1. What is the problem you want to solve?

2. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?

3. What data are you going to use for this? How will you acquire this data?

4. In brief, outline your approach to solving this problem (knowing that this might change later).

5. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

1) A lot of music recommendation algorithms rely primarily on a social aspect or some sort of user interactivity or history to determine other songs that we might be interested in. This works to a degree as some of us are likely to check out albums that our friends recommend to us. Music tastes can change though, and a song that may have been a favorite a few months ago may no longer be. Since these tastes are both fluctuating and unique, I believe that it may be best to utilize a recommendation system based solely on audio features, allowing a playlist to be curated on just the audio similarities from one song to the next, allowing for a solo listening experience that can change as often as we would like.

2) The ultimate client would be any listener of music. Currently many people's listening habits are based on others without even realizing, the most popular songs that are chart toppers are due to the fact that everyone is listening to them, leaving some quality music that may not be as well-known entirely un-listened to. By implementing this type of recommendation system, it benefits both the consumer and the music maker, by allowing listeners to experience music that they otherwise would not have, and the music maker reaches a wider audience. More immediately, music streaming and library services such as Spotify, Pandora, YouTube, and iTunes could implement this type of system in addition to existing algorithms to generate a unique listening experience. Though since these services already have a recommendation algorithm in place, it would be most beneficial for a new streaming service to be created that focused their recommendations solely on audio, allowing independent on smaller artists to increase their audience, purely on the music's own merit.

3) After quite a while of going back and forth for a while on how to acquire the data and whose to use, I ultimately decided to go with the Million Song Dataset. Spotify's search which I intended to scrape had a filter already set in based on popularity so I decided to settle on an already curated list. I will begin my work on a 10,000 song portion of the 1 million songs, which was selected randomly from the list, and due to large numbers, should be widespread enough to not

accidentally have bias. If time and space allow, I would like to work on the entire million song set to give the best recommendation system possible.

4) My approach is to first clean and trim down the data to both get a more reasonable dataset size as well as to only extract the audio features from the dataset. Then the plan is to do unsupervised learning and trying to cluster around the most similar features of songs such as tempo and key. My hope is to showcase several examples of playlists that would be created based on a few starting songs. I'm not entirely sure what type of clustering would be a best fit, so I will probably be trying several different types and seeing which ones work best.

5) Deliverables will be the notebook, a report, and most likely a slide deck. If I had more front end ability and more time, I would want to implement this into a live music library.