ERASMUS UNIVERSITEIT ROTTERDAM

MSc in Economics and Informatics

January 8, 2014

The influence our friends have on our music taste: An empirical analysis

by

Harwick Schob

325050

Supervisor: T. Tervonen

Co-reader: N. Baştürk

Table of Contents

Abstract

In this thesis, we will examine the relationship of social influence and music taste, and how weather influences one's music taste. This thesis reports results of an empirical investigation of interpersonal relationships on Last.fm, a music-based social network site. In addition, we have shown that temperature, cloud coverage, and sunshine duration can influence the amount of music and also the type of music one listens to. The chief goals of this study were to examine the degree to which music taste is characterized by social influence. Results indicate that although there is some evidence suggesting that one's peers influence their music taste, we did not observe sufficient significant evidence to conclude this.

Acknowledgements

Thank Thanks goes first and foremost to our supervisor Tommi Tervonen, who was always ready to help when we needed advice. Second, special thanks goes to Carlos da Silva Lourenço, who took time out of his busy schedule to give helpful advice and comments during various phases of the work. Additionally, we want to thank Nalan Baştürk for taking the time to read the thesis and give helpful feedback.

Finally, we would like to thank our families and friends for supporting us for the last four years, in many ways.

1 Introduction

The music industry has suffered a lot of losses since the mid-1990s when the music industry was without a doubt a structured and thriving sector. Compact discs were introduced in 1983 as a result of a joint venture between Philips and Sony. Due to this innovation and the growth in music television and music videos had brought a new impetus to a previously stagnating music industry (Sanjek and Sanjek 1991; Denisoff 1988). The shift from vinyl records to CDs had tripled the world wide sales for the recording industry from $12.3 billion (1985) to $39.7 billion (1995) (IFPI, 1999). Since the mid-1980's the global market has been dominated by five record companies, namely Universal, Polygram, Sony Music Entertainment, EMI, Warner Music Group, and Bertelsmann Group (BMG) (Tschmuck, 2012). According to McChesney (1997), it was estimated that these *majors* controlled between 80 to 90 percent of the global music market. These companies were vertically integrated and had total control over the selection and management of musicians, recordings and copyrights. They also produced music in their own recording studios and pressing plants and regulated their products' global distribution through their own distribution systems (Dolata, 2011).

More than a decade later, the music industry has changed completely. The four major record companies that remained still control around 70% of the global music market[1]. In spite of these figures, the music industry has suffered a lot of losses since the end of the 1990s. The global recording sales have decreased from $40.5 billion in 1999 to $27.8 billion in 2008 (IFPI, 2010). The record sales in the US alone fell from $14.3 billion in 2000 to $7.7 billion in 2009 (RIAA 2008, 2010). The main reasons for this decrease in record sales were as a result of the drastic drop of traditional CDs' sales and the rise of the digital market.

The internet was a technology that transformed the music industry. To understand the technology-driven transformation behind the music industry, one must go back in time when the music industry still operated in the traditional way. Contrary to DVDs that were introduced in the mid-1990s, CDs were introduced back in 1983 without any copy restrictions. This made it possible, with the introduction of CD recorders and writable CDs in the mid-1990s, to copy any digital media without use restrictions. Besides this, in the second half of the 1990s, the introduction of MP3s made it possible to not only share music files

[1] http://www.copynot.org/Pages/The%20big%20four%20Record%20Companies.html

online, but also convert music data onto CDs. Therefore, it could be said that the music industry had underestimated these two technological developments (Dolata, 2011).

The music industry and their interest groups, specifically the International Federation of the Phonographic Industry (IFPI) and the Recording Industry Association of America (RIAA), devised by the second half of the 1990s two containment strategies to deal with the new technological challenges. The first containment strategy was signed by the end of 1996 by the US government and the World Intellectual Property Organization (WIPO). The strategy consisted of two treaties, namely the WIPO Copyright and the WIPO Performances and Phonograms. These treaties gave music companies and artists the exclusive rights to publish their music online. On top of this, they also supplied legal protections to protect their products against bypassing the technical protections. At the end of 1998 the Digital Millennium Copyright Act (DMCA) was implemented by the influence of the Motion Picture Association of America (MPAA) and the RIAA. Due to the DMCA act was the WIPO treaties transposed into national law (U.S. Copyright Office, 1998).

The second containment strategy was implemented at the end of 1998 by the RIAA and the IFPI. They founded the Secure Digital Music Initiative (SDMI) to control the illegal distribution of music files through technical restrictions. Their aim was to develop a universally applicable technical standard for digital music in order to manage the legal sales of music files. This created a partnership between more than 200 companies, which included not only large music companies, but also leading producers of consumer electronics as well as hardware and software producers of the information technology industry. Their mission was to develop and implement a compatible Digital Rights Management standard (DRM) for digital music, recorders and players. However, the strategy was implemented too late. Most music was by the end of the 1990s obtainable in unprotected formats (Dolata, 2011).

By the end of the 1990s was the music industry taken by surprise with the unforeseen fame of file sharing networks such as Napster, Kazaa, Gnutella, and Freenet. These networks made it possible to exchange free online music files directly with one another. By the beginning of 2001, the exchange of online free music through Napster alone reached approximately 44.6 million consumers worldwide (Alderman, 2002). Already prior to these networks was it without a doubt possible to copy and exchange music. However, the exchange was local in nature, restricted to family and friends and limited to the physical copies of music available to them. These file sharing networks removed these restrictions and made it easy to download free music instead of purchasing music to make copies of it. The exchange of music shifted from local contexts to the global scale of digital community networks (Hughes and Lang, 2003).

The music industry and the RIAA reacted to the exchange of free music online by filing a suit against Napster in December 1999. They claimed damages caused by continuous copyright infringement. In mid-2001 Napster had to shut down their ongoing operations after a numerous defeats in court. Simultaneously, the RIAA filed suits against other file sharing networks such as Kazaa and Gnutella and was capable of shutting down other providers (Alderman, 2002).

Furthermore, the music industry attempted between 2000 and 2003 to succeed with their own commercial downloads and tried to bring the digital marketplace under their control (Dolata, 2011). However, there did not exist a notable market for digital music between 2000 and 2003. Although early commercial providers such as eMusic were launched at the end of the 1990s within the independent label business, songs from the major music companies were not available for digital purchase until 2000 (Tschmuck, 2012).

According to Dolata (2011), the breakthrough of commercial digital music distribution occurred thanks to an industry outsider, namely Apple Computers. In 2003 Apple introduced the iTunes music store in the US and one year later also in the UK, Germany and France. The iTunes music stores were among the first to offer songs from all the major record companies and from over 1000 independent labels, making Apple the first that could offer commercial downloading of music in combination with their iPod digital music player. By the end of 2007 iTunes had sold over 3 billion song titles and Apple had sold over 100 million iPods (Apple, 2007). Even though there existed more than 500 different online digital music stores in 2007, Apple still claimed since 2003 more than 80 percent of the leading digital music markets in 2007 (IFPI, 2008).

In spite of the fact that the music industry has suffered a lot of challenges and changes over the last decades, little is known about how individual’s music taste is influenced by either friends or external factors. The purpose of this research is to identify whether social influence or external factors play a role in influencing our music taste.

The remainder of this thesis is structured as follows:

Chapter 2 is a brief introduction of the theoretical background, which consists of music taste, social networks, social influence, and homophily.

Chapter 3 explores the modeling of music taste and gives an introduction of the variables being studied in this research.

Chapter 4 describes the research methodology, source of the research, how the data was collected, the sample construction, and the hypothesis development.

Chapter 5 presents the results of the analysis and explains the additional tests that have been performed.

Chapter 6 discusses the conclusions and future research.

2 Theoretical background

2.1 Music taste

As Levitin (2011) points out, there are numerous factors involved when an individual selects a specific song, album, or genre of music to be played. However, not much has been identified regarding the fundamental principles serving as a basis for individual musical taste. Due to the fact that music is used for many different purposes, it is even more difficult to investigate individual's music taste: Kohut and Levarie (1950) identified that music was used in contemporary society for pure enjoyment and aesthetic appreciation, whereas Dwyer (1995) and Large (2000) found that music has the ability to inspire dance and physical movement. According to Rentfrow and Gosling (2003), music is also used by many individuals for mood regulation and enhancement.

Cattell and Saunders (1954) were among the first to investigate individual differences in music preferences. They intended to create a method for evaluating dimensions of unconscious personality traits by creating a music preference test comprising of 120 classical and jazz music excerpts, where respondents had to report their degree of liking for each. They sought to understand 12 factors to explain them with regard to unconscious personality traits. For example, musical excerpts identified as melancholy and slow tempos were classified as the factor *sensitivity*. The excerpts with fast tempos characterized another factor classified as *surgency*. As Rentfrow, Goldberg and Levitin (2011) emphasize in their report, *"The structure of musical preferences: A five-factor model"*, Cattell's music taste measure never gained traction, but his results were among the first to suggest a latent structure to music taste.

Research on individual differences in music preferences reemerged nearly 50 years later. Rentfrow and Gosling (2003) explored individual differences in music preferences by investigating not only the lay beliefs about music, but also the structure underlying music preferences, and links between music preferences and personality. Results from the music preferences of 3500 individuals converged to show 4 music-preference factors, namely *Reflective and Complex, Intense and Rebellious, Upbeat and Conventional,* and *Energetic and Rhythmic*. The *Reflective and Complex* factor comprised of genres that appear to facilitate introspection and are structurally complex such as blues, jazz, classical, and folk music. The *Intense and Rebellious* factor, on the other hand, consisted of genres that are full of energy and emphasized themes of rebellion such as rock, alternative, and heavy metal. Then again, the *Upbeat and Conventional* factor is composed of genres that emphasize positive emotions and are structurally simple like country, sound track, religious, and pop music. The fourth and last factor named *Energetic and Rhythmic* included genres that are lively and often emphasize the rhythm, some of which are: rap/hip-hop, soul/funk, and electronica/dance music.

Obviously, some individuals tend to have stronger preferences towards a certain type of music than others. The most important questions that arise from this fact are: What influences a person's preferences, and are there specific individual differences relating people to a particular genre of music? As mentioned in previous section, there are studies that indicated links to personality when it comes to music preferences (Cattell & Saunders, 1954; Litle & Zuckerman, 1986; McCown, Keiser, Mulhearn & Williamson, 1997). Other studies revealed links between music preferences and physiological arousal (Oyama et al., 1987; Rider, Floyd & Kirkpatrick, 1985; McNamara & Ballard, 1999). Social identity was also associated with music preferences in numerous studies (North & Hargreaves, 1999; Tarrant, North & Hargreaves, 2000; North, Hargreaves & O'Neill, 2000). The following paragraphs will explain these links to music preferences in more detail.

Personality. Cattell was one of the first researchers to attempt to understand personality through music preferences. He assumed that an individual who likes a specific type of music may uncover vital information regarding unconscious aspects of personality (Cattell and Saunders, 1954). They developed a test, named I.P.A.T Music Preference Test consisting of 120 classical and jazz music items, where individuals had to specify how much they like each item. Cattell and Saunders (1954) applied factor analysis in order to establish 12 music-preference factors and regarded every one as an unconscious reflection of particular personality characteristics. As Rentfrow and Gosling (2003) emphasize, in contrast to Cattell who assumed that music preferences provide a window into the unconscious, most researchers have considered music preferences as a manifestation of more explicit personality traits. Litle and Zuckerman (1986), for instance, point out that sensation seeking seems to be positively correlated with preferences for punk, rock, and heavy metal music. On the other hand, they also revealed that preferences for religious music and sound tracks are negatively correlated.

Physiological Arousal. Other studies focusing on music preferences have shifted their attention to the physiological correlates of music preferences. To give an example, individuals who listen to heavy metal tend to perceive higher resting arousal than individuals listening to country music. Moreover, Gowensmith and Bloom (1997) have shown that heavy metal music increases the arousal level of individuals listening to heavy metal music surpassing that of individuals who listen to country music. Likewise, McNamara and Ballard (1999) established that individuals with preference for genres such as heavy metal, rock, alternative, rap, and dance seems to be positively correlated with resting arousal, sensation seeking, and antisocial personality. These genres are classified as highly arousing music.

Social Identity. The relationship between music preferences and personality has also been associated with research on social identity. North and Hargreaves (1999) point out that individuals consume music as a “badge” to convey their values, attitudes, and self-views. They investigated the typical attributes of rap and pop music fans. These individuals’ preferences were associated with the extent to which their self-views corresponded with the attributes of the prototypical music fan. Nonetheless, it was concluded that individuals with higher self-esteem had more similarity with one another than individuals with low self-esteem. The idea that people’s self-views and self-esteem is influenced by music preferences have been contributed by findings in various populations, age groups, and cultures (North, Hargreaves & O'Neill, 2000).

Despite the fact that the outcomes from these researches provide compelling information on music preferences and personality, they still present an incomplete picture. The majority of the researches considered merely a small collection of music genres. To give a few examples, Cattell and Saunders (1954) studied only classical and jazz music, Gowensmith and Bloom (1997) considered heavy metal and country music, and North and Hargreaves (1999) analyzed pop and rap music. Furthermore, nearly all research on music preferences investigated a small number of personality aspects. For instance, Little and Zuckerman (1986) focused on sensation seeking, McCown et al. (1997) studied Extraversion and Psychoticism, and McNamara and Ballard (1999) explored antisocial personality.

2.2 Social Networks

In this last decade, studying networks has become very popular in a wide range of fields such as sociology, psychology, anthropology, biology, communication studies, economics, geography, information science, and organizational studies. The exchange of information between the various disciplines has been slow due to the diverse backgrounds of researchers and numerous scientific communities. Figure 2.1 illustrates the wide range of applications and fields in the discipline of network research.

Graph theory is the groundwork of network research and was first introduced by the distinguished researcher Euler (1736) in his report, *"Solutio Problematis Ad geometriam Situs Pertinentis"*, where he introduced a solution to the Königsberg bridge problem to connect different parts of the city. Social networks have been analyzed by sociologists since the 1930s. They have contributed significantly to the statistical methods and empirical analysis for studying networks.

In the last decade, researchers such as mathematicians and physicists have also become part of the network research. This current field study is often referred to as *complex networks research*, which is interested in analyzing the similarities and differences of the countless types of networks that exists (Costa et al., 2007).

The rest of this chapter will examine some basic concepts of networks in order to create a clear understanding of this concept. A network has been formally defined as a set of pairwise relations between entities. In the context of graph theory, these "graphs" consists of a set of *vertices* or *nodes*, and the relations between these vertices are described as *edges*.

The *degree* of the node can be described as the amount of edges connected to a specific node. A *fully connected* network or *dense network* is when each node is connected to an edge. This can be calculated with the following equation where a network with N nodes the number of edges is $E = N (N - 1) / 2$. However, most real-world networks are connected by a low number of edges, and are called *sparse networks*. These can be found in networks in which links are more difficult to create such as citation networks, friendship networks, and power-line networks.

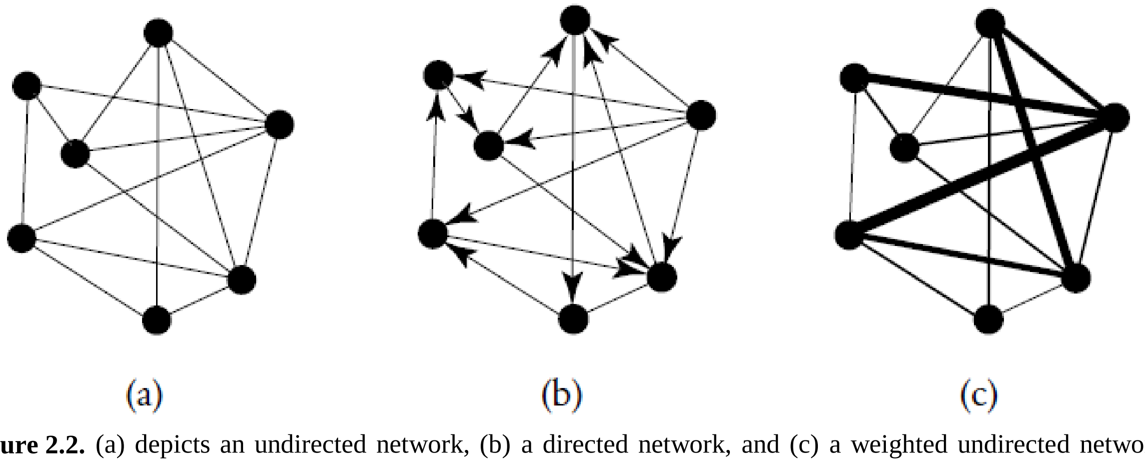


Figure 2.2. (a) depicts an undirected network, (b) a directed network, and (c) a weighted undirected network. Source Boccaletti et al. (2006)

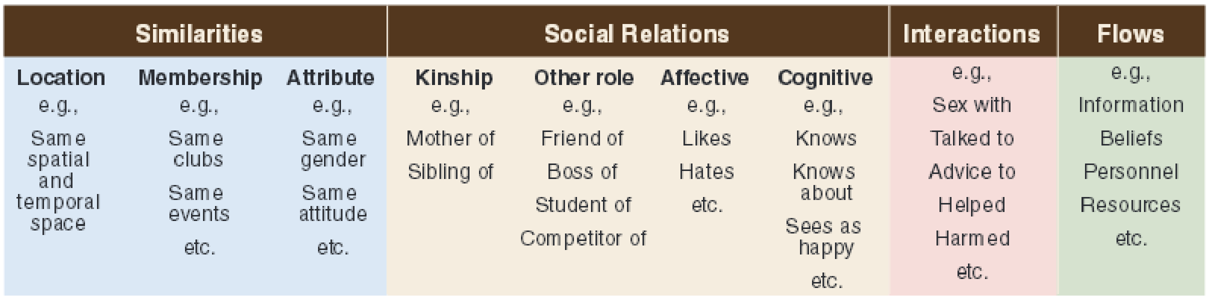
A network can be depicted as any system where individual elements interact with one another. These include web pages (Watts, 1999), journal articles (White, Wellman and Nazer, 2004), countries, neighborhoods, departments within organizations (Quan-Haase and Wellman, 2005), or positions (Boorman and White, 1976; White et al., 1976). Figure 2.2 illustrates some common categorization of networks, which are *undirected, directed, and weighted* networks or graphs.

2.2.1 Social network analysis

Wasserman (1994) stated in his book, *"Social network analysis: Methods and applications"*, that the term social network was first introduced in 1954 by Barnes. According to Marin and Wellman (2011) a social network is a set of socially-relevant nodes connected by one or more relations. They also emphasized that the nodes or actors are the units that are connected by the relations whose patterns they study. Usually these units are either persons or organizations, however, any objects or nodes that can be connected to one another can be analyzed. This field is known as *social network analysis* (SNA).

Wasserman and Faust (1994) point out that the goal of social network analysis is to create a model of the social interactions between individuals, and then study how this structure influences the functioning of these individuals and groups in the network. However, an initial challenge in social network analysis is to establish which nodes to include. Laumann et al. (1989) established three techniques to address this problem, which he named the *boundary specification problem*. His first proposition was a *position-based* approach. This approach regards those actors who are members of an organization or hold particular formally defined positions to be network members and all others would be excluded (Marin & Wellman, 2011). The second approach was defined as an *event-based* approach, which tries to establish the boundaries of the network by looking at who had participated in key events believed to define the population of the network (Marin & Wellman, 2011). The last approach that Laumann (1989) identified was the *relation-based* approach. This approach starts with a small set of nodes considered to be within the population of interest and then includes others sharing particular types (Marin and Wellman, 2011). Marin and Wellman (2011) emphasized that most of the time these three approaches will not be used exclusively, but rather in combination to define network boundaries.

In addition to establishing network members, researchers must determine the relations between these nodes. According to Wasserman and Faust (1994) these could include collaborations, friendships, trade ties, web links, citations, resource flows, information flows, exchanges of social support, or any other possible connection between these particular units. Borgatti et al. (2009) describe in their report, *“Network Analysis in the Social Sciences”*, four types of relations or ties, namely *similarities, social relations, interactions, and flows*. Figure 2.3 illustrates these relations studied in social network analysis.



Similarities			Social Relations				Interactions	Flows
Location	Membership	Attribute	Kinship	Other role	Affective	Cognitive	e.g.,	e.g.,
e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	Sex with	Information
Same spatial and temporal space	Same clubs	Same gender	Mother of	Friend of	Likes	Knows	Talked to	Beliefs
	Same events	Same attitude	Sibling of	Boss of	Hates	Knows about	Advice to	Personnel
				Student of	etc.	Sees as happy	Helped	Resources
	etc.	etc.		Competitor of			Harmed	etc.
						etc.	etc.	

Figure 2.3. A typology of ties studied in social network analysis (Borgatti et al. 2009).

Similarities take place if a couple of nodes have the same attributes such as demographic features, attitudes, locations, or group memberships. These attributes are often examined in variable-based approaches. Network analysts often treat only group memberships as relations. For example, Mizruchi and Stearns (1988) have analyzed the structure of industries by investigating networks created by interlocking directorates.

According to Marin and Wellman (2011), s*ocial relations* encompass kinship or other types of commonly-defined role relations (e.g. friend, student); affective ties, which are based on network members’ feelings for one another (e.g. liking, disliking); or cognitive awareness (e.g. knowing). To give an example, Casciaro et al. (1999) examined how positive affectivity (liking) influences people’s perception of the patterns of social relationship around them.

Interactions, on the other hand, specify behavior-based ties like sex with, talking to, or inviting into one’s home. According to Marin and Wellman (2011) these mostly take place in context of social relations and interaction-based measures.

Flows are similarities established on exchanges or transfers among nodes. These may include relations in which resources, information, or influence flow through networks. According to Marin and Wellman (2011) flow-based relations frequently takes place within other social relations, and researchers often assume or study their co-existence.

Social network analysis became known as a vital technique in the study of human social behavior. It also became popular in fields such as marketing, communication studies, psychology, economics, biology, political science and information science. An example of such studies is that of Milgram (1967), who stated that the likelihood that any pair of actors (nodes) on the planet are separated by at most six degrees of separation. This paradox is also known as the "*small world"* phenomenon. The study of McPherson, Smith-Lovin & Cook (2001), *"Birds of a feather: Homophily in social networks"*, argues that a contact between similar people occurs at a higher rate than among dissimilar people. The term *"homophily"* was first used by sociologists in the 1950s to describe the phenomenon (Lazarsfeld & Merton, 1954).

Online social networks have made it possible to test these type of hypotheses. Leskovec and Horvitz (2008) proved in their report, *"Planetary-scale views on a large instant-messaging network"*, that the average path length between two MSN messenger users is 6.6 degrees. The availability of a wide variety of online data has made it possible to also verify other hypotheses such as *homophily* (McPherson, Smith-Lovin & Cook, 2001) or *shrinking diameters* (Leskovec, Kleinberg, & Faloutsos, 2005). In general, the availability of massive amounts of data in an online setting has given a new impetus towards a scientific and statistically robust study of the field of social networks (Aggarwal, 2011).

2.2.2 Online Social Networks

Online social networks have grown increasingly popular over the last decades and have attracted the attention of billions of social network users worldwide since their introduction. According to eMarketer in their report, *"Worldwide Social Network Users: 2013 Forecast and Comparative Estimates"*, nearly one in four people around the world will use social networks in 2013[2]. They also pointed out that the number of social network users worldwide increased 18% from 1.47 billion in 2012 to 1.73 billion social network users in 2013. It is estimated that by 2017 the total of users globally will account for 2.55 billion. Nielson and NM Incite emphasizes in their annually Social Media Report that there are two factors still driving the continued growth of social network sites, namely *mobile* and *proliferation*. Smartphones and tablets are being used more by people to access social network sites and new social network sites continue to emerge and catch on. While Facebook and Twitter carried on in 2012 to be among the most popular social network sites, Pinterest

[2] http://www.newmediatrendwatch.com/world-overview/137-social-networking-and-ugc

came out to be the social network site in 2012 with the largest year-over-year increase in both unique audience and time spent of any social network site across PC, mobile web, and apps (Nielsen & NM Incite, 2012).

As Ellison (2007) points out, social network sites are web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. Huberman, Romero & Wu (2008), on the other hand, stated that while the standard definition of a social network embodies the notion of all the people with whom one shares a social relationship, in reality people interact with very few of those “listed” as part of their network. The key factors that make social networks so unique is not that they let individuals meet strangers, but rather that they allow users to articulate and make their social networks visible (Ellison, 2007).

2.3 Social Contagion

As Latané (2000) points out, social contagion occurs when individuals change their behavior as a result of interaction with others. Other researchers have described social contagion as an actors’ adoption of behavior as a function of their exposure to other actors’ knowledge, attitudes, or behavior (Van den Bulte & Lilien, 2001). The theory behind the social contagion phenomenon is that information, ideas, and even behavior can spread through networks of people just the way that infectious diseases do[3].

The mechanisms of social contagion are the diverse social forces that make information, ideas, and behavior spread from one person to another. Social influence tends to be the mechanism most frequently related to social contagion. Nevertheless, there are other mechanisms besides social influence such as local information, social identity, social exclusion, homophily, and environmental factors that cause social contagion among individuals (Barash, 2011). This thesis will focus on social influence, homophily and environmental factors. In the subsequent sections we will describe these mechanisms.

2.3.1 Social influence

Social influence is possibly the most studied mechanism in the research field of social contagion. Social influence is described as the phenomenon that actions of a user can persuade his/her friends to behave in a similar way (Anagnostopoulos, Kumar & Mahdian,

[3] http://www.poptech.org/e1_duncan_watts

2008). An example of this scenario is when a user buys an iPod because one of his/her friends recently bought an iPod.

Negative influence is another type of influence, where users influence their friends to become less like them. The literature has studied this phenomenon much less, however it has received some attention in cases such as teenage rebellion and contentious relationships. Barash (2011) emphasized that negative influence does not play a strong role in social contagion, as individuals would be influenced to not adopt the contagious phenomenon their friends adopt.

There exists two main types of influence, namely social and interpersonal influence. Interpersonal influence is the ability of highly influential individuals to dictate their friends' behavior. Most of the time, all that is required is for one significant individual to purchase a product to effect his/her friends' behavior. Social influence, on the other hand, refers to the ability of groups to exert pressure on an individual. Thus, the impact of social influence on behavior changes increase with each individual that adopts the behavior.

An important aspect in social contagion research is the econometric identification of social influence. As Aral (2011) points out, there are two basic reasons for this. First, it is essential to estimate casual empirical effects of social influence to formulate effective social contagion management policies. While several studies proved the clustering of human behaviors amongst peers (e.g., Christakis and Fowler 2007, Crandall et al. 2008, Aral and Van Alstyne 2009), they do not show whether such behavioral clustering takes place due to social influence in order to formulate effective social contagion strategies. Iyengar, Van den Bulte & Valente (2011), on the other hand, assessed whether new product adoption was subject to social contagion operating through network ties such that better connected adopters exert more influence than less connected ones, and whether such contagion operates over and above the effect of targeted marketing efforts and system-wide influences that vary over time. The second reason highlighted by Aral (2011) is that causal empirical estimation is closely attached to our most fundamental perception of how one individual can influence the other. These definitions of social influence are being guided by the presumptions when social contagion is "taking place". One must recognize some essential presumptions regarding what it means for one individual to influence the other to estimate social influence and social contagion, and besides the causal structure of peer-to-peer induction in diffusion processes, there are also other similarly vital modeling choices and estimation strategies (Aral, 2011).

According to Iyengar et al. (2011) the most fundamental assumption of network marketing is that social influence or social contagion at work is among customers. Although this is regularly presumed, this does not always have to be the case. For example, many studies have found overestimated results of social contagion because of estimation difficulties or making use of theoretically over-determined models (Iyengar et al. 2011).

2.3.2 Homophily

As McPherson, Smith-Lovin and Cook (2001) point out in their report, *"Birds of a Feather: Homophily in Social Networks"*, homophily is the principle that a contact between similar people occurs at a higher rate than among dissimilar people. This is supported by Currarini and Vega-Redondo (2011), who emphasize that a pervasive feature of social and economic networks is that contacts tend to be more frequent among similar agents than among dissimilar ones. One common fact of homophily is that cultural, behavioral, genetic, and material information that flows through a social or economic network has a tendency to be localized (McPherson et al., 2001).

The basic idea here is that *birds of a feather may flock together*, meaning that we may choose friends who are more like ourselves, thus creating correlations in the behaviors amongst people that are connected in a social or economic network. This quote was first attributed to Robert Burton (1577 – 1640). However, before Robert Burton it was Aristotle who noted that people *"love those who are like themselves"* (Aristotle 1934, p. 1371). Before that it was Plato who observed that *"similarity begets friendship"* (Plato 1968, p. 837). Researchers frequently imply that modern social network analysis started with Jacob L Moreno's book, *"Who Shall Survive"* in 1934 (Alba, 1982; Freeman, White and Romney, 1992; Wasserman and Faust, 1994).

3 Modeling Music taste

Music taste can be described as the musical preferences of individuals, which is influenced by numerous factors when they select a specific song, album, or genre. Before the rise of the Internet it was not feasible to capture and analyze what individuals where listening to, and when they were listening to these songs, albums, or genres. Due to the fact that social music sites have made these type of data publicly available, this research contemplates to create a music taste model that summarizes which genres are being listened to by whom, and which social or external factors influence the genres individuals listen to.

3.1 Music taste drivers

Based on the literature reviewed in section two, we concluded that there are numerous factors involved that influences an individual's music taste (Levitin, 2011). For example, according to Rentfrow and Gosling (2003), people prefer to listen to music that reflects their specific personality characteristics. For this reason, we have chosen to use the genre categorizations of Rentfrow and Gosling (2003) for this research. These categorization will be further explained in detail in the following sections.

3.1.1 User demographics and summary statistics

Individuals listen to certain types of music mainly because they have greater preferences to some genres than others. Nevertheless, a few underlying factors to individual's music taste will have impact on the genres that they choose to listen to. For the purpose of this research, we have chosen to look at age and gender as user demographics. By analyzing these two factors we can also see whether homophily plays a role. As discussed in the previous chapters, homophily is the principle that a contact between similar people occurs at a higher rate than among dissimilar people. Thus, people of the same age or gender should have similar musical tastes.

3.1.1.1 Age

One underlying factor that should greatly influence music taste is age. For example, young people tend to listen to music such as rock, dance, hip-hop, or rap. On the other hand, you will not find many young people listening to genres such as classical, jazz, or blues. These genres tend to be listened to by older people with different musical tastes than younger people. As Harrison and Ryan (2010) point out in their report, *"Musical taste and ageing:*

Ageing and Society", music taste begins with fairly narrow tastes in young adulthood, and then expands into middle age, and narrows again later in life. He found this pattern in three separate surveys extending over 20 years. What was also remarkable was that the middle-aged and the oldest study groups did not like any of the genres that the youngest group liked.

3.1.1.2 Gender

Christenson and Peterson (1988) have suggested that for a variety of reasons, gender is central to the ways in which popular music is used and tastes are organized. They have found in their analysis that there is a key difference among males and females in how they "map" different music types. Significantly, it was noted by Roe (1984) that females tend to pay more attention to lyrics and listen to music to minimize the feelings of loneliness. He also emphasized that females, in contrast to males, like in general "pop hits" or mainstream music, folk, and classical music. Males, on the other hand, liked rock, hard rock, jazz, and harder forms of popular music. This is supported by Warner (1984), who argued that males prefer "macho/aggressive" styles of music, while females fancy for the most part romantic type of music.

3.1.2 External conditions

Previous studies have shown that weather seems to influence human behavior and consumer decision making in numerous ways (Murray, Di Muro, Finn & Popkowski Leszczyc, 2010). Studies in psychology, for example, hold the view that temperature greatly influences mood, and mood changes in turn cause behavioral changes (Cao & Wei, 2005). Despite the fact that the influence of weather on human behavior has been extensively investigated in fields like psychology and finance, the marketing field has not paid much attention in the past to the influence of weather. More recently, marketeers have started to incorporate weather variables in their models to, for example, predict sales. For instance, in June 2006 Wal-Mart had reduced its sales forecast due to an abnormal cool weather in the summer that could have resulted in an unfavorable sales of air conditioners, and swimming pool supplies (Murray et al., 2010).

Howarth and Hoffman (1984) established in their study, *"A multidimensional approach to the relationship between mood and weather"*, that humidity, temperature, and sunshine duration had the greatest effects on mood. For this reason, we have chosen to use these variables in our study to analyze whether the weather has influence on the genres that individuals choose to listen to.

Even though we did not include mood characteristics of individuals in our analysis, it would be fascinating to evaluate whether there is a direct correlation between external conditions and music taste of individuals. Our research included three external conditions, namely weekly average temperature, weekly cloud coverage, and the weekly sunshine duration.

These indicators were derived from the KNMI weather stations. For example, the minimum and maximum of the weekly average temperature was -2,87°C and 20,89°C, with a mean of 10,54°C during the year of 2009. The cloud coverage is measured in an ordinal scale from 0 to 8, where 8 indicates that the sky is totally invisible. After computing the weekly cloud coverage, we ended up with a minimum of 2,29 and a maximum of 7,57 with a mean of 5,19. This is due to the fact that we computed the average cloud coverage for an entire week. The last external condition was the average sunshine duration. This variable is measured on an hourly basis of global radiation per day. The minimum of weekly sunshine duration was 0,86 and the maximum 11,1 with a mean of 5,16. The histogram, normal q-q plot, detrended normal q-q plot, and boxplot of all these three external conditions can be found in Appendix A.

3.1.3 Social influence

Social influence is described as the phenomenon that actions of a user can persuade his/her friends to behave in a similar way (Anagnostopoulos, Kumar & Mahdian, 2008). There exists two main types of influence, namely social and interpersonal influence. Interpersonal influence is the ability of highly influential individuals to dictate their friends' behavior. Most of the time, all that is required is for one significant individual to purchase a product, to induce a friend's behavior. Social influence, on the other hand, refers to the ability of groups to exert pressure on an individual. Thus, the impact of social influence on behavior changes increase with each individual that adopts the behavior.

For the purpose of our research, we assume that friends past listening behavior will have a positive correlation on which genres the individuals will listen to in the future. We have decided to look at what the friends have listened to both one week and two weeks in the past. By looking at one week and two weeks in the past we can determine whether there is a decay effect if there is a correlation between the genres that friends listen to and the individuals' music taste.

3.2 Music genres

For the purpose of this research we have decided to analyze which factors impact the music genres that individuals choose to listen to. These factors include, as mentioned in the previous sections, user characteristics, external conditions and social influence. We presume that these factors should model the music taste of individuals.

For example, Rentfrow and Gosling (2003), explored individual differences in music preferences by investigating not only the lay beliefs about music, but also the structure underlying music preferences, and links between music preferences and personality. Results from the music preferences of 3500 individuals converged to show 4 music-preference factors, namely *Reflective and Complex, Intense and Rebellious, Upbeat and Conventional,* and *Energetic and Rhythmic*. The *Reflective and Complex* factor comprised of genres that appear to facilitate introspection and are structurally complex such as blues, jazz, classical, and folk music. The *Intense and Rebellious* factor, on the other hand, consisted of genres that are full of energy and emphasized themes of rebellion such as rock, alternative, and heavy metal. Then again, the *Upbeat and Conventional* factor is composed of genres that emphasize positive emotions and are structurally simple like country, sound track, religious, and pop music. The fourth and last factor named *Energetic and Rhythmic* included genres that are lively and often emphasize the rhythm, some of which are: rap/hip-hop, soul/funk, and electronica/dance music.

We have decided to use these 4 categorizations as the main genres. The reason for this was due to the fact that there exists a lot of genres and these genres also consists of sub genres. Also another reason is that songs, albums, and artists sometimes fall into different genres making it difficult to classify them in the right genres. By using Rentfrow and Gosling´s music-preference factors makes it much easier to classify and quantify the genres individuals listen to.

4 Research methodology

This chapter will describe the research methodology used in the study. The source of the data, the study design and the population and sample will be described. The techniques used to collect the data and analyze the data will also be described.

4.1 Hypothesis development

This research is focused on examining the relationship between social influence and music taste. We will determine empirically whether social influence and external conditions influence music taste. With the aim of accomplishing the proposed research objectives, the following research question was formulated:

RQ1: *Does social influence or external conditions play a role in influencing ones music taste?*

According to Rubin and Rubin (2005, p. 40), research questions are better presented as hypothesis. They argue that hypothesis are statements that imply how to or more concepts or underlying ideas are related. Based on the literature reviewed in section two, the modeling choices in section three, and RQ1, the following hypothesis were formulated and will be tested:

H1: Music taste is not influenced by what your peers have listened to;

H1a: Music taste is not influenced by what your peers have listened to one week ago;

H1b: Music taste is not influenced by what your peers have listened to two weeks ago;

H3: Music taste is not influenced by external conditions;

H4: Music taste is not influenced by playcounts per week.

4.2 Methodology

We have decided to employ four multiple linear regression model to capture the relationship between music taste and what influences it. The base model employed for this research is as following:

$$\begin{aligned} GenreX_{it} \; = \; & \beta_0 + \beta_1 . age_i + \beta_2 . gender_i + \beta_3 . average\_cloud\_coverage_t \\ & + \beta_4 . average\_temperature_t + \beta_5 . average\_sunshine\_duration_t \\ & + \beta_6 . friends\_genreX_{i,t-1} + \beta_7 . friends\_playcount\_week_{i,t-1} \\ & + \beta_8 . friends\_genreX_{t-2} + \beta_9 . friends\_playcount\_week_{t-2} \end{aligned}$$

$GenreX_{it}$ = The counts of a hyper genre of music listened by $user_i$ at $week_t$
i = The number of users in the sample (i = 1, 2, ..., 1295).
t = The number of weeks in the particular year (t = 1, 2, ..., 53).

Dependent variables

These are the four hyper genres we have used to classify each track the users have listened to in a particular week. These are the counts of a genre of music listened by each user in a particular week. As mentioned before, we have decided to use Rentfrow and Gosling’s (2003) 4 music taste factors, namely *Reflective and Complex, Intense and Rebellious, Upbeat and Conventional,* and *Energetic and Rhythmic*. The *Reflective and Complex* factor comprised of genres that appear to facilitate introspection and are structurally complex such as blues, jazz, classical, and folk music. The *Intense and Rebellious* factor, on the other hand, consisted of genres that are full of energy and emphasized themes of rebellion such as rock, alternative, and heavy metal. Then again, the *Upbeat and Conventional* factor is composed of genres that emphasize positive emotions and are structurally simple like country, sound track, religious, and pop music. The fourth and last factor named *Energetic and Rhythmic* included genres that are lively and often emphasize the rhythm, some of which are: rap/hip-hop, soul/funk, and electronica/dance music.

Independent variables

The independent variables used in this research to establish what influences music taste are: user’s age, user’s gender, average cloud coverage per week, average temperature per week, average sunshine duration per week, friends genres listened to one week ago, total friends playcount one week ago, friends genres listened to two weeks ago, and total friends playcount two weeks ago.

The user's age is a self explanatory variable. During the course of the year 2009, the user's age do not change because the data was collected at one point in time. The gender variable is a dummy variable, being that male = 1 and female = 0. The user's total playcount for each week is a sum of the 4 main genres played for that particular week. The average cloud coverage, temperature, and sunshine duration per week has been collected from the KNMI database and we have computed these variables for each week of the year 2009. For the sake of proving whether social influence plays a role on what individuals listen to, we have decided to analyze what the individual's friends have listened to one week ago, and also two weeks ago. We have done the same with the total friends playcount.

The following figure illustrates how the model looks like and what are the dependent and independent variables:

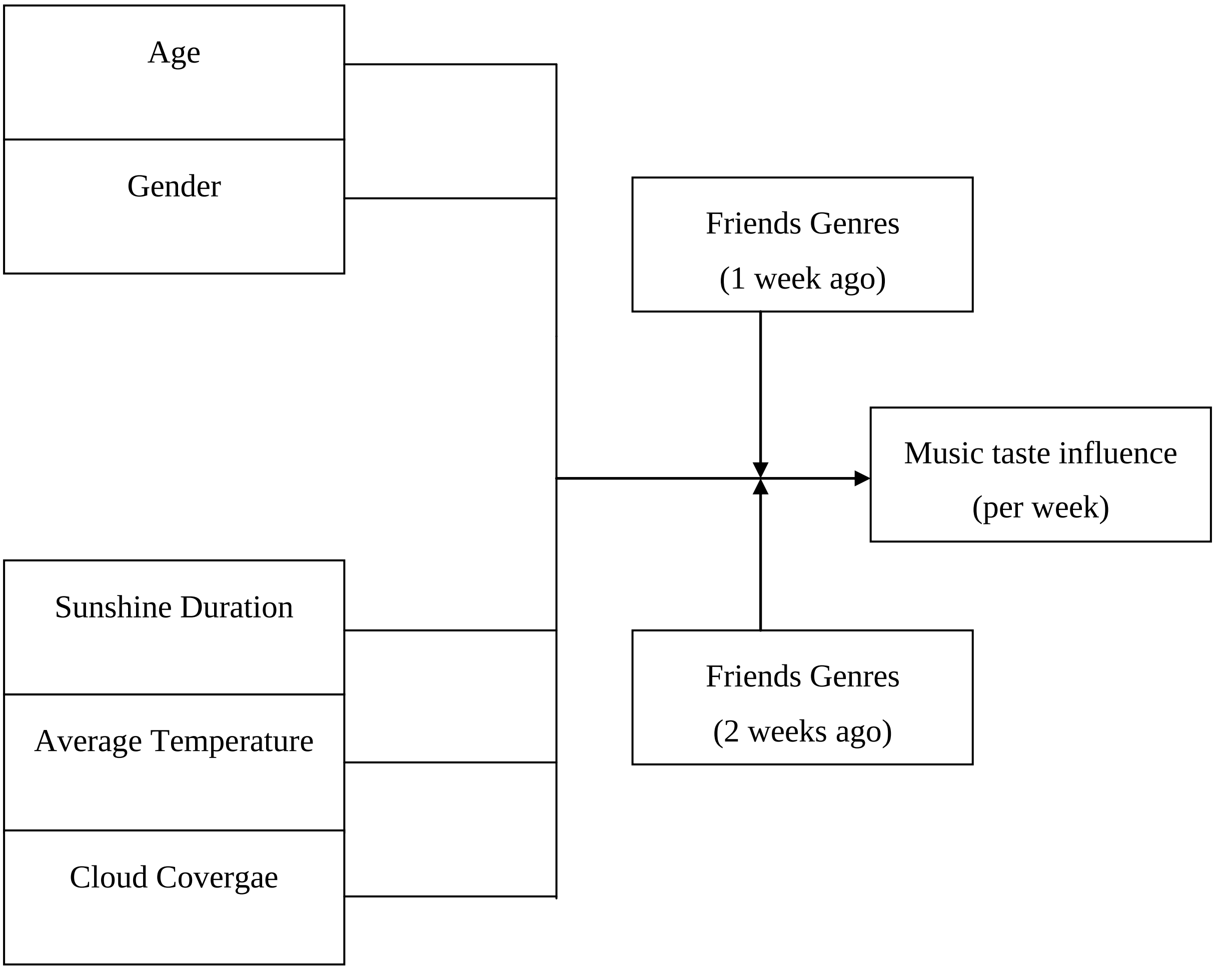


Figure 4.1 Model for predicting music taste.

4.3 Source of the research

Last.fm is a social music site, established in the UK in 2002. It allows its users to listen to radio stations that are based on their preferred genre, artists, albums, and tracks played. However, Last.fm became really popular for its *"Audioscrobbler"* software. This is a music recommender system that establishes a comprehensive profile of every single Last.fm user's musical taste. It accomplishes this by recording every track a user has listened to via the internet radio stations, the user's personal computer or portable device. The information transmitted to Last.fm's database, which is called a *scrobble*, is utilized to recommend the user new artists and songs they might enjoy. Last.fm also automatically generates charts and statistics for its users on the Top Artists, Top Albums, and Top Tracks they have most frequently listened to. These charts and statistics are presented on the user's profile, which can be shared among users to promote their musical taste.

The ability to track what users have listened to and when offers a great opportunity for social scientists to study the user's listening behaviors and tastes. Besides presenting a user's Top Artists, Top Albums, and Top Tracks, Last.fm also calculates how compatible their music taste is to another user. These users are called neighbors, which are other people on Last.fm who have the most similar taste to a specific user. In addition, these users can become friends with one another. Users can also create groups with other users who have something in common like a genre of music, or fans of an artist. Last.fm also allows its users to tag artists, albums, and tracks. Thus, creating site-wide folksonomy of music, which makes it much easier to search and discover new music. The tagging of music can be classified by either genre, mood, artist characteristic, or any other user-defined categorization.

4.4 Data Collection

Last.fm has reached more than 50 million active users since its release in 2002 (Skilledtest, 2012). The scope of this paper focuses on Dutch Last.fm user's listening behavior and habits and how this is influenced by peers. Therefore, a sample of these users and their listening history was used for this study. This is why the 50 million users are mostly not relevant anymore.

A web spider was created to systematically crawl each user profile on Last.fm and save all Dutch user profiles it encounters. The web spider ended up with a list of 18.050 Dutch users. For each of these users, their friends were extracted resulting in a list of 104.428

users and friends. As mentioned before, we determined to work only with a sample of the dataset.

We discarded all the non-Dutch users and decided to maintain Dutch users who had between 5000 and 10.000 scrobbles since they registered. This guarantees that there is a small variance among users, thus making the comparison between them much reliable. This resulted in a list of 1295 users and their listening history for the year 2009, which comprised of 6.722.166 scrobbles. For the purpose of this research, we have chosen to only use data of the listening history of the users in 2009. After doing an extensive analysis, we came to the conclusion that in this year the most users were listening more actively and constantly to Last.fm. As it can be seen in the following section, after applying the chosen restrictions, we were left with 3.687 users. We extracted all of the listening history of these 3.687 users since they had registered. The plan was to analyze their listening behavior from 2009 until 2013. However, when we started to analyze the data we could see that a lot of users did not use Last.fm actively all those five years. This gap of no listening behaviors would for certain create bias in our research. Thus, we started to look at which period(s) had the least gap of no listening behaviors and which period(s) still had the most users left. At the end of this analysis we had chosen the year 2009 which had the least amount of no listening behaviors and still had 1295 users left, which was still quite a substantive amount of users for our research.

After we collected each user's listening history, we still had to collect the item tags in order to categorize which genres the users were listening to. Tags are keywords that can be assigned to an item, which best describes this item and makes it easy to search and find the item again. For our research we decided to collect the top 10 artist tags because in our opinion this would better describe which genre belongs to each artist listened to.

In sum, the following information was collected for each user chosen for this study:

- The complete information on each user, which included their username, real name, country, age, gender, registration date, number of play counts, and their profile image;
- A complete list of their friends and information of these friends;
- A comprehensive list of the listening history of each user, from 01 January 2009 to 31 December 2009;
- For each artist listened to, a list of top 10 tags in order to be able to categorize each listening to genres based on the corresponding artist's tags.

4.5 Sample & Data

Data was collected over a period of several months on 104.428 Last.fm users. The sample users, which consisted of 1295 Dutch Last.fm users, had listened to over 19 million tracks since 2002. For this study we focused on tracks listened to from 01 January 2009 to 01 January 2010, which still was a comprehensive size of tracks. The tracks listened to in 2009 were over 6 million tracks. Table 4.2 illustrates the sample construction and the filters, which resulted in a sample of 1295 users.

Table 4.2 Sample construction

	Criteria	Eliminated	Sample
1	All entries		104.428
2	Country = NL	(67.608)	36.820
3	Playcount BETWEEN 5000 AND 10.000	(32.404)	4.416
4	Gender not null	(28)	4.388
5	Age BETWEEN 15 AND 77	(701)	3.687
6	YEAR 2009	(2392)	1295

Let us take a look at some descriptive statistics of the users and the predictors. We start by looking at the friend relations left after applying all of these restrictions. There are still 174 friend relations left among the 1295 users of our sample data. As it can be seen in table 4.3 most of our sample users have no friends. This group is 89,1 % of our sample users. Most of our sample users have 1 friend, which is 9,7% of our sample users. There is one user with the highest number of friends, twelve.

Table 4.3 Friend relations among sample users

Users	**# of friends**
1154	0
126	1
9	2
3	3

1	4
1	5
1	12

For this research we have only extracted users who have made their age publicly available on their Last.fm profile. Nevertheless, there were users who claimed to be either younger than 5 years old or older than 100 years old. Therefore, we decided to filter out these users and only take users into account that were between the age of 15 and 77 years old, with a mean of 29,2. In order to get a feeling how the distribution of the sample population's age was, we created the following histogram:

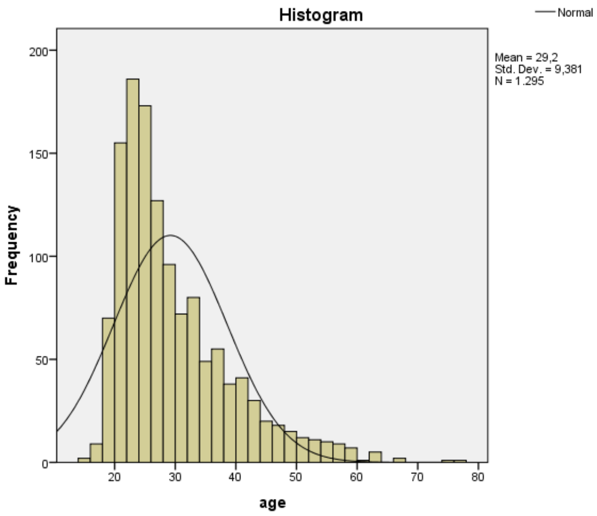
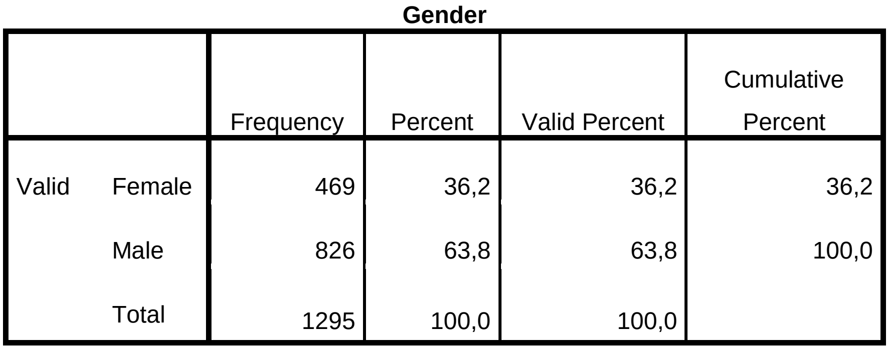


Figure 4.4 User's age histogram.

As it can be seen in the histogram, the majority of the users were between the age of 20 and 30. After analyzing this histogram, one could say that this histogram is right skewed. The proper way to analyze whether this histogram is right skewed is by looking at mode, median, and mean. Thus, for a right skewed distribution, mode < median < mean. In this case mode = 24, median = 26, and the mean = 29,2. This has proved that most of the data falls to the left of the mean, resulting in a right skewed distribution. We have also computed the log of age in order to get a better distribution. The histogram and descriptive statistics can be found in Appendix A.

Table 4.5 depicts the percentages between female and male users. As it can be noted, there were more male users than female users. The population comprised of 63,8% male users and merely 36,2% female users.

Table 4.5 User’s gender descriptive statistics



Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	469	36,2	36,2	36,2
	Male	826	63,8	63,8	100,0
	Total	1295	100,0	100,0	

4.6 Statistical methods

In this section we will briefly discuss the statistical methods used for this study. We have chosen to use two statistical methods, namely multiple linear regression model and panel data analysis.

4.6.1 Multiple linear regression model

The notion of “regression” was first proposed by Francis Galton (1986) who examined the relationship between fathers’ and sons’ heights. He noted in his research that there was no relationship between sons’ and their fathers’ heights but rather that the sons’ heights “regress to” the mean of the population. This notion in combination with the establishment of Carl Friedrich Gauss’ (Myers, 1990) method of least squares procedures became a widely used statistical technique. This technique was called multiple regression analysis employing ordinary least squares procedures (OLS) to investigate relationships between variables.

As Baker (2006) points out, multiple regression is a regression with two or more independent variables on the right-hand side of the equation. He also noted that one must use multiple regression if more than one cause is associated with the effect one wishes to understand. There are two reasons why one would use multiple regression, namely for prediction and explanation.

The multiple linear regression model can be specified as following:

$y = \beta_0 + \beta_1.X_{1i} + \beta_2.X_{2i} + \ldots + \beta_k.X_{ki} + \varepsilon_i$ (i = 1, 2, ... , n)

According to Ethington, Thomas & Pike (2002), that the outcome y is produced by two components. They noted that the first component defines the "best" linear relationship between the outcome (y) and the predictors ($\beta_0 + \beta_1.X_{1i} + \beta_2.X_{2i} + \ldots + \beta_k.X_{ki}$). They go about by arguing that for any given level of X, there is a corresponding "predicted" level of y, (E[y] = $\beta_0 + \beta_1.X_{1i} + \beta_2.X_{2i} + \ldots + \beta_k.X_{ki}$). The second component they mention is ε_i, the stochastic or random source of variation. It is a random variable for which outcomes are governed by a probability distribution and Ethington et al., (2002) summarized the Gauss-Markov assumptions to consist of the following properties:

1. E(ε_i) = 0, the mean of the error term is always equal to 0;
2. Var(ε_i) = σ^2, the variance of the error is the same at any level of X, (homoschedasticity);
3. Cov(ε_i, ε_j) = 0, the error terms for any two observations are uncorrelated (independence);
4. E($\varepsilon_i|X_{1i},X_{2i},\ldots, X_{Ki},$) = 0, all explanatory variables are uncorrelated with the error term (exogeneity);
5. ε_i is normally distributed.

4.6.2 Panel data analysis

When one uses *panel data* or also known as *longitudinal data*, they are interested in studying subjects over time as opposed to regression data. Time series data focuses on the same subject over time as opposed to panel data which focuses on many subjects over time. According to Maddala (2001) panel (data) analysis is a statistical method that measures two-dimensional (cross-sectional/time-series) panel data, which is most commonly used in social science, epidemiology, and econometrics.

The general notation for a basic regression prediction equation is as following:

$\boldsymbol{Y_{it} = \beta_0 + X_{it}\beta + Z_i y + \alpha_i + \varepsilon_{it}}$

Where i and t are indices of individuals and time.

$\boldsymbol{\beta_0}$: the constant term of the formula;

$\boldsymbol{X_{it}\beta}$: Observed variables and can be estimated by both fixed and random effects models, they are time-variant factors;

$\boldsymbol{Z_i y}$: Observed variables which can NOT be estimated directly by fixed model but can be estimated by random effects model (time-invariant factors);

$\boldsymbol{\alpha_i}$: These are the un-observed individual specific effect, a fixed value for each individual across repeated measures;

$\boldsymbol{\varepsilon_{it}}$: Un-observed random error terms (residuals).

When discussing whether to use fixed effects or random effects model, the issue is how to deal with the un-observed individual specific effect (α_i). There are two strategies that one can use to deal with this, that are basically assumptions that one must make about the un-observed individual specific effect.

The first assumption is the fixed effect assumption, which states that the individual specific effect is correlated with the independent variables. Hence, time-invariant factors will be excluded from the model by taking the difference between each observation with the within-group mean values in order to get rid of the individual specific effect term α_i. Thus, in a fixed effect model, α_i and $\mathrm{Z}_i y$ will be excluded from the model and therefore,

$$E(\alpha_i \mid X_{it}, \mathrm{Z}_i) \neq 0.$$

Another type of model or assumption that can be made is the random effect model. The random effect assumption is that the individual specific effects are uncorrelated with the independent variables. Coefficients of time-variant as well as time-invariant variables will be estimated. In a random effect, there is no fixed individual specific effect and therefore,

$$E(\alpha_i \mid X_{it}, \mathrm{Z}_i) = 0.$$

From this it follows that α_i and ε_{it} can be combined together to form a new error term ε_{it} = α_i + ε_{it}. Thus, it is not necessary to take the difference here, and all variables, time-variant and time-invariant, will be included in the model.

5 Results

This chapter will present the results of the analysis we have employed on the Last.fm data set. We have a total of six models where we try to understand what influences individuals music tastes. For each model, we started with a multiple linear regression analysis to estimate the relationship among several variables. For these analysis we treated the data as pooled data. However, we know that the data was categorized as panel data. In order to account for the individual effects and time effects, we also did panel data analysis for each model. Thus, for each model we will present first the results of the multiple linear regression analysis followed by the results of the panel data analysis. Furthermore, for all our models, we did a collinearity diagnostics to check for multicollinearity. In order to check whether users listen the most or the least to a particular genre when they are young or old, we decided to add the variable Age² to see if there is a quadratic effect. However, we had to drop Age² from all our models because the variance inflation factors were higher than 10, indicating that we had a standard error problem. According to O'brien (2007) the variance inflation factor and tolerance are both widely used measures of the degree of multi-collinearity of the $_i$th independent variable with the other independent variables in a regression model. O'brien (2007) noted that tolerance for the $_i$th independent variable is 1 minus the proportion of variance it shares with the other independent variable in the analysis (1 - R_i^2). He also argued that the variance inflation factor (VIF) is the reciprocal of tolerance: 1/ (1 - R_i^2) (O'brien, 2007). In order to test whether there is a quadratic effect, we will perform some additional tests with Age² and some other independent variables of interest to show whether individuals listen the most or the least to certain genres in their middle years.

Model 1

Model 1 starts by looking at what variables influence the users' playcount per week. We have transformed this variable into ln(playcount_week + 1), because the playcount per week contained a lot of zeros and extremes. The independent variables used in this model are ln(Age), Gender, Average temperature per week, Average Cloud Coverage per week, and Average Sunshine Duration per week. We also filtered out all observations that had a playcount per week equal to zero, because this would create bias in our results. A playcount equal to zero would mean that the respondent has not used Last.fm in that particular week. We started with 68635 observations (1295 users x 53 weeks equals 68635 observations) prior to the filter and ended up with 37596 observations after.

The degree of explanation for model 1 of playcount per week was $R^2 = 0.01$ and $F(5, 37590) = 89.90$ with a *p*-value smaller than 0.001 ($p < 0.001$). We see for the variable ln(Age) significant results. Age has a significant negative influence on playcount per week, with a coefficient of -0.04 ($p < 0.001$). On average, the effect of an increase of 1 year of age, would result in -18%[4] less playcounts per week. Thus, it can be said that older people listen less often to Last.fm each week. The variable gender is also significant. It can be concluded that males listen on average 13% less than females. The difference of gender is significant for playcount per week with a coefficient of -0.05 ($p < 0.001$).

The external factors influencing playcount per week were the temperature, cloud coverage, and sunshine duration. The variable average temperature per week has a negative effect on playcount per week. The increase of 1 degree Celsius in a week, would result on average in -3% less playcounts per week. The average temperature per week has a significant influence on the play count per week with a coefficient of -0.12 ($p < 0.001$). On the other hand, cloud coverage has a positive effect on playcount per week. The increase of cloud coverage per week, results on average in 7% more playcounts per week. Cloud coverage has also a significant influence on playcount per week, with a coefficient of 0.08 and $p < 0.001$. The average sunshine per week has also a significant influence on the playcount per week, with a coefficient of 0.14 ($p < 0.001$). For an increase of 1 hour per week of sunshine there is on average 6% more playcounts being listened per week.

Model 1 (Pooled)			
	Ln(Playcount_week)		
Variable	***B***	***SE B***	***Stand. β***
Ln(Age)	-0.18	0.02	-0.04***
Gender (male=1)	-0.13	0.02	-0.05***
Average Temperature	-0.03	0.00	-0.12***
Average Cloud Coverage	0.07	0.01	0.08***
Average Sunshine Duration	0.06	0.01	0.14***
R²	0.01		
Adjusted R²	0.01		
F	89.90***		
N	37596		

[4] This percentage increases are used throughout the model for the ease of presentation although they do not correspond exactly to the true percentage increase from the model (this has to do with the logs transformed variables).

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$;

Let us look now at the results of the panel data analysis for the same model. This model has also as dependent variable playcount per week and as independent variables ln(Age), Gender, Average Temperature per week, Average Cloud Coverage per week, and Average Sunshine Duration per week. In the table below you can see that for the panel data analysis of this model we execute a fixed effect model and a random effect model. After getting the results we employ a Hausman Test to check which model is appropriate.

The Hausman test indicates that for this model the fixed effect model should be used. A low *p*-value indicates that a fixed effects model is more appropriate for this model. Here we can see that the average temperature per week has a negative effect (-2.7%) on the playcount per week and it is significant with a $p < 0.001$. On the other hand, average cloud coverage per week has a positive effect on playcount per week with a $p < 0.001$. An increase in cloud coverage per week, results on average in 7.5% more playcounts per week. Average sunshine per week has also a significant influence ($p < 0.001$) on playcount per week. This would mean that 1 hour extra of sunshine per week is on average 6.3% more playcounts per week.

Model 1 (Panel)				
Ln(Playcount_week)				
	Fixed effect		**Random effect**	
Variables	***B***	***SE B***	***B***	***SE B***
Ln(Age)	-	-	-0.227	0.068**
Gender (male = 1)	-	-	-0.109	0.041**
Average Temperature	-0.027	0.002***	-0.027	0.002***
Average Cloud Coverage	0.075	0.008***	0.073	0.008***
Average Sunshine Duration	0.063	0.005***	0.061	0.005***
F(3, 36298)	103.63***			
Wald chi2 (5)			343.63***	
Hausman Test (Prob>chi2)	0.001			
N	37596		37596	

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$;

Model 2

In model 2 we will start looking at music taste. For model 2 we run four separate multiple regressions each having one dependant variable. The 4 dependant variables for these regressions are ln(Genre_1 + 1), ln(Genre_2 + 1), ln(Genre_3 + 1), and ln(Genre_4 + 1). Here we transformed these four genres to Ln(X + 1) because the genres contained a lot of zeros and extremes. This would smooth our data and get rid of the long tail in our variables. Each hyper genre is the count of a genre of music listened by a user in the particular week. The independent variables used in this model are ln(Age), Gender, Average temperature per week, Average Cloud Coverage per week, and Average Sunshine Duration per week. We also filtered out all observations that had a playcount per week equal to zero, similarly to model 1. We started with 68635 observations (1295 users x 53 weeks equals 68635 observations) prior to the filter, and ended up with 37596 observations after filtering.

We see for the variable age that only two of the four hyper genres is significant, namely hyper genre_1 and hyper genre_2. Age has a significant influence on the total plays per week of hyper genre_1, with a coefficient of 0.16 ($p < 0.001$). On average, the increase of 1 year of age would mean 86% more listening of hyper genre_1. This was expected being that hyper genre_1 consists of genres such as blues, jazz, classical, and folk that tend to be listened to by older people. On the other hand, age has a negative effect on the total plays per week of hyper genre_2, with a coefficient of -0.12 ($p < 0.001$). The effect of an increase of 1 year of age, would result on average in -56% less listening of hyper genre_2. Hyper genre_2 tends to be listened to by younger people than older people. Thus, the older people become the less they are going to listen to genres such as rock and heavy metal. Neither hyper genre_3 nor hyper genre_4 were significant and both had a very small effect.

On the other hand, the variable gender was significant for all the genres. It can be concluded that males listen on average less to hyper genre_1 (-42%) and hyper genre_3 (-42%) than females. This difference of gender is significant for hyper genre_1 with a coefficient of -0.13 ($p < 0.001$) and hyper genre_3 with a coefficient of -0.14 ($p < 0.001$). Males listen also less to hyper genre_2 (-11%) than females. This difference of gender is significant for hyper genre_3, however the influence is smaller than the other hyper genres with a coefficient of -0.04 ($p < 0.001$). Hyper genre_4 is the only hyper genre that we can conclude that males listen on average 5% more than females. So, gender has a significant influence on the total plays per week of hyper genre_4, with a coefficient of 0.02 ($p < 0.001$).

The average temperature per week has a significant influence on all the four hyper genres. It has also a negative effect on all the hyper genres. An increase of the average temperature per week of 1 degree Celsius, results on average in -3% less total plays of all the hyper genres per week. The hyper genre with the highest negative influence of the average temperature per week was hyper genre_2, with a coefficient of -0.11 (p < 0.001). All the other hyper genres had a coefficient of -0.010 (p < 0.001). Thus, it could be that when it is hotter people in the Netherlands tend to be more outside and less behind their computers listening to music.

The average cloud coverage has also a significant influence on all the four hyper genres. However, in contrast to the average temperature per week, average cloud coverage per week has a positive effect on all the four hyper genres. On average there is 7% more total plays per week for hyper genre_2 and hyper genre_4, if there would be an increase in cloud coverage per week. For hyper genre_3 this would be 6% and for hyper genre_1 this is 5%.

There is a significant influence of the average sunshine duration per week on all the four hyper genres. The hyper genre with the highest influence of average sunshine duration per week was hyper genre_2, with a coefficient of 0.13 (p < 0.001), followed by hyper genre_4 with a coefficient of 0.12 (p < 0.001) and hyper genre_3 (coefficient = 0.10; p < 0.001). The hyper genre with the least influence of average sunshine duration was hyper genre 1 (coefficient = 0.08; p < 0.001). An increase of 1 hour of average sunshine duration per week, would result on average in 6% more total plays for hyper genre_2 and hyper genre_4, followed by hyper genre_3 (5%) and hyper genre_1 (4%).

Model 2 (Pooled)												
Variable	**Ln(Genre 1 + 1)**			**Ln(Genre 2 + 1)**			**Ln(Genre 3 + 1)**			**Ln(Genre 4 + 1)**		
	B	***SE B***	***Stand. B***	***B***	***SE B***	***Stand. β***	***B***	***SE B***	***Stand. β***	***B***	***SE B***	***Stand. β***
Ln(Age)	0.86	0.03	0.16***	-0.56	0.03	-0.12***	0.04	0.03	0.01	-0.001	0.03	0.00
Gender (male = 1)	-0.42	0.02	-0.13***	-0.11	0.02	-0.04***	-0.42	0.02	-0.14***	0.05	0.02	0.02***
Average Temperature	-0.03	0.00	-0.10***	-0.03	0.00	-0.11***	-0.03	0.00	-0.10***	-0.03	0.00	-0.10***
Average Cloud Coverage	0.05	0.01	0.05***	0.07	0.01	0.08***	0.06	0.01	0.06***	0.07	0.01	0.07***
Average Sunshine Duration	0.04	0.01	0.08***	0.06	0.01	0.13***	0.05	0.01	0.10***	0.06	0.01	0.12***
R²	0.04			0.02			0.02			0.01		
Adjusted R²	0.04			0.02			0.02			0.01		
F	295.87***			174.04***			176.42***			38.65***		
N	37596			37596			37596					

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$;

The following table presents the results of the panel data analysis for model 2. For this model we executed both fixed effects and random effects model for all the hyper genres. The independent variables are ln(Age), Gender, Average Temperature per week, Average Cloud Coverage per week, and Average Sunshine Duration per week. After executing the fixed effects and random effects model, we check which is more appropriate by performing a Hausman Test.

The Hausman Test shows that for this model it is more appropriate to use the fixed effects models for all four hyper genres. The *p*-value of the Hausman test was the lowest for hyper genre_1 ($p < 0.001$) and hyper genre_3 ($p < 0.001$), followed by hyper genre_4 ($p < 0.01$) and hyper genre_2 ($p < 0.01$). A low *p*-value when executing a Hausman Test tells you that a fixed effects model is more appropriate.

Here we can see just as in the multiple regression analysis that average temperature per week has a negative effect on the total plays for all the hyper genres. The highest negative effect is for hyper genre_2 (-2.7%), which is followed by hyper genre_3 (-2.5%) and hyper genre_1 (-2.4%). The hyper genre with the lowest negative effect was hyper genre_4 (-2.3%). Average temperature per week has a significant influence on all four hyper genres ($p < 0.001$).

The average cloud coverage has a significant influence on the total plays per week for all four hyper genres ($p < 0.001$). On average, the effect of 1 hour of extra of cloud coverage per week, gives 4.9% extra listens to hyper genre_1. For hyper genre_2 this would be on average 7.5%, and for hyper genre_3 this is 6.5% extra plays per week. Last but not least, for hyper genre_4 this would mean 7.4% extra plays per week.

For the average sunshine duration we see the same results what we already concluded when we executed the multiple regression analysis. Average sunshine duration has a positive effect on the total plays per week for all the hyper genres. The highest effect of average sunshine duration was on hyper genre_4 (6.5%) and for hyper genre_2 (6.4%), followed by hyper genre_3 (5.5%) and hyper genre_1 (4.1%). Average sunshine duration has a significant influence on all the four hyper genres ($p < 0.001$).

Model 2 (Panel)																
	Ln(Genre 1 + 1)				Ln(Genre 2 + 1)				Ln(Genre 3 + 1)				Ln(Genre 4 + 1)			
	Fixed effect		Random Effect		Fixed effect		Random Effect		Fixed effect		Random Effect		Fixed effect		Random Effect	
Variable	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***
Ln(Age)	-	-	0.817	0.095***	-	-	-0.597	0.076***	-	-	-0.036	0.090	-	-	-0.035	0.095
Gender (male = 1)	-	-	-0.389	0.057***	-	-	-0.080	0.046	-	-	-0.390	0.054***	-	-	0.059	0.057
Average Temperature	-0.024	0.002***	-0.024	0.002***	-0.027	0.002***	-0.027	0.002***	-0.025	0.002***	-0.025	0.002***	-0.023	0.002***	-0.024	0.002***
Average Cloud Coverage	0.049	0.009***	0.048	0.009***	0.075	0.008***	0.073	0.008***	0.065	0.009***	0.064	0.009***	0.074	0.009***	0.073	0.009***
Average Sunshine Duration	0.041	0.006***	0.040	0.006***	0.064	0.005***	0.062	0.005***	0.055	0.006***	0.054	0.006***	0.065	0.006***	0.064	0.006***
F(3, 36298)	80.26***				102.15***				82.21***				72.78***			
Wald chi2 (5)			349.66***				382.82***				308.94***				226.72***	
Hausman Test (Prob>chi2)	0.001				0.0045				0.001				0.002			
N	37596		37596		37596		37596		37596		37596		37596		37596	

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$;

Model 3

In model 3 we included the friends' genres for each individual. These variables were included to see whether peers had an influence on what individuals listened to. We decided to check if what your friends have listened to the previous week would have an influence on what one would listen the current week. For this model we added the independent variables $Ln(friends\_genre1)_{t-1}$,$Ln(friends\_genre2)_{t-1}$,$Ln(friends\_genre3)_{t-1}$,and $Ln(friends\_genre4)_{t-1}$. Thus, for each hyper genre we compared them with their friends' genres of the previous week. These four friends' genres were also transformed to Ln(X + 1) similarly to model 2. Like in the prior models, we also filtered out all observations that had a playcount per week equal to zero and the friends' playcount per week that equaled zero. We started with 68635 observations (1295 users x 53 weeks equals 68635 observations) prior to the filter and ended up with 2479 observations after the filtering.

Model 2 has already shown us how age, gender, and the weather variables effect the total plays per week for each hyper genre. Let us focus now in model 3 on whether the friends influence what the individuals listen to in a particular week. The friends' genre 4 has a significant influence on the total plays per week on hyper genre_4 of the individuals, with a coefficient of 0.08 ($p < 0.001$). On average, the effect of 1 extra play of genre 4 per week of the friends, gives 7% extra plays of hyper genre_4 for the individuals. Friends' genre 1 has also a significant influence on the total plays per week on hyper genre_1 of the individuals, with a coefficient of 0.04 ($p < 0.05$). An increase of 1 play per week of genre 1 of the friends, would result on average in 4% extra listens of hyper genre_1 for the individuals. Neither hyper genre_2 nor hyper genre_3 had a significant influence.

Model 3 (Pooled)												
Variable	**Ln(Genre 1 + 1)**			**Ln(Genre 2 + 1)**			**Ln(Genre 3 + 1)**			**Ln(Genre 4 + 1)**		
	B	***SE B***	***Stand. β***	***B***	***SE B***	***Stand. β***	***B***	***SE B***	***Stand. β***	***B***	***SE B***	***Stand. β***
Ln(Age)	0.69	0.11	0.13***	-0.86	0.09	-0.18***	-0.18	0.10	-0.04	0.07	0.10	0.02
Gender (male = 1)	-0.65	0.06	-0.21***	-0.09	0.06	-0.03	-0.61	0.06	-0.20***	-0.20	0.06	-0.07***
Average Temperature	-0.03	0.01	-0.10***	-0.03	0.01	-0.12***	-0.02	0.01	-0.06*	-0.03	0.01	-0.10***
Average Cloud Coverage	0.04	0.04	0.04	0.09	0.04	0.10*	0.07	0.04	0.07	0.09	0.04	0.10*
Average Sunshine Duration	0.01	0.03	0.03	0.06	0.02	0.14**	0.03	0.03	0.07	0.06	0.02	0.12*
Ln(friends_genre1)t-1	0.04	0.02	0.04*									
Ln(friends_genre2)t-1				0.02	0.02	0.02						
Ln(friends_genre3)t-1							0.03	0.02	0.03			
Ln(friends_genre4)t-1										0.07	0.02	0.08***
R²	0.07			0.04			0.05			0.02		
Adjusted R²	0.06			0.04			0.05			0.02		
F	29.40***			18.63***			21.04***			7.50***		
N	2479			2479			2479			2479		

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$;

Let us take a look at the results from the panel data analysis. When we take the individuals effects and time effects into account, we unfortunately do not see significant influence on any of the hyper genres. Panel data analysis is most definitely the more appropriate statistical method for this data. Thus, if we had to choose between the two results, we would accept the hypothesis that music taste is **not** influenced by what your peers listen to. What must be added is that when we do not apply the filters of playcount > 0, we see significant influence for some hyper genres. However, these results would be biased. So we prefer to present negative results, than positive results that are biased.

Model 3 (Panel)																
	Ln(Genre 1 + 1)				**Ln(Genre 2 + 1)**				**Ln(Genre 3 + 1)**				**Ln(Genre 4 + 1)**			
	Fixed effect		**Random Effect**		**Fixed effect**		**Random Effect**		**Fixed effect**		**Random Effect**		**Fixed effect**		**Random Effect**	
Variable	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***
Ln(Age)	-	-	0.828	0.249**	-	-	-0.860	0.225***	-	-	0.004	0.243	-	-	0.152	0.246
Gender (male = 1)	-	-	-0.551	0.152***	-	-	-0.095	0.138	-	-	-0.577	0.149***	-	-	0.200	0.151
Average Temperature	-0.019	0.007**	-0.021	0.007**	-0.024	0.006***	-0.025	0.006***	-0.023	0.007**	-0.022	0.007**	-0.032	0.007***	-0.031	0.007***
Average Cloud Coverage	0.049	0.034	0.050	0.034	0.090	0.033**	0.091	0.032**	0.097	0.035**	0.093	0.035**	0.122	0.034***	0.119	0.034***
Average Sunshine Duration	0.025	0.022	0.024	0.022	0.061	0.021**	0.062	0.021**	0.053	0.022*	0.050	0.022*	0.083	0.022***	0.079	0.021***
Ln(friends_genre1)t-1	0.009	0.020	0.014	0.020												
Ln(friends_genre2)t-1					0.036	0.020	0.032	0.020								
Ln(friends_genre3)t-1									0.026	0.020	0.027	0.020				
Ln(friends_genre4)t-1													0.021	0.020	0.030	0.020
F(4, 2334)	3.49**				5.11**				4.98**				7.41***			
Wald chi2 (6)			39.32***				37.72***				34.17***				32.22***	
Hausman Test (Prob>chi2)			0.1442				0.8987				0.7915				0.2556	
N	2479		2479		2479		2479		2479		2479		2479		2479	

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

Model 4

Model 4 has actually the same dependent and independent variables as model 3. The only difference is that here we did not use the whole sample. We decided here to use only individuals with 1 friend. After analyzing the friend network, we came to the conclusion that 89.1% of the individuals had no friends. The second largest group was individuals with only one friend. This group comprised of 9.7% of the total sample of users. Thus, for this reason we chose to see whether there would be influence of music taste for this group. Here we applied the same filters as before, playcount per week of the individuals and their friends had to be larger than zero. We ended up with 2089 observations of the original 68635 observations.

The friends' genre 4 has the highest significant influence on the total plays per week of hyper genre_4 of the individuals, with a coefficient of 0.08 (p < 0.01), followed by the friends' genre 3 on hyper genre_3 of the individuals with a coefficient of 0.06 (p < 0.01). Thus, an increase of 1 total play per week of genre 4 in the previous week of one's friend, would result on average in 8% more listens of hyper genre_4 in the next week by the individuals. For hyper genre_3 this would result on average in 6 % more listens. Unfortunately, there is no significant influence for hyper genre_1 and hyper genre_2. Thus, for hyper genre_1 and hyper genre_2, music taste is not influenced by what their peers have listened to 1 week ago.

Model 4 (Pooled)												
	Ln(Genre_1 + 1)			Ln(Genre_2 + 1)			Ln(Genre_3 + 1)			Ln(Genre_4 + 1)		
Variable	*B*	*SE B*	*Stand. β*	*B*	*SE B*	*Stand. β*	*B*	*SE B*	*Stand. β*	*B*	*SE B*	*Stand. β*
Ln(Age)	1.11	0.11	0.21***	-0.83	0.10	-0.18***	0.06	0.11	0.01	0.27	0.11	0.06*
Gender (male = 1)	-0.80	0.07	-0.26***	-0.12	0.06	-0.04	-0.63	0.07	-0.21***	-0.27	0.07	-0.09***
Average Temperature	-0.04	0.01	-0.14***	-0.03	0.01	-0.13***	-0.02	0.01	-0.09**	-0.03	0.01	-0.14***
Average Cloud Coverage	0.06	0.04	0.06	0.08	0.04	0.09*	0.10	0.04	0.10*	0.10	0.04	0.10*
Average Sunshine Duration	0.03	0.03	0.05	0.06	0.03	0.13*	0.05	0.03	0.11*	0.06	0.03	0.13*
Ln(friends_genre1)t-1	0.04	0.02	0.04									
Ln(friends_genre2)t-1				0.03	0.02	0.03						
Ln(friends_genre3)t-1							0.06	0.02	0.06**			
Ln(friends_genre4)t-1										0.08	0.02	0.08**
R²	0.12			0.05			0.05			0.03		
Adjusted R²	0.11			0.04			0.05			0.03		
F	45.47***			16.23***			19.48***			10.10***		
N	2089			2089			2089			2089		

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$;

The panel data results for this model indicate that friend’s genre 3 has a significant influence on the individuals’ hyper genre_3 with a *p*-value smaller than 0.05 when we executed the random effects. The Hausman Test indicated that for this model, random effects is more appropriate than fixed effect model.

The hyper genre with the highest Wald chi2 was hyper genre_1 (Wald chi2 = 53.39; *p* < 0.001), followed by hyper genre_3 (Wald chi2 = 38.31; *p* < 0.001) and hyper genre_2 (Wald chi2 = 37.43; *p* < 0.001). The hyper genre with the lowest Wald chi2 was hyper genre_4 (Wald chi2 = 35.36; *p* < 0.001). The Wald chi2 is used to test the hypothesis that at least one of the predictors’ regression coefficients is not equal to zero.

Model 4 (Panel)																
	Ln(Genre_1 + 1)				**Ln(Genre_2 + 1)**				**Ln(Genre_3 + 1)**				**Ln(Genre_4 + 1)**			
	Fixed effect		**Random Effect**		**Fixed effect**		**Random Effect**		**Fixed effect**		**Random Effect**		**Fixed effect**		**Random Effect**	
Variable	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***
Ln(Age)	-	-	1.070	0.256***	-	-	-0.883	0.239***	-	-	0.115	0.251	-	-	0.252	0.262
Gender (male = 1)	-	-	-0.617	0.158***	-	-	-0.010	0.148	-	-	-0.576	0.155***	-	-	-0.255	0.162
Average Temperature	-0.024	0.007**	-0.026	0.007***	-0.027	0.007***	-0.028	0.007***	-0.029	0.008***	-0.027	0.007***	-0.038	0.007***	-0.037	0.007***
Average Cloud Coverage	0.048	0.037	0.051	0.037	0.076	0.036*	0.078	0.036*	0.115	0.038**	0.111	0.038**	0.121	0.037**	0.118	0.037**
Average Sunshine Duration	0.024	0.024	0.025	0.024	0.051	0.023*	0.052	0.023*	0.064	0.024**	0.061	0.024*	0.084	0.024***	0.081	0.024**
Ln(friends_genre1)t-1	0.014	0.023	0.020	0.022												
Ln(friends_genre2)t-1					0.038	0.022	0.036	0.022								
Ln(friends_genre3)t-1									0.037	0.022	0.042	0.021*				
Ln(friends_genre4)t-1													0.020	0.022	0.030	0.022
F(4, 1959)	4.47**				5.21**				6.24**				7.73***			
Wald chi2 (6)			53.39***				37.43***				38.31***				35.36***	
Hausman Test (Prob>chi2)			0.0739				0.9850				0.7523				0.2930	
N	2089		2089		2089		2089		2089		2089		2089		2089	

$*p < 0.05$; $**p < 0.01$; $***p < 0.001$;

Model 5

For model 5 we have decided to work once again with the whole sample of users. In this model we are still interested at looking whether friends influence what the individuals listen to. Hence, we have used the same dependent and independent variables as the previous models, but for this model we have also added the friends' genres from 2 weeks ago. So, we included friends' genres with one and two week lags. This was done to see whether there is a decay effect or not. However, we decided to use only 2 lags, because if one includes too much lags there is a risk of multicollinearity and also the standard error would be inflated.

Most definitely we can see here a decay effect that is taking place of the influence friends' genres has on the individuals' hyper genres. Previously in model 3 we saw that the friends' genre 1 of the previous week had a significant influence on the individuals' hyper genre_1 with a coefficient of 0.04 (p < 0.05), and also that friends' genre 4 of the previous week had a significant influence on the individuals' hyper genre_4 with a coefficient of 0.08 (p < 0.001). Now in model 5 we see that the friend's genre 1 of the previous week has no influence on the individuals' hyper genre_1 anymore plus the coefficient is smaller (0.02). This is also the case for the friends' genre 4 of the previous week, which has no influence anymore on hyper genre_4, and additionally the coefficient is now smaller (0.03). However, we do see that the friends' genre 4 of 2 weeks in the past has a significant influence on the individuals' hyper genre_4, with a coefficient of 0.05 (p < 0.05). Thus, an increase of 1 total play per week of the friends' genre 4 of 2 weeks in the past, results on average in 5% more listens of the individuals' hyper genre_4 for that particular week. Nevertheless, none of the other friends' listenings of 2 weeks in the past had a significant influence on the individuals' hyper genres.

Model 5 (Pooled)												
	Ln(Genre_1 + 1)			**Ln(Genre_2 + 1)**			**Ln(Genre_3 + 1)**			**Ln(Genre_4 + 1)**		
Variable	***B***	***SE B***	***Stand. β***	***B***	***SE B***	***Stand. β***	***B***	***SE B***	***Stand. β***	***B***	***SE B***	***Stand. β***
Ln(Age)	0.65	0.11	0.12***	-0.89	0.10	-0.19***	-0.19	0.10	-0.04	0.01	0.10	0.00
Gender (male = 1)	-0.65	0.06	-0.21***	-0.13	0.06	-0.05*	-0.65	0.06	-.021***	-0.18	0.06	-0.06**
Average Temperature	-0.03	0.01	-0.09***	-0.02	0.01	-0.09**	-0.02	0.01	-0.06*	-0.02	0.01	-0.09**
Average Cloud Coverage	0.04	0.04	0.04	0.11	0.04	0.13**	0.06	0.04	0.06	0.08	0.04	0.09*
Average Sunshine Duration	0.01	0.03	0.02	0.06	0.02	0.14**	0.03	0.03	0.06	0.04	0.03	0.09
Ln(friends_genre1)t-1	0.01	0.02	0.02									
Ln(friends_genre2)t-1				0.01	0.02	0.01						
Ln(friends_genre3)t-1							0.02	0.02	0.02			
Ln(friends_genre4)t-1										0.03	0.02	0.03
Ln(friends_genre1)t-2	0.04	0.02	0.04									
Ln(friends_genre2)t-2				-0.002	0.02	-0.002						
Ln(friends_genre3)t-2							-0.01	0.02	-0.01			
Ln(friends_genre4)t-2										0.05	0.02	0.05*
R²	0.07			0.05			0.05			0.02		
Adjusted R²	0.06			0.05			0.05			0.01		
F	24.56***			17.35***			19.58***			5.55***		
N	2448			2448			2448			2448		

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

The panel data results for this model indicate that none of the friend's genres has a significant influence on the individuals' hyper genres. Neither for the previous week nor the week before that. However, when we compare these results with the panel data results of model 3, it can be concluded that also in this model there is a decay effect present.

The Hausman test starts by indicating that for this model it is more appropriate to use the random effects model for hyper genre_1, hyper genre_2, and hyper genre_3. However, for hyper genre_4 the more appropriate model would be the fixed effect model. This has probably to do with the omitted variable bias when using the random effects model.

Previously in model 3 we saw that there was a greater effect for all the hyper genres than in model 5, namely 1.4% for the friends' genre 1 of the previous week on the individuals' hyper genre_1, 3.2% for the friends' genre 2 of the previous week on the individuals' hyper genre_2, 2.7% for the friends' genre 3 of the previous week on the individuals' hyper genre_3, and 3.0% for the friends' genre 4 of the previous week on the individuals' hyper genre 4. However, none of the friends' genres had a significant influence on the individuals' hyper genres. In model 5 we see that these values are significantly lower after adding the independent variables, friends' genres of two weeks in the past. The effect of the friends' genre 1 of the previous week on the individuals' hyper genre_1 is 0.9%, for the friends' genre 2 of the previous week on the individuals' hyper genre_2 this is 2.0%, for the friends' genre 3 of the previous week on the individuals' hyper genre_3 this is 1.7%, and for the friends' genre 4 of the previous week on the individuals' hyper genre_4 this is 1.8%. Hence, there is most definitely a decay effect when the friends' genres of week 1 and week 2 in the past are introduced.

Model 5 (Panel)																
	Ln(Genre_1 + 1)				**Ln(Genre_2 + 1)**				**Ln(Genre_3 + 1)**				**Ln(Genre_4 + 1)**			
	Fixed effect		**Random Effect**		**Fixed effect**		**Random Effect**		**Fixed effect**		**Random Effect**		**Fixed effect**		**Random Effect**	
Variable	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***
Ln(Age)	-	-	0.816	0.251**	-	-	-0.920	0.224***	-	-	0.004	0.234	-	-	0.058	0.251
Gender (male = 1)	-	-	-0.586	0.155***	-	-	-0.141	0.138	-	-	-0.629	0.144***	-	-	-0.199	0.155
Average Temperature	-0.021	0.007**	-0.022	0.007**	-0.022	0.007**	-0.023	0.007**	-0.026	0.007***	-0.024	0.007**	-0.029	0.007***	-0.028	0.007***
Average Cloud Coverage	0.034	0.035	0.037	0.035	0.090	0.034**	0.096	0.033**	0.063	0.036	0.064	0.035	0.092	0.035**	0.092	0.034**
Average Sunshine Duration	0.018	0.022	0.017	0.022	0.058	0.021**	0.060	0.021**	0.044	0.022	0.042	0.022	0.063	0.022**	0.061	0.022**
Ln(friends_genre1)t-1	0.007	0.018	0.009	0.018												
Ln(friends_genre2)t-1					0.023	0.016	0.020	0.016								
Ln(friends_genre3)t-1									0.017	0.018	0.017	0.017				
Ln(friends_genre4)t-1													0.020	0.018	0.023	0.018
Ln(friends_genre1)t-2	0.009	0.021	0.014	0.021												
Ln(friends_genre2)t-2					-0.002	0.022	-0.001	0.021								
Ln(friends_genre3)t-2									-0.018	0.022	-0.016	0.021				
Ln(friends_genre4)t-2													-0.011	0.021	-0.001	0.020
F(5, 2302)	3.32**				3.49**				3.61**				4.65**			
Wald chi2 (7)			43.16***				38.94***				36.23***				26.02**	
Hausman Test (Prob>chi2)			0.1206				0.2081				0.3789		0.0119			
N	2448		2448		2448		2448		2448		2448		2448		2448	

Model 6

Model 6 has actually the same dependent and independent variables as model 5. The only difference is that here we did not use the whole sample. We decided here to use only individuals with 1 friend. After analyzing the friend network, we came to the conclusion that 89.1% of the individuals had no friends. The second largest group was individuals with only one friend. This group comprised of 9.7% of the total sample of users. Here we applied the same filters as before, playcount per week of the individuals and their friends playcount of 2 weeks in the past had to be larger than zero. We ended up with 2060 observations of the original 68635 observations.

Only the friends' genre 3 of the previous week has a significant influence on the individuals' hyper genre_3, with a coefficient of 0.05 (p < 0.05). An increase of 1 total play per week of the friends' genre 3 of the previous week, would result on average in 4% extra listening to hyper genre_3. Once again we could see in this model a decay effect taking place when compared to model 4. There are less significant influences on the hyper genres and also the effects are smaller than the effects in model 4.

Model 6 (Pooled)												
	Ln(Genre_1 + 1)			**Ln(Genre_2 + 1)**			**Ln(Genre_3 + 1)**			**Ln(Genre_4 + 1)**		
Variable	***B***	***SE B***	***Stand. β***	***B***	***SE B***	***Stand. β***	***B***	***SE B***	***Stand. β***	***B***	***SE B***	***Stand. β***
Ln(Age)	1.07	0.11	0.21***	-0.87	0.10	-0.19***	0.05	0.11	0.01	0.20	0.11	0.04
Gender (male = 1)	-0.80	0.07	-0.25***	-0.16	0.06	-0.06**	-0.68	0.07	-0.23***	-0.26	0.07	-0.09***
Average Temperature	-0.04	0.01	-0.13***	-0.03	0.01	-0.11**	-0.02	0.01	-0.08**	-0.03	0.01	-0.12***
Average Cloud Coverage	0.05	0.04	0.05	0.11	0.04	0.13**	0.08	0.04	0.09	0.09	0.04	0.10*
Average Sunshine Duration	0.02	0.03	0.04	0.06	0.03	0.13*	0.05	0.03	0.10	0.05	0.03	0.11*
Ln(friends_genre1)t-1	0.01	0.02	0.01									
Ln(friends_genre2)t-1				0.01	0.02	0.02						
Ln(friends_genre3)t-1							0.04	0.02	0.05*			
Ln(friends_genre4)t-1										0.03	0.02	0.03
Ln(friends_genre1)t-2	0.04	0.02	0.04									
Ln(friends_genre2)t-2				0.01	0.02	0.01						
Ln(friends_genre3)t-2							-0.002	0.02	-0.002			
Ln(friends_genre4)t-2										0.04	0.02	0.04
R^2	0.11			0.05			0.06			0.02		
Adjusted R^2	0.11			0.05			0.06			0.02		
F	37.53***			15.37***			18.21***			7.2***		
N	2060			2060			2060			2060		

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

The panel data results for this model indicate that none of the friend's genres has a significant influence on the individuals' hyper genres. Neither for the previous week nor the week before that. However, when we compare these results with the panel data results of model 4, it can be concluded that also in this model there is a decay effect present. The coefficient of the random effect of $Ln(friends\_genre1)_{t-1}$ in model 4 was 2.0% for Ln(Genre_1 +1) and in model 6 it was 1.3%. Similarly, the coefficient of $Ln(friends\_genre2)_{t-1}$ in model 4 was 3.6% for Ln(Genre_2 + 1) and in model 6 it was 2.4%. This is also the case for the coefficient of $Ln(friends\_genre3)_{t-1}$, which was 4.2% in model 4 for Ln(Genre_3 + 1) and in model 6 it was 3.1%. The coefficient for $Ln(friends\_genre4)_{t-1}$ was for Ln(Genre_4) in model 4 3.0%, and in model 6 this was 2.1%. When we look at the coefficients of $Ln(friends\_genreX)_{t-2}$ for each hyper genre, the coefficients are even smaller.

The Hausman test starts by indicating that for this model it is more appropriate to use the random effects model for hyper genre_2 and hyper genre_3. However, for hyper genre_1 and hyper genre_4 the more appropriate model would be the fixed effect model. This has probably to do with the omitted variable bias when using the random effects model.

Previously in model 4 we saw that there was a greater effect for all the hyper genres than in model 6, namely 2.0% for the friends' genre 1 of the previous week on the individuals' hyper genre_1, 3.6% for the friends' genre 2 of the previous week on the individuals' hyper genre_2, 4.2% for the friends' genre 3 of the previous week on the individuals' hyper genre_3, and 3.0% for the friends' genre 4 of the previous week on the individuals' hyper genre_4. However, none of the friends' genres had a significant influence on the individuals' hyper genres. In model 6 we see that these values are significantly lower after adding the independent variables, friends' genres of two weeks in the past. The effect of the friends' genre 1 of the previous week on the individuals' hyper genre_1 is 1.0%, for the friends' genre 2 of the previous week on the individuals' hyper genre_2 this is 2.4%, for the friends' genre 3 of the previous week on the individuals' hyper genre_3 this is 3.1%, and for the friends' genre 4 of the previous week on the individuals' hyper genre_4 this is 1.9%. Hence, there is most definitely a decay effect when the friends' genres of week 1 and week 2 in the past are introduced.

Model 6 (Panel)																
	Ln(Genre_1 + 1)				**Ln(Genre_2 + 1)**				**Ln(Genre_3 + 1)**				**Ln(Genre_4 + 1)**			
	Fixed effect		**Random Effect**		**Fixed effect**		**Random Effect**		**Fixed effect**		**Random Effect**		**Fixed effect**		**Random Effect**	
Variable	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***	***B***	***SE B***
Ln(Age)	-	-	1.060	0.258***	-	-	-0.949	0.236***	-	-	0.120	0.239	-	-	0.150	0.268
Gender (male = 1)	-	-	-0.652	0.160***	-	-	-0.156	0.147	-	-	-0.639	0.149***	-	-	-0.268	0.167
Average Temperature	-0.026	0.008**	-0.028	0.008***	-0.024	0.008**	-0.025	0.007**	-0.031	0.008***	-0.029	0.008***	-0.033	0.008***	-0.032	0.008***
Average Cloud Coverage	0.030	0.038	0.034	0.038	0.079	0.037*	0.086	0.037*	0.073	0.039	0.075	0.039	0.088	0.038*	0.090	0.038*
Average Sunshine Duration	0.016	0.024	0.017	0.024	0.048	0.024*	0.051	0.023*	0.052	0.025*	0.051	0.025*	0.061	0.024*	0.060	0.024*
Ln(friends_genre1)t-1	0.010	0.020	0.013	0.020												
Ln(friends_genre2)t-1					0.027	0.017	0.024	0.017								
Ln(friends_genre3)t-1									0.029	0.019	0.031	0.018				
Ln(friends_genre4)t-1													0.019	0.020	0.021	0.020
Ln(friends_genre1)t-2	0.008	0.023	0.016	0.023												
Ln(friends_genre2)t-2					-0.006	0.024	-0.003	0.023								
Ln(friends_genre3)t-2									-0.032	0.023	-0.026	0.023				
Ln(friends_genre4)t-2													-0.016	0.023	-0.005	0.022
F(5, 1929)	4.18**				3.52**				4.58**				4.65**			
Wald chi2 (7)			57.55***				38.43***				39.98***				27.38**	
Hausman Test (Prob>chi2)	0.0447						0.1357				0.1117		0.0187			
N	2060		2060		2060		2060		2060		2060		2060		2060	

6 Conclusions

This dissertation has investigated whether social influence or external factors play a role in influencing one’s music taste. We mainly focused on the relationship between the friends musical tastes and whether they influence the music taste per week for each of the 1295 Dutch Last.fm users we have used in our subsample. Furthermore, we set out to determine if external factors such as temperature, cloud coverage, and sunshine duration also have an influence on the music listening behavior of individuals.

According to past studies, individuals’ music taste is not only influenced by one’s personality traits, but it can also be influenced by many other factors (Levitin, 2011). After analyzing the data, we can conclude that the data stood up well against most of our expectations. Firstly, music taste is indeed strongly influenced by the age and gender of the respondent. The older respondents listened more to genres such as blues, jazz, classical, and folk. In contrast to the younger respondents who listened more to genres such as rock, alternative, and heavy metal. This is supported by the literature, which states that music taste begins with fairly narrow tastes in young adulthood, and then expands into middle age, and narrows again later in life (Harrison & Ryan, 2010). In addition, the older the respondents become, the less music they listen to per week. Younger respondents listen much more to genres such as rock, alternative, and heavy metal.

The results also show that females have listened more to genres such as pop music, classical, folk, and jazz. This is also supported by Roe (1984), which argued that females in general like “pop hits” or mainstream music, folk, and classical music. According to Roe (1984), males listen much more to genres like rock, alternative, and hard rock. Due to the fact that males prefer “macho/aggressive” styles of music. This opposes our findings, because in our results females listen slightly more to genres such rock, alternative, and heavy metal.

Moreover, we have proven that external conditions, like weather, have a significant effect on the music taste of individuals. Studies in psychology hold the view that temperature greatly influences mood, and mood changes in turn cause behavioral changes (Cao & Wei, 2005). In our results we did see behavioral changes, people listened to all hyper genres less when the temperature increased. The most obvious explanation for this observation is that Dutch people tend to be more outside when there is nice weather. Another interesting finding was that cloud coverage has a positive effect on the amount of plays per hyper genre in a particular week. The respondents listened more to each hyper genre when the sky was cloudy and miserable. Thus, once again proving that weather greatly influences the mood of the respondents, which in turn results in behavioral changes.

As for social influence from one's friends, the multiple regression analysis provided some significant evidence that friends can influence an individual's music taste. Nevertheless, when we accounted for the individual effects and time effects by performing a panel data analysis, suddenly there was no significant evidence anymore for social influence from one's friends. The results of the panel data analysis would be the more appropriate analysis for this study, due to the way the data was categorized by individuals and weeks. However, we must add that in model 4, where we only used users with one friend, there was a significant evidence for hyper genre 3. This gives us an indication that if we have users with more friends, we probably will get more significant evidence of social influence playing a role in one's music taste.

Above all, it can be concluded that most of the variables in this study have an effect on the music taste of an individual. However, the one effect that we were the most interested in, social influence, did not provide sufficient significant evidence to conclude that it plays a role in influencing one's music taste.

6.1 Limitations

There are a number of concerns with the conclusions provided in the preceding section, which may have influenced the soundness of the results:

1. *More assumptions could have been looked into*. While the research attempted to investigate several assumptions, due to time constraints it was not possible to explore various options of these assumptions. Specifically, the following options of the assumptions for:
   - Date friends: In order for two users to become 'friends' on Last.fm, one user has to invite the other user to become 'friends'. Thus, the other user has to accept the invitation before they become 'friends' of one another. However, this date, when they became 'friends', was not available on Last.fm. Would the results have been different if we knew exactly when the users became friends?
   - Musical neighbors: Last.fm points out to its users other people on Last.fm who have the most similar tastes to them. These musical neighbors are

automatically recommended to Last.fm users. Would the results have been different if we took musical neighbors into account?

- Time interval: A time interval was defined in this study to investigate whether friends could influence someone's music taste. We decided to use a week as time interval. Would the results be different if we decided that the time interval should be one day or one month?

2. *The sample users must contain enough friend relationships.* Last.fm users that were used for this study did not have enough friend relations. After analyzing the friend network we came to the conclusion that the majority of our sample users (1295) did not have any friends. The second largest group (126) had only one friend. Would the results have been different if our sample users had enough friend relationships?
3. *The sample users must have listened actively throughout the chosen time period.* The sample users used in this study did not listen actively throughout the time period (1 year). We had 1295 users and for each there was 53 weeks (1 year) of observations. Thus, in total we had 68635 (1295 x 53) observations. However, when we applied the filter that the playcount for each user for each week had to be larger than zero (playcount > 0), we ended up with 37596 observations. This is almost half of the observations. Would the results have been different if the sample users were actively listening to Last.fm throughout the year?

6.2 Future research

There are a number of ways in which the model for social influence in the music industry could be investigated to a greater extent:

1. *Further investigate the assumptions recognized in the preceding section*. One must take into account different time interval (day or month), include musical neighbors in the study, and include when users became friends with one another.
2. *Investigate data from different sources*. Nowadays, there are more and more social music sites making these types of data publicly available. By combining data from social music sites such as Spotify, YouTube, Pandora, and Rhapsody.

References

Aggarwal, C. C. (2011). *An introduction to social network data analytics* (pp. 1-15). Springer US.

Alba, R. D. 1982. Taking stock of network analysis: a decade's results. *Research in the Sociology of Organizations* 1:39-74

Alderman, J. (2002). *Sonic boom: Napster, MP3, and the new pioneers of music*. HarperCollins UK.

Anagnostopoulos, A., Kumar, R., & Mahdian, M. (2008, August). Influence and correlation in social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 7-15). ACM.

Apple, 2007: *Q4 2007 Unaudited Summary Data*. Apple: http://images.apple.com/investor/

Aristotle. 1934. *Rhetoric. Nichomachean ethics*. In *Aristotle in 23 volumes*. Rackman transl. Cambridge: Harvard Univ. Press

Baker, S. (2006). Multiple Regression Theory.

Barash, V. (2011). *The dynamics of social contagion* (Doctoral dissertation, Cornell University).

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. U. (2006). Complex networks: Structure and dynamics. *Physics reports*, *424*(4), 175-308.

Boorman, S. A., & White, H. C. (1976). Social structure from multiple networks. II. Role structures. *American Journal of Sociology*, 1384-1446.

Cao, M., & Wei, J. (2005). Stock market returns: A note on temperature anomaly. *Journal of Banking & Finance*, *29*(6), 1559-1573.

Casciaro, T., Carley, K. M., & Krackhardt, D. (1999). Positive affectivity and accuracy in social network perception. *Motivation and Emotion*, *23*(4), 285-306.

Cattell, R. B., & Saunders, D. R. (1954). Musical preferences and personality diagnosis: I. A factorization of one hundred and twenty themes. *The Journal of Social Psychology*, *39*(1), 3-24.

Christenson, P. G., & Peterson, J. B. (1988). Genre and gender in the structure of music preferences. *Communication Research*, *15*(3), 282-301.

Costa, L. D. F., Rodrigues, F. A., Travieso, G., & Villas Boas, P. R. (2007). Characterization of complex networks: A survey of measurements. *Advances in Physics*, *56*(1), 167-242.

Currarini, S., & Vega-Redondo, F. (2011). *A simple model of homophily in social networks*. mimeo.

Denisoff, R. S. (1988). *Inside Mtv*. Transaction Publishers.

Dolata, U. (2011). The music industry and the internet: a decade of disruptive and uncontrolled sectoral change.

Dwyer, J. J. (1995). Effect of perceived choice of music on exercise intrinsic motivation. *Health Values: The Journal of Health Behavior, Education & Promotion*.

Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer‐Mediated Communication*, *13*(1), 210-230.

Ethington, C. A., Thomas, S. L., & Pike, G. R. (2002). Back to the basics: Regression as it should be. In *Higher education: Handbook of theory and research* (pp. 263-293). Springer Netherlands.

Euler, L., 1736. Solutio Problematis Ad geometriam Situs Pertinentis. *Commenrarii Academiae Scientiarum Imperialis Petropolitanae*, 8:128–140. Reprint in English in N. Biggs, E. Lloyd, and R. Wilson (1976), editors, *Graph Theory 1736-1936*, Clarendon Press, Oxford, UK.

Freeman, L. C., White, D., & Romney, A. K. (Eds.). (1992). *Research methods in social network analysis*. Transaction Books.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, *15*, 246-263.

Gowensmith, W. N., & Bloom, L. J. (1997). The effects of heavy metal music on arousal and anger. *Journal of Music Therapy*, *34*, 33-45.

Harrison, J., & Ryan, J. (2010). Musical taste and ageing. *Ageing and Society*, *30*(4), 649.

Howarth, E., & Hoffman, M. S. (1984). A multidimensional approach to the relationship between mood and weather. *British Journal of Psychology*, *75*(1), 15-23.

Hughes, J., & Lang, K. R. (2003). If I had a song: The culture of digital community networks and its impact on the music industry. *International Journal on Media Management, 5*(3), 180-189.

IFPI (International Federation of the Phonographic Industry), 1999: *The Recording Industry in Numbers '99*. London: IFPI.

IFPI (International Federation of the Phonographic Industry), 2008: *IFPI Digital Music Report 2008*. London: IFPI.

IFPI (International Federation of the Phonographic Industry), 2010: *IFPI Digital Music Report 2010*. London: IFPI.

Iyengar, R., Van den Bulte, C., & Valente, T. W. (2011). Opinion leadership and social contagion in new product diffusion. *Marketing Science, 30*(2), 195-212.

Kohut H, Levarie S. (1950). On the enjoyment of listening to music. The Psychoanalytic Quarterly.

Large, E. W. (2000). On synchronizing movements to music. *Human Movement Science, 19*(4), 527-566.

Latané, B. (2000). Pressures to uniformity and the evolution of cultural norms: Modeling dynamic social impact.

Laumann, E. O., Marsden, P. V., & Prensky, D. (1989). The boundary specification problem in network analysis. *Research methods in social network analysis, 61*, 87.

Levitin, D. J. (2011). *This is your brain on music: Understanding a human obsession*. Atlantic books.

Litle, P., & Zuckerman, M. (1986). Sensation seeking and music preferences. *Personality and individual differences, 7*(4), 575-578.

Maddala, G. S., (2001). Introduction to econometrics. (Third ed.). New York: Wiley.

Marin, A., & Wellman, B. (2011). Social network analysis: An introduction. *The Sage Handbook of Social Network Analysis, London, Sage*, 11-25.

McChesney, Robert W. *Corporate Media and the Threat to Democracy.* New York: Seven Stories Press, 1997.

McCown, W., Keiser, R., Mulhearn, S., & Williamson, D. (1997). The role of personality and gender in preference for exaggerated bass in music. *Personality and Individual Differences, 23*(4), 543-547.

McNamara, L., & Ballard, M. E. (1999). Resting arousal, sensation seeking, and music preference. *Genetic, Social, and General Psychology Monographs*.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 415-444.

Mizruchi, M. S., & Stearns, L. B. (1988). A longitudinal study of the formation of interlocking directorates. *Administrative Science Quarterly*, 194-210.

Murray, K. B., Di Muro, F., Finn, A., & Popkowski Leszczyc, P. (2010). The effect of weather on consumer spending. *Journal of Retailing and Consumer Services*, *17*(6), 512-520.

Myers, R. H. (1990). *Classical and modern regression with applications* (Vol. 2). Belmont, CA: Duxbury Press.

North, A. C., & Hargreaves, D. J. (1999). Music and adolescent identity. *Music education research*, *1*(1), 75-92.

North, A. C., Hargreaves, D. J., & O'Neill, S. A. (2000). The importance of music to adolescents. *British Journal of Educational Psychology*, *70*(2), 255-272.

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, *41*(5), 673-690.

Oyama, T., Hatano, K., Sato, Y., Kudo, M., Spintge, R., & Droh, R. (1987). Endocrine effect of anxiolytic music in dental patients. In *Musik in der Medizin/Music in Medicine* (pp. 223-226). Springer Berlin Heidelberg.

Plato. 1968. Laws. *Plato in Twelve Volumes*, *Vol. 11*. Bury translator. Cambridge: Harvard Univ. Press

Quan-Haase, A., & Wellman, B. (2005). Local virtuality in an organization: Implications for community of practice. In *Communities and technologies 2005* (pp. 215-238). Springer Netherlands.

Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: the structure and personality correlates of music preferences. *Journal of personality and social psychology*, *84*(6), 1236.

Rentfrow, P. J., Goldberg, L. R., & Levitin, D. J. (2011). The structure of musical preferences: A five-factor model. *Journal of personality and social psychology*, *100*(6), 1139.

RIAA (Recording Industry Association of America), 2008: *2007 Year-End Shipment Statistics*.

RIAA (Recording Industry Association of America), 2010: *2009 Year-End Shipment Statistics*.

Rider, M. S., Floyd, J. W., & Kirkpatrick, J. (1985). The effect of music, imagery, and relaxation on adrenal corticosteroids and the re-entrainment of circadian rhythms. *Journal of Music Therapy*.

Roe, K. (1984). Youth and Music in Sweden. Results from a Longitudinal Study of Teenagers' Media Use. Media Panel Report No. 32.

Sanjek, R., & Sanjek, D. (1991). *American popular music business in the 20th century*. Oxford University Press.

Saunders, E. M. (1993). Stock prices and Wall Street weather. *The American Economic Review*, *83*(5), 1337-1345.

Skilledtest,2012.
http://www.skilledtests.com/wiki/Last.fm_statistics#user_accounts_at_last.fm

Tarrant, M., North, A. C., & Hargreaves, D. J. (2000). English and American adolescents' reasons for listening to music. *Psychology of Music*, *28*(2), 166-173.

Tschmuck, P. (2012). *Creativity and innovation in the music industry* (pp. 225-251). Springer Berlin Heidelberg.

U.S. Copyright Office, 1998: *The Digital Millenium Copyright Act of 1998. U.S. Copyright Office Summary*. U.S. Copyright Office

Wasserman, S. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge university press.

Wasserman, Stanley and Faust, Katherine (1994) *Social Network Analysis*. Cambridge: Cambridge University Press.

Watts, D. J. (1999). *Small worlds: the dynamics of networks between order and randomness*. Princeton university press.

White, H. D., Wellman, B., & Nazer, N. (2004). Does citation reflect social structure?: Longitudinal evidence from the “Globenet” interdisciplinary research group. *Journal of the American Society for information Science and Technology*, *55*(2), 111-126.

White, H. C., Boorman, S. A., & Breiger, R. L. (1976). Social structure from multiple networks. I. Blockmodels of roles and positions. *American journal of sociology*, 730-780.

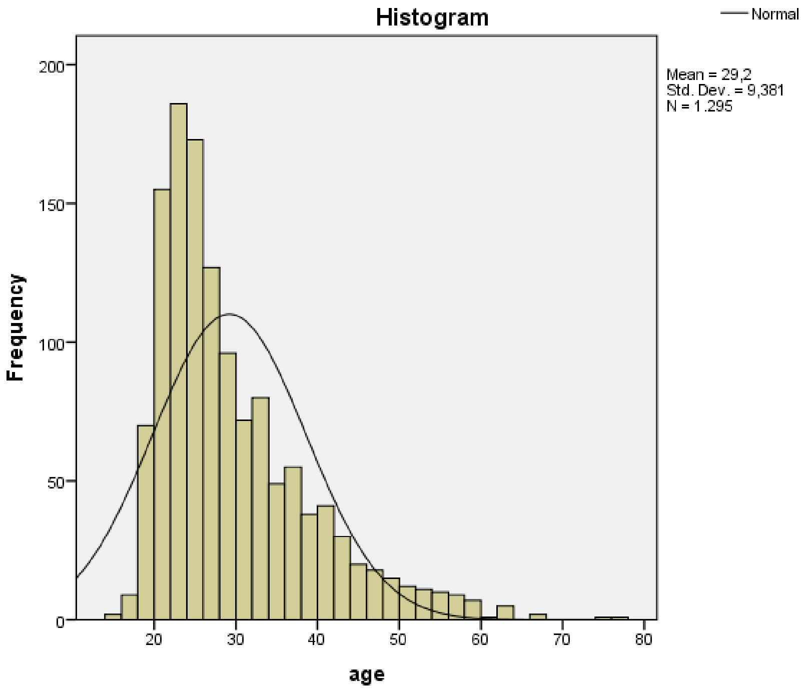
Appendix A.

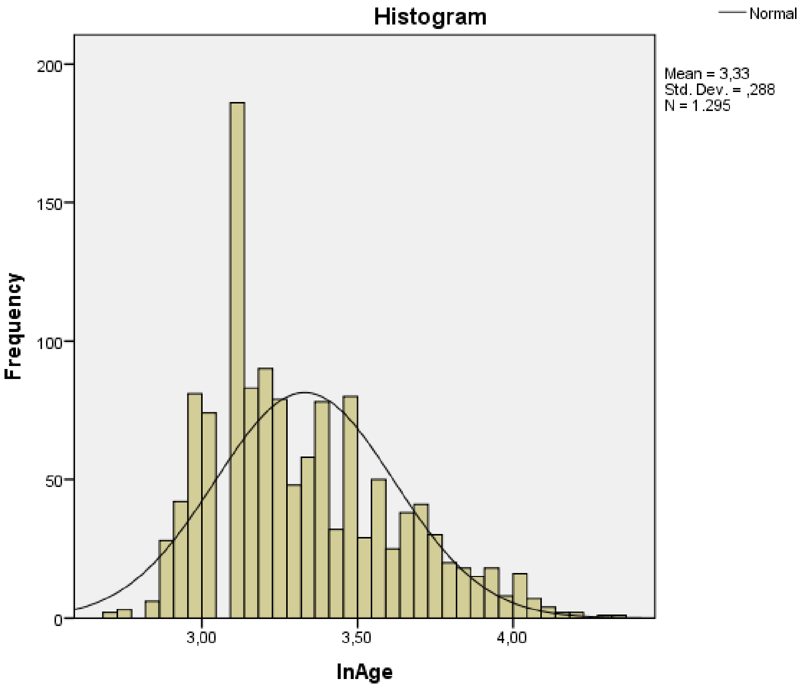
Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic
Age	1295	15	77	29,20	9,381	1,390
lnAge	1295	2,71	4,34	3,3299	,28849	,689
Valid N (listwise)	1295					

Descriptive Statistics

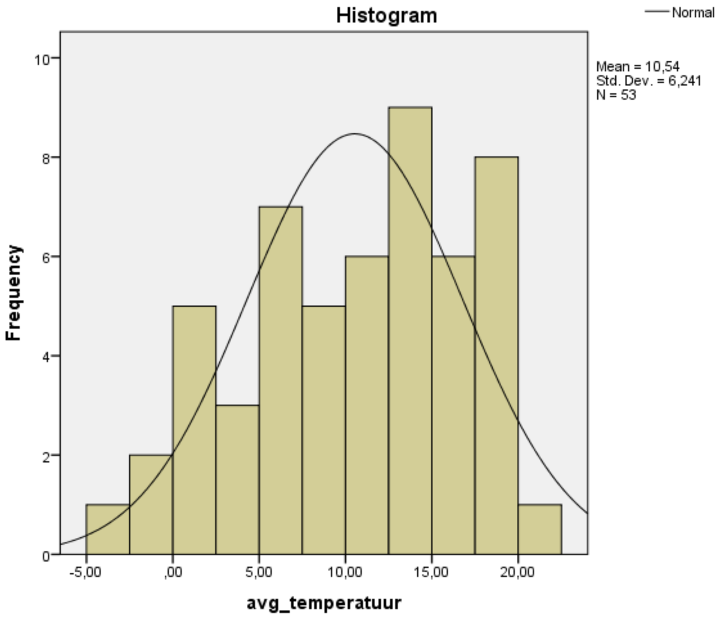
	Skewness	Kurtosis	
	Std. Error	Statistic	Std. Error
Age	,068	2,034	,136
lnAge	,068	-,074	,136
Valid N (listwise)			

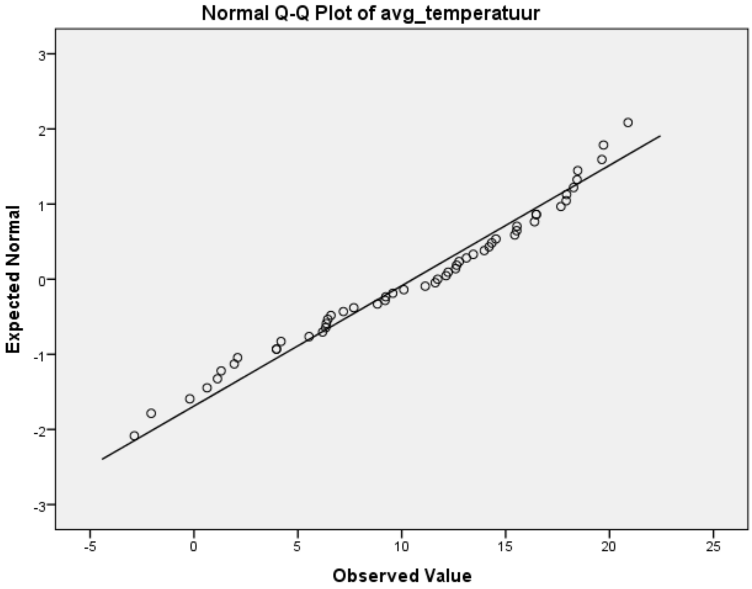


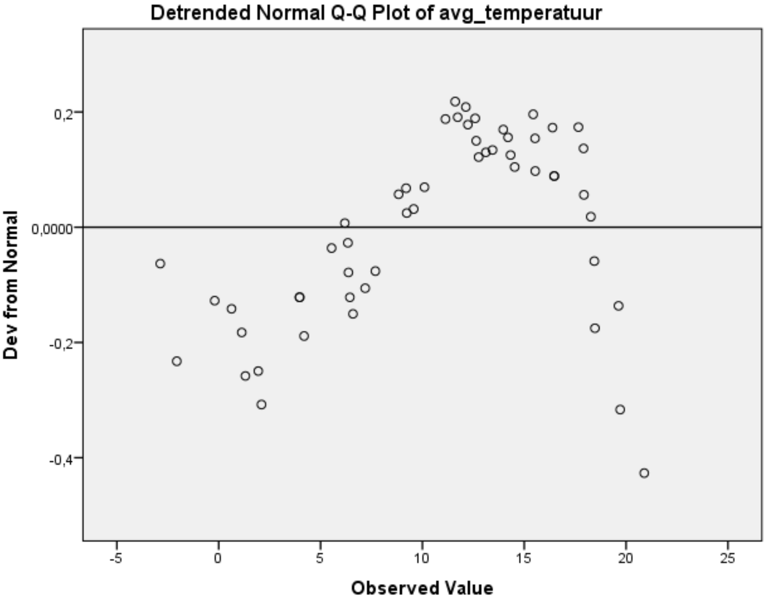

Histogram
Normal
Mean = 3,33
Std. Dev. = ,288
N = 1.295
Frequency
200
150
100
50
0
3,00
3,50
4,00
lnAge

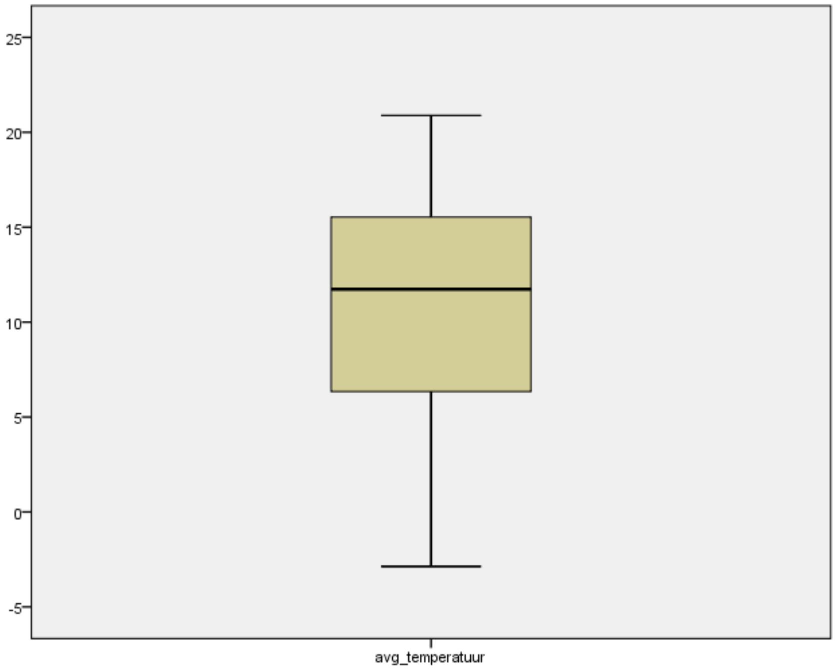
Appendix B.

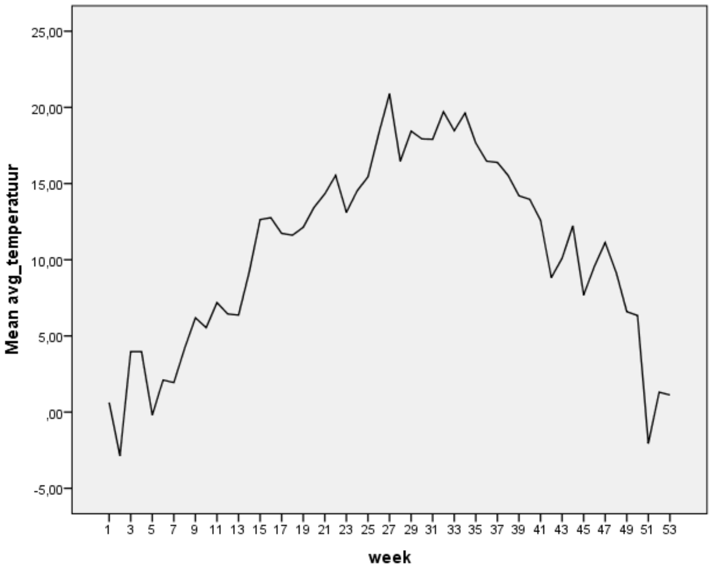
Weekly average temperature



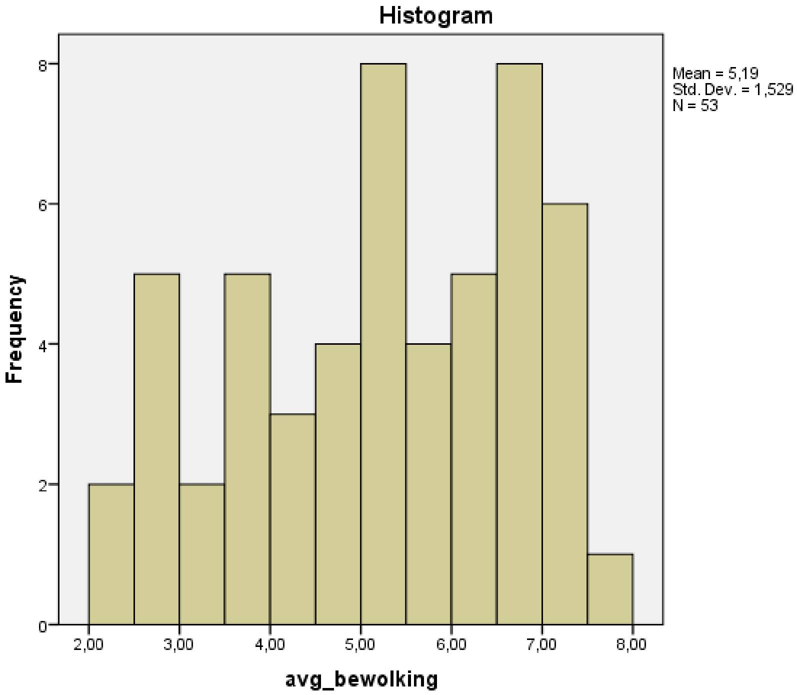


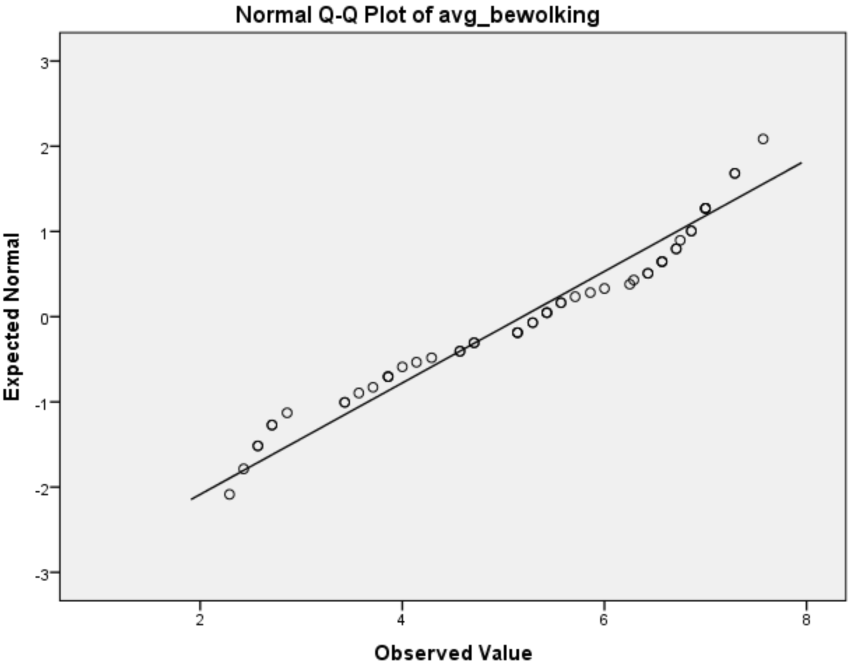

Detrended Normal Q-Q Plot of avg_temperatuur
Dev from Normal
0,2
0,0000
-0,2
-0,4
-5
0
5
10
15
20
25
Observed Value

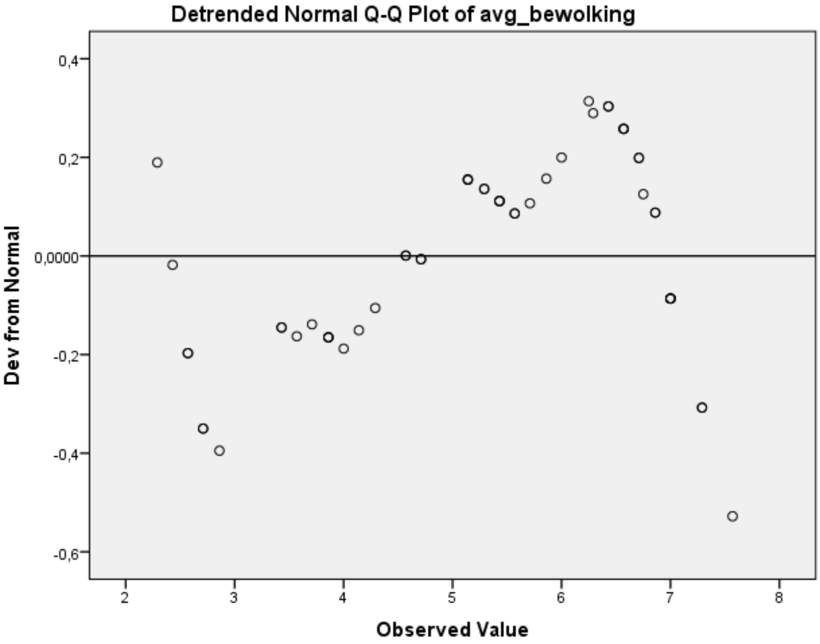

25
20
15
10
5
0
-5
avg_temperatuur

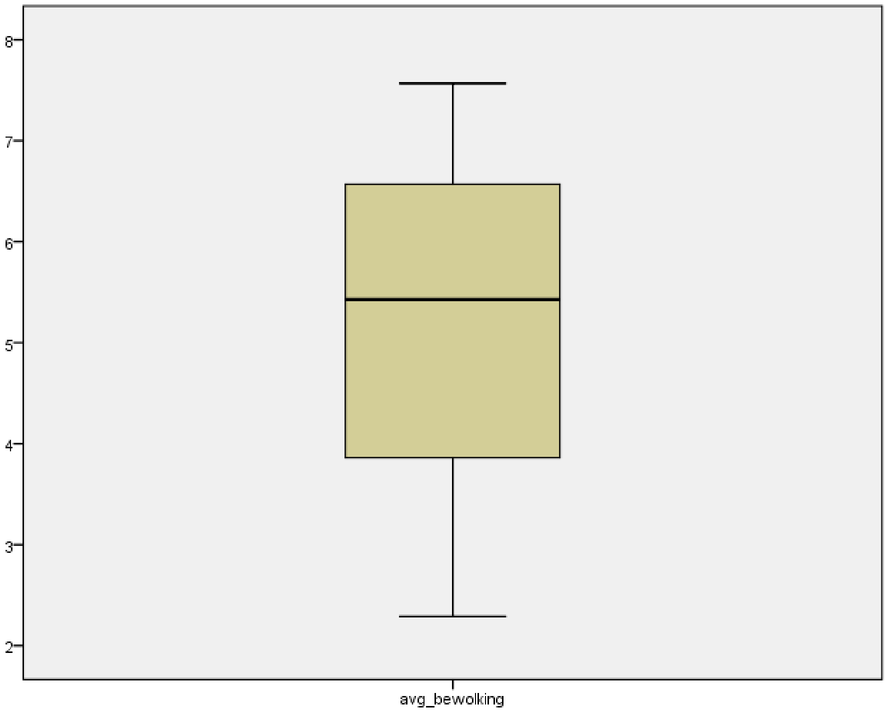


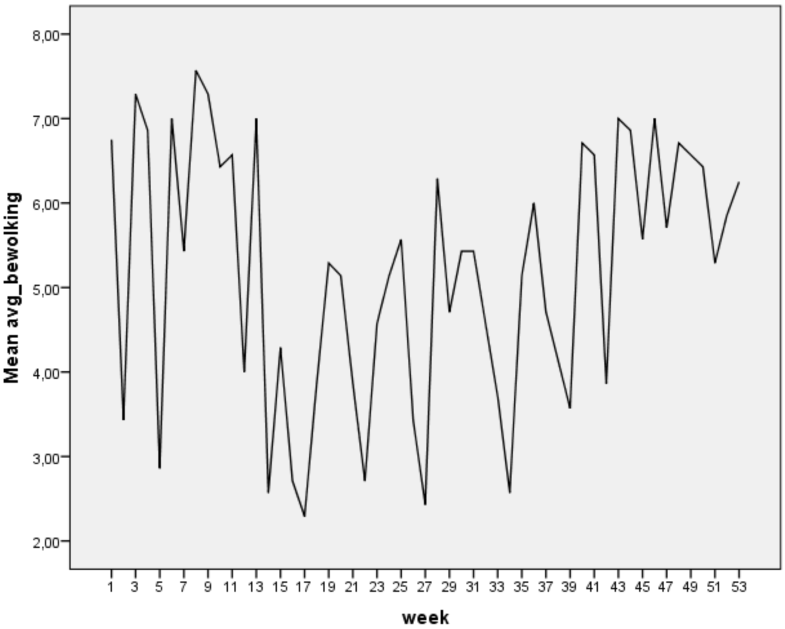
Weekly average cloud coverage



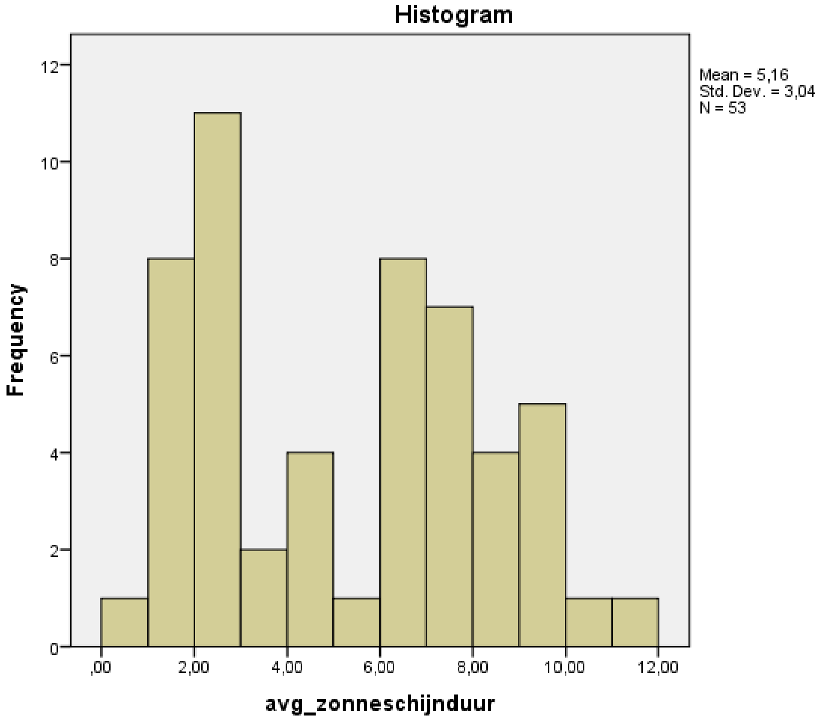

Normal Q-Q Plot of avg_bewolking
Expected Normal
Observed Value
3
2
1
0
-1
-2
-3
2
4
6
8

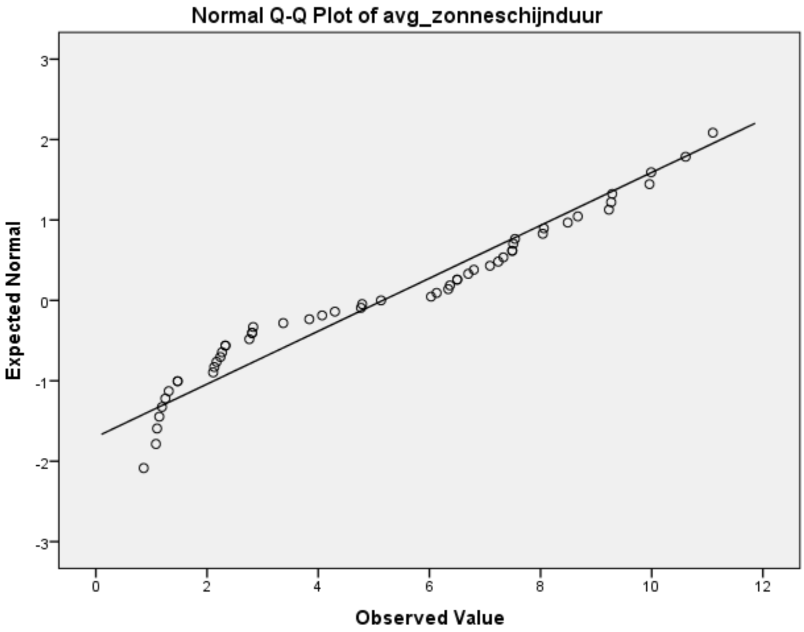

Detrended Normal Q-Q Plot of avg_bewolking
Dev from Normal
Observed Value
0,4
0,2
0,0000
-0,2
-0,4
-0,6
2
3
4
5
6
7
8

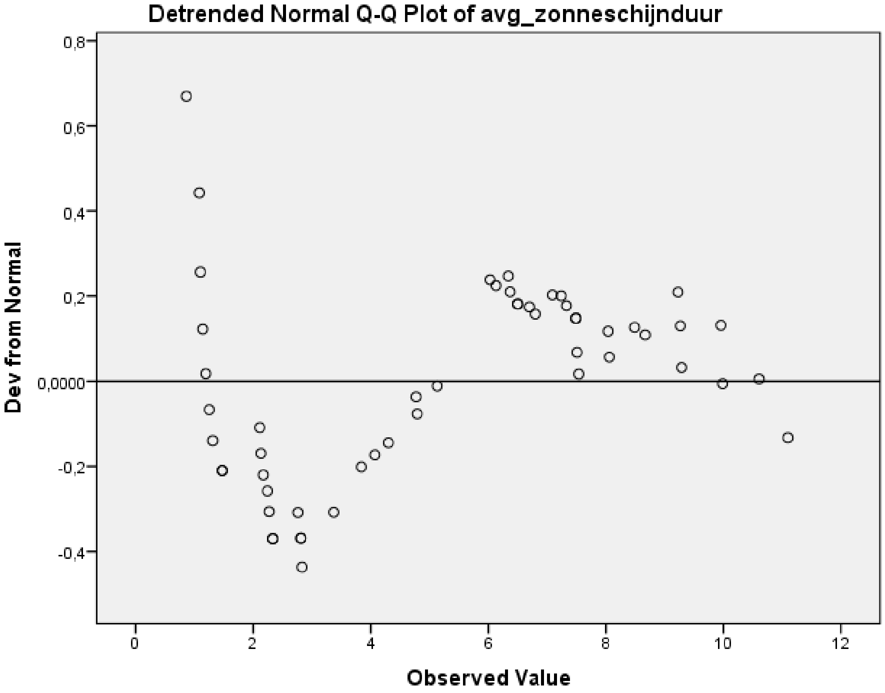

8
7
6
5
4
3
2
avg_bewolking

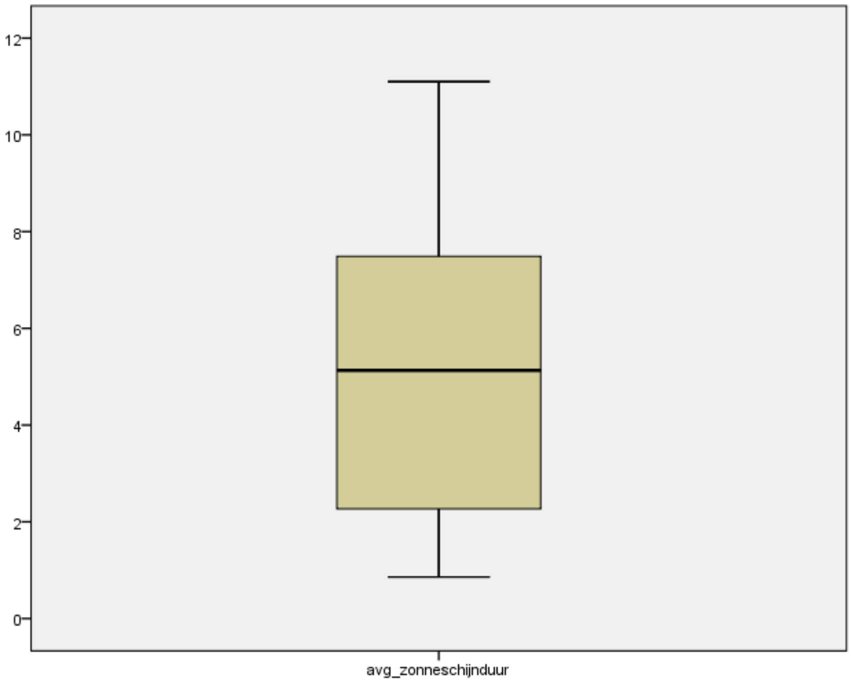

8,00
7,00
6,00
5,00
4,00
3,00
2,00
Mean avg_bewolking
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53
week

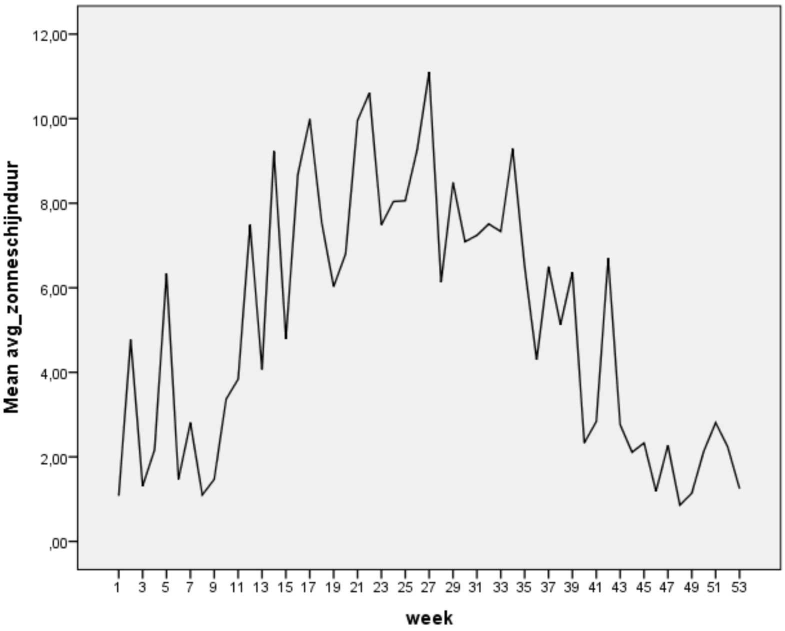
Weekly average sunshine duration






Detrended Normal Q-Q Plot of avg_zonneschijnduur
Dev from Normal
0,8
0,6
0,4
0,2
0,0000
-0,2
-0,4
0
2
4
6
8
10
12
Observed Value


12
10
8
6
4
2
0
avg_zonneschijnduur


12,00
10,00
8,00
6,00
4,00
2,00
,00
Mean avg_zonneschijnduur
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53
week

Appendix C.

Correlations

		lnPlaycountWeek	lnAge	lnAge2
Pearson Correlation	lnPlaycountWeek	1.000	-.050	-.048
	lnAge	-.050	1.000	.999
	lnAge2	-.048	.999	1.000
	gender	-.059	.266	.264
	avg_temperatuur	-.063	-.027	-.027
	avg_bewolking	.008	.003	.004
	avg_zonneschijnduur	-.009	-.009	-.009
Sig. (1-tailed)	lnPlaycountWeek	.	.000	.000
	lnAge	.000	.	.000
	lnAge2	.000	.000	.
	gender	.000	.000	.000
	avg_temperatuur	.000	.000	.000
	avg_bewolking	.057	.253	.230
	avg_zonneschijnduur	.039	.049	.039
N	lnPlaycountWeek	37596	37596	37596
	lnAge	37596	37596	37596
	lnAge2	37596	37596	37596
	gender	37596	37596	37596
	avg_temperatuur	37596	37596	37596
	avg_bewolking	37596	37596	37596
	avg_zonneschijnduur	37596	37596	37596

Correlations

		gender	avg_temperatuur	avg_bewolking
Pearson Correlation	lnPlaycountWeek	-.059	-.063	.008
	lnAge	.266	-.027	.003
	lnAge2	.264	-.027	.004
	gender	1.000	-.001	-.004
	avg_temperatuur	-.001	1.000	-.354
	avg_bewolking	-.004	-.354	1.000
	avg_zonneschijnduur	.005	.632	-.849
Sig. (1-tailed)	lnPlaycountWeek	.000	.000	.057
	lnAge	.000	.000	.253
	lnAge2	.000	.000	.230
	gender	.	.393	.213
	avg_temperatuur	.393	.	.000
	avg_bewolking	.213	.000	.
	avg_zonneschijnduur	.168	.000	.000
N	lnPlaycountWeek	37596	37596	37596
	lnAge	37596	37596	37596
	lnAge2	37596	37596	37596
	gender	37596	37596	37596
	avg_temperatuur	37596	37596	37596
	avg_bewolking	37596	37596	37596
	avg_zonneschijnduur	37596	37596	37596

Correlations

		avg_zonneschijnduur
Pearson Correlation	lnPlaycountWeek	-.009
	lnAge	-.009
	lnAge2	-.009
	gender	.005
	avg_temperatuur	.632
	avg_bewolking	-.849
	avg_zonneschijnduur	1.000
Sig. (1-tailed)	lnPlaycountWeek	.039
	lnAge	.049
	lnAge2	.039
	gender	.168
	avg_temperatuur	.000
	avg_bewolking	.000
	avg_zonneschijnduur	.
N	lnPlaycountWeek	37596
	lnAge	37596
	lnAge2	37596
	gender	37596
	avg_temperatuur	37596
	avg_bewolking	37596
	avg_zonneschijnduur	37596