

My first project proposal for the data science capstone is to continue working with the top 50 songs from each country and globally, as per the dataset from Spotify, to look further into a prediction model. In theory, the prediction model to build would predict songs to be a number one hit within a country based on how it is categorized within Spotify alone. As an avid music listener and Spotify user, I was able to find a comma-delimited file from [www.kaggle.com](https://www.kaggle.com/datasets/miquelneck/worlds-spotify-top-50-playlist-musicality-data) with the top 50 Spotify songs ranked by country and globally. Within the file, there is 3,588 total data entries of songs from Spotify. At the first point the project was assembled, the dataset held songs from early November 2023, being November 6th to November 12th, the 45th week of 2023. However, I plan on using more updated songs from the original link, as it is expected to be updated weekly. The file contains thirteen columns created by Spotify to analyze tracks, copied and defined by the author of the dataset.

One technique I plan to use is clustering, in the sense of grouping number one songs together based on variables they share, and then re-clustering with different variables, to draw conclusions about what variables matter more when producing and releasing a song to reach the top of the charts. In terms of predictive modeling, the first time I used a trained and tested decision tree. However, as this model only produced around 50% accuracy, I am planning to use different predictive modeling techniques, starting with random forest, in order to attempt to generate a more accurate and much more precise prediction model for songs. Overall, I estimate that this project should take around 2-3 weeks working with what is already in place from my original R code.

Link to dataset: <https://www.kaggle.com/datasets/miquelneck/worlds-spotify-top-50-playlist-musicality-data>

My second project proposal involves using the dataset for the Selected Bridges from around the greater Providence area. Within this project, one of the techniques that I would like to use would be to attempt to build a model predicting how long bridge construction work would take, based on past results on other local bridges. There seems to be a trend with the DOT giving an estimated completion time, and not actually meeting the completion time as expected, usually taking an excessive amount of time longer to complete a repair project. Other things to measure within the project would be variables such as the average amount of traffic going both ways, as well as average time during a rush hour period that it would take one to cross the bridge, in order to gauge a sense of how many people rely on one certain route for commuting in order to show the impact of repair times. With this, I would need to look for another supplementary dataset from a source such as EZPass or Google maps in order to see the full effect on traffic. Overall, I estimate that this project, much like the first one, would take approximately 2-3 weeks to develop and troubleshoot.