**DATS 6103 Final Project Proposal**

*Alex Khater, Vaishnavi Nagraj, Aditya Nayak, Pooja Chandrashekara*

Project Proposal: Using a dataset of nearly 11,000 songs from Spotify, we would like to compare traditional music theory metrics (Key, Tempo, Time Signature, Duration, Tempo, Energy, Explicit, Mode) and algorithmically generated metrics from Spotify, and see which are better predictors of a song’s streaming popularity. We would like to see what traits make a popular song and (if it is possible to predict a songs popularity using only information about the song).

**S:** This question is tied to 3 specific columns in the dataset and a clear response variable, it is quite a specific question.

**M:** Our response variable is clearly numeric and shows how many times a song is streamed. It is easily measurable.

**A:** This question is quite attainable with relatively basic modeling and regression methods as all variables are pretty standard integer or categorical data.

**R:** This question will attempt to distill a popular song down to its core components from a music theory standpoint and give insight into creating one.

**T:** This project will be easily doable with a team of 4 as the dataset is quite clean from the start and the team is big enough to finish the project quickly

We will use Multiple Linear Regression and any other models that reveal anything interesting.

Dataset Link: <https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset>

Github Link: <https://github.com/awinkhater/DATS6103T3>