‘Any suggestions?’

A quantitative within-subject experiment study on the effect of different filter types on the consumer utility of music recommendation

Student Name: Zoë Voyle
Student Number: 431764

Supervisor: Dr. João Ferreira Gonçalves
Second reader: Dr. Julia Kneer

Media, Culture & Society (MCS)
Erasmus School of History, Culture and Communication
Erasmus University Rotterdam

Master Thesis
July, 2019

‘Any suggestions?’

A quantitative within-subject experiment study on the effect of different filter types on the consumer utility of music recommendation

Abstract

Nowadays a large portion of music consumption is done via streaming. Platforms that enable music streaming have a hand in what is recommended and through which method, and therefore have a hand in what is consumed. This in turn results in undemocratic recommendation methods. Consumers are subsequently left largely unaware of the processes behind their song recommendations on music streaming platforms. This research has aimed to dissect the processes behind song recommendation via separating Content Based Filtering (CBF) and Collaborative Filtering (CF) mechanisms. Dissecting these mechanisms was necessary to assess whether the different filter methods resulted in different levels of consumer utility and to what extent. Consumer utility was assessed via looking at respondents' likelihood to listen to filter recommended songs present in the within-subject experiment survey, and whether these songs fit within their taste. In order to test this a total of 302 participants took part in the survey of this study. Within the survey the respondents could pick one out of five music genre paths to go down, within these paths they came across five songs, two recommended via CBF, two recommended via CF, and one Randomly recommended. The CBF recommendations were made via attributing and cross referencing meta-data the those respective songs. CF recommendations were created via utilising prewritten code for the Million Song Dataset. Random recommended songs were recommended via generating a random number and corresponding that number with music chart positions. Results indicate that there is a difference in consumer utility between CBF and CF recommended songs. In fact, this research concludes that in general consumers have a preference towards CBF recommended songs. Alongside this the study found no significant results supporting the hypotheses that different listener types, heavy or light, would result in a preference towards songs recommended by a particular filter type. The higher consumer utility from CBF recommendations would infer that within the field of music, experts and curators are still much needed on the production side and much appreciated on the consumption side.

Keywords: Music recommendation, Content based filtering, Collaborative filtering, Algorithms, Consumer utility

Table of Content

Chapter 1. Introduction

Streaming has become an integral part of the consumption of online entertainment. Streaming takes up this pivotal position as an affordance of consumption because through streaming online data is transmitted directly to the user, without the need to download it (Ciocca, 2017; Ricci, Rokach, Shapira & Kantor, 2011). If we take Spotify as an example, the platform is primarily used by consumers for streaming songs online. Consequently, whatever these consumers are listening to and the frequency hereof is documented (Burke, 2007; Ciocca, 2017; Ricci et al. 2011). Platforms and streaming services such as Spotify in turn analyse this user data with the aim to maximise consumer utility. In order to be able to make conclusions and predictions about such utility it is in the streaming platforms' interest to add as much metadata to its available content as possible. As a result, the consumption patterns can be analysed on the basis of said metadata. This is mostly done via algorithms. There are two main ways that such algorithms can work by: Content Based Filtering (CBF) or Collaborative Filtering (CF). Content Based Filtering is done via the back end. This infers, that it is not the producers (record labels/bands/artists/etc.) of the content themselves per se that label the content; but it is the platforms, the facilitators that do so. For example, streaming platforms such as Apple Music, Deezer, and Spotify can decide to attribute both objective and subjective metadata to its content. Objective metadata would entail things as the nationality of the artist and/or band, release year, etc. Subjective metadata would cover aspects such as the possible mood of a song and whether the song contains explicit content or not. Having worked at a music company Z (which will remain anonymous due to confidentiality reasons) my first hand anecdotal work experience ongoing for about two years and further supported by theory and research done in this thesis, I have observed that record labels tend to add mainly objective metadata to their content and it is the facilitators that add the more subjective metadata. In other words, it is the facilitators and their curators that are in charge of assigning metadata, be it objective or subjective, and not the consumers. By adding such characteristics the streaming platforms are able to filter their content and make recommendations on the basis of comparison with other songs, artists, and albums which contain similar or perhaps even the same attributed metadata.

Collaborative Filters, on the other hand, are generated via consumers. Collaborative Filtering occurs through collecting and tracking the consumption behaviour/patterns of consumers. By doing so, the consumption choices can be mapped. With insight on more "individualistic" consumer choices, recommendations can be made via comparing this data to the data of other consumers with similar past consumption choices (Burke, 2007; Logg, 2017; Prey, 2017; Ricci et al. 2011). It is important to note that CF cannot be achieved on its own as it needs a basis to start from, which can be minimal content based metadata.

In reality, streaming platforms do not make use of just one of these two, i.e. CF and CBF strategies, instead they typically employ hybrid recommender systems (Burke, 2007; Ricci et al. 2011). The algorithms of such platforms have become increasingly skilled in combining both and other methods for content recommendation. Although these filtering mechanisms seem inherently intertwined, this thesis aims to examine whether there would be a difference in consumer utility if these filtering methods, CBF or CF were isolated. This brings us to the research question: *To what extent might either Content Based Filtering or Collaborative Filtering result in the highest consumer utility on music streaming platforms?*

If this research demonstrates that one type of filtering results in a higher consumer utility than the other, this might have fruitful implications for future recommendation of music content. However, larger streaming platforms tend not to be transparent in the ways their recommendation systems are run. It is, of course, in their interest to have the best coded algorithms and most finetuned datasets in their competitive markets. Furthermore, it is also in their interest to keep this information as confidential as possible in order to remain competitive within their markets. According to The Recording Industry Association of America, "[o]verall market trends in the first half of 2018 continued to reflect the music industry's rapid transition from unit based physical and digital sales towards streaming music sources (Friedlander & Bass, 2018, p. 1). Consequently, music streaming makes up 75% of total music sale revenue in the U.S. (Friedlander & Bass, 2018). Researching consumption choices following either CBF or CF based recommendations would shed light on the relevance and accuracy of both types. Furthermore, such an analysis can help unveil what is happening on the backend of such streaming platforms and how their algorithms might be run, and which strategies might be employed.

The results of this study aim to clarify whether aggregated consumer preferences and behaviour actually lead to better product recommendation and consumer utility. For example, if it is the case that Content Based Filtering is preferred, it would infer that those at the backend of the facilitating platforms are better at assessing what consumers would enjoy than users who share similar consumption choices. Such information can in turn be fruitful for smaller businesses, so that they know whether to invest more in curating efforts (CBF) or audience expansion efforts (CF).

This thesis is organised as follows: Chapter 2 provides an overview of the related literature related and relevant to this research following the literature, this chapter introduces the hypotheses of this study. Chapter 3 explains the methods used to conduct this research and provides a sample description. Chapter 4 reports and analyses the results of the statistical analyses done in order to accept or reject the hypotheses. Finally, Chapter 5 concludes and proposes future research directions.

Chapter 2. Theoretical Framework

In this chapter I will expound upon existing literature surrounding recommender systems, consumer decision-making processes, and consumer utility. This paper specifically focuses on the decision-making processes for selecting songs for the consumer. These processes are further explored through three different lenses: the (platform dependent) song availability, the consumer's taste, and their mood. Furthermore, this section will discuss the contributing factors to both parties - the recommender and the consumer - that may have an impact on the consumer's consumption choices. It is also important to highlight and frame this theoretical framework within the field of online music consumption, specifically streaming, especially when almost three quarters of music revenue is made through streaming (Friedlander & Bass, 2018). It is necessary to dissect the mechanics and the contributing factors behind these processes in order to answer the main research question of this thesis, as this research ultimately aims to discover which filter method, Content Based Filtering or Collaborative Filtering, results in the highest consumer utility.

2.1 Recommender types

Before exploring which filter method could lead to a higher consumer utility, it is important to distinguish the different kinds. According to Ricci et al. (2011) and Burke (2007) there are four different classes of recommendation techniques based upon knowledge source:

There are four main types of filter methods worth highlighting for this research, the first being Collaborative Filtering. With CF the recommender system generates output by looking at information based solely on the profiles for different users; it locates peer users with similar consumption patterns and decisions and consequently generates the recommendation based on the highest degree of similarities. An example of this is the existence of music charts. These charts exist through so-called people-to-people correlation (Ricci et al., 2011); the more people listen to the same song, the more likely others will enjoy it too. Consequently charts can be made which serve as excellent recommenders due to high consumer utility (Ricci et al., 2011) – the concept consumer utility will be expanded on later. The second type of filter method is Content Based Filtering. With this filter type the recommender generates output coming from two sources: the meta-data of products and that of the user ratings. CBF recommendations are based on user-specific classification problems and learn from users' own consumption behaviour and choices. For example, if a user frequently listens to Panic! At The Disco, a band with the following characteristics: pop/rock and alternative/indie rock, American, male vocalist, active from early 2000's (AllMusic, n.d.); then based on these characteristics this user is likely to also enjoy Fall Out Boy, a band with very similar characteristics. The third method to be discussed is the Demographic recommender type, this type provides recommendation based upon the demographic profile of the user. The generated recommendations cater to demographic niches by combining the ratings of users within them, as a form of clustering. For example, if on YouTube a user has provided profile information stating that they are below the age of 18 some explicit content on the platform will not available or recommended to them (Ricci et al., 2011). The last filter mention worth mentioning is the Knowledge-Based recommender generates output via the construed user's needs and preferences. Here the recommender generates output by matching problem description with solutions. For instance, if a music streaming platform that utilises this technique is connected to a user's digital agenda it can make specific recommendations based upon the type of activity - i.e. if the agenda lists going to the gym as an activity, the knowledge-based recommender system can thus recommend upbeat and high-tempo music (Ricci et al., 2011).

This research however is mostly concerned with the first two recommendation techniques mentioned above: Collaborative and Content Based. In practice CBF and CF recommenders are often mixed, which results in hybrid recommender systems; when hybridised, this type makes recommendations via combining user data from CF and product meta-data from CBF (Burke, 2007; Logg, 2017; Ricci et al., 2011). Most scholars tend to only explore or emphasise the effectiveness of CF filters (Herlocker, Konstan, Terveen & Riedl, 2004; Logg, Minson, Moore, 2019; Sarwar, Karypis, Konstan & Riedl, 2001; Sundar, Oeldorf-Hirsch & Xu, 2008). This effectiveness is usually connected to CF's ability to document and track a large number of consumer behaviour, thus recommending products based upon this aggregated information.

However, as Yeomans, Shah, Mullainathan & Kleinberg (2019) rightfully point out, humans possess a wealth of knowledge that recommender systems do not have. CBF recommendations are in general based on the knowledge that humans have and the metadata that they have attributed to the products (Pazzani & Billsus, 2007; Yeomans et al., 2019). By contrast, CF recommenders solely operate on objective data collected through user tracking (Logg, Minson, Moore, 2019). As Yeomans et al (2019) succinctly point out: CF recommenders "only know *what* we like, not *why* we like it" (p. 4). Human recommendations, largely done through CBF, are able to take contextuality and more subjective estimations into consideration when making recommendations. It is still necessary to note that both processes are done via digital algorithms, however CBF recommendations include more human input than CF recommendations.

As mentioned earlier, these two filter types are commonly hybridised (Burke, 2007; Ricci et al. 2011). Moreover, research on the subject tends to focus on the two recommender techniques independent of each other (Herlocker et al., 2004; Pazzani & Billsus, 2007; Sarwar et al., 2001; Sundar et al., 2008; Yeomans et al., 2019), or solely describe the existing types of techniques and systems(Burke, 2007; Ricci et al. 2011). Thus, this research aims to dissect them in order to investigate whether there is a difference in consumer utility between CBF or CF recommendations. This leads to the first hypothesis of this research, *H1: There is a difference within consumer utility depending upon which recommender is used: either CBF or CF.*

To give a concrete example of how this works in reality, the following will serve as a succinct exploration of Spotify's recommendation models. According to Burke (2007, p. 380), the hybrid recommender system can be differentiated into seven different types: weighted, switching, mixed, feature combination, feature augmentation, cascade, and meta-level. Within music streaming platforms it is most likely that the mixed hybrid type is employed. This mixed types makes "[r]ecommendations from different recommenders [that] are presented together" (p. 380). Here from it can be deduced that Spotify uses a mixed hybridised recommendation technique as characterised by Burke (2007) and Ricci et al. (2011). Spotify uses three main types of recommendation models: Collaborative Filtering (CF), Natural Language Processing (NLP), and raw audio models (Ciocca, 2017). The CF that Spotify utilises works the same as described before. Second, the NLP model works on a semantic basis, by basing song suggestions upon matching song lyrics to track and/or artist related internet sources, such as blogs or news articles, and then match them against each other. The complex mechanisms behind NLP are beyond the scope of this paper, but it is still worth mentioning as Spotify is able to scour the web for written texts concerning the respective songs and/or artists. Attention is paid to semantic choices such as adjectives and reoccurring discourse. These recurring semantic analyses allow for *cultural vectors* or *top terms* to be determined (Ciocca, 2017). The reoccurrence of each term denotes the probability that the song and/or artist is related to them. Ultimately, "much like in collaborative filtering, the NLP model uses these terms and weights to create a vector representation of the song that can be used to determine if two pieces of music are similar" (Ciocca, 2017, para 31).

Third, the raw audio models bring a fruitful addition to the mix of recommendations. Such models are able to take new songs into account. The model matches the audio of one track in its database to the rest. Through convolutional neural networks, audio tracks can be identified and coded by what are called convolutional layers (Ciocca, 2017). After processing these layers will bring forward new recommendations based on the song's key characteristics and similarities. In theory, through the raw audio model, a new song with a low stream count and without any references on the internet, can thus become part of recommended songs based on its audio characteristics and similarities. Ciocca (2017) lists these recommendation models as three separate mechanisms. Though, deducing from the basis of their workings the NLP model and the raw audio model can both be attributed to Content Based Filtering. This is due to both of the models making recommendations starting from the metadata of a track and consequently matching it against other track metadata, without involving user data. Thus, it becomes apparent that Spotify makes use of a mixed hybridised recommendation technique by mixing CF and CBF methods (Burke, 2007; Ciocca, 2017; Ricci et al. 2011).

Algorithms currently in use on music streaming platforms can only learn from and improve by the consumption that takes place within their domain. Hence, user-tracking methods are employed to follow and consequently document the movements that consumers make. Prey (2017) labels this as algorithmic individuation. This process "should be understood as a dynamic socio-technical process engaged in enacting the individual" (p. 1095). This infers that one's musical profile, or user-identity, on a music streaming platforms is in constant development. Arguably, in the same way that consumers can become more knowledgeable and develop likings to new products over time, with increased use algorithms are able to facilitate and even estimate that as well.

A new or generic consumer on a music streaming platform will be faced with many choices on what to listen to first. In this stage most options for the new user will come from popularity driven recommendations. This is the case because a recommender system aims to recommend content with the biggest consumption probability and utility (Ricci et al., 2011). Therefore, popularity is a logical way to start and the utility of popular songs is estimated to be reasonably high (Ricci et al., 2011). Ricci et al. (2011) determine the utility value of a product, in this case a song, by the degree of appreciation – whether something is liked or not and to what extent. Thus, consumer utility is determined via combining the consumption probability - which is in turn based on other consumers , consumer profile – which is based on their profile settings and information, and on their previous consumption choices. Once the consumer will have made a choice, either by selecting a song recommended from the platforms catalogue or via explicitly searching for a song and/or artist that they want to listen to, the filter models start to kick in. Over time, following the consumer's long-term use of the platform the algorithms inevitably become more and more informed about the user's behaviour and taste. The logical assumption is that a more informed algorithm is a better – perhaps even "smarter" – algorithm. Consequently, the consumer will receive more fitting recommendations, related to their taste and if advanced enough even related to their mood. The goal of such a streaming platform is to know their consumers as well as possible. In turn the consumers can then ponder in awe "How does [streaming platform] know me so well". Collaborative filters compute the degree of utility for a specific product for a "user *u* for the item *i* as a (real valued) function *R(u, i)*" (Ricci et al., 2011, p. 10). Mathematically, Ricci et al. (2011) would calculate the degree of utility in the case of music consumption by considering the listener count. This is a fundamental process within CF recommender systems, they predict the estimated value ^*R* over pairs of users *u* and items *i*. Thus to compute the value of a specific music product, the formula would go as follows: ^R(u, i) = ^R (u, i1), …, ^R(u,iN). The system will recommend a (perhaps predetermined) limited amount of songs with the largest predicted utility *K*, and *K* is always smaller than the total amount (*N*) of songs available in the database (Ricci et al., 2011).

A Content Based Filter can generate recommendations for a new or generic user based upon the knowledge the recommender system has. This knowledge consists of product knowledge and user knowledge. Once both are combined, the algorithm will be able to generate utility predictions and subsequently recommendations. The CBF recommendations are thus still able to cross-reference existing product labels combined with the consumer choices of other users in order to make a recommendation without needing any or little input from the user who is requesting a product (Logg, 2017; Logg, Minson, Moore, 2019; Ricci et al., 2011).

2.3 Consumer decision making

2.3.1 Consideration set

For a product to be consumed, it has to first be considered by the consumer. This means that it must be part of a consumer's consideration set (Shocker, Ben-Akiva, Boccara & Nedungadi, 1991). The term 'consideration set' has its roots in the fields of economics and psychology. A consideration set is a dynamic non-tangible set of consumption products existing in the consumers mind. The content of this set can change with time and occasion, and might be affected by consumer contexts and purposes (Shocker et al., 1991). For example, if a person wants to listen to music for exercising, they would most likely opt for songs with an upbeat tempo and consistent rhythm. Considering this, ballads would be immediately disregarded for this activity, as ballads would not be part of the consumer's consideration set at that moment. Instead, they might prefer to opt for a song belonging to the dance genre, which tends to be upbeat and have a consistent rhythm. A consideration set is part of the universal set, which refers to all of the alternatives that can be used, and is obtained or purchased by a consumer (Shocker et al., 1991). The products within the consideration set are subset to the universal set; this consideration set consists of the products consumers are aware of within the respective area of consumer interest. For example, in case a consumer is thinking of using a music streaming service, Spotify or Apple Music are very likely to be part of their consideration set, as these platforms are heavily advertised. In contrast, the music streaming service Deezer could potentially not be part of their consideration set as it is less well-known. This means that it subsequently cannot be part of the consideration set and thus belongs in the universal set.

Having explored what goes on within a recommender system in order to make a music recommendation, it is also necessary to explore what goes on within consumers to make a choice. This thesis suggests that there are three main contributing factors: availability, taste, and mood. These three factors are not exhaustive, however they are deemed most important in this study on music recommendation and consumption of them. When combining all three factors, availability denotes the practicalities of possible consumption, which is related to the universal set. Taste denotes what lies within consumer's potential liking, and subsequent consideration set. Lastly mood denotes what is part of a consumer's consideration set based upon their personal emotional context.

2.3.2 Availability

First and foremost, a song, album or artist can only be listened to if streaming platform has the song. This is a very large determining factor for whether the product of the respective artist can be consumed and ultimately recommended. It is not only vital for a streaming platform to have a large music catalogue, but also to differentiate themselves from their competitors within the same market. For a platform as big as Spotify, which despite its 200 million active listeners per month in 2019 (Spotify, 2019), they must also have the flexibility to create their own content in order to not only differentiate themselves from competitors, but to maintain existing customers and attract new ones as well (Morris & Powers, 2015). Thus, it is clear that availability is the most important factor concerning recommendation and choice - particularly when exclusive content is only available on their platform.

2.3.3 Taste

Taste is a more abstract and multifaceted contributing factor that can lead to consumption choices of individuals. According to Bourdieu (1984), taste is (pre)determined by the degree of possession of three kinds of capital: economic capital, social capital, and cultural capital. Economic capital is the amount of economic resources an individual has such as money, assets, or property. Social capital are the actual and potential social relationships an individual might have in their life, be it via their social network, institutions, acquaintances or recognition. Lastly, cultural capital consists of an individual's education in the form of knowledge and intellectual abilities, which tend to provide advantages in life and help individuals attain higher social-status. However, it should be mentioned that Bourdieu's work has often been criticised for being highly deterministic, meaning that in his theory there is little room for people to ascend or descend from their given social strata. In other words, according to Bourdieu, people can only ever play with the cards they are dealt with.

In contemporary society, however, taste is considered something more fluid and hybrid (Savage & Gayo, 2011). Individuals from various social strata are able to and do consume cultural products of a variety of genres, outside those of their traditional paradigms. Peterson and Kern (1996) coin the term cultural omnivore as "omnivorous inclusion seems better adapted to an increasingly global world managed by those who make their way, in part, by showing respect for the cultural expressions of others" (p. 906). Omnivorous consumption and taste denote the mixed consumption of cultural products that of high- and lowbrow status. For example, highbrow music is often typified by opera, and lowbrow can be considered as popular chart music. Those who consume and enjoy opera music do not do so because it is typically easy to listen to, nor is it so easily accessible to attend live events. As those who tend to enjoy opera have the means to understand and access the genre, they are more likely to appreciate it in turn. This is an example of status-based-culture and its subsequent status-based-consumption. This was more prevalent in earlier centuries and is a fundamental component of the work of Max Weber and Pierre Bourdieu (Savage & Gayo, 2011). Even though Weber's and Bourdieu's work is still important in academia, when it is applied in the context of this paper it becomes apparent that status-based-culture has largely ceased to exist in Western society (Savage & Gayo, 2011). The cultural omnivore has become more tolerant and liberalised in its consumption behaviour and choices. In other words, the contemporary consumer is not faced with a mutually exclusive highbrow-versus-lowbrow consumption-decision based upon their Bourdieusian capital - they are instead faced with an inclusive highbrow-versus-lowbrow decision. This decision is based upon their own mood, taste, and the product's availability. Despite this, Savage and Gayo (2011) state that it is still possible to draw the distinction between a highbrow and lowbrow consumer. The authors agree that it is possible to enjoy classical music as a contemporary consumer, however they argue that there are two different types of consumers that differentiate the highbrow from the lowbrow. Experts have a degree of 'mastery' over their musical interest and thus their enjoyment of classical music can feed into enjoyment of jazz, rock, and some forms of contemporary music, or so-called "easy listening" (Savage & Gayo, 2011, p. 353). These enthusiasts in turn might be fond of what is institutionalised as 'light classical' within the main genre and thus do not infer highbrow taste (Savage & Gayo, 2011). In this research, building upon Savage and Gayo's classification (2011) the heavy listener can be separated from the light listener by their knowledge of a large number of music genres – and their canon – that include both contemporary and classical reference points. This leads us to the second hypothesis of this research, *H2: Being a heavy listener will result in a preference towards CBF recommendations.*

Here the assumption is that a consumer who listens to music a lot would be able to recognise music of quality based upon their knowledge and consequent expertise in the area. Furthermore, their acquired taste and knowledge within a particular field of music allows the consumer to enjoy the canonical works of a variety of genres, especially outside the mainstream. Arguably, these kinds of consumers would also be aware of critics' reviews and recommendations as they are well versed in the matter and up-to-date with the latest developments within the field. It is necessary to note that heavy listeners are not per se the opposite to light listeners, they do not have to listen to more hours of music or are hardcore fans of music in order to be classified as heavy listeners. In this research the distinction between the two is more focussed towards the way either listener is aware of their consumption ways or not. In general, a heavy listener is more aware of and consciously busy with what type of music content they consume.

As mentioned, Content Based Filtered songs are recommended to the consumer via checking corresponding meta-data of a song and then by cross-referencing existing songs within the database. Following this, a person with a more advanced taste will receive and, or be aware of, more advanced recommendations made by the backend curators. Savage and Gayo (2011) state that it is not just the appreciation and consumption of classical music that determines the high from the lowbrow, but instead the intensity of consumption of all music. This differentiation is what signifies the true form of the omnivore. Additionally, according to

Goldberg, Hannan and Kovács (2016) omnivoresness - which they typify as variety seeking and liking – concerns not boundary/genre erosion but rather the protection thereof, "to make their breadth of consumption socially meaningful" (p. 233). Omnivores require boundaries to distinguish themselves. They will look for, and be aware of, critically acclaimed content before and when consuming music. This reflects a process which is similar to Bourdieu's description of social distinction practices (Bourdieu, 1984; Goldberg, Hannan & Kovács, 2016).

Conversely, it can be argued that the light listener will be happy to enjoy anything that is at the top of the charts, especially when and if they know their peers enjoy it too (Surowiecki, 2005). Therefore, the third and final hypothesis of this research is the following, *H3: Among light consumers, those who listen to music recommended by peers will have a preference towards CF recommendations.*

Preferring peer recommendations over curated recommendations could also be due to what is typified as homophily; this implies that individuals with common characteristics such as religion, education, and gender have an easier time communicating and forming relationships, because they like the same things (McPherson, Smith-Lovin & Cook, 2001). When it comes to music taste, individuals who tend to have and prefer homophilic relationships would often end up liking what those within their peer group like. Arguably, a Collaborative Filter in this case would be an excellent means to recommend musical content to such consumers. In this way, light listeners would be supporters of the wisdom of the crowd notion which denotes that aggregated preferences are better than expert knowledge (Surowiecki, 2005).

Furthermore, according to Fleder and Hosanagar (2009), contrary to reality, consumers feel that recommendations systems have increased the range of products they consume. The first explanation this they present for this phenomenon is that consumers perceive their recommended product diversity at an increased level. However, ultimately the level of diversity is still decreased in total aggregated products. Thus, even though individuals may be exploring more choices, they are in effect eventually being pushed toward the same choices. The second explanation is that as more products from the production side, i.e. songs, become available. Then the recommendations can in turn spark a feeling within consumers that they are consuming more diverse products simply because there are more products to be consumed. By eliminating a system which only pushes products that are popular among critics the platforms can allow for more inherent diversity as there are generally more non-expert consumers than critics. Finally, if products are added at a higher rate than being recommended then this can lead to a situation where recommended products turn out to be less concentrated. This outweighs the recommendation system and leaves the consumer with a feeling of more diverse product recommendation. Therefore, the consumers that are in fact concerned with filter bubbles caused by algorithmic recommendations, have valid reasons to be concerned. In fact, Fleder and Hosanagar (2009) have concluded that "[s]everal common recommenders were found to exert a concentration bias" (p. 711). However, there is an argument to be made for the social benefit of recommendation systems pushing what is already popular and placing consumers in such filter bubbles. Some consumers, light listeners, would and do in fact appreciate receiving recommendations of popular goods if this means that they can discuss their experiences of the goods among their peers. In this instance filter bubbles enable consumers to stay close to the latest consumer hype and, consequently close to being and discussing what is socially relevant (Fleder and Hosanagar, 2009). Hence, there is a case to be made for the benefit of filter bubbles as described above and for the efficiency of the wisdom of the crowd. Both can be deemed successful in the music industry, but most notably in the form of music charts. Such charts are built via aggregating music sales, listening counts, and airtime of songs.

Overall, consumers would rarely be aware whether their recommended content was Content Based Filter driven or Collaborative Filter driven. Though a lack of awareness of which filter – the curated CBF or user driven CF – recommended a song to the consumer does not necessarily mean a lack of care. Holbrook, Lacher & LaTour (2006) came to the conclusion that non-expert music consumers do recognise the aesthetic excellence of songs, as assessed and attributed by experts. This aesthetic excellence is also associated with audience appeal and moderated by aesthetic merit which they have denoted as audience judgements. Thus, according to Holbrook et al., consumers do care and appreciate expert judgements on the content they may want to consume. It is important to note that Spotify consumers, for example, cannot rate songs by making reviews themselves. The popularity and subsequent rating of a song or album comes from stream counts on the platform, but Spotify does make Top 50 or Viral 50 chart lists. Interestingly, the Viral 50 chart follow a "new metric" (Helgadóttir, 2018). The Viral 50 charts are made up of tracks divided by three main characteristics: big national or international hit, heavily shared tracks on social media, or songs that appeared in pop culture - such as films or memes (Helgadóttir, 2018). Arguably, the aggregated preferences on the platform cannot be misguided by bad reviews, as stream counts would signify popularity more objectively. In an ideal scenario, what determines the charts are the aggregated consumer preferences of the platform, which would result in only the top songs and artists reigning the charts and the streaming platforms. In reality, though, it is not so democratic. Given the success of music streaming platforms, they now have the opportunity to create content of their own. Consequently, it is within their interest to push this content to the foreground and increase its visibility for their consumers (Morris & Powers, 2015). Moreover, linking back to availability, creating platform unique content is another way to attract and keep users as well.

Spotify has their own 'Spotify Sessions', which feature popular artists at their own 'Spotify Studio'. At these sessions the artists perform covers or acoustic renditions of their own songs (Dillet, 2016). By creating such content, Spotify is actively trying to differentiate itself from other competing platforms by offering exclusive content to its consumers (Morris & Powers, 2015). For the consumer being able to discuss exclusive content can also create buzz and excitement amongst consumers, which consequently benefits the platform and consumer in a reciprocal manner. Thus, the bandwagon effect, which entails consumers basing their consumption choices on that of other consumers, can still be present even without reviews on the platform itself (Sundar et al., 2008; Hills, 2002). The bandwagon effect also brings bandwagon fans and such fans tend to not to be as loyal to the artists of the moment when compared to those who are active fans of the artists. They become fans of the trendiest performers of the time, switching between artists as their popularity rises and falls (Sundar et al., 2008; Hills, 2002).

2.3.4 Mood

The last main characteristic, mood, is just as important as the previously discussed factors when it comes to recommending a song. There is evidence for consumer choices that lie outside of conscious decision-making processes. Fitzsimons et al. (2002) have found that non-conscious influences also affect consumer choice. Fitzsimons et al. state that even when consumers believe to have made a truly conscious and well thought-out decision, non-conscious influences can still play a significant part in their decision making. Such non-conscious influences take many forms, including asking the consumer a hypothetical question, exposing them to facial expressions/clear emotions, or semantic primes (Fitzsimons et al., 2002).

These non-conscious influences take place in the individual's automatic processing of cognitive information. Knobloch & Zillmann (2002) recognise the importance of non-conscious influences on consumer mood and choices, following the mood management theory. The abundance of entertainment products available at any time and place affords consumers to let their mood, and the surroundings the consumer finds themselves in, to be part of the deciding factors in entertainment – in this case music – consumption. The mood management theory as conceptualised by Zillmann (1988a, 1988b), is based upon the assumption that "individuals seek to experience the highest degree of pleasure attainable under given circumstances, this theory posits that persons are motivated to make entertainment choices that will help them to diminish or terminate negative moods and to extend and enhance good moods" (Knobloch & Zillmann 2002, p. 352). Knobloch & Zillmann (2002) find that consumers in three basic moods – good, neutral or bad – all acted accordingly as the mood management theory suggests. Respondents in a bad mood purposefully chose to listen to highly energetic-joyful music, in a higher degree than respondents who were already in a good mood. Those in a neutral mood also sought out highly energetic-joyful music. In other words, the participants in neutral or bad moods were aware of their emotional states and had an understanding of their states as well. They knew which music to choose, highly energetic-joyful music, to increase their chances of altering their emotional state. Contrastingly, participants already in good moods took more liberty in exploring more emotionally diverse music, as their mood was already (close to) optimal (Knobloch & Zillmann, 2002).

2.4 Chapter summary

Based on the literature discussed in this chapter three hypotheses were made. The first one concerns itself with dissecting the two filter methods and testing them against one another. Moreover, based on previous literature discussed in this chapter, the consumer utility would be highest with CF recommendations. Based on the music consumption profiles of heavy listeners, which will be discussed in the methodology section, H2 would result in a preference towards CBF recommendations. This will be due to the individual's heavy consumption habits and thus will want to be recommended more novel and expert appreciated content rather than popular, consumer driven content. The third and final hypothesis would be favoured toward CF recommendations based on discussed literature and based on the light listener profiles of the respondents who prefer peer suggested music. Both methods have been researched via within-subject survey experiments containing songs recommended via both types of filter methods. Ultimately, the considering the main research question of this study, the three hypotheses aim to test whether there is a different level of consumer utility depending upon the filter type behind a song and the type of listener it is recommending this to.

Chapter 3. Methods

3.1 Design

The research method that has been chosen is a quantitative analysis with within-subject survey experiments, as characterised by Neuman (2014). Quantitative research is deductive in nature and when combined with surveys it makes it possible to translate theoretical concepts into concrete and measurable variables (Bryman, 2012; Neuman, 2014). Survey experiments aim to test, in a controlled yet artificial environment, whether there is a causal relationship between one choice present in the survey versus another, in this case the choices shall be filter recommended songs. The options of the songs that the respondents will be faced with will come to them one by one in a linear manner in the survey. All the respondents were subjected to the within-subject experiments as all respondents came across songs recommended to them via three types of filtering: Content Based Filtering, Collaborative Filtering, and Random filtering. The respondents were not aware via which filter type the recommendations were made, this is the experimental element.

For this research within-subject survey experiments were chosen because this method was deemed the best way to test the consumer utility of the respondents based on the filter recommendations. The respondents were not made aware, as on real music streaming platforms, how the song recommendations were made for them and based on what criteria or via which method. A within-subject survey experiments method allows all respondents to be exposed to such filter types in a controlled environment, the survey. By doing so, all survey input can be standardised and documented, from which conclusions can be drawn and the hypothesis can be answered (Neuman, 2014). All respondents came across the same number of songs (6 in total) and the same amount of filter types (2 CBF, 2 CF, 1 Random). The only difference between the respondents were the genre specific paths they could go down, of which there were five in total. Moreover, the order in which the respondents encountered the songs was randomised for each respondent and their respective genre path. The following will be a detailed description of the methods employed for making the survey, selecting the filtered songs and approaching the respondents.

3.2 Procedure

An online experimental survey was created to effectively and quantifiably measure the dependent variable ‘consumer utility’ of the independent variables which were the types of filter recommendations (CBF, CF and Random). And the consumer utility of the different filter types for both ‘heavy listeners’, and ‘light listeners’. The survey programme ‘Qualtrics’ was used for making and publishing the survey. The survey could be answered via an anonymous web-link.

3.2.1 Survey design

The survey consisted of 153 questions in total. However, because respondents go down different paths within the survey based on the selection of their preferred main genre the respondents only had to answer 41 questions each. The survey was pretested by 11 respondents and some changes were made to the survey based on their feedback; which changes were made will be discussed at various times in this chapter. The survey questions can be found in Appendix A. At first the respondents needed to provide informed consent that they will be participating in anonymised, academic research. Moreover, they were provided with a succinct description of the research and its purpose at the beginning of the survey. Subsequently, the participants were asked if they made use of any music streaming platforms, a list of alphabetically arranged names was provided but the participants were also free to type in the names of other music streaming platforms they made use of (1 = Apple Music, 2 = Deezer, 3 = Google Play Music, 4 = Pandora, 5 = SoundCloud, 6 = Spotify, 7 = Tidal, 8 = YouTube, 9 = Other, please specify: ____, 10 = I do not make use of any music streaming platforms.). The latter option was used as a filter in order to exclude respondents who did not make use of music streaming platforms as these respondents would not be useful given the purpose of this research. Then they were asked if they had a paid subscription for any of their music streaming platforms on a dichotomous nominal answer key (1 = yes, 2 = no).

Then respondents were asked questions about their demography. First, their age was inquired via an open answer key. Respondents below the age of 18 were also filtered out and consequently excluded from the survey due to ethical standards. Second, their gender was inquired via a nominal answer key (1 = male, 2 = female, 3 = other, 4 = prefer not to say). Third, the respondents were asked to select their native country via a dropdown list of 195 countries (1 = Afghanistan …, 195 = Zimbabwe). Fourth, the respondents were asked to select the country in which they currently reside with the same list. This list of countries was imported from the pre-made Qualtrics library questions. Lastly, the respondents were asked to select their level of education via a ordinal answer key (1 = No formal education, 2 = Secondary education, 3 = Vocational education, 4 = University degree (BA), 5 = University degree (MA), 6 = University degree (PhD)).

After the demographic questions the respondents came across questions pertaining to their own music listening habits and behaviour. The respondents were asked to estimate on average how many songs they listen to per day and how many hours they listen to music per day via an open answer key. To assess the respondents' perception of their own listening habits and behaviour, they were asked via a 7-point Likert scale (1 = Strongly disagree, …, 7 = strongly agree) to what extent they agreed with six separate statements. These statements included but were not limited to whether the respondents regarded themselves as heavy music listeners, would listen to music suggested by their peers, and would listen to music suggested by (journalistic) music sources. From these statements the operationalisation of an omnivore can also take place. As stated, an omnivore enjoys consuming a wide variety cultural goods, credited for example by a variety of sources such as peers or curators. Previously, omnivores and their disputed existence used to be predominantly studied by qualitative scholars such as DeNora (1999; 2000). However, as Savage and Gayo (2011) point out the omnivore concept lends itself to clear and definite forms of quantifiable measurement within surveys, such as simply inquiring respondents whether they consume a particular (highbrow/lowbrow) cultural good or not.

Following this, the participants were asked to indicate their mood via four categories. The options were nervous/calm, upset/happy, indifferent/excited, fatigued/energised. The categories were adopted and modified from Barrett and Russell's (1999) study on the independent dimensions of experience, pleasure and activation. The respondents could indicate their mood via a matric table with three scale points. Initially the matrix table consisted of solely a dichotomous two-point scale, as done by Barrettt and Russell (1999), but this was changed to a three-point scale due to feedback from respondents who filled in the pilot survey. This change was made because the pilot survey respondents expressed that they would like to have a neutral mood option. Therefore, in the final survey the respondents could for example indicate their mood between upset, neutral or happy; which is arguably more nuanced and thus more accurate. The questions pertaining to music listening habits and behaviour and mood were asked before the recommendations so that the respondent might be prone to think that these also affected the recommendations presented to them. This was done purposefully so that the respondent would think they had more of an influence in their recommendations, in the sense that they might be even more tailored to them specifically.

Consequently, the respondents arrived at the main part of the survey. Here the respondents were asked to select which of the five following music genres and their corresponding songs appealed to them the most:

1 = Pop: Lady Gaga – Alejandro,

2 = Rock: Red Hot Chili Peppers – Snow (Hey Oh),

3 = Hip Hop: Kanye West ft. Chris Martin – Homecoming,

4 = R&B: Leona Lewis - Bleeding Love,

5 = Dance: Major Lazor ft. Vybz Kartel & Afrojack – Pon De Floor

In order to recreate a layout that looks as close as a music streaming service as possible, along with the artist and title of the songs, respondents will also be shown cover art of either the respective singles or albums, together with audio fragments of up to 30 seconds. Most of the audio tracks included in the survey covered the first seconds of the songs, however in some cases fragments in the middle of a song were chosen due to increased recognisability chances. This is because song that has a long instrumental introduction could take longer to identify than a song that starts with lyrics. These five options lead up to the five song recommendations, two of which have been recommended via Content Based Filtering, two via Collaborative Filtering and one was Random filtering. Before the respondents received their five recommendations they first came across the following statement: "Based on your music listening habits and music consumption 5 song recommendations have been made. Please note: due to technical limitations the songs presented in this survey will have been released in or before the year 2011." This statement was included to inform the respondents shortly of what they would come across in the survey after having provided information based on their demographics and music listening habits and behaviour. The section "5 song recommendations" was highlighted bold in the survey to emphasise that the respondent would come across no more than five songs. This was done to give the respondent of the survey an indication of how long the rest of the survey would take. It is important to note that the five recommended songs were already preselected for the survey respondents, the only liberty they had was to pick their preferred main genre/main song out of the five mentioned above. The respondents were not made aware of this however, due to the fact that on a music streaming platform consumers are also not aware how recommendations are made exactly, thus neither were the survey respondents in order to simulate authenticity.

Distinguishing the genre of music artist or song has long been a topic of discussion amongst scholars (Frow, 2006; Hennion, 2001; Savage & Gayo, 2011). This is because music genres tend not to be entirely pure or unambiguous. The songs within this research were attributed to five separate main genres, namely: pop, rock, hip hop, r&b, and dance. This was done to give the respondents a frame of reference to what kind of recommendations might follow from their choice of song and its respective genre. Because even though the boundaries of genres might be difficult to define because they are constructed by "human, institutional and technical agencies" (Savage & Gayo, 2011, p. 340), and these boundaries are constructed it makes them fluid and complex. Arguably, the main genres mentioned above would strike enough resemblance with the respondents, therefore they were mentioned and utilised in the survey. Subgenres such as alternative rock, or dance-pop make labelling and sorting matters a lot more complex, thus no mention of subgenres was made in the survey, also not for the recommended songs.

After having chosen which one of first five songs they would prefer to listen to, their path was paved in the direction of that corresponding genre. From then on, they came across five other songs – in a linear manner – whereof two will have been preselected via CBF, another two will have been preselected via CF and the remaining one was Randomly preselected. These options will be presented in a random order. Following this, a repeated measures experiment (Bryman, 2012; Neuman, 2014) can take place with the presentation of different CBF, CF and Random preselected songs.

As stated, songs selected for the survey will have been released in or before the year 2011. To increase the chances of the survey respondents being familiar with the songs, popular music charts have been consulted, this was the case for both CBF and CF songs. By doing so, this will hopefully decrease the amount of times the respondents will answer that they do not know a song. The audio fragments have been included for the same purpose as well. Furthermore, if respondents answer that they do not know the songs in the survey, then the recommended options (from either CF or CBF methods for instance) can be deemed less appropriate. This is because the level of familiarity the respondents will have with the presented songs also provides information about the utility of the song recommendations in the survey and, ultimately, the filters. An overview of all the songs present in the survey can be found in Appendix C.

Content Based Filtering was operationalised via ascribing both objective and subjective metadata to the total of ten preselected CBF songs in the survey. Each genre contained of two CBF recommended songs. First and foremost, objective metadata will be ascribed to the songs. This consists of the song title, artist/band name, release year, country of origin, main, subgenre and instrumentation. Taking the pop song Alejandro by Lady Gaga as the starting position and example again; the objective metadata would be:

Table 1. *Example of ascription of objective CBF metadata.*

Objective metadata	**Starting song**	**Objective metadata**	**Recommendation**
Song title	Alejandro	Song title	Princess of China
Artist	Lady Gaga	Artist	Coldplay ft. Rihanna
Release year	2009	Release year	2011
Country of origin	US	Country of origin	GB & US
Main genre	Pop	Main genre	Pop
Subgenre	Dance-pop	Subgenre	Electropop
Instrumentation*	Violins, synthesizer	Instrumentation*	Synthesizer

* for the instrumentation only the most stand out instruments were noted down

Following this, subjective metadata will be ascribed. The subjective metadata consists of song energy, song mood, and song speed. Song energy is characterised mostly by the instrumentation in a song. Thereof, it is possible to have ‘chill’ with simple instrumentation such as soft synths and/or acoustic guitar, ‘normal’, and ‘powerful’ loud/strong song with voluminous vocals and/or instrumentation. Song mood consists of ‘sad’, ‘melancholic’, ‘normal’, ‘uplifting’, and ‘happy’. Song speed is divided by the labels ‘ballad’, ‘relaxed’, ‘normal’, ‘up-tempo’, and ‘extra fast’. It is often the case that certain instrumentation goes hand in hand with song mood, energy or tempo; for example, songs that predominantly contain piano or string instrumentation often tend to be downtempo and sad or melancholic. Such conventions will ease the process of ascribing metadata to the songs.

Table 2. *Example of ascription of subjective CBF metadata.*

Lady Gaga - Alejandro		Coldplay ft. Rihanna – Princess of China	
Subjective metadata	**Starting song**	**Subjective metadata**	**Recommendation**
Song energy	Powerful	Song energy	Powerful
Song mood	Melancholic	Song mood	Melancholic
Song speed	Normal	Song speed	Normal

As can be seen from the tables above, most metadata match between both songs, in the case of the subjective metadata they match completely. Both songs contain heavy synthesized beats. The genres dance-pop and electropop share a lot of similarities too (Mackay, 1981); dance-pop is a subgenre from dance and pop, the genre is generally up-tempo and intended to be danceable; it is often characterised by a strong beat with conventional pop song structures (Mackay, 1981) which fits with Alejandro; electropop is a derivative of synthpop, but this subgenre places the emphasis on a harder electronic sound such as a heavy synthesizer (Mackay, 1981), as is present in Princess of China.

Collaborative Filtering was operationalised via utilising The Million Song Dataset by Bertin-Mahieux, Ellis, Whitman & Lamere (2011). The survey consists of ten CF songs in total, with two CF songs per genre. This dataset was deliberately made charge free and open access for anyone. The purpose of the dataset is to encourage research on music algorithms, to provide a reference dataset, and to give a shortcut alternative to creating a music dataset with API's. The dataset is said to contain the "metadata for a million contemporary popular music tracks" (Bertin-Mahieux et al., 2011). The dataset contains one million songs, 44,745 unique artists, and 515,576 dated track starting from 1922 (Bertin-Mahieux et al., 2011). It is important to note that the dataset was built/finalised in 2011, this means that songs after 2011 are not included in the dataset. Considering this, the songs presented in the survey will also not surpass the year 2011, also not for the CBF and random songs. All songs had to adhere to this rule because this ensured that a user of The Million Song Dataset could encounter the song herein, thus making it part of their universal set or consideration set, especially considering the importance of availability for consumption. Furthermore, considering the dataset was built in 2011, another guideline for the recommended songs was implemented that they were released between 2011 and 2006. This period of 5 years was chosen to increase the probability of the respondents' familiarity with the songs. This rule was adhered to for all songs in the survey except for one, Guns N' Roses – Sympathy for the Devil which was released in 1994. This exception was made to enhance the validity of the study due to feedback from the respondents of the pilot survey, whom thought the initial selection of rock songs did not feature enough "pure rock", but mostly alternative rock.

The complete code used for Collaborative Filtered songs can be found in Appendix B. In short, CF recommendations are made via mathematical processes estimating the probability of consumption for a user (Burke, 2007; Ricci et al., 2011). The coding was done via the general-purpose Python programming language (Python, n.d.). The code used for this research was prewritten by Narayana Swami (GitHub, n.d.). GitHub is an open source website where users can upload their own software or codes. Within the code used for CF firstly it is necessary to define the sample size n for the entire population N (which is approximately one million songs) of the data set. The number of *n* denotes the top songs up to and including *n*, for example if *n* 5000 the code will include the top 5000 songs for the final ten recommendations it will put out. Now the code will collect the top 5000 songs from the top 5000 users in the dataset. Then it proceeds to calculate the probability between a recently listened to song (1) and *n*:

```
user_data = {'Snow [Hey Oh] (Album Version) - Red Hot Chili Peppers': 0,
 'Alejandro - Lady GaGa':1,
'Bleeding Love - Leona Lewis': 0,
'Homecoming - Kanye West': 0,
 "Pon De Floor - Major Lazor / Vybz Kartel / Afrojack": 0}
```

Here 1 denotes a song listened to and the recommendations starting point, in this case it will be exemplified by the song Alejandro by Lady Gaga from the main genre pop, and 0 denotes to not include the song in the recommendation. Only one of the five songs needed to be included each time for the recommendations, because the respondents in the survey go down a separate path based on their preferred genre and song, thus only one song is denoted with a 1 and the rest with 0.

From the top 5000 most listened to songs by the top 5000 users and starting from Lady Gaga's Alejandro, the coding proposed ten top recommendations from the dataset: Love Story - Taylor Swift, Reelin' In The Years - Steely Dan, Catch You Baby (Steve Pitron & Max Sanna Radio Edit) - Lonnie Gordon, Boys Boys Boys - Lady Gaga, Crescendolls - Daft Punk, A Thousand Miles - Vanessa Carlton, Monster - Lady Gaga, Just Dance - Lady Gaga / Colby O'Donis, Sehr kosmisch – Harmonia, and Whataya Want From Me - Adam Lambert. From these ten two songs had to be chosen for the CF recommended songs (as the other three songs came from CBF (2) and randomly (1) selected recommendations). The selection of the two CF songs that could be included in the survey was done largely via a process of elimination. To avoid artist bias and to ensure artist variation the other Lady Gaga songs could be disregarded. Considering the rule of the 5-year period the songs A Thousand Miles – Vanessa Carlton and Crescendolls - Daft Punk among others could also be left out. Lastly, the songs were checked for popularity – and their consequent familiarity – within the charts, such as Billboard. Following all these steps the songs Love Story – Taylor Swift (2009) and Whataya Want From Me – Adam Lambert (2009) were selected for the CF songs in the survey.

In addition, a random song was added per every path as well. There were a total of five Random filtered songs, with one Random song per genre. This is firstly to test whether the songs are truly tailored by either filtering recommendations and not just random. The second reason is to also simulate a path - and thus present a song - that the respondent could stumble upon, outside of filter recommendations. For example, a consumer can explicitly look for a song, regardless of any suggestions made by music streaming platforms, as it was part of their consideration set. At first a random number between 1 and 100 was selected via a random number generator. Then the Billboard Charts Archive was consulted (Billboard, n.d.). The number from the generator was checked with the position of songs in the Billboard Hot 100 charts in the case of Pop this was number 39 in the Year-End Charts Pop Songs of 2010 which denoted Britney Spears – 3 (2009). For the other genres, such as rock or dance, at times genre specific Top 100 or Year-End Charts were consulted. It is important to note that all the randomly selected songs had to be available in the dataset. This meant that in some cases a different number had to be generated or another year had to be selected between 2006 and 2011 if they corresponded to a song that was not available within the dataset.

3.3 Independent variables

3.3.1 Consumer utility

The concept consumer utility was tested by combining two variables from the survey, namely the variables connected to the question "How likely will you be to listen to this song?" which could be answered via a 7-point Likert scale (1 = extremely unlikely …, 7 = extremely likely) and the question "Does this song fit in your musical taste?" which could be answered via a 5-point Likert scale (1 = not well at all …, 5 = extremely well). As mentioned, to analyse a repeated measures ANOVA will be employed to answer the hypothesis as well as the main research question of this thesis. Ultimately, the correlation between the independent variables of CBF, CF, Random filter, were tested against the main dependent variable 'consumer utility' composed by variables connected to song familiarity and song taste. These correlations were tested by conducting repeated measures ANOVA on the programme SPSS.

3.3.2 Heavy listener

Whether a respondent was a heavy listener depended on the answers they provided to particular questions within the survey. For example, one question in the survey directly asked the respondents whether they agreed with the statement "I consider myself a heavy music listener." (1 = strongly disagree …, 7 = strongly agree). The former question denotes the type of listener a respondent is. Additionally, the respondents were asked to which extent they agreed (on the same scale as mentioned) which statements regarding where they preferred their consumed music came from, such as "I like to find music on my own." or "I will listen to music recommendations suggestions by (journalistic) music sources.”. As discussed in the previous chapter heavy listeners tend to be more omnivorous in their consumption and tend to enjoy and prefer distinguishing themselves from general consumers (Goldberg, Hannan & Kovács, 2016; Savage & Gayo, 2011). Thus, such heavy listening consumers will be operationalised via looking at the extent to which they agree with the statements mentioned above.

Another way to identify heavy listeners is if they (positively) consider themselves fans of music. Fans, according to Hills (2002) and Jenkins (1992), can be characterised as active consumers. There tends to be a level of variety between researchers when it comes to the exact criteria for the distinction between active consumers, and casual consumers. Typically, to be considered an active consumer or fan of any media product, individuals are ought to consume the media product in question at least multiple times a week, at minimum. Though with music the distinction might go even further than consuming it multiple times a week because a person can hear music at many places, such as commercial establishments or through the next-door neighbours wall, without actively having asked for it or cared. Thus, to be an active consumer and a fan of music there must be a degree of decisiveness and commitment to listen to music, especially of one’s own collection or playlist. Furthermore, Hills (2002) and Jenkins (1992) state that fans often are actively involved in the discovery of more (new) content.

A new variable was created combining the variables heavy music listener, fan of music, and lover of music. This new variable, denoted as heavy listener scale, had a reliability of α = .775. The variable for heavy music listener had a Skewness of -.90 and a Kurtosis of -.08; this means that the variable is normally distributed among the 302 respondents. The variable lover of music had a Skewness of -1.69 and a Kurtosis of 3.57, and the variable fan of music had a Skewness of -1.61 and a Kurtosis of 2.86, though according to the Central Limit Theorem and due to the sample size this is not a serious violation of the assumption of normality, also because they do not differ too much from the Skewness and Kurtosis limits of -3 and 3.

3.3.3 Light listener

Conversely, light listener will be operationalised via looking at the extent they disagree with such statements. Furthermore, as the literature suggests light listeners tend to listen to music suggested by their peers, and even prefer to do so rather than by curators. Following this, light listeners would positively agree with the survey statement “I would prefer to get music recommendations based on the music taste of my peers”. This would fall in line with the wisdom of the crowd (Surowiecki, 2005) and homophily (McPherson, Smith-Lovin & Cook, 2001).

3.4 Participants

In order to take part in this research, respondents were asked if they are 18 years or older of age. Secondly, they were asked if they use any type of streaming platform(s) when consuming music. If both questions are answered positively, they will be allowed to continue the survey. If any of the questions were answered negatively, meaning they do not make use of any music streaming platforms or they are below the age of 18, then the participants were automatically directed to the end of the survey and thanked for their participation. This was done because if a respondent did not make use of music streaming platforms then they would not encounter a scenario in which songs are recommended to them via different types of filter methods. Thus, to ensure authenticity and real-life probability these respondents were filtered out. Secondly, respondents below the age of 18 were filtered out in the same manner too but this was due to ethical standards. The survey responses were collected between Wednesday 15th May, 2019 and Thursday 31st May, 2019. The survey took around 10 minutes to fill in. A total of 389 surveys were collected. However, 73 survey had to be discarded due to significantly incomplete answers. This results in 316 completed surveys, though of the 316 (N = 316) completed surveys another 14 were finished incompletely because these surveys did not meet the streaming platform consumer and/or age requirements and thus were sent to the end of the survey automatically. Of these 14, five people did not make use of music streaming platforms and nine people were below the age of 18. The survey respondents were made up of 184 males, 111 females, 5 who specified their gender as "Other" and 2 people whom withheld their answer. The respondents were between the age of 14 and 66 (M = 26.95, SD = 9.96), the mode was 21 years old. The level of education of the respondents varied substantially: 5 (1.6%) respondents obtained no formal education, 61 (19.3%) obtained secondary education, 29 (9.2%) obtained vocational education, 154 (48.7%) obtained a BA university degree, 50 (15.8%) obtained a MA university degree, and 3 obtained a PhD (0.9%). The sample obtained a total of 185 nationalities, the most prominent being the United States (110 respondents, 34.8%), The Netherlands (60, 19.0%), and Canada (15, 4.7%). Most respondents currently resided in the United States (112, 35.4%), The Netherlands (70, 22.2%), and the United Kingdom (19, 6.0%). The possible explanations for the skewness of the demographics will be discussed more in depth in the Discussion and Conclusion chapter.

Participants were reached mainly via the means of non-random convenience and criterion sampling, as characterised by Bryman (2012). The survey was published and shared on two social media platforms: Facebook and Reddit. The survey was posted within groups related to events, work, study and hobbies – preferably with a distinct relationship with and/or interest in music. On reddit for example, the survey was shared on the fora of the platforms listed as question two in the survey, among which: r/GooglePlayMusic, r/Spotify, and r/Tidal. In most cases the respondents participated in the survey out of their free will, or what can be called self-selected sampling (Bryman, 2012). The online survey programme Qualtrics was used in order to ensure anonymity of the respondents. Furthermore, Qualtrics was also chosen because of the many distinct features this survey programme has that will be most appropriate in designing the experiment within the survey. The following formula was used to estimate a representative sample from a population of – approximately – two billion music streaming platform users (McIntyre, 2018): Sample Size Formula = [z2 * p(1-p) ÷ e20] ÷ [1 + (z2 * p(1-p)) ÷ e2 N]). In this formula z refers to the critical value of the normal distribution, p is the sample proportion, e the margin of error, and N id the population size. With a 5.0% margin of error and 95.0% confidence level, the response rate was estimated at 20.0%, this meant that at least 1925 had to be approached for a representative sample size of 385. As stated for this thesis 389 people had started the survey but only 316 responses could be deemed valid for analysis.

Chapter 4. Results

4.1 Descriptive statistics

Out of 316 respondents 248 (78.5%) have a paid subscription to a music streaming platform, 63 (19.9%) respondents did not. There are 5 missing answers because out of the total 316 surveys five people did not meet the music streaming platform requirement as mentioned in the Methods chapter.

The most popular path respondents went down was rock, with 49.7% (*n* = 157) choosing this path. Following this 15.5% (*n* = 49) chose pop, 14.6% (*n* = 46) chose hip hop, 8.2% (*n* = 26) chose dance, and lastly 7.6% chose r&b (*n* =24).

24 respondents chose the r&b path. The favourite song of the 5 r&b songs in the survey was Rihanna - Te Amo (CF) 33.3% of the respective respondents chose this song. 25.0% chose D'Angelo - Untitled (How Does It Feel) (CBF), 16.7% chose Beyoncé - Sweet Dreams (CBF), 16.7% chose Keyshia Cole - Fallin' Out (R), and lastly 8.3% chose Justin Bieber – One Time (CF).

46 respondents chose the hip hop path. The favourite song of the 5 hip hop songs in the survey was Lupe Fiasco ft. Matthew Santos – Superstar (CF) 37.0% chose this song as their favourite out of the 5 hip hop songs. 34.8% chose Drake ft. Lil Wayne & Young Jeezy – I'm Goin In (CBF), 13.0% chose Soulja Boy – Crank That (Soulja Boy) (CBF), 10.9% chose DJ Khaled ft. T-Pain, Ludacris, Snoop Dogg & Rick Ross - All I Do Is Win (CF), and lastly 4.3% chose Young Money ft. Lloyd – BedRock (R).

The most popular path was rock, 157 respondents chose this path. The favourite rock song out of the 5 was Muse – Uprising (CBF) with 31.2% of respondents selecting this song. Furthermore, 23.6% chose The Black Keys – Tighten Up (R), 21.0% chose Kings of Leon – Use Somebody (CBF), 15.3% chose Rise Against – Behind Closed Doors (CF), and finally 8.9% chose Guns N' Roses – Sympathy For The Devil (CF).

The second most chosen path was pop with 49 respondents. The favourite pop song among respondents was Coldplay ft. Rihanna – Princess of China (CBF), with 30.6%. Next came Britney Spears – 3 (R) with 26.5%, Timbaland ft. OneRepublic – Apologize (CBF) with 18.4%, and Taylor Swift – Love Story (CF) and Adam Lambert - Whataya Want From Me (CF) with 12.2% each.

Finally, 26 respondents chose the dance path. The most favourite song was Crookers ft. Kid Cudi - Day 'N' Nite (CBF) with 30.8%. Following this 26.9% chose The Black Eyed Peas – I Gotta Feeling (CBF), 23.1% chose Crystal Castles – Vanished (CF), 15.4% chose Justin Timberlake - LoveStoned/I Think She Knows (R), and 3.8% chose California Swag District - Teach Me How To Dougie (CF).

4.2 Hypotheses Testing

4.2.1 Hypothesis 1

In order to test H1: *There is a difference within consumer utility depending upon which recommender is used: either CBF or CF,* a repeated measures ANOVA was conducted for each of the five main genres present in the survey. The repeated measures ANOVA is a test to detect any overall differences in the means of related variables (Field, 2013; Pallant, 2013). For a repeated measures ANOVA to take place a number of assumptions have to be met: the dependent variables need to be normally distributed, the independent variables have to be categorical and the dependent variables need to be scale variables, the variance levels between the different conditions need to be close to equal (Field, 2013; Pallant, 2013). The mean and standard deviations of the variables denoting likelihood to listen to a respective song, answerable via a 7-point Likert scale (1 = extremely unlikely, …, 7 = extremely likely), and the variable denoting whether a song fit within a respondents taste, answerable via a 5-point Likert scale (1 = not well at all, …, 5= extremely well), were both independently compared against the related groups which were five songs per genre. Both results were combined into one table per genre. Note, the order that the songs appear in the tables is not necessarily the order that the respondents encountered the songs in the survey, as they were randomised per survey respondent. The following will be a reporting of the results of the repeated measures ANOVA per genre.

Table 3. *Reports of means and standard deviations for R&B likelihood and taste.*

	Filter type	Likelihood		Taste	
R&B		*M*	*SD*	*M*	*SD*
1. D'Angelo - Untitled (How Does It Feel)	CBF	3.96	2.07	2.46	1.41
2. Rihanna - Te Amo	CF	4.63	1.86	2.96	1.27
3. Beyoncé - Sweet Dreams	CBF	5.04	1.81	3.13	1.33

4. Keyshia Cole - Fallin' Out	R	4.21	1.87	2.79	0.93
5. Justin Bieber – One Time	CF	2.63	1.86	1.88	1.04

Out of the 302 respondents 24 participants chose the r&b path. Mauchly's test for sphericity has been met for the main effects of likelihood to listen to the five r&b songs from the survey, $\chi^2(9) = 14.31$, $p = .113$. The repeated measures ANOVA revealed a significant main effect on the likelihood to listen to differently filtered r&b songs, F (4, 92) = 7.00, $p < .001$. Two significant differences were found with the Bonferroni Correction between songs and respondents' likelihood to listen to the r&b songs. The first between song 5 (Justin Bieber – One Time, CF) and song 2 (Rihanna – Te Amo, CF) $p = .001$. The second between song 5 and song 3 (Beyoncé – Sweet Dreams, CBF) $p < .001$.

Regarding whether the respective r&b songs fit within the musical taste of the respondents, Mauchly's test for sphericity has been met for the main effects of taste on the five r&b songs from the survey, $\chi^2(9) = 14.12$, $p = .119$. The repeated measures ANOVA revealed a significant main effect of taste to differently filtered r&b songs, F (4, 92) = 4.92, $p = .001$. Three significant differences were found with the Bonferroni Correction between songs and whether the songs fit within the respondents' taste. The first between song 5 (Justin Bieber – One Time, CF) and song 2 (Rihanna – Te Amo, CF) $p = .007$. The second between song 5 and song 3 (Beyoncé – Sweet Dreams, CBF) $p = .002$. The third between song 5 and song 4 (Keyshia Cole – Fallin' Out, R) $p = .042$.

Table 4. *Reports of means and standard deviations for hip hop likelihood and taste.*

	Filter type	Likelihood		Taste	
Hip hop		*M*	*SD*	*M*	*SD*
1. DJ Khaled ft. T-Pain, Ludacris, Snoop Dogg & Rick Ross - All I Do Is Win	CF	2.98	1.99	2.04	1.15
2. Soulja Boy – Crank That (Soulja Boy)	CBF	3.07	1.94	2.07	1.04
3. Young Money ft. Lloyd – BedRock	R	3.11	2.06	2.20	1.24
4. Lupe Fiasco ft. Matthew Santos – Superstar	CF	4.46	2.01	2.80	1.22

5. Drake ft. Lil Wayne & Young Jeezy – I'm Goin In	CBF	3.74	2.03	2.54	1.21

Out of the 302 respondents 46 participants chose the hip hop path. Mauchly's test for sphericity has been met for the main effects of likelihood to listen to the five hip hop songs from the survey, $\chi^2(9) = 12.59$, $p = .182$. The repeated measures ANOVA revealed a significant main effect on the likelihood to listen to differently filtered hip hop songs, F (4, 180) = 17.75, $p < .001$. Three significant differences were found with the Bonferroni Correction between songs and their likelihood to listen to the hip hop songs. The first between song 4 (Lupe Fiasco ft. Matthew Santos – Superstar, CF) and song 1 (DJ Khaled ft. T-Pain, Ludacris, Snoop Dogg & Rick Ross - All I Do Is Win, CF) $p < .001$. The second between song 4 and song 2 (Soulja Boy – Crank That (Soulja Boy), CBF) $p = .005$. The third between song 4 and song 3 (Young Money ft. Lloyd – BedRock, R) $p = .001$.

Concerning whether the respective hip hop songs fit within the musical taste of the respondents, Mauchly's test for sphericity has been met for the main effects of taste on the five hip hop songs from the survey, $\chi^2(9) = 13.14$, $p = .156$. The repeated measures ANOVA revealed a significant main effect of taste to differently filtered hip hop songs, F (4, 180) = 5.71, $p < .001$. Two significant differences were found with the Bonferroni Correction between songs and whether the songs fit within the respondents' taste. The first between song 4 (Lupe Fiasco ft. Matthew Santos – Superstar, CF) and song 1(DJ Khaled ft. T-Pain, Ludacris, Snoop Dogg & Rick Ross - All I Do Is Win, CF) $p = .005$. And the second between song 4 and song (Soulja Boy – Crank That (Soulja Boy), CBF) $p = .017$.

Table 5. *Reports of means and standard deviation for rock likelihood and taste.*

	Filter type	Likelihood		Taste	
Rock		*M*	*SD*	*M*	*SD*
1. The Black Keys – Tighten Up	R	4.60	1.97	3.02	1.31
2. Kings of Leon – Use Somebody	CBF	4.31	2.06	2.75	1.26
3. Rise Against – Behind Closed Doors	CF	3.82	2.00	2.47	1.17
4. Muse – Uprising	CBF	4.93	1.84	3.11	1.16
5. Guns N' Roses - Sympathy For The Devil	CF	3.81	1.98	2.49	1.22

Out of 302 respondents 157 people chose the rock path. Mauchly's test indicated that the assumption of sphericity had been violated for the main effect likelihood to listen to the five rock songs, $\chi^2(9) = 32.27$, $p < .001$. Therefore degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity ($\varepsilon = .90$ for the main effect of likelihood to listen). The repeated measures ANOVA revealed a significant main effect on the likelihood to listen to differently filtered rock songs, $F\ (3.59, 560.38) = 13.84$, $p < .001$. Five significant differences were found with the Bonferroni Correction between songs and their likelihood to listen to the rock songs. The first between song 1 (The Black Keys – Tighten Up, R) and song 3 (Rise Against – Behind Closed Doors, CF) $p = .004$. The second between song 1 and song 5 (Guns N' Roses - Sympathy For The Devil, CF) $p = .001$. The third between song 2 (Kings of Leon – Use Somebody, CBF) and song 4 (Muse – Uprising, CBF) $p = .008$. The fourth between song 3 and song 4, $p < .001$. And finally, between song 4 and song 5, $p < .001$.

Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of taste on the five rock songs, $\chi^2(9) = 29.81$, $p < .001$. Therefore, the degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity ($\varepsilon = .91$ for the main effect of taste). The repeated measures ANOVA revealed a significant main effect of taste to differently filtered rock songs, $F\ (3.63, 565.92) = 12.20$, $p < .001$. Five significant differences were found with the Bonferroni Correction between rock songs and whether they fit within the respondents' taste. The first between song 1 (The Black Keys – Tighten Up, R) and song 3 (Rise Against – Behind Closed Doors, CF) $p = .001$. The second between song 1 and song 5 (Guns N' Roses - Sympathy For The Devil, CF) $p = .001$. The third between song 2 (Kings of Leon – Use Somebody, CBF) and song 4 (Muse – Uprising, CBF) $p = .020$. The fourth between song 3 and song 4, $p < .001$. And finally, between song 4 and song 5, $p < .001$.

Table 6. *Reports of means and standard deviation for pop likelihood and taste.*

	Filter type	Likelihood		Taste	
Pop		*M*	*SD*	*M*	*SD*
1. Coldplay ft. Rihanna – Princess of China	CBF	3.98	2.00	2.51	1.24
2. Britney Spears – 3	R	3.78	1.93	2.37	1.09
3. Timbaland ft. OneRepublic – Apologize	CBF	4.04	2.05	2.55	1.21
4. Taylor Swift – Love Story	CF	3.29	1.90	2.14	1.17

5. Adam Lambert - Whataya Want From Me	CF	3.78	2.13	2.53	1.21

Out of the 302 respondents 49 people chose the pop path. Mauchly's test for sphericity has been met for the main effects of likelihood to listen to the five pop songs from the survey, $\chi^2(9) = 16.14$, $p = .064$. The repeated measures ANOVA revealed a non-significant main effect on the likelihood to listen to differently filtered pop songs, $F\ (4, 192) = 1.42$, $p = .228$.

Mauchly's test indicated that the assumption of sphericity had been violated for the main effect taste on the five pop songs, $\chi^2(9) = 27.31$, $p = .001$, Therefore degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity ($\varepsilon = .80$ for the main effect of likelihood to listen).

The repeated measures ANOVA revealed a non-significant main effect of taste to differently filtered pop songs, $F\ (3.22, 154.37) = 1.70$, $p = .165$.

Table 7. *Reports of means and standard deviation for dance likelihood and taste.*

	Filter type	Likelihood		Taste	
Dance		*M*	*SD*	*M*	*SD*
1. Crystal Castles – Vanished	CF	4.12	2.01	2.42	1.24
2. California Swag District - Teach Me How To Dougie	CF	3.77	2.10	2.35	1.23
3. The Black Eyed Peas – I Gotta Feeling	CBF	3.96	2.18	2.54	1.36
4. Crookers ft. Kid Cudi - Day 'N' Nite	CBF	4.77	1.99	3.00	1.33
5. Justin Timberlake - LoveStoned/I Think She Knows	R	3.42	2.02	2.19	1.23

Out of 302 respondents 26 people chose the dance path. Mauchly's test for sphericity has been met for the main effects of likelihood to listen to the five dance songs from the survey, $\chi^2(9) = 11.70$, $p = .232$. The repeated measures ANOVA revealed a non-significant main effect on the likelihood to listen to differently filtered dance songs, $F\ (4, 100) = 2.22$, $p = .072$. However, one significant difference was found with the Bonferroni Correction between the pairwise comparisons of the songs and their likelihood to listen to the dance songs. Namely between song 4 (Crookers ft. Kid Cudi - Day 'N' Nite, CBF) and song 5 (Justin Timberlake - LoveStoned/I Think She Knows, R) $p = .044$.

Whether the dance songs fit within the musical taste of the respondents, Mauchly's test for sphericity has been met for the main effects of taste on the five hip hop songs from the survey, $\chi^2(9) = 12.95$, $p = .166$. The repeated measures ANOVA revealed a non-significant main effect of taste to differently filtered dance songs, $F\ (4, 100) = 2.01$, $p = .099$. However, one significant difference was found with the Bonferroni Correction between the pairwise comparisons of the dance songs and whether the songs fit within the respondents' taste. Namely between song 4 (Crookers ft. Kid Cudi - Day 'N' Nite, CBF) and song 5 (Justin Timberlake - LoveStoned/I Think She Knows, R) $p = .033$.

4.2.1.2 Comparison of means

To get an overview of the results, the means of the independent variable filter type and the dependent variable likelihood to listen were compared. The overall means per filter type and the likelihood to listen to a song of that filter type differed slightly per filter type, as stated the answers were given on a 7-point Likert scale (1 = extremely unlikely, …, 7 = extremely likely). CBF had an overall mean of 4.18, CF of 3.73, and Random of 3.82. Though the results differ minimally the results do show that overall the respondents tend to be more likely to listen to CBF recommended songs over CF or Random recommended songs. Interestingly, looking at the means it becomes apparent that after CBF songs, the respondents were more likely to listen to Randomly recommended songs than CF recommended songs.

The same test was done to compare the means of the independent variable filter type and the dependent variable taste. Again, the overall means per filter type and taste differed somewhat, here the answers to whether a song fit in a respondent's taste could be given on a 5-point Likert scale (1 = not well at all, …, 5 = extremely well). CBF had an overall mean of 2.67, CF of 2.41, and Random of 2.51. Considering the minimal differences, the respondents still overall gave a preference to the CBF recommendation. In this case as well, the respondents answered that second to CBF songs, Randomly recommended songs fit better within their musical taste over CF recommended songs.

In general the results seem to support H1, indicating that CBF generates a higher utility level for consumers than CF. Based on likelihood to listen and taste: CBF was superior to CF on six occasions, CF was superior to CBF on two occasions, CF was superior to R on one occasion, and R was superior to CF on five occasions. Moreover, it occurred two times that one CBF song had more consumer utility than the other CBF song, and it occurred four times that a CF song had more consumer utility over the other CF song. Interestingly, it is even the case that the R filtered songs were superior to the CF filtered songs.

4.2.2 Hypothesis 2

To assess H2: *Being a heavy music consumer would result in a preference towards CBF recommendations,* a Chi-square test was conducted. Chi-squares test the significance level of the association between categorical variables (Field, 2013; Pallant, 2013). For H2 it was first necessary to distinguish the heavy listeners from the light listeners. In order to do this the 'heavy listener scale' variable (see Chapter 3: Methods) was created. Consequently, the variable was divided into two groups with a median split (0 = light listener, 1 = heavy listener). For this hypothesis the listeners (heavy or light) were sorted into the columns as independent variables, and the filter types (CBF, CF, R) were sorted into the columns as dependent variables.

Table 8. C*rosstabulation of heavy and light listeners and their choice of favourite song out of five, recommended to them by the three filter types.*

		Heavy listener	Light listener	Total
Filter type				
CBF	Count	78	75	153
	% within	52.0%	49.3%	50.7%
CF	Count	44	45	89
	% within	29.3%	29.6%	29.5%
R	Count	28	32	60
	% within	18.7%	21.1%	19.9%
Total	Count	150	152	302
	% within	100.0%	100.0%	100.0%

The Chi-square test revealed that the type of listener (heavy or light) a respondent was, is insignificantly related to the kind of filter types they prefer in recommended songs, χ^2 (N = 302, 2) = .32, p = .851. Though the results are not significant, the crosstabulation does show that overall 52.0% of heavy listeners and 49.3% light listeners preferred Content Based Filtered songs over songs recommended to them by the other filter types. Considering this, due to the insignificant *p* value and minimal differences between the respondents' listener type and their respective filter preference, H2 can be rejected. Thus, being a heavy music consumer does not result in a preference towards CBF recommendations.

4.2.3 Hypothesis 3

In order to accept or reject H3: *Among light consumers, those who listen to music recommended by peers will have a preference towards CF recommendations;* a Chi-square test was conducted. A Chi-square was chosen for H3 because a Chi-square is able to test relationships between categorical variables (Field, 2013; Pallant, 2013). In this case the independent variable in the columns denotes whether respondents tend to, or not, listen to peer suggestions, and the rows denote the dependent variables namely the filter types (CBF, CF, R). For this analysis the heavy listeners have been excluded as this hypothesis concerns only light listeners, this is why the in the crosstabulation below *n* = 150. The light listeners were selected specifically via filtering out the respondents who were determined not to be heavy listeners based on the 'heavy listener scale' variable (see Chapter 3: Methods).

Table 9. *Crosstabulation of light listeners who either tend to listen to peer recommendations, or not, and their choice of favourite song out of five, recommended to them by the three filter types.*

Filter type		Tend not to listen to peer recommendations	Tend to listen to peer recommendations	Total
CBF	Count	44	34	78
	% within	54.3%	49.3%	52.0%
CF	Count	23	21	44
	% within	28.4%	30.4%	29.3%
R	Count	14	14	28
	% within	17.3%	20.3%	18.7%

Total	Count	81	69	150
	% within	100.0%	100.0%	100.0%

The Chi-square test revealed that respondents who tend to listen to peer suggestions, is insignificantly related to the kind of filter types they prefer in songs, χ^2 (N = 302, 2) = .42, p = .812. Though the results are not significant, the crosstabulation does show that overall 28.4% of respondents who tend not to listen to peer recommendations still chose CF recommended songs, and 30.4% of who did tend to listen to peer recommended songs chose the CF song as their favourite out of five. Moreover, 52.0% of light listeners preferred Content Based Filtered songs over songs recommended to them by the other filter types, regardless if they listened to peer recommendations or not. However, due to the insignificant p value and minimal differences between the respondents' preference, H3 can be rejected. Thus, being a light music consumer does not result in a preference towards CF recommendations.

Chapter 5. Discussion and Conclusion

This chapter offers a conclusion to the findings of this research, and it will answer the main research question of this thesis. Furthermore, the chapter will expound upon unexpected results and the limitations of the study, Finally, this chapter will make recommendations for future research on this field of study.

5.1 Discussion

As became apparent in the Results chapter, CBF was the most successful recommender type. These results followed the expectation that either CBF or CF had a different level of consumer utility over the other, what was not expected is the fact that the R filtered songs had a higher consumer utility than CF recommended songs. This could potentially be due to the fact that R recommended songs were chosen via also looking at Billboard charts. This would infer that chart popularity of the respective R filtered songs could have an impact on their consumer utility. Through the CF code of the Million Song Dataset the output was also popularity driven, but solely based on the users within that dataset, and not considering a song's popularity outside of it. This could be an explanation to why the songs that came from the R filter attained a higher consumer utility among respondents over the CF filtered songs, as were connected to (international) Billboard music charts.

Respondents were unaware of the filter types behind the recommendations. In this same way that awareness might have an influence on their choices, unawareness might also influence the choices that respondents made. For instance, if a respondent considers themselves more in line with a heavy listener then they might make more conscious choices to select CBF choices over CF or R recommended songs. However, it was purposefully chosen to leave the respondents unaware of these processes as in reality this is the case too.

5.2 Conclusion and implications

In order to answer the main research question for this research: *To what extent might either Content Based Filtering or Collaborative Filtering result in the highest consumer utility on music streaming platforms?*, three hypotheses were setup up. From the analysis, only H1: *There is a difference within consumer utility depending upon which recommender type is used: either CBF or CF*, was accepted as overall the results seem to support H1, indicating that CBF generates more utility for consumers than CF. Based on the results of the repeated measures ANOVA the respondents' likelihood to listen to songs and their taste seemed to generally be highest in the cases where CBF was applied. This means that there is in fact a difference between consumer utility between the two filter types. Alongside this, with CBF being the more successful recommender, the results align with the findings of Holbrook et al. (2006), who have found that consumers do appreciate expert judgements, and Yeomans et al. (2019) who point out the fruitful ability that human curators and critics have to consider contextuality and subjectivity.

H2: *Being a heavy music consumer would result in a preference towards CBF recommendations*, and H3: *Among light consumers, those who listen to music recommended by peers will have a preference towards CF recommendations*, both had to be rejected as there were insignificant results for either hypothesis. This would imply that regardless whether a consumer is a heavy or light listener this does not affect the consumer utility of the different recommender types behind each song. However, the insignificance of these results could also be due to the relatively small sample size per genre, this point will be expanded upon in the limitations section of this chapter.

This study offers a novel and systematic way to conduct research on the recommendation of songs via either CBF or CF methods. As explained in detail in the Methods chapter, a step by step description is provided on what methods were utilised to make the recommendations. In this way the process could be replicable and useful for future research on the subject. As stated, the Million Song Dataset is open access for everyone and was made to encourage research on recommendation algorithms (Bertin-Mahieux et al., 2011), this research has aimed to do exactly that. This research has aimed to contribute to the field of research on (online) product recommendation, especially in the case of music. One of the aims of this research was to provide a replicable method for research on any type of entertainment product recommendation. This would infer that if the dataset was made for books, films, or any other entertainment products, the research could be conducted in a similar manner. What would differ most herein would be the labels of the CBF meta-data, as these would are highly contextual. The CF recommendations are most generally standardised in their process (Burke, 2007; Ricci et al., 2011), thus they would not differ much in their workings, solely in their output.

Considering the main findings of this research, CBF recommendations were superior to CF or R recommendations. This could infer that the role of critics and curators is still very valuable and necessary in the music industry. As in this research CBF recommendation have proven to be most preferable, it is in fact the critic and curator that knows best in terms of attaining the highest consumer utility. Following this, the music industry would benefit from hiring and training skilled critics and curators, as Prey (2017) and Yeomans et al. (2010) argue as well. Logically, satisfied consumers would make more revenue. Therefore, investing in the best methods for song recommendation, in this case CBF, would result in higher consumer utility. In this case music streaming platforms should invest in hiring such skilled professionals, who with their knowledge will be able to provide well-suited meta-data to songs and in turn can make the algorithms smarter and more advanced.

5.3 Limitations

This study was conducted through making thoughtful considerations of appropriate recommendation methods and creating relevant relations between existing literature surrounding the subject of study. However, limitations are inevitable; hence, in this section the most significant limitations of this research will be discussed.

One of the main limitations for the selection of the songs for the survey was the fact that the dataset was rather outdated. As stated, the One Million Song Dataset consisted of songs up to and including the year 2011. Even though the feedback received from the respondents, especially on Reddit, was quite positive, the survey was regarded as containing songs more appropriate in a 'Throwback Thursday' playlist due to the dataset limitation. If the dataset had been more contemporary, including eight more years of music, then recommendations had more chances to be novel rather than serving as a trip down memory lane. Though, it must be noted that some respondents exclaimed that they had found new songs to add to their personal playlists as well, especially in the rock path which had the most respondents.

The creators, Bertin-Mahieux et al., also discuss a number of limitations of the dataset. They point out that "the dataset is currently lacking album and song-level metadata and tags" (2011, p. 594). Thus, Content Based Filtered recommendations could not be done through this dataset. As the CBF recommended songs were labelled by myself complete objectivity in the labelling process is largely unattainable. At times I would come across songs that would match perfectly, based on CBF characteristics, to the starting song in a particular genre but it could happen that were not present within the Million Song Dataset, thus another song had to be opted for. In this way more secondary CBF recommendations were made. Furthermore, the creators of the dataset rightfully state that "[d]iversity is another issue: there is little or no world, ethnic, and classical music" (2011, p. 594). If the dataset had been up to date and no world music or ethnic music was included, the dataset could miss the mark with mark even more, as in an increasingly globalised society more people are open to listen to non-Western music (Straubhaar, 2014).

The sampling methods, selective and convenience, utilised for this research are generally not advised when conducting a quantitative study. Although, respondents also took part in self-selected sampling, the survey was purposefully posted in a variety of social media platforms and their respective within platform domains (i.e. Reddit threads) therefore, the respondents still were subject to selective and convenience sampling. This in turn could be one of the explanations for the unbalanced demographic, especially concerning nationality and place of residence. As stated, the United States (110 respondents, 34.8%), The Netherlands (60, 19.0%), and Canada (15, 4.7%) were the most represented nationalities, and the represented places of residence were United States (112, 35.4%), The Netherlands (70, 22.2%), and the United Kingdom (19, 6.0%). This disbalance thus must be concluded as a sample bias as neither literature nor statistics support such a national divide.

Mood was listed as one of the contributing factors in the decision-making processes of consumers when selecting a song to listen to. However, as discussed in the Methods chapter, the respondents found it difficult to express themselves regarding mood in the survey. Due to this, and the fact that there was no previous literature nor exemplary method on how mood could be incorporated in song recommendation this element was left out from the final analysis and deemed beyond the scope of this research.

Overall, this study reached a decent number of respondents to conduct this research. However, once the respondents chose one of the five genre paths to go down the units of analysis shrunk per genre, and were divided unequally. This could be one of the main explanations why H2 and H3 had to be rejected. Thus, for future research a larger sample size, for the entire survey or per genre path, would be advisable.

5.4 Future research

More research should be done exploring the kinds of labels being attributed to entertainment content, the reasoning behind these existing labels, and the level of subjectivity within them as well.

Another direction for future research could be a study on respondents' song choices when made aware of the filter types and mechanisms behind their recommendations. As Prey (2017) and Yeomans et al. (2019) point out consumers are unaware of the processes behind their recommendations and the recommender types at play. Furthermore, consumers are also unaware that undemocratic processes are at hand as well, as platform it is in a platform's interest to promote their own platform exclusive content via their own algorithms (Morris & Powers, 2015).

As the Million Song Dataset was created as an open access, free of charge research project, it would benefit from updating its content. In this way newer songs can be taken into consideration for the CF recommendations. With the addition of these new songs, studies using this dataset would not have these similar year-related limitations. Moreover, if metadata would be applied to the dataset as well, CBF recommendations could be done too. Subsequently, if this process is thus standardised via two sets of code, one for CF and one for CBF, then the levels of subjectivity within recommendations will arguably also decrease.

All in all, this research has aimed to provide a methodological basis upon which future researchers could be able to build. The results from this study and the theoretical framework which this thesis is based upon can hopefully serve as a point to advance from for future researchers.

References

AllMusic. (n.d.). *Panic! At the Disco*. Retrieved from https://www.allmusic.com/artist/panic%21-at-the-disco-mn0000318085

Barrett, L. F., & Russell, J. A. (1999). The structure of current affect: Controversies and emerging consensus. *Current Directions in Psychological Science, 8(1)*, 10–14. https://doi.org/10.1111/1467-8721.00003

Bertin-Mahieux, T., Ellis, D. P. W., Whitman, B. & Lamere, P. (2011). The Million Song Dataset. In Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011).

Billboard. (n.d.). *Billboard charts archive*. Retrieved from https://www.billboard.com/archive/charts

Bourdieu, P. (1984). *Distinction: A Social Critique on the Judgement of Taste*. Cambridge, MA: Harvard University Press.

Burke, R. (2007). Hybrid web recommender systems. In Brusilovski, P., Kobsa, A. & Nejdl, W. (Ed.) *The adaptive web:* Methods and strategies of web personalization (pp. 377-408). Berlin: Springer Science & Business Media.

Bryman, A. (2012). *Social Research Methods*. New York: Oxford University Press.

Ciocca, S. (2017, October 10). How does Spotify know you so well? *Medium, 10.* Retrieved from https://medium.com/s/story/spotifys-discover-weekly-how-machine-learning-finds-your-new-music-19a41ab76efe

DeNora, T. (1999). Music as a technology of the self. *Poetics* (27). New York: Elsevier.

DeNora, T. (2000). *Music in everyday life*. Cambridge: Cambridge University Press.

Dillet, R. (2016, December 1). Spotify launches new series of original recordings called Spotify Singles. *Tech Crunch.* Retrieved from https://techcrunch.com/2016/12/01/spotify-launches-new-series-of-original-recordings-called-spotify-singles

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage.

Fitzsimons, G. J., Hutchinson, J. W., Williams, P., Alba, J. W., Chartrand, T. L., Huber, J., ... & Shiv, B. (2002). Non-conscious influences on consumer choice. *Marketing Letters*, *13*(3), 269-279.

Fleder, D. & Hosanagar, K. (2009). Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management science*, *55*(5), 697-712.

Friedlander, J. P. & Bass, M. (2018). *Mid-year 2018 RIAA music revenues report*. Retrieved from https://www.riaa.com/wp-content/uploads/2018/09/RIAA-Mid-Year-2018-Revenue-Report.pdf

Frow, J. (2006). *Genre: The New Critical Idiom*. London: Routledge.

GitHub. (n.d.). *kswamy15 / song-recommender.* Retrieved from https://github.com/kswamy15/song-recommender/blob/master/song%20recommender%20swamy%20083016.ipynb

Hennion, A. (2001). Music lovers: taste as performance. *Theory, Culture and Society 18*, 1–22.

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, *22*(1), 5-53.

Hills, M. (2002). *Fan Cultures*. London: Routledge.

Holbrook, M. B., Lacher, K. T. & LaTour, M. S. (2006). Audience judgments as the potential missing link between expert judgments and audience appeal: An illustration based on musical recordings of “My funny Valentine”. *Journal of the Academy of Marketing Science*, *34*(1), 8-18.

Jenkins, H. (1992). *Textual poachers: Television fans & participatory culture*. New York: Routledge.

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

Logg, J. M. (2017, March 13). *Theory of machine: When do people rely on algorithms?* Retrieved from https://dash.harvard.edu/handle/1/31677474

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Mackay, A. (1981). *Electronic music (No. 2)*. London: Control Data Publishing.

McIntyre, H. (2018, May 25). *The top 10 streaming music services by number of users*. Retrieved from https://www.forbes.com/sites/hughmcintyre/2018/05/25/the-top-10-streaming-music-services-by-number-of-users/

Neuman, W.L. (2014). Experimental Research. In: WL. Neuman (Ed.) *Social Research Methods: Qualitative and Quantitative Approaches* (7th Ed.), 281-313. Essex: Pearson.

Pallant, J. (2013). *SPSS survival manual*. McGraw-Hill Education. United Kingdom: Cambridge

Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In Brusilovski, P., Kobsa, A. & Nejdl, W. (Ed.) *The adaptive web:* Methods and strategies of web personalization (pp. 325-341). Berlin: Springer Science & Business Media.

Peterson, R. A. & Kern, R. M. (1996). Changing highbrow taste: From snob to omnivore. *American Sociological Review*, 900-907.

Python. (n.d.). *Python about*. Retrieved from https://www.python.org/about/

Ricci, F., Rokach, L., Shapira, B. & Kantor, P. B. (2011). *Recommender systems handbook*. Boston, MA: Springer.

Sarwar, B. M., Karypis, G., Konstan, J. A., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. *Www*, *1*, 285-295.

Shocker, A. D., Ben-Akiva, M., Boccara, B. & Nedungadi, P. (1991). Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions. *Marketing letters*, *2*(3), 181-197.

Straubhaar, J.D. (2014). Mapping “global” in global communication and media studies. In: K.G. Wilkins, J.D. Straubhaar & S. Kumar (Eds.), *Global communication*. New agendas in communication (pp.10-34), London, UK: Routledge

Sundar, S. S., Oeldorf-Hirsch, A. & Xu, Q. (2008). The bandwagon effect of collaborative filtering technology. In *CHI'08 extended abstracts on Human factors in computing systems* (pp. 3453-3458). ACM.

Surowiecki, J. (2005). *The wisdom of crowds*. New York, NY: Anchor.

Wlömert, N. & Papies, D. (2016). On-demand streaming services and music industry revenues—Insights from Spotify's market entry. *International Journal of Research in Marketing*, *33*(2).

Yeomans, M., Shah, A., S., Mullainathan, S. & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making*, 1– 12. https://doi.org/10.1002/bdm.2118

Zillmann, D. (1988a). Mood management through communication choices. *American Behavioral Scientist*, *31*, 327–340.

Zillmann, D. (1988b). Mood management: Using entertainment to full advantage. In L. Donohew, H. E. Sypher, & E. T. Higgins (Eds.), *Communication, social cognition, and affect* (pp. 147–171). Hillsdale, NJ: Erlbaum.

Appendix A. Survey

Start of Block: Introduction

Q24 Thank you for taking the time to fill out this survey!
My name is Zoë Voyle, I am a Master student at Erasmus University Rotterdam, and I am currently conducting a research project about algorithmic recommendation types within music streaming platforms. I am interested in your music preferences between the choices put forward within the survey. In the following, you will be asked about your music listening habits. Consequently, you will come across a total of 5 songs recommended to you via two types of filter methods. The aim of this survey is to assess whether one of the filter methods makes more accurate recommendations than the other.

Please answer each question truthfully and according to your own opinion, there are no correct or incorrect answers.

The data from this survey shall only be used for research purposes and the responses will remain anonymous. Note that you are free to abandon the survey at any time.
The survey will take around 10 minutes.

If you have any questions or comments, do not hesitate to contact me via email: 431764zv@student.eur.nl
With kind regards, Zoë Voyle
By clicking "I agree" you consent to taking part in this research survey.

○ I agree (1)

Page Break

Q20
Which of the following music streaming platform(s) do you make use of?
(Multiple answers possible)

- ☐ Apple Music (1)
- ☐ Deezer (2)
- ☐ Google Play Music (3)
- ☐ Pandora (4)
- ☐ SoundCloud (5)
- ☐ Spotify (6)
- ☐ Tidal (7)
- ☐ YouTube (8)
- ☐ Other, please specify (9) ____
- ☐ I do not make use of any music streaming platforms. (10)

Skip To: End of Survey If Which of the following music streaming platform(s) do you make use of?(Multiple answers possible) = I do not make use of any music streaming platforms.

Q125 Do you have a paid subscription for any of the music streaming services you use?

- ○ Yes (5)
- ○ No (6)

Page Break

Q1 What is your age?

____

Skip To: End of Survey If What is your age? <= 17

Page Break

Q2 What is your gender?

- ○ Male (1)
- ○ Female (2)
- ○ Other (3)
- ○ Prefer not to say (4)

Page Break

Q19
What is your native country?

▼ Afghanistan (1) ... Zimbabwe (1357)

Q25 In which country do you currently reside?

▼ Afghanistan (1) ... Zimbabwe (1357)

Page Break

Q4 What is your level of education?

- ◯ No formal education (1)
- ◯ Secondary education (2)
- ◯ Vocational education (3)
- ◯ University degree (BA) (4)
- ◯ University degree (MA) (5)
- ◯ University degree (PhD) (6)

Page Break

End of Block: Introduction

Start of Block: Music profiling & Consumption

Q9 On average, how many songs do you listen to per day?

____

Q10 On average, how many hours do you listen to music per day?

____

Page Break

Q11 Please indicate to which extent you disagree or agree with the following statements:

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
I consider myself a heavy music listener. (1)	○	○	○	○	○	○	○
I consider myself a fan of music. (2)	○	○	○	○	○	○	○
I consider myself a lover of music. (3)	○	○	○	○	○	○	○
I will listen to music recommendations suggested by my peers. (4)	○	○	○	○	○	○	○
I will listen to music recommendations suggested by (journalistic) music sources. (5)	○	○	○	○	○	○	○
I like to find music on my own. (6)	○	○	○	○	○	○	○

Page Break

Q124 Please give an indication about your current mood:

	1 (1)	2 (2)	3 (3)	
Nervous	○	○	○	Calm
Upset	○	○	○	Happy
Indifferent	○	○	○	Excited
Fatigued	○	○	○	Energised

End of Block: Music profiling & Consumption

Start of Block: Filters

Q17
Which of the following music genres and its song appeal to you the most?
Please pick one of the five songs identified below:

Feel free to listen to the audio tracks if you do not recognise some of the following songs.

○ Pop: Lady Gaga - Alejandro (1)

○ Rock: Red Hot Chili Peppers – Snow (Hey Oh) (2)

○ Hip Hop: Kanye West ft. Chris Martin - Homecoming (3)

○ R&B: Leona Lewis - Bleeding Love (4)

○ Dance: Major Lazor ft. Vybz Kartel & Afrojack – Pon De Floor (5)

Page Break

Q27
Based on your music listening habits and music consumption **5 song recommendations** have been made.
Please note: due to technical limitations the songs presented in this survey will have been released in or before the year 2011.

End of Block: Filters

Start of Block: R&B, Song 1: D'Angelo - Untitled (How Does It Feel)

Q108 D'Angelo - Untitled (How Does It Feel)

Are you familiar with this song?

○ Yes (1)

○ No (2)

Q109 Are you familiar with this artist/band?

○ Yes (1)

○ No (2)

Q110 How likely will you be to listen to this song?

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Q111 Does this song fit within your musical taste?

- Not well at all (1)
- Slightly well (2)
- Moderately well (3)
- Very well (4)
- Extremely well (5)

Q130 Would you add this song to your playlist?

- Yes (5)
- No (6)

End of Block: R&B, Song 1: D'Angelo - Untitled (How Does It Feel)

Start of Block: R&B, Song 2: Rihanna - Te Amo

Q112 Rihanna - Te Amo

Are you familiar with this song?

- Yes (1)
- No (2)

Q113 Are you familiar with this artist/band?

- Yes (1)
- No (2)

Q114 How likely will you be to listen to this song?

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Q115 Does this song fit within your musical taste?

- Not well at all (1)
- Slightly well (2)
- Moderately well (3)
- Very well (4)
- Extremely well (5)

Q131 Would you add this song to your playlist?

- Yes (5)
- No (6)

End of Block: R&B, Song 2: Rihanna - Te Amo

Start of Block: R&B, Song 3: Beyoncé – Sweet Dreams

Q116 Beyoncé – Sweet Dreams

Are you familiar with this song?

- Yes (1)
- No (2)

Q117 Are you familiar with this artist/band?

- Yes (1)
- No (2)

Q118 How likely will you be to listen to this song?

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Q119 Does this song fit within your musical taste?

- Not well at all (1)
- Slightly well (2)
- Moderately well (3)
- Very well (4)
- Extremely well (5)

Q132 Would you add this song to your playlist?

- Yes (5)
- No (6)

End of Block: R&B, Song 3: Beyoncé – Sweet Dreams

Start of Block: R&B, Song 4: Keyshia Cole – Fallin' Out

Q120 Keyshia Cole – Fallin' Out

Are you familiar with this song?

- Yes (1)
- No (2)

Q121 Are you familiar with this artist/band?

- Yes (1)
- No (2)

Q122 How likely will you be to listen to this song?

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Q123 Does this song fit within your musical taste?

○ Not well at all (1)

○ Slightly well (2)

○ Moderately well (3)

○ Very well (4)

○ Extremely well (5)

Q133 Would you add this song to your playlist?

○ Yes (5)

○ No (6)

End of Block: R&B, Song 4: Keyshia Cole – Fallin' Out

Start of Block: R&B, Song 5: Justin Bieber – One Time

Q124 Justin Bieber – One Time

Are you familiar with this song?

○ Yes (1)

○ No (2)

Q125 Are you familiar with this artist/band?

○ Yes (1)

○ No (2)

Q126 How likely will you be to listen to this song?

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Q127 Does this song fit within your musical taste?

- Not well at all (1)
- Slightly well (2)
- Moderately well (3)
- Very well (4)
- Extremely well (5)

Q134 Would you add this song to your playlist?

- Yes (5)
- No (6)

End of Block: R&B, Song 5: Justin Bieber – One Time

Start of Block: R&B, final questions

Q128 Out of the five previous songs, which one is your favourite?

- ◯ D'Agelo - Untitled (How Does It Feel) (1)
- ◯ Rihanna - Te Amo (2)
- ◯ Beyoncé - Sweet Dreams (3)
- ◯ Keyshia Cole - Fallin' Out (4)
- ◯ Justin Bieber – One Time (5)

Q122 As a consumer of music streaming platforms, please indicate to what extent you disagree or agree with the following statements:

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
I am satisfied with the recommendations music streaming platforms provide me. (1)	◯	◯	◯	◯	◯	◯	◯
I would prefer to get recommendations based on the music taste of my peers. (2)	◯	◯	◯	◯	◯	◯	◯
I would prefer to get music recommendations based on the expertise of curators. (3)	◯	◯	◯	◯	◯	◯	◯
I would prefer to get a mix of music recommendations based on listening habits from my peers and curated content. (4)	◯	◯	◯	◯	◯	◯	◯

Q123 Did you search for any additional information about the artists, cover art and/or songs you came across while answering the survey?

- Yes (1)
- No (2)

End of Block: R&B, final questions

Start of Block: Hip hop, Song 1: DJ Khaled ft. T-Pain, Ludacris, Snoop Dogg & Rick Ross - All I

Q86 DJ Khaled ft. T-Pain, Ludacris, Snoop Dogg & Rick Ross - All I Do Is Win

Are you familiar with this song?

- Yes (1)
- No (2)

Q87 Are you familiar with this artist/band?

- Yes (1)
- No (2)

Q88 How likely will you be to listen to this song?

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Q89 Does this song fit within your musical taste?

- Not well at all (1)
- Slightly well (2)
- Moderately well (3)
- Very well (4)
- Extremely well (5)

Q135 Would you add this song to your playlist?

- Yes (5)
- No (6)

End of Block: Hip hop, Song 1: DJ Khaled ft. T-Pain, Ludacris, Snoop Dogg & Rick Ross - All I

Start of Block: Hip hop, Song 2: Soulja Boy – Crank That (Soulja Boy)

Q90 Soulja Boy – Crank That (Soulja Boy)

Are you familiar with this song?

- ◯ Yes (1)
- ◯ No (2)

Q91 Are you familiar with this artist/band?

- ◯ Yes (1)
- ◯ No (2)

Q92 How likely will you be to listen to this song?

- ◯ Extremely unlikely (1)
- ◯ Moderately unlikely (2)
- ◯ Slightly unlikely (3)
- ◯ Neither likely nor unlikely (4)
- ◯ Slightly likely (5)
- ◯ Moderately likely (6)
- ◯ Extremely likely (7)

Q93 Does this song fit within your musical taste?

○ Not well at all (1)

○ Slightly well (2)

○ Moderately well (3)

○ Very well (4)

○ Extremely well (5)

Q136 Would you add this song to your playlist?

○ Yes (5)

○ No (6)

End of Block: Hip hop, Song 2: Soulja Boy – Crank That (Soulja Boy)

Start of Block: Hip hop, Song 3: Young Money ft. Lloyd - BedRock

Q94 Young Money ft. Lloyd - BedRock

Are you familiar with this song?

○ Yes (1)

○ No (2)

Q95 Are you familiar with this artist/band?

○ Yes (1)

○ No (2)

Q96 How likely will you be to listen to this song?

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Q97 Does this song fit within your musical taste?

- Not well at all (1)
- Slightly well (2)
- Moderately well (3)
- Very well (4)
- Extremely well (5)

Q137 Would you add this song to your playlist?

- Yes (5)
- No (6)

End of Block: Hip hop, Song 3: Young Money ft. Lloyd - BedRock

Start of Block: Hip hop, Song 4: Lupe Fiasco ft. Matthew Santos - Superstar

Q98
Lupe Fiasco ft. Matthew Santos - Superstar

Are you familiar with this song?

- Yes (1)
- No (2)

Q99 Are you familiar with this artist/band?

- Yes (1)
- No (2)

Q100 How likely will you be to listen to this song?

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Q101 Does this song fit within your musical taste?

- Not well at all (1)
- Slightly well (2)
- Moderately well (3)
- Very well (4)
- Extremely well (5)

Q138 Would you add this song to your playlist?

- Yes (5)
- No (6)

End of Block: Hip hop, Song 4: Lupe Fiasco ft. Matthew Santos - Superstar

---

Start of Block: Hip hop, Song 5: Drake ft. Lil Wayne & Young Jeezy – I'm Goin In

Q102 Drake ft. Lil Wayne & Young Jeezy – I'm Goin In

Are you familiar with this song?

- Yes (1)
- No (2)

Q103 Are you familiar with this artist/band?

- Yes (1)
- No (2)

Q104 How likely will you be to listen to this song?

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Q105 Does this song fit within your musical taste?

- Not well at all (1)
- Slightly well (2)
- Moderately well (3)
- Very well (4)
- Extremely well (5)

Q139 Would you add this song to your playlist?

- Yes (5)
- No (6)

End of Block: Hip hop, Song 5: Drake ft. Lil Wayne & Young Jeezy – I'm Goin In

Start of Block: Hip hop, final questions

Q106 Out of the five previous songs, which one is your favourite?

- DJ Khaled ft. T-Pain, Ludacris, Snoop Dogg & Rick Ross - All I Do Is Win (1)
- Soulja Boy – Crank That (Soulja Boy) (2)
- Young Money ft. Lloyd - BedRock (3)
- Lupe Fiasco ft. Matthew Santos - Superstar (4)
- Drake ft. Lil Wayne & Young Jeezy – I'm Goin In (5)

Q124 As a consumer of music streaming platforms, please indicate to what extent you disagree or agree with the following statements:

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
I am satisfied with the recommendations music streaming platforms provide me. (1)	○	○	○	○	○	○	○
I would prefer to get recommendations based on the music taste of my peers. (2)	○	○	○	○	○	○	○
I would prefer to get music recommendations based on the expertise of curators. (3)	○	○	○	○	○	○	○
I would prefer to get a mix of music recommendations based on listening habits from my peers and curated content. (4)	○	○	○	○	○	○	○

Q125 Did you search for any additional information about the artists, cover art and/or songs you came across while answering the survey?

○ Yes (1)

○ No (2)

End of Block: Hip hop, final questions

Start of Block: Rock, Song 1: The Black Keys – Tighten Up

Q64 The Black Keys – Tighten Up

Are you familiar with this song?

○ Yes (1)

○ No (2)

Q65 Are you familiar with this artist/band?

○ Yes (1)

○ No (2)

Q66 How likely will you be to listen to this song?

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Q67 Does this song fit within your musical taste?

- Not well at all (1)
- Slightly well (2)
- Moderately well (3)
- Very well (4)
- Extremely well (5)

Q140 Would you add this song to your playlist?

- Yes (5)
- No (6)

End of Block: Rock, Song 1: The Black Keys – Tighten Up

---

tart of Block: Rock, Song 2: Kings of Leon – Use Somebody

Q68
Kings of Leon – Use Somebody

Are you familiar with this song?

- Yes (1)
- No (2)

Q69 Are you familiar with this artist/band?

- Yes (1)
- No (2)

Q70 How likely will you be to listen to this song?

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Q71 Does this song fit within your musical taste?

◯ Not well at all (1)

◯ Slightly well (2)

◯ Moderately well (3)

◯ Very well (4)

◯ Extremely well (5)

Q141 Would you add this song to your playlist?

◯ Yes (5)

◯ No (6)

End of Block: Rock, Song 2: Kings of Leon – Use Somebody

Start of Block: Rock, Song 3: Rise Against – Behind Closed Doors

Q72 Rise Against – Behind Closed Doors

Are you familiar with this song?

◯ Yes (1)

◯ No (2)

Q73 Are you familiar with this artist/band?

◯ Yes (1)

◯ No (2)

Q74 How likely will you be to listen to this song?

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Q75 Does this song fit within your musical taste?

- Not well at all (1)
- Slightly well (2)
- Moderately well (3)
- Very well (4)
- Extremely well (5)

Q142 Would you add this song to your playlist?

- Yes (5)
- No (6)

End of Block: Rock, Song 3: Rise Against – Behind Closed Doors

Start of Block: Rock, Song 4: Muse – Uprising

Q76 Muse – Uprising

Are you familiar with this song?

- Yes (1)
- No (2)

Q77 Are you familiar with this artist/band?

- Yes (1)
- No (2)

Q78 How likely will you be to listen to this song?

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Q79 Does this song fit within your musical taste?

- Not well at all (1)
- Slightly well (2)
- Moderately well (3)
- Very well (4)
- Extremely well (5)

Q143 Would you add this song to your playlist?

- Yes (5)
- No (6)

End of Block: Rock, Song 4: Muse – Uprising

Start of Block: Rock, Song 5: Guns N' Roses - Sympathy For The Devil

Q80 Guns N' Roses - Sympathy For The Devil

Are you familiar with this song?

- Yes (1)
- No (2)

Q81 Are you familiar with this artist/band?

- Yes (1)
- No (2)

Q82 How likely will you be to listen to this song?

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Q83 Does this song fit within your musical taste?

- Not well at all (1)
- Slightly well (2)
- Moderately well (3)
- Very well (4)
- Extremely well (5)

Q144 Would you add this song to your playlist?

- Yes (5)
- No (6)

End of Block: Rock, Song 5: Guns N' Roses - Sympathy For The Devil

---

tart of Block: Rock, final questions

Q84 Out of the five previous songs, which one is your favourite?

- The Black Keys – Tighten Up (1)
- Kings of Leon – Use Somebody (2)
- Rise Against – Behind Closed Doors (3)
- Muse – Uprising (4)
- Guns N' Roses - Sympathy For The Devil (5)

Q126 As a consumer of music streaming platforms, please indicate to what extent you disagree or agree with the following statements:

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
I am satisfied with the recommendations music streaming platforms provide me. (1)	○	○	○	○	○	○	○
I would prefer to get recommendations based on the music taste of my peers. (2)	○	○	○	○	○	○	○
I would prefer to get music recommendations based on the expertise of curators. (3)	○	○	○	○	○	○	○
I would prefer to get a mix of music recommendations based on listening habits from my peers and curated content. (4)	○	○	○	○	○	○	○

Q127 Did you search for any additional information about the artists, cover art and/or songs you came across while answering the survey?

○ Yes (1)

○ No (2)

End of Block: Rock, final questions

Start of Block: Pop, Song 1: Coldplay ft. Rihanna – Princess of China

Q5
Coldplay ft. Rihanna – Princess of China

Are you familiar with this song?

○ Yes (1)

○ No (2)

Q6
Are you familiar with this artist/band?

○ Yes (1)

○ No (2)

Q7
How likely will you be to listen to this song?

- ○ Extremely unlikely (1)
- ○ Moderately unlikely (2)
- ○ Slightly unlikely (3)
- ○ Neither likely nor unlikely (4)
- ○ Slightly likely (5)
- ○ Moderately likely (6)
- ○ Extremely likely (7)

Q8
Does this song fit within your musical taste?

- ○ Not well at all (1)
- ○ Slightly well (2)
- ○ Moderately well (3)
- ○ Very well (4)
- ○ Extremely well (5)

Q145 Would you add this song to your playlist?

- ○ Yes (5)
- ○ No (6)

End of Block: Pop, Song 1: Coldplay ft. Rihanna – Princess of China

Start of Block: Pop, Song 2: Britney Spears - 3

Q19 Britney Spears - 3

Are you familiar with this song?

- Yes (1)
- No (2)

Q20
Are you familiar with this artist/band?

- Yes (1)
- No (2)

Q21
How likely will you be to listen to this song?

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Q22
Does this song fit within your musical taste?

- ○ Not well at all (1)
- ○ Slightly well (2)
- ○ Moderately well (3)
- ○ Very well (4)
- ○ Extremely well (5)

Q146 Would you add this song to your playlist?

- ○ Yes (5)
- ○ No (6)

End of Block: Pop, Song 2: Britney Spears - 3

Start of Block: Pop, Song 3: Timbaland ft. OneRepublic – Apologize

Q23 Timbaland ft. OneRepublic – Apologize

Are you familiar with this song?

- ○ Yes (1)
- ○ No (2)

Q24
Are you familiar with this artist/band?

- ○ Yes (1)
- ○ No (2)

Q25
How likely will you be to listen to this song?

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Q26
Does this song fit within your musical taste?

- Not well at all (1)
- Slightly well (2)
- Moderately well (3)
- Very well (4)
- Extremely well (5)

Q147 Would you add this song to your playlist?

- Yes (5)
- No (6)

End of Block: Pop, Song 3: Timbaland ft. OneRepublic – Apologize

---

Start of Block: Pop, Song 4: Taylor Swift – Love Story

Q27
Taylor Swift – Love Story

Are you familiar with this song?

- ○ Yes (1)
- ○ No (2)

Q28 Are you familiar with this artist/band?

- ○ Yes (1)
- ○ No (2)

Q29 How likely will you be to listen to this song?

- ○ Extremely unlikely (1)
- ○ Moderately unlikely (2)
- ○ Slightly unlikely (3)
- ○ Neither likely nor unlikely (4)
- ○ Slightly likely (5)
- ○ Moderately likely (6)
- ○ Extremely likely (7)

Q30 Does this song fit within your musical taste?

- Not well at all (1)
- Slightly well (2)
- Moderately well (3)
- Very well (4)
- Extremely well (5)

Q148 Would you add this song to your playlist?

- Yes (5)
- No (6)

End of Block: Pop, Song 4: Taylor Swift – Love Story

Start of Block: Pop, Song 5: Adam Lambert - Whataya Want From Me

Q31 Adam Lambert - Whataya Want From Me

Are you familiar with this song?

- Yes (1)
- No (2)

Q32 Are you familiar with this artist/band?

- Yes (1)
- No (2)

Q33 How likely will you be to listen to this song?

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Q34 Does this song fit within your musical taste?

- Not well at all (1)
- Slightly well (2)
- Moderately well (3)
- Very well (4)
- Extremely well (5)

Q149 Would you add this song to your playlist?

- Yes (5)
- No (6)

End of Block: Pop, Song 5: Adam Lambert - Whataya Want From Me

Start of Block: Pop, final questions

Q26 Out of the five previous songs, which one is your favourite?

- Coldplay ft. Rihanna – Princess of China (1)
- Britney Spears - 3 (2)
- Timbaland ft. OneRepublic – Apologize (3)
- Taylor Swift – Love Story (4)
- Adam Lambert - Whataya Want From Me (5)

Q128 As a consumer of music streaming platforms, please indicate to what extent you disagree or agree with the following statements:

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
I am satisfied with the recommendations music streaming platforms provide me. (1)	○	○	○	○	○	○	○
I would prefer to get recommendations based on the music taste of my peers. (2)	○	○	○	○	○	○	○
I would prefer to get music recommendations based on the expertise of curators. (3)	○	○	○	○	○	○	○
I would prefer to get a mix of music recommendations based on listening habits from my peers and curated content. (4)	○	○	○	○	○	○	○

Q129 Did you search for any additional information about the artists, cover art and/or songs you came across while answering the survey?

○ Yes (1)

○ No (2)

End of Block: Pop, final questions

Start of Block: Dance, Song 1: Crystal Castles - Vanished

Q130 Crystal Castles - Vanished

Are you familiar with this song?

○ Yes (1)

○ No (2)

Q131 Are you familiar with this artist/band?

○ Yes (1)

○ No (2)

Q132 How likely will you be to listen to this song?

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Q133 Does this song fit within your musical taste?

- Not well at all (1)
- Slightly well (2)
- Moderately well (3)
- Very well (4)
- Extremely well (5)

Q150 Would you add this song to your playlist?

- Yes (5)
- No (6)

End of Block: Dance, Song 1: Crystal Castles - Vanished

Start of Block: Dance, Song 2: California Swag District - Teach Me How To Dougie

Q134 California Swag District - Teach Me How To Dougie

Are you familiar with this song?

- Yes (1)
- No (2)

Q135 Are you familiar with this artist/band?

- Yes (1)
- No (2)

Q136 How likely will you be to listen to this song?

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Q137 Does this song fit within your musical taste?

- Not well at all (1)
- Slightly well (2)
- Moderately well (3)
- Very well (4)
- Extremely well (5)

Q151 Would you add this song to your playlist?

- Yes (5)
- No (6)

End of Block: Dance, Song 2: California Swag District - Teach Me How To Dougie

---

Start of Block: Dance, Song 3: The Black Eyed Peas – I Gotta Feeling

Q138 The Black Eyed Peas – I Gotta Feeling

Are you familiar with this song?

- Yes (1)
- No (2)

Q139 Are you familiar with this artist/band?

- Yes (1)
- No (2)

Q140 How likely will you be to listen to this song?

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Q141 Does this song fit within your musical taste?

- Not well at all (1)
- Slightly well (2)
- Moderately well (3)
- Very well (4)
- Extremely well (5)

Q152 Would you add this song to your playlist?

- Yes (5)
- No (6)

End of Block: Dance, Song 3: The Black Eyed Peas – I Gotta Feeling

Start of Block: Dance, Song 4: Crookers ft. Kid Cudi - Day 'N' Nite

Q142
Crookers ft. Kid Cudi - Day 'N' Nite

Are you familiar with this song?

- Yes (1)
- No (2)

Q143 Are you familiar with this artist/band?

- Yes (1)
- No (2)

Q144 How likely will you be to listen to this song?

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Q145 Does this song fit within your musical taste?

- Not well at all (1)
- Slightly well (2)
- Moderately well (3)
- Very well (4)
- Extremely well (5)

Q153 Would you add this song to your playlist?

- Yes (5)
- No (6)

End of Block: Dance, Song 4: Crookers ft. Kid Cudi - Day 'N' Nite

---

Start of Block: Dance, Song 5: Justin Timberlake - LoveStoned/I Think She Knows

Q146 Justin Timberlake - LoveStoned/I Think She Knows

Are you familiar with this song?

- Yes (1)
- No (2)

Q147 Are you familiar with this artist/band?

- Yes (1)
- No (2)

Q148 How likely will you be to listen to this song?

- Extremely unlikely (1)
- Moderately unlikely (2)
- Slightly unlikely (3)
- Neither likely nor unlikely (4)
- Slightly likely (5)
- Moderately likely (6)
- Extremely likely (7)

Q149 Does this song fit within your musical taste?

- Not well at all (1)
- Slightly well (2)
- Moderately well (3)
- Very well (4)
- Extremely well (5)

Q154 Would you add this song to your playlist?

- Yes (5)
- No (6)

End of Block: Dance, Song 5: Justin Timberlake - LoveStoned/I Think She Knows

Start of Block: Dance, final questions

Q150 Out of the five previous songs, which one is your favourite?

- Crystal Castles - Vanished (1)
- California Swag District - Teach Me How To Dougie (2)
- The Black Eyed Peas – I Gotta Feeling (3)
- Crookers ft. Kid Cudi - Day 'N' Nite (4)
- Justin Timberlake - LoveStoned/I Think She Knows (5)

Q124 As a consumer of music streaming platforms, please indicate to what extent you disagree or agree with the following statements:

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
I am satisfied with the recommendations music streaming platforms provide me. (1)	○	○	○	○	○	○	○
I would prefer to get recommendations based on the music taste of my peers. (2)	○	○	○	○	○	○	○
I would prefer to get music recommendations based on the expertise of curators. (3)	○	○	○	○	○	○	○
I would prefer to get a mix of music recommendations based on listening habits from my peers and curated content. (4)	○	○	○	○	○	○	○

Q127 Did you search for any additional information about the artists, cover art and/or songs you came across while answering the survey?

- ○ Yes (1)
- ○ No (2)

End of Block: Dance, final questions

Appendix B. CF Code

The hashtag (#) symbol and grey text denote the description of what is entered or needed to enter in the code, and are not included during the coding process.

Black text is the prewritten code.

Green text is the data that needs to be adjusted for the CF process, such as the songs.

Input:

```
# -*- coding: utf-8 -*-

# imports

import pandas as pd

import numpy as np

from sklearn.metrics.pairwise import cosine_similarity


# read data into a DataFrame

song_data = pd.read_csv('song_data.csv')

song_data.head()


# let's limit things to the top 250 songs (put 5000)

n = 5000

top_n = song_data.song.value_counts().index[:n]

song_data_top = song_data[song_data.song.isin(top_n)]

print(song_data_top.shape)


# collect the top 250 users from this top 250 songs data to be used
as test users in the sql data
```

```
top_users_n = song_data_top.user_id.value_counts().index[:n]
user_data_top =
song_data_top[song_data_top.user_id.isin(top_users_n)]
print(user_data_top.shape)


user_data_top.to_csv('user_data_top.csv')


print("melting...")
song_wide = pd.pivot_table(song_data_top, values=["listen_count"],
                   index=["song", "user_id"],
                   aggfunc=np.sum).unstack()


# any cells that are missing data (i.e. a user didn't listen to a
song)
# we're going to set to 0
song_wide = song_wide.fillna(0)
# this is the key. we're going to use cosine similarity from scikit-
learn
# to compute the distance between all songs
print("calculating similarity")
dists = cosine_similarity(song_wide)


# stuff the distance matrix into a dataframe so it's easier to
operate on
dists = pd.DataFrame(dists, columns=song_wide.index)
```

```
# give the indicies (equivalent to rownames in R) the name of the product id
dists.index = dists.columns

dists.to_pickle('song_similarity.pkl')

a = np.zeros(shape=(1,len(dists)))
user_df = pd.DataFrame(a,columns = dists.index)

user_data = {'Snow [Hey Oh] (Album Version) - Red Hot Chili Peppers': 0,
 'Alejandro - Lady GaGa':1,
'Bleeding Love - Leona Lewis': 0,
'Homecoming - Kanye West': 0,
 "Pon De Floor - Major Lazer / Vybz Kartel / Afrojack": 0}
for key in user_data.keys():
    user_df.loc[0,key] = user_data[key]

single_user_matrix_multiply = user_df.dot(dists)

single_user_matrix_transpose = single_user_matrix_multiply.transpose()
song_reco = single_user_matrix_transpose[0].sort_values(ascending=False)
```

```
song_reco_10 =
song_reco.index[song_reco.index.isin(user_data.keys())==False][:10]


print(song_reco_10.values)
```

Output:

```
runfile('C:/Users/XXX/XXX/XXX/col_filtering.py',
wdir='C:/Users/XXX/XXX/XXX)

(879756, 6)

(174779, 6)

melting...

calculating similarity

['Love Story - Taylor Swift'

"Reelin' In The Years - Steely Dan"

'Catch You Baby (Steve Pitron & Max Sanna Radio Edit) - Lonnie
Gordon'

'Boys Boys Boys - Lady GaGa'

'Crescendolls - Daft Punk'

'A Thousand Miles - Vanessa Carlton'

'Monster - Lady GaGa'

"Just Dance - Lady GaGa / Colby O'Donis"

'Sehr kosmisch - Harmonia'

'Whataya Want From Me - Adam Lambert']
```

Appendix C. Songs present in survey

Pop

	Artist & Title	Release year	Filter type
Start	Lady Gaga - Alejandro	2009	X
1	Coldplay ft. Rihanna – Princess of China	2011	CBF
2	Timbaland ft. OneRepublic – Apologize	2007	CBF
3	Taylor Swift – Love Story	2008	CF
4	Adam Lambert - Whataya Want From Me	2009	CF
5	Britney Spears - 3	2009	Random

Rock

	Artist & Title	Release year	Filter type
Start	Red Hot Chili Peppers – Snow (Hey Oh)	2006	X
1	Kings of Leon – Use Somebody	2008	CBF
2	Muse – Uprising	2009	CBF
3	Rise Against – Behind Closed Doors	2006	CF
4	Guns N' Roses - Sympathy For The Devil	1994	CF
5	The Black Keys – Tighten Up	2010	Random

Hip Hop

	Artist & Title	Release year	Filter type
Start	Kanye West ft. Chris Martin – Homecoming	2007	X
1	Soulja Boy – Crank That (Soulja Boy)	2007	CBF
2	Drake ft. Lil Wayne & Young Jeezy – I'm Goin In	2009	CBF
3	DJ Khaled ft. T-Pain, Ludacris, Snoop Dogg & Rick Ross - All I Do Is Win	2010	CF
4	Lupe Fiasco ft. Matthew Santos – Superstar	2007	CF
5	Young Money ft. Lloyd – BedRock	2009	Random

R&B

	Artist & Title	Release year	Filter type
Start	Leona Lewis – Bleeding Love	2008	X
1	D'Angelo - Untitled (How Does It Feel)	2006	CBF
2	Beyoncé – Sweet Dreams	2008	CBF
3	Rihanna – Te Amo	2009	CF
4	Justin Bieber – One Time	2009	CF
5	Keyshia Cole – Fallin' Out	2007	Random

Dance

	Artist & Title	Release year	Filter type
Start	Major Lazor ft. Vybz Kartel & Afrojack – Pon De Floor	2009	X
1	The Black Eyed Peas – I Gotta Feeling	2009	CBF
2	Crookers ft. Kid Cudi - Day 'N' Nite	2009	CBF
3	Crystal Castles – Vanished	2008	CF
4	California Swag District - Teach Me How To Dougie	2011	CF
5	Justin Timberlake - LoveStoned/I Think She Knows	2006	Random