Playlist Recommendation

회귀분석팀 권남택 윤주희 진효주 한유진 황유나

[CONTENTS]

1

주제 선정

- 주제 선정 배경
- 분석 개요

2

데이터 확인

- 데이터 소개
- 데이터 연결

3

데이터 탐색

- 메타데이터
- 학습데이터
- TVT 비교
- LDA

4

다음주 예고

- Word2vec
- 3주차 예고

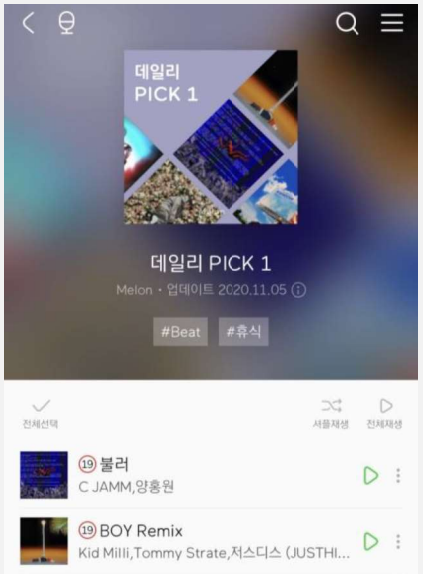
01. 주제 선정

문제 정의 분석 개요

- 음악 추천이 중요한 이유

음악이 필요한 순간

Melon

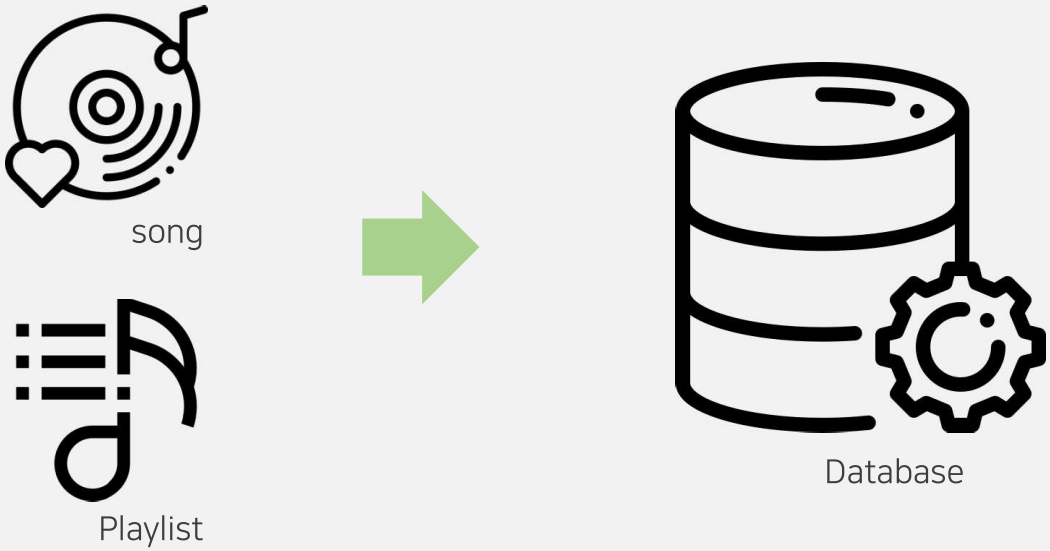


멜론에서 서비스하는 곡은 수천만…!
모든 곡을 다 들을 수 없다!

01. 주제 선정

문제 정의 분석 개요

- 멜론이 제공하는 플레이리스트 추천



곡과 플레이리스트는 저장되기 때문에 시간에 민감하게 변하지 않음!

Like Netflix…!

02. 데이터 확인

데이터 소개 데이터 연결

genre_gn_all.json

	곡 장르코드
1	gnr_code : 장르 고유 코드
2	gnr_name : 장르명

train/val/test.json

	플레이리스트
1	tags
2	id
3	songs
4	like_cnt
5	updt_date

song_meta.json

	곡 별 메타데이터
1	song_gn_dtl_gnr_basket
2	issue_date
3	album_name
4	album_id
5	artist_id_basket
6	song_name
7	song_gn_gnr_basket
8	artist_name_basket
9	Id

02. 데이터 확인

데이터 소개 데이터 연결

- train데이터에 곡 메타데이터 조인

song_meta.json

아티스트_id	곡 명	장르	아티스트_id	곡_id
[2727]	Feelings	['GN0900]	['Various Artists']	0
[29966]	Bach : Partita No. 4 In D Major, BWV 828	['GN1600]	['Murray Perahia']	1
[437]	스치듯 안녕	['GN0100]	['윤종신']	707986
[729868]	숲의 빛	['GN1800]	['Nature Piano']	707987
[895]	Queen 명곡 멜로디	['GN0600]	['김경호']	707988

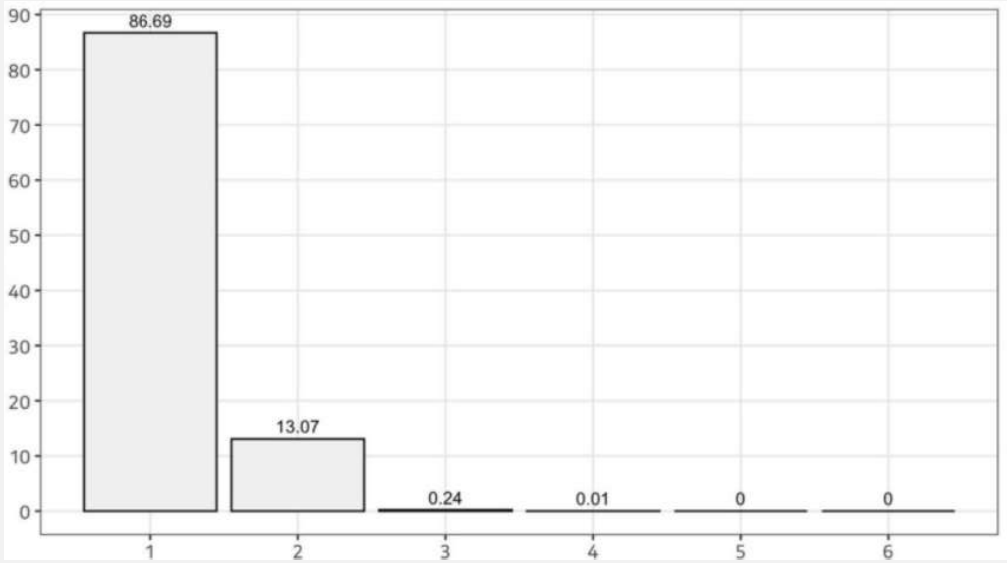
train.json

playlist name	수록 곡 id	좋아요 개수
여행같은 음악	[525514, 129701, 383374, 562083, 297861, 13954...	71
요즘 너 말야	[432406, 675945, 497066, 120377, 389529, 24427...	1
편하게, 잔잔하게 들을 수 있는 곡.-	[83116, 276692, 166267, 186301, 354465, 256598...	17
퇴근 버스에서 편히 들으면서 하루를 마무리하기에 좋은 POP	[533534, 608114, 343608, 417140, 609009, 30217...	4
FAVORITE POPSONG!!!	[26008, 456354, 324105, 89871, 135272, 143548,...	17

03. 데이터 탐색

- 곡 메타데이터(한 곡이 포함하는 장르 수)

장르 개수	곡 수	비율
1	612806	86.686
2	92378	13.067
3	1694	0.24
4 이상	52	0.007



대체로 노래 한 곡에 한 개의 장르가 할당되지만
약13%의 비율로 2개 이상의 장르를 가진다는 것을 확인할 수 있었다!

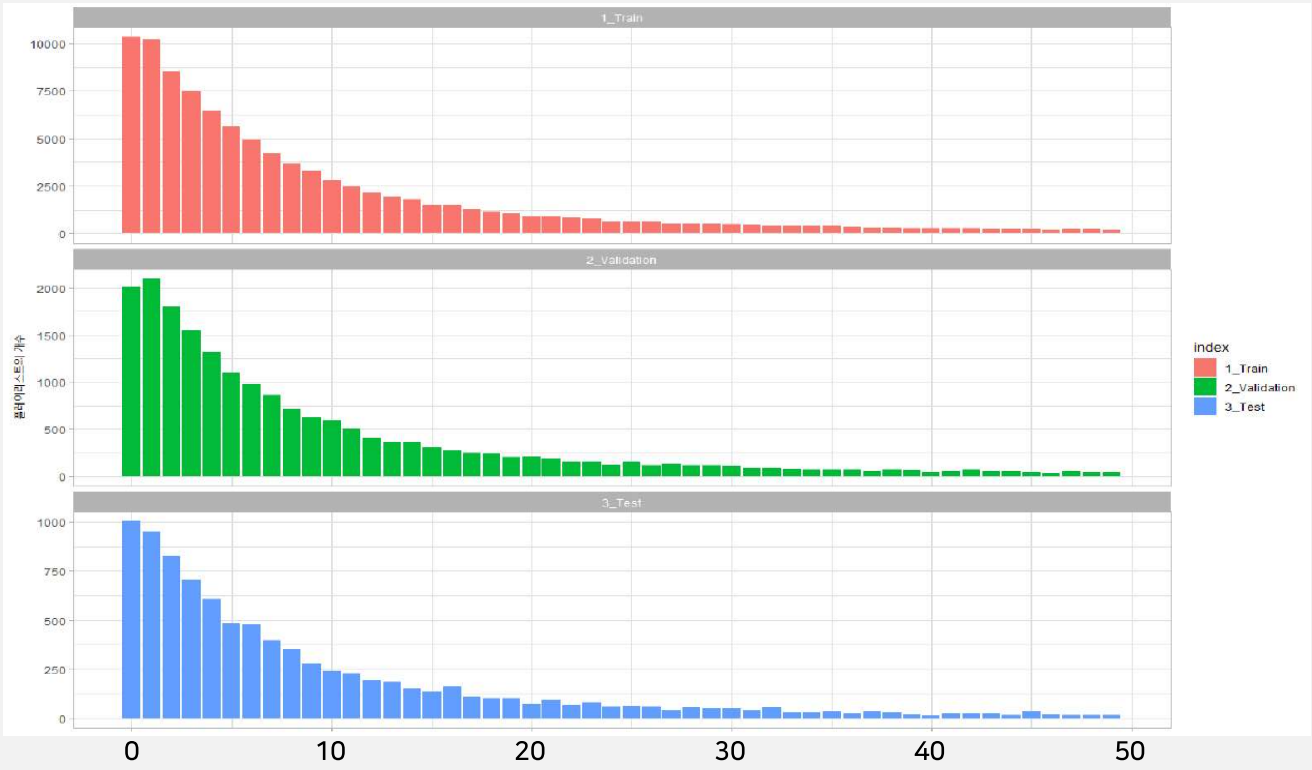
03. 데이터 탐색

- 플레이리스트 전체 태그 확인

워드클라우드한 결과를 살펴보면
'기분전환', '감성', '휴식 ' 과 같은
태그들이 많은 모습이다

03. 데이터 탐색

- TVT 플레이리스트별 '좋아요 개수' 분포



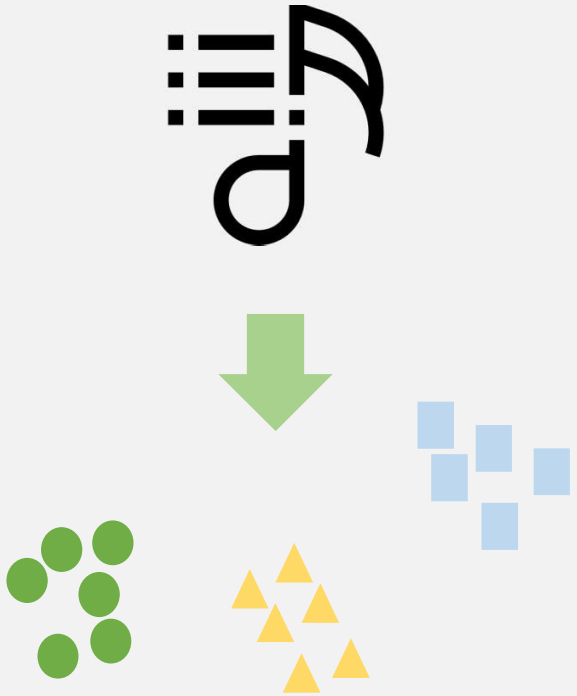
03. 데이터 탐색

- LDA (Latent Dirichlet Allocation)

✓ 비지도 학습 방법

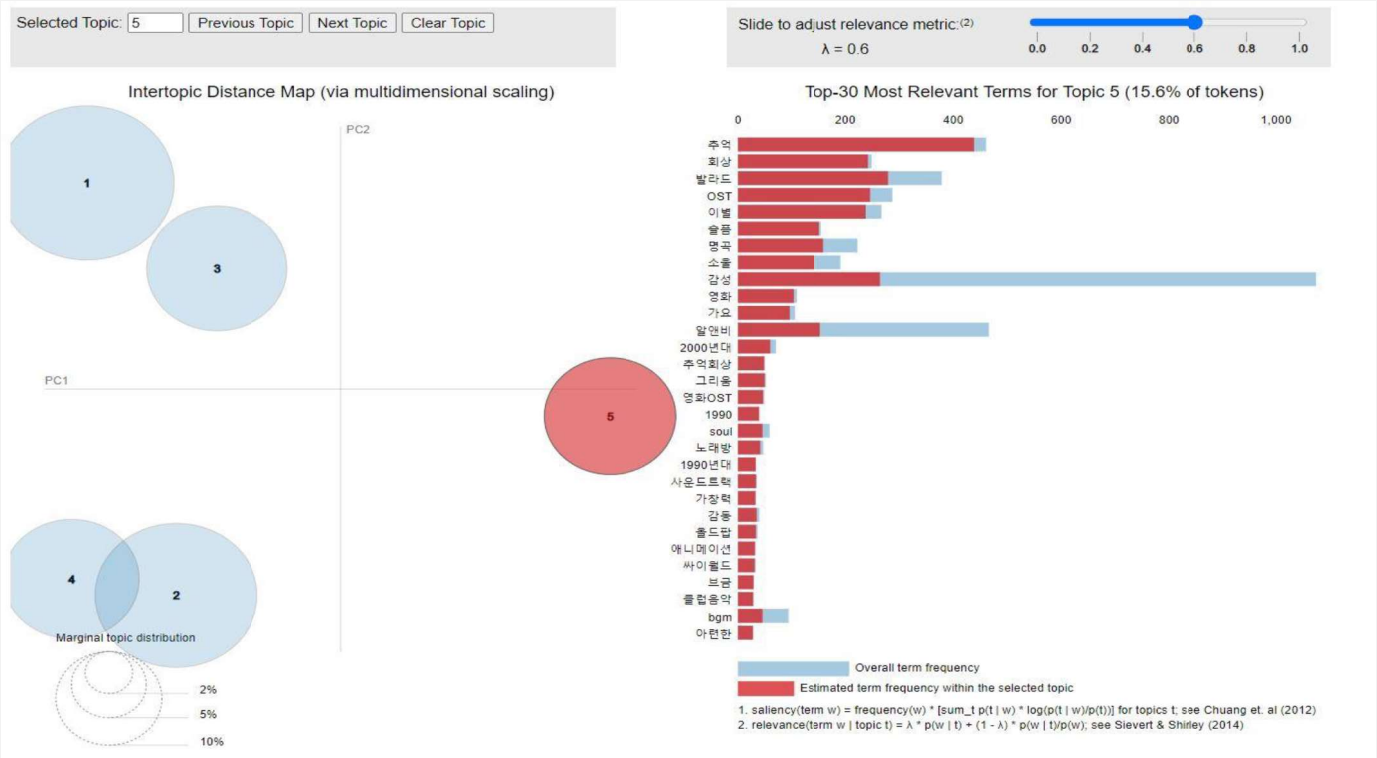
✓ 토픽개수 k는 분석가가 결정
(혹은 perplexity로 결정)

태그들을 잠재변수인 '토픽'의 실현으로 이해
각 토픽의 단어는 디리클레분포를 따름



03. 데이터 탐색

- 토픽 5 : 추억의 노래



04. 3주차 예고

- Word2vec 임베딩

예시

"배가 고파서 나누미 떡볶이를 3인분을 먹었다."

CBOW 적용

"____ 고파서 나누미 떡볶이를 3인분을 먹었다."

빈 칸의 앞 뒤 문맥을 통해서
"배가" 가 들어가야함을 추론

Word2vec이란 ?

두 단어의 유사한 정도를 이용해 주변 단어인지 아닌지를 예측하는 모델

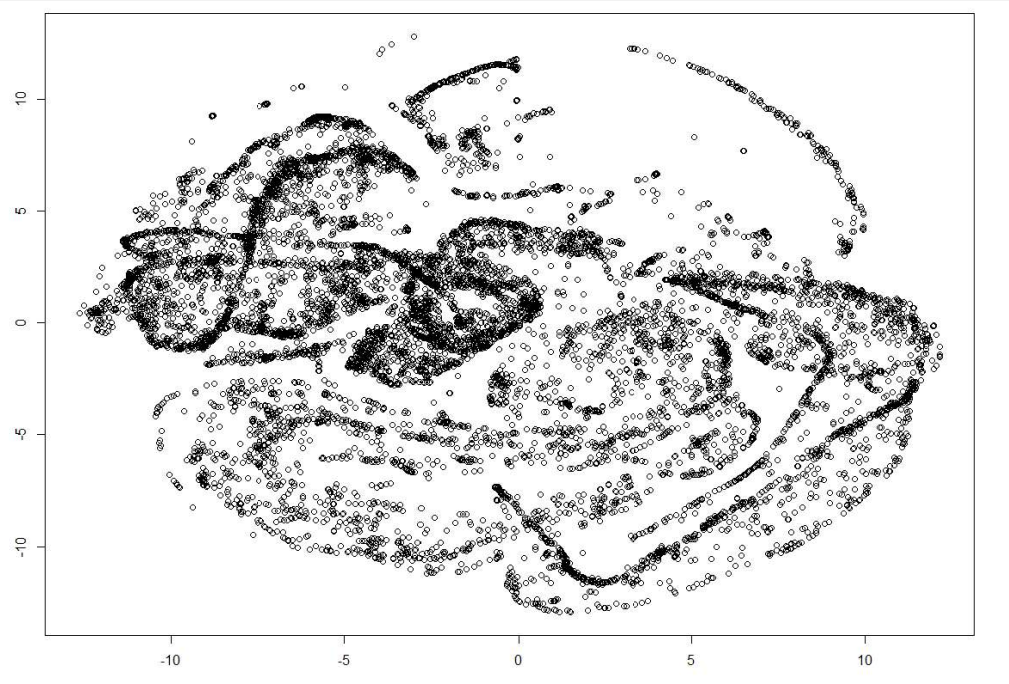
Skip-gram

CBOW(Continuous Bag of Words)

주어진 단어에 대해 앞 뒤로 C/2개씩
총 C개의 단어를 사용하여
단어를 맞추는 모델

04. 3주차 예고

- 플레이리스트 곡 tsne 시각화, 하지만…



명확한 군집형태가 드러나지 않음!