Plan:

* Option 1: focus solely on billboard data. Use Genre, Artist, Lyrics, etc. to predict rank
* Option 2: focus solely on spotify data and create the dataset to compare hit songs v non-hit songs from the same album...predicting song likeability using 5-7 factors

Outline:

* Wednesday 7/15
  + Finish gathering data
  + Begin machine learning / analytics
* Saturday 7/18
  + Finish machine learning / analytics
* Monday 7/20
  + Hosting / displaying data
  + Javascript
* Wednesday 7/22
  + HTML / CSS / Bootstrap design
  + Finishing javascript touches

----------

Interesting bits - since there are 6, we could each take on 2 of these?

Outline

* The Question: Can we predict popularity based on song attributes?
* The Data:
  + Spotify API
  + Musical attributes
  + Artist attributes
  + A word on “popularity” being included
* The Analysis:
  + A word on “class weights” imbalance and why that matters
* The Results:
  + A word on random states and variability in results
    - Random State = using the same data in training to predict the test results
    - Different “data slices” will produce different results when there is limited data (our set is ~2500 songs)
  + Overall - mixed
  + A word on score vs precision vs recall and how it relates to our research question
    - Precision = of the total number of predicted “top songs” were actually “top songs”
    - Recall = the number of top songs actually selected vs the the total number of top songs in the test set
  + A word on the behavior of different models - Flatterer v Sour Puss
  + Show some actual v predictions
* Conclusion:
  + It won’t help you identify songs that will make it to the Top 10 BEFORE they are produced
  + But once you produce a song and it’s on spotify for enough time, it may help you decide whether to invest in marketing to make it BIG
* Additional Fun with Songs:
  + Can ML models identify the artist?
  + Mainly, yes
  + A word on the false positives and how that might lead to interesting findings