Hot Streaks in the Brazilian Music Market: A Comparison Between Physical and Digital Eras

Gabriel R. G. Barbosa[1] **, Bruna C. Melo**[1] **, Gabriel P. Oliveira**[1] **,**
Mariana O. Silva[1] **, Danilo B. Seufitelli**[1] **, Mirella M. Moro**[1]

[1]Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6627 – 31270-901 Belo Horizonte, MG

grgb@ufmg.br, {brunacamposmelo,gabrielpoliveira,mariana.santos,daniloboechat,mirella}@dcc.ufmg.br

Abstract. *Consuming music through streams has made huge volumes of data available. We collect a part of such data and perform cross-era comparative analyses between physical and digital media for successful artists within the music market in Brazil. Given an artist's career, we focus on hot streak periods defined as high-impact bursts occurring in sequence. Specifically, we construct artists' success time series to detect and characterize hot streak periods for both physical and digital eras. Then, we assess their features, analyze them in the genre scale, and perform a cluster analysis to identify groups of artists with distinct success levels. For both physical and digital eras, we find the same clusters: Spike Hit Artists, Big Hit Artists, and Top Hit Artists. Our insights shed light on significant changes in the dynamics of the music industry over the years, by identifying the core of each era.*

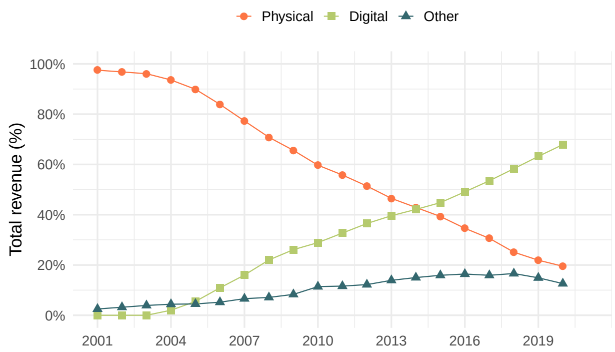


Figure 1: Global recorded music revenues by segment (2001–2020) – source: IFPI GMR[1]

1. Introduction

Physical media is constantly making room for the consolidation of the digital era. Figure 1 shows this process in the music industry, 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 62% of the music industry revenue in 2020.[1] The scenario in Brazil is similar. As Latin America's largest music market, about 72.4% of music revenue comes from digital media in Brazil, against 1.4% from physical, according to the most recent Pró-Música report.[2]

The current streaming dominance has proved beneficial in several ways, such as promoting local artists and increasing listeners' engagement. Also, while the world contends with the COVID-19 pandemic, people have once again been reminded of the timeless power of music to console, heal and lift their spirits, which strengthens the connection between artists and fans. Indeed, with the world in lockdown and live music shut down, most fans around the globe enjoy music via streaming. Still, streaming popularization has also brought new challenges due to the massive volume of music-related data to process and analyze.

An example of a task that has become more complex and important in recent years is finding and promoting artists with promising careers. If a contract with a major label was essential to be successful in the physical era, artists from smaller or independent labels can now go viral and become popular thanks to streaming services. Thus, Artificial Intelligence (AI) tools and techniques come to hand and assist in this type of analysis, creating significant benefits to 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 [1, 2] and success prediction [3, 4]. Regarding the latter, identifying upcoming artists with outstanding success is crucial, as it helps planning and adjusting marketing directions for their careers.

Overall, musical careers usually present continuous periods of success above average, defined as *hot streaks* (HS). This concept has been explored in several domains, including science [5], social media [6], and creative careers [7, 8]. In such a context, our goal is to identify and characterize hot streak periods in the music market in Brazil. We assess the evolution of successful careers by comparing data from 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) to 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 about temporal evolution of musical careers.

The remainder of this paper is organized as follows. First, we discuss related work in Section 2. Then, we describe the data acquisition process in Section 3. We detail the methodology used to identify hot streaks in Section 4. Next, we apply such methodology in the physical and digital eras in Sections 5 and 6, respectively, with a cross-era comparison in Section 7. Finally, we conclude with future directions in Section 8.

[1]IFPI Global Music Report: `https://gmr2021.ifpi.org/`
[2]Pró-Música Brasil Report: `http://bit.ly/ProMusica2019`

2. 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 [9]. 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 [9, 10, 11] but also on the success patterns shaping this dynamic market [4, 12, 13].

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 [14], automatic genre classification [1, 2, 15], 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 [16, 17, 18]. 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 [7].

For individual and creative careers, research assessing impact is much more recent. Liu et al. [7] 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 use data from Twitter, one of the most popular online social networks, and define users' impact as the reach of their content [6]. Janosov et al. [8] 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. 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 [13, 19], 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.

3. 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). Then, we also use the Pró-Música Brasil platform to describe the Physical Era, with data from 1990 to 2015. Next, we detail data acquisition processes for both physical (Section 3.1) and digital media (Section 3.2).[3]

Table 1: Pro-Música Brasil certification levels for Brazilian and foreign artists (A&S is Albumns and Singles).

	Brazilian		Foreign	
Certification	**A&S**	**DVDs**	**A&S**	**DVDs**
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

3.1. 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 website[4] 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.

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, its nationality and PMB's sales metric for the disc award. Finally, we use Spotify's API[6] to associate each artist to 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 (left).

3.2. Digital Media

Between 2016 and 2017, there was a key change in PMB's metric, which moved from physical media (i.e., DVD and CD) towards digital media (i.e., Singles and Albums), as depicted in Figure 2 (right). Meanwhile, streaming was already the main revenue source for digital media (58.3%).[7]

[3]Link to whole datasets omitted for blind review

[4]PMB Certificates: `https://bit.ly/CertificatesPMB`

[5]Wikipedia Python Library: `https://github.com/goldsmith/Wikipedia`

[6]Spotify API: `https://developer.spotify.com/`

[7]Pró-Musica Brasil 2016: `https://bit.ly/ProMusica2016`

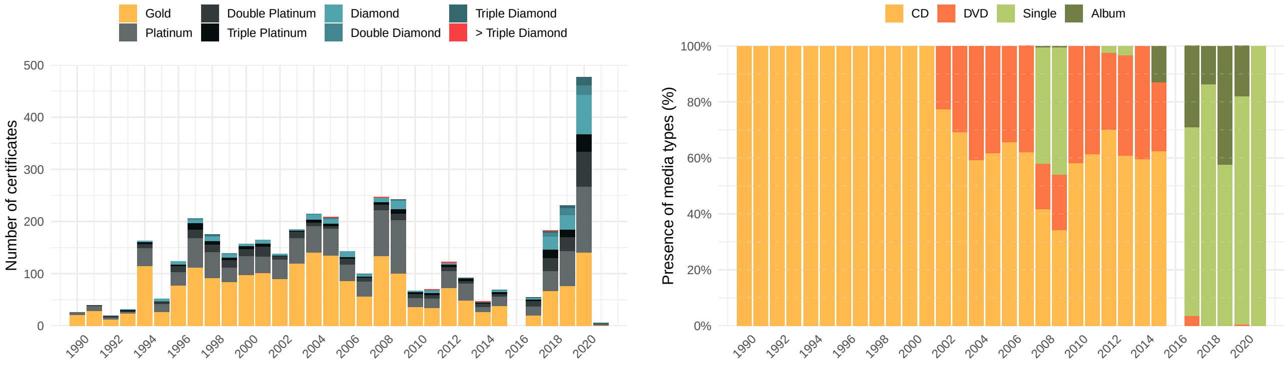


Figure 2: Discs certificated (left) and media type (right) in Pró-Música Brasil (1990–2021). In 2016, there was a metric change in the certification, hence the lack of data.

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.

4. Hot Streak Detection

After collecting the data, we are able to build artists' success time series for both eras (details in Sections 5 and 6). To detect hot streaks in such time series, we rely on previous work that shows the most successful points in professional careers tend to happen close to each other [6, 7]. Hence, we use a technique to reduce the time series dimensionality to delimiter continuous 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) [20]. Given a time series $X = x_1, x_2, \cdots, x_n$ of length n, PAA reduces it into a new series $\overline{X} = \overline{x_1}, \overline{x_2}, \cdots, \overline{x_N}$ 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 produces 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.

$$\overline{x_i} = \frac{n}{N} \sum_{j=\frac{n}{N}(i-1)+1}^{\frac{n}{N}i} x_j \qquad (1)$$

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* [21], 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, for each artist, we chose a specific threshold for defining the hot streak periods. Such an individualized approach is based on the percentiles of the success metric, and it allows analyzing 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 received only a few certificates and streams.

5. Physical Era

Here, we deepen the analysis on the PMB dataset by building the time series for each artist in Section 5.1, and characterizing the hot streaks and understanding their relationship to music genres in Section 5.2. Furthermore, we cluster artists into different success groups 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 artists' time series corresponds to the number of sales achieved by the artist in that year. This number is equivalent to the aggregate of certificates received (Table 1).

From the artists' time series, we detect the hot streaks periods using PAA (see Section 4). To do so, we first set the number of segments in which the series will be split, as this is the only parameter of the method. Since we deal with yearly success time series, the minimum size of each segment must be two years. After extensive empirical experiments, we set the window to the minimum, which is enough to validate a hot streak. Hence, the number of segments is calculated by dividing the time series length by such a 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.

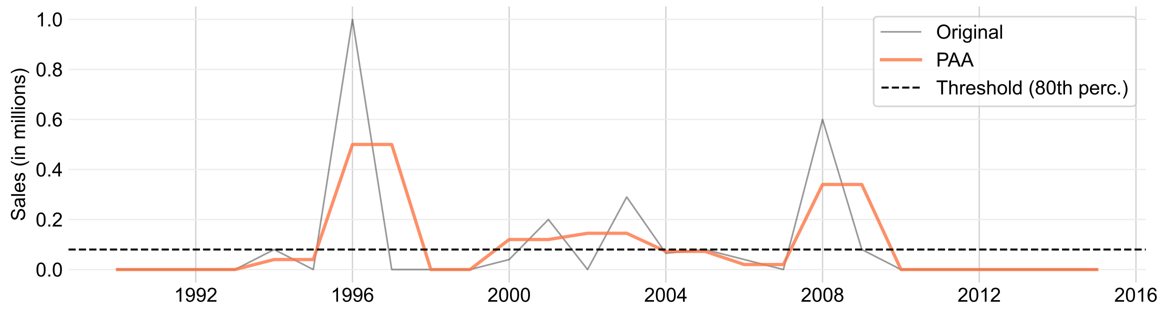


Figure 3: Piecewise Aggregate Approximation (PAA) applied to Skank's success time series in the Physical Era (1990–2015). Periods above the threshold are considered hot streaks.

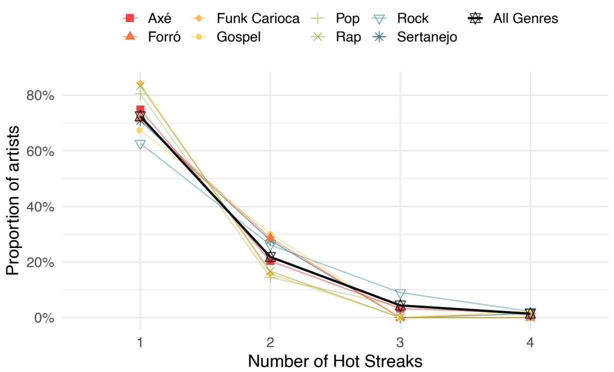


Figure 4: Hot streaks detected for artists of distinct genres in the Physical Era.

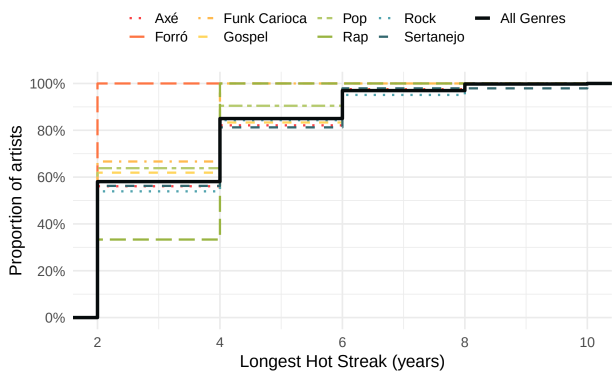


Figure 5: Cumulative distribution function of the duration of the longest hot streak in the Physical Era (in years).

In order to showcase how the hot streak detection works, we use the time series of the Brazilian band Skank. Figure 3 shows their time series after applying PAA. The threshold is set as the 80th percentile, as aforementioned. We observe three distinct hot streak periods. The first one from 1996 to 1997, when they released *O Samba Poconé*, a diamond disc awarded album. The second hot streak lasted from 2000 to 2003, when they released the albums *Maquinarama* and *Cosmotron*, which granted them platinum and gold discs, respectively. The last one occurred from 2008 to 2009, when they released the album *Estandarte*. Skank is the perfect example of a successful album that does not lead to a HS: in 1994, their album *Calango* got two golden discs, which is still inferior to their greatest hit *Garota Nacional* released in 1996.

5.2. Hot Streak Characterization

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

Figures 4 and 5 assess each genre's performance regarding the number of hot streaks as well as their maximum duration, respectively. In Figure 4, the majority of artists (about 70% to 94%) have up to two hot streak periods, and only 6% of the artists have between three and four hot streaks. Although such a pattern happens in all considered 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.

Next, we analyze the duration of hot streaks. We consider only the longest HS for each artist, as they may experience more than one. In Figure 5, we use a Cumulative Distribution Function (CDF), which informs the proportion of artists who have the longest HS with duration up to that value for a given number of years (x-axis). Note that the size of a HS is always be a multiple of two, as the PAA segment is set to two years (see Section 5.1). In general, only *Pop*, *Gospel* and *Funk Carioca* genres follow a similar behavior when compared to the average (i.e., *All Genres*). While the *Rap* genre has artists with shorter HS (up to four years), artists from the other genres have a wider range (between one and ten years). Therefore, in contrast to the previous analysis, we note that genre is relevant to describe the longevity of hot streaks.

5.3. Cluster Analysis

We now move to the cluster analysis, which helps better understand the characteristics of the different success levels of artists achieved during the Physical Era. We use the K-Means algorithm, which is the most commonly used clustering method for dividing a dataset into a set of k groups. The considered features for the algorithm include

Table 2: Main statistics on the artist clusters in the Physical Era.

	All	SHA	BHA	THA
Number of Artists	574	527	38	9
Average Number of HS	1.3	1.3	1.8	1.9
Median Sales (10^4)	8.5	8	152	507
Median Threshold	0	0	72,500	300,000

the total number of hot streaks, total sales and the time series threshold. In order to determine the optimal number of clusters, we use the Elbow Method [22] that plots the explained variation as a function of the number of clusters and consider the curve *elbow* as the optimal k. In our case, the method outcome suggests $k = 3$. We name the resulting clusters according to the success metric (i.e., number of sales): *Spike Hit Artists* (SHA), *Big Hit Artists* (BHA), and *Top Hot Artists* (THA). The main statistics of the clusters are presented in Table 2 and summarized as follows.

Spike Hit Artists (SHA). This cluster contains most artists (527). The median PAA threshold for such artists is 0, indicating that their sales are, in general, much lower than the artists from other clusters. In addition, the average number of hot streaks for SHA is 1.3. Such a result suggests that PMB certificates happen sparsely for such artists because there are not many hot streaks even with a low threshold. Artists in this cluster include Banda Eva, Claudinho e Bochecha, Coldplay, Paramore, and Latino.

Big Hit Artists (BHA). This cluster represents a bridge between the most and less successful ones, with 38 artists. Although BHA presents a higher number of sales when compared to SHA, they 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. Besides, there is a substantial increase in the threshold for artists in this cluster, reaching 72,500 sales. Madonna, Legião Urbana, Skank, and U2 are examples of BHA.

Top Hit Artists (THA). This is the cluster with the most successful artists, as they have the highest median number of sales. The average number of hot streaks for THA is very close to the value for BHA. However, the artists in this cluster achieved major sales success throughout the Physical Era, as they have a very high median threshold (300,000 sales). Examples of THA include Ivete Sangalo, Zeca Pagodinho, Roberto Carlos, and Sandy e Junior.

Figure 6 presents a scatter plot comparing the number of hot streaks and the total number of sales per artist. The results confirm the artists are clustered according to success level, i.e., total number of sales. Regardless of their number of hot streaks, SHA generally present the lowest, followed by BHA and THA.

6. Digital Era

We now deepen the analysis on the Digital Era by building the artists' time series that achieve Spotify's Brazil Top 200 Charts in Section 6.1. Next, from the constructed time series, we detect and characterize hot streaks in Section 6.2. Finally, we perform a cluster analysis to group similar artists based on their success levels in Section 6.3.

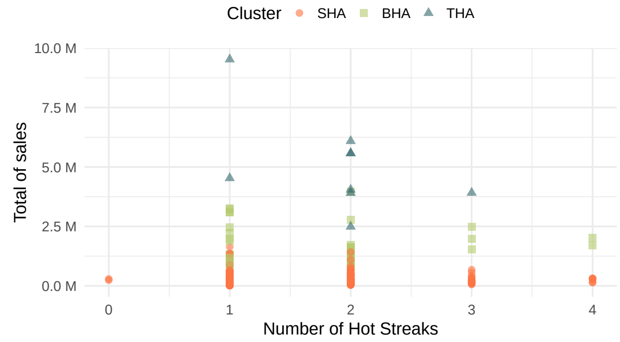


Figure 6: Number of hot streaks and total sales for clusters SHA, BHA, and THA.

6.1. Artists' Time Series

In the Digital Era, we use the Spotify Top 200 Chart in Brazil as our basis to model artists' success over time. For each artist, 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 the Spotify) per week. For example, Figure 7 presents the time series of the Brazilian singer Anitta.

Similar to the Physical Era, we apply PAA (Section 4) in each time series to detect the hot streaks periods. This method helps balance weeks with little or no success metric value, which was quite common in the data. In the Digital Era, songs and artists achieve success much faster. In addition, the digital nature of streaming platforms allows for successful weekly data to be made available almost in real-time. Hence, according to our experiments, we define 12 weeks as the size of each PAA segment. Each segment comprises a three-month period, which is a reasonable time to analyze the continuous periods of great success in streaming platforms such as Spotify.

We define hot streaks as the periods in which PAA is above a threshold, defined per each artist. However, as we are dealing with numerous single weeks on the Spotify Chart, we consider the artists' Activity Rate (AR) based on the threshold. The AR is the number of weeks that an artist appears on the chart divided by the total number of weeks. Hence, if AR $< 10\%$, the threshold is set as 95% of the success measure; if $10\% \leq$ AR $< 15\%$, the threshold is set as 90%; if $15\% \leq$ AR $< 20\%$, the threshold is set as 85%; finally, if AR $\geq 20\%$, the threshold is set as 80%.

Figure 7 depicts PAA applied to Anitta's career, currently one of the most influential Brazilian artists worldwide. She is present in all considered weekly charts (AR $= 100\%$) and thus, we set the 80th percentile as the threshold for defining hot streaks. There are three HS in her time series. The first one from November 2017 to April 2018, when she released *Vai Malandra*, which became the most streamed song on its release date.[8] The second HS

[8] No Brasil, "Vai Malandra" supera Taylor Swift e é a música

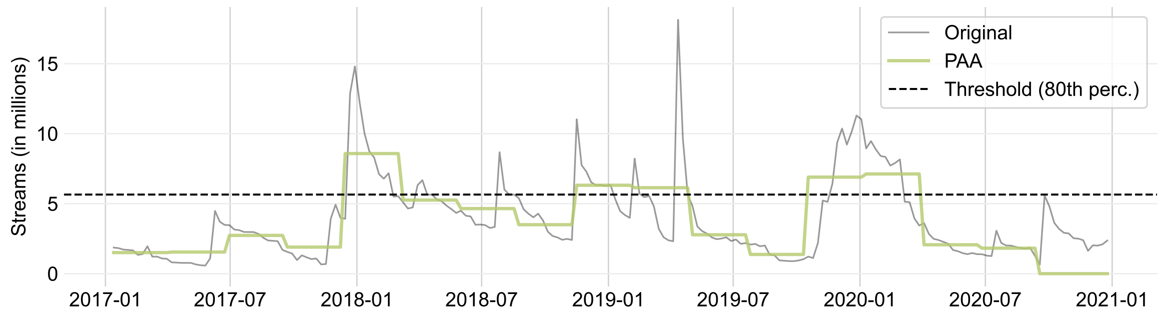


Figure 7: Piecewise Aggregate Approximation (PAA) applied to Anitta's success time series in the Digital Era (2017–2020). Periods above the threshold are considered hot streaks.

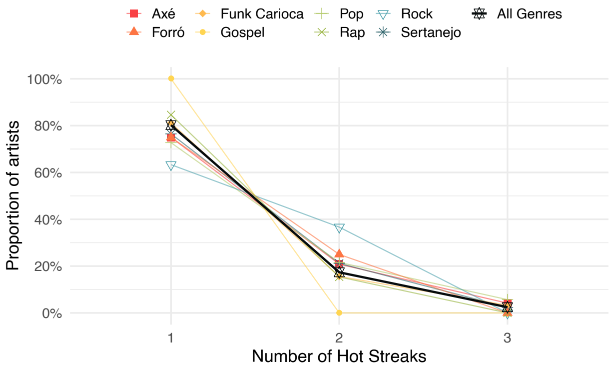


Figure 8: Hot streaks detected for artists of distinct genres in the Digital Era.

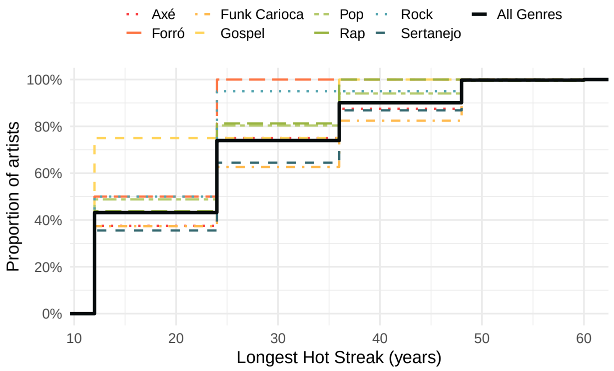


Figure 9: Cumulative distribution function of the duration of the longest hot streak in the Digital Era (in weeks).

period coincides with *Veneno* and *Não Perco Meu Tempo* single releases. Finally, the third one comprises the period in which she collaborated with both national and international famous artists, such as Marília Mendonça (*Some Que Ele Vem Atrás*) and Black Eyed Peas (*Explosion*).

6.2. Hot Streak Characterization

We characterize the hot streak periods identified in the previous section. As in the Physical Era, we also consider the same super-genres in our analyses. Figures 8 and 9 present the HS analysis in quantity and duration perspectives, respectively. Regarding the number of hot streaks, most artists (about 80% to 98%) have up to two hot streak periods, similar to the Physical Era. Besides, only about 2% of artists achieved three hot streaks in the Digital Era. Although all genres follow a similar trend in general, *Rock* presents a higher percentage of artists with two HS. In contrast, all *Gospel* artists have only one HS in such period.

Next, we assess the duration of hot streaks. Analogous to the Physical Era, we consider only the longest HS for each artist. Overall, about half of the artists from all genres have three-month long hot streaks. However, when analyzing the genres individually, there are specific hot streak patterns. For instance, most *Axé* artists have shorter HS periods, as more than 60% of them present a 12-week hot streak. On the other hand, *Sertanejo* artists have longer hot streaks, as about 60% of them have HS periods with up to 24 weeks, i.e., six months.

mais tocada no Spotify em sua estreia: `https://bit.ly/AnittaVaiMalandra`. Accessed on June 26, 2021.

Table 3: Main statistics on the artist clusters in the Digital Era.

	All	SHA	BHA	THA
Number of Artists	1,018	940	70	8
Average Number of HS	0.6	0.5	1.6	1.6
Median Streams (10^5)	7.6	3.9	2,190	10,857
Median Threshold	0	0	2,052,131	7,145,116

6.3. Cluster Analysis

We now perform cluster analysis over the Digital. We also apply the K-Means algorithm in the time series, and the Elbow method to find its optimal number of clusters. As the Physical Era, the optimal k number found is three. We name the resulting clusters according to the success metric (i.e., number of streams); as the results are similar to the Physical Era, the names remain the same. The clusters main statistics are in Table 3 and summarized as follows.

Spike Hit Artists (SHA). This cluster comprises most artists (940). The median PAA threshold for such artists is 0, indicating that their streams are, in general, lower than other clusters. In addition, the average number of hot streaks for SHA is 0.5. However, we cannot say that such artists were not successful because they are in 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.

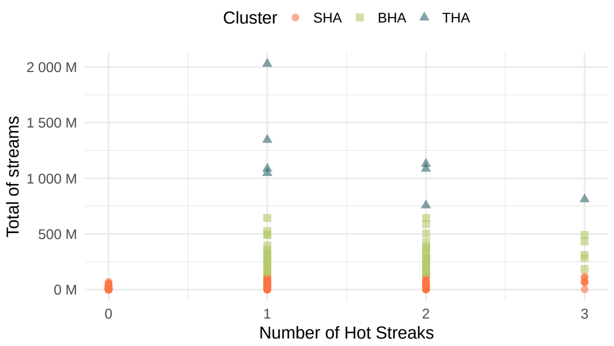


Figure 10: Number of hot streaks and streams for clusters SHA, BHA, and THA.

Big Hit Artists (BHA). This group represents a bridge between the most and less successful ones, comprising 70 artists with an average of 1.6 hot streaks between 2016 and 2020. Besides, the median streams are considerably higher compared to the previous group, as is the threshold, reaching more than 2M streams. Barões da Pisadinha, Dua Lipa, and Pabllo Vittar are examples of BHA.

Top Hit Artists (THA). This cluster contains only eight artists, with an average of 1.6 hot streak periods per artist. 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 may be considered highly successful artists, as their songs achieve a higher stream count throughout the weeks. All THA in the Digital Era are Brazilian, and examples include Anitta, Marília Mendonça, and Zé Neto & Cristiano.

Figure 10 presents a scatter plot comparing the number of hot streaks and the total number of streams for each artist. The results confirm that the artists are clustered according to their success level. Overall, all clusters patterns are similar to the Physical Era. Specifically, SHA generally present lower sales, while THA are the ones with higher sales. Furthermore, BHA have an intermediate level of sales. Hence, our clustering approach is coherent, as it manages to distinguish the levels of successful artists.

7. Cross-era Comparison

Music is part of people's daily lives regardless of the era experienced, whether Physical or Digital. With musical consumption constantly on the rise, we may notice similarities between both eras. From the results obtained in Sections 5 and 6, we 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. We highlight the BHA and THA clusters, which comprise the most successful artists in both eras (i.e., higher sales and streams), including Skank and Anitta. In particular, the second one represents the paramount artists, who are all 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.

As for the differences, Figure 11 shows a rotation in the preference of the musical genre. Specifically, there was an increase in *Gospel* sales in the Physical Era but not in the Digital one. This fact may be because *Gospel* listeners still consumed physical media by 2015, while audiences from other genres had already migrated to streaming in previous years (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* (Ivete Sangalo), *Sertanejo* (Zezé di Camargo & Luciano) and *Rock* (Skank); whereas in the Digital Era, the most successful artists (THA) come from one style: *Sertanejo*, with for more than 50% of streamings in late 2020.

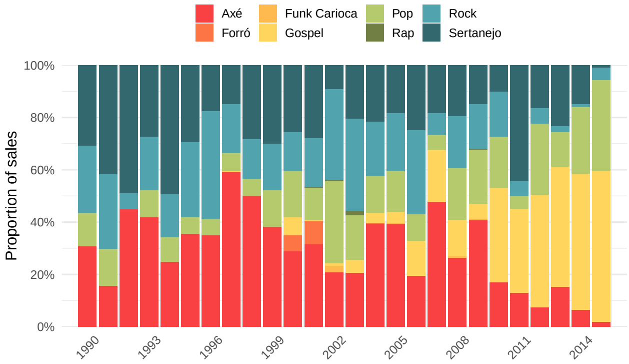
Overall, the Digital Era allows the appearance of new popular genres, as well as the decline of previously popular ones. For example, the prevalence of *Sertanejo* is remarkable over time, while *Pop* decreases from 2016 to 2020. Moreover, we highlight the rise of *Forró* in mid-2020 as a popular genre, following the growth of popular artists who have bursted 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 to new realities. In fact, Marília Mendonça had the most-streamed YouTube live worldwide in 2020, with over 3.31 million viewers.

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 entirely 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.

8. Conclusion

Here, we performed a cross-era comparative analysis between physical and digital musical media in Brazil, the largest market for recorded music in Latin America. First, we built artists' success time series to identify hot streak periods for both eras, defined as continuous high-impact bursts. Next, we characterized such periods to understand the dynamics of success among artists from different musical genres. Our results showed that, although there are similarities among all music styles, 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. Finally, a profiling analysis uncovered three different clusters in both eras: Spike Hit Artists (SHA), Big

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



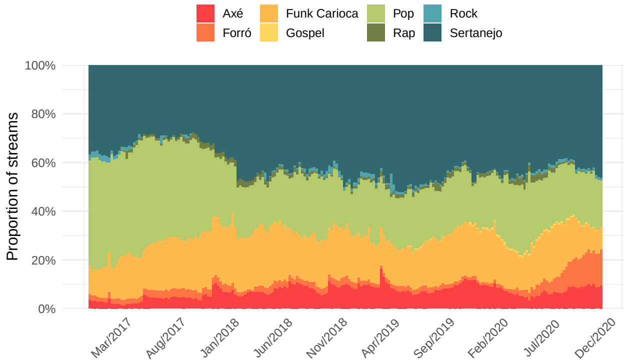


Figure 11: Genre evolution in (left) Physical and (right) Digital Eras.

Hit Artists (BHA), and Top Hit Artists (THA), which acted as class descriptors of successful artists.

Overall, our findings offer meaningful insights for MIR tasks, such as prediction and recommendation. For example, the identified clusters information may be used as features in 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 as a whole. Analyzing the evolution of artist careers reveals success trends in Brazil from what people consume. Indeed, our results uncover that although Brazilians are connected 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 not only the human element in the music industry (e.g., A&R executives and record labels CEOs) but also people's relation 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.

Acknowledgments. Work supported by CAPES and CNPq, Brazil.

References

[1] Débora C. Corrêa and Francisco Ap. Rodrigues. A survey on symbolic data-based music genre classification. *Expert Systems with Applications*, 60:190–210, 2016.

[2] Vítor Shinohara, Juliano Foleiss, and Tiago Tavares. Comparing meta-classifiers for automatic music genre classification. In *SBCM*, pages 131–135, 2019.

[3] Carlos Soares Araujo, Marco Cristo, and Rafael Giusti. Predicting music popularity on streaming platforms. In *SBCM*, pages 141–148, 2019.

[4] D. Martín-Gutiérrez et al. A multimodal end-to-end deep learning architecture for music popularity prediction. *IEEE Access*, 8:39361–39374, 2020.

[5] R Sinatra, D Wang, P Deville, C Song, and A-L Barabási. Quantifying the evolution of individual scientific impact. *Science*, 354(6312), 2016.

[6] Kiran Garimella and Robert West. Hot streaks on social media. In *ICWSM*, pages 170–180, 2019.

[7] Lu Liu et al. Hot streaks in artistic, cultural, and scientific careers. *Nature*, 559(7714):396–399, 2018.

[8] Milan Janosov, Federico Battiston, and Roberta Sinatra. Success and luck in creative careers. *EPJ Data Sci.*, 9(1):9, 2020.

[9] Leonardo De Marchi and João Martins Ladeira. Digitization of music and audio-visual industries in brazil: new actors and the challenges to cultural diversity. *Cahiers d'Outre-Mer*, 71(277):67–86, January 2018.

[10] Marcelo Kischinhevsky, Eduardo Vicente, and Leonardo De Marchi. Em busca da música infinita: os serviços de streaming e os conflitos de interesse no mercado de conteúdos digitais. *Fronteiras-estudos midiáticos*, 17(3):302–311, 2015.

[11] Joel Waldfogel. How digitization has created a golden age of music, movies, books, and television. *Journal of economic perspectives*, 31(3):195–214, 2017.

[12] Mariana O. Silva, Laís M. Rocha, and Mirella M. Moro. Collaboration Profiles and Their Impact on Musical Success. In *ACM/SIGAPP SAC*, pages 2070–2077, 2019.

[13] Gabriel P. Oliveira et al. Detecting collaboration profiles in success-based music genre networks. In *ISMIR*, pages 726–732, 2020.

[14] R Borges and Marcelo Queiroz. A probabilistic model for recommending music based on acoustic features and social data. In *SBCM*, pages 7–12, 2017.

[15] R de Araújo Lima et al. Brazilian lyrics-based music genre classification using a BLSTM network. In *ICAIS*, 2020.

[16] Darryll Hendricks, Jayendu Patel, and Richard Zeckhauser. Hot hands in mutual funds: Short-run persistence of relative performance, 1974–1988. *The Journal of finance*, 48(1):93–130, 1993.

[17] Matthew Rabin and Dimitri Vayanos. The gambler's and hot-hand fallacies: Theory and applications. *The Review of Economic Studies*, 77(2):730–778, 2010.

[18] Markus Raab, Bartosz Gula, and Gerd Gigerenzer. The hot hand exists in volleyball and is used for allocation decisions. *Journal of Experimental Psychology: Applied*, 18(1):81, 2012.

[19] Gabriel B. Vaz de Melo, Ana F. Machado, and Lucas R. de Carvalho. Music consumption in Brazil: an analysis of streaming reproductions. *PragMATIZES - Revista Latino-Americana de Estudos em Cultura*, 10(19):141, 2020.

[20] Eamonn J. Keogh and Michael J. Pazzani. Scaling up dynamic time warping for datamining applications. In *SIGKDD*, pages 285–289. ACM, 2000.

[21] R Tavenard et al. Tslearn a machine learning toolkit for time series data. *J.Mach.Learn.Res*, 21:118:1–118:6, 2020.

[22] Purnima Bholowalia and Arvind Kumar. Ebk-means: A clustering technique based on elbow method and k-means in wsn. *Int'l J. of Computer Applications*, 105(9), 2014.