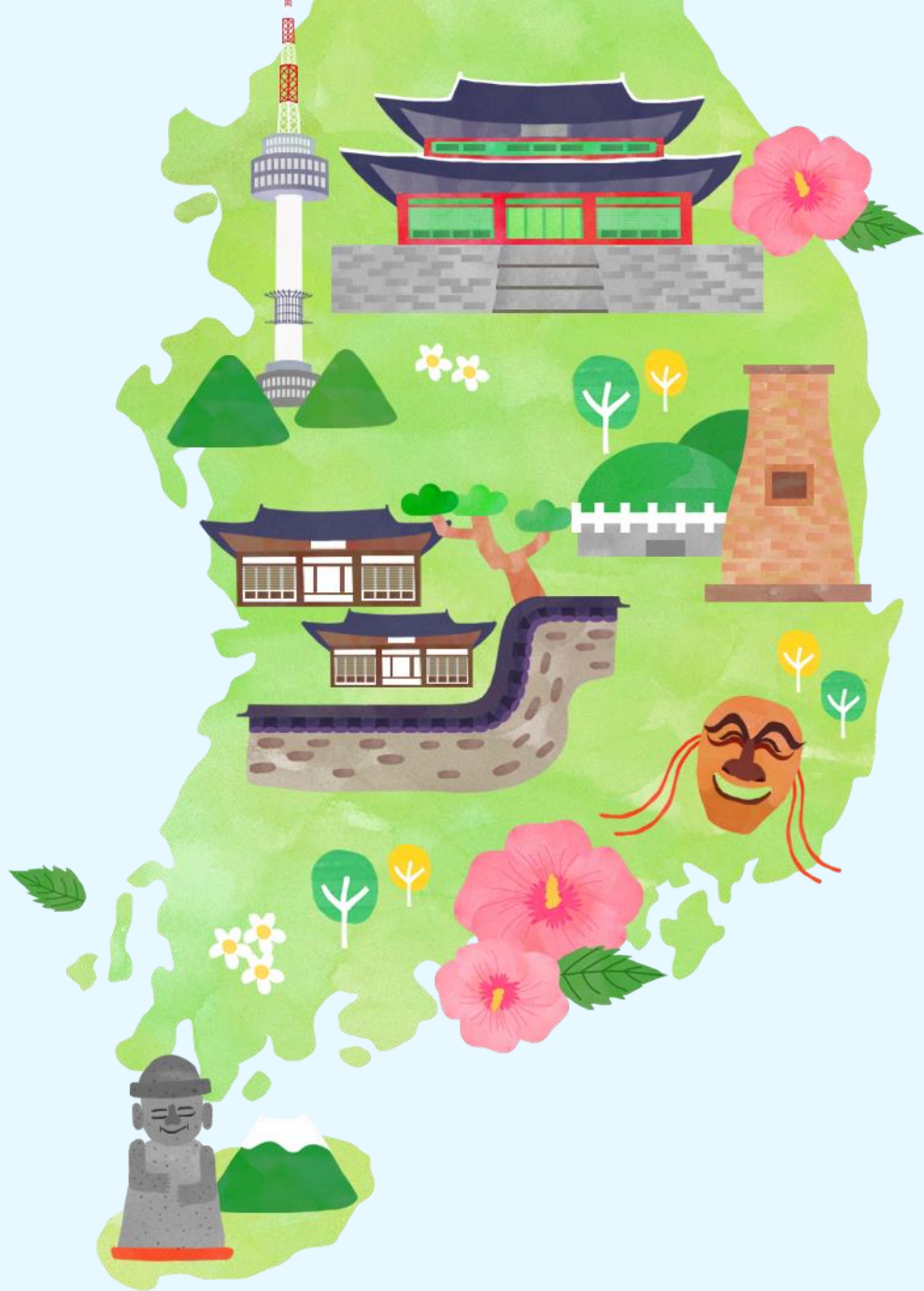


서울 말고 다른 곳은 어때?

- 외국인들을 위한 뉴노멀 분산 관광 제언





Contents

1. 주제 선정 및 분석 방향

2. 전체 모델링 구조

3. 활용 데이터

4. 데이터 모델링

5. 숨은 명소 지수

6. 결론 및 활용 방안

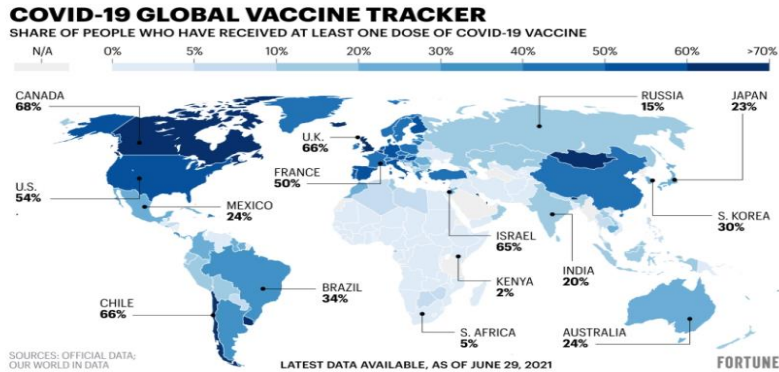
1. 주제 선정 배경



1. 세계적으로 해외여행 본격화 준비

"백신 맞았으니 해외여행 가자"...여권 신청 폭증하는 미국

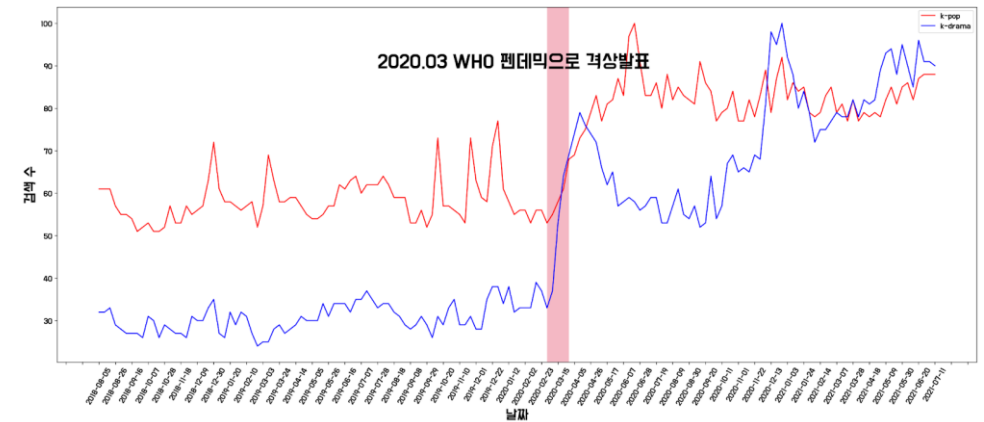
입력 2021.07.15 10:14 | 수정 2021.07.15 10:23



- ✓ 올해 초, 백신 개발 이후 많은 국가들이 일상 복귀를 위해 백신 접종 가속화 진행중
- ✓ 세계적으로 백신 접종률이 높아짐에 따라 각 나라별로 국내/해외 관광산업 활성화 추진중

2. 코로나를 극복하는 문화적 백신 '한류'

k-pop 및 k-drama 검색 건 수(google trend)



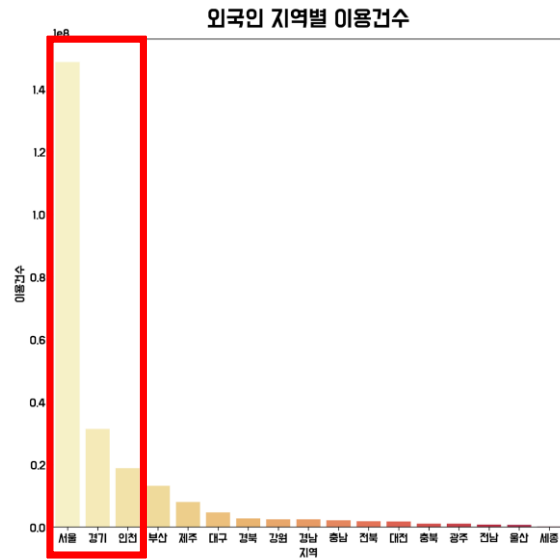
- ✓ WHO의 팬데믹 격상발표(20.03.11) 이후 한류 관련 콘텐츠 검색 건수 급상승
- ✓ 올해 방탄소년단(BTS)의 빌보드 1위, 윤여정의 아카데미 시상식 수상, NETFLIX의 한국 드라마 인기 등으로 한국에 대한 관심도 증가

코로나 19가 종식되지 않았지만 국가 간 여행이 점차 회복됨과
한국에 대한 관심도가 높아짐에 따라 외래 관광객 유입이 높아질 전망

1. 주제 선정 배경



3. 수도권 중심의 오버 투어리즘 문제



- ✓ 방한 외래 관광객의 약 82%가 서울, 경기, 인천(수도권)에 집중적으로 분포
- ✓ 수도권 중심의 오버 투어리즘으로 인해 수도권을 제외한 지역의 관광 산업은 침체 진행중

4. 코로나 시국 수도권 위주 관광의 문제점

코로나19 신규확진 수도권 비율 60% '대책은?'

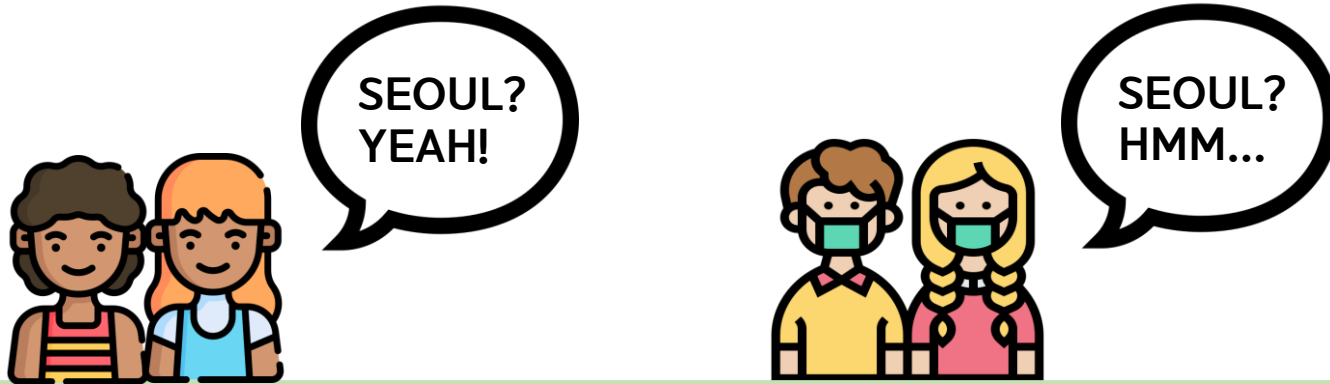
2021/07/27 08:41 송고



- ✓ 높은 밀집도는 코로나19 감염의 주요 원인
- ✓ 인구가 밀집된 수도권 중심의 관광은 팬데믹 상황에서 코로나 19 감염 위험을 높일 뿐 아니라 안전한 여행을 보장해주지 못함

이러한 흐름에 발맞추어 외국인들의 한국행을 대비하는 안전 + 분산 관광 정책 필요

1. 분석 방향



유명 명소와 비슷한 특징을 가지면서 숨어있는 명소들을 발굴하여 외국인 관광객 유도

예시)

외국인 관광객들의 주된 한국 문화 체험 장소 : 서울 종로구 북촌 한옥마을

마찬가지로 한국 문화를 체험할 수 있지만 외국인 관광객들은 잘 모르는 지역
: 안동 하회마을이나 거창 한옥마을

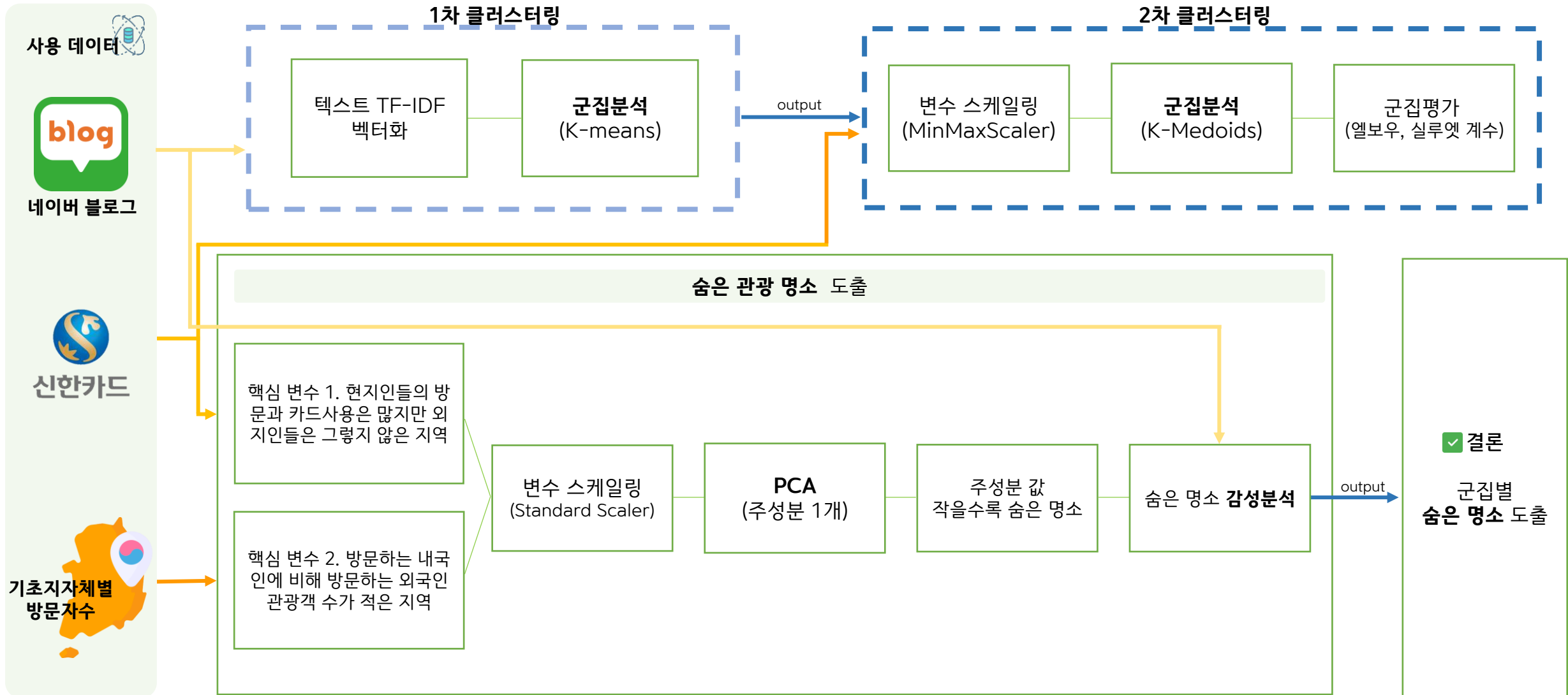
분석 방향 : 지역 특성으로 군집화 한 후 숨은 명소 도출



2. 전체 모델링 구조



모델링 목표 : 지역별 관광 특성 알아보기 -> 유사한 특징을 가진 지역끼리 묶기



3. 활용 데이터



네이버 블로그

기간	2019년~2021년
군집화	비슷한 특성을 가진 지역 클러스터링
숨은 명소	감성분석을 통한 여행 만족도 조사
방법	K-Means clustering/ 감성분석

신한카드

기간	군집화: 전체 숨은 명소: 2019년
군집화	업종별 매출액 이용해 지역 클러스터링
숨은 명소	카드 이용 건수를 이용해 숨은 관광 명소 찾기
방법	K-Medoids clustering/ 데이터 전처리

기초지자체별 방문자수

기간	숨은 명소: 2019년
군집화	사용 안함
숨은 명소	외국인/내국인 방문자 수 추출해 숨은 관광 명소 찾기
방법	데이터 전처리

3. 활용 데이터

신한카드 데이터 (내국인/외국인)

변수명	변수 설명	비고
V1	거주지	내국인 (외국인: 202007이후 V2로 수정)
V2	카드 이용 지역	공통
V3	카드 이용 지역	공통
GB3	업종 대분류	내국인 : 4개 분류 외국인: 7개 분류
GB2	업종 소분류	내국인 : 28개 분류 외국인: 51개 분류
...	...	
VLM	취급액(단위:원)	공통
USEC	이용건수(단위:건)	공통

추가한 변수

변수명	변수 설명	비고	
		내국인	외국인
V4	카드 이용지역 (229개 지역)	V2+V3 (202007 이전 데이터) V1+V2 (202007 이후 데이터)	V2+V3
GB_NEW	업종 새 분류 (8개 업종)	쇼핑, 숙박, 스포츠, 음식, 의료및미용 체험, 문화, 관광으로 새로 분류	



변수명	변수 설명	비고
V4	이용 지역	기초지자체 단위 (229개 지역)
S1 ~ S8	업종별 매출액	S1 = 쇼핑, S2 = 숙박, S3 = 스포츠, S4 = 음식, S5 = 의료및미용 S6 = 체험, S7 = 문화, S8 = 관광

3. 활용 데이터

기초 지자체 방문자수 데이터



전처리한 데이터

변수명	변수 설명	비고
지역	지역	251개 시군구
구분	방문자 특징	현지인/외지인/외국인 외부 방문자 수/전년외부 방문자 수 외부 방문자 수 전년대비 증감률
2019	2019년 방문자 수	공통

변수명	변수 설명	비고
지역	방문 지역	기초지자체 단위 (229개 지역)
외국인	외국인 방문자 수	19년 외국인 방문자 수 합계
외부 방문자	외부 방문자 수	19년 외부 방문자 수 합계
외지인	외지인 방문자 수	19년 외지인 방문자 수 합계
현지인	현지인 방문자 수	19년 현지인 방문자 수 합계

4. 데이터 모델링 - 1차 클러스터링



STEP 1. 데이터 수집



```
def naver_api(where, what, how):  
    encText = urllib.parse.quote(what)  
  
    for i in range(1,1001,100):  
        url = "https://openapi.naver.com/v1/search/" + where + "?query=" + encText + "&display=100" + '&start=' + str(i) + "&sort=sim"  
  
        request = urllib.request.Request(url)  
        request.add_header("X-Naver-Client-Id", client_id)  
        request.add_header("X-Naver-Client-Secret", client_secret)  
        response = urllib.request.urlopen(request)  
        rescode = response.getcode()  
        if(rescode==200):  
            response_body = response.read()  
            results = response_body.decode('utf-8') # 사람 언어로 바꾸는 코드  
            # print(response_body.decode('utf-8'))  
            items = json.loads(results)['items'] # 딕셔너리로 바꿔주는 코드  
  
            for item in items:  
                f = open(how, 'a', encoding = "utf-8")  
                title = item['title']  
                description = item['description']
```

네이버 API Key를 발급받아 우리나라
기초지자체 229개의 '지역명+여행' 키워드를 검색

키워드 정확도(sim) 기준으로
기초지자체별(지역별) 1000개의
블로그 글 데이터 수집

4. 데이터 모델링 - 1차 클러스터링



STEP 2. 데이터 전처리



Kkma 형태소 분석기

- 1) 블로그 데이터: 블로그 글 특성상 맞춤법 오류 가능성 ↑
 - 2) Kkma 형태소 분석기: 가장 많은 형태소 범주로 분석가능,
띄어쓰기 오류에 덜 민감
- Kkma 형태소 분석기 적합 후 명사 추출

```
# 텍스트 데이터 전처리
import re
def clean_str(text):
    pattern = '([a-zA-Z0-9_+~]+@[a-zA-Z0-9-]+\.([a-zA-Z0-9-]+))' # E-mail 제거
    text = re.sub(pattern=pattern, repl='', string=text)
    pattern = '(http|ftp|https)://(?:[-\w.]+|(?%[\da-f]{2}))+' # URL 제거
    text = re.sub(pattern=pattern, repl='', string=text)
    pattern = '([ㄱ-ㅎㅏ-ㅣ]+)' # 한글 자음, 모음 제거
    text = re.sub(pattern=pattern, repl='', string=text)
    pattern = '<[^>*>' # HTML 태그 제거
    text = re.sub(pattern=pattern, repl='', string=text)
    pattern = '["\w\s]' # 특수기호 제거
    text = re.sub(pattern=pattern, repl='', string=text)
    pattern = '[0-9]+' # 숫자 제거
    text = re.sub(pattern=pattern, repl='', string=text)
    return text
```

불용어 처리

- 1) 특수문자, 숫자, 한국어 자/모음, 영어 제거
- 2) 고유명사인 지역명, 의미 없는 명사 불용어
사전에 추가 후 제거

4. 데이터 모델링 - 1차 클러스터링



STEP 3. 데이터 모델링

TF-IDF

모든 문서에 나타나는 단어보다
해당 문서 고유의 단어를
도드라지게 만들어주는 알고리즘



지역별 특성을 추출하여 유사한 지역끼리
군집화하기 위한 분석의 목적과 부합

```
from sklearn.feature_extraction.text import TfidfVectorizer  
tfidf_vectorizer = TfidfVectorizer() # TF-IDF 만드는 모델  
tfidf_vectorizer.fit(info) # 전체 문서인 info를 모델에 적용
```

	0	1	2	3	4	5	6	7	8	9	...
0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	...
1	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	...
2	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	...
3	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	...
4	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	...
...
224	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	...
225	0.0	0.0	0.0	0.0	0.0	0.010133	0.0	0.0	0.0	0.0	...
226	0.0	0.0	0.0	0.0	0.0	0.004905	0.0	0.0	0.0	0.0	...
227	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	...
228	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	...

“ 229개 지역별 명사 코퍼스에
대해 **TF-IDF 벡터화** 진행 ”
(229 x 119600 TF-IDF matrix 생성)

4. 데이터 모델링 - 1차 클러스터링



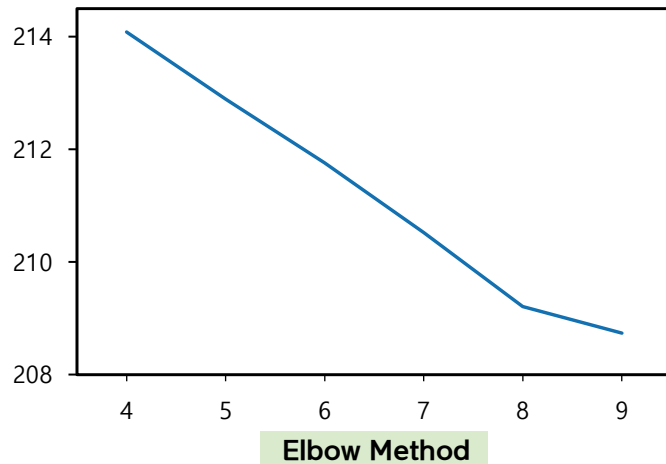
STEP 3. 데이터 모델링

K-Means

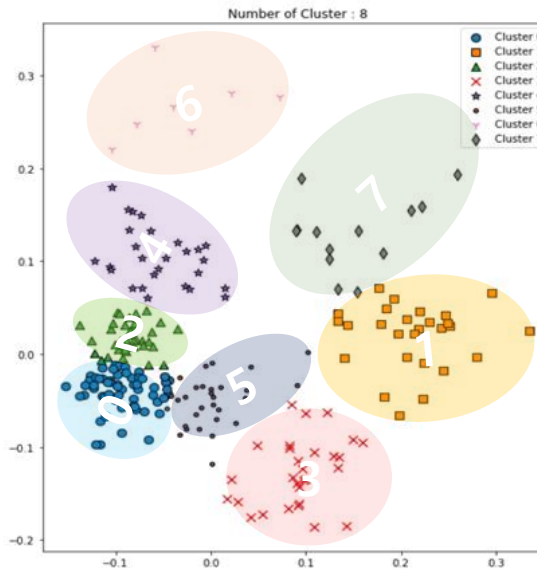
- 장점 : 범주에 대한 사전지식이 없어도 관측값 사이의 거리를 측정할 수 있음
- 단점 : 이상치에 민감함



불용어 처리를 진행하였기 때문에 이상치가 존재할 확률이 낮은 데이터이므로 K-Means로 분석 효과 극대화 가능



✓ Inertia 값을 기준으로 K-Means Cluster수 K=8로 결정



Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
경기 시흥시	경기 과천시	강원 양구군	경기 파주시	경기 부천시	경기 광명시	경북 안동시	강원 강릉시
경기 오산시	경남 창원군	강원 영월군	경남 거창군	경기 성남시	경기 화성시	서울 강남구	강원 고성군
부산 서구	전남 강진군	강원 원주시	경남 남해군	경북 포항시	경북 구미시	서울 강동구	강원 동해시
부산 중구	전남 무안군	경기 양평군	경북 김천시	부산 기장군	광주 광산구	서울 광진구	강원 삼척시
서울 중구	전남 순천시	경기 여주시	경북 봉화군	부산 해운대구	대구 중구	서울 서초구	강원 속초시
인천 강화군	전남 장성군	경기 연천군	경북 성주군	울산 북구	부산 사하구	서울 성동구	강원 양양군
인천 계양구	전북 부안군	경남 밀양시	전남 곡성군	제주 서귀포시	서울 서대문구	서울 성북구	경남 고성군
인천 남동구	전북 정읍시	경북 문경시	전남 구례군	제주 제주시	울산 중구	서울 송파구	경북 영덕군
인천 동구	충남 당진시	전남 신안군	전북 진안군	충남 보령시	전북 군산시	서울 용산구	경북 울릉군
인천 미추홀구	충남 서산시	충남 부여군	충남 계룡시	충남 태안군	충북 청주시	서울 은평구	경북 울진군

✓ K-Means Clustering (K=8) 결과

4. 데이터 모델링 - 2차 클러스터링



STEP 1. Mixed Type Data의 클러스터링

사용 데이터 : 1차 클러스터링 결과(범주형) + 신한카드 내국인 매출액(수치형)

✓ K-Medoids 방법

데이터의 중간값을 중심으로 K개의 cluster로 나누어 준다.
이상치에 robust하고, 데이터 실제값을 기반으로 하기 때문에 클러스터의 실제 대표 데이터를 살펴볼 수 있음

Gower Distance (by MinMaxScaler)

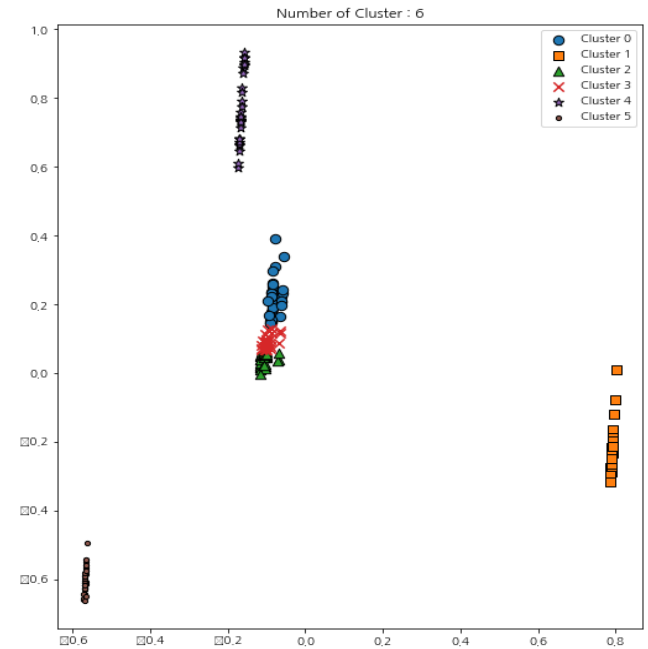
: K-Medoids Method의 Mixed Type Data의 스케일링

$$d(i, j) = \frac{1}{p} \sum_{i=1}^p d_{ij}^{(f)}$$

- Gower distance : 각 데이터의 the average of partial dissimilarities를 계산한 거리
수치형인 경우 [0,1]사이 값으로 변환, 범주형인 경우 0 또는 1

region	s1	s8	c1	c2	c7	c8		region	s1	s8	c1	c8
강원 강릉시	3.56E+10	7.54E+09	0	0	0	1	스케일링 →	강원 강릉시	0.039553	0.098351	0	1
강원 속초시	6.05E+10	...	0	0	...	1		강원 속초시	0.011541	...	0	1
강원 원주시	6.16E+09	3.56E+08	0	0	0	1		강원 원주시	0.114623	0.018476	0	1
강원 춘천시	1.4E+10	6.6E+08	0	0	0	1		강원 춘천시	0.298674	0.042447	0	1
...								...				

- s1 ~ s8 : 스포츠, 문화, 쇼핑, 의료 및 미용, 숙박, 관광, 음식, 체험
- c1 ~ c8 : 클러스터 0 ~ 7



✓ 1차 클러스터 c1 ~ c8 (원핫인코딩 형태) + 업종별 매출액 s1 ~ s8
(MinMaxScaler) 데이터에 Kmedoids 적용 후 2차원 그래프로 나타낸 결과 데이터가 극단적으로 나누어짐

4. 데이터 모델링 - 2차 클러스터링



STEP 2.

텍스트 데이터 클러스터링 결과를 원핫인코딩이 아닌 **소프트 군집화** 방법으로 보정
(데이터에서 각 클러스터까지의 거리로 표현)

region	s1	s8	c1	c2	c7	c8
강원 강릉시	3.56E+10	7.54E+09	1.010331	1.004398	1.021441	0.890139
강원 속초시	6.05E+10	38148695	1.028365	1.015917	1.035463	0.912191
강원 원주시	6.16E+09	3.56E+08	1.043952	1.030522	1.050999	0.870006
강원 춘천시	1.4E+10	6.6E+08	1.022731	1.010181	1.03268	0.889642

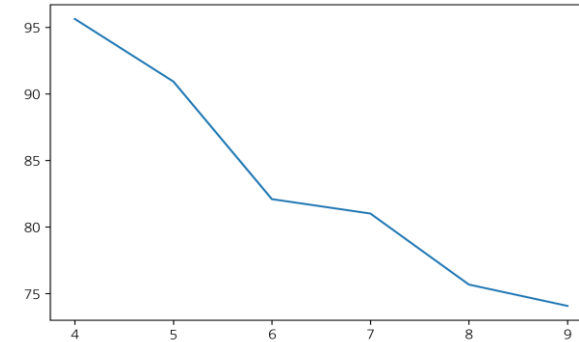


데이터 스케일링 (MinMaxScaler)
: Gower Distance

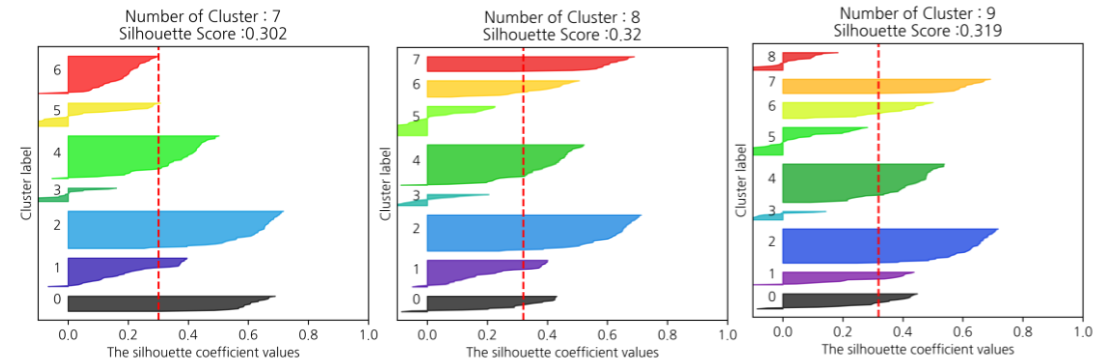
region	s1	s8	c1	c8
강원 강릉시	0.039553	0.098351	0.739059	0.098164
강원 속초시	0.067205	0.000265	0.849803	0.205686
강원 원주시	0.006822	0.004419	0.945516	0
강원 춘천시	0.015553	0.008397	0.815202	0.095742

이 데이터로 Kmedoids Clustering

STEP 3. Cluster 수(K)의 결정



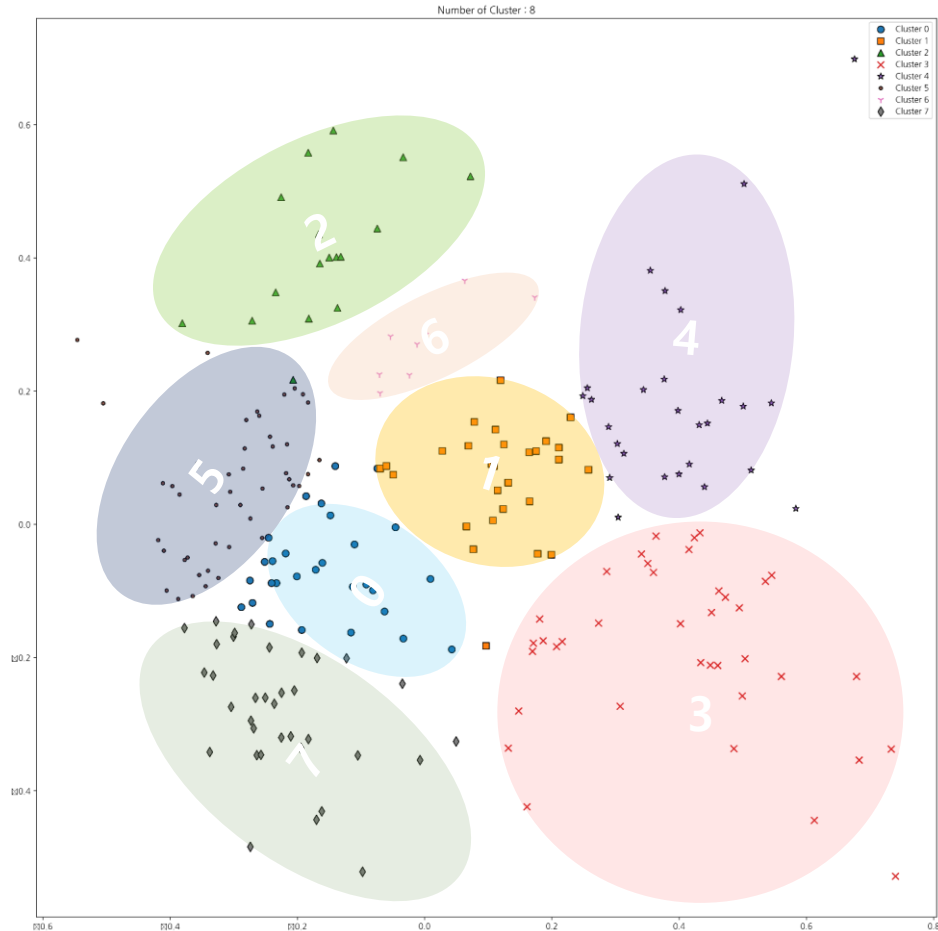
Elbow Method



Silhouette Coefficient Graph

✓ 실루엣 점수가 0.32로 가장 크고, Elbow를 형성하는 구간인 K=8을 클러스터 수로 정하였다.

4. 데이터 모델링 - 2차 클러스터링 결과



✓ Kmedoids Clustering (K=8) 결과

Cluster 0

울산 북구, 부산 동래구, 부산 연제구, 경북 포항시, 부산 사상구, 부산 남구, 부산 수영구, 부산 영도구, 부산 진구, 경남 거제시, **부산 해운대구**, 부산 서구, 부산 동구, **경남 통영시**, 부산 기장군, 제주 제주시, **충남 보령시**, 제주 서귀포시, 충남 태안군

Cluster 1

울산 중구, 경기 군포시, 경기 수원시, 경남 창원시, 경기 안양시, 경기 안산시, 충북 청주시, 광주 남구, 경기 동두천시, 경기 화성시, 경기 용인시, 경기 광주시, 경기 시흥시, 경기 구리시, 경기 광명시, 강원 동해시, 대전 중구, 울산 울주군, 경기 부천시, 경기 김포시, 대구 중구, 경남 사천시, 서울 서대문구, 경기 하남시, 전남 완도군, 경북 영덕군, 경기 성남시, 인천 용인군, 강원 양양군, 강원 고성군, 경북 울릉군, 전남 담양군

Cluster 2

충북 증평군, 전북 익산시, 전북 전주시, 경기 파주시, 전남 광양시, 경남 함안군, 경남 의령군, 전북 완주군, 경남 거창군, 경북 김천시, 경남 산청군, 경북 예천군, 경북 영천시, 경북 칠곡군, 경북 영양군, 전남 영암군, 충남 계룡시, 전남 고흥군, 경북 봉화군, 전남 장흥군, 경남 하동군, 전북 남원시, 경남 합천군, 충북 보은군, 전북 진안군, 전남 보성군, 경남 함양군, 경북 청송군, 경북 고령군, 전남 구례군, 경남 남해군, 충북 영동군, 전북 순창군, 충남 서천군, 전북 장수군, 경북 의성군, 전북 임실군, 충북 옥천군, 전남 곡성군, 경북 성주군, 충남 금산군, 경북 군위군, 전북 무주군

Cluster 3

대구 달성군, 전남 목포시, 경북 구미시, 경남 김해시, 대전 대덕구, 전남 여주시, 경남 고성군, 전북 군산시, 강원 삼척시, 강원 강릉시, 세종 세종시, 경북 상주시, 인천 강화군, 강원 속초시

Cluster 4

경기 의왕시, 경남 진주시, 경기 양주시, 경기 남양주시, 경기 고양시, 강원 태백시, 강원 춘천시, 강원 원주시, 경북 영주시, 서울 관악구, 경기 포천시, 경남 밀양시, 경기 안성시, 경남 양산시, 충북 진천군, 충북 제천시, 경기 연천군, 경기 이천시, 충북 충주시, 충남 아산시, 강원 영월군, 충남 천안시, 경기 여주시, 강원 양구군, 충남 논산시, 경북 경산시, 충북 음성군, 충남 예산군, 경기 양평군, 전남 나주시, 충남 청양군, 경북 경주시, 강원 화천군, 경북 울진군, 강원 철원군, 전남 신안군, 전남 진도군, 경기 가평군, 경북 문경시, 강원 정선군, 충남 부여군, 강원 횡성군, 강원 인제군, 강원 홍천군, 충북 단양군, 충남 공주시, 강원 평창군, 경북 청도군

Cluster 5

서울 도봉구, 인천 동구, 서울 강북구, 경기 오산시, 인천 남동구, 서울 은평구, 서울 중랑구, 서울 노원구, 경기 의정부시, 대구 남구, 인천 미추홀구, 경기 평택시, 서울 성북구, 부산 금정구, 서울 동대문구, 경북 안동시, 서울 성동구, 서울 양천구, 서울 광진구, 부산 중구, 서울 강동구, 서울 금천구, 부산 강서구, 서울 강서구, 부산 북구, 서울 서초구, 서울 송파구, 서울 영등포구, 서울 마포구, 인천 중구, 서울 강남구, 서울 구로구, 서울 동작구, 서울 용산구, 서울 중구

✓ 수도권, 대도시 위주의 클러스터이므로 **여행지 분산** 목표에 맞지 않아
추후 분석에서 Cluster 5는 제외

Cluster 6

울산 동구, 대구 서구, 대구 달서구, 부산 사하구, 울산 남구, 광주 북구, 광주 광산구, 인천 서구, 인천 부평구, 대구 북구, 인천 계양구, 광주 서구, 대구 수성구, 인천 연수구, 대전 서구, 광주 동구, 대구 동구, 대전 유성구, 서울 종로구, 대전 동구

Cluster 7

전남 순천시, 전북 김제시, 전남 무안군, 충남 서산시, 전북 정읍시, 경기 과천시, 충남 홍성군, 충남 당진시, 경남 창녕군, 전남 해남군, 전남 강진군, 전북 부안군, 전남 화순군, 전남 영광군, 전북 고창군, **전남 함평군**, 충북 괴산군, 전남 장성군

5. 숨은 명소 지수



숨은 명소

외국인/내국인 관광객이 잘 알지 못해 방문객이 적어 홍보 및 인프라 구축이 필요한 곳

1. 내국인 관광객들에 비해 외국인 관광객의 수가 적은 지역
2. 현지인들의 카드사용액에 비해 외지인들의 카드사용액이 적은 지역

숨은 명소 지수



관광객의 만족도가 높은 지역

→ 비율로 나타낸 척도여서 방문·이용객 수 자체가 매우 작은 지역의 경우 구분 x

숨은 명소를 제안하는 이유

- ✓ 외국인들의 주요관광코스들이 서울에 집중되는 문제의 해결방안으로 제안

a. 변수 생성

사용한 데이터와 변수

1. 신한카드 내국인 데이터 → 외지인/현지인으로 ratio_ex_in 변수 생성

- 외지인: 거주지(v1)와 이용지역(v2)이 다른 경우
- 현지인: 거주지 (v1) 와 이용지역 (v2) 이 같은 경우

- ✓ 사용 이유: 국내 관광객들에게 숨은 명소인 지역을 찾기 위해

2. 관광데이터랩 기초지자체 방문자 수 데이터(2019년) → 외국인여행객/한국여행객(외지인)으로 ratio_visit 변수 생성

- ✓ 사용 이유: 외국인들에게 숨은 명소인 지역을 찾기 위해

cluster_	region	ratio_ex	ratio_vis
0	경남 거제	0.258151	0.018911
0	경남 통영	0.404013	0.029052
0	경북 포항	0.189096	0.00622
0	부산 기장	0.414124	0.01231
0	부산 남구	0.20348	0.01457
0	부산 동구	0.305256	0.022582
0	부산 동래	0.132709	0.002733
0	부산 부산	0.250857	0.016301
0	부산 사상	0.201093	0.005841
0	부산 서구	0.287612	0.025087
0	부산 수영	0.239618	0.006726
0	부산 연제	0.134351	0.004835

5. 숨은 명소 지수 - PCA



b. 숨은 명소 지수 생성

숨은 명소 지수를 하나의 값으로 나타내기 위해 차원축소 기법인 PCA 사용

cluster_lakregion	ratio_ex_irratio_visipca1
0 울산 북구	0.128955 0.012259 -0.52864
0 부산 동래	0.132709 0.002733 -0.52494
0 부산 연제	0.134351 0.004835 -0.52328
0 경북 포항	0.189096 0.00622 -0.46853
0 부산 사상	0.201093 0.005841 -0.45654
0 부산 남구	0.20348 0.01457 -0.45411
0 부산 수영	0.239618 0.006726 -0.41801
0 부산 영도	0.239974 0.038141 -0.4175
0 부산 부산	0.250857 0.016301 -0.40672
0 경남 거제	0.258151 0.018911 -0.39942
0 부산 해운	0.277698 0.020436 -0.37986
0 부산 서구	0.287612 0.025087 -0.36993
0 부산 동구	0.305256 0.022582 -0.3523
0 경남 통영	0.404013 0.029052 -0.25351
0 부산 기장	0.414124 0.01231 -0.24348
0 제주 제주	0.700592 0.09379 0.043374
0 충남 보령	0.868141 0.008488 0.210516
0 제주 서귀	1.195404 0.059764 0.538019
0 충남 태안	1.247911 0.035255 0.590408

숨은명소 지수 낮은 지역은 네이버 블로그 크롤링 데이터
감성 분석을 통해 각 군집별 대표적인 숨은명소 추출

숨은 명소 지수 높은 지역은 네이버 블로그 크롤링 데이터
워드클라우드로 시각화하여 군집별 관광 키워드 추출

✓ PCA (n_component = 1) Code

```
import pandas as pd
data_pca = pd.read_csv("C:\\Users\\SamSung\\tour\\pca_data1.csv")
data_pca = pd.DataFrame(data_pca)
x = data_pca.drop("region",axis=1)
x = x.drop("place_num",axis=1)

from sklearn.decomposition import PCA
pca = PCA(n_components=1) # 주성분을 몇개로 할지 결정
principalComponents = pca.fit_transform(x)
pcaDF = pd.DataFrame(data=principalComponents, columns = ['pca1'])
# 주성분으로 이루어진 데이터 프레임 구성

data_pca['pca1'] = pcaDF['pca1']
```

5. 숨은 명소 지수



숨은 명소 후보군

Cluster 0

부산광역시 동래구
부산광역시 연제구
울산광역시 북구

Cluster 1

경기도 군포시
울산광역시 중구

Cluster 2

전라북도 익산시
전라북도 전주시
충청북도 증평군

Cluster 3

대구광역시 달성군
전라남도 목포시

Cluster 4

경기도 양주시
경기도 남양주시
경상남도 진주시
경기도 의왕시

Cluster 6

대구광역시 달서구
대구광역시 서구
울산광역시 동구

Cluster 7

전라남도 순천시
전라북도 김제시

5. 숨은 명소 지수 - 감성분석



c. 감성 분석

숨은 명소 지수가 작은 지역들 중 관광 만족도가 높은 지역을 군집별 대표적인 숨은 명소로 판단

관광 만족도를 확인하기 위해 네이버 블로그 본문 데이터를 이용한 감성분석 실시

1. 사용 데이터

감성 사전 : KNU 한국어 감성사전 사용

분석 데이터 : 지역별 네이버 블로그 본문 크롤링 데이터

2. 분석 방법

word	polarity
아름다운	2
자연을	0
느끼며	0
도시의	0
생활에	0
지친	-1
몸과	0
마음을	0



content	score
화천에서 증평으로 왔습니다. 춘천역 > 청	-2
충청북도 여행 명소 best 11을 뽑아 보았	23
전국 5일장을 다 돌아보고 싶은 해피콜입	1
증평군 여행을 하게 된 것은 그곳에 사는	26
내일로 3일차 증평에 도착! 버스를 타고 지	-3
세상이 초록의 싱그러움으로 물들기 시작	16
신록이 아름다운 푸른색으로 바뀌어 가는	40
증평가볼만한곳, 충북가볼만한곳 언택트	2
안녕하세요 충청북도 여행 가이드 들락날	0
서천사랑이 전하는 청주여행이야기 오늘	30



군집	지역	점수
0	부산광역시 동래구	13.18421
0	부산광역시 연제구	11.86842
0	울산광역시 북구	17.31707
1	경기도 군포시	7.264706
1	울산광역시 중구	14.56522
2	전라북도 익산시	14.56522

✓ R tidytext 패키지의 `unnest_tokens()` 함수 이용해
텍스트를 개별 토큰으로 분해

✓ 감성 사전을 이용해 감성 점수 부여

✓ 사분위수 Q1, Q3 사이의 값만 추출해 구한 평균값을
지역별 감성 점수로 사용

5. 숨은 명소 지수 - 감성분석



c. 감성 분석 - 결과



각 군집별로 숨은 명소 후보군 중 감성 분석 점수가 가장 높은 지역을 그 군집을 대표하는 숨은 명소로 지정

Cluster 0

부산광역시 동래구 13.18421
부산광역시 연제구 11.86842
울산광역시 북구 17.31707 ✓

Cluster 4

경기도 양주시 10.46667
경기도 남양주시 9.285714
경상남도 진주시 11.16 ✓
경기도 의왕시 10.21429

Cluster 1

경기도 군포시 7.264706
울산광역시 중구 14.56522 ✓

Cluster 6

대구광역시 달서구 10.65
대구광역시 서구 7.057
울산광역시 동구 11.77778 ✓

Cluster 2

전라북도 익산시 13.075
전라북도 전주시 9.416667
충청북도 증평군 15.37838 ✓

Cluster 7

전라남도 순천시 7.076923
전라북도 김제시 13.28125 ✓

Cluster 3

대구광역시 달성군 10.88889 ✓
전라남도 목포시 7.655172

바다, 해수욕장, 해안, 힐링, 카페, 맛집,
가족, 숙소, 식당, 횃집, 음식, 서해안, 쇠
소각, 백사장, 사구, 갯벌, 축제, 해변

5. 숨은 명소 지수

CLUSTER 1 관광 키워드

자연, 바다, 산, 힐링, 축제, 문화, 식당,
숙소, 해안, 전망대, 등대, 해안, 산, 나
무, 죽녹원, 화진포, 금강산

#강원 고성군 #경북 울릉군 #전남 담양군

CLUSTER 2 관광 키워드

문화재, 축제, 문화, 고택, 체험, 자연, 힐
링, 휴양림, 가족, 적상산, 진악산, 대둔산,
산성, 안국사, 화산, 인삼, 반딧불, 삼국유
사, 석굴암, 나제통문

#경북 군위군 #충남 금산군 #전북 무주군

CLUSTER 3 관광 키워드

시장, 카페, 맛집, 숙소, 바다, 해안, 나
들이, 먹방, 해수욕장, 설악산, 횃집, 대
게

#인천 강화군 #강원 속초시

CLUSTER 4 관광 키워드

마을, 문화, 체험, 축제, 월정사,
마곡사, 운문사, 사찰, 백제, 오대산, 계곡,
케이블카, 대관령, 무령왕릉, 빛축제

#강원 평창군 #경북 청도군 #충남 공주시

CLUSTER 6 관광 키워드

도시, 맛집, 데이트, 미식, 전시, 관람,
문화, 문화재, 경복궁, 북촌, 창덕궁, 대
동하늘공원, 대청호

#서울 종로구 #대전 동구

CLUSTER 7 관광 키워드

문화, 축제, 생태, 가을, 단풍, 자연, 체
험, 역사, 문화재, 산, 속리산, 임꺽정,
홍길동, 옛길, 백양사

#전남 함평군 #전남 장성군 #충북 괴산군

6. 결론 - 키워드 활용



군집별 관광 키워드를 이용 + 대표 명소의 외국인 유치 정책을 벤치마킹
→ 숨은 명소의 외국인 대상 관광 정책을 개선

벤치마킹 예시

✓ CLUSTER 4의 숨은 명소: 경남 진주

경남 진주의 진주 남강 유등축제



출처 : 진주 남강유등축제 홈페이지

✓ 벤치마킹

CLUSTER 4의 대표 지역: 충남 공주시
충남 공주의 홈스테이 제도



① 한국문화교류 : 전세계에서 한국을 방문하는 외국인들이 한국에 머무르는 동안, 한국의 가정에서 생활하게 함으로써 한국과 한국문화에 대한 이해를 높이고 친분을 쌓음으로 민간외교의 역할까지 할 수 있는 좋은 제도입니다.
② 문화체험 기회 : 외국인의 경우는 이러한 국제민박제도가 활성화되어 있어 우리의 유학생이나 어학연수생 또는 여행객이 저렴한 가격으로 이용하여 그 나라의 문화와 언어에 좀 더 많이 접할 수 있는 기회를 제공하고 있습니다.
출처 : 공주시 홈페이지



6. 결론 - 활용 방안

〈 어플리케이션 개발〉



분석 결과 도출된 숨은 명소를 활용하여 외국인 대상으로
가까운 숨은 명소를 추천해주는 어플리케이션 개발



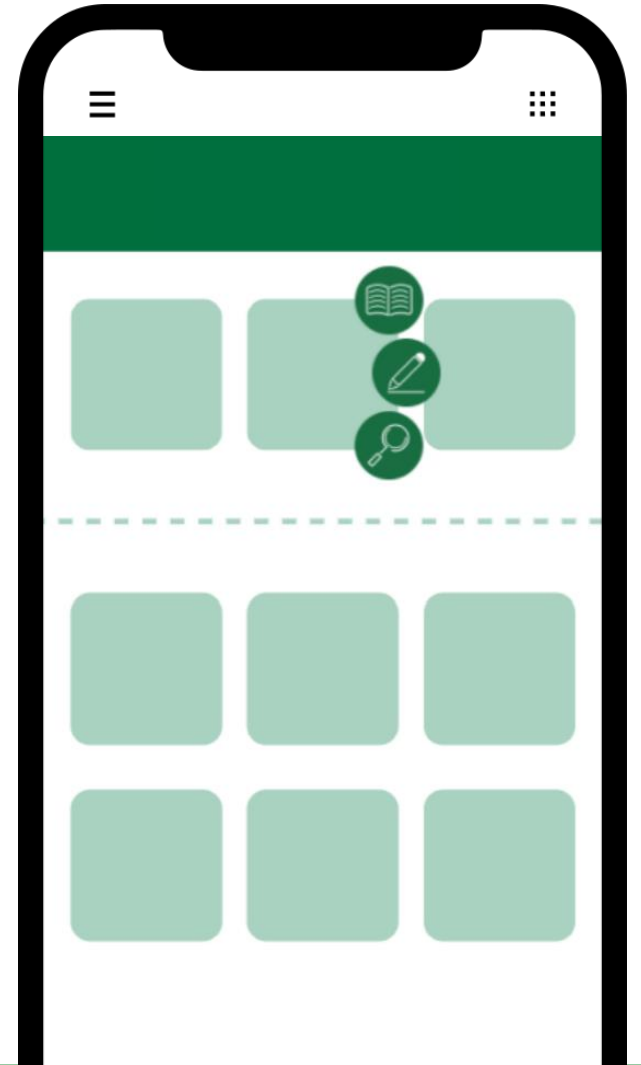
한국 여행 또는 계획 중인 외국인들은 해당 어플리케이션을 통해
현지인(내국인)만 아는 숨은 명소를 검색할 수 있다.



작동 방식 : 어플리케이션을 사용하는 외국인 관광객의 위치 정
보를 받아 가까운 여행지를 추천한다.



기대 효과 : 외국인 관광객들은 편리하고 색다른 관광을 할 수 있
으며, 코로나로부터 안전한 여행을 기대할 수 있다.



6. 결론 - 활용 방안



〈 한류 스타와의 협업 〉

BTS와 대구 곳곳을 둘러보자! (BTS gives ARMY Permission to Daegu!)

- ✓ **제작 이유** : 외국인들은 한류스타가 등장하는 여행 홍보 영상의 주요 시청층으로 외국인 대상의 홍보용으로 적합하다 판단.
e.g VisitSeoulTV의 “ 24시간이 모자라! 차은우가 서울 액티비티 다 알려줌 ” → 댓글의 대부분이 외국어
- ✓ **제작 목적** : 숨은 명소 지역이 포함된 대구광역시의 관광지들을 외국인들에게 소개
- ✓ **‘방탄소년단’ 선정 이유** : 세계적인 인기를 누리고 있고 특히나 대구는 몇몇 멤버들의 고향이므로 콘텐츠의 주인공으로 선정
- ✓ **콘텐츠 내용** : ‘방탄소년단’과 대구광역시의 곳곳을 여행하는 컨셉의 유튜브 영상을 제작하여 외국인 관광객들의 대구 여행을 유도
- ✓ **기대 효과** : 영상을 접한 외국인들이 한국 여행 시, 서울 뿐만 아니라 대구에도 방문하여 외국인 관광객들이 서울로 집중되는 문제 해소

〈 제작 예시 〉



[360°VR DAEGU] 24Hours are not enough! Enjoying with BTS | BTS와 함께하는 대구 투어

조회수 1394만회 • 1개월 전



Visit Daegu TV ✓

“다시 여행이 시작되면 그 시작은 대구가 되길 바래“ BTS의 고향에서 함께하는 360°VR 대구 랜선여행.

6. 결론 - 활용 방안

〈 콘텐츠 제작사와의 협업 〉

나만 알기 아까운 우리 동네, 넷플릭스에서 보고싶어!

주관: 문화체육관광부 (이하 '문체부')

심사: 넷플릭스

✓ 공모전 참여가능 대상: 군집별 숨은 명소로 선정된 기초지자체

(울산광역시 북구, 중구, 동구, 충청북도 증평군, 대구광역시 달성군, 경상남도 진주시, 전라남도 김제시)

✓ 공모전 내용

1. 각 기초지자체에 거주하는 주민들이 '넷플릭스에서 보고싶은 우리만 알기 아까운 멋진 장소'를 촬영하여 출품
2. 공모전의 우승 지역은 넷플릭스 제작 콘텐츠인 '넷플릭스 오리지널'의 촬영지로 활용

✓ 공모전의 기대 효과

1. 공모전의 우승 지역은 넷플릭스 콘텐츠의 촬영지로 선정되어 외국인 관광객 유치 및 홍보 정책에 활용
2. '넷플릭스'는 공모전에서 선정된 지역에서 영화 촬영을 하는 것에 대한 지원을 '문체부'로부터 받는다.
3. '문체부'는 외국인 관광객들이 서울로 집중되는 오버투어리즘 문제의 해결책으로 활용



이미지 출처 : 드라마 '도깨비'

6. 결론



코로나 이후의 관광 정책

: 서울 외 다양한 지역에서 외국인 관광객 유치에 힘써 **여행지 분산을 유도** 해야 할 것으로 보임

외국인 관광객들을 대상으로 홍보할 **숨은 명소를 데이터 분석을 통해 선정**

군집별 숨은 명소에 외국인 대상 여행 상품, 관광 인프라 확보 + 외국인을 대상으로 하는 홍보

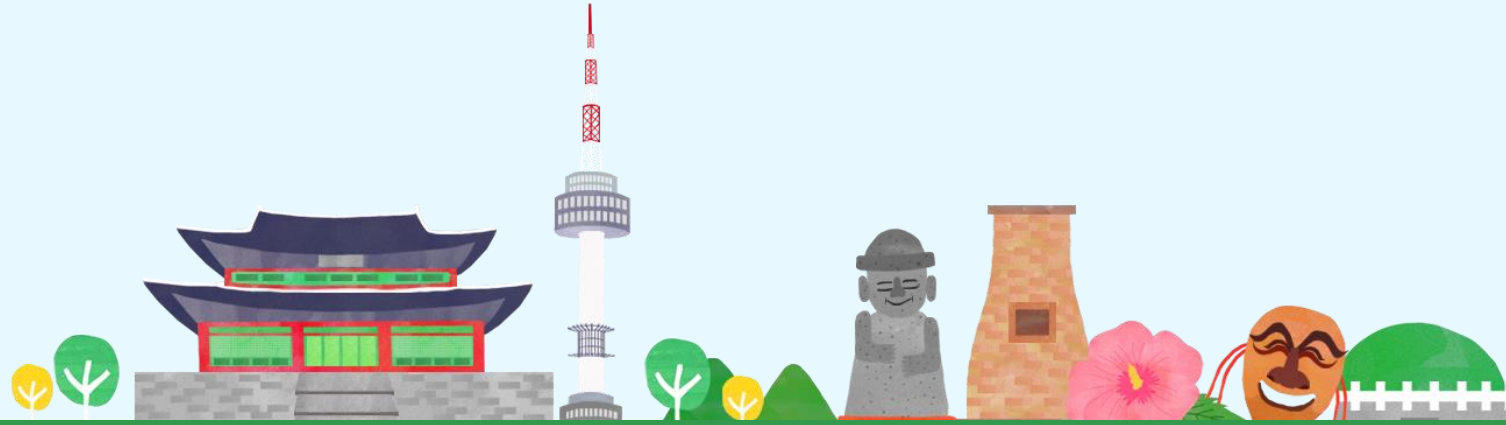
숨은 명소에 외국인 관광객 유치 가능

∴ 수도권 지역의 외국인 관광객 오버투어리즘 해소 & 보다 안전한 방향의 여행

7. 출처



"백신 맞았으니 해외여행 가자"...여권 신청 폭증하는 미국[웹사이트]. (2021년07월15일). Retrieved from https://www.kvina.co.kr/newsCenter/view.asp?articleId=202107152505i&pgCode=0303&utm_source=dable
<https://trends.google.co.kr/trends/explore?geo=KR&q=k-drama> (구글트렌드)
<https://trends.google.co.kr/trends/explore?q=k-pop&geo=KR> (구글트렌드)
<https://datalab.visitkorea.or.kr/datalab/> (한국관광 데이터랩)
<http://dilab.kunsan.ac.kr/knusl.html> (KNU 한국어 감성사전)
Aurelien Geron. (2020). *핸즈온 머신러닝(2판) (사이킷런, 케라스, 텐서플로 2를 활용한 머신러닝, 딥러닝 완벽 실무)*. 한빛미디어.
권철민. (2020). *파이썬 머신러닝 완벽가이드*. 위키북스.
<https://yudeung.com/origin> (진주 남강유등축제 홈페이지)
http://www.gongju.go.kr/prog/tursmCn/tour/sub03_02_07/view.do?cntno=82 (공주시 홈페이지)



감사합니다.