

# Predicting Song Popularity

## Analysis and Modeling of Data Scraped from Spotify

J. Bockman, N. Bridges, Z. Prinsloo & M. Vincent  
a1627392, a1720433, a1703479 & a1148120

School of Mathematical Sciences, The University of Adelaide

October 26, 2018

Report submitted for **STATS 2107 Statistical Modelling and Inference** School of  
Mathematical Sciences, University of Adelaide



### Abstract

This paper examines data taken from the online music streaming platform Spotify, and constructs a linear regression model for predicting song popularity. A description of the data used in the analysis is provided, along with the steps taken to clean it. Both univariate and bivariate analyses were undertaken prior to selecting the model. Model selection incorporated forward, backward and stepwise algorithms, and both the Akaike and Bayesian information criteria. The cross validated mean square error of each model in conjunction with complexity were used to select the final model. This resulted in a model that uses five of the eight available predictor variables; decade, danceability, loudness, energy and duration, with two interaction terms between energy and duration, and decade and loudness. After the linear assumptions of the model were found to be valid, a prediction was made using new data.

---