

국방과제 해결을 위한 크롤링 자동화 프로젝트

# 뉴스기사 크롤링을 통한 북한 도발 징후 포착

1등이조 : 김영성, 김지민, 신경민, 현지영

# 목차

01

프로젝트 개요

02

프로젝트 수행

03

시행

04

기대효과

05

개발후기

# 1. 프로젝트 개요

# 프로젝트 기획 배경

신년 초 공개된 당 중앙위원회 제8기 9차 전원회의(2023.26~30) 결정문과 1월 15일 최고인민회의 시정연설 내용을 둘러싸고 국내 정책당국자와 북한 전문가들 간에 뜨거운 논쟁이 벌어지고 있다. 이는 거칠고 과격한 언사뿐만 아니라 북한의 기존 입장을 완전히 뒤엎는 것이기 때문이다. 특히 김일성-김정일 정권의 연장선에서 김정은 정권과 북한 체제를 보아왔던 전문가들은 인식체계의 혼란에 빠져있는 것으로 보인다.

김정은 위원장은 당 전원회의 결정문에서 남북관계에 대해 "적대적인 두 개의 국가관계이며 전쟁 중에 있는 두 개의 교전국가간 관계"로 재규정하고, "일단 전쟁이 우리앞의 현실로 다가온다면 (중략) 전쟁은 대한민국이라는 실체를 끔찍하게 괴멸시키고 끝나게 만들 것"이라고 밝혔다. 최고인민회의 시정연설에서는 "독립적인 사회주의국가로서의 조선민주주의인민공화국의 주권행사령영역을 합법적으로 정확히 규정하기 위한 법률적 대책"을 세우라고 지시했다.

## 우리나라와의 관계를 적대적인 관계를 규정한 김정은

대륙간탄도미사일(ICBM) 발사를 비롯한 북한의 잇따른 군사적 도발에 기민하게 대처하기 위해 한국·미국·일본이 북한 미사일 경보정보(warning data)의 실시간 공유체계를 가동했다. 북한이 쏜 발사체에 대한 정보를 세 나라가 공유해 대응속도를 높일 수 있게 된 것이다.

오로지 북한 미사일이 발사된 직후부터 탄착할 때까지의 정보만 실시간으로 공유하며, 북한의 미사일 발사 징후 등 사전정보나 우리 측 탐지·추적 자산의 종류·위치 등의 정보는 공유하지 않는다.

국방부 관계자는 "경보정보만 공유하기 위해서는 정보를 필터링해야 하고, 제대로 걸러졌는지 검증도 해야 한다"며 "기술적 사안을 고려할 때 수개월 내 진전이 있도록 한다는 것이 목표"라고 말했다.

## 북한의 도발을 감지하고 대응하기 위해 위성, 레이더, 정보공유 시스템을 사용중

# 북한의 도발에 대한

신년 초 공개된 당 중앙위원회 제8기 9차 전원회의(2023.26~30) 결정문과 최고인민회의 시정연설 내용을 놀려싸고 국내 정책당국자와 북한 전문가들 간에 뜨거운 논쟁이 벌어지고 있다. 이는 거칠고 과격한 언사뿐만 아니라 북한의 기존 입장을 완전히 뒤엎는 것이기 때문이다. 특히 김일성-김정일 정권의 연장선에서 김정은 정권은 정권과 북한 체제를 보아왔던 전문가들은 인식체계의 혼란에 빠져있는 것으로 보인다.

북한의 도발에 대한 대처하기 위한 미사일(CB) 발사를 비롯한 북한의 잇따른 군사적 도발에 기민하게 대처하기 위해 미국·일본이 북한 미사일 경보정보(warning data)의 실시간 공유체계를 구축했다. 북한이 손발사체에 대한 정보를 세 나라가 공유해 대응속도를 높일 수 있게 된 것이다.

# 사전 징후나 준비를 파악하기에는 부족하다

김정은 위원장은 당 전원회의 결정문에서 남북관계에 대해 "절대적인 두 나라의 국적 관계이며 전쟁 중에는 구체적인 전국가간 관계"로 재규정하고, "일단 전쟁이 우리앞의 현실로 다가온다면 (중략) 전쟁은 대한민국이라는 실체를 끔찍하게 괴멸시키고 끝나게 만들 것"이라고 밝혔다. 최고인민회의 시정연설에서는 "독립적인 사회주의 국가로서의 조선민주주의인민공화국의 주권행사령영역을 합법적으로 정확히 규정하기 위한 법률을 제정해 시행하고자 했다.

북한의 도발에 대한 대처하기 위한 미사일(CB) 발사를 비롯한 북한의 잇따른 군사적 도발에 기민하게 대처하기 위해 미국·일본이 북한 미사일 경보정보(warning data)의 실시간 공유체계를 구축했다.

국방부는 "전쟁 예상 시나리오에 대해서는 정보를 필터링해야 하고, 제대로 걸러졌는지 검증도 해야 한다. 예상 시나리오가 정해지면 이를 고려할 때 수개월 내 진전이 있도록 한다는 것이 목표"라고 말했다.

# 프로젝트 기획 배경

## 하인리히 법칙

어떤 대형사고가 발생하기 전에  
그와 관련된 수십차례의 경미한  
사고와 수백번의 징후가 반드시  
나타난다는 법칙

### 제 1 절 실험 모델

문제 발생의 징후를 가장 잘 설명하는 통계학적 접근으로 1:29:300의 법칙이라고 불리는 **하인리히의 법칙(Heinrich's law)**이 있다[33]. 이는 어떤 대형사고가 발생하기 전에는 그와 관련된 수십 차례의 경미한 사고와 수백 번의 징후들이 반드시 나타난다는 것을 뜻한다. 북한 도발도 사건이 발생하기 전에 우리가 인지하지 못하거나 간과하는 몇 가지 징후들이 있을 것이고, 도발 발생시점이 다가오면 더 명확하고 뚜렷한 징후가 빈번하게 나타날 수 있다는 것을 전제로 모델을 구상하였다.

출처: 김동훈(2021). BERT 기반 언어 모델링을 활용한 북한 도발 징후 포착 = BERT-Based Catching Signs of Provocation of North Korea Using Language Modeling. 서울대학교 석사학위논문. 27페이지

## 프로젝트 목표

북한 도발 징후 포착

# 타겟

국방 및 안보 연구자

안보 교육 및 홍보기관

# 구성원 및 역할

신경민

- 대남침투 데이터 수집
- 도발 사례(1~14) 한 달 전 북한 뉴스(KCNA) 국가별(미국, 러시아, 중국, 한국) 기사 크롤링
- 도발 없는 시기 6개월 간 북한 뉴스(KCNA) 국가별(미국, 러시아, 중국, 한국) 기사 크롤링
- 남북한 기사 워드클라우드 생성, 도발 유형 변화 및 정권별 특성 분석
- DB 저장 및 생성, 메인페이지 대시보드 구현, 도발 유형 비율 도넛차트 생성
- 서브페이지(attack, nuclear) 구현, 14건의 도발 사건 지도 시각화
- 정권별 도발 차트 구현(스택형 그래프), 연도별 도발 누적 비율 라인차트 구현

김영성

- 오물풍선 데이터 수집
- 도발사례(1~14) 한 달 전 남한 뉴스(국방일보) 국가별(미국, 러시아, 중국, 한국) 기사 크롤링
- 도발 없는 시기 6개월 간 남한 뉴스(국방일보) 국가별(미국, 러시아, 중국, 한국) 기사 크롤링
- 오물풍선 및 대남침투 데이터 전처리, 남북한 기사 전처리
- 미사일 도발 시기별 분석, 도발 유형 변화 및 정권별 특성 분석
- 도발사건(1~14건) 정보 정리, 발표 ppt 준비
- 데이터 명세서 작성

# 구성원 및 역할

김지민

- 미사일 데이터 수집
- 도발사례(1~14) 한 달 전 북한 뉴스(KCNA) 전체 기사 크롤링
- 도발 없는 시기 6개월 간 북한 뉴스(KCNA) 전체 기사 크롤링
- 미사일 및 핵실험 데이터 전처리. 남북한 기사 전처리
- 남북한 기사 워드클라우드 생성. 미사일 도발 시기별 분석
- 도발 시기별 뉴스 기사량 분석 및 시각화. 도발 유형 변화 및 정권별 특성 분석
- 데이터 명세서 작성

현지영

- 핵실험 데이터 수집
- 도발사례(1~14) 한 달 전 남한 뉴스(국방일보) 전체 기사 크롤링
- 도발 없는 시기 6개월 간 남한 뉴스(국방일보) 전체 기사 크롤링
- 도발 시기별 뉴스 기사량 분석 및 시각화. 도발 유형 변화 및 정권별 특성 분석
- DB 저장 및 생성. 메인 페이지 대시보드 구현. 발표 ppt 준비
- 서브페이지(missile, balloon) 구현. 도발 유형별 지도 시각화
- 도발 전/평시 뉴스 비교 시각화. 정권별 도발 횟수 barplot 시각화

## 2. 프로젝트 수행

# 대표 도발 사건들과 선정 이유

시켜야 한다. 연구에서는 표 2-2에서 보는 바와 같이 주요 북한 도발 사례 총 11건을 선정하였다. 선정 기준은 원활한 데이터 수집을 위해 인터넷 포털 사이트와 온라인 뉴스를 이용하기 시작한 1990년대 후반 이후의 도발 사례, 우리 정부와 군의 대응이 있었으며 언론에서 집중 보도한 도발 사례를 선정하였고, 전략·기술적 도발을 모두 포함하였다. 선정한 주

## 참고 논문의 대표 도발 사건 선정 기준:

우리 정부와 군의 대응이 있었던 도발 사건 /  
언론에서 집중 보도한 도발 사건

구분	도발 사례	일시
1	제 2 연평해전	2002.06.29
2	북한 1차 핵실험	2006.10.09
3	북한 2차 핵실험	2009.05.25
4	천안함 피격	2010.03.26
5	연평도 포격	2010.11.23
6	북한 3차 핵실험	2013.02.12
7	DMZ 목함지뢰 매설	2015.08.04
8	북한 4차 핵실험	2016.01.06
9	북한 5차 핵실험	2016.09.09
10	북한 6차 핵실험	2017.09.03

# 대표 도발 사건들과 선정 이유

## 2014년 들어 미사일 14차례 97발 쏘아댄 북한

관련된 논란에 대한 분석을 시도한다. 하노이 북미정상회담이 결렬된 후 북한은 2019년 5월 4일, 522일 만에 미사일 도발을 재개하였다. 이후 북한은 2020년 8월까지 자위적 국방력 강화라는 목표 아래,

## 北, 사상 첫 NLL 이남으로 탄도미사일 발사…軍 "단호히 대응"(종합)

'자동차 박살에 지붕 파손'…北 오물풍선으로 피해액 1억원 넘어

구분	도발 사례	일시
11	동해안 단거리 미사일 발사	2014.03.28
12	신형 단거리 탄도 미사일 발사	2019.05.09
13	NLL 이남 대륙간 탄도 미사일 발사	2022.11.02
14	오물풍선	2024.05.29
15	도발 없던 시기	2011.03.01 ~ 2011.09.30

# 국가 키워드 선정 과정

“**北, 中과 학술교류로 미사일 기술 공공연히 수입**”…**日매체 분석**

미 전문가 “**미국과 관계 정상화 포기한 북한, 전쟁하기로 결정한 듯**”

4월 26일 발레리 게라시모프 러시아군 총참모장이 “북한 군인과 장교들은 우크라이나 습격을 격퇴하는 동안 러시아군과 어깨를 나란히 해 전투 임무를 수행했다”며 **북한군 러시아 파병을 공식 확인한 것이다**. 북한 노동신문과 조선중앙통신, 조선중앙TV 등 북한 매체들도 일제히 북한군 파병 사실을 확인했다.

북한과 러시아가 북한군 파병을 인정한 의도에 대해 많은 관측이 제기되는 상황에서, **러.북 간의 포괄적 전략 동반자 관계 조약**, 북한군 포로 송환 그리고 푸틴의 공동 전범 김정은 국제형사재판소 (ICC) 제소 관련해 간략히 살펴본다.

• 북한과 밀접한 관련이 있는 미국, 중국, 러시아

는데, 이런 상황에 놓이게 되면 외부위협수준을 완화하는 것은 정권의 생존과 직결되는 문제가 된다. 따라서 **북한은 한국·미국·중국**을 비롯한 주변국들에게 다양한 방법을 통해 외부위협 수준을 완화할 것을 압박한다. 하지만 북한의 이러한 압박에도 주변국들이 ‘외

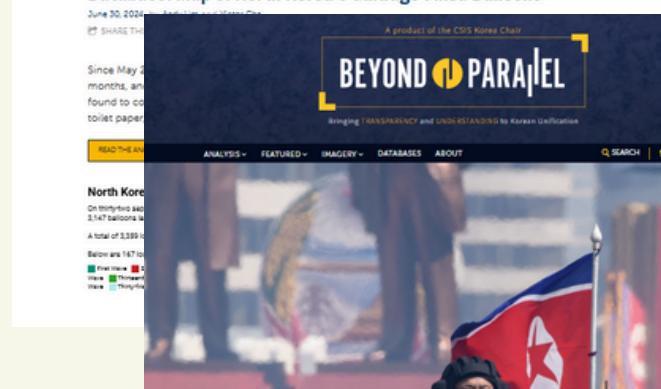
- **북한 도발에 영향을 주는 주요주변국**

→ 미국, 중국, 러시아, 북한, 한국 검색어를 입력한 뉴스 기사 크롤링 추가 진행

# 북한 도발 관련 데이터



Database: Map of North Korea's Garbage-Filled Balloons



Database: North Korean Provocations

## 북괴의 대남도발 주요 일지

이 자료는 북괴의 끊임없는 적화통일야욕과 수단 방법을 가리지 않는 대리 공작에 대한 안보의식을 강화하기 위해, 해방 이후부터 전쟁기간을 제외한 1980년까지 일어났던, 북괴의 각종 도발사건 중 주요사항을 선별하여 일자로 엮은 것이다.

1945. 9. 5. 조선정판사 위조지폐 사건

이완수·권오직·박락종 등이 경제적인 혼란을 일으키기 위하여 5회에 걸쳐 1,200만원의 위조지폐를 발행하였다.

학병동맹사건

신탁동치 반대시위를 하면 학생들에게 좌익학병동맹이 총기를 난사하여 유혈충돌을 일으켰다.

정계요인 암살 및 경찰서 방화기도 사건

한민당의 수석총무인 송진우를 암살하려다 실패하고 구속된, 고사대의 거짓말을 구출하기 위하여 공산당원들이 둘대문서

News Experts Events Analysis Blog Countries and Areas Education Center

Nuclear Biological AI & Tech Get Involved Impact About

DATABASE — Nov 12, 2024  
The CNS North Korea Missile Test Database  
Part of The CNS Missile and SLV Launch Databases

## • 데이터 수집 공식 문서, 연구 사이트

출처: CNS, The CNS North Korea Missile Test Database, NTI. <https://www.nti.org/analysis/articles/cns-north-korea-missile-test-database/>

출처: Andy Lim and Victor Cha, Database: Map of North Korea's Garbage-Filled Balloons, Beyond Parallel. <https://beyondparallel.csis.org/map-of-north-koreas-garbage-filled-balloons/>

출처: CSIS, Database: North Korean Provocations, Beyond Parallel. <https://beyondparallel.csis.org/database-north-korean-provocations/>

출처: 國防部 戰史編纂委員會 編. 北괴의 대남도발 주요일지. 국회도서관. 1986.09.

출처: 위키피디아, List of nuclear weapons tests of North Korea languages. [https://en.wikipedia.org/wiki/List\\_of\\_nuclear\\_weapons\\_tests\\_of\\_North\\_Korea](https://en.wikipedia.org/wiki/List_of_nuclear_weapons_tests_of_North_Korea)

A	B	C	D	E	F	G	H	I
Wave number	Location	LAT	LONG	Province	Korean Name	Date	Time	Source
First Wave	Duma-myeon, G	36.2542683	127.2608089	Chungcheongna	충남 계룡시 두마면	May 29, 2024	오전 3:05	Donga Ilbo
First Wave	Tanhyeon-myeon	37.80919	126.71015	Gyeonggi Provin	파주시 탄현면	May 29, 2024	오전 5:15	Newsis
First Wave	Wicheon-myeon	35.75086	127.83267	Gyeongsangnan	경남 거창군 위천면	May 29, 2024	오전 5:30	Donga Ilbo
First Wave	Mujueup, Muju- <i>g</i>	35.93172	127.71118	Jeollabuk Provin	전북자치도 무주군	May 29, 2024	오전 5:45	<a href="https://www.news1.kr">https://www.news1.kr</a>
First Wave	Seongnam-si, Si	37.4503386	127.1462933	Gyeonggi Provin	성남시 수정구	May 29, 2024	오전 5:49	<a href="https://mobile.news1.kr">https://mobile.news1.kr</a>
First Wave	Sahudo Reservc	37.10531	127.08763	Gyeonggi Provin	경기도 평택시 진	May 29, 2024	오전 7:03	<a href="https://mobile.news1.kr">https://mobile.news1.kr</a>
First Wave	Seoreung-ri, Dai	36.4485683	128.3559773	Gyeongsangbuk	경북 의성군 대	May 29, 2024	오전 7:30	<a href="https://www.news1.kr">https://www.news1.kr</a>

날짜	사건명	사건장소	위도	경도	내용
1945. 9. 5.	조선정판사, 위조지폐 사건				이만출, 권오직, 박락종 등이 경제적인 혼란을 일으키기 위해 5회에 걸쳐
1946. 1. 18.	학병동맹사건				신탁동치 반대시위를 하던 학생들에게 좌익학병동맹이
1946. 1. 29.	정계요인 암살 및 경찰서 방화기도 사건				한민당의 주석홍무의 송진우를 암살하려다 실패하고 구속된 공산당원 김길중을 구출하기 위
1946. 10. 1.	대구폭동 사건	대구	35.87156652	128.601259	공산당의 선동하에 대구를 중심으로 한 경북, 우도도지에서 폭동이 발생하여, 피살자 83명이
1946. 10. 16.	조병옥 경무부장 암살미수 사건				공산당 산하인 민청 소속의 유옥렬, 이충락 등이 조병옥
1946. 11. 13.	장택상 경무총감 피격 사건				민중동맹원 김용화, 최인석 등이 수도경구 경찰정장 및 제1경무총장 장택
1946. 12. 15.	대남 송전 중단 사건				북괴가 남한지역 송전선의 대부분을 북한이 공급하던 것을 이때부터 송전중단(71천kW). 전력분으로 1948
1946. 5. 8.	연백평야 통구(通水)단절 사건				북괴는 38도선 이남에 있는 농민들에게 공급 중이던, 38도선 이북에 위치한 황해도 연백수리조합 구역자수지
1946. 7. 12.	국대안 반대 사건				공산당은 국립서울대학교의 설치를 반대하는 여론을 이
1946. 8. 15.	보성 적색테리 사건	전남 보성	34.77021733	127.0790385	전남 보성에서 2,000여 명의 좌익 폭동분자가 경찰 1명
1946. 8. 23.	이스마리 헬기 격추 사건				고사단의 기관총 기여로 드디어 이스마리 헬기를 아사시키기 위해선 헬기

1	9-4월-84	23-12월-16	Scud-B	SRBM	Tonghae Satellite Launching Gro
2	9-4월-84	23-12월-16	Scud-B	SRBM	Tonghae Satellite Launching Gro

Date	Type	Event	Description	Resources	Latitude	longitude
2006-10-0	Nuclear Provocation	Nuclea 풍계리 핵실험장에서 시행된 북한의 1차 핵실험	<a href="http://georgewbush-whit">http://georgewbush-whit</a>	41.28505	129.1084	
2009-05-0	Nuclear Provocation	Nuclea 풍계리 핵실험장에서 시행된 북한의 2차 핵실험	<a href="https://www.whitehouse">https://www.whitehouse</a>	41.29142	129.0816	
2013-02-0	Nuclear Provocation	Nuclea 풍계리 핵실험장에서 시행된 북한의 3차 핵실험	<a href="http://www.wsj.com/artic">http://www.wsj.com/artic</a>	41.26809	129.0807	
2016-01-0	Nuclear Provocation	Nuclea 풍계리 핵실험장에서 시행된 북한의 4차 핵실험	<a href="http://www.reuters.com/">http://www.reuters.com/</a>	41.309	129.0339	
2016-09-0	Nuclear Provocation	Nuclea 풍계리 핵실험장에서 시행된 북한의 5차 핵실험	<a href="http://www.cnn.com/201">http://www.cnn.com/201</a>	41.298	129.015	
2017-09-0	Nuclear Provocation	Nuclea 풍계리 핵실험장에서 시행된 북한의 6차 핵실험	<a href="http://www.cnn.com/201">http://www.cnn.com/201</a>	41.343	129.036	

## • 수집한 엑셀 파일

# 뉴스 데이터 - 남한(국방일보)

The screenshot shows the homepage of the Kookbang news website. At the top, there is a navigation bar with categories: 국방, 기획, 국내외, 포토라운지, 오피니언, 엔터스포츠, 라이브리리, 무기백과, and Shorts. Below the navigation bar is a search bar with a magnifying glass icon. The main content area is titled '날짜별 기사보기' (Article by Date) and shows a date range from 2025.05.12 to 2025.05.13. There are five tabs at the top of the list: All, 국방, 기획, 국내외, and 오피니언. The 'All' tab is selected. Below the tabs, there is a sub-tab for '엔터스포츠'. The list of news articles includes:

- 국가유공자·제대군인 취·창업 향한 발길 이어졌다** (2025.05.09 17:14)
- 육군9보병사단 황금박쥐여단, 모범 간부에 표창·노트북 전달** (2025.05.11 13:14)
- 국방정신전력원 '프라미스 작전' 참가 CCT 초청 특별초빙교육** (2025.05.11 14:48)
- 고온·사막 극한 환경서 능숙한 공중동작·정확한 착지** (2025.05.11 14:48)
- 최북단 관측소 '금강산전망대' 견학 재개**

- 기간 전체 뉴스 url 수집 화면

```

start_date = '2010-02-05'
end_date = '2010-03-05'
date_range = pd.date_range(start=start_date, end=end_date, freq='B')
date_list = date_range.strftime('%Y.%m.%d').tolist()
print(date_list)

['2010.02.05', '2010.02.08', '2010.02.09', '2010.02.10', '2010.02.11', '2010.02.12',
 '2010.02.17', '2010.02.18', '2010.02.19', '2010.02.22', '2010.02.23', '2010.02.24', '2010.02.25',
 '2010.03.02', '2010.03.03', '2010.03.04', '2010.03.05']

```

- 날짜 설정 코드

```

name_list, year_list, url_list = [], [], []
for date in date_list:
    date_input = driver.find_element(By.ID, 'news_s_date')
    driver.execute_script("document.getElementById('news_s_date').value = ''")
    time.sleep(2)
    date_input.send_keys(date)
    time.sleep(2)

    try:
        driver.find_element(By.XPATH, '//*[@id="mainForm"]/div/button').click()
    except Exception as e:
        print("연결 실패:", e)
        time.sleep(3)

    html = driver.page_source
    soup = BeautifulSoup(html, 'html.parser')

    for i in [1, 2, 3, 5]:
        article = soup.find_all('div', {'id': f'TODAY_ATCE_CTGR_00{i}0000000'})
        for art in article:
            cart = art.find('h3')
            articles = art.find_all('ul', {'class': 'basic_list'})
            for ul in articles:
                titles = ul.find_all('h3', {'class': 'eps1'})
                date_tags = ul.find_all('span')
                a_tags = ul.find_all('a')
                for pn in titles:
                    name_list.append(pn.text.strip() + '<' + cart.text.strip() + '>')
                for dat in date_tags:
                    year_list.append(dat.text.strip()[:13])
                for a in a_tags:
                    href = a.get('href')
                    if href and href.startswith('/newsWeb/'):
                        url_list.append('https://kookbang.dema.mil.kr' + href)
                    else:
                        print(f"잘못된 href: {href}")

```

- url 수집 코드

# 뉴스 데이터 - 남한(국방일보)


≡
[국방](#)
[기획](#)
[국내·외](#)
[포토라운지](#)
[오피니언](#)
[엔터·스포츠](#)
[라이브러리](#)
[무기백과](#)
[Shorts](#)



지난 26일 대전시 유성구 국립대전현충원 현충광장에서 열린 제1회 순직의무군경의 날 기념식에서 한덕수 국무총리와 강정애 국가보훈부장관 등 주요 참석자들이 국기에 대한 경례를 하고 있다. 이경원 기자

꽃다운 나이 국가의 부름을 받고 병역의무를 이행하다가 안타깝게 사망한 순직의무군경을 기억하고 추모하기 위한 기념식이 정부기념행사로 올해 처음 거행됐다.

국가보훈부(보훈부)는 지난 26일 대전시 유성구 국립대전현충원 현충광장에서 제1회 순직의무군경의 날 기념식을 엄수했다. ►관련 기사 3면

순직의무군경은 나라를 지키기 위해 의무복무를 하는 과정에서 직무 수행이나 교육훈련 중 순직한 사람을 뜻한다. 현역병과 상근예비역, 의무경찰, 의무소방, 경비교도대 등이 포함된다. 보훈부에 따르면 순직의무군경 수는 1만 6419명으로 군 1만6355명, 경찰 59명, 교정 4명, 소방 1명이다.

정전(停戰)이라는 특수한 상황에서 젊은 나이에 국방의 의무를 수행하다가 사망한 의무군경에 대한 기념일 지정이 필요하다는 의견은 줄곧 제기됐다. 유가족 측에서도 다른 유공자 등과 형평성 차원에서 전체 순직군경을 추모할 수 있는 순직의무군경의 날 제정이 필요하다는 입장을 계속 밝혀 왔다.

- **기사 내용 수집 화면**

```

bad_keywords = ['페이스북', '트위터', '밴드', '카카오톡', 'url복사', '기자', '글=', '사진']
con_list = []

for i, url in enumerate(url_list):
    try:
        print(f"({i}/{len(url_list)}) 크롤링 중")
        soup = BeautifulSoup(urlopen(url), 'html.parser')
        contents = soup.find('div', {'id': 'article_body_view'})

        if not contents:
            con_list.append('본문 없음')
            continue

        for br in contents.find_all('br'):
            br.replace_with('\n')

        full_text = contents.get_text(separator='\n').strip()
        lines = [line.strip() for line in full_text.split('\n') if line.strip()]
        clean_lines = [line for line in lines if not any(bad in line for bad in bad_keywords)]

        final_text = re.sub(r'\s+', ' ', ''.join(clean_lines)).strip()
        con_list.append(final_text)
        print("완료")

    except Exception as e:
        print(f"에러 발생: {e}")
        con_list.append('에러 발생')

```

- **url 접속 후 기사 내용 수집 코드**

# 뉴스 데이터 - 북한(KCNA)

**Search Results**

Sort by **Relevance** Date Range **01-01-1990** TO **12-05-2025**

1,088,861 results for  
Want more refined results? Try putting "" marks around your search term: i.e. "ministry of state security"

**AD**

**선군시대 《아리랑》 류의 노래들**  
(평양 6월 28일발 조선중앙통신)오늘 조선에서는 민요풍인 《아리랑》 류의 노래들이 우리 인민의 남다른 관심과 사랑속에 선군시대를 장식  
June 29, 2006

**《필승불패의 조선식사회주의는 영원불멸할것이다》 - 《빠뜨리오뜨》 책임주필**  
(평양 6월 29일발 조선중앙통신)29일부 《로동신문》은 로씨아신문 《빠뜨리오뜨》의 책임주필 앤.아.젬스꼬브가 쓴 글 《필승불패의 조선  
June 29, 2006

**최태복비서가 수리아소년선봉대대표단을 만났다**  
(평양 6월 28일발 조선중앙통신)조선로동당 중앙위원회 비서 최태복은 28일 만수대의사당에서 수리아 바스소년선봉대 위원장 아흐마드  
June 29, 2006

**음악일화집 《장군님과 노래》 (상) 출판**  
(평양 6월 28일발 조선중앙통신)조선의 문학예술출판사에서 음악일화집 《장군님과 노래》 (상)를 출판하였다. 6개편으로 되여있는 도서에는  
June 29, 2006

**주조 무관단성원들 신천박물관 참관**  
(신천 6월 28일발 조선중앙통신)6.25-7.27반미공동투쟁월간에 즈음하여 주조 무관단성원들이 28일 신천박물관을 참관하였다. 무관단성원들

- 기간 전체 뉴스 url 수집 화면

```
actions = ActionChains(driver)
actions.send_keys(Keys.ESCAPE).perform()
time.sleep(1)
```

- 팝업, 알림창 방지 코드

```
while True:
    try:
        more_button = driver.find_element(By.CSS_SELECTOR, 'a.more-full-link')

        if more_button.is_displayed():
            driver.execute_script("arguments[0].scrollIntoView(true);", more_button)
            time.sleep(2)
            current_scroll = driver.execute_script("return window.scrollY;")
            scroll_to = current_scroll - 300
            driver.execute_script(f"window.scrollTo(0, {scroll_to});")
            time.sleep(2)
            driver.execute_script("arguments[0].click()", more_button)
            print(f'More Articles 클릭')
            time.sleep(2)

    else:
        print('버튼이 더 이상 보이지 않음 → 종료')
        break
```

- 더보기 누르는 코드

# 뉴스 데이터 - 북한(KCNA)

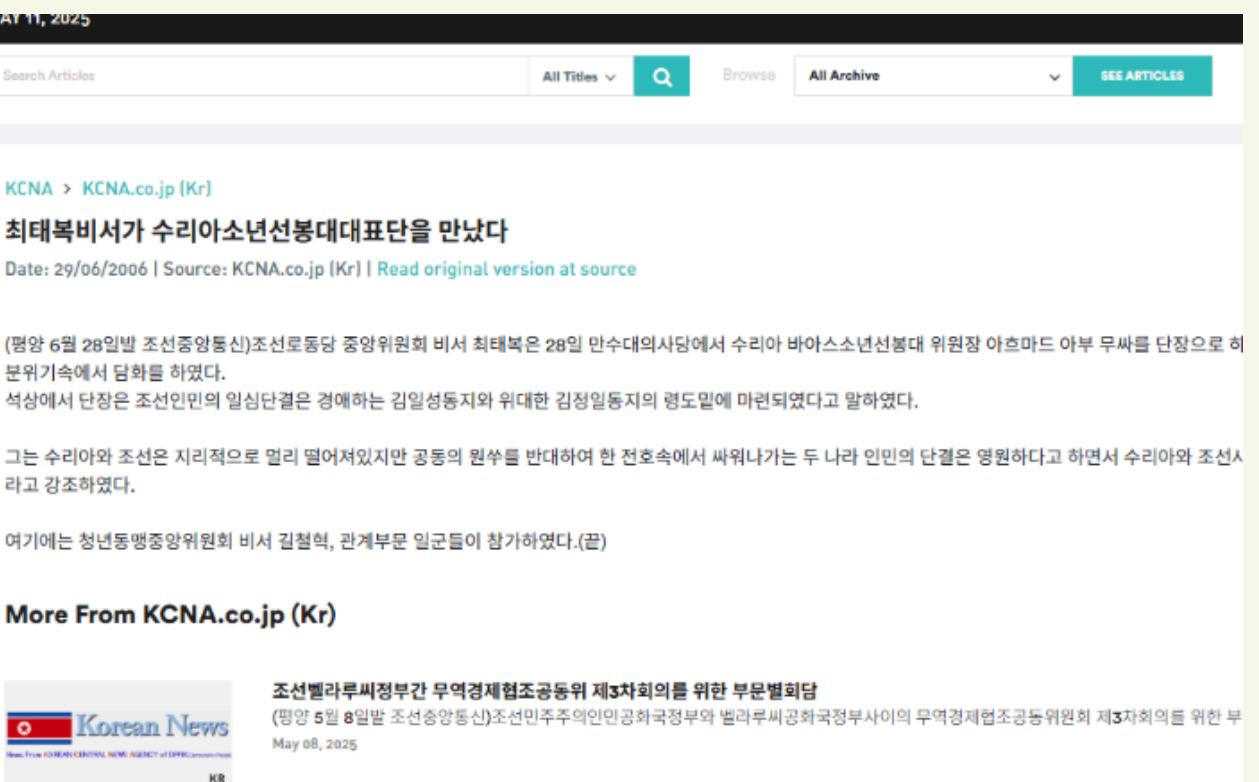
```

html = driver.page_source
soup = BeautifulSoup(html, 'html.parser')
actions.send_keys(Keys.ESCAPE).perform()
time.sleep(1)
articles = soup.find('div', {'id': 'search_wrapper'})
article_list = articles.find_all('div', {'class': 'article-desc'})

name_list, year_list, url_list = [], [], []

for article in article_list:
    span_tag = article.find('span')
    if span_tag:
        year_list.append(span_tag.text.strip())
for article in article_list:
    a_tag = article.find('a')
    if a_tag:
        name_list.append(a_tag.text.strip())
for article in article_list:
    a_tags = article.find_all('a')
    url_found = False
    for a_tag in a_tags:
        if a_tag.has_attr('href') and 'newstream' in a_tag.get('href'):
            url_list.append(a_tag.get('href'))
            url_found = True
            break
    if not url_found:
        url_list.append('')

```



## • 북한 뉴스 기사 내용 수집 화면

### • url 수집 코드

```

con_list = []
driver = webdriver.Chrome()
for url in url_list:
    driver.get(url)
    html = driver.page_source
    soup = BeautifulSoup(html, 'html.parser')
    contents = soup.find('div', {'class': 'article-content'})
    if contents:
        text = contents.get_text(separator='\n').strip()
        con_list.append(text)
    else:
        con_list.append('본문 없음')

```

### • 북한 뉴스 기사 내용 수집 코드

## 데이터 명세서

## 데이터 목록

No	카테고리	데이터명(한글)	데이터크기	개수	데이터유형	작성일자: 2025-05-12	제공처
1_01	도발유형 (4)	피습	17KB	252	xlsx	국회 도서관, 국가 정보원	
1_02		핵실험	2KB	6	csv	Beyond Parallel, WIKIPEDIA	
1_03		미사일 도발	213KB	309	xlsx	NTI(Nuclear Threat Initiative)	
1_04		오를풀선 살포	31KB	168	csv	Beyond Parallel	
2_01			735				
2_02		제2연평해전 한국기사 전체	1200KB	620	csv	국방일보	
2_03		북한 1차 핵실험 한국기사 전체	1500KB	810	csv	국방일보	
2_04		북한 2차 핵실험 한국기사 전체	1800KB	889	csv	국방일보	
2_05		천안함 피격 한국기사 전체	2100KB	893	csv	국방일보	
2_06		연평도 포격 한국기사 전체	2400KB	969	csv	국방일보	
2_07		북한 3차 핵실험 한국기사 전체	2800KB	905	csv	국방일보	
2_08		동해안 단거리 미사일 발사 한국기사 전체	3700KB	1074	csv	국방일보	
2_09		DMZ 목할지로 매설 한국기사 전체	4100KB	1025	csv	국방일보	
2_10		북한 4차 핵실험 한국기사 전체	3300KB	1005	csv	국방일보	
2_11		북한 5차 핵실험 한국기사 전체	3700KB	1105	csv	국방일보	
2_12		북한 6차 핵실험 한국기사 전체	4100KB	1472	csv	국방일보	
2_13		신형 단거리 탄도 미사일 한국기사 전체	5900KB	1295	csv	국방일보	
2_14		NLL 이남 대륙간 탄도 미사일 한국기사 전체	6000KB	966	csv	국방일보	
2_15		오를풀선 한국기사 전체	3700KB	943	csv	국방일보	
			15080				
3_01	남한 기사_국가별 (15)	제2연평해전 한국기사 국가별	576KB	168	csv	국방일보	
3_02		북한 1차 핵실험 한국기사 국가별	663KB	194	csv	국방일보	
3_03		북한 2차 핵실험 한국기사 국가별	1039KB	335	csv	국방일보	
3_04		천안함 피격 한국기사 국가별	1094KB	292	csv	국방일보	
3_05		연평도 포격 한국기사 국가별	1802KB	424	csv	국방일보	
3_06		북한 3차 핵실험 한국기사 국가별	1132KB	351	csv	국방일보	
3_07		동해안 단거리 미사일 발사 한국기사 국가별	1364KB	366	csv	국방일보	
3_08		DMZ 목할지로 매설 한국기사 국가별	978KB	227	csv	국방일보	
3_09		북한 4차 핵실험 한국기사 국가별	827KB	218	csv	국방일보	
3_10		북한 5차 핵실험 한국기사 국가별	1303KB	386	csv	국방일보	
3_11		북한 6차 핵실험 한국기사 국가별	2272KB	769	csv	국방일보	
3_12		신형 단거리 탄도 미사일 한국기사 국가별	1616KB	420	csv	국방일보	
3_13		NLL 이남 대륙간 탄도 미사일 한국기사 국가별	877KB	416	xlsx	국방일보	
3_14		오를풀선 한국기사 국가별	588KB	357	xlsx	국방일보	
3_15		평시상황 한국기사 국가별	8681KB	326	csv	국방일보	
			5249				

## 데이터 목록

No	카테고리	데이터명(한글)	데이터크기	개수	데이터유형	작성일자: 2025-05-12	제공처
4_01	북한 기사_전체 (15)	제2연평해전 북한기사 전체	1300KB	844	csv	KCNA	
4_02		북한 1차 핵실험 북한기사 전체	1500KB	956	csv	KCNA	
4_03		북한 2차 핵실험 북한기사 전체	2300KB	1045	csv	KCNA	
4_04		천안함 피격 북한기사 전체	1500KB	863	csv	KCNA	
4_05		연평도 포격 북한기사 전체	1800KB	1037	csv	KCNA	
4_06		북한 3차 핵실험 북한기사 전체	2800KB	1626	csv	KCNA	
4_07		동해안 단거리 미사일 발사 북한기사 전체	5700KB	3318	csv	KCNA	
4_08		DMZ 목할지로 매설 북한기사 전체	13100KB	5555	csv	KCNA	
4_09		북한 4차 핵실험 북한기사 전체	15200KB	5998	csv	KCNA	
4_10		북한 5차 핵실험 북한기사 전체	19500KB	7804	csv	KCNA	
4_11		북한 6차 핵실험 북한기사 전체	26000KB	9940	csv	KCNA	
4_12		신형 단거리 탄도 미사일 북한기사 전체	21200KB	9338	csv	KCNA	
4_13		NLL 이남 대륙간 탄도 미사일 북한기사 전체	5300KB	7021	csv	KCNA	
4_14		오를풀선 북한기사 전체	5200KB	3208	csv	KCNA	
4_15		평시상황 북한기사 전체	17400KB	17178	csv	KCNA	
			75731				
5_01	북한기사_국가별 (15)	제2연평해전 북한기사 국가별	592KB	255	csv	KCNA	
5_02		북한 1차 핵실험 북한기사 국가별	985KB	229	csv	KCNA	
5_03		북한 2차 핵실험 북한기사 국가별	1475KB	335	csv	KCNA	
5_04		천안함 피격 북한기사 국가별	951KB	260	csv	KCNA	
5_05		연평도 포격 북한기사 국가별	1131KB	310	csv	KCNA	
5_06		북한 3차 핵실험 북한기사 국가별	1579KB	592	csv	KCNA	
5_07		동해안 단거리 미사일 발사 북한기사 국가별	3720KB	1268	csv	KCNA	
5_08		DMZ 목할지로 매설 북한기사 국가별	14806KB	3374	csv	KCNA	
5_09		북한 4차 핵실험 북한기사 국가별	12834KB	2968	csv	KCNA	
5_10		북한 5차 핵실험 북한기사 국가별	22616KB	5628	csv	KCNA	
5_11		북한 6차 핵실험 북한기사 국가별	25830KB	6958	csv	KCNA	
5_12		신형 단거리 탄도 미사일 북한기사 국가별	11636KB	3786	csv	KCNA	
5_13		NLL 이남 대륙간 탄도 미사일 북한기사 국가별	5173KB	1530	csv	KCNA	
5_14		오를풀선 북한기사 국가별	972KB	447	csv	KCNA	
5_15		평시상황 북한기사 국가별	17500KB	6205	csv	KCNA	
			34145				

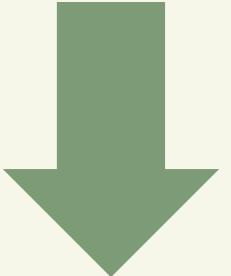
→ 총 130,940개의 데이터

# 북한 도발 유형별 데이터

## 예시) 피습 사건 수집 데이터

날짜	사건명	사건장소	위도	경도	내용
0 1945. 9. 5.	조선정판사, 위조지폐 사건	NaN	NaN	NaN	이만출, 권오지, 박탁용 등이 경제적인 혼란을 일으키기 위해 5회에 걸쳐 1,200...
1 1946. 1. 18.	학병동맹사건	NaN	NaN	NaN	신탁통치 반대시위를 하던 학생들에게 좌익학병동맹이 총기를 난사하여 유혈충돌을 일으켰다.
2 1946. 1. 29.	정계요인 암살 및 경찰서 방화기도 사건	NaN	NaN	NaN	한민당의 주석홍무익 송진우를 암살하려다 실패하고 구속된 공산당원 김길홍을 구출하기 ...
3 1946. 10. 1.	대구폭동 사건	대구 35.871567	128.601259		공산당의 선동하에 대구를 중심으로 한 경북, 우도도지에서 폭동이 발생하여, 피살자 ...

- 전처리 전



Type	Date	Case	Latitude	Longitude	Year	n_gov	s_gov
피습 사건	1945-09-05	조선정판사, 위조지폐 사건			1945	임시정부	임시정부
피습 사건	1946-01-18	학병동맹사건			1946	임시정부	임시정부
피습 사건	1946-01-29	정계요인 암살 및 경찰서 방화기도 사건			1946	임시정부	임시정부
피습 사건	1946-05-08	연백평야 통구(通水)단절 사건			1946	임시정부	임시정부
피습 사건	1946-07-12	국대안 반대 사건			1946	임시정부	임시정부
피습 사건	1946-08-15	보성 적색테러 사건	34.77021733	127.0790385	1946	임시정부	임시정부

- 전처리 후

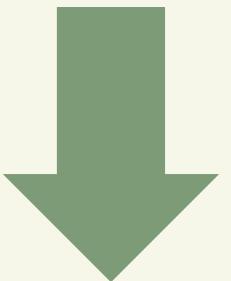
과정	내용
정제	불필요한 컬럼 제거, 컬럼명 통일
변환	날짜 형식 통일 Year, n_gov, s_gov 컬럼 추가
병합	유형별 데이터를 하나의 테이블로 병합

# 뉴스 데이터

예시) 한국 뉴스(국방일보) 크롤링 데이터

	Date	Title	content
1	2017. 09. 01	미 전폭적 지지…안보공약 재확인 <국방>	맹수열 기자 <guns13@dema.mil.kr>
2	2017. 08. 30	빅터 차 주한 미국대사 내정 <국방>	이영선 기자 <ys119@dema.mil.kr>
3	2017. 08. 29	중형 전술차량에 레이저무기 설치 <국방>	윤병노 기자 <trylover@dema.mil.kr>
4	2017. 08. 22	상어 학체로 생화학 위협 탐지 연구 <국방>	윤병노 기자 <trylover@dema.mil.kr>
5	2017. 08. 21	통같은 안보공약 변함없다 재확인" <국방>	김용호 기자 <yhkim@dema.mil.kr>
6	2017. 08. 21	정관, 한·미의 전략적 공조방안 논의 <국방>	김용호 기자 <yhkim@dema.mil.kr>

• 전처리 전



	Date	Title	Content	Category	Country
1	2017-08-01	정영칼럼] '창조 도발'의 종말	는 전략이 나오길 기대한다.	오피니언	미국
2	2017-08-01	활동…미·중 이어 세계 3위	<guns13@dema.mil.kr>	국방	미국
3	2017-08-01	대체 위해 민간 부문과 협력	국방기술품질원 수석연구원	기획연재	미국
4	2017-08-01	국성 위해 공개평정 도입해야	력이 될 것이라고 확신한다.	오피니언	미국
5	2017-08-01	특이한 北 참수함 활동 감지"	발언을 한 바 있다. 연합뉴스	국정일반	미국
6	2017-08-01	간미, 어느 때보다 긴밀 공조"	<jataka@dema.mil.kr>	국정일반	미국

• 전처리 후

과정	내용
정제	<p>불필요한 문자 제거</p>
변환	<p>날짜 형식 통일 국가 검색어 입력 뉴스 데이터 프레임 → Country(미국/중국/러시아/북한/한국) 컬럼 추가</p> <p>기간별 한국 전체 뉴스 데이터 프레임 → Category(국방/기획/국내.외/오피니언) 컬럼 추가</p>

# 뉴스 데이터

## - 워드 클라우드 생성을 위한 텍스트 전처리

```

for line in text_list:
    try:
        token_pos = okt.pos(line)
        tokens += [token for token, tag in token_pos if tag in ['Noun', 'Verb', 'Adjective']]
    except:
        continue

```

- 형태소 분석 및 단어 추출

```

stopwords = [
    '하였다', '하여', '하는', '하고있다', '다음', '였다', '되었다', '하기', '있다', '많은',
    '할수', '하면서', '대해', '한다', '하시였다', '따라', '하며', '누구', '때문', '위해',
    '어떤', '무슨', '위대한', '있는', '지금', '아니라', '그것', '입니다', '모든', '없다',
    '하신', '대하', '되어', '되며', '아니라', '바로', '지난해', '하면', '자기', '사실', '지난',
    '즈음', '되는', '의해', '되고', '않는', '합니다', '한다고', '같은', '있었다', '여러', '해야',
    '없는', '더욱', '대한', '이어', '가장', '관련', '오늘', '이런', '하게', '비롯',
    '다른', '높이', '하고', '계속', '미니', '하고있는', '자주', '우리', '여기', '최근', '이번', '있습니다',
    '전체', '되었습니다', '했습니다', '또한', '깊은', '되고있다', '로부터', '하지', '올해', '보고', '어떻게',
    '된다', '미려한', '있게', '그미', '않고', '하였습니다', '그대로', '했다', '무단', '배포', '금지', '저작권',
    '공유', '했다', '통해', '기사', '댓글', '스크랩', '국방일보', '이었다', '밝혔다'
]

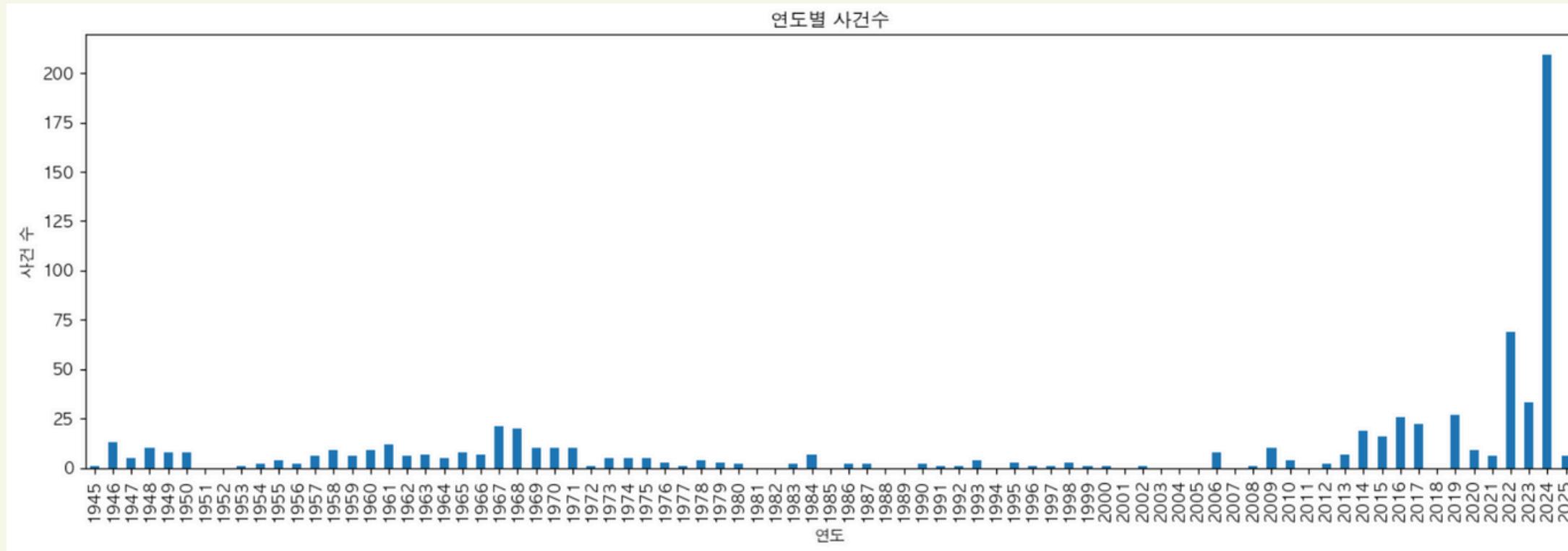
# 동적으로 제거할 단어 결정
if '전체' in filename:
    dynamic_stopwords = stopwords
elif nation == '한국':
    dynamic_stopwords = stopwords.union({'한국', '남조선'})
elif nation == '러시아':
    dynamic_stopwords = stopwords.union({'러시아', '로씨야'})
else:
    dynamic_stopwords = stopwords.union({nation})

# 불용어 적용
tokens = [token for token in tokens if len(token) >= 2 and token not in dynamic_stopwords]
cnt = Counter(tokens)

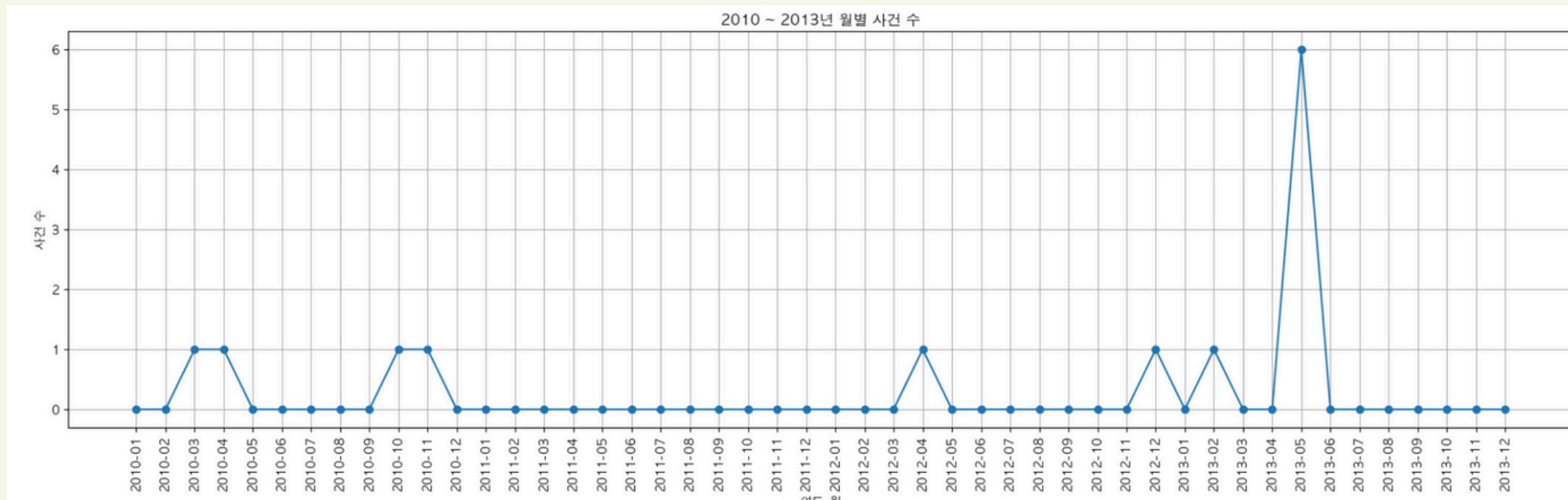
```

- 불용어 지정을 활용한 단어 정제

# 시기별 북한 도발 빈도



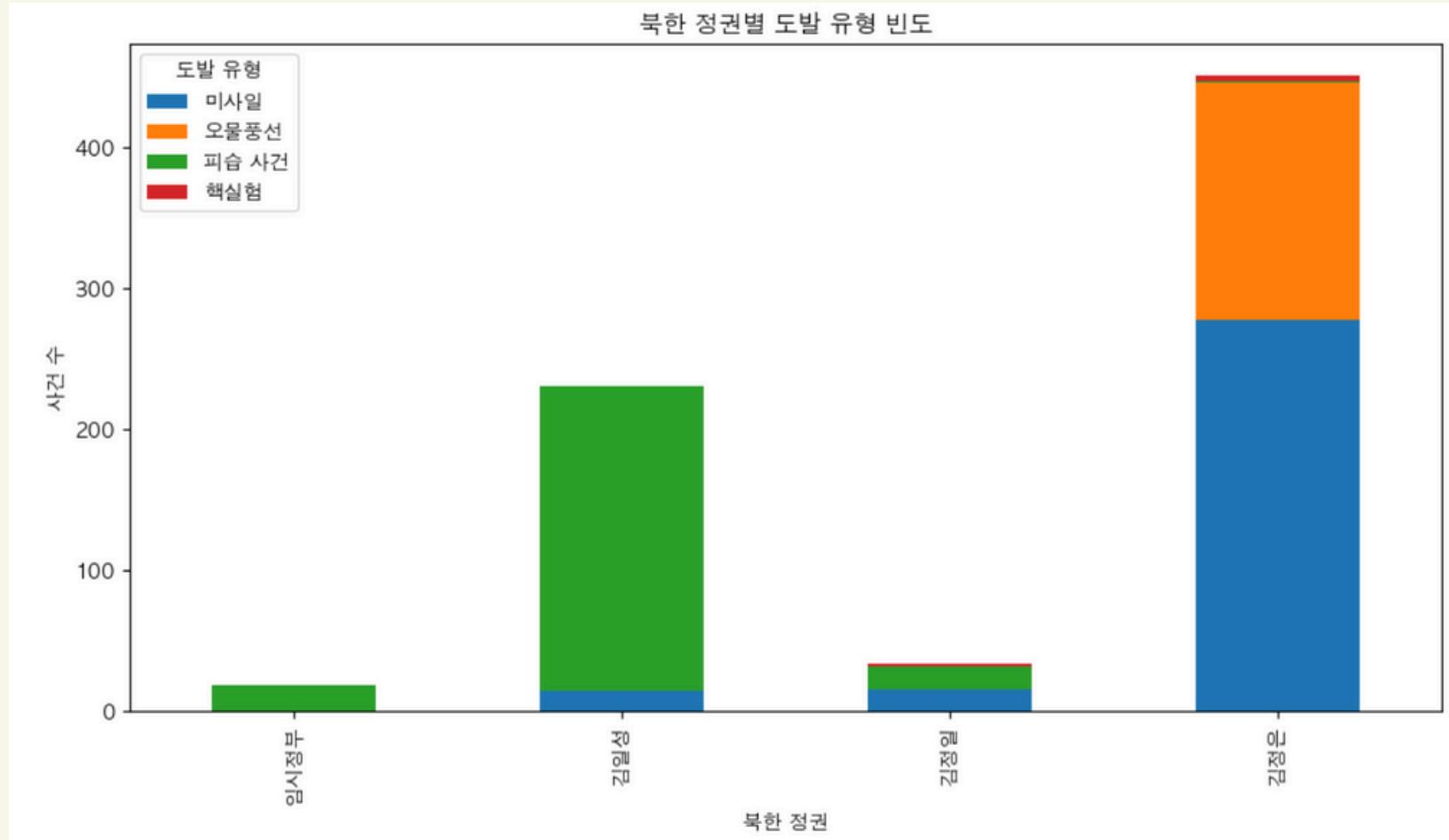
• 정전 이후 연도별 북한의 도발 빈도



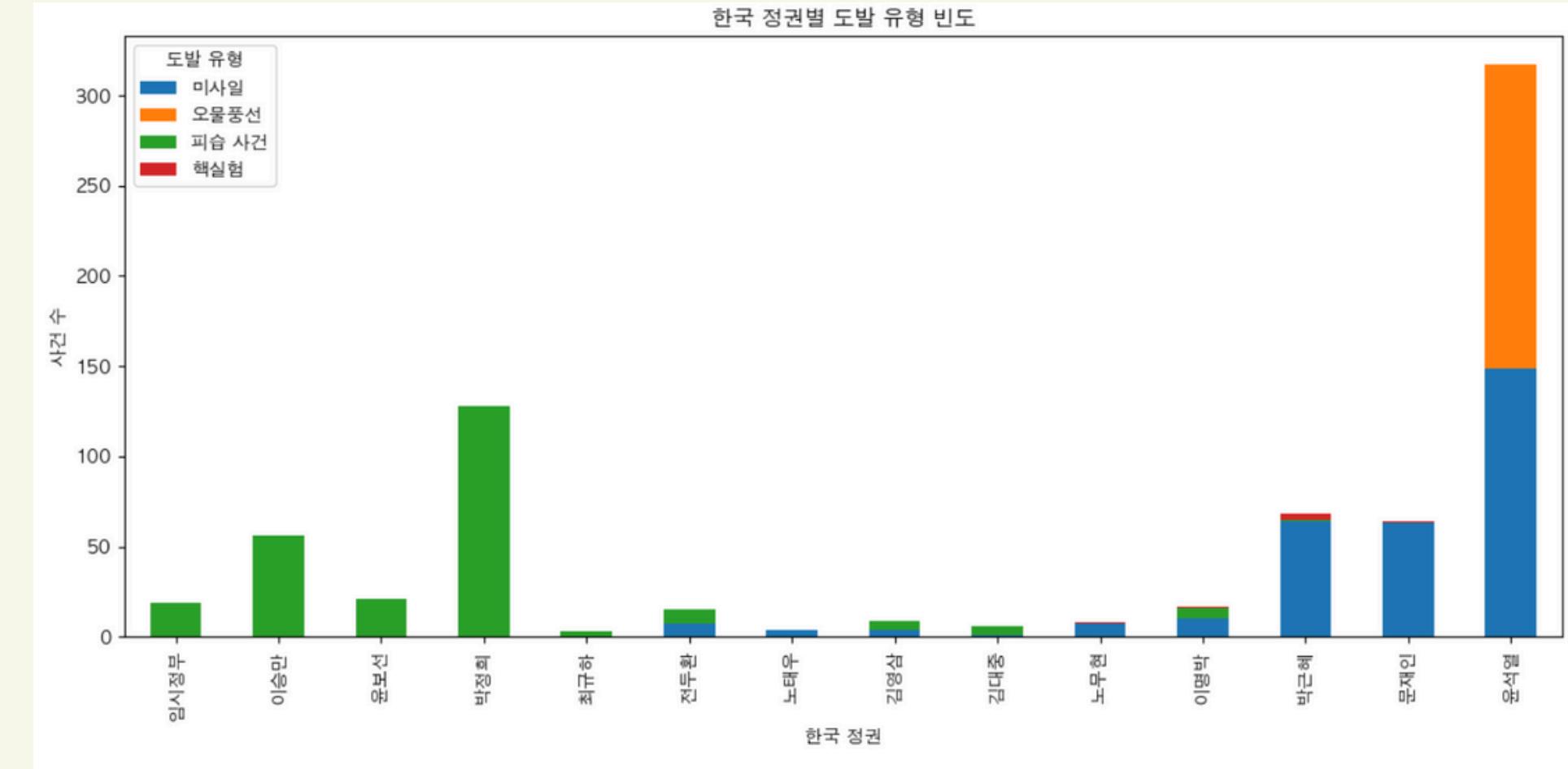
• (2010~2013년) 연도별 북한의 도발 빈도

→ 도발 징후 분석을 위한 평온 시기 설정  
 : 도발이 발생하지 않은 시기인  
 2010년 12월부터 2012년 3월 사이의  
 중간 기간(2011년 3월~9월)

# 남북한 정권별 북한 도발 유형 변화

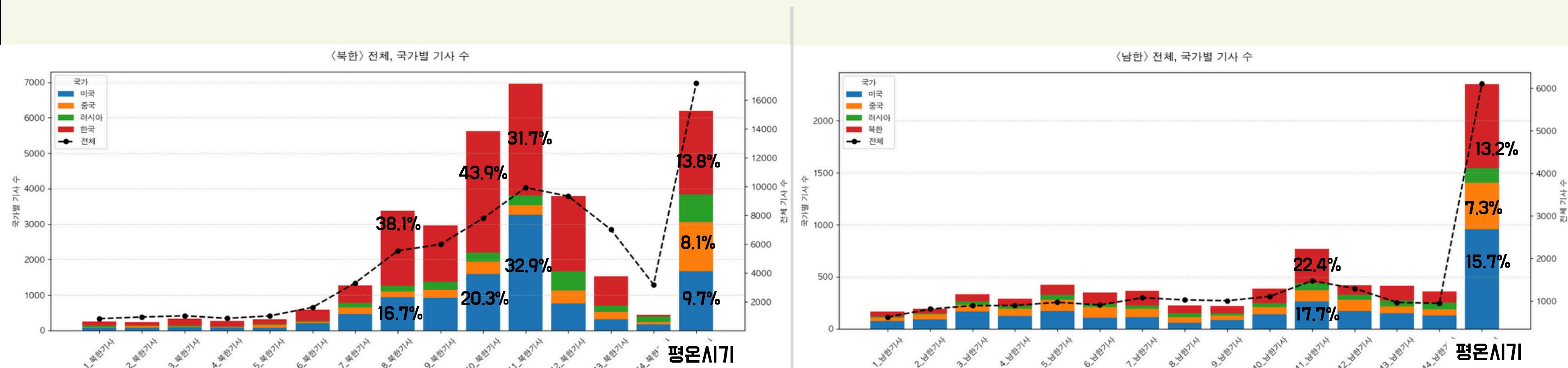


• 북한 정권별 도발 유형 빈도



• 남한 정권별 도발 유형 빈도

# 도발 전/ 평온시기 기사 수 비교



case 8(DMZ 목함지로 설치), 10(5차 핵실험), 11(6차 핵실험)  
 → 평온 시기 보다 증가한 한국, 미국 단어 포함 기사 비율

case 11(6차 핵실험)  
 → 평온 시기 보다 증가한 북한, 미국 단어 포함 기사 비율

# 워드 클라우드



• 북한 기사 워드 클라우드

• 남한 기사 워드 클라우드

## 메인 페이지

Deploy :

main

attack

balloon

missile

nuclear

## 📰 뉴스기사 크롤링을 통한 북한 도발 징후 포착 🚨

(2)

도발 위치 지도

안내: 지도 위 마커를 클릭하면 아래에 사건 페이지로 이동할 수 있는 버튼이 표시됩니다.

(1)

성권별 도발 유형 보기

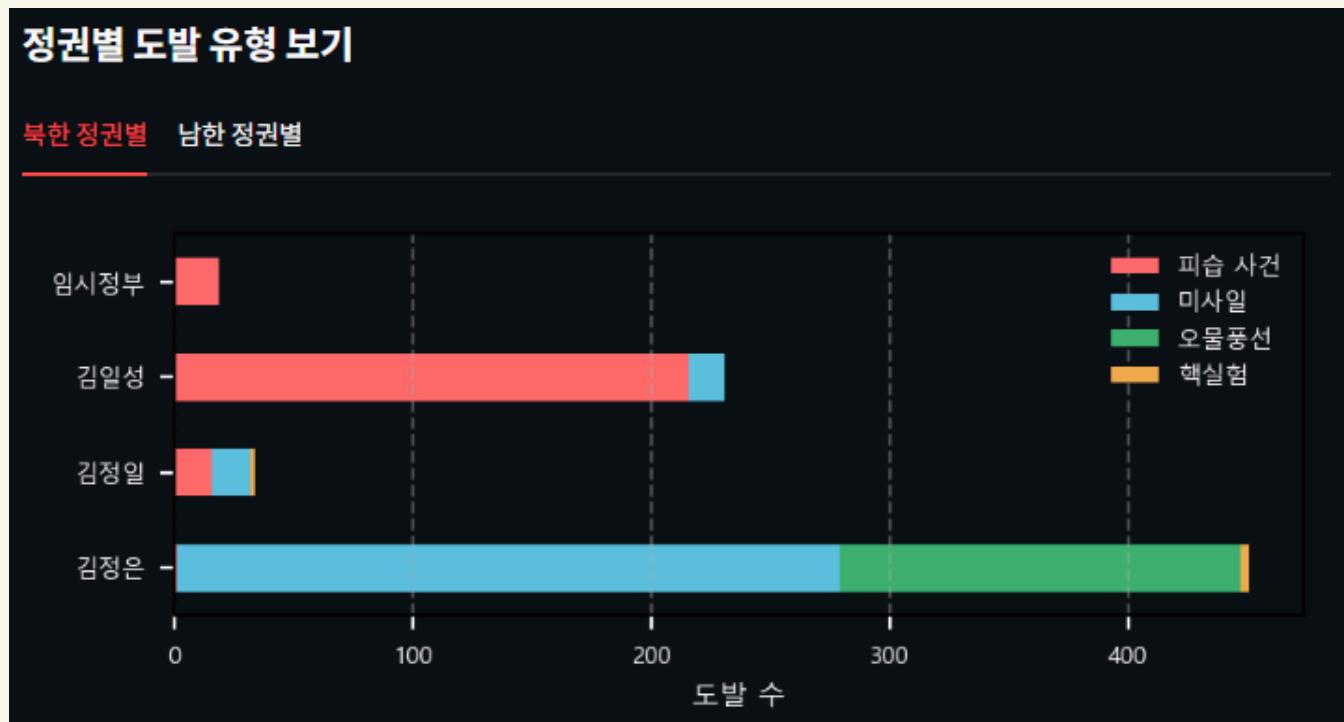
북한 정권별 남한 정권별

도발 비율 차트

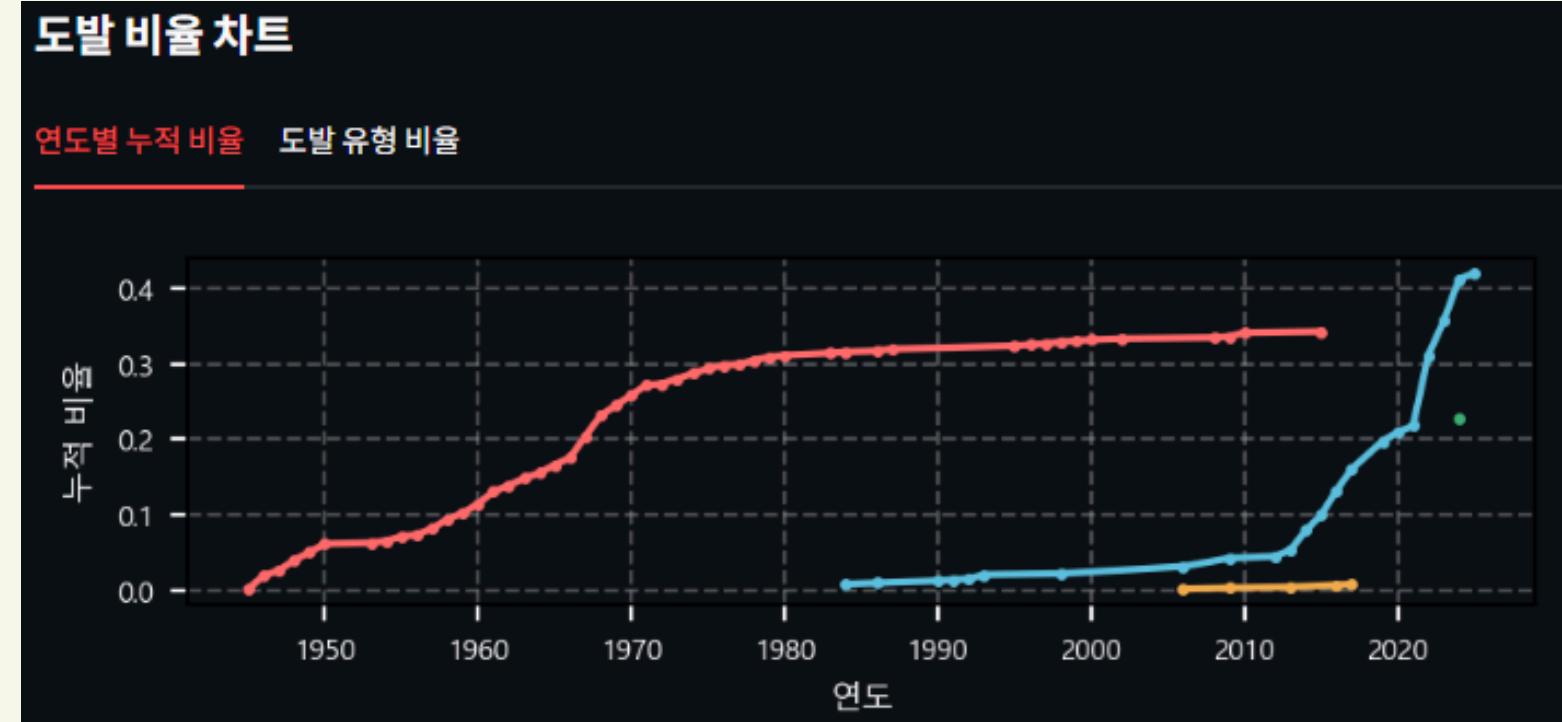
연도별 누적 비율 도발 유형 비율

# 메인 페이지 - ① 북한 도발에 대한 분석

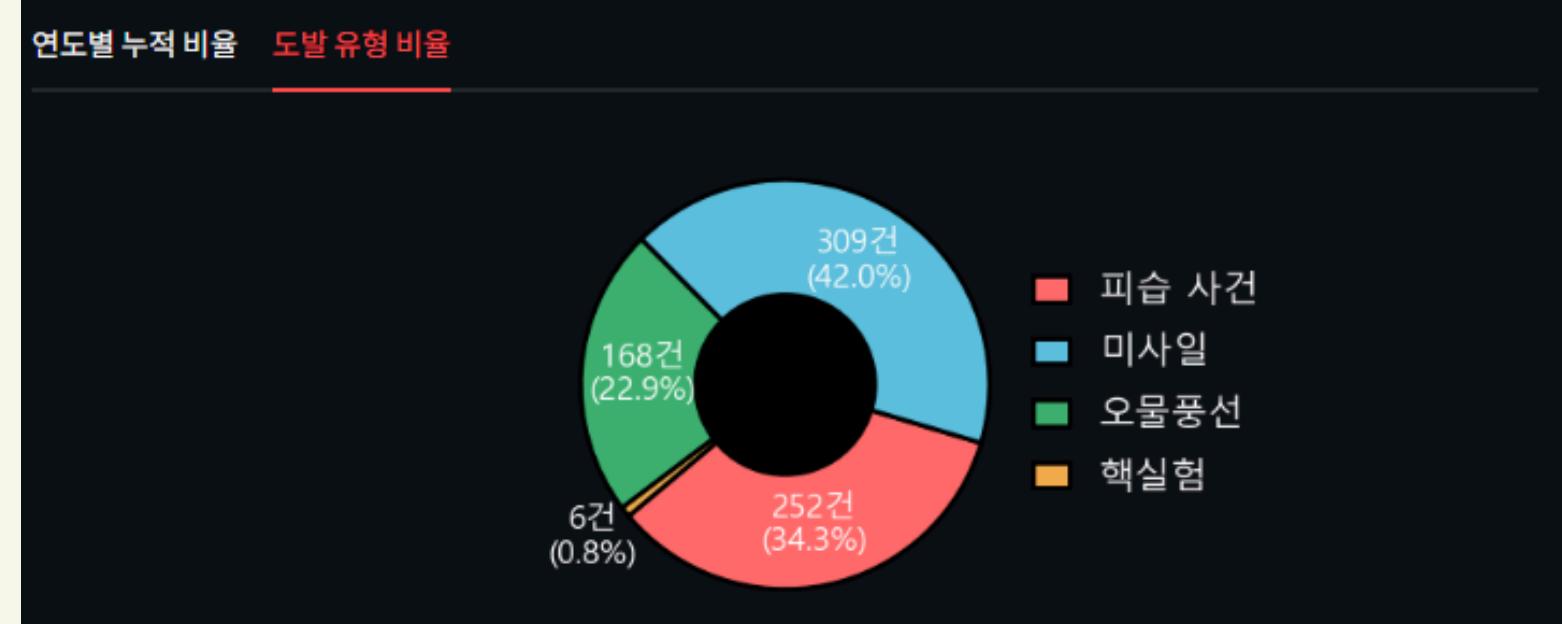
## - 정권별 도발 유형



## - 도발 유형별 비교



## 도발 비율 차트



# 메인 페이지 - ② 대표 도발 사건 위치 지도

**도발 위치 지도**

지도 위 마커를 클릭하면 아래에 사건 페이지로 이동할 수 있는 버튼이 표시됩니다.

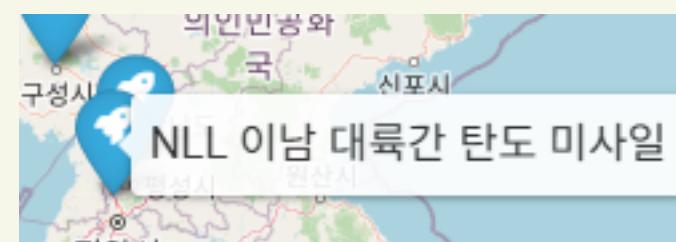
선택된 사건: NLL 이남 대륙간 탄도 미사일

Missile 사건 페이지로 이동

## - 마커 아이콘 설명창



## - 마커 상호작용



### • 마커에 마우스를 올렸을 때



### • 마커 클릭하였을 때

```
with col_icon:
    st.markdown('')

    <div style="position: relative; display: inline-block; cursor: pointer;">
        <i class="fa fa-info-circle" style="font-size: 26px; text-shadow: 0 0 5px black;"></i>
        <div style="position: absolute; top: 50%; right: 10%; /* 오른쪽에 붙었으니 왼쪽으로 뜨게 */ margin-right: 10px; transform: translateY(-50%); opacity: 0; transition: opacity 0.3s ease-in-out; font-size: 13px; box-shadow: 0 0 10px rgba(0,0,0,0.7);">
            <div style="margin-bottom: 6px;">
                <i class="fa fa-user" style="color: red; font-size: 16px;"></i> : 피습 사건
            </div>
            <div style="margin-bottom: 6px;">
                <i class="fa fa-bomb" style="color: orange; font-size: 16px;"></i> : 핵실험
            </div>
            <div style="margin-bottom: 6px;">
                <i class="fa fa-circle" style="color: green; font-size: 16px;"></i> : 오물풍선
            </div>
            <div>
                <i class="fa fa-rocket" style="color: blue; font-size: 16px;"></i> : 미사일
            </div>
        </div>
        <style>
            div:hover > div {
                visibility: visible !important;
                opacity: 1 !important;
            }
        </style>
    </div>
    <link href="https://cdnjs.cloudflare.com/ajax/libs/font-awesome/4.7.0/css/font-awesome.min.css" rel="stylesheet">
    <!--
    unsafe_allow_html=True
    -->
```

### • i 아이콘 생성 코드

# 메인 페이지 - ② 대표 도발 사건 위치 지도



## - 서브 페이지 이동

```

if 'selected_case' not in st.session_state:
    st.session_state.selected_case = None

selected_case = None
if map_data and map_data.get("last_object_clicked_tooltip"):
    selected_case = map_data["last_object_clicked_tooltip"]
    st.session_state.selected_case = selected_case

if selected_case:
    st.session_state.selected_case = selected_case

current_case = st.session_state.selected_case

if current_case:
    st.markdown(f"선택된 사건: {current_case}", unsafe_allow_html=True)
    category = marker_category_map.get(current_case)
    if category == 'attack':
        if st.button("Attack 사건 페이지로 이동"):
            st.switch_page("pages/attack.py")
    elif category == 'nuclear':
        if st.button("Nuclear 사건 페이지로 이동"):
            st.switch_page("pages/nuclear.py")
    elif category == 'balloon':
        if st.button("Balloon 사건 페이지로 이동"):
            st.switch_page("pages/balloon.py")
    elif category == 'missile':
        if st.button("Missile 사건 페이지로 이동"):
            st.switch_page("pages/missile.py")
    else:
        st.info("카테고리를 찾을 수 없습니다.")

```

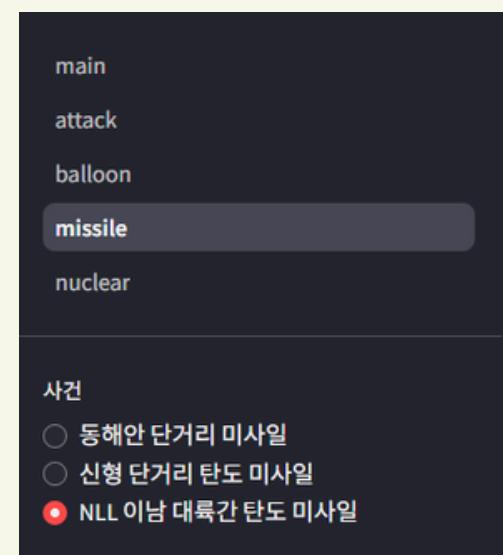
## • 메인 페이지 코드

```

selected_case = st.session_state.get('selected_case', None)
if selected_case and selected_case in reverse_mapping:
    expectedSubview = reverse_mapping[selected_case]
    if st.session_state.get('Subview') != expectedSubview:
        st.session_state['Subview'] = expectedSubview
        st.rerun()

```

## • 서브 페이지 코드



## 뉴스기사 크롤링

### 미사일 도발 징후 분석

#### 3) NLL 이남 대륙간 탄도 미사일



## • 서브 페이지 이동 후 화면



# 서브 디자인

## 예시) 미사일 도발 유형 서브 페이지

```
# 페이지 시작
subview = reverse_mapping[subview_label]

st.title("▣ 뉴스기사 크롤링을 통한 북한 도발 징후 포착")
st.header('미사일 도발 징후 분석', divider=True)

m_icon = get_icon_name('미사일')
m_icon_color = case_color_map('미사일')

if subview == "missile_case_1":
    render_event_case(
        case_title='1) 동해안 단거리 미사일',
        date='2014년 3월 28일',
        content_lines=[
            "2014년 한미연합훈련 '키 리졸브'와 '독수리'에 대한 반발로 미사일을 발사함",
            "해당 미사일은 scud 계열의 단거리 탄도미사일로 추정됨"
        ],
        coords=coords_missile[0],
        casenum=1,
        event_type='미사일'
    )

elif subview == "missile_case_2":
    render_event_case(
        case_title='2) 신형 단거리 탄도 미사일',
        date='2019년 5월 9일',
        content_lines=[
            "오후 4시 29분과 4시 49분경 평안북도 구성 지역에서 불상의 발사체를 각각 등쪽으로
            "군 당국은 이를 지난 2017년 화성-15형 이후 1년 5개월 만에 발사된 미사일로 규정",
            "코로나19로 국제사회 관심이 분산된 틈을 타 체계 건재함을 과시하려는 의도가 있음",
            "미국 및 한국 정부를 대상으로 한 외교적·군사적 압박 목적으로 내포됨",
            "정밀도와 신뢰성 향상을 위해 반복 실험을 수행한 것으로 분석됨",
            "통상적인 군사훈련의 일환으로 무기 성능 실험을 진행한 것으로 보임",
            "(김보미, 북한 단거리 미사일 시험발사의 배경과 함의, INSS 전략보고, 2020.09.)"
        ],
        coords=coords_missile[1],
        casenum=2,
        event_type='미사일'
    )

elif subview == "missile_case_3":
    render_event_case(
        case_title='3) 미니 이란 대류간 탄도 미사일'
```

### • 서브 페이지(미사일) 코드

# 서브 페이지 - ① 도발 사건 정보 제공

예시) 미사일 도발 유형 서브 페이지

Deploy

■ 뉴스기사 크롤링을 통한 북한 도발 징후 포착

미사일 도발 징후 분석

3. NLL 이남 대륙간 탄도 미사일

일시 | 2022년 11월 2일

내용

북한은 오전 8시 51분 강원도 원산 일대에서 SRBM 3발을 동해상으로 발사함  
그 중 한 발은 분단 이후 처음으로 동해 NLL 이남 공해상에 낙하하여 실질적 영토 침해로 간주됨  
해당 도발은 한미 연합공중훈련 '비질런트 스톰'에 대한 강력한 반발로 분석됨  
훈련에는 미국의 최신 스텔스기 F-35B가 참가하고 있었으며, 이에 대한 대응성 도발임  
(BBC뉴스 코리아, 북한, 한국 겨냥 탄도미사일 발사... '변화 바라지 말라'는 시그널, 2022.11.02)

- 위치(지도), 일시, 내용

```
def render_event_case(case_title, date, content_lines, coords, casenum, event_type):
    m_icon = get_icon_name(event_type)
    m_icon_color = case_color_map(event_type)

    st.subheader(case_title)
    col1, col2 = st.columns([1, 2], gap='medium')

    with col1:
        with st.container():
            create_folium_map_stop(
                lat=coords[0], lon=coords[1],
                width=300, height=300,
                zoom_start=5.5,
                icon_name=m_icon, icon_color=m_icon_color,
                tooltip=None, popup_text=None
            )

    with col2:
        table_html = (
            '<table border="1" style="width:100%; border-collapse:collapse;">'
            '<tr>'
            f'<td style="width:20%; font-weight:bold; vertical-align:top;">일시</td>'
            f'<td>{date}</td>'
            '</tr>'
        )

        if content_lines:
            table_html += (
                '<tr>'
                f'<td style="width:20%; font-weight:bold; vertical-align:top; border-right:1px solid #444;">내용</td>'
                f'<td style="padding:5px; word-wrap:break-word; overflow-wrap:break-word; '
                f'white-space:normal; border-bottom:1px solid #444;">{content_lines[0]}</td>'
                '</tr>'
            )

            for line in content_lines[1:]:
                table_html += (
                    '<tr>'
                    f'<td style="padding:5px; word-wrap:break-word; overflow-wrap:break-word; '
                    f'white-space:normal; border-bottom:1px solid #444;">{line}</td>'
                    '</tr>'
                )

            table_html += '</table>'

    st.markdown(table_html, unsafe_allow_html=True)
```

• 도발 사건별 페이지 생성 함수

→ **render\_event\_case** ( 사건명, 일시, 내용, 사건\_df, 사건 번호, 도발 유형 )

```
def create_folium_map_stop(lat, lon,
                           width=420, height=420,
                           zoom_start=6,
                           icon_name='rocket', icon_color='darkred',
                           map_center=[38.03375, 128.02025],
                           tooltip=None, popup_text=None):
    m = folium.Map(
        location=map_center,
        zoom_start=zoom_start,
        scrollWheelZoom=False,
        dragging=False
    )
    folium.Marker(
        [lat, lon],
        tooltip=tooltip,
        popup=popup_text,
        icon=folium.Icon(icon=icon_name, prefix='fa', color=icon_color)
    ).add_to(m)
    st.markdown('<div style="text-align:center">', unsafe_allow_html=True)
    st_data = st_folium(m, width=width, height=height)
    st.markdown('</div>', unsafe_allow_html=True)
```

• 맵 생성 함수

→ **create\_folium\_map\_stop()**

# 서브 페이지 - ① 도발전/ 평온 시기 비교

## 예시) 미사일 도발 유형 서브 페이지

■ 웹 크롤링을 통한 도발전/ 평온 시기 비교

1) 뉴스기사 선정 시기

도발전 수집기간	도발일 기준 1개월 전
비교군 수집기간	2011-03-01 ~ 2011-09-30
비교 방법	워드 클라우드

2) 뉴스기사 비교

북한 뉴스 남한 뉴스

전체 뉴스기사

도발 전

평온 시기

검색어 뉴스기사(북한뉴스)

검색어: 한국, 남조선

검색어: 미국

검색어: 중국

검색어: 루시아

```

st.markdown("<hr style='border:2px dashed #aaa;'>", unsafe_allow_html=True)
st.header('* 웹 크롤링을 통한 도발전/ 평온 시기 비교')
st.subheader('뉴스 선정 시기')
st.markdown(
    """
    <div style='width:50%;'>
        <table style='font-size:20px; border-collapse:collapse; width:100%; table-layout:fixed;'>
            <tr>
                <td style='font-weight:bold; padding:8px; border:1px solid #444;'>뉴스 수집 기간(도발전)</td>
                <td style='padding:8px; border:1px solid #444;'>도발일 기준 1개월 전</td>
            </tr>
            <tr>
                <td style='font-weight:bold; padding:8px; border:1px solid #444;'>뉴스 수집 기간(비교군)</td>
                <td style='padding:8px; border:1px solid #444;'>2011-03-01 ~ 2011-09-30</td>
            </tr>
            <tr>
                <td style='font-weight:bold; padding:8px; border:1px solid #444;'>비교 방법</td>
                <td style='padding:8px; border:1px solid #444;'>워드 클라우드</td>
            </tr>
        </table>
    </div>
    """
)
st.subheader('뉴스 비교')
wordcloud(casenum)

```

- 도발 사건별 페이지 생성 함수  
→ `reder_event_case()`

- 도발 전 1달 기간 / 평온 시기(7개월) 비교 워드 클라우드

```

def wordcloud(casenum, ):
    direc = 'images/' + str(casenum)

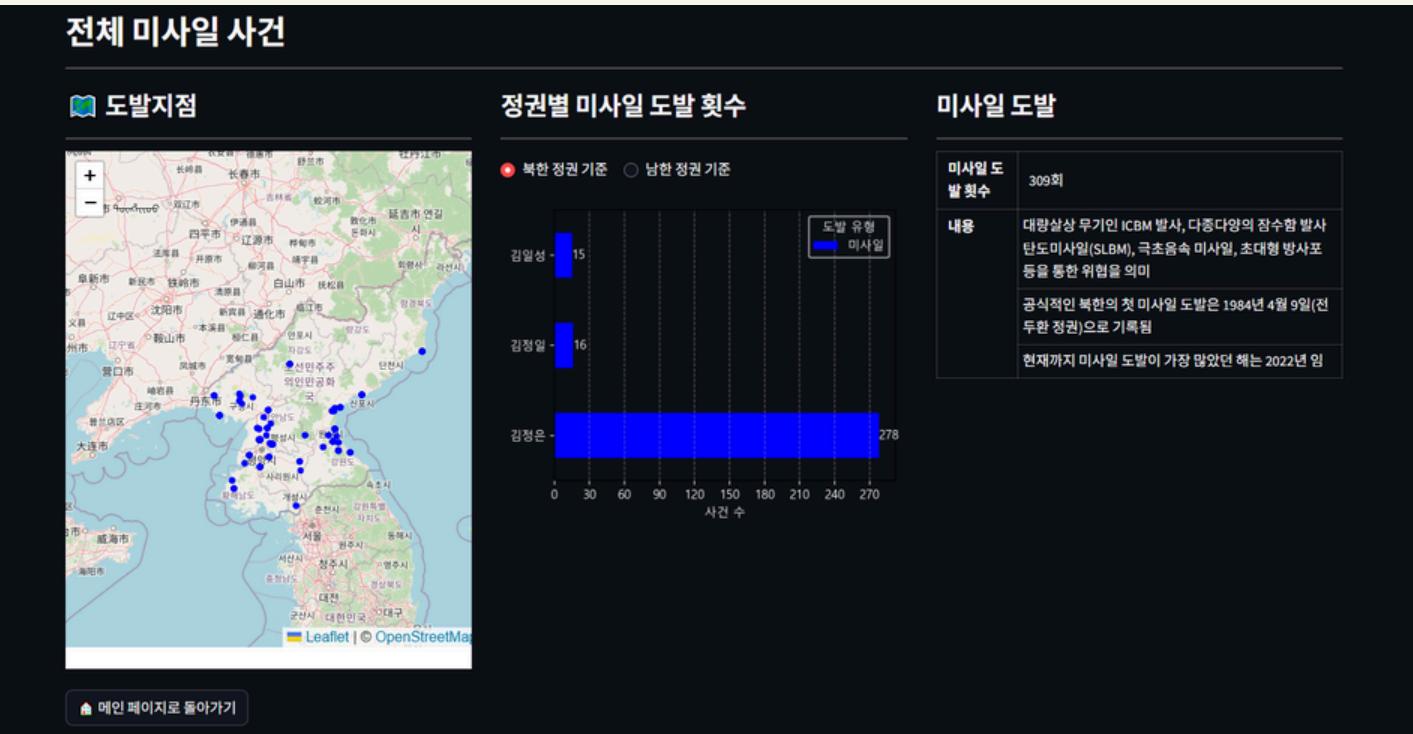
    tab_title = ['북한 기사', '한국 기사']
    [tab1, tab2] = st.tabs(tab_title)
    # 북한 기사 워드 클라우드
    with tab1:
        with st.container():
            st.subheader('전체 기사')
            with st.container(border=True):
                col1, col2 = st.columns(2)
                with col1:
                    st.image(direc + '_북한기사_전체_wordcloud.png')
                    st.markdown(
                        "<div style='text-align: center; font-size: 28px; font-weight: bold;'>도발 전</div>",
                        unsafe_allow_html=True
                    )
                with col2:
                    st.image('images/15_북한기사_전체_wordcloud.png')
                    st.markdown(
                        "<div style='text-align: center; font-size: 28px; font-weight: bold;'>평온 시기</div>",
                        unsafe_allow_html=True
                    )
            st.subheader('검색어 기사(북한뉴스)')
            row = st.columns(4)
            with row[0]:
                with st.container(border=True):
                    st.write('검색어 : 남한')
                    st.image(direc + '_북한기사_한국_wordcloud.png')
                    st.image('images/15_북한기사_한국_wordcloud.png')
            with row[1]:
                with st.container(border=True):
                    st.write('검색어 : 미국')
                    st.image(direc + '_북한기사_미국_wordcloud.png')
                    st.image('images/15_북한기사_미국_wordcloud.png')
            with row[2]:
                with st.container(border=True):
                    st.write('검색어 : 중국')
                    st.image(direc + '_북한기사_중국_wordcloud.png')
                    st.image('images/15_북한기사_중국_wordcloud.png')
            with row[3]:
                with st.container(border=True):
                    st.write('검색어 : 러시아')
                    st.image(direc + '_북한기사_러시아_wordcloud.png')
                    st.image('images/15_북한기사_러시아_wordcloud.png')
            st.subheader('검색어 기사(한국뉴스)')
            row = st.columns(4)
            with row[0]:
                with st.container(border=True):
                    st.write('검색어 : 미국')
                    st.image(direc + '_한국기사_미국_wordcloud.png')
                    st.image('images/15_한국기사_미국_wordcloud.png')
            with row[1]:
                with st.container(border=True):
                    st.write('검색어 : 중국')
                    st.image(direc + '_한국기사_중국_wordcloud.png')
                    st.image('images/15_한국기사_중국_wordcloud.png')
            with row[2]:
                with st.container(border=True):
                    st.write('검색어 : 러시아')
                    st.image(direc + '_한국기사_러시아_wordcloud.png')
                    st.image('images/15_한국기사_러시아_wordcloud.png')
            with row[3]:
                with st.container(border=True):
                    st.write('검색어 : 남한')
                    st.image(direc + '_한국기사_한국_wordcloud.png')
                    st.image('images/15_한국기사_한국_wordcloud.png')
    
```

- 워드 클라우드 생성 함수

→ `wordcloud(사건 번호)`  
: 워드클라우드 이미지파일 사용

# 서브 페이지 - ② 도발 유형별 정보 제공

예시) 미사일 도발 유형 서브 페이지



## - 도발 유형별 정보 제공

→ 도발 지점(지도).

정권별 유형의 도발 횟수. 내용

```
with st.container():
    acol1, acol2, acol3 = st.columns(3, gap='medium')
    with acol1:
        st.subheader('도발지점', divider=True)
        m = all_map_markers(missile_df, m_icon_color, center=(39.9642135, 127.0016985))
        st_folium(m, width=550, height=550)

    with acol2:
        st.subheader('정권별 미사일 도발 횟수', divider=True)
        regime_choice = st.radio(
            "#### 정권별 미사일 도발 횟수 보기",
            ('북한 정권 기준', '남한 정권 기준'),
            horizontal=True,
            key='missile_radio',
            label_visibility='collapsed'
        )
        draw_regime_chart(
            df=missile_df,
            regime_type='북한' if regime_choice == '북한 정권 기준' else '남한',
            attack_type='미사일'
        )

    with acol3:
        st.subheader('미사일 도발', divider=True)
        content_lines = [
            "대량살상 무기인 ICBM 발사, 다종다양의 잠수함 발사 탄도미사일(SLBM),",
            "극초음속 미사일, 초대형 방사포 등을 통한 위협을 의미",
            "공식적인 북한의 첫 미사일 도발은 1984년 4월 9일(전두환 정권)으로 기록됨",
            "현재까지 미사일 도발이 가장 많았던 해는 2022년임"
        ]
        st.markdown(make_summary_table("미사일 도발", missile_count, content_lines), unsafe_allow_html=True)

if st.button("메인 페이지로 돌아가기"):
    st.switch_page("main.py")
```

## • 서브 페이지(미사일) 코드

```
def all_map_markers(df, color, center=(36.53375, 127.82025), zoom=6):
    m = folium.Map(
        location=center,
        zoom_start=zoom,
        scrollWheelZoom=False,
        dragging=False
    )
    coords = df[['latitude', 'longitude', 'case_description']].dropna().values.tolist()
    for lat, lon, _ in coords:
        try:
            lat_f, lon_f = float(lat), float(lon)
            folium.CircleMarker(
                location=[lat_f, lon_f],
                radius=2,
                color=color,
                fill=True,
                fill_color=color,
                fill_opacity=0.8
            ).add_to(m)
        except (ValueError, TypeError):
            continue
    return m
```

- 맵 생성 함수
- all\_map\_markers()

# 대시보드 시연

### 3. 7대효과

# 인사이트 도출(북한 6차 핵실험)

## 〈 북한 기사 워드클라우드 〉

- 검색어 : 한국

## • 검색어 : 중국

- 검색어 : 러시아

- **검색어** : 미국

출처: 신지홍, 강영두, 트럼프 "北, 미국 더 위협말라...'화염과 분노'에 직면할 것"(종합), 연합뉴스, 2017.08.09.  
출처: 차두현, 박지현, 북한 6차 핵실험의 의미: 도박인가, 승부수인가?, 아산정책연구원, 2017.09.04.

## 트럼프 "北, 미국 더 위협말라…'화염과 분노'에 직면할 것"(종합)

송고 2017-08-09 05:13

# 트럼프 “화염과 분노” 경고에 북한은 “괌 폭격 검토”

북한, 미국 예방전쟁 시사에 괄 폭격으로 대응  
미 국방정보국, 북한의 핵탄두 소형화 인정

첫째, 자신들은 결코 ‘겁쟁이’가 아니라는 워싱턴에 대한 메시지이다. 7월 28일 북한의 ‘화성-14’호 미사일 발사로 고조되기 시작한 한반도의 긴장은 8월 8일 트럼프 대통령의 ‘화염과 분노(fire and fury)’ 발언, 그리고 북한 전략군사령부의 ‘괌 포위사격’ 발언으로 정점을 향해 치닫는 양상을 보였다. 그러나 8월 14일 김정은이 전략군 사령부의 ‘괌 포위사격’ 계획을 보고받는 자리에서 상황을 관망하겠다는 뜻을 시사함으로써 긴장은 상대적으로 완화되었고, 미국 역시 이러한 신호를 긍정적으로 평가했다. 이런 고양새는 김정은으로 하여금 마치 자신들이 미·북간의 전략적 게임에서 패배하는 것으로 여겨졌을 것이다. 즉, 북한은 자신들이 ‘치킨’이 아니며 미국의 태도변화가 없이는 어떠한 긴장도 감수할 수 있다는 인상을 주려 한 것이다 이와 관련하여 미국 국무부는 헤더 노트 대변인 명의의 외신기자 브리핑을 통해 미·북 대화가 이루어지려면 북한이 먼저 “핵실험을 중단하고, 탄도미사일 발사실험 역시 하지 않으며, 동북아시아의 평화를 약화시키는 언행을 자제해야 한다”는 전제를 제시한 바 있다.<sup>4</sup> 북한은 이러한 미국의 요구를 알아들일 생각이 없으며, 만일 대화가 필요하다면 미국이 먼저 머리를 숙이고 나와야 한다는 점을 강조한 것이다.

출처: 정의길, 이용인, 트럼프 “화염과 분노” 경고에 불화은 “괌 폭격 검토”, 한겨레 신문, 2017.08.09.

# 인사이트 도출(북한 전체 사건 키워드)

평온시기															도발전 단어 순위						
조선	2259	조선	3685	조선	5714	인민	6582	조선	8343	조선	10377	인민	10489	인민	9368	중앙	2451	조선	16236	1	11 김정은
평양	1811	평양	3231	평양	5238	조선	4991	청년	6289	인민	7969	최고	9462	혁명	6034	위원회	2403	평양	16041	3	12 애하
국가	1325	중앙	2389	인민	4745	김정	4238	평양	6267	전쟁	7766	조선	8014	건설	5093	평양	2379	중앙	10784	5	22 최고
사람	1187	사람	2299	전쟁	3968	평양	4123	인민	5817	제재	5363	령도자	7096	생산	5024	생산	2269	국가	7830	10	29 청년
김정은	1063	위원회	2060	수령	3593	장군	3982	혁명	5618	공화국	5236	김정은	6733	애하	4968	사람	2139	인민	7650	2	30 조국
위원회	1052	인민	1963	나라	3354	민족	3749	나라	4497	나라	5052	애하	6358	중앙	4948	국가	2048	사람	7543	7	33 주체
중앙	1021	혁명	1923	사람	3221	혁명	3330	민족	4470	혁명	5007	평양	6229	사람	4733	건설	1965	통신	7499	24	36 장군
인민	974	국가	1919	김성	3025	사람	3301	중앙	4220	사람	4849	위원회	6225	사업	4615	인민	1886	나라	7450	6	령도자
혁명	958	나라	1751	중앙	2973	나라	3194	박근혜	3826	청년	4828	중앙	6134	평양	4212	김정은	1694	위원회	7080	9	43 원수
김정	892	통신	1667	괴뢰	2879	김정은	3165	사람	3565	위원회	4797	나라	6057	위원회	4169	기술	1626	지역	5762	38	47 공장
나라	854	김정은	1319	조국	2767	애하	2925	김정은	3514	평양	4772	국가	5900	나라	3993	나라	1562	대통령	5620	34	54 사랑
위성	769	최고	1169	민족	2743	중앙	2863	김성	3511	세계	4509	민족	5525	김정은	3652	지역	1533	김정	5334	25	괴뢰
세계	726	사회주의	1158	혁명	2715	국가	2773	사회주의	3500	군사	4395	수령	5078	사회주의	3520	조선	1516	건설	5250	13	56 성과
건설	710	세계	1144	승리	2457	당국	2673	국가	3443	전략	4120	혁명	5022	조선	3489	발전	1377	발전	4851	26	60 전략
사회주의	655	회의	1090	세계	2333	세계	2370	위원회	3264	국가	4080	건설	4969	주체	3357	사업	1271	전쟁	4817	15	63 해방
결의	607	문제	1032	위원회	2321	위원회	2271	투쟁	3235	중앙	4052	사회주의	4758	국가	3267	혁명	1259	김성	4585	20	64 학생
발사	606	선거	1026	통신	2242	원수	2046	건설	3212	평화	4001	사람	4640	기술	3250	진행	1215	당국	4547	44	과학
통신	602	건설	992	국가	2232	수령	2042	세계	3025	민족	3887	김성	4560	진행	3204	새로운	1208	세계	4232	17	침략
평화	552	전쟁	984	김정은	2147	문제	2023	애하	2970	최고	3883	평화	4181	발전	3053	사회주의	1202	문제	4214	30	조선로동당
공화국	542	정책	925	대통령	2117	투쟁	1946	조국	2944	당국	3828	발전	3980	공장	2886	농장	1200	혁명	4190	4	조선반도
유엔	539	이데올로기	920	공화국	2070	사회주의	1863	평화	2925	수령	3545	력사	3811	수령	2756	공장	1148	국제	4055	42	령도자
전쟁	525	발전	895	해방	2067	건설	1822	전쟁	2921	투쟁	3533	진행	3780	조선로동당	2686	지방	1084	평화	3856	27	제재
제재	513	사회	886	애하	1937	조국	1815	공화국	2854	괴뢰	3454	대통령	3775	농장	2620	계획	1038	투쟁	3763	21	계획
준비	483	대통령	872	투쟁	1607	주체	1773	통신	2822	애하	3425	회의	3711	사랑	2497	국제	1037	회의	3690	35	대회
발전	476	평화	865	건설	1543	통신	1715	김정	2780	령도자	3380	투쟁	3477	투쟁	2456	농업	1017	정부	3680	57	높은
군대	469	인권	857	력사	1409	평화	1690	수령	2674	김정은	3203	자유	3442	원수	2418	과학	993	군사	3592	35	준비
국제	468	군사	847	인권	1398	전쟁	1663	배치	2613	조국	3180	공화국	3091	문제	2142	노동자	954	강조	3582	정신	침략
정책	461	요구	845	사회주의	1397	력사	1586	괴뢰	2612	승리	3067	생산	3039	장군	2102	애하	950	대표	3559	46	훈련
지역	460	김성	836	진행	1351	공화국	1585	군사	2452	진행	2969	선언	2928	조직	2094	부문	945	요구	3493	45	부문
강조	443	지역	826	원수	1350	김성	1551	지역	2416	로케트	2952	대표	2830	성과	2073	공무원	937	사회주의	3389	15	박근혜
김성	435	강조	819	김정	1338	사랑	1541	해방	2415	문제	2940	사업	2799	회의	2015	주체	913	관계	3247	59	보장
문제	427	진행	802	문제	1333	진행	1505	진행	2232	조선반도	2934	통신	2795	김성	1940	세계	893	진행	3238	23	해방
성명	409	김정	799	당국	1290	생산	1385	문제	2190	국제	2835	조국	2773	계획	1881	수행	888	민족	3190	12	위업
군사	407	생산	774	군대	1269	발전	1336	청소년	2152	발사	2806	세계	2757	수행	1858	축하	886	방문	3124	84	하나

김정은	애하
최고	청년
조국	장군
주체	령도자
장군	원수
령도자	공장
원수	사랑
공장	괴뢰
사랑	성과
괴뢰	전략
성과	해방
전략	학생
해방	과학
학생	침략
과학	조선로동당
침략	조선반도
조선로동당	령도자
조선반도	제재

- 북한 도발 14가지 사례와 평온 시기 북한 전체 뉴스 기사
- 상위 100개 단어의 빈도수 추출

- 도발 전 기간에서만 언급된 단어

# 인사이트 도출(북한 전체 사건 키워드)

- '조국', '주체'

→ 이는 집단 내부 결속력을 높이기 위한 모습

- '김정은'

→ 언급된 당시의 통치자 이기에 많이 언급

- '성과'

→ 도발 직전에 결과와 업적을 강조하는 보도가 증가

이와 더불어 북한은 '정권의 취약성'이 높은 수준에 있을 경우 이를 완화하기 위한 전략으로 대남군사도발을 선택한다. 북한 '정권의 취약성'은 정권의 이념적·실질적 정통성이 취약하고 지배엘리트와 주민들이 정권에 대해 이완되어 있을 경우 높은 수준에 놓이게 되는데, 이러한 상황에서 북한의 대남군사도발은 세 가지 측면에서 정권의 취약성을 완화하는데 효과적 수단이 된다. 첫째로 '선군사상'을 중심으로 '강성대국 건설'을 이뤄야 하는 북한에게 대남군사도발은 취약한 이념적·실질적 정통성을 극복할 수 있는 실제적 업적이 될 수 있다. 둘째로 '외부의 적과 위협'에 대한 실질적 대응인 대남군사도발을 통해 내부적으로 지배엘리트들의 단결을 도모함과 동시에 상대적으로 위험감수적인 정책인 군사도발을 통해 충성스러운 엘리트와 정권에 이완되어 있는 엘리트를 가려냄으로서 추후 발생할 수 있는 내부 분열도 방지할 수 있다. 마지막으로 북한은 대남군사도발을 통해 경제난 등의 북한내부문제로 인한 주민들은 높은 불만을 외부위협으로 분산시킴으로써 정권에 대한 사회결속을 도모하여 정권의 취약성을 완화할 수 있다.

다양한 대내외 요인에 의해 발생하는 것으로 첫째, 주요 도발의 요인은 체제결속·내부 안정성 증대의 필요성 때문이다. 북한은 탈냉전 이래 만성적인 식량난과 에너지난에 의한 민생고와 함께 지배층의 체제 불안감 증대, 외부 사조 유입과 사회주의 일탈 현상의 만연 등 체제의 이완 현상이 증대해 왔다. 따라서 핵·미사일 도발과 열병식 등을 통해 북한 외부 세력의 위협에 대한 지배층과 주민들의 불안 심리를 잠재우고 탈북을 방지함으로써 체제결속과 김정은 정권의 안정화를 나타낼 필요가 있었다.

# 향후 개선 사항

- 관련국 뉴스까지 수집 시 도발 징후 분석 및 국제 정세 파악에 더 정밀한 접근 가능
- 기사 외에도 정부 발표문, 국제기구 보고서 등 공식 문서 포함 고려
- 워드클라우드 정제를 위한 자동 불용어 필터링 알고리즘 개발
- 사이버 공격 등 새로운 도발 유형 확장 가능성 검토

## 기대효과

- 북한 도발 징후 조기 포착을 통한 대응 전략 수립에 기여
- 국방·안보 교육 및 국제관계 연구 자료로 활용 가능
- 향후 도발 예측 및 정책 수립을 위한 데이터 기반 마련

# 4. 개발후기

# 개발후기

## 어려웠던 점

- 크롤링해야 할 뉴스 기사 양이 많아 전체 수집에 시간이 오래 걸림
- 워드클라우드 및 키워드 추출 과정에서 다수의 파일을 반복 처리해야 해 작업량이 많았음
- 초기에 통일부 뉴스 데이터를 사용했으나, 제목만 제공되고 본문이 없어 분석에 한계가 있었음  
→ 기존에 사용했던 북한기사 제공 사이트를 활용하여 재수집함

## 좋았던 점

- 이번 프로젝트를 통해 북한의 도발 실태에 대해 더 자세히 알게 됨
- 수업 때 배운 내용을 더 많은 사이트들과 데이터에 활용할 수 있었음
- 조원들과 데이터 수집부터 시각화, 대시보드 구현까지 함께 전과정을 해볼 수 있었음

# 질의응답

감사합니다