

## Group Assignment 1 - CPSC6300 - Applied Data Science

1. List the names and GitHub user names for all members of your group.

Nirali Bandaru - niralibandaru

Rohan Gangisetty - rgangis

2. What is your main question?

How is popularity associated with danceability in different genres/moods of music?

3. Provide a brief motivation for your project question. Why is this question important? What can we learn from your project?

Music is known to serve as a sort of therapy to people. We like to listen to sad music when we're sad and happy music when happy. Danceability is generally associated with happy songs. We are curious about what moods or genres of music are more likely to produce danceable songs, and whether negative moods/sad music also has its share of popular danceable songs. This can prove to be important in not only music therapy, but can also be extended to a larger project where one could investigate whether high danceability in sad songs or moody genres improves a person's mood after listening to it. In addition, future AI could help improve a person's mood with its music recommendations depending on their choice of songs (if they have consistently been listening to sad music, for instance).

4. Briefly describe the data source(s) you intend to use. Where is the data available for download? How big is the data in terms of data points and/or file size? If the data is not already available, how do you plan to collect the data?

We intend to use the Kaggle data set of Spotify songs containing 160,000 data points from 1921 - 2020.

Link: <https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks>

This dataset is 28.3 MB.

An alternative dataset is another Kaggle dataset of Spotify songs containing 19,000 data points.

Link: <https://www.kaggle.com/edalrami/19000-spotify-songs>

This dataset is 3 MB.

We have chosen the one with a higher number of data points for more reliable results.

5. What is the outcome (or response or target) in your project? How is the outcome measured / how do you plan to measure it?

The response variable is the popularity of the song. The dataset contains a popularity feature that spans from 1 to 100.

6. Briefly discuss what you think are going to be the key predictors in your project and how those predictors are measured. If you don't have the data yet, what do you think are some key predictors you will need to collect?

Key predictors in this project will be:

1. Acousticness
2. Danceability
3. Energy
4. Instrumentalness
5. Tempo
6. Liveness
7. Loudness
8. Speechiness
9. Year
10. Mode
11. Key

All predictors are given a value spanning a range pre-defined by the dataset.

7. [Optional] Are there any roadblocks you expect your team will face? Are there any problems we should know about? Do you have any questions about your project?
- One possible roadblock is having too big a dataset, and pre-processing it
  - We would like to associate “moods” with danceability and popularity. Would it be better to just choose genres, rather than figure out the mood of a song by some means?
  - We have three major variables we’re investigating- genre/mood, danceability and popularity. Is it okay to have three of them, or would it be better to focus on the association between two? We chose three as we saw fit with the question we came up with.