POPULAR SONGS ANALYSIS

USING 2023 SPOTIFY DATA

Calvin Gerard

Project Objective

A local studio musician and songwriter is looking to write a song that will perform well on Spotify. Achieving this will allow the artist to gain more visibility and career momentum if many users are streaming their music.

Therefore, my goal was to analyze aspects of the most streamed songs on Spotify in 2023 to find insights that I can use to advise the client.

Process

The “Most Streamed Spotify Songs 2023” dataset is available as this file on Kaggle. This dataset contains information on what the top streaming songs and artists are along with measured attributes of the music that I used for my analysis. The aforementioned attributes include danceability, energy, acousticness, and instrumentalness (all measured in percentages), which I determined were the clearest factors that can be analyzed to gain insights on the project objective. Additionally, the dataset contains information on the key, mode, and bpm that each of the top streamed songs are written in, which can also be used to advise the client on what key and mode to write in.

Process (cont.)

The Kaggle post containing the Spotify dataset defines the musical attributes in the data as the following:

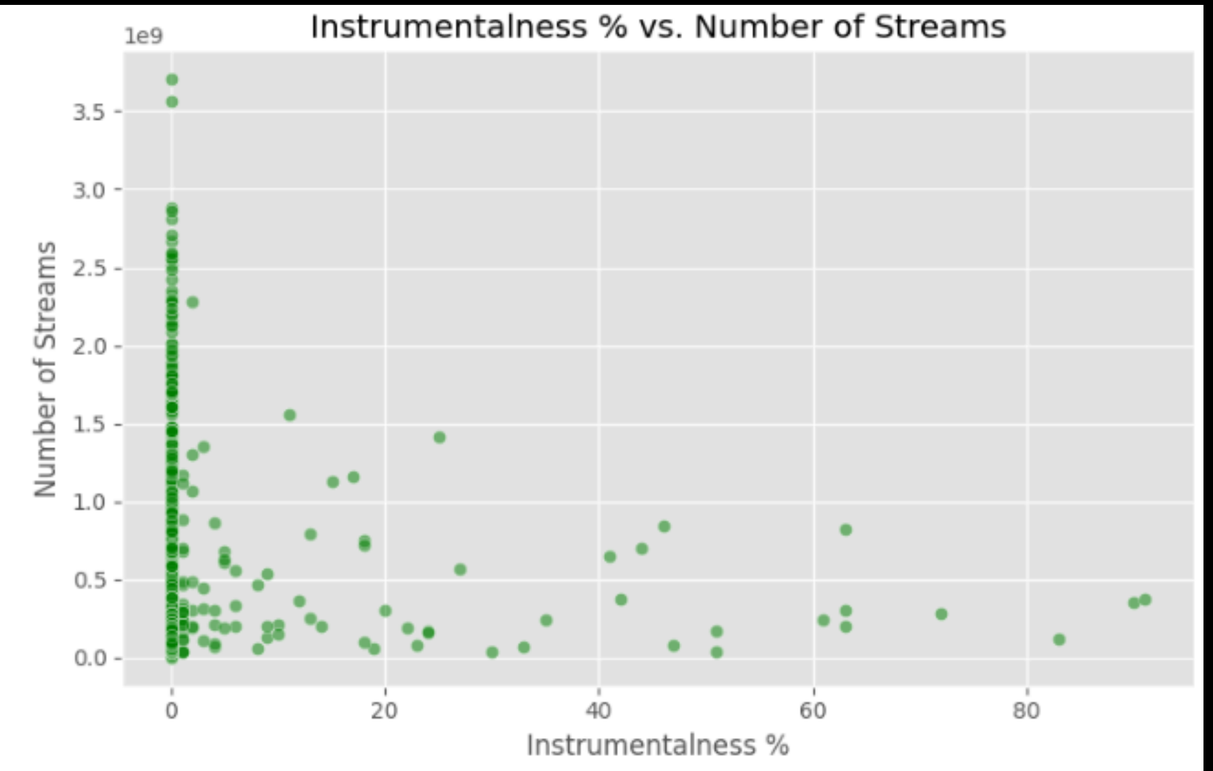
- **key: Key of the song**
- **mode: Mode of the song (major or minor)**
- **danceability_%: Percentage indicating how suitable the song is for dancing**
- **valence_%: Positivity of the song's musical content**
- **energy_%: Perceived energy level of the song**
- **acousticness_%: Amount of acoustic sound in the song**
- **instrumentalness_%: Amount of instrumental content in the song**
- **liveness_%: Presence of live performance elements**
- **speechiness_%: Amount of spoken words in the song**

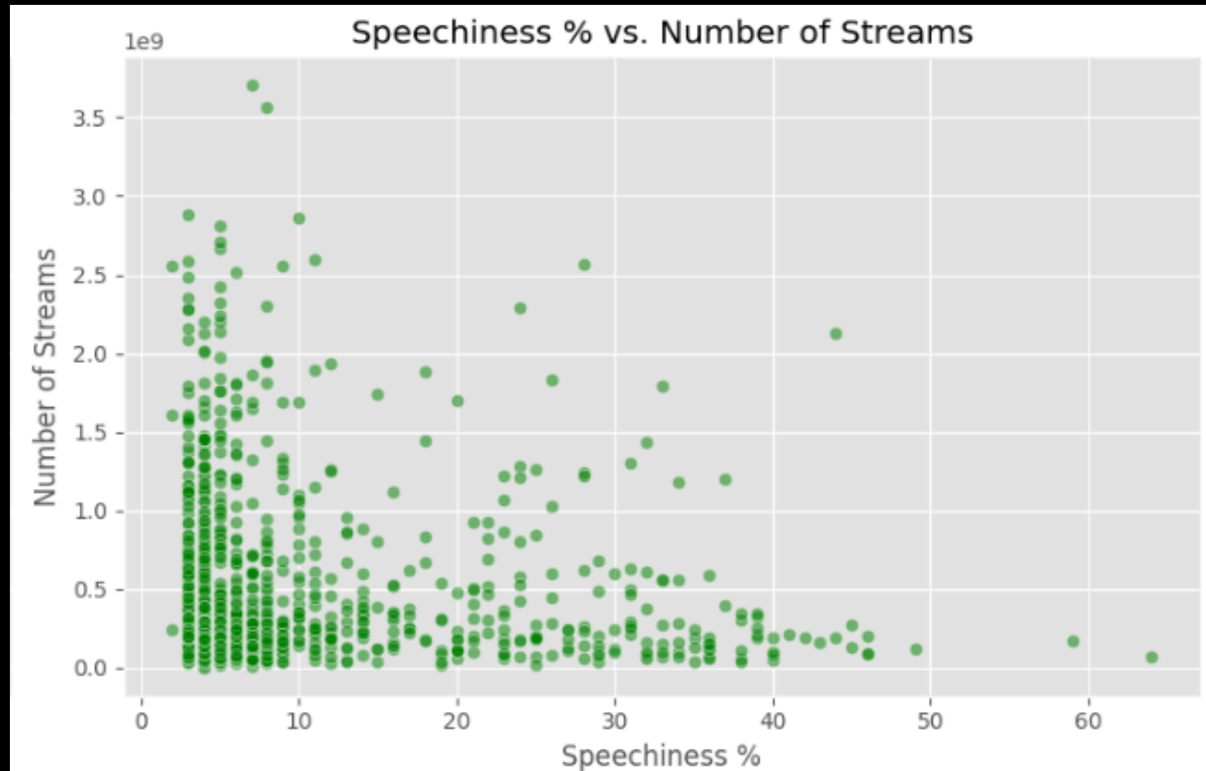
Process (cont.)

I included danceability, energy, instrumentalness, and speechiness as the four attributes measured by percentage that are the clearest defined and most applicable to the client. Liveness is not relevant since the recorded song will not be a live performance, while acousticness is not clearly defined enough (are “acoustic elements” the inclusion of acoustic instruments or are they measurable qualities of the recording environment?). Valence may be useful to analyze, but it is also unclear -does “positivity of musical content” include lyricism, positivity reflected in musical key, or both? Given that lyrics are subject to interpretation and we don’t know the method used to measure this data, instead I chose to advise the client about musical key, mode, and bpm, which is more clearly represented and the artist can use to make their own judgments regarding how those factors affect a song’s positivity.

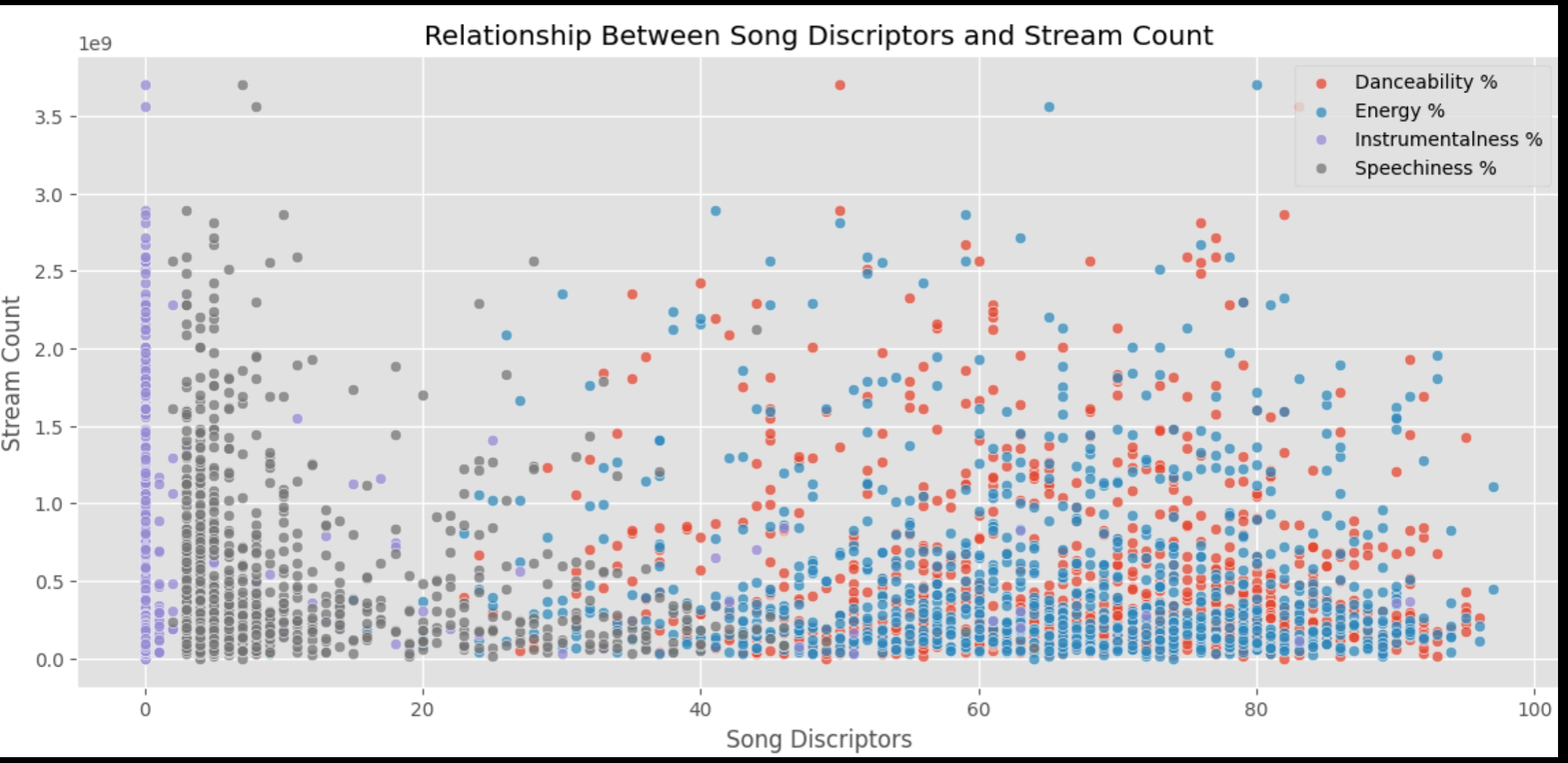
Instrumentalness vs Speechiness

Here are graphs comparing Instumentalness and Speechiness percentages individually against the number of streams a song has. Songs appear to get more streams with lower levels of both of these factors, especially with Instrumentalness.

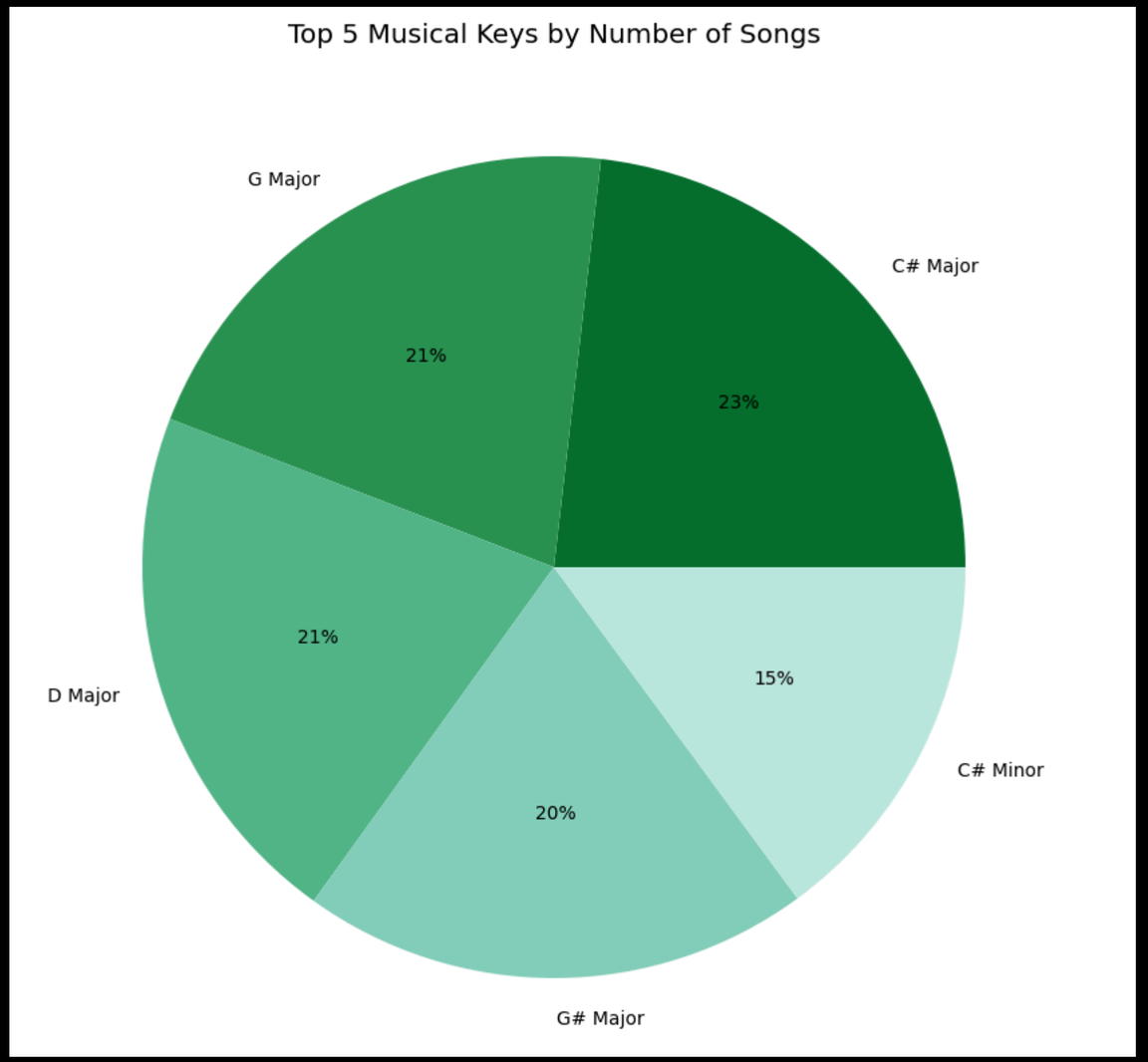




Song Discriptors and Stream Count



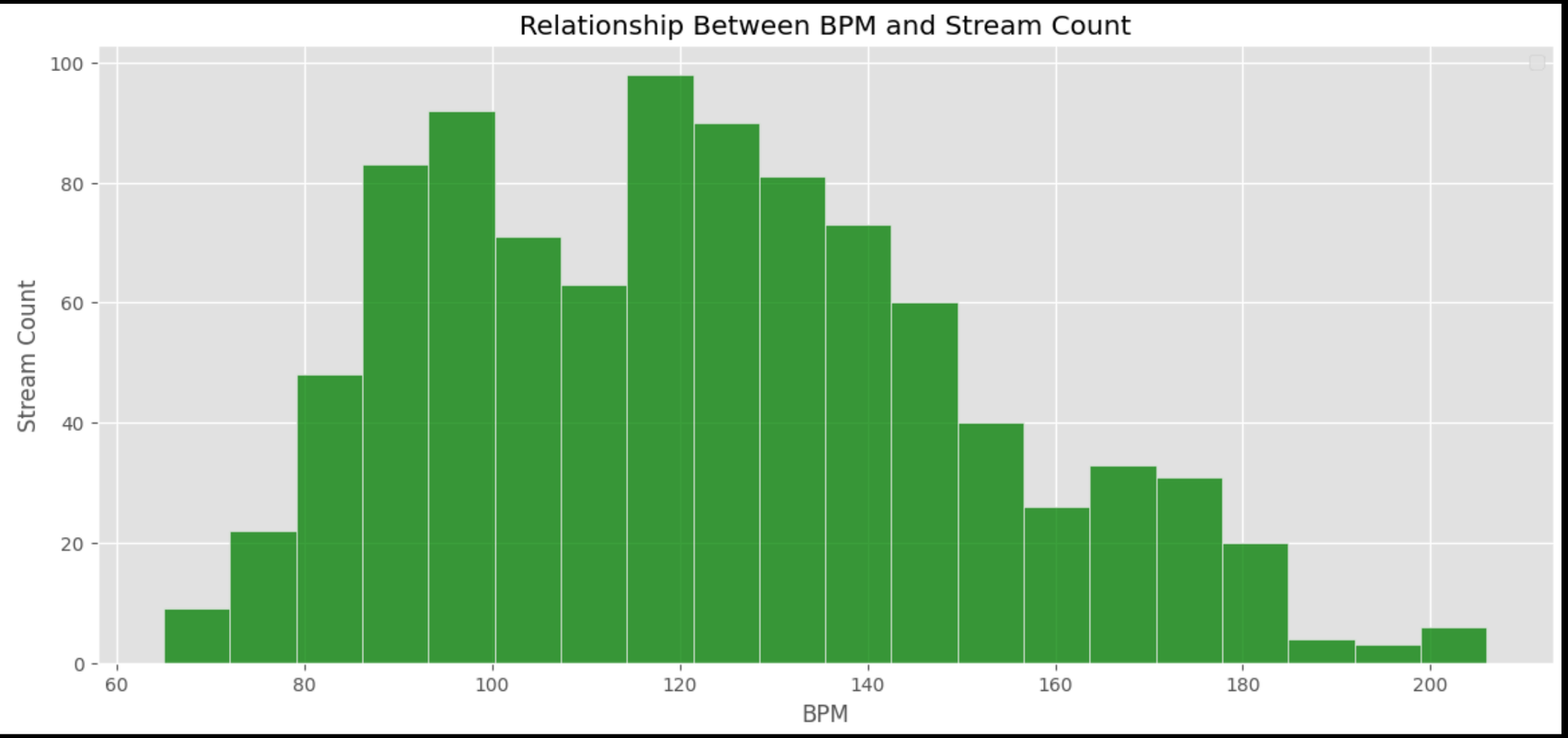
Instead of comparing variables one by one, I took the four aforementioned song descriptors recorded in percentages and put them in one graph to compare their relevance to stream count. Here we see that high Danceability and Energy are related to each other (shown by their many overlapping points) and correlated to high stream count in comparison to songs with low Danceability and Energy. We see the same results for Instrumentalness and Speechiness as the first graph, but here we see a better representation of how low percentages of these attributes relates to high stream count and high percentages of the other attributes.



Most Popular Keys & Modes

This graph shows us the top 5 musical keys & modes that are most often represented in the data. As you can see, the incremental percentage difference between each of these top five keys is not overly significant. Thus I advised the client to pick one of these five keys instead of a specific one, so the client could make a decision while keeping in mind their songwriting style and the effect of musical keys on valence that we covered previously.

Most Popular BPM



Finally, I made this graph to determine what the best BPM would be for the client to write a higly streamed song. To do this, I made a histogram showing the most popular BPM represented in the data. Given that there are multiple different BPMs that are highly popular according to the data, I would recommend the client write their song at somewhere around 120 BPM to capitalize on the most represented portion of the data. However, this is also open for the client to determine for themself regarding their own taste and what will match the other goals we have outlined here.

Conclusion

To reiterate, after analyzing the data, my advice to the client to achieve their goal of writing a highly streamed song is as follows:

- Write in one of the following keys: C# Major, G Major, D Major, G# Major, or C# Minor.
- Make sure the song is around 120 BPM.
- When writing and recording, aim for the song to be energetic and danceable, while avoiding emphasis on instrumentality and speechiness.

Improvements

The main factor that was not included in the dataset that would have been useful would be genre data. This was not collected by the poster of the dataset and collecting the data from Spotify was not within the scope of this project. Additionally, the song attributes data can be used to approximate genre, since genre can be somewhat subjective despite it often being labeled on streaming platforms. Nevertheless, if I were to do a project like this in the future, I would like to include genre data for clearer representation of the factors that the client wanted me to illuminate in this project.

THANK YOU