Understanding Pop Music in 2018 - STAT 410 Project Proposal

Question

What makes a song popular? Energetic party music or chill acoustic music for studying? Do we empathize with heartbreaking love songs more, or like cheerful music better? In my project, I want to investigate the relationship between a song's attributes and its popularity. Specifically, what attributes, if any, are the determining factors for a song's popularity? What attributes have a positive correlation with popularity?

Method

I will take advantage of Spotify's API endpoint, which allows easy access to the information of over 35 million songs in its database. As one of the major music player, I assume that it reflects the world's music taste. Spotify has advanced audio analysis algorithms that parameterize each song with semantic characteristics like acousticness, danceability, and energy etc. The API also provides a value for each song's popularity, calculated from the number of the streams and how recent the streams are.

1000 songs with different popularities will be chosen randomly to ensure the variance in response and the results will generalize to other songs. Below is a table that lists the regressors I plan to use. I will first fit an MLR model on the dataset, and then investigate the significance level, the interaction between regressors (e.g. danceability and tempo), and try to reduce the number of parameters.

	Data	Range	Example [Billie Jean - Michael Jackson]
Regressor	acousticness	[0.0, 1.0]	0.0236
	danceability	[0.0, 1.0]	0.92
	duration_ms	positive int measured in ms	293827
	energy	[0.0, 1.0]	0.654
	instrumentalness (Predicts whether a track contains no vocals)	[0.0, 1.0]	0.0158
	liveness (the presence of an audience in the recording.	[0.0, 1.0]	0.0359
	loudness (in dB)	[0.0, 1.0]	-3.051
	speechiness (the presence of spoken words in a track)	[0.0, 1.0]	0.0401
	tempo	float	117.046
	valence (the musical positiveness conveyed by a track)	[0.0, 1.0]	0.847
Response	popularity	int between [0,100]	83