Data Visualization on most streamed spotify songs 2024

Siddharth Menon
IIIT Bangalore
Bengaluru, India
Siddharth.Menon@iiitb.ac.in

Shreyank Bhat
IIIT Bangalore
Bengaluru, India
Shreyank.Bhat@iiitb.ac.in

Vrajnandak Nangunoori
IIIT Bangalore
Bengaluru, India
Vrajnandak.Nangunoori@iiitb.ac.in

***Abstract*—This document contains the visual analytics workflow in visualizing the kaggle dataset ”Most streamed Spotify Songs 2024”, incorporating additional datasets.**

I. Datasets Used:

1) Most Streamed Spotify Songs in 2024
2) Spotify Tracks Genre

Since the date cutoff for *Spotify Tracks Genre* dataset is 2022, even the *Most Streamed Spotify Songs in 2024* dataset had to be cut down to only have songs from 2022 or earlier, to facilitate a merge of the two datasets.

The following two workflows were aimed primarily at determining how much of a correlation the technical characteristics of a song has with its popularity, and how its popularity is affected by it. This also serves as the justification for the choice of secondary dataset.

A merge was performed between these datasets, using a string similarity function to compare names in the string columns.

II. Tasks

The objective for the analysis is to be able to observe and figure out the factors that contribute to the success or failure of songs and artists on streaming platforms. The factors considered include:

- Song information. The details include track information, various things like danceability, energy, loudness, liveness, tempo, genre etc.
- Artist information. The genres associated with the artists etc.

III. Visual Analytics Workflow

A. Workflow 1: Artist Characteristics and popularity

*1) **Iteration 1: Platform Relevance**:* For much of this analysis, we will be focusing only on the most relevant music streaming platforms. The ’relevance’ of a platform is influenced mainly by two factors:

- What are the total number of streams/views received on a platform?
- Can the popularity of songs on one platform be directly attributed to its overall popularity which is reflected in other platforms too? A good example for this is, songs that do well on TikTok aren’t necessarily popular music on other platform, and their popularity is attributed to their performance on one platform.

Identify applicable funding agency here. If none, delete this.

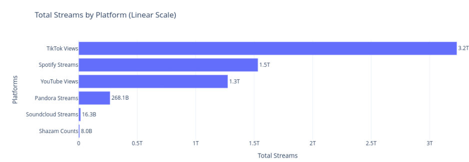

Fig. 1. Total Streams by Platform

Data

The data for loop 1 is a subset of the assignment 1 dataset alone. It includes:

- Track Name
- Spotify Streams
- YouTube Views
- TikTok Views
- SoundCloud Streams
- Pandora Streams
- Shazam Counts

Visualisation

Three visualisations were created with the above data. Among the two metrics listed earlier for gauging the ’relevance’ of a platform, for the first metric we have a histogram with the total number of streams on each platform. We see that TikTok, Spotify and YouTube far outperform any other platforms in figure 1.

As for the second metric that we used, i.e. the relevance of the platform, we used a sunburst chart to check how the top ten songs by total streams perform on all of the platforms. If we find that certain platforms are key contributors to the overall streams of a song, we can take this as a sign of their relevance, as their streams on those platforms decide the popularity of the song.

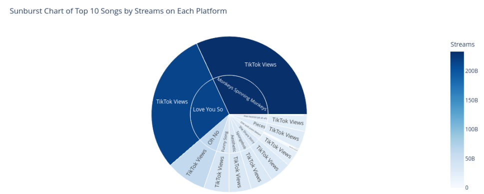

Fig. 2. Sunburst Chart: Top 10 songs by streams across all platforms

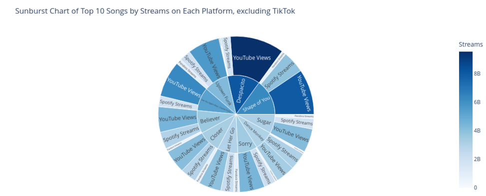

Fig. 4. Sunburst Chart: Top 10 songs calculated without TikTok Views

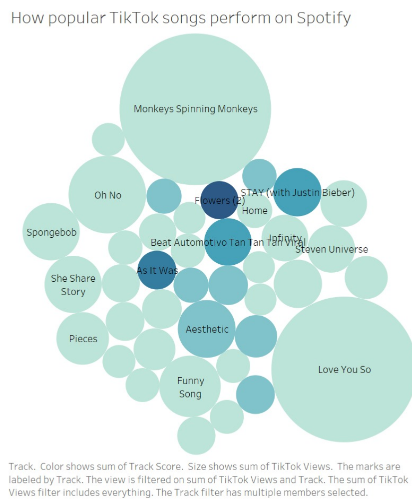

Fig. 3. Track Score vs TikTok popularity. Track Score (indicated by the colour channel), is a function of the Spotify Popularity of the song. The songs with the highest TikTok popularity aren't close to as popular on Spotify.

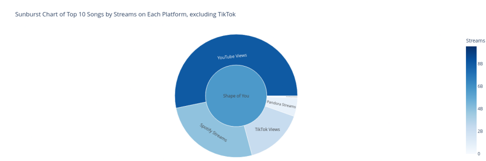

Fig. 5. Sunburst Chart: Platform relevance for a single song, Shape og You

Total streams were initially calculated by adding TikTok Streams as well. We quickly ran into a problem, as we found that TikTok seemed to far outweigh other platforms, and the top 10 songs were dominated by songs with many streams on TikTok but very few on the other platforms.
This is consistent with our observations from assignment 1, where we found that songs that do really well on TikTok do not do particularly well on Spotify and YouTube, possible because of the kind of music being used in TikTok, with their purpose often being only to complement the actual visual content. On the other hand, songs on Spotify and music videos on YouTube offer the music itself as their primary content. This can be seen in figure 3.
Therefore, for a more fair reflection, we also did the same by calculating total streams without TikTok views, to also study the class of songs which are popular, but due to TikTok's dominance, it gets drowned out in total streams.
Figures 4 and 5 show that after TikTok views are removed in the calculation of total streams, other platforms find better representation. It is worth noting that TikTok is included in the visualisation though, showing that songs that are popular on other platforms are popular on TikTok too, although it wasn't true for the converse.
Spotify and YouTube were the two best performing songs in the new visualisation, and this is seen in the analysis of a single artist, i.e. Ed Sheeran, when we drill down.

Knowledge

TikTok, Spotify and YouTube are the most popular platforms, that is backed by sufficient statistical evidence.

*2) **Iteration 2: Artist popularity analysis**:* In this analysis we will be studying the popularity of artists across different streaming platforms. Our objective in this section are the following:

- How well have the top artists performed in the streaming platforms as compared to each other?
- Contribution of streams of the top 60 artists sorted on Total Streams
- Which of the platforms turn out to be most popular for the top 60 artists.
- Do some artists have a monopoly over a platform?

Data:

The data for iteration 2 is taken from the 'Most Streamed Spotify Songs 2024' which requires the following columns:

- Artist
- Youtube Views
- TikTok Views
- Spotify Streams
- And an additional column that takes the aggregate of the streams in each of these streaming platforms for each row 'Total Streams'
- Dataframe for the top 60 songs were taken sorted on total streams.

Visualisations:

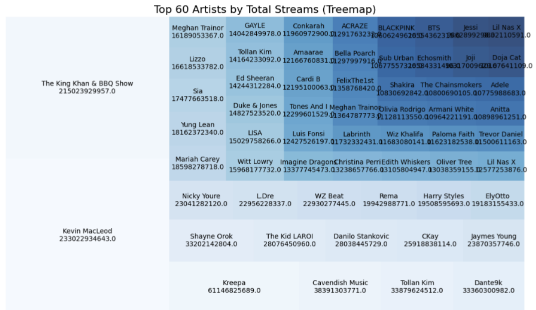

Fig. 6. Tree Map for top 60 artists

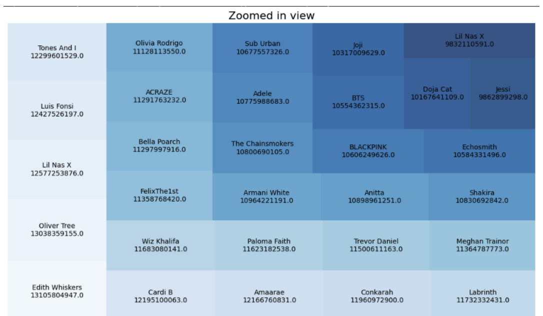

Fig. 7. Tree map for bottom half of top 60 artists(Zoomed in view)

We have plotted the following visualisations for this loop

- To begin the analysis, a general Tree Map was created to provide an overview of the streaming success of various artists. This initial visualization, highlights the top 60 artists based on their total streams, depicted in Figure 6. Figure 7 shows the bottom half of the top 60 artists for a more clear view. By offering a clear comparison of streaming counts, this plot serves as an introductory perspective on how certain artists have achieved higher streaming numbers relative to others.

- We aimed to identify artists who demonstrated a monopoly over a particular platform and those who contributed significantly across multiple platforms. To analyze this, we aggregated the data for the top 60 artists and divided them into batches of 20. For each batch, we calculated the average number of streams across the platforms under consideration: YouTube Views, Spotify Streams, and TikTok Views. Using these averages, we plotted a pie chart for each batch, which visualizes the proportional contribution of each platform to the total streams for that group of artists.
  The pie charts provide an initial understanding of how streams are distributed across platforms. By analyzing these distributions, we can assess whether certain platforms dominate specific batches or whether there is a balanced contribution from multiple platforms. For instance, the presence of a dominant platform in the pie chart indicates that a particular group of artists relies heavily on that platform for their streams. On the other hand, a more balanced distribution suggests that the artists have a diverse audience spread across multiple platforms.
  This analysis serves as a foundational step to decide how to further investigate these patterns.Figure 8 - Figure 10 shows the visualizations.

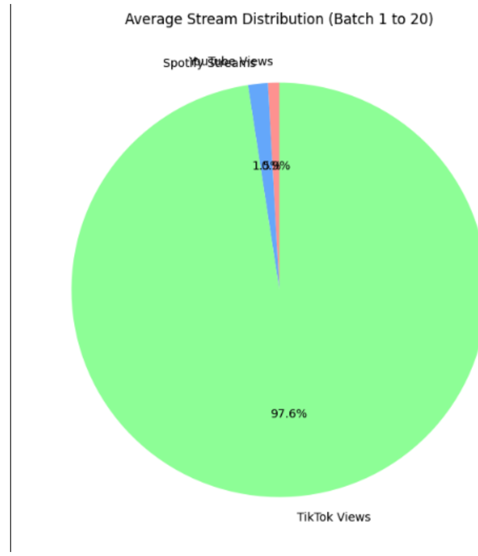

Fig. 8. Pie chart for artists ranked 1-20

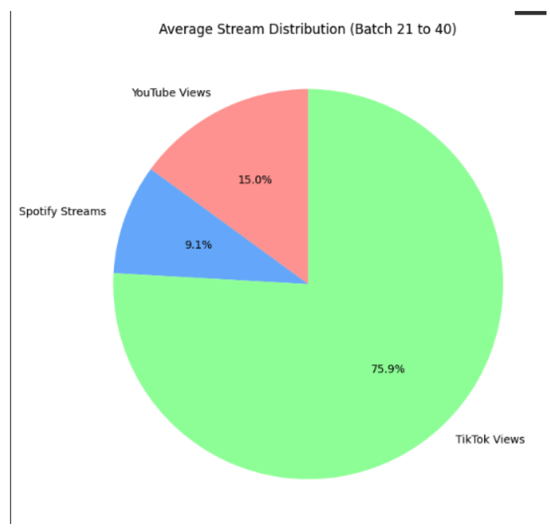

Fig. 9. Pie chart for artists ranked 21-40

- We plotted a bar chart at batches of twenty inferring from the pie chart. These bars were grouped for each artist showing the spotify, you tube and Tik Tok streams. This analysis helps in finding out those artists that had better distribution across platforms and those that focused on a single platform. Figure 11 - Figure 13 shows the bar charts.

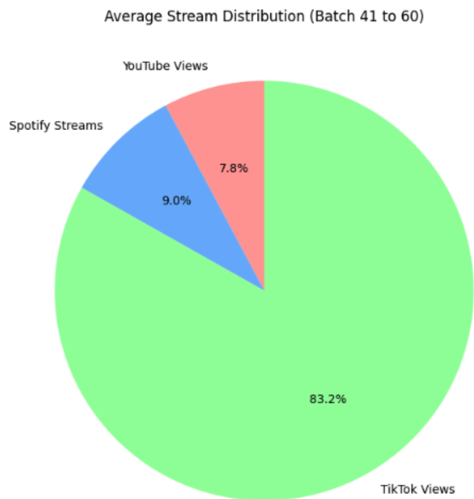

Fig. 10. Pie chart for artists ranked 41-60

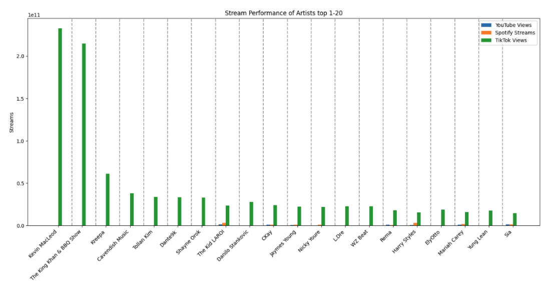

Fig. 11. Grouped bar charts for artists ranked 1-20

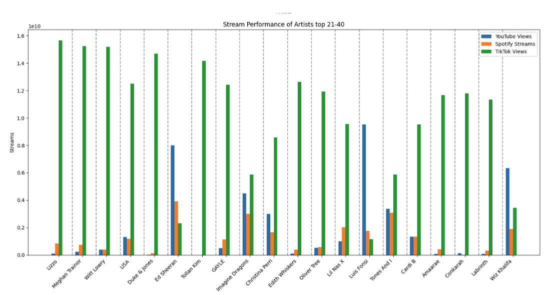

Fig. 12. Grouped bar charts for artists ranked 21-40

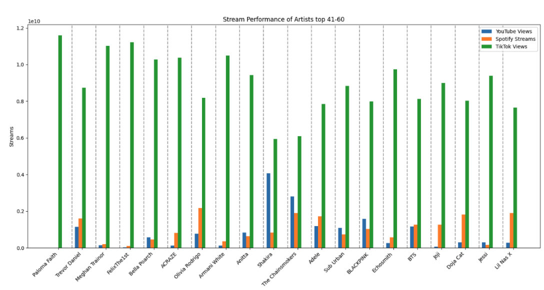

Fig. 13. Grouped bar charts for artists ranked 41-60

Knowledge:

If we examine the three bar charts that were plotted, we observe that the artists depicted in Figure 12 and Figure 13 exhibit a better distribution of streams across the platforms. In contrast, the artists in Figure 11 tend to focus their streams predominantly on a single platform. For example, Kevin MacLeod, The King Khan & BBQ Show, Kreepa, and Cavendish Music showed poor distribution of streams. In contrast to this, a little more even distribution of streams across platforms could be noted for Imagine Dragons, Wiz Khalifa, Ed Sheeran and The Chainsmokers. This difference portrays the diversity with respect to the level of platform dependency of the artists and their streaming strategies.

*3) **Iteration 3: Predicting Artist Popularity Distribution**:* This iteration of the workflow builds on the knowledge about the varying nature of artist dominance on different platforms - about how it can be very even sometimes, and also really skewed. Different machine learning models are experimented with to try and predict how an artist would fare on different platforms, by making predictions on their songs. The aim, very broadly, was to create a model that would tell which 'category' a song belonged to, with respect to its popularity on different platforms (performed much better in 1 platform vis-à-vis other platforms/performed well in 2 but not in a third, performed well in all 3 platforms almost equally) and then predicting the overall dominance of an artist as an aggregate/average of the performance of their songs. This would then be used to determine the strength of the correlation between the characteristics of song, in terms of the nature of the song, with its popularity on a platform, if at all there was any. The quality of the predictions made by our model would be an indicator of the presence and extent of said correlation.

Data:

A merge was performed on the 2 datasets being used in this assignment. The merge was on the (Track, Artist) tuple, where a string similarity check was performed to account for the cases where the same song appeared on two different datasets with slightly different track names or artist names. Filtering was performed based on the inferences from the first iteration, keeping only Spotify, YouTube and TikTok streams/views for further analyses.

Model

Two different approaches were tried for the creation of the model. One was to perform unsupervised clustering on the data, with the assumption that music that perform differently on different platforms must be fundamentally different from each other in terms of its characteristics (metrics such as tempo, energy, acousticness etc.).

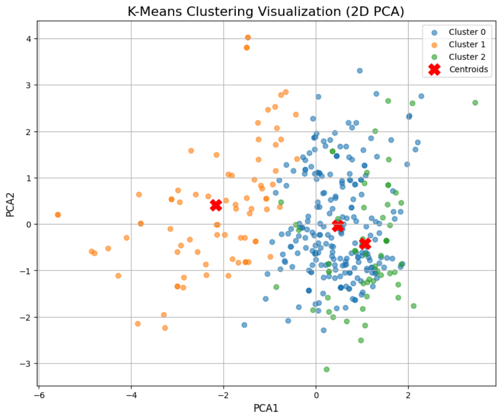

Fig. 14. Result of Unsupervised Clustering on Data

Therefore, an unsupervised clustering was done using the K-Means clustering algorithm. Although the result (represented visually in 14 after reducing to 2 dimensions using PCA) does have 2-3 clusters which are vaguely separable, a count of the categories that points in each cluster belong to, it proved to be of no real significance, as all clusters were evenly composed.
A switch was made to then using a supervised model, the random forest classifier in this case. The model would output 3 True or False values [x y z] - with x corresponding to the prediction about whether or not the song performed well in Spotify, y about TikTok and z about YouTube.

Accuracy: The model yielded accuracies of 71%, 70% and 77% respectively for predicting performances on each of the platforms. This accuracy corresponds to predicting the dominance of each song on different platforms. However, when an aggregation of these results is used to predict the artist performance on different platforms, this accuracy was found to be higher, as was seen in manually tested cases, with data from outside the dataset being used to verify model predictions.

Knowledge:

The Random Forest Classifier model proves the existence of a weak, although existent correlation between song performance across platforms and the nature of the song itself. This correlation is stronger for songs and artists that do well on YouTube than Spotify or TikTok. The exact attributes that impact the performance of a song on a platform will be studied in the next iteration of this workflow.

*4) **Iteration 4**: Mapping artists and their song nature to the categories:* In this iteration, we build upon the insights gained from the third iteration to plot a Parallel Coordinates Plot (PCP). This visualization allows us to analyze the correlations between song characteristics and their nature. By examining these relationships, we aim to understand how various attributes influence a song's dominance across different platforms.

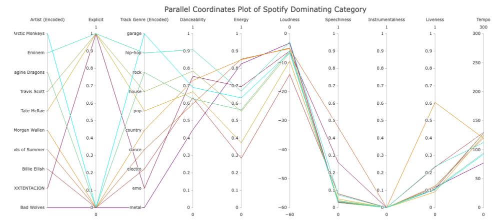

Fig. 15. PCP for songs in Spotify dominating category

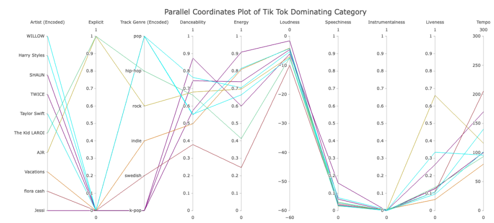

Fig. 16. PCP for songs in Tik Tok dominating category

Data:

We proceeded to make use of the merged data in iteration 3. Then we divided the data into 4 different csv files, where each csv file contained songs and its information belonging to a particular category, namely 'Spotify Dominant', 'Youtube Dominant' , 'Tik Tok dominant' , 'No dominance'. 10 artists from each category were taken to plot the PCP and to analyse the correlation. Due to the presence of greater variance in the 'Valence' and 'Acousticness', we decided to drop those columns from this pcp.

Visualizations:

Visualizations for each category has been plotted using the parallel coordinates plot which was built using the Plotly.js and D3.js api.

Figure 15 to Figure 18 shows the pcp for different categories.

For the nature of the songs like Danceability, Energy,

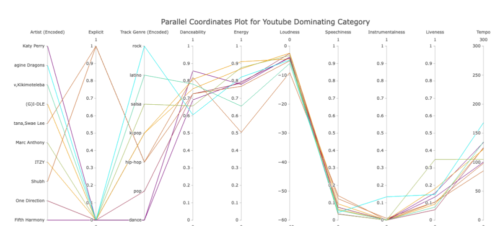

Fig. 17. PCP for songs in Youtube dominating category

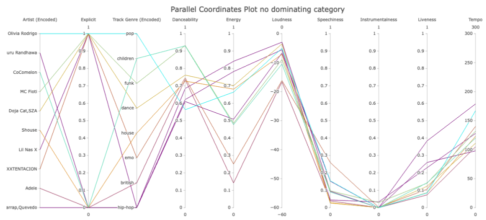

Fig. 18. PCP for songs in undefined dominating category

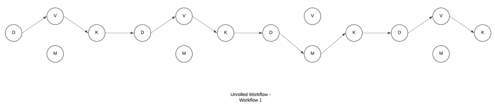

Fig. 19. Unrolled Workflow - Workflow 1

Loudness, Speechiness, Liveness and Tempo we can compare to show how these influence the category that the songs belong to. Instrumentalness was close to 0 for all categories and The track Genres had no real pattern so we don't include it in the comparison below.

- For the **Spotify** dominant category(Figure 15), Danceability and Energy displayed no significant patterns due to high variance. However, Loudness consistently ranged between -15 and -5. Speechiness averaged around 0.1, while Liveness and Tempo were approximately 0.1 and 100 to 125, respectively.

- For the **YouTube** dominant category(Figure 17), there were fewer explicit songs, with non-explicit songs being more prevalent. Other features exhibited lower variance. Danceability averaged around 0.7, and Energy was between 0.8 and 0.9 for the majority of the songs. Loudness ranged from -10 to 0, Speechiness was between 0.05 and 0.15, Liveness averaged around 0.1, and Tempo averaged approximately 125.

- For the **TikTok** dominant category(Figure 16), explicit songs were again fewer in number. Danceability, Energy, Liveness, and Tempo showed high variance, making them less contributory to the classification of songs in this category. Loudness ranged from -10 to 0, while a significant portion of Speechiness fell between 0 and 0.1.

- For the **No dominant** category(Figure 18), Tempo primarily ranged between 100 and 150, and Liveness was predominantly close to 0.1. However, other features did not exhibit meaningful patterns due to high variance.

Knowledge

The classification of songs into distinct categories was clearly demonstrated in the PCP, highlighting a correlation between the nature of the songs and their respective categories.

B. *Workflow 2*

The insights from **Assignment - 1** were considered for songs released prior to the year 2024. However, since the additional dataset we are using contains songs released prior to 2022, we only consider the merge between the two datasets, taking into account songs released prior to the year 2022. In addition, the following analysis uses the knowledge from Workflow 1 that the TikTok, Spotify, YouTube are the most relevant platforms.

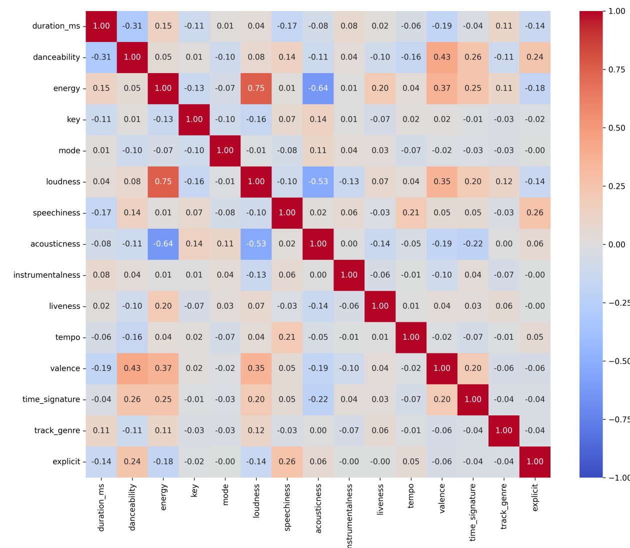

Fig. 20. Correlation Matrix for Dataset Features

Dataset Description: The merged dataset contains information regarding songs spanning approximately 45 genres. The following features will be used in the workflow:

- **track_name**: Name of the individual track.
- **duration**: Duration of the song in milliseconds.
- **danceability**: A score from 0 to 1 representing how suitable a track is for dancing based on various musical elements.
- **energy**: A measure of the intensity and activity of a track in the range 0–1.
- **key**: An estimated overall key of the track, represented by an integer where values map to pitches using standard pitch class notation (e.g., 0=C, 2=D, etc.).
- **mode**: Indicates the modality (major or minor) of a track, with major represented by 1 and minor by 0.
- **loudness**: The loudness of the track in decibels.
- **speechiness**: A score from 0 to 1 that represents the presence of spoken words in a track. Values closer to 1 indicate mostly spoken content.
- **acousticness**: A score from 0 to 1 that represents the extent to which a track possesses an acoustic quality.
- **instrumentalness**: A score from 0 to 1 representing the likelihood of a track being instrumental.

- **liveness**: A measure detecting the presence of an audience, where higher values indicate a higher probability of live performance.
- **tempo**: The overall estimated tempo of a track in beats per minute (BPM).
- **valence**: A score from 0 to 1 representing the positiveness conveyed by a track.
- **time_signature**: The number of beats in each bar of the track.
- **track_genre**: The genre of the track.

To understand the features in the additional dataset, we first analyze the correlation matrix of the numerical attributes to get a basic idea of the features themselves.

Insights from the Correlation Matrix

1) **Energy and Loudness**: Tracks with higher energy tend to be louder (**correlation: 0.75**).
2) **Acousticness and Energy**: Tracks with higher acousticness are less energetic (**correlation: -0.64**).
3) **Valence and Danceability**: Danceable tracks are often more positive in emotion (**correlation: 0.43**).
4) **Speechiness and Explicit**: Songs with spoken words are slightly more likely to be explicit (**correlation: 0.26**).
5) **Tempo and Other Features**: Tempo has no significant correlation with other features, suggesting it may vary independently.

From the insights of the previous workflow, we understand that TikTok, Spotify, and YouTube are the more relevant platforms among the various platforms available. This workflow aims to conclusively derive insights regarding what things contribute to a song's success or failure.

*1) **Iteration 1: Exploring the influence of features on platform-specific top performing songs**:* We first try to figure out how the core features, namely 'duration_ms', 'energy', 'loudness', 'speechiness', 'tempo' and 'instrumentalness' influence a song's popularity for the top-performing songs.

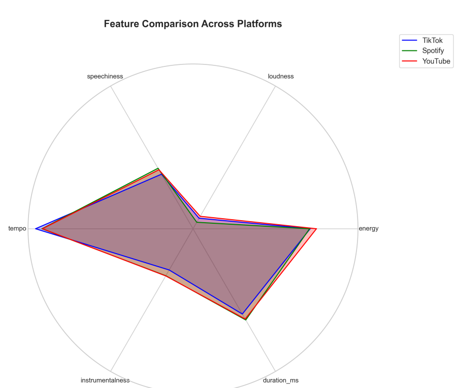

Fig. 21. Radar Chart: Min-Max scaled down means of the factors for the 3 platforms.

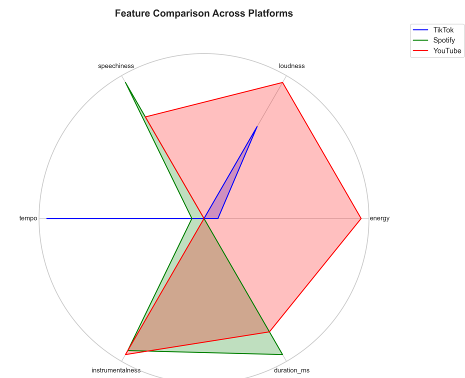

Fig. 22. Radar Chart: Min-Max normalized means of the factors for the 3 platforms.

Inorder to figure out which factors are more prominent among the 3 platforms, we visualize the differences of means by using radar plots as shown in figures 21 and 22 (where figure 21 shows the scaled down values of the mean values of the factors for a better idea of the actual values and figure 22 shows the min-max normalized values for the mean value of the features among the 3 platforms)

Insights

We draw our insights from Table 1, showing the values obtained from the plots.

- **Speechiness:** Songs with a relatively less speechiness score tend to perform better.
- **Tempo:** The 3 platforms have a high average, indicating that fast-paced songs tend to perform well, on spotify more than the other 2 platforms.
- **Instrumentalness:** We observe a very low average across the platforms, showing that songs with more vocals perform better.
- **Duration:** The better performing songs are close to 3 and a half minutes.
- **Energy:** A medium energy tends to influence the songs to getting more views, close to 0.65.
- **Loudness:** The 3 platforms have a loudness value close to -5 DB, indicating that louder songs are preferred.

	Ideal value	Min value	Max value	The Influence (3 platforms)
Speechiness	≈ 0.1	0.024	0.463	Spt $>$ YT $>$ TT
Tempo	≈ 130	48.718	205.561	TT $>$ Spt $>$ YT
Instrumentalness	≈ 0.003	0	0.703	YT $>$ Spt $>$ TT
Duration (sec)	≈ 203	94.2	436.7	Spt $>$ TT $>$ YT
Energy	≈ 0.651	0.142	0.974	YT $>$ TT $>$ Spt
Loudness (DB)	≈ -5.93	-16.6	-0.17	YT $>$ TT $>$ Spt

TABLE I
THE INFLUENCE OF THE 6 FACTORS ON THE TOP 3 PLATFORMS (TIKTOK - TT, SPOTIFY - SPT, YOUTUBE - YT)

We observe that a lower instrumentalness score is preferred and a low speechiness score is also preferred. However this

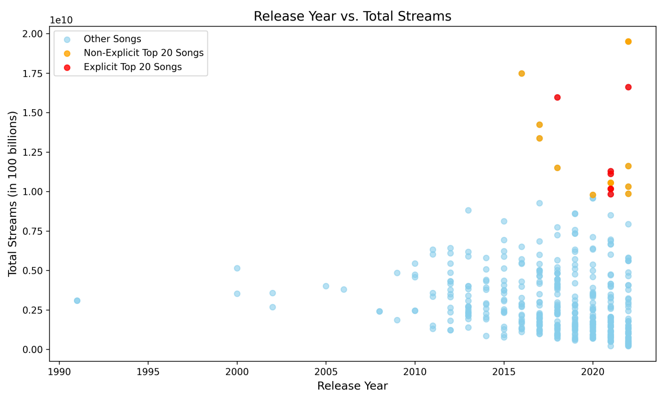

Fig. 23. Scatter Plot: Release Years of songs against combined streams.

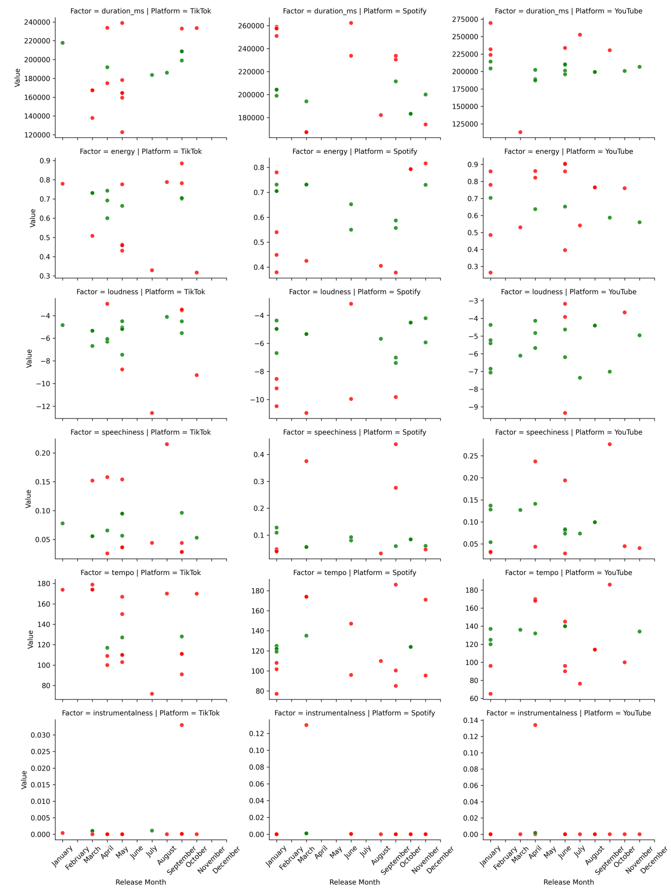

Fig. 24. Facet Grid of Scatter Plots: Release months of songs against the 6 factors's values.

would mean to have more vocals in the song and at the same have less spoken words indicating that songs with more humming noises could prove to influence a song's performance in a good way.

2) Iteration 2: An year-wise analysis of songs: We now try to analyse if the release years of the songs influence their reach by checking their combined views. We additionally, also check for the same using the 'popularity' score for performance on spotify.
Using a scatterplot, as shown in figure 23, we observe the release years of the songs and color the top performing songs.

Insights

- We observe from figure 23 that the songs released over the recent years (2018-2022) have more combined views compared to the songs that were released earlier. This could be due to many reasons, such as more content of songs due to advent in ease of sharing videos etc, or some other reasons.
- We also observe that there are some songs with explicit content (6 songs), mainly in the year 2021 when considering the top 20 songs. This could mean that a song would perform well being non-explicit rather than containing explicit content.

3) Iteration 3: Combined Influence of the factors and Release months of songs: We've understood from the previous loop that the songs released in the recent years performed the best. Now, we try to go a step further and check if there are any other reasons, such as festivals or any other holidays, that play a role in influencing a song's reach. For this, we try to analyse the trends in factors for top 20 songs, disregarding the influence of year for now as we consider the **years 2018-2022** only.
The figure 24 represents a grid a scatter plots, plotting the values of the 6 factors for the 3 platforms. The red colored scatter-points indicate that the points lie far away from the threshold values (the same as the ideal values in Table 1) and the green points indicate that the points lie close to the threshold values. The margins used for the 6 factors were: speechiness - 0.05, tempo - 15, instrumentalness - 0.002, duration - 20 seconds, energy - 0.1, loudness - 2

	TikTok	Spotify	YouTube
Speechiness	0.05 - 0.15	0.05 - 0.15	0.05 - 0.15
Tempo	100-120,160-180	100-140	100-140
Instrumentalness	0.001	0.001	0.001
Duration (sec)	160-200, 230-240	180-220	180-220
Energy	0.55-0.75	0.55-0.75	0.55-0.75
Loudness (DB)	-4 to -8	-4 to -8	-4 to -8
Best Months	Mar-May, Sep	Jan, Sep-Nov	Jan, June (others too)

TABLE II
BEST VALUES FOR PLATFORM SPECIFIC SONGS

Insights

- **Seasonal Preferences:**
  - **TikTok:** The best seasons to release songs are **spring (March to May)** and **early autumn (September)**, aligning with increased engagement during these times.

- **Spotify:** Songs tend to perform better during **winter (January)** and **autumn (September to November)**, suggesting a preference for new music in colder, reflective months.
- **YouTube:** While **January** and **June** stand out as optimal months, other months seem to work reasonably well, indicating a consistent performance across the year.

- **Threshold Comparisons:**
  - Across all platforms, songs with specific characteristics tend to align closely with the threshold values:
    - **Speechiness:** Ideal range is **0.05–0.15**.
    - **Tempo:** Optimal ranges vary:
      - **TikTok:** Prefers **100–120 bpm** or **160–180 bpm**, possibly for its short-form content and dance trends.
      - **Spotify and YouTube:** Favor a broader range of **100–140 bpm**.
    - **Instrumentalness:** Low instrumentalness (close to **0.001**) is common across platforms, suggesting vocals are key to popularity.
    - **Duration:**
      - **TikTok:** Prefers extremes, with durations clustering around **160–200 seconds** or **230–240 seconds**.
      - **Spotify and YouTube:** Stick to the mid-range, preferring **180–220 seconds**.
    - **Energy and Loudness:** Energy levels of **0.55–0.75** and loudness between **-4 to -8 dB** are universally favored, reflecting a balance of vibrant yet not overwhelming audio.
- **Platform-specific Trends:**
  - TikTok's song preferences exhibit greater variability, as seen by clusters of red points far from the thresholds. This may reflect the platform's trend-driven, dynamic nature.
  - Spotify and YouTube have more green points, indicating a stronger adherence to threshold ranges, possibly due to their more consistent, curated algorithms and user behavior.

*4) **Iteration 4: Genre analysis of songs**:* We now try to look at a different aspect of songs on each platform to figure out whether a certain genre dominates any platforms or if there is equal preference for different genre.
We do this by considering pie charts, one for each platform, and check the percentages of the different genres while considering top 20 songs and then top 50 songs to get a better understanding of how genre distribution varies with the number of songs considered. By analyzing the genre breakdowns for both smaller and larger sample sizes, we can assess whether platforms tend to favor a specific genre or whether there is a more balanced representation across multiple genres.

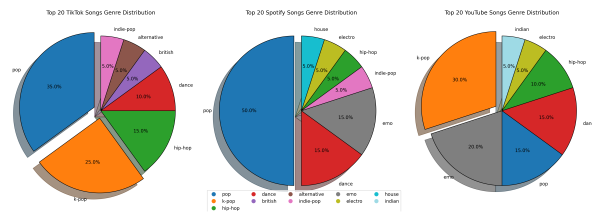

Fig. 25. Pie Chart: Genres of top 20 platform-specific songs.

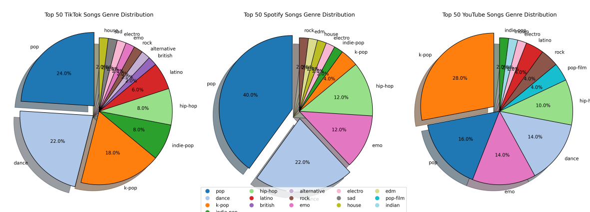

Fig. 26. Pie Chart: Genres of top 50 platform-specific songs.

Thus, we observe from figures 25 and 26, that while considering top 20 songs, Pop, K-pop tend to dominate, with TikTok favoring both, Spotify favoring Pop, and YouTube favoring K-pop in addition to having a better diversity of genres. When considering 50 songs, Pop continues to be dominant for the most part, with K-pop, dance, and emo also having a notable percentage. Other niche genres also tend to emerge such as 'rock', 'electro', 'indie-pop' with youtube showing a better percentage as compared to the other 2 platforms. Thus, considering everything, the genres 'pop', 'K-pop' and 'dance' tend to perform well, albeit different platforms favoring some of these.

We now try to confirm if the values of the 6 factors obtained from data exploration (in Table 2), still hold for the best genres to see their combined influence on a song's popularity. We do this by using a small multiples chart of violin plots to see the distribution of the values for the factors for the top 3 genres (considering all the songs) and check if they match with the values obtained in Table 2.

Thus, we observe from figure 27 that the 3 genres tend to exhibit values that are very close to the values obtained from table 2 for the 6 factors. However, we do notice an outlier in the 'dance' genre but can ignore it for the most part.

Thus, in conclusion, given a song with it's features (of genre, speechiness, loudness etc) the likelihood of it performing well can be maximised if it aligns with the following conditions:

- Considering the 6 factors, should have values close to those mentioned in Table 2.
- Considering when to release the song: Releasing the song in Autumn would likely result in better performance in all the 3 platforms. Releasing in months specified in Table 2 would give better results for a platform-specific popularity.
- Considering the genre of the song: The song is likely to perform well on TikTok if it is 'pop', 'dance' or

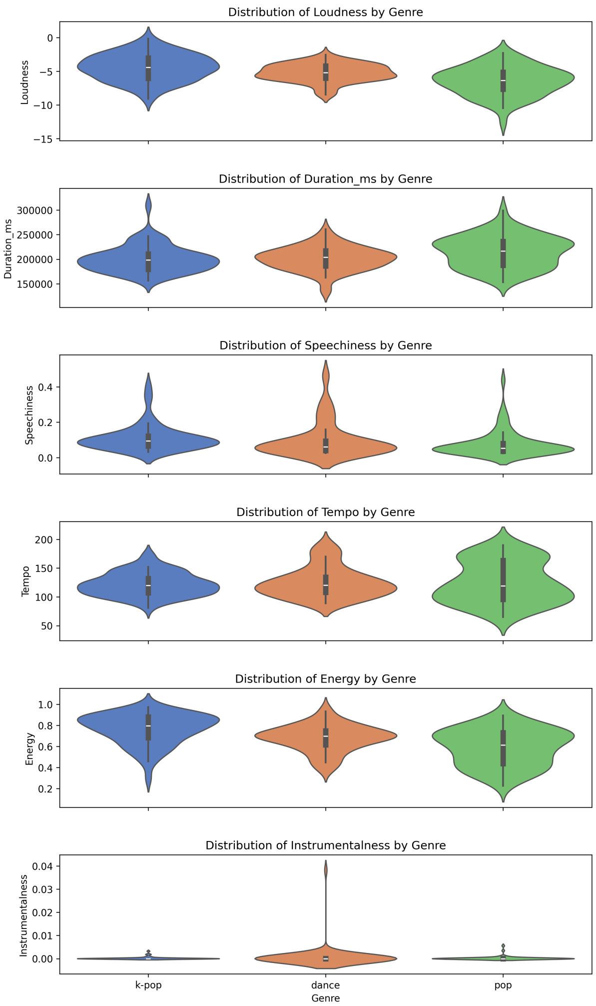

Fig. 27. Small Multiples Violin Chart: Distribution of the 6 factors for the 4 major performing genres.

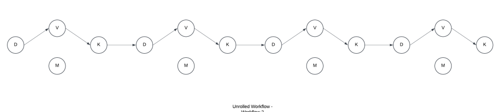

Fig. 28. Unrolled Workflow - Workflow 2

'k-pop', on Spotify if it is 'pop', 'dance' or 'emo' (but only to a certain degree), on YouTube if it is 'K-pop', 'emo', 'Pop' or even 'dance'.

IV. Work Distribution

- Siddharth Menon - IMT2022001
  - Preprocessing and Dataset Merging
  - Iterations 1 and 3 of Workflow 1
- Shreyank Bhat - IMT2022516
  - Workflow 1 - Iteration 2 - Artist Analysis
  - Workflow 1 - Iteration 4 - Mapping artists and their song nature to the categories
- Vrajnandak Nangunoori - IMT2022527
  - Workflow 2 - Iteration 1 - Influence of audio attributes of songs.
  - Workflow 2 - Iteration 2 - Influence of release year of songs.
  - Workflow 2 - Iteration 3 - Combined influence of the release months and the audio attributes.
  - Workflow 2 - Iteration 4 - Role of genre in determining song success.