경진대회

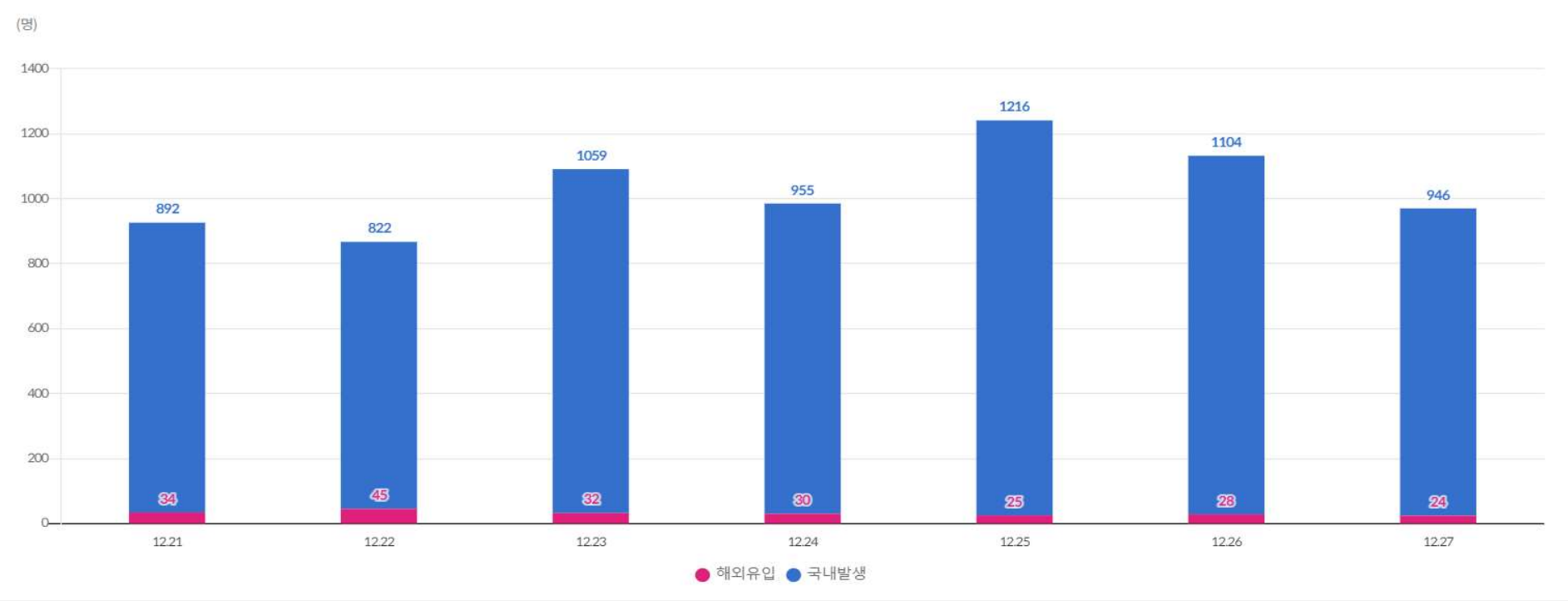
보자보자 어디보자

5639190 통계학전공 최지우
5343113 경영정보학전공 남은수

INDEX
목차

PART 1
코로나 현황

감염경로구분에 따른 신규 확진자 현황(12.27. 00시 기준)



PART 1
코로나 현황

일일 및 누적 확진 환자 추세 (12.27. 00시 기준)

PART 1
코로나 현황

서울시 코로나 확진자 현황 (12.28. 00시 기준)

PART 2
문제 제시

문제 제시

- 코로나는 전 세계적으로 큰 이슈로 떠오르고 있음
- 우리나라도 코로나에 대해 민감하게 대응하고 있지만 코로나 확진자는 꾸준히 나오고 있는 추세임
- 특히 서울, 수도권 지역을 중심으로 많이 나타나고 있음
- 서울 내에서도 구별로 확진자 수 차이가 크게 나타나고 있음
- 구별로 코로나 확진의 요인들을 찾아내고 확진자 수를 예측함으로써 코로나를 미리 예방할 필요가 있음

PART 3
데이터 설명

데이터 설명

서울시 코로나 확진자 현황 ✓ 확진자 개개인별 정보 제공

- 연번, 확진일, 환자번호, 국적, 환자정보
- 지역, 여행력, 접촉력, 조치사항, 상태
- 이동경로, 등록일, 수정일, 노출여부

SKT 유동인구 데이터 ✓ 지역별 유동인구 데이터

- 일자, 시간(1시간단위)
- 연령대(10대단위), 성별
- 시 , 군구, 유동 인구수

네이버 검색 데이터 ✓ 네이버에서 검색 키워드 사용 트랜드 데이터

- 날짜, 코로나
- 사회적거리두기, 우울

PART 3
데이터 설명

데이터 설명

구글 이동성 데이터

- ✓ 구글에서 제공하는 이동 경향을 나타내는 데이터
- ✓ "구글 모빌리티 리포트로 알아본 k-방역" 참조

- ✓ **Date** – 날짜
- ✓ **grocery_and_pharmacy_percent_change_from_baseline** – 식료품 매장, 식자재 창고, 농산물 시장, 전문 식품 매장, 드럭스토어, 약국과 같은 장소에서 나타난 이동 추이
- ✓ **parks_percent_change_from_baseline** – 지역 공원, 국립 공원, 공용 해수욕장, 마리나, 반려견 공원, 광장, 공공 정원과 같은 장소에서 나타난 이동 추이
- ✓ **transit_stations_percent_change_from_baseline**– 지하철, 버스, 기차역 등의 대중교통 허브와 같은 장소에서 나타난 이동 추이
- ✓ **retail_and_recreation_percent_change_from_baseline**– 식당, 카페, 쇼핑센터, 놀이공원, 박물관, 도서관, 영화관과 같은 장소에서 나타난 이동 추이
- ✓ **residential_percent_change_from_baseline**– 거주지에서 나타난 이동 추이
- ✓ **workplaces_percent_change_from_baseline** – 직장에서 나타난 이동 추이

PART 3

누적 확진자 현황 (12.27. 00시 기준)

확진환자				격리해제		격리중		사망	
누적	전일대비			누적	전일대비	누적	전일대비	누적	전일대비
	소계	국내발생	해외유입						
56,872	+970	946	24	39,040	+508	17,024	+447	808	+15

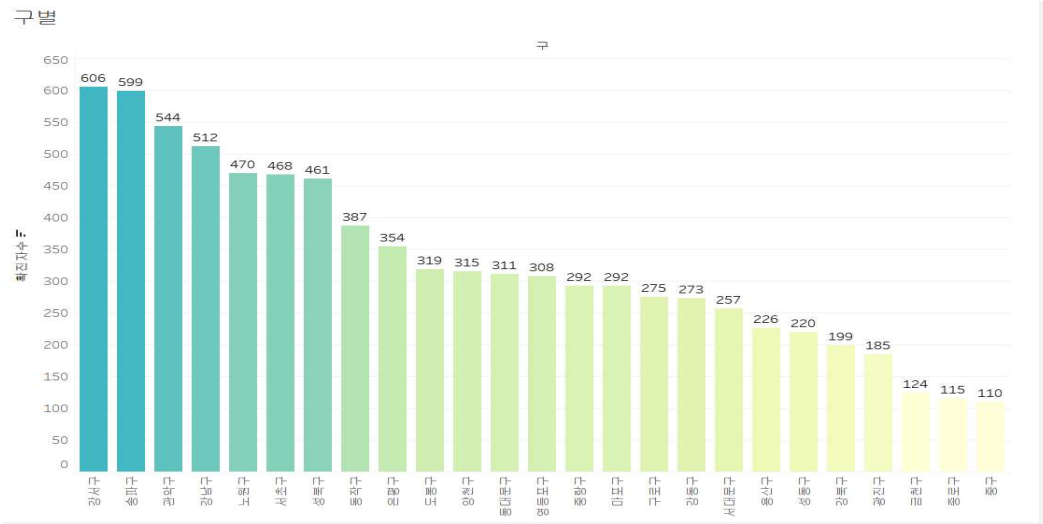
전처리

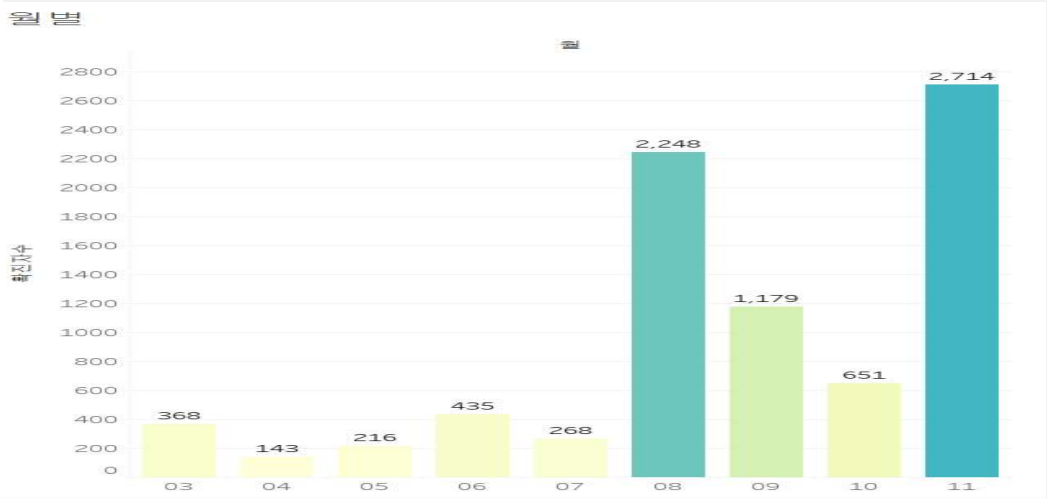
서울시 코로나 현황	구글 이동성 데이터	SKT 유동인구 데이터	네이버 검색 데이터
확진일, 지역을 가지고 월, 일, 구별 환자 수로 정리	월,일,구별로 데이터 정리	월,일,구별로 데이터 정리	월,일별로 데이터 정리

전처리를 한 네 가지 데이터를 하나의 데이터로 합침

PART 5 시각화 – 코로나 현황

시각화





✓ 강서구, 송파구가 코로나 확진자 수가 많음

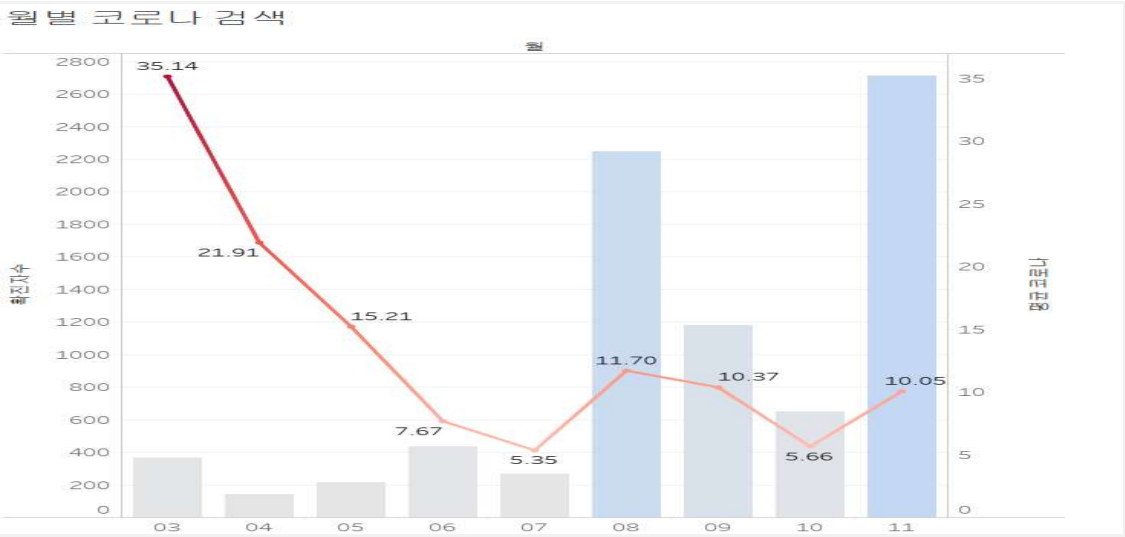
✓ 중구, 종로구가 확진자 수가 가장 적음

✓ 3월부터 11월 까지 월별 코로나 확진수

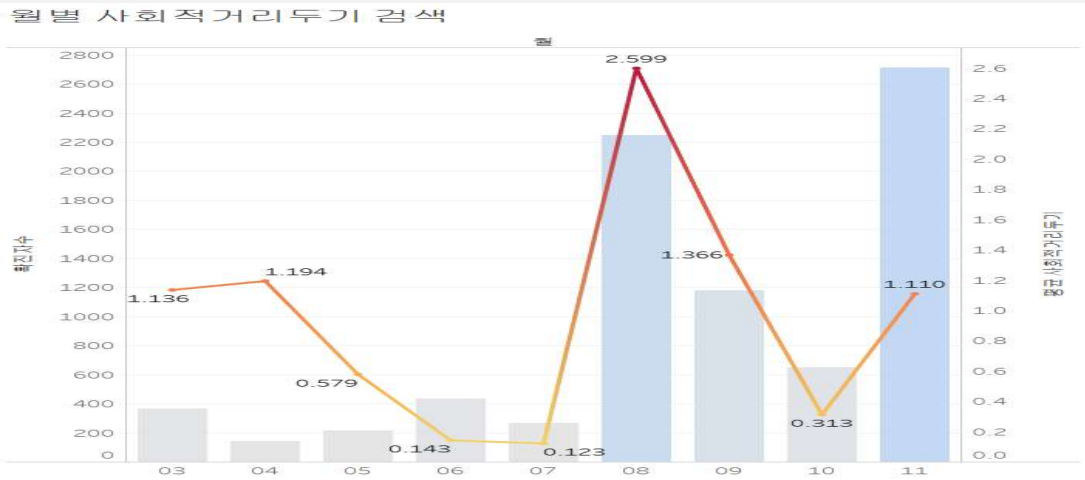
✓ 8월과 11월이 2천명대로 가장 높다

PART 5 시각화

시각화



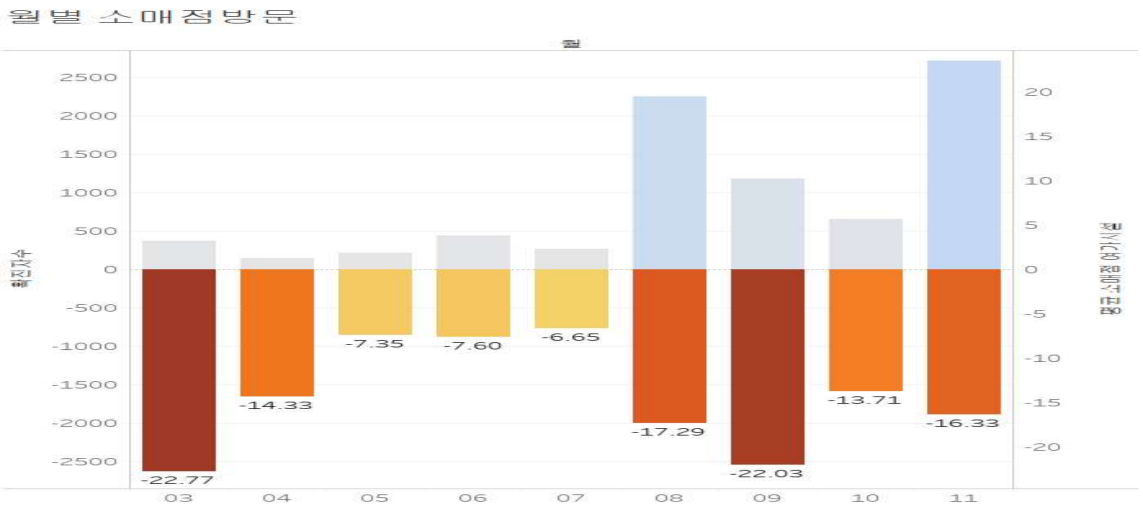
코로나 확진자 수 증감 추세를 따라서
코로나 검색 빈도 또한 같은 추세로 증감함



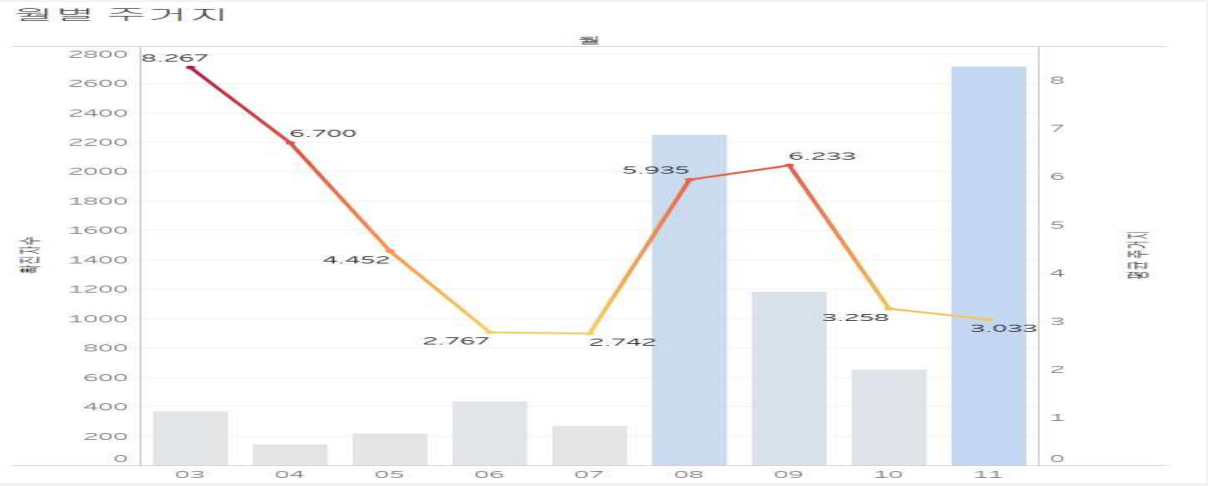
코로나 확진자 수 증감 추세를 따라서
사회적 거리두기 검색 빈도 또한
같은 추세로 증감함

PART 5 시각화

시각화



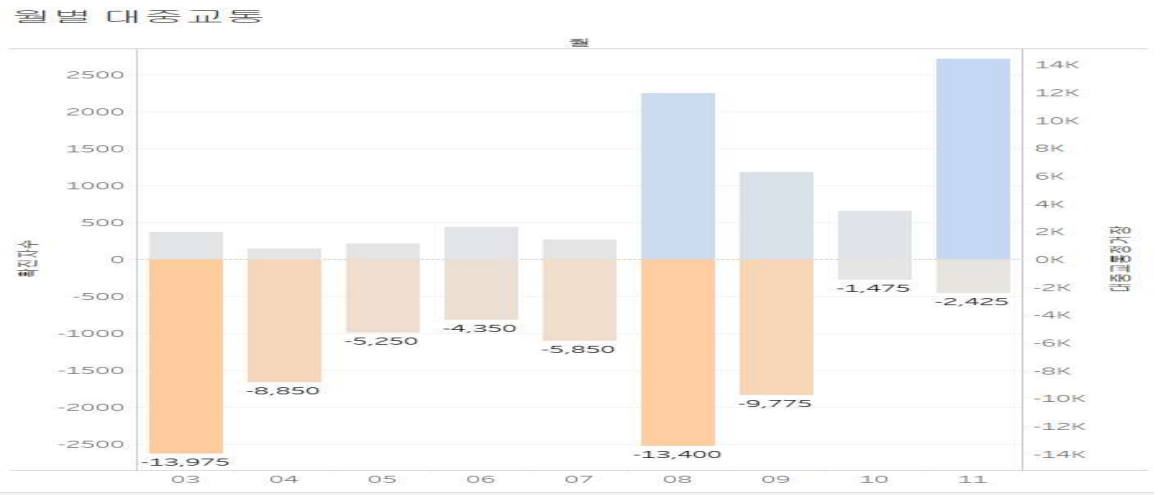
월별 소매점 방문 지수는 코로나 확진자 수의 증감 추세를 따라감



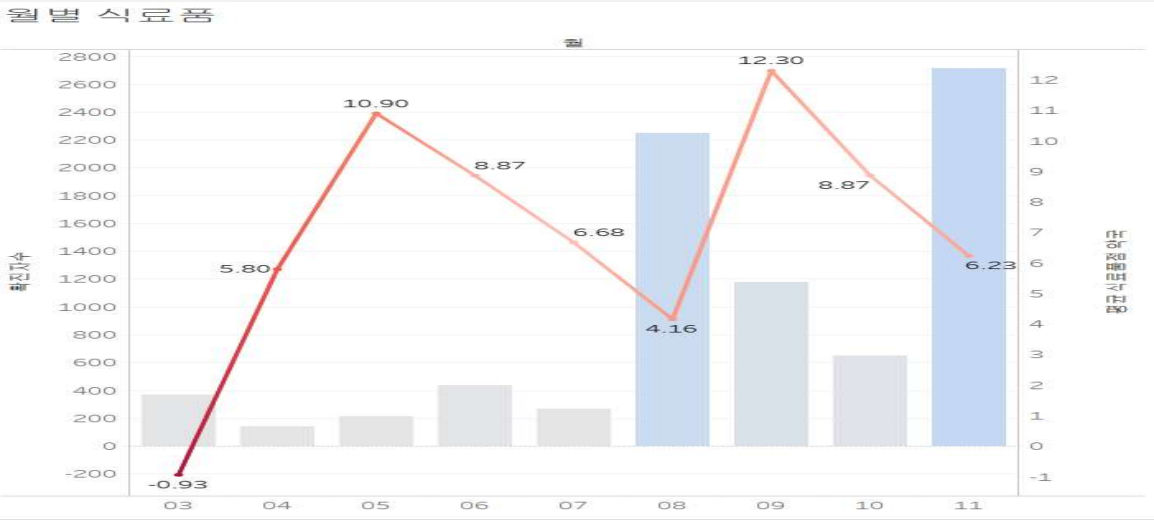
월별 주거지 방문 지수는 코로나 확진자 수의 증감 추세를 따라감

PART 5
시각화

시각화



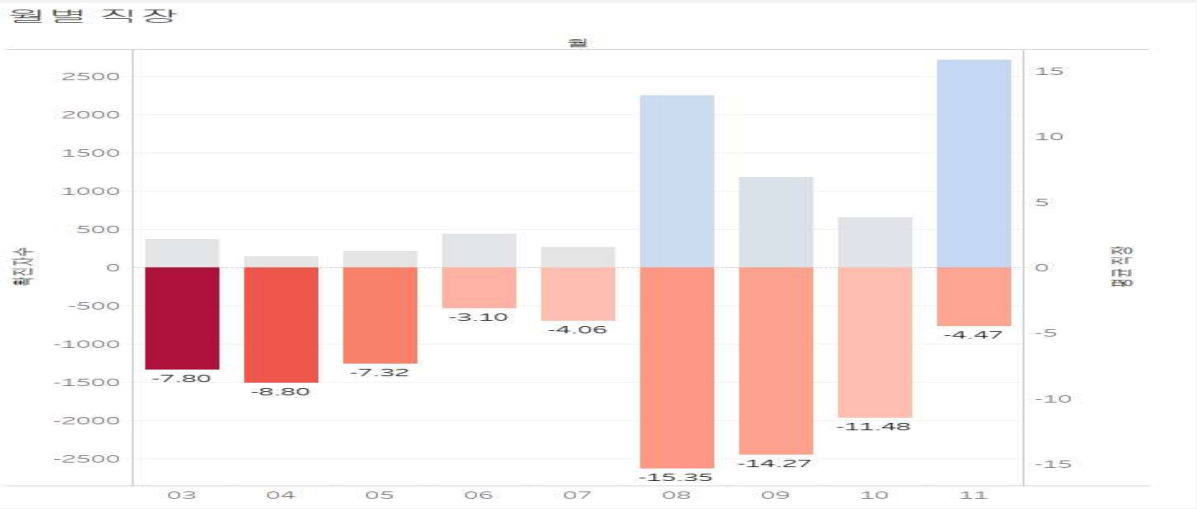
월별 대중교통 정거장 방문 지수는 코로나 확진자 수의 증감 추세를 따라감



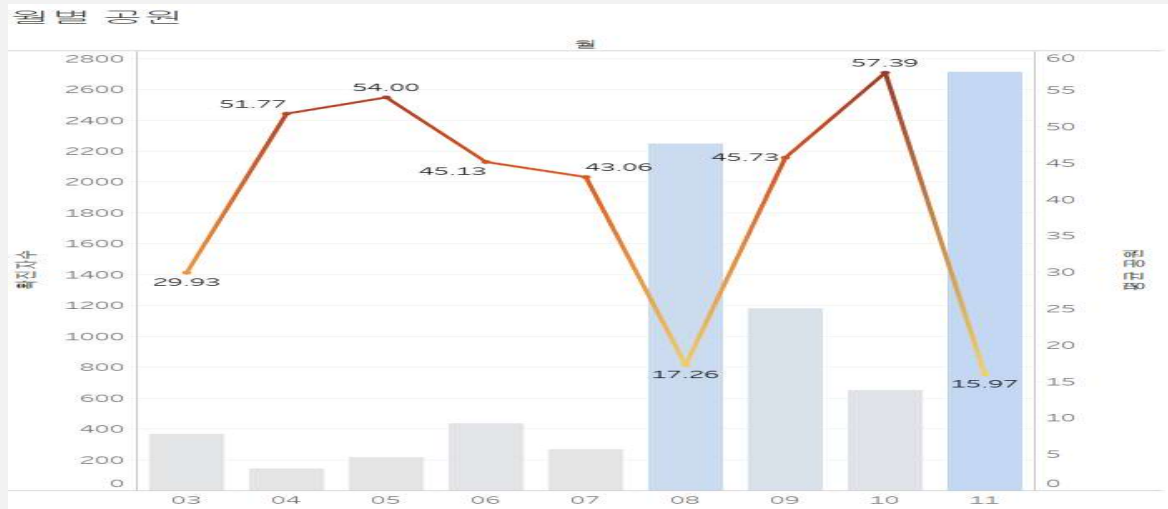
월별 식료품점 방문 지수는 코로나 확진자 수의 증감 추세와 반대로 나타남

PART 5
시각화

시각화

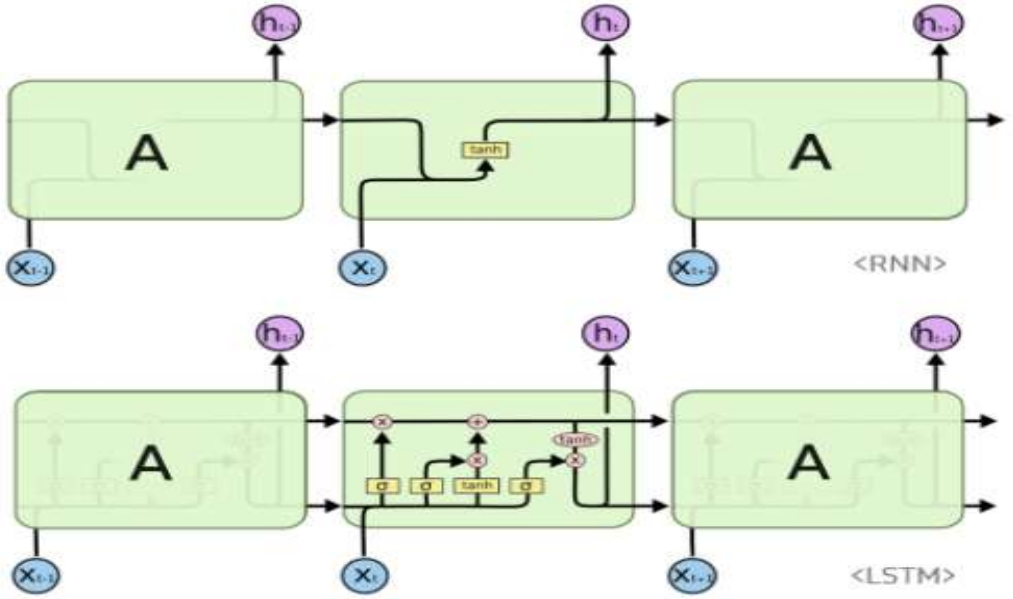


월별 직장 방문 지수는 코로나 확진자 수의 증감 추세와 반비례하여 증감함



월별 공원 방문 지수는 코로나 확진자 수의 증감 추세와 반대로 증감함

모델링 - LSTM



- LSTM은 순차적으로 등장하는 데이터 처리에 적합한 모델인 RNN을 보완하여 만들어진 모델이다.
- RNN의 vanishing gradient problem문제를 극복하기 위해 RNN의 hidden state에 cell-state를 추가한 형태이다.

출처: https://ratsgo.github.io/natural%20language%20processing/2017/03/09/rnnlstm/

PART 6 모델링

모델링

모델에 들어가는 변수(구별로)

기간: 3월 ~ 11월

예측 기간: 11월 16 ~ 11월 30일
- 유동 인구수
- 코로나 검색빈도
- 사회적거리두기 검색 빈도
- 우울 검색빈도
- 소매점 여가시설 이동 데이터
- 식료품점약국 이동 데이터
- 공원 이동 데이터
- 대중교통정거장 이동 데이터
- 직장 이동 데이터
- 주거지 이동 데이터

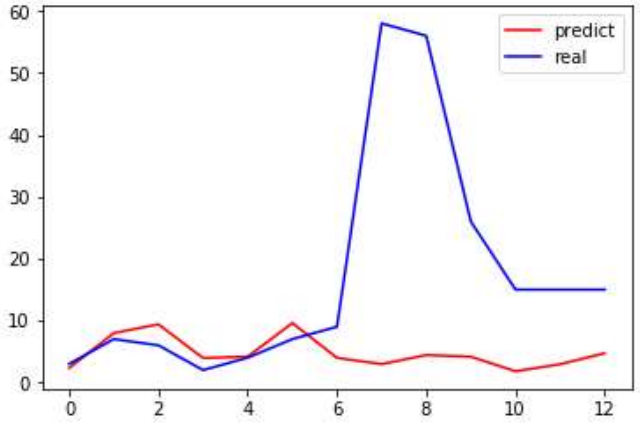
- 구별로 확진자 수가 많은 상위 5구와 하위 5구에 대해서 각각 모델 적용
- 2일의 데이터를 사용하여 다음날의 확진자 수를 예측하는 모델을 만듬

<모델 요약>

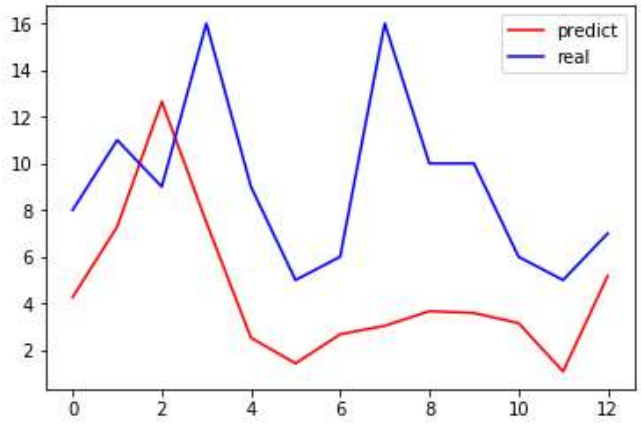
```
model.summary()
```

```
Model: "model_30"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_31 (InputLayer)        [(None, 2, 10)]           0
_________________________________________________________________
lstm_30 (LSTM)               (None, 64)                19200
_________________________________________________________________
dense_60 (Dense)             (None, 32)                2080
_________________________________________________________________
dense_61 (Dense)             (None, 1)                 33
=================================================================
Total params: 21,313
Trainable params: 21,313
Non-trainable params: 0
_________________________________________________________________
```

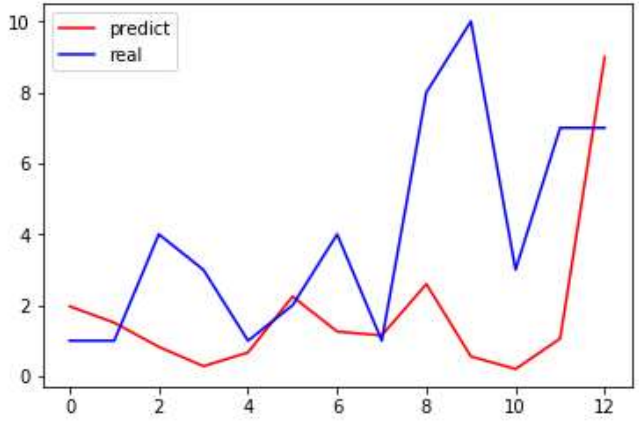
PART 7 결과 확진자 수 Top5. 구



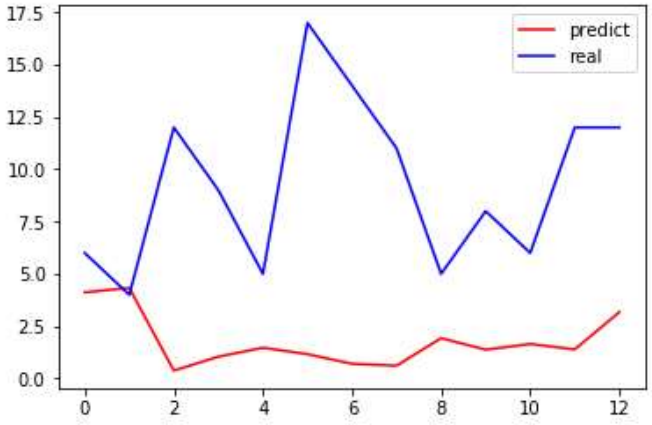
강서구



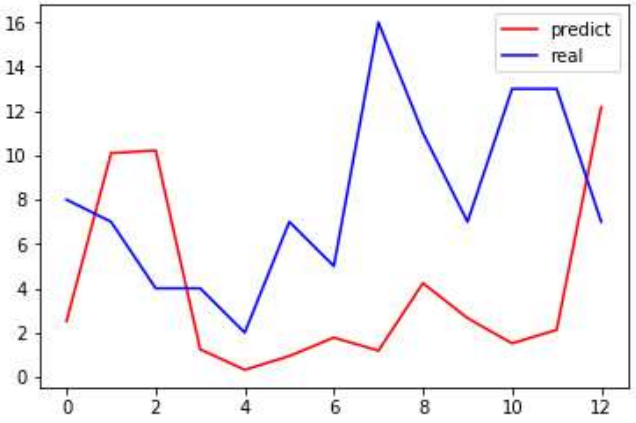
송파구



관악구



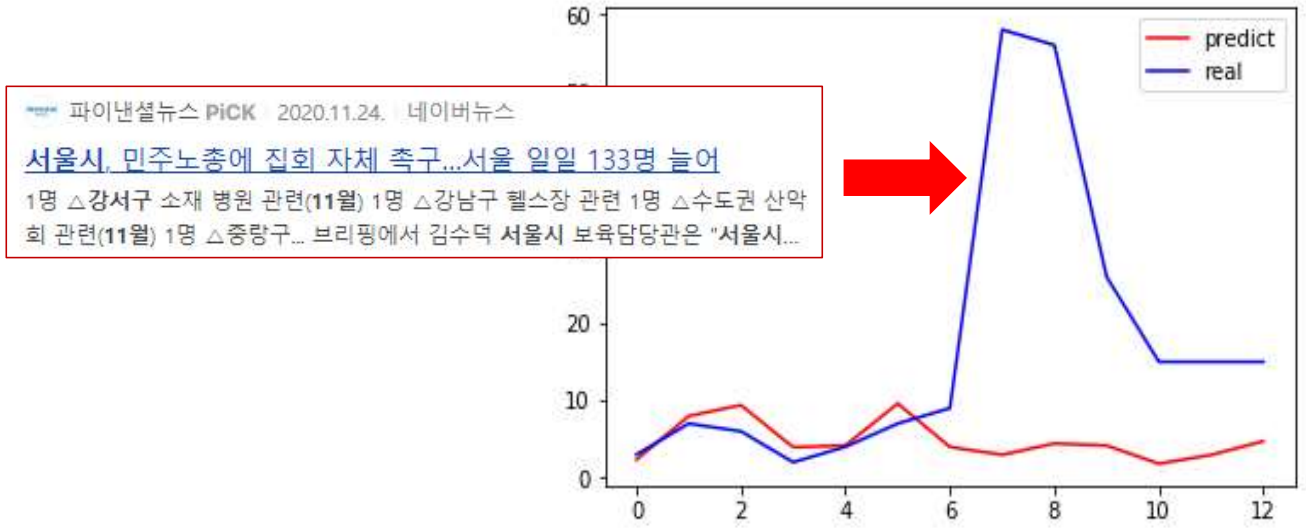
강남구



노원구

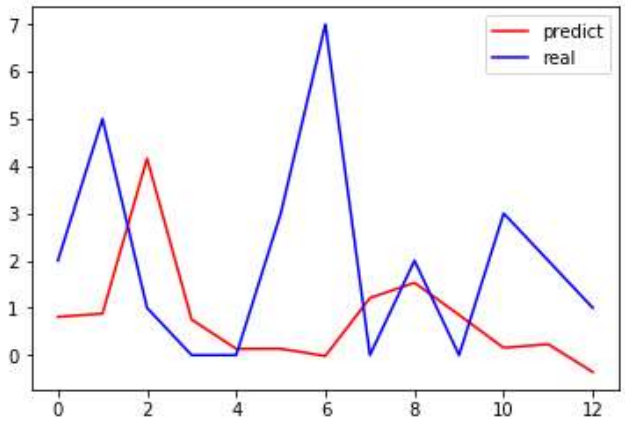
PART 7 결과 확진자 수 Top5. 구

- 확진자 수가 많은 5개의 구를 예측한 결과 대략적인 확진자 수 증감을 따라감
- 하지만 급격한 증가나 감소를 보이는 구간은 잘 따라가지 못하는 모습을 보임
- 확진자 수가 급증하거나 급감하는 원인을 나타내는 요인을 찾을 필요가 있음
- 특히 강서구와 같이 확진자가 급격하게 발생하는 이벤트성 사건을 예측할 수 있는 요인이 필요함

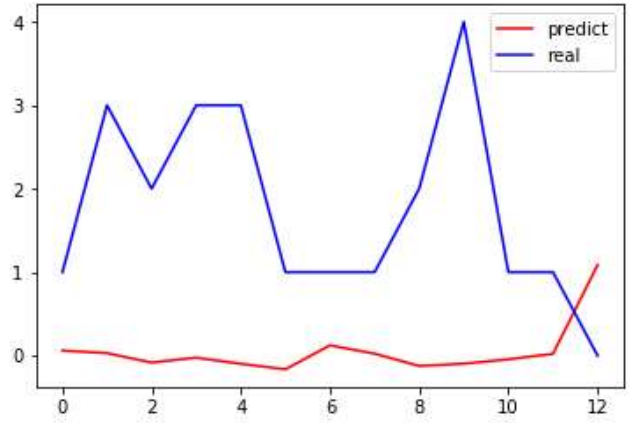


PART 7 결과

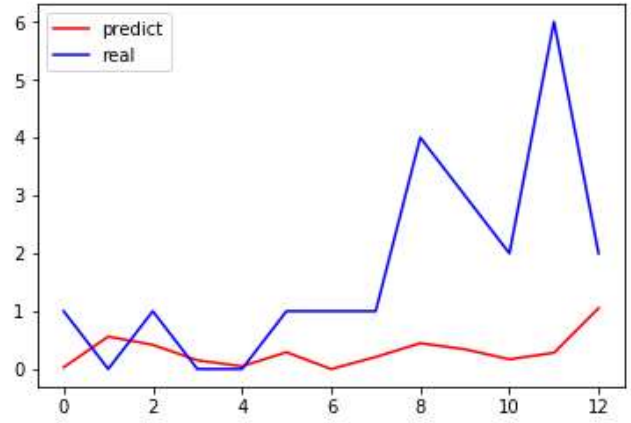
확진자 수 bottom5. 구



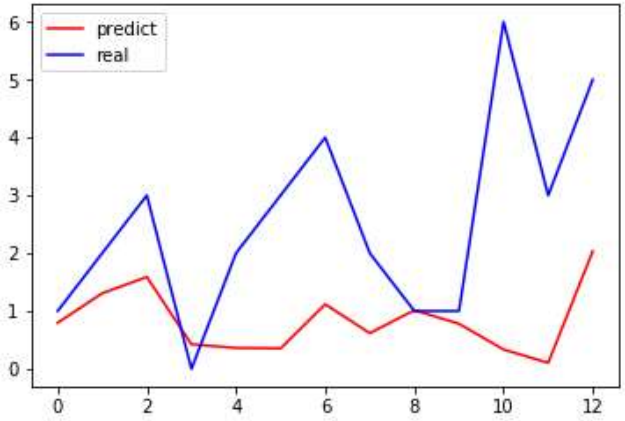
중구



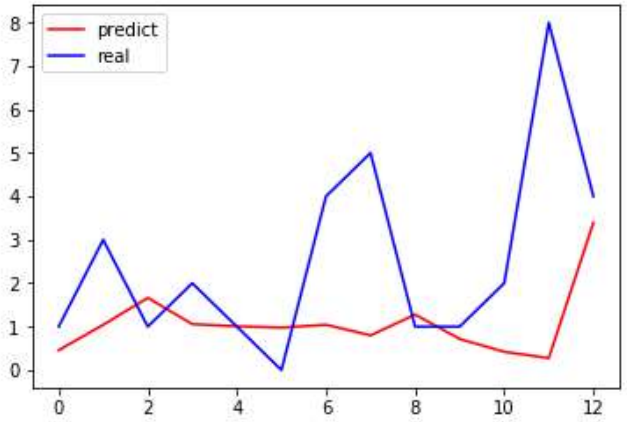
종로구



금천구



광진구



강북구

PART 7 결과 확진자 수 bottom5. 구

- 확진자 수가 적은 5개 구의 경우 증감 추세를 따라가지만 예측값이 거의 0에 가까움
- 확진자 수의 증감의 단위가 작아서 예측값과의 오차는 작게 나옴
- 확진자 수가 적은 구의 경우도 확진자가 나오게 되는 원인을 나타내는 요인을 찾을 필요가 있음

PART 8 결론

결론 및 한계점

- 확진자 수가 많은 상위 5개 구와 하위 5개구를 시계열 딥러닝 모델인 LSTM을 통해 약 2주간의 확진자 수를 예측한 결과 대략적인 증감 추세는 따라감
- 하지만 확진자 수가 급증하거나 급감하는 경우에는 모델의 예측값과의 차이가 많이남
- 확진자 수가 급증 및 급감하는 이벤트성 사건을 나타낼 수 있는 요인들을 찾는 것이 중요함
- 시계열 모델을 사용하였지만 분석에 사용한 기간이 3월~11월로 학습할 수 있는 기간이 부족함
- 더 긴 시간의 학습기간이 주어진다면 증감의 추세뿐 아니라 확진자 수의 오차도 줄어들 수 있음