



재난안전 팀프로젝트

DART-B TEAM PROJECT PRESENTATION

발표 목차

01

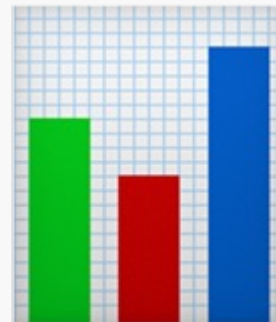
주제 선정



긴급재난문자

02

분석 내용과 인사이트



데이터 전처리
데이터 EDA
데이터 분석

03

최종 목표



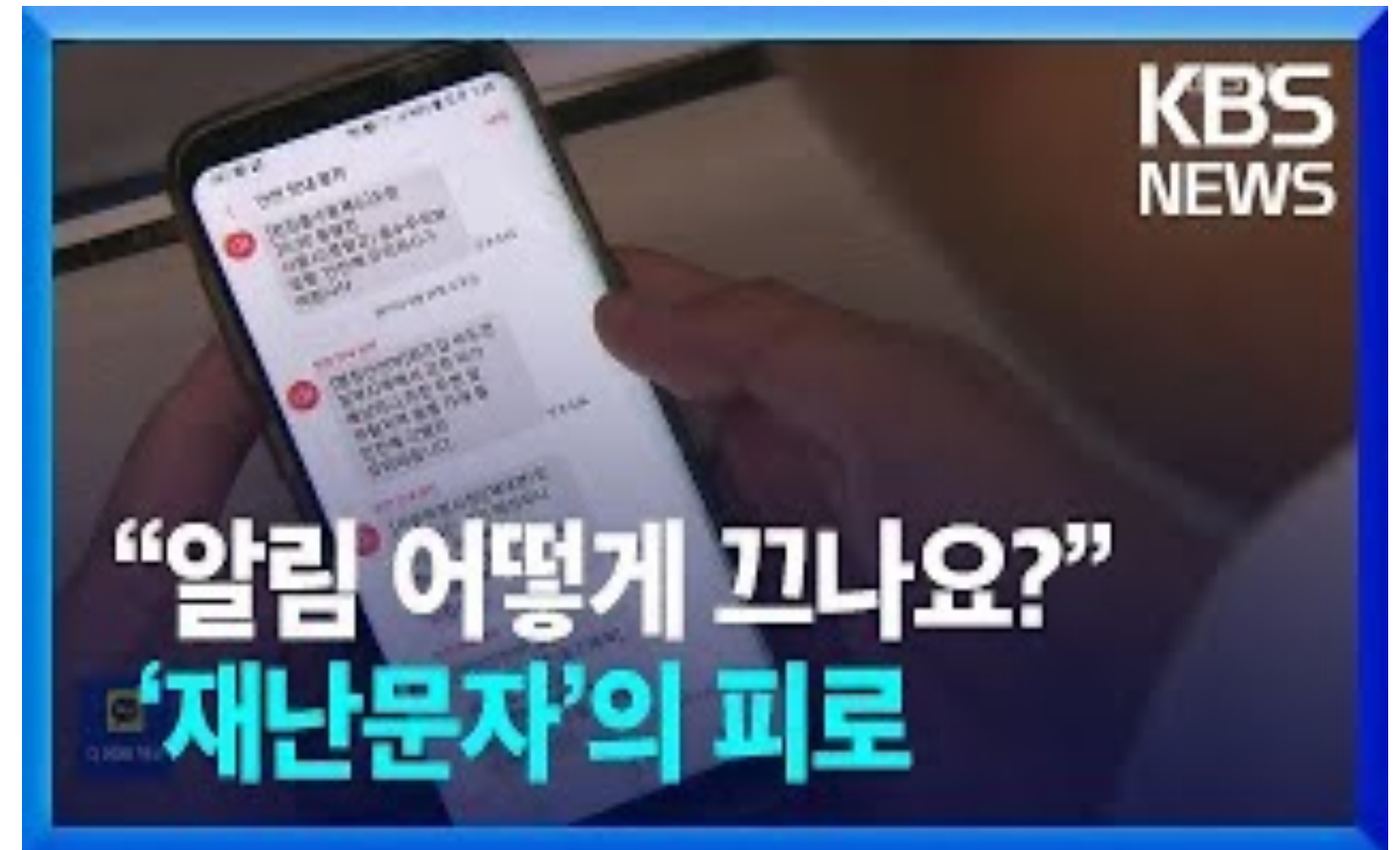
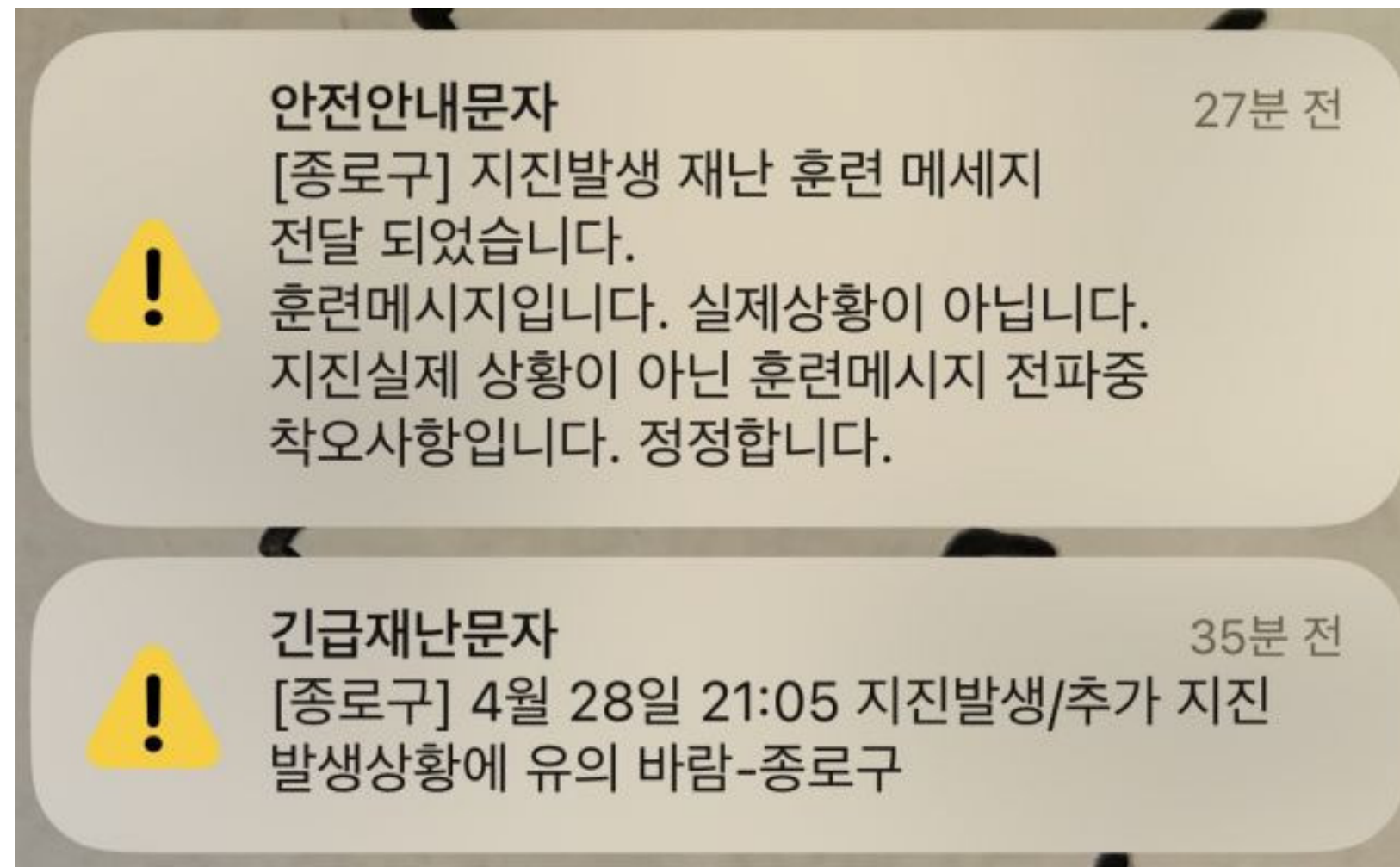
학술제까지의 목표
추후 방향성



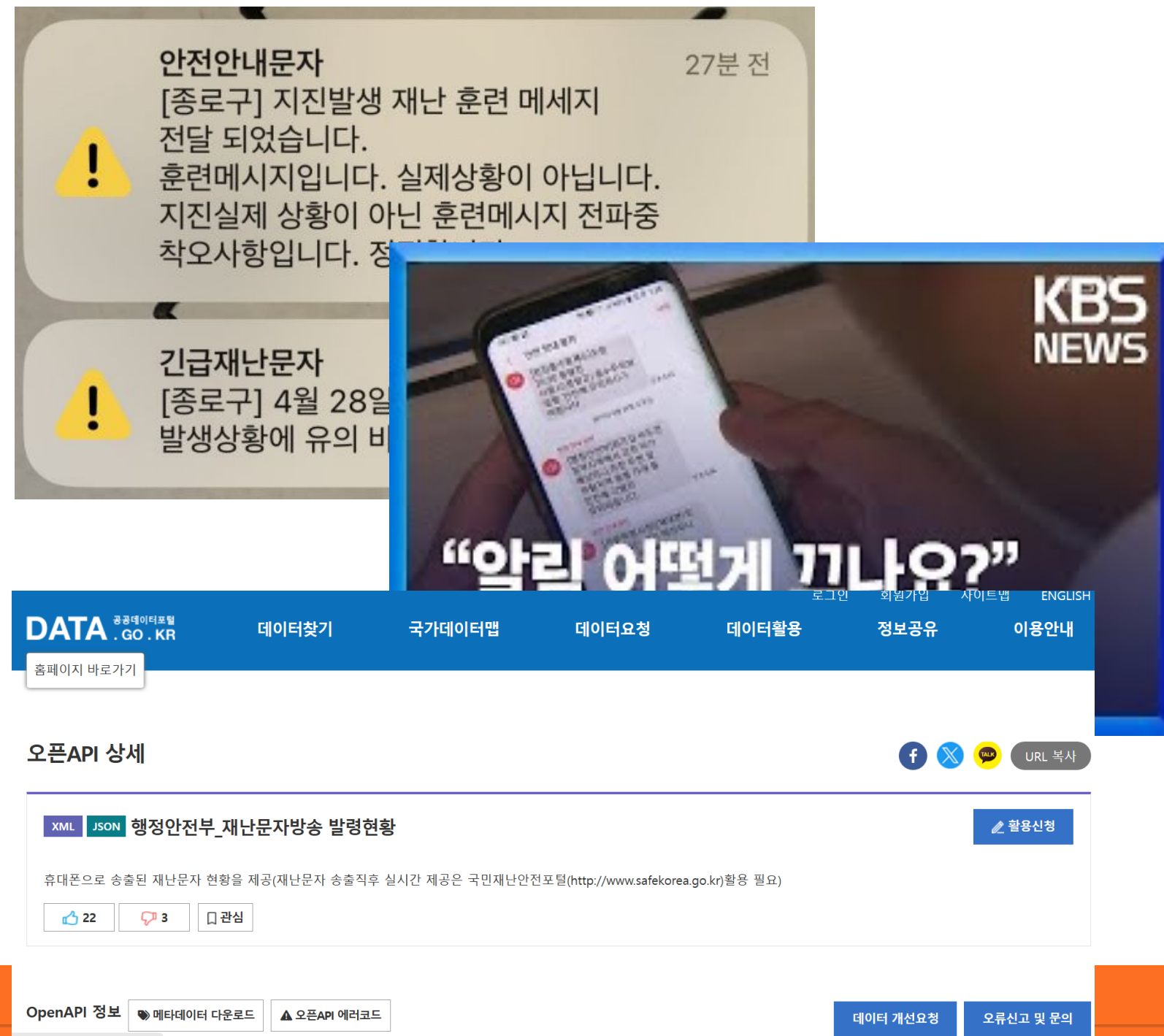
주제 선정

- 긴급 재난 문자

주제 선정



주제 선정



1 재난 문자에 대한 신뢰감 하락

2 과도한 발송으로 인한 재난문자에 대한 피로감

3 접근하기 쉬운 공공데이터의 장점

주제 긴급 재난문자 선별 프로젝트 !!

이용할 데이터

create_date	location_id	location_name	md101_sn	msg	send_platform
2023/11/26 17:15:18	151	서울특별시 서초구	207012	[서울경찰청] 서초구에서 배회중인 이 갑세씨(남,81세)를 찾습니다 -171cm, 남색...	cbs
2023/11/26 16:49:36	50	경기도 포천시	207011	[경기북부경찰청] 포천시에서 실종된 한말녀씨(여,76세)를 찾습니다 -140cm,4...	cbs
2023/11/26 16:07:58	9736,9737,9738,9739,9744,9746,9747,9748	경기도 이천시,경기도 이천시,경기도 이 천시,경기도 이천시,경기도 이천시,경기 도 이천...	207010	[이천시] 오늘 15:30경 모가면 소사리 폐차장에서 화재 사고 발생. 검은 연기 가...	cbs
2023/11/26 13:41:35	137	서울특별시 강남구	207009	[서울경찰청] 강남구에서 실종된 이갑 세씨(남,81세)를 찾습니다 -163cm,남 색패...	cbs
2023/11/26	174	인천광역시 부평구	207008	[인천경찰청] 부평구에서 실종된 조태 용씨(남,70세)를 찾습니다	cbs

행정안전부_재난문자방송 발령현황 API
공공데이터포털 data.go.kr

df.shape

(185354, 6)

이용할 데이터



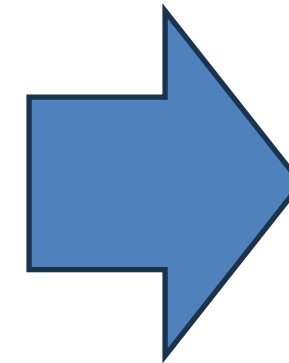
관련 코드

```
df_cols=['create_date', 'location_id', 'location_name', 'md101_sn',
'msg', 'send_platform']

row = []
for i in trange(1000):
    url =
    f'http://apis.data.go.kr/1741000/DisasterMsg3/getDisasterMsg1List
?ServiceKey={encoding}&type=xml&pageNo={i+1}&numOfRows=1000'
    response = requests.get(url)

# xml 내용
content = response.text
xml_obj = bs4.BeautifulSoup(content, 'lxml-xml')
rows = xml_obj.findAll('row')

for node in rows:
    t = {}
    for col in df_cols:
        t[col] = node.find(col).text
    row.append(t)
```



(185354개
의 데이터
수집)



재난안전데이터공유플랫폼



분석 내용과 인사이트

- 데이터 전처리 .

데이터 전처리

1. [](대괄호) 내용 추출

create_date	location_id	location_name	md101_sn	msg	send_platform	기관
2023/11/26 17:15:18	151	서울특별시 서초구	207012	[서울경찰청] 서초구에서 배회중인 아랑씨(남,81세)를 찾습니다 -171cm,남색...	cbs	서울 경찰청
2023/11/26 16:49:36	50	경기도 포천시	207011	[경기북부경찰청] 포천시에서 실종 된 한민씨(남,55세)를 찾습니다 -140cm,4...	cbs	경기 북부 경찰청
2023/11/26 16:07:58	9736,9737,9738,9739,9744,9746,9747,9748	경기도 이천시,경기도 이천시,경기도 이천시,경기도 이천시,경기도 이천시, 경기도 이천...	207010	[이천시] 오늘 15:30경 모가면 소사 리 패환자에서 화재 사고 발생. 검은 연기가...	cbs	이천 시
2023/11/26 13:41:35	137	서울특별시 강남구	207009	[서울경찰청] 강남구에서 실종된 이 강씨(남,81세)를 찾습니다 -163cm,남색패...	cbs	서울 경찰청
2023/11/26 13:14:23	174	인천광역시 부평구	207008	[인천경찰청] 부평구에서 실종된 조 재민씨(남,79세)를 찾습니다 -160cm,50k...	cbs	인천 경찰청
...
2011/11/20 12:10:41	3,4,5,6,7,9	강원도 강릉시,강원도 고성군,강원도 동해시,강원도 삼척시,강원도 속초시, 강원도 양양군	10	[국민안전처]20일 12시 동해중부면 바다 풍랑경보 4호 발령.어선은 출항을 금지 하고 출어선은...	cbs	국민 안전처

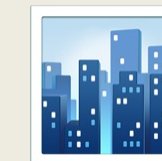
기관이 발송하는 문자 내용에
큰 특징이 있는 경우



경찰청 - 실종, 배회



국민안전처 - 자연재해



시청 - 지역 내 주민 대상

대괄호에 있는 내용을 추출하여 발송 기관 컬럼
생성

데이터 전처리

2. 실종자 문자: ()-소괄호 내용 추출

```

1 def extract_gender_age(text):
2     match = re.search(r'#[([^\]]+)]#', text)
3     return match.group(1) if match else None
4
5 # Apply the function to create a new column
6 df['성별 나이'] = df['msg'].apply(extract_gender_age)
7 df.head()

```

	location_id	location_name	md101_sn	msg	send_platform	기관	성별 나이
create_date							
2023/11/26 17:15:18	151	서울특별시 서초구	207012	[서울경찰청] 서초구에서 실종 증인 이갑 씨(남,81세)를 찾 았습니다. 112로 연락해...	cbs	서울 경찰청	남,81 세
2023/11/26 16:49:36	50	경기도 포천시	207011	[경기북부경찰청] 포천시에서 실종된 한말 씨(여,76세)를 찾 았습니다. 112로 연락해...	cbs	경기 북부 경찰청	여,76 세
2023/11/26 16:07:58	9736,9737,9738,9739,9744,9746,9747,9748	경기도 이천시,경기도 이천시,경기 도 이천시,경기도 이천시,경기도 이 천시,경기도 이천...	207010	[이천시] 오늘 15:30경 모가면 소사리 폐차장에서 화재 사고 발생. 검은 연기가...	cbs	이천 시	None
2023/11/26 13:41:35	137	서울특별시 강남구	207009	[서울경찰청] 강남구에서 실종 된 이갑 씨(남,81세)를 찾습니 다. 112로 연락해...	cbs	서울 경찰청	남,81 세

소괄호에 있는 내용을 추출하여 컬럼 생성



주로 실종자에 대한
성별, 나이 () 안에 있음

데이터 전처리

2. 실종자 문자: CM, KG, 이름, 찾습니다

create_date	location_id	location_name	md101_sn	msg	send_platform	기관	height	weight	name
2023/11/26 17:15:18	151	서울특별시 서초구	207012	[서울경찰청] 서초구에서 배회중인 이갑세씨(남,81세)를 찾습니다 -171cm,남색...	cbs	서울경찰청	171cm	None	[이갑세씨]
2023/11/26 16:49:36	50	경기도 포천시	207011	[경기북부경찰청] 포천시에서 실종된 한말녀씨(여,76세)를 찾습니다 -140cm,4...	cbs	경기북부경찰청	140cm	43kg	[한말녀씨]
2023/11/26	경기도 이천시,경기도 이천시 경기도 이천시			[이천시] 오늘 15:30 경 모가면 소사리 폐		이			

KG, CM, 씨를 기준으로 새로운 컬럼 생성
메시지 내용에 '배회', '실종'가 있을 경우, 실종
데이터로 취급



실종자에 대한 인적사항 추출

데이터 전처리

3. 단체 문자

```
1 def count_commas(location_name):
2     return str(location_name).count(',')
3
4 # 'location_name' 컬럼에 함수 적용하여 개수를 저장하는 '쉼표_개수' 컬럼 생성
5 df['단체 문자'] = df['location_name'].apply(count_commas) + 1
6
7 df
```

create_date	location_id	location_name	md101_sn	msg	send_platform	기관	실종	height	weight	name	단체 문자
2023/11/26 17:15:18	151	서울특별시 서초구	207012	[서울경찰청] 서초구에서 배회중인 이갑세씨(남,81세)를 찾습니다 -171cm,남색...	cbs	서울경찰청	1	171cm	None	이갑세씨]	1
2023/11/26 16:49:36	50	경기도 포천시	207011	[경기북부경찰청] 포천시에서 실종된 한말녀씨(여,76세)를 찾습니다 -140cm,4...	cbs	경기북부경찰청	1	140cm	43kg	한말녀씨]	1
2023/11/26 16:07:58	9736,9737,9738,9739,9744,9746,9747,9748	경기도 이천시,경기도 이천시,경기도 이천시,경기도 이천시,경기도 이천...	207010	[이천시] 오늘 15:30 경 모가면 소사리 폐차장에서 화재 사고 발생. 검은 연기가...	cbs	이천시	0	None	None	one	8



단체 문자 일수록
내용의 중요도가
높을거라 가정

Location_name 컬럼의 내용 속 ,(쉼표)의 개수 + 1 = 발송 대상의 수

데이터 전처리

4. 시도 별

```
1 def location_city(location_name, target_city):
2     return 1 if target_city in str(location_name) else 0
3
4 시도 = ['서울', '부산', '대구', '광주', '인천', '대전', '울산', '세종', '경기도',
5         '강원도', '충청남도', '충청북도', '전라남도', '전라북도', '경상남도', '경상북도', '제주']
6
7 for city in 시도:
8     df[city] = df['location_name'].apply(location_city, target_city=city)
9
10 df
```

1 df[시도]	서울	부산	대구	광주	인천	대전	울산	세종	경기도	강원도	충청남도	충청북도	전라남도	전라북도	경상남도	경상북도	제주
create_date																	
2023/11/26 17:15:18	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2023/11/26 16:49:36	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
2023/11/26 16:07:58	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
2023/11/26 13:41:35	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2023/11/26 13:14:23	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
...
2011/11/20 12:10:41	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
2011/11/18 21:12:20	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
2011/11/18 11:38:36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
2011/11/18 08:20:05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
2011/11/18 07:43:44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

185350 rows × 17 columns

'서울' 컬럼의 총합: 13279
 '부산' 컬럼의 총합: 9238
 '대구' 컬럼의 총합: 3578
 '광주' 컬럼의 총합: 6060
 '인천' 컬럼의 총합: 6094
 '대전' 컬럼의 총합: 2668
 '울산' 컬럼의 총합: 3117
 '세종' 컬럼의 총합: 2194
 '경기도' 컬럼의 총합: 37762
 '강원도' 컬럼의 총합: 13998
 '충청남도' 컬럼의 총합: 16027
 '충청북도' 컬럼의 총합: 11215
 '전라남도' 컬럼의 총합: 21940
 '전라북도' 컬럼의 총합: 15039
 '경상남도' 컬럼의 총합: 20323
 '경상북도' 컬럼의 총합: 16550
 '제주' 컬럼의 총합: 2837



경기도가 월등하게 높고
 수도권보다는 충청도,
 경상도, 전라도의 빈도
 높음

Location_name 컬럼의 내용 속 시별 or 도별을 나타낸
 더미변수



분석 내용과 인사이트

- 데이터 EDA

데이터 EDA

1. 긴급재난문자의 길이

```
1 df['msg_len'] = df.msg.apply(len)
2 df
```

create_date	location_id	location_name	md101_sn	msg	send_platform	기관	성별 나이	msg_len
2023/11/26 17:15:18	151	서울특별시 서초구	207012	[서울경찰청] 서초구에서 배회중인 이갑세씨(남,81 세)를 찾습니다 -171cm,남 색...	cbs	서울 경찰 청	남,81 세	82
2023/11/26 16:49:36	50	경기도 포천시	207011	[경기북부경찰청] 포천시 에서 실종된 한말녀씨 (여,76세)를 찾습니다 -140cm,4...	cbs	경기 북부 경찰 청	여,76 세	89
2023/11/26 16:07:58	9736,9737,9738,9739,9744,9746,9747,9748	경기도 이천시,경기도 이천 시,경기도 이천시,경기도 이 천시,경기도 이천시,경기도 이천...	207010	[이천시] 오늘 15:30경 모 가면 소사리 폐차장에서 화재 사고 발생. 검은 연기 가...	cbs	이천 시	None	86

```
1 df.msg_len.describe()
```

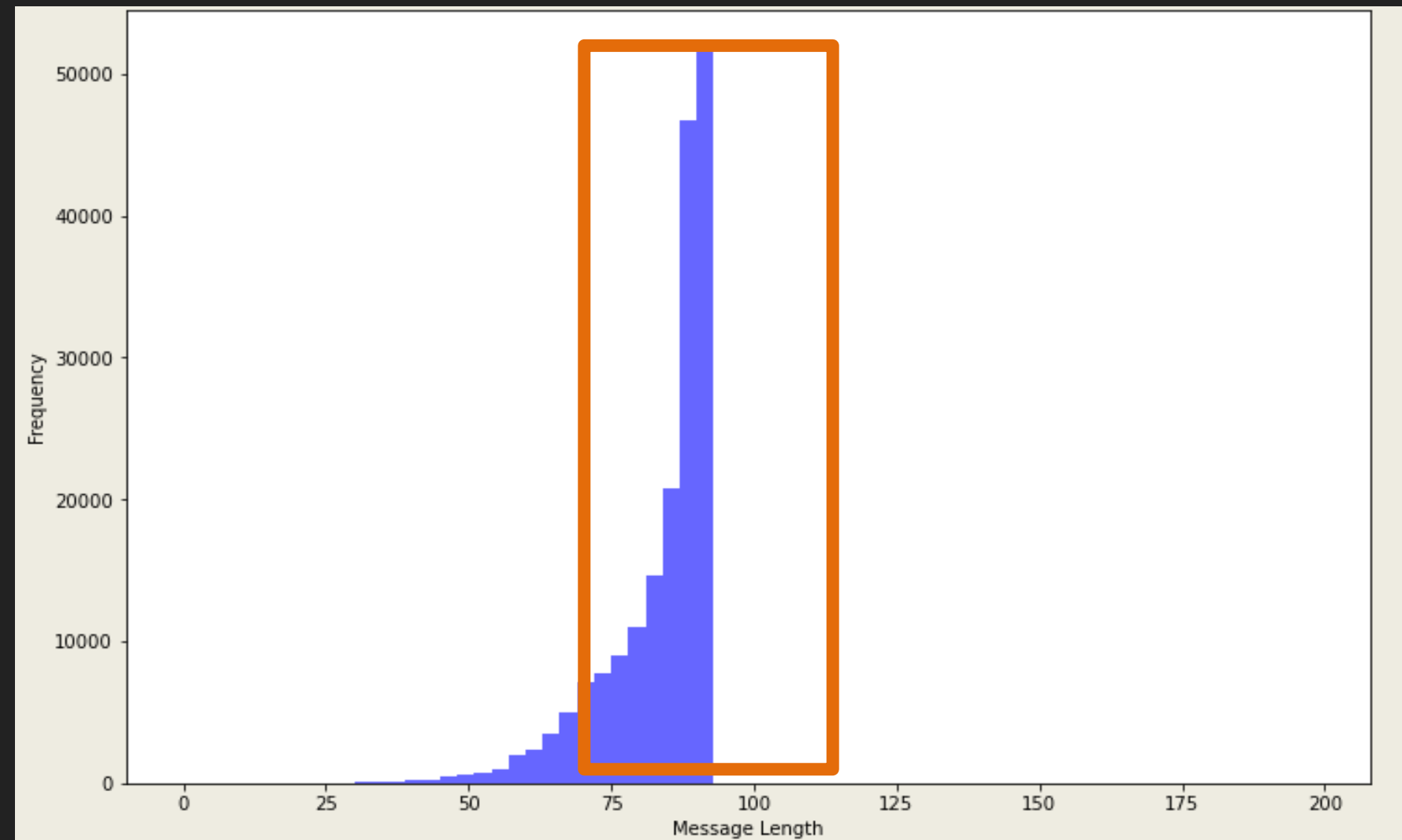
```
count    185350.000000
mean       83.339250
std        12.438484
min         0.000000
25%        79.000000
50%        87.000000
75%        90.000000
max       915.000000
Name: msg_len, dtype: float64
```

짧고 굵을수록 긴급한 재난문자일거라 생각했지만, 대부분 정정이나 오류의 내용이 있었음.

1분위, 중위값, 3분위 수를 보면 주로 80~90에 집중되어 있음

데이터 EDA

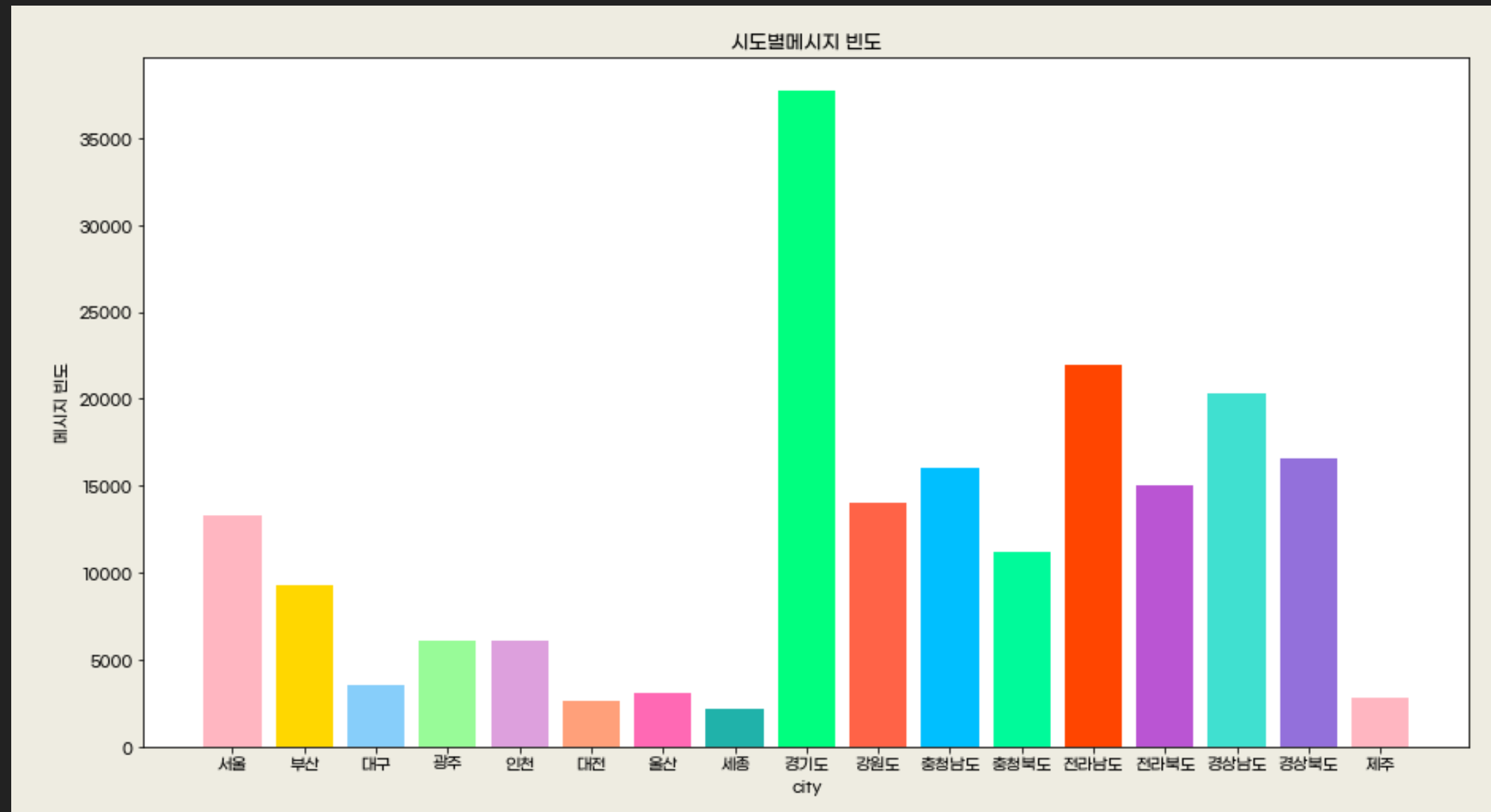
1. 긴급재난문자의 길이



MMS 문자 byte 수 제한 때문에 주로 80~90에 집중되어 있음

데이터 EDA

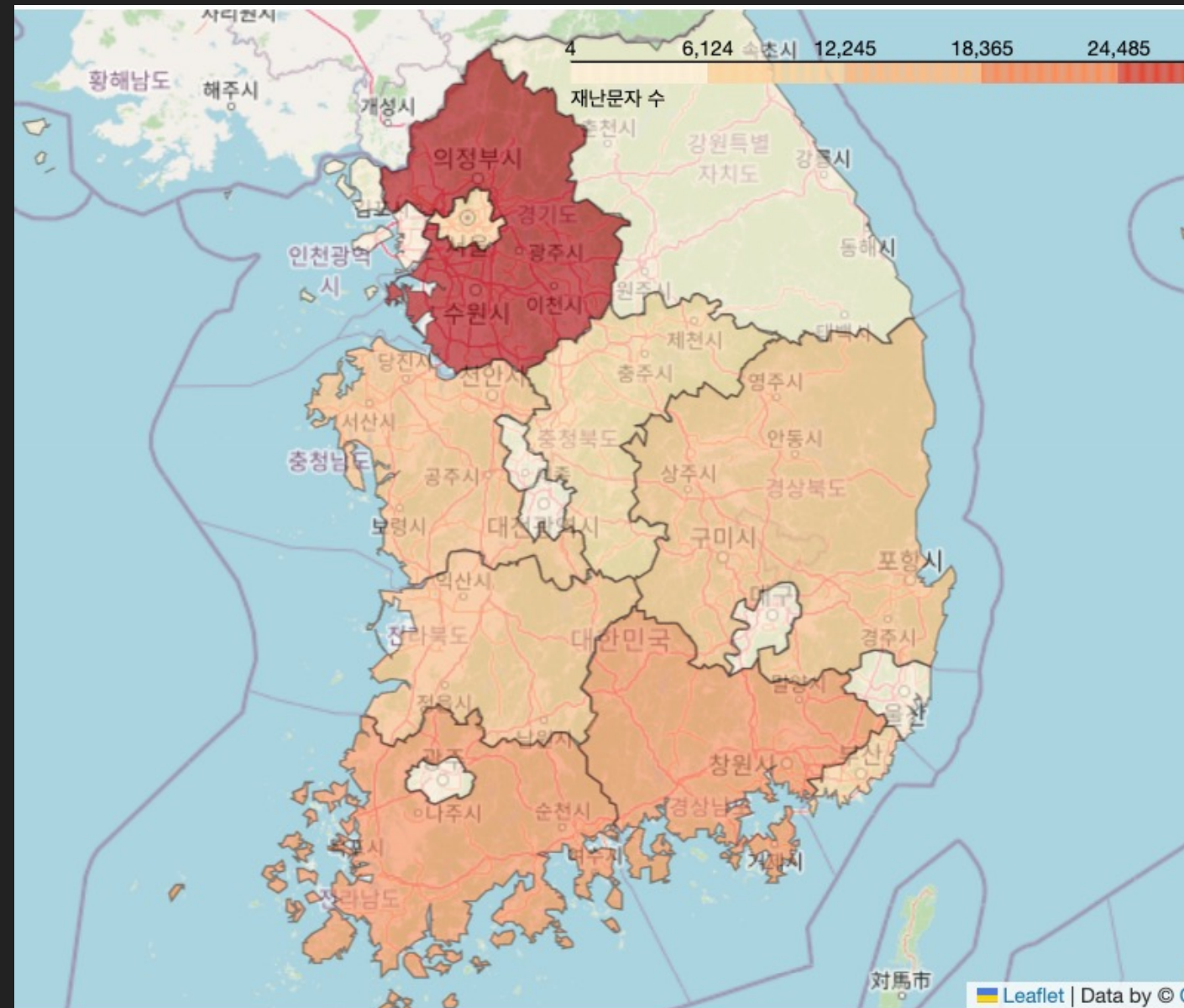
2. 시도 별 긴급재난문자 빈도



시도별 긴급재난 문자 빈도 시각화 -> 경기도 압도적, 세종, 제주, 대전이 적은 편

데이터 EDA

3. 시군구 별 긴급재난문자 빈도 지도에 시각화





분석 내용과 인사이트

- 데이터 분석

KEY WORD 분석

KeyBert model

```
df['keyword'] = df.msg.progress_apply(lambda x: kw_model.extract_keywords(x, keyphrase_ngram_range=(2,4), u
```

271m 30.9s

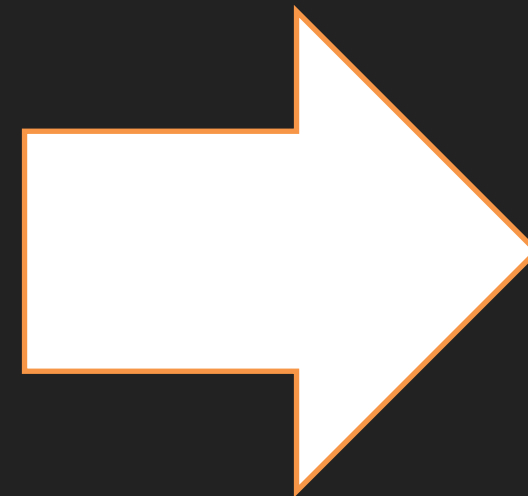
Pyth

17% | 31052/185292 [4:31:30<15:50:52, 2.70it/s]

- 분석의 정교화를 위해 키워드 추출 후 시각화 해볼예정

추후 계획

- 재난 문자 관련 전처리 정교화
- keyword 추출을 위한 다양한 알고리즘 고려
- 키워드 바탕으로 EDA 진행 -> 인사이트 도출
- 지도에 시각화 하는 방법 좀 더 연구
 - 재난문자 카테고리나눠서 시각화



대시보드 제작

- streamlit
- tableau



질문 또는 피드백

감사합니다.

안전 재난 팀프로젝트

팀장 김성연
팀원 문예준
팀원 성유진

