**Capstone Project 2: Proposal**

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

Based on current data, try analyzing which features play an essential role in determining the song's popularity. Divide the popularity into ten categories. And determine the rating outcome based on the other features.

**Who is the client, and why do they care?**

Spotify is most beneficial from the project because they can predict the song become popular or not.

**What data do I want to use for this? How do I acquire it?**

I'm going to use a dataset from the Kaggle. <https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks>

**My approach to solving the problem is.**

1. Comprehensive Data Exploration with Python
   * Understand how variables are distributed and how they interact
   * Apply different transformations before training machine learning models
   * Covers both univariate and multivariate approaches
   * Includes visualizations using matplotlib and seaborn
2. Spotify EDA
   * Learn to use visualization techniques to study missing data and distributions
   * Covers both continuous and categorical data
   * Includes correlation heatmaps, pairplots

**What are my Deliverables?**

Code and paper.