Hello Professor Zu,

Johnny Arguedas and I are working together for this project. Here are the questions we're thinking of focusing on:

1. What is the makeup of songs in our dataset, in terms of genre?
2. What are the characteristics of each genre?
3. Which artists are more or less popular?
4. Are there any cross-genre artists?
5. What dates are covered in the dataset?
6. Does the release date of a song influence its popularity?
7. Do we see relationships between individual variables?
8. What variables influence popularity overall?
9. What variables influence popularity between each genre?
10. Do album ratings and song popularity have any correlation?

I think our overall goal will be to try to determine if different combinations of song aspects (such as danceability, acousticness, etc.) can be used to predict popularity. We suspect that any multiple linear regression model that we come up with will most likely not predict popularity with a high degree of reliability (we're expecting low values for R-squared). Rather than finding a fool-proof popularity "formula", I think our goal will be to determine the effect that different levels of the various aspects have on popularity, and the degree to which certain combinations can be correlated with popularity, whether positively or negatively.

We will likely limit ourselves to one dataset (the Spotify data), but I believe Johnny has an idea about a couple of external data sets that may prove useful.