

## SI 206 Final Project Plan

- A. What is your group's name?
  - a. Spot the song
- B. Who are the people in the group (first name, last name, umich email)?
  - a. AJ deVaux - [ajdv@umich.edu](mailto:ajdv@umich.edu)
  - b. Siya Shah - [siya@umich.edu](mailto:siya@umich.edu)
- C. What APIs/websites will you be gathering data from? (The base URLs for the APIs/websites must be different for them to count as different APIs).
  - a. [Spotify](#) (python version)
  - b. [Kaggle - Album Duration Database](#)
- D. What data will you collect from each API/website and store in a database? Be specific.
  - a. From the Spotify API, we plan to collect data from 100+ songs including the artist that produced it, their duration, the language they are in, and the release year.
  - b. From Kaggle, we plan to collect information about song duration, genre and the year they were released along with the artists that produced the album.
  - c. We also plan on using demographic data of the artist such as their race/ethnicity, age, country of origin, etc, Grammys/awards won, etc., and ?we are looking into ways of collecting and calculating data for duration of the music/duration of words vs music within a song?
- E. What data will you be calculating from the data in the database? Be specific.
  - a. From the data we collected from Spotify, we plan to calculate average duration for songs in different languages and average duration for songs from different periods of time.

- b. From the data from Kaggle, we will calculate the average duration of top albums from given years.
- F. What visualization package will you be using (Matplotlib, Plotly, Seaborn, etc)?
  - a. As of right now, we are planning to use Matplotlib.
- G. What graphs/charts will you be creating?
  - a. We will create a bar graph from the Spotify data showing the average duration of a song based on the language it is in, and also a Matplotlib scatterplot showing the variation in average duration of songs over time.
  - b. We will create another scatterplot showing the album duration as a function of the year it was released to hopefully see if there has been any change over the past 30 years.
- H. Who is responsible for what? Please note that all team members should do an equal amount of programming and total work.
  - a. AJ is working on API1. This included collecting the data, doing the calculations, and creating the visualizations.
  - b. Siya is working on API2. This included collecting the data, doing the calculations, and creating the visualizations.