Analyzing Top Spotify Tracks of 2017

Capstone Project Proposal

1. What is the problem you want to solve?

Spotify is a digital music, podcast, and video streaming service that gives you access to millions of songs and other content from artists all over the world. They currently have over 30 millions songs and over 140 million active users in 61 countries. At the end of each year, Spotify compiles a playlist of the top 100 songs streamed most often over the course of that year. My goal is to do an intensive analysis of the audio features of these songs to determine if there are any similarities in them. Once we have this information, would it be possible to predict if a song might make it in the top 10 or top 100 songs list?

2. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?

With physical music sales declining year after year, the entire music industry is transitioning to a more digital music experience. Music streaming services are becoming more popular than ever and only showing signs of growth. The data in this analysis would be valuable for any music steaming service (Spotify, Apple Music, Youtube, Amazon Music, Pandora, Soundcloud, etc.) to help them easier determine what songs to include in their more popular suggested playlists and emails. This could also help record labels and artists choose the right single to release off an album.

3. What data are you going to use for this? How will you acquire this data?

I will be using the 'Top Tracks of 2017' Spotify playlist as a reference for this analysis. I will also be using the Spotify Web API along with a Python specific API called Spotipy to extract all the available audio features Spotify uses to categorize these songs, such as danceability, tempo, key, etc.

4. In brief, outline your approach to solving this problem (knowing that this might change later).

After extracting and cleaning the necessary data, I will do a full analysis of the Spotify 2017 Top 100 Tracks list and try to find any similarities in audio features among these songs. Once we have some indication of the characteristics of these songs, I'd like to test some recently released songs and see if we can predict if they might be in the top 100 tracks list for this year based on a comparison of their audio features and our findings.

5. **What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.**

   I will be working on this analysis using Jupyter notebooks, which will allow my code & plots/graphs to be easily seen along with notes of my findings as we go through the analysis process. Once this process is complete, I will work on a powerpoint slide deck which I will use for the presentation of my findings.