Victoria Dechamps
Progress Report

Project scope update

The scope of my project has not changed since I started working on the project. It is going slowly, and I have only been able to complete the first part (extracting data from Kworb and Spotify API), but it is currently working as planned. The main update is that instead of retrieving all 2,500 datapoints from Kworb, I am retrieving the top 400 songs. This is to be more efficient and avoid rate limits.

Data sources

- Kworb Top 400 Songs - I scraped 400 of the top songs into a .csv file
- Spotify Web API - used to collect metadata for the 400 top songs (based on Kworb dataset)

Issues / difficulties

Spotify seems to limit the number of requests that can be made in a short period of time. To handle this I added checkpoints to save data between batches. As mentioned, retrieving thousands of songs was taking too long, so I'm now working with a smaller section of the Kworb dataset.

I also hadn't realized that we can't commit the data/ folder. I initially planned to manually copy/paste my Kworb data into a csv file, but eventually realized that I had to create this csv file via scraping. This is not necessarily an issue, it's just a work-around.