|  |  |
| --- | --- |
| 1 | 안녕하세요 세계주요 10개 관광지의 호감도 및, 호감요인에 대한 특징분석을 주제로 발표하게된 김진호, 이영석입니다. |
| 2 | 저희 조는 데이터수집부터 분석 및 시각화까지 모두 직접하였습니다.  데이터는 네이버블로그를 기준으로 크롤링하였습니다. 이후 전처리를 거쳐 감성분석을 진행하였습니다.  감성분석에 사용한 감성사전으로는 대표적인 한국어 감성사전인 KNU사전를 사용하였습니다.  이를 워드클라우드와 SNA로 시각화하여 더 많은 인사이트를 도출하고자 하였습니다. |
| 3 | 데이터수집입니다.  먼저 주요 관광도시의 기준을 선정해야 했는데요, 이 기준은 유명 비행 티케팅사이트인 스카이스캐너에서 발표한 주요 관광지 상위 10개로 하였습니다.  여기에는 오사카, 도쿄, 방콕, 후쿠오카, 다낭, 타이베이, 홍콩, 세부, 하노이, 삿포로가 있습니다.  본 도시에 어떤 호감요인과 비호감요인이 존재하는지 파악하고자 하였습니다. |
| 4 | 크롤링은 네이버 블로그에서 11월 25일을 기준으로 10개의 도시명에, 여행이라는 키워드를 붙여 검색하였을 때 나오는 최근 90개의 글, 총 900개의 블로그 글을 수집하였습니다. 단어수로 하면 대략 63만 단어입니다.  더 많은 글을 수집할수록 분석의 신뢰도가 높아지겠지만, 컴퓨터 리소스로 인해 도시당 90개의 블로그 글만 크롤링하였습니다. |
| 5 | 이렇게 수집된 데이터를 전처리하였습니다.  먼저 명사분리를 하였고 이때 트위터에서 만든 한국어 분석 라이브러리인 Okt를 사용하였습니다.  한나눔이나 미캡 등도 사용해봤으나 Okt가 비교적 본 데이터에 알맞은 명사분리를 한 것으로 보였기 때문입니다.  그 다음으로 불용어와 한글자 단어를 제거했습니다. 그 목록에는 보시는 바와 같이 도시이름, 도시를 가진 국가이름, 기타 명사분리가 제대로 되지 않은 단어들이 있습니다.  위와 같은 과정을 거친 후 토큰화를 진행했습니다. |
| 6 | 빈도분석을 수행했습니다. 단어의 빈출 정도를 토대로 그래프로 도식화하였으며 워드클라우드로 시각화하였습니다.  지면상의 이유로 본 페이지에는 방콕만 표시했습니다만, 모든 동남아지역에 ‘호텔’이라는 단어가 눈에 띄게 많이 등장했습니다. 이는 저렴한 가격으로 높은 서비스를 이용하는 호텔과 같은 활동이 동남아에 주력 상품으로 작용하고 있음을 의미합니다. |
| 7 | 실제로 ‘호텔’이라는 단어가 들어간 문장들을 보면 ‘럭셔리 호텔’, ’고급’이라는 단어 등이 등장한다는 점을 볼 수 있습니다. |
| 8 | 반면 비교적 가격대가 있는 일본의 3개지역은 ‘호텔’이라는 단어의 빈도수가 매우 적으며, 그에 비해 ‘사진’이라는 단어가 타 지역들에 비해 확연하게 빈번히 등장하였습니다. 이는 비용이 덜 드는 활동중심적이며, 주변에 알리고 추억하기를 더 희망한다고 해석할 수 있습니다. |
| 9 | ‘사진’이라는 단어가 들어간 문장들을 보면 인스타 사진찍기, 기억해야지 등의 단어가 있었습니다. |
| 10 | 다음으로 도시별 특이점들도 비교해보았습니다.  일본 도시들에서는 ‘투어’라는 단어가 거의 등장하지 않았습니다. 그러나 타이페이, 홍콩, 다낭에서는 투어라는 단어가 어느정도 등장했으며, 세부와 하노이에서는 투어라는 단어가 눈에 띄게 많이 등장했습니다.  값싼 여행지에서 투어형 여행이 많다는 점을 확인할 수 있었고, 그 중에서도 세부와 하노이에 투어형 관광이 집중된다는 점을 파악할 수 있었습니다. |
| 11 | 도시별 특징을 살펴보았지만, 그래서 어느 도시가 종합적으로 만족도가 높았는지는 확인하기 어려웠습니다.  그래서 KNU감성사전을 통해 만족도를 긍부정지수로 수치화하였습니다.  먼저 생소하신 분이 계실 테니 KNU감성사전에 대해서 설명드리겠습니다.  KNU감성사전은 대표적인 한국 일상어 감성사전입니다. 국립국어원의 표준국어대사전을 구성하는 각 단어의 뜻풀이를 분석하여 긍정과 부정으로 명확하게 분류되는 단어들을 추출합니다. 이를 토대로 30만여개의 한국어 단어 중 약 1만 5천 단어에 매우 부정부터 매우 긍정까지의 수준을 부여하였습니다.  그렇게 하여 ‘맛집’, ’사랑’과 같은 단어는 긍정으로, ‘지옥’과 같은 단어는 부정으로 분류합니다. |
| 12 | 로우데이터인 블로그 글 특성상 글마다 단어수가 다릅니다. 그래서 저희는 부정단어 대비 긍정단어의 비율을 계산하였습니다.  관광 후기인 만큼 모든 도시에서 긍정단어가 부정단어보다 월등히 높게 나왔습니다. 그러나 후쿠오카나 세부에서는 비교적 낮은 긍정도를 보인 반면, 다낭과 홍콩에서는 높은 긍정도가 나왔습니다. |
| 13 | 긍정점수가 높았던 다낭과 홍콩에서 많이 등장한 키워드가 무엇인지 살펴보았습니다. |
| 14 | 그 결과 ‘재미, 맛집, 할인’이라는 단어가 다른 도시들에 비해 많았습니다. |
| 15 | 반대로 부정점수가 높았떤 후쿠오카와 세부에서 많이 등장한 단어는 무엇인지 살펴보았고, ‘걱정’이라는 단어가 많이 등장했습니다. |
| 16 | 걱정이 들어간 문장이 무엇이 있는지 살펴보았고, 보시는바와 같이 위생이나 인파에 대한 걱정들이 있었습니다.  (빨간표시는 ‘걱정’만 표시하는게 나을것같아) |
| 17 | 하지만 빈출키워드 분석만으로는 단어간 연관관계를 알기 어려웠습니다. 그래서 같은 문장에 등장하는 단어들을 기준으로 연결망을 만들었습니다.  시각화를 함에 있어 각 노드별로 중요도 및 빈출정도를 가중치로 수치화하여 노드의 크기로 표현하였고, 도시별로 엣지의 색을 다르게 하였습니다. |
| 18 | 보다 효과적으로 분석하기 위해 900개의 블로그 제목들을 연결망 분석하였습니다.  일정 이상의 지지도를 가질떄 노드가 표시되고, 일정 이상의 신뢰도를 가질 때 링크가 표시됩니다.  분석 결과 모든 여행지에서 ‘맛집’이라는 키워드가 자주 등장하였으며, 동남아 여행지들은 골프 단어가 자주 등장하였습니다.  이를 통해 여행에서 다른 요소보다 식도락이 매우 중요시되며, 그래서 음식이 뛰어난 도시는 관광산업 개발에 용이하다는 점을 파악할 수 있었습니다.  또한 동남아관광지는 투어형 관광이나 골프를 위한 관광수요가 많은 만큼, 이들을 더욱 발전시키는게 필요하다고 보았습니다. |
| 19 | 한편 일본의 경우 ‘일상’이라는 키워드가 높은 신뢰도를 보였습니다. 이는 우리나라와 일본이 지리적/인문적으로 가까운 만큼 일상을 공유하는 관광, 보다 자연스러운 여행이 많다는 것을 볼 수 있었습니다.  추가로, 유별나게 ‘버블’과 ‘트래블’이라는 단어 간 신뢰도가 높음을 확인했습니다. 이는 저희가 데이터를 모은 시기에 코로나가 있었기 때문에, 코로나와 함꼐 등장한 ‘트래블 버블’이 해당 기간동안 빈번하게 등장했음을 볼 수 있습니다. |
| 20 | 위에서 주장하는 바를 확인하기 위해 지지도를 높여 SNA를 작성했고 재확인하였습니다. 키워드 연관성을 도시별로도 분석하여 확인해보았고 |
| 21 | 긍정과 부정만 도시별로 나눠서도 확인했습니다. |
| 22 | 결론 및 활용방안입니다.  첫번째로 동남아지역은 호텔, 골프, 투어와 연관성이 높았습니다. 이를 보았을 때 동남아지역의 관광상품을 개발할 때 가격경쟁력을 앞세운 럭셔리를 즐길 수 있는 상품을 개발 및 홍보하는 것이 필요하다고 보았습니다. 또한 골프와 투어형 관광이 인기있는 만큼 이들을 묶은 여행상품을 개발하는것도 의미있을것이라고 보았습니다.  두번쨰로 일본지역은 사진, 일상이라는 단어와 연관성이 높았고 투어라는 단어가 거의 등장하지 않았던 것으로 보아, 투어형 관광개발의 필요성이 적음을 보았습니다. 그보다 자유여행과 일본의 일상 및 감성을 즐기는데 중점을 둔 교통권 및 상품권 패키지개발이 필요하다고 보았습니다.  이 점은 외국인의 한국관광을 진흥하는데 참고할 필요가 있다고 느꼈습니다. 우리나라도 일본처럼 가격경쟁력이 낮은 만큼, 보다 ‘일상에 아름다움’과 감성에 중점을 둔 관광상품 개발이 필요합니다. |
| 23 | 종합적으로 한국인의 관광만족도가 높은 지역은 다낭과 홍콩이었으며, 낮은 지역은 후쿠오카와 세부였습니다. |
| 24 | 모든 지역에서 맛집이라는 단어의 빈출도가 매우 높은 것을 보아, 우리나라도 관광개발을 위해서는 식도락 개발 및 홍보가 필요하며, 외국관광시에도 식도락을 강조하는 것이 필요하다고 보았습니다. |
| 25 | 마지막으로 한계입니다.  저희가 크롤링한 최근 90개의 글들은 대부분 코로나시대에 속한 후기들이었습니다. 그러나 앞으로의 관광은 에프터코로나 시대의 관광일 것임으로 이에 맞는 크롤링 기간을 조정할 필요가 있습니다.  더불어, 코로나시대에 여행이 제한적이었다는 점을 감안할 때, 데이터수집시 크롤링 기준시점을 조정한다면 더 많은 인사이트를 추출할 수 있을 것이라고 보았습니다. |
| 26 | 이상 발표 들어주셔서 감사합니다. |