This song is WHACK! Skip!

Investigating the Causal Influence of Song Transitions on User Behavior

Amie Thomas & Aabha Latkar

Introduction

In the realm of music streaming platforms, understanding user behavior and preferences is paramount for enhancing user experience. One intriguing question that arises is whether the transition between songs with different vibes influences user engagement, particularly in the context of skipping tracks. This project aims to delve into this question by conducting a causal analysis on the Music Streaming Sessions dataset released in 2018. The dataset provides insights into user listening records, including session dates, song sequences, and user attributes such as subscription status. Additionally, it encompasses song features like danceability, mood, and energy levels. By leveraging this comprehensive dataset, we seek to explore whether deviations in song vibes prompt users to skip tracks. There's a lot of analysis and work done on music recommendation systems and predicting if a user will skip songs or not. We believe it would be really cool to conduct a causal analysis on the same to definitively ascertain the factors that cause users to skip tracks.This investigation not only addresses an intriguing aspect of user behavior but also holds relevance for improving recommendation algorithms and enhancing user satisfaction on music streaming platforms.

Method and Analysis

The data used for this project originates from the Musical Streaming Sessions datasets released by Spotify in 2018. The first dataset contains details concerning user listening behavior, encompassing features such as subscription type, user skip behavior, and playlist switching within sessions. Complementing this, the second dataset comprises song features, encompassing 30 features related to song composition, such as danceability, energy, and beat strength. Employing Principal Component Analysis (PCA), we reduced the dimensionality of these features while preserving their essential information. Subsequently, we clustered these components, resulting in 10 distinct clusters representing unique song "vibes". The causal parameter of interest in this project is the causal effect of song transition (changing from one cluster to another) on the likelihood of a user skipping a song within a listening session. In other words,

we investigated whether the change in song cluster causally influences the decision to skip a song.

Discussion

In this project we assume several assumptions. Our first assumption is the Stable Unit Treatment Value Assumption (SUTVA). According to the assumption no interference between units means the treatment status of one unit does not affect the outcomes of other units. For our project this means that the transition of one song does not directly influence the likelihood of skipping other songs in the session. We assume Positivity, which states that there is a nonzero probability of receiving each level of treatment for all units in the population, given their observed covariates. In the context of this project, there are no unreachable combinations of song transitions and user characteristics.

Consistency is another of our assumptions. Consistency assumes that the potential outcome under a given treatment level is well-defined and consistent across units. We assume that the effect of transitioning from one song cluster to another on song skipping is consistently defined for all users. Exchangeability implies that, conditional on the covariates, the treatment assignment is independent of the potential outcomes. For us, it means that users who experience different song transitions would be comparable if they had the same covariate values. Lastly, we assume no selection bias. No selection bias states that there are no systematic differences between the treated and control groups, except for the treatment itself. This means that users who experience different song transitions are similar in all aspects except for the transition itself.

The risk ratio is the most suitable method for analyzing this scenario due to its compatibility with binary outcome variables, such as the decision to skip a song within a listening session. By calculating the risk ratio, we can directly quantify the relationship between transitioning songs and the likelihood of song skipping in a clear and interpretable manner. Unlike other measures the risk ratio provides a straightforward interpretation. It represents the relative change in the risk of the outcome (song skipping) between the exposed (those experiencing a song transition) and unexposed (those not experiencing a song transition) groups. This direct comparison of risks offers a more intuitive understanding of the causal effect of song transitions on user behavior within the context of a listening session. Moreover, the risk ratio is particularly advantageous when the outcome is relatively common, as it ensures a more accurate estimation of the causal effect without the need for additional transformations or adjustments. Thus, employing the risk ratio as the primary analytical method allows for a comprehensive assessment of the impact of song transitions on song skipping behavior, facilitating informed decision-making and effective intervention strategies.

In our analysis, we utilized the risk ratio as our estimator to quantify the relationship between song transitions and the likelihood of song skipping within listening sessions. This method allows us to directly estimate the risk ratio while accounting for potential confounders, such as

subscription type and context type, thereby providing a robust assessment of the causal effect of song transitions on song skipping behavior. To ascertain the precision of our estimates, we obtained confidence intervals through bootstrapping techniques. By resampling the data with replacement and recalculating the risk ratio for each iteration, we generated a distribution of estimates from which we derived the confidence intervals. This approach accounts for the variability in our sample and provides a reliable measure of the uncertainty surrounding our point estimates, ensuring the robustness and validity of our findings.

Results

In our analysis, we initially obtained a risk ratio of 1.065 without conditioning on covariates, indicating that individuals exposed to the treatment—transitioning to a song with a different mood than the previous one—are only 1.065 times more likely to skip the song compared to those not exposed to the treatment. The corresponding confidence interval for this non-conditional risk ratio was [1.052282, 1.079062], and the results demonstrated robustness. Upon further investigation while conditioning on covariates—subscription type and context switch—we observed a slight increase in the risk ratio to 1.086, with a confidence interval of [1.0827, 1.0889]. This finding suggests that even after considering these covariates, there remains little to no discernible causal effect of mood change on user skip behavior.

Conclusion

This project offered insights into the dynamics between music characteristics, user behavior, and platform functionality within streaming services. By exploring the causal relationship between song transitions and user skip behavior, we gained a deeper understanding of how subtle changes in the music listening experience can influence user engagement and satisfaction. Moreover, by incorporating covariates such as subscription type and context switch, we uncovered nuanced patterns that shed light on the diverse preferences and behaviors of different user segments. Additionally, the use of advanced analytical techniques, such as Principal Component Analysis (PCA) for feature reduction and robust regression models for causal inference, equipped us with methodological tools to navigate complex datasets and derive meaningful conclusions. Ultimately, findings from this project contributed to refining recommendation algorithms, enhancing user experience design, and informing strategic decision-making in the digital music streaming industry.

References

https://engineering.atspotify.com/2014/02/how-to-shuffle-songs/

https://austinpoor.com/blog/predict-spotify-skips/

https://repositori.upf.edu/bitstream/handle/10230/47327/ferraro_wsdm_skip.pdf?sequence=1&isAllowed=y

https://www.technologyreview.com/2023/04/04/1070885/complex-math-counterfactuals-spotify-predict-finance-healthcare/

https://www.nature.com/articles/s42256-023-00611-x.epdf?sharing_token=YPp6X7lZZdhHH5yN53Lw6NRgNyFTVk4gLzvrYxChOGvTK0yL3Ot61HD06QKpdNiyjm6vO38H8Nq8wa2Kx5WQLHibAHqtMNQ_dlle_UkgWcazSu4NMPVeQcO17bBPu8FhX1I-N5YLkxzDhbifAAAGelPdpG9SX2Nbuv9eqK7-QUSa6vyTsOCPBlhpIMJG5iuBO4JzbdHmZa2SJ8Cox0%3D&tracking_referrer=www.technologyreview.com