

Melón

멜론 플레이리스트 데이터 Visualization



주제 소개

멜론 플레이리스트 데이터를 활용한 시각화

Melón

플레이리스트에 들어 있는 곡의 절반만 보여주고,
나머지 절반을 예측하는 모델을 만드는 ‘멜론 플레이리스트 예측 대회’의 데이터셋을 활용함

데이터셋 목록: 1) 장르 분류 데이터 2) 플레이리스트 데이터 (대회에서의 train set) 3) 곡 데이터

데이터 소개 - 장르 데이터 (df_genre)

	Code	Genre
0	GN0100	발라드
1	GN0101	세부장르전체
2	GN0102	'80
3	GN0103	'90
4	GN0104	'00
5	GN0105	'10-
6	GN0200	댄스
7	GN0201	세부장르전체
8	GN0202	'80
9	GN0203	'90

- 254개의 행, 2개의 열
- 00으로 끝나는 대분류 30개와 그 외의 소분류 224개로 구분됨
- 트리맵 형태를 통해 장르를 한 눈에 파악하고자 하였음.

데이터 소개 - 플레이리스트 데이터 (df_pl)

	tags	id	plylst_title	songs	like_cnt	updt_date
0	[락]	61281	여행같은 음악	[525514, 129701, 383374, 562083, 297861, 13954...	71	2013-12-19 18:36:19.000
1	[추억, 회상]	10532	요즘 너 말야	[432406, 675945, 497066, 120377, 389529, 24427...	1	2014-12-02 16:19:42.000
2	[카페, 잔잔한]	76951	편하게, 잔잔하게 들을 수 있는 곡.-	[83116, 276692, 166267, 186301, 354465, 256598...	17	2017-08-28 07:09:34.000
3	[연말, 눈오는날, 캐럴, 분위기, 따듯한, 크리스마스캐럴, 겨울노래, 크리스마스,...]	147456	크리스마스 분위기에 흥뻑 취하고 싶을때	[394031, 195524, 540149, 287984, 440773, 10033...	33	2019-12-05 15:15:18.000
4	[댄스]	27616	추억의 노래 ㅋ	[159327, 553610, 5130, 645103, 294435, 100657,...]	9	2011-10-25 13:54:56.000

- 115071개의 행, 6개의 열
- [tags] : 플레이리스트에 어울리는 태그
- [plylst_title] : 플레이리스트 이름
- [songs] : 플레이리스트에 수록된 곡 아이디
- [like_cnt] : 플레이리스트의 좋아요 수

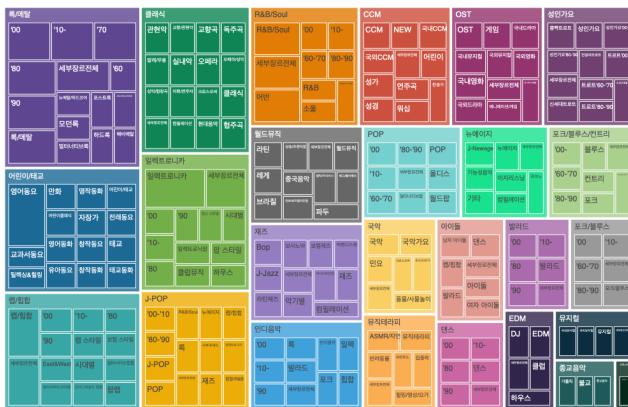
데이터 소개 - 노래 메타데이터 (df_meta)

	song_gn_dtl_gnr_basket	issue_date	album_name	album_id	artist_id_basket	song_name	song_gn_gnr_basket	artist_name_basket	id
0	[GN0901]	20140512	불후의 명곡 - 7080 추억의 알개시대 팝송베스트	2255639	[2727]	Feelings	[GN0900]	[Various Artists]	0
1	[GN1601, GN1606]	20080421	Bach : Partitas Nos. 2, 3 & 4	376431	[29966]	Bach : Partita No. 4 In D Major, BWV 828 - II....	[GN1600]	[Murray Perahia]	1
2	[GN0901]	20180518	Hit	4698747	[3361]	Solsbury Hill (Remastered 2002)	[GN0900]	[Peter Gabriel]	2
3	[GN1102, GN1101]	20151016	Feeling Right (Everything Is Nice) (Feat. Popc...	2644882	[838543]	Feeling Right (Everything Is Nice) (Feat. Popc...	[GN1100]	[Matoma]	3
4	[GN1802, GN1801]	20110824	그남자 그여자	2008470	[560160]	그남자 그여자	[GN1800]	[Jude Law]	4

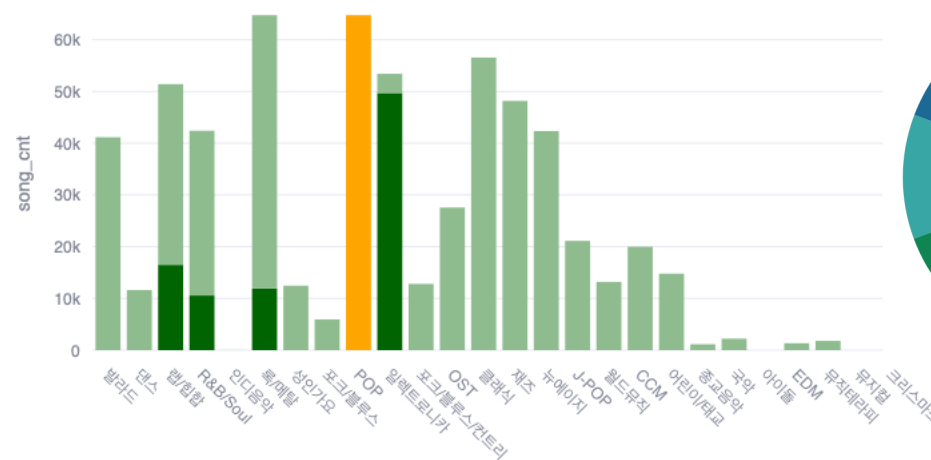
- 707989개의 행, 9개의 열
- 모든 곡은 1개의 대장르[song_gn_gnr_basket]와 1개 이상의 소장르[song_gn_dtl_gnr_basket]에 속함
- [artist_name_basket] [artist_id_basket] : 가수 및 가수 아이디
- [album_name] [album_id] : 곡이 수록된 앨범명 및 아이디
- [issue_date] : 곡 발매일, string 형식으로 되어 있어 year과 month만 추출해서 시각화 진행

장르 분석

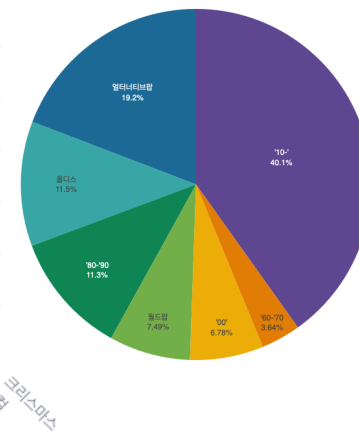
전체 장르 트리맵



대분류 장르(x)별 곡 수(y)



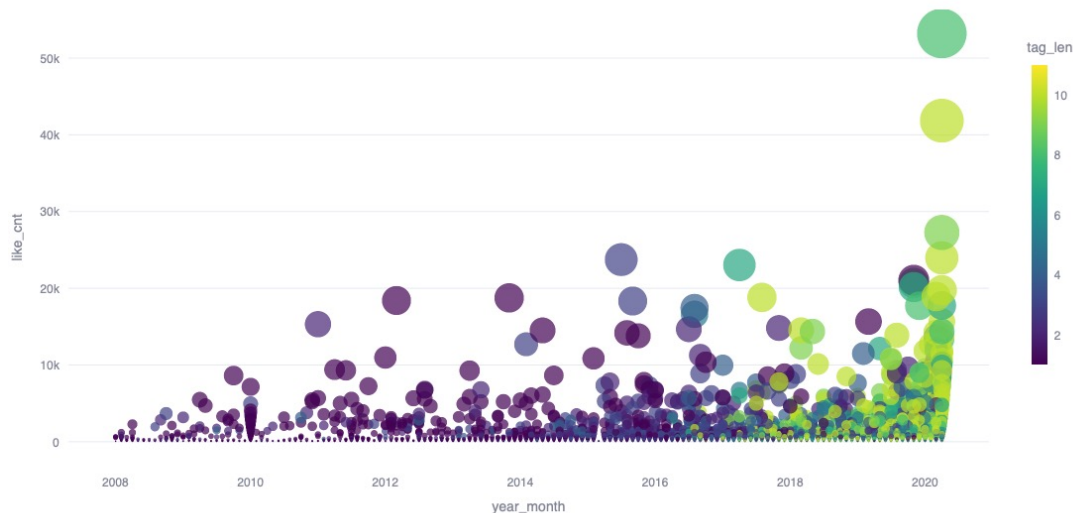
POP 장르의 소분류



- 가장 먼저 트리맵을 통해 전체적인 그림을 파악한 뒤 구체적으로 각각의 비율을 살펴봄
- 대분류 장르의 곡 수를 나타낼 때 파이 차트를 활용하려 했으나, 비율이 높은 장르들이 꽤 많아 bar chart가 더 효과적임을 확인함
- 국내와 해외를 별개의 대분류 장르로 보았을 때, 단일 장르로서 비중이 가장 큰 것은 'POP'
- POP 장르의 소분류를 확인해 보았을 때 2010년 이후가 40.1%, 얼터너티브팝이 19.2%를 차지함

플레이리스트 분석

연도(x)에 따른 태그 수(color) & 좋아요 수 (y)



- 좌측 그래프를 통해 아래와 같은 사실을 확인할 수 있음
 - 1) 최근에 만들어진 플레이리스트들의 태그 수가 많다.
 - 2) 좋아요 수가 많은 플레이리스트는 최근에 만들어진 경우가 많다.

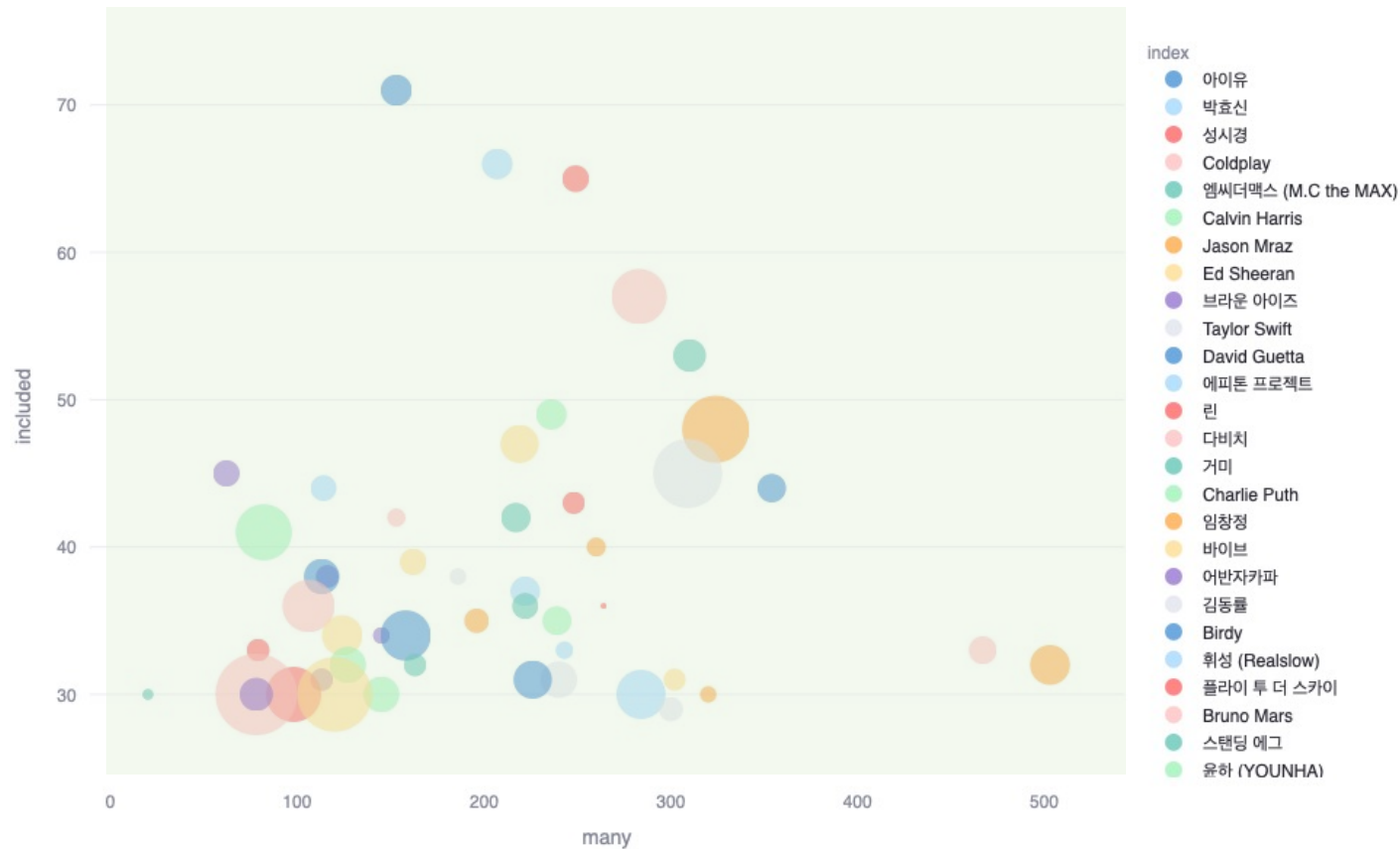
연도(x)별 발매된 곡 수



- 이유를 파악하기 위해 플레이리스트 데이터와 곡 데이터를 결합함
- 그 결과 최근 발매된 노래의 비율이 더 많은 것을 확인할 수 있었고, 최근 플레이리스트일수록 최신 노래를 수록할 확률이 높으므로 위와 같은 결과가 나타날 수 있겠다고 해석함

가수 분석 (상위 150개 플레이리스트)

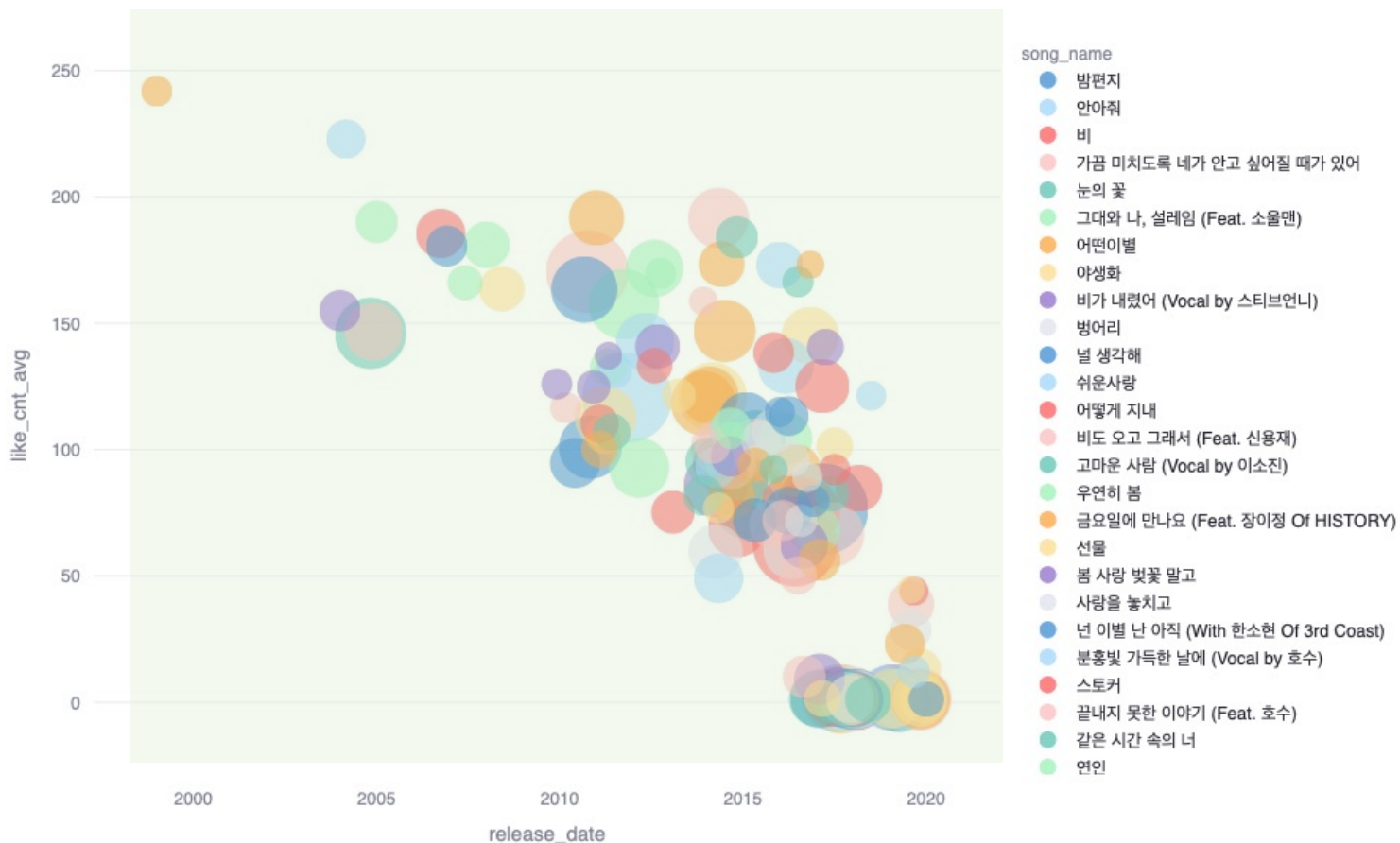
가수별 보유곡 수(x) & 수록곡 수(y)
& 수록된 플레이리스트의 평균 좋아요 수(size)



- 좋아요 수 기준 상위 150개 플레이리스트에 가장 많이 수록된 가수들을 확인함 (아이유, 박효신, 성시경, Coldplay 등)
- 특히 보유 곡이 많다고 해서 플레이리스트에 많이 수록되지는 않는 것을 확인할 수 있음

곡 분석 (상위 150개 플레이리스트)

곡 발매연도(x)에 따른
곡의 플레이리스트 수록 횟수(size)와 플레이리스트의 평균 좋아요 수(y)



- 최신 발매된 곡일 수록 해당 곡이 포함된 플레이리스트의 수가 적어 평균 좋아요 수가 낮은 것으로 보임
- 전체에서 가장 많이 수록된 곡:
아이유 - 밤편지 (2175회)
정준일 - 안아줘 (2121회)

가수 & 장르 & 곡 분석



네트워크 그래프



가수 이름: 아이유



대장르: 발라드, OST 등



소장르: '10-, '00, 국내 드라마 등



곡 이름: 나만 몰랐던 이야기, Someday 등

- 점이 가장 많이 모여있는 부분을 보면 해당 가수가 어떤 장르를 많이 불렀는지 한 눈에 파악 가능하며, 대장르와 소장르의 계층 구조 또한 쉽게 이해 가능함

TOP 10 플레이리스트 분석

Playlist Title	Like Count
HOT TRENDY POP: 놓쳐선 안될 'POP' (매주 업데이트)	53211
❤️때겔룩님 TAKE A LOOK 플레이리스트❤️	41844
듣다보면 '우와!' 하고 제목을 보게되는 팝	27268
약속 있어? 외출 전, 기분 UP 하고 싶을 때 들어봐! [매주 목요일]	23965
감성이 터지는 팝음악들	23732
신비롭거나 몽환적 느낌이 나는 곡	23032
♪중독성강한음악들♪	21045
제목이 기억안나서 못들었던 뉴에이지	20677
여자들 취향저격하는 달달한 노래	20149
불 꺼진 방 침대 속 자기 직전... 듣기 좋은 노래 🌟	19758



- 좋아요 수가 가장 많은 상위 10개의 플레이리스트에는 '드라이브', '신나는', '감성', '기분전환' 등 태그가 많음을 알 수 있음
- 호기심을 자극하는 플레이리스트 이름과 감정과 관련된 태그들이 두드러짐

Word2Vec를 활용한 태그 데이터 분석 및 추천

1. 한국어 Word2Vec 만들기

tag 데이터와 plyst_title을 하나의 문장으로 만들고, 정규 표현식을 통해 한글이 아닌 경우 제거하는 전처리 진행

2. 형태소 분석기를 사용하여 토큰화 후 불용어 제거

```
stopwords = ['의','가','이','은','들','는','좀','잘','강','과','도','를','으로','자','에','와','한','하다']
```

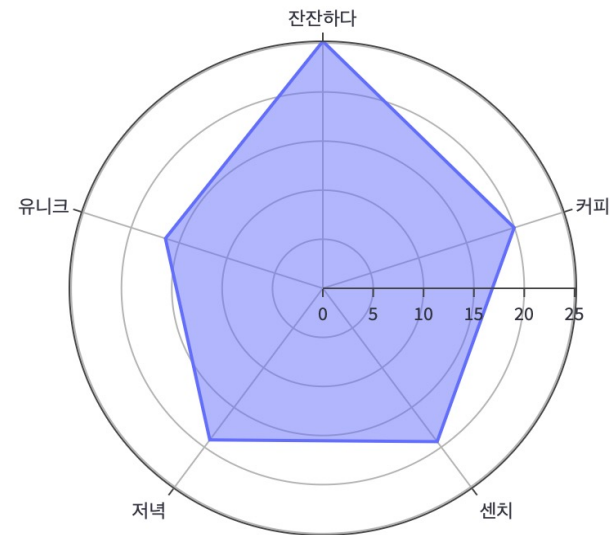
3. Word2vec의 CBOW 알고리즘을 사용하여 토큰화 된 단어 학습

4. Input으로 태그 데이터를 받아 다음과 같이 유사도를 기준으로 제일 비슷한 태그 도출

```
ex) print(model.wv.most_similar("드라이브"))
```

```
[('리드미컬', 0.6551075577735901), ('산책', 0.6256625652313232), ('운전 ', 0.620262861251831), ('여행', 0.6087409257888794), ...]
```

Word2Vec를 활용한 태그 데이터 분석 및 추천



- 사용자의 플레이리스트들의 태그 데이터를 하나씩 input으로 받아서 각 output 태그 데이터마다 유사도를 통합하여 상위 5개의 태그 데이터를 라디오 차트로 시각화해서 추천
- 입력 태그를 바탕으로 유사 태그를 추천해줌으로써 사용자의 취향과 유사하면서 보다 다양한 플레이리스트를 제공할 수 있음

분석 및 시각화 결과 종합

분석 결과 요약

- (장르) 플레이리스트에 가장 많은 대분류 장르: POP
- (플레이리스트) 최근에 생성된 플레이리스트일수록 태그 수가 많음
- (플레이리스트) 좋아요 수가 많은 플레이리스트일 경우 최근에 만들어졌을 확률이 높음
- (가수) 보유 곡이 많다고 해서 플레이리스트에 많이 수록되지는 않는 것을 확인할 수 있음
- (곡) 플레이리스트에 가장 많이 수록된 곡: 아이유 - 밤편지
- (곡) 최신 발매된 곡일 수록 해당 곡이 포함된 플레이리스트의 평균 좋아요 수가 낮음

분석 의의

무엇보다 구축한 시스템을 통해 사용자가 입력한 태그를 바탕으로 유사한 태그를 추천할 수 있으며,
분석 결과를 통해 입력 태그가 포함된 인기 플레이리스트 및 가수를 추천할 수 있어 사용자에게 다양한 음악 경험을 제공할 수 있음

1

주제 소개

2

데이터 소개

3

시각화 결과

4

결론

5

Streamlit

Streamlit

Q & A

감사합니다