We decided to use Spotify data that focused on different categories that make up a song such as energy, loudness, beats per minute etc. The purpose of the project is to see if there are any similarities in the most popular songs and if those similarities could be judged against the song’s popularity score. We pulled three data sets from Kaggle, which provided reliable data. Tables were created in pgAdmin then the Kaggle data was imported the first two data sets were the top 50 songs from 2019 and 2021, those data tables were cleaned and then brought together as one table using the UNION function in SQL. The third data set consisted of songs from 2010 to 2019 the data was cleaned and put into a table in SQL that used the same column headings as the previous top 50 songs to make the process as uniform as possible. All three table were uploaded to GitHub and then used in the machine learning process.