**Scope of Work**

**Group # 33**

**Hartin Code, Daniel Nightingale, Will Cunningham**

**We communicated to out TF**

**Project Question:**

Our goal is to come up with a model for generating playlists that will maximize followers/number of listens for the most common user-specified genres or search filters. We will present to Spotify how much potential revenue can be gained through increasing followers on these playlists and will provide insight into the features / song qualities that users care the most about.

In addition to variables obtained from Spotify’s API, we will investigate other potential predictors, potentially using the ISRC number provided in the data to identify songs on alternate databases such as Wikipedia, Billboard, and others. These databases will provide information about artist and song popularities, as well as additional song characteristics, including tempo and loudness.

**Preliminary EDA Plans:**

We will start by collecting and cleaning the data. Data will come from Spotify and other sources. We recognize that the regression will be predicting the success of a playlist, which can range from 5 - 500 songs, and that our dataset will vastly outnumber requirements needed to make a prediction. We will limit our data to a train dataset to perform our initial visualizations and model building. **(Deadline: November 22; Hartin and Danny)**

We will continue by plotting the number of followers of a playlist (on the y-axis) by a number of different features (on the x-axis). We will start with the Spotify-provided predictors and will then migrate to a variety of other predictors as well (as soon as we have identified and cleaned other useful predictors and datasets). Some of these variables will perform better normalized, and some regularized. We will continue to think of other useful **(Deadline: November 26; All members) (we will split up the EDA more precisely once we have a better sense of the data we are working with)**

We will also need to come to a decision regarding the definition of a “successful” playlist: number of plays per day/week/month, a “momentum” factor that measures virality, number of followers (or growth rate of followers), etc.

The end goal of our EDA process will be to extract the variables necessary to create a viable prediction of any given playlist by minimizing model complexity and maximizing model interpretability. **(Dealine: By Milestone #3; All members)**

The writeup will be jointly written. **(Dealine: By Milestone #3; All members)**