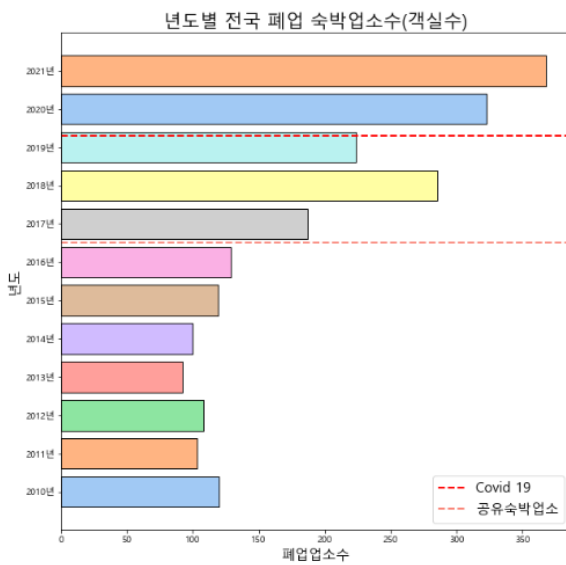


오늘의 집 포스팅 데이터 분석을 통한 코로나 이후 숙박업소 방향성 제시

AI빅데이터융합경영학과

20172834 백찬진

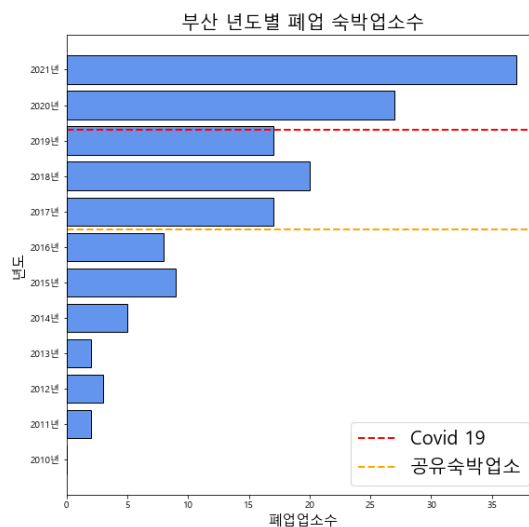
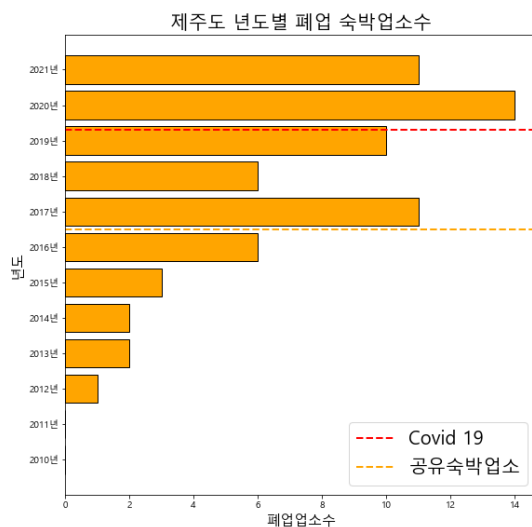
1. 연구 배경



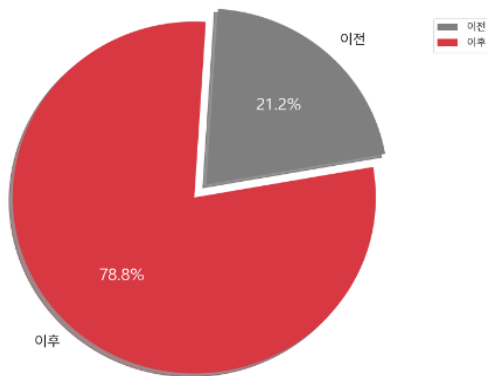
1) 공유숙박업의 장려와 코로나로 인한 숙박업소의 타격

왼쪽의 그래프는 2010년부터 2021년까지 년도별 전국 폐업 숙박업소 수를 나타낸 것이다. 2016년은 박근혜 정부의 추진으로 '우버', 'Airbnb' 등으로 대표되는 공유경제가 활성화되었다. 이에 따라 국내의 많은 숙박업소들이 타격을 입게 되었다. 그래프에서도 그 모습이 잘 나타나 있는데, 2016년 이전까지는 비교적 일정하던 폐업 업소의 수가 2017년에 눈에 띄게 증가한 것을 볼 수 있다. 또, 2019년 12월 코로나19로 인하여 국내 경제에 큰 타격이 있었다. 이에 따라 국내 숙박업소들 또한 많은 타격을 입게 되었고, 많은 수의 업소들이 폐업을 하게 되었다.

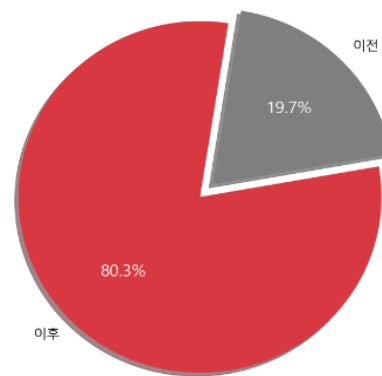
2) 주요 관광도시의 타격



제주도 숙박업소 폐업 현황



부산 숙박업소 폐업 현황



숙박업소의 타격은 주요 관광도시에서 더욱 두드러지는 모습을 보인다. 공유 숙박업 정책이 시작되기 전인 2016년 이전과 이후를 나누어 2010년부터 2021년까지 폐업 업소의 비율을 파이차트로 확인해 보았다. 매년 폐업 업소의 수가 동일하다고 가정할 때, 단순 비율로 본다면 2016년 이후는 총 12년 중 5년이기 때문에 5/12의 비율이 나올 것이라고 예상할 수 있다. 하지만 약 80%의 폐업 업소들이 2016년 이후에 폐업한 것으로 보아 공유숙박업과 코로나의 영향이 매우 크게 작용했음을 알 수 있다. 단순히 머무르다 가는 숙소의 경쟁력은 떨어지기 마련이고, 위치럼 주변 환경의 변화에 타격이 굉장히 크다. 따라서 단순히 '머물다가는' 숙소를 넘어 경쟁력을 가져야한다.

3) 오늘의 집 '집들이' 포스팅 크롤링을 통한 최근 거주공간 트렌드 분석

그렇다면 어떤 경쟁력을 가져야 할까? 다양한 방법이 있겠지만 오늘의 집에 있는 '집들이' 포스팅을 통해 '인기있는 집', '살아보고 싶은 집'의 특징에 대해 분석해 보고 이를 통하여 숙박업소들이 경쟁력을 갖출 수 있는 하나의 방향을 제시하고자한다.

2. 관련연구

1) 크롤링을 통하여 데이터 가져오기

분석을 위하여 오늘의 집 페이지에 있는 데이터를 가져와야한다. 이를 위해 웹 페이지에서 데이터를 추출하는 방법인 크롤링을 사용한다. 크롤링은 정적 크롤링과 동적 크롤링으로 나눌 수 있다. 정적 크롤링은 한 페이지에 모든 정보가 담겨있는 페이지에서 정보를 가져오는 방법이다. 파이썬 bs4의 모듈인 BeautifulSoup을 사용하여 정적 크롤링을 할 수 있다. 동적 크롤링은 사용자가 마우스를 내릴 때마다 페이지에 새로운 정보가 생기는 페이지에서 데이터를 가져오는 방법이며 selenium을 통하여 수행할 수 있다.

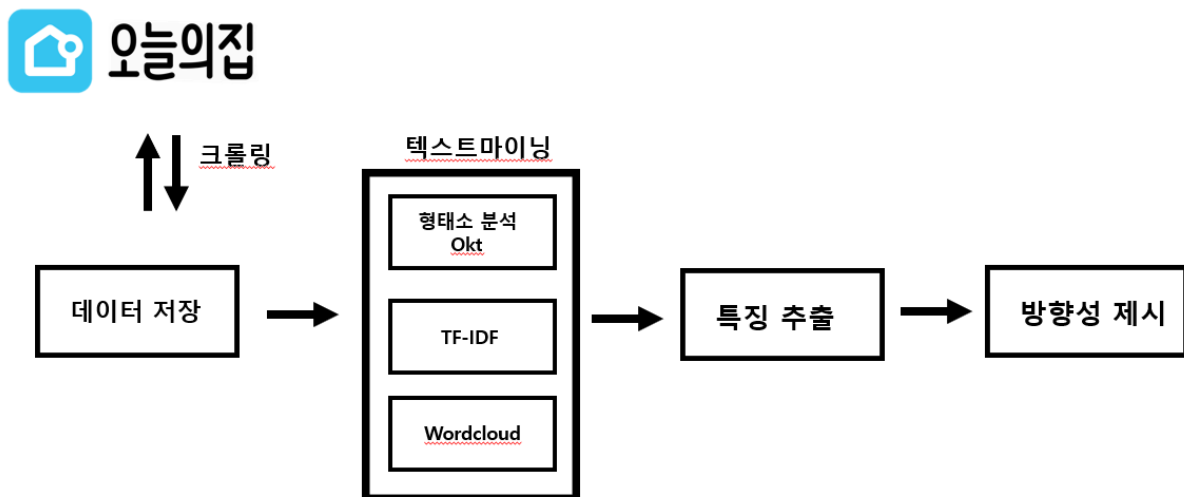
2) 텍스트 데이터 분석을 위한 Oki, TF-IDF

크롤링을 통하여 가져온 데이터를 분석하기 위하여 Oki, TF-IDF방식을 사용한다. Oki는 가져온 데이터에서 필요한 명사만 가져오기 위한 방식이며 다른 방법보다 이모티콘과 같이 인터넷에 사용되는 언어를 해석하기에 용이하다. TF-IDF는 모든 문서에 많이 나온 단어의 중요도를 낮게 평가하고 희귀한 단어의 중요도를 높게 판단하여 단어의 중요도를 판단할 수 있게 해주는 방법이다. 단어의 빈도수를 통하여 중요도를 판단한 것보다 TF-IDF를 사용하면 우리가 문장을 사용할 때 많이 사용하는 단어가 아닌 실제로 중요한 단어를 찾아낼 수 있다. 예를 들어 문장을 사용할 때, 항상 등장하는 단어인 '입니다', '합니다' 등의 중요도는 빈도수를 기준으로 본다면 중요한 단어라고 판단될 수 있지만 실제로는 중요한 단어가 아니다. 이를 TF-IDF를 통하여 판단할 수 있다.

3. 제안 방법론 & 실험 내용

1) 개요

본 장에서는 경쟁력 있는 숙박업소들의 특징을 파악하기 위한 방안을 제시한다. 전체적인 개요는 다음 그림과 같다. 우선 '살고 싶은' 집의 특징을 파악하기 위하여 오늘의 집 '집들이' 페이지에서 최근 인기가 있었던 포스팅 12개, 역대 가장 인기가 있었던 포스팅 12개, 최근에 올라온 포스팅 12개를 크롤링하여 포스팅 내용을 분석하였다. 이후 텍스트 마이닝을 통하여 의미있는 특징을 추출하고 이를 통하여 숙박업소들이 경쟁력을 갖추기 위한 하나의 방향을 제시한다.



2) 데이터 수집을 위한 크롤링

오늘의 집 '집들이' 페이지에서 정렬을 통해 '최근 인기순', '역대 인기순', '최근순'의 순서로 12개씩의 포스팅을 데이터로 가져온다. 포스팅을 가져올 때, 동적 크롤링을 통해 인기글들의 url을 가져온 후 정적 크롤링을 통해 포스팅 데이터를 추출한다. 이를 통하여 최근에 인기가 많은 집의 특징을 분석할 수 있고, 가장 인기가 많았던 집에서도 특징을 찾아낼 수 있다. 마지막으로 최근의 포스팅을 통해 최근 집의 트렌드를 알아보려고 하였다.

3) 포스팅 내용을 통해 '집'에서 중요하게 생각한 요소 파악



수집한 데이터를 Okt 를 통해 명사만 추출한 이후 아파트, 원룸, 오피스텔, 가족, 신혼집, 등 중요한 단어가 아닌 특정 사용자의 경우 빈도수가 높게 책정된 단어는 불용어로 제외를 하였다.

빈도수를 기준으로 wordcloud를 그린 결과가 왼쪽의 그림이고 TF-IDF의 점수를 기준으로 wordcloud를 한 결과가 오른쪽의 그림이다.

이를 통해 사람들이 집에서 중요하게 생각하는 요소를 알 수 있다. 집의 구조적인 측면에서는 주방, 거실, 욕실 순의 순서로 빈도수와 TF-IDF가 높게 나온 것을 보아 주방, 거실, 욕실의 순서로 중요하게 생각한다고 볼 수 있다. 또, 수납의 크기가 두 그림 모두에서 크게 나온 것을 볼 때, 수납이 굉장히 중요한 것을 알 수 있다. 마지막으로 화이트, 인테리어, 조명 등 집의 인테리어를 중요하게 생각한다고 볼 수 있다.

4) 포스팅 제목을 통한 '인기 있는 집', '살고 싶은 집'의 특징 분석



포스팅의 제목은 해당 포스팅을 대표하는 하나의 문장으로 볼 수 있다. 따라서 제목에는 본인이 중요하게 생각하는 것 또는 본인의 집의 컨셉을 담기 마련이다. 왼쪽의 그림이 단순 빈도수로

wordcloud를 한 결과이고 오른쪽의 그림이 TF-IDF를 통하여 결과를 도출한 그림이다.

위의 포스팅 내용과는 다르게 방의 컨셉, 인테리어에 대한 자세한 내용을 확인할 수 있다. 인테리어, 화이트, 우드, 심플, 호텔, 수납 등의 단어를 보면 자연스럽게 '깔끔'이라는 단어가 생각난다. 이를 통해 다수의 사람들에게 인기 있는 집은 '깔끔한 집'이라고 볼 수 있다.

4. 결론



오늘의 집 '집들이' 페이지의 인기 있는 집들의 제목과 포스팅을 분석해본 결과 사람들이 살고 싶고 사람들에게 인기 있는 집은 화이트와 우드를 기본 바탕으로 인테리어 돼있는 호텔 느낌의 심플한 공간으로 볼 수 있다. 왼쪽의 그림은 실제 오늘의 집 페이지에서 크롤링한 포스팅 중 하나의 집 사진이다. Wordcloud를 통해 확인한 결과와 비슷한 형태의 모습을 보인다.

숙박업소들이 폐업을 하게 된 이유에는 다양한 이유가 있을 것이다. 위치, 가격, 서비스, 등 고려해야 할 것들이 많다. 또, 경쟁력을 가질 수 있는 요소 또한 무궁무진하다. 단순히 오늘의 집 포스팅 분석을 통하여 숙박업소들이 나아가야 하는 방향성을 제시하기에는 무리가 있을 것이다. 하지만 오늘의 집 포스팅 데이터를 통하여 분석해본 결과 많은 요소 중 사람들이 좋아하는 숙소의 조건 중 하나가 인테리어에 있다는 것을 확인하였다.

5. 참고자료 출처

출처: 박근혜 정부 공유경제: <https://www.yna.co.kr/view/AKR20160216143700002>

결론 사진: <https://ohou.se/cards/512052>