TREBLE MAKERS PLAYLIST

Search

Home

Library

THE
TREBLE MAKERS!

Jui Nagarkar, Adawn Symonette, Injuu Jyenis, & Jack Deveney

Search

Home

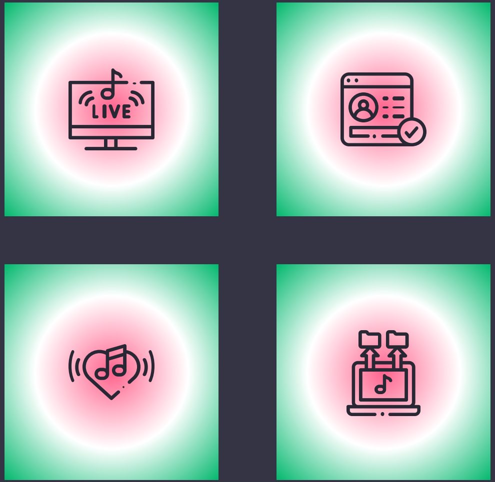
Library

Our Music Playlists

01

Introduction

Introduction



In our capstone project, we worked to build a recommender system using the music dataset. Recommender systems play a vital role in helping users navigate through vast options, enhancing user engagement and satisfaction in various digital platforms. This presentation will explore how we can utilize data from over 92,834 user-artist listening interactions, coupled with tagging and social networking information from 1,892 users, to craft personalized music recommendations.

Search

Home

Library

Our Music Playlists

02

Data Exploration

Tools Used

- Python (PyCharm & Google Colab)
- SQL (DBeaver)

Data Overview

Artists
- ArtistID, Name, URL, PictureURL

Tags
- TagID, TagValue

User_friends
- UserID, friendID

UserArtist
- UserID, ArtistID, Weight

UserTaggedArtist
- UserID, ArtistID, TagID, Day, Month, Year

We started off by cleaning the data to complete a thorough exploration of the data. This is necessary To ensure consistency and facilitate accurate data comparison. The purpose was to:

Correct Data Errors and Inconsistencies

3:15

ABC Name
Sara Tavares
CÃ©line Dion

Standardize Formats

3:20

ABC Name
MALICE MIZER
Diary of Dreams

Check and Remove Duplicates

3:10

ABC Name	123 COUNT(*)
The K	2

Search

Home

Library

Our Music Playlists

03

Collaborative-Based

Search

Home

Library

Our Music Playlists

UserArtistTable

- UserID is a number that represents the user
- ArtistID represents a specific artist
- Weight is how much the user listened to a specific artist
- This is the main table that is used for collaborative based filter

	123 UserID	123 ArtistID	123 Weight
1	2	51	13,883
2	2	52	11,690
3	2	53	11,351
4	2	54	10,300
5	2	55	8,983
6	2	56	6,152
7	2	57	5,955
8	2	58	4,616

Search

Home

Library

Our Music Playlists

To begin the collaborative filtering we used the min-max scaling for each user so each user will have a max scaling weight of 1.

	UserID	ArtistID	scaled_weight
49	2	99	0.0011935073
50	2	100	0
51	3	101	1
52	3	102	0.0455342842

Search

Home

Library

Our Music Playlists

We made a similarity matrix by finding the Euclidean distance between all of the different user's scaled listening weights.

UserID	2	3	4	...	2097	2099	2100
UserID				...			
2	0.000000	2.242183	2.218897	...	2.274518	2.833143	2.907011
3	2.242183	0.000000	1.528234	...	1.543416	2.238771	2.331558
4	2.218897	1.528234	0.000000	...	0.811267	2.309721	2.399755

Search

Home

Library

Our Music Playlists

- We then use that similarity matrix to find the most similar user by find the user that has the minimum value within our chosen user's column.
- We can then use the user that is most similar to find the artist we are recommending by finding their most listened to artists.

```
>>> print(id_name_map)
{55: 'Kylie Minogue', 229: 'The Killers', 289: 'Britney Spears', 292: 'Christina Aguilera', 455: 'Backstreet Boys'}
```

Our Music Playlists

Search

Home

Library

04

Content-Based

Search

Home

Library

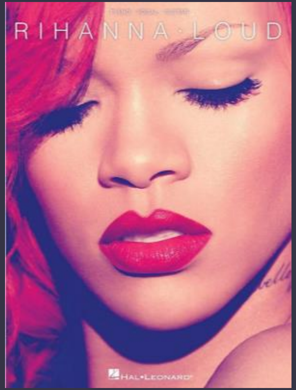
Our Music Playlists

TagsTable

- TagID is a number that represents the tag
- TagValue is posts users made
- Category represents the refined music genres

ArtistsCategoriesTable

- Created Table that has Artist ID, Artist Name, Tag Value & Category for the convenience



For the content-based filtering, we did some data cleaning to emphasize music-related hashtags. This involved eliminating irrelevant hashtags and grouping similar ones.

TagID	TagValue
200	hot
201	guilty pleasures
202	electro pop
203	disco
204	fierce

223	umtalented
224	bad
225	money
226	california
227	comedy
228	catchy

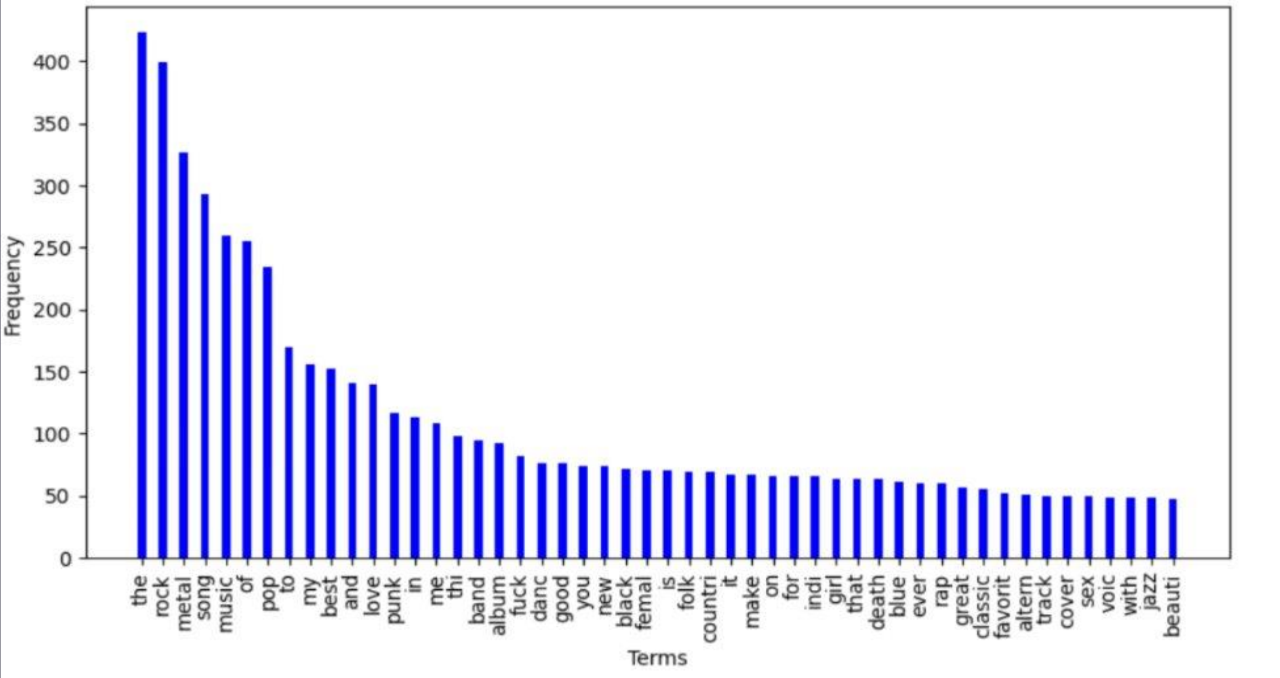
The images are examples of the hashtags in the tag value column that we wanted to clean to give more relevance and refinement to our search.

The list was refined using machine learning techniques. We created a terms versus frequency graph for bigrams and unigrams to find the top 10 values that match in both the graphs and further modify the database.

Followed Pre-Processing Steps:

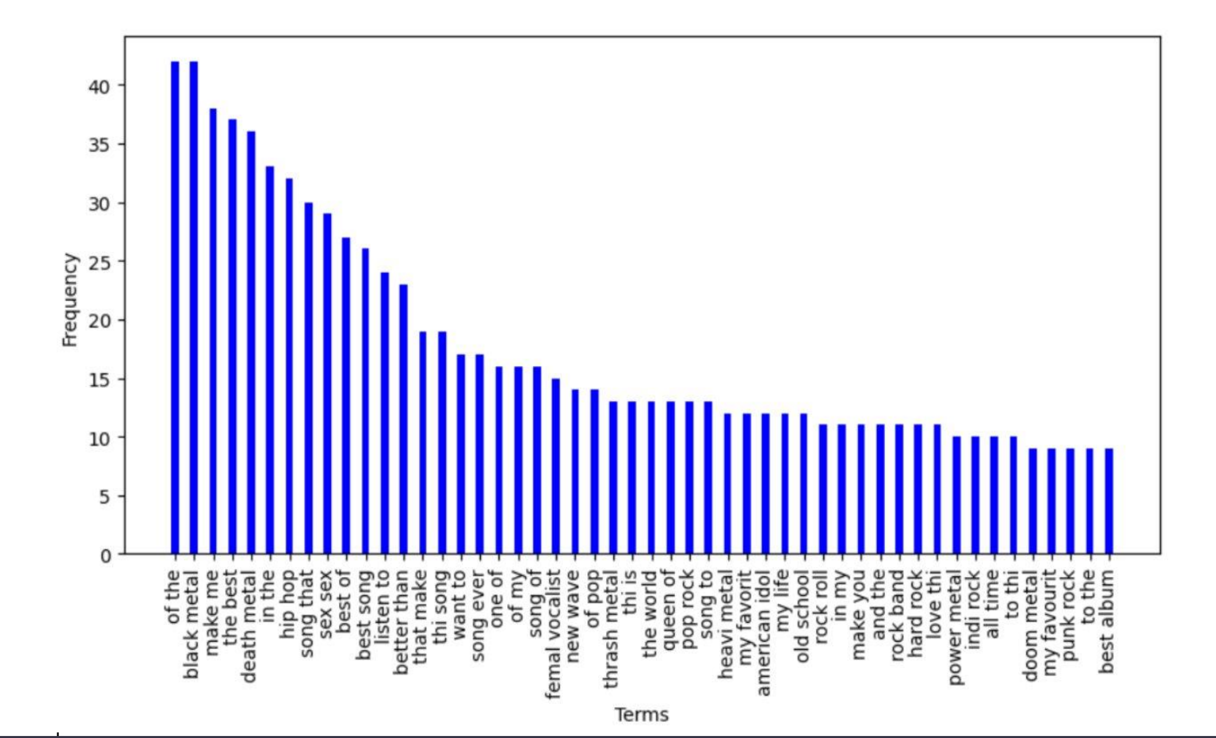
- make everything lowercase
- remove numbers
- remove punctuation
- remove extra spaces
- remove stopwords

UNIGRAMS



Frequency
0
50
100
150
200
250
300
350
400
Terms
the
rock
metal
song
music
of
pop
to
my
best
and
love
punk
in
me
thi
band
album
fuck
danc
good
you
new
black
femal
is
folk
countri
it
make
on
for
indi
girl
that
death
blue
ever
rap
great
classic
favorit
altern
track
cover
sex
voic
with
jazz
beauti

BIGRAMS



Frequency
Terms
0
5
10
15
20
25
30
35
40
of the
black metal
make me
the best
death metal
in the
hip hop
song that
sex sex
best of
best song
listen to
better than
that make
thi song
want to
song ever
one of
of my
song of
femal vocalist
new wave
of pop
thrash metal
thi is
the world
queen of
pop rock
song to
heavi metal
my favorit
american idol
my life
old school
rock roll
in my
make you
and the
rock band
hard rock
love thi
power metal
indi rock
all time
to thi
doom metal
my favourit
punk rock
to the
best album

To help improve the accuracy and relevance of our music recommendations system we grouped the hashtags into categories. A new column was created, and we identified key terms that are indicative of the specific genre.

3	goth rock	Rock
4	black metal	Metal
5	death metal	Metal
6	industrial metal	Metal
7	gothic metal	Metal
8	terror ebm	Electronic
9	electro-industrial	Electronic

Other examples of these categories were 'Hip Hop and R&B', 'Pop', 'Classical', 'Decades' and 'Foreign Music' to name a few.

Then, we created a table called 'ArtistsCategories' where we have Artist ID, Artist's Name, Tag Value related to them and which Category they belong to. The following is an example of Coldplay:

	123 ID	ABC Name	ABC TagValue	ABC Category
1447	65	COLDPLAY	rock	Rock
1448	65	COLDPLAY	alternative rock	Rock
1449	65	COLDPLAY	alternative	Other
1450	65	COLDPLAY	indie	Rock
1451	65	COLDPLAY	indie rock	Rock

We tried filtering by Category, User ID and Artist Name --> Unfortunately, we got thousands of results for the recommendations

- One hot encoding
- Similarity Matrix
- Recommending additional artists for specific users who listened to an artist "X"

Part of our huge one hot encoding table:

Name	ID	is_Metal	is_Electronic	is_Rock
ALMAMEGRETTA	18,432	0	0	1
ALMORA	9,063	1	0	0
ALOHA	12,202	0	0	1
ALOHA FROM HELL	2,435	0	0	1
ALPH LYLA	7,901	0	0	0
ALPHA	9,355	0	1	0
ALPHA BLONDY	16,103	0	0	0
ALPHA BOY SCHOOL	7,538	0	0	0
ALPHA QUADRANT	1,250	0	1	0

Created a table specifically for the artist "RIHANNA":

ABC Name	123 ID	123 is_Metal	123 is_Electronic	123 is_Rock	123 is_HipHop_and_RnB
RIHANNA	288	1	1	1	1

Part of our similarity score table:

Name	ID	SimilarityScore
BEFORE THE DAWN	10,956	0
BEFORE THEIR EYES	2,618	1
BEFORU	9,103	1
BEFOUR	9,232	1
BEHEMOTH	12	1
BEHERIT	8,341	0
BEHIND CRIMSON EYES	15,610	1
BEHOLD... THE ARCTOPUS	9,453	1
BEIRUT	196	2
BEKKI WILLIAMS	12,744	1
BEL CANTO	11,897	4
BELA B.	12,538	1

Search

Home

Library

Our Music Playlists

We then chose a specific User ID to generate recommendations for Rihanna:

ABC Name	123 SimilarityScore
AVRIL LAVIGNE	7
CORPORE	7
ELVIS PRESLEY	7
LADY GAGA	7
MADONNA	7

Search

Home

Library

Our Music Playlists

05

Hybrid Filtering

Checked recommendation for a specific user:

Based on what similar users have listened to:

Jonas Brothers

Chris Brown

Justin Bieber

Selena Gomez
and the Scene

Bruno Mars

COLLABORATIVE

Checked recommendation for a specific user:

Since you listened to this artist, here are 5 others we think you would like!

Avril Lavigne

Corpore

Elvis Presley

Madonna

MONO

CONTENT

Conclusion

Our GOAL as The Treble Makers is to make the BEST hybrid music recommendation system

Clean and preprocess the data to make sure we can use it

Collaborative Based Filtering

- Min-Max scaled the listening weights for each user
- Created similarity matrix
- Found recommendations from similarity matrix

Content-Based Filtering

- Created categories for each tag
- Hot-encoded each artist to each category
- Used dot product to find most similar artist to give recommendations

Used both filtering methods to create hybrid recommendations

Search

Home

Library

Our Music Playlists

"Thank you for listening!"

Do you have any questions?

—The Treble Makers