MUSIC WORDS SLAP

By: Cole Gaito
GitHub Location: https://github.com/gaitocole/Final_Project.git
Video Location: https://youtu.be/KyqpACFpE2E
(See Music_Trends)

Research Questions

1. Can lyrics predict Billboard chart success?
2. Are thematic correlations evident among top hits?
3. Do world events influence lyrical content?

- Data set Location from Kaggle/Github/Billboard:

    https://www.kaggle.com/datasets/dhruvildave/billboard-the-hot-100-songs

    https://github.com/utdata/rwd-billboard-data

    Original Source: https://www.billboard.com/charts/

    Lyrics: Genuis API (built inhouse): https://genius.com

    Song Metadata: https://developer.spotify.com/dashboard/create

The Whose Who in Group Two

Cole Gaito

Where you at?

Saved Locally

The What's What

■ This project, I wish to answer the following questions:

Is there a thematic correlation between top chart hits?

Is this derived from the words themselves, or correlated to world events?

Novel Ideas and Prior Works

■ While I didn't explicitly look for the question there are similar themes and projects such as:

https://cs.brown.edu/courses/cs100/students/project11/

https://www.reddit.com/r/musicleague/comments/kq3p3b/theme_ideas/

What to do in world in tune?

Anticipated Work (done during the Genuis API phase):

Data Cleaning – Almost None

There are a few errors that have been mentioned throughout discourse on the data set, but overall, the data is quite clean.

Data Incorporation – Significant

Work has already begun on obtaining the lyrics to all the associated songs. I expect that this along with any word identification to take the longest of all associated tasking.

Data Preprocessing – Some

Here I expect that I will need to do some stastical analysis to perform and derive further data. Overall while there is significant volume of data ~52weeks x 60+ years, once established the process will not take long.
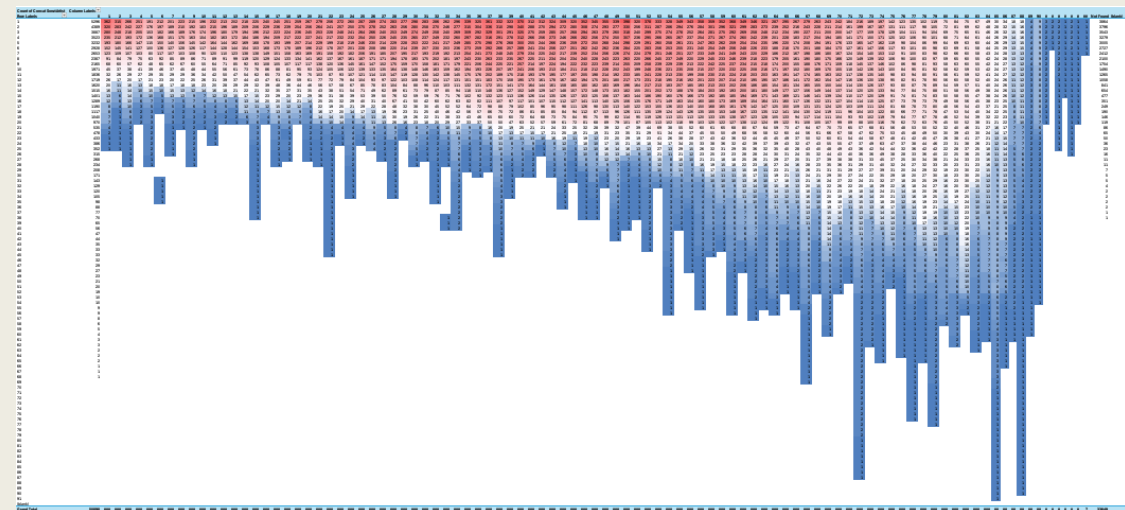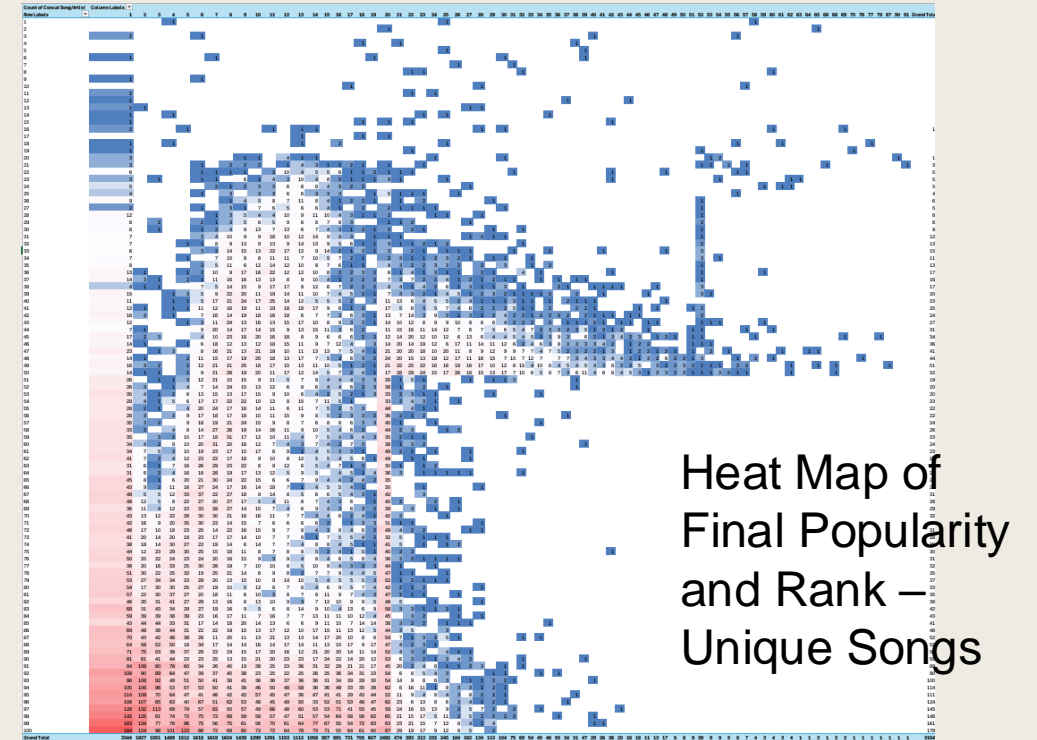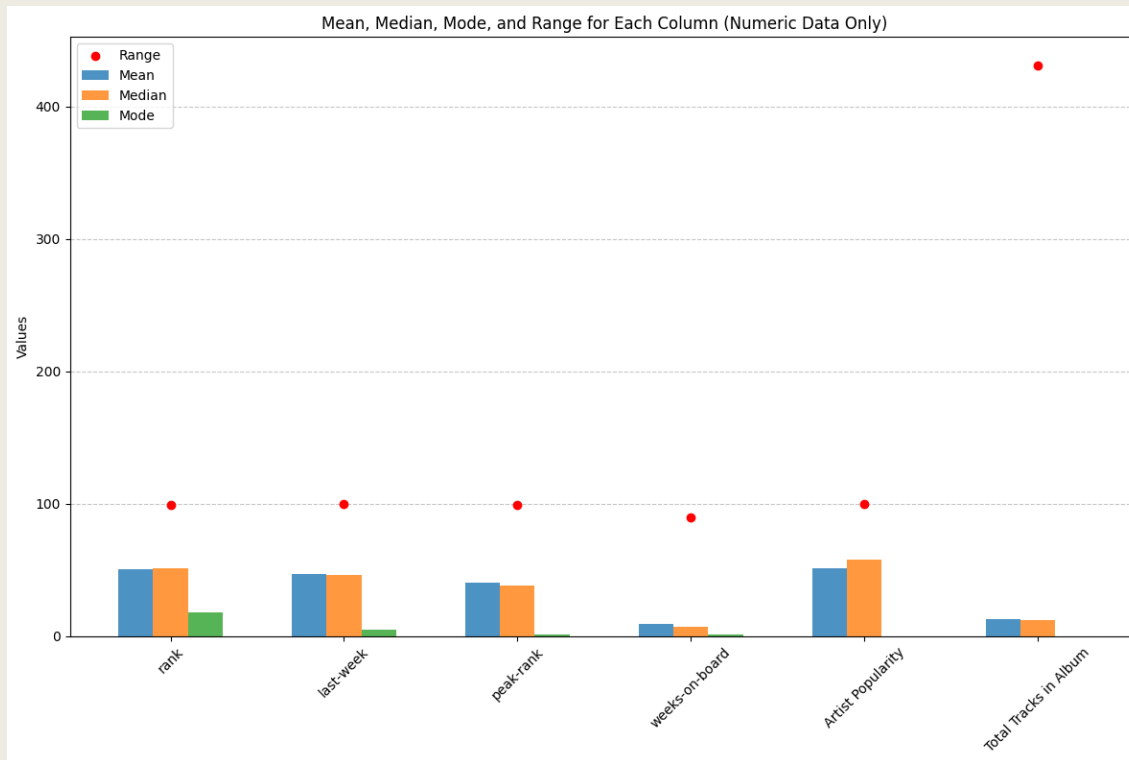
When do we know we have a hit?

Tools:

Ultimately, I am sure that there are trends within the music hits. I plan on utilizing Jupyter Notebooks to stage all the associated code. Python3 will be the base coding language. I assume that plotting will need to occur so ScikitLearn will also be a functional library.
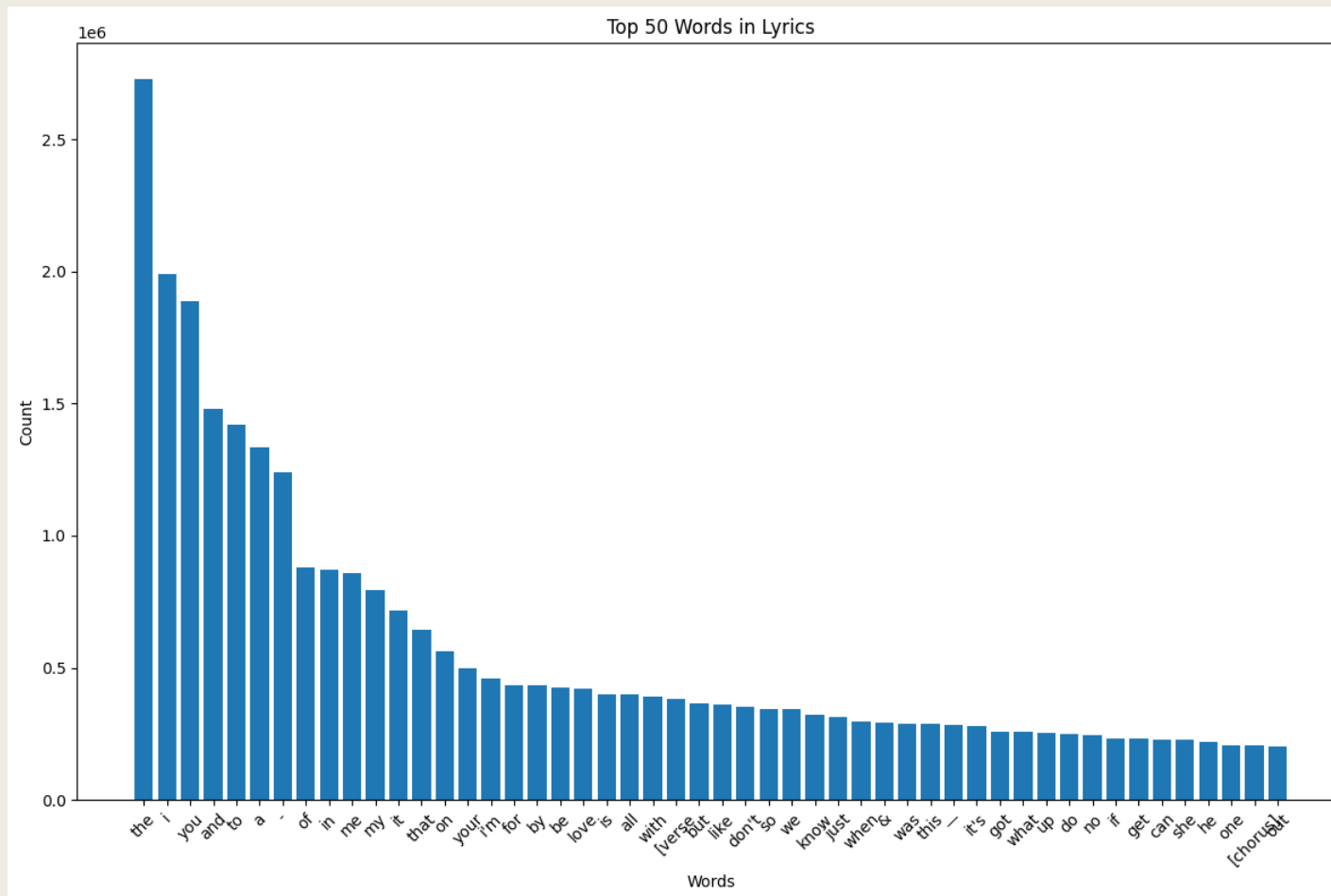
Evaluation:

In terms of finishing, being able to capture trends, or dispel the idea that trends exist within the charts among lyrics. If the data isn't dense enough there may be other methods of evaluation like beat and chord progression, but for the moment, I don't plan on analyzing that.

Heat Map of Final Popularity and Rank – Unique Songs

Heat Map of Final Popularity and Rank – All Songs

Results

Top 50 Words in Lyrics

Results

Prediction: Range of +/- 19% confidence
Epic Computation Time : 7-15 seconds per step totaling 8 days for 5 Epics

Application of Results

What was learned?
1. Can lyrics predict Billboard chart success?
   – Yes, though with limited accuracy t0 19%.

2. Are thematic correlations evident among top hits?
   – Yes, love is the highest scoring non-trivial word.

3. Do world events influence lyrical content?
   - Undetermined, available for further research.

How can this be applied?
• From a financial perspective, you can bet love songs will be more successful.
• From a RNN, we can determine that the lyrics may not be enough…