The Analysis of K-pop Music

SI 618 Project 2 Report Ping-Lun Lai

Motivation

In recent years, the Hallyu (Korean Wave) had a significant influence in Asia. As a K-pop lover, I have listened to and watched plenty of K-pop music videos on YouTube. Thus, I am interested in interpreting them to have more understanding of my interest.

This project performs an analysis of a K-pop information dataset to realize the growth of K-pop music. To be more specific, the following are the main questions and several questions I scaled down to:

1. How's the evolution of K-pop groups?
   - The total amount of new debut groups
   - The average number of members per group
   - The average percentage of the non-Korean in a group
   - The average age of groups when they debuted
2. How's the growth of K-pop music videos?
   - The number of released music videos per year for groups versus solo artists
   - The number of released music videos per year for the artist type
   - The average total views per music video for the major and the minor album
   - The average video length during the four seasons for each year and predict the length in the future
3. What's the distribution of famous K-pop artists?
   - The comparison between artists in groups or solo, and the gender of the group
   - The highest likeCount, commentCount, and viewCount from these music videos

Data Source

The dataset I used is K-Pop Database (1992-2020) | Kaggle from Kaggle (https://www.kaggle.com/datasets/kimjihoo/kpopdb) and YouTube Data API | Google Developers (https://developers.google.com/youtube/v3). Generally, the dataset contains information between 03/1992 – 05/2020.

In the K-Pop Database dataset, there are four CSV files:

1. kpop_idols.csv
   There are 1310 records in this file. The variables that I used are "Date of Birth", "Group", and "Country".

2. kpop_idols_boy_groups.csv
   There are 147 records in this file. The variables I used are “Name”, “Debut”, “Members”
3. kpop_idols_girl_groups.csv
   There are 152 records in this file. The variables I used are as same as those in boy groups.
4. kpop_music_videos.csv
   There are 3772 records in this file. The variables I used are “Date”, “Artist”, “Video”, and “Type”.

Methods

Question 1. How’s the evolution of K-pop groups?

For this question, I concatenated the boy groups’ and girl groups’ data into a data frame *kpop_groups* and added a column “gender” to label the group’s gender. Besides, since I only need the debut year, I made a column “debut_year” to retrieve it from “Debut”. In this data frame, the variables I used have no missing or incomplete data.

Question 2. How’s the growth of K-pop music?

For this question, since there is no information about music videos’ views and video duration in ‘kpop_music_videos.csv’, I used the YouTube API to retrieve them. In this file, there are some duplicate records. Thus, I kept the last record that appeared in the file. In addition, I added “year” as the music video release year, “is_group” and “gender” as some features of the artist.

Question 3. What’s the distribution of famous K-pop artists?

For this question, I defined famous K-pop artists represent they released at least one music video that has over 100,000,000 views. After filtering my definition, I used the “statistics” information retrieved from YouTube API and group them by “Artist”, “Type”, etc. To have a better visualization, I used the mosaic to compare the percentage of some variables.

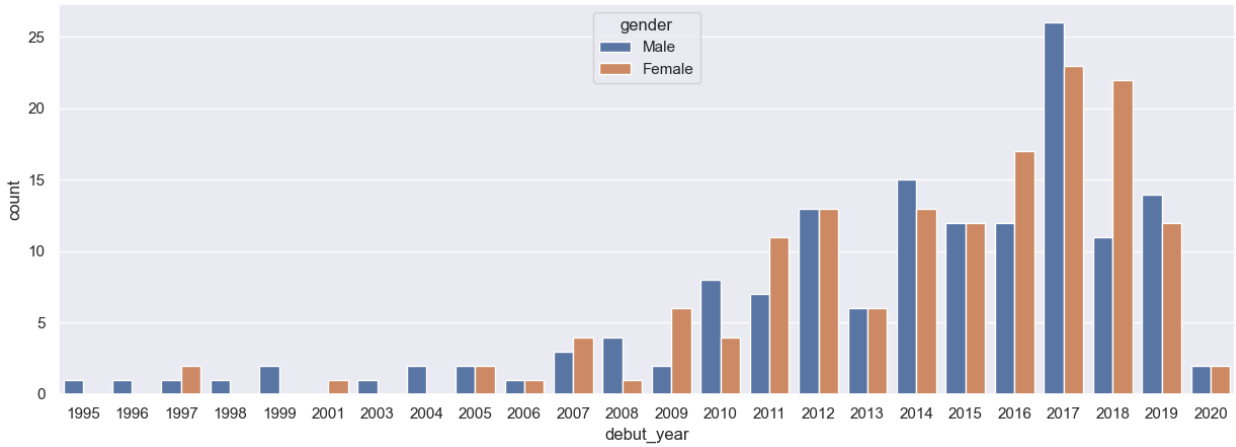
The challenge I met in this project is the usage of YouTube API. Due to the daily request restriction of the API and I had to learn how to use it, I spent a lot of time trial and error. Sometimes a misclick could cause another three days of waiting.

Analysis and Results

To extract the target feature list in the motivation section, I used the groupby function in pandas to calculate the result and visualize them with seaborn and plotnine. In these analyses, I neglected the data recorded in 2020 because the time duration in the original dataset only records the data to May 2020.

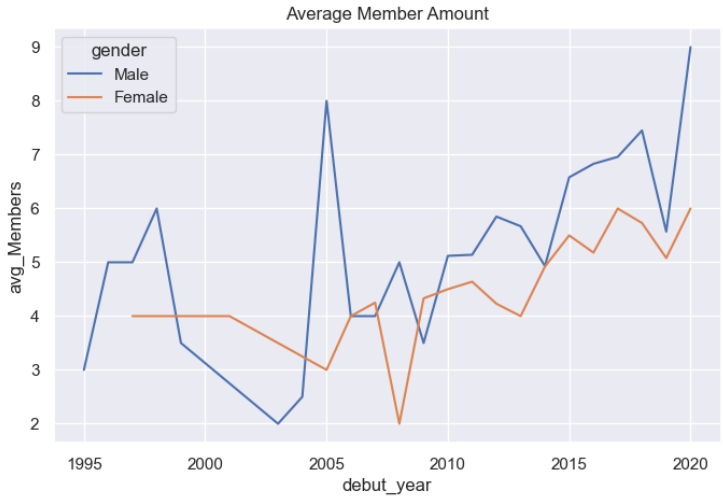
Question 1. How's the evolution of K-pop groups and idols?

- The total amount of new debut groups



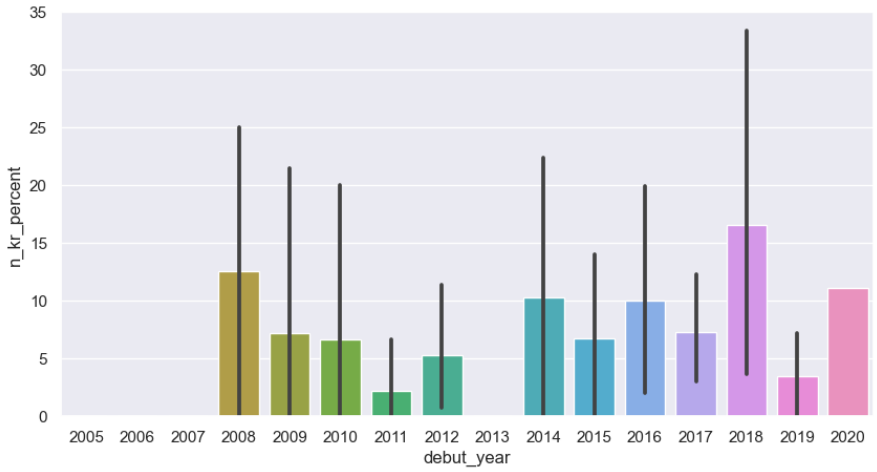
From the figure above, it is clear that the new debut groups start ascending around 2007 – 2008, which might be the time when K-pop starts flourishing.

- The average number of members per group



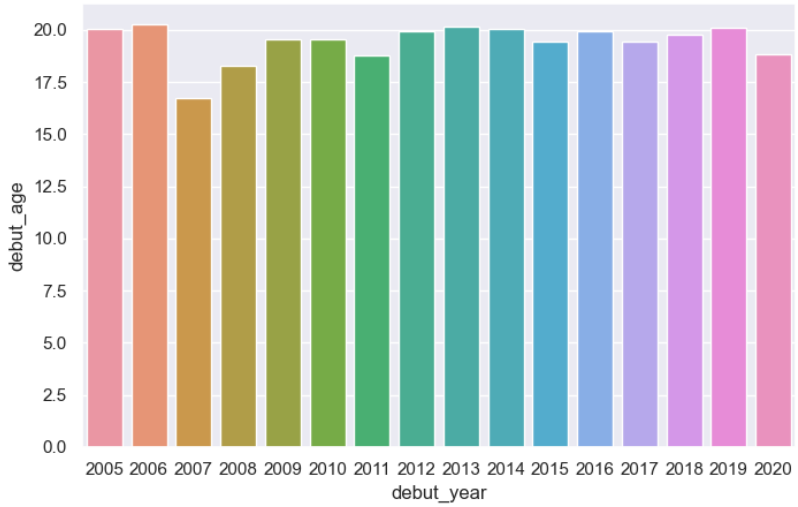
From the figure, it is obvious that both male and female groups have an increasing trend. Interestingly, it also shows that male groups have one more member than female groups on average.

- The average percentage of non-Korean in a group



From the figure above, although the variance is quite high, the average percentage of non-Korean in a K-pop group is growing since 2012 generally, which indicates that K-pop might have started to increase its customer base in international markets since then.

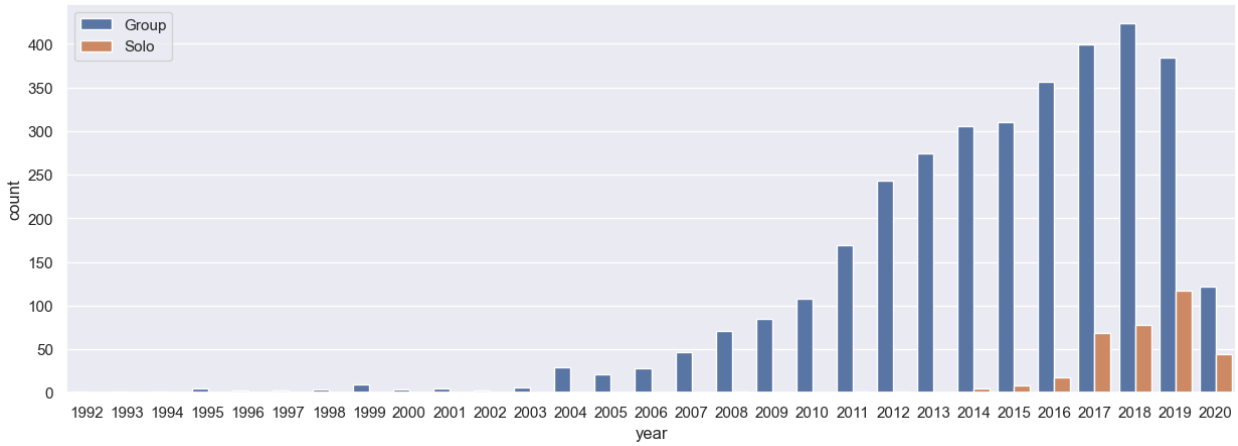
- The average age of groups when they debuted



This figure shows that across the time duration that has the records in the data source, the average age of a group is around 17 to 20. It means that K-pop idols often start working before they come of age.

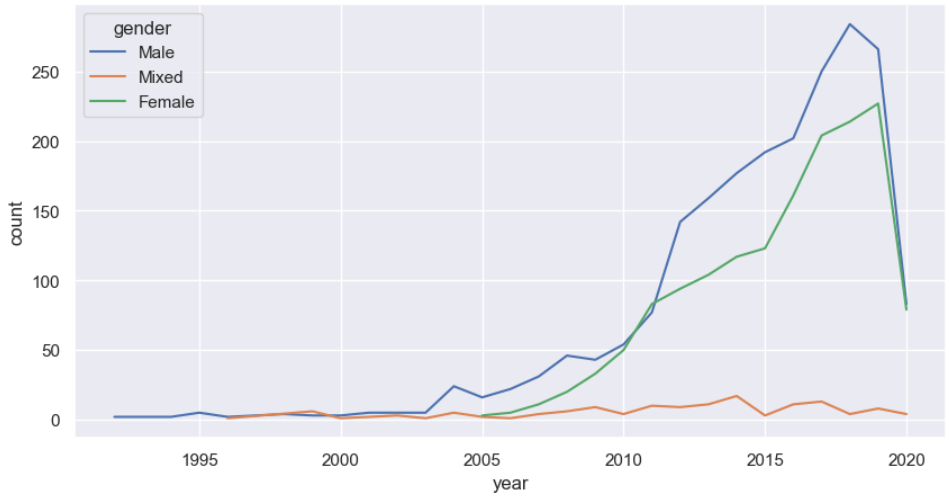
Question 2. How's the growth of K-pop music?

- The number of released music videos per year for groups versus solo artists



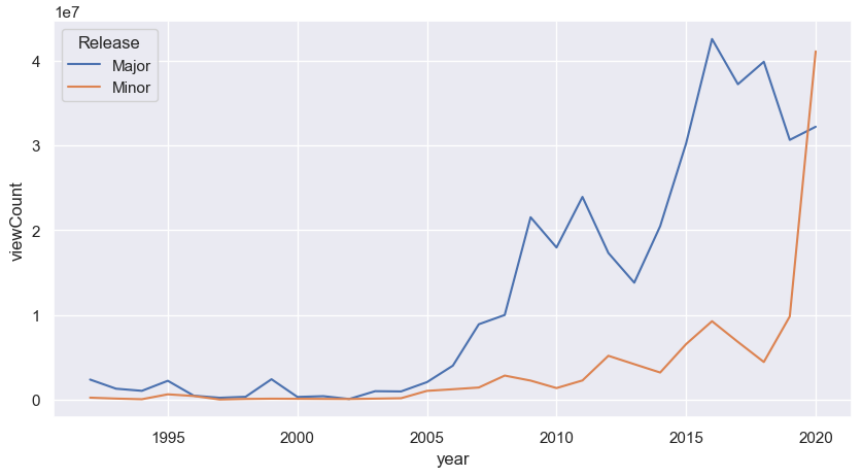
The figure above represents the fact that the concept of boy solo and girl solo started in 2014. The growth of the released music videos can also prove the analysis in the first part of question 1.

- The number of released music videos per year for the artist type



From the figure above, it is clear that boy groups produced more music videos than girl groups. The number of music videos made by mixed groups doesn't have a significant change in these years.

- The average total views per music video for the major and the minor album



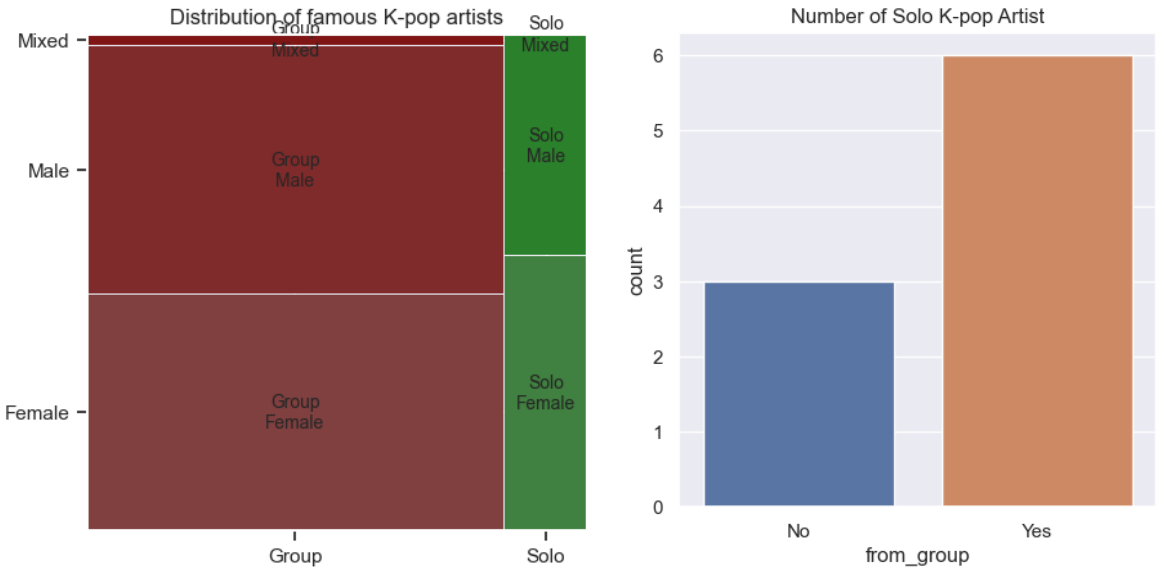
The figure indicates that after K-pop start flourishing (2007-2008), music videos in major albums receive more attention than those in minor albums. However, in 2019 and 2020, music videos in minor albums start having huge success. It can reflect the fact that during the covid-19 epidemic, people tend to watch and listen to more K-pop than its expected growth rate obtained from previous years.

- The average video length during the four seasons for each year and predict the length in the future

In this figure, I find out that the video length in the early progress of K-pop has a larger variance and a longer average duration. Besides, the video length is decreasing gradually, which might show that audiences nowadays are apt to have shorter musical entertainment.

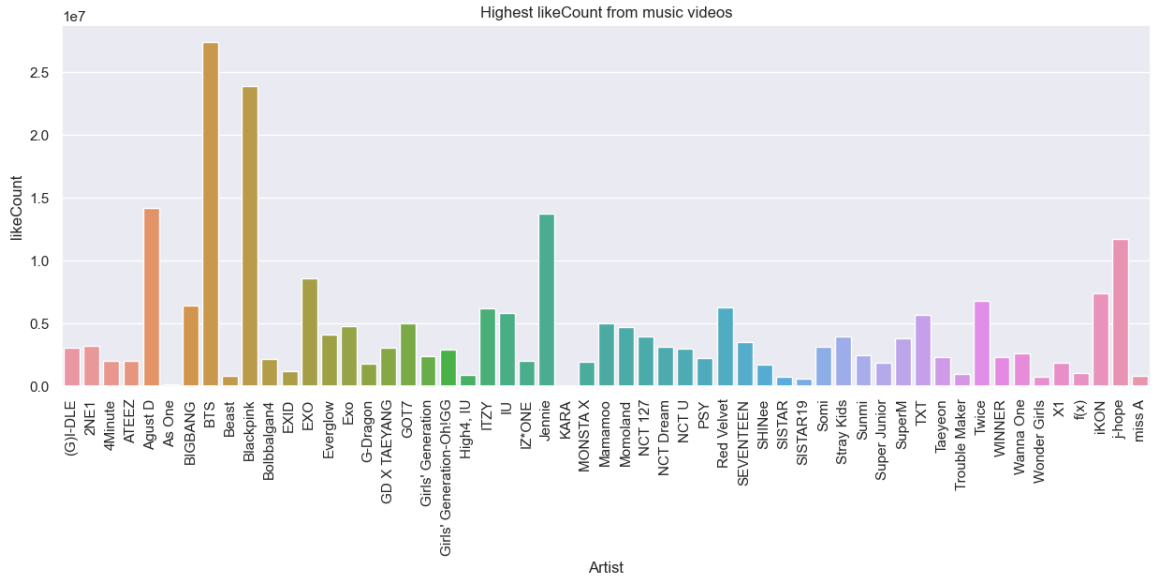
Question 3. What's the distribution of famous K-pop artists?

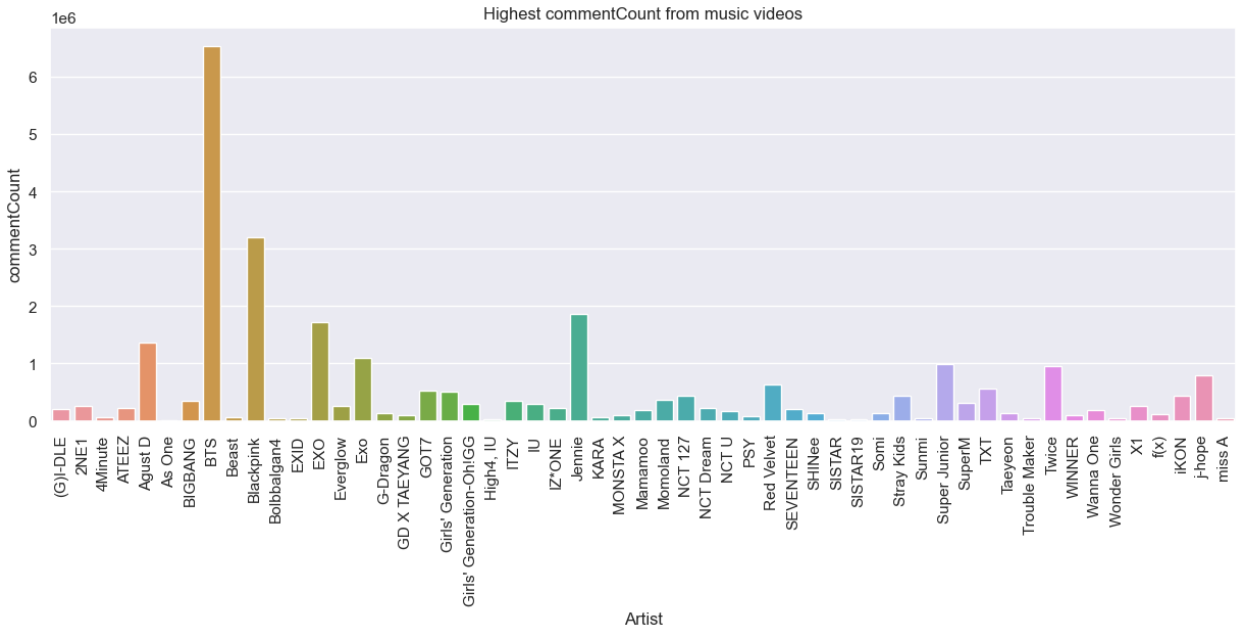
- The comparison between artists in groups or solo, and the gender of the group

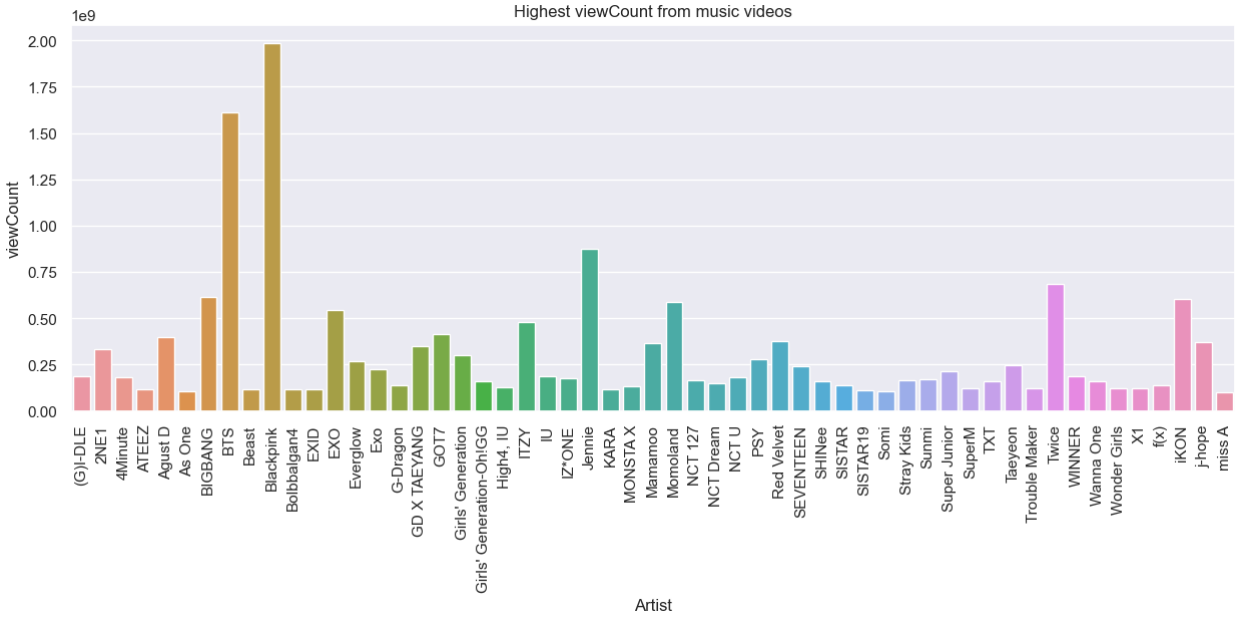


From the top left mosaic plot, it shows that music videos released by groups have more success than those by solos and gender has nearly no influence on the distribution in groups. However, when it comes to solos, it's obvious that there are more female artists than males. Furthermore, from the top right figure, we can know those solo artists who debut in a group have a higher chance to succeed.

- The highest likeCount, commentCount, and viewCount from these music videos







From the three figures above, the top ten in “Highest viewCount” are ‘Blackpink’, ‘BTS’, ‘Jennie’, ‘Twice’, ‘BIGBANG’, ‘iKON’, ‘Momoland’, ‘EXO’, ‘ITZY’, ‘GOT 7’. In this list, only Jennie is a solo artist and she is in the group ‘Blackpink’, which can corroborate the analysis in the last section. Compared viewCount with likeCount and commentCount, it is clear that they have a positive correlation since they have a similar pattern.