[홈냐링 팀 보고서]

**1. 연령대별 국내여행 분석**

**(1) 연령대별 평균여행횟수**

15~19세/20대/30대/40대/50대/60대/70세 이상의 7개 그룹별로 국내여행에 대해 분석했다.

먼저, 연령대별 평균여행횟수를 파악했다.

[용어 정의]

1. 관광여행

▸ 관광·휴양 활동 방문

- 관광/자연 감상/거리 산책

- 박물관/미술관/동물원/식물원/수족관/테마파크 방문

- 해수욕/트레킹·등산/ 에코 투어

- 아마추어 스포츠 활동(골프·테니스 등)/레포츠 활동/스포츠 관람

- 온천 - 낚시·크루즈

- 캠핑

- 체험학습/수학여행

- 이벤트·축제 참가 - 신혼여행

- 종교 및 순례

- 건강·치유 여행 - 인센티브 관광

- 콘서트·영화·연극 감상을 겸한 나들이

- 맛집 투어를 겸한 나들이

- 쇼핑을 겸한 나들이

1. 기타여행

▸ 가족‧친구‧친지 방문

- 명절 귀성, 가족 행사, 결혼식, 장례식, 집들이 등 포함

▸ 출장·업무 방문

- 연수·세미나 참석/회의·대회·학회·컨벤션 참여/박람회 및 전시회 참여

- 연구·교육 활동

- 자원 봉사 활동

- 본사·지점·공장 등의 방문/거래처 방문/건설·토목 등의 감독/시찰·취재 - 강연·연주회 개최·출연/프로 스포츠 활동 - 기타 업무 목적의 모든 활동

▸ 기타여행 목적으로 행정구역상 현 거주지(일상생활권)를 벗어났다 하더라도 여행 방문지에서

관광‧휴양 활동을 병행했다면 관광여행으로 분류

* 전체 여행에 대해 연령대별 평균여행횟수 파악 결과

|  |  |
| --- | --- |
| 연령대 | 국내여행횟수 |
| 15~19세 | 5.655815 |
| 20대 | 8.367467 |
| 30대 | 9.189228 |
| 40대 | 8.638880 |
| 50대 | 8.199340 |
| 60대 | 6.881230 |
| 70세 이상 | 4.904285 |

30대의 여행횟수가 약 9.2번으로 가장 많고, 40대, 20대가 그 뒤를 잇는다.

* 관광 여행에 대해 연령대별 평균여행횟수 파악 결과

|  |  |
| --- | --- |
| 연령대 | 국내관광여행횟수 |
| 15~19세 | 4.542091 |
| 20대 | 6.942927 |
| 30대 | 7.124723 |
| 40대 | 6.601072 |
| 50대 | 6.300356 |
| 60대 | 5.332804 |
| 70세 이상 | 3.783295 |

관광 여행 횟수의 경우에도 30대가 가장 많으며, 약 7.1번이다. 그 뒤를 20대, 40대가 잇는다. 전체 여행은 40대가 앞섰으나 관광 목적의 여행의 경우 20대가 앞섬을 파악할 수 있었다.

* 기타 여행에 대해 연령대별 평균여행횟수 파악

|  |  |
| --- | --- |
| 연령대 | 국내기타여행횟수 |
| 15~19세 | 1.113724 |
| 20대 | 1.424540 |
| 30대 | 2.064505 |
| 40대 | 2.037808 |
| 50대 | 1.898984 |
| 60대 | 1.548426 |
| 70세 이상 | 1.120990 |

기타 목적의 여행의 경우 관광 목적의 여행보다는 적게 가는 것을 알 수 있다. 30대, 40대, 50대 순으로 기타 여행 횟수가 많다.

**(2) 연령대별 평균여행일수**

* 전체 여행에 대해 연령대별 평균여행일수 파악 결과

|  |  |
| --- | --- |
| 연령대 | 국내여행일수 |
| 15~19세 | 9.650628 |
| 20대 | 14.029326 |
| 30대 | 15.625341 |
| 40대 | 14.296744 |
| 50대 | 13.344029 |
| 60대 | 10.805357 |
| 70세 이상 | 8.034646 |

30대의 여행일수가 약 15.6일로 가장 많고, 40대, 20대가 그 뒤를 잇는다.

* 관광 여행에 대해 연령대별 평균여행일수 파악 결과

|  |  |
| --- | --- |
| 연령대 | 국내여행일수 |
| 15~19세 | 7.613811 |
| 20대 | 11.626904 |
| 30대 | 12.180706 |
| 40대 | 11.040954 |
| 50대 | 10.349554 |
| 60대 | 8.422585 |
| 70세 이상 | 6.090042 |

30대의 여행일수가 약 12.2일로 가장 많고, 20대, 40대가 그 뒤를 잇는다.

* 기타 여행에 대해 연령대별 평균여행일수 파악 결과

|  |  |
| --- | --- |
| 연령대 | 국내여행일수 |
| 15~19세 | 2.036817 |
| 20대 | 2.402422 |
| 30대 | 3.444635 |
| 40대 | 3.255690 |
| 50대 | 2.994475 |
| 60대 | 2.382771 |
| 70세 이상 | 1.944604 |

30대의 여행일수가 약 3.4일로 가장 많고, 40대, 50대가 그 뒤를 잇는다.

**2. 유형에 따른 만족도 비교**

만족도는 1~5 사이 숫자로 표시되며, 각 숫자는 아래와 같은 의미를 가진다.

1 매우 불만족, 2 불만족, 3 보통, 4 만족, 5 매우 만족

가족외동행여행만족도는 4.13, 가족여행만족도(타가구포함)는 4.14, 가족여행만족도(가구원)는 4.17, 나홀로여행만족도는 4.05이다. 나홀로여행보다는 동행자가 있는 경우 만족도가 더 높음을 알 수 있다.

이 차이가 유의한 것인지를 파악하기 위해 먼저 등분산 검정과 정규성 검정을 수행했으나, 등분산성과 정규성 가정을 만족하지 않아 비모수 검정인 크루스칼 왈리스 검정을 수행했다. 영가설은 ‘모든 집단의 평균은 동일하다.’로, p-value 2.203231003234039e-10으로 유의수준 0.05보다 매우 작아 영가설이 기각된다. 즉, 각 그룹(가족외동행여행만족도, 타가구포함가족여행만족도, 가구원한정가족여행만족도, 나홀로여행만족도)별로 유의미한 차이가 있다는 결론을 얻었다.

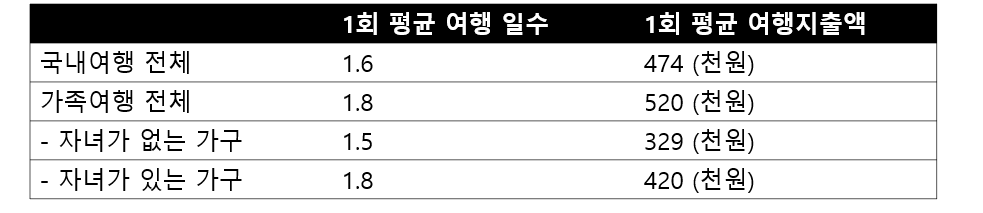
**3. 국내 가족여행 분석**

**(1) 사용한 데이터셋**

2019년 국민여행 실태조사 데이터를 분석에 사용함. 해당 데이터는 2019년 국민여행 조사의 결과를 수록한 것으로, 국내편과 해외편으로 구성되어 있음. 본 연구에서는 국내여행만 고려하므로 국내편 데이터만 사용함. 또한 본 데이터에 수록된 여행의 종류로는 관광여행, 기타여행이 있는데, 관광여행은 여행의 주된 목적이 관광/휴양인 경우이며 기타여행은 여행의 주된 목적이 출장/업무. 단순 귀성이나 단순 친구/친지 방문인 경우에 해당한다. 본 연구에서는 관광여행에 대해서만 고려했다.

**(2) 여행일수/지출액 분석**

가족 여행시, 1회 평균 여행 일수는 1.80일로 국내여행 전체보다 높다. 또한 가족 여행시 1회 평균 여행 지출액은 520천원으로 국내여행 전체보다 높다.



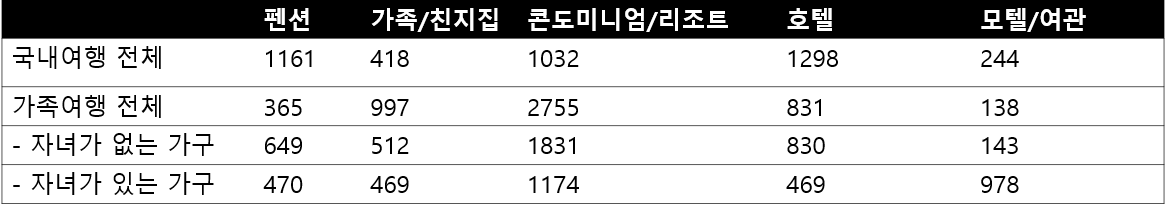
**(3) 이동수단 분석**

본 연구에서 사용한 데이터셋에서 구성된 이동수단 카테고리는 총 11개로, 자가용, 철도, 항공기, 선박/해상교통, 지하철, [정기]고속/시외/시내버스, [부정기]전세/관광버스, 차량대여/렌트, 자전거, 택시, 기타로 이루어져 있다. 전체 분포를 확인한 결과 하위 6개의 범주에 대해서는 아주 낮은 비율을 보여, 해당 6개는 기타에 포함시키고 분석했다. 또한 본 데이터에서는 1순위 이동수단, 2순위 이동수단, 3순위 이동수단으로 구성된 구분단위를 각 가중치를 지정하여 하나의 이동수단 칼럼으로 생성했다. 가족여행 전체에서의 자가용 이용 비율이 국내여행 전체보다 높았다.



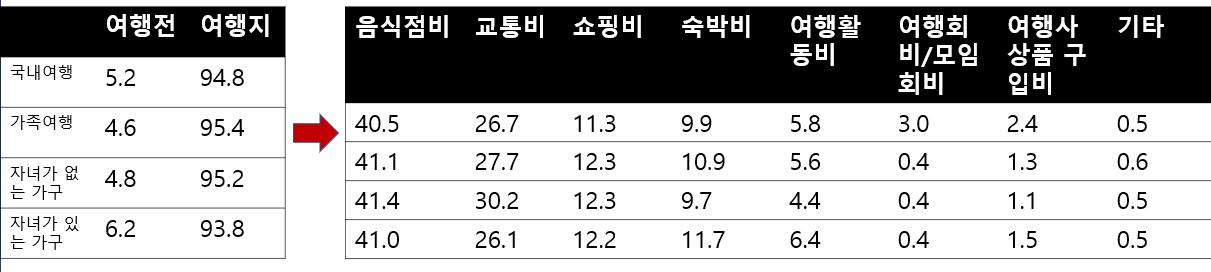
**(4) 숙박시설 분석**

숙박시설의 경우 총 14개의 카테고리, 호텔, 모텔/여관, 펜션, 콘도미니엄/ 리조트, 유스호스텔, 야영(캠핑), 연수원/수련원, 전통 한옥 숙박시설, 게스트 하우스, 민박(도시형/농촌형), 자연 휴양림 내 숙박시설, 무박여행, 기타로 이루어져 있다. 이 중 비율이 낮은 하위 7개를 기타로 통합하여 분석을 진행했다. 가족여행시, 가족/ 친지집, 콘도 미니엄/ 리조트에서의 숙박 비율이 국내여행 전체보다 높았다.



**(5) 여행 전/여행지 지출**

가족 여행시 여행 전 지출 비중이 국내여행 전체보다 높다. 또한 가족 여행시 음식점비, 교통비, 쇼핑비, 숙박비의 비중이 국내여행 전체보다 높았다.



**4. 여행지역에 따른 만족도 차이 분석**

**(1) 사용한 데이터셋 & 변형 방법**

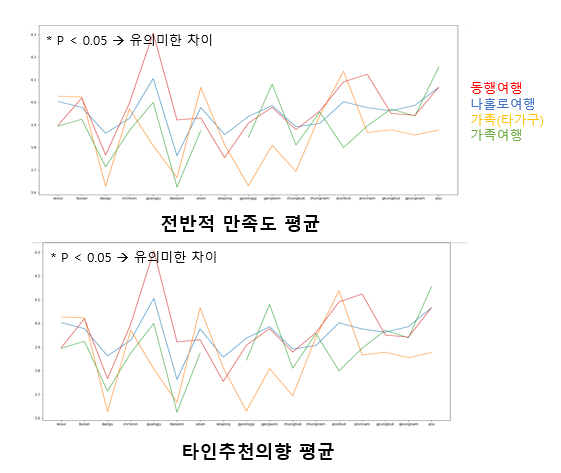
2019 국민여행 실태조사 데이터를 사용하여 분석함. 여행 구분에서 ‘해외여행’, ‘가구 국내 관광여행/해외여행 없음’, ‘기타여행’, ‘북한여행’을 제외한 순수 국내 관광여행만을 고려하여 분석함. 또한 설문지 데이터에서 응답자 id가 없으면 응답에 신빙성이 없다고 판단하여 데이터를 제외했다. 데이터 레코드의 경우는 ‘명’이 기준으로 되어있다. 이 경우, 한 레코드에 여행한 횟수만큼 문항이 다중응답이 되어있으므로, 분석에 어려움이 있어 데이터 레코드 형식을 건으로 바꾸어 활용했다. 마지막으로는 세가지 파생변수를 생성했다. 첫 번째 파생변수는 Group으로 여행종류, 즉 동행여행, 나홀로 여행, 타가구 포함 가족여행, 가구원만 있는 가족여행의 4가지 그룹으로 선정하여 새로운 변수를 생성했다. 두 번째 파생변수는 COUNT로, 지역별 여행지에서의 활동별 합계를 계산했다. 세번째 파생변수는 여행한 시기로, GOLD변수에 황금연휴 시기를 직접 계산하여 구성했다.

**(2) 그룹과 지역에 따른 차이 검정**

여행 유형 (동행여행/나홀로여행/타가구포함 가족여행/가족여행)에 따라 여행한 지역을 분석한 결과이다. 이 결과, 동행여행의 경우는 강원도, 나홀로 여행의 경우는 서울, 타가구포함 가족여행의 경우는 경남, 가족여행의 경우는 경기의 빈도가 가장 높았다.

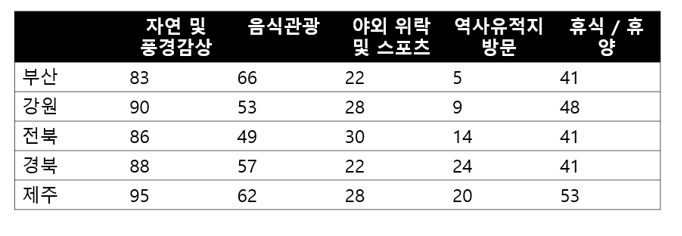


아래의 그림은 여행 지역간 만족도의 평균 차이를 보이고 지역과 그룹 간 상호작용의 효과가 있음을 보이는 그림이다. 실제로 각 지역별 만족도의 평균간 정규성과 등분산성을 만족함을 확인했고, ANOVA검정시 영가설 기각으로 유의미한 차이가 있음을 확인한 바가 있다



**(3) 그룹별 여행지역 (특정지역 선정)에 따른 만족도 차이 분석**

앞선 분석 결과, 전반적인 만족도가 높은 제주를 제외하고, 나홀로여행보다 다른 유형의 여행의 만족도가 높을 것 같은 부산, 전북이 나홀로 여행의 만족도가 가장 높았다. 부산 및 전북이 다른 그룹에 비해 만족도가 높음을 확인했고, 강원(휴식/휴양)에서는 단체 여행을 이용한 가족여행이 다른 그룹에 비해 만족도가 높았다.

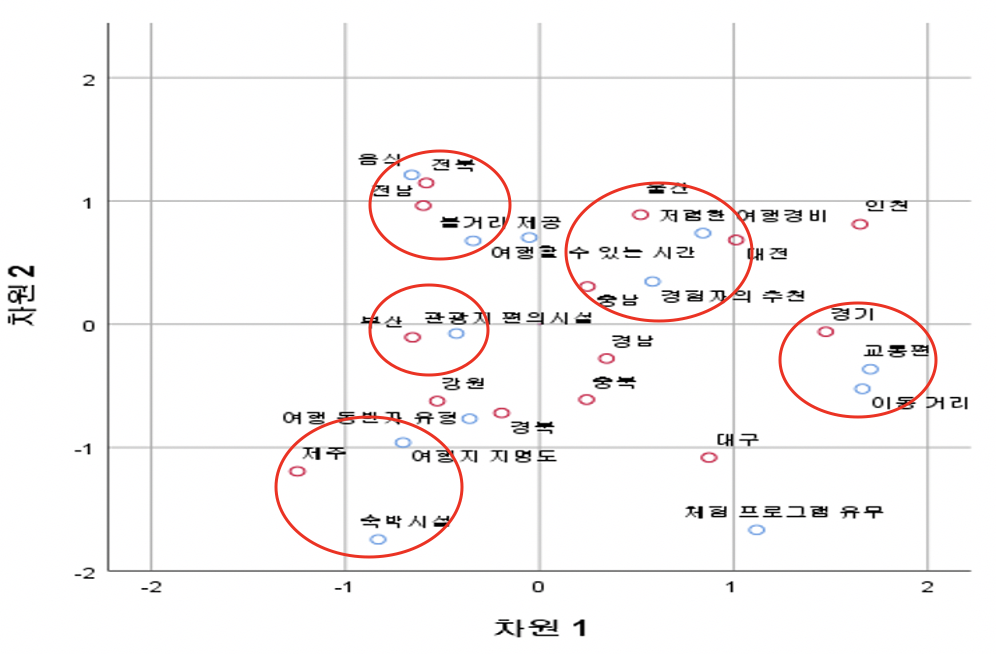


**5. 여행지역에 따른 선택 이유 및 활동 분석**

**(1) 사용한 데이터셋 & 분석방법**

2019 국민여행 실태조사 데이터를 사용하여 분석했다. 여행지역에 따른 선택 이유 및 활동을 분석하고자 SPSS 최적화 척도법을 활용해 다중대응일치분석을 진행했다. 분석을 여러 번 시행하면서 동떨어져 있는 데이터를 제거하였고, 최종적으로 분석이 완료된 표를 구성했다.

1. 여행지역에 따른 선택 이유

  
경기 지역은 교통편과 이동 거리를 이유로 여행지로 선택한 것으로 보인다.

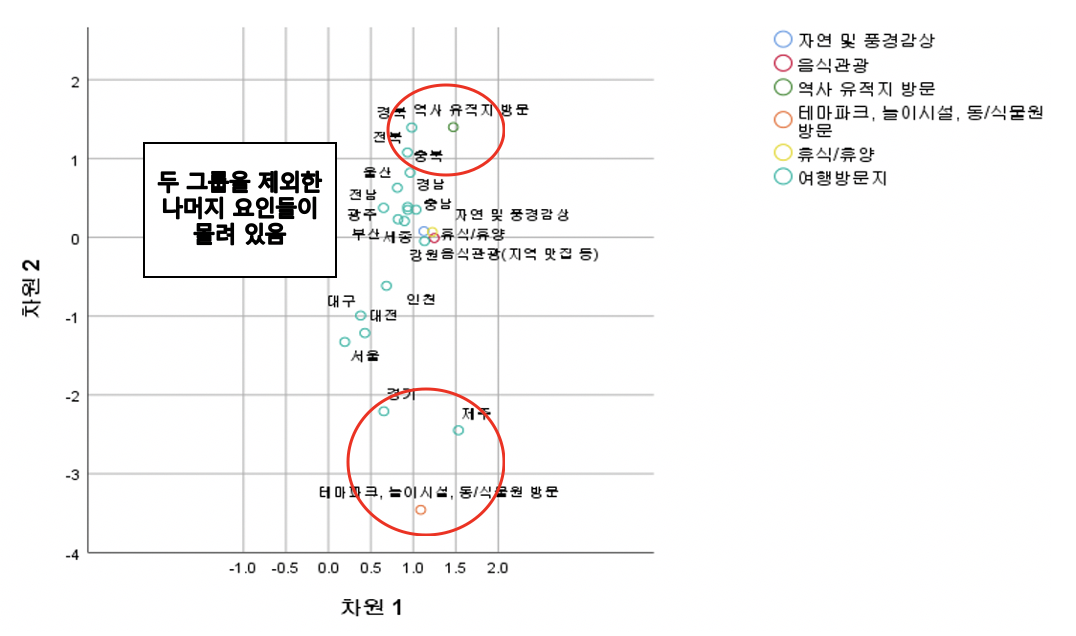
전남 전북 지역은 음식(먹거리) 및 볼거리 제공을 이유로 여행지로 선택한 것으로 보인다.

부산 지역은 관광지 편의 시설을 이유로 여행지로 선택한 것으로 보인다.

울산 충남 대전 지역은 저렴한 여행경비 및 여행 경험자의 추천을 이유로 여행지로 선택한 것으로 보인다.

제주 지역은 숙박 시설 및 여행지 지명도를 이유로 여행지로 선택한 것으로 보인다.

1. 여행지역에 따른 활동 분석



경북, 전북, 충북 지역은 역사 유적지 방문을 주로 한 것으로 보인다.

경기, 제주 지역은 테마파크, 놀이시설, 동/식물원 방문을 주로 한 것으로 보인다.

위 5개 지역을 제외한 나머지 지역은 자연 및 풍경 감상, 음식관광, 휴식/휴양을 주로 한 것으로 보인다.

**6. 여행 시기에 따른 여행 활동, 만족도, 여행지 분석**

**(1) 사용한 데이터셋 & 분석방법**

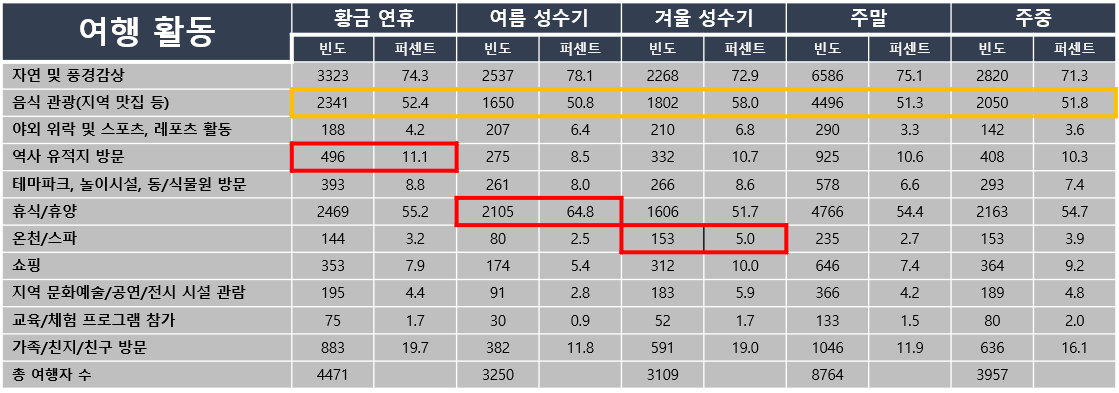
2019 국민여행 실태조사 데이터를 사용하여 분석했다. 여행 시기에 따른 여행 활동, 만족도, 여행지 분석하고자 시기 별로 데이터를 분류하여 개별의 데이터셋을 구성하였다. 여행시기는 황금연휴, 여름 성수기, 겨울 성수기, 주말(황금연휴와 성수기 제외), 주중(황금연휴와 성수기 제외)로, 총 5 그룹으로 나누어 데이터 시각화를 통해 분석을 진행하였다.

**(2) 5 그룹으로 구성된 여행 시기의 정확한 기간**



2019년의 황금연휴, 여름 성수기, 겨울 성수기, 주말, 주중은 위와 같은 기간으로 선정하였고, 연차 휴일일 경우를 고려하여 기간을 선정하였다. 황금 연휴, 여름 성수기, 겨울 성수기, 주말, 주중에 대한 여행 데이터셋을 각각 생성하여 분석에 사용하였다.

**(3) 여행 시기에 따른 여행 활동, 만족도, 여행지 분석**

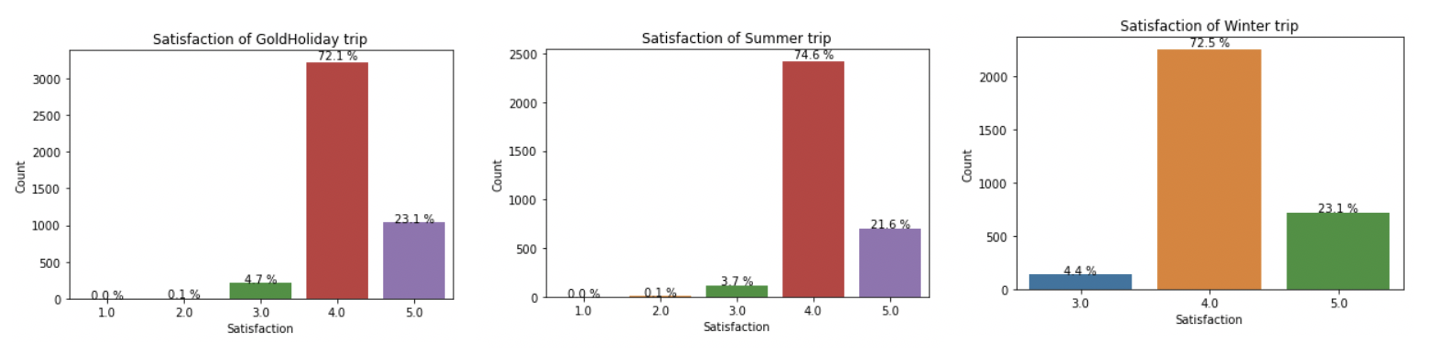
1. **여행지역에 따른 선택 이유**  
   

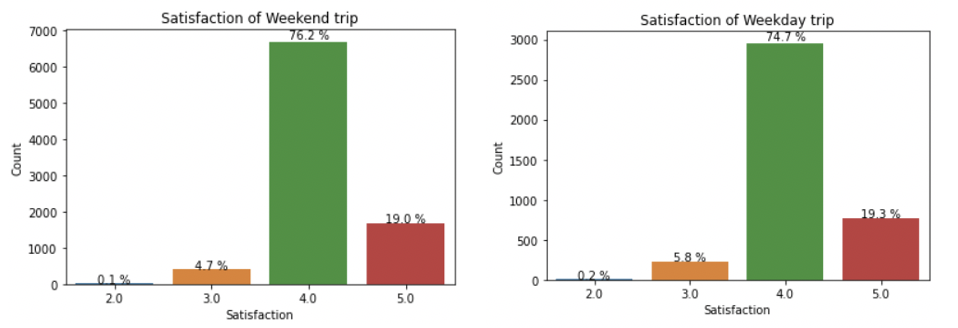
여행 시기별 여행 활동을 분석해 보면 위와 같이 11가지 활동(①자연 및 풍경 감상, ②음식 관광, ③야외 위락 및 스포츠, 레포츠 활동, ④역사 유적지 방문, ⑤테마파크, 놀이시설, 동/식물원 방문, ⑥휴식/휴양, ⑦온천/스파, ⑧쇼핑, ⑨지역 문화예술/공연/전시 시설 관람, ⑩교육/체험 프로그램 참가, ⑪ 가족/친지/친구 방문)에 대해 여행 시기별 빈도수를 알 수 있다.

빈도수, 퍼센트를 통해 다음과 같은 분석 결론을 도출할 수 있다.

* 음식관광은 전체적으로 유사한 비중을 나타냄.
* 황금연휴때는 역사 유적지 방문이 다른 때보다 높은 비중을 나타냄.
* 여름 성수기때는 휴식/휴양이 다른 때보다 높은 비중을 나타냄.
* 겨울 성수기때는 온천/스파가 다른 때보다 높은 비중을 나타냄.

1. **시기별 여행 만족도 차이 분석**

****

****

여행 시기별(황금연휴, 여름 성수기, 겨울 성수기, 주말, 주중)에 따른 전반적인 여행의 만족도를 시각화하면 위와 같다.

차트를 통해 다음과 같은 분석 결론을 도출할 수 있다

* 겨울 성수기의 경우 만족도가 나쁨(만족도 1,2)인 경우는 없다.
* 황금연휴와 여름 성수기에 다녀온 여행의 경우 매우 비슷한 만족도를 보인다.

1. **여행 시기별 주 이용지역 Top5 선정**

**테이블이(가) 표시된 사진

자동 생성된 설명**

**테이블이(가) 표시된 사진

자동 생성된 설명**

여행 시기별(황금연휴, 여름 성수기, 겨울 성수기, 주말, 주중)에 따라 주 이용지역 Top5를 정리해보면 위와 같다.

* 여름, 겨울 성수기에 강원도가 가장 높은 이용 빈도를 보인다.
* 황금연휴, 주말의 경우 경기도가 가장 높은 이용 빈도를 보인다.
* 주중에는 서울특별시가 유일하게 top5안에 포함. 전라남도는 모든 기간에 top5안에 포함된다..

**7. 코로나 데이터 탐색적 분석**

코로나 데이터를 바탕으로 탐색적 분석을 통하여 전반적인 코로나 감염 현황을 확인한 후 군집분석(clustering)을 사용하여 위험군을 선정한다. 이후에 시계열 모델과 순차(sequential) 모델인 양방향 LSTM모델을 사용하여 확진자를 예측했다.

**(1) 일자별 신규 & 누적 확진자수 확인**

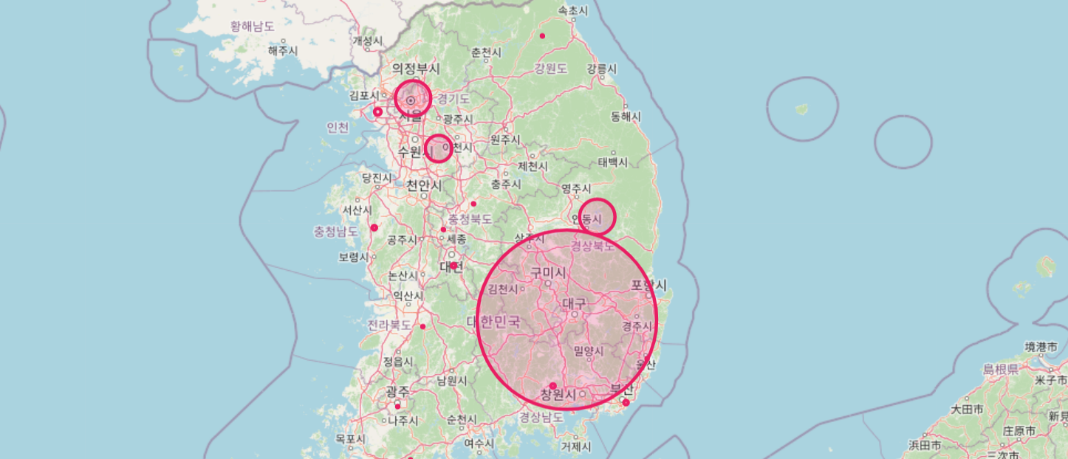
**(1)-1. 일별 확진자수를 기존 데이터에 추가**

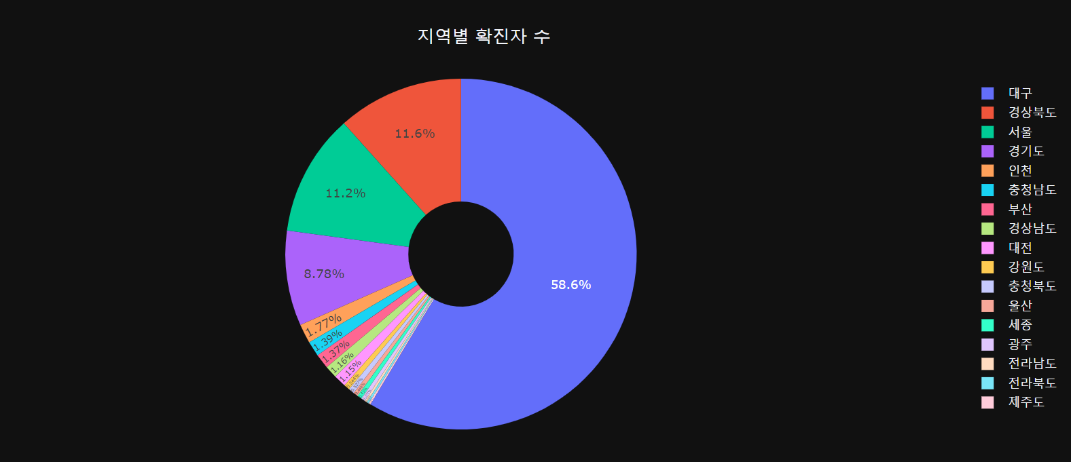
가장 먼저, 기존 데이터에 일별 확진자 데이터가 없음을 확인한 수에 일별 확진자수를 기존 데이터에 추가하도록 하는 함수를 작성하여 데이터를 구축했습니다. 이를 바탕으로, 일자별 누적 확진자수와 일자별 확진자수를 시각화한 결과, 3월에 가장 많은 확진자가 나왔음을 확인할 수 있고, 다시 잠잠하다 2020년 6월부터 재확산 되었음을 확인할 수 있다.

가장 먼저, 기존 데이터에 일별 확진자 데이터가 없음을 확인한 수에 일별 확진자수를 기존 데이터에 추가하도록 하는 함수를 작성하여 데이터를 구축했습니다. 이를 바탕으로, 일자별 누적 확진자수와 일자별 확진자수를 시각화한 결과, 3월에 가장 많은 확진자가 나왔음을 확인할 수 있고, 다시 잠잠하다 2020년 6월부터 재확산 되었음을 확인할 수 있다.



**(1)-2. 코로나 19 확진자 map**

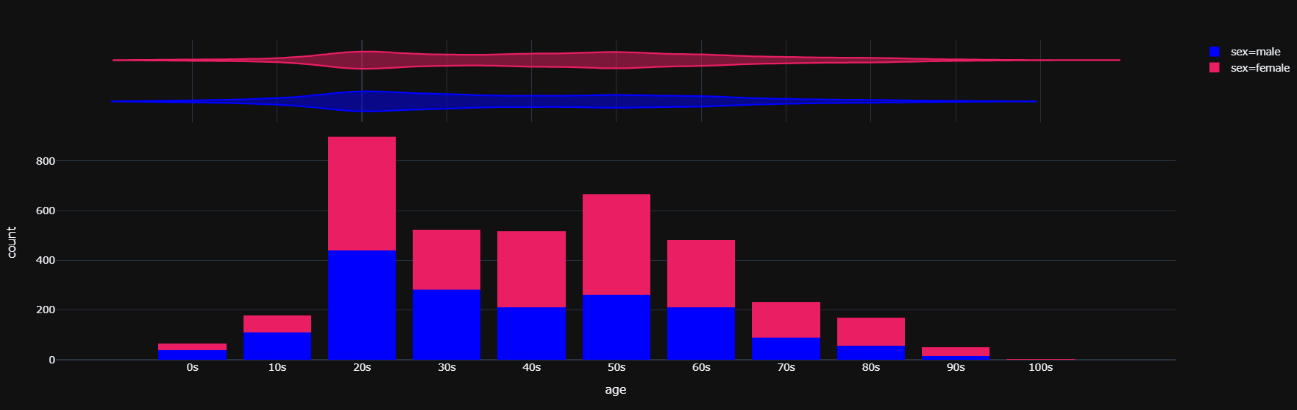
각 지역의 위,경도 데이터를 활용하여 지역별 감염 추세를 살펴보았다.



위의 pie chart를 확인하면 대구, 경북, 경기 순으로 코로나 확진자가 있음을 확인할 수 있다. 대구는 신천지 집단 감염 사태로 인해 확진자가 폭증했는데 이와 인접한 결과이다.

최근 확진세에 의해 대구,경북을 제외하면 수도권 지역 확진자수가 가장 많다.

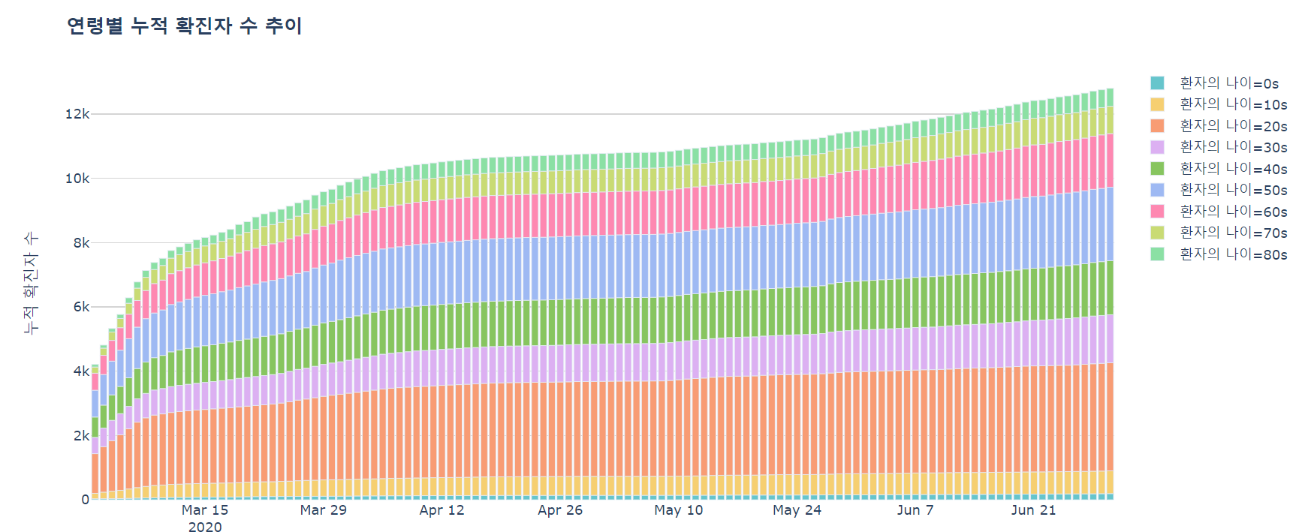
**(1)-3. 성별 / 나이대별 확진자수 분석**

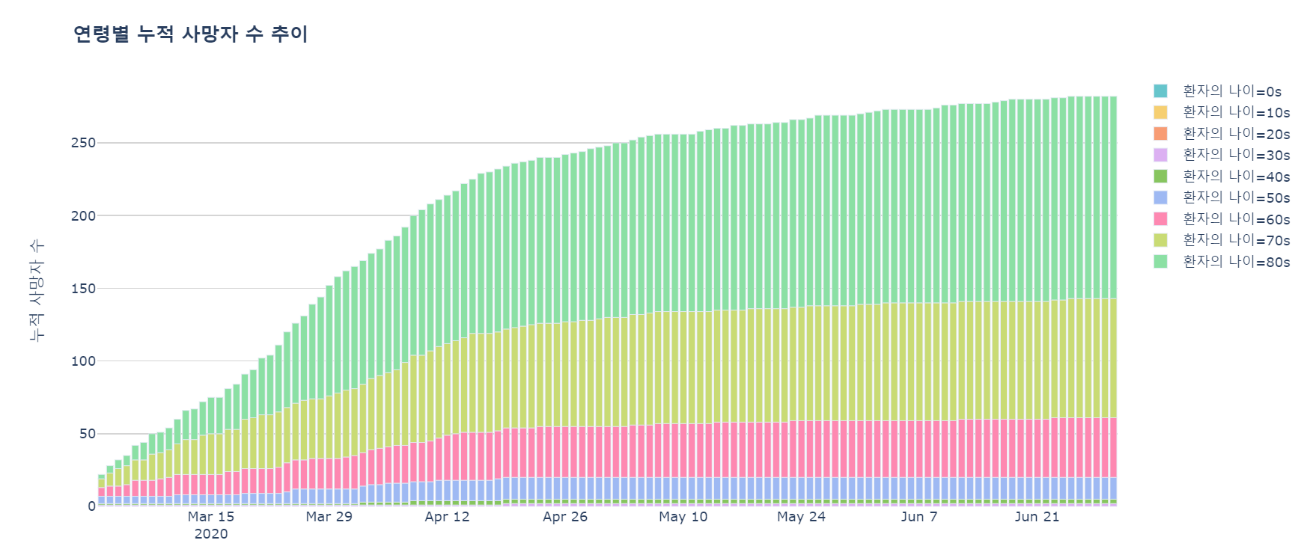


위의 barplot을 확인하면 20대의 확진자 비율이 가장 높음을 확인해 볼 수 있다. 이는 20대가 10대보다 비교적 활동하는 반경도 없고, 30대보다 비교적 여유로운 시간을 가진다는 점에서 나온 결과이다.

**(2) 연령별 누적 확진자 및 사망자 분석**

확진자 데이터에서 환자의 나이를 연령대별로 카테고리를 나누어 범주형 변수로 바꿔줬다. 이를 바탕으로 연령별 누적 확진자수와 사망자수 추이를 살펴본 결과는 아래와 같다.

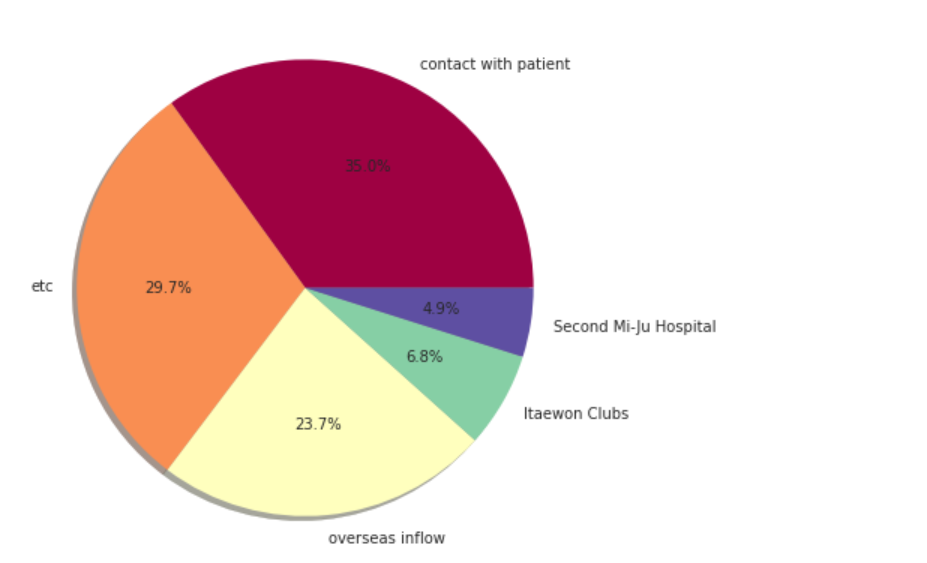




위의 결과를 확인하면, 연령별 확진자수는 청년층 38%, 중년층 31%, 고령층 24%, 미성년자 7% 수준이다. 반면 사망자수를 확인하면 고령층이 압도적으로 많음을 확인할 수가 있다. 즉 고령층은 다른 연령대보다 코로나가 생사에 더 치명적이라고 할 수가 있다.

**(3) 감염이유 분석**

코로나 감염 이유 데이터를 사용하여 짐단 감염 사례 중 상위 6개를 확인했다. 6개를 확인한 결과 신천지 집단 감염 사례는 독보적이므로 그를 제외한 나머지 다섯가지 사례에 대한 pie chart를 생성하여 시각화 했다.

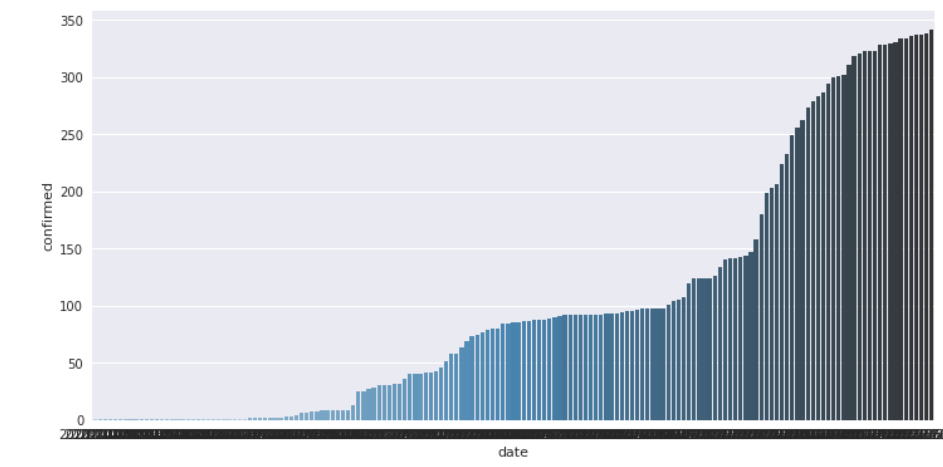


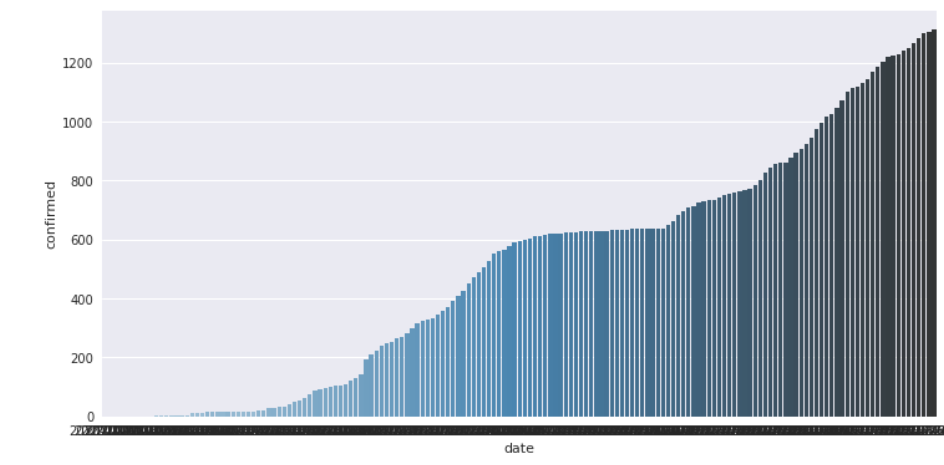
위의 감염 이유를 확인하면, 해외유입을 제외하면 대부분 집단감염이 가장 큰 원인이었음을 확인할 수가 있다.

(**4) 확진자수를 기준으로 위험군 선정**

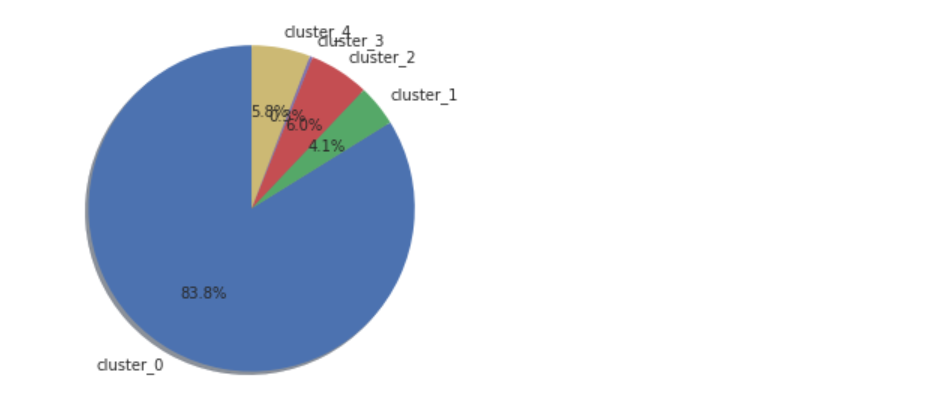
코로나로 인한 사회적 거리두기 기간이 길어짐에 따라 사람들이 코로나의 위험성에 대한 경각심이 사라지고 있는 상황이다. 따라서 이동하는 지역이 확진자 수가 높은 위험 지역으로 분류된다면 알림을 주도록 하는 기능을 추가하고자, 군집분석(clustering)을 사용하여 위험군을 선정했다.

위험군을 선정하기 전, 몇 지역에 대해 확진자수 추이를 살펴본 결과는 아래와 같다. 대표적으로 인천, 서울지역 확진자에 대한 시각화를 순서대로 첨부한다.

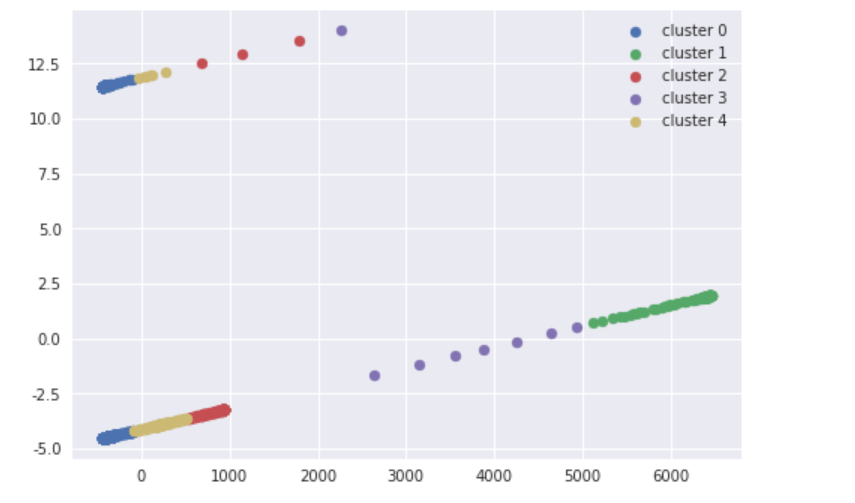




지역별 확진자수 데이터를 바탕으로 k-means clustering을 적용하여 총 5개의 군집을 만들다. 차례대로 최고위험군, 고위험군, 보통, 저위험군, 최저위험군으로 구분했다. 클러스터링 결과는 아래와 같다.

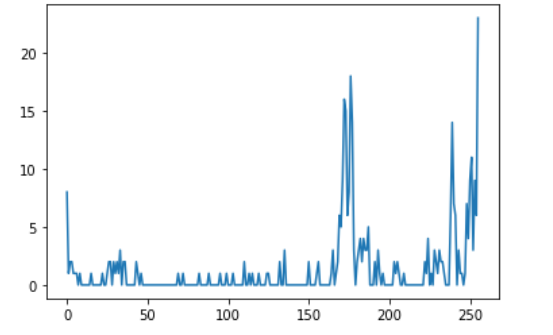


클러스터링이 잘 되었는지를 확인하기 위해 PCA를 적용한 결과는 아래와 같고, 클러스터링이 잘 이루어 졌음을 확인할 수 있다.



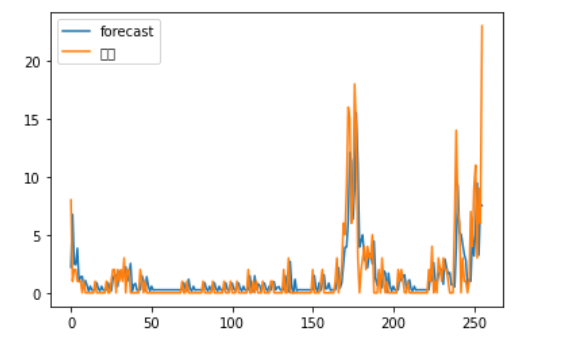
**8. 코로나 확진자 예측 모델링**

먼저 전체 확진자 데이터를 지역별로 분할하여 따로 데이터를 저장했다. 이를 바탕으로 시간대별 확진자 추이를 지역별로 확인했다. 총 17개의 지역이 있었으며, 이 보고서에서는 강원지역에 대한 시각화와 모델링 결과를 첨부한다.



날짜에 따른 서울지역 확진자 추이는 위와 같다. 위의 그래프를 확인하면 중간중간 (3월, 6월)등에 이벤트가 생기면서 확진자가 급증하는 형태를 확인해볼 수가 있다. 따라서 그런 불규칙한 이벤트를 고려하는 시계열 모델을 사용해야 한다고 판단했다.

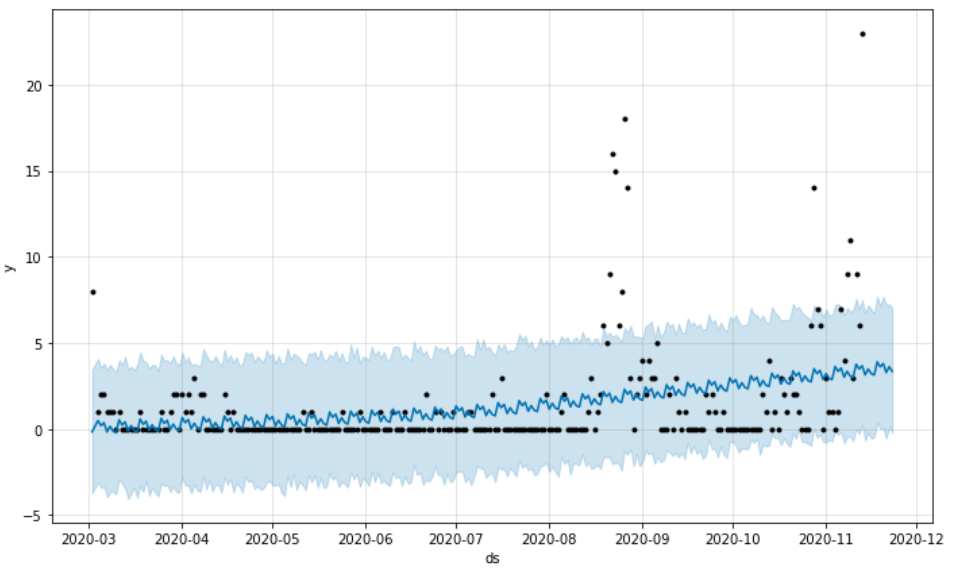
사용한 모델은 총 3가지로, ARIMA, facebook prophet, bidirectional LSTM모델이다. ARIMA의 경우는 기초적인 시계열 분석 통계모델로 사용되는 AR, MA, ARMA모델들이 분석하는 데이터의 추세와 계절성 뿐 아니라 데이터의 자기상관성까지 고려하는 모델이므로 사용했다, 아래의 결과는 ARIMA 모델을 적용하여 예측한 결과를 시각화한 형태이다.



(\* 파란색 선은 실제 결과이고 주황색 선은 예측한 결과이다)

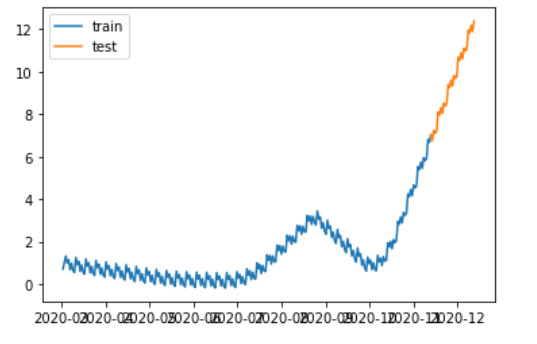
위의 결과를 확인하면 예측결과가 그다지 좋지 않음을 확인할 수 있다. 예측이 잘 되지 않은 이유를 불규칙한 이벤트 발생이라고 보고, non-linear growth과 불규칙 이벤트를 고려하는 모델인 facebook prophet을 사용했다.

facebook prophet는 growth, seasonality, holidays의 세가지 요소를 고려한다. 먼저 growth의 경우는 linear growth, non-liear growth를 고려한다. seasonality의 경우는 사용자들의 행동양식으로 주기적으로 나타나는 패턴을 나타내며, 푸리에 급수를 사용해 패턴의 근사치까지 찾아준다. 마지막으로 holidays는 주기성을 가지진 않지만 전체 추이에 큰 영향을 주는 이벤트를 분석해 준다.



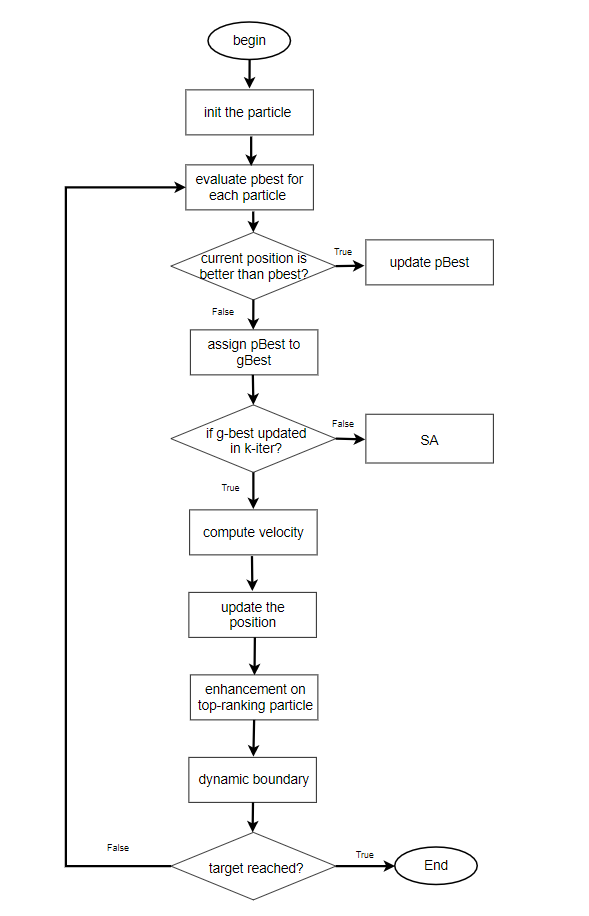
위의 결과는 facebook prophet 모델을 사용하여 확진자 수를 예측한 결과를 시각화한 형태이다. 검은색 점은 실제 확진자를 의미하여, 파란색 선은 예측한 결과를 나타낸다. 비교적 ARIMA에 비해 이벤트 예측이 잘 되며, 빨간색 점선으로 change point를 그려본 결과, 포인트를 잘 잡아냄을 확인했다.

마지막 bidirectional LSTM 모델을 사용하여 예측한 결과는 아래와 같다. 먼저 train, test데이터를 만들기 위해 데이터를 섞지 않고 특정 시점을 기준으로 학습용, 검증용 데이터를 구분했다. min-max scaler를 사용하여 스케일링을 진행했으며, 모델에 적용이 가능한 데이터 형태로 가공했습니다. Bidirectional-LSTM모델의 경우는 이전 시점의 결과를 다음 시점에 반영하므로 시계열 모델에 최적화되어 사용하게 되었다. 아래는 앙뱡항 LSTM을 사용하여 예측한 결과를 시각화한 형태이다.

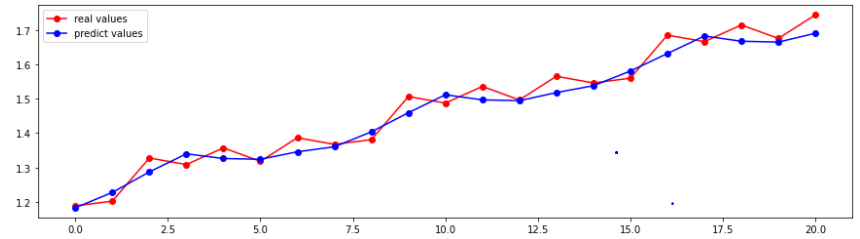


학습 결과는 위와 같다. loss값은 0.0062로 학습이 잘 되었음을 확인할 수가 있다.

위의 결과는 양방향 LSTM모델을 사용하여 predict한 결과를 시각화한 형태이다. 또한 하이퍼파라미터 튜닝 방법으로는 AutoML 방법중 Particle swarm Optimization(PSO)에 Genetic Algorithm의 연산 concept을 사용하여 변형한 방법을 사용했다. 아래는 사용한 알고리즘을 flowchart 그린 그림이다.



마지막으로 세가지 모델을 사용한 결과, 학습 결과가 좋았던 Facebook prophet, RNN 모델을 앙상블 하여 예측 결과를 얻었다. 예측결과는 아래와 같다.



**9. 결론**

**‘2. 유형에 따른 만족도 비교’**에 따르면, 나홀로여행보다 가족들(가구원)과 함께 떠나는 여행이 만족도가 높았다. 가구원의 경우 불가피하게 거리 두기를 할 수 없는 경우이므로, 높은 만족도를 위해 가구원과 함께 떠나는 것을 권장하는 바이다.

또한 **‘6. 여행 시기에 따른 여행 활동, 만족도, 여행지 분석’**에서 겨울 성수기에는 강원도, 전라남도, 경기도, 경상북도, 충청남도에 많은 여행객들이 몰리는 것을 확인할 수 있었다. Modeling 파트에서 지역별 확진자 수를 예측해본 결과, 특히 강원도에서 확진자 수가 증가하는 경향을 확인했다.

반면 대구광역시, 대전광역시, 인천광역시, 광주광역시, 울산광역시에는 상대적으로 적은 여행객들이 방문하는 것을 알 수 있었다.

따라서 위험한 TOP5 지역에 여행객들이 몰리는 현상을 방지할 필요가 있다. (포스트 코로나 시대의 안전한 여행을 위해서는 오버 투어리즘 방지가 이전보다 더 중요하다고 볼 수 있다.) 🡪 여행객 수 하위 5개 지역의 여행상품을 홍보하여 TOP5 지역에만 여행객들을 몰리는 것을 방지한다. ‘**5. 여행지역에 따른 선택 이유 및 활동 분석’**에 의하면 대구광역시, 대전광역시, 인천, 광주광역시, 울산광역시의 경우 음식 관광, 휴식, 휴양을 목표로 가는 경우가 많으므로 이를 중심으로 홍보를 하면 된다.