# Analysis of Billboard Top Songs and Spotify Top Songs
## from 2010 to 2019

Alvin Victor & Hannah Moy

**Introduction:**
The music industry has changed dramatically over the past decade due to the advancement of technology. Billboard began their notable top charts in 1958, using Nielson Soundscan, record sales data and radio airplay impressions. In 2003, physical record sales faced a major decrease with the creation of iTunes and digital purchasing. Soon later in 2006, Spotify was formed, eventually replacing both physical and digital sales with online streaming. This project will use ETL to analyze the top songs ranked on both Billboard, and the top songs ranked on Spotify (2010 – 2019) to discover music trends throughout the growth of technology.

**Extract – Data Sources:**
Kaggle – Top Spotify Songs from 2010 – 2019:
https://www.kaggle.com/leonardopena/top-spotify-songs-from-20102019-by-year

Data World – Billboard Hot Weekly Charts:
https://data.world/kcmillersean/billboard-hot-100-1958-2017

**Transform – Process:**
1. Microsoft Excel was used for the first stage of cleaning process. Our team manipulated the basic formatting to merge data sets by similar columns. This was also crucial to account for case sensitivities in both pgAdmin and Python.

2. pgAdmin was used to our team created schemas for both data sets. In these schemas, we listed the appropriate variable character fields for all columns and transformed parts of the data to identical formats. Due to differences in our column names, we faced a number of difficulties accounting for the case sensitivity differences between pgAdmin and Python and had to go back to Excel to fix.

3. Jupyter Notebook was an integral part in the final cleaning process and data transformation. Using Sql in Jupyter, we merged both datasets, dropped unnecessary columns, and reindexed columns. We faced a few difficulties when merging the data sets by date due to formatting issues but eventually learned how to reformat to matching sets. This cleaning process allowed us to read both data sets side by side.

**Load – Results:**
After cleaning and transforming the data sets, the final tables were loaded through pgAdmin. When analyzing both data sets, Spotify has a small, but more significant change in variety of genres starting from 2017 to 2019 while Billboard rarely charts a song of a different genre than pop. The development of Spotify and interactive streaming has offered audiences to actually choose when and what to listen to, which has opened up the industry game to a more diverse players.