

1. 개인 감염보다 '그룹 감염' 비율이 더 높을 것이다

- group 변수가 TRUE일 때가 FALSE보다 더 많을 것이다
- 그룹 감염일 경우 가장 높게 나타난 infection_case는 Shincheonji Church, 신천지 교회이다
- 개인 감염일 경우 가장 높게 나타난 infection_case는 overseas inflow, 해외유입이다
- 지역별 코로나 집단감염 환자 수는 인구밀도(인구수/지역 면적)와 비례할 것이다.

2. '확진자 수'는 유동성이 높은 20대, 그리고 남성일 때 가장 클 것이다.

- age 변수가 20s일 때 발생빈도가 가장 높을 것이다
- sex 변수가 male일 때 female보다 발생빈도가 높을 것이다

3. '확진자 대비 사망비율'은 면역력이 약한 노인층(70-100세), 그리고 남성일 때 가장 클 것이다.

- age 변수의 값이 높을수록 state 변수의 deceased/confirmed 비율이 높을 것이다(70-100세에서 가장 높을 것이다)
- sex 변수가 male일 때 female보다 state 변수의 deceased/confirmed 비율이 높을 것이다
- 격리해제 평균 기간과 사망 평균 기간은 연령과는 비례하고, 성별과는 무관할 것이다.
- '코로나 확진자/해당 연령대의 인구수' 비율은 연령과 비례할 것이다.

4. 시계열 자료 분석

- 시간이 지날수록 확진비율(confirmed/test)과 사망비율(deceased/confirmed)은 감소, 격리해제비율(released/confirmed)은 증가할 것이다. (연령별, 성별, 지역별로 비교)
- 특정 시점에서 확진자 또는 사망비율이 급증한 원인에는 그룹 감염의 영향이 클 것이다.

5. 서울에서 유동인구가 많은 곳이 집단감염이 더 많을 것이다.

- fp_num(유동인구)이 높은 지역일수록 Case에서 발생빈도가 높을 것이다

6. 서울에서 기온이 낮고, 습도가 높을수록 감염 가능성이 더 클 것이다.

- Weather에서 서울의 avg_temp가 낮은 날일수록 확진자 수 증가할 것이다.
- Weather에서 서울의 avg_relative_humidity가 낮은 날일수록 확진자 수 증가할 것이다.

7. 정책 변화에 따라 확진자 수에 유의미한 영향이 있다.

- 10대 확진자는 초중고 학교 온라인 수업 기간(4/9-5/19)보다 오프라인 등교 기간(5/20-6/30)에 더 많다.
- 바이러스 확진자는 사회적 거리 두기 강화 기간(2/29-4/19)보다 완화 기간(4/20-6/30)에 더 많다. 사회적 거리 두기 정책이 유의미한 정책이다.

8. '코로나 바이러스' 검색량과 확진자의 수는 유의미한 관계가 있다.

- '코로나 바이러스' 검색량은 바이러스가 등장했던 1월에 가장 많을 것이다.
- '코로나 바이러스' 검색량이 전일 대비 급증한 date에는 코로나 확진 판정수 또한 전일 대비 급증했을 것이다.
- '코로나 바이러스' 검색량은 기온이 낮을수록 증가하고, 특정 Policy가 시행된 date에 증가할 것이다.