**Capstone Project Proposal**

**What is the problem you want to solve?**

The problem I am looking to solve is what makes a song popular throughout the world. What are the most common attributes of popular songs and has the importance of certain attributes changed over the past decade?

**Who is your client and why do they care about this problem? What will your client do or decide based on your analysis that they wouldn’t have done otherwise?**

My clients would include recording artists/musical groups who want to gain insight into how they can possibly tailor/alter their music so it has more worldwide appeal and higher chances of topping the Billboard charts. In addition, other clients would be recording labels and music producers who want to develop and release more mainstream songs. Based on my analysis, both artists and producers can ensure the songs they are recording and releasing have the attributes that have proven to contribute to a song’s popularity and in turn financial success.

**What data are you using? How will you acquire the data?**

<https://www.kaggle.com/leonardopena/top-spotify-songs-from-20102019-by-year>

I will be using the dataset above from Kaggle which contains attributes of the most popular songs in the world from 2010-2019. The data contains the following columns/attributes: title (song’s title), artist, top genre (genre of the song), year, bpm (beats per minute), nrgy (how energetic the song is), dnce (danceability), dB (loudness), live (how likely the song is a live recording), val (valence - how positive the mood of the song is), dur (duration), acous (how acoustic the song is), spch (how much spoken word the song contains), and pop (popularity).

Overall, there are 603 rows of data with 14 attributes each that can be analyzed.

**Briefly outline how you’ll solve this problem**

After loading the data into a dataframe, I will first manipulate the categorical variables into a usable form for analysis. I will perform some exploratory data analysis to get further insight into the characteristics of each variable. Then, I will run regressions to see how each variable/combinations of variables impact the popularity of a song. NLP will also be used to determine if certain words associated with song titles or genres affect popularity as well. Along with providing a model that can be used to predict the popularity of a song, I will also provide visualizations to show how the determination of popularity has changed over the past decade and how certain variables relate to each other.

**What are your deliverables?**

My deliverables will include the code used which will be provided on Github. Along with the code, I will provide a slide deck outlining my process and conclusions.