# COVID-19

코딩해조

팀원:김도경, 윤준영, 이호진, 황예진

## **Table of Contents**

**01** 팀 소개 및 역할

**02** 데이터 시각화

**03** 프로젝트 진행 배경 및 가설

**04** 상관관계 분석 **05** 결론

**06** 해커톤 후기

# 01. 팀 소개 및 역할

### 01. 팀 소개 및 역할

김도경

데이터 전처리

데이터 시각화

이호진

데이터 시각화

추가 데이터 수집

발표

윤준영

데이터 분석

황예진

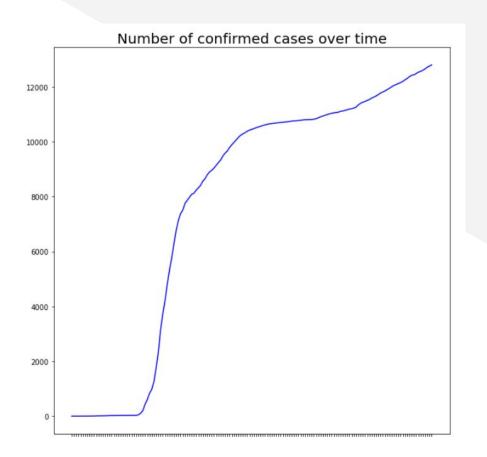
데이터 시각화

발표자료 준비

□ 0 🕶 🖿 / aiffel / Hackathon / data
□
☐ Case.csv
☐ PatientInfo.csv
☐ Policy.csv
☐ Region.csv
☐ Region_final.csv
☐ SearchTrend.csv
☐ Seoul_covid2021.csv
☐ SeoulFloating.csv
☐ Time.csv
☐ TimeAge.csv
☐ TimeGender.csv
☐ TimeProvince.csv
☐ ☐ Weather.csv

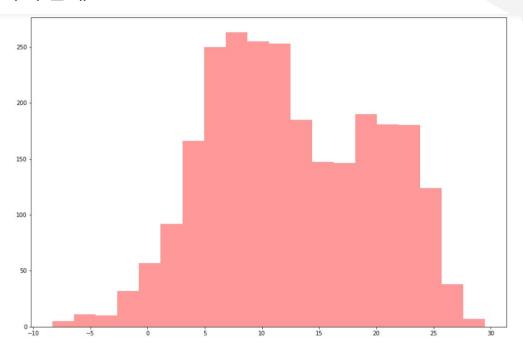
시간의 흐름에 따른 코로나 확진자 수

데이터셋기간: 2020.01.20~2020.06.30



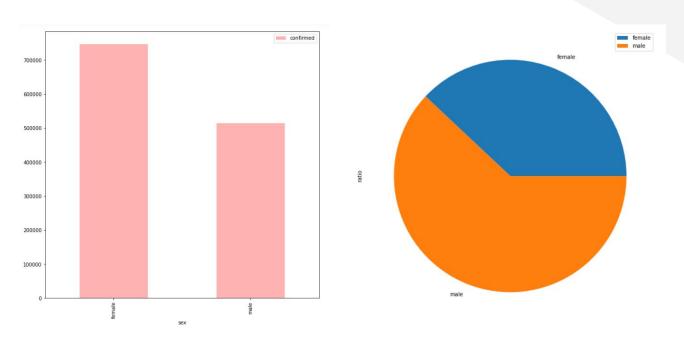
2020년 2월에 발발한 대구 신천지교회 집단감염으로 인해 확진자 수가 급증했다.

기온과 확진자 수와의 관계



기온과 확진자 수는 유의미한 관계가 없음을 확인 할 수 있었습니다.

성별과 치사율과의 관계



확진자 수는 여성이 남성보다 많았지만, 치사율은 반대

#### 동아사이언스

코로나19 감염자 생사 가르는 2가지 요인 '나이' '기저질환'



신종 코로나바이러스 감염증(COVID-19·코로나19) 바이러스. 미국 MIT 제공



기저질환과 나이가 신종 코로나바이러스 감염증(COVID-19·코로나19)으로 인한 사망에 큰 영향을 미친다는 사이이 여그로 인증됐다.



다는 사실이 연구로 입증됐다.



서울대병원은 이호진 마취통증의학과 교수 연구팀이 건강보험심사평가원이 공개한 자료를 토대로 기저질 환과 나이가 코로나19로 인한 사망에 미치는 영향을 분석한 결과를 국제학술지 '사이언티픽 리포트' 18일 자에 공개했다고 31일 밝혔다.



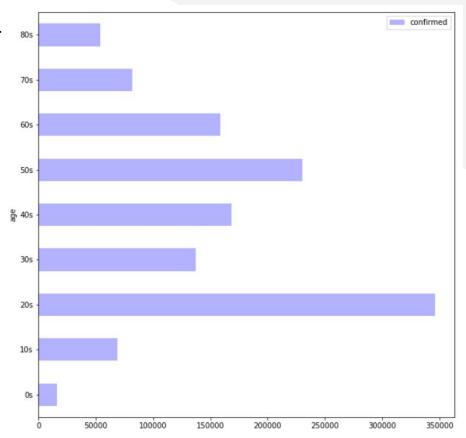
연구팀은 기저질환과 나이가 코로나19 환자의 사망에 얼마나 영향을 줬는지 알아보기 위해 '연령보정 찰슨 동반상병지수'를 활용했다. 연령보정 찰슨동반상병지수는 환자가 앓고 있는 기저질환이 환자의 사망에 미치는 영향을 나타낸 지표다. 심근경색과 심부전, 당뇨병, 치매, 간질환, 신장질환 등 19개 동반질환이 사망에 미친 영향을 점수로 나타낸다.

-젊은 사람들에 비해서, 연령이 많은 분들의 치사율이 더 높다??

-> 나이와 치사율에 대한 데이터 분석을 진행

나이와 치사율과의 관계

- 연령대별 코로나 확진자 수

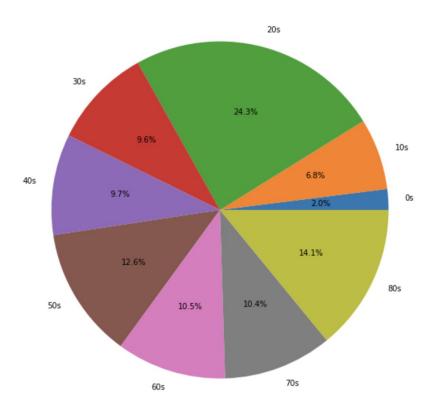


<연령대별 코로나 확진자 수>

나이와 치사율과의 관계

- 연령대별 코로나 확진자 수

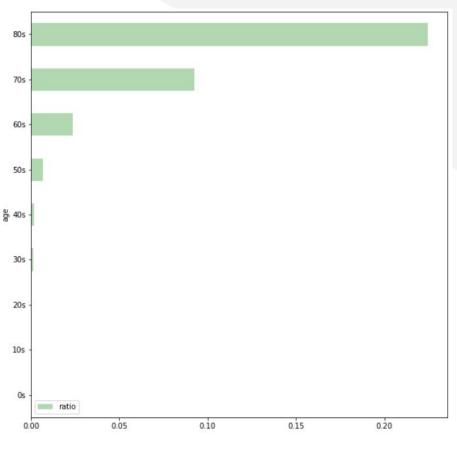
The ratio of COVID-19 confirmed patients to the number of people in each age group.



<각 연령층 대비 코로나 확진자 수의 비율>

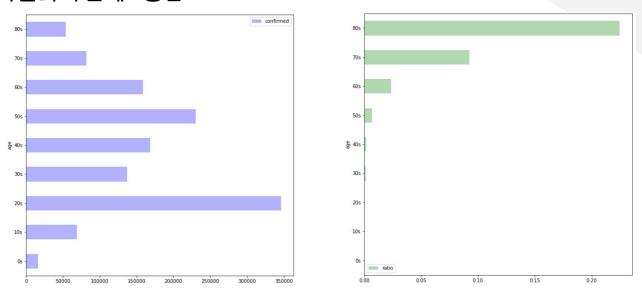
나이와 치사율과의 관계

- 연령대별 치사율



<연령대별 치사율>

나이와 치사율과의 관계 - 종합

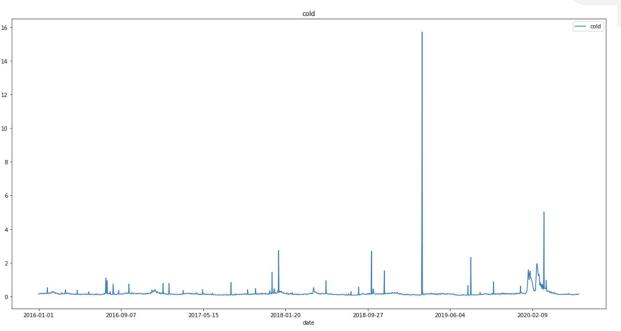


20대에서 확진자 수가 가장 많았으나, 80대에서 치사율이 가장 높았다.

연령대가 높아질수록 치사율이 높아진다.

### 검색량추이

#### cold (감기) 에 대한 검색량을 보여주는 그래프



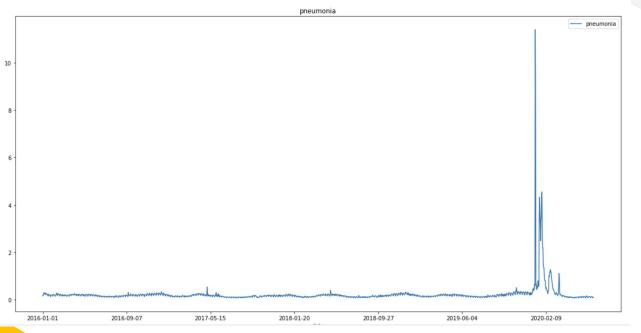
cold

max:16

2016-2020년까지 검색량 -> 사람들의 관심도 지표

검색량추이

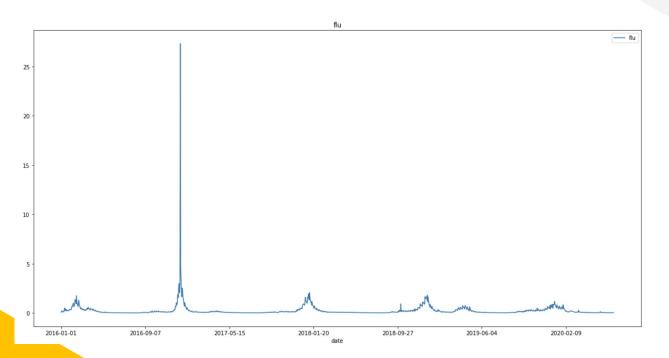
#### pneumonia (폐렴)에 대한 검색량을 보여주는 그래프



pneumonia max:15

검색량추이

flu(독감)에 대한 검색량을 보여주는 그래프

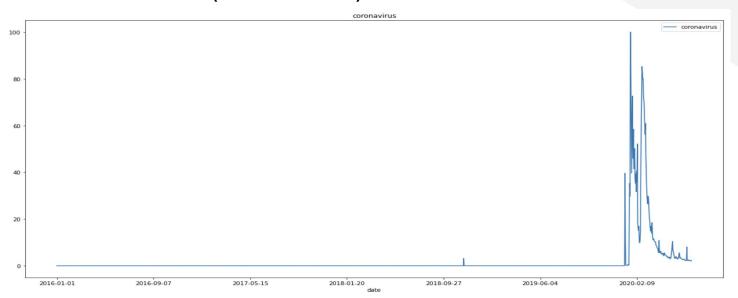


pneumonia max:30

검색량 추이

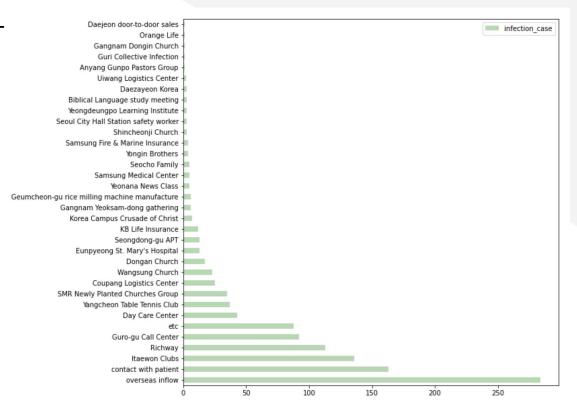
최근 :10이하/즉 감기 정도의 관심도로 변해가는중

### coronavirus(코로나 바이러스)에 대한 검색량을 보여주는 그래프



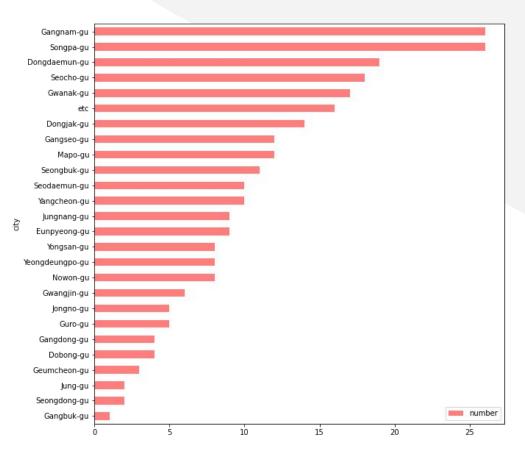
<mark>코로나</mark> 바이러스에 대한 검색량이 줄어드는 것으로 보아 **사람들의 관심도는 줄어드는 것**으로 보이며, 다른 바이러스들 처럼 특정 시기에만 **패턴적으로 관심도가 증가할 것으로 예상** 

서울 내 코로나 감염 원인 분석



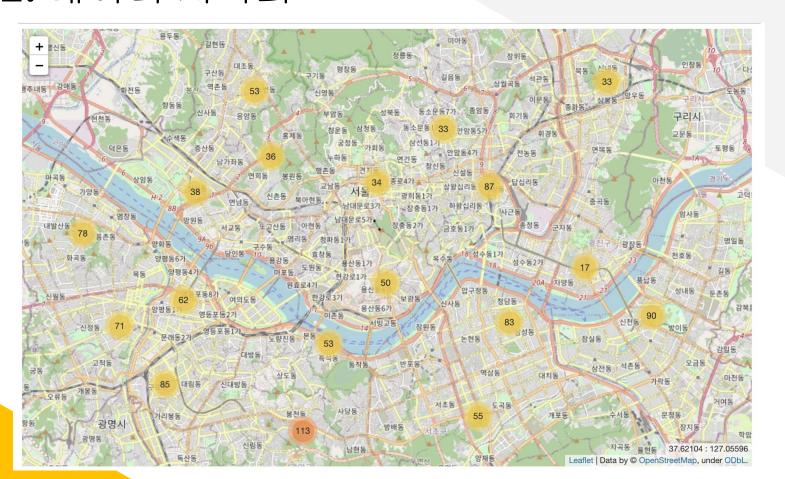
overseas inflow, contact with patient, Itaewon Club 순으로 많다.

서울 구별 확진자 수

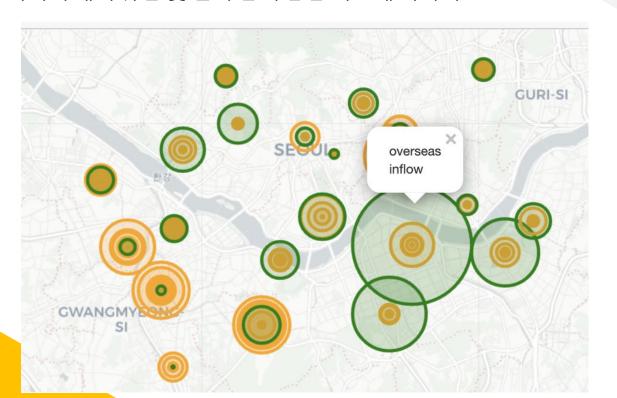


확진자 수가 많은 지역은 강남구, 송파구, 동작구, 서초구 순으로 많다.

## 02. 데이터 시각화 서울시구별 확진자 수를 지도로 시각화



서울 지역의 해외 유입 및 환자 접촉 감염 지도에 시각화



overseas inflow

contact with patient

# 03. 프로젝트 진행 배경 및 가설

### 03. 프로젝트 진행 배경 및 가설

### 프로젝트 진행 배경

- 사회적 거리두기의 효과 및 정부에서 규제를 가한 <sup>88</sup> 장소가 확진자 수와 연관이 보이는 지 궁금증을 가졌다.

#### ■ 사회적 거리두기 4단계 요약(10.18.~10.31.)

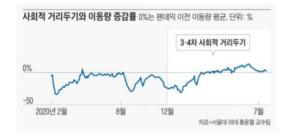
구분	변경
사적모임	- <u>시간 구분 없이</u> 미경증거 4 <u>인</u> - 모든 다중이용시설 등에서 집중 완료자 포함 <u>8인</u> 까지 가능 - 예방검증 원료자 는 2차 검증 후(안선 백신은 1회) 연역 형성 기간인 14일이 경과한 사람을 의미
운영시간	- 독서실, 스터디카페, 광연장, 영화관 24시 운영 제한 - 방문판매 동을 위한 직접 판매 홍보관운영시간 제한 해제
스포츠 관람	집중 완료자로만 구성 시 <u>수용인원의 20%(설비), 30%(설외)</u> 까지 가능 • 야구, 축구, 배구, 농구 등 스포츠 관망 해당
스포츠 대회	- 접종 완료자 등 • 으로 최소 인원 참여하는 경우 개최 하용 • 참여 연형 및 전체 규모에 따라 PCR 음성 확인자(48시간 내)도 인청
종교활동	실대 99인 실한 해계 하여 전체 수용인원 10%까지 가능 또는 <u>결혼 원로자로만 구성 100%</u> 까지 가능 소민생명사사식 학급 지
숙박시설	- 객실 운영 제한 해제

#### \* 다중이용시설 그룹별 분류

구분	주요 시설
1그룹	△유흥시설, △홀덤펍·홀덤게임장, △콜라텍·무도장
2그룹	$\triangle$ 노래연습장, $\triangle$ 식당ㆍ카페, $\triangle$ 목욕장업, $\triangle$ 방문판매 등을 위한 직접판매홍보관, $\triangle$ 실내체육시설(고강도ㆍ유산소)
3그룹	소영화관·공연강, △학원 등, △결혼식강, △강례식장, △이미용업, △PC방, △오락실·멀티방, △독서실·스터디카페, △ 높이공원·워터파크, △상정·마트·백화점(300㎡ 이상) △카지노 △실내체육시설(고강도·유산소 외)

서울대 연구팀 "사회적 거리두기 효과 소멸… 4단계에도 이동량 변화 거의 없어"

송주상 기자 입력 2021 08 23 22:18



23일 채널A 보도. 홍윤철 교수 연구팀의 연구 결과를 바탕으로 거리두기 조치를 기준으로 날짜별 시람들의 이동량을 나타냈다. 이 표에 따르면, 1·2차 사회적 거리두기 조치(빨강) 때는 시행 직후 이동량은 감소했다. 반면, 3차와 최근 4차 사회적 거리두기 조치(파랑) 때는 시행 전후로 이동량 변화가 거의 없다. 또 3차와 4차 사이 이동량이 증가하며 팬데믹 이전 평균 이동량을 넘어섰다.

사회적 거리두기 규제 효과가 소멸했다는 연구 결과가 국내 연구진에게서 나왔다.

23일 사회적 거리두기와 이동량 추이에 관한 흥윤철 서울대 의대 예방의학교실 교수 연구팀의 연구 결과에 따르면 정부의 거리두기 조치가 거듭될수록 실제 이동량 감소에 뚜렷한 영향을 주지 못하는 것으로 밝혀졌다고 채널 A가 보도했다.

흥 교수 연구팀은 구글 위치기록 데이터 등을 바탕으로 사회적 거리두기 조치에 따라 사람들 이 얼마나 이동하는지 비교했다.

### 03. 프로젝트 진행 배경 및 가설

#### 서울로 한정 지은 이유

- 2021년 10월 기준 대한민국 내에서
- 누적 확진자의 수가 가장 많기 때문

• 지도 • **지역별 표** • 누적확진 추이 • 일별 신규 확진자수

지역	누적확진자	신규확진자
서울	113,891	519 ▲
경기	102,965	518 ▲
인천	18,320	121 🔺
대구	16,588	21 🛦
부산	13,541	44 🔺
경남	12,514	34 🔺
충남	10,100	27 🔺
경북	9,393	45 ▲

( 1/3 >

국내현황 2021.10.22. 00:00 집계 기준 🛈

가설 1:정부가 규제한 장소와 코로나 바이러스의 감염의 연관이 미미할 것이다.

(유흥 업소 수, 식당 수, 카페 수, 체육 시설 수와 확진자 수 간의 관계를 중심으로)

## 데이터 프레이밍

```
In [128]: st 1=starbucks[starbucks['address'].str.contains(province)]
        st 1.shape[0]
        # st 1.count()
Out[128]: 80
In [119]: list province =['강남구','강동구','강서구','강북구','관악구','광진구','구로구','금천구','노원구','동대문구','동대문구','동작구','마포구','서대문구
 In [ ]: list province_eng =['강남구','강동구','강서구','강북구','관악구','광진구','구로구','금천구','노원구','동대문구','도봉구','동작구','마포구','서
In [120]: list num=[i for i in range(len(list province))]
In [140]: st list=[0]*25
In [141]: st list
In [143]:
        for i in range(25):
           province=list province[i]
            st 1=starbucks[starbucks['address'].str.contains(province)]
           num=st 1.shape[0]
            st list[i]=num
In [136]: list province[0]
Out[136]: '강남구'
```

```
[145]: star dict=dict(zip(list province,st list))
146]: star dict
[146]: {'강남구': 80,
       '강동구': 14,
       '강서구': 20,
       '강북구': 5,
       '관악구': 11,
       '광진구': 15,
       '구로구': 10,
       '금천구': 10,
       '노원구': 13,
       '동대문구': 8,
       '도봉구': 2,
       '동작구': 11,
       '마포구': 32,
       '서대문구': 18,
       '성동구': 11,
       '성북구': 13,
       '서초구': 46,
       '송파구': 30,
       '영등포구': 34,
       '용산구': 19,
       '양천구': 15,
       '은평구': 8,
       '종로구': 38,
       '중구': 86,
       '중랑구': 6}
```

```
print(pd.DataFrame.from dict({
  '강남구': 80,
'강동구': 14,
'강서구': 20,
'강북구': 5,
'관악구': 11,
'광진구': 15,
'구로구': 10,
'금천구': 10,
'노원구': 13,
'동대문구': 8,
'도봉구': 2,
'동작구': 11,
'마포구': 32,
'서대문구': 18,
'성동구': 11,
'성북구': 13,
'서초구': 46,
'송파구': 30,
'영등포구': 34,
'용산구': 19,
'양천구': 15,
'은평구': 8,
'종로구': 38,
'중구': 86,
'중랑구': 6}, orient='index').rename(columns={0:'starbucks num'}))
```

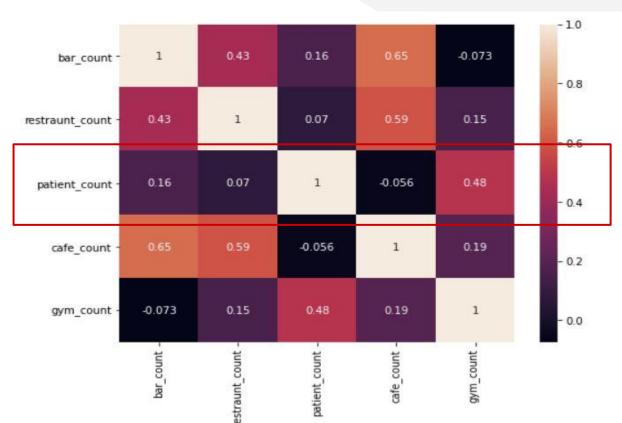
	province	city	bar_count	restraunt_count	patient_count	cafe_count	gym_count
1	Seoul	Gangnam-gu	179.0	380.0	83.0	80.0	292.0
2	Seoul	Gangdong-gu	116.0	257.0	35.0	14.0	64.0
3	Seoul	Gangbuk-gu	64.0	248.0	24.0	5.0	130.0
4	Seoul	Gangseo-gu	126.0	411.0	78.0	20.0	166.0
5	Seoul	Gwanak-gu	188.0	309.0	113.0	11.0	158.0
6	Seoul	Gwangjin-gu	14.0	254.0	17.0	15.0	104.0
7	Seoul	Guro-gu	32.0	504.0	85.0	10.0	156.0
8	Seoul	Geumcheon-gu	48.0	274.0	32.0	10.0	88.0
9	Seoul	Nowon-gu	21.0	376.0	43.0	13.0	232.0
10	Seoul	Dobong-gu	21.0	247.0	62.0	2.0	100.0
11	Seoul	Dongdaemun-gu	67.0	266.0	38.0	8.0	52.0
12	Seoul	Dongjak-gu	47.0	269.0	53.0	11.0	122.0
13	Seoul	Ma <mark>po-g</mark> u	51.0	504.0	38.0	32.0	127.0
14	Seoul	Seodaemun-gu	29.0	388.0	36.0	18.0	140.0
15	Seoul	Seocho-gu	45.0	450.0	55.0	46.0	203.0
16	Seoul	Seongdong-gu	22.0	300.0	49.0	11.0	55.0
17	Seoul	Seongbuk-gu	8.0	202.0	33.0	13.0	158.0
18	Seoul	Songpa-gu	92.0	590.0	55.0	30.0	147.0
19	Seoul	Yangcheon-gu	4.0	281.0	71.0	15.0	158.0
20	Seoul	Yeongdeungpo-gu	139.0	620.0	62.0	34.0	87.0
21	Seoul	Yongsan-gu	24.0	207.0	50.0	19.0	40.0
22	Seoul	Eunpyeong-gu	81.0	153.0	53.0	8.0	48.0
23	Seoul	Jongno-gu	192.0	438.0	22.0	38.0	38.0
24	Seoul	Jung-gu	217.0	585.0	12.0	86.0	32.0

## 04. 상관관계 분석

<유흥업소, 음식점, 카페, 체육시설>

각 데이터 간의 상관관계

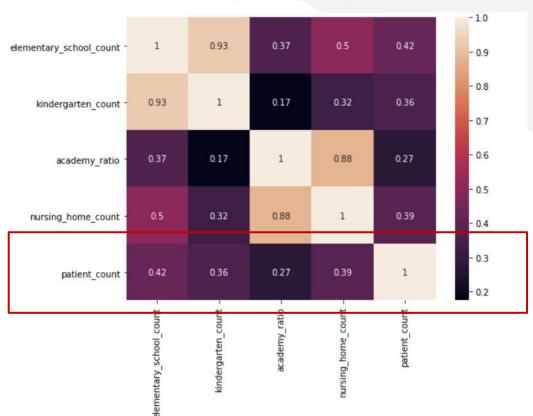
(정부 규제 장소)



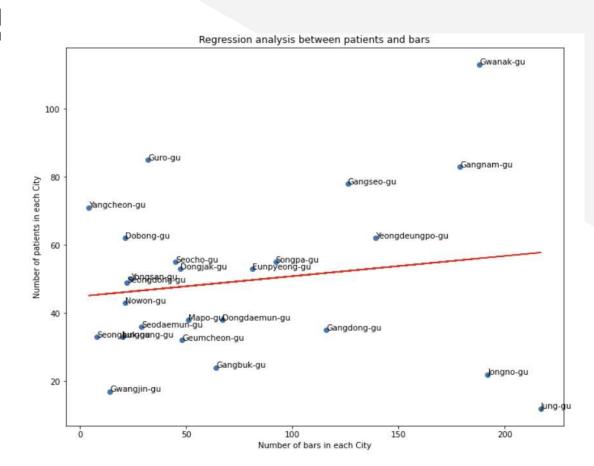
각 데이터 간의 상관관계

(정부 규제 없는 장소)

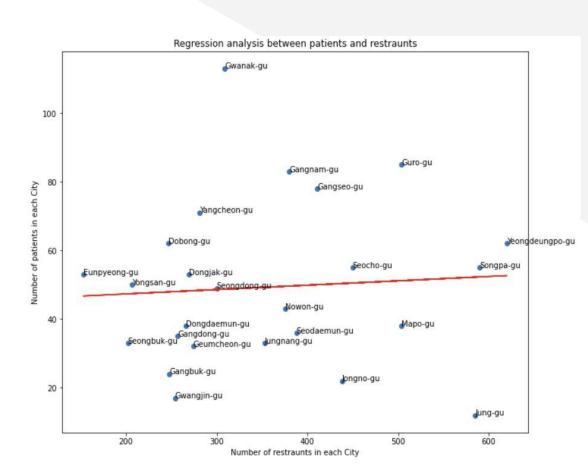
<초등학교, 유치원, 학원, 요양원>



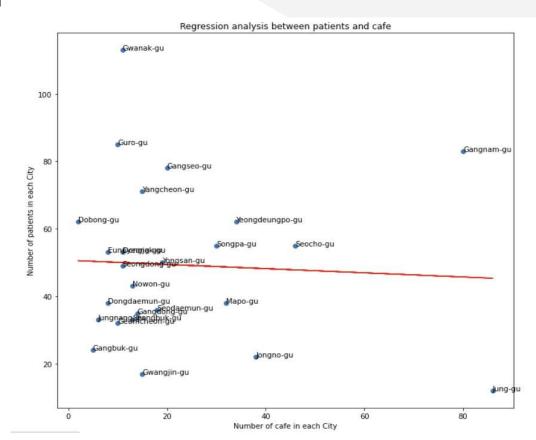
유흥업소 수와 확진자 수의 관계



식당 수와 확진자 수의 관계

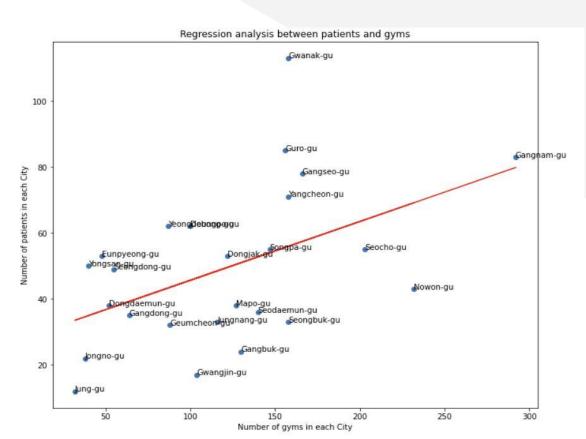


카페 수와 확진자 수의 관계



### 04 상관관계 분석

체육 시설 수와 확진자 수의 관계



### 가설 2: 사회적 거리두기 단계가 효과가 미미할 것이다.

### "획일적 거리두기 효과 없다" 전문가들, 방역체계 전환 주문

용 김나현 기자 │ ② 입력 2021.08.25 04:46 │ ■ 댓글 0







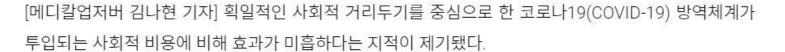








민주당 공공의료TF, 신종감염병 대응 주제로 심포지엄 개최 김윤 교수 "거리두기 줄이고 확진자 검사, 접촉자 격리 주력해야" 조심스러운 복지부 "백신접종과 사회적 거리두기 균형 필요"



전문가들은 확진자 중심이 아닌 중증환자, 사망자 중심의 정책을 준비하고 '9월말 10월초' 시행을 목표로 지속가능한 방역체계를 지금부터 검토해야 한다고 조언했다.











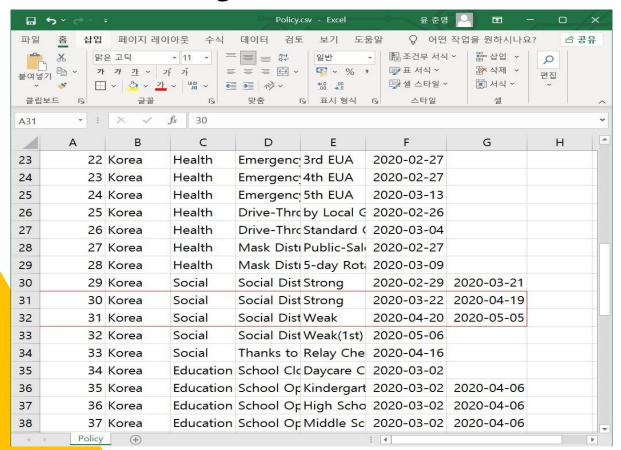








## Policy.csv



사회적 거리두기: Weak

- 2020-04-20 ~ 2020-05-05

사회적 거리두기: Strong

- 2020-03-22 ~ 2020-04-19

### Social distance: weak

	date	test	negative	confirmed	released	deceased	date_2	
91	2020-04-20	563035	540380	10674	8114	236	20200420	
92	2020-04-21	571014	547610	10683	8213	237	20200421	
93	2020-04-22	577959	555144	10694	8277	238	20200422	
94	2020-04-23	583971	563130	10702	8411	240	20200423	
95	2020-04-24	589520	569212	10708	8501	240	20200424	
96	2020-04-25	595161	575184	10718	8635	240	20200425	
97	2020-04-26	598285	578558	10728	8717	242	20200426	
98	2020-04-27	601660	582027	10738	8764	243	20200427	
99	2020-04-28	608514	588559	10752	8854	244	20200428	
100	2020-04-29	614197	595129	10761	8922	246	20200429	
101	2020-04-30	619881	600482	10765	9059	247	20200430	
102	2020-05-01	623069	603610	10774	9072	248	20200501	
103	2020-05-02	627562	608286	10780	9123	250	20200502	
104	2020-05-03	630973	611592	10793	9183	250	20200503	
105	2020-05-04	633921	614944	10801	9217	252	20200504	
106	2020-05-05	640237	620575	10804	9283	254	20200505	

# Social distance: Strong

	date	test	negative	confirmed	released	deceased	date_2								
62	2020-03-22	331780	308343	8897	2909	104	20200322	86	2020-04-15	534552	508935	10591	7616	225	20200415
63	2020-03-23	338036	315447	8961	3166	111	20200323	87	2020-04-16	538775	513894	10613	7757	229	20200416
64	2020-03-24	348582	324105	9037	3507	120	20200324	88	2020-04-17	546463	521642	10635	7829	230	20200417
65	2020-03-25	357896	334481	9137	3730	126	20200325	89	2020-04-18	554834	530631	10653	7937	232	20200418
66	2020-03-26	364942	341332	9241	4144	131	20200326	90	2020-04-19	559109	536205	10661	8042	234	20200419
67	2020-03-27	376961	352410	9332	4528	139	20200327								
68	2020-03-28	387925	361883	9478	4811	144	20200328								
69	2020-03-29	394141	369530	9583	5033	152	20200329								
70	2020-03-30	395194	372002	9661	5228	158	20200330								
71	2020-03-31	410564	383886	9786	5408	162	20200331								
72	2020-04-01	421547	395075	9887	5567	165	20200401								
73	2020-04-02	431743	403882	9976	5828	169	20200402								
74	2020-04-03	443273	414303	10062	6021	174	20200403								
75	2020-04-04	455032	424732	10156	6325	177	20200404								
76	2020-04-05	461233	431425	10237	6463	183	20200405								
77	2020-04-06	466804	437225	10284	6598	186	20200406								
78	2020-04-07	477304	446323	10331	6694	192	20200407								
79	2020-04-08	486003	457761	10384	6776	200	20200408								
80	2020-04-09	494711	468779	10423	6973	204	20200409								
81	2020-04-10	503051	477303	10450	7117	208	20200410								
82	2020-04-11	510479	485929	10480	7243	211	20200411								
83	2020-04-12	514621	490321	10512	7368	214	20200412								
84	2020-04-13	518743	494815	10537	7447	217	20200413								
85	2020-04-14	527438	502223	10564	7534	222	20200414								

### 데이터 분석 결과

#### < Social distance: Weak >

```
< Social distance : Strong >
```

```
In [51]:
         weak_test_total = weak['test'].sum()
                                                                      In [54]: strong_test_total = strong['test'].sum()
          weak confirmed total = weak['confirmed'].sum()
                                                                               strong_confirmed_total = strong['confirmed'].sum()
          weak_test_total, weak_confirmed_total
                                                                               strong_test_total, strong_confirmed_total
Out [51]: (9678959, 171875)
                                                                     Out [54]: (13151736, 290549)
         weak_ratio = weak_confirmed_total / weak_test_total
                                                                      In [55]:
                                                                              strong_ratio = strong_confirmed_total / strong_test_total
          weak ratio
                                                                               strong ratio
Out [52]: 0.017757591493052095
                                                                     Out [55]: 0.022092064500078164
```

- 데이터 분석 결과 사회적 거리두기 단계가 약했을 때보다 강했을 때의 기간이 더 길어서 단순히 인원수로 따지는건 무의미해 보였고.
- '확진자 수'의 비율로 따진 결과, 사회적 거리두기 단계가 약했을 때보다 강했을 때 확진자 수의 비율이 높음을 확인

## 05. 결론

#### 05. 결론

데이터 분석 결과 정리

1. 정부가 규제를 가한 장소인 유흥업소, 음식점, 카페, 체육 시설과 확진자 수의 상관 관계를 분석해 본 결과, **체육시설을 제외한 나머지 유형의 장소들에서는 코로나** 바이러스의 감염과 무의미하였음을 알 수 있었습니다.

2. 사회적 거리두기가 약했을 때보다 강했을 때 오히려 확진자 수 비율이 높았음을 알 수 있었습니다.

### 05. 결론

### 가설: (서울 기준) 정부에서 올바른 코로나 관련 정책을 내리지 않았을 것이다.

정부에서 규제를 가한 장소와 확진자의 수와의 연관성은 없었고,

오히려 사회적 거리두기를 강화했을 때 확진자 수가 많았다.



정부가 내린 사회적 거리두기에 따른 장소의 규제 및 단계가 코로나 바이러스 감염 방지에 효과적이지 않다.

# 06. 해커톤 후기



데이터 분석을 진행해야할 양에 비해서 3일이라는 시간이 짧게 느껴져서 데이터 분석을 충분히 진행하지 못한 점이 아쉬움이 남았었습니다.

#### 이호진



지금까지 Ims에서 학습 시엔, 노드에서 시키는 내용만을 진행하느라고, 제가 그 내용을 정확히 숙지 했는 지에 대한 판단을 하지 못 했었는데, 거기에 대한 메타인지를 높이는 좋은 기간이 되었다고 생각합니다.

#### 김도경



3일동안 협업을 하면서 분업을 통해 데이터처리를 해서 효율적으로 진행할 수 있었고, 현 시점에서 가장 이슈인 코로나 데이터를 통해 인사이트를 도출해냈다는 점에서 좋았습니다. 다만 원하는 데이터를 유연하게 가공하지 못한 점은 아쉬웠습니다.

#### 윤준영



처음에는 어떻게 시작해야 할 지도 모르겠고 감도 안 잡혔었지만, 팀원 분들과 이야기를 나누면서 재밌는 주제들을 선정할 수 있었고, 힘든 일들을 알아서 맡아서 해주시다보니 정말 재밌게 진행 할 수 있었습니다. 또한 수업 시간에 배웠던 내용들을 이렇게 실습해 볼 수 있어서 정말 뜻깊은 시간이였습니다.

# 감사합니다

## Q&A