National University of Singapore

AY 20/21, SEMESTER 2

GET1030: Computers and the Humanities

Spotify in the Covid-19 Era

Group Members:

Chang An Le Harry Jr (A0201825N)

Eldora Boo Puay Eng (A0221833M)

Enerio Johannine Salvilla (A0204690J)

Joel Yap Wen-Liang (A0184382Y)

Justin Eng Wei Jie (A0167946L)

Project Team:

13

Introduction

Impact of music

The effects of music on an individual's mental state and cognitive function have long been studied and observed, with a variety of test results published in research papers. In a study, Finnish researchers used fMRI scans to show how the brain responded differently towards distinct aspects of music (Alluri 2012). These aspects include timbre, tempo and tonality, which were associated in activating cognitive, motor and emotion-related circuits in the brain. There also exists other research that studied the effects that "sedative" or "stimulative" music genres have on performance, anxiety and concentration in students (Dolegui 2013). That being said, it is apparent how music has an effect on a listener's brain and emotion.

Noting how music today has become more readily available and accessible than ever, it has consequently become a common sight to see people listening to music just about anywhere. This could be during their daily morning commute, while exercising, while doing work or while studying… the list goes on. It is also a given that many people often listen to music to lower their stress levels and to improve their mood, so our group wanted to see whether there was any form of trend or correlation between the type, characteristics, and amount of music that people listen to, relative to a particular period of time.

The Covid-19 pandemic has no doubt changed the consumption patterns of consumers, including music. In fact, a recent study conducted in Belgium discovered that there has been an increase in nostalgia consumption based on the analysis of a dataset which contains over 17 trillion plays of songs on Spotify in six European countries (Yeung 2020). As such, we were inspired by this study to explore how Covid-19 has altered music consumption on Spotify for our project, and conversely, if we could analyse how music might be a predictor of group or individual sentiment.

Covid-19 in Singapore

Our chosen time period is between January 2020 and March 2021, coinciding with the timeline of the Covid-19 pandemic. In this project, we attempted to observe if there is an increasing trend of Spotify streams for a certain genre or type of music.

Due to the coronavirus outbreak around the world, much of the global population was faced with stay-at-home measures for both work and school, resulting in an inordinate amount of time spent at home. In the privacy of home, people will be able to play music while doing their work with impunity, and thus the usage of music streaming services such as Spotify has risen. At the same time, the overall mood of people forced to stay at home has gone down due to the frustration of being confined and unable to socialize physically. Thus, it is logical that people would make use of the stress-relieving qualities of music as well, similarly resulting in an increased usage of Spotify and other music streaming services.

While the correlation between an increased consumption of music and the onset of the coronavirus and lockdown measures seems fairly logical, our project aims to investigate if there is a deeper link between the Coronavirus safety measures and the trends of specific features in the music streamed by people during this time period, such as the genre, type or even language of songs. We will analyze data provided by Spotify and attempt to discover if such trends exist.

The impact of this study could be beneficial to helping people to de-stress during these troubling times, if a trend could be found about the type of music that is preferred during this time period, it can be applied as a guide to what type of music will help to lift their mood. This could also be applied at a higher level, where organisations such as shopping malls could make use of this research to play music with a more calming effect to soothe the visitors in public spaces.

Implementation

<u>The data</u>

In this visualization, metadata about the songs and artists that were on the Singapore top 200 charts during the Covid-19 pandemic and data behind the Covid-19 cases in Singapore are represented in green and red respectively. In addition, variables that contribute to the musicality of the songs are also represented in the visualization.

The data was sourced from Spotify, OurWorldInData (OWID) and Kaggle. The metadata for the songs and artists were scraped from the Spotify top 200 charts[1]. It is generated by analysing listeners' activity, transformed into the number of streams, and further evaluated to obtain the most listened artists and songs.

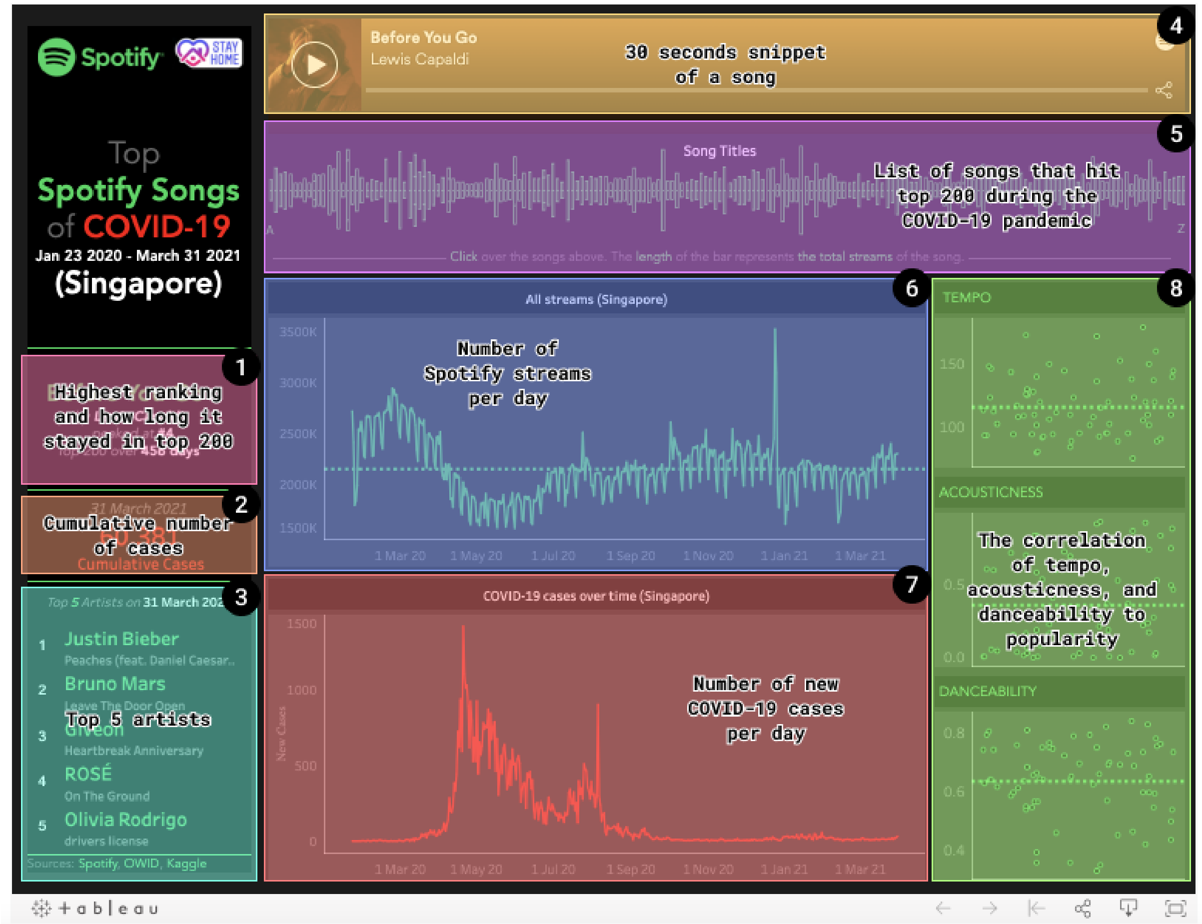
The detailed audio features (tempo, acousticness, danceability) of each song are obtained via a dataset on Kaggle[2] which has already been precalculated using Spotify Audio Analysis. They are consolidated into a single dataset such that each row represents a song and its corresponding audio features.

Lastly, the data regarding the number of Covid-19 cases a day comes from OWID[3] which consolidates data from health institutions and government bodies. As the data is specific to Singapore, the data is obtained from the Ministry of Health in Singapore. It is updated daily and includes data on confirmed cases, deaths, hospitalizations, and testing. The number of new cases is calculated from the results of confirmed cases, deaths, hospitalizations, and testing.

---

[1] Spotify Charts: www.spotifycharts.com
[2] Kaggle Dataset: https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks
[3] OWID data: https://ourworldindata.org/coronavirus/country/singapore

Data Visualisation



The output produced by the interactive data visualization can be divided into 8 segments (numbered from 1-8). The first segment displays the highest ranking of a song and the duration it has stayed in the top 200 songs on Spotify. The second segment presents the cumulative number of cases in a day. The third segment lists the top 5 artists in a day. The fourth segment provides a 30 second snippet of a song. The fifth segment is a list of top 200 songs that has hit top 200 during the Covid-19 pandemic in Singapore. The sixth segment shows the trend of Spotify streams across the duration of 23 January 2020 to 21 March 2021. The seventh segment shows the trend of new Covid-19 cases across the same duration. The eighth segment graphs the correlation of tempo, acousticness, and danceability to popularity.

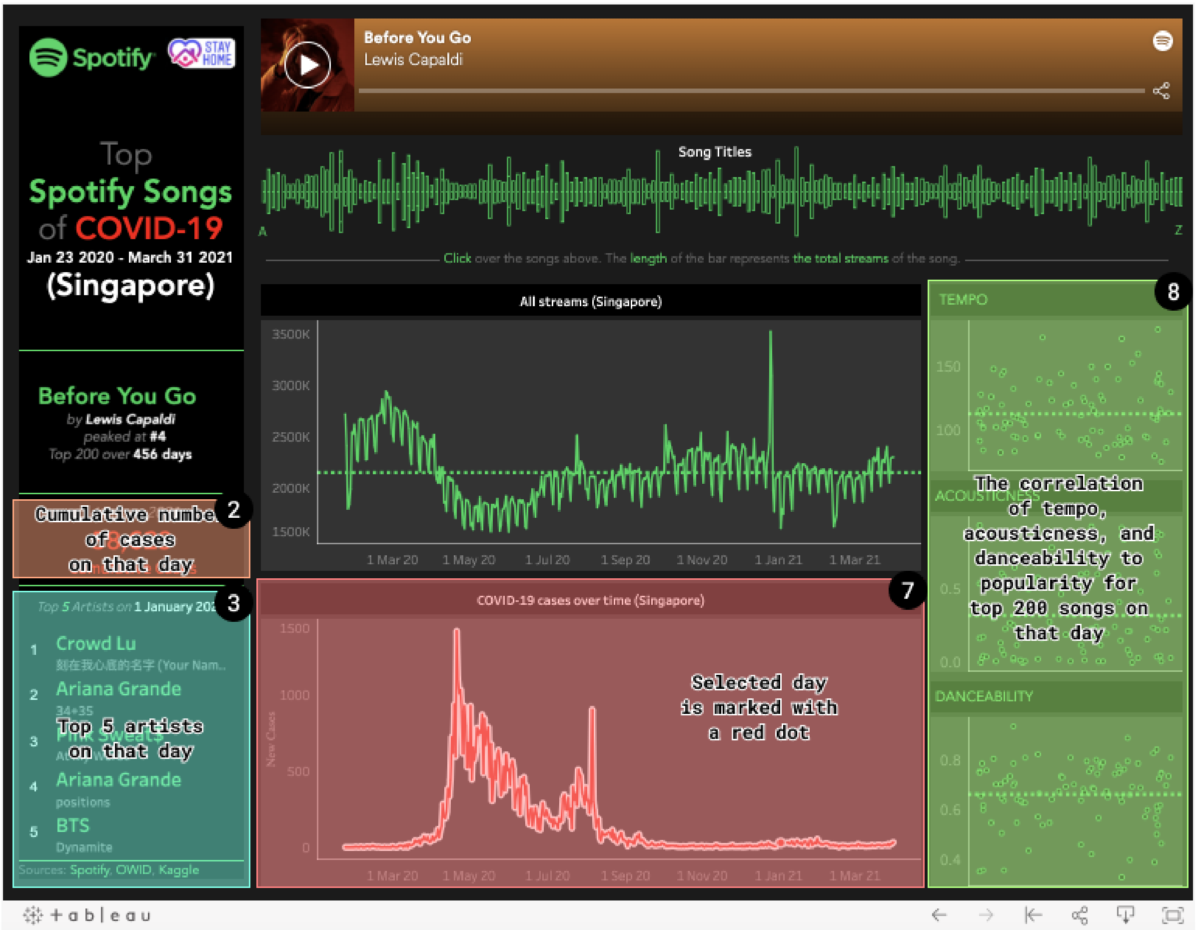
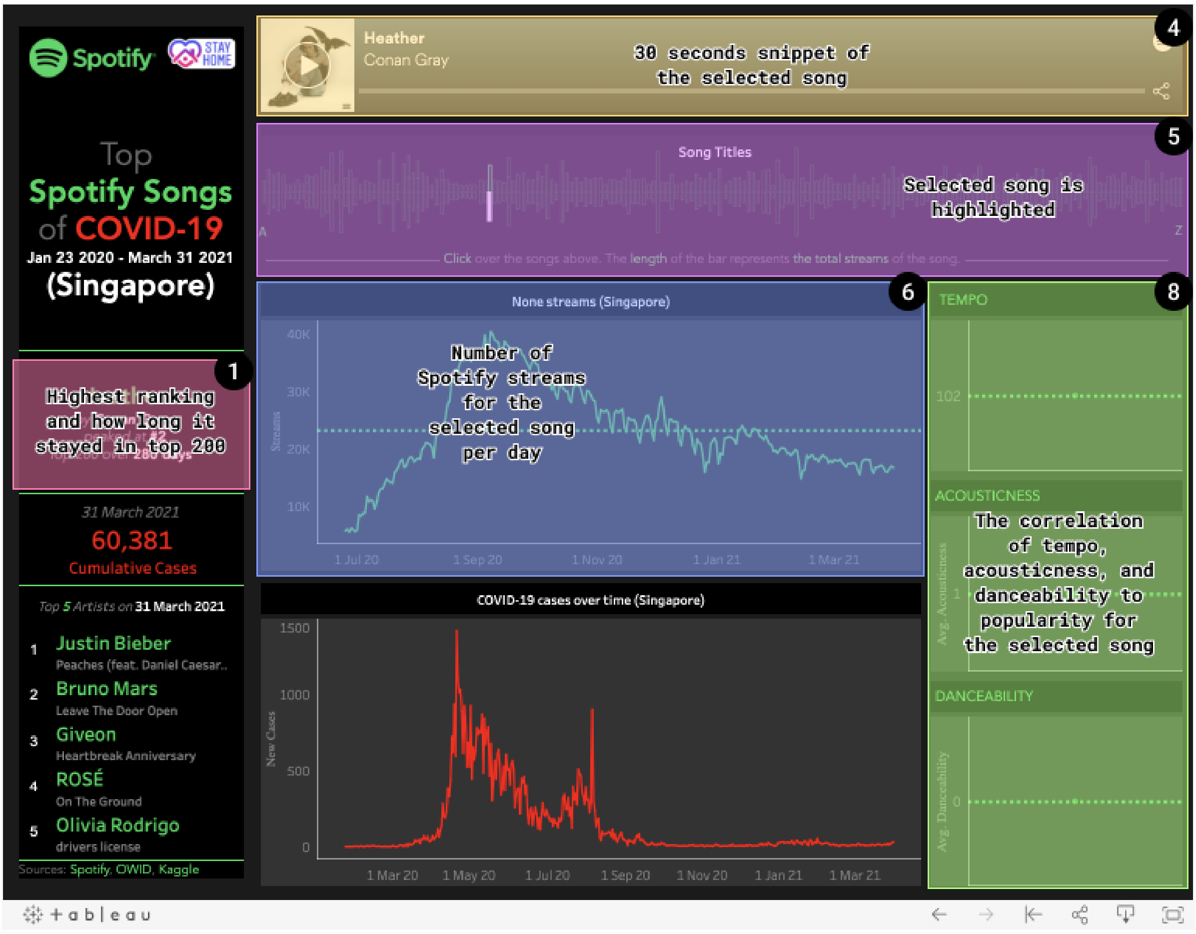


Tableau is used to combine the visualizations into an interactive dashboard. Line graphs were used to show the trends of streams and new Covid-19 cases over time. This is useful to show data variables and trends very clearly and could help to make predictions. Scatter plots were used to show the correlation of tempo, acousticness and danceability to popularity. It is useful as it shows the relationship between two variables. In the case that the relationship is not linear, it is the best method to show a non-linear pattern.

The design of the data visualization is largely inspired by Spotify's user interface. The colour scheme follows its signature green and black in most of the visualisations within the main data visualisation. This would allow readers to roughly deduce what the data visualisation would be about at first sight from the appearance of it. This is largely inspired by modern data visualizations that tend to put the overall design and theme into the context of the data, in order to give readers a prelude to what the data is trying to communicate.



The visualization is insightful as we are able to directly interact with it. For instance by selecting a specific day on the Covid-19 cases chart, we would be able to view the cumulative cases up to that day, the metadata of the top 5 artists of the day, as well as the correlation of tempo, acousticness and danceability to popularity on that day. From here, we can analyse the type of songs that are trending based on a certain number of cases.

Similarly, by selecting a song in the “soundwave", we are able to see the streams for that song over time and observe the relationship with Covid-19 cases. In addition, we can also refer to the song’s tempo, acousticness and danceability and find patterns in songs with similar values.

The data sources are combined through a left join of the Kaggle dataset onto the Spotify dataset on it’s unique id field. A left join means that the integrity of the Spotify charts data is kept intact, however this would result in null values for attributes where the Spotify data did not match with any song in the Kaggle dataset. This would result in some missing song attributes from the data.

Analysis

From observing the visualisation, we can draw that while the number of streams per song over time generally and naturally decreases, there seems to be a decrease in streams especially corresponding with the time of the first spike in the number of Covid-19 cases. This would mean that our initial hypothesis that people would be listening to more music during this period of time (especially during the Circuit Breaker period) is actually proven wrong.

A possible explanation for the data observed is the great reduction of travelling to and from workplaces. In 2018, an average of 7.5 million trips were taken daily in Singapore, with many commuters plugged into their music devices. Thus, one possible inference could be that the lack of commuting to and from home results in a downturn of the amount of music streamed. Furthermore, from our literature we gleaned that there was an uptick of nostalgic songs across Europe. As our study only examines the top 200 charts, we were not able to account for overall streams.

We do however, acknowledge that correlation is not causation, and that these results are not conclusive, especially since the results vary over different types of songs, depending on genre, feel and even language. Furthermore, there could be many more external or temporal factors which may have an effect on the number of music streams in that time period.

Another interesting factor was that there was no significant change between the tempo, acousticness, or danceability from the start of Covid-19 (23 Jan 2020), to the peak of cases (20 April 2020 & 5 August 2020), through to March 31 2021. This means that although music may affect mood, it is not clear how the mood of an individual affects the music they listen to. Interestingly enough, much research has been done on mood-based music, enabling individuals to choose songs based on what mood they are feeling, however, not much research has been done to determine how people are feeling based on the music they are listening to (Meyers 2007). This tells us that more research can be done in this field, with new machine learning and predictive analytics models being used. Furthermore, since our data is taken from aggregate Spotify datasets, it may not be that there are no ways to predict the mood of an individual based on music type, but that we are not able to determine the significant factors due to the granularity of the data.

Reflections

The visualisation itself has certain limitations. For starters, our visualisation only explores the analysis of Covid-19 cases in Singapore. Some may argue that perhaps this analysis can be expanded and even compared to other bigger countries such as the United States to reaffirm the trend analysis that we can conclude from Singapore's context. Nonetheless, we felt that Singapore was an exemplary case study to use in our analysis given its status as a highly globalised country with many users streaming music on Spotify, which is what we aimed to explore in this project.

Furthermore, since the data in Spotify prioritizes the top 200 songs of each day, even if a song were to consistently rank 201st over a period of time, it would not be considered as part of the data. This could mean that we could potentially miss out on significant data in the process.

Another point to note is that there are also other music streaming apps such as Apple Music and SoundCloud, which offer their respective APIs for data analysis. Perhaps, another area of improvement would be to perform additional analysis using datasets from these other apps to ensure a consistent trend in our analysis to reaffirm our hypothesis. In addition, given that we are still living in the Covid-19 era, it would be better if the visualisation can be updated on a daily basis. The restriction of the scope of analysis from Jan 2020 to Mar 2021 is merely used as a means to come up with the visualisation, which can be improved with live updates if possible to reflect the ever-changing trends in the current Covid-19 situation.

In spite of the above-mentioned limitations, we hope that our interactive dashboard that we built using Tableau provides users with a visual appeal to perform their own exploratory data analysis to investigate the possible correlation between the number of daily Covid-19 cases and the genres of songs streamed by Spotify users.

References:

Alluri, Vinoo, Petri Toiviainen, Iiro P. Jääskeläinen, Enrico Glerean, Mikko Sams, and Elvira Brattico. “Large-Scale Brain Networks Emerge from Dynamic Processing of Musical Timbre, Key and Rhythm.” *NeuroImage* 59, no. 4 (February 15, 2012): 3677–89. https://doi.org/10.1016/j.neuroimage.2011.11.019.

Dolegui, Arielle S. 2013. The Impact of Listening to Music on Cognitive Performance. *Inquiries Journal/Student Pulse* 5, no. 09 (2013) http://www.inquiriesjournal.com/a?id=1657

Yeung, Timothy Yu-Cheong, Did the COVID-19 Pandemic Trigger Nostalgia? Evidence of Music Consumption on Spotify (August 21, 2020). *Faculty of Law, KU Leuven*. http://dx.doi.org/10.2139/ssrn.3678606

Meyers, Owen Craigie, A mood-based music classification and exploration system (2007). Massachusetts Institute of Technology. Dept. of Architecture. Program In Media Arts and Sciences https://dspace.mit.edu/handle/1721.1/39337