HOW CURATORS MODERATE THE EFFECT OF VARIETY ON PLAYLIST SUCCESS

A quantitative analysis in the music streaming industry

Name:	Lieke Jacobs
Program:	MSc Marketing Analytics
Supervisor:	Dr. Hannes Datta
Second assessor:	Nazli Alagöz
Date:	January 6th, 2020

MANAGEMENT SUMMARY

In this thesis, we studied the supply of variety in playlists on Spotify as well as how this variety affects playlist success. Playlists have become an increasingly important medium for music discovery. However, most current work on automatic playlist generation (APG) focus on approaches using content-based similarity which maximize semantic cohesion. While transition effects play an important role in modeling playlist continuity and cohesion plays a significant part in the assessment of playlist quality, the importance of diversity is often overlooked. Moreover, these approaches typically rely on predefined features to measure variety and thus assume that all of these features are predictive of the users' taste. Likewise, these studies often assume that the similarity between consecutive songs gives a more accurate representation of cohesion than the overall similarity of tracks within a playlist.

We aim to overcome these problems and test these assumptions by collecting observational data related to playlist success on Spotify using the Chartmetric API. We used two multivariate time series regression models to test the effect of within-playlist variety and transition variety on playlist success while differentiating between the types of playlist curators to account for differences in consumers' tastes. Variety was measured along twelve different dimensions including eleven audio features (acousticness, danceability, energy, speechiness, instrumentalness, key, liveness, loudness, mode, tempo, and valence) and the number of unique artists. We distinguished between 5 types of playlist curators: Spotify's human curators, Spotify's AI, Spotify's personalized AI, major labels, and independent curators.

The results indicated that within-playlist variety and transition variety have a significant effect on playlist success and that this effect differs per playlist curator. In contrast to the dominant view, the results indicated that consumers had different preferences for within-playlist variety and transition variety. Furthermore, the variety dimensions had different effects within each model including positive, negative, non-linear and insignificant effects. The observed patterns show how some of our expectations are confirmed, while others outline avenues for future research opportunities. This thesis shows that the relationship between variety and playlist success is much more complex than existent literature suggests. While the practical implications of this research are therefore limited, this study provides useful new insights into the understanding of consumer preferences for variety in music.

Table of Contents

1 INTRODUCTION

Music recommendation systems (MRSs) have gained substantial interest in recent years due to the rapid growth of on-demand streaming services such as Spotify, Deezer, Amazon Prime Music, Apple Music, and SoundCloud (Aguiar & Waldfogel, 2018). According to Nielsen (2018a) US on-demand audio streams increased by 59% throughout 2017, resulting in a 12.5% growth of overall music consumption. In the first half of 2018, US on-demand audio streaming volume had exceeded 268 billion, an increase of roughly 45% (Nielsen, 2018b). In revenue, the RIAA reports that 75% of the US music industry's total revenue in 2018 belonged to streaming services, resulting in a total streaming revenue of $7.4 billion (RIAA, 2019).

Initially, streaming platforms differentiated themselves based on their available music libraries. As consumers are fond of variety when it comes to consumption goods, streaming platforms with larger music libraries were more successful (Datta, Knox, & Bronnenberg, 2018; Kahn, 1995; Lee, Bare, & Meek, 2011; Schedl, Zamani, Chen, Deldjoo, & Elahi, 2018). Over the years, however, the available music libraries on these platforms have grown to include tens of millions of songs (Schedl et al., 2018). Consequently, the point of differentiation shifted from a focus on song availability to the extent to which the platform is successful in reducing choice overload for their consumers through the use of MRSs (Bonnin & Jannach, 2015). In other words, music streaming platforms are nowadays evaluated based on their ability to assist consumers in exploring songs, e.g., through the creation and sharing of playlists (Hogan, 2015; Slaney & White, 2006). According to the Music Business Association (2016), playlists account for 31% of music listening time among listeners in the USA, whereas albums and single tracks account for 22% and 46% respectively. Despite the importance of playlists in music discovery, little progress has been made in MRS research concerning the drivers and evaluation of playlist success.

In a recent study, Datta et al. (2018) investigated the relationship between the adoption of music streaming platforms and music consumption on an individual level through measuring the quantity, variety, and discovery of music on Spotify. Due to music streaming platforms charging a subscription fee for access to the entire library instead of a fee per song, the monetary cost of marginal variety on these platforms is practically zero (aside from search costs). As a result, Datta et al. (2018) found that users listened to more unique artists, songs, and genres after the adoption of Spotify. Furthermore, users spent less time listening to superstars and personal favorites and thus consumed music across a wider set of songs. Lastly, new music discoveries were consumed more after the adoption of Spotify.

Thus, streaming providers lower search costs for their consumers through the recommendations of playlists. This lower marginal cost for variety is expected to increase the demand for variety. Yet, current automatic playlist generation (APG) and automatic playlist continuation (APC) approaches focus on semantic cohesion as the main metric for playlist quality and thereby assume that users prefer playlists where tracks are similar (Kamehkhosh, Jannach, & Bonnin, 2018; Knees, Pohle, Schedl, & Widmer, 2006; McFee & Lanckriet, 2012; Schedl et al., 2018). As a result, MRSs often lack diversity and produce unsatisfactory recommendations (Schedl et al., 2018). Furthermore, these APG approaches typically use a set of predefined audio features and thus assume that all of these features are predictive of consumer preferences. In practice, however, it has been shown that the perception of the audio features is subjective (Schedl et al., 2018). Thus, the question remains what varieties in audio features are actually predictive of consumer preferences, and how these preferences differ across these features. Another limitation of previous research is that they focus on APG approaches to maximize cohesion within playlists, but do not observe how variety in playlists is supplied in reality.

From a business perspective, defining which aspects of variety drive playlist success would support streaming providers in building playlists that offer a better fit to the consumers' preferences for variety. This applies to the creation of both handcrafted playlists and algorithm-generated playlists. Furthermore, these attributes may provide insight to record labels or publishing companies into the preferences of consumers for album compositions. Concerning consumers, enhancing the variety within playlists may benefit the individual through offering a better fit to their preferences (Boughanmi, Ansari, & Kohli, 2018). On a different note, listening to diverse music may improve the mental health of consumers as it promotes feelings of social affiliation, identity, and social relatedness (Schäfer et al., 2016; Schäfer & Mehlhorn, 2017). This in turn, may reduce conflict. From a society perspective, enhancing the musical diversity within playlists may lead to consumers discovering more unique artists which in turn would allow the market to support more artists.

In this thesis, we provide a first look at how variety in acoustic features and artists is supplied on Spotify. We chose Spotify because it is the market leader in terms of subscriber share of music streaming services with 36% in the first half of 2019 (Statista, 2019). The streaming platform hosts over 50 million tracks and 3 billion playlists (Spotify, 2019). Furthermore, we investigate how variety in these dimensions affects playlist success. Here, we make a distinction between variety in consecutive tracks (transition variety) and variety within the playlist overall (within-playlist variety). Additionally, we investigate how these effects differ per type of playlist curator to account for differences in listeners' tastes and perception of variety. Concerning the playlist curator, we differentiate between Spotify's human curators, its AI, its personalized AI, major labels, and independent curators. The effect of variety in acoustic features has recently been explored in relation to album success (Boughanmi et al., 2018), however, it has not yet been applied to playlists. While both albums and playlists are sequences of tracks intended to be listened to together, albums often focus on a particular artist. More importantly, playlists, unlike albums, are dynamic in nature; their composition changes frequently as a result of track additions and removals. Thus, it is unclear whether the findings in Boughanmi et al. (2018) apply to playlists with variety in artists and where the level of variety is dynamic. As opposed to previous research related to APG, this thesis will not use pre-defined features combined into a single variety score. Instead, we test the significance of each dimension in a multivariate time series regression model. With regards to the data, we obtain a list of the top 1 million playlists using the API of Chartmetric[1]. For each playlist, the playlist characteristics such as the content type and owner name are recorded. Furthermore, for the top 1000 playlists, we obtain the daily number of followers as well as their current and historic track information including detailed track characteristics (e.g., the audio features, artist name, and release dates).

This thesis is organized as follows. First, we present an overview of existing literature related to the drivers and evaluation of playlist success. Next, our research method and expectations for the results are presented. Then, an overview of the data collection, preparation and operationalization is given. Subsequently, descriptive statistics will be presented as well as the model specification. Finally, we discuss our results and conclusions.

[1] https://api.chartmetric.io/apidoc/

2 LITERATURE REVIEW

In this chapter, an overview of existing literature is given that relates to the drivers and evaluation of playlist success. First, previous research on APG and the implications of semantic cohesion as a metric of playlist quality is outlined in paragraph 2.1. In paragraph 2.2., we look at the importance of diversity in the evaluation of APG approaches based on qualitative research. Finally, the last paragraph gives a critical overview of these streams of research and discusses the contributions of this thesis.

2.1 SEMANTIC COHESION

While the link between variety and playlist success is not as clearly defined in the literature, it is suggested that playlists with more variety are less successful. McFee and Lanckriet (2011) categorize the main approaches to APC evaluation as human evaluation, sequence prediction, and semantic cohesion. Human evaluation directly measures human response through asking test subjects to rate the quality of playlists generated by an algorithm. While this approach comes closest to measuring user satisfaction, it suffers from problems of scale and reproducibility. Sequence prediction compares the algorithm's prediction of the next song in a playlist to the observed ranking of songs in an existing playlist. While this method is more scalable, it is flawed by the assumption that an inaccurate prediction is a bad recommendation. Finally, semantic cohesion as a metric of playlist quality is easily measurable and reproducible. In McFee and Lanckriet (2011), semantic cohesion is not formally defined. Generally speaking, it refers to the coherence of tracks within a playlist based on some semantic tag. Lee et al. (2011) mentioned that similarity is important and tracks not complying negatively affect the flow of the playlist. Most researchers contributing to this stream of literature studied the effective generation of playlists using content-based methods. Knees et al. (2006) presented a playlist generation approach using a combination of audio-based similarity and artist similarity measures. This approach advanced automatic playlist generation and improved playlist quality in stylistic consistency while focusing on transition variety. Similarly, Maillet, Eck, Desjardins, and Lamere (2009) demonstrate a method for learning song transition probabilities from track similarity based on radio station playlists. The use of content-based similarity to generate playlists with a certain target characteristic is also discussed in McFee and Lanckriet (2012), Panteli, Benetos, and Dixon (2016), Ragno, Burges, and Herley (2005), Reynolds, Barry, Burke, and Coyle

(2007), and Slaney and White (2006). Examples of target characteristics include a single seed song, a start and end song, maximum similarity between consecutive songs, or explicit constraints such as musical attributes. Recently, Kamehkhosh et al. (2018) found that while transition effects are an indicator of playlist quality, homogeneity of musical features within the playlist overall play a larger role. Furthermore, Vall, Quadrana, Schedl, Widmer, and Cremonesi (2017) shows that transition effects only significantly impact playlists with tracks from the long tail.

2.2 DIVERSITY

Another school of thought argues that current APG approaches overlook the importance of diversity (Schedl et al., 2018). This stream of research is based on the theory that consumers prefer variety in their choices of services or goods (Kahn, 1995; Kahneman & Snell, 1990; McAlister & Pessemier, 1982; Simonson, 1990). This variety-seeking behavior can have various motivations such as an internal desire for change, preference uncertainty, or external factors (Kahneman & Snell, 1990; Simonson, 1990). McAlister (1982) further argues that once a consumer reached an optimal level of an attribute, he may choose to satiate a different attribute afterwards. The author maintains that consumers seek a balance of attributes to maximize their utility. The desire for variety may be larger in certain conditions such as 1) when consumers select items simultaneously as a bundle, rather than as a sequential series of decisions (Adomavicius, Bockstedt, & Curley, 2015) or 2) when the product's attributes interact with the senses (Inman, 2001). As playlists can be seen as a "bundle of tracks" which interact with the sense of hearing, these findings would suggest that consumers' preferences for variety are greater in playlists than in a general consumption good. The conclusion that emerges from these studies is supported by Slaney and White (2006) who found that users of WebJay, a web playlist community, indeed have an interest in diverse music. Lee et al. (2011) further supports the idea that consumers prefer diversity through conducting a qualitative user study in which participants stated that they were bothered by the lack of variety both within a playlist and between consecutive songs. A brief overview of relevant studies is shown in Table 1.

Table 1

COMPARISON OF THIS THESIS WITH EXISTING STUDIES RELATED TO THE DRIVERS AND EVALUATION OF PLAYLIST SUCCESS

	Effect of	Effect on
Slaney and White (2006)	Variety in genre (Tzanetakis' GenreGram)	Playlist volume
Boughanmi et al. (2018)	Genre (Discogs), album tags (Last.fm), and acoustic features (Spotify API)	Album success
McFee and Lanckriet (2012)	Similarity in audio (Echo Nest), lyrics (musiXmatch), social tags (Last.fm), transition effects	MRS accuracy
Kamehkhosh et al. (2018)	Similarity in audio (Spotify API), artist diversity, transition effects	Playlist quality
Knees et al. (2006)	Similarity in audio (MFCC) and artists	Playlist consistency
This study	Variety in artists and acoustic features (Spotify API)	Playlist success

2.3 CONTRIBUTIONS OF THIS THESIS

Both of these schools of thought have tended to focus on finding a relationship between variety and consumption rather than on the definition of "variety" and how it exists on music streaming platforms. The first school of thought uses content-based methods to compute a single similarity score, but at the same time uncertainty arises regarding the interpretation of this score. As an illustration, one pair of tracks is very similar based on some attributes, but dissimilar on others, while another pair of tracks are moderately similar on all attributes. Using a single similarity score may result in both pairs having a comparable similarity, whereas in reality the pairs are very different. The second school of thought acknowledges that consumers seek variety in attributes. However, it does not indicate what these attributes are.

Thus, the main contribution of this thesis will contain a descriptive overview of the supply side of variety through several steps. First, the variety will be measured using several dimensions, rather than a single score, to identify variety more accurately. An extensive explanation of these dimensions is given in Chapter 3. Furthermore, as mentioned in paragraph 2.1, many studies assume that transition variety is a good measure for semantic cohesion, however, recent studies argue that the order of tracks within a playlist does not always matter. Thus, the focus on transition variety instead of within-playlist variety to measure semantic cohesion may not be justified for playlists in general. To verify this claim, this thesis measures variety both between sequential tracks and within the entire playlist. Since musical diversity is a rather unexplored topic, it is unclear which measure of variety will have the best fit. Therefore, this study will compare the results of several measures, such as the median absolute deviation centered on the mean, median absolute deviation centered on the median, interquartile range, and coefficient of variation. Next, the within-playlist variety is plotted over time to show changes in variety as a result of playlist updates as opposed to previous studies which assumed a static supply of variety. Furthermore, as the perception of variety differs person-to-person, it would be of interest to learn how consumers of different types of playlist curators perceive and/or appreciate variety differently. Thus, rather than solely measuring the effect of variety on playlist success, this thesis will also test whether this effect differs depending on the curator of the playlist.

The gap regarding the definition of variety is not sufficient to explain why there is discrepancy in the expectations regarding the relationship between variety and playlist success. In an attempt to reconcile these differences, Schedl et al. (2018) claim that, as music is often consumed passively, "not skipping a song" may be wrongly interpreted as a positive signal in inferring consumer preferences for music. Ultimately, consumers may choose playlists according to certain characteristics, but in doing so, are constrained to the availability of playlists. Hence, research studies that only measure the number of streams of a playlist may overestimate the consumer preferences for content-based similarity within playlists (including transition effects). It would thus be of interest to find a different measure for playlist success which better represents consumer preferences than the number of streams of a playlist or its skip-ratio. When a Spotify user follows a playlist, the playlist is saved to his side bar and he will receive notifications when songs are added to the playlist. Following can thus be seen as a more active way of expressing engagement with the playlist as opposed to passive listening. According to Joven (2018), a follower is a vote of long-term confidence. Based on this article, the number of followers is a more meaningful metric to measure playlist success than the number of streams as used in previous studies.

Lastly, this thesis builds upon the research study by Boughanmi et al. (2018). The article studied the success of albums over time in relation to several musical characteristics, such as genre, energy, or valence. The authors find that, e.g., rock music had its peak in the 60s and since declined in popularity until making its comeback in the 2010s. In addition, the authors show that loud music was very popular in the 80s and 90s, but diminished in popularity since. In contrast, Serrà, Corral, Boguñá, Haro, and Arcos (2012) found that popular songs have become louder over the years. Although this body of literature does not prioritize a variety perspective, it does provide an initial idea of which dimensions are relevant with regards to measuring variety. For the purpose of this thesis, the variety in audio attributes as well as the variety in artists will be taken into account. The latter is taken into account to test the generalizability of research on album success to playlist success. Furthermore, this thesis challenges the assumptions made by Boughanmi et al. (2018) which state that they did not use playlist data due to it being difficult to obtain and it being inconsistent in objectivity of the measures. The first issue is addressed through collecting playlist data using a modern music data tool, Chartmetric. The second problem is solved by using a more objective measure for playlist success, namely its number of followers, rather than a subjective measure such as album scores used by Billboard magazine (which are used in Boughanmi et al. (2018)). In addition, the article briefly investigated the relationship between variety and album success over time. However, the article categorized variety in “low”, “moderate”, and “high” levels (which were undefined) which blurs the practical implications of their results. Therefore, this thesis will use a numerical measure of variety to make the results more accessible. Finally, Boughanmi et al. (2018) aggregated time to 10-year intervals. Since this thesis focuses on playlists which are dynamic in nature, daily intervals are used instead to capture changes in variety as a result of playlist updates.

In conclusion, previous quantitative studies on APG focus on semantic cohesion as a metric of playlist quality. These studies often do not make a clear distinction between cohesion in consecutive tracks and within a playlist overall. Furthermore, qualitative research suggests that these APG approaches overlook the consumers’ preferences for diversity. Both streams of research fail to define the dimensions of variety. In the next chapter, we provide a more extensive overview of the variety dimensions and formulate our expectations.

3 CONCEPTUAL FRAMEWORK AND EXPECTATIONS

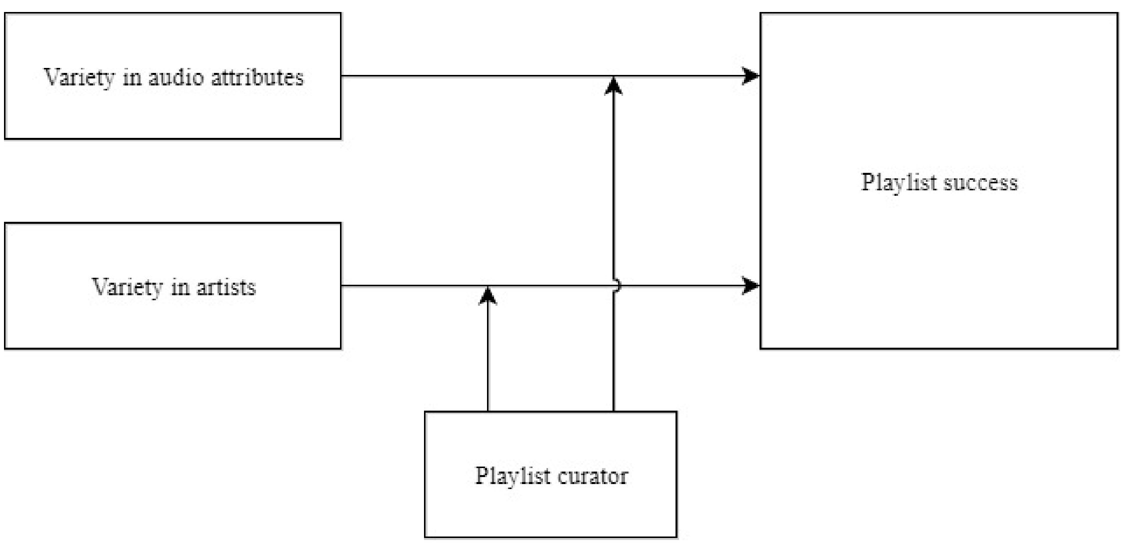
This chapter presents the proposed conceptual framework and expectations of this thesis. First, the conceptual model is shown in paragraph 3.1. In paragraph 3.2, we formulate our expectations of the effects of variety on playlist success and how the playlist curator moderates this effect.

3.1 CONCEPTUAL MODEL

This thesis investigates the relationship between variety and playlist success, as well as the influence of the playlist curator on this relationship. The variety dimensions include acoustic features, like previous studies, and variety in artists. Figure 1 shows the proposed conceptual framework. The definitions of the variables and the expected relations, as are depicted in the conceptual model, will be discussed in the next section.

Figure 1

THE PROPOSED CONCEPTUAL FRAMEWORK



3.2 EXPECTATIONS

In this subchapter, we formulate our expectations with regard to the relationship between variety and playlist success. In paragraph 3.2.1, we look at the twelve different dimensions of variety and how we expect these dimensions to relate to playlist success. The final paragraph gives an overview of the different types of playlist curators and explains how consumers of these curators are expected to have different preferences for variety.

3.2.1 *Variety*

Individuals have a tendency to seek variety, i.e., diversity in their choices of services or goods (Kahn, 1995; Kahneman & Snell, 1990; McAlister, 1979; McAlister & Pessemier, 1982; Simonson, 1990). Moreover, the discrepancy hypothesis in psychology suggests that consumers like new things that are sufficiently different, but not too different, from familiar ones (Haber, 1958). Like Haber (1958), Lee et al. (2011) found that although similarity seems to positively affect the user rating of a playlist, higher similarity does not always correspond to higher rated playlists. Furthermore, several research studies stated that consumers may foresee that listening to their favorite song too much will cause reduced marginal pleasure or even discontent (Inman, 2001; Kahn, 1995; Wang, Sun, & Keh, 2013). Thus, consumers may seek variety in playlists to savor their preferred option, i.e., protect them from oversaturation with favorite choices. In another scenario, consumers may be content with their choices, but seek variety to satisfy a desire for novelty, complexity, or excitement (Kahn, 1995). Lastly, a study by Slaney and White (2006) showed that there is a positive correlation between the number of tracks in a playlist and its variety. This suggests that as consumers listen to more tracks, their preference for variety grows.

Based on the findings discussed in the previous paragraph as well as paragraph 2.2, we expect a positive association between variety and playlist success in general. However, some dimensions of variety may have a different relation to playlist success. Table 2 provides an overview of the acoustic attributes we take into account when measuring variety. As stated in the literature review, transition variety and within-playlist variety are both used as a metric for semantic cohesion, but there is no clear distinction between its effects in existent literature. Due to this limited availability of empirical research, we expect both measures to have the same type of relationship with playlist success. Thus, the next sections will discuss the expected associations between the dimensions of variety and playlist success for both measures.

Table 2

DESCRIPTIONS OF ACOUSTIC FEATURES

Variable	**Description**
Acousticness[a]	A confidence measure of whether the song is acoustic
Danceability[a]	The suitability of a track for dancing based on a combination of music elements including tempo, rhythm stability, beat strength, and overall regularity
Energy[a]	The perceptual measure of intensity and activity

Variable	Description
Instrumentalness[ab]	Prediction of whether a track contains no vocals
Key	The estimated overall key of the section
Liveness[a]	The presence of an audience in the recording
Loudness	The overall loudness of the track in decibels (dB)
Mode	The modality (major or minor) of a track
Speechiness[ab]	The presence of spoken words in a track.
Tempo	The overall estimated tempo of the track in beats per minute (BPM)
Valence[a]	The musical positiveness conveyed by a track

Notes. Adapted from Spotify. Retrieved from https://developer.Spotify.com/documentation/web-api/reference/tracks/get-audio-analysis/ and https://developer.Spotify.com/documentation/web-api/reference/tracks/get-audio-features/

[a]These dimensions are measured on a scale from 0 to 1. The closer the value is to 1, the higher the confidence the track loads highly onto this attribute. For example, a danceability score of 1 means the song is very suitable for dancing.

bThe instrumentalness and speechiness dimensions both measure the presence of vocals. The difference lies in that instrumentalness strictly measures vocal content (including non-word vocalizations). Speechiness, on the other hand measures the extent to which the audio is exclusively speech-like. The two dimensions are thus indeed related, but they penalize the presence of words differently.

Loudness, tempo, danceability, and energy

Under the assumption that consumers' preferences for playlists are similar to those for albums, uniformly loud playlists can irritate a listener (Boughanmi et al., 2018). Similarly, although consumers have a preference for highly danceable and energetic songs as shown by the distribution curve of these dimensions[2], listening to a series of such songs may become tiresome for listeners as it is often associated with a fast tempo (Boughanmi et al., 2018). Therefore, we expect variety in these dimensions to have a positive effect on playlist success.

Valence, mode, and speechiness

Gatewood (1927) found that consumers prefer music that generates stronger emotions. In Western cultures, people consume music specifically because of its emotion-inducing and mood-regulating properties (Knobloch & Zillmann, 2002; Panksepp, 1995; Sloboda, 1992). Furthermore, Boughanmi et al. (2018) indicated that listeners prefer an emotional balance within an album. Several research studies, e.g., Scherer (2004) and Zentner, Grandjean, and Scherer (2008) support this claim. Likewise, Koelsch, Fritz, von Cramon, Müller, and Friederici (2006) found that although sad music evokes sadness, it is more enjoyable as it activates other positive emotions and has a greater aesthetic appeal

[2] https://developer.Spotify.com/documentation/web-api/reference/tracks/get-audio-features/

Furthermore, several research studies stated that consumers prefer happy over sad-sounding music (Gosselin et al., 2005; Hunter, Schellenberg, & Schimmack, 2008, 2010; Husain, Thompson, & Schellenberg, 2002; Schellenberg, Peretz, & Vieillard, 2008; Thompson, Schellenberg, & Husain, 2001). However, the preference for happy- rather than sad-sounding music reduces as consumers listen to a series of happy songs, as well as when listeners are in a negative mood state or are tired (Ladinig & Schellenberg, 2012). The first part of this statement is supported by Kahn and Isen (1993) who found that positive affect can increase variety-seeking behavior. Typically, tracks with a high valence and major mode sound more joyful, while tracks with low valence and minor mode sound more somber. According to Hunter et al. (2010) these associations extend beyond listeners' perception of emotions to actual feelings. Additionally, a track's lyrics can convey emotions. The speechiness dimension measures the presence of spoken words in tracks[3]. As a track's lyrics can strengthen its ability to convey emotions (Boughanmi et al., 2018), higher levels of speechiness of a track may indicate that the track induces more emotion. Thus, the valence, speechiness, and mode dimensions of variety each capture an aspect of the track's intended emotions. In summary, consumers prefer music featuring stronger emotions (Gatewood, 1927), as well as variety in these emotions (Kahn & Isen, 1993; Ladinig & Schellenberg, 2012) to create an emotional balance (Boughanmi et al., 2018). Thus, consumers appreciate variety in these dimensions, but only to a certain extent. Hence, we expect variety in these dimensions to have an inverse U-shaped relationship with playlist success.

Keys, instrumentalness, and acousticness

With regards to keys, Boughanmi et al. (2018) found that consumers generally prefer variety, however, recently, successful albums have songs with low variety in the keys. Their results further indicated the same development occurred when looking at the instrumentalness dimension. Higher levels of variety in instrumentalness were previously related to higher levels of album success, however, lower levels of variety on this feature appeared more appealing recently. In contrast, albums with a high level of variety in the acousticness attribute of its songs have grown to be more popular recently. Thus, Boughanmi et al. (2018) suggest that the association between these dimensions and playlist success changes over time. While it is reasonable that consumer preferences for music change over time, it seems odd that these

[3] https://developer.Spotify.com/documentation/web-api/reference/tracks/get-audio-features/

preferences change abruptly as described in their study. Instead, we expect that the authors failed to recognize a non-linear relationship. Specifically, we expect consumers to have a high appreciation for low and high levels of variety in key, instrumentalness, and acousticness, but a rather low appreciation for moderate variety in these dimensions. Thus, we expect variety in these dimensions to have a U-shaped relationship with playlist success.

Liveness

This dimension measures the likelihood of there being an audience in the recording. Due to the less controlled environment and thus less isolation between instruments in live recordings, this dimension, in a way, captures the audio quality of a track. Plambeck (2010) argued that the sound quality decreased in the mobile age as well as that consumers happily traded sound quality for portability. However, recently, Lunny (2019) stated that due to new technological developments and an increase in the number of people buying high-end audio devices, demand for high quality audio streaming is growing. Thus, a shift is occurring in which consumers prefer high quality audio (and thus studio recordings) more and more over live recordings. Due to these recent developments it is difficult to determine whether consumers still appreciate variety in this dimension or not. Boughanmi et al. (2018) argued that variety in liveness is preferred in some genres, e.g., rock. However, this genre has lost popularity in the last decades (Boughanmi et al., 2018). In general, we expect variety in liveness to have a negative effect on playlist success due to consumer preferences shifting towards higher levels of audio quality.

Artists

Since the adoption of streaming technology, consumers listen to more unique artists (Datta et al., 2018; Hagen, 2015). As mentioned in Chapter 1, playlists dominate the way to explore music on streaming platforms and have a higher music listening share than albums (Music Business Association, 2016). The main difference between playlists and albums is that albums often feature a single artist. Based on these findings, we expect that variety in a playlist's number of unique artists has a positive effect on its success.

The expected associations between the dimensions of variety and playlist success are summarized in Table 3. However, we would like to emphasize that our expectations are constrained by the limited available literature on the effects of the dimensions of variety on playlist success. Thus, it is likely that our expectations will be proven wrong in the following of this thesis. Therefore, despite expecting positive and negative associations between some dimensions of variety and playlist success, we test for non-linear effects in all dimensions.

Table 3

EXPECTED ASSOCIATIONS BETWEEN VARIETY AND PLAYLIST SUCCESS

Expected association	**Dimensions**
Positive	Artists, danceability, energy, loudness, tempo
Negative	Liveness
U-shape	Acousticness, instrumentalness, key
Inverse U-shape	Mode, speechiness, valence

3.2.2 Playlist curator

As mentioned in Chapter 1, the perception of acoustic features is subjective. Thus, the term variety can have different meanings for different listeners, e.g., variety in genres or artists, or lyrics (Schedl et al., 2018). As music streaming providers now differentiate themselves by their ability to assist consumers in exploring songs through the creation and sharing of playlists (Hogan, 2015; Slaney & White, 2006), it would be interesting to study how these providers perceive variety. However, since this study focuses on a single music streaming platform (Spotify), we will test how consumers of different types of playlist curators value variety differently instead. We will make a distinction between Spotify's human curators, its AI, its personalized AI, major label curators, and independent curators.

In an interview, Athena Koumis, a Spotify editor, stated that to them [Spotify] the order of tracks within a playlist is very important and that transitions play a key role in making sure that the playlist flows well (Koumis, 2017). Furthermore, she stated that they [Spotify] always try to make more playlists to cater to different audiences. Thus, consumers of Spotify playlists are expected to satisfy their need for variety across playlists rather than within. In other words, they are expected to have a lower appreciation for variety within playlists. Hence, we expect that the relationship between variety and playlist success weakens when the curator is Spotify. Moreover, when this playlist is generated by the Spotify AI, this relationship is expected to weaken further as these playlists typically focus on a certain target characteristic (such as a specific artist). Consumers of these playlists likely have a preference for this specific characteristic and therefore will not appreciate deviation (i.e., variety) from it as much as consumers of other playlist curators. A further distinction is made for Spotify's personalized playlists. While these playlists are also generated by an algorithm, they are different for every user. Consumers of these playlists are expected to appreciate variety more than consumers of Spotify's human curated or AI-generated playlists as the tracks in the playlist are tailored to them.

Major labels, on the other hand, are expected to prioritize placements of their own tracks (and thus profits) in the creation of playlists. To maximize profits, major labels are more likely to include popular tracks that are less specific in taste and therefore cater to a wider audience. Thus, consumers of such playlists are expected to have a rather "neutral" taste with no particular interest in variety. Thus, we expect that the relationship between variety and playlist success weakens when the playlist is owned by a major label. However, this effect is expected to be smaller than that of Spotify. Moreover, these consumers might be a fan of a particular artist that is signed with the major label. This would result in a particularly weak or even negative association between the variety in artists and playlist success when the playlist is owned by a major label.

Independent playlist creators (including indie labels and Spotify users), on the other hand, are expected to focus more on the playlist as a whole and thus put a higher importance on variety. Indie labels, in comparison with major labels, typically put less creative control on artists they sign with and are more dedicated to their unique styles (McDonald, 2019). Thus, we expect these playlists to feature less 'generic' songs and more diversity. Likewise, Spotify users are more likely to incorporate variety in their playlists as they are less concerned with featuring a particular artist because of a contract they may have with them. Typically, playlists in this category are less popular as they offer a more diverse set of tracks which may not cater to the average taste. Consumers of these playlists are thus expected to have a higher appreciation for variety in playlists. A possible explanation for this is that consumers of these playlists have a greater interest in discovering new music to develop their own taste. All in all, we expect that the impact of variety on playlist success is greatest when the playlist is owned by an independent creator.

4 DATA

This chapter describes the data collection, preparation and exploration processes. First, the raw data collection is described in paragraph 4.1. Next, the data parsing process is outlined in paragraph 4.2. In paragraph 4.3, the data cleaning, variable operationalization, and data aggregation steps are discussed. We present an overview of the supply of variety in paragraph 4.4. Finally, paragraph 4.5 provides an overview of the descriptive statistics of our final data sets.

4.1 DATA COLLECTION

With the purpose of investigating the proposed problem and expectations, this thesis will make use of an observational research approach. Using the API of Chartmetric.com, we aimed to obtain a list of the top 1.1 million playlists on Spotify (ranked by number of followers) along with the playlists' characteristics. These characteristics include information on the number of followers and listeners at data retrieval, whether the playlist was editorial (i.e., owned by Spotify), whether the playlist was personalized, and whether the playlist tends to feature catalogue (i.e., older than 18 months) or frontline tracks. We also observe the name of the playlist (e.g., "Today's Top Hits", "Running Playlist") and the name of the playlist owner.

To investigate our research problem, we select the top 1000 playlists. These playlists account for 52% of the total number of followers. Focusing on the top 1000 playlists is important because these playlists have the most followers. As a result, there is more variety in the number of followers of these playlists over time relative to playlists that have a consistently low number of followers. Thus, these playlists are more responsive to changes in variety as opposed to a random sample. We then obtain, for each of these playlists, data on all songs that have been listed in these playlists between January 8th, 2016 and September 27th, 2019. For each song, we observe its current position within the playlist as well as its historic positions, when it was added to the playlist, and, if applicable, when it was removed. The data also provides detailed track characteristics (e.g., the associated label, release date, as well as the acoustic features). Lastly, we collected the daily number of followers for each of these playlists in the period of November 3rd, 2016 until October 8th, 2019.

4.2 DATA PARSING

The raw data obtained from the Chartmetric API is stored in 3 separate JSON files. These files will be referred to as the playlists data, the placements data, and the followers data. The playlists data includes the playlist-level data of the top 1.1 million playlists. The placements data consists of the top 1000 playlists' current and historic track data. Finally, the followers data includes the daily follower changes of the top 1000 playlists.

The JSON files were parsed as completely as possible into fitting data structures which were then written into CSV files. In the parsing of the placements data, a data error was found. For some tracks in a playlists, we observe position changes before the track was added to the playlist. This indicates that the 'added_at' field of the playlist-track object is incorrect. For the placements data frame, we instead set the date the track was added to the playlist as the earliest date in a list of the value of the 'added_at' field and the dates of the position changes. Furthermore, to reduce data redundancy, we created a separate data frame ('positions') for the position data. Here, all historic and current position stats are parsed from the raw placements data to retrieve a dataset with 5 variables: playlist id, track id, the position of the track, the date the track was added to this position, and the date it was removed from it.

4.3 DATA PREPARATION AND VARIABLE OPERATIONALIZATION

This section outlines the data cleaning, variable operationalization, and data aggregation steps.

4.3.1 Data cleaning

As mentioned in the previous section, we aimed to obtain four different types of data: playlist data, placements data, position data, and followers data. For the first dataset, we aimed to obtain a list of the top 1.1 million playlists on Spotify. However, due to playlists changing in rank during data collection, some playlists had multiple entries. For these observations, the first record (i.e., highest rank) is retained while the others were filtered out, resulting in 1,000,725 unique playlists. Starting the 908,325th rank, some playlist observations recorded a follower value of -1 or NA. Here, the follower value is recoded to 0 which is the correct follower value based on the rank of the playlists. A quick scan of a random sample of these playlists on Spotify shows that these playlists indeed have 0 followers. Due to a data error in the raw position data (presumably caused by a fault in the API), some track objects were duplicated within a playlist object. As a result, our placements data contained 1,462 duplicated rows. These rows were excluded. The resulting dataset contains 1,537,711 observations. A loudness of greater than 0 was found in 819 observations, or 105 unique tracks. As loudness is constrained to the [-60, 0] interval per definition, these observations are incorrect. Here, unlike the incorrect follower values, there is no way to determine the correct loudness value, and thus the observations were removed instead. Tracks with incomplete information with regards to the acoustic attributes were also removed to ensure that each track had the same weight on every acoustic attribute. Concerning the position data, we encountered a serious issue. As mentioned before, the placements data had some duplicated track objects. While this issue was of a relatively small magnitude in the placements data, it led to serious issues for the position data. In the placements data set, each track object within a playlist had 1 observation. However, in the position data set, each track object within a playlist had a number of observations equal to its position changes. Thus, if a track's position within a playlist would have changed 80 times over its lifetime in the playlist, the track-playlist object would have 80 rows in the parsed data set. As a result, the parsed position data contained 87,377,572 million observations. After filtering out duplicated observations, only 8,210,363 observations remained. We decided against filtering out the duplicated track objects during the parsing process to keep a clear separation between the parsing and data cleaning processes. Finally, in the followers dataset, duplicate entries based on the playlist, date, and follower combinations were excluded. This reduced the number of observations from 922,546 to 902,409.

4.3.2 Data aggregation

To measure the effect of variety on playlist success, two separate data frames are created. The first data frame contains the within-playlist variety aggregated to a date-playlist level. To create this data frame, a subset of the placements data is created containing only tracks which are present in any playlist on a specific date. Next, this subset is grouped per playlist. In this step, an additional error was found in the data collected using the Chartmetric API; some playlists featured the same track multiple times on the same date. In contrast to the duplicated track objects in the historic position data, these observations had slightly different dates on which they were added to and removed from the playlist. While it is possible that the same track is added to a playlist multiple times, these tracks were duplicated dozens of times and were most of the time on the same track position. Differences in the date the track was added and removed made it impossible to identify the correct observations. Thus, we removed all of these observations from the subset. Then, the number of unique artists as well as the average and coefficient of variation of the tracks' acoustic attributes within each playlist are measured.

A similar approach was used to generate the date-playlist level data frame for measuring the transition variety. First, the track's acoustic features are added to the position data set. Then, a subset is created containing only tracks which are present in any playlist on a specific date. The same data error concerning the duplicated tracks was found and these observations were again excluded. Next, this subset is grouped per playlist and ordered by the track's position. For each observation, the absolute difference was computed in the track's position and track's acoustic attributes. As the interest of this study lies in transition variety, observations with an absolute difference in track position unequal to one were filtered out. For the remaining observations, the mean of the absolute differences in the track's attributes within each playlist is measured.

The aggregation process for both the within-playlist variety and transition variety data sets is summarized in Appendix A. This process is repeated for each date between November 3rd, 2016 and September 27th, 2019 (i.e., the dates that are present in both the placements and follower datasets). Next, both the within-playlist variety and transition variety data frames are split into a training and test set using the 80/20 rule and a set seed to enforce reproducibility of our results.

4.3.3 Variable operationalization

In the playlists data, we create a new variable containing the type of playlist curator. As mentioned in Section 3.2.2, we initially distinguish between Spotify, major labels, and independent curators. Next, the Spotify playlists are broken down into human generated-, AI generated-, and personalized playlists. In the dataset, the default type of curator is 'Independent'. Next, playlists of which the playlist's owner's name starts with 'Filtr', 'Digster', or 'Topsify' are classified as 'MajorLabel'-curated. These labels belong to Sony Music Entertainment, Universal Music Group, and Warner Music Group respectively. Afterwards, playlists which are 'editorial' are classified as curated by 'SpotifyHuman'. Then, editorial playlists of which the name starts with 'This is' or ends with 'Top 50' or 'Viral 50', are reclassified as 'SpotifyAI'-curated. Lastly, playlists that are both 'editorial' and 'personalized', are reclassified as 'SpotifyPersonalized'. The owner classes of each playlist are added to the aggregated datasets based on the playlist id.

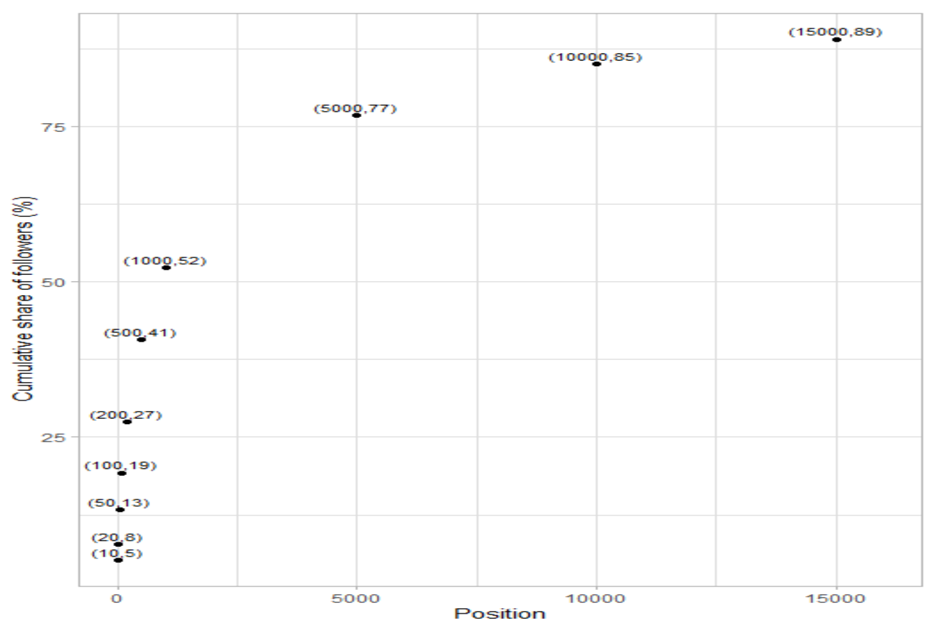
Concerning the operationalization of playlist success, we add the follower data to the aggregated datasets based on the playlist id and date. Next, as there are huge differences in the number of followers of playlists in the top 1000, we normalize the number of followers. The actual number of followers is likely more related to the popularity of the curator as well as marketing efforts, so using a relative number of followers which accounts for the baseline popularity of the playlist is more effective in measuring the effect of variety on playlist success. To normalize the number of followers, the mean of followers is calculated per playlist in the training set. For each playlist, this mean is copied to the test data. Next, for each observation in both the training and test sets, the relative followers is computed as the number of followers of a playlist on a specific date divided by the average followers of the playlist.

4.4 OVERVIEW OF THE PLAYLISTS ON SPOTIFY

In this paragraph we give an overview of the playlist and track data before aggregation as well as the supply of variety over time using the aggregated training data sets. These outputs are used to characterize the market for playlists on Spotify. The number of followers of all retrieved playlists lies between 0 and 24,129,529, with an average of 2,026.709. Figure 2 shows the cumulative market share of the top 15,000 playlists. Based on this figure, it seems that the market for playlists is highly concentrated.

Figure 2

THE CUMULATIVE MARKET SHARE OF THE TOP 15,000 PLAYLISTS



Notes. The data points represent the (x, y) coordinates. For example, the top 1,000 playlists have a cumulative market share of 52%.

Of all retrieved playlists, 98% is owned by independent curators, 1.2% by Spotify, and 0.8% by major labels. In total, we observe 248,770 unique owners in these playlists. In contrast, Spotify owns 85.2% of the top 1000 playlists. The remaining 148 playlists are owned by 77 unique owners. Concerning content types, Chartmetric differentiates between frontline-, catalogue-, and mixed content. When at least 75% of a playlist's current tracks are frontline (i.e., less than 18 months old), the playlist's content type is classified as 'frontline'. When it features less than 25% frontline tracks, it is classified as 'catalogue'. The remaining playlists have a 'mixed' content. The majority (55.8%) of all retrieved playlists feature mixed content, 41.9% features catalogue content, and the remaining 2.2% features frontline content. The distribution within the top 1000 playlists is very different. Here, 38.5% of the playlists feature frontline content, 38.4% features catalogue, and 23.1% features mixed content. Thus, playlists in the top 1000 feature more frontline tracks. This may suggest that playlist success may be driven by the age of its tracks.

Table 4 provides some descriptive statistics with regards to the track characteristics before aggregation. To compare the differences in the supply of variety across attributes, we computed the coefficient of variation (CV). As shown, the supply of variety differs per acoustic feature, e.g., the tracks have a relatively high variety in instrumentalness and a low variety in tempo. This may indicate that consumers, in general, have strong preferences for the tempo of a track.

Table 4

DESCRIPTIVES OF THE TRACK CHARACTERISTICS IN THE PLACEMENTS DATA SET

Statistic	Mean	SD	Min	Max	CV
Key	5.342	3.566	.000	11.000	.668
Mode	.602	.489	.000	1.000	.813
Danceability	.604	.177	.000	.988	.293
Energy	.639	.244	.000	1.000	.383
Speechiness	.098	.101	.000	.966	1.034
Acousticness	.275	.308	.000	.996	1.122
Instrumentalness	.152	.303	.000	1.000	1.994
Liveness	.196	.165	.000	1.000	.842
Valence	.461	.254	.000	.995	.550
Tempo	118.886	30.525	.000	248.066	.257
Loudness	-7.892	4.869	-60.000	.000	-.617

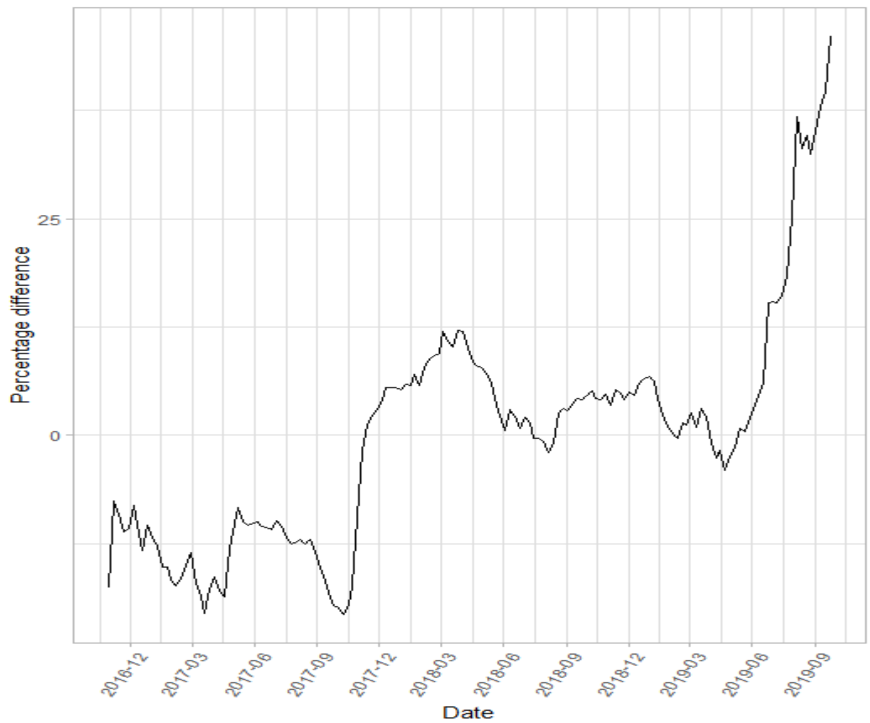
Notes. The measures are computed using a subset of the placements data containing only unique track observations (246,415). CV = SD/Mean

For the initial overview of the supply side of variety, the training data of the within-playlist variety and transition data frames are aggregated to a weekly level by computing the averages of the corresponding variables.

Figure B1 and B2 in Appendix B show the changes in the average acoustic features within playlists over time. We find that most acoustic attributes have a relatively stable supply, however, some variables show high differences in their average supply over time, e.g., acousticness, instrumentalness, and speechiness. Changes in the within-playlist variety of the acoustic features over time are shown in Figure B3 and B4. Here, we see much larger differences over time in the supply of the acoustic features, especially in danceability, acousticness, instrumentalness, loudness, and tempo. Figure B5 and B6 show the transition-variety of the acoustic features over time. Here, we see that for the danceability, energy, tempo, and loudness attributes, transition variety and within-playlist variety peak around the end of 2017. This suggests that higher levels of within-playlist variety in these attributes were associated with higher levels of variety between consecutive songs as opposed to it being built up using smooth transitions. Furthermore, it seems that higher levels of variety are not necessarily associated with higher levels of the corresponding attribute within a playlist. Concerning the within-playlist variety in artists over time, Figure 3 shows a very high variety and clear positive trend in the average unique number of artists. This may indicate that the growth in playlist popularity over albums may be related to playlists featuring variety in artists.

Figure 3

PERCENTAGE DIFFERENCE IN THE AVERAGE NUMBER OF UNIQUE ARTISTS WITHIN A PLAYLIST OVER TIME

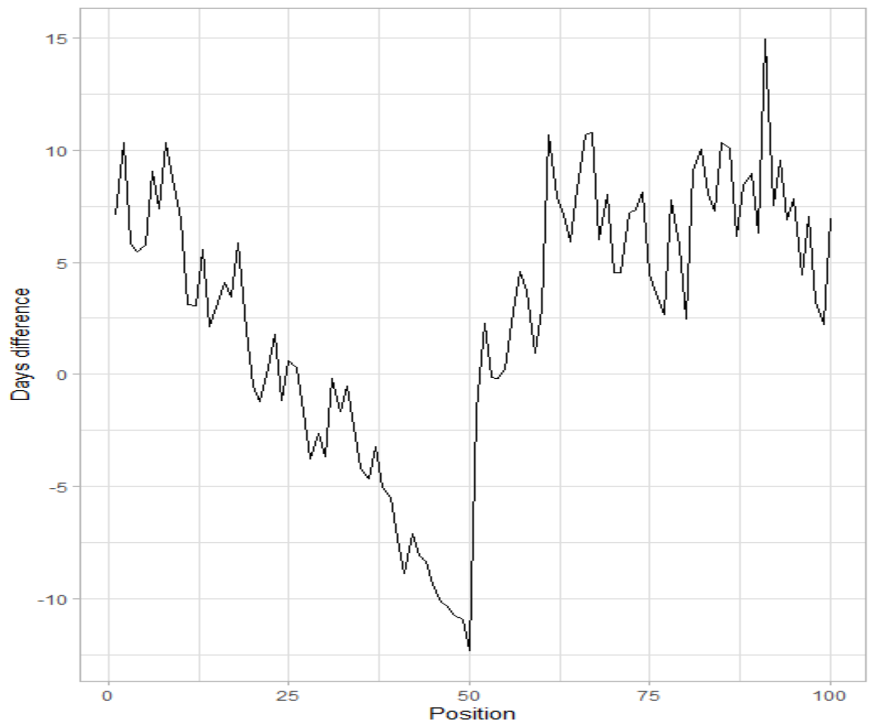


Notes. The computation of the percentage difference is explained in Appendix B.

Concerning track placement, we observe that tracks generally are added to either the front or the end of a playlist as shown in Figure 4. In additional explorative plots, Appendix C explores the placement of tracks in more detail. We observe that older tracks are typically added to the end in Figure C1. Playlists with frontline content typically have larger differences in the dates in which tracks were added than playlists with catalogue content as shown in Figure C2. Lastly, in Figure C3, we observe that playlists owned by major labels have large differences in dates compared to playlists owned by other creators. This may suggest that major labels typically maintain their playlists over a longer period of time (so that tracks may be added over a wider period), whereas other playlist curators typically either update their playlists more frequently or not at all after creation (so that the date tracks were added are closer). Furthermore, we see that the curve for playlists owned by the Spotify AI is much more instable after position 50, and drops almost vertically after position 88. Based on the operationalization of owner classes, Spotify AI playlists contain mostly playlists focusing on the top 50 tracks in a certain category. Thus, it seems likely that the estimates after the 50th position are much more instable. Furthermore, this suggests that the sudden shift after the 50th position in Figure 4 may be driven by Spotify's AI-generated playlists no longer being present.

Figure 4

DIFFERENCE IN DAYS BY TRACK POSITION



Notes. The computation of the days difference is explained in Appendix C.

4.5 OVERVIEW OF THE FINAL DATA SET

This subsection provides a general overview of the variation in the data along the different dimensions of variety as outlined in paragraph 3.2.1. Our complete training data contains 645,823 observations in the within-playlist variety model and 631,087 observations in the transition variety model. The coefficient of variation is used to measure the within-playlist variety while the mean absolute deviation is used to measure the transition variety. The decision process for the use of these measures is explained in Chapter 5. As shown in Table 5, instrumentalness generally has a relatively high level of within-playlist variety and a low level of transition-variety compared to the other acoustic attributes. This suggests that playlists typically build up a high within-playlist variety in instrumentalness through smooth transitions.

Table 5

DESCRIPTIVE STATISTICS OF THE AGGREGATED DATA SET

	Within-playlist variety				Transition variety			
Statistic	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
Quantity of consumption								
Relative followers	1.003	.357	.000	4.439	1.000	.347	.000	4.012
Breadth of variety								
Key	.676	.135	.000	2.000	4.038	.866	.000	11.000
Mode	.767	.319	.000	4.899	.416	.168	.000	1.000
Danceability	.230	.118	.000	2.387	.142	.047	.000	.859
Energy	.297	.193	.000	2.162	.169	.059	.000	.933
Speechiness	.722	.256	.000	2.355	.063	.045	.000	.878
Acousticness	1.007	.573	.000	9.737	.211	.103	.000	.990
Instrumentalness	3.432	1.850	.000	14.375	.086	.116	.000	.983
Liveness	.725	.168	.000	1.663	.146	.071	.000	.925
Valence	.457	.226	.000	4.072	.221	.065	.000	.868
Tempo	.255	.102	.000	2.216	31.642	11.395	.000	186.001
Loudness	-.364	.123	-3.316	.000	2.867	1.402	.000	23.573
Artists	43.530	37.981	1.000	721.000				
Playlist curators								
SpotifyHuman	.406	.491	.000	1.000	.400	.490	.000	1.000
SpotifyAI	.116	.321	.000	1.000	.118	.323	.000	1.000
SpotifyPersonalized	.333	.471	.000	1.000	.337	.473	.000	1.000
MajorLabel	.065	.247	.000	1.000	.064	.245	.000	1.000
Independent	.079	.270	.000	1.000	.080	.271	.000	1.000

Notes. In the within-playlist variety model, the variety is measured using the coefficient of variation. In the transition variety model, the mean absolute deviation is used instead.

5 MODEL

We use a multivariate time series regression model to estimate the effect of variety on playlist success as well as how this effect is influenced by the type of playlist curator. The model is estimated using the training data sets as described in paragraph 4.3. Since we focus on the top 1000 playlists, it is likely that playlists have a higher share of relative followers because the number of followers grew over time. We include date as a covariate variable to control for this effect. Furthermore, as mentioned in paragraph 2.3, it is unclear which measure of variety will have the best fit as previous studies mostly used single measures of variety instead of testing the significance of each of its dimensions. Before arriving at the model specification below, we tested the model fit using the median absolute deviation centered on the mean, median absolute deviation centered on the median, interquartile range, and coefficient of variation (CV). The fit of the within-playlist models (as measured by R^2) were .5689, .5683, .5684, and .5704 respectively. Thus, the model fit hardly changes. In the final model specification, the CV is chosen to measure within-playlist variety as it allows for easier comparison across attributes with different means. With regards to transition variety, the mean absolute deviation is chosen instead as it allows us to easily adjust the center point from the mean of the playlist to the previous track without losing interpretability. Furthermore, from a practical point of view, transition variety is more easily incorporated than within-playlist variety. Thus, choosing the MAD to measure transition-variety enhanced the practical implications for playlist curators.

The final model is then:

$$
\begin{aligned}
Y_{pd} = \beta_0 + \sum_{i=1}^{N} (\beta_i Variety_{i,p,d} + \gamma_i {Variety^2}_{i,p,d} + \delta_i Variety_{i,p,d} \times SpotifyHuman_p \\
+ \eta_i Variety_{i,p,d} \times SpotifyAI_p + \phi_i Variety_{i,p,d} \times SpotifyPersonalized_p \\
+ \psi_i Variety_{i,p,d} \times MajorLabel_p) + \delta_0 SpotifyHuman_p + \eta_0 SpotifyAI_p \\
+ \phi_0 SpotifyPersonalized_p + \psi_0 MajorLabel_p + \omega Date + \varepsilon_{pd}
\end{aligned}
$$

where Y_{pd} is the relative followers of playlist p on date d (i.e., $\frac{followers_{p,d}}{average\ followers_p}$), β_0 is the constant, N is the number of track attributes, and $Variety_{i,p,d}$ is the variety in attribute i (see Table 6) in playlist p on date d. The variables $SpotifyHuman_p$, $SpotifyAI_p$,

$SpotifyPersonalized_p$, $MajorLabel_p$ are dummy variables indicating the type of playlist curator of playlist p and ε_{pd} is the error.

In the within-playlist variety model, the variety is measured by the coefficient of variation, whereas in the transition variety model, it is measured by the mean absolute deviation. An exception occurs for the artist attribute (i = 12), which is measured in number of uniques in the within-playlist variety model, and is excluded from the transition variety model.

Table 6

ATTRIBUTE INDICES

i	Attribute
1	Key
2	Mode
3	Danceability
4	Energy
5	Speechiness
6	Acousticness
7	Instrumentalness
8	Liveness
9	Valence
10	Tempo
11	Loudness
12	Artists

6 RESULTS

In this section, we present the results of the analyses conducted using the model specification described in the previous section. First, we elaborate on the results regarding model fit in paragraph 6.1. Next, we investigate the results regarding the main effects of the variety dimensions on playlist success in paragraph 6.2. Finally, we discuss the moderating effect of the playlist curator in paragraph 6.3.

6.1 MODEL FIT

As mentioned in the previous chapter, we used a regression analysis to test if within-playlist variety in acoustic attributes significantly predicted playlist success, and if this relationship is influenced by the type of playlist curator. The results of the regression indicated that the predictors explained 57.0% of the variance (R^2 = .570, R^2_{adj} = .570, F(77;645,745) = 11,135.230, p < .001). A similar model was run in the case of transition variety, here the predictors explained 55.7% of the variance (R^2 = .557, R^2_{adj} = .557, F(71;631,015) = 11,192.310 , p < .001). The full model results for the within-playlist and transition variety models are shown in Appendix D and E respectively. As expected, the control variable date has a significant positive effect on playlist success (within-playlist variety: ω = .001, p < .001; transition-variety: ω = .001, p < .001). This is a face-valid effect as playlists in the dataset are likely to have grown in followers over time to secure their place in the top 1000. We tested both models for overfitting and multicollinearity. Table 7 summarizes the results for overfitting. As shown in this table, the RMSE and R^2 scores are very similar for the test and training sets in both models. Hence, no evidence for overfitting was found. Next, to test for multicollinearity, the VIF scores of the predictors were computed in both models. The results of this computation are documented in Appendix F. As many studies use a VIF greater than 10 as an indication for multicollinearity, no evidence for multicollinearity was found since none of the VIF values exceeded 5 (Menard, 1995; Neter et al., 1989; Hair et al., 1995; Marquardt, 1970; Mason et al., 1989).

Table 7

TEST OF MODEL FIT ON TRAINING AND TEST SETS

	Within-playlist variety		Transition variety	
Statistic	**Training set**	**Test set**	**Training set**	**Test set**
RMSE	.234	.233	.231	.232
R^2	.570	.573	.557	.556

6.2 THE MAIN EFFECT OF VARIETY ON PLAYLIST SUCCESS

Table 8 shows the effects of the twelve dimensions of variety on playlist success as estimated by the regression model. In contrast to our expectations, within-playlist variety and transition-variety have very different effects on playlist success with the exception of the key, instrumentalness, and energy dimensions. This finding does not necessarily oppose previous research which assumed that both measurements of variety have similar effects. Rather, it provides a more in-depth view as to how this effect is shaped through the individual contribution of each variety dimension. It may be the case that consumers preferences for within-playlist variety and transition variety is indeed similar overall, but the composition of these effects differ. However, in both models the vast majority of the effects are positive or of quadratic nature. Consequently, the assumption that semantic cohesion (i.e., similarity) is a good evaluation metric for playlist quality does not seem to have face validity. Furthermore, as shown in Table 8, the acousticness, danceability, and tempo dimensions of variety are not significant in one of the models. This finding may be explained by consumers not perceiving the variety in the associated acoustic attribute in the model in which the effect is insignificant. This is in keeping with the ideas expressed by Schedl, Flexer, and Urbano (2013) who found that differences in consumer preferences may influence the perception of item similarity. Furthermore, this finding discredits the assumption of most existing approaches of measuring track similarity which rely on a number of predefined audio attributes and thereby assume that all of these features are predictive of the consumers' preferences (Schedl et al., 2018).

As we expected, variety in artists has a positive effect (β_{12}= .0001, p < .001) on playlist success. This finding supports the idea that playlists may be more successful than albums because albums often feature a single artist. Likewise, within-playlist variety in energy (β_4= .061, p = .001) and transition variety in energy (β_4= .261, p < .001) both have a positive effect on playlist success. This result supports the notion that although consumers like highly energetic songs, listening to a series of such songs can become tiresome. However, contrary to our expectations, this relationship is non-linear in the within-playlist model (γ_4 = .021, p < .001). Thus, within a playlist consumers have an increasing appreciation for variety in energy as this variety increases. Furthermore, as we anticipated, consumers seem to prefer either a relatively low or high variety in the instrumentalness (within-playlist variety: β_7 = -.003, p = .001, γ_7 = .0005, p < .001; transition variety: β_7 = -.140, p < .001; γ_7 = .253, p < .001) and key (within-playlist variety: β_1 = -.084, p < .001, γ_1 = .036, p < .001; transition variety: β_1 = -.038, p < .001, γ_1 = .003, p < .001) dimensions as opposed to previous studies estimating a linear relationship.

Furthermore, transition variety in liveness has a negative association with playlist success (β_8= -.053, p = .008, γ_8 = -.050, p < .001) as expected. However, this changes to a positive association for within-playlist variety (β_8= .116, p < .001). This suggests that consumers prefer playlists with a higher level of variety in liveness, but the difference in liveness between consecutive tracks should be low. As mentioned in Chapter 3, it was difficult to estimate the effects of variety in liveness beforehand as recent developments have shifted demand towards higher quality audio. It is intuitive to find a negative effect with regards to the transition-variety as it may be annoying for users to listen to live recordings and studio recordings in alternation. However, it seems that despite live recordings losing popularity (Boughanmi et al., 2018) people still prefer some variety in liveness in a playlist over no variety.

Like liveness, the remaining dimensions of variety show different consumer preferences for within-playlist variety and transition variety. To our knowledge, this thesis is the first to examine the individual effects of the dimensions of variety on playlist success. The expectations were, for the most part, based on indirect qualitative research and Boughanmi et al.'s (2018) study on the effect of variety on album success. Due to this limited availability of related research, it is difficult to provide reasoning for the findings which contradict our expectations. We will elaborate on these limitations in Chapter 7.

Table 8

EFFECTS OF VARIETY ON PLAYLIST SUCCESS (WITHIN-PLAYLIST VARIETY VS. TRANSITION VARIETY)

Dimension	**Expected association**	**Within-playlist variety model**	**Transition variety model**
Acousticness	U-shape	Insignificant	Negative but decreasing effect
Danceability	Positive	U-shape	Insignificant
Energy	Positive	Positive but increasing effect	Positive linear effect
Instrumentalness	U-shape	U-shape	U-shape
Key	U-shape	U-shape	U-shape
Liveness	Negative	Positive linear effect	Negative but increasing effect
Loudness	Positive	Inverse u-shape	Positive but increasing effect

Dimension	Expected association	Within-playlist variety model	Transition variety model
Mode	Inverse u-shape	Positive linear effect	Inverse u-shape
Speechiness	Inverse u-shape	U-shape	Inverse u-shape
Tempo	Positive	Inverse u-shape	Insignificant
Valence	Inverse u-shape	Inverse u-shape	Negative but increasing effect
Artists	Positive	Positive linear effect	

Notes. The classification process of these effects is explained in Appendix G.

6.3 THE MODERATING EFFECT OF PLAYLIST CURATORS

Table 9 shows the interaction effects of each variety dimension with the type of playlist curator. The entries in the table are the estimated values of δ_i, η_i, ϕ_i, and ψ_i as shown in the model specification in Chapter 5. Thus, the higher the estimated value, the more the consumers of the associated playlist curator appreciate the variety dimension with regards to playlist success. For example, within a playlist, the relationship between energy and playlist success is strongest for playlists created by independent curators and weakest for playlists created by the Spotify AI. An exception occurs with regard to the loudness dimension. Here, the within-playlist variety, i.e., the coefficient of variation, is always 0 or negative. Thus, in this case, the relationship is reversed; lower estimated values reflect stronger relationships between variety in loudness and playlist success. To further improve the understanding of the influence of the playlist curator on the relationship between variety and playlist success, Figure 5 shows the inverse U-shaped effect of within-playlist variety in loudness as well as the U-shaped effect of transition variety in key on playlist success. The remaining 5 effects (see Table 8) are illustrated in Figure H1 (Appendix H).

Table 9

MODERATING EFFECT OF PLAYLIST CURATOR ON THE RELATION BETWEEN VARIETY AND PLAYLIST SUCCESS

	Within-playlist variety model				Transition variety model			
	Human curated	AI	Personalized	Major label	Human curated	AI	Personalized	Major label
Key	-.032**	-.023	.082***	.135***	.013***	.025***	.013***	.006**
Mode	-.014**	-.008	-.042***	-.127***	-.036***	-.064***	-.004	-.019
Danceability[b]	-.079**	.467***	.133***	.422***	.022	-.043	-.009	.336***
Energy	-.072***	-.234***	-.148***	-.114***	-.420***	-.274***	-.355***	.137**
Speechiness	-.034***	.066***	-.011	-.055***	-.054	-.112**	-.088**	-.222***
Acousticness[a]	.013***	.014**	-.017***	.021***	.395***	.240***	.430***	.069**
Instrumentalness	.002***	.007***	.007***	-.026***	.004	.017	.011	-.135***
Liveness	-.152***	.023*	-.130***	-.169***	-.060**	.173***	-.021	.551***

	Within-playlist variety model				Transition variety model			
	Human curated	**AI**	**Personalized**	**Major label**	**Human curated**	**AI**	**Personalized**	**Major label**
Valence	.020	-.255***	-.054***	-.402***	.544***	.775***	.631***	.482***
Tempo[b]	.087***	.018	-.093***	-.003	-.001***	-.0006	-.001***	-.001***
Loudness	.051***	.160***	.102***	-.537***	-.037***	-.040***	-.037***	.002
Artists	-.0003***	-.0002*	.0004***	-.002***				

Notes. The ‘Human curated’, ‘AI’, and ‘Personalized’ playlist curators refer to the different types of Spotify-owned playlists as discussed in Chapter 3.2.2.

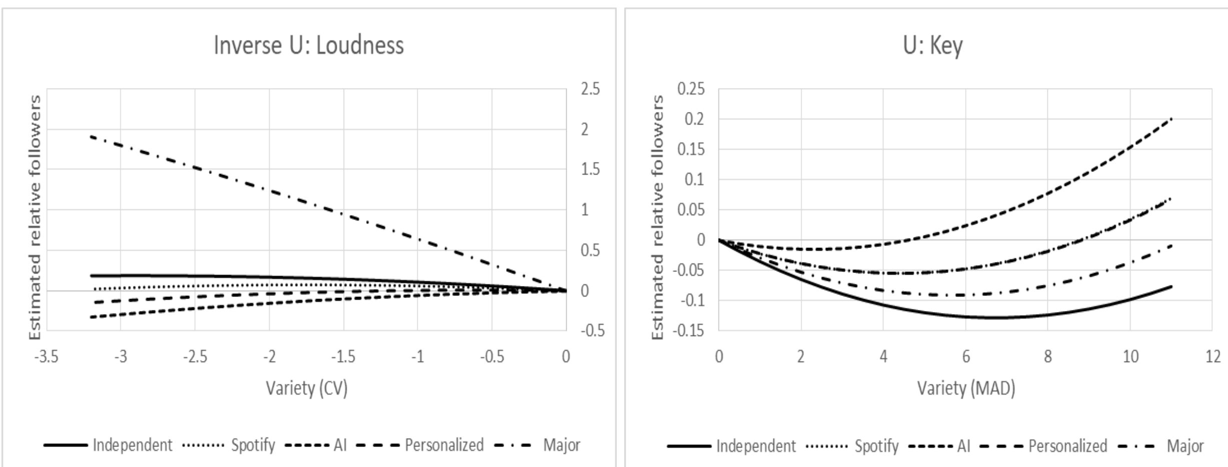
[a]The main effect is insignificant at a 0.05 level in the within-playlist variety model.

[b]The main effect is insignificant at a 0.05 level in the transition variety model.

*p < 0.05; **p < 0.01; ***p < 0.001.

Figure 5

EXAMPLES OF THE INFLUENCE OF THE TYPE OF PLAYLIST CURATOR ON THE EFFECT OF VARIETY ON PLAYLIST SUCCESS



The results indicate that while the main effects of within-playlist variety in acousticness and transition variety in danceability and tempo are insignificant, variety in these dimensions does seem to be impact playlist success depending on the curator of the playlist. Furthermore, several patterns can be extracted from this table. We will discuss these patterns per type of curator.

First, consumers of Spotify’s human generated playlists have consistent preferences for variety relative to consumers of independent curators (i.e., most interaction effects have the same sign in both models). They have the lowest appreciation for within-playlist variety in key and danceability and for transition variety in energy and liveness. However, they have the highest appreciation for within-playlist variety in tempo.

Second, in contrast to our expectations, consumers of Spotify's AI-generated playlists have the highest preferences for within-playlist variety over all. In a third of the significant effects, the estimated interaction effects of Spotify AI playlists is the highest. However, they have the lowest appreciation for within-playlist variety in energy. Interestingly, they have the highest appreciation for transition variety in key and the lowest appreciation for the transition variety in mode, whereas these interaction effects were insignificant in the within-playlist variety model.

Third, consumers of Spotify's personalized playlists have the lowest preferences for variety in tempo and for within-playlist variety in acousticness. However, they have the highest appreciation for transition variety in acousticness and for within-playlist variety in instrumentalness and artists.

Fourth, consumers of major label playlists have the lowest appreciation for within-playlist variety. Here, in 6 out of 11 significant effects, the estimated interaction effect of major label is the lowest. As expected, consumers of major labels have the lowest appreciation for variety in artists. This may be due to consumers of these playlists liking a particular artist that is signed with the major label. Surprisingly, they have very different preferences for within-playlist variety and transition variety. In 3 out of 8 significant effects, a relatively low appreciation for within-playlist variety is associated with a relatively high appreciation for transition variety. This is the case for the energy, liveness and valence dimensions. As a result, they have the second highest appreciation for transition variety in general. Here, in a third of the significant effects, the estimated interaction effect of major label playlists is the highest. Based on the estimated interaction effects, it seems that consumers of major label playlists have more specific preferences for transition variety than for within-playlist variety.

Finally, consumers of independent curators have the highest preferences for transition variety overall. Here, consumers have highest appreciation for variety in 5 out of 11 attributes. Surprisingly, they also have the lowest appreciation for transition-variety in 4 out of 11 attributes. Thus it seems that, although consumers of playlists owned by independent curators do not have particularly strong preferences for within-playlist variety, they have specific preferences for transition variety.

7 DISCUSSION

7.1 SUMMARY OF MAIN FINDINGS

This thesis provided a first overview of how variety is incorporated in playlists. We showed that the supply of variety changes over time and differs per acoustic attribute. Furthermore, we observed that these changes are not necessarily associated with changes in the mean of the attribute. Likewise, we observe that within-playlist variety and transition-variety may have a different supply over time. This may suggest that for some attributes consumers like smooth transitions to build up variety, whereas for other attributes rougher transitions are preferred.

Furthermore, this thesis argued that the dimensions of variety have different effects on playlist success. We found evidence that playlists have relatively more followers on days in which it incorporates high variety in its tracks' energy, whereas variety in its tracks' tempo is only appreciated in moderation. Moreover, the results of this thesis indicate that within-playlist variety and transition similarity have very different effects on playlist success. For example, our findings state that within-playlist variety in liveness has a positive effect on playlist success, but transition variety in liveness has a negative effect. Concerning the playlist curators, we found that consumers of major label playlists typically have the lowest appreciation for within-playlist variety, whereas consumers of Spotify's AI-generated playlists have the highest. Consumers of independent curators do not have strong preferences for within-playlist variety, but have very specific preferences for transition variety relative to consumers of other playlist curators.

7.2 THEORETICAL AND MANAGERIAL TAKE-AWAYS

With regards to the main effects of the variety dimensions on playlist success, our expectations were often wrong or insignificant in one of the models. As existent literature does not make a clear distinction between the effects of transition variety and within-playlist variety on playlist success, it is difficult to provide support for our findings. This difficulty increases as research on the individual effects of the dimensions of variety is scarce. Despite this limitation, this thesis provides evidence that the relationship between variety and playlist success is far more complex than current research on APG approaches acknowledge. While variety can have a negative effect on flow and therefore playlist quality, we show that variety can have positive effects as well and that this effect differs per dimension and per type of playlist curator. Furthermore, the use of predefined features to measure variety shows flaws as some features had an insignificant effect on playlist success in one of the models.

Concerning the practical implications, this thesis provides new insights into consumer preferences for music diversity with respect to the playlist curator. For example, we found that consumers of major label playlists have the lowest appreciation for within-playlist variety and variety in artists, but have the second highest appreciation for transition variety. Human curators and AIs can use this information to tailor their playlists to their consumers' perception of and preferences for variety. However, it may be difficult to apply the optimal levels of variety in each attribute as in reality these attributes are not independent. For example, we would advise major labels to minimize variety in artists and transition variety in speechiness and instrumentalness while maximizing transition variety in energy and danceability. Trade-offs need to be made.

7.3 LIMITATIONS AND FUTURE RESEARCH

This thesis has a number of limitations. First, the data collected using the Chartmetric API had many errors, as explained in paragraph 4.3 (the data preparation section). Aside from these data errors, many observations were excluded in the data aggregation of the transition-variety dataset due to missing position data. Although we cleaned the data to the best of our ability, it is likely that there are data errors left in the dataset which are not obvious. Furthermore, 'relative followers' is not an ideal measure for consumer preferences. It shows us how many consumers like a playlist and intent to listen to it. However, it does not tell us how many of these consumers actually listen to the playlist. According to Joven (2018), a playlist's number of followers almost always outnumbers its estimated listeners count. Thus, although 'following' a playlist is a better measure for engagement than the number of streams, there is room for improvement. Joven (2018) suggests a follower-to-estimated-listeners (FEL) ratio as a more meaningful playlist engagement metric. This metric takes both the followers' intentions and behaviors into account and may thus provide a more holistic view of consumer preferences. As we did not have access to the estimated listeners count of each playlist over time, the use of FEL may be an interesting avenue for future research. Another avenue for future research includes testing the importance of the individual attributes. Finally, this thesis is a highly correlational study. It aimed to describe and predict behavior, but not to explain it. Thus, an area for future research would be investigating whether the discovered relationships in this thesis are causal relationships.

References

Adomavicius, G., Bockstedt, J., & Curley, S. P. (2015). Bundling effects on variety seeking for digital information goods. *Journal of Management Information Systems, 31*(4), 182-212.

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. doi:10.1016/j.ijindorg.2017.06.004

Bonnin, G., & Jannach, D. (2015). Automated generation of music playlists: Survey and experiments. *ACM Computing Surveys (CSUR), 47*(2), 26.

Boughanmi, K., Ansari, A., & Kohli, R. (2018). *Dynamics of musical success: A bayesian nonparametric approach*. Retrieved from http://www.columbia.edu/~kb2662/music_dynamics.pdf

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. doi:10.1287/mksc.2017.1051

Gatewood, E. L. (1927). An experimental study of the nature of musical enjoyment. In M. Schoen (Ed.), *The effects of music* (pp. 78-127). Freeport, NY: Books for Libraries Press.

Gosselin, N., Peretz, I., Noulhiane, M., Hasboun, D., Beckett, C., Baulac, M., & Samson, S. (2005). Impaired recognition of scary music following unilateral temporal lobe excision. *Brain, 128*(3), 628-640.

Haber, R. N. (1958). Discrepancy from adaptation level as a source of affect. *Journal of Experimental Psychology, 56*(4), 370.

Hagen, A. N. (2015). The playlist experience: Personal playlists in music streaming services. *Popular Music and Society, 38*(5), 625-645.

Hogan, M. (2015). How playlists are curating the future of music. Retrieved from https://pitchfork.com/features/article/9686-up-next-how-playlists-are-curating-the-future-of-music/

Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2008). Mixed affective responses to music with conflicting cues. *Cognition & Emotion, 22*(2), 327-352.

Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2010). Feelings and perceptions of happiness and sadness induced by music: Similarities, differences, and mixed emotions. *Psychology of Aesthetics, Creativity, and the Arts, 4*(1), 47.

Husain, G., Thompson, W. F., & Schellenberg, E. G. (2002). Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music Perception: An Interdisciplinary Journal, 20*(2), 151-171.

Inman, J. J. (2001). The role of sensory-specific satiety in attribute-level variety seeking. *Journal of Consumer Research, 28*(1), 105-120.

Joven, J. (2018). Measuring attention on Spotify playlists: Followers vs. listeners. Retrieved from https://blog.chartmetric.io/measuring-attention-on-spotify-followers-vs-listeners-da0383d90796

Kahn, B. E. (1995). Consumer variety-seeking among goods and services: An integrative review. *Journal of retailing and consumer services, 2*(3), 139-148. doi:10.1016/0969-6989(95)00038-0

Kahn, B. E., & Isen, A. M. (1993). The influence of positive affect on variety seeking among safe, enjoyable products. *Journal of Consumer Research*, *20*(2), 257-270.

Kahneman, D., & Snell, J. (1990). Predicting utility. In R. M. Hogarth (Ed.), *Insights in decision making: A tribute to Hillel J. Einhorn*. Chicago, IL, US: University of Chicago Press.

Kamehkhosh, I., Jannach, D., & Bonnin, G. (2018). *How automated recommendations affect the playlist creation behavior of users.* Paper presented at the IUI Workshops.

Knees, P., Pohle, T., Schedl, M., & Widmer, G. (2006). *Combining audio-based similarity with web-based data to accelerate automatic music playlist generation.* Paper presented at the Proceedings of the 8th ACM international workshop on Multimedia information retrieval.

Knobloch, S., & Zillmann, D. (2002). Mood management via the digital jukebox. *Journal of communication*, *52*(2), 351-366.

Koelsch, S., Fritz, T., von Cramon, D. Y., Müller, K., & Friederici, A. D. (2006). Investigating emotion with music: An fMRI study. *Human brain mapping, 27*(3), 239-250.

Koumis, A. (2017) *I create Spotify playlists for a living/Interviewer: C. Allen*. BuzzFeed.

Ladinig, O., & Schellenberg, E. G. (2012). Liking unfamiliar music: Effects of felt emotion and individual differences. *Psychology of Aesthetics, Creativity, and the Arts, 6*(2), 146.

Lee, J. H., Bare, B., & Meek, G. (2011). *How similar is too similar?: Exploring users' perceptions of similarity in playlist evaluation.* Paper presented at the 12th International Society for Music Information Retrieval Conference (ISMIR), Miami, FL, USA.

Lunny, O. (2019). Why high-fidelity streaming is the audio revolution your ears have been waiting for. *Forbes*. Retrieved from https://www.forbes.com/sites/oisinlunny/2019/01/18/why-high-fidelity-streaming-is-the-audio-revolution-your-ears-have-been-waiting-for/#1062c43273ff.

Maillet, F., Eck, D., Desjardins, G., & Lamere, P. (2009). *Steerable Playlist Generation by Learning Song Similarity from Radio Station Playlists.* Paper presented at the ISMIR.

McAlister, L. (1979). Choosing multiple items from a product class. *Journal of Consumer Research, 6*(3), 213-224. doi:10.1086/208763

McAlister, L. (1982). A dynamic attribute satiation model of variety-seeking behavior. *Journal of Consumer Research, 9*(2), 141-150. doi:10.1086/208907

McAlister, L., & Pessemier, E. (1982). Variety seeking behavior: An interdisciplinary review. *Journal of Consumer Research, 9*(3), 311-322. doi:10.1086/208926

McDonald, H. (2019). The pros and cons of independent record label deals. Retrieved from https://www.thebalancecareers.com/indie-record-deals-assesing-the-pros-and-cons-2460376

McFee, B., & Lanckriet, G. R. (2011). *The Natural Language of Playlists.* Paper presented at the ISMIR.

McFee, B., & Lanckriet, G. R. (2012). *Hypergraph models of playlist dialects.* Paper presented at the 13th International Society for Music Information Retrieval Conference (ISMIR), Porto, Portugal.

Music Business Association. (2016). Playlists overtake albums in listenership, says loop study. Retrieved from https://musicbiz.org/news/playlists-overtake-albums-listenership-says-loop-study/

Nielsen. (2018a). 2017 U.S. music year-end report. Retrieved from https://www.nielsen.com/us/en/insights/reports/2018/2017-music-us-year-end-report.html

Nielsen. (2018b). U.S. music mid-year report 2018. Retrieved from https://www.nielsen.com/us/en/insights/reports/2018/us-music-mid-year-report-2018.html

Panksepp, J. (1995). The emotional sources of" chills" induced by music. *Music Perception: An Interdisciplinary Journal, 13*(2), 171-207.

Panteli, M., Benetos, E., & Dixon, S. (2016). *Learning a feature space for similarity in world music.* Paper presented at the 17th international society for music information retrieval conference (ISMIR), New York, NY, USA.

Plambeck, J. (2010). In mobile age, sound quality steps back. *The New York Times*. Retrieved from https://www.nytimes.com/2010/05/10/business/media/10audio.html

Ragno, R., Burges, C. J., & Herley, C. (2005). *Inferring similarity between music objects with application to playlist generation.* Paper presented at the Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval.

Reynolds, G., Barry, D., Burke, T., & Coyle, E. (2007). *Towards a personal automatic music playlist generation alogorithm: The need for contextual information.* Paper presented at the 2nd international audio mostly conference: interaction with sound, Ilmenau, Germany.

RIAA. (2019). RIAA 2018 year-end music industry revenue report. Retrieved from http://www.riaa.com/wp-content/uploads/2019/02/RIAA-2018-Year-End-Music-Industry-Revenue-Report.pdf

Schäfer, T., Auerswald, F., Bajorat, I. K., Ergemlidze, N., Frille, K., Gehrigk, J., . . . Sanahuja, A. (2016). The effect of social feedback on music preference. *Musicae Scientiae, 20*(2), 263-268.

Schäfer, T., & Mehlhorn, C. (2017). Can personality traits predict musical style preferences? A meta-analysis. *Personality and Individual Differences*, *116*, 265-273.

Schedl, M., Flexer, A., & Urbano, J. (2013). The neglected user in music information retrieval research. *Journal of Intelligent Information Systems, 41*(3), 523-539.

Schedl, M., Zamani, H., Chen, C.-W., Deldjoo, Y., & Elahi, M. (2018). Current challenges and visions in music recommender systems research. *International Journal of Multimedia Information Retrieval, 7*(2), 95-116. doi:10.1007/s13735-018-0154-2

Schellenberg, E. G., Peretz, I., & Vieillard, S. (2008). Liking for happy-and sad-sounding music: Effects of exposure. *Cognition & Emotion, 22*(2), 218-237.

Scherer, K. R. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *Journal of new music research, 33*(3), 239-251.

Serrà, J., Corral, Á., Boguñá, M., Haro, M., & Arcos, J. L. (2012). Measuring the evolution of contemporary western popular music. *Scientific reports, 2*, 521.

Simonson, I. (1990). The effect of purchase quantity and timing on variety-seeking behavior. *Journal of Marketing Research, 27*(2), 150-162. doi:10.2307/3172842

Slaney, M., & White, W. (2006). *Measuring playlist diversity for recommendation systems.* Paper presented at the Proceedings of the 1st ACM workshop on Audio and music computing multimedia.

Sloboda, J. A. (1992). Empirical studies of emotional response to music.

Spotify. (2019). Company info. Retrieved from https://newsroom.spotify.com/company-info/

Statista. (2019). Share of music streaming subscribers worldwide as of the first half of 2019, by company. Retrieved from https://www.statista.com/statistics/653926/music-streaming-service-subscriber-share/

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2001). Arousal, mood, and the Mozart effect. *Psychological science, 12*(3), 248-251.

Vall, A., Quadrana, M., Schedl, M., Widmer, G., & Cremonesi, P. (2017). *The Importance of Song Context in Music Playlists.* Paper presented at the RecSys Posters, Como, Italy.

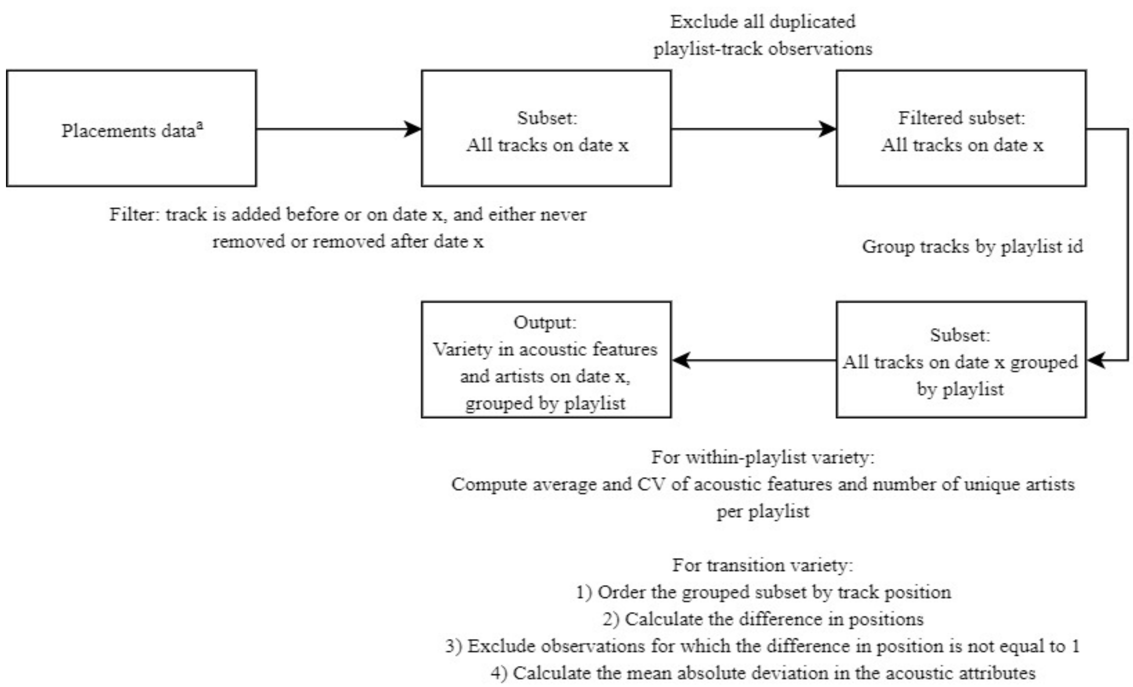
Wang, X., Sun, L., & Keh, H. T. (2013). Consumer responses to variety in product bundles: The moderating role of evaluation mode. *International Journal of Research in Marketing, 30*(4), 335-342.

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion, 8*(4), 494.

APPENDIX A: SUMMARY OF THE DATA AGGREGATION PROCESS

Figure A1

THE AGGREGATION PROCESS



[a]In the aggregation process of the transition variety data, the placements data includes the parsed position stats, whereas for the computation of the within-playlist variety the position data is left out.

APPENDIX B: AVERAGE PLAYLIST ATTRIBUTES OVER TIME

In the plots below, the Y-axis features the percentage difference of an attribute Z. This measure is calculated as follows:

$$Percentage\ difference_Z = \frac{Average\ Z\ in\ week\ X - Average\ Z\ across\ all\ dates}{Average\ Z\ across\ all\ dates} * 100\%$$

Figure B1

AVERAGE ACOUSTIC ATTRIBUTES WITHIN A PLAYLIST OVER TIME (PART I/II)

Figure B2

AVERAGE ACOUSTIC ATTRIBUTES WITHIN A PLAYLIST OVER TIME (PART II/II)

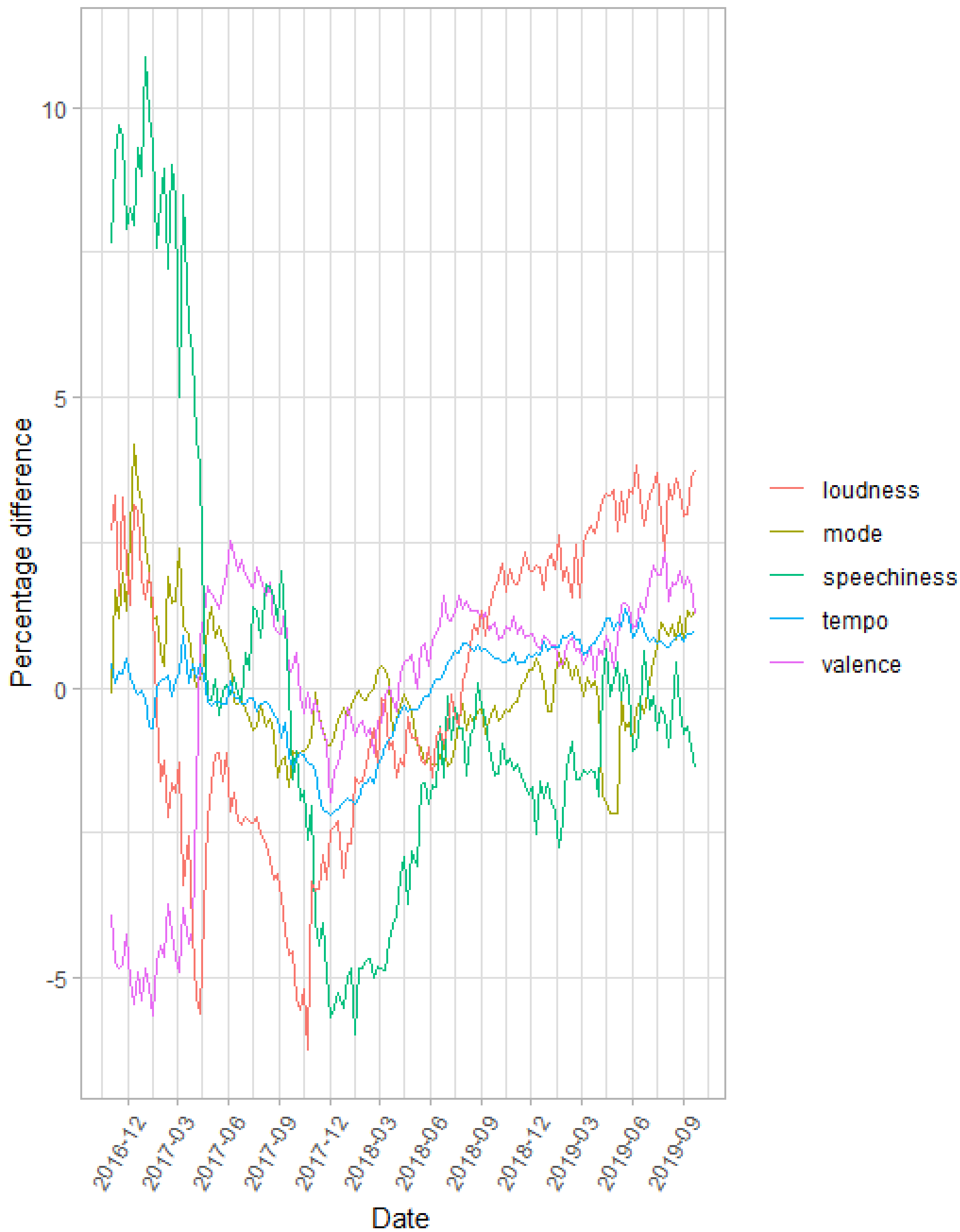


Figure B3

WITHIN-PLAYLIST VARIETY OF ACOUSTIC ATTRIBUTES OVER TIME (PART I/II)

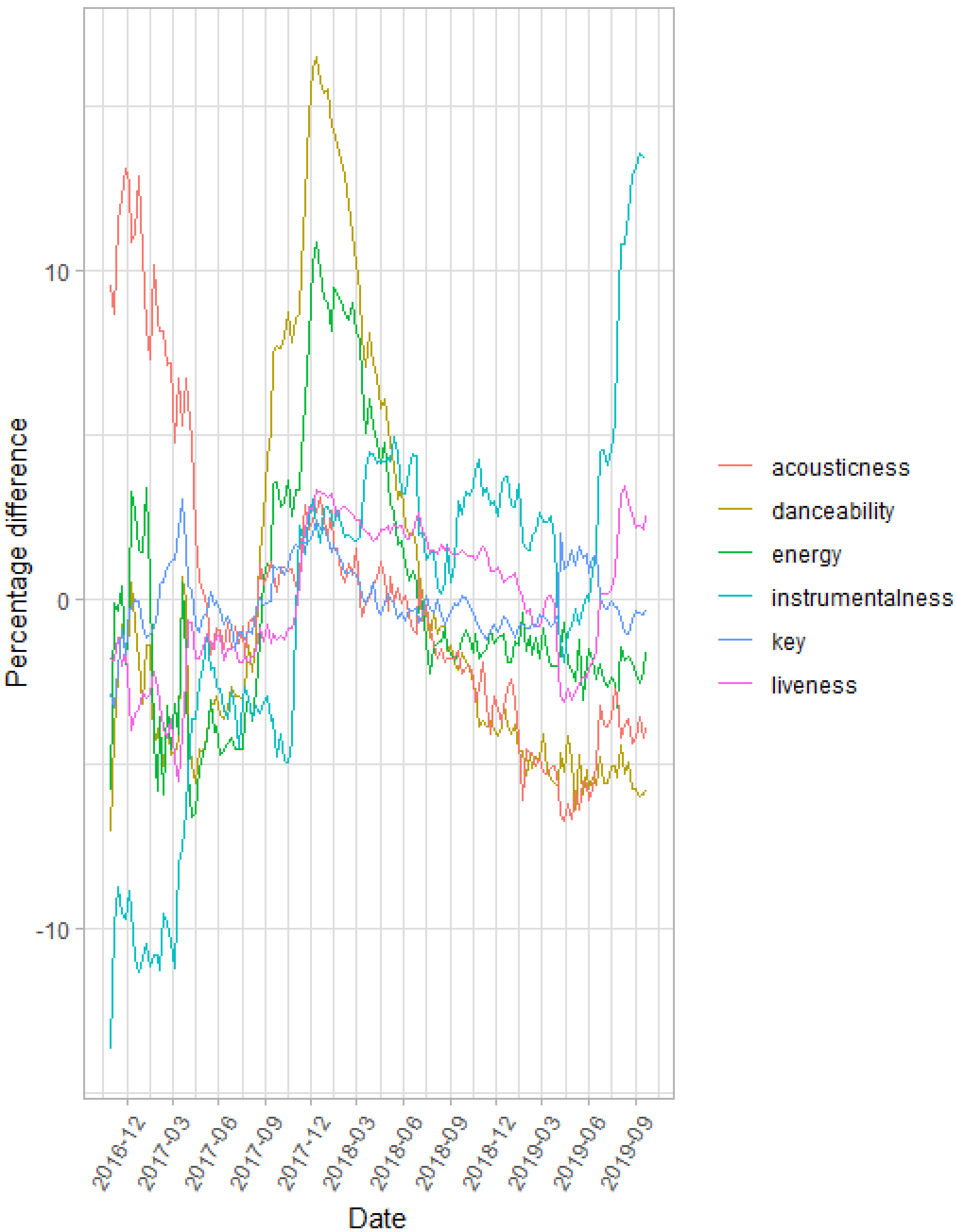


Figure B4

WITHIN-PLAYLIST VARIETY OF ACOUSTIC ATTRIBUTES OVER TIME (PART II/II)

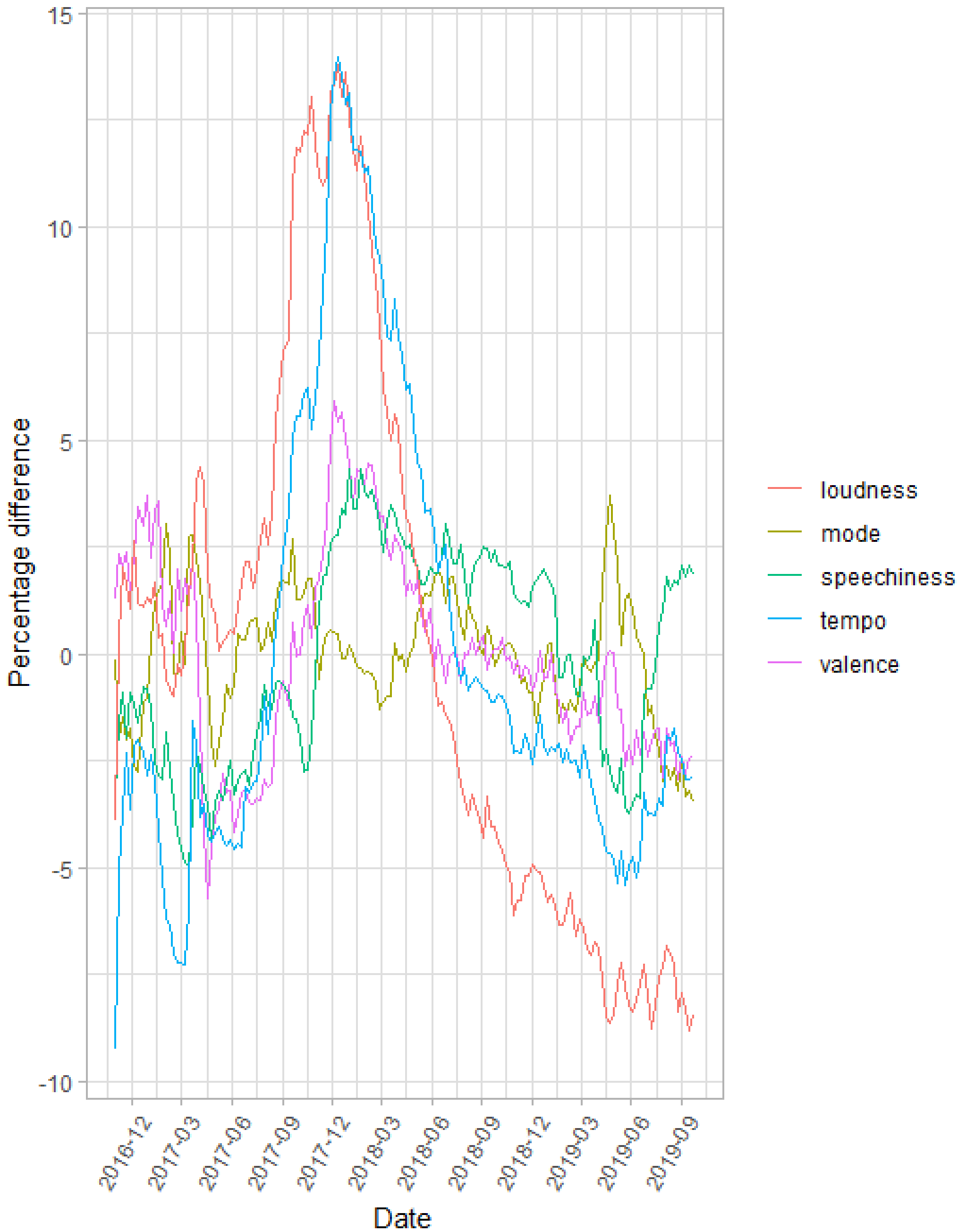


Figure B5

TRANSITION VARIETY OF ACOUSTIC ATTRIBUTES OVER TIME (PART I/II)

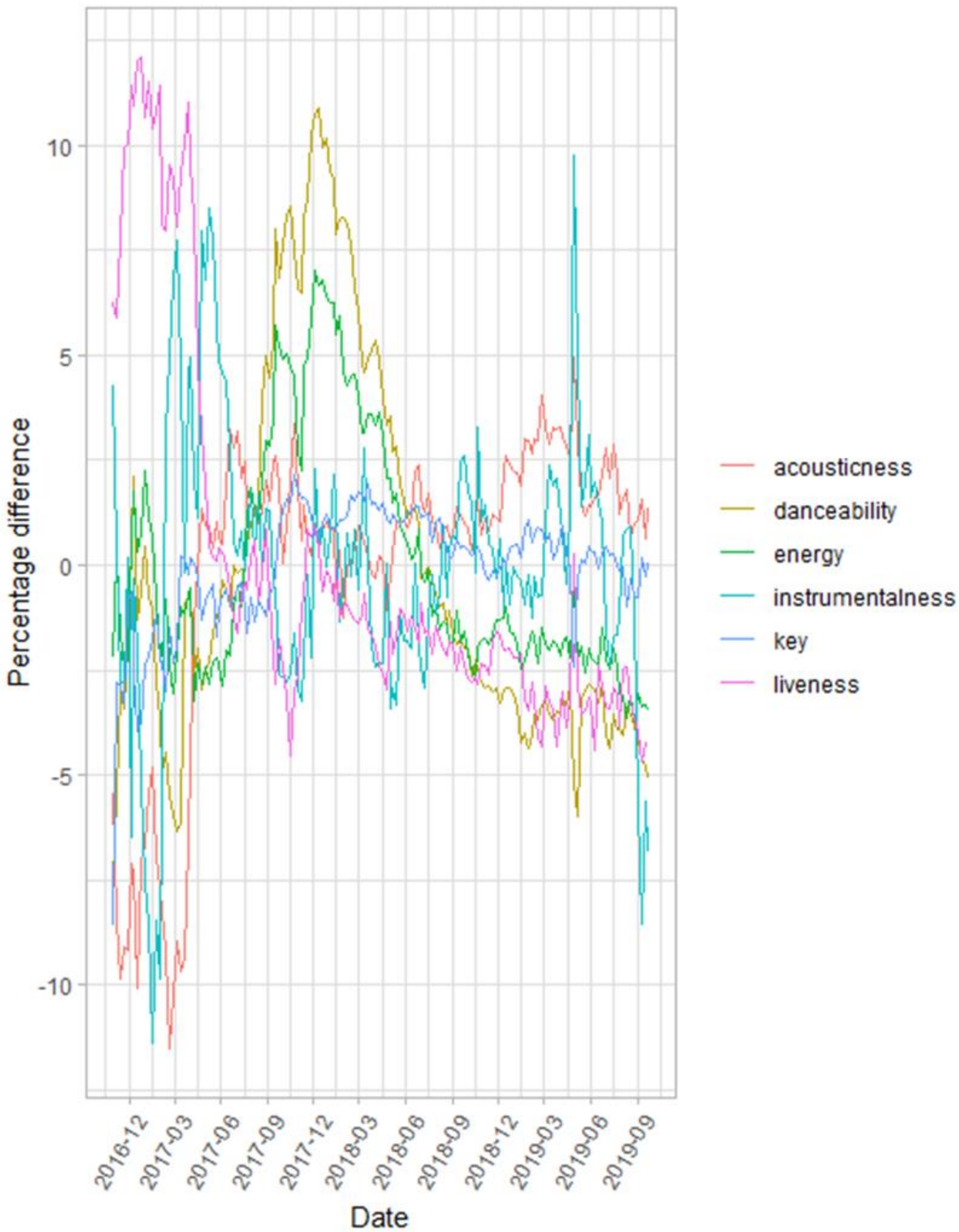
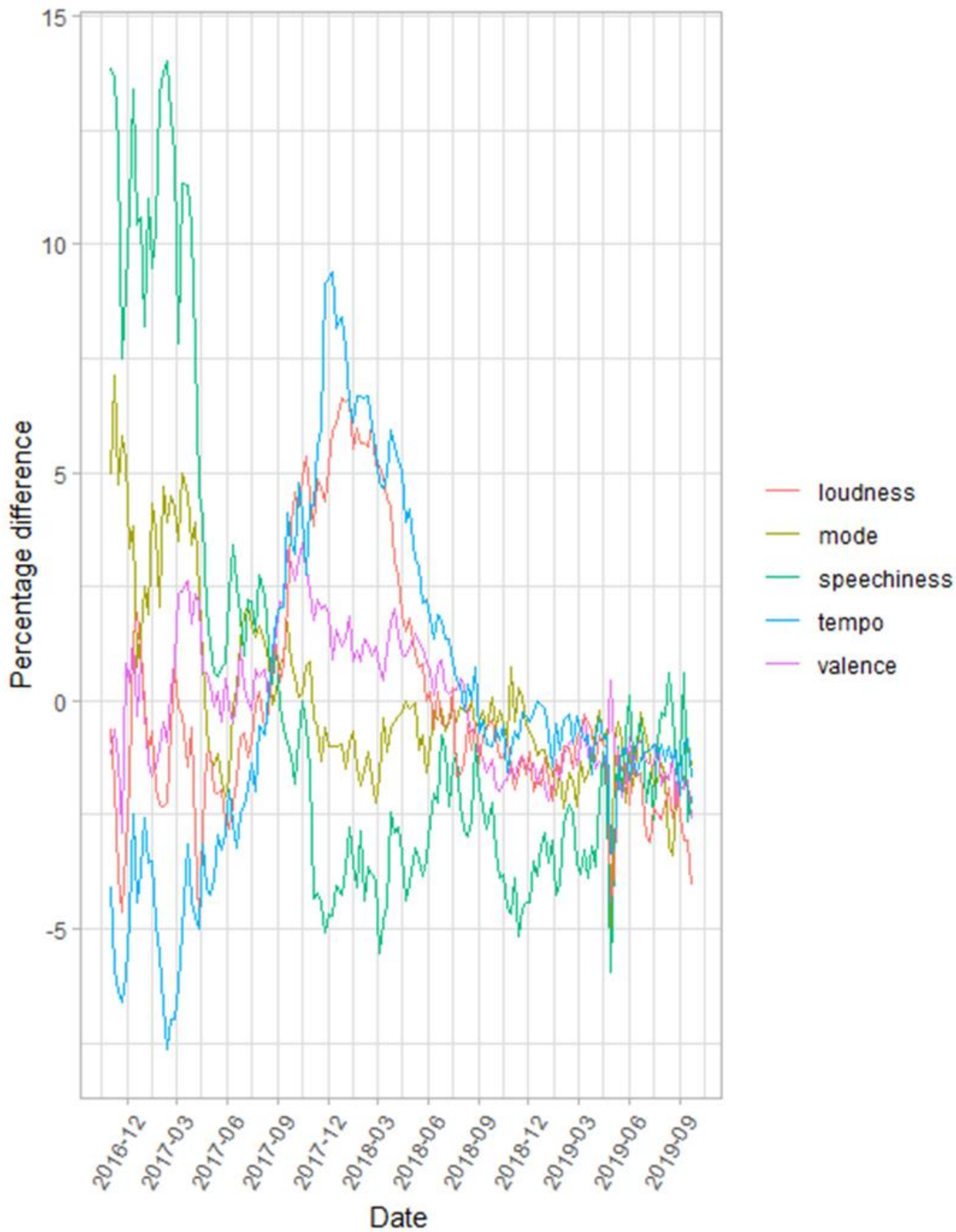


Figure B6

TRANSITION VARIETY OF ACOUSTIC ATTRIBUTES OVER TIME (PART II/II)



APPENDIX C: TRACKS PLACEMENT

In the plots below, the Y-axis features the days difference of a track in a playlist. This measure is calculated as follows:

$$Days\ difference = Date\_added_{t,p} - Date\_added_p$$

where $Date\_added_{t,p}$ = the date track t was added to playlist p,

and $Date\_added_p$ = the average date tracks were added to playlist p.

This measure is then grouped by track position (Figure 4), by track position and age (Figure C1), by track position and content type of the playlist (Figure C2), or by track position and curator type of the playlist (Figure C3). For example: a track t was added to playlist p on position 1 on April 5, 2017. The average date tracks were added to playlist p is April 3, 2017. This would result in a data value (x,y) = (1, 2). Then per position (x), the average difference in days (y) is calculated and plotted in the graph. Thus, if days difference > 0 for a certain track position, tracks on this position are typically newer to the playlist than the average.

Figure C1

DIFFERENCE IN DAYS BY TRACK POSITION AND AGE

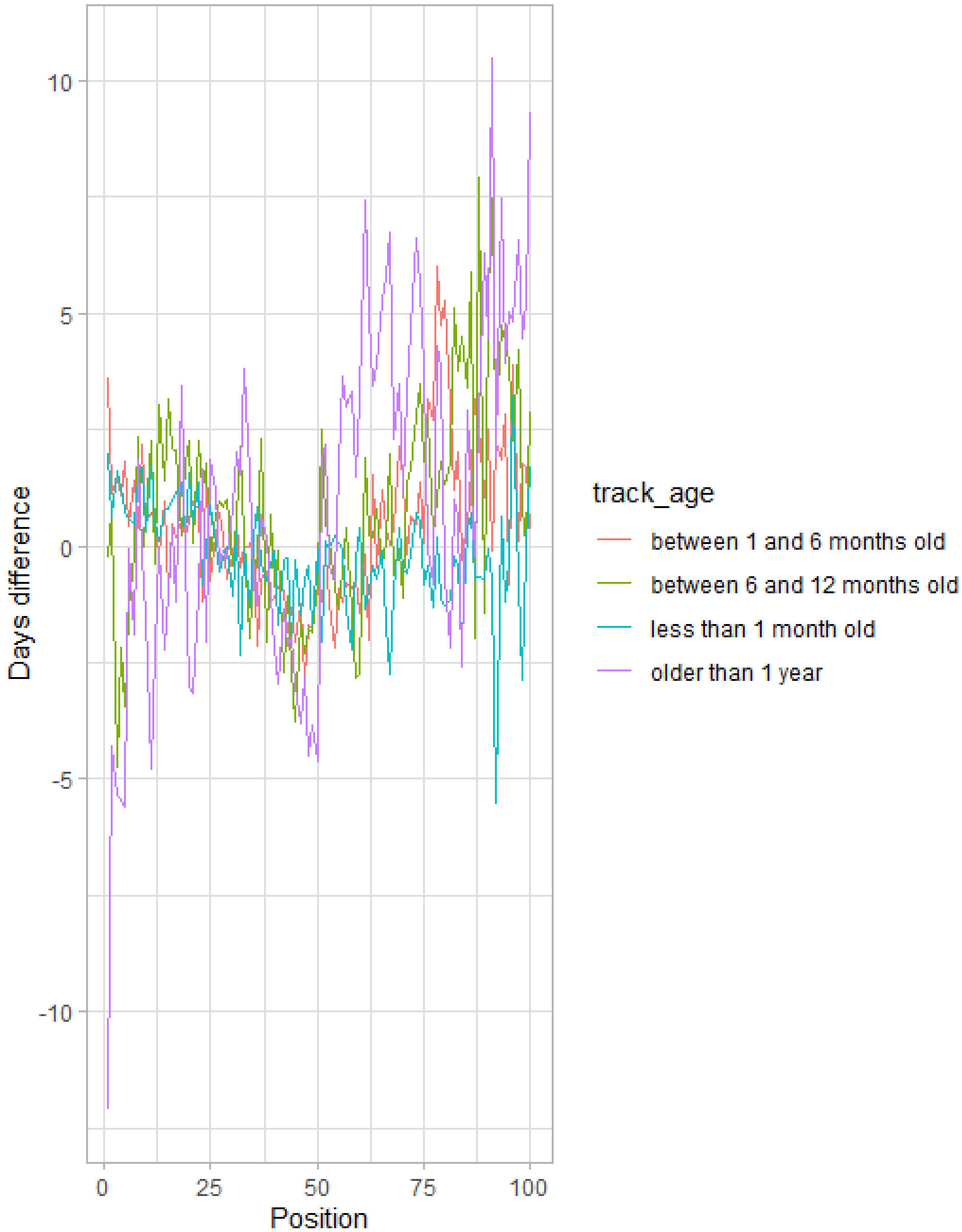


Figure C2

DIFFERENCE IN DAYS BY TRACK POSITION AND CONTENT TYPE OF THE PLAYLIST

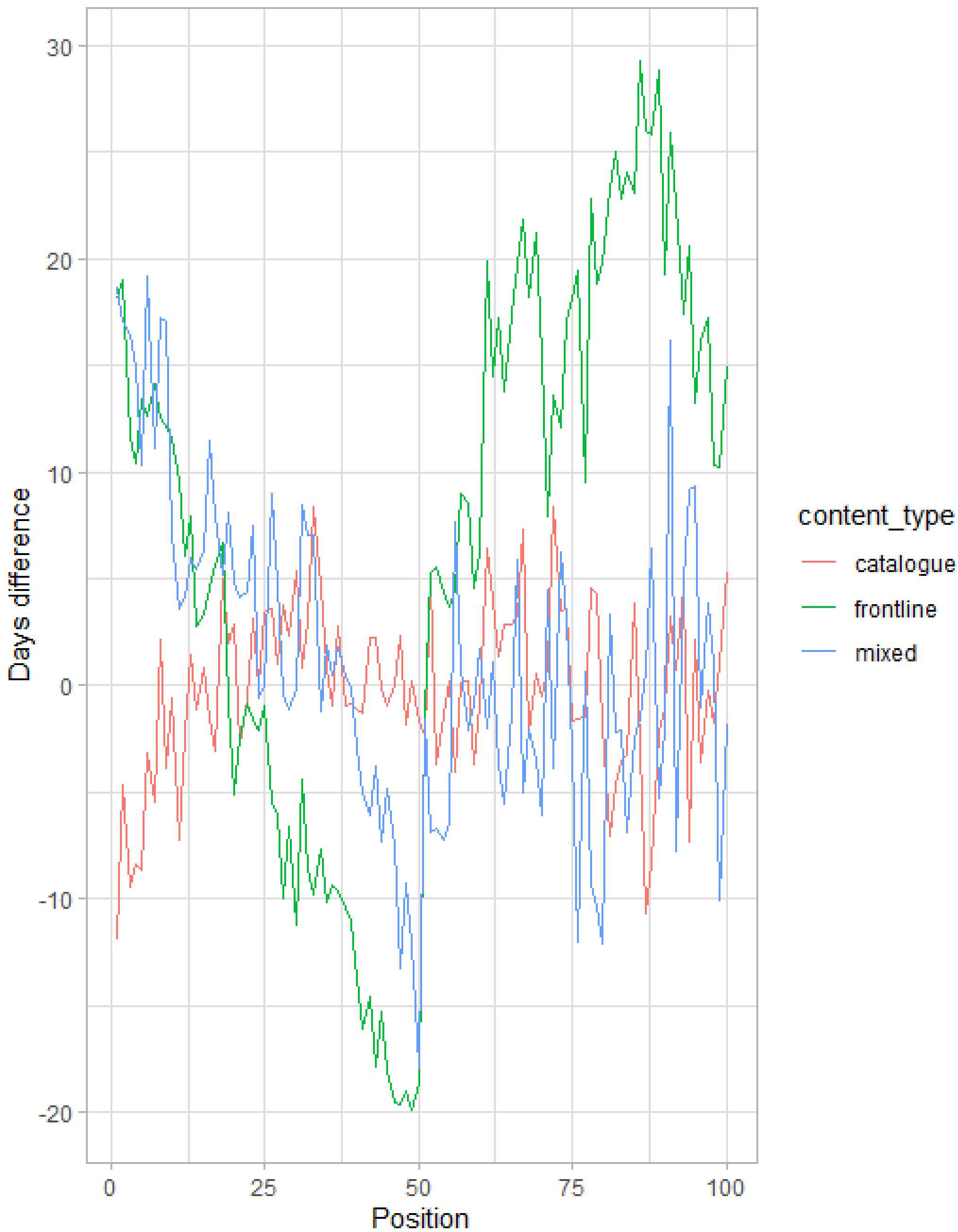


Figure C3

DIFFERENCE IN DAYS BY TRACK POSITION AND OWNER CLASS OF THE PLAYLIST

APPENDIX D: COMPLETE RESULTS OF THE WITHIN-PLAYLIST VARIETY MODEL

	Dependent variable:
	Relative followers
Date	0.001*** (0.00000)
Key	-0.084*** (0.014)
Mode	0.025*** (0.006)
Danceability	-0.247*** (0.027)
Energy	0.061*** (0.017)
Speechiness	-0.074*** (0.008)
Acousticness	0.006 (0.004)
Instrumentalness	-0.003*** (0.001)
Liveness	0.116*** (0.014)
Valence	0.127*** (0.013)
Tempo	0.052** (0.023)
Loudness	-0.126*** (0.012)
Artists	0.0001*** (0.00003)
SpotifyHuman	0.202*** (0.012)
MajorLabel	0.358*** (0.018)
SpotifyAI	-0.031** (0.014)
SpotifyPersonalized	0.166*** (0.013)
Key^2	0.036*** (0.006)
$Mode^2$	-0.001 (0.001)
$Danceability^2$	0.259*** (0.022)
$Energy^2$	0.021*** (0.006)
$Speechiness^2$	0.048*** (0.003)
$Acousticness^2$	-0.002*** (0.0004)
$Instrumentalness^2$	0.0005*** (0.0001)
$Liveness^2$	0.010 (0.006)
$Valence^2$	-0.062*** (0.003)
$Tempo^2$	-0.138*** (0.020)
$Loudness^2$	-0.021*** (0.004)
$Artists^2$	0.00000* (0.00000)
Key × SpotifyHuman	-0.032*** (0.011)
Key × MajorLabel	0.135*** (0.017)
Key × SpotifyAI	-0.023* (0.012)
Key × SpotifyPersonalized	0.082*** (0.011)
Mode × SpotifyHuman	-0.014*** (0.005)

Mode × MajorLabel	-0.127^{***} (0.007)
Mode × SpotifyAI	-0.008 (0.006)
Mode × SpotifyPersonalized	-0.042^{***} (0.005)
Danceability × SpotifyHuman	-0.079^{***} (0.026)
Danceability × MajorLabel	0.422^{***} (0.039)
Danceability × SpotifyAI	0.467^{***} (0.031)
Danceability × SpotifyPersonalized	0.133^{***} (0.026)
Energy × SpotifyHuman	-0.072^{***} (0.017)
Energy × MajorLabel	-0.114^{***} (0.027)
Energy × SpotifyAI	-0.234^{***} (0.021)
Energy × SpotifyPersonalized	-0.148^{***} (0.017)
Speechiness × SpotifyHuman	-0.034^{***} (0.006)
Speechiness × MajorLabel	-0.055^{***} (0.010)
Speechiness × SpotifyAI	0.066^{***} (0.007)
Speechiness × SpotifyPersonalized	-0.011^{*} (0.006)
Acousticness × SpotifyHuman	0.013^{***} (0.004)
Acousticness × MajorLabel	0.021^{***} (0.006)
Acousticness × SpotifyAI	0.014^{***} (0.004)
Acousticness × SpotifyPersonalized	-0.017^{***} (0.004)
Instrumentalness × SpotifyHuman	0.002^{***} (0.001)
Instrumentalness × MajorLabel	-0.026^{***} (0.001)
Instrumentalness × SpotifyAI	0.007^{***} (0.001)
Instrumentalness × SpotifyPersonalized	0.007^{***} (0.001)
Liveness × SpotifyHuman	-0.152^{***} (0.010)
Liveness × MajorLabel	-0.169^{***} (0.015)
Liveness × SpotifyAI	0.023^{**} (0.011)
Liveness × SpotifyPersonalized	-0.130^{***} (0.010)
Valence × SpotifyHuman	0.020 (0.014)
Valence × MajorLabel	-0.402^{***} (0.021)
Valence × SpotifyAI	-0.255^{***} (0.017)
Valence × SpotifyPersonalized	-0.054^{***} (0.014)
Tempo × SpotifyHuman	0.087^{***} (0.023)
Tempo × MajorLabel	-0.003 (0.030)
Tempo × SpotifyAI	0.018 (0.027)
Tempo × SpotifyPersonalized	-0.093^{***} (0.023)
Loudness × SpotifyHuman	0.051^{***} (0.012)
Loudness × MajorLabel	-0.537^{***} (0.016)
Loudness × SpotifyAI	0.160^{***} (0.012)

Loudness × SpotifyPersonalized	0.102*** (0.012)
Artists × SpotifyHuman	-0.0003*** (0.00003)
Artists × MajorLabel	-0.002*** (0.0001)
Artists × SpotifyAI	-0.0002** (0.0001)
Artists × SpotifyPersonalized	0.0004*** (0.00003)
Constant	-16.702*** (0.024)
R^2	0.570
Adjusted R^2	0.570
Residual Std. Error	0.234 (df = 645745)
F Statistic	11,135.230*** (df = 77; 645745)
Observations	645,823

Notes. The table shows a regression with standard errors in parentheses. Estimates are calculated on the aggregated within-playlist variety data set as described in paragraph 4.3.2. The dependent variable is the ratio of relative followers of a playlist on a certain day. The independent variables are the 12 dimensions of variety including 11 acoustic features and artists. The acoustic features are measured by the coefficient of variation and artists is measured by the number of uniques. The multiple regression model includes both their linear and quadratic terms. Furthermore, it includes the interaction effects of the 12 dimensions of variety with the playlist curator. The playlist curators are classified through several steps as outlined in paragraph 4.3.3. In this model, they are included via dummy variables which are equal to 1 if the playlist is owned by the respective curator type.

*p<0.1; **p<0.05; ***p<0.01

APPENDIX E: COMPLETE RESULTS OF THE TRANSITION VARIETY MODEL

	Dependent variable:
	Relative followers
Date	0.001^{***} (0.00000)
Key	-0.038^{***} (0.002)
Mode	0.088^{***} (0.010)
Danceability	0.062^{*} (0.036)
Energy	0.261^{***} (0.033)
Speechiness	0.165^{***} (0.032)
Acousticness	-0.460^{***} (0.016)
Instrumentalness	-0.140^{***} (0.011)
Liveness	-0.053^{***} (0.020)
Valence	-0.393^{***} (0.026)
Tempo	0.00003 (0.0001)
Loudness	0.027^{***} (0.001)
SpotifyHuman	-0.043^{***} (0.010)
MajorLabel	-0.187^{***} (0.014)
SpotifyAI	-0.238^{***} (0.011)
SpotifyPersonalized	-0.089^{***} (0.010)
Key^2	0.003^{***} (0.0001)
Mode^2	-0.046^{***} (0.007)
Danceability^2	-0.101^{*} (0.054)
Energy^2	0.071^{*} (0.042)
Speechiness^2	-0.636^{***} (0.082)
Acousticness^2	0.165^{***} (0.017)
$\text{Instrumentalness}^2$	0.253^{***} (0.014)
Liveness^2	-0.050^{**} (0.024)
Valence^2	-0.415^{***} (0.031)
Tempo^2	0.00001^{***} (0.00000)
Loudness^2	0.0004^{***} (0.0001)
Key × SpotifyHuman	0.013^{***} (0.001)
Key × MajorLabel	0.006^{***} (0.002)
Key × SpotifyAI	0.025^{***} (0.002)
Key × SpotifyPersonalized	0.013^{***} (0.001)
Mode × SpotifyHuman	-0.036^{***} (0.009)
Mode × MajorLabel	-0.019 (0.012)
Mode × SpotifyAI	-0.064^{***} (0.010)

Mode × SpotifyPersonalized	-0.004 (0.009)
Danceability × SpotifyHuman	0.022 (0.032)
Danceability × MajorLabel	0.336*** (0.047)
Danceability × SpotifyAI	-0.043 (0.037)
Danceability × SpotifyPersonalized	-0.009 (0.033)
Energy × SpotifyHuman	-0.420*** (0.030)
Energy × MajorLabel	0.136*** (0.044)
Energy × SpotifyAI	-0.274*** (0.035)
Energy × SpotifyPersonalized	-0.355*** (0.030)
Speechiness × SpotifyHuman	-0.054* (0.029)
Speechiness × MajorLabel	-0.222*** (0.040)
Speechiness × SpotifyAI	-0.112*** (0.034)
Speechiness × SpotifyPersonalized	-0.088*** (0.030)
Acousticness × SpotifyHuman	0.395*** (0.015)
Acousticness × MajorLabel	0.069*** (0.023)
Acousticness × SpotifyAI	0.240*** (0.017)
Acousticness × SpotifyPersonalized	0.430*** (0.015)
Instrumentalness × SpotifyHuman	0.004 (0.010)
Instrumentalness × MajorLabel	-0.135*** (0.025)
Instrumentalness × SpotifyAI	0.017 (0.013)
Instrumentalness × SpotifyPersonalized	0.011 (0.010)
Liveness × SpotifyHuman	-0.060*** (0.019)
Liveness × MajorLabel	0.551*** (0.027)
Liveness × SpotifyAI	0.173*** (0.021)
Liveness × SpotifyPersonalized	-0.021 (0.019)
Valence × SpotifyHuman	0.544*** (0.023)
Valence × MajorLabel	0.482*** (0.032)
Valence × SpotifyAI	0.775*** (0.026)
Valence × SpotifyPersonalized	0.631*** (0.023)
Tempo × SpotifyHuman	-0.001*** (0.0001)
Tempo × MajorLabel	-0.001*** (0.0002)
Tempo × SpotifyAI	-0.001*** (0.0002)
Tempo × SpotifyPersonalized	-0.001*** (0.0001)
Loudness × SpotifyHuman	-0.037*** (0.001)
Loudness × MajorLabel	0.002 (0.002)
Loudness × SpotifyAI	-0.040*** (0.002)
Loudness × SpotifyPersonalized	-0.037*** (0.001)
Constant	-16.050*** (0.022)

R^2	0.557
Adjusted R^2	0.557
Residual Std. Error	0.231 (df = 631015)
F Statistic	11,192.310*** (df = 71; 631015)
Observations	631,087

Notes. The table shows a regression with standard errors in parentheses. Estimates are calculated on the aggregated transition variety data set as described in paragraph 4.3.2. The dependent variable is the ratio of relative followers of a playlist on a certain day. The independent variables include the 11 acoustic features as dimensions of variety. These features are measured by the mean absolute deviation of consecutive songs. The multiple regression model includes both their linear and quadratic terms. Furthermore, it includes the interaction effects of the 11 dimensions of variety with the playlist curator. The playlist curators are classified through several steps as outlined in paragraph 4.3.3. In this model, they are included via dummy variables which are equal to 1 if the playlist is owned by the respective curator type.

APPENDIX F: TEST FOR MULTICOLLINEARITY

Table F1

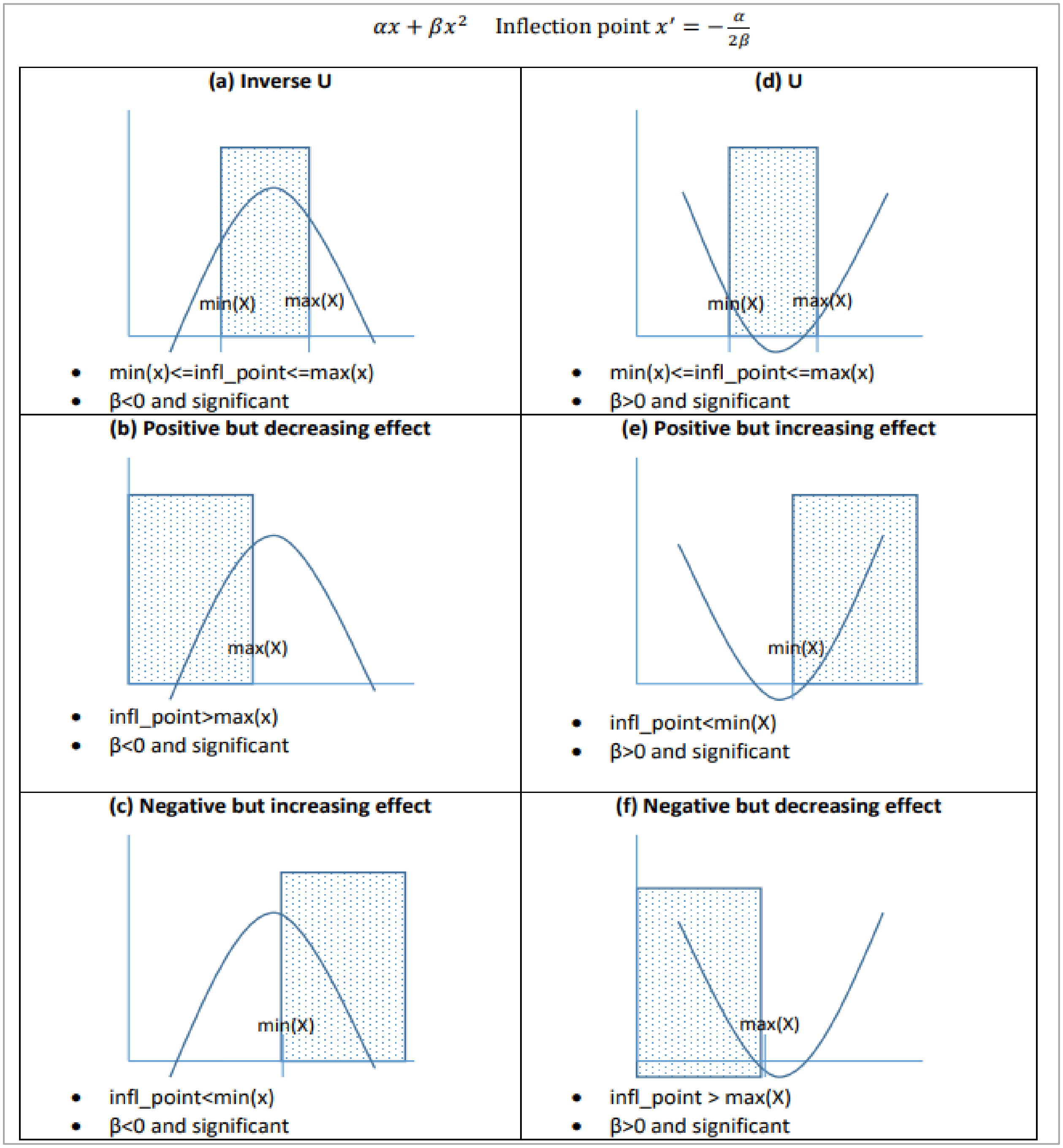
MULTICOLLINEARITY TESTS USING VIF VALUES FOR BOTH MODELS

	Within-playlist variety model	Transition variety model
Statistic	**VIF**	**VIF**
Date	1.080	1.030
Key	1.130	1.020
Mode	1.102	1.122
Danceability	4.843	1.541
Energy	3.683	1.743
Speechiness	1.157	1.196
Acousticness	1.683	1.414
Instrumentalness	1.501	1.270
Liveness	1.090	1.084
Valence	2.962	1.271
Tempo	2.727	1.396
Loudness	1.341	1.507
Artists	1.274	
CuratorType	1.473	1.179

APPENDIX G: CLASSIFICATION OF THE EFFECTS OF THE VARIETY DIMENSIONS ON PLAYLIST SUCCESS

Figure G1

CLASSIFICATION OF NON-LINEAR EFFECTS



Notes. As shown in Chapter 5, our model uses β and γ to denote the effects of the linear and quadratic terms respectively. Thus, when interpreting this figure, α is replaced by β and β is replaced with γ. For example, the inflection point $x' = -\frac{\beta}{2\gamma}$ in our model. Adapted from H. Datta (personal communication, December 16, 2019)

Table G1

TESTING FOR NON-LINEAR EFFECTS – WITHIN-PLAYLIST VARIETY MODEL

Variety dimension	β	γ	Inflection point	Min.	Max.	Classification
Acousticness	.006[a]	-.002	1.374	.000	9.737	Insignificant
Danceability	-.247	.259	.477	.000	2.387	U-shape
Energy	.061	.021	-1.411	.000	2.162	Positive but increasing effect
Instrumentalness	-.003	.0005	3.536	.000	14.375	U-shape
Key	-.084	.036	1.171	.000	2.000	U-shape
Liveness	.116	.010[a]	-5.850	.000	1.663	Positive linear effect
Loudness	-.126	-.021	-2.940	-3.316	.000	Inverse U-shape
Mode	.025	-.001[a]	16.534	.000	4.899	Positive linear effect
Speechiness	-.074	.048	.768	.000	2.355	U-shape
Tempo	.052	-.138	.187	.000	2.216	Inverse U-shape
Valence	.127	-.062	1.030	.000	4.072	Inverse U-shape
Artists	.0001	.0000001[a]	-497.062	1.000	721.000	Positive linear effect

[a]This effect is insignificant at a 0.05 level.

Table G2

TESTING FOR NON-LINEAR EFFECTS – TRANSITION VARIETY MODEL

Variety dimension	β	γ	Inflection point	Min.	Max.	Classification
Acousticness	-.460	.166	1.388	.000	.990	Negative but decreasing effect
Danceability	.062a	-.101a	.305	.000	.859	Insignificant
Energy	.261	.071a	-1.840	.000	.933	Positive linear effect
Instrumentalness	-.140	.253	.277	.000	.983	U-shape
Key	-.038	.003	6.731	.000	11.000	U-shape
Liveness	-.053	-.050	-.532	.000	.925	Negative but increasing effect
Loudness	.027	.0004	-34.442	.000	23.573	Positive but increasing effect
Mode	.088	-.046	.952	.000	1.000	Inverse U-shape
Speechiness	.165	-.636	.130	.000	.878	Inverse U-shape
Tempo	.00003a	.00001	-1.538	.000	186.000	Insignificant
Valence	-.393	-.415	-.474	.000	.868	Negative but increasing effect

[a]This effect is insignificant at a 0.05 level.

APPENDIX H: VISUALIZATION OF THE MODERATING EFFECT OF THE TYPE OF PLAYLIST CURATOR ON THE RELATIONSHIP BETWEEN VARIETY AND PLAYLIST SUCCESS

Figure H1

EXAMPLES OF THE MODERATING EFFECT OF THE PLAYLIST CURATOR