Project Proposal

Group 14

Aditya Ratna

Umang Mistry

617-642-8352 (Aditya Ratna)

571-430-5577 (Umang Mistry)

ratna.a@northeastern.edu

mistry.u@northeastern.edu

Percentage of Effort Contributed by Student 1: 50%

Percentage of Effort Contributed by Student 2: 50%

Signature of Student 1: 

Signature of Student 2: 

Submission Date: 09/19/2021

Problem Setting

Spotify is a music streaming service with the largest user base with over 365 million active users and more than 165 million premium subscribers. The major challenges for online streaming services are enhancing the user experience and delivering relevant content efficiently. Spotify music suggestions especially for new users is a crucial part of user retention. Premium subscriptions and ads are the major sources of revenue which are again based on number of active users and subscribers.

Problem Definition

Through this project we aim to identify trends within the user base of Spotify. Another goal is to determine what makes a song or album popular.

The main objective of this project is:

1. To create a recommendation system based on the features in the data set and group songs based on similarity and popularity.
2. Implement this model on a user's action on the platform and compare it with Spotify's own recommendation to evaluate the performance.

Data Sources

The dataset is a collection of tracks listed on Spotify. This dataset contains 232,000 rows and 18 columns. The dataset is a collection of categorical both and numeric data.
Categorical features include:
- Genre: It is the genre of track such as R&B, Hip-hop, Country etc.
- Artist Name
- Track Name
Dataset Link: https://www.kaggle.com/zaheenhamidani/ultimate-spotify-tracks-db
For additional data we can leverage Spotify's API to make requests to obtain the data in the required format.

Data Description

Numeric features will be the most important features for our recommendation system. For example:

- Acousticness: It rates the how acoustic the song is on a scale of 0 to 1
- Dancebility: It rates the track on how suitable it is to dance to
- Liveness: It identifies the presence of audience in the soundtrack
- Tempo: Overall tempo of the track given by Beats per Minute
- Time Signature: Specifies how many beats are there in each bar
- Valence: This metric conveys the positiveness described by the track.