Project Proposal
Team: Who's That Artist?!

Name	NetID
Hyo Sup Kim (Captain)	hyosup2
Yutong Dai	ymdai2
Chang Hun Park	chp2

- *What is your free topic?*
  - Our free topic is an application-based project, in which we are aiming to create a web application that allows users to gain additional insights into their favorite artists and their music through easy-to-understand visualizations on an interactive dashboard.

- *Detailed Topic Description - What is the task?*
  - We will conduct sentiment analysis and natural language processing on the song lyrics, as well as statistical analysis and data visualization on the audio features of the songs.
  - Song lyrics will be extracted using the Genius API and undergo a comprehensive data cleaning process, including filters stopword removal and spelling correction, to derive their most important features for use in our sentiment analysis and natural language processing of the lyrics.
  - The audio features of the songs will be extracted using the Spotify API, at which point the features will be combined into a format that would allow our code to analyze, draw statistical insights, and visualize information our users may find interesting.
  - The APIs will be connected to a Dash web app, which would prompt users to enter an artist's name (e.g. Taylor Swift), which would output a UI that displays the artist's name, the artist's face, an analysis of the artist's discography, a statistical analysis of the artist's songs' audio features, and data visualizations.

- *Why is it important or interesting?*
  - Data visualization and analysis has permeated through our society and is no longer limited to research and professional fields. Spotify in particular, is notable for releasing Spotify Wrapped, which summarizes the trends and tastes of their users at the end of every year in an easy to understand manner that many users end up talking about on social media- demonstrating its popularity and success. Our goal is to further extend this system functionality to include audio features of songs and song lyrics to see if we can provide further insights into the relationship between a user and their favorite artist.

- *What is your planned approach?*
    - Our team has been meeting every week since Week 5 to discuss the project, divide the work and provide feedback for each other. Notes are taken during the meeting to ensure everyone knows exactly what tasks they are responsible for.
    - We created a mock UI in Week 6 that served as a rough schematic for our vision for the project.
    - All files are stored in a Google Drive folder so that everyone can access the material as needed.
    - Our timeline is as follows:
        - Decide on project topic (W4 ~ W5)
        - Research APIs needed for the project (W5 ~ W6)
        - Work on code that would pull data from API (W6 ~ W9)
        - Transform API data into something useful and interesting (W9 ~ W12)
        - Create Dash app and begin working on connecting dashboard UI to API & analysis code (W10 ~ W13)
        - Tutorial presentation, documentation, and final test run (W14 ~ W15)

- *What tools, systems or datasets are involved?*
    - Google Drive
        - File storage and meeting notes.
    - Spotify API
        - Audio features of songs, artist's face and artist information.
    - Genius API
        - Song lyrics and artist information.
    - Toolkits:
        - NLTK, TextBlob, spaCy, gensim

- *What is the expected outcome?*
    - Our users will be prompted to enter the name of a musical artist, which would output the artist's name, a picture of the artist's face, a list of their songs, an analysis of the song lyrics, a statistical analysis of the audio features of their songs, and data visualizations of our findings.

- *How are you going to evaluate your work?*
    - We will evaluate our work by using informal test cases through the input of musical artists that we are familiar with. For example, entering Chopin's (a Polish composer and classical pianist in the Romantic period) name will result in our code only displaying numerical/statistical information about his discography since his songs do not contain any lyrics. We are also *expecting* a noticeable difference in metrics such as danceability, loudness and energy in his discography than a musical artist such as Martin Garrix, who is a DJ that specializes in progressive house music.

- *Which programming language do you plan to use?*
    - We plan on using Python for the majority of our project; some Javascript may be utilized for the UI of our dashboard.

- *Justify that the workload of your topic is at least 20*N hours, N being the total number of students in your team. You may list the main tasks to be completed and the estimated time cost for each task.*
    - Our team meets at least once a week for about an hour to discuss our progress and planning ahead. A breakdown of our past progress as well as our plans for the future has been provided below.
    - Breakdown: **~ 29 hours/ person (total)** + **~ 1 hour/ week** for our weekly meetings.
        - Decide on project topic (W4 ~ W5) → **~ 1.5 hours**
        - Research APIs needed for the project (W5 ~ W6) → **4 hours**
        - Work on code that would pull data from API (W6 ~ W9) → **8 hours**
        - Transform API data into something useful/usable/interesting (W9 ~ W12) → **8 hours**
        - Create Dash app and begin working on connecting dashboard UI to API & analysis code (W10 ~ W13) → **6 hours**
        - Tutorial presentation, documentation and final test run (W14 ~ W15) → **~ 1.5 hours**