

2023 새만금 공공·빅데이터 활용 공모전

떠나보새만금

지역문화 활성화와 관광 산업 발전을 위한 관광 상품 추천 및 정책 제안

불패경통 | 김해우 이은지 임형빈

INDEX

01. 주제 선정 배경

02. 분석 활용 데이터

03. 데이터 분석 방법

04. 분석 결과

05. 활용방안

01

주제 선정 배경

주제 선정 배경

프로젝트 배경 및 주제

관광 도시로서의 충분한 경쟁력

전북 새만금 '이차전지 특화단지 지정'...이차전지 핵심기지로 도약

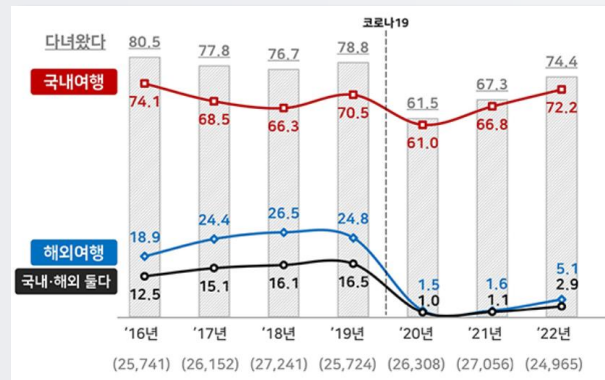
새만금 땅·바다·하늘 연결하는 교통·물류 기반시설 건설 착착 진행 중

기사입력 2023-05-27 24:00:19

LS그룹, 새만금에 1조8천400억 투자...이차전지 소재 공장 건립

- 최신 공업 단지 형성으로 인프라 구축
- 세계 최장 길이의 방조제와 바다를 포함한 자연 경관을 갖춘

코로나 이후 국내 여행 수요 증가



- 관광객 유치 → 지역 경제 활성화



새만금 지역문화 활성화와 관광 산업 발전을 위한 **관광 상품 추천 및 정책 제안**

주제 선정 배경

전체 로드맵

프로젝트 배경 및 목적

관광 도시로서의 충분한 경쟁력
코로나 이후 국내 여행 수요 증가



관광객 유치로 지역 경제 활성화



새만금 지역의
관광 상품 추천 및 정책 제언

1차 크롤링 및 분석



관광 지역 + 여행
키워드로 25개 지역
게시물 1000개 크롤링



2차 크롤링 및 분석



관광 도시 별 관광 상품 6개 + 여행
키워드로 26개 지역
총 5000개 이상의 게시물 내용 추출



자연어 처리
데이터 전처리
군집화



클러스터 결과

활용방안

기대효과

02

분석 활용 데이터

분석 활용 데이터

(1) 활용한 공공데이터 및 민간 데이터

공공 데이터

- ✓ 새만금개발청_새만금지역 축제현황_20220818.csv
- ✓ 한국관광데이터랩_지역별 AI 관광 분석 데이터
- ✓ 한국관광데이터랩_전국 관광지검색순위

민간 데이터

- ✓ 관광 도시 25개 키워드로 네이버 블로그 데이터 크롤링 → 10만 개 이상의 데이터 수집

03

데이터 분석 방법

데이터 분석 방법

데이터 1차 크롤링



한국관광데이터랩

- 유명 관광 도시 25개 추출
- 지역별 1000개의 게시물 크롤링

< 크롤링 검색어 예시 >

강릉
양양
수원
속초
전주
경주
⋮

+ 여행



네이버 블로그

게시물 제목, 내용, 링크 추출

각 지역 당 약 1000개 데이터 크롤링 후
데이터프레임 작성



인스타그램

게시물 본문 내용, 해시태그, 작성일자 추출

각 지역 당 약 1000개 데이터 크롤링 후
데이터프레임 작성

데이터 1차 크롤링

blog 네이버 블로그

	title	content	link
0	청주 가볼만한곳 청남대 : 과거 대통령 별장, 현재는 대통령기념관 및 산책코스 (...	청주 가볼만한곳 청남대와거 대통령 별장,현재는 대통령기념관 및 산책코스(+ 입장료)...	https://blog.naver.com/ellasha123/223169599078
1	따뜻한 남쪽의 청남대/대통령 기념관 : 네이버 블로그	청남대/대통령 기념관충북 청주시 상당구 문의면 청남대길 646청남대 본관은 1993...	https://blog.naver.com/parkbg0714/223160811307
2	청남대 봄기운 가득, 축제로 흥겨운 가득 : 네이버 블로그	지난 주말 영춘제가 열리고 있는 청남대를 다녀왔습니다. 영춘제는 봄을 맞이한다는 의...	https://blog.naver.com/alice8565/223088318794
3	청주 청남대 맛집 부농장 : 네이버 블로그	안나페어 붐어예요여기 제가 엄청 어려울 때 비닐하우스? 같은데서먹은 기억이 있는데 ...	https://blog.naver.com/shmwlove6/223066757792

[illegible]

데이터 1차 분석

```
1 import copy
2
3 with open('stopwords.txt', 'r', encoding='utf-8') as f:
4     stopwords = f.readlines()
5     stopwords = [x.replace('\n', '') for x in stopwords] # stopword 파일의 줄바꿈 문자 제거
6     okt = Okt()
7
8 def preprocess_korean(text):
9     my_text = copy.copy(text)
10    my_text = my_text.replace('\n', '')
11    spacer = pykospacing.Spacing()
12    my_text = spacer(my_text) # 띄어쓰기 교정
13    sents = kss.split_sentences(my_text) # sentence tokenization
14
15    p = re.compile('[^ㄱ-ㅎㅌ-ㅣ가-힣]*')
16    results = []
17    for sent in sents:
18        result = []
19        tokens = okt.morphs(sent, stem=True) # word tokenization
20        for token in tokens:
21            token = p.sub('', token) # 특수문자 제거
22            if token not in stopwords:
23                result.append(token) # stopword removal
24        results.extend(result)
25    result = ' '.join(results)
26
27    return result
28
```

분석 방법

네이버 오픈 API, 구글 크롬 드라이버, 파이썬 셀레니엄 사용

자연어 처리

- (1) 정규표현식 → 한글을 제외한 숫자·영어·특수문자 제거
- (2) 불용어처리 및 띄어쓰기 교정 후 토큰화 진행
ex. 고군산 군도 → 고군산군도
- (3) 명사만 추출

데이터 1차 분석

[illegible]

맛집, 카페

리조트, 펜션, 호텔, 터미널

갯벌, 크루즈, 생태체험

영화세트, 산, 바다

데이터 1차 분석

분석 결과

- 대부분의 도시들을 특정 테마로 묶을 수 있음을 확인했다.
- 특정 지역 명 + 인접한 다른 지역들의 언급 비중 ↑
- 인스타그램 → 트렌드 파악이 어려운 포괄적인 해시태그가 대부분
- 블로그 크롤링 결과가 더 유의미하다고 판단

데이터 분석 방법

데이터 2차 크롤링



한국관광데이터랩

유명 관광 도시 26개

관광 도시 별 대표 인기 관광 상품 6개 선별

< 크롤링 검색어 예시 >

인기 관광 상품

- 4. 인천광역시(완)
 - 월미도
 - 울왕리해수욕장
 - 소래포구종합어시장
 - 송도센트럴파크
 - 동막해수욕장
 - 인천대공원
- 5. 광주광역시(완)
 - 무등산 국립공원
 - 광주 시립 민속 박물관
 - 5·18 기념공원
 - 역사문화마을
 - 송정역시장
 - 광주기아챔피언스필드
- 6. 대전광역시(완)
 - 대전오월드
 - 유성온천지구
 - 장태산 자연 휴양림
 - 대전 엑스포 과학공원
 - 엑스포 아쿠아리움
 - 한밭수목원

+ 여행



네이버 블로그

게시물 제목, 내용, 링크 추출

약 150개의 관광 키워드 별
1000개의 게시글 크롤링

✓ 총 5000개 이상의 게시글 내용 추출
(전처리 방식은 앞의 방법과 동일)

데이터 분석 방법

데이터 2차 분석

지역별 언급 빈도 상위 30개의 키워드를 추출

	게이볼 카	공원	백화	바다	전망 대	풍경	카페	스카이라 인	바위	자연	...	문화역사체험 관	동식물 원	항구	과학기술체험 관	산	바다액티비 티	강	액티비 티	상물	도리
tongyoung	0.13	0.10	0.08	0.07	0.04	0.03	0.03	0.03	0.04	0.02	...	0.00	0.00	0.00	0.00	0.15	0.01	0.00	0.16	0.04	0.00
jeju	0.00	0.00	0.00	0.09	0.00	0.03	0.07	0.00	0.00	0.03	...	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.03	0.00
siheung	0.00	0.20	0.00	0.04	0.03	0.02	0.04	0.00	0.00	0.00	...	0.08	0.00	0.00	0.00	0.00	0.14	0.00	0.04	0.00	0.00
seoul	0.00	0.18	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	...	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17
seogwipo	0.00	0.00	0.00	0.05	0.02	0.03	0.06	0.00	0.19	0.02	...	0.00	0.00	0.00	0.00	0.14	0.00	0.03	0.00	0.00	0.00
pohang	0.00	0.03	0.00	0.07	0.02	0.02	0.05	0.00	0.03	0.02	...	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.00
paju	0.00	0.11	0.00	0.00	0.03	0.03	0.10	0.00	0.00	0.00	...	0.02	0.10	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00
mokpo	0.06	0.04	0.00	0.05	0.04	0.03	0.04	0.00	0.03	0.00	...	0.07	0.00	0.01	0.00	0.00	0.00	0.00	0.12	0.00	0.00
taean	0.00	0.00	0.00	0.07	0.00	0.00	0.03	0.00	0.02	0.00	...	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
gyungju	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.00	0.00	0.00	...	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00
chungju	0.00	0.03	0.03	0.00	0.05	0.02	0.06	0.00	0.00	0.00	...	0.02	0.07	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00
changwon	0.00	0.12	0.00	0.02	0.00	0.00	0.02	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.04
bosung	0.00	0.03	0.00	0.04	0.03	0.03	0.02	0.00	0.00	0.02	...	0.02	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.04	0.07
buan	0.00	0.01	0.00	0.08	0.00	0.02	0.03	0.00	0.00	0.00	...	0.00	0.00	0.09	0.00	0.00	0.00	0.13	0.00	0.02	0.00
hwasung	0.04	0.00	0.00	0.07	0.00	0.02	0.06	0.00	0.02	0.00	...	0.00	0.00	0.13	0.00	0.00	0.02	0.00	0.00	0.00	0.00
yeosu	0.06	0.09	0.00	0.09	0.02	0.02	0.05	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.04	0.00
gangneung	0.00	0.00	0.00	0.06	0.00	0.00	0.06	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
gunsan	0.00	0.06	0.00	0.05	0.00	0.03	0.05	0.00	0.02	0.00	...	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00
yangyang	0.02	0.03	0.00	0.05	0.00	0.00	0.03	0.00	0.02	0.00	...	0.00	0.00	0.00	0.01	0.07	0.00	0.00	0.03	0.00	0.00
ulsan	0.00	0.11	0.00	0.08	0.00	0.02	0.04	0.00	0.22	0.00	...	0.06	0.00	0.00	0.00	0.00	0.08	0.00	0.01	0.00	0.01
gwangju	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
yongin	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.02	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.24	0.00	0.05
daejeon	0.00	0.05	0.00	0.00	0.02	0.00	0.02	0.00	0.00	0.00	...	0.00	0.10	0.00	0.12	0.07	0.00	0.00	0.02	0.02	0.08
daegu	0.09	0.02	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	...	0.00	0.07	0.00	0.07	0.10	0.00	0.00	0.02	0.00	0.02
namyangju	0.00	0.32	0.00	0.00	0.00	0.04	0.03	0.00	0.00	0.02	...	0.05	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.10	0.00
incheon	0.00	0.03	0.00	0.11	0.00	0.00	0.06	0.00	0.02	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

→ 인덱스를 지역 이름으로 갖는 통합 데이터프레임 작성

데이터 전처리

- ✓ 결측치 → 지역에서 해당 키워드의 언급X → 0으로 대체
- ✓ 행 별로 합이 1이 되도록 정규화

군집화

- ✓ kmeans clustering 진행

```
from sklearn.cluster import KMeans
km = KMeans(n_clusters = num, random_state=42)
km.fit(keyword_ratio)
```

데이터 2차 분석

클러스터 개수 선정

- ✓ 군집 개수를 조절해보며 모델링 실험
- ✓ 군집이 9개일 때 가시적으로 군집 테마 분석이 용이

0	
사찰	0.11
문화역사체험관	0.10
첨성대	0.09
동굴	0.08
석굴암	0.06

1	
동식물원	0.0600
과학기술체험관	0.0475
카페	0.0475
공원	0.0475
산	0.0425

2	
음식	0.45
야구장	0.29
시장	0.14
경기장	0.09
카페	0.03

3	
공원	0.233333
바다엑티비티	0.060000
문화	0.056667
문화역사체험관	0.050000
체험	0.046667

4	
음식	0.130000
해수욕장	0.113333
해산물	0.096667
시장	0.093333
바다	0.073333

5	
엑티비티	0.126
공원	0.074
체험	0.060
카페	0.044
호수	0.044

6	
녹차	0.31
문화	0.07
해수욕장	0.07
음식	0.06
체험	0.05

7	
해수욕장	0.130000
바다	0.071667
카페	0.048333
사찰	0.046667
동굴	0.043333

8	
바위	0.205
해수욕장	0.080
산	0.070
바다	0.065
공원	0.055

04

분석 결과

분석 결과

클러스터 결과

Cluster 1

역사체험

Cluster 2

동물, 수목원, 과학 체험관

Cluster 3

스포츠 직관, 맛집

Cluster 4

예술 문화 생활

Cluster 5

바다 관광

Cluster 6

내륙 액티비티 및 생태 체험

Cluster 7

지역 특산물

Cluster 8

바다 힐링, 지역 문화 체험

Cluster 9

자연 경관, 힐링

분석 결과

클러스터 결과

✓ 관광지들이 해당 군집에 맞게 분류됨

```
keyword_ratio[keyword_ratio['cluster'] == 0 ].index  
Index(['gyungju'], dtype='object')
```

```
keyword_ratio[keyword_ratio['cluster'] == 4 ].index  
Index(['gangreung', 'yangyang', 'incheon'], dtype='object')
```

- ✓ 전국 유명 관광지들의 여행 테마 별 인기 관광 트렌드 파악
- ✓ 새만금 지역의 특성을 살릴 수 있는 관광 상품 제안

05

최종 인사이트 및 활용 방안

최종 인사이트 및 활용 방안

활용 방안

Cluster 1 : 역사 체험

- ✓ **청자박물관**
→ 부안 인기 관광지 및 중심 관광지 각 30, 21위
- ✓ 새만금의 역사 축제 + 부안 청자 박물관
- ✓ 청자박물관 연관 관광지 → 여행 코스 구성 적합

부안청자박물관 연관 관광지 ⓘ 총 50건			
전체 21 22 23 24 ...			
순위	이름	주소	유형
1	슬지제병소	전라북도 부안군	음식
2	내소사	전라북도 부안군	역사관광
3	소노벨/변산	전라북도 부안군	숙박
4	채석강	전라북도 부안군	기타관광
5	부안누에타운	전라북도 부안군	기타관광



제안 상품

**청자박물관을 중심으로 한
청자 축제 개최**

최종 인사이트 및 활용 방안

활용 방안

Cluster 2 : 동물, 수목원, 과학 체험관

- ✓ 전시 혹은 박람회를 개최하여 **다른 명소와 연계**
- ✓ 과학테마 파크와 같은 관광 상품 개발 시,
국제적인 관심을 받아 **외국 관광객 유치 기여도↑**
- ✓ **이차전지 분야의 중심지**로서 이미지 강화



제안 상품
이차전지 국립과학관

최종 인사이트 및 활용 방안

활용 방안

Cluster 3: 스포츠 직관, 맛집

- ✓ 광주 여행: 야구 언급량 ↑
- ✓ 새만금 지역의 해양 특성을 활용한
해양 스포츠 및 액티비티 활성화
- ✓ 해양 레저 스포츠(4군집)
→ 단순 체험에서 **축제 형태로 확대**



제안 상품

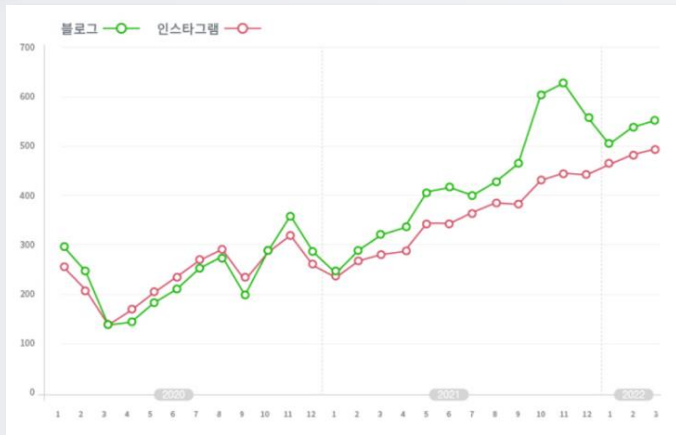
해양 스포츠와 액티비티 축제

최종 인사이트 및 활용 방안

활용 방안

Cluster 4 : 예술 문화 생활

- ✓ 블로그와 인스타그램에서 전시회 언급량 2배↑



- ✓ 다양한 전시 형태의 관광 상품 → 젊은 층의 관심↑
- ✓ 전시회 방문 사진의 SNS 공유 → 홍보 효과



제안 상품
미디어 아트 전시관

최종 인사이트 및 활용 방안

활용 방안

Cluster 5 : 바다 관광

- ✓ 운동 및 피트니스 앱의 월간 사용자 수의 꾸준한 증가
- ✓ 해양레저장비, 서핑, 요트 관련 업종 매출 증가



제안 상품

해양 레저 스포츠, 요트 투어
(서핑, 패러세일링과 제트스키)

최종 인사이트 및 활용 방안

활용 방안

Cluster 6 : 내륙 액티비티 및 생태 체험

- ✓ 현재 운영중인 새만금 지역의 환경생태단지 관광 상품
+ 철새 먹이터 조성, 철새조망대
- ✓ 관광 측면 : 철새 도래지로 관광객 유입 증가
- ✓ 환경 측면 : 다양한 생물종 서식지 보호에 관심 증가



제안 상품
환경생태단지 활성화

최종 인사이트 및 활용 방안

활용 방안

Cluster 7 : 지역 특산물, 녹차밭

- ✓ 보성 : 산악 배경, 녹차밭 → 녹차 : 주요 특산물
- ✓ 새만금 간척지 → 농경지로 활용 예정
- ✓ 새만금에서 과거 보리 관련 축제 다수 진행
- ✓ 단기 행사가 아닌 고정 관광상품으로 활성화



제안 상품
전망대 건설
(보리밭, 일몰 감상 목적)

최종 인사이트 및 활용 방안

활용 방안

Cluster 8 : 바다 힐링, 지역 문화 체험

✓ **새만금 자전거 도로 및 무인대여시스템 구축**

: 관광객의 편의를 위하여 여러 위치에
자전거 무인대여시스템을 구축

✓ **다양한 라이딩 행사 개최**

: 라이딩 행사를 주기적으로 개최
→ 관광객들에게 힐링의 기회와 여유로운 시간을 제공



제안 상품

라이딩 관련 관광상품

최종 인사이트 및 활용 방안

활용 방안

Cluster 9 : 자연 경관, 힐링

✓ **교통편 추가 지원**

: 새만금의 선유도 무녀도 등의 섬을 추가로 관광
할 수 있도록 접근성을 늘린다.

✓ **공원과 같이 사람들이 머물 수 있는 장소 제공**



제안 상품

**교통편 추가 지원,
대형 카페 및 공원 시설 유치**

최종 인사이트 및 활용 방안

기대 효과



여러가지 관광 시설 설립

→ 지역 일자리 창출 및 경제 활성화 유도



크롤링 기법으로 가장 최신의 트렌드 반영

→ 계절에 상관없이 일반화된 트렌드 파악 가능

참고 자료 및 출처

- 1 지역별 AI 관광 분석 데이터
(관광 데이터랩)
- 2 시각에서 체감으로 변화한 전시회 그래프
(바이브컴퍼니 생활변화관측소)
- 3 2016~22년 국내 및 해외 여름휴가 여행률 추이
(컨슈머인사이트 '연례 여름휴가만족도 조사')
- 4 해양레저장비 시장 규모 그래프
(KB국민카드)