**Project Proposal**

**Project Type (Analysis, App Dev?)**

**DSC 478**

**Crystal Contreras, Tavis Sotirin-Miller**

**Overview**

Highlights of ideas, methods and data description here.

**Analysis Approach**

DM Task 1: Pre-processing (general or detailed)

* Sub DM task A: Understanding our Data
  + What type of data is it?
  + What are the features? What types are they? Do they need to be converted into different types? e.g., categorical to numeric
  + Are some features correlated?
  + Is the data skewed?
  + Are there outliers?
  + If you have class labels, is it **imbalanced**? e.g., a few transactions labeled as fraud in a sea of “normal” transactions
  + Part of visualization is about whether you can easily digest and convey the characteristics of your data
* Sub DM task B: Data Pre-Processing
  + Use PCA or Forward/Backward Selection or DT to determine which features are most important to determining popularity of the song per genre (though there already is a “popularity” feature…)

DM Task 2: Exploratory phase e.g., clustering

* Sub DM task A: Explore Spotify API
  + Objective is to find more tracks grouped by genre to add to our dataset for more accurate prediction of popularity. Must have popularity feature.

DM Task 3: Supervised Learning : e.g., classification:

DM Task N: …

**Data Schema and Size**

**Plan for Evaluation – Analysis of Results/ Discussion**

**Plan Work Distribution**