

What is your dataset? Why?

<https://www.kaggle.com/datasets/salvatorerastelli/spotify-and-youtube>

Most classmates use Spotify and/or YouTube. Our group is curious about music statistics on both platforms and how they correlate with one another.

Question to Research:

Q1: What features are the most prevalent in popular songs (top 10%) on YouTube vs Spotify? (Matthew)

Q2: For popular songs (with a lot of streams), what is the distribution of the feelings of the song? (utilizing quartiles) (Marc)

Q3: Do the attributes of a song play a role in the amount of streams across platforms? (Chase).

Q4: Are high liveness songs generally more popular on YouTube, and are low liveness songs generally more popular on Spotify? (Khalil)

Visualization Proposals

Q1: Various Visualizations based on the categories. (Grouped Bar Chart and Leaderboard).

Q2: Box Plot / Quartiles

Q3: Heatmap with ANOVA and T-test

Q4: Scatter Plot (liveness vs. popularity by platform – side-by-side)

Inspiration - other code/articles similar to your dataset

Recommend code tab on Kaggle datasets

<https://www.kaggle.com/code/aleksapet98/youtube-spotify-eda-and-simple-regression>

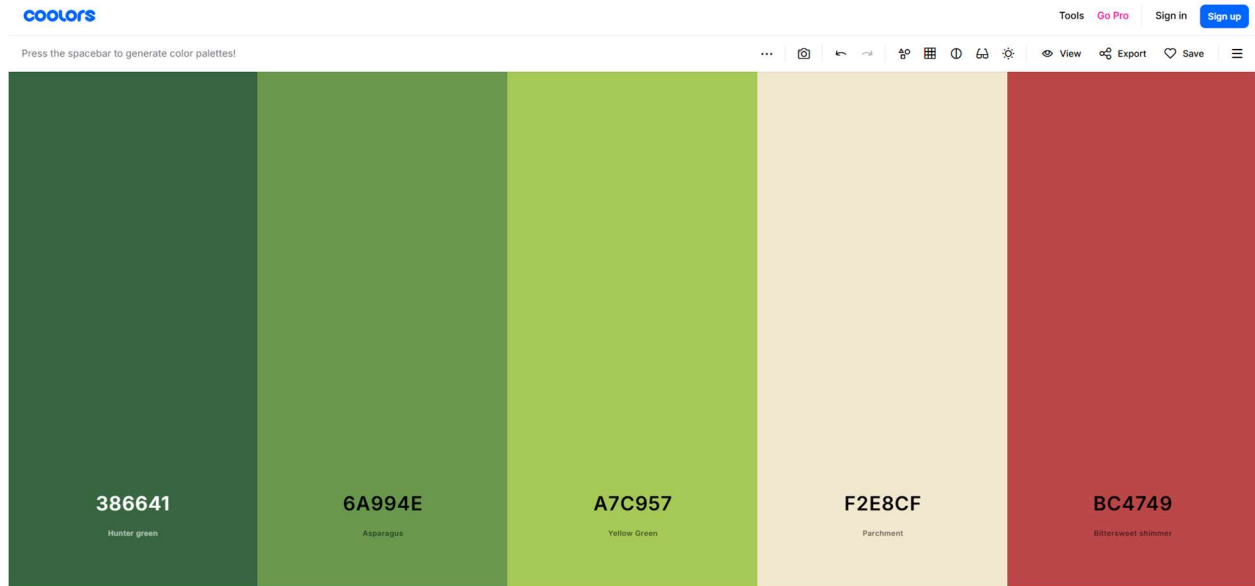
<https://www.kaggle.com/code/themeeemul/youtube-spotify-eda-and-views-prediction>

What are you regressing?

Q4 scatterplot, linear regression to see correlation between liveness and popularity by platform.

Color palette

We utilized the [Coolers.co](https://coolers.co) tool for generating palette to fit our needs for the project. Since YouTube and Spotify utilize red and green for their color palettes, we will do the same. See the image below for reference.



Roles & Responsibilities

Matthew	Chase	Khalil	Marc
GitHub Manager	Final Report Writer	QA Lead on Data Viz -- Plots	QA on Data Viz -- Code
Question 1 Lead	Question 3 Lead	Question 4 Lead	Question 2 Lead

GitHub Link

<https://github.com/silvertimmy26/project1-group4>