Group 4 Project Report

- **E**xtract: your original data sources and how the data was formatted (CSV, JSON, pgAdmin 4, etc).
    - We extracted data from Spotify API and CSV.
    - The csv was a list of the top 50 songs from 2019 and their attributes (rhythm, tempo, etc..).
    - The Spotify API was used to extract information from user playlists.
    - The Spotify API was also used to extract current information for the artists who charted on the 2019 csv.
- **T**ransform: what data cleaning or transformation was required.
    - Chris
        - After the csv data was imported into pandas, he cleaned it to keep only the data he wanted (artist name).
        - He then used a for loop using a list of the 2019 artist names to grab their unique artist IDs using the Spotify API.
        - He then created a new data frame using the artist IDs to grab current attributes about the artists, such as popularity, genre, and number of followers.
        - These new attributes were placed in a new data frame and pushed to a relational SQL database.
    - Lawrence
        - Using his current playlist, Lawrence filtered through the Spotify API calls in json format to pick out the artist's name, song name, track ID, artist ID, album the song is from, and the length of the song.
        - He then organized all of this information in a pandas data frame.
        - He then created a SQL table to hold the data and pushed the data he cleaned to a relational SQL database.
    - Dora
        - Using her current playlist, Dora filtered through the Spotify API in json format and pulled the track ID, using the track ID to find the audio features for each song. These features included song name, artist, album, energy, loudness, and track ID.
        - She then cleaned the data and added a track ID column to her data frame.
        - She then created a SQL table to hold the data and pushed the data she cleaned to a relational SQL database.

- **L**oad: the final database, tables/collections, and why this was chosen.
    - We choose SQL as the preferred database because the data we pulled from Spotify API help common attributes like artist name, artist id, song name, and track id, so we thought a relational database appropriate.
    - After pulling all three tables into one database, we created an ERD to show how the databases were related. The main database contained artist names and artist ids, and we used artist name as the primary key to connect the three tables.

/s/

Dora, Chris, Lawrence