

민원데이터를 활용한 빅데이터 분석 및 시각화

주제 : 코로나 확산 방지 목적의 효율적인 재난문자 활용

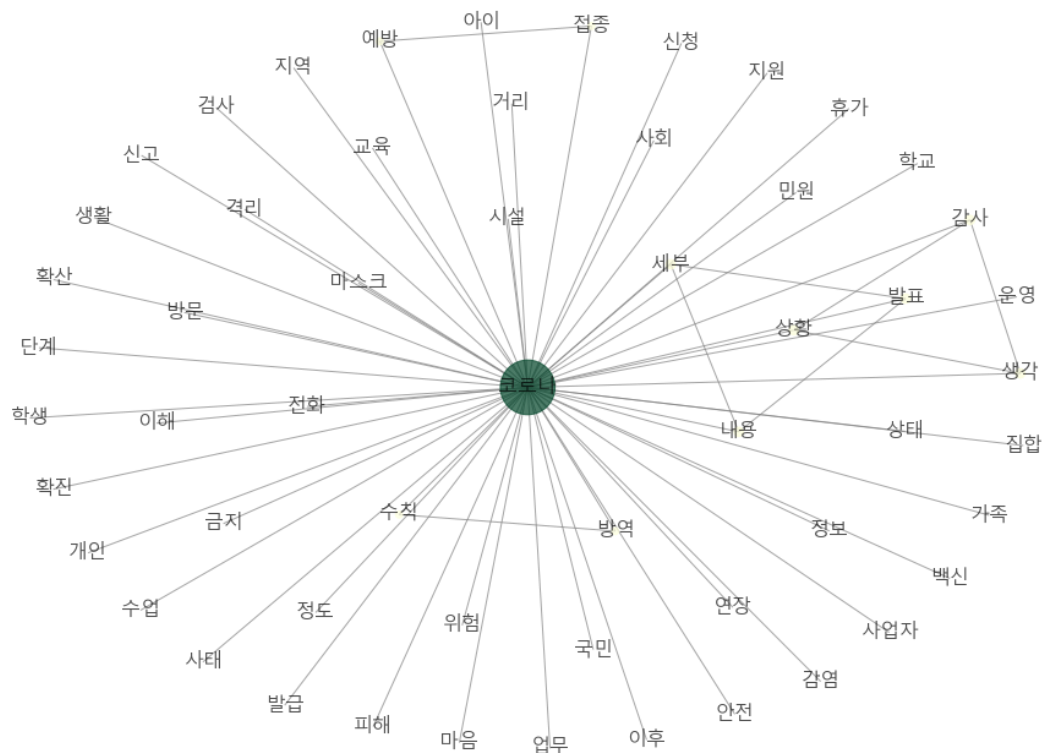
팀명 : 기파랑간부
팀원 : 황현수, 유민서

목 차

- 01. 주제 선정 배경
- 02. 분석 목적(인사이트)
- 03. 데이터 설명
- 04. 분석 결과 및 한계점

주제 선정 과정 : 코로나관련 민원 분석(연관어 분석)

- 2020년부터 현재까지 지속되는 코로나 사태를 증점으로 두고 민원 데이터 *전처리 후 연관어분석 실시
* 전처리 : 키워드(코로나, 코로나19, 확진) / 형태소분석 : mecab 사용 / 불용어사전 만든 후 불용어 제거
- 전처리 후 나온 코로나 데이터의 연관어 분석 결과만으로는 주제 선정이 어려워 토픽 모델링 실시



< 코로나 데이터 연관어 분석 결과 >

주제 선정 과정 : 코로나관련 민원 분석(토픽모델링)

• 토픽모델링(8개의 토픽) 분석결과 :

Topic 0: #격리 #해제 #접촉 #확진 #코로나 #기준 #필요 #검사 #접촉자 #자녀
확진자가 격리해제된 후 다른 확진자와 접촉하면 어떡하나요
확진자 격리해제 후 재검출되면 어떡하나요
코로나 확진자 격리해제기준
코로나19 자가격리통지서
자가격리 생활지원비 신청

Topic 1: #문자 #재난 #지진 #긴급 #발송 #무엇 #문의 #방송 #방법 #어디
찾은 휴대폰 재난문자
코로나19 안내문자
재난문자 수신 문의
예비저감조치 시에는 재난문자를 발송하나요?
폭설 긴급재난문자 추가내용

Topic 2: #단계 #두기 #거리 #피해 #필요 #감염 #발생 #코로나 #대구 #이상
전북 전주시 거리두기 2단계 상향 신중히 생각해주십시오.
사회적거리 두기 4단계 실시 요청
대구 사회적 거리두기 2단계
거리두기 2.5단계 유지에 따른 영내휴가 사용 허가 및 추가 연가저축 허용 검토 요청드립니다.
코로나 관련 불법영업문의

Topic 3: #학교 #코로나 #개방 #운동장 #외부 #수업 #확진 #증가 #지속 #접촉자
학교 운동장 개방 관련 질의
학교 운동장 개방 관련 질의
동경한국학교의 전면 원격수업 운영에 대한 문의
000학원 방역지침 위반으로 인한 폐쇄
학교내 코로나19 확진자가 발생하였는데 타학년 학생들은 등교수업을 해도 되는지 여부



주제 선정 과정 : 코로나관련 민원 분석(토픽모델링)

- 토픽모델링(8개의 토픽) 분석결과 :

Topic 4: #정보 #공개 #관련 #동선 #이동 #확진 #방통 #삭제 #세부 #발표

[보도자료] 방통위, 확진자 동선정보 삭제관련 대응 강화키로

확진자 이동동선 정보공개 지침 개정 요구

[보도자료] 방통위, 코로나19 확진자 동선정보 관련 사업자와 이용자의 자발적인 삭제 협조

확진자 공개 기준 근거 지침

확진환자의 이동경로 등 정보공개 불만

Topic 5: #시험 #시험장 #수능 #방역 #응시 #경우 #가족 #가격 #리자 #확진

[수능시험장 방역]시험 전 또는 당일 자가격리자 또는 확진자 통지를 받은 경우 시험 응시는

[수능시험장 방역]가족 중 확진자가 발생할 경우 시험 응시를 할 수 있나요?

[수능시험장 방역]제가 시험 보려는 시험장(**학교)에서 얼마전 확진자가 나왔다는 보도..

코로나-19 확진자와 동선 겹친 가족(학생) 등교 여부

코로나 확진자입니다.

Topic 6: #등교 #환자 #수업 #발생 #조치 #학생 #경우 #중지 #출결 #처리

(코로나19)등교 후(수업중) 확진환자 발생시 대응 문의

코로나 관련 귀가조치 및 등교중지의 경우 출결처리 문의

확진환자, 의사환자, 유증상자 등 등교중지 학생의 출결 문의

학교내 코로나19 확진자가 발생하였는데 타학년 학생들은 등교수업을 해도 되는지 여부

코로나19 예방 관련 조치

Topic 7: #마스크 #주민 #착용 #방역 #시간 #사람 #상황 #확진 #부탁 #시설

공무원도 시설 방문시 KF94 마스크 착용해주세요.

군부대 마스크 미착용

교통정리하시는분이 1시간째 마스크안쓰고계시네요

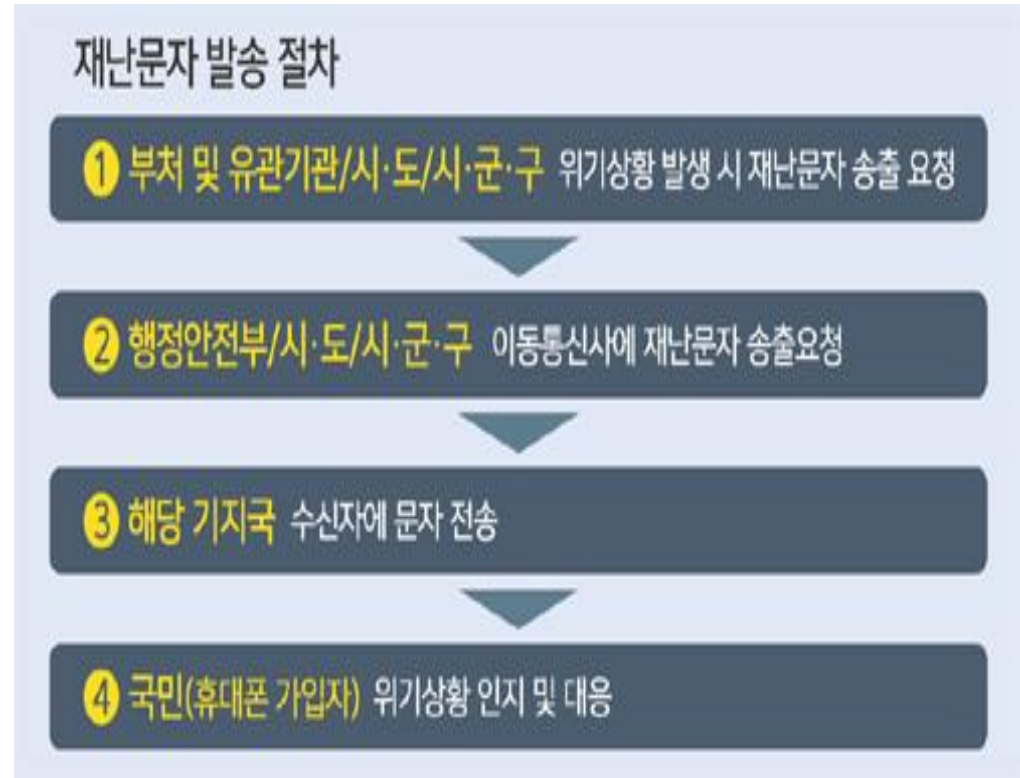
현장, 사무실 내에서 마스크 미착용자 다수로 신고

동네 평상 좀 치워주세요.



주제 선정 이유 : 재난문자는 어떤 기준으로 발송 될까?

- 세부 토픽주제 내 '재난문자', '확진자동선' 관련 불만을 확인 후, 사람들의 이동경로를 파악해 재난문자를 이용하여 코로나 감염확산 방지를 목적으로 빅데이터 분석 시작
- 행정안전부에 따르면, 재난문자는 해당지역의 통신3사 LTE 기지국을 기반으로 전파 수신 범위 약 15km 안에 있는 모든 휴대전화를 대상으로 재난문자를 전송. 즉, 현재 위치를 중심으로 수신.
- 다른 지역에서 코로나 위험 지역으로 이동하려는 대상에게도 경고 필요하다고 생각.



분석 목적(인사이트) : 장거리 여행,출장자들에게 이동자제 권고

- 지역에서 집단감염이 시작되면 근처에 있는 사람들은 재난문자를 받을 수 있지만 그 지역으로 장거리 여행이나 출장 가는 사람은? 사전에 알기 어렵다. 그렇다고 **전부 재난문자 발송은 비효율**.
- 특정지역, 특정기간 코로나에도 불구하고 장거리 출장이나 여행을 한 사람들의 데이터를 이용해 또다시 집단감염이 발생했을 때, 그지역으로 이동을 많이 했던 **지역별,연령별,성별 효율적인 범위를 추정, 선별** 해서 경고문자를 보내면 **효율적으로 이동자제를 하고 감염확산 예방 가능**

↓↓ 같은 날 안동, 세종의 재난 문자 수신내용 ↓↓

■ 위치 : 안동

< 안전 안내 문자

2021년 11월 14일 일요일

- [경상북도청] 11/14 0시 기준, 코로나19 신규확진자 42명(영천10,경주7,안동6,경산5,구미3,고령3,성주3,포항2,청도2,칠곡1), 시군홈페이지 참조 09:43
- [안동시청] 11.4(목) 12:00~13:00 안동참마손국수 종가식당(퇴계로141, 신안동) 방문자는 가까운 선별진료소에 방문하여 코로나19검사 바랍니다. 09:48
- [안동시청] 11.2(화) 12:00~13:00 보리밀식당(태사1길24-1, 북문동) 방문자는 가까운 선별진료소에 방문하여 코로나19검사 바랍니다. 09:51
- [포항시청] 11.9(화)~11.13(토) 11:00~17:00 보가주단(북구 죽도동 593-3번지) 해당시간 방문자는 보건소에서 검사 바랍니다. 10:08
- [안동시청] 11.11(월)~11.12(금) 대안식당(광석3길 6, 광석동) 방문자는 가까운 선별진료소에 방문하여 코로나19검사 바랍니다. 12:59
- [포항시청]확진자 버스동선 안내 동시간 탑승객은 보건소에서 검사바랍니다. <https://m.blog.naver.com/inpohang/222567716801> 14:56

■ 위치 : 세종

< 안전 안내 문자

2021년 11월 14일 일요일

- [충북도청] 안전한 일상회복을 위한 ▲하루 3번, 10분 이상 창문열기 ▲환풍기 수시 가동 바랍니다. ▲13일 도내 확진자 27명 발생 오전 9:35
- [세종시청] 11월 14일 0시 기준, 전일 확진자 2명(누적1,402명)발생. 타지역 확진자 접촉 2. 자세한 사항은 시 홈페이지 참고 바랍니다. 오전 10:01
- [청주시청] 11월13일 확진자 13명 발생(상당구5,서원구2,흥덕구4,청원구2)▲휴폐이지 참조▲주기적 환기▲마스크 착용▲인후통 등 유증상 시 무료검사 바랍니다. 오전 10:32

2021년 11월 15일 월요일

- [충북도청] 단계적 일상회복을 위해 ▲마스크 착용 ▲손씻기 ▲손 소독 ▲하루 3번 환기하기 ▲주기적 소독 등 방역수칙 준수, ▲14일 도내 확진자 18명 발생 오전 9:30
- [세종시청] 11월 15일 0시 기준, 전일 확진자 1명 발생(누적1,403명). 타지역 확진자 접촉 1, 자세한 사항은 시 홈페이지 참고 바랍니다. 오전 10:01
- [청주시청] 11월 14일 확진자 7명 발생(상당구2,서원구3,청원구1,타지역1)▲휴폐이지 참조▲코로나19 백신 미접종자는 예방접종에 적극 참여해 주시기 바랍니다. 오전 10:25

제1회 민원 데이터 분석 경진대회

데이터 설명

- 사용데이터 : Sk텔레콤에서 제공하는 무료데이터
- 변수 시작일, 여행지, 거주지 등 7개, 레코드 27270 개

| Dt | sido_cd | sido_nm | home_sido_nm | age_cd | sex_cd | cnt_aprx |
|-----------------|------------|------------|-----------------|-------------------|------------------|----------------|
| 주간: 시작일 (날짜) | 여행지(지역 코드) | 여행지(지역 이름) | 거주지역 (지역 이름) | 연령대 (...,70이상) | 성별 (1:남, 2:여) | 추정 여행 객수(명) |

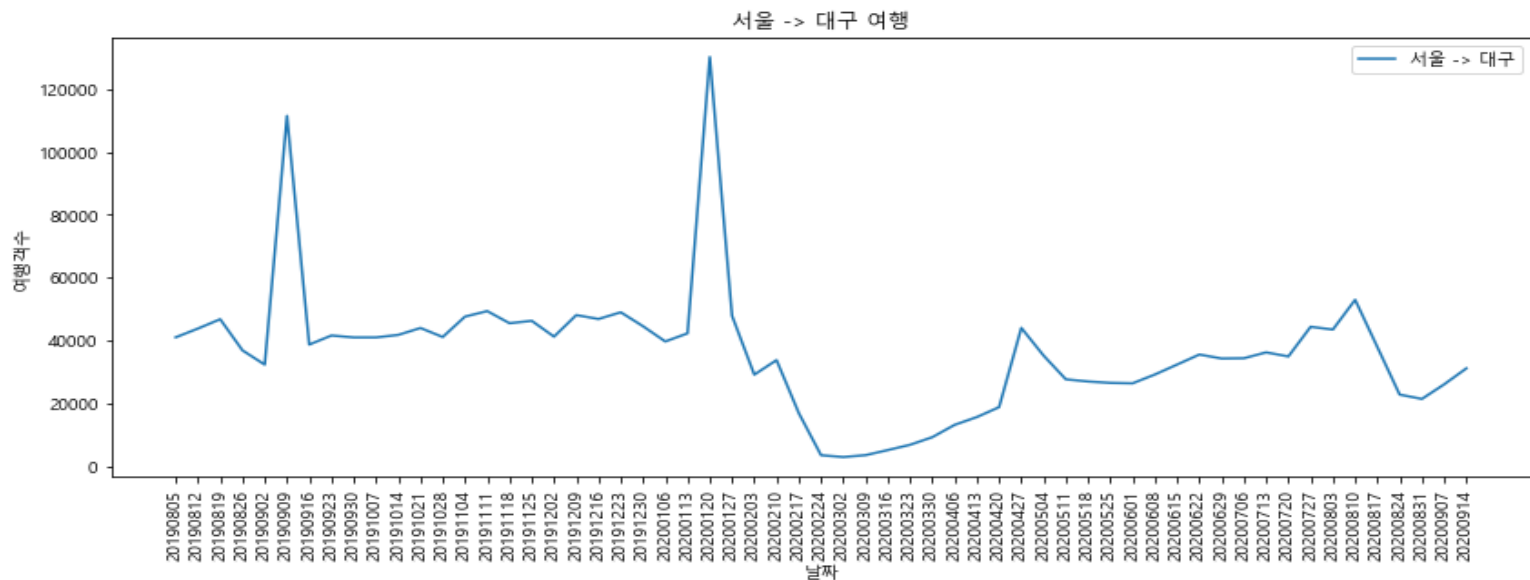
- 20190805 ~ 20200914 기간동안 광역시도별 여행 또는 단기출장 등으로 추정되는 방문자수
- 추정여행객수는 개인이 '집' 과 '회사'(또는 학교) 등 상시적인 경로 외, 이례적 이동수를 집계해 모수를 추정하는 방식
 - 추정 근거 : 1) 연령대별 스마트폰보급율(과기부, 2019)
 - 추정 근거 : 2) 이동통신 가입자 점유율(skt 자체 집계, 2019)

| Index | dt | sido_cd | sido_nm | home_sido_cd | home_sido_nm | age_cd | sex_cd | cnt_aprx |
|-------|----------|---------|---------|--------------|--------------|--------|--------|----------|
| 0 | 20190805 | 36 | 세종특별자치시 | nan | nan | nan | nan | 71455 |
| 1 | 20190805 | 26 | 부산광역시 | nan | nan | 60 | nan | 48673 |
| 2 | 20190805 | 29 | 광주광역시 | nan | nan | 30 | nan | 36233 |
| 3 | 20190805 | 29 | 광주광역시 | nan | nan | 60 | nan | 20192 |
| 4 | 20190805 | 43 | 충청북도 | nan | nan | 0 | nan | 14733 |
| 5 | 20190805 | 44 | 충청남도 | nan | nan | 10 | nan | 82124 |
| 6 | 20190805 | 11 | 서울특별시 | nan | nan | nan | 1 | 349816 |
| 7 | 20190805 | 27 | 대구광역시 | nan | nan | nan | 1 | 160154 |
| 8 | 20190805 | 41 | 경기도 | nan | nan | nan | 2 | 344912 |

| Index | dt | sido_cd | sido_nm | home_sido_cd | home_sido_nm | age_cd | sex_cd | cnt_aprx |
|-------|----------|---------|---------|--------------|--------------|--------|--------|----------|
| 27262 | 20200914 | 29 | 광주광역시 | 26 | 부산광역시 | nan | nan | 3687 |
| 27263 | 20200914 | 29 | 광주광역시 | 50 | 제주특별자치도 | nan | nan | 1839 |
| 27264 | 20200914 | 36 | 세종특별자치시 | 50 | 제주특별자치도 | nan | nan | 391 |
| 27265 | 20200914 | 41 | 경기도 | 31 | 울산광역시 | nan | nan | 7901 |
| 27266 | 20200914 | 42 | 강원도 | 43 | 충청북도 | nan | nan | 25619 |
| 27267 | 20200914 | 42 | 강원도 | 48 | 경상남도 | nan | nan | 9249 |
| 27268 | 20200914 | 50 | 제주특별자치도 | 43 | 충청북도 | nan | nan | 5591 |
| 27269 | 20200914 | 50 | 제주특별자치도 | 48 | 경상남도 | nan | nan | 11463 |
| 27270 | 20200914 | 43 | 충청북도 | 11 | 서울특별시 | nan | nan | 137640 |

데이터 추출, 그래프 표현 과정

- 데이터 추출 [서울(거주지) -> 대구(여행지)]
- 추정여행객수 자료형(숫자형) 변경
- 서울에서 대구로 이동한 추정여행객수를 분류하여 날짜별로 시각화

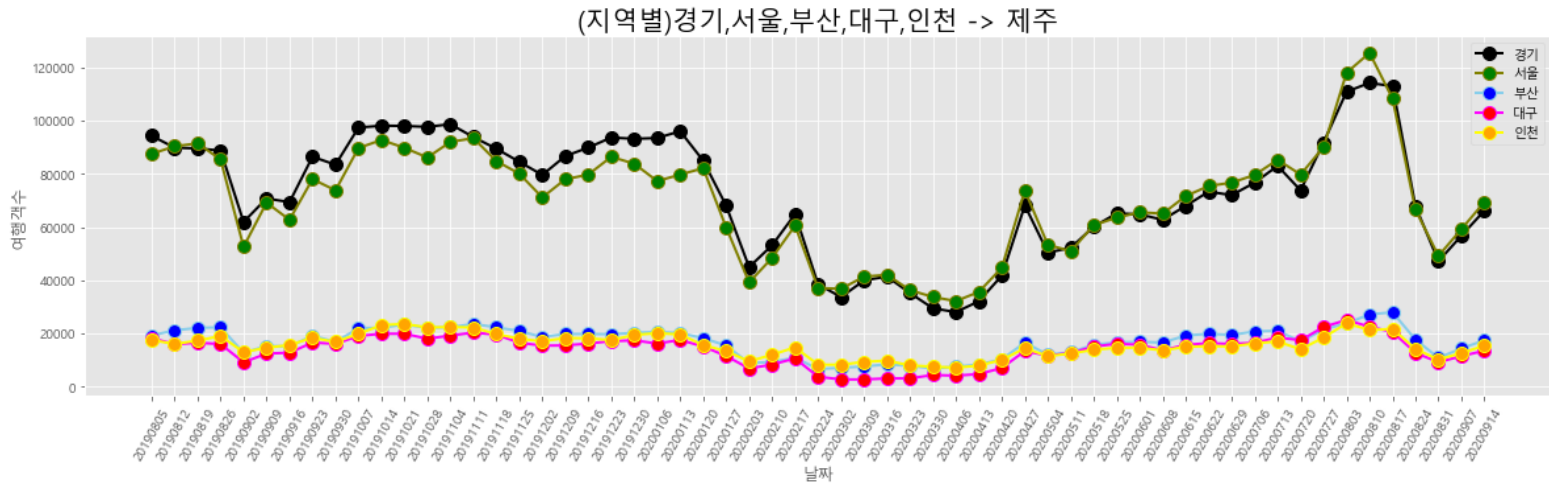


지역별 여행지(제주도) : 대표 관광지

- 데이터 추출 : [여행지(제주도)]
- ➔ 대표 관광지인 제주도를 여행지로 하고 **지역별 그래프** 표현.
- ➔ 여행지를 제주도로 하는 데이터만 추출 하여 **지역별로 분리**
- 추정여행객수 지역별 분리

| dt | sido_cd | 여행지 | ome_sido_c | age_cd | sex_cd | 총 여행객수 |
|----------|---------|---------|------------|--------|--------|--------|
| 20190805 | 50 | 제주특별자치도 | 11 | nan | nan | 87565 |
| 20190812 | 50 | 제주특별자치도 | 11 | nan | nan | 90390 |
| 20190819 | 50 | 제주특별자치도 | 11 | nan | nan | 91519 |
| 20190826 | 50 | 제주특별자치도 | 11 | nan | nan | 85427 |
| 20190902 | 50 | 제주특별자치도 | 11 | nan | nan | 52594 |
| 20190909 | 50 | 제주특별자치도 | 11 | nan | nan | 69194 |
| 20190916 | 50 | 제주특별자치도 | 11 | nan | nan | 62750 |
| 20190923 | 50 | 제주특별자치도 | 11 | nan | nan | 78046 |
| 20190930 | 50 | 제주특별자치도 | 11 | nan | nan | 73831 |

- 지역별 추정여행객수 그래프

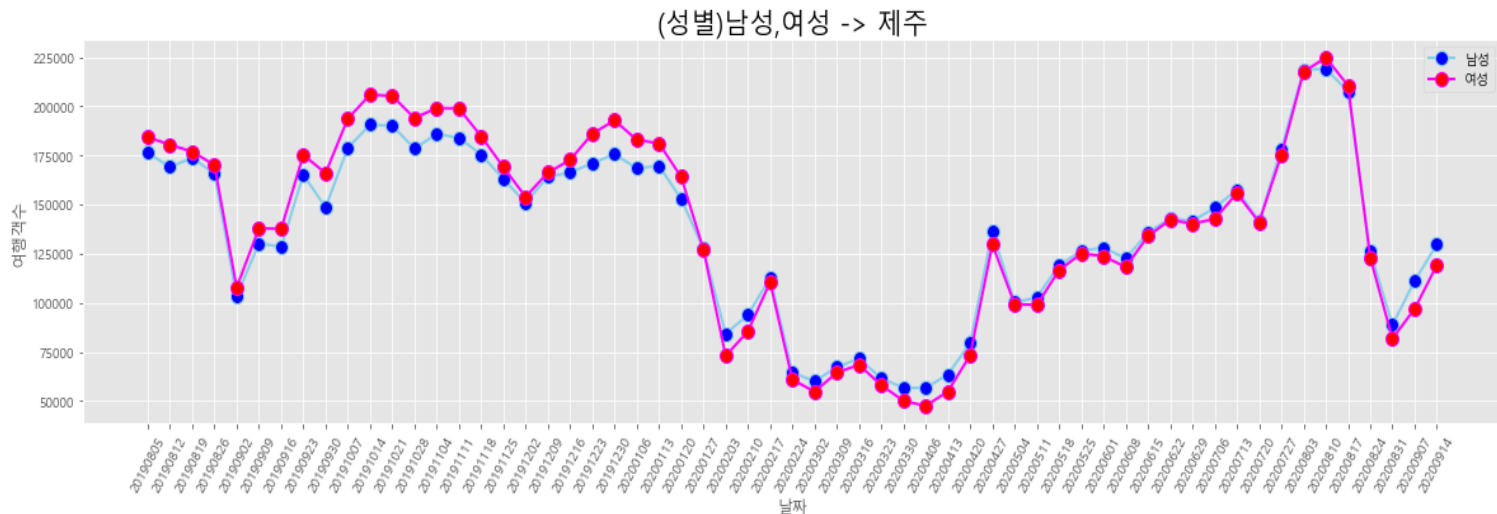


➔ 결과 : 인구가 많은 경기, 서울 여행객이 확실히 많고 다른 지역은 모두 비슷.

성별 여행지(제주도)

- 데이터 추출 [여행지(제주도)]
- 추정여행객수 성별 분리
- 성별 추정여행객수 그래프

| dt | sido_cd | 여행지 | ome_sido_c | 거주지역 | age_cd | 추정 여행객수 |
|----------|---------|---------|------------|------|--------|---------|
| 20190805 | 50 | 제주특별자치도 | nan | nan | nan | 176298 |
| 20190812 | 50 | 제주특별자치도 | nan | nan | nan | 169112 |
| 20190819 | 50 | 제주특별자치도 | nan | nan | nan | 173850 |
| 20190826 | 50 | 제주특별자치도 | nan | nan | nan | 165653 |
| 20190902 | 50 | 제주특별자치도 | nan | nan | nan | 103335 |
| 20190909 | 50 | 제주특별자치도 | nan | nan | nan | 130262 |
| 20190916 | 50 | 제주특별자치도 | nan | nan | nan | 128481 |



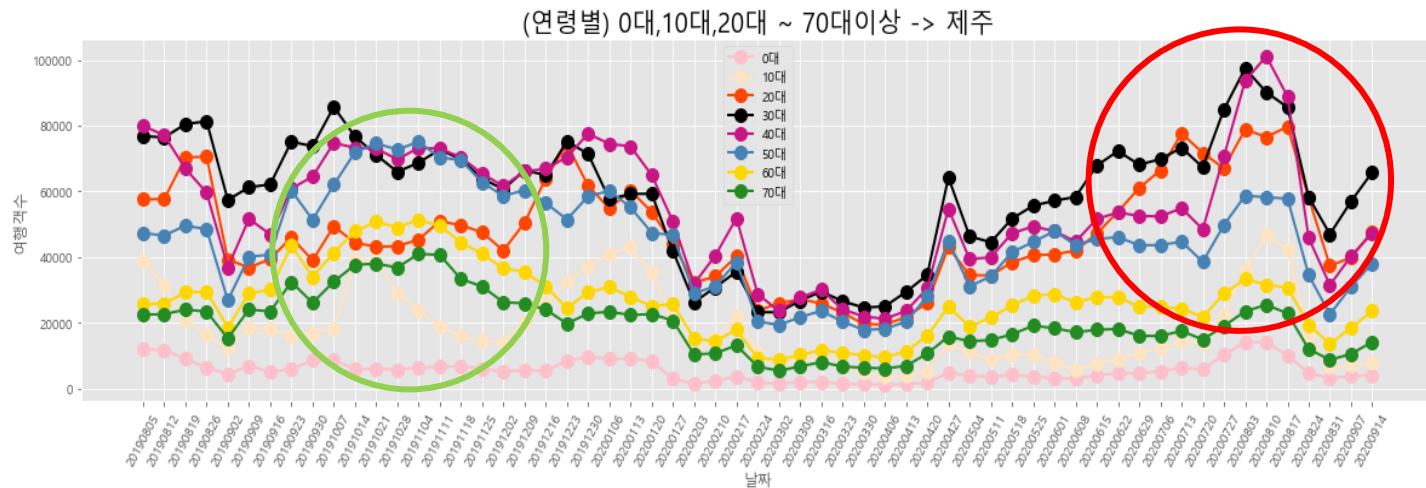
➔ 결과 : 다른 지역에 비해 제주도에서는 여성 여행객이 남성 여행객보다 조금 더 많은 경향

연령별 여행지(제주도)

- 데이터 추출 [여행지(제주도)]
- 추정여행객수 연령별 분리

| dt | sido_cd | 여행지 | ome_sido_c | 거주지역 | sex_cd | 총 여행객수(|
|----------|---------|---------|------------|------|--------|---------|
| 20190805 | 50 | 제주특별자치도 | nan | nan | nan | 12075 |
| 20190812 | 50 | 제주특별자치도 | nan | nan | nan | 11595 |
| 20190819 | 50 | 제주특별자치도 | nan | nan | nan | 9089 |
| 20190826 | 50 | 제주특별자치도 | nan | nan | nan | 6339 |
| 20190902 | 50 | 제주특별자치도 | nan | nan | nan | 4419 |
| 20190909 | 50 | 제주특별자치도 | nan | nan | nan | 6903 |
| 20190916 | 50 | 제주특별자치도 | nan | nan | nan | 5150 |
| 20190923 | 50 | 제주특별자치도 | nan | nan | nan | 5925 |
| 20190930 | 50 | 제주특별자치도 | nan | nan | nan | 8613 |

- 연령별 추정여행객수 그래프 (참고 : 빨간색 선 : 20대)

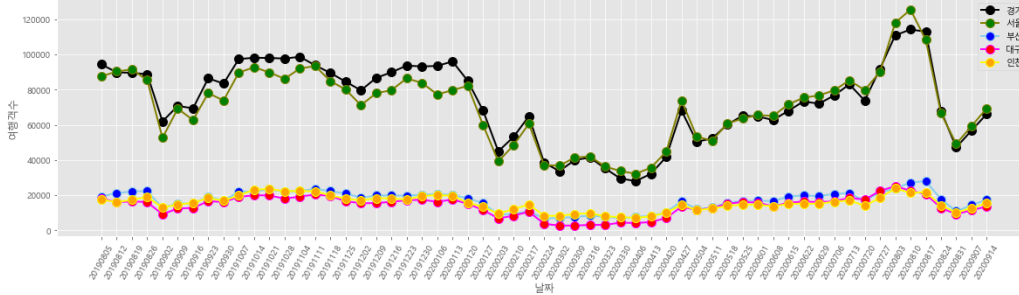


➔ 결과 : 여름 휴가 기간에만 유독 20대의 여행객이 다른 연령대에 비해 높아지는 추세

(빨간색 동그라미 : 여름 휴가철 20대 여행객, 녹색 동그라미 : 다른 기간 20대 여행객)

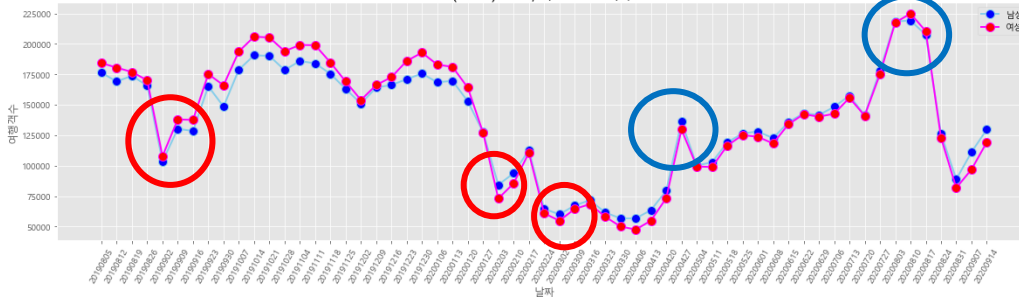
종합 그래프(제주도) 설명

(지역별)경기,서울,부산,대구,인천 -> 제주



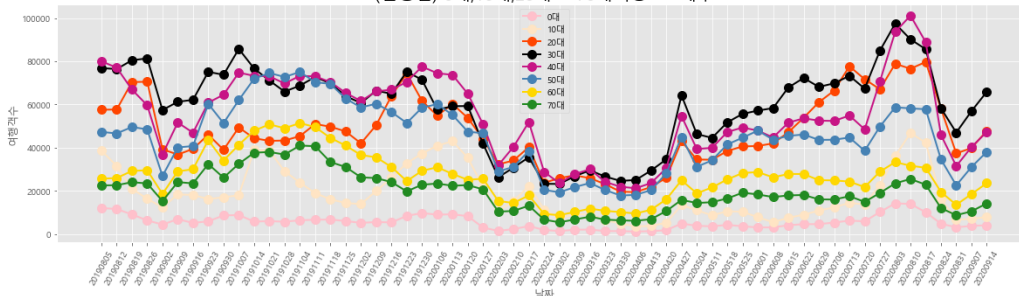
지역별로 분리한 것을 보면
서울,경기가 압도적,
하지만 원래 서울,경기인구가 압도적으로 많기 때문에 상대적으로 비교하기
힘들어 상대적 크기차이를 제거하고 비교하는 정규화 적용(14p).

(성별)남성,여성 -> 제주



설, 추석 명절기간에는 여행객수 감소.
20년 2월말 신천지 집단감염사태일 때
여행객수가 줄어드는 것을 볼 수 있다.
(빨간 동그라미)

(연령별) 0대,10대,20대 ~ 70대이상 -> 제주

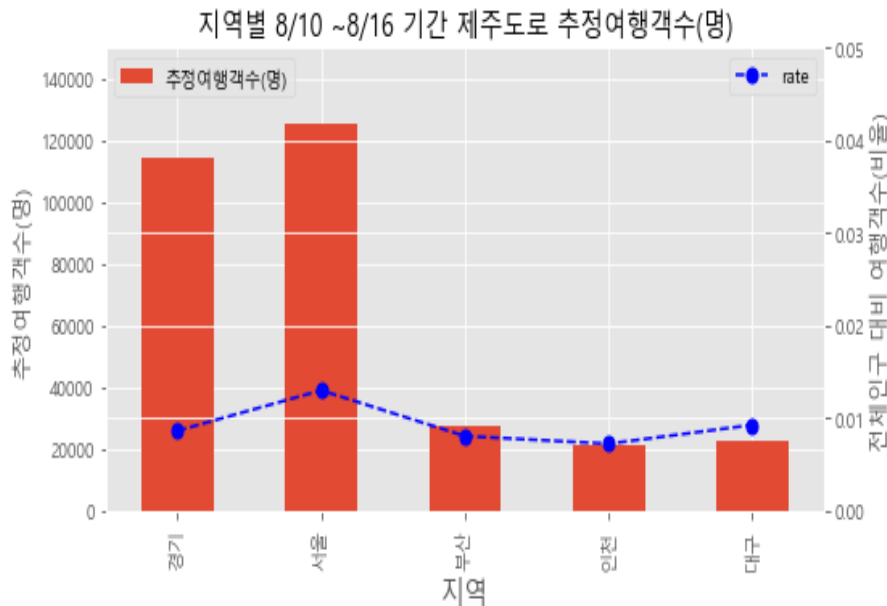


부처님 오신 날과 어린이날이 포함된
4월말 연휴기간과 여름휴가철에는
여행객수가 증가했음을 알 수 있다.
(파란 동그라미)

정규화(제주도)

- 정규화 : 데이터의 상대적인 크기 차이를 제거하는 것,
각 열(변수)에 속하는 데이터 값을 동일한 크기 기준으로 나눈 비율로 나타내는 것
- 남녀비율과 나이대비율은 전체인구비율에 대해 지역별로 큰 차이가 없어 정규화 해도 큰 영향이 없다.
- 따라서 지역별 인구데이터 정규화.
- 정규화 방법 : 해당 열의 최대값(의 절대값)으로 나누는 방법
- 최대값 : 각 시도별 전체인구 =>
- 각 시도별 이동인구에서 그 지역 최대값(전체인구)을 나누면 모든 데이터를 0~1 값으로 상대적인 크기 차이를 제거하고 비교 가능하다.

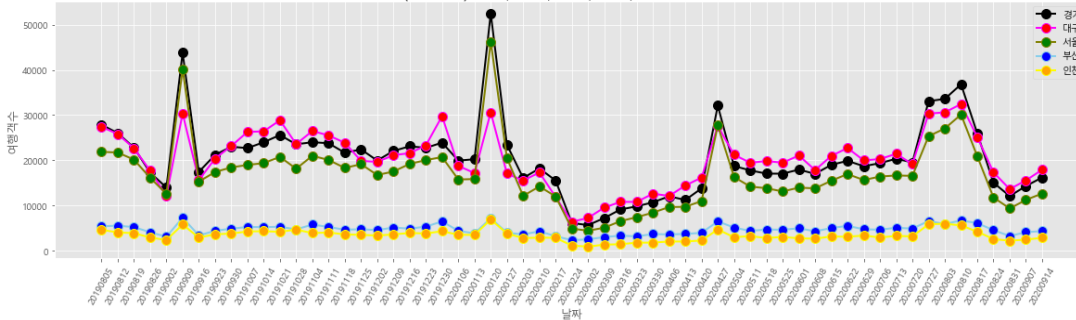
```
#####
# 2019년 시도별 인구
# 경기 : 13238000(명)
# 서울 : 9662000(명)
# 부산 : 3373000(명)
# 인천 : 2944000(명)
# 대구 : 2432000(명)
#####
```



➔ 결과 : 제주도에 있는 실제 여행객 수는 서울, 경기도 압도적으로 많지만 여행비율은 지역별로 비슷비슷한 것을 알 수 있다.

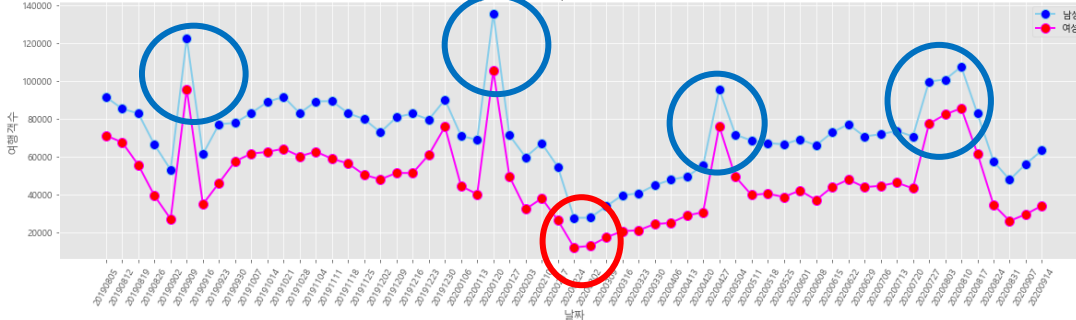
종합 그래프(울산) 설명

(지역별)서울,경기,부산,대구,인천 -> 울산



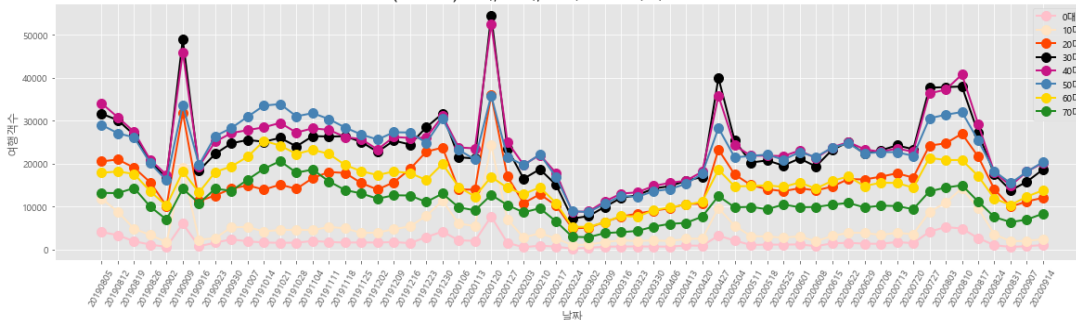
지역별그래프를 보면 울산과 가까운 부산보다 특이하게 대구가 서울,경기만큼 높은 것을 알 수 있음.

(성별)남성,여성 -> 울산



성별그래프는 다른 지역에 비해 남성이 더 높은 것을 알 수 있음
관광지인 제주도와 다르게 울산은 설,추석 명절에 추정여행객 증가,
4월 연휴기간과 여름휴가기간은 제주도와 마찬가지로 증가

(연령별) 0대,10대,20대 ~ 70대이상 -> 울산



연령별그래프는 다른 지역에 비해 40대·50대 비율이 높은 것을 알 수 있다.

정규화(울산)

- 정규화 : 데이터의 상대적인 크기 차이를 제거하는 것,
각 열(변수)에 속하는 데이터 값을 동일한 크기 기준으로 나눈 비율로 나타내는 것

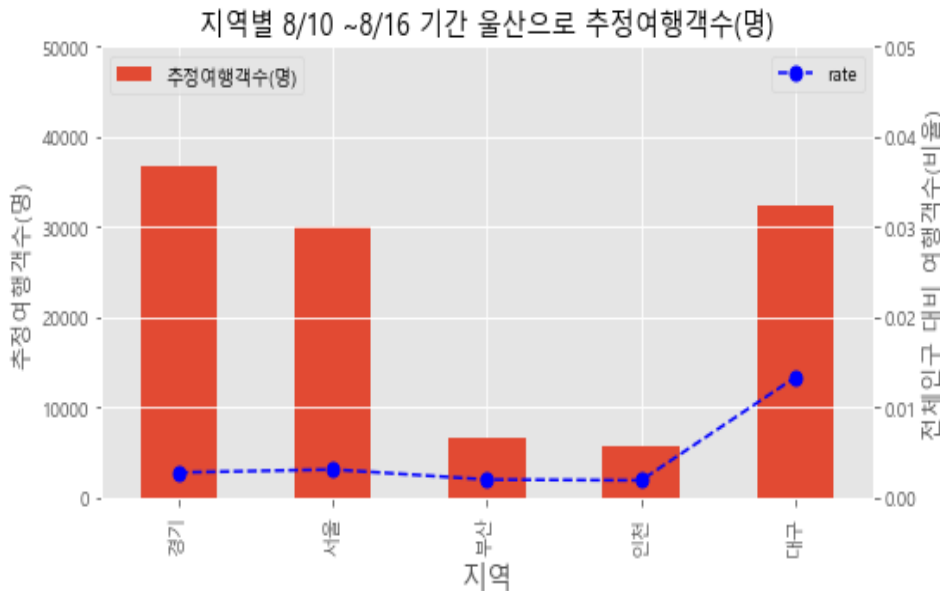
- 정규화 방법 : 해당 열의 최대값(의 절대값)으로 나누는 방법

- 최대값 : 각 시도별 전체인구

=>

- 각 시도별 이동인구에서 그 지역 최대값(전체인구)을 나누면 모든 데이터를 0~1 값으로 상대적인 크기 차이를 제거하고 비교 가능하다.

```
#####
# 2019년 시도별 인구
# 경기 : 13238000(명)
# 서울 : 9662000(명)
# 부산 : 3373000(명)
# 인천 : 2944000(명)
# 대구 : 2432000(명)
#####
```



➔ 결과 :

추정여행객 수는 서울, 경기, 대구가 비슷하게 높지만 정규화한 여행비율은 대구가 가장 높음.

특이점 :

반대로 울산->대구로 가는 비율은 높지 않음.

이유 :

울산은 자동차를 완성하는 현대자동차 공장이 많고 대구에는 자동차 부품 공장이 많아

대구->울산으로 이동률이 높은 것으로 추측

분석결과 및 한계점

< 분석결과에 따른 적용 >

- 대구에서 울산으로 다른 지역에 비해 이동률 높음(특히 40대, 50대 남성)
-> 울산에서 감염병 확산시 대구지역의 40대, 50대 남자 회사원들을 상대로 경고문자를 보내면 효과적인 이동자제 가능.
- 관광도시 성격의 제주도, 부산(여름철), 강원도(겨울철)는 휴가철 20대,30대 비율이 다른 지역에 비해 상대적으로 높음
: 젊은 층의 여행이 많다 추측하여
-> 해당지역 집단감염 발생시 젊은 층에 재난문자 발송,
젊은 층이 많이 이용하는 SNS나 인터넷을 통해 이동자제 권고
- 공업도시 성격의 대구, 경상남도, 울산은 40대, 50대비율이 다른 지역에 비해 상대적으로 높음
: 회사원들의 단기출장 많다 추측
-> 해당지역 집단감염 발생시 40,50대 회사원들에 재난문자 발송,
회사들을 상대로 단기출장 자제를 권고

➔ 특정 기간별 · 성별 · 연령별 이동경로를 예측하여 재난 문자를 이용해 미리 주의를 준다면
코로나 확산 방지 및 코로나 관련 민원 방지 효과를 기대.

분석결과 및 한계점

<분석 도중 나타난 한계점>

- 추정여행객수 데이터는 거주지역, 여행지역 변수가 큰 단위의 17개의 광역자치단체로 나누어져 있어 서로 가까운 거리의 지역을 제외하고, 뚜렷한 관계가 있는 지역은 대구와 울산 정도로 분석 전 기대했던 것 보다 많이 적었음.

➔ 시·군·구 226개의 기초자치단체로 변수가 나누어져 있다면 더 다양하고 정확한 분석가능.

감사합니다.