**Tidy Tuesday Step 1: Project Proposal Submission**

Group Members: Yash Agarwal, Diaa Bakir, Angelo Gener, Steven Hua, Muhammad Iqbal, Jia Ye

The data set chosen for the TidyTuesday project is the Spotify dataset from 2020. This dataset contains 12 audio features such as liveliness, energy, duration used by Spotify to make content such as playlists such as Discover Weekly and Recommended for You. Spotify utilizes these features in order to understand the relationships between users and their choice and music. The dataset showcases songs from six genres: pop, rap, rock, latin, EDM, and R&B. The data used in these studies were achieved by getting the top 20 playlists for each genre thus resulting in 5000 songs. After accumulating the data, the correlation between features and genres. The outliers were removed, and the songs were later classified into genres by utilizing the findings and relevant audio features. Songs were then categorized into what genre they should belong to by using decision trees. Additionally, a random forest (multiple decision trees were used to find a bootstrap aggregate to improve prediction accuracy). The findings show that the most optimal method was random\_forest that had an accuracy of 54.32%.

The variables we will be track album release date and song popularity, the relationship between Valence, Tempo, Danceability and lastly, the connection between Liveness and Instrumentalness. We have chosen track album release date and song popularity to determine what albums were the most popular when they were released in order to determine what the most popular albums were and the impact, they had on the overall music environment in these 6 genres. Additionally, the relationship between valence, tempo and danceability was a particularly interesting topic as the relationship between how cheerful or happy the music is and how fast the song is and how that relates to the danceability of songs. The impact that the pace of a song and how happy it is estimated to be a big factor that can suggest how danceable a song is. The connection between how wordy they are and lively the song is will be explored by our group. Often times, the wordier the songs are, the less likely they are to be performed live and this will be researched throughout our TidyTuesday project. However, EDM is a big exception to this believed hypothesis since many EDM songs have a high instrumentalness score but are performed live. In conclusion, the relationship between track album dates and song popularity in order to find the most popular albums upon release, how the valence and tempo affects a song’s danceability and whether or not it can be used as a predictor for a song’s danceability score and lastly, how likely a song with a high instrumentalness score has on a song’s liveliness index.