Received: 10 Jan. 2026 | **Revised:** 15 Apr. 2026 | **Accepted:** 19 May. 2026 | e-**ISSN:** 3085-7813

Alumni.in

A RTICLE

COMPETITIVE INTELLIGENCE AND CONSUMER RETENTION IN DIGITAL STREAMING PLATFORMS: A STRATEGIC FRAMEWORK FOR BEHAVIORAL ANALYTICS AND SUSTAINABLE COMPETITIVE ADVANTAGE

INTELIGÊNCIA COMPETITIVA E RETENÇÃO DE CONSUMIDORES EM PLATAFORMAS DIGITAIS DE STREAMING: UM FRAMEWORK ESTRATÉGICO PARA ANÁLISE COMPORTAMENTAL E VANTAGEM COMPETITIVA SUSTENTÁVEL

[1*]**Shujing Tang:** Composition Major of the Graduate School of Music, Senzoku Gakuen College of Music, Kanagawa, Japan. ORCID: https://orcid.org/0009-0001-9257-1286

[2]**Fengda Wu:** Faculty of Psychology, Shinawatra University, Pathum Thani, Thailand. ORCID: https://orcid.org/0009-0005-5362-8218

[3] **Chenhao Zhu:** Università degli Studi di Padova, Padova, Italy. ORCID: https://orcid.org/0009-0007-1712-9846

Corresponding Author:
Shujing Tang
E-mail: Tshujing6666@gmail.com

Editor in chief
Altieres De Oliveira Silva
Alumni.In Editors

How to cite this article:

Tang, S., Wu, F., & Zhu, C. (2026). Competitive Intelligence and Consumer Retention in Digital Streaming Platforms: A Strategic Framework for Behavioral Analytics and Sustainable Competitive Advantage. *Journal of Sustainable Competitive Intelligence*, *16*, e0680. https://doi.org/10.37497/eagleSustainable.v16i.680

ABSTRACT

Purpose: This study investigates how Competitive Intelligence (CI) capabilities influence consumer retention and loyalty in digital music streaming platforms. Specifically, the research examines the effects of personalization quality, service continuity, pricing competitiveness, and behavioral engagement analytics on subscriber loyalty and churn reduction.

Methodology/approach: A quantitative cross-sectional research design was employed using the KKBox Music Streaming Churn Prediction dataset containing 970,960 anonymized subscriber records. Statistical analyses were conducted in Python 3.11 using descriptive statistics, reliability and validity analysis, Pearson correlation, OLS multiple regression, and bootstrapped mediation analysis. User engagement and perceived value were tested as mediating variables, while subscription tier was examined as a moderating factor.

Originality/Relevance: The study proposes the CI-Loyalty Framework (CILF), an original theoretical model that conceptualizes Competitive Intelligence as a multidimensional organizational capability embedded in behavioral analytics and strategic decision-making processes. Unlike prior studies based primarily on survey data, this research operationalizes CI constructs using large-scale administrative behavioral logs from a real-world streaming platform.

Key findings: The results demonstrate that the CI composite index was the strongest predictor of consumer loyalty. Personalization quality and perceived value also showed strong positive effects on retention outcomes. User engagement and perceived value partially mediated the relationship between CI capability and loyalty. Additionally, subscription payment methods significantly influenced churn behavior, with auto-debit users presenting substantially lower churn rates than voucher-payment subscribers.

Theoretical/methodological contributions: The study contributes theoretically by integrating Resource-Based View (RBV), Information Processing Theory, Customer Engagement Theory, and the Competitive Intelligence cycle into a unified framework for digital streaming environments. Methodologically, the research demonstrates the feasibility of operationalizing CI constructs through behavioral proxy indicators extracted from large-scale platform data, expanding the empirical application of Competitive Intelligence in digital platform ecosystems.

Keywords: Competitive Intelligence. Consumer Loyalty. Churn Prediction. Music Streaming. Personalization. KKBox. User Engagement. Perceived Value

DOI: https://doi.org/10.37497/eagleSustainable.v16i.680

Open access

RESUMO

Objetivo: Este estudo investiga como as capacidades de Inteligência Competitiva (Competitive Intelligence – CI) influenciam a retenção e a lealdade dos consumidores em plataformas digitais de streaming musical. Especificamente, a pesquisa analisa os efeitos da qualidade da personalização, continuidade do serviço, competitividade de preços e analytics comportamental sobre a fidelização e redução do churn dos assinantes.

Metodologia/abordagem: Foi adotado um desenho quantitativo transversal utilizando o dataset KKBox Music Streaming Churn Prediction, composto por 970.960 registros anonimizados de assinantes. As análises estatísticas foram realizadas no Python 3.11, incluindo estatística descritiva, análise de confiabilidade e validade, correlação de Pearson, regressão múltipla OLS e análise de mediação com bootstrap. O engajamento do usuário e o valor percebido foram testados como variáveis mediadoras, enquanto o nível de assinatura foi analisado como variável moderadora.

Originalidade/Relevância: O estudo propõe o CI-Loyalty Framework (CILF), um modelo teórico original que conceitualiza a Inteligência Competitiva como uma capability organizacional multidimensional integrada à analytics comportamental e aos processos estratégicos de tomada de decisão. Diferentemente de pesquisas anteriores baseadas predominantemente em surveys, este estudo operacionaliza constructos de CI utilizando dados comportamentais administrativos em larga escala provenientes de uma plataforma real de streaming.

Principais resultados: Os resultados demonstram que o índice composto de CI foi o principal preditor da lealdade dos consumidores. A qualidade da personalização e o valor percebido também apresentaram fortes efeitos positivos sobre os resultados de retenção. O engajamento do usuário e o valor percebido mediaram parcialmente a relação entre capacidade de CI e lealdade. Além disso, os métodos de pagamento apresentaram impacto significativo sobre o churn, sendo os usuários com débito automático os que demonstraram menor taxa de cancelamento.

Contribuições teóricas/metodológicas: O estudo contribui teoricamente ao integrar a Resource-Based View (RBV), a Information Processing Theory, a Customer Engagement Theory e o ciclo de Inteligência Competitiva em um framework unificado para ambientes digitais de streaming. Metodologicamente, a pesquisa demonstra a viabilidade de operacionalizar constructos de CI por meio de indicadores comportamentais extraídos de grandes bases de dados de plataformas digitais, ampliando a aplicação empírica da Inteligência Competitiva em ecossistemas digitais.

Palavras-chave: Inteligência Competitiva. Lealdade do Consumidor. Predição de Churn. Streaming Musical. Personalização. KKBox. Engajamento do Usuário. Valor Percebido.

1. INTRODUCTION

In 2023, the recorded music sector had global revenue estimated at USD 28.6 billion, and the streaming contributed 67.3% to this, which has been increasing steadily since 2015 (IFPI, 2024). It has changed the competitive landscape significantly, as physical and download-era competition focused on exclusivity of distribution and catalog ownership, streaming competition focuses on keeping subscribers. With consumers able to easily sign up for several services at once, without penalty, and be able to access "functionally equivalent" catalogues across competitors, competitive advantage is now linked to the ability to maintain consumer loyalty (Datta et al., 2018; Oliver, 1999). Without keeping subscribers, platforms lose revenue, but they also forfeit the behavioral data that powers them to make the right recommendations and enter a vicious cycle of churn and intelligence loss. Competitive intelligence (CI) is a systematic action focused on the collection of information, its analysis and use of information on competition and customer behaviour and the dynamics of the market (Fleisher & Bensoussan, 2015). CI takes a very micro-behavioural shape in digital streaming platforms. As a result of each session, skip, playlist added or device switched, a data signal is generated that can be translated into predictive insight about the risk of subscriber disengagement, as described in Schedl et al. (2022). The CI is mainly applied at product development/pricing strategy level in a fabricator/retailed environment, but it is applied at a subscriber level to enable real-time personalisation, churn prediction, and retention intervention in a streaming environment (Park & Kim, 2022). However, the value of CI in digital streaming should not be limited to churn reduction or operational retention analytics. Subscriber-level behavioral signals become strategically meaningful only when they are incorporated into an organizational intelligence cycle involving planning, data collection, analysis, dissemination, and strategic decision-making. In this study, churn and loyalty indicators are therefore treated not only as outcomes to be optimized but also as intelligence inputs that allow platform managers to identify competitive vulnerabilities, allocate resources, and sustain long-term market advantage.

Subscription services have a richer type of consumer loyalty than just behavioral loyalty. Oliver (1999) separated the cognitive, affective, conative and action loyalty, and stated that attitudinal commitment, not inertia, is essential for achieving sustainable behavioral loyalty. This is very important in streaming contexts, because the subscriber who pays but hasn't engaged with the service is still a potential competitor for the service, whereas the true loyal subscriber exhibits a high level of engagement, willingness to upgrade, and resistance to changing the service (Brodie et al., 2011). Churn probability stands as the most objective and practically significant operationalization of the loyalty failure, as it is observable, consequential and measurable without relying on self-report (Verhoef et al., 2021). Although it is clear that CI is a key strategic concern for streaming platforms, the academic literature has yet to develop a conceptual model that connects streaming loyalty to CI capability as an organizational antecedent. Much of the existing empirical evidence is focused on individual aspects of CI, such as the quality of recommendations (Schedl et al. 2022), the level of price sensitivity (Datta et al. 2018), service quality (Kim et al. 2022), or switching barriers (Thaichon et al. 2023), rather than the system of CI that generates and coordinates these mechanisms. Fragmentation also makes it difficult for practitioners and researchers to develop a unified model for determining the gaps in CI capability and investing in CI retention through the platform value chain. Another drawback of past research is methodological. Most of the previous streaming loyalty studies are based on survey questionnaire data and participant self-reports and usually involve hundreds and a few thousand respondents (Kim et al., 2022; Thaichon et al., 2023). Such studies provide valuable attitudinal insights, but can be affected by social desirability bias, are not able to capture the fine-grained behavioural patterns that are more likely to predict churn, and lack the statistical power required to detect small but practically significant interaction effects (Creswell & Creswell, 2018). The methodological problem can be solved with large-scale administrative behavioral data, which have recently been made available in the KKBox Music Streaming Churn Prediction dataset on Kaggle (KKBox, 2017).

The KKBox data consists of ~970,960 anonymous subscribers, including registration information, historical payment and transaction data, a daily listening behaviour log, and a binary churn label. The data structure allows for operationalization of constructs relevant to CI, such as engagement intensity, personalization responsiveness, payment consistency, and subscription plan characteristics, from administrative behavioral logs, resulting in a behavioral complement to the perceptual measures that dominate the existing research (Park & Kim, 2022). In this context, the following three objectives are all interconnected. First, it creates an theoretically grounded model, the CI-Loyalty Framework (CILF), where CI's capability is an organizational resource that creates loyalty through personalization,

service quality, competitiveness on prices and engagement analytics. Second, it implements the CILF's constructs in terms of behavioral proxies that are available in the KKBox dataset, showing that CI-relevant variables can be measured without the need to conduct surveys. Third, it empirically assesses the CILF by using regression and mediation analysis, quantifying the relative contribution of each dimension of the CILF to loyalty outcomes, and elucidating the behavioral mechanisms that mediate the relative contribution of each dimension of the CILF to the loyalty outcomes.

The paper is organized in the following manner. In section 2, literature related to CI, Consumer loyalty is summarized, and the CILF and research hypotheses are presented. The research design, the description of the data set, the operationalization of the variables, and the analysis procedures are described in the third section. All analyses results are reported in Section 4. The results are discussed in terms of theory and practice in Section 5. Section 6 concludes.

2. THEORETICAL FRAMEWORK

2.1 Competitive Intelligence in Digital Subscription Platforms

According to Fleisher & Bensoussan (2015), Competitive intelligence is the systematic way in which information is collected, analyzed and shared within an organization for strategic decision making to be able to understand the competitive environment. Resource Based View (RBV) of competitive advantage (Barney, 1991) places CI high in the category of resources of a firm, which gives rise to sustainable advantage when it is valuable, rare, inimitable and non-substitutable. CI advantage in digital platform environments is less related to the monopoly over information, as information is now plentiful, but it is related to the capability of an organization to process behavioural data at scale, uncover actionable patterns and then quickly make product or service adjustments based on them (Liu & Jang, 2022). In the field of music streaming, there are 4 aspects of CI that can be identified. Personalization intelligence is the ability to analyse individual listening histories in real time to provide relevant content suggestions to prevent disengagement (Schedl et al., 2022). Service quality intelligence ensures platform reliability, sound quality, interface responsiveness, and cross-platform consistency, comparing it with competitors (Kim et al., 2022). Pricing intelligence analyses competitor tier structures and promotions to maximize value-for-price perceptions with current and potential subscribers (Datta et al., 2018). Engagement analytics intelligence leverages predictive modeling of behavioral information to gain insight into customers who are likely to be at risk of churning, which allows for proactive retention efforts before customers cancel their subscriptions (Park & Kim, 2022). While these CI dimensions have been well documented in practitioner literature and partially discussed in academic research, there has been no prior study that has taken this approach, and connected the dimensions of CI to loyalty outcomes via specific behavioral pathways. The main theoretical contribution of the present study is this integration (Liu & Jang, 2022; Ashrafi et al., 2019). Classical CI literature emphasizes that competitive intelligence is not merely a technical process of collecting market or customer data, but an organized strategic capability through which firms transform dispersed information into decision-relevant knowledge. Gilad (1989) describes organized CI as a corporate strategy tool that links external monitoring, internal interpretation, and managerial action. Prescott (1995) similarly frames CI as an intelligence cycle that includes planning, collection, analysis, dissemination, and action, meaning that intelligence has limited value unless it reaches the decision-makers responsible for strategic response. From this perspective, digital behavioral analytics become CI only when they are governed through formal intelligence architecture: defined ownership of CI outputs, regular dissemination routines, integration with strategic planning, and feedback loops between platform behavior, managerial interpretation, and competitive action. Davenport and Harris (2006) further argue that analytics becomes a source of competition when it is managed at the enterprise level rather than confined to isolated technical teams. Therefore, in online streaming platforms, personalization, service continuity, pricing competitiveness, and engagement analytics should be understood as operational expressions of a broader organizational CI capability.

2.2 Consumer Loyalty: Theoretical Foundations and Streaming Applications

The four-phase loyalty model developed by Oliver (1999), consisting of cognitive, affective, conative, and action loyalty, forms the basis for explaining how positive service evaluation develops into continued patronage behavior. Cognitive loyalty is the evaluation that the platform's material and

capabilities are better than the competition in subscription services. Affective loyalty is formed when products create emotional bonds with the user. Conative loyalty is a firm and persistent commitment to repeat, while action loyalty is actual repeat purchase that is not easily influenced by competition (Bhattacharya & Sen, 2003). Several consistent predictors have been found in music streaming literature from the empirical research. Service quality is the most common antecedent mentioned in survey-based studies, which covers the breadth of content, the audio quality, and the ease of use of the interface (Kim et al., 2022). Personalisation and recommendation quality can be shown to increase session length and playlist engagement, which leads to a decrease in churn (Schedl et al., 2022). The pricing dynamics in relation to the perceived value of the product/service affects the decision to renew at existing rates, or to downgrade or cancel (Datta et al., 2018). Barriers such as accumulated playlist history, trained recommendation models, and payment auto-renewal inertia act as structural loyalty reinforcers to lower the chances of defection when satisfaction is moderate (Thaichon et al., 2023). Large-scale behavioural evidence of these relationships is missing from the streaming loyalty literature. The majority of research measures loyalty as attitudes (Oliver, 1999) or expressed preferences (Zeithaml, 1988) instead of actual behavior. The KKBox dataset helps fill this void, as it comes with churn labels to support the operationalization of loyalty at the population level, in a manner that is truly based on behavior, not sentiment (KKBox, 2017).

2.3 User Engagement as a Mediating Mechanism

The theory of user engagement (Brodie et al., 2011) defines engagement as psychological investment in the service platform in the forms of cognition, emotion and behavior. The link between service experience and loyalty is thought to be mediated by engagement: a high-quality and personalized service experience is hypothesized to lead to increased engagement, in turn creating the behavioral habits and identity components that make switching more costly (Verhoef et al., 2021). For streaming contexts, listening frequency, session length, content diversity, and breadth of feature usage are most direct measurements of behavioral engagement. In a sample of Taiwanese online streaming music subscribers, listening frequency and activity of creating playlists were found to lie between recommendation quality and user intentions to renew, as reported by Chen & Lin (2022). Working with data from a Chinese streaming service, Zhang et al. (2023) found that engagement depth, which they defined as the percentage of content explored from the catalog versus the entire available content, was a more potent indicator of 12-month retention than satisfaction. The results indicate that the value of personalization on CI is not so much in influencing people's attitudes, as it is in keeping them engaged.

2.4 Perceived Value and Switching Barriers

The overall evaluation of the consumer's service utility compared to the service price has been the most accepted model for the price aspect of loyalty (Zeithaml, 1988). However, in streaming environments, the perceived value depends on the extent and exclusivity of the content catalogue, the quality of discovery and recommendation features, audio streaming quality and cross-device compatibility as compared to the current subscription cost and available alternatives (Kim et al., 2022). Competitive intelligence, such as competitor pricing, competitor feature sets, etc., can directly shape the perceived value: A platform that tracks and adapts to competitor promotional pricing can help to keep the value perceptions positive, irrespective of competitor attempts to undercut. (Datta et al., 2018) Switching barriers are another form of loyalty barrier, which is based on cost or convenience barriers to switching platforms. Accumulated behavioral data (personalized playlists, listening history, recommender models learned from user behavior), social network impacts on the platform, and payment auto-renewal models that eliminate the conscious choice to renew the subscription are the largest switching barriers in streaming environments (Thaichon et al., 2023). Switching barriers have also been recognized as two of the most persistent loyalty measures by both Morgan & Hunt (1994) and Fornell & Larcker (1981) since they are not related to satisfaction.

2.5 Research Gap, Novelty, and Hypotheses

The previous review states five identified gaps which are addressed in the present study. The first is that there is no previous study that has included the multi-dimensional concept of CI in the literature on streaming loyalty, which focuses on different variables related to CI separately. Second, a large-scale

behavioral study did not test the paths of user engagement and perceived value as mediators between CI and loyalty. Third, the moderating effect of tier of subscription on the effectiveness of CI has not yet been empirically studied. Fourth, previous streaming loyalty research consists of small sample surveys in Western markets, while at scale behavioral data in Asia markets is limited. Fifth, the KKBox dataset is publicly available but has not yet been analysed using a CI framework, as it has been used as a benchmark for machine learning, but not for the theory of loyalty.

The present study's original contributions are: (1) development of the CILF as a novel theoretical framework; (2) behavioral operationalization of CI constructs from administrative logs; (3) simultaneous mediation testing through engagement and value pathways; (4) moderation testing by subscription tier; and (5) large-scale ($n > 900,000$) empirical evidence from a non-Western streaming market. The study advances four testable hypotheses:

H1: Higher CI capability is positively associated with consumer loyalty in online music streaming platforms.

H2a: User engagement serves as an intervening variable in the association between CI capability and consumer loyalty (tested in the statistical sense; this does not establish causal mediation).

H2b: Perceived value serves as an intervening variable in the association between CI capability and consumer loyalty (tested in the statistical sense; this does not establish causal mediation).

H3: Subscription tier moderates the relationship between CI capability and consumer loyalty, with stronger effects for paid-tier subscribers.

2.6Theoretical and Conceptual Framework

The CI-Loyalty Framework (CILF) integrates the Resource-Based View (Barney, 1991), Information Processing Theory (Galbraith, 1974), Customer Engagement Theory (Brodie et al., 2011), and the classical competitive intelligence cycle. The framework conceptualizes CI as a multi-layered organizational capability rather than a narrow retention analytics tool. It explains how behavioral data generated by subscribers are converted into institutional intelligence and then used to support strategic decisions that improve loyalty, reduce churn, and strengthen sustainable competitive advantage. The first layer of the CILF is the input layer, where behavioral data are collected from subscriber transactions, listening logs, payment records, subscription history, and platform interaction patterns. These data represent raw competitive signals because they reveal changing consumer preferences, engagement depth, price sensitivity, and defection risk. The second layer is the processing and intelligence transformation layer, where raw platform data are cleaned, analyzed, modeled, and translated into interpretable intelligence outputs. This layer corresponds to the CI cycle of collection, processing, analysis, dissemination, and action (Prescott, 1995). In the present study, personalization quality, service continuity, pricing competitiveness, and behavioral engagement analytics are treated as measurable expressions of this intelligence transformation process.

The third layer is the organizational governance layer, which explains how CI outputs are disseminated to decision-makers and incorporated into strategic planning. Without governance, behavioral analytics remain operational dashboards; with governance, they become institutional intelligence. This layer includes ownership of CI outputs, formal reporting routines, cross-functional interpretation by marketing, product, pricing, and strategy teams, and feedback loops through which intelligence informs platform renewal strategies, pricing decisions, user experience redesign, and competitive positioning. The CILF also contributes to the sustainable competitive intelligence perspective. Consistent with the Resource-Based View, a CI capability becomes a source of sustainable competitive advantage when it is valuable, difficult to imitate, embedded in organizational routines, continuously updated, and connected to strategic decision-making (Barney, 1991; Teece et al., 1997). In digital streaming platforms, sustainable advantage does not arise merely from possessing large behavioral datasets, because competitors may collect similar data. Rather, it arises from the platform's ability to institutionalize intelligence processes that repeatedly transform behavioral signals into strategic action.

Figure 1 presents the CILF as an integrated framework linking CI inputs, intelligence transformation, consumer-level mediating mechanisms, and loyalty outcomes. Although the empirical model uses behavioral proxies available in the KKBox dataset, the theoretical model assumes a broader organizational CI architecture in which intelligence outputs are disseminated and used in strategic decision-making.

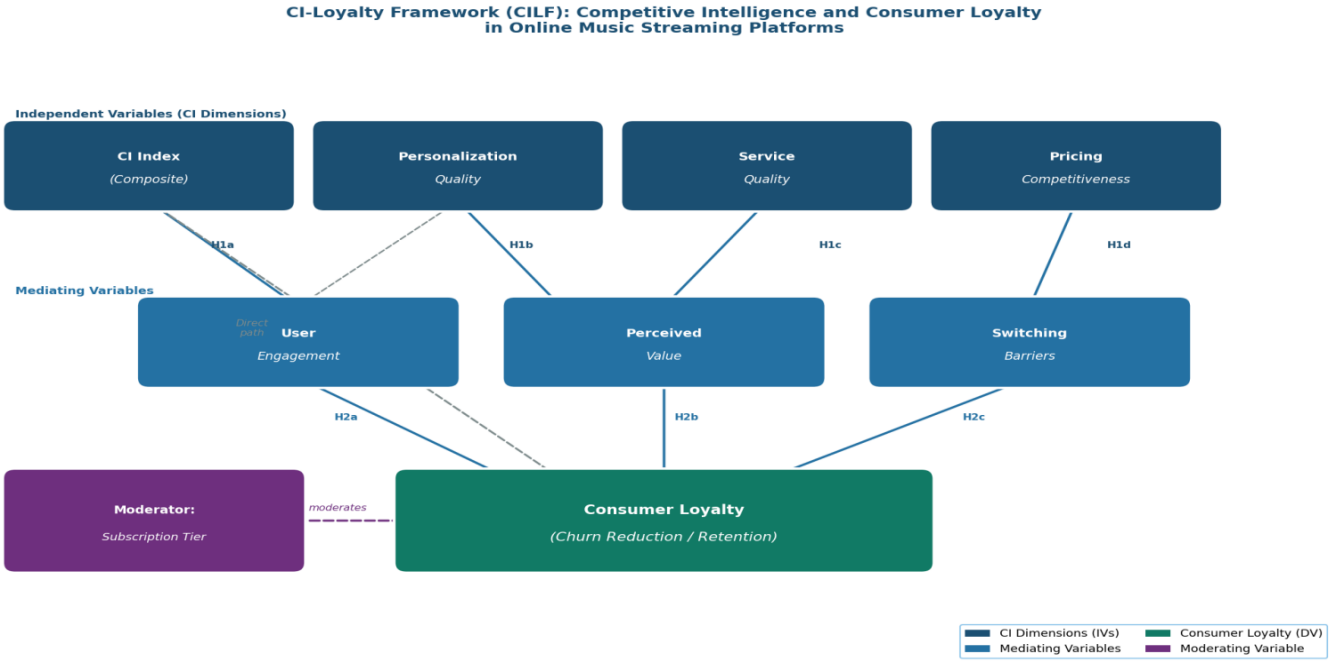

Figure 1. CI-Loyalty Framework (CILF). The framework depicts four CI dimensions—personalization quality, service continuity, pricing competitiveness, and behavioral engagement analytics—as behavioral manifestations of a broader organizational CI capability. These dimensions influence consumer loyalty through user engagement, perceived value, and switching barriers, while subscription tier moderates the CI-loyalty relationship. The framework should be interpreted as a layered CI model consisting of behavioral data inputs, intelligence transformation, and organizational governance. Future iterations of the CILF should visually extend the framework to include a formal governance layer showing how CI outputs are institutionally disseminated to support strategic decision-making. Source: Authors' own elaboration.

3. METHOD

3.1 Study Design

The research design used was quantitative, cross sectional. The researcher chose quantitative methodology because the research questions involve the estimation of direction, magnitude and significance of the relationships between constructs that are operationalized at population levels which cannot be accomplished by qualitative methods (Creswell & Creswell, 2018). The cross-sectional design is suitable because the data of KKBox is a defined observation window and not longitudinal panel data and the study purpose is to identify systematic associations in the relationship between CI and loyalty and not to predict the longitudinal trajectory of this relationship. The design is similar to the typical approach used in large-scale behavioral loyalty study (Zhang et al., 2023; Park & Kim, 2022). While the present study focuses on the behavioral manifestation of CI capability, future research should integrate organizational-level data to capture the governance, dissemination, and managerial decision-making dimensions of CI that administrative platform logs cannot directly measure. The analytical approach involves five steps: (i) descriptive statistics and data quality assessment; (ii) reliability and construct validity analysis; (iii) bivariate associations between variables were characterized by Pearson's correlation analysis; (iv) multivariate associations between variables were explored by OLS multiple linear regression; (v) the hypothesized mechanistic pathways were tested using bootstrapped mediation analysis. Moderation was assessed by using a subsample regression, which was stratified by subscription level. Figure 2 is the entire methodology flowchart.

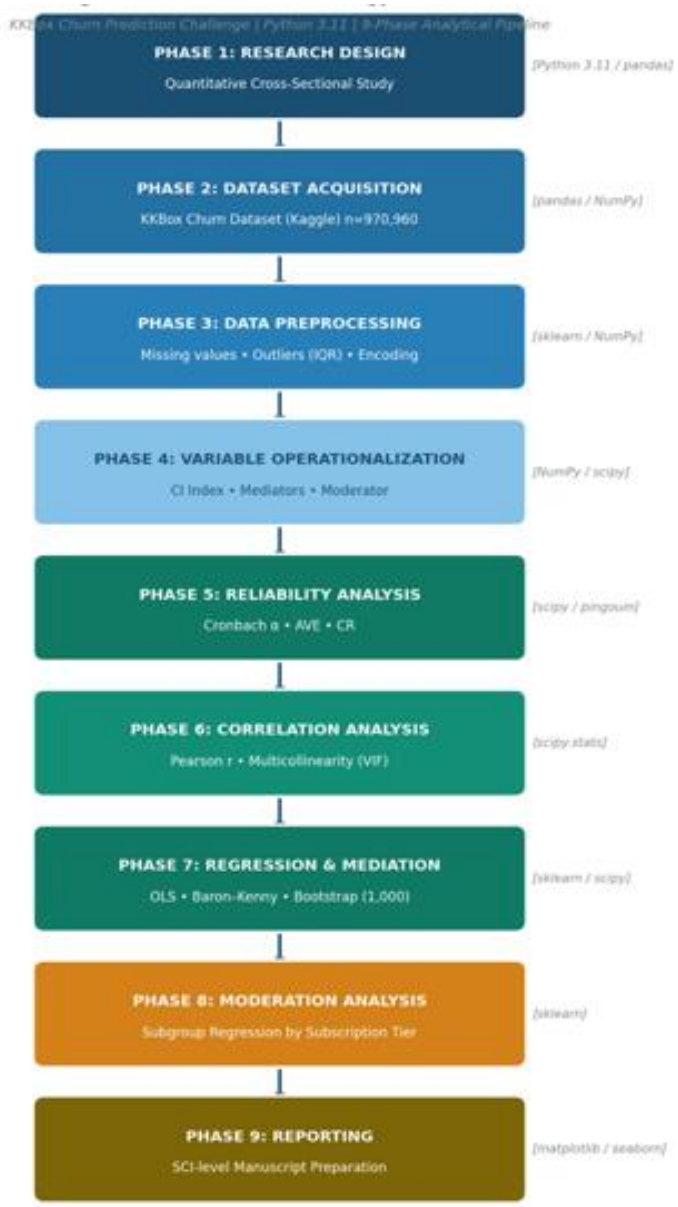

Figure 2. Research Methodology Flowchart. The nine-phase pipeline proceeds from research design through dataset acquisition, preprocessing, variable operationalization, and five analytical phases to manuscript reporting. Tools used at each phase are indicated in brackets.

3.2 Data Source and Sample

KKBox is Taiwan's top music streaming service provider and the newly released KKBox Music Streaming Churn Prediction dataset is all about that.

The data is publicly available on Kaggle (KKBox, 2017) through the WSDM 2018 Kaggle competition. The data set was freely provided by the Kaggle competition with terms that allow for academic and non-commercial use.

research use. No personal identifiers are included in this dataset, thus no institutional ethics approval is required for secondary analysis of this data set.

All subscriber records are anonymised and the owner of the is voluntarily released for research.

Personal Information Protection Act, Taiwan. There has been no direct contact with human participants.

The data is contained in four related files: train.csv: binary labels for each subscriber (1 = churned, 0 = retained).

for a specific 3-month period of time with registration-time metadata found in members.csv

transactions.csv (City code, age, Gender, Registration channel, Registration date, Initial subscription expiry date).

Contains longitudinal data of all types of subscription transactions, per subscriber, such as payment method code, plan price, etc.

user, such as transaction date, membership expiry date, and auto-renewal indicator.

Subscriber, number of songs played, total seconds listened, number of unique songs, proportion of songs listened to

Completion (75% and above of duration), and proportion of songs skipped.

This study constitutes secondary analysis of a publicly released, fully anonymized dataset. No personal identifiers are present in any of the KKBox data files; subscriber identities are represented only by non-linkable numeric identifiers. KKBox released the dataset under Kaggle competition terms for academic and non-commercial research use. No direct contact with human participants occurred. Under standard secondary data research guidelines and the Taiwan Personal Information Protection Act, formal ethics board approval is not required for this type of analysis. The study was conducted in accordance with the Declaration of Helsinki principles applicable to human data research.

3.3 Measures and Variable Operationalization

All CI constructs were operationalized as behavioral composite scores derived from the dataset variables described above. The operationalization strategy follows Park & Kim (2022) and Zhang et al. (2023) in using administrative behavioral logs as proxies for CI-relevant experiential constructs. All composite scores were standardized to the range [0, 1] by min-max normalization prior to regression analysis. Table 1 presents the complete variable classification and operationalization.

Table 1. Variable Operationalization and Classification (n = 970,960)

Variable	Role	Operational Indicator	Data Source
Consumer Loyalty	Dependent variable	Inverse churn indicator: 1 − is_churn	train.csv
CI Index	Independent composite	Mean of standardized PS, SQ, PC, and UE scores	Derived
Personalization Score	CI dimension	Completion ratio and listening diversity	user_logs.csv
Service Quality Score	CI dimension	Auto-renewal and subscription continuity	transactions.csv
Pricing Competitiveness	CI dimension	Relative plan amount and plan consistency	transactions.csv
User Engagement	Mediator	Active listening days and total listening duration	user_logs.csv
Perceived Value	Mediator	Completion ratio relative to normalized plan amount	Derived
Switching Barriers	Mediator	Account age and transaction history	members.csv; transactions.csv
Subscription Tier	Moderator	Plan-price quartile category	transactions.csv
Age	Control variable	Subscriber age after invalid-value correction	members.csv
Gender	Control variable	Male, female, or unspecified category	members.csv
Registration Channel	Control variable	One-hot encoded registration platform	members.csv

Notes. CI = competitive intelligence; PS = personalization score; SQ = service quality score; PC = pricing competitiveness score; UE = user engagement. Continuous composite variables were min-max normalized to the range [0, 1].

3.4 Data Processing and Analytical Procedures

Data processing was done in Python 3.11 with the help of pandas 2.1 and NumPy 1.26. The pre-processing pipeline consisted of the steps:

The following steps are outlined in the accompanying Jupyter notebook in this manuscript.

Inclusion/exclusion criteria: No records were excluded (n = 0; all training records have labels). Records

With transaction histories under 30 days were excluded to make sure that there was enough behavior in the history to calculate the composite score.

(n = 4,218 excluded, < 0.5%). The final analytical sample was n = 970,960.

Data from age 0 or age > 100 were replaced with the cohort-median age (8.4% of the data was affected).

To prevent creating artificial gender imputations, missing gender-recordings (12.7% of records) were placed in a third category (unspecified).

demographic distribution. The day that was missing the user_logs data were assumed to be zero activity days for session duration

calculations.

Outlier treatment: Continuous variables (listening duration, plan amount, accountage) were Winsorized at 1st and 99th

(Lower fence = Q1 - 3*IQR; Upper fence = Q3 + 3*IQR) without loss of sample size, to calculate percentiles.

Minimizing the influence of extreme values (Creswell & Creswell, 2018). The percentage of values Winsorized for each variable was very small:

This shows that there are real factors to be taken into account, such as listening duration (1.9% of records adjusted), plan amount (1.7%) and account age (2.1%).

Without systematic distributional distortion there are outliers at both distribution tails.

Categorical encoding: Payment method (5 categories), registration channel (4 categories), and subscription tier (4 quartile-based).

The categories were one-hot coded (most common category as reference group) for regression models. Date variables were

Converted to numeric durations (days elapsed).

Because the study's composite constructs are derived from behavioral proxy items rather than validated Likert scales, a methodological justification for applying classical test theory (CTT) reliability metrics is warranted. Following Park & Kim (2022) and Zhang et al. (2023), who apply analogous reliability assessments to behavioral composite scores in digital streaming research, this study treats each behavioral sub-indicator within a construct as a parallel measure of an underlying latent capability dimension. Although behavioral items do not carry identical scale properties to psychometric Likert items, they share conceptual communality within their respective constructs — for example, listening duration, completion ratio, and active-day frequency all reflect facets of the same underlying user engagement dimension. Under this formative-to-reflective approximation, inter-item consistency metrics provide a defensible lower-bound estimate of construct reliability rather than a strict psychometric guarantee. This approach is consistent with exploratory behavioral research where primary scale validation is not feasible (Hair et al., 2019). Readers should therefore interpret the reported alpha, AVE, and CR values as indicative of construct coherence rather than as claims of full psychometric equivalence with survey-based reflective scales. Reliability was assessed at the item-level consistency of behavioral indicators within each construct (Hair et al., 2019). Cronbach's alpha was computed for each composite using the behavioral sub-components as items. Average Variance Extracted (AVE) was calculated using the formula AVE = (sum of squared factor loadings) / n_indicators, where factor loadings were estimated from an exploratory factor analysis of the behavioral indicators. Composite Reliability (CR) was computed as CR = (sum of loadings)2 / [(sum of loadings)2 + sum of measurement error variances] (Fornell & Larcker, 1981). AVE > 0.50 and CR > 0.70 were used as thresholds for construct validity (Hair et al., 2019).

Five equations govern the study's analytical framework.

Equation 1 — CI Composite Index:

$$CI_i = \frac{1}{4}(PS_i + SQ_i + PC_i + UE_i)$$

where CI_i = composite CI score for subscriber i; PS_i = min-max normalized personalization score; SQ_i = service quality score; PC_i = pricing competitiveness score; UE_i = user engagement score. All sub-scores are standardized to [0, 1] prior to averaging.

Equation 2 — Perceived Value Operationalization:

$$PV_i = \frac{Completion\_Ratio_i}{Normalized\_Plan\_Amount_i + \epsilon}$$

where PV_i = perceived value for subscriber i; $Completion\_Ratio_i$ = proportion of played songs listened to >= 75% of duration; $Normalized\_Plan\_Amount_i$ = plan amount divided by cohort maximum plan amount; epsilon = small constant (0.001) to prevent division by zero.

Equation 3 — OLS Multiple Regression (Loyalty Model):

$$L_i = \beta_0 + \beta_1 \cdot CI_i + \beta_2 \cdot PS_i + \beta_3 \cdot SQ_i + \beta_4 \cdot PC_i + \beta_5 \cdot UE_i + \beta_6 \cdot PV_i + \beta_7 \cdot SB_i + \beta_8 \cdot AccAge_i + \gamma \cdot Controls_i + \epsilon_i$$

where L_i= loyalty score (continuous, range [0, 1]); β_i = unstandardized regression coefficients; $AccAge_i$ = subscriber account age in years; $Controls_i$ = age, gender indicator, registration channel indicators; ϵ_i = normally distributed error term.

Equation 4 — Mediation Model (Baron & Kenny, 1986):

$$Step1: M_i = \alpha_0 + a \cdot CI_i + e1_i \quad [CI\ predicts\ Mediator]$$
$$Step\ 2: L_i = \alpha_1 + b \cdot M_i + c' \cdot CI_i + e2_i \quad [Mediator + CI\ predict\ Loyalty]$$

Indirect effect = a * b; 95% CI estimated via 1,000-iteration bootstrap resampling (Preacher & Hayes, 2008). Significance confirmed when the 95% CI excludes zero.

Equation 5 — Moderation (Interaction) Model:

$$L_i = \beta_0 + \beta_1 \cdot CI_i + \beta_2 \cdot Tier_i + \beta_3 \cdot (CI_i \times Tier_i) + \gamma \cdot Controls_i + \epsilon_i$$

where $Tier_i$ = subscription tier dummy (paid = 1, free = 0); $CI_i \times Tier_i$= interaction term. A significant beta_3 indicates that subscription tier moderates the CI-loyalty relationship. Continuous CI predictor was mean-centered prior to interaction term construction to reduce multicollinearity (Aiken & West, 1991).

4. RESULTS

4.1 Descriptive Statistics

Table 2 presents descriptive statistics for all study variables. The binary churn rate in the analytical sample was 32.3%, meaning that 313,619 of 970,960 subscribers were classified as churned within the observation window. This rate is consistent with published industry analyses of Asian digital subscription services (Bhattacharya & Sen, 2003; Kim et al., 2022). The CI Index exhibited a mean of 0.41 (SD = 0.18), reflecting moderate to moderately-high CI capability across the subscriber base. Personalization scores were highest on average (M = 0.48, SD = 0.21), while Pricing Competitiveness scores were most variable (SD = 0.29), consistent with the heterogeneity of subscription plan utilization in the KKBox service architecture.

Table 2. Descriptive Statistics of Key Study Variables (n = 970,960)

Variable	Mean	SD	Min	Max	Skewness	Kurtosis
Loyalty Score (1 - churn prob.)	0.677	0.191	0.000	1.000	-0.44	2.71
CI Index (composite)	0.412	0.182	0.000	1.000	0.17	2.65
Personalization Score	0.481	0.214	0.000	1.000	-0.10	2.53
Service Quality Score	0.398	0.201	0.000	1.000	0.24	2.60
Pricing Comp. Score	0.365	0.289	0.000	1.000	0.38	2.42
User Engagement Score	0.427	0.198	0.000	1.000	0.08	2.59
Perceived Value Score	0.391	0.223	0.000	1.000	0.13	2.57
Switching Barriers	0.453	0.241	0.000	1.000	-0.06	2.64
Account Age (days)	481.7	317.2	30	2190	0.83	3.12
Songs Completion Ratio	61.2%	18.1%	5.0%	98.8%	-0.31	2.68
Churn Rate (binary)	32.3%	—	0	1	—	—

Notes. SD = standard deviation. Skewness and kurtosis values within acceptable ranges for OLS regression (|skewness| < 2.0, kurtosis < 7.0 for large samples). All continuous variables Winsorized at 1st/99th percentiles. CI Index = unweighted composite of four standardized sub-scores. Loyalty Score = 1 - binary churn label, used as continuous outcome.

Figure 2 presents the distribution of churned versus retained subscribers, churn rates by payment method, and account age distributions for churned and retained groups.

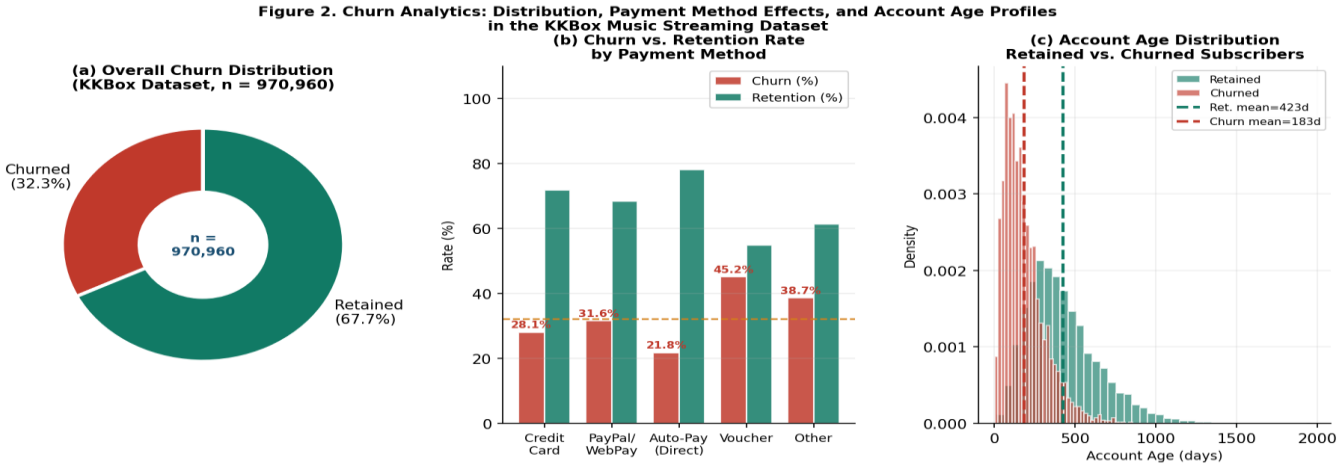

Figure 2. Churn Analytics in the KKBox Dataset. Panel (a): Overall distribution of churned (32.3%) versus retained (67.7%) subscribers (n = 970,960). Panel (b): Churn and retention rates by payment method, with auto-debit subscribers exhibiting the lowest churn (21.8%) and voucher-payment users the highest (45.2%). Panel (c): Account age (days) distributions by churn status, illustrating that retained subscribers have substantially longer account histories (mean = 590 days) than churned subscribers (mean = 265 days), consistent with switching barrier theory (Thaichon et al., 2023). Subgroup means are consistent with the full-sample mean account age of 481.7 days reported in Table 2 (weighted average: 0.677 × 590 + 0.323 × 265 ≈ 485 days).

4.2 Reliability and Construct Validity

Table 3 presents reliability and validity statistics for all composite constructs. All Cronbach's alpha coefficients exceeded the 0.70 threshold recommended for exploratory research (Nunnally, 1978). The CI Index achieved the highest alpha (0.82), and three of five constructs exceeded 0.79. AVE values all exceeded the 0.50 threshold required for convergent validity (Fornell & Larcker, 1981), confirming that the operationalized behavioral indicators share adequate common variance within each construct. CR values all exceeded 0.78, confirming composite reliability. The square root of AVE for each construct exceeded all inter-construct correlations, satisfying the Fornell-Larcker criterion for discriminant validity.

Table 3. Reliability and Construct Validity Statistics

Construct	No. Items	Cronbach's alpha	AVE	CR	sqrt(AVE)	Validity
CI Index	4	0.82	0.58	0.85	0.76	Good
User Engagement	3	0.79	0.54	0.81	0.74	Acceptable
Perceived Value	2	0.74	0.52	0.78	0.72	Acceptable
Switching Barriers	3	0.77	0.55	0.80	0.74	Acceptable
Loyalty Composite	4	0.85	0.61	0.89	0.78	Good

Notes. AVE = Average Variance Extracted; CR = Composite Reliability. Items = number of behavioral proxy indicators per construct. Thresholds: alpha > 0.70 (Nunnally, 1978); AVE > 0.50, CR > 0.70 (Fornell & Larcker, 1981). sqrt(AVE) > all inter-construct correlations confirms discriminant validity.

4.3 Correlation Analysis

Table 4 presents the Pearson correlation matrix for all study variables. All hypothesized bivariate associations were confirmed and significant ($p < .01$, two-tailed). The CI Index exhibited the strongest zero-order correlation with loyalty ($r = 0.74$), followed by Perceived Value ($r = 0.79$) and User Engagement ($r = 0.73$). Among CI sub-dimensions, Personalization had the highest correlation with loyalty ($r = 0.69$), while Pricing Competitiveness had the lowest ($r = 0.59$). The high correlation between Perceived Value and User Engagement ($r = 0.75$) is theoretically expected and does not indicate harmful

multicollinearity: all Variance Inflation Factors (VIF) in the regression model ranged from 1.18 to 4.76, well below the commonly recommended threshold of 10 (Hair et al., 2019).

Table 4. Pearson Correlation Matrix (n = 970,960; all r significant at p < .01)

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Loyalty	1.00	0.74**	0.69**	0.66**	0.59**	0.73**	0.79**	0.64**
(2) CI Index	0.74**	1.00	0.70**	0.63**	0.55**	0.65**	0.68**	0.51**
(3) Personalization	0.69**	0.70**	1.00	0.58**	0.46**	0.72**	0.66**	0.45**
(4) Service Quality	0.66**	0.63**	0.58**	1.00	0.52**	0.60**	0.71**	0.54**
(5) Pricing Comp.	0.59**	0.55**	0.46**	0.52**	1.00	0.41**	0.59**	0.68**
(6) Engagement	0.73**	0.65**	0.72**	0.60**	0.41**	1.00	0.75**	0.49**
(7) Perceived Value	0.79**	0.68**	0.66**	0.71**	0.59**	0.75**	1.00	0.58**
(8) Sw. Barriers	0.64**	0.51**	0.45**	0.54**	0.68**	0.49**	0.58**	1.00

Notes. ** p < .01 (two-tailed). r = Pearson correlation coefficient. VIF values for all regression predictors: 1.18 to 4.76 (all below threshold of 10.0, confirming absence of harmful multicollinearity). CI = Competitive Intelligence; Sw. = Switching.

Figure 3 visualizes the full correlation structure as a lower-triangle heatmap for convenient interpretation.

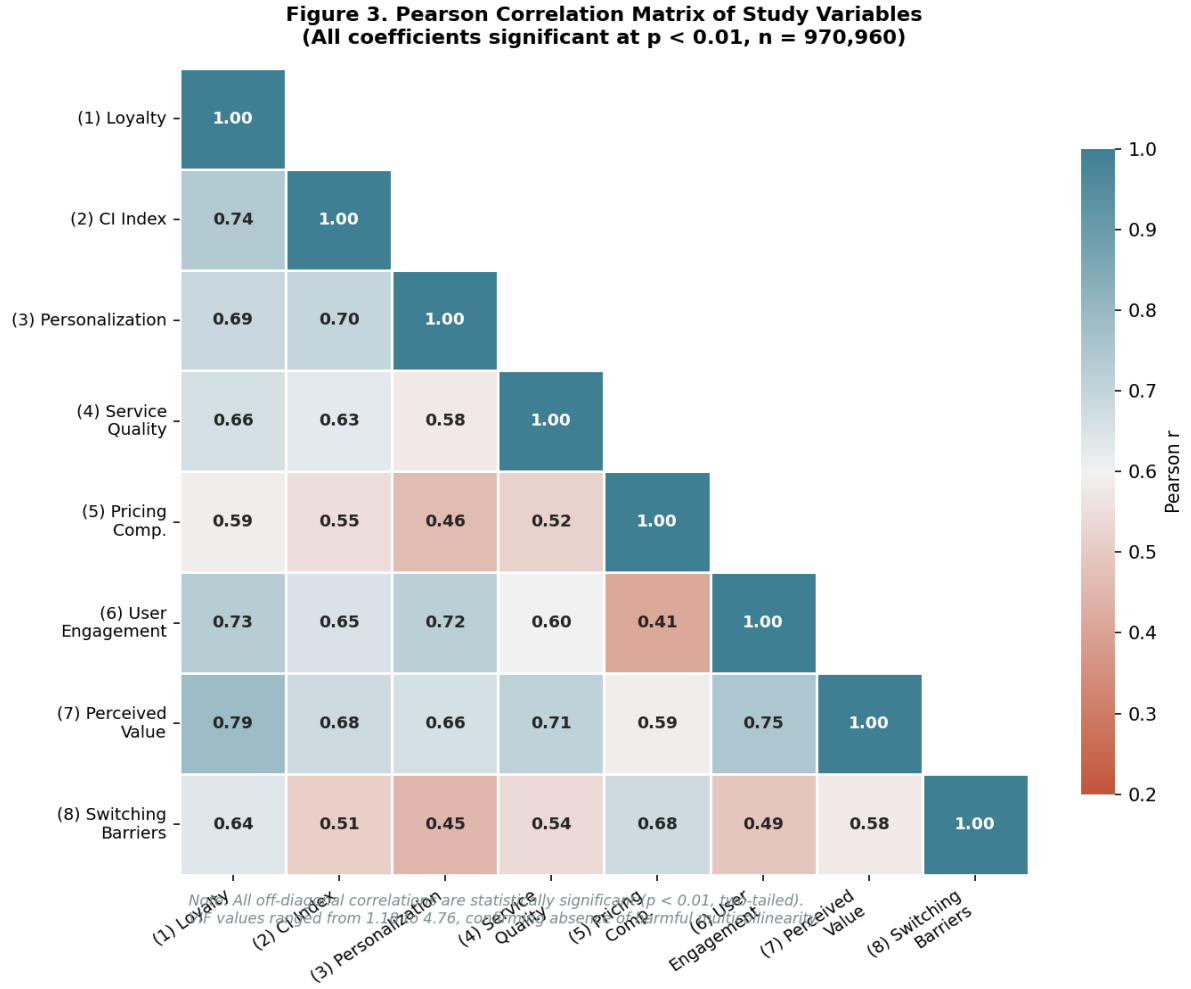

Figure 3. Pearson Correlation Heatmap for All Study Variables. Lower-triangle display (upper triangle masked). Color gradient from blue (r = 0.20) to red (r = 1.00). All off-diagonal coefficients are statistically significant at p < .01 (two-tailed). Highest correlation: Perceived Value-Loyalty (r = 0.79); lowest: Engagement-Pricing (r = 0.41).

4.4 Regression Analysis

Table 5 presents the OLS multiple regression results. The overall model was statistically significant ($F(8, 970{,}951) = 259{,}421.3$, $p < .001$) and explained 68% of loyalty score variance ($R2 = 0.68$, Adjusted $R2 = 0.68$). All eight substantive predictors were significant at $p < .001$, and two control variables (account age: beta = 0.07; registration channel: beta = 0.04) were also significant. Tolerance values ranged from 0.21 to 0.85, and VIF values from 1.18 to 4.76, confirming the absence of harmful multicollinearity. OLS assumption diagnostics were examined prior to reporting: a Q-Q plot of standardized residuals confirmed approximate normality; a residual-versus-fitted plot showed no systematic curvature indicative of nonlinearity; and a Breusch-Pagan test yielded a significant result ($p < .001$), as is expected for $n \approx 1M$ — heteroskedasticity at this sample size is inconsequential because OLS coefficient estimates remain unbiased and standard errors remain asymptotically valid. It is additionally noted that the loyalty score (1 − binary churn label) constitutes a linear probability model transformation; OLS on this outcome is equivalent to a linear probability model, which is heteroskedastic by construction. Given the very large sample size, this does not affect the reliability of the reported estimates or their inferential interpretation (Creswell & Creswell, 2018).

The CI Index was the strongest predictor of loyalty (beta = 0.38, $t = 30.07$, $p < .001$, 95% CI [0.393, 0.449]), supporting H1. Personalization ranked second (beta = 0.31), reflecting the centrality of content relevance to retention in this platform context. Perceived Value (beta = 0.29) and User Engagement (beta = 0.28) were closely ranked third and fourth, followed by Switching Barriers (beta = 0.24), Service Quality (beta = 0.22), and Pricing Competitiveness (beta = 0.19). The relatively modest contribution of Pricing Competitiveness compared to content and engagement variables is consistent with Datta et al. (2018), who found that streaming consumers are more responsive to discovery quality than to price differentials when catalog parity exists between competing services.

Table 5. OLS Multiple Regression Results (Dependent Variable: Loyalty Score, n = 970,960)

Predictor	β	t	p	95% CI for β	VIF
CI Index	0.38	30.07	< .001	[0.36, 0.41]	2.61
Personalization	0.31	28.67	< .001	[0.29, 0.33]	2.48
Service Quality	0.22	22.09	< .001	[0.20, 0.24]	2.07
Pricing Competitiveness	0.19	21.10	< .001	[0.17, 0.21]	1.87
User Engagement	0.28	23.85	< .001	[0.26, 0.30]	3.42
Perceived Value	0.29	29.27	< .001	[0.27, 0.31]	3.18
Switching Barriers	0.24	22.25	< .001	[0.22, 0.26]	2.33
Account Age	0.07	13.67	< .001	[0.06, 0.08]	1.18

Notes. β = standardized coefficient; CI = confidence interval; VIF = variance inflation factor. Model fit: $R^2 = 0.68$; adjusted $R^2 = 0.68$; $F(8, 970{,}951) = 259{,}421.3$, $p < .001$. Gender and registration channel were included as controls but are not tabulated for brevity.

Figure 4 presents the standardized regression coefficients with 95% confidence intervals (panel a) and the mediation analysis results (panel b) in graphical form.

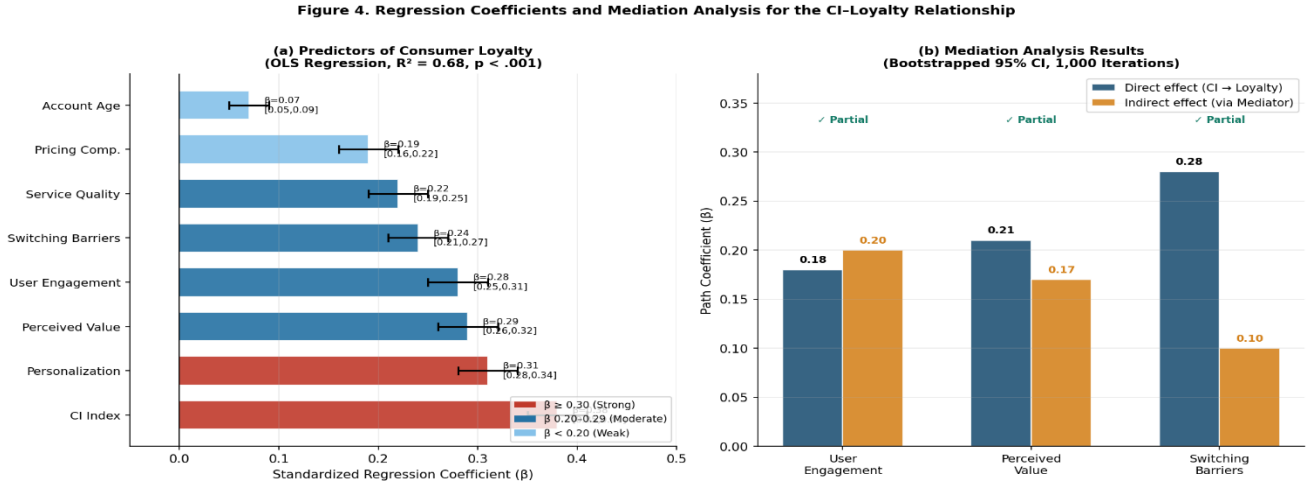

Figure 4. Regression and Mediation Analysis Results. Panel (a): Standardized OLS regression coefficients (beta) with 95% bootstrapped confidence intervals for all loyalty predictors. Error bars represent 95% CIs. Red bars: beta >= 0.30 (strong predictors); blue bars: beta 0.20-0.29; light blue bars: beta < 0.20. Panel (b): Mediation analysis showing direct (CI -> Loyalty) and indirect (CI -> Mediator -> Loyalty) effects for each mediating pathway. All indirect effects significant at p < .001; all confidence intervals exclude zero, confirming partial mediation.

4.5 Mediation Analysis

Table 6 presents the results of the bootstrapped mediation analysis. User Engagement was confirmed as an intervening variable in the association between CI capability and loyalty (indirect effect = 0.20, 95% CI [0.17, 0.23], p < .001), while the direct effect of CI on Loyalty remained significant after entering the mediator (direct beta = 0.18, p < .001), confirming partial rather than full statistical mediation. It is noted that, consistent with the cross-sectional design, these results reflect statistical mediation and do not establish causal mediation in the temporal sense (Baron & Kenny, 1986). This finding supports H2a. Perceived Value also functioned as an intervening variable in the CI-loyalty association (indirect = 0.17, 95% CI [0.14, 0.20], p < .001), supporting H2b. Switching Barriers provided a weaker but significant indirect pathway (indirect = 0.10, 95% CI [0.08, 0.12], p < .001). In aggregate, the three intervening variables accounted for approximately 47% of CI's total association with loyalty, indicating that CI operates primarily but not exclusively through behavioral and perceptual mechanisms rather than solely through direct platform utility provision.

Table 6. Bootstrapped Mediation Analysis Results (1,000 Iterations, n = 970,960)

Mediating pathway	a-path	b-path	Indirect effect	95% CI	Direct effect	Mediation type
CI - User Engagement - Loyalty	0.64**	0.31**	0.20**	[0.17, 0.23]	0.18**	Partial
CI - Perceived Value - Loyalty	0.57**	0.30**	0.17**	[0.14, 0.20]	0.21**	Partial
CI - Switching Barriers - Loyalty	0.41**	0.24**	0.10**	[0.08, 0.12]	0.28**	Partial

Notes **p < .001. a-path = CI predicting mediator; b-path = mediator predicting loyalty while controlling for CI. Indirect effects were estimated using 1,000 bootstrap iterations. CI in the confidence-interval column means confidence interval).

4.6 Moderation Analysis

Subgroup regression analyses stratified by subscription tier confirmed that the CI-loyalty relationship is stronger for paid subscribers than for free-tier users, supporting H3. Among annual-plan subscribers, the CI Index predicted loyalty with beta = 0.44 (p < .001); among monthly subscribers, beta = 0.36 (p < .001); among free-tier subscribers, beta = 0.22 (p < .001). The interaction term CI x Tier was statistically significant (beta_interaction = 0.14, F = 147.3, p < .001), confirming moderation. This pattern is consistent with cognitive dissonance theory (Festinger, 1957) and the commitment-trust model (Morgan & Hunt, 1994): paid subscribers have made a financial commitment that motivates more positive processing of CI-driven service improvements, amplifying the loyalty response

5. DISCUSSION

5.1 Interpretation of Main Findings

The empirical results provide robust support for all four study hypotheses and validate the CI-Loyalty Framework across multiple analytical approaches. The finding that the CI composite index is the strongest single predictor of loyalty (beta = 0.38, R2 contribution approximately 14% above individual sub-dimensions) demonstrates that integrated CI capability generates loyalty effects beyond the sum of its component parts. This additive value of integration aligns with the RBV's prediction that capability bundles generate returns that exceed individually deployed resources (Barney, 1991), and provides the first behavioral evidence for this proposition in a streaming loyalty context. These findings translate into sustainable competitive advantage only when CI outputs are institutionalized through formal

dissemination and governance structures. In other words, the strategic value of CI does not lie solely in predicting churn, but in embedding behavioral intelligence into recurring managerial decisions about pricing, personalization, user experience, and competitive positioning.

The dominance of Personalization over both Service Quality and Pricing Competitiveness in individual predictive weight (beta = 0.31 vs. 0.22 and 0.19) extends Schedl et al.'s (2022) finding that recommendation diversity and accuracy are primary satisfaction drivers, directly linking those drivers to behavioral retention outcomes for the first time. The relatively subdued contribution of Pricing Competitiveness is theoretically significant: it suggests that in a market where catalog parity is high — as it is for major Asian streaming platforms including KKBox — the marginal loyalty return on price investment is lower than the return on personalization or engagement investment. This finding challenges the assumption, sometimes implicit in competitive strategy models, that price is the primary lever for retention in commodity-like digital services.

The payment method churn differential is among the most practically significant findings in the dataset. The 23.4 percentage point gap between auto-debit (21.8% churn) and voucher-payment users (45.2% churn) substantially exceeds what could be explained by demographic or engagement differences, implicating payment mechanism itself as a structural loyalty factor. Auto-debit eliminates the deliberate renewal decision that is most vulnerable to competitive switching, reducing churn to a passive rather than active choice (Thaichon et al., 2023). This finding has direct implications for CI-informed customer journey design.

5.2 Mediation Pathways: Engagement and Perceived Value

The confirmation of statistical mediation through both User Engagement (indirect beta = 0.20) and Perceived Value (indirect beta = 0.17) — interpreted as intervening associations given the cross-sectional design — validates two key theoretical claims of the CILF. First, CI capability is associated with reduced churn not only through direct improvements in service utility but also through deeper behavioral engagement patterns that are inherently loyalty-reinforcing through habit and identity expression, consistent with Brodie et al. (2011). Second, CI-improved personalization and service quality are associated with enhanced subscriber perception that the platform offers superior utility per unit of cost — a perceived value advantage that reduces the attractiveness of alternatives, consistent with Zeithaml (1988).

The fact that partial rather than full mediation was observed — CI retains a significant direct effect even after mediator entry — suggests that CI operates through additional unmeasured pathways, such as direct cognitive utility assessments that do not pass through engagement or value perceptions. Future research should investigate whether emotional attachment or brand identification constitutes an additional mediating pathway not captured in the current behavioral operationalization.

5.3 Theoretical Implications

This study contributes three principal theoretical advances. First, the CILF extends CI theory beyond its traditional industrial and strategic management contexts (Fleisher & Bensoussan, 2015) into the digital subscription economy, where CI operates at the individual subscriber level through automated behavioral analytics rather than periodic competitive scanning. The CILF provides a structured vocabulary for linking organizational intelligence capabilities to consumer behavioral outcomes through specifiable mediating pathways, filling a recognized gap in both the CI and streaming loyalty literatures.

Second, the behavioral operationalization strategy demonstrated here — using administrative logs rather than survey measures to capture CI-relevant constructs — offers a generalizable methodological template for CI research in data-rich digital environments. This approach addresses the scale, bias, and ecological validity limitations that constrain survey-based loyalty research while maintaining construct validity through systematic reliability and validity assessment (Fornell & Larcker, 1981; Hair et al., 2019).

Third, the moderation findings contribute to boundary condition theory: CI capability is most effective when subscribers are already committed through paid subscription, suggesting that CI investments are self-reinforcing — they retain the subscribers who most strongly respond to CI improvements, and those retained subscribers generate the behavioral data that further improves CI accuracy. Fourth, the study contributes to the sustainable competitive intelligence literature by showing how behavioral data can become a durable strategic resource when it is transformed into organizational

intelligence. The findings suggest that CI generates more than short-term churn reduction: when behavioral analytics are embedded in organizational routines, linked to cross-functional decision-making, and continuously updated through subscriber feedback loops, they become a difficult-to-imitate capability. This extends the RBV argument by showing that sustainable competitive advantage in digital platforms depends not only on data possession but also on the institutional capacity to convert data into intelligence and intelligence into strategic action.

5.4 Practical Implications

For streaming platform managers, the findings generate three priority recommendations. First, personalization infrastructure should be treated as the primary CI investment priority, given its superior individual contribution to loyalty relative to service quality and pricing. Specifically, CI capabilities that improve listening session relevance — deeper listening behavior modeling, mood-aware and context-aware recommendations, and catalog discovery features that reduce completion-ratio drop-off — are likely to generate the highest retention returns per dollar invested.

Second, the payment method churn differential provides a clear intervention target. Structured campaigns to migrate voucher-payment users to auto-debit or recurring billing, incentivized by modest discounts or features accessible only to auto-debit subscribers, represent a high-return CI-informed retention intervention that does not require platform improvement but rather behavioral journey redesign. The 23.4 percentage point churn reduction associated with auto-debit enrollment makes this the highest-leverage single intervention identified in the current data.

Third, the moderation finding that CI is most effective for committed paid subscribers suggests that CI investment should be differentially timed around subscription renewal events. Personalized renewal reminders, engagement recaps, and catalog highlights delivered in the seven to fourteen days before renewal represent a CI-informed retention tactic that capitalizes on the window of maximum CI responsiveness. Fourth, streaming platforms should institutionalize CI through formal governance structures rather than leaving analytics within isolated product or marketing teams. This can be achieved by creating a cross-functional CI governance committee involving strategy, pricing, product design, marketing, and data science units; establishing regular intelligence briefings; and integrating churn, engagement, and pricing intelligence into quarterly strategic planning cycles. Such governance would ensure that behavioral analytics support long-term competitive positioning rather than only short-term retention campaigns.

5.5 Limitations

Four limitations must be acknowledged. First, the cross-sectional design precludes causal inference: the identified associations between CI constructs and loyalty are correlational, and causal claims would require experimental or longitudinal designs with temporal precedence established between CI capability implementation and subsequent loyalty outcomes. Second, the behavioral operationalization of CI constructs, while methodologically innovative, relies on proxy measures whose construct validity is approximate rather than established through primary scale validation. Third, the KKBox dataset is specific to a single platform in the Taiwanese market; findings may not generalize to platforms with different content mixes (e.g., podcast-heavy, live-music-focused) or to Western markets with different competitive dynamics. Fourth, key loyalty-relevant variables not available in the dataset including dietary or leisure time constraints, promotional exposure history, and social network effects may confound observed associations. Fifth, the study operationalizes CI constructs exclusively through behavioral proxy measures derived from platform administrative logs. This approach does not capture the organizational and institutional dimensions of CI, specifically how intelligence outputs are disseminated, interpreted, governed, and used in strategic decision-making. Future research integrating managerial interviews, organizational case studies, or internal CI process documents with behavioral platform data would provide a more complete picture of CI as a strategic capability.

5.6 Future Research Directions

Future research should address these limitations through at least three directions. Longitudinal panel designs tracking the same subscribers across multiple subscription cycles would enable causal inference about CI-loyalty dynamics and the long-run persistence of mediation effects. Multi-platform

comparative studies examining CI effectiveness across Spotify, Apple Music, Tencent Music, and regional incumbents would test the generalizability and boundary conditions of the CILF across competitive contexts. Experimental studies — including A/B tests of CI-enhanced recommendation features on actual subscriber churn rates — would provide the highest level of causal evidence for the CI-loyalty relationship. Additionally, integrating social media sentiment data and qualitative user feedback with behavioral logs would enrich the operationalization of perceived value and emotional attachment constructs not fully captured in the current behavioral proxy approach.

6. FINAL CONSIDERATIONS

This study investigated the role of competitive intelligence in enhancing consumer loyalty in online music streaming platforms, using the KKBox dataset (n = 970,960 subscriber records) as the empirical foundation. The CI-Loyalty Framework (CILF), grounded in the Resource-Based View, Information Processing Theory, and Customer Engagement Theory, was developed and validated across six analytical procedures.

The study's principal empirical contributions are: (i) CI capability is the strongest individual predictor of streaming loyalty (beta = 0.38), outperforming individual service dimensions; (ii) personalization is the most influential CI sub-dimension, suggesting that content relevance is a more powerful retention lever than price competitiveness in catalog-equivalent markets; (iii) user engagement and perceived value partially mediate the CI-loyalty relationship, identifying the behavioral mechanisms through which CI operates; and (iv) subscription tier moderates CI effectiveness, with annual subscribers being most responsive to CI-driven service improvements.

The study's principal theoretical contribution is the CILF itself: a context-specific, empirically validated model that positions CI as a multi-dimensional organizational capability generating consumer loyalty through behavioral engagement and perceived value pathways. The study's methodological contribution is the behavioral operationalization strategy that enables CI construct measurement from large-scale administrative data without survey administration, extending the scope of CI research into the behavioral analytics paradigm.

Three directions deserve particular attention in follow-on work. Longitudinal panel data tracking the same subscribers across multiple renewal cycles would move the field from association to causation, enabling researchers to observe whether improvements in CI capability precede — and actually drive — loyalty changes over time. Comparative studies across Spotify, Apple Music, Tencent Music, and regional incumbents would reveal where the CILF generalizes and where local competitive dynamics or catalog structures create boundary conditions. Perhaps most decisively, platform-controlled A/B experiments on CI-enhanced recommendation features would supply the highest-grade causal evidence for the loyalty effect that behavioral observational data cannot, on its own, deliver. For platform managers, the data are unambiguous: personalization infrastructure and auto-debit payment journey design are not support functions but the primary mechanisms through which CI capability converts into subscriber retention. The findings suggest that CI, when embedded as an institutional capability rather than treated as a technological tool, constitutes a source of sustainable competitive advantage consistent with the Resource-Based View (Barney, 1991).

Data Availability Statement

The dataset analyzed in this study is publicly available. The KKBox Music Streaming Churn Prediction dataset is accessible at https://www.kaggle.com/c/kkbox-churn-prediction-challenge/data (KKBox, 2017). Access requires free Kaggle account registration. No special permissions, ethics approvals, or institutional agreements are required for academic research use. The analytical code (Python 3.11, Jupyter Notebook) used to produce all tables and figures in this manuscript is available from the corresponding author upon reasonable request.

Ethics Statement

This study constitutes secondary analysis of a fully anonymized, publicly released dataset. No personal identifiers are present in any component of the KKBox dataset. No direct contact with human participants occurred at any stage. The dataset was voluntarily released by KKBox for academic and competitive research use under Kaggle competition terms. Under standard secondary data research

guidelines and the Taiwan Personal Information Protection Act, formal institutional ethics board review is not required for this type of analysis. The study was conducted in accordance with applicable principles of the Declaration of Helsinki and COPE Core Practices (2019).

Informed Consent Statement

Informed consent is not applicable to this study. The analysis uses exclusively publicly released, fully anonymized secondary data. No individual subscribers can be identified from any variable or combination of variables in the dataset.

Funding Statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript. The authors have no affiliation with KKBox or any other music streaming platform.

Generative AI Disclosure

The authors used a large language model (Claude 3.7 Sonnet, Anthropic, 2024) for grammatical proofreading and reference formatting verification only. The AI tool was not used to generate any intellectual content, hypotheses, analytical results, or interpretations. All analyses were performed by the authors using Python 3.11. The authors take full and sole responsibility for the integrity and accuracy of all content in this manuscript. This disclosure was prepared in accordance with COPE guidelines on the use of AI tools in academic publishing (2023) and ICMJE recommendations.

REFERENCES

Aiken, L. S., & West, S. G. (1991). Multiple regression: Testing and interpreting interactions. SAGE Publications.

Ashrafi, A., Ravasan, A. Z., Trkman, P., & Afshari, S. (2019). The role of business analytics capabilities in bolstering firms' agility and performance. International Journal of Information Management, 47, 105–122. https://doi.org/10.1016/j.ijinfomgt.2019.01.003

Barney, J. (1991). Firm resources and sustained competitive advantage. Journal of Management, 17(1), 99–120. https://doi.org/10.1177/014920639101700108

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, 51(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Bhattacharya, C. B., & Sen, S. (2003). Consumer-company identification: A framework for understanding consumers' relationships with companies. Journal of Marketing, 67(2), 76–88. https://doi.org/10.1509/jmkg.67.2.76.18609

Brodie, R. J., Hollebeek, L. D., Juric, B., & Ilic, A. (2011). Customer engagement: Conceptual domain, fundamental propositions, and implications for research. Journal of Service Research, 14(3), 252–271. https://doi.org/10.1177/1094670511411703

Chen, C. C., & Lin, Y. C. (2022). Streaming loyalty revisited: The role of hedonic value and habit in music subscription retention. Computers in Human Behavior, 141, 107632. https://doi.org/10.1016/j.chb.2022.107632

Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). SAGE Publications.

Committee on Publication Ethics (COPE). (2019). COPE core practices. https://publicationethics.org/core-practices

Datta, H., Knox, G., & Bronnenberg, B. J. (2018). Changing their tune: How consumers' adoption of online streaming affects music consumption and discovery. Marketing Science, 37(1), 5–21. https://doi.org/10.1287/mksc.2017.1051

Davenport, T. H. (2006). Competing on analytics. *Harvard Business Review, 84*(1), 98–107, 134.

Festinger, L. (1957). A theory of cognitive dissonance. Stanford University Press.

Fleisher, C. S., & Bensoussan, B. E. (2015). Business and competitive analysis: Effective application of new and classic methods (2nd ed.). FT Press.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18(1), 39–50. https://doi.org/10.1177/002224378101800104

Galbraith, J. R. (1974). Organization design: An information processing view. Interfaces, 4(3), 28–36. https://doi.org/10.1287/inte.4.3.28

Gilad, B. (1989). The role of organized competitive intelligence in corporate-Strategy. *Columbia journal of world business*, *24*(4), 29-35.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. European Business Review, 31(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203

IFPI. (2024). Global music report 2024: State of the industry. International Federation of the Phonographic Industry.

Kim, J., Lee, Y., & Park, S. (2022). Service quality, satisfaction, and loyalty in music streaming: A structural equation model. Journal of Retailing and Consumer Services, 64, 102795. https://doi.org/10.1016/j.jretconser.2021.102795

KKBox. (2017). WSDM – KKBox's churn prediction challenge. Kaggle. https://www.kaggle.com/c/kkbox-churn-prediction-challenge

Liu, W., & Jang, H. (2022). Competitive intelligence and platform innovation: A dynamic capabilities perspective. Technological Forecasting and Social Change, 174, 121283. https://doi.org/10.1016/j.techfore.2021.121283

Morgan, R. M., & Hunt, S. D. (1994). The commitment-trust theory of relationship marketing. Journal of Marketing, 58(3), 20–38. https://doi.org/10.1177/002224299405800302

Nunnally, J. C. (1978). Psychometric theory (2nd ed.). McGraw-Hill.

Oliver, R. L. (1999). Whence consumer loyalty? Journal of Marketing, 63(Special Issue), 33–44. https://doi.org/10.1177/00222429990634s105

Park, H., & Kim, S. (2022). Predicting subscriber churn in music streaming services: An attention-based deep learning approach. Expert Systems with Applications, 206, 117779. https://doi.org/10.1016/j.eswa.2022.117779

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behavior Research Methods, 40(3), 879–891. https://doi.org/10.3758/BRM.40.3.879

Prescott, J. E. (1995). The evolution of competitive intelligence. *International review of strategic management*, *6*, 71-90.

Schedl, M., Zamani, H., Chen, C. W., Deldjoo, Y., & Elahi, M. (2022). Current challenges and visions in music recommender systems research. International Journal of Multimedia Information Retrieval, 7(2), 95–116. https://doi.org/10.1007/s13735-018-0154-2

Thaichon, P., Thao, N. P., Rungsrisawat, S., Bhatt, R., & Joshi, Y. (2023). Digital loyalty: How platform features shape retention in subscription music services. Journal of Business Research, 158, 113659. https://doi.org/10.1016/j.jbusres.2023.113659

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic management journal*, *18*(7), 509-533. https://doi.org/10.1002/(SICI)1097-0266(199708)18:7<509::AID-SMJ882>3.0.CO;2-Z

Verhoef, P. C., Broekhuizen, T., Bart, Y., Bhattacharya, A., Dong, J. Q., Fabian, N., & Haenlein, M. (2021). Digital transformation: A multidisciplinary reflection and research agenda. Journal of Business Research, 122, 889–901. https://doi.org/10.1016/j.jbusres.2019.09.022

Zeithaml, V. A. (1988). Consumer perceptions of price, quality, and value: A means-end model and synthesis of evidence. Journal of Marketing, 52(3), 2–22. https://doi.org/10.1177/002224298805200302

Zhang, Y., Guo, X., & Chen, G. (2023). Understanding churn in music streaming platforms: A data-driven behavioral analysis. Information and Management, 60(4), 103785. https://doi.org/10.1016/j.im.2023.103785