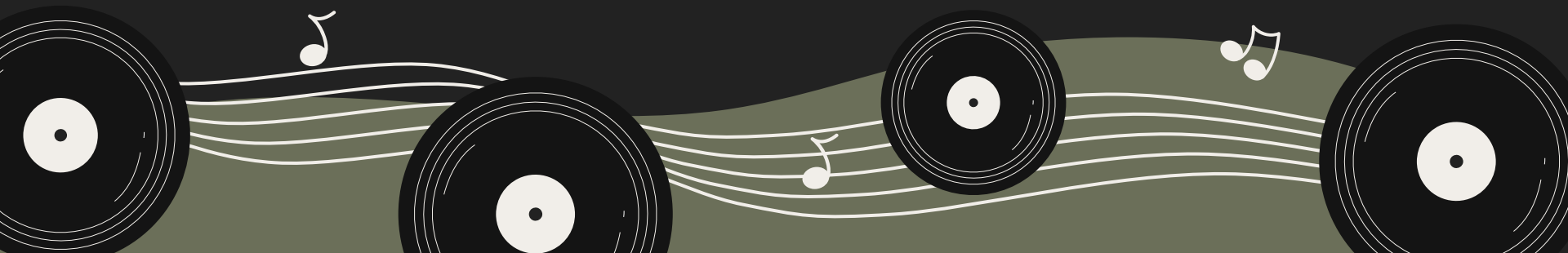
Melodify

Music Similarity through Embedding and Metadata

Gautham, Saumya, Pradeep, Yukta, Parikha



Introduction

Problem Statement:

- **Modern systems** struggle to identify similar songs due to a lack of understanding of a song's content and emotional feel. Other important limitation is cold-start problem and same artist song recommendation irrespective of song understanding.
- **Traditional methods** rely on surface features like tempo and key, missing deeper meanings.
- **Goal:** Embed songs into a vector space using audio content and metadata to capture rich features.

Motivation:

- Music similarity is complex, involving lyrics, vocals, and instruments, beyond simple features.
- Traditional systems focus on user behavior and ignore intrinsic song characteristics.
- A better method should integrate audio, metadata, and emotion to capture a song's true essence.

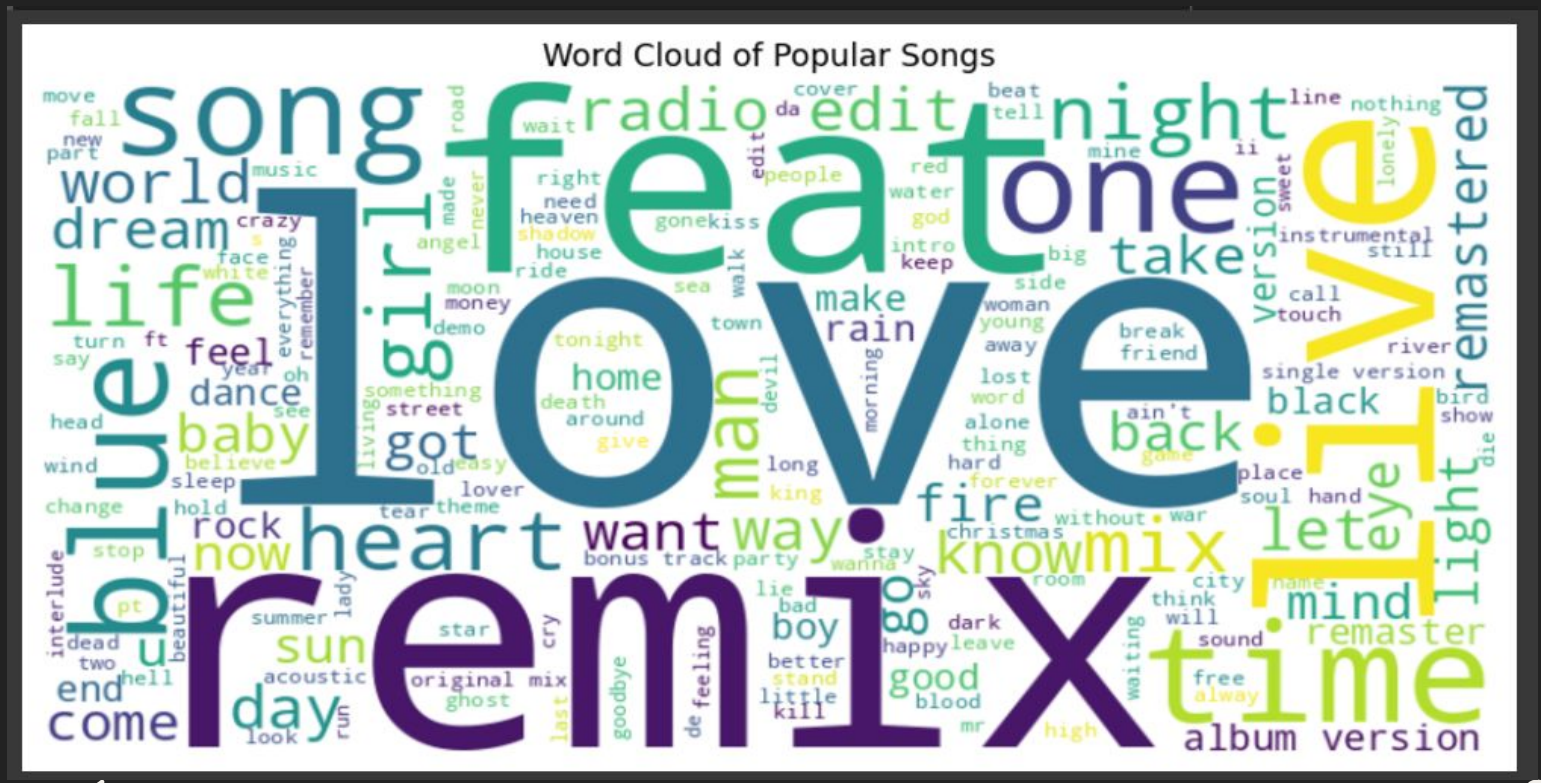
Data - Source and Preparation

Dataset	Key Characteristics	Pre Processing
Million Song Dataset	• Comprehensive dataset with audio features of over 1 Million Songs • Subset of 17,636 songs selected for detailed analysis	Non Numeric columns and low variance columns are removed to enhance feature relevance.
Last FM	• Rich metadata about artist, album and user interactions • Subset of 28,743 songs utilized for training ML Models	Focused on top 40 popular genre tags to create a streamlined subset of album, artist and user data.

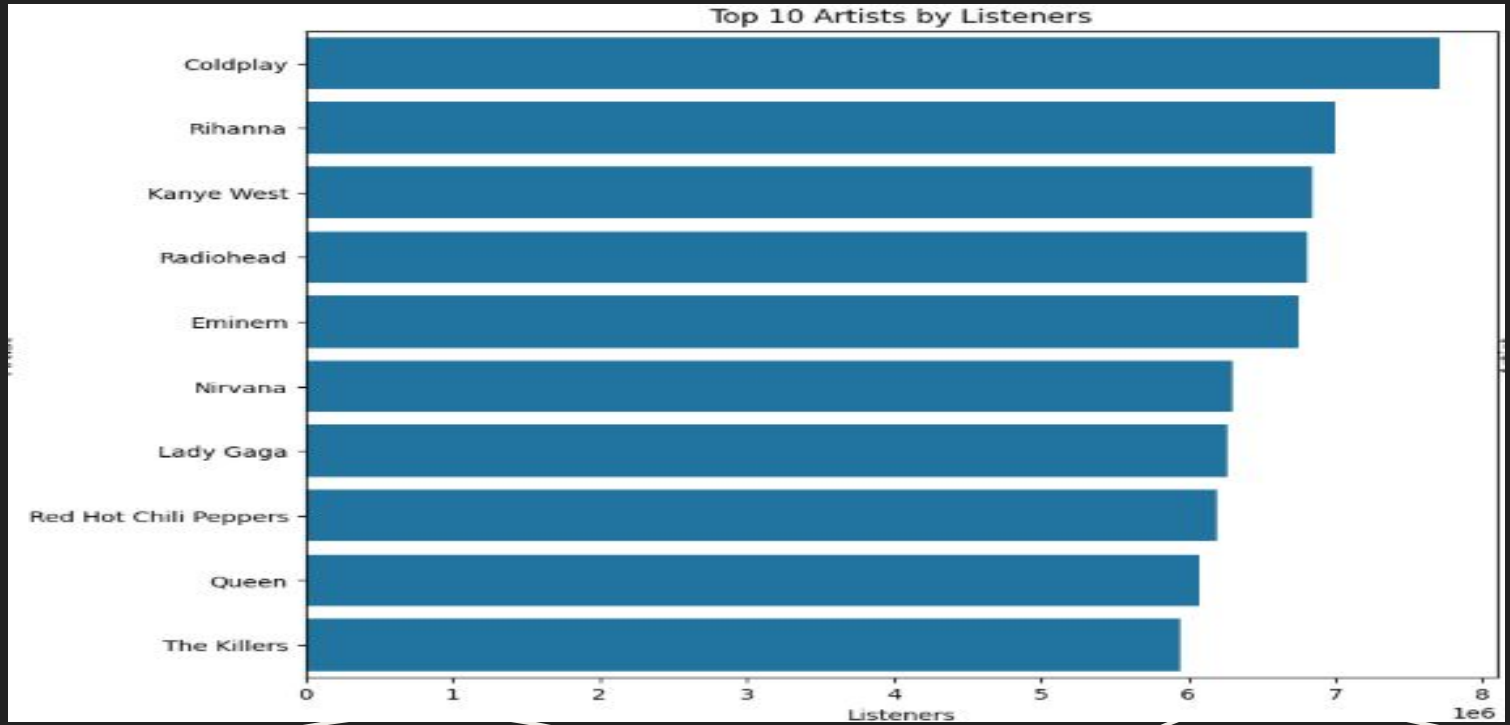
Note: For testing, Lowered similarity threshold to 0.3 for low play count tracks, ensure fairness

Data Visualization

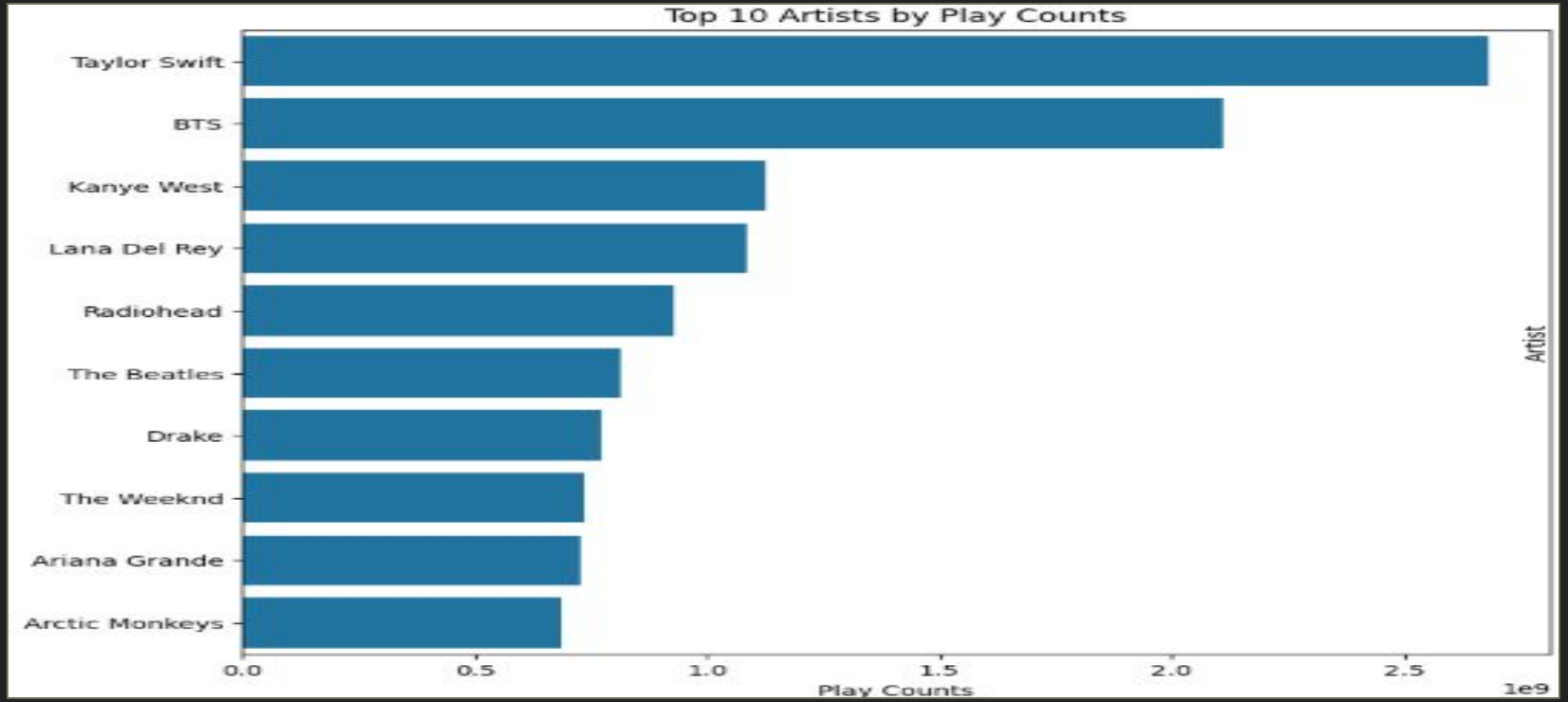
Data Visualization



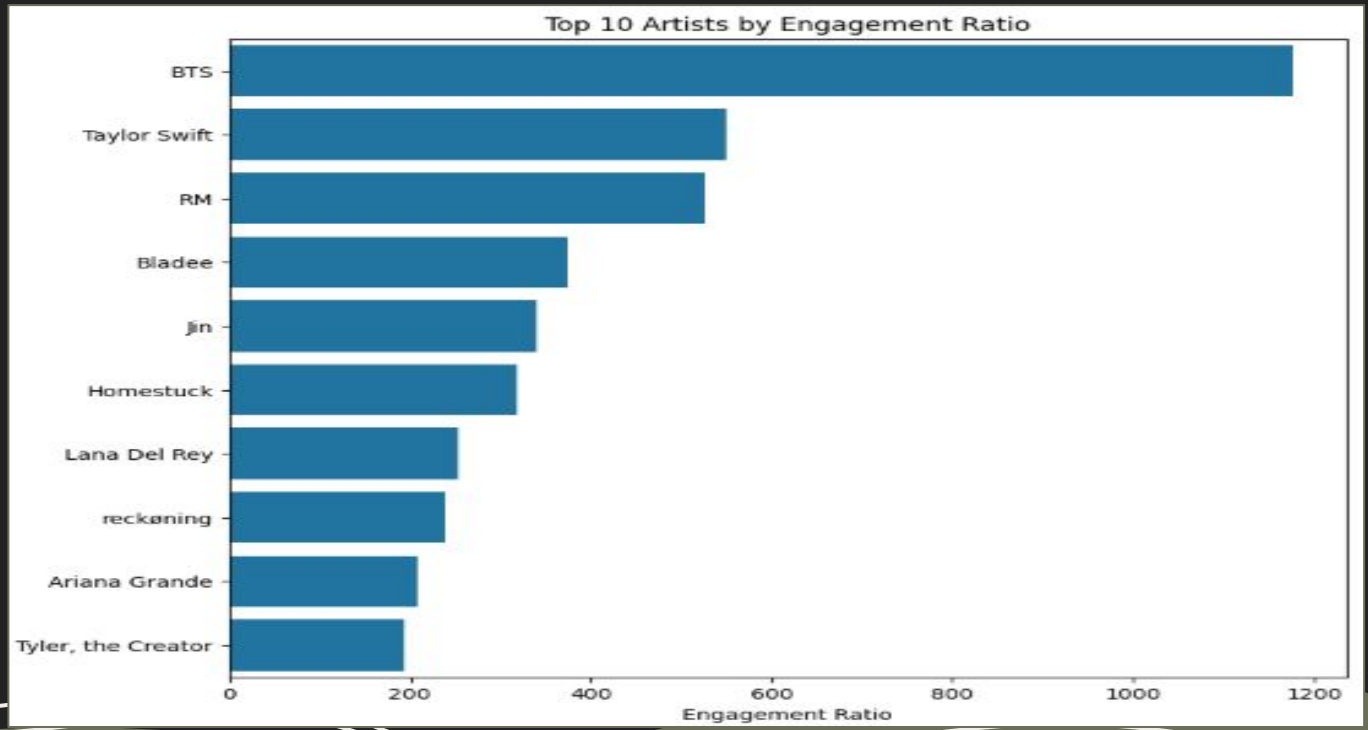
Data Visualization



Data Visualization



Data Visualization



Methodologies

Audio and Tag Embedding

- Audio - Numerical columns are padded to uniform dimensions and passed as an NxD matrix to a CNN architecture, outputting a fixed-size embedding vector of 50 dimensions.
- Tag - Tags from the Last.fm Dataset are processed using Word2Vec to generate tag embeddings.

Graph Embedding

- Heterogeneous Graph Construction on Artists, Albums, Tracks, and Tags from the Last.fm Dataset.
- Node2Vec Algorithm: Predicts embeddings for songs based on relationships between artists, albums, and tags.

XgBoost

- Converted summary, songs, and album name into vector using Word2Vec
- Do inverse transformation on listeners count and play count so that artist with lesser value get more weightage
- Converted tags into embedding, using pre trained Sentence transformer (SBERT).
- Performed agglomerative clustering on tags to cluster them into 10 groups
- Assigned a precedence order according to each clusters frequency

Methodologies

RNN

Audio Embedding
More importance was given to important features by given weights to them

Dimensionality reduction and **normalizations** were performed on numbered features

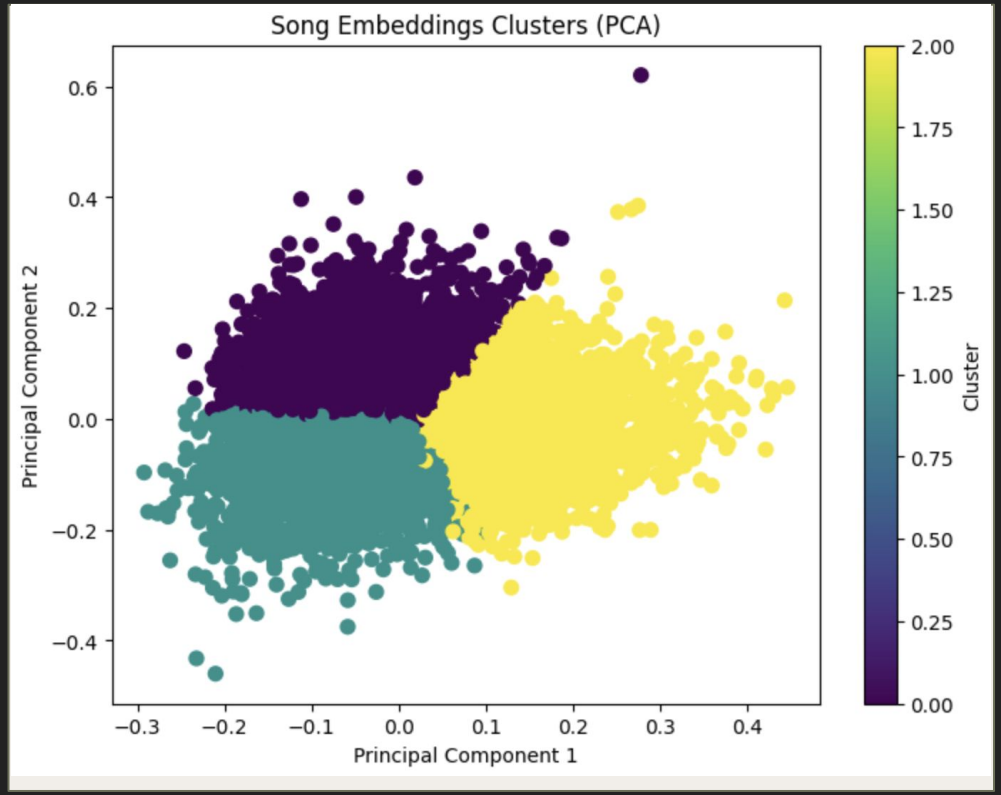
RNN was trained to recommend 5 similar songs to current song

Random Forest

Million Song Subset

- **Nodes:** Artists, Albums, Tracks, and Tags from the Last.fm Dataset.
- **Node2Vec Algorithm:** Predicts embeddings for songs based on relationships between artists, albums, and tags.

Results (Audio Embedding)




```
Cluster 0:
Loud And Clear
Boss-Eyed Whelk
Summer Summer
Separate Ways
Body
Freiwild
El Gatito
Just A Phase
What If I Kissed You Right Now?
Bewitched_ Bothered & Bewildered


Cluster 1:
Baby My Love
Focus
Chim Chim Cheree
Girls Against Boys (LP Version)
Saturday Night Special
The Three-Dimensional Shadow
I Wanna Sing (Autobiography Album Version)
Dedicado
Where in the World
Una dolce melodia


Cluster 2:
I've Won (Introduction - Speaking)
The Big Stall
Container
Money Blues
Style
High On the Mountain
We Will Glorify (Key-B-Db-Premiere Performance Plus)
Poaki
Poppin' Them Thangs
Think About Me
```

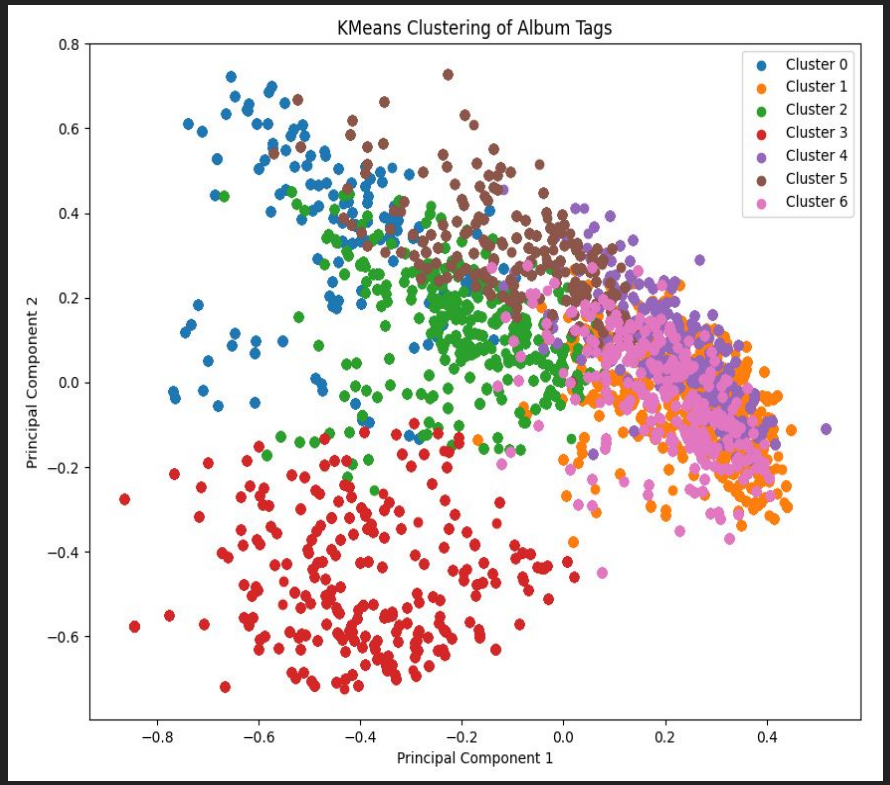
- **Cluster 0:** Classical, R&B, French chanson, and hip-hop
- **Cluster 1:** Lo Fi, Non english, Global
- **Cluster 2:** Indie Fusion, folk, rock

Results (Audio Embedding)

Poppin them Thangs - G Unit
161 M Views
Genre - Rap

Money Blues - S.A.M.
156 K Views
Genre - House

Results (Tag Embedding)




```
--------------------
Cluster 4:
13 Pieces, Op. 76: No. 11 Linnaea (Twinflower of
Next Year, Baby
Get Some Rest
Nocturne
Swan Lake, Op.20, Act III: 20. Hungarian Dance:
You Love Jazz Now
Sirènes de la Fête
These Foolish Things (Remind Me Of You)
Si tu n'étais pas là
Kiss Me On My Neck (Hesi)
--------------------
Cluster 5:
カンタンナコト
Cut Your Hair
Bloodhound
Blow Ya Mind
(i would have followed you)
Out Of My Dreams (bonus track)
I Can Play The Piano
The Big House (feat. Frances Quinlan)
How'd You Like That
Yes I'm A Mess
--------------------
Cluster 6:
Break & Enter (2005 Live Edit)
Regeneration (2009 Remaster)
I Love My Harbour
Not Giving In (feat. John Newman & Alex Clare)
Tour De France (2009 Remaster)
Delilah (Pull Me Out Of This)
Tyranny
What If I Go?
J'adore Hardcore [Explicit]
is to create
```

```
Cluster 0:
Out Of My Face
Romantic
Romeo & Rebecca
Like the Angel
Nitro (Youth Energy)
Fahrradsattel
I'm Gonna Be Your God
Queen Wasp - Fox Studio 1983
Portrait of Authority
I Fought the Law
--------------------
Cluster 1:
Wretched Shades
On My Own (feat. Kid Cudi)
Você fez Merda
I Know What You Are
Can’t Punk Me
No Exchange
Scheamin'
Long Was the Year
Sunday Hotel
Just Lose It
```

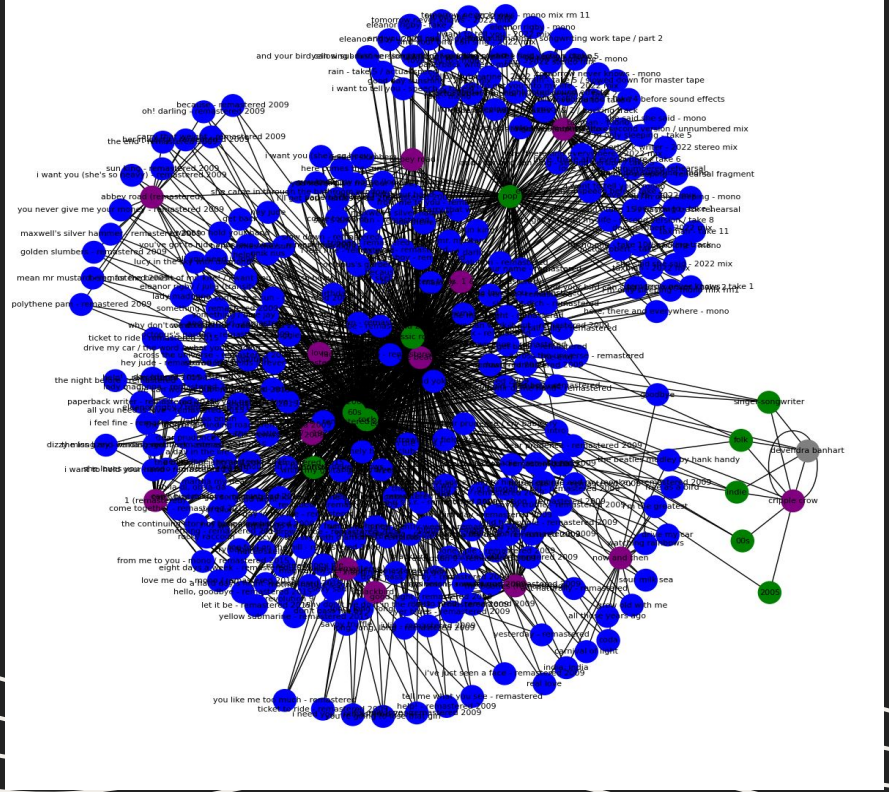
Results (Tag Embedding)

Yes I’m a Mess - AJR
8M Views
Genre - Pop & Rock

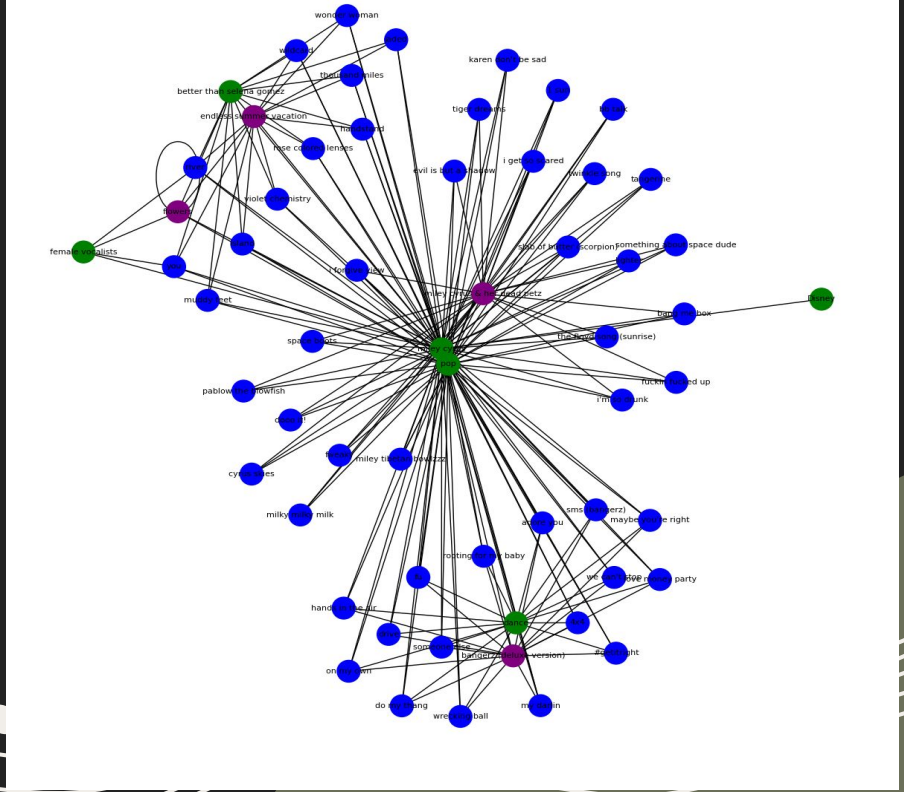
Cut your Hair - Pavement
3.6 M Views
Genre - Pop & Rock

Results (Graph Embedding)

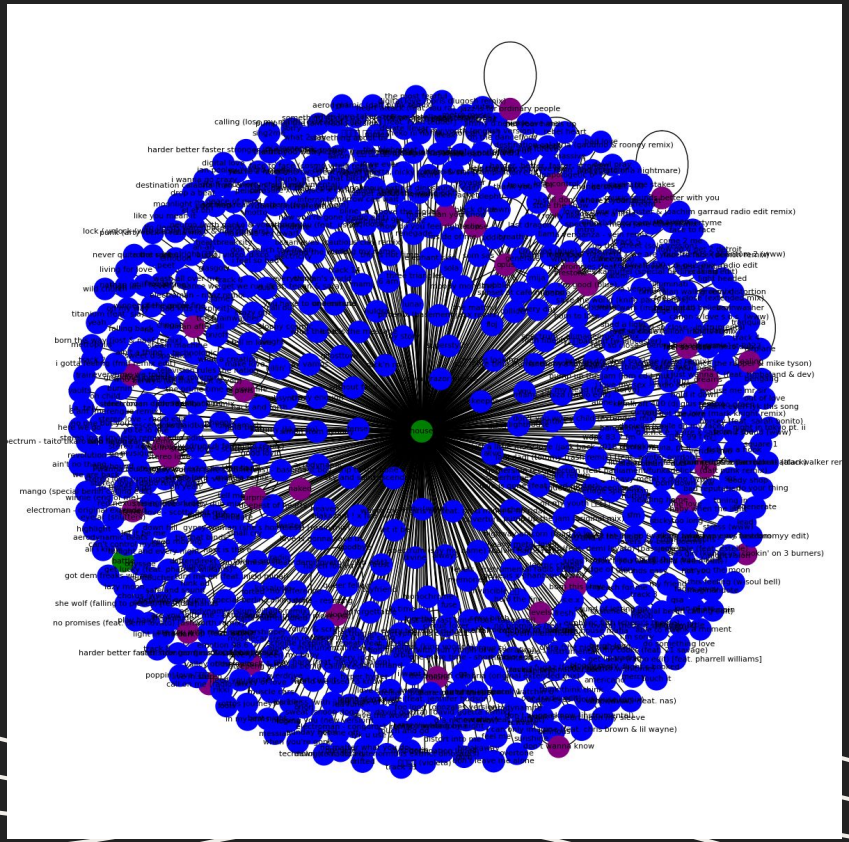
The Beatles - subgraph



Miley Cyrus - Subgraph



Results (Graph Embedding)



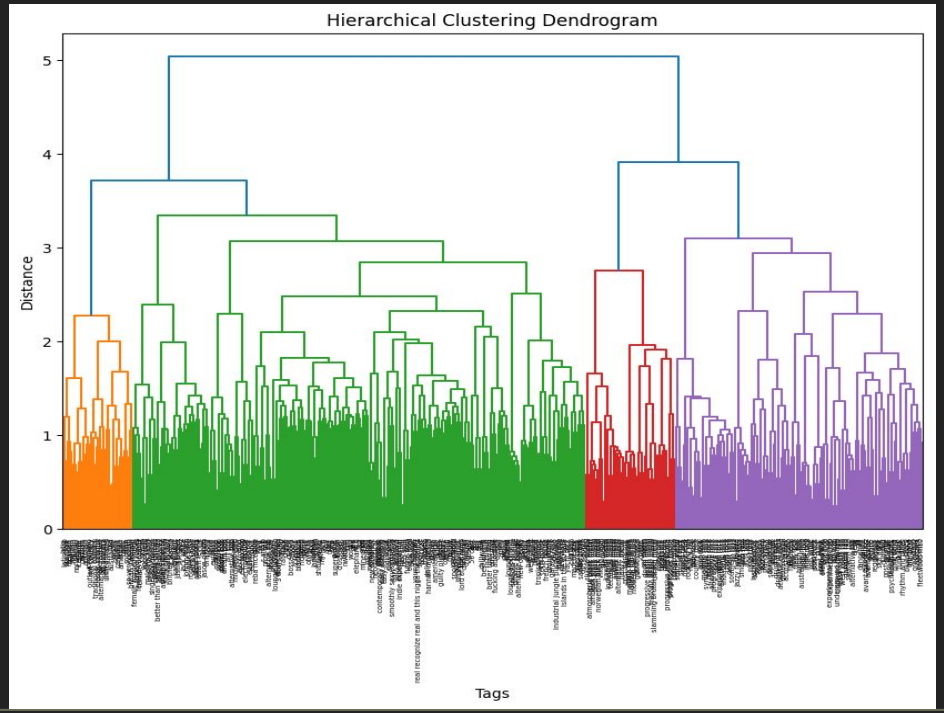
We see that tag nodes have very high centrality in the graph,

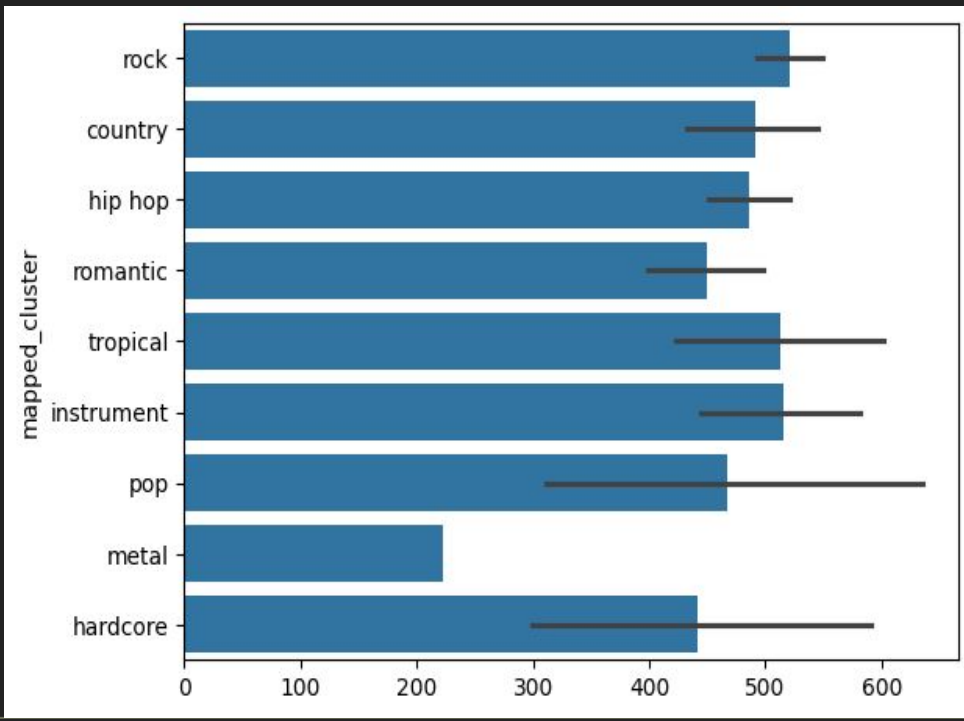
as they are common nodes between all tracks, albums and artists

and provide a meaningful way to model the Node2Vec for a track

Subgraph - network for the tag “house”

Results (XGBoost)






```
dict_keys(['hip hop', 'romantic', 'tropical', 'instrument', 'singer', 'rock', 'pop', 'country', 'hardcore', 'metal'])
```

```
Artist Recommended for tag pop - ['Of Monsters and Men', 'Kings of Leon', 'Johnny Cash', 'Sade', 'a-ha',
'OutKast', 'Ed Sheeran', 'Shakira', 'Duffy', 'Edward Sharpe & The Magnetic Zeros']
```

Results (RNN)

title	I Didn't Mean To	Soul Deep	Amor De Cabaret	Something Girls	Face the Ashes	The Moon And I (Ordinary Day Album Version)	Keepin It Real (Skit)	Drop of Rain	Pink World	Insatiable (Instrumental Version)	...	20.000 Seconds	Pastel
I Didn't Mean To	1.000000	0.067220	0.999999	-0.917843	0.646928	0.868617	0.341102	0.988153	0.710170	0.010337	...	0.593782	0.682947
Soul Deep	0.067220	1.000000	0.068583	-0.457742	0.804313	0.552751	-0.914971	-0.086699	-0.654701	0.998380	...	-0.762892	0.774724
Amor De Cabaret	0.999999	0.068583	1.000000	-0.918385	0.647969	0.869293	0.339818	0.987943	0.709207	0.011703	...	0.592682	0.683944
Something Girls	-0.917843	-0.457742	-0.918385	1.000000	-0.896468	-0.993933	0.060058	-0.846051	-0.372365	-0.406409	...	-0.225608	-0.916792
Face the Ashes	0.646928	0.804313	0.647969	-0.896468	1.000000	0.939765	-0.496149	0.522236	-0.077430	0.769198	...	-0.229434	0.998837
The Moon And I (Ordinary Day Album Version)	0.868617	0.552751	0.869293	-0.993933	0.939765	1.000000	-0.169480	0.782285	0.268030	0.504436	...	0.117090	0.955154
Keepin It Real (Skit)	0.341102	-0.914971	0.339818	0.060058	-0.496149	-0.169480	1.000000	0.481327	0.904048	-0.936450	...	0.958910	-0.453705
Drop of Rain	0.988153	-0.086699	0.987943	-0.846051	0.522236	0.782285	0.481327	1.000000	0.809804	-0.143247	...	0.710233	0.562751
Pink World	0.710170	-0.654701	0.709207	-0.372365	-0.077430	0.268030	0.904048	0.809804	1.000000	-0.696652	...	0.988167	-0.029264

Results (RNN)

Neva Play - Megan Thee Stallion
28 M Views

Superconfidential - C.L.P.
642 Views

Note

Since, the concept of similarity in music is a highly subjective,

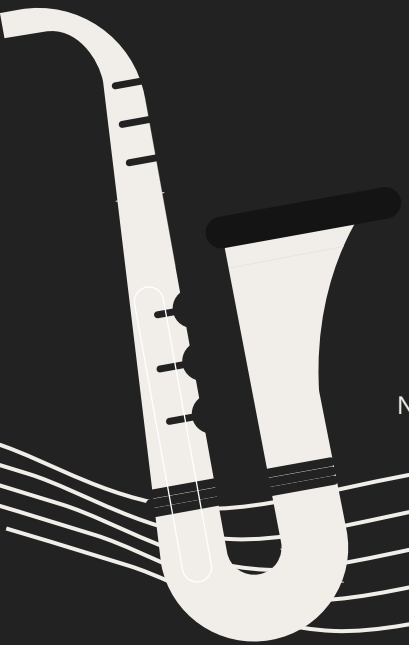
While evaluating our models for accuracy, we rely on pre existing similar song sets from the LastFM dataset, which in itself have drawbacks of being biased towards popular artists and tracks.

So, the accuracies presented in the next slide are just for a baseline evaluation.

Accuracy

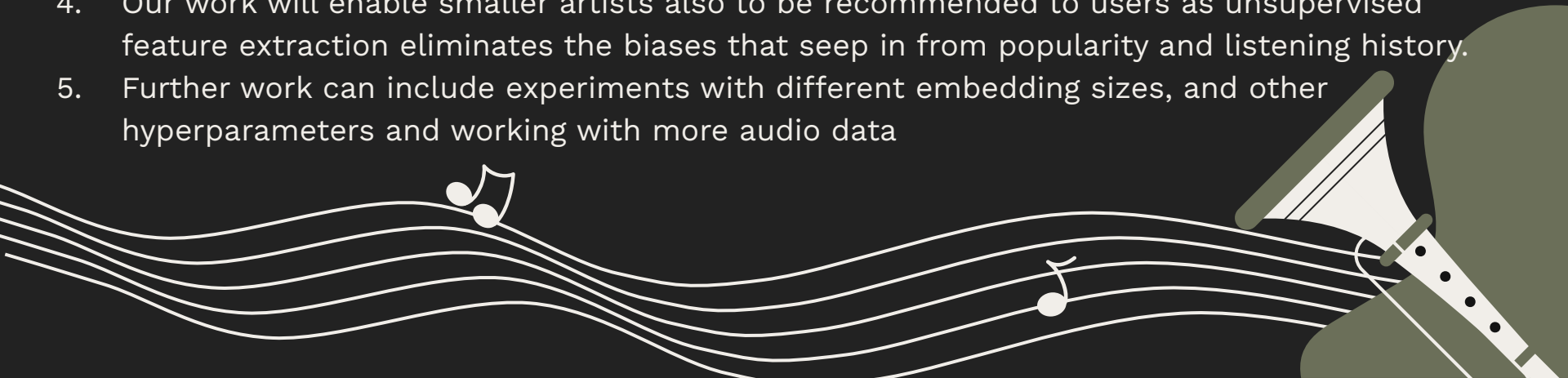
25%	Graph Node2Vec
30%	Audio Embedding
53%	Tag Embedding
52%	XGBoost

Note: In RNN and RF, aim is to recommend songs, no accuracy parameters, relying solely on user feedback



Conclusion and Future Scope

1. We believe that there is real value in extracting features from songs similar to word embeddings in NLP, as an unsupervised task.
2. Modern systems (such as the spotify method) use predefined tags which work well so far, but are limited to the capabilities of human tagging.
3. Our approach proves that even with a relatively smaller dataset, it is possible to perform unsupervised feature extraction from songs and can be used to improve recommender systems.
4. Our work will enable smaller artists also to be recommended to users as unsupervised feature extraction eliminates the biases that seep in from popularity and listening history.
5. Further work can include experiments with different embedding sizes, and other hyperparameters and working with more audio data



Thank You

Melodifiers:

Gautham

Saumya

Yukta

Pradeep

Parikha