**DSCI 303 Final Project Proposal/Outline**

* **Data set:** [**link**](https://www.kaggle.com/mrmorj/dataset-of-songs-in-spotify?select=genres_v2.csv)
* **Which problems are you trying to solve? Why is it important to solve the problem?**
  + We would like to come up with some model that can help a Spotify user to find their new favorite songs.
  + Listening to music is a big part of lots of people’s lives.
  + Finding new favorite songs may help someone on a stressful day.
  + Helping new artists gain new listeners!
* **Any prior related work [research papers, which methods did they use? How is your project different from others?]**
  + Algorithms other people used for recommendation projects:
    - K means clustering: [link](https://thecleverprogrammer.com/2021/03/03/spotify-recommendation-system-with-machine-learning/)
    - Spotify uses collaborative filtering
    - kNN: [link](https://towardsdatascience.com/how-to-build-an-amazing-music-recommendation-system-4cce2719a572)
* **What kinds of data preprocessing do you need? (if needed)**
  + Standardizing data (normalize values)
  + Drop rows with missing values (some song names are missing)
  + Possibly reduce features after explorations
* **Which algorithms are you planning to use?**
  + KNN
  + Clustering algorithms
  + Regression models
* **Which is your hypothesis about your models and results?**
  + People have certain tendencies to like a song based on one or two features(ex: genre, danceability)
* **How do you evaluate your models?**
  + Measuring some sort of distance function for the user’s inputted song/playlist to the recommended song(s)
  + Does the user like the recommendation?
* **Do you need additional computational resources?** No
* **Timeline**
  + Oct 26 - Final project proposal + exploratory data analysis presentation
  + Nov 9 - Most of the algorithm coding will be complete
  + Nov 23 - Complete code, testing, and generate figures for presentation
  + Nov 30 - Final Project presentation due