2024.11.14

강원도 관광지 토픽 모델링

키워드에 맞는 관광지 추천 모델

» Tourist attraction in Gangwon-do
» Topic modeling

VISION

데융스 20
20203223 변지민
20203204 고승균
20203217 박세현
20203228 오진성
20203251 정현우

「한림대학교 데이터사이언스학부 학술제 <VISION>」 결과보고서

팀명	데융스 20
주제	강원도 관광지 추천과 키워드 검색을 위한 토픽 모델링

목 차

제1장 서론

◦ 연구 배경

최근 몇 년간 강원도는 아름다운 자연경관과 다채로운 문화적 매력을 갖춘 관광지로 자리 잡아왔지만, 관광객 수와 그에 따른 관광객 소비가 감소하는 추세를 보인다. 이에 따라 강원도의 관광 산업은 경제적 문제에 직면하고 있으며, 지역 경제 활성화에도 부정적인 영향을 미치고 있다. 코로나19 팬데믹과 같은 외부 요인도 관광 감소에 영향을 주었지만, 팬데믹이 잠잠해졌음에도 변화하는 관광객의 요구와 선호를 제대로 반영하지 못한 것이 중요한 요인으로 작용했다.

(강원도 관광객 감소 관련자료)

https://news.kbs.co.kr/news/pc/view/view.do?ncd=8101739&ref=A

https://www.kwnews.co.kr/page/view/2021122700000000114

관광객의 관심과 선호도를 파악하여 이를 관광 산업의 개선과 활성화에 반영하는 것이 중요해지고 있는 시점입니다. 이에 따라 데이터 분석과 토픽 모델링 기법은 강원도 관광지와 관련된 주요 관심사를 파악하는 데 유용한 도구로 활용될 수 있다고 생각했다.

◦ 현황 및 필요성

관광객들의 여행 트렌드가 점점 더 다양해지고 맞춤형 경험을 중시하는 방향으로 변화함에 따라, 기존의 전통적 마케팅과 서비스로는 충분히 대응하기 어려운 게 현실이다. 이에 따라, 빅데이터 기반의 토픽 모델링을 통해 관광객들이 실제로 어떤 주제와 경험을 선호하는지를 파악하는 것은 매우 중요하다고 느꼈다. 이러한 분석은 강원도의 관광 정책 관련 기관과 기업들이 보다 효과적인 관광 상품을 개발하고, 새로운 마케팅 전략을 수립하는 데 필요한 인사이트를 제공할 수 있을 것으로 기대한다.

◦ 연구 목적

강원도 관광지에 대한 온라인 데이터(방문 리뷰)를 기반으로 LDA 토픽 모델링을 수행하여 관광객들이 선호하는 주제와 그들의 주요 관심사를 도출하는 것이다. 이를 통해서 강원도의 관광 전략을 보다 효과적으로 설계하고, 관광객의 수요를 충족시킬 수 있는 맞춤형 관광 상품과 마케팅 전략을 개발하는 데 기여하고자 한다. 궁극적으로, 강원도의 관광 활성화 방안을 제시하여 지역 경제에 긍정적인 영향을 미치고자 한다.

제2장 연구 방법

◦ 데이터 수집

강원도 관광지에 대한 분석을 위해 주요 온라인 플랫폼인 '구글맵'과 '네이버지도'의 공개된 24곳의 관광지 방문 리뷰(장소 당 100개) 데이터를 활용했다. 이 두 플랫폼은 다양한 사용자층의 리뷰를 포함하고 있어, 관광객들이 실제로 경험한 내용과 주요 관심사를 파악하기에 적합한 데이터 소스로 판단되었다.

- 조사 관광지

강촌 레일파크, 경포대, 국립춘천박물관, 김유정 문학촌, 낙산사, 남이섬, 대관령 양떼목장, 레고랜드, 뮤지엄산, 비발디파크, 양양 서피비치, 설악산, 설악 워터피아, 소금산 출렁다리, 소양강 스카이워크, 아르떼뮤지엄, 알파카월드, 오대산, 오션월드, 오죽헌, 젊은달와이파크, 촛대바위, 치악산, 하슬라 아트월드 (총 24곳)

◦ 데이터 전처리 과정

	장소	리뷰
0	경포대	강원도 강릉시 저동 경포호수 옆 언덕위에 있는 유명한 곳. 호수와 달, 인근 지역,...
1	경포대	경포호를 한눈에 조망할 수 있는 곳입니다. 경포대 안으로 신발 벗고 들어가 볼 수도...
2	경포대	모르고, 지나쳐는데, 가고 보니,경포대의 의미가 뜻 깊어 좋아습니다. 5개 단이라....
3	경포대	경포대에서는 다섯 개의 달을 볼 수 있었단다. 하늘의 달, 호수에 비친 달, 바다에...
4	경포대	과격한 높이에서 바라보는 낭떠러지의 경치가 아니라 나즈막한 언덕 위에서 멀찌감치 바...

<기존 데이터 형태>

	장소	리뷰
0	경포대	강원도 강릉시 저동 경포호수 옆 언덕위에 있는 유명한 곳 호수와 달 인근 지역 바닷...
1	경포대	경포호를 한눈에 조망할 수 있는 곳입니다 경포대 안으로 신발 벗고 들어가 볼 수도 ...
2	경포대	모르고 지나쳐는데 가고 보니경포대의 의미가 뜻 깊어 좋아습니다 5개 단이라 진짜 달...
3	경포대	경포대에서는 다섯 개의 달을 볼 수 있었단다 하늘의 달 호수에 비친 달 바다에 비친...
4	경포대	과격한 높이에서 바라보는 낭떠러지의 경치가 아니라 나즈막한 언덕 위에서 멀찌감치 바...

텍스트에서 불필요한 특수문자와 공백을 제거하여 한글과 숫자를 따로 추출하였다.

	장소	리뷰
0	경포대	[강원도, 강릉시, 저동, 경포, 호수, 언덕, 호수, 인근, 지역, 바닷가, 건물...
1	경포대	[경포호, 한눈, 조망, 경포대, 신발, 무료, 주차장, 주차, 사람, 불만]
2	경포대	[경포대, 의미, 호수, 바다, 술잔, 그대]
3	경포대	[경포대, 다섯, 하늘, 호수, 바다, 술잔, 당신]
4	경포대	[높이, 낭떠러지, 경치, 언덕, 운치, 호수, 계층, 양반]

Mecab 형태소 분석기를 사용하여 문장에서 길이가 2 이상인 명사만 추출하였다.

```
stop_words = ['시간', '추천', '느낌', '최고', '방문', '생각', '정도', '입장료','장소',
              '스팟', '날씨', '마지막', '가격', '가지', '사람', '가능','코로나', '이용',
              '이곳', '위치', '여기', '기분', '우리', '하나','주차', '주차장']
```

< 불용어 >

	장소	리뷰	불용어 제거 후
0	경포대	[강원도, 강릉시, 저동, 경포, 호수, 언덕, 호수, 인근, 지역, 바닷가, 건물...	[강원도, 강릉시, 저동, 경포, 호수, 언덕, 호수, 인근, 지역, 바닷가, 건물...
1	경포대	[경포호, 한눈, 조망, 경포대, 신발, 무료, 주차장, 주차, 사람, 불만]	[경포호, 한눈, 조망, 경포대, 신발, 무료, 불만]
2	경포대	[경포대, 의미, 호수, 바다, 술잔, 그대]	[경포대, 호수, 바다, 술잔, 그대]
3	경포대	[경포대, 다섯, 하늘, 호수, 바다, 술잔, 당신]	[경포대, 다섯, 하늘, 호수, 바다, 술잔, 당신]
4	경포대	[낭떠러지, 경치, 언덕, 운치, 호수, 계층, 양반]	[낭떠러지, 경치, 언덕, 운치, 호수, 계층, 양반]

도출된 키워드 중 연구 목적과 관련이 없거나 지나치게 일반적인 유의미한 정보를 제공하지 않는 키워드를 제거하였다.

- 이것은 모델의 해석력을 높이고, 분석 결과가 더 구체적이고 의미 있는 정보를 제공할 수 있게 조정한 것이다.
- 2번째 줄에서 '주차장', '주차', '사람' 이라는 단어들이 제거된 것을 알 수 있다.

◦ 사용 토픽 모델링 방법

LDA(Latent Dirichlet Allocation) 토픽 모델링 기법 중 LDA를 사용했다. LDA는 토픽 모델링의 가장 대표적인 알고리즘으로 텍스트 기반의 문서 데이터에서 핵심 주제(Topic)를 찾는 데이터 분석 방법론이다. 각 리뷰에 어떤 토픽이, 어떤 비율로 구성되어 있는지 분석하는 데에 적합하다고 생각했다.

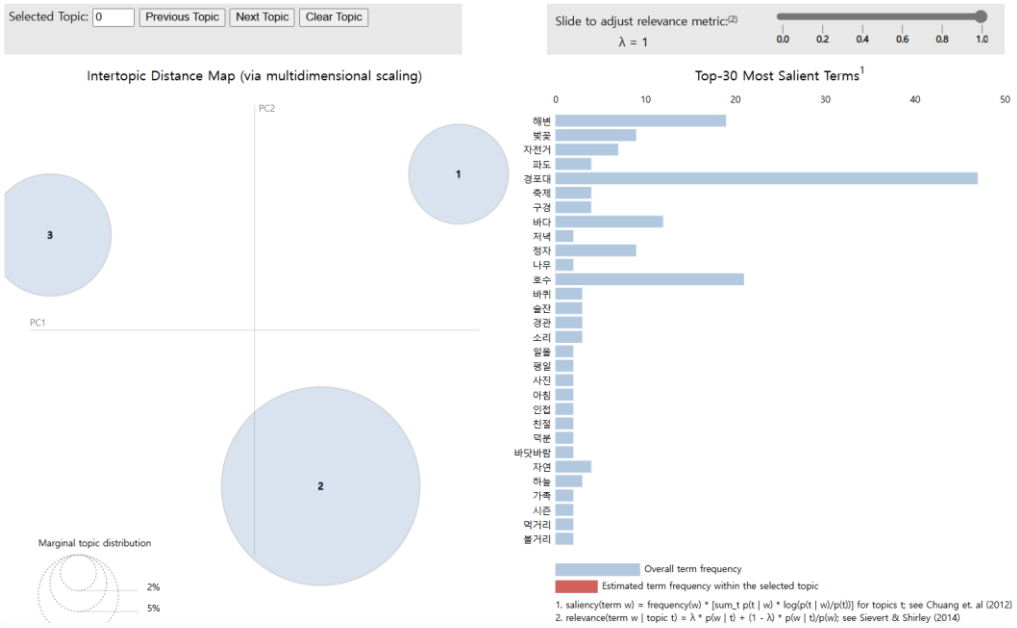
◦ 사용 소프트웨어 및 라이브러리

SW: Python

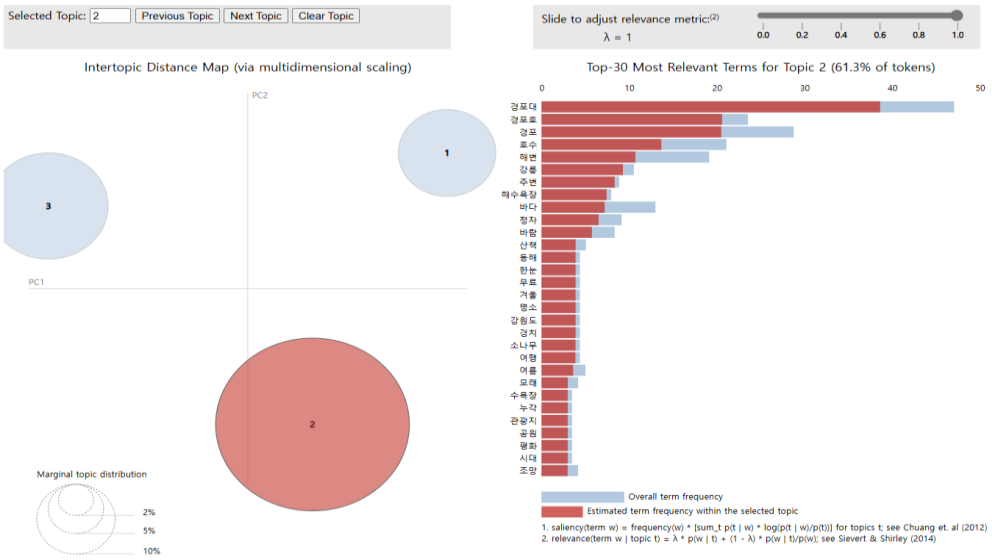
Library: Mecab, ldamodel, pandas 등

제3장 결과

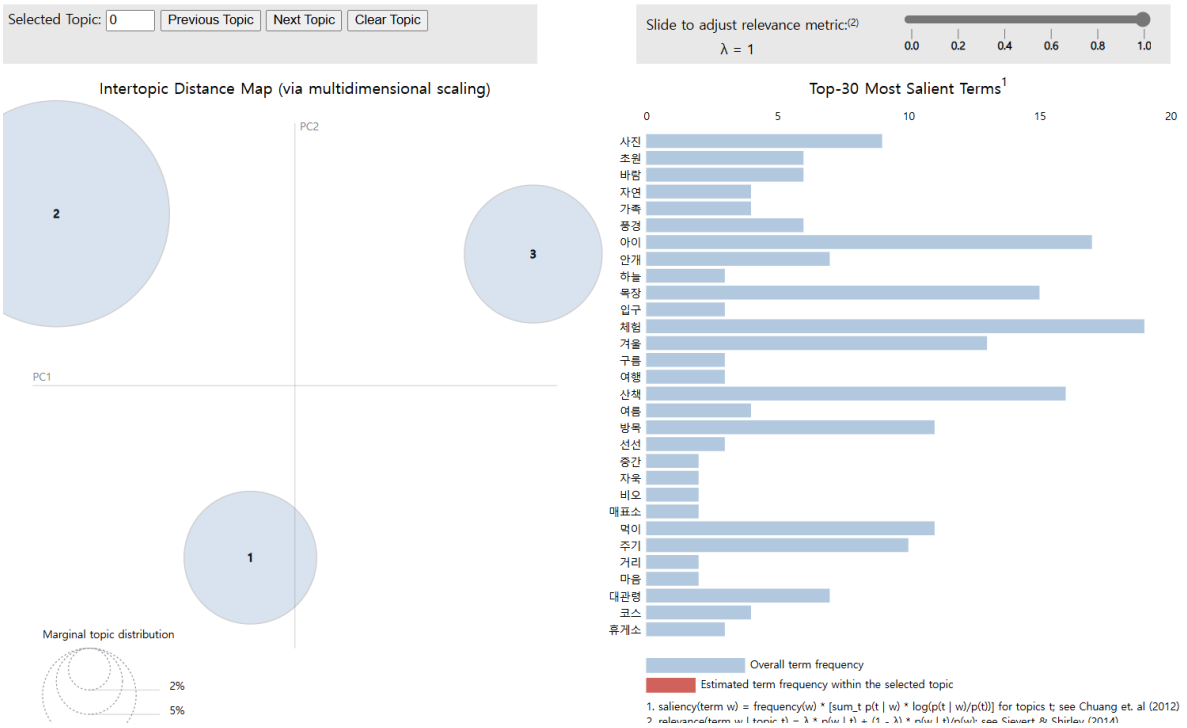
◦ 연구 결과



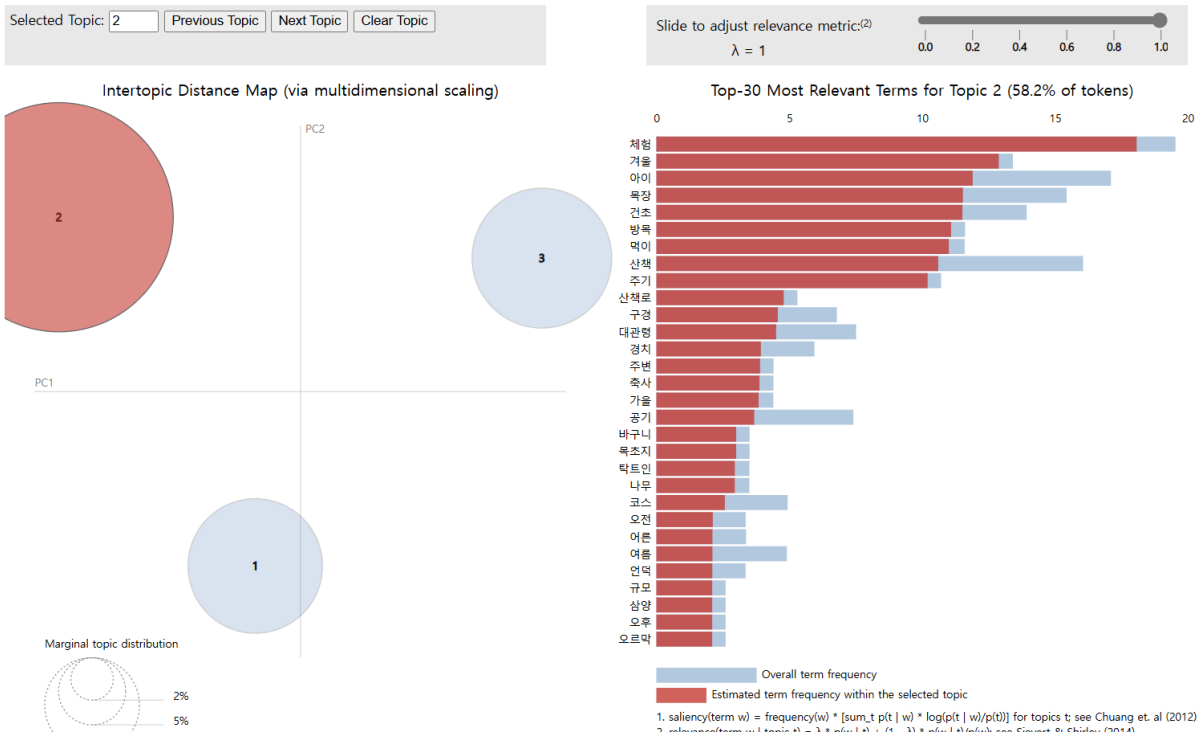
<경포대>



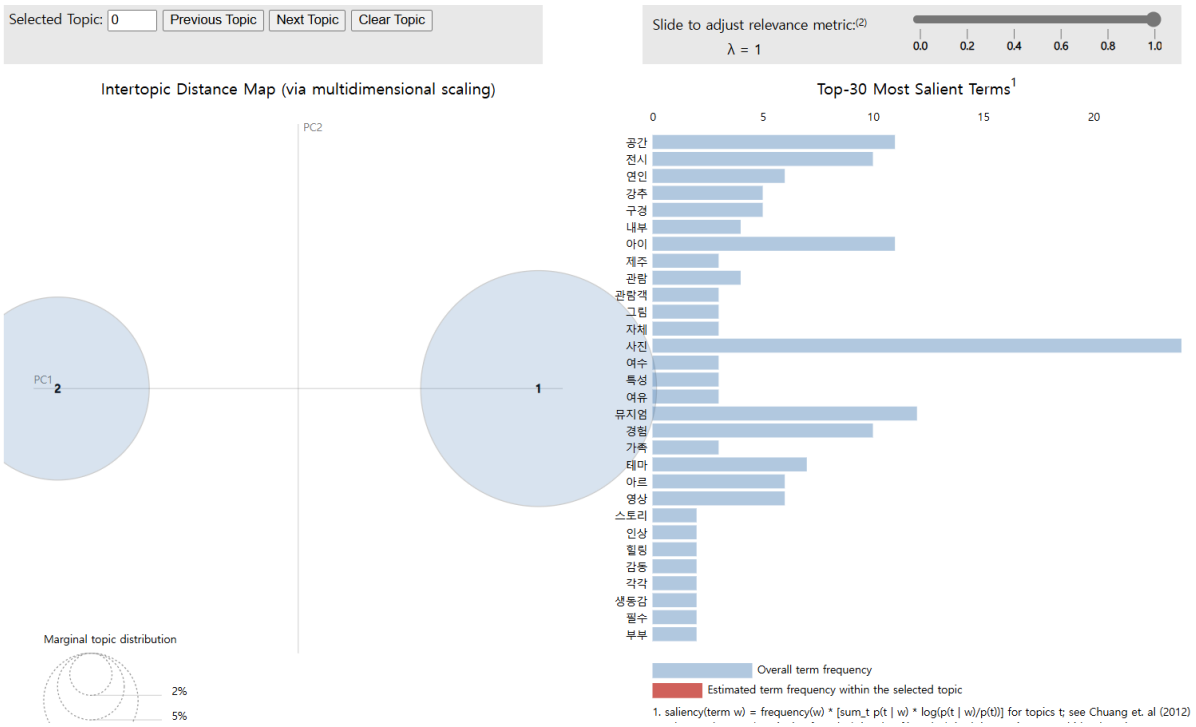
<경포대 중요 토픽>



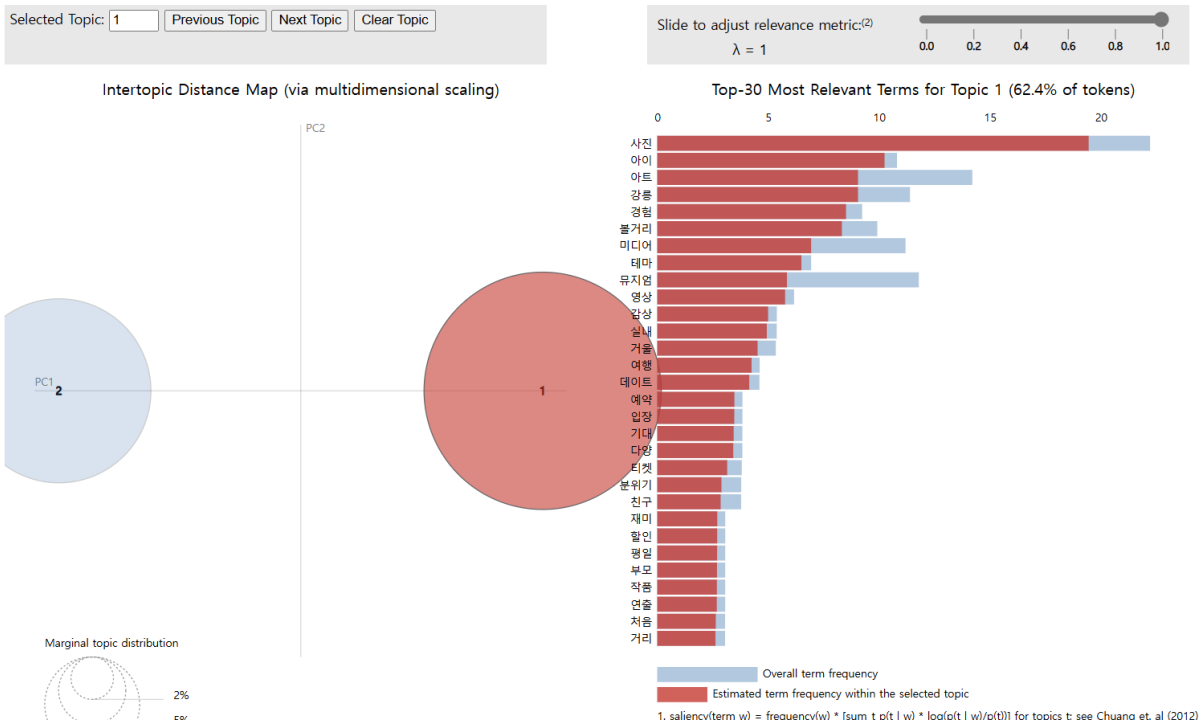
<대관령 양떼목장>



<대관령 양떼목장 중요 토픽>



<아르떼 뮤지엄>



<아르떼 뮤지엄 중요 토픽>

◦ 연구 결과 분석

<관광지 별 주요 토픽>

- **경포대:** 해변, 바다, 산책, 바람, 조망
- **대관령 양떼목장:** 아이, 체험, 사진, 먹이, 야외
- **아르떼뮤지엄:** 아이, 공간, 관람, 내부, 사진

<주요 토픽 별 관광지>

- **자연, 경치:** 남이섬, 오대산, 치악산, 설악산
- **체험, 활동:** 양떼목장, 알파카월드, 스카이워크
- **역사, 문화:** 오죽헌, 김유정 문학촌, 국립춘천박물관
- **바다, 해변:** 경포대, 서피비치, 촛대바위
- **아이, 가족:** 레고랜드, 비발디파크, 레일바이크
- **야외:** 오션월드, 출렁다리, 설악 워터피아
- **실내, 공간, 사진:** 아르떼뮤지엄, 하슬라아트월드, 뮤지엄산

◦ 주요 토픽과 그 해석

- **자연과 경치**
  이 토픽은 남이섬, 오대산, 치악산, 설악산 등 자연경관이 아름다운 관광지에서 자주 언급되었다. 이 관광지들의 공통된 키워드는 '자연', '경치'로, 관광객들이 강원도 방문 시에 자연의 아름다움과 계절별로 바뀌는 전경에 감동받고 이를 즐기는 모습을 보여준다. 특히 가을 단풍이나 산에서의 산책, 트레킹이 주요 활동으로 드러난 것으로 볼 수 있다.

- **체험과 활동**
  양떼목장, 알파카월드, 스카이워크 등은 체험형의 관광지로, 방문객들은 '체험', '활동'이라는 키워드를 많이 언급했다. 이 토픽은 가족 단위 관광객이나 아이들과 함께하는 여행에서 특히 중요한 요소로 나타나고, 동물 먹이 주기나 스카이워크에서의 스릴 넘치는 경험 특징이다.

- **역사와 문화**
  오죽헌, 김유정 문학촌, 국립춘천박물관 등에서는 '역사', '문화', '교육'이 주요 키워드로 도출되었다. 관광객들이 역사적 인물과 관련된 장소를 방문하며 교육적인 가치와 역사적인 배움을 느끼고, 가족 단위로 학생을 동반한 여행에 적합한 장소임을 보여준다.

- **바다와 해변**
  경포대, 서피비치, 촛대바위에서는 '바다', '해변'이 주요 키워드로 나타났다. 여름철 해변 활동이

나 해안 경치 감상이 강원도의 주요 매력 중 하나를 나타낸 것으로 볼 수 있다. 관광객들은 해양 스포츠, 해변 산책, 일몰 관람 등의 활동을 즐기는 것으로 분석할 수 있다.

- **아이와 가족 중심 관광지**
  레고랜드, 비발디파크, 강촌 레일바이크는 '아이', '가족'과 관련된 키워드가 두드러진다. 이러한 관광지는 가족 단위 방문객들이 즐길 수 있는 다양한 활동과 시설을 제공하며, 놀이와 체험을 통해 가족 간의 유대감을 강화하는 특징이 있다.

- **야외 활동**
  오션월드, 출렁다리, 설악 워터피아는 '야외'라는 키워드로 묶이고, 방문객들이 액티브한 야외 활동을 선호하는 것으로 나타났다. 특히, 여름철이나 따뜻한 계절에 물놀이, 하이킹, 야외 스포츠를 즐기는 모습을 볼 수 있다.

- **실내 공간과 사진**
  아르떼뮤지엄, 하슬라아트월드, 뮤지엄산은 '실내', '공간', '사진' 키워드가 주요 특징으로 나타났다. 이 장소들은 실내에서 관람할 수 있는 예술적이고 창의적인 공간을 제공하고, 방문객들이 사진을 찍고 SNS에 공유할 만큼 인상적인 장소임을 보여준다. 예술 작품이나 독특한 전시 공간은 방문객들에게 시각적 즐거움을 선사하는 것을 볼 수 있다.

◦ 최종 프로그램

```
검색할 단어를 입력하세요 (종료하려면 '종료' 입력): 역사
'역사'와(과) 관련된 장소:
['오죽헌']
검색할 단어를 입력하세요 (종료하려면 '종료' 입력): 힐링
'힐링'와(과) 관련된 장소:
['강촌 레일파크', '김유정 문학촌', '서피비치', '오대산', '치악산']
검색할 단어를 입력하세요 (종료하려면 '종료' 입력): 카페
'카페'와(과) 관련된 장소:
['뮤지엄산', '젊은달와이파크']
검색할 단어를 입력하세요 (종료하려면 '종료' 입력): 가족
'가족'와(과) 관련된 장소:
['남이섬', '레고랜드', '비발디파크', '설악 워터피아', '소금산 출렁다리', '오션월드', '하슬라 아트월드']
검색할 단어를 입력하세요 (종료하려면 '종료' 입력): 종료
프로그램을 종료합니다.
```

최종적으로 단어(키워드)를 입력하면 여행지를 추천해주는 프로그램을 만들었다.

사용자가 특정 키워드를 입력하면, 그와 관련된 강원도 관광지를 추천해주는 시스템이다.

제4장 심층 분석 및 해석

◦ 주요 발견에 대한 심층 분석

계절적 요소의 중요성

- 특히 가을, 겨울 등 특정 계절에 관광객의 언급이 많이 나온 것을 보아, 시즌별 관광 프로모션

이나 계절별 체험 프로그램을 개발할 필요성이 있다고 생각한다.

아이와 가족 중심 관광지의 인기

- '아이', '가족' 키워드가 레고랜드, 양떼목장 등에서 많이 언급된 것은 가족 단위 방문객이 강원도의 주요 타겟층임을 알 수 있다. 이를 통해서, 어린이와 가족이 함께 즐길 수 있는 체험과 프로그램을 추가하거나 개선하는 전략이 필요하다는 아이디어를 제공한다.

SNS에 최적화된 관광지의 부각

- '사진'과 관련된 키워드가 아르떼뮤지엄, 뮤지엄산 등에서 두드러진 것은 젊은 연령층의 방문객들이 소셜 미디어에 공유할 수 있는 시각적으로 인상적인 장소를 선호한다는 것을 나타낸다. 이는 관광지의 마케팅 전략으로 사진 촬영 포인트나 독창적인 장식을 강조할 필요가 있다는 것을 보여준다.

해양 관광의 계절적인 트렌드

- '바다', '해변' 키워드는 여름철 해양 관광지의 인기를 보여주며, 강원도가 계절에 맞는 해양 스포츠 및 해변 관련 이벤트를 더 강화할 필요가 있음을 보여준다고 볼 수 있다. 특히 해변에서의 축제, 캠핑 등 계절 한정 이벤트는 방문객 수를 크게 증가시킬 수 있을 것이다.

제5장 결론

◦ 연구 결론

<관광지 특징>

키워드를 통해 관광지의 특징과 이미지를 알 수 있었다.

<공통된 키워드>

24개 관광지 리뷰 속 공통된 키워드는 '구경', '날씨', '가족', '여행'

<차별화된 키워드>

관광지별로 차별화된 키워드는 '역사', '실내', '체험', '자연'

◦ 연구의 활용 가능성

맞춤형 마케팅 전략 수립: 분석 결과로 도출된 주요 관심사와 선호 주제는 관광객들의 요구를 파악하는 데 도움이 된다. 이를 기반으로 관광지의 장점을 강조하는 마케팅 캠페인이나, 특정 계절이나 트렌드에 맞게 홍보 콘텐츠를 제작하기에 용이하다고 생각한다.

관광 상품 개발: 관광객들이 리뷰에서 언급한 활동이나 개선점을 토대로 새로운 관광 상품이나 체험 프로그램을 기획할 수 있다. 예를 들어, 자연경관을 즐기는 관광객이 많다면 생태 체험 투어를 강화할 수 있고, 지역 특산물에 대한 관심이 높다면 음식 체험 투어를 추가하는 것처럼 말이다.

관광지 개선 및 관리: 리뷰에 언급된 문제점이나 불만 사항을 분석해 관광지 관리와 서비스 개선에 활용한다면 관광지의 서비스 품질을 높이고, 재방문율을 증가시킬 것이다.

정책 결정 지원: 관광객의 피드백을 반영하여 시설 투자, 기반 시설 확충, 환경보호 정책 등 실질적인 개선 방안을 마련하는 것처럼, 분석 결과는 지방자치단체나 관광 관련 공공기관이 관광 정책을 수립하는 데 근거자료로도 활용될 수 있다.

관광 트렌드 분석: 장기적으로 데이터 분석을 반복 수행해서 관광 트렌드의 변화를 모니터링하고, 관광객의 관심이 이동하는 패턴을 파악할 수도 있을 것이다. 이는 관광 시장의 경쟁력을 유지하고 빠르게 대응하는 데 도움이 될 것으로 보인다.

◦ 결과에 대한 비판적 분석

구글맵과 네이버 지도의 방문 리뷰 데이터를 사용했지만, 이 리뷰는 특정 관광지나 특정 시점에 집중되었을 수도 있다. 예를 들어서, 인기 있는 관광지나 계절적 요인에 따라 특정 기간에 많은 리뷰가 작성될 수 있으므로 데이터가 편향될 가능성이 있다.

방문자들이 남긴 리뷰가 짧고 비정형적인 경우가 여럿 있어서 일부 중요한 정보가 누락되었을 수 있다. 또한, 사용자가 직접 작성한 리뷰는 문법적으로 오류가 있거나 의미가 불명확한 경우도 있어서 자연어 처리에서 어려움을 겪었을 수도 있다.

◦ 향후 연구 제안

리뷰 데이터 외에도 인스타그램, 네이버 블로그 등의 소셜 미디어에서 언급되는 관광지와 관련된 데이터도 추가적으로 분석하면 더 좋은 효과를 볼 수 있을 것 같다. 소셜 미디어는 실시간 트렌드와 감성 분석에 유리하기 때문에 더욱 풍부한 관광객의 의견을 반영할 수 있다고 본다.

여건이 된다면 감성 분석의 기능을 추가하여 각 관광지에 대한 긍정적, 부정적, 중립적인 감정을 분류하고, 이를 통해 관광지에 대한 전반적인 이미지나 평판을 파악하고, 그에 따른 마케팅 전략을 세우면 더 도움이 될 것 같다.

추천 알고리즘을 적용하여 특정 관광지나 활동을 개인화된 추천 목록으로 제공하는 시스템을 개발할 수도 있다. 리뷰 분석을 통해 관광객의 취향을 파악하고, 이를 바탕으로 강원도 내의 더욱 다양한 관광지나 활동, 체험을 추천하는 서비스를 만들 수 있을 것이다.

향후 연구에서는 타지역의 관광지와 강원도의 주요 키워드를 비교 분석해서 강원도 관광의 차별성과 독창적인 매력을 강조한다면 또 다른 홍보 방법 중 하나가 될 것이다. 이를 통해 강원도의 경쟁력 있는 관광 요소를 부각하고, 지역 맞춤형인 전략적 수립에 기여할 수 있을 것으로 본다.

제6장 참고 문헌 및 부록

LDA 토픽 모델링으로 콘텐츠 리뷰를 분석하자

https://velog.io/@mare-solis/LDA-%ED%86%A0%ED%94%BD-%EB%AA%A8%EB%8D%B8%EB%A7%81%EC%9C%BC%EB%A1%9C-%EC%BD%98%ED%85%90%EC%B8%A0-%EB%A6%AC%EB%B7%B0%EB%A5%BC-%EB%B6%84%EC%84%9D%ED%95%98%EC%9E%90

BERT 기반 복합 토픽 모델

https://wikidocs.net/161310

파이썬 텍스트 마이닝 토픽 모델링

https://blog.naver.com/j7youngh/222915890492

경포대.html

대관령 양떼목장.html

아르떼뮤지엄.html