March 9, 2020

… So I Can Sing Along to Katy Perry, that’s why!

For my ETL project, I chose to make a database of song lyrics. For my first data source, I chose a csv that I downloaded from Kaggle. This csv contained a list of all the songs that made it into the top ten songs on Spotify for the years 2010-2019.

The first step was loading the downloaded csv into the Jupyter notebook, then dropping NAN values and selecting only the title, artist, year and genre columns. From this dataframe I created a lyrics dataframe, taking the artist and title columns, then adding a column for lyrics.

Next, I ran an iterative loop making calls to an API called apiseeds.com, using each artist and title in the lyrics dataframe to search, then placing the lyrics into the empty column. This API was only able to supply lyrics for about 50% of the songs.