

From Compact Discs to Streaming: A Comparison of Eras within the Brazilian Market

Danilo B. Seufitelli, Gabriel P. Oliveira, Mariana O. Silva, Gabriel R. G. Barbosa, Bruna C. Melo, Juliana E. Botelho, Luiza de Melo-Gomes, Mirella M. Moro

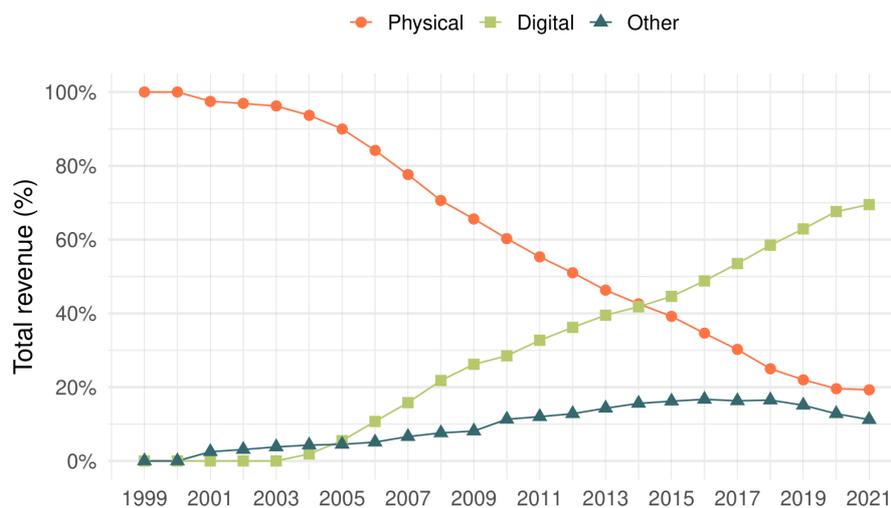
Universidade Federal de Minas Gerais | Brazil

Abstract: The music industry has undergone many changes in the last few decades, notably since vinyl, cassettes and compact discs faded away as streaming platforms took the world by storm. This Digital evolution has made huge volumes of data about music consumption available. Based on such data, we perform cross-era comparisons between Physical and Digital media within the music market in Brazil. First, we build artists' success time series to detect and characterize hot streak periods, defined as high-impact bursts that occur in sequence, in both eras. Then, we identify groups of artists with distinct success levels by applying a cluster analysis based on hot streaks' features. We find the same clusters for both Physical and Digital eras: Spike Hit Artists, Big Hit Artists, and Top Hit Artists. Our results reveal significant changes in the music industry dynamics over the years by identifying the core of each era.

Keywords: Brazilian music market, musical success, music information retrieval, time series analysis, hot streaks.

Music has experienced huge transformations in recent decades, shifting from ownership to access, from compact discs to streams, and mostly from Physical to Digital. Physical media is indeed constantly making room for the consolidation of the Digital era. Figure 1 shows this process, with a turning point in the early 2010s. In such a market, Physical media sales are still going on; whereas streaming services dominate music consumption, accounting for over 65% of the music industry revenue in 2021.¹ The scenario in Brazil is similar: as Latin America's largest music market, about 85.6% of music revenue comes from Digital media, against 0.6% from Physical, according to the most recent *Pró-Música* report.²

FIGURE 1 – Global recorded music revenues by segment (1999–2021).



Source: IFPI GMR (2022, p. 11)

Digital media dominance has proved beneficial in several ways, such as promoting local artists and increasing listeners' engagement. Also, while musicians struggled with the COVID-19 pandemic in 2020-2021 (SANDRONI et al., 2021), people have been reminded of the timeless healing power of music, positively impacting listeners' well-being and strengthening the connection between artists and fans. Indeed, with live music on hold and the world in lockdown, most music fans worldwide prefer to consume music through Digital ways (streaming and short video apps).³

¹ IFPI Global Music Report: <https://gmr2022.ifpi.org/>

² Pró-Música Brasil Report: <https://bit.ly/ProMusica2021>

³ IFPI Engaging with Music report, 2021: <https://bit.ly/IFPI-engaging-2021>

Still, streaming popularization has brought new challenges due to the massive volume of music-related data to process and analyze.

Finding and promoting artists with promising careers is an example of a task that has become more complex and important. In the Physical era, having a major label was essential to breakthrough; but now, artists from small or independent labels can go viral and become popular thanks to streaming services, showing just how inherently dynamic the music industry is. In such a context, combining Artificial Intelligence (AI) tools and techniques may be the key to facing the existing challenges of this task, creating significant benefits for both the artists and the A&R (Artists & Repertoire) executives. In fact, many applications use AI-powered technology in the music industry, such as genre classification (SHINOHARA; FOLEISS; TAVARES, 2019) and success prediction (ARAUJO; CRISTO; GIUSTI, 2019; MARTÍN-GUTIÉRREZ et al., 2020). Regarding the latter, identifying upcoming artists with outstanding success is crucial, as it helps planning and adjusting marketing directions for their careers.

Generally, musical careers present continuous periods of success above average, defined as hot streaks (HS). The concept has been investigated in many domains, including science (SINATRA et al., 2016), social media (GARIMELLA; WEST, 2019), and creative careers (LIU et al., 2018; JANOSOV; BATTISTON; SINATRA, 2020; LIU et al., 2021). In such a context, we explore over three decades in the Brazilian music market, assessing the evolution of successful careers by comparing data from the Physical (1990–2015) and Digital (2016–2020) eras. In particular, we build artists' success time series based on sales (Physical era) and streams (Digital era). Based on such time series, we investigate whether the most successful periods in an artist's career occur chronologically close and detect hot streak periods. Then, we perform a cluster analysis to group artists according to their success level. Finally, we characterize such hot streaks to extract insights into the temporal evolution of musical careers.

This article extends a paper from the 18th Brazilian Symposium on Computer Music (BARBOSA et al., 2021). As new material, we perform a clustered success analysis to investigate if the success is grouped in time for most artists. In addition, we deepen the previous analyses on hot streak periods by analyzing other dimensions such as artist nationality and the position of the first hot streak. The remainder of this paper is organized as follows. First, we discuss related work in

Section 1. Then, we describe the data acquisition process in Section 2. We present a clustered success analysis in Section 3. We detail the methodology used to identify hot streaks in Section 4. We overview the identified clusters and previous analyses in Section 5. Next, we further enrich the cross-era comparison by including the evolution of genre preference, artist nationality preference in the Brazilian Market, and the first hot streak analysis in Section 6. Finally, we conclude with future directions in Section 7.

1. Related Work

Although streaming platforms are inherently designed to not interfere with the music production process, their leading role in the music industry is unquestionable: they determine the amount paid to music content producers, and dictate the type of music accessible through their recommendation algorithms (DE MARCHI; LADEIRA, 2018). Such constant changes in the music market reinforce the investigation demand on the implications of these new media players' insertion. Specifically, the Physical to Digital era shift requires attention mainly for Music Information Retrieval (MIR), not only on technological-driven factors (KISCHINHEVSKY; VICENTE; DE MARCHI, 2015; WALDFOGEL, 2017; DE MARCHI; LADEIRA, 2018) but also on the success patterns shaping this dynamic market (SILVA; ROCHA; MORO, 2019; MARTÍN-GUTIÉRREZ et al., 2020; OLIVEIRA et al., 2020).

After decades of intense transformations in the music market, the Digital era brought novel challenges, including a substantial volume of data. As human inspection is almost impossible for music big data scale, specialized algorithms can help with several tasks in MIR, including music recommendation (BORGES; QUEIROZ, 2017), automatic genre classification (CORRÊA; RODRIGUEZ, 2016; SHINOHARA; FOLEISS; TAVARES, 2019; ARAÚJO LIMA et al., 2020), algorithmic composition (HOLOPAINEN, 2021) and so on. Another possible benefit is to feed machine-learning models for musical success early prediction, contributing to identify trends and new talent. Indeed, evaluating the impact of human performance is a common practice in many research fields (HENDRICKS; PATEL; ZECKHAUSER, 1993; RABIN; VAYANOS, 2010; RAAB, GULA; GIGERENZER, 2012). The term *Hot Streak* emerges in such context, as the

reference to a specific period within professional careers when the success is significantly higher than the average (LIU et al., 2018).

For individual and creative careers, research assessing impact is much more recent. Liu et al. (2018) consider large-scale careers of artists, film directors and scientists to demonstrate that hot streaks are remarkably universal across diverse domains, yet usually unique across different careers. In this sense, Garimella and West (2019) use data from Twitter, one of the most popular online social networks, and define users' impact as the reach of their content. Janosov et al. (2020) also consider luck as a crucial ingredient to achieve impact in creative domains. Regarding music, they model the historical artist timelines based on the release year of songs and measure success by the total play counts obtained from Last.fm.

Nonetheless, to the best of our knowledge, no previous studies address the dynamics of music artists' success periods (i.e., hot streaks) within the Brazilian market. Also, although Brazil's high rates of music consumption, little is known about the key factors driving musical success and defining artists' promising careers. As regional markets have their own success patterns and behavior (OLIVEIRA et al., 2020; DE MELO; MACHADO; DE CARVALHO, 2020), such individual analyses are crucial. Therefore, this work is a step forward towards understanding the specific dynamics of music artist success within the Brazilian market.

2. Data Acquisition

To perform a cross-era comparative analysis between Physical and Digital media, we focus on musical success in Brazil. Our first data source is Spotify, the most popular global audio streaming service. However, its Charts only comprise data from 2016 onwards. Hence, to describe the Digital Era, we consider the range period available (2016–2020). Therefore, we also use the Pró-Música Brasil platform to describe the Physical Era, with data from 1990 to 2015. Next, we detail the data acquisition processes for both Physical (Section 2.1) and Digital media (Section 2.2).

2.1. Physical Media

Pró-Música Brasil (PMB) is the official representative body of the record labels in the Brazilian phonographic market. It represents artists in legal and financial instances and issues certification awards, as authorized by record companies. The certification awards recognize the work of performers according to sale numbers in the form of “special discs”, i.e., Gold, Platinum and Diamond discs. The data on such awards is available on its website⁴ and was collected on February 5th, 2021. The final dataset comprises information on awarded artists, release year, disc category, song/album name and media type since 1990.

TABLE 1 – Pro-Música Brasil certification levels for Brazilian and foreign artists (A&S is Albums and Singles)

<i>Certification</i>	Brazilian		Foreign	
	<i>A&S</i>	<i>DVDs</i>	<i>A&S</i>	<i>DVDs</i>
Gold	40,000	25,000	20,000	15,000
Platinum	80,000	50,000	40,000	30,000
Double Platinum	160,000	100,000	80,000	60,000
Triple Platinum	240,000	150,000	120,000	90,000
Diamond	300,000	250,000	160,000	125,000
Double Diamond	600,000	500,000	320,000	250,000
Triple Diamond	900,000	750,000	480,000	375,000
Quadruple Diamond	-	1,000,000	-	500,000
Quintuple Diamond	-	1,250,000	-	625,000

In PMB, the threshold sales number for each certificate depends on whether the artist is Brazilian or not, as shown in Table 1. However, as such information is not available in PMB, we crawled it from Wikipedia using a Python library.⁵ Next, we collect the total sales for each musical work based on the certification awarded, nationality, and PMB's sales metric for the disc award. Finally, we use Spotify's API⁶ to associate each artist with their respective genres specified on the streaming platform. Hence, our final dataset contains information about 4,198 musical works from 780 artists. Considering only the period between 1990 and 2015 (i.e., Physical Era), there are 3,243 musical works from 574 artists. Quantitative information on the certificates is shown in Figure 2.

⁴ PMB Certificates: <https://bit.ly/CertificatesPMB>

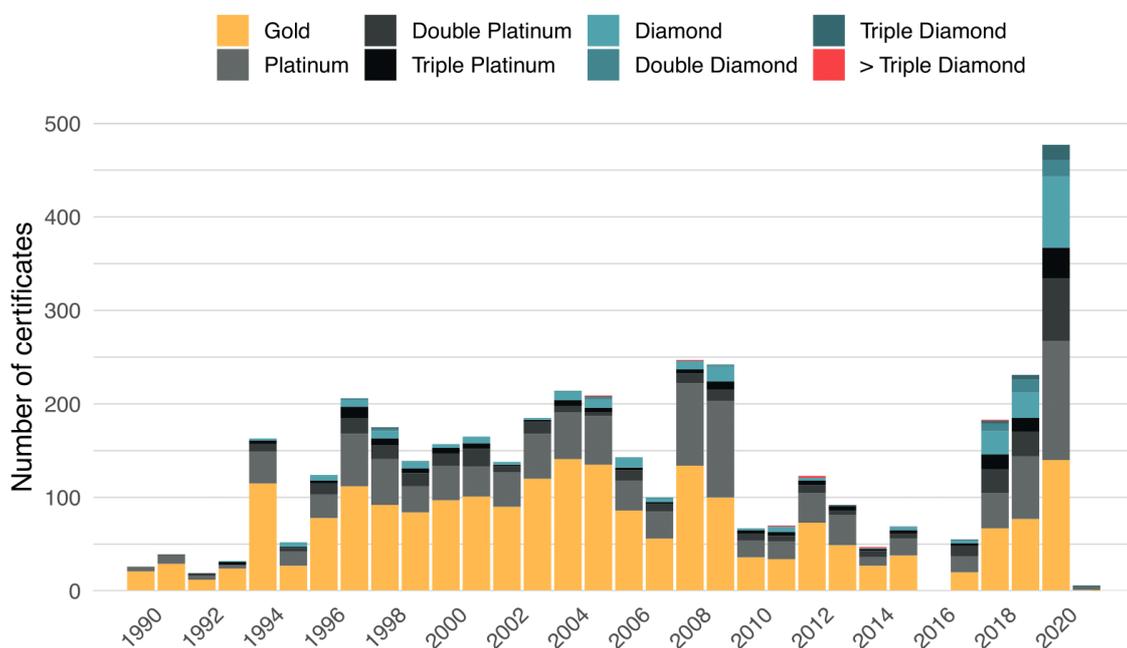
⁵ Wikipedia Python Library: <https://github.com/goldsmith/Wikipedia>

⁶ Spotify API: <https://developer.spotify.com/>

2.2. Digital Media

Between 2016 and 2017, there was a crucial change in PMB's metric, which moved from Physical media (i.e., DVD and CD) to Digital media (i.e., Singles and Albums), as depicted in Figure 3. Meanwhile, streaming was already the primary revenue source for Digital media (58.3%).⁷

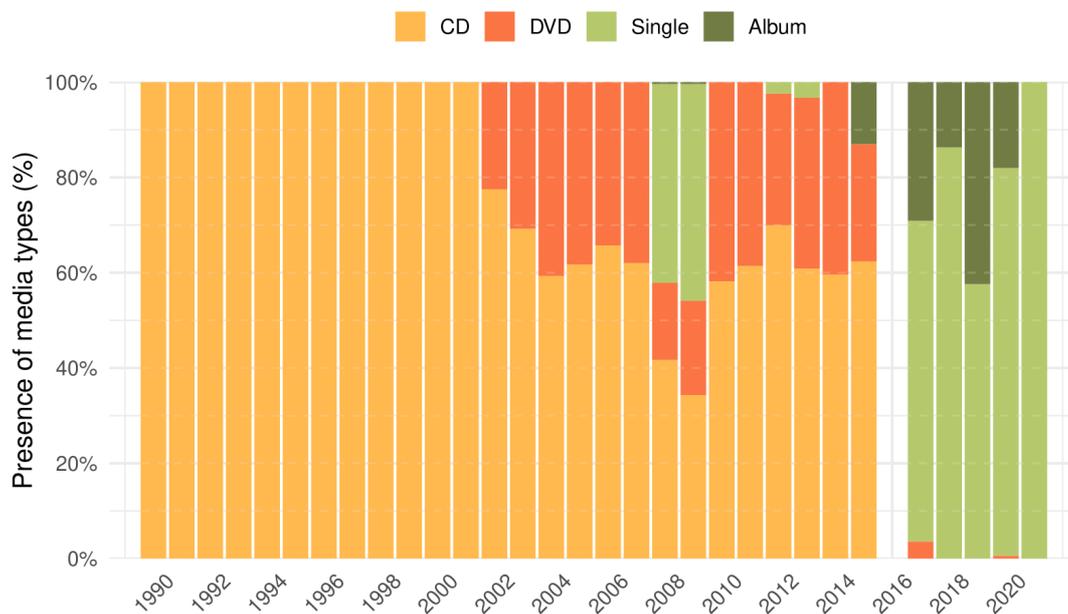
FIGURE 2 – Discs certificated in Pró-Música Brasil (1990–2021). In 2016, there was a metric change in the certification, hence the lack of data.



Given its relevance, we extract data referring to the Digital Era from the weekly Spotify Top 200 Chart, which corresponds to the most streamed songs in Brazil. Each chart entry contains the song's name and its artist(s), the number of streams, the song's Spotify URL and its position on the chart. We collect data from January 2017 to December 2020. We also collect artist data using the Spotify API: name, number of followers, and their list of genres. Our final dataset for Digital Era comprises 2,595 songs from 1,018 artists obtained from 108 weekly charts.

⁷ Pró-Música Brasil 2016: <https://bit.ly/ProMusica2016>

FIGURE 3 – Media type in Pró-Música Brasil (1990–2021). In 2016, there was a metric change in the certification, hence the lack of data.



3. Clustered Success

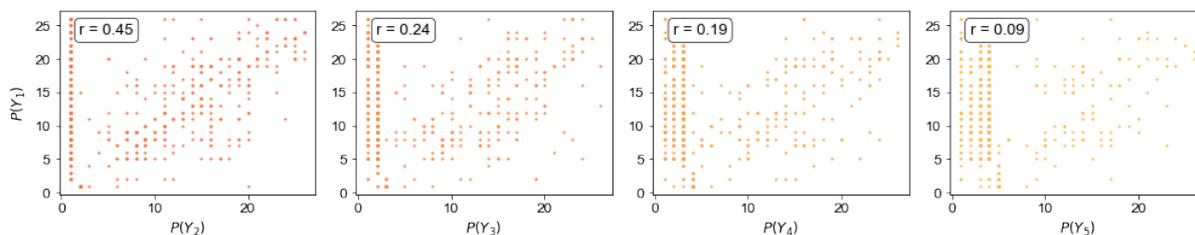
In this section, we present an analysis of the distribution of success over time. We follow the methodology used by Garimella and West (2019) to investigate whether the most successful periods of each era (years or weeks) occur close to each other in artists' careers. We define artists' career as their success time series from their debut on charts to the last date of our collection (2015 for Physical Era, and December 2020 for Digital Era). To the Physical Era, we set the position $P(y_i)$ of a year y_i within a time series as its index i . The k most successful years (i.e., with the highest sales values) are denoted by Y_1, Y_2, \dots, Y_k . Likewise, in Digital Era, we set the position $P(w_i)$ of a week w_i within a time series as its index i . Here, the k most successful weeks (i.e., with the highest stream values) are denoted by W_1, W_2, \dots, W_k .

Our analyses focus on two main points. First, we investigate the timing of the most successful periods of an artist's career for both eras. Then, we look at the distribution of the difference between the positions of the two most successful periods within artists' careers. Such analyses are all made compared to shuffled careers to check the robustness of our findings, that is, if the observed effects still happen.

3. 1. Timing of the most impactful periods

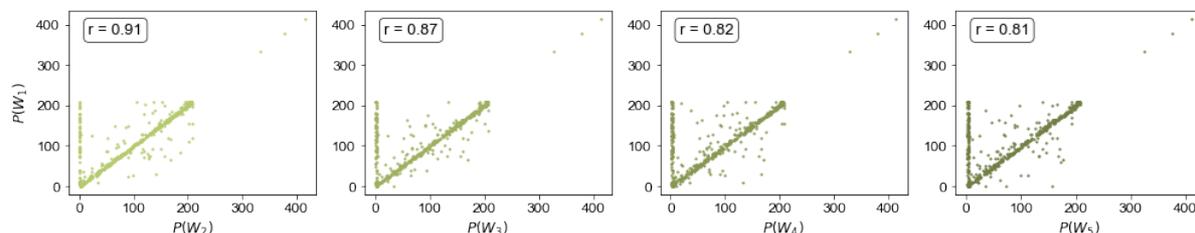
First, we analyze the positions of the five most successful years (Physical Era) and the five most successful weeks (Digital Era) within artists' careers. Figure 4 presents scatter plots of the first year of Physical Era $P(Y_1)$ versus the other considered years $P(Y_i) \forall i \in [2, 5]$, as well as the Pearson correlation coefficient (r) for each plot – correlation values are statistically significant ($p < 0.05$). We consider all artists from our dataset. The results show linear correlation for all artists' careers and higher Pearson coefficient when comparing the first and second most popular years. Compared with the third, fourth and fifth years, the correlation decreases its values. Such a finding reinforces the hypothesis that the most impactful years within an artist's career are more likely to happen close to each other.

FIGURE 4 – Scatter plots with Pearson correlation (r) of the position of the most successful year (Physical Era) in artist careers (Y_1) with Y_2 , Y_3 , Y_4 and Y_5 , respectively. Each point represents an artist.



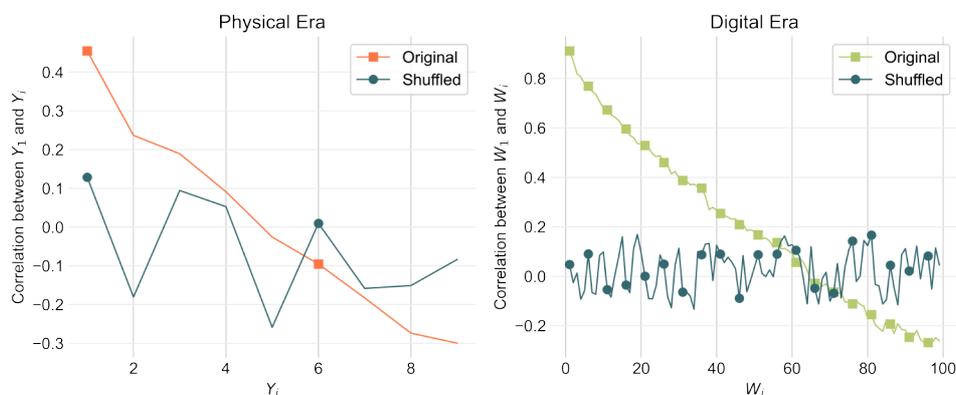
Regarding the Digital Era, Figure 5 presents scatter plots of the first week $P(W_1)$ versus the other considered weeks $P(W_i) \ i \in [2,5]$, as well as the Pearson correlation coefficient (r) for each plot – correlation values are statistically significant ($p < 0.05$). Here, we also consider the totality of artists from our dataset. In contrast to the Physical Era, the results show clearer and stronger linear correlation for all artists' careers. Likewise, there is a higher Pearson coefficient concerning the first and second most popular weeks. However, although the third, fourth and fifth weeks decrease their correlation values, their value remains high. These results are different from those found for the Physical Era, and the greater granularity of the Physical data may directly influence them. Nevertheless, it is still possible to verify that the first two weeks are more correlated than others, in both cases, reinforcing the premise that artists' most successful periods group in time.

FIGURE 5 – Scatter plots with Pearson correlation (r) of the position of the most successful week (Digital Era) in artist careers (W_1) with W_2 , W_3 , W_4 and W_5 , respectively. Each point represents an artist.



We then expand the investigation on correlation values to compare the positions of successive periods. Figure 6 shows a decrease in the correlation in all artists' careers for both eras. Still, this pattern is not observed in shuffled careers, in which the correlation is always between -0.1 and 0.1 for the Physical Era, and -0.2 and 0.2 for Digital Era. Therefore, there is a general trend of clustering within the most successful periods (years or weeks) in artist careers, as they tend to happen close to each other in the success time series.

FIGURE 6 – Correlation between the first and i -th most successful years to the Physical Era (left), and the i -th most successful weeks to the Digital era (right).

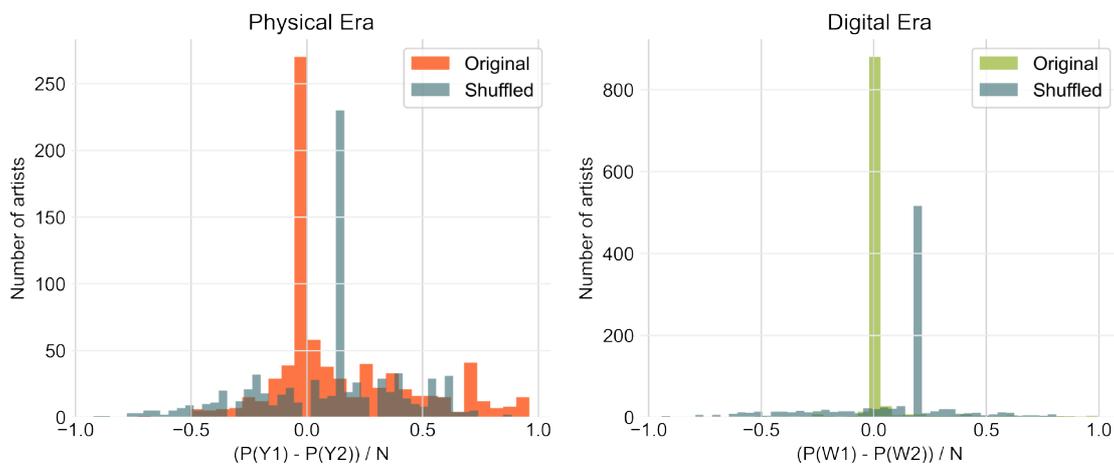


3.2. Difference in positions of the most successful periods

In this section, we complement the previous analysis by taking each artist's first and second most successful periods to calculate their differences in positions and verify if they happen near each other. For instance, if each artist's most successful periods are close together (years 2 and 3 for one group of artists, and years 5 and 6 for another), subtracting their positions results in 1; the final result is close to zero when normalizing such values; i.e., if such position differences have the value of 1 for most artists, their success occurs in a consecutive period, or they are grouped in time.

To calculate the difference in the positions of the top two most successful periods for the artist's careers, we consider $P(Y_1)$ and $P(Y_2)$ for the Physical Era and $P(W_1)$ and $P(W_2)$ for Digital Era. We normalize such a difference by the number N of years/weeks of the artist's time series in the corresponding era. Figure 7 shows the distribution has a peak around zero for all artists' careers, suggesting that these two periods (years/weeks) are close to each other on the timeline.

FIGURE 7 – The normalized difference between the positions of the first and second most successful periods within artists' careers. Years for Physical Era (right), and weeks for Digital Era (right).



Note that the results are similar to both eras. Such outcome agrees with the findings of the previous analysis. Further, when we shuffle artists' careers, the distribution of these differences is much distinct from the original, demonstrating that this behavior of musical careers is not random. Hence, there is strong evidence that artists may experience periods of outstanding success, or *hot streaks*, which we investigate in the next section.

4. Hot Streak Detection

To detect hot streaks in the artist's time series, we rely on previous work that shows the most successful points in professional careers tend to happen close to each other (GARIMELLA; WEST, 2019). Hence, we use a technique to reduce the time series dimensionality to continuous delimited periods within careers. Then, we define a hot streak as the periods in which the success (i.e., Physical sales or Digital streams) is above a certain threshold obtained from the career itself. In

other words, the hot streak detection does not consider external factors (e.g., genre and time) because artists reach different levels of success, and choosing a single threshold would make the comparison unfair.

To reduce time series dimensionality, we use Piecewise Aggregate Approximation – PAA (KEOGH; PAZZANI, 2000). Given a time series $A = a_1, a_2, \dots, a_n$ of length n , PAA reduces it into a new series $\bar{A} = \bar{a}_1, \bar{a}_2, \dots, \bar{a}_n$ with N dimensions, $1 \leq N \leq n$. The intuition behind this method is that dividing the original time series into N equal-sized segments produce N new points. The value of each segment is defined as the average of the points within such a frame (Equation 1). Hence, the approximation of each point on the original time series is made by simply assigning the PAA value of its corresponding segment.

$$\bar{a}_i = \frac{n}{N} \sum_{j=\frac{n}{N}(i-1)+1}^{\frac{n}{N}i} a_j \quad (1)$$

Note that artists' careers may contain points with extreme values for the success metric. Therefore, PAA is a helpful tool to smooth such differences and delimit periods in the careers. Regarding code, we use the PAA implementation of *tslearn* (TAVENARD et al., 2020), a Python package for time series analysis. Its only parameter is the number of segments to split the series into (further information on values next).

Finally, we chose a specific threshold for defining the hot streak periods for each artist. Such an individualized approach is based on the percentiles of the success metric, and it allows analyzing the careers of artists with different levels of success. In other words, as success is relative for each artist, we detect HS for widely known artists with higher sales and streams, as well as independent artists who have received only a few certificates and streams.

5. Existing Clusters and Previous Analyses

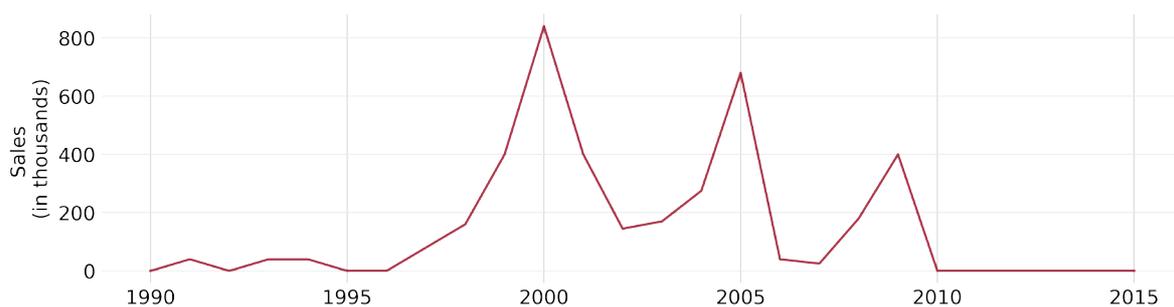
Here, we review the previous analysis for both Physical and Digital Eras (BARBOSA et al, 2021). In Section 5.1, we use the PMB data and Spotify's Brazil Top 200 Charts for building the time series for each artist, respectively, to such eras. Next, we characterize the hot streaks for both

eras and understand their relationship to music genres in Section 5.2. Finally, we perform a cluster analysis to group similarly artists based on their success levels in Section 5.3.

5.1. Artists' Time Series

In the Physical Era, the evolution of an artist's success is represented by the certificates received at the PMB from 1990 to 2015. We build the time series with an annual granularity because PMB provides data for the year the artist got the certificate. Hence, each point in the artist's time series corresponds to the number of sales achieved in that year. On the other hand, we use Brazil's Spotify Top 200 Chart in the Digital Era as our basis to model artists' success over time. For all artists, each point in their time series represents the accumulated success in a given week, according to the chart. In this case, the success measure considered is the total number of streams (i.e., the number of times the song was listened to on Spotify) per week. For example, Figure 8 presents the time series of *Sandy & Junior*. Their last album as a band was *Acústico MTV: Sandy & Junior*, released in 2007. Nevertheless, they continued to sell many records even after two years of the duo's disbandment.

FIGURE 8 – Sandy & Junior's success time series in the Physical Era (1990–2015).

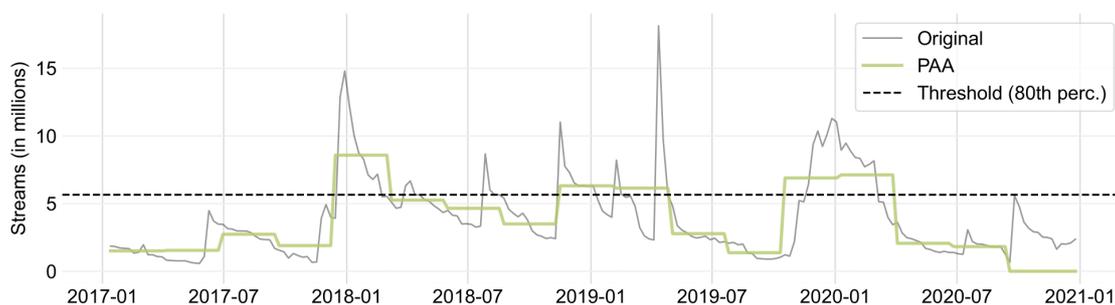


From the artists' time series, we detect the hot streaks periods by first applying PAA (see Section 4). To do so, we set the number of segments in which the series will be split, as this is the only parameter of the method. After extensive experiments to reach meaningful values, we set the size of each PAA segment equal to two years for the Physical Era and 12 weeks for the Digital Era. Hence, we calculate the number of segments by dividing the time series length by the predefined

size. In addition, we set the 80th percentile of the success metric in artists' time series as the threshold for defining the hot streak periods.

Figure 9 illustrates PAA applied to *Anitta's* career, currently one of the most influential Brazilian artists worldwide. On March 24, 2022, *Anitta* broke the record as the first Brazilian female artist to achieve the first position on Spotify Top 50 Global Chart with her single *Envolver*. Being present in all considered weekly charts, there are three HS in her time series. The first HS is from November 2017 to April 2018, the period when she released *Vai Malandra*, which became the most streamed song on its release date. The second HS period coincides with *Veneno* and *Não Perco Meu Tempo* single releases. Finally, the third one comprises the period in which she collaborated with famous national and international artists, such as Marília Mendonça (*Some Que Ele Vem Atrás*) and Black Eyed Peas (*Explosion*).

FIGURE 9 – Piecewise Aggregate Approximation (PAA) applied to Anitta's success time series in the Digital Era (2017–2020). Periods above the threshold are considered hot streaks.



5.2. Hot Streak Characterization

We characterize the hot streak periods identified for artists according to their musical genres. As individually considering closely related music styles may create artist overlapping and bias within the results, we define *super-genres* for this analysis. For example, we verify that *Indie Folk* is more frequently associated with *Rock* than any other super-genre and is then incorporated into *Rock*. Finally, the top eight prominent super-genres considered in Brazilian music are *Rock*, *Axé*, *Pop*, *Sertanejo*, *Gospel*, *Rap*, *Funk Carioca* and *Forró*.

TABLE 2 – Main statistics on Hot Streaks grouped by Genres.

Genre	Number of Hot Streaks		Proportion of Artists (%)		Median of Hot Streaks		Max of Hot Streaks	
	Physical	Digital	Physical	Digital	Physical (year)	Digital (weeks)	Physical (year)	Digital (weeks)
Rock	164	26	23.90	4.97	2	18	8	48
Axé	162	31	28.80	5.96	2	24	8	48
Pop	131	219	24.60	41.70	2	24	8	60
Sertanejo	64	95	11.20	18.90	2	24	10	48
Gospel	57	4	9.83	1.00	2	12	8	36
Rap	4	20	0.70	3.98	4	24	4	36
Funk Carioca	3	104	0.70	22.60	2	24	4	48
Forró	2	5	0.23	1.00	2	18	2	24

Table 2 assesses each genre's performance regarding the number of hot streaks as well as the proportion of artists who have achieved Hot Streaks by genre, and the values of median and maximum duration of a hot streak, respectively. All genres analyzed in the Physical Era have at least two hot streak periods. Indeed, the genres of artists that concentrate higher proportions are *Axé* (~29%), *Pop* (~25%) and *Rock* (~24%). Although this pattern happens in such genres, there is no clear correlation between genre and the number of HS. As we consider a yearly time window, achieving a high number of hot streaks is not an easy task. Regarding the number of hot streaks in the Digital Era, the genres have at least four Hot Streaks. Although all genres follow a similar trend in general, *Pop* presents a higher percentage of artists that have achieved more HS, around 42%. In contrast, all *Gospel* and *Forró* artists have only a 1% of presence in HS.

Next, we analyze the duration of hot streaks. We consider the median and the longest HS for each artist, as there may be more than one (HS size is always a multiple of two for Physical Era because PAA segment is set to two years). Overall, all genres have a median number of Hot Streaks of two, except for *Rap*, with four Hot Streaks. Nevertheless, *Axé*, *Gospel*, *Pop* and *Rock* get a remarkable maximum of eight years of Hot Streak, and *Sertanejo* reaches 10 years of success.

In the Digital Era, about half of the artists from all genres have 12-month long hot streaks. However, there are specific hot streak patterns when analyzing the genres individually. For instance, most *Gospel* artists have shorter HS periods, as the median time presents a 12-week hot streak. On the other hand, genres such as *Axé*, *Funk Carioca*, *Pop*, *Rap* and *Sertanejo* artists have longer hot streaks, as they have median HS periods of up to 24 weeks, i.e., six months. Still, only *Pop* achieves a

maximum duration of Hot Streak of 60 weeks. Therefore, in contrast to the previous analysis, we note that genre is relevant to describing hot streaks' longevity.

5.3. Cluster Analysis

We now move to the cluster analysis, which helps to better understand the characteristics of different success levels of artists achieved during the Physical and Digital Eras. We apply the *K*-Means algorithm in the time series and the Elbow method to find its optimal number of clusters. The considered features for the algorithm include the total number of hot streaks, total sales and the time series threshold. As a result, the method outcome suggests three clusters. We name the resulting clusters according to the success metric (i.e., number of sales and streams): Spike Hit Artists (SHA), Big Hit Artists (BHA), and Top Hot Artists (THA). The main statistics of the clusters in both eras are presented in Table 3 and summarized as follows.

TABLE 3 – Main statistics on the artist clusters in the Physical and Digital Eras.

	All		SHA		BHA		THA	
	Physical	Digital	Physical	Digital	Physical	Digital	Physical	Digital
Number of artists	574	1,018	527	940	38	70	9	8
Average number of HS	1.3	0.6	1.3	0.5	1.8	1.6	1.9	1.6
Median sales (10 ⁴) / Streaming (10 ⁶)	8.5	7.6	8	3.9	152	2,190	507	10,857
Median threshold	0	0	0	0	72,500	2,052,131	300,000	7,145,116

Spike Hit Artists (SHA). This cluster contains most artists in Physical (527) and Digital (940) Eras. The median PAA threshold for such artists is 0, indicating that their sales and streams are, in general, much lower than the artists from other clusters. In addition, the average number of hot streaks for SHA is 1.3 and 0.5 for Physical and Digital Eras, respectively. To the former, the result suggests that PMB certificates happen sparsely for such artists because there are few hot streaks even with a low threshold. Artists in this cluster include *Banda Eva*, *Claudinho & Bochecha*, *Coldplay*, *Paramore*, and *Latino*. To the latter, we cannot say that such artists were not successful, as they are

on Spotify's Top Charts. On the other hand, they do not have many hot streaks (one HS on average). Artists in this cluster include *Billie Eilish*, *Leo Santana*, and *Naiara Azevedo*.

Big Hit Artists (BHA). This cluster represents a bridge between the most and less successful ones, with 38 artists in the Physical era and 70 artists in the Digital era. Although BHA presents a higher number of sales when compared to SHA, its artists do not present a proportional increase in the average number of hot streaks. In fact, they have, on average, 1.8 hot streaks between 1990 and 2015 in the *Compact Disc* Era and 1.6 hot streaks between 2016 and 2020 in *Streaming* Era. Also, there is a substantial increase in the threshold for artists in this cluster, reaching 72,500 sales and more than 2M streams in respective eras. *Madonna*, *Legião Urbana*, *Skank*, and *U2* are examples of Physical BHA, and *Barões da Pisadinha*, *Dua Lipa*, and *Pablo Vittar* are of Digital BHA.

Top Hit Artists (THA). This is the cluster with the most successful artists, as they have the highest median number of sales in both scenarios. The average number of hot streaks for THA is very close to the value for BHA. However, the artists in this cluster have achieved major sales success throughout the Physical Era, as they have a very high median threshold (300,000 sales). Similar, the number of Spotify streams is much higher when compared to the previous groups, which is also observed with the threshold. Therefore, Top Hit Artists in Digital Era may be considered highly successful artists, as their songs achieve a higher stream count throughout the weeks. Examples of Physical THA include *Ivete Sangalo*, *Zeca Pagodinho*, *Roberto Carlos*, and *Sandy & Junior*. All Digital THA are Brazilian, and examples include *Anitta*, *Marília Mendonça*, and *Zé Neto & Cristiano*.

6. Cross-era Comparison

Music is part of people's daily lives regardless of the era experienced, whether Physical or Digital. With musical consumption constantly rising, we may notice similarities between both eras. From the results obtained in (BARBOSA et al., 2021), we have explored Hot Streaks (HS) in musical careers within the Brazilian market. Such HS periods provide valuable information used in

cluster analysis, in which we also notice cross-era similarities. Here, we take a step further by deepening the previous analyses on hot streak periods. Specifically, we discuss Brazilian listeners' main musical genre preferences in Section 6.1. Next, we explore the presence of Brazilian artists as the preferred consumption in the local market in Section 6.2. Finally, we explore when the first Hot Streaks occurs in the artists' careers in Section 6.3.

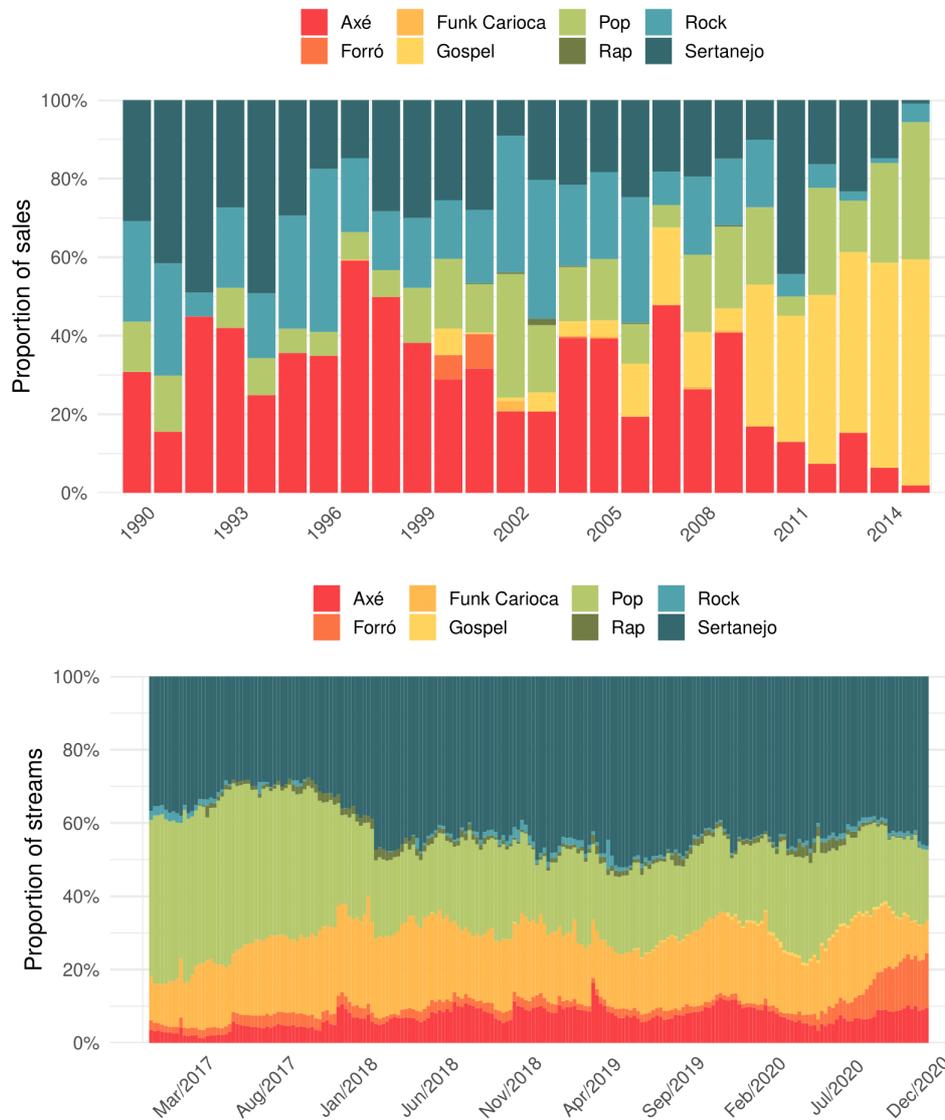
6.1. Genre Evolution

Musical genres express the cultural diversity existing in the country. Such diversity can be observed by the number of rhythms and musical styles as well as the specific characteristics that each of them retains. The most popular genres usually oscillate in the music market. Several factors may influence this issue, including the help provided by streaming platforms to spread cultural diversity worldwide in the form of musical genres. Hence, we analyze the temporal evolution of consumption of the main genres in the Brazilian market, both for the Physical and Digital eras.

Figure 10 shows the transformation in the listeners' preference regarding musical genre. For instance, there was an increase in *Gospel* sales in the Physical Era, but not in the Digital one. A potential reason is *Gospel* listeners still consume Physical media by 2015, whereas audiences from other genres had already migrated to streaming (until 2016, the PMB methodology still favored Physical sales). However, the transition of preference for musical genres over the years is notorious: in the Physical Era, the predominant rhythms were *Axé* (e.g., *Ivete Sangalo*), *Sertanejo* (e.g., *Zezé di Camargo & Luciano*) and *Rock* (e.g., *Skank*); whereas in the Digital Era, the most successful artists (THA) come from one style, *Sertanejo*, with more than 50% of streams in late 2020.

Overall, the Digital Era allows the appearance of new popular genres, as well as the decline of previously popular ones. For example, the prevalence of *Sertanejo* is remarkable over time, while *Pop* decreases from 2016 to 2020. Moreover, we highlight the rise of *Forró* in mid-2020 as a well liked genre, following the growth of popular artists who have burst the regional bubble, such as *Barões da Pisadinha*, *Solange Almeida*, and *Wesley Safadão*. Such a significant boost for regional artists may have been enhanced by the remarkable lives of *Forró* and *Sertanejo* artists during the COVID-19 pandemic, showing the music industry's ability to adapt. In fact, *Marília Mendonça* had the most-streamed YouTube live worldwide in 2020, with over 3.31 million viewers.

FIGURE 10 – Genre evolution in (top) Physical and (bottom) Digital Eras.



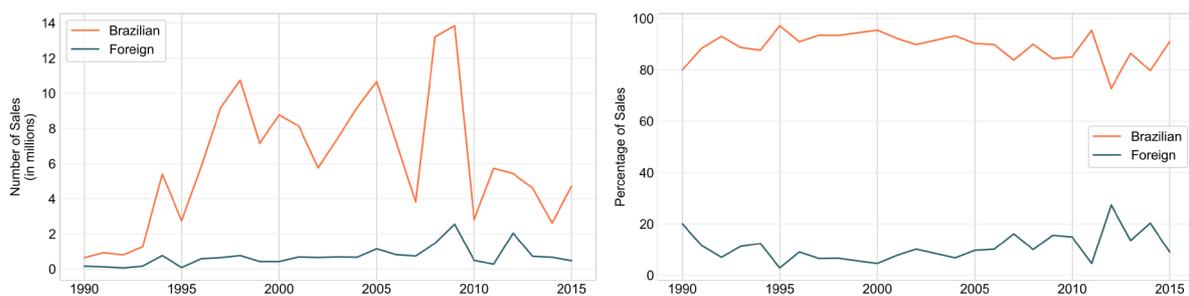
6.2. Brazilian Artists vs. Foreign Artists

Record companies have been actively working to promote local artists aiming to expand the music ecosystem. Furthermore, according to the IFPI 2022 report, fans are listening to more local artists than ever before, and their music also has the power to go global from day one. As such, the music industry needs to invest in discovering and nurturing the artists of tomorrow. There is a strong predominance of local artists' consumption in Brazil compared to other important

countries, such as the USA and European countries.⁸ Therefore, our purpose is to identify whether the Brazilian market follows this trend of consuming local artists in the Physical and Digital eras. Thus, we analyze the distribution of consumption of Brazilian and foreign artists in each period. Finally, we identify the representativeness of Brazilian and foreign artists in each cluster identified in the previous analyses, also separated by era.

In the Physical Era market, we notice a strong sales presence of artists whose nationality is Brazilian. Figure 11 shows the sales evolution in the Brazilian market, comparing sales of local artists versus foreign artists. Figure 11 shows the number of sales in millions on the left, and the percentage of total sales on the right, for both Brazilians and foreign artists. There is a constant evolution of the preference for Brazilian artists, while there is a tendency for stability in foreign artists. Indeed, the sales representativeness by Brazilian artists corresponds to around 90% throughout the entire period of the Physical era. Only around 2012, there was a slight drop in consumption of local artists, but sales still represented more than 70% in such a period. Hence, we show that Brazilians do indeed favor consuming local musical artists.

FIGURE 11 – Comparison of the sales evolution by Brazilian artists versus foreign artists in the Physical Era in absolute values (left) and percentage values (right).

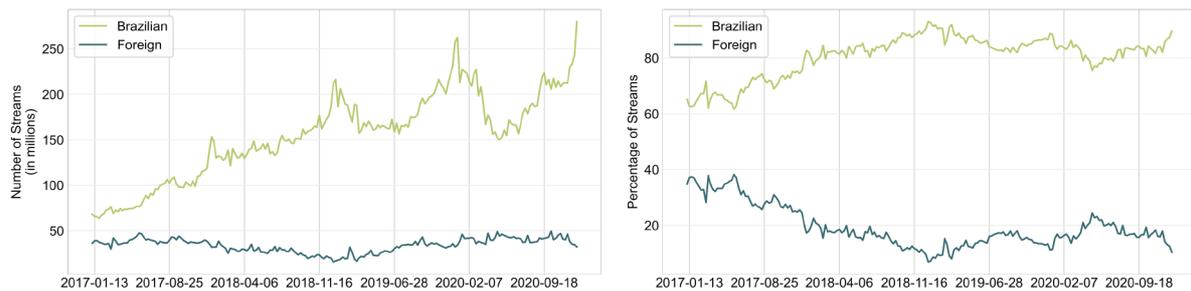


Concerning the Digital Era, Figure 12 compares the streaming evolution by Brazilian and foreign artists by the number of streams (left) and its corresponding percentage (right). The Brazilian market in Digital Era has its beginning marked by a lower preference for foreign artists when compared to the Physical Era. In such a scenario, the sales of foreign artists in the Brazilian market reached around 40% of total streaming. One possible explanation is that different artists and

⁸ *Brasileiros são os que mais ouvem a própria música entre todos os países*. Link: <https://bit.ly/folha-sp-musica>. Access: March 26, 2022.

genres spread more quickly worldwide with the advent of streaming platforms. However, there is a clear trend of growth in the consumption of Brazilian artists, reaching higher levels in 2020 and revealing a movement of a more significant decline in the consumption of foreign artists.

FIGURE 12 – Comparison of the streams evolution by Brazilian artists versus foreign artists in the Digital Era in absolute values (left) and percentage values (right).



Finally, we also analyze the distribution of the presence of Brazilian artists in each of the clusters for the Physical and Digital eras identified in Section 5. Table 4 summarizes the number of Brazilian and foreign artists by era and cluster. In general, there is a strong trend in local music consumption in the Brazilian market. We highlight the BHA and THA clusters, which comprise the most successful artists in both eras (i.e., higher sales and streams), including *Sandy & Júnior* and *Anitta*. In particular, the second cluster represents the paramount artists. Although the Physical era has one more foreign artist (five) than Brazilian (four), in the Digital era, all artists are Brazilian, indicating a strong preference for local artists and genres. As a result, the SHA cluster indicates regular success, accounting for over 90% of the artists.

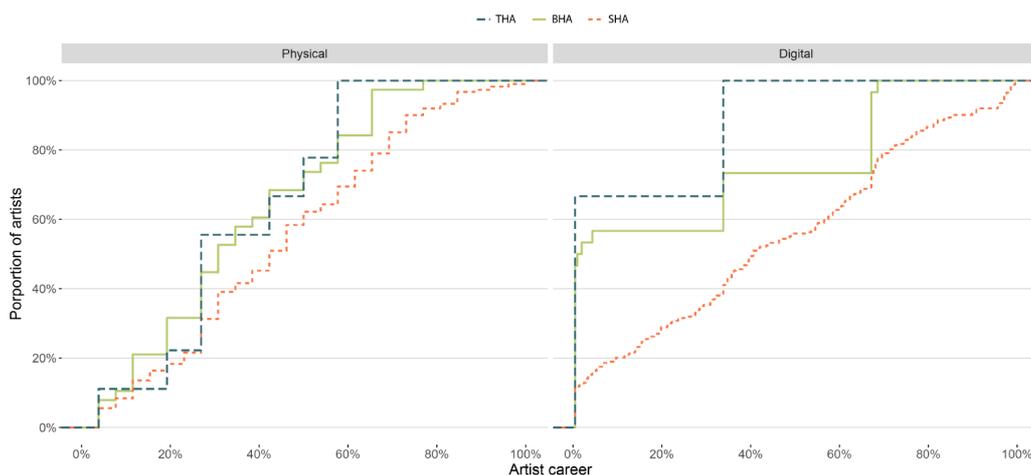
TABLE 4 – Number of Brazilian and foreign artists by cluster and era, with a majority of Brazilian artists in all groups.

	All		SHA		BHA		THA	
	Physical	Digital	Physical	Digital	Physical	Digital	Physical	Digital
Brazilian	366	587	342	522	20	57	4	8
Foreign	208	431	185	418	18	13	5	0
TOTAL	574	1,018	527	940	38	70	9	8

6.3. First Hot Streak Analysis

The last comparative aspect is the position of hot streak periods within artist careers. Previous studies on other domains show that such periods are temporally localized and happen at any point in an individual's works sequence. Thus, to assess whether there is a similar behavior in the music domain, in both Physical and Digital eras, we investigate in which point artists experience their first stardom. Here, we focus only on the first point, as we detect more than one hot streak for several artists. Figure 13 shows the cumulative distribution of the position of the first hot streak within artist careers in both Physical (left) and Digital (right) eras, grouped by cluster. In the Physical era, in general, the cumulative distribution of the position is very similar among the three identified clusters, probably due to the data granularity (i.e., considered in years and not weeks). Nonetheless, for the BHA and THA clusters, the first explosion of success still occurs faster than the artists of the SHA cluster, considering the artists' career time.

FIGURE 13 – Cumulative Distribution Function (CDF) of the position of the first hot streak within artist careers in both Physical and Digital eras, grouped by clusters. Artist timelines are described in percentages, in which 0% represent the debut week and 100% is the last year/week collected in our dataset.



On the other hand, regarding the Digital era, the two clusters that group artists with the highest success levels (i.e., BHA and THA) stand out much more in comparison to the SHA cluster. Almost 80% of the BHA/THA artists have their first burst of success early in their careers (i.e., in the first 30% of their timelines), whereas most SHA artists reach their first hot streak much later.

Therefore, our results indicate that artists who have achieved their first stardom peak earlier in their careers have a higher overall success, regardless of the musical era. It is important to note that artists may have careers of different sizes depending on their debut date, as the last date in the time series is always the same (i.e., the collection date). However, the main objective of this analysis is to offer a big-picture comparison between the Physical and Digital eras, seeking to highlight the relationship between artists' levels of success and the speed of achieving the first burst of success.

7. Conclusion

This article evaluated the success of musical artistic careers in the Brazilian market, comparing them in different eras: Physical Era, when listeners purchase Physical media to engage their favorite artists (e.g., LPs, CDs, and DVDs); and Digital Era, when music consumption took place mainly through streaming, which have democratized access to music, as streaming services do not necessarily require payment (just an internet connection). In this sense, comparing these eras becomes particularly relevant and valuable, as it enables to identify similar or divergent patterns that record companies can use to generate valuable insights during decision making. Thus, we performed a cross-era comparative analysis between the Physical and Digital media in the Brazilian music market, which is the largest market for the music industry in Latin America.

First, we found that the artists' most successful periods tend to group in time. Motivated by such results, we built artists' success time series for both eras to identify hot streak periods, defined as continuous high-impact bursts. Next, we characterized such periods to understand the dynamics of success among artists from different musical genres. Although there are similarities among all music styles, our results showed that some genres have meaningful specific patterns for both eras. Therefore, as in other studies in the MIR field, considering music genre information can be relevant for both the predictive and descriptive models. We also performed a profiling analysis that uncovered three different clusters in both eras: Spike Hit Artists (SHA), Big Hit Artists (BHA), and Top Hit Artists (THA), which acted as class descriptors of successful artists. In addition, we found that artists who have achieved their first stardom peak earlier in their careers have a higher overall success, regardless of the musical era. Finally, we discover that Brazilians prefer to consume

music by Brazilian artists, and we highlight the proportion of Brazilian and foreign artists within each identified cluster. Such a pattern repeats in both eras.

Overall, our results shed light on meaningful insights for MIR tasks, such as prediction and recommendation. For example, the identified clusters may serve as input features for musical success prediction models. In addition, they may also help in recommending potentially successful musical partnerships and collaborations. Besides helping the scientific community, this work also benefits the music industry. Analyzing the evolution of artist careers reveals success trends in Brazil from what people consume. Indeed, our results demonstrate that although Brazilians connect with international hit songs, they still have a strong preference for local artists regardless of the era. Hence, considering individually regional markets is crucial for better comprehension of specific factors driving musical success. Finally, understanding hot streak periods and success patterns can enhance the human element in the music industry (e.g., A&R executives and record label CEOs) and people's relationships with music. Our findings may help describe the listeners' behavior and musical trends, allowing the music industry to connect people to songs relevant to them.

Threats to Validity

Here, a limiting factor is that piracy had a high impact in the music consumption in Brazil, mainly in the late 2000s and early 2010s. Therefore, data collected from PMB may not precisely reflect Brazilian preference in music.⁹ In addition, although we found similar patterns between the Physical and Digital eras, each data source used considers its own success measure, which can cause biased results. Finally, we only consider artists who are recognized as successful (either through their sales or their position in stream rankings). Future work should overcome all such limitations to enhance the results and further advance in the state-of-the-art.

ACKNOWLEDGMENT

Work supported by CAPES and CNPq, Brazil.

⁹ For reference, 52% of the music consumption in Brazil in 2005 came from piracy: <https://bit.ly/PiracyReportIFPI>

REFERENCES

- ARAUJO, Carlos Soares; CRISTO, Marco; GIUSTI, Rafael. Predicting music popularity on streaming platforms. In: *Proceedings of the 17th Brazilian Symposium on Computer Music*. Porto Alegre: SBC, 2019. p. 141-148.
- ARAÚJO LIMA, Raul de et al. Brazilian lyrics-based music genre classification using a BLSTM network. In: *International Conference on Artificial Intelligence and Soft Computing*. Cham: Springer, 2020. p. 525-534.
- BARBOSA, Gabriel R. G.; MELO, Bruna C.; OLIVEIRA, Gabriel P.; SILVA, Mariana O.; SEUFITELLI, Danilo B.; MORO, Mirella M. Hot Streaks in the Brazilian Music Market: A Comparison Between Physical and Digital Eras. In: *Proceedings of the 18th Brazilian Symposium on Computer Music*. Porto Alegre: SBC, 2021, p. 155-162.
- BHOLOWALIA, Purnima; KUMAR, Arvind. EBK-means: A clustering technique based on elbow method and k-means in WSN. *International Journal of Computer Applications*, v. 105, n. 9, 2014.
- BORGES, R.; QUEIROZ, Marcelo. A probabilistic model for recommending music based on acoustic features and social data. In: *16th Brazilian Symposium on Computer Music*, 16, São Paulo, Brazil. 2017. p. 7-12.
- CORRÊA, Débora C.; RODRIGUES, Francisco Ap. A survey on symbolic data-based music genre classification. *Expert Systems with Applications*, v. 60, p. 190-210, 2016.
- DE MARCHI, Leonardo; LADEIRA, João Martins. Digitization of music and audio-visual industries in Brazil: new actors and the challenges to cultural diversity. *Les Cahiers d'Outre-Mer. Revue de géographie de Bordeaux*, v. 71, n. 277, p. 67-86, 2018.
- DE MELO, Gabriel Borges Vaz; MACHADO, Ana Flávia; DE CARVALHO, Lucas Resende. Music consumption in Brazil: an analysis of streaming reproductions. *PragMATIZES - Revista Latino-Americana de Estudos em Cultura*, v. 10, n. 19, p. 141-169, 2020.
- GARIMELLA, Kiran; WEST, Robert. Hot streaks on social media. In: *Proceedings of the AAAI International Conference on Web and Social Media*. Palo Alto: AAAI Press, 2019. p. 170-180.
- HENDRICKS, Darryll; PATEL, Jayendu; ZECKHAUSER, Richard. Hot hands in mutual funds: Short-run persistence of relative performance, 1974–1988. *The Journal of finance*, v. 48, n. 1, p. 93-130, 1993.
- HOLOPAINEN, Risto. Making Complex Music with Simple Algorithms, is it Even Possible?. *Revista Vórtex*, v. 9, n. 2, 2021.
- JANOSOV, Milán; BATTISTON, Federico; SINATRA, Roberta. Success and luck in creative careers. *EPJ Data Science*, v. 9, n. 1, p. 9, 2020.
- KEOGH, Eamonn J.; PAZZANI, Michael J. Scaling up dynamic time warping for datamining applications. In: *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2000. p. 285-289.

- KISCHINHEVSKY, Marcelo; VICENTE, Eduardo; DE MARCHI, Leonardo. Em busca da música infinita: os serviços de streaming e os conflitos de interesse no mercado de conteúdos digitais. *Fronteiras-estudos midiáticos*, v. 17, n. 3, p. 302-311, 2015.
- LIU, Lu et al. Hot streaks in artistic, cultural, and scientific careers. *Nature*, v. 559, n. 7714, p. 396-399, 2018.
- LIU, Lu et al. Understanding the onset of hot streaks across artistic, cultural, and scientific careers. *Nature Communications*, v. 12, n. 1, p. 1-10, 2021.
- MARTÍN-GUTIÉRREZ, David et al. A multimodal end-to-end deep learning architecture for music popularity prediction. *IEEE Access*, v. 8, p. 39361-39374, 2020.
- OLIVEIRA, Gabriel P.; SILVA, Mariana O.; SEUFITELLI, Danilo B.; LACERDA, Anisio; MORO, Mirella M. Detecting collaboration profiles in success-based music genre networks. In: *Proceedings of Int'l Society for Music Information Retrieval Conference (ISMIR)*, 21, 2020, Montreal, Canada. p. 726-732.
- RAAB, Markus; GULA, Bartosz; GIGERENZER, Gerd. The hot hand exists in volleyball and is used for allocation decisions. *Journal of Experimental Psychology: Applied*, v. 18, n. 1, p. 81, 2012.
- RABIN, Matthew; VAYANOS, Dimitri. The gambler's and hot-hand fallacies: Theory and applications. *The Review of Economic Studies*, v. 77, n. 2, p. 730-778, 2010.
- SANDRONI, Clara et al. A Covid-19 e seus efeitos na renda dos músicos brasileiros. *Revista Vórtex*, v. 9, n. 1, 2021.
- SHINOHARA, Vítor; FOLEISS, Juliano; TAVARES, Tiago. Comparing Meta-Classifiers for Automatic Music Genre Classification. In: *Proceedings of the 17th Brazilian Symposium on Computer Music*. Porto Alegre: SBC, 2019. p. 131-135.
- SILVA, Mariana O.; ROCHA, Laís M.; MORO, Mirella M. Collaboration profiles and their impact on musical success. In: *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*. New York: ACM, 2019. p. 2070-2077.
- SINATRA, Roberta et al. Quantifying the evolution of individual scientific impact. *Science, American Association for the Advancement of Science*, v. 354, n. 6312, 2016.
- TAVENARD, Romain et al. Tslern, a machine learning toolkit for time series data. *J. Mach. Learn. Res.*, v. 21, n. 118, p. 1-6, 2020.
- WALDFOGEL, Joel. How digitization has created a golden age of music, movies, books, and television. *Journal of economic perspectives*, v. 31, n. 3, p. 195-214, 2017.

ABOUT THE AUTHORS

Danilo B. Seufitelli is a Ph.D. student in Computer Science at Universidade Federal de Minas Gerais (UFMG). He received the M.Sc. in Computer Science at the same university in December 2016. He received his Bachelor degree in Information Systems from Universidade Federal do Espírito Santo (UFES) in 2014. He is a member of the Laboratory of Interdisciplinary Computer Science (CS+X), and his research focuses on data analysis on collaborative domains, such as music, social networks, and open government data. Recently, he has been working on Både Project (Metrics on Data from/to Social Networks in Different Contexts). ORCID: <https://orcid.org/0000-0002-0155-7631>. E-mail: daniлоboechat@dcc.ufmg.br

Gabriel P. Oliveira is a Ph.D. student in the Computer Science Graduate Program at Universidade Federal de Minas Gerais (UFMG). He received the B.Sc. and M.Sc. degrees in Computer Science from UFMG in 2018 and 2021, respectively. He is currently a member of the Laboratory of Interdisciplinary Computer Science (CS+X) and a Data Scientist in the Analytical Capabilities Project. His research interests include Social Computing, Data Science and Social Network Analysis, with a strong emphasis in collaborative domains. ORCID: <https://orcid.org/0000-0002-7210-6408>. E-mail: gabrielpoliveira@dcc.ufmg.br

Mariana O. Silva is a Ph.D. student in the Computer Science Graduate Program at Universidade Federal de Minas Gerais (UFMG). She received the B.Sc. and M.Sc. degrees at UFMG in 2017 and 2020, respectively. She was part of the Apoena Project as an undergraduate research assistant and the Både Project as a M.Sc. student. She is currently a member of the Laboratory of Interdisciplinary Computer Science (CS+X) and a Data Scientist in the Analytical Capabilities Project. Her research interests include Data Science, Machine Learning and Social Network Analysis. ORCID: <https://orcid.org/0000-0003-0110-9924>. E-mail: mariana.santos@dcc.ufmg.br

Gabriel R. G. Barbosa is an undergraduate student in Electrical Engineering at Universidade Federal de Minas Gerais (UFMG). During this work, he was a research assistant sponsored by a CNPq scholarship. ORCID: <https://orcid.org/0000-0001-7930-4506>. E-mail: grgb@ufmg.br

Bruna C. Melo is an undergraduate student in Computer Science at Universidade Federal de Minas Gerais (UFMG). During this work, she was a research assistant sponsored by a CNPq scholarship. Her research interests include Databases, Social Networks and Software Development. ORCID: <https://orcid.org/0000-0002-4535-0288>. E-mail: brunacamposmelo@dcc.ufmg.br

Juliana E. Botelho is an undergraduate student in Computational Math at Universidade Federal de Minas Gerais (UFMG). She's been working with Gender Diversity in Computing within the Project Bytes & Elas since 2019. During this work, she was a research assistant sponsored by a CNPq scholarship. Her research interests include Math Education, Data Science and Team Management. ORCID: <https://orcid.org/0000-0002-2497-7559>. E-mail: juliana.botelho@dcc.ufmg.br

Luiza de Melo-Gomes is an undergraduate student in Information Systems at Universidade Federal de Minas Gerais (UFMG) and a research assistant. Her research interests include Data Science, Artificial Intelligence and Software Development. ORCID: <https://orcid.org/0000-0002-0756-2992>. E-mail: luizademelo@dcc.ufmg.br

Mirella M. Moro is an associate professor at the Computer Science department at UFMG (Belo Horizonte, Brazil). She holds a Ph.D. in Computer Science (University of California Riverside - UCR, 2007), and MSc and BSc in Computer Science as well (UFRGS, Brazil). She was a member of the ACM Education Council (2009-2018) and the Education Director of SBC (Brazilian Computer Society, 2009-2015), where she's currently a Council Member and part of the SBC Meninas Digitais (*Digital Girls*) Steering Committee. Her research interests include Data-driven Research, Social Analysis, Gender Diversity, and Computer Science Education. She is also an advocate for increasing women participation in Computing, coordinating projects such as *BitGirls* and *Bytes & Elas*. ORCID: <https://orcid.org/00000-0002-0545-2001>. E-mail: mirella@dcc.ufmg.br