UNIVERSIDADE CATÓLICA PORTUGUESA

Data Analytics in Budgeting for Cultural Institutions

The Case Study of Casa da Música

Luís Maria Cardoso Ayres Gagliardini Graça

Católica Porto Business School
2025

UNIVERSIDADE CATÓLICA PORTUGUESA

Data Analytics in Budgeting for Cultural Institutions

The Case Study of Casa da Música

Master's Final Assignment – Project
Presented to *Universidade Católica Portuguesa*
to obtain a Master's Degree in Management
with specialisation in Business Analytics

by

Luís Maria Cardoso Ayres Gagliardini Graça

Under the supervision of
Vera Lúcia Miguéis Oliveira e Silva, PhD

Católica Porto Business School
April 2025

Agradecimentos

Obrigado, Meu Pai, Leonor, Pe. Jorge e Pais.

Obrigado a quem rezou por mim.

Obrigado, Professora Vera, pela orientação, apoio e, acima de tudo, estímulo.

Obrigado, Casa da Música, Isabel, Rita, Marta e Filipe, pelo desafio, envolvimento e acompanhamento.

Obrigado.

Abstract

This study explores the application of data analytics to support budgeting and management control processes within cultural institutions, through the case study of Fundação Casa da Música. By adopting the CRISP-DM methodology, the project developed a twofold solution: (i) a descriptive model to automate classification and reporting of accounting transactions, and (ii) a predictive model to forecast monthly financial performance. The automation reduced the time spent on transaction classification by over 98%, significantly enhancing efficiency and scalability. Additionally, statistical forecasting techniques—namely Holt-Winters Exponential Smoothing—were applied to project future revenues, expenses, and earnings. The outputs were integrated into a Power BI environment, offering dynamic, interactive dashboards that supported financial oversight and strategic planning. Results indicate that data analytics can not only streamline reporting workflows but also foster proactive decision-making in the cultural sector. The findings highlight the transformative potential of analytical tools in bridging operational efficiency with institutional sustainability.

Keywords: Data Analytics, Business Analytics, Cultural Institutions, Budgeting, Management Control, Power BI, Forecasting, Casa da Música

Word count (from the Table of Contents to Chapter 5):
9.021 words

Resumo

Este estudo explora a aplicação de *data analytics* para apoiar os processos de orçamentação e controlo de gestão em instituições culturais, através do estudo de caso da Fundação Casa da Música. Recorrendo à metodologia CRISP-DM, o projeto desenvolveu uma solução com duas vertentes: (i) um modelo descritivo para automatizar a classificação e o reporte de transações contabilísticas; e (ii) um modelo preditivo para prever o desempenho financeiro mensal. A automatização permitiu reduzir em mais de 98% o tempo despendido na classificação das transações, aumentando significativamente a eficiência e a escalabilidade do processo. Adicionalmente, foram aplicadas técnicas estatísticas de previsão — nomeadamente o método de Holt-Winters — para projetar receitas, despesas e resultados. Os outputs foram integrados num ambiente Power BI, proporcionando *dashboards* dinâmicos e interativos que apoiam o acompanhamento financeiro e o planeamento estratégico. Os resultados obtidos demonstram que o uso de *data analytics* pode não só agilizar os fluxos de trabalho de reporte, como também promover uma tomada de decisão mais proativa no setor cultural. As conclusões evidenciam o potencial transformador das ferramentas analíticas na articulação entre eficiência operacional e sustentabilidade institucional.

Palavras-chave: Data Analytics, Business Analytics, Instituições Culturais, Orçamentação, Controlo de Gestão, Power BI, Forecasting, Casa da Música

Contagem de palavras (do Índice ao Capítulo 5):
9.021 palavras

Table of Contents

Table of Figures

Table of Tables

Introduction

This study was proposed as an opportunity for the author to explore how data analytics can be applied to strengthen budgeting and management control practices within cultural institutions, with Casa da Música serving as a practical and illustrative case.

The project focuses on Fundação Casa da Música, a Portuguese cultural institution renowned for its commitment to musical excellence and public service. Despite having implemented a modern ERP system, its Management Control department still relied heavily on manual classification of transactions and static reporting formats, which introduced inefficiencies, operational risks, and scalability challenges.

Leveraging the CRISP-DM methodology, this study aimed to design and implement a data-driven solution that addresses these issues in two complementary ways: through automation of transaction classification and through predictive modeling of financial trends. The solution was operationalized using Python and Power BI, integrating statistical models and business rules into a coherent and user-friendly analytical environment.

Given this context, the structure of the document is organized into six chapters. Chapter 1 presents a literature review on the application of data analytics in the budgeting processes of cultural institutions. Chapter 2 introduces Fundação Casa da Música and outlines the main challenges that motivated the project. Chapter 3 details the methodological framework adopted —CRISP-DM— and describes the applied research strategy. Chapter 4 presents the results and discusses the implications of the proposed solution. Chapter 5 offers final reflections and recommendations for future research.

Finally, all the project deliverables, including dashboards, documentation, and questionnaire, were produced in Portuguese to ensure effective communication with the partner institution. Additionally, all the values presented throughout this work are fictitious, in order to preserve the confidentiality agreed upon under the terms of the signed Non-Disclosure Agreement (NDA).

Data Analytics in Budgeting for Cultural Institutions: The Case Study of Casa da Música

1. The Importance of Data Analytics for Cultural Institutions

1.1. Introduction and Literature Review

The integration of data analytics into budgeting processes has become increasingly vital for enhancing operational efficiency, strategic planning, and financial health in various sectors, including cultural institutions. This literature review examines the application of data analytics in budgeting for cultural institutions, to answer the case study of Casa da Música. The aim is to explore how data analytics can transform budgeting processes, address challenges, and foster innovation in cultural management. To achieve this, a thorough exploration of existing literature on these topics is essential.

For example, in the field of predictive analytics, by analyzing historical data, institutions can better predict cash flows, identify potential financial risks, and make informed decisions regarding resource allocation. This approach not only improves financial planning but also supports continuous performance management by enabling institutions to adjust their strategies in response to real-time data (Pushkarna & Walcott, 2023). Predictive models help in anticipating donor behaviors and funding cycles, which are critical for the sustainability of cultural institutions (Wnuczak, 2014).

Furthermore, the implementation of Business Intelligence (BI) systems in budgeting processes allows for comprehensive data management and analysis. BI systems integrate various data sources, perform complex analyses, and

generate actionable insights (Correia et al., 2021). The use of data warehouses, OLAP (Online Analytical Processing), and data mining techniques enables institutions to create detailed financial models and forecasts (Wang, 2016). Such capabilities are critical for institutions like Casa da Música, which must navigate fluctuating funding sources and diverse revenue streams.

Having these ideas in mind, Data Analytics facilitates a shift from intuitive to evidence-based decision-making in cultural institutions. This transition is particularly important in budgeting, where accurate and timely data can significantly impact financial outcomes, especially in institutions that have a need to allocate resources (public and private) in an efficient manner (Gałecka & Smolny, 2017). By leveraging data analytics, cultural managers can develop more effective budgeting strategies, optimize resource allocation, and improve overall financial performance. Real-time data analytics provide insights into current spending patterns and highlight areas where cost savings can be achieved without compromising the quality of cultural offerings (Pushkarna & Walcott, 2023).

A relatively new way to manage businesses, rolling forecasting - a data-driven approach - allows cultural institutions to extend their planning horizons and respond agilely to changing market conditions (Miller et al., 2013). Unlike traditional annual budgets, rolling forecasts continuously update financial plans based on the latest data, providing a more dynamic and responsive approach to financial management (Miller et al., 2013). This method is especially beneficial for Casa da Música, enabling it to align its financial strategies with long-term organisational goals (Wang, 2016).

Data Analytics can also play a significant role on non-financial departments. For example, it enhances the ability of cultural institutions to understand and engage with their audiences. By analysing audience data, institutions can tailor their offerings to meet the preferences and expectations of different segments

(Nuccio & Bertacchini, 2022). This not only improves customer satisfaction but also drives higher attendance and revenue. Data-driven insights enable institutions to measure the social and economic impacts of their activities, thereby demonstrating their value to funders and stakeholders (Nuccio & Bertacchini, 2022). For example, Casa da Música can use audience analytics to identify trends in ticket sales and customize marketing campaigns to boost attendance during off-peak seasons. The adoption of data analytics fosters innovation in business models for cultural institutions. Data-driven strategies can help institutions like Casa da Música develop new revenue streams through targeted marketing, personalized experiences, and enhanced membership programs (Nuccio & Bertacchini, 2022). Additionally, data analytics can support the creation of sustainable business models that balance cultural missions with financial viability. Innovations such as dynamic pricing models, where ticket prices fluctuate based on demand, can also be explored to maximize revenue and accessibility (Nuccio & Bertacchini, 2022).

Despite the clear benefits, the adoption of data analytics in cultural institutions faces several challenges. One major barrier is the financial investment required for implementing and maintaining advanced data analytics systems. Many institutions struggle with limited budgets and competing priorities, making it difficult to allocate resources for such initiatives (Nuccio & Bertacchini, 2022). Technological and technical barriers, including the need for specialised skills and infrastructure, pose additional challenges (Rafi et al., 2020).

Cultural and organisational resistance is another significant hurdle. Many cultural institutions have entrenched practices and a reluctance to adopt new technologies (Nuccio & Bertacchini, 2022). This resistance can stem from a fear of losing the human touch in arts management or a lack of understanding of the potential benefits of data analytics. Overcoming this resistance requires a concerted effort to build a data-driven culture and demonstrate the tangible

benefits of data analytics in enhancing operational efficiency and financial health (Nuccio & Bertacchini, 2022). Training programs and workshops can help staff develop the necessary skills and confidence to develop the budgeting process, utilising data analytics effectively (Rafi et al., 2020).

Wrapping up, the integration of data analytics into budgeting processes offers substantial benefits for cultural institutions like Casa da Música. By enhancing financial forecasting, supporting strategic planning, and fostering innovation, data analytics can significantly improve the financial health and operational efficiency of these institutions (Wnuczak, 2014). However, addressing the financial, technological, and cultural barriers is crucial for the successful adoption of data analytics. As cultural institutions continue to navigate the complexities of modern financial management, the strategic use of data analytics will undoubtedly play a pivotal role in their sustainability and growth.

In this context, the case of Casa da Música illustrates the transformative potential of data analytics in budgeting for cultural institutions. By embracing data-driven strategies, cultural managers can not only optimize their financial performance but also enhance their ability to fulfill their cultural missions and engage with their communities (Nuccio & Bertacchini, 2022). This comprehensive approach ensures that cultural institutions remain vibrant and sustainable in an increasingly data-driven world.

1.2. Gap Identification and Defining the Research Question(s)

Despite the growing recognition of data analytics as a transformative tool in various sectors, its adoption in the budgeting processes of cultural institutions remains limited and underexplored (Wnuczak, 2014). Budgeting methods which rely on historical data and static annual budgets remain the traditional approach and often fail to address the dynamic and unpredictable nature of revenue and

costs in cultural institutions (Wnuczak, 2014). There is a significant gap in understanding how data analytics can be effectively integrated into the financial management practices of these institutions to enhance accuracy, agility, and strategic decision-making. The case of Casa da Música provides a unique opportunity to explore these aspects in depth, as there is a lack of comprehensive studies that examine the specific benefits, challenges, and outcomes associated with data analytics in the budgeting processes of cultural institutions.

Accordingly, the central research question guiding this study is:

How can data analytics be used to support the budgeting and management control processes of cultural institutions, specifically in the context of Casa da Música?

To answer this question, three sub-questions are posed:

1. *In what ways can data-driven decision-making affect strategic planning and resource allocation?*
2. *What challenges and barriers has Casa da Música encountered in implementing data analytics, and how have they been addressed?*
3. *What are the tangible benefits observed from the integration of data analytics in terms of financial performance and operational efficiency?*

To conclude, the motivation behind this research is rooted in the need for more dynamic and responsive financial management practices within cultural institutions. Casa da Música, like many cultural organizations, operates in an environment where funding sources are unpredictable, and financial sustainability is a constant challenge, and traditional budgeting methods are often insufficient to navigate these complexities effectively.

2. The Case Study of Casa da Música

As previously mentioned, this study will use the case study of Casa da Música as a practical example. Accordingly, it will be necessary to contextualize Fundação Casa da Música as a cultural institution and to identify the specific challenge that justifies its relevance for analysis within the field of Business Analytics.

2.1. Fundação Casa da Música

Fundação Casa da Música is a cultural institution governed by private law and recognized as serving the public interest. It was established by the Portuguese State and the Municipality of Porto, with additional support from Private Founders. Its governance is defined by the statutes outlined in Decreto-Lei No. 18/2006 of January 26 (Fundação Casa da Música, 2024b).

Conceived to commemorate Porto's designation as the European Capital of Culture in 2001, Casa da Música stands as the first building in Portugal designed exclusively for music. It serves as a space dedicated both to public performances and enjoyment, as well as to artistic education and musical creation (Fundação Casa da Música, 2024a).

Casa da Música project was conceived in 1999 following an international architecture competition, which selected the design proposed by Rem Koolhaas and the Office for Metropolitan Architecture. Construction began the same year on the site of Porto's former Remise, located at Rotunda da Boavista. The building was officially inaugurated on April 15, 2005, during the spring season (Fundação Casa da Música, 2024a).

2.2. Identifying the Challenge

Fundação Casa da Música implemented a renewed ERP system in 2020, featuring clean and up-to-date data where all financial and accounting transactions are recorded. The Management Control team manually classifies transactions into categories and subcategories based on specific characteristics. Once all transactions are classified, the department manually prepares its tables, transferring them to the monthly budget execution report.

The institution's objective is to automate this process and develop methods to forecast the performance of future months, dividing this study's scope of application into a descriptive component and a predictive component, respectively.

To address this issue, it is essential to consider the maturity level in adopting Business Analytics systems and practices, in Fundação Casa da Música. A solution must be designed to encompass the entire data lifecycle, employing a consistent and robust methodology that enables medium-term results while also delivering quick wins.

The Case Study of Casa da Música can thus serve as an example of how Business Analytics can positively impact operational efficiency and reporting effectiveness in cultural institutions.

In the following chapter, the research approach and methods used to address this challenge will be explored in greater detail.

3. Research Approach and Methods

The complexity of data-driven decision-making highlighted the need for structured methodologies to ensure the reliability and replicability of data mining projects. Among the various process models proposed, the CRISP-DM (Cross Industry Standard Process for Data Mining) has emerged as a de facto standard due to its adaptability across industries and its structured yet flexible approach (Wirth & Hipp, 2000). Originally developed by a consortium of industry leaders, in 1999, CRISP-DM provides a systematic framework that guides the entire lifecycle of a data mining project (Wirth & Hipp, 2000). It is composed of six iterative phases that will be explored in the following subchapters: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment (Wirth & Hipp, 2000).

Despite its widespread adoption, recent literature suggests that while CRISP-DM remains highly relevant, many projects neglect the deployment phase, which is crucial for translating analytical insights into actionable strategies (Schröer et al., 2021). Nevertheless, its structured methodology continues to offer a robust foundation for research and industry applications, making it an appropriate choice for guiding case studies in diverse domains (Schröer et al., 2021).

In this chapter, CRISP-DM will serve as the guiding framework for conducting the Case Study of Casa da Música, ensuring a structured approach to data collection, processing, and analysis. By adhering to this well-established methodology, this study aims to provide reproducible and reliable insights, while addressing potential challenges associated with real-world data mining applications.

3.1. Business Understanding

The Business Understanding phase is the first and foundational step of the CRISP-DM methodology. It focuses on defining the objectives of the project from a business perspective and translating them into a data mining problem (Wirth & Hipp, 2000). This phase involves identifying key stakeholders, understanding business constraints, and determining the success criteria for the project (Wirth & Hipp, 2000). By establishing a clear connection between business needs and analytical goals, this phase ensures that the data mining process remains aligned with the broader organizational objectives (Wirth & Hipp, 2000). Additionally, a preliminary project plan can be developed, outlining the necessary resources, assumptions, and potential risks involved (Schröer et al., 2021).

In the Case Study of Casa da Música, the Business Understanding phase was conducted through regular meetings with the Management Control team to gain insights into the main activities involved in the preparation of monthly reports for monitoring budget execution. These key tasks consist of manually extracting accounting transactions from the ERP system into Excel, followed by a line-by-line manual classification into subcategories, which are then consolidated into broader categories and, ultimately, into aggregators - levels created by the Management Control team for reporting to the administration.

After the classification and review process, Pivot Tables for each reporting level are updated and then copied as images into PowerPoint, which is later exported as a PDF and printed for distribution, in both digital and physical formats.

In general, business process modelling can play a key role in optimizing and automating workflows. Business Process Model and Notation (BPMN) diagrams provide a clear, standardized way to represent operations, making it easier for teams to communicate and improve efficiency (Object Management Group, 2013). By mapping out processes, organizations can identify inefficiencies and

opportunities for improvement, highlighting bottlenecks and automation gaps that BPM systems and workflow engines can help address (Object Management Group, 2013). BPMN is also valuable for assessing operational risks, as it offers a visual representation of dependencies, critical paths, and potential failure points within a process (Object Management Group, 2013).

Accordingly, this methodology was used to produce an *As-Is* Process Flowchart which, combined with the Data Understanding phase, enabled the development of an initial *To-Be* process design, also in flowchart form. This preliminary mapping laid the foundation for subsequent stages of the project. A more detailed analysis of these diagrams will be presented in the *Results and Implications* chapter.

The following subchapter explores the approach to understanding the available data in order to develop the most suitable solution for executing the previously mentioned process.

3.2. Data Understanding

The Data Understanding phase is the second step of the CRISP-DM methodology and plays a crucial role in ensuring the quality and relevance of the data used for analysis (Wirth & Hipp, 2000). This phase begins with the collection of initial data from various sources, followed by an exploratory analysis to assess its structure, completeness, and potential limitations (Wirth & Hipp, 2000). Key activities include data description, data exploration, and quality verification, which help identify inconsistencies, missing values, or biases that could impact the results. Through statistical summaries and visualization techniques, analysts gain preliminary insights into the data, enabling them to refine the problem definition and adjust the project approach if necessary (Schröer et al., 2021). By thoroughly understanding the dataset, this phase lays the foundation for effective data preparation and modeling.

In the Case Study of Casa da Música, it was essential to understand the format of the data exported from the ERP, the main rules for classifying transactions into subcategories, and the implications of modifying existing files or creating new working files.

In 2024, the file *Conciliacao2024.xlsx* was used to process ERP-extracted data and convert it into reportable results. This file comprises 28 sheets: one containing raw data, another integrating the data into subcategories to feed 10 sheets with embedded Pivot Tables, and 16 additional sheets with supporting information.

The file is heavily reliant on formulas and Pivot Tables, making it complex to track and modify. With a size of approximately 60MB, it is inefficient to use collaboratively or to distribute tasks. Moreover, navigating the file requires extensive tacit knowledge, accumulated over time, which is difficult to transfer to other team members.

These limitations introduce significant operational risks, dependencies, and inefficiencies. The manual process of importing and classifying transactions—assisted by formula-based hints in auxiliary columns—is not easily replicable by others. Consequently, it became essential to identify the critical information and dependencies required for effective Data Preparation.

Prior to the reporting phase, the most critical task was the classification of accounting transactions into management control subcategories. This step represented the most resource-intensive component in terms of both computational memory and manual effort. It was performed within a table functioning as a dimensional database, with an order of magnitude between 100,000 and 200,000 rows and 35 columns. Of these, 12 contained lookup formulas, 3 required manual input, and 20 were sourced directly from the ERP system.

Once all transactions were classified, the subsequent tasks primarily involved presenting the data through aggregation and disaggregation across various reporting levels, followed by analysis, evaluation, and commentary.

3.3. Data Preparation

The Data Preparation phase is a critical step in the CRISP-DM methodology, as it involves transforming raw data into a structured format suitable for analysis. This phase includes data selection, cleaning, transformation, integration, and formatting, ensuring that the dataset is optimized for the modeling stage (Wirth & Hipp, 2000). Handling missing values, removing inconsistencies, and engineering relevant features are key activities that improve data quality and enhance model performance. Given that data-related issues often arise at this stage, an iterative approach is typically required to refine the dataset and address unexpected challenges (Schröer et al., 2021). Effective data preparation is essential for building reliable models and extracting meaningful insights from the analysis.

Several iterative meetings were held with Casa da Música's Management Control team to thoroughly understand the dataset and refine it to better align with the team's needs.

As outlined in Chapter 3.2, this study consists of two main components: a descriptive analysis of budget execution and a predictive analysis for budget forecasting.

For the descriptive analysis, initial meetings were conducted to gather information and define the classification rules for assigning transactions to subcategories. It was determined that five key attributes—Account, Group, Nature, Event, and Area—when combined in a specific way, segment transactions into distinct categories.

To systematize this process, an inventory of classification rules was developed, serving as a mapping framework for transaction categorization, as illustrated in **Table 1**.

ID_Regra	ID_Conta	ID_Grupo	ID_Natureza	ID_Evento	ID_Area	ID_Subrubrica
1	721	2	51	*null*	*null*	19
2	721	4	51	*null*	*null*	19
3	721	7	51	*null*	*null*	19
…	…	…	…	…	…	…
1050	62	*null*	*null*	2512-267	23	50

Table 1: Example of the created rules inventory

In this mapping, the completed fields are restrictive, whereas *null* fields can take on any value from the transaction data. Additionally, the values in ID_Conta on the mapping side must match the beginning of the corresponding values in the transaction data. For example, account 61 in the mapping should correspond to 612 in the transaction data, rather than only transactions with an exact 61 match. All other mappings follow a direct correspondence.

Appendix *1* presents the Python code developed to automate this classification process. This automation was built iteratively over several hours through collaboration between the study's author and the Management Control team.

As shown, the code performs a column-by-column and row-by-row mapping between the rules table and the transactions table. A rule is added to the regras_possiveis object only if all columns in a given row meet the required conditions; a single mismatch is enough to exclude it. Once all applicable rules have been identified, the most specific one is selected, following the principle that the rule with the fewest empty columns in the rules table takes precedence.

An alternative approach was considered, involving the creation of an intermediate mapping matrix to filter applicable rules using vectorized operations. However, implementing this would have required a *cross-join* operation, which—given the scale of 1,053 rules and 150,000 transactions—proved significantly slower and more computationally demanding than the iterative approach used.

For the predictive analysis, data processing was more straightforward, as it only required extracting the total monthly values for Expenses, Revenue, and EBT over the past four years. In summary, it was necessary to retrieve the values of all transactions in accounts starting with 6 (expenses) and 7 (revenue) for each month from January 2021 to December 2024. As a validation step, the annual totals were compared against the corresponding Financial Statements, confirming the accuracy of the extracted values. The data in this extraction enabled the development of predictive models for Casa da Música's monthly budget forecasting, as explained in the next subchapter.

3.4. Modelling

The Modelling phase of the CRISP-DM methodology involves selecting and applying appropriate data mining techniques to extract patterns and insights from the prepared dataset. This step requires choosing the most suitable algorithms, defining parameter settings, and building models that address the business objectives outlined in earlier phases (Wirth & Hipp, 2000). Depending on the nature of the problem—such as classification, regression, clustering, or association rule mining—different modeling approaches may be explored and compared. Since some techniques require specific data formats, there is often a feedback loop between this phase and Data Preparation, where adjustments may be necessary to improve model performance (Schröer et al., 2021). The ultimate goal of this phase is to develop robust and interpretable models that effectively capture relationships within the data, serving as the foundation for evaluation and decision-making.

With this objective in mind, the data model for the descriptive analysis was first developed in Power BI, as shown in **Figure *1***.

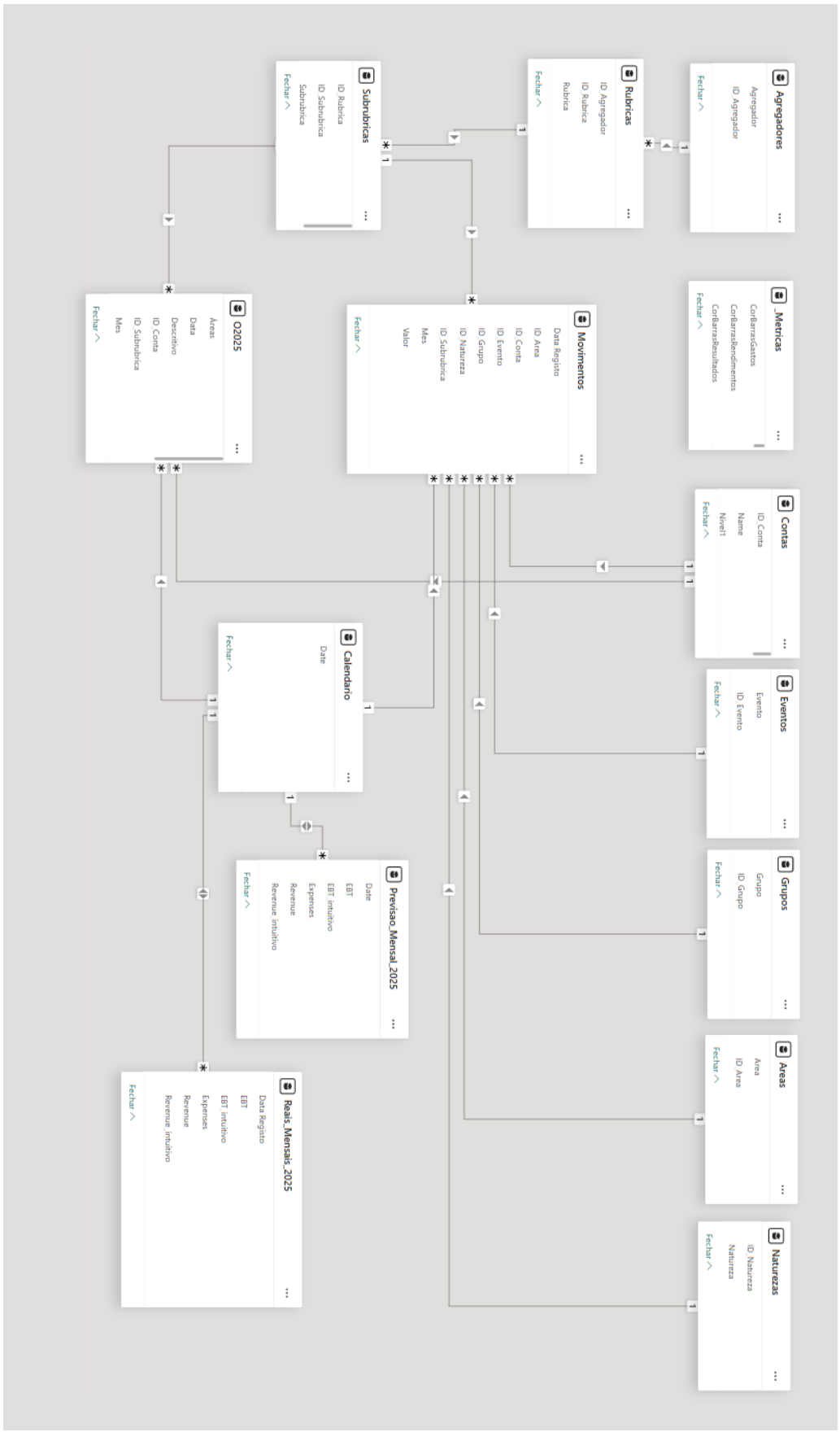


Figure 1: Data Model for the Management Control reporting process

This data model consists of several interconnected tables, organized in a hierarchical structure, illustrated in **Appendix 2**, and explained in the *Results and Implications* chapter. The metrics used are also represented in **Appendix 2**.

Regarding the forecasting component, historical data from 2021 to 2024 was used to generate monthly performance forecasts for Casa da Música using the Holt-Winters Exponential Smoothing method (Kalekar, 2004). In fact, Exponential Smoothing is particularly suitable for short-term forecasting with time series that can exhibit trend and seasonality (Kalekar, 2004), making it an appropriate choice for the Case Study of Casa da Música.

Using the Python script shown in **Appendix 3** , the RMSE for different variations of the Holt-Winters Exponential Smoothing models (Kalekar, 2004) were calculated, to assess their forecasting accuracy.

The following subchapter will explain the evaluation of the models and the selection of the most accurate predictive model.

With both the descriptive and predictive models in place, three Power BI dashboards were developed as prototypes for further analysis and team evaluation.

3.5. Evaluation

The Evaluation phase in the CRISP-DM methodology is essential for assessing the performance and validity of the models developed in the Modelling phase. This step ensures that the selected models align with business objectives and provide meaningful, actionable insights (Wirth & Hipp, 2000). Qualitative assessment, including business validation and expert feedback, plays a crucial role in determining the model's practicality and reliability (Schröer et al., 2021). If the models do not meet the expected performance criteria, iterations with further data refinement or alternative modeling approaches may be necessary.

Ultimately, this phase facilitates informed decision-making on whether the model is suitable for deployment or requires further adjustments.

In the domain of predictive modelling, evaluation typically involves quantitative metrics, such as accuracy, precision, recall, and F1-score for classification tasks, or RMSE and R^2 for regression models. In this particular study, for the predictive component, the model with the lowest RMSE was selected. Once the model was selected, monthly forecasts for 2025 were generated, as presented in the *Results and Implications* chapter.

For the descriptive component, validation was required only from the Management Control team to ensure that the data model was appropriate and that the necessary data processing steps had been correctly implemented. After addressing minor adjustments, the model was finalized.

3.6. Deployment

The Deployment phase is the final step of the CRISP-DM methodology, where the validated model is integrated into a real-world environment to generate actionable insights. Deployment can take various forms, ranging from generating reports and visualizations to fully automating model predictions within an operational system (Wirth & Hipp, 2000). This phase also includes planning for monitoring and maintenance, ensuring that the model remains accurate and relevant over time. Despite its importance, recent literature indicates that many data mining projects do not fully implement this phase, limiting the practical impact of analytical findings (Schröer et al., 2021). A well-executed deployment strategy ensures that the insights derived from data mining contribute effectively to decision-making and business objectives.

In the Case Study of Casa da Música, the Deployment phase primarily involved integrating the automation and process changes into the Management Control team's online OneDrive folders. Additionally, it included explaining the

new steps to the Management Control team for generating the monthly budget execution reports.

The published Power BI report and the developed automation ran in parallel with the previous process for the March 2025 period, reported in April. This approach allowed the Management Control team to gain confidence in the results before fully transitioning to the new process in subsequent periods.

To enable a more quantitative assessment of this Deployment phase, a questionnaire was created for the four main points of contact within the Management Control and IT teams. The questionnaire is available in **Appendix 4**.

4. Results and Implications

This chapter outlines the main results obtained throughout the project, following a logical sequence from process understanding to solution evaluation and implementation.

It begins by presenting the business processes mapped in the initial phases of the project, which served as the foundation for identifying inefficiencies and defining the target "To-Be" processes. This mapping exercise was crucial to ensure that the analytical solution responded to real operational needs.

Then, it describes the data preparation and modelling efforts, detailing the steps taken to clean, transform, and structure the data in a way that supported the development of relevant analyses.

Following this, the dashboards created in Power BI are presented. These visual tools were designed to faster the reporting process and to facilitate decision-making, by providing clear and timely insights to the organization's key stakeholders.

Finally, the results of the questionnaire are analyzed to assess the perceived usefulness, clarity, and usability of the developed solution. This feedback is essential to understand the practical impact of the project and to identify opportunities for refinement and further development.

4.1. Business and Data Understanding

To apply the principles outlined in 3.1. Business Understanding and 3.2. Data Understanding to the Case Study of Casa da Música, a high-level flowchart of the As-Is Reporting Process was systematically mapped, as illustrated in **Figure 2**.

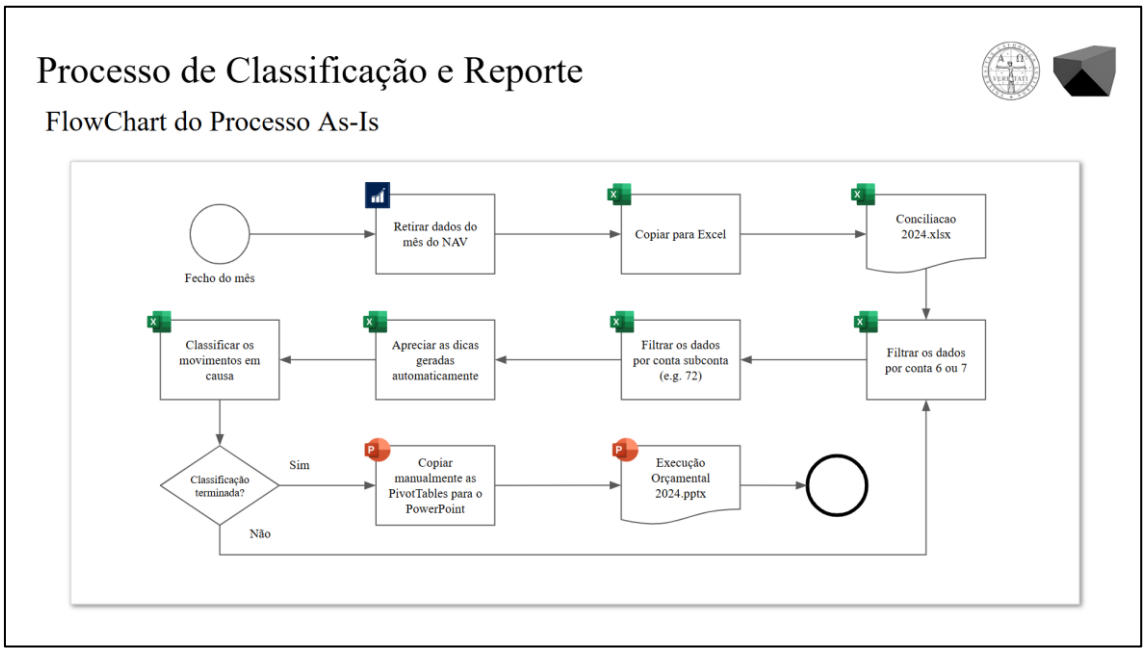


Figure 2: Flowchart of the Management Control reporting process (*as-is*)

This process involves extracting monthly data from the NAV system, transferring it manually into Excel files, and performing a series of repetitive filtering and classification tasks. Although partially supported by automated suggestions, the classification step requires significant manual input and expert judgment. Once completed, the data is manually copied into PowerPoint presentations to report on budget execution.

This workflow is highly dependent on individual effort, spread across multiple tools (Excel, PowerPoint), and lacks integration or centralized validation mechanisms. The absence of automation introduces several limitations: it increases the risk of human error, reduces consistency, and limits the scalability and agility of the reporting process.

After confirmation of the Management Control team, it was possible to develop the flowchart in **Figure 3**, to map the main proposed enhancements. These

improvements mitigate the risk of operational errors and accelerate execution, freeing up resources - both time and availability - that can be redirected toward result analysis rather than mere data preparation.

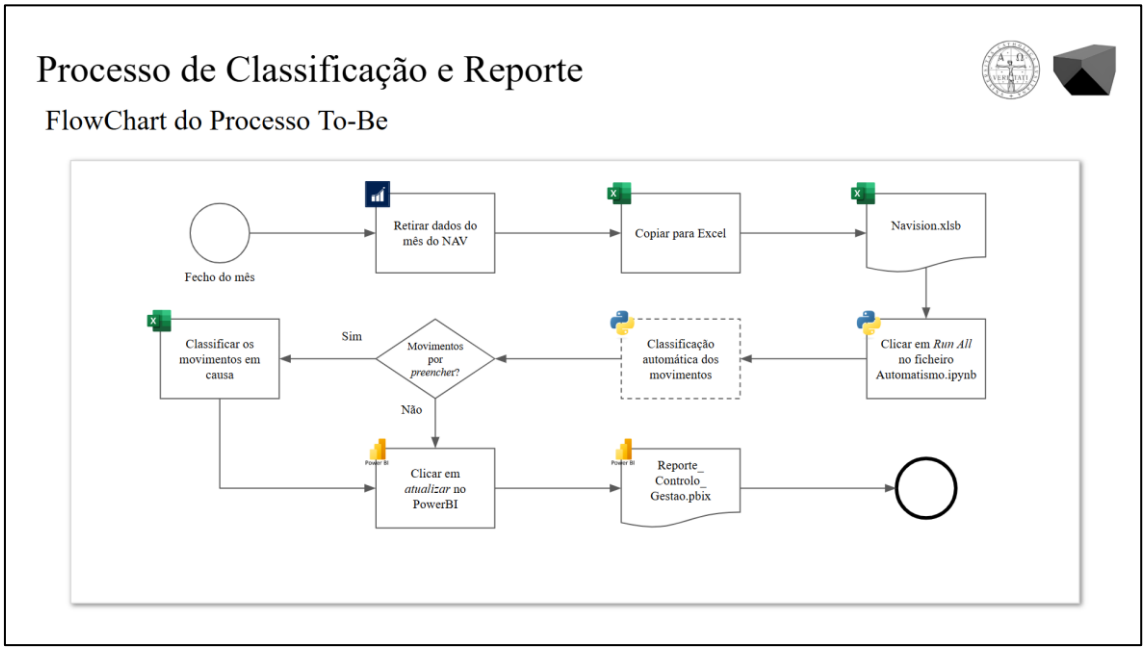


Figure 3: Flowchart with the efficiency benefits of Data Analytics for the reporting process

The transition from the *As-Is* to the *To-Be* process reflects a substantial improvement in efficiency, reliability, and overall usability of the reporting workflow. In the original *As-Is* process, most tasks were manual and fragmented across multiple tools, requiring users to extract data from NAV, manipulate it in Excel through filters and pivot tables, and finally transfer the results into PowerPoint for presentation. This workflow was not only time-consuming, but also prone to human error, heavily reliant on implicit knowledge, and difficult to scale or standardize.

In contrast, the *To-Be* process introduces a higher level of automation and integration. After the data is exported from NAV and saved into an Excel file, the user simply runs a Python script (Automatismo.ipynb), which automatically classifies the transactions according to pre-defined rules and structures the data for analysis. If there are transactions that require manual classification, they are flagged for user input. Once all transactions are processed, the Power BI

dashboard (Reporte_Controlo_Gestao.pbix) is refreshed with a single click, providing real-time, interactive reporting across multiple dimensions.

This redesigned workflow significantly reduces the number of manual steps, mitigates the risk of inconsistency, and ensures that the reporting logic is centralized and reusable. Moreover, the output moves from static files to a dynamic, filterable Power BI interface, enhancing data accessibility and decision-making. The process is thus more sustainable, scalable, and better aligned with the organization's long-term digital transformation efforts, as presented in **Table 2**.

Dimension	Figure 2	Figure 3
Data Extraction	Manual export from NAV and copy into Excel	Manual export from NAV and copy into Excel
Classification	Manual, based on individual judgment and Excel filters	Automated via Python script; manual input only for exceptions
Tools Involved	Excel, PowerPoint	Excel, Python, Power BI
Repetition and Effort	High – repeated tasks every month, including reclassification	Low – once rules are defined, only incremental adjustments are needed
Error Risk	High – due to manual steps and copy-paste operations	Low – logic is embedded in the script and dashboards, reducing human intervention
Output Format	Static PowerPoint presentation	Dynamic, interactive Power BI dashboard
Know-How Dependency	High – relies on tacit knowledge and manual interpretation	Lower – logic formalized in automated classification and centralized models
Updating Process	Manual – copy of PivotTables to presentation format	Automated – one-click refresh updates the dashboard
Scalability	Limited – difficult to expand without adding manual workload	High – new data can be processed and visualized with minimal additional effort

Table 2: Comparison between the *As-Is* and *To-Be* reporting processes

4.2. Data Preparation

In this project, the data preparation phase played a foundational role in enabling both the descriptive and predictive components of the solution.

This section outlines the main transformations carried out to ensure data quality, consistency, and analytical usability.

The output of the automation presented in **Appendix *1*** is the table exemplified in **Table *3***, which contains the same number of rows as the transactions table but only 25 columns: 20 corresponding to transaction data, 4 auxiliary columns for filtering and visualization in Excel by the Management Control team, and 1 containing the subcategory classification (semi-automated, because there are ID_Subrubricas that need manual input, identified with *[Preencher]*).

Transactions Information			**Auxiliar Columns**				**Classification**
ID_Movimento	**…**	**Valor**	**Obs.**	**Mês_Mov**	**Mês_Evento**	**Ano_Evento**	**ID_Subrubrica**
1	…	-53,25	Pautas	01	11	23	19
2	…	3,84	Economato	01	01	24	23
3	…	9,99	Ciclo de Piano	01	05	24	[Preencher]
…	…	…	…	…	…	…	…
149 873	…	-232,76	*null*	12	02	25	132

Table 3: Example of the output from the semi-automatic classification process

The automation developed for the classification of accounting entries led to a significant productivity improvement in the monthly reporting process. Traditionally, this task required one to two full working days (8 to 16 hours) from a single employee—the only one with the necessary expertise to complete it within that timeframe. Each month, new transactions were manually classified, and previous months' classifications were reviewed and adjusted when needed.

With the implementation of the automated classification system, it was possible to test its performance using the full set of transactions from December 2024, corresponding to the total volume of movements for the entire year. The system successfully classified approximately 150,000 transactions in just 8.5 minutes (0.1417 hours), using over 1,050 predefined rules.

The productivity gain was calculated by subtracting the time taken by the automated process from the time typically spent on manual classification, dividing the result by the manual time, and multiplying by 100 to obtain the percentage. Based on this formula, the estimated productivity increase ranges between 98.2% (assuming 8 hours of manual work) and 99.1% (assuming 16 hours). These results highlight the potential of the automated system to drastically reduce manual workload, ensure standardization, and free up valuable human resources for more analytical and strategic tasks.

With this data, and following the steps outlined in the Python code in **Appendix 5**, the table presented in **Table 4** was generated and then also used in the Data Model, for the Reais_Mensais_2025 table.

Data	Expenses	Revenue	EBT
2025-01	-40 722,89	40 250,84	-472,05
2025-02	-10 957,19	10 490,96	-466,23
2025-03	-13 773,08	12 134,92	-1638,16
2025-04	-21 051,28	21 105,65	54,37
2025-05	-25 029,16	31 340,36	6 311,20

Table 4: Illustrative output produced by the Python code, based on non-sensitive dummy data

4.3. Modelling and Evaluation

Following the data preparation phase, the project advanced to the development and evaluation of both the data and predictive models, through an iterative process. The primary objective of this phase was to design models and visualizations capable of supporting planning and strategic decision-making.

The implications of these modelling efforts are also discussed, particularly in terms of their added value for proactive financial management and long-term scalability. The data model was structured as illustrated in **Figure 1**, and its constituent tables are described below.

The Agregadores table contains the IDs and names of aggregators, representing the highest level of budget execution reporting at Casa da Música. The Rubricas table consists of category IDs and names, with a foreign key linking

them to the corresponding aggregator. Similarly, the table Subrubricas contains subcategory IDs and names, linked to their respective categories. These three tables establish the fundamental budget classification hierarchy.

The Contas, Grupos, Naturezas, Eventos, and Areas tables store the IDs and descriptive names of their respective dimensions and maintain one-to-many relationships with the central Movimentos table. This table aggregates all transaction records and includes foreign keys referencing the relevant subrubrica, account, group, nature, event, and area. The Movimentos table also contains the transaction value and a month indicator, allowing for temporal analysis.

The Calendario table was introduced to ensure consistent time-based reporting. It establishes relationships with multiple fact tables: Movimentos (explained above), Previsao_Mensal_2025 (monthly forecasts obtained through predictive models), Reais_Mensais_2025 (monthly actual performance), and O2025 (containing budget information for 2025). These tables include values such as Revenue, Expenses, and EBT (Earnings Before Taxes), and are essential for comparative performance analysis across budgeted, forecasted, and actual data.

For the predictive component, the model with the lowest RMSE was selected, as shown in **Table *5***.

Model	RMSE_Revenue	RMSE_Expenses	RMSE_EBT
ETS(alfa,None,None)	374302,4358	245882,5046	394028,7491
ETS(alfa,beta,None)	376472,9379	245450,2391	396197,9397
ETS(alfa,beta,gama)	**346431,4474**	**139551,7805**	**374604,2089**

Table 5: The RMSE results for the nine models created, with the three best in bold

Once the model was selected, monthly forecasts for 2025 were generated, as presented in **Table *6***. As a validation step, the estimated total annual EBT for 2025 was checked against previous years' EBTs and was confirmed to be consistent, receiving approval from the business team.

Date	Revenue	Expenses	EBT
2025-01-31	1 260 622,62	-1 259 030,89	3 994,75
2025-02-28	1 176 533,42	-1 325 195,94	-145 914,34
2025-03-31	1 315 333,54	-1 422 101,08	-97 476,37
2025-04-30	1 200 424,93	-1 386 074,77	-181 371,35
2025-05-31	1 247 605,51	-1 294 096,21	-39 314,82
2025-06-30	1 350 752,66	-1 287 422,86	68 858,85
2025-07-31	1 310 944,99	-1 315 454,50	4 253,55
2025-08-31	1 240 861,17	-1 152 945,96	93 190,55
2025-09-30	1 256 609,79	-1 327 387,59	-64 848,66
2025-10-31	1 345 321,55	-1 482 908,67	-130 751,30
2025-11-30	1 477 145,94	-1 429 556,44	54 428,20
2025-12-31	1 627 631,31	-1 928 253,27	-291 101,99

Table 6: The forecasts for each of the three selected models

The data model, the monthly forecasts, and the four prototype Power BI dashboards were designed according to the needs of the Management Control team, collected in the kick-off meetings. In order to properly evaluate and further develop them, they were refined based on the feedback gathered during a 3-hour and 30-minute collaborative working session. This iterative process led to their final version, as presented below.

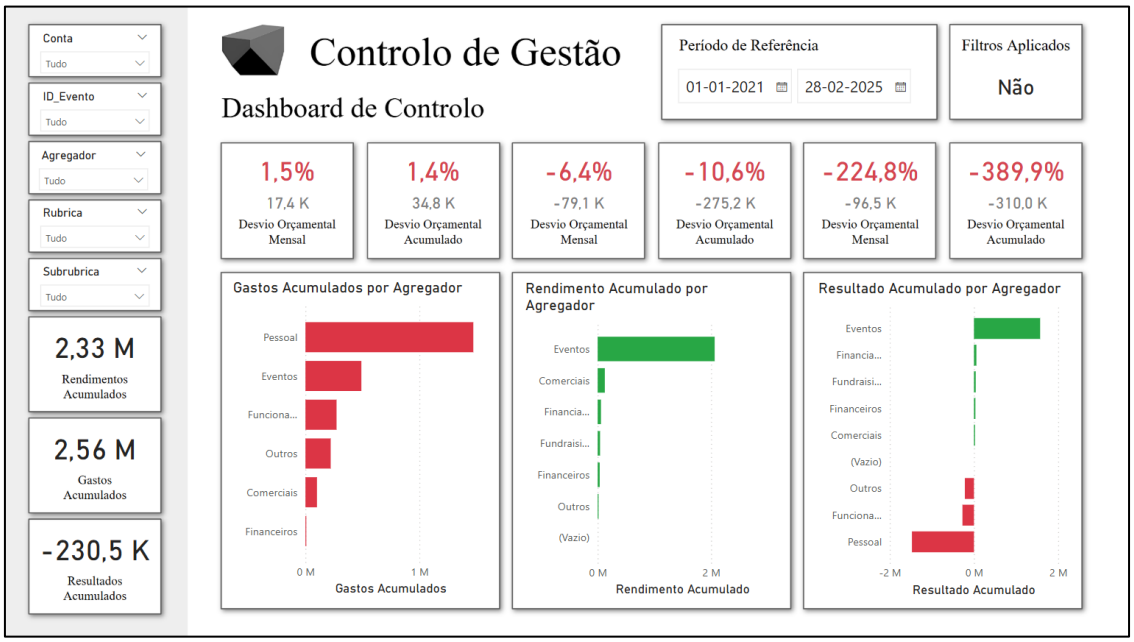


Figure 4: The Management Control Dashboard

The management control dashboard provides a comprehensive and integrated view of budget execution, enabling the continuous monitoring of revenues, expenses, and deviations from the budget. This panel plays a central role in

supporting decision-making by offering aggregated and segmented financial indicators across different areas of activity (referred to as "Aggregators").

At the top of the dashboard, key performance indicators (KPIs) are displayed, highlighting both monthly and cumulative budget deviations in absolute and percentage terms. These metrics allow for the quick identification of imbalances in budget execution over the selected period.

The dashboard also includes visualizations of accumulated revenues, expenses, and financial results, using horizontal bar charts to show the main aggregators driving each dimension. This segmentation facilitates the identification of areas with the highest expenditures (such as "Personnel") or the most significant sources of revenue (notably "Events").

In addition, the dashboard features dynamic filters by account, event, category, and subcategory, allowing users to explore the data at varying levels of granularity. The inclusion of a reference period selector further enhances analytical flexibility, making the dashboard a versatile tool for financial oversight and for promoting more efficient and transparent management practices.

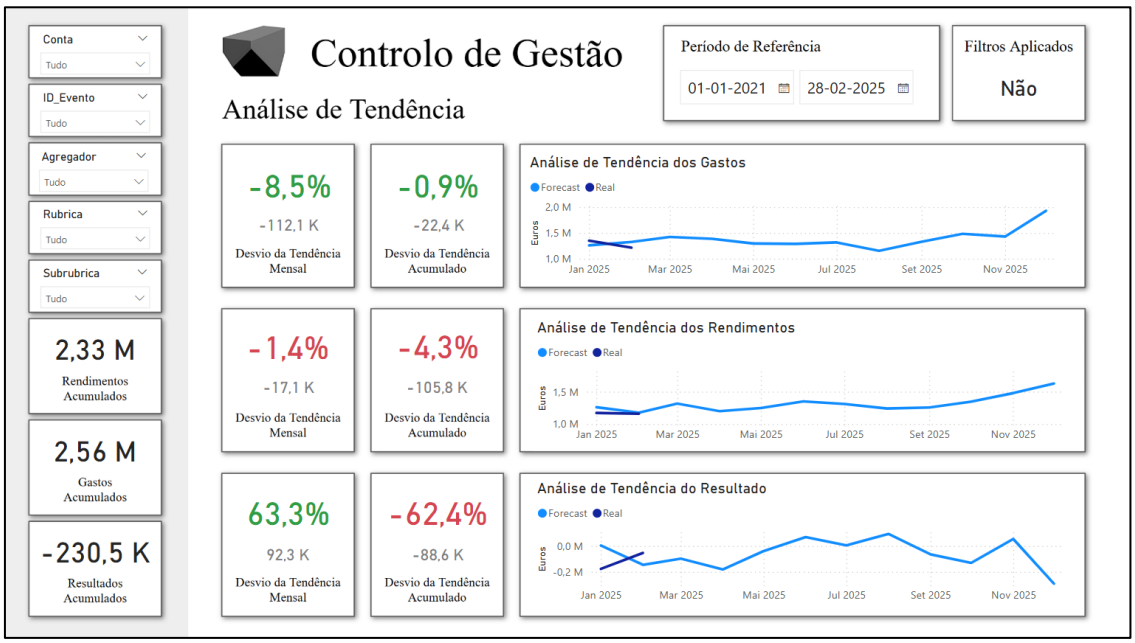


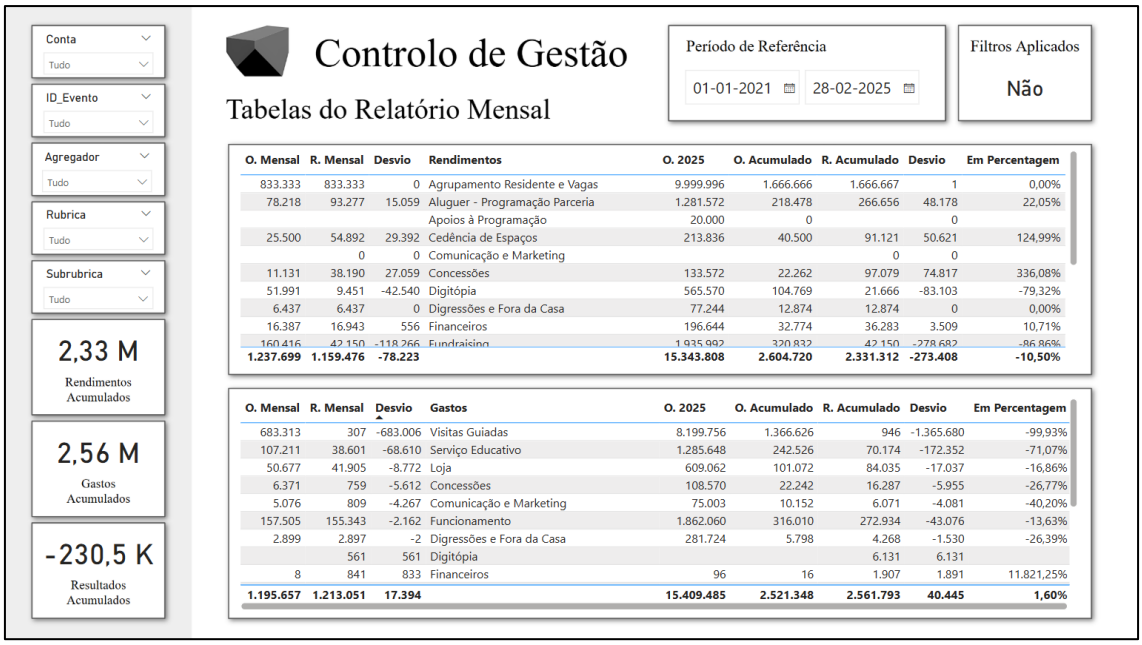
Figure 5: The Trend Analysis Dashboard

The trend analysis dashboard offers a forward-looking perspective on financial management by comparing actual values with statistical forecasts. This panel is particularly useful for identifying deviations from expected financial behavior, enabling proactive decision-making and improving the institution's ability to anticipate risks and opportunities.

The dashboard presents key metrics on monthly and cumulative deviations from forecasted trends, segmented across three core dimensions: expenses, revenues, and results. These indicators quantify the variance between actual and predicted values, both in absolute terms and as a percentage, helping assess the reliability of planning and execution.

Complementing these indicators, line charts provide a visual comparison of real versus forecasted amounts over time, offering an intuitive view of performance dynamics throughout the year. Separate graphs are dedicated to each financial dimension, allowing users to identify seasonal patterns, outliers, or significant divergences.

As with the control dashboard, interactive filters enable detailed analysis by account, event, category, and subcategory, while the reference period selector supports temporal flexibility. By integrating predictive analytics into financial oversight, this dashboard promotes more informed and responsive management practices.



O. Mensal	R. Mensal	Desvio	Rendimentos	O. 2025	O. Acumulado	R. Acumulado	Desvio	Em Percentagem
833.333	833.333	0	Agrupamento Residente e Vagas	9.999.996	1.666.666	1.666.667	1	0,00%
78.218	93.277	15.059	Aluguer - Programação Parceria	1.281.572	218.478	266.656	48.178	22,05%
			Apoios à Programação	20.000	0	0		
25.500	54.892	29.392	Cedência de Espaços	213.836	40.500	91.121	50.621	124,99%
	0	0	Comunicação e Marketing			0	0	
11.131	38.190	27.059	Concessões	133.572	22.262	97.079	74.817	336,08%
51.991	9.451	-42.540	Digitópia	565.570	104.769	21.666	-83.103	-79,32%
6.437	6.437	0	Digressões e Fora da Casa	77.244	12.874	12.874	0	0,00%
16.387	16.943	556	Financeiros	196.644	32.774	36.283	3.509	10,71%
160.416	42.150	-118.266	Fundraising	1.935.992	320.832	42.150	-278.682	-86,86%
1.237.699	**1.159.476**	**-78.223**		**15.343.808**	**2.604.720**	**2.331.312**	**-273.408**	**-10,50%**

O. Mensal	R. Mensal	Desvio	Gastos	O. 2025	O. Acumulado	R. Acumulado	Desvio	Em Percentagem
683.313	307	-683.006	Visitas Guiadas	8.199.756	1.366.626	946	-1.365.680	-99,93%
107.211	38.601	-68.610	Serviço Educativo	1.285.648	242.526	70.174	-172.352	-71,07%
50.677	41.905	-8.772	Loja	609.062	101.072	84.035	-17.037	-16,86%
6.371	759	-5.612	Concessões	108.570	22.242	16.287	-5.955	-26,77%
5.076	809	-4.267	Comunicação e Marketing	75.003	10.152	6.071	-4.081	-40,20%
157.505	155.343	-2.162	Funcionamento	1.862.060	316.010	272.934	-43.076	-13,63%
2.899	2.897	-2	Digressões e Fora da Casa	281.724	5.798	4.268	-1.530	-26,39%
	561	561	Digitópia			6.131	6.131	
8	841	833	Financeiros	96	16	1.907	1.891	11.821,25%
1.195.657	**1.213.051**	**17.394**		**15.409.485**	**2.521.348**	**2.561.793**	**40.445**	**1,60%**

Figure 6: The Monthly Report Tables Dashboard

The monthly report tables dashboard provides a detailed breakdown of budget execution at the subcategory level, offering a granular view of both revenues and expenses. By presenting actual values alongside planned (budgeted) amounts, the dashboard facilitates a precise assessment of monthly and cumulative deviations.

The upper section of the dashboard focuses on revenues, presenting side-by-side comparisons of budgeted and realized amounts for each income source. Deviations are calculated in both absolute terms and percentages, allowing users to identify areas of overperformance or underperformance—such as significant positive deviations in "Space Rentals" or negative deviations in "Digital Projects".

The lower section of the dashboard mirrors this structure for expenses, enabling detailed tracking of spending across various activities. Key cost centers such as "Guided Tours" or "Educational Services" are monitored to assess their adherence to the planned budget.

This tabular layout supports accountability by making deviations transparent and traceable. When combined with the dashboard's filtering capabilities (by

account, event, category, and subcategory), it allows for flexible and in-depth financial analysis. Overall, the dashboard complements the more visual and aggregated panels by offering a precise, data-driven representation of monthly financial execution.

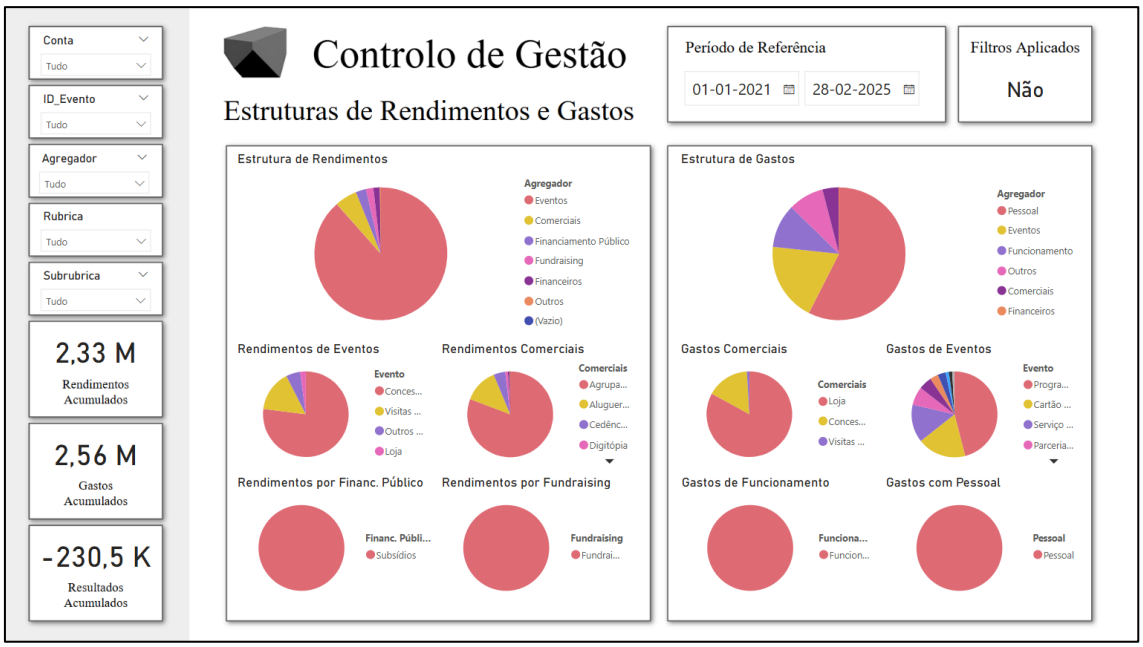


Figure 7: The Revenue and Expense Structures Dashboard

The revenue and expense structures dashboard provides a visual breakdown of financial flows by category, offering insights into the composition of both income and expenditure. Through a series of pie charts, the dashboard reveals how different activities and aggregators contribute to the institution's financial structure.

On the left-hand side, the dashboard displays the structure of revenues, highlighting the dominant role of "Events" as a source of income, followed by commercial activities, public funding, and other categories. Complementary pie charts decompose revenues into further detail—distinguishing between types of event-related income, commercial operations (such as space rentals and concessions), public funding (primarily subsidies), and fundraising.

On the right-hand side, the expense structure is similarly represented, with the largest share attributed to "Personnel", followed by operational costs, events,

and other categories. Detailed charts further break down spending within each aggregator, including “Commercial Expenses”, “Event Expenses”, “Operating Expenses”, and “Personnel Costs”.

This dashboard plays a key role in enhancing transparency by making the distribution of resources immediately visible. The intuitive visual format facilitates comparative analysis between income and expenditure categories, supporting more informed financial planning and resource allocation.

Figure 8: The Commentary Section Dashboard

The commentary section dashboard is designed as a dedicated space for qualitative analysis and interpretation of financial results. Unlike the other dashboards, which are data-driven and automatically populated, this panel requires manual input, enabling the integration of contextual information, narrative insights, and strategic reflections.

Users can insert free-text observations, key takeaways, and recommendations based on the quantitative findings presented in the other dashboards. This may include explanations for significant budget deviations, unforeseen events affecting financial performance, or proposals for corrective measures.

The flexibility of this section allows it to be tailored to different reporting needs and audiences, making it particularly valuable for monthly or quarterly financial reviews. By combining data with commentary, the dashboard fosters a more holistic approach to financial management—balancing numbers with interpretation and strategic foresight.

4.4. Deployment

The final phase of the project focused on the deployment of the developed solution, ensuring its integration into the organization's operational context and its accessibility to end users. This stage aimed not only to deliver a technically functional product, but also to guarantee that the tools created—dashboards, models, and automation scripts—could be adopted effectively and sustainably by the target stakeholders.

One of the main challenges encountered during this phase was the functional implementation itself. Specific authorizations and configuration changes to institutional computers were required to allow the Python scripts to run smoothly. Additionally, two full days of on-site work were necessary to thoroughly explain, demonstrate, and implement the new reporting process, which initially ran in parallel with the previous process during a trial phase in April.

All requirements of the Management Control team were respected, and final feedback and requests were incorporated into the solution. These included the creation of the Structure and Commentary dashboards—features not originally planned—as well as the development of a reference period filter allowing users to select a custom start and end date, replacing the initially proposed single-month filter.

To quantitatively assess this final phase—alongside earlier stages—, as mentioned before, a feedback questionnaire was designed and distributed, as

described in the chapter Research Approach and Methods. The results of this questionnaire are presented in the table below.

Business Understanding

Affirmation	Insuffient	Regular	Excellent
The project clearly identified the business objectives.	-	-	4
The project was aligned with Casa da Música's strategic goals.	-	-	4
The institution's specific needs were well understood.	-	-	4
This phase was conducted with rigour and attention.	-	-	4

Data Understanding

Affirmation	Insuffient	Regular	Excellent
The data was properly understood before model development began.	-	-	4
An appropriate assessment of the data quality was carried out.	-	-	4
The information available was sufficient to address the project's goals.	-	-	4
This phase was conducted with rigour and attention.	-	-	4

Data Preparation

Affirmation	Insuffient	Regular	Excellent
The data was correctly cleaned and prepared for analysis.	-	-	4
Relevant variables were selected based on clear criteria.	-	-	4
The data preparation was adequate for the needs of the data and predictive models.	-	-	4
This phase was conducted with rigour and attention.	-	-	4

Modelling

Affirmation	Insuffient	Regular	Excellent
The modelling techniques used were appropriate for the problem under study.	-	1	3
The models were properly trained and validated using the available data.	-	2	2
The assumptions underlying the models were clearly analysed and communicated.	-	-	4
This phase was conducted with rigour and attention.	-	1	3

Evaluation

Affirmation	Insuffient	Regular	Excellent
The results presented in Power BI were interpreted clearly and rigorously.	-	1	3
The dashboards effectively responded to the identified business challenges	-	-	4
A critical, collaborative, and comparative evaluation of possible solutions was conducted.	-	1	3
This phase was conducted with rigour and attention.	-	-	4

Deployment

Affirmation	Insuffient	Regular	Excellent
The project deliverables (e.g., automation, dashboards, and the new reporting process) were presented in a clear and accessible way	-	-	4
The solution was successfully implemented within the organization.	-	2	2
The team was adequately involved in adopting and applying the solution.	-	-	4
This phase was conducted with rigour and attention.	-	-	4

Table 7: Count of the answers for each affirmation in the questionnaire

Lastly, this section also incorporates a qualitative statement provided by the Director of Management Control and Systems at Casa da Música. To preserve the authenticity and original tone of the contribution, the comment is first presented in its original language, Portuguese, and subsequently translated into English. The full transcript and the translation of the statement are available below.

[Portuguese version]

"Em termos gerais o projeto decorreu de forma exemplar. Houve algumas questões no formulário que respondi com "regular" em vez de "excelente", pelo facto de não ter sido possível no espaço de tempo do projeto aplicar os modelos a dados reais e testar as regras para classificação dos dados. Só iremos fazê-lo com os dados do fecho de contas do mês de março. Outra dimensão que não foi possível analisar mais de perto e avaliar foi a dos modelos preditivos. Em suma, acho que teria sido útil mais uma sessão de trabalho com a avaliação dos modelos aplicados à população, e não apenas a uma amostra, e com dados reais/totais no que respeita às regras. Estou consciente que o facto da equipa da Casa da Música estar ocupada com o seu trabalho diário também pode ter limitado a disponibilização atempada de informação necessária ao projeto.

Estamos imensamente gratos ao Luís, pela dedicação, empenho e profissionalismo demonstrado. Tenho a certeza que este trabalho irá impactar de forma positiva todo o trabalho do Controlo de Gestão. O automatismo criado para a classificação dos movimentos contabilísticos e os dashboards produzidos em Power BI são sem dúvida uma mais valia para o trabalho da equipa do Controlo de Gestão da Casa da Música.

Muito obrigada e Parabéns."

[English version]

"In general, the project went really well. There were a few questions in the form where I answered "fair" instead of "excellent", simply because it wasn't possible within the project's timeframe to apply the models to real data and test the rules for data classification. We'll only be able to do that with the March month-end data. Another area we couldn't look at more closely or evaluate was the predictive models.

All in all, I think an extra working session would've been useful — one where we could assess the models applied to the full dataset, not just a sample, and test the rules using real/complete data. I'm aware that the Management Control team's day-to-day workload may also have limited the timely availability of the info needed for the project.

We're extremely grateful to Luís for his dedication, effort, and professionalism. I'm sure this work will have a positive impact on the entire Management Control team's work. The automation built for classifying accounting entries and the dashboards created in Power BI are definitely a valuable asset for the team at Casa da Música.

Thank you very much, and congratulations."

5. Final Reflections and Directions for Future Research

Fundação Casa da Música served as the case study for this project, motivated by the organization's need to reinforce its budgeting and management control processes through analytical and automated tools.

Using the CRISP-DM methodology, a twofold solution was developed: a descriptive approach, enabling the automation of accounting transaction classification, and a predictive component, focused on modelling monthly financial performance. These solutions were integrated into a Power BI environment, resulting in dynamic dashboards that streamline financial reporting and support strategic and proactive planning.

Based on the organization's internal data, it was possible to significantly reduce manual effort—achieving productivity gains of over 98%—standardize reporting logic, and provide faster access to reliable insights. The solution was well received by end users, as reflected in both the evaluation questionnaires and qualitative feedback, confirming its practical relevance and usability.

Considering the above, the main contributions of this project can be summarized as follows: i) an automated system for classifying accounting transactions was developed, delivering substantial efficiency gains; ii) interactive dashboards for the budgeting and management control processes, tailored to end-user needs, were created and implemented in the institution; iii) and the CRISP-DM methodology was practically applied in the real-world context of a cultural institution.

Nevertheless, some limitations must be acknowledged. The inability to test the models using actual month-end data from 2025 within the project timeframe restricted their full validation. Additionally, technical and operational constraints—such as hardware limitations and restricted access to data—affected the speed of implementation and the ability to iterate even more.

In this context, the positive impact of this case study opens the door to further work in this domain, including the i) ongoing validation of the predictive models, using real and complete data from 2025 to assess forecasting accuracy; ii) the improvement and expansion of the classification rules and system in order to be fully automated; iii) the integration of cultural programming, audience, and ticketing data into management dashboards, enabling cross-analysis between cultural activity and financial performance; iv) the development of more advanced predictive models, including machine learning approaches for medium-term forecasting; v) the design of a data governance model, ensuring long-term sustainability and data quality maintenance throughout the years.

In conclusion, this project successfully achieved its goals and made a valuable contribution to the modernization of Casa da Música's management control practices. Even more importantly, it lays a solid foundation for the digital transformation of budgeting and management control processes in cultural institutions.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of my written work, ChatGPT and Microsoft Copilot were used for the following tasks:

- Refining research questions and structuring the overall outline of the thesis;
- Clarifying theoretical concepts and summarizing academic literature;
- Supporting the development of data analysis logic and DAX formulas for Power BI;
- Task 4: Providing suggestions for improving the clarity and coherence of written sections;
- Task 5: Reviewing grammar and style in the final proofreading stage;

The prompts used are listed at the end of the document in the Prompts List section.

After using these tools, I reviewed and edited the content as necessary, and I take full responsibility for the content of the work presented.

I also declare that I am aware of and respect the Artificial Intelligence Rules of Conduct of Católica Porto Business School.

Bibliography

Correia, A., Agua, P. B., & Luis, A. (2021, June 23). Business Intelligence supporting Budget Management. *Iberian Conference on Information Systems and Technologies, CISTI*. https://doi.org/10.23919/CISTI52073.2021.9476430

Fundação Casa da Música. (2024a). *A Casa*. Https://Casadamusica.Com/a-Casa/.

Fundação Casa da Música. (2024b). *A Fundação*. Https://Casadamusica.Com/Fundacao-Casa-Da-Musica/.

Gałecka, M., & Smolny, K. (2017). Financing rules of the activity of cultural institutions in the context of economic efficiency. *Ekonomia i Prawo*, *16*(4), 387. https://doi.org/10.12775/eip.2017.027

Kalekar, P. S. (2004). *Time series Forecasting using Holt-Winters Exponential Smoothing*.

Miller, D., Allen, M., Schnittger, S., & Hackman, T. (2013). *How rolling forecasting facilitates dynamic, agile planning.*

Nuccio, M., & Bertacchini, E. (2022). Data-driven arts and cultural organizations: opportunity or chimera? *European Planning Studies*, *30*(9), 1638–1655. https://doi.org/10.1080/09654313.2021.1916443

Object Management Group. (2013). *Business Process Model and Notation (BPMN) Version 2.0.2*.

Pushkarna, S., & Walcott, T. H. (2023). Managerial Budget Improvements Using Data Analytics. *Proceedings - 2023 International Conference on Computing, Electronics and Communications Engineering, ICCECE 2023*, 121–126. https://doi.org/10.1109/iCCECE59400.2023.10238582

Rafi, M., Ahmad, K., Bin Naeem, S., & Jianming, Z. (2020). Budget harmonization and challenges: understanding the competence of professionals in the budget process for structural and policy reforms in public libraries.

Performance Measurement and Metrics, *21*(2), 65–79. https://doi.org/10.1108/PMM-09-2019-0048

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, *181*, 526–534. https://doi.org/10.1016/j.procs.2021.01.199

Wang, Y. (2016). Application of budget system based on business intelligence. *2016 7th IEEE International Conference on Software Engineering and Service Science*, 808–812.

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (Vol. 1, pp. 29–39).

Wnuczak, P. (2014). *Zeszyty Teoretyczne Rachunkowości* (Vol. 79, Issue 135).

Appendices

Appendix 1: Python code for the semi-automatic classification of transactions

```
import pandas as pd
from datetime import datetime
import os

nome_do_utilizador = "lgraca"
os.chdir("Path_to_the_selected_folder")

movimentos_inicial = pd.read_excel("Input_Automatismo.xlsx",
sheet_name="Movimentos", dtype=str)
regras_inicial = pd.read_excel("Input_Automatismo.xlsx",
sheet_name="Regras", dtype=str)

colunas_filtragem = ["ID_Conta", "ID_Grupo", "ID_Natureza", "ID_Evento",
"ID_Area"]

movimentos = movimentos_inicial[["ID_Movimento", "Data"] +
colunas_filtragem].copy()
regras = regras_inicial[colunas_filtragem + ["ID_Subrubrica"]].copy()

regras = regras.fillna("any")

# Criar dicionário para correspondência de prefixos de ID_Conta
conta_map = {}
for _, row in regras.iterrows():
    conta_regra = row["ID_Conta"]
    conta_map.setdefault(conta_regra, []).append(row)

# Função para encontrar a melhor correspondência sem cross join
def encontrar_correspondencia(row):

    conta_mov = row["ID_Conta"]
    if conta_mov in conta_map:
        possiveis_regras = conta_map[conta_mov]
    else:
        possiveis_regras = [
            regra for prefixo, regras in conta_map.items()
            if conta_mov.startswith(prefixo) for regra in regras
        ]

    possiveis_regras = [
        regra for regra in possiveis_regras
        if all(regra[col] == "any" or regra[col] == row[col] for col in
colunas_filtragem if col != "ID_Conta")
    ]

    if possiveis_regras:
```

```
        melhor_regra = max(possiveis_regras, key=lambda x: sum(x[col] !=
"any" for col in colunas_filtragem))
        return melhor_regra["ID_Subrubrica"]
    return "[Preencher]"

# Aplicar a correspondência usando .apply()
movimentos.loc[:, "ID_Subrubrica"] =
movimentos.apply(encontrar_correspondencia, axis=1)

# Criar tabela de estatísticas
total_movimentos = len(movimentos)
preencher = movimentos["ID_Subrubrica"].value_counts().get("[Preencher]",
0)
classificadas = total_movimentos - preencher

tabela_stats = pd.DataFrame({
    "Classificação": ["[Preencher]", "Classificadas", "Total"],
    "Movimentos": [preencher, classificadas, total_movimentos],
    "%": [f"{preencher/total_movimentos:.0%}",
f"{classificadas/total_movimentos:.0%}", "100%"]
    })

# Preparar dados para exportação
stats_extrair = tabela_stats
classificacao_extrair = movimentos
movimentos_extrair = movimentos_inicial
regras_extrair = regras_inicial

# Gerar timestamp para o nome do arquivo
data_exportacao = datetime.now().strftime("%Y%m%d")
nome_arquivo = f"Output_Automatismo_{data_exportacao}.xlsx"

# Exportar para Excel
with pd.ExcelWriter(nome_arquivo, engine='openpyxl') as writer:
    stats_extrair.to_excel(writer, sheet_name="Estatisticas",
index=False)
    classificacao_extrair.to_excel(writer, sheet_name="Classificacao",
index=False)
    regras_extrair.to_excel(writer, sheet_name="Regras", index=False)
    movimentos_extrair.to_excel(writer, sheet_name="Movimentos",
index=False)

print(f"☑ Classificação concluída e guardada em '{nome_arquivo}'")
```

Appendix 2: Tables and Metrics from the PowerBI Data Model

Subrubricas (table)

ID_Rubrica	ID_Subrubrica	Subrubrica
1	0	Rend Visitas Guiadas
1	1	Posterior - Rend Visitas Guiadas
1	2	Gastos Visitas Guiadas
2	4	Rend Digitópia
2	5	Gastos Digitópia
3	6	Bilheteira Serviço Educativo

Rubricas (table)

ID_Agregador	ID_Rubrica	Rubrica
1	1	Visitas Guiadas
2	2	Digitópia
2	3	Serviço Educativo
2	4	Aluguer - Programação Parceria
2	5	Programação Própria
3	6	Pessoal
1	7	Loja
4	8	Funcionamento
2	9	Parceria - Programação Parceria
2	10	Digressões e Fora da Casa
1	11	Concessões
2	12	Comunicação e Marketing
2	13	Cedência de Espaços
2	14	Cartão Amigo
2	15	Agrupamento Residente e Vagas
5	16	Subsídios
6	17	Fundraising
7	18	Apoios à Programação
1	19	Outros Rendimentos Comerciais
8	20	Financeiros
9	21	Outros Rendimentos
1	22	Outros Gastos Comerciais
9	23	Outros Gastos

Reais_Mensais_2025 (table)

Expenses	Revenue	EBT	Data Registo	Revenue_intuitivo	EBT_intuitivo
1348742,22	-1171835,91	176906,31	1 de janeiro de 2025	1171835,91	-176906,31
1213050,87	-1159476,22	53574,65	1 de fevereiro de 2025	1159476,22	-53574,65

Previsao_Mensal_2025 (table)

Date	Revenue	Expenses	EBT	EBT_intuitivo	Revenue_intuitivo
31/01/2025 00:00:00	-1260622,61544049	1259030,89222743	-3994,74555320394	3994,74555320394	1260622,61544049
28/02/2025 00:00:00	-1176533,41848419	1325195,93792123	145914,34019376	-145914,34019376	1176533,41848419
31/03/2025 00:00:00	-1315333,53561079	1422101,07625113	97476,3719245939	-97476,3719245939	1315333,53561079
30/04/2025 00:00:00	-1200424,92779122	1386074,76769654	181371,354448158	-181371,354448158	1200424,92779122
31/05/2025 00:00:00	-1247605,51299657	1294096,21189244	39314,8218292116	-39314,8218292116	1247605,51299657
30/06/2025 00:00:00	-1350752,65839514	1287422,85658877	-68858,8522690994	68858,8522690994	1350752,65839514
31/07/2025 00:00:00	-1310944,99472616	1315454,50274915	-4253,5525618248	4253,5525618248	1310944,99472616
31/08/2025 00:00:00	-1240861,17332344	1152945,96318993	-93190,5524305014	93190,5524305014	1240861,17332344
30/09/2025 00:00:00	-1256609,79341391	1327387,58820112	64848,6602367921	-64848,6602367921	1256609,79341391
31/10/2025 00:00:00	-1345321,54730154	1482908,66904049	130751,303193296	-130751,303193296	1345321,54730154
30/11/2025 00:00:00	-1477145,9360051	1429556,43740873	-54428,1999653789	54428,1999653789	1477145,9360051
31/12/2025 00:00:00	-1627631,31324687	1928253,26625137	291101,990288737	-291101,990288737	1627631,31324687

O2025 (table)

Subcapítulo	Pivot info	Áreas	Descritivo	Tipo	ID_Subru	Valor	Mes	ID_Conta	Data
Outros Serviços Prestadc	Outros Serviços P	Outros Serviços	Rendimentos Con	Rendimento	6	625	07-julho	7	1 de julho de 2025
Outros Serviços Prestadc	Outros Serviços P	Outros Serviços	Rendimentos Con	Rendimento	6	625	02-fevereiro	7	1 de fevereiro de 2025
Outros Serviços Prestadc	Outros Serviços P	Outros Serviços	Rendimentos Con	Rendimento	6	625	03-março	7	1 de março de 2025
Outros Serviços Prestadc	Outros Serviços P	Outros Serviços	Rendimentos Con	Rendimento	6	625	04-abril	7	1 de abril de 2025
Outros Serviços Prestadc	Outros Serviços P	Outros Serviços	Rendimentos Con	Rendimento	6	625	05-maio	7	1 de maio de 2025
Outros Serviços Prestadc	Outros Serviços P	Outros Serviços	Rendimentos Con	Rendimento	6	625	06-junho	7	1 de junho de 2025
Outros Serviços Prestadc	Outros Serviços P	Outros Serviços	Rendimentos Con	Rendimento	6	625	01-janeiro	7	1 de janeiro de 2025
Outros Serviços Prestadc	Outros Serviços P	Outros Serviços	Rendimentos Con	Rendimento	6	625	08-agosto	7	1 de agosto de 2025
Outros Serviços Prestadc	Outros Serviços P	Outros Serviços	Rendimentos Con	Rendimento	6	625	09-setembro	7	1 de setembro de 2025
Outros Serviços Prestadc	Outros Serviços P	Outros Serviços	Rendimentos Con	Rendimento	6	625	10-outubro	7	1 de outubro de 2025
Outros Serviços Prestadc	Outros Serviços P	Outros Serviços	Rendimentos Con	Rendimento	6	625	11-novembro	7	1 de novembro de 2025
Outros Serviços Prestadc	Outros Serviços P	Outros Serviços	Rendimentos Con	Rendimento	6	625	12-dezembro	7	1 de dezembro de 2025
Bilheteira Serviço Educat	Bilheteira Serviço	Receitas Serviço	Rendimentos de E	Rendimento	5	623	07-julho	7	1 de julho de 2025
Tributação Autónoma (D			Outros (inclui Dep	Gastos	137	600	12-dezembro	6	1 de dezembro de 2025
2412-600 GRAVAÇÕES | E	Programação Pró	Gravações e Edi	Gastos Eventos	Gastos	9	583	07-julho	6	1 de julho de 2025
2412-600 GRAVAÇÕES | E	Programação Pró	Gravações e Edi	Gastos Eventos	Gastos	9	583	02-fevereiro	6	1 de fevereiro de 2025

Naturezas (table)

ID_Natureza	Natureza
10	MUSICOS.Coros
11	DIREÇÃO/MAESTROS
12	MUS.EXTRA
13	SOLISTAS
14	ALUGUER PARTITURAS
15	DES|VIAG|REF Eventos
16	MUS. SUBSTITUIÇÃO
21	EQUIP.+TÉCNICOS
22	OUTROS Gst Prod Eventos
23	TÉCNICOS de SUBTITUIÇÃO
24	TÉCNICOS, Contratação de
25	EQUIPAMENTOS, Aluguer de

Grupos (table)

ID_Grupo	Grupo
2	Piano
4	Recitais e Música de Câmara
7	CCM
12	REMIX
14	Orquestra Sinfónica
15	Orquestra Barroca
16	Outros Ensembles
17	Orquestras Barrocas Convidadas
18	Exposições, Residências e Outros
19	Palestras Pré-Conc e Conferências

Eventos (table)

ID_Evento	Evento
2501-000	2025-01 Custos Comuns aos Eventos
2501-012	OS_FS_Concerto de Ano Novo
2501-013	OS_FS_Concerto de Ano Novo II
2501-014	P_Francisco Costa
2501-015	OS_CI_Cruzamentos Ibéricos
2501-016	OB_Triplo de Bach
2501-017	RE_O regresso de Steve Reich
2501-018	OS_FS_World as Lover / Músicas do Mundo
2501-019	CCM_Bohemian Rhapsody
2501-020	OS_Dc_Dies Irae

Contas (table)

ID_Conta	Nivel1	Nivel2	Nivel3	Nivel4	Nivel5	Nivel6	Name
6	6						GASTOS
61	6	61					CMVMC
611	6	61	611				Mercadorias
612	6	61	612				Matérias-primas, subs e cons
62	6	62					Fornecim e serviços externos
621	6	62	621				Subcontratos
622	6	62	622				Serviços especializados
623	6	62	623				Materiais
624	6	62	624				Energia e fluidos
625	6	62	625				Desloc., estadas e transp
626	6	62	626				Serviços diversos
63	6	63					Gastos com o pessoal

Calendario (table)

Date
01/01/2021 00:00:00
02/01/2021 00:00:00
03/01/2021 00:00:00
04/01/2021 00:00:00
05/01/2021 00:00:00
06/01/2021 00:00:00

Areas (table)

ID_Area	Area
1	Orgãos Sociais
10	Programação
11	Produção
12	Orquestra Sinfónica - Custos Fixos
13	REMIX - Custos Fixos
14	Orquestra Barroca - Custos Fixos
15	Coro CM - Custos Fixos
16	Apoio à Orquestra Sinfónica
20	Educação
31	Assessoria de Comunicação

Agregadores (table)

ID_Agregador	Agregador
1	Comerciais
2	Eventos
3	Pessoal
4	Funcionamento
5	Financiamento Público
6	Fundraising
7	Apoios Diretos à Programação
8	Financeiros
9	Outros

_Metricas (blank table, with all the metrics stored)

```
CorBarrasGastos =
IF(
   [Valor_Real_Acumulado_Conta6] <= 0,
   "#28a745",   -- Verde
   "#dc3545"    -- Vermelho
)

CorBarrasRendimentos =
IF(
   [Valor_Real_Acumulado_Conta7] >= 0,
   "#28a745",   -- Verde
   "#dc3545"    -- Vermelho
)

CorBarrasResultados =
IF(
   [Resultado_Real_Acumulado] >= 0,
   "#28a745",   -- Verde
   "#dc3545"    -- Vermelho
)

Desvio_Acumulado_Conta6 =
[Valor_Real_Acumulado_Conta6] - [Valor_Orcamento_Acumulado_Conta6]

Desvio_Acumulado_Conta7 =
[Valor_Real_Acumulado_Conta7] - [Valor_Orcamento_Acumulado_Conta7]

Desvio_Acumulado_Perc_Conta6 =
DIVIDE([Desvio_Acumulado_Conta6],
Valor_Orcamento_Acumulado_Conta6])

Desvio_Acumulado_Perc_Conta7 =
DIVIDE([Desvio_Acumulado_Conta7],
[Valor_Orcamento_Acumulado_Conta7])

Desvio_EBT_Acumulado =
[EBT_Real_Acumulado] - [EBT_Prev_Acumulado]

Desvio_EBT_Acumulado_Perc =
DIVIDE([Desvio_EBT_Acumulado], ABS([EBT_Prev_Acumulado]))
```

Desvio_EBT_Mensal =
[EBT_Real_Mes] - [EBT_Prev_Mes]

Desvio_EBT_Mensal_Perc =
DIVIDE([Desvio_EBT_Mensal], ABS([EBT_Prev_Mes]))

Desvio_Exp_Acumulado =
[Expenses_Real_Acumulado] - [Expenses_Prev_Acumulado]

Desvio_Exp_Acumulado_Perc =
DIVIDE([Desvio_Exp_Acumulado], ABS([Expenses_Prev_Acumulado]))

Desvio_Exp_Mensal =
[Expenses_Real_Mes] - [Expenses_Prev_Mes]

Desvio_Exp_Mensal_Perc =
DIVIDE([Desvio_Exp_Mensal], ABS([Expenses_Prev_Mes]))

Desvio_Mensal_Conta6 = [Valor_Real_Mes_Conta6] -
[Valor_Orcamento_Mes_Conta6]

Desvio_Mensal_Conta7 = [Valor_Real_Mes_Conta7] -
[Valor_Orcamento_Mes_Conta7]

Desvio_Mensal_Perc_Conta6 =
DIVIDE([Desvio_Mensal_Conta6], [Valor_Orcamento_Mes_Conta6])

Desvio_Mensal_Perc_Conta7 =
DIVIDE([Desvio_Mensal_Conta7], [Valor_Orcamento_Mes_Conta7])

Desvio_Rev_Acumulado =
[Revenue_Real_Acumulado] - [Revenue_Prev_Acumulado]

Desvio_Rev_Acumulado_Perc =
DIVIDE([Desvio_Rev_Acumulado], ABS([Revenue_Prev_Acumulado]))

Desvio_Rev_Mensal =
[Revenue_Real_Mes] - [Revenue_Prev_Mes]

Desvio_Rev_Mensal_Perc =
DIVIDE([Desvio_Rev_Mensal], ABS([Revenue_Prev_Mes]))

```
EBT_Prev_Acumulado =
SUM('Previsao_Mensal_2025'[EBT_intuitivo])

EBT_Prev_Mes =
VAR DataMax = [MesMaisRecente]
RETURN
CALCULATE(
   SUM('Previsao_Mensal_2025'[EBT_intuitivo]),
   MONTH('Previsao_Mensal_2025'[Date]) = MONTH(DataMax)
)

EBT_Real_Acumulado =
SUM('Reais_Mensais_2025'[EBT_intuitivo])

EBT_Real_Mes =
VAR DataMax = [MesMaisRecente]
RETURN
CALCULATE(
   SUM('Reais_Mensais_2025'[EBT_intuitivo]),
   MONTH('Reais_Mensais_2025'[Data Registo]) = MONTH(DataMax)
)

Expenses_Prev_Acumulado =
SUM('Previsao_Mensal_2025'[Expenses])

Expenses_Prev_Mes =
VAR DataMax = [MesMaisRecente]
RETURN
CALCULATE(
   SUM('Previsao_Mensal_2025'[Expenses]),
   MONTH('Previsao_Mensal_2025'[Date]) = MONTH(DataMax)
)

Expenses_Real_Acumulado =
SUM('Reais_Mensais_2025'[Expenses])

Expenses_Real_Mes =
VAR DataMax = [MesMaisRecente]
RETURN
CALCULATE(
   SUM('Reais_Mensais_2025'[Expenses]),
```

```
    MONTH('Reais_Mensais_2025'[Data Registo]) = MONTH(DataMax)
)

MesMaisRecente =
MAX('Movimentos'[Data Registo])

MesMaisRecenteTexto =
VAR DataMax = MAX('Movimentos'[Data Registo])
RETURN
FORMAT(DataMax, "MMMM yyyy")

Resultado_Desvio_Acumulado = [Resultado_Real_Acumulado] -
[Resultado_Orcamento_Acumulado]

Resultado_Desvio_Acumulado_Perc =
DIVIDE([Resultado_Desvio_Acumulado],
ABS([Resultado_Orcamento_Acumulado]))
Resultado_Desvio_Mensal =
[Resultado_Real_Mes] - [Resultado_Orcamento_Mes]

Resultado_Desvio_Mensal_Perc =
DIVIDE([Resultado_Desvio_Mensal], ABS([Resultado_Orcamento_Mes]))

Resultado_Orcamento_Acumulado = [Valor_Orcamento_Acumulado_Conta7]
- [Valor_Orcamento_Acumulado_Conta6]

Resultado_Orcamento_Mes = [Valor_Orcamento_Mes_Conta7] -
[Valor_Orcamento_Mes_Conta6]

Resultado_Real_Acumulado = [Valor_Real_Acumulado_Conta7] -
[Valor_Real_Acumulado_Conta6]

Resultado_Real_Mes = [Valor_Real_Mes_Conta7] - [Valor_Real_Mes_Conta6]

Revenue_Prev_Acumulado =
- SUM('Previsao_Mensal_2025'[Revenue])

Revenue_Prev_Mes =
VAR DataMax = [MesMaisRecente]
RETURN
- CALCULATE(
    SUM('Previsao_Mensal_2025'[Revenue]),
```

```
    MONTH('Previsao_Mensal_2025'[Date]) = MONTH(DataMax)
)

Revenue_Real_Acumulado =
- SUM('Reais_Mensais_2025'[Revenue])

Revenue_Real_Mes =
VAR DataMax = [MesMaisRecente]
RETURN
- CALCULATE(
    SUM('Reais_Mensais_2025'[Revenue]),
    MONTH('Reais_Mensais_2025'[Data Registo]) = MONTH(DataMax)
)

Texto_Filtros_Ativos =
VAR Conta = IF(ISFILTERED('Contas'[Nivel2]), 1, 0)
VAR Evento = IF(ISFILTERED('Eventos'[ID_Evento]), 1, 0)
VAR Agregador = IF(ISFILTERED('Agregadores'[Agregador]), 1, 0)
VAR Rubrica = IF(ISFILTERED('Rubricas'[Rubrica]), 1, 0)
VAR Subrubrica = IF(ISFILTERED('Subrubricas'[Subrubrica]), 1, 0)
VAR Todos = Conta + Evento + Agregador + Rubrica + Subrubrica
RETURN
IF(Todos > 0, "Sim","Não")

Valor_Orcamento_Acumulado_Conta6 =
SUMX(
    FILTER(
        O2025,
        RELATED(Contas[Nivel1]) = 6
    ),
    O2025[Valor]
)

Valor_Orcamento_Acumulado_Conta7 =
SUMX(
    FILTER(
        O2025,
        RELATED(Contas[Nivel1]) = 7
    ),
    O2025[Valor]
)
```

```
Valor_Orcamento_Anual_Conta6 =
CALCULATE(
   SUMX(
      FILTER(
         O2025,
         RELATED(Contas[Nivel1]) = 6
      ),
      O2025[Valor]
   ),
   ALL(Calendario[Date])
)

Valor_Orcamento_Anual_Conta7 =
CALCULATE(
   SUMX(
      FILTER(
         O2025,
         RELATED(Contas[Nivel1]) = 7
      ),
      O2025[Valor]
   ),
   ALL(Calendario[Date])
)

Valor_Orcamento_Mes_Conta6 =
VAR DataMax = [MesMaisRecente]
RETURN
SUMX(
   FILTER(
      O2025,
      RELATED(Contas[Nivel1]) = 6 &&
      MONTH(O2025[Data]) = MONTH(DataMax)
   ),
   O2025[Valor]
)

Valor_Orcamento_Mes_Conta7 =
VAR DataMax = [MesMaisRecente]
RETURN
SUMX(
   FILTER(
      O2025,
```

```
        RELATED(Contas[Nivel1]) = 7 &&
        MONTH(O2025[Data]) = MONTH(DataMax)
    ),
    O2025[Valor]
)

Valor_Real_Acumulado_Conta6 =
SUMX(
    FILTER(
        Movimentos,
        RELATED(Contas[Nivel1]) = 6
    ),
    Movimentos[Valor]
)

Valor_Real_Acumulado_Conta7 =
- SUMX(
    FILTER(
        Movimentos,
        RELATED(Contas[Nivel1]) = 7
    ),
    Movimentos[Valor]
)

Valor_Real_Mes_Conta6 =
VAR DataMax = [MesMaisRecente]
RETURN
SUMX(
    FILTER(
        Movimentos,
        RELATED(Contas[Nivel1]) = 6 &&
        MONTH(Movimentos[Data Registo]) = MONTH(DataMax)
    ),
    Movimentos[Valor]
)

Valor_Real_Mes_Conta7 =
VAR DataMax = [MesMaisRecente]
RETURN
- SUMX(
    FILTER(
        Movimentos,
```

```
        RELATED(Contas[Nivel1]) = 7 &&
        MONTH(Movimentos[Data Registo]) = MONTH(DataMax)
    ),
    Movimentos[Valor]
)
```

Appendix 3: Python code to generate the 2025 monthly forecasts

```
import pandas as pd
import numpy as np
from statsmodels.tsa.holtwinters import ExponentialSmoothing
from sklearn.metrics import mean_squared_error

# Paths necessários
file_path = 'Resultados_2124.xlsx'
output_file_path = 'Resultados_Modelagem_ETS_3.xlsx'

print("Carregando os dados do ficheiro...")
# Carregar os dados do ficheiro
df = pd.read_excel(file_path)
print("Dados carregados com sucesso! Exibindo as primeiras linhas:")
print(df.head())

# Definir os modelos ETS apenas aditivos (excluindo multiplicativos,
porque há valores negativos)
ets_models = [
    ('alfa', None, None),
    ('alfa', 'beta', None),
    ('alfa', 'beta', 'gama')
]

# Função para ajustar o modelo ETS e calcular o RMSE
def fit_ets_model(data, error, trend, seasonal, seasonal_periods):
    print(f"Ajustando o modelo ETS({error},{trend},{seasonal})...")
    model = ExponentialSmoothing(
        data,
        trend=trend,
        seasonal=seasonal,
        seasonal_periods=seasonal_periods,
        initialization_method='estimated',
        use_boxcox=False
    )
    fit = model.fit(optimized=True, method='L-BFGS-B', remove_bias=True)
    predictions = fit.fittedvalues
    rmse = np.sqrt(mean_squared_error(data, predictions))
    print(f"Modelo ETS({error},{trend},{seasonal}) ajustado com sucesso!
RMSE: {rmse}")
    return fit, rmse

# Avaliar cada modelo e armazenar os resultados para Revenue, Expenses e
EBT
results = []
forecasts = {'Mensal': {}}
```

```
for error, trend, seasonal in ets_models:
        try:
            fit_revenue, rmse_revenue =
fit_ets_model(data_source['Revenue'], error, trend, seasonal,
seasonal_periods)
            fit_expenses, rmse_expenses =
fit_ets_model(data_source['Expenses'], error, trend, seasonal,
seasonal_periods)
            fit_ebt, rmse_ebt = fit_ets_model(data_source['EBT'], error,
trend, seasonal, seasonal_periods)

            results.append({
                'Model': f'ETS({error},{trend},{seasonal})',
                'RMSE_Revenue': rmse_revenue,
                'RMSE_Expenses': rmse_expenses,
                'RMSE_EBT': rmse_ebt
            })

            # Previsão para 2025
            forecast_periods = 12
            forecasts['Revenue'] = fit_revenue.forecast(forecast_periods)
            forecasts['Expenses'] =
fit_expenses.forecast(forecast_periods)
            forecasts['EBT'] = fit_ebt.forecast(forecast_periods)

        except Exception as e:
            print(f"Erro ao ajustar o modelo {analysis_type}
ETS({error},{trend},{seasonal}): {e}")

# Converter resultados para DataFrame para melhor visualização e ordenar
por Analysis_Type
results_df = pd.DataFrame(results)
print("\nResultados da Modelagem ETS:")
print(results_df)

# Preparar previsões para exportação
monthly_forecast_df = pd.DataFrame(forecasts['Mensal'])
monthly_forecast_df.index = pd.date_range(start='2025-01-01', periods=12,
freq='M')
monthly_forecast_df.reset_index(inplace=True)
monthly_forecast_df.rename(columns={'index': 'Date'}, inplace=True)

quarterly_forecast_df = pd.DataFrame(forecasts['Trimestral'])
quarterly_forecast_df.index = ['2025-Q1', '2025-Q2', '2025-Q3', '2025-
Q4']
quarterly_forecast_df.reset_index(inplace=True)
quarterly_forecast_df.rename(columns={'index': 'Period'}, inplace=True)
```

```
# Exportar resultados para Excel com várias sheets
print("Exportando os resultados para o ficheiro Excel...")
with pd.ExcelWriter(output_file_path) as writer:
    df.to_excel(writer, sheet_name='Dados Mensais', index=False)
    results_df.to_excel(writer, sheet_name='Resultados Modelagem',
index=False)
    monthly_forecast_df.to_excel(writer, sheet_name='Previsão Mensal
2025', index=False)

print(f"Resultados exportados com sucesso para {output_file_path}!")
```

Appendix 4: Questionnaire to the Management Control team (portuguese version)

Projeto Business Analytics na Casa da Música

O projeto ***Business Analytics na Casa da Música*** foi conduzido com base na metodologia ***CRISP-DM*** (Cross Industry Standard Process for Data Mining), que contempla as seguintes fases:

1. Business Understanding (Compreensão do Negócio);
2. Data Understanding (Compreensão dos Dados);
3. Data Preparation (Preparação dos Dados);
4. Modelling (Desenvolvimento de Modelos);
5. Evaluation (Avaliação dos modelos e escolha da solução);
6. Deployment (Lançamento da solução escolhida).

Este **questionário** visa avaliar o **projeto** e o seu **impacto**, de forma a contribuir para o desenvolvimento de estudos sobre Business Analytics aplicados à Orçamentação em Instituições Culturais, usando o exemplo da Casa da Música.

Nota de confidencialidade:
Todos os dados serão tratados com confidencialidade e usados apenas para fins académicos.

Next Clear form

Business Understanding (Compreensão do Negócio)

Esta fase foca-se em entender os objetivos estratégicos e operacionais da organização. Envolve a identificação das necessidades do negócio, a definição clara dos problemas a resolver e a tradução desses objetivos em metas analíticas concretas. O sucesso do projeto depende de um alinhamento sólido entre os desafios da instituição e as soluções a desenvolver.

Por favor, avalie as afirmações abaixo. *

	Insuficiente	Regular	Excelente
O projeto identificou claramente os objetivos do negócio.	○	○	○
Houve alinhamento entre o projeto e os objetivos estratégicos da Casa da Música.	○	○	○
As necessidades específicas da instituição foram bem compreendidas.	○	○	○
Esta fase foi conduzida com rigor e atenção.	○	○	○

Data Understanding (Compreensão dos Dados)

Após a definição do problema, é essencial recolher, descrever e explorar os dados disponíveis. Esta fase permite avaliar a qualidade dos dados, identificar padrões iniciais e compreender as suas limitações. O conhecimento obtido orienta as etapas seguintes e garante que a solução se baseia numa compreensão sólida da informação existente.

Por favor, avalie as afirmações abaixo. *

	Insuficiente	Regular	Excelente
Os dados foram devidamente compreendidos antes do desenvolvimento dos modelos.	○	○	○
Foi realizada uma análise adequada à qualidade dos dados existentes.	○	○	○
A informação disponível mostrou-se suficiente para responder aos objetivos do projeto.	○	○	○
Esta fase foi conduzida com rigor e atenção.	○	○	○

Data Preparation (Preparação dos Dados)

Esta etapa consiste na seleção, limpeza, transformação e integração dos dados a utilizar. Envolve decisões críticas sobre que variáveis manter, como lidar com valores em falta ou inconsistências, e como estruturar os dados para a modelação. Uma preparação cuidadosa garante a fiabilidade e a robustez da análise.

Por favor, avalie as afirmações abaixo. *

	Insuficiente	Regular	Excelente
Os dados foram corretamente limpos e preparados para análise.	○	○	○
As variáveis relevantes foram selecionadas com base em critérios claros.	○	○	○
A preparação dos dados foi adequada às necessidades dos modelos descritivos e preditivos.	○	○	○
Esta fase foi conduzida com rigor e atenção.	○	○	○

Modelling (Desenvolvimento de Modelos)

Aqui são aplicadas técnicas analíticas e estatísticas adequadas ao problema identificado. A fase inclui o desenvolvimento de diferentes modelos, com a escolha criteriosa de dimensões, parâmetros e algoritmos. O objetivo é desenvolver soluções descritivas e/ou preditivas que respondam com precisão aos desafios definidos anteriormente.

Por favor, avalie as afirmações abaixo. *

	Insuficiente	Regular	Excelente
As técnicas de modelação aplicadas foram adequadas ao problema em estudo.	○	○	○
Os modelos foram corretamente treinados e validados com base nos dados disponíveis.	○	○	○
Os pressupostos subjacentes aos modelos foram devidamente analisados e partilhados.	○	○	○
Esta fase foi conduzida com rigor e atenção.	○	○	○

Evaluation (Avaliação dos modelos e escolha da solução)

Depois de criados os modelos, é necessário avaliá-los de forma crítica. Esta fase verifica se os resultados obtidos estão alinhados com os objetivos do negócio, e compara alternativas possíveis para a solução. A decisão final a adotar é tomada com base numa análise técnica e estratégica, feita com os principais intervenientes.

Por favor, avalie as afirmações abaixo. *

	Insuficiente	Regular	Excelente
Os resultados espelhados em PowerBI foram interpretados de forma clara e rigorosa.	○	○	○
Os dashboards desenvolvidos responderam eficazmente aos desafios de negócio identificados.	○	○	○
Foi realizada uma avaliação crítica, colaborativa e comparativa entre diferentes soluções possíveis.	○	○	○
Esta fase foi conduzida com rigor e atenção.	○	○	○

Deployment (Lançamento da solução escolhida)

Refere-se à fase em que os resultados do projeto são integrados nas operações da organização. Esta etapa inclui a implementação prática da solução desenvolvida, a explicação do novo processo a realizar e a comunicação dos resultados às partes interessadas. O objetivo é garantir que o conhecimento gerado seja utilizado de forma eficaz e sustentável pela instituição.

Por favor, avalie as afirmações abaixo. *

	Insuficiente	Regular	Excelente
Os resultados do projeto (e.g. automatismo, dashboards e novo processo de reporte) foram apresentados de forma clara e acessível.	○	○	○
A solução foi implementada com sucesso na organização.	○	○	○
A equipa foi envolvida de forma adequada na adoção e aplicação da solução.	○	○	○
Esta fase foi conduzida com rigor e atenção.	○	○	○

Pedido de Feedback

Nesta secção, convidamo-lo(a) a partilhar, de forma aberta, as suas opiniões e sugestões sobre o projeto. Comentários qualitativos são fundamentais para compreender aspetos que não são captados pelas respostas anteriores. A sua perspetiva é muito importante para a compreensão e aperfeiçoamento deste e de futuros projetos.

Pode partilhar aqui o seu feedback, sugestões ou reflexões sobre o projeto.

Your answer

Appendix 5: Python code to generate the table with the actual Expenses, Revenue and EBT for each month

```
movimentos = pd.read_excel("Output_Automatismo.xlsx",
sheet_name="Movimentos", dtype=str)

# Converter a coluna 'Valor' para numérica
movimentos["Valor"] = pd.to_numeric(movimentos_inicial["Valor"],
errors='coerce')

# Filtrar apenas os movimentos de 2025
movimentos['Data'] = pd.to_datetime(movimentos['Data'], errors='coerce')
movimentos_2025 = movimentos[movimentos['Data'].dt.year == 2025].copy()

# Adicionar coluna de tipo de valor
movimentos_2025['Tipo_Valor'] = movimentos_2025['ID_Conta'].apply(lambda
x: 'Expenses' if str(x).startswith('6') else ('Revenue' if
str(x).startswith('7') else 'Other'))

# Calcular os resultados mensais
resultados_mensais =
movimentos_2025.groupby([movimentos_2025['Data'].dt.to_period('M'),
'Tipo_Valor'])['Valor'].sum().unstack(fill_value=0)
resultados_mensais['EBT'] = resultados_mensais.get('Revenue', 0) +
resultados_mensais.get('Expenses', 0)
resultados_mensais.index = resultados_mensais.index.astype(str)

resultados_mensais.to_excel("Resultados_Mensais_2025.xlsx")

print(f"☑ Cálculo concluído e guardado em
'Resultados_Mensais_2025.xlsx'")
```

Prompts List

Below is a list of prompts used during the preparation of the written work:

1. "Help me define and refine research questions related to the use of data analytics in management control."
2. "Suggest a possible structure or outline for a master's thesis on data analytics in a cultural organization."
3. "Summarize key concepts from the academic literature I gave you, on budgeting, forecasting, and business intelligence."
4. "Explain the difference between [insert] and [insert] in a Power BI context."
5. "Help me write or revise the DAX functions used to [insert objective]."
6. "Review this section for clarity and coherence: [insert paragraph]."
7. "Correct grammar and suggest improvements in style for this paragraph: [insert paragraph]"