Erasmus
School of
Economics

Does offensive language sell?

Investigating the influence of offensive language on popular music.

Master Thesis
MSc. Data Science & Marketing Analytics
Erasmus University Rotterdam

Justin Albert-Paul Bor
409733

Supervisor	Dr. R. Karpienko
Second Reader	Prof. Dr. P. Groenen

The views stated in this thesis are those of the author and not necessarily those of Erasmus School of Economics or Erasmus University Rotterdam.

I. Abstract

This paper investigates the relationship between the use of offensive language and song popularity for music data from 2010 till 2019. This investigation is extended further by looking at the moderating influence of song genre on this relationship. Because no direct measure of offensive language is found, hate speech use is employed as a proxy variable. To answer the questions, data is gathered from the Spotify and Genius APIs. An NLP Transformer model, BERT, is used to classify the use of hate speech. The ANOVA and Tukey HSD test is used to compare the level of hate speech use between genres. Multiple Linear Regression models help answer if the direct and moderated relationship of offensive language on song popularity is found. The results show that an ambiguous relationship is found for the direct relationship between hate speech and song popularity, therefore lacking a conclusion regarding this relationship. Clear, distinct and significant relationships are found for the moderating relationships of song genres. Increased hate speech use in the Country genre yields higher song popularity, whereas this increased use leads to lower song popularity for Hip-Hop/Rap, Metal, and Rock music.

Keywords: Offensive Language, Song Popularity, NLP, BERT, Spotify, Genius

II. Table of Contents

1. Introduction

Offensive language is, according to the Oxford Dictionary, language that is "*rude in a way that causes somebody to feel upset or annoyed because it shows a lack of respect"* and *"extremely unpleasant"* (Oxford University Press, 2021). According to the Merriam-Webster Dictionary, offensive language is language that "*causes displeasure or resentment*" (Merriam-Webster Inc., 2021). Offensive language has been used for music for a long period. Although the artistic process that surrounds the production of a new song can be difficult to grasp, the linguistic aspect of it can prove to be insightful for the music industry when the right investigation techniques are applied. On one side, as with all aspects of the music-making process, the use of language can be seen as a form of artistry. Artists possibly make use of certain language because it aligns with their views of life. Furthermore, artistic decisions can be influenced and inspired, at an early stage, by their environment, family, or mentors (Weller, 2013).

On the other side, a reason for using offensive language in music could be the popularity an artist hopes to gain from the use of such language. Media presence plays an important role in the process of gaining popularity (Budzinski & Pannicke, 2017). As the use of offensive language leads to more shock amongst exposed listeners, the complimentary media presence received can be a strong factor in the choice for use of this language in music. Moreover, as the popularity of music is linked to industry revenues, it is an indirect measure of an artist's ability to generate streams, sell music, merchandise, and concert tickets (Kamara II, 2018; Galuszka & Wyrzykowska, 2016). If using more offensive language leads to more popularity, musical artists and publishers have a direct incentive to make more use of it, as this drives revenues. In this thesis paper, the relationship between the use of offensive language in songs and the popularity of these songs is investigated. The main research question is:

Main Research Question: *Does the use of offensive language increase the popularity of songs?*

A careful consideration of offensive language leads to a conclusion that is associated with other types of languages. Examples of these types of language include hate speech, but also sexual, violent, or profane language. According to the Oxford Dictionary, hate speech is "*speech or writing that attacks or threatens a particular group of people, especially based on race, religion or sexual orientation*" (Oxford University Press, 2021). While offensive language does not exclusively equal hate speech, hate speech can be considered offensive language. Other types of language that can be considered offensive language are sexual, violent, or profane language. Sexual language is "*connected with the physical activity of sex*", violent language describes "*the* *use of harmful or destructive physical force*" and profane language has or shows "*a lack of respect for God or religion*".

Since the 1950s, with the rise of Rock & Roll and later with the rise of Soul music in the 1960s, the use of offensive language in music has grown. During these first decades, this 'offensive' music mostly used sexual language. Examples of these artists include Elvis Presley, Curtis Mayfield, and Marvin Gaye (Marshall Cavendish Corporation, 2010). Although the use of these kinds of languages has not been popular with all listeners from the onset, the use of offensive language has become embroiled in controversy during the 1980s. Discussions regarding the explicit nature of song lyrics became a heated topic of discussion in the US. Parent and teacher organizations lobbied heavily for more stringent rules on what language could be used by artists in their music (Rolling Stone, 2015; NPR, 2010). In 1985, under pressure from lawmakers, the recording industry created the Parental Advisory Label (PAL). Although the PAL explicit label is found on many album covers, the placement is voluntary (NPR, 2010).

An important argument from advocates for labeling music is the protection of its listeners. The science relating to the psychological effects of music finds that ideas discussed by artists become activated within their listeners. This activation is triggered by regular exposure to these ideas (Gannon, 2009; Luszcynska & Schwarzer, 2005). Furthermore, popular music artists play an important role in the lives of their fans. These listeners infer behavior and acquire beliefs based on these role models, including music artists (Martino et al., 2009). Another aspect of labeling music is the influence music could have on the views of children and young adults. The long-term exposure to hate speech can lead to radicalization or reduce empathy for other societal groups and leads to the erosion of anti-discriminatory norms in society (Bilewicz & Soral, 2020).

The use of offensive language is not equally distributed across music genres (Hart & Day, 2020). Specifically, when looking at the use of hate speech, research shows that racist or antisemitic hate speech is present in some Country and Rock music. "Nazi Rock" is a prominent form of a hate speech music genre that emerged in Great Britain during the 1970s. Over time, the rest of Europe has become influenced by this genre (Brown, 2004). Music was seen as a vehicle of promoting neo-Nazi ideas that align with far-right extremist views (Cotter, 1999). Furthermore, even before the emergence of "Nazi Rock", "Country hate music" made its entrance in the 1960s. Most of this hate speech-oriented music in the country genre was released by smaller, less known music labels (Messner et al., 2007). To investigate if the level of offensive language differs between genres, the first sub-question is:

Sub-question 1: *Do genres differ in the use of offensive language?*

If different levels of hate speech are found between genres, it will become meaningful to investigate if the song genre has a moderating influence on the relationship between offensive language and song popularity. Therefore, the second sub-question states:

Sub-question 2: *Does the song genre moderate the influence of hate speech use on song popularity?*

The research question is answered by performing research on data that is gathered using APIs from Spotify and Genius. Multiple data sources are needed, since not one source of data is available that contains both lyrical data on songs and their popularity. A description of the gathering, manipulation, and application of the data can be found in the data section. Furthermore, a Natural Language Processing (NLP) method is employed to classify the type of language used in the lyrical data. With the use of this data set, statistical methods are employed to answer the research questions and test the hypotheses. In this thesis paper, the Pearson Correlation Coefficient (PCC), Multiple Linear Regression (MLR), one-way Analysis of Variance (ANOVA) test, and Tukey HSD (T-HSD) test are used. These methods are discussed in the methodological section. The findings are presented in the results section. The results are interpreted in light of existing research in the discussion section. Afterward, the main findings, but also limitations to this paper and the data set used are discussed in the conclusion and limitation section.

2. Theoretical Framework

As the music industry generates a high amount of revenue yearly, the popularity of the artist in the music industry is an important aspect for them, managers, the record labels, and the publishing companies. Popularity is directly associated with the reach and potential fandom of an artist. More popularity allows for higher marketing potential and therefore revenues (Bellogín et al., 2013). If popularity does not directly translate to more revenue, it indirectly increases the chance of an artist to market him or herself to their audience.

Sources of income for artists are both music distribution (physical or digital), but also concerts (Galuszka & Wyrzykowska, 2016). Nowadays, most of an artist's income originates from concerts. Artists try to be as popular as possible since the music industry is characterized as a superstar industry. Superstar industries have a small portion of 'stars' that earn most of the income within a certain industry. In the case of the music industry, the top 5% of artists in the music industry have a better negotiation position with music publishers and are responsible for 60 – 80% of revenues of concert tickets in the music industry (Connolly & Krueger, 2006). As concerts are the most profitable income source for an artist, becoming more popular should be a core focus of musicians.

Besides profiting from an artist's popularity with concerts, new contracts in the music industry, known as 360-degree deals, are becoming commonplace. These 360-degree deals are more complex and capture a bigger part of artist revenue, compared to simple publishing contracts that were used in the past decades (Galuszka & Wyrzykowska, 2016). Publishing labels are taking cuts of multiple sources of revenue, including concerts, merchandising, brand endorsements, and other types of income from the artist (Byun, 2016). Artists at the start of their career have a weak negotiation position compared to big music labels, that are operating in an oligopolistic market. Only artists with an established fan base, that rise to become the most popular artists, receive a relatively high percentage of the music royalties from the sale of music (Connolly & Krueger, 2006). An important reason for the increasing use of these 360-degree deals by record labels and music publishers is the mitigation of sales losses. These sale losses occur due to the switch from physical sales of music to a streaming-based revenue model (Rogers, 2017).

Not only popularity is an important aspect in the music industry. Elements like tempo, valance, or danceability are parts that make up a song. Another element that makes up the art form referred to as music, is language. Spoken language is used in music, mostly in the form of singing or rapping. Not all forms of music make use of spoken words. Furthermore, the use of language is an important element of music because it can transfer a linguistic message to its listeners, allowing the artist to transfer idea's, events, or emotions to its listeners and making it more than just a collection of sounds (Bright, 1963; Powers, 1980). With their music, artists can convey acts or ideas with which they can offend. They can use strong language to insult or shock.

A type of offensive language that can be used to insult or shock, is hate speech. It is a type of language that is negatively targeted at specific groups. Examples include minorities, people of a different race, gender, or other elements that deviate from what is considered normal in society (Levy et al., 2000). Although generally hate speech is used by a select few that chose to publicly preach their hate ideas, the internet has allowed the existence of a global racist subculture. (Back et al. 1998). The increasing use of the internet has allowed many hate speech-related websites to take off. Starting with the launch of Stormfront.org in 1995 by a Klu Klux Klan member, the use of and exposure to hate speech has increased. In 2010, the Simon Wiesenthal Center estimated around 8000 hate-related websites were live. Next to the web, social media channels have become breeding grounds for hate speech, allowing users to more easily and unanimously spread their hate ideas to other users on these social media platforms (Banks, 2010).

Offensive language, like hate speech, is not only used on online websites, but also in music. Research points to some Country, Rock, and Hip-Hop/Rap artists making use of offensive language, hate speech in particular. Messner et al. (2007) find that some Country music is used by white nationalists to "dehumanize" African-Americans and labels this music as 'Country Hate Music'. Although this music does focus heavily on hate against African-Americans, it also features negative sentiment against the federal government or hippies. Messet et al. (2007) explicitly identify confrontational hate music as the explicit use of hate, racism, or white supremacy in music. 'Hate Rock' describes the use of hate speech in some of the Rock genre. Where Country hate music focuses mostly on people that identify as 'Dixies' (a popular nickname for the population of the Southern United States), hate rock is usually played by racist skinheads or neo-nazis (Messner et al., 2007). Hate Rock originates from the early 1960s in Great Britain and was successful because of its ability to create a source of identity for skinheads. Hip-Hop/Rap is another genre that is associated with hate speech. A large amount of Hip-Hop contains hate speech, mostly directed at groups of people. Some media outlets have even speculated that hate speech could be to blame for the use of violence or vandalism. The use of hate speech is partly the result of the disconnect that exists between modern-day artists and their fans, mostly due to the invention of modern-day recorded music. A few trends in the use of hate speech in Hip-Hop/Rap have become visible over time. In the early 1990s, rap was used as an artistic form of rebellion, usually directed at white authoritarian figures. Later misogyny created anti-feminist sentiment. Lastly, homophobia was another prominent element of Hip-Hop/Rap songs (Kenvarg, 2021).

Hate speech is just one type of language that is considered offensive. Other forms of offensive language include, but are not limited to sexual, violent, or profane language. Offensive language and the music industry have a long history when it comes to censorship (Marshall Cavendish Corporation, 2010). Research on the use of sexual, violent, and profane language in music reveals that this type of language is still frequently used in music. Close to 33% of all music contains sexual references (Primack et al., 2009). Hart & Day (2020) find lower numbers in general, although the researchers are quick to note that implicit references, sarcasm, or slang are not taken into account. No research is found that focuses specifically on the use of violent or profane language in music.

Since offensive language, like hate speech or sexual, violent, or profane language can be a cause for labeling songs as explicit music, research into this field might extend the idea about how much offensive language is used in music. Research papers that use explicit musical or lyrical material focus on explicitness detection using a wide variety of machine learning algorithms. These research papers make use of unbalanced databases, to accurately represent the level of explicit materials in music overall. Based on lyrical data from Lyricfind.com, 13% of all lyrics contain an explicit label (Bergelid, 2018). For Spotify data, researchers found that 7,7% of all lyrics contain an explicit label (Rospocher, 2021). Based on the explicitness classification by the Korean Ministry of Gender Equality and Family, classified songs contain 11,9% explicitly labeled content (Kim & Yi, 2018). Using the WASABI database, an average of 9,9% of all songs are labeled as containing explicit material (Fell et al., 2019). It can therefore be concluded that in general, 7% to 15% of all lyrics have an explicit label, containing some form of offensive language.

Since no other research focuses on the relationship between the use of offensive language in song lyrics and the popularity of these songs, conclusions might be drawn from other research that focuses on violence, profanity, or sexuality in other types of creative industries. Lang & Switzer (2008) focus on the effect of sexual or violent content on the box office revenues of movies in the United States and other (foreign) countries. They find that even though less violent or sexual movies return higher movie revenues, movie producers choose to produce increasingly more sexual or violent movies. This artistic choice is explained by the fact that the non-American movie audience becomes an increasingly important part of the movie revenues earned. This foreign audience is more attracted to violent and sexual movies. In particular, violent movies positively impact movie revenues in non-American movie theaters. An increase in sexual content attracts the middle-aged population, but negatively impacts viewership by elders. For the advertisement industry, effects are many-sided and do not provide clear results. Differing effects are found for research relating to the use of sexual content in advertising and gender (Gramazio et al., 2021; Zawisza et al., 2018) or age (Lundstrom & Sciglimpaglia, 1977). Other researchers prove that using sexual content in advertising has adverse effects (Lull & Bushman, 2015; Parker & Furnham, 2007).

Little research exists regarding the popularity of music and how it is affected by elements like offensive language. Careful consideration must be made about what effect occurs between the use of offensive language in songs and the popularity of these songs. Based on findings by Lang & Switzer (2008) concerning the use of sexual or violent content in movies for non-American audiences and some research findings concerning the use of sexual advertising, this thesis paper hypothesizes that the use of offensive language in music positively relates to the popularity of that music. Therefore, the first hypothesis states that:

Hypothesis 1: *The use of offensive language in a song positively relates to the popularity of that song.*

Not all song genres make equal use of offensive language or content. Even though close to one-third of music contains sexual content (Primack et al., 2009), this explicit material is not uniformly distributed across genres (Hart & Day, 2020). Some genres show to contain more sexual material than others. Especially Hip-Hop/Rap features a masculine, tough, and manly culture. The portrayal of gender inequality that is a result of this masculine culture (Iwamoto, 2003), leads to the idea that Hip-Hop/Rap contains more offensive language as compared to other genres. This belief is enhanced by the fact that a large amount of Hip-Hop contains hate speech (Kenvarg, 2021). For Rock or Metal, the use of violent terms can be overrepresented compared to other genres. A significant difference is found for sexual content between different musical genres. Rap contains 3 times as much sexual content as Pop and 13 times more sexual content as compared to Country music (Hart & Day, 2020). The topics that are explicitly mentioned more in Hip-Hop/Rap music are sex, objectification of women, and general explicitness (Smiler et al., 2017).

Hip-Hop/Rap is overrepresented when it comes to the portrayal of violent and sexual content in music videos. Especially, discussions about gun use, drug use, or (physical) grabbing are overrepresented in Hip-Hop/Rap music videos. Moreover, Hip-Hop/Rap is more likely to contain curse words compared to other genres. For other genres, like Rock or Country music, the use of alcohol is a reoccurring theme. When it comes to sexual themes, Hip-Hop/Rap is more likely to discuss sexual acts, feature men without shirts, but also have females dance sexually in the music videos. Although Rock is generally not seen as a sexually expressive genre compared to the Rock and Roll of the 1950s, nudity is still an important part of the Rock culture. Rock music specifically features a lot of artists that are undressed, therefore also showing sexual content to its viewers. Country music shows more cleavage than genres like Rap, Hip-Hop, or Rock music (Jones, 1997).

Musical genres differ in the themes and topics they address, including the use of offensive language. Research proves that a difference between musical genres is found, for both the language used in music, but also the video clips. Hip-Hop/Rap music lays focus on the masculinity of its rappers, while Rock features undressed singers. Therefore, it is expected that the level of offensive language will differ between genres. The second hypothesis states that:

Hypothesis 2: *The amount of offensive language used in music differs per musical genre.*

As musical genres contain differing levels of offensive language, it is important to understand what lies at the basis of this difference. As an artist starts their career or looks for a boost in popularity, they might look at other, more developed artists for a basis of what kind of musical material to produce. Furthermore, artists might use established norms in their respective genres to decide on what themes are important for them. An important aspect in people's behavior and origins of norms is what they see in their surroundings. Norms that are established by a group of people are called social norms. In general, a reason for sticking to social norms is the fact that collective wisdom is helpful to individuals, as well as to society (Lapinski & Rimal, 2005).

The music industry might also harbor continuing social norms about what is expected of artists. Especially within specific genres, there is space for assumptions about what an artist should or is expected to do. Certain genres might be associated with a specific tempo, dress code, outlook on the world, but also the use of language within its music. Social norms could even be understating the level of mimicry that takes place within the music industry. This can be derived from the fact that musical stereotypes are broadly researched within the scientific field. Stereotypes mostly focus on the behavior of listeners that relate to a specific musical genre. Listening behavior, for example, can reveal information about one's personality traits or strengthen beliefs about one's personality or identities. People with an energetic personality will be more inclined to like commercial music, whereas people with an open personality might prefer to listen to more complex genres, like Jazz or Classical music (Rentfrow & Gosling, 2007). Furthermore, listening behavior can be linked to being part of a certain community or group. Research suggests that music listening behavior differs based on income or social class. High social class is associated with genres like Classical or Big Band music, whereas low social class individuals listen to genres like Country or Rap music (Katz-Gerro, 2002).

Because of their need to associate with their base of dedicated listeners, artists might want to conform to these stereotypes to fit the picture of a certain type of artist, consistent with Jones (1997). A Hip-Hop/Rap artist might want to be associated with the use of alcohol, drugs, expensive cars, or want to sexualize women. Rock artists might want to associate their acts with violence, anger, or pain and be topless while performing on stage. Stereotypes could be problematic because of behavior that is activated in their listeners. They create an association between social groups and a listener's/artist's view of a certain musical genre. Furthermore, because of the incomprehensible use of language in Rock and Hip-Hop music listeners are more inclined to base their understanding of a genre on the stereotypes that are present (Neguţ & Sârbescu, 2014). Extending this further to artists that are at the start of their career, they also could be influenced to adapt their style or artistry to the stereotypes that are present in their genre.

Artistic musical decisions are influenced by these social norms and stereotypes. Elements of their image or musical performance are based on what artists perceive as to be 'standard' within their type of genre. The level of offensive language in their music can be attributed to the social norms and stereotypes within a certain musical genre. It is therefore assumed that the level of offensive language within a particular genre follows a certain base level, based on these norms and stereotypes. Therefore, it is assumed that if offensive language use influences the popularity of a song, this relationship is moderated by the genre of the song. The third hypothesis states that:

Hypothesis 3: *The genre of a song moderates the effect between the level of offensive language and popularity.*

Part of the research design of this thesis paper is shown in Figure 1. A complete description of the research approach used can be found in the Methodology section.

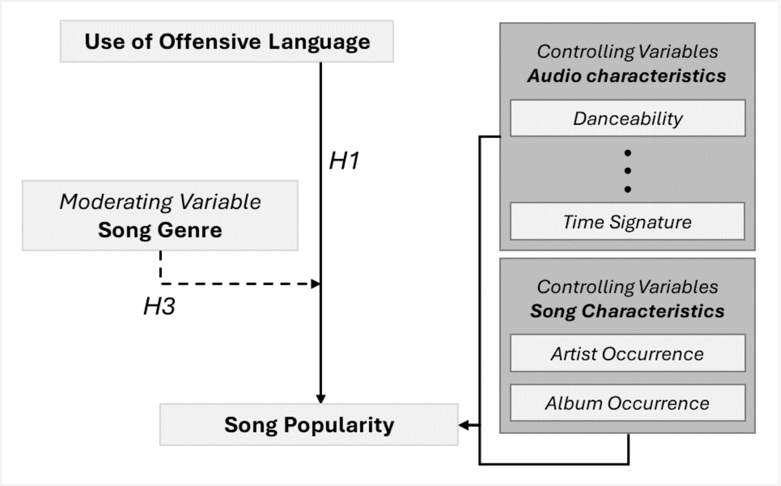


FIGURE 1: VISUAL REPRESENTATION OF THE CONCEPTUAL FRAMEWORK

3. Data

3.1 Data collection

As there is no one collection of music features and lyrics, multiple sources of data are needed to answer the research question. Many different sources of musical data are available for research, varying widely in their use case and quality of data. While selecting appropriate data sources in this paper, sources were weighed based on the available amount of data, user base, the accuracy of its information, and access to the API. After considering multiple sources of musical data, the Spotify and Genius API were selected. The Spotify API allows for the collection of random songs and specific information about them, like the song's popularity or release date. The Genius API returns a collection of individual song lyrics.

Spotify is the world's largest streaming provider. According to Statista (2021), Spotify had 172 million paying premium subscribers in the third quarter of 2021. In terms of its availability of music, Spotify currently has the biggest collection of digital music that is available for streaming, offering over 70+ million songs in its library. Other worldwide competitors of Spotify are Apple Music, Deezer, Tidal. Furthermore, Spotify also has regional competitors, like the Chinese QQ Music. Not only streaming services are competing with Spotify. Currently, YouTube, a video-sharing service, is the largest music listening platform in the world (IFPI, 2021). Next to music, Spotify has invested heavily in the podcasting industry over the last few years, as its goal is to become a center point of the general listening experience of consumers.

Genius (Genius Media Group Inc.) is a multimedia company that started as RapGenius, mostly basing its operations on providing Hip-Hop/Rap based lyrics to listeners. Currently Genius produces a multitude of audiovisual content surrounding music, including interviews with artists. Next to music lyrics, Genius provides transcripts of news events, sports, and historic events. Competitors of Genius include Musixmatch and Lyrics.com. Genius stands out from other lyric-focused companies by the big network of users that update lyrics and fill in explanations for each specific sentence in a song. Artists of these songs can verify these explanations or fill in their own. This leads to higher accuracy of lyrics compared to other lyrical platforms, decreasing the chance of mistakes in the lyrics.

4.1.1 Spotify API

The Spotify API lets users input requests on a song, album, or artist level. Random songs were collected with the API's search function. The search was not completely random, some arguments were passed on in the search. First, songs were gathered that have a release date between 2010 and 2019. This period was chosen because of the makeup of the popularity variable that is scraped from Spotify, which is described further in the data description section. Because this variable is biased towards plays in the near past (Spotify, 2021), an earlier period is selected to partly negate this effect. The US music market is selected, as the focus of the research lies on English lyrics. In line with the research goal and based on already identified trends in the use of hate speech in music (e.g., Brown, 2004; Cotter, 1999; Messner et al., 2007), song data was collected for a selected number of genres. These genres are Hip-Hop/Rap, Country, Rock, and Metal. To limit the number of duplicated values and increase randomness, an offset was used in the search request. Even though an offset was used, a high number of duplicated values were returned. To lower the number of duplicated values that were found further, multiple random vowels (wildcards) were passed through in the search request. With the introduction of wildcards, the randomness of results that were returned by the API increased. Even after the use of wildcards, a significant proportion of the returned observations were duplicates, as described in the following sections.

Two rounds of API requests were needed to gather all essential information. Using the API's search function, songs' related genres could not be gathered. Therefore, during the second round of requests, songs' genres were collected. After the song genres were collected, a check was performed to see if the songs, as scraped with the search function, had the initiated genre. In total 18,000 requests were made to the API in the first round, collecting 50 songs each, which would lead to a theoretical data set of 900,000 songs. As mentioned earlier, not all data returned by the API was useable. Some requests returned empty cells, mostly caused by network interruptions. Moreover, duplicated data was collected due to the randomness of the search. Data collected with the request included song names and IDs, artists' names and IDs, album names, and the popularity score of the song.

Furthermore, data was collected regarding the audio features of each song, e.g., describing the danceability or tempo of a song. These variables are specific to each song and are created with help of Spotify's data science team. The variables collected can be found in Table 1. Because of the way the API returned data and the specification for songs between 2010 and 2019 for four different musical genres, some collected songs share an album or artist name with other songs. In total 74.03% of songs share the album with at least one other song in the data set. For artists, 94.27% of songs share the artist with at least one other song in the data set. To control for the effect of an artist or album on the popularity of a song, control variables were introduced in the data set, as discussed in the conceptual framework and the data description. After successfully executing the data collection using the Spotify API, 388,950 songs were eventually found to be unique. This means that a loss of 56.78% occurred compared to the theoretically viable set of 900,000 songs. These observations were passed on to the Genius API.

4.1.2 Genius API

Scraping of lyrics from the Genius website was done with the use of the *geniusr* R-library (Henderson, 2021). Genius does not make use of the industry-standard song identifier, the ISRC. Therefore, two rounds of data gathering were needed to gather both Genius' assigned song identifier and the lyrics. During the first round of data gathering the song and artist name, as returned from the Spotify API, were input in the search function of the Genius API. Not all song titles and artists' names matched exactly. Therefore, to control for any matching errors, a similarity measure was employed. The *levenshteinSim*-function was used from R's *RecordLinkage* library (Sariyar, 2021). This function is based on the Levenshtein distance and is calculated by:

$$1 - \frac{d(string\ 1, string\ 2)}{max(length\ string\ 1, length\ string\ 2)} \tag{1}$$

where d is the Levenshtein distance function. This distance function is a metric that describes the minimal number of insertions, deletions, or substitutions that are needed to change string 1 into string 2, or vice versa. The function outputs a value between 0 and 1 based on the similarity value calculated with the function. An accuracy level of 0.75 for the song name and 0.85 for the artist's name was maintained. Songs that did not exceed this boundary were removed from the data set. After this removal, 168,320 observations were left, meaning a further loss of 56.72%. During the second round of data collection, Genius' song IDs were used to gather lyrics of Genius' website. A loss of data occurred because not all Genius IDs contained lyrics, some IDs returned empty observations. Because of this loss, a total of 98,414 observations remained, meaning an additional loss of 41.53%. Furthermore, not all songs returned by the Spotify API use the English language. Therefore, after the data collection was completed, the R-library *textcat* was used (Hornik, et al., 2013). The *textcat* package employs an n-gram based text categorization that can filter for 74 different languages. Whereas N-gram usually split a sentence into groups of words, in the *textcat* function N-grams are an N-character slice of a string. The function simultaneously employs bi-, tri-, and quad-grams (e.g., 'HE', 'ELO' and/or 'ELLO') and compares the frequency of these N-grams of a text to averages of languages. With this comparison, the function can estimate a text's language (Cavnar & Trenkle, 1994). After filtering for the English language, 45,212 observations were left (54.06% loss). The lyrics contain a combined total of 14,848,953 words, making up the data set as used further in this thesis paper. Data loss occurred at multiple points in the data-gathering phase. Figure 2 gives an overview of the steps that were taken during the data gathering phase and the occurrence of data loss at some of these steps.

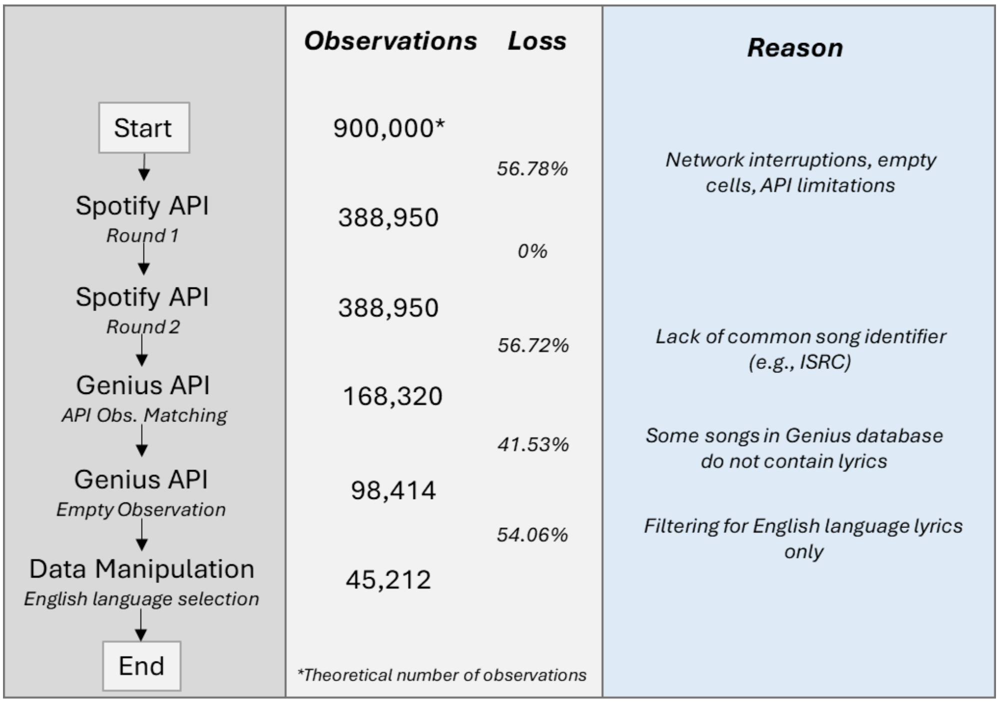

Observations
Loss
Reason
Start
900,000*
56.78%
Network interruptions, empty cells, API limitations
Spotify API
Round 1
388,950
0%
Spotify API
Round 2
388,950
56.72%
Lack of common song identifier (e.g., ISRC)
Genius API
API Obs. Matching
168,320
41.53%
Some songs in Genius database do not contain lyrics
Genius API
Empty Observation
98,414
54.06%
Filtering for English language lyrics only
Data Manipulation
English language selection
45,212
End
*Theoretical number of observations

FIGURE 2: AN OVERVIEW OF SCRAPING PROCESS AND OBSERVATION LOSS DURING DATA GATHERING

3.2 Data Description & Data Manipulation

In total, the data set contained 26 variables. A description of these variables can be found in Table 1. These variables originated from the Spotify API, Genius API, data manipulation phase, or the result from the analysis phase. Because songs from the same artist or album were part of the data set, 5 descriptive identifiers were part of the data set. The Spotify *(spotify_id)* and Genius *(genius_id)* identifiers helped identify the song within the respective APIs, furthermore the song *(song_name)*, album *(album_name)* and artist *(artist_name)* names were part of the data set. These three last identifiers originated from the Spotify API. Another descriptive variable is *release_date*, which describes the year in which a certain song was released. This variable varies between 2010 and 2019, which aligns with the search specifications as described in the data collection.

The dependent variable of this thesis paper is *ln_popularity,* which is ranges from 0 to 4.331. This variable is based on the *popularity variable*, which is an integer value that theoretically should range between 0 and 100 and derives from Spotify. The *popularity* variable was left-skewed, which is visible in Figure 3. To transform the *popularity* variable to *ln_popularity*, 1 was added to the *popularity* variable, in accordance with Bellego et al. (2021), after which the natural logarithmic value of this resulting value was calculated. According to Spotify, the *popularity* measure is based on the number of plays a song has received and the recent nature of those

	Variable name	Type of data	Description	Data Source
Identifier	*spotify_id*	character	Unique identifier	Spotify API
	genius_id	character	Unique identifier	Genius API
Song Descriptives	*song_name*	character	Song name	Spotify API
	artist_name	character	Artist name	Spotify API
	album_name	character	Song name	Spotify API
	lyric	character	Song lyric	Genius API
Variables of Interest	*ln_popularity*	integer	Natural log. value of the (song popularity + 1), values between 0 and 4.331	Spotify API
	hate	double	Hate speech probability, values between 0 and 1	Data Analysis
Other Variables	*release_date*	integer	Release date, values between 2010 and 2019	Spotify API
	genre	categorical	Genres included: Country, Hip-Hop/Rap, Metal, Rock	Spotify API
	lyric_length	integer	Length of the song lyric, values between 1 and 7079	Data Manipulation
Song-specific Variables	*artist_occ*	integer	Co-occurrence of song' artist in data set, values between 1 and 210	Data Manipulation
	album_occ	integer	Co-occurrence of songs' album in data set, values between 1 and 58	Data Manipulation
Audio-based Variables	*danceability*	double	Ability to dance to a song, values between 0 and 1	Spotify API
	energy	double	Intensity and activity of song, values between 0 and 1	Spotify API
	key	categorical	Key of the song, values between 0 and 11	Spotify API
	loudness	double	Loudness in dB, values between -60 and 0	Spotify API
	mode	boolean	Modality of the song, major is 1 and minor is 0	Spotify API
	speechiness	double	Number of spoken words in the song, values between 0 and 1	Spotify API
	acousticness	double	Confidence measure of acousticness, values between 0 and 1	Spotify API
	instrumentalness	double	Amount of instrumentalness, values between 0 and 1. Values towards 1 indicate instrumental songs.	Spotify API
	liveness	double	Presence of audience in a recording, values between 0 and 1. Values towards 1 indicate a live performance recording	Spotify API
	valence	double	Positivity in song, values between 0 and 1	Spotify API
	tempo	double	Average tempo in a song (measured in BPM), values between 0 and 220 BPM	Spotify API
	duration_ms	integer	Duration of a song is measured in milliseconds. Values between 4,027 and 2,172,760	Spotify API
	time_signature	integer	Number of beats in each bar, values between 0 and 5	Spotify API

TABLE 1: DESCRIPTION OF VARIABLES INCLUDED IN THE DATA SET. DESCRIPTIONS ARE PARTIALLY BASED ON SPOTIFY (2020).

plays, therefore adding a discriminative factor to songs that were released at an earlier date compared to newer songs. Because of the make-up of this variable, this thesis paper focuses solely on data that is less recent (e.g., not released in 2020 or 2021). This is done to combat the heavyweight of more recently released and therefore popular songs on the popularity variable. Descriptive statistics for the *ln_popularity* and other numeric variables can be found in Table 2.

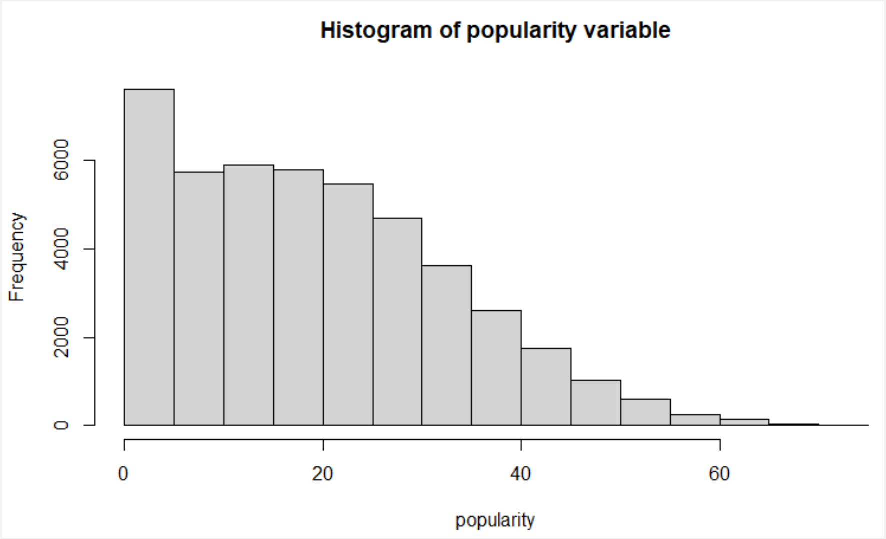


FIGURE 3: HISTOGRAM OF THE DISTRIBUTION OF THE POPULARITY VALUES

		Minimum	Maximum	Median	Mean	Std. Dev.
Variables of interest	*ln_popularity*	0	4.331	2.944	2.706	0.974
	hate	0.022	0.908	0.142	0.26	0.252
Other Variable	*length*	1	7079	269	328.429	220.758
Song-specific Variables	*artist_occ*	1	210	11	17.421	22.268
	album_occ	1	58	3	3.859	3.752
Audio-based Variables	*danceability*	0	0.985	0.535	0.534	0.169
	energy	0	1	0.735	0.701	0.219
	key	1	12	6	6.243	3.581
	loudness	-30.107	2.363	-6.343	-7.035	3.209
	mode	0	1	1	0.68	0.467
	speechiness	0	0.964	0.056	0.107	0.112
	acousticness	0	0.996	0.054	0.193	0.266
	instrumentalness	0	0.994	0	0.082	0.213
	liveness	0	0.999	0.146	0.223	0.186
	valence	0	0.983	0.428	0.446	0.233
	tempo	0	220.085	122.012	123.186	29.954
	duration_ms	4027	2172760	225200	238731.5	87884.003

TABLE 2: DESCRIPTIVE STATISTICS OF NUMERICAL VARIABLES IN THE DATA SET

The *lyric* variable contains the song lyrics, as collected with help of the Genius API. To ease the work of the word embedding model and prevent the failure of recognizing the same words, but spelled with a capital or lower-case letter, all words are converted to a lower-case form. Furthermore, since the relationship between words can be clouded by the introduction of excess punctuation, these are removed, as well as all other non-ASCII text in the lyrics. After the manipulation is complete, the *lyric_length* variable is created. This variable indicates the word length of each lyric.

This thesis paper investigates the use of offensive language and its effects on the song's popularity. Offensive language consists of a broad group of languages that are considered rude, upsetting, or unpleasant. Therefore, hate speech is used as a measure of offensive language use. The remainder of this thesis paper will refer mostly to hate speech use for analytical steps and interpretation, but this variable should be interpreted in light of its function as a proxy of offensive language. The *hate* variable is generated with the use of the BERT NLP model and describes the hate speech probability of the song lyrics. A extensive description of the BERT NLP model can be found in the Methodology. The *hate* variable is right-skewed and has a median value of 0.142.

Since this thesis paper investigates the effect of offensive language, specifically hate speech for different musical genres, a categorical variable *(genre)* is present that indicates if a song is part of the Country, Hip-Hop/Rap, Metal, or Rock genre. Each genre is represented relatively equally in the data set, not making the overrepresentation of one specific category problematic. 20.44% of all observations originate from the country genre, 25.98% from the Hip-Hop/Rap genre, 24.81% from the Metal genre, and 28.77% from the Rock genre.

Because songs in the data set originate from the same artist or album, two additional variables are created, the artist occurrence *(artist_occ)* and album occurrence *(album_occ)*. The distribution of these variables can be found in Figure 4. These variables control for the effect of the presence of the same album or artists for multiple different observations in the data set. Next to artist and album occurrence, other controlling variables are used in this thesis paper, namely the audio features of each song. These variables are included to control for the effect the musical side of a song has on its popularity.

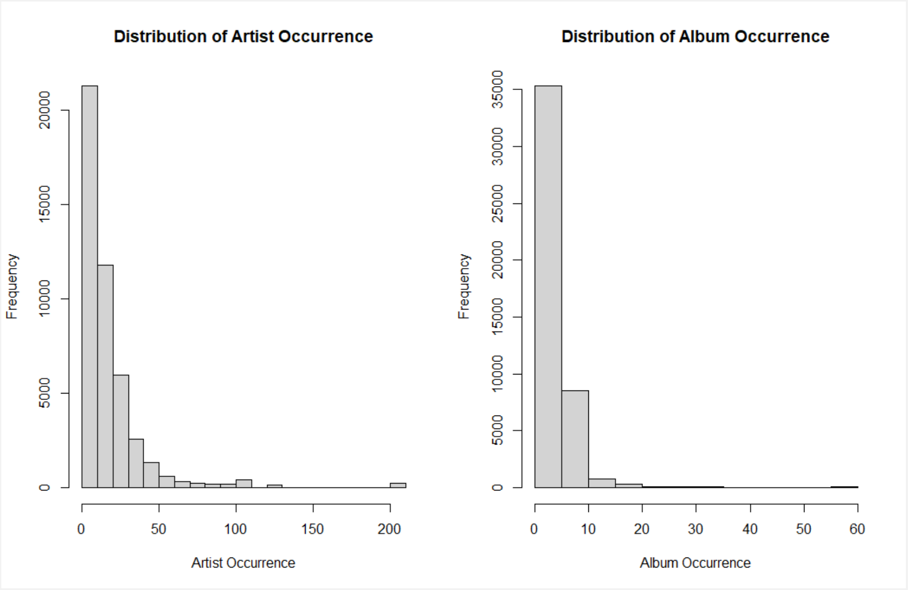


FIGURE 4: DISTRIBUTION OF THE ARTIST AND ALBUM OCCURRENCE VARIABLE

4. Methodology

After data was gathered with the use of the two APIs, as described in the Data section, the task of classifying song lyrics as hate speech remained. The classification of song lyrics was done with the use of the Bidirectional Encoder Representations from Transformers (BERT) model. The hate speech probability that resulted from this classification task was input as an independent variable into two of the final models, which were multiple linear regression (MLR) models. These multiple linear regressions allowed for capturing of the possible relationship between the use of hate speech in songs and the popularity of these songs. Moreover, the moderating effects of the song genre could be investigated with these models. Other tests, like the Pearson Correlation Coefficient (PCC) test, an Analysis of Variance (ANOVA) test, and a Tukey HSD (T-HSD) test were also employed. The PCC test was used for an exploratory investigation of relationships at play in the data set. To determine if levels of hate speech differ significantly per genre and answer the first sub-question, a one-way Analysis of Variance (ANOVA) test was employed. Furthermore, a Tukey HSD test was performed, allowing for the testing of hate speech probability differences between genres. The ANOVA test's and Tukey HSD test's results helped answer if a difference in the use of hate speech probability exists between various musical genres and what the size of this difference is. The BERT, PCC, ANOVA, T-HSD, and MLR models are described in the following section, together with a description of the steps that were taken during the analysis phase.

4.1 Bidirectional Encoder Representations from Transformers

For the classification of lyrics into hate or non-hate speech, a Bidirectional Encoder Representations from Transformers (BERT) model (Devlin et al., 2018) was used. This model was pretrained by Aluru et al. (2020) and relied on the PyTorch Tensor library (PyTorch, 2021). During the pretraining process, Aluru et al. passed six different hate speech labeled data sets into the BERT model. Five out of six data sets find their origins on Twitter. One dataset originates from Stormfront.org, the earlier described forum for white supremacists. All data used in the pre-trained model had a binary hate or non-hate label (Aluru et al., 2020).

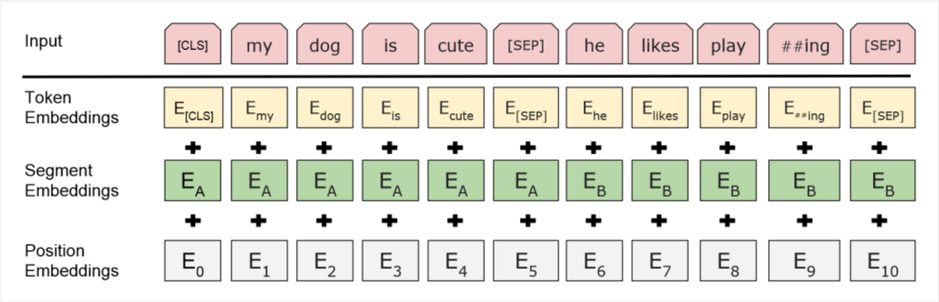
After the BERT model was pretrained on hate speech labeled text, the song lyrics, as present in the data set, could be input into the pre-trained model to retrieve the hate speech probability of each song. Where Aluru et al.'s most complex model could identify hate speech from multiple languages, the model used in this thesis paper was only able to recognize English language hate speech, and is known as a monolingual model (Aluru et al., 2020). The section below describes the BERT model and elaborates on how data was passed through the model in order to generate hate speech probabilities.

BERT is part of a group of Natural Language Processing (NLP) methods. Different types of NLP methods exist, like the Word2Vec (Mikolov et al., 2013) or Global Vectors (GloVe; Pennington et al., 2014) model. These models use words in the near vicinity to a focal word (e.g., the window of 2 words before and after a focal word) to model and infer the mathematical meaning of this focal word. This mathematical meaning is represented in the form of a word embedding. Word embeddings are vectors that contain predefined dimensions of values, with each value representing the weight of a focal word for a certain dimension. Although these types of NLP models have become popular and are relatively new (Word2Vec was released in 2013 and GloVe in 2014), rapid innovations in the NLP space have led to the development of more complex models.

One such complex model is BERT. BERT originates from Google and is a state-of-the-art Transformer model. Other examples of Transformer models are the Generative Pre-trained Transformer (GPT; Radford et al., 2018), DistilBERT (Sanh et al., 2019), and GPT-2 (Radford et al., 2019). These models are based on an Encoder or Decoder architecture. In some cases, an Encoder-Decoder architecture is used. Different architectures are used for different NLP tasks. A Decoder-based Transformer is useful for text generation, while an Encoder-Decoder Transformer is used for text summarization or translation. Sentence classification or name entity recognition is done with an Encoder-based Transformer (Gugger, 2021). Because this thesis paper aimed to classify text, the use of the Encoder-based BERT Transformer model was deemed appropriate.

The architecture of BERT is more complex than just a single Encoder (block). In total, the BERT model in this paper made use of 12 encoder layers, also known as Transformer blocks. This means that classified text input into the model passed through 12 encoders before being output by the BERT model. Encoder mechanisms in Transformer models consist of a self-attention layer and a feed-forward neural network. The goal of the self-attention layer is to imitate the understanding that humans have regarding the context of words. Similar to humans, Transformer models can understand not only the context between words but also the context of words in relation to other sentences and the relationship sentences have with each other.

To input information into the first self-attention layer of the BERT model, the input of word tokens is required. Tokens contain information representing a specific word or contain information representing the position of a sequence or sentence. These informative tokens are the sequence-start ([CLS]) and separator ([SEP]) tokens. The sequence-start token indicates the start of a text, and the separator token indicates the start of a new sentence (Devlin et al., 2018). Furthermore, next to inputting token information regarding the position of words compared to other words, additional information is created based on the input token. This additional information indicates what sentence a token is a part of, known as Segment Embeddings. Moreover, the token contains information about how its position, relative to other tokens in a text, known as the Position Embedding. The Token Embedding, Segment Embedding, and Position Embedding are grouped into a Tensor, which is then input into the BERT model. In this thesis paper, instead of inputting complete song lyrics, the length of each song had been limited to a maximum of the first 512 words. This word limit was caused by the BERT model's constrained input of a maximum of 512 words. Therefore, during the tokenization phase, which translated the lyric's words into a mathematical representation, the *truncation* argument (HuggingFace, 2021) was used to cut down the size of the lyrics into a maximum of the first 512 words of a song lyric.



Input
[CLS]
my
dog
is
cute
[SEP]
he
likes
play
##ing
[SEP]
Token Embeddings
$E_{[CLS]}$ E_{my} E_{dog} E_{is} E_{cute} $E_{[SEP]}$ E_{he} E_{likes} E_{play} $E_{\#\#ing}$ $E_{[SEP]}$
+
Segment Embeddings
E_A E_A E_A E_A E_A E_A E_B E_B E_B E_B E_B
+
Position Embeddings
E_0 E_1 E_2 E_3 E_4 E_5 E_6 E_7 E_8 E_9 E_{10}

Figure 5: Graph from Devlin et al. (2018), showing the architecture of a BERT Tensor

The self-attention layers make use of two different mechanisms to optimize its understanding of an input word compared to other words and sentences in a text. These mechanisms are employed during the pretraining of the model, after which weights in the self-attention layers are established. After pretraining, passing data through the self-attention layer makes use of these pre-established weights. First, a masking technique is employed. A random percentage (usually 15 %) of words in a text or sentence is masked, using a mask ([MASK]) token. This method is called Masked LM (MLM). To introduce randomness into this masking technique, some words are replaced with a random word or remain unchanged. This helps improve the predictive accuracy of the model compared to the use of a more straightforward masking approach (Devlin et al., 2018). The meaning of words is distilled from looking at many of the neighboring words on both the left and right sides of the masked word, known as Bi-Directional Self-Attention. This approach is in contrast with other leading NLP Transformer methods, like the GPT (Radford et al., 2018) which only allows for a left-directed look at neighboring words, called Left-to-Right LM (Devlin et al., 2018). Secondly, understanding the relation between sentences in the context of a sequence is modeled. For this purpose, BERT makes use of Next Sentence Prediction (NSP). During the pretraining phase, the sequential next sentence that is passed on is either the real next sentence or interchanged with another sentence. This is done with a 50/50 ratio (Devlin et al., 2018).

After passing these tokens through the first self-attention layer, word vectors are created. Like the Word2Vec or GloVe model, these word vectors contain information about the word in context to other words. Whereas these older NLP models use only a selected view for understanding word context, BERT can create richer word vectors that contain information about a focal word in relation to all other words that are part of a text. Whereas Word2Vec interprets *'row'* the same in the context of a *'theatre row'* and *'row a boat'*, BERT can make a distinction between these meanings. Moreover, these vectors contain information about how words in a particular sentence relate to other words in different sentences, leading to the self-attention mechanism. These vectors are then passed on through to a feed-forward neural network layer. This feed-forward network aids in improving the results of the BERT model. In this thesis paper, each feed-forward layer contained 768 hidden units. After this step is completed, the data is passed through the first Encoder. The subsequent 11 Encoders, with their pre-established weights, make use of the established embeddings to further improve the accuracy of the model. When this task is finished the final word vectors are established.

To complete the classification task that is at hand in this thesis paper, HuggingFace's *BertForSequenceClassification* framework (HuggingFace, 2021) was added to the model. This framework adds an additional classification layer to the model that outputs logits for the classification task. After the raw logits were output by the BERT model, they had to be transformed into probabilities. This was done by applying a *SoftMax* activation function to the raw logits from the model. The *SoftMax* activation function normalized the prediction (Goodfellow et al., 2016) as was output by the BERT model. In essence, the *SoftMax* activation function generates probabilities, in this case, the probabilities of a text containing hate speech or not containing hate speech.

Because of the immense size of the BERT model and the limited computational power available during this research, the BERT model was run in batches of 25 observations. The outputs were then concatenated. Because the model was already pretrained and therefore the weights within the model were already established, running the model in batches did not affect the probabilities that were output for each observation. After running the BERT model, the probabilities of songs using hate speech were input into the data set for further analysis.

4.2 Pearson Correlation Coefficient

Before other, more advanced, statistical methods were employed, research was done with the use of the Pearson Correlation Coefficient (PCC). This statistical method offers insights into what relationships could be at play in the data set. The PCC measures the degree of linearity between two variables in a data set. The PCC is defined by the following formula:

$$r_{ab} = \frac{cov(a,b)}{\sigma_a \cdot \sigma_b} \tag{2}$$

Where r is the correlation coefficient, $cov(a, b)$ is the covariance between variable a and b and σ is the standard deviation of variable a or b (Profillidis & Botzoris, 2019). PCC values range from -1 to 1, where 1 indicates a perfect positive correlation. A perfect positive correlation means that when one variable moves in a positive direction, the other variable moves in exactly the same direction. If the PCC is 0, there is no correlation between the two variables and -1 indicates a perfect negative correlation. Weak, moderate, or strong levels of correlation are interpreted differently by different researchers. In this thesis paper, the PCC values were interpreted in line with Dancey & Reidy (2007).

While using the PCC, only a potential association between variables could be found, since the PCC did not allow for more complex (multivariate) relationships to be investigated. Therefore, care must be exerted when deducting relationships or effects from the PCC values. Strong levels of correlation between variables could be problematic when using these same variables in linear regression. In this thesis paper, the PCC was used to offer an exploratory answer to what relationships could be at play in the data set. All numeric variables in the data set were used as inputs, the PCC was calculated between all the variables and a correlogram was plotted. This correlogram can be found in Figure 6. Some of the weak, moderately strong, or strong correlations were then interpreted. Interpretations of these correlation coefficients can be found in the results.

4.3 Analysis of Variance and Tukey Honestly Significant Difference

To investigate if the level of hate speech probability differs between genres, a one-way Analysis of Variance (ANOVA) test was employed. This analysis helped answer the first sub-research question and test the second hypothesis. A one-way ANOVA tests if the means of different groups within the data differ significantly. The test does this by comparing if the means of at least two classes in a categorical variable differs significantly (Howell, 2012). The formula of the one-way ANOVA is:

$$F = \frac{MS_b}{MS_f} \tag{3}$$

Where MS_b is the mean sum of squared between groups and MS_f is the mean sum of squared within groups. The F-value must be compared to the critical F-value in order to reject or accept the test. In essence, the one-way ANOVA performed here tested if the mean value of $hate$, the song hate speech probability value, differed significantly for songs that belonged either in the Country, Rock, Metal, or Hip-Hop/Rap genre. Based on the significance of the test, using a significance boundary of $p < 0.05$, conclusions could be formed about the difference of *hate* for the genres under investigation.

After the one-way ANOVA test was performed, a Tukey Honestly Significant Difference (HSD) test (Tukey, 1949) was used to analyze the differences of the *hate* variable between each of the song genres. The Tukey HSD is very similar to a t-test. It makes use of the mean of both factor levels, as explained in the formula below:

$$q = \frac{M_a - M_b}{SE} \tag{4}$$

Where M is the mean value of mean level a or b (e.g., Metal or Country music), SE is the standard error and q is the value that should be compared to the critical q value. The differences between each of two genres were output and based on the significance value conclusions could be formed about the difference in hate speech probability between two different song genres.

4.4 Multiple Linear Regression

After the difference between hate speech probability was investigated for different song genres, the focus was directed towards answering the main research question and second sub-research question. It also allowed for testing of the first and third hypotheses. This was done with the use of multiple linear regression models. The first model was used to investigate how the probability of hate speech used in song lyrics related to the popularity of a song. A second model investigated the moderating influence that the song genre had on the relationship under investigation in the first model. These models used the *ln_popularity* variable as the dependent variable. Control variables were introduced to adjust for the effects of other elements in music and their influence on the popularity of that music. Two groups of control variables were used in this research: audio-based features and song-specific features. Audio-based variables included features like tempo, danceability, and valance (see Data section). Because of these inclusions, any effects of these audio-based effects on the dependent variable were controlled and not captured by the effect of independent variables of interest on the dependent variable. Song-specific variables consisted of the occurrence of the same artist or album within the data set. Furthermore, the inclusion of multiple songs from the same artists or album could have influenced the effect of our variable of interest on the dependent variable. To combat this indirect effect, the occurrence of songs from the same artist or same album were included as control variables in the linear regression.

A linear regression model is part of the family of generalized linear regression models, which also include the logistic regression model (Nelder & Wedderburn, 1972). The multiple linear regression is an extension to the simple linear regression, which assumes the use of one single independent variable. In a multiple linear regression, more than one independent variable is used. A multiple linear regression assumes the following relationship:

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n + \epsilon \tag{5}$$

Where Y is the dependent variable, $\beta_0$is the constant, $x_n$are the independent variables, β_n are the coefficients for each independent variable, n is the number of independent variables and ϵ is the residual or error term (Lacey, 1998). A linear regression model is based on the Ordinary Least Squared (OLS) process, which minimizes the squared errors (MSE) by fitting a linear line through the data. The lower the MSE, the better the data fits the model (Quaedvlieg, 2020). Furthermore, because a linear model is used to fit the data, using a linear regression model assumes that the relationship between the dependent and independent variables is linear.

This thesis paper dealt with multiple independent variables at a single time. Therefore, a decision needed to be made about what number of variables were optimal for use in the multiple linear regression model. For this selection process, a stepwise selection was used, which employed both a forward and backward selection approach and returned the best fitting model. Yang (2005) suggests that Akaike Information Criterion (AIC) is best used for the selection of the best fitting linear regression model. Therefore, during the stepwise selection approach, a final model was chosen, based on the lowest AIC score. The AIC uses the following formula (Burnham et al., 2011):

$$AIC = 2k - 2 \ln (\hat{L}) \tag{6}$$

Where k is the number of parameters in the model and $\hat{L}$ the maximum likelihood value of the model. Another measure that was employed during the stepwise selection approach is the Variance Inflation Factor (VIF). The VIF is a measure to identify (multi)collinearity in a model. In case of extreme collinearity, e.g., a VIF value higher than 5, variables can be excluded from the regression to combat the presence of collinearity. The VIF is calculated during a two-step approach. In the first step, each variable is regressed on all other variables of the linear model of interest. During the second step, the R^2 (R-squared) of this regression is employed in the following formula:

$$VIF = \frac{1}{1-R^2} \tag{7}$$

For this analysis, this thesis paper made use of the *vif* function from the *car* R-library (Fox, 2021). In case that one of the variables in the function had degrees of freedom higher than 1, the function reverted to the Generalized VIF (GVIF) and its extension, $GVIF^{\left(\frac{1}{2\cdot Df}\right)}$. These generalized forms could be interpreted similarly to the VIF.

To answer the main research question and first hypothesis, all variables in the data set, excluding the identifiers and song descriptives, were input into a *glm* function (Dobson, 1990), after which the *stepAIC* formula was used from the *MASS* R-library (Ripley, 2021). This function employed a stepwise algorithm to find the model with the lowest AIC score. The first model, as under investigation with the use of the *glm* function, can be found below (variables of interest in bold):

$$\begin{aligned}\boldsymbol{ln\_popularity} = {} & \beta_0 + \boldsymbol{\beta_1 \cdot hate} + \beta_2 \cdot genre + \beta_3 \cdot length + \beta_4 \cdot artist_{occ} + \\ & \beta_5 \cdot album_{occ} + \beta_6 \cdot danceability + \beta_7 \cdot energy + \beta_8 \cdot key + \beta_9 \cdot loudness + \\ & \beta_{10} \cdot mode + \beta_{11} \cdot speechiness + \beta_{12} \cdot acousticness + \beta_{13} \cdot instrumentalnes + \\ & \beta_{14} \cdot liveness + \beta_{15} \cdot valence + \beta_{16} \cdot tempo + \beta_{17} \cdot duration_{ms} + \epsilon\end{aligned} \tag{8}$$

After the lowest AIC model was found, this model was then checked for multicollinearity based on the VIF. If variables with a VIF value higher than 5 were found, these variables were excluded from the regression. The results were then output.

For the second multiple linear regression model that was employed, the lowest AIC model in the first regression was used as a base. Interaction effects between the hate speech probability variable, *hate*, were then added to this model. This model was then run using the *glm* function. The variables in the model were checked with the use of the VIF. Similarly, to the first model, variables were excluded if the boundary of 5 was violated. The second model, as under investigation with the use of the *glm* function, can be found below (variables of interest in bold):

$$\begin{aligned}\boldsymbol{ln\_popularity} = {} & \beta_0 + \boldsymbol{\beta_1 \cdot hate} + \boldsymbol{\beta_2 \cdot genre} + \beta_3 \cdot length + \beta_4 \cdot artist_{occ} + \\ & \beta_5 \cdot album_{occ} + \beta_6 \cdot danceability + \beta_7 \cdot energy + \beta_8 \cdot key + \beta_9 \cdot loudness + \\ & \beta_{10} \cdot mode + \beta_{11} \cdot speechiness + \beta_{12} \cdot acousticness + \beta_{13} \cdot instrumentalnes + \\ & \beta_{14} \cdot liveness + \beta_{15} \cdot valence + \beta_{16} \cdot tempo + \beta_{17} \cdot duration_{ms} + \\ & \boldsymbol{\beta_{18} \cdot hate * genre} + \epsilon\end{aligned} \tag{9}$$

The result of both multiple linear regression models can be found in Table 6 of the Results section.

5. Results

5.1 Pearson Correlation Coefficient

Based on the Pearson Correlation Coefficient, which is shown in Figure 6, a few weak and moderate correlations between variables were found. The hate speech probability variable has a moderately positive correlation with the *speechiness* and *length* variables. This indicates that songs that contained longer lyrics have a higher probability of containing hate speech. The hate speech variable also has a weak positive correlation with *danceability*. This indicates that songs that contain more hate speech are also more danceable on average.

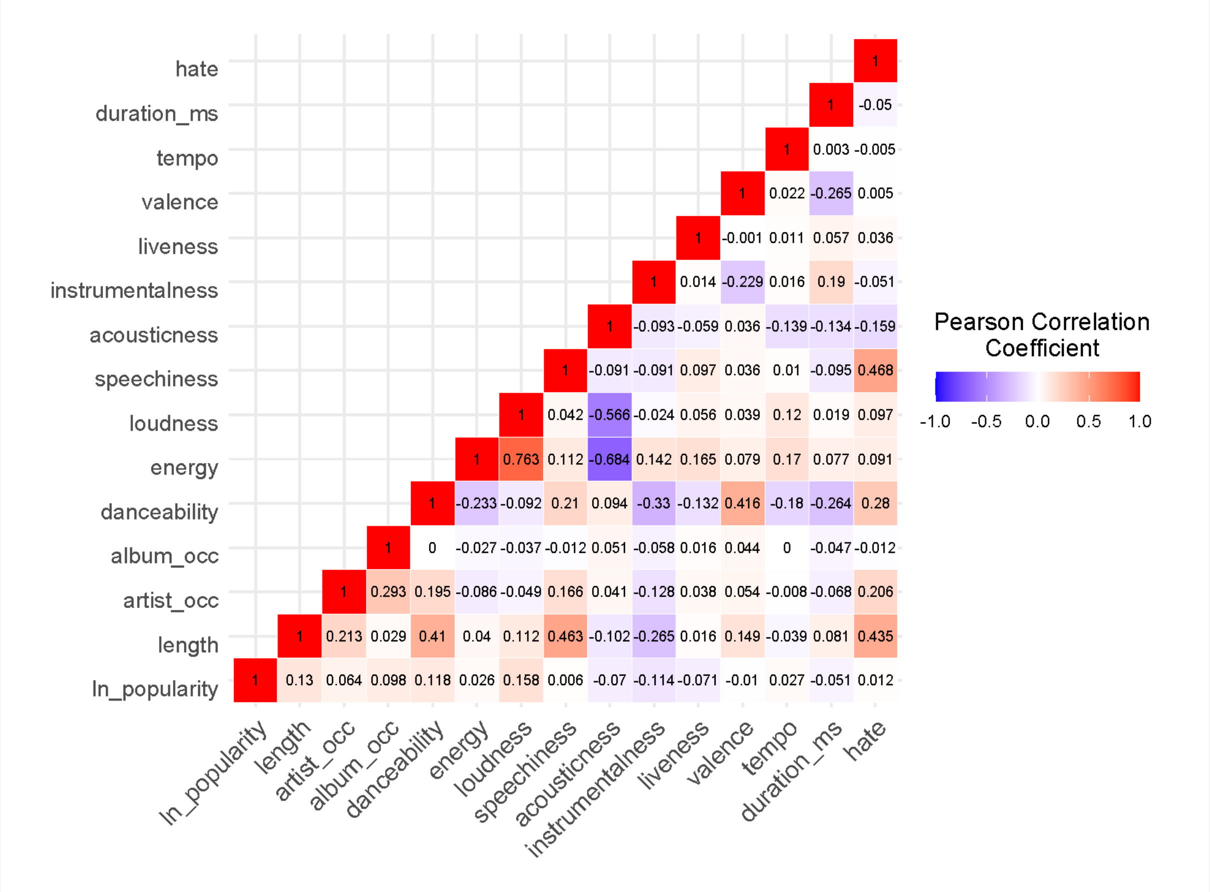


FIGURE 6: CORRELOGRAM BASED ON ALL NUMERIC VARIABLES IN THE DATA SET

Furthermore, the weakly positive correlation of the hate variable with the *artist_occ* variable indicates that some artists that appear more often in the data set, also more frequently feature songs that contain hate speech. Furthermore, some strong or moderate negative correlations were found between the audio-based variables. The correlation between *acousticness* and *loudness* and *energy* variables are strongly negative. This indicates that, on average, songs that contain less electronic music, are less loud (measured in Db) and contain less energetic tones. The correlation between the *loudness* and *energy* variable is strongly positive, indicating that loud music contains more energetic tones.

5.2 Difference in hate speech probability per genre

Based on the results of the one-way ANOVA, found in Table 4, a significance level of p < 0.01 was found. This means that a significant difference between the levels of hate speech probability was found for at least two genres in the data set (Country, Hip-Hop/Rap, Metal, and Rock). The difference between the levels of hate speech probability for each genre was visualized and can be found in Figure 7. This figure shows that the average levels of hate speech probability differ between genres. Furthermore, Table 3 shows the mean and standard error of the mean of the *hate* variable. Hip-Hop/Rap has the highest mean value of hate probability (0.491), compared to the other genres. Moreover, the Hip-Hop/Rap genre has the highest spread of the Q1 to Q3, indicating that the hate speech probability varies the most for this genre on average. Country, Metal, and Rock music have a mean value of the *hate* variable between 0.13 and 0.25, with a smaller spread between their Q1 and Q3. These three genres show to have outliers that contain a higher level of hate speech than the genres' respective mean or median values. This indicates that, although the genres have lower values of hate speech probability on average, compared to the Hip-Hop/Rap genre, they do have individual songs that score high for hate speech probability.

Genre	**Mean**	**St. Error of the Mean**
Country	0.161	0.002
Hip-Hop/Rap	0.491	0.003
Metal	0.240	0.002
Rock	0.139	0.001

TABLE 3: MEAN AND SME VALUES OF THE FOUR GENRES USED FOR THE HATE VARIABLE

	Degrees of freedom	**Sum of Squares**	**Mean Square**	**F-test**	**P-value**
genre	3	912.1	304.03	7012	*<2e-16****
Residuals	45208	1960.0	0.04		
*Note: *, **, *** significance at a 10%, 5% and 1%-level*					

TABLE 4: ONE-WAY ANOVA TEST RESULTS FOR THE GENRES UNDER INVESTIGATION

Factors	Difference	P-value
Hip-Hop/Rap - Country	0.330	*<2e-16****
Metal – Country	0.079	*<2e-16****
Rock – Country	-0.022	*<2e-16****
Metal – Hip-Hop/Rap	-0.251	*<2e-16****
Rock – Hip-Hop/Rap	-0.352	*<2e-16****
Rock – Metal	-0.101	*<2e-16****
*Note: *, **, *** significance at a 10%, 5% and 1%-level*		

TABLE 5: TUKEY HSD TEST RESULTS FOR GENRES UNDER INVESTIGATION

The difference that was apparent in Figure 7 had to be investigated analytically. The Tukey HSD (Honest Significant Difference) test results in Table 5 showed that all differences between groups of two genres are significant at a p < 0.01 level. The difference in hate speech probability between the two genres is the highest between Rock and Hip-Hop/Rap, being -0.352. This means that compared to Hip-Hop/Rap, the hate speech probability of Rock music is 0.352 lower. In general, Hip-Hop/Rap shows to differ most compared to the other musical genres, which is in line with Figure 6. The difference between Hip-Hop/Rap and Country music is 0.330. This difference meant that on average, Hip-Hop/Rap has a 0.330 higher level of hate speech probability as compared to the Country genre. Compared to Hip-Hop/Rap, Metal songs have a 0.251 lower hate speech probability.

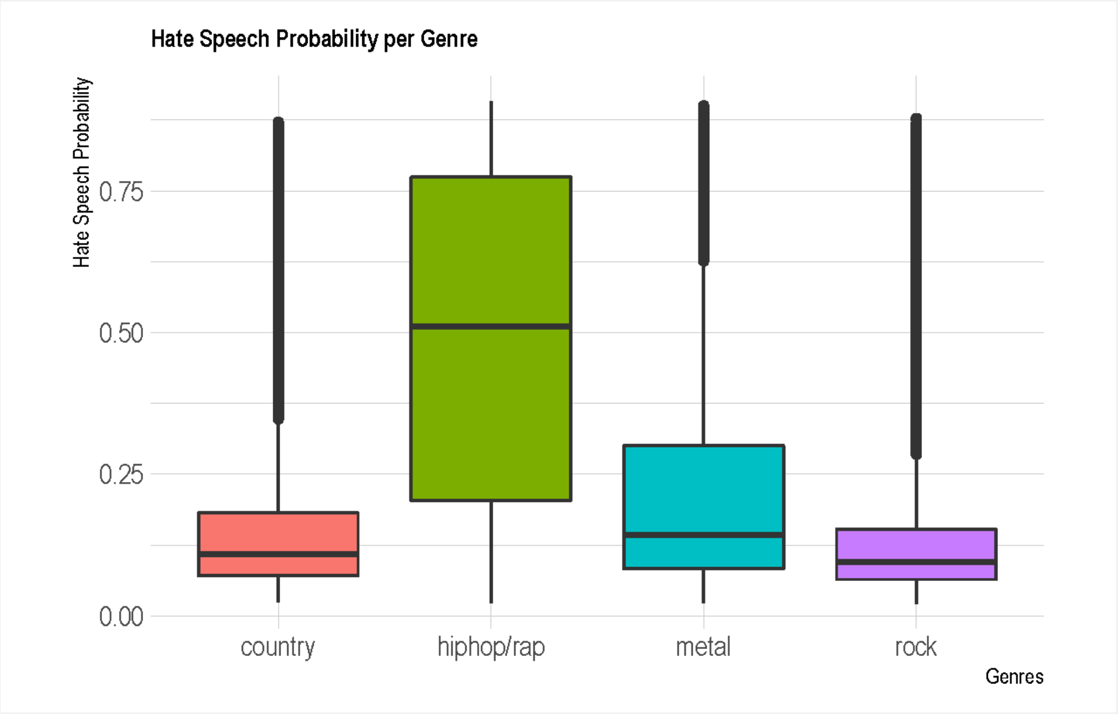


FIGURE 7: HATE SPEECH PROBABILITY DISTRIBUTION PER GENRE

5.3 Hate speech probability and popularity, moderated by song genre

The model selected with the *StepAIC* function was made up of all available variables, excluding the artist's occurrence variable. This model can be found as model 1 in Table 6. Based on Tabe 8, none of the variables in the first regression model have a (G)VIF higher than 5, indicating that no problematic (multi)collinearity is present in the model. All variables in the model are significant, most at a $p < 0.01$ level. Only the *key* variable is significant for some levels, as compared to the base 0-key level. An attempt was made to improve the significance of the key variable by grouping the *key* levels, e.g. key levels 0 and 1 became level 0-1, and so on. These merged *key* levels did not improve the significance of the key variable and did not yield a lower AIC score for that model. Therefore, the choice was made to continue with the original model, without grouped *key* levels.

This thesis paper hypothesizes that song genre (*genre)* has a moderating effect on the relationship between the use of hate speech and song popularity. Based on the significant existence of this moderating effect, the results in the first regression model should or should not be interpreted. For the second regression model, none of the variables in the model have a (G)VIF higher than 5, indicating that (multi)collinearity should not be present in the model (see Table 8). This second regression model has a lower AIC score as compared to the first linear regression model, indicating that the second model more accurately described the data set.

When looking at the influence of hate speech probability (*hate*) on the song popularity (*ln_popularity)*, the results in the first multiple linear regression model show that the relationship between the hate speech probability and the song popularity is significant ($p < 0.01$) and negative. Based on the second multiple linear regression model, displayed in Table 6, it can be concluded that keeping all other variables constant, the use of hate speech has a significantly ($p < 0.01$) positive effect on the song's popularity. These two *hate* coefficient values are in direct contradiction to each other. The switch of the hate speech coefficient's sign must be interpreted in light of the interaction effects that were included in the second regression model.

Based on the significant coefficients of the interaction effects in the second multiple linear regression model, it can be concluded that song genre has a significant moderating effect on the relationship between the hate speech probability and song popularity. Furthermore, based on the interaction plot in Figure 8, it can be concluded that for the Hip-Hop/Rap, Metal and Rock genre, an increasing level of hate speech use yields a lower song popularity. This effect is not as strong for each song genre. It can be concluded that this effect is strongest for Metal song, followed by Rock and then Hip-Hop/Rap songs. Country music is the only genre under investigation in this thesis paper that has an increasing level of song popularity for an increasing level of hate speech use. This result shows that Country songs that make heavy use of hate speech, are more popular than songs with a lower level of hate speech use. All these interaction effects are significant at a p < 0.01 level

Songs in the Hip-Hop/Rap, Metal, or Rock genres have a decreasing popularity with an increasing level of hate speech probability. The Country genre, however, has an increasing level of popularity with an increasing level of hate speech probability. Because this moderating relationship is only partly captured in the first regression model, it causes the hate speech probability coefficient to be negative. The inclusion of this significant moderating effect in the second model leads to a sign switch of the hate speech probability coefficient. This coefficient represents the effect of hate speech use on the song popularity coefficient, when not being part of the Hip-Hop/Rap, Metal, or Rock genre. Therefore, since the Country genre has an increased popularity for an increased level of hate speech use, the *hate* coefficient in the second regression model is positive.

Because of the changing sign of the *hate* coefficient, which is caused by the moderating influence of song genre, it can be concluded that the influence of the hate speech variable on the song popularity variable is ambiguous. The influence of hate speech on song popularity is influenced by the song genre. Therefore, only the influence of the hate speech variable on the song popularity can be determined for a specific genre and no general effect is found. Hence, this thesis paper refrains from interpreting the *hate* coefficient in a general context.

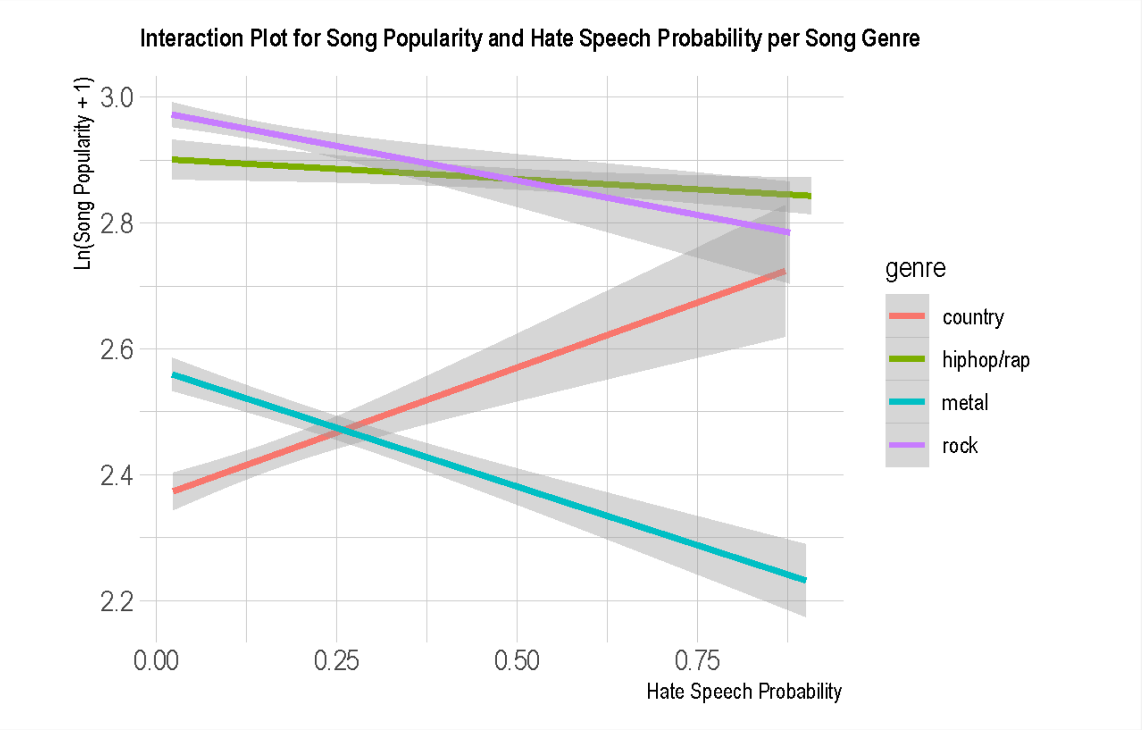


FIGURE 8: INTERACTION PLOT FOR SONG POPULARITY AND HATE SPEECH PROBABILITY PER SONG GENRE

	Dependent Variable *ln_popularity*				
	Model 1 *Model excl. interaction effects*	**Model 2** *Model incl. interaction effects*		**Model 1** *Model excl. interaction effects*	**Model 2** *Model incl. interaction effects*
Constant	3.066***	3.009***	*key8*	0.089***	0.089***
hate	**-0.210*** **	**0.219*** **	*key9*	-0.027	-0.027
genrehiphop/rap	0.289***	**0.368*** **	*key10*	0.024	0.025
genremetal	0.037**	**0.135*** **	*key11*	0.003	0.003
genrerock	0.522***	**0.598*** **	*loudness*	0.071***	0.071***
length	0.0004***	0.0004***	*mode1*	-0.021**	-0.018*
artist_occ	0.001***	0.001***	*speechiness*	-0.413***	-0.424***
album_occ	0.032***	0.032***	*acousticness*	-0.155***	-0.148***
danceability	0.539***	0.523***	*instrumentalness*	-0.162***	-0.160***
energy	-0.482***	-0.481***	*liveness*	-0.240***	-0.244***
key1	0.026	0.026	*valence*	-0.426***	-0.428***
key2	-0.011	-0.013	*tempo*	0.001***	0.001***
key3	0.080***	0.081***	*duration_ms*	-0.000***	-0.000***
key4	-0.014	-0.014	***hate:genrehiphop/rap***		**-0.428*** **
key5	0.021	0.022	***hate:genremetal***		**-0.553*** **
key6	0.073***	0.073***	***hate:genrerock***		**-0.485*** **
key7	-0.012	-0.012			

	Model 1	**Model 2**
Observations	45,212	45,212
Log Likelihood	-59,695.590	-59,666.590
Akaike Inf. Crit.	119,451.200	119,399.200
Note: **, **, *** significance at a 10%, 5% and 1%-level*		

TABLE 6: OVERVIEW OF OUTPUT MULTIPLE REGRESSION MODE

Next to the interpretation of the *hate* coefficients, the interaction coefficients and the interaction plot, the direct effect of the song genre on the song popularity can be interpreted. The Hip-Hop/Rap, Metal, and Rock genres have a significantly positive influence on song popularity as compared to the Country genre. If a song is part of the Rock genre, it has an 81.85% ($e^{0.598} - 1$) higher popularity score as compared to being part of the Country genre. This effect is smaller for the Hip-Hop/Rap and Metal genre. When a song belongs to the Hip-Hop/Rap genre, the popularity score is 44,48% higher as compared to being part of the Country genre. For the Metal genre, this score is 14,45% higher as compared to being part of the Country genre. Lastly, because none of the genre coefficients are negative, it can be concluded that if a song is part of the Country genre, it has a lower popularity score as compared to the other song genres, keeping all other variables constant.

5.4 Implications for Hypotheses

Firstly, results from the Tukey HSD test showed that the level of hate speech probability differs significantly between genres, leading to the acceptance of the second hypothesis. Secondly, based on the ambiguity of the relationship between the hate speech probability variable and the song popularity variable, no clear conclusions can be formed about this relationship. This leads to the rejection of the first hypothesis. Lastly, because the moderating influence of the song genres is found to be significant and distinct effects are identified for each song genre, the third hypothesis is accepted.

Nr	**Hypothesis**	**Result**
1	The use of hate speech in a song positively relates to the popularity of that song.	Rejected
2	The amount of hate speech used in music differs per musical genre.	Accepted
3	The genre of a song moderates the effect between the level of hate speech and popularity.	Accepted

TABLE 7: OVERVIEW OF REJECTED OR ACCEPTED HYPOTHESES

6. Discussion

This thesis paper finds evidence that different levels of offensive material, measured in the form of hate speech, were found between genres. Interpreting this result from the perspective of an artist, certain base levels of offensive language are at play for song genres. This is finding is in line with literature concerning the use of offensive language across genres. Hart & Day (2020) stated that the level of explicit material is not uniformly distributed across genres. This thesis paper finds that Hip-Hop/Rap music had a significantly higher level of hate speech probability than other genres. Kenvarg (2021) states that Hip-Hop/Rap, compared to other genres, contains more hate speech. Iwamoto (2003) points out that Hip-Hop/Rap is a masculine genre and uses this as an explanation for its more male-dominant use of offensive language, which is also concluded in this thesis paper. Lastly, in contrast to Kenvrag's finding, this thesis paper did not find evidence confirming that the use of offensive language is higher for the Rock and Metal genre, as compared to other song genres. Although a significantly higher level of hate speech probability was found for the Metal genre, Rock had a significantly lower level of hate speech probability as compared to the Country genre.

Next to the difference of song genres, a significant interaction effect of the song genres was found for the relationship between the level of hate speech probability and song popularity. This indicated that the relationship between the use of hate speech and the song's popularity was significantly influenced by the genre of a song. Although no direct research focuses on the moderating influence a song genre has on the relationship between the use of hate speech and song popularity, the findings seem to be in line with other research. Taking Jones' (1997) findings into account, it could be argued that norms are in place for artists from different song genres. Combined with the results from hypothesis 2, it can be concluded that base levels of hate speech use are in place for song genres. Artists at the beginning of their career, look at more established ones to fit a particular stereotype, as described by Lapinski & Rimal (2005). Therefore, these artists feel the need of adapting to the norms and base levels of offensive language that are established within social groups, in this case, groups formed by fans or listeners of different song genres. These norms then inform artists about the base level of offensive language they should use within a particular genre, in this case, measured as hate speech.

When looking at the direct relationship between hate speech use and song popularity, it can be concluded that this relationship is ambiguous. Therefore, no inferences or conclusions can be made about the sign and size of the influence of the use of hate speech on the popularity of a song. This conclusion seems to mirror findings by Lang & Switzer (2008). They find a complex relationship when it comes to the use of offensive content, in this case, sexual or violent content, on (non-American) movie revenues. For non-American audiences, the use of sexual and especially violent content seems to increase movie revenues. For a general audience, which includes a large American public, the increased use of sexual or violent content seems to negatively influence movie revenues. This finding suggests that the relation between the use of offensive content and a measure of success, be it popularity or revenue, can be regionally different and ambiguous in a general form. The findings of this ambiguous relationship are further exemplified by the fact that, for the use of sexual content in the advertising industry, conflicting results were found (Gramazio et al., 2021; Zawisza et al., 2018; Lundstrom & Sciglimpaglia, 1977; Lull & Bushman, 2015; Parker & Furnham, 2007).

7. Conclusions

7.1 Main Conclusions

This thesis paper finds that the level of hate speech use differs between genres and base levels of offensive language, measured in the form of hate speech use, exist. Especially Hip-Hop/Rap makes extensive use of hate speech in its lyrics, which is in line with previous research (Iwamoto, 2003; Kenvarg, 2021) that paints Hip-Hop as a masculine and female-unfriendly genre. When looking closely at the other genres, Metal contains the second most amount of hate speech, Country the third-highest level, and Rock contains the lowest amount of hate speech of all genres in this thesis paper. Furthermore, this thesis paper finds that song genre has a significant moderating influence on the relationship between the use of hate speech and song popularity. Using an increased amount of hate speech in the Country genre yields more popular songs. For the other three song genres, an increase in hate speech use leads to less popular songs. This thesis paper concludes that significant differences in norms exist between song genres, consistent with Lapinski & Rimal (2005). From a qualitative perspective, this difference is attributed to stereotypes and mimicking behavior of artists, consistent with Jones (1997).

When looking at the direct and unmoderated influence of hate speech use on the song's popularity, no evidence is found for the confirmation of this direct relationship. Literature finds both positive and negative effects for the use of offensive language in other creative sectors and its relationship to popularity. It suggests regional differences are present (Lang & Switzer, 2008) and is ambiguous in general (Gramazio et al., 2021; Zawisza et al., 2018; Lundstrom & Sciglimpaglia, 1977; Lull & Bushman, 2015; Parker & Furnham, 2007) indicating no clear conclusions can be formed about the direct and unmoderated relationship under investigation.

7.2 Managerial Implications

The music industry is a big creative sector. Revenues in the worldwide music industry were $ 23.1 billion in 2020 (Statista, 2021). Although many musical artists are actively recording and releasing new music, the music industry is a superstar industry. Within a superstar industry, most of the revenues earned are by the top 5% of artists. Therefore, the listening behavior of superstars' fans can have consequences for the music industry's ability to generate revenues (Kamara II, 2018; Galuszka & Wyrzykowska, 2016). Furthermore, because of the increasing popularity of 360-degree deals, music publishers are becoming increasingly dependent on an artist's ability to generate revenue in all types of ways. This stands in stark contrast to older types of contracts that were solely focused on generating revenue from music sales and radio plays (Galuszka & Wyrzykowska, 2016).

The research performed in this thesis paper answers how the use of offensive language, specifically hate speech, influences song popularity. No clear relationship is found between these two variables. Therefore, no direct implications can be derived from this relationship to form recommendations for managers in the music industry. Based on results regarding the first hypothesis and previous research, it can be concluded that the effect of offensive language or content on song popularity is ambiguous. Therefore, management within the music industry must exert caution when advising artists on what levels of offensive language are appropriate.

Clear evidence is found that base levels of offensive language, specifically hate speech, exist within the musical genres under investigation. The level of hate speech differs significantly between musical genres. Furthermore, the song genre has a significant influence on the relationship between the use of hate speech and song popularity. This implies that differences exist between song genres in terms of how hate speech use affects the song's popularity. Knowing about the existence of such differences could be useful for managers in the music industry. A new artist or an artist looking for a new image could be advised to follow a genre's base level to align more with audience expectations. To known what audience expectations are, managers in the music industry are advised to engage in the collection of customer preferences on a local level. This collection could be done in the form of a customer panel, survey, or other types of customer feedback. With this information, managers in the music industry can advise local artists on customer preferences with a certain genre. Lastly, next to investigating consumer preferences, managers could make use of genre median levels to establish what these base levels are. Although the levels of hate speech for the Metal, Country, and Rock genres differ significantly, they are relatively similar and low (median levels under 0.25). The use of hate speech in the Hip-Hop/Rap genre is relatively higher than in these other song genres. Therefore, especially an artist looking to enter the Hip-Hop/Rap genre should focus on the extensive use of offensive language in their songs.

Distinct effects of different song genres on the relationship between hate speech use and song popularity are found. These effects are modeled in the second linear regression model in Table 6. Next to finding this moderating influence, multiple audio-based control variables are used in this model. With the use of this linear regression model, managers in the music industry can predict an artist's song popularity, based on these factors. This predictive opportunity that this linear regression model offers could serve two use cases. First, artists that have already recorded a song, can input information about the song, i.e. hate speech use, tempo, or danceability, into the model and create an estimation of how well their song will perform. This could aid the decision to publish a certain song as a single, offering the song actively to radio stations or creators of popular playlists, or make the song a less important part of an album. Secondly, artists that are looking to create popular music, could use the predictive ability of this second linear regression model to foresee what elements their next song should contain to become popular. Although, creating music is an artistic and creative process, being informed as on artists on what elements make a popular song could be influential in this process. Lastly, what must be taken into account when using this model is that the only language modeled in this thesis paper is hate speech. Many different forms of offensive language exist, therefore including different types of language would aid the benefit of this model to artists. This does require a scientific effort since the influence of these other types of language must be established scientifically first before additional conclusions can be formed.

7.3 Limitations

7.3.1 Data Limitations

One might question the high level of loss that occurred throughout the data collection phase. Although in total 5,0% of theoretically possible observations make up the final data set, multiple choices and collection limitations caused the low number of observations. Firstly, the Spotify Search API is efficiently designed to find materials that are input into the search function. It, however, has not been designed to find randomized songs. Therefore, a randomized search using the Search API yielded fewer random song as expected at first. Secondly, the fact that both Spotify's and Genius' API do not make use of industry-standard identifiers made it necessary to perform an error check, to remove incorrectly gathered Genius API data. Thirdly, although Genius does maintain a large library of song lyrics, more than 40% of all songs input into the API do not contain song lyrics. Lastly, the fact that this thesis paper focuses on songs that make use of the English language further limited the number of songs that could be included in the data set. The limited scope of observations that were gathered during the data collection process did not hinder the research. In total, a combined 14.848.953 words were available for use in an NLP method.

An important note to the data collection approach is that Spotify does not discriminate between songs that are released for the first time on the platform and re-released, usually in the form of a remix, remaster, or best-of album. Therefore, not all release dates are accurate. This however is not expected to be problematic for the research. The choice for a song between 2010 and 2019 is double-sided. On one hand, the data set needs to be restricted to a certain size, as the API allows the collection of unlimited observations. Furthermore, as mentioned earlier in the data collection, songs from 2020 and 2021 should be excluded to combat the bias in Spotify's popularity measure.

Moreover, most of the data is collected from just two sources, most of which is Spotify. Since the values of the audio features that are collected are based on technologies that are owned and operated by Spotify, a high dependence is put on Spotify's platform for the outcomes in this thesis paper. This should be noted as a weakness, especially since Spotify does not exactly specify how the audio features are generated and with what accuracy these automated features are generated. Furthermore, even though Genius is one of the biggest user-supported lyric platforms in the world, it does not guarantee the accuracy of its lyrics. This could mean that some of the words gathered in the lyrics could be non-existent or inaccurate. The inaccuracy of lyrics, however, is a problem that all lyric providers deal with since the share size of lyrics available makes it impossible to manually review every lyric.

7.3.2 Methodological Limitations

To classify the presence of hate speech in the song lyrics a state-of-the-art NLP model, BERT, was used. In this case, this specifically concerned a pre-trained BERT model. Generally, pre-trained BERT models are known for their good performance for certain text-based tasks. These tasks, however, usually lie in the field of classification of a piece of text, like a review or transcript. Although extremes exist, these types of texts are comparable to the way language is used in all-day tasks. Musical lyrics, however, are not comparable to these types of text. Lyrics rely heavily on the artistic vision of an artist and contain abbreviations, slang, or other types of differing language. Furthermore, for some musical genres, lyrics usually are written in a rime form, further distinguishing these types of language from normal-day language. Based on these differences, one could argue that to potentially create a model that is better able to classify lyrics, the pretraining of such a model must also be done with the use of song lyrics. Therefore, the fact that this thesis makes use of Aluru et al.'s (2020) model, which is mostly trained on hate speech from Twitter, can be seen as a limitation to the research in this thesis paper.

Furthermore, although Aluru et al. (2020) extensively describe their approach taken during the pretraining of the BERT model, the ability to tweak or improve the BERT model is lost, since only a final pretrained model is available for use. This limits the ability of the researcher to improve or adapt the model based on the NLP task that is at hand. Moreover, most pre-trained BERT models that are available for use contain a word limit of 256 or 512 words. This is partially done because of the immense size of the available BERT models but does limit the ability to classify all of the available lyrics within the data set. Although 512 words seem to capture a sizeable proportion of a song lyric, songs that are longer than 512 words and heavy on hate speech use towards the end of their lyrics, are particularly affected by this limitation.

Lastly, this thesis paper makes use of hate speech as a measure of offensive language, an umbrella form of language that describes the rude, upsetting, or unpleasant language. Hate speech, however, is not a perfect proxy of offensive language. Many other forms of offensive language exist, which were not investigated during this thesis paper. To improve the relevance between existing literature and the research in this paper, other types of language classifications could have been performed (see Future Research).

7.4 Future Research

To improve the research performed in this thesis paper, a few points are noted. The BERT model used in this thesis paper could be improved because it is mostly pretrained on Twitter data. In order to more accurately classify music lyrics, future research could employ a BERT or other state-of-the-art NLP model that is pretrained on music lyrics specifically. With this different pretraining step, potential improved results could be achieved. What must be noted is that to the knowledge of this author, no or little language-labeled music lyrics are available for research. Therefore, such a research design would require significant manual labor and could be costly.

Two other steps that could improve the current research design are the inclusion of more song genres and using a larger song sample. With the inclusion of more song genres, broader research could be done into the use of offensive language in music, since the research focused on 4 song genres only. By investigating more song genres, advice to managers within the music industry could improve, since most music labels or publishers sign artists from multiple genres. By using a larger song sample, for example by broadening the years during which the songs in the data set were released, more general effects could be investigated. Since the songs in the data set now are specifically from 2010 till 2019, relations that were existent for this period could only be investigated. What must be noted is that if such a research design was chosen, the inclusion of a time-variant variable would be advisable, since potential effects over a larger amount of time could be present.

Although relations regarding the use of hate speech were investigated, this thesis' literature study also included sources that looked at the influence of sexual, violent, or profane language. In this thesis paper, these types of language were grouped under the umbrella of offensive language. Many different types of offensive language exist. Therefore, a wider study that looks at different types of offensive language could add valuable knowledge to the scientific field. This would be especially valuable because limited research on the use of offensive language in songs exists. What must be considered when employing such a research design, is the fact that only relatively few NLP methods have pre-trained models that focus on offensive language, with most of the research and articles focused specifically on the sole classification of explicit music. Therefore, such a research design could become more labor-intensive.

8. References

Alekseevna, M. O., & Gennadyevna, K. M. (2016). Slang in rock-music. *In The Eighth International Congress on Social Sciences and Humanities*, 130.

Aluru, S. S., Mathew, B., Saha, P., & Mukherjee, A. (2020). Deep learning models for multilingual hate speech detection. *arXiv*.

Back, L., Keith, M., & Solomos, J. (1998). *Racism on the Internet: Mapping neo-fascist subcultures in cyberspace. Nation and race: The developing Euro-American racist subculture.* Lebanon, NH: Northeastern University Press.

Banks, J. (2010). Regulating hate speech online. *International Review of Law, Computers & Technology*, 233-239.

Bellego, C., Benatia, D., & Pape, L. D. (2021). Dealing with logs and zeros in regression models. *CREST-Série des Documents de Travail*.

Bergelid, L. (2018). Classification of explicit music content using lyrics and music metadata. *Doctoral dissertation, KTH*.

Bilewicz, M., & Soral, W. (2020). Hate speech epidemic. The dynamic effects of derogatory language on intergroup relations and political radicalization. *Political Psychology*, 3-33.

Bright, W. (1963). Language and music: Areas for cooperation. *Ethnomusicology*, 26-32.

Brown, T. S. (2004). Subcultures, pop music and politics: Skinheads and" Nazi rock" in England and Germany. *Journal of Social History*, 157-178.

Budzinski, O., & Pannicke, J. (2017). Does popularity matter in a TV song competition? Evidence from a national music contest. Evidence from a National Music Contest. *Ilmenau Economics Discussion Papers*, 21.

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral ecology and sociobiology*, 23-25.

Butler, A. (2019). Why Streaming is a Good Thing for the Music Industry. *Backstage Pass*, 22.

Byun, C. (2016). *The economics of the popular music industry: Modelling from microeconomic theory and industrial organization.* . Springer.

Cambridge University Press. (2021, November 4). Retrieved from Cambridge Dictionary: https://dictionary.cambridge.org/dictionary/english

Capital FM. (2020, August 25). *The WAP Dance Challenge Is Taking Over TikTok And James Charles & Addison Rae's Epic Moves Are Seriously Impressive*. Retrieved from Capital FM: https://www.capitalfm.com/news/wap-tiktok-dance-addison-rae/

Cavnar, W. B., & Trenkle, J. M. (1994). N-gram-based text categorization. *In Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval* .

CNBC. (2020, May 19). *Joe Rogan's podcast moves exclusively to Spotify, stock soars*. Retrieved from CNBC: https://www.cnbc.com/2020/05/19/spotify-lands-exclusive-rights-to-joe-rogan-podcast.html

Coleman, J. (2014). *Global English Slang: Methodologies and Perspectives*. London: Routledge.

Connolly, M., & Krueger, A. B. (2006). Rockonomics: The economics of popular music. 1, 667-719. In M. Connolly, & A. B. Krueger, *Handbook of the Economics of Art and Culture* (pp. 667-719). Amsterdam: Elsevier.

Cotter, J. M. (1999). Sounds of hate: White power rock and roll and the neo-nazi skinhead subculture. *Terrorism and Political Violence*, 111-140.

Dancey, C., & Reidy, J. (2007). *Statistics without Maths for Psychology.* Pearson Education.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv*.

Dobson, A. J. (1990). *An Introduction to Generalized Linear Models.* London: Chapman and Hall. Retrieved from RDocumentation.

Fell, M., Cabrio, E., Corazza, M., & Gandon, F. (2019). Comparing Automated Methods to Detect Explicit Content in Song Lyrics. *In Proceedings of the International Conference on Recent Advances in Natural Language Processing*, 338-344.

Firth, J. (1957). *"A synopsis of linguistic theory 1930–1955". Studies in Linguistic Analysis: 1–32. Reprinted in F.R. Palmer, ed. (1968).* . London: Longman.

Fox, J. (2021, December 17). *car: Companion to Applied Regression*. Retrieved from CRAN R-Project: https://cran.r-project.org/web/packages/car/index.html

Furer, K. S. (1991). Warning: Explicit Language Contained Obscenity and Music. *NYL Sch. J. Hum. Rts.*, 461.

Galuszka, P., & Wyrzykowska, K. M. (2016). Running a record label when records don't sell anymore: empirical evidence from Poland. *Popular Music*, 23-40.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). 6.2.2.3 Softmax units for multinoulli output distributions. . In *Deep learning* (p. 180). Boston, MA: MIT Press.

Gramazio, S., Cadinu, M., Guizzo, F., & Carnaghi, A. (2021). Does sex really sell? Paradoxical effects of sexualization in advertising on product attractiveness and purchase intentions. *Sex Roles*, 701-719.

Gugger, S. (2021). *Summary of Transformer Models*. Retrieved from Hugging Face: https://huggingface.co/course/chapter1/9?fw=pt

Hart, C., & Day, G. A. (2020). Linguistic Analysis of Sexual Content and Emotive Language in Contemporary Music Genres. *Sexuality & Culture* , 516-531.

Henderson, E. (2021, December 22). *GeniusR*. Retrieved from GitHub: https://ewenme.github.io/geniusr/index.html

Holody, K. J., Anderson, C., Craig, C., & Flynn, M. (2016). "Drunk in Love": The Portrayal of Risk Behavior in Music Lyrics. *Journal of Health Communication*, 1098-1106.

Hornik, K. (2021, December 18). *textcat: N-Gram Based Text Categorization*. Retrieved from CRAN R-Project: https://cran.rstudio.com/web/packages/textcat/

Hornik, K., Mair, P., Rauch, J., Geiger, W., Buchta, C., & Feinerer, I. (2013). The textcat Package for n-Gram Based Text Categorization in R. *Journal of Statistical Software*, 1-17.

Howell, D. C. (2012). *Statistical methods for psychology.* Boston, MA: Cengage Learning.

Hugginface. (2021, February). Retrieved from Github: https://github.com/huggingface/tokenizers

HuggingFace. (2021, December 17). *BERT*. Retrieved from HuggingFace: https://huggingface.co/docs/transformers/model_doc/bert

HuggingFace. (2021, December 17). *HuggingFace*. Retrieved from Tokenizer: https://huggingface.co/docs/transformers/main_classes/tokenizer#transformers.PreTrainedTokenizerBase.__call__.truncation

IFPI. (2021). *Global Music Report*. Retrieved from https://www.ifpi.org/wp-content/uploads/2020/03/GMR2021_STATE_OF_THE_INDUSTRY.pdf

International Federation of the Phonographic Industry. (2021, Juli 2). *Industry*. Retrieved from ifpi.org: https://www.ifpi.org/our-industry/industry-data/

Iwamoto, D. (2003). Tupac Shakur: Understanding the identity formation of hyper-masculinity of a popular hip-hop artist. *The Black Scholar*, 44-49.

Jones, K. (1997). Are rap videos more violent? Style differences and the prevalence of sex and violence in the age of MTV. *Howard Journal of Communications*, 343-356.

Kamara II, K. (2018). Music Artists' Strategies to Generate Revenue Through Technology. *Doctoral dissertation, Walden University*.

Katz-Gerro, T. (2002). Highbrow cultural consumption and class distinction in Italy, Israel, West Germany, Sweden, and the United States. *Social forces*, 207-229.

Kenvarg, E. (2021). “Thisniggaugly”: Understanding Violent Hate Speech in Rap Music. *In Peace in the Twenty-First Century* .

Kids-In-Mind. (2021, September 15). *Our Mission*. Retrieved from Kids-In-Mind: https://kids-in-mind.com/about.htm#methodologyofnumbers

Kim, J., & Yi, M. Y. (2018). A hybrid modeling approach for an automated lyrics-rating system for adolescents. *Advances in information retrieval – 41st European conference on IR research*, 779-786.

King, P. (1988). Heavy metal music and drug abuse in adolescents. *Postgraduate medicine*, 295-304.

Lacey, M. (1998). *Multiple Linear Regression*. Retrieved from Yale STAT 101: http://www.stat.yale.edu/Courses/1997-98/101/linmult.htm

Lang, D. M., & Switzer, D. M. (2009). Does Sex Sell? A look at the effects of sex and violence on motion picture revenues. *Working paper*.

Lapinski, M., & Rimal, R. (2005). An Explication of Social Norms. *Communication Theory*, 127-147.

Levy, L. W., Levy, L. W., Karst, K. L., & Winkler, A. (2000). *Encyclopedia of the american constitution*. USA: Macmillan.

Lull, R. B., & Bushman, B. J. (2015). Do sex and violence sell? A meta-analytic review of the effects of sexual and violent media and ad content on memory, attitudes, and buying intentions. *Psychological bulletin*, 1022.

Lundstrom, W. J., & Sciglimpaglia, D. (1977). Sex role portrayals in advertising. *Journal of marketing*, 72-79.

Maiden, M., Smith, J., & Ledgeway, A. (2010). *The Cambridge History of the Romance Languages*. Cambridge: Cambridge University Press.

Marshall Cavendish Corporation. (2010). *Sex and society*. New York: Marshall Cavendish.

McRae, R. (2001). "What is hip?" and other inquiries in jazz slang lexicography. *Notes*, 574-584.

Merriam-Webster Inc. (2021, 11 15). *Definition of obscene*. Retrieved from Merriam-Webster Dictionary.

Messner, B. A., Jipson, A., Becker, P. J., & Byers, B. (2007). The hardest hate: A sociological analysis of country hate music., 30(4), 513-531. *Popular Music and Society*, 513.

Metro. (2020, August 10). *Viola Davis is living her WAP fantasies as fans edit her into Cardi B and Megan Thee Stallion's music video*. Retrieved from Metro: https://metro.co.uk/2020/08/10/viola-davis-living-wap-fantasies-fans-cut-cardi-b-megan-thee-stallions-music-video-13108113/

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv:1301.3781*.

Mikolov, T., Yih, W. T., & Zweig, G. (2013). Linguistic regularities in continuous space word representations. *In Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies*, 746-751.

Negus, K. (2018). From creator to data: the post-record music industry and the digital conglomeratres. *Media, Culture & Society*, 367-384.

Neguţ, A., & Sârbescu, P. (2014). Problem music or problem stereotypes? The dynamics of stereotype activation in rock and hip-hop music. *Musicae Scientiae*, 3-16.

Nelder, J., & Wedderburn, R. (1972). "Generalized Linear Models". Journal of the Royal Statistical Society. *Blackwell Publishing*, 370-384.

New York Times. (2020, October 27). Cardi B's 'WAP' Proves Music's Dirty Secret: Censorship Is Good Business. *New York Times*.

NPR. (2010, October 29). *You Ask, We Answer: 'Parental Advisory' Labels — The Criteria And The History*. Retrieved from NPR: https://www.npr.org/sections/therecord/2010/10/29/130905176/you-ask-we-answer-parental-advisory---why-when-how?t=1628693439444

Oxford University Press. (2021, June 11). Oxford English Dictionary. Oxford, England.

Parker, E., & Furnham, A. (2007). Does sex sell? The effect of sexual programme content on the recall of sexual and non-sexual advertisements. . *Applied Cognitive Psychology*, 1217-1228.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. *In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532-1543.

Powers, H. S. (1980). Language models and musical analysis. *Ethnomusicology*, 1-60.

Primack, B. A., Douglas, E. L., Fine, M. J., & Dalton, M. A. (2009). Exposure to sexual lyrics and sexual experience among urban adolescents. *American journal of preventive medicine*, 317-323.

Profillidis, V. A., & Botzoris, G. N. (2019). *Chapter 5—Statistical Methods for Transport Demand Modeling. Modeling of Transport Demand;*. Profillidis, VA: Botzoris.

PyTorch. (2021, December 17). *PyTorch*. Retrieved from GitHub: https://github.com/pytorch/pytorch

Quaedvlieg, R. (2020). *Erasmus University Rotterdam*. Retrieved from Financial Methods & Techniques: https://canvas.eur.nl/courses/32151

Radford, A., Narasimhan, K., & Salimans, T. &. (2018). Improving language understanding by generative pre-training.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*.

Reichert, T., Childers, C. C., & Reid, L. N. (2012). How sex in advertising varies by product category: An analysis of three decades of visual sexual imagery in magazine advertising. *Journal of Current Issues & Research in Advertising*, 1-19.

Rentfrow, P. J., & Gosling, S. D. (2007). The content and validity of music-genre stereotypes among college students. *Psychology of music*, 306-326.

Ripley, B. (2021, December 18). *MASS: Support Functions and Datasets for Venables and Ripley's MASS*. Retrieved from CRAN R-Project: https://cran.r-project.org/web/packages/MASS

Rogers, J. (2017). Deconstructing the music industry ecosystem. *In Media convergence and deconvergence*, 217-239.

Rolling Stone. (2015, September 17). *PMRC's 'Filthy 15': Where Are They Now?* . Retrieved from Rolling Stone: https://www.rollingstone.com/music/music-lists/pmrcs-filthy-15-where-are-they-now-60601/

Rospocher, M. (2021). Explicit song lyrics detection with subword-enriched word embeddings. *Expert Systems with Applications*, 163.

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv*.

Sariyar, M. (2021, December 17). *RecordLinkage: Record Linkage Functions for Linking and Deduplicating Data Sets*. Retrieved from CRAN R-Project: https://cran.r-project.org/web/packages/RecordLinkage/index.html

Smiler, A., Shewmaker, J., & Hearon, B. (2017). From "I Want To Hold Your Hand" to "Promiscuous": Sexual Stereotypes in Popular Music Lyrics, 1960–2008. *Sexuality & Culture*, 1083-1105.

Spotify. (2021, September 27). *Spotify for Developers*. Retrieved from Web API Reference : https://developer.spotify.com/documentation/web-api/reference/#endpoint-get-audio-features

Statista. (2021, July 14). *Global recorded music revenue from 1999 to 2020*. Retrieved from Statista: https://www.statista.com/statistics/272305/global-revenue-of-the-music-industry/

Statista. (2021, November 18). *Number of Spotify premium subscribers worldwide from 1st quarter 2015 to 3rd quarter 2021* . Retrieved from Statista: https://www.statista.com/statistics/244995/number-of-paying-spotify-subscribers/

Thompson, K. M., & Yokota, F. (2004). Violence, sex, and profanity in films: correlation of movie ratings with content. *Medscape General Medicine*, 6.

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 99-114.

Vaglio, A., Hennequin, R., Moussallam, M., Richard, G., & d'Alché-Buc, F. (2020). Audio-based detection of explicit content in music. *In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing*, 526-530.

VICE. (2020, December 28). *Mums and Dads Were the Real Stars of TikTok in 2020*. Retrieved from VICE: vice.com/en/article/5dp5z3/mums-and-dads-were-the-real-stars-of-tiktok-in-2020

Weller, J. F. (2013). How popular music artists form an artistic and professional identity and portfolio career in emerging adulthood. *Doctoral Dissertation*.

Wlömert, N., & Papies, D. (2016). On-demand streaming services and music industry revenues—Insights from Spotify's market entry. *International Journal of Research in Marketing*, 314-327.

Yang, Y. (2005). Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation. *Biometrika*, 937-950.

Zawisza, M., Luyt, R., Zawadzka, A. M., & Buczny, J. (2018). Cross-cultural sexism and the effectiveness of gender (non) traditional advertising: A comparison of purchase intentions in Poland, South Africa, and the United Kingdom. *Sex Roles*, 738-751.

Zhou, Y., & Fan, Y. (2013). A sociolinguistic study of American slang. *Theory and practice in language studies*, 2209.

Appendix

A. GVIF values for both multiple linear regression models

	Model 1 *Model excl. interaction effects*			**Model 2** *Model incl. interaction effects*		
Variable	**GVIF**	**Degrees of Freedom**	**GVIF^ (1/(2*Df))**	**GVIF**	**Degrees of Freedom**	**GVIF^ (1/(2*Df))**
hate	1.608	1	1.268	14.034	1	3.746
genre	4.818	3	1.3	27.118	3	1.733
hate:genre				90.587	3	2.119
length	2.001	1	1.414	2.015	1	1.42
artist_occ	1.217	1	1.103	1.229	1	1.109
album_occ	1.13	1	1.063	1.131	1	1.063
danceability	2.099	1	1.449	2.111	1	1.453
energy	4.332	1	2.081	4.39	1	2.095
key	1.195	11	1.008	1.196	11	1.008
loudness	2.717	1	1.648	2.724	1	1.651
mode	1.165	1	1.079	1.168	1	1.081
speechiness	1.817	1	1.348	1.827	1	1.352
acousticness	2.095	1	1.447	2.106	1	1.451
instrumentalness	1.281	1	1.132	1.285	1	1.134
liveness	1.081	1	1.04	1.082	1	1.04
valence	1.557	1	1.248	1.561	1	1.249
tempo	1.074	1	1.037	1.076	1	1.037
duration_ms	1.291	1	1.136	1.292	1	1.137

TABLE 8: SHOWING THE GENERALIZED VARIANCE INFLATION FACTORS (GVIF) FOR BOTH MULTIPLE REGRESSION MODELS