Dallin Nielson

DSC 530 – Data Exploration and Analysis

Final Project Summary

For my final project, I explored factors that influence the popularity of Irish tunes on [TheSession.org](http://www.thesession.org/), in hopes that I could identify ways I could boost the popularity of tunes that I publish there. I learned several interesting things about my data during my exploratory analysis. First, none of my variables followed a normal distribution, which means that future analysis would need to look at different types of distributions. Second, two types of tunes (jigs and reels) make up over half of the tunes on the site. Finally, most tunes that are published on the site are on the low end of popularity measures, indicating that it would be unlikely for me to publish a tune that gains widespread popularity.

One large assumption I made during my analysis was that the Books variable is a good variable to use as a proxy for the popularity of a tune. However, some of the other variables such as Sets and Recordings, might also be good proxies for popularity, or perhaps all three of them are related to popularity of a tune. Since the website doesn’t have any upvoting features, I think that using these variables was the best option available as a proxy for tune popularity.

My analysis would have benefited from having a date/time variable for when tunes were first published. It’s very likely that tunes that were published earlier have had more time for users to add them to their tunebooks. This variable could also be used to detect tunes that achieved more popularity in a short amount of time.