Data Science Project Proposal – The Intermediate Wizards

Project Title: Factors That Attribute to a Song Becoming Popular

1. **Members' Duties**

   Justin Hoffman – Lead Data Wrangler/Explorer

   Luke Bondi – Lead Creative Mind (slides)

   Kyle Forsberg – Primary Visualizer (data visualization)

   Sam Wolf – Secretary (helps with each and keeps us organized)

2. **Description**

   The first data set is a big collection of songs that appear on Spotify. The data was collected by scraping Spotify of songs and using these songs to create ratings for each song based on different categories, such as tempo, instrumental, etc. The second data set is Billboard's top 100 songs, and these rankings are taken from each week for the entire year.

   To start each data set is a subset of the original csv as it contained a lot of observations. The first data set is the charts data set. This contains the top 100 weekly songs that Billboard magazine released from 1958 to 2021. This data set contains 4500 observations and 7 variables. The data types for these variables consist of numerical data and characters or categorical data. Each row represents a song that made the billboard during that time frame and its position in the top 100. For the second data set it is the tracks. This contains songs that are on Spotify and their own characteristics. This data contains 1661 observations with 21 variables to describe each song. The data types are both numerical and categorical. Each row represents a unique song that is in Spotify and the given characteristics of the song.

3. **Goals**
   Our goal of the assignment is to try and decipher if there are any attributes that influence a song being in the top 100. If there are any attributes that are, we would like to learn which ones that they are and see if any attributes complement each other. The variables we are interested in include but are not limited to:
   Weeks on board, energy, valance, duration, danceability, genre, explicit