

# Music Recommendation System

Group: Double Eleven

Yang Xu, Muyan Li

# 1. EDA

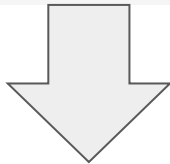
We have 2 datasets: one of users' playlists, one of songs' features.

	user_id	artists	track_name	playlist_name
0	9cc0cfd4d7d7885102480dd99e7a90d6	Elvis Costello	(The Angels Wanna Wear My) Red Shoes	HARD ROCK 2010
1	9cc0cfd4d7d7885102480dd99e7a90d6	Elvis Costello & The Attractions	(What's So Funny 'Bout) Peace, Love And Unders...	HARD ROCK 2010
2	9cc0cfd4d7d7885102480dd99e7a90d6	Tiffany Page	7 Years Too Late	HARD ROCK 2010
3	9cc0cfd4d7d7885102480dd99e7a90d6	Elvis Costello & The Attractions	Accidents Will Happen	HARD ROCK 2010
4	9cc0cfd4d7d7885102480dd99e7a90d6	Elvis Costello	Alison	HARD ROCK 2010

track_id	artists	album_name	track_name	popularity	duration_ms	explicit	danceability	energy	...	loudness	mode	speechiness
5SuOikwiRyPMVolQDJUGSV	Gen Hoshino	Comedy	Comedy	73	230666	False	0.676	0.4610	...	-6.746	0	0.1430
4qPNDBW1i3p13qLCt0Ki3A	Ben Woodward	Ghost (Acoustic)	Ghost - Acoustic	55	149610	False	0.420	0.1660	...	-17.235	1	0.0763
1iJBSr7s7jYXzM8EGcbK5b	Ingrid Michaelson;ZAYN	To Begin Again	To Begin Again	57	210826	False	0.438	0.3590	...	-9.734	1	0.0557
6lfxq3CG4xtTiEg7opyCyx	Kina Grannis	Crazy Rich Asians (Original Motion Picture Sou...	Can't Help Falling In Love	71	201933	False	0.266	0.0596	...	-18.515	1	0.0363
5vjLSffimilP26QG5WcN2K	Chord Overstreet	Hold On	Hold On	82	198853	False	0.618	0.4430	...	-9.681	1	0.0526

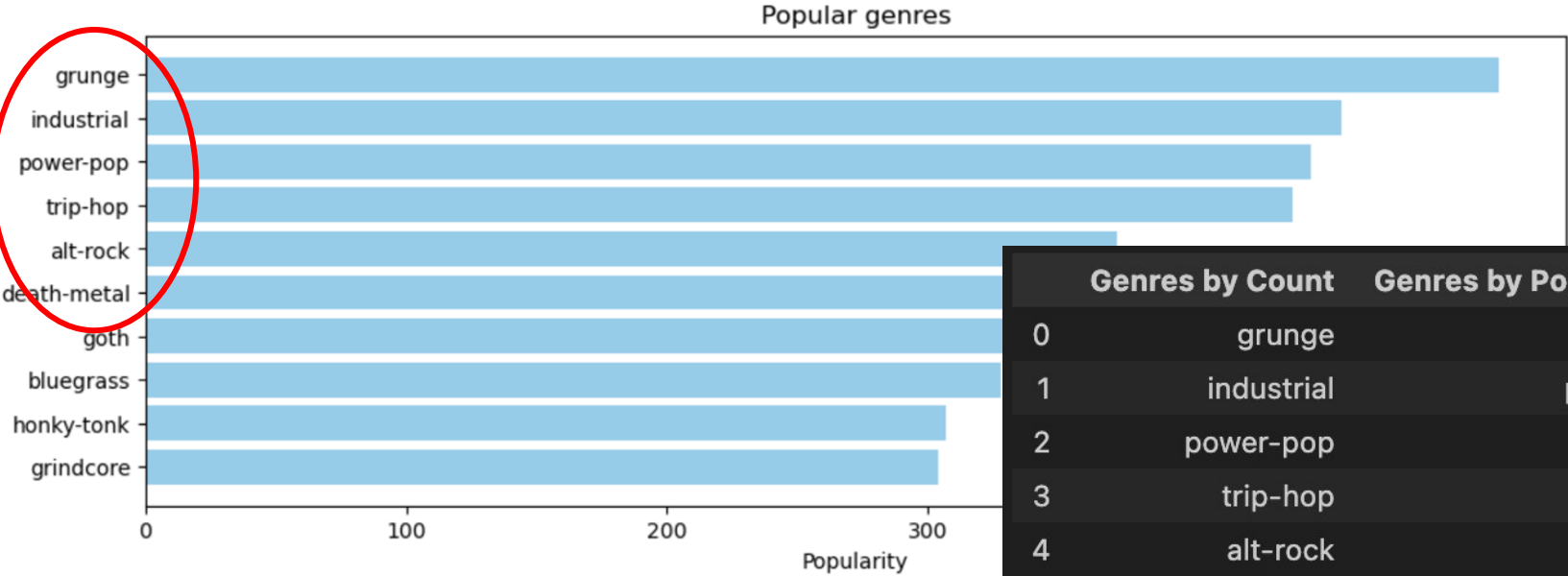
Duplicate songs exist in either dataset:

track_id	artists	album_name	track_name	popularity	duration_ms	explicit	danc
5SuOikwiRyPMVoIQDJUgSV	Gen Hoshino	Comedy	Comedy	73	230666	False	
5SuOikwiRyPMVoIQDJUgSV	Gen Hoshino	Comedy	Comedy	73	230666	False	
5SuOikwiRyPMVoIQDJUgSV	Gen Hoshino	Comedy	Comedy	73	230666	False	
5SuOikwiRyPMVoIQDJUgSV	Gen Hoshino	Comedy	Comedy	73	230666	False	



Drop songs with same artists and name for the model.

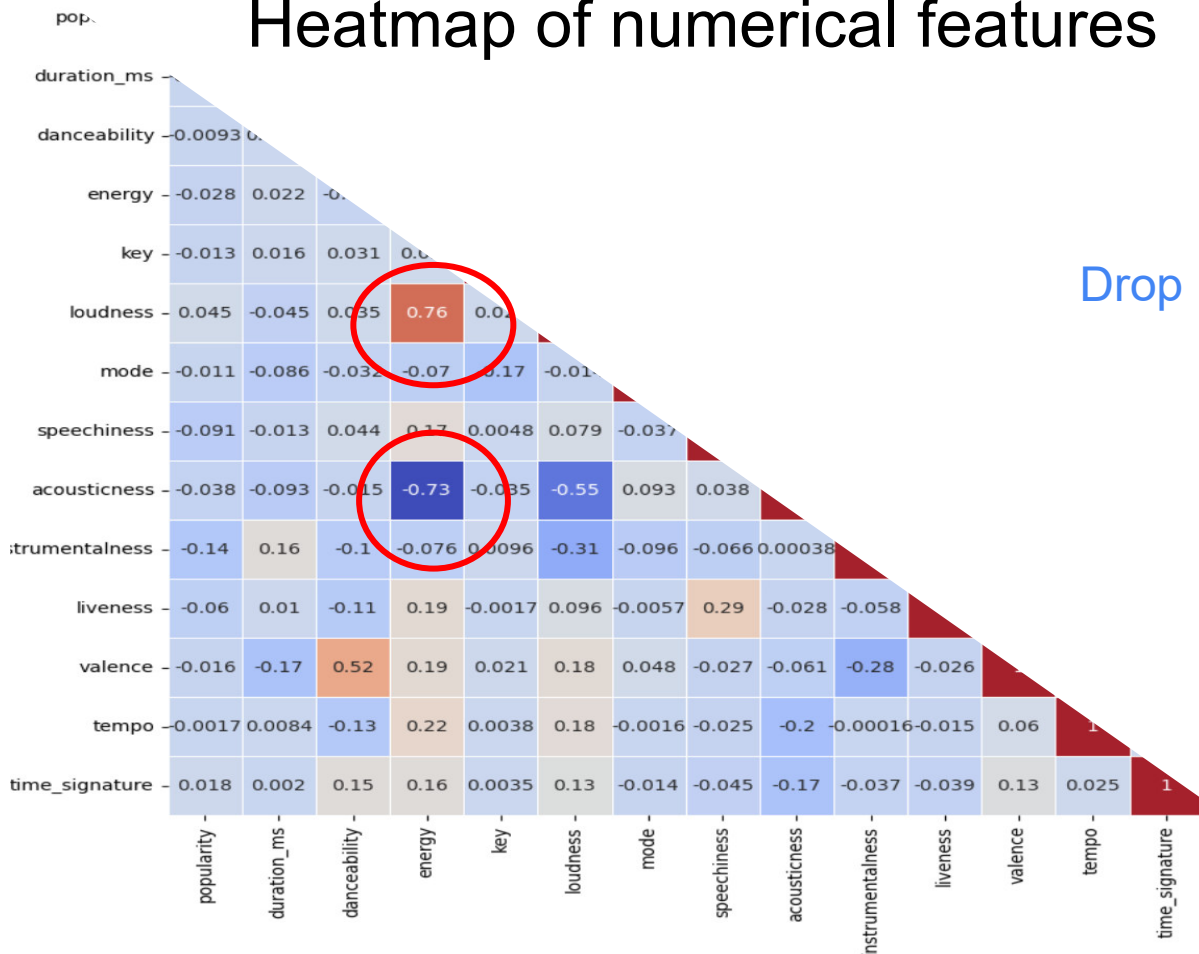
First let's take a look at popular genres.



*Rock wins!*

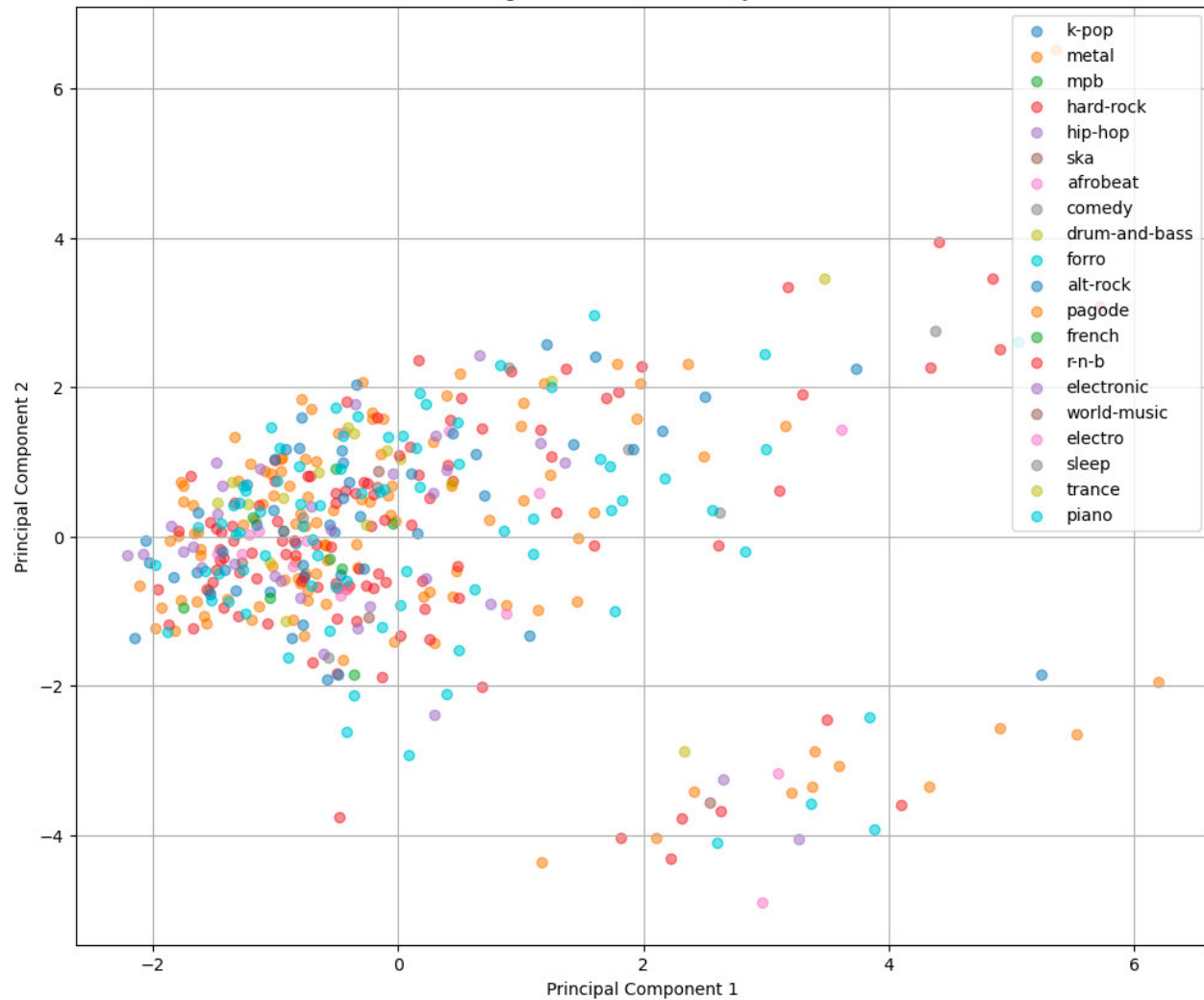
	Genres by Count	Genres by Popularity
0	grunge	chill
1	industrial	pop-film
2	power-pop	metal
3	trip-hop	k-pop
4	alt-rock	ambient
5	death-metal	singer-songwriter
6	goth	hard-rock
7	bluegrass	cantopop
8	honky-tonk	anime
9	grindcore	indie-pop

# Heatmap of numerical features



Drop 'energy' in the model.

2D PCA of Song Features for Randomly Selected Genres



## 2. Model - Main idea

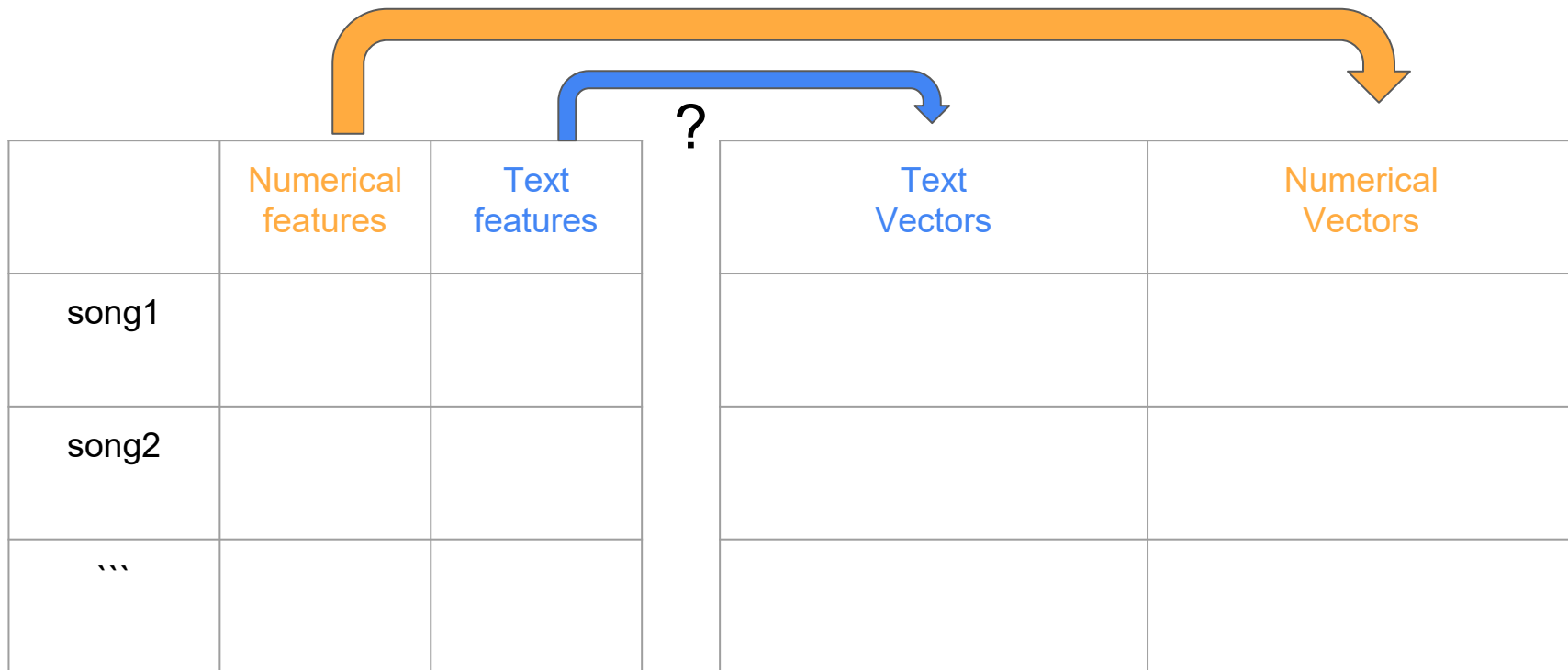
- Use vectors to represent songs
- For a user
  - Get his/her taste vector based on the song he/she likes
  - Identify  $k$  song vectors closest to the taste vector
  - Recommend them !

## 2. Model - Main idea

- Use vectors to represent songs
- For a user
  - Get his taste vector based on the song he/she likes
  - Identify the song closest to the taste vector
  - Recommend

How to define  
the vector for  
each song?





Our target .....

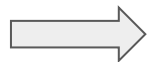
The songs liked by the same user should be closer !

Our target .....

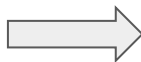
The songs liked by the same user should be closer !

Word2vec !

Words in  
similar contexts



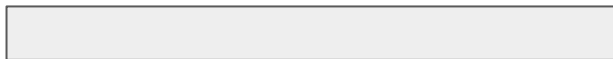
**Word2vec**



Word vectors  
with closer distance

Sentence 1

[



,

Sentence 2



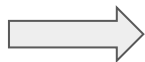
,

Sentence n

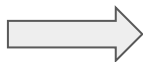


]

Song features  
liked by same user



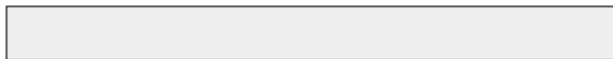
**Word2vec**



Feature vectors  
with closer distance

User 1

[



,

User 2



,

User n



]

# Example

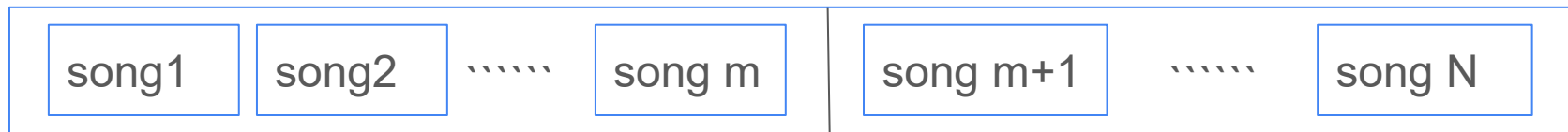
```
interest = 'taylorswift'  
similar_words = vectorizer.model.wv.most_similar(interest, topn=10)  
for word in similar_words:  
    print(word)
```

✓ 0.0s

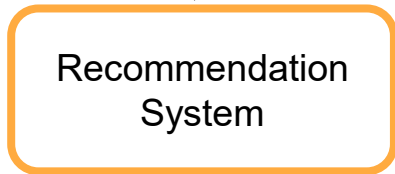
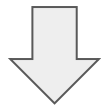
```
('justinbieber', 0.8739178776741028)  
( 'jasonderulo', 0.7950416207313538)  
( 'selenagomezthescene', 0.7633613348007202)  
( 'vanessacarlton', 0.7504104971885681)  
( 'christinaperri', 0.7465274333953857)  
( 'krisallen', 0.7457529902458191)  
( 'lifehouse', 0.7431434988975525)  
( 'shaniatwain', 0.7302798628807068)  
( 'neyo', 0.7079509496688843)  
( 'gleecast', 0.7035359144210815)
```

# Hit@ K

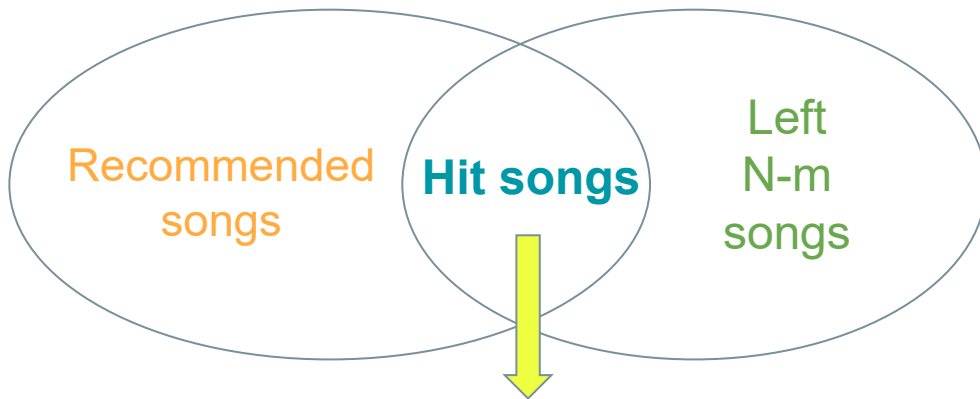
A user's playlist with N songs



Taste vector



Recommend K songs



$$\text{Hit@K} = \text{hit songs} / (\text{N-m})$$

# Test Result

Hit@K for different models

