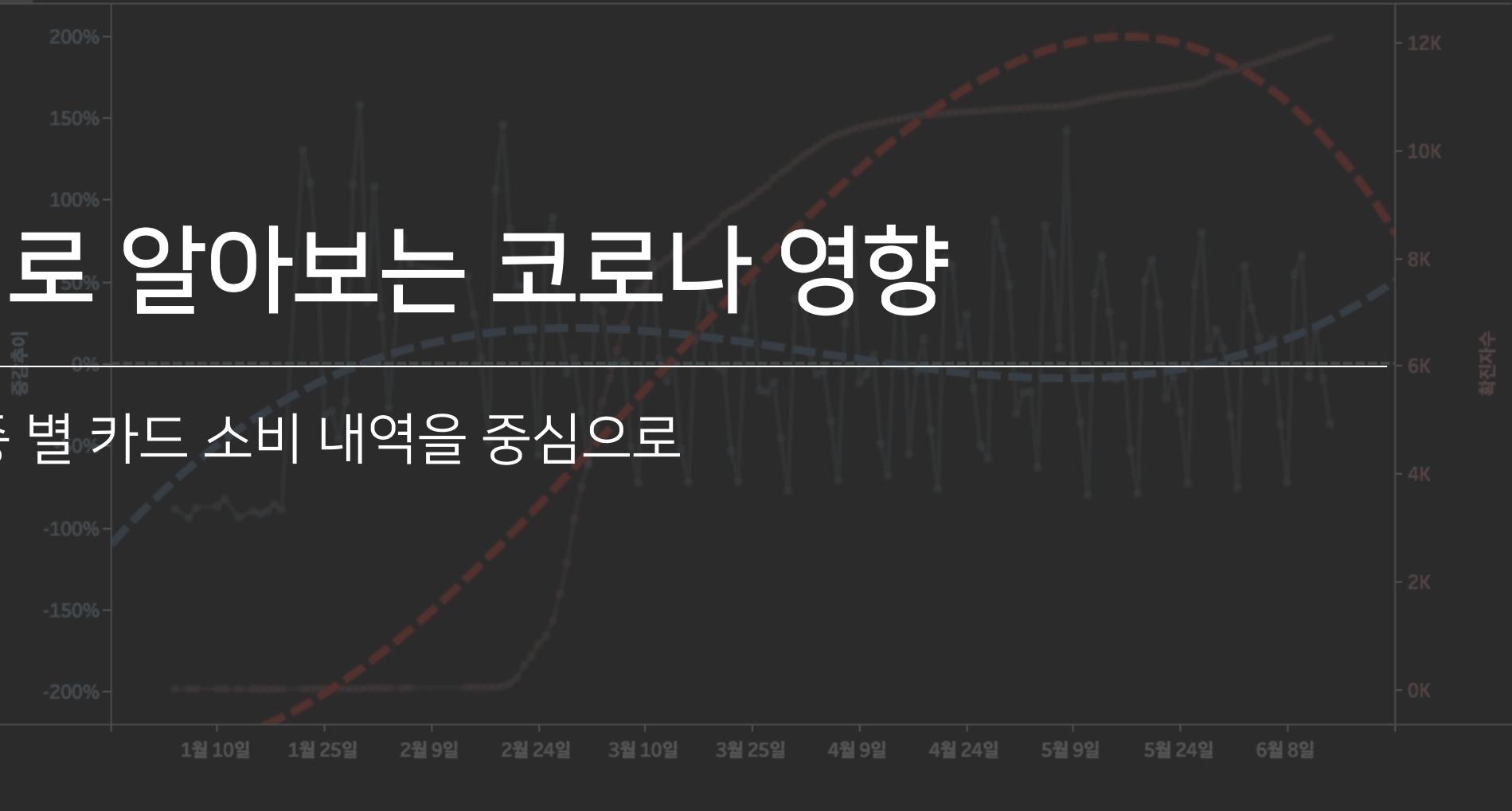
포스트 코로나 데이터 시각화 경진대회

군집 별로 알아보는 코로나 영향



Our visualizations introduce clusters to identify the effects of COVID. Through the results, you can see the consumption pattern. It is predictable and understandable for industries that changes.

Check the details of the mouse over the chart



EXPORT IMAGE

GOAT Tableau Viz

Data Source Dacon, Seoul Open Data, Seoul TOPIS

Periods 2020. 01 - 2020. 06

이종호 | 박은빈

Date of Visualization 2020. 07. 31.

Table of Contents

1. 과제 목적 및 현황

- 1-1. 내부데이터
- 1-2. 외부데이터

2. 데이터 분석 및 시각화 방법

- 2-1. 군집화
- 2-2. 변화추이 계산 방법

3. 인사이트 분석

- 3-1. 해석에 필요한 주요 이슈들
- 3-2. 군집별 시각화 해석

4. 한계 및 추가 연구 내용

- 4-1. 시각화의 한계점
- 4-2. 추가 연구 내용

과제 목적 및 현황

과제 목적 및 현황

포스트 코로나 시대 유의미한 인사이트 도출과 빅데이터 기반 생태계 활성화

- 데이터 기간 : 2020년 01월 - 2020년 06월

내부 데이터

- 업종 별 결제금액 데이터
 - 가맹점 업종명과 매출발생건수만을 차출
- 코로나 바이러스 확진자 수 데이터

외부 데이터

- 서울시 열린 데이터 광장
 - 버스 및 지하철 데이터
- 서울 TOPIS
 - 교통량 데이터

데이터 분석 및 시각화 방법

데이터 분석 및 시각화 방법

군집화

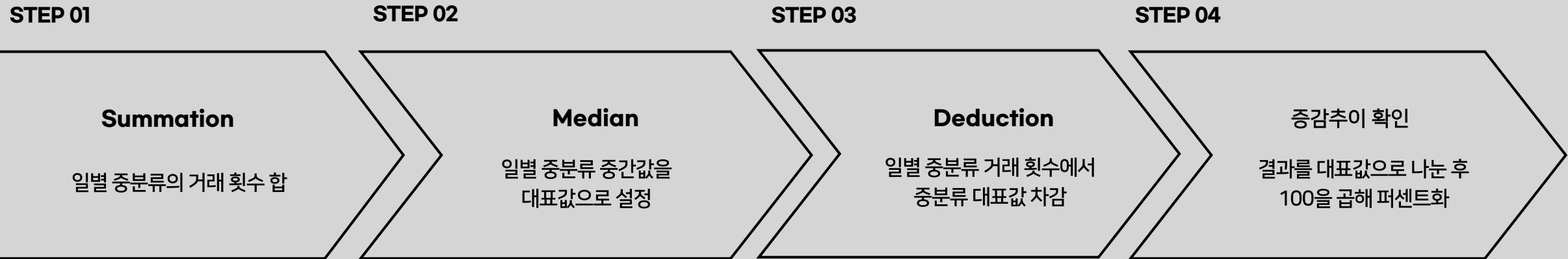
중분류
버스
노래방
유치원
1급 호텔
정장
⋮



중분류	대분류
버스	대중교통
노래방	실내활동
유치원	교육
1급 호텔	관광
정장	의류
⋮	

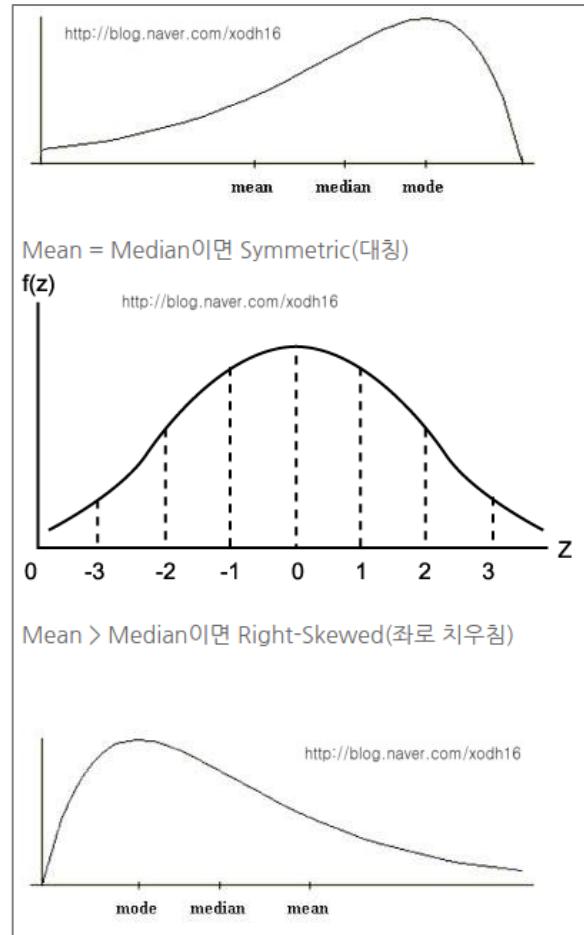
데이터 분석 및 시각화 방법

변화추이 계산 방법



데이터 분석 및 시각화 방법

대표값 산출 방법



- 표본의 편차 혹은 왜곡이 심하게 나타나는 경우 **중앙값이 유용합니다.**
 - 현재 데이터는 2020년 01월부터 06월 중순까지의 데이터로 데이터가 적습니다.
 - 코로나라는 큰 변수가 있기에 거래 횟수가 급격히 증가하거나 감소하는 경우가 많습니다.
-

많이 이용하는 평균값을 대표값으로 사용하는 것보다
중앙값을 **대표값**으로 사용하는 것이 좋다고 판단하였습니다.

인사이트 분석

인사이트 분석

주요 이슈

KBS NEWS

분야별 ▾ 시사·다큐 ▾

뉴스9 취재K 정보K 글로벌K 스포츠 크랩 #안전 대한민국

‘코로나19’ 팬데믹
[코로나19-서울] 서울시 유흥업소 19일까지 집합금지 - 4월 8일 11시 브리핑

입력 2020.04.08 (14:09) 케이아

뉴스룸 | 쇠인기자

코로나19에 예체능학원 '고사 직전'..."원생 50명중 1명 온다"

송고시간 | 2020-03-23 06:15



임성호 기자

| 입시와 직접 관련 없고 온라인 교습도 불가…감염 걱정에 학생 발길 '뚝'

HOME > 소식 > 행정 > 포토뉴스

코로나 경보 '심각' 격상, 학교 개학 전격 연기

곽우석 기자 | 승인 2020.02.23 18:58 | 댓글 0

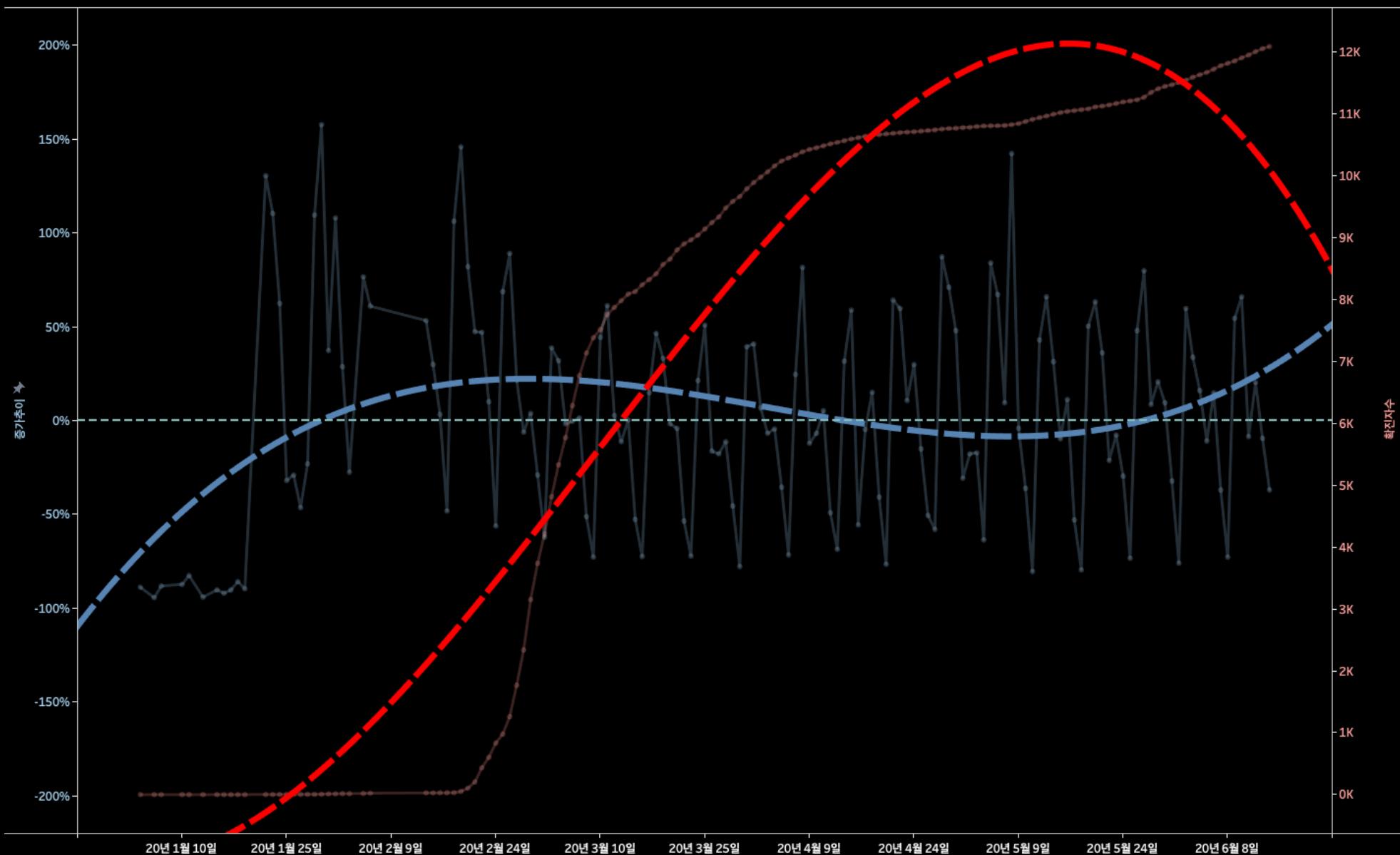
정부, 코로나19 위기 대응경보 단계 '경계'→'심각' 조처
유치원·초·중·고 개학 3월 2일→9일 연기...상황 따라 추가 연기도 검토

군집 별 시각화 해석

관광

1급 호텔

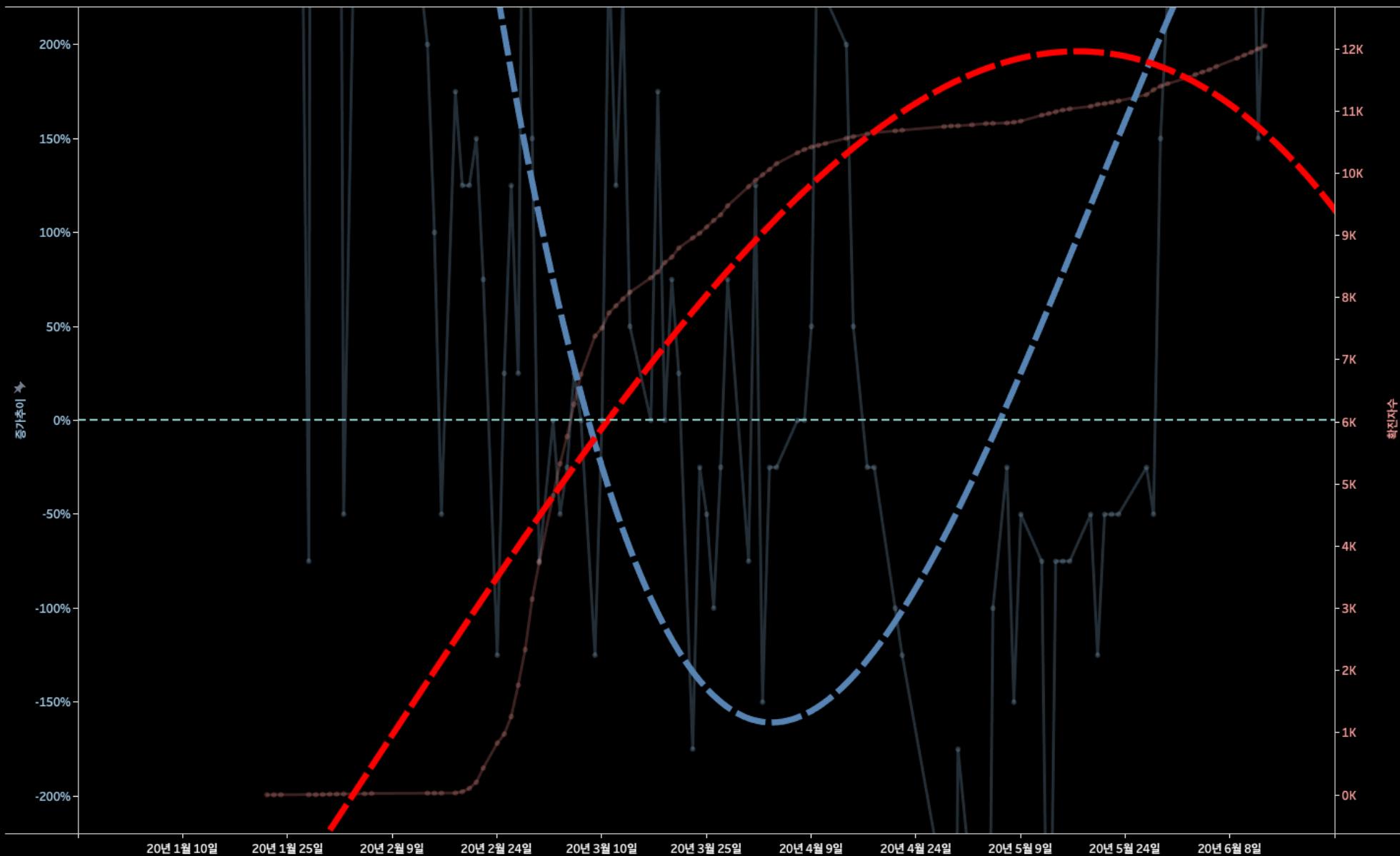
코로나의 중반에는 감소추이를
보이며 후반으로 가며 다시
증가하는 모습입니다.



교육

유치원

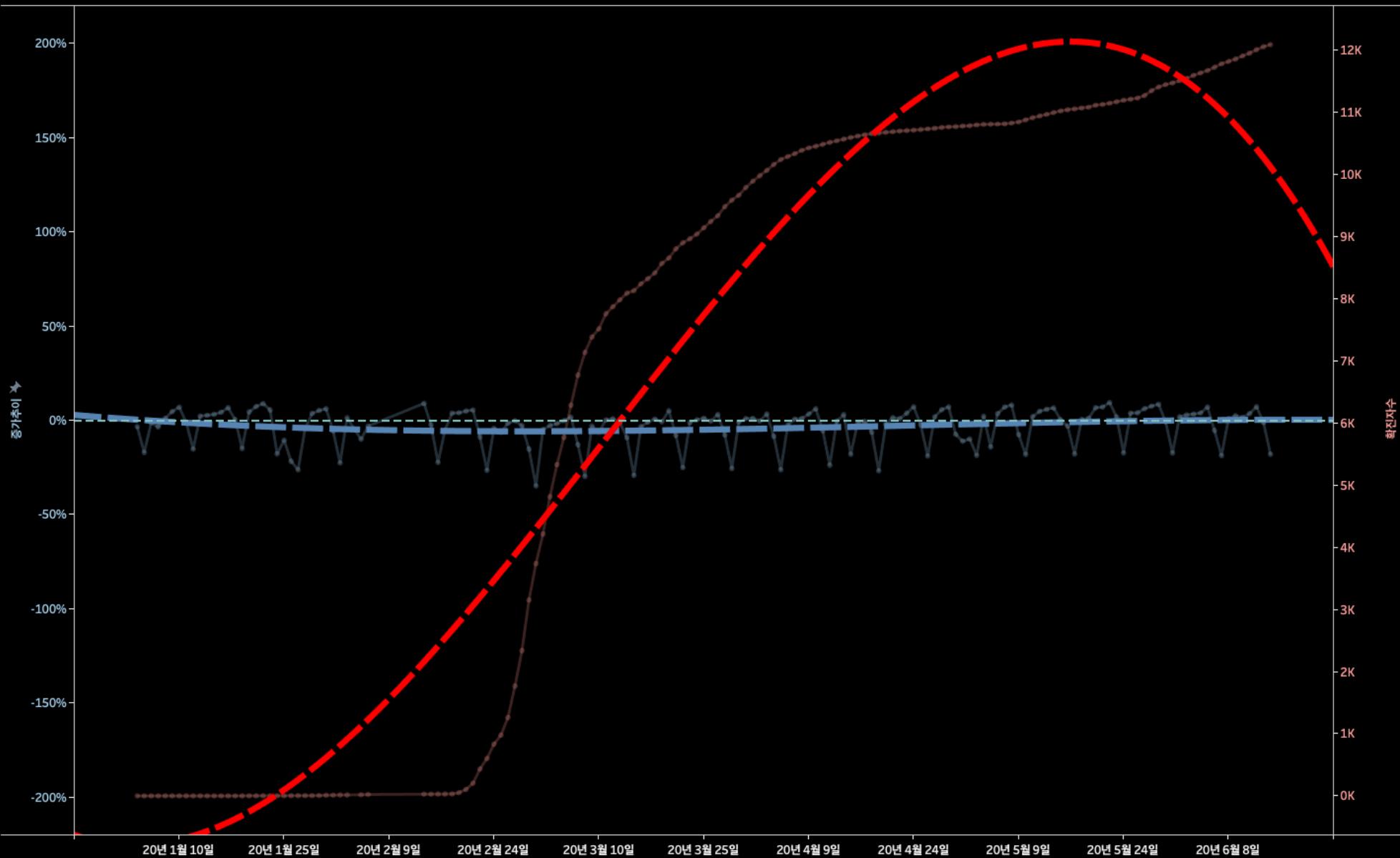
바이러스에 취약한 아동이기 때문에 급격한 감소추이를 보인 것으로 예상됩니다. 이후 코로나가 안정세를 취하자 다시 증가추이를 보였습니다.



교통

교통량

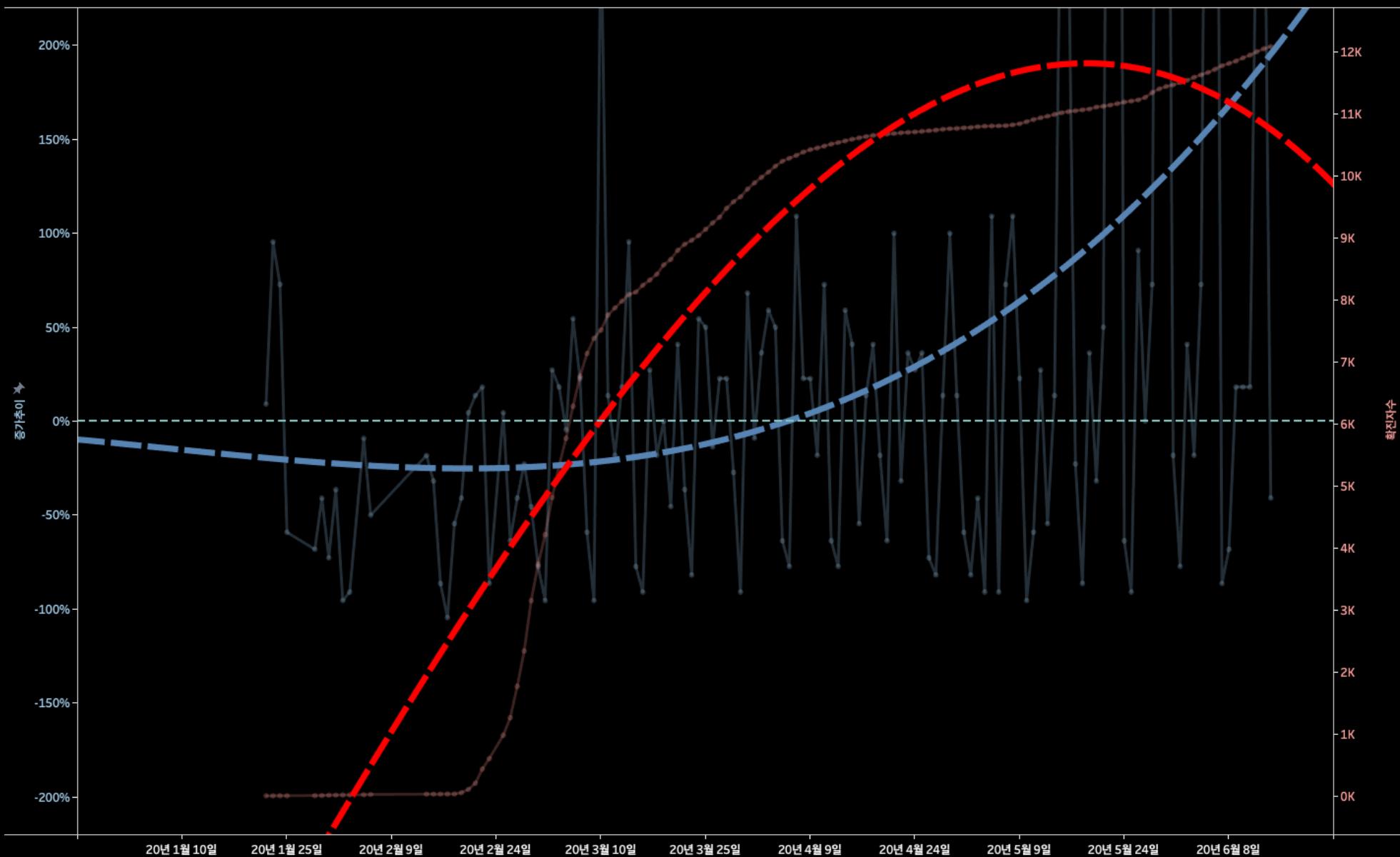
코로나보다는 평일과 주말의 영향이 더 크다고 생각되며 코로나에는 거의 영향을 받지 않았다고 볼 수 있습니다.



농업

기타농업관련

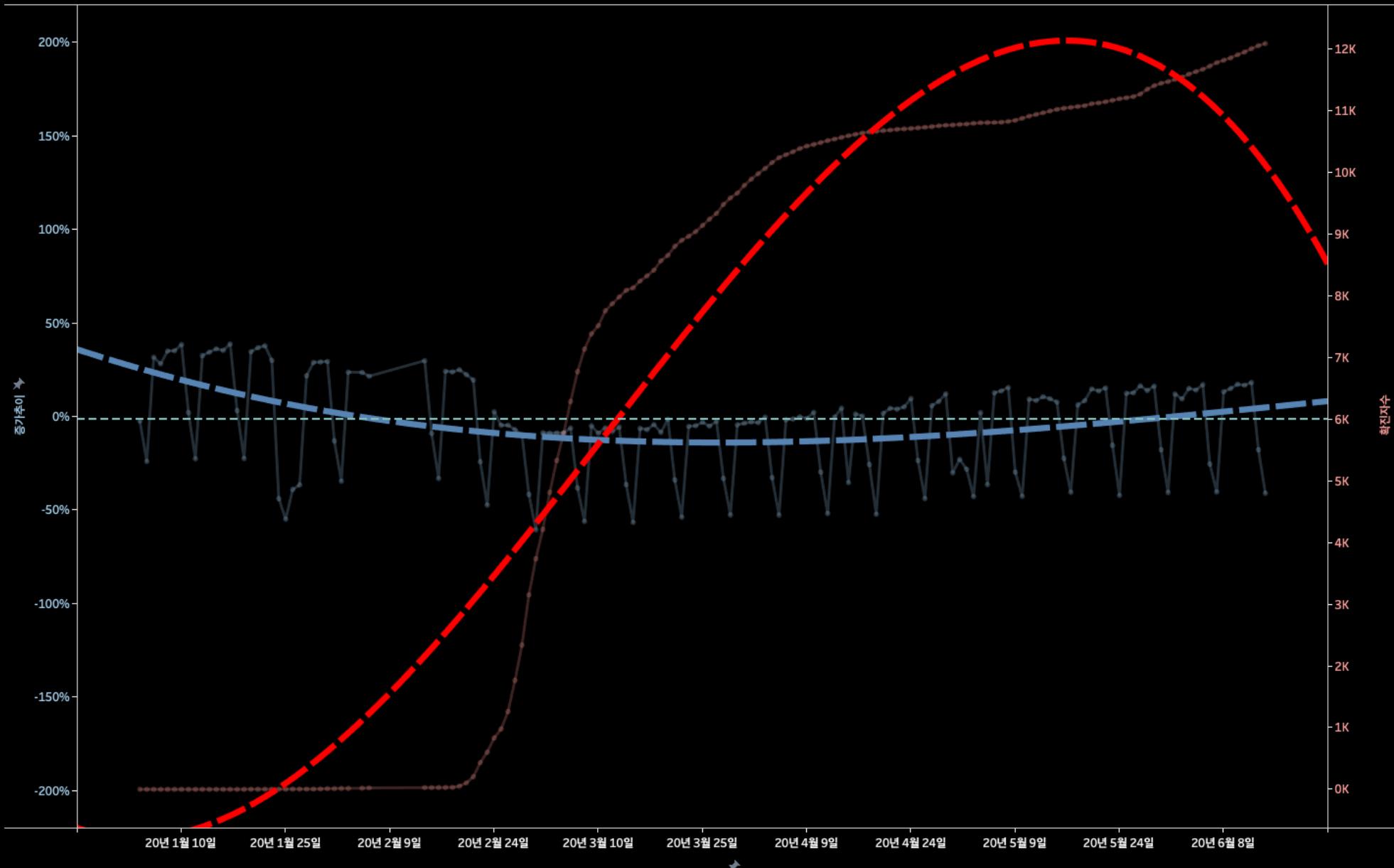
농업 카테고리의 기타농업관련은
코로나의 영향을 받기보다는
농업이 수확을 하는 시기에
영향을 받았다고 생각합니다.



대중교통

버스

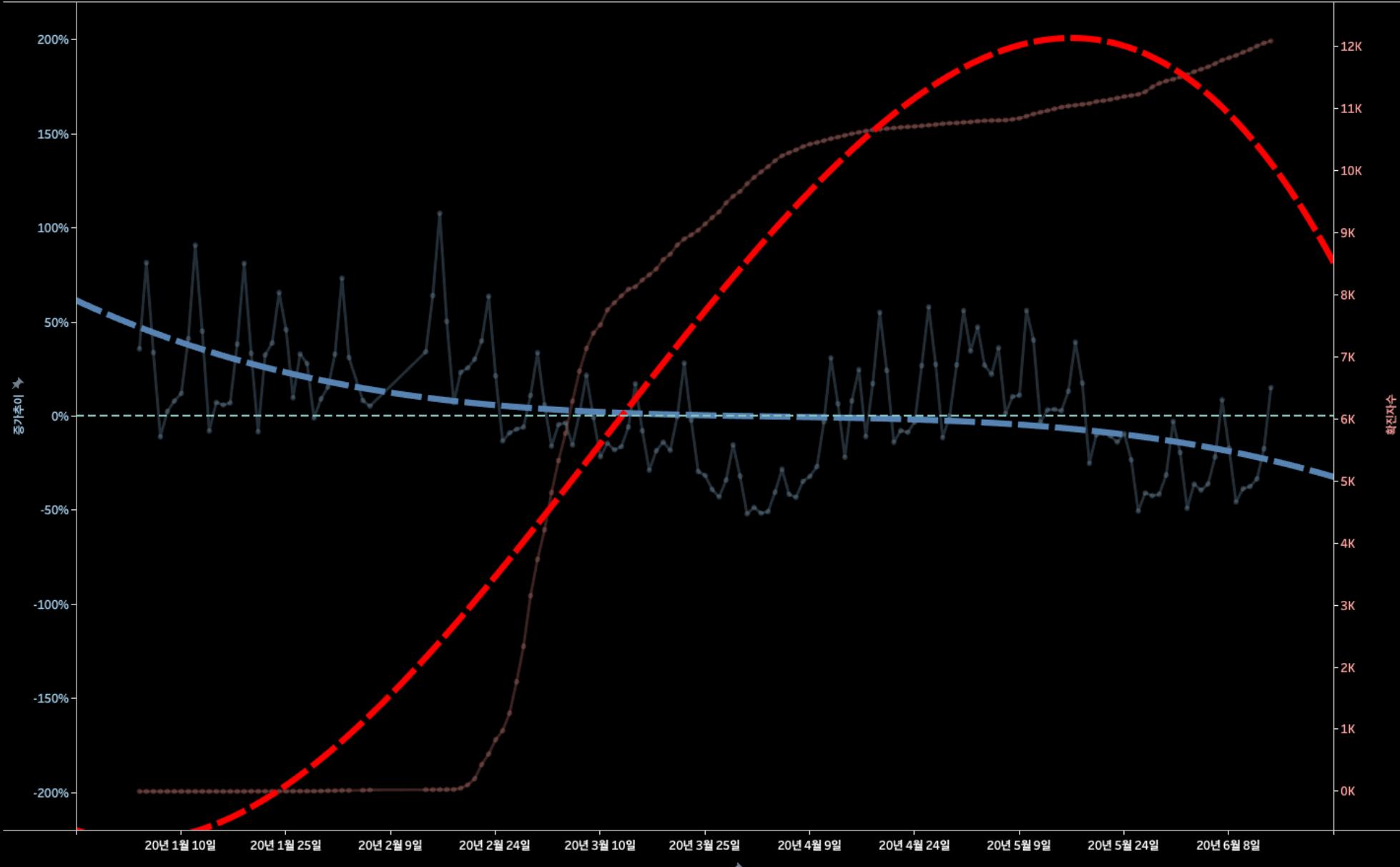
코로나의 초반에는 대중교통 확산을 우려해 자차 혹은 자택근무 등으로 인한 감소추이가 일어난 것이라고 생각합니다. 전체적인 증감추이의 변화는 많이 없었으며 평일과 주말에 의한 규칙적인 변화를 확인할 수 있었습니다.



실내활동

노래방

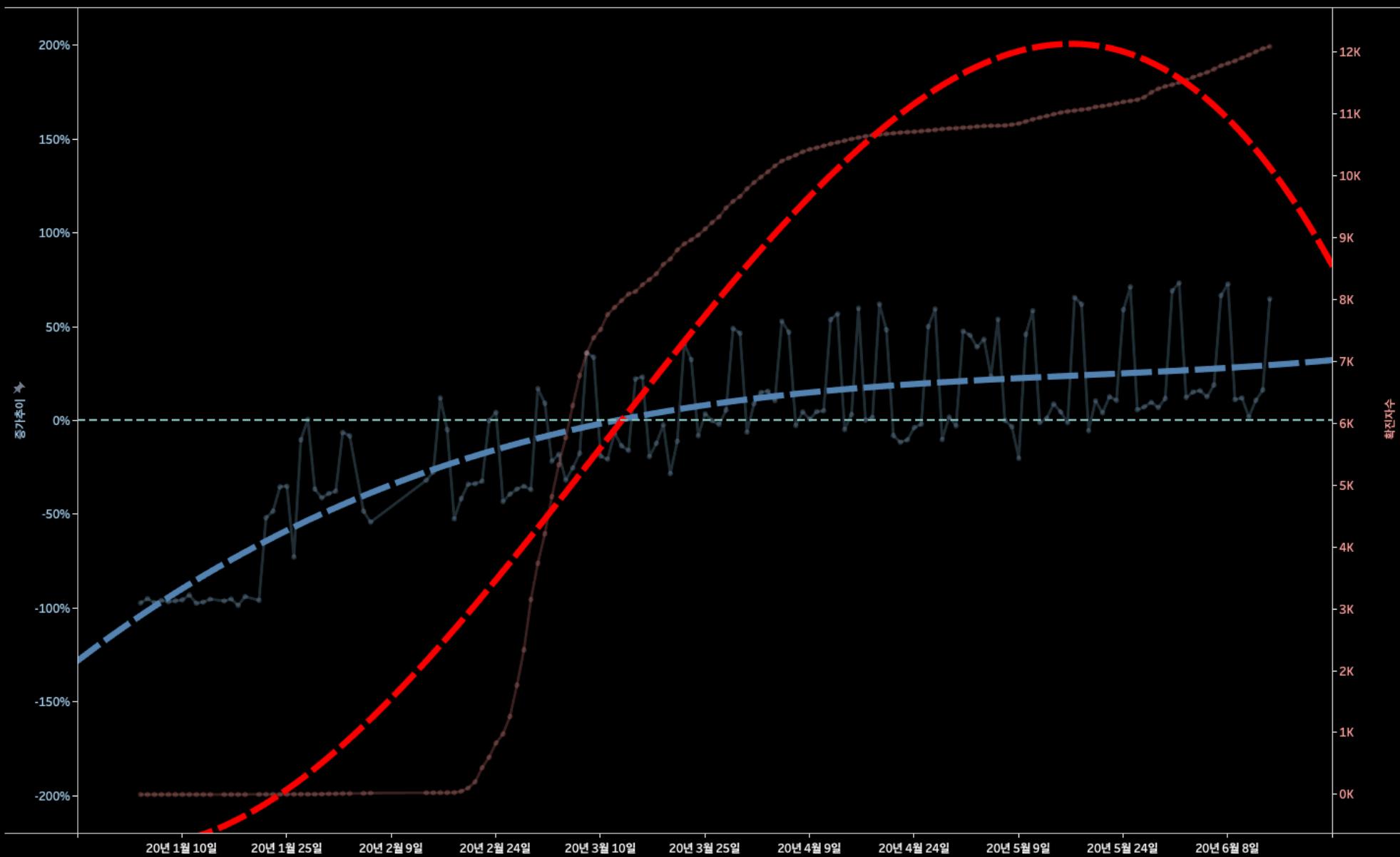
정부에서 유흥시설 등에 코로나 확산 위험이 있다고 발표하면서 감소추이가 꾸준하게 보인 것으로 판단됩니다.
집합금지명령 이후의 그래프를 확인하면 눈에 띄게 감소하는 것을 확인할 수 있습니다.



실외활동

골프연습장

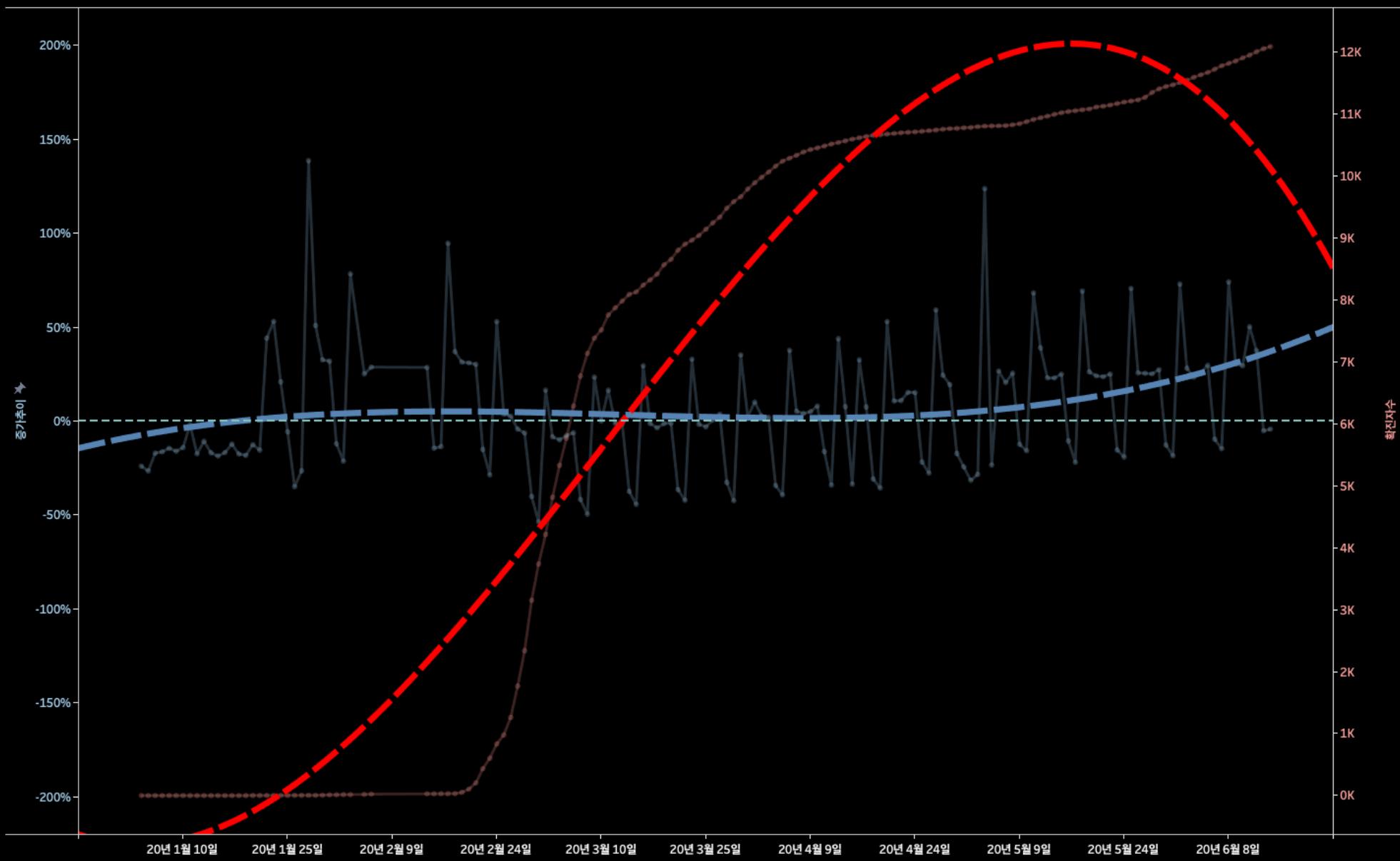
코로나 이후로 꾸준히
증가추세를 보이고 있습니다.
코로나가 확산됨에 따라
실내활동이 제한되며 실외활동이
상대적으로 증가되었다고
생각합니다.



온라인

인터넷Mall

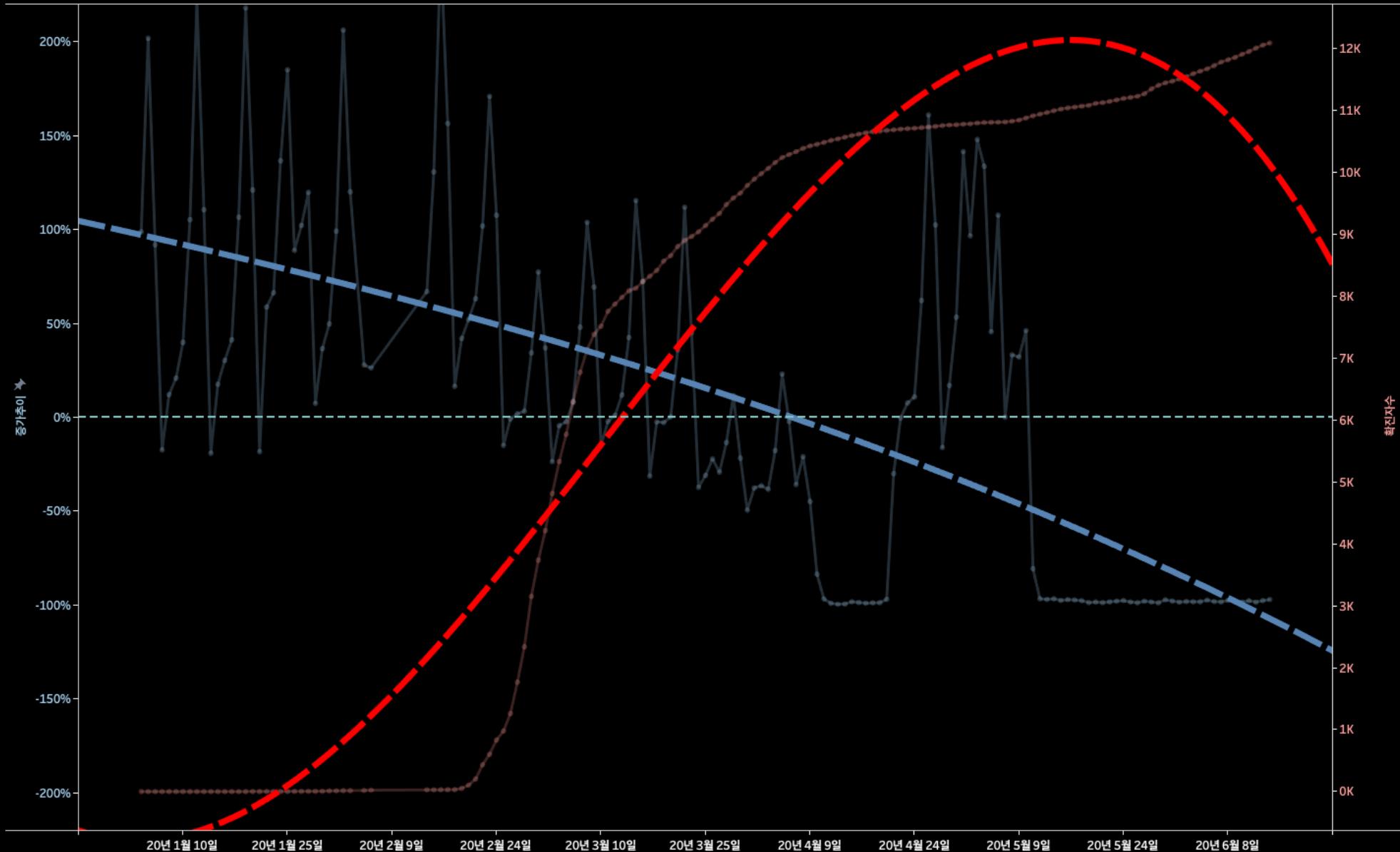
코로나가 확산됨에 따라
사람들이 오프라인으로
구매하는 것이 아닌
온라인으로 구매하는 양이
늘어났다고 판단됩니다.



유흥시설

유흥주점

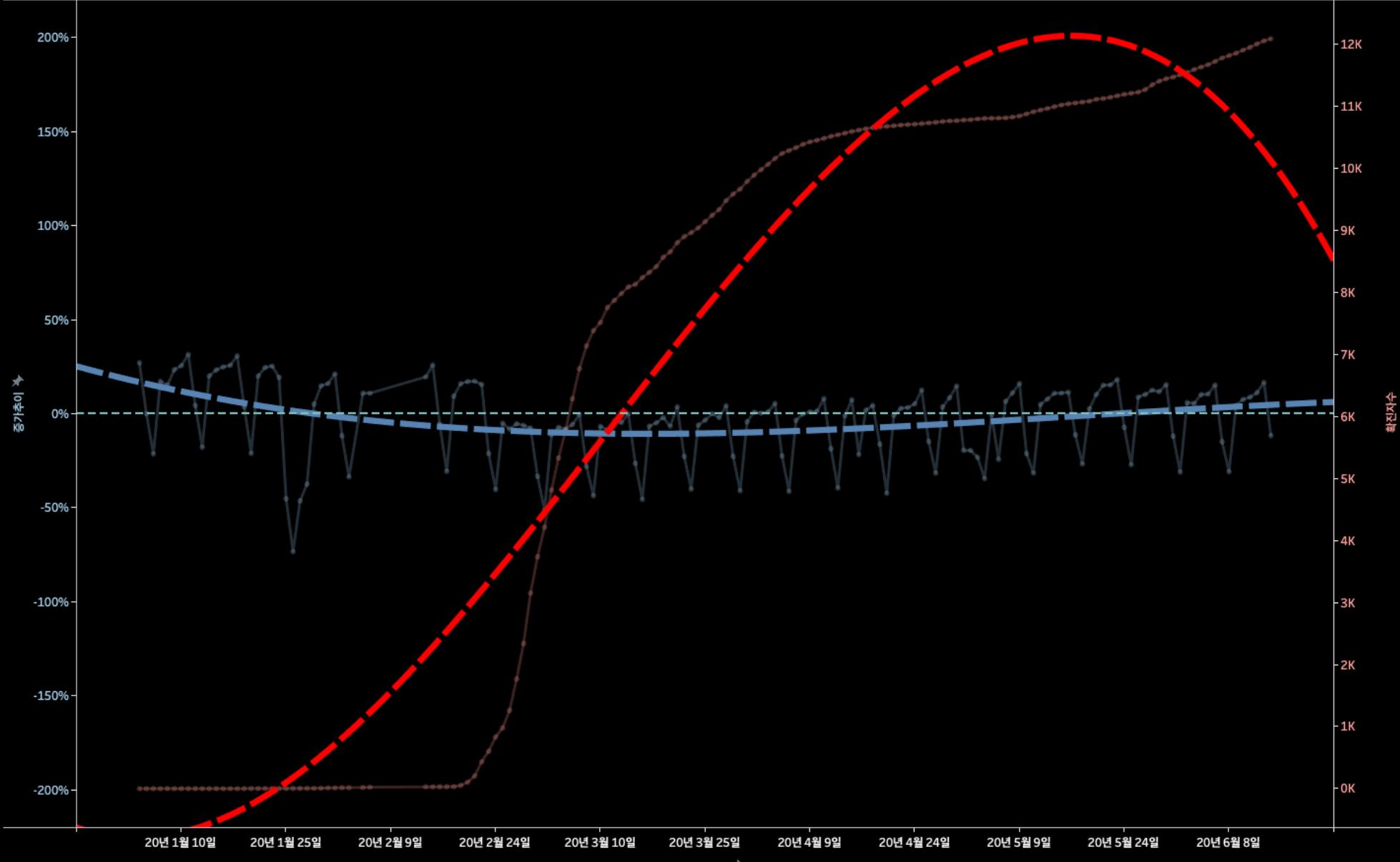
사회적 거리두기로 인해
감소추이가 나타나는 것으로
판단되며
집합금지명령 시행 및 이행 여부
점검으로 인해 4월 이후 급격하게
감소하는 구간이 보입니다.



음식점

일반한식

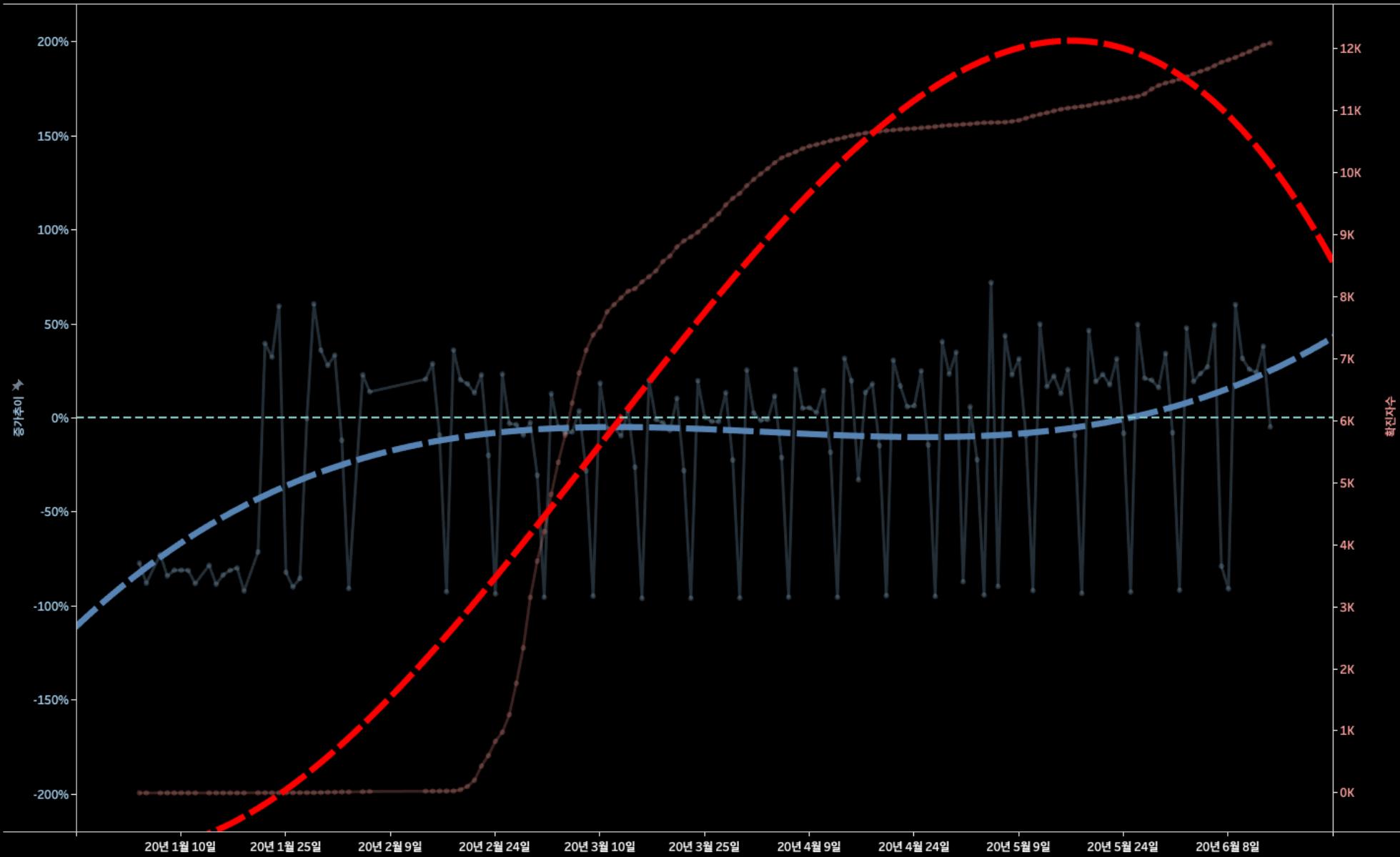
사회적 거리두기로 코로나
초반에 감소추이를 보인 것으로
판단됩니다.
국민 정부재난지원금으로 인해
음식점에 대한 감소추이의
정도가 줄고,
다시 증가추이를 보이며 정상화
되는 것으로 판단하였습니다.



의료

병원

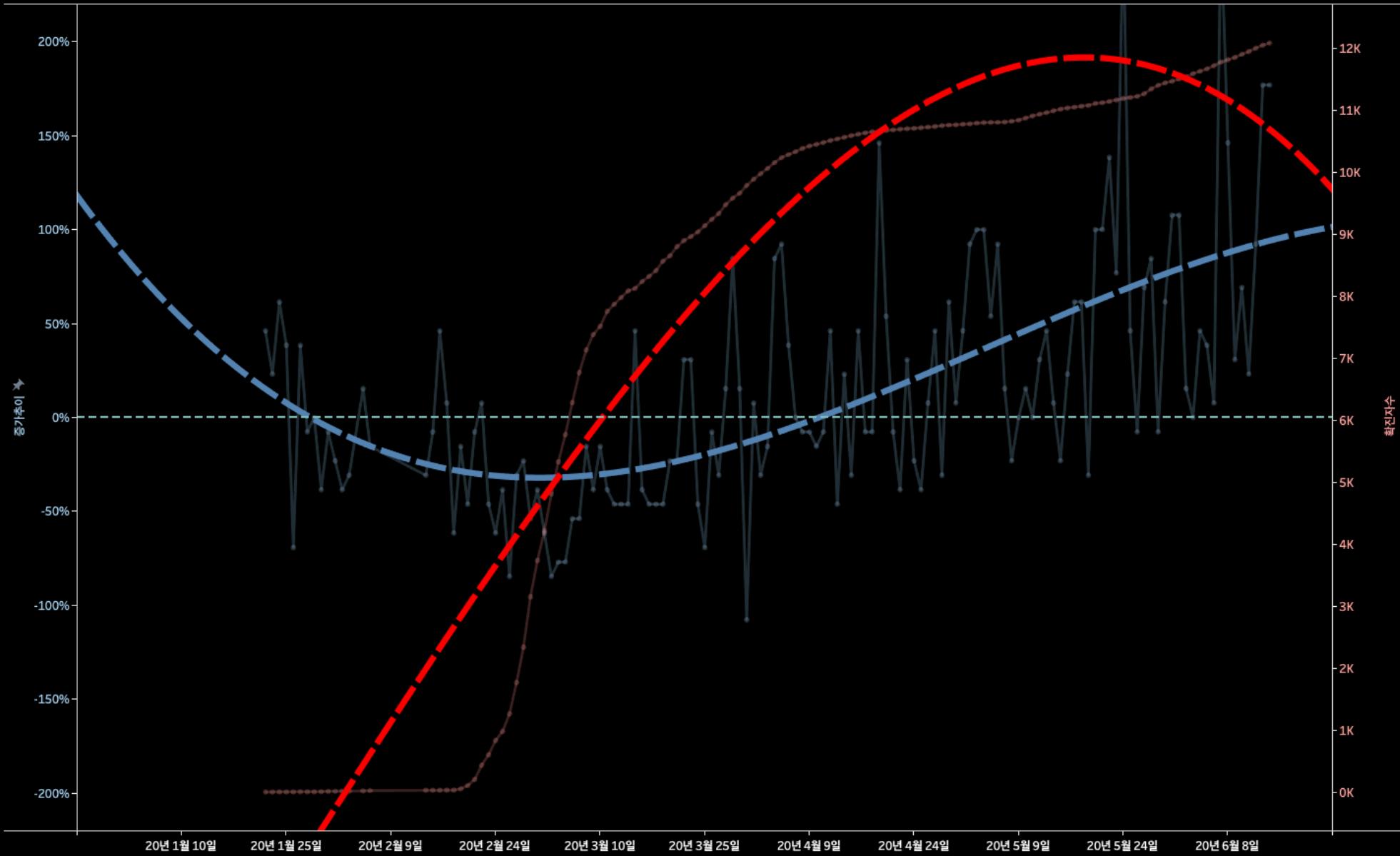
코로나 의심환자들이 발생하기 시작하며 병원 환자들이 증가하는 현상이라고 생각합니다. 병원은 평일과 주말에 영향을 많이 받는 것으로 판단됩니다.



의류

캐쥬얼의류

코로나 초반 사회적 거리두기로
인해 감소추이가 나타난 것으로
생각합니다.
후반에 들어서며 사람들이
보복소비로 인해 추이가 급증한
것으로 생각됩니다.



한계 및 추가 연구 내용

한계 및 추가 연구 내용

1. 데이터 부족 문제

- 계절성을 보이는 데이터가 많았으나 계절성과 코로나의 영향을 비교하기에 한계가 있었습니다.
- 연도별 데이터가 있었다면 시계열 예측 모델 구축이 가능했을 거라 생각합니다.
- 현재는 약 6개월 정도의 데이터만 존재하기 때문에 학습과 테스트데이터를 나눠 모델링 했다고 가정한다면 계절, 날씨, 온도 등에 대한 데이터가 부족합니다.

2. 데이터 스타일 문제

- 데이콘에서 제공되는 B2C에 해당하는 데이터들을 제공하고 있습니다.
- 주제에 맞게 시대 유망 품목 및 산업 발굴을 위해서는 B2C뿐만 아니라 B2B, C2C 등 더 다양한 형태의 데이터들이 있었다면 전체 산업을 파악하는데 용이할 것입니다.

한계 및 추가 연구 내용

1. 예측 모델링

- 추가 데이터가 수집되었다는 가정하에
- 전세계가 이슈화되는 전염병이 터지는 주제들을 분석 및 예측하여 예방할 수 있고
- 국민의 생활 / 소비 변화를 예측하여
정부에서 급변하는 변화를 대응할 수 있을 것이라고 생각합니다.

2. 배달

- 우리나라에서 급변하는 산업 중 하나가 배달이라고 생각합니다.
- 음식, 자동차, 옷, 이제는 서비스를 배달도 증가하는 추세입니다.
- 이처럼 배달산업을 좀 더 분석한다면 생활/소비변화에 이용할 수 있을 것이라고 생각합니다.

THANK YOU