**Project Title: Spotify Machine Learning**

**Team Members:** Michael Willis, Gregory Smith, Joe Cotten, Sara Kleine-Kracht

**Project Description/ Outline**

The team will analyze popular song data from Spotify to create and train a supervised learning model that will help predict the popularity of a song. We will then use Spark to query the data and create visualizations that explore trends across the song data.

**Questions to Answer**

* Use the model to predict a song’s ranking on Spotify
* What are the common characteristics that appear in the most popular songs?
  + Characteristics include, but are not limited to: danceability, energy, loudness, tempo, etc.
* What genre of songs are most popular at different times of the year? Can the supervised learning model predict timing of popularity?

**Datasets to be Used**

* <https://www.kaggle.com/datasets/sashankpillai/spotify-top-200-charts-20202021>
* <https://www.kaggle.com/datasets/brunoalarcon123/top-200-spotify-songs-dataset>
* Spotify Web API

**Rough Breakdown of Tasks**

* Research and dataset selection
* Data cleaning
* ETL
* Create, train, fit the model
* Optimize the model
* Data analysis
* Data visualization
* Presentation