A Work Project, presented as part of the requirements for the Award of a Master's degree in Business Analytics from the Nova School of Business and Economics.

Improving Personalized Recommendations for Cold-Start Users on the NetEase Cloud Music Platform: User Preference Elicitation Recommendation

Cristina Pereira Tiago

Work project carried out under the supervision of:

Yufei Shen

17-12-2024

Abstract

The rise of big data and rapid digitization in the digital media industry have made recommendation systems essential for delivering relevant, personalized content to users. In the music streaming sector, platforms like NetEase face the cold-start problem when recommending content to new users with minimal interaction data. To address this, we developed and compared various techniques, including a User Similarity-Based Recommendation Algorithm, a User Preference Elicitation Recommendation Algorithm, a DeCS-Inspired Recommendation Algorithm, and a Discriminative Frequent Itemsets model. Our findings show that the DeCS-Inspired model performs best in data-rich scenarios, while Demographic-Based methods excel in cold-start situations. To optimize performance, we propose a hybrid approach that combines Demographic-Based techniques for cold-starts and transitions to the DeCS-Inspired model as user data grows.

Key Words: Recommendation System, Information System, Cold-Users, Music Streaming Service

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

1. Introduction

"We engage with automation and machine learning every day in ways we may not realize. Consider the most common online habits: browsing social media, shopping on Amazon, or binge-watching on Netflix. These platforms rely on "recommendation systems"" to analyze tastes and preferences and try to serve up the products, content, or people they believe you want to see." (What Netflix's Recommendation Systems Can Teach Us About The Computing Challenges Of The Near Future s.d.). The swift transition to digital systems and services has transformed the way users interact with information across the internet. Information systems now play a critical role in managing vast catalogs of media content, catering to diverse user profiles with varying preferences and behaviors. The diverse nature of both content and user needs makes delivering relevant items in response to a user query a complex challenge. At the core of these information systems lie recommendation systems, which aim to present users with the most personalized and relevant items. *Figure 1* provides an abstract overview of the key components that constitute a recommendation system.

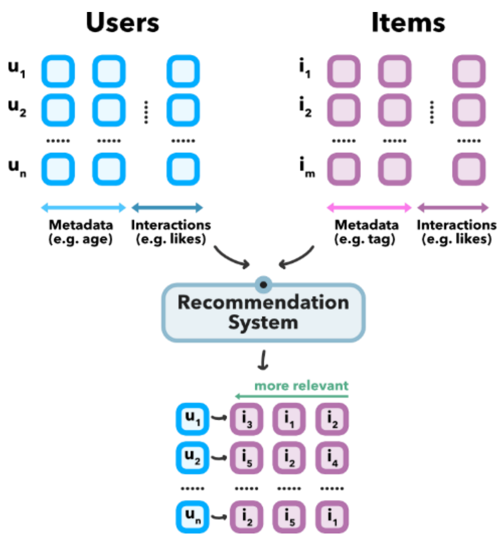


Figure 1 - Recommendation System Key Components Overview

The growing reliance on recommendation systems in the music streaming industry highlights the importance of personalization in attracting users. The global market for music streaming is expected to generate $29.60 billion in 2024 and $33.97 billion in 2027, growing through increased affordability, advancement in mobile technology, and greater access in emerging markets (Chadha e Jain 2022; Music Streaming - Worldwide | Statista Market Forecast s.d.). A major challenge recommender systems face is the **Cold Start Problem** (Chadha e Jain 2022), which can be divided into three categories: the **new community problem (**no interactions between any user and items), the **new-item problem** (recently added items), and the **cold-user problem** (brand new users on the platform). To address the cold-user problem, **Preference Elicitation** is a popular technique (Pu e Chen 2008). This method gathers user information during the registration process, either by directly asking users to provide their preferences or by leveraging existing data, such as from their social media accounts. This initial data helps generate an embedding for the user, serving as a recommendation starting point. The "cold-start problem" remains a critical bottleneck in the effectiveness of recommendation systems. The consequences are significant: users may exit platforms after unsatisfactory early experiences, and there is little visibility for new talent in an increasingly competitive market. In this work project, we have investigated the challenge of providing accurate/meaningful recommendations to cold users on the NetEase Cloud Music platform, where traditional approaches struggle because of minimal interaction records.

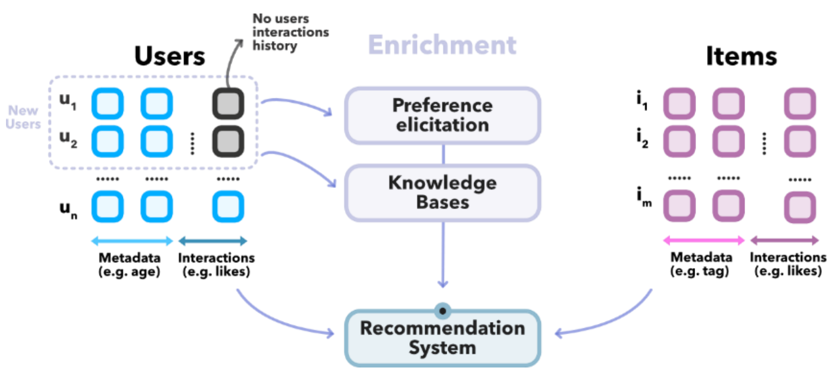


Figure 2 - Cold Users Problem *Abstract Components*

2. Literature Review

The cold-start problem can be a challenge in recommendation systems, particularly when limited user data is available, making it difficult to generate accurate and personalized suggestions. This is a common issue when new users or items lack interaction and therefore restricts recommendation accuracy. Researchers have proposed several innovative approaches to address this problem, these methods aim to extract more meaningful insights from limited data, allowing recommendation systems to provide relevant suggestions even in the absence of extensive user history. In traditional collaborative filtering (CF) systems, the absence of historical data constrains the system in detecting user preferences, while content-based filtering (CBF) approaches have limitations when item features don't represent user preferences properly (Pandey e Rajpoot 2016), (Li, et al. 2017). This section discusses key approaches that have been suggested by researchers, for instance, recommendations based on demographic attributes, such as age and gender, to enhance initial recommendations or combining several approaches to reduce the impact of data sparsity (Lam, et al. 2008), (Lika, Kolomvatsos e Hadjiefthymiades 2014), (Tahmasebi, et al. 2021).

2.1 Demographic Approaches

Demographic data allows systems to group users and items with similar characteristics, improving recommendation accuracy when historical data is lacking (Zhu, et al. 2020). Studies also highlight incorporating Deep Learning using auxiliary information, including but not limited - to text, user metadata, and other additional information that could generate more accurate predictions (Roy e Dutta 2022), (Steck, et al. 2021). One example of a Deep Learning based approach is DropoutNet which focuses on training neural networks for cold-start situations, using dropout techniques to create the effect of missing data during the training phase (Steck, et al. 2021).

2.2 Meta-Learning and Reinforcement Learning in Cold-Start Scenarios

Meta-learning has become an essential tool in addressing cold start challenges by allowing recommendation models to dynamically adapt to sparse data environments. (Wang, et al. 2022) showcase how reliable metadata and meta-learning approaches are in an overview that includes the expansion of existing collaborative filtering recommender systems with an enhanced correlation, inspired by (Vozalis e Margaritis 2003). Meanwhile, (Son 2016) conducted a comparative review on the new user cold-start problem in recommender systems. In this review, the model in question was found to perform less efficiently than its peers— NHSM, FARAMS, and HU-FCF—based on error metrics and computational time. On the opposite end, NHSM (Liu, et al. 2014) emerged as a robust solution to the cold-start problem, emphasizing improved accuracy in collaborative filtering. This new approach evaluated the distance between rating pairs, deviation from the median, and uniqueness of rating pairs, 7 while using the mean and variance of user behavior globally—rather than locally—leading to its high accuracy and reduced reliance on metadata. Reinforcement learning (RL) is also increasingly applied in recommender systems, enabling them to adapt dynamically to continuously evolving information and user preferences (Singh e Singh 2024).

2.3 Recommendation system techniques and related issues: a survey

There are numerous methods and models that, while attempting to solve certain issues in item recommendation, inadvertently exacerbate others. Common issues (Kumar e Thakur 2018) include data sparsity and high dimensionality, which are pervasive across different recommendation problems. Another issue is limited content analysis, where content-based filtering is restricted by relying solely on explicit characteristics of the recommended item.

2.4 Emerging Trends

(Karabila, et al. 2024) proposes an approach that takes on deep advanced learning and selfsupervised learning techniques, that promotes the implementation of a recommendation system that is an ensemble of three models: BERT-enhanced deep neural networks, selfsupervised learning and collaborative filtering approaches, capturing relationships in textual context, enhancing the user-item relationship understanding. Here, the combination of item and user similarity metrics with other content-based methods is key into providing accurate and well-founded recommendations. On the other hand, (Hafnar e Demšar 2024) draws a rather different perspective at the possibility of mitigating this issue, through the usage of the GPT 4 LLM (Large Language 8 Models) for personalized content generation in a context where there is a need of creating tailored game levels within minimal player interaction data. The relevance this paper brings into this problem is of a complete bypass on the user-item data need through the zero shot capabilities of LLMs, that allow for personalized and relevant outputs with minimal inputs.

2.5 Addressing the Cold-Start Problem in Recommender Systems Based on Frequent Patterns

(Panteli e Boutsinas 2023) proposed a methodology to solve the cold-start problem in recommender systems by integrating user clustering with the extraction of frequent discriminative patterns. Unlike conventional collaborative filtering approaches that struggle to recommend items to new users or products due to a lack of historical interactions, their method employs user features and previous purchase data from so-called "hot" users to generate recommendations. Users are first grouped based on demographic or contextual characteristics. Within each group, frequent item sets are extracted to identify combinations of products that existing users consistently like. Discriminative frequent item sets, which are patterns that are significantly more prevalent in one group of users than in others, are then extracted. These discriminative item sets serve as 'mind-blowing' buying behaviors for new and cold users who share similar characteristics with the group. By assigning relevant frequent items to these users, the method can effectively improve the user-product matrices, reducing dispersion problems and allowing for precise recommendations, even without initial rankings. Empirical evaluations on the MovieLens dataset show that this hybrid approach can achieve higher accuracy and handle both out-of-matrix (pure cold-start) and in-matrix (sparsity-related) prediction scenarios. The main contributions of this work include improving the diversity of recommendations, handling zero-rating scenarios without requiring extensive user interaction.

2.6 Limitations of Current Approaches

Though current approaches have improved greatly concerning the cold-start problems, there are still limitations. Approaches that are mostly based on demographic information and auxiliary data-based approaches may face issues of data availability or insufficient personalization if the additional data lacks relevance. Also, the incorporation of deep learning, although appropriate, can increase the intricacy and resource implications rendering them impractical for smaller resource-constrained (Steck, et al. 2021). Thus, in future work, the emphasis may be on developing low-cost yet effective models that include strategies for mitigating biases in recommendations without compromising their efficiency (Pandey e Rajpoot 2016), (Silva, et al. 2019).

2.7 Summary

The more recent strategies have shown us that the cold-start problem has several means to be solved, with the arrival of deep learning, contextual embeddings and adequately leveraged one-shot learning being key into creating what is, as of this day, the state-of- the-art panoply of means to tackling this problem. It portrays something that can be summed up as the contemporary existence of highly contextual, data broad approaches to reach a highly reasoned recommendation and a zero shot, more aggressive approach that requires very limited amounts of data, and accounts for the creation of recommendations by focusing on the abilities to be consistent, while having a tailoring aspect to it.

3. Context and Data

3.1 NetEase Research Context

NetEase Cloud Music (NCM), launched in 2013 by NetEase, Inc., has grown into one of China’s leading music streaming platforms. The platform stands out for its innovative features, fostering a community-oriented ecosystem on short-form music content and interactive functionalities. This dataset contains an impression-level interface of over 57 million recordings, reflecting users' behavioral data (clicks, likes,shares, comments) in addition to demographic/content metadata, all collected in November 2019.

3.2 Data Description

The **interaction_data** dataset from NetEase Cloud Music encompasses over 57 million impressions within the “cloud village” tab of NetEase’s app, on a sample with over 2 million users where each impression consists of music content card appearance to any user. Additionally, other datasets capture impression-related data, offering insights into users, content creators, and the impressions themselves. These datasets provide an added layer of detail, enriching the understanding of interactions and its stakeholders.

3.3 Metadata

The dataset captures all aforementioned user interactions through a binary field for each action in particular (**isClick, isLike, isShare, and isComment**). Additionally, the view time a field corresponds to the number of seconds a user has spent on a card after clicking it. On the card level, variables such as **mlogId, songId,** and **artistId** uniquely identify each card. Other attributes, including **publishTime**, **contentId**, **talkId**, and **cardType** respectively 10 describe the card’s publication timeline, content, topics and the format of each card. The dataset also includes demographic, use metrics(such as **age**, **gender** or **province**). (*Appendix 1*) provides a comprehensive overview of these metadata variables, summarizing the features that were used in our recommender systems.

3.4 Cold-Start Problem: Definition and Challenges

The cold-start problem is a restriction on the effectiveness of recommender systems, especially when new users or new items are involved. For this thesis, we focus on the **New User Problem**, where little to no data exists for a user’s engagement history. Hence, to define cold-start users in our context, we analyzed the available data and established criteria to identify a cold-start user, to then implement our recommender systems:

- **Registration Date:** Users who registered within the last six months.
- **Interaction Count:** Users with four or fewer actual interactions (e.g., clicks, likes, comments).
- **Activity Level:** Users with an activity level of three or lower on the platform.

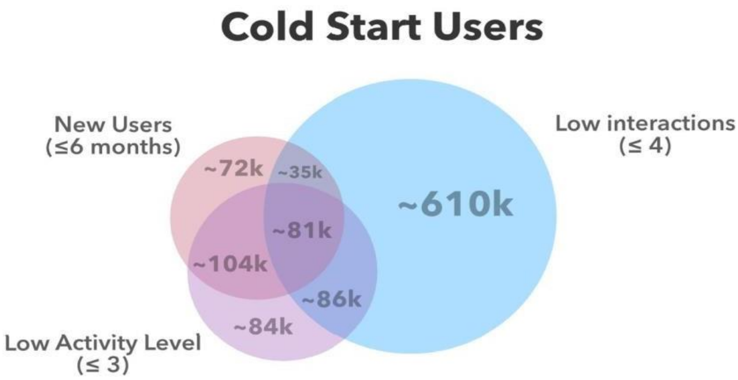


Figure 3 - Cold start users definition

This group accounts for only **3.88%** (*See Figure 3*) of the total user base and includes users who have few records available for making recommendations. The focus on users from the past six months was necessary, as older users lacked sufficient interaction records dating from November 2019, and thus their engagement history could not be measured effectively. The interaction limit was established by analyzing user engagement (*See Table 1*):

- **25%** of users had interactions less than or equal to four.
- **50%** had below seven interactions.

- **75%** had more than seventeen, with the average being 26.

count	2085480
mean	26.08
std	112.52
min	1
25%	4
50%	7
75%	17
max	22637

Table 1 - User impressions statistics

Users with ≤4 interactions and platform activity level ≤3 form a low-engagement group that poses significant personalization challenges due to limited data for reliable recommendations. Implicit feedback, such as "isClick" and "isLike," lacks explicit preference details, complicating user profiling and often leading to inaccurate suggestions. Limited demographic data, combined with the platform's dynamic nature and constant influx of new users and content, further hinders prediction accuracy. Addressing these issues requires tailored solutions to mitigate the cold-start problem for both users and new content effectively.

4. Exploratory Data Analysis

In the direction of understanding user behavior and data quality within the NetEase Cloud Music platform, we carefully analyzed key variables and interaction patterns present in the dataset. The analysis also addressed data inconsistencies, outliers, and the distribution of user engagement across the platform. Below is a detailed breakdown of the key findings and steps taken during the EDA.

4.1 Feature Explanation

The analysis of temporal variables, like impressTime converted to datetime format, revealed engagement patterns, with Friday and Saturday showing peak activity and hourly engagement peaking between 2 PM and 3 PM, as well as notable late-night usage after 10 PM. Discrepancies between the dt column and impressTime indicated potential data flaws. Invalid impressionPosition values (3.21% of records) and logical inconsistencies, such as clicks with zero or missing view times and positive view times without clicks, were identified. These records were excluded to ensure data integrity and accurate user behavior insights.Notably, the mlogViewTime column, representing the time users spent on viewing cards, showed that 95.2% of mlogViewTime were 0 or null, which suggests that, in most cases, the impression generated did not result in any meaningful view time. This finding highlights the platform's overall low engagement levels and the challenges of capturing user attention. Furthermore, the analysis revealed a classic long-tail distribution in user and content engagement, where a small subset of users and items accounted for the majority of interactions (*See Figure 4*).

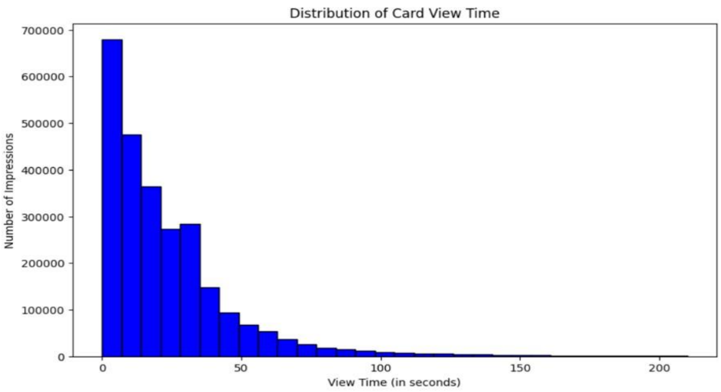


Figure 4 - Long-tail distribution on mlogViewTime

Other key user interaction features such as **isViewComment**, **isComment**, **isLike**, **isShare**, and **isIntoPersonalHomepage** were analyzed in detail. Logical errors were identified where interactions like likes, comments, or shares were recorded without an initial click (e.g., "isClick" = 0). Since such actions require opening the card, these discrepancies were anticipated as data collection artifacts and were removed. User behavior was further analyzed using the **userId** column, where we calculated the ratio of clicks to impressions for each user. This metric allowed us to better understand the engagement levels across the user base. These were combined into a new data frame and subsequently analyzed on the distribution of user interactions and engagement rates. The same type of analysis was performed on **mlogId** to identify cards with high impressions and clicks, along with their corresponding engagement rates. When examining the **songId** column, we found that the most used song and the most popular artist were both identified by the same ID: "AFBHAHLH". Besides, we have analyzed the distribution of **songId** to identify popular songs, showing that the average popular song was used around 5 times per month, with some notable outliers indicating highly popular tracks. Similarly, we analyzed the distribution of **artistId** to identify popular artists. This revealed that some artists have more songs used by users, indicating potential artist popularity or a larger catalog of appealing music. To understand the types of content shared, we analyzed the distribution of **type** column, which revealed that image-based content is more prevalent (53.9%) than video content (46.1%) in music logs. Additionally, the **publishTime** column revealed that the average creation date for content is 2 months before the data collection period. Demographic features, such as **age** and **gender**, provided more insights about the users. The **age** feature was categorized into age groups, creating a new variable that enabled a deeper exploration of user preferences and behavior across demographics. Summary

The EDA revealed significant insights into user behavior and data quality on the NetEase Cloud Music platform. This was aimed at understanding the key variables structure and their relationship with one another to help explain user activity and identify some data quality concerns. The data showed a very low rate of user engagement, with only 3.82% of impressions resulting in actions, which brings out the difficulties encountered in eliciting and predicting actions when designing recommendation systems. There is a need to understand the drivers of engagement as well as develop advanced techniques towards the enhancement of engagement levels. Additionally, we observed some discrepancies in the data where users appeared to engage (like, share, comment) before clicking on the card, which was flagged for deletion in order to preserve the data for future research. to formulate better and more precise strategies for modeling and making predictions.

5. Method

5.1 Evaluating Recommender Systems

The dataset includes dense demographic data, creator statistics, impression data, and detailed activity logs, making computations highly demanding due to its size and complexity. Key challenges involved high RAM requirements, with a maximum of 51 GB used, along with 225 GB of disk storage and 100 compute units. To address these constraints, a 5% sample of the data was used for exploratory analysis and model testing, avoiding memory overload. While a larger subset or full dataset might yield deeper insights, computational limitations restricted this possibility. The sampling approach allowed for methodology testing and validation in manageable segments. It also provided a foundation for scaling up as resources improve.

6. Users Preference Elicitation Recommendation

For this algorithm, we follow a content-based recommendation system. It was assumed that a cold user has already demonstrated a positive interaction. We have defined a positive interaction as being a weighted score of a user’s engagement with metadata associated with a card (e.g., clicking on a card). Given that the system's data is static, this interaction serves as a form of **preference elicitation**. As previously mentioned, preference elicitation, involves leveraging user preferences – either requested or inferred before user-item interaction - to improve recommendations in cold-user scenarios. In this particular case, we do not have the ability to request a priori users preferences. Therefore, we rely on users with at least one positive interaction to infer their preferences (e.g., family of item categories) from the available data. The elicited preferences are derived from the metadata of the interacted card. This metadata is subsequently utilized to identify and recommend other cards with similar characteristics, thereby tailoring recommendations to the inferred preferences of the user. The structure of the algorithm is illustrated in *Figure 5,* which will be explained and discussed in the following sections.

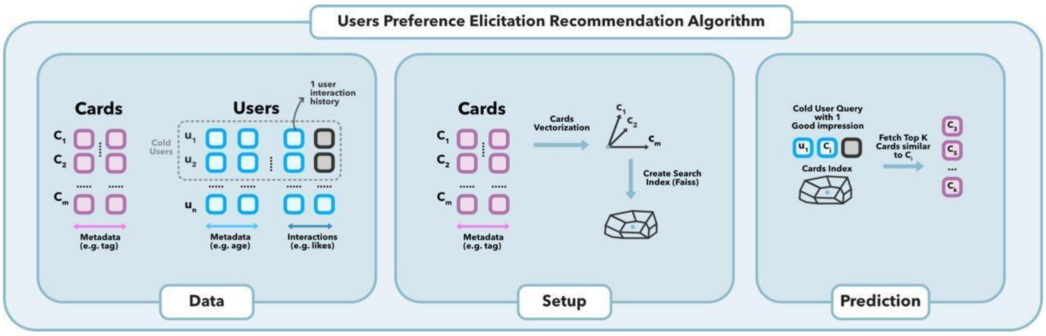


Figure 5 - Users Preference Elicitation Recommendation Algorithm Structure

6.1 Data: Preprocessing

Effective data preprocessing is essential for extracting meaningful insights from raw interaction data and enabling accurate recommendations. Our dataset did not include explicit user ratings for card interactions, which are often essential for personalized recommendations. To bridge this gap, we introduced an induced scoring mechanism. This approach assigns a meaningful score to each user-card interaction by leveraging weighted metadata associated with the interactions. This induced score serves as a proxy for user preferences and provides a foundation for building more effective recommendations. The mathematical formulation of this scoring mechanism is detailed below:

$$S_{weighted} = C \cdot wC + M \cdot wM + H \cdot wH + S \cdot wS + V \cdot wV + L \cdot wL$$

Where:

$S_{weighted}$ = Weighted Interaction Score; C = Indicator for "isClick" (binary: 0 or 1); M = Indicator for "isComment" (binary: 0 or 1); H = Indicator for "isIntoPersonalHomepage" (binary: 0 or 1); S = Indicator for "isShare" (binary: 0 or 1); V = Indicator for "isViewComment" (binary: 0 or 1) L = Indicator for "isLike" (binary: 0 or 1)

And the weights are defined as:

$w_C = 1/0.05$, $w_M = 1/0.0001$, $w_H = 1/0.0003$, $w_S = 1/0.0003$, $w_V = 1/0.0072$, $w_L = 1/0.0024$

We formulate a positive interaction as being an interaction where:

$$Sweighted \geq 1$$

The positive interactions are important as they enable us to filter users who have demonstrated meaningful engagement, providing a dataset for testing our algorithm. Specifically, we focused on users with at least 10 positive interactions. This threshold was established to ensure that, during evaluation, there is a sufficiently large pool of items to assess whether the recommended items align with user preferences. Without this threshold, the evaluation of our system could generate pessimistic results due to an insufficient item pool. Each user-card interaction is linked to a specific card. To identify cards similar to those associated with positive interactions, we need a method to quantify similarity between cards. To achieve this, we first establish a similarity distance measure and then transform the cards into vector representations. This transformation allows us to calculate similarity distances between cards in a quantifiable and operational manner.

To ensure the data is appropriately prepared for similarity computation, we transformed the cards categorical columns (e.g. contentId, talkId) into a one-hot encoded format. This approach transforms each category into a binary representation, allowing numerical algorithms to process these variables appropriately. Since there is no inherent order in these categories, one-hot encoding is a suitable choice.

6.2 Setup: Similarity Index

Pre-processing the cards table provides a vector representation. This vector is unique for each card and is a numerical compact way to represent a card. User similarity was calculated using the FAISS (Facebook AI Similarity Search) library (Douze, et al. 2024). Initially, we explored building a cosine similarity matrix (cards × cards) by computing the cosine of the angle between every pair of card vectors. While cosine similarity is a widely used method, we chose FAISS for its efficiency and speed in handling large-scale data. FAISS operates as an indexing mechanism, clustering vectors and enabling rapid retrieval of similar items using k-nearest neighbor searches. Furthermore, FAISS supports inner product indexing, which is equivalent to cosine similarity when vectors are normalized. Since our vectors are already normalized, FAISS provides a highly optimized solution for identifying similar users and items, even in extensive datasets. This scalability and performance advantage made it the preferred choice over a traditional cosine similarity matrix.

6.3 Recommendation based on similarity

For each cold user, the system begins by identifying a card that the user has positively interacted with. This positive interaction serves as the basis for understanding the user's initial preferences (preference elicitation). Once this card is identified, the system uses the FAISS indexing mechanism to search for and retrieve the **k** most similar cards from the dataset. FAISS organizes and indexes all cards as vectors in a way that allows for quick and efficient similarity searches. By comparing the user's positively interacted card to the other cards, the system identifies those that are most alike in terms of features or content. This approach allows the system to recommend cards that are likely to align with the user's preferences, even when the user has interacted with only a limited number of items. It ensures that recommendations are relevant and personalized, making the most of the available data to provide a meaningful experience for cold users.

6.4 Recommendation Evaluation

The system's performance is evaluated by comparing the recommended items to the actual interactions of the target cold users. To accurately assess our algorithm, we simulate cold- user scenarios using normal users. Specifically, we select users who already have at least
10 positive interactions, ensuring they meet the criteria for meaningful evaluation.
However, to remain consistent with the definition of cold users, we limit the information (preference elicitation) used for these users to the bare minimum required for cold-users recommendations. For evaluation, we used standard metrics commonly used in recommendation systems, which provide a reliable benchmark for assessing the algorithm's effectiveness.

- **Precision@k:** Measures the proportion of recommended items that are relevant to the user.

- **Recall@k:** Measures the proportion of relevant items that were effectively recommended.
- **Hit Rate@k:** Indicates whether at least one relevant item was recommended.
- **MRR@k (Mean Reciprocal Rank):** Evaluates the position of the first relevant item in the recommendations, assigning greater weight to items ranked higher.
- **nDCG@k (Normalized Discounted Cumulative Gain):** Assesses the quality of recommendations by considering the position of relevant items, with higher weight for those ranked closer to the top.

The parameter k in the evaluation metrics denotes the number of items recommended to a user. Incorporating k into the evaluation allows for a less pessimistic assessment of the system's performance, as it considers an item being in the top-k recommendations for a user as equivalent to being the top-ranked recommendation. In this work, we evaluate the system using k=1, 3, 5, 10. These values are practical, as interfaces often display between 1 to 10 items on a screen.

6.5 Results and interpretation

For each user evaluated, a pool of their interactions is established (e.g., top_m = 20). The parameter m defines the size of this pool and influences the algorithm's behavior, making it either more pessimistic or optimistic, and thereby impacting the quality of the results. It is important that m is greater than or equal to k, where k represents the number of items to be recommended. The value of m limits the size of the interaction pool (i.e., the set of items the user has previously interacted with), while ensuring that recommendations are drawn from a sufficiently large subset of the pool. This alignment is critical for producing reliable results, as it guarantees that the recommendations are evaluated within the context of the user's historical interactions, which affects performance according to the evaluation metrics. We conducted an evaluation involving 1000 users to ensure a representative and diverse sample. The results of this algorithm are presented in *Table 2*:

Metric	k=1	k=3	k=5	k=10
Precision	0.48544	0.5178	0.46602	0.4901
Recall	0.02427	0.07767	0.11651	0.11651
HitRate	0.48544	1.35922	2.13592	2.13592
MRR	0.48544	0.82524	0.98058	0.98058
nDCG	0.48544	0.50225	0.4649	0.30169

Table 2 - Users Preference Elicitation Recommendation Results

The observed metric values are presented as percentages, indicating that overall performance may not appear very high at first glance. However, it is important to note that the algorithm's results significantly outperform random guessing. Given that the impressions dataset includes interactions with approximately 200,000 unique cards, the algorithm achieves results that are at least 1,000 times better than chance.

Additionally, it is worth acknowledging that achieving high metric values in cold-user scenarios is inherently challenging. Despite its simplicity and low computational complexity, this algorithm manages to achieve a HitRate of around 2%. This translates to 2–3 out of every 100 users seeing at least one item they like on their device screen—a promising outcome given the difficulty of the cold-start problem. A notable critique of this method is that, for a given user, it will consistently produce the same K recommendations. While this limitation can be mitigated through techniques such as random sampling, it is important to highlight that this algorithm is specifically designed to target cold-start users. As such, it is intended to be used sparingly, typically only a few times per user. For users with a greater number of interactions, alternative algorithms better suited to their interaction history should be employed.

6.6 Summary

Inspired by the concept of preference elicitation, this content-based recommendation algorithm leverages a single positive interaction to infer user preferences and effectively address the cold-start problem. Positive interactions, determined through a weighted scoring mechanism, allow the system to extract preferences from metadata-rich items (e.g., cards). Card metadata is preprocessed using one-hot encoding, and FAISS is employed for efficient similarity indexing and retrieval. For each user with at least one positive interaction, the algorithm identifies the k most similar cards, aiming to deliver relevant and personalized recommendations. Evaluation metrics, including **precision@k**, **recall@k**, **nDCG@k**, and **hit rate@k**, demonstrate its effectiveness, achieving a precision@5 of 0.47% and a hit rate@5 of 2.14%. While the recommendation scores may seem modest, this is consistent with the challenges inherent in the cold-start problem. Additionally, recommendations may lack variability, as the algorithm is specifically optimized for early-stage engagement, with the expectation of transitioning to more advanced methods as users accumulate richer interaction histories.

7. Conclusion and Discussion

7.1 Key Findings

This document explores how data manipulation, exploration, and machine learning can be leveraged to develop recommendation systems for the **NetEase Cloud Music Platform**, specifically targeting *cold-start users*. Cold-start users are those with low activity levels and recent interactions, making it difficult to provide accurate and meaningful recommendations due to the lack of sufficient user information. To address this challenge the **User Preference Elicitation Recommendation** algorithm was used, it provides key insights for addressing the cold-start problem in recommendation systems. First, the model employs a **custom predefined weighted scoring mechanism** to robustly identify meaningful user interactions based on metadata (e.g., clicks, likes) in the absence of explicit ratings. While this static weighting approach proved effective, it also highlights an opportunity for future research into data-driven methods, such as machine learning, to dynamically optimize these weights for improved performance. Second, the study emphasizes the **computational challenges** of large-scale similarity computations. Although traditional cosine similarity matrices are intuitive, they are prohibitively expensive for large datasets. By leveraging **FAISS** for efficient k-nearest neighbor searches, the algorithm achieved both scalability and high performance, mitigating the risk of computational exhaustion. Third, aligning the **recommendation pool size (m)** with the **evaluation set (k)** is shown to be critical for reliable testing, ensuring that the algorithm produces results that accurately reflect user preferences. Finally, the deterministic nature of the recommendations highlights a promising opportunity to introduce **diversity** or stochastic elements. Incorporating these enhancements could make the system more adaptable, ultimately improving user experience and satisfaction.

The results of the algorithm are now consolidated into a single representation for clarity and comparison (*Table 3*).

Metric	k	Users Preference Elicitation
Precision	1	0.49
	3	0.52
	5	0.47
	10	0.49
Recall	1	0.02
	3	0.08
	5	0.12
	10	0.12
HitRate	1	0.49
	3	1.36
	5	2.14
	10	2.14
MRR	1	0.49
	3	0.83
	5	0.98
	10	0.98
nDCG	1	0.49
	3	0.5
	5	0.46
	10	0.3

Table 3 - Comparison of Recommendation Models Across Evaluation Metrics

The **User Preference Elicitation** model rely heavily on user attributes and static preferences. This model is particularly advantageous in cold-start scenarios, where interaction data is scarce. This algorithm achieves a commendable level of personalization early on, however it struggles with broader recommendation lists. Its heavy dependence on static preference capturing limits its ability to dynamically adapt to broader patterns or evolving user behavior.

7.2 Limitations

The **User Preference Elicitation Recommendation Model,** while effective for cold- start users, produces repetitive user experiences by delivering the same **top-k recommendations** without variety. Its performance depends on the quality and depth of content features; when these are insufficient, the model's usefulness diminishes. Additionally, it struggles to remain relevant as users become more active, making it less sustainable for long-term engagement and less adaptable to diverse user preferences.

7.3 Research Contributions

Despite the limitations discussed earlier, this study makes significant contributions to the field of personalized recommendation systems, specifically in tackling the cold-start user problem on the NetEase Cloud Music platform. A major contribution of this study is the demonstration that integrating diverse data sources offers a holistic approach to understanding user preferences. The tested model showcased how combining multiple data layers—such as user behavior, metadata, and side information—can produce nuanced user 25 profiles. These profiles serve as the foundation for generating personalized recommendations, particularly in cold-start environments. Complementing this, the study highlights the effective use of metadata fields to tailor recommendations and evaluate initial recommendation sets. This research also introduces a rigorous set of evaluation metrics to systematically assess model performance. These carefully selected metrics not only highlight each algorithm's strengths and weaknesses but also provide deeper insights into their real-world applicability. By establishing an evaluation framework, this study paves the way for scalable solutions, offering a standardized methodology that remains insightful regardless of the computational complexity or sophistication of the model. Beyond academic contributions, this work delivers a practical framework for digital platforms like NetEase Cloud Music to improve user retention and engagement. By specifically addressing the needs of cold-start users, the methods developed in this study aim to foster long-term user satisfaction, which is critical in maintaining a competitive advantage in the fast-paced digital music industry.

7.4 Implications for Practice

A new user joins NetEase Cloud Music, hungry to know more yet not knowing how to begin. They open the app, and, instantaneously, the platform must ask itself: How do we make this moment matter? Without a listening history, the system relies on what it knows—demographic data, registration context, and perhaps a hint of regional trends. Referred to as the cold start problem, it is a problem faced by every music streaming service. But it is also an opportunity to **build trust, encourage user participation,** and create the foundation for a long-term relationship. Considering this particular situation as an increasingly common scenario in any digital, online platform, this research seeks to contribute to both academic knowledge and practical industry solutions by advancing the understanding of recommender systems. With a particular focus on addressing the cold- start problem, the conducted study approaches a **critical challenge** that hinders effective personalization in both regional and global markets. From a practical perspective, the findings of this study aim to provide actionable insights for enhancing user experiences on NetEase's digital music platform, fostering greater user engagement, and boosting retention in the competitive and volatile digital media landscape. It is not a merely technical problem, but it is very crucial when it comes to starting the user's journey on a platform like NetEase Cloud Music. With some demographics involved, such as age and geographic and cultural context, the platform can perhaps make recommendations that seem to reside within easy reach in a very intuitive way. Possibly, an

18-year-old who lives in a teeming city might be pulled into the trendiest pop hits, while someone who is older and comes from a quieter environment might appreciate instrumental or folk tracks. These would give someone that personal welcome—a very impersonal content library becomes a more curated invitation into exploration. The implications of solving the cold-start problem extend far beyond that first impression. It should incite exploratory activity and discovery. A listener on pop should be interested into a more nearby genre, such as indie or electronica, which would broaden horizons and create a deeper experience. Meanwhile, based on demographic trends, new content—emerging artists or new releases without much interaction data—could find the audience compellingly through publicity in an already congested crowd. This serves both users and creators, thereby strengthening the platform's ecosystem.

Cold-start solutions are also important for returning users because their preferences might have changed or evolved. The same strategy can be employed on returning users as on new users, to spark that interest at the beginning and ensure that submissions on the platform remain relevant and consistently in touch with time. For the music streaming industry, the cold-start solution is a step toward making the whole system understood by users and embracing all creators with equal opportunities to be discovered. This cold-start problem is not solely about filling the normatively structured void in the data; **it is also about shaping the user's experience from that initial interaction**. For NetEase Cloud Music and similar platforms, solving this challenge transforms the act of streaming into a **journey of connection and discovery**, one that builds loyalty and fosters a deeper love for music.

7.5 Future Research Directions

Addressing the need for personalized user journeys requires exploring data strategies, model development, and privacy concerns. Existing data lacks critical components like activity levels and creator types, limiting analysis and model efficiency. Song-related data should include attributes like genre or energy to enhance clustering and embeddings for cold-start solutions. The dataset's one-month span restricts temporal insights, but broader data could improve recommendations by revealing long-term user patterns. Transitioning to online recommendation systems could ensure real-time updates and increased user satisfaction. The User Preference Elicitation method's reliance on static data highlights the need for machine learning techniques like dynamic metadata weighting for better personalization. Larger datasets and computational resources are essential for tackling cold-start challenges. Data privacy remains critical, requiring protection against threats like adversarial attacks to maintain user trust. This research lays the groundwork for solving the cold-start problem and advancing beyond it. Future directions will involve smarter algorithms, secure and ethical data management, and a vision for inclusive, user-centric platforms. For NetEase Cloud Music and the broader music streaming industry, this represents not just a technical challenge but an opportunity to redefine how music connects users in the digital age. **In an industry where the sheer abundance of content can overwhelm users, recommendation systems serve as a critical and indispensable differentiator.**

8. References

s.d. “AI in Music Streaming: Personalization at its Best.” *AI in Music Streaming: Personalization at its Best.*

s.d. “AI-based Recommendation System Market Report.” *AI-based Recommendation System Market Report.*

Bobadilla, JesúS, Fernando Ortega, Antonio Hernando, e Jesús Bernal. 2012. “A collaborative filtering approach to mitigate the new user cold start problem.” *Knowledge-based systems* (Elsevier) 26: 225–238.

Chadha, Aman, e Vinija Jain. 2022. “Recommender Systems Primer - Cold Start.” *Recommender Systems Primer - Cold Start.* www.vinija.ai.

Chakravorti, Bhaskar. 2018. “Competing in the Huge Digital Economies of China and India.”

Harvard Business Review 6.

Douze, Matthijs, Alexandr Guzhva, Chengqi Deng, Jeff Johnson, Gergely Szilvasy, Pierre-Emmanuel Mazaré, Maria Lomeli, Lucas Hosseini, e Hervé Jégou. 2024. “The Faiss library.”

Elahi, Mehdi, Farshad Bakhshandegan Moghaddam, Reza Hosseini, Mohammad Hossein Rimaz, Nabil El Ioini, Marko Tkalcic, Christoph Trattner, e Tammam Tillo. 2021. “Recommending videos in cold start with automatic visual tags.” *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization.* 54–60.

Feng, Philip J., Pingjun Pan, Tingting Zhou, Hongxiang Chen, e Chuanjiang Luo. 2021. “Zero shot on the cold-start problem: Model-agnostic interest learning for recommender systems.” *Proceedings of the 30th ACM international conference on information & knowledge management.* 474–483.

Hafnar, Davor, e Jure Demšar. 2024. “Zero-Shot Reasoning: Personalized Content Generation

Without the Cold Start Problem.” *arXiv preprint arXiv:2402.10133.*

Han, Cuize, Pablo Castells, Parth Gupta, Xu Xu, e Vamsi Salaka. 2022. “Addressing cold start in product search via empirical bayes.” *Proceedings of the 31st ACM International Conference on Information & Knowledge Management.* 3141–3151.

s.d. “How Retailers Can Keep Up with Consumers.” *How Retailers Can Keep Up with Consumers.*

Karabila, Ikram, Nossayba Darraz, Anas El-Ansari, Nabil Alami, e Mostafa El Mallahi. 2024. “Addressing data sparsity and cold-start challenges in recommender systems using advanced deep learning and self-supervised learning techniques.” *Journal of Experimental & Theoretical Artificial Intelligence* (Taylor & Francis) 1–31.

Kowald, Dominik, Markus Schedl, e Elisabeth Lex. 2020. “The unfairness of popularity bias in music recommendation: A reproducibility study.” *Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part II 42.* 35–42.

Kumar, Pushpendra, e Ramjeevan Singh Thakur. 2018. “Recommendation system techniques and related issues: a survey.” *International Journal of Information Technology* (Springer) 10: 495–501.

Lam, Xuan Nhat, Thuc Vu, Trong Duc Le, e Anh Duc Duong. 2008. “Addressing cold-start problem in recommendation systems.” *Proceedings of the 2nd international conference on Ubiquitous information management and communication.* 208–211.

Li, Jingjing, Ke Lu, Zi Huang, e Heng Tao Shen. 2017. “Two birds one stone: on both cold-start and long-tail recommendation.” *Proceedings of the 25th ACM international conference on Multimedia.* 898–906.

Lika, Blerina, Kostas Kolomvatsos, e Stathes Hadjiefthymiades. 2014. “Facing the cold start problem in recommender systems.” *Expert systems with applications* (Elsevier) 41: 2065–2073.

Lin, Xixun, Jia Wu, Chuan Zhou, Shirui Pan, Yanan Cao, e Bin Wang. 2021. “Task-adaptive neural process for user cold-start recommendation.” *Proceedings of the Web Conference 2021.* 1306–1316.

Liu, Haifeng, Zheng Hu, Ahmad Mian, Hui Tian, e Xuzhen Zhu. 2014. “A new user similarity model to improve the accuracy of collaborative filtering.” *Knowledge-based systems* (Elsevier) 56: 156–166.

Melville, Prem, e Vikas Sindhwani. 2010. “Recommender systems.” *Encyclopedia of machine learning* 1: 829–838.

s.d. “Music Streaming - China | Statista Market Forecast.” *Music Streaming - China | Statista Market Forecast.*

s.d. “Music Streaming - Worldwide | Statista Market Forecast.” *Music Streaming - Worldwide | Statista Market Forecast.*

Nguyen, An-Te, Nathalie Denos, e Catherine Berrut. 2007. “Improving new user recommendations with rule-based induction on cold user data.” *Proceedings of the 2007 ACM conference on Recommender systems.* 121–128.

Pandey, Anand Kishor, e Dharmveer Singh Rajpoot. 2016. “Resolving Cold Start problem in recommendation system using demographic approach.” *2016 International Conference* on Signal Processing and Communication (ICSC). 213-218. doi:10.1109/ICSPCom.2016.7980578.

Panteli, Antiopi, e Basilis Boutsinas. 2023. “Addressing the cold-start problem in recommender systems based on frequent patterns.” *Algorithms* (MDPI) 16: 182.

Pu, Pearl, e Li Chen. 2008. “User-involved preference elicitation for product search and recommender systems.” *AI magazine* 29: 93–93.

Roy, Deepjyoti, e Mala Dutta. 2022. “A systematic review and research perspective on recommender systems.” *Journal of Big Data* (Springer) 9: 59.

Silva, Nícollas, Diego Carvalho, Adriano C. M. Pereira, Fernando Mourão, e Leonardo Rocha. 2019. “The pure cold-start problem: A deep study about how to conquer first-time users in recommendations domains.” *Information Systems* (Elsevier) 80: 1–12.

Singh, Neetu, e Sandeep Kumar Singh. 2024. “A systematic literature review of solutions for cold start problem.” *International Journal of System Assurance Engineering and Management* (Springer) 1–35.

Son, Le Hoang. 2016. “Dealing with the new user cold-start problem in recommender systems: A comparative review.” *Information Systems* (PERGAMON-ELSEVIER SCIENCE LTD THE BOULEVARD, LANGFORD LANE, KIDLINGTON …) 58: 87–104.

Song, Yading, Simon Dixon, e Marcus Pearce. 2012. “A survey of music recommendation systems and future perspectives.” *9th international symposium on computer music modeling and retrieval.* 395–410.

Steck, Harald, Linas Baltrunas, Ehtsham Elahi, Dawen Liang, Yves Raimond, e Justin Basilico. 2021. “Deep learning for recommender systems: A Netflix case study.” *AI Magazine* 42: 7–18.

Sun, Anchen, e Yuanzhe Peng. 2022. “A survey on modern recommendation system based on big data.” *arXiv e-prints* arXiv–2206.

Tahmasebi, Faryad, Majid Meghdadi, Sajad Ahmadian, e Khashayar Valiallahi. 2021. “A hybrid recommendation system based on profile expansion technique to alleviate cold start problem.” *Multimedia Tools and Applications* (Springer) 80: 2339–2354.

Vozalis, Emmanouil, e Konstantinos G. Margaritis. 2003. “Analysis of recommender systems algorithms.” *The 6th Hellenic European Conference on Computer Mathematics & its Applications.* 732–745.

Wang, Chunyang, Yanmin Zhu, Haobing Liu, Tianzi Zang, Jiadi Yu, e Feilong Tang. 2022. “Deep meta-learning in recommendation systems: A survey.” *arXiv preprint arXiv:2206.04415.*

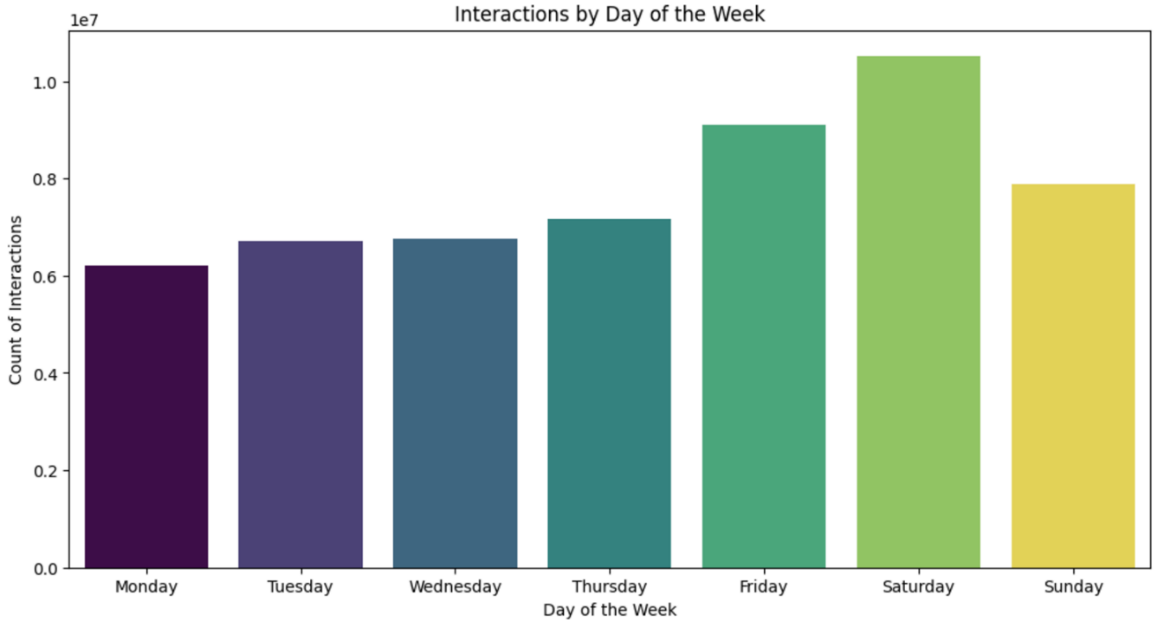
Wang, Shixiong, Wei Dai, e Geoffrey Ye Li. 2024. “Distributionally Robust Outlier-Aware Receive Beamforming.” *2024 IEEE 34th International Workshop on Machine Learning for Signal Processing (MLSP).* 1–6.

s.d. “What Netflix's Recommendation Systems Can Teach Us About The Computing Challenges Of The Near Future.” *What Netflix's Recommendation Systems Can Teach Us About The Computing Challenges Of The Near Future.* https://www.forbes.com/councils/forbestechcouncil/2021/02/19/what-netflixs- recommendation-systems-can-teach-us-about-the-computing-challenges-of-the-near- future/.

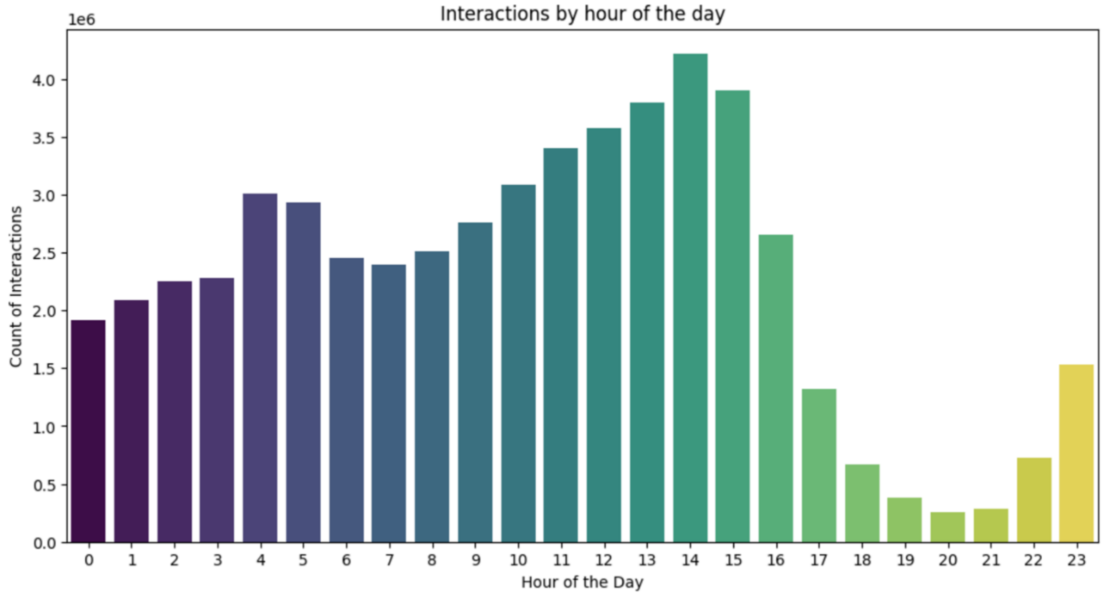
Zhang, Dennis J., Ming Hu, Xiaofei Liu, Yuxiang Wu, e Yong Li. 2022. “NetEase cloud music data.” *Manufacturing & Service Operations Management* (INFORMS) 24: 275–284.

Zhu, Ziwei, Shahin Sefati, Parsa Saadatpanah, e James Caverlee. 2020. “Recommendation for new users and new items via randomized training and mixture-of-experts transformation.” *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval.* 1121–1130.

9. Appendix



Appendix 1 - Interactions by day of week



Appendix 2 - Interactions by hour of the day