INFSCI 2415 Information Visualization

Final Report

Chahee Park

November 29, 2022

**FIGURE**

**Chart, line chart

Description automatically generated**

**Chart, line chart

Description automatically generated**

Each line is color-coded by the member’s official color.

**FINDINGS**

* This multinational 9-member girl group debuted with two-superstars
* No one has received stable line distributions – changes occur
* 2018-2019 had the most noticeable changes among members with line distributions and the group’s hits frequency in the chart

**DATA**

Two datasets and a Spotify playlist by Spotify, and background knowledge were used for this project. The primary dataset is Korean pop girl group TWICE’s Korean single lyrics from 2015 to 2022. The variables include *start\_time, end\_time, lyric, vocal1, vocal2, duration, song\_name, and song\_year*. The dataset was created by the owner using timing data from Youtube music video subtitles and scraping blog entries from colorcodedlyrics.com for the song lyrics. The secondary dataset contains hits (popular songs) from the era (90s) or by the year from 2000 to 2021. These playlists were created officially by Apple Music and transferred in Spotify for comparison. For the sake of coordinating datapoints for the two datasets, additional Apple Music official playlist was used for a missing year, 2022, for a manual data entry and cross-checked with a Spotify official playlist. The variables include *title, artist/s, danceability, energy, key, loudness, mode, speechiness, acousticness, valence, tempo, duration\_ms, and time\_signature*. Background knowledge were used to explain outstanding datapoints and they were supported by the online sources listed.

**Dataset sources:**

Peralta, JC. (August, 2022). TWICE Korean singles lyrics. *Kaggle*. <https://www.kaggle.com/datasets/jcacperalta/twice-korean-singles-lines>

Berjamin, Sandra Angela. (February, 2022). K-Pop Hits Through The Years. *Kaggle*. <https://www.kaggle.com/datasets/sberj127/kpop-hits-through-the-years>

Apple Music. (2022). Top 100: South Korea. *Apple Music*. <https://music.apple.com/us/playlist/top-100-south-korea/pl.d3d10c32fbc540b38e266367dc8cb00c>

Spotify. (2022). K-POP ON!. *Spotify*. <https://open.spotify.com/playlist/37i9dQZF1DX9tPFwDMOaN1?si=6d24bbdfaa244f9d>

**Supportive online sources for the background knowledge:**

Allkpop. (November 5, 2020). Netizens start rumors that TWICE's Chaeyoung is dating a tattoo artist. *Allkpop*. <https://www.allkpop.com/article/2020/11/netizens-start-rumors-that-twices-chaeyoung-is-dating-a-tattoo-artist>

R. Jun. (January 6, 2020). TWICE’s Jihyo Posts Letter Apologizing For Recent Comment, Explains How She’s Been Feeling The Past Year. *Soompi*. <https://www.soompi.com/article/1375547wpp/twices-jihyo-posts-letter-apologizing-for-recent-comment-explains-how-shes-been-feeling-the-past-year>

Yonhap. (January 2, 2020). Super Junior's Hee-chul, TWICE's Momo in romantic relationship. *The Korea Herald*. <https://m.koreaherald.com/view.php?ud=20200102000082>

General, Ryan. (October, 2018). K-Pop Group TWICE Finally Allowed to Date People After 3-Year Dating Ban. *Nextshark*. <https://nextshark.com/twice-dating-ban-k-pop/#:~:text=Members%20of%20K%2Dpop%20girl,three%20years%20from%20their%20debut>.

Joan, Roby. (September, 2022). 4 Female K-pop Idols Who Drew Criticism for Weight Gain. *KpopStartz*. <https://www.kpopstarz.com/articles/308972/20220918/4-female-kpop-idols-criticism-weight-gain.htm>

**METHOD** (terms are mentioned in *italics*):

For the primary dataset, only vocal 1 (main vocal for the lyric), duration, song\_name, and song\_year were used for this analysis. From the secondary dataset, frequency of artists name was measured and used.

**Manipulating Dataset and Line Plot**

The primary dataset’s data points are sorted by lyric and by song. As there are multiple singles in a year, it is decided to group them (*groupby*) by vocal 1, song\_year, and sum of duration of the year. Then each member’s duration was represented as a fraction of the entire song duration of the year. *Sns.lineplot* was used to present the line distribution (percent) by year.

**Bar Chart**

Among the secondary dataset, specific range of years was used (2015-2021) since TWICE has debuted. Then, additional research helped adding a datapoint of 2022 manually (mentioned in the “data” part of the report). As the secondary dataset contains multiple csv files, each year’s data frame was created and the *frequency* of TWICE’s single in the chart was recorded manually using *Dictionary*. Then bar chart has been created (*plt.subplots, sns.barplot*) and the figure has been saved to be used as overlay.

Overlaying the bar chart figure as a *background*, line plot has been added and arranged accordingly (sharing the same x-axis).

**SIGNIFICANCES OF THE FIGURE AND TOPIC**

Throughout the 8 years of TWICE history, some years had outstanding changes in terms of the line distribution which often has to do with the member’s popularity. Based on the presented figure with supportive online sources, it is plausible to say that each member’s scandalous news including tattoo, dating, gaining weight (regardless of the cause) plays a negative and pivotal role in their line distributions and overall the group’s image. It is not only TWICE’s own matter but as dating and tattoo are still considered to be taboo within K-pop idols and as a “shown” job, in K-pop industry, it is required for them to maintain skinny body regardless of their health issues. Therefore, it is expected that analogous patterns would appear in another research on other K-pop groups – leading members, impact on reputation due to scandalous incidents, line distribution competition, and such.

**GITHUB PAGE**

the used Python code, data source, and the report itself as the documentation