**#0. start (주제 소개)**

안녕하세요, 팀케일리의 **첫 번째 발표 주제인 코로나 19 공공데이터 api 활용 가이드**를 소개하겠습니다. 저희 작업에는 14명의 케일리 팀원들이 참여했으며 저는 발표 전반부를 맡은 중앙대학교 문헌정보학과 19학번 팀케일리 설지은입니다.

**#1-1. 문제의식 소개(초록)**

**기존에 제공되고 있는 정보보다 더 의미 있는 정보를 만들 수 있지 않을까?**

**기존의 배포된 시각화를 더 개선할 수 없을까?**

이러한 문제의식 아래에서 저희 작업은 시작되었습니다.

그 이유는 현재 코로나 정보 제공에 있어서 아쉬운 점이 있었기 때문입니다.

**#1-2. 기존 코로나 정보 제공의 문제점**

여러분들께서는 네이버나 보건복지부 같은 코로나 상황판이 있는데 굳이 더 정보가 필요한가? 라는 생각을 하실 수도 있는데요. 그럼에도 불구하고 저희 팀이 발표 주제로 이것을 선택한 데는 이유가 있습니다. 우선 보건복지부 코로나 상황판을 보시면서 이에 대해 설명해드리겠습니다.

보시는 사이트는 보건복지부 코로나 발생현황 제공 사이트인데요.

보시다시피 표로 제공되는 정보들이 많았고

무엇보다 차트로 제공되는 정보들은 모두 누적값들뿐이었습니다. (시도별 파트 보여주기) 이렇게 결과적인 수치는 보여주지만 어떤 기간에 가장 많이 확진자가 증가했는지는 한눈에 알 수 없었습니다.

누적값을 기반으로 한 시각화들은 상황을 전반적으로 한눈에 볼 수 있다는 장점이 있지만 보다 세부적으로 상황을 파악할 수 없고 최신 동향을 알 수 없다는한계가 있습니다. 몇몇 데이터들은 표 형식으로만 제공되고 있어 한눈에 이해하기 어려웠습니다. 무엇보다 기간별 동향을 살펴볼 수 없다는 문제가 있는데요. 저희는 이러한 문제에 주목했습니다.

오늘 저희는 이러한 문제를 개선한 데이터 분석과 시각화를 보여드리겠습니다.

또한 저희 팀 발표는 공공데이터 api 가이드라인과 활용 예시를 제공할 것입니다. 뒤에 이어질 세미나에 선행하여 코로나19 데이터 분석의 기본이 되는 내용을 통해 이번 세션의 문을 열고자 합니다.

**#2. 목차 설명**

발표는 크게 세 부분으로 구성됩니다.

첫 번째로 OpenAPI 살펴보기에서는 본 발표에서 주로 사용된 보건복지부 데이터를 얻고 전처리하는 과정을 알아보겠습니다.

두 번째로 코로나 1년 돌아보기에서는 앞의 보건복지부 API를 통해 얻은 데이터를 기반으로 할 수 있는 다양한 분석과 시각화들을 보여드리겠습니다.

세 번째로 포스트 코로나로 나아가기에서는 보건복지부 api 이외의 데이터로 코로나 상황을 살펴보고 앞으로 나아갈 방향에 대해 고찰하도록 하겠습니다.

네 번째로 결론과 제언에서는 저희 작업의 의미와 저희 팀내에서 작업을 하며 느낀 점을 이야기하며 발표 마무리하도록 하겠습니다.

**#3. API 불러오기 및 전처리**

먼저 첫 번째로 OpenAPI 살펴보기에 대해서 발표할 텐데요

보건복지부에서 제공하는 api를 통해 데이터를 얻고 전처리하는 과정을 알아보겠습니다.

**#3-1 데이터 초기 형태 설명**

우선 공공데이터 포털에서 가져와 사용한 데이터는 코로나19 감염현황, 시도발생 현황, 연령별 성별 감염 현황, 해외발생 현황 네 가지 데이터 파일입니다. 화면에 보시다시피 각 데이터에는 다음과 같은 정보들이 있습니다.

굵직하게 각 데이터를 설명드리자면 감염현황 데이터에는 검사 수, 확진자 수, 사망자 수 등 감염과 관련된 수치들이, 시도발생 현황에는 확진자, 사망자가 시도별로 구분된 수치들이, 연령별 성별 감염 현황에는 확진자, 사망자가 성별, 연령별로 구분된 수치들이 포함되어 있습니다.

특이점이 몇 가지 있다면 검사 수 데이터가 감염현황 파일에만 포함되어 있어 검사 수 데이터를 시도 별로 제공하지 않는다는 것입니다. 또한 지역발생, 해외유입 수 데이터는 시도 별 파일로만 제공한다는 점이 있습니다.

이 데이터들은 모두 보건복지부에서 제공하며, 현재는 하루에 한 번씩 집계되고 있습니다. 데이터는 공공데이터 포털에서 활용 신청 후 파이썬을 활용해 xlsx 형식으로 다운 받은 후 사용했습니다.

**3-2. 컬럼 이름변경,값밀림 수정**

openapi에서 가져온 원본 데이터는 약간의 전처리 과정이 필요했는데요. 네 가지 데이터 중 감염현황 데이터 전처리 과정만 간단히 설명하겠습니다. 전처리는 크게 3가지로 나뉩니다. 컬럼 이름 변경해주고 컬럼과 데이터의 불일치 수정하고 하루에 한 개의 데이터만 들어가도록 수정하는 것인데요.

이 표가 원본 데이터 인데, (ppt\_raw 데이터) 보시는 것 처럼 컬럼명이 이해하기 어렵게 표시되어 있습니다. 이 부분을 openapi 활용 가이드 문서를 참고해 익숙한 컬럼명으로 수정해주었습니다.

또한 시간이 지나며 데이터 수집 형식에 변화가 생겨 데이터의 컬럼 수가 늘어나서인지 3/2일 이전에는 값이 밀려 컬럼과 해당 데이터가 불일치하는 부분이 있었습니다. 따라서 이 부분을 일치하도록 옮겨주는 처리도 진행하였습니다.

마지막으로, 날짜는 ‘기준일’ 기준으로 정렬하였는데, 기준 시간에 따라 하루에 한 번 이상 집계된 날짜도 있었습니다. 그런 날짜들을 모두 찾아 한 번씩만 들어가게 해 주었습니다.

**3-3. 필요한 컬럼 추가**

이처럼 전처리 과정을 거쳤지만 API 데이터의 확진자 수, 사망자 수 등 대부분의 데이터는 모두 누적값, 합계치만을 집계하고 있어 일자별 데이터는 파악할 순 없었습니다. 따라서 데이터 활용과정에서 추가적으로 필요한 ‘일별 확진자 수’, ‘일별 사망자 수’, ’일별 격리 해제수’, ‘일별 검사 완료수’ ‘일별 신규 검사수’, ‘기준월’을 직접 구해서 추가했습니다. (일별 데이터는 (전날 데이터-당일 데이터) 로 계산합니다.) 다른 데이터들도 이와 비슷한 방법으로 진행했습니다.

**#3-4. 작업환경 소개**

**미리 로그인 해두기**

본격적으로 저희의 작업물에 소개하기 전에 저희 팀의 작업환경을 간단히 이야기하겠습니다. 팀 케일리는 구글 코랩으로 오늘 보여드리는 모든 발표 내용을 작업하였습니다. 코랩을 이용하여 브라우저에서 파이썬을 작성하고 실행할 수 있습니다.

**#4. 코로나19 : 1년 돌아보기**

두 번째 파트 코로나 1년 돌아보기에서는

앞에서 전처리한 데이터들을 갖고 코로나 현황을 크게 감염현황과 검사현황으로 나누어 시각화 했는데요. 먼저 감염 현황 시각화를 소개하도록 하겠습니다.

**#4-1. 감염현황**

**#4-1-1.감염현황 - 시도별 누적 확진자(지도, 꺾은선, 애니메이팅)**

앞의 포털에서 세부적으로 상황을 파악하는 데 한계가 있었는데요. 저희는 이러한 점들을 개선하기 위해서 여러가지 시도를 해보았습니다. 먼저 전국 확진자 누적값을 꺾은선 그래프로 나타내보았는데요. 보시다시피 초반엔 녹색인 대구가, 그리고 지금은 빨간색 그래프인 수도권을 중심으로 유행했습니다.

그리고 확진자를 시도별로 구분하여 지도로 누적 시각화해보았습니다. 색이 진한 곳일 수록 확진자 수가 많다는 것을 의미합니다. 하지만 이와 같은 시각화는 기존의 것들과 비슷하기 때문에 무엇보다 기간별 동향을 살펴볼 수 없다는 문제가 있는데요.

저희는 이러한 문제에 주목하여 월별로 기간을 나눠 동향을 살펴보려 했습니다. 이 그래프는 애니메이션 차트로 누적 확진자 값이 월별로 어떻게 변하는지 보여줍니다. 실제 그래프는 코랩 링크에 다 확인 가능하니 참고해주시길 바랍니다.

**4-1-3. 감염현황 - 월별 시도별 감염현황**

**기간별로 동향을 디테일하게 파악하기 위해서 저희는 감염현황에 대해 월별, 시도별 두가지 기준을 설정한 후 시각화 하였습니다.**

**이 그래프를 통해 월별 확진자수 현황에서는 지역과 시간에 따라 달리 나타나는 확진자 수를 확인할 수 있는데요. 각 지역의 확진자 수를 월별로 합계 내어 시각화한 애니메이션 그래프입니다.**

한번 보실까요?

**혹시 3월과 12월에는 확진자가 가장 많이 나온 지역을 보셨나요?**

못보신 분들 위해서 3**월에 대구 지역이 타지역에 비해 압도적으로 확진자 수가 많고, 12월에 수도권 지역이 타지역에 비해 확진자 수가 많았습니다.** 이는 기간별 이슈에 따른 지역별 집단 감염에 비롯된 것으로 대구 내 신천지 집단 감염, 수도권 내 일상생활 간 집단 감염으로 인한 것으로 보입니다. 이것에 대해서는 세번째 챕터에소 소개할 타임라인에서 다시 이야기하겠습니다.

이 그래프는 집단 감염과 지역 별 확진자 수 변화의 연관성을 직관적으로 보여줍니다.

**무엇보다 지역에 따라 확진자 수가 천차만별이고, 기간에 따른 추이가 다르다는 것을 쉽게 보여줄 수 있습니다.**

**목적: 앞에서는 단순히 확진자 수만을 살펴보았는데요. 애초에 그 지역의 인구가 많아서 확진자도 많이 나오는 걸 수도 있지 않을까라는 생각에 그 지역의 인구수를 고려하여 이러한 시각화를 만들어봤는데요** 통계청의 인구수 데이터를 이용해 각 지역의 월별 확진자 합계를 각 지역의 인구수로 나누어 주었습니다. 원의 크기는 인구수에 비례하는데요. 인구수 대비 월별 확진자 수가 기간별로 어떤 양상을 보이는지 확인하시겠습니다. (말에 강약을 준다..?)

.

**다른 기존의 시각화들은 지역의 인구수를 고려하지 않은 것들인데 지금 보시는 그래프는 인구수는 물론 그 지역의 확진자수를 직관적으로 볼 수 있습니다.**

이외에도 사망률, 사망자수, 격리해제수와 관련하여 비슷한 방식으로 시각화를 하였는데요 코랩을 참고해주시면 감사하겠습니다.

**4-1-4. 감염현황 - 월별 연령별 감염현황**

**다음으로는 나이와 시간에 따라 달리 나타나는 감염현황을 확인할 수 있는 월별 연령별 확진률그래프입니다.**

**목적: 기존 다른 사이트의 시각화에는 연령별 감염현황에 대해 표나, 누적치로만 보여주었기 때문에 시간에 따른 변화를 알 수 없었습니다. 저희는 그러한 한계에 주목하였는데요.**

조금 설명 드리자면 보시면 초반엔 20대의 확진률이 높았고 뒤로 갈 수록 5-60대의 확진률이 높아집니다. 아마 초반에 20대 비중이 높은 것은 집단 감염 사례인 신천지와 이태원 클럽 집단 감염의 영향으로 보입니다. 이후 5 - 60 대의 확진률이 높아지는 것은 기저 질환의 영향과 요양병원 집단 감염 사례의 영향으로 볼 수 있을 것 같습니다.

이외에도 사망률, 치명률과 관련하여 비슷한 방식으로 시각화를 하였는데요 코랩을 참고해주시면 감사하겠습니다.

**4-1-5. 감염현황 - 월별 성별 감염현황**

**목적 :월별 성별 감염현황 시각화도 앞서 보여드렸던 기존의 연령별 차트와 같은 한계점을 보충하고자 했습니다.**

성별 확진률 차트에 대해 조금 설명 드리자면 보시면 여성의 확진률이 코로나 확산 초기에 남성보다 상대적으로 높은데요. 이에 대한 이유를 3월 발생한 신천지 집단 감염에서 여성의 비중이 높았던 것 그리고 구로 콜센터 집단 감염 등 여성 근로자의 비중이 높은 직군에서 집단감염이 발생했던 것을 바탕으로 추측해볼 수 있습니다.

이외에도 성별 사망률에 대해서도 비슷한 방식으로 시각화를 했으니 코랩을 참고해주시면 될 것 같습니다.

**#4-2. 검사현황**

**#4-2-1. 일별/ 월별 검사현황**

지금부터는 코로나 19 검사현황에 대해 이야기 하겠습니다.

보여지는 이 그래프는 일별 검사현황과 확진자 수를 알려주는 그래프인데요 주황색 선은 신규 검사자 수를, 회색 바 그래프는 확진자 수를 뜻합니다.

**일별 신규 검사자수는 의심 신고 검사자 수이므로 확진자 수와 긴밀히 연결되어 있을거라 생각해 일별 확진자수와 함께 시각화했습니다.**

**분석을 진행할 당시 검사를 받은 후 확진 판정이 나오기 때문에 검사 건수의 증가 양상이 먼저 보일거라고 추측하고 있었는데, 막상 확인해 보니 확진자 수가 증가한 후에 검사 건수가 증가함을 확인했습니다.**

**#4-2-2. 요일별 검사현황**

다음으로는 요일별 검사현황을 살펴보도록 하겠습니다.

**목적: '주말은 검사를 적게해서 확진자가 적다는 말이 사실인가?'에 대한 의문이 들어 분석해보았습니다.**

**결론적으로, '일요일과 월요일은 전날 검사를 적게해서 확진자 수가 적게 나타난다' 라는 주장은 어느 정도 타당하다고 볼 수 있습니다**

**전날 혹은 당일의 검사완료 수에 영향을 받는 것을 고려해야 하기 때문입니다**

월요일을 살펴보면, 신규검사수가 일주일 중 가장 많음에도 검사 완료수는 그렇지 않습니다. 그러나 화요일의 검사 완료수가 가장 많은 것으로 보아 월요일의 신규 검사수가 화요일에 완료되었다고 볼 수 있습니다. **따라서 월요일의 확진자수가 적게 나타나는 이유는 월요일 당일의 신규검사수보다, 전날인 일요일의 신규검사수가 적기 때문에 그 영향이 월요일에까지 미쳤을 것으로 예상됩니다.**

또한 일요일을 보시면 당일 검사 완료수도, 신규 검사수도, 확진자 수도 가장 적습니다. **당일의 검사수와 전날의 토요일의 검사수가 모두 적기 때문에 일요일의 확진자 수도 적은 것으로 예상됩니다.**

**그러나 토요일은 당일의 신규검사수와 검사 완료수는 적지만, 확진자수는 비교적 많다는 점에서 약간의 예외 현상이 발견되지만, 평일에 비하면 적은 확진자 수를 보입니다.**

이처럼 요일별 검사수를 확인해보았습니다만 아쉬웠던 점은 데이터가 상세하지 않았기 때문입니다. 각각 케이스의 검사시작일과 검사완료일에 대한 데이터가 제공된다면 더 정확하고 상세한 분석이 가능할 것입니다.

**#4-2-3. 익명 검사 현황 (api 사용 x)**

**목적:**  익명 검사는 작년 말부터 수도권을 중심으로 실시되고 있는데요. 익명 검사가 얼마나 이용되고 있는지, 전체 검사 중 얼마나 차지하는지 궁금했지만 다른 곳에서 시각화를 제공하고 있지 않기 때문에 직접 시각화해 보았습니다**. 다만 익명 검사 데이터의 경우 api에서 제공하고 있지 않기 때문에 보도자료 pdf에서 하나하나 직접 데이터를 입력하여 시각화하였습니다.**

다음의 그래프는 총 검사수와 익명검사수를 보여주는데요. **익명검사 건수가 일별 총 검사 건수의 절반 가량을 차지**할 정도로 많은 시민분들이 익명검사를 이용한 것을 알 수 있습니다.

또한 아래 그래프는 전체 신규확진자 중 익명 검사 신규확진자 수가 얼마나 되는지를 확인하기 위한 그래프입니다. 익명검사자 중에서도 90~150명 정도의 신규 확진자가 발생했는데요.. 익명 검사를 통해 발견한 신규확진자는 전체 신규확진자의 약 10% 가량인 것을 확인할 수 있습니다.

여기까지 코로나19 1년 돌아보기 파트였고요 뒤이어서 김가윤님이 발표를 이어나가 주시겠습니다.

**<Intro>**

#23 안녕하세요, 후반부 API이외 데이터를 통해 포스트코로나로 나아가기의 발표를 맡은 중앙대학교 사회학과 17학번 팀케일리 김가윤입니다.

코로나 발병 이후 우리는 새로운 변수, 문제를 마주할 때마다 신속한 대응을 위해 정책 시행에 급급한 모습이 있었습니다. 하지만 이러한 정책이 실제로 효과가 있는지, 또 발병이후 어떤 일이 있었는지에 대해서 살펴볼 기회는 없었던 듯합니다. 또한 앞서 API로 많은 것들을 살펴보기도 했지만, 코로나 감염현황을 통한 더 많은 이슈파악을 위해선 현재 API만으로는 부족했습니다. 정례브리핑과 보건복지부 자료를 보면 선별진료소, 중환자 병상 데이터 등 다양한 정보가 있었지만, 이를 API로 제공하고 있지 않는 것이 아쉬웠습니다. 따라서 이번 세션은 저희가 이를 수집한 자료를 통해 여러분들과 함께 과거와 현재의 우리 그리고 정책들을 돌아보고 더 나은 미래를 위해 사회를 조망하는 시각과 새로운 관점을 제공하고자 합니다.

살펴볼 항목들은 선별진료소와 중환자 병상 데이터, 그리고 코로나 관련 타임라인입니다. 주제는 크게 두 개로 나누어보았습니다. 세부적으로 소개해드리자면,

우선 선별진료소와 중환자 병상 데이터를 통해 이들이 모두 고르게 분포되어 있는지, 혹은 패턴이나 어떠한 양상을 보이는지를 알아봅니다. 또한 코로나가 발병한 이후 정부가 실시한 방역정책을 모두 정리해 시각화한 그래프를 살펴보면서 지금까지 정부가 실시하고 있는 정책에 대한 실효성을 알아봅니다.

그 다음으로는 코로나 발병 이후 현재까지의 시도별 타임라인을 살펴봅니다. 이 타임라인은 시도별 감염현황을 시기별로 살펴볼 수 있음과 동시에 집단감염이 일어난 사례와 이슈 등의 전반적인 코로나와 관련된 기사를 아카이빙했습니다.

**<선별진료소>**

#24 그럼, 선별진료소는 전국에 고르게 설치되어있을까? 라는 질문 아래 선별진료소 데이터를 시각화한 그래프를 살펴보겠습니다.

#25 저희는 우선 시군구별 선별진료소의 수를 지도에 시각화해 나타내 보았는데요, 각 지역에 설치된 선별진료소의 위치가 핀으로 고정되어 있고, 마커 클러스터는 마커들 사이 거리에 따라 클러스터링 된 개수를 의미합니다. 이를 보면 수도권에 수가 집중되어 있는 모습을 볼 수 있습니다.

#26 왼쪽에 보시는 그래프는 색의 채도를 기준으로 선별진료소의 수를 나누어 시각화해본 것인데요, 서울특별시 중구, 진주시의 경우 8개로 가장 많은 선별진료소가 설치되어 있었습니다. 또한 앞선 질문에 대한 답을 조금 더 정밀하게 알아보기 위해 오른쪽처럼 시군구별 인구수 대비 선별진료소 비율을 살펴보았습니다. 선별진료소가 가장 많이 설치되어 있는 중구와 진주시가 연하게 표출된 반면, 구례군, 진안군 등 군 단위에서 인구수 대비 선별진료소의 비율이 높게 나타났습니다.

#27 더불어 응급의학계에 따르면 신종 코로나 감염 환자가 발생 한 뒤 보건복지부에서 지역응급의료센터급 의료기관에 선별진료소를 설치할 것을 지시했다는 기사를 찾아볼 수 있었습니다. 추가적으로 응급의료센터가 없는 지역의 경우에도 설치해달라고 권유를 받은 곳도 있었고, 의료기관에 따라 자체 설치한 곳도 있었습니다. 질병관리청 중앙방역대책본부가 각 지자체에 배포한 신종 코로나 대응지침에 따르면 보건소를 포함한 지자체별 선별진료소의 목적, 역할, 요건이 담겨있지만 선별진료소를 어떤 형태로 만들어 운영해야하고 어느 정도 인력과 장비를 갖춰야 한다는 구체적인 기준은 지침에 담겨 있지 않다고 합니다.

결론적으로 전국 선별진료소는 인구 이동을 고려하여 적절하게 설치된 것 같아 보이지만 이는 필요한 때마다 추가적으로 세워진 것을 추측해볼 수 있습니다. 또한 선별진료소에 대한 특정 기준과 지침이 없기 때문에 이와 더불어 법적 정의와 역할을 마련하고 지역 별로 협의체를 구성해 논의해야할 필요가 있다고 생각합니다.

**<중환자병상>**

#28 또 지난 11월/ 확진자 수가 급증해 병상이 부족해진 부산에서 / 처음 / 타시군구의 병원으로 환자를 이송해서 치료하는 등 의료 시설과 병상 부족 문제가 대두되기도 했는데요. 이와 관련해 일일 확진자 수와 해당 지역의 가용 병상수를 한번 비교해보면서 위중증환자의 발생을 커버할 수 있을만큼 가용 중환자 병상이 확보되어 있었는지를 확인해보았습니다. 데이터는 정례브리핑자료를 바탕으로 한 '중증도별 병상현황' 자료를 직접 구축하여 사용하였습니다. 데이터가 11월 7일부터 존재하기도 하고, 현황이 발표되지 않은 날이 많아 누락된 데이터가 꽤 있었습니다. 브리핑자료에 발표된 형식에 따라 이번에는 시도별이 아닌 지역권별로 분류했습니다.

# 29 발표에서는 수도권만 간단하게 살펴보겠습니다. 수도권에서는 12/5-12/22에 가용 중환자 병상 수가 20 이하였던 것을 확인할 수 있습니다. 같은 시기 수도권의 신규확진자 수는 급증하는 추세를 보이는 것 또한 확인할 수 있어, 신규 확진자는 많고 가용 중환자 병상은 비교적 적은 상황이었습니다. 이 그래프는 y축이 다르기 때문에 잘못해석할 수 있는 여지가 있어, 다음 그래프처럼 직관적인 시각화를 만들었습니다.

# 30 이 그래프는 '필요한 가용 중환자병상 수'와 '보유한 중환자 병상 수'의 차이를 나타낸 그래프인데요, 필요한 가용 중환자병상의 수는 중앙사고수습본부의 브리핑 내용의 위중증환자는 확진자 수의 2.2% 인것을 참고하여 조금 더 여유 있게 일일 지역 확진자 수의 2.5%로 계산 했습니다. 필요한 가용 병상 수와 실제 가용 병상 수의 차가 1보다 작은 경우에 부족하다고 판단되어 주황색으로 표시했습니다. 12월에 지역 전반적으로 부족한 모습을 보였고, 여기서는 보실 수 없지만 강원과 제주의 경우 1월에도 부족한 날이 있었습니다. 따라서 위와 같이 언제든 병상이 부족한 상황이 생길 수 있기 때문에 이를 충분하게 확보하는 것이 좋을 것 같다고 판단이 되는데요, 인구 수와 비용 등의 여러 변수를 고려하면 추가 설치보다는 지역 간의 연결 활성화가 더욱 합당하다고 보입니다. 또한 인구에 따른 것도 있지만, 지역별 의료 현황이 다 다른 모습을 볼 수 있었습니다.(가지각색이다) 따라서 주요 시설들이 수도권에 집중되어 있는 현상을 해소하기 위해 이와 관련된 기관들이 지역별로 대책을 마련하고, 이에 관련한 논의가 더욱 필요해 보입니다. 코로나를 통해 다시 한번 위 안건에 대해 생각해볼 수 있는 시간이었던 것 같습니다.

**<타임라인>**

#31 다음은 정부 정책과 시도 별 이슈를 알아볼 수 있는 코로나 타임라인을 각각 살펴보겠습니다.

**정책 별 타임라인**

#32 우선 정책 별 타임라인을 살펴보겠습니다. 이 그래프에서는 일별 확진자 수와 함께 정부가 실시했던 거리두기, 마스크 5부제, 시설 제한 정책 등의 방역 정책을 한 눈에 볼 수 있습니다. 또 앞에서/ API데이터를 통해 유추한 것들이 실제로 코로나 감염에 얼마나 영향을 미쳤을까요? 보도 되었던 집단 감염이 어느정도의 파장을 불러일으켰는지 궁금하진 않으셨나요? 이와 같은 시각화를 통해 위 질문에 대한 답을 확인해 볼 수 있습니다. 또 주목해야할 부분은 11월 1일 거리두기 단계 별 변화가 생기게 되면서 거리두기의 격상이 전과 같은 비율로 적용되지 않은 것이 그래프 상으로 나타나는 모습인데요, 거리두기 정책이 조금 늦게 시행된 것과 동시에 코로나가 오래 지속되다 보니 거리두기에 대한 시민들의 무뎌짐과 안전불감증과 같은 심리적 요인이 적용된 것으로도 추측해볼 수 있습니다.

시도 별 타임라인

# 33 다음은 시도 별 타임라인을 살펴봅니다. 각 시도별 일별 확진자 수와 함께 코로나 집단감염과 이슈 관련 기사를 아카이빙해 그래프에 표시했습니다. 대표적으로 서울의 타임라인을 가져와봤는데요, 보이는 빨간 마커를 누르면 각 일자별로 어떤 감염 사례 혹은 이슈가 있었는지 확인해볼 수 있습니다. 이는 두가지 측면에서 의의를 가집니다. 우선, 시도 별 각자 차별화된 집단 감염 사례와 지역적 특성을 알아볼 수 있기 때문에 수도권, 비수도권으로 나눈 이분법적인 정책과는 다른 세부적인 시도별 코로나와 관련한 정책 세우기에도 도움이 될 것 같다고 생각했습니다.

또한 시민들도 언론이나 포털 사이트에서 당일의 기사와 확진자 수 등의 정보를 단면적으로 접하게 됩니다. 이 때문에 막연한 공포심에 휩싸이거나 충분한 정보를 얻지 못하는 경우가 있습니다. 시민들 또한 이 타임라인을 통해 당시 기사와 함께 쉽게 상황을 돌아볼 수 있고, 코로나에 대한 경각심과 동시에 발병 이후의 우리 사회를 총체적이고 객관적으로 돌아볼 수 있는 기회를 가질 수 있습니다. 또한 포스트 코로나를 살아가야 하는 우리의 마음가짐과 자세를 확립할 수 있습니다.

**-------------------------------------------------------------------------------------------------------------------------**

**결론과 제언**

**결론**

# 34 마지막으로 이번 저희 팀의 프로젝트의 결론과 제언을 이야기해 보고자 합니다.

#35 우선 프로젝트를 통해 저희는 코로나 사회에 대한 심층적인 분석을 했습니다. 기존에 제공 되었던 보건복지부 API를 통해 전체 기간이 아닌, 특정 기간 데이터를 선택가능한 형태로 각 감염현황별로 심도있게 살펴보았습니다. 그 이외 저희가 수집하거나 구축한 데이터들로 정책 실효성을 되짚어보고 현재의 문제점을 발견하기도 했습니다. 또 특정 기간에 있었던 대유행이나 사회이슈를 한번에 시각화해서 볼 수 있도록 하여 새로운 인사이트를 제공했습니다. 기존보다 개선된 시각화를 통해 더 많은 정보를 쉽고, 직관적으로 이해할 수 있도록 하기도 했습니다. 마지막으로 데이터 기반으로 알아보는 객관적인 정책 실효성을 통해 앞으로 데이터의 중요성과 더불어 이를 기반한 의사결정과 정책 시행의 필요성을 깨닫는 시간이었습니다.

#36 이를 위해서 정부에서도 개선해야할 사항들을 몇가지 제언하고자 합니다. 첫째, 의미 있는 분석을 위한 추가적인 데이터를 개방해야 합니다. 많은 데이터가 제공될수록 더 나은 사회를 위한 다양한 정책 아이디어와 세부적인 개선 방향이 생겨난다고 생각합니다. 두 번째, 정부에서도 개선된 시각화를 국민에게 제공해야 합니다. 모든 국민은 충분한 정보를 얻을 권리가 있습니다. 하지만 정제되지 않은 데이터나 텍스트, 숫자형과 같은 보기 어렵고 이해하기 힘든 데이터는 이를 저해합니다. 정확성과 무결성을 보장하는 데이터를 제공하면서 국민들에게 보다 이해하기 쉽고 새로운 인사이트를 제공하는 시각화를 추가적으로 제공할 것을 제안합니다. 마지막으로 데이터 기반 정책 실효성과 행정을 위한 세부적인 기준이 필요하다는 것을 알려드리고 싶습니다. 현재 정부는 시행중인 디지털 뉴딜 정책을 통해 국민생활과 밀접한 분야의 데이터 구축과 개방을 약속했고 공공데이터를 전면 개방하며 분야별 데이터의 수집과 활용 확대하면서 빅데이터 플랫폼을 구축할 것을 예고한 바 있습니다. 효과적인 성과 도출을 위해 데이터를 수집하고 구축하는데에 더욱 세부적인 기준과 정제 방법을 명시하는 것이 필요합니다. 또한 앞서 발표에서와 같이 정책의 기준이 모호한 모습을 볼 수 있었습니다. 정책을 시행하면서 데이터를 기반으로 한 실효성을 검증하고 이를 개선해야할 구체적인 기준을 마련해야 합니다.

**<마무리>**

# 36 지금까지 api 전처리부터 그 외 자료까지 다루어보면서 코로나 발병 이후 1년이 넘는 시간을 다양한 시각화와 함께 돌아 보았습니다. 다음 이어지는 발표에서는 집단감염 발표로 이전의 타임라인에서 다루었던 각종 집단감염 이슈를 실제 데이터로 구현한 다양한 시각화와 분석을 살펴보도록 하겠습니다. 이상으로 첫 번째 주제의 발표를 모두 마치겠습니다. 감사합니당 ^^\*