

Toronto Metropolitan University

# Social Media Analytics

Analyze data from Twitter by using Netlytics and Excel

Ethan Duong  
9-12-2022

## Section 1: Topic Selection and Data Collection

**Topic: How people reacted to BTS newest song, Yet to Come, on Twitter.**

### Introduction

This report focuses on analyzing the influence of Yet To Come, a BTS newest song, to Twitter's user all over the world. Additionally, this project will also investigate the type of mobile operating system of BTS's followers. The data for the report is collected from Twitter via Twitter API by using Netlytic. This data and information will only be used for this assignment purposes.

### Basic Assumptions

There are three assumptions will be used before analysis. First of all, people who tweet about Yet To Come is reasonably considered BTS's followers. Additionally, each author of a tweet is an individual, even the fact that one user can have multiple accounts. Twitter users who use web application to tweet is not considered as both iOS and Android user. They will be classified web application users.

### Search Query Criteria

For this assignment, these criteria will be used in Twitter Search Query: #BTS #Yettocome @BTS\_twt lang:en -filter:replies min\_retweets:1 min\_faves:1. I will explain each component and the reason for using it. First of all, #YetToCome hashtag is used to indicate the name of the song, which is the main criteria to gather tweets that mentioned information related to the song. Moreover, #BTS is a hashtag of band name to ensure the song comes from BTS band. @BTS\_twt: this indicates BTS official account. Lang:e: ensure tweets are in English, which helps us to be more convenient while analyzing this project. Another criteria is "-filter:replies", which helps to make sure the data is not overwhelmed by the large number of replies from BTS fans. This could make our data more diverse as Netlytic only allows us to collect 2500 tweets. Lastly, "Min\_retweets:1" and "min\_faves:1" is used to set the minimum number of like and retweet to 1. This could reduce the number of tweeter bots as Twitter bots usually automatically like and retweet posts, which likely to have no like or any retweet. Setting minimum number of likes and retweet will definitely reduce the large number of Twitter bots. Before importing the data, I was considering some other criteria, but decided not to use as it can cause confusions and bias in the collected sample data. "#TheMostBeautifulMoment" (album of Yet To Come) is one of it. This hashtag is not used as it can easily mention other songs from this album while this report only focuses on the impact of Yet To Come song. Secondly, "#Sugar" (BTS member) also not be used as this hashtag might focus on individual member's fan rather than the whole band's fan.

### Analyzing Process

This section outlines the process of collecting and analyzing the data. First step is choosing topic and brainstorming most suitable searching criteria. When the topic is chosen and criteria are clear, the next step is to configure criteria to gather data via web based Netlytics after linking Twitter account. Before downloading, I tested the query to measure the quantity and the quality of data. In detail, I was looking for a range of data, where there is diversity in public date and tweet types. The last step is downloading collected data and uploading data to Google Sheet to perform Exploratory Data Analysis by Explore function.

## Section 2: Dataset Description

This section will provide several significant parameters that worth considering and explanation about the sample data. First, according to the Table 1 below, we can see that there are 24 data fields that give us different aspects of 2500 tweets and their authors. There are several data fields that will be used frequently in this assignment, such as author, favorite\_count, pubdate, tweet\_type, user\_friends\_count, user\_followers\_count and source. I will also add a column of definition to explain about each data fields.

Data fields	Type	Definition	Data fields	Type	Definition
Id	Int	Row id	In_reply_to_screen_name	Text	
Tweetid	Int	Tweet Id	In_reply_to_user_id	Int	Id of replier
Guid	Text	Link to guid	In_reply_to_status_id	Int	Id of replier status
Link	Text	Link to the tweet	Retweeted_screen_name	Text	Name of retweet
Author	Text	Author name	Retweeted_status_id	Text	Retweeted status id
Title	Text	Tweet title	Retweeted_user_id	Int	Retweeted user id
description	Text	Tweet plot	User_id	Int	User id
Pubdate	Date	Tweet publishes date	Profile_image_url	Text	Link to profile image
Source	int	User application	User_status_count	Int	Number of user status
Favorite_count	Int	Number of likes	User_friends_count	Int	Number of user friends
Retweet_count	Int	Number of tweets	User_followers_count	Int	Number of user followers
lang	Language	Tweet language	User_created_at	Date	User profile created time
Quoted_text	Text	Quote of tweet	User_bio	Text	User bio
Tweet_type	Int	Type of tweet	User_verified	Boolean	User verification status

Table 1: Data output fields

There are several parameters that should be noticed after collecting the data. First of all, there are a total of 2500 tweets. In these tweets, there are 2222 original tweets and 278 quote tweets, which is showed in tweet\_type field. Additionally, these 2500 tweets are belonging to 1010 different users. User had a highest number of like at 14572 likes and highest number of followers at 408178. It is also worth noting that tweets that highest number of being retweeted again at 4243. There are no constrains in geographic aspects as I want to collect tweets from all around the world to measure the impact of Yet to come in the global scale. At the end, the public date of these tweets starts from 10 a.m. September 29 to 3 p.m. October 15, while the data is collected and exported in 3p.m October 15.

According to these parameters, we can see that, with only 7 days, the song have quickly spread on Twitter with a large number of interaction due to the large number of retweets and likes. There are a significant number of BTS's fan on Tweeter.

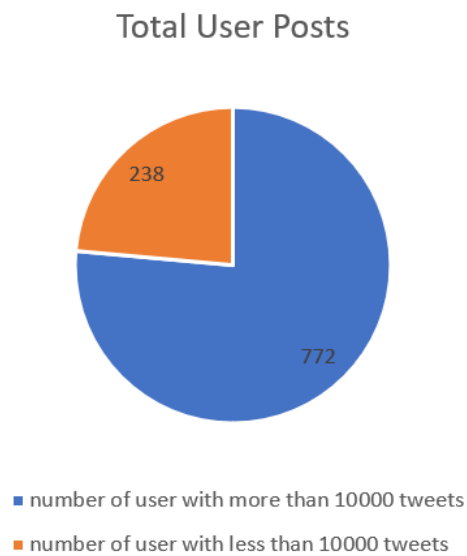
## Section 3: Exploratory Data Analysis (EDA) (approximate length: 3 pages)

In this section, we will explore 3 different areas to investigate why and how BTS newest songs spreading on Twitter as well as the type of operating system used for posting these tweets. Three areas are: different types of tweets and its popularity, audience interaction and posting times, and user applications. At the end, the report will point out who will interested in these data and the reasons for it.

### Tweet Types and its popularity

According to the Tweet Type chart (appendix), the number of original tweets covers most of the proportion with 2222 tweets. Meanwhile, the quote tweets, which allows user to tweet another person's Tweet with their own comment added, just take around 11%, which is only 278 tweets. This could mean that most fan decided to post their own tweets about the song rather than quoting.

Let dig deep into the users, who post tweets related to BTS newest song. According to the data below, more than 75% of user has more than 10000 posts in their posting history. This is mind-blowing as 10000 posts indicates that the posters are significantly active on this social media platform. Meanwhile, 238 authors seem to have less than 10000 posts. These authors still play a vital role in helping the song spreading on Twitter. In general, based on the data collected, we can argue that BTS followers are remarkably active on Twitter and they contribute significantly to the success of the song.



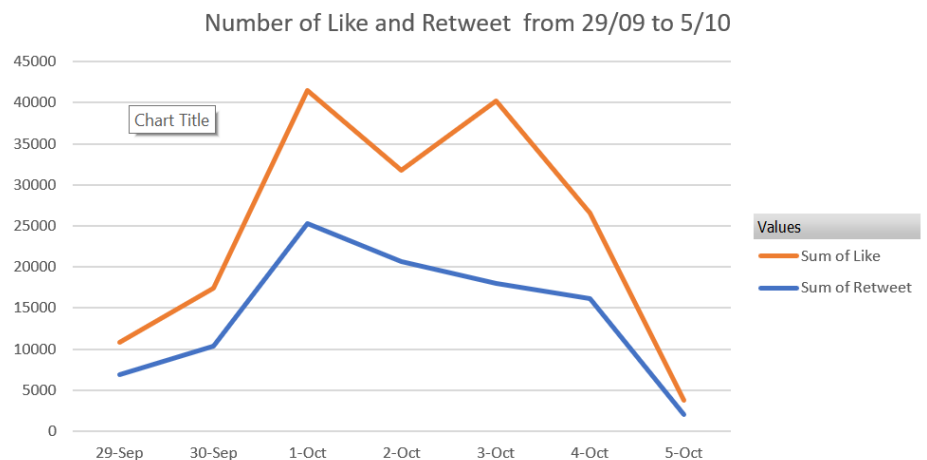
Additionally, more than 60% of these followers stay in a strong network with more than 300 friends and followers (User followers and popularity, Appendix). They can be considered as the main force that help the song spreading remarkably on Twitter. The other 405 users only have less than 300 friends as well as 300 followers on Twitter. Even having less than 300 friends and followers, they still made a huge contribution to the spread of Yet To Come song.

In conclusion, BTS has the strong Twitter followers, who like to post their own tweets on this social media than quoting other people tweets. They are considered strong influencer on social media. That is why their posts get so many interactions, which helps BTS song, Yet To Come, spreading on social media in a short period of time

### Interaction and Posting times

The Number of Like and Retweet chart is clearly indicated how audience react with BTS song. The orange line indicates the sum of all the likes of tweets related to the song, while the blue line shows the total number of retweets. According to the chart, the total number of like peaked at 40000 likes in Oct 1, and Oct 3. While the total of retweet only peaked in Oct 1 at 25000 tweets. It seems like the songs was mostly spreading from Oct 1 to Oct 3, then went down quickly after that.

The data also indicates that the total number of likes, and followers peaked in the same time period (October first). After that, the data reveals the decrease in both line, dropped less than 5000 in the last day of the given period. This fluctuation of both variables might happen as the result of BTS marketing campaign.



The Number of posting time chart (appendix) indicates that people who post the tweets related to the song could be categorized into 3 types: posted 1 time, posted 2 to 4 times and post more than 4 times. Remarkably, 40% of the tweet are posted more than 4 times. This showed that there are a large number of fans regularly posting about the song in just a few days.

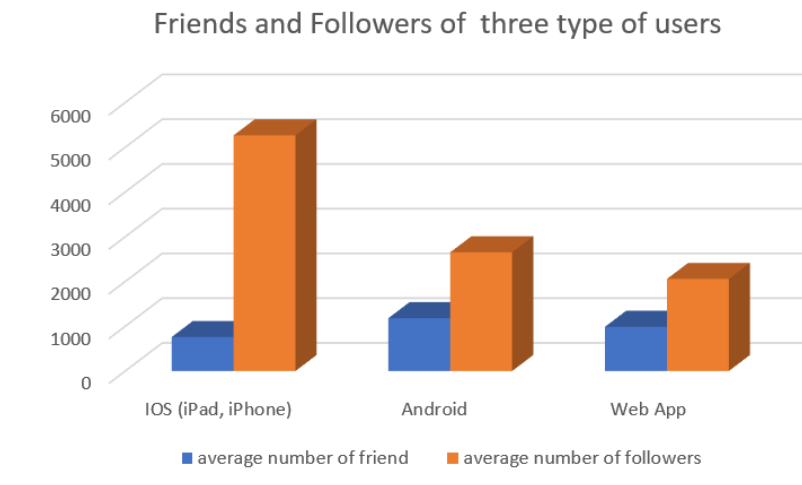
This pie chart also demonstrates that there is the same amount of portion when it comes to author posting only one time and author posting two to four times. The portion for these both author groups are around 30% of the total number of authors. Even though these authors did not post more than four times, they still considered a strong force that help the song spreading significantly on Tweeter.

In conclusion, the popularity of Yet To Come on Twitter is highly depend on the fans. Most of them are post frequently, which increase the number of likes and retweets for the song. Now, let explore what operating system that these fans used for posting.

## User Application

As we can see from the application report (appendix), in 7 days, most of the user was using Android application to post their tweets. In the opposite, there are only nearly 600 tweets are posted by using IOS system. This could happen as the result of the fact that most of BTS fans are Koreans, who might still prefer using Samsung (biggest Android application) rather than using Apple products.

The data from the below table shows that Android author and Web App author have a similar pattern with average number of friends and followers are around 1000 and 2000 respectively. Meanwhile, IOS's average followers reached significantly to 5000, but their average number of friends is only about 500.



Thanks to the data from two chart, we can conclude that BTS fans are mostly Android user, which make perfect sense as the most famous application is Korean is Samsung (John, 2022), which run on Android application. However, IOS BTS's followers also play a significant role in increasing the popularity of the song among Twitter users as they have much higher of followers, compared to Android and Web applications.

## So, who might user the data?

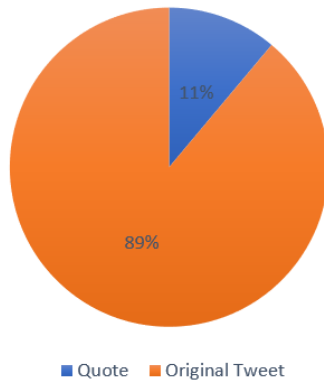
The information generated by this reported can be helpful for various groups. First of all, BTS artist manager can benefit from this as it helps them to measure the popularity of the song and navigate the area, which the song has the most impact on. With this information, they can decide to execute their marketing campaign in the most effective way. Secondly, application provider also gains advantages by using this information as it will help them to understand more about what their customers' need and care about, so that they can be provided suitable advertiser, which is more related to their interests. As an example, Samsung can run more ads that related to BTS products to gain better outcomes. Lastly, this information might be most beneficial for BTS product sellers. This data is likely to be directly related to their customer target. These sellers can know where and when there is a highest need for the product so they can target this area in the hot time. This could increase their sale significantly that target general customers.

## References

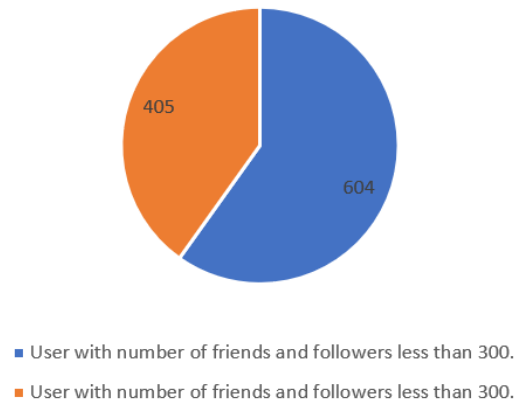
John, 2022, *Smartphone Use in Korea is Getting Out of Hand and Hurting Koreans*, <https://www.seoulz.com/smartphone-use-in-korea-is-getting-out-of-hand-and-hurting-koreans/#:~:text=In%20fact%2C%20a%20little%20under,the%20Galaxy%2021%20in%202021.>

## Appendix

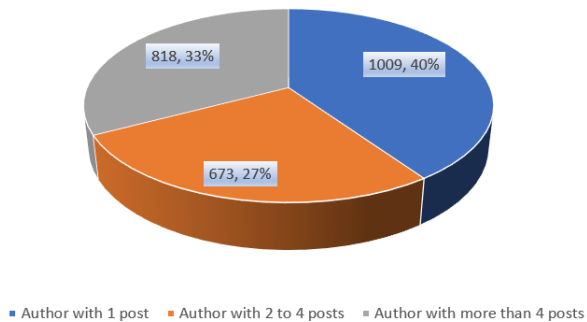
Chart Area **Tweet Types**



**User followers and popularity**



**number of posting times**



**Applications report**

