Group 4 - YouTube vs Spotify EDA

Matthew Adent, Marcanthony Solorzano, Chase Mueller, Khalil Locke

Data Cleaning/Feature Engineering

- Dropped certain columns from the dataset:
  - Unnamed: Obvious reasons; no data relevance.
  - Url_spotify: Link to the song provides no value.
  - Uri: No need for an embedded link in the dataset.
  - Url_youtube: … do we need to say it again
  - Description: Strings are not feasible for data analysis.

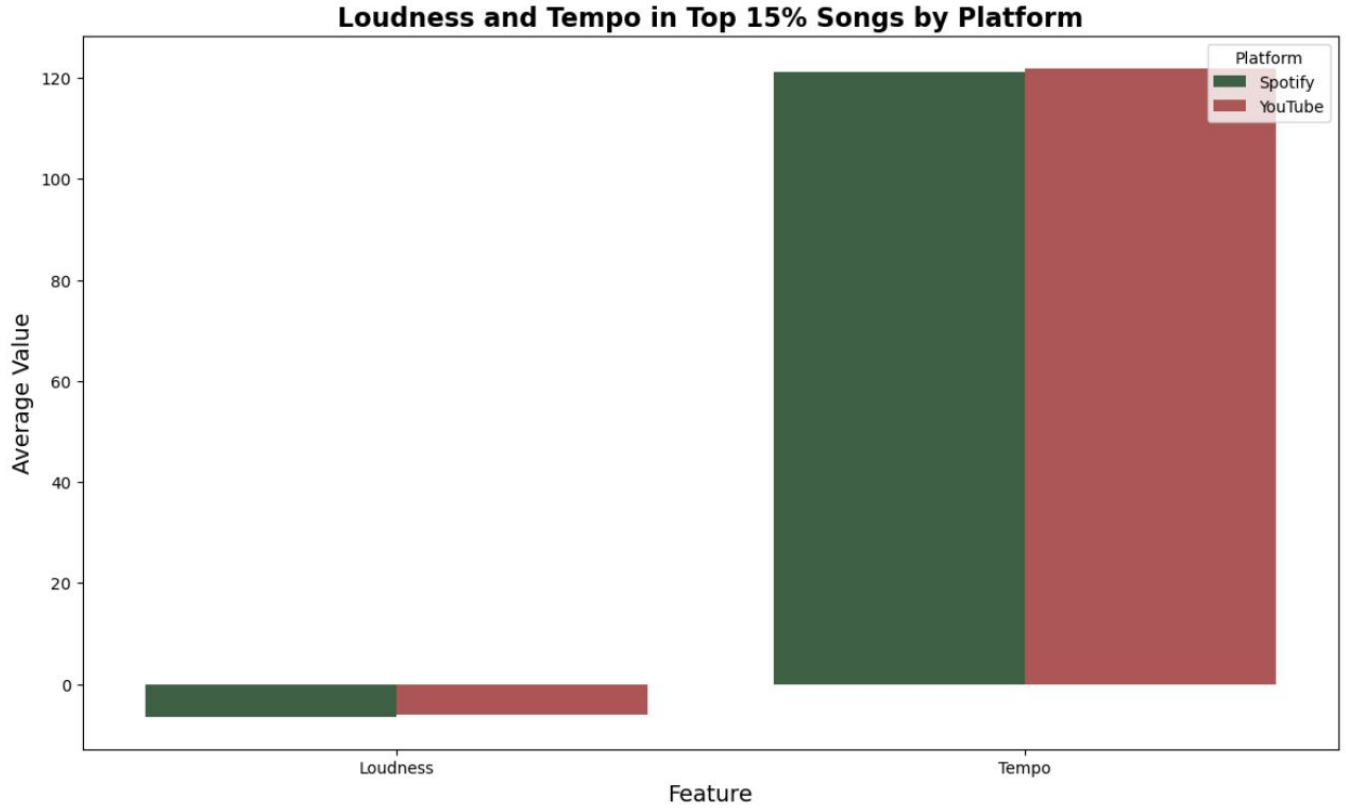
The “why” behind our Data Set.

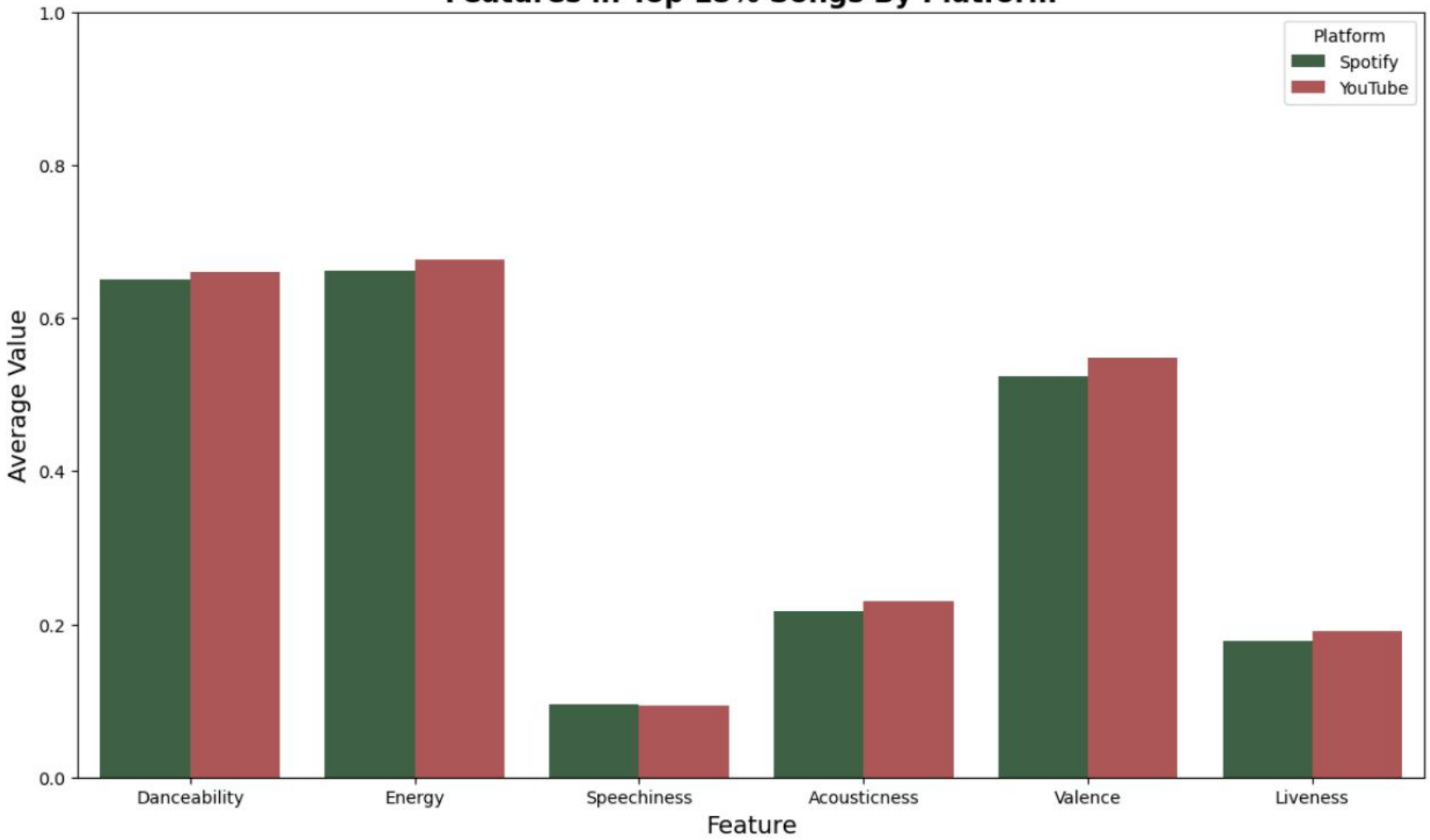
- Music is something that a majority of people listen to on a daily basis, whether they know it or not.
- We wanted to explore the difference between YouTube users and Spotify Users.
- Understanding the emotional tones and popularity trends on these platforms helps reveal how digital streaming influences global music trends.
- Both platforms significantly influence the music industry's revenue streams. Insights from this dataset can help optimize strategies for artists and record labels.

Key Terms:

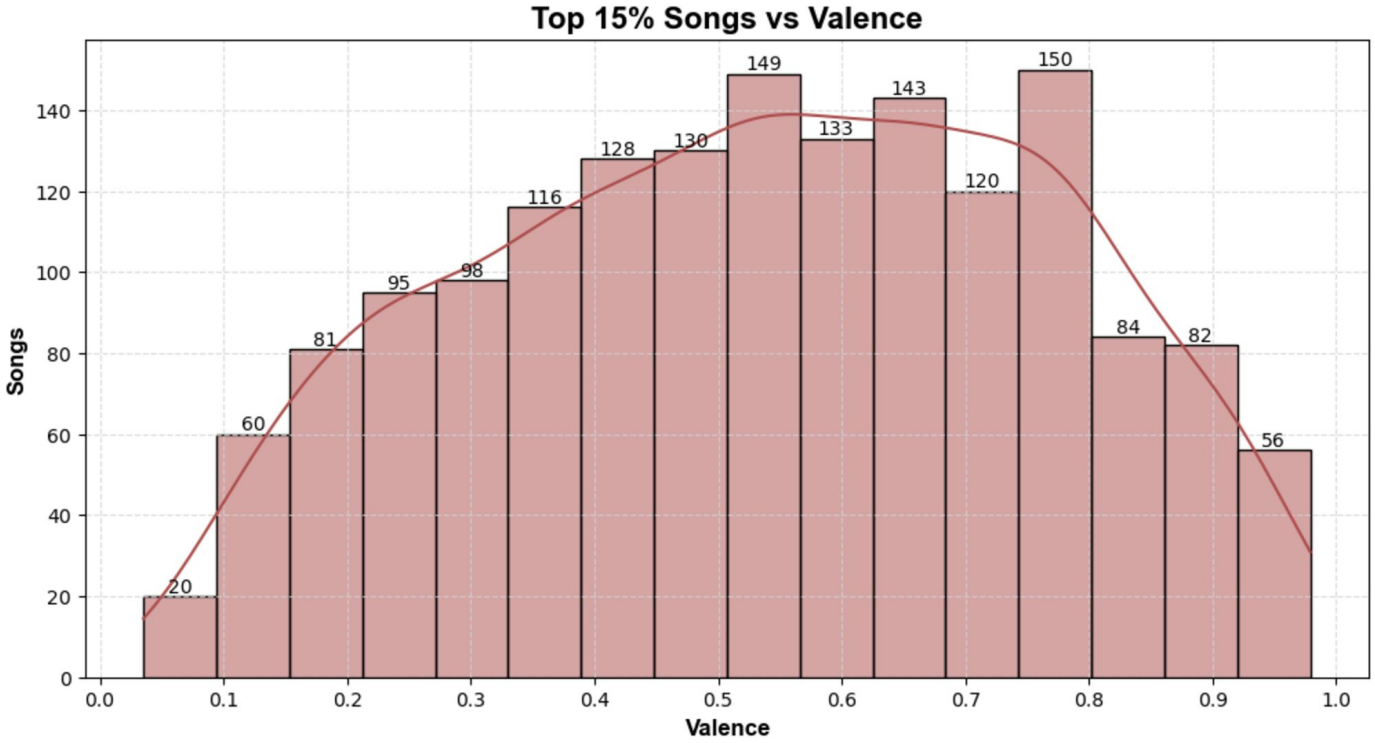
- **Danceability:** Measures how suitable a track is for dancing (0.0 = least, 1.0 = most).
- **Energy:** Measures intensity and activity (0.0–1.0). High-energy tracks feel fast and loud, while low-energy tracks feel calm.
- **Loudness:** Average loudness of a track in decibels (dB), typically ranging from -60 to 0 dB.
- **Speechiness:** Detects spoken content. Values >0.66 = mostly speech; 0.33–0.66 = mixed; <0.33 = mostly music.
- **Acousticness:** Confidence (0.0–1.0) that a track is acoustic (1.0 = highly acoustic).
- **Instrumentalness:** Predicts if a track has no vocals. Values >0.5 likely indicate instrumental tracks.
- **Liveness:** Measures audience presence; >0.8 suggests a live performance.
- **Valence:** Measures musical positivity (0.0 = negative, 1.0 = positive).

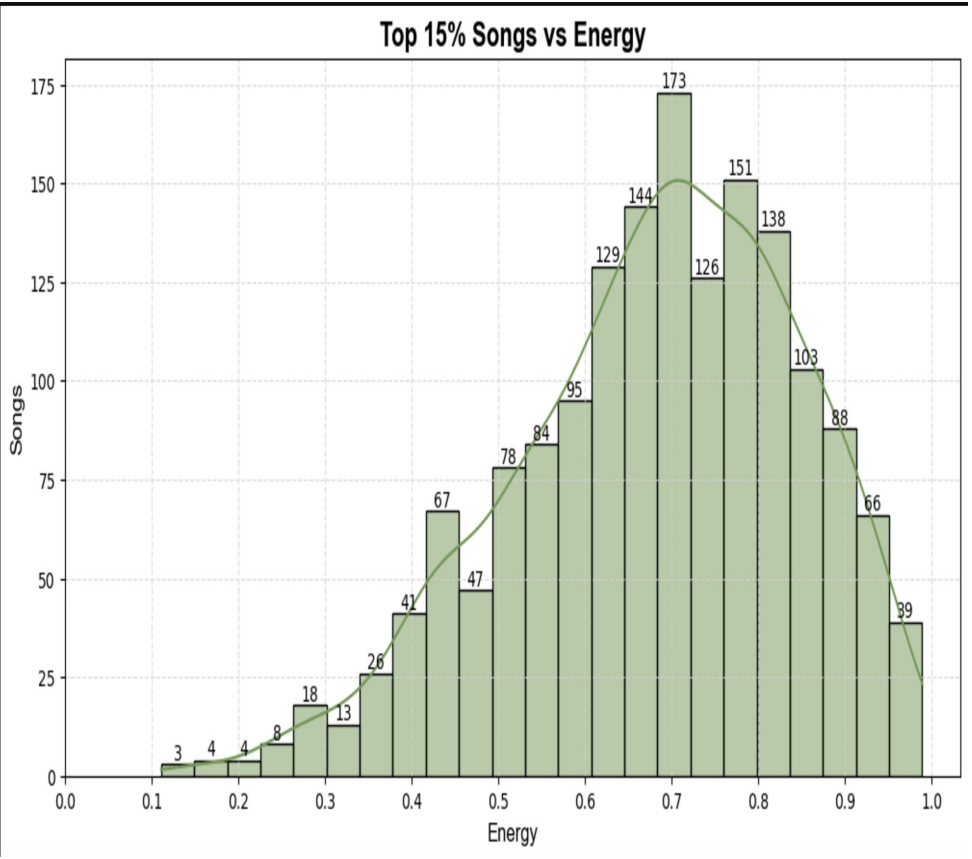
Q1: How do the features differ between the top 15% of songs on Youtube and Spotify?

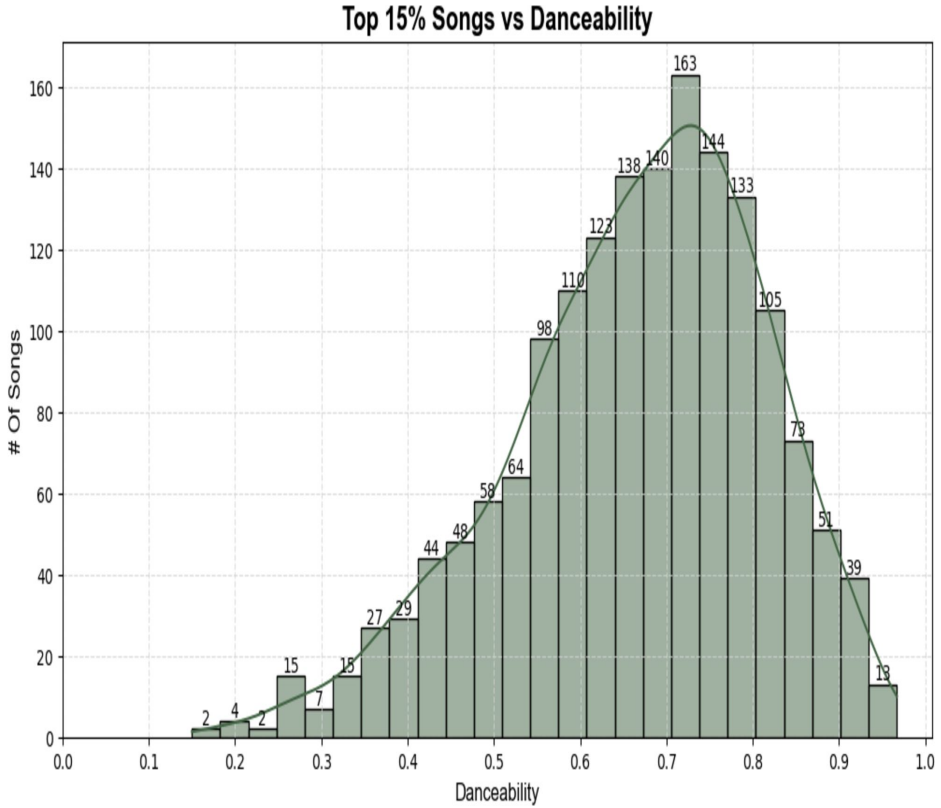


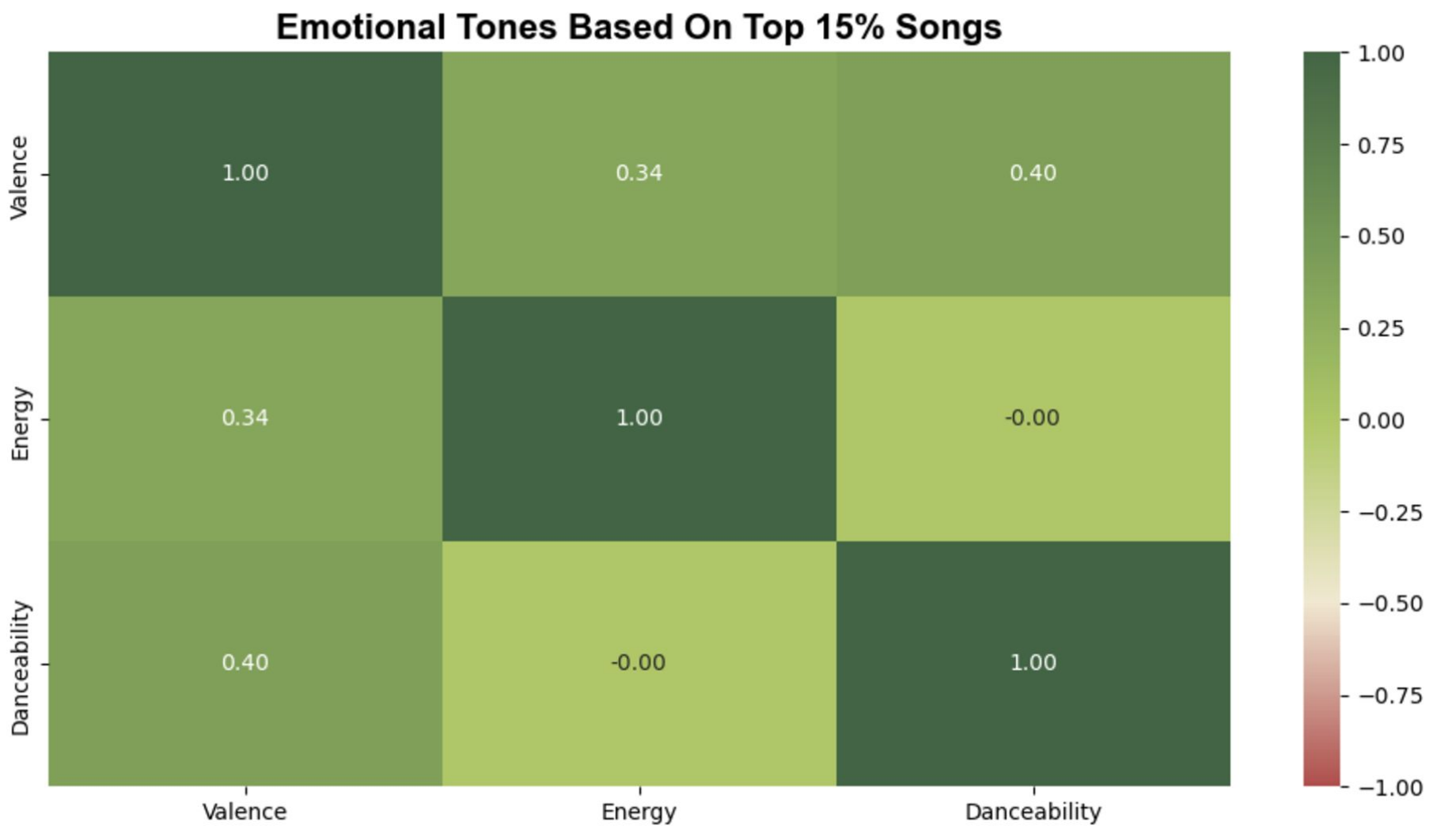

Features in Top 15% Songs By Platform
Platform
Spotify
YouTube
Average Value
1.0
0.8
0.6
0.4
0.2
0.0
Danceability
Energy
Speechiness
Acousticness
Valence
Liveness
Feature

Q2: How are the emotional tones of top-streamed and most-viewed songs distributed?

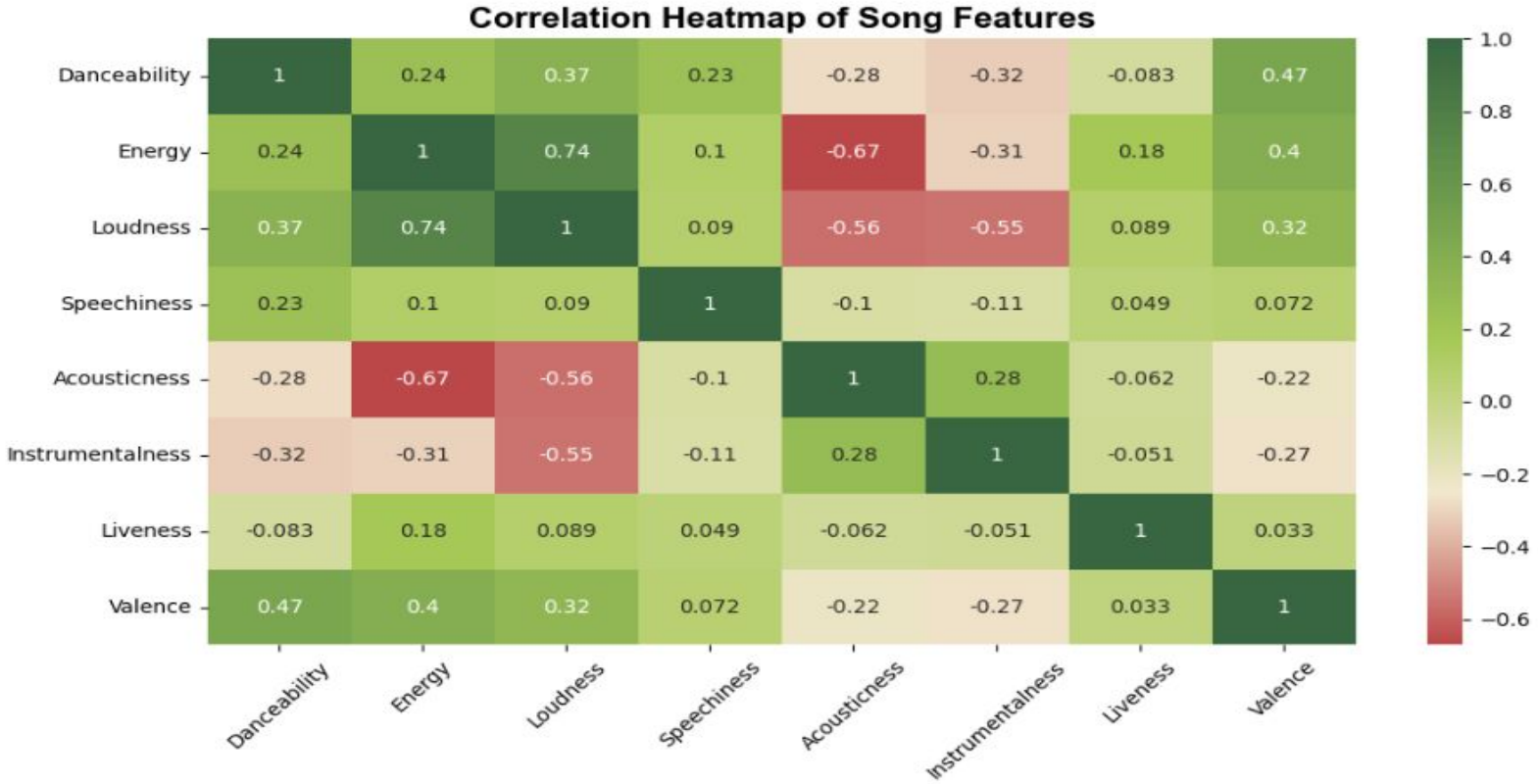



Top 15% Songs vs Energy
Songs
Energy
0
25
50
75
100
125
150
175
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
3
4
4
8
18
13
26
41
67
47
78
84
95
129
144
173
126
151
138
103
88
66
39

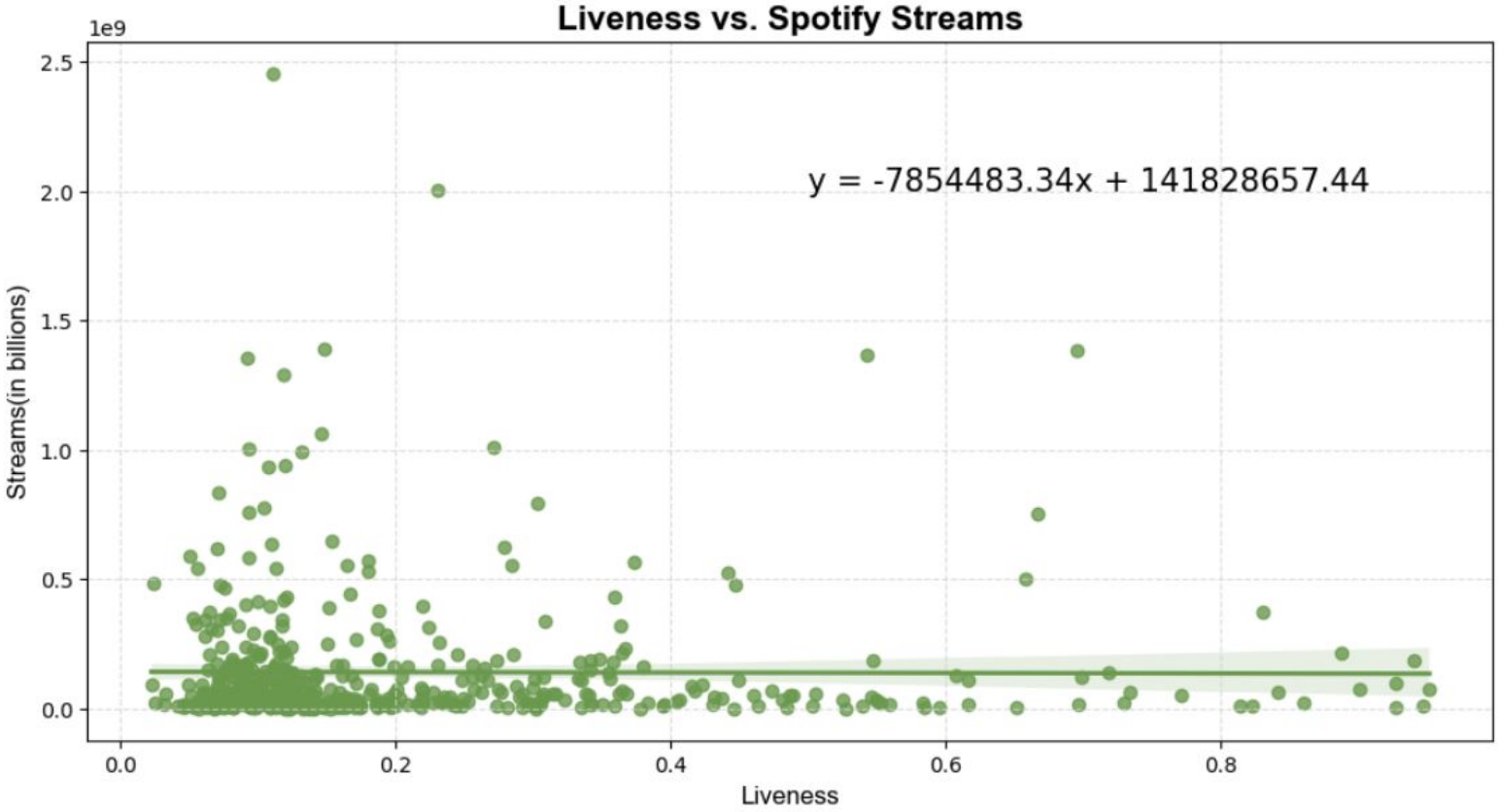

Top 15% Songs vs Danceability
Of Songs
Danceability
0
20
40
60
80
100
120
140
160
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
2
4
2
15
7
15
27
29
44
48
58
64
98
110
123
138
140
163
144
133
105
73
51
39
13

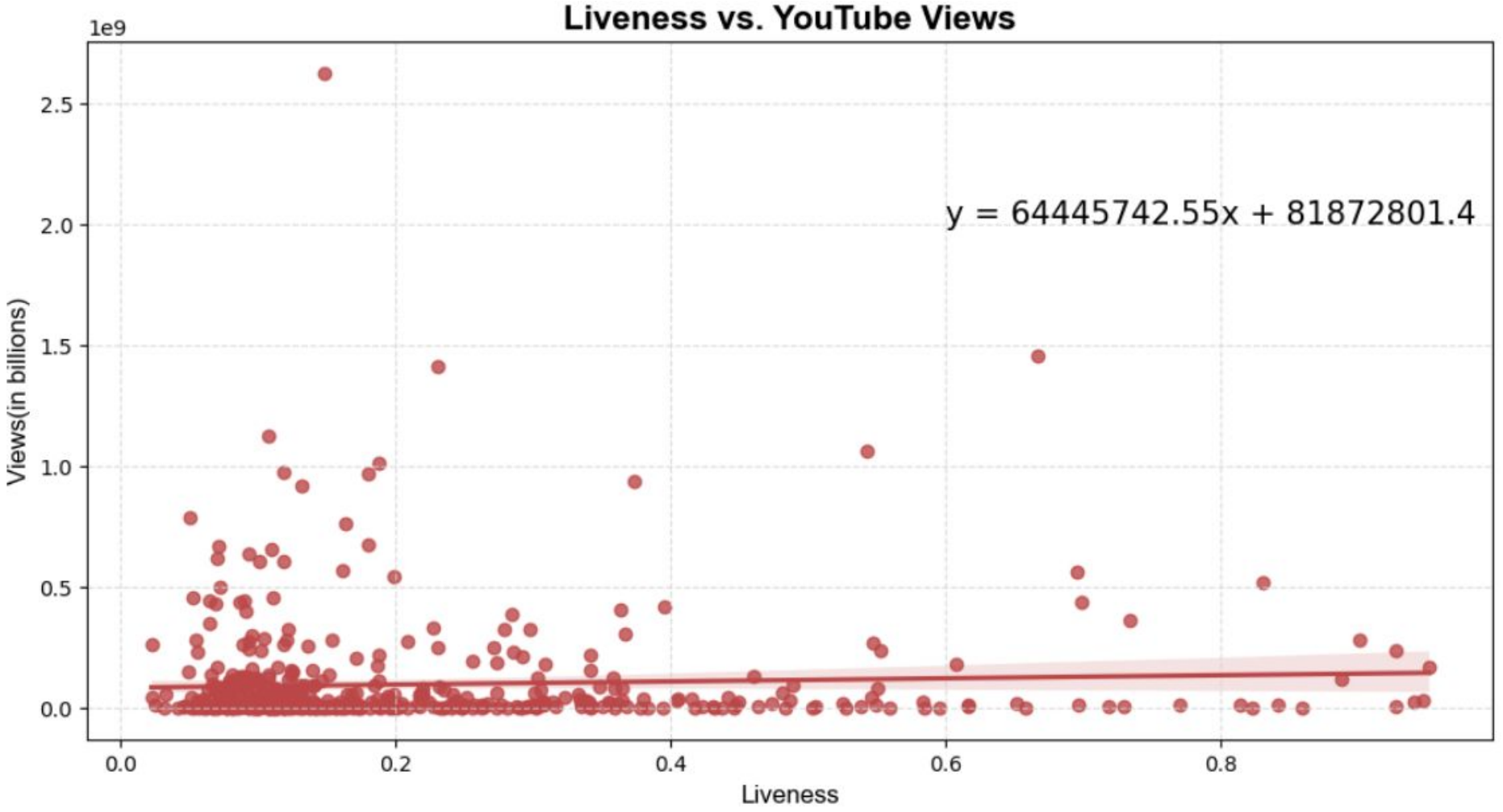

Emotional Tones Based On Top 15% Songs
Valence
Energy
Danceability
1.00
0.34
0.40
0.34
1.00
-0.00
0.40
-0.00
1.00
Valence
Energy
Danceability
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

Q3: What kind of correlation do the features of a song have with each other?



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




Liveness vs. YouTube Views
1e9
2.5
2.0
1.5
1.0
0.5
0.0
Views(in billions)
y = 64445742.55x + 81872801.4
0.0
0.2
0.4
0.6
0.8
Liveness

Call to Action

- Several trends can be noted from our dataset that artists can use to maximize their engagement across platforms:
  - For more live, energetic, or high-energy music/audio -- Youtube is the place to go.
  - For more relaxed and casual listening -- Spotify will be your best platform.
  - Perhaps this is what upcoming artists might consider when crafting their next big hit.

Bias and Limitations

- The dataset does not account for users that use both platforms when it comes to views and counts.
- The dataset does not provide a genre for each song, so there was no way to analyze the datasets that way.
- The dataset does not include timestamps for when views or streams occurred, making it impossible to analyze trends over time.
- The dataset does not specify the geographic regions where the songs are most popular, which may skew the results toward regions with high user activity on Spotify or YouTube.

Future Work

- Introduce dates to analyze how emotional tones and streaming/viewing patterns evolve over time. This could highlight seasonal or event-driven trends.
- Explore user-specific metadata, such as age, location, or gender, to examine how different demographics engage with music and whether emotional tones resonate differently across groups.
- Build a predictive model using features like energy, danceability, and valence to forecast a song's likelihood of achieving high engagement on YouTube and Spotify.

Works Cited

1. Pandas Development Team. *Pandas Documentation*. https://pandas.pydata.org/docs/reference/frame.html.
2. "Python | Pandas DataFrame." *GeeksforGeeks*. https://www.geeksforgeeks.org/python-pandas-dataframe/.
3. OpenAI. "ChatGPT." OpenAI, https://openai.com/chat/.
4. Module 5 & 6, Data Analytics Bootcamp.

Questions?