Project Title: Does a song/track belong in this playlist?

Henry Talasan

Problem Statement:

The main question I want to explore is **if a song/track belongs in a particular playlist?** This may be done by first analyzing the particular playlist and determining a specific 'theme' based on the current tracks of the playlist. Then see if a particular track fits in that overarching theme. If it does, it would be recommended that the track shouldn't be included, but if it does it is recommended to be included.

The client would be any user interested in music, but specifically a user that needs to curate a playlist like a DJ. By implementing the above, this may help the user better curate a playlist for a particular sound, mood, or intangible 'thing' that describes the playlist by trying to recommend avoiding including a song that may not mesh well with the theme of the playlist.

Dataset:

The dataset/s to be used is a customized playlist/s provided by the client/s assuming that all the tracks (or songs) are part of Spotify's catalog. Each playlist's tracks can can be accessed by using Spotify's Developer API. There are two types of information for each track that is provided by Spotify, Audio Analysis and Audio Features. The Analysis gives data on 'low-level audio analysis for all of the tracks in the Spotify catalog. The Audio Analysis describes the track's structure and musical content, including rhythm, pitch, and timbre.' Whereas the Features gives track data on 'audio feature information' which includes but not limited to: acousticness, danceability, liveness, etc.

Initial Approach:

- Data wrangling. First step in the process is to acquire the data. As mentioned above in the dataset section, this involves the use of Spotify's Developer API. There is a python wrapper, spotipy, that can help in this stage of the process. It seems that the track data comes in a JSON format. A challenge would be getting tracks from a particular playlist. Also if I were to use both features and data analysis, how would I integrate both?
- Exploratory data analysis (EDA).
- Statistical Inferences. This stage would incorporate figuring out which features can best define a given playlist for its theme. This may include but not limited to correlations of features between tracks, .
- Modeling. Create a classification model for a playlist using the current tracks' features or analysis information determining a theme for the playlist. Using that 'theme', songs can be classified to be within that theme or not. A further goal might be to retrospectively look at the playlist and see if there are any songs that does not fit in the theme.

Deliverables:

Deliverables of this project will include a report, a slide deck and documentation of the code utilized.