## **Data Cleanup & Analysis**

Once you have identified your datasets, perform ETL on the data. Make sure to plan and document the following:

* The sources of data that you will extract from.
  + <https://data.world/kcmillersean/billboard-hot-100-1958-2017> - import.io
  + <https://www.kaggle.com/nadintamer/top-spotify-tracks-of-2018> - API

* The type of transformation needed for this data (cleaning, joining, filtering, aggregating, etc).
  + Cleaning: Removed unused columns, Song titles, column names, removing duplicates while maintaining the lowest peak position and the highest amount of weeks on the chart(using a sort function), transposed dates from string to an integer date, dropped songs that did not chart.
  + Inner joining: On song name to combine song attributes with Billboard position information.
  + Correlating each list of attributes with peak position and weeks on the Billboard Hot 100.

* The type of final production database to load the data into (relational or non-relational).
  + Relational - PGAdmin SQL

* The final tables or collections that will be used in the production database.
  + migoss.csv