문화데이터 공모전 대본

P.16

앞에서 합쳐진 데이터프레임을 토대로 별점별 맛집을 지도에 시각화 해보았습니다. 먼저, 카카오맵 크롤링 데이터의 위도와 경도의 값이 전부 NaN이기 때문에 지오코딩을 진행하였습니다. 지오코딩은 주소와 장소 이름과 같은 텍스트 기반의 위치정보를 지리적 좌표로 변환해주는 기술로, 저희가 갖고 있는 데이터프레임의 주소 정보를 가진 ‘ADDR’ 열을 통해 위도와 경도값을 알아냈습니다. 카카오맵에서 크롤링한 데이터를 활용했기에 카카오 개발자 콘솔에서 자바스크립트 API를 발급받아 지오코딩을 하였습니다.

p.17

알아낸 위도와 경도 값으로 지도 시각화 코드를 실행한 결과, 데이터가 약 17,000개로, 범위가 너무 겹쳐져 해석에 어려움이 생긴 문제가 발생하였습니다. 이를 해결하고자 별점 데이터가 있는 Score 값이 각각 3.5이상, 4.0이상, 4.5이상인 경우를 기준으로 데이터프레임을 분할하였고, 옆에 보이시는 사진과 같이 마커 색상도 다르게 지정하였습니다.

p.25

다음은 촬영지 데이터를 동적 지도로 구현해보았는데요.

p.26

활용한 데이터는 한국문화정보원에서 제공한 미디어콘텐츠 영상 내 유명지 데이터로, 한국 드라마, 영화, 티비 프로그램, 아티스트 촬영지 정보를 제공합니다. 뒤에 나올 전처리 과정에서 자주 쓰이는 컬럼들을 짧게 설명드리겠습니다. MEDIA\_TY는 촬영 유형을 나타낸 것으로, 값은 drama, movie, show, artist가 존재하며, TITLE\_NM고 PLACE\_NM은 각각 촬영한 미디어 제목과 장소를 의미합니다. PLACE\_NM 옆에 있는 PLACE\_TY는 촬영한 장소의 유형을 의미하며, cafe(카페), restaurant(식당) , stay(숙박) 등의 값을 가집니다. 이 데이터를 통해 지역별, 아티스트명별 지도 시각화를 진행하고자 다음과 같은 전처리 과정을 거쳤습니다.

p. 27

먼저 지역별 시각화를 위해 주소 열 값 중 앞에서 두 텍스트만을 추출해 지역명 컬럼을 새로 생성하였고, MEDIA\_TY 열을 기준으로 군집화하여 밑에 보이시는 4개의 데이터프레임을 생성하였습니다. 마지막으로 NaN값이 많거나 불필요한 컬럼을 제거하였습니다.

P. 28

지역명 컬럼의 unique값을 뽑은 결과, 총 14개의 지역이 출력되는데, 대한민국의 행정구역 구분을 토대로, 1개의 특별시와 특별자치도, 특별자치시와 6개의 광역시, 8개의 도로 나누어 진행하였습니다. 8개의 도의 경우, 경상북도와 경상남도는 경상도로 통일하여 진행하여 도는 총 5개로 분류하여 진행하였다는 점 참고해주시면 감사하겠습니다. 기본적으로 folium 라이브러리를 활용해 지도를 구현하였고, 지도 검색의 편리성을 더하기 위해 minimap을 사용하여 지도 하단에 작은 맵을 추가하였고, 여러 지역이 겹치는 문제 해결을 위해 markercluster로 여러 마커를 클러스터로 묶어 표시하였습니다. 다음은 경기 지역을 시각화한 코드 예시이며, 위도와 경도 값이 잘못표시된 값들은 구글맵스를 통해 직접 위도와 경도 값을 찾아내어 수정하였습니다.

p.29

다음은 서울 특별시를 시각화한 결과이며, 클러스터를 반복적으로 클릭하여 해당 장소 마커를 누르면 우측 사진과 같이 장소명, 장소유형, 촬영 내용, 주소 정보를 사용자에게 보여줍니다. 서울 지역 외에도 13곳 시각화한 결과를 웹페이지에서 확인해볼 수 있습니다.

p.30

다음은 아티스트명별로 지도 시각화하는 과정을 설명드리겠습니다. 앞에서 보여드린 한류(K-POP) 키워드 데이터에서 상위 10개의 아티스트명을 다음과 같이 추출하여 촬영지 데이터에 해당 아티스트명이 존재하는지 여부를 파악하였습니다.

p.31

파악한 결과, 각각 44위와 45위에 해당하는 jyp와 bigbang 값이 존재하지 않아 51위에 해당하는 세븐틴으로 대체한 결과 다음과 같이 상위 10명의 아티스트가 확정되었습니다. 이를 토대로 검색한 아티스트명과 일치하는 장소를 필터링하는 함수를 우측과 같이 구현하였으며, 해당 장소의 개수 정보를 제공하여 전국과 지역별로 지도에 시각화하였습니다.

p.32

다음은 TITLE\_NM의 값을 방탄소년단일 경우의 시각화 결과이며, 다른 아티스트명을 입력할 시, 해당 함수가 실행되어 다른 지도도 볼 수 있습니다.