What influences our choices in music?

Sankalp Khanna (20371974)

Aditya Lamba (20381682)

Wai Lok Siu (20191881)

Chun Hei Mok (20450134)

COMP4462: DATA VISUALISATION

**TABLE OF CONTENTS**

[Abstract 3](#_Toc40976593)

[1. Introduction 3](#_Toc40976594)

[2. Song Popularity over Time 3](#_Toc40976595)

[2.1 Dataset 3](#_Toc40976596)

[2.2 Technology 3](#_Toc40976597)

[2.3 Visualizations and Findings 3](#_Toc40976598)

[3. Similarity of Features amongst Popular Songs 5](#_Toc40976599)

[3.1. Dataset 5](#_Toc40976600)

[3.2. Technology 6](#_Toc40976601)

[3.3. Visualisation and Findings 6](#_Toc40976602)

[4. Feature Analysis of Top Songs 8](#_Toc40976603)

[4.1. Dataset 8](#_Toc40976604)

[4.2. Technology 8](#_Toc40976605)

[4.3. Visualisation and Findings 8](#_Toc40976606)

[5. Conclusion 9](#_Toc40976607)

[Reference 10](#_Toc40976608)

# Abstract

Music is an irreplaceable part of our society. It is not only a cultural activity; music is a vital part of our way of life. It has a variety of genres – from Modern Electronic to Classical music with more than a billion songs. However, people easily parse through these to choose specific songs based on their tastes without a second thought. These behaviours can be uncanny reflections of complex and subtle elements such as the liveness of that song. This project aims at investigating these factors that could possibly affect people’s taste in music.

# 1. Introduction

Music is often viewed as lifestyles by many. According to the statistics by Spotify, there are 256 million monthly listeners on the platform. An average consumer listens up to 25 hours each month. While a lot of research has been conducted on how music is related to our personalities and our moods, there has been no major headway when it comes to the reasons behind our choices in music.

Thus, the objective of this project is to find a quantifiable solution. In order to achieve our goal, data are transformed into different sets of visualizations for further analysis. The scope of this project would focus on the most popular songs and how their features correlate with their popularity. This project is divided into three different tasks:

1. Song Popularity
2. Feature Similarity
3. Features analysis

# 2. Song Popularity over Time

Our project aims at finding the factors determining the people's taste in music. Since Spotify has an enormous database with more than 50 million songs, the scope is narrowed down to the most popular ones. Therefore, the daily top 15 songs in 2019 are used in this task. This task focuses on the popularity of the songs and their trends throughout the whole year.

## 2.1 Dataset

In order to analyse the popular songs, a dataset containing the Daily Top 15 songs in 2019 was created. It also contains information such as the names of soundtracks, the artists, and the numbers of streams, which indicate the popularity of songs, and the themes of the songs.

## 2.2 Technology

Tableau was used to create visualizations of this task. A Streamriver graph was made with interactive features.

## 2.3 Visualizations and Findings

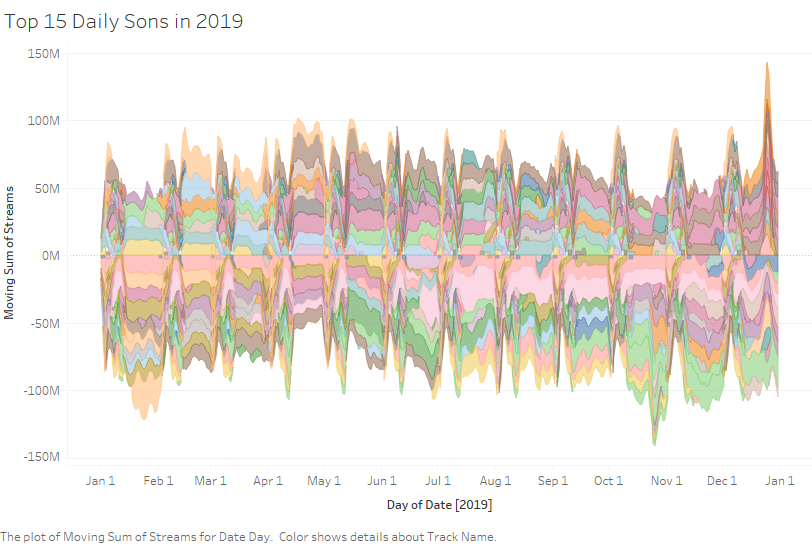


Figure 1: Daily Top 15 Songs in 2019 in Streamriver visualization

Figure 1 visualizes how the counts of the streams of songs change over time in 2019. Each colour represents a song in this Sreamriver graph. The first finding is related to the lifespans of popular songs. It shows most popular songs can only last for around 4 to 6 months. The global market would be tired of these songs after several months in most cases. The theme, artist and other factors would not help them to keep thier momentum after that period of time, except one exceptional case that would be explained in the second finding. This might help the music industry to make some financial decisions.



Figure 2: The count of stream of ‘All I Want For Christmas Is You’ by Mariah Carey, released in 1994

Besides the lifespans of songs, our second finding focuses on another vital component of music – the theme. Most people might assume there are some trends in the theme. However, this expected result does not hold in the dataset. In fact, songs with all kinds of themes, including love, friendship, life, have similar chances to become popular. The global market did not show a specific tendency on the themes except one specific case: festival.

Figure 2 shows the sudden surge in December 2019. All songs with keywords ‘Christmas’ had increased dramatically in popularity during that month. ‘All I Want For Christmas Is You’ by Mariah Carey, which was released in 1994, had more than 9 million streams around the world on 25 December 2019.

Therefore, it is obvious that festivals would have a huge impact on the taste of music. People would be more likely to select related songs during festival time, even if the song was released 25 years ago.

# 3. Similarity of Features amongst Popular Songs

Song features are one of the most important aspects of a song that affects people’s perception of the song. In this section, the similarity of popular songs would be inspected. We shall compare the feature values of the 250 most popular songs for the year of 2019 and see how they correlate to each other.

## 3.1. Dataset

For the purpose of this visualisation, Spotify library in Python was applied to extract features for our songs. A dataset containing the names, IDs and features of the most popular songs of 2019 was created.

## 3.2. Technology

This visualisation based on the Hierarchical Edge Bundling model with D3.js. The code was forked from a pre-existing L2 distance-calculating algorithm on Observable.

## 3.3. Visualisation and Findings

Building the visualisation, L2 distance finding theorem was applied to develop vectors between songs based on their differences in features. This algorithm was then applied in the whole dataset, creating a matrix of Objects that contained the index of the compared songs and the calculated differences.

Finally, two datasets was made using the matrices developed. A threshold was set at the half of the total average of all calculated differences. For the first one, only songs with differences less than the threshold were selected, which includes songs with high similarities. The second dataset was conducted in the opposite way. Only songs with differences greater than 1.5 of the total average were kept. This kept songs that were dissimilar from the norms.

The visualisations for both datasets were developed using Hierarchical Edge Bundling technique. This method was used because of its ability to handle large number of objects as well as its capability to showcase the locations of nodes in each dataset.

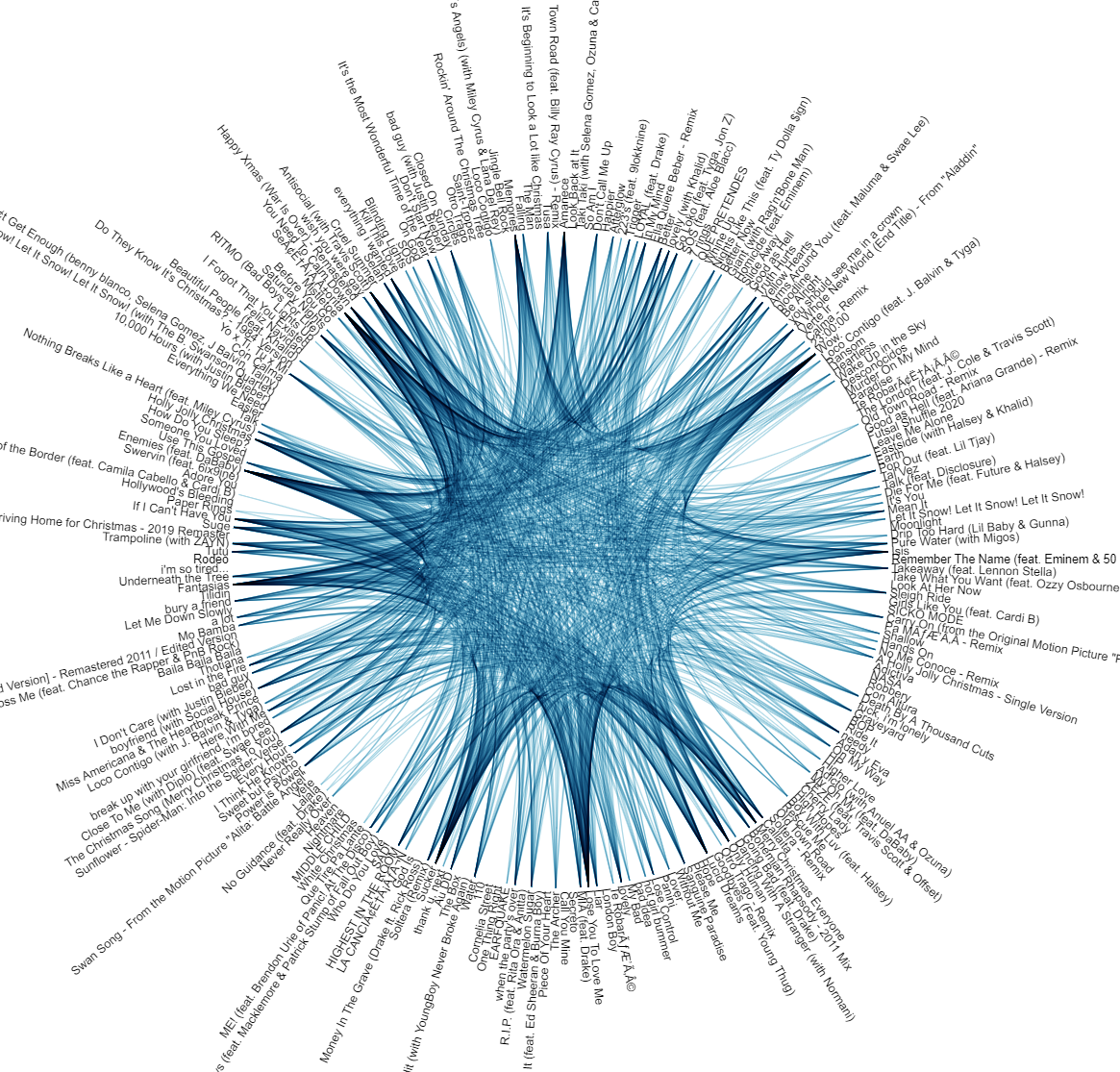


Figure 3: Songs with the least similar features

Figure 3 visualises the second dataset developed. The nodes are built with songs that are different from one another. Some may believe songs are getting similar. However, there were some popular songs sharing low similarities in features. One of the most popular songs of 2019, Senorita stood out from the norms because of its rather different features.

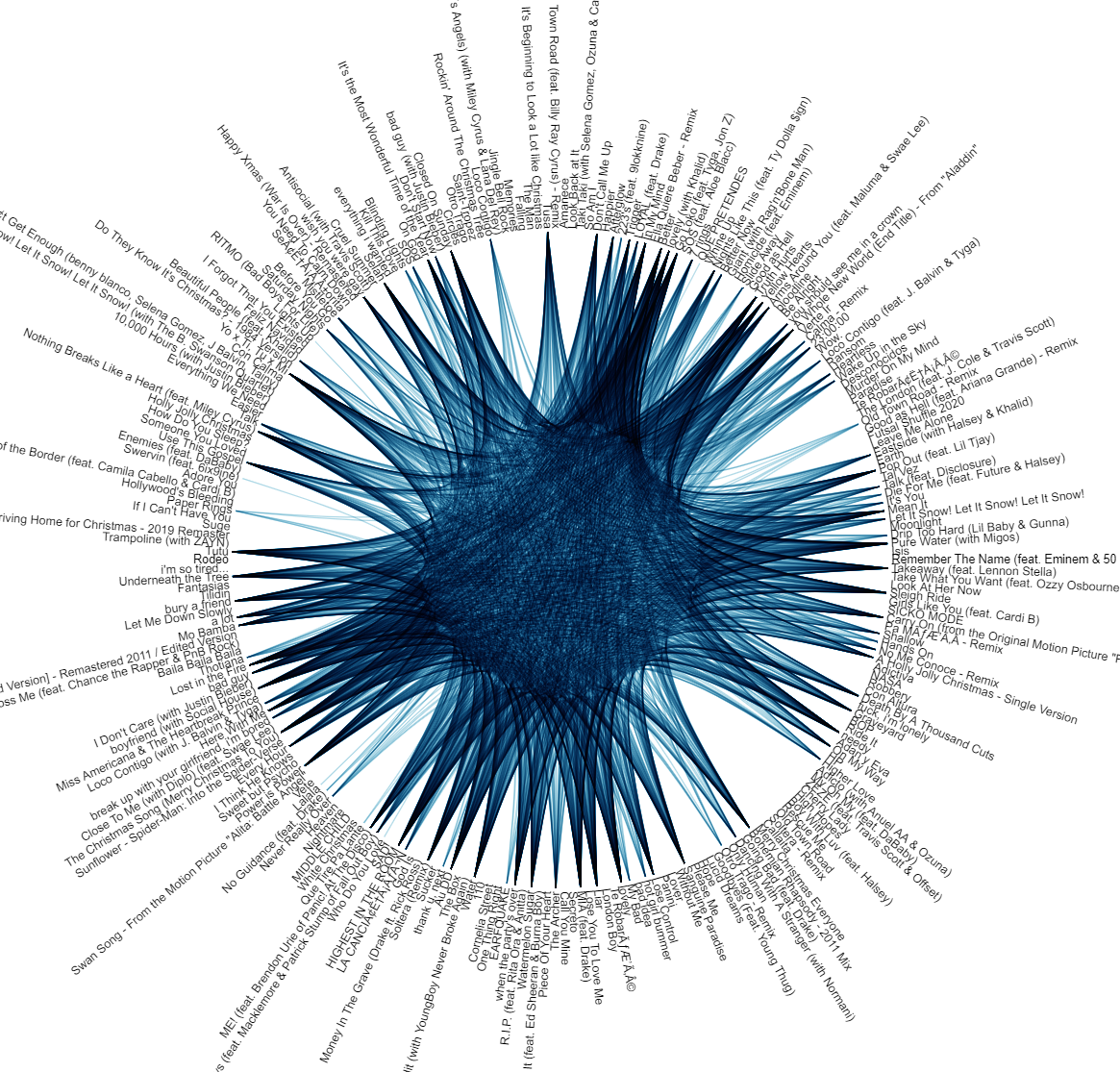


Figure 4: Songs with similar features

A more important topic would be how similar are the top music. As shown in Figure 4, a lot of songs have similar features. Despite being lauded for its fresh beats and tempo, ‘Bad Guy’ by Billie Eilish shares lots of similarities with many other songs in the dataset.

# 4. Feature Analysis of Top Songs

This task analyses the features of top songs globally to plot a trend of the features that are common among them. A Parallel Coordinate graph has been implemented to compare and observe any relationship amongst the features of global top songs in 2019.

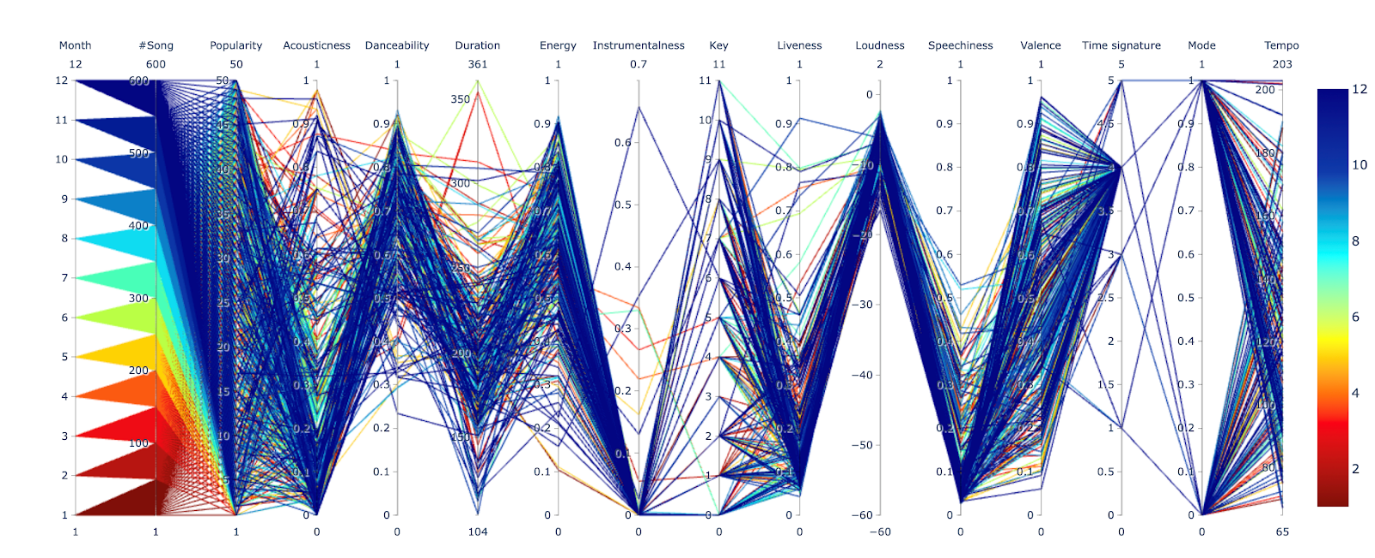
## 4.1. Dataset

Since the data is cumulative, popular 200 songs were extracted for each month using Spotify Charts. Spotify library in Python was used to extract features corresponding to each unique ID of a song. All duplicates were removed. After getting rid of duplicate songs, for a focussed study, features of top 50 songs for each month for the year 2019 has been chosen and used.

## 4.2. Technology

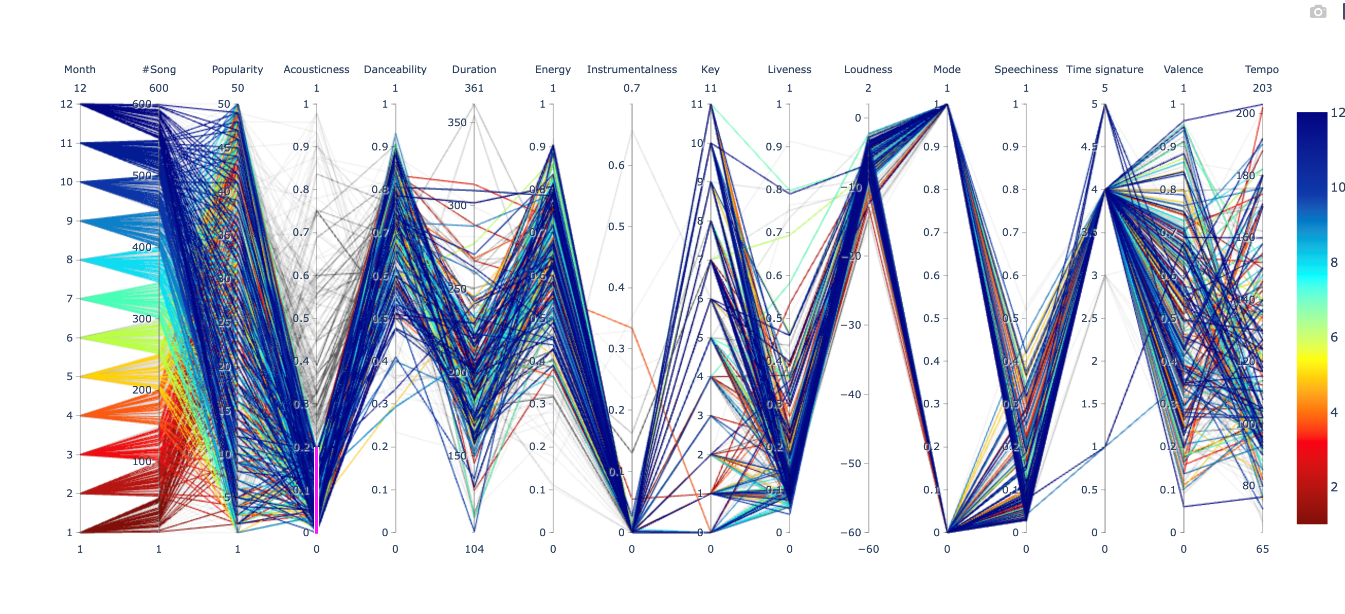
Python was used to develop this visualisation. We specifically made use of Pandas and Plotly libraries to achieve our goals.

## 4.3. Visualisation and Findings

Figure 5: Features of global top 50 songs arranged monthly

As each feature represents a different dimension of a song, we have made use of Parallel Dimensions for this visualisation. Each y-axis is a different dimension and has its own unit of measurement. To make it slightly easier to differentiate, each month has been given its own colour and numbered from 1 to 12. 1 represents January while 12 represents December. The overview of the visualisation can be seen in Figure 5.

Since a lot of features are dealt in the graph, a multi-filtering option has been implemented. The relationship between adjacent features are easier to perceive, thus, the graph also allows users to do side-by-side feature analysis by re-ordering the respective axes. Furthermore, to remedy the problem of cluttering, a technique called “Brushing” has been implemented allowing users to highlight a line or a collection line while fading out the others. This allows to isolate the sections of interesting plots and find patterns while filtering noise.

Figure 6: Filtering and findings

From Figure 6, we can observe that people usually prefer songs that have high danceability, high energy, low duration, and fewer lyrics. Another interesting find was that people generally songs that had electric and amplified musical instruments as compared to stringed instruments. This can be proven in the figure where the acoustics of most popular songs accumulate towards the lower side.

# 5. Conclusion

Our aim of this project is to find how different features of song would alter our choices in music. This is achieved by studying popular songs and some conclusions have been made based on the findings.

The popularity and the momentum of popular songs rarely exceed 4 to 6 months, irrespective to other factors of the songs. Festive songs are exceptional to this finding. They always gain popularity even after decades.

In terms of features, most popular songs are extremely similar. Only a small proportion of them do not share high similarities. This implies that some specific features might account for the commercial success of these songs.

The features were analysed in the last part. By mapping the trend these features followed, songs with high danceability, energetic, low duration, performing by electrical instruments, and having not too many lyrics are the key success factors.

This is understandable as these a high danceability means these songs are played often in clubs, thus exposing it to more people.

Through the course of this study, we have aimed to identify these key features and their correlation with popularity. We believe that this study and visualisation can be used by individuals to better understand their preferences and can also be used by record labels to decide what kind of songs they should produce to maximise their sales. Not only this, but this study will also help pave the path for future music related studies that will help us come to an understanding about how we respond to external stimulus.

# Reference

1. **Observable notebook for Edge Bundling:** <https://observablehq.com/@sankomil/hierarchical-edge-bundling/2>
2. **Github repository:** <https://github.com/sankomil/COMP4462-Project>
3. **Parallel Dimensions Collab notebook:** <https://colab.research.google.com/drive/12PmnofvZpJ6wC4sJcwuaw3Uec_i7_8bX>