Omni Lyric Data Science

-Alex Brand ablerex@umich.edu, Roger Barber rebarber@umich.du

1. From Spotify: We’ll pull the top 20 songs from the last 50 years (20 / decade) and store the following info:
   1. Genre
   2. Audio Features – represented as scores which include:
      1. Mood
         1. Danceability
         2. Valence
         3. Energy
         4. Tempo
      2. Properties
         1. Loudness
         2. Speechiness
         3. Instrumentalness
      3. Context
         1. Liveness
         2. Acousticness
2. From Genius: We’ll pass the top 20 songs / decade compiled from Spotify to find the following:
   1. Genre
   2. Lyrics
   3. Artist information
3. We will explore the data to see if there is a correlation between the top 10 words (or phrases if possible) used in music for the top 20 songs per decade (accounting for and removing common words like to, the, and, from, if etc.) and the various audio features described by Spotify.
4. What we’ll graph:
   1. We’ll graph the top 10 words used across all 100 samples to show which words are most used in lyricism for the most popular songs. – Pie Chart
   2. We will then use word counts from the top 10 songs to compare those songs’ top 10 words to their Audio feature scores. – Histogram
5. We want to see if and how music has changed over time, how lyricism has evolved, and how the audio features might relate to the most common words.
6. We’ll be using Matplotlib or Pandas for visualization.
7. We will create an equal split of work, Roger will process song information from Spotify pulling the top 100 songs and their audio features, as well as genre if available. Alex will work with the Genius API to pass in the songs returned from Spotify and process lyrical information to pull the top words used for each song, and genre if available.