

접수번호

※작성하지 않음

「제4회 대구 빅데이터 분석 경진대회」 과제 분석 계획서

신청자명(팀명)	대도시
분석주제명	데이터분석 기반 맞춤형 대구 관광지 경로 추천 서비스 구축

※ 5장 내외 작성 및 목차 준용 (이미지 첨부 가능)

1. 기획 배경 및 개요

우리 삶을 뒤바꿔 놓았던 코로나 19 팬데믹이 엔데믹으로 전환되면서 우리는 과거의 일상을 조금씩 되찾고 있다. 그동안 자제했던 야외활동 수요가 증가함과 동시에 ‘관광’은 큰 화두에 올랐다. 실제로 여행 플랫폼의 이용자 수는 전월 대비 30%가 증가했다. 대구는 오래도록 여행을 손꼽아 기다린 우리에게 좋은 여행지일까?

대구是我们에게 최고의 여행지가 되어줄 것이다. 코로나 이전 2019년도 대구 관광객 대상 실태조사에서 만족도는 5점 만점에 4.23을 기록하며 뚜렷한 관광실적을 자랑했다. 특히나 이미지 개선은 84.4%, 재방문의사 91.3%를 기록한 경험이 있다. 대구에 한 번이라도 방문한다면 이미지 개선은 물론이고 지속적인 방문을 기대할 수 있을 것이다.

대구是我们에게 여행객들에게 편리하고 즐거운 여행 이미지를 심어줄 필요가 있다. 여행을 가기 전 가장 힘든 것은 여행 코스를 짜는 것이다. 대구是我们에게 여행객의 편리한 여행을 위해 이를 도와줄 필요가 있다. 사용자가 선호하는 대구 여행 코스는 데이터를 통해 찾아낼 수 있다. 우리는 그 간 축적된 데이터를 통해 사용자에게 맞춤형 여행 코스를 제공하고자 한다. 유명한 명소뿐만 아니라 신생 명소, 리뷰 대비 인지도가 낮은 명소를 함께 추천하여 균형 있는 관광산업의 발전과 대구여행 진입장벽 감소를 기대할 수 있을 것이다.

2. 분석 내용 요약

통신사 관광지 방문인구 데이터를 활용해서 연령, 성별을 기준으로 그룹을 나누고 각 그룹에 해당하는 빈도수를 구한다. 이 빈도수를 각 성별과 연령대에 맞는 맞춤 관광 코스를 구성하는데 활용할 수 있다.

3. 분석 방법 및 계획

○ 분석에 활용할 데이터

- 데이터 확보 계획: 공공데이터포털 이용, 대구 빅데이터 활용센터 직접 방문 또는 온라인 수령 등 데이터 확보

○ 활용 데이터명

2.관광

연령별 방문 빈도

2-1. 통신사 관광지방문인구 데이터

(파일명: 1. 유니크인구 및 관광지 인구 / 2021.02~12 /daegu_tour_pop_2021MM.CSV
/ 2021.01~12 / daegu_unique_pop_2021MM.CSV)

관광지 별 리뷰

2-2. 맛집, 호텔, 명소 리뷰 데이터

(파일명: 1. A사/ A사 리뷰_대구 (맛집, 호텔, 명소)_외국인구분.csv

2. B사/ B사 리뷰_대구 (맛집, 호텔, 명소)_외국인구분.csv

3. C사/ C사 리뷰_대구 (맛집, 호텔, 명소)_외국인구분.csv)

2-3. 대구시 관광지 데이터(블로그, 카페)

(파일명: 1. 포털_블로그_대구 관광.csv

2. 포털_카페_대구 관광.csv)

2-4. 대구 관광 리뷰데이터

(파일명: 1. 포털A_대구관광_리뷰데이터(utf-8).csv

2. 포털B_대구관광_장소데이터(utf-8).csv

3. 포털B_대구관광_리뷰데이터(utf-8).csv

4. 포털C_대구관광_리뷰데이터(utf-8).csv)

장소별 동선

8-1. 행정구역 지리정보

(파일명: 1. 시군구/ bnd_sigungu_22_2018_2018_2Q.shp

2. 시도/ bnd_sido_22_2018_2018_2Q.shp

3. 읍면동/ bnd_dong_22_2018_2018_2Q.shp

4. 대한민국 전체 행정구역/ bnd_*_00_2018_2018/ *shp)

9-1. 유동인구 소블럭 기준

(파일명: DMM_BLOCK.DBF

DMM_BLOCK.PRJ

DMM_BLOCK.shp

DMM_BLOCK.SHX

대구_BLOCK_좌표.CSV)

◦ 분석에 적용할 빅데이터 통계 . 분석 기법, 방법론

프로토타입 모델의 분석방법론을 활용할 예정이다. 계속적으로 분석을 시도하고 반복적으로 개선하여 해결책 모형을 제시하는 상향식 접근 방법을 활용할 것이다.

가설 설정 후 결과물을 통해 가설을 확인하고 검증을 통해 가설의 실험을 거쳐 결과물을 완성 시킬 것이다.

예를 들어 A라는 관광명소에 여성 20대의 수가 상대적으로 다른 그룹에 비해 높다면 20대 여성 이용객에게 A 관광명소를 여행 코스로 추천해 줄 수 있다. 반대로 20대 남자의 수가 상대적으로 다른 그룹에 비해 낮다면 20대 남자 이용객에는 A 관광명소를 제외하고 다른 명소를 여행 코스로 짤 수 있다.

2-2. 맛집, 호텔, 명소 리뷰데이터에서 2-1 데이터와 중복되는 관광명소가 있다면 해당되는 리뷰 데이터를 추출해서 텍스트마이닝을 실행한 뒤 자주 등장하는 단어를 바탕으로 여행 관광객에게 맞춤 관광 코스 서비스를 제공할 수 있다.

예를 들어서 B라는 관광맛집에 리뷰이벤트에서 ‘맵다’ ‘맛있다’ 라는 단어의 빈도수가 많았고 2-1, 2-2, 2-4 데이터에서 특정 연령과 성별의 빈도수가 높게 나왔다면 해당 연령과 성별의 관광객 중에서 매운 음식을 좋아한다면 B관광 명소를 추천해주는 분석 결과가 나올 수 있다.

더하여 각 명소끼리의 거리를 계산하여 최단 거리에 있는 명소를 우선적으로 추천하는 서비스를 제공한다.

◦ 분석 결과에 대한 시각화 기법 및 기능 구현 방법

classification 모델, 텍스트 크롤링, 최단 경로 알고리즘

4. 분석 결과 활용 및 시사점

◦ 분석 결과 활용 방안 및 적용 대상

통신사 관광지 방문인구 데이터를 기반으로 성별, 연령대 별 관광지 방문빈도를 산출한다. 이를 기반으로 방문자 특성 별 빈도수가 가장 대구 관광지를 선정하여 ‘맞춤형 경로 추천’ 서비스를 제공한다. 관광지 방문빈도 분석 결과를 통해 성별, 연령대 별 선호하는 관광지 유형을 파악할 수 있으며, 이는 대구 관광코스 추천 서비스에 활용된다.

‘맞춤형 대구 관광지 경로 추천 서비스’는 전 연령대의 방문자에 적용 가능하며, 본 서비스를 통해 사용자 특성 별 대구 명소를 추천받을 수 있다. 이와 더불어 행정구역 지리정보 데이터를 활용해 관광지 간 거리산출 서비스 또한 수행할 수 있다. 관광지 간 거리를 고려한 경로 추천 서비스에 활용 가능하다.

◦ 분석 결과 적용 시 기대효과

‘맞춤형 대구 관광지 경로 추천’ 분석 결과 적용으로 대구 관광산업 활성화를 도모할 수 있다. 이는 코로나19의 장기화로 인한 관광산업 침체 상황에서 재도약의 기반이 될 것으로 예상된다.

현재 일상회복 단계로 접어들며 국내 관광객 증가하는 추세임에 따라 관광객 유치를 위한 관광 발전 전략이 수립되어야 할 것이다. 본 관광데이터 분석 결과 적용으로 대구 관광객 유치에 큰 파급효과가 발생할 것으로 예상된다.

사용자는 맞춤형 관광지 경로 추천 서비스를 통해 여행 일정 수립의 부담을 완화할 수 있으며, 여행의 만족도 또한 증진될 수 있다. 관광객의 여행 진입장벽을 낮추고, 동시에 관광 명소 및 발전가능성 있는 명소를 발견함으로써 관광지로서의 대구의 입지를 다질 수 있을 것이다.

5. 참고문헌 출처 등

- 백운용 기자, 대구광역시, “내국인관광객 91.3% 외국인관광객 83%...대구관광지 다시 방문하고 싶다!”, 2020.08.12

<https://www.gukjenews.com/news/articleView.html?idxno=2054243>