









INDEX

- 1. 주제 선정 배경
- 2. 국내 관광지 추천 시스템 소개
- 3. 데이터 설명
- 4. 데이터 분석
- 5. 최종 시스템 알고리즘
- 6. 국내 관광지 추천 시스템 활용예시
- 7. 분석 의의 및 활용방안



주제 선정 배경



- 한국관광공사 2020년 국내여행 트렌드 중 하나는 인기 여행지보다 덜 알려진 여행지로 가려는 경향
- 대도시나 인기 여행지보다 조용하고 쾌적한 여행을 즐길 수 있는 숨은 명소 관광 수요가 증가



- 언택트 관광은 비대면, 비접촉 관광을 의미
- COVID-19로 인해 사람들이 여행객 밀집도를 중요시하면서 언택트 관광 수요가 증가



나만 아는 관광지를 찾는 트렌드가 포스트 코로나 시대에 더욱 가속화될 전망

주제 선정 배경

2

오버투어리즘 문제 심화

- 오버투어리즘 : 과도한 관광객 유입으로 지역주민의 삶과 생활환경에 부정적인 영향을 미치는 현상
- 국내는 제주도, 해운대 지역의 오버투어리즘 문제가 심각



2019년 국민여행조사 여행지 선택 이유 1순위



• 오버투어리즘 문제 해결을 위해 '관광객 분산 유도'가 필요



경기연구원 설문조사 오버투어리즘 정책방향

오버투어리즘 문제가 심화되고 있으며, 이를 해결하기 위한 정책수단 필요

국내 관광지 추천 시스템

나만 아는 관광지, 언택트 관광 트렌드



오버 투어리즘 문제 해결 필요성

국내 관광지 추천 시스템

- 시스템 사용자 성향, 소비 패턴 분석
- 사용자 성향 부합 & 대중에게 잘 알려지지 않은 국내 관광지 추천

관광객

본인의 여행 성향과 관광 트렌드 동시에 충족하는 여행지 검색 가능

비인기 관광지

침체된 관광산업 활성화와 주요 관광객 정보 활용한 투자,개발

인기 관광지

오버투어리즘 문제 해결과 방문 관광객 만족도 상승

국내 관광지 추천 시스템







가고 싶은 여행지, 선호하는 관광지

연령

입력

사용자 선호 관광 특성, 소비 패턴 분석 결과

선호 관광지 특성과 유사하고 사용자 연령과 예상 경비에 적합한, 잘 알려지지 않은 국내 관광지 추천

예상경비

데이터 설명 - 데이터분석과정

데이터 선택

데이터 분석

결과 도출

네이버 블로그 데이터



여행지별 유사도 측정



• 네이버 api를 통한 블로그 크롤링

- KoNLPy 형태소 분석
- Kmeans Clustering

관광지 추천 시스템

신한카드 데이터



소비패턴 분석



• 2019년 데이터

• 지역, 연령별 소비패턴 분석

1 크롤링

네이버 블로그

블로그 1-10 / 280,138건



<u>순천여행, 순천만습지 근처 돌마루펜션 솔직후기</u> 3일 전

하지만 가격이 착한만큼 큰 기대는 하지 않아야해요ㅎㅎ 1박2일의 짧은 순천여행이었지만 맛있는 것도 먹고 순천만습지도 다녀온 가족여행. 곧 맛있게 먹었던 식당과... 하루하루 추억 한자락、 blog,naver,com/young... | 블로그 내 검색 | ■약도 ▼



<u>주연이랑 함께 한 2박 3일 여수/순천 여행 (•∪•)♥</u> 2020,07,21,

사실 이 포스팅은 삿포로 여행기였어야 했다 (매국노 드립 pass) 원래는 제헌절 끼고... 순천가는 날이라 아침거리 사러 또 여수당 포기할 수 없는 쑥 아이스크림도 한 번... 별것없고새로물것없... blog.naver.com/mere0606/222037538708 | 블로그 내 검색



순천여행 조계산 선암사 승선교 2020,07,27,

여행 준비할때 순천에 갈만한 곳을 간난이에게 보내줬는데 그 리스트 중 승선교를 보고 반 가워했던 간난이, 간난이가 예전에 티비CF에 어떤 돌다리가 나온 것을 보고... 제목없음 blog.naver.com/lovec... │ 블로그 내 검색 │ ▲ 약도 ▼



·순천 여행 꼬막맛집 명품식당 필수코스 2020 05 22

작년 낙안읍성을 함께 다녀왔던 언니에게서 **순천여행**을 가자며 오랜만에 연락이 왔다. 처음 계획은 순천만 다녀오자 였었지만 자유여행의 묘미는 뭐다? 발길 닿는대로...

- 내가 만드는 일상 blog naver com/youli - 트를로그 내 검색 | 🚺 약도



순천 여행, 엄마랑 낙안읍성 구경 그리고 주변 맛집 2020,07.07.

엄마랑 순천 여행, 낙안 읍성 근처에서 꼬막 정식 먹고 낙안 읍성 구경 원래 지난... 다녀왔 드아~ 순천은 몇번 여행을 다녀왔지만, 이번에는 좀 다른 루트로 돌아보고...

■방랑써니의 여행일... sunliy,blog,me/222023536277 │ 블로그 내 검색



 0 주연이랑 함께 한 2박 3일 여수/순천 여행 (•○•) ○ 사실 이 포스팅은 삿포로 여행기 1 홀딱 반한 순천 맛집 지난 주말엔 여름 휴가 겸 전라도에 갔다가, 유명한 순천 맛집 0 2 만족했었던 순천 맛집 몇일전 친구들고 약속이 있어서 순천에서 만났습니다.점심으 3 소담스런 순천 한정식 얼마 전 순천으로 여행을 가면서 한끼 정도는 순천의 손맛이 4 순천 꼬막맛집 밥도둑이야 주말에 친한 언니가 출장차 순천 간다길래 따라가게 심심 5 고급졌던 순천 맛집 #밥꽃이야기들마루 #순천맛집 #한식상호명 : 밥꽃이야기 들마 6 순천마사지 제대로 받았어요 이번에 순천으로 이사를 가게 되어서 원래 다니던 마시 7 완전정복 순천 꼬막정식 지난주 친구 커플과 함께 순천에 다녀왔는데요.제대로된 순 	베도 C 로 소. 담긴 I해서 루주소 I지안
2 만족했었던 순천 맛집 몇일전 친구들고 약속이 있어서 순천에서 만났습니다.점심으 3 소담스런 순천 한정식 얼마 전 순천으로 여행을 가면서 한끼 정도는 순천의 손맛이 4 순천 꼬막맛집 밥도둑이야 주말에 친한 언니가 출장차 순천 간다길래 따라가게 심성 5 고급졌던 순천 맛집 #밥꽃이야기들마루 #순천맛집 #한식상호명 : 밥꽃이야기 들마 6 순천마사지 제대로 받았어요 이번에 순천으로 이사를 가게 되어서 원래 다니던 마시	로 소. 담긴 해서 루주소 지안
3 소담스런 순천 한정식 얼마 전 순천으로 여행을 가면서 한끼 정도는 순천의 손맛이 4 순천 꼬막맛집 밥도둑이야 주말에 친한 언니가 출장차 순천 간다길래 따라가게 심선 5 고급졌던 순천 맛집 #밥꽂이야기들마루 #순천맛집 #한식상호명 : 밥꽂이야기 들마 6 순천마사지 제대로 받았어요 이번에 순천으로 이사를 가게 되어서 원래 다니던 마시	담긴 해서 루주소 지안 천 꼬
4 순천 꼬막맛집 밥도둑이야 주말에 친한 언니가 출장차 순천 간다길래 따라가게 심심 5 고급졌던 순천 맛집 #밥꽃이야기들마루 #순천맛집 #한식상호명 : 밥꽃이야기 들마- 6 순천마사지 제대로 받았어요 이번에 순천으로 이사를 가게 되어서 원래 다니던 마시	- 니해서 루주소 나지안 - - 천 꼬
5 고급졌던 순천 맛집 #밥꽂이야기들마루 #순천맛집 #한식상호명 : 밥꽂이야기 들마 6 순천마사지 제대로 받았어요 이번에 순천으로 이사를 가게 되어서 원래 다니던 마시	루주소 }지안 ·천 꼬
6 순천마사지 제대로 받았어요 이번에 순천으로 이사를 가게 되어서 원래 다니던 마시	· · ·지안 ·천 꼬
	· · · · · 천 꼬
7 완전정복 순천 꼬막정식 지난주 친구 커플과 함께 순천에 다녀왔는데요.제대로된 순	
	순천 :
8 확실했던 순천 꼬막 정식 지난 주말에는 온가족이 순천 여행을 다녀왔어요. 식사는	
9 순천 한정식 맛집 발견! 얼마 전 오랜만에 부모님을 모시고 순천에 다녀왔는데요특!	별한 선
10 행복했던 순천 맛집 지난 주에는 가족들과 바람을 쐬러 전라도에 갔다가, 현지인들!	∈ 많0
11 순천여행 겨울여행 1일차 올해 1월,코로나사태가 발생하기 전겨울휴가기간 다녀온	순천(
12 월등했던 순천 맛집 #순천맛집#나눌터이번에 여행을 한 바퀴 돌고 오다가 발견한 유	음식점
13 정성 가득했던 순천 맛집본문 기타 기능본문 보기날씨 좋은 어느날에 저는 순천지	인이 .
14 으뜸이였던 순천 맛집 얼마전 친구들과 시간을 내서 순천 드라마세트장을 갔다미리	점찍
15 순천 신상카페 용당동 카페 블루밍 (리버뷰) 출근 길 날씨가 너무 좋네요!새벽엔 천둥	·번개
16 순천 신대지구 맛집 대하구이 시작! ** 내돈주고 먹은 리얼 후기! **정말 오랜만에 돼	스팅
17 순천 용당 남양휴튼 용당동 아파트 정보안내! 반갑습니다! 지하 1F ~ 지상 15F의 구	성을 :
18 포레나 순천 인기 주거브랜드 8월 분양 한화건설은 다음달 전라남도 순천시 서면 선	평리
19 순천 용당동 남양휴튼 용당 아파트 정보안내 오늘 소개해 드릴 곳은 지역 주택 조힙	에서
20 다채롭던 순천 조례동 맛집 며칠 전에 친한 지인과 함께 순천 조례동 맛집에 다녀욌	어요.
21 극찬했던 순천 호수공원 맛집 얼마전에는 친구들과 순천 호수공원 맛집에 다녀왔어	요. 지

'지역명 + 여행'을 키워드로 지역당 1000개의 블로그 크롤링

2 형태소 분석

```
주연 여수 순천 여행 사실 삿포로 여행기 하 매국노 드립 원래 제헌절 끼 여름휴가 겸 주
```

KoNLPy의 kkma 형태소 분석기를 통해 일반명사, 고유명사, 형용사, 동사, 어근만 추출

```
stop_words = []
for words in list(map(lambda x : x.split(' ')[:-1], CITY)):
  for word in words :
    if not word in stop_words:
        stop_words.append(word)
stop_words = list(map(lambda x : normalize('NFC', x), stop_words))
stop_words += ['강원도', '경기도', '전라도', '경상도', '총청도']
```

지역 고유명사는 분석에 필요하지 않으므로 제거

stop_words += ['여행','사진', '시간','그리','들어가','이렇', '정도','가보','보이','나오','위하','가능']

상위 빈도 단어 중 의미없는 단어도 제거

데이터 설명 - 신한카드데이터

기존 변수

변수명	변수설명	변수값
V1	이용회원 거주지	17개 시도
V2	카드이용 가맹점 주소	17개 시도
V3	카드이용 가맹점 주소	시군구
GB3	가맹점 업종 대분류	
GB2	가맹점 업종 소분류	
SEX_CCD	성별	M:남자 F:여자
CLN_AGE_R	연령대	
TA_YM	이용년월	
DAW_CCD_R	요일	RED:주말 WHITE: 평일
APV_TS_DL_TM_R	이용시간대	
VLM	취급액(단위: 원)	
USEC	이용건수(단위: 건)	

신규 변수

변수명	변수설명	변수값
rel	카드 이용 가맹점 주소	V2+V3
GB2	가맹점 업종 소분류	
CLN_AGE_R	연령대	
VLM	취급액(단위: 원)	

V1과 V2가 다르면 여행으로 판단
GB2에서 여행과 관련 있는 변수 값만 사용
→ 숙박, 외식, 관광쇼핑, 레저스포츠, 공연관람, 골프, 체험

TF-IDF 벡터화

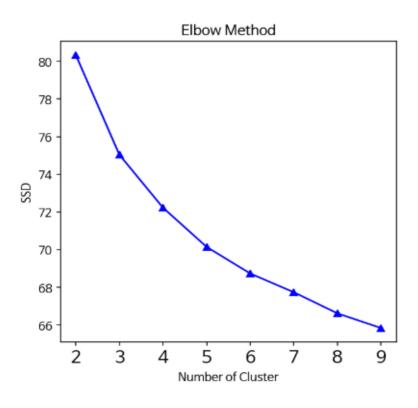
모든 문서에서 자주 사용되는 단어에는 페널티를 부여하고, 특정 문서에만 자주 나오는 단어에 가중치를 주는 방식 → 여행지별 유사도 판별에 용이

분석 방식

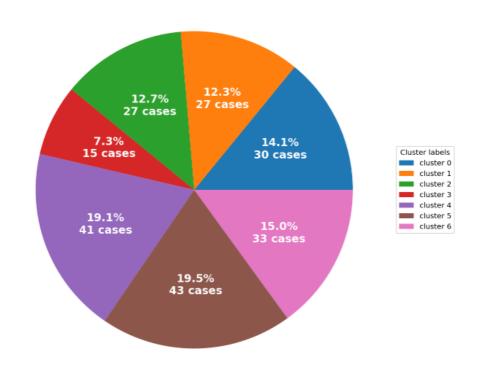
여행지별 상위 250개 단어들에 대해 tf-idf 벡터화 진행
→ 220 rows x 250 columns 의 TF-IDF matrix 생성

0 1 2 3 246 247 248 249 0 0.020927 0.053586 0.026159 0.027110 0.026476 0.014585 0.039634 0.056281 1 0.013365 0.044459 0.036822 0.034367 0.021139 0.010774 0.058779 0.038868 2 0.034145 0.039541 0.020189 0.019817 0.009622 0.006234 0.010606 0.028469 3 0.034749 0.044424 0.035976 0.040472 0.016625 0.011583 0.040609 0.041154 4 0.024561 0.075289 0.028735 0.038688 0.030019 0.016856 0.023919 0.042701 </th <th></th>										
1 0.013365 0.044459 0.036822 0.034367 0.021139 0.010774 0.058779 0.038868 2 0.034145 0.039541 0.020189 0.019817 0.009862 0.006234 0.010606 0.028469 3 0.034749 0.044424 0.035976 0.040472 0.016625 0.011583 0.040609 0.041154 4 0.024561 0.075289 0.028735 0.038688 0.030019 0.016856 0.023919 0.042701 0.030019 0.016856 0.023919 0.042701		0	1	2	3		246	247	248	249
2 0.034145 0.039541 0.020189 0.019817 0.009862 0.006234 0.010606 0.028469 3 0.034749 0.044424 0.035976 0.040472 0.016625 0.011583 0.040609 0.041154 4 0.024561 0.075289 0.028735 0.038688 0.030019 0.016856 0.023919 0.042701 .	0	0.020927	0.053586	0.026159	0.027110		0.026476	0.014585	0.039634	0.056281
3 0.034749 0.044424 0.035976 0.040472 0.016625 0.011583 0.040609 0.041154 4 0.024561 0.075289 0.028735 0.038688 0.030019 0.016856 0.023919 0.042701 .	1	0.013365	0.044459	0.036822	0.034367		0.021139	0.010774	0.058779	0.038868
4 0.024561 0.075289 0.028735 0.038688 0.030019 0.016856 0.023919 0.042701 0.030019 0.016856 0.023919 0.042701	2	0.034145	0.039541	0.020189	0.019817		0.009862	0.006234	0.010606	0.028469
.	3	0.034749	0.044424	0.035976	0.040472		0.016625	0.011583	0.040609	0.041154
215 0.030780 0.073951 0.028381 0.042172 0.044570 0.046769 0.015790 0.046969 216 0.011863 0.043192 0.034067 0.053382 0.021748 0.020684 0.035588 0.035882 217 0.010516 0.029446 0.027042 0.036957 0.027943 0.025990 0.028244 0.029145 218 0.025133 0.067875 0.030946 0.036588 0.026842 0.019149 0.021200 0.033510 219 0.020679 0.030542 0.027360 0.028315 0.023065 0.016862 0.031337 0.036905	4	0.024561	0.075289	0.028735	0.038688		0.030019	0.016856	0.023919	0.042701
216 0.011863 0.043192 0.034067 0.053382 0.021748 0.020684 0.035588 0.035882 217 0.010516 0.029446 0.027042 0.036957 0.027943 0.025990 0.028244 0.029145 218 0.025133 0.067875 0.030946 0.036588 0.026842 0.019149 0.021200 0.033510 219 0.020679 0.030542 0.027360 0.028315 0.023065 0.016862 0.031337 0.036905										
217 0.010516 0.029446 0.027042 0.036957 0.027943 0.025990 0.028244 0.029145 218 0.025133 0.067875 0.030946 0.036588 0.026842 0.019149 0.021200 0.033510 219 0.020679 0.030542 0.027360 0.028315 0.023065 0.016862 0.031337 0.036905	215	0.030780	0.073951	0.028381	0.042172		0.044570	0.046769	0.015790	0.046969
218 0.025133 0.067875 0.030946 0.036588 0.026842 0.019149 0.021200 0.033510 219 0.020679 0.030542 0.027360 0.028315 0.023065 0.016862 0.031337 0.036905	216	0.011863	0.043192	0.034067	0.053382		0.021748	0.020684	0.035588	0.035892
219 0.020679 0.030542 0.027360 0.028315 0.023065 0.016862 0.031337 0.036905	217	0.010516	0.029446	0.027042	0.036957		0.027943	0.025990	0.028244	0.029145
	218	0.025133	0.067875	0.030946	0.036588		0.026842	0.019149	0.021200	0.033510
220 rows × 250 columns	219	0.020679	0.030542	0.027360	0.028315		0.023065	0.016862	0.031337	0.036905
	220 rows × 250 columns									

TF-IDF Matrix를 K-means 알고리즘을 통해 유사한 여행지끼리 군집화



Elbow method를 이용해 클러스터 개수 7개로 결정



각 클러스터별로 최소 15개에서 최대 43개의 지역들로 분류

클러스터링

Cluster 0

창녕 여행,함안 여행,인천 동구 여행,익산 여행,이천 여행,충주 여행,합천 여행,증평 여행,시흥 여행,부천 여행,논산 여행, 나주 여행,김천 여행,과천 여행,의왕 여행,공주 여행,음성 여행,용인 여행,예산 여행,부여 여행,연천 여행,고령 여행, 김해 여행,상주 여행,수원 여행,여주 여행,세종 여행,진천 여행,문경 여행,안산 여행 양주 여행

Cluster

하남 여행,대구 수성구 여행,광주 서구 여행,인천 남동구 여행,인천 서구 여행,부산 부산진구 여행,대전 서구 여행, 대구 남구 여행,인천 부평구 여행,광주 광산구 여행,광주 동구 여행,광주 북구 여행,부산 수영구 여행,부산 연제구 여행, 대구 달서구 여행,대구 서구 여행,인천 계양구 여행,인천 미추홀구 여행,의정부 여행,부산 동래구 여행,울산 북구 여행, 대전 유성구 여행,대구 북구 여행,서울 강남구 여행,서울 서초구 여행,대구 동구 여행,부산 금정구 여행

Cluster 2

태안 여행,제주 여행,통영 여행,신안 여행,완도 여행,고성 여행,거제 여행,강릉 여행,영덕 여행,포항 여행,보령 여행,삼척 여행,부산 해운대구 여행,부산 기장 여행,울산 동구 여행,서귀포 여행,양양 여행,동해 여행,고흥 여행,목포 여행,속초 여행,여수 여행,울진 여행,사천 여행,울릉 여행,진도 여행,인천 응진 여행,부안 여행

Cluster 3

전주 여행,의성 여행,아산 여행,부산 사하구 여행,곡성 여행,영양 여행,영주 여행,봉화 여행,군위 여행,부산 영도구 여행,부산 서구 여행,예천 여행,안동 여행,남해 여행,광주 여행,광주 여행,광주 여행

Cluster 4

정읍 여행 철원 여행 태백 여행,해남 여행,청송 여행 장흥 여행,청양 여행,무안 여행,구례 여행,진안 여행,영천 여행, 고창 여행,성주 여행,인제 여행,완주 여행,서천 여행,옥천 여행,임실 여행,화천 여행,영동 여행,장성 여행,산청 여행,양구 여행, 강진 여행,보은 여행,함명 여행,울산 울주 여행,괴산 여행,라산 여행,횡성 여행,광양 여행,순창 여행,김제 여행,의령 여행,거창 여행, 홍성 여행,장수 여행,영양 여행,화순 여행,함양 여행,안성 여행

Cluster 5

평택 여행,대전 대덕구 여행,서울 서대문구 여행,서울 노원구 여행,경산 여행,고양 여행,대전 동구 여행,대구 달성 여행, 울산 남구 여행,부산 동구 여행,서울 금천구 여행,부산 사상구 여행,군포 여행,서울 마포구 여행,안양 여행,서울 관악구 여행, 성남 여행,구리 여행,동두천 여행,서울 강동구 여행,서울 동대문구 여행,부산 북구 여행,서울 구로구 여행,서울 영등포구 여행, 부산 남구 여행,서울 광진구 여행,서울 중랑구 여행,서울 도봉구 여행,인천 연수구 여행,서울 은평구 여행,오산 여행,서울 종로구 여 창원 여행,계룡 여행,서울 송파구 여행,서울 등작구 여행,서울 용산구 여행,서울 성동구 여행,서울 강북구 여행,서울 강서구 여행, 서울 성북구 여행,서울 양천구 여행,부산 강서구 여행

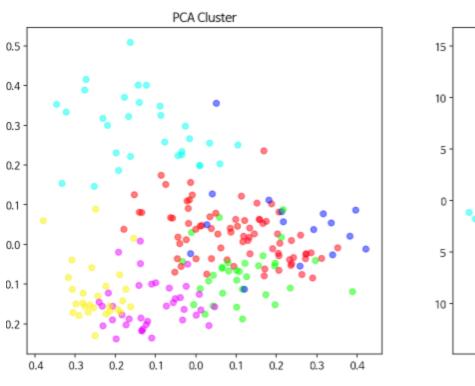
Cluster

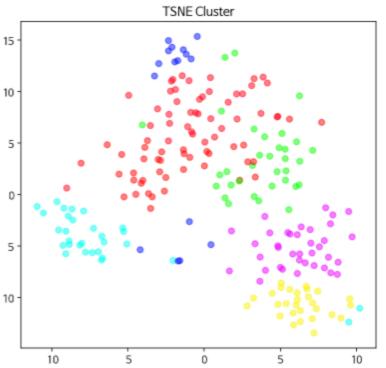
정선 여행,춘천 여행,제천 여행,천안 여행,칠곡 여행,평창 여행,하동 여행,가평 여행,군산 여행,영월 여행,홍천 여행,청주 여행, 진주 여행,양평 여행,밀양 여행,인천 강화 여행,경주 여행,포천 여행,무주 여행,김포 여행,보성 여행,구미 여행,당진 여행,서산 여행 담양 여행,남양주 여행,양산 여행,남원 여행,광명 여행,단양 여행,화성 여행,순천 여행,원주 여행 220개 지역을 총 7개 클러스터로 분류



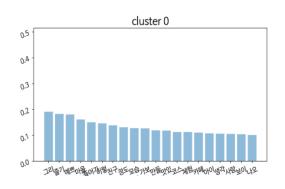
각 클러스터 내 지역들은 유사한 특징을 지니는 여행지라고 판단 가능

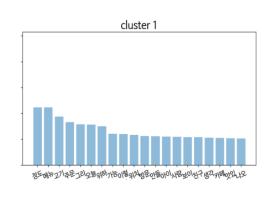
PCA & T-SNE Cluster

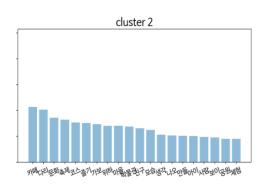


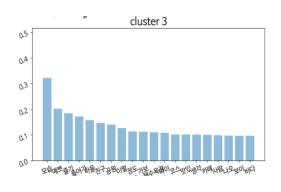


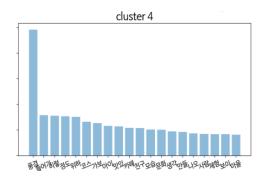
클러스터별로 구분이 잘 된 것을 확인할 수 있다.

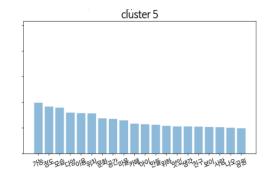


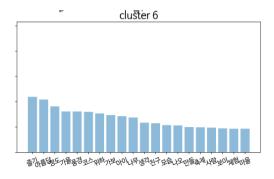












클러스터별 **단어 분포**가 상이함

단어 분포를 통해 지역별 특성 파악이 가능하다.

데이터그룹화

	rel	gb2	cln_age_r	vim
2138	가평	골프	20	20313896.0
2139	가평	골프	30	103202160.0
2140	가평	골프	40	223245963.0
2141	가평	골프	50	273584292.0
2142	가평	골프	60	107932322.0
344253	홍성	종합쇼핑	20	96831469.0
344254	홍성	종합쇼핑	30	178433894.0
344255	홍성	종합쇼핑	40	185375265.0
344256	홍성	종합쇼핑	50	112353248.0
344257	홍성	종합쇼핑	60	51439077.0

지역(rel), 소비항목(gb2), 연령대(cln_age_r)별로 데이터 그룹화

> 지역 내 소비항목, 연령대별로 총 소비금액(vlm)을 파악



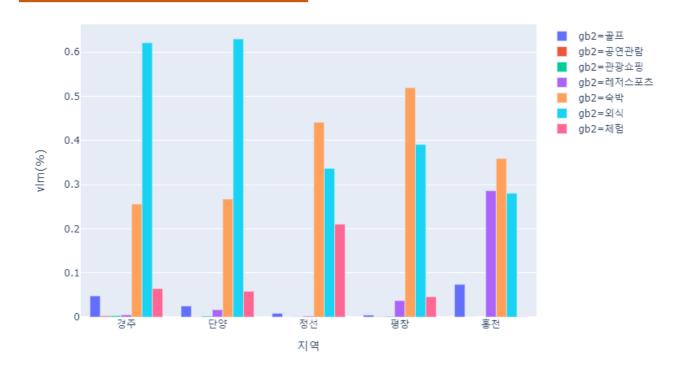
지역 여행객들의 소비 패턴 파악 가능



동일 클러스터 내에서 지역별, 연령별 소비패턴을 파악한다.

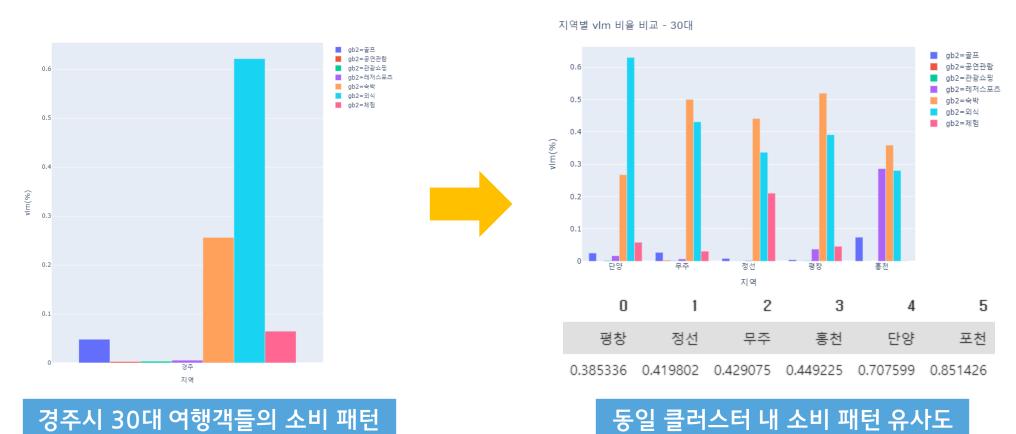
소비패턴이 상이하다면, 같은 클러스터 내에서도 소비패턴별 유사도를 판단할 수 있다.

지역별 30대 소비패턴



같은 클러스터 내에서도 상이한 소비패턴을 보인다.

총 경비에서 **개별 항목들의 비율**을 소비 벡터로 파악 벡터간 거리(민코프스키 거리)를 계산하여 **소비 패턴 유사도** 도출



단순히 소비 패턴의 유사도만을 비교할 시 비유명 여행지 추천이 어려움 총 여행 소비금액이 낮은 지역을 비유명 여행지라 판단하고 가중치 부여

유사도 = 민코프스키 거리 * 클러스터 내 소비금액 비율

0	1	2	3	4	5
평창	정선	무주	홍천	단양	포천
0.385336	0.419802	0.429075	0.449225	0.707599	0.851426



 0
 1
 2
 3
 4
 5

 평창
 정선
 홍천
 무주
 단양
 가평

 12.03
 13.3632
 13.8938
 14.4308
 23.2735
 27.0988

가중치 부여 전

가중치 부여 후

최종 시스템 알고리즘

1

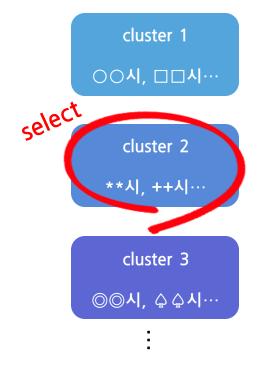
USER'S INPUT

##본래 가고 싶었던 여행지 ₩ANNA_CITY = '**시'

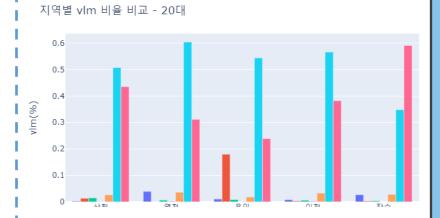
이용자의 연령대 USERS_AGE = 20

rooms = 10 #숙박비로 사용하는 금액
food = 5 #외식비로 사용하는 금액
shopping = 10 #쇼핑으로 사용하는 금액
leisure = 0 #레저스포츠로 사용하는 금액
perform = 5 #공연관람으로 사용하는 금액
golf = 0 #골프로 사용하는 금액
art = 0 #미술 공예 참여로 사용하는 금액

가고자 했던 여행지, 사용자 예상 경비 내역, 나이를 입력값으로 받음 2



해당 여행지가 속하는 클러스터 도출 3



클러스터 내 타 지역의 동일 연령대 소비 패턴 파악

최종 시스템 알고리즘

4

```
tour_feature = tour_feature[tour_feature.rel.isin(in_cluster) & (tour_feature.cln_age_r == age)]
tour_feature = tour_feature.groupby(['rel','gb2']).sum().reset_index()
tour_feature = tour_feature.pivot(index = 'gb2', columns = 'rel', values = 'vlm')
tour_feature = tour_feature.reindex(index = features)
tour_feature.reset_index(inplace=True)
tour_feature.fillna(0, inplace=True)

DIST = []
for rel in in_cluster:
    vlm_sum = tour_feature[rel].sum()
    error = np.linalg.norm(tour_feature[rel].values/vlm_sum - user_vlm_ratio)
    DIST.append(error)
    DIST
result_sort = []
for i in np.argsort(DIST):
    result_sort.append(in_cluster[i])
return result_sort
```





지역간 유사도 계산

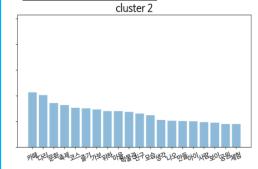
5

blog crawling 결과와 소비 패턴의 유사도를 비교해보았을 때, 당신에게 다음의 5개 여행지를 추천합니다

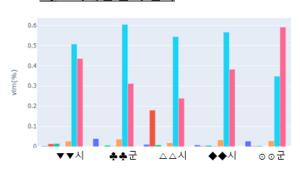
- 1. ▼▼시
- 2. ♣♣군
- 3. △△시
- 4. ◆◆시
- 5. ⊙⊙군

위 여행지들은 다음과 같은 특징을 갖습니다

1. 블로그 크롤링 결과



2. 소비패턴 분석 결과



최종 여행지 추천

국내 관광지 추천 시스템 활용 예시

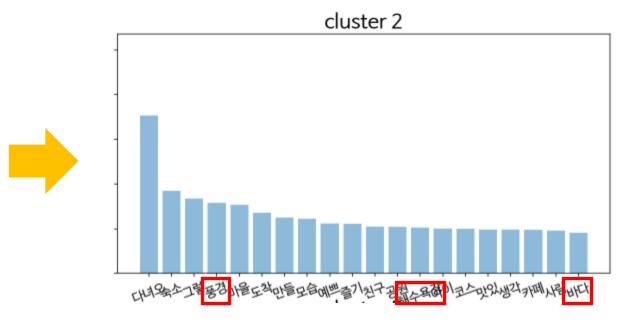


해운대 방문 예정 40대

🚺 해당 클러스터 파악

Cluster 2

태안 여행,제주 여행,통영 여행,신안 여행, 완도 여행,고성 여행,거제 여행,강릉 여행, 영덕 여행,포함 여행,보령 여행,삼척 여행, 부산 해운대구 여행,부산 기장 여행, 울산 동구 여행,서귀포 여행,양양 여행, 동해 여행,고흥 여행,목포 여행,속초 여행, 여수 여행,울진 여행,사천 여행,울릉 여행, 진도 여행,인천 응진 여행,부안 여행



풍경, 해수욕장, 바다와 같은 특성을 가지는 지역들임을 확인

국내 관광지 추천 시스템 활용 예시

2 소비 패턴 분석

소비 예상 금액 입력

```
#'숙박', '외식', '쇼핑', '레저스포츠', '공연관람', '골프', '체험'에 대한 소비를 얼마나 하시나요?

rooms = 12 #숙박비로 사용하는 금액

food = 15 #외식비로 사용하는 금액

shopping = 0 #쇼핑으로 사용하는 금액

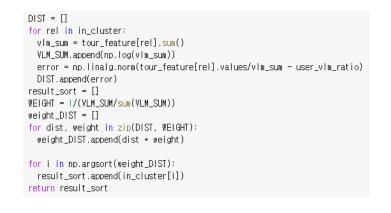
leisure = 2 #레저스포츠로 사용하는 금액

perform = 0 #공연관람으로 사용하는 금액

golf = 0 #골프로 사용하는 금액

art = 1 #체험으로 사용하는 금액
```

비유명 여행지 가산점 부여



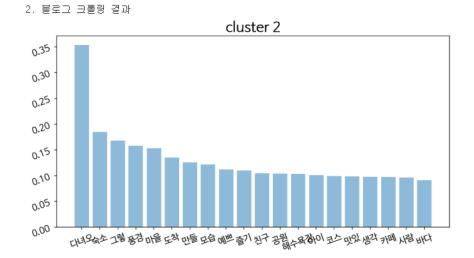
유사도 순위 계산



국내 관광지 추천 시스템 활용 예시

3 최종 여행지 추천





거제, 부안, 삼척, 서귀포, 신안을 유사 여행지로 추천

분석 의의 및 활용방안

의의



지역간 여행객 격차를 해소하여 특정 지역 편중으로 인한 여러 문제점 해결

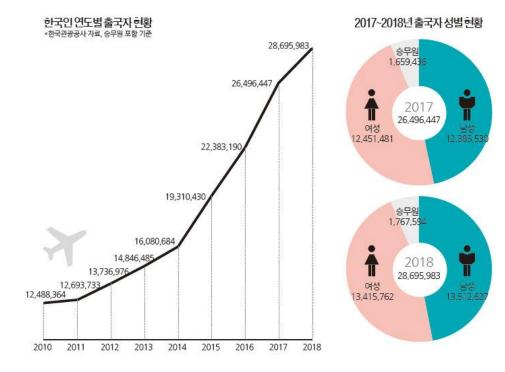
여행객에게는 인원 분산을 통한 쾌적한 여행 환경 제공



단순히 유명하지 않은 여행지를 추천해주는 것이 아니라 사용자 취향에 맞는 여행지 추천이 가능

분석 의의 및 활용방안

활용방안



해외 관광지와 국내 관광지의 유사도를 계산하여 내국인 해외여행 수요자들의 국내 관광 활성화



축처 : 내일로

지역간 클러스터링을 활용하여 여행 concept별 코스 제공

