MSBX 5420-003 - April 15th, 2022

Personal Streaming Database

Project Presentation

Presented by: Dock Dock Goose (Group 11)

- Joey Athanasiou
- Marie Reine Axalan
- Marcos A. Fontes
- Jordyn Gerstle-Goodman
- Talia Zalesne

TABLE OF CONTENTS

I. Introduction

Problem to solve:
Analyze and develop a recommendation model using preliminary Spotify Data capturing feature based recommendations opposed to the current recommendation algorithm, enhancing general user experience.

Call to Action:
Most Spotify users would argue that although Spotify has capitalized on the industry issue that is connecting people with new music, recommendations are never generated off what the consumer wants but rather what is popular.

Dataset:
'Music Recommendation System using Spotify Dataset' From Kaggle.

- 42305 songs
- 15 genres
- 12 Song Attributes

The data

Outlining to develop our model on variables including Genres, Artist, and System interaction we are confident in our ability to transform and translate our project into every user's favorite song.

Size:

- 12.4 MB
- 169,220 Objects
- 22 columns

Genres:

Trap, Techno, Techhouse, Trance, Psytrance, Dark Trap, Hardstyle, Underground Rap, Trap Metal, Emo, Rap, RnB, Pop, Hiphop and more...

Data Dictionary

Danceability	Danceability is measured using a mixture of song features such as beat strength, tempo stability, and overall tempo.
Energy	Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy.
Key	A system of functionally related chords deriving from the major and minor scales, with a central note, called the tonic (or keynote).
Loudness	Loudness is a way to measure audio levels based on the way humans perceive sound.
Mode	Any of several ways of ordering the notes of a scale according to the intervals they form with the tonic, thus providing a theoretical framework for the melody.
Speechiness	Detects the presence of spoken words in a track". If the speechiness of a song is above 0.66, it is probably made of spoken words, a score between 0.33 and 0.66 is a song that may contain both music and words, and a score below 0.33 means the song does not have any speech

Data Dictionary Cont.

Acousticness	The closer it is to 1.0, the more instrumental the song is. Acousticness: This value describes how acoustic a song is. A score of 1.0 means the song is most likely to be an acoustic one.
Instrumentalness	This value represents the amount of vocals in the song. The closer it is to 1.0, the more instrumental the song is.
Liveness	This value describes the probability that the song was recorded with a live audience.
Valence	A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).

Data Dictionary Cont.

Tempo	How fast or slow a piece of music is performed.
Time_signature	Indicate how many beats are in each measure of a piece of music, as well as which note value is counted as a beat.
Genre	A conventional category that identifies some pieces of music as belonging to a shared tradition or set of conventions. It is to be distinguished from musical form and musical style, although in practice these terms are sometimes used interchangeably.

Call to action

Abstract: Spotify Algorithm Listening Habits

Known to be composed of music that the algorithm **thinks** that user is likely to enjoy, based on their particular Spotify listening habits. These habits are shared to include:

- Any artists/albums/tracks a user likes
- any artists/albums/tracks a user shares
- any tracks a user saves to their own playlists

What users desire…

Looking further into the neurology that comes with this advancement in user demand, our team found across multiple studies that engaging the brain at its highest extent (or in other words 'exercising your brain') the personality tends to leave people satisfied and craving more. When looking at the abstract of music, genres typically limit the neurological stimulation you receive from listening to music.

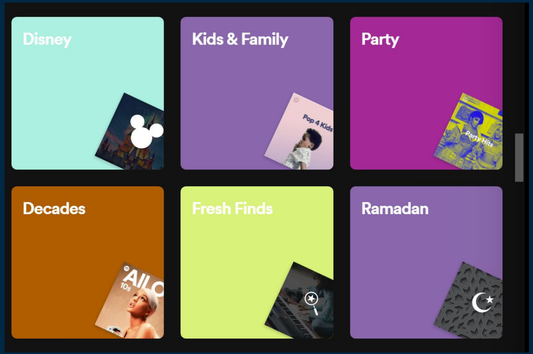
….Or in other words **'Catching a Vibe'**

II. Use Case

Motivation and Background for the business problem.

By analyzing song attributes and clustering them together, we believe we can create the perfect playlists for a music listener. More specifically by generating playlists that are focused on the compilation of the songs presented through different variables that stimulate the brain in different ways (Tone, Tempo, Volume ect).

Spotify playlists are curated by a group playlist editors who review song pitches by artists and choose what songs to put in their playlists. It is impossible for every song that is pitched to be listened to which puts some artists at a disadvantage when it comes to human bias.



What is a vibe? Can the algorithm catch one?

User playlists are curated for a variety of reasons, only a few of these reasons require the genres of the songs. These reasons include:

- General feeling that user is trying to convey
- Playlists for a specific event or gathering
- Keeping track of songs that user enjoys
- Songs with specific BPM for workouts

The overall benefits of this model include but are not limited to:

- Automating parts of playlist creation to save time
- Commiting to greater efforts in leaving users more satisfied by using the platform and its services.
- Better help music listeners expand their music knowledge beyond genre
- Provide a deeper understanding of music and the role it plays in our daily lives

Data Analysis

Classification Clusters

Examined the song characteristics averaged out by genre in order to understand how our algorithm is separating songs into playlists. (baseline to observe our clusters against.)

<u>We chose this specific dataset mostly due to the depth of information provided.</u>

Basing recommendations off of technical music attributes would be best. After searching through kaggle, we were able to find this dataset which included the genre for each song. We decided this attribute would be important in our analysis when comparing the success of our model hypothesis.

III. The Clusters

Cluster 00

Majority Genre: Trance/Psytrance
Defined by: Loud and high energy, mostly instrumental
Notable songs:
Pump it - Black Eyed Peas &
Locked out of heaven- Bruno Mars

genre	count(genre)
Dark Trap	184
Trap Metal	72
psytrance	418
Emo	59
Rap	71
trap	717
hardstyle	785
Underground Rap	244
RnB	69
Pop	12
Hiphop	83

Cluster 01

Majority Genre: Trap
Defined by: Low valence but high energy
Notable songs:
Radioactive - Imagine Dragons & Viva La Vida- Coldplay

genre	count(genre)
Dark Trap	287
Trap Metal	89
psytrance	476
Emo	59
Rap	86
trance	569
trap	156
Underground Rap	256
RnB	61
Pop	13
techno	150
Hiphop	60

Cluster 02

Majority Genre: RnB
Defined by: heavy bass, low valence, but high energy songs
Notable songs:
Still don't Know My Name- Labrinth & Centuries- Fall Out Boy

genre	count(genre)
Dark Trap	108
Trap Metal	60
Emo	87
dnb	906
Rap	40
Underground Rap	214
RnB	86
Pop	9
Hiphop	183

Cluster 03

Majority Genre: Dark Trap
Defined by: Faster tempo, higher acousticness with little spoken word
Notable songs:
Crazy in Love- Beyonce & In The End- Linkin Park

genre	count(genre)
Dark Trap	135
Trap Metal	47
Emo	34
Rap	2
Underground Rap	125
RnB	74
Pop	20
Hiphop	73

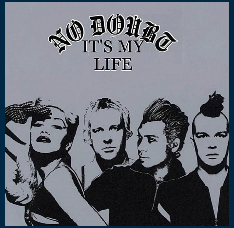
Cluster 04

Majority Genre: Trance/Psytrance
Defined by: the highest acousticness and low speechiness
Notable songs:
Stronger -Kelly Clarkson & All Along the watchtower- Jimi Hendrix

genre	count(genre)
Dark Trap	76
Trap Metal	58
Emo	54
Rap	75
Underground Rap	165
RnB	58
Pop	15
Hiphop	84

Cluster 05

Majority Genre: Tech House
Defined by: largest variety of genres and the highest danceability
Notable songs:
The edge of glory-Lady Gaga & It's my life -No doubt



genre	count(genre)
Dark Trap	328
techhouse	933
Trap Metal	129
psytrance	3
Emo	59
Rap	87
trance	346
Underground Rap	295
RnB	112
Pop	48
techno	736
Hiphop	79

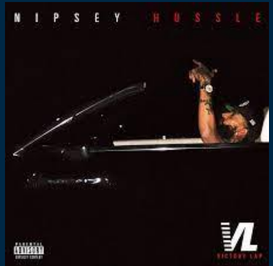
Cluster 06

Majority Genre: Rap
Defined by: highest speechiness, low energy, and low valence
Notable songs:
Nights- Frank Ocean & X Gon' Give it Ya- DMX

genre	count(genre)
Trap Metal	1
Rap	137
Underground Rap	1
RnB	1
Hiphop	1

Cluster 07

Majority Genre: Underground Rap
Defined by: the highest valence and higher speechiness
Notable songs:
Right Hand 2 God - Nipsey Hussle and Become What you hate - Midtown



genre	count(genre)
Dark Trap	68
Trap Metal	29
Emo	56
Rap	3
Underground Rap	226
RnB	87
Pop	15
Hiphop	220

Cluster 08

Majority Genre: Underground Rap
Defined by: being the middle-ground of all attributes compared to the other clusters
Notable songs:
Live your life - TI and Steal My Girl - One Direction

prediction	genre	count(genre)
8	Pop	8
8	Hiphop	97
8	Underground Rap	281
8	Trap Metal	85
8	RnB	75
8	trap	53
8	hardstyle	113
8	Emo	74
8	psytrance	13
8	Rap	61
8	Dark Trap	182

IV. CONCLUSION

Song similarities go beyond genre

While there is room to further support this claim we are confident in our Models ability to generate playlists on the criteria set by the 'vibes'

Looking at clusters is a more efficient way of creating playlists

Versus having playlist editors keep up with current and popular songs that might fall into a specific genre, making it more efficient opposed to current process(s)

Next Steps

Expand data set

Our data set was limited to songs with low speechiness, which excludes many songs Our dataset also lacked acoustic music, which could match with other songs in our dataset

Develop a metric capturing users response to playlists

While our data skills proved to be sufficient in supporting our hypothesis, start expanding into new metrics that will capture the support for the project.

Improve our model

Define more distance between each cluster

THANK YOU!

ARE THERE ANY QUESTIONS?

Dock Dock Goose (Group 11)

- Joey Athanasiou
- Marie Reine Axalan
- Marcos A. Fontes
- Jordyn Gerstle-Goodman
- Talia Zalesne