

빅데이터 분석과 응용 팀 프로젝트

여행, 어디까지 가 봤니?

빅포텐셜

60191096 고수빈 60180800 박재연
60171100 윤민혁 60180909 이신행

여행, 어디까지 가 봤니 ?

CONTENTS

01 주제 선정 이유

02 주제 소개

03 데이터 수집 및 전처리

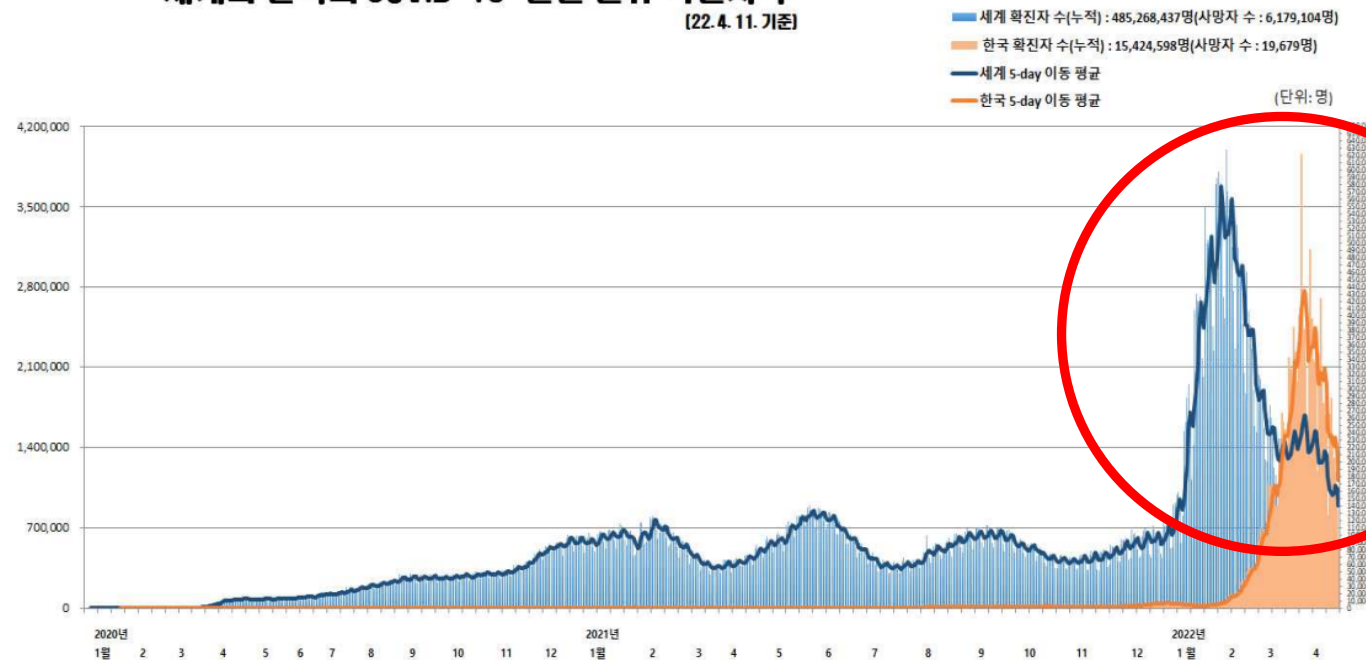
04 실행

05 결론

01. 주제 선정

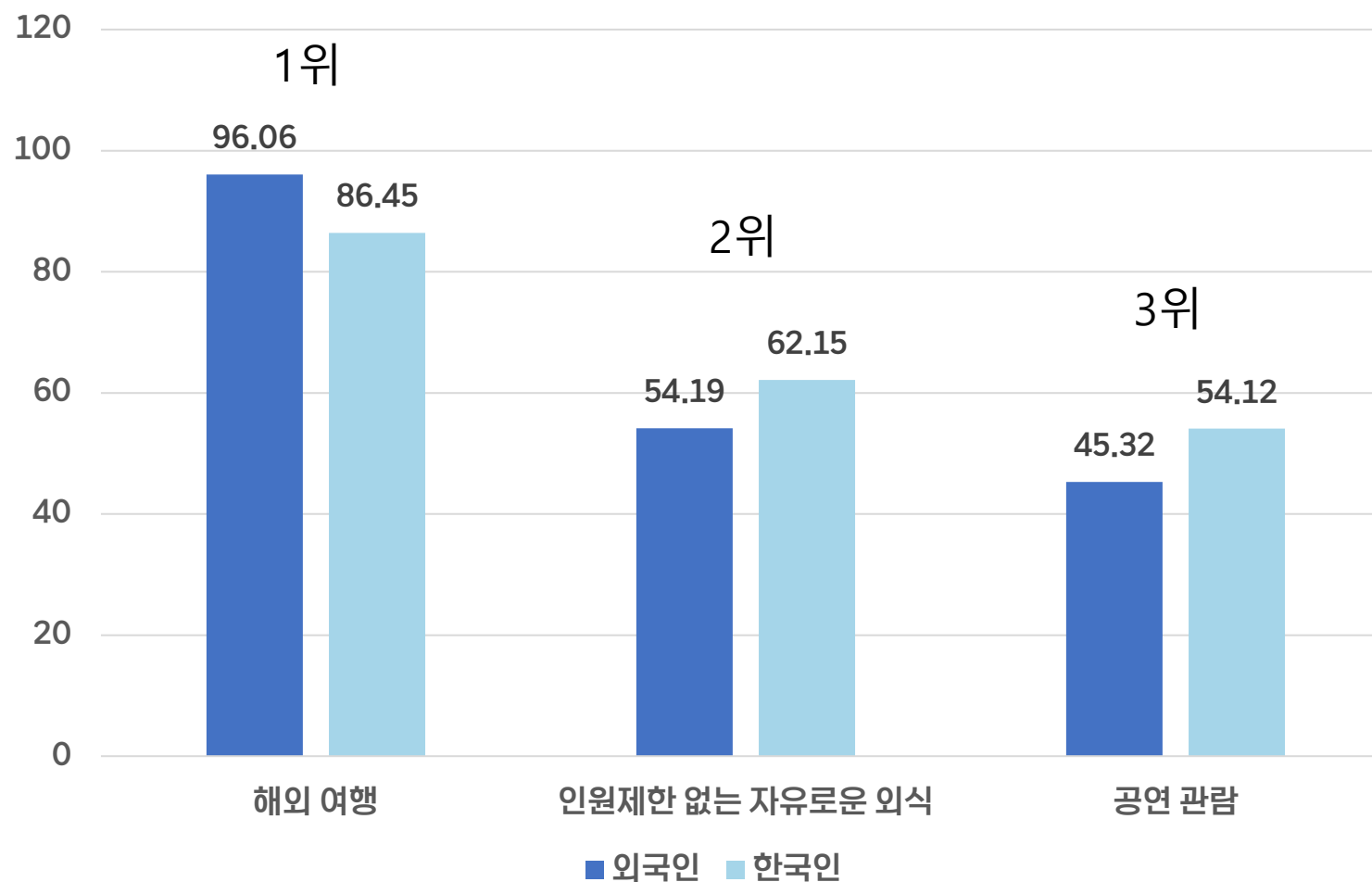
출처: 한국관광 데이터랩 - 2020 잠재 방한 여행객 조사

세계와 한국의 COVID-19 일별 신규 확진자 수
[22. 4. 11. 기준]



- 코로나 감소 추세로 인한 '앤데미크' 시대 기대감에 불붙은 여행 심리

코로나가 끝나면 하고싶은 것

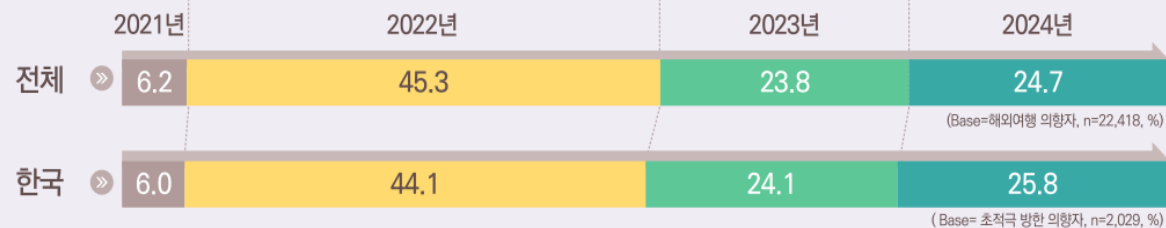




- 외국 관광객들의 한국 여행 증가 예상

-> BTS, 드라마 '오징어 게임' 등
K-콘텐츠의 전 세계적 인기로
한국을 관광 목적지로 고려하는데
긍정적 영향

향후 3년 내(~2024년) 첫 번째 해외여행 희망 목적지 방문 예상 시기(1순위 기준)



향후 3년 내(~2024년) 첫 번째 해외여행 희망 목적지(1순위 기준)

(Base=해외여행 의향자, n=22,418, %)

전체(상위 5위)



- 외국 관광객들의 한국 여행 증가 예상

-> 향후 3년 내 방한 여행 의향 있는 외국인
" 47% "

-> 한국은 첫 번째 해외여행 희망 목적지
" 2위 "

외국인의 한국 관광을 위한 "여행지 추천 시스템"



02 . 주제 소개

여행, 어디까지 가 봤니?

- 여기 가봤으면, 여기는 어때? -

03. 데이터 수집 및 전처리

1. 국내 여행지 목록 데이터 수집

- 나무위키의 '죽기 전에 가야할 국내 여행 1001'



| 여행지 | |
|------|------------|
| 0 | 경복궁 |
| 1 | 삼청동 거리 |
| 2 | 인사동 |
| 3 | 조계사 |
| 4 | 청와대 |
| ... | ... |
| 1000 | 쇠소깍 |
| 1001 | 백조일손지묘 |
| 1002 | 휴애리 자연생활공원 |
| 1003 | 안덕계곡 |
| 1004 | 제주 감귤박물관 |

999 rows × 1 columns

2. 국내 여행지에 대한 리뷰 데이터 수집

- '트립 어드바이저' 사이트에서 웹 크롤링



 Tripadvisor



서울 호텔 즐길거리 음식점 항공권 쇼핑 항공+호텔 패키지 크루즈 렌터카 ...

아시아 > 대한민국 > 서울 > 서울의 즐길거리 > 경복궁

경복궁

●●●●● 10,656 • 서울의 즐길거리 1,010개 중에서 7위 • 역사적인 장소 • 역사박물관

영업시간 종료 • 오전 9:00~오후 5:00 웹사이트 접속 > 전화



경복궁

모든 결과

호텔

음식점

즐길거리

도시 및 지역

"경복궁"과(와) 일치하는 검색 결과



평점

경복궁



10,656건의 리뷰

종로구 사직로 161, 서울, 대한민국

리뷰

주소

카테고리

역사적인 장소

① 크롤링 기본 코드

```
1
2 options=webdriver.ChromeOptions()
3 options.add_argument('user_agent=Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/101.0.495
4 options.add_experimental_option("excludeSwitches", ["enable-logging"])
5
6
7 driver=webdriver.Chrome()
8 url='https://www.tripadvisor.co.kr/'
9
10 reviews=[]
11 for i in range(len(df_c)):
12     driver.get(url)
13     search=df_c.여행지[i]
14     driver.find_element_by_css_selector('#lithium-root > main > div.dqRmR.dcDXR.dJjeH > div > div > div.weilG.ZD.Wh.fRhgZ > form >
15     driver.find_element_by_css_selector('#lithium-root > main > div.dqRmR.dcDXR.dJjeH > div > div > div.weilG.ZD.Wh.fRhgZ > form >
16     driver.switch_to.window(driver.window_handles[-1])
17     time.sleep(2)
18     try:
19         driver.find_element_by_css_selector('#BODY_BLOCK_JQUERY_REFLOW > div.page > div > div.ui_container.main_wrap > div > div >
20
21         driver.switch_to.window(driver.window_handles[-1])
22         soup=BeautifulSoup(driver.page_source,'html.parser')
23         mydata=soup.find_all('span','NejBf')
24         for item in mydata:
25             reviews.append(search + '-->' + item.get_text().strip())
26         driver.switch_to.window(driver.window_handles[-1])
27         driver.close()
28         driver.switch_to.window(driver.window_handles[0])
29         driver.back()
30         time.sleep(1)
31     except:
32         pass
```

② 리뷰 추출

```

| reviews
executed in 21ms, finished 22:06:03 2022-05-14
['경복궁-->대한민국의 역사',
'경복궁-->대한민국의 역사가 잠들어 있는 곳. 서울을 방문했다면 꼭 방문해야 되는 곳. 경복궁은 우리의 역사다. 넓은 경복궁을 산책할 수 있는 것은 언제나 행복이다.',
'경복궁-->국민이 공감하는 장소',
'경복궁-->경복궁은 국민들이 자주 찾는곳으로 작성자는 주말에 자주 가족들과 방문하고 있음.특히 가족들과 방문시 옛고궁의 멋을 즐길수 있으며 쾌적한 공기와 역사를 체험하고 힐링할수 있는곳임.',
'경복궁-->산책하기 좋은 경복궁',
'경복궁-->날씨 좋은 날 종종 산책하러 경복궁에 가는데 마음이 편해지는 기분이라고 할까요? 특히 봄이나 가을에 산책가는 걸 추천해요!',
'경복궁-->Good',
'경복궁-->Gooooooooo 다 좋습니다 다음에 또 오고 싶네요 근처 관광지도많고 먹을거리도많네요 서울올때마다 자주이용중인데 또오고싶습니다.',
'경복궁-->가족단위로 방문하기 좋은곳',
'경복궁-->요새 더더욱 코로나로 인해 사람 방문이 적음. 두자녀 동반시 성인 입장무료. 지금 시기에 나들이 하기 딱 좋은곳. 지하철 광화문역이나 종각역이용 편리.',
'경복궁-->하늘이 내린 큰 복',
'경복궁-->조선 개국 4년째인 1395년에 처음으로 세운 으뜸 궁궐이다.'하늘이 내린 큰 복' 이라는 뜻으로 경복궁이라 이름지었다.가장 아름답고 웅장한 이 경복궁은 임진왜란, 일제강점기 2번에 걸쳐 일본의 약탈하고 잔인한 파괴를 당한 악연을 갖고있다.향원정은 보수 공사로 인해 볼 수 없지만 아름다운 경회루만으로도 충분한 경복궁이다.',
'경복궁-->곳',
'경복궁-->한국 건축물의 아름다움을 느낄 수 있는 곳이고, 한복을 입었을 때 입장료는 무료.주변에 한복 대여점이 많으니 한번쯤 대여해서 입어보는것도 나쁘지 않을듯',
'경복궁-->조선의 집',
'경복궁-->조선의 성곽 안에는 박물관도 있고 경회루 등 볼것이 많다. 외국인들 및 연인들도 주변에서 한복 및 특수복장 빌려서 성을 둘러보며 데이트를 즐기기에 딱 좋다. 야간개장도 하니 잘 알아보고 올것 한국사람이라면 한번이상은 꼭와봤고 , 외국인들에게도 좋은 관광지라고 생각한다. 앞에 광화문 광장 청계천 위로 삼청동 이주변에 연계된 명소들도 많기 때문에 꼭 추천한다.',
'경복궁-->대표 문화재',
'경복궁-->대부분의 사람들이 여행을 간다면 그 나라의 유서 깊은 궁전을 찾게 된다.한국에서는 그 곳이 바로 경복궁이다.지금의 경복궁은 원래의 것보다 많이 축소되었지만 그래도 깨끗이 잘 정돈되어 있으며 위엄이 있다.',
'경복궁-->경복궁',
'경복궁-->몇번을 가도 감탄하게 되며 산책하기 최고인 경복궁 매번 갈때마다 정교함과 운치에 또한번 반하게 되는곳 주변에 볼거리도 많음',
'삼청동 거리-->기대 이상 만족',
'삼청동 거리-->저번달에 아주 오랜만에 다시 방문. 10년만에 재방문인데 그사이에 카페나 작은 삼점 옷가게가 눈에 많이 띄었다. 서울내에 한옥건물 제발 보존 해주세요',
'삼청동 거리-->삼청동길',
'삼청동 거리-->삼청동길 한옥건물들을 볼수있는 거리중 한곳다만 예전에 비해 상권이 많이 없어졌어요..그래도 걸으면서 구경하기에는 좋아요.',
'삼청동 거리-->전통과 현대의 건축양식의 조화가 주는 이색적인 도시 풍경',
'삼청동 거리-->이 곳에서는 기와나 목재 등을 이용해 한국의 전통적인 느낌을 살린 현대식 건물이 많아서 이색적인 도시 풍경을 감상할 수 있다. 걷기만 해도 재미있다.',

```


③ 평점 카테고리 크롤링

```

1 options=webdriver.ChromeOptions()
2 options.add_argument('user_agent=Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/101.0.495
3 options.add_experimental_option("excludeSwitches", ["enable-logging"])
4
5
6 driver=webdriver.Chrome()
7 url='https://www.tripadvisor.co.kr/'
8
9 category=[]
10 score = []
11 for i in range(len(df_2)):
12     driver.get(url)
13     search=df_2.여행지[i]
14     driver.find_element_by_css_selector('#lithium-root > main > div.dqRmR.dcdXR.dJjeH > div > div > div.weilG.ZD.Wh.fRhgZ > form >
15     driver.find_element_by_css_selector('#lithium-root > main > div.dqRmR.dcdXR.dJjeH > div > div > div.weilG.ZD.Wh.fRhgZ > form >
16     driver.switch_to.window(driver.window_handles[-1])
17     time.sleep(2)
18     try:
19         driver.find_element_by_css_selector('#BODY_BLOCK_JQUERY_REFLOW > div.page > div > div.ui_container.main_wrap > div > div >
20
21         driver.switch_to.window(driver.window_handles[-1])
22         soup=BeautifulSoup(driver.page_source, 'html.parser')
23         mydata=soup.find('div', 'pIRBY_T KRIav')
24         category.append(search + '-->' + mydata.get_text())
25         mydata2=soup.find('div', 'WIVyy cPsXC fksET cMKsg')
26         score.append(search + '-->' + mydata2.get_text())
27         driver.switch_to.window(driver.window_handles[-1])
28         driver.close()
29         driver.switch_to.window(driver.window_handles[0])
30         driver.back()
31         time.sleep(1)
32     except:
33         category.append('결과 없음')
34         score.append('결과 없음')

```

④ 평점과 카테고리 크롤링 결과

| | 카테고리 | 평점 |
|----|--|--------------------------|
| 0 | 암사동 선사주거지-->역사적인 장소 | 암사동 선사주거지-->4.5 |
| 1 | 화성-->역사적인 장소 | 화성-->4.0 |
| 2 | 대부도-->섬 | 대부도-->4.0 |
| 3 | 국립현대미술관(과천관)-->아트뮤지엄 | 국립현대미술관(과천관)-->4.5 |
| 4 | 독립기념관-->역사박물관 | 독립기념관-->4.5 |
| 5 | 공산성-->역사적인 장소 • 줄길거리 & 랜드마크 | 공산성-->4.5 |
| 6 | 무령왕릉-->역사적인 장소 | 무령왕릉-->4.0 |
| 7 | 충청수영성-->역사적인 장소 • 성 • 경관이 좋은 산책로 | 충청수영성-->5.0 |
| 8 | 천주교 대전교구 공세리성당-->종교적인 장소 | 천주교 대전교구 공세리성당-->4.5 |
| 9 | 팜카밀레 허브농원-->공원 • 가든 | 팜카밀레 허브농원-->5.0 |
| 10 | 천주교 전주교구 전동성당-->종교적인 장소 | 천주교 전주교구 전동성당-->4.0 |
| 11 | 무주 구천동계곡 방화동 가족휴양촌-->국립공원 • 자연 & 야생동물 서식지 • 계곡 | 무주 구천동계곡 방화동 가족휴양촌-->5.0 |
| 12 | 순천만 갈대밭-->공원 | 순천만 갈대밭-->4.5 |
| 13 | 용천사-->역사적인 장소 • 종교적인 장소 • 경관이 좋은 산책로 | 용천사-->4.0 |
| 14 | 계림-->유서 깊은 산책로 | 계림-->4.0 |
| 15 | 금오산-->공원 | 금오산-->4.5 |
| 16 | 빙계계곡-->물길 • 산림 • 계곡 | 빙계계곡-->5.0 |
| 17 | 군위 삼존석굴-->역사적인 장소 • 종교적인 장소 | 군위 삼존석굴-->4.5 |
| 18 | 주남저수지-->자연 & 야생동물 서식지 | 주남저수지-->4.0 |
| 19 | 청곡사-->종교적인 장소 | 청곡사-->3.5 |
| 20 | 고성 탈박물관-->전문 박물관 | 고성 탈박물관-->4.0 |

⑤ 주소 크롤링 결과

| | 0 | 행정구역 | 관광지 |
|----|---|------|---------------------|
| 0 | 암사동 선사주거지-->강동구 올림픽로 875, 서울, 대한민국 | 서울 | 암사동 선사주거지 |
| 1 | 화성-->팔달구 정조로 825, 수원, 경기도, 대한민국 | 경기도 | 화성 |
| 2 | 대부도-->2-3 Kitanomaruken, 지요다, 도쿄, 도쿄도, 관동, 일본 | 관동 | 대부도 |
| 3 | 국립현대미술관(과천관)-->광명로 313, 과천, 경기도, 대한민국 | 경기도 | 국립현대미술관(과천관) |
| 4 | 독립기념관-->충남 천안시 동남구 목천읍 독립기념관로 1 | NaN | 결과 없음 |
| 5 | 공산성-->280 Ungjin-ro, Gongju-si, Chungcheongnam... | 충청남도 | 공산성 |
| 6 | 무령왕릉-->왕릉로 37, 공주, 충청남도, 대한민국 | 충청남도 | 무령왕릉 |
| 7 | 충청수영성-->오천면 소성리 931, 보령, 충청남도, 대한민국 | 충청남도 | 충청수영성 |
| 8 | 전주교 대전교구 공세리성당-->인주면 공세리성당길 10, 아산, 충청남도, 대한민국 | 충청남도 | 전주교 대전교구 공세리성당 |
| 9 | 팜카밀레 허브농원-->남면 우운길 56-19, 태안, 충청남도, 대한민국 | 충청남도 | 팜카밀레 허브농원 |
| 10 | 전주교 전주교구 전동성당-->200-1, Jeon-dong, Wansan-gu, 전... | 전라북도 | 전주교 전주교구 전동성당 |
| 11 | 무주 구천동계곡 방화동 가족휴양촌,-->설천면 구천동1로 159, 무주, 전라북도,... | 전라북도 | 무주 구천동계곡 방화동 가족휴양촌, |
| 12 | 순천만 갈대밭-->대대동 162-2, 순천, 전라남도, 대한민국 | 전라남도 | 순천만 갈대밭 |
| 13 | 용천사-->해보면 용천사길 209, 함평, 전라남도, 대한민국 | 전라남도 | 용천사 |
| 14 | 계림-->1, Gyo-dong, 경주, 경상북도, 대한민국 | 경상북도 | 계림 |
| 15 | 금오산-->경북 구미시 남통동 288-2 | NaN | 결과 없음 |
| 16 | 빙계계곡-->춘산면 빙계리 896, 의성, 경상북도, 대한민국 | 경상북도 | 빙계계곡 |
| 17 | 군위 삼존석굴-->302, Namsan-ri, Bugye-myeon, 군위, 경상북... | 경상북도 | 군위 삼존석굴 |
| 18 | 주남저수지-->의창구 등을 주남로101번길 26, 창원, 경상남도, 대한민국 | 경상남도 | 주남저수지 |
| 19 | 청곡사-->금산면 월아산로1440번길 138, 진주, 경상남도, 대한민국 | 경상남도 | 청곡사 |
| 20 | 고성 탈박물관-->고성을 읍대리 666-18, 경상남도, 대한민국 | 경상남도 | 고성 탈박물관 |

트립 어드바이저 사이트에서의 크롤링 과정

| | 여행지 | 주소 | 행정구역 | 시군구 | 분류 | 트립어드바이저 평점 | 카테고리 |
|-----|------------------|---------------------|---------|------|-------------|---------------|---------|
| 0 | 63스퀘어 | 영등포구 63로 50, 서울 | 서울시 | 영등포구 | 건축물 | 4 | 랜드마크 |
| 1 | BIFF 광장 | 중구 남포동, 부산 | 부산광역시 | 중구 | 즐길거리 & 랜드마크 | 4 | 랜드마크 |
| 2 | SEA LIFE 부산아쿠아리움 | 해운대구 해운대해변로 266, 부산 | 부산광역시 | 해운대구 | 수족관 | 4 | 랜드마크 |
| 3 | 가야산 국립공원 | 경남 합천군 가야면 | 경상남도 | 합천군 | 국립공원 | 4.5 | 공원 |
| 4 | 간월재 | 울주군 상북면, 울산 | 울산광역시 | 울주군 | 산림 | 4.5 | 산 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 653 | 활지 | 활지동 25-4, 태백 | 강원도 | 태백시 | 물길 | 4 | 자연경관 |
| 654 | 회암사지 | 회암사길 281, 양주, 경기도 | 경기도 | 양주시 | 종교적인 장소 | 5 | 종교적인 장소 |
| 655 | 휴애리 자연생활공원 | 남원읍 신례동로 256, 서귀포 | 제주특별자치도 | 서귀포시 | 공원 | 4 | 공원 |

--> " 최종 수집 데이터 "

- 여행지
- 주소
- 평점
- 카테고리

04. 실행

```
counts=Counter(noun_list)
tags=counts.most_common(70)
tags
```

```
[('방문', 164),
 ('마을', 125),
 ('아이', 118),
 ('정말', 113),
 ('사진', 112),
 ('사람', 109),
 ('생각', 94),
 ('추천', 91),
 ('산책', 85),
 ('바다', 84),
 ('장소', 81),
 ('가족', 76),
 ('전망대', 74),
 ('시간', 72),
 ('구경', 71),
 ('한번', 68),
 ('체험', 67),
 ('공원', 63),
```

```
from PIL import Image
image=Image.open('cloud_img.jpg')
mask=np.array(image)
```

```
## 워드 클라우드 생성
```

```
font_path='C:/Windows/Fonts/HMFMOLD.ttf'
```

```
from wordcloud import WordCloud
```

```
wordcloud=WordCloud(font_path=font_path,
                    background_color='Black',width=1200,height=800,colormap='spring',mask=mask)
```

```
print(dict(tags))
```

```
{'방문': 164, '마을': 125, '아이': 118, '정말': 113, '사진': 112, '사람': 109, '추천': 91, '산책': 85, '바다': 84, '장소': 81, '가족': 76, '전망대': 74, '시간': 72, '구경': 71, '한번': 68, '체험': 67, '공원': 63, '시장': 63, '부산': 60, '서울': 59, '거리': 59, '바람': 58, '근처': 55, '건물': 52, '풍경': 50, '조금': 49, '여행': 49, '이용': 46, '경치': 46, '코스': 44, '위치': 43, '모습': 43, '보고': 43, '언덕': 42, '볼거리': 41, '주변': 41, '정도': 40, '하나': 40, '카페': 40, '입구': 39, '날씨': 39, '주차': 37, '역사': 36, '식당': 36, '겨울': 35, '가게': 34, '바로': 34, '촬영': 33, '한국': 33, '입장료': 33, '골목': 32, '가격': 32, '주차장': 32, '다만': 31, '타고': 31, '경험': 31, '타워': 31, '테마': 31, '보기': 30, '시설': 30, '친구': 29, '먹거리': 29, '자연': 29}
```

토픽 모델링 과정

```
landmark=list(df3['여행지'].unique())
history=list(df4['여행지'].unique())
history
```

```
C:\Users\S00BIN\anaconda3\lib\site-packages\ipykernel\ipykernel.py:287: DeprecationWarning: `should_run_async` result to `transformed_cell` argument and any exception that happen during the transform in `preprocess` and should_run_async(code)
```

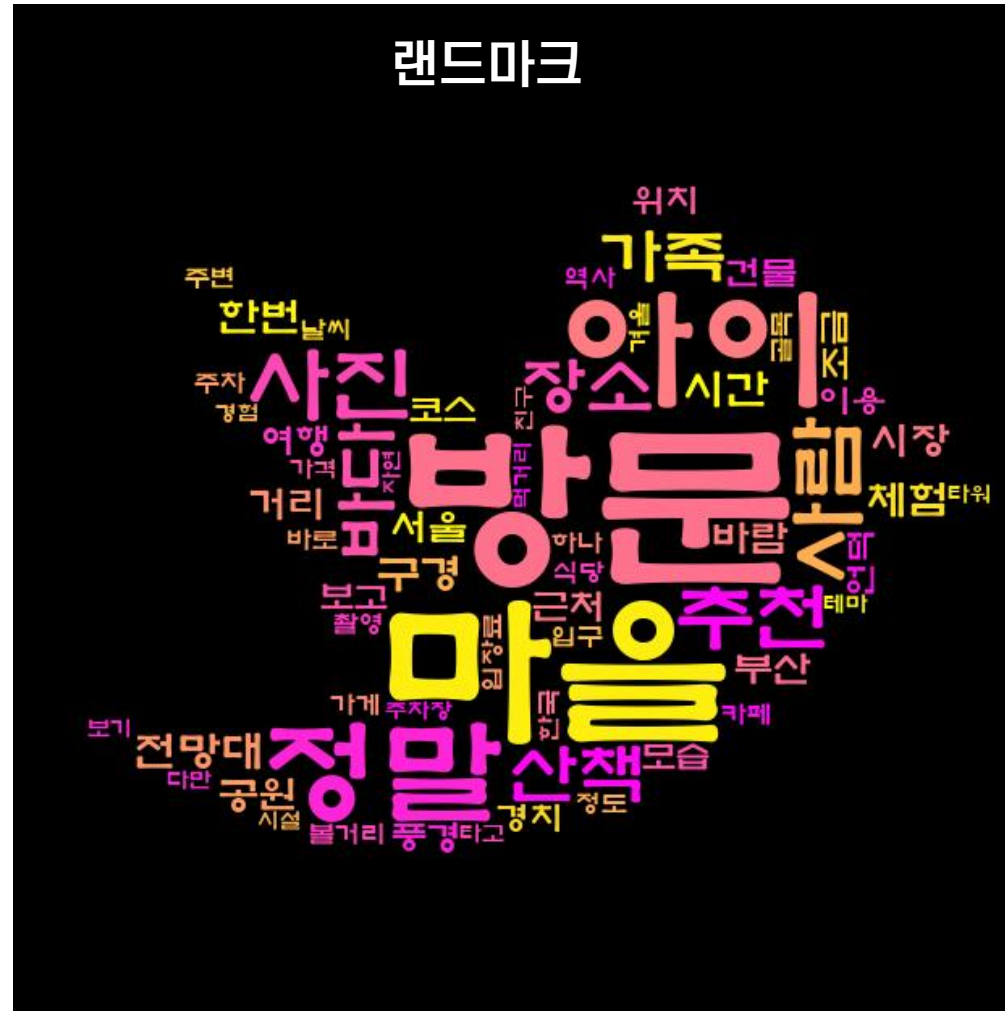
```
['감은사지',
'갑곶돈대',
'강릉선교장',
'개실마을',
'건봉사',
'경교장',
'경기전',
'경복궁',
'고려궁지',
'고모산성',
'고창읍성',
'광성보와 용두돈대',
'괘릉',
'과시리 전통마을',
'그지보']
```

```
pd.DataFrame(tokenized_data)
```

```
C:\Users\S00BIN\anaconda3\lib\site-packages\ipykernel\ipykernel.py:287: DeprecationWarning: `should_run_async` result to `transformed_cell` argument and any exception that happen during the transform in `preprocess` and should_run_async(code)
```

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | ... | 131 | 132 | 133 | 134 | 135 |
|-----|------|------|------|------|------|------|------|------|------|------|-----|------|------|------|------|------|
| 0 | 대한민국 | 역사 | 서울 | 방문 | 방문 | 경복궁 | 우리 | 역사 | 경복궁 | 산책 | ... | None | None | None | None | None |
| 1 | 경복궁 | 국민 | 자주 | 작성자 | 주말 | 자주 | 가족 | 방문 | 가족 | 방문 | ... | None | None | None | None | None |
| 2 | 날씨 | 종종 | 산책 | 경복궁 | 마음 | 해지 | 기분 | 가을 | 산책 | 추천 | ... | None | None | None | None | None |
| 3 | 다음 | 근처 | 관광지 | 거리 | 자주 | 이용 | None | None | None | None | ... | None | None | None | None | None |
| 4 | 요새 | 더욱 | 코로나 | 사람 | 방문 | 자녀 | 반시 | 성인 | 입장 | 무료 | ... | None | None | None | None | None |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 792 | 완전 | None | None | None | None | None | None | None | None | None | ... | None | None | None | None | None |
| 793 | 제주도 | 과거 | None | None | None | None | None | None | None | None | ... | None | None | None | None | None |
| 794 | 보기 | 적합 | None | None | None | None | None | None | None | None | ... | None | None | None | None | None |
| 795 | 역사 | None | None | None | None | None | None | None | None | None | ... | None | None | None | None | None |
| 796 | 일본 | 식민지 | 시대 | 군용 | 비행장 | None | None | None | None | None | ... | None | None | None | None | None |

워드 클라우드 - 랜드마크

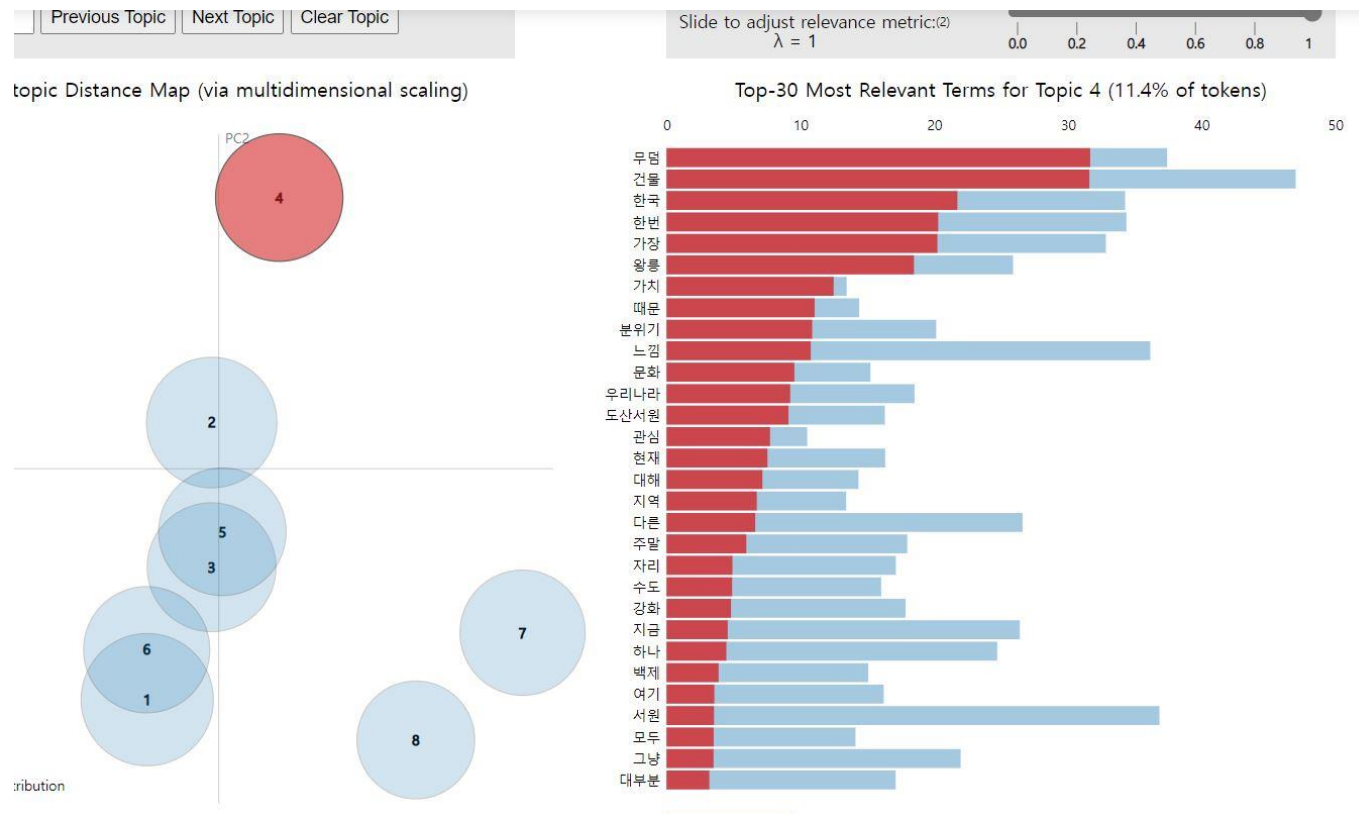


토픽 모델링 - 랜드마크

토픽명

단어

| | | |
|---|------------|--------------------|
| 0 | 편의시설 좋은 곳 | [주차장, 구경, 관람, 무료] |
| 1 | 서울 관련 랜드마크 | [입구, 버스, 서울, 모습] |
| 2 | 지형 관련 랜드마크 | [무덤, 건물, 한국, 한번] |
| 3 | 역사 관련 랜드마크 | [여행, 코스, 고분, 길이] |
| 4 | 세계적 관광지 | [최고, 세계, 유적지, 연꽃] |
| 5 | 소풍가기 좋은 곳 | [가을, 경주, 계곡, 나들이] |
| 6 | 체험 가능한 곳 | [사진, 체험, 주차, 전시] |
| 7 | 경치 좋은 곳 | [산성, 경치, 관리, 거리] |
| 8 | 풍경 좋은 곳 | [공원, 풍경, 가기, 조선시대] |



워드 클라우드 – 역사적인 장소



토픽 모델링 – 역사적인 장소

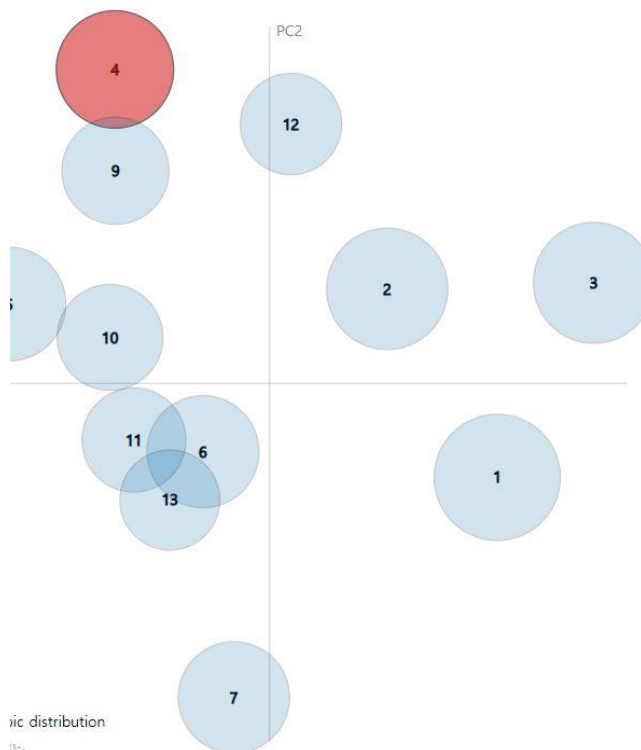
| 토픽 명 | 단어 |
|---------------|--------------------|
| 0 세계적인 역사 관광지 | [서울, 문화유산, 길이, 세계] |
| 1 | [한번, 바다, 구경, 전통] |
| 2 계절별 좋은 코스 | [코스, 여름, 가을, 겨울] |
| 3 편의시설 좋은 곳 | [무료, 의미, 근처, 주차] |
| 4 시대별 고분 | [모습, 무덤, 주차장, 시대] |
| 5 왕릉 주변 자연 | [왕릉, 나무, 연꽃, 주말] |
| 6 | [유적지, 가면, 산성, 사진] |
| 7 | [공원, 한국, 무덤, 바람] |
| 8 체험 가능 관광지 | [마음, 체험, 보기, 나들이] |
| 9 조선시대 관련 관광지 | [건물, 다산, 도산서원, 최고] |
| 10 | [서원, 여행, 가장, 고분] |
| 11 방문 느낌, 후기 | [풍경, 느낌, 공부, 기대] |
| 12 가까운 역사 관광지 | [버스, 위치, 강화, 유적] |

4 Previous topic Next topic Clear topic

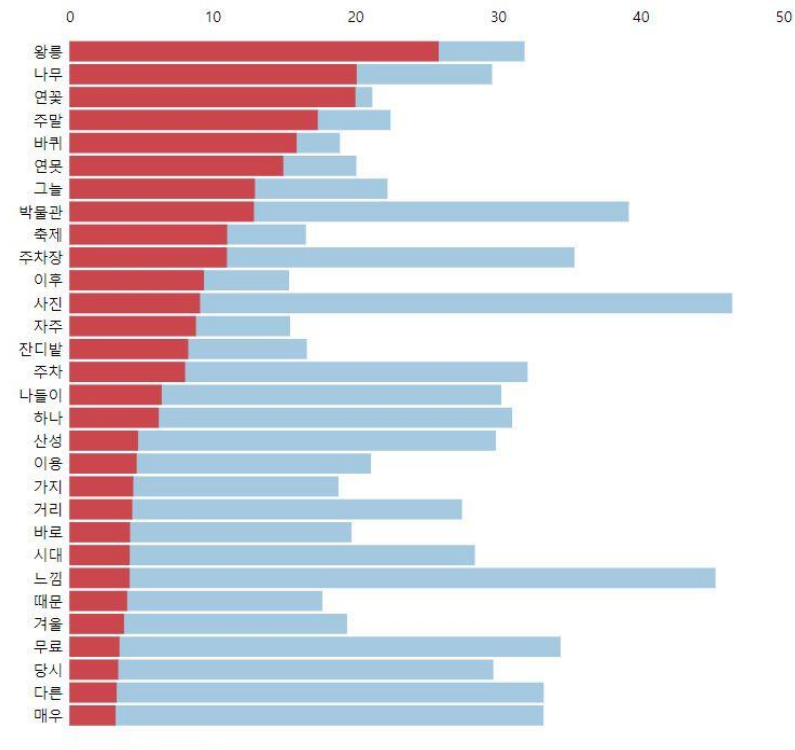
Slide to adjust relevance metric:(2)
 $\lambda = 1$

0.0 0.2 0.4 0.6 0.8 1

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 4 (8.5% of tokens)



```
pip install pororo # Pororo 설치
```

```
Attempting uninstall: regex  
Found existing installation: regex 2019.12.20  
Uninstalling regex-2019.12.20:  
Successfully uninstalled regex-2019.12.20  
Attempting uninstall: torchvision  
Found existing installation: torchvision 0.12.0+cu113  
Uninstalling torchvision-0.12.0+cu113:  
Successfully uninstalled torchvision-0.12.0+cu113  
Attempting uninstall: nltk  
Found existing installation: nltk 3.2.5  
Uninstalling nltk-3.2.5:
```

```
import time  
from pororo import Pororo
```

```
def sentiment_analysis(comment_list):  
    sa=Pororo(task='sentiment', lang='ko') # 한국어 형태로 분석  
  
    pos_prob, neg_prob=0.0, 0.0  
    for comment in review['전체내용']: #리뷰내용 칼럼에 있는 리뷰들 분석  
        sa_result=sa(comment[0], show_probs=True)  
  
        pos_prob += sa_result['positive']  
        neg_prob += sa_result['negative']  
  
    total_num = len(comment_list) # 크롤링한 댓글 수  
    pos_prob, neg_prob = pos_prob/total_num, neg_prob/total_num  
    pos_perc, neg_perc = round(pos_prob*100), round(neg_prob*100) #긍정확률, 부정확률  
  
    print('##### 댓글 선호도 분석 결과 #####')  
    print(f'긍정: {pos_perc}%, 부정: {neg_perc}%') #결과 출력
```

```
review['리뷰내용_1']=review['전체내용'].str[1]
review['리뷰내용_0']=review['전체내용'].str[0]
```

```
review['리뷰내용_0']
```

```
0          대한민국의 역사
1          국민이 공감하는 장소
2          산책하기 좋은 경복궁
3          Good
4          가족단위로 방문하기 좋은곳
...
5231      저는 녹차밭을 별로 기대하고 가지 않았습니...
5232          보성 녹차밭
5233      참 멋지고 이쁘네요말로만듣고 이번에 가봤는데 좋았어요또 가보고싶었고요추천해요혼자갔는...
5234          비오는 겨울 풍경도 아름다운 다원...
5235      비오는 겨울 다원 산책은 무모한 도전이었지만 또 다른 경험...다우너의 초록색은 없...
Name: 리뷰내용_0, Length: 5236, dtype: object
```

```
review['리뷰내용_1']
```

```
0      대한민국의 역사가 잠들어 있는 곳. 서울을 방문했다면 꼭 방문해야 되는 곳. 경복...
1      경복궁은 국민들이 자주 찾는곳으로 작성자는 주말에 자주 가족들과 방문하고 있음.특...
2      날씨 좋은 날 종종 산책하러 경복궁에 가는데 마음이 편온해지는 기분이라고 할까요?...
3      Gooooooooo 다 좋습니다 다음에 또 오고 싶네요 근처 관광지도많고 먹을거리도...
4      요새 더더욱 코로나로 인해 사람 방문이 적음. 두자녀 동반시 성인 입장무료. 지금...
...
5231      NaN
5232      NaN
5233      NaN
5234      NaN
5235      NaN
Name: 리뷰내용_1, Length: 5236, dtype: object
```

```
sentiment_analysis(review['리뷰내용_0'])
sentiment_analysis(review['리뷰내용_1'])
```

댓글 선호도 분석 결과

긍정: 67%, 부정: 33%

댓글 선호도 분석 결과

긍정: 67%, 부정: 33%

review

| | 관광지 | 전체내용 | 리뷰내용_1 | 리뷰내용_0 | 긍정확률_0 | 부정확률_0 | 긍정확률_1 | 부정확률_1 |
|------|--------------|--|--|--|----------|----------|----------|----------|
| 0 | 경복궁 | [대한민국의 역사, 대한민국의 역사가 잠들어 있는 곳. 서울을 방문했다면 꼭 방... | 대한민국의 역사가 잠들어 있는 곳 서울을 방문했다면 꼭 방문해 야 되는 곳 경복궁은... | 대한민국의 역사 | 0.381490 | 0.618510 | 0.951248 | 0.048752 |
| 1 | 경복궁 | [국민이 공감하는 장소, 경복궁은 국민 들이 자주 찾는곳으로 작성자는 주말에 자주... | 경복궁은 국민들이 자주 찾는곳 으로 작성자는 주말에 자주 가족 들과 방문하고 있음특히... | 국민이 공감하는 장소 | 0.850068 | 0.149932 | 0.974223 | 0.025777 |
| 2 | 경복궁 | [산책하기 좋은 경복궁, 날씨 좋은 날 중 종 산책하러 경복궁에 가는데 마음이 편... | 날씨 좋은 날 종종 산책하러 경복 궁에 가는데 마음이 편운해지는 기분이라고 할까요 ... | 산책하기 좋은 경복궁 | 0.916744 | 0.083256 | 0.976418 | 0.023582 |
| 3 | 경복궁 | [Good, Gooooood 다 좋습니다 다음에 또 오고 싶네요 근처 관광지도... | 다 좋습니다 다음에 또 오고 싶네 요 근처 관광지도 많고 먹을거리 도많네요 서울을때마... | | 0.137770 | 0.862230 | 0.984774 | 0.015226 |
| 4 | 경복궁 | [가족단위로 방문하기 좋은곳, 요새 더 더욱 코로나로 인해 사람 방문이 적음. 두... | 요새 더더욱 코로나로 인해 사람 방문이 적음 두자녀 동반시 성인 입장무료 지금 시... | 가족단위로 방문하기 좋은곳 | 0.931031 | 0.068969 | 0.690581 | 0.309419 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 5231 | 별교, 대한 다원 | [저는 녹차밭을 별로 기대하고 가지 않았 습니다.. 처음에는 안 들리고 지나가려고 했... | 리뷰없음 | 저는 녹차밭을 별로 기대하고 가지 않았습니다 처음에는 안 들리고 지 나가려고 했었는데... | 0.905600 | 0.094400 | 0.005893 | 0.994107 |
| 5232 | 별교, 대한 다원 | [보성 녹차밭] | 리뷰없음 | 보성 녹차밭 | 0.582564 | 0.417436 | 0.005893 | 0.994107 |

여행지 추천 시스템 도출

여행지 추천

```
data=pd.read_csv('진짜최종 데이터.csv')
data
```

| | 여행지 | 주소 | 행정구 역 | 시군 구 | 분류 | 트립어드 바이저평 점 | 카테고 리 | 최종긍정 | 최종부정 | Unnamed: 9 | Unnamed: 10 | Unnamed: 11 | Unnamed: 12 |
|---|----------------------|-------------------------|-----------|----------|--------------------|-------------------|----------|----------|----------|---------------|----------------|----------------|----------------|
| 0 | 63스퀘어 | 영등포구 63로 50, 서 울 | 서울시 | 영등 포구 | 건축물 | 4 | 랜드마 크 | 0.792686 | 0.207314 | NaN | NaN | NaN | NaN |
| 1 | BIFF 광장 | 중구 남포동, 부산 | 부산광 역시 | 중구 | 줄길거리 & 랜드마 크 | 4 | 랜드마 크 | 0.544369 | 0.455631 | NaN | NaN | NaN | NaN |
| 2 | SEA LIFE 부 산아쿠아리움 | 해운대구 해운대해변 로 266, 부산 | 부산광 역시 | 해운 대구 | 수족관 | 4 | 랜드마 크 | 0.714060 | 0.285940 | NaN | NaN | NaN | NaN |
| 3 | 가야산 국립공 원 | 경남 합천군 가야면 | 경상남 도 | 합천 군 | 국립공원 | 4.5 | 공원 | 0.897052 | 0.102948 | NaN | NaN | NaN | NaN |
| 4 | 간월재 | 울주군 상북면, 울산 | 울산광 역시 | 울주 군 | 산림 | 4.5 | 산 | 0.802439 | 0.197561 | NaN | NaN | NaN | NaN |

```
from sklearn.metrics.pairwise import cosine_similarity
cosine_sim=cosine_similarity(df,df)
cosine_sim
```

```
array([[1.          , 0.55069625, 0.56685942, ..., 0.35621017, 0.12458964,
        0.13979261],
       [0.55069625, 1.          , 0.7738173 , ..., 0.33308688, 0.1131351 ,
        0.11445813],
       [0.56685942, 0.7738173 , 1.          , ..., 0.34997955, 0.12135456,
        0.13224839],
       ...,
       [0.35621017, 0.33308688, 0.34997955, ..., 1.          , 0.12383507,
        0.13799701],
       [0.12458964, 0.1131351 , 0.12135456, ..., 0.12383507, 1.          ,
        0.12315416],
       [0.13979261, 0.11445813, 0.13224839, ..., 0.13799701, 0.12315416,
        1.          ]])
```


여행지 추천

```
get_recommendations('경복궁')
```

| | 여행지 | 행정구역 | 시군구 | 트립어드바이저평점 | 카테고리 | 최종긍정 | 최종부정 |
|---|--------------|------|-----|-----------|---------|----------|----------|
| 0 | 창덕궁 | 서울시 | 종로구 | 4.5 | 역사적인 장소 | 0.916274 | 0.083726 |
| 1 | 창경궁 | 서울시 | 종로구 | 4.5 | 역사적인 장소 | 0.936209 | 0.063791 |
| 2 | 국립현대미술관(서울관) | 서울시 | 종로구 | 4.5 | 박물관 | 0.951715 | 0.048285 |
| 3 | 낙산공원 | 서울시 | 종로구 | 4.5 | 공원 | 0.945897 | 0.054103 |
| 4 | 북악산 서울 한양도성 | 서울시 | 성북구 | 4.5 | 역사적인 장소 | 0.914828 | 0.085172 |
| 5 | 국립서울현충원 | 서울시 | 동작구 | 4.5 | 역사적인 장소 | 0.908730 | 0.091270 |
| 6 | 경교장 | 서울시 | 종로구 | 4 | 역사적인 장소 | 0.902907 | 0.097093 |
| 7 | 조계사 | 서울시 | 종로구 | 4.5 | 종교적인 장소 | 0.882525 | 0.117475 |
| 8 | 운현궁 | 서울시 | 종로구 | 4 | 역사적인 장소 | 0.874813 | 0.125187 |
| 9 | 종묘 | 서울시 | 종로구 | 4 | 역사적인 장소 | 0.865868 | 0.134132 |

```
data[data['카테고리']=='자연 경관']
```

| | 여행지 | 행정구역 | 시군구 | 카테고리 |
|-----|-----------|------|------|-------|
| 37 | 서울숲 | 서울시 | 성동구 | 자연 경관 |
| 87 | 임진각 평화누리 | 경기도 | 파주시 | 자연 경관 |
| 111 | 허브아일랜드 | 경기도 | 포천시 | 자연 경관 |
| 113 | 국망봉 자연휴양림 | 경기도 | 포천시 | 자연 경관 |
| 120 | 유명산 자연휴양림 | 경기도 | 가평군 | 자연 경관 |
| ... | ... | ... | ... | ... |
| 626 | 빙계계곡 | 경상북도 | 의성군 | 자연 경관 |
| 628 | 주남저수지 | 경상남도 | 창원시 | 자연 경관 |
| 634 | 제주 올레 | 제주도 | 서귀포시 | 자연 경관 |
| 638 | 탄금대 | 충청북도 | 충주시 | 자연 경관 |
| 642 | 천제연폭포 | 제주도 | 서귀포시 | 자연 경관 |

05. 결론

경복궁



창덕궁

```
get_recommendation('경복궁')
```

| | 여행지 | 행정구역 | 시군구 | 트립어드바이저평점 | 카테고리 | 최종긍정 | 최종부정 |
|---|--------------|------|-----|-----------|---------|----------|----------|
| 0 | 창덕궁 | 서울시 | 종로구 | 4.5 | 역사적인 장소 | 0.916274 | 0.083726 |
| 1 | 창경궁 | 서울시 | 종로구 | 4.5 | 역사적인 장소 | 0.936209 | 0.063791 |
| 2 | 국립현대미술관(서울관) | 서울시 | 종로구 | 4.5 | 박물관 | 0.951715 | 0.048285 |
| 3 | 낙산공원 | 서울시 | 종로구 | 4.5 | 공원 | 0.945897 | 0.054103 |
| 4 | 북악산 서울 한양도성 | 서울시 | 성북구 | 4.5 | 역사적인 장소 | 0.914828 | 0.085172 |
| 5 | 국립서울현충원 | 서울시 | 동작구 | 4.5 | 역사적인 장소 | 0.908730 | 0.091270 |
| 6 | 경교장 | 서울시 | 종로구 | 4 | 역사적인 장소 | 0.902907 | 0.097093 |
| 7 | 조계사 | 서울시 | 종로구 | 4.5 | 종교적인 장소 | 0.882525 | 0.117475 |
| 8 | 운현궁 | 서울시 | 종로구 | 4 | 역사적인 장소 | 0.874813 | 0.125187 |
| 9 | 종묘 | 서울시 | 종로구 | 4 | 역사적인 장소 | 0.865868 | 0.134132 |

창덕궁



종묘

```
1 get_recommendations('창덕궁')
```

| | 여행지 | 행정구역 | 시군구 | 트립어드바이저평점 | 카테고리 | 최종긍정 | 최종부정 |
|---|--------------|------|-----|-----------|---------|----------|----------|
| 0 | 경복궁 | 서울시 | 종로구 | 4.5 | 역사적인 장소 | 0.909808 | 0.090192 |
| 1 | 창경궁 | 서울시 | 종로구 | 4.5 | 역사적인 장소 | 0.936209 | 0.063791 |
| 2 | 국립현대미술관(서울관) | 서울시 | 종로구 | 4.5 | 박물관 | 0.951715 | 0.048285 |
| 3 | 낙산공원 | 서울시 | 종로구 | 4.5 | 공원 | 0.945897 | 0.054103 |
| 4 | 북악산 서울 한양도성 | 서울시 | 성북구 | 4.5 | 역사적인 장소 | 0.914828 | 0.085172 |
| 5 | 국립서울현충원 | 서울시 | 동작구 | 4.5 | 역사적인 장소 | 0.908730 | 0.091270 |
| 6 | 경교장 | 서울시 | 종로구 | 4 | 역사적인 장소 | 0.902907 | 0.097093 |
| 7 | 조계사 | 서울시 | 종로구 | 4.5 | 종교적인 장소 | 0.882525 | 0.117475 |
| 8 | 운현궁 | 서울시 | 종로구 | 4 | 역사적인 장소 | 0.874813 | 0.125187 |
| 9 | 종묘 | 서울시 | 종로구 | 4 | 역사적인 장소 | 0.865868 | 0.134132 |

종묘



인사동길

```
1 get_recommendation('종묘')
```

| | 여행지 | 행정구역 | 시군구 | 트립어드바이저평점 | 카테고리 | 최종긍정 | 최종부정 |
|---|--------|------|-----|-----------|---------|----------|----------|
| 0 | 운현궁 | 서울시 | 종로구 | 4 | 역사적인 장소 | 0.874813 | 0.125187 |
| 1 | 경교장 | 서울시 | 종로구 | 4 | 역사적인 장소 | 0.902907 | 0.097093 |
| 2 | 보신각 | 서울시 | 종로구 | 4 | 역사적인 장소 | 0.708608 | 0.291392 |
| 3 | 창경궁 | 서울시 | 종로구 | 4.5 | 역사적인 장소 | 0.936209 | 0.063791 |
| 4 | 창덕궁 | 서울시 | 종로구 | 4.5 | 역사적인 장소 | 0.916274 | 0.083726 |
| 5 | 경복궁 | 서울시 | 종로구 | 4.5 | 역사적인 장소 | 0.909808 | 0.090192 |
| 6 | 송례문 | 서울시 | 중구 | 4 | 역사적인 장소 | 0.820792 | 0.179208 |
| 7 | 인사동길 | 서울시 | 종로구 | 4 | 랜드마크 | 0.817851 | 0.182149 |
| 8 | 태릉 | 서울시 | 중랑구 | 4 | 역사적인 장소 | 0.815358 | 0.184642 |
| 9 | 이화벽화마을 | 서울시 | 종로구 | 4 | 랜드마크 | 0.804984 | 0.195016 |

인사동길



?

```
1 get_recommendations('인사동길')
```

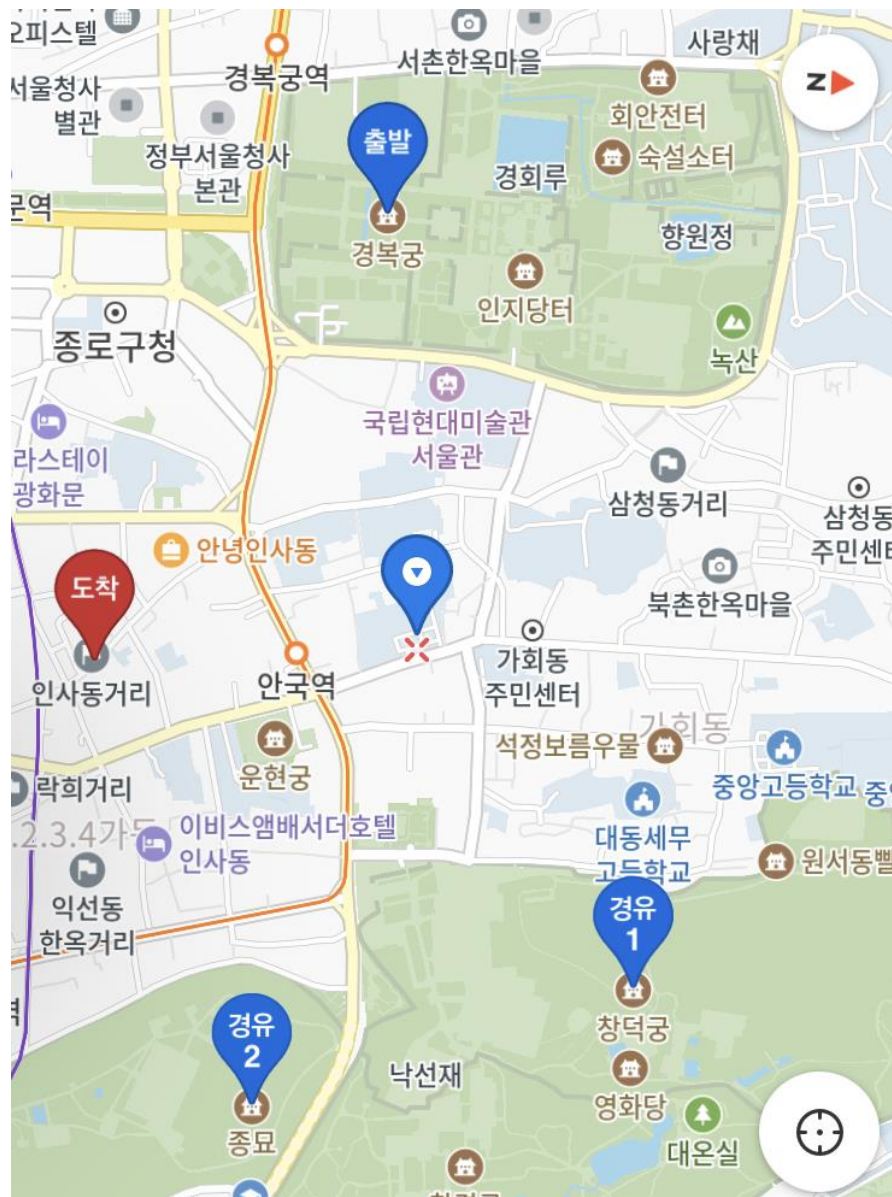
| | 여행지 | 행정구역 | 시군구 | 트립어드바이저평점 | 카테고리 | 최종공정 | 최종부정 |
|---|--------|------|-----|-----------|---------|----------|----------|
| 0 | 이화벽화마을 | 서울시 | 종로구 | 4 | 랜드마크 | 0.804984 | 0.195016 |
| 1 | 북촌한옥마을 | 서울시 | 종로구 | 4 | 랜드마크 | 0.773710 | 0.226290 |
| 2 | 청와대 | 서울시 | 종로구 | 4 | 랜드마크 | 0.735838 | 0.264162 |
| 3 | 경교장 | 서울시 | 종로구 | 4 | 역사적인 장소 | 0.902907 | 0.097093 |
| 4 | 서울광장 | 서울시 | 중구 | 4 | 랜드마크 | 0.916131 | 0.083869 |
| 5 | 운현궁 | 서울시 | 종로구 | 4 | 역사적인 장소 | 0.874813 | 0.125187 |
| 6 | 종묘 | 서울시 | 종로구 | 4 | 역사적인 장소 | 0.865868 | 0.134132 |
| 7 | 청계천 | 서울시 | 종로구 | 4.5 | 랜드마크 | 0.847465 | 0.152535 |
| 8 | 삼청동길 | 서울시 | 종로구 | 4.5 | 랜드마크 | 0.815358 | 0.184642 |
| 9 | 사직단 | 서울시 | 종로구 | 4 | 공원 | 0.802912 | 0.197088 |

경복궁

① 창덕궁

② 종묘

서울 종로구 인사동길



참고 문헌 및 출처

참고문헌 및 출처

- 주제 선정 : (출처) 한국관광 데이터랩 – 2020 잠재 방한 여행객 조사
한국이미지커뮤니케이션연구원 – 코로나가 끝나면 하고 싶은 것
- 토픽 모델링 :
 - 정지우 외, 2021, 「외국인 관광객 리뷰데이터를 활용한 토픽모델링 기반의 공간분석: 대구광역시를 사례로」, Journal of Digital Convergence.
 - 이혜진, 강영옥, 2020, 「토픽모델링과 LSTM기반 텍스트 분석을 통한 부산방문 외국인 관광객의 선호관광지 및 관광매력요인 분석」, 한국도시지리학회지.
- 추천 시스템 :

김상모, 김형준, 한인규, 2013, 「코사인 유사도 기법을 이용한 뉴스 추천 시스템」, 제25회 한글 및 한국어 정보처리 학술대회 논문집.
- 감성분석 :

승현수 외, 2021, 「검색환경 개선을 위한 자연어 처리 기반맞춤형 추천 검색시스템」, ACK 2021 학술발표대회 논문집 (28권 2호).

감사합니다
