

A/B Test: Briefing: Predicting Music Success On YouTube

Student Name: Helge Moes

Student Number: 11348801

Date: 29-02-2024

This A/B Test is an initial attempt and is therefore subject to change. Due to the freedom given in this assignment, an exploratory analysis shall be attempted in order to examine which variables are best fitted to examine user engagement of music on YouTube. Additionally, an example dataset is included to add depth and clarity to the research. Therefore, the case and the initial data presented in this briefing can be considered as preliminary.

Case Description

Throughout history, music has always been a prominent medium that has been consumed by many, whether live or on the radio. However, through the introduction of platforms, such as Applemusic, Spotify and YouTube, it has shifted to another medium: online streaming services (Simon, 2019). YouTube in particular is interesting, since it started with the intention to enable regular users to publish videos (Liikkanen & Salovaara, 2015).

While their initial emphasis was on videos, the online streaming platform shifted to presenting music videos, which made it accessible for everyone anywhere to listen or watch music. Eventually the platform became the most recognized digital music brand in 2014 (Liikkanen & Salovaara, 2015). This popularity was accompanied with a gradual development of the platform into a professional media outlet that has been used by brands in order to interact with their audience. As a result, the user has access to a vast library of millions of music at any time and as many times as they want.

Besides this unlimited access, the rise of online streaming platforms made it also possible to analyze the data of music and whether users engage more with certain content. By analyzing music data on YouTube, it can provide insights into listener preferences and behavior, which not only can improve YouTube's value proposition as a company that offers targeted content based on the user's preferences, but it can also help the creators and users to improve their music production on the platform.

The relevance of this research is to provide insights to how YouTube can improve their music recommendation based on engagement metrics of users and understand how this relates to the engagement, which in turn shall improve the user experience and increase user engagement. Due to the challenge of defining engagement, this study shall regard engagement mainly through metrics, such as viewCount, likeCount and commentCount. Moreover, the results of this research can also serve as a means to understand how successful music content (lyrics) can be created based on engagement for other platforms that also provide recommendations of music to users, such as Tidal, Applemusic and Spotify, since they also rely on customer feedback for a strong presence of a certain content (Smith et al., 2012).

Based on these grounds, this research shall attempt to uncover the impact of language in music and how it affects user behavior and engagement. More specifically, it shall explore whether music with profound or simple language holds an impact on user engagement. Therefore, the following question is formulated:

Research Question: To what extent does language in music have an impact and can we use metadata to predict the success of music on YouTube?

In order to answer this question, the following sub questions shall be explored:

- What features are associated with increasing engagement on YouTube?
- What models are best suited to predict popularity on YouTube based on language and user feedback?
- How can the insights be used to improve music marketing and content creation on the YouTube platform?

Set up

In order to answer these questions, the data collection for this study emphasizes music quality, performance metrics and engagement indicators that highlight language specifically. Therefore, the A/B testing shall compare the different videos and their data, where engagement metrics such as viewCount are considered as dependent variables. Moreover, the independent variable shall be the transcript of

the music video, because it represents a characteristic or attribute that may be subject to manipulation for research purposes and filtering out noise. In this case, the language in the transcript refers to the complexity or linguistic features present in the lyrics.

Furthermore, this briefing also includes data that was collected through YouTubeDataTools from Digital Methods Initiative of the Universiteit van Amsterdam in order to define the relevant variables for this exploratory analysis (*YouTube Data Tools*, n.d.). The query “music” was used in the example dataset and English music content was filtered between January 2023 and January 2024, which resulted in a dataset with 445 songs and 29 variables. Besides this dataset, a separate dataset of the transcript is needed to assess whether the content of the music resonates with the users. Therefore, the audio transcripts of the videos still need to be retrieved from the youtube-transcript-api (Depoix, n.d.).

Since this research is of an exploratory nature, the different variables presented in this table shall be explored in order to determine which variables seem to be most relevant to measure user engagement and whether language can also be considered to be a factor. Consequently, the variables presented in *Appendix A* are still subject to change and shall be further examined for this particular analysis.

Possible recommendations

Through this A/B test, I expect to be able to evaluate the accuracy of machine learning models in predicting popularity using metadata. Moreover, I aim to identify the most effective model by comparing performance metrics such as like count, comment count, video categories, and duration in seconds. Consequently, I intend to utilize insights gained from the analysis to enhance music marketing and content creation on YouTube. This involves tailoring content to align with engagement-driving features and experimenting with various content formats, particularly focusing on lyrics.

Additionally, I plan to explore opportunities to enhance content recommendation algorithms on YouTube. This includes considering the integration of features such as video category, duration, and language to personalize user experiences and increase engagement with music content. Lastly, I aim to

investigate the impact of language in music on user engagement metrics. My goal is to determine whether views, likes, tags, and comments can be used to predict this accurately. This research aims to refine content creation strategies and recommend language styles that effectively resonate with the target audience.

Bibliography

- Depoix, J. (n.d.). *youtube-transcript-api: This is an python API which allows you to get the transcripts/subtitles for a given YouTube video. It also works for automatically generated subtitles, supports translating subtitles and it does not require a headless browser, like other selenium based solutions do!* (0.6.2) [Python; OS Independent]. Retrieved February 26, 2024, from <https://github.com/jdepoix/youtube-transcript-api>
- Liikkanen, L. A., & Salovaara, A. (2015). Music on YouTube: User engagement with traditional, user-appropriated and derivative videos. *Computers in Human Behavior*, 50, 108–124. <https://doi.org/10.1016/j.chb.2015.01.067>
- Simon, J. P. (2019). New players in the music industry: Lifeboats or killer whales? the role of streaming platforms. *Digital Policy, Regulation and Governance*, 21(6), 525–549. <https://doi.org/10.1108/DPRG-06-2019-0041>
- Smith, A. N., Fischer, E., & Yongjian, C. (2012). How Does Brand-related User-generated Content Differ across YouTube, Facebook, and Twitter? *Journal of Interactive Marketing*, 26(2), 102–113. <https://doi.org/10.1016/j.intmar.2012.01.002>
- YouTube Data Tools*. (n.d.). Retrieved February 26, 2024, from <https://ytdt.digitalmethods.net/>

Appendix A - Preliminary Table of Music Evaluation

Variable Name	Definition
channelId	Identifier for the publisher of the video

channelTitle	The title of the published video
videoId	Identifier for the YouTube video
videoTitle	The title of the YouTube video
publishedAt	Date and time the video was published on YouTube
videoDescription	The description of the
tags	Tags assigned by the publisher to the video
videoCategoryLabel	Classification tag of the video based on the content
durationSec	The duration of the YouTube video
viewCount	The number of views the video has
likeCount	The number of users who have liked the video
commentCount	The number of comments the video has
transcript	The raw audio transcript from the video