

[한국 관광 100선] 관광지 추천 서비스

20231026 경북대학교 AI·빅데이터 전문가 양성과정 4기 박 소 원

목차

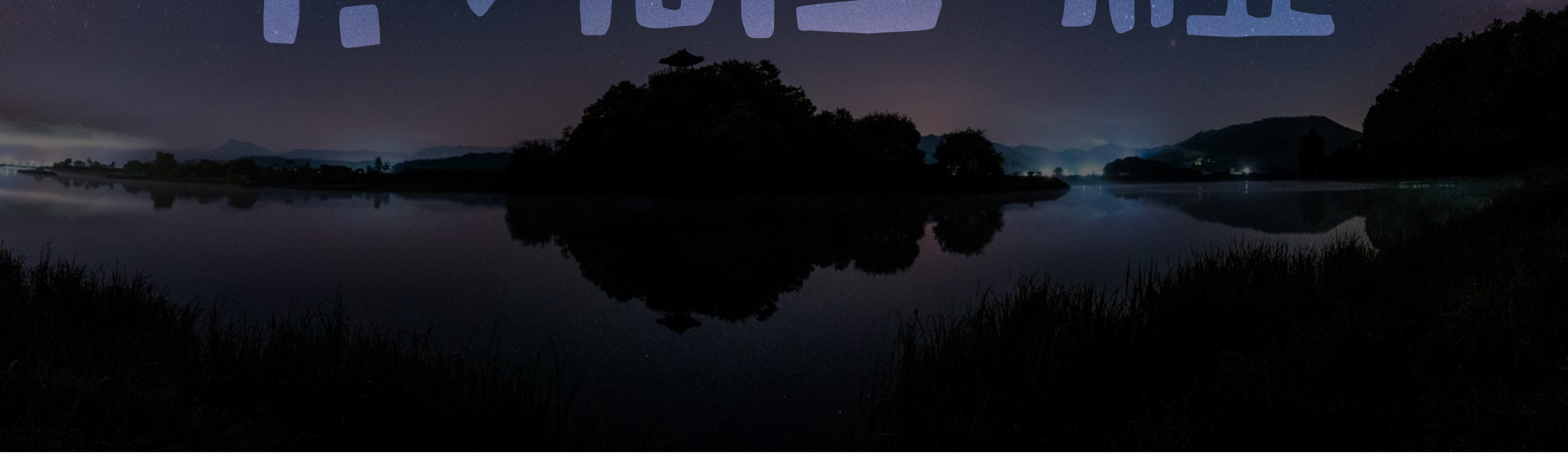
서비스
개요

개발
과정

주요
기능

서비스
시연

1. 서비스 개요



국내 여행, 어디로 가야 할까 ?

[국내 여행 방문지 선택 이유]

볼거리 제공

21.8%

여행지 지명도

17.4%

여행할 수
있는 시간

12.3%

2022 국민여행조사. 문화체육관광부

볼거리가 많고
유명한 관광지를
어디서, 어떻게 찾을 수 있을까 ?

국내 여행, 어디로 가야 할까 ?

[여행 정보 획득 경로]

과거 방문 경험

33.1%

주변인

27.1%

인터넷 사이트
모바일 앱

17.4%

본인 / 주변인의 경험

⇒ 한정된 정보

인터넷 / 모바일 검색

⇒ 과다한 정보

서비스 개요

개발 과정

주요 기능

서비스 시연

A-Ha

한국 관광 100선

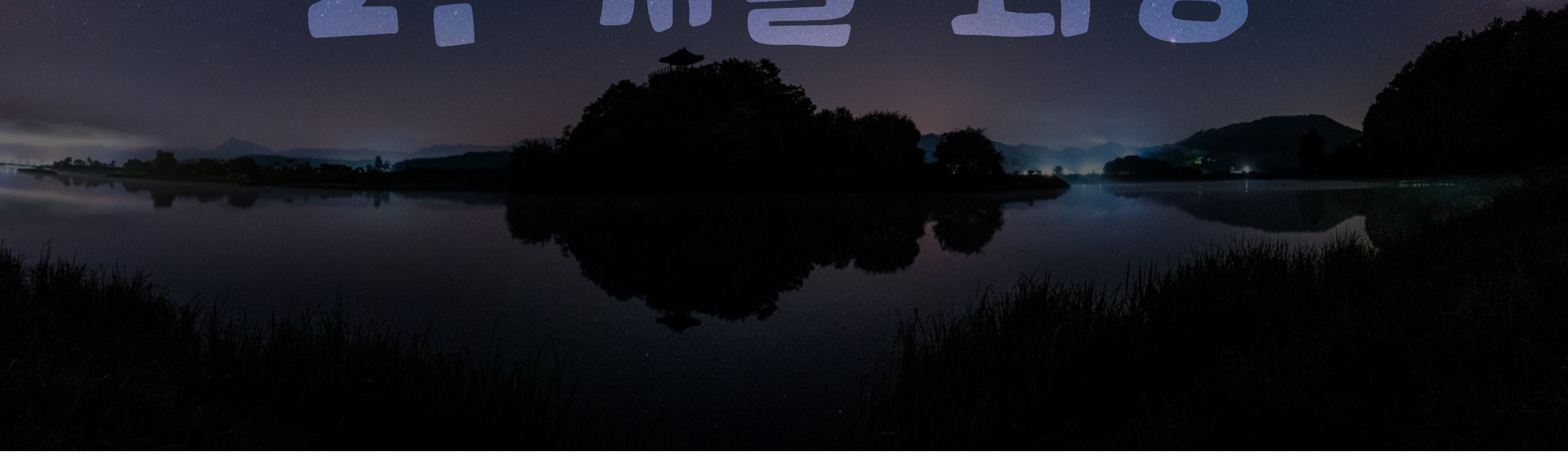
문화체육관광부와 한국관광공사에서 선정한 한국 대표 관광지 100선
객관적이고 엄선된 정보를 제공



원하는 여행을 검색하면 [한국 관광 100선]의
관광지 중 5곳을 추천해주는 서비스

A-Ha

2. 개발 과정



[illegible]

1. 데이터 수집 _ 데이터 정제

>> 원본 데이터

2023~2024 한국관광 100선					
번호	지역		관광지	주소	비고
1	수	서울	서울 5대 고궁	서울특별시 종로구 사직로 161(경복궁)	
		-9	(경복궁, 창덕궁, 창경궁, 덕수궁, 경희궁)	서울특별시 종로구 율곡로 99(창덕궁)	
	도			서울특별시 종로구 창경궁로 185(창경궁)	
				서울특별시 종로구 세종대로 99(덕수궁)	
	권			서울특별시 종로구 새문안로 55(경희궁)	

>> 정제 데이터

관광지	주소
경복궁	서울특별시 종로구 사직로 161(경복궁)
창덕궁	서울특별시 종로구 율곡로 99(창덕궁)
창경궁	서울특별시 종로구 창경궁로 185(창경궁)
덕수궁	서울특별시 종로구 세종대로 99(덕수궁)
경희궁	서울특별시 종로구 새문안로 55(경희궁)

- 여러 개의 관광지가 하나로 합쳐져서 선정된 경우
⇒ 각 하나의 관광지로 분리
- 필요 없는 데이터 제거

1. 데이터 수집 _ 위도·경도 데이터 추가

1 위도, 경도 변환 - GEO 코딩

```
1 from geopy.geocoders import Nominatim

1 geo_local = Nominatim(user_agent='South Korea')
2
3 # 위도, 경도 반환하는 함수
4 def geocoding(address):
5     try:
6         geo = geo_local.geocode(address)
7         x_y = [geo.latitude, geo.longitude]
8         return x_y
9
10    except:
11        return [0,0]
```

	관광지	주소	latitude	longitude
2	창경궁	서울특별시 종로구 창경궁로 185(창경궁)	37.580246	126.994833
3	덕수궁	서울특별시 종로구 세종대로 99(덕수궁)	0.000000	0.000000
4	경희궁	서울특별시 종로구 새문안로 55(경희궁)	0.000000	0.000000

```
1 len(data[data['latitude']==0])
```

53

- 변환하지 못하는 주소가 많음

1. 데이터 수집 - 위도, 경도 데이터 추가

1 위도, 경도 변환 -

관광지

주소

latitude longitude

2 창경궁

185(창경궁)

37.580246 126.994833

3 덕수궁

세종대로 99(덕수궁)

0.000000 0.000000

종로구 새문안로 55(경희궁)

0.000000 0.000000

```
1 from geopy.geocoders import Nominatim
```

```
1 geo_local = Nominatim(user_agent='South Korea')
```

```
2  
3 # 위도, 경도 반환하는 함수
```

```
4 def geocoding(address):
```

```
5     try:
```

```
6         geo = geo_local.geocode(address)
```

```
7         x_y = [geo.latitude, geo.longitude]
```

```
8         return x_y
```

```
9
```

```
10    except:
```

```
11        return [0,0]
```

```
1 len(data[data['latitude']==0])
```

```
58
```

못하는 주소가 많음

1. 데이터 수집 _ 위도·경도 데이터 추가

» 구글 스프레드시트 확장프로그램

[Geocode by Awesome Table]



Geocode by Aw...

Geocode is a tool that helps you get latitudes & longitudes from addresses in a Google Sheet to display them on a map you can share.

개발자: [Talarian](#)

정보 업데이트: 2023년 1월 10일

관광지

주소 Latitude Longitude

경복궁	서울특별시 종로구 사직로 161(경복궁)	37.579617	126.977041
창덕궁	서울특별시 종로구 율곡로 99(창덕궁)	37.581122	126.991112
창경궁	서울특별시 종로구 창경궁로 185(창경궁)	37.578771	126.994870
덕수궁	서울특별시 종로구 세종대로 99(덕수궁)	37.565805	126.975146
경희궁	서울특별시 종로구 새문안로 55(경희궁)	37.570755	126.968834

1. 데이터 수집 _ 크롤링



- Selenium 이용
각 관광지명 네이버 블로그 검색
⇒ 스크롤 다운 30회
⇒ 텍스트 크롤링

2. 데이터 전처리

》 한글과 띄어쓰기만 남기기

```
# 한글과 띄어쓰기만 남기기 -> 정규식
```

```
kr_texts = [re.sub(r'^가-힣\s', '', text) for text in texts]
```

》 불용어 & 한 글자 단어 제외

```
# 불용어 제외
```

```
text = text.replace('\n', '').replace('블로그', '').replace('문서', '').replace('저장하기', '').replace(n, '')
```

```
stop_words = ['가끔', '가도', '갑자기', '개구', '개월', '거기', '곳도', '과정', '관련', '관리', '구간', '구매', '구성',  
              '구역', '군데', '궁리', '그대로', '그동안', '기본', '기자', '나름', '나중', '내내', '내용', '내일', '년전',  
              '누구', '대리', '대부분', '대신', '대인', '덕분', '두번째', '둘러보기', '디피', '따라서', '랍니', '린지',  
              '마감', '마련', '마이', '마침', '매년', '매장', '며칠', '무엇', '무조건', '무척', '문의', '보통', '부근',  
              '비롯', '솔직', '안녕', '얼마나', '완전', '우리', '우선', '월월', '월일', '전날', '주의', '평소']
```

```
# 한 글자 단어 삭제
```

```
input_data = [word for word in nouns if len(word) > 1]
```

```
input_data = [' '.join(input_data)]
```


3. 정규화·명사 추출 및 벡터화

[벡터화]	[정규화·명사 추출]	
	okt	kkma
TFIDF	성능 저하	성능저하
Countvectorizer	성능 양호	성능 양호



비슷한 성능
&

처리 속도가 조금 더 빠른
okt & Countvectorize 로
정규화·명사 추출 및 벡터화

4. 빈도 분석

» 가장 많이 나온 1000개 / 10개의 키워드 추출

	관광지	주소	total_1000	top10	top1000
0	경복궁	서울특별시 종로구 사직로 161(경복궁)	[('맛집', 403), ('서울', 250), ('야간', 189), ('카페', ...	맛집 서울 야간 카페 개장 시간 광화문 한복 관람 근처	맛집 서울 야간 카페 개장 시간 광화문 한복 관람 근처 가을 여행 데이트 위치 삼...
1	창덕궁	서울특별시 종로구 율곡로 99(창덕궁)	[('후원', 288), ('서울', 281), ('카페', 219), ('창경궁'...	후원 서울 카페 창경궁 관람 궁궐 달빛 여행 기행 경북궁	후원 서울 카페 창경궁 관람 궁궐 달빛 여행 기행 경북궁 시간 고궁 매화 안국역 ...
2	창경궁	서울특별시 종로구 창경궁로 185(창경궁)	[('서울', 290), ('창덕궁', 258), ('궁궐', 178), ('야간'...	서울 창덕궁 궁궐 야간 온실 개장 여행 시간 고궁 스냅	서울 창덕궁 궁궐 야간 온실 개장 여행 시간 고궁 스냅 한복 경북궁 문화 나들이 ...
3	덕수궁	서울특별시 종로구 세종대로 99(덕수궁)	[('서울', 394), ('미술관', 197), ('돌담', 189), ('여행'...	서울 미술관 돌담 여행 맛집 정동 장육진 석조전 국립 카페	서울 미술관 돌담 여행 맛집 정동 장육진 석조전 국립 카페 현대 와플 궁궐 시간 ...

5. 코사인 유사도

» 코사인 유사도 함수 적용

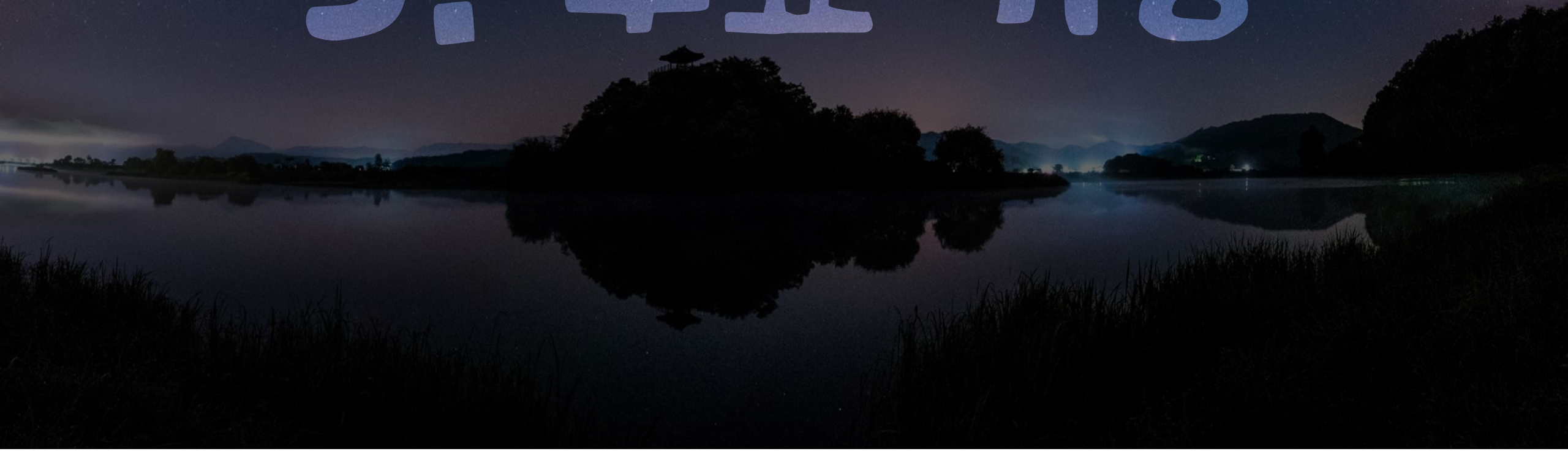
```
def make_cosine(a, b):  
    return np.dot(a,b) / (np.linalg.norm(a) * np.linalg.norm(b))
```

```
sample=cv.transform(input_data).toarray().reshape(len(df_dtm.columns))
```

```
box = []  
for target in tqdm(df_dtm.iloc):  
    cosine = make_cosine(target, sample)  
    box.append(cosine)
```

```
data['유사도분석'] = box  
data.sort_values("유사도분석", ascending=False)['관광지'][:5]
```


3. 주요 기능



1. 찾고 싶은 관광지를 검색
2. [한국관광 100선] 중 검색 조건에 맞는 5개의 관광지 추천
3. 각 관광지의 정보 제공
 - 위치 정보
 - TOP 10 키워드
4. TOP 10 키워드 관련 이미지 제공

4. 서비스 시연



서비스 개요

개발 과정

주요 기능

서비스 시연

A-Ha



1



2



3



4

》》 좋았던 점

- 허접하지만 뿌듯함 :D
- 처음부터 끝까지 다 해보니 전체적인 흐름을 알 수 있었음
- 부족한 점이 너무 많아서 더 열심히 해야겠다고 생각함
- 웹, Flask 수업 때 멘붕... 이었는데 거부감이 사라짐

》》 개선할 점

- [한국관광 100선] 내에서만 크롤링, 관광지 추천을 했는데 더 많은 관광지를 추가하면 좋을 듯
- 크롤링을 네이버 블로그 외 다양한 채널에서 했으면 더 좋았을 듯
- 웹은 자바스크립트가 필요하다...

감사합니다

