Project Proposal:

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

Analysis: Create a dataset that will allow for analysis of Spotify popularity compared to Billboard Top 100 charts: including different variables to explore (danciness, liveliness, and duration)

- The sources of data that you will extract from.
    - [Billboard Charts CSV](#)
    - [Top Spotify songs from 2010-2019 - BY YEAR](#)
- The type of transformation needed for this data (cleaning, joining, filtering, aggregating, etc).
    - Remove years before 2015
    - Remove unnecessary columns that don't support analysis statement above
    - Join on song
    - Support filtering based on level of danciness, liveliness and duration
    - Support aggregating by year
- The type of final production database to load the data into (relational or non-relational).
    - Relational
- The final tables or collections that will be used in the production database.
    - Not fully flushed out yet--but high level:
        - Artist table
        - Genre table
        - Song/Billboard Charts