

숙박 추천 서비스

TEAM : HPC

Github :

https://github.com/pwh9882/HPC_accommodation_recommender



요구사항 분석

서비스 적용 범위 : 전세계에 있는 호텔

주요 서비스

호텔 등급의 범위를 설정하는 기능

사용자가 요구사항을 직접 입력 가능

Process

1

CSV 파일 전처리

전세계 호텔 정보(이름, 호텔등급, 나라, 도시)가 담긴 CSV 파일을 텍스트로 변환

2

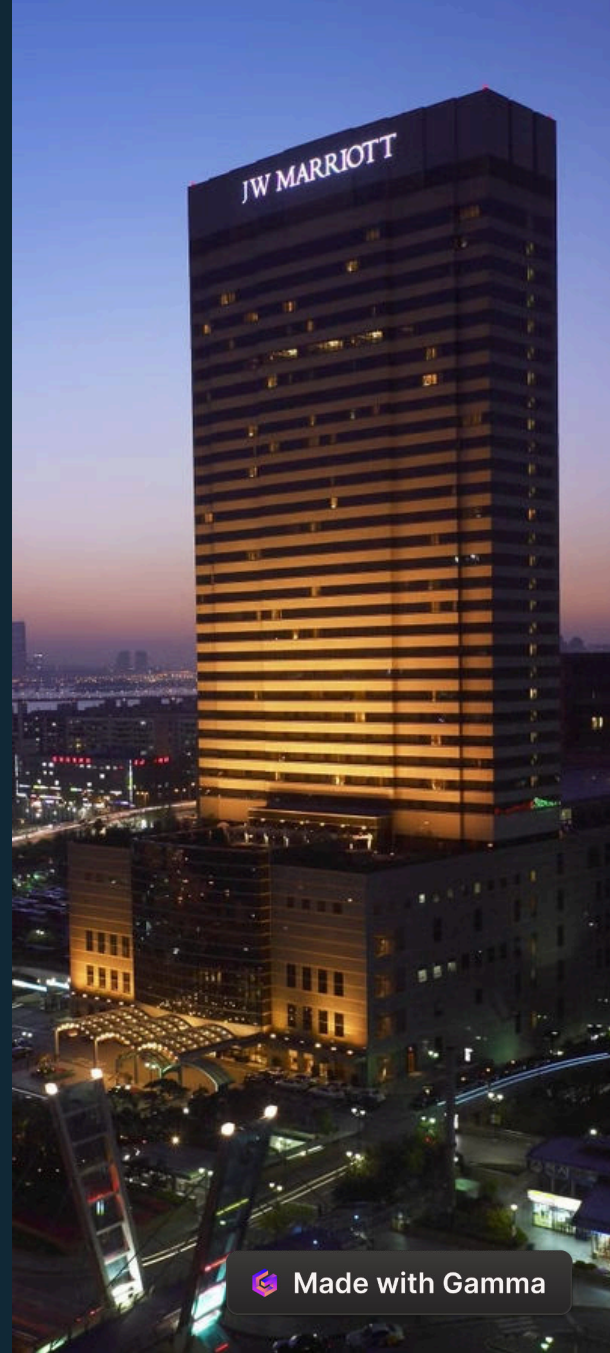
LLM 및 RAG 분석

텍스트 파일을 LLM과 RAG 기술을 활용하여 분석

3

사용자 요구사항 충족

사용자의 요구사항을 충족하는 최적의 호텔 정보 제시



데이터 소스 선택 및 변경

프로젝트 초기에는 '여기어때' 사이트에서 서울 지역의 호텔 데이터를 수집하려 했습니다. 하지만 더 넓은 범위의 데이터 분석을 위해 Kaggle의 호텔 데이터셋으로 변경하기로 결정했습니다. 이 결정은 다음과 같은 이점을 제공했습니다:

1 **글로벌 관점**
다양한 국가의 호텔 정보를 포함하여 더 넓은 시각에서 분석할 수 있게 되었습니다.

2 **데이터 품질**
Kaggle의 데이터셋은 이미 일정 수준의 품질 관리가 이루어진 상태로, 더 신뢰할 수 있는 데이터를 제공합니다.

3 **풍부한 정보**
호텔의 다양한 특성과 설명을 포함하고 있어, 더 깊이 있는 분석이 가능해졌습니다.

4 **접근성**
공개된 데이터셋을 사용함으로써 데이터 수집 과정에서의 법적, 윤리적 문제를 피할 수 있었습니다.





데이터 탐색 및 구조 파악

1

데이터 로딩

pandas 라이브러리를 사용하여 데이터셋을 로드하고, 기본적인 정보(행 수, 열 이름 등)를 확인

2

결측치 확인

각 열의 결측치 비율을 계산하여 데이터의 완전성을 평가

3

데이터 유형 검사

각 열의 데이터 유형을 확인하고, 필요한 경우 적절한 유형으로 변환

4

고유값 분석

특히 '국가' 열의 고유값을 확인하여 데이터셋에 포함된 국가의 수와 종류를 파악

호텔 데이터 선별 기준 설정

국가별로 호텔 데이터를 필터링한 후, 분석에 사용할 호텔 데이터를 선별하기 위한 기준을 설정했습니다. 이 기준은 데이터의 품질과 분석의 목적을 고려하여 결정되었습니다:

설명 존재 여부

호텔에 대한 설명(description)이 있는 경우만 선택했습니다. 이는 텍스트 분석을 위해 중요한 기준이었습니다.

무작위 선택

각 국가별로 30개의 호텔을 무작위로 선택했습니다. 이는 데이터의 다양성을 유지하면서도 관리 가능한 크기로 데이터셋을 유지하기 위함입니다.

데이터 완전성

선택된 호텔 데이터에 중요한 정보(예: 이름, 주소, 평점 등)가 누락되지 않았는지 확인했습니다.

국가별 CSV 파일 생성

pandas 라이브러리를 사용하여 호텔 데이터를 국가별로 CSV 파일로 저장하는 작업을 수행했습니다. 이 과정은 다음과 같은 단계로 진행되었습니다

1

데이터 정제

각 국가별로 선별된 30개의 호텔 데이터에서 불필요한 열을 제거(description 존재x)하고, 필요한 정보만 남겼습니다.

2

파일명 설정

각 국가의 이름을 사용하여 CSV 파일명을 생성했습니다. (예: "South Korea.csv")

3

CSV 저장

pandas의 to_csv() 함수를 사용하여 각 국가별 데이터프레임을 CSV 파일로 저장했습니다.

4

저장 확인

각 파일이 올바르게 저장되었는지 확인하고, 파일 크기와 내용을 검증했습니다.

그리고 마지막으로 통합.csv 파일을 생성했습니다. (모든 나라의 호텔 정보가 들어있는)

데이터 전처리 결과 및 향후 계획

데이터 전처리 과정을 통해 다음과 같은 결과를 얻었습니다:

총 처리된 국가 수	142개국
국가별 선별된 호텔 수	50개
총 호텔 데이터 수	6254개
생성된 CSV 파일 수	51개 (국가별 50개 + 통합 1개)

이제 전처리된 데이터를 바탕으로 다음과 같은 분석을 진행할 계획입니다:

- 고객의 니즈에 따라 설명을 분석하여 추천

이러한 분석을 통해 글로벌 호텔 산업의 트렌드와 특성을 파악하고, 유용한 정보를 도출했습니다.

홈 화면

Deploy

필터 옵션

국가 이름

Albania

도시 이름

Albanien

☐ 모든 HotelRating

HotelRating

135

검색

숙박 추천 앱

원하는 조건을 선택하고 추천 숙소를 확인하세요.

Windows 정품 인증

[설정]으로 이동하여 Windows를 정품 인증합니다.

검색 결과

<

필터 옵션

국가 이름

South Korea

도시 이름

Andong

☐ 모든 HotelRating

HotelRating

135

tv

검색

Deploy

숙박 추천 앱

원하는 조건을 선택하고 추천 숙소를 확인하세요.

조건에 맞는 숙소

★ AI 추천 숙소!

안동에 위치한 3성급 호텔 중에는 다음과 같이 TV가 구비되어 있는 호텔들이 있습니다:

- 안동 호텔 (Andong Hotel)
 - 객실: LCD TV가 구비되어 있습니다.
 - 주소: 40-11 Munhwagwangjang-gil, Andong, 36701, Gyeongsangbuk-do
- 안동 리버 섀넌 펜션 (Andong River Sunset Pension)
 - 객실: 평면 TV가 구비되어 있습니다.
 - 주소: 1796 Pungsaesa-ro, Pungsan Eup, Pungsan-eup, Andong, 36622

이 두 호텔 모두 TV가 구비된 객실을 제공합니다.

Andong Hotel

countyName: South Korea

cityName: Andong

HotelRating: ThreeStar

Address: 40-11 Munhwagwangjang-gil
Andong 36701Gyeongsangbuk-do

Andong River Sunset

countyName: South Korea

cityName: Andong

HotelRating: ThreeStar

Address: 1796 Pungsaesa-ro Pungsan Eup
Pungsan-eupAndong 36622

Windows 정품 인증
[설정]으로 이동하여 Windows를 정품 인증합니다.

Made with Gamma

감사합니다

