TrendTune - Analysis and forecast of music popularity trends by genre and location

Larissa Terto Alvim
School of Arts, Sciences and Humanities
University of São Paulo
São Paulo, Brazil
larissa_terto@usp.br

Rafaela Elache Maciel
School of Arts, Sciences and Humanities
University of São Paulo
São Paulo, Brazil
rafaelaelache@usp.br

Thaís Risa Fujissaka
School of Arts, Sciences and Humanities
University of São Paulo
São Paulo, Brazil
thais.fujissaka@usp.br

Livia Vanessa Carlini Martins
School of Arts, Sciences and Humanities
University of São Paulo
São Paulo, Brazil
liviavcarlini@usp.br

Laila Malafaia Vieira
School of Arts, Sciences and Humanities
University of São Paulo
São Paulo, Brazil
lailamalafaiav@usp.br

Regis Rossi Alves Faria
Lab of Audio and Music Technologies
University of São Paulo
São Paulo, Brazil
regis@usp.br

***Abstract*—With the growth and diversification of the digital music market, there is a need for tools to help music producers and independent artists understand the music consumption behavior of their target audience. We introduce an online tool designed to identify music consumption patterns and then display music popularity by genre and location, allowing for analysis of consumption information over time. Additionally, it enables the estimation of popularity trends using predictive models, allowing artists to target their campaigns with greater precision and thereby increase their chances of success in the market. A functional prototype based on the retrieval of information from music public performances and from music platforms is presented, and future perspectives are addressed.**

***Index Terms*—Digital music market, music popularity analysis, music trend forecasting, Music Information Retrieval, Internet of Musical Things.**

I. Introduction

The music industry has undergone significant transformations with the advent of social media and streaming platforms, which now dominate the distribution and consumption of music in digital format. According to the International Federation of the Phonographic Industry (IFPI) Global Music Report 2024, global recorded music revenues reached $28.6 billion in 2023, representing a 10.2% increase over the previous year. Of this total, 48.9% of revenues came from streaming subscriptions, with over 500 million paid subscriptions, solidifying streaming as the main driver of the music industry [1].

However, as noted in a 2022 article by Music Business Worldwide (MBW), a respected platform for in-depth coverage of the music industry, market saturation has significantly increased, with approximately 100,000 new tracks being uploaded daily to streaming services like Spotify [2]. MBW is known for its comprehensive analysis of trends in the music business, including data on streaming platforms, artist releases, and industry shifts. This underscores that, while digitalization has democratized access and entry for artists, it has also posed challenges for their visibility and the discovery of new talent.

In this context, there is a need for tools that help music artists understand the music consumption behavior of their target audience. This article presents an application that allows artists to identify music consumption patterns based on geographic and musical genre data over a period of time. In addition, predictive models are used to enable them to target their marketing campaigns more accurately, increasing their chances of success in the musical market [3], [4].

The developed tool fits into the fields of Music Information Retrieval (MIR) and Internet of Musical Things (IoMusT), by exploring ways to retrieve, organize, analyze, and process large volumes of music-related data. The prediction of music trends has gained relevance in this field, and this application has the potential to benefit the music market by providing valuable insights into audience behavior and upcoming trends for artists and entrepreneurs, offering a competitive advantage by predicting consumption trends [4], [5].

The application was designed with a focus on independent and emerging musical artists, particularly those facing challenges in entering the music market. This group includes students and independent musicians without contracts with major record labels, who typically lack support for conducting market analyses, and the developed application emerges as a means to help identify trends that would otherwise be monopolized by large record labels or marketing departments.

The application is also useful for artists whose musical styles do not follow general trends, who explore alternative, experimental, or niche genres. These artists often have difficulties identifying and connecting with their target audience; therefore, the tool is designed to assist them in understanding which regions consume their music the most.

This article presents the goals and methods that guided the development of the proposed application. The two main goals to ensure its utility for music artists are the following:

1) **Analysis of musical genre popularity by location and time**: This goal involves offering a detailed graphical

analysis of how specific music genres are consumed through a segmentation by location and period of time. The tool is to allow for an accurate understanding of listeners' consumption profiles. This functionality could assist both emerging artists and those whose musical styles are not very popular and face challenges reaching their target audience. Based on the analysis of this information, artists and/or their music producers are expected to be able to adjust their marketing campaigns and strategically expand their reach, ensuring that their marketing resources are utilized efficiently [3].

2) **Prediction of musical genre popularity by location and time**: This goal involves creating a computational model capable of making useful predictions about the popularity of different music genres in various geographic locations. The accuracy of these predictions is to increase by integrating historical and behavioral data extracted from both music streaming platforms — like Spotify — and music public performance analysis offices — which provide rankings of the most listened songs in the digital domain and most played songs in public music performances, respectively. In Brazil, the Brazilian Central Office of Collection and Distribution, aka ECAD (from *Escritório Central de Arrecadação e Distribuição* in Portuguese), is an official institution responsible for collecting and distributing royalties from public musical performances, counting the number of performances in concert halls, events, bars, restaurants, radio and TV stations [6]. This will allow artists and their representatives to understand which musical styles are rising in certain areas, facilitating the adaptation of marketing strategies.

II. Related Works

Music marketing models are rapidly being renewed and expanded, and their precision to explain (and predict) market trends seems to benefit from the aggregation of data from cultural dimensions. Cultural association with music attributes is found to improve the marketing success [31]. Related works developed by specialists who seek to understand music trends and the factors that influence audiences' listening patterns helped the creation of the predictive and analytical tool proposed by this article.

The work of Katz (2010) [7] provides the context of how technologies impact the music industry, enabling the democratization of access and personalized consumption between artists and listeners. Thus, information about the popularity of songs and predictive tools for music trends becomes important to gather and understand the influences of historical and social factors in the music industry, as well as to comprehend the role of technology as a driving force behind the emergence of trends and artists [4]. To gather information about the popularity of songs, many works utilize the Application Programming Interface (API) of the Spotify platform [8], which consists of request methods available for developers to access information about the acoustic characteristics of songs, and the popularity of tracks and artists in different global markets.

One approach to this research problem consists of creating or manipulating databases of information about the popularity of songs. The project by Mondelli, Gadelha Jr., and Ziviani (2019) [9] seeks to map data based on information obtained from Spotify. The API does not provide methods to request the most popular music genres in a specific country, but it is possible to obtain the genre associated with an artist, and then, the most popular artists in a location. Thus, the researchers mapped the most listened genres in various music markets around the world and presented consumption patterns in different countries. Moorhouse (2021) [10] also used the Spotify API to perform an analysis on users' listening habits based on the acoustic features of their most listened songs. One of the features retrieved was the popularity of a track: a value between 0 and 100 calculated by the Spotify algorithm based on the number of recent plays. The calculation behind this feature inspired the approach described in this article, which created a metric for the popularity score of a musical genre based on the number of songs played in that genre.

Another approach to this research problem is the development of algorithms for predicting music trends based on consumption data across different digital platforms. The work of Saragih (2023) [11] utilized regression and machine learning models to predict the popularity of songs based on their audio features and consumption data from Spotify in Indonesia. The prediction model was also used to evaluate the relevance of each feature in determining the popularity of a song.

Valcarcel (2019) [12] obtained a result similar to that of [9], but within the scope of cities. They used the *Every Place at Once* website[1], a georeferenced compilation of the most played songs with access to the audio tracks, to create a unified database that associates musical styles with their popularity in a city. To obtain the data from the site, they performed a web scraping technique, in which the algorithm seeks the information contained in a website's HTML elements [13].

Methods commonly reported in the literature for predicting music popularity include decision trees (DT), Random Forest models, KNN (K-nearest neighbors), linear and logistic regression, neural networks and machine learning classifiers, and Support Vector Machines, among others. Lee & Lee (2018) present an interesting perspective on (possible) metrics for assessing popularity and methods for classification [34]. Temporal data models such as non-linear auto-regressive (neural network) classifiers have been used for the prediction with prediction accuracy around 50% [30].

Finally, besides the scope within MIR, we find a close affinity with the context of the Internet of Musical Things (IoMusT) [33], an extension of the Internet of Things (IoT) dealing with a distributed network of metadata objects about music-related features and metrics, serving a musical (promotion) purpose.

[1] https://everynoise.com/everyplace.cgi

III. Methodology and Implementation

Firstly, we conducted a literature review of related works and developed a survey to assess our primary users' interests, through which we built a ranked list of desired features. To achieve the two goals listed in Section I, the methodology comprised the development of a predictive model (Section III-D) and the implementation of a user interface with information visualization capabilities (Section III-C). Both phases were preceded by the construction of a database of the popularity of musical genres across different locations in previous years (Section III-B).

The construction of a database of the popularity of musical genres was accomplished by collecting monthly data of music consumption segmented by city in Brazil from ECAD [6] and from Spotify Charts [15] — a website with data-driven rankings showcasing the most popular songs and artists played on the Spotify platform. For the forecasting module, the next phase consisted of developing and evaluating the performance of a prediction model against a real/known scope of time. The user interface was implemented and tuned after completing the popularity analysis and forecast modules. The visualization choice was to show popularity information for each location using a geographical map interface. For this, we created visualizations using the Power BI software [14] to display data of music genre popularity for each location, and dynamic tables and graphs to display the most to the least popular genres for a given period.

Figure 1 shows a block diagram for the whole application, and the following sections detail the components implementation.

A. Features selection and design concepts

A brief survey was carried out with music students and emerging artists to determine what features would be most desired in a platform designed to provide information about the music market and predict the popularity of a musical genre. This was conducted through a Google Forms survey, and, though small, it was useful to establish the design requirements for the features and functionalities and for adapting the visualizations for a proof-of-concept.

The first key information acquired was the primary digital platform used for music promotion among our target users, with 77.8% of responses indicating that they used Spotify regularly for this purpose. This was significant for choosing it as a main online data source, employing the music charts provided by this platform. The participants also indicated that they were not aware of any tools for predicting music trends or popularity, and 66.7% considered "predicting popularity trends" of music styles to be very useful (maximum score on a 5-level importance grading scale), indicating a strong demand for this feature.

Out of several options for visualizing information about music trends, the "Interactive Geographic Map" received the highest number of votes for best choice in terms of practicality and ease of reading data, with 55% votes for a maximum score of 5 (on a 5-level scale) and 33% votes for a score of 4. Line graphs came in second, with 22% votes for a score of 5 and 44% votes for a score of 4.

The survey also demonstrated the participants' interest in information about "trends by music genres", with 100% of votes considering it the most valuable. This information was essential for determining the visualization scheme to be used on the interfaces designed to display forecasts and historical popularity of music genres in specific locations.

For this project, we considered the Brazilian music market as the target, covering the scope of its locations and musical genres according to existing classifications in the Brazilian context, and considering the availability of data for this market. The application was designed to be accessible online, built as a web service.

B. Data Collection

To provide graphs and analyses based on the most listened genres in Brazil, it was necessary to find databases as reliable as possible about listeners' play counts. A key innovation of our approach is incorporating live-concert data to gauge genre popularity, offering a perspective on musical trends that extends beyond digital streams. Specifically, for each month of 2023 and 2024, we compiled the top 10 most-performed songs at concerts held throughout Brazil, using data collected from ECAD.

The second database of music popularity used to support the application was Spotify Charts [15], which provided the 100 most listened to songs weekly in Brazil and in specific cities, as well as the number of listeners for each song in the ranking. For standardization purposes, only the last week of each month was selected to reduce the influence of the previous month's ranking on the popularity and capture a consolidated view of the performance of the various genres during the month, just as ECAD's rankings do. In a first sampling to study, we compiled the rankings of the most played songs in Brazil in 2023 and 2024, and in the city of São Paulo in 2023.

The ECAD and Spotify Charts databases do not directly present the genres of each song; therefore, the Spotify API was used to retrieve the associated music genre for each song in the sample. The API was accessed using the Python programming language and the *spotipy* [16] and *pandas* libraries. At this point, a limitation was found based on the registered genre for a song in the Spotify API: it simply associates the musical style of the artist with the song. This means that the popularity of certain music genres is strongly tied to the artist's own descriptions, making their actual popularity harder to predict. This issue will be addressed in the discussions later in Section V.

After this initial collection, the search patterns involved in the task were identified, allowing the creation of an automation algorithm developed in Java for retrieving data from the Spotify Charts. This code fetches the *json* object returned from the webpage access request. With this data, which includes the song and the artist, the code makes a request to the Spotify API to obtain the genre associated with each song in the ranking. Finally, the list with the maximum ranking

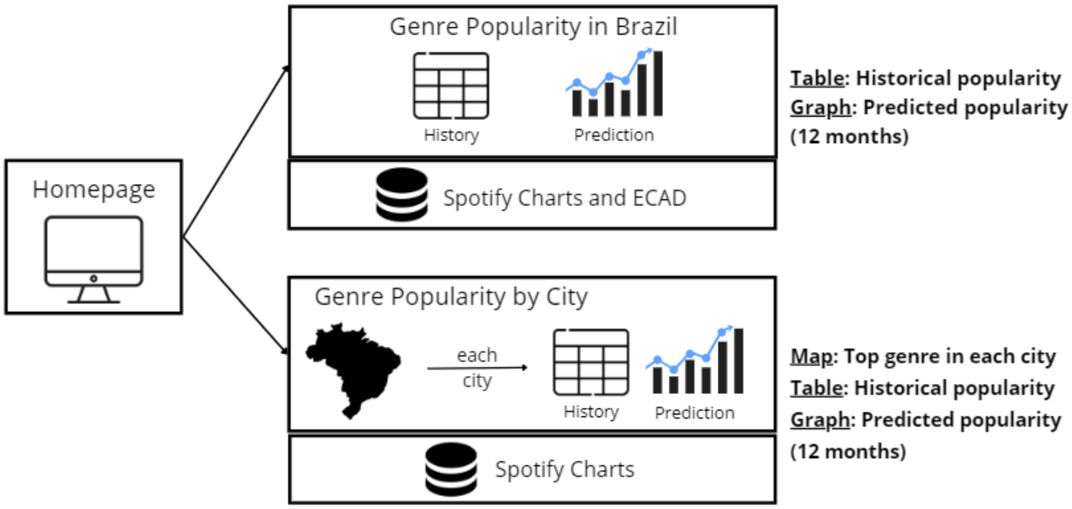

Fig. 1. Block diagram of the developed application, illustrating its two main modules (”Genre Popularity in Brazil” and ”Genre Popularity by City”) and their data sources. The application integrates historical music consumption data from Spotify Charts and ECAD to generate both historical and 12-month predicted popularity visualizations (Tables and Graphs) at the national level and city-specific geographical map interfaces.

for each music genre is compiled into a CSV file for each month. To expedite this development and ensure scalability, two widely adopted tools were used: *Spring Boot*, a framework that simplifies the creation of stand-alone, production-grade Spring-based applications [17]; and *Apache Maven*, a build automation and dependency management tool for Java projects [18].

Through the use of this web-scrapping code we finally gathered the city-specific rankings from Spotify Charts in 2023 and 2024 for two cities with different cultural profiles and preferences: São Paulo and Cuiabá. While Cuiabá is more influenced by regional prevalences, São Paulo offers a more diverse music market, encompassing both regional and international genres, justifying this selection for comparisons. The data was organized into CSV files with the following structure: music genre, year, month, ranking, and the genre popularity.

The genre popularity score created is based on the ranking from the Spotify Charts, and is computed as follows. The genre ranking is proportional to the number of plays of the most-played song belonging to that genre. The genre that received the most plays was ranked first in the list. The popularity column is a custom score from 0 (lowest) to 10 (highest), where genres ranked 1 receive 10 points, and genres ranked below 10 receive 0 points; the calculation performed is 11 minus the ranking of the musical genre. There were genres that appeared in most of the months but not all; to avoid excluding them from the prediction model due to a lack of data for all months, we assigned a score of 0 to the missing data.

To combine monthly data for the 10 most-played genres in Brazil, from both ECAD and Spotify, a weighted average between the rankings from the different sources was calculated. The weight used for the ECAD ranking was 0.7, and 0.3 for the Spotify Charts. This difference was an empirical attempt to reduce an inherent inaccuracy in assessing the popularity of actual (real) digital music consumption, as digital rankings can be easily manipulated through algorithms to automate music playback, known as *bots* [19].

C. Historical popularity module

Based on the collected data described in Section III-B, data visualizations were created using the commercial software Power BI [14]. A graphical user interface (GUI) was implemented to support navigation through the different visualizations designed for this implementation, providing an intuitive walkthrough of popularity analysis across different geographical segmentations.

The GUI was implemented with JavaScript, CSS, and HTML, along with the *Bootstrap* framework, a customizable toolkit for front-end development [20]. The incorporation of the Power BI graphics into the GUI was achieved using HTML iframe elements, which embed the live visualization into the interface, maintaining data synchronization and interactivity.

The GUI consists of two main webpages: ”Popularity Brazil”, which presents an analysis of popularity data for the country as a whole, and ”Popularity by City.” Below, we describe the graphs and filters presented on each page:

- **Popularity Brazil**:
  - **Annual Popularity by Genre in Brazil**: This graph, shown in Figure 2, displays the evolution of music genre popularity in Brazil over time, based on the consolidated real data. This graph is also a line graph, providing a clear view of historical variations. Filters by year and genre are available, allowing for more refined and specific analysis.
  - **Monthly Popularity Score in Brazil**: This visualization, shown in Figure 3, consists of a

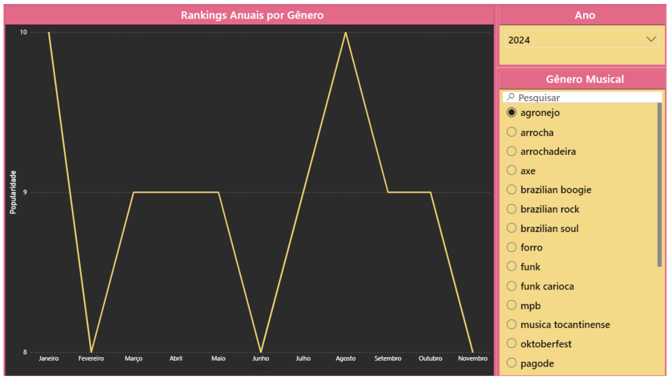

Fig. 2. Power BI graph visualization with the annual popularity of music genres in Brazil. The line graph illustrates the evolution of the popularity of a selected musical genre over the twelve months of the selected year, based on consolidated historical data. The selected filters for year and genre are 2024 and 'agronejo', respectively.

dynamic table presenting the most popular music genres in Brazil in descending order of popularity. The table is interactive and offers filters by year, month, and genre, enabling detailed analysis over different periods. This format allows for both the visualization of general patterns and specific insights, depending on the user's focus.

The visualizations related to Brazil enable users to identify genre popularity in specific periods, assess the impact of events or music releases on the consumption of certain genres, and understand national preferences comprehensively.

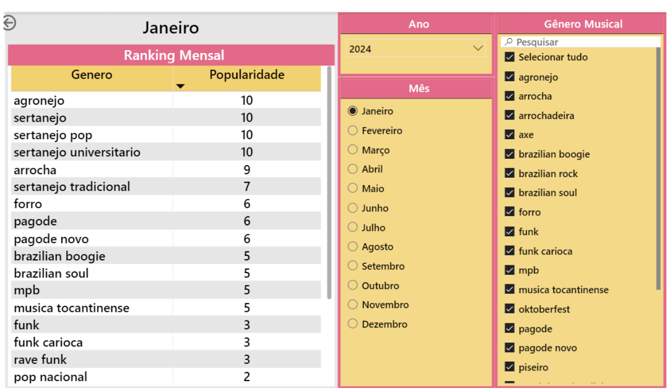

Fig. 3. Power BI table visualization with the monthly popularity ranking of music genres in Brazil. The selected filters for year, month, and genre are 2024, January, and All, respectively. This means that the visual displays the popularity score of all available genres for the selected time period.

- **Popularity by City**
  - **Most Played Genres by City**: This graph, displayed in Figure 4, shows the distribution of music genre popularity in specific cities, allowing for comparisons of music consumption in different locations. Available filters include year, month, and Spotify ranking. Additionally, if cities have the same genre for the same rank, the size of the bubbles differs based on the number of weeks that the genre was in that position in the ranking.

This page also contains a selector tool that allows the user to navigate to the page of a specific city. On the city page, it is possible to see the details of music genre popularity in the same way presented on the Popularity Brazil page: there is a line graph for the historical popularity of music styles (not shown, similar to that of Figure 2), and a table that allows viewing the historical rankings of all available music genres (similar to that of Figure 3).

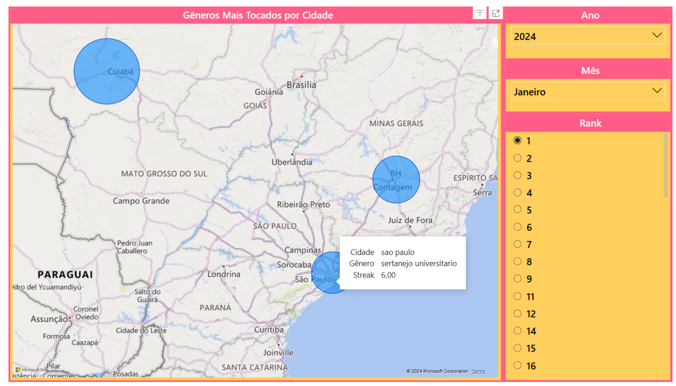

Fig. 4. Power BI map visualization showing the most popular genres by city. The visual presents the cities São Paulo, Cuiabá, and Belo Horizonte. The selected values of the available filters, year, month, and rank, are 2024, January, and 1, respectively. This means that the visual displays the top-ranked music genre in each location for the selected time period. When hovering over São Paulo, the tooltip reveals that the leading genre is 'sertanejo universitário' with a six-week streak, meaning that it has kept the first position in the rankings for six consecutive weeks.

Both visualizations provide clear and objective insights, catering to different levels of geographic segmentation. The national view allows for a broader approach, while the city view is ideal for specific and detailed analyses. In addition, the conversion of the number of listeners to the popularity score, described in Section III-B, has been incorporated into all graphs, allowing the user to have direct information on the popularity of musical styles during navigation.

D. Forecast module

The prediction model used to implement the forecast module was the Extreme Gradient Boost (XGBoost) [21], which employs a gradient boosting technique. It generates and trains decision trees repeatedly, using the previous results to minimize a loss function and fine-tune the model [22].

XGBoost is a scalable machine learning model widely adopted across industry applications for trend prediction tasks, such as forecasting consumer behavior, which is the goal of the proposed solution [21]. Its claimed scalability and high performance stem from a variety of algorithmic and statistical optimizations, including the mechanisms for handling sparse data and capturing seasonal patterns, which are particularly relevant when working with popularity rankings, as done in this application [21].

It is possible to adapt XGBoost depending on the nature of the data and the prediction to be made: while XGBoost Classifier uses loss functions suitable for classification problems that utilize categorical data, XGBoost Regressor uses functions appropriate for continuous values, such as the monthly scores compiled in this project. The construction of the XGBoost Regressor model was carried out using the Python 3.12 language and the *xgboost*, *matplotlib*, *pandas*, *numpy*, and *sklearn* libraries.

Firstly, the data gathered from 2023 and 2024 were imported and pre-processed to eliminate null values that could harm the model. Then, the popularity of music genres in 2023 and 2024 was used to train the model so to capture seasonal variations in music genre popularity. To preserve the simplicity of the model and avoid interference with the baseline results, no hyperparameter tuning was applied during the implementation of the regressor. The default configuration includes a learning rate of 0.3, a maximum tree depth of 6, 100 estimators, and full sampling of both rows and features per tree.

By training the XGBoost Regressor using a rolling window, where each month's data was used to predict the following month, the model was able to reflect temporal dynamics and evolving trends. This approach allowed us to observe seasonality effects in the data while preserving the integrity of baseline results. This process was carried out for the data from Brazil, and for the cities São Paulo and Cuiabá.

Upon reaching a stopping point, defined by the convergence of the loss function, XGBoost provided a set of predicted data for 2025, which was compiled into CSV files to create the visualizations described in Section III-C.

The algorithm uses loss functions suitable for regression tasks, including the use of metrics for error estimating, such as the Mean Squared Error (MSE) and Mean Absolute Error (MAE) [23]. Therefore, to assess the model's reliability, we calculated the MSE and MAE between the popularity predicted by the model and the actual popularity of each musical genre. The correlation coefficient (r) between predicted and actual values was also calculated to quantify the strength and direction of a linear relationship between the predicted values from the model and the actual observed data [24].

To evaluate the accuracy of the model through direct comparison, we configured the model to predict months for which data already existed. In the first setup, the algorithm was trained on data from January 2023 to September 2024 (21 months), and tasked with predicting October and November 2024.

To assess year-long predictive performance, we later adopted a second setup: training the model exclusively on 2023 data to forecast all twelve months of 2024. However, this approach limited the model's ability to capture seasonal trends across different years, as it relied on only one full-year input. As expected, the reported accuracy was lower using a smaller period sample to train the model (which may underestimate the model's true potential).

The MSE for each music genre was calculated using the equation 1, with $n = 12$ months to evaluate the prediction performance. The MSE was calculated by averaging the squared errors across monthly predictions for a given genre. The variables y' and y represent the predicted value and the actual value, respectively.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y'_i - y_i)^2 \tag{1}$$

To assess the reliability of predictions, we defined the MSE threshold based on the universal thresholding formula proposed by Donoho and Johnstone, and referenced by Jansen in his work on wavelet denoising and minimax risk bounds [25]. The formula is given by equation 2, where σ represents the standard deviation of the prediction errors and n is twelve, as explained above.

$$\lambda = \sigma\sqrt{2\log n} \tag{2}$$

Based on the MSE of each genre, we calculated a reliability rate equal to the number of genres with MSE below the threshold (reliable prediction) divided by the total number of genres within the aimed popularity score range. We also calculated the global accuracy of predictions equal to the number of predictions with a squared error below the threshold, and the total number of predictions. The results regarding the reliability and accuracy are described in Section IV-A.

E. Forecast visualization module

This module presents visualizations derived from the prediction model detailed in Section III-D. The predicted data were integrated into the GUI to provide forward-looking insights into music genre popularity trends. Below, we present the forecast visualization pages for the country and for the cities.

- **Popularity Brazil**:
  - **Popularity Forecast Brazil**: This graph, shown in Figure 5, presents the forecast of music genre popularity in Brazil for 2025. The estimate is based on the predictive model detailed in the previous section. The visualization uses a line graph, allowing for a temporal analysis over 12 months. This setup enables the identification of a growth or decline trend in popularity for the selected genre. Users can apply filters to select the music genre of interest, adjusting the analysis according to their needs.

- **Popularity by City (Forecast View)**: In addition to historical data, city-specific pages also include a line graph for the forecast of music genre popularity for the selected city, as shown in Figure 6.

This predictive layer enables users to anticipate regional trends and potential shifts in musical preferences at a local level, complementing the historical and ranking views presented in the main GUI.

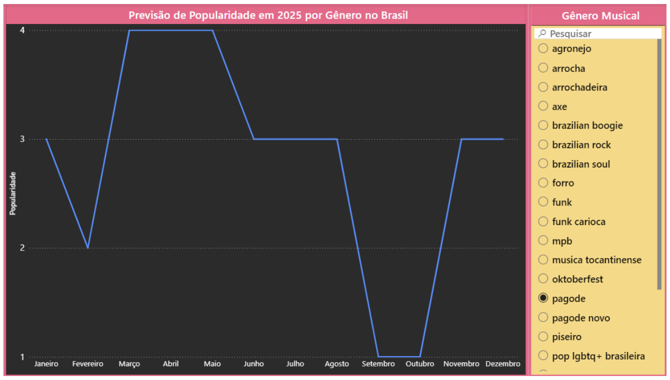

Fig. 5. Power BI visualization with the forecast of the popularity score for a selected musical genre in Brazil for 2025. The selected value for genre in the presented visual is 'pagode'.

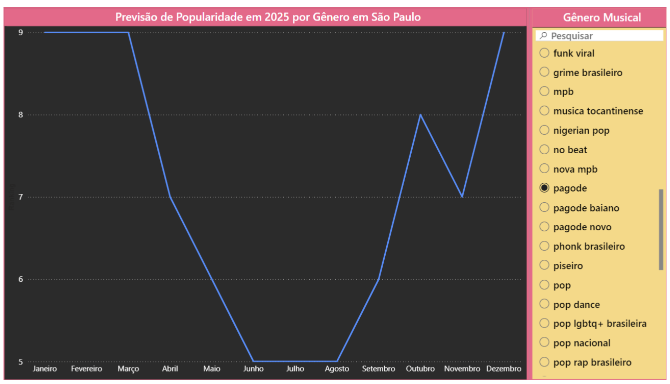

Fig. 6. Power BI visualization with the forecast of the popularity score for a selected musical genre in São Paulo for 2025. The selected value for genre in the presented visual is 'pagode'.

IV. Results

The development of the application was supported by GitHub, and the code implementation is maintained in a separate repository [27]. A public release of this proof-of-concept as a web service is available at GitHub Pages [28].

The information displayed in the music popularity visualizations achieves the purpose of allowing users to assess the popularity of musical genres by location and months of the years covered by the data collection. Similarly, the popularity prediction tool allows users to estimate future popularity levels by genre, location, and time of year.

No extensive end-user testing was conducted at this stage. A validation of the implemented functionalities was performed as a functional proof-of-concept through a demonstration of the application for a group of IT professionals who evaluated the achievement of the implementation goals and the tool's computational performance. The prediction model was evaluated by its accuracy upon testing it with known popularity scores. This is detailed below.

A. Prediction Model

The runtime for the XGBoost model was approximately 16 seconds, and approximately 25 seconds for the overall code, including r for each music genre obtained from the testing procedure described in Section III-D and presented in Table I.

Using the residuals from the forecasts for the entire year of 2024, we estimated the standard deviation $\sigma = 2.31$, and with $n = 12$ months, the resulting threshold is $\lambda = 5.1497$. This value provides a theoretically grounded upper bound for acceptable error (slightly higher than the threshold previously obtained of 2.0 for $n = 2$), and reinforces the reliability of our model with a formal justification.

Table I presents the results from the second setup, in which the model was trained on 2023 data to predict all twelve months of 2024. In this configuration, MSE of predictions for 6 out of 23 music genres exceeded the defined threshold, and 17 genres stayed within the reliable prediction threshold, resulting in a reliability rate of approximately 73.9% (17/23). For comparison, the first setup, in which the model was trained on 21 months of data and predicted for 2 months, achieved a higher confidence rate of 82% (with a smaller threshold: 2.1), highlighting the positive impact of feeding the model with more data. Additionally, the average MSE across all genres in the second setup was 4.8351, which remains below the calculated threshold, reinforcing the overall reliability of the predictions.

The total number of predictions was 276 (12 predictions for 23 genres). Out of these predictions, 207 had an squared error below the defined threshold of $\lambda = 5.1497$, therefore, our overall accuracy was 75%. The databases with the popularity scores, and the reliability and accuracy measures of the predictions, are available in a public repository [29].

V. Discussions

Given the early nature of the current code and the interest in its application with creative industry collaborators, reproducibility checking is not currently fully available as in an open implementation. Power BI has proven to be an effective and fast solution for the integration of visualizations at this stage, and the dashboards have their own links that can be shared publicly, but sharing data access requires a Pro license (which is no longer available).

Popularity and revenue forecasts are found as solutions marketed by some businesses that feed popular charts (e.g. Billboard's) with measurements of digital, retail, and airplay of music. Such information can be an additional basis for orienting campaign promotion for music producers and artists as final users. However, in this stage we were interested in capturing numbers of audience, not considering sales numbers, which imply further investigation and validation requirements.

Additional information may or may not help improve musical popularity prediction accuracy. Araújo et al. (2019) used an SVM classifier with prediction accuracy ranging from 70% to 98%, depending on the use of additional information (such as acoustic features) and the method's baseline. They introduced a method for predicting whether a song will go "viral" and predicting consistent popularity growth thereafter, when the prediction window is a few weeks further out from the training period [30]. Cases of sudden popularity observed in "viral"

TABLE I
MEAN SQUARED ERROR OF THE POPULARITY PREDICTION FOR DIFFERENT MUSIC GENRES FOR 2024

Music Genre	MSE	MAE	r	MSE Threshold Status
agronejo	0.8611	0.7164	0.0533	below
arrocha	12.6637	3.1069	-0.6161	**above**
arrochadeira	1.2318	0.8174	0.2870	below
axe	0.3846	0.5286	—	below
brazilian boogie	4.8222	1.7749	-0.0889	below
brazilian rock	3.6965	1.0733	—	below
brazilian soul	4.8222	1.7749	-0.0889	below
forro	24.6966	4.5740	—	**above**
funk	0.5308	0.3163	—	below
funk carioca	3.5848	1.6326	0.2341	below
mpb	7.1523	2.3578	-0.4137	**above**
musica tocantinense	8.0631	2.5268	-0.1687	**above**
oktoberfest	1.2543	1.1200	—	below
pagode	13.0456	3.0496	-0.0834	**above**
pagode novo	17.2140	3.6325	—	**above**
piseiro	0.9947	0.8918	0.2794	below
pop lgbtq+ brasileira	0.5666	0.5568	-0.0893	below
pop nacional	0.5246	0.5586	0.4604	below
rave funk	2.3706	1.2529	-0.3407	below
sertanejo	0.2818	0.4063	0.3842	below
sertanejo pop	1.6860	1.1101	-0.4752	below
sertanejo tradicional	0.1478	0.3754	—	below
sertanejo universitario	0.6119	0.5050	0.3264	below

lists but not in “most popular” lists (while not addressing the underlying reasons for popularity) may increase prediction accuracy based on previous popularity reports, and deserve further investigation.

Boosting algorithms are reported to exhibit outstanding prediction performance [35], especially dealing with complex data challenges, and we decided to undertake them on this task of predicting music popularity, which is simultaneously influenced by dimensions not easily observable or tractable. Achieved prediction accuracy figures were better or within the range reported by other studies, and we assess the performance reached in the baseline configuration as promising. Aiming at improvements on its predictive power and robustness, it is suggested as a future work to explore hyperparameter optimization techniques.

This case study focused solely on Brazil, although, due to its geographic scope and market relevance, it should not be strictly classified as restricted. A key reason for this was the requirement to use official data from ECAD, which operates exclusively in the country. The 70/30 ratio adopted, with a greater weight given to official ECAD data than to Spotify charts, was an arbitrary initial choice to study the attribution of trust in the data. A study of the variation of this rate on the accuracy of the model is considered a necessary future study, to adjust the model to observable reality.

ECAD’s criteria and process for verifying musical plays are not open but are auditable. The use of the organization’s consolidated data is officially valid in the country, and we assume it adds greater predictive power by taking into account verified performance reports of songs played during a given period on radio/TV and streaming platforms, as well as event itineraries, including musical setlists performed at shows and concerts.

To contextualize similarities with Internet of Musical Things (IoMusT), according to [32], we can raise that the manipulation and visualization of descriptive metadata such as popularity scores, music consumption numbers, genre classifications, and geographic coordinates — particularly data originating from online music databases, music platforms and applications — falls within the context of IoMusT. Nevertheless, in its application scenarios [33], we foresee that the use of music-related resources for artist promotion finds a niche for further exploration.

While distortions resulting from the use of artists’ own genre assignments affect the quantification of their popularity, the tool does not currently aim to quantify this impact or solve the challenging problems and taxonomic difficulties of artist positioning. Future work may address alternatives to mitigate this issue and even optimize artist genre assignment to more stable and accurate labels, for example, using musical descriptors for feature-based genre assignment, external databases and listener behavior patterns. Overcoming artist-based genre classification may help to improve the accuracy of predicting genre popularity. However, it should also be considered that the tool is useful to artists insofar as it addresses the genre(s) to which they believe they belong.

Since we find that many popular genres present predictions beyond the tolerable threshold, this suggests the need to pursue a future threshold calibration policy, possibly tied to practical decision factors.

Model enhancements for the future may also consider adding information from music festivals, seasonal events and social factors as exogenous regressors, and clustering locations, allowing for zooming in/out the forecasting patterns. The highest position on the Billboard charts of a song / album and the number of weeks it has been on the chart could also be used to enhance or to cross-check the consistency of genre popularity.

VI. Conclusions

We presented a proof-of-concept of a data analysis and prediction application for the music industry called *TrendTune*, which provides data visualizations of the popularity of musical genres in Brazil, segmented by time and geographic location. The application provided a gradient boosting prediction model to estimate the popularity of a genre for the following 12 months, which achieved a global accuracy of 75%. A data collection automation was also implemented, allowing for faster compilation of future data to be included in the database supporting the prediction model and visualizations.

A preliminary survey previously conducted with young emerging artists, as a requirements analysis for the design of TrendTune, assisted in the selection and specification of a set of highly rated features for prototype development.

TrendTune has been qualitatively evaluated to date in terms of its feasibility and potential value to musical artists and producers. The application has been pointed as a highly useful resource by consulted professionals in the music market, but it is also recognized that known issues related to music genre labeling and artificial popularity inflation — due to the use of bots in the digital domain and paid promotion in music public performance — must be properly managed with an appropriate algorithm to quantify and compensate for the biases imposed by these practices.

At the end of the current phase of investigation, the project opens more avenues for exploration and revision than a single perspective for a definitive application. The results obtained are promising and encourage the continuity of the project. User testing with groups of artists and music producers is a future goal, which will allow us to validate the features based on case studies, assess their ability to assist artists in their music market planning, and define the scope for a second round of development. The prediction algorithm is also expected to be improved by increasing the training datasets and tuning hyperparameters to enhance the accuracy and robustness of the model.

It is worth noting that no databases were found that provided information on the genre and age group of listeners, with geographical and temporal segmentation. Future works might involve collecting demographic data from listeners, such as gender and age group, as well as popularity information beyond live concerts and digital platforms, which may include festivals and radio rankings, for example. With this information, properly segmented by time and location, it is believed that it will be possible to more faithfully provide an in-depth analysis of musical genres for specific audiences.

References

[1] International Federation of the Phonographic Industry, "Global Music Report 2024: State of the Industry", IFPI, 2024. [Online]. Available: https://www.ifpi.org/ifpi-global-music-report-global-recorded-music-revenues-grew-10-2-in-2023/

[2] T. Ingham, "It's happened: 100,000 tracks are now being uploaded to streaming services like Spotify each day", Music Business Worldwide, 2022. [Online]. Available: https://www.musicbusinessworldwide.com/its-happened-100000-tracks-are-now-being-uploaded/

[3] M. Darvish and M. Bick, "The Role of Digital Technologies in the Music Industry—A Qualitative Trend Analysis", Information Systems Management, vol. 41, no. 2, pp. 181–200, 2023. [Online]. Available: https://doi.org/10.1080/10580530.2023.2225129

[4] N. S. Jung and F. Mayer, "Beyond Beats: A Recipe to Song Popularity? A machine learning approach", arXiv, Mar. 2024. [Online]. Available: https://arxiv.org/abs/2403.12079

[5] Y. Kakkad, P. Savsani, B. Dave, and K. Bhattacharyya, "Predicting Song Popularity Using Machine Learning Algorithms", GitHub, 2021. [Online]. Available: https://github.com/yashrajkakkad/song-popularity-prediction

[6] ECAD, "Ranking ECAD", 2024. [Online]. Available: https://www4.ecad.org.br/ranking/

[7] M. Katz, "Capturing Sound: How Technology Has Changed Music", Rev. ed., University of California Press, 2010.

[8] Spotify, "Spotify Web API Documentation", Spotify for Developers, 2024. [Online]. Available: https://developer.spotify.com/documentation/web-api

[9] M. L. B. Mondelli, L. M. R. Gadelha Jr., and A. Ziviani, "What Countries Listen to: Analyzing the Network of Music Genres around the World", iSys - Brazilian Journal of Information Systems, vol. 12, no. 3, 2019. [Online]. Available: https://journals-sol.sbc.org.br/index.php/isys/article/view/597

[10] B. Moorhouse, "Spotify: An Analysis of Personal and Global Listening Trends, A Data Driven Approach", Medium, Feb. 2021. [Online]. Available: https://brendanmoorhouse4.medium.com/spotify-an-analysis-of-personal-and-global-listening-trends-96f50641a9c6

[11] H. S. Saragih, "Predicting song popularity based on Spotify's audio features: insights from the Indonesian streaming users", J. Manage. Analytics, vol. 10, no. 4, pp. 693–709, Oct. 2023. [Online]. Available: https://ideas.repec.org/a/taf/tjmaxx/v10y2023i4p693-709.html

[12] C. Valcarcel, "Mapping music with every noise and Plotly", Medium, Better Programming, Aug. 2019. [Online]. Available: https://betterprogramming.pub/mapping-music-with-everynoise-9ccb03db4f08

[13] Kinsta, "O que é web scraping?", Kinsta, Aug. 2023. [Online]. Available: https://kinsta.com/pt/base-de-conhecimento/o-que-e-web-scraping/

[14] S. Carlisle, "Software: Tableau and Microsoft Power BI", Technology—Architecture+Design, Taylor & Francis, vol. 2, n. 2, pp. 256—259, 2018.

[15] Spotify, "Spotify Charts Overview", 2024. [Online]. Available: https://charts.spotify.com/charts/overview/global

[16] P. Lamere, "Spotipy: A lightweight Python library for the Spotify Web API", GitHub, 2020. [Online]. Available: https://github.com/plamere/spotipy

[17] Spring Boot, "Spring Boot", Spring.io. [Online]. Available: https://spring.io/projects/spring-boot

[18] Apache Maven, "Apache Maven Project", The Apache Software Foundation. [Online]. Available: https://maven.apache.org/

[19] Amazon Web Services, "What is a Bot?", AWS. [Online]. Available: https://aws.amazon.com/what-is/bot/

[20] Bootstrap, "Bootstrap: The most popular HTML, CSS, and JS library in the world", 2025. [Online]. Available: https://getbootstrap.com/

[21] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system", in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, San Francisco, CA, USA, 2016, pp. 785–794. doi:10.1145/2939672.2939785

[22] J. Brownlee, "A gentle introduction to the gradient boosting algorithm for machine learning", 2020. [Online]. Available: https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/

[23] J. Brownlee, "A gentle introduction to XGBoost loss functions", 2021. [Online]. Available: https://machinelearningmastery.com/xgboost-loss-functions/

[24] J. Li, "Assessing the accuracy of predictive models for numerical data: Not r nor r², why not? Then what?", *PLOS ONE*, vol. 12, no. 8, pp. e0183250, 2017. [Online]. Available: https://doi.org/10.1371/journal.pone.0183250

[25] M. Jansen, "Estimating the minimum MSE threshold", in Noise Reduction by Wavelet Thresholding, New York, NY, USA: Springer, 2001, pp. 81–100. doi:10.1007/978-1-4613-0145-5_4

[26] S. Beheshti, M. Hashemi, E. Sejdic, and T. Chau, "Mean square error estimation in thresholding", IEEE Signal Process. Lett., vol. 18, pp. 103–106, 2011. doi:10.1109/LSP.2010.2097590

[27] L. M. Vieira and T. R. Fujissaka, "TrendTune user interface", GitHub, 2024. [Online]. Available: https://github.com/thais-fujissaka/TrendTune

[28] T. R. Fujissaka, "TrendTune website", GitHub Pages, 2024. [Online]. Available: https://thais-fujissaka.github.io/TrendTune/

[29] L. T. Alvim, "TrendTune", GitHub Pages, 2024. [Online]. Available: https://github.com/LarissaTerto/TrendTune

[30] C. Araujo, M. Cristo, and R. Giusti. "Predicting Music Popularity on Streaming Platforms", in Proceedings of the 17th Brazilian Symposium on Computer Music, São João del-Rei, 2019, pp. 141-148, doi: https://doi.org/10.5753/sbcm.2019.10436

[31] T. Bandyopadhyay, A. B. Woszczynski, D. Crowell. "Musical attributes, cultural dimensions, social media: Insights for marketing music to consumers", Journal of Retailing and Consumer Services, vol. 82, 2025, doi.org/10.1016/j.jretconser.2024.104103

[32] P. Lončar, "Internet of Musical Things and Music Data Visualization," 2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 2022, pp. 1501-1506, doi: 10.23919/MIPRO55190.2022.9803404.

[33] L. Turchet, C. Fischione, G. Essl, D. Keller and M. Barthet, "Internet of Musical Things: Vision and Challenges", IEEE Access, vol. 6, pp. 61994–62017, 2018.

[34] J. Lee and J. -S. Lee, "Music Popularity: Metrics, Characteristics, and Audio-Based Prediction," in IEEE Transactions on Multimedia, vol. 20, no. 11, pp. 3173-3182, Nov. 2018, doi: 10.1109/TMM.2018.2820903.

[35] C. Zhang, Y. Zhang, X. Shi et al. "On Incremental Learning for Gradient Boosting Decision Trees". Neural Process Lett. 50, pp. 957–987, 2019. https://doi.org/10.1007/s11063-019-09999-3