



HAL
open science

The Impact of Algorithmic and Human Recommendations on Platform User Satisfaction

François Moreau, Patrik Wikström

► To cite this version:

François Moreau, Patrik Wikström. The Impact of Algorithmic and Human Recommendations on Platform User Satisfaction. 2024. <hal-04553197>

HAL Id: hal-04553197

<https://hal.science/hal-04553197v1>

Preprint submitted on 19 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

The Impact of Algorithmic and Human Recommendations on Platform User Satisfaction

François MOREAU¹ & Patrik WIKSTRÖM^{2,3}

¹University Sorbonne Paris Nord, France
CEPN (UMR 7234) & LabEx ICCA
<https://orcid.org/0000-0002-5818-5680>

²Queensland University of Technology, Australia
QUT Digital Media Research Centre,
ARC Centre of Excellence for Automated Decision-Making and Society

³Inland Norway University of Applied Sciences
<https://orcid.org/0000-0003-4720-0416>

Abstract. *Music streaming platforms use two main mechanisms for music curation: One mechanism is the human-curated playlist – a sequence of songs focused on a specific theme. The second mechanism relies on a so-called recommender system, which automatically suggests what a user should listen to next based on the user’s past behavior on the platform. While platforms invest considerable resources to develop their recommender systems to maximize consumer satisfaction and loyalty, there is limited research on the actual impact of these systems on consumer satisfaction. This paper addresses this gap and assesses the impact of algorithmic and human music recommendations on listener satisfaction. Our approach builds on an inherent feature of music listening experiences – that people tend to listen to songs they like multiple times. Our dataset is user-level data of 9,778 random premium subscribers of a leading European music streaming platform who discovered 4,136 unique new songs during three months, yielding a dataset of 1,952,195 song discoveries. Our findings show that on-platform curation is less effective than off-platform information in suggesting satisfying content to users and that human-curated playlists are, on average, more effective than algorithmic recommendations. Further, we find that on-platform curation is more effective than off-platform information in suggesting relevant, unfamiliar content, especially content from foreign low-popularity artists which the users are unfamiliar with. Finally, within the on-platform curation discovery modes, human-curated playlists are always more effective than algorithmic recommendations in suggesting relevant, albeit more familiar, content. Conversely, algorithmic recommendations do better than human-curated playlists when suggesting relevant unfamiliar content.*

Keywords. Recommender systems; Human curation; User satisfaction; Music streaming

1 Introduction

Human-curated music playlists and algorithmic recommendations are crucial mechanisms for making the music streaming platforms' enormous song catalogs manageable for listeners. Without effective recommendations that direct users to lesser-known songs that fit their music preferences, most songs on the platform catalog would never be found or played.

These music recommendation mechanisms are pivotal parts of the platforms' user experience designs, ensuring music listeners remain happy, satisfied, and loyal platform subscribers. However, while platforms invest considerable resources to develop their recommender systems to maximize consumer satisfaction and loyalty, limited research exists on the actual impact of these systems on consumer satisfaction.

It is well-established that there is a salience effect associated with recommendations, which means that the very fact that an item is recommended serves as a signal to the consumer and impacts their purchase intention (Lee & Hosanagar, 2021). Adomavičius et al. (2018) show that consumers' willingness to pay is positively impacted by recommendations, even when recommendations are artificially biased to be irrelevant to the consumer. Adomavičius et al. (2014) also provide evidence that consumer preference ratings are malleable and can be significantly influenced by the recommendation. A key conclusion from these studies is that just because a user responds positively to human or algorithmic recommendations – for instance, by watching the recommended movie or purchasing the recommended product – does not systematically mean that the recommended product is a good fit with consumers' preferences.

In this paper, we control for this salience effect and study whether recommendations lead to user satisfaction. We exploit an inherent feature of recorded music consumption, which provides us with an innovative method to assess user preferences. The fundamental feature we are utilizing is that, unlike most products, when music listeners like a song, they often listen to it repeatedly (Schedl et al., 2018; Sguerra et al., 2022). This feature separates music from most other products, specifically from

entertainment products, where the utility commonly peaks quickly and then declines with the number of plays (Kahn et al., 1997; Hennig-Turau & Houston, 2019).

Our paper reports on the analysis of a unique dataset from one of France's leading music streaming services. The dataset shows how 9,778 randomly selected users discover newly released songs (through an algorithmic recommendation, a platform playlist, or outside the platform) and how many times they actively choose to relisten to these songs in the 24 months following the discovery.

This unique dataset allows us to meet two conditions that are necessary to empirically test the impact of human and algorithmic recommendations on consumer satisfaction. First, our data contrasts users exposed to platform recommendations with those unexposed to platform recommendations (Lee & Hosanagar, 2019). Second, our dataset is *user-level* data, where each case represents a single stream of a specific song by a specific user. This data granularity is unusual, as most music streaming research data available so far is at the *song level*, which means that each case represents the number of streams of a song during some time period – without any specific information about the individual streams or the individual users engaged in the listening. User-level data is crucial for the study we are reporting on in this paper, and it is recognized as the “next step for music research” to address research questions with greater ecological validity (e.g., Greenberg & Rentfrow, 2017).

With aggregate data alone, there is no conclusive evidence of what individual users actually do. However, user-level data allows us to effectively disentangle both individual music consumption from recommender system rankings (Aguiar & Waldfogel, 2021) as well as repeat consumption of a song from the number of unique listeners of that song (Aguiar & Waldfogel, 2018; Chen et al., 2016; Datta et al., 2018). The disentanglement of these phenomena is essential to properly studying recommendations' impact on user satisfaction (Sim et al., 2022).

Indeed, some studies using aggregate data have attempted to address questions of discovery and utility - for instance, Sim et al. (2022) can separate discovery from repeat consumption in their analysis. However, the authors measure repeat consumption at an aggregate level only and not at the individual level, making it impossible to assess the satisfaction the recommendation brings to the individual user.

With our novel approach and unique dataset, we can overcome these challenges and provide a path forward for assessing user satisfaction from algorithmic and human recommendations.

Our findings show that on-platform curation is less effective than off-platform information in suggesting satisfying content to users and that human-curated playlists are, on average, more effective than algorithmic recommendations. Further, we find that on-platform curation is, in relative terms, more effective than off-platform information in suggesting relevant, *unfamiliar* content, especially content from low-popularity artists and foreign origins. Finally, within the on-platform curation discovery modes, human-curated playlists are always more effective than algorithmic recommendations in suggesting relevant, albeit more familiar, content. Conversely, algorithmic recommendations do better than human-curated playlists (or less bad) when suggesting relevant unfamiliar content.

In the next section, the paper provides a background to the curation of musical experiences, music discovery, and the assessment of music recommendation system effectiveness. The paper then continues with a presentation of a literature review and a theoretical framework, followed by a section on our empirical strategy. We then present the results from the study and argue the robustness of our results. Lastly, we discuss our findings and make conclusions and recommendations for future research.

2 Background

2.1 Human and algorithmic curation of music experiences

When enjoying recorded music, people are rarely satisfied by listening to a single song. Instead, a music listening session often lasts for hours and consists of a *sequence* of songs. The curation of this sequence is, to some extent, controlled by the artist when they organize the songs on an album or plan the set list for a live show. However, the sequence of songs can also be curated by other actors, such as a DJ in a club, a radio station programmer, a friend who is recording a mixtape as a gift, or the music listeners themselves.

The introduction of music streaming platforms (e.g., Spotify, Apple Music, or Deezer), which make millions of songs instantly available to their users, has dramatically changed the rules for curating personalized musical experiences. Music streaming platforms use two main curatorial mechanisms: One is the *playlist*, which (as one might assume) is a sequence of songs focused on a specific theme, such as a genre, period, mood, activity, etc. Streaming platforms maintain hundreds of playlists. Most of these lists are curated by a human music expert, who maintains and updates a number of lists within their area of expertise. About 8 percent of all streams in our dataset have been initiated as part of a playlist officially curated by the streaming platform, and another 2 percent by other major stakeholders (mainly major labels). Table 1 shows the ten playlists that reached the highest share of users in our sample. All the playlists in this table are curated in-house by the streaming platform. 82% of all users in the sample have listened at least once to the playlist with the highest reach – “*Les titres du moment*” and 1.5% of all the streams users have listened to have been initiated from one of the top three playlists. The table shows that the most popular playlists are all human-curated, except for “*Top France*,” an automated list of the most streamed songs in France during the previous day. The table also shows that the most popular playlists are either current hits, a selection of new releases, or songs to make the listener relax and feel good. One exception is *Les Hits de Noël*, which is the only playlist in the top ten focused on a specific event.¹

Table 1 - Top ten playlists with the highest reach among users in our sample

Title	User reach	Stream share	Title/description (translated)
Les titres du moment	82%	0.71%	“Sounds of the moment”
Deezer Hits	62%	0.26%	“Essential hits of the year”
Top France	54%	0.51%	Daily chart - The most streamed songs in France
Les nouveautés du vendredi	51%	0.13%	“New releases every Friday”
Chill relax	51%	0.17%	“The ideal playlist to relax”
Poptop	50%	0.10%	“The best new pop songs”
Feel Good	47%	0.12%	“A selection to put you in a good mood”
Bleu blanc hits	43%	0.15%	“The latest French hits”
Soirée au top	39%	0.06%	“No ideas for tonight? We have the playlist you need.”
Les hits de Noël	38%	0.06%	“Christmas Classics”

¹ Another kind of playlist is the user-generated playlist, where music listeners save a collection of songs of a certain type; for some kind of event; or simply the songs they currently like. About 41 percent of all streams in our dataset have been initiated as part of a user-generated playlist.

The second mechanism is the platform's recommender system (Resnick & Varian, 1997).

Recommender systems generate personalized, unprompted recommendations based on a user's past behaviors and characteristics. They answer the "what's next" question for a user – such as the next book to read, the next movie to watch, or the next song to play. The context for this paper is music streaming platforms. The major music streaming platforms offer song catalogs with more than 100 million songs, adding approximately 60,000 new songs per day. It is obviously impossible for users to sift through all these songs without help from the platform. Algorithmic recommendations can be incorporated into the user experience in several different ways. For instance, the platform can enable users to launch an algorithmically curated sequence of songs based on a specific song or an artist as a seed, e.g., "Lady Gaga Radio." The platform can also use its recommender system to curate a personalized playlist based on the user's listening practices during the past week. Recommender systems combine two basic principles when making a recommendation: One principle, conventionally referred to as collaborative filtering (Goldberg et al., 1992), makes the recommendation based on patterns of listener practices across a section of the user population. For example, suppose users who frequently listen to Lady Gaga are prone to also listen to Beyoncé. In that case, it makes sense to recommend Beyoncé to other Lady Gaga fans who have not yet discovered Beyoncé. The second principle, the so-called "content-based" recommender system, is based on an individual listener's engagement with diverse types of musical content. For example, if a user frequently listens to 1960s jazz during the past week, it is reasonable to recommend more jazz from the 1960s to this user during the week ahead. Recommender systems are far more sophisticated than these two examples suggest, and today, they are used by most digital platforms to recommend to users what to watch, hear, read, or buy next. About 16 percent of all streams in our dataset have been initiated by the platform's recommendation system.

2.2 Streaming platforms and music discovery

The question of how music listeners discover new songs and artists is firmly based at the heart of the music industrial logic. Traditionally, songs and artists are introduced to new listeners by broadcast

radio or television, by music journalists' reviews, or by listeners' friends and families. However, these structures have been overturned since the introduction of music streaming platforms. Studies from France and the US show (see Table 2 and Table 3) that music listeners state music streaming platforms as the most common place for music discovery, although broadcast media, friends, and family remain essential.

Table 2 – Sources of musical discoveries in France (16-64 years old)

Source	% individuals who declare using it
Audio and/or video streaming	48.0
Radio	47.0
Word-of-Mouth	34.0
Television	27.0
Social Media (Instagram, Facebook, Twitter, etc.)	19.0

Source: IFPI Consumer Study 2021

Table 3 – Sources of musical discoveries in the USA (18+ years old)

Source	% individuals who declare using it
Music apps (Spotify, Apple Music, etc.)	35.0
Radio	35.0
Word-of-Mouth	31.0
Movies/Television shows	33.0
Social Media	30.0

Source: YouGov, MarketingCharts.com

2.3 Assessing the effectiveness of recommender systems

A music recommender system aims to predict a listener's music preferences. When evaluating the effectiveness of recommender systems, the *accuracy* of those predictions has traditionally been the measure to gauge the recommender system's performance. However, an accurate recommender system has significant issues, primarily as it tends to confine listeners to a music experience that is gradually becoming less diverse and less serendipitous. Users who enjoy listening to new songs and artists are unlikely to be satisfied with such a limited experience. To ensure that these users do not leave the platform, platform operators try to balance the accuracy measure with other measures, primarily novelty, serendipity, and diversity (e.g., Adamopoulos & Tuzhilin 2014). Balancing these ideals is difficult, however, as recommendations of less-popular songs intended to increase the diversity of a user's music listening experience run the risk of simply being too far away from the users'

preferences. The optimal balance, which maximizes listener satisfaction, is not easy to find – particularly in a real-world production environment.

This paper addresses this challenge and uses a novel approach to assess the impact of algorithmic and human music recommendations on listener satisfaction. Our approach builds on an inherent feature of music listening experiences – that people tend to listen to songs they like multiple times. In the dataset we use in this paper, more than 40 percent of the songs streamed at least once after their discovery have been streamed six times or more by the same user over a 24-month period, and more than 20 percent have been streamed more than twenty times. We use repeat listening of a song as an indicator of how much a user likes the specific song. Combined with an examination of the different recommendation mechanisms that lead users to music discovery, we can determine how algorithmic and human recommendations impact user satisfaction.

3 Literature review

We structure the literature that is relevant to our paper into two streams. One stream deals with how curation impacts the market outcomes (e.g., sales, discovery of new products, diversity), and a second stream consists of studies assessing users’ consumption preferences.

3.1 The impact of curation on market outcomes

The literature focused on the economic outcomes of curation deals with questions such as how curation impacts sales, the discovery of new products, consumption diversity, or how curation affects search vs. experience products. This literature states that recommender systems can increase consumers’ buying propensity and sales (De et al., 2010; Pathak et al., 2010; Adomavicius et al., 2018; Lee & Hosanagar, 2019). This salience effect observed for algorithmic recommendations is also evident for human-curated playlists. For example, Aguiar & Waldfogel (2021) and Knox & Datta (2022) have shown that songs included on such a human-curated playlist benefit from a substantial increase in the stream count. There is also literature on the impact of recommender systems on the discovery of new products. For instance, Datta et al. (2018) show that adopting a music streaming

platform favors consumers' discovery of new products. Likewise, Sim et al. (2022) identify a positive impact of reaching the top 100 charts as a discovery-based consumption at the individual level.

The impact of curation on consumption diversity, however, is less clear. Curation is supposed to help consumers discover products that fit their personal consumption preferences (Dzyabura & Hauser, 2019). Based on this proposition, there is an argument that a recommender system will differentiate the consumption patterns of individual users from “mainstream preferences” and from other users, thereby contributing to an *increase* in market-level diversity and the “long tail” phenomenon in online retailing, as discussed by e.g., Anderson (2006), Brynjolfsson et al. (2011), Oestreicher-Singer & Sundararajan (2012), and Zhou et al. (2010). Others, however, find the opposite. For instance, Fleder & Hosanagar (2009) suggest that recommender systems can lead to a rich-get-richer effect for popular products, decreasing market-level sales diversity. A study of recommender systems on the Spotify platform focuses on consumption *genre* diversity and concludes that “algorithmically driven listening through recommendations is associated with reduced consumption diversity.” (Anderson et al., 2020) Lee and Hosanagar (2019), as well as Holtz et al. (2020), show that recommender systems can increase individual diversity while reducing aggregate diversity.²

Studies focused on human-curated music playlists are also in disagreement. Aguiar et al. (2021), studying playlists on Spotify, show that playlists favor independent labels. Wlömert et al. (2021) agree and conclude that playlists benefit relatively unknown artists more as playlists cater to niches that have previously been neglected because of high search costs. On the other hand, Mariuzzo & Ormosi

² The lack of consensus between studies on the economic impact of recommender systems can be explained by the fact that different studies focus on different recommender systems, which in turn follow different algorithmic logics (e.g., Hosanagar et al., 2014; Lee & Hosanagar, 2019). For example, studies of content-based recommenders conclude that this type of recommender system increases individual-level consumption diversity (e.g., Wu et al., 2011). Lee & Hosanagar (2019), on the other hand, analyze the use of traditional collaborative filters across a wide range of product categories and find that this type of recommender system is associated with a *decrease in market-level sales diversity* relative to a world without product recommendations. The authors' explanation of this finding is that recommender systems are effective in making individual consumers discover products that are unknown to them but popular at a global level rather than aiding the discovery of niche products.

(2020) argue that major labels benefit the most and have disproportionate access to Spotify-generated playlists.

There are no clear results either on the impact of recommender systems on product attributes such as experience vs. search goods. We can imagine that the impact of recommender systems on consumer choices should be more significant with experience goods than search goods because, with the former, the product's fit with a consumer's preferences is uncertain. However, Dai et al. (2020) claim that consumers rely less on other consumers' reviews when shopping for experience products. They show that consumers of experience products believe other consumers' reviews are incompatible with their own assessments. In laboratory experiments with two products (one search and one experience), Senecal & Nantel (2004) and Aggarwal & Vaidyanathan (2005) find opposing results. More recently, using field data and a wide variety of products, Lee & Hosanagar (2021) found no statistically significant difference between experience and search products. According to them, although recommender systems are considered "digitized word-of-mouth" (Chen et al., 2009), simple signals such as "other consumers who have purchased this item also purchased..." do not provide enough details to be effective for experience products. Moreover, it has been widely confirmed that online reviews can be unreliable (among many, Mayzlin et al., 2014; He et al., 2022).

One of the possible explanations for the contradictory results obtained when studying the impact of recommender systems lies in how this impact is measured. As previously discussed, the effectiveness of a recommender system is traditionally assessed based on its "accuracy," i.e., the correspondence between the recommender system's prediction and the consumer's actual preferences (e.g., using consumer product ratings as an indicator of actual preferences). However, the measures becoming increasingly crucial for recommender system design – diversity, novelty, or serendipity – are much more complex to assess quantitatively than accuracy. Moreover, limiting the evaluation to quantitative measures means forgoing another vital factor: the actual user experience (Schedl et al., 2018). To address this challenge, it is necessary to generate user-level evidence (as we do in this study) to test if recommender systems give users a useful personalized experience (Knijnenburg et al., 2012). Methods and approaches for obtaining and assessing such user-level evidence are discussed in the next section.

3.2 Assessing users' preferences

User preference data can be obtained using explicit or implicit methods. Explicit (*stated preference*) methods include user studies (Barrington et al., 2009) where users directly provide feedback on recommended items, e.g., offline or online questionnaires. Implicit (*revealed preference*) methods assess user preferences by monitoring their behavior, e.g., purchase transactions or clickstream data (Iwanaga et al., 2019), and assuming that user behavioral patterns reveal their true consumption preferences.

Relying on explicit strategies is still the predominant way of evaluating recommender systems in academia, not least due to the lack of access to (large numbers of) real users (Schedl et al., 2018). However, the strategy to infer the quality of recommender systems from answers to questions after an experiment or a session of actual consumption is far from perfect. The strategy simply does not provide enough information on the perceived quality of recommendations and their actual usefulness for the user. High recommender accuracy does not always correlate with user satisfaction. When it comes to discovery, the fit with users' preferences requires direct feedback from the user about a specific recommendation. Finally, such user studies hardly gather enough participants to draw significant and usable conclusions (Schedl et al., 2018). Loepp et al. (2019) show that if the experiment design does not allow users to test the recommended product, the reliability of the answers to the questionnaire is strongly negatively impacted, which qualifies the usefulness of such approaches.

Some studies using explicit (stated preference) methods address very similar questions as we do in our paper. These studies are specifically interested in the impact of recommender systems on user satisfaction and rely on user data from a music streaming platform (e.g., Garcia-Gathright et al. 2018; Mehrotra et al. 2018; 2019). The overall conclusion from these studies is that the relevant interaction signals between users and recommender systems, which are used to measure user satisfaction, strongly depend on the users' listening intent (e.g., passive vs active listening). However, while these results are interesting, these studies ultimately have the same limitations and caveats as other studies using explicit, stated preference, methods. Mehrotra et al. (2019) note the limitations of this approach,

particularly in predicting users' satisfaction with recommender systems to discover new music for immediate listening. The studies are still relevant to our study since they are also relying on user data from music streaming platforms. Other studies that focus on industries other than music streaming are also relevant to our work since they compare algorithmic and human recommendations. For instance, Senecal & Nantel (2004) and Krishnan et al. (2008), found that algorithms outperform humans, while Amatrian et al. (2009) and Sinha & Swearingen (2011) found the opposite. However, they all rely on small samples and on explicit, stated preference methods.

While implicit (revealed preference) methods are less burdened by the limitations discussed above, the main deficit with these methods is that they use data from one-shot purchase transactions, which are imperfect in predicting satisfaction (especially for experience goods). Adomavičius et al. (2018) show that when purchasing digital songs, consumers' willingness to pay is positively impacted by recommendations, even when the researchers add a bias to render the recommendations irrelevant. This salience effect of recommender systems seems strong enough to make users unsure whether the recommended items they purchase are something they genuinely like. Likewise, Sim et al. (2022) find that for songs included in playlists, the increase in streaming consumption is more consistent with a salience effect of being included in the list of most popular songs than with observational learning. Cowgill (2020), Yeomans et al. (2019), and Peukert et al. (2023) use implicit methods to compare algorithmic and human recommendations, and the conclusions from their studies are mixed. The two first studies find that algorithmic recommendations outperform human recommendations. Peukert et al. (2023) however, arrive at a more nuanced conclusion. By using the context of online news and a field experiment design, they show that algorithmic recommendations do better than human curation when enough personal data on users has been previously gathered. Conversely, human curation can outperform algorithms when the human curator is experienced enough and when there is strong and sudden variation in users' preferences.

The Peukert et al. (2023) study is highly relevant to our work and we extend their findings in two ways. First, we compare algorithmic and human on-platform recommendations, but we also analyze *off*-platform recommendations. Second, Peukert et al. (2023) consider that users are satisfied with a

recommendation when they click on the suggested article, which ignores the salience effect as discussed above (Lee & Hosanagar, 2021; Adomavičius et al., 2018; Adomavičius et al., 2014). The design of our study addresses this deficiency and offers a robust and valid indicator of user satisfaction.

The inherent nature of music consumption allows us to bypass these limitations and implement an innovative implicit method to assess user preferences. The essential feature we are utilizing in this study is that unlike most entertainment products (movies, books, series), when music listeners like a song, they often listen to it repeatedly. The utility of most entertainment product usage peaks quickly and then rapidly declines with the number of usages (Kahn et al., 1997; Hennig-Turau & Houston, 2019). However, this is less true for songs than for movies or books (Schedl et al., 2018). While repeat consumption of the same song also follows an inverted-U shape function, the peak is reached much later for music than for books or movies, usually after ten or more exposures to new songs (Sguerra et al., 2022).

A song appearing *once* in the user's streaming history does not necessarily mean the user actively likes the song, and a song that has been skipped does not mean that they do not (Schedl et al., 2018).

However, while this may be the case, it is more probable that *the number of repeated streams* is a strong indicator of whether the user likes the song. This is because we can think of the first stream as an action based on predicted satisfaction, while repeated streams of the same song signal actual ex-post satisfaction. Repeated consumption is already an established satisfaction indicator in the marketing and psychology literature (e.g., La Barbera & Mazursky, 1983; Bhattacharjee, 2001; Liu et al., 2011; Rauyruen et al., 2007; Zeithaml et al., 1996; Paul et al., 2009). In the music industry, Fricke et al. (2019) or Garcia-Gathright et al. (2018) also use playback counts as a measure of music preferences.

Jannach and Abdollahpouri (2023) stress that the long-term effect of different recommendation strategies is insufficiently explored. We answer to their call and contribute to the literature discussed above in several ways. First, we study whether curation at the individual level and not only the economic outcomes of curation at the aggregate level (e.g., Aguiar & Waldfogel, 2021; Wlömert et al.,

2021). Second, we add to previous studies using individual-level data from a streaming platform, which either do not focus on curation at all (e.g., Way et al., 2020; Datta et al., 2020; Knox & Datta, 2022), or do not analyze the impact of curation on user satisfaction (e.g., Villermet et al., 2021). Third, we add to previous studies that focus on algorithmic recommendations but do not make comparisons with other curation strategies, such as human on-platform recommendations or off-platform recommendations (e.g., Garcia-Gathright et al., 2018; Mehrotra et al., 2018, 2019). Fourth, we add to previous studies that compare algorithmic and human recommendations (Senecal and Nantel, 2004; Krishnan et al., 2008; Amatrian et al., 2009; Sinha and Swearingen, 2011; Cowgill, 2020; Yeomans et al., 2019; Peukert et al., 2023) as we compare on-platform (human and algorithmic) recommendations with off-platform recommendations. Thereby, we are able to address one of the key limitations of studies based on field data, namely that they do not provide a contrast between users exposed to online recommendations and those who are not (Lee & Hosanagar, 2019). Fifth, we provide a measure of users' satisfaction, which is robust to the salience effect, which is not the case for studies relying on purchase or click data (e.g., Peukert et al., 2023).

4 Theoretical framework

4.1 Research questions and hypotheses

We argue that the effectiveness of a discovery mode of new songs can be determined by its ability to entice users toward songs that truly fit their preferences, i.e., songs they will repeatedly actively choose to listen to. These streams are called *organic* and contrasted with *curated* streams initiated via platform recommendation mechanisms. Organic streams are (1) streams that the user has initiated after having actively searched for the song or album or (2) streams that have been initiated as part of the user's personal playlist.

We acknowledge that some listeners are satisfied with listening to music through human-curated playlists or algorithmic recommendations. However, from an artist's perspective, only organic streams and engaged fans can help build a sustainable career beyond the realms of the streaming platform, such as concert ticket sales. Moreover, as we will discuss later, users in our sample primarily engaged

in passive music listening behavior only represent a small share of all users. They also tend to be low-intensive users, i.e., they use the service less than the average subscriber.

From the previous literature, the first research question pertains to the comparative effectiveness of the various modes of discovery. Users typically receive two types of information about the songs through two main modes of discovery: One type of information is *on-platform curation*, manifested by information from platform features discussed above, such as playlists and recommender systems. The second type of information is *off-platform information*, manifested by various information sources, including traditional media, social media, friends, family, etc.

We expect that on-platform curation is more effective than off-platform information in helping users discover content that truly fits their preferences. There are several reasons why this ought to be the case. First, song sequences generated by on-platform curation can potentially be much more attuned to users' personal preferences (1-to-1 recommendations) than traditional media such as radio or TV broadcasts (1-to-many recommendations) (e.g., Villermet et al., 2021). Further, the set of songs from which platform curated streams are drawn is much larger³ than those that friends and relatives know and can recommend (e.g., Peukert et al., 2023). Based on this reasoning, we establish our first hypothesis:

Hypothesis 1: On-platform curation is more effective than off-platform information in helping users discover content that fits their preferences.

Algorithmic recommendations are tailored to each user, while human-curated playlists are mainly untailored⁴. Since we expect tailored information to fit user preferences better than untailored, algorithmic recommendations should drive users toward more satisfying songs than playlists providing that enough information is available on users' preferences (e.g., Peukert et al., 2023). Hence, our second hypothesis:

³ Major streaming platforms (e.g., Apple, Deezer, Spotify) claim to have more than 100 million songs in their catalogs.

⁴ Some playlists are curated by combining inputs from humans and algorithmic personalization.

Hypothesis 2: Algorithmic recommendations are more effective than human-curated playlists in helping users discover content that fits their preferences.

A second research question deals with the interaction between the *type* of content and the comparative effectiveness of different discovery modes. Both human and algorithmic recommendations are supposed to help suggest content that a user would not otherwise have chosen (e.g., Fleder & Hosanagar, 2009). As mentioned above, human and music streaming platforms try to include this type of content in their algorithmic recommendations by assessing measures such as diversity, novelty, and serendipity in their recommender system designs (e.g., Adamopoulos & Tuzhilin, 2014). If successful, this approach should increase the number of *unfamiliar* songs suggested to the music listener. Conversely, more familiar content, because of the popularity of the artist or its nationality, is more likely to be promoted by common off-platform gatekeepers, such as friends, family, or traditional media.

This reasoning leads us to the following hypothesis:

Hypothesis 3: On-platform curation is more effective than off-platform information in helping users discover unfamiliar content that fits their preferences.

From the above observation that personalized recommendations should do better than untailed ones, we posit the following hypothesis:

Hypothesis 4: Algorithmic recommendations are more effective than human-curated playlists in helping users discover unfamiliar content that fits their preferences.

4.2 The notion of familiarity

To tackle this notion of “familiarity,” we argue that three artist-level dimensions are especially relevant to distinguish familiar from unfamiliar content: the artist’s overall popularity, the artist’s country of origin (domestic, US, or other), and the novelty of the artist from the perspective of the individual user.

Artist popularity: Popular artists benefit from traditional marketing promotion, TV interviews, and radio broadcasting. Their new releases do not need curation by streaming platforms to be discovered by users. Conversely, in line with the long tail hypothesis (Anderson, 2006), on-platform curation should be especially helpful for low-popularity artists who might fit with users' preferences but are lost among the millions of songs on the platforms. However, as illustrated in the literature review above, there are also arguments for a "rich-get-richer" effect with online recommendations that could conversely favor already popular artists (Fleder & Hosanagar, 2009; Lee & Hosanagar, 2019). Nevertheless, we posit that on-platform curation is more helpful than off-platform information, particularly for "low-popularity" artists (Aguiar et al., 2021; Wlömert et al., 2021) and that for this category of artists, personalized algorithmic recommendations are more helpful than human-curated playlists.

The artist's country of origin: Domestic content dominates most major national music markets – a fact that is even more true for countries such as France, Australia, Canada, or New Zealand with radio quotas (Bourreau et al., 2022; Waldfogel, 2018). Moreover, non-domestic content is less familiar to users because they benefit less from off-platform promotion. We thus argue that on-platform curation is more helpful than off-platform information for foreign artists (Aguiar et al., 2018; Waldfogel, 2018; Bourreau et al., 2022) and that for this category of artists, personalized algorithmic recommendations are more helpful than human-curated playlists.

The artist's novelty to the individual user: Finally, besides the above market familiarity factors, we also consider an individual familiarity feature. Novelty and serendipity are recognized as required characteristics of useful recommendations (e.g., Adamopoulos & Tuzhilin, 2014). Users expect more from curation than just being recommended new releases from artists they already know and have listened to recently. Thus, we expect on-platform curation to be more helpful than off-platform information when recommending music by artists who are entirely novel and unknown to a user. We also posit that for such "novel" artists, personalized algorithmic recommendations are more helpful than human-curated playlists.

Lastly, the impact of these criteria of familiarity is cumulative. “Least familiar” is the first stream of a song released by a foreign low-popularity artist not previously streamed by the user. “Most familiar” is the first stream of a song released by a high-popularity artist from France or the USA who has already streamed at least once by the user.

5 Empirical strategy

5.1 Data

The unit of observation in this study is a “song discovery,” i.e., the event when a user listens to a song for the first time on the music streaming platform. To generate the dataset, we start with 9,778 random premium subscribers of the music streaming platform Deezer. Deezer is a European music streaming service with 9.4 million users worldwide. The company is based in Paris, and while the service is available in 170 countries, France is where the platform has the largest market share. With its history as a large mainstream online music platform in the French music market, it is reasonable to assume that the characteristics of the platform users are similar to user characteristics on other large music platforms such as Spotify, Apple Music, or Amazon Music Unlimited. Furthermore, France is the world’s fifth-largest music market. While all national music markets are idiosyncratic, France shares many characteristics with other similar-sized markets, such as Germany, the Netherlands, or Canada, which means that the findings from this study are transferable to other advanced music markets around the world.

The streaming platform anonymized the user data, and the only demographic data provided to the research team was the users’ year of birth and gender. However, such user data is entered by the user when registering to the platform, and it is difficult to ascertain the accuracy of the data provided. For this reason, we do not use this data in this study. The original dataset consists of users streaming logs over three years, 2018-2020 inclusive. During this period, the users streamed 4,542,167 unique songs recorded by 434,798 artists at least once. In total, data on 323,779,669 streams are included in the original dataset.

We only want to analyze newly released songs and, at the same time, include as much usage data as possible about these songs. We do this by only focusing on songs first listened to by a user in the dataset during the three months of September, October, and November 2018. This yields an 8-month period prior to this period (January – August 2018) when the songs were never streamed by any user (since they had not yet been released). We then capture streaming data for the 24 months following the first stream, which allows a thorough examination of the users’ streaming behaviors. We argue that the selection of this period provides the right balance between ensuring the discovered song is indeed a new release and retaining enough usage data after its release. A closer examination of the release date of these songs shows that some of them were not recently released after all but are so-called back catalog, i.e., they were released more than 18 months before September 2018. It is not apparent why these songs have not been listened to for several months and why they were “re-discovered” by the users in our dataset in September 2018. We choose to drop these older releases and only focus on songs released after 1 January 2018, which are more likely to be discovered for the first time by our users. This procedure leaves us with 48,970 songs.

This paper is focused on song discoveries, and we are only interested in songs that (a) have indeed been discovered by multiple users and (b) are newly released. In order to remove songs that do not fit these criteria, we first select songs streamed by at least one percent of the users in our cohort (i.e., by at least 100 users). Music streaming is highly skewed towards a small number of successful artists and songs, and most songs in the platform’s catalog are rarely streamed. Consequently, only 4,136 songs (~8.4%) meet this one-percent criterion. 1,534 artists performed these songs, and they have been streamed 15,122,302 times by the 9,778 users in our dataset.

The songs were combined with users to generate a dataset with 1,952,195 song discoveries, and key variables were calculated for subsequent analysis. These variables are defined in the section below.

5.2 Variables

As an indicator of the fit of the recommended song with the user’s preferences – our dependent variable – we use the number of organic repeat streams (*ORGANIC_STREAMS*), which is the number

of times the user has organically streamed the song *after* the first stream. We consider only organic streams (defined as streams initiated by the user after having actively searched for the song or the album, as well as streams initiated as part of a user’s personal playlists) since streams initiated within a platform playlist or by an algorithmic recommendation reflect the platform’s prediction of the user’s preferences rather than the user’s genuine appreciation of a song.

Our main explanatory variables are three dummies representing the song discovery mode, i.e., how a user discovers a song or, specifically, the context of the user’s first stream of a song. We consider three possibilities:

OFF_PLATFORM represents when the first stream was initiated after the user actively searched for a song or album. This indicates that the user has discovered the song off-platform⁵.

PLAYLIST represents when the first stream was initiated within a human-curated playlist maintained by the streaming platform or by other major stakeholders (mainly major labels).

RECO represents a case when an algorithmic recommendation initiated the first stream.

In some treatments, we combine *PLAYLIST* and *RECO* in a single category labeled *CURATION*. Half of the discoveries in our sample are off-platform, while algorithmic recommendations and playlists each represent around a quarter of the discoveries (Table 4).

Table 4 – Distribution of the song discoveries (first streams) by mode (in percentage)

Off-platform information (OFF_PLATFORM)	50.87%
Human-curated playlists (PLAYLIST)	27.27%
Algorithmic recommendations (RECO)	21.86%
Total	100.00%

⁵ Broadcast media, expert reviews, friends, and relatives, etc. But also, possibly, another online platform such as YouTube or TikTok.

We include three artist-related variables – artist popularity, artist nationality, and artist novelty – that allow us to distinguish familiar from unfamiliar content, which may impact the effectiveness of on-platform curation. Firstly, we consider the artist’s popularity (*ARTIST_POPULARITY*). This variable is measured as an artist’s share of the total number of streams over the period. We argue that on-platform curation is especially useful in suggesting songs from low-popularity artists. Secondly, we consider the artist’s country of origin. Most music markets are dominated by music from domestic artists and artists from the world’s leading music exporter, the US. This is also the case for the dataset used in this study, and French and American content combined represents nearly 70% of all observations. This distribution is consistent with data from other related studies (e.g., Aguiar & Waldfogel, 2021; Bourreau et al., 2022). We argue that songs by artists from nations with a smaller market share benefit the most from on-platform curation. Artist nationality is represented in the model by three dummies: *FR*, *US*, and *OTHER_NATIONALITY*. The third variable captures whether a user is already familiar with the artist who recorded the song they discovered. We argue that on-platform curation is more effective in suggesting a song by an unfamiliar artist than a song by an artist that the user already knows. Finally, we use a dummy (*NOVEL_ARTIST*) representing whether the user knows the artist or not. This is determined by controlling if the dataset contains evidence that the user has streamed other songs by the artist in question prior to the song discovery. Since the song discoveries are all happening in September, October, and November 2018, and the dataset contains data from the beginning of that year, we can only search for evidence during this 8–10-month period. Therefore, we cannot ascertain that the user has not listened to music by the artist in question in 2017 or earlier since this period is outside the scope of our dataset.

Finally, we use a set of control variables that could contribute to explaining the number of repeat streams made by a user. Some musical genres, for instance, hip-hop & rap, could be more favorable to intensive repeat streams. We thus include eight dummies for each musical genre in the sample: *CLASSICAL*, *JAZZ*, *POP*, *ROCK*, *HIPHOP_RAP*, *RNB_SOUL*, *ELECTRO*, and *OTHER_GENRE*. We also include three dummies accounting for the type of provider that makes the song available on the

platform: major labels (*MAJOR*), digital aggregators such as Believe⁶ (*AGGREGATOR*), and independent labels (*INDIE*). Of course, the song’s duration (*SONG_DURATION* is a continuous variable in seconds) and the song’s popularity (*SONG_POPULARITY* is a continuous variable that accounts for a song’s share of all the streams over the period) can also have an impact on the number of repeat streams independently from how the user has discovered the song. Finally, we also include two user-level variables. The first one considers the user’s habit of repeatedly listening to the same song. *RELISTENING_HABIT* is a continuous variable that measures each user’s average repeat stream per song. The second variable measures the completion ratio (*COMPLET_RATIO*) of the first time a user listens to a specific song. We expect that the lower this completion ratio, the lower the likelihood that the user will organically relisten to it frequently. Finally, we include dummies for the month when the user discovered the song.

Table A1 in the appendix provides the correlation between the main independent variables, and Table 5 summarizes the descriptive statistics of our sample.

Table 5 – Descriptive statistics

Variable	Dummy/variable name	Counts	Mean	St Dev	Min	Max
Dependent variable	ORGANIC_STREAMS	1,952,195	4.74291	17.60554	0	1915
Discovery mode	RECO	1,946,249	.2176849	.4126721	0	1
	PLAYLIST	1,943,547	.2719559	.4449674	0	1
	CURATION	1,938,351	.4912567	.4999237	0	1
	OFF-PLATFORM	1,939,074	.5085536	.499927	0	1
User familiarity	NOVEL_ARTIST	1,952,195	.2956421	.456331	0	1
Artist popularity	ARTIST_POPULARITY	1,952,195	.0073941	.0113152	3.32e-06	.0499953
Artist nationality	US	1,952,195	.1939273	.3953728	0	1
	FR	1,952,195	.5049485	.4999756	0	1
	OTHER_NATIONALITY	1,952,195	.3011223	.4587459	0	1
Label category	MAJOR	1,952,195	.7730729	.4188452	0	1
	AGGREGATOR	1,952,195	.0882202	.2836149	0	1
	INDIE	1,952,195	.1387069	.3456405	0	1
Genre	CLASSICAL	1,952,195	.0118713	.1083067	0	1
	ELECTRO	1,952,195	.0813694	.2734017	0	1
	POP	1,952,195	.3448703	.4753261	0	1
	ROCK	1,952,195	.1325969	.3391386	0	1
	JAZZ	1,952,195	.0108053	.1033853	0	1
	RNB_SOUL	1,952,195	.0464682	.2104969	0	1
	OTHER_GENRE	1,952,195	.0537723	.2255679	0	1
	HIPHOP_RAP	1,952,195	.2859576	.4518694	0	1

⁶ Believe is a French music company (<https://www.believe.com/>).

User listening profile	RELISTENING_HABIT	1,952,195	2.303671	.9194589	1.037572	16.57937
Song characteristics	SONG_POPULARITY	1,952,195	.0010906	.0020698	1.61e-06	.0138559
	SONG_DURATION	1,952,195	208.7675	53.08466	10	839
Completion ratio	COMPLET_RATIO	1,952,195	.5520284	.3995572	.0011919	1
Month	SEPT	1,952,195	.2518975	.4341028	0	1
	OCT	1,952,195	.3380733	.4730538	0	1
	NOV	1,952,195	.4100292	.4918388	0	1

5.3 Econometric model

Our dependent variable is a count variable that presents a lot of zeros: *ORGANIC_STREAMS* is zero for 69.7% of all observations. Put another way, almost 70% of all song discoveries do not entice the music listener to organically stream the song even once. This large number of zeros means that the standard deviation is much greater than the mean, suggesting overdispersion. Therefore, a negative binomial model is likely more appropriate than a Poisson model (this will be properly checked later). However, since the mere observation of a large number of zeros does not necessarily mean an “excess” of zeros exists, we must first check whether this is the case. If so, a *zero-inflated* negative binomial regression model (ZINB) can potentially be relevant (Lambert, 1992; Cameron & Trivedi, 2005).

A zero-inflated model requires that there exist two types of zeros generated by two distinct processes. The first process is binary and predicts “certain zeros.” In this context, a certain zero would represent a song recommendation that had no chance of leading to an organic stream. One such scenario could be a song recommendation that the user simply did not hear because they were not in the room when the song was playing. The second process determines the frequency of the count variables, including possible “regular” zeros but excluding “certain zeros.” A regular zero in our setting represents a song recommendation that could have led to an organic stream but did not – perhaps due to the song’s poor fit with the user’s music preferences. In this two-stage process, the first equation (inflate regression) is a Logit that only aims to explain the zero observations. The second equation (main model) is a negative binomial regression explaining the count variable.

In our setting, it is impossible to know whether the zeros in the variable are “certain” or “regular” since all recommendations, regardless of whether they are human-curated, algorithmic, or off-

platform, are intended to fit a user's preferences. To test for that, we use the *countfit* procedure in Stata (Long & Freese, 2014), which supports the zero-inflated negative binomial model over the simple negative binomial model (see Table A2 in the appendix).⁷

A limitation of this analysis is that the direct comparison between on-platform and off-platform discoveries is subject to selection bias. Baseline characteristics of treated subjects (on-platform discoveries) often differ systematically from those of control subjects (off-platform discoveries). We thus implement a Propensity Score Matching (PSM) analysis which is widely used to reduce this selection bias and identify a causal inference (Austin, 2011).

6 Results

6.1 The global effect of on-platform curation vs. off-platform information

Table 6 shows the results of the zero-inflated negative binomial model. The value of $\ln(\alpha)$ confirms that there is overdispersion and that the zero-inflated negative binomial should be preferred to a zero-inflated Poisson model. The main model uses a user's number of organic repeat streams of the same song as the dependent variable. The inflate regression highlights the variables that strongly explain the excess in zeros. It confirms that first streams (potential discoveries) initiated by on-platform curation are more likely to be "certain zeros" than those sparked by off-platform information. However, even when we control for these "certain" zeros, on-platform curation is still less effective than off-platform information for suggesting songs that fit users' preferences.

⁷ Our data confirm that the excess in zeros is clearly linked to on-platform curation. 47.3% of songs discovered off-platform are organically restreamed at least once. This should be compared to songs discovered through on-platform curation: only 12.6% of songs discovered through an algorithmic recommendation are organically streamed at least once after discovery, with the corresponding value for songs discovered via a human-curated playlist being 15.6%.

Table 6 – Results of the Zero Inflated Negative Binomial model (on-platform curation vs. off-platform information)

Dependent variable: ORGANIC_STREAMS		
	Main	Inflate
CURATION	0.715*** (0.00514)	21.54*** (0.263)
OFF-PLATFORM	Ref.	Ref.
Constant	0.390*** (0.00859)	0.113*** (0.00412)
ln_alpha	4.264*** (0.0113)	
N	1,938,351	

Exponentiated coefficients; robust standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: The regression includes control for the type of provider, the music genre, the country of origin, the average number of repeat streams per song (user level), the artist's popularity, the familiarity of the user with the artist, the song popularity, the song duration, the completion rate of the first stream of the song (user level), and month dummies.

Since the coefficients are exponentiated, the interpretation of Table 6 is straightforward. The main regression shows that, on average, a song discovered through on-platform curation generates only 71.5% of the repeat streams generated by a song discovered off-platform and initiated organically by the user. In other words, hypothesis 1 is not supported since on-platform curation is not more effective than off-platform information in suggesting content that really fits user preferences.

We now focus specifically on comparing the two modes of on-platform curation. Table 7 shows no significant difference between these two types of curation regarding the excess of zeros (see the table column for inflate regression). Moreover, it shows that, on average, a song discovered through an algorithmic recommendation generates ~20% fewer repeat streams than a song discovered within a playlist. Hence, opposite to hypothesis 2, personalized algorithmic recommendations do not do better than human-curated playlists in suggesting songs that fit users' preferences.

Table 7 – Results of the Zero Inflated Negative Binomial model (algorithmic recommendations vs. playlists)

Dependent variable: ORGANIC_STREAMS		
	Main	Inflate
RECO	0.804*** (0.0119)	0.889*** (0.0145)
PLAYLIST	Ref.	Ref.
Constant	0.0623*** (0.00460)	0.535*** (0.0518)
ln_alpha	8.636*** (0.0761)	
N	952,228	

Exponentiated coefficients; robust standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: The regression includes control for the type of provider, the music genre, the country of origin, the average number of repeat streams per song (user level), the artist's popularity, the familiarity of the user with the artist, the song popularity, the song duration, the completion rate of the first stream of the song (user level), and month dummies.

6.2 The role of familiarity

In the tables below, we analyze the relative performances of on-platform curation in providing music that really fits users' preferences along three familiarity criteria: the artist's popularity, the artist's country of origin, and the novelty of the artist to the user.

To account for the first dimension of familiarity, the sample has been divided into three approximately equivalent subsets according to artists' popularity measured by their market share in streams over the whole period. To account for the impact of the country of origin, we distinguish domestic content (50.5% of our observations), from US content (19.4%), and those from other foreign countries (30.1%). Finally, to account for the third dimension of familiarity (whether the artist is novel to the user or not), we distinguish song discoveries where the user has already listened to songs by this artist (i.e., not a novel artist) from song discoveries where there is no evidence that the user has ever listened to the artist (i.e., it is a novel artist). The impact of the three familiarity criteria being cumulative, we compare the effectiveness of on-platform curation in recommending (a) songs that a user is least likely to be familiar with and (b) songs that a user is most likely to be familiar with. We create these two categories by combining our three familiarity measures: "Least familiar" is a user's first stream of a song released by a foreign low-popularity artist not previously streamed by the user. "Most familiar" is a user's first stream of a song released by a domestic high-popularity artist who has already been streamed at least once by the user. We also create a third category, "Intermediate", for songs that belong neither to the "Least familiar" nor to the "Most familiar". The implementation and interpretation of interaction terms in count data regressions are especially complex (Leitgöb, 2014), which is why we split our large sample into three subsamples: least familiar, most familiar and intermediate (the remaining of discoveries).

Table 8 – Results of the Zero Inflated Negative Binomial model for most and least familiar songs

Dependent variable: ORGANIC_STREAMS

	Least familiar		Intermediate		Most familiar	
	Main	Inflate	Main	Inflate	Main	Inflate
CURATION	1.126*** (0.0324)	40.75*** (3.285)	0.742*** (0.00666)	25.32*** (0.551)	0.332*** (0.00438)	4.069*** (0.105)
OFF_PLATFORM	Ref.		Ref.		Ref.	
Constant	0.232*** (0.0160)	0.0699*** (0.00727)	0.405*** (0.0109)	0.0856*** (0.00426)	0.647*** (0.0310)	6.867*** (0.938)
ln_alpha	6.099*** (0.0975)		4.520*** (0.0171)		4.047*** (0.0138)	
N	266,926		1,271,537		399,888	

Exponentiated coefficients; robust standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: regressions include control for the type of provider, the music genre, the country of origin, the average number of repeat streams per song (user level), the song popularity, the song duration, the completion rate of the first stream of the song (user level), and month dummies.

Table 8 (above) shows that on-platform curation of the least familiar songs is significantly more effective than off-platform information (+12.6% more organic streams). Conversely, on-platform curation of the most familiar songs generates fewer repeat streams (-66.8%) than off-platform information. In other words, on-platform curation is more effective than off-platform discoveries in recommending non domestic long-tail artists that the user is not already familiar with. Conversely, off-platform curation does better for more famous domestic artists that the user is already familiar with.

We now turn to the relative performances of both modes of on-platform curation when content familiarity is considered. Table 9 (below) also distinguishes the “least,” “intermediate,” and “most” familiar content. While there is no significant difference between algorithmic recommendation and playlists in recommending relevant unfamiliar discoveries to users, algorithmic recommendations have a clear disadvantage in recommending new releases much more familiar to the users (-23.7% of organic streams generated by the recommendations).^{8,9}

⁸ We provide in appendix (Tables A3 to A5 and Tables A6 to A8) the results of the regressions for the three dimensions separately. They are in line with the results at the aggregate level.

⁹ An obvious explanation of these results would be that playlists tend to focus on presumably more familiar content. If such content is over-represented in playlists, it should automatically increase the likelihood of recommending content that fits users’ preferences. However, as illustrated by Table A9, this is not what our data shows. We observe the opposite. Playlists

Table 9 – Results of the Zero Inflated Negative Binomial model for most and least familiar songs

Dependent variable: ORGANIC_STREAMS						
	Least familiar		Intermediate		Most familiar	
	Main	Inflate	Main	Inflate	Main	Inflate
RECO	0.962 (0.0526)	0.485*** (0.0306)	0.781*** (0.0137)	0.667*** (0.0160)	0.763*** (0.0195)	0.845*** (0.0428)
PLAYLIST	Ref.		Ref.		Ref.	
Constant	0.0155*** (0.00272)	2.51e-28*** (6.96e-28)	0.0385*** (0.00285)	7.96e-30*** (1.34e-29)	0.0144*** (0.00180)	3.59e-31*** (1.21e-30)
ln_alpha	26.22*** (0.530)		10.78*** (0.0637)		6.659*** (0.0572)	
N	175,737		623,330		153,161	

Exponentiated coefficients; robust standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: regressions include control for the type of provider, the music genre, the country of origin, the average number of repeat streams per song (user level), the song popularity, the song duration, the completion rate of the first stream of the song (user level), and month dummies.

To sum up, overall, on-platform curation is less effective than off-platform information in suggesting relevant content (Hypothesis 1 is not supported). Moreover, human-curated playlists are, on average, more effective than algorithmic recommendations (Hypothesis 2 is not supported either). Secondly, on-platform curation is more effective in suggesting relevant, unfamiliar content. Hypothesis 3 is thus supported. Human-curated playlists are always more effective than algorithmic recommendations in suggesting relevant, more familiar content within the on-platform curation discovery modes.

Conversely, algorithmic recommendations do better (or less bad) for more unfamiliar content, supporting Hypothesis 4.

6.3 Causal inference

A limitation of the above analysis is that, alike most observational studies, we cannot attribute the differences in the outcome between the on-platform and off-platform discoveries to the pure effect of curation. The treatment allocation (on-platform discovery) is not random but potentially influenced by

tend to recommend a higher proportion of unfamiliar content than algorithmic recommendations. For instance, while discoveries made within playlists account for only 27.2% of the sample, their share is almost 10 points higher in the subsamples constituted by songs recorded by low-popularity artists, by non-US and non-domestic countries or by artists that are novel to the user. Nevertheless, these least familiar songs clearly do not fit users' preferences better since they do not generate more repeat streams than discoveries made through algorithmic recommendations.

subject characteristics. Although we control for a lot of user-song characteristics in the previous regressions, we lack counterfactuals to ensure a causal inference. When the user receives an on-platform recommendation we cannot compare the output with what would have happened if the user had preferred to rather follow at this very moment an off-platform recommendation of the same kind of song (that would have satisfied the same expectations).

Methods based on the propensity score are widely used to reduce this selection bias (Austin, 2011). Propensity Score Matching (PSM) relying on the nearest neighbor matching aims at building a specific control group that allows us to match each treated observation, i.e. an on-platform discovery (resp. algorithmic recommendation), with one control observation, i.e. an off-platform discovery (resp. human curated playlist), that has the closest propensity score; that is to say, almost the same characteristics in terms of artist's popularity and nationality, music genre, type of music label, user's repeat listening behaviour, etc. Hence, the only remaining explanation to a difference in the two groups in expected outcomes would be the mode of discovery.

Implementing PSM with nearest neighbor matching, we display on Table 10 the average treatment effects (ATE) for the various cases we studied previously and show that our main conclusions still hold. off-platform recommendations do better than on-platform curation to help users discovering content that fit with their preferences. In average, content suggested by the platform is listened to 5 times less than content suggested outside the platform. However, online curation does relatively better for the least familiar content (the difference is then only 3 times less). Likewise, while the difference between algorithmic recommendations and human curated playlist is much less important (-0.12 organic streams), algorithmic recommendations do also better for unfamiliar content (+0.18) than for familiar content (-0.66). This supports our previous results and suggests that there is indeed a causal effect between the discovery mode and the fit with users' preferences.

Table 10 – Average Treatment Effects of Propensity Score Matching (nearest neighbor matching)

Dependent variable: ORGANIC_STREAMS

	CURATION (treated) vs. OFF-PLATFORM (matched control)			RECO (treated) vs. PLAYLIST (matched control)		
	Whole sample	Least familiar	Most familiar	Whole sample	Least familiar	Most familiar
ATE	-5.0058*** (.03014)	-3.0739*** (.06320)	-6.7816*** (.08252)	-.11584*** (.03253)	.17729*** (.03809)	-.65543*** (.09626)
N	1,938,351	266,926	399,888	952,228	175,737	623,330

AI robust standard error in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

However, PSM requires some conditions to be fulfilled. Firstly, the quality of matching has to be checked. This is done by the analysis of the standardized differences in the means of all the covariates before and after PSM. In all the cases presented on Table 10, the standardized differences are lower than 0.1, which indicates a negligible difference in the mean or proportion of a covariate between the treatment and control groups (Ho, 2007). We are thus confident on the quality of the matching. Secondly, the causal interpretation of PSM results also requires that two main assumptions are fulfilled. On the one hand, it is assumed that each observation has a positive probability to be in the treatment or in the control group in order to avoid comparing non-comparable individuals (Common Support Assumption). In our case, of course some songs have a higher probability to be recommended to a specific user, but there is no reason to believe that some songs are excluded. On the other hand – and this is clearly the strongest assumption (Caliendo & Kopeining, 2008) – selection is supposed to be solely based on observable characteristics and all variables that influence treatment assignment and potential outcomes simultaneously have been accounted for (Conditional Independence Assumption or CIA). Hence, no unobservable characteristics could influence both the treatment and the outcome. Although CIA is an untestable assumption, our covariates including all the most important characteristics used by the platform to characterize a song and thus its potential fit with users' preferences, we are confident that the potential unobserved confounders should be highly correlated with the covariates in our model. The outcome measure should not be related to the treatment anymore, other than via the efficiency of the recommender system.

6.4 Robustness

Firstly, we have included a set of control variables that can contribute to explaining the impact of the discovery mode on listener satisfaction. However, we did not consider the user's time on the platform before the data collection commenced. Alike Peukert et al. (2023), we could posit that it is easiest to suggest relevant music to long-time subscribers rather than newcomers (because more data has been collected on the former). However, including the number of days since the user has subscribed to the platform as a control variable does not change our results (see Table A10 in the appendix).

Fraud is widely acknowledged as a significant issue in the streaming market (e.g., Smith Burton, 2021). While a robot that is using a subscription account to stream the same set of artists 24 hours a day is probably detected by platforms, we cannot exclude that some songs extensively streamed by the same account could be fraudulent. In our sample, some songs are repeatedly streamed very frequently by the same individual (the maximum being 1,915 streams for the same song by the same user over the period). To control for this potential issue, we ran our main regression excluding song discoveries where the number of repeat streams exceeded 50 (which only accounts for 2.5% of the whole sample). Table A11 in the appendix shows that the results remain qualitatively similar to the regression in Table 6.

A third potential issue relates to the way we measure satisfaction provided by a song. As noticed previously, repeat consumption is widely recognized as an indicator of satisfaction (La Barbera & Mazursky, 1983; Bhattacharjee, 2001; Liu et al., 2011; Rauyrueen et al., 2007; Zeithaml & Berry, 1996; Paul et al., 2009), especially for music (Schedl et al., 2018). However, to check for the robustness of our results, we ran our main regression with a simple probit and a zero-inflated ordered probit. The first regression distinguishes songs that have been streamed once and never again from songs that have been organically streamed at least one more time. The second regression separates songs that have been streamed once and never again from those streamed only once after the discovery stream and from those that have been streamed again at least twice. This allows us to account for the fact that all users do not have the same satiety function for music and that a single user may not have the same satiety function for all genres of music and all songs. Tables A12 and A13 in the appendix show that

our main result remains unchanged. On the whole sample, songs discovered through on-platform curation remain less frequently streamed than songs discovered off-platform.

A fourth robustness check relates to the dimensions of familiarity we consider. We claim that superstardom (popularity), domestic content, and artist novelty are relevant dimensions of familiarity in the music industry but acknowledge that there are other relevant dimensions. For instance, there is a traditional opposition between broad music genres (pop, hip-hop & rap) and niche genres. Within our sample, the two dominant genres represent 35.6% and 29.4% of the observations, respectively, while the other smaller genres (classical, jazz, etc.) account for only a third of the observations. Tables A14 and A15 in the appendix provide the results of zero-inflated negative binomial regressions that consider this alternative dimension of familiarity. The results are consistent with those obtained with the three original dimensions of familiarity. On-platform curation is more effective for niche genres, and, within on-platform curation, algorithmic recommendations do relatively better for niche genres than for dominant genres.

A fifth issue concerns the possible asymmetry between on-platform curation and off-platform recommendations. It is reasonable to assume that out of all the songs someone is exposed to off-platform – on broadcast radio and television, on YouTube, on TikTok, recommended by friends and relatives, etc. – only a few are “good enough” to entice the music listener to do an active search for the song on their music streaming platform. We can thus study all songs discovered via on-platform curation, but only those songs discovered off-platform that result in on-platform listening. This deficiency can explain the lower effectiveness of on-platform curation and especially the higher rate of certain zeros. To control for this, we have analyzed only those song discoveries that generated at least one organic stream after the first stream. Thereby, we only include songs discovered either on- or off-platform that are “good enough” to spur the music listener to actively search on the platform for more. We thus run a Negative Binomial regression on only 615,944 song discoveries. Even with this subsample, recommendations via on-platform curation are less effective than off-platform discoveries (cf. Table A16 in the appendix). A song discovered through on-platform curation still generates 20%

fewer organic streams than one discovered off-platform (-7% for a song discovered through algorithmic recommendations as compared to human edited playlist).

A sixth issue pertains to the fact that we withdrew the least popular songs from the dataset, consumed by less than 1% of our users. Only 8.4% of the songs are listened to by more than 1% of the users, but on the other hand, these songs account for 88% of the streams and 75% of the user discoveries.¹⁰

Nevertheless, it could be that these least popular songs are precisely those that benefit the most from curation since they are more likely to be unfamiliar to users. While this would have a low impact at the market level (because of the low number of streams concerned), this could be important for individual user satisfaction. To check for that, we run our main regressions in this new sample of users discovering niche songs. As shown in Table 11, our main results still hold, and the new results align with the previous results related to discoveries of unfamiliar songs. Curation remains less efficient than off-platform discoveries to drive users to content that fits their preferences. However, the difference is less significant than in the sample of more popular songs and the magnitude is lower (cf. Table 6): -7.9% vs. -28.5% of organic streams per discovery compared to off-platform discoveries. Furthermore, within these very unpopular songs, algorithmic recommendations appear to be more efficient than playlists in driving users toward content that fits their preferences. On average, a discovery through algorithmic recommendation leads to 84% more organic streams than a discovery through a playlist. This analysis of the sample of very unpopular songs gives additional support to hypothesis 4 (algorithmic recommendations do better than playlists for more unfamiliar content).

Table 11 – Zero-Inflated Negative Binomial regression on the sample of songs that gather less than 1% of the users

DV: ORGANIC_STREAMS				
	All discoveries (on- and off-platform)		On-platform discoveries only	
	Main	Inflate	Main	Inflate
CURATION	0.921*	30.77***		

¹⁰ Including the least popular songs would have made the # songs increased from 4,136 to 48,620, # artists increased from 1,534 to 15,257, # streams increased from 15,122,302 to 17,125,002, and # discoveries increased from 1,952,195 to 2,419,983.

OFF_PLATFORM	(0.0298) Ref.	(0.955)		
RECO			1.836*** (0.0939)	1.526*** (0.129)
PLAYLIST			Ref.	
Constant	2.465*** (0.132)	1.028*** (0.0701)	3.154*** (0.535)	15.89*** (3.306)
ln_alpha	5.512*** (0.0276)		30.29*** (0.536)	
N	467,765		136,880	

Exponentiated coefficients; robust standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: regressions include control for the type of provider, the music genre, the country of origin, the average number of repeat streams per song (user level), the song popularity, the song duration, the artist's popularity, the completion rate of the first stream of the song (user level), and month dummies.

Finally, we have not taken into account that some consumers may prefer passive listening. However, it is important to note that such passive listeners make up only 30% of our observations and belong to the category of least intensive users. Moreover, the existence of passive listeners does not explain our results. As mentioned previously, even when we exclude passive listeners and limit the analysis to users who organically relisten to each song at least once over the period, our results still hold (Table A16 in the appendix). Curated discoveries are still more likely to be certain zeros, and even when they are not, they generate fewer repeat streams than off-platform discoveries.

7 Discussion

7.1 Opening the black box of curation

Our first main result is that despite the huge number of data that music streaming platforms are able to leverage, human offline curation remains more effective than online curation (both human and algorithmic) in driving users to relevant discoveries, i.e., content that fits their preferences and that they are prone to relisten again. As mentioned earlier, 85.4% of new songs discovered through on-platform curation are *never* relistened to organically, while this figure drops to 52.2% for songs discovered off-platform. There are at least two ways to explain this finding. One explanation is that on-platform recommendations are simply not a good enough fit with user preferences to ignite a positive response. An alternative explanation is that users are not paying attention when they listen to

music on a playlist or via algorithmic recommendations. So, even though these songs could potentially fit their preferences, they do not elicit a reaction from users. Our analysis shows, however, that even when we restrict the sample only to songs actively noticed by the users, songs discovered via on-platform curation still generate fewer repeat streams than off-platform discoveries. Hence, the alternative explanation does not hold. The fact that so many songs discovered via on-platform curation never result in a single organic repeat stream can thereby only be explained by recommendations that are a poor fit with user preferences. In other terms, there seems to be room for improvement in the quality of algorithmic recommendations and the composition of playlists. Why are personalized recommendations less effective than mainly untargeted playlist recommendations? Since the most popular playlists are constituted by already popular songs (within a specific theme or at a general level), being included in a playlist could be a “quality” signal in the same vein as user ratings. For users sensitive to network effects in cultural consumption and who like consuming what others are also consuming (Adler, 1985), playlists could thereby turn out to be more valuable than recommendations, which can be perceived to be based on obscure algorithms. A second plausible explanation is that streaming platforms do not design algorithmic recommendations to introduce new songs to users. As discussed in section 2.3, at the core of all recommender system design sits the challenge of balancing the accuracy and the serendipity of the algorithmic recommendations. When deciding where on this “diversity spectrum” the recommendations should be placed, it is generally riskier (from the platform’s point of view) to design a system that frequently offers serendipitous discoveries since a song in the flow that is too far off the user’s preferences, can have a significantly negative impact on the user experience and cause them to stop listening and increase churn. It is safer to lean towards the accuracy end of the diversity spectrum, where users may not discover anything new but where the songs are sure not to upset the user’s musical enjoyment. This reasoning explains the findings by Anderson et al. (2020) and can also explain why users in our sample are more likely to encounter useful recommendations via human-curated playlists.

Our second main result is that this domination of human experience over data curation is less prominent when it comes to unfamiliar content. Our results support our hypotheses about the higher

effectiveness of on-platform curation over off-platform discoveries and algorithmic recommendations compared to human-curated playlists when recommending unfamiliar songs to users. Clearly, the platform's algorithmic recommendation mechanisms can identify and suggest unfamiliar songs to users that they most likely would not have discovered otherwise. These results are also in line with Peukert et al. (2023). For unfamiliar content, the platform curation benefits from more relevant data than offline human curation. Niche content is less likely to be recommended by both word-of-mouth and radio broadcasting. Furthermore, online human curation through playlists could be less effective than algorithmic recommendations because unfamiliar content is less likely to meet the expectations of a large number of users, which is the aim of most human-curated playlists. One implication of these results is that the platform recommendation mechanisms are indeed contributing to the expansion of the users' musical experiences and a wider distribution of streaming royalties among rightsholders. Multiple streaming platforms¹¹ regularly claim this is happening and use the claim as evidence of their positive impact on the industry. It should be noted, however, that our findings do not say anything about the diversity of the individual users' music listening experience. Further, our findings do not contradict Anderson et al. (2020), who showed that users listening to music via algorithmic recommendations lead to a consumer experience characterized by diminishing genre diversity.

7.2 Managerial implications

Our results have managerial implications, both at the platform and the right-holder level. As far as digital media platforms are concerned, our paper highlights that for on-platform curation to be effective, the recommendations need to be a good fit with the user's preferences, but the user, first and foremost, must be able to perceive the recommendations. Most of the research on platform curation, especially recommender systems, has been devoted to the first issue, but our results show how crucial the second issue is. Platforms should try to identify patterns (e.g., context, time of day, day of the week) that maximize the likelihood that users will actively notice the recommendations, which means

¹¹ E.g., <https://loudandclear.byspotify.com/>

that it has the potential to lead to repeated organic streams. There are indications that some platforms are working on these questions, and our work further highlights the urgency of such initiatives.

From the right-holder perspective (also present in academic literature), there seems to be a misconception about the significance of getting an artist's song added to popular playlists. According to Wlömert et al. (2021), playlists are much more effective in shaping demand than traditional forms of marketing and promotion, such as TV advertising or radio airplay, particularly because a song included in a playlist is considered "a free trial" opportunity. However, Wlömert et al. (2021) rely on aggregate data that do not allow them to isolate repeat consumption from the same users, which can explain the discrepancy between their results and what is found in the present study. Perhaps on-platform curation can shape demand better than broadcast TV and radio in the short run. However, offline promotion tools still do better when driving users toward content that genuinely fits their preferences. Likewise, Aguiar & Waldfogel (2021) show that being included in a popular playlist increases the number of song streams, but mostly only during the period the song belongs to the playlist. This result is consistent with our claim that songs discovered within playlists generate much less repeat consumption than off-platform discoveries. Since we obtained the same result with algorithmic recommendations, it appears that having music curated by platforms may not be as effective as musicians and music labels believe. Our results show that, except for unfamiliar content, curated streams do not generate a considerable number of repeat streams, which suggests that platforms may have a limited ability to influence users' music consumption practices.

For music labels and artists, it is essential to have a high number of streams to generate revenues. However, it is also necessary to build a sustainable and active fanbase. While streams originating from playlists or recommender systems generate streaming royalties, there is a risk that listeners only pay limited attention to which artist or which song is played. Therefore, it is unlikely that curated streams make a significant contribution to an artist's fanbase and long-term career. Artists need an intimate relationship with their fans to cultivate multiple on- and off-platform revenue streams (e.g., live concerts). On a streaming platform, such a relationship is manifested by a sizable number of organic streams and less by inclusion in playlists and algorithmic flows.

8 Conclusion

Our paper reported on the analysis of the quality of algorithmic and human recommendations using a stream-level longitudinal dataset from a music streaming platform. The data enables the analysis of individual users and the various ways they discover and listen to new songs on the platform. We exploit an inherent feature of recorded music consumption, namely that, unlike most other experience products, when music listeners like a song, they tend to listen to it many times. This allows us to account for the saliency effect of online recommendations that makes unsure that a recommendation-induced consumption really fits user preferences. Using this novel approach and unique dataset, we study how users discover newly released songs and how many times they organically relisten to these songs in the two years following the discovery.

Our findings show that on-platform curation is less effective than off-platform information in recommending satisfying content to users and that human-curated playlists are, on average, more effective than algorithmic recommendations. Further, on-platform curation effectively recommends relevant, *unfamiliar* content, especially from low-popularity artists and foreign origins. Human-curated playlists are always more effective than algorithmic recommendations in suggesting relevant, more familiar content within the on-platform curation discovery modes. Conversely, algorithmic recommendations do better (or less bad) recommending relevant, unfamiliar content.

Based on our findings, we have identified a number of managerial implications both for platforms and for rightsholders. We show that off-platform recommendations still significantly impact user listener preferences, and we suggest that platforms need to focus more on identifying the ideal conditions for recommendations to maximize the long-term impact on music listener satisfaction and subscriber loyalty.

Finally, we do acknowledge that our research has certain limitations. Given the large sample of users and the high granularity of the data gathered, we are confident of the internal validity of our research. The external validity is more questionable. Our study focused solely on a single music streaming service in one country. While this service is one of the two primary services in France, which is the

fifth largest market worldwide, it would be interesting to replicate our research in other countries and with other music streaming services. For instance, Spotify has announced that it has introduced a level of bias to its recommendations through its “Discovery Mode” program. This program allows artists to influence the radio and autoplay algorithm and increase their exposure on the platform.¹² Our research design makes it possible to evaluate the extent to which the discovery mode impacts user satisfaction and whether both stakeholders’ and users’ expectations can be fulfilled at the same time. In addition to this research opportunity, our approach can also be extended to products other than music, for which repeat consumption also is a signal of user satisfaction. Undoubtedly a promising avenue for future research.

References

- Adamopoulos, P. & Tuzhilin, A. (2014). On Unexpectedness in Recommender Systems: Or How to Better Expect the Unexpected. *ACM Transactions on Intelligent Systems and Technology*, 5(4):1-32
- Adler, M. (1985). Stardom and talent. *American Economic Review*, 75: 208–12.
- Adomavicius, G. & Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems: A Survey of the State-Of-The-Art and Possible Extensions. *IEEE transactions on knowledge and data engineering*, 17(6): 734–749.
- Adomavicius, G., Bockstedt, J.C., Curley, S.P. & Zhang, J. (2018). Effects of online recommendations on consumers’ willingness to pay. *Information Systems Research*, 29(1): 84-102.
- Adomavicius, G., Bockstedt, J.C., Curley, S.P. & Zhang, J. (2014). Do recommender systems manipulate consumer preferences? A study of anchoring effects. *Information Systems Research*, 24(4): 956–975.
- Aggarwal, P. & Vaidyanathan, R. (2005). Perceived effectiveness of recommendation agent routines: Search vs. experience goods. *International Journal of Internet Marketing and Advertising*, 2(1/2).
- Aguiar, L., & Waldfogel, J. (2018). As streaming reaches flood stage, does it stimulate or depress music sales? *International Journal of Industrial Organization*, 57: 278–307.
- Aguiar, L., & Waldfogel, J. (2021). Platforms, power, and promotion: Evidence from Spotify playlists. *Journal of Industrial Economics*, 69(3): 653–691.
- Aguiar, L., Gomez-Herrera, E. & Waldfogel, J. (2018). Does digitization threaten local culture? Music in the transition from iTunes to Spotify, mimeo.
- Aguiar, L., Waldfogel, J. & Waldfogel, S. (2021). Playlisting favorites: Measuring platform bias in the music industry. *International Journal of Industrial Organization*, 78: 102765.
- Amatriain, X., Lathia, N., Pujol, J. M., Kwak, H., & Oliver, N. (2009). The wisdom of the few: a collaborative filtering approach based on expert opinions from the web. *32nd international ACM SIGIR conference on Research & Development in Information Retrieval*, 532–539.
- Anderson, C. (2006). *The Long Tail: Why The Future Of Business Is Selling Less Of More*. Hyperion Books.
- Anderson, A., Maystre, L., Anderson, I., Mehrotra, R., & Lalmas, M. (2020). Algorithmic Effects on the Diversity of Consumption on Spotify. *Proceedings of The Web Conference 2020*, 2155–2165.

¹² <https://www.billboard.com/pro/spotify-discovery-mode-expands-access-stream-on-event/>

- Austin, P.C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research*, 46(3): 399-424.
- Barrington L., Oda R. & Lanckriet G.R.G. (2009). Smarter than genius? Human evaluation of music recommender systems. In: Proceedings of the 10th international society for music information retrieval conference, *ISMIR 2009*, 357–362
- Bhattacharjee, A. (2001). Understanding information systems continuance: an expectation-confirmation model. *MIS Quarterly*, 25: 351–370.
- Bourreau, M. & Gaudin, G. (2022). Streaming platform and strategic recommendation bias. *Journal of Economics & Management Strategy*, 31(1): 25-47.
- Bourreau M., Moreau F. & Wikström, P. (2022). Does Digitization lead to the homogenization of cultural content? *Economic Inquiry*, 60(1): 427-453.
- Brynjolfsson, E., Hu, Y. & Simester, D. (2011). Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science*, 57(8): 1373-1386.
- Caliendo, M. & Kopeinig, S. (2008). Some practical guidance for the implementation of Propensity Score Matching. *Journal of Economic Surveys*, 22(1): 31-72.
- Cameron A.C & Trivedi, P.K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.
- Chen, P.-Y., Chou, Y.-C. & Kauffman, R. J. (2009). Community-Based Recommender Systems: Analyzing Business Models from a Systems Operator’s Perspective. *HICSS’09*, 1–10.
- Chen, J., Wang, C., Wang, J. & Yu, P.S. (2016). Recommendation for Repeat Consumption from User Implicit Feedback. *IEEE Transactions on Knowledge and Data Engineering*, 28(11): 3083–3097.
- Cowgill, B. (2020). Bias and Productivity in Humans and Machines. Available at SSRN: <https://ssrn.com/abstract=3584916>.
- Dai, H., Chan, C. & Mogilner, C. (2020). People rely less on consumer reviews for experiential than material purchases. *Journal of Consumer Research*, 46(6): 1052–1075.
- 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.
- De, P., Hu, Y. & Rahman, M.S. (2010). Technology usage and online sales: An empirical study. *Management Science*, 56(11): 1930-1945.
- Dzyabura, D. & Hauser, J.R. (2019). Recommending products when consumers learn their preference weights. *Marketing Science*, 38(3): 417-441.
- Fleder, D. & Hosanagar, K. (2009). Blockbuster culture’s next rise or fall: The effect of recommender systems on sales diversity. *Management Science*, 55(5): 697–712.
- Fricke, K.R., Greenberg, D.M., Rentfrow, P.J. & Herzberg, P.Y. (2019). Measuring musical preferences from listening behavior: Data from one million people and 200,000 songs. *Psychology of Music*, 49(3): 371–381.
- Garcia-Gathright, J., St Thomas, B., Hosey, C., Nazari, Z., & Diaz, F. (2018). Understanding and Evaluating User Satisfaction with Music Discovery. *41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 55–64.
- Goldberg, D., Nichols, D., Oki, B.M. & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12), 61–70.
- Greenberg, D.M. & Rentfrow, P.J. (2017). Music and big data: A new frontier. *Current Opinion in Behavioral Sciences*, 18: 50–56.
- He, S., Hollenbeck, B. & Proserpio, D. (2022). The Market for Fake Reviews. *Marketing Science*, 41(5): 871-1027.
- Hennig-Thurau, T. & Houston, M.B. (2019). *Entertainment Science: Data Analytics and Practical Theory for Movies, Games, Books, and Music*, Springer.
- Ho, D.E., Imai, K., King, G., & Stuart, E.A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15:199–236.
- Holtz, D., Carterette, B., Chandar, P., Nazari, Z., Cramer, H., & Aral, S. (2020). The Engagement-Diversity Connection: Evidence from a Field Experiment on Spotify. *21st ACM Conference on Economics and Computation*, 75–76.
- Hong, Y. & Pavlou, P.A. (2014). Product fit uncertainty in online markets: Nature, effects, and antecedents. *Information Systems Research*, 25(2): 328–344.

- Hosanagar, K., Fleder, D., Lee, D. & Buja, A. (2014). Will the global village fracture into tribes? Recommender systems and their effects on consumer fragmentation. *Management Science*, 60(4): 805-823.
- Iwanaga, J., Nishimura, N., Sukegawa, N. & Takano, Y. (2019). Improving collaborative filtering recommendations by estimating user preferences from clickstream data. *Electronic Commerce Research and Applications*, 37: 100877.
- Jannach, D., & Abdollahpouri, H. (2023). A survey on multi-objective recommender systems. *Frontiers in Big Data*, 6: 1157899.
- Kahnx, B., Ratner, R. & Kahneman, D. (1997). Patterns of Hedonic Consumption Over Time. *Marketing Letters*, 8: 85–96
- Knijnenburg B. P., Willemsen M.C., Gantner Z., Soncu H. & Newell C. (2012). Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction*, 22(4–5): 441–504.
- Knox, G. & Datta, H. (2022). Platform-Guided Consumption Similarity. SSRN 4236521
- Krishnan, V., Narayanashetty, P.K., Nathan, M., Davies, R.T., & Konstan, J.A. (2008). Who predicts better? results from an online study comparing humans and an online recommender system. *ACM conference on Recommender systems*, 211–218.
- LaBarbera, P. A. & Mazursky, D. (1983). A Longitudinal Assessment of Consumer Satisfaction, Dissatisfaction: The Dynamic Aspect of Cognitive Process, *Journal of Marketing Research*, 20: 393-404.
- Lambert, D. (1992). Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics*, 34(1): 1-14.
- Lee, D. & Hosanagar, K. (2019). How Do Recommender Systems Affect Sales Diversity? A Cross-Category Investigation via Randomized Field Experiment. *Information Systems Research*, 30(1): 239-259.
- Lee, D. & Hosanagar, K. (2021). How Do Product Attributes and Reviews Moderate the Impact of Recommender Systems Through Purchase Stages? *Management Science*, 67(1): 524-546.
- Leitgöb, H. (2014). Modeling interactions in count-data regression: Principles and implementation in Stata, German Stata Users' Group Meetings, Stata Users Group.
- Liu, C.T., Guo, Y.M., & Lee, C.H. (2011). The effects of relationship quality and switching barriers on customer loyalty. *International Journal of Information Management*. 31: 71–79.
- Loepp, B., Donkers, T., Kleemann, T. & Ziegler, J (2019). Impact of Consuming Suggested Items on the Assessment of Recommendations in User Studies on Recommender Systems. Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19), 6201-6205.
- Long, S. & Freese, J. (2014). Regression Models for Categorical Dependent Variables Using Stata. 3rd Edition, Stata Press, College Station.
- Mariuzzo F. & Ormosi, P. (2020). Independent v major record labels: Do they compete on a level-playing streaming field? SSRN 37229966.
- Mayzlin, D., Dover, Y., & Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 8: 2421–2455.
- Mehrotra, R., McInerney, J., Bouchard, H., Lalmas, M., & Diaz, F. (2018). Towards a Fair Marketplace: Counterfactual Evaluation of the trade-off between Relevance, Fairness & Satisfaction in Recommendation Systems. *27th ACM International Conference on Information and Knowledge Management (CIKM'18)*.
- Mehrotra, R., Lalmas, M., Kenney, D., Lim-Meng, T. & Hashemian, G. (2019). Jointly Leveraging Intent and Interaction Signals to Predict User Satisfaction with Slate Recommendations. *World Wide Web Conference (WWW'19)*.
- Moulton, B.R. (1990). An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units. *Review of Economics and Statistics*, 72(2), 334–338.
- Oestreicher-Singer, G. & Sundararajan, A. (2012). Recommendation networks and the long tail of electronic commerce. *MIS Quarterly*. 36(1): 65-83.
- Pachali, M.J. & Datta, H. (2022). What Drives Demand for Playlists on Spotify? SSRN 4079693
- Papassapa, R. & Miller., K.E. (2007). Relationship quality as a predictor of B2B customer loyalty. *Journal of Business Research*, 60(1): 21-31.

- Pathak, B., Garfinkel, R., Gopal, R.D., Venkatesan, R. & Yin, F. (2010). Empirical Analysis of the Impact of Recommender Systems on Sales. *Journal of Management Information Systems*, 27(2): 159-188.
- Paul, M., Hennig-Thurau, T., Gremler, D.D., Gwinner, K.P. & Wiertz, C. (2009). Toward a theory of repeat purchase drivers for consumer services. *Journal of the Academy of Marketing Science*, 37(2): 215-237.
- Peukert, C., Sen, A. & Claussen, J. (2023). The Editor and the Algorithm: Recommendation Technology in Online News. *Management Science*, doi.org/10.1287/mnsc.2023.4954.
- Rauyruen, P. & Miller, K.E. (2007). Relationship quality as a predictor of B2B customer loyalty. *Journal of Business Research*, 60(1): 21-31.
- Resnick, P. & Varian, H.R. (1997). Recommender systems. *Communications of the ACM*, 40(3), 56–58.
- Schedl, M., Zamani, H., Chen, CW., Deldjoo, Y. & Elahi, M. (2018). Current challenges and visions in music recommender systems research. *International Journal of Multimedia Information Retrieval*, 7: 95–116.
- Senecal, S. & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, 80(2): 159–169.
- Sguerra, B., Tran V.A. & Hennequin, R. (2022). Discovery Dynamics: Leveraging Repeated Exposure for User and Music Characterization. *Proc. 16th ACM Conference on Recommender Systems – RecSys '22*.
- Sim, J., Park, J. G., Cho, D., Smith, M. D. & Jung, J. (2022). Bestseller lists and product discovery in the subscription-based market: Evidence from music streaming. *Journal of Economic Behavior & Organization*, 194: 550-567.
- Sinha, R., & Swearingen, K. (2011). Comparing recommendations made by online systems and friends, mimeo.
- Smith Burton, C. (2021). 'Playola' and Fraud on Digital Music Platforms: Why Legislative Action is Required to Save the Music Streaming Market, *Journal of Business & Technology Law*, 16(2): 9.
- Villermet, Q., Poiroux, J., Moussallam, M., Louail, T. & Roth, C. (2021). Follow the guides: disentangling human and algorithmic curation in online music consumption. *Proc. 15th ACM Conference on Recommender Systems – RecSys '21*, 380-389.
- Waldfoegel J., (2018). *Digital Renaissance. What Data Tell Us about the Future of Popular Culture*. Princeton University Press.
- Way, S.F., Garcia-Gathright, J., & Cramer, H. (2020). Local Trends in Global Music Streaming. *Proceedings of the Fourteenth International AAAI Conference on Web and Social Media (ICWSM 2020)*.
- Wlömert, N., Papies, D. & van Heerde, H.J. (2021). How Curation Shapes Demand for Digital Information Goods: Estimating Playlist Elasticities for Music Streaming Services, <https://sites.insead.edu/facultyresearch/research/file.cfm?fid=68433>
- Wu, L.L., Joung, Y.J. & Chiang, T.E. (2011). Recommendation Systems and Sales Concentration: The Moderating Effects of Consumers' Product Awareness and Acceptance to Recommendations. *HICSS 2011*, 1–10.
- Xiao, B. & Benbasat, I. (2007). E-commerce product recommendation agents: use, characteristics, and impact. *MIS Quarterly*, 31(1): 137-209.
- Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making*, 32(4): 403–414.
- Zeithaml, V.A., Berry, L.L. & Parasuraman, A. (1996). The behavioral consequences of service quality. *Journal of Marketing*, 60(2): 31-46.
- Zhou, R., Khemmarat, S. & Gao, L. (2010). The Impact of YouTube Recommendation System on Video Views. *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*.

Appendix

Table A1 – Correlation table

	CURATION	PLAYLIST	RECO	NOVEL_ARTIST	ARTIST_POPULARITY	FR	US	OTHER_NAT	HIPHOP_RAP	POP	OTHER_GENRE
CURATION	1.0000										
PLAYLIST	0.6139	1.0000									
RECO	0.5591	-0.3112	1.0000								
NOVEL_ARTIST	0.1169	0.1522	-0.0192	1.0000							
ARTIST_POPULAR.	-0.1146	-0.0629	-0.0718	-0.1986	1.0000						
FR	-0.1161	-0.1642	0.0327	-0.1609	0.0340	1.0000					
US	0.0313	0.0700	-0.0358	0.1102	-0.0618	-0.5588	1.0000				
OTHER_NAT	0.1043	0.1234	-0.0040	0.0819	0.0189	-0.6272	-0.2954	1.0000			
HIPHOP_RAP	-0.0679	-0.0877	0.0104	-0.0771	-0.0630	0.2795	-0.0508	-0.2743	1.0000		
POP	0.0235	-0.0132	0.0421	-0.0514	0.1570	0.1381	-0.1649	-0.0042	-0.4792	1.0000	
OTHER_GENRE	0.0412	0.0970	-0.0522	0.1252	-0.0974	-0.4054	0.2140	0.2661	-0.4737	-0.5460	1.0000

Table A2 – Countfit procedure in STATA to test whether a Zero-inflated negative binomial model should be preferred to a Negative binomial model

Regression model	BIC	AIC	Difference	Prefer	Over	Evidence
NBRM	5.917e+06	5.917e+06	163857.467	ZINB	NBRM	Very strong
vs ZINB	5.753e+06	5.753e+06	164130.387	ZINB	NBRM	

Table A3 – Results of the Zero Inflated Negative Binomial model for various levels of artist's popularity

Dependent variable: ORGANIC_STREAMS

	Low popularity artists		Medium popularity artists		High popularity artists	
	Main	Inflate	Main	Inflate	Main	Inflate
CURATION	0.921*** (0.0164)	42.63*** (2.620)	0.677*** (0.00888)	46.62*** (6.159)	0.562*** (0.0292)	17.66*** (2.376)
OFF-PLATFORM	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Constant	0.284*** (0.0129)	0.0429*** (0.00330)	0.413*** (0.0144)	0.0388*** (0.00520)	0.587*** (0.0873)	0.191*** (0.0612)
ln_alpha	5.549*** (0.0533)		4.902*** (0.0498)		3.785*** (0.0213)	
N	512,738		652,369		638,784	

Exponentiated coefficients; robust standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Note: Regressions include control for the type of provider, the music genre, the country of origin, the average number of repeat streams per song (user level), the song popularity, the familiarity of the user with the artist, the song duration, the completion rate of the first stream of the song (user level), and month dummies.

Table A4 – Results of the Zero Inflated Negative Binomial model for different countries of origin

Dependent variable: ORGANIC_STREAMS						
	Domestic		US		Other foreign	
	Main	Inflate	Main	Inflate	Main	Inflate
CURATION	0.616*** (0.00634)	20.72*** (0.428)	0.773*** (0.0137)	24.10*** (0.921)	0.766*** (0.0110)	16.16*** (0.327)
OFF_PLATFORM	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Constant	0.442*** (0.0154)	0.171*** (0.0124)	0.202*** (0.00959)	0.0240*** (0.00179)	0.552*** (0.0302)	0.251*** (0.0181)
ln_alpha	4.167*** (0.0142)		4.613*** (0.0329)		3.964*** (0.0226)	
N	978,812		375,840		449,239	

Exponentiated coefficients; robust standard errors in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.001.

Note: Regressions include control for the type of provider, the music genre, the average number of repeat streams per song (user level), the song popularity, the artist's popularity, the familiarity of the user with the artist, the song duration, the completion rate of the first stream of the song (user level), and month dummies.

Table A5 – Results of the Zero Inflated Negative Binomial model for novel artists

Dependent variable: ORGANIC_STREAMS				
	Artist is novel to the user		Artist is known to the user	
	Main	Inflate	Main	Inflate
CURATION	0.768*** (0.0131)	22.43*** (0.724)	0.687*** (0.00583)	31.72*** (1.192)
OFF_PLATFORM	Ref.	Ref.	Ref.	Ref.
Constant	0.234*** (0.0120)	0.108*** (0.00865)	0.393*** (0.0101)	0.0896*** (0.00578)
ln_alpha	5.481*** (0.0461)		4.246*** (0.0174)	
N	465,679		1,338,212	

Exponentiated coefficients; robust standard errors in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.001.

Note: Regressions include control for the type of provider, the music genre, the country of origin, the average number of repeat streams per song (user level), the artist's popularity, the song popularity, the song duration, the completion rate of the first stream of the song (user level), and month dummies.

Table A6 – Results of the Zero Inflated Negative Binomial model for various levels of artist's popularity

Dependent variable: ORGANIC_STREAMS						
	Low popularity artists		Medium popularity artists		High popularity artists	
	Main	Inflate	Main	Inflate	Main	Inflate
RECO	0.819*** (0.0288)	0.575*** (0.0226)	0.946* (0.0217)	1.220*** (0.0488)	0.803*** (0.0158)	0.725*** (0.0270)
PLAYLIST	Ref.		Ref.		Ref.	
Constant	0.0215*** (0.00289)	1.35e-30*** (3.56e-30)	0.325*** (0.0332)	6.113*** (0.993)	0.0266*** (0.00253)	1.34e-33*** (3.88e-33)
ln_alpha	18.45*** (0.234)		13.92*** (0.103)		6.613*** (0.0428)	
N	296,558		311,949		249,263	

Exponentiated coefficients; robust standard errors in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.001

Note: regressions include control for the type of provider, the music genre, the country of origin, the average number of repeat streams per song (user level), the song popularity, the familiarity of the user with the artist, the song duration, the completion rate of the first stream of the song (user level), and month dummies.

Table A7 – Results of the Zero Inflated Negative Binomial model for different countries of origin

Dependent variable: ORGANIC_STREAMS						
	Domestic		US		Other foreign	
	Main	Inflate	Main	Inflate	Main	Inflate
RECO	0.784*** (0.0159)	0.949* (0.0249)	0.735*** (0.0247)	0.687*** (0.0309)	0.978 (0.0265)	0.970 (0.0232)
PLAYLIST	Ref.		Ref.		Ref.	
Constant	0.0345*** (0.00365)	0.330*** (0.0604)	0.0282*** (0.00400)	6.52e-30*** (1.73e-29)	0.149*** (0.0210)	1.497*** (0.165)
ln_alpha	8.655*** (0.0783)		11.67*** (0.306)		7.308*** (0.133)	
N	413,338		190,184		254,248	

Exponentiated coefficients; robust standard errors in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.001

Note: regressions include control for the type of provider, the music genre, the average number of repeat streams per song (user level), the familiarity of the user with the artist, the song popularity, the song duration, the completion rate of the first stream of the song (user level), and month dummies.

Table A8 – Results of the Zero Inflated Negative Binomial model for artist novelty

Dependent variable: ORGANIC_STREAMS				
	Artist is novel to the user		Artist is known to the user	
	Main	Inflate	Main	Inflate
RECO	0.888*** (0.0291)	0.589*** (0.0258)	0.819*** (0.0134)	1.060** (0.0189)
PLAYLIST	Ref.		Ref.	
Constant	0.0196*** (0.00237)	6.86e-27*** (1.49e-26)	0.0677*** (0.00559)	0.670*** (0.0651)
ln_alpha	13.54*** (0.148)		8.047*** (0.0572)	
N	267,512		590,258	

Exponentiated coefficients; robust standard errors in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.001

Note: regressions include control for the type of provider, the music genre, the country of origin, the average number of repeat streams per song (user level), the artist's popularity, the song popularity, the song duration, the completion rate of the first stream of the song (user level), and month dummies.

Table A9 – Familiarity offered by the different discovery modes

Share within:	Whole sample (%)	Low-popularity artists (%)	Non-domestic (%)	Novel artist (%)
First stream initiated by:				
Algorithmic recommendation	21.8	23.2	20.3	20.2
Playlist	27.2	37.1	35.7	39.7
Organic search	51.0	39.7	44.0	40.1
Total	100.0	100.0	100.0	100.0

Table A10 - Regression on the whole sample, including as a control the log of the number of days since the user joined the platform (Log(REGISTRATION))

DV: ORGANIC_STREAMS		
	Main	Inflate
CURATION	0.715*** (0.00513)	21.50*** (0.261)
OFF_PLATFORM	Ref.	
Log(REGISTRATION)	0.987*** (0.00314)	1.073*** (0.00556)
Constant	0.433*** (0.0139)	0.0665*** (0.00367)
ln_alpha	4.261*** (0.0112)	
N	1,938,351	

Exponentiated coefficients; robust standard errors in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.001

Note: The regression includes control for the type of provider, the music genre, the country of origin, the average number of repeat streams per song (user level), the artist's popularity, the familiarity of the user with the artist, the song popularity, the song duration, the completion rate of the first stream of the song (user level), and month dummies.

Table A11 – Regression on the whole sample, excluding observations in which the number of organic repeat streams exceeds 50

DV: ORGANIC_STREAMS		
	Main	Inflate
CURATION	0.701*** (0.00455)	17.91*** (0.208)
OFF_PLATFORM	Ref.	
Constant	1.013*** (0.0217)	0.192*** (0.00719)
ln_alpha	3.699*** (0.00719)	
N	1,891,294	

Exponentiated coefficients; robust standard errors in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.001

Note: The regression includes control for the type of provider, the music genre, the country of origin, the average number of repeat streams per song (user level), the artist's popularity, the familiarity of the user with the artist, the song popularity, the song duration, the completion rate of the first stream of the song (user level), and month dummies.

Table A12 – Regression on the whole sample with a Probit model

DV: ORGANIC_STREAMS	
	yes/no
CURATION	0.358*** (0.000792)
OFF_PLATFORM	Ref.
Constant	0.611*** (0.00450)
N	1,938,351
Pseudo R2	0.1678
Wald chi (20)	341150.51
Prob > chi2	0.0000
Log pseudolikelihood	-1008522.9

Exponentiated coefficients; robust standard errors in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.001

Note: The regression includes control for the type of provider, the music genre, the country of origin, the average number of repeat streams per song (user level), the artist's popularity, the familiarity of the user with the artist, the song popularity, the song duration, the completion rate of the first stream of the song (user level), and month dummies.

Table A13 – Zero Inflated Ordered Probit model (on-platform curation vs. off-platform information)

DV: ORGANIC_STREAMS		
	Main	Inflate
CURATION	0.398*** (0.00107)	0.411*** (0.00224)
OFF_PLATFORM	Ref.	
Constant		7.49137e+11*** (2.12455e+11)
Cut1	2.705*** (0.0237)	
Cut2	4.310*** (0.0380)	
N	1,938,351	

Exponentiated coefficients; robust standard errors in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.001

Note: The regression includes control for the type of provider, the music genre, the country of origin, the average number of repeat streams per song (user level), the artist's popularity, the familiarity of the user with the artist, the song popularity, the song duration, the completion rate of the first stream of the song (user level), and month dummies.

Table A14 – Zero-Inflated Negative Binomial regression with an alternative dimension of familiarity: music genres (whole sample)

DV: ORGANIC_STREAMS				
	Neither pop nor hiphop_rap		Pop or hiphop_rap	
	Main	Inflate	Main	Inflate
CURATION	0.763*** (0.00996)	21.67*** (0.463)	0.684*** (0.00589)	21.83*** (0.345)
OFF_PLATFORM	Ref.		Ref.	
Constant	0.337*** (0.0132)	0.0647*** (0.00383)	0.565*** (0.0144)	0.159*** (0.00722)
ln_alpha	4.425*** (0.0225)		4.212*** (0.0134)	
N	715,911		1,222,440	

Exponentiated coefficients; robust standard errors in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.001

Note: The regression includes control for the type of provider, the music genre, the country of origin, the average number of repeat streams per song (user level), the artist's popularity, the familiarity of the user with the artist, the song popularity, the song duration, the completion rate of the first stream of the song (user level), and month dummies.

Table A15 – Zero-Inflated Negative Binomial regression with an alternative dimension of familiarity: music genres (curated discoveries only)

DV: ORGANIC_STREAMS

	Neither pop nor hiphop_rap		Pop or hiphop_rap	
	Main	Inflate	Main	Inflate
RECO	0.847*** (0.0220)	0.677*** (0.0215)	0.799*** (0.0137)	1.049** (0.0185)
PLAYLIST	Ref.		Ref.	
Constant	0.0744*** (0.00941)	0.427*** (0.0709)	0.0898*** (0.00727)	0.888 (0.0801)
ln_alpha	10.22*** (0.137)		7.820*** (0.0702)	
N	379,644		572,584	

Exponentiated coefficients; robust standard errors in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.001
 Note: The regression includes control for the type of provider, the music genre, the country of origin, the average number of repeat streams per song (user level), the artist's popularity, the familiarity of the user with the artist, the song popularity, the song duration, the completion rate of the first stream of the song (user level), and month dummies.

Table A16 – Results with songs that have been organically relistened at least once (Negative Binomial Regression)

Dependent variable: ORGANIC_STREAMS

	All discoveries (on- and off-platform)	On-platform discoveries only
CURATION	0.807*** (0.00441)	
OFF-PLATFORM	Ref.	
RECO		0.931*** (0.00915)
PLAYLIST		Ref.
Constant	1.546*** (0.0245)	1.097* (0.0466)
ln_alpha	1.424*** (0.00213)	1.582*** (0.00477)
N	615,944	138,664

Exponentiated coefficients; robust standard errors in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.001
 Regressions include control for the type of provider, the music genre, the country of origin, the average number of repeat streams per song (user level), the artist's popularity, the familiarity of the user with the artist, the song popularity, the song duration, the completion rate of the first stream of the song (user level), and month dummies.