Music Recommendation System

Team

Simran Goindani
Charvi Madan
Sumukh Vasisht Shankar

Course Instructor

Prof. Yehoshua Roi

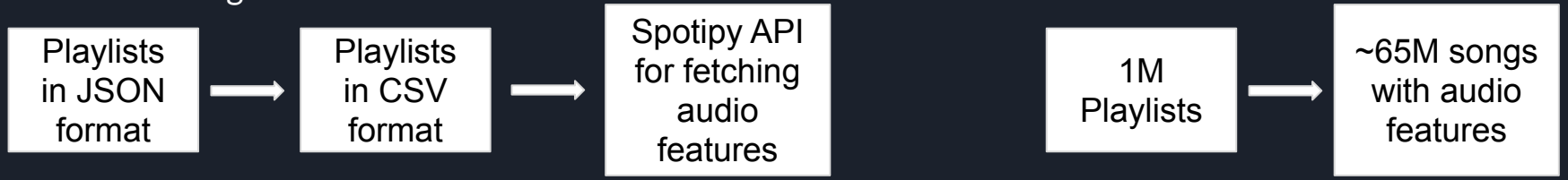
Motivation

- Develop a recommender system to recommend songs given a playlist.
- This will help reduce the human efforts
- Help in engaging the audience more with a platform
- Input data - songs or a playlist
- Output - songs

Technical Problem Formulation

Data Source: Spotify Million Playlist Challenge (https://research.atspotify.com/2020/09/the-million-playlist-dataset-remastered/)
Data consists of data and metadata related to 1M playlists. This includes song names, artist for each song, song duration, track ID, number of songs in the playlist, artist ID, etc., in a json format.

Data Processing:



Audio Features include key, tempo, danceability score, acousticness, loudness, speechiness, etc.

Challenges faced with the dataset

- Working with the whole dataset was computationally too intensive!
- Restrictions on the number of Spotipy API calls that can be made in a minute. This hindered our data collection process a lot even though we could implement parallel processing.
- Clustering and DImensionality Reduction Algorithms took forever to run even on 10% of the data (6.5M songs - 100k Playlists)

Solution?

- We choose to go ahead with a small sample size of ~15k playlists (150k songs).
- Found a much more articulated dataset on Kaggle with the same audio features we wanted.

Clustering Algorithms

KMeans

DBSCAN

Spectral Clustering

Gaussian Mixture Model

Agglomerative Clustering

Input Features

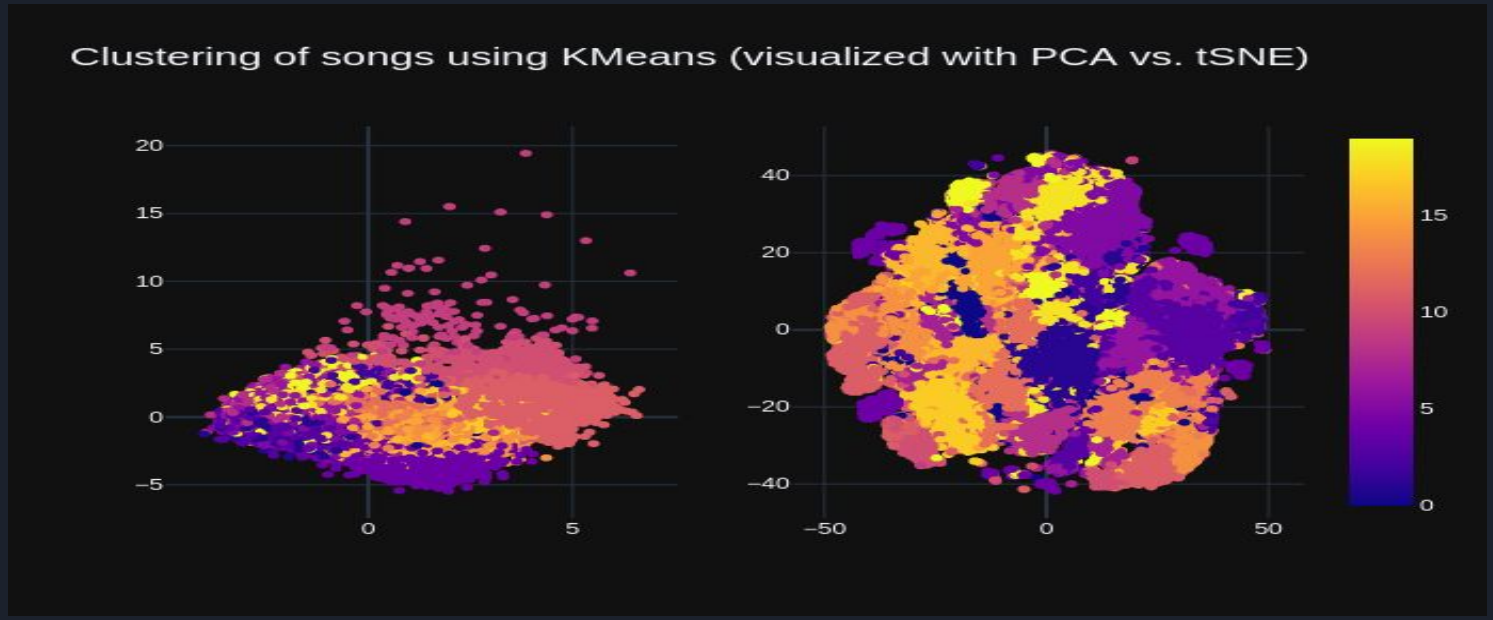
Valence , Year, Accousticness, Danceability

Duration in ms, Energy, Explicit,

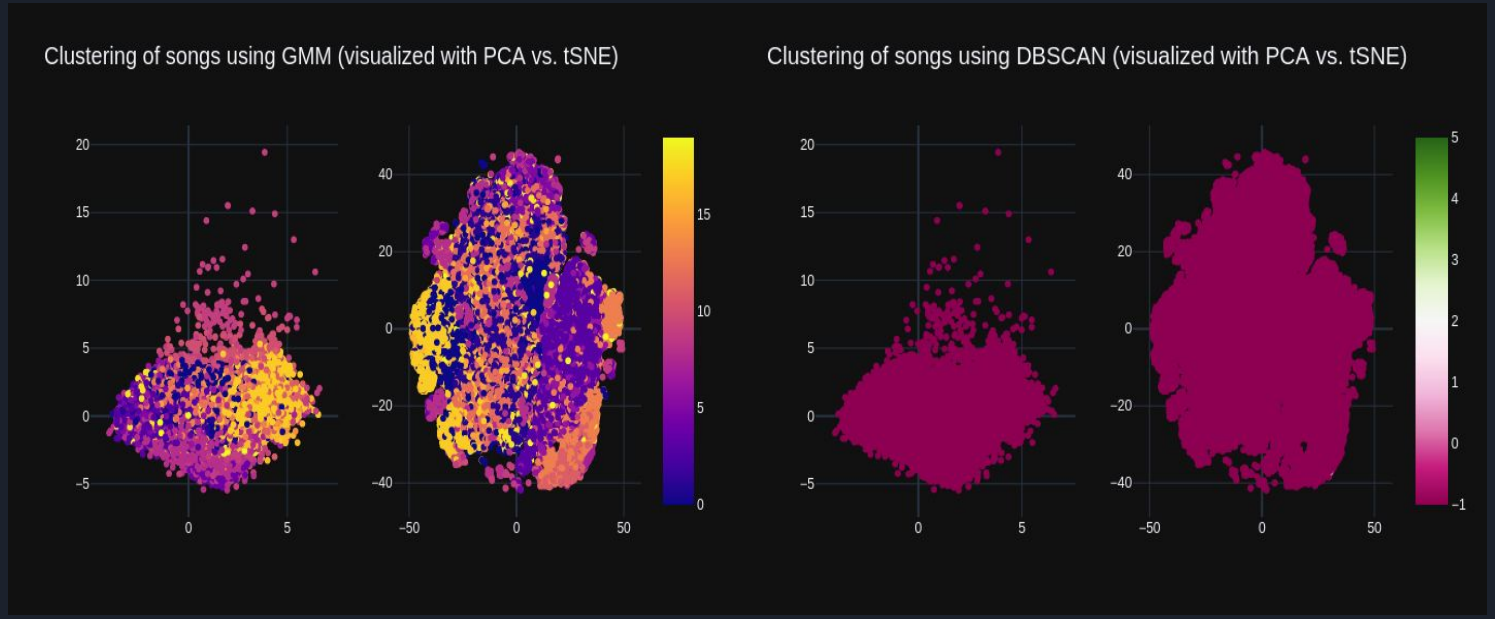
Instumentalness, Key, Liveliness, Loudness,

Mode, Popularity, Speechiness, Tempo

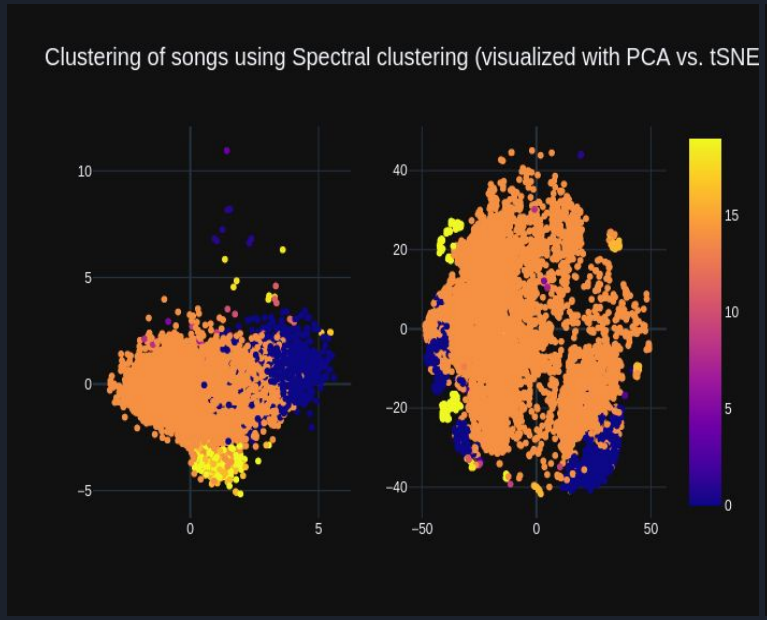
Results

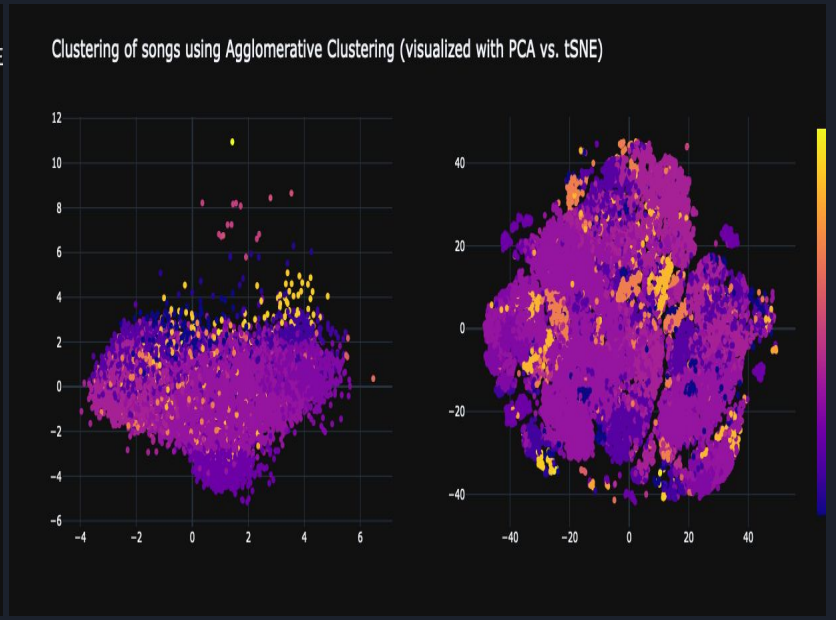

Clustering of songs using KMeans (visualized with PCA vs. tSNE)
20
15
10
5
0
−5
0
5
40
20
0
−20
−40
−50
0
50
15
10
5
0

Results

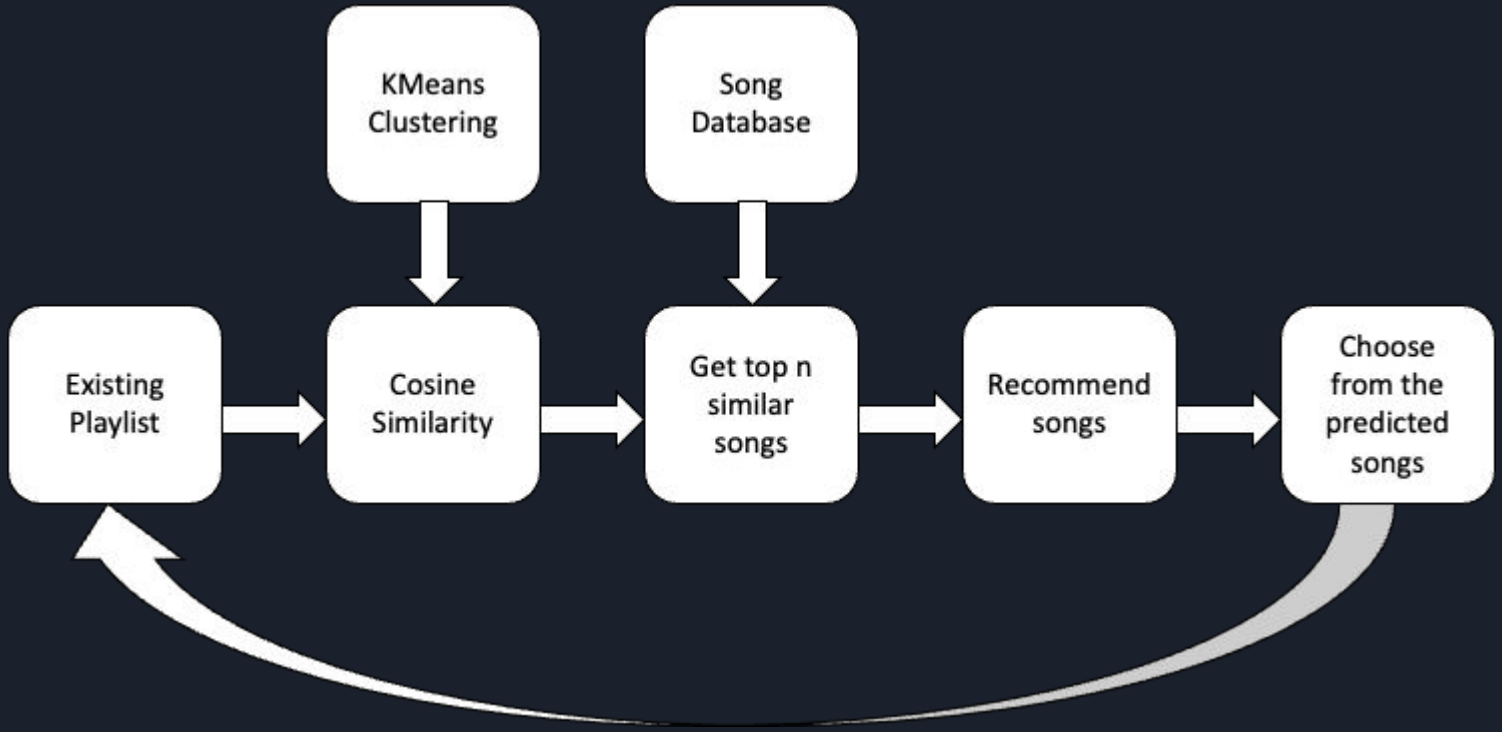

Clustering of songs using GMM (visualized with PCA vs. tSNE)
Clustering of songs using DBSCAN (visualized with PCA vs. tSNE)

Results


Clustering of songs using Spectral clustering (visualized with PCA vs. tSNE
10
5
0
−5
0
5
40
20
0
−20
−40
−50
0
50
15
10
5
0


Clustering of songs using Agglomerative Clustering (visualized with PCA vs. tSNE)
12
10
8
6
4
2
0
−2
−4
−6
−4
−2
0
2
4
6
40
20
0
−20
−40
−40
−20
0
20
40

Recommender System



RS Results

Tracks added to playlist

Sl. No.	Song Name	Year
1	believer	2021
2	it's a trip!	2017
3	wow.	2019
4	tap in	2020
5	for free (feat. drake)	2016

Recommended Tracks

Sl. No.	Song Name	Year	Artist(s)	
1	Address It	2020	['LPB Poody']	Add
2	Satisfacción	2020	['Arcangel', 'Myke Towers']	Add
3	That's How You Feel	2018	['Drake']	Add

Future Work

Collaborative filtering requires no domain knowledge for extracting features for the model. Also users can explore new interests. This approach can be used by adding certain extensions-:

1. Including user profile data into the system
2. Make the algorithms scalable by using clustering approach before hand

Content based recommendation system using clustering techniques produced cohesive and well separated clusters and as a result, returned good recommendations to user instance. This can be further enhanced by -:

1. Incorporating the use of collaborative filtering in the current model such that clusters are formed based on similarities between users in terms of their ratings and reviews provided.