

Project 2

Milestone 1 – Proposal

Laura Hoffmann

DSC680

Topic

The streaming business is booming, not only in terms of television and movie content but also in the music industry.

Business Problem

Dozens of audio streaming platforms sharing everything from songs to podcasts and standup routines are being utilized by millions around the globe every day. To keep the business of their subscribers and continue making money, these platforms must ensure that customers are satisfied with their services, not only of quality and quantity of the streaming content but also of recommending new content that users enjoy. Exploring a data set and building a functional model in the prediction for recommendation of Spotify songs could help the business in how it interacts with its users. Being able to successfully suggest new content for users to listen to can act as a type of customer retention for the business.

Datasets

[Spotify Songs](#)

[Spotify Song Attributes](#)

[Spotify's Million Playlist Dataset Challenge](#) (Account created to download private data)

Methods

Cleaning the data: Check for duplicate values and drop. Check for missing values and outliers, handle them accordingly. Variable creation or deletion if not necessary for analysis and modeling portion.

Visualizing the data: Get a good feel for the data using distribution and correlation visuals from the plotly and matplotlib packages.

Modeling the data: Use sklearn package to standardize and classify for modeling predictions on whether a song will be liked or not. Print confusion matrix to show how successful the model is visually.

Ethical Considerations

There is no personal information shared throughout the dataset so confidentiality will not be an issue. Other than confidentiality (and that's not a problem) for this project, it will be a rather simple analysis and modeling process for an interesting and entertaining subject so ethics are marginally involved.

Challenges

Spotify already has many algorithms and models set in place in its program that recommend new songs to its users. However, my Discover Weekly Playlist, a playlist created specifically for me as a Spotify user, based on those algorithms, is usually filled with songs I don't particularly like. Clearly, suggesting songs for millions of users all with different tastes and preferences can pose quite a challenge. I will only be working with one set of data which contains several different categories of ratings from one user and attempting to build a model from that data so what I am working with is limited. So rather than being able to suggest new songs I will be able to build a model that predicts whether a user will or will not like new songs that could be recommended to them.

References

[Spotify's Datasets](#)

[Spotify's Recommendations \(Article\)](#)

[Spotify's Algorithms \(Toward Data Science\)](#)

[EDA and Predictions](#)

[Predictions](#)