Spotify®

Trending 163,528

Contributors = [Sarah Caldwell,
Rainier Radkowski,
John Silberstein,
Conrad Urffer,
Clara Bucar]

The Dataset

Data Summary:

- Data on Spotify Top 200 songs from 35 countries + global for the range 2017-2020.
- Two CSV Files ~ 1.64GBs total.

Source:

- Kaggle.com - Link included in project repository.

Tempo / BPM Analysis

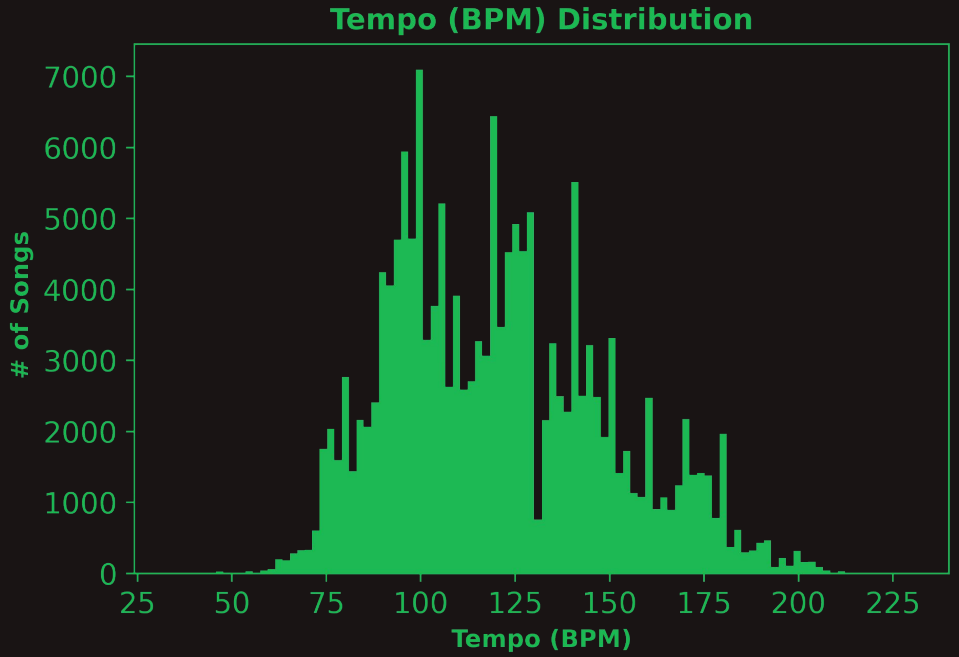
Question:

Does choosing a popular tempo increase the popularity of a song?

Data used in Analysis:

- Tempo - The overall estimated tempo of a track in beats per minute (BPM).

- Popularity - Weighted score based upon the number of days and in the Top 200, and the ranks it had during those days.



Top_5_BPMs = {
***"100 BPM"** : 5816,*
***"120 BPM"** : 5520,*
***"140 BPM"** : 4934,*
***"130 BPM"** : 4340,*
***"105 BPM"** : 3649,*
}

- **163,527 Songs**
- **Total of 174 different BPMs**

- **Mode: 100 BPM (5816 songs) = 3.6%**
  - Mean: 121 BPM
  - Median: 120 BPM

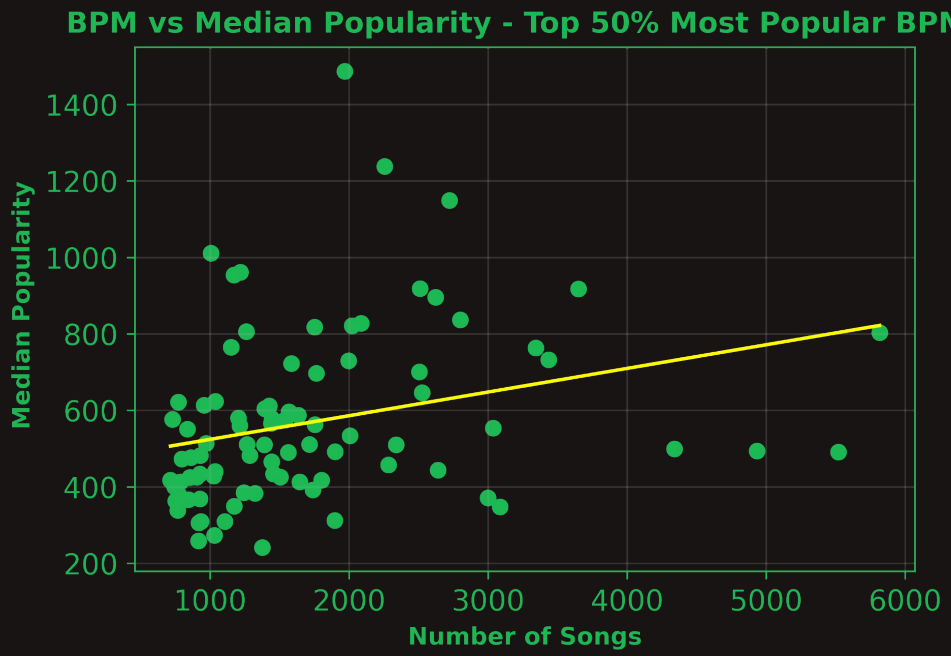
- **Not normally distributed.**
  - **p-value = 0.0**

- **The top 10 most popular BPMs represent 24.56% of all songs.**

- **The bottom half of BPMs represent 7.95% of all songs.**

*** A single song had 0 BPM. This outlier was removed.*

*** All BPMs were rounded to nearest whole number.*
Ex. 187.23 BPM >>>187 BPM.



Top_2_BPMs = {
***"100 BPM"** : 5816,*
***"120 BPM"** : 5520,*
}

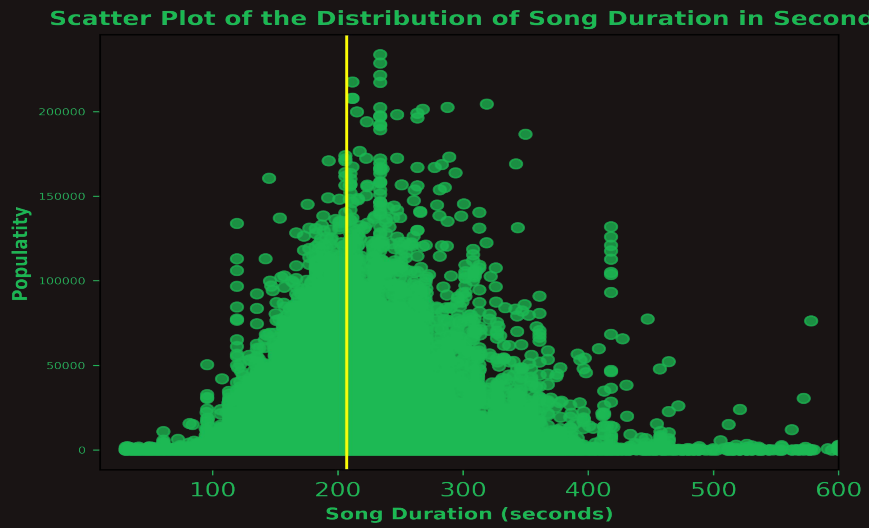
- **Analyze top 50% of BPMs (87 Total)**
  - Represents 92.05% of Songs
- **Pearson R = 0.283**

Conclusion:
There is a weak positive correlation between the popularity of a tempo, and the popularity of a song.

Songs with popular tempos are marginally more popular on average compared to songs based upon less popular tempos.

*** Median rather than mean was used to avoid the impact of exceptional outliers of popularity within the dataset.*

Song Length / Popularity Analysis



- Population Mean = 207 seconds
- Population Mode = 192 second
- Population Median = 202 Seconds
- Standard Deviation = 50 seconds
- Pearson Correlation Coefficient = 0.018
- The graphic has been truncated after 600 seconds (songs greater than 10 minutes in length)

Question: Is there a correlation between song length and the popularity of a song?

Data used in Analysis: **Song length** - which was recorded in milliseconds but for the purposes of this analysis converted to seconds.
Popularity - Weighted score based upon the number of days and in the Top 200, and the ranks it had during those days.

Conclusion: There is not a relationship between song length and popularity. The song length has been determined historically by the technology of the recording media.

Release Type / Popularity Analysis

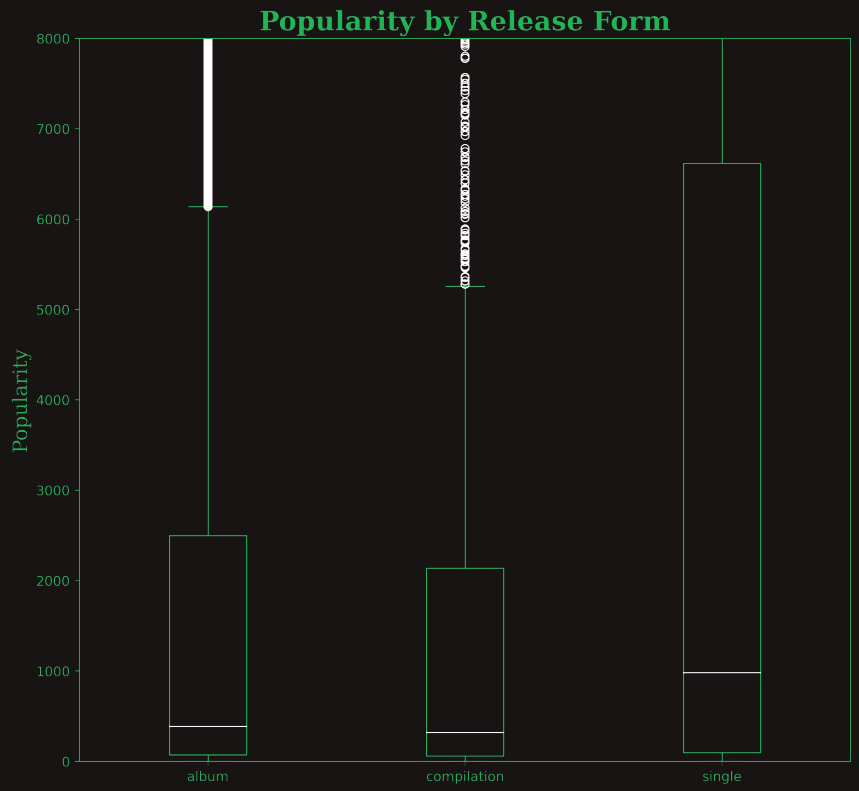
Question:

Is there a difference in song popularity by release type?

Data used in Analysis:

- Album/Single - Was the song released as a single, album, or compilation?

- Popularity - Weighted score based upon the number of days and in the Top 200, and the ranks it had during those days.



- Popularity Means by Subgroup :
  - Album = 4,686
  - Compilation = 3,455
  - Single = 6,748

- One Way ANOVA
  - P-value = 5.7e-221

- Independent T Tests:
  - Album/Compilation
    - P-value = 3.3 e-09
  - Compilation/Single
    - P-value = 3.4e-53
  - Album/Single
    - P-value = 1.9 e-206

Conclusion:

All groups show significant differences in popularity. Songs released as singles are significantly more popular than those released within an album or compilation. Songs released in an album are significantly more popular than those released within a compilation.

Explicit / Popularity Analysis

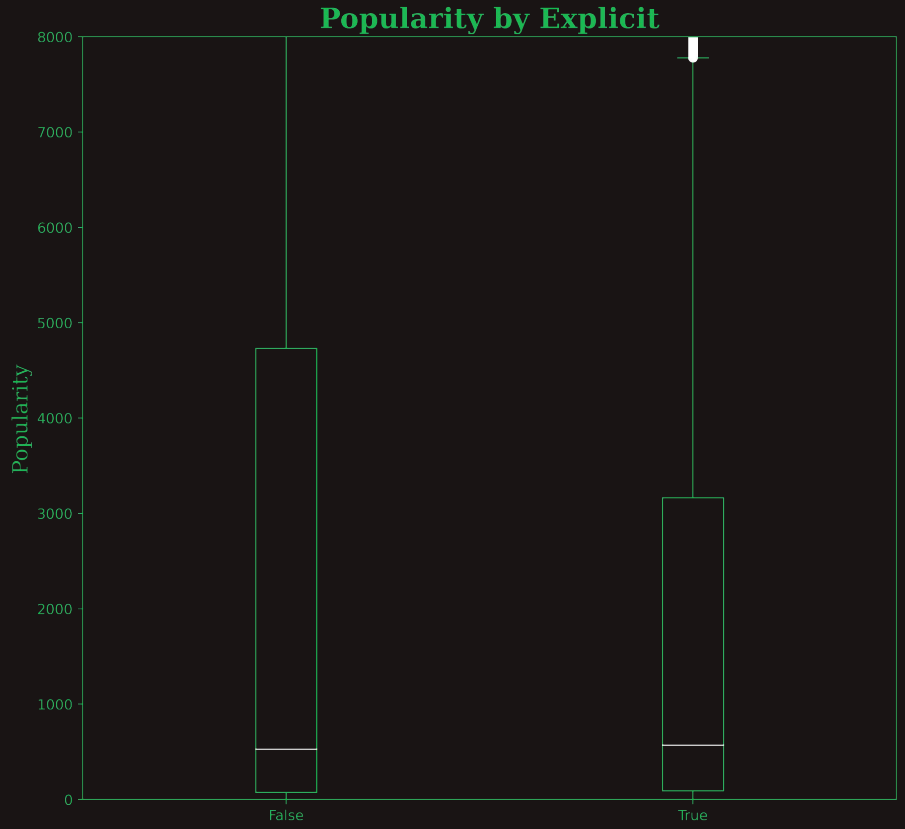
Question:

Is there a difference in song popularity by whether or not it is explicit?

Data used in Analysis:

- Explicit - Is the song Explicit? (T/F)

- Popularity - Weighted score based upon the number of days and in the Top 200, and the ranks it had during those days.



- **Popularity Means by Subgroup :**
  - **False = 6,181**
  - **True = 4,476**

- **Independent T Test:**
  - **P-value = 2.84e-161**

<u>Conclusion:</u>

Songs that are not explicit are significantly more popular than songs that are explicit.

Emotions in Songs Analysis

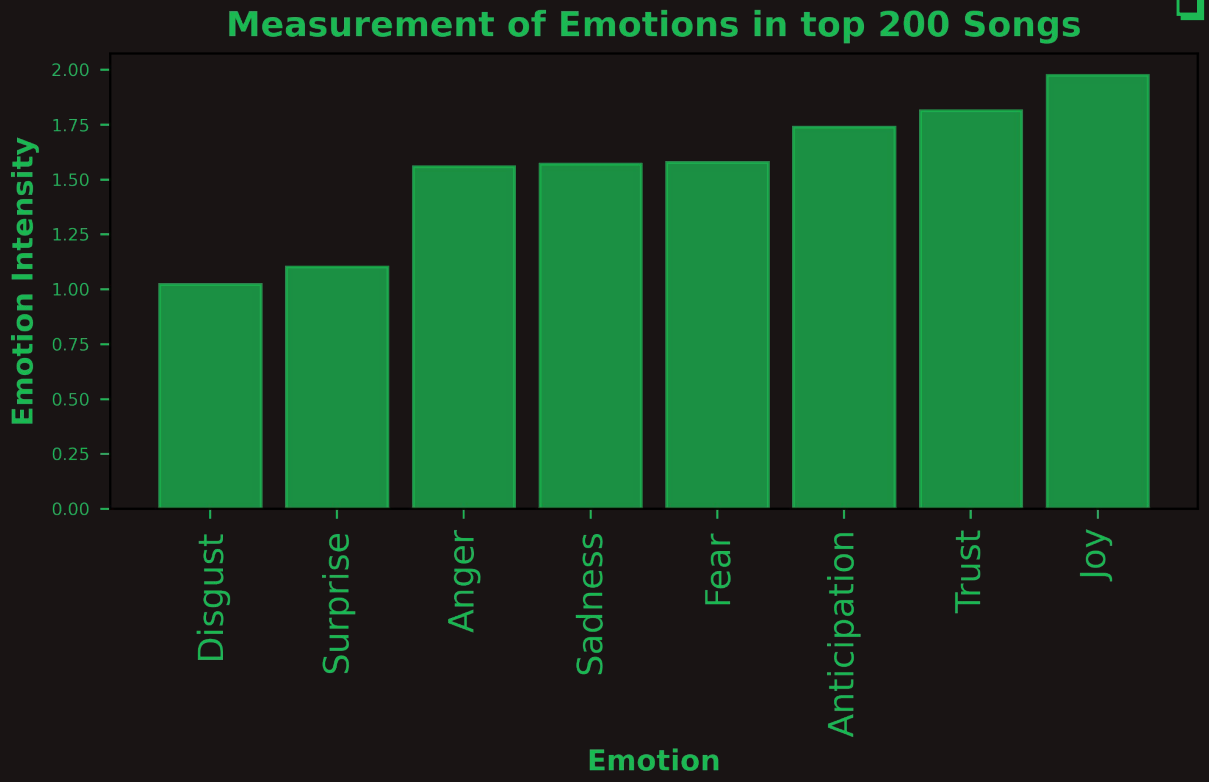
Question:

How do emotions detected in English speaking songs in the top 200 correlate with popularity of a song and genre?

Data used in Analysis:

- Popularity
- Genre
- Number of words related to emotions (anger, disgust, joy, surprise, anticipation, fear, sadness and trust) divided by the total number of words found by the dictionary
- Number of words related to emotions (anger, disgust, joy, surprise, anticipation, fear, sadness and trust) divided by the total number of emotional words found in the top 200 songs

Analysis limitation: Only applicable for songs in English language, all 35 countries + Global

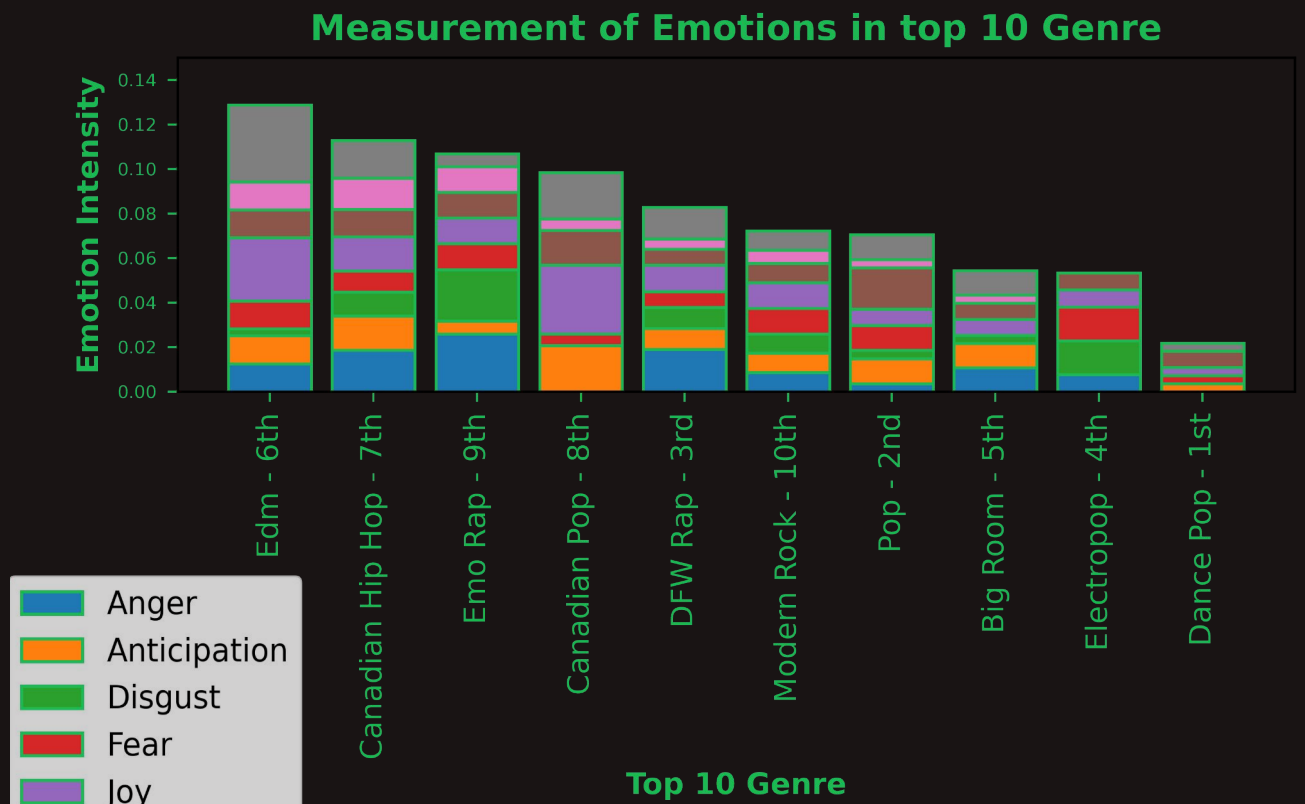


- **Analyze top 200 songs by emotion**
  - Only applicable for songs in English language, all 35 countries + Global considered
  - Songs were grouped by title and emotions rates were averaged throughout all 200 top songs

Conclusion:

There was no recognizable trend of preponderant emotions in most popular songs, the measured emotions were dispersed throughout the data sample.

The most detected emotion in the sample was joy, followed by trust. Overall, positive emotions were predominant over negative emotions.



- Analyze top 10 genres by emotion (45 Total)
  - Popularity per country was summed to reflect weight
  - Emotions rate per country were averaged

Conclusion:

There is significant difference in how emotions are expressed throughout popular genres. The distribution of the emotions can be considered counter intuitive, showing modern rock, for instance, as noticeably less emotional than edm (electronic dance music).

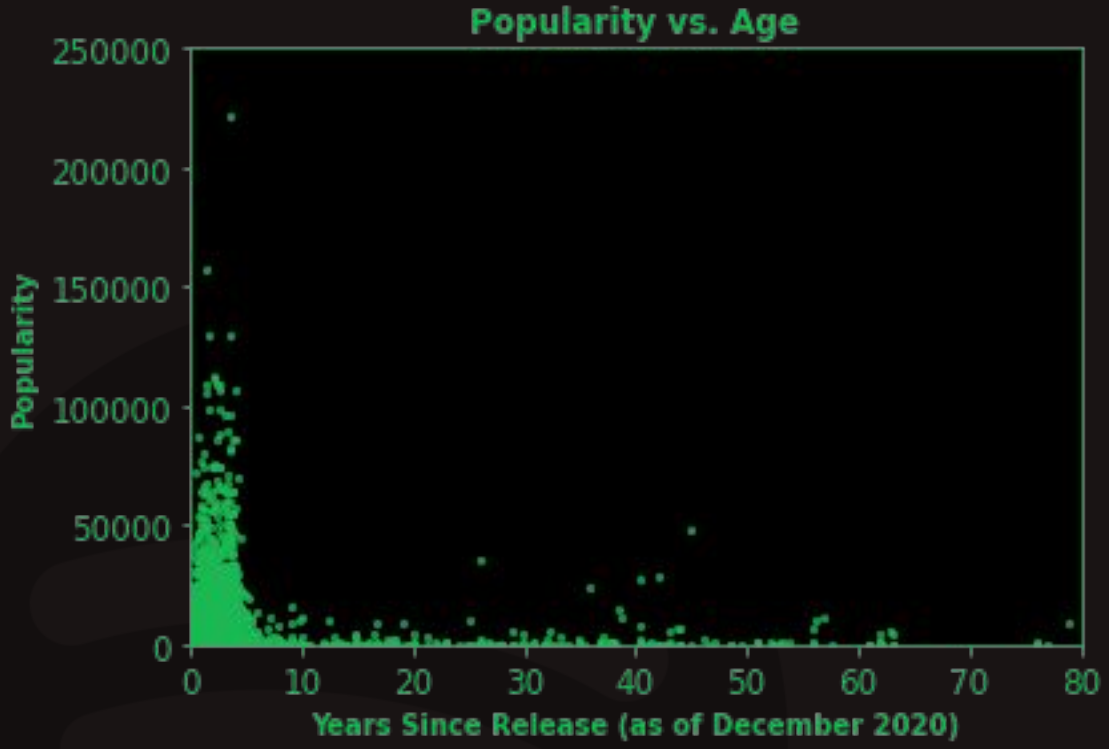
Time of Release Analysis

Question:

Are there trends over time in the overall popularity of Spotify songs?

Data used in Analysis:

- Time Since Release - How old the song is
- Popularity - Weighted score based upon placement in the Top 200, and how long it stayed there.
- Reminder - data includes top 200 songs from 2017-2020



Conclusion:

- Our dataset does not cover a large enough time frame to identify large-scale changes in popularity over time
- Popularity charts are dominated by recent songs
- Several decades-old songs become moderately popular each year
- Total number of streams would also be necessary for further analysis

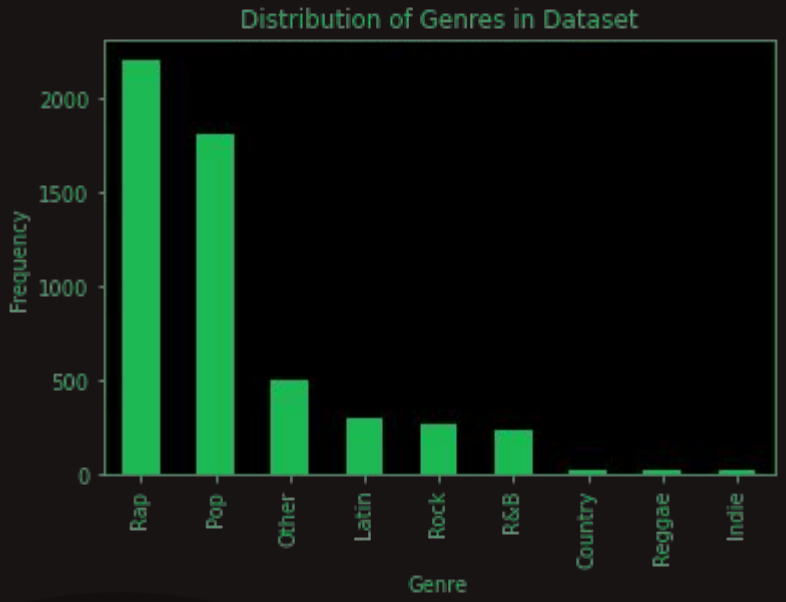
Genre & Popularity

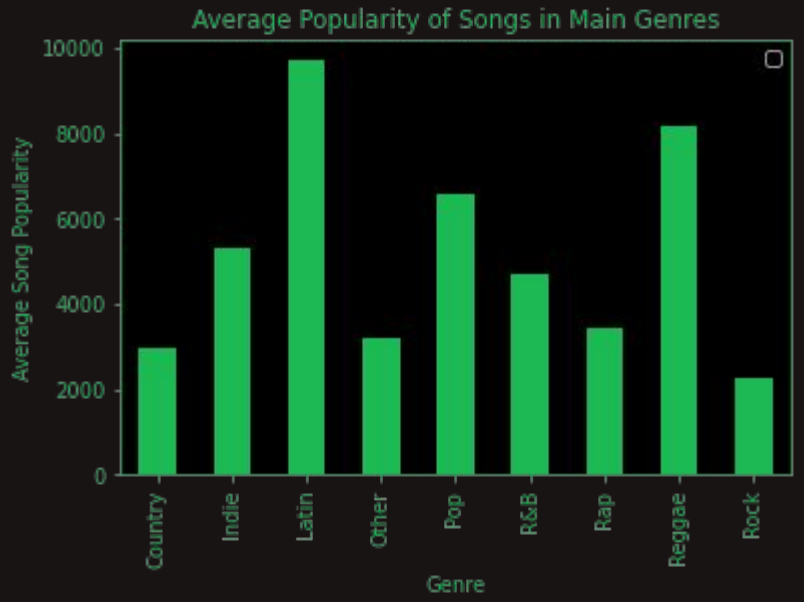
Question:
Which genres are most likely to be popular?

Data used in Analysis:

- Genre - data was grouped into broad genres

- Popularity - Weighted score based upon placement in the Top 200, and how long it stayed there.





- Rap and pop were by far the most common genres
- However, Latin and Reggae were the most popular
  - The “Latin” genre does not include songs that could fall into another category
- Difference in popularity among genres is even more pronounced when the median is used
  - Most of the difference comes from outliers in “expected” popular genres

- One way ANOVA F-statistic: 17.65
  - P-value ~10^{-26}
- Rap & Pop have the most popular songs - but they are outliers
- Reggae and Latin are more consistently popular