

Factors Affecting Number of Views on YouTube Videos*

An Exploration of Official Music Videos uploaded by Genierock in 2015

Samita Prabhasavat

30 April 2022

Abstract

Both domestically and internationally, Thailand's entertainment and media sector is expanding. This report is based on information obtained from Genie Records' marketing department. The data is cleaned, altered, and investigated, yielding solid proof that both the time and the song itself have an impact on the number of views on a music video. The impact of the time and the song on the number of views on the music video will be discussed in this paper.

keywords: music, YouTube, views, date, Thai entertainment and media industry

1 Introduction

GMM Grammy Public Company Limited is Thailand's largest media and entertainment conglomerate. It controls 70 percent of the Thai entertainment market. The organisation has 15 music subsidiaries that specialise in various kinds of music. Genie Records, founded in 1998, is a Thai record company and a subsidiary of GMM Grammy that specialises in rock music. Klear, Cocktail, and Bodyslam are among the 23 Thai artists in the company.

The data for this report was gathered by Genie Records' marketing department with the initial goal of increasing engagement on official music videos posted on YouTube. From 2015 through 2021, the dataset includes the date the video was uploaded, the name of the band, the name of the song, and the number of views tallied in each month. The goal of this analysis is to look into the aspects that may influence how many people watch each music video. The major criteria in this report will be the time since the song was released and the song itself.

The data was received from Genie Records' marketing department and was altered to ensure that it was in the correct format. The data was then analysed, leading to the conclusion that the time since the song was released as well as the song have a significant impact on the number of views. In this work, a prediction model is proposed to further investigate factors affecting the number of views and provide insights for the marketing team. Understanding these elements can aid us in comprehending the effects of events that occurred during that time period, as well as the direction of the Thai entertainment industry, which contributes significantly to the Thai economy.

This report will first, look at the data collected by the marketing department of Genie Records and the methodology used to collect the data in Section 2 to provide readers with the context of the data used in this report. Second, it will discuss the insights found from exploring the data in Section 3. Last, it will conclude the insights found from exploring this data set and provide plans for further exploration of this data set in Section 4.

*Code and data are available at: https://github.com/PSamita/music_video.git

2 Data

The R statistical programming language (R Core Team 2020) is used to analyse the data in this report. The here package (Müller 2020) is used to access files from another location in the same R Project. The janitor package (Firke 2021) is used to tidy up the column names. Data manipulation is done with the tidyverse package (Wickham et al. 2019). To make a table from a data frame, use the knitr package (Xie 2014). Graphs are made with the ggplot2 package (Wickham 2016). The RColorBrewer (Neuwirth 2014) is used to assign colours to graphs and figures. (Xie 2016) The bookdown package is used to cross-reference graphs and figures.

2.1 Overview of the data

The marketing department of GMM Grammy Public Company Limited provided the data for this study. The data was collected in order to increase the engagement of official music videos released by artists signed to the GMM Grammy label and its sublabels. The dataset includes the number of monthly views of four official music videos, the title of the song, and the name of the band collected from 2015 to 2020.

2.1.1 Data Sources

GMM Grammy Public Company Limited is the largest media conglomerate and entertainment company in Thailand and is home to many of Thailand’s top artists, such as Bird Thongchai, Bie Sukrit, Tata Young, and Bodyslam. The company is also active in concert production, artist management, film and television production, and publishing, in addition to music. GMM Grammy has 15 music subsidiaries focusing on different styles of music, including Genie Records, Grammy Gold, White Music, Musiccream, and Genelab.

Genie Records is a Thai record label that concentrates on rock music and is a subsidiary of GMM Grammy. Genie Records was founded on January 1, 1998, by GMM Grammy, and Wichian Rerkpaisan was appointed the company’s first managing director. Sumet and the Pang, a musical duo who became popular with their first album, was one of the first artists released by the record label, which covered a variety of music genres. It also introduced new performers to the entertainment industry, like Tanawut Chawathanavorakul, Palaphol Pholkongseng, Noppasin Sangsuwan, Budokan, Venus, and Dajim, who went on to make hit songs. It went up against More Music, the country’s premier rock music record label at the time, which featured experienced rock performers like Asanee–Wasan, Loso, Nuvo, and others.

The record label was rebranded “Genie Rock” in 2012 in order to concentrate on producing rock songs. Labanoon, The Yers, Palmy, 25 Hours, Paper Planes, and Potato, who returned to Genie Records after a brief stay on We Records and WerkGang, were among the new and returning rock musicians. As of 2022, there are 23 artists under Genie Records, including 7 soloists and 16 bands.

Four songs by four bands under Genie Records were chosen to be used in this report. The first song is ‘Leave’ (Thai: Ting Wai Glang Tang) by Potato. The second song is ‘Soulmate’ (Thai: Koo Chee Wit) by Cocktail. The third song is ‘Object’ (Thai: Sing Khong) by Klear. The last song is ‘Never’ (Thai: Mai Keay) by 25 Hours. The music videos for these songs were released on Genierock’s official YouTube channel.

Potato is a Thai rock band that was established in 2001. Patchai Pukdesusook, guitarist Teekatas Taviararakul, bassist Piyawat Anukul, drummer Kan Uamsupan, and pianist Kiattiyos Malathong make up the band’s current lineup. It has eight studio albums, one extended play, and a number of compilation albums to its credit. It is regarded as one of Thailand’s most successful rock bands. The chosen song, “Leave,” was released in 2015 but was only included in a Chudteejed album in 2019. The song is about a couple’s split, and how it’s intended to be this way, and that it’s best for both of them to break up.

Cocktail is a Thai rock band, founded in 2002. The band currently has four members, including lead vocalist Panthaphon Prasarnrakit, guitarist Chawarat Hadsunthathai, bassist Krirkkiat Sawangwong, and drummer Philips Prem Sirikorn. It has released seven studio albums, one extended play, and several compilation albums. It is considered one of the most successful Thai rock bands, with lots of hit songs. The chosen

song, “Soulmate,” was released in 2015 but was first heard by the public at the lead vocalist Panthaphon’s wedding in 2012, as it was the song he wrote for his wife. The song expresses the feeling of how much the song writer loves his lover and that he is willing to do anything for her.

Klear is a Thai rock band, founded in 2007. The band currently has four members, including lead vocalist Rannapun Yungyeunpoonchai, guitarist Nathawat Sangvichit, bassist Keyapatr Sangwonpailert, and drummer Nut Nilvichian. It has released six studio albums and one extended play. It is considered one of the most successful Thai rock bands, with lots of hit songs. The chosen song, “Object,” was released in 2015 on the Silver Lining album. The song expresses the feelings of a person who has been unloved by everyone in his or her life until that person gets to meet someone who really loves him or her, and the person feels very thankful for that love.

25hours is a Thai rock band, founded in 2008. The band currently has five members, including lead vocalist Somphol Rungpanich, guitarists Prateep Siri-issarnunt and Piyawat Mekreu, bassist Aeksiri Kumbungpai, and drummer Krittapong Sakulnarmanake. It has released four studio albums. The chosen song, “Never,” was released in 2015 on the Mom & Popshop album. The song expresses the feelings of missing someone who is not in your life anymore.

2.1.2 Data Collection Methodology

The marketing department of the GMM Grammy, including the data scientists and data analysts, collected the view count provided on YouTube Analytics. The number of views was recorded at the end of each month after the song was released. Normally, the marketing team will analyze the number of views along with data collected from the comments under each music video and data collected from other social media platforms.

On YouTube, a view is defined as: 1) a viewer-initiated intended play; and 2) a video that has been despammed by YouTube’s algorithm. To put it another way, the video was viewed by a human on a single device, and YouTube acknowledged it.

2.1.3 Error Detection

The data was already reviewed and cleaned by the data team because it was acquired by the Marketing Department for business purposes. As a result, there is no error detected in this dataset as expected. There are some null values in the dataset, which is because the music video has not yet been uploaded on YouTube, so it does not have a number of views.

2.1.4 Disclosure Control

The Marketing Department of GMM Grammy contributed this dataset only for the purpose of this study. For business purposes, no additional information will be shared or released to the public.

2.2 Variables

There are five variables in the original dataset:

- Date
 - The date the number of views was recorded.
- Song
 - The name of the song in Thai.

- English Name
 - The name of the song in English.
- Band
 - The name of the band.
- Views
 - The number of views

2.3 Strengths and Weaknesses

The strength of this dataset is the quality of the data. The data in this dataset is very accurate and clean because it was collected and cleaned by data analysts and data scientists in the Marketing Department. The dataset is extracted from the real dataset the Marketing Department uses to make business decisions and increase engagement.

The fact that the data cannot be traced back is a weakness in this dataset. Because the information is not available to the public, it is impossible to verify that the numbers of views are correct. However, if applicable, the data's validity, outliers, and null values can still be recognised using statistical methods.

2.4 Focused Aspects

This paper will focus on the numbers of views on four music videos uploaded on the Genierock channel from 2015 to 2020 by month.

2.4.1 Data Exploration

Table 1 shows the example of cleaned data generated from the raw data provided by the Marketing Department of GMM Grammy. The table includes 5 variables which are year, month, song, band, and views.

Table 1: Number of Views on the Official Music Videos from 2015 to 2020

Year	Month	Song	Band	Views
2015	03	Leave	Potato	NA
2015	03	Soulmate	Cocktail	22085749
2015	03	Object	Klear	NA
2015	03	Never	25hours	NA
2015	04	Leave	Potato	NA
2015	04	Soulmate	Cocktail	51963702

The line graphs in Figure 1 represent the data provided by the Marketing Department of GMM Grammy which is used in this report to further explore the factors affecting the number of views on the official music videos. The y-axis shows the number of views counted on official music videos uploaded on YouTube while the x-axis shows the year between 2015 and 2020. Each line graph shows the trend of the number of views for each song.

A box plot is used to assess the data's normality before selecting the suitable statistical summary for this dataset. The data points in this dataset lie between the upper thin horizontal line and the bottom thin

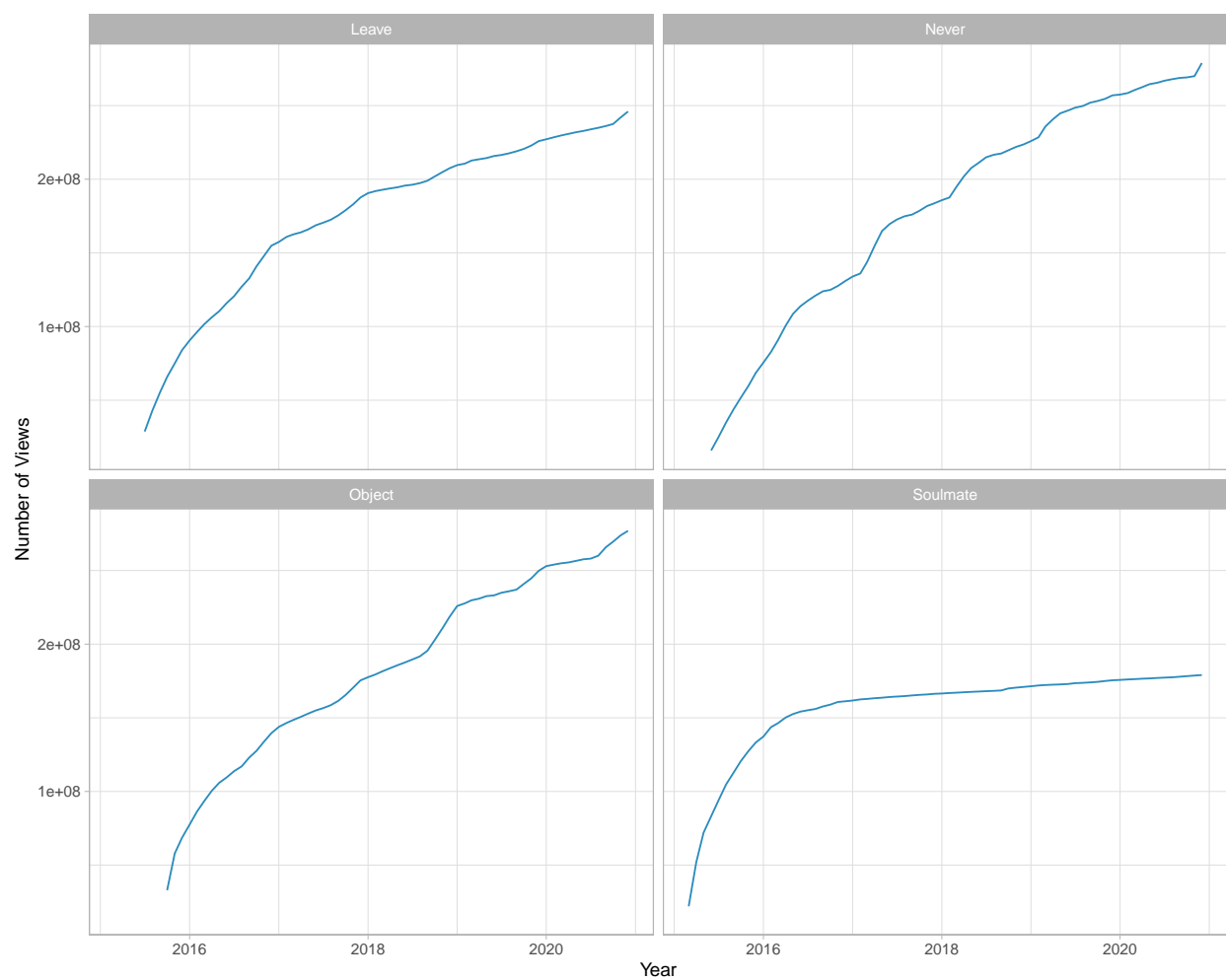


Figure 1: Number of Views from 2015 to 2020 by Song

horizontal line, according to a box plot in Figure 2. The median number of views from this dataset is represented by a thick black horizontal line across the blue box, with 50% of the data above the median and 50% of the data below the median. Within the blue box, 25% of the data is above the median, while the remaining 25% is below the median. The thick black horizontal line, or the median line, in this box plot is a little low, indicating a possible skew in the data. Furthermore, six outliers were found in this sample. However, it is confirmed that they are all correct, so all of them are not excluded from the dataset.

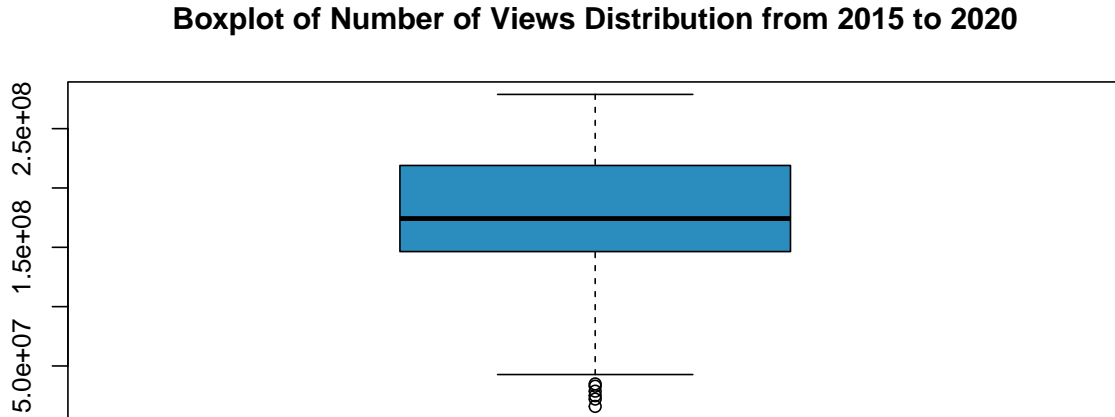


Figure 2: Boxplot of Number of Views Distribution from 2015 to 2020

In order to further confirm the non-normality of the data from looking at the box plot above, a histogram is plotted as in Figure 3. It can be seen that the histogram is left skewed and does not look like a typical bell-shaped diagram, which indicates that the data is not normally distributed.

The median has been chosen as a measure of central tendency for exploring this dataset as the data is not normally distributed.

3 Results

3.1 Number of Views by Year

A bar graph in Figure 4 shows the median number of views on four chosen music videos uploaded on Genierock’s official YouTube channel by year. It can be seen that the median number of views increased every year from 2015 to 2020. In 2020, the median number of views is the highest among the six years with around 250 million views, followed by 2019, 2018, 2017, 2016, and 2015 at around 50 million views, in that order.

3.2 Number of Views by Month

A bar graph in Figure 5 shows the median number of views on four chosen music videos uploaded on Genierock’s official YouTube channel by month. It can be seen that the median number of views hit the peak around the end of the year to the beginning of the year. The median number of views in March is the

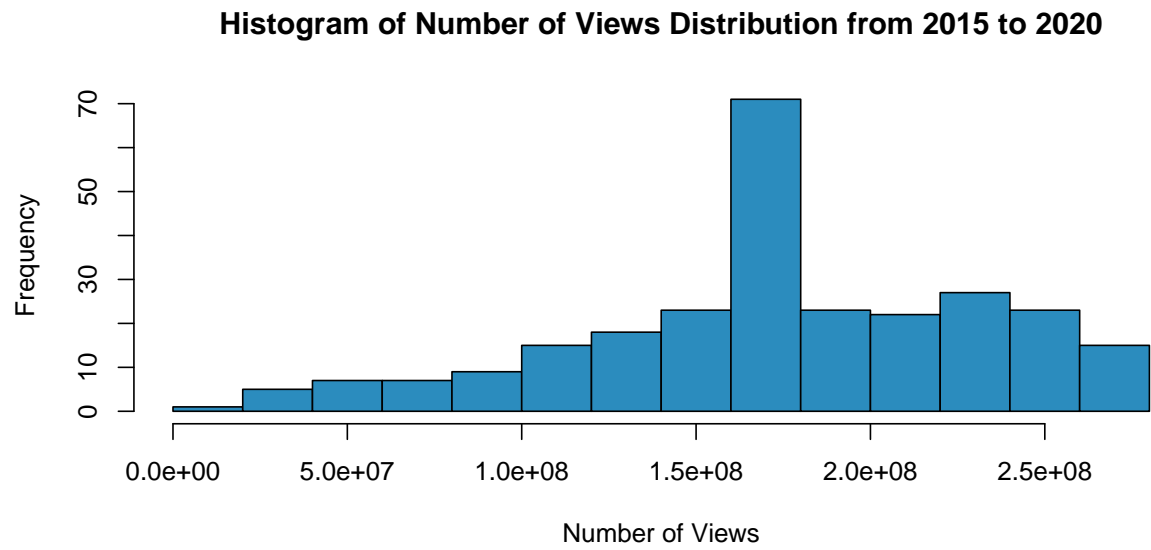


Figure 3: Histogram of Number of Views Distribution from 2015 to 2020

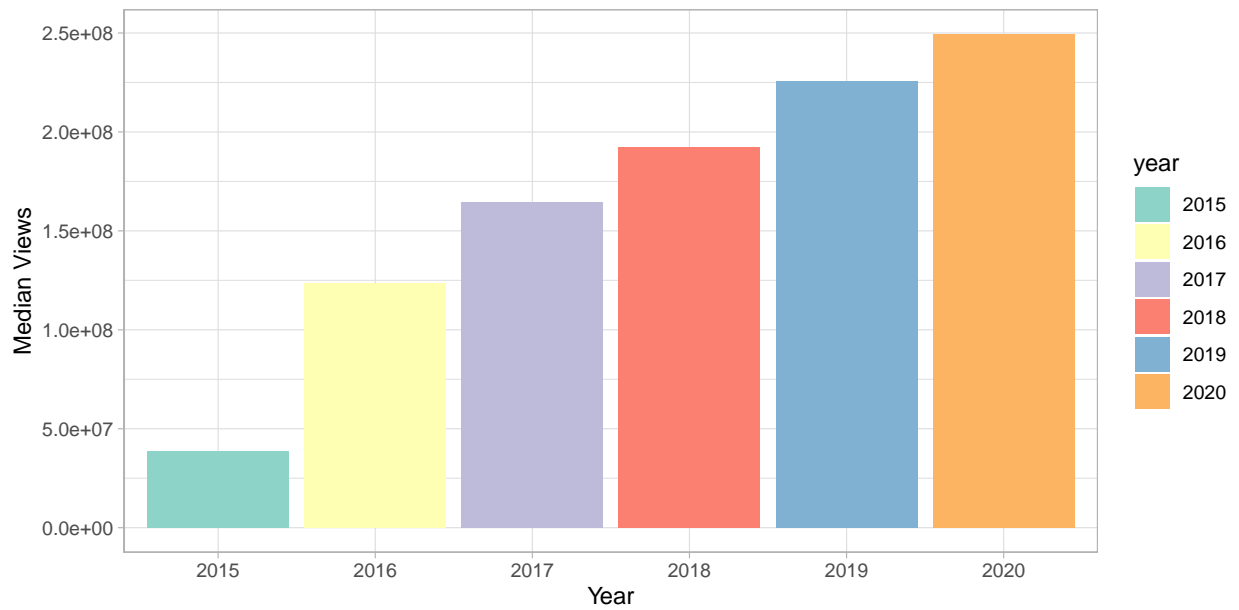


Figure 4: Number of Views by Year from 2015 to 2020

lowest. The median number of views increases from March until reaches its peak at the end of the year and remain high until February. However, the median number of views for every month are still within the range of 150 million to 200 million views.

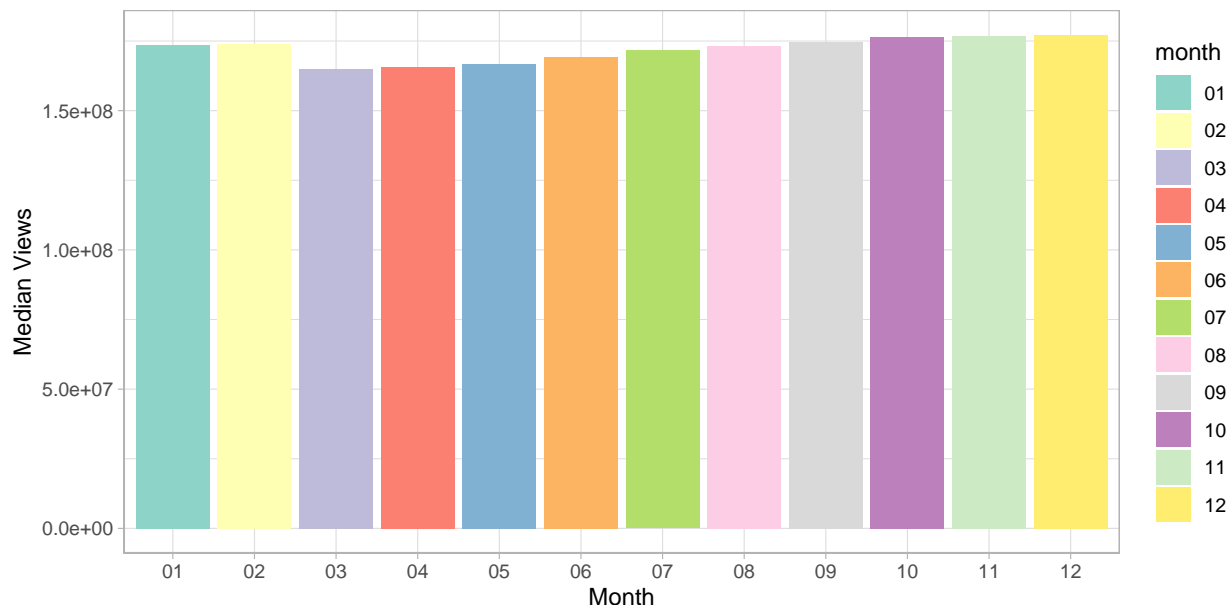


Figure 5: Number of Views by Month from 2015 to 2020

3.3 Number of Views by Song

A bar graph in Figure 6 shows the median number of views on four chosen music videos uploaded on Genierock's official YouTube channel by song. It can be seen that, between 2015 and 2020, 'Leave' has the highest median number of views compared to the other three songs at almost 200 million views. 'Never' has the second highest median number of views at about 180 million views. The third highest median view belongs to 'Object' at around 175 million views. 'Soulmate' has the lowest median number of views at approximately 160 million views.

3.4 Number of Views by Year and Song

A bar graph in Figure 7 shows the median number of views on four chosen music videos uploaded on Genierock's official YouTube channel by year and song. It can be seen that the overall trend of the median number of views is positive. In 2015, the median number of views for 'Object' is zero because the music video was uploaded in the later half of the year. Most of the view count for 'Object' in 2015 is zero which resulted the median equal to zero. The medians number of views for 'Soulmate' are the highest compared to other songs in 2015 and 2016. However, from 2017 to 2020, 'Never' has the highest medians number of views.

3.5 Number of Views by Month and Song

A bar graph in Figure 8 shows the median number of views on four chosen music videos uploaded on Genierock's official YouTube channel by month and song. It can be seen that the medians number of views of 'Leave' are the highest compared to other songs from January to April. On the other hand, 'Never' has

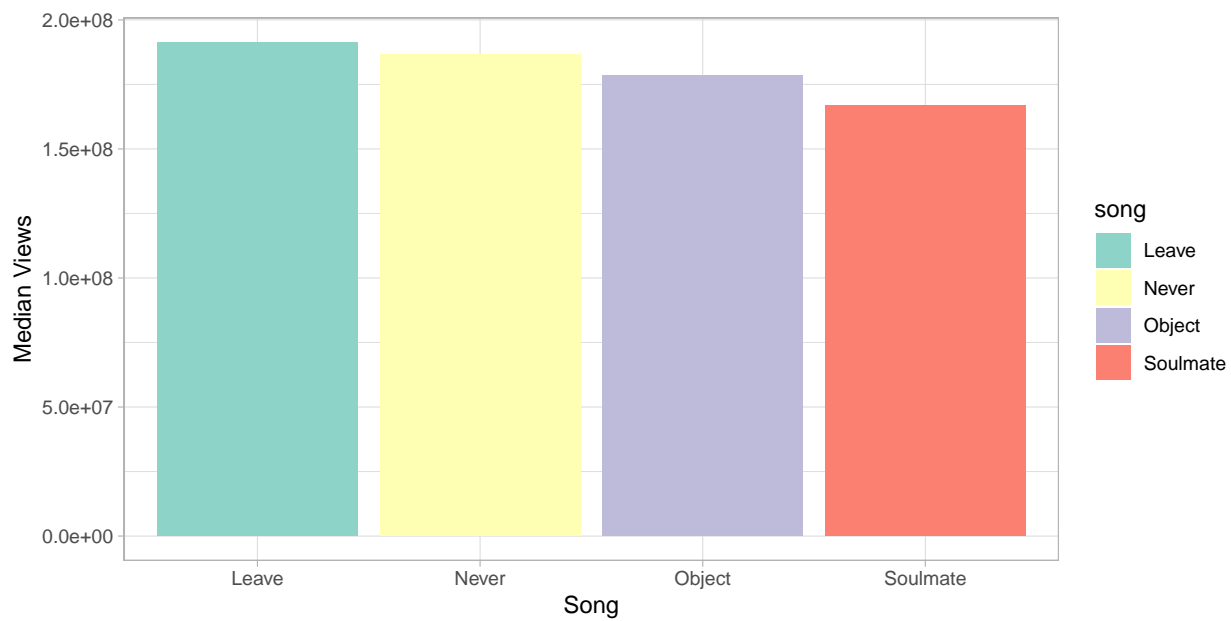


Figure 6: Number of Views by Song from 2015 to 2020

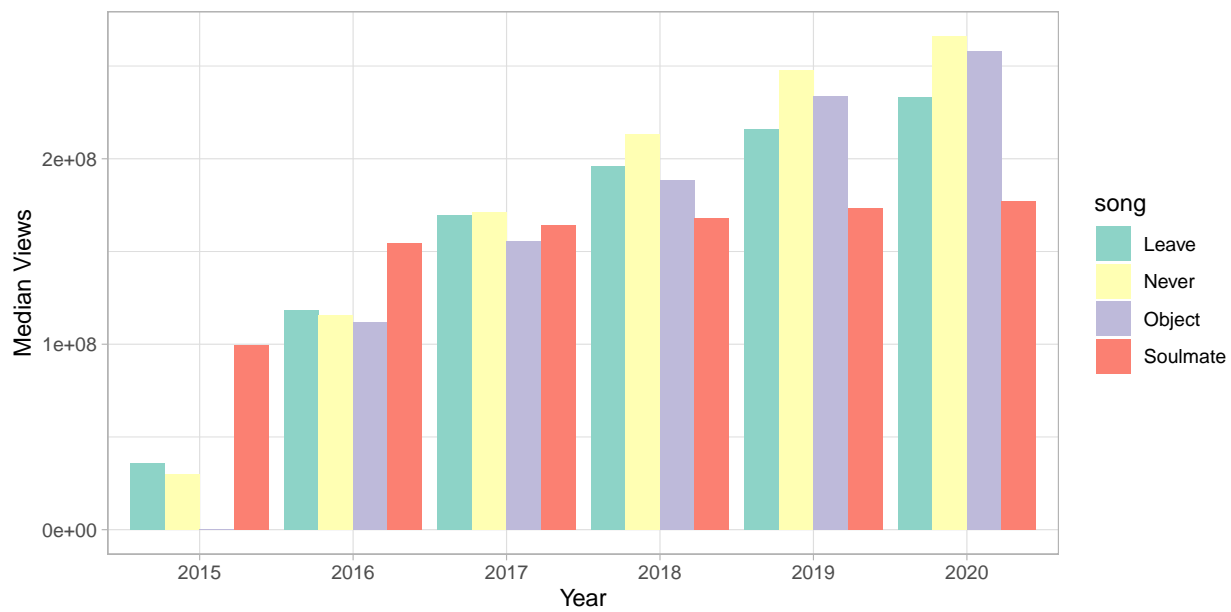


Figure 7: Number of Views by Year and Song from 2015 to 2020

the highest medians number of views in the rest of the year. The medians number of views for ‘Object’ are high from September to February, while the medians number of views for ‘Soulmate’ are higher than usual in February, April, June, and September to December.

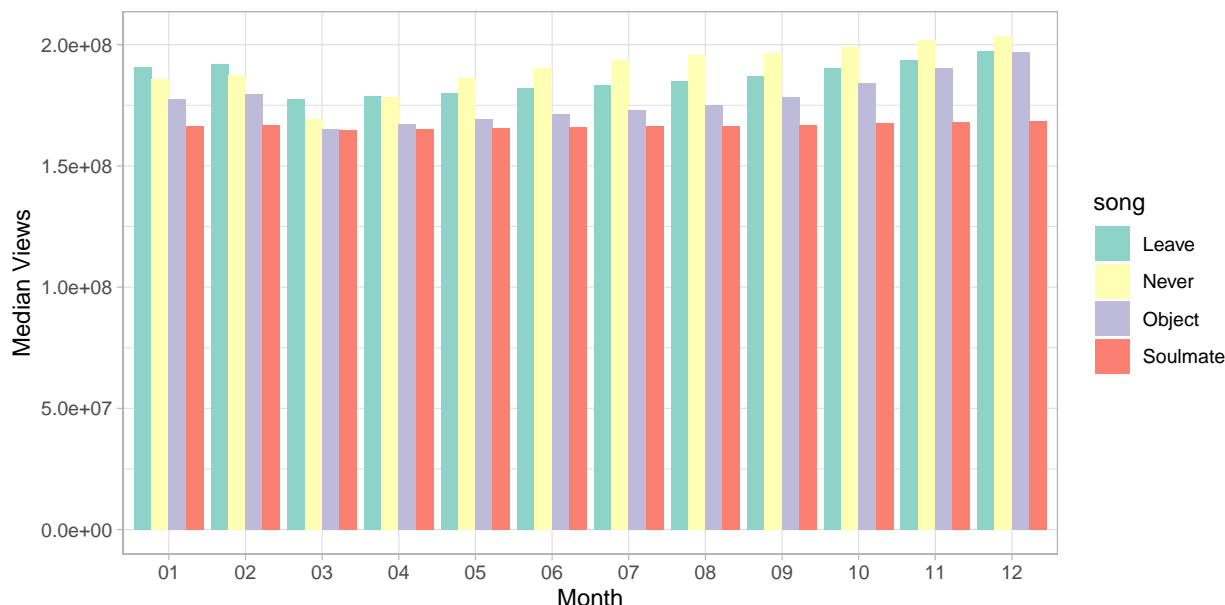


Figure 8: Number of Views by Month and Song from 2015 to 2020

4 Discussion

4.1 Methodology

This paper was created by downloading raw data provided by the Marketing Department of GMM Grammy as a csv file into RStudio software. The data was then filtered to keep only variables of interest and ensured that all data was consistent and in the right format. For example, the number of views should be in numeric format. If there is data containing characters, that data must be examined. Moreover, outliers were detected if applicable. In this case, six outliers were detected. However, after consulting with the data analyst in charge of the dataset, the data is all correct and the outliers should not be removed in this case.

After the data is ready to be used, the next step is data exploration. The normality of the data was first tested using a box plot and a histogram, which led to the decision of which statistical summary would be utilised in this study. Second, to investigate the data, a statistical summary was used. Third, graphs are created to aid in the finding of data insights. Finally, the plan for the future study is provided.

4.2 The Impact of Time on Number of Views

Figure 4 shows that the number of views increases as the year progresses. This is definite because the more people that click on the music video and watch it, the larger the amount of views will be. With each passing year, more people get access to the internet, personal computers, or mobile phones and tablets, making it easier to watch YouTube videos from anywhere. In addition, since 2017, there have been music shows in Thailand. Artists will perform the song they just released at the music show, and the artist with the most music video views combined with album sales for that week will win an award. For this reason, fans are encouraged to increase the number of views on the official music video.

Furthermore, as the year progresses and as a result of globalization, businesses will have more capacity and resources to market the music both domestically and worldwide. As a result, more individuals are visiting Genierock's YouTube account, following the company's musicians, and watching official music videos to show their support. Since 2016, GMM Grammy has been attempting to break into the worldwide market. In 2017, the good outcomes began to emerge. In China, Hong Kong, Taiwan, Indonesia, Malaysia, and the Philippines, many GMM Grammy artists have a sizable fanbase audience. Especially in Southeast Asia, where many prefer YouTube over other platforms like Apple Music or Spotify for music streaming.

Looking at Figure 5, the number of views on music videos is not that different between months. However, according to Figure 1, "Leave", "Object," and "Never" have fluctuations in the trend, which means there are months that people watch the music videos more than the rest of the months. According to Figure 8, there are fluctuations in the trends for every song. In Thailand, music festivals usually occur in the winter, when the weather is nice and not too hot. Winter in Thailand starts in October and ends in February. When people hear a song at a music festival, they usually go to YouTube and search for it to listen to it again because of post-concert depression, or PCD. PCD occurs when a person has a happy time, like going to a concert, which makes the body release serotonin, dopamine, endorphins, and oxytocin. However, after the concert, the sudden stop of those chemicals makes the person feel depressed and want to go to the concert more. Because the concert just ended, people usually listen to the song they just listened to during the concert to keep the memories and feelings.

4.3 The Impact of Song on Number of Views

According to Figure 1 and Figure 8, there are fluctuations in the trend for every song, which are affected by the song itself. The lead vocalist first performed "Soulmate" at his wedding to express his feelings to his wife. As a result, the song has been played or performed very often at weddings. In Thailand, according to Thai norms, people normally hold their weddings in February, April, June, September, and October to December. It can be seen in Figure 8 that the number of views increases in these months. In 2019, two singers under GMM Grammy made a cover of 'Soulmate' which made the song go viral again. After listening to the cover version, people still go back to listen to the original version. Also, another singer under GMM Grammy chose 'Soulmate' to sing in "The Mask Singer." The Mask Singer is a television show where artists are invited to sing while wearing masks so that the audience does not know who is singing and will only judge the artist by his or her singing skills. The performance and the rearrangement of 'Soulmate' went viral, which made people revisit the original music video.

In 2020, the lead vocalist of Potato got married, which made people revisit the music video of 'Leave.' Even though the song is a sad song about breaking up, people still revisit the music video to congratulate the singer on his wedding.

The song 'Object' is often performed at music festivals where people enjoy listening to both love songs and breakup songs. The lead vocalist of Klear has invented a new performing trick in 2018 where she will call a partner of one of the audience members and sing the song for him or her. The audience really love this trick as they want to tell their lovers how thankful they are to be together.

The song 'Never' is often used around the graduation time of the students because the song is about missing someone who is not in your life anymore. In Thailand, the graduation ceremonies are held in either March, April, or May. However, the song is also widely used as a background music when someone passed away which can happen anytime.

4.4 Weaknesses

This report was created by the data that were already cleaned and provided by the Marketing Department of GMM Grammy. As a result, it is difficult to trace back to the original raw data which gives limitations to some statistical methods that cannot be used without knowing the original data. Also, there are only 5 variables in this dataset which limits the number of aspects that can be explored in this dataset.

4.5 For the future

In order to explore more about this data, it is better to have access to the original data. In that case, more variables can be used to explore about the factors affecting the number of views on official music videos.

Also, it would be better to compute the change rate in the number of views to see how fast the number of views increases by year and month to further explore the factors.

Moreover, after exploring all the possible factors, a prediction model will be provided to predict the number of views to help making marketing decisions or creating the plan to promote the songs

Appendix

A Datasheet

The dataset was created to help GMM Grammy make more accurate and precise business decisions which will help the company to gain more profits both domestically and worldwide, and will help contributing to Thailand's economy. The dataset is not available publicly, but was provided by the Marketing Department specifically for this study. However, there are limitations on exploring the dataset because of the limited amount of data.

The Marketing Department of GMM Grammy, including the data analysts and the data scientists, created and provided the dataset for this study.

The data was cleaned, manipulated, and explored using R Studio Software.

No direct funding was received for this study.

There are only one table in this dataset containing 281 rows of data which are all possible instances.

Each instance consists of the date, the name of the song, the name of the band, and the number of views.

The number of views are not available in some cases because the music video was not yet uploaded at that time.

There are no errors, sources of noise, or redundancies in the dataset.

The dataset is self-contained.

The data is confidential but is allowed to be used in this project.

The dataset does not identify any individuals.

References

- Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- Neuwirth, Erich. 2014. *RColorBrewer: ColorBrewer Palettes*. <https://CRAN.R-project.org/package=RColorBrewer>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- 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 Golemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. <http://www.crcpress.com/product/isbn/9781466561595>.
- . 2016. *Bookdown: Authoring Books and Technical Documents with R Markdown*. Boca Raton, Florida: Chapman; Hall/CRC. <https://bookdown.org/yihui/bookdown>.