Genre compatibility in consumer profiles

Mapping listening behaviour patterns to identify genre compatibility in the profiles of the Last fm users.

Student name: Marloes de Vries
Student number: 463284

Supervisor: Koen van Eijck
Second reader: Frank Weij

Erasmus School of History, Culture and Communication
Master Arts, Culture and Society
Erasmus University Rotterdam
Master of Arts Thesis
14-06-2019

Genre compatibility in consumer profiles
MA Thesis, Marloes de Vries (463284)

Abstract

In a world where every type of music is available from every online medium, listeners can be overwhelmed by the amount of content. Several online platforms and music providers offer a recommendation system and a personal account for their users, acting as digital cultural intermediaries by recommending songs while mapping the users' listening behaviour. One of these online platforms is Last fm, a website that connects different music streaming services into one user account, in which they recommend music and analyse the listening behaviour for an improved experience for the user, and provide live statistics on listening behaviour. This vast amount of music listening behaviour data of Last fm is anonymized into a dataset made available for non-commercial research. In this thesis, the Last fm database is used to map consumption patterns and find genre compatibility using cluster analysis and linear regression models. The cluster membership is then linked to the user data consisting of age, gender, country, listening intensity and diversity score. Five clusters of Listening behavioural patterns are found and are compared. The way different genres correlate with each other and with diversity gives insight into the genres compatibility and patterns.

Keywords:
music listening behaviour, audience studies, consumption patterns, music genre compatibility

Acknowledgements

A master thesis, seeing it as the largest paper I’ve ever written is exciting. Realising it’s the last and final assignment of my 6 yearlong study period, is nerving. Thinking it’s the end of my school life phase that started when I was four, now that's quite scary. But I have done it, and my MA thesis lays here before you now.

First I’d like to thank dr. Koen van Eijck for his excellent guidance and his valuable and constructive suggestions for this thesis research and for his lectures throughout the studies that sparked many interesting discussions and a lasting interest in sociology. Second to the department MMC of the TU Delft for providing me with the last fm data and specifically to dr. Cynthia Liem, Andrew Demetriou and Jaehun Kim for their guidance and motivation.

Thirdly, thank you to all teachers, friends and classmates that had faith in me. Thank you as well to all the people that did not, for giving me the possibility to prove you wrong.

Thank you to my parents and brother for motivating my ambitions and supporting me through every setback, and to Jantien and Mirre who provided me with valuable feedback on my thesis and for their emotional support throughout life. And to Bart, for his feedback, support and existence a great thank you.

And finally to the reader of this thesis thank you for your time and concentration, I hope you will find it time well spent.

Index

1. Introduction

Whether it is dance music from the 50s, or an upcoming vocal quartet that you are listening to, you have a music taste. All the music you choose to listen to together creates a music listening profile. You might be a frequent listener to both metal and R&B, but not many people are. Which genres are compatible with each other in listening behaviour profiles has been a subject that has popped up in several sociological theories. Whether taste differences are explained in terms of upper and lower class, life stage or subculture, most theorists draw from sociological theorizing in explaining musical tastes, typically finding that different genres cluster into taste patterns.

Most of these studies have been done using surveys. Such surveys on taste have several limitations. The genres used to assess 'taste' may not correspond to which music is actually listened to. The matter of genrefication as addressed by Lena (2016) is often problematic; do we file Madonna under pop, funk or dance? In survey studies, different ideas on genre content and characteristics may exist among the researcher, the respondent and the interpreters. Hence the relevance of doing research that involves a minimum level of personal interpretation and a wider scope of both respondents and actual songs than most surveys ever reach. The quantitative data available for this research allow us to analyze what people actually listen to and not just look at what they say they like.

This research undertakes a journey into the listening behaviour of users of online music streaming services in which Madonna is categorized under the same genre for each and every user, providing less genrefication biased data due to the less ambiguous genre label per song.

The aim of this research is to gain better insight into the nature of music listening patterns and their corresponding audiences, and to map listening behaviour patterns of consumers to identify genre compatibility in the consumer profiles of the Last fm database. This leads to the research question of this thesis: "What listening behaviour patterns can be found and how do these patterns correlate with listening diversity and intensity in the music listening behaviour profiles of Last fm?". This main research question is separated into three sub questions:

1. Which listening behavioural patterns can be found within the Last fm database considering genre compatibility?
2. How do the users differ between the clustered listening profiles?
3. How do specific genre preferences relate to diversity and intensity of music listening?

These sub questions will be discussed separately in the results section of this thesis. First, the studies that have been done previously on taste patterns, music consumption and genre compatibility will be addressed and compared within the greater sociological context in the theory chapter, chapter 2. For this research, several analysis methods were used on a sample of the Last fm dataset, all of which is described in chapter 3, the methods chapter, which is followed by the results in chapter 4, describing the findings and comparing these to previous studies. Finally, the conclusion and discussion are presented in chapter 5, which will also include the research limitations and future prospects

2. Theory

Taste patterns

The cultural taste of individuals and social groups has been of interest to sociologists for a long time. Peterson and Bourdieu both provided the world with insights combining taste content with taste diversity.

Bourdieu (1973;1984) found, in 1960's Paris, that social class was closely related to cultural taste. The upper class consumed more difficult elite types of culture that Bourdieu would call 'highbrow'. Within the segregation of society he also distinguished 'middlebrow' and 'lowbrow' art taste patterns. According to Bourdieu each individual had three types of capital that together formed the habitus of a person. The three types of capital are, economic capital, social capital and cultural capital. In short, economic capital is the amount of money a person has, social capital is one's network and social acceptability and cultural capital is the knowledge of and interest in and visiting of cultural scenes. These capitals together form the habitus. The habitus decides in which class of society a person is situated. It also provides the person with, so called, symbolic capital that allows a person to connect himself in different social fields and to certain groups, having the sufficient capital to manage in a certain social field. In turn, the habitae of the people are influenced and shaped within these social fields. To elaborate, the development of children happens within these fields and families, building the child's habitus by developing the three types of capital from an early age. Class segregation is reproduced within the social fields by excluding people with different amounts of capital and by developing and influencing the habitae of the people in the fields, thus maintaining and reproducing class segregation. A 'highbrow' taste in culture and visits of often expensive cultural venues show a high amount of cultural capital within the social field resulting in an upper class position. The large amount of money and social connections needed to get into these venues and circles keeps the class separation intact. The line between highbrow, middlebrow and lowbrow cultural outings is still quite distinct. classical music, ballet, opera and art museums were considered 'highbrow'. Local music, jazz, blues etc. were considered ' lowbrow' and more entertainment than art.

Peterson (1992) recognised a cultural taste pattern which is characterized by a wide orientation in different cultural genres, interests and knowledge. The omnivorousness taste pattern developed to be the new indicator of upper class cultural consumption. The cultural omnivore is the opposite of the cultural univore, who has a very singular taste pattern. Univores taste patterns are not part of upper class society, but are connected to the lower social status. His research was later complemented and substantiated by many sociologists, making it an important new theory on cultural consumption and distinction (Hazir, 2015; Ferrant, 2018; Van den Haak, 2014). The omnivore and univore distinction and its link to societies' segregation was revisited by several researchers who connected it back to Bourdieu (1973;1984). Van Eijck and Lievens (2008) used a well-known distinction between highbrow, folk and pop music, and aimed their research on finding the different types of omnivores differently combining the three cultural schemes. The tripartite division had been used and found fruitful before by Frith (1990,1987), Peterson and Simkus (1992) and Schulze (1992). The segregation of society upheld by music taste patterns exclusivity was also recognized in Michael (2017) research on high potentials' highbrow taste. where her interviewees considered some art forms higher and more difficult to understand or appreciate which for them indicated a higher social position. Ollivier (2008) compared several typologies and studies concerning omnivorism and class differentiation. She concluded that many researchers found the same patterns when comparing levels of education and differences in age. These sociodemographic characteristics linked to specific taste patterns in multiple researches that she compared. She noted that younger people where more open to diversity than the older generation.

On gender differences and their effect on listening profiles and openness, Ollivier found that the aspect of gender was underrepresented in the studies. There were very few indicators that female consumers are over represented in the exclusive highbrow group, and underrepresented in the omnivore highbrow taste category, which may be due to the increase in feminisation of the highbrow cultural field according to Donnat (2005)(Corrêa et al., 2016)).

Both taste diversity, taste patterns content and sociodemographics and group indicators are found and used for gaining understanding and to map cultural consumption differentiations by sociologists. Besides the vertical class segregation, music consumption and cultural participation are also horizontally divided. The more singular taste patterns, such as the by Peterson described univores, distinguish themselves from other univore or singular taste patterns, creating different music scenes and groups.

Music scenes

Within the vast amount of music items there are several music scenes existing of music styles and genres. The participation in and consumption of a specific music scene is often seen as a marker of identity, and can indicate a membership of a certain group or subculture. People with the same taste pattern or music scene attract each other and form groups, and within these groups taste patterns and cultural ideas are exchanged and normalized within the groups' identity. When a new member wants to be part of the group, with its subculture, they adjust their taste pattern and behaviour to fit the subcultures/group identity which is often created by cultural outings and taste patterns, often distinguishable by music taste profile/ music genre (Bennet & Peterson, 2004). These subcultures' taste patterns can be linked to the layers in society, where economic status, social capital and education level are often connected to these group identities.

Young people in particular often wear their taste pattern as a badge of identity, using their clothes, images such as buttons and band t-shirts to indicate their belonging to a subculture or identification pattern (Frith, 1987). Michael (2017) found in her research on high potentials vision on and consumption of highbrow culture that they connect specific music tastes to certain prejudices and images of subcultures or class positions that they do not want to be associated with. The taste patterns that subcultures represent are quite singular because these subcultures often distance themselves from the masses, and prefer to remain recognisable. The respondents in Michael's research publicly distanced themselves from several types of music for that reason as they did not want to be associated with the subculture and social image.

Digitalisation of consumption

The digital revolution has rapidly made culture, and specifically music, more widely available through a fast development of cd's, mp3 files and, more recently, digital music streaming. There is by far more music available online than one can listen to in a lifetime, even if you would play music 24 hours a day (Nell, 2014).

Location and time are no longer limitations for music access, and neither are expensive entrance fees. Money, economic capital, or knowing the right people to get into a venue or social capital are no longer needed to acquire a certain taste pattern or to consume certain types of music, as everything is just one click away on the world wide web (Nell, 2014). This digitalisation of culture consumption has led to a situation that Webster (2014) named 'the marketplace of attention'. In the marketplace of attention a vast amount of providers and platforms try to reach and maintain the largest number of worldwide users and a competition of screams for attention has developed.

Digital music streaming services such as Spotify and Deezer self-proclaim to be an easy way to discover, manage and share music (Aguiar, 2017). Datta et al. (2018) found that the use of these streaming services for digital music consumption indeed do add to the discovery of more music, and lead to a more diverse music listening behaviour and provide wider orientation possibilities in the music industry for their users. Especially within the first month of use of the Deezer service the diversity increases rapidly, and although this effect decreases after that first month, the diversity and wider orientation is maintained and further broadened after half a year of use. Datta et al. provide the explanation that when music had to be bought, diversifying and trying something new was an expensive endeavour, while in the digital musical omnipresence experimentation with and broadening of music consumption and music taste is easier and accessible.

Through the digitalisation of music consumption, a vast amount of data on both artists, genres, and listeners has become widely available. Several services and companies in the commercial sector use data analysis methods to group and profile users, needing to predict the users' behaviour as closely as possible to make their service valuable enough for the user to pay for (Prey, 2016). These streaming services are the future of music distribution and consumption in the post-download era, where music is no longer owned but merely provided to only consume (IFPI 2015 in Vlegels and Lievens, 2017).

Prediction and recommendation

What used to be a question of persuasion, has become a problem of prediction', (Harvey, 2014).To get the internet users the content that they might find interesting, the algorithm called collaborative filtering selects content for them from the large number of possibilities based on their previous consumer behaviour. Their previous behaviour is compared to other users, and these users are grouped together into collaborations and every user from that collaboration contributes to the individual recommendation based on the group's preferences (Sánchez-Moreno et al., 2016 ;Webster, 2014). This digital recommendation algorithm can be considered a cultural intermediary if we consider the definition by Wright (2005: 118): *'Intermediaries work between the production and consumption of cultural goods'*. This algorithm in the function of a cultural intermediary has two major setbacks considering cultural consumption.

First, there is the phenomenon of the filter bubble, where the users in a collaboration look so much alike and do not create new input. This narrows their consumption behaviour into a certain singular direction and the user gets trapped in a so called 'bubble' by the automated recommendation system. This can be a cramped and even lead to dangerous situations in the case of political opinion and information bubbles (Rowland, 2011). Within the music field this filtering bubble also occurs, especially in less advanced versions of the collaborative filtering model (idem).

The other phenomenon is the long tail effect. In the collaborative filtering mechanism the already popular items are recommended more often which further enlarges their popularity, marginalizing other content, creating a feedback loop that leads to a relatively small number of items which are consumed often, and an extremely large number items which are rarely or even never consumed (Kagie et al., 2008) . In the recommendation circle of popular items, no new items ever make it into the filtering mechanism. The exception to this effect are the very few items that go ‘viral’. Those that do not end up in the never recommended long tail. This long tail effect leads to a shallowing cultural consumption and more easily developed filter bubbles (Rowland, 2011; Jaeger and Katz-Gerro, 2008).

Many streaming services aim to help the users explore, Spotify’s system results in users getting in contact with 27 new artists a month, by the way of popular users creating playlists and mixing new artists in the daily mixes of users. This may be diversitizing, but the danger of financial motivation to advertise and recommend certain artists or albums more than others hovers. Since Spotify pays the music companies per played song on the service, giving a financial incentive to influential users to tag and boost a specific artist could turn out beneficially for the investor and can influence the collaborative recommendation system to flatten and popularize taste and create a beneficial filter bubble (Kissel, 2015).

Researchers and developers such as Liem et al. (2011) and Lu & Tseng (2008) are developing and researching a recommendation system of online platforms and developing a filtering technique called automated tagging. Certain tags belonging to songs are recommended in the system. A new song with similar tags can therefore appear in the recommendations and also different musical outings with several comparable tags might be recommended, leading to a broadening of consumer behaviour and taste. For example, someone who listens to a lot of symphony orchestra’s and upbeat cello concertos might be recommended a song by Ali B made in cooperation with the Rotterdam Symphonic Orchestra. Since many of the tags on both songs are similar and a few are different, through this crossover music the listener may become more open to rap songs, Dutch poetic rhyme music or spoken word. This would lead to more diverse listening behaviour, and would bring the digital recommendation system more closely to the function of a cultural intermediary.

What's in a Genre?

In many studies music is categorized by genre. For example a study might use a survey question listing about 10-15 genres asking the respondent to note on a 1-5 Likert-scale how fond they are of that music genre. But what genre definitions do the respondents use to fill in these surveys, and how do the researchers interpret these genre listings as representation of taste? In this type of studies, where three people all listening to the same music and sharing an identical taste pattern may all fill in their taste under different genres, a genre content bias occurs, called genrefication by Lena (2016).
Questions on music taste in surveys especially using a predefined list of music genres create fewer distinctions and depth of knowledge than the consumers and actual users of culture do (Dimaggio, 1987).

Last fm

Several commercial music streaming services such as Spotify, Apple-music and Deezer, use data analysis to improve their market position and to recommend their users the music they are most likely to like and listen to. Last fm is an online service that took this user tracking to a higher level, allowing their users to scrobble (collect/ add to their account) all music consumption data from other services and their offline collection as well to create a more personal recommendation algorithm, and to collect and review their music consumption collection/ pattern in one place. Premium members can see a lot of live statistics and compare their user profiles to those of others, when they are truly interested can see the level of mainstreamness of their listening profile, which is an indicator given by the website to give the user insight in how much his account data overlaps with the main group of users (Schedl et al., 2013).

Within the social sciences, this large amount of data on listening behaviour instead of proclaimed taste preference has created many new opportunities for quantitative research on music listening behaviour and cultural consumption. Demetriou et al. (in press) took a closer look into the Last fm user profiles and found that comparing it to the more well-known MGM database which had a limited amount of information on genre spread and listening amount per user, the Last fm's provided detailed track record of each user . In this very new possibility of large scale data analysis within the social sciences, this thesis research will contribute to the exploration of the newly developed field and aims to combine user listening profiles clustering, analysis of genre compatibility and socio demographics within these data-driven behavioural patterns to gain more insight into listening profiles, and genre listening correlations to diversity in behaviour.

3. Methods

3.1 Type of method

To gain more insight into the listening behaviour of a population, a quantitative research method is the best fit. This quantitative statistical research in the Last fm database will provide insight into the users' listening behaviour. Compared to the more researched music preferences and tastes assessed in population surveys, this comprehensive data analysis will provide a more detailed and accurate view of listening behaviour than a survey method. It does so by providing actual listening behaviour instead of self-reports, and including more detailed knowledge spread over more genres than can be included or gotten from survey research. Finally, the listening behaviour database of Last fm shows listening behaviour of users without dual interpretation except for the genre classification made by the platform and the inevitable interpretation of the researcher of the data findings. What makes this data analysis more reliable is the fact that within the service provider the songs are listed on a genre tag, which leads to an elimination of genre interpretation by the listeners.

3.2 Measuring instrument

The Last fm platform

The Last fm online music streaming service was one of the first online music recommendation and collection platforms. Where a user could scrobble a data collection from offline digital data and services of later included the scrobbling of digital providers for example iTunes or Spotify. Last fm offers their users a personalised profile and the possibility to collect their music listening data on different devices and within different streaming services, to enlarge their personal listening profile, allowing for personal recommendations and listening advice. It allows users to have power over their own account, by scrobbling music they like. Another feature many users like is their live charts and statistics, showing most listened to tracks/ albums of upcoming artists or newest least listened to artists. Last fm promotes itself to music lovers, and people who actively want to take part in the growth of their profile and portray themselves as intermediary within music consumption/ taste development. Around last fm there is quite an active user fanbase communicating on forums and social media, and several (homemade) developed apps and websites that visualize Last fm listening history (Reddit, n.d.).

The last fm database

The Last fm database includes the number of songs listened to within a specific genre and the users' age, country and gender are represented in the dataset as well. Next to that a listening diversity score is calculated, representing the spread of the listening intensity over different genres.

For this thesis a cooperation with the TU Delft, the multimedia and computing group was established via assistant professor Liem. Dr.Liem arranged access to the Last fm dataset and Jaehun Kim MSc and Andrew Demetriou MA, two PhD candidates who both work with the dataset as well, provided guidance and additional knowledge and explanations where needed. They cleaned the original Last fm dataset to the raw version that was used for this thesis research. They took out two genres that were very rarely listened to, namely children's music and spoken word. Kim made several adaptations of the dataset that all represented a different possibility of research (Demetriou et al., in press).

The preliminary findings, data assimilation test of this research was done with the dataset adaptation which added a scale of 0 to 7 to each genre, that might have been comparable to the Likert scores and scale variables of many surveys, yet the calculation in the dataset to get to the scale variable was based on the amount of songs listened to in that genre, which concluded in fact that intense users of the service had 7’s on all genres, because they had reached the threshold to ‘score’ a seven on those genres even though it was their least listened to genre. the ‘the user listened to this genre a lot’ was undefined and had no connection to the distribution of listening counts over the genres.

For this thesis research the original data unit was used, namely the total songs listened to per genre. The total amount of songs listened to per user can be found in the intensity variable, to get a good comparable genre listening profile per user, the percentages of the intensity of listening per genre was calculated for each genre per user, leading to a list of new variable, namely the percentage of intensity that is filled by each specific genre. The by Kim added diversity variable is used as well, which represents the categorical distribution of listening counts over the different genres. For a more detailed explanation see operationalisation.

Several adaptations of the database are available. All having a translation step from the raw data. All adaptations will be shortly described and compared below.
The version of the dataset called Likert scale, where the users behaviour has been brought back to a more comparable scale score of 1-7. However, this data translation does not include an absolute frequency and is limited in the sense that a ‘super user’ scores high on all genres because the scale is set as the amount of times songs from a genre are listened to. Data set 2 is more suitable for portraying genre diversity and the spread of the listeners intensity across the different genres. The data interpretation of version2 is based on the proportions of the listening intensity and its distribution over the different genres. which is very similar to version3 were the percentages of distribution of listening behaviour are represented. In the final analysis the diversity indicator is compared to the genres percentages to see how diverse specific genre listeners are.

The Last fm users

The 1020.000 users of Last fm that are included in the Last fm database are on average 25,4 years old, with a median of 23 and the 25th and 75th percentile at respectively 20 and 28 years old. The distribution of the age is skewed to the older ages.

The service is worldwide available, yet the countries most represented are the US, Russia, Germany the UK, Poland and Brazil. Only 38,31% of all users provided their age information. Half of the users provided their gender information out of which 28% is female. The most listened to genres are rock (18:27%), alternative (16:75%), and pop (13:64%)(Schedl, 2016).

3.3 Operationalisation

The music classification into genres

The Last fm data frame collects every song played by each user, these loose songs have related tags which combine them in groups, or sub genres. This data set has been gathered by the department of multimedia and computing of the TU where the sub genres were grouped into wider genres which made possible research outcomes more comparable to other researches. They used the genre grouping defined by Schedl and Ferwerda (2017). Jaehun Kim cleaned the dataset, and calculated the diversity score for the user profiles, to be able to indicate their listening diversity in the researches on the dataset that had yet to come. The different genres in the for this thesis used dataset are: R&B, rap, electronic, rock, new age, classical, reggae, blues, country, world, folk, easy listening, jazz, vocal, punk, alternative, pop, heavy metal

Intensity and diversity

Each user of the Last fm service listens to songs, all songs listened to together count towards a grand total, which represents the intensity of that user account. The intensity variable literally represents the amount of audio files listened to.

The variable diversity shows the listening diversity by calculating the flatness of the categorical distribution of listening per genre. To calculate the diversity variable Jaehun Kim assumed each genre as an independent taste/genre. The distribution of the total listening counts over the different genres was set into a categorical distribution, which have a certain flatness or peak level. This flatness or spread of listening distribution is the diversity index. The variable is not normalized and does not have a set range for its size or largeness. The relative distance can be used to compare listeners and listeners profiles to each other. The range in the used data set sample is (lowest-highest) = 2.33 - 16.28. The use of the entropy based diversity calculation is suitable for music listening diversity to calculate the flatness of the categorical distribution. The main limitation of this variable's calculation it the assumption that the values per genre are independent (Demetriou et al, in press).

Percentages per genre variable

The dataset included total listening counts per user per genre, which are hard to compare since the absolute frequency is unknown. Therefore the total listening counts were calculated to percentages by seeing how much percent of the listening intensity was spend in a certain genre. With this division of intensity over genres variable, the user accounts are more directly comparable in how they spend their listening time. both the cluster analysis and correlation matrices use this percentage per genre variable.

3.4 Sampling

Sample within the data users

In this research on listening behaviour, the Last fm user dataset is used for the research, which funnels the population of music listeners to those that use the Last fm service and are included in their dataframe. Within the data set, 120,173 user profiles are included. When mapping these users in a histogram considering their intensity score of listening, it is seen that several of them are extreme users (figure 1). Looking at where and by whom the Last fm service is used, several of the mass user accounts can be linked to cafes or companies that continuously play music to a larger group of people.

This research aimed at single user profiles. Including the superusers would not help us understand individual profiles and lead them to disproportionately influence the findings. Therefore, the super users are excluded from this research, to create the possibility for nuance and omit institutional / commercial users. Another large influence on the findings of the research is the large amount of accounts that do not have a significant listening intensity. Some accounts have listened to only eight songs in total, all from one genre, which would lead to a lot of univores found, while that very low intensity account is not a good representation of an users listening profile.

To gain the most insight in listening profiling the group of users with an intensity between 20,000 and 60,000 is selected (figure 2), within this sample the diversity and formation of their diversity and listening behaviour will be further looked into. After this selection, 30,520 user profiles from the database are included in the sample. When selecting this sample, I found that a multiple accounts without a gender/ age were filtered out. This seems to confirm the assumption that those might have belonged to either institutes or groups or to inactivated accounts rather than active personal users.

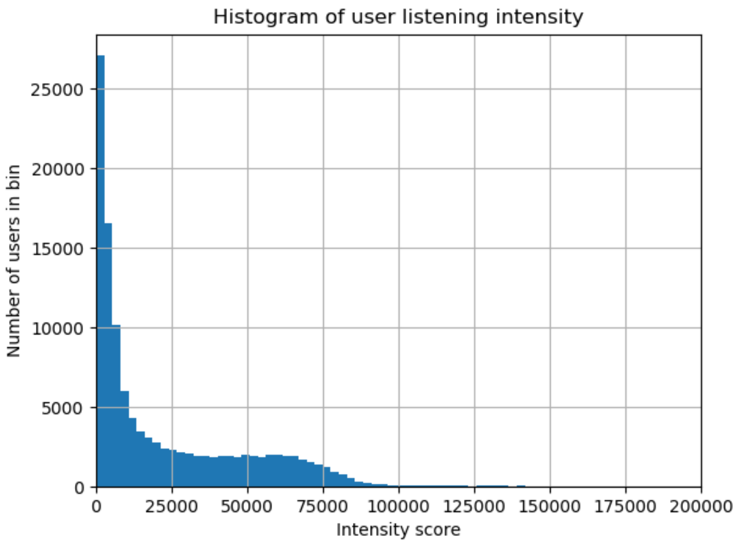

Figure 1 Histogram of user listening intensity

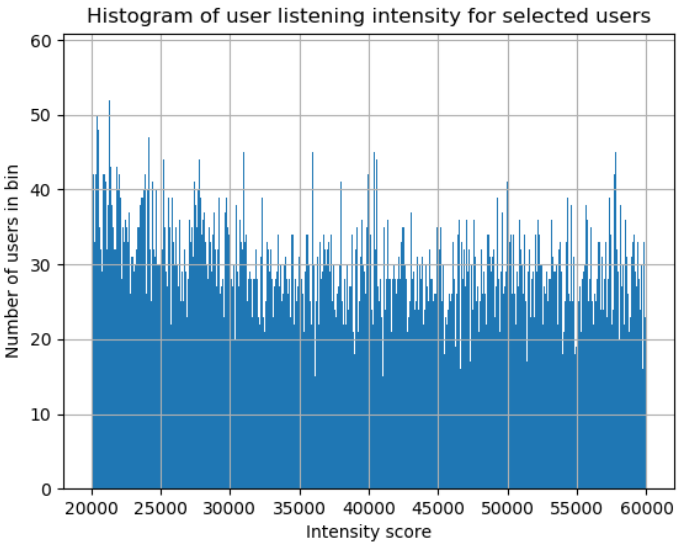

Figure 2 Histogram of user listening intensity for selected users

3.5 Methods

Clustering (k-means)

The k-means clustering method groups data input (either respondents or variables) into groups. For this study, genres were clustered in order to find sets of related genres representing musical listening patterns. The mean values of these variables are then represented per cluster providing a segmentation of the database based on these values. In this research the k-means cluster analysis of genres is used to and users are linked to one of the clusters depending on their listening frequency of (sets of) specific genres. Providing information to answer; What listening behavioural patterns can be found within the last fm database considering genre compatibility?

Compare means

A compare means table compares the means of several variables, in this research the means of age, diversity, intensity and gender are compared between the clusters membership. To include gender in this comparison the variable of gender (m/f/n) has been converted into numeral variables, m=1 f=2 n=3, where the neutral has been excluded from the comparison leaving the means of male and female values which can then be easily set to percentages. For example: a mean gender score of 1.35 indicates that 35% of the members of that cluster are female.

Multinomial logistic regression analysis

A multinomial regression analysis shows the linear regression within multi-class comparisons, appointing one of the clusters as the reference category. The variables age, gender, intensity and diversity score are used to describe and predict cluster membership. This comparison and regression analysis using these variables gives us more insight into the composition of the different clusters. Giving insight into the question: How do the users differ between the clustered listening profiles?

Compare means

A compare means table compares the means of several variables, in this research the means of age, diversity, intensity and gender are compared between the clusters membership. To include gender in this comparison the variable of gender (m/f/n) has been converted into numeral variables, m=1 f=2 n=3, where the neutral has been excluded from the comparison leaving the means of male and female values which can then be easily set to percentages. For example: a mean gender score of 1.35 indicates that 35% of the members of that cluster are female.

Bivariate Correlation

A bivariate correlation measures the linear relationship between to variables, in this research the relationship between the listening to the different genres and the diversity score will be calculated, providing a good estimate on which genres relate to a wider or more shallow listening diversity. Illustrating how specific genre preferences relate to diversity and intensity of music listening.

4. Results

4.1 Clusters

To find behavioural patterns of music listening behaviour in the database, considering genre compatibility and combination. A cluster analysis is done to group the behaviour into user clusters based on their listening behaviour. After running the K-means cluster analysis, an optimum of five clusters was found which showed a clear 5th cluster compared to the four cluster calculation, while the step towards six clusters watered out the results interpretability greatly. The five cluster option is the chosen optimum which is described below.

In the cluster comparison table, the percentages listened to by cluster members for different genres are represented (table 1). Although the weight in the calculation is heavier for some genres than for others, this had developed since the k-means analysis strives to find the greatest stable difference between the clusters. A visual representation of table 1 is added in figure 1 which clarifies the differences between the percentages and the major genres and genre differences between the clusters. Following the 5 listening pattern clusters will be described. In part 4.2 the cluster members and their socio demographics will be described further.

Table 1: Cluster division based on genre percentage

	Cluster				
	1	2	3	4	5
R&B	.03	.10	.00	.01	.02
Rap	.02	.14	.01	.02	.03
Electronic	.10	.13	.06	.09	.26
Rock	.18	.13	.29	.23	.15
Newage	.01	.00	.01	.00	.03
Classical	.02	.01	.01	.01	.03
Reggae	.01	.02	.00	.01	.01
Blues	.04	.02	.02	.02	.01
Country	.02	.01	.01	.01	.00
World	.02	.01	.01	.01	.02
Folk	.08	.03	.05	.04	.04
Easy Listening	.03	.02	.01	.01	.02
Jazz	.05	.05	.02	.02	.04
Vocal	.01	.01	.00	.00	.01
Punk	.05	.04	.07	.14	.04
Alternative	.17	.15	.15	.21	.17
Pop	.16	.14	.05	.10	.10
Heavy Metal	.01	.01	.23	.06	.02
Nr. cluster members	12688	2415	3700	7583	4134

In the table above, the 5 clusters and their proportion of listeners per genre is given. The cells representing the cluster with the most frequent listeners to each genre have been made green. The orange cells are shared first places or large values. The yellow fields are other remarkably high scores. This colouring gives a more visual display of the division of listening behaviour within the clusters and which genres are linked within the cluster suggesting listening compatibility. The bar graph below in figure 3 shows the cluster divisions representing the genre percentages in the height of the bars, showing the division of the genres per cluster.

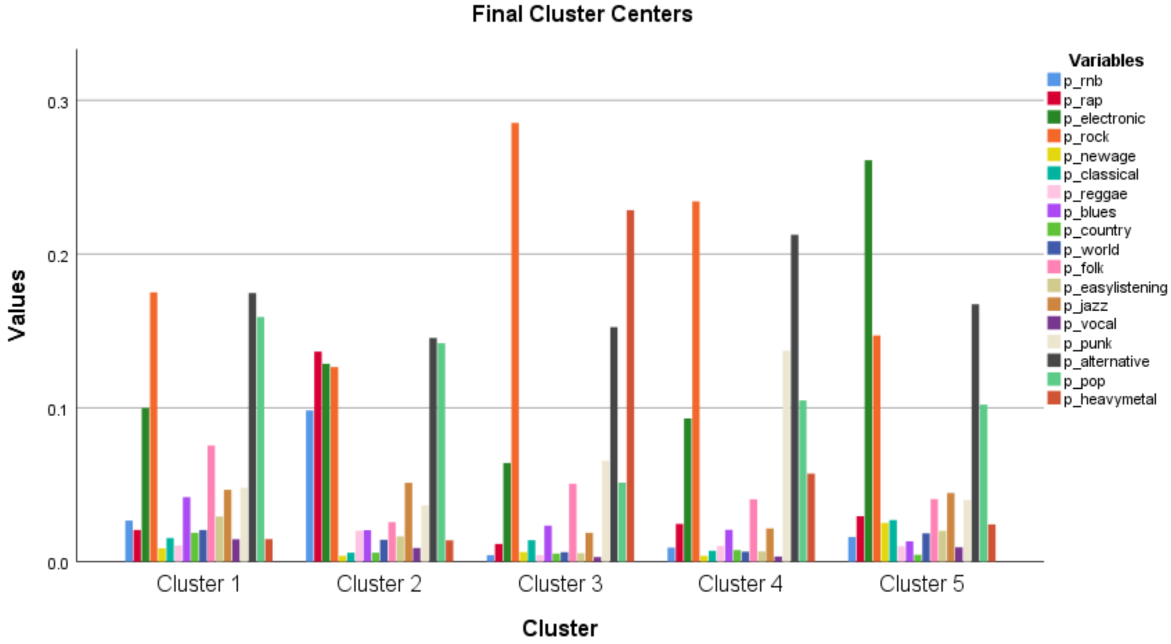

Figure 3 Bar graph of five cluster division based on genre percentage

Cluster 1 Omnivorous; Pop blues folk

The biggest of the five clusters is number one, were no extreme numbers are found, and many genres are listened to. This omnivorous group scores highest on pop, blues and folk music, but only with small differences in comparison to the other clusters. This omnivorous pattern of combining folk, pop and several ‘highbrow’ music genres is comparable to a taste pattern identified by Van Eijck and Lievens (2008) who described those that combine these three meta-genres as omnivores.

Cluster 2 Rap R&B reggae

The users in cluster 2 are the most frequent listeners to Rap and R&B. They listen to R&B over 3 times more than those in cluster one and ten times as much as cluster 4. And even more to Rap in comparison to the other clusters. This profile is very heavily skewed to rap and R&B. Suggesting that the listener of rap and R&B have a quite singular non diverse and univore taste which was also found in the research of Ollivier (2008) and Gelder (2007).

Cluster 3 Rock and heavy metal

Heavy metal shows to be a niche genre. Where those in cluster 3 listen to heavy metal for 23% while the other clusters do not get higher than 6% and most only have 1% or 2%. The listening to heavy metal and Rock music falls in the same cluster within this analysis, showing that it occurs often in the database that listeners score high both on listening to Rock as to heavy metal. They mix it up a little with alternative, some folk and punk, but are mainly listening to these two genres, which seem compatible within the listening profiles. Noteworthy is also their tendency to rarely listen to pop and to not listen to vocal music, reggae, or R&B at all. The singularity of this listening profile, might indicate the individuality of the music scene, and the often attached subculture where the distinction of the listening profile to other profiles and genres is important. To create an identity by opposing themselves against other groups and distinguished groups such as rap and R&B (Gelder, 2007).

Cluster 4 Alternative, Punk

The second largest cluster of user, combines the consumption of Punk and alternative music with Rock, electronic and a bit of pop. What sets them aside is the large amount of time they spend on listening to Punk music. Their combination of genres is les singular as for example cluster three is yet their combination of 21% punk with the highest score on alternative with 21% makes their cluster interpretable as having a distinct direction, and they might be considered univores. The music scene of punk is recognisable in this clusters listening pattern, where punk taste and punk subculture is individual direction of music taste that distinguishes itself from other subcultures and scenes.

The punk subculture started from politically loaded lyrics and get togethers at concerts strengthening the group structure and identity. Through the digitalisation of music and communication the music pattern spread to other countries and users. For the original punk subculture the digital communication means are used to continue conversations and discussions that originated from face to face meets and concerts. The 'only online' punk fans are dismissed by them as fake punkers', this lead to a duality within the punk music listeners, between the spread music listening pattern subculture, and the tight knit core of 'punkers' (Williams, 2006). Cluster 4 might be considered a representation of that through digitalisation grown group of punk listeners, representing not per se the political movement but the singular music taste pattern of punk, considering the members of cluster 4 are with thousands of last fm users, spread across the globe.

Cluster 5 Electronic, new age, classical

Cluster 5 stands out because of its highest listening to electronic music. They are also the most frequent listeners to classical and new age music and although they combine their consumption with some rock, alternative and pop music, their focus lies with electronic music. Though electronic music is recognized as a subculture, it has quite a spread across genres, even though the singularity of electronic music taste is very recognizable in the cluster's listening profile (Gelder, 2007).

4.2 Users up close

Intensity, diversity, age, gender

Out of the 30520 users in the sample, 17910 filled in their age (58,68%) and 22883 noted their gender (74,98%) as male or female. The means compared by age and gender, are based on these present numbers of users information. Noticeable was how only 63% of the users in cluster 1 provided information on their gender while 90% of the users in cluster 3 filled in their gender information. All differences in the means compared table are significant, except for the intensity of listening since those are not really different between the clusters. This is possibly because of the sample criteria which included users with a listening intensity between 20000 and 60000,which leads to an average listening intensity of 39200 songs which all the intensity means in the table are very similar to (Table 2).

Table 2: Comparison of cluster means

Cluster		intensity	diversity	age	percentage female
1-Pop, blues & folk	Mean	38602	10.29	26.0	48.33%
2-Rap, R&B & reggae	Mean	38807	9.75	22.6	35.65%
3-Rock & heavy metal	Mean	37392	7.20	24.2	24.00%
4-Alternative & punk	Mean	40878	8.15	23.4	34.96%
5-Electronic, new-age & classical	Mean	39804	8.90	25.6	30.27%
Total	Mean	39200	9.15	24.7	37.42%

Table 3:

Diversity means per gender

Gender	Diversity mean
Male	8.84
Female	9.08
Total	8.90

Table 4: Multinomial regression analysis of user information between clusters

Cluster		B	Wald	Sig.
1	Intercept	-6.620	936.334	.000
	Age	-.007	7.157	.007
	Gender	1.242	349.219	.000
	Intensity	.000	.314	.575
	Diversity	.638	1185.101	.000
2	Intercept	-3.182	109.638	.000
	age	-.086	184.733	.000
	gender	.423	21.718	.000
	intensity	.000	.054	.816
	diversity	.432	294.398	.000
3	Intercept	6.051	734.088	.000
	age	-.002	.505	.477
	gender	-.078	.874	.350
	intensity	.000	11.023	.001
	diversity	-.689	1091.625	.000
4	Intercept	2.618	179.845	.000
	age	-.026	62.379	.000
	gender	.479	50.156	.000
	intensity	.000	49.308	.000
	diversity	-.298	299.754	.000

a. The reference category is: 5.

In table 4 we see the differences between clusters considering the user information; age, gender, intensity and diversity calculated in a multinomial regression analysis. Showing the parameters of cluster differences. The green values are those that are looked into and together with the compared means table give more insight into how the users differ between the clustered listening profiles. The intensity variable show insignificant for cluster 1 and 2 and shows a neutral association with cluster 3 and 4. Therefor intensity is not further described as difference between the clusters. The results of the comparison of the cluster users is described based on both the compared means analysis (Table 2) and the multinomial regression analysis (Table 4).

In the tables (2 & 4) we see that the pop, blues and folk listeners have the largest diversity. This corresponds to the cluster analysis finding them omnivorous and their taste spread out. The lowest diversity score is found in the singular listening pattern of rock and heavy metal listeners which includes 13.42% less women than the total average does.

Cluster 5 also has quite a small percentage of women in their cluster with 7.18% less than average. These two predominantly male clusters, together with cluster 5, have quite a low diversity compared to the two clusters which include the most women, which have high diversity scores. Which would lead to the assumption that female users have a higher diversity, which is also shown in the compared means table for gender and diversity (Table 3). The correlation between being female and diversity is a significant positive of +.056, indicating that women have a larger tendency to have a diverse listening profile in opposition to men.

The cluster with the youngest users, Rap, R&B and reggae, cluster 2, also show a high diversity and reasonably high amount of female users in their cluster, within the age range of these cluster-means we cannot assume that the younger users have the highest diversity score, since the cluster 1 with the highest diversity score has the oldest age average.

Cluster Centers comparison male-female

The aspect of gender differences within listening profiles is an underrepresented aspect of the study to music taste and consumption (Ollivier, 2008). The following comparison between the cluster centers for both male and female users will provide some insight on gender differences and their correspondence to differences in listening profiles.

Out of the 30520 users in the data sample, 16033 are male (70,1%), 5138 are female, making up 22,5% of the users that provided their gender. 1712 categorized themselves as neutral. 7637 of the users did not provide the music service with any gender information, and are considered as missing values. To be able to compare the division of the users across the cluster the amount of users has been set to percentages of the total female and male user group. The differences of the genre profiles and the genre compatibility between male and female users are small but present.

In table 5 the percentages of genre consumption of both male and female group are put together. The green values are those that were the highest score of that genre within one gender group. red scores are the second highest, if the difference between the second highest and other scores was still significant, or there was a shared first place. and the yellow values are other quite high scores that might provide a more in depth image of the genre profiles. The noticeable differences between the male and female users are lined in the table.

Table 5: Cluster centers compared per gender

	1		2		3		4		5	
	M	F	M	F	M	F	M	F	M	F
R&B	.02	.02	.08	.10	.00	.01	.01	.01	.02	.02
Rap	.02	.02	.19	.08	.01	.01	.03	.02	.03	.02
Electronic	.11	.10	.13	.13	.06	.07	.09	.10	.28	.24
Rock	.18	.18	.12	.14	.29	.27	.24	.23	.14	.16
Newage	.01	.01	.00	.00	.01	.01	.00	.00	.02	.02
Classical	.02	.01	.00	.01	.01	.02	.01	.01	.02	.03
Reggae	.01	.01	.02	.02	.00	.00	.01	.01	.01	.01
Blues	.04	.04	.02	.02	.02	.02	.02	.02	.01	.01
Country	.01	.02	.00	.01	.01	.00	.01	.01	.00	.00
World	.02	.02	.01	.02	.01	.01	.01	.01	.02	.02
Folk	.07	.08	.02	.03	.05	.06	.04	.04	.04	.05
Easy Listening	.02	.03	.01	.02	.01	.01	.01	.01	.02	.02
Jazz	.05	.04	.06	.04	.02	.01	.02	.02	.04	.04
Vocal	.01	.01	.01	.01	.00	.00	.00	.00	.01	.01
Punk	.05	.05	.04	.04	.06	.06	.15	.14	.04	.04
Alternative	.18	.19	.15	.15	.15	.17	.21	.22	.16	.18
Pop	.15	.16	.12	.17	.05	.06	.09	.13	.10	.11
Heavy Metal	.02	.01	.02	.01	.24	.19	.07	.04	.03	.02
Nr. users	6172	2127	900	577	2511	610	4120	1258	2330	566
% users	38,5 %	41,4 %	5,6 %	11,2 %	15.7 %	11,9 %	25,7 %	24,5 %	14,5 %	11,0 %

In many of the more singular genre such as cluster 2 and 3 men are more extreme in their listening behaviour, where although the female users, make a distinctive choice to listen to specific genres their behaviour is less extreme. We see this in the cluster 2 rap R&B and reggae, in the genres Rap R&B and Jazz. And within the heavy metal and rock listener, the same pattern is seen, in heavy metal which is also the case for heavy metal in cluster 4, and with Electronic music in cluster 5. The woman in cluster 5 include more classical music in their profile than man do. The female listeners mix more pop into their profiles than males do, where female listeners have a higher listening intensity in pop in all clusters, especially within cluster 2 and 4 (Table 5).

4.3 Genres and diversity: a correlation

Several genres indicate a positive correlation with a wider diversity while listening to others suggest a smaller diversity. Some of the main positive correlations of genres with a wider diversity are: R&B, blues, easy listening, jazz, vocal, and world. While the listening to several others correlate negatively with the diversity of a listening profile; electronic, punk, and alternative, but the largest negative correlation is that for rock (-.564). Suggesting that the more rock is included in users' profiles, the less diverse their account is. Considering cultural omnivorousness the genres rock, punk, alternative and electronic may suggest a less diverse more singular univore taste pattern, which might be connected to them being ,according to Peterson (1996), 'lowbrow' , or their connection to subcultures which also show singularity in taste, we see the singularity and low diversity pattern in the correlation matrix as well as in the cluster analysis (Table 6 & Table 1)

Table 6: Genre Correlations

	Diversity	R&B	rap	electronic	rock	new age	classical	Reggae	blues	country	world	folk	easy listening	jazz	vocal	punk	alternative	pop	heavy metal
R&B	.302	1																	
rap	.059	.444	1																
electronic	-.063	~~.010~~	.091	1															
rock	-.564	-.426	-.368	-.427	1														
new age	.171	-.113	-.119	.271	-.273	1													
classical	.225	-.121	-.144	.124	-.196	.551	1												
reggae	.258	.179	.287	-.032	-.240	-.088	-.104	1											
blues	.470	.051	-.176	-.385	-.082	-.086	-.019	.030	1										
country	.344	~~-.009~~	-.149	-.302	-.094	-.081	-.026	~~-.015~~	.576	1									
world	.449	.072	-.076	.035	-.362	.329	.221	.153	.132	.035	1								
folk	.267	-.175	-.295	-.218	-.104	.065	.111	-.104	.285	.419	.251	1							
easy listening	.562	.198	-.118	-.058	-.440	.262	.201	.065	.364	.227	.321	.139	1						
jazz	.566	.219	.117	.066	-.470	.094	.103	.168	.390	.137	.331	.063	.491	1					
vocal	.478	.187	-.108	-.049	-.368	.170	.241	~~.007~~	.342	.196	.366	.156	.642	.352	1				
punk	-.398	-.294	-.066	-.258	.348	-.250	-.251	.033	-.222	-.133	-.316	-.225	-.402	-.387	-.341	1			
alternative	-.298	-.308	-.122	~~-.017~~	.250	-.243	-.285	-.111	-.240	-.131	-.305	-.138	-.369	-.323	-.413	.326	1		
pop	.305	.339	~~.015~~	-.038	-.391	-.169	-.161	~~.012~~	.099	.183	.049	.088	.271	.036	.208	-.201	.086	1	
heavy metal	-.479	-.285	-.167	-.280	.615	-.081	-.017	-.181	-.153	-.217	-.225	-.129	-.314	-.331	-.247	.084	-.174	-.589	1

<-.350
-.200- -.350
-.050- -.200
-.050- .050
.050-.200
.200 - .350
> .350

In table 6 the correlations between the different genres is shown in color gradient to visualize the positive and negative correlations. The insignificant outcomes are crossed out and left white.

The genre correlation table shows how the genres correlate within the database, meaning that a positive correlation show these genres are connected or compatible with each other, for example one that listens to heavy metal is very likely to also listen to rock and slightly likely to include punk in its listening profile. The heavy metal genre does not correlate well with listening country and even less to pop. Heavy metal shows to be quite an individual genre that does not correlate to many others, which shows in in his correlation to the diversity variable, which shows little diversity.

Several genres have an overall negative correlation with the other genres and are thus more self-contained, showing the more you listen to the genres the lesser the chance you listen to other genres, except for the few positive correlations.

The Omnivorous group of users on Last fm are those in cluster 1, where pop, blues and folk are most frequently listened to but only with a slight difference to other genres and cluster centers. The omnivorousness or compatibility of these genres in the overall profile of diversity is very clearly seen in table x. where all three main genres have a lot of positive correlations to other genres, only some genres that also showed to be negatively correlated to diversity in table 6 would a negative correlation with pop, blues and folk. Namely, punk, alternative, heavy metal and electronic.

Rap and R&B correlate very positively with each other, making the combining of these genres very likely. They both correlate with reggae as well, which we also see in our cluster division where rap R&B and combined with reggae form the main genres in cluster 2.

The two main genres from the earlier described cluster 3 ‘rock and heavy-metal’ indeed relate very positively to each other and to very little other, only punk is slightly positively related to these two genres. Rock is also positively related to alternative while heavy metal is not. Yet rock has some very negative correlations while heavy metal is only slightly negatively related to the other genres and even neutrally related to classical music while rock listeners have a negative correlation to classical music. Bothe genres are over all quite singular in their correlations which reflects the negative correlation with diversity as seen in table 6.

Alternative is only positively related to two other genres namely rock and punk. Punk is only positively correlated with rock, alternative and very slightly to reggae and heavy metal. In cluster 4 Punk and alternative make the most important genres, with a very non diverse profile, including rock, some pop and heavy metal. The large amount of negative correlations and the singularity of cluster 4 matches the correlation of the genres to the diversity variable, being negative for both punk and alternative (Table 6).

5. Conclusion and discussion

The aim of this research was to gain better insight into music listening behaviour patterns of consumers and to find how these patterns correlate with listening diversity and listening intensity within the listening behavioural data of Last fm. Three sub questions were formulated to find these patterns, correlations and different users within them:

1. Which listening behavioural patterns can be found within the Last fm database considering genre compatibility?
2. How do the users differ between the clustered listening profiles?
3. How do specific genre preferences relate to diversity and intensity of music listening?

In the research, an optimum of five listening profiles was found. Within the user description of these clusters, some differences and similarities between the user's listening profiles were found, next to genre compatibilities and combinations which had influence on the diversity of the listening profiles, both per cluster and in the correlation regression. In this chapter, the three research questions of interest are addressed in order, after which the significance, contribution and limitations of this research to sociology is discussed in combination with a discussion on future prospects of research.

5.1 Conclusion: patterns and people

Listening behavioural patterns and users up close

Which listening behavioural patterns can be found within the last fm database considering genre compatibility?

How do the users differ between the clustered listening profiles?

By means of a k-means cluster analysis five listening behavioural patterns were found in the database, all focussing their listening behaviour more on certain genres than others. Within the users clustered into these five consumer profiles several differences can be seen, considering age, gender and diversity. For the intensity score no significant difference was found, probably due to intensity being the indicator of the sample.

The five clusters of listening behaviour are:

- Cluster 1 Omnivorous: Pop blues folk

As the largest cluster, with almost 50% female listeners, this cluster combines pop,blues and folk with a little of many other genres as well. It is a diverse listening profile (diversity score 10.29) with a widespread consumption pattern and, in comparison to the other clusters, a quite old user group with an average age of 26 years. The female and male users in this cluster show very little difference in listening intensity of genres or genre combinations.

- Cluster 2: Rap R&B reggae

This young (22.6 yr) and diverse (diversity score 9.75) listening profile combines Rap and R&B with reggae. When we separate the male and female listeners (35,65%), jazz comes into play in this cluster for both male and female listeners, while female users mix more pop into this listening profile.

- Cluster 3: Rock and heavy metal

The Rock and heavy metal listeners show the least diversity of all clusters with only a diversity score of 7.20, which is very low compared to the overall average of 9.15 and with the second lowest, cluster 4, scoring 8.15. Cluster three has very few women in its user base (24%). These women show a slightly different listening pattern than the men. The females in cluster 3 spend less time listening to heavy metal, 19% compared to the male 24%, the female's listening is a bit more spread across other genres, such as folk and alternative.

- Cluster 4: Alternative, Punk

With a distinct direction of Punk and alternative music consumption, cluster five shows a quite singular taste pattern. A quarter of all male users finds himself in this cluster, as well as a quarter of all female users. Males and females differ slightly when comparing the distribution of listening intensity across the genres. Women in this cluster listen to heavy metal half as much as the male users do, while mixing more pop into the pattern than men do. An explanation for the popularity of punk in the database can be the digitalisation of the music and the easy spread of music, which watered out the subculture into a music scene which attracted a lot of fans and followers.

- Cluster 5: Electronic, new age, classical

The electronic music listeners mix new age and classical music in their listening profile. Cluster 5 has a below average amount of female users, and an above average age. The genders in the cluster show a female group of users, spreading more evenly, listening to slightly less electronic and more classical music than the men do. Folk, alternative and pop are also included a bit more by the female users.

Gender in music profiles

What role does gender play in listening pattern, cluster membership and listening diversity?

When comparing the cluster centers of male and female users, we see women mixing more genres and being less singular in their listening behaviour. In total females show a larger diversity average in their profiles, and have a positive correlation with the listening diversity variable. This indicates that women are more diverse consumers of music than man.

Genre combinations, compatibility and diversity correlation

'How do specific genre preferences relate to diversity and intensity of music listening?'

The correlations between the separate genres and the diversity score of users showed several patterns. Many of those are relatable to the cluster divisions, as rock, heavy metal, punk and alternative had a negative correlation with diversity and very few positive connections to other genres except the main ones in their own cluster. The genres that did not stand out at the identifiers for the clusters, showed a very positive correlation with diversity and with each other; blues, country, easy listening, folk and vocal formed a 'green block' of correlations with each other and with the diversity indicator, further combined with positive correlations with jazz, classical and new age music. This 'green block' showed many connections with each other and to diversity indicating a more omnivore, and widespread taste pattern.

5.2 Discussion

Contribution and significance

This research on genre compatibility and consumer profiles within the Last fm database found several listening behavioural patterns and correlations within music taste and consumptions, contributing to music consumption and taste patterns theory within sociology. The five clusters described several music scenes, that can be linked to subcultures and specific tastes patterns, and the combinations of genres within. Van Eijck and Lievens pattern (2008) on the triptych of music scenes linked to omnivorism were reinforced by the clustering as well as by the genre correlations matrix. Also several genre's linked to subculture and distinctive styles, such as described by; Gelder (2007); Frith (1987); Williams (2006). were indeed shown to have less diversity than other genres and listening profiles in the study. This research is adding sociological analysis and insights of this data on user grouping and genre correlations to diversity to the ongoing research of the MMC department of the TU Delft.

The findings from this large dataset analysis include listening profiles and correlations that would not have been possible or less specific when using a research method such as a survey or qualitative research methods. By researching the actual consumption and listening behaviour instead of asking questions on music taste in a survey, more distinctions and depth of knowledge is reached (Dimaggio, 1987). Because real listening behavioural data from digital platform was used, there was less human interpretations bias or genrefication (Lena, 2016).

This research also looked into the gender differences considering combining genres and listening diversity, a matter that has been underrepresented in many music theory studies (Ollivier, 2008). This research indicates that female music listeners have a more diverse taste than men, and tend to mix more pop into the more singular taste patterns than men do. This finding is in contradiction to the few indicators that Ollivier found, where women were underrepresented in the omnivore highbrow taste category, at least on the topic of listening diversity and the mixing of genres. Overall the research contributed a statistical insight on music listening behaviours and genre correlations to the theory of music patterns and behaviour.

Limitations

This research is based on data from the online music platform last fm, which might attract a certain type of person, or type of music listener creating a sampling bias. On last fm a recommendation algorithm is actively recommending diverse types of music, and new artists to users, which combined to the charts and statistics one can see on their profile motivates diversitizing the users profile. This was also described by Demetriou et al. (in press), showing that the higher a person's intensity the wider it's diversity.

Future prospects

From the results of this thesis research, a number of interesting future research options sprouted. First, the combination of qualitative research methods combined with statistical analysis can prove fruitful. For example, interviewing several users from the dataset about their cultural participation and social background, or the way they perceive music listening and their choices. This would provide for an in depth insight into the user including a longitudinal retrospective statistical study on their listening behaviour.

In this thesis research, a sample based on listening intensity was made, excluding low intensity users and 'super' users. Comparing the findings from this study with the same analysis for users with a smaller listening intensity, and super users (more than 60000 songs listened) might show differences in the clustering or correlations and can possibly add to the question on the development of taste through digital intermediaries. Another possible research within the last fm database is to add the timestamps of each song into the database. This enables the mapping of listening behaviour over the different hours of the day, and allows for the researcher to follow an user's music behaviour development. Such research could answer follow up questions such as:

- To what extent are the univore patterns in cluster heavy metal & rock and rap and R&B comparable to the by Peterson created theory of upper class omnivorousness and lower class univores and to what extent are collaborative filtering or filter bubbles responsible for the singularity in these clusters?
- Do personal tastes dissolve by means of programmed general broadening?
- Are the users in cluster one omnivorous in their taste pattern and culturally wide orientated as a person or is the mainstream listening behaviour including the recommendation influence just quite broad within the different genres?

6. References

Aguiar, L. (2017). Let the music play? Free streaming and its effects on digital music consumption. *Information Economics and Policy*, *41*, 1–14. https://doi.org/10.1016/j.infoecopol.2017.06.002

Bennett, A., & Peterson, R. A. (2004). *Music Scenes: Local, Translocal and Virtual*. Vanderbilt University Press.

Bourdieu, P. (1973). Cultural reproduction and social reproduction. London: Tavistock, 178.

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

Corrêa, D. C., & Rodrigues, F. (2016). A survey on symbolic data-based music genre classification. *Expert Systems with Applications*, *60*, 190–210. https://doi.org/10.1016/j.eswa.2016.04.008

Datta, H., Knox, G., & Bronnenberg, B. J. (2018). Changing Their Tune: How Consumers' Adoption of Online Streaming Affects Music Consumption and Discovery. *Marketing Science*, *37*(1), 5–21. https://doi.org/10.1287/mksc.2017.1051

Demetriou et al. (in press) Beyond Explicit Reports: Comparing Data-Driven Approaches to Studying Underlying Dimensions of Music Preference.

DiMaggio, P. (1987). Classification in Art. *_American Sociological Review_*, *52(4)*:440-455.

Donnat, O. (2005). 49. La féminisation des pratiques culturelles. Dans : Margaret Maruani éd., Femmes, genre et sociétés: *L'état des savoirs* (pp. 423-431). Paris: La Découverte.

Eijck, Koen & Lievens, John. (2008). Cultural Omnivorousness as a Combination of Highbrow, Pop, and Folk Elements: The Relation between Taste Patterns and Attitudes Concerning Social Integration. *Poetics. 36*. 217-242. 10.1016/j.poetic.2008.02.002.

Ferrant, C. (2018). Class, culture, and structure: Stratification and mechanisms of omnivorousness. *Sociology Compass*, *12*(7). https://doi.org/10.1111/soc4.12590,

Frith, S. (1981). Sound Effects; Youth, Leisure, and the Politics of Rock'n'roll. Retrieved from https://www.cabdirect.org/cabdirect/19841806277

Frith, S. (1987). Music and identity. In *Questions of cultural identity* (pp. 108–150). Washington: Georgetown university.

Frith, S (1990). What is good music? Canadian university music review, 10 (2), 92-102

Gelder, K. (2007). Subcultures. London: Routledge, https://doi.org/10.4324/9780203446850

Hazir, I. K., & Warde, A. (2015). The cultural omnivore thesis: Methodological aspects of the debate. In L. Hanquinet & M. Savage (Red.), *Routledge International Handbook of the Sociology of Art and Culture (1ste edition)*. London: Routledge.

van den Haak, M. A. (2014). Disputing about taste: Practices and perceptions of cultural hierarchy in the Netherlands. Retrieved from https://dare.uva.nl/search?identifier=a5e9d9fc-6d54-4007-bbb7-b0aa53b8102c

Harvey, E. (2014) Station ot station: The past, present and future of dtreaming music. Retrieved May 27, 2019, from https://pitchfork.com/features/coverstory/reader/streaming/

Jæger, M. M., & Katz-Gerro, T. (2008). The rise of the cultural omnivore. *Research department of social policy and welfare services.* Consulted on https://pure.sfi.dk/ws/files/444063/WP_09_2008.pdf

Kagie, M, van der Loos, M.J.H.M, & van Wezel, M.C. (2008). *Including Item Characteristics in the Probabilistic Latent Semantic Analysis Model for Collaborative Filtering*

Kissel, C. (2015, 23 juli). *Spotify Listeners Discover Roughly 27 New Artists a Month*. Consulted on 3 juni 2019, van https://diffuser.fm/spotify-listeners-discover-roughly-27-new-artists-a-month/

Lena, J. C. (2016). Genre: Relational approaches to the sociology of music. In L. Hanquinet & M. Savage (Eds.),*Routledge International Handbook of the Sociology of Art and Culture* (pp. 149-160).Routledge

Liem, C. S., Müller, M., Eck, D., Tzanetakis, G., & Hanjalic, A. (2011). The need for music information retrieval with user-centered and multimodal strategies. *Proceedings of the 1st international ACM workshop on Music information retrieval with user-centered and multimodal strategies - MIRUM '11*, 1–6. https://doi.org/10.1145/2072529.2072531

Lu, C. C., & Tseng, V. S. (2009). A novel method for personalized music recommendation. *Expert Systems with Applications*, *36*(6), 35–44. https://doi.org/10.1016/j.eswa.2009.01.074

Michael, J. (2017). Highbrow culture for high-potentials? Cultural orientations of a business elite in the making. *Poetics, 61*, 39 - 52.

Nell, A. V. (2014). Recommended Culture. Distinctions in cultural consumption and taste in a digitized and recommendation driven age (Master's thesis).

Ollivier, M. (2008). Modes of openness to cultural diversity: Humanist, populist, practical, and indifferent. *Poetics*, *36*(2–3), 120–147. https://doi.org/10.1016/j.poetic.2008.02.005

Peterson, R. A. (1992). Understanding audience segmentation: From elite and mass to omnivore and univore. *Poetics, 21*(4), 243-258.

Peterson, M., & Simkus, A. (1992). How musical tastes mark occupational status groups. In M. Lamont & M. Fournier (Red.), *Cultivating Differences: Symbolic Boundaries and the Making of Inequality* (pp. 152–186). University of Chicago Press.

Peterson, R. A., & Kern, R. M. (1996). Changing Highbrow Taste: From Snob to Omnivore. *American Sociological Review*, *61*(5), 900–907. https://doi.org/10.2307/2096460

Prey, R. (2016). Musica Analytica: The Datafication of Listening. *Networked Music Cultures,* 31–48. https://doi.org/10.1057/978-1-137-58290-4_3

Van Rees, K., & Van Eijck, K. (2003). Media repertoires of selective audiences: the impact of status, gender, and age on media use. *Poetics*, *31*(5-6), 465–490. https://doi.org/10.1016/j.poetic.2003.09.005

Reddit, Lastfm (n.d.) - *List of best programms/sites to analyze Last.fm data*. . Consulted on 4 juni 2019, van https://www.reddit.com/r/lastfm/comments/7xb65y/list_of_best_programmssites_to_analyze_lastfm_data/

Rowland, F. (2011). The Filter Bubble: What the Internet is Hiding from You (review). portal: Libraries and the Academy 11(4), 1009-1011. Johns Hopkins University Press. Retrieved November 6, 2018, from Project MUSE database.

Sánchez-Moreno, D., Gil González, A. B., Muñoz Vicente, M. D., López Batista, F., & Moreno García, M. N. (2016). A collaborative filtering method for music recommendation using playing coefficients for artists and users. *Expert Systems with Applications*, *66*, 234–244. https://doi.org/10.1016/j.eswa.2016.09.019

Schedl, M. (2016). The LFM-1b Dataset for Music Retrieval and Recommendation. *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval - ICMR '16*. https://doi.org/10.1145/2911996.2912004

Schedl, Marcus, Nicola Orio, Cynthia C. S. Liem, and Geoffroy Peeters. 2013. A professionally annotated and enriched multimodal data set on popular music. *In Proceedings of the 4th ACM Multimedia Systems Conference (MMSys '13*). ACM, New York, NY, USA, 78-83. DOI=http://dx.doi.org/10.1145/2483977.2483985

Schedl M. and Ferwerda B. 2017. Large-Scale Analysis of Group-Specific Music Genre Taste from Collaborative Tags. IISM . IEEE Computer Society, 479–482.

Schulze, G., 1992. Die Erlebsnisgesellschaft: Kultursoziologie der Gegenwart. Campus Verlag, Frankfurt.

Vlegels, J., & Lievens, J. (2017). Music classification, genres, and taste patterns: A ground-up network analysis on the clustering of artist preferences. *Poetics*, *60*, 76–89. https://doi.org/10.1016/j.poetic.2016.08.004

Webster, J. G. (2014). The Marketplace of Attention: How Audiences Take Shape in a Digital Age. Cambrige, Massachussets, United States of America: The MIT press.

Williams, J. P. (2006). Authentic Identities: Straightedge Subculture, Music, and the Internet. Journal of Contemporary Ethnography, 35(2), 173–200. https://doi.org/10.1177/0891241605285100

Wright, David. (2005). Mediating production and consumption: Cultural capital and 'cultural workers'. *The British journal of sociology*. *56.* 105-21. 10.1111/j.1468-4446.2005.00049.x.