

Spotify Analysis on Kpop Female Groups: Twice, Aespa and Red Velvet*

Anqi Xu, Yunkai Gu, Yitong Wang

October 10, 2024

The musical patterns of three major K pop female groups, Aespa, Red Velvet, and TWICE, are examined in this study using Spotify data on song duration, valence, and album release date. We identified variations in song duration and emotional expression throughout time, as well as significant disparities in each group’s contribution to the broader K pop music landscape. In recent years, Aespa’s music has taken on a more upbeat tone, whereas TWICE has demonstrated a wide range of emotional diversity. The data provide insights into these organizations’ shifting music production techniques, demonstrating how trends in the K pop business change over time and influence worldwide music consumption habits.

Table of contents

1	Introduction	2
2	Data	2
2.1	Overview	2
2.2	Measurement	3
2.3	Outcome variables	3
3	Results	5
	Appendix	6
A	Additional data details	6
	References	7

*Code and data are available at: <https://github.com/demainwang/Spotify-Analysis-Kpop.git>

1 Introduction

The global popularity of K-pop has surged in recent years, particularly driven by the dynamic performances, innovative music production, and unique branding of female groups such as TWICE, Red Velvet, and Aespa. As these groups achieve international recognition, understanding the musical trends they represent can provide valuable insights into the evolving music tastes of global audiences. Spotify provides large-scale access to music consumption data, and enable quantitatively analyze the evolution of their music.

This paper explores key musical characteristics, such as song duration and valence, alongside the frequency of album releases, for the three selected groups using Spotify data from 2015 to 2024. Through these analyses, we aim to identify trends and patterns in their musical evolution and contribute to a deeper understanding of the K-pop genre’s development over the past decade. Specifically, we investigate how the duration of songs has changed, how emotionally positive or negative the music is, and how each group’s contribution to the total number of releases has evolved over time.

This research is important because it presents the evolving strategies of K-pop groups in the global music market, highlighting how they adapt to shifting consumer preferences. By analyzing these musical trends, we provide a quantitative basis for understanding the global success of these K-pop groups and offer insights into broader trends within the global music industry.

The remainder of this paper is structured as follows. Section 2 provides details on the data source, the specific variables used and analysis of constructed data visualization. Data tools and methodologies applied in this paper are stated in Section 2.1, and Section 2.2 explains the variables in detail. Section 2.3 presents analysis and visualization with multiple plots and corresponding descriptions. Section 3 presents the overall results of the whole research. Section A provides additional details for the paper, including process of data downloads. The comprehensive structure ensures that the reader gains a full understanding of what was analyzed, the findings, and their significance.

2 Data

2.1 Overview

The dataset is originally obtained from Spotify (Spotify 2024), a music streaming platform. In the analysis procedure, We use the statistical programming language R (R Core Team 2023) to conduct the analyses in the paper. The `spotifyr` packages (Thompson et al. 2022) were used to download the raw data from Spotify. Data cleaning process was done through `tidyverse` package (Wickham et al. 2019). `ggplot2` package (Wickham 2016) and `lubridate` package (Grolemund and Wickham 2011) are then used to make graphs.

2.2 Measurement

The production and release of music could be affected by multiple factors in real life, such as shifts in musical styles, changes in listener preferences, impact of technological advancements and so on. As a large music steaming platform, Spotify gathers a large dataset with various data types from its users, which provides information of each singer or music group on such dynamics.

The analysis in paper mainly focuses on three primary variables of the female Kpop groups Twice, Red Velvet and Aespa: album counts each year, song length, and valence. “Album counts” each year measures the number of albums the group released in one-year basis; “Song length” measures the duration of each song released by the groups; “Valence” captures the emotional tone of the song which has value between zero and one. It “signals the musical positiveness of the track”. The higher value of valence, the more positive the song is.

By employing statistical and visual analysis techniques, the paper aims to derive meaningful insights into the interrelationships among these variables. Section A includes additional data details.

2.3 Outcome variables

Analysis and data visualisation are conducted following with multiple plots and charts.

Figure 1 represents the album release distribution of each group from 2015 to 2024. The number of albums released by each group is broken down by year and visualized by histogram. Different color is applied to identify different groups which red represents Aespa, blue for Red Velvet and green for TWICE.

TWICE has been a significant contributor to the total number of albums released every year, particularly between 2016 and 2018, where they were the primary group releasing albums. While their contribution slightly decreased after 2018, TWICE remains the dominant group in terms of album releases. Red Velvet had a steady but lower album release contribution compared to TWICE. Aespa’s contribution starts in 2022, reflecting their later debut; nonetheless, Aespa quickly contributed a notable portion of albums, especially in 2023 and 2024, where their album releases were comparable to or even exceeded those of Red Velvet.

Figure 2 illustrates the variation in song durations over time, with most songs concentrating between 180,000 to 220,000 milliseconds. There is not clear trend appear to indicate a obvious increase or decrease in song duration, the data shows occasional outliers with extremely longer durations, especially around 2018 and 2024. These outliers could be experimental tracks or unusual releases. Despite these exceptions, the bulk of songs have remained relatively stable in length throughout time, implying that the musician has consistently adhered to a set range of normal track lengths.

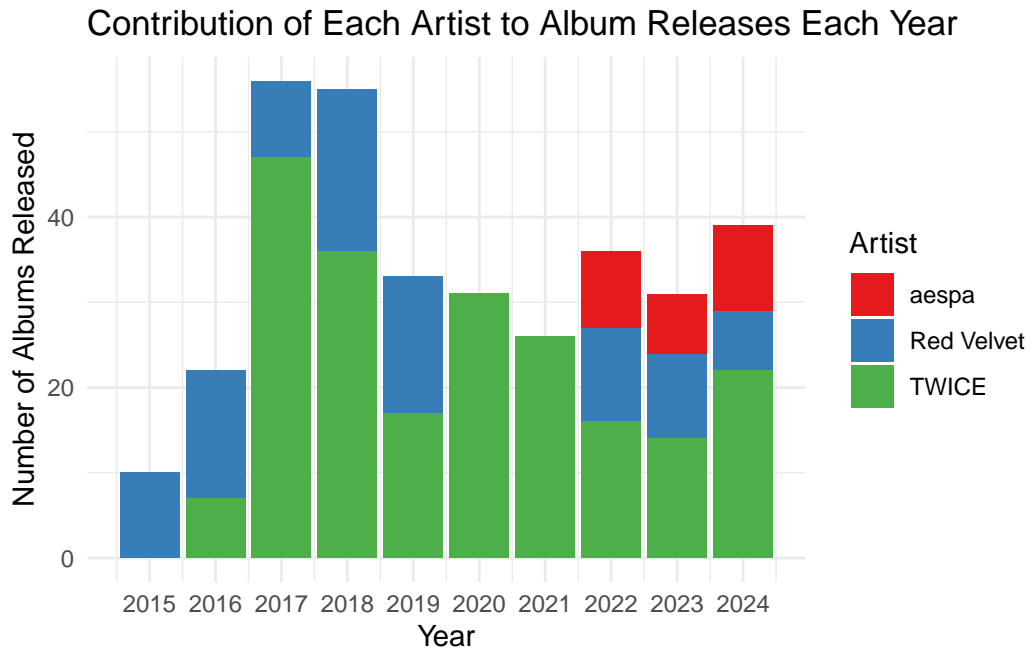


Figure 1: Length of each Twice song over time

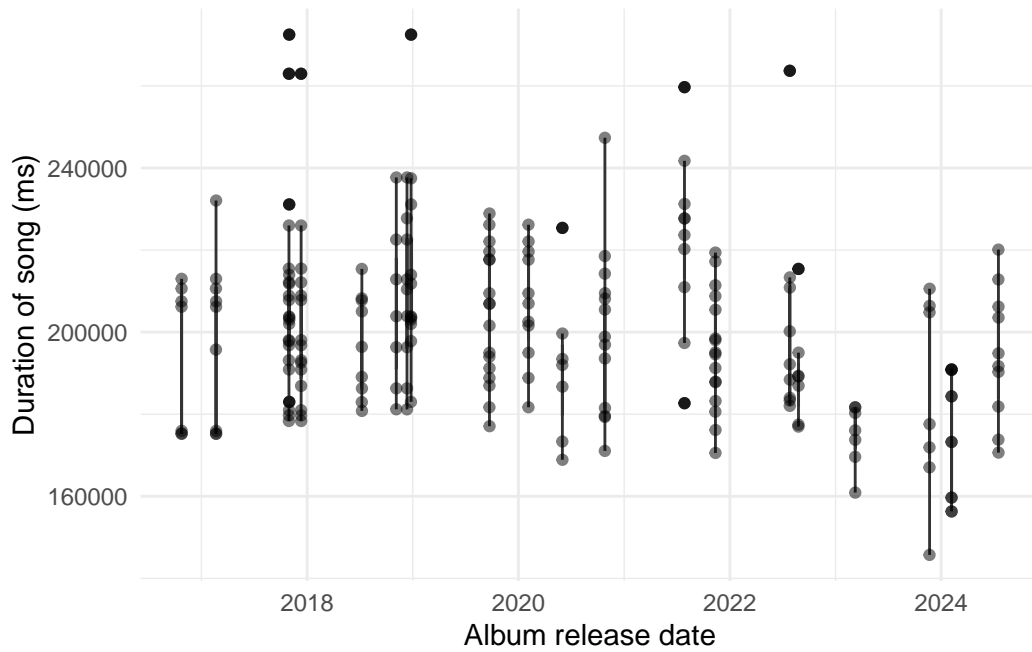


Figure 2: Length of each Twice song over time

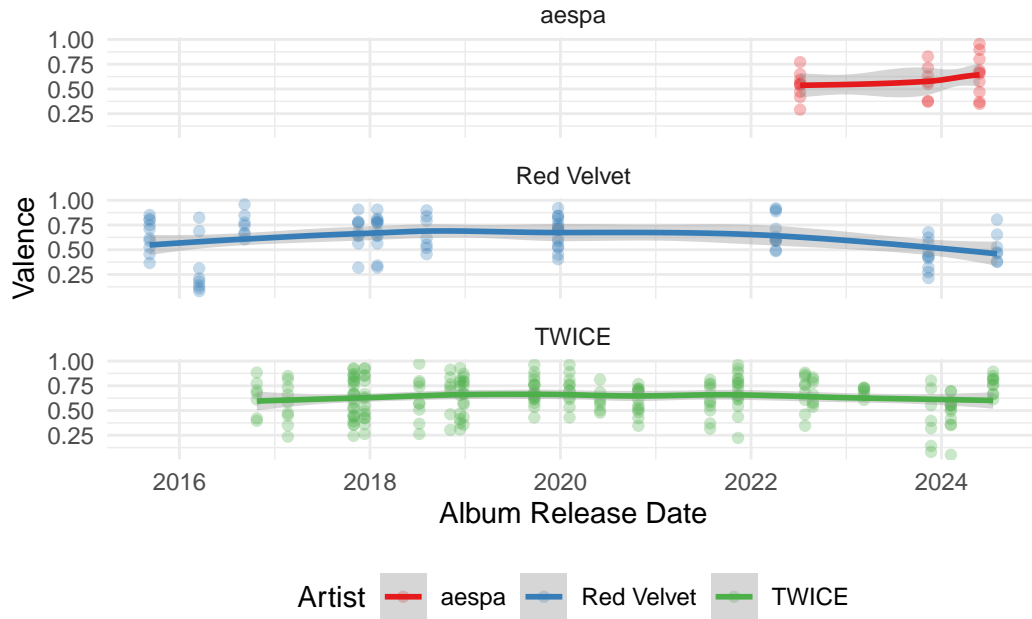


Figure 3: Comparing valence over time, for aespa, Red Velvet and Twice

The scatter plot Figure 3 presents the valence trend based on the album release date for three K-pop groups, Aespa, Red Velvet, TWICE. The red line represents Aespa, blue line for Red Velvet and green line for TWICE.

Aespa’s valence shows a clear upward trend from 2022 to 2024, indicating their songs are becoming more positive or “happier” over time. Red Velvet has maintained a relatively stable valence over the years, though there is a slight downward shift around 2022. However, the overall valence of their music remains high, suggesting a consistently positive emotional tone. TWICE exhibits a more varied valence with fluctuations across the years but recent data from 2022 to 2024 suggests a possible slight rise in valence.

3 Results

Our findings reveal distinct trends among the groups. For instance, Aespa’s recent music shows an upward trend in emotional positivity, while TWICE’s emotional range is more variable, showing a mix of positive and neutral tones. Red Velvet, on the other hand, maintains a relatively stable musical valence over time. Additionally, we observe a decline in average song length since 2020, with shorter songs becoming more common, and identify how Aespa’s recent rise in album releases compares to the longstanding dominance of TWICE.

Appendix

A Additional data details

The original dataset is obtained from Spotify, a music streaming platform. The raw dataset contains 39 variables, each column representing certain piece of information for the music group.

After downloading the raw dataset by `spotifyr` package (Thompson et al. 2022), the raw dataset for each group is converted into rds file respectively. It is then made into tibble to make the analysis easier.

For the figures plotting three groups together, mutation process contributes to the data cleaning, creating tibble which contains the variables needed for analysis.

References

- Grolemund, Garrett, and Hadley Wickham. 2011. “Dates and Times Made Easy with lubridate.” *Journal of Statistical Software* 40 (3): 1–25. <https://www.jstatsoft.org/v40/i03/>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Spotify. 2024. “Spotify Web API.” <https://developer.spotify.com/documentation/web-api/>.
- Thompson, Charlie, Daniel Antal, Josiah Parry, Donal Phipps, and Tom Wolff. 2022. *Spotifyr: R Wrapper for the 'Spotify' Web API*. <https://github.com/charlie86/spotifyr>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.