

코로나19 API를 통한 데이터 수집 및 태블로 시각화

오구 김민형 박경빈 조민호 최원희



목 차

- 프로젝트 개요
- 흐름도
- 데이터 구성
- 분석 결과 및 시각화
- 결론
- 시사점



- 프로젝트 개요

03

ChosunBiz PICK ①
"신천지 사태, 대구·경북 지역 1명당 3.5명
감염"
입력 2020.06.29. 오후 3:39 수도권 > 수도권일반
입력 2020.06.29. 오후 3:39 이태원 클럽발 서울 확진자 2명 늘어 99명...구로콜센터 넘어
섰다
KBS
현대백화점 무역센터점 47명 집단감염."방문객 특정 안 돼"
최유경 입력 2021.07.06.20:59 수정 2021.07.06.21:31 댓글 1개

2020년부터 본격적으로 시작된
코로나19



코로나19 백신과 치료제



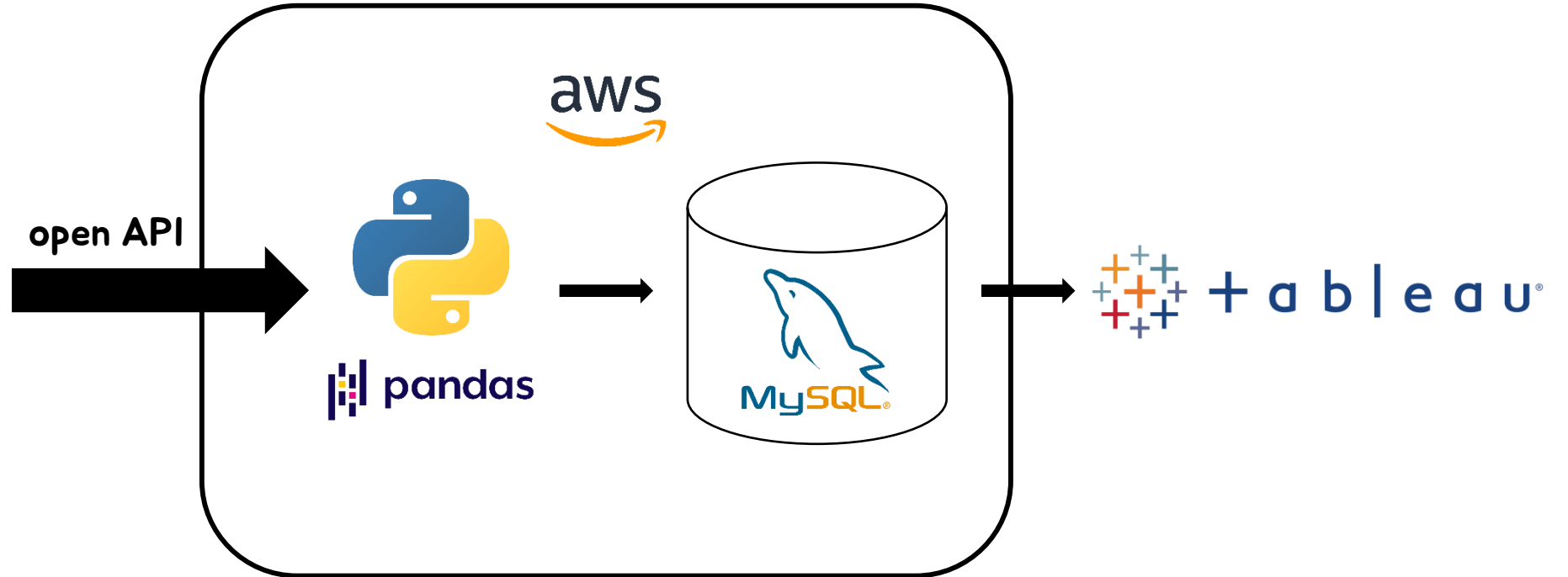
온라인 활동의 확산



코로나19 데이터 시각화를 통한 우리나라 코로나19 관련 현황 및 인사이트 도출

- 흐름도

04



- 흐름도

```
root@ip-172-31-11-17: /home/ubuntu
root@ip-172-31-11-17:/home/ubuntu# mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 63
Server version: 5.7.35-0ubuntu0.18.04.1 (Ubuntu)

Copyright (c) 2000, 2021, Oracle and/or its affiliates.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> use mysql;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> select user, host, authentication_string from user;
+-----+-----+-----+
| user          | host      | authentication_string |
+-----+-----+-----+
| root          | localhost | *THISISNOTAVALIDPASSWORDTHATCANBEUSEDHERE |
| mysql.session | localhost | *THISISNOTAVALIDPASSWORDTHATCANBEUSEDHERE |
| mysql.sys     | localhost | *THISISNOTAVALIDPASSWORDTHATCANBEUSEDHERE |
| debian-sys-maint | localhost | *C1cAlF97984EE2E4DAB0B5CF7295C06DD50C2823 |
| ogu           | %        | *A4B6157319038724E3560894F7F932C8886EBFCF |
+-----+-----+-----+
5 rows in set (0.00 sec)

mysql>
```

MySQL-server 내 ogu계정 사용

MySQL workbench 내부 DB table 구성

MySQL Workbench interface showing the 'Decide' table structure and query results.

Table: Decide

Columns:

- 날짜: datetime
- 당일 확진자 수: bigint(20)
- 검사 진행률 수: bigint(20)
- 누적 확진자 수: bigint(20)
- 누적 사망자 수: bigint(20)
- 누적 격리 해제 수: bigint(20)
- 지표 종합 점수: double
- 누적 확진률: bigint(20)
- 누적 검사 수: bigint(20)
- 누적 검사 완료 수: bigint(20)

Result Grid:

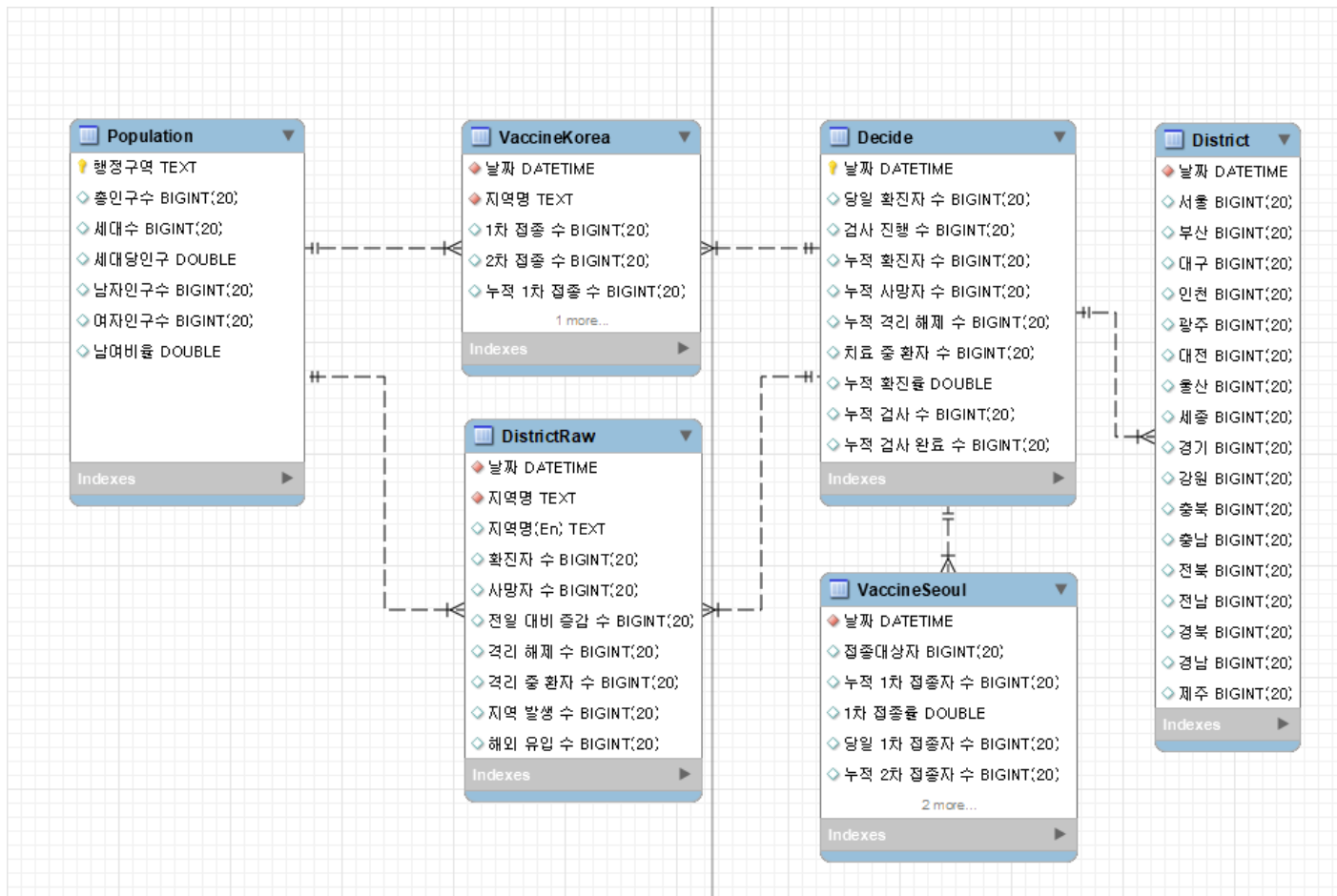
날짜	당일 확진자 수	검사 진행률 수	누적 확진자 수	누적 사망자 수	누적 격리 해제 수	지표 종합 점수
2021-08-18 00:00:00	2152	652740	230808	2191	201235	27382
2021-08-17 00:00:00	1805	634727	228656	2178	199582	26896
2021-08-16 00:00:00	1372	620680	226851	2173	197224	27454
2021-08-15 00:00:00	1555	595225	225479	2167	196198	27114
2021-08-14 00:00:00	1816	585131	223924	2156	195103	26665
2021-08-13 00:00:00	1928	562705	222108	2148	193778	26182
2021-08-12 00:00:00	1990	539535	220180	2144	192248	25788
2021-08-11 00:00:00	1987	510096	218190	2138	190536	25516
2021-08-10 00:00:00	2222	488067	216203	2135	189506	24562
2021-08-09 00:00:00	1537	465922	213981	2134	187524	24323
2021-08-08 00:00:00	1491	467508	212444	2125	186243	24076
2021-08-07 00:00:00	1728	466279	210953	2121	185003	23829
2021-08-06 00:00:00	1823	444937	209225	2116	183788	23321
2021-08-04 00:00:00	1704	427050	207402	2113	182054	23235
2021-08-03 00:00:00	1775	415489	205698	2109	180721	22868
2021-08-02 00:00:00	1200	398300	202198	2104	177910	22184
2021-08-01 00:00:00	1218	387170	200998	2099	176600	22299
2021-07-31 00:00:00	1440	378600	199780	2098	175673	22009

Action Output:

#	Time	Action	Message	Duration / Fetch
3	10:44:28	SELECT * FROM corona19: #Tableau_18_sid_00000057_1_Group	15 row(s) returned	0.015 sec / 0.000 sec
4	10:44:36	SELECT * FROM corona19: #Tableau_16_sid_00000018_1_Group	Error Code: 1146. Table 'corona19.#Tableau_16_sid_00000018_1_Group' ...	0.000 sec
5	10:44:42	SELECT * FROM corona19: #Tableau_17_sid_0000001C_2_Group	Error Code: 1146. Table 'corona19.#Tableau_17_sid_0000001C_2_Group' ...	0.000 sec
6	10:44:52	SELECT * FROM corona19: #Tableau_20_sid_0000001E_1_Group	15 row(s) returned	0.000 sec / 0.000 sec
7	10:45:55	SELECT * FROM corona19: Decide	535 row(s) returned	0.016 sec / 0.000 sec

- 데이터 구성

ERD (Entity Relationship Diagram)



Decide - 검진 및 확진 관련 데이터

Population - 지역별 인구 현황

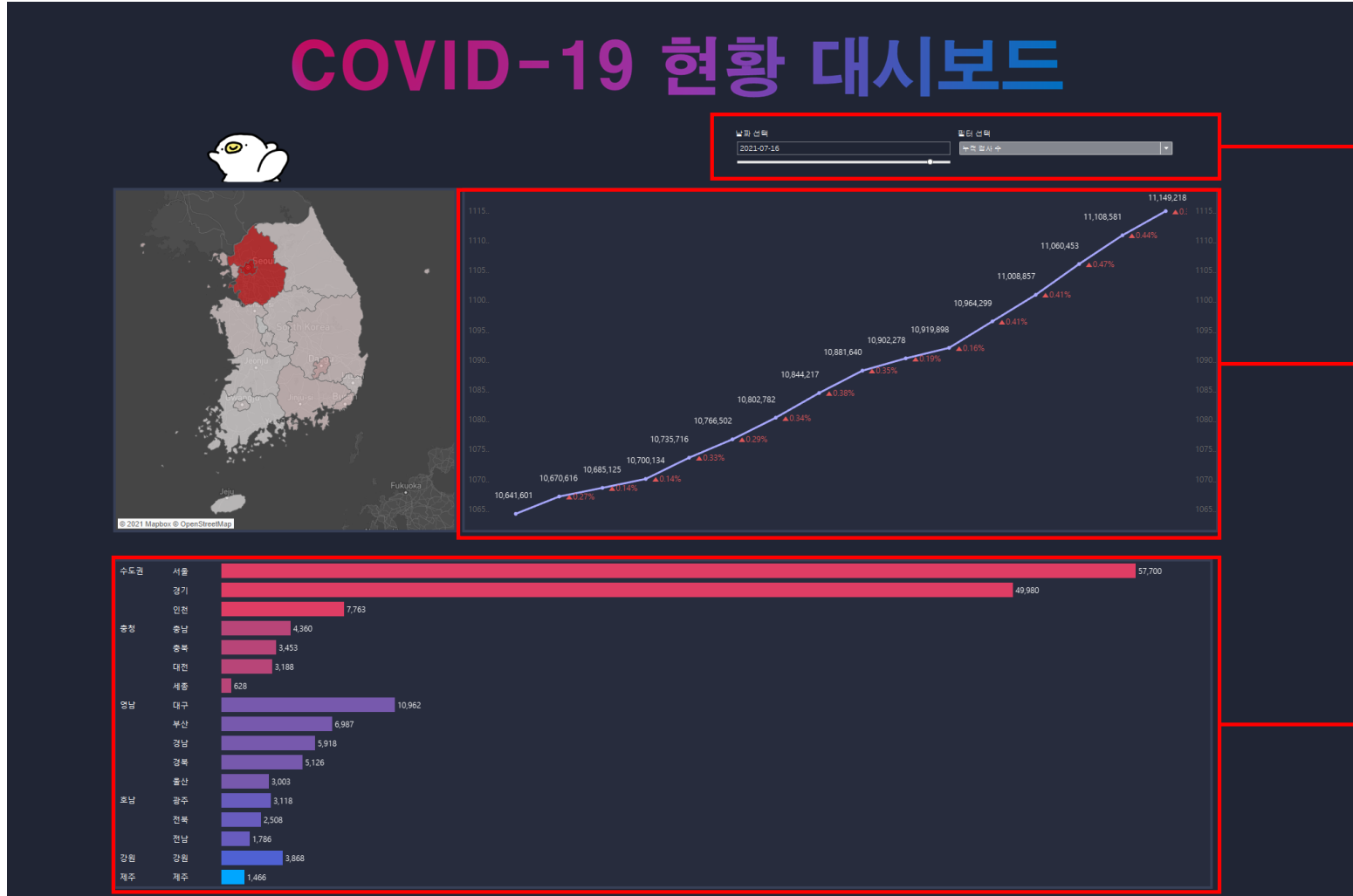
DistrictRaw - 각 지역 일자별 확진자

District - 지역별 신규 확진자

VaccineSeoul - 서울 백신 접종 현황

VaccineKorea - 전국 백신 접종 현황

- 분석 결과 및 시각화



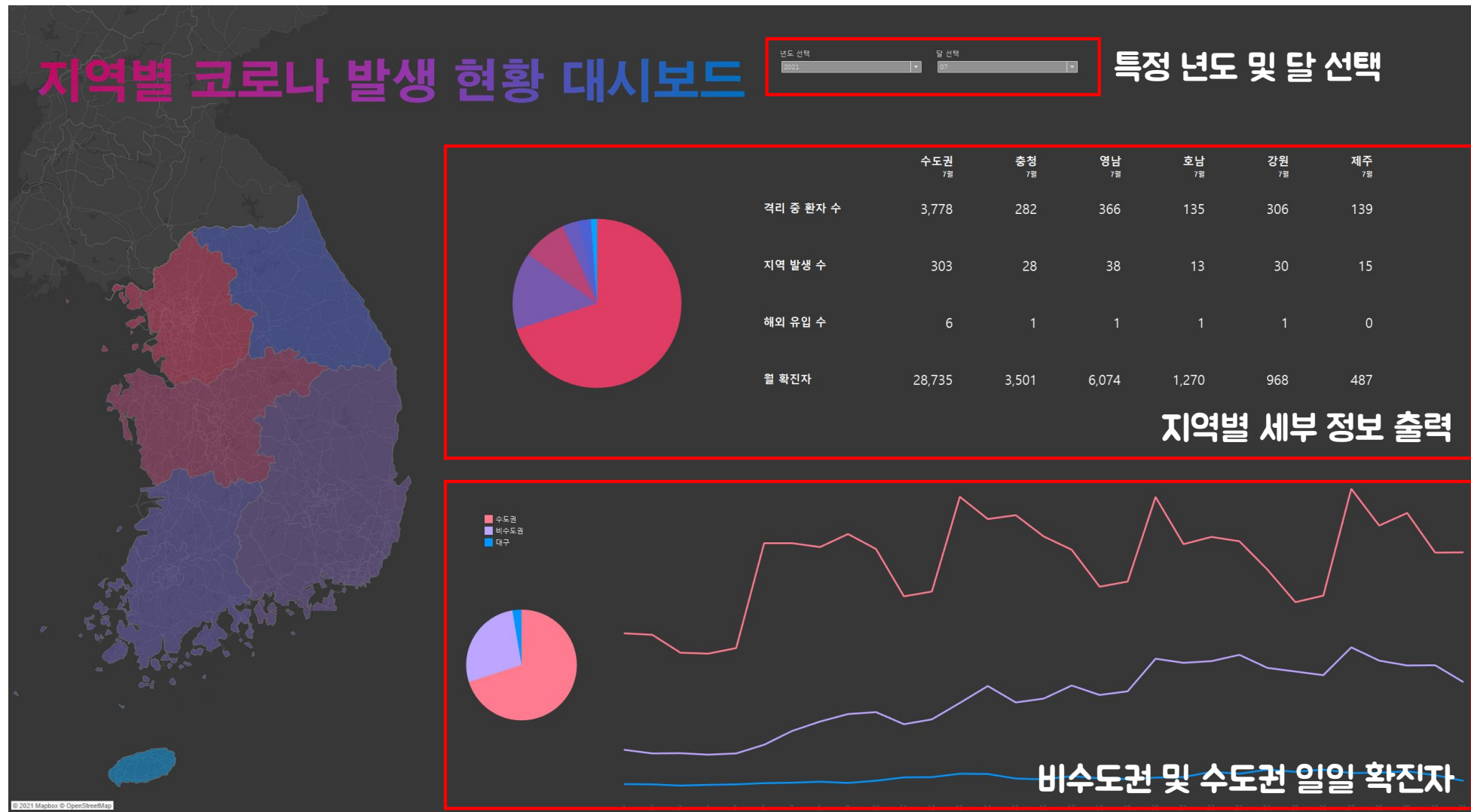
07

날짜와 원하는 출력 데이터 선택

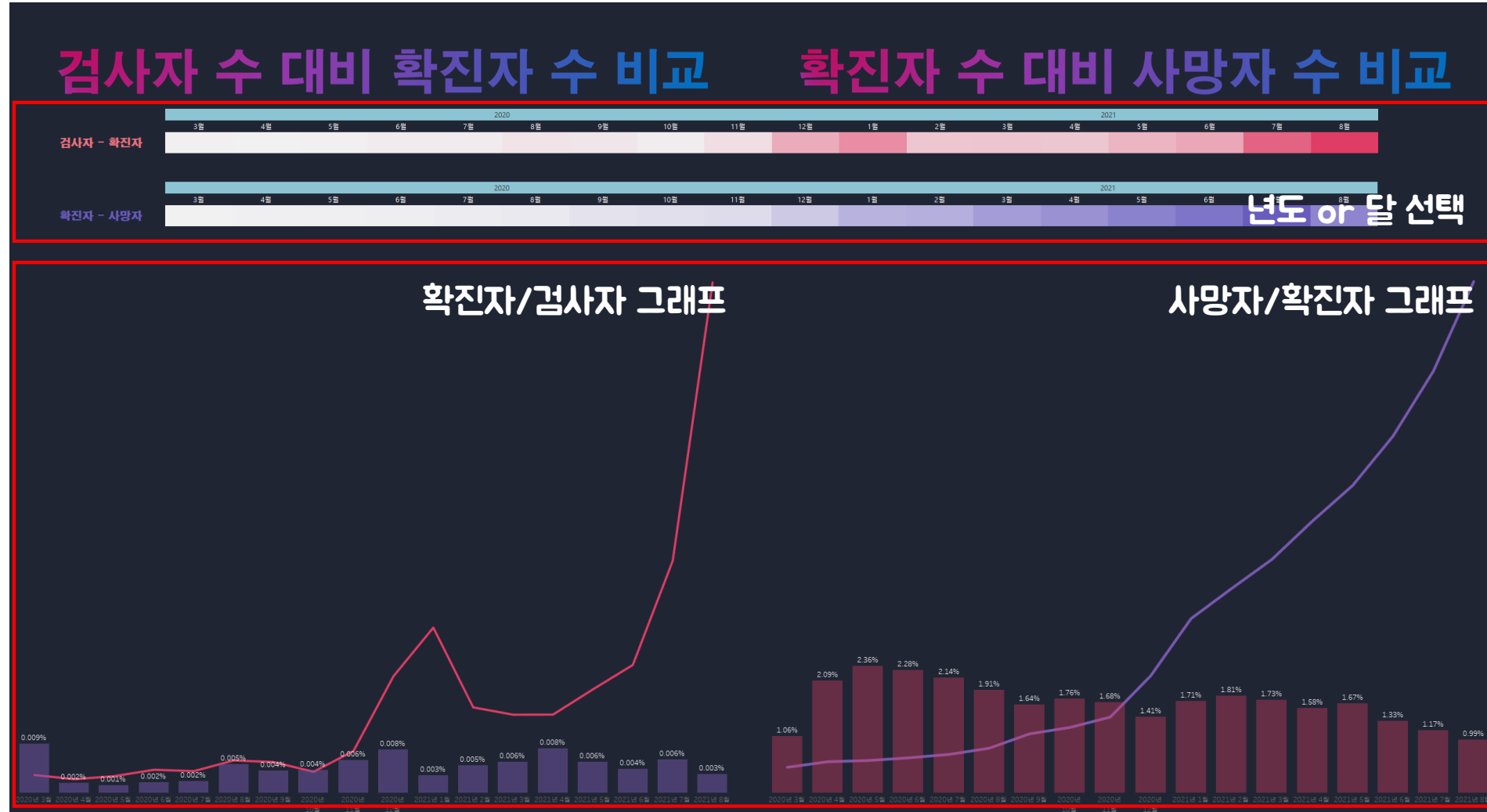
선택한 날짜에 대한
데이터 출력

선택 날짜 기준 전국 시도별
누적 확진자 통계

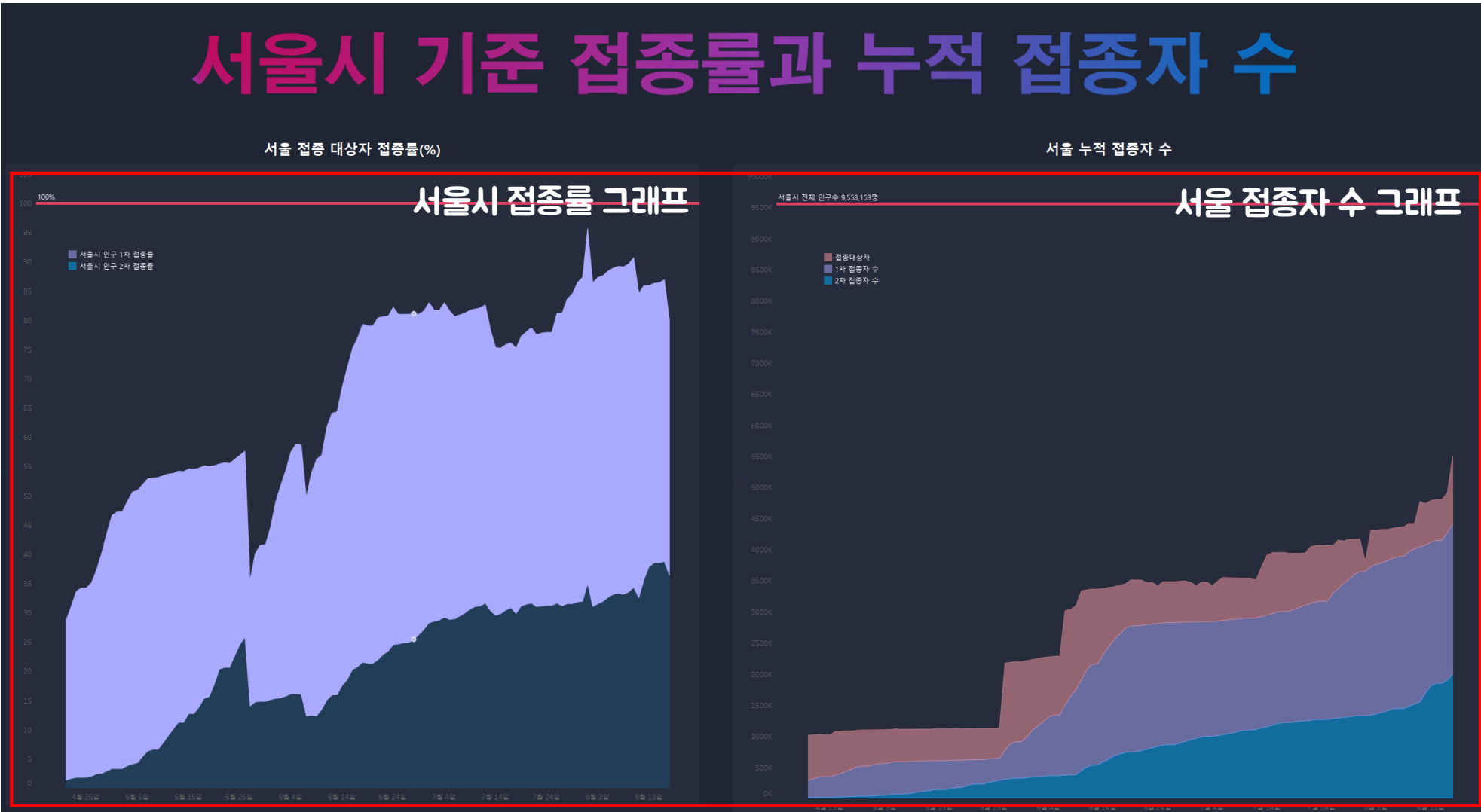
- 분석 결과 및 시각화



- 분석 결과 및 시각화

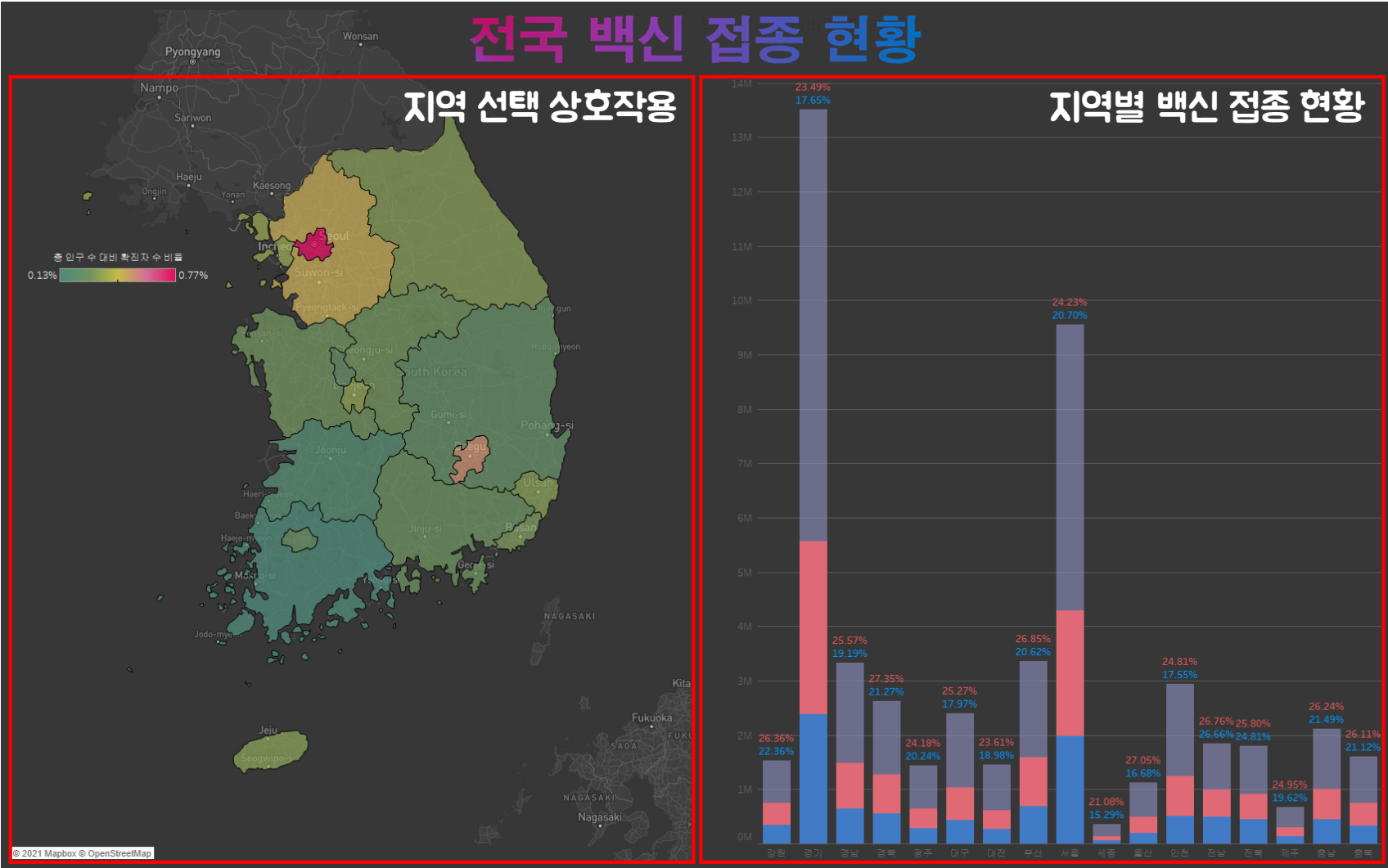


- 분석 결과 및 시각화

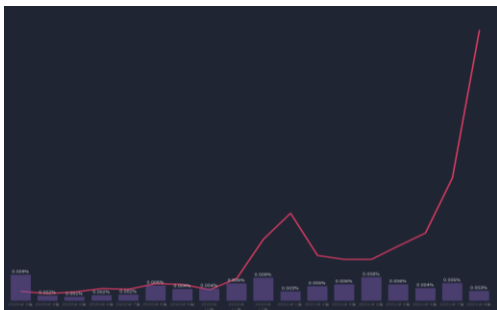


010

- 분석 결과 및 시각화

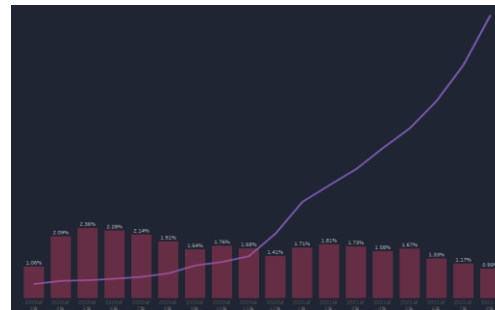


- 결론



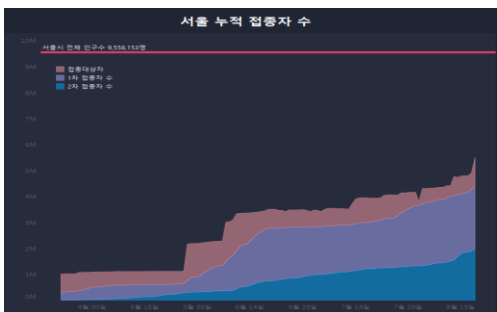
검사자 수 증가에 따른 확진자율 변화

검사자 수의 폭증에 비해
확진자율은 큰변화를 보이지 않음
증상 발현 시 PCR검사 권고



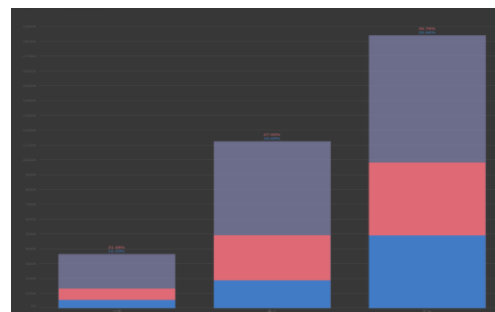
확진자 수 대비 사망률 변화

확진자 수가 급격하게 늘어나는 것에 비해
사망률은 전체적으로 감소하는 추세를 보임



서울특별시 백신 접종률 현황

데이터 상 접종률은 90%에 육박하지만
실제 전체 인구 대비 접종률은 50% 미만



지역별 백신 접종 현황 비교

전남의 경우 총 접종률 약 53%로 높은 편이며
울산은 1차에 비해 2차가 약 11%p가 낮고
세종시는 총 접종률 약 36%로 전국에서 가장 낮음

– 시사점

013

– Spark Streaming

Spark Streaming을 통해 open api로 제공되는 데이터 뿐 아니라 안전 대책 문자 등
코로나 실시간 데이터를 받아올 경우 실시간 변화 구현 가능

– 데이터 범위의 한계

영국, 인도 등 전세계 각국의 코로나 데이터 수집 시 다각도로 분석 가능

– 웹 페이지를 통한 외부 사용자 접근

외부 접근 가능한 대시보드 구현을 위한 웹페이지 환경 구성

-느낌점

우선 좋은 분들과 프로젝트를 할 수 있게 되어 너무 행복한 시간이었습니다=) 저는 태블로 시각화를 맡았는데 이 프로젝트 덕분에 태블로라는 프로그램을 공부할 수 있는 시간이 되어 뜻깊었습니다. 미숙한 실력이지만 제 태블로 작품을 보시고 칭찬해주신 우리 조원 여러분 정말 감사합니다~ 여러분덕분에 이번 프로젝트 동안 컴퓨터 전반에 대한 여러가지들을 공부할 수 있었습니다. 무엇보다도 태블로를 자유자재로 쓸 수 있게 되어 너무나도 기쁩니다! 모두 수고하셨습니다.



김민형



박경빈

조민호

최원희

본 교육을 지원하게 된 계기인 데이터 엔지니어링 수업을 통해 Spark와 SQL을 배운 후 프로젝트를 통해 직접 사용을 해보았습니다. AWS를 사용해 내부에 직접 MySQL 서버를 구축하고 open API를 통해 데이터를 가져와 저장하는 코드 작성을 하였는데 AWS내에서 데이터 수집, 처리, 저장의 모든 과정을 수행한다는 부분이 프로젝트 내에서 가장 뜻깊은 부분이었다고 생각이 듭니다.

spark를 통해 HDFS를 사용한 데이터 저장을 완료하지 못한 부분이 많이 아쉬웠지만 아쉬운만큼 시각화에서 좋은 성과를 낸것같아 결과적으로 만족스러운 프로젝트 였습니다. 프로젝트 기간 동안 모든 팀에 대해 꼼꼼하게 신경써주신 김성환 강사님께 감사드립니다.



이번에 배운 Spark, Hadoop, SQL 등을 직접 프로젝트에 접목해볼 수 있는 너무 좋은 기회였다. AWS 서버를 활용해 데이터베이스를 서버 내에 구축하고 서버 내에서 직접 데이터 저장 및 처리를 하는 것이 색달랐다.백엔드 뿐만 아니라 요즘 핫한 'Tableau'를 활용한 프론트엔드 구현까지, 전체적인 과정을 다 경험해볼 수 있었다는 것이 좋았다.2주라는 기간동안 저와 함께해준 팀원들에게 감사하고, 어떤 질문이든 마다하지 않고 친절하게 받아주시고도와주신 김성환 강사님 감사합니다!

우선 프로젝트를 같이 진행하며 고생한 저희 5조 조원분들에게 감사하다는 말 드리고 싶습니다.이번 프로젝트를 진행하면서 평소 관심이 많았던 Tableau를 사용할 수 있어 즐겁게 진행했습니다. AWS서버의 mysql server와 Tableau를 연결한 상태에서 시각화를 진행하며 많은 것을 배울 수 있었습니다. 또한 DE과정에서 이론으로 배운 Spark, Hadoop과 AWS서버를 직접 활용하여, 전체적인 프로그램 사용을 경험 할 수 있는 기회였습니다. 2주 동안 진행된 프로젝트 동안 도움을 주신 김성환 강사님 감사합니다.



THANK YOU

오구