A Work Project, presented as part of the requirements for the Award of a Master's degree in Business Analytics from the Nova School of Business and Economics.

Creating a Product to Segment Donors and Predict Donor Churn

Requirements Engineering and Data Visualization

BOWEN CHEN
44994

Work project carried out under the supervision of:

Qiwei Han

14-12-2021

Abstract

The present work explores the application of Machine Learning in a Non-Governmental Organization (NGO) - as proof of the benefits of extending Artificial Intelligence to non-profits - to support fundraising actions, by identifying donors' similarities and churn probability. Requirements engineering is the very first step of the development of a machine learning project. It is a systematical method to avoid many possible problems with early intervention. As the requirements engineering is managed and supervised, the data visualization, which usually as an outcome, can meet the requirements more accurately.

This paper introduces the concepts of requirements engineering and data visualization and applies the theories to a real-world business project.

Keywords: Machine Learning, Requirements Engineering, Requirements Management, Data Visualization, Power BI

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

Contents

I. Introduction

The following project is the result of collaboration between students of the Master's in Business Analytics and the non-profit organization Assistência Médica Internacional (AMI), with the advisory of faculty members and external consultants from Accenture. Fundraising actions represent heavy expenditures for charities (Andreoni and Payne 2013). As Non-Governmental Organizations (NGO) depend on external giving to guarantee their financial sustainability and aptitude to operate, it becomes crucial to find efficient solutions for these activities. In fact, to make a better use of limited resources, we propose targeting a restrained number of donors whose characteristics are in line with the NGO's growth strategy. To decide who to approach, when and how, we suggest grouping donors into different segments, for which tailored actions should be applied. NGOs might also consider the likelihood of them never contributing again, since costs incurred in for churned donors cannot generate return. We recognize Machine Learning (ML) as a possible option to accomplish these goals and consider the requirements implied for its appropriate use - a deeper examination of the data needs is performed in this work.

Section II defines the aim and motivation of the project. The Literature Review provides background on the use of data by NGOs, their limitations and how they can integrate ML. In section III, we individualize AMI's use case, briefly introducing the NGO and the agreed upon project scope. Subsection 3 regards the substantiation of these objectives, including the Methodology, data and models used. Subsection 4 contains preliminary conclusions and limitations. Based on these conclusions, in section IV & V, we study potential solutions to enhance the business requirements engineering process and design the final visualization report for better decision-making.

II. Motivation and Literature Review

The fast evolution of technology in recent years makes it an essential tool for business strategies. Artificial Intelligence (AI) encompasses the ability of a machine to learn tasks that would typically demand human intelligence (Turing 1950). AI englobes Machine Learning (ML), a process based on the perception of regularities through the training of data that has been increasingly adopted by companies (Jordan e Mitchell 2015). The commercial sector makes the most of technology to improve their performance in a multitude of areas, widely applying it to marketing to create high-level segmentation structures that could also be directed to social marketers (Dietrich, Rundle-Thiele and Kubacki 2017). Among Non-Profit Organizations' (NPO's) activities, we can identify two main stakeholders: the target group that receives free support, and the donor, who provides the help. Reaching out to the latter, through marketing and communication strategies, is vital to raise donations. Nonetheless, only a small fraction of NGOs acknowledges this priority (Dolnicar and Lazarevski 2009). Amid historical donors, optimal targeting has shown to perform better to increase donations than untargeted communication. Therefore, charities that ignore the need for differentiated approaches wind up wasting resources. (Cagala, Glogowsky, et al. 2021).

ML can be a good solution to identify resemblances among a population (donors) more automatically and efficiently, resorting typically to measures of distances to group similar objects (Berry, Mohamed and Yap 2020). Furthermore, it can observe historical data associated with a label (churned) as a pair, learn from it, and predict the probability for new donors with similar backgrounds (Jordan e Mitchell 2015). Despite these advantages, NPOs often do not possess the necessary means for the same technological innovation as other sectors due to budget restrictions, lack of trained staff and restrained data resources. Consequently, they rely mostly on simple segmentation methods (Dietrich, Rundle-Thiele and Kubacki 2017). In fact, higher granularity and complex combinations require more data collection and preparation, and the inherent expenses (Panda, et al. 2021). For better funds allocation, the organization should define what features should be gathered in advance. Optimally, data used in segmentation studies should reflect consumer (donor) behaviour. Demographic information, socioeconomic characteristics, historical and geospatial information severely influence the algorithms' performance (Cagala, Glogowsky, et al. 2021). However, the NGO is dependent on the giver's willingness to provide these details, and some might be perceived as intrusive or time-consuming, encumbering their collection. Thus, alternative solutions and data sources are essential (Chen, et al. 2018).

This paper reconvenes the benefits of adopting ML for segmentation and churn prediction to increase the values of fundraising for NGOs, using AMI's example. It aims to understand if, in fact, more complex algorithms can be used for NGOs, despite their above-mentioned limitations, to increase the efficiency of their processes. Furthermore, it relies on the importance of data quality, namely data collection and cleaning techniques, to reach the anticipated results in data science. Finally, this work investigates the ability of features that are widely recognized as beneficial for clustering to enhance the performance of the models proposed. According to the results, we intend to define whether these should be a part of the NGO's data collection strategy.

III. Creating a Product to Segment Donors and Predict Donor Churn

1. NGO Presentation

This project uses AMI's data to create models capable of being generalized to other NGOs. AMI is a private Portuguese organization whose mission is to provide humanitarian aid and to promote human development (AMI: Vision, Mission and Values 2021). Although AMI has supported over 10 000 people in Portugal through 15 social facilities, the organization is globally oriented, having carried intervention activities in 82 countries. As per mentioned by the NGO's representatives, it is composed of a small technical team, most collaborators work full-time and have non-technical background and little experience in computer science.

2. Context and Problem Definition

AMI has around 70 000 donors - 93% individuals and 7% companies - and a vast majority of Portuguese contributors (99%). Although the foundation relies on monetary donations to maintain its activity, there is a notorious absence of recurrence: donors are unlikely to donate more than once. The lack of regular donations, and consequential financial instability, challenges AMI's viability. To increase monetary contributions, the NGO should grow the number of active donors - those who not only registered, but also contributed monetarily -, the number of donations and the amount donated per individual. As a result, the following project scope was defined:

- **Create a product to segment donors:** group donors based on similarities. Consequently, the possibility of predicting fallbacks will be improved by virtue of a better understanding of each donor's willingness to donate and his/her characteristics. Furthermore, it will allow to identify the donors who are more likely to make regular donations and target them. Ultimately, this will improve the ability to retain donors and increment donations.
- **Develop a model to predict donor churn:** The model will render, for each donor, their probability of churn (the risk of not contributing to AMI again). This metric will be important to understand the willingness to donate, prevent churn, and allow more stability in the management of ongoing activities.
- **Create an interactive dashboard for data visualization:** This tool will help visualize data intuitively, providing easiness of interpretation. So, AMI's managers will be able to extract lessons and make data-driven decisions.

3. Data and Methodology

3.1 Data Description

The information was extracted in September 2020 from the NGO's own Customer Relationship Manager (CRM), where inputs are made by all business areas and the outputs are used by the departments of Communication, Image and Sensibilization, and Marketing. It contains details on the individuals and companies registered to AMI since 1990, as well as information about the donations, including their value and type (monetary, volunteering, goods, or services). Following General Data Protection Regulation (GDPR) rules, all the data delivered was anonymized and, therefore, could be used with no restrictions. It is recorded in six excel files containing only active donors.

Before an in-depth review of these records, it is relevant to establish two important distinctions. First, AMI can receive both monetary and non-monetary contributions. The first will generate a monetary flow, therefore including sales, donations, and funding. For the latter, the value of the offer is evaluated by AMI considering the time granted by the donor, in case of volunteer work, or the value of the good or service provided. Additionally, we should distinguish potential contributions from the ones that are finalized. AMI may contact previous donors to assess their interest in participating in new initiatives. If the donor does not respond, it will be classified as “potential”.

3.2 Exploratory Data Analysis

With the intent of gaining a better understanding of the data provided, an Exploratory Data Analysis (EDA) was performed. The main purpose is to understand the data prior to making any assumptions that might affect future decision making. It focuses on finding relevant patterns, correlations, outliers, and mistakes. To simplify interpretation, visualizations were created using the matplotlib and seaborn libraries in Python. The inferred conclusions from the current analysis are useful to support the decisions on the procedures to fill missing data. The general approach to dealing with missing values started by defining a threshold of 70%. Following GDPR rules, AMI can only contact individuals who gave their consent - hence, records whose consent date is not Null. Consequently, AMI only has permission to approach

0.17% of their single donors, blocking the implementation of targeted communication. Since this is a legal requirement, it cannot be deleted.

(i) Analysis at Donor Level

AMI provided data on 62 032 single donors and 3 164 companies, made up of their characteristics and demographics. However, from a total of 31 variables regarding information on individuals, as per the established threshold, only 14 variables were considered usable. In reference to the companies, we preserved 10 out of the original 14 features. Thus, decisive characteristics that could help improve segmentation (individual's age or companies' sector) were disregarded. The amount of missing data is one of the biggest obstacles to the project and must be considered throughout all decisions. There is a balanced distribution of donors by gender (51.8% men, against 48.1% women). We can also assume that AMI's appeal is mainly national: 99.75% of single contributors are Portuguese and only one company is foreign (Spanish). On average, an individual will donate 130 € to AMI in their donor journey, and a company 3 652 € [1]. The total value offered by a company varies significantly, ranging from 1 to 4 103 803.93 €. Most single donors (91.8%) are not registered as "AMI Friends", a loyalty program that encourages frequent donations, and 3.5% of companies partake in this initiative. AMI allocated every individual that did not have a known registration date to 1990, ergo the peak in enrolments in that year. Registration sources were grouped into 13 categories [1] for a simpler analysis of its impact.

[1] These values refer to the sum of all contributions ever made by a donor - single and company, respectively.

(ii) Analysis at Donation Level

Potential Contributions: AMI initiated the recording of potential contributions once the NGO transitioned to CRM, in 2018. Since CRM implied a higher granularity and maintenance of information, data quality improved. However, pattern recognition is hampered by the low number of observations - there was only one full year of data (2019) at the time of the analysis. Most contributions are set as “Lost”, most likely because AMI makes untargeted contacts, automatically creating leads that do not generate response. This enhances the value of the present project: targeted decisions on segments help trigger actions by the donors.

Finalized Contributions: These records were either imported from AMI’s old database, or already had a receipt associated, indicating the transaction was completed. These contemplate 170 750 logs from individuals between January 2004 and September 2020. As for companies, the team had access to information updated in February 2021 to improve model performance. It portrayed a total of 16 979 company contributions starting in 2004. Less than 50% of the individuals in AMI’s database contributed at least once during the time frame mentioned. The maximum number of contributions made by a donor was 399. On average, each person donated 6 times and each company 4 times. However, most are one-time donors (61% of single donors and 57% of companies). The number of contributions per year has been tendentiously decreasing since 2011. GDPR rules restricted proactive contacts to subscribers to raise awareness and propose contributions, aggravating this trend. On average, a person will assist AMI 125 days after the previous contribution. Over 90% of donors had not made any endowment in the past 1.5 years, at the time of the analysis, coherent with the fact that most of them are one-time donors. Over 99% of contributions from individuals (60% for companies) refer to donations -voluntary transfers of money. The average value of received from individuals is of 54 €. If we consider only non-monetary contributions (goods and services, volunteer work, visibility, and rec cling), this comes up to 718 €. As for monetary, the average value collected is 50 €. The average value of a giving made per company is 1 137 € and has been increasing since 2018. Monetary ones are lower (932 €) than non-monetary (1 489 €). The number of contributions made by a donor and their value have a correlation of -0.04, indicating independence.

3.3 Methodology

To achieve the goals propositioned in subsection 4, we will resort to ML algorithms. Their application requires an understanding of the data used, as well as its cleaning. The decision on the best models requires the testing of multiple algorithms and the comparison of their outputs. The final verdict will be based on predefined measures - suitable for each algorithm - and important criteria, according to the project scope.

3.4 Pipeline and Data Preparation

(i) Data Cleaning

Once data ingestion was finalized and information from the different sources was joined, it was prepared for the model. Consolidation and cleaning of the data involved the deletion of outliers, including errors - for instance, donors registered in 1930 (prior to the NGO's foundation) - and numerical variables whose z-score[2] was higher than five. Missing values

[2] Numerical measurement that represents the distance of a value to the group's mean.

were filled according to context inferred internally and, when necessary, an external data source, the Iberian Balance Sheet Analysis System (SABI)[3] was used to get firmographics on the companies. We transformed the data types to meet the model's requirements and performed feature engineering, including upstream categorization (grouping values into wider categories) and the creation of additional variables (e.g., days since donors' last contribution). Finally, we resorted to the "scikit-learn" library to encode categorical variables (one-hot-encoding) and standardize numerical features (standard scaler).

(ii) Feature Selection

After feature engineering, 494 variables integrated the final table. Feature selection followed a multidimensional approach that combined analytical decisions - deleting attributes based on the lack of data quality or high levels of correlation - and functional evaluation, based on AMI's experience. It is proposed as functional because it considers a practical interpretation of the problem, centering on an intuitive view of what characteristics should be key in a segmentation. At last, we tested different combinations of variables and selected the one that provided the best result[4].

---

[3] SABI is a database with information on companies from Portugal and Spain.

[4] Defined in subsection (iii) and specified in Appendix.

4. Model Application and Evaluation

4.1 Segmentation Model

(i) Algorithms attempted

Since there are no previously defined clusters, we can consider there is no target variable for the segmentation. As such, we tested several unsupervised learning models[5]. The first attempt used DB-Scan [2], a density-based clustering algorithm (Scikit-learn 2021). For this algorithm, a minimum number of donors is defined for each cluster. Similar donors are grouped into the same cluster, and those that are in regions with a density below the minimum threshold are marked as outliers. However, even with a low number of necessary neighbors for each cluster, most donors were classified as outliers, which is hardly interpretable.

K-Means [3] is a fast, widely used method that minimizes the inertia (maximizes the cohesion within the same group), by partitioning the observations to the clusters with the nearest mean (Scikit-learn 2021). The algorithm was experimented with multiple combinations of variables and the elbow method (Soppin, Ramachandra and Chandrashekar 2021) was applied to derive the optimal number of clusters for each trial. PCA [4] was applied before K-Means, for a two-dimensional representation of the model, facilitating the comparison of the performance of the different sets of attributes (Brownlee 2020).

[5] Algorithms that find patterns and train on data that does not possess a label associated (Jordan e Mitchell 2015)

(ii) Model Evaluation

The best option should be heterogeneous inter-cluster, whilst coherent intra-cluster, so that it is easy to characterize a donor based on the group he/she is inserted in. Separability and interpretability are key when evaluating the model. To ensure these, two-sample t-tests and one-way ANOVA were employed at each attempt. The former evaluates if feature values were different from one segment to another, by examining whether the differences are statistically significant (P-value< 0.05) (Webster 2013). The latter tests the null hypothesis that two or more groups have the same population mean. The silhouette-score - a similarity score where "the best value is 1 and the worst value is -1" (Scikit-learn 2021) - was also considered when comparing the distinguishability of the outcomes, i.e., whether the object resembles its own cluster and not the others. Finally, a homogenous size of the clusters was used as cohesion indicator.

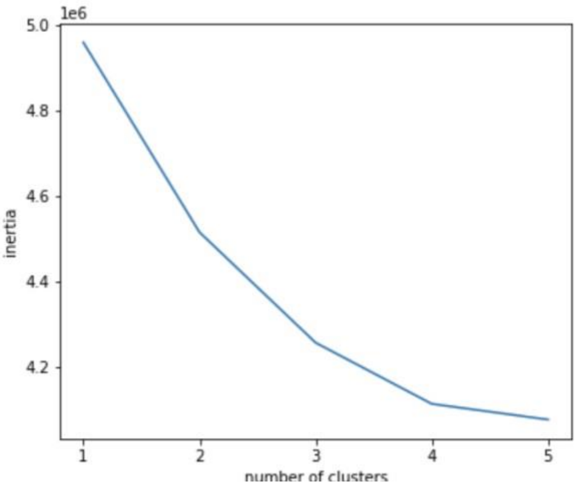
(iii) Model Decision

The best model focuses mostly on interpretability, to be user-friendly for the client. It outperformed the others for all metrics [5], but the silhouette-score. Standardizing the data can make separability more difficult, which could explain this lower score. The final model used 4 different clusters and a set of features[6] and characteristics heterogeneous enough to embark efficient marketing actions. Cluster 0 ended up with approximately 7k donors, 1 with 5k, 2

[6] Included demographics (gender and region); firmographics (field and legal form); consent; information on the value, frequency, and recency of the contributions. The name of the variables can be found in appendix [7].

[7] Supervised machine learning retrieves an output from an input based on pair examples (Jordan e Mitchell 2015).

with 24k and 4 with 31k. Cluster 0 was found the persona most important to retain, as they donate more than once and a relatively high value in their lifetime as donors [6].



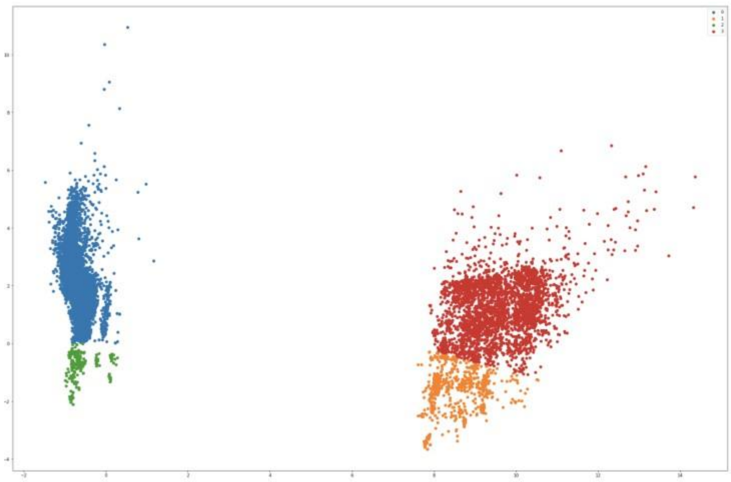


Figure 1 Elbow method with selected features.

Figure 2 Clusters from K-Means application (k=4) with selected features.

4.2 Churn Prediction Model

(i) Churn Definition

For exiting prediction, the team used a supervised model[7] where the target variable should indicate whether the donor is churned. Since AMI did not have a definition of churn prior to the project, the team defined, together with the NGO, a list of rules for this classification. First, only people who contributed with a monetary donation at least once should be eligible for this evaluation. Then,

- If the donor only donated once, then the time since last contribution was taken into account. If the number of days since that donation was higher than the average number of days between contributions for all donors, the donor was classified as churned.
- If the donor contributed multiple times, the team compared the last contribution to the donor s usual frequency. If the number of days since the last time the donor

contributed was higher than the average number of days between contributions for that donor, he/she was classified as churned.

(ii) Pipeline for Classification Model

Most of AMI's givers are one-time donors and, consequently, have churned. As such, our data is imbalanced. We employed synthetic minority over-sampling technique (SMOTE) (Imbalanced-Learning 2021) [8] to generate the same number of observations for the minority class (not churned) as for the majority class (churned), so the model can learn better. Synthetic points are generated from the potential points that underly the pairs of actual data, by assuming they would follow a similar distribution, i.e., anything between 2 points from the minority class is likely to be from that class too. The resulting samples were used to train multiple classification models. The team used grid search [9] to decide on the optimal values for a set of hyperparameters including the penalty to apply (L1 or L2), to avoid overtraining (Géron 2017) (Scikit-learn 2021).

(iii) Model Evaluation

The model is designed to identify new donors who are at risk of churn. The data was split into a training set and a test set to evaluate the capability of generalizing the model to donors not used to train it. Accuracy and precision were chosen as the most relevant metrics to indicate the quality of the predictions. The first returns the percentage of correctly predicted points out of the total points, the second returns the percentage of points predicted as positive (churned) that are correct (True Positives). If a person is at risk of churn but is not predicted, AMI will apply to them undifferentiated actions. If they are wrongly predicted as churned, AMI will use part of their resources for targeted actions and increase their costs, thus False Positive cost is higher.

(iv) Model Decision

Since the results for the baseline model (Logistic Regression [10]) and the more complex ones (XG-Boost [11]) were not significantly different, the team opted to use the Logistic Regression. This still allowed to get a precision of 99.9% and accuracy of 99.73% for the test set, while performing faster and implying lower costs than the other options. We should consider that the results of the model are highly influenced by the fact that the dataset usable for the churn prediction is very small. For the test set, a total of 5 285 donors were predicted as churned, versus 5 213 not churned (only monetary contributors considered). As AMI considered that the probability of churn would be more useful for their business decisions than a binary classification, the final output was changed to meet those needs.

4.3 Fairness and Bias

An important part of the evaluation of the model is understanding if our classification model is fair and unbiased. We used two variables that could create an unfair model: gender and region. We calculated the recall parity with male and Lisbon as reference values, using the "Aequitas Audit Toolkit". Balanced outcomes between the different categories of the chosen variables lead us to believe the outputs of the model are fair. The segmentation was not accessed in this metric because it is, by definition, an agglomerate of similar individuals (intra-cluster) that should be addressed differently (inter-cluster).

5. Partial Conclusions

ML allows to make prompt predictions of the likelihood of a giver not donating again, based on his characteristics and previous records. The model is learning from a highly imbalanced dataset: there is a higher percentage of individuals churned than not churned. Two factors justify the unevenness: i) Regarding the clustering algorithm (K-Means), the quality of the results was conspicuously affected by the high percentage of missing values and lack of variance in the features. The model's outcomes presented low separability, translated into a low silhouette score. While attempting profiling of the groups, the chosen model allowed to segregate the clusters regarding behavioral characteristics (e.g., the number of contributions), helping to recognize the most valuable segment. Some traits were also identifiable (for instance, it was composed of a higher percentage of women - 71% - than men), however others, including the pattern in region distribution, were common to all clusters. This, along with the fact that there was a small number of usable demographics, hinders the profiling ability. ii) Most individuals contributed once to AMI (therefore, churned afterwards); secondly, the dataset includes givers who registered over 20 years ago, hence with a distant last contact. Even so, by creating synthetic points of the minority group, the algorithm learned to make accurate predictions of churn probability. Thereupon, AMI can define whether to invest in communicating with certain donors, according to their value. To further improve the decision making, requirements engineering was done to validate the models' functions and avoid possible future conflicts, and data visualization was produced for results presenting, which will be covered in section IV & V.

IV. Requirements Engineering Introduction

1.1 Business Requirements

In the area of software engineering, business requirements is the process of eliciting and documenting business requirements of users in the first development step of the system design Business requirements from all the stakeholders are the keys to fully understand the needs of the project, so that to initiate the project with a clear target from a holistic view of the whole process system. Requirements can be converted into the construction goals of the software system, which can be reflected in two aspects: problems and opportunities. (Beal, 2012).

The problems could happen during the procedure of demand capture, which is used to describe what results the users need to obtain and how to achieve these goals. Different distributed users will put forward demands from different angles, levels, and perspectives. However, sometimes since users are at different levels of the organizations or departments, there will be the phenomenon of the requirements are not collected or organized by the real end-users, instead, only several representatives are involved in the requirements acquisition. Even further, the large stakeholder base involved in collecting the requirements might cause conflicts in interests. Or the requirements collected from different users are unambiguous and contradictory.

Throughout the process of business requirements, there are some opportunities to be seized. Firstly, early defined requirements can help reduce failures of the project those due to unmatched requirements and expectations. Secondly, a well-structured and documented business requirements can create healthy consensus and improve collaboration amongst all stakeholders. Finally, it can largely reduce the costs which may occur for the changed requests and related items. (Goldsmith, 2004)

Reflected in the requirements document, it can be expressed as follows: business requirements are generally reflected in the project scope document, user requirements are presented from the user's perspective through the use case document, and software requirements need to be written in the project scope document.

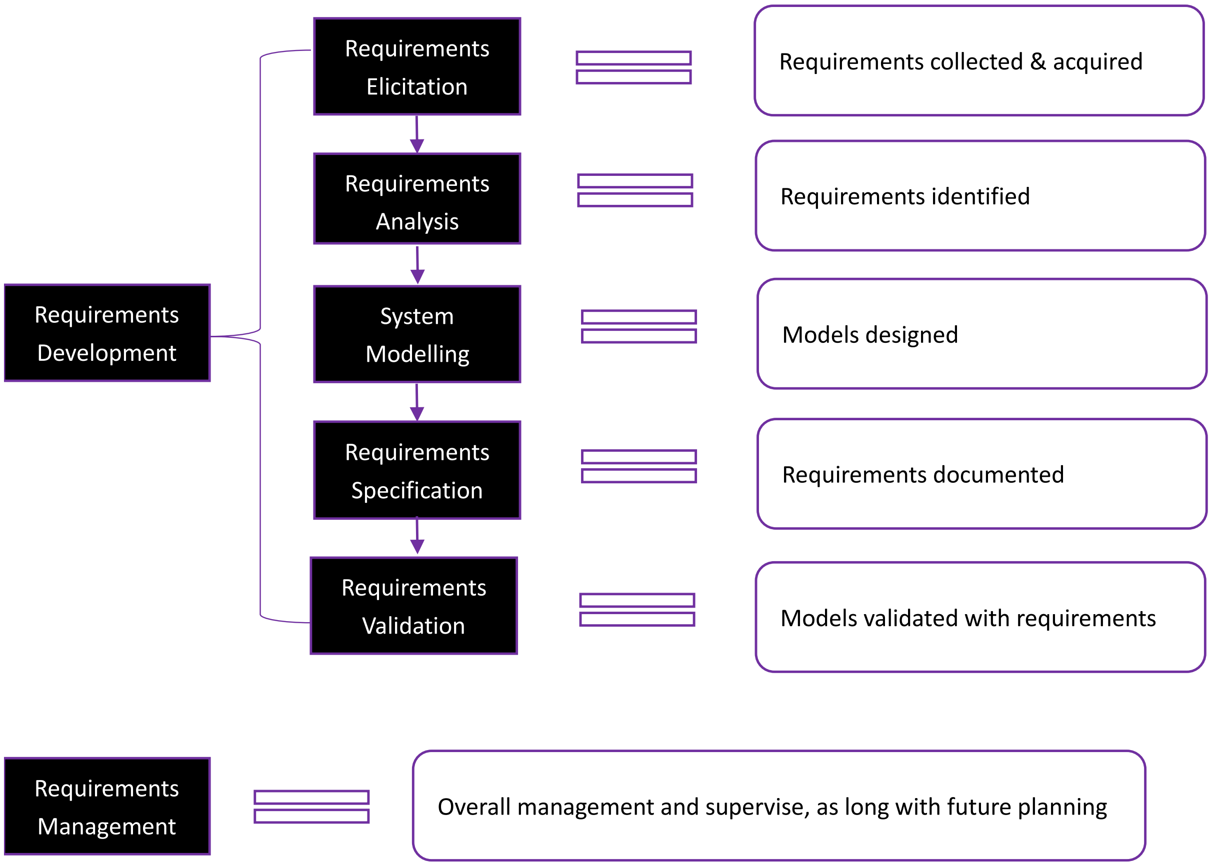
2. Requirements Standards

Excellent requirements must meet the following standards: a) completeness: verify that the subject areas, business events, business activities and business steps listed are complete; b) correctness: strive for the information transmission process with no distortion, namely, correct and unambiguous; c) priority: distinguish effectively the priorities to better manage the needs; d) early technical intervention: the solution must be both feasible and verifiable in case the requirements were incomplete and distorted.

3. Requirements Engineering

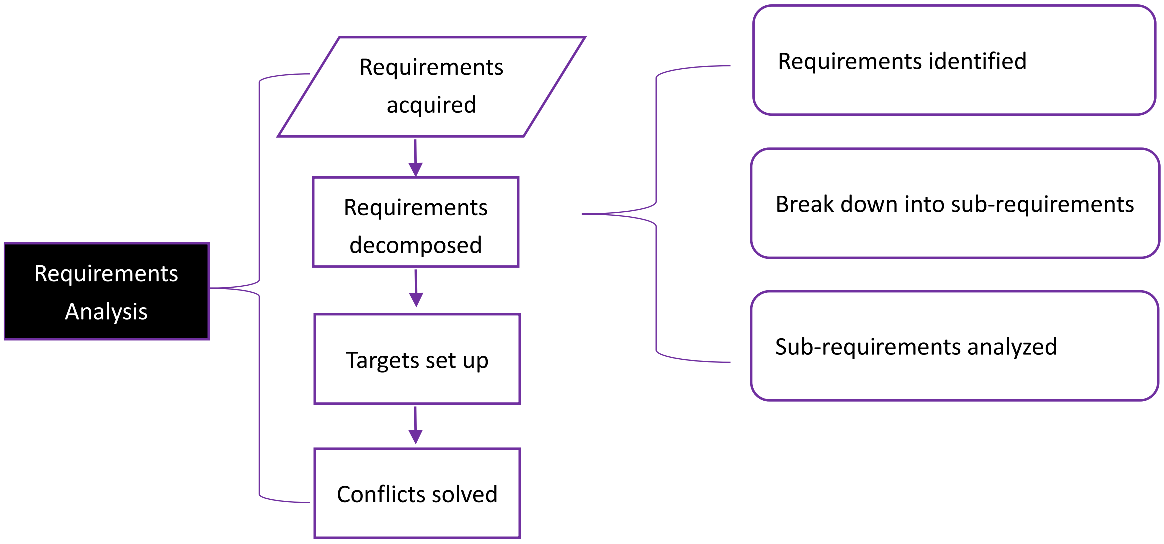
Requirements engineering is a series of activities aiming to define, document, and maintain requirements. (Kotonya & Sommerville, 1998) It includes two process areas, which are requirements development as well as requirements management. First, requirements development is the process of collecting, analyzing, organizing, compiling and verifying requirements, with the focus on developing high-quality requirements specifications. Second, requirements management is the procedure of supervising and tracking the realization and changes of requirements, emphasizing on ensuring that the software developed meets these requirements.

The requirements development is carried out in a multi-period form of five specific activities: requirements acquisition, requirements analysis, system modelling, requirements specification, and requirements verification. (Sommerville, 2009)



Requirements elicitation is also called requirements inception, which is the process of the acquisition of the demands of the output from stakeholders.

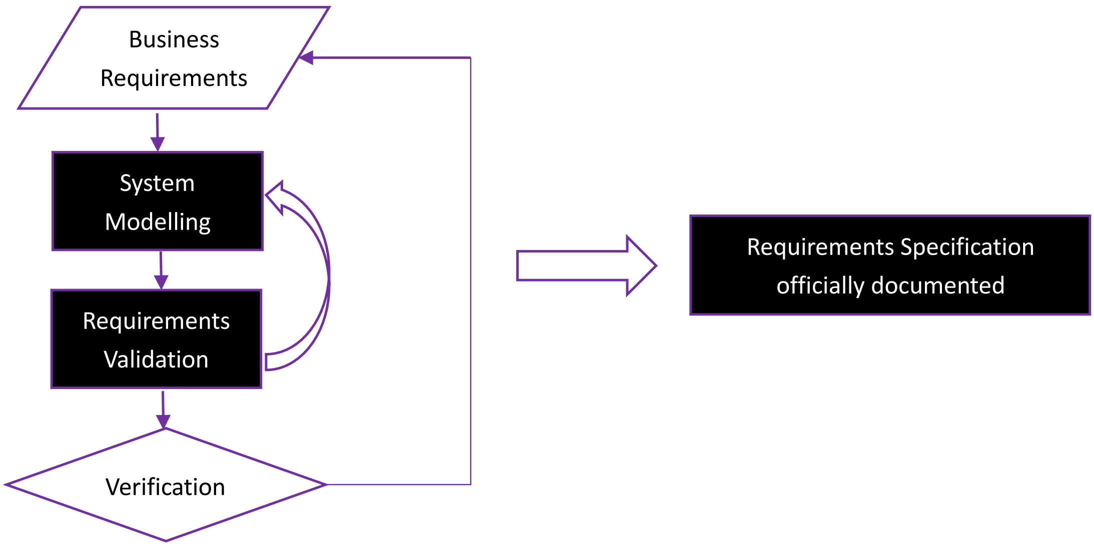
Requirements analysis is the core task in requirements development, but it is often neglected in practice, namely, the information captured by the requirements is directly classified into the requirements specification. Requirement analysis is first business analysis, which is to study the problem domain from business clues instead of the system structure, so that the general background information is acknowledged. Requirements analysis is a decomposition activity, which divides the system to be developed into different subject areas according to responsibilities. It can be perceived as subsystems, which are carried out according to the business perspective when dividing. Moreover, during requirements analysis, contradictions, and conflicts of requirements from various stakeholders should be identified and solved. (Berenbach, Paulish, Katzmeier, & Rudorfer, 2009)



System modelling is presented here for some engineering fields and some specific circumstances which may require the model to be entirely contrived before its construction kicks off. However, in some areas such as software engineering, system modelling process are categorized as part of design events, thus not as requirements engineering activities.

Requirements specification is an official artifact where requirements are documented, and will turn out to be authorized after the process of requirements validation. A requirements specification may consist of both written and visual information in the form of graphics.

Requirements validation is applied to confirm that the requirements documented and models designed are consistent and are accordant to stakeholders' demands. After the stage of validation, the documented requirements specification becomes official and authorized, which represents the accomplishment of the process of requirements development. Note that the different between requirements validation and verification. Validation is to validate the result of the model with the model itself. And verification is to verify the result with the initial business requirements.



Requirements management is the final step for requirements engineering. All processes, which associated to the requirements since the project initiation until it is put into use should be managed by supervising while the system is developed, in the event of modifications, postponements, etc. In this manner, it is possible to trace and prioritize requirements to limit changes may suggested by stakeholders.

4. RE analysis of AMI and system design

4.1 Ami Requirements Background

AMI, as previously introduced, has around 70 000 donors in total, with about 93% and 7% companies. Normally, an NGO pursues to make changes in minor scales straightly through different projects. There, it is crucial for an NGO to successfully launch programs to mobilize funds, materials, and volunteers, so that further local programs can be continued and thus, forming a virtuous cycle. In some occasions, an NGO conducts fund raising events and applies to governments or other organizations for grants or fundings to raise money for further projects. Although NGOs mostly are based as community-wise, they can also be either national or international. Therefore, as an NGO, AMI relies heavily on monetary donations to maintain its routine operation and future projects. Nevertheless, it is difficult for AMI to take accurate marketing actions because they know neither who are donating nor how the donors are. It is said that they have large information in their CRM database whilst they cannot find a way to utilize them. The fact of losing donors and lacking for new donors are the main problems they run into.

4.2 Assess risks and limitations

Imitating the definition of risk, the risk of a requirements engineering could be diverse. Many factors that happen during the activities in requirements engineering could be interfering the outcome. However, it is indicated that the most common problems are insufficient requirements, constantly changed goals, while relatively rare problems are blemished communications, deficient traceability, terminology problems, and ambiguous duties. (Méndez Fernández & Wagner, 2015)

Also, compliance with EU data collection and processing rules on GDPR is considered the most influential constraint. In addition, it is assumed that AMI will be updated regularly, because there is no new information, the model may be out of date. The team is also responsible for ensuring that clients understand how the project will produce results and educating collaborators to use the tool.

4.3 Requirements engineering for AMI

(i) Requirements elicitation

In the project scoping process, the background information of AMI has been collected and communicated. It is known that AMI relies on monetary donations to maintain its activities, however, AMI is under the situation of lacking repetition of donations, meaning that donors are unlikely to donate multiple times. The irregular monetary donations and the consequent financial instability challenged the viability of AMI. Therefore, AMI feels urgent to find an approach to increase monetary donations from the public as long with the amount of donation value per donor. Besides, AMI is facing the problem that some donors are only registered in their database without any donation records. A flow chart is drawn to concretize the processes.

(ii) Requirements analysis

Beforehand, it is crucial to identify all the stakeholders and to document their respective demands. In the AMI project, there is one main target or demand, which is to improve the total monetary donation for AMI. Under this demand, there are several ways to resolve including their direct requirements.

First, in order to accurately target the potential donors and marketing straightly towards them, the team needs to create a model to divide donors into various segmentations to perform the marketing. The segmentations are ought to be precise, so that the receivers should be persuaded whilst not be bothered and then churned. Also, a better understanding of each donor's willingness to donate and its characteristics will be instrumental in preparing future marketing campaign. In addition, the donors who are most likely to make regular donations can be identified and located. Afterall, AMI will obtain the ability to retain donors and increase monetary donations.

Second, since there is some one-time or even zero-time donation, it is required to produce a model to predict the probability of the donors' churn rate. The model should also provide a way for AMI to distinguish their potential monetary donors who would possibly donate multiple times and start donation launch for whom never donated, and to prevent in time those are more likely to churn.

Thirdly, since AMI is an NGO, we are interested in their project launching situations, which are important not only to fund raising, donors or volunteers mobilizing, but also to form a healthy virtuous cycle in order to maintain the operation of the organization. In addition, the recognition of the advantageous projects and campaigns can direct AMI new movements to attract more donors.

Lastly, in light of the experience for users from AMI, it is pivotal for them to fully understand the results produced by the models. Thus, the design of the final report, which is a power BI dashboard, should be highly inclusive and intuitive in a manner of user end friendly. The related topic such as data visualization will be further discussed later in the paper.

(iii) System modelling

Echoing the previous requirements analysis on the marketing strategy for AMI, from the perspective of requirements engineering, it was significantly valued about the campaigns and projects those donors was registered through. However, since this variable is far from the threshold agreed, it was discarded. Afterwards, the model is conducted through python. An unsupervised model is created for donors' segmentation and a supervised model is created for donors' churn rate prediction.

(iv) Requirements Specification

The requirements specification consists of functional and non-functional requirements. Functional requirements indicate the function of the system, and non-functional requirements suggest the properties of the system. Therefore, the system designed for AMI is specifically creating the two models required and the final dashboard report as presented in project scoping process. And from a general view of the system, a pipeline is created.

Also, in terms of the two-parties confirmation on the requirements specification, every process of the project is confirmed and signed by the team and AMI by the means of interim reports.

(v) Requirements Validation

To validate the model, we need to find an acceptable solution between bias and variance. Cross-validation is used here, and it divides the data into three parts: training data set, validation data set and test data set. The training data set is used to feed different models. The performance of the model on the validation data set is used as the criterion for model selection. And the test data can get the results from the trained model to see if the model is doing properly. In this case, the two models for AMI are both working as designed to segment donors and to predict their churn rate, which is consistent with requirements collected and analyzed.

V. Data visualization

1. Visualization Technical Methods

Data visualization is a systematic and technical research on the visual depictions of data. Data visualization can be reflected as a mapping between the curated data and graphic elements. The concept of data visualization is constantly expanding, referring to newly innovative methods. These methods realize usages of graphics, various charts, moving visions, and interactive interfaces to explain ideas or conclusions drawn from the visual representation of data through expression, modelling, three-dimensional, curved surfaces, attributes, and even animation. (Nussbaumer Knaflic, 2015)

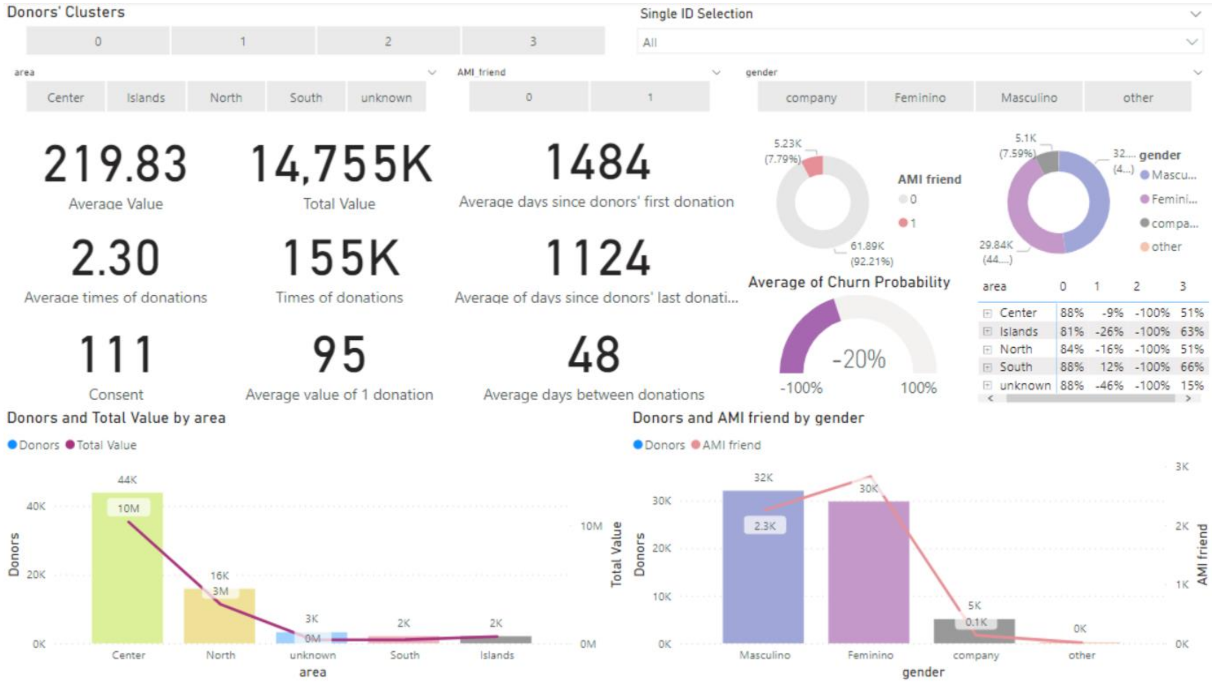
2. Value of Power-BI

The tool used here for data visualization is Power BI, which is a business analytics service aiming to offer interactive visualizations and business intelligence competences, and being end users friendly. The advantages of Power-BI are abundant. Firstly, it extends the data from various databases directly to end users, realizing the end-to-end informatization. Secondly, information Search engine can be a timely and accurate presentation of key information, allowing decision makers to stay away from the torrent of information. With external links and drilling of relevant information, users are no longer subject to traditional restrictions. Lastly, it can help make decisions driven by data, instead of instinct. It is necessary to not only know the results, but also the reasons.

With highly plentiful libraries and powerful visualization functions, Power BI can make interactive dashboards and, so as to analyze the data of the organization, simultaneously, taking advantages of visualization. Furthermore, the dashboards created and designed can publish and share across global data centers internally and externally, which can effectively meet the needs of AMI for cross-department cooperation.

3. Planning Report Structure

To design the dashboard report for AMI, the report structure is planned based on the requirements engineering analysis in the previous steps. Firstly, to manage the content structure in each topic, the main page has several keep indicators selected, such as *total value, average days between donations, AMI friend, etc.*, and arranged according to the importance level of the analyzed content. The layout method considers equal division, symmetry, and golden ratio. Secondly, to ensure the alignment of related topics, two pages are created for the two different indexes: count of donors and total value donated, so that end users can analyze the data from different aspects. Thirdly, to maximize the use of report space, enhance the cohesion of content, and avoid a highly compact layout, bookmarks are used to fold, expand, and switch the same content and clusters are used as slicers to present the same KPI with different perspectives. In addition, the overall report employs a navigational connection structure, and a single report page uses drill-through for cross-analysis. Next, adequate report descriptions are provided, such as report page descriptions, indicator descriptions, and interaction descriptions. Following is the main page of the report.



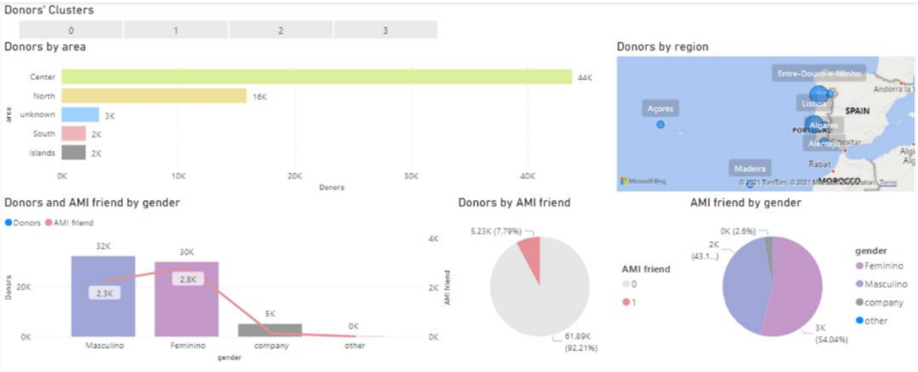
4. Data Transportation and Visual Design

After the dashboard structure planned, it is crucial to introduce data transportation. Considering the IT ability of AMI, the procedure of data transportation from their internal database to the final visualization report should be as concise as possible. During the process, the team can extract, transform, and load (ETL) massive amounts of data from any data source into the data warehouse according to different analysis topics and through full and incremental data update strategies. In this case, AMI would like to use data from their internal CRM system. Thus, there are several steps for them to follow: a) extract data as csv. files from their system; b) run the data curation process in the form of python notebook; c) run the model in the form of python notebook; d) refresh the dashboard in Power BI.

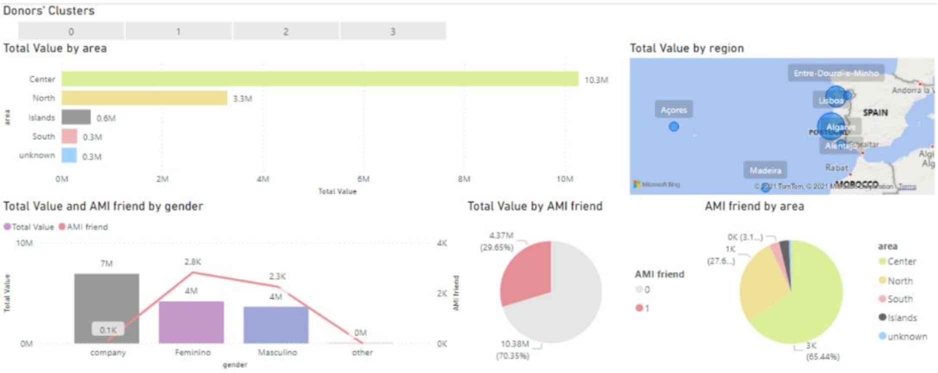
In order to improve the practicality and design sense of the report, some optimizations have been made in the visual design. The following are some principles that should be followed: a) clear content level: The commonly considered method is to build a content hierarchy and highlight key content by changing the background color and divide the content with dividing lines and titles. b) appropriate white space and strict alignment of visual elements. c) semantic unification: including color meaning, unit, display style, terminology, etc. d) unified layout: including unified page size, view type and background settings; unified page margins, printing plate spacing, etc. e) unified format: unified chart general settings (background, shadow, border, etc.) ; unified chart format settings (legend, X/Y-axis, data labels, etc.) ; unified font size, etc. f) consistent design language: consistent screening area, Consistent navigation design, consistent functional design. (Munari, 2006)

5. Enhanced interaction

Appropriate interaction can help readers better explore the dashboard, where interactions belong to the following categories: content jump, level drill and drill, filters, and parameters. For example, each report page will provide other dimensions of filtering, so that report users can better segment and explore the data. Also, multiple places in the report will use the drill-down function to achieve cross-analysis between different parts. In addition, the most used interaction method in the entire report is to use bookmarks to expand and collapse content, switch, and jump content, which helps users understand the data from different perspectives. Moreover, being accordant to the requirements, an interactive map is designed for AMI to understand the data more intuitively and geographically.



gender / area	company Donors_count	company Churn_prob	Feminino Donors_count	Feminino Churn_prob	Masculino Donors_count	Masculino Churn_prob	other Donors_count	other Churn_prob	Total Donors_count	Total Churn_prob
Center	3219	-9%	19991	-8%	20599	-32%	40	-35%	**43849**	**-19%**
Islands	216	-26%	870	19%	983	-15%	3	100%	**2072**	**-2%**
North	1143	-16%	6622	-10%	8112	-33%	14	-10%	**15891**	**-22%**
South	157	12%	939	1%	986	-22%	1	1%	**2083**	**-9%**
unknown	362	-46%	1417	-45%	1431	-52%	15	-100%	**3225**	**-48%**
Total	**5097**	**-13%**	**29839**	**-9%**	**32111**	**-32%**	**73**	**-37%**	**67120**	**-20%**



gender / area	company total_value	company avg_value	company Churn_prob	Feminino total_value	Feminino avg_value	Feminino Churn_prob	Masculino total_value	Masculino avg_value	Masculino Churn_prob	other total_value	other avg_value	other Churn_prob	Total total_value	Total avg_value	Total Churn_prol
Center	4,917,919.39	1,527.78	-9%	2,894,690.19	144.80	-8%	2,464,297.80	119.63	-32%	1,624.00	40.60	-35%	**10,278,531.38**	**234.41**	**-19**
Islands	326,552.05	1,511.82	-26%	114,886.80	132.05	19%	135,117.88	137.45	-15%	163.00	54.33	100%	**576,719.73**	**278.34**	**-2**
North	1,428,154.88	1,249.48	-16%	967,848.15	146.16	-10%	895,192.96	110.35	-33%	1,335.00	95.36	-10%	**3,292,530.99**	**207.19**	**-22**
South	99,708.44	635.09	12%	131,312.74	139.84	1%	77,517.26	78.62	-22%	29.00	29.00	1%	**308,567.44**	**148.14**	**-9**
unknown	151,000.64	417.13	-46%	69,394.73	48.97	-45%	78,091.18	54.57	-52%	0.00	0.00	-100%	**298,486.55**	**92.55**	**-48**
Total	**6,923,335.40**	**1,358.32**	**-13%**	**4,178,132.61**	**140.02**	**-9%**	**3,650,217.08**	**113.67**	**-32%**	**3,151.00**	**43.16**	**-37%**	**14,754,836.09**	**219.83**	**-20**

VI. Conclusion

Requirements engineering is a pivotal concept when starting a project or designing a system. From an overall view of system, requirements engineering is indispensable, it manages the whole procedure of any projects relating to creation and maintenance of a new system. Early intervention is in its nature, which helps avoid problems regarding undocumented requests and constant change of requirements. And all activities taken place in this process are advancing the successes and preventing any possible failures.

Data visualization is the end of the lifecycle of a business analytics project, since it will be or be inside the final report directly facing to end users. visualization requires the logic of storytelling, from macroscopic issues, in-depth and detailed to all aspects of the problem. Data visualization is very different from other information visualization. Data visualization is more instrumental and requires more logical levels. The main point of its production lies in the chart, but it is outside the chart. In many cases, either draw simple diagrams or visualize them. The real core of data visualization is to describe status and solve problems through chart tools.

The implementation of data visualization also needs to be contemplated. Since the product delivered shall be able to produce updated result, new data is needed to be refreshed. Therefore, the procedure of ETL should be as simple and intuitive as possible, so that the product is truly valuable for end users.

References

AMI. (2021). *Vision, Mission and Values.* https://ami.org.pt/en/aami/vision-mission-and-values/.

Andreoni, J., & Payne, A. (2013). Charitable Giving. *Handbook of Public Economics*, 1-50.

Beal, A. (2012). *Requirement is what we have to do to achieve an objective.*

Berenbach, B., Paulish, D., Katzmeier, J., & Rudorfer, A. (2009). *Software & Systems Requirements Engineering: In Practice.* New York: McGraw-Hill Professional.

Berry, M., Azlinah, M., & Bee, Y. (2020). Supervised and Unsupervised Learning for Data Science. *Springer*.

Buchmann, Robert, A., Andrea, P., Johansson, B., & Dimitris, K. (Perspectives in Business Informatics Research). *2020.* 19th International Conference on Business Informatics Research.

Cagala, T., Ulrich, G., Johannes, R., Anthony, & Strittmatter. (2021). *Optimal Targeting in Fundraising: A Machine Learning Approach.*

Chen, Y., Nicole, I., Brendan, L., & Vasilis, S. (2018). Optimal Data Acquisition for Statistical Estimation. *Proceedings of the 2018 ACM Conference on Economics and Computation*, (pp. 27-44). Ithaca.

Demissie, K., Hanley, J. A., Dick, M., Lawrence, J., & Pierre, E. (2000). Agreement in Measuring Socio-economic Status: Area-based Versus Individual Measures. *Chronic Dis. Can. 21(1)*, 1-7.

Dietrich, T., Sharyn, R.-T., & Kubacki, K. (2017). *Segmentation in Social Marketing : Process, Methods and Application.* Singapore: Springer.

Ding, S., Hong, H., Tao, Z., & Xiaoming, F. (2019). Estimating Socioeconomic Status via Temporal-Spatial Mobility Analysis - A Case Study of Smart Card Data. *2019 28th International Conference on Computer Communication and Networks (ICCCN)*, 1-9.

Dolnicar, S., & Lazarevski, K. (2009). Marketing in non-profit organi□ations: an international perspective. *International Marketing Review*, 26 (3), 275-291.

Dy, J. G., & Brodley, C. E. (2003). *Feature Selection for Unsupervised Learning.* https://ece.northeastern.edu/fac-ece/jdy/papers/jmlr_tr.pdf.

Goldsmith, R. F. (2004). *Discovering Real Business Requirements for Software Project Success.* Artech House.

Herzog, T. N., Scheuren, F. J., & Winkler, W. E. (2007). *Data Quality and Record Linkage Techniques.* New York: Springer.

Ian, F., Schellinck, D. A., & Kendall, K. W. (1983). ASSESSING THE RELIABILITY OF PSYCHOGRAPHIC ANALYSES. *Marketing Science 2(1)*, 57-73.

Imbalanced-Learning. (n.d.). *2. Over-sampling.* https://imbalancedlearn.org/stable/over_sampling.html.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science 349 (6245)*, 255-260.

Kotonya, G., & Sommerville, I. (1998). *Requirements Engineering: Processes and Techniques.* John Wiley & Sons.

Li, Y., Xiaohui, Y., & Koudas, N. (2021). *Data Acquisition for Improving Machine Learning Models.* PVLDB, 14(10): 1832-1844.

Méndez Fernández, D., & Wagner, S. (2015). *Naming the pain in requirements engineering: A design for a global family of surveys and first results from Germany.* Information and Software Technology.

Munari, B. (2006). *Design and visual communication.* Chronicle Books.

Nussbaumer Knaflic, C. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professionals.*

Panda, S., Vaibhav, M., Balamurali, R., & Ahmed, E. (2021). *Artificial Intelligence and Machine Learning in Business Management: Concepts, Challenges and Case Studies.* Boca Raton: CRC Press.

Scikit-learn. (n.d.). *2.3. Clustering.* https://scikitlearn.org/stable/modules/clustering.html.

Scikit-learn. (n.d.). *sklearn.metrics.silhouette_score.* https://scikitlearn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html.

Scikit-learn. (n.d.). *sklearn.model_selection.GridSearchCV.* https://scikitlearn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html.

Shah, A., Sendhil, M., & Eldar, S. (2012). Some Consequences of Having Too Little. *Science 338(6107)*, 682-685.

Sheehy-Skeffington, J. (2020). The effects of low socioeconomic status on decisionmaking processes. *Current Opinion in Psychology 33*, 183-188.

Sommerville, I. (2009). *Software Engineering (9th ed.).* Addison-Wesley.

Turing, A. M. (1950). *I.-COMPUTING MACHINERY AND INTELLIGENCE.* Mind IX (236): 433-460.

Walliman, N. (2017). *Social Research Methods.* London: Routledge.

Appendix

1. REGISTRATION SOURCE GROUPS

Registration Source Group Name	*Original values included in the group*
OTHERS	'OUTROS'
CORPORATE PARTNERSHIPS	'Giro 2.1'
	'AMIARTE'
	'NBup'
	'AÇÃO FRESENIUS'
	'Pontos Meos Voluntariado'
	'PARCERIA'
	'BARCLAYS/CABAZ NATAL12'
ENVIRONMENT	'AMBIENTE'
	'No planet B'
TRAINING	'Formação aos voluntários'
	'FORMACAO'
	'REUNIÃO NÚCLEO'
	'SOCORRISMO'
SOCIAL CAMPAIGNS	'PRENDA SOLIDÁRIA/09'
	'AÇÃO SOCIAL'
	'Fundo Universitário'
	'Iniciativa ODS'
	'Taleigo Amigo'
	'Projeto'
	'Voluntariado Nacional'
	'PAPORTO'
	'PREMIO JORNALISMO'
	'AVENTURA SOLIDÁRIA'

	'Giving Tuesday'
	'MARVILARTE'
INTERNATIONAL CAMPAIGNS:	'Campanha FNAC Moçambique'
	'CAMPANHA ZIMBABUÉ'
	'CAMPANHA TERCEIROS'
	'MAILLING SRI LANKA'
	'Voluntariado Internacional'
EMERGENCY CAMPAIGNS	'CAMPANHA EMERGÊNCIA HAITI'
	'CAMPANHA EMERGÊNCIA MADEIRA'
	'Emergência Moçambique 2019'
	'Emergência Covid-19'
	'CAMPANHA SOS LÍBANO'
	'MAILING EMERGÊNCIA LÍBANO'
	'EMERGENCIA FILIPINAS'
	'MAILLING EMERGÊNCIA/09'
CHRISTMAS CAMPAIGNS	'MAILING NATAL 07'
	'MAILING NATAL 10'
	'MAILING NATAL 06'
	'MAILLING NATAL 05'
	'M. NATAL'
DONORS	'DOADORES 04'
	'DOADORES 09'
	'DOADORES 08'
	'DOADORES 02'
	'DOADORES 10'
	'DOADORES 97'
	'DOADORES 07'
	'DOADORES 03'
	'DOADORES 05'

	'DOADORES/NATAL/08'
	'M. DOADORES'
	'DOADORES 98'
AUCHAN CAMPAIGN	'CAMPANHA AUCHAN 06'
	'AUCHAN'
	'AUCHAN 10'
	'AUCHAN 09'
VOLUNTEER REGISTRATION	'ESPONTÂNEO'
AMI RESOURCES/DEPARTMENTS	'RECURSOS HUMANOS'
	'PUBLICIDADE'
	'REVISTA'
	'SITE AMI',
	'LOJA AMI'
	'FICHA AMIGO'
	'Noticias AMI'
	'Recrutamento'
	'Campanha face to face'
	'COLABORADORES'
	'ADMINISTRAÇÃO',
	'PLATAFORMA'
	'Órgão de Comunicação Social'
	'FICHA DE INSCRIÇÃO AA'
	'Departamento Internacional'
	'PRENDA AMIGA'
	'DEL NORTE'
	'DEL FUNCHAL'
	'DEL CENTRO'
	'Delegação da Terceira'
TMN Campaigns:	'PONTOS TMN'
	'PONTOS TMN MADEIRA'

'Pontos TMN MEO Pedrogão'
'PONTOS TMN MADEIRA'
'Pontos TMN Moçambique'

2. DBSCAN ALGORITHM

The Density-Based Spatial Clustering of Applications with Noise is an alternative algorithm to the K-means. It clusters data into high-density form free shapes surrounded by lower density areas. The procedure is usually based on Euclidean distance and a radius, ε. For an example x_i, surrounded by n_{min} points, it becomes a core point:

$$N(d(\bar{x}_i, \bar{x}_j) \leqslant \epsilon) \geqslant n_{min}$$

A certain point is said to be close to another if the Euclidean distance between the two points in minor than the radius, ε. All samples belong to a cluster. After accounting for all directly reachable points the remaining ones are considered *noise*. In the end we can visualize conglomerates of points surrounded by some noisy samples.

3. K-MEANS

K-Means is an unsupervised learning algorithm proposed by Thomas Cover and used for the first time by James MacQueen in 1967. James investigated a method for partitioning a basic sample into k sets in order to achieve an efficient within-class variance (MacQueen, 1967). Nowadays, the goal of this algorithm is to minimize the average squared Euclidean distance of objects to their cluster centers (Manning et al.), to acquire qualitative insights on data by dividing the information according to the similarities of each data point.

The less variation exists between clusters, the more similar data points are with the same cluster (Imad Dabbura, 2018).

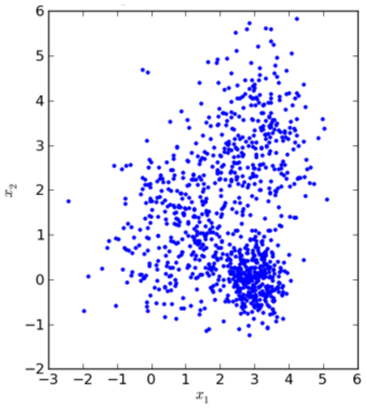


Figure 5 Example of unclustered data

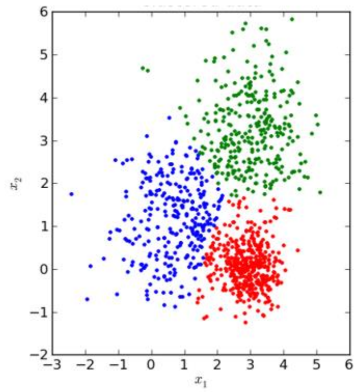


Figure 6 Clustered data after k-means algorithm

Above we can see a visual representation of the K-means algorithm and its purpose. The number of clusters is chosen prior to the implementation of the algorithm. It can be determined by different techniques such as the elbow method (Kodinariya & Makwana, 2013) and silhouette analysis. To conclude, K-means is one of the most used clustering algorithms due to how easily it can be implemented. It is a computationally faster method that produces tight clusters and find more sub-cluster if data larger cluster number is pre specified.

4. PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal Component Analysis (PCA) can be considered as an unsupervised learning algorithm, commonly used for dimensionality reduction to help in important tasks, including, "visualization, compressing the data, and finding a representation that is more informative for further processing (Müller & Guido, 2016). PCA is an attribute aggregation technique, initially proposed in 1901 by Karl Pearson, that consists of projecting data to a new dataset with a smaller number of features and, although it is considered simple, it requires some strong mathematical assumptions (Mendes Moreira, C. P. L. F. de Carvalho, & Horváth, 2018).

A data projection consists of transforming attributes in one space (sources) to another (signals), while trying to keep as much as possible the information from the original set of attributes and eliminating its redundancy. PCA achieves this elimination of noise by combining the original features into new ones (the principal components) that allow for a reduction of the covariance, using matrix operations from linear algebra. High covariance of variables mean that they will influence the model in the same way, therefore should not be chosen together. In PCA, "each principal component is a linear combination of the original attributes" (Mendes Moreira, C. P. L. F. de Carvalho, & Horváth, 2018). The linear algebra used by PCA is provided by Singular Value Decomposition (SVD). Firstly, the mean of each dimension is calculated for the whole dataset (d dimensionns), used to compute the covariance matrix afterwards. Then, *eigenvectors* and corresponding *eigenvalues* are calculated and sorted, from highest to lowest. The predefined number (k) of largest eigenvalues are chosen to form new a dimensional matrix dimensional (d x k), used to transform the samples onto the new subspace (Dubey, 2018).

An Eigenvalue increases the Eigenvectors along their span by the value of the first, when transformed linearly (Adewumi, 2019). An eigenvalue (λ) of a square matrix (A) will do such that any transformation on it with an eigen vector x will equal the scalar multiplication of λ with A ($Ax = \lambda x \; where \; x \neq 0$) (Dadhich, 2018).

In sum, when using the PCA method, the components are ranked from the highest variance value to the lowest and selected by prioritizing the largest variances. The number of components chosen is based on a predefinition of that value, or when the increase in variance of the data when adding the component becomes irrelevant. By using this process, there is the guarantee that any feature that is not contributing significantly to the results is ignored and the model can benefit from the advantages of a smaller number of dimensions.

5. METRICS – T-TEST AND ANOVA RESULTS

ANOVA Results

	p-value
gender_Feminino	0
gender_Masculino	0
gender_company	0
gender_other	0,00011
AMI_friend	0
consent_boolean	0
total_value	0
registration_source_group_AMI RESOURCES/DEPARTMENTS	0
registration_source_group_AUCHAN CAMPAIGN	0
registration_source_group_CHRISTMAS CAMPAIGNS	0
registration_source_group_CORPORATE PARTNERSHIPS	0
registration_source_group_DONORS	0
registration_source_group_EMERGENCY CAMPAIGNS	0
registration_source_group_ENVIRONMENT	0
registration_source_group_INTERNATIONAL CAMPAIGNS	0

registration_source_group_NATIONAL CAMPAIGNS	0
registration_source_group_OTHERS	0
registration_source_group_SPONTANEOUS REGISTRATION	0
registration_source_group_TMN CAMPAIGNS	0
registration_source_group_TRAINING	0
1st_value	0
avg_value_cont	0
avg_days_between_contrib	0
days_since_donor_last_contribution	0
everismonetary	0
total_contributions	0
region_Alentejo	0
region_Algarve	0,00577
region_Açores	0,00135
region_Beira Interior	0
region_Beira Litoral	0
region_Entre-Douro-e-Minho	0
region_Lisboa	0
region_Madeira	0
region_NOT FOUND	0
region_Trás-os-Montes e Alto Douro	0
region_Vale do Tejo	0
individual/collective	0
field_description_ACTIVIDADES ADMINISTRATIVAS E DOS SERVIÇOS DE APOIO	0
field_description_ACTIVIDADES ARTÍSTICAS, DE ESPECTÁCULOS, DESPORTIVAS E RECREATIVAS	0
field_description_ACTIVIDADES DE CONSULTORIA, CIENTÍFICAS, TÉCNICAS E SIMILARES	0
field_description_ACTIVIDADES DE INFORMAÇÃO E DE COMUNICAÇÃO	0
field_description_ACTIVIDADES DE SAÚDE HUMANA E APOIO SOCIAL	0

field_description_ACTIVIDADES FINANCEIRAS E DE SEGUROS	0
field_description_ACTIVIDADES IMOBILIÁRIAS	0
field_description_ADMINISTRAÇÃO PÚBLICA E DEFESA; SEGURANÇA SOCIAL OBRIGATÓRIA	0
field_description_AGRICULTURA, PRODUÇÃO ANIMAL, CAÇA, FLORESTA E PESCA	0
field_description_ALOJAMENTO, RESTAURAÇÃO E SIMILARES	0
field_description_CAPTAÇÃO, TRATAMENTO E DISTRIBUIÇÃO DE ÁGUA; SANEAMENTO,GESTÃO DE RESÍDUOS E DESPOLUIÇÃO	0
field_description_COMÉRCIO E VEÍCULOS	0
field_description_CONSTRUÇÃO	0
field_description_EDUCAÇÃO	0
field_description_ELECTRICIDADE, GÁS, VAPOR, ÁGUA QUENTE E FRIA E AR FRIO	0
field_description_INDÚSTRIAS EXTRACTIVAS	0
field_description_INDÚSTRIAS TRANSFORMADORAS	0
field_description_OUTRAS ACTIVIDADES DE SERVIÇOS	0
field_description_TRANSPORTES E ARMAZENAGEM	0
field_description_individuals	0
field_description_not found	0
Legal Form_Agrupamento Complementar de Empresas	0,00683
Legal Form_Associação	0
Legal Form_Autarquia	0
Legal Form_Cooperativa	0
Legal Form_Empresário em Nome Individual	0,00683
Legal Form_Entidade Estrangeira	0
Legal Form_Fundação	0
Legal Form_Organismo de Administração Pública	0
Legal Form_Sociedade Anónima	0
Legal Form_Sociedade Unipessoal por Quotas	0
Legal Form_Sociedade por Quotas	0

Legal Form_individuals	0
Legal Form_unknown	0

T-TEST: Cluster 0 VS Cluster 1	*statistic*	*pvalue*
gender_Feminino	-53,1744	0
gender_Masculino	-95,6949	0
gender_company	1,43E+18	0
gender_other	-1,94184	0,05216
AMI_friend	28,35666	0
consent_boolean	-0,99153	0,32143
total_value	40,5415	0
registration_source_group_AMI RESOURCES/DEPARTMENTS	39,29291	0
registration_source_group_AUCHAN CAMPAIGN	6,25992	0
registration_source_group_CHRISTMAS CAMPAIGNS	21,14016	0
registration_source_group_CORPORATE PARTNERSHIPS	33,78672	0
registration_source_group_DONORS	20,46315	0
registration_source_group_EMERGENCY CAMPAIGNS	24,8972	0
registration_source_group_ENVIRONMENT	31,77094	0
registration_source_group_INTERNATIONAL CAMPAIGNS	12,6955	0
registration_source_group_NATIONAL CAMPAIGNS	2,39966	0,01642
registration_source_group_OTHERS	-98,5558	0
registration_source_group_SPONTANEOUS REGISTRATION	46,18694	0
registration_source_group_TMN CAMPAIGNS	31,88278	0
registration_source_group_TRAINING	12,89019	0
1st_value	161,2823	0
avg_value_cont	-2,34313	0,01913
avg_days_between_contrib	-0,89154	0,37265
days_since_donor_last_contribution	159,5334	0
everismonetary	156,5473	0

total_contributions	52,39363	0
region_Alentejo	1,0014	0,31664
region_Algarve	0,51775	0,60464
region_Açores	3,80495	0,00014
region_Beira Interior	4,13714	0,00004
region_Beira Litoral	10,10547	0
region_Entre-Douro-e-Minho	-3,98539	0,00007
region_Lisboa	-5,77327	0
region_Madeira	7,37926	0
region_NOT FOUND	5,31359	0
region_Trás-os-Montes e Alto Douro	3,5627	0,00037
region_Vale do Tejo	-5,33711	0
individual/collective	-1,5E+18	0
field_description_ACTIVIDADES ADMINISTRATIVAS E DOS SERVIÇOS DE APOIO	29,42304	0
field_description_ACTIVIDADES ARTÍSTICAS, DE ESPECTÁCULOS, DESPORTIVAS E RECREATIVAS	18,4215	0
field_description_ACTIVIDADES DE CONSULTORIA, CIENTÍFICAS, TÉCNICAS E SIMILARES	52,51137	0
field_description_ACTIVIDADES DE INFORMAÇÃO E DE COMUNICAÇÃO	34,89604	0
field_description_ACTIVIDADES DE SAÚDE HUMANA E APOIO SOCIAL	33,74669	0
field_description_ACTIVIDADES FINANCEIRAS E DE SEGUROS	19,73143	0
field_description_ACTIVIDADES IMOBILIÁRIAS	25,82775	0
field_description_ADMINISTRAÇÃO PÚBLICA E DEFESA; SEGURANÇA SOCIAL OBRIGATÓRIA	7,4181	0
field_description_AGRICULTURA, PRODUÇÃO ANIMAL, CAÇA, FLORESTA E PESCA	12,62946	0

field_description_ALOJAMENTO, RESTAURAÇÃO E SIMILARES	44,26436	0
field_description_CAPTAÇÃO, TRATAMENTO E DISTRIBUIÇÃO DE ÁGUA; SANEAMENTO,GESTÃO DE RESÍDUOS E DESPOLUIÇÃO	14,01943	0
field_description_COMÉRCIO E VEÍCULOS	85,40046	0
field_description_CONSTRUÇÃO	32,86214	0
field_description_EDUCAÇÃO	24,43724	0
field_description_ELECTRICIDADE, GÁS, VAPOR, ÁGUA QUENTE E FRIA E AR FRIO	6,99316	0
field_description_INDÚSTRIAS EXTRACTIVAS	7,4181	0
field_description_INDÚSTRIAS TRANSFORMADORAS	57,06631	0
field_description_OUTRAS ACTIVIDADES DE SERVIÇOS	19,08719	0
field_description_TRANSPORTES E ARMAZENAGEM	21,9879	0
field_description_individuals	-1,5E+18	0
field_description_not found	125,1038	0
Legal Form_Agrupamento Complementar de Empresas	2,47076	0,01349
Legal Form_Associação	17,55551	0
Legal Form_Autarquia	6,05507	0
Legal Form_Cooperativa	12,1316	0
Legal Form_Empresário em Nome Individual	2,47076	0,01349
Legal Form_Entidade Estrangeira	9,25655	0
Legal Form_Fundação	11,61284	0
Legal Form_Organismo de Administração Pública	6,05507	0
Legal Form_Sociedade Anónima	85,18258	0
Legal Form_Sociedade Unipessoal por Quotas	50,84022	0
Legal Form_Sociedade por Quotas	135,1748	0
Legal Form_individuals	-1,5E+18	0
Legal Form_unknown	126,5396	0

T-TEST: Cluster 0 VS Cluster 2	*statistic*	*pvalue*
gender_Feminino	-110,669	0
gender_Masculino	-45,9474	0
gender_company	6,83E+17	0
gender_other	-2,15633	0,03108
AMI_friend	-76,3643	0
consent_boolean	-5,78894	0
total_value	7,11267	0
registration_source_group_AMI RESOURCES/DEPARTMENTS	-69,7666	0
registration_source_group_AUCHAN CAMPAIGN	-27,9922	0
registration_source_group_CHRISTMAS CAMPAIGNS	-12,5716	0
registration_source_group_CORPORATE PARTNERSHIPS	16,63064	0
registration_source_group_DONORS	-7,67463	0
registration_source_group_EMERGENCY CAMPAIGNS	0,11975	0,90468
registration_source_group_ENVIRONMENT	18,86661	0
registration_source_group_INTERNATIONAL CAMPAIGNS	8,85951	0
registration_source_group_NATIONAL CAMPAIGNS	-1,47759	0,13954
registration_source_group_OTHERS	78,33962	0
registration_source_group_SPONTANEOUS REGISTRATION	15,72508	0
registration_source_group_TMN CAMPAIGNS	3,82486	0,00013
registration_source_group_TRAINING	6,55274	0
1st_value	72,15897	0
avg_value_cont	-20,4625	0
avg_days_between_contrib	-50,4798	0
days_since_donor_last_contribution	-20,8749	0
everismonetary	-85,6798	0
total_contributions	-31,7696	0
region_Alentejo	6,50341	0
region_Algarve	1,05602	0,29098

region_Açores	2,78634	0,00534
region_Beira Interior	5,52118	0
region_Beira Litoral	-0,31446	0,75318
region_Entre-Douro-e-Minho	-6,09929	0
region_Lisboa	2,71943	0,00655
region_Madeira	3,91225	0,00009
region_NOT FOUND	17,30239	0
region_Trás-os-Montes e Alto Douro	1,73753	0,08232
region_Vale do Tejo	-10,4186	0
individual/collective	-6,9E+17	0
field_description_ACTIVIDADES ADMINISTRATIVAS E DOS SERVIÇOS DE APOIO	13,53269	0
field_description_ACTIVIDADES ARTÍSTICAS, DE ESPECTÁCULOS, DESPORTIVAS E RECREATIVAS	8,4727	0
field_description_ACTIVIDADES DE CONSULTORIA, CIENTÍFICAS, TÉCNICAS E SIMILARES	24,15183	0
field_description_ACTIVIDADES DE INFORMAÇÃO E DE COMUNICAÇÃO	16,04992	0
field_description_ACTIVIDADES DE SAÚDE HUMANA E APOIO SOCIAL	15,52129	0
field_description_ACTIVIDADES FINANCEIRAS E DE SEGUROS	9,07518	0
field_description_ACTIVIDADES IMOBILIÁRIAS	11,8791	0
field_description_ADMINISTRAÇÃO PÚBLICA E DEFESA; SEGURANÇA SOCIAL OBRIGATÓRIA	3,41185	0,00065
field_description_AGRICULTURA, PRODUÇÃO ANIMAL, CAÇA, FLORESTA E PESCA	5,80874	0
field_description_ALOJAMENTO, RESTAURAÇÃO E SIMILARES	20,35874	0
field_description_CAPTAÇÃO, TRATAMENTO E DISTRIBUIÇÃO DE ÁGUA; SANEAMENTO,GESTÃO DE RESÍDUOS E DESPOLUIÇÃO	6,44803	0
field_description_COMÉRCIO E VEÍCULOS	39,27869	0
field_description_CONSTRUÇÃO	15,11446	0
field_description_EDUCAÇÃO	11,23955	0

field_description_ELECTRICIDADE, GÁS, VAPOR, ÁGUA QUENTE E FRIA E AR FRIO	3,2164	0,0013
field_description_INDÚSTRIAS EXTRACTIVAS	3,41185	0,00065
field_description_INDÚSTRIAS TRANSFORMADORAS	26,24681	0
field_description_OUTRAS ACTIVIDADES DE SERVIÇOS	8,77887	0
field_description_TRANSPORTES E ARMAZENAGEM	10,11301	0
field_description_individuals	-6,9E+17	0
field_description_not found	57,53967	0
Legal Form_Agrupamento Complementar de Empresas	1,13639	0,25582
Legal Form_Associação	8,0744	0
Legal Form_Autarquia	2,78494	0,00536
Legal Form_Cooperativa	5,57975	0
Legal Form_Empresário em Nome Individual	1,13639	0,25582
Legal Form_Entidade Estrangeira	4,25741	0,00002
Legal Form_Fundação	5,34115	0
Legal Form_Organismo de Administração Pública	2,78494	0,00536
Legal Form_Sociedade Anónima	39,17848	0
Legal Form_Sociedade Unipessoal por Quotas	23,38321	0
Legal Form_Sociedade por Quotas	62,17166	0
Legal Form_individuals	-6,9E+17	0
Legal Form_unknown	58,20002	0

T-TEST: Cluster 0 VS Cluster 3	*statistic*	*pvalue*
gender_Feminino	-83,7429	0
gender_Masculino	-60,6352	0
gender_company	1,28E+18	0
gender_other	-3,03872	0,00238
AMI_friend	-7,99039	0
consent_boolean	-3,60846	0,00031

total_value	32,89276	0
registration_source_group_AMI RESOURCES/DEPARTMENTS	-10,2729	0
registration_source_group_AUCHAN CAMPAIGN	-12,234	0
registration_source_group_CHRISTMAS CAMPAIGNS	0,70229	0,4825
registration_source_group_CORPORATE PARTNERSHIPS	14,5002	0
registration_source_group_DONORS	5,97419	0
registration_source_group_EMERGENCY CAMPAIGNS	-2,70583	0,00682
registration_source_group_ENVIRONMENT	33,2853	0
registration_source_group_INTERNATIONAL CAMPAIGNS	1,96111	0,04988
registration_source_group_NATIONAL CAMPAIGNS	-5,30748	0
registration_source_group_OTHERS	48,27228	0
registration_source_group_SPONTANEOUS REGISTRATION	1,49289	0,13548
registration_source_group_TMN CAMPAIGNS	-46,0711	0
registration_source_group_TRAINING	2,06296	0,03913
1st_value	141,3907	0
avg_value_cont	-25,0227	0
avg_days_between_contrib	-29,1678	0
days_since_donor_last_contribution	-9,90513	0
everismonetary	-47,126	0
total_contributions	10,64753	0
region_Alentejo	-1,18746	0,23506
region_Algarve	-1,1656	0,24378
region_Açores	2,53397	0,01128
region_Beira Interior	-0,33972	0,73407
region_Beira Litoral	0,82436	0,40974
region_Entre-Douro-e-Minho	0,37469	0,70789
region_Lisboa	4,47068	0,00001
region_Madeira	-0,72238	0,47007
region_NOT FOUND	6,94774	0
region_Trás-os-Montes e Alto Douro	0,13137	0,89549

region_Vale do Tejo	-7,33717	0
individual/collective	-1,3E+18	0
field_description_ACTIVIDADES ADMINISTRATIVAS E DOS SERVIÇOS DE APOIO	26,01949	0
field_description_ACTIVIDADES ARTÍSTICAS, DE ESPECTÁCULOS, DESPORTIVAS E RECREATIVAS	16,29057	0
field_description_ACTIVIDADES DE CONSULTORIA, CIENTÍFICAS, TÉCNICAS E SIMILARES	46,43705	0
field_description_ACTIVIDADES DE INFORMAÇÃO E DE COMUNICAÇÃO	30,85939	0
field_description_ACTIVIDADES DE SAÚDE HUMANA E APOIO SOCIAL	29,843	0
field_description_ACTIVIDADES FINANCEIRAS E DE SEGUROS	17,44898	0
field_description_ACTIVIDADES IMOBILIÁRIAS	22,8401	0
field_description_ADMINISTRAÇÃO PÚBLICA E DEFESA; SEGURANÇA SOCIAL OBRIGATÓRIA	6,56	0
field_description_AGRICULTURA, PRODUÇÃO ANIMAL, CAÇA, FLORESTA E PESCA	11,16853	0
field_description_ALOJAMENTO, RESTAURAÇÃO E SIMILARES	39,14402	0
field_description_CAPTAÇÃO, TRATAMENTO E DISTRIBUIÇÃO DE ÁGUA; SANEAMENTO,GESTÃO DE RESÍDUOS E DESPOLUIÇÃO	12,39771	0
field_description_COMÉRCIO E VEÍCULOS	75,52166	0
field_description_CONSTRUÇÃO	29,06077	0
field_description_EDUCAÇÃO	21,61043	0
field_description_ELECTRICIDADE, GÁS, VAPOR, ÁGUA QUENTE E FRIA E AR FRIO	6,18422	0
field_description_INDÚSTRIAS EXTRACTIVAS	6,56	0
field_description_INDÚSTRIAS TRANSFORMADORAS	50,46509	0
field_description_OUTRAS ACTIVIDADES DE SERVIÇOS	16,87925	0
field_description_TRANSPORTES E ARMAZENAGEM	19,44442	0
field_description_individuals	-1,3E+18	0
field_description_not found	110,6323	0
Legal Form_Agrupamento Complementar de Empresas	2,18495	0,0289

Legal Form_Associação	15,52475	0
Legal Form_Autarquia	5,35464	0
Legal Form_Cooperativa	10,72826	0
Legal Form_Empresário em Nome Individual	2,18495	0,0289
Legal Form_Entidade Estrangeira	8,18578	0
Legal Form_Fundação	10,26951	0
Legal Form_Organismo de Administração Pública	5,35464	0
Legal Form_Sociedade Anónima	75,32898	0
Legal Form_Sociedade Unipessoal por Quotas	44,95921	0
Legal Form_Sociedade por Quotas	119,5383	0
Legal Form_individuals	-1,3E+18	0
Legal Form_unknown	111,902	0

T-TEST: Cluster 1 VS Cluster 2	*statistic*	*pvalue*
gender_Feminino	-54,2038	0
gender_Masculino	54,21749	0
gender_company	0	1
gender_other	-0,45798	0,64697
AMI_friend	-205,366	0
consent_boolean	-12,998	0
total_value	-79,7395	0
registration_source_group_AMI RESOURCES/DEPARTMENTS	-201,024	0
registration_source_group_AUCHAN CAMPAIGN	-70,6774	0
registration_source_group_CHRISTMAS CAMPAIGNS	-44,7362	0
registration_source_group_CORPORATE PARTNERSHIPS	0,47722	0,63321
registration_source_group_DONORS	-34,2072	0
registration_source_group_EMERGENCY CAMPAIGNS	-25,1784	0
registration_source_group_ENVIRONMENT	4,69533	0
registration_source_group_INTERNATIONAL CAMPAIGNS	2,05357	0,04002
registration_source_group_NATIONAL CAMPAIGNS	-5,03984	0

registration_source_group_OTHERS	307,8531	0
registration_source_group_SPONTANEOUS REGISTRATION	-21,1004	0
registration_source_group_TMN CAMPAIGNS	-25,8452	0
registration_source_group_TRAINING	-0,13572	0,89204
1st_value	-47,8683	0
avg_value_cont	-50,5081	0
avg_days_between_contrib	-124,72	0
days_since_donor_last_contribution	-234,332	0
everismonetary	-1359,17	0
total_contributions	-94,5557	0
region_Alentejo	7,05116	0
region_Algarve	0,86597	0,38651
region_Açores	-0,12674	0,89915
region_Beira Interior	3,04643	0,00232
region_Beira Litoral	-11,6579	0
region_Entre-Douro-e-Minho	-3,96126	0,00007
region_Lisboa	9,97505	0
region_Madeira	-1,94003	0,05238
region_NOT FOUND	14,88147	0
region_Trás-os-Montes e Alto Douro	-1,33655	0,18138
region_Vale do Tejo	-8,4793	0
individual/collective		
field_description_ACTIVIDADES ADMINISTRATIVAS E DOS SERVIÇOS DE APOIO	0	1
field_description_ACTIVIDADES ARTÍSTICAS, DE ESPECTÁCULOS, DESPORTIVAS E RECREATIVAS	-176,378	0
field_description_ACTIVIDADES DE CONSULTORIA, CIENTÍFICAS, TÉCNICAS E SIMILARES	-53,4602	0
field_description_ACTIVIDADES DE INFORMAÇÃO E DE COMUNICAÇÃO	176,3784	0
field_description_ACTIVIDADES DE SAÚDE HUMANA E APOIO SOCIAL	-25,8542	0
field_description_ACTIVIDADES FINANCEIRAS E DE SEGUROS	0	1

field_description_ACTIVIDADES IMOBILIÁRIAS	-176,378	0
field_description_ADMINISTRAÇÃO PÚBLICA E DEFESA; SEGURANÇA SOCIAL OBRIGATÓRIA	81,12737	0
field_description_AGRICULTURA, PRODUÇÃO ANIMAL, CAÇA, FLORESTA E PESCA	-81,1274	0
field_description_ALOJAMENTO, RESTAURAÇÃO E SIMILARES	-176,378	0
field_description_CAPTAÇÃO, TRATAMENTO E DISTRIBUIÇÃO DE ÁGUA; SANEAMENTO,GESTÃO DE RESÍDUOS E DESPOLUIÇÃO	-176,378	0
field_description_COMÉRCIO E VEÍCULOS	-25,8542	0
field_description_CONSTRUÇÃO	25,85424	0
field_description_EDUCAÇÃO		
field_description_ELECTRICIDADE, GÁS, VAPOR, ÁGUA QUENTE E FRIA E AR FRIO	-39,5317	0
field_description_INDÚSTRIAS EXTRACTIVAS	81,12737	0
field_description_INDÚSTRIAS TRANSFORMADORAS	-81,1274	0
field_description_OUTRAS ACTIVIDADES DE SERVIÇOS	0	1
field_description_TRANSPORTES E ARMAZENAGEM	0	1
field_description_individuals		
field_description_not found	0	1
Legal Form_Agrupamento Complementar de Empresas	-81,1274	0
Legal Form_Associação	0	1
Legal Form_Autarquia	0	1
Legal Form_Cooperativa	59,70641	0
Legal Form_Empresário em Nome Individual	19,17306	0
Legal Form_Entidade Estrangeira	39,53172	0
Legal Form_Fundação		
Legal Form_Organismo de Administração Pública	0	1
Legal Form_Sociedade Anónima	0	1
Legal Form_Sociedade Unipessoal por Quotas		
Legal Form_Sociedade por Quotas	0	1

Legal Form_individuals		
Legal Form_unknown	59,70641	0

T-TEST: Cluster 1 VS Cluster 3	*statistic*	*pvalue*
gender_Feminino	-53,4961	0
gender_Masculino	53,75435	0
gender_company	0	1
gender_other	-3,59622	0,00032
AMI_friend	-41,516	0
consent_boolean	-7,86767	0
total_value	-53,4051	0
registration_source_group_AMI RESOURCES/DEPARTMENTS	-55,7926	0
registration_source_group_AUCHAN CAMPAIGN	-33,3952	0
registration_source_group_CHRISTMAS CAMPAIGNS	-20,5055	0
registration_source_group_CORPORATE PARTNERSHIPS	-17,7399	0
registration_source_group_DONORS	-13,3442	0
registration_source_group_EMERGENCY CAMPAIGNS	-30,8926	0
registration_source_group_ENVIRONMENT	5,13447	0
registration_source_group_INTERNATIONAL CAMPAIGNS	-13,4238	0
registration_source_group_NATIONAL CAMPAIGNS	-14,7276	0
registration_source_group_OTHERS	272,8065	0
registration_source_group_SPONTANEOUS REGISTRATION	-60,7689	0
registration_source_group_TMN CAMPAIGNS	-133,007	0
registration_source_group_TRAINING	-11,4695	0
1st_value	-59,3529	0
avg_value_cont	-61,6362	0
avg_days_between_contrib	-72,0583	0
days_since_donor_last_contribution	-193,478	0
everismonetary	-317,546	0
total_contributions	-105,603	0

region_Alentejo	-3,92857	0,00009
region_Algarve	-3,04985	0,00229
region_Açores	-1,87888	0,06027
region_Beira Interior	-7,52527	0
region_Beira Litoral	-15,2425	0
region_Entre-Douro-e-Minho	7,79229	0
region_Lisboa	18,19211	0
region_Madeira	-12,8084	0
region_NOT FOUND	2,82785	0,00469
region_Trás-os-Montes e Alto Douro	-5,63474	0
region_Vale do Tejo	-3,88468	0,0001
individual/collective		
field_description_ACTIVIDADES ADMINISTRATIVAS E DOS SERVIÇOS DE APOIO	76,76889	0
field_description_ACTIVIDADES ARTÍSTICAS, DE ESPECTÁCULOS, DESPORTIVAS E RECREATIVAS		
field_description_ACTIVIDADES DE CONSULTORIA, CIENTÍFICAS, TÉCNICAS E SIMILARES	-44,7884	0
field_description_ACTIVIDADES DE INFORMAÇÃO E DE COMUNICAÇÃO		
field_description_ACTIVIDADES DE SAÚDE HUMANA E APOIO SOCIAL	0	1
field_description_ACTIVIDADES FINANCEIRAS E DE SEGUROS	76,76889	0
field_description_ACTIVIDADES IMOBILIÁRIAS		
field_description_ADMINISTRAÇÃO PÚBLICA E DEFESA; SEGURANÇA SOCIAL OBRIGATÓRIA	-76,7689	0
field_description_AGRICULTURA, PRODUÇÃO ANIMAL, CAÇA, FLORESTA E PESCA	-155,978	0
field_description_ALOJAMENTO, RESTAURAÇÃO E SIMILARES		
field_description_CAPTAÇÃO, TRATAMENTO E DISTRIBUIÇÃO DE ÁGUA; SANEAMENTO,GESTÃO DE RESÍDUOS E DESPOLUIÇÃO		
field_description_COMÉRCIO E VEÍCULOS	-44,7884	0
field_description_CONSTRUÇÃO	0	1

field_description_EDUCAÇÃO		
field_description_ELECTRICIDADE, GÁS, VAPOR, ÁGUA QUENTE E FRIA E AR FRIO	0	1
field_description_INDÚSTRIAS EXTRACTIVAS	-76,7689	0
field_description_INDÚSTRIAS TRANSFORMADORAS	-155,978	0
field_description_OUTRAS ACTIVIDADES DE SERVIÇOS	0	1
field_description_TRANSPORTES E ARMAZENAGEM	0	1
field_description_individuals		
field_description_not found	0	1
Legal Form_Agrupamento Complementar de Empresas	0	1
Legal Form_Associação	46,94736	0
Legal Form_Autarquia	0	1
Legal Form_Cooperativa	76,76889	0
Legal Form_Empresário em Nome Individual	0	1
Legal Form_Entidade Estrangeira	71,32745	0
Legal Form_Fundação		
Legal Form_Organismo de Administração Pública	0	1
Legal Form_Sociedade Anónima	0	1
Legal Form_Sociedade Unipessoal por Quotas		
Legal Form_Sociedade por Quotas	-76,7689	0
Legal Form_individuals		
Legal Form_unknown	-233,687	0

T-TEST: Cluster 2 VS Cluster 3	*statistic*	*pvalue*
gender_Feminino	18,8219	0
gender_Masculino	-18,7012	0
gender_company	0	1
gender_other	-1,60642	0,10819
AMI_friend	124,512	0
consent_boolean	4,9311	0

total_value	57,91092	0
registration_source_group_AMI RESOURCES/DEPARTMENTS	99,57734	0
registration_source_group_AUCHAN CAMPAIGN	32,84313	0
registration_source_group_CHRISTMAS CAMPAIGNS	22,60579	0
registration_source_group_CORPORATE PARTNERSHIPS	-8,66688	0
registration_source_group_DONORS	20,13625	0
registration_source_group_EMERGENCY CAMPAIGNS	-3,16249	0,00157
registration_source_group_ENVIRONMENT	-2,66593	0,00768
registration_source_group_INTERNATIONAL CAMPAIGNS	-7,92277	0
registration_source_group_NATIONAL CAMPAIGNS	-4,33393	0,00001
registration_source_group_OTHERS	-37,6302	0
registration_source_group_SPONTANEOUS REGISTRATION	-17,8279	0
registration_source_group_TMN CAMPAIGNS	-54,8472	0
registration_source_group_TRAINING	-5,42686	0
1st_value	25,40449	0
avg_value_cont	26,38235	0
avg_days_between_contrib	58,30931	0
days_since_donor_last_contribution	17,4458	0
everismonetary	42,36033	0
total_contributions	73,67835	0
region_Alentejo	-8,96019	0
region_Algarve	-2,64634	0,00814
region_Açores	-1,0116	0,31174
region_Beira Interior	-6,98471	0
region_Beira Litoral	1,34052	0,18009
region_Entre-Douro-e-Minho	8,8417	0
region_Lisboa	1,22822	0,21938
region_Madeira	-5,52297	0
region_NOT FOUND	-13,5495	0
region_Trás-os-Montes e Alto Douro	-2,09639	0,03606

region_Vale do Tejo	5,81238	0
individual/collective		
field_description_ACTIVIDADES ADMINISTRATIVAS E DOS SERVIÇOS DE APOIO	39,25762	0
field_description_ACTIVIDADES ARTÍSTICAS, DE ESPECTÁCULOS, DESPORTIVAS E RECREATIVAS	155,9757	0
field_description_ACTIVIDADES DE CONSULTORIA, CIENTÍFICAS, TÉCNICAS E SIMILARES	39,25762	0
field_description_ACTIVIDADES DE INFORMAÇÃO E DE COMUNICAÇÃO	-155,976	0
field_description_ACTIVIDADES DE SAÚDE HUMANA E APOIO SOCIAL	25,54988	0
field_description_ACTIVIDADES FINANCEIRAS E DE SEGUROS	39,25762	0
field_description_ACTIVIDADES IMOBILIÁRIAS	155,9757	0
field_description_ADMINISTRAÇÃO PÚBLICA E DEFESA; SEGURANÇA SOCIAL OBRIGATÓRIA	-81,1269	0
field_description_AGRICULTURA, PRODUÇÃO ANIMAL, CAÇA, FLORESTA E PESCA		
field_description_ALOJAMENTO, RESTAURAÇÃO E SIMILARES	155,9757	0
field_description_CAPTAÇÃO, TRATAMENTO E DISTRIBUIÇÃO DE ÁGUA; SANEAMENTO,GESTÃO DE RESÍDUOS E DESPOLUIÇÃO	155,9757	0
field_description_COMÉRCIO E VEÍCULOS	0	1
field_description_CONSTRUÇÃO	-25,5499	0
field_description_EDUCAÇÃO		
field_description_ELECTRICIDADE, GÁS, VAPOR, ÁGUA QUENTE E FRIA E AR FRIO	39,25762	0
field_description_INDÚSTRIAS EXTRACTIVAS	-81,1269	0
field_description_INDÚSTRIAS TRANSFORMADORAS		
field_description_OUTRAS ACTIVIDADES DE SERVIÇOS	0	1
field_description_TRANSPORTES E ARMAZENAGEM	0	1
field_description_individuals		
field_description_not found	0	1
Legal Form_Agrupamento Complementar de Empresas	81,1269	0

Legal Form_Associação	25,54988	0
Legal Form_Autarquia	0	1
Legal Form_Cooperativa	0	1
Legal Form_Empresário em Nome Individual	-18,895	0
Legal Form_Entidade Estrangeira	0	1
Legal Form_Fundação		
Legal Form_Organismo de Administração Pública	0	1
Legal Form_Sociedade Anónima	0	1
Legal Form_Sociedade Unipessoal por Quotas		
Legal Form_Sociedade por Quotas	-39,2576	0
Legal Form_individuals		
Legal Form_unknown	-168,668	0

6. SEGMENTATION MODEL RESULTS

Cluster 0

	Rel. Freq. of Donors in each Region
Vale do Tejo	0,34
Entre-Douro-e-Minho	0,25
Lisboa	0,18
Beira Litoral	0,12
Algarve	0,03
Alentejo	0,02
Beira Interior	0,02
Madeira	0,01
NOT FOUND	0,01
Açores	0,01
Trás-os-Montes e Alto Douro	0,01

	Rel. Freq. of Donors per Gender
Feminino	0,71
Masculino	0,29
other	0

	Rel Freq. of Donors Registered as AMI_friend
1	0,57
0	0,43

	Rel Freq. of Donors per Registration Source Group
AMI RESOURCES/DEPARTMENTS	0,56
AUCHAN CAMPAIGN	0,14
SPONTANEOUS REGISTRATION	0,11
CHRISTMAS CAMPAIGNS	0,06
TMN CAMPAIGNS	0,04
DONORS	0,04
EMERGENCY CAMPAIGNS	0,03
NATIONAL CAMPAIGNS	0,01
OTHERS	0,01
INTERNATIONAL CAMPAIGNS	0
CORPORATE PARTNERSHIPS	0
TRAINING	0
ENVIRONMENT	0

	Rel Freq. of Donors per Company Field Description
Individuals	1

	Rel Freq. of Donors per Legal From
individuals	1

Avg Total Value Donate Per Donor	Avg Total Number of Contributions Per Donor	Avg Number of Days Since Donor Last Contribution (Recency)	Median Number of Days Between Contribution (Frequency)
801,34	15,37	2505	127,67

Cluster 1

	Rel Freq. of Donors per Region
Vale do Tejo	0,25
Entre-Douro-e-Minho	0,21
Lisboa	0,19
Beira Litoral	0,12
NOT FOUND	0,07
Alentejo	0,04
Algarve	0,03
Beira Interior	0,03
Madeira	0,02
Açores	0,02
Trás-os-Montes e Alto Douro	0,02

	Rel Freq. of Donors per Gender
company	1

	Rel Freq. of Donors Registered as AMI_friend
0	0,97
1	0,03

	Rel Freq. of Donors per Registration Source Group
OTHERS	0,5
SPONTANEOUS REGISTRATION	0,21
TMN CAMPAIGNS	0,05
ENVIRONMENT	0,05

AMI RESOURCES/DEPARTMENTS	0,05
CORPORATE PARTNERSHIPS	0,04
EMERGENCY CAMPAIGNS	0,03
INTERNATIONAL CAMPAIGNS	0,02
CHRISTMAS CAMPAIGNS	0,01
DONORS	0,01
NATIONAL CAMPAIGNS	0,01
TRAINING	0,01
AUCHAN CAMPAIGN	0

	Rel Freq. of Donors per Company Field Description
not found	0,33
COMÉRCIO E VEÍCULOS	0,19
INDÚSTRIAS TRANSFORMADORAS	0,09
ACTIVIDADES DE CONSULTORIA, CIENTÍFICAS, TÉCNICAS E SIMILARES	0,08
ALOJAMENTO, RESTAURAÇÃO E SIMILARES	0,06
ACTIVIDADES DE INFORMAÇÃO E DE COMUNICAÇÃO	0,04
ACTIVIDADES DE SAÚDE HUMANA E APOIO SOCIAL	0,04
CONSTRUÇÃO	0,03
ACTIVIDADES ADMINISTRATIVAS E DOS SERVIÇOS DE APOIO	0,03
ACTIVIDADES IMOBILIÁRIAS	0,02
EDUCAÇÃO	0,02
TRANSPORTES E ARMAZENAGEM	0,02
ACTIVIDADES FINANCEIRAS E DE SEGUROS	0,01
OUTRAS ACTIVIDADES DE SERVIÇOS	0,01
ACTIVIDADES ARTÍSTICAS, DE ESPECTÁCULOS, DESPORTIVAS E RECREATIVAS	0,01

CAPTAÇÃO, TRATAMENTO E DISTRIBUIÇÃO DE ÁGUA; SANEAMENTO,GESTÃO DE RESÍDUOS E DESPOLUIÇÃO	0,01
AGRICULTURA, PRODUÇÃO ANIMAL, CAÇA, FLORESTA E PESCA	0,01
INDÚSTRIAS EXTRACTIVAS	0
ADMINISTRAÇÃO PÚBLICA E DEFESA; SEGURANÇA SOCIAL OBRIGATÓRIA	0
ELECTRICIDADE, GÁS, VAPOR, ÁGUA QUENTE E FRIA E AR FRIO	0

Avg Total Value Donate Per Donor	*Avg Total Number of Contributions Per Donor*	*Avg Number of Days Since Donor Last Contribution (Recency)*	*Median Number of Days Between Contribution (Frequency)*
1358,32	2,31	1764,72	-1

Cluster 2

	Rel Freq. of Donors per Region
Vale do Tejo	0,3
Entre-Douro-e-Minho	0,21
Lisboa	0,17
Beira Litoral	0,11
NOT FOUND	0,05
Alentejo	0,05
Algarve	0,03
Beira Interior	0,03
Madeira	0,03
Trás-os-Montes e Alto Douro	0,02
Açores	0,01

	Rel Freq. of Donors per Gender
Feminino	0,58
Masculino	0,42
other	0

	Rel Freq. of Donors Registered as AMI_friend
0	0,95
1	0,05

	Rel Freq. of Donors per Registration Source Group
TMN CAMPAIGNS	0,37
SPONTANEOUS REGISTRATION	0,2
OTHERS	0,2
AMI RESOURCES/DEPARTMENTS	0,09
AUCHAN CAMPAIGN	0,04
EMERGENCY CAMPAIGNS	0,03
NATIONAL CAMPAIGNS	0,02
CHRISTMAS CAMPAIGNS	0,01
CORPORATE PARTNERSHIPS	0,01
INTERNATIONAL CAMPAIGNS	0,01
DONORS	0,01
TRAINING	0,01
ENVIRONMENT	0

	Rel Freq. of Donors per Company Field Description
individuals	1

Avg Total Value Donate Per Donor	*Avg Total Number of Contributions Per Donor*	*Avg Number of Days Since Donor Last Contribution (Recency)*	*Median Number of Days Between Contribution (Frequency)*

101	1,62	2046,48	-1

Cluster 3

	Rel Freq. of Donors per Region
Vale do Tejo	0,29
Entre-Douro-e-Minho	0,23
Lisboa	0,23
Beira Litoral	0,07
NOT FOUND	0,05
Alentejo	0,04
Algarve	0,03
Beira Interior	0,02
Açores	0,01
Madeira	0,01
Trás-os-Montes e Alto Douro	0,01

	Rel Freq. of Donors per Gender
Masculino	0,64
Feminino	0,36
other	0

	AMI_friend
0	1
1	0

	Rel Freq. of Donors per Registration Source Group
OTHERS	0,94
SPONTANEOUS REGISTRATION	0,04

NATIONAL CAMPAIGNS	0,01
TMN CAMPAIGNS	0
ENVIRONMENT	0
INTERNATIONAL CAMPAIGNS	0
CORPORATE PARTNERSHIPS	0
EMERGENCY CAMPAIGNS	0
AUCHAN CAMPAIGN	0
TRAINING	0
CHRISTMAS CAMPAIGNS	0
DONORS	0

	Rel Freq. of Donors per Field Description
individuals	1

Avg Total Value Donate Per Donor	*Avg Total Number of Contributions Per Donor*	*Avg Number of Days Since Donor Last Contribution (Recency)*	*Median Number of Days Between Contribution (Frequency)*
0,12	0,01	3,73	-1

7. FEATURES SELECTED DESCRIPTION

Context	Feature Name	Description
Demographics	gender_Feminino	Donor's gender is Female
	gender_Masculino	Donor's gender is Male
	gender_company	Company (gender not applicable)
	gender_other	Gender not defined
	region_Alentejo	Donor comes from Alentejo
	region_Algarve	Donor comes from Algarve
	region_Açores	Donor comes from Açores
	region_ Beira Interior	Donor comes from Beira Interior
	region_ Beira Litoral	Donor comes from Beira Litoral
	region_Entre-Douro-e-Minho	Donor comes from Entre-Douro-e-Minho
	region_Lisboa	Donor comes from Lisboa
	region_Madeira	Donor comes from Madeira
	region_Trás-os-Montes e Alto Douro	Donor comes from Trás-os-Montes e Alto Douro
	region_Vale do Tejo	Donor comes from Vale do Tejo
	field_description_ACTIVIDADES ADMINISTRATIVAS E DOS SERVIÇOS DE APOIO	Company's work field is related to administrative work and support
	field_description_ACTIVIDADES ARTÍSTICAS, DE ESPECTÁCULOS, DESPORTIVAS E RECREATIVAS	Company's work field is related to arts and sports
	field_description_ACTIVIDADES DE INFORMAÇÃO E DE COMUNICAÇÃO	Company's work field is related to communication
	field_description_ACTIVIDADES DE SAÚDE HUMANA E APOIO SOCIAL	Company's work field is related to social work and health
	field_description_ACTIVIDADES FINANCEIRAS E DE SEGUROS	Company's work field is related to financial services

Firmographics	field_description_ACTIVIDADES IMOBILIÁRIAS	Company's work field is related to real estate
	field_description_ADMINISTRAÇÃO PÚBLICA E DEFESA; SEGURANÇA SOCIAL OBRIGATÓRIA	Company's work field is related to public administration and safety
	field_description_AGRICULTURA, PRODUÇÃO ANIMAL, CAÇA, FLORESTA E PESCA	Company's work field is related to primary sector activities
	field_description_ALOJAMENTO, RESTAURAÇÃO E SIMILARES	Company's work field is related to restaurants
	field_description_CAPTAÇÃO, TRATAMENTO E DISTRIBUIÇÃO DE ÁGUA; SANEAMENTO,GESTÃO DE RESÍDUOS E DESPOLUIÇÃO	Company's work field is related to water and waste treatment
	field_description_COMÉRCIO E VEÍCULOS	Company's work field is related to vehicles
	field_description_CONSTRUÇÃO	Company's work field is related to construction
	field_description_EDUCAÇÃO	Company's work field is related to education
	field_description_ELECTRICIDADE, GÁS, VAPOR, ÁGUA QUENTE E FRIA E AR FRIO	Company's work field is related to water, electricity and gas
	field_description_INDÚSTRIAS EXTRACTIVAS	Company's work field is related to extractive industries
	field_description_INDÚSTRIAS TRANSFORMADORAS	Company's work field is related to manufacturing industries
	field_description_OUTRAS ACTIVIDADES DE SERVIÇOS	Company's work field is not defined (others)
	field_description_TRANSPORTES E ARMAZENAGEM	Company's work field is related to transports and storage
	field_description_individuals	Donor is an individual (work field not applicable)
	Legal Form_Agrupamento Complementar de Empresas	Company's legal form is Complementary Company

	Legal Form_Associação	Company's legal form is Association
	Legal Form_Autarquia	Company's legal form is autarchy
	Legal Form_Cooperativa	Company's legal form is Collective company
	Legal Form_Empresário em Nome Individual	Company's legal form is Sole proprietorship
	Legal Form_Entidade Estrangeira	Company's legal form is Foreign entity
	Legal Form_Fundação	Company's legal form is Foundation
	Legal Form_Organismo de Administração Pública	Company's legal form is Public Administration
	Legal Form_Sociedade Anónima	Company's legal form is Public limited company
	Legal Form_Sociedade Unipessoal por Quotas	Company's legal form is Private Limited Company
	Legal Form_individuals	Donor is an individual (legal form not applicable)
	Legal Form_unknown	Not defined or unknown legal form
Relation with AMI	AMI_friend	Donor is registered as AMI Friend
	consent_boolean	Donor gave consent to be contacted by AMI
	registration_source_AMI RESOURCES/DEPARTMENTS	Donor registered with AMI through channels with direct contact with departments
	registration_source_AUCHAN CAMPAIGN	Donor registered with AMI through an Auchan campaign
	registration_source_CHRISTMAS CAMPAIGNS	Donor registered with AMI through a Christmas campaign
	registration_source_CORPORATE PARTNERSHIPS	Donor registered with AMI through a corporate partnership
	registration_source_DONORS	Donor registered with AMI through a donation
	registration_source_EMERGENCY CAMPAIGNS	Donor registered with AMI through a subscription to an Emergency campaign
	registration_source_ENVIRONMENT	Donor registered with AMI through a campaign related to the Environment
	registration_source_INTERNATIONAL CAMPAIGNS	Donor registered with AMI through an application related to an International Campaign

	registration_source_NATIONAL CAMPAIGNS	Donor registered with AMI through an application related to a National Campaign
	registration_source_SPONTANEOUS REGISTRATION	Donor registered with AMI through an spontaneous registration
	registration_source_TMN CAMPAIGNS	Donor registered with AMI through a TMN Campaign
	registration_source_TRAINING	Donor registered with AMI through a campaign related to a Training
Contributions details	total_value	Total value every contributed to AMI by the donor
	1st_value	Value of first contribution the donor ever made to AMI
	avg_value_cont	Average value of donor's contributions
	total_contributions	Total number of contributions ever made to AMI by the donor
	everismonetary	If the donor ever made a monetary contribution
	avg_days_between_contrib	Average number of days between contributions by the donor (frequency)
	days_since_donor_last_contribution	Number of days since the donor's last contribution (recency)

8. SMOTE RESAMPLING

This algorithm is called Synthetic Minority Over-sampling Technique and it was design to generate samples that are coherent with the minor class distribution. SMOTE considers the existing relationships between samples and then creates synthetics points that connect neighbors.

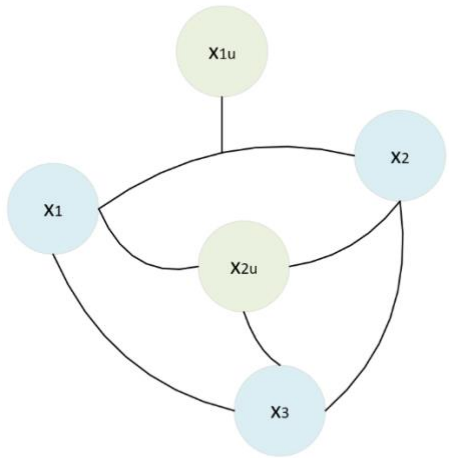


Figure 1 - neighborhood containing three points (x1, x2, x3) and two synthetic ones (O'Reilly)

As we can see from the figure above, SMOTE upsamples the classes by generating the synthetic samples x_{1u} and x_{2u}. This procedure is based on the assumption that properties of the samples are not different within a certain neighborhood radius, allowing the creation os synthetic points that follow the same original distribution.

9. GRID SEARCH: Hyperparameter Tuning

To customize a machine learning model to a specific dataset and met necessary needs, it is necessary to previously set hyperparameters. An hyperparamenter is a model characteristic that is external to the model and it can not be calculated from data. The best approach to define hyperparameters starts by testing different values for model hyperparameters and eventually choose the one that achieves the best performance (Brownlee, 2020). Grid Search is a technique used to perform an exhaustive search over hyperparameter values for an estimator (Scikit-learn.org, 2019). The output of this technique is the optimal combination of one or multiple hyperparameters that ensure the best performance of the model.

10. LOGISTIC REGRESSION

Logistic Regression (LR) is frequently used for classification problems. It entails approximately the same assumptions as linear regression, since it also attempts to find a relationship between a dependent variable and one or more independent features. However, it allows to retrieve binary results (Hosmer et. all 2013). In LR input features are linearly scaled and the result is fed to the logistic function. The logistic function transforms the input in nonlinear and the output represents the probability, within an interval [0,1], of the input belonging to class 1 (H0). The regression uses the logistic function (Sigmoid Function) to find the solution to this hypothesis:

$$H_\theta(x) = P(y = 1|x) = \sigma(\theta^T x) = \frac{1}{1 + e^{(-\theta^T x)}}$$

. The loss function of logistic regression (Logistic Loss) aims to measure the difference between the actual output and the one given by the model : the error or loss (Luo 2019). The cost function will sum the errors made for all training data with m samples, following the formula below:

$$J(\theta) = \frac{1}{m}[\sum_{i=1}^{m} -y^{(i)} log(h_\theta(x^{(i)})) + (1 - y^{(i)}) log(1 - h_\theta(x^{(i)}))]$$

The goal is to minimize this Loss Function, through a set of iterations, in order to minimize the errors made by the model, using a gradient descent function.

11. XG-Boost

XGBoost, short for Extreme Gradient Boosting, is an ensemble method - it combines multiple machine learning models (the base learners) to enhance the results when compared to their individual performance - with a gradient boosting framework (Wade & Glynn, 2020). Decision Trees are the most common baseline models for the case and make predictions by splitting data into branches and following them until the leave level. XGBoost conglomerates trees through boosting: the model will consecutively learn from the errors made by each individual tree and adapt the future predictions, so these mistakes are not repeated. (Wade & Glynn, 2020) More specifically, gradient boosting, involves the optimization of a loss function, a weak learner (decision trees), and an additive model (adds tree each time) - which will join all the weak learners until minimizing the loss function (Brownlee, 2016). The loss function considers the residuals, which indicate how different the model's predictions are from reality. Minimizing the square of these residuals is the objective of a linear regression model (Wade & Glynn, 2020).

XGBoost presents multiple advantages when compared to other classification models, including speed - thanks to block compression (blocks are compressed and decompressed when loading into main memory, which is faster than disk reading) and sharding (data is split into multiple disks, to increase performance) for instance - and accuracy - due to the possibility of regularization, and, as mentioned before, because it is an ensemble model and presents the inherent benefits.

12. PORDATA DATA

Municipality	Total Population	No Schooling	1st-4-th grade	5th-6th grade	7th - 9th grade	High school	Medium	Superior	Purchasing Powe (2017)	Average number of people per household (2020)
Arcos de Valdevez	20 268	5 060	6 972	1 998	2 801	2 038	93	1 306	67,8	1,2
Caminha	14 650	1 641	4 382	1 813	2 766	2 284	134	1 630	78,9	1,1
Melgaço	8 392	1 892	3 135	836	1 047	887	77	518	62,0	1,1
Monção	17 149	3 163	5 585	2 166	2 359	2 292	166	1 418	70,2	1,3
Paredes de Coura	8 071	1 690	2 611	1 049	1 325	923	65	408	66,4	1,4
Ponte da Barca	10 522	1 989	3 503	1 227	1 747	1 282	78	696	64,4	1,4
Ponte de Lima	36 762	5 406	10 761	6 088	7 269	4 399	243	2 596	71,0	1,8
Valença	12 256	1 454	4 150	1 762	2 248	1 698	108	836	82,7	1,6
Viana do Castelo	76 229	7 975	19 882	10 883	14 608	11 702	607	10 572	93,1	1,7
Vila Nova de Cerveira	8 023	1 161	2 344	1 078	1 543	1 130	82	685	84,2	1,4
Amares	15 750	1 847	4 714	2 380	3 007	2 316	156	1 330	70,9	1,8
Barcelos	100 389	9 333	30 188	20 688	20 058	12 261	688	7 173	78,9	2,4
Braga	151 827	10 441	33 911	19 775	30 771	26 282	1 435	29 212	107,0	2,1
Esposende	28 599	2 736	7 707	5 321	5 449	3 881	256	3 249	83,8	1,6
Terras de Bouro	6 308	1 019	2 218	827	1 121	769	41	313	63,2	1,3
Vila Verde	39 890	5 661	12 106	6 404	7 512	5 100	279	2 828	66,8	2,0
Cabeceiras de Basto	13 987	2 282	4 048	2 399	2 554	1 620	102	982	65,4	1,6
Fafe	42 815	4 981	14 080	7 845	7 224	4 792	233	3 660	75,3	1,8
Guimarães	133 412	12 043	41 867	21 501	27 000	17 593	939	12 469	91,4	2,2
Mondim de Basto	6 368	1 246	2 095	943	1 019	648	55	362	59,6	1,5
Póvoa de Lanhoso	18 316	2 692	5 901	3 109	3 191	2 092	110	1 221	69,4	1,8
Vieira do Minho	11 219	1 804	3 706	1 863	1 739	1 334	72	701	67,2	1,4
Vila Nova de Famalicão	112 215	9 573	32 200	20 163	22 038	15 972	1 024	11 245	88,8	2,3
Vizela	19 758	1 947	6 784	3 446	3 906	2 355	121	1 199	84,1	2,5
Arouca	18 896	2 537	6 273	3 438	3 123	1 943	189	1 393	70,8	1,9
Espinho	27 743	2 719	8 545	3 530	4 947	3 867	177	3 958	103,0	1,8
Gondomar	142 182	11 149	40 945	20 441	29 037	23 336	1 455	15 819	84,1	2,2

Maia	112 530	6 634	27 601	12 891	22 026	20 416	⊥ 1 435	21 527	110,7	2,3
Matosinhos	150 366	11 185	40 246	18 208	27 428	24 602	⊥ 1 677	27 020	123,0	2,1
Oliveira de Azeméis	58 932	5 420	18 867	10 614	11 371	7 358	⊥ 502	4 800	83,1	2,2
Paredes	70 716	6 763	23 256	13 611	13 444	8 205	⊥ 474	4 963	79,8	2,4
Porto	209 212	13 814	50 642	21 333	35 585	32 936	⊥ 1 917	52 985	157,8	1,5
Póvoa de Varzim	52 991	4 357	15 051	9 383	9 972	7 441	⊥ 417	6 370	95,5	1,7
Santa Maria da Feira	117 270	10 799	35 941	19 825	21 968	15 581	⊥ 980	12 176	84,8	2,2
Santo Tirso	61 648	5 920	21 374	9 750	11 039	7 599	⊥ 492	5 474	85,8	2,2
São João da Madeira	18 587	1 285	5 089	2 697	3 787	3 037	⊥ 210	2 482	135,4	2,1
Trofa	32 924	2 644	10 209	5 425	6 656	4 714	⊥ 361	2 915	92,7	2,4
Vale de Cambra	19 965	2 458	6 491	3 079	3 349	2 505	⊥ 195	1 888	86,9	1,8
Valongo	78 319	5 562	21 977	11 843	15 730	12 981	⊥ 852	9 374	91,7	2,4
Vila do Conde	66 602	5 488	20 096	11 110	12 488	9 091	⊥ 594	7 735	96,2	2,1
Vila Nova de Gaia	255 654	18 851	70 853	33 970	49 801	41 476	⊥ 2 752	37 951	100,1	2,1
Boticas	5 169	1 269	2 053	575	513	475	⊥ 26	258	59,0	1,1
Chaves	36 213	4 900	11 561	3 950	5 899	5 326	⊥ 219	4 358	79,2	1,4
Montalegre	9 534	2 193	3 565	1 069	1 132	951	⊥ 39	585	62,4	0,8
Ribeira de Pena	5 725	1 468	1 805	815	831	499	⊥ 23	284	63,3	1,2
Valpaços	15 166	3 478	6 048	1 701	1 760	1 332	⊥ 54	793	59,4	1,1
Vila Pouca de Aguiar	11 652	2 074	4 488	1 351	1 690	1 195	⊥ 42	812	63,5	1,2
Amarante	47 227	6 125	16 181	6 200	8 652	5 745	⊥ 307	4 017	71,8	1,8
Baião	17 410	3 429	6 123	2 568	2 825	1 634	⊥ 82	749	58,2	1,6
Castelo de Paiva	14 030	1 466	4 238	3 361	2 683	1 454	⊥ 99	729	65,7	2,0
Celorico de Basto	17 037	2 727	5 673	2 672	2 951	1 947	⊥ 114	953	56,1	1,7
Cinfães	17 405	3 155	6 269	3 222	2 432	1 524	⊥ 76	727	56,9	1,5
Felgueiras	48 098	5 231	16 485	8 338	9 147	5 717	⊥ 318	2 862	81,6	2,3
Lousada	38 572	4 033	12 660	7 747	7 844	3 904	⊥ 189	2 195	71,1	2,4
Marco de Canaveses	43 795	5 107	14 693	8 170	8 209	4 755	⊥ 311	2 550	74,1	2,1
Paços de Ferreira	46 020	4 417	15 865	9 976	8 191	4 516	⊥ 225	2 830	79,4	2,6
Penafiel	59 509	6 801	18 873	10 471	11 719	7 240	⊥ 330	4 075	78,9	2,2
Resende	9 661	2 047	3 463	1 194	1 461	961	⊥ 26	509	59,5	1,3
Douro	178 041	27 711	58 194	21 672	28 155	22 799	⊥ 1 083	18 427	76,1	1,3

Alijó	10 596	1 868	4 161	1 370	1 455	1 082	⊥ 62	598	63,6	1,2
Armamar	5 468	875	2 105	746	811	611	⊥ 32	288	63,3	1,1
Carrazeda de Ansiães	5 737	1 195	2 318	592	687	533	⊥ 31	381	63,0	1,1
Freixo de Espada à Cinta	3 351	759	1 273	398	454	279	⊥ 21	167	63,2	1,0
Lamego	22 973	3 373	6 674	3 120	3 938	3 060	⊥ 130	2 678	81,4	1,5
Mesão Frio	3 854	687	1 385	531	606	444	⊥ 22	179	69,5	1,6
Moimenta da Beira	8 754	1 412	3 043	1 056	1 381	1 084	⊥ 63	715	66,3	1,2
Murça	5 287	1 058	1 867	682	760	520	⊥ 31	369	61,2	1,3
Penedono	2 590	489	953	351	315	283	⊥ 15	184	62,1	1,0
Peso da Régua	14 809	2 184	4 804	1 892	2 545	1 961	⊥ 84	1 339	82,5	1,7
Sabrosa	5 571	1 048	2 086	660	847	567	⊥ 22	341	61,9	1,3
Santa Marta de Penaguião	6 470	1 406	2 326	800	875	633	⊥ 30	400	58,7	1,4
São João da Pesqueira	6 815	1 236	2 503	1 032	1 005	671	⊥ 31	337	64,6	1,3
Sernancelhe	4 954	854	1 846	601	744	618	⊥ 35	256	58,8	1,2
Tabuaço	5 560	964	1 973	775	923	610	⊥ 24	291	55,3	1,4
Tarouca	6 770	990	2 374	1 014	1 139	784	⊥ 35	434	63,4	1,1
Torre de Moncorvo	7 826	1 641	2 876	790	1 003	886	⊥ 58	572	62,3	1,0
Vila Nova de Foz Côa	6 520	1 204	2 498	678	844	747	⊥ 45	504	68,0	1,0
Vila Real	44 136	4 468	11 129	4 584	7 823	7 426	⊥ 312	8 394	98,1	1,6
Alfândega da Fé	4 608	930	1 624	487	729	488	⊥ 31	319	66,0	1,2
Bragança	30 964	4 022	7 999	3 157	5 137	4 749	⊥ 338	5 562	96,5	1,3
Macedo de Cavaleiros	13 928	2 323	4 808	1 473	2 163	1 616	⊥ 91	1 454	73,5	1,2
Miranda do Douro	6 752	1 322	2 167	676	1 090	838	⊥ 48	611	72,7	1,2
Mirandela	20 954	3 292	6 334	2 592	3 273	2 861	⊥ 139	2 463	83,0	1,4
Mogadouro	8 655	1 830	2 916	913	1 267	912	⊥ 73	744	67,9	1,1
Vila Flor	5 976	1 068	2 188	725	887	647	⊥ 32	429	62,4	1,1
Vimioso	4 264	991	1 628	475	546	362	⊥ 19	243	62,3	0,8
Vinhais	8 372	1 779	3 338	1 189	876	731	⊥ 44	415	57,9	1,1
Alcobaça	48 411	5 701	14 968	6 075	9 487	7 184	⊥ 437	4 559	86,7	1,5
Alenquer	36 130	4 111	9 935	4 823	7 417	6 173	⊥ 342	3 329	89,2	1,9
Arruda dos Vinhos	10 929	1 039	2 909	1 366	2 003	1 957	⊥ 141	1 514	91,0	2,3

Bombarral	11 425	1 458	3 663	1 570	2 115	1 550	⊥ 104	965	83,5	1,5
Cadaval	12 206	1 823	4 128	1 609	2 072	1 671	⊥ 99	804	70,8	1,5
Caldas da Rainha	44 190	4 697	11 760	4 907	9 132	7 513	⊥ 500	5 681	98,1	1,6
Lourinhã	21 767	2 945	6 415	3 060	4 047	3 179	⊥ 215	1 906	78,3	1,5
Nazaré	13 052	1 406	4 073	1 805	2 427	1 913	⊥ 152	1 276	85,6	1,0
Óbidos	10 049	1 370	3 070	1 339	1 800	1 457	⊥ 89	924	75,5	1,3
Peniche	23 634	2 850	7 204	3 465	4 631	3 158	⊥ 232	2 094	85,9	1,2
Sobral de Monte Agraço	8 466	951	2 447	1 212	1 612	1 324	⊥ 89	831	103,4	2,0
Torres Vedras	67 324	7 616	18 649	9 056	13 580	10 130	⊥ 669	7 624	95,3	1,7
Águeda	41 087	4 116	12 548	6 537	7 883	5 589	⊥ 418	3 996	86,5	2,0
Albergaria-a-Velha	21 359	2 092	6 203	3 770	4 253	2 783	⊥ 225	2 033	84,5	1,9
Anadia	25 411	3 342	8 197	3 272	4 226	3 287	⊥ 266	2 821	78,7	1,8
Aveiro	67 019	4 769	15 037	8 228	13 423	10 510	⊥ 651	14 401	123,1	1,9
Estarreja	23 054	2 402	7 246	3 594	4 637	2 878	⊥ 151	2 146	82,1	2,0
Ílhavo	32 643	2 602	9 022	4 449	6 551	4 761	⊥ 310	4 948	88,6	1,7
Murtosa	8 946	1 199	3 198	1 487	1 411	880	⊥ 51	720	69,3	1,3
Oliveira do Bairro	19 401	2 341	5 575	2 970	3 501	2 628	⊥ 189	2 197	79,7	2,1
Ovar	46 815	3 949	13 114	7 139	9 496	7 022	⊥ 388	5 707	88,9	1,9
Sever do Vouga	10 716	1 212	3 377	1 728	1 996	1 381	⊥ 95	927	74,0	1,6
Vagos	19 446	2 666	5 850	2 936	3 561	2 486	⊥ 162	1 785	72,1	1,7
Arganil	10 795	1 952	3 811	1 193	1 967	1 249	⊥ 54	569	68,2	1,0
Cantanhede	31 872	4 846	10 106	3 782	5 284	4 096	⊥ 287	3 471	79,8	1,7
Coimbra	125 559	9 331	26 950	11 196	21 811	21 230	⊥ 1 235	33 806	128,7	1,7
Condeixa-a-Nova	14 340	1 705	3 676	1 307	2 580	2 438	⊥ 160	2 474	77,9	2,0
Figueira da Foz	54 060	6 642	14 323	6 071	10 593	8 623	⊥ 457	7 351	95,0	1,3
Góis	3 793	830	1 335	368	660	393	⊥ 25	182	64,0	0,7
Lousã	14 824	1 314	4 161	1 839	3 179	2 466	⊥ 145	1 720	80,6	1,6
Mealhada	17 597	1 807	5 251	2 175	3 301	2 644	⊥ 233	2 186	86,9	1,9
Mira	10 905	1 664	3 489	1 348	1 683	1 340	⊥ 120	1 261	72,2	1,3
Miranda do Corvo	11 269	1 366	3 524	1 494	2 175	1 640	⊥ 87	983	67,3	1,7
Montemor-o-Velho	22 789	3 265	6 505	2 776	4 256	3 402	⊥ 171	2 414	71,0	1,9
Mortágua	8 595	1 437	3 154	1 062	1 271	934	⊥ 54	683	74,1	1,4
Oliveira do Hospital	18 068	2 695	6 335	2 341	3 124	2 123	⊥ 127	1 323	74,5	1,4

Pampilhosa da Serra	4 160	1 213	1 617	385	470	350	⊥ 10	115	64,6	0,7
Penacova	13 377	2 163	4 791	1 698	2 342	1 437	⊥ 85	861	63,7	1,6
Penela	5 252	916	1 856	523	857	632	⊥ 53	415	70,0	1,1
Soure	16 987	3 205	5 060	1 539	3 112	2 579	⊥ 160	1 332	70,7	1,4
Tábua	10 434	1 575	3 760	1 233	1 945	1 201	⊥ 81	639	68,6	1,3
Vila Nova de Poiares	6 185	803	1 914	797	1 354	867	⊥ 49	401	70,5	1,5
Alvaiázere	6 500	1 364	2 369	697	911	708	⊥ 51	400	66,5	1,1
Ansião	11 445	2 027	3 736	1 263	1 901	1 566	⊥ 136	816	73,5	1,4
Batalha	13 335	1 478	3 997	1 742	2 599	2 013	⊥ 174	1 332	84,8	1,9
Castanheira de Pêra	2 853	518	1 133	288	454	294	⊥ 20	146	65,6	0,9
Figueiró dos Vinhos	5 512	1 037	1 934	626	860	634	⊥ 56	365	65,5	1,1
Leiria	107 580	10 931	26 728	13 224	21 749	17 851	⊥ 1 301	15 796	103,4	1,8
Marinha Grande	32 879	3 198	8 708	3 972	7 173	5 809	⊥ 369	3 650	98,5	1,7
Pedrógão Grande	3 481	697	1 260	392	490	392	⊥ 61	189	67,9	0,9
Pombal	47 489	8 928	14 253	5 282	8 332	6 453	⊥ 458	3 783	82,2	1,5
Porto de Mós	20 684	2 569	6 507	2 754	3 999	2 801	⊥ 222	1 832	80,2	1,8
Aguiar da Beira	4 862	1 236	1 743	455	655	514	⊥ 33	226	67,4	0,9
Carregal do Sal	8 488	1 317	2 934	1 170	1 477	993	⊥ 39	558	70,9	1,4
Castro Daire	13 308	2 765	4 786	1 739	2 005	1 262	⊥ 49	702	64,8	1,1
Mangualde	17 207	2 272	5 717	2 312	3 109	2 135	⊥ 94	1 568	82,4	1,4
Nelas	12 111	1 529	4 141	1 463	2 194	1 601	⊥ 77	1 106	77,1	1,5
Oliveira de Frades	8 718	1 041	2 815	1 355	1 585	1 132	⊥ 82	708	77,5	1,7
Penalva do Castelo	6 977	1 509	2 572	873	956	640	⊥ 39	388	58,4	1,3
Santa Comba Dão	10 064	1 432	3 306	1 415	1 730	1 255	⊥ 93	833	71,5	1,5
São Pedro do Sul	14 706	2 341	5 117	2 026	2 227	1 779	⊥ 109	1 107	68,7	1,3
Sátão	10 717	1 913	3 548	1 475	1 605	1 249	⊥ 78	849	61,9	1,2
Tondela	25 503	3 832	8 966	3 323	4 134	3 073	⊥ 151	2 024	75,0	1,5
Vila Nova de Paiva	4 476	888	1 378	639	726	518	⊥ 22	305	62,7	1,0
Viseu	84 115	8 360	21 578	9 538	15 184	13 792	⊥ 682	14 981	94,4	1,7
Vouzela	9 232	1 360	3 392	1 261	1 453	1 038	⊥ 68	660	64,0	1,4
Castelo Branco	49 002	6 030	13 389	4 514	9 045	8 265	⊥ 449	7 310	95,8	1,3

Idanha-a-Nova	8 870	2 491	3 249	751	1 155	747	⊥ 29	448	67,6	0,7
Oleiros	5 327	1 381	2 137	433	618	432	⊥ 40	286	63,6	1,0
Penamacor	5 267	1 451	1 915	529	683	456	⊥ 17	216	60,6	0,7
Proença-a-Nova	7 511	1 451	2 650	677	1 165	937	⊥ 61	570	69,0	1,1
Vila Velha de Ródão	3 258	616	1 423	288	440	309	⊥ 7	175	71,7	0,9
Abrantes	34 378	4 396	10 675	3 954	6 642	4 933	⊥ 253	3 525	89,2	1,4
Alcanena	12 005	1 342	3 814	1 669	2 201	1 807	⊥ 96	1 076	86,1	1,6
Constância	3 437	404	1 025	387	677	556	⊥ 36	352	83,1	1,8
Entroncamento	16 951	1 002	3 620	1 749	3 764	3 615	⊥ 219	2 982	98,5	2,0
Ferreira do Zêzere	7 525	1 453	2 632	876	1 228	892	⊥ 63	381	67,8	1,0
Mação	6 672	1 346	2 358	687	1 202	653	⊥ 38	388	68,8	0,9
Ourém	39 265	5 947	11 214	5 102	7 213	5 559	⊥ 437	3 793	83,6	1,5
Sardoal	3 458	451	1 115	419	665	498	⊥ 22	288	68,6	1,2
Sertã	13 900	2 592	4 722	1 369	2 438	1 785	⊥ 169	825	72,6	1,2
Tomar	35 415	4 049	10 698	4 109	6 629	5 456	⊥ 318	4 156	85,0	1,4
Torres Novas	31 654	3 412	8 944	3 800	5 954	5 155	⊥ 296	4 093	96,8	1,7
Vila de Rei	3 088	686	1 185	246	455	328	⊥ 48	140	66,2	1,1
Vila Nova da Barquinha	6 320	626	1 885	831	1 204	1 061	⊥ 69	644	72,3	1,8
Almeida	6 650	1 150	2 591	608	1 005	794	⊥ 37	465	74,2	0,9
Belmonte	6 051	988	2 069	704	954	796	⊥ 53	487	71,3	1,3
Celorico da Beira	6 755	1 312	2 482	598	1 127	722	⊥ 51	463	65,0	1,1
Covilhã	45 428	5 682	13 591	4 863	8 242	6 817	⊥ 412	5 821	86,6	1,3
Figueira de Castelo Rodrigo	5 588	1 150	2 014	652	797	528	⊥ 32	415	66,8	1,0
Fornos de Algodres	4 447	866	1 786	415	631	423	⊥ 31	295	58,8	1,1
Fundão	25 779	4 506	8 022	2 813	4 640	3 099	⊥ 249	2 450	77,9	1,2
Gouveia	12 557	2 126	4 818	1 294	1 857	1 279	⊥ 77	1 106	65,9	1,1
Guarda	36 708	3 848	9 423	3 333	7 689	5 812	⊥ 373	6 230	96,2	1,4
Manteigas	3 087	436	1 288	311	435	368	⊥ 25	224	63,9	1,1
Mêda	4 680	1 017	1 836	474	580	426	⊥ 32	315	62,1	0,9
Pinhel	8 627	1 642	3 243	828	1 351	941	⊥ 51	571	62,8	1,0
Sabugal	11 540	2 783	4 415	895	1 605	1 092	⊥ 80	670	63,4	0,7
Seia	21 941	3 096	8 118	2 332	3 539	2 768	⊥ 166	1 922	76,2	1,2
Trancoso	8 773	1 620	3 218	754	1 429	1 002	⊥ 59	691	66,7	1,1
Alcochete	14 237	1 090	2 920	1 501	2 803	2 926	⊥ 172	2 825	118,8	2,2

Almada	148 447	11 069	35 248	16 036	31 068	28 068	⊥ 1 701	25 257	108,7	1,6
Amadora	149 233	11 752	37 290	16 934	30 929	28 390	⊥ 1 677	22 261	100,6	2,1
Barreiro	67 543	5 249	18 670	7 477	14 327	12 494	⊥ 685	8 641	100,0	1,8
Cascais	173 824	9 767	30 426	15 484	34 083	37 171	⊥ 2 709	44 184	122,1	1,9
Lisboa	477 239	32 413	96 850	38 897	74 655	80 869	⊥ 5 142	148 413	219,6	1,6
Loures	172 998	13 596	43 559	20 600	35 867	31 177	⊥ 1 942	26 257	92,3	2,1
Mafra	62 320	4 889	14 391	7 953	12 556	11 779	⊥ 828	9 924	96,3	1,9
Moita	55 480	5 139	14 878	7 257	12 827	9 906	⊥ 622	4 851	82,0	1,8
Montijo	42 716	4 124	9 825	5 005	8 654	8 215	⊥ 542	6 351	99,2	2,1
Odivelas	122 637	8 274	30 273	14 102	25 506	23 533	⊥ 1 525	19 424	89,3	2,2
Oeiras	145 561	7 196	23 443	11 384	26 408	30 518	⊥ 1 922	44 690	156,5	2,0
Palmela	52 151	5 271	12 467	6 087	11 256	9 448	⊥ 583	7 039	98,1	1,9
Seixal	132 522	8 558	30 804	15 780	30 994	26 710	⊥ 1 624	18 052	89,7	2,0
Sesimbra	40 885	3 167	9 470	5 153	9 370	7 998	⊥ 574	5 153	90,0	1,6
Setúbal	101 628	8 977	22 976	11 561	22 508	19 196	⊥ 1 154	15 256	107,5	1,8
Sintra	311 202	18 351	64 044	39 362	76 335	66 819	⊥ 4 091	42 200	94,1	2,1
Vila Franca de Xira	113 372	7 469	25 084	13 649	26 041	24 215	⊥ 1 419	15 495	98,4	2,2
Alcácer do Sal	11 361	2 142	3 661	1 489	1 853	1 387	⊥ 64	765	81,5	1,3
Grândola	12 989	2 404	3 733	1 790	2 175	1 862	⊥ 82	943	85,9	1,2
Odemira	22 904	5 247	6 074	2 520	4 176	3 114	⊥ 163	1 610	81,7	1,2
Santiago do Cacém	26 088	3 881	6 630	2 790	5 106	4 602	⊥ 302	2 777	92,3	1,5
Sines	12 170	1 278	3 142	1 500	2 622	2 198	⊥ 173	1 257	128,7	1,6
Aljustrel	8 194	1 329	2 566	952	1 638	1 018	⊥ 37	654	87,9	1,4
Almodôvar	6 566	1 602	1 788	593	1 411	767	⊥ 20	385	79,5	1,2
Alvito	2 179	399	601	258	421	316	⊥ 14	170	66,7	1,4
Barrancos	1 588	241	437	242	288	281	⊥ 5	94	63,4	1,3
Beja	30 480	3 598	7 010	3 316	6 309	5 036	⊥ 311	4 900	105,3	1,6
Castro Verde	6 320	926	1 904	729	1 200	869	⊥ 65	627	102,1	1,4
Cuba	4 241	647	1 254	487	805	633	⊥ 36	379	67,9	1,5
Ferreira do Alentejo	7 243	1 407	2 231	899	1 289	845	⊥ 50	522	73,0	1,5
Mértola	6 609	1 524	2 235	826	967	692	⊥ 28	337	66,6	0,7
Moura	12 765	2 449	3 690	1 714	2 352	1 580	⊥ 71	909	77,5	1,3
Ourique	4 839	1 106	1 474	520	839	619	⊥ 16	265	77,3	1,1
Serpa	13 670	2 791	3 877	1 779	2 467	1 648	⊥ 89	1 019	72,8	1,4
Vidigueira	5 114	877	1 518	628	918	714	⊥ 52	407	72,8	1,4
Almeirim	19 837	2 907	5 841	2 523	3 493	2 809	⊥ 188	2 076	86,2	1,8
Alpiarça	6 558	1 004	2 168	687	1 178	850	⊥ 49	622	77,1	1,7

Azambuja	18 608	2 301	5 493	2 420	3 740	2 914	⊥ 171	1 569	106,2	1,9
Benavente	23 873	2 341	5 887	3 439	5 186	4 237	⊥ 276	2 507	94,4	2,1
Cartaxo	20 865	2 083	5 793	2 692	4 310	3 433	⊥ 236	2 318	86,5	1,7
Chamusca	8 943	1 565	3 138	1 197	1 513	927	⊥ 46	557	71,7	1,5
Coruche	17 556	3 667	5 921	1 854	2 715	2 058	⊥ 127	1 214	76,8	1,4
Golegã	4 730	615	1 424	605	910	684	⊥ 40	452	82,6	1,7
Rio Maior	17 993	2 253	5 309	2 412	3 474	2 736	⊥ 206	1 603	90,2	1,6
Salvaterra de Magos	18 900	3 248	5 348	2 542	3 522	2 705	⊥ 176	1 359	77,5	1,8
Santarém	53 309	5 821	14 305	5 809	10 165	8 656	⊥ 516	8 037	101,4	1,6
Alter do Chão	3 178	640	1 043	359	496	398	⊥ 10	232	72,9	1,0
Arronches	2 833	644	891	321	441	331	⊥ 12	193	72,0	1,1
Avis	4 052	847	1 305	493	646	473	⊥ 29	259	74,1	1,1
Campo Maior	7 140	1 154	1 926	850	1 408	1 125	⊥ 64	613	94,0	1,6
Castelo de Vide	3 063	578	910	294	586	394	⊥ 14	287	81,6	1,0
Crato	3 357	703	1 214	377	540	312	⊥ 18	193	72,8	0,9
Elvas	19 507	2 652	5 155	2 472	3 978	3 175	⊥ 183	1 892	89,0	1,5
Fronteira	2 972	590	848	395	518	371	⊥ 15	235	79,0	1,1
Gavião	3 774	814	1 448	438	607	303	⊥ 19	145	71,8	0,9
Marvão	3 179	667	1 019	404	512	336	⊥ 12	229	64,8	1,0
Monforte	2 840	678	899	310	424	327	⊥ 23	179	73,9	1,3
Nisa	6 745	1 419	2 356	668	1 057	729	⊥ 50	466	73,2	0,8
Ponte de Sor	14 609	2 696	4 520	1 873	2 386	1 964	⊥ 81	1 089	84,4	1,4
Portalegre	21 680	2 767	5 454	2 287	4 164	3 304	⊥ 148	3 556	104,1	1,5
Sousel	4 432	895	1 381	502	809	515	⊥ 25	305	69,0	1,2
Alandroal	5 178	1 050	1 841	702	788	518	⊥ 21	258	64,5	1,1
Arraiolos	6 451	1 106	1 862	812	1 222	916	⊥ 46	487	73,3	1,4
Borba	6 483	1 243	1 899	848	1 057	929	⊥ 49	458	73,9	1,6
Estremoz	12 652	2 244	3 551	1 341	2 279	1 978	⊥ 97	1 162	93,9	1,3
Évora	48 448	4 684	10 870	5 281	9 072	9 078	⊥ 508	8 955	117,3	1,7
Montemor-o-Novo	15 342	2 901	4 723	1 700	2 476	2 081	⊥ 116	1 345	86,8	1,5
Mora	4 474	960	1 579	454	729	453	⊥ 19	280	82,4	1,0
Mourão	2 251	434	705	320	400	280	⊥ 12	100	69,9	1,3
Portel	5 627	1 080	1 926	770	872	668	⊥ 45	266	64,8	1,4
Redondo	6 130	1 078	1 938	877	1 045	751	⊥ 42	399	71,9	1,4
Reguengos de Monsaraz	9 286	1 622	2 726	1 185	1 508	1 345	⊥ 76	824	89,4	1,5
Vendas Novas	10 175	1 489	3 036	1 207	1 934	1 511	⊥ 91	907	95,2	1,7

Viana do Alentejo	4 920	933	1 360	671	894	689	⊥ 34	339	79,3	1,4
Vila Viçosa	7 258	1 082	2 119	996	1 234	1 156	⊥ 60	611	83,4	1,6
Albufeira	34 328	3 083	7 184	4 315	7 958	7 354	⊥ 488	3 946	112,0	1,0
Alcoutim	2 687	760	958	305	295	237	⊥ 11	121	67,5	0,6
Aljezur	5 216	980	1 274	510	950	969	⊥ 50	483	63,1	0,9
Castro Marim	5 909	971	1 778	804	1 061	807	⊥ 37	451	71,7	0,7
Faro	55 160	4 508	11 908	5 258	11 008	10 922	⊥ 622	10 934	132,5	1,6
Lagoa	19 377	1 864	4 813	2 452	4 559	3 547	⊥ 197	1 945	89,8	1,1
Lagos	26 179	2 769	5 709	3 009	5 768	5 378	⊥ 328	3 218	93,8	1,1
Loulé	60 330	6 866	15 520	6 823	12 642	11 076	⊥ 714	6 689	105,9	1,0
Monchique	5 446	1 088	1 825	517	829	778	⊥ 43	366	61,9	1,1
Olhão	37 894	3 878	10 137	4 880	8 232	6 417	⊥ 410	3 940	81,1	1,7
Portimão	46 899	4 668	10 594	5 385	10 600	8 648	⊥ 620	6 384	103,5	1,2
São Brás de Alportel	9 127	906	2 611	1 010	1 672	1 687	⊥ 108	1 133	84,8	1,5
Silves	31 997	4 207	8 524	3 947	6 896	5 327	⊥ 321	2 775	76,9	1,1
Tavira	22 654	3 195	6 209	2 589	4 093	3 815	⊥ 235	2 518	92,0	0,9
Vila do Bispo	4 647	632	1 329	593	916	761	⊥ 45	371	64,6	0,8
Vila Real de Santo António	16 182	1 591	4 574	2 187	3 505	2 712	⊥ 168	1 445	91,8	0,9
Vila do Porto	4 589	476	1 332	787	892	675	⊥ 56	371	89,7	1,5
Lagoa [R.A.A.]	11 413	1 208	3 527	2 422	2 180	1 128	⊥ 86	862	74,2	2,8
Nordeste	4 061	424	1 373	844	777	419	⊥ 13	211	62,5	1,8
Ponta Delgada	56 380	5 343	12 375	10 332	11 801	7 944	⊥ 662	7 923	107,8	2,3
Povoação	5 209	732	1 689	1 105	879	485	⊥ 26	293	66,3	1,6
Ribeira Grande	24 623	3 192	7 146	5 556	4 366	2 467	⊥ 192	1 704	71,1	2,8
Vila Franca do Campo	9 045	1 352	2 931	1 890	1 477	799	⊥ 50	546	65,0	2,6
Angra do Heroísmo	29 609	2 810	8 851	4 842	5 547	3 758	⊥ 264	3 537	94,7	2,2
Vila da Praia da Vitória	17 661	1 879	5 910	2 847	3 346	2 265	⊥ 118	1 296	74,7	2,2
Santa Cruz da Graciosa	3 741	529	1 357	583	631	353	⊥ 20	268	73,1	1,5
Calheta [R.A.A.]	3 216	536	1 177	478	570	253	⊥ 13	189	77,1	1,3
Velas	4 620	497	1 659	658	915	520	⊥ 38	333	83,1	1,6

Lajes do Pico	4 088	409	1 541	628	682	484	⊥ 27	317	70,0	1,4
Madalena	5 170	435	1 941	651	949	738	⊥ 62	394	89,1	1,7
São Roque do Pico	2 941	221	1 077	422	529	408	⊥ 27	257	81,5	1,4
Horta	12 591	914	3 704	2 009	2 531	1 868	⊥ 137	1 428	90,7	2,0
Lajes das Flores	1 283	141	479	186	227	143	⊥ 11	96	74,8	1,4
Santa Cruz das Flores	1 966	158	690	277	433	221	⊥ 8	179	88,4	1,9
Corvo	369	42	135	64	57	46	⊥ 1	24	76,2	2,4
Calheta [R.A.M.]	9 806	1 926	3 369	1 265	1 363	1 067	⊥ 79	737	62,8	1,4
Câmara de Lobos	28 221	4 927	9 039	4 973	4 838	2 813	⊥ 234	1 397	58,3	2,5
Funchal	95 487	8 827	24 428	13 207	17 527	15 081	⊥ 1 241	15 176	114,3	2,0
Machico	18 370	2 508	6 335	2 820	3 021	2 055	⊥ 190	1 441	78,2	2,0
Ponta do Sol	7 246	1 249	2 363	1 123	1 098	791	⊥ 54	568	55,2	1,8
Porto Moniz	2 380	610	842	331	263	197	⊥ 11	126	56,9	1,2
Ribeira Brava	11 005	2 191	3 429	1 626	1 595	1 249	⊥ 79	836	68,4	1,8
Santa Cruz	34 964	2 837	7 860	5 431	7 273	6 284	⊥ 679	4 600	71,5	2,2
Santana	6 708	1 512	2 307	863	853	676	⊥ 69	428	58,3	1,4
São Vicente	4 921	1 140	1 646	696	589	508	⊥ 33	309	61,0	1,3
Porto Santo	4 665	399	1 240	698	975	852	⊥ 58	443	93,5	1,1

13. SEGMENTATION MODEL RESULTS AFTER ADDING NEW SOCIO-ECONOMIC FEATURES

Cluster 0

	Rel Freq. of Donors per ***Region***
Vale do Tejo	0,33
Entre-Douro-e-Minho	0,28
Lisboa	0,14
Beira Litoral	0,11
Algarve	0,03
Alentejo	0,02
NOT FOUND	0,02
Beira Interior	0,02
Trás-os-Montes e Alto Douro	0,02
Açores	0,01
Madeira	0,01

	Rel Freq. of Donors per ***Gender***
Feminino	0,71
Masculino	0,29
other	0

	Rel Freq. of Donors Registered as AMI_friend
1	0,59
0	0,41

	Rel Freq. of Donors ***per Registration Source Group***
AMI RESOURCES/DEPARTMENTS	0,57
AUCHAN CAMPAIGN	0,14
SPONTANEOUS REGISTRATION	0,1
CHRISTMAS CAMPAIGNS	0,06
TMN CAMPAIGNS	0,05
DONORS	0,03
EMERGENCY CAMPAIGNS	0,03
NATIONAL CAMPAIGNS	0,01
OTHERS	0,01
INTERNATIONAL CAMPAIGNS	0
CORPORATE PARTNERSHIPS	0
TRAINING	0
ENVIRONMENT	0

	Rel Freq. of Donors ***per Company Field Description***
individuals	1

	Rel Freq. of Donors ***per Company Legal Form***
individuals	1

Avg Total Value Donate Per Donor	*Avg Total Number of Contributions Per Donor*	*Avg Number of Days Since Donor Last Contribution (Recency)*	*Median Number of Days Between Contribution (Frequency)*	*Avg Purchasing Power*	*Avg Number of People per Household*	*Avg Education Score*
818,16	16,01	2506,6	127	120,22	1,81	2,86

Cluster 1

	Rel Freq. of Donors ***per Region***
Vale do Tejo	0,28
Lisboa	0,25
Entre-Douro-e-Minho	0,23
Beira Litoral	0,07
NOT FOUND	0,05
Alentejo	0,04
Algarve	0,03
Beira Interior	0,02
Açores	0,01
Madeira	0,01
Trás-os-Montes e Alto Douro	0,01

	Rel Freq. of Donors ***per Gender***
Masculino	0,63
Feminino	0,37
other	0

	Rel Freq. of Donors Registered as AMI_friend
0	1
1	0

	Rel Freq. of Donors per Registration Source Group
OTHERS	0,91
SPONTANEOUS REGISTRATION	0,06
NATIONAL CAMPAIGNS	0,01
INTERNATIONAL CAMPAIGNS	0
ENVIRONMENT	0
TMN CAMPAIGNS	0
AMI RESOURCES/DEPARTMENTS	0
CORPORATE PARTNERSHIPS	0
AUCHAN CAMPAIGN	0
EMERGENCY CAMPAIGNS	0
TRAINING	0
DONORS	0
CHRISTMAS CAMPAIGNS	0

	Rel Freq. of Donors ***per Company Field Description***
individuals	1

Avg Total Value Donate Per Donor	*Avg Total Number of Contributions Per Donor*	*Avg Number of Days Since Donor Last Contribution (Recency)*	*Median Number of Days Between Contribution (Frequency)*	*Avg Purchasing Power*	*Avg Number of People per Household*	*Avg Education Score*
0,35	0,01	13,21	-0,9	136,8	1,76	2,96

Cluster 2

	*Rel Freq. of Donors **per Region***
Vale do Tejo	0,25
Entre-Douro-e-Minho	0,21
Lisboa	0,19
Beira Litoral	0,12
NOT FOUND	0,07
Alentejo	0,04
Algarve	0,03
Beira Interior	0,03
Madeira	0,02
Açores	0,02
Trás-os-Montes e Alto Douro	0,02

	*Rel Freq. of Donors **per Gender***
company	1

	Rel Freq. of Donors Registered as AMI_friend
0	0,97
1	0,03

	*Rel Freq. of Donors **per Registration Source Group***
OTHERS	0,5
SPONTANEOUS REGISTRATION	0,21
TMN CAMPAIGNS	0,05
ENVIRONMENT	0,05
AMI RESOURCES/DEPARTMENTS	0,05
CORPORATE PARTNERSHIPS	0,04
EMERGENCY CAMPAIGNS	0,03
INTERNATIONAL CAMPAIGNS	0,02
CHRISTMAS CAMPAIGNS	0,01
DONORS	0,01
NATIONAL CAMPAIGNS	0,01
TRAINING	0,01
AUCHAN CAMPAIGN	0

	*Rel Freq. of Donors **per Company Field Description***
not found	0,33
COMÉRCIO E VEÍCULOS	0,19
INDÚSTRIAS TRANSFORMADORAS	0,09

ACTIVIDADES DE CONSULTORIA, CIENTÍFICAS, TÉCNICAS E SIMILARES	0,08
ALOJAMENTO, RESTAURAÇÃO E SIMILARES	0,06
ACTIVIDADES DE INFORMAÇÃO E DE COMUNICAÇÃO	0,04
ACTIVIDADES DE SAÚDE HUMANA E APOIO SOCIAL	0,04
CONSTRUÇÃO	0,03
ACTIVIDADES ADMINISTRATIVAS E DOS SERVIÇOS DE APOIO	0,03
ACTIVIDADES IMOBILIÁRIAS	0,02
EDUCAÇÃO	0,02
TRANSPORTES E ARMAZENAGEM	0,02
ACTIVIDADES FINANCEIRAS E DE SEGUROS	0,01
OUTRAS ACTIVIDADES DE SERVIÇOS	0,01
ACTIVIDADES ARTÍSTICAS, DE ESPECTÁCULOS, DESPORTIVAS E RECREATIVAS	0,01
CAPTAÇÃO, TRATAMENTO E DISTRIBUIÇÃO DE ÁGUA; SANEAMENTO,GESTÃO DE RESÍDUOS E DESPOLUIÇÃO	0,01
AGRICULTURA, PRODUÇÃO ANIMAL, CAÇA, FLORESTA E PESCA	0,01
ADMINISTRAÇÃO PÚBLICA E DEFESA; SEGURANÇA SOCIAL OBRIGATÓRIA	0
INDÚSTRIAS EXTRACTIVAS	0
ELECTRICIDADE, GÁS, VAPOR, ÁGUA QUENTE E FRIA E AR FRIO	0

Avg Total Value Donate Per Donor	*Avg Total Number of Contributions Per Donor*	*Avg Number of Days Since Donor Last Contribution (Recency)*	*Median Number of Days Between Contribution (Frequency)*	*Avg Purchasing Power*	*Avg Number of People per Household*	*Avg Education Score*
1358,32	2,31	1764,72	-1	128,8	1,77	2,91

Cluster 3

	Rel Freq. of Donors ***per Region***
Vale do Tejo	0,31
Entre-Douro-e-Minho	0,21
Lisboa	0,15
Beira Litoral	0,12
NOT FOUND	0,05
Alentejo	0,05
Algarve	0,04
Beira Interior	0,03
Madeira	0,03
Trás-os-Montes e Alto Douro	0,02
Açores	0,01

	Rel Freq. of Donors ***per Gender***
Feminino	0,57
Masculino	0,43
other	0

	Rel Freq. of Donors Registered as AMI_friend

Avg Total Value Donate Per Donor	Avg Total Number of Contributions Per Donor	Avg Number of Days Since Donor Last Contribution (Recency)	Median Number of Days Between Contribution (Frequency)	Avg Purchasing Power	Avg Number of People per Household	Avg Education Score
111	1,72	2118,04	-1	120,28	1,78	2,85

0	0,94
1	0,06

	Rel Freq. of Donors ***per Registration Source Groups***
TMN CAMPAIGNS	0,38
OTHERS	0,2
SPONTANEOUS REGISTRATION	0,18
AMI RESOURCES/DEPARTMENTS	0,09
AUCHAN CAMPAIGN	0,04
EMERGENCY CAMPAIGNS	0,03
NATIONAL CAMPAIGNS	0,02
CHRISTMAS CAMPAIGNS	0,01
CORPORATE PARTNERSHIPS	0,01
INTERNATIONAL CAMPAIGNS	0,01
DONORS	0,01
TRAINING	0,01
ENVIRONMENT	0

	Rel Freq. of Donors ***per Company Field Description***
individuals	1