**□ 모듈프로젝트 산출물**

1. 클라우드기반 데이터 전문가 양성 과정 – 기반기술 인프라

|  |
| --- |
| 모듈프로젝트 |

**클라우드 데이터 전문가**

이름 : 이호연

1. **모듈프로젝트 내용**

|  |
| --- |
|  |

**1)코로나 종식을 위해 정부는 무엇을 하는가?**

* 1. 정부 정책 효율성
* Import 선언 및 일별 코로나 확진자, 사망자, 누적확진자 수

|  |
| --- |
|  |
|  |
|  |
| **라이브러리 설명**  시각화를 위해 사용한 라이브러리는  import matplotlib.pyplot as plt,  import seaborn as sns ,  import folium,  import plotly  import plotly.graph\_objs as go  import plotly.express as px  from plotly import graph\_objects as go 입니다.  from matplotlib import font\_manager, rc로 폰트를 설정했습니다.  **그래프 설명**  코로나 현황을 계기판 형태로 표현한 차트입니다. plotly 내의 go를 사용했습니다. 숫자로 시선을 확 끌 수 있는 차트를 만들기 위해 계기판 아이디어를 사용했습니다. 2021년 8월 1일 기준 신규 확진자수, 격리 해제수, 검사 수, 사망자 수를 표현했습니다. 이 그래프를 통해 현재 코로나의 대유행 상황에 대해 설명하였습니다.  **코드 설명**  go.Figure()에 add\_trace를 이용해 그래프를 추가해 총 네 개의 그래프를 그렸습니다. Indicator를 이용해 추가 옵션을 설정했습니다. Value는 본래 값, delta = 'reference' 코드는 참조로 이 값을 참고하여 본래 값과의 차이가 얼마인지 확인해 화살표 형태로 차이를 나타냅니다. |

* 일별 확진자 및 사망자 데이콘 그래프 작성

|  |
| --- |
|  |
|  |
|  |
|  |
| **그래프 설명**  데이콘 그래프를 그래프로 시각화했습니다. 첫 번째는 일별 확진자를 플로틀리를 사용해 바 그래프로 만들었고, 두 번째 그래프는 히트맵을 사용해 일별 사망자 수의 분포를 나타냈습니다.  확진자와 사망자 모두 특정기간 폭증하는 형태를 띄고 있고 사망자의 경우 확진자가 폭증하고 2주 정도 뒤 같은 양상으로 폭증하는 것을 볼 수 있습니다. 이로서 확진 후 사망까지 2주 정도가 걸린다는 사실을 알 수 있습니다. 일별 확진자의 경우 plotly.express의 bar 그래프를 사용하였고 일별 사망자의 경우 plotly.express의 density\_heatmap 그래프를 사용하였습니다.  **코드 설명**  데이콘 데이터의 경우 확진자, 사망자, 완치자가 일일이 아닌 누적으로 되어있었지만 확진자가 폭증하는 양상을 확인하기 위해서는 일일 확진자를 산출해야 했기 때문에 for 반복문을 사용했습니다. 먼저 sum을 이용한 groupby로 날짜별로 묶어주고 각 컬럼들을 리스트 형태로 바꿔주었습니다. 이후 for 문에서 ‘당일 누적 확진자 – 전날 누적 확진자’ 식을 이용해 당일 일일 확진자, 완치자, 사망자 수를 산출했습니다. 날짜의 경우 date 타입으로 바꿔서 그래프에서 날짜로 인식할 수 있게 전처리 해주었습니다. |

* 사회적 거리두기 시행단계별 서울시 유동인구 변화 추이

|  |
| --- |
|  |
|  |
| **그래프 설명**  데이콘의 SeoulFloating 데이터에서 fp\_num를 이용해 일별 유동 인구 변화량을 구했습니다. matplotlib.pyplot을 이용해 그래프를 시각화했으며 빨간 선은 유동 인구량의 평균이고 파란 선은 재난 경보를 시행한 시기를 마크해 둔 것 입니다. 이를 통해 1월 3일에 시행한 재난 레벨 1 경보는 유동인구 감소에 효과가 있었지만 제난 레벨 2와 3은 유동 인구량에 변화가 거의 없고 재난 레벨 4의 경우 오히려 유동 인구량이 증가한 것을 볼 수 있습니다. 재난 경보 정책의 경우 처음에만 효과가 있었고 이후 사람들이 재난 경보에 무뎌지면서 효과가 줄어든 것으로 볼 수 있습니다.  **코드 설명**  Mean을 이용한 groupby로 유동 인구의 평균을 구하고 2월 23일 데이터의 경우 극단적으로 높은 수치가 나와서 이상치로 판단하고 제거했습니다. 넘파이를 이용해 데이터를np형태로 바꿔준 다음 싸이킷런의 MinMaxScaler를 이용해 데이터를 정규화 해준 후 이 데이터를 이용해 그래프를 그렸습니다. Plt. Annotate를 이용해 정책이 시행된 날짜에 마크를 추가 했으며 이를 통해 해당 날짜에 어떤 정책이 시행됐는지 알 수 있으며 유동 인구 감소에 정책 별 영향을 확인할 수 있었습니다. |

* 북촌 CCTV로 확인한 유동인구 추이

|  |
| --- |
|  |
|  |
| **그래프 설명**  공공데이터 센터에서 서울시 북촌의 CCTV 인원수 데이터를 사용해 유동인구 추이를 나타낸 그래프입니다. Iplot을 사용한 선그래프로 나타냈으며 날짜별 카메라 통과 인원을 통해 유동인구의 추이를 대략적으로 확인할 수 있습니다. 개학 지연을 시행한 3월 2일 이후에도 지속적으로 유동인구가 늘어나는 것을 확인할 수 있습니다. 일별 데이터라 주말과 평일의 유동인구 차이가 눈에 띄는 특징이 있습니다.  **코드 설명**  데이콘 데이터의 경우 유동 인구량이 2월 27일까지 밖에 나와있지 않습니다. 하지만 사회적 거리두기 정책과 개학 지연 정책의 경우 정책 시행이 3월 이후 부터라서 3월 이후의 데이터가 필요했습니다. 그래서 공공 데이터와 통계청 데이터를 활용해 그래프를 시각화했습니다. 측정 시작 시간의 경우 초 단위까지 나와있었는데 이를 str.slice 를 이용해 일자 까지만으로 자르고 그룹 바이를 이용해 날짜 별로 묶어줬습니다. |

* 월별 부산 10대 유동인구 이동량 조사

|  |
| --- |
|  |
|  |
| **그래프 설명**  부산시의 10대 이동량 그래프입니다. 코로나 발병 이후인 2020년도 데이터와 코로나 발별 전인 2019년도 데이터를 비교할 수 있습니다. 뚜렷한 차이가 보이지 않는 점을 볼 수 있습니다. 심지어 2020년도의 10대 이동량이 더 많습니다. 이를 통해 개학 지연을 했음에도 10대들은 학원을 가고 친구들을 만나러 가는 등의 활동을 통해 여전히 활기차게 다니고 있는 모습을 볼 수 있습니다. Iplot의 bar를 사용해 시각화했습니다.  **코드 설명**  CCTV 데이터에서 평일과 주말이라는 특징 때문에 유동 인구가 차이가 나는 것을 보완하기 위해 부산시에서 연령별 부산시 유동인구 데이터를 다운받아 사용했습니다. 10대부터 70대까지 있는 데이터였고 개학 지연 정책의 효과를 보고 싶었기 때문에 10대만 뽑아서 사용했으며 2019년도와 2020년도 데이터를 나눠서 같은 그래프에 시각화함으로써 그 차이를 한눈에 볼 수 있게 표현했습니다. |

* SKT 모바일 인구이동으로 보는 사회적 거리두기 정책 효과

|  |
| --- |
|  |
|  |
| **그래프 설명**  2020년 2월부터 10월까지 SKT 모바일 인구이동 그래프입니다. 사회적 거리두기가 시행된 3월 2주차부터 4월 2주차 약 한달간 인구 이동이 안정적으로 유지되는 것을 볼 수 있습니다. 사회적 거리두기를 2단계에서 2.5단계로 상향했던 8월 2주차에도 모바일 인구 이동이 급진적으로 하락하는 모습을 볼 수 있습니다. 하지만 2.5단계에서 2단계로 단계를 낮췄을 경우인 4월 말과 9월 초반의 경우 모바일 유동 인구가 급격하게 증가하는데, 이를 통해 사회적 거리두기 상향은 효과가 있지만 하향 시에도 인구 이동이 증가하는 양상이 있음을 알 수 있으며 대유행이 완전히 끝날 때 까지 사회적 거리두기 단계를 낮추면 안된다는 것을 확인할 수 있습니다.  **코드 설명**  지역이나 연령, 성별에 국한되지 않는 최대한 넓은 범위의 국민 이동량을 볼 수 있는 데이터는 없을까 하고 찾다가 SKT의 모바일 인구이동 데이터를 통계청에서 발견했습니다. 해당 데이터를 이용해 2020년 2월부터 10월까지 인구 이동량 그래프를 그렸습니다. 한글 컬럼으로 이루어진 데이터라 인코딩을 cp949로 했주었습니다. |

* 특별 출입국 절차 시행 이후 해외 유입 확진자수

|  |
| --- |
|  |
|  |
| **그래프 설명**  특별 출입국 절차 시행 이후 해외 유입 확진자수를 날짜별로 시각화한 그래프입니다. 특별 출입국 절차를 시행한 이후임에도 확진자가 폭증하는 모습을 볼 수 있으며 국내 일일 확진자수 그래프와 동일한 양상을 띄고있습니다. 이를 통해 특별 출입국 절차의 효과가 미미했음을 알 수 있습니다.  **코드 설명**  데이콘의 Patientinfo 데이터를 사용해 해외 유입 확진자수를 확인했습니다. Infection\_case가 overseas inflow인 환자만 추출했으며 필요없는 컬럼들은 drop을 사용해 삭제했습니다. sns.countplot을 사용해 날짜별 해외 유입 확진자수 그래프를 시각화했습니다. |

* 코로나 코로나 정책별 네이버 검색량

|  |
| --- |
|  |
|  |
| **그래프 설명**  코로나 정책별 네이버 검색량을 네이버 데이터랩에서 추출해 일자별로 나타낸 관심도 그래프입니다. 국민들이 가장 관심을 많이 가졌던 정책은 개학 연기이고 그 다음은 사회적 거리두기입니다. 나머지 해외 유입, 진단 키트, 재난 문자 정책의 경우 두 가지 정책에 비해 관심도가 매우 떨어지는 것을 볼 수 있습니다 |

* 1. 병상 부족 해결을 위한 지역별 의료 인프라 분석
* 인구 천명당 의사 수 데이터 분석

|  |
| --- |
|  |
|  |
| 그래프 설명  인구 천 명당 의사수, 의사 천 명당 의료 기관수, 응급의료시설 5분내 접근 가능 인원수를 지역별로 묶어서 plotly.ff의 create\_table을 사용해 테이블로 시각화했습니다. 그래프 뿐만 아니라 테이블도 시각화해서 보여드리고 싶어서 표현했습니다.  코드 설명  통계청 SGIS에서 의료 격차 데이터를 다운받아서 각각 다른 데이터를 지역별로 merge한 다음 같은 수치 범위로 표준화시켜주었습니다. 필요없는 컬럼인 위도 경도등을 삭제하였습니다. |

* 지역별 의료 인프라 구축 현황

|  |
| --- |
|  |
|  |
| **그래프 설명**  지역별 의료 인프라 구축 현황을 한 그래프 내에서 볼 수 있게 나타냈습니다. 전체적인 인프라 점수가 가장 높은 지역은 광주이며 가장 낮은 지역은 세종입니다. 서울의 경우 인당 의사수는 많지만 인당 의료기관 수는 부족함을 볼 수 있습니다. 이 그래프를 통해 의료 인프라가 잘 구축되어있는 지역과 그렇지 않은 지역을 추출할 수 있으며 이후 정부의 의료 지원에서 어떤 지역이 1순위가 되어야하는지 볼 수 있습니다.  **코드 설명**  인구 천 명당 의사수, 의사 천 명당 의료 기관수, 응급의료시설 5분내 접근 가능 인원수를 표준화한 데이터를 사용해 한 그래프에 표현함으로써 전체 지표를 확인할 수 있게 하였습니다. |

* 종합 인프라 및 의료기관 주제별 상 ,하위 3개 도시 도출

|  |
| --- |
|  |
|  |
| **그래프 설명**  위의 그래프로 표현한 결과를 테이블로 옮겨서 텍스트로 읽을 수 있게 표현했습니다.  종합 인프라 상위 3개 도시는 광주, 대전, 부산, 하위 3개 도시는 세종 충청남도, 제주도 입니다.    **코드 설명**  go.Figure 을 사용해 테이블을 시각화 했습니다. cells=dict(values)를 이용해 테이블 안에 값을 직접 넣어주어 원하는 테이블을 나타낼 수 있었습니다. |

* 지역별 확진자 수 파악

|  |
| --- |
|  |
|  |
| **그래프 설명**  지역별 누적 확진자수를 시각화한 지도 그래프입니다. 서울 경기권인 수도권 지역이 1400명 이상으로 가장 심했고 그 다음이 경남쪽이 도합 1500명 가량, 세종의 경우 지역 의료 인프라가 가장 낮음에도 불구하고 세번째로 높은 누적 확진자수를 보여 가장 의료적으로 취약한 지역임을 알 수 있었습니다. 수치가 높은 경우 붉은색이 진해지고 낮은 경우 초록색으로 표시되어 색깔로도 심각도를 확인할 수 있습니다.  **코드 설명**  데이콘의 Patientinfo 데이터와 Region 데이터를 merge하여 지역별 확진자수를 산출했으며 이를 folium을 이용해 지도로 시각화했습니다. 중앙값은 [35.820308, 127.108791]로 설정했으며 zoom\_start=7로 설정하여 한눈에 전체 지역의 확진자를 확인할 수 있게 설정하였습니다. 지역별 위도와 경도를 사용해 folium내에서 위치를 설정했습니다. |

**2)개인은 코로나에 어떻게 반응하고 있는가?**

2-1) 연령대별 행동 요인

* 연령대별 확진자 행동경로 분석(막대그래프)

|  |
| --- |
|  |
|  |
| **그래프 설명**  확진자들의 이동 경로를 통해 행동 특성을 확인할 수 있습니다. 전체 연령대에서 가장 높은 비율을 차지한 장소는 병원, 식당, 상점 등이고 10대는 피시방이 높은 비율을 차지했고 20대와 30대의 경우 대중교통이 높게 나왔으며 하숙이나 리얼 에스테틱 샵의 비율도 다른 연령대보다 높게 나왔음을 알 수 있습니다.  **코드 설명**  PatientRoute 데이터와 Patientinfo 데이터를 patient\_id로 merge하여 사용했습니다. 연령대 별로 나누어서 각 연령대별 이동 경로를 알 수 있게 시각화했으며 subplot을 사용해 여러 개의 그래프를 한번에 나타냈습니다. |

* 연령대별 확진자 행동경로 분석(파이 차트)

|  |
| --- |
|  |
|  |
| **그래프 설명**  현재 코로나 대유행에서 가장 큰 확진자 비율을 차지하고 있는 20대와 30의 이동 경로 특징을 파이차트 형태로 나타냈습니다. 20대의 경우 대중교통의 비율이 매우 높았으며 하숙과 같은 특이 케이스도 다른 연령대에 비해 높았습니다. 30대의 경우 리얼 에스테틱 샵의 비율이 높았습니다. |

2-2) 코로나 관심도와 확진자 추이의 상관관계 분석

* 코로나 정책별 네이버 검색량 및 확진자 수

|  |
| --- |
|  |
|  |
| **그래프 설명**  코로나 관련 검색어와 확진자 수의 상관관계를 알기 위해 그린 날짜별 그래프입니다. 빨간 선이 코로나 바이러스에 대한 검색량이고 파란선이 일별 확진자 수 입니다. 명확한 상관관계를 도출할 순 없지만 코로나 바이러에 대한 검색량과 일별 확진자수의 그래프 모양이 상당히 비슷한 것을 볼 수 있습니다. 이후 코로나 검색량에 대한 추이를 통해 일별 확진자수를 어느정도 예측해 볼 수 있습니다.  **코드 설명**  코로나 관련 검색어는 SearchTrend 데이콘 데이터에서 추출했습니다. 해당 그래프와 일별 확진자 그래프를 한 그래프 상에서 동시에 표현했습니다. |

**3)결론**

3-1) 정책별 분석 및 의견

* 정책상 분석의견 및 효과가 있다고 도출된 정책

|  |
| --- |
|  |
|  |
| **그래프 설명**  결론적으로 재난 경보, 개학 지연, 특별 출입국 절차는 그 효과가 미비했으며 사회적 거리두기만이 직접적인 효과를 눈으로 확인할 수 있었습니다.  **코드 설명**  Go.Figure을 사용해 values를 직접 넣어 테이블을 생성했습니다. |

3-2) 최종결론

|  |
| --- |
|  |