Developing Marketing Analytics Tools in a Non-Profit Organization

Manuel Carlos Oliveira Rocha

Master's Dissertation

Advisor at FEUP: Prof. Luís Guimarães

U. PORTO

FEUP **FACULDADE DE ENGENHARIA**
UNIVERSIDADE DO PORTO

Integrated Master in Industrial Engineering and Management

2021-06-28

Abstract

One of the major struggles a Non-Profit Organization has to face is the one related to revenues. "Histórias da Ajudaris" is a project aimed at the participation of students from the preschool level up until the sixth grade. These students write several stories which are then submitted by the schools. The best stories are selected by Ajudaris and are then published in books. The sales of these books constitute a significant portion of the organization's revenues. The overall goal of the present dissertation was to find ways of making the project more efficient and more relevant among schools.

There were several opportunities for improvement. One included the need to automate the data collection and organization processes, as the manual tasks took too long and led to mistakes or data loss. Then, there was also room for improvement in the targeted marketing components to identify and approach the right targets.

The present dissertation was thus organized into two main components - data management and marketing analytics. In the first, there was the need to list every stakeholder of the project and all the relevant information to collect from each one. With all the information requirements, a database was created. There was also the concern of connecting the questionnaires' answers to the database with an automated data flow. Further in the project, it was crucial to automate the process of assigning tasks to specific stakeholders. The second part included identifying participants that left the project and potentially new participants that should be contacted. This is a targeted marketing approach towards guaranteeing a significant number of participants. The identification of new stakeholders included the development of clustering techniques.

The database and the data flow created allowed the automation of the data collection. Microsoft Access as a front end for the database facilitated its usage. Through the clustering process, 23 school groups were identified as relevant contacts. The remainder of the contacts management activities resulted in automated ways of contacting and attributing tasks to each stakeholder.

Future works can include ways of improving the clustering procedures. There can also exist the focus on the further stages of the project. This includes the analysis of the buyers of the books and options to improve the different selling channels. Then, it can be advantageous to develop a marketing campaign strategy to foster sales.

Resumo

Uma das maiores dificuldades sentidas por associações sem fins lucrativos está relacionada com as receitas que consegue obter. "Histórias da Ajudaris" é um projeto dirigido a alunos desde o nível da pré-escola até ao segundo ciclo. Os estudantes escrevem histórias que são submetidas pelas respetivas escolas. As melhores são selecionadas pela Ajudaris e são, posteriormente, publicadas em livros. A venda destes livros constitui uma porção significativa das receitas da associação. O objetivo geral da presente dissertação era o de encontrar maneiras de tornar o mesmo projeto mais eficiente e mais relevante entre as escolas.

Foram identificadas várias oportunidades de melhorias. Uma incluía a necessidade de automatizar os processos de recolha e organização de dados, já que os processos manuais ocupavam demasiado tempo e podiam conduzir a erros ou perda de informação. Posteriormente, também havia espaço para melhorias nas componentes de *targeted marketing* para identificar e abordar os alvos corretos.

A presente dissertação foi então organizada em duas componentes principais - gestão de dados e marketing analítico. Na primeira parte, havia a necessidade de listar todos os intervenientes do projeto e toda a informação relevante a recolher de cada um. Com os requisitos de informação, uma base de dados foi construída. Também havia a preocupação de ligar as respostas dos questionários à base de dados com um fluxo automatizado. Numa fase posterior, foi importante automatizar o processo de alocar tarefas aos intervenientes. A segunda parte incluiu a identificação de participantes que deixaram este projeto e de potenciais novos participantes que deveriam ser contactados. Isto diz respeito a uma abordagem de *targeted marketing* para garantir um número significativo de participantes no projeto. A identificação destes intervenientes incluiu o desenvolvimento de técnicas de segmentação (*clustering*).

A base de dados e o fluxo de dados criados permitiram a automatização do processo de recolha de informação. O uso do Microsoft Access como interface para o utilizador facilitou o uso da base de dados. Através da segmentação, 23 agrupamentos de escolas foram identificados como contactos potencialmente relevantes. As restantes atividades de gestão de contactos resultaram em maneiras automatizadas de contactar e alocar tarefas a cada participante.

Trabalhos futuros podem incluir melhorias ao processo de segmentação. Também pode existir um foco em fases mais avançadas do projeto. Isto inclui a análise dos compradores dos livros e oportunidades de melhoria dos diferentes canais de venda. Pode ainda ser vantajoso desenvolver uma estratégia de campanha promocional para potenciar as vendas dos livros.

Acknowledgments

Contents

Acronyms and Symbols

KPI   Key Performance Indicator
NPO   Non-Profit Organization
ODBC Open Database Connectivity
SQL   Structured Query Language
UML   Unified Modeling Language
VBA   Visual Basic for Applications

List of Figures

List of Tables

Chapter 1

Introduction

The dissertation project was developed with FEUP within the Integrated Master in Industrial Engineering and Management and in cooperation with Ajudaris, a Non-Profit Organization (NPO) based in Porto.

This first chapter introduces the organization, as well as the project in more detail and the motivations behind it. Furthermore, the goals and methodology of the project are approached and the structure of the remaining of the dissertation is presented.

1.1 The Organization - Presenting Ajudaris

Ajudaris is a non-profit organization founded in 2008 which is focused on addressing societal problems such as hunger, poverty and social exclusion. The organization's mission is to increase the quality of the lives of those that benefit from it while tackling loneliness and promoting social interaction within the society. To do so, Ajudaris regularly organizes several events or initiatives involving schools (with competitions to motivate and/or educate students) and families or people in need (with the collection and distribution of food, for example). The organization relies on different types of income sources, including the sale of products through the online store, individual donations, membership fees, partnerships with companies and volunteerism. Besides these contributions, there is also the book "Histórias da Ajudaris", which is going to be further developed in the next sections (Ajudaris 2021a).

1.2 The Project "Histórias da Ajudaris"

The core idea of the project "Histórias da Ajudaris" is the creation and sale of books containing small stories written by young students from schools that apply to enter the competition. By selling the resulting books, Ajudaris guarantees one of its main revenue streams necessary to sustain its mission and activities. This way, Ajudaris can appeal to communities' sense of volunteerism and, simultaneously, incentive or strengthen reading habits, creativity and citizenship of young

children. "Histórias da Ajudaris" is directed for students from preschool until the sixth grade (Ajudaris 2021b).

For a school (or school group) to participate in this project, it has to fill in and submit an initial form containing important information, such as the address and relevant contacts. Later on, and until an established deadline, these schools need to submit their students' stories to be evaluated and graded by a jury selected by Ajudaris. These stories should be written in groups (of students supported by teachers) and should address a specific theme that Ajudaris establishes each year. According to some evaluation components, the judges select the best stories among the ones received to be published in one of the books that will enter the market. Then, the chosen stories are illustrated by solidary artists and divided into different book volumes, according to the region of the schools. Finally, the books are sold through three main channels - online shop, Pingo Doce stores and school events (Ajudaris 2021b).

1.3   Project Motivation

The idea for the present dissertation was initially developed following an approach made by Ajudaris, which wanted to address a specific problem within the project "Histórias da Ajudaris". Starting from there, a broader project was established to help the organization tackle more of its issues, which were aggravated in the context of the pandemic.

"Histórias da Ajudaris" is a fundamental project for Ajudaris and, as years go by, some schools already get to know it and decide to participate regularly without having to be contacted by the organization. However, Ajudaris still needs to reach out to several schools every year to guarantee that the project has a considerable number of participants.

This contact can be made to bring in new schools or to bring back some schools that have already participated but did not renew the application. This is where one of the main problems exists. Up until this point, the usual process for Ajudaris was to collect information from the participants and to add it manually to a separate spreadsheet. Thus, from year to year, some of the data from schools and individuals was being lost, which complicated the process of contacting them. Therefore, it became relevant to address this situation and improve the way information was being collected and stored for further use.

A more effective usage of data can allow the organization to perceive the project as a whole and to save precious time for its key activities within the community. This involves not only the collection and storage of data (so that the process is more automated and the organization does not lose any records), but also the way new potential participants are identified and contacted. The latter is about continuously trying to guarantee more and more participants each year. By focusing on these issues repeatedly, decreases in the number of participants during a specific edition can be avoided more easily.

Not only the schools participate in the project. Other stakeholders, such as judges or illustrators, intervene and have to be contacted by Ajudaris. The process of assigning tasks (i.e. sending stories) to these stakeholders involves a high degree of complexity. The facts that there are several

constraints affecting the process and that all the contacts were, up until this moment, made individual and manually justify the previous statement. So, it also became important to address these issues to guarantee more effective and intuitive ways of assigning tasks to stakeholders and then contacting them.

Having this said, the current dissertation aims to address problems related to "Histórias da Ajudaris" so that the organization can manage the project in a more efficient manner.

1.4   Project Goals

Considering the main challenges mentioned above, the same type of division can be established for the goals of this project.

First, a data management component aims at helping the organization to have a clear understanding and complete perception of the project "Histórias da Ajudaris". This demands that the organization recognizes every single group of stakeholders involved in the project as well as their roles in each of the project's stages. Only that way it is possible to list the information needed regarding the interaction between each stakeholder and the project itself. Ultimately, the main objective beyond the clear understanding of the project is that Ajudaris is able to keep track of and organize all the data collected, thus being able to contact more schools and guarantee more participants each year. This can be achieved through a more organized and automated database that improves the methods currently used by Ajudaris. The goal here is to make the data collection and storage more automated while also developing a new database, making use of the organization's resources.

Secondly, the focus on the concept of marketing analytics has one main concern. At an early stage of the project "Histórias da Ajudaris", the goal is to get more schools to participate, either by identifying new contacts with higher chances of participating or by avoiding or reducing the churn of previous participants. Therefore, in terms of schools/school groups, it is necessary to analyze and understand the differences between these stakeholders so that the right targets are identified. So, the objective is to understand the processes of segmenting stakeholders and making contacts (mainly related to schools).

There is also a concern relating to other stakeholders, such as the judges, reviewers and illustrators. Here, the automation and management of the contacts made are part of the marketing analytics component. So, the goals refer to the way contacts are managed and established throughout the project - for schools it is necessary to identify the desired profiles before making contacts and for other stakeholders (judges, for example) the objective is to make the contacting process more efficient and automated.

1.5   Methodology

The project started with a component of literature review of concepts such as Marketing, Marketing Analytics and Non-Profit Organizations, among others.

Moving on to the more practical part of the project, the plan was to divide it into the main stages already mentioned. The first involves mapping the different stages of the project, as well as all the stakeholders involved in each stage and how they intervene. From here, a clearer perception of the structure of a possible database for the organization to keep data from stakeholders was achieved. So, the following step was to organize this information into a Unified Modeling Language (UML) diagram, which served as the basis for the said database. It was also necessary to guarantee a structured and automated data flow.

After the information management and database development stage, tasks were more directly related to the marketing analytics components and made use of the information from the previous steps. It was necessary to clearly define the process of contacting schools - the ones that leave the project and the ones that are targeted as potentially new participants. This part involves segmentation and targeting of these stakeholders and the further automation of the way contacts are made.

Then, the focus should be on the latter parts of the project, more specifically on the judges, reviewers and illustrators of the stories. Stories are allocated to them and this task is to be developed and automated.

1.6   Structure of the dissertation

The second chapter of this dissertation consists on the state-of-the-art around the issues of Marketing, Non-Profit Organizations and Marketing Analytics, among others. So, several sources of literature are reviewed to present those same topics. In terms of the application of marketing concepts to the non-profit sector, other issues such as stakeholders' segmentation and online campaigns are addressed through existing literature.

In Chapter 3, the current situation and functioning of the project "Histórias da Ajudaris" is presented in more detail. The process is mapped and all the stakeholders of the project are listed. Also, the main problems from the current situation are described.

Chapter 4 consists of the definition and proposal of a solution to the mentioned problems. The organization of a database and the tools used to design and implement it are also detailed. Furthermore, a proposal regarding contact management and approaches to different stakeholders is described. This is preceded by an analysis of the evolution of the participation in the project in terms of schools and school groups.

Finally, Chapter 5 includes the conclusions regarding what was described in the remaining chapters, as well as an analysis of future work and potential improvements to what was achieved during the project that originated this dissertation.

Chapter 2

State of the Art

This chapter aims at introducing some of the key concepts for this particular project. To do so, existing literature approaching these issues was reviewed and is presented below.

The first main topic reviewed is related to the non-profit sector and its organizations. Secondly, a general overview of the marketing concept is made and followed by a more specific approach to marketing analytics. The third main issue relates to the application of marketing in Non-Profit Organizations (NPOs). This third part also considers the application of marketing analytics, stakeholder segmentation and Customer Relationship Management (CRM) processes. Finally, the subject of social media campaigns is also included in this chapter, with particular focus on Facebook, and is followed by references to some analytical methods or procedures relevant to the context of this dissertation.

2.1 Non-Profit Organizations

As stated by Anheier (2005), the non-profit sector includes different types of organizations besides the non-profit ones: private and voluntary organizations. This definition can be extended to specify the type of organizations and associations that belong to this sector: the non-profit sector includes entities that are organized, private, not profit-distributing, self-governing and voluntary (Campos et al. 2005).

In terms of funding, there are some key differences between NPOs and for-profit organizations. The latter tends to fund their activities with bank loans, investors and sales of products/services, while NPOs usually benefit from donations (of time or money), sponsorships from companies or even government grants. Moreover, for-profit companies intend to create profit and usually have a clear target audience. On the other hand, NPOs focus more on social causes rather than profits and their audience is broader - includes stakeholders like volunteers and donors, among others (Norwich University Online 2016).

Some other revenue sources can be listed for NPOs. Organizations might also get revenue from the sales of products or services, federated funds, fundraising, membership fees, crowdfunding and other types of events. Another important contribution for organizations is the work from

volunteers, which technically is not a revenue income but is fundamental for the functioning of any of these organizations (Young 2007; Fritz 2019; GoFundMe Charity 2019).

Foster, Kim, and Christiansen (2009) contributed to this matter by classifying NPOs into ten different funding models, which are listed below:

- Heartfelt connector - organizations that grow by addressing causes that are meaningful to a large portion of the population;
- Beneficiary builder - institutions, such as hospitals and universities, that count on donations from people who have already benefited from them;
- Member motivator - rely on contributions from people that feel benefited by the services (e.g. churches);
- Big bettor - rely on major contributions from a few donors;
- Public provider - receive previously defined funding from government agencies to provide social services;
- Policy innovator - are funded by the government, but outside the usual programs;
- Beneficiary broker - organizations that compete to deliver services compensated with government funds;
- Resource recycler - collect donations and distribute those through other organizations;
- Market maker - organizations that operate in areas where it could be unlawful or unseemly to have for-profit businesses (e.g. organ donations);
- Local nationalizer - deal with local issues across the country and rely on local financial support.

Having considered the possible revenue sources for a NPO, another concern emerges: should NPOs focus on a single source of revenue or diversify and consider several sources? As Chang, Tuckman, and Chikoto-Schultz (2018) conclude, the diversification of revenue sources may carry several benefits but, at the same time, can involve risks for the company.

From one side, there are studies suggesting that diversification leads to a stronger position for organizations. Chang and Tuckman (1994) present evidence that support the perspective that companies with more sources are in a financially stronger position. The authors also refer that diversification can help organizations achieve objectives more efficiently and be more stable because of a lower financial risk. Nevertheless, they also recognize, from a social perspective, that diversification can lead to a loss of control over donors.

From the opposite perspective, there are also articles concluding that diversification affects NPOs negatively, as some sources of revenue might not be compatible with the association's mission or even because some NPOs do not have sufficient skills to deal with multiple sources (de los Mozos, Duarte, and Ruiz 2016).

2.2 Marketing Analytics applied to NPOs

2.2.1 Introduction to Marketing and Marketing Analytics

According to the American Marketing Association (2017), marketing can be defined as "the activity, set of institutions, and processes for creating, communicating, delivering and exchanging offerings that have value for customers, clients, partners and society at large". This definition, revised in 2017, is the most recent one and recognizes, since 2007, the sector of NPOs and its importance within the concept of marketing (Ahmed 2013). Generically, the goal of marketing can be to bring in "new customers by promising superior value" and, simultaneously, to try keeping or even growing "current customers by delivering value and satisfaction" (Kotler and Armstrong 2018, p. 27).

Like all other activities within a company/organization, marketing's performance should be monitored to allow potential corrections or improvements in the future. That is where marketing analytics comes into play. Marketing analytics can possibly be defined as the "process of identifying metrics that are valid indicators of marketing's performance in pursuit of its objectives" - these metrics are the ones that need to be followed to assess the effectiveness of marketing related activities (Rackley 2015, p. 2). The same author states that without analytics a certain marketing organization acts out of pure instinct - "flying blind".

The adequate metrics for a company obviously depend on each one's situation and strategy. Nevertheless, it is possible to list broad categories of metrics that can be used by most (if not all) companies. The first one is related to the marketing process and consists of metrics that must indicate the productivity of the company. Secondly, metrics oriented to customers are vital for any organization and can include concepts such as customer satisfaction or customer churn. The third category relates to lead generation. Finally, there can also be metrics related to digital marketing, which includes metrics regarding websites (e.g. traffic sources), social media (e.g. number of followers), email marketing (e.g. open rate) and digital advertising (e.g. conversion rate) (Rackley 2015, chap. 5).

Germann, Lilien, and Rangaswamy (2013), in a study about the implications of marketing analytics, conclude that it has a positive impact on a firm's performance. In addition to that, it is mentioned in the same book that marketing analytics can be an important tool to achieve competitive advantage over other companies within the same industry.

An article by Kotler and Levy (1969) was one of the first to recognize that the concept of marketing should not be applied exclusively to firms in the for-profit sector, as it is stated that all organizations need to deal with activities within the scope of marketing. The authors conclude by saying clearly that "no organization can avoid marketing" (Kotler and Levy 1969, p. 15), even the ones outside the scope of business firms.

As Kotler (1979) explains, the paper written ten years before generated quite some controversy and he adds that it took some time until new studies about marketing in the non-profit segment were developed. The conclusion of this paper is that businessmen were starting to understand the

importance of marketing in organizations and that marketing would "lead to a better understanding of the needs of different client segments" (Kotler 1979, p. 44).

Marketing has been gaining importance in the non-profit context ever since. This need for marketing in the non-profit sector can be a consequence of several factors, the first one being the assistance marketing activities can provide to address financial issues (low funding, for example). Moreover, NPOs can create awareness through marketing initiatives. Another important reason is the increased competition NPOs might face, which illustrates the importance that marketing strategies can have towards the survival of an organization (Ahmed 2013; McLeish 2010).

2.2.2   Marketing in the non-profit sector VS Marketing in the for-profit sector

As expected, there are some noticeable differences between the marketing process in the sectors of non-profit and for-profit. The first difference relates to the nature of products, as NPOs offer mainly services in detriment of physical products. Secondly, another difference can be related to the public scrutiny, mainly due to the tax status that these organizations have. Finally, NPOs need to account for multiple stakeholders, including people that are served by the organization, donors, staff and volunteers. As a consequence of the presence of these stakeholders, NPOs usually have multiple objectives as well (Ahmed 2013).

According to Bruce (1995), there are two groups of customers served by NPOs. The first one is composed mainly by the stakeholders mentioned above, which the author calls end customers. The second group mentioned in the same paper refers to the intermediary customers (including, for example, government agencies that may lead clients to specific associations), which are essential in a way that allow organizations to reach its prime customers.

While the end goal of for-profit organizations is, above anything else, the creation of profit, that is not the case of non-profit organizations. As it was already mentioned, these organizations usually have several different goals, which can include increasing the number of donations/donors, raising funds, attracting more volunteers or even creating awareness into the organization or its causes (Decker 2021).

The presence of multiple stakeholders impacts the way an organization has to work. It's fundamental for a NPO to always maintain contact with its stakeholders through the right channels and in the correct moments/contexts. This is even more relevant when talking specifically about donors, as NPOs must understand the motivations these individuals have to donate and then develop marketing strategies that address these motivations (Pope, Sterrett Isely, and Asamoa-Tutu 2009).

In a general way, communication with stakeholders can be done through social media, newsletters, special events or even content marketing. Each organization has to study and understand the different stakeholders and segment them into separate groups with specific characteristics, so that messages are crafted to each one of the groups (Decker 2021).

Having discussed the importance of marketing in non-profit organizations, it becomes clear that marketing analytics will also be fundamental to track the organization's performance. In a report titled "Fundraising and analytics", marketing analytics is said to have "the combined goal

of enhancing the constituent experience and optimizing the non-profit's marketing and fundraising resources" (Pritcher and Patrick 2015, p. 3). These authors justify the emergence of marketing analytics in this sector with different motives. The first is related to the increased number of channels through which associations must engage with individuals, which makes the collection of data more complex. The second is that analytics make computation of data much more efficient. Then, as NPOs are growing and diversifying, marketing analytics is an important tool to "maximize the effectiveness of spend across channels" (Pritcher and Patrick 2015, p. 4).

The inclusion of analytics is intended to answer several questions to help the well functioning of the associations that use it. By using the correct metrics, organizations can separate donors into two groups (the ones they want to keep and the ones that should be discarded). Furthermore, analytics can also be used in this context for associations to understand where revenues might come from and, by continuing to analyze stakeholders, understand why donors donate (i.e. their motivations to give money to a certain cause or association) (Bennett 2007).

In addition to this, marketing analytics help identify the audience (types of donors, for example), the correct offer (what the association presents to its audience) and also the correct moment and channels for the engagement to happen (Pritcher and Patrick 2015).

In this context, McLeish (2010) mentions what he believes NPOs should measure. That list includes concepts such as growth, employee welfare, brand strength, website traffic, financial stability and product quality.

More specific metrics that might be useful for any NPO are listed below, divided into four groups - web analytics; email analytics; fundraising and donor analytics; business and financial analytics (DNL OmniMedia Inc. 2018).

- Web analytics - website traffic sources, average session duration, conversion rate, Google AdWords Analytics;
- Email analytics - open rate, click-through rate, conversion rate;
- Fundraising and donor analytics - retention rate, major gift indicators, demographics, online giving rate;
- Business and financial analytics - non-profit revenue reliability, cost per dollar raised.

2.2.3   Stakeholder segmentation in NPOs

Before discussing the concept of segmentation specifically in NPOs, it is important to introduce the broader concept of market segmentation - it consists of dividing potential customers of a company into different groups. Each resulting group should be composed by customers with similar characteristics, interests and needs (McDonald and Dunbar 2012, p. 14).

By dividing the market, a company can address the smaller segments in a more efficient and effective way by being better positioned and corresponding to the needs of specific consumers. When segmenting consumer markets, four main variables can be considered - geographic, demographic, psychographic and behavioral variables. Demographic variables might include age,

gender and income, while behavioral variables are composed of issues such as occasion, user status, usage rate or even loyalty status (Kotler and Armstrong 2018, p. 212).

Specifically regarding NPOs, segmentation of stakeholders is fundamental as well. As it is mentioned by Randle and Dolnicar (2009), NPOs are continuously trying to connect with the right group of volunteers rather than just increasing the number of volunteers. This is a development in terms of marketing techniques for NPOs, for which the two authors consider the increasing competition in the non-profit sector as the main cause.

The segmentation of stakeholders can thus be seen as a crucial marketing activity for NPOs to perform and function in a sustainable way. As presented by Shelley and Polonsky in an article from 2002 entitled "Do charitable causes need to segment their current donor base on demographic factors?: An Australian examination" of the volume 7 of the International Journal of Nonprofit and Voluntary Sector Marketing (quoted in Srnka, Grohs, and Eckler 2003), the competition within the non-profit sector has been increasing and donors receive more solicitations to help. This fact can strengthen the argument that organizations need to segment their stakeholders to overcome problems brought about by the competition. As mentioned in the same article by Shelley and Polonsky (quoted in Srnka, Grohs, and Eckler 2003), this happens because segmentation allows NPOs to understand the different types of donors and align its messages to each group (Srnka, Grohs, and Eckler 2003).

Going from the four types of variables listed by Kotler and Armstrong (2018), Rupp, Kern, and Helmig (2014) specify five factors that can be used to segment stakeholders of a NPO. The first one agglomerates demographic and socio-economical variables, being designated by socio-demographic segmentation - this is a term used, for example, by Van Slyke, Ashley and Johnson in the article "Nonprofit performance, fund-raising effectiveness, and strategies for engaging African Americans in philanthropy", of the volume 37 from September 2007 of the American Review of Public Administration (quoted in Rupp, Kern, and Helmig 2014). Then, they consider the same psychographic and behavioral variables as Kotler and Armstrong. The fourth factor is related to the financial value of a stakeholder. Finally, the fifth includes the possibility of having more factors not contemplated by the previous four – segmentation criteria designated as other.

Besides some of the already mentioned criteria (e.g. geographic or demographic variables), a report by Abercrombie, Bagwell, and Baumgartner (2014) also lists the following:

- Attitudinal segmentation - relates to values and reasons to support an association;
- Selection segmentation - involves segmentation and targeting of groups that are more likely to respond to a campaign;
- Channel preference segmentation - considers communication channels stakeholders prefer.

Moreover, as mentioned by Moore in a work entitled "Organizational Imagery and Voluntarism" in the Managing Voluntary Organizations Conference Proceedings from October 1983 (quoted in Harvey 1990), there are even more approaches NPOs can and usually follow regarding marketing segmentation. Some of the new criteria, when comparing with the previously mentioned literature, are listed below:

- Type of donation - can include money or even assets;
- Amount of the donation - refers to the value donated;
- Temporal dimension of the donation - considers that donations can be made through bequests or funds, for example;
- Giver class - can involve individuals and companies, among other types of donors.

This concept of segmentation can also be included in a broader concept that all organizations (both in the for-profit and non-profit sectors) need to consider, which is Customer Relationship Management (CRM). The main intention of organizations when considering CRM is the one of developing and maintaining relationships with customers. To do so, an organization has to study its market and all the stakeholders in it so that it can adapt to the needs and wants of specific segments - mainly through segmentation, as mentioned above. By adapting to these needs and providing customized solutions, a closer relationship and stronger sense of loyalty can be created between organizations and stakeholders. Furthermore, to succeed in a CRM initiative, it is also fundamental that a strategy is well developed beforehand and the entire organization is involved in the initiative. Also, CRM technology should not be perceived as the strategy itself and, above all, the organization needs to clearly understand its customers (Kristoffersen and Singh 2004).

A possible definition of CRM is the one that describes it as the "mode of operation deployed to maintain and develop relationships with profitable customers, and manage the cost of doing business with less profitable customers" (Foss and Stone 2002, p. 14). This reinforces the power of understanding customers and to address efforts to those more likely to respond to a particular cause (Foss and Stone 2002; Polonsky and Sargeant 2007).

2.2.4 Social media campaigns for NPOs

After the segmentation process, it is necessary for any company or organization to connect with the identified stakeholders through the right channels. One of the most commonly used channels is social media, as the majority of organizations are using it to establish/maintain relationships with stakeholders. This is particularly important for non-profit organizations, mainly because these depend on donations, volunteerism or other forms of public support. In this context, NPOs can use social media to create awareness towards a specific cause or to raise funds (through donations). Besides that, creating more powerful relationships with stakeholders by sharing information and stories and establishing communications is viewed as a fundamental virtue of social media in this sector (Auger 2013; Lozano, Balonas, and Ruão 2020; Young 2017; Carboni and Maxwell 2015).

Obviously, different social media platforms can have different benefits for an organization, due to the types of publications that can be made on each. The specific case of Facebook can be related more with a two-way communication strategy with an organization's followers, which can lead to a stronger connection between organizations and the ones that are targeted by this Facebook content. This interaction can happen in different ways as well, as Facebook users can engage with an organization's publication by just liking it, commenting or even sharing it with others. This

might be one of the reasons why some consider Facebook as the most used social media platform by organizations in general (Auger 2013; Cho, Schweickart, and Haase 2014; Lucas 2017).

2.3   Methods for NPO marketing analytics

Having explored concepts such as marketing analytics (and its importance for NPOs), it is essential to list and explore some specific methods that can be used in this context.

The upcoming methods are thus fundamental to the activities related to marketing analytics within the context of NPOs. These include prescriptive, predictive and descriptive methods. For each one of the methods, some examples are given and the utility of each in the context of marketing analytics is also approached for some of those.

2.3.1   Clustering and K-Means

Clustering is an unsupervised classification process used to divide data into separate groups according to similarities. The goal is that the data objects within one group (cluster) are as similar as possible to each other and as distant as possible from objects or records belonging to a different group so that the groups created are clearly distinguishable and separated from each other (Tan et al. 2019; Jain, Murty, and Flynn 2000).

There are several different clustering algorithms used in different contexts. The algorithms are divided into distinct categories as well. These categories include partitional clustering, density-based clustering and hierarchical clustering. The main example of algorithms belonging to the first category is K-Means clustering. DBSCAN is an example of density-based clustering and agglomerative hierarchical clustering can illustrate the last category mentioned above. To these categories, one can also include grid-based, model-based methods and constraint-based clustering methods (Tan et al. 2019; Velmurugan and Santhanam 2011).

One of the most important methods from all the categories considered is K-Means. K-Means is one of the partitioning methods and divides the data into a predefined set of K clusters. This K has to be defined by the user before the execution of the clustering itself. The right K can be found through several methods. One consists of a plot of the Sum of Squared Errors (SSE) against the number of clusters (K) and another uses the silhouette coefficient as measure. In these cases, both measures are plotted against the number of clusters (K) - Figure 2.1 illustrates the first measure.

Furthermore, K initial points (centroids) are then established. All the data points in analysis are assigned to the nearest centroid. Based on the groups created, the centroids are redetermined and the process is repeated (i.e. the data points being assigned to the nearest centroid) until no modifications occur in these steps. The proximity between points can be calculated with different methods. The most common one is probably the Euclidean distance, but there are other measures such as the Manhattan distance that evaluate the desired proximity. By performing the mentioned steps, the target is to minimize the distance of each point to the centroid it has been assigned to. K-Means, like any other method, has its own advantages and disadvantages. It is one of the easiest methods of clustering and can be very efficient as well. However, it does not handle clusters
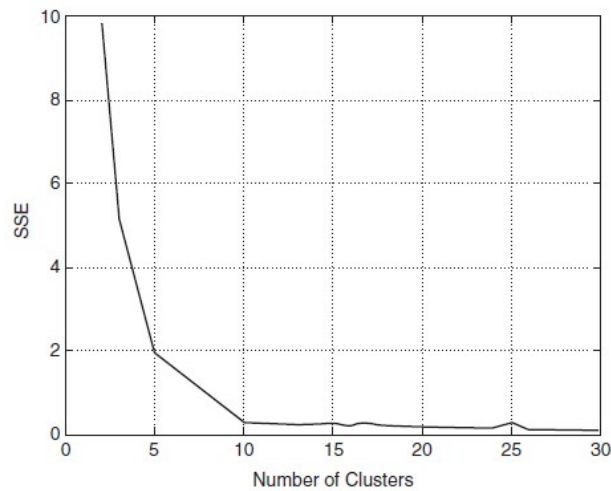
Figure 2.1: Determining the right k (SSE versus number of clusters). Source: Tan et al. (2019)

containing different shapes and also outliers as well as other methods, which can be more difficult to be applied (Tan et al. 2019; Kanungo et al. 2002).

One way of evaluating the quality of the clustering process is through the already mentioned silhouette coefficient. This measure considers the cohesion of the clusters created in the process and also how well the clusters are separated from each other. The value of the coefficient varies between -1 and 1. To indicate a good clustering, the silhouette score should be above 0 and, preferably, close to 1. A negative value would indicate that one data point could be closer to the centroid of a different cluster than the one from the cluster it was assigned to (Tan et al. 2019).

The clustering procedures described are of great importance in the context of marketing analytics. By segmenting stakeholders (i.e. dividing stakeholders into separate clusters), it is possible to improve the performance of marketing related activities. This segmentation can help identifying the right targets of the marketing initiatives, thus enabling a better allocation of efforts and resources of a company or organization. Instead of targeting a vast group of stakeholders or customers, a more restrict set of targets can then be identified (Yang et al. 2016).

2.3.2 Logistic Regression

Logistic regression is a supervised classification algorithm that uses as inputs a set of independent data variables to predict the desired dependent variable. When compared to linear regression models, the logistic regression takes the logit transformation of the variable to be predicted (the dependent variable). This means that the values predicted are all within the interval between 0 and 1 (i.e. the values are probabilities). If the classification problem has more than one class that can be predicted, there is a particular case of this logistic regression model - the multiple (or multinomial) logistic regression. This is also a supervised algorithm that allows to classify and distinguish between more than two classes by taking an input of independent variables that are

used as predictors of the dependent variable (Brownlee 2021; Erdfelder et al. 2009; Krishnapuram et al. 2005; Karp 1998).

Logistic regression can be very useful within marketing (analytics) initiatives. It is one of the models that can be used to improve the quality of the information provided by a research in the marketing context (Constantin 2015). Some more specific examples of this utility include the fact that this method can be applied to create models aimed at increasing customer retention. Moreover, logistic regression sheds light on issues surrounding the comprehension of customers behavior. This can then be used to improve marketing strategies, which strengthens the importance of logistic regression within the context of marketing analytics (Karp 1998).

The applications of logistic regression include studies made on customers (regarding customer behavior, for example) and other more business oriented concerns. The latter includes different types of decisions made within a marketing strategy (Akinci et al. 2007).

2.3.3 Feature Selection

Feature selection is one method that enables the improvement of one classifier through the selection of the most appropriate features or variables within a model to predict one desired dependent variable. By selecting the most significant variables or reducing the dimension of data sets, the performance of the classification model can increase. This improvement can be the result of a higher accuracy of the classification model or even the reduction of the time required to train the model in question, among other reasons. There are several techniques within this feature selection methods, among which one can mention the L1-based feature selection and the Chi-Square based feature selection. In the first of these two alternatives, the model attributes weights to data variables according to the importance they have in predicting the dependent variable. Weights of zero are attributed to those variables with low or none significance in the model and higher weights to those that have higher importance. The second alternative is based on a statistical procedure that tests the dependence between two variables - one of the independent variables and the variable that is being predicted (the dependent variable). The variables without any dependency with the dependent variable are thus eliminated from the model, reducing the number of independent variables (scikit-learn - Machine Learning in Python 2020; Mushtaq et al. 2020; Abbad Ur Rehman, Lin, and Mushtaq 2021).

By removing from one model those variables that do not add to the performance of the predictions, one can understand the impact this method can have in the marketing context. Moreover, there is a trade-off when selecting features to improve the performance. There should be a balance, in the sense that, in most of the situations, one should not select too few or too many variables to feature in the model (Tan et al. 2013).

2.3.4 General Assignment Problem

General assignment problems (GAP) are, in a general sense, the types of problems that allocate different tasks/jobs/time slots (among other possibilities) to different workers/machines (among

others). The assignments are subject to a set of constraints and, most of the times, to an objective function. This objective can be, for example, the minimization of costs or the maximization of profits. The constraints can reflect production capacities of workers or machines, time constraints or other limitations that are specific to every single problem and context (Ramesh and Razak 2019; Feltl and Raidl 2004; Chauvet, Proth, and Soumare 2000).

Chapter 3

Problem Description

This chapter aims to explain the current situation of Ajudaris when it comes to the specific project "Histórias da Ajudaris". A more thorough analysis of the current situation is made and is followed by the recognition of the main problems and improvement areas.

3.1   The project's current situation

"Histórias da Ajudaris" is one of the main revenue sources for the organization and is centered on stories written by students and submitted by the respective schools to contest. Then, the best stories among the ones received by Ajudaris are selected and edited in books.

To truly understand the way this project works and the amount of data involved, the entire process was mapped. Hence, the first step was to list all the stakeholders involved in some way with the project (i.e. everyone that participates/interacts with the project at some point in time). Besides Ajudaris, schools/school groups are some of the most important stakeholders.

The organization contacts directly with schools and receives the stories and other requested data from them. Each one of the schools has at least one interlocutor for the project, which is the person that is responsible for the development of the project within each school and is the one that Ajudaris contacts when necessary. Another obvious group is composed by students and teachers, as they are the ones that write the stories sent out to the contest. Here, parents of participating students can also be mentioned in the sense that they can be the ones contacting Ajudaris for their children's schools to participate. The next set of actors relates to judges and illustrators. As expected, judges are the ones that receive the stories from Ajudaris and grade them so that Ajudaris can select the best stories to be in the books. The illustrators are given stories from the group selected and illustrate each one. Moreover, a broader group of stakeholders was considered and denominated with the general term "partners". These partners include companies, institutions or individuals that contribute to the development of the project in some way. Some examples of partners are the reviewers of the stories, graphic designers, distributors, municipalities and other types of sponsors and supporting companies. The last set of stakeholders is more involved in the

selling part of the process and include the Pingo Doce's stores and, obviously, the buyers of the books.

From all these stakeholders, the first contact (the one that initiates the participation of a school in the project) can be made by Ajudaris, schools or parents. Every year, Ajudaris reaches out to several schools to get more participants. Nevertheless, some schools are already familiar with the project and, as a consequence, apply without having to be contacted. It can also be the case that some schools or teachers get to know the project from other schools. In the case of parents, they can directly contact Ajudaris (by email) if they want their children's school to get to know "Histórias da Ajudaris". In the case of illustrators, the initial contact can be made either by Ajudaris or illustrators themselves. The illustrators that get to know the project in some way and want to help and participate can have the initiative to contact the organization. From the other side, Ajudaris can also obtain the contacts of potentially interested illustrators through suggestions from other illustrators or even schools.

Before the beginning of the COVID-19 pandemic, the normal procedure would be to accept registrations until the end of January and submission of stories until the end of March. Then, stories would be evaluated and selected throughout April and the illustrating procedure would happen in May. From June to September books would be produced, published and prepared to be sold through three main channels - online shop, Pingo Doce's stores and school events. Obviously, 2020 and 2021 were atypical years, in the sense that some deadlines had to be delayed and several events cancelled. The different stages of the project are further explored in this chapter.

When it comes to numbers, the normal scenario involves printing around 20 000 books per year and it is not normal for Ajudaris to sell all of those during one edition. Currently, most of the sales occur in events organized by schools, with both the online shop and stores falling behind expectations. Moreover, in each edition there are approximately 900 schools all around the country and even from abroad participating. The selected stories among the ones submitted by these schools are divided according to the schools' regions. So, different volumes of the book are produced, each one containing around 100 stories from specific regions.

3.1.1 Project's phases and stakeholders

To summarize who are the stakeholders involved in the project and when they intervene, a responsibility matrix was developed and is shown in Figure 3.1. This can be an important tool to visualize all of those involved and to serve as the basis for the next step of the analysis - to deconstruct each one of the identified phases into more detailed tasks.

On the left side of the matrix, the already mentioned stakeholders are listed. On the top, six main stages were identified within the context of "Histórias da Ajudaris" and are explored in this section. On the bottom of the matrix there are milestones, which represent the outputs of each one of the phases. The transition from one phase to another only happens when the milestone is achieved. Finally, the dots placed on some of the lines' interceptions identify which stakeholders participate in each phase of the project. For each of the phases, a more detailed swimlane was
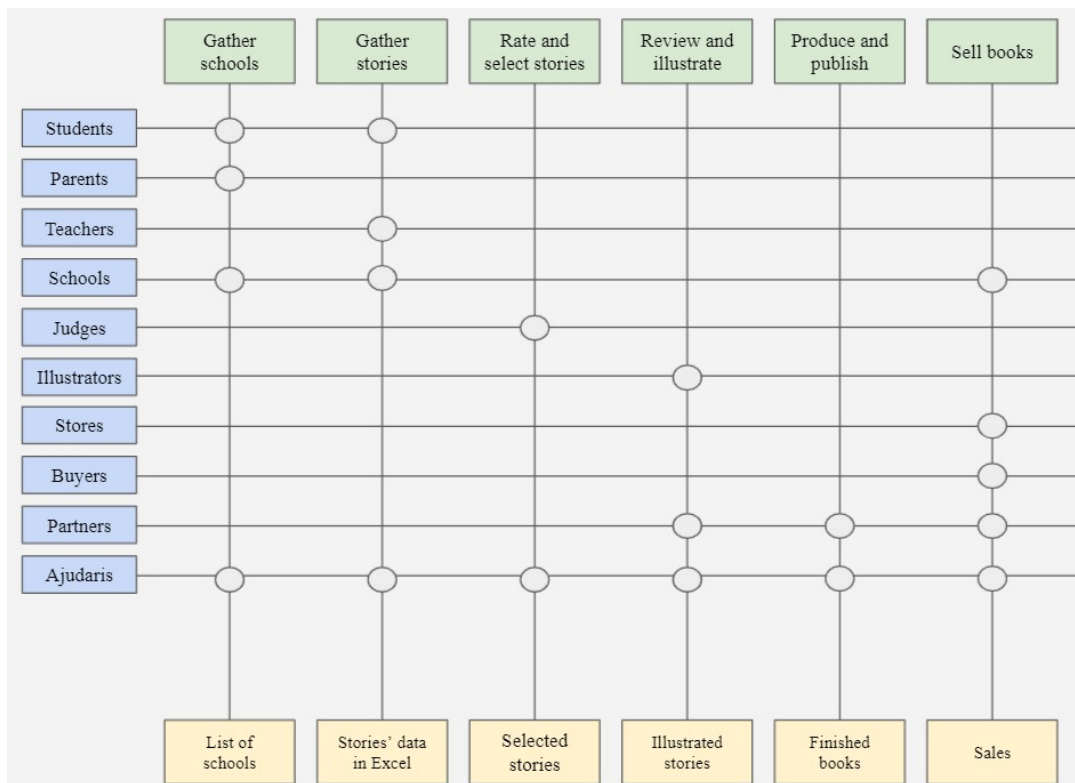
Figure 3.1: Responsibility matrix of the project

created using the software Bizagi Process Modeler. These swimlanes are presented in Appendix A. In all of the 6 figures, Ajudaris is referenced as NPO (Non-Profit Organization).

Phase 1 - Gather schools

As it was mentioned, and also according to Figure 3.1, the first contact can be made by parents or even students. However, the core of this phase is performed by Adjudaris. The initial list of schools can either include the ones that have already participated and for which the organization already has some information or some other schools that have never participated. In the second case, those schools can be suggestions from other participants or can be found in official lists of schools in Portugal. From there, Ajudaris has to analyze schools and choose which ones to contact (this will be approached in the following chapter).

   If a specific school is not eligible for the project (due to the students' ages, for example), it is discarded from the possible contacts. Then, the organization faces two scenarios, depending on if the school being considered has ever entered the project or not. If a school has participated before but did not apply for the newest edition of the project, it is of the best interest of Ajudaris to recontact and remind about the project. If a potential new participant is on the table, it is necessary to establish contact to pitch the idea for the first time and convince the school's representatives to participate.

From this point forward, the main actors are the school groups. If the schools' representatives agree to participate in the project, they have to fill in and submit a Google Forms, made available by Ajudaris, containing some of the following information:

- Name, address and relevant contact information of the school group;
- Name of the school group's director;
- Name, address and relevant contact information of the participating schools;
- Name and relevant contact information of the schools' interlocutors;
- Name and relevant contact information of the schools' participating teachers;
- Relevant contact information of the parents' association;
- Contacts' suggestions.

This initial Google Forms is filled in by school groups, which means that one school group only submits one answer, even if it has more than one school participating in the project.

Phase 2 - Gather stories

After the registration deadline, it is up to the teachers and their students to write their stories according to the predefined yearly theme. Then, each school sends its stories to Ajudaris and has to indicate the year and grade of the writers of each one of the stories. After confirming the reception of the stories, Ajudaris makes available to each school a second Google Forms to collect more information, including:

- Names of the school group, school and interlocutor(s);
- Motivation to participate in the project;
- Number of students and teachers involved and number of stories submitted;
- Will to organize events;
- Satisfaction level and improvement suggestions;
- Suggestion of themes and possible contacts;
- Number of books requested.

Phase 3 - Rate and select stories

As soon as stories are received, Ajudaris divides and distributes them to the judges according to specific constraints (which are going to be explored in the next chapter). Then, all the stories are sent via email to the judges they were assigned to. The judges grade the stories according to several criteria so that the best stories among all submitted can later be selected by Ajudaris.

Phase 4 - Review and illustrate stories

Before sending the selected stories to the illustrators, there is the need to review them one last time to verify if all the stories respect the rules of the contest (e.g. no plagiarism is allowed). The reviewers are mainly partners that volunteer to help in the process. This is one of the possible

contributions of partners to the project. The partners in this stage can then be teachers, students or other stakeholders of the project that are available to review a certain number of stories. The stories approved are illustrated and gathered by the organization at the end.

Similarly to what happens with schools, illustrators also have to fill in and submit a Google Forms after submitting the illustrations. This form includes the following information fields:

- Name, address, small résumé and relevant contact information of the illustrator;
- Value that the illustrator attributes to the illustration submitted;
- Permission to sell the original illustration in an auction;
- Will to participate in events;
- Suggestions of themes and possible contacts;
- Satisfaction level.

Phase 5 - Design, produce and publish books

This phase of the project is mainly performed by the organization's partners. The final group of stories and the respective illustrations are divided into different book volumes according to the schools' regions. The output of this stage is precisely the set of finished books that will be sold later on.

The partners in this stage are mainly companies that want to provide help in these tasks. Partners' contributions can then include, besides what was mentioned in phase 4, the provision of money or facilities/assets to allow the production or even the distribution of books.

Phase 6 - Distribution and sales

At this point, Ajudaris has to allocate the books produced into the three main selling channels. Therefore, schools and stores participate by selling books and/or organizing events and partners contribute by distributing books to the different destinations. The other obvious stakeholders of this phase are the end consumers (i.e. the buyers). Partners can also be involved in this phase. Here, contributions come mainly in the form of monetary donations or the provision of transportation.

3.2   Current problems

From the previous analysis, it becomes clearer that there is a large amount of data being collected from stakeholders at each one of the six main phases of the process. This information is either demographic or relates to the interaction of the stakeholders with a specific part of the project. The latter can be exemplified by the information about the number of stories submitted that a school indicates in the second Google Forms. The amount of information and the way the organization currently deals with it is one of the main challenges within this dissertation project.

Currently, after receiving the responses to the registration forms, Ajudaris would have to manually organize the data in an Excel sheet. The same would happen upon the reception of the stories

with information related to those stories and its writers. All in all, these have been some of the most time consuming activities performed by Ajudaris, which needs time to perform the core and value adding activities. Moreover, with these procedures, the organization tends to lose records and its contacts, which makes it more difficult for Ajudaris to understand which schools to contact at each moment and how to contact them.

There is also the concern of which schools should or should not be contacted. Trying to contact all the schools is not a feasible procedure and selecting some without a specific criteria may be an unrewarding task, as not all the schools have the right profile/characteristics that justify an approach.

Further in the project, the attribution of stories to the respective judges, reviewers and illustrators is also performed manually. As there are some constraints in the evaluation process, the fact that this stage is performed manually can lead to mistakes and requires that each of these stakeholders is contacted individually.

The second main issue is more related with the sales part of the process. From one side, two of the selling channels have been falling behind expectations - the one through which more books are sold relates to the events organized by schools that participated in the project. From the other side, public health related restrictions due to the COVID-19 pandemic also cause problems when it comes to selling books. One specific situation mentioned by Ajudaris' director consists of the difficulty of reaching out to the parents of the young writers and other potential buyers. As many presentations and events in schools had to be cancelled, those parents could not truly perceive what exactly was their children's project and, as a consequence of that, they would not buy the book. This is one of the examples that justifies a drop of sales - if potential buyers are not aware of the books and the overall cause, they will not buy or contribute in some other way.

Finally, it is also important to mention the issue of the organization of the data. All the information collected from all the stakeholders is fragmented and dispersed, which makes it harder to consult and manage it. Some important information can thus be lost and, in most of the situations, the processes that deal with all the information management are consequently more complex.

3.3 Main goals

It is necessary to address the main problems mentioned in the previous section. The first one is related to the data collection and management, as well as how stakeholders are approached and how the processes can be automated. The goals here are:

- to develop a more automated database that also presents information in a clearer way than the one obtained with the current organization in an Excel file;
- to facilitate the management of contacts also in terms of the judges, reviewers and illustrators of the stories in competition and how the communication is established with them.

Before the development of the database, it is also necessary to represent all the necessary and useful information in a UML diagram. The objective with this diagram is to start structuring the

database by visualizing the necessary information related to each stakeholder and its interaction with the project "Histórias da Ajudaris".

The second issue is more related to a marketing oriented approach. This involves the segmentation of specific stakeholders (schools) to help in the marketing analytics procedures. Some specific goals are detailed below:

- Increase the number of participants;
- Reduce time spent collecting and organizing data;
- Reduce the number of dropouts from the project;
- Identify schools/school groups for Ajudaris to contact and automate the way all these contacts are established.

3.4 Performance measurement

By analyzing the process divided in the previous six main phases, it is possible to define examples of Key Performance Indicators (KPIs). During the first phase, some metrics might include the number of schools contacted by Ajudaris, the number of schools participating in a given edition, the number of new participants or even the number of dropouts (i.e. schools that participated at least once but did not renew the application in a given edition).

In the second stage, the KPIs are more related to the number of stories received by the organization, which is an indicator that can be subdivided by schools or regions. The number of stories selected for the books is the most obvious indicator to be used in the third phase, which can also include the number of judges involved in the evaluation of the stories. The fourth one can contemplate the number of illustrators involved and the number of stories discarded in the reviewing stage.

When it comes to the phase of designing, producing and publishing books, it might be useful to highlight the number of different volumes edited in one edition, the number of copies printed and the overall costs associated with the production. Finally, the last phase can include KPIs such as the number of books sold and revenues obtained from those sales.

The same procedure can be applied to the project as a whole instead of specifying each phase. In this case, KPIs include some of the examples mentioned above for each phase and also broader ones:

- Number of books produced and sold;
- Number of schools participating;
- Number of stories submitted;
- Total costs incurred;
- Total revenues obtained;
- Satisfaction level of stakeholders.

Chapter 4

Methodology

The present chapter approaches the solution that was developed to tackle Ajudaris' main concerns, as well as detailed descriptions of what was done to achieve that. First, the focus is on the necessary information from all the stakeholders and how it is collected and stored. Then, solutions related to the construction of a database were also presented. Finally, the entire contact management process is detailed.

4.1 Data collection and management

4.1.1 External data

Following the six main stages and respective tasks, all the important types of information necessary to fully understand the process had to be listed for each one of the stakeholders in each stage. Some of these data inputs were already being collected and, therefore, are listed on the third chapter of the present dissertation. However, more inputs were listed to improve the comprehension and evolution of the project. For example, in terms of schools/school groups, the identification code of the establishment is a useful addition to the further development of the desired database.

When it comes to the teachers, information about the current school where they work is important for the tasks of contacting the stakeholders - it is especially important in moments when a teacher changes from one school to another. Moving on to the judges and illustrators, it is also necessary to detail which stories were evaluated/illustrated by each stakeholder. In terms of the partners, information regarding the type of contribution (and the respective amount/value) to the project should also be included in the database.

Finally, in the moment of the sales of the books, information about the buyers and sellers could be registered by Ajudaris. For sellers, the obvious information refers to the number of books sold (through each selling channel). About the buyers, it could also be useful to understand the connection they have with the project (e.g. their children participated) and which books they bought. However, this final stage (selling process) was not included in the solution developed for the UML diagram and the database.

Having detailed all these information requirements, a UML diagram was created using the software Draw.io to set the ground for the creation of the desired database. The diagram is shown in Appendix B.

There is a class for each one of the stakeholders identified for the project containing the most relevant information attributes about them. This includes demographic information and other data entries that were already being collected by the organization through the Google Forms. Some other attributes or even classes were added to the diagram so that it would include all the stakeholders and all the necessary information regarding their interactions with the project "Histórias da Ajudaris".

It is also important to mention the ternary associations present in the diagram. The first two include the schools and project (edition) alongside teachers, in one case, and interlocutors, in the other. The third one also includes the schools and the project, which are then connected with the stories written by the students. Furthermore, there are two additional ternary associations involving both the stories and the project - one connects with the illustrators (which represents the illustration made) and the other with the judges (which relates to the stories' evaluations/grades). Finally, one involving the partners and the type of contribution made by each in a specific year was created as well. In this last case, it is relevant to explain that companies or individuals that are partners of Ajudaris can contribute in different ways to help in the execution of the project. For example, a partner can provide the transportation of books or even the revision of the stories, among other tasks.

4.1.2   Data collection

To make the best use of Ajudaris' resources, which include a Microsoft Office 365 license, the forms were copied into the format of Microsoft Forms (instead of Google Forms). The three forms that were already being used were thus kept in this new format with approximately the same information as the previous ones, with only a few fields being added and some others being removed. In terms of information added, the main question is related to school groups' (and schools') identification codes. These codes exist in official databases made available by the government's education branches (Instituto de Gestão Financeira da Educação (IGeFE) 2021; Associação dos Trabalhadores da Educação 2018).

The first form (the one in which school groups register to enter the project) collects data into the following tables: School Group, School, Teacher, Interlocutor (plus the connection tables related to these ones). This registration form continued to be directed at school groups, which can register one or more schools to participate. The second one ("Boas Práticas - Histórias", which schools fill in after submitting their stories), flows data into the table that connects the school to the project edition, containing information about the participation of the schools in a said year. The third form ("Boas Práticas - Ilustradores", directed at supportive illustrators that collaborate with Ajudaris) collects data to be stored in the tables for illustrations and for the relation between illustrators and the current project's edition. The table for illustrations represents a ternary relationship between the tables of the illustrators, stories and project (edition). The main additional

field added to this form is the one referring to identification codes of the stories attributed to each illustrator - this issue will be further explored during the present chapter.

Two new forms were then proposed to complement the collection of information. The first of these is directed at schools and it is made available at the same moment as the form "Boas Práticas - Histórias". Until this point, schools used to send their students' stories via email in a document that had to contain the references to the school, school group, grade and class of the students. The new form replaces this type of submission, which is now made directly in this Microsoft Form (each school submits one form per story delivered). Part of this form is represented in Figure 4.1. The alternative here is to maintain the submission of the forms via email (preferably containing a Google Drive or Microsoft Drive link to access the online document of the story). In this case, Ajudaris still has to register the stories manually, but now an intuitive and easy to use interface (in MS Access) that was developed makes this process relatively quicker than it was before.

Figure 4.1: Form for schools to submit stories

The second additional form is directed at the judges. The judges submit one form with the grades of the stories evaluated, making reference to the codes of the stories that were given to them, as well as their own identification codes. Part of this form can be seen in Figure 4.2. The

attribution of stories to judges, reviewers and illustrators will be further developed in this chapter - this will include the explanation of the said identification codes (stories' codes and judges' codes).

Figure 4.2: Form for judges to evaluate stories

The remaining necessary information will continue to be collected in the same way as before by the organization. This includes demographic information about the judges, partners and illustrators, which is collected upon the moment one of these stakeholders contacts the organization and is then inserted manually into the database by Ajudaris.

4.1.3   Data management

Initially, Ajudaris would make the Google Forms available and then organize the answers one by one in an Excel file - this process would be done manually. The new procedure proposition involved using Microsoft Forms instead of Google Forms.

The answers submitted through the mentioned forms are now collected into a Microsoft Access (also known as MS Access) database. For the data to reach the said database in a more automated way, the Microsoft Power Automate service was used to collect answers. However, this service does not allow a direct connection between Microsoft Forms and MS Access. Thus, an intermediary step had to be taken and it consisted in the MySQL Workbench software.

The data flow created within Microsoft Power Automate starts with a trigger - in this case, the trigger is the submission of answers to the Forms. Then, Power Automate retrieves all the answers

to each of the questions and, in the next step, connects those answers to specific columns of tables created in the MySQL Workbench environment.

MySQL Workbench database

Following the structure of the UML diagram, the database was created within the environment of MySQL Workbench. Here, a new connection was created and the credentials were made available for Ajudaris. Then, a new schema was also established to save the relevant tables and data entries. From the five Microsoft Forms considered (the three original ones and the two proposed), only the form "Boas Práticas - Histórias" flows data directly to its final table already mentioned in this chapter. The remaining forms input data (through Microsoft Power Automate) into intermediate tables with the same entry fields as the forms. Then, these tables have a set of "After insert" or "After update" triggers to pass each one of the data entries to the respective tables as answers are being submitted in the Microsoft Forms. This way, all the relevant tables are populated appropriately.

The remaining tables correspond to the ones also represented in the UML diagram. Besides the mentioned triggers, there are also several stored procedures used to separate strings into different entries. One of the cases in which this is necessary happens when the responsible of a school group fills in the registration form - there is a question relative to the participating schools in which it is requested to separate the schools by semicolon. With these procedures, the several schools can occupy one row each in the respective table. The same happens with teachers, interlocutors or even contacts' suggestions made by participants.

All in all, this MySQL database is a system that allows to automatize the processes of collecting and storing data from the project's stakeholders. The data is then organized and separated into the respective data tables already considered.

Microsoft Access database

MySQL Workbench, as a database, is not that much user friendly in terms of interfaces and data consultation. Therefore, the MS Access software was chosen in this context to serve as an user front-end. Thus, a connection between the two software had to be created so that there could exist the link between tables and a new external data source had to be added to the MS Access file from an Open Database Connectivity (ODBC) database (the ODBC database is, in this case, the one created in the MySQL Workbench software).

The first step taken, even before what was just mentioned, was the installation of the MySQL's ODBC connector on Windows to allow the already mentioned connection between the two environments. This connector had then to be configured with the corresponding parameters of the MySQL Workbench database and the schema containing the tables created in that environment (MySQL 2021a; MySQL 2021b).

This connection made the tables created in MySQL available in MS Access. The same tables that were created and populated in MySQL Workbench become also available in MS Access - any update made in Access changes the tables in MySQL Workbench as well. Having the said

tables available in this software, user interfaces (forms) can be created, making also use of queries from the available data. Some of the user forms created are directly related to specific tables and contemplate the registration of new project editions, stories, illustrators, judges, partners and contributions.

Other forms are the result of Structured Query Language (SQL) queries, such as one that presents the schools that participated in previous editions of the project but still have not applied to the newest edition - this is the reminder query, which has a corresponding user form that shows the churners from the project and presents the relevant contacts so that they can be contacted and reminded of the project. Another example can be the query that sums all the grades given from different judges to the same story, in order to obtain the final grade. Then, the majority of the queries connect data from different tables. Some more details regarding the MS Access environment will be mentioned further in the present chapter.

While MySQL is a database that collects information and automatizes its storage and organization, MS Access is a user-oriented database. Then, all the operations performed by Ajudaris have to be facilitated by this solution. Summing up what was mentioned above, MS Access' major roles include:

- the data consultation, including tables and relevant measures/indicators (the number of participants in a specific year, for example);
- the creation of new records, including the registration of new stakeholders participating in one edition;
- the facilitation of the process of contacting stakeholders, both by listing those that should be contacted and by automatizing the way these contacts are established (will be further developed in this chapter).

Figure 4.3 summarizes the data flow through the several software and tools mentioned in this section. The end goal of the flow is to present relevant data to the user of the database. Therefore, it culminates in the Microsoft Access software.

4.2 Analytics for school participation management

4.2.1 Evolution of the participation in "Histórias da Ajudaris"

By analyzing the list of participants from previous editions (from 2017 to 2020) provided by Ajudaris, it is possible to study the evolution of the participation in the project in these past years.

In 2017 there were 423 participants (332 of those were school groups from mainland Portugal and the remaining ones were either ungrouped schools or other establishments) in the project.

In 2018, the total number decreased to 419. Nevertheless, there was an increase in terms of school groups - from 332 to 346. From these 346, 70 were new participants and 276 had already participated in 2017 - thus, there were 56 school groups that dropped out from the project from 2017 to 2018.
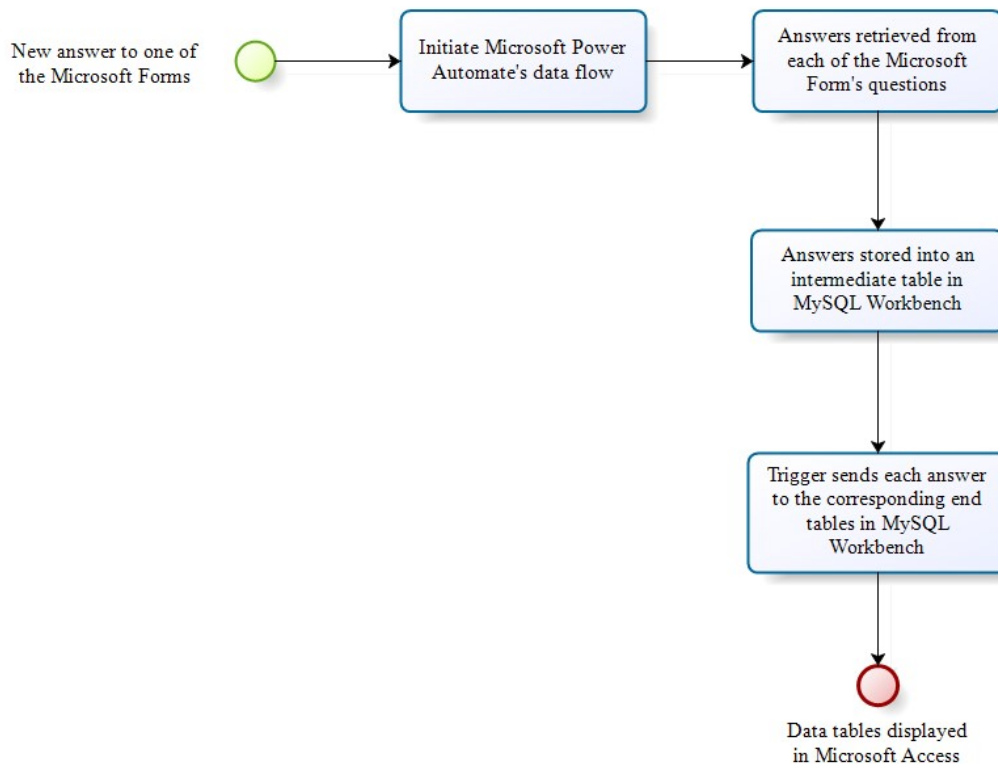
Figure 4.3: Data flow from the Microsoft Forms to Microsoft Access

There was another participation decrease in 2019, as there were, in total, 379 teaching establishments in the project. This decrease happened mostly in terms of school groups, as the number of those participants went from 346 to 305. In this year, relative to 2018, there were 58 new school groups and 247 renovated their registrations - 99 school groups left the project when compared with the previous year, which justifies the participation decrease.

Finally, 2020 had 375 participants, from which 303 were school groups. Therefore, the final number was almost the same as 2019. However, there were still significant changes, as 81 of the school groups from 2019 left the project, with 79 entering in 2020 - 224 school groups stayed from the previous year. It is then obvious by looking at Figure 4.4 that the majority of each year's participants (when considering the school groups) come from the year before.

It was also possible to identify that 166 school groups participated in all of the four mentioned editions of the project. Furthermore, 104 participated in 3 of those editions, 111 participated in 2 and 88 participated only once during this period.

The remainder of the chapter will be focused on the participation of only school groups from mainland Portugal, as the information available about the other institutions (school groups, schools or other establishments) is not entirely the same - for many of the cases, there is a considerable amount of missing data that would not be compatible with the desired analysis.
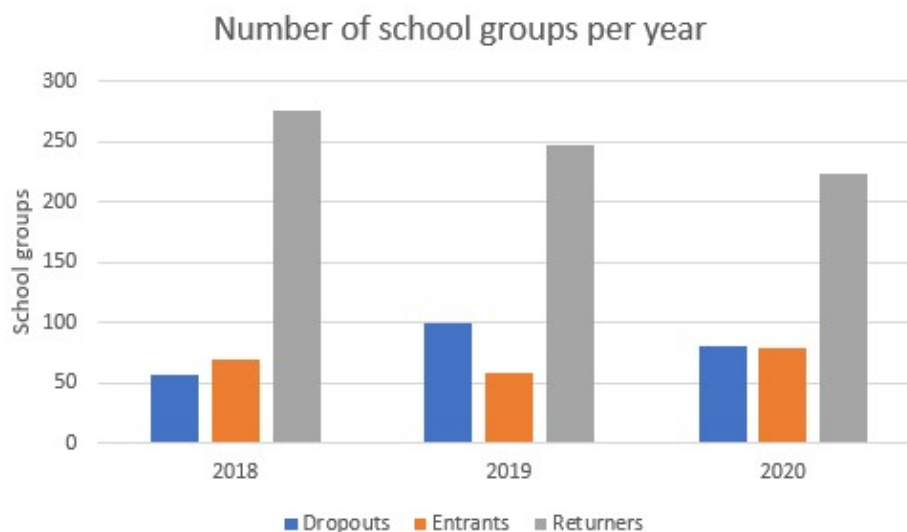
Figure 4.4: Evolution of the participation of school groups in the project

4.2.2 Increasing the number of participating schools

Even though many schools already know about the project and decide to participate, Ajudaris has to contact several other schools every year. This is done either by sending reminders to schools that used to participate but did not sign up for one specific edition or by contacting new schools (those which have never participated in "Histórias da Ajudaris"). Moreover, there is also the scenario in which a teacher responsible for the development of the project in a school changes to a different school in the following year. Here, two possible actions are raised - ask the teacher or contact the old school to make sure the project is passed on to someone else; ask the teacher or contact the new school to guarantee another participant for the most recent edition.

Previous years' participants - School churn

When it comes to reminding previous participants, it is necessary to be aware of the school churn (i.e. schools that have already participated in the project but did not renew the registration in one edition).

The churn rate can increase as a consequence of the following main scenarios:

- One school group did not renew the registration (i.e. school group churn);
- One school group registered less schools in the project (i.e. school churn);
- One school group registered different schools in the project (i.e. school churn).

Thus, the following step is to contact these churners as an attempt of guaranteeing another participant in the project. The interlocutors (from the last participation) of the schools in question are approached by Ajudaris. Here, one of the possible justifications for the school churn can be that the said interlocutor went to work in a different school. In this particular case, it is important not only to convince the interlocutor to pass the project to someone else at the previous school

but also to try to get the new school to participate as well. There can be other justifications for the school churn, but the goal is still the same for those other cases - guarantee the renewal of the schools' applications.

The solution developed includes automated reminders that alert Ajudaris about the churners. Thus, a SQL query was developed in MS Access to select those schools that are present in the list of participants from previous years but are not yet registered for the most recent edition. These schools are then displayed in a table alongside the respective contact information (phone number and/or email) of the schools and its last interlocutors. Then, the possibility of sending automated and personalized emails to the respective interlocutors was also developed from a template email. To do so, a Visual Basic for Applications (VBA) code module was created with a function that allows sending an email via Outlook (which is within the Microsoft account) with custom parameters retrieved from the table containing data from the churners (both schools and the respective interlocutors). This fits within the contacting procedures mentioned above.

Identification of potentially new participants - targeted marketing

When it comes to the contacts made to get new participants, it is useful to know what kind of schools are more likely to accept. Thus, it was necessary to build a profile for those schools. To do so, the list of Portuguese school groups was gathered from official databases, as well as the respective identification codes and basic information - address, contact information and number of schools within the school group (Instituto de Gestão Financeira da Educação (IGeFE) 2021; Associação dos Trabalhadores da Educação 2018).

This information was complemented with other school groups' statistics, including the number of students and teachers per school group and per teaching level. Furthermore, statistics regarding Portuguese municipalities were also collected and related to the school groups according to each one's location. It is important to mention that these variables were collected either from Pordata or DGEEC (Direção-Geral de Estatísticas da Educação e Ciência (DGEEC) 2021; PORDATA 2021).

The procedure involved the selection of geo-demographic variables that could have an influence on school groups' decision about whether to participate in the project or not, thus distinguishing those willing to participate from those that should not be contacted in the first place. Having this said, the selected variables about the municipalities are listed below:

- Percentage of inhabitants enrolled in different school levels;
- Number of teachers in different teaching levels;
- Percentage of inhabitants that benefit from *Rendimento Mínimo Garantido* (RMG) and *Rendimento Social de Inserção* (RSI);
- Percentage of inhabitants that benefit from unemployment insurance;
- Percentage of inhabitants that benefit from social security support;
- Per capita purchasing power;
- Number of unemployed inhabitants per 100 employed;
- Difference between minimum wage and average monthly wage;

- Town Hall's spends in culture and sports in proportion to the total amount spent;
- Spendings, Income and Balance of each Town Hall;
- Real Enrollment rate per teaching level;
- Ageing index.

All the data was stored into two tables (one containing information about the school groups and the other with statistics per municipality) created inside the MS Access environment. These tables were then connected in a simple SQL query through each school group's geographical variables. The table that resulted from this query was the one used in the Python script to execute the clustering procedure. The table has the identification code of each one of the 713 school groups and 33 other columns containing the variables already listed (with the respective subdivisions). This connection between the Python script (in Visual Studio) and MS Access was done through a VBA function ("Call shell") in Access that calls the Python script through an additional intermediate file in the '.bat' format. This file contains the paths to the Python executable (python.exe) and to the Python script created.

Before executing the clustering itself, it was necessary to process and manipulate the data. Firstly, there was the issue of missing values with the variable of the enrollment in the high school level. More specifically, 22 of the municipalities from the list did not have values for this variable. In these cases, the missing values were replaced by the values that corresponded to the regions (NUTS III[1]) where each of the municipalities belonged to - the value of each one of the regions represents the average of the municipalities that it includes.

Secondly, some of the mentioned variables were already present in the ratio format. However, some other variables consisted of absolute values. Thus, for these variables, a ratio was made according to the municipalities' number of residents. Those variables were the following:

- Number of teachers in different teaching levels;
- Amounts spent, Income and Balance of each Town Hall.

Then, as the variables had different units and weights, all of them went through a standardization process. This way, all the numerical values were within the same boundaries - 0 and 1. From this point forward, the clustering process could be initiated. The method chosen was K-Means, which meant that the k (number of clusters) had to be determined in advance. To do so, the 'elbow-curve' was used. The obtained diagram can be seen in Figure 4.5.

Even though the diagram is not entirely clear in terms of the best k, the one chosen in this situation was k=10. Probably the most adequate k that one can see in the plot would be around 4 or 5, due to a clearer elbow shape in that zone. However, the intention was to separate the school groups into a larger number of clusters.

So, 713 school groups (from mainland Portugal alone) were then divided into 10 clusters according to the above-mentioned variables. The output of this entire Python script was the allocation of each school group to one of the clusters. This was then printed in a separate table

---

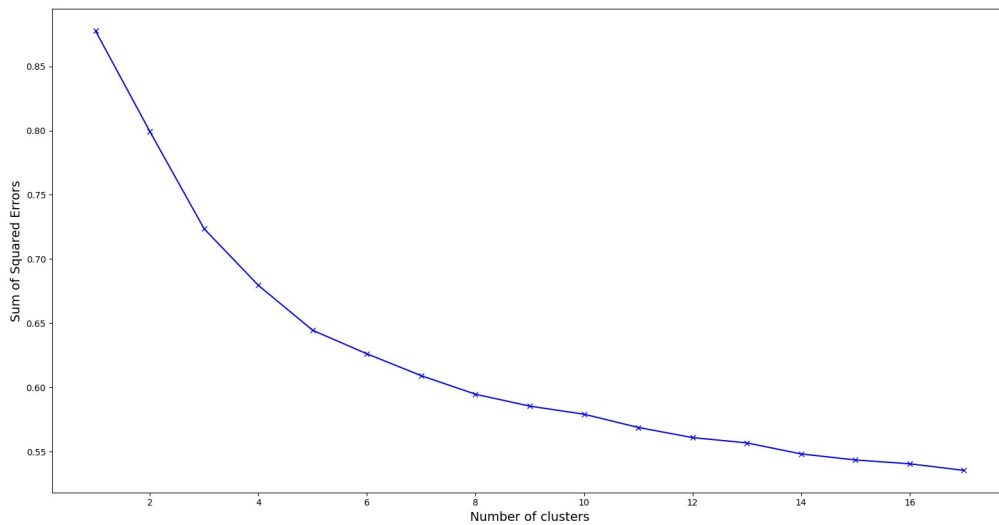[1]NUTS III divides the country into 39 regions

Figure 4.5: Determining the right number of clusters

created in MS Access. When calculating the silhouette index to evaluate the quality of the upper mentioned process, the result came up too low (silhouette score of around 0.15).

Therefore, with the results from the clustering, a different analysis was made to assess the importance of all the variables considered. To do so, one of the feature selection functions made available by Python's 'scikit-learn' package was used. The method used was the L1-based feature selection to attribute weights to all the variables according to each one's importance within the model (scikit-learn - Machine Learning in Python 2020; Mushtaq et al. 2020; Abbad Ur Rehman, Lin, and Mushtaq 2021).

From here, 7 of the 33 variable columns were selected as the ones with more importance among the entire set of variables initially considered, including:

- the percentage of each municipality's inhabitants enrolled in the first teaching level (primary schools);
- the difference between minimum wage and average monthly wage of each municipality;
- the real enrollment rate for the four teaching level (from preschool to high school).

Then, the same process was executed using the already processed data variables. Once again, the number of clusters used was 10 as the interpretation of Figure 4.6 was the same as the one from Figure 4.5.

The silhouette score was then recalculated and the result was 0.32 (more than the double of the previous score). Even though the result is still far from ideal (closer to 1), this clustering process was considered for further analysis.

Obviously, some of the variables have more impact than others in the clustering process. Table 4.1 shows the squared values of the correlation rations for each one of the seven variables used. This indicates the proportion of the variance of these variables that can be explained through the created model. The variables with higher values can be considered as the ones with more impact
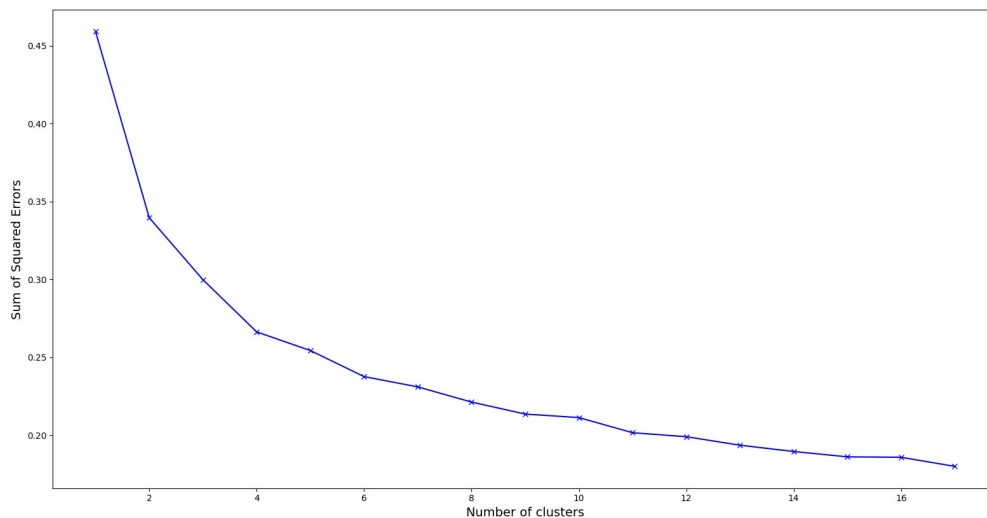
Figure 4.6: Determining the right number of clusters (with the 7 variables)

in the definition of the clusters. In this case, the three variables with higher values are related to the real enrollment rate (for preschool (*taxa_preescola*), second (*taxa_2ciclo*) and third (*taxa_3ciclo*) Portuguese education levels [2]) (Rakotomalala 2017).

Table 4.1: Squared correlation ratios for each variable.

Variable name	(Corr. ratio)^2
%matriculados_1ciclo	0.7049
diferenca_salario	0.7163
taxa_preescola	0.8218
taxa_1ciclo	0.7677
taxa_2ciclo	0.8130
taxa_3ciclo	0.8304
taxa_secundario	0.6922

To add to the evaluation process of the clustering as a whole, a Multinomial Logistic Regression model was also used to calculate the final accuracy and verify the results of the clustering process. Here, the output of the K-Means (1 column per variable and 1 column with the cluster attributed to each school group) was given as the input for the model. Then, the function divided, in each iteration, the data set into 10 separate folds, using 9 of those to create and train the regression model and the other to test and calculate the accuracy. Having the values for each variable and the respective attribution of a cluster (this is the output of the clustering process), the model is constructed in a sense that the variables should allow it to distinguish between clusters and compare them. Then, the records that are not used to build the model are given as an input. Here, based on the values for the set of variables, the model will predict the cluster those records (school groups) belong to. By comparing these results with the actual clusters attributed in the K-Means

---

[2]The second level includes the 5th and 6th grades, while the third level includes the 7th, 8th and 9th grades

method, it is possible to calculate the accuracy to understand if these procedures go in line with the results obtained in the previous stages (clustering). Having this said, this model calculated a final accuracy (after considering all the iterations of this supervised model) of 81.3% (Brownlee 2021; Erdfelder et al. 2009; Krishnapuram et al. 2005).

With the list of participants in previous editions of the project (2017-2020) made available by Ajudaris, the clusters with larger proportions of participants were identified. The assumption was that the school groups belonging to these clusters are more likely to participate in the project. So, those school groups that belonged to the mentioned clusters but did not belong to the list of previous participants were pointed out as the ones that could be contacted by Ajudaris to potentially feature in upcoming editions of the project. This proportion was calculated by dividing the number of school groups from each cluster that have participated in at least one of those four years by the total number of school groups from the same cluster.

The clusters 8, 9 and 6 have the highest proportions, with 88%, 87% and 76%, respectively. Then, cluster 5 has a proportion of 69%, followed by cluster 2 (with 68%), cluster 0 (with 66%), cluster 3 (with 62%) and cluster 4 (with 59%). Finally, the clusters with lower proportions are clusters 7 and 1, respectively with 40% and 18%. Therefore, three separate levels can be seen in terms of the proportion and are listed in Table 4.2.

Table 4.2: Clusters divided into groups.

Group	Clusters
A	8, 9, 6
B	5, 2, 0, 3, 4
C	7, 1

Having formed the 10 clusters, the average values of each one of the seven variables were calculated for each of the clusters. Then, the variances of the same variables were also calculated using the said averages (variances between the clusters).

Figure 4.7 illustrates these variances and its analysis goes in line with the one made with the squared correlation ratios. The three variables with higher variances between clusters are the same three variables with higher correlation ratios, which helps to conclude about the higher significance of these variables to distinguish the ten clusters when compared with the remaining variables.

Then, to visualize what the top three clusters (in terms of the proportion of participants) have in common, Figure 4.8 also represents the variances for all the variables, this time only between clusters 8, 9 and 6.

By comparing Figure 4.7 and Figure 4.8, one possible interpretation could be that some of the variables that separate the 10 clusters bring the top 3 clusters together. This appears to happen for the first four variables of the list - the percentage of each municipality's inhabitants enrolled in the first teaching level ('%matriculados_1ciclo'), the difference between minimum wage and average monthly wage of each municipality ('diferenca_salario') and the real enrollment rate for preschool and primary school ('taxa_preescola' and 'taxa_1ciclo').
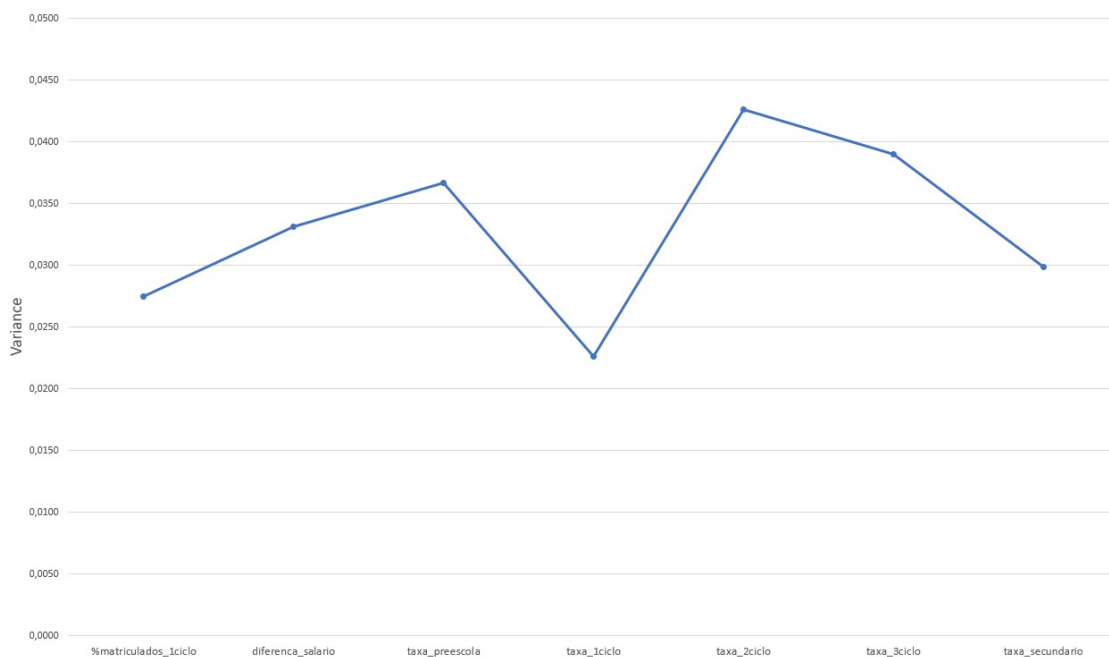
Figure 4.7: Variances of each variable between the 10 clusters

The averages of the variables considered were also calculated for the three groups (A, B and C). From the first four variables already considered, and by looking at Figure 4.9, group A (clusters 8, 9 and 6) appears to stand out as the group with:

- the lowest percentage of each municipality's inhabitants enrolled in the first teaching level (*%matriculados_1ciclo*);
- the highest difference between minimum wage and average monthly wage of each municipality (*diferenca_salario*).

Furthermore, a geographical analysis was also attempted - the variable of the specific region (NUTS II[3]) of each school group was used. Even though the analysis is not too clear, as the majority of the clusters contain many school groups from each region, the tendency seems to be that there are more school groups from the North and Center of Portugal belonging to the top clusters (in terms of the proportion of participants). The bottom clusters (7 and 1) contain almost exclusively school groups from the region of Lisbon. The number of school groups per region for each cluster is presented in Table 4.3 - only 5 of the 7 regions are listed as the clustering process only took into account the school groups from mainland Portugal.

By doing what has been described so far (i.e. the clustering techniques), instead of contacting all the schools or choosing some of them randomly, the ones closer to the profile built and that are not part of the list of participants (from 2017 to 2020) can now be identified and contacted by the organization. This process aims at increasing the efficiency of Ajudaris when contacting potentially new participants, saving precious time for other more value adding activities.

---

[3]NUTS II divides Portugal into 7 broad regions identified by codes: North (101), Center (102), Region of Lisbon (103), Alentejo (104), Algarve (105), Região Autónoma dos Açores (201) and Região Autónoma da Madeira (301)
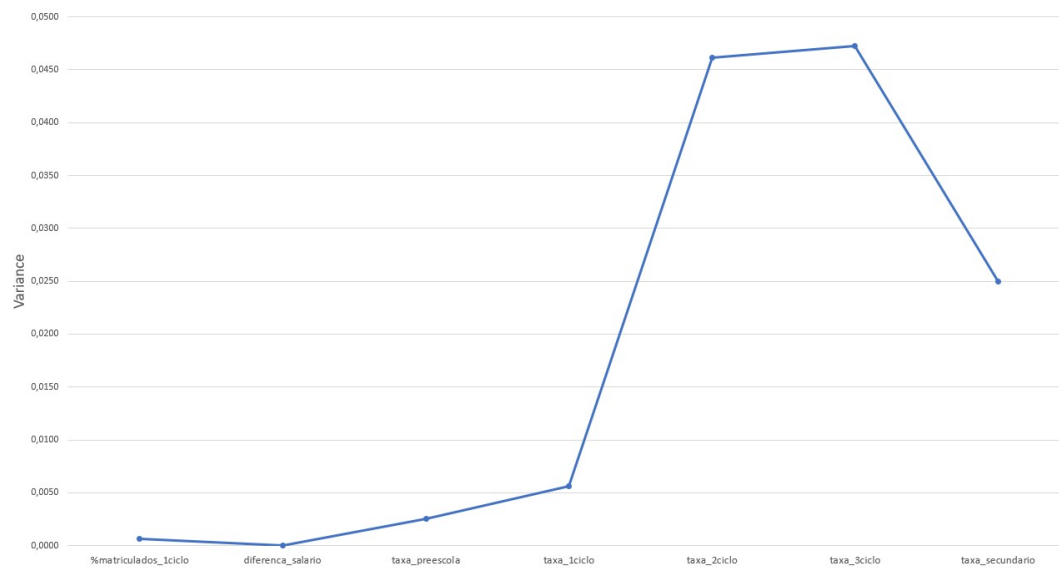
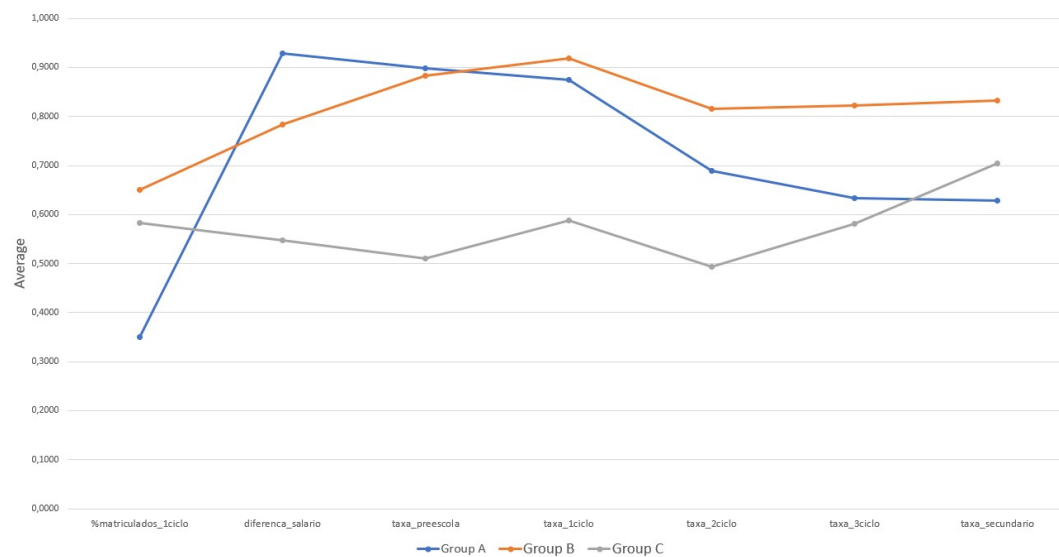Figure 4.8: Variances of each variable between the top-3 clusters

Figure 4.9: Average of each variable per group

The result included 23 school groups belonging to the top three clusters (group A). These school groups were thus identified as the potential main targets for Ajudaris to contact in future editions of "Histórias da Ajudaris". From the twenty three targets, eight belong to cluster 8, six to cluster 9 and nine to cluster 6.

To facilitate the approach to the selected school groups, the output of this entire process described above was made available in a simple MS Access form, which presents the records containing the necessary contact information so that Ajudaris can contact them in an easy and appropriate manner (either by email or phone call).

Moreover, Ajudaris can also use contacts' suggestions provided by other stakeholders to contact new potential participants. These suggestions are requested in the Forms that the several

Table 4.3: School groups from each region (NUTS II) per cluster.

Cluster	101	102	103	104	105
0	78	37	10	21	6
1	1	2	74	0	0
2	60	9	29	8	6
3	18	0	28	1	0
4	33	24	34	8	5
5	16	0	22	4	16
6	8	15	6	8	1
6	0	0	11	0	0
6	24	32	4	6	2
6	15	10	11	8	2

participants have to fill in throughout the project in order to participate. Alongside the remaining answers given to those Forms, the suggestions (names and relevant contacts) are also stored in the developed database in a separate table built for that specific utility.

4.3 Analytics for the rating, revision and illustration process

Having completed the stages of collecting data and, then, stories from the participants, the process focuses on the tasks in which the judges, reviewers and illustrators participate. There is the need to send those stakeholders a set of stories for them to execute their parts in the process. Up until now, all the stories were assigned manually in an Excel spreadsheet containing information about the stories and the involved stakeholders - a code referring to a judge, reviewer or illustrator (depending on the stage of the process) would be added to each of the stories, so that it would be possible to identify who would be in charge of doing what. Only then the stories would be sent to the respective targets via email.

The first assignment process refers to the judges. When the evaluations from the judges would be sent back to Ajudaris, the grades would also be added to the spreadsheet. Then, those stories selected according to the final grades and the regions would be sent to the reviewers after another assignment process. This stage eventually discards some stories that do not follow the rules of the project. Then, those stories that remain eligible are finally assigned to illustrators.

In order to optimize what was described above, the General Assignment Problem was considered. In this case, three separate linear programming problems were defined, each one with specific constraints. These are linear programming problems responsible for matching the identification codes of stories and stakeholders while obeying the existing constraints. For each of the problems, an arbitrary objective function was created, as in these cases there are no costs of assigning a specific story to a specific judge/reviewer/illustrator. The problems were executed in separate Python scripts (one for each group of stakeholders), using the software Visual Studio and making use of another Python's package - 'pulp'.

4.3.1 Assign stories to the judges

In this problem, the main constraints are the following:

- each story is evaluated by a group of three independent judges;
- each group of three judges evaluates the same set of stories;
- judges should evaluate, as much as possible, stories from different districts than their own.

The entire problem was formulated in a Python script, which could be initialized with a button in one of the user forms created in MS Access. As it was already described in this chapter, the registrations of judges in the database is still a manual process, but now there is a specific user form in MS Access to add a new record to the respective table, which creates auto incremental identification codes. The stories are received through the submission of answers to the respective Microsoft Forms and are stored automatically in the database.

The Python script accesses the list of stories submitted in the latest edition of the project and the list of judges available to cooperate in the same year. These records are stored in data frames with additional information about each. The list of judges is divided into groups of three and the list of stories is divided into the same number of groups as the judges, so that each group of stories is assigned to one group of judges - this guarantees the first two constraints. To minimize the number of cases in which a judge evaluates a story submitted by a school from the same region, the lists of judges and stories were sorted by the corresponding regions (each region has a numerical code). In the case of judges, the list was ordered from the smallest number (identification code of the region) to the largest. The opposite was done in terms of stories.

For each one of the groups created, a separate linear programming model was defined. The structure of each of the models is formulated below.

Decision variables

- J = Number of judges within a group;
- S = Number of stories within a group;
- $X_{sj} = \begin{cases} 1, \text{ if story s is assigned to judge j;} \\ 0, \text{ otherwise.} \end{cases}$

Problem formulation

$$
\begin{aligned}
\text{minimize} \quad & 0 \\
\text{subject to} \quad & \sum_{j=1}^{J} X_{sj} = J, \quad s = 1, \dots, S \\
& \sum_{s=1}^{S} X_{sj} = S, \quad J = 1, \dots, J
\end{aligned}
$$

The output of this problem is a table with two columns - the identification codes of stories and the identification codes of judges they were assigned to. The correspondence between these two is then printed in a table created in MS Access. Then, a separate Python script was created

to automatically send out the stories to the respective judges, making use of the correspondence listed in the table. This script accesses the last mentioned table and matches the identification codes with the necessary information about judges (name and email address) and stories (link to the document, title and authors). The procedure involves creating a new data frame containing as many rows as there are judges cooperating in the current edition. This data frame stores all the stories that were attributed to each judge in the same line. Furthermore, a Python function specifically related to Microsoft Outlook was also used to allow Ajudaris to send personalized emails to each one of the judges - all judges receive one single email containing the information and files of all the stories that were assigned to them. The instruction to send out these emails is also given through a button created in MS Access.

Later on, judges submit the grades for the stories assigned to them through one Microsoft Forms, making reference to their own identification codes and the identification codes of the stories (these codes are included in the automatic emails alongside the remaining information). The grades are thus saved into the database. An SQL query sums the grades given by the different judges to each one of the stories, allowing Ajudaris to select the best stories among the ones still in competition.

4.3.2   Assign stories to the reviewers

This linear programming problem is simpler than the previous one. The goal here is to distribute those stories that were selected in the previous step. The main restriction here refers to the number of stories each of the stakeholders is available to review. This information is included in the database during the registration process of each reviewer. Reviewers should not be given more stories than the number they indicated in the first place. If there are not enough reviewers for all the stories selected, more reviewers should be contacted or the ones already collaborating should be asked to review more stories, if possible. The same happens in the case of illustrators.

The rest of the procedures are the same as with the judges. The output is inserted in a table in MS Access and automatized and customized emails can be sent to all of the reviewers containing the necessary information about the stories assigned to them.

This linear programming problem is formulated below.

Decision variables

- S = Number of stories to be reviewed;
- R = Number of reviewers cooperating in the latest edition;
- C_r = Maximum number of stories to be attributed to each reviewer r;
- $X_{sr} = \begin{cases} 1, \text{ if story s is assigned to reviewer r;} \\ 0, \text{ otherwise.} \end{cases}$

Problem formulation

$$
\begin{aligned}
\text{minimize} \quad & 0 \\
\text{subject to} \quad & \sum_{r=1}^{R} X_{sr} = 1, \qquad s = 1,\ldots,S \\
& \sum_{s=1}^{S} X_{sr} <= C_r, \quad r = 1,\ldots,R
\end{aligned}
$$

4.3.3 Assign stories to the illustrators

Finally, the stories that were not discarded in the previous stage are then to be assigned to the illustrators. Similarly to what happens with the reviewers, each illustrator is available to illustrate a certain amount of stories - this information is also stored in the registration process. Thus, this is one of the constraints of this specific linear programming problem.

An additional constraint is established with the same logic as one of the constraints present in the allocation of stories to the judges but in the opposite way - it is preferred that illustrators work with stories from the same districts as them. To maximize the cases in which that correspondence is verified, both the data frames with information from illustrators and judges were sorted according to the respective district's code - this time, both were sorted in ascending order. The rest of the procedures (presenting the output in MS Access and sending out automatic and personalized emails with the stories to the illustrators) are the same as with the judges and reviewers. The illustrators have then to submit an additional form (one of the original three) in which it is necessary to make reference to the identification codes of the illustrators and the stories (the codes are detailed in the emails alongside the remaining information about the stories).

This linear programming problem is formulated below.

Decision variables

- S = Number of stories to be illustrated;
- I = Number of illustrators cooperating in the latest edition;
- C_i = Maximum number of stories to be attributed to each illustrator i ;
- $X_{si} = \begin{cases} 1, \text{ if story s is assigned to illustrator i;} \\ 0, \text{ otherwise.} \end{cases}$

Problem formulation

$$
\begin{aligned}
\text{minimize} \quad & 0 \\
\text{subject to} \quad & \sum_{i=1}^{I} X_{si} = 1, \qquad s = 1,\ldots,S \\
& \sum_{s=1}^{S} X_{si} <= C_i, \quad i = 1,\ldots,I
\end{aligned}
$$

Chapter 5

Conclusions

This dissertation project was included within the broad area of marketing analytics and aimed mainly at increasing the efficiency of the procedures executed by Ajudaris in several of its tasks.

There were issues related to the fact that many of the procedures of the organization were performed manually, such as the registration of the answers to different Google Forms in a separate Excel spreadsheet and the individual contacts made towards each stakeholder. Still regarding the contacts, there was no exact criteria to segment the different types of stakeholders to understand which of them should or should not be contacted. Other concerns were more related with further stages of the project "Histórias da Ajudaris" - communication with the end customers (buyers of the books) and selling activities.

Initially, the entire process of the project was mapped and all the stakeholders involved were listed as well. This was done to facilitate the full comprehension of the project and to set the ground for the creation of the desired database. The structure of the database was described in a UML diagram and used to develop the final solution using the software MySQL Workbench. Microsoft Access was used as a front-end solution to present the data tables and to allow the insertion and manipulation of data within the database.

Later on, there was the need to automate the collection and organization of the answers to the Forms used by Ajudaris (the Google Forms were converted into Microsoft Forms). Microsoft Power Automate was the tool selected for this task and it allowed the mentioned automation. One answer to one of the forms sets one trigger that starts the flow of the data to the database, where it is divided into the corresponding tables. Then, several user forms were also created to allow the consultation, insertion and editing of the same tables and their records. Some SQL queries were created to help keeping track of the evolution of the project, with information such as schools that should be contacted or the final evaluation of one particular story, among other examples.

Moreover, the focus was then on the process of guaranteeing more participants for the project. Previous participants that did not renew the participation were identified and listed and the possibility of contacting them automatically with customized emails was also provided.

When it came to establishing new contacts, the procedure was to identify those school groups that could be more prone to participate in "Histórias da Ajudaris". This is also aimed at increasing

the efficiency of Ajudaris in the sense that it attempts to reduce the number of contacts made in which the answer received is a negative one. In this context, the clustering separated the school groups into segments and a further analysis of the results identified the school groups that should be prioritized when contacting new potential participants.

From the 713 school groups in mainland Portugal alone, 469 had already participated at least once from 2017 to 2020, which were the four years under consideration during the clustering procedures. Therefore, there were 244 potential contacts to be made and the ones closer to the desired profile should be prioritized. The result of the clustering process, which divided the 713 school groups into 10 different clusters, indicated that 23 school groups belonging to the clusters with higher participation levels had never participated in any of the years in question. So, these school groups were listed as the main output of the clustering activities.

Further in the project, the focus went from the schools to other stakeholders - judges, reviewers and illustrators. Here, the main concern was to automate the process of assigning stories to each stakeholder according to specific constraints of each of the cases. So, linear programming problems were formulated so that each story could be attributed to a specific number of stakeholders - each story is attributed to three judges, the ones selected are attributed to one reviewer each and the ones that remain in competition are then assigned to one illustrator each. Once these assignments have been dealt with, the idea was to also automate the communication with these same stakeholders. The results of the assignment problems are presented in MS Access and the stakeholders can receive personalized emails generated by a separate Python script containing the necessary information for the execution of their tasks - the links to the documents of the stories, relevant identification codes and links to Microsoft Forms (if applicable).

In terms of limitations, it is notorious that the evaluation of the clustering process (through the silhouette coefficient) is not ideal (it is lower than 0.5). This opens the way to future improvements of the process by further exploring the types of variables that could be selected to better separate the school groups. Another limitation was related to the availability of data for every single school/school group in the country. As the amount of data was not the same for every single institution, the clustering process described in this dissertation only considered the school groups from mainland Portugal.

5.1   Future Work

One of the possible improvements for further developments of this project includes the consideration of other schools and teaching establishments alongside a more significant set of variables to describe each one of these. This is aimed at increasing the quality of the clustering procedures described in the previous chapter. Furthermore, in the procedures of allocating stories to the jury and to the illustrators, the constraints related to the regions of the stories and participants were not entirely satisfied. Therefore, this could also be something to explore further.

One of the main issues that was not dealt with in the present dissertation was the idealization and development of an online campaign to increase the number of sales of books and donations

to Ajudaris. Further works related to this project could thus focus more on the end part of the project "Histórias da Ajudaris" (i.e. the stages of the process in which the main stakeholders are the buyers of the books). Then, a deeper understanding of who the end buyers are and why they buy these books would be essential. This includes the ability to understand the motivations behind a purchase, the preferred selling channels, the volumes bought (and respective amounts) and also the buyers' connection to the project. Moreover, ways of improving the efficiency of the several selling channels could be identified to help in the sales. This demands the identification of the main limitations of each one of the channels.

The development of the online campaigns requires a thorough identification of the target audiences and the definition of a maximum budget to perform the activities associated with the campaigns. Therefore, it can be important to explore Facebook's (or Instagram's) functionalities and costs, as well as to understand the type of profiles that should be targeted by the campaigns. The intention behind the development of these campaigns is to increase the number of sales or donations and to increase social media interactions and website visits.

Bibliography

Abbad Ur Rehman, Hafiz, Chyi Yeu Lin, and Zohaib Mushtaq. 2021. "Effective K-Nearest Neighbor Algorithms Performance Analysis of Thyroid Disease." *Journal of the Chinese Institute of Engineers, Transactions of the Chinese Institute of Engineers,Series A* 44 (1): 77–87.

Abercrombie, Rob, Sally Bagwell, and Lena Baumgartner. 2014. "Fundraising Perspectives: Donor Segmentation & Money for Good UK." Technical Report, NPC.

Ahmed, Shamima. 2013. "Non-Profit Marketing." Chapter 8 of *Effective Non-profit Management - Context, Concepts, and Competencies*, edited by David H. Rosenbloom, 1, 217–237. Boca Raton: CRC Press.

Ajudaris. 2021a. Quem Somos. http://www.ajudaris.org/site/home/ajudaris/. Last visited 2021-05-30.

———. 2021b. Regulamento 2021. https://drive.google.com/file/d/1T5IQTteXvX-QIQRw2TV1pT67_FTDWBRy/view. Last visited 2021-06-12.

Akinci, Serkan, Erdener Kaynak, Eda Atilgan, and Şafak Aksoy. 2007. "Where does the logistic regression analysis stand in marketing literature?: A comparison of the market positioning of prominent marketing journals." *European Journal of Marketing* 41 (5-6): 537–567.

American Marketing Association. 2017. What is Marketing? — The Definition of Marketing. https://www.ama.org/the-definition-of-marketing-what-is-marketing/. Last visited 2021-02-17.

Anheier, Helmut K. 2005. *Nonprofit Organizations: Theory, Management, Policy*. 1. Edited by Regina List. London: Routledge.

Associação dos Trabalhadores da Educação. 2018. Agrupamentos de Escolas . https://www.ate.pt/horizontes-da-educacao-2-2/agrupamentos-de-escolas/. Last visited 2021-06-12.

Auger, Giselle A. 2013. "Fostering democracy through social media: Evaluating diametrically opposed nonprofit advocacy organizations' use of Facebook, Twitter, and YouTube." *Public Relations Review* 39 (4): 369–376 (nov).

Bennett, Roger. 2007. "The use of marketing metrics by British fundraising charities: a survey of current practice." *Journal of Marketing Management* 23 (9-10): 959–989.

Brownlee, Jason. 2021, jan. Multinomial Logistic Regression With Python. https://machinelearningmastery.com/multinomial-logistic-regression-with-python/. Last visited 2021-06-15.

Bruce, Ian. 1995. "Do not-for-profits value their customers and their needs?" *International Marketing Review* 12 (4): 77–84.

Campos, Raquel, Franco S Wojciech Sokolowski, Eileen M H Hairel, and Lester M Salamon. 2005. "The Portuguese Nonprofit Sector in Comparative Perspective." Technical Report, Universidade Católica Portuguesa, John Hopkins University, Porto.

Carboni, Julia L., and Sarah P. Maxwell. 2015. "Effective social media engagement for nonprofits: What matters?" *Journal of Public and Nonprofit Affairs* 1 (1): 18–28.

Chang, Cyril F., and Howard P. Tuckman. 1994. "Revenue diversification among non-profits." *Voluntas* 5 (3): 273–290 (oct).

Chang, Cyril F., Howard P. Tuckman, and Grace L. Chikoto-Schultz. 2018. "Income diversity and nonprofit financial health." Chapter 1 of *Handbook of Research on Nonprofit Economics and Management: Second Edition*, edited by Bruce A. Seaman, Andrew Young, and Dennis R. Young, 2, 11–34. Cheltenham: Edward Elgar Publishing.

Chauvet, F., J. M. Proth, and A. Soumare. 2000. "The simple and multiple job assignment problems." *International Journal of Production Research* 38 (14): 3165–3179.

Cho, Moonhee, Tiffany Schweickart, and Abigail Haase. 2014. "Public engagement with nonprofit organizations on Facebook." *Public Relations Review* 40 (3): 565–567.

Constantin, Cristinel. 2015. "Using the Logistic Regression model in supporting decisions of establishing marketing strategies." *Bulletin of the Transilvania University of Braşov Series V: Economic Sciences* • 8 (57): 1–8.

Decker, Allie. 2021. The Ultimate Guide to Nonprofit Marketing in 2021. https://blog.hubspot.com/marketing/how-to-run-a-lean-mean-nonprofit-marketing-machine. Last visited 2021-02-17.

de los Mozos, Ignacio Sacristán López, Antonio Rodríguez Duarte, and Óscar Rodríguez Ruiz. 2016. "Resource Dependence In Non-profit Organizations: Is It Harder To Fundraise If You Diversify Your Revenue Structure?" *Voluntas* 27 (6): 2641–2665.

Direção-Geral de Estatísticas da Educação e Ciência (DGEEC). 2021. Estatísticas - Educação Pré-escolar, Ensino Básico e Secundário. https://www.dgeec.mec.pt/np4/17/. Last visited 2021-05-30.

DNL OmniMedia Inc. 2018. 13 Key Analytics for Nonprofits: Why These Metrics Matter. https://www.dnlomnimedia.com/blog/nonprofit-analytics/#skip-link. Last visited 2021-02-17.

Erdfelder, Edgar, Franz FAul, Axel Buchner, and Albert Georg Lang. 2009. "Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses." *Behavior Research Methods* 41 (4): 1149–1160.

Feltl, Harald, and Günther R. Raidl. 2004. "An improved hybrid genetic algorithm for the generalized assignment problem." *Proceedings of the ACM Symposium on Applied Computing*, Volume 2. Nicosia: Association for Computing Machinery (ACM), 990–995.

Foss, Bryan, and Merlin Stone. 2002. "The state of CRM in financial services in the UK: promise vs reality." Chapter 1 of *CRM in Financial Services: A Practical Guide to Making Customer Relationship Management Work*, edited by Ann Rodrigues and Merlin Stone, 1, 11–20. London: Kogan Page.

Foster, William Landes, Peter Kim, and Barbara Christiansen. 2009. "Ten Nonprofit Funding Models." Technical Report, Stanford Graduate School of Business, Stanford.

Fritz, Joanne. 2019. How Nonprofits Generate Revenue Streams. https://www.thebalancesmb.com/where-do-nonprofits-get-their-revenue-2502011#fees-for-services-and-sales-of-products. Last visited 2021-02-17.

Germann, Frank, Gary L. Lilien, and Arvind Rangaswamy. 2013. "Performance implications of deploying marketing analytics." *International Journal of Research in Marketing* 30 (2): 114–128.

GoFundMe Charity. 2019. Top Nine Nonprofit Funding Sources for Any Organization. https://charity.gofundme.com/c/blog/nonprofit-funding-sources. Last visited 2021-02-17.

Harvey, James W. 1990. "Benefit segmentation for fund raisers." *Journal of the Academy of Marketing Science* 18 (1): 77–86 (dec).

Instituto de Gestão Financeira da Educação (IGeFE). 2021. Pesquisa da Rede Escolar. http://www.igefe.mec.pt/PesquisaRede. Last visited 2021-05-30.

Jain, A K, M N Murty, and P J Flynn. 2000. "Data Clustering: A Review." Technical Report.

Kanungo, Tapas, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu. 2002. "An efficient k-means clustering algorithms: Analysis and implementation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (7): 881–892 (jul).

Karp, Andrew H. 1998. "Using Logistic Regression to Predict Customer Retention." Technical Report, Sierra Information Services, Inc., San Francisco, California.

Kotler, Philip. 1979. "Strategies for Introducing Marketing into Nonprofit Organizations." *Journal of Marketing* 43 (1): 37–44.

Kotler, Philip, and Gary Armstrong. 2018. *Principles of Marketing*. 17. Edited by Linda Albelli. Harlow: Pearson.

Kotler, Philip, and Sidney J. Levy. 1969. "Broadening the Concept of Marketing." *Journal of Marketing* 33 (1): 10–15.

Krishnapuram, Balaji, Lawrence Carin, Mário A.T. Figueiredo, and Alexander J. Hartemink. 2005. "Sparse multinomial logistic regression: Fast algorithms and generalization bounds." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (6): 957–968 (jun).

Kristoffersen, Line, and Sangeeta Singh. 2004. "Successful Application of a Customer Relationship Management Program in a Nonprofit Organization." *Journal of Marketing Theory and Practice* 12 (2): 28–42.

Lozano, Emilia Smolak, Sara Balonas, and Teresa Ruão. 2020. "Public relations strategies in social media: analysis of campaigns for social change in the education sector in Spain and Portugal." *Comunicação e Sociedade* especial (jul): 175–196.

Lucas, Evie. 2017. "Reinventing the rattling tin: How UK charities use Facebook in fundraising." *International Journal of Nonprofit and Voluntary Sector Marketing* 22 (2): 1–9 (may).

McDonald, Malcolm, and Ian Dunbar. 2012. *Market Segmentation: How to Do It and How to Profit from It*. 4. West Sussex: John Wiley & Sons, Inc.

McLeish, Barry. 2010. *Successful Marketing Strategies for Nonprofit Organizations - Winning in the Age of the Elusive Donor*. 2. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Mushtaq, Zohaib, Akbari Yaqub, Shaima Sani, and Adnan Khalid. 2020. "Effective K-nearest neighbor classifications for Wisconsin breast cancer data sets." *Journal of the Chinese Institute of Engineers, Transactions of the Chinese Institute of Engineers,Series A* 43 (1): 80–92 (jan).

MySQL.   2021a.   Configuring a Connector/ODBC DSN on Windows with the
    ODBC Data Source Administrator GUI.   https://dev.mysql.com/doc/connector-odbc/en/
    connector-odbc-configuration-dsn-windows-5-2.html. Last visited 2021-06-12.

———.   2021b.   Installing Connector/ODBC on Windows.   https://dev.mysql.com/doc/
    connector-odbc/en/connector-odbc-installation-binary-windows.html. Last visited 2021-06-
    12.

Norwich University Online.   2016.   7 Key Differences Between Nonprofit and For-
    profit   Organizations.   https://online.norwich.edu/academic-programs/resources/
    7-key-differences-between-nonprofit-and-profit-organizations. Last visited 2021-02-16.

Polonsky, Michael Jay, and Adrian Sargeant. 2007. "Managing the donation service experience."
    Nonprofit Management and Leadership 17 (4): 459–476.

Pope, Jennifer A, Elaine Sterrett Isely, and Fidel Asamoa-Tutu. 2009. "Developing a Marketing
    Strategy for Nonprofit Organizations: An Exploratory Study." *Journal of Nonprofit & Public
    Sector Marketing* 21 (2): 184–201.

PORDATA. 2021. Base de Dados dos Municípios. https://www.pordata.pt/Municipios. Last
    visited 2021-05-30.

Pritcher, Cristopher, and Jeffrey M. Patrick. 2015. "Fundraising and Analytics - Data-driven
    insights drive measurable net revenue." Technical Report, Merkle, Maryland.

Rackley, Jerry. 2015. *Marketing Analytics Roadmap*. 1. Berkeley, CA: Apress.

Rakotomalala, Ricco. 2017. "Cluster Analysis with Python - HAC and K-Means." Technical
    Report, Université Lumière Lyon 2, Lyon.

Ramesh, P, and K Abdul Razak. 2019. "A Mixed Objective Assignment Problem." *Journal of
    Physics: Conference Series*, Volume 1362. IOP Publishing, 12089.

Randle, Melanie, and Sara Dolnicar. 2009. "Not just any volunteers: Segmenting the market
    to attract the high contributors." *Journal of Nonprofit and Public Sector Marketing* 21 (3):
    271–282.

Rupp, Christine, Sarah Kern, and Bernd Helmig. 2014. "Segmenting nonprofit stakeholders to
    enable successful relationship marketing: A review." *International Journal of Nonprofit and
    Voluntary Sector Marketing* 19 (2): 76–91.

scikit-learn - Machine Learning in Python. 2020. Feature selection. https://scikit-learn.org/
    stable/modules/feature_selection.html. Last visited 2021-06-12.

Srnka, Katharina J., Reinhard Grohs, and Ingeborg Eckler. 2003. "Increasing Fundraising Effi-
    ciency by Segmenting Donors." *Australasian Marketing Journal* 11 (1): 70–86.

Tan, Ding-Wen, William Yeoh, Yee Ling Boo, and Soung-Yue Liew. 2013. "The Impact of
    Feature Selection: A Data-mining Application in Direct Marketing." *Intelligent Systems in
    Accounting, Finance and Management* 20 (1): 23–38 (mar).

Tan, Pang-Ning, Michael Steinbach, Anuj Karpatne, and Vipin Kumar. 2019. "Cluster Analysis:
    Basic Concepts and Algorithms." Chapter 7 of *Introduction to Data Mining*, 2, 525–612.
    New York: Pearson.

Velmurugan, T., and T. Santhanam. 2011. "A Survey of partition based clustering algorithms in
    data mining: An experimental approach." *Information Technology Journal* 10 (3): 478–484.

Yang, Jingyuan, Chuanren Liu, Mingfei Teng, March Liao, and Hui Xiong. 2016. "Buyer target-
    ing optimization: A unified customer segmentation perspective." *Proceedings - 2016 IEEE*

International Conference on Big Data, Big Data 2016. Institute of Electrical and Electronics Engineers Inc., 1262–1271.

Young, Dennis R. 2007. *Financing Nonprofits: Putting Theory into Practice.* 1. Lanham: AltaMira Press.

Young, Jimmy A. 2017. "Facebook, Twitter, and Blogs: The Adoption and Utilization of Social Media in Nonprofit Human Service Organizations." *Human Service Organizations Management, Leadership and Governance* 41 (1): 44–57 (jan).

Appendix A

Swimlanes of the phases of the project "Histórias da Ajudaris"
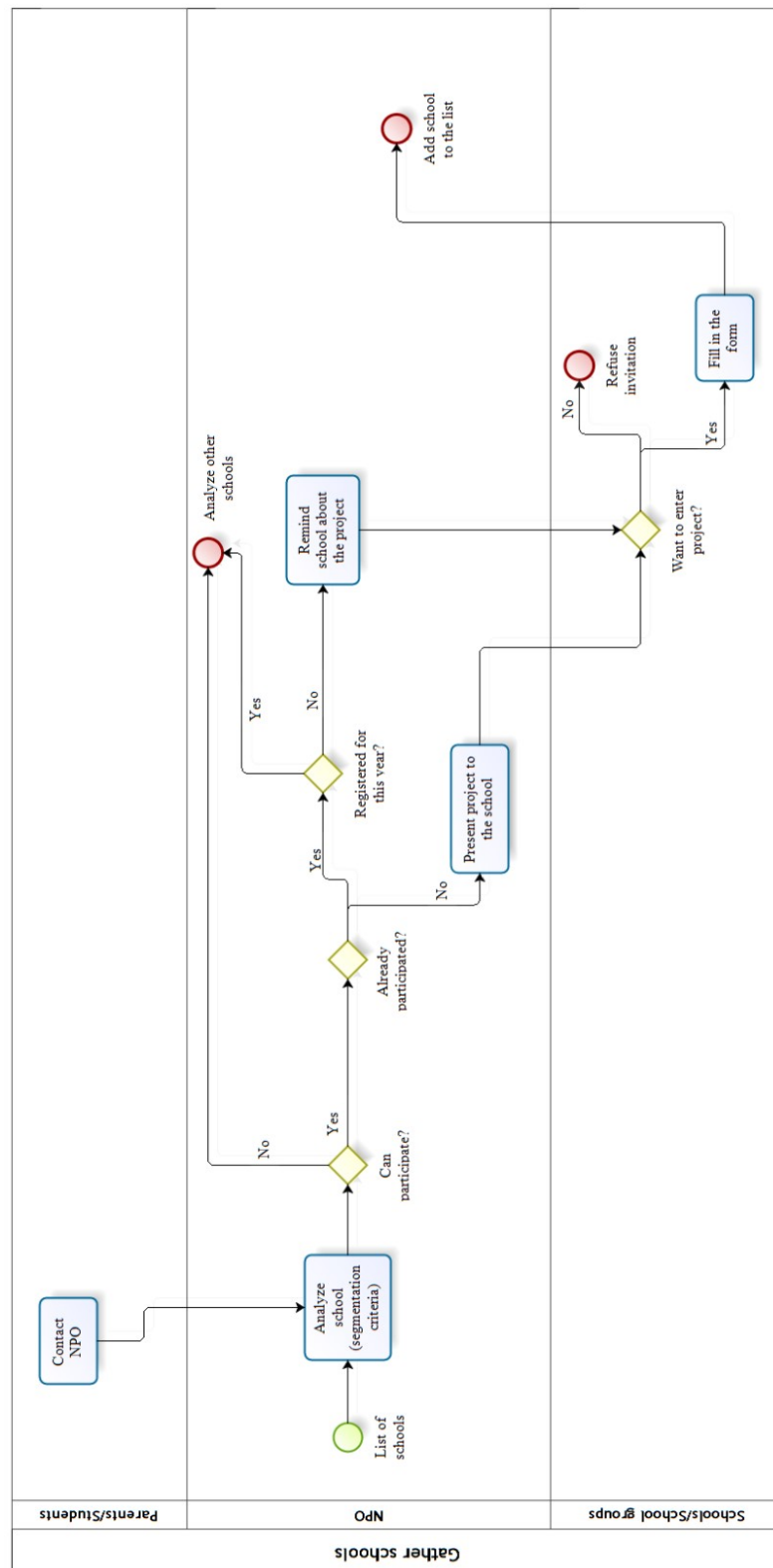
Figure A.1: Phase 1 - Gather schools
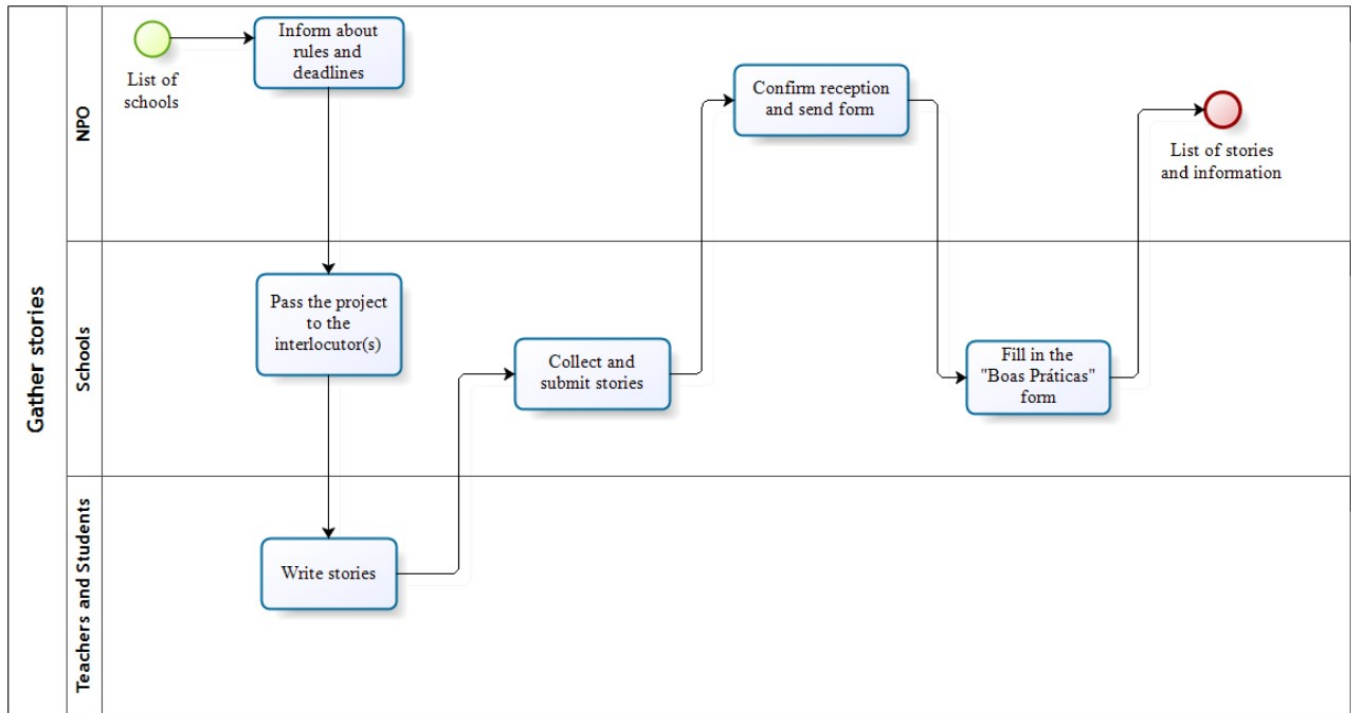
Figure A.2: Phase 2 - Gather stories
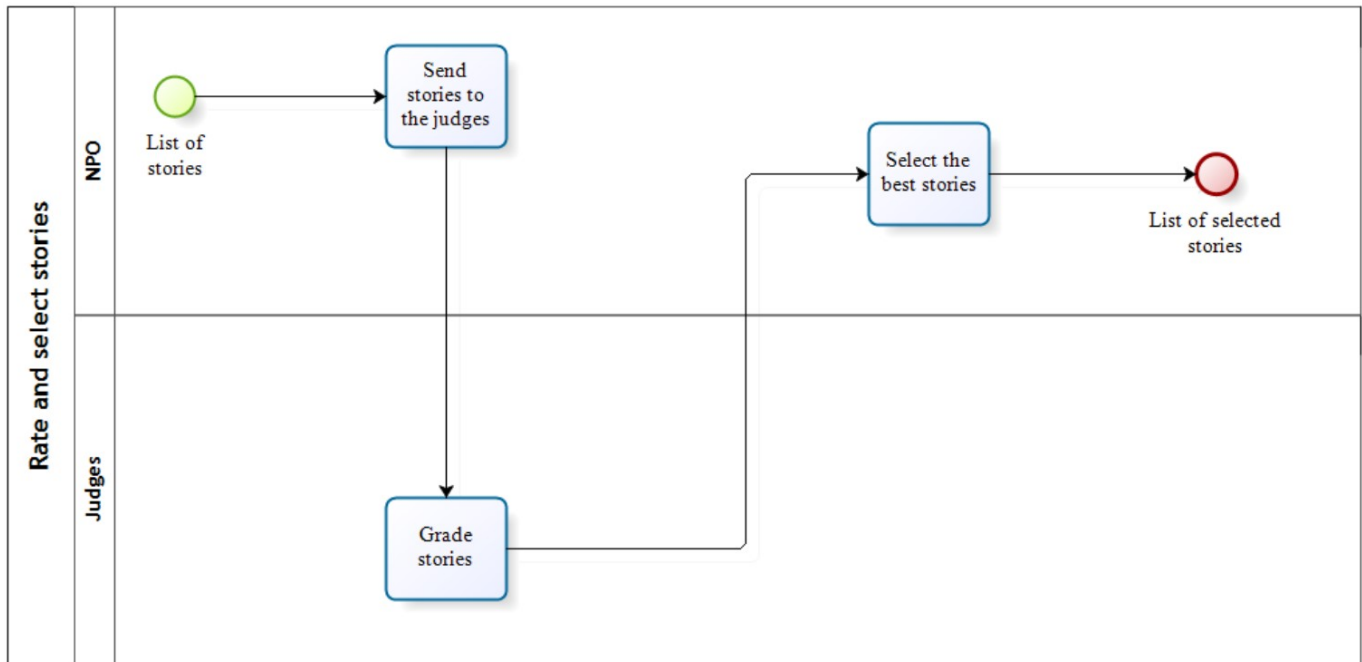
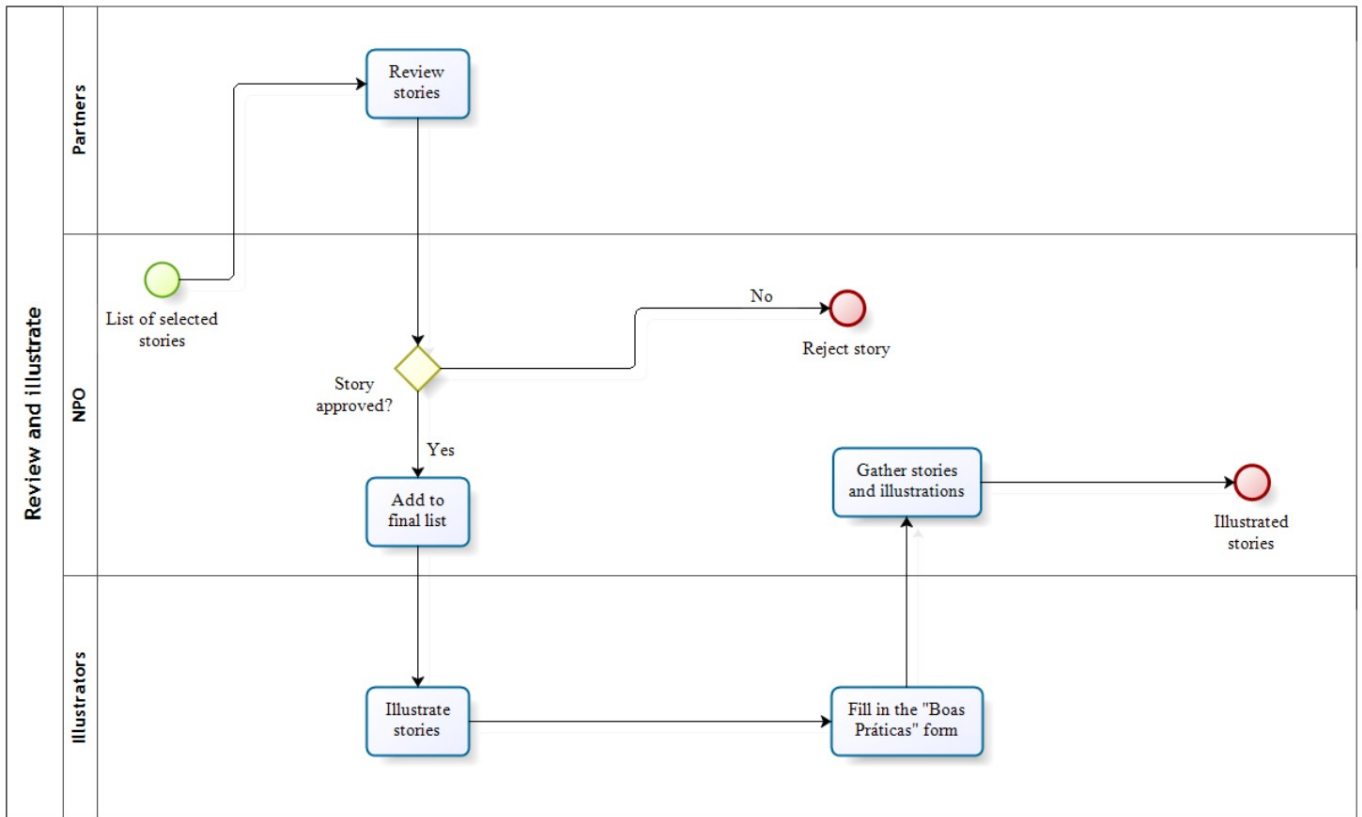Figure A.3: Phase 3 - Rate and select stories

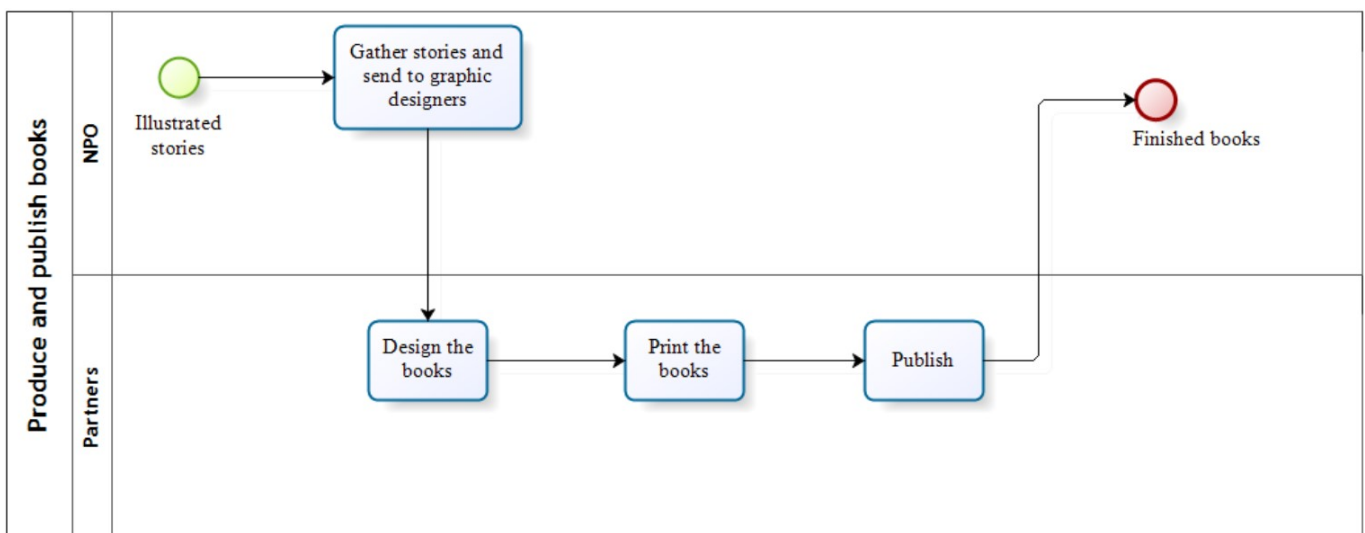Figure A.4: Phase 4 - Review and illustrate stories
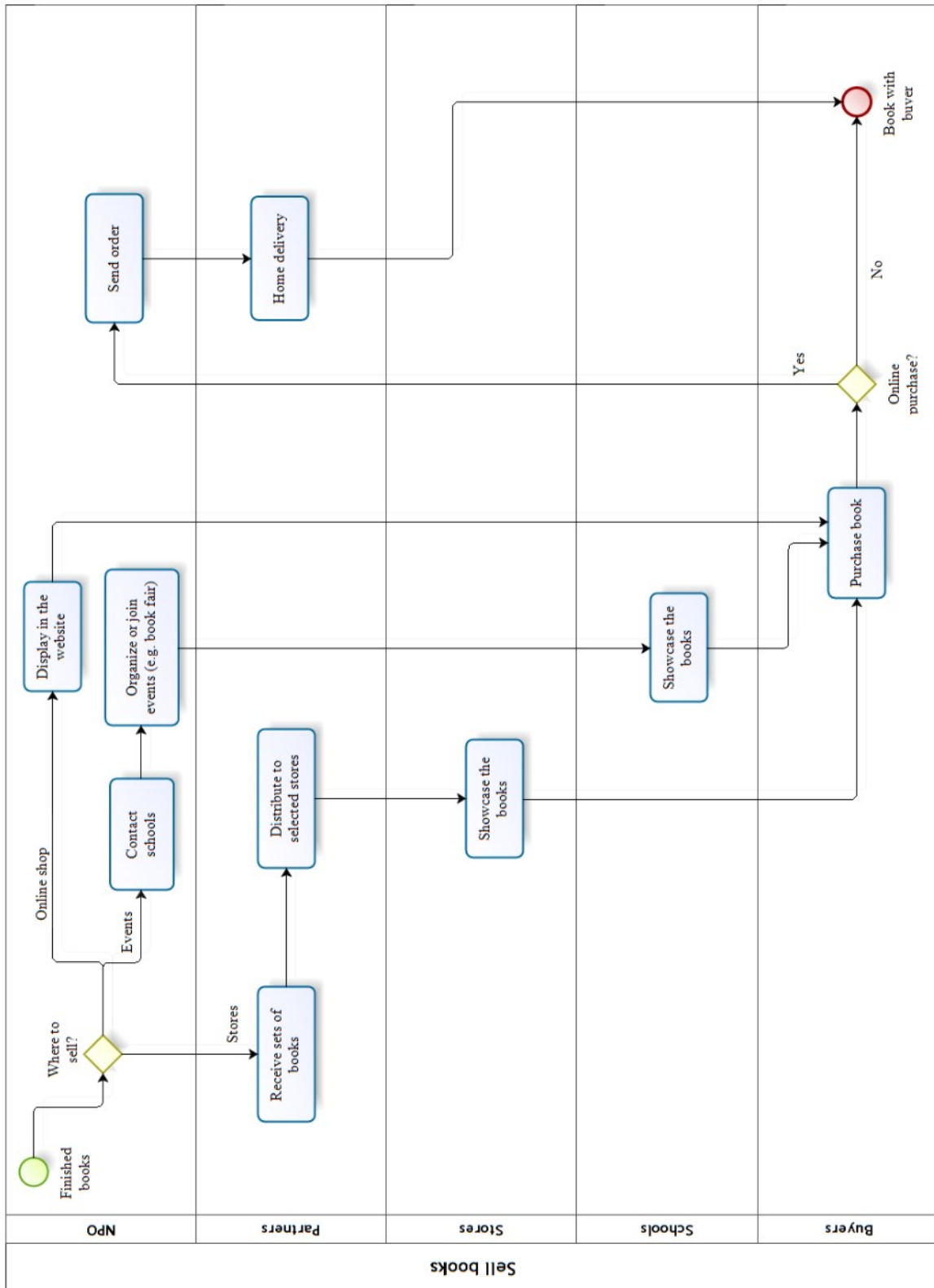

Figure A.5: Phase 5 - Produce and publish books

Figure A.6: Phase 6 - Distribution and sales

Appendix B

UML Diagram

Figure B.1: UML Diagram