ITM 학부생 연구원 활동 결과 보고서

- 성명: 황재훈
- 학번: 17102070 (3학년)

국내 코로나 확진자 데이터 분석을 통한 지역간 감염 전파 시뮬레이션 연구

서론

저는 이번 방학동안 같은 연구실의 이성호 학생과 함께 국토교통부와 DACON 에서 개최하는 `국토교통 빅데이터 온라인 해커톤 경진대회` 에 참여하였습니다. 본 대회의 목적은 국토교통 데이터와 코로나 데이터 등을 융합분석하여, 국민의 안전한 이동을 위한 새로운 통찰과 창의적 아이디어를 도출하는 것입니다. 저희는 코로나 확진자 데이터와 코레일에서 제공하는 철도 이용량 데이터를 분석하여, 간선 철도 (경부, 호남) 의 이용이 코로나 바이러스의 지역간 전파에 미치는 파급력을 알아보고자 하였습니다.

저희의 분석 로드맵을 간단히 정라하면, (1) 코로나 데이터에서 감염자간 전파 경로를 분석, (2) 단순 감염자 및 최초 감염자가 평균적으로 1 주일 동안 몇명을 추가로 감염시켰는지 산출, (3) 이를 이용한 감염시킬 확률 분포 구함, (4) 철도 여객 수송 데이터를 분석하여 각 탑승 경로별 이동자수를 구함, (4) 확률 분포와 수송 데이터를 결합하여 특정지역에서 다른 지역으로 옮겨간 감염자가 그 지역에 코로나를 얼마나 전파시키는지 확인, 정도가 되겠습니다.

본론

코로나 데이터 전처리

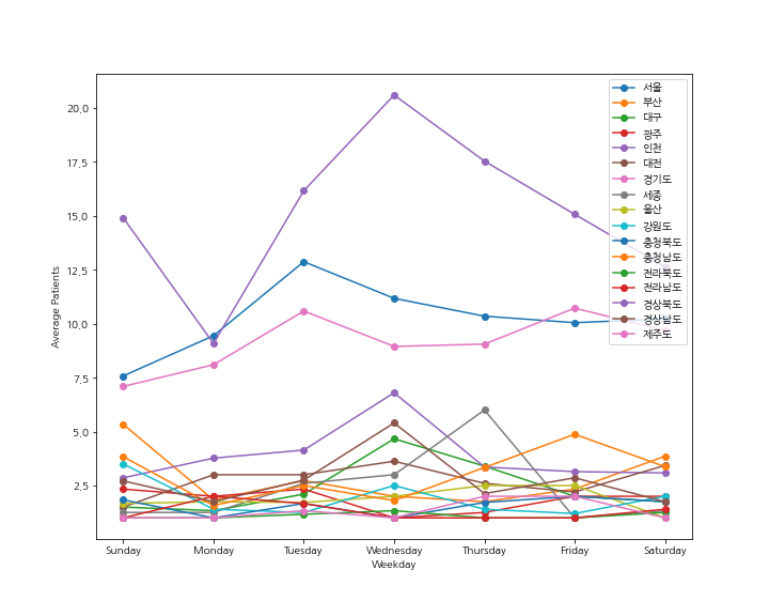
본 코로나 확진자 데이터는 2020년 1월 23일 ~ 6월 18일 사이에 확진된 환자 5164명의 기록으로 이루어져 있습니다. 각 환자는 고유 id 로 구분되어있고, 이들의 성별, 연령대, 국적, 주소, 감염 원인, 감염시킨 사람의 id, 확진날짜 등이 함께 기록되어 있습니다. 이중애 성별의 경우 약 1000 개, 나이대의 경우는 약 1500 가량의 관측치가 비어있었습니다만 다행히 환자의 소재지 같은 경우 null 값이 하나도 없었습니다.

안타까운점은 감염시킨 사람 id 가 null 이 아닌 경우, 즉 감염경로가 밝혀진 경우가 1346건 밖에 존재하지 않았다는 것입니다. 이마저도 데이터 내에 오류가 있어 하나하나 수정하는 작업이 필요했습니다. 대표적인 오류로는 감염은 시켰는데 환자 id 에는 없거나, 두명의 환자로 부터 한명이 감염된 경우, 자기자신에게 감염된 경우, 두명의 환자가 서로를 감염시킨 경우 가 있었습니다. 아마 데이터셋에 감염원인 환자를 한명만 특정해서 저장할수 밖에없고, 일부 단체로 감염된 환자들의 경우 정확한 감염원 환자를 찾기 어려웠던 부분이 한계로 작용했던것 같습니다. 특히 서로를 감염시킨 경우를 찾아본 결과, 대부분은 부부 또는 형제 관계였으며 이들이 거의 같은 경로로 이동하였기에 단순히 누가 먼저 감염됬는지는 알길이 없었습니다. 따라서 이경우에는 확실한 직업을 가진 사람을 감염원으로 설정하였습니다.

또한 데이터를 살펴보면서 일부 굵직했던 집단감염 사건들이 누락되있음을 알게되었습니다. 평택 군부대 와인바 집단감염의 경우 국적이 잘못 표기되있거나 해외유입으로 분류된 경우가 있었고, 천안 줌바댄스 강사 모임에서 시작된 집단감염은 워크숍에 참석했던 강사들이 전혀 이어져 있지 않고 단순 환자 접촉으로만 분류되어 있었습니다. 이 경우들은 기사에 나온 사실을 토대로 데이터를 수정하였습니다. 전처리 결과 감염 경로가 밝혀진 확진자는 1325명으로 감소 하였습니다.

코로나 데이터 분석

저희는 계보를 찾아내기 전에 우선 지역별로 나누어 요일 평균 확진자수를 알아 보았습니다. 데이터를 요일별로 나눈후, 지역별로 전체 확진자수 / 특정 요일이 포함된 횟수를 계산하였습니다.

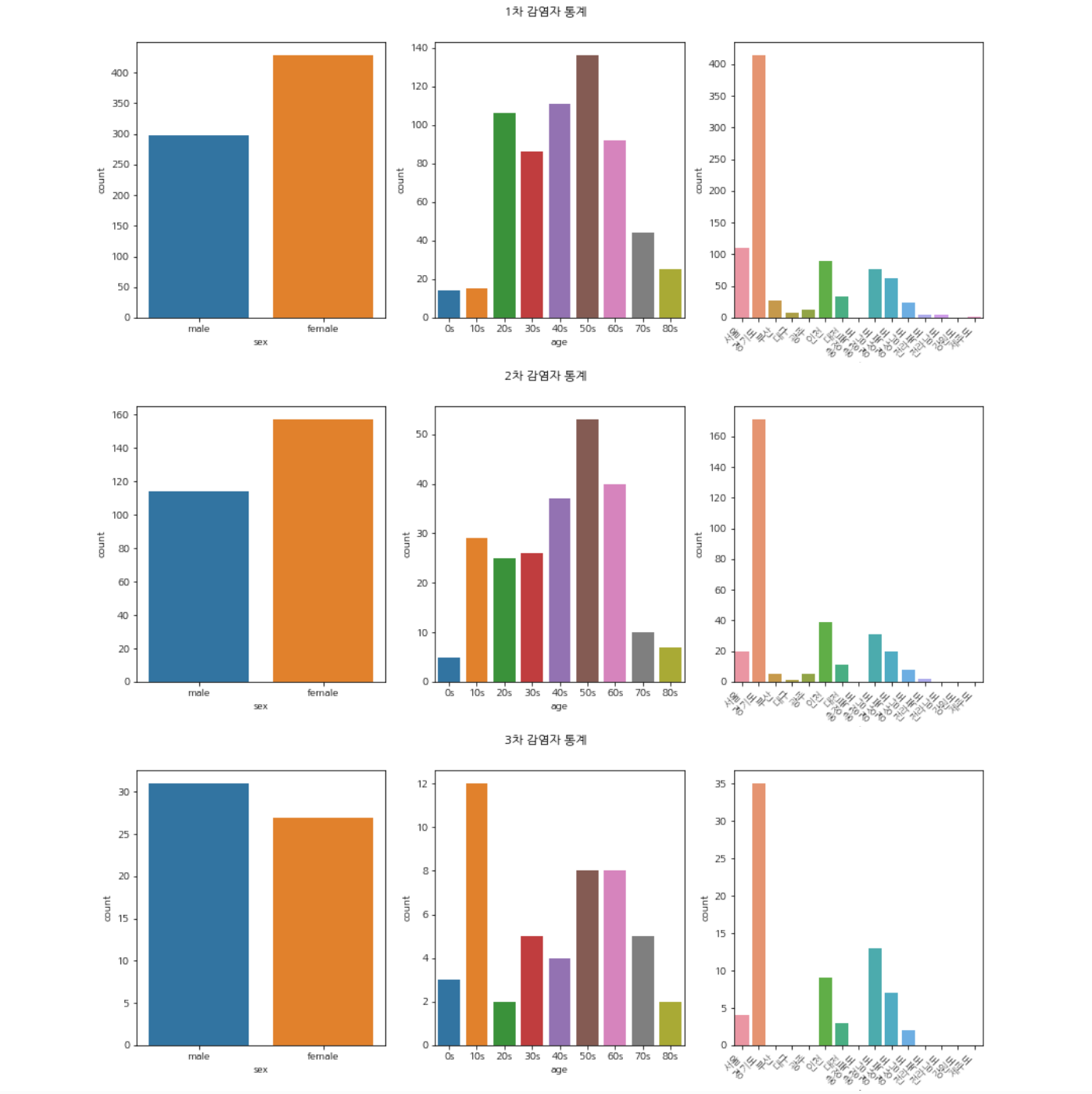


그래프를 보시면 경상북도, 서울, 경기도, 다른 지역 순으로 전체적인 확진자수 순위를 알수 있습니다. 실제로 전체 기간동안 데이터내 서울 확진자수 1312명으로 제일 많고, 그다음이 경상북도 1254명, 경기도가 1208명으로 3워, 인천이 343명으로 4위, 나머지 지역은 100명 전후의 확진자가 나타나 TOP3 와 나머지 간의 차이가 매우 큼을 알수있습니다. 여기서 주의해야 할 점은 경상북도 확진자수가 서울지역 확진자수보다 적게 나왔음에도 불구하고 요일별 평균 확진자수는 더 크다는 것입니다. 이는 그만큼 경상북도 확진자가 특정 기간에 몰려서 확진되었음을 뜻합니다. 또한 데이터가 수집된 기간 (~ 6월 18일) 동안 신천지로 인한 대구 & 경북 지역의 집단 감염이 극에 달했을 시기임을 감안했을때, 일부 데이터가 누락 되었음을 알수 있습니다. 실제로 6월 18일 기준 국내 누적 확진자는 12257명 이었으며 그중 대구의 확진자가 6896명으로 서울의 6배가 넘었고, 과반수 이상의 환자정보가 데이터셋에 포함되지 않았음을 확인하였습니다.

그 다음으로 저희는 재귀함수를 사용하여 환자들 간의 확진 계보를 구하였습니다. 그 결과 감염 경로의 꼭대기에 있는 최초 감염자 378명을 찾아낼수 있었습니다. 이들의 남/녀 비율은 거의 동일했으며, 20대가 58명으로 제일 많았고, 50대, 60대가 약 50명, 30대, 40대가 약 40명으로 대부분을 차지 하였습니다. 이러한 특징은 감염 원인으로부터 설명되는데, 원인의 대부분이 환자 접촉, 해외 유입, 쿠팡 물류 센터, 이태원 클럽, 신천지 로 이루어져 있었습니다. 다만 여기서도 이들을 단순히 최초감염자라고 보기에는 무리가 있습니다. 왜냐하면 같은 감염원인 내에서 감염경로가 밝혀진 경우가 거의 없기 때문입니다. 이는 집단 감염 사건을 하나의 시작점으로 보고 역학조사를 진행하기 때문에 그 전의 발생한 경로는 알기 힘들었던것으로 보입니다.

계보를 자세히 살펴보면, 최대 7차 감염까지 발생하였으며, 횟수로는 7차, 6차, 5차 감염이 각각 1건으로 제일 작았고, 2차 감염이 58건, 1차 감염이 287건으로 제일 많았습니다. 7차 감염의 경우 경북 예천에서 40대 여성에게서 시작되여 약 20일 동안 가족, 직장동료, 친구 등을 통해 빠르게 번져 37명이 감염되었으며 (기사), 6차 감염 사례의 경우 인천 부평구의 개척교회 모임 예배에서 시작되어 총 66명을 감염되었습니다 (기사). 계보별 총 감염 환자수는 인천 부평구 확진자와 서울시 3월 9일 확진자 를 통해 각각 66명씩 감염되어 최대였으며, 3월 8일 구로구 콜센터 첫번째 확진자로 부터 총 48명이 추가로 감염되어 2위를 차지했고, 평균적으로 계보 한건당 3.5명이 감염된 것으로 확인되었습니다 (중앙값, 최빈값 1명).

이어서는 n 차 감염자별 통계에 대해서도 알아보았습니다.



위 그래프 들은 1차 (900명) / 2차 (317명) / 3차 (74명) 로 감염된 확진자들의 인구통계학적 특성을 나타내고 있습니다. 우선 세 부분의 공통점은 경기도 환자가 압도적으로 많다는 것입니다. 이는 n 차 감염이 차수에 상관없이 수도권에서 높게 나타남을 나타내기 때문에 다른지역에 비해 수도권의 방역이 매우 중요함을 나타내는 지표입니다. 성별의 경우 1차와 2차에서는 여성 감염자가 더 많았지만, 3차 감염에서는 남성의 비율이 약간더 높게 나타났습니다. 나이대에서는 변화가 더 확연하게 보였습니다. 대부분의 경우에서 50, 60대 중장년층의 비율이 비중있게 나타났으며, 20, 30, 40대는 차수가 늘어날수록 그 비율이 상당히 감소하였습니다. 또한 3차 감염이 되면서 10대 확진자 비율이 크게 증가하였습니다. 이를 통해 사회에서 경제, 생산 활동을 하는 인구가 감염의 시작이며, 그 영향이 형제/자녀, 노년층 으로 퍼져나가는 양상이 보이므로, 생산인구의 적극적인 방역 참여가 꼭 필요함을 알수있습니다.

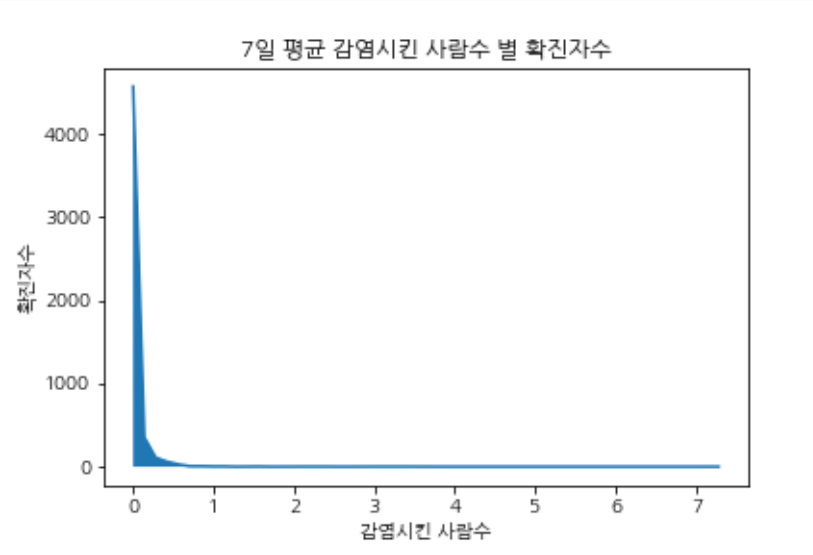
그 다음으로는 지역간의 감염을 설명할 확률 분포를 구할때, 특정 경로로 나눠서 하기 위해서 지역간 감염자수를 알아보았습니다. 우선 경로가 밝혀진 1325 건중 다른지역 사람에게 감염시킨 경우는 184 건으로 약 14% 밖에 차지하지 않았습니다.

```
Ori: 서울          Dest: 경기도        #: 110
Ori: 서울          Dest: 충청남도      #: 1
-----------------------------------------
Ori: 경기도        Dest: 충청남도      #: 2
Ori: 경기도        Dest: 서울          #: 1
-----------------------------------------
Ori: 부산          Dest: 경기도        #: 1
-----------------------------------------
Ori: 대구          Dest: 대전          #: 1
Ori: 대구          Dest: 울산          #: 1
Ori: 대구          Dest: 경기도        #: 4
Ori: 대구          Dest: 경상북도      #: 4
Ori: 대구          Dest: 경상남도      #: 1
-----------------------------------------
Ori: 광주          Dest: 전라남도      #: 2
Ori: 광주          Dest: 전라북도      #: 2
-----------------------------------------
Ori: 인천          Dest: 경기도        #: 30
-----------------------------------------
```

```
Ori: 대전          Dest: 서울          #: 1
Ori: 대전          Dest: 경기도        #: 4
Ori: 대전          Dest: 전라북도      #: 2
Ori: 대전          Dest: 세종          #: 2
Ori: 대전          Dest: 충청남도      #: 4
Ori: 대전          Dest: 대구          #: 1
-----------------------------------------
Ori: 세종          Dest: 충청남도      #: 1
-----------------------------------------
Ori: 울산          Dest: 전라남도      #: 1
Ori: 울산          Dest: 서울          #: 1
-----------------------------------------
Ori: 충청남도      Dest: 대전          #: 2
-----------------------------------------
Ori: 전라북도      Dest: 대전          #: 1
-----------------------------------------
Ori: 경상북도      Dest: 경상남도      #: 1
Ori: 경상북도      Dest: 울산          #: 1
-----------------------------------------
Ori: 경상남도      Dest: 부산          #: 1
Ori: 경상남도      Dest: 충청남도      #: 1
-----------------------------------------
Total: 184
```

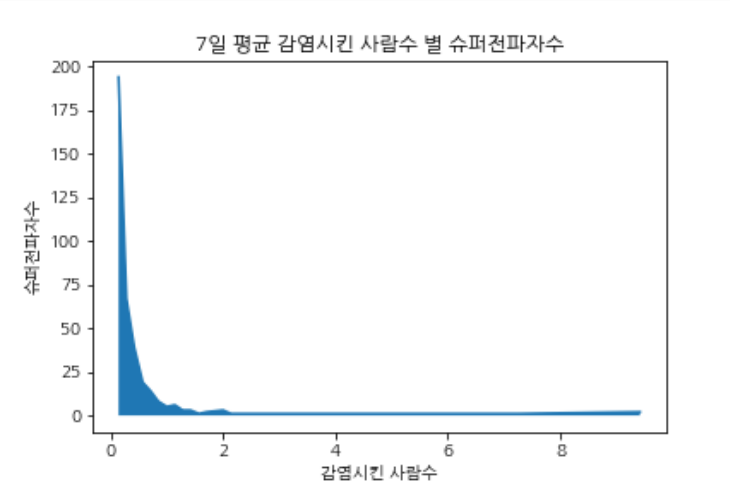
대부분의 경우가 서울, 경기, 인천 사이에 발생한 통학/통근 범위내에서 발생한 감염이고, 지역간 감염은 대구 ~ 경기, 대전 ~ 경기에서 각각 4건 정도로 매우 미미한 수준으로 나타났습니다. 이중 대구 4건은 신천지 교인인 대구 31번째 확진자가 증상이 나타난후 약 10일간 병원, 호텔등을 방문 (기사) 했을때 접촉했던 사람들이 경기도에 있는 거주지로 돌아간후 확진판정을 받았으며, 대전 4건의 경우 대전 다단계 판매업에 관련된 환자 2명과 각각 접촉하면서 확진된 것으로 보입니다 (기사1, 기사2). 이들은 공통적으로 다중이용시설을 아무렇지 않게 이용하였고, 같은지역 내부에도 많은 확진자를 발생시켰습니다. 한 가지 다행이었던 점은 지역간 감염전파에서 KTX 이용했던 사례를 발견한 것입니다. 대전 ~ 서울간 전파 (1건) 는 대전 2번 확진자가 KTX 를 이용해서 서울 관악구에 있는 부동산을 방문하면서 발생했습니다. 대부분의 경우 이용한 교통수단을 찾을수 없는 경우가 많았지만, 이를통해 철도를 통해서 감염전파가 일어날수 있음을 알수 있었습니다. 다만 데이터가 너무 적은 관계로 경로별로 확률 분포를 구해서 사용하기에는 한계가 있었습니다.

마지막으로는 7일 평균 감염시킨 사람수별로 환자들을 나눠서 확률분포를 구하였습니다.




```
p = ~0.8855        0.0명
-----------------------------
p = ~0.9539        0.14명
-----------------------------
p = ~0.9752        0.28명
-----------------------------
p = ~0.987         0.42명
-----------------------------
p = ~0.9922        0.57명
-----------------------------
p = ~0.9936        0.71명
-----------------------------
p = ~0.9949        0.85명
-----------------------------
p = ~0.9957        1.0명
-----------------------------
p = ~0.9965        1.14명
-----------------------------
p = ~0.9967        1.28명
-----------------------------
p = ~0.997         1.42명
-----------------------------
p = ~0.9976        1.57명
-----------------------------
p = ~0.9978        1.71명
-----------------------------
p = ~0.9982        2.14명
-----------------------------
p = ~0.9986        2.42명
-----------------------------
p = ~0.9988        2.57명
-----------------------------
p = ~0.9992        3.0명
-----------------------------
p = ~0.9996        3.42명
-----------------------------
p = ~0.9998        3.85명
-----------------------------
p = ~0.9999        7.28명
-----------------------------
```

왼쪽의 그래프는 전체 확진자 5164명을 기준으로 각 확진자가 감염시킨 사람의 수를 세서 나타낸 것입니다. 감염 시킨 사람 id 에 나타나지 않은 확진자는 아무도 감염시키지 않은것 (0명) 으로 처리하였습니다. 자세히는 0.14 명을 감염시킨 사람이 353명, 0.28명이 110명 이었으며,1명이 최대 7.28명을 감염시킨것으로 나타났습니다. 이는 전체 환자의 약 90% 가량이 무전파자로서 전파 확률 분포가 매우 치우쳐져있음을 나타냅니다.



p = ~0.5132	0.14명	p = ~0.9656	2.14명
p = ~0.6904	0.28명	p = ~0.9682	2.42명
p = ~0.7936	0.42명	p = ~0.9708	2.57명
p = ~0.8439	0.57명	p = ~0.9735	2.71명
p = ~0.8809	0.71명	p = ~0.9761	2.85명
p = ~0.9021	0.85명	p = ~0.9788	3.14명
p = ~0.9153	1.0명	p = ~0.9814	3.28명
p = ~0.9312	1.14명	p = ~0.9841	4.71명
p = ~0.9391	1.28명	p = ~0.9867	4.85명
p = ~0.947	1.42명	p = ~0.9894	5.28명
p = ~0.9497	1.57명	p = ~0.992	6.85명
p = ~0.955	1.71명	p = ~0.9947	7.28명
p = ~0.9629	2.0명	p = ~1.0	9.42명

또한 확진자가 다른 청정지역으로 이동해서 그 지역의 감염을 촉진할 가능성이 있으므로, 위에서 잠깐 설명했던 최초 감염자 (슈퍼전파자) 378명 만을 가지고도 분포를 나타내 보았습니다. 슈퍼전파자수의 비율은 기존의 분포와 거의 일치하는 수준이었지만 (0.14 명 감염시킨 사람 194명, 0.28명이 67명 등), 감염시킨 사람수가 n차 감염까지 포함된 수치이므로 최대 9.42명 까지 감염시키는 경우가 발견되었습니다. 다만 이 분포의 경우 감염을 시키지 않는 상황은 가정하고 있지 않기 때문에, 이후 시뮬레이션에서는 기존 분포를 사용하기로 결정하였습니다.

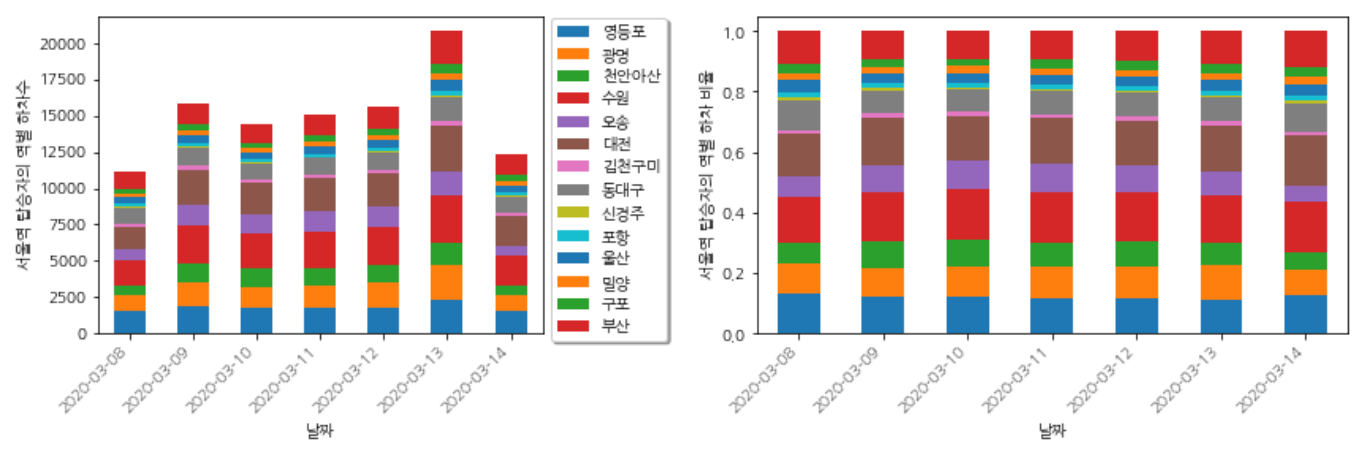
한국철도공사 역별 승하차 데이터 분석

저희는 특정기간에 특정지역간 철도를 이용해 이동한 사례수를 구하기 위해 간선철도역별 승하차 데이터를 추가로 분석하였습니다. 이 데이터 셋은 2020년 1월 1일 부터 5월 31일 사이에 매일 코레일에서 운영하는 243개의 철도역에 대한 승/하차 이용자수를 기록한 것입니다. 저희는 대도시 중심에 있는 수요가 충분한 역에 대해서 알아보기위해, 우선 역의 범위를 경부선에 속하는 역으로 한정하였고, 그중에서도 KTX 정차역만 뽑아내었습니다. 철도여객수송에서 경부선과 KTX 가 대부분을 차지한다는 근거는 국토교통부 철도통계 (KRIC) 웹페이지 에서 찾을수 있었는데, 1월부터 5월 사이 전체 철도 이용자 약 3661만명 중 KTX 이용자가 약 1766만명으로 48%의 비중을 차지하였으며, 전체 노선중 경부선의 노선 부담률이 44.88%로 두번째로 분담률이 높은 호남선의 12.88%에 비해 약 4배가량 많은것을 알수 있었습니다.

문제는 승하차 데이터에는 승객이 어떤역에서 승차하고 하차했는지에 대한 경로를 알수 없다는 것입니다. 인터넷에서는 해당 데이터를 구할 수 없었기 때문에, 저희는 역별 승하차 비율을 활용하였습니다.

> 예를 들어 경부선에 역이 `서울`, `대전`, `동대구`, `부산` 만 존재한다고 가정하고 `서울` → `부산` 여객 이용량을 구해보도록 하겠습니다.
>
> **(공식)** 서울역 승차 & 부산역 하차한 인원수 = 서울역승차인원수 × $\frac{\text{부산역하차인원수}}{\text{대전역하차인원수} + \text{동대구역하차인원수} + \text{부산역하차인원수}}$
>
> 출발역 승차인원수에 나머지 노선의 하차 인원수중 도착역이 차지하는 비율을 곱하여 중간에 하차한 인원을 고려하였습니다.

위방식을 토대로 3월 8일 (일요일) ~ 3월 14일 (토요일) 일주일 동안 서울역 탑승자의 다른역 하차 횟수와 비율들 뽑아보았습니다.



요일별 건수를 살펴보면 주말 (8일, 14일) 에는 탑승자가12000명 가량으로 평소보다 적게 나타났고, 월 ~ 목요일은 15000명 전후를 유지하였으며, 금요일 (13일) 탑승자가 20848명으로 다른날에 비해 5000명 이상 급격히 증가한것을 알수있었습니다. 그런데 하차수에는 어느정도 변화가 있었지만, 하차역간 비율은 일별로 거의 동일했습니다. 대부분의 인원이 수원 (16%), 대전 (14%), 동대구 (8%), 부산 (10%) 에 집중되어 있었으며, 금요일의 경우에도 비슷한 비율을 보여주었습니다. 이때 평일에는 수도권역 (영등포 ~ 수원) 하차 비율이 높고 주말에는 타지역 비율이 약간 높게 나타나기는 했으나 수치가 1 ~ 2% 정도 밖에 차이나지 않았으므로 의미있다고 보기에는 힘들었습니다. 다만, 이 한 주간은 특별한 이벤트가 없었고, 이후의 3월 셋째주나 넷째주의 상황이 비슷한것 또한 고려했을때, 금요일에는 기차를 이용해 여행을 가는 사람들이 많다고 추측해볼수 있습니다. (이외에 다른 특별한 이유를 찾지 못했습니다.)

여기까지 시뮬레이션을 위한 분석을 모두 마쳤습니다.

시뮬레이션

지금 부터는 앞부분에서 알게된 내용을 기반으로 7일간의 시뮬레이션을 직접 진행해 보도록 하겠습니다.

(기본 가정)

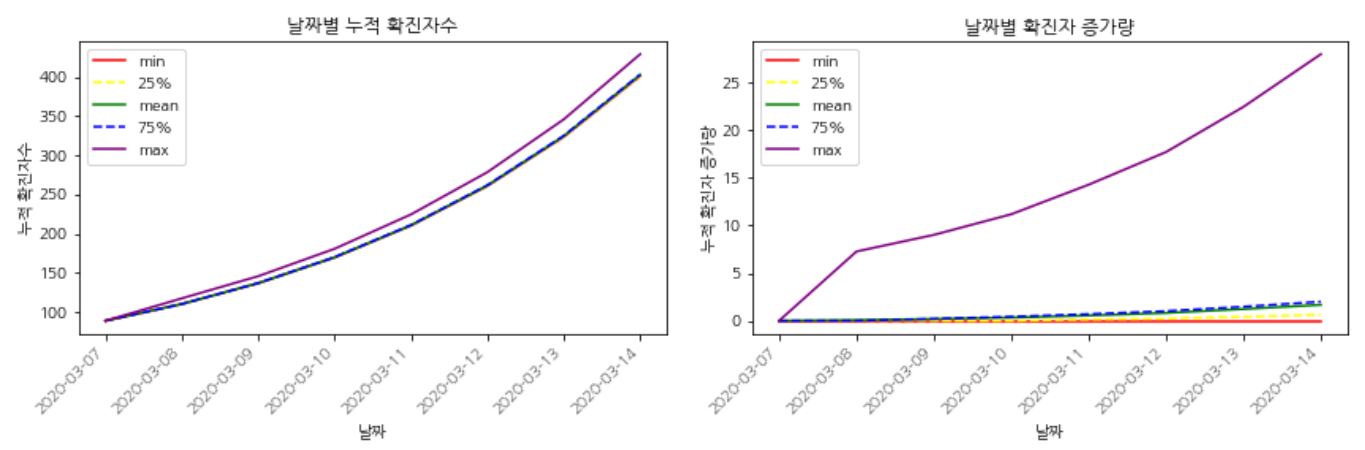
- 기간: 3월 8일 ~ 3월 14일 (7일, 가장 평범한 특별한 이벤트가 없는 일주일)
- 장소: 서울 (서울역) ~ 부산 (부산역)

(과정)

- 2020년 3월 기준 서울시 인구는 약 970만명 입니다.
- 3월 8일 기준 전날 7일까지 서울시 확진자는 120명으로 전체 서울시 인구의 0.0012% 입니다.
- 앞에서 계산한 비율 대로라면, 3월 8일 서울역에서 기차를 탑승하여 부산역에서 하차한 사람은 1221명입니다.
- 따라서 저희는 기차를 이용한 1221명 중에서 0.0012%, 즉 올림해서 1명을 코로나 감염자라고 가정하였습니다.
    - 반올림을 할경우 7일 내내 탑승한 감염자가 0명이 되므로, 부득이하게 올림을 선택하였습니다.
- 감염자로 뽑힌 1명은 각자 0 ~ 1 사이의 랜덤값 (확률) 을 부여받고, 그 확률을 이전에 구해놓은 감염 확산 확률분포에 대응하여 몇명을 감염시킬지 정하게 됩니다.
    - 만약 랜덤값으로 0.9998이 뽑히게 될 경우 이 환자는 3.85명을 감염시키는 것으로 정해집니다.
- 단순히 서울에서 내려온 환자들 이외에 기존 부산에 있었던 환자들로부터 또한 감염이 발생할 것이므로, 부산 지역내에서 일어나는 감염도 함께체크합니다.
    - 2020년 3월 기준 부산시 인구는 약 341만명 입니다.
    - 3월 8일 기준 전날 7일까지 부산시 내의 확진자는 89명으로 전체 부산시 인구의 0.003% 이었습니다.

    - 전체 기간 (1월 23일 ~ 6월 18일) 에서 부산시 확진자는 151명 발생하였으며, 이중 감염경로가 밝혀진 사람은 32명이었고, 그중 같은 부산사람을 감염시킨 경우가 31명이었습니다.
    - 31명을 가지고 알아본 결과, 부산내 확진자 한명이 다른 시민을 감염시키는 명수는 하루평균 1.19명 이었으며, 나머지 120명을 아무도 감염시키지 않은 사람 (0명 전파) 이라고 가정하고 포함시켰을때는 하루평균 0.24명 이었습니다.
        - 각 확진자가 전파한 사람수를 전파된 날짜 횟수로 나눈후 평균내었습니다.
        - 외부인을 감염시킨 1명의 경우 부산내 감염이 아니기 때문에 아무도 감염시키지 않은것으로 하였습니다.
- 결과적으로 확진자가 하루동안 $(3.85) + (89 \times 0.24) = 25$ 명 증가한것으로 볼수있습니다. (누적 확진자 약 114명)
- 이러한 방식으로 7일 간의 확진자 수를 시뮬레이션 하여, 얼마만큼 증가하였는지 알아보았습니다.
- 일부 서울에서 온 확진자수가 줄어들 경우 (타지에서 온 사람이 아무도 감염을 시키지 않을 경우) 얼마만큼 확진자가 증가하는지화 비교해 보았습니다.

(결과)



시뮬레이션을 1000회 반복한 결과, 감염확률이 매우 낮은 관계로, 제 3 사분위수 까지는 기차를 통해 감염된 확진자수가 현저히 적었으며, 대부분의 감염은 같은 지역 내에서만 발생하였습니다. 다만 이례적으로 기차를 이용한 사람이 슈퍼감염자일 경우 (max 일때), 기존에 비해 최대 26명까지 추가로 확진자가 나오는 경우도 있었습니다. 하지만 이마저도 7일후 누적확진자 429명의 6% 수준으로 많은 부분을 차지한다고 보기 힘들으로 보입니다.

결론

시뮬레이션 결과를 토대로 정리하자면 간선철도 이용량은 지역간 감염에 크게 영향을 미치지 않는것으로 나타났습니다. 그러나 단순히 이렇게 결론짓기에는 문제점이 많습니다. 우선 경부선 일부구간에 특정 시간만을 이용해서 실험을 했기때문에 일반화하기위한 근거가 부족합니다. 또한 시뮬레이션 설계에서 집단감염이 발생했던 시점과 하지않았던 시점을 나눠서 분포를 구하는등 상황을 좀더 구체화할 필요가 있어보입니다. 물론 단순히 실험의 문제만은 아닙니다. 앞에서 설명했다시피 실제 확진자의 반 이상이 데이터 셋에서 누락되어 있었으며, 확진자간 감염경로를 알수있는 데이터도 전체의 20% 수준으로 매우 작았고, 그 마저도 수도권에 치우쳐져 있어서 지역간 감염사례를 조사하기 쉽지 않았습니다. 확실한 분석을 위해서는 좀더 고도화된 데이터가 절실해 보입니다. 애초에 아이디어 문제였을수도 있습니다. 국내 코로나 발병 시점부터 지금 (9월) 까지의 확진자 기사를 찾아보면 (전체를 찾아보지는 않았습니다만) 확진자가 기차를 이용했다는 내용을 거의 찾아볼수가 없습니다. 이는 물론 확진자가 실제로 기차를 안탔을수도 있고, 역학조사에서 누락됐을수도 있습니다. 또한 철도 이용자수도 급격히 감소했는데, 2020년 3월 철도 전체 수송인원은 약 400만명으로, 전년 동기 1200만명에 비해 1/3 으로 감소하였습니다. 애초에 감염병이라는 것이 사람들이 많이몰리면 감염자수가 증가하는것인데, 이런 사실을 미리 알았더라면 차라리 다른 분야에서 사람이 많이 몰리는 경우를 분석하는 것도 나쁘지 않았을것 같다고 생각합니다.

그래도 이번기회를 통해 데이터를 다양한 방법으로 분석해보고 시각화해볼수 있어서 개인적으로는 좋은 경험이 됐다고 생각합니다. 감사합니다.