What makes a song popular?

Analysis on Spotify data

ERASMUS UNIVERSITEIT ROTTERDAM

Master thesis

MSc Data Science and Marketing Analytics

Student: Francesco Vitolo, 496228

Supervisor: Dr. Vardan Avagyan

Abstract

This paper investigates which ones are the most important factors that make a song more popular than other ones, on data from 2010 until January 1st, 2021. This investigation is further extended by analysing whether external elements, recognized as "context", moderate the relationship between these important variables and popularity. To answer this question, data is extracted using the Spotify Web API, having as predictors a series of audio-based variables, and as response the "popularity" variable. A LightGBM is used to assess which ones of these variables are the most important in making a song more popular than other ones. A first analysis found the two important variables to be *tempo* and *loudness*. A second analysis included their interactions with two context variables: *liveness* and *acousticness*. This showed that *loudness* is important, and that its effect is enhanced together with *liveness*. Other important positive predictors are *energy, instrumentalness,* and *speechiness,* with an emphasis on the value of *instrumentalness* over *speechiness*.

Table of Contents

1. Introduction

1.1 Research Problem

Music has always been part of our society, and in our everyday life we are constantly exposed to music. It is enough to say that Americans listen on average 18 hours a week to recorded music (Rentfrow et al., 2011) and music on the radio, and if we consider background music as well, the amount goes up to 5 hours a day (Levitin, 2006). The pervasive presence of music within our lives is representative of the effects that music has on our brains: it enables the same euphoric responses as food, sex, and drugs (Blood el al., 2001).

The music industry keeps on getting bigger and in constant evolution throughout the years. The global revenue in 2021 totalled 26 billion USD, and global recorded music reported a growth of 18.5% in 2022 compared to 2021. Internet has completely changed the business model of the industry, as we see that nowadays 65% of global revenue comes from streaming of songs rather than with the physical sale of individual records or albums. Record companies are always looking out for new talent, totalling a global spend of 5.9 billion USD in A&R (artist and repertoire) expenditure (IFPI, 2022).

The question of what makes such products, songs or albums, a success or a "hit" compared to other ones is of particular importance for many different reasons. Music executives have spoken about this topic and given their own interpretation. Steve Greenberg, Founder and CEO of S-Curve Records has talked about the importance of elements such as catchiness and emotional resonance in successful songs, and he mentions the importance of "writing a song that people can't get out of the head" (Business Insider, 2018). Lucian Grainge, CEO of Universal Music Group has been an advocate for the use of data analytics in the music industry. He believes that by analysing data on streaming numbers, social media engagement and fan demographics, record labels can gain better insights on what are the preferences of their audience, enabling for a better allocation of resources. In particular, the emphasis is on understanding the anatomy of a song. He said that by analysing the structure, tempo and other features of successful songs, music executives can gain insights into what do the listeners find to be appealing and can better decide on how to market and sell songs to the public.

Assessing which ones are the factors that determine product success within the creative industries represents a challenge, and today this can only be explained partly, by reading data such as artists and production houses' past successes (Boughanmi and Ansari, 2021). Businesses that thrive in this industry want to gain insight into what elements make a song famous, to make more profitable choices and to always keep up with the current market trends (Herremans et al., 2014). This question is interesting also from an academic perspective. Many academics focused on how innovation affects production of diverse music (Lopes, 1992), and new research fields born: “hit song science” is defined as “the emerging field of investigation that aims at predicting the success of songs before they are released on the market” (Pachet and Sony, 2012).

The focus of this study is to shed a light on which ones are the most important factors that contribute to making a song successful or not. This will be done by analysing audio features of tracks to learn about their characteristics, extracted through the Spotify web API. These ones refer to all the elements that make up a song, which are linked to the mood of the song (danceability, valence, energy, tempo, key), to its properties (loudness, speechiness, instrumentalness), and to its context (acousticness, liveness) (Spotify for Developers, 2023). Further analysis will be delved into understanding how context affects the popularity predictors, and how much listeners are influenced by external factors. The dependent variable will be the “popularity” variable given by the dataset, which ultimately shows how much a song is popular o not compared to other ones.

Hence, our RQ is:

“*What are the drivers that influence the success of a song?”*

1.2 Managerial motivation

Being able to identify which ones are the important characteristics that make one song more famous than another one is industry motivated (Borg et al., 2011). The success of a song can have a significant impact on the revenue generated by a record label, music streaming service, or other music industry players. By understanding the drivers that influence the success of a song, music executives can make more informed decisions about which songs to promote, how to market them, and how to allocate resources such as advertising and tour support. They can also use this information to identify emerging trends and potential new stars, and to better predict which songs are likely to be successful in the future. It is important for big music companies and music investors to be always progressive by bringing new and innovative content, and to be aligned with current trends. Moreover, understanding the drivers that influence the success of a song can also help music executives to identify areas where they can improve their products or services. For example, they may discover that certain genres or types of songs are more popular among certain demographics or in certain regions, which can help them tailor their offerings to better meet the needs and preferences of their target audience.

The music industry heavily relies on streaming services like Apple Music and Spotify. These services use algorithms to recommend music to users based on their listening history and preferences. By understanding the drivers that influence the success of a song, music executives can provide streaming services with more accurate information about which songs are likely to be popular among certain demographics or in certain regions. For example, if a study found that songs with a certain tempo or a particular type of instrumentation were more likely to be successful among young listeners, streaming services could use this information to better tailor their recommendations to that audience. Likewise, if a study found that certain types of songs were more popular in certain regions, streaming services could use that information to feature those songs more prominently in those areas. By leveraging the findings of such studies, streaming services can improve their recommendations, attract more users, and increase engagement. At the same time, music executives can benefit from increased exposure for their artists and increased revenue from streaming royalties.

1.3 Academic motivation

Many academics have been focusing on analysing the anatomy of a song to try to understand what are the key elements that drive its success, using many different techniques and different datasets. However, this study will represent a more complete view of the topic for many different reasons.

Boughanmi and Ansari (2021) analysed multimedia data (metadata, acoustic features, user-generated textual data) of a dataset extracted through the Spotify web API to try to predict success of albums and playlists, using as popularity index data extracted through Billboard 200 album ranking data. Our paper will focus on individual songs rather than albums. By doing so, we will have many more songs covering a broader spectrum of musical genres, artists, and album of provenience.

Berns and Moore (2012) focused on creating a predictor of music popularity by analysing brain responses to different songs using fMRI, using for their experiment songs downloaded from mySpace in the year 2006. However, this study does not establish a relationship between likability of songs and sale patterns and its conclusions cannot be generalised for business purposes. Furthermore, the experiment was applied to a non-representative population sample, and subjective likeability cannot be generalised to markets' preferences. Our paper will have a strong focus on the application of the results from a business perspective, giving a more tangible view of why and how a song is more popular than other one, comparing different markets worldwide.

Lastly, Pachali and Datta (2022) explored which ones are the factors that drive demand for playlists on Spotify and found that customers that value personalisation and ability to discover more music are more likely to listen to playlists. Our paper will expand such research by focusing on what drives demand for individual songs. These potential findings give better and more precise insights on what are users' preferences, and businesses can leverage such results to develop tools such as Spotify recommendations more accurately and more in line with current trends.

2. Literature Review

2.1 Drivers of popularity in creative industries

For creative industries we refer to all the ones based on individual creativity, skill and talent, or the ones that have power to create jobs and wealth through development of intellectual property. These industries can be very diverse from each other, but having in common some key characteristics, which are the use of creative inputs, the production of creative outputs, which are often highly subjective and dependent on individual taste, and capture of economic value through channels such as royalties, advertising, and ticket sales (Flew, 2002). The sectors that fall under this term are namely music and art, film and photography, publishing (books, magazines etc), and electronic publishing (software, videogames, etc).

These industries are important for the economy for many different reasons. Together they account for 3% of the global GDP, employing more than 20 million people across the six and the three largest European and Asian economies, with 8 million more jobs created and an overall growth of up to 40% by 2030. This growth is driven by the fact that when consumers have more to spend and have already sated their demand for primary goods, they are likely to spend their remaining part of income on outputs of creative economy. The higher purchase power of consumers the higher growth rate of the industries (Deloitte, 2021).

This growth is not only limited to creative outputs and everything that is directly related to them. There are major spillover effects on the overall economy, which stimulate entire sectors and foster their growth as well. Creative industries act as a source of new ideas, knowledge and techniques that can be applied to many other sectors: for example, media firms that develop new approaches to content creation or distribution that is then applied to other industries (Bakhshi et al., 2008). New developments in software, music or craft sectors can make a region or a country more appealing to foreign investors, fostering the economic growth of that area and stimulating global competitiveness (Deloitte, 2021).

Sectors within the creative industries are constantly changing, and new ones keep on getting created (eg. the creation of video games in 20th century and podcasts in 21st century). Firms and individuals try to stay ahead of competition by bringing to the market new and innovative products, often leading to the disruption of existing industries and markets. This process, known as "creative destruction", is very important as these industries are characterized by short product life cycles and by the need of adapting to the changing consumer taste, hence it is essential that new products challenge existing innovations (Potts, 2009). An important characteristic is the existence of a unique environment for innovation, as there is tension between the desire for innovation and the need of commercial success (Jones et al., 2016). To facilitate this process, managers foster creativity by encouraging a culture of risk-taking and experimentation, easing communication between workers, sharing ideas, and providing support and resources (Bilton and Leary, 2002).

As the creative industries operate under unique market conditions compared to other industries, it is very difficult to establish objective metrics on what are the key elements that make new products more successful than other ones. Given the subjective nature of work it becomes challenging to predict consumer preferences accurately, creating an environment where the success of a product cannot be determined solely based on its quality or merits. Reputation and branding become vital factors, as consumers rely on established names or on artists with proven track records. These high levels of unpredictability and uncertainty mean that we very often see few products dominating entire markets (Caves, 2000).

Popularity is driven by a mixture of factors, including cultural, economic, and social dynamics. A key aspect is known as cultural resonance, which means that innovations that align with existing cultural values and trends are more likely to gain attention and become popular. Products that connect with local cultures evoke a sense of familiarity and identification, which makes them feel more authentic as well. Social factors play a pivotal role in determining success and adoption, such as word-of-mouth and social influence. Endorsement from influential groups or individuals like celebrities significantly impacts the popularity of that given product (O' Connor, 2009).

We see this by taking as an example a product like a movie. When it comes to sponsor a new movie coming up in cinemas, movie studios need to consider many aspects before releasing a movie in each country, such as cultural influences, and having a well-tailored marketing campaign. The timing of the publication is also very important, as holidays for example attract more customers. The movie genre is also a key factor, as comedies tend to attract more people than documentaries, and lastly the presence of a worldwide-known movie star in the cast makes the difference (Elberse and Eliashberg, 2003).

To tailor their products, businesses need to consider the role that subjectivity and personal preferences play in shaping commercial success, for which there is involved a psychological component. Knowing the target audience from this perspective allows for better tailoring of products and subsequent advertising campaign. For example, movies that have high emotional engagement, such as laughter, suspense, or empathy, are the ones that give the best experience to the viewers. Furthermore, movies that have a high degree of realism are the ones that have higher engagement, as viewers are more likely to immerge themselves into the plot and the characters (Eliashberg and Sawhney 1994).

These insights do not provide companies with objective metrics on why we see some movies and other products fail and some of them succeed, which makes very hard to assess what are the key elements that lead to popularity. For this, creative business executives of any sector invest time and resources in trying to understand what ultimately leads to commercial success. One of the most important sectors with similar characteristics is the music sector, for which it is very hard to pinpoint which ones are the key elements that make a newly published song or album more famous than other ones.

2.2 Drivers of popularity in music

Being a sector within the creative industries, the music industry is characterized by very strong branding and reputation, which drive commercial success of artists and every music-related product. The market poses very high barriers to entry for new artists or labels, which most of the times struggle to achieve commercial success. Music streaming platforms, which are the main way to reach out to the audience, have biased music recommendation algorithms, and they keep on recommending songs released by well-known artists or labels. This happens because such algorithms are most often based on a certain number of times a song is played, and the likeability to be recommended is increased by the fact that labels publishers of these songs put in more money and effort for marketing and exposure (Kowald et al., 2020).

Music executives are interested in having a deep understanding of why some products are more successful than others to overcome the uncertainty, by trying to understand what are the factors that influence popularity on music-related products. However, music is very subjective, and it is very hard to account for all different perceptions of music by individuals, and then to tailor successful products. Music has a very strong psychological influence on the human brain, and it has a profound ability to evoke a wide range of emotions. Elements such as melody, tempo, and harmony in different combinations can elicit different specific emotional reactions, for example fast music sparks joy while melancholic sparks sadness and nostalgia. Furthermore, according to the intensity with which we feel such emotions, out body has different physiological responses, such as increased heart rate or changes in skin conductance. All these combinations of factors spark in people many different reactions to music listening (Thompson, 2015).

Repetitive exposure to music sparks different physiological and subjective changes to people and it acts as a mood enhancer. It can cause people to have higher levels of liking of the music, increasing appreciation for a single piece and having overall feelings of positive emotions. Physiological responses tend to decrease over time as listeners become more familiar with the music, hence their bodies tend to become less reactive (Iwanaga et al., 1996).

Because of this strong psychological component, the way people engage with music is on a very introspective level and it plays a key role in their lives. People tend to attach meanings and personal interpretations to the music they listen to, which they use to navigate their personal experiences and connect with their inner selves. It serves as a powerful tool for creating personal narratives and for finding inspiration or solace. Music becomes a shared experience within social groups, and it serves to shape collective identities and foster sense of belonging. It is a mean for which people connect by finding a common ground and engaging in music-related activities, such as going to a concert or to any other event (Shankar, 2000).

Trying to have a deep understanding of what music if published in each market at a given time would lead to commercial success can be very hard because of such complicated relationships that each listener has with music. However, it is also thanks to these characteristics that companies can create clusters based on cultural, social, and psychological factors, giving fruitful insight into what drives popularity for each of them.

The image of an artist does not depend solely on his talent in performing. The reputation is built by many factors, such as stylistic choices, political affiliations, and personal relations. The ideology of an artist shapes entirely his image and depending on his actions, such as participating in social movements or engage in activism, he will attract a specific kind of audience (Gerstin, 1998).

People tend to use music as a mean to reinforce their position within the society. Musical taste can serve as symbolic boundaries that distinguish and differentiate occupational status groups. People’s preferences for certain genres of music reflect their cultural capital and social identities, including their aspiration to belong to specific social groups. Musical tastes are not merely a representation of personal preferences, but rather they represent a signature to a social group and are used to differentiate themselves to the rest (Peterson and Sikmus, 1992).

Music listening is an activity that fulfils many functions in people’s everyday life. Among the many reasons already mentioned (mood enhancement, self-expression, bonding with others etc), listeners choose to listen to different songs and music genres according to the activity they are doing in a given moment of the day. For instance, people that seek to boost their motivation to do their physical exercise might listen to more upbeat and energetic music which could match with hip-hop or rock and roll, while some others might seek to increase focus and relaxation while studying with some classical music (Sloboda, 1999).

Thanks to these insights, music publishers can segment and target the market with more accuracy, knowing what certain audiences would enjoy listen to. Nevertheless, such information helps in understanding which are the past or present preferences of the audience and does not give any insight into what the next trends could be. This is to say, music executives cannot understand a priori what will make a song successful and cannot make solid predictions in order for them to stay ahead of competition, which is essential to survive in businesses such as the creative industries.

With the rise of streaming platforms as main tool for music listening, companies had access to a wide range of insights with a better accuracy than ever. Metrics such as number of streams, per song or musical genre, allow for a better visualisation and understanding of the market, and people have started to leverage data analytics to extract patterns, allowing for more profitable decisions and resource allocation. Moreover, businesses have started to focus on the "anatomy" of a song. This refers to the analysis of the elements that make a song, such as its tempo, its key, and its loudness, thanks to which more meaningful insights can be extracted on what exactly becomes commercially successful. Such analysis is done also for predictive purposes, as companies can see how these elements put in different combinations are likely to correspond to a high number of plays or sales.

This application of data analytics has received a wide interest from the academia, for which many researchers have experimented many machine learning techniques focusing every time on a different aspect of music data (Borg et al., 2011).

Weihs et al. (2007) explored the application of classification techniques in music search, which are very useful for the purpose of gaining insights into patterns, structures, and relationships within musical compositions. They emphasized on data pre-processing and on feature selection to obtain accurate results, and they presented case studies in which they showcase these classification models for music recommendation systems, music genre classification and composer attribution. The same techniques can be used to account for the dynamic nature of music and for individual taste and preferences.

Aljanaki et al. (2017) developed a benchmark for emotional analysis in music. They gathered data from various musical genres, and they collected some listeners' individual thoughts on the songs just listened. To evaluate the performance of the algorithm, they measured the accuracy of recognizing specific emotional labels and the intensity and valence dimensions associated with each emotion. This algorithm has many possible applications, including music therapy, better music recommendation systems, and companies can assess which ones are the trends that lead the most to popularity.

Herremans et al. (2014) conducted a study on predicting dance hit songs with research focused on developing a computational model that could identify potential hit songs within the dance musical genre. They gathered a songs dataset with metadata, such as artist name, album name, and audio features, such as the tempo, the BPMs and the key. In their experiment, they found that rhythmic qualities of dance songs are the crucial factors for its commercial success. Second, they found that a machine learning technique, in this case a Support Vector Machine (SVM) can effectively predict success of songs, and it can be implemented to assess which songs have a higher commercial potential over other ones.

Lee and Lee (2018) have analysed a huge dataset of songs with the goal to shed a light on the relationship between popularity and acoustic features, lyrics, and social factors. Their findings are that popularity as a general meaning of a song, which is not exactly corresponding to the number of times a song has been streamed, is a combination of these factors, and that the loudness, tempo, and energy out of all the acoustic features are the ones that tend to affect popularity the most. The authors have shown that the same model can be used for prediction and could be used as a very helpful asset by firms and artists that try to maximize the commercial success of their products.

Yang et al. (2017) drew several conclusions regarding audio-based hit song prediction using convolutional neural networks (CNN). They concluded that neural networks are able to extract meaningful patterns and representation from audio data, and that they perform with better accuracy compared to more traditional machine learning techniques. They say that feature engineering for audio-based predictions is very important: the model will result more accurate if the features to be analysed are chosen carefully, which requires a deep understanding of the domain and of its market. Lastly, they confirm that audio features are crucial in determining a song’s success, and similarly to other research they find rhythm, tempo, and energy to be the most important ones, as they strongly influence the appeal and popularity of the song itself.

To contribute to all the past research done in this topic, in this paper we will carry out an analysis on a more comprehensive dataset. The dataset will feature a huge variety of songs, published in many different markets and having different music genres, with the aim of identifying which ones are the root factors that make a song more streamed and commercially successful than another one.

3. Data

3.1 Data Collection

To extract the dataset containing all the needed information we first had to choose the source. This was done based on criteria such as accuracy of data, its completeness, available amount of data, user base, and accessibility to its API. After careful consideration we narrowed down the choice to the Spotify web API, which allows for collection of random songs and information about them such as its key and its popularity.

Spotify is the world's biggest music streaming provider. The platform features a library of more than 100 million songs and the biggest market share sizes of monthly active users, with more than 510 million active users in Q1 2023 (Statista, 2023). Its main competitors are other platforms like Apple Music and Tidal. Furthermore, not just as streaming provider, Spotify's role has been pivotal in the breakthrough of the podcast industry, leading all the competitors to adapt to the demand for such product.

The Spotify API allows users to retrieve soundtracks, albums, and playlists based on certain criteria inserted. Our goal was to randomize the search as much as possible, and for this we used the Spotify's search command, which sends back results based on specific parameters, such as type (eg. album, playlist, or track, in our case track), keyword and market in which that content is available on Spotify. This function sends back a maximum of 50 elements per batch, and to build our dataset we needed multiple batches of 50 elements each. We started with requesting 2000 batches, theoretically leading to a dataset of 100,000 songs. To ensure a high level of randomness and importantly to avoid duplicates, we used as keywords random combinations of all the letters of the alphabet, and we also used an offset. However, many duplicates were returned, which led to a reduction in observations compared to our initial forecast. We further reduced the size of the dataset by selecting songs that were published between 1st of January 2010 up until 1st of January 2021. This was done because of how the variable popularity is accounted for by Spotify. The songs that are being played a lot recently have a bigger popularity than the ones that are played a lot in the past (Spotify, 2021), and to partly try to delete this effect we did not include songs that were published in the last 3 years. This filtering led to a dataset of 12,314 observations, each observation representing a single track.

3.2 Data Description

The dataset contains in total 14 variables, 2 identifiers, and 4 time-based variables per song. Below we can see Table 1 with a detailed description of each variable, together with Table 2 showing the descriptive statistics of the numerical variables.

	Name	Type	Description
Identifiers	*id*	character	Song unique identifier
	name	character	Song name
Time-based variables	*album.release_date*	date	Release date of the album of the song
	release_month	date	Release month of the album of the song
	release_year	date	Release year of the album of the song
	months_since_2010	numerical	Number of months that has passed since January 2010
Audio-based variables	*danceability*	numerical	Description of how suitable a track is for dancing, based on combination of many elements including tempo and rhythm stability. Values from 0.0 to 1.0
	energy	numerical	Representation of perceptual measure of intensity and activity. Values from 0.0 to 1.0
	key	integer	The key the track is in. Values from 0 to 11
	loudness	numerical	Overall loudness in decibel (dB) of a song. Values from -60 to 0
	mode	integer	Indicates the modality (major or minor) of a track. Major is 1 and Minor is 0
	speechiness	numerical	Detects presence of spoken words in a track. Values from 0.0 to 1.0. The closer to 1 the more likely the recording is made entirely of words
	acousticness	numerical	Confidence measure of whether the track is acoustic. Values from 0.0 to 1.0
	instrumentalness	numerical	Predicts whether a recording contains no vocals, with values from 0.0 to 1.0. The closer to 1.0 the more instrumental the recording will be
	liveness	numerical	Detector of audience in the recording. The higher the more likely the song was live recorded. Values from 0.0 to 1.0
	valence	numerical	Descriptor of musical positiveness of a recording. Values from 0.0 to 1.0. Valence closer to 1 stands for more cheerful and happy songs, while the opposite stands for sad and angry songs.
	tempo	numerical	Estimated tempo of a recording in beats per minute (BPMs). Values from 0 to 235.135
	duration_ms	integer	Duration of track in milliseconds. Values from 7,907 to 3,851,651
	time_signature	integer	Estimated time signature, specifying how many beats are in each bar. Values from 0 to 5

Variable of Interest	*popularity*	integer	The popularity of a track is a value between 0 and 100, with 100 being the most popular. The popularity is calculated by algorithm and is based, in the most part, on the total number of plays the track has had and how recent those plays are. Generally speaking, songs that are being played a lot now will have a higher popularity than songs that were played a lot in the past.

Table 1: Description of variables and identifiers included in the dataset, based on Spotify documentation (2023)

	Minimum	**Maximum**	**Mean**	**Median**	**Std. Dev.**
danceability	0.000	0.986	0.603	0.628	0.185
energy	0.000	1	0.597	0.631	0.245
key	0.000	11	5.225	5	3.628
loudness	-55.628	4.363	-9.096	-7.543	5.734
mode	0.000	1	0.603	1	0.489
speechiness	0.000	0.963	0.135	0.063	0.167
acousticness	0.000	0.996	0.321	0.191	0.327
instrumentalness	0.000	1	0.178	1.67E-05	0.333
liveness	0.000	0.993	0.202	0.125	0.174
valence	0.000	0.993	0.464	0.457	0.257
tempo	0.000	235.135	120.261	120.075	30.724
duration_ms	7907.000	3851651	221861.3	206795.5	116782
time_signature	0.000	5	3.889	4	0.515
popularity	0.000	83	28.625	24	21.622

Table 2: Descriptive statistics of variables included in the dataset

The first and second group represent time-based variables and songs identifiers, which are *id*, *name*, *album.release_date*, *release_month*, *release_year*, and *months_since_2010*. The *id* is the unique song identifier, and it was needed to extract all the audio features related to the specific song and to check for duplicates. The variable *album.release_date* was needed to filter out songs published earlier than 2010 and later than 2021 , and to have variables such that we are able to account for the presence of any trends across the years. Since these songs were released on a span of 10 years, in which the type of audience has changed, it is important that we account for the publication year in our analysis. Furthermore, because we know that such products are highly subjective to seasonality (Elberse and Eliashberg, 2003), accounting for the month of publication is also very important. The third group represents audio-variable features, which are all the variables that represent the “anatomy” of a song. The fourth one is made by our variable of interest, which is *popularity*.

We can see some interesting insights from our dataset before diving deep with the simple and main analysis. Figure 1 below shows the mean value of the duration per song based on their release year.

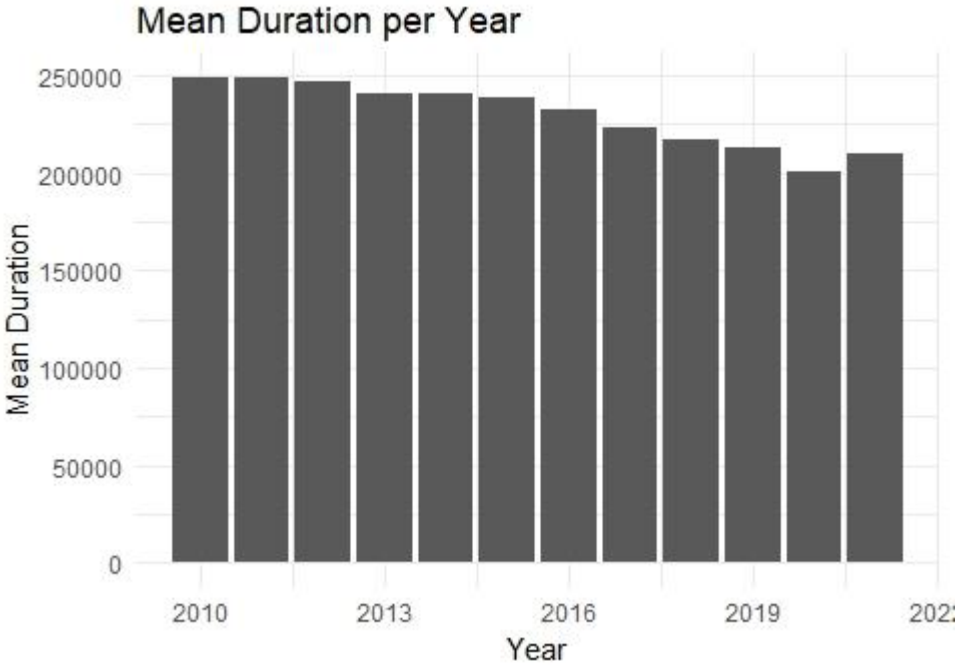

Figure 1: Mean value of duration_ms per release year

We see that the mean value of duration of songs is descending, meaning that the closer to the present a song is released, the shorter will be its duration. As shown by research published by UCLA (2020), the average song length has fallen from 4.20 to 3.15 minutes from 1990 until 2020, possibly done due to a overall reduction of the attention span by listeners. We also see that most of the songs tend to have a high energy level, with a mean value of 0.59. A song with high energy is usually associated with high popularity, as it evokes feelings of joy and happiness. For this similar reason, we see in the dataset that the mean danceability value is at 0.6, showing a high value and expressing the presence of many danceable songs (Namica, 2020). Furthermore, we notice that these songs that we will analyse are mostly instrumental, as the mean speechiness value is 0.134, which is quite low. Songs that have high popularity values are generally associated with catchiness, corresponding with how much people are likely to remember the song, which is given mostly by the lyrics of the song itself (Namica, 2020). Given these characteristics, it will be very interesting to explore the relationships that the characteristics of these songs have with the popularity value.

4. Methodology

4.1 Simple Analysis

To try to give an answer to our main research question we need to establish whether there is a relationship between the audio-based variables of our dataset and the variable of interest, popularity. Before stepping into more complex analysis, we need to gather some preliminary insights on our data, which is needed to also check if these preliminary results are in line or show some variation compared to what has been already done in the literature. Hence, we decided to perform a multiple linear regression. A multiple linear regression establishes whether there is a relationship between many variables and one dependent variable, and it is an extension of the linear regression. The formula looks as follows:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n + \varepsilon$$

We have Y, which is the dependent variable, β_0 the constant, β_n which are the coefficients of the independent variables and x_n which are the independent variables. This model aims to reduce the squared errors by fitting a linear line through the data, with a process called OLS (ordinary least squared), assuming that the relationship between the variables is linear (Echambadi et al., 2007). To have the best overview of all the variables, we will include all the variables to assess their relationship with popularity. However, to avoid overfitting due to the presence of many variables, we decided to leave out the time-dependent variables, and to focus on this section of preliminary insights only on the influence of audio-based variables on popularity. Table 3 below shows the results of the formula called:

	Dependent variable:		
	popularity		
danceability	0.097	liveness	-3.179***
	(1.288)		(1.135)
energy	-9.280***	valence	-2.647***
	(1.470)		(0.887)
key	0.074	tempo	0.009
	(0.052)		(0.006)
loudness	0.567***	duration_ms	-0.00001***
	(0.059)		(0.00000)
mode	0.754*	time_signature	0.174
	(0.388)		(0.380)
speechiness	-22.656***	months_since_2010	-0.027***
	(1.248)		(0.006)
acousticness	2.099***	Constant	47.858***
	(0.810)		(2.283)
instrumentalness	-12.893***		
	(0.684)		

Table 3: Linear regression results

We notice some interesting insights, as we see some differences with the model-free evidence described in the previous section. From the table we see that the most significant coefficients are *energy*, *loudness*, *speechiness*, *instrumentalness*, and the *duration*. The biggest difference is made by the coefficient of *energy*, as an increase in *energy* by 1 point leads on average to a reduction in *popularity* by more than 9 points. This is followed by the coefficient of *duration*: it is very interesting to see that the more a song last, the less popular it will be by more than 9 points. Furthermore, if a song contains more vocals and it is more instrumental popularity is reduced by respectively 22 and 13 points. *Loudness* has a positive impact, and it says that the louder is the recording the more popular it will be by more than 0.5 points. Other significant coefficients are *acousticness*, *liveness*, *valence*, and *mode*. The first two are more significant, with *acousticness* with a positive effect on *popularity* and *liveness* with a negative one. *Valence* has a negative effect and *mode* a positive one.

These preliminary results give some very interesting insights, some of them in line with the existent literature. Like it was found out by Lee and Lee (2018), *energy* and *loudness* play a pivotal role in determining commercial success of a recording. It is interesting to see that in this dataset, songs that are more sad or melancholic tend to be more listened to rather than more joyful ones, as shown by the negative coefficient of *energy*. *Duration* plays a pivotal role too, and the fact that its coefficient is negative shows that listeners enjoy listening to shorter songs rather than longer ones. Research published by UCLA has shown how the average length of a song has fallen from 4:20 to 3:15 from 1990 until 2020 (UCLA, 2020).

Multiple linear regression can be very efficient in doing a “quick-and-dirty” analysis, as it gives immediate insights and allows for easy interpretations. However, there are some shortcomings of using this method with any dataset, specifically a dataset like the one in question. First, linear regression does not handle well unbalanced data. In this case, even though there are no missing data, more than 10% of the total number of observations the popularity value is 0, which can negatively affect the model’s accuracy in predicting relationships between variables. Furthermore, in the dataset there are observations of songs which were published in different years and at different points in time, which cannot be compared all to the same level, that is something that linear regression does. For instance, if there are any major changes in music listening trends linear regression would not account for those (Detienne et al., 2003).

In the following section we will present some methods which we would use to answer our main research questions. We will describe the pros and the cons of each one and at the end, we will choose the most appropriate one for our dataset and our research.

4.2 Candidate methods

Balanced random forest trees

Balanced random forest is a variation of the original technique, which aims to handle unbalanced data, and it is very straightforward and easy to understand technique due to their linguistic nature (Martens et al., 2008). Random forest is a machine learning algorithm used for classification and regression tasks, which is an ensemble of decision trees. Decision trees are flowchart-like structures where internal nodes represent features, branches represent decision rules, and leaf nodes represent class labels or output values. Random forests improve the accuracy by combining many decision trees. Each tree is trained on a random subset of the data and random subset of features, and the final prediction in a random forest is obtained by aggregating the individual predictions of all the decision trees (Ali et al., 2012). In traditional decision trees, each node split is based on a feature that maximizes the information gain. However, this leads to biases towards the majority class, having little to learn about the minority class. Balanced random decision trees introduce modification in the splitting criterion taking class imbalance into account with a combination of random oversampling of the minority class at each decision node.

This technique will tackle the issue of having one class which is consistently bigger than the other ones, in our case the number of songs that have a popularity value of 0. However, balanced random forest trees are not particularly suitable for handling time series data. This means that it will not explicitly highlight any trend differences across time, which is something we would like to see in songs published across a span of 10 years. This happens because this technique treats each row independently, ignoring any kind of time series correlation.

LightGBM

LightGBM is an implementation of gradient boosted decision trees (GBDT), meaning that it is a model that uses weaker learners, in this case decision trees, to create a stronger ensemble model. The framework starts with the initial model (weaker learner) which is trained on the data, of which it computes differences between actual values and prediction of the initial model, representing the errors that the model was unable to capture. The gradient descent algorithm is in charge of finding the best direction for improving the model, determining in which directions the model's predictions need to be adjusted to reduce errors. New base learners are added with each iteration, but this time they are trained to predict the negative gradient of the loss function instead of original target values, with the goal of reducing residuals. All these predictions are combined in the end to form the ensemble model, which is used to make predictions on unseen data. These predictions are weighted based on their performance during the training phase. The algorithm prevents overfitting by employing regularization techniques like shrinkage and feature subsampling, which control for the contribution of each weaker learner (Ke et al., 2017).

LightGBM is part of the gradient boosting framework family of algorithms, including other frameworks like XGBoost and AdaBoost. LightGBM shares many of the advantages of XGBoost, such as sparse optimization, parallel training, multiple loss functions, regularization, bagging, and early stopping. The major difference between the two methods lies in the construction of trees, as LightGBM does not grow tree-wise (row by row), but rather it grows leaf-wise, choosing the leaf that according to the algorithm will lead to largest decrease in loss. LightGBM does not look for the best splitting point on sorted feature values, but rather it implements a highly optimized histogram-based decision tree learning algorithm, yielding great advantages on both efficiency and memory consumption (Ke et al., 2017).

LightGBM is a very suitable method for our analysis. Firstly, it is suitable for unbalanced data for many reasons. It assigns higher weights to minority classes and lower weights to majority classes, correcting for any differences. The model supports balanced bagging, which involves balancing class distribution with each bagging iteration, and its algorithm enables to assign higher gradients to misclassified samples from the minority class, which allows the framework to learn about the patterns of the minority class itself.

LightGBM is suitable also for tackling the time-series data issue. The algorithm can learn temporal patterns and include the influence of time in the predictions thanks to time-related features included in our dataset, such as the release month and year of the album containing the song.

LASSO Regression

LASSO (Least Absolute Shrinkage and Selection Operator) is a linear regression technique that performs variable selection and regularization, and it was originally introduced to address the limitations of traditional linear regression, specifically when it comes to dealing with big datasets with many observations. LASSO regression minimizes sum of squared errors adding an additional penalty term, according to the formula:

$$\sum_{i=1}^{n}\Big(y_i - \sum_{j} x_{ji}\beta_{ji}\Big)^2 + \lambda \sum_{j=1}^{\rho} \parallel \beta_j \parallel$$

where y is the vector of observed values of the dependent variable, x is the predictor variable, β is the vector coefficient to be estimated, λ is the parameter that regulates the strength of the penalization term, and ‖ β_j ‖ represents the sum of absolute values of the coefficients. The penalty term shrinks some of the coefficient toward zero, having as a result some of the predictors to be exactly zero hence, to be excluded from the model, and this process is known as variable selection. The penalty term imposes a constraint on the magnitude of coefficients. As λ increases, more coefficients are shrunk to zero, leading to a simpler model with fewer predictions (Tibshirani, 1996).

LASSO is an easy-interpretable method, and it can analyse big datasets. It suits our purpose because it eliminates the variables that it considers to be non-relevant, in our case non-relevant in making popularity higher or lower. In doing so, it can also account for time differences, and it can highlight whether timing has an influence or not, after the correct data pre-processing. However, the shortcoming of this model in our case is that it does not handle well unbalanced data. Unlike LightGBM and Balanced decision trees, where class weighting is done automatically, in LASSO we would need to do class weighting by ourselves, making the model more suitable for our needs (Tibshirani, 1996).

ARIMA

ARIMA (Autoregressive Integrated Moving Average) is a model used for understanding and describing the underlying structure of data, and to forecast future points in the series. As the name suggests, the model is divided into three parts. The Auto-Regressive part (AR) represents the relationship between an observation and a number of lagged observations, and it is based on the idea that the current observation has a linear relationship with the lagged ones. The AR part captures the trend in the time series, assuming that if there has been a recent increase or decrease it is likely that it will keep doing so for future values. The Integrated (I) part refers to making the time series stationary, that is a time series with its statistical properties (mean, median, etc) non-variable over time. The Moving Average (MA) part refers to modelling the relationship between an observation and a residual error from moving average model applied to the lagged observations.

ARIMA does an excellent job in understanding time series data and in extracting valuable insights regarding future trends, which is very important in understanding where the trends are going. However, the model is not properly suitable to our research because it does not pinpoint which variables are the most important themselves.

4.3 Table summary

Table 3 shows an overview of the methods described in the previous section. These methods are suitable for our main research question, while they address the issues of linear regression.

Model Name	Class Imbalance	Time-series data	Feasibility	Interpretability: assessing variable influence	Interaction effects
Balanced Random Forest trees	Suitable	Not particularly suitable	Very easy to implement and to run	Very straightforward to understand and interpret	Yes
LightGBM	Suitable	Suitable	Easy to implement and run	Depends on the complexity of the model	Yes
LASSO Regression	Not suitable	Suitable	Less easy to run due to data pre-processing	Easy to interpret	Yes
ARIMA	Not suitable.	The most suitable method of all four	Requires data pre-processing	Not suitable	Yes

Table 4: Comparison of candidate methods for main analysis

4.4 Model selection

After having reviewed the four models the one that we will use for our analysis is LightGBM. LightGBM will highlight which ones are the variables that affect the most popularity, which is the main goal of this research. While doing so, it will account for other factors, like unbalanced data and time series data. LightGBM is designed to handle well unbalanced data, in this case it will take care of the fact that there are many observations with a popularity value of 0. LightGBM is a technique that learns about time factors that might influence the predictions, hence it will account for the fact that a song published in 2010 might have a higher popularity value than a song published in 2021. Lastly, the model allows for the introduction of interaction effects, which will help us to better understand the drivers of popularity and which factors combined might have a higher influence.

The model will be implemented as follows. First, we will split the data into train and test sets, which is needed to assess the performance of the model. The data will be transformed into the “lgb.Dataset” format, which is the preferred format for LightGBM, as it is optimized for both memory efficiency and training speed. The model has a set of parameters, such as number of boosting rounds and the early stopping rounds. To make sure we get the best results out of it, we need to finetune these parameters, which we will be able to do by trying many different combinations on the dataset itself. Once we have the results of the model, we will plot the graph of importance of the features, which will give us clear visualization on the variables of interest, helping us in answering to the main research question of this paper. Subsequently, we will elaborate on the interaction effects to see if there are other variables that moderate the relationship between popularity and the main variables that affect its value. All of this will be shown more in depth in the following section, the results section.

5. Results

5.1 Main Analysis

Figure 2 shows the ranking of feature importance, which shows the features ranked according to their contribution to the model. This is needed to know which of these features we will consider to understand what drives popularity in songs. Figure 3 shows the SHAP values of the variables of the LightGBM, which we will use to understand which ones of the relevant predictors has a positive effect on *popularity*. A positive value means that the predictors have a positive effect on the predictions, in this case on the value of *popularity*, and inversely the negative ones have a negative effect.

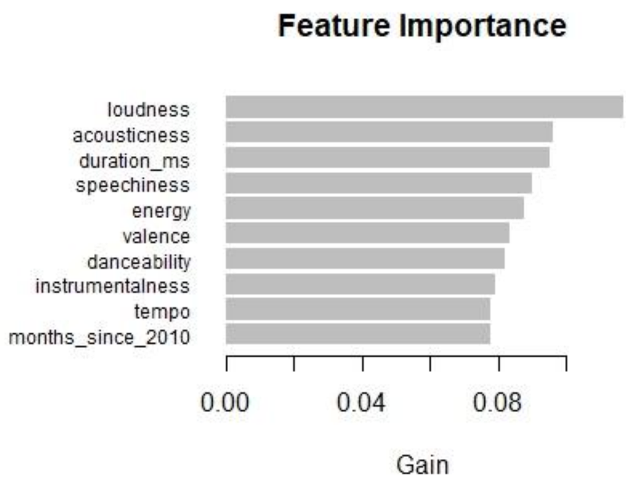

Figure 2: Feature Importance ranking

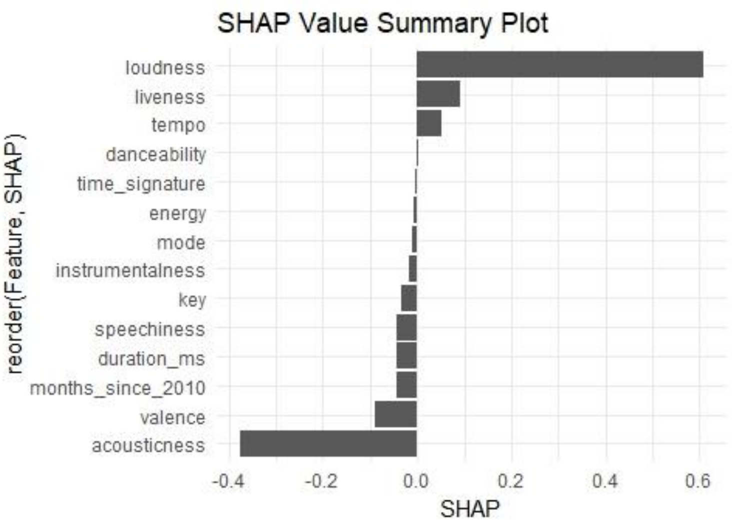

Figure 3: SHAP Values of the LightGBM model

From this graph we can see that the predictors that have a positive effect on the *popularity* value are *loudness, liveness, tempo,* and *danceability*. However, according to the importance graph we see that the two relevant ones are *loudness, tempo,* and *danceability,* with the last one having a positive effect close to zero. We can see that there are some different results compared to the previously done linear regression model. While they both captured *loudness* as the most important variable leading to a positive *popularity* value, this model shows other two variables to be positively correlated with *popularity*. Linear regression showed *acousticness* as the other positively correlated variable, while in this model it is the one with the biggest negative effect. We see another interesting insight with *tempo*, meaning that this model predicts that if a song has a higher tempo (i.e., it is faster) it is more likely to be more popular than a song with a lower rhythm.

5.2 Interaction effects

After having found the predictors that are responsible for a high level of popularity, we will delve deeper into some of the interactions of these predictors, and whether and how such have an effect or not on the popularity itself. For this analysis, we will use the context variables, which indicate whether there are other factors in the recording of a song. The variables are *liveness*, and *acousticness*, where the first one indicates whether a song has been recorded live or not, and the second one indicates whether the song is acoustic or not, which means that the song was played with instruments that have no electronic amplification. We will see the interactions between these two variables and *loudness* and *tempo*, which are the two of the three top contributors to a high level of popularity. This will lead to the addition of 4 predictors to our model. This is done for us to see whether the context of a song, which is whether there are external factors that might have an influence, for example the presence of a live crowd cheering, might moderate the effect that variables like loudness have on popularity.

Figure 4 shows the ranking of feature importance, while figure 5 shows the SHAP values of the LightGBM model, including these 4 new predictors.

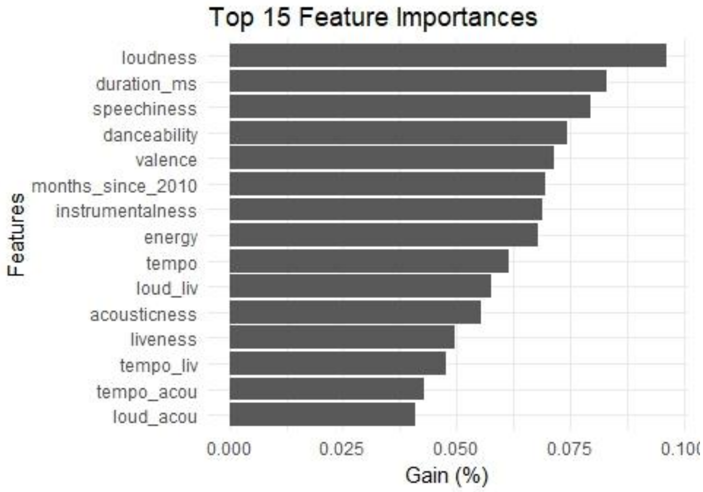

Figure 4: Feature Importance ranking

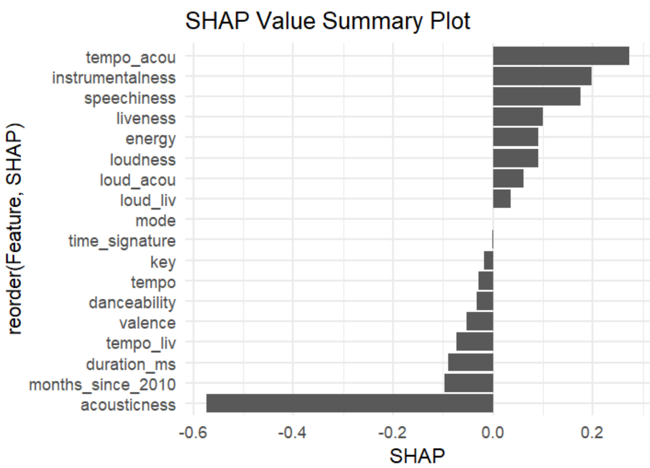

Figure 5: SHAP values of the LightGBM model with interaction effects

The introduction of these new predictors brought significant changes to the original model. We can see now that variables like *instrumentalnes, speechiness,* and *energy* are at the top of the chart, and they are shown in the importance plot. *Instrumentalness* has now a strong positive effect compared to the previous model, followed by *speechiness* and *energy*, which all are shown in the importance plot as well. *Loudness* has still a positive impact, but of a much less magnitude. *Tempo* itself now has a negative impact, while *acousticness* now does not appear in the importance plot, meaning that it is not significant for this analysis. We see that most of the interaction effects themselves are not significant, except for *loud_liv*, which is the interaction between *loudness* and *liveness*, which contributes positively to popularity. This new predictor is telling us that the effect of loudness has a strong and significant effect on popularity when songs are live recorded.

6. Conclusion

6.1 Summary of the research

The goal of this research paper is to assess the elements that are responsible for making a song more popular than other ones. The angle from which the analysis is done is through analysing what are the elements that make up a song, or the "anatomy" of a song. The contributions of this research are both managerially and academically justified. Understanding what makes a song popular can help businesses in the music industry in making more profitable decisions regarding publishing their songs or albums. Furthermore, this research adds to the previous literature on how it is possible to predict whether a song will be a "hit" or not even before it gets published, known as "hit song science" (Pachet and Sony, 2012).

The elements analysed are audio-based variables, which can be grouped into variables that are linked to the mood, to the properties, and to the context of the songs (Spotify for Developers, 2023). The dataset in question was extracted through the Spotify Web API, which allows through API calls the possibility to access its whole songs, albums, and podcasts database. By means of a random search, we extracted a dataset of 12,314 songs, published at any time between January 1st, 2010, and January 1st, 2021. All the audio-based variables were used in our analysis, and in order to account for the effect that time throughout the years, we added a variable called *months_since_2010* for our main analysis.

We carried out firstly a preliminary analysis to explore the data and whether these first results are in line or not with the previous literature. We chose the multiple linear regression method, which gave back some interesting insights. The most interesting ones are related to *energy*, *duration*, and *loudness*. The more energetic a song is, the less popular it is (an increase by 1 point in *energy* leads on average to a reduction in *popularity* by more than 9 points). An increase of *duration* by 1 point corresponds to less *popularity* by more than 9 points. *Loudness* has a positive impact, and it says that the louder is the recording by 1 point the more popular it will be by more than 5 points. Lee and Lee (2018) found similarly that *energy* and *loudness* play a pivotal role in determining commercial success of a recording. *Duration* shows that listeners enjoy listening to shorter songs rather than longer ones, like research published by UCLA, that has shown how the average length of a song has fallen from 4:20 to 3:15 from 1990 until 2020 (UCLA, 2020).

Our main analysis consisted in applying a LightGBM to the dataset. We chose this method as it can give better insights on the relationship between the variables and popularity, and because it can capture time trends that linear regression is not able to do. Because the songs were published in a span of 10 years, LightGBM can learn from the data whether there are some significant time-related effects that might influence the popularity score of songs that are published earlier. Furthermore, it can handle unbalanced classes, which is useful in our case because more than 10% of total observations have 0 as *popularity* value, that might cause our predictions to be less accurate.

We performed a first analysis on the dataset, which gave *loudness* to be the most important predictor of a popular song, together with *tempo* and *danceability.* To further delve deeper into what other combinations of factors might affect the popularity of a song, we explored the interaction between two of the top elements that make a song popular together with the two context variables, which are *liveness* and *acousticness.* This is done to see whether the context, that is other factors that might affect the listeners' experience apart from just the song itself, moderates the relationship between the key elements and the popularity itself. We saw the interactions between these two variables context variables, together with *loudness* and *tempo.* By adding 4 new predictors we notice that the original graph showing the predictors according to their contribution has changed. Variables like *instrumentalnes, speechiness,* and *energy* are at the top of the chart, and they are shown in the importance plot. The other positive predictors are *loudness* and *loud_liv,* the last one being the interaction effect between *loudness* and *liveness,* meaning that if songs are live recorded loudness has a strong effect on popularity.

6.2 Interpretation of results

An important element that occurred throughout the various analyses to positively influence *popularity* is *loudness*, meaning that in this dataset, songs that are recorded at a higher volume tend to be more popular than other ones. This is in line with what has been the trend of the last years, known as "loudness war" (Devine, 2013). There has been a trend in the music industry consisting in increasing the loudness of recorded music over time, consisting in making the quieter parts of a song louder. This is because of many factors, namely technological improvement for music recording and attempt to grabbing the attention of listeners. Thanks to technology, it started to become possible for recording music at a higher volume without distorting the sound. Because typically the louder songs within an album are the ones that stand out, all the songs were recorded at a generally higher volume than before. Louder songs also can win the attention of listeners when they are in busy spaces, such as gym or public spaces, being more suitable for listening throughout the entire day and not just at home (Devine, 2013). People tend to better remember a louder song rather than a less loud one, because, if recorded properly, it is easier to catch all the instruments playing at the same time, and because it overall gives to the listeners a better musical enjoyment (Vickers, 2011).

However, through the results we notice that the effect of *loudness* on *popularity* relies on the surrounding context, which in this dataset seems to have a positive effect. Live music has the power to engage better with the crowd of listeners, creating a unique experience and bond. During live performances, people tend to show all together this engagement with physical movements, like moving the head back and forth at the same time, which is given by many factors like the presence of many other people, like a big social event, the presence of performers, and the general amplitude of the sound (Swarbrick et al., 2019). Comparing this to our results, it might be that listeners tend to value more loud songs which are live recorded because they give them a better engaging feeling, and they feel like they are at a live performance, which can be explained by the *loud_liv* score leading towards a positive *popularity* value.

Another important element that we see is *energy*. Energy represents whether a track feels "energetic" or not, and by that we mean a track that feels fast, loud, and noisy. In our dataset, we have a positive impact on popularity by energy, which means that the more energetic a track is the more popular it will be. Similarly, loudness and loudness and liveness together also have a positive contribution, contributing to the conclusion that more energetic tracks tend to have a higher popularity than less energetic and more "relaxed" tracks. Similar findings are from the experiment carried by Lee and Lee (2018).

We notice that in this dataset there are significant and positive indicators of *speechiness* and *instrumentalness*. Songs that are more mainstream than other ones are energetic songs, as stated before, which people "cannot get out of their head" (Business Insider, 2018). Songs that can be considered such are considered more wordy songs rather than instrumental. This is because songs with lyrics can be interpreted in many and more easily ways, and we as humans tend to be more psychologically drawn to the human voice rather than simple instruments. Songs with more words can be sung along, it connects more with the crowd, and it is easier for people to remember parts or all the songs in their heads (Musical Mum, 2021). However, in this dataset we see that *instrumentalness* has a higher value than *speechiness*, suggesting that songs with more instrumental over speech tend to be more popular. This means that listeners in this dataset valued more the musical composition part rather than the singing part, which opens to a series of many different music genres to eventually take over the market. This represents an important shift in consumer preferences, which might represent a broader or cultural musical trend, where listeners are looking for more immersive and melodious experiences, rather than mere storytelling.

6.3 Managerial Implications

There are many categories of people that work within the industry for which the results of this research have implications. The main ones are music executives, artists, and streaming platforms, like Spotify.

This research showed that loudness itself has a stronger effect on popularity if the song is live recorded, which might give to the listeners more of a "live" feeling and a better immersion into the song. From the Billboard 200 chart (Billboard, 2023) we can see that in the top 10 albums there are no live recorded tracks at any live shows of the artists. With this regard, the advice to music executives is to include more live recorded tracks in the albums at a high loudness. With a high loudness, songs are delivered with a higher intensity, creating a stronger bond between the artist and the listeners, and it ensures that all the instruments used for the recording can be listened one by one, being able to better capture the attention of the listeners. This is enhanced by the live recording, which sparks a higher energy and a higher engagement with the song itself. Artists are advised to take this aspect into consideration when recording their songs, as this might influence their whole style, image, and the way they interact with the crowd. If more live performances are scheduled, it gets more and more important for an artist to engage with its crowd, and hence to adapt their behaviour to such occasions. Streaming platforms such as Spotify and Apple Music can use insights about song popularity to adjust their recommendation algorithms, making sure that users end up listening songs that match the available preferences. Such knowledge can direct both algorithmic and human curated playlists, prioritizing songs aligned with identified popular brands. Additionally, these platforms can actively expose users to new potential hits, especially from up-and-coming artists. By incorporating discovered analytics into their programs, they can increase user engagement, simplify music discovery, and provide valuable feedback to artists and bands.

Another contribution of this research is that it highlighted a small but rather important difference between instrumental and more vocal songs, where in this dataset instrumental songs seem to be more famous than the other ones. Considering that this might be part of a much broader shift in consumer preferences, there are many things that music executives can do to stay ahead of competition. Album should feature more instrumental tracks, and marketing campaigns should focus on highlighting the instrumental composition, using them as soundtracks on commercials, movies, and videogames. New collaborations should be explored between mainstream artists and more instrumental musicians, such as electronic music producers, which can create unique sounds and explore new realities. Furthermore, even though in the western world the mainstream music is with English lyrics, instrumental music resents less of language barriers, and can be more easily marketed and published in worldwide markets. Live performances incorporate more instrumental music, which is done to create a higher level of engagement with the crowd, hence there should be more live recordings in albums and more live performances of artists planned, which is in line with what previously stated about the live recordings.

6.4 Limitations & Further research

This research focused on the elements that contribute the most in making a song popular, by means of an analysis on audio-based variables extracted from Spotify. This makes the whole research based entirely on the accuracy of the data that the Data Science team of Spotify generated and made available for the public. This needs to be noted as a weakness for the research, because there is no specification on their website on how these data are created and with what accuracy they are generated. Furthermore, the dataset is of 12.314 observations, which can be thought of a limited amount of data, considering the nature of the research question and the implications of this research to music executives. However, due to many duplicates returned and network interruptions it is not always easy and possible to extract a considerable number of songs at once.

Song preference is a personal choice, which makes very hard to pinpoint factors that can go well universally and understand what makes a song famous for “everybody”. We have seen that music is used by people for many different activities through their daily life. For instance, while studying or while going on a run or having a workout session. This influences the occasions for which a listener would switch to a certain type of music throughout the day, based on what he is doing in that moment. The analysed dataset does not account for external factors such as the context or the current activity the person is doing while listening to that specific song. One way to try to account for these differences is trying to analyse the different music genres, which we will see later in the “Further research” section.

Further research should focus on gathering data from many different sources and not just one. For instance, the popularity score can be gathered using the Billboard API, which allows for the visualization of many different charts throughout the years, as the first Billboard Music Popularity Chart started on 1940. This allows for a diversification of data collection and hence there can be done more generalised predictions and with more solid implications for managers.

To have a more comprehensive understanding of the causes that make songs popular and how that varied over time there should be done analysis on a bigger time span. In this case, even though the span was of 10 years, the method LightGBM highlighted some factors which are of interest. However, carrying an analysis on 30 to 40 years using time series data would lead to a more global overview and more generalise insights. Research should focus on see how the music taste changed over time and how did music recording and publishing vary.

Music genres differ completely between each other. There are huge differences in tempo, energy, and instruments used to compose songs. Rather than just trying to generalize the idea of popularity with songs, researchers should focus on trying to gather insights per music genre. By accounting for music genres, better decisions can be made in order to target the right group of people. It becomes then possible to try to infer on the context based on where people do listen to specific music or not, and how the genre becomes relevant and how do songs become successful in that regard.

7. References

Ali, J., Khan, R., Ahmad, N., and Maqsood, I. (2012). Random forests and decision trees. *International Journal of Computer Science Issues (IJCSI)*, *9*(5), 272.

Aljanaki, A., Yang, Y. H., and Soleymani, M. (2017). Developing a benchmark for emotional analysis of music. *PloS one*, *12*(3), e0173392.

Bakhshi, H., McVittie, E., and Simmie, J. (2008). *Creating Innovation: Do the creative industries support innovation in the wider economy?* (pp. 1-40). London: Nesta.

Berns, G. S., and Moore, S. E. (2012). A neural predictor of cultural popularity. *Journal of Consumer Psychology*, *22*(1), 154-160.

Billboard (2023): “Billboard Top Chart 200”. Retrieved 9th Sept., 2023, from: https://www.billboard.com/charts/billboard-200/

Bilton, C., and Leary, R. (2002). What can managers do for creativity? Brokering creativity in the creative industries. *International journal of cultural policy*, *8*(1), 49-64.

Blood, A. J., and Zatore, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. Proceedings of the National Academy of Science, 98(20), 11818–11823.

Borg, Nicholas, and George Hokkanen. "What makes for a hit pop song? What makes for a pop song." *Unpublished Thesis, Stanford University, California, USA* (2011).

Boughanmi, K., and Ansari, A. (2021). Dynamics of musical success: a machine learning approach for multimedia data Fusion. *Journal of Marketing Research*, *58*(6), 1034-1057.

Caves, R. E. (2000). *Creative industries: Contracts between art and commerce* (No. 20). Harvard University Press.

Deloitte (2021): “The future of Creative Economy”. Retrieved 9th Sept., 2023, from https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/technology-media-telecommunications/deloitte-uk-future-creative-economy-report-final.pdf

Detienne, K. B., Detienne, D. H., and Joshi, S. A. (2003). Neural networks as statistical tools for business researchers. *Organizational Research Methods*, *6*(2), 236-265.

Devine, K. (2013). Imperfect sound forever: loudness wars, listening formations and the history of sound reproduction. *Popular Music*, *32*(2), 159-176.

Echambadi, R., and Hess, J. D. (2007). Mean-centering does not alleviate collinearity problems in moderated multiple regression models. *Marketing Science*, *26*(3), 438-445.

Eliashberg, J., and Sawhney, M. S. (1994). Modeling goes to Hollywood: Predicting individual differences in movie enjoyment. *Management science*, *40*(9), 1151-1173.

Flew, T. (2002). Beyond ad hocery: defining creative industries. In *Cultural Sites, Cultural Theory, Cultural Policy, The Second International Conference on Cultural Policy Research*.

Gerstin, J. (1998). Reputation in a musical scene: The everyday context of connections between music, identity and politics. *Ethnomusicology*, *42*(3), 385-414.

Herremans, D., Martens, D., and Sörensen, K. (2014). Dance hit song prediction. *Journal of New Music Research*, *43*(3), 291-302.

IFPI (2023). "Industry data". Retrieved 9th Sept., 2023, from ***https://www.ifpi.org/our-industry/industry-data/***

Iwanaga, M., Ikeda, M., and Iwaki, T. (1996). The effects of repetitive exposure to music on subjective and physiological responses. *Journal of Music Therapy*, *33*(3), 219-230.

Jones, C., Svejenova, S., Pedersen, J. S., and Townley, B. (2016). Misfits, mavericks and mainstreams: Drivers of innovation in the creative industries. *Organization Studies*, *37*(6), 751-768

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... and Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, *30*.

Kowald, D., Schedl, M., and Lex, E. (2020). The unfairness of popularity bias in music recommendation: A reproducibility study. In *Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part II 42* (pp. 35-42). Springer International Publishing.

Lee, J., and Lee, J. S. (2018). Music popularity: Metrics, characteristics, and audio-based prediction. *IEEE Transactions on Multimedia*, *20*(11), 3173-3182.

Levitin, D. J. (2006). This is your brain on music: The science of a human obsession. New York: Dutton

Lopes, P. D. (1992). Innovation and diversity in the popular music industry, 1969 to 1990. *American sociological review*, 56-71.

Martens, D., and Provost, F. (2014). Explaining data-driven document classifications. *MIS quarterly*, *38*(1), 73-100.

Musical Mum (2023): "Instrumental Music vs Vocal Music". Retrieved 9th Sept., 2023, from : https://www.musicalmum.com/instrumental-music-vs-vocal-music/#:~:text=Vocal%20music%20is%20generally%20more,psychologically%20to%20the%20human%20voice

Musician Wave (2023). "How long should a song be". Retrieved 9th Sept., 2023, from https://www.musicianwave.com/how-long-should-a-song-be/#:~:text=The*%20average%20length%20of%20a%20song%20can%20change%20over%20time,minutes%20and%2030%20seconds%20long.*

Namica (2023). "Music that makes us feel good". Retrieved 9th Sept., 2023, from *https://namica.org/blog/music-that-makes-us-feel-good-playlist/#:~:text=Our%20favorite%20melodies%20release%20dopamine,energize%20us%20for%20the%20day.*

O'Connor, J. (2009). Creative industries: a new direction?. *International journal of cultural policy*, *15*(4), 387-402.

Pachali, M. J., and Datta, H. (2022). What drives demand for playlists on Spotify? *Available at SSRN.*

Pachet, F., and Sony, C. S. L. (2012). Hit song science. *Music data mining*, 305-326.

Peterson, R. A., and Simkus, A. (1992). How musical tastes mark occupational status groups. *Cultivating differences: Symbolic boundaries and the making of inequality*, *152.*

Potts, J. (2009). Why creative industries matter to economic evolution. *Economics of innovation and new technology*, *18*(7), 663-673.

Rentfrow, P. J., Goldberg, L. R., and Levitin, D. J. (2011). The structure of musical preferences: A five-factor model. Journal of Personality and Social Psychology, 100(6), 1139–1157

Shankar, A. (2000). Lost in music? Subjective personal introspection and popular music consumption. *Qualitative Market Research: An International Journal*, *3*(1), 27-37.

Sloboda, J. A. (1999). Everyday uses of music listening: A preliminary study. *Music, mind and science*, *354*, 369.

Spotify (2023). "Spotify for Developers". Retrieved 9th Sept., 2023, from https://developer.spotify.com/discover/

Statista (2023): "Spotify global MAU". Retrieved 9th Sept., 2023, from https://www.statista.com/statistics/367739/spotify-global-mau/

Swarbrick, D., Bosnyak, D., Livingstone, S. R., Bansal, J., Marsh-Rollo, S., Woolhouse, M. H., and Trainor, L. J. (2019). How live music moves us: head movement differences in audiences to live versus recorded music. *Frontiers in psychology*, *9*, 2682.

Thompson, W. F. (2015). *Music, thought, and feeling: Understanding the psychology of music*. Oxford university press.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267-288.

UCLA (2020): "Spotify Trend Analysis". Retrieved 9th Sept., 2023, from https://ucladatares.medium.com/spotify-trends-analysis-129c8a31cf04

Vickers, E. (2011). The loudness war: Do louder, hypercompressed recordings sell better?. *Journal of the audio engineering society*, *59*(5), 346-351.

Weihs, C., Ligges, U., Mörchen, F., and Müllensiefen, D. (2007). Classification in music research. *Advances in Data Analysis and Classification*, *1*, 255-291.

Yang, L. C., Chou, S. Y., Liu, J. Y., Yang, Y. H., and Chen, Y. A. (2017, March). Revisiting the problem of audio-based hit song prediction using convolutional neural networks. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 621-625). IEEE.