ETL Project Proposal

E – Extract:

* Two datasets, both from Kaggle
* One dataset contains information about 500 of the top songs (all-time)
* The other dataset is information on thousands on tv shows and movies on Netflix
* Both datasets are formatted in csv

T – Transform:

* Transform in Pandas
* Delete unnecessary columns in each dataset
* Filter tv shows/movies dataset by year, rating, and number of seasons
* Filter songs dataset by month, year, streak (# of weeks on Billboard), and position (highest position a song landed on Billboard).

L – Load:

* All final databases will be loaded into Postgres
* Postgres was chosen because of familiarity
* Database name will be “ETLProject.”