Spotify and YouTube Dashboard Visualization

Josh DeKeersgieter
The Ohio State University
Columbus, Ohio
dekeersgieter.2@osu.edu

Ryan Kline
The Ohio State University
Columbus, Ohio
dekeersgieter.2@osu.edu

Hrushik Chiluvuri
The Ohio State University
Columbus, Ohio
dekeersgieter.2@osu.edu

Abstract- We have created an interactive dashboard using Tableau to visualize data regarding song popularity and attributes on both Spotify and YouTube. Special interest is taken in examining which attributes could potentially influence popularity on Spotify, and whether anything correlates with success on both platforms. The dashboard itself accomplishes this by using a combination of more simplistic visualizations, namely a scatterplot where each point is an individual song, as well as more complex visualizations, namely treemaps and parallel coordinate plots. Finally, an interactive filter was added where subsets of artists can be selected, both to enable users to explore their personal favorite bands and ensure that the visualizations did not become too crowded. The end result is an interactive visualization tool that we believe is highly effective in enabling song attribute exploration and discovery.

Introduction- Our interactive data visualization dashboard offers a user-friendly and intuitive way to explore music attributes and compare personal taste with the broader music industry. Our tool allows users to interactively explore factors such as energy, danceability, and liveness, gaining insights into how these attributes impact song popularity. By visualizing these factors, enthusiasts, researchers, and industry professionals have a powerful tool to gain insight and influence decision making. Additionally, casual users are able to examine their own music preferences in comparison to the broader music landscape, enabling them to gain a unique perspective on their individual taste in music. There is also potential for them to see artists that produce similar music, with similar levels of energy and danceability, and find songs that could closely match their musical taste.

Our dashboard is designed to be visually appealing and informative, with an emphasis on user-friendliness and interactivity. We believe that our dashboard has the potential to enhance understanding of the factors that drive music popularity and provide valuable insights into personal music preferences in relation to the broader music landscape.

The specific goals of our dashboard can be summarized in the following elementary (e) and synoptic (s) tasks that we hoped to accomplish: (1e) Find the 10 most popular artists on Spotify; (2e) Determine the most viewed track on Youtube; (1s) Determine the relationship between various song success metrics such as Youtube views and spotify streams relative to each artist; (2s) Display popularity of each track and show extent to which each track contributed to the artist's popularity; (3s) Find the relationship between metrics such as energy, danceability, liveness, etc. to the selected groups of

artists in order to see how they may affect a song's success.

Related Work- Our first related work is "Predicting Music Popularity on Streaming Platforms" by Carlos Araujo, Marco Cristo, and Rafael Giusti. It explores how song attributes like danceability, energy, and loudness can impact a song's popularity over time, as well as predict if sudden spikes in popularity can lead to long-term popularity. They showed that these attributes can be successful in predicting popularity, which is why we decided to include them as important factors to visualize in our dashboard. Our second related work is "Song hit prediction: Predicting billboard hits using Spotify data" by Kai Middlebrook and Kian Sheik. It has a similar goal of predicting which songs will be chart-popping hits, using the Spotify Web API. The main takeaway from this paper is that Spotify is a very powerful tool for analyzing music on an objective level, since they create so many attributes to describe each song and can be used in a highly analytical way.

Implementation- The raw data can be loaded into Tableau; no pre-processing in a programming language (ie. Pandas in Python) is necessary. Tableau is capable of doing some basic data aggregation necessary for the plots. For example, in the parallel coordinate plot, the average number of Spotify Streams grouped by artist is calculated, and then those aggregated values are re-scaled to be between 0 and 1 before they are displayed.
To create the dashboard, each visualization is generated in its own tab, as is standard in Tableau. All of the visualizations have off-the-shelf "generators" that are used to generate the plots themselves, once they are given instructions on what they should visualize. Once the atomic visualizations are generated, a dashboard can be created that takes those atomic plots, combines them, and re-scales the plot size if necessary to make them fit together.

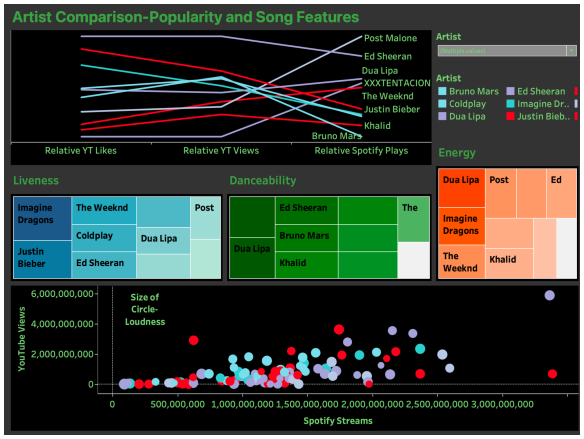
<u>Plots</u>

Parallel Coordinate Plot (1 plot):
- Multi-Variant (4) - Displays Relative Youtube Views, Relative Youtube Likes, and Relative Spotify Streams for each selected Artist

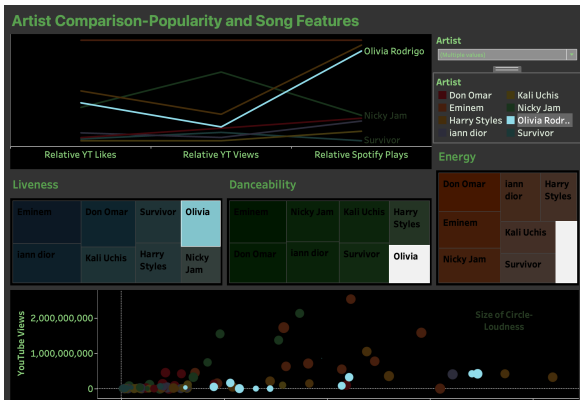
Treemaps (3 plots):
- Double-Variant - Each treemap shows a new metric grouped by Artist. The 3 metrics are Energy, Danceability, Liveness

Scatterplot (1 plot):
- Multi-Variant (5) - X and Y axes are Youtube Views and Spotify Streams respectively. Each pixel represents a different Track. Pixels are colored by Artist and sized by Loudness

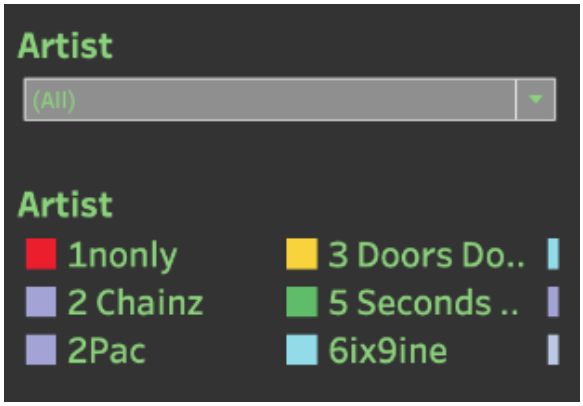
Results- Please see below for a set of pictures demonstrating the dashboard created, the individual plots made available, and the filtering & highlighting features made available when working in Tableau.



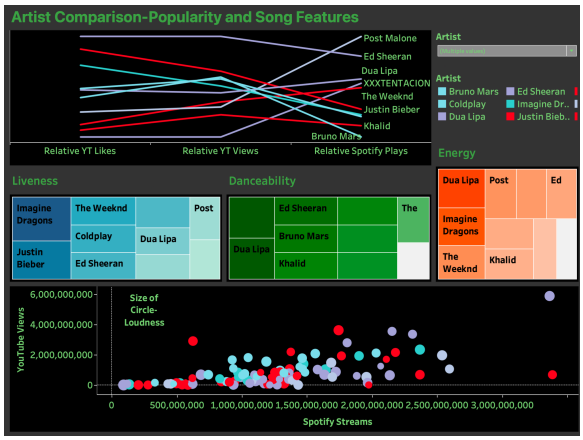
The Overall Look of The Dashboard



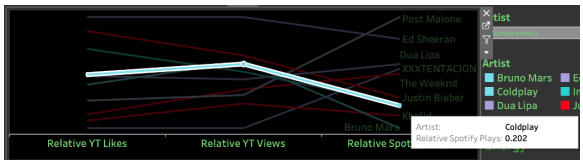
Showcasing linked highlighting Olivia Rodrigo's Data

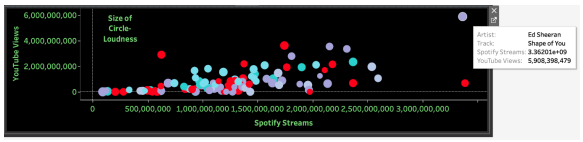


The Artist Selection and Color Guide For Artist Selection, Also Where Selection of Artists is Completed

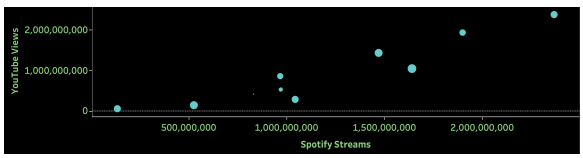


View of The Dashboard Comparing The Top 10 Most-Played Artists on Spotify





The 2 Figures above Demonstrate The Tooltip Feature of The Dashboard, Showcasing The Numeric Values of The Artist (above) and Each Song (Below)



Demonstrating One Artist's Song Popularity

Demonstrating The View With All Artists Selected

The elementary tasks can be answered without much trouble. The most viewed track on YouTube is the highest dot on the y-axis of the Scatterplot (Despacito for those wondering). The 10 most streamed artists on Spotify are the 10 highest values on the Relative Spotify Streams portion of the parallel coordinate plot. They are sub-selected and shown directly to the left for convenience.

The synoptic tasks are a little more complex. Tracks that stand out from the crowd on the scatterplot are very likely candidates to contribute greatly to an artists' success. For example, Despacito is a likely contributor to Luis Fonsi's success, and Gangman Style is another outlier that made Psy very well known internationally.

The metrics of energy, danceability, and liveness can also be matched to popular bands. High levels of energy seem to correlate with very successful metal or rock bands, such as Slayer, Five Finger Death Punch, or Sum 41. Artists with a high Danceability value include Megan Thee Stallion and Pharrell, who make music commonly played in fitness centers or clubs / bars.

Finally, in determining a relationship between success on Spotify and YouTube, we found that *in general*, there is about a 1-to-1 correlation between Spotify streams and YouTube views, meaning that an average song that gets about 1 million streams also gets about 1 million views. However, it is much easier for a song to get above 2 million Youtube views than it is for a Spotify song to get more than 2 million streams. We believe this is because YouTube

is more available internationally than Spotify is, and accessible by a wider audience. There are also some songs that may be impacted by the time of their release- Justin Bieber's *Baby* has an incredible number of YouTube views because it was released before Spotify was popular, but songs from his newest album are some of the most-streamed Spotify songs in our database.

Future Directions- Our dashboard and analysis make use of the various song success metrics such as youtube views and likes and spotify streams, however, we used only a subset of the other various attributes that describe each song. Out of 20 or so columns that describe each song, only danceability, energy, liveness, and loudness are used in the treemaps and scatterplot shown in the *implementation* section. In the future, we would like to extend our scatterplot to see how scores of acousticness, valence, etc. effect the popularity of songs and their corresponding artists.

The other method of improvement that our team plans to undertake is enabling additional filters for our dashboard. As of its current state, the filters only allow viewers to pick their favorite artists, but there are a plethora of other useful filters that could help answer a variety of other questions for users. For example, a filter to display whether an official youtube music video was released could better assist viewers in answering the question of the extent of the impact that an official music video may have in measuring a track's youtube success. Other filter options such as sliders to view only songs with x amounts of danceability, energy, valence, etc. could also provide useful depending on what kinds of questions users hope to answer.

References-

Middlebrook, Kai, and Kian Sheik. "Song hit prediction: Predicting billboard hits using Spotify data." arXiv preprint arXiv:1908.08609 (2019).

Araujo, Carlos & Cristo, Marco & Giusti, Rafael. (2019). "Predicting Music Popularity on Streaming Platforms". 141-148. 10.5753/sbcm.2019.10436.

Our dataset is available on Kaggle, at https://www.kaggle.com/datasets/salvatorerastelli/spotify-and-youtube.