section3 프로젝트 발표를 시작하겠습니다

저는 원래 국내의 음원 사이트 멜론,지니,벅스,플로 등의 순위와 좋아요 수를 통해서 빌보드 코리아의 순위를 예측해 보려고 했습니다. 참고로 빌보드 코리아는 데이터 집계를 네이버 바이브, 한테오 글로벌, yg plus에서 하고 있습니다. 하지만 프로젝트를 진행하면서 시간이 부족한 관계로 멜론의 좋아요수를 통한 순위 예측으로 주제를 급하게 바꾸었다는 점 참고해주시기 바랍니다.

데이터를 가져올 때 웹크롤링을 통해 각 음원 사이트에서 순위, 아티스트명, 앨범명, 발매일로부터의 날 수, 장르를 가져왔습니다.

이렇게 가져온 데이터를 db파일로 저장하여 dbeaver를 통해 csv 파일로 변환하여 파이썬을 통해 전처리를 진행하였습니다. 발매가 된지 오래되었지만 차트에 계속 있는 음악도 있기 때문에 좋아요의 수를 발매일로부터 지난 날의 수로 나누었습니다. 사람들이 좋아요를 음악이 발매되고 나서 초반에 누르기 때문에 발매된지 100일에서 1년사이의 음악은 모두 100으로 나누었고 그 이후는 365일로 나누었습니다. 이렇게 전처리한 데이터를 다시 db파일에 넣고 메타베이스에 업로드 하였습니다.

API 서비스 개발에서 로컬호스트를 통해 사용자에게 좋아요수와 일별 좋아요수를 입력받게 했고

그것을 토대로 순위를 예측하게 했습니다.

데이터 분석용 대시보드는 메타베이스를 통해서 만들었습니다.

이렇게 멜론차트와 좋아요수

일별 좋아요수

또한 지니는 곡에 대한 좋아요 수가 없고 앨범에 대한 좋아요 수만 얻을 수 있었기 때문에

예측하려는 결과가 거의 무의미 하다고 볼 수 있습니다.

총 좋아요 수의 상위 10개 항목을 정리한 그래프이고요

일일 좋아요 수의 상위 10개 항목의 그래프입니다.

이런 그래프들을 보아 저의 모델이 매우 성능이 떨어지는 부분도 있지만 날짜별 자세한 좋아요수의 데이터를 알 수 있다면 더 정확했을 것이라고 추정됩니다. 결과적으로 좋아요수는 차트 순위에 영향이 조금 있다고 볼 수 있습니다.