MGI

Master Degree Program in

Information Management

The use of Power BI as a Supportive Tool in the Calibration Process of Customer Risk Rating Models

Pedro Almeida Barrocas

Dissertation

presented as partial requirement for obtaining the Master Degree Program in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

THE USE OF POWER BI AS A SUPPORTIVE TOOL IN THE CALIBRATION PROCESS OF CUSTOMER RISK RATING MODELS

By

Pedro Almeida Barrocas

Master Thesis presented as partial requirement for obtaining the Master's degree in Information Management, with a specialization in Business Intelligence and Knowledge Management

Supervised by

Pedro Maia Malta

February, 2024

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledge the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Pedro Almeida Barrocas

Lisbon, 28th of February, 2024

ACKNOWLEDGEMENTS

I want to express my gratitude to my parents and brother for always being there for me on this journey, their belief in me has been my greatest motivation.

Additionally, I could not have achieved this final work without the support and endless availability of the teacher Pedro Malta, whose guidance and expertise were fundamental for taking me through this process.

I would also like to mention all of the NOVA Information Management School members, being them professors, staff, or colleagues, for providing me with knowledge over the past five years.

Finally, I am deeply grateful to my family, friends, and girlfriend for the strength, unwavering support, and motivation they transmitted to me on my academic journey, particularly on this master's thesis development.

This thesis is a testament to the collective support and encouragement I have received from my family, friends, and academic mentors. Your unwavering support, encouragement, and belief in my potential have been a constant source of strength and inspiration, and I am humbled by the opportunity to have undertaken this research.

ABSTRACT

Money Laundering is a crucial topic in the financial sector, and it is necessary to have the right tools to combat this problem in our society. The Customer Risk Rating model is a technique that can impact and reduce the exposure of cases of ML. With all the increments in machine learning and data visualization, it is possible to improve the efficiency and effectiveness of these models. Studying how a specific tool can impact the calibration process of weighted models is crucial for keeping this anti-money laundering technique updated and adapted to each problem faced. The main objective of this study is to develop and provide dashboards with a comparison of results between two different visions, where one is based on real-time choices relative to the weights applied in the model. This solution will allow the presentation of a dynamic report with relevant statistics about the distributions and characteristics of the clients. Plus, it will improve and reduce the time of the decision-process of weights, since it will present the impact of each choice in real time. In this study a specific dataset was used, but the purpose of this solution is to be adaptive to new models or new risk factors in order to not be dependent on a specific model. The method of constructing the dashboards followed the Methodology presented in this thesis, and it took into consideration the principles of Data Visualization. Once the development of all the dashboards was complete, it was possible to evaluate the results obtained. This study can provide results that will guide further research regarding the calibration process of CRR Models.

KEYWORDS

Data Visualization; Anti-Money Laundering; Customer Risk Rating Models; Dashboard; Weights; Parameters; Power BI

INDEX

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

ML – Money Laundering

AML – Anti Money Laundering

DV – Data Visualization

CFT - Countering the Financing of Terrorism

GDP – Gross Domestic Product

FATF - Financial Action Task Force

AI - Artificial Intelligence

1 INTRODUCTION

Money laundering has been a threat and has been present in our society for many years, and that is why organisations are investing in techniques to prevent and identify these situations. We can define it as the process of concealing the real source or origin of illegally obtained assets/money to make it look as if they were obtained through legitimate sources. It is a complex process that can involve several transactions or activities, making it difficult to trace its origins.

Nowadays, financial institutions and governments have decided to invest in anti-money laundering techniques to combat this critical crime, but it is still very difficult to prevent all the situations that can happen. According to FATF, new technologies seek to improve the speed, quality, efficiency, and cost of some AML/CFT measures, as well as the complexity of implementing the AML/CFT framework, compared to the use of traditional methods and processes.

An estimated USD 2 trillion is laundered annually through the global banking system, translating to roughly 2-5% of the global GDP (Cameron, 2021).

Data Visualization is present in several areas of our society, and it is fundamental to improve the efficiency and effectiveness of Data Analysis. With techniques of DV, it is possible to perform several reports based on large quantities of data. In the specific area of AML, customer risk rating models are one of the tools that lead to more data in the area of Compliance. Using client risk rating models is one of the primary techniques that financial institutions use to detect money laundering. Unfortunately, while operationalizing client risk rating models, financial institutions often face challenges relating to inaccurate risk ratings and the resulting increased level of effort (Cameron, 2021).

The advantage of DV in this topic is its capacity to lead with large amounts of data, perform visualizations with the results of the analyses, and, specifically in this area, help the calibration process of a Customer Risk Rating Model and its efficiency and accuracy in the results. With the challenges and complexity presented in this topic, improving the quality and efficiency of CRR Models seems to be a relevant topic for the financial institution market. The result of "Analysing the use of Power BI as a supportive tool in the calibration process of CRR Models" seems to be aligned with reducing the issues in this topic as well as useful for forward applications in the AML area.

It is expected with this study to analyse the data collected for calculating the risk of a client and perform analysis and visualizations that can help the calibration process of a CRR Model. Since Power BI is one of the most powerful tools for Data Visualization it is expected to replicate the logic of the model in the software in order to achieve precise and correct results. By presenting the results in this software, it would make the process of discussion the methods and parameters used in the CRR Model easier.

Finally, I also hope to contribute to the literature review on this topic and present concrete results for future studies.

2 LITERATURE REVIEW

In this topic, all the introductions and necessary notions will be explained for a better understanding of the thesis. It will be composed of some theoretical concepts and some legal obligations, as well as technical concepts about Data Visualization and Customer Risk Rating Models. This chapter will explain the topics that connect the area of DV with AML techniques, and how both areas can contribute together to the efficiency and effectiveness of CRR Models, which consequently will increase the accuracy of identifying ML situations.

2.1. Business Concepts

This first group is where most of the topics will be introduced and defined. It is fundamental to give a better context about the main subjects of the study and the impact that this thesis will have in the area.

Money Laundering:

Money laundering is the process of converting large sums of money obtained through criminal activity, such as drug trafficking, into funds originating from a legitimate source (Financial Crime Academy, 2023).

The proceeds mentioned clearly can be very diverse. However, generally, it consists in three stages, which are:

Placement:

During the initial placement stage, illicit funds or money acquired through illegal methods are introduced into a legal financial transaction system. Different financial instruments, such as cash deposits and wire transfers, are utilized to transfer the funds away from any direct connection to the criminal activity.

Layering:

In the subsequent stage, legal funds are "layered" with the illegally obtained money that was previously placed into the financial system. The aim during this phase is to “hide” the financial trail of the combined sum, which can resemble the actions of a criminal engaging in cross-border trading of stocks, commodities, and real estate.

Integration:

The concluding phase of money laundering is attained once the "dirty" money and "clean" money have been merged to the extent that all funds appear legal. When the criminals can provide a legitimate rationale for the financial assets that were placed and layered, the funds can be obtained from their original illicit source using inconspicuous means. This permits the criminals to use the funds freely within the conventional monetary system without drawing attention to their illegal origin.

Implications of Money Laundering:

The negative implications of money laundering in our society can be grouped into some categories, as will be described below.

In terms of economic impact, Money laundering goes hand-in-hand with tax evasion and duty evasion, which is the non-payment of import and export duties by smuggling goods in and out of a nation. Such activities deprive public service departments of important revenue sources (Santosh & Mohanty, 2011).

Furthermore, money launderers tend to transfer and reroute their funds across various asset classes, generating a cycle of financial transactions that conceals the illicit source of the money.

Money launderers conceal their illicit funds and place them in locations where the risk of being detected is minimal, in order to use the funds to generate profits. Such logic-defying investment activities have the ability to destabilize the overall financial system by weakening investor confidence due to unnecessary volatility, especially when the fund inflows and outflows of these laundering activities are disproportionately high (Santosh & Mohanty, 2011).

Money laundering activities can have further detrimental effects on a nation, such as the escalation of interest rates and instability in foreign exchange due to the unpredictability of irregular capital flows. Legal Impact Money laundering and criminal activities form a vicious cycle. Several large-scale illegal activities such as arms dealing, organized crime, terrorist financing, as well as drug and sex trafficking, do not just drive money laundering but thrive on it.

A strong correlation usually exists between countries with inadequate anti-money laundering regulations and the high incidence of illegal and criminal activities related to money laundering. Nonetheless, in some countries where anti-money laundering regulations have deliberately been relaxed to attract capital, the adverse consequences of heightened illegal activities may not be experienced inside of those societies but could potentially facilitate and increase illicit activities in other nations. The social impact of strong illegal businesses includes increased drug addiction, rampant corruption, and criminals empowered with economic powers (Santosh & Mohanty, 2011).

Anti-Money Laundering:

Anti-Money Laundering can be defined as a combination of measures and regulations to detect and prevent illegal origin and use of funds.

To identify and notify users of any potential money laundering schemes, the typical first generation of anti-money laundering systems used an if-then, rule-based approach. These rule-based methods examined data for possible instances of questionable money laundering based on user-generated scenarios that combined monetary thresholds with known laundering trends. However, as the sophistication of money laundering evolved over the years, conspirators began outwitting these solutions and the need for second-generation anti-money laundering solutions emerged (Santosh & Mohanty, 2011).

AML procedures require a multifaceted strategy that enlists the assistance of governments, financial institutions, and other stakeholders. Governments are crucial in the creation and enforcement of anti-money laundering legislation and regulations, as well as in advising and assisting financial firms. The implementation of AML procedures, including client due diligence, transaction monitoring, and customer risk rating models, is the responsibility of financial institutions.

In order to combat these types of crimes, anti-money laundering techniques have gained significant importance. That relevance is also related to the application of data visualization techniques since their capacities contribute to improving the quality and effectiveness of the analysis and data presented.

Customer Risk Rating Models:

Client risk rating models typically classify the client into a certain risk category (e.g., low, medium, high) based on a set of client risk factors, such as source of wealth, source of funds, client domicile, transaction behaviour, complex ownership structures, high-risk industries, negative news, wealth volume, and politically exposed person status (McKowen, 2023).

According to the FATF 2006 report, it is expected that financial institutions know their customers before establishing business or relations with them, to avoid situations of providing services to criminals and fraudsters, or even more importantly, to not be part of financing activities of terrorism or ML.

The moment of collection of this information is when the first interaction is established between the client and the bank, which means when an account is open. However, it is necessary to keep updated that information to also measure the level of risk for the client every time that the model runs. However, most of financial institutions do not have automatic processes for requesting information from the clients, which represents a problem for the performance of the CRR Models.

One of the key responsibilities of the banks in fighting ML according to the research of De Koker, (2006) is Customer Due Diligence (CDD) and Know Your Customer (KYC). It was explained that these actions can be done by these organisations if their employees receive regular training in dealing with the characteristics of the customers and their transactions. By doing this, they will create awareness for reporting to the authorities situations that represent an action of money laundering. In conclusion, together, banks can easily collaborate in the fight against money laundering.

Inside each bank, and particularly in the compliance department, compliance officers use the information presented in the system of the bank about the customers, to decide the level of risk associated with dealing/establishing relations with customers, and provide the final advice on whether they must or not accept a client as their customer and which services they should allow him to use it (Favarel et al, 2008). To help this process, it is relevant to have a CRR Model, for providing a group of clients that should be evaluated by these officers, the high-risk clients, and also to exclude clients that are not relevant or do not have characteristics that need to be analysed, which means, medium and low-risk clients.

In line with what was mentioned above, KYC has relevance in mitigating the risk of customers, by being a tool to identify situations where, for example, customers who are PEP (Politically Exposed Person), who need Customer Due Diligence to deeply understand their wealth source.

It is the responsibility of the banks to have the necessary resources for verifying the documents and information given by the clients, in order to avoid situations of fake IDs or fiscal documents that can hide the real identity of the customer.

These models can represent a fundamental tool for identifying potential cases of money laundering. It allows financial institutions to group their clients by different levels of risk, and, in necessary cases, apply due diligence in order to prevent illegal activities. While going through the logic besides a CRR Model, it is important to understand which risk factors can be relevant to include in the model (KYC information). To guarantee a correct result performed by the model the data needs to be in the right format and without incoherencies to get a precise final score.

The key benefits of deploying such methodologies include reducing 'noise' to focus on real high-risk customers and transactions, decreased operational costs due to the smaller number of customers for high-risk review, a better capture rate of bad customers, enhanced efficiency of AML processes, and ultimately, improved customer experience (Cameron, 2021).

Data Visualization:

Data visualization is the practice of translating information into a visual context, such as a map or graph, to make data easier for the human brain to understand and pull insights from. The main goal of data visualization is to make it easier to identify patterns, trends and outliers in large data sets (Brush & Burns, 2022). Regarding the principles of DV, several articles mentioned different guidelines that should be followed in order to create compelling data visualizations (Hammond, 2023):

1. Know the audience;
2. Keep things simple;
3. Use the right chart type;
4. Use colors wisely;
5. Highlight the most essential information; and,
6. Avoid clutter.

The increase in data visualization techniques is also connected with the mechanisms applied in AML/CFT. The area of data visualization has exponentially increased, justified by the investments in AI tools and the development of data analysis. These powerful techniques applied in the areas that lead to large amounts of data have a recognized impact.

In the area of Data Visualization, several graphs and visualizations are used to present a correct and dynamic vision of the results performed by a model/measures/formulas. DV can be applied to AML to evaluate massive amounts of financial data, spot trends, distributions, scores and pinpoint abnormalities. This can decrease the possibility of misunderstanding or wrong analysis while also increasing the effectiveness and accuracy of AML measures.

The area of data visualization improves the capacity to analyse real-time data and increases the performance of dynamic reports based on clients' or users' choices, which can lead to a faster interpretation of the results, as well as clarity and objectivity in the analysis performed.

2.2. Technical Concepts

Client risk rating models typically classify the client into a certain risk category (McKowen, 2023). Typically, it uses categorical categories like Low, Medium, and High risk for the different levels of risk. The complexity of the model can influence the definition of these levels, where usually models that have more variables or more steps of calculation can evaluate the clients in a more detailed and precise way. Most AML models are overly complex. The factors used to measure customer risk have evolved and multiplied in response to regulatory requirements and perceptions of customer risk but still are not comprehensive. Models often contain risk factors that fail to distinguish between high- and low-risk countries, for example (Mikkelsen et al., 2019). These types of models can receive input data from different variables. However, it depends on each industry or specific case to identify which areas need more focus based on the universe of clients and their characteristics. In the specific business line of banks or financial institutions, normally these models are created for analyzing the exposition of creating a relationship with a certain client.

Normally, regulators require that financial institutions perform anti-money laundering and Know Your Customer risk assessments (AdvisoryHQ, n.d.). With these tools, it is possible to calculate the exposition and risk associated with money laundering.

In order to apply these models, financial institutions can use external platforms or internal software as a tool to develop them. Several solutions offer tools to create models or to use already developed and tested demos in the real market. For this study, it was analyzed some solutions that consulting companies offer in the business context of AML/CDD solutions. Since the focus was not on the model, but the calibration of one, it was analysed what are the usual requirements and implementation steps of these types of models in order to perform one that fits the necessities of the supportive solution that will be created. To decide which factors could be relevant to apply in the model, it was analysed the usual information that financial institutions collect regarding KYC information, and which information is mandatory concerning the obligations provided by the National Bank or FATF. In this step of collecting information, the Financial Action Task Force (FATF) recommends that where firms cannot apply the appropriate level of CDD, they should not enter into the business relationship, or should terminate the business relationship (Cameron, 2023).

According to diverse studies, there are four main pillars to consider in terms of variables in the CRR Models. The first one is the customer risk, like for example, the occupation or professional activity of individual clients, corporate entities usually consider factors like, for example, the economic activity or if the company does not have a beneficial owner. The second pillar is the geographic risk, like the citizenship, residence or where a company has its headquarters. The third one is the type of product or service (type of account) required by the client. Lastly, transactional behaviour, which just can be considered in an ongoing vision since it takes into consideration the transactional history of the client. This last pillar can be relevant in order to identify potential money laundering situations like, for example, transfers to countries evaluated as high-risk because legal obligations are less restrictive (Ibitola, 2023).

Nowadays, the most common form of the customer risk rating model is a score-based risk rating model (Ratanachu-Ek, 2019). For developing this study, it was decided to consider a model that used a weighted risk scoring logic, because it is still being used by several institutions and it can have some issues of optimization in the calibration process. This was a requirement assumed since the original purpose of this study was to create a tool that could be used by consulting companies in different financial projects developed. Since it was not possible to use an actual model developed by any company, it was analysed several logics developed in different studies in order to create a model that could be used in this thesis. Firstly, to access the data and the respective score was used the method mentioned in the KYC_2.0_Risk_Assessment_Guide by Oracle, particularly in the topic of risk assessment parameters. It explains how to access each parameter and the respective risk score regarding the dataset considered. The process of accessing the scores attributed to each risk factor is based on the type of variable, where, for example, for the citizenship the method of collect considers the max score between the first and second citizenship. By achieving the code of the country, it is used a lookup table that takes into lookup value the code or a continuous value for giving the respective score. For numeric variables, as variables like relation duration, a different method is followed. It is based on creating intervals of values with the respective score and the lookup is done regarding the value observed and the interval where it places (Oracle® Financial Services Know Your Customer Risk Assessment Guide, n.d.).

The dataset used in this study had random codes or numeric values and the column with the score was also already filled, so it was not necessary to create this lookup table with the respective score for each observation. After having the dataset in the right structure, it was necessary to investigate and present the logic used for the weighted rating model.

This study used the same logic as was mentioned in the article "A Framework to Score the Risk Associated with Suspicious Money Laundering Activity and Social Media Profile" (Parida & Kumar, 2020) about the definition of weights and parameters. Basically, each risk factor created has a respective weight and the value of the weight is used in the formula of calculation the risk. For calculating the final score, it is obtained by summing all the multiplications between the scores and each weight defined.

$$Potential\,Risk = \sum_{i=1}^{n} (W_i \times \; R_i)$$

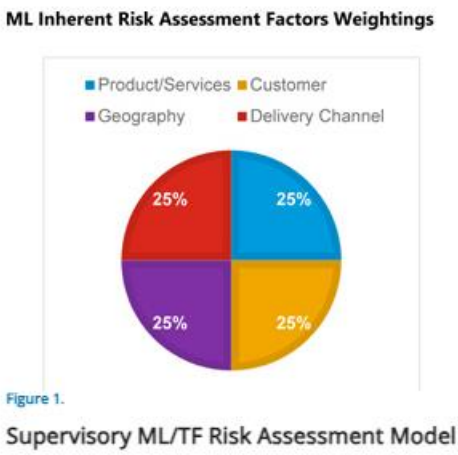
Where Potential Risk = Total risk score of customer
Wi = Risk weightage for each filter i = I,..., n
Ri = Risk assessment score for each filter i = I,..., n
n = Number of risk assessment filter

Source: This figure was extracted from the article "A Framework to Score the Risk Associated with Suspicious Money Laundering Activity and Social Media Profile" published by Parida & Kumar in 2020.

Figure 2.1 - Method of Calculating the Risk

Both legal entities and natural persons undergo a rigorous CRRM process where anonymity, business relationships, and country-based risk factors are scrutinized. Since different attributes must just be analysed for individuals' clients and others for corporate entities, it was necessary to create two separate models for each one of the types of clients (Customer Risk Rating Methodology (CRRM) - Anti-Money Laundering, 2023).

As mentioned, since there are four pillars of customer risk rating models, they were considered in the calculation of the final score. The final risk was assumed by the formula in Figure 2.1 inside each group of risk factors, and by summing all the scores of each group of risk factors based on a weighted model. This step was developed based on the Figure 2.2, where a risk assessment model is presented. It was followed the same logic of distributing weights per group of risk factors in order to give more importance to specific groups of risk (Iceland: Financial Sector Assessment Program-Technical Note on Anti-Money Laundering/Combating the Financing of Terrorism, 2023). In the end it was applied a combination of the mathematical formulas mentioned in Figure 2.1 and Figure 2.2. This logic is also presented in the article “AML KYC Risk Rating Assessment Template, Methodology, Rating Matrix” where different weights are also applied to different risk factors/groups (AdvisoryHQ, n.d.).



Source: This figure was extracted from IMF Staff Country Reports 2023, 277.

Figure 2.2 - Risk Assessment Model

As will be mentioned in the next steps, the mathematical formula was applied to each group of risk factors, and then the weights per each group are distributed to calculate the final risk. This step was added to the model in order to include the impact of the behavioural risk for clients that are not considered in the onboarding moment, which means, it is just applied to clients that already have relationship with a bank (Ibitola, 2023).

It was created a model with variables that were in the dataset used and that are usually incorporated in the CRR Models. Since their nature is also related to the four pillars mentioned, the variables were grouped into four categories, and they were selected based on the available ones presented in the dataset.

It is important to mention that the mathematical formula used and the logic besides the model were created as an illustrative example without any techniques of optimization or feature importance. Additionally, the tool presented in this master thesis was developed to support the calibration process of all the weighted models that can be created following a similar structure of the one considered.

To calculate the risk associated with each client, different risk factors that characterise the clients and the business relationship were considered and grouped into four Risk Groups: i) Country, ii) Entity/Customer, iii) Product, and iv) Behavioural (Cameron, 2023). As it was already mentioned, the factors considered in each CRR model change according to the type of client, i.e. Individuals and Corporate Entities.

It will be possible to index a certain weight to each Risk Group in order to adjust its contribution to the final risk level. Likewise, the risk factors included within each Risk Group are also multiplied by weights whose sum totals 100% (Parida & Kumar, 2020).

Regarding what was mentioned about including the fourth pillar in the calculations of the risk, it was identified as a relevant aspect of this study to include the analysis of clients that are already on an ongoing vision, which means, the calculation risk for the clients already presented in the institution. As well, an ongoing customer relationship for a certain length of time, such as three to five years or more, is also a mitigating factor (Alessa, 2023). With this information, there is a column in the dataset that divides the universe of clients in Onboarding or Ongoing, so it was considered that column to apply different models regarding the indicator verified. The clients ongoing are identified by their business relation with an institution, that can be defined regarding different intervals. Since the information in the dataset was already with this indicator it was not necessary to perform calculations in order to divide the universe of clients (Central Bank of Ireland, 2021).

In short, the changes made to the model between onboarding and ongoing are just related to the transactional history, which characterizes the Behavioural Risk. When applying the onboarding model, the Behavioural Risk group will not weigh in the risk calculation, thus increasing the weights assigned to the other risk groups.

Based on the paper of SAS, there are also heuristic rules variants used in these types of models that represent an attribution of a specific score if a client accomplishes one of the rules defined (Rivera, E., West, J., & Suplee, C. (2015). Usually, these rules are applied in situations of more exposition to potential ML. In order to include also this variant in the model defined, the following heuristic rules were created regarding the information presented in the dataset:

- For Individuals, if the client is identified as a Politically Exposed Person ("PEP") or is associated with Suspicious Activity Reports ("SARs") or has Residence Permit for Investment (“ARI”) the risk is automatically set to high (auto high).
- For Corporate Entities, if a client is identified as a Correspondent Bank, if it has no associated Ultimate Beneficial Owner ("UBO"), or if it is associated with a SAR, the risk is automatically defined as high.

The final risk calculated by the CRR model should obtain a final score (i.e., continuous value) which will be classified qualitatively into three risk levels (i.e. low, medium and high) through value ranges.

All the values parametrized must work as a support for the calculation of the final risk and it must have functional purpose and meaning. To define the right score to attribute to each observation, usually, it is consulted the exercises of Risk Assessment, where several aspects are evaluated, and a final score is defined (Cams, 2024).

For the parametrization of the countries scores, it was taken into consideration some research that can connect the countries to possible illegal activities, one of the examples used was the latest version of the list of high-risk third countries provided by the European Commission (High Risk Third Countries and the International Context Content of Anti-money Laundering and Countering the Financing of Terrorism, n.d.). Since the data does not present business meaning or real information because it is based on random codes and scores, the rest of the parametrizations defined did not have any research associated.

Risk Assessment Score is calculated for a particular Risk Assessment Filter and to be anywhere between 0 to 100 (Parida & Kumar, 2020). The scores defined were based on a scale of 20 – Low Risk, 40 – Medium Risk and 100 – High Risk.

For the Behavioural Group, normally some measures are created as risk factors. For example, it can be analysed transfers to high-risk countries as was already mentioned, or, for example, it can be identified relevant patterns that can associate a client to suspicious transactions. (Cameron, 2021).

Since the dataset does not have transactional information, it was assumed a fixed score for this group as it will be possible to define in the Demonstration Chapter.

2.3. Final Remarks

Considering the information presented, as well the importance of the CRR Model as an anti-money laundering tool it is fundamental to have a calibration process of parameters well defined and adapted to specific necessities.

With the help of Data Visualization, it will be possible to improve the process of calibration of the parameters and present in real-time results updated and aligned with the choices applied.

3 METHODOLOGY

This master thesis aims to evidence the supportive capacity of data visualization and data analysis in the calibration process of a customer risk access model, particularly using the Power BI tool. The methodology decided to use was the Design Science Research Methodology (DSRM) (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2008). This methodology is based on six steps, which will be enumerated and explained in order to show how this methodology fits this problem and provides guidance to achieve results. The methodology chosen was based on the problem that this study is facing, and it takes into consideration the interactivity between each one of the steps, which means, the possibility of cycling back to earlier activities, in order to achieve complete results.

In the first moment, it will be explained the problem and the importance of developing this study, which means, it will go through the CRR Model and present the benefits of developing a tool that supports the calibration process and how to improve the efficiency of the model regarding the characteristics of the clients supported by real-time results.

Secondly, the objectives of this study will be pragmatically defined to conclude and measure the efficiency and quality of the solution developed. In this specific chapter, it was aligned what are the main topics to cover by developing this supportive tool.

After defining the objectives and requirements, the design phase will be presented being also connected with prototyping the solution. This chapter is where the initial steps were taken in the software to understand if some limitations would be faced.

For analysing the impact of this study, it was used a problem situation defined to demonstrate the utility of this tool. After developing this phase, it helped to complete the prototyping of the solution, since some analyses were identified as relevant to present in this chapter.

In conclusion, the evaluation and communication phases are where the results will be detailed, and the feedback obtained will be presented regarding the analysis of the final solution.

3.1. Identify Problem and Motivate

As it was already mentioned before, the purpose of this research is to present a solution that improves the efficiency and effectiveness of the calibration process concerning the Customer Risk Rating models. Since the model is defined by a mathematical formula, the process of calibration of parameters can be hard and exhaustive for present results that fit the necessities of the client. Concerning the dimension of the bank's clients, some requests are usually made in order to present an output that fits the capacity of maintenance and sustainability of those models. For example, if a bank has more than 1 million clients it is expected a distribution by each level of risk different than a bank that has 30.000 clients.

Facing these challenges of fitting and fulfilling the necessities of each client can represent a difficult process because the basis of the mathematical formula applied in the models is the same for all the weighted models. However, the calibration of the parameters and the process of choosing the risk factors can be done by analysing the necessities presented by the client. Most of the risk factors must be considered because of the ones defined by FATF, which can influence and contribute to similarities between the models implemented. However, banks can identify some specific risk factors that improve the accuracy of ranking the clients in the different risk levels. For answering the necessities of calibration, it was important to have a tool that could support the decision process of giving each parameter the right weight in the mathematical formula of the model, to achieve a distribution of clients by each level of risk aligned with the expectations of each client.

The reality of applying a heavier weight to one parameter instead of distributing the same weight to all the parameters can be difficult to measure. Additionally, in terms of the distribution of scores of each risk factor can be difficult to understand the impact of applying a certain combination of weights in the final risk of the clients. That is the reason why sometimes the calibration process of a CRR Model is not efficient or aligned with the distributions analysed.

Developing this tool will be a resourceful and adaptative solution that can be applied in future projects provided by consultant companies or directly considered by financial institutions. Since the logic behind the model is the same for most weighted models, also this tool will complement and make the models even more robust and aligned with the expectations of the clients.

3.2. Define Objectives of the Solution

As providing precise and supported results is fundamental and a requirement in the AML area, by developing this solution it will be possible to create different variations of results performed by the Customer Risk Rating model. The solution has the objective of comparing and analysing the results of two variations based on the definition of each parameter utilized in the model. Concerning those two views, the report will be able to present in real-time how changing the weight of one or even all the parameters can influence the final distribution of the clients regarding the scale defined. Also, the different analyses presented in the final solution can support the decision of the scale that defines each level of risk.

The powerfulness of this tool is to have a dynamic variation of the parameters that are used in the model, and show, by comparing with a static variation, how the decision of combining different weights can influence the final score of each client.

The opportunity to apply this tool in different models that have the same mathematical logic gives a huge capacity for utility and optimization of the services provided. Another objective of developing this solution is to improve the precision of the results and provide the opportunity to decide with a client, in real-time, which are the parameters that make more sense to apply regarding the clients' characteristics and the reality of the bank in terms of reviewing alerts generated by the high-risk clients when running the CRR Model. By providing this solution to different clients, the adaptive capacity of the solution will be tested, since it will be possible to identify the aspects that are relevant to each one of them. Plus, it will evidence how it supports the calibration process by presenting relevant patterns and analyses that justify the decision to apply the right weights to each parameter.

Concluding, this study aims to provide a complete tool that improves the process of defining each parameter in the mathematical model of the calculation of risk of clients, using a dynamic and real-time updated tool.

3.3. DESIGN AND DEVELOPMENT

After identifying the problem and the introduction of the solution, it needs to be developed the final tool that fits all the points mentioned above. In order to present a solution that can contribute to different problems, there was the necessity of identifying what would generally be useful for several potential clients and analyse the relevant topics to include in the PowerBI Dashboards. To do it, it was created the structure of the solution for answering the following points:

- What are the risk factors that have more influence on the final score;
- How is the distribution of clients by each strategic rule defined;
- Which risk factors should have heavier weight;
- How the behavioural group can impact the final score of each client; and,
- What is the distribution of clients by each level of risk.

During this process was possible to understand what the expectations were for developing this solution and how to proceed in the prototype phase.

This study was also developed based on my professional experience, but it was not possible to apply real data into this process. However, as was mentioned synthetic data was used in order to show the capacities of the tool and also to take the necessary conclusions of this study.

The data used has randomly generated information about 40.000 clients, with information for each one of the risk factors presented. This dataset does not have information regarding any client or confidential information since the information does not have a business meaning.

Regarding the logic besides the CRR Model, the method to select the groups and the risk factors for this study took into consideration several articles presented in the literature review and also the information available in the dataset. The final list defined is composed by the ones presented below.

For Individuals clients, it was decided the following list:

- Country Group:
    - Citizenship; and,
    - Residence.
- Entity Group:
    - Adverse Media;
    - CAE (In the case of being an individual entrepreneur);
    - Occupation;
    - Professional Activity;
    - Business Duration Relation;
    - SAR;
    - PEP;
    - ARI;
    - Channel; and,
    - Opening Purpose.
- Product Group:
    - Product/Service Associated;
- Behavioural Group.

For Corporate Entities, the risk factors considered were:

- Country Group:
  - Headquarters.
- Entity Group:
  - Adverse Media;
  - CAE;
  - Constitution Date;
  - Business Duration Relation;
  - SAR;
  - Correspondent Bank;
  - No UBO;
  - Channel; and,
  - Opening Purpose.
- Product Group:
  - Product/Service Associated;
- Behavioural Group.

The respective mathematical formula to calculate the final risk it is achieved by the formula presented in the Figure 2.1 of the chapter Literature Review and it will be presented in the prototype phase of this study.

Regarding the risk factors chosen, some analyses were defined as relevant to present in this report. As this methodology allows, in the moment of prototyping the tool, additional analysis and visualizations were identified as fundamental in order to support the decisions taken for each one of the variations. In conclusion, some of the points that were defined as relevant to accomplished were the following ones:

- Distribution of the clients by each level of risk, type of model and type of client;
- Average of final scores by each level of risk, type of model and type of client;
- Average of scores per group of risk factors by each type of model and type of client;
- Distribution of the Auto-Highs by risk factor;
- The possibility of changing each weight of the parameter by risk factor and group;
- The possibility of including the transactional behaviour for calculating the impact in the final risk; and,
- Several analyses regarding the characteristics of the clients.

Concerning the structure of the report, since this study was developed for also future applications, it was agreed that some fundamentals needed to be followed:

- A resume page with the “big numbers” and the distribution of the metrics calculated for each vision presented.
- Two pages with detailed analysis regarding each one of the visions;
- One page for each type of clients in order to change the parameters of each group/factor risk;
- Controls for giving the possibility of each person of the team or the client to change the parameters and not achieve wrong results;
- A vision without the application of the strategic rules, for calculating the impact of having the Auto-Highs in the distribution of the risk; and,
- Distribution of scores in each risk factor.

3.4. Prototyping the Solution

After defining the requirements considered relevant for the construction of the solution, the still main objective was to have the possibility of changing each parameter used by the model and replicating the mathematical logic of the model in PowerBI.

The dataset used, as was already mentioned, has information regarding 40000 clients with information for each one of the risk factors. After analysing potential integrity or quality problems, it was possible to conclude that the data in general was in the right format for application in the model. However, some steps were applied in order to avoid problems in the performance of the model. In the area “Power Query Editor”, besides the mandatory step of defining the source, the following steps were applied:

1. Source: The dataset was in CSV format, and it was used the encoding 65001.
2. Promoted Headers: For defining the headers of each column with the values presented in the first line of the dataset, was applied this step.
3. Changed Type: Since after uploading the dataset all the variables were in text format, some changes were needed. All the variables that represented the score attributed to each of the risk factors were converted to integers (Int64.Type).
4. Replaced Value: Some scores were defined as “NULL”. Since this column needed to be in the integer format, it was replaced the NULL by the High Score defined (100 points).

The data model view will be added in the annexes as it was not relevant to include in this chapter, since it was not created a relational model, because all the data was in one dataset and in one external file that has the parameters for the baseline view.

After defining the source of the data, it was possible to prototype one of the requirements defined, which was, to present several statistics about the information and distribution of scores by each risk factor. This page has relevance for the decision of weights since it is where the characteristics of the clients can be analysed. It is important to mention that all the parametrizations were already presented in the dataset, and just the scores of countries of citizenship and residence were reviewed based on personal experience and articles analysed. Those parametrizations were not presented in this study since each code does not have a business meaning. It can be adapted in future projects, since its purpose is to identify clients with characteristics that can relate to illegal activities, such as country of nationality, occupation, the product account that the client requires, etc. Even though this study faced the limitations already described, the scores defined are based on the scale already defined: 20 points for lower risk, 40 points for medium risk and 100 points for high risk.

Concerning these parametrizations, each parameter of the risk factor will be multiplied by the score evaluated and contribute to the final score as was already mentioned in the mathematical formula of the model.

As it is possible to analyse in Figure 3.1, some cards were utilized for presenting the number of clients associated to strategic rules, and in the bar charts it is possible to analyse the distribution of clients per score by each risk factor:

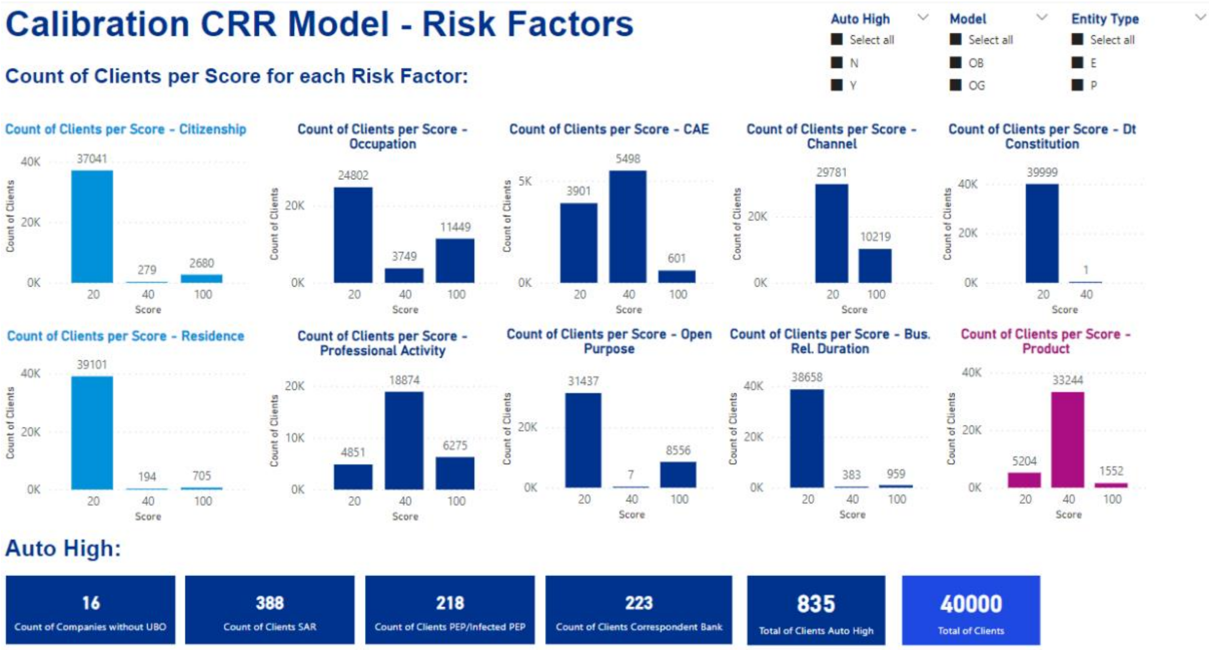


Figure 3.1 - Dashboard of Exploratory Analysis - Factors

Regarding the requirements defined, the following step was the most challenging one. Replicate the CRR Model in Power BI with the possibility of applying dynamic changes in the parameters, which represented the aim of this study and the most technical part of prototyping this solution.

For following the exact same methodology of calculating the score as it was mentioned, it was necessary to create several parameters in the report. These parameters work as variables that can be used in measures, and also the value considered for the calculation of measures can be selected in the different options of a slicer. Since there is a specific model for each entity type, it was necessary to create parameters regarding the entity type, and also for each model, onboarding and ongoing. In total, it was concluded that were needed more than 50 parameters. Since all the parameters needed to take a value between 0 and 1, the process of generation was the same for all of them, and they were created as it is possible to observe in the Figure 3.2:

Figure 3.2 – Structure of an Example of a Parameter Created

For the identification of each parameter, the same structure was used for the key names: the type of entity, the model of onboarding/ongoing, the group of risk factors, and finally the name of the risk factor.

The DAX code used for generating these parameters was the one presented in Figure 3.3.

```
P_Ongoing_Country_Citizenship = GENERATESERIES(0, 1, 0.05)
```

Figure 3.3 – DAX Formula of the Generation of the Parameter

As it is possible to check, it generates a series between 0 and 1 with intervals of 0.05. It was decided to use this interval amplitude in order to have an easier interpretation and also to limit the list of possibilities. This decision can be changed in future projects in order to adapt the amplitude of the interval and adjust the minimum and maximum values.

In order to apply the value generated by the parameter, it was used a measure as it is in Figure 3.4.

```
P_Ongoing_Country_Citizenship Value = SELECTEDVALUE('P_Ongoing_Country_Citizenship'[P_Ongoing_Country_Citizenship])
```

Figure 3.4 – DAX Formula of Accessing the Value of the Parameter

When all the parameters were created, it was possible to start the process of replicating the mathematical formula of the CRR Model using the parameters and the selected weights for each one of them.

As mentioned at the beginning of this thesis, two visions would be created: one static with the parameters defined based on professional experience and the best practices used in the area and one dynamic that it would be possible to select each weight for each one of the parameters used in the formula.

Concerning those two visions it will be now described how both were developed based on the data considered and parameters already performed.

3.4.1. Baseline

As the name describes it, this vision represents the first calculated view of the final risk done in Power BI, and it is where the parameters of each risk factor and group of risk factors were static. So, this chapter will explain how all the necessary information was calculated for presenting in the required visualizations the relevant information and analysis.

This process did not use the parameters created, since all the weights were defined in an Excel file that was used as the source of data to complement the data model. This external file has all the weights for each one of the risk factors and was defined after analysing the data and characteristics of the clients. The information used does not represent real data so it was difficult to find patterns or identify situations that could influence the decision process of attributing the right weight to each risk factor. However, one of the purposes of this study is to compare two visions and see the impact of changing one or even all the weights. For defining these weights, it was analysed the risk factors that can expose more the client, which means, it was given more relevance to risk factors that analyse the exposition of a client to lead with money or for example the citizenship associated to a country that has several cases of money laundering. It was important to distribute higher values of weights inside of the entity group associated with risk factors that analyse the professional activity, occupation, economic activity and after higher importance to risk factors related to the reason of establishing relation and the type of opening channel used. After defining the parameters, it will be possible to analyse the distribution of clients per each level of risk and take the right conclusions about the accuracy of the results.

In technical terms of developing this vision, in the table of parameters (appendixes) it is possible to see that each weight of each risk factor has a key that was used to multiplicate the score analysed. Concerning this, new columns were created with the result of the score multiplied by the parameter, giving the necessary results for calculating the final score of this static vision.

As it is possible to analyse in Figure 3.5, the new columns followed the same method of creation as it is described below:

```
score_pond_business_duration = if(and(Scores_Power_BI[Model] = "OB", Scores_Power_BI[Entity Type] = "P"),Scores_Power_BI[score_dur_rel]
*CALCULATE(max(Parameters[Parameter]), Parameters[Key] = "P_Onboarding_Entidade_Dur_Rel_Neg"),
if(and(Scores_Power_BI[Model] = "OG", Scores_Power_BI[Entity Type] = "P"),Scores_Power_BI[score_dur_rel]*CALCULATE(max(Parameters
[Parameter]), Parameters[Key] = "P_Ongoing_Entidade_Dur_Rel_Neg"),
if(and(Scores_Power_BI[Model] = "OG", Scores_Power_BI[Entity Type] = "E"),Scores_Power_BI[score_dur_rel]*CALCULATE(max(Parameters
[Parameter]), Parameters[Key] = "E_Ongoing_Entidade_Dur_Rel_Neg"),
if(and(Scores_Power_BI[Model] = "OB", Scores_Power_BI[Entity Type] = "E"),Scores_Power_BI[score_dur_rel]*CALCULATE(max(Parameters
[Parameter]), Parameters[Key] = "E_Onboarding_Entidade_Dur_Rel_Neg"),0))))
```

Figure 3.5 – DAX Formula of an Example of One Calculated Column

Additionally, for each group of risk factors, the same logic was followed, it was created a new column for each result of the total score of each group of risk factors:

```
score_pond_entity_group = if(and(Scores_Power_BI[Model] = "OB", Scores_Power_BI[Entity Type] = "P"),(Scores_Power_BI[score_pond_cae]
+Scores_Power_BI[score_pond_profissao]+Scores_Power_BI[score_pond_open_purpose]+Scores_Power_BI[score_pond_adv_media]+Scores_Power_BI
[score_pond_canal_contrat]+Scores_Power_BI[score_pond_ativ_prof]+Scores_Power_BI[score_pond_business_duration])*CALCULATE(max(Parameters
[Parameter]), Parameters[Key] = "P_Onboarding_Entidade_Macro_Ponderador"),
if(and(Scores_Power_BI[Model] = "OG", Scores_Power_BI[Entity Type] = "P"),(Scores_Power_BI[score_pond_cae]+Scores_Power_BI
[score_pond_profissao]+[score_pond_open_purpose]+Scores_Power_BI[score_pond_adv_media]+Scores_Power_BI[score_pond_canal_contrat]
+Scores_Power_BI[score_pond_ativ_prof]+Scores_Power_BI[score_pond_business_duration])*CALCULATE(max(Parameters[Parameter]), Parameters
[Key] = "P_Ongoing_Entidade_Macro_Ponderador"),
if(and(Scores_Power_BI[Model] = "OG", Scores_Power_BI[Entity Type] = "E"),(Scores_Power_BI[score_pond_cae]+Scores_Power_BI
[score_pond_dt_const]+Scores_Power_BI[score_pond_open_purpose]+Scores_Power_BI[score_pond_adv_media]+Scores_Power_BI
[score_pond_canal_contrat]+Scores_Power_BI[score_pond_business_duration])*CALCULATE(max(Parameters[Parameter]), Parameters[Key] =
"E_Ongoing_Entidade_Macro_Ponderador"),
if(and(Scores_Power_BI[Model] = "OB", Scores_Power_BI[Entity Type] = "E"),(Scores_Power_BI[score_pond_cae]+Scores_Power_BI
[score_pond_dt_const]+Scores_Power_BI[score_pond_open_purpose]+Scores_Power_BI[score_pond_adv_media]+Scores_Power_BI
[score_pond_canal_contrat]+Scores_Power_BI[score_pond_business_duration])*CALCULATE(max(Parameters[Parameter]), Parameters[Key] =
"E_Onboarding_Entidade_Macro_Ponderador"),0))))
```

Figure 3.6 – DAX Formula of the Score per Entity Group

Finally, all the results of each group were summed, and it was possible to achieve the final score for each client. It is important to mention, that this was the phase where the strategic rules were implemented, which means that if a client has one of the flags mentioned before, the final score automatically turns 500 points (High-Risk).

Below, it describes the calculation of the final score and the application of the strategic rules.

Firstly, a column was created with a flag of “Y” or “N” that indicates if a client has one of the strategic rules defined as “Y”:

```
Auto High = if(or(or(or(Or(OR(Scores_Power_BI[x_suspicious_activity_ind] = "Y", Scores_Power_BI[x_correspondent_bank_ind] = "Y"), Scores_Power_BI[x_no_ubo_ind] = "Y"),
Scores_Power_BI[score_pep] <> 0),Scores_Power_BI[x_ari_ind]="Y"),Scores_Power_BI[x_no_ubo_ind]="Y"), "Y","N")
```

Figure 3.7 – DAX Formula of the Auto High Column Created

Secondly, the final score was calculated by summing all the results obtained from each group of risk factors:

```
Final_Score_Baseline = Scores_Power_BI[score_pond_country_group] + Scores_Power_BI[score_pond_entity_group] + Scores_Power_BI[score_pond_product_group] +
Scores_Power_BI[score_pond_behavioural_group]
```

Figure 3.8 – DAX Formula of the Final Score Baseline

Lastly, it was converted to the final score, if the client had the flag “Auto High” marked as “Y”:

```
Final_Score_Baseline_AH = if(Scores_Power_BI[Auto High] = "Y", 500, Scores_Power_BI[Final_Score_Baseline])
```

Figure 3.9 – DAX Formula of the Final Score Baseline with Auto-High

After all the calculations and columns were created it was possible to calculate the final score for each client based on his entity type and model.

3.4.2. Variation

As was already described, the objective of the variation vision is to provide the possibility of analysing the impact in real-time by changing the parameters that were attributed to the baseline vision. However, it was necessary to create several measures in the Power BI solution for achieving the final objective, which means, the final score also with the strategic rules applied.

It was in this variation that the parameters created were applied. Each calculation between the score and the parameter chosen for the risk factor was created as a measure following the DAX structure of the figure below:

```
Measure Open Purp = sumx(Scores_Power_BI,if(and(Scores_Power_BI[Model] = "OB", Scores_Power_BI[Entity Type] = "P"),Scores_Power_BI[score_op_purpose]
*P_Onboarding_Entity_Prop_Bus_Rel[P_Onboarding_Entidade_Prop_Rel_Neg Value],
if(and(Scores_Power_BI[Model] = "OG", Scores_Power_BI[Entity Type] = "P"),Scores_Power_BI[score_op_purpose]*P_Ongoing_Entity_Prop_Bus_Rel
[P_Ongoing_Entidade_Prop_Rel_Neg Value],
if(and(Scores_Power_BI[Model] = "OG", Scores_Power_BI[Entity Type] = "E"),Scores_Power_BI[score_op_purpose]*E_Ongoing_Entity_Prop_Rel_Neg
[E_Ongoing_Entidade_Prop_Rel_Neg Value],
if(and(Scores_Power_BI[Model] = "OB", Scores_Power_BI[Entity Type] = "E"),Scores_Power_BI[score_op_purpose]*E_Onboarding_Entity_Prop_Bus_Rel
[E_Onboarding_Entidade_Prop_Rel_Neg Value],0)))))
```

Figure 3.10 – DAX Formula of the Score per Risk Factor Variation

As it is possible to analyse, it was used the function “sumx” for calculating the score multiplied by the parameter chosen in the slicer. This process was repeated for each risk factor in order to calculate the aggregation of scores per group of factors.

Since one of the requirements was to present some statistics regarding the average of the scores multiplied by the weight of each risk factor, it was also created a measure for each risk factor using the function "AVERAGEX" as it is presented in the figure below:

```
Measure AVG Open Purp = AVERAGEX(Scores_Power_BI,if(and(Scores_Power_BI[Model] = "OB", Scores_Power_BI[Entity Type] = "P"),Scores_Power_BI[score_op_purpose]
*P_Onboarding_Entity_Prop_Bus_Rel[P_Onboarding_Entidade_Prop_Rel_Neg Value],
if(and(Scores_Power_BI[Model] = "OG", Scores_Power_BI[Entity Type] = "P"),Scores_Power_BI[score_op_purpose]*P_Ongoing_Entity_Prop_Bus_Rel
[P_Ongoing_Entidade_Prop_Rel_Neg Value],
if(and(Scores_Power_BI[Model] = "OG", Scores_Power_BI[Entity Type] = "E"),Scores_Power_BI[score_op_purpose]*E_Ongoing_Entity_Prop_Rel_Neg
[E_Ongoing_Entidade_Prop_Rel_Neg Value],
if(and(Scores_Power_BI[Model] = "OB", Scores_Power_BI[Entity Type] = "E"),Scores_Power_BI[score_op_purpose]*E_Onboarding_Entity_Prop_Bus_Rel
[E_Onboarding_Entidade_Prop_Rel_Neg Value],0)))))
```

Figure 3.11 – DAX Formula of the Average Score per Risk Factor

In most of the calculations, it was used 0 as the else value, because it is not expected to have a NULL value as in the weight as in the score. Additionally, these measures were created for all the risk factors present in the CRR Model.

Finally, the same method was followed to calculate the aggregated score per group and the final score. As an example of the DAX syntax applied to the groups:

```
Measure Entity Group = sumx(Scores_Power_BI, if(and(Scores_Power_BI[Model] = "OB", Scores_Power_BI[Entity Type] = "P"),([Measure CAE] + [Measure Open Purp] +
[Measure Adv Media]+[Measure Canal Contrat]+[Measure Ativ Prof]+[Measure Dur Rel]+[Measure Profissao])*P_Onboarding_Entity_Macro_Weighter
[P_Onboarding_Entidade_Macro_Ponderador Value],
if(and(Scores_Power_BI[Model] = "OG", Scores_Power_BI[Entity Type] = "P"),([Measure CAE] + [Measure Open Purp] +[Measure Adv Media]+[Measure Canal Contrat]+
[Measure Ativ Prof]+[Measure Dur Rel]+[Measure Profissao])*P_Ongoing_Entity_Macro_Weighter[P_Ongoing_Entidade_Macro_Ponderador Value],
if(and(Scores_Power_BI[Model] = "OG", Scores_Power_BI[Entity Type] = "E"),([Measure CAE]+[Measure Dt Constituicao]+[Measure Open Purp]+[Measure Adv Media]+
[Measure Canal Contrat]+[Measure Dur Rel])*E_Ongoing_Entity_Macro_Weighter[E_Ongoing_Entidade_Macro_Ponderador Value],
if(and(Scores_Power_BI[Model] = "OB", Scores_Power_BI[Entity Type] = "E"),([Measure CAE]+[Measure Dt Constituicao]+[Measure Open Purp]+[Measure Adv Media]+
[Measure Canal Contrat]+[Measure Dur Rel])*E_Onboarding_Entity_Macro_Weighter[E_Onboarding_Entidade_Macro_Ponderador Value],0)))))
```

Figure 3.12 – DAX Formula of the Score per Entity Group

```
Measure Avg Entity Group = AVERAGEX(Scores_Power_BI, if(and(Scores_Power_BI[Model] = "OB", Scores_Power_BI[Entity Type] = "P"),([Measure CAE] + [Measure Open
Purp] +[Measure Adv Media]+[Measure Canal Contrat]+[Measure Ativ Prof]+[Measure Dur Rel]+[Measure Profissao])*P_Onboarding_Entity_Macro_Weighter
[P_Onboarding_Entidade_Macro_Ponderador Value],
if(and(Scores_Power_BI[Model] = "OG", Scores_Power_BI[Entity Type] = "P"),([Measure CAE] + [Measure Open Purp] +[Measure Adv Media]+[Measure Canal Contrat]+
[Measure Ativ Prof]+[Measure Dur Rel]+[Measure Profissao])*P_Ongoing_Entity_Macro_Weighter[P_Ongoing_Entidade_Macro_Ponderador Value],
if(and(Scores_Power_BI[Model] = "OG", Scores_Power_BI[Entity Type] = "E"),([Measure CAE]+[Measure Dt Constituicao]+[Measure Open Purp]+[Measure Adv Media]+
[Measure Canal Contrat]+[Measure Dur Rel])*E_Ongoing_Entity_Macro_Weighter[E_Ongoing_Entidade_Macro_Ponderador Value],
if(and(Scores_Power_BI[Model] = "OB", Scores_Power_BI[Entity Type] = "E"),([Measure CAE]+[Measure Dt Constituicao]+[Measure Open Purp]+[Measure Adv Media]+
[Measure Canal Contrat]+[Measure Dur Rel])*E_Onboarding_Entity_Macro_Weighter[E_Onboarding_Entidade_Macro_Ponderador Value],0)))))
```

Figure 3.13 – DAX Formula of the Average Score per Entity Group

And for the final score for each client:

```
Measure Final Score Variation = [Measure Country Group] + [Measure Entity Group] + [Measure Product Group] + [Measure Behavioural Group]
```

Figure 3.14 – DAX Formula of the Final Score Variation

```
Measure Final Score Variation AH = sumx(Scores_Power_BI, if(Scores_Power_BI[Auto High] = "Y",500, [Measure Final Score Variation]))
```

Figure 3.15 – DAX Formula of the Final Score with Auto-High Variation

3.4.3. Definition of the Parameters – Variation

In order to support the process of defining each parameter for each variable of the CRR Model, two pages in the report were created to make the process of choosing easily understandable and intuitive. As it is possible to analyse, for each risk factor and group there is a drop-down box where the user that is analysing the report can choose a parameter between 0 and 1. Additionally, two controls were created to guarantee that the sum of parameters defined is 1 for the entity group as well as the sum of the groups' parameters.

Each slicer contains in its field the parameter value and impacts instantly and in real-time all the graphs presented in the rest of the dashboards since all these slicers are synced with the other pages. Concerning visual aspects, the colours defined have the purpose of associating all the same risk factors that are part of one group with the same colour.

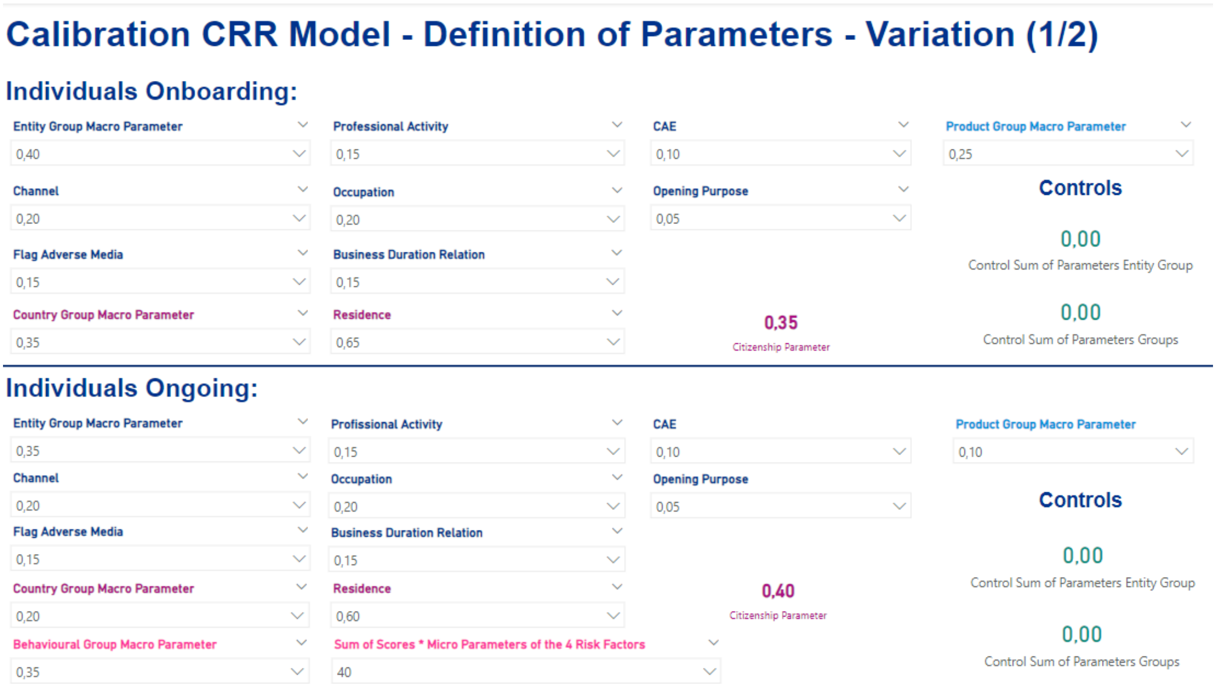


Figure 3.16 – Dashboard of Definition of Parameters Individuals

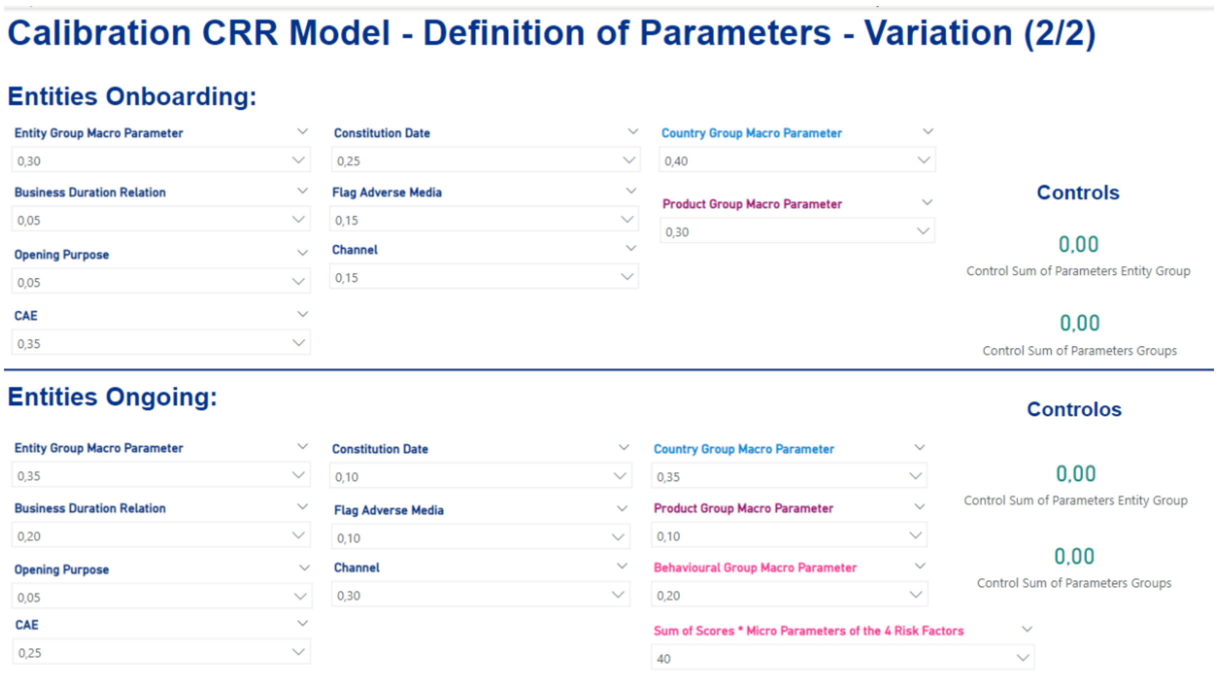


Figure 3.17 – Dashboard of Definition of Parameters Corporate Clients

After developing what was described, it was important to define a specific problem that this solution could solve and also to prototype the rest of the pages with the relevant information to conclude if the objective was fulfilled or if some adjustments were needed.

3.5. Demonstration

The specific situation identified to analyse the performance of this tool was defined based on the analysis presented in Figure 3.1, where it is possible to conclude that most of the clients have a business relationship with a low score (20 points). It was possible to verify that this low score was associated to older durations in business relations, which means, that most of the clients are evaluated in the ongoing model. Since the ongoing model is usually for clients with a relation higher than 182 or 365 days and most of the clients do not have a recent relation, it was decided that this solution would evaluate the impact of just changing the macro parameters (weights of the groups) in the model and keeping the micro-ones (weights of each risk factor) the same in both visions. Additionally, it was decided to create the baseline with no behavioural group, which means, the weight of the parameter of the behavioural group is 0. This decision was placed to show the impact of introducing the behavioural group in the variation. Since most of the clients achieve the condition of being on the ongoing model, it was decided to give more than 50% to the transactional group in order to understand the impact of including transactional history in the risk calculation of the clients. However, it was not possible to present values based on variables, because the data randomly created only had KYC information. So, for this study, it was considered a dynamic parameter: the average score for each client in terms of metrics calculated, where it is possible to choose between 20, 40 and 100 points. In other words, 40 points represent that the average of each client in the behavioural group is 40 – Medium Risk. This allows us to be conservators and analyse the impact on the distribution of clients by each level of risk if, for example, the average of all the clients in the behavioural group would be 100 points.

Even though the universe of clients at a bank can be very diverse, it is not expected that most clients are associated to suspicious transactions. However, it was decided for this study to consider a constant value of 40, which means medium risk for all the clients, for analysing from a conservative perspective the distribution of clients per level of risk.

The scale that defines the level of risk was an assumption used in this demonstration. As it is possible to understand the dashboard with detailed information about the clients and their characteristics, the tendency that we concluded was that most of the clients have low risk in most of the risk factors. To be able to present a distribution that identifies clients in all the risk levels, it was decided to use the following scale: 0 < Final Score <= 35 the client is low risk, 35 < Final Score <= 60 the client is medium risk and higher than 60 the client is high risk. The maximum score used for the strategic rules was defined as 500. Lastly, if a client has some information missing or is not parametrized, the score applied is 100 (as the high risk). This assumption was created based on the characteristics of the clients, and it can be adjusted for further work in the case that the clients present different characteristics.

With this situation defined, it was possible to develop the rest of the solution in order to present the analysis required based on the parameters chosen.

3.5.1. Prototyping the Final Solution

After developing the steps described, it was necessary to finish the prototype phase of the solution, for doing it, the parameters chosen in Figures 3.18 and 3.19 were considered for presenting the following results. In the other vision, it is possible to analyse in the appendices the parameters used for the baseline vision.

This chapter will present the results of the pages developed to answer the requirements that were defined in the design phase.

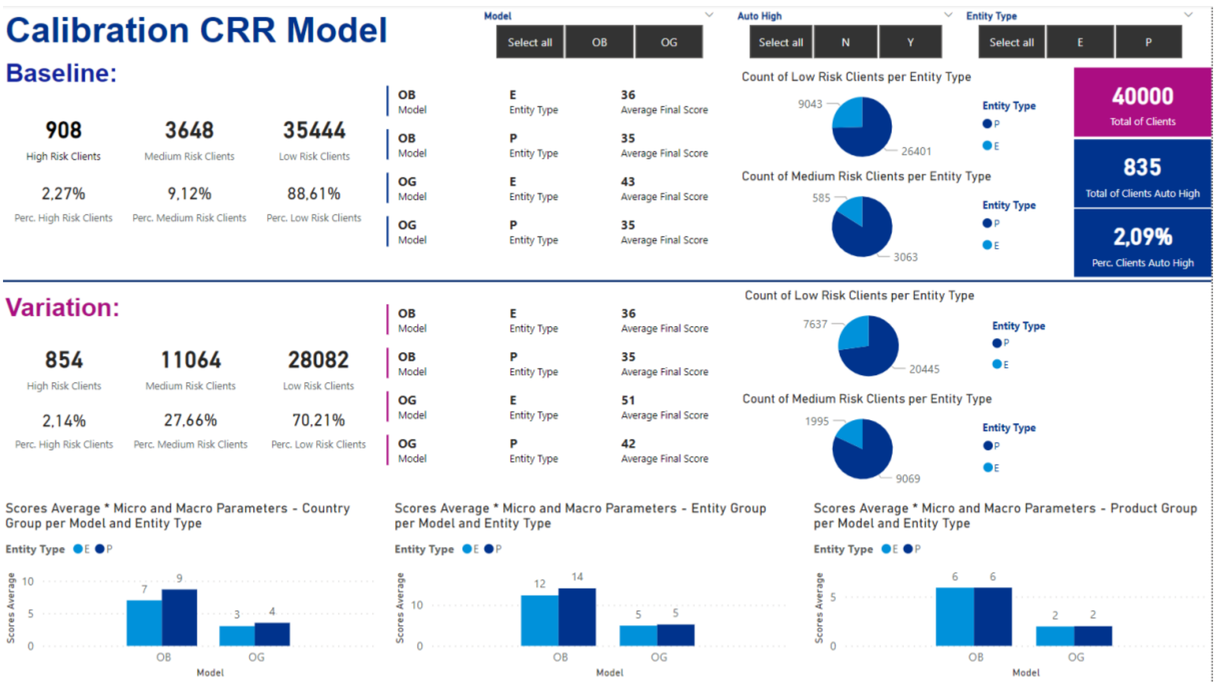


Figure 3.18 – Dashboard Resume

The first page works as a resume of the main statistics about the distribution of clients and the final score. For presenting all these results, the principles of DV were followed, like "Highlighting the most important information" and "Avoiding clutter", which means, cards for the main numbers and the slicers to choose each model, the entity type and remove the auto-high clients. Also, both visions are clearly illustrated in this dashboard, and it was tried to keep a similar structure in each one of them to make the dashboards easier to understand. In the case of applying selections in the slicers, it would show a clear vision of the statistics. In the bar charts, it represents the global average of the score multiplied by the parameter chosen for each model and entity type. Additionally, this page presents the count of clients per level of risk and also the averages of final scores achieved for both visions. With this information, some requirements that were established in this study were achieved.

As one of the requirements was to present a vision without the application of strategic rules, which means, the same universe of clients but without heuristic rules, the dashboard presented before was replicated for presenting the statistics for this vision:

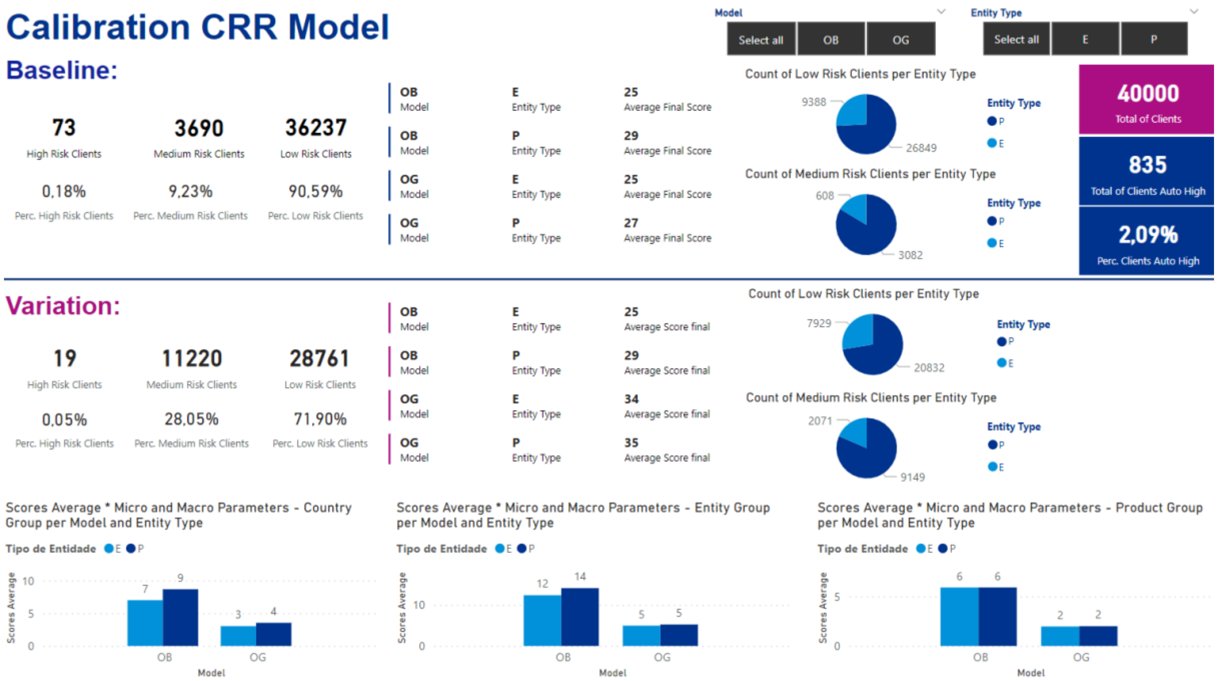


Figure 3.19 – Dashboard Resume Without Auto-High Rules

As it is understandable from the figure, the structure was kept from the dashboard presented below in order to have a clear comparison between both results. The application of strategic rules influences mainly the high-risk clients, as it is possible to observe in the statistics. While going through the analysis of this dashboard, a financial institution can evaluate the impact of defining strategic rules and their effectiveness, which means, a client can have a low score in all the other factors, but if he has a" Y" in one of the strategic rules, it makes him an automatically high-risk.

The third and fourth dashboards are where the parameters were chosen, they have the same structure as it was mentioned in the prototyping phase since it was not identified any relevant topic or necessity of change regarding the structure and appearance of the page. Although below it is possible to analyse the final parameters used in this demonstration:

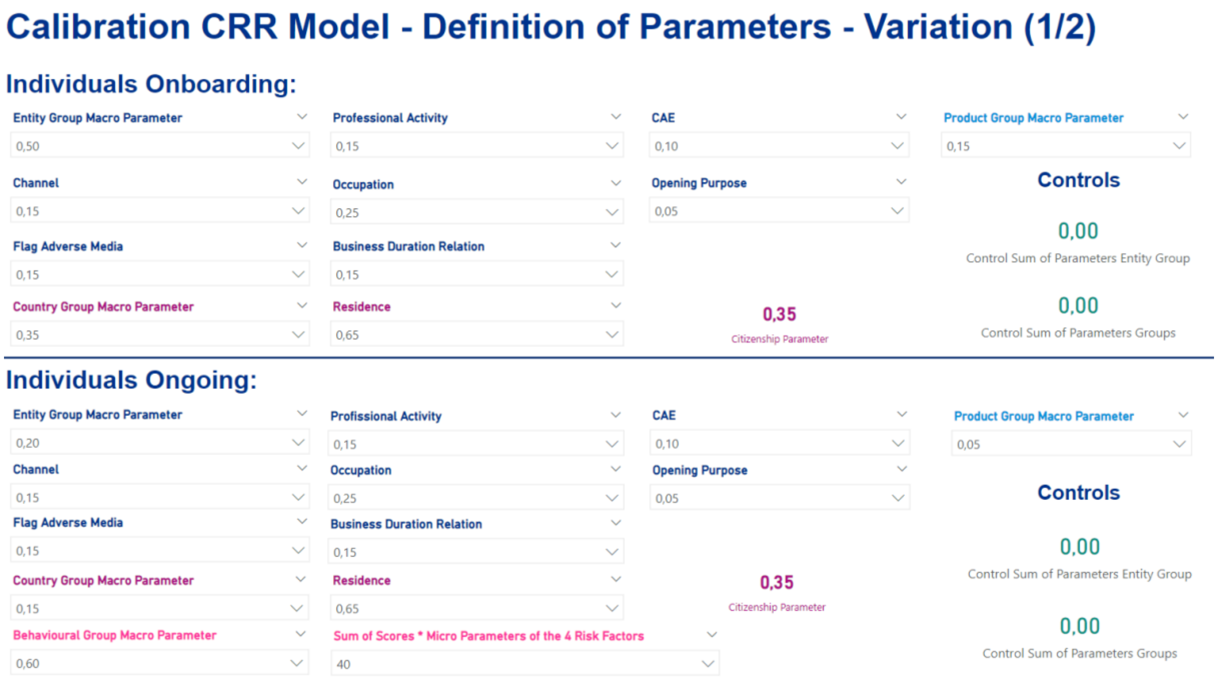


Figure 3.20 – Final Parameters Considered Individuals Clients

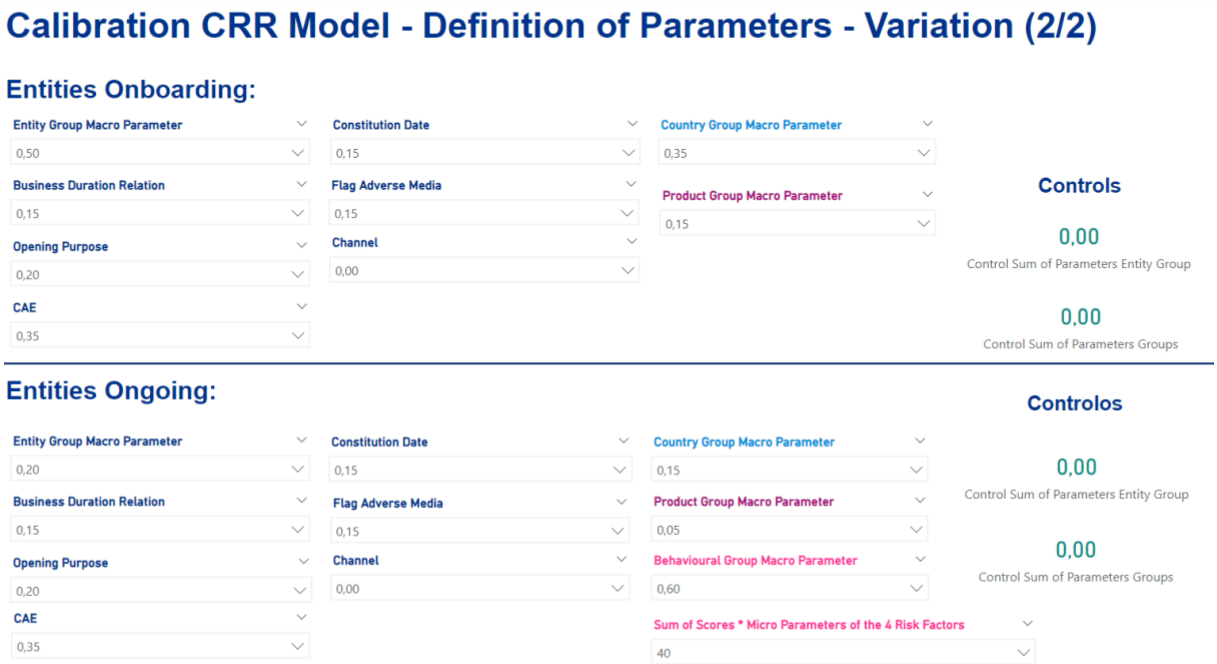


Figure 3.21 – Final Parameters Considered Corporate Clients

As was mentioned at the beginning of this chapter, it was supposed to define a weight heavier than 50% to the Behavioural Group to analyse the impact of the transactional history in the clients' risk. The rest of the parameters were defined based on the ones of the Baseline vision, where they were kept in the same proportion and with the same distribution. Additionally, as it was mentioned in one of the requirements, for creating a dynamic report where the user can define the parameters used, it is important to keep the coherency in the results for creating comparisons. In order to maintain coherency, it is necessary to have controls that alert if the user's choices are correctly applied. In the right corner, it was created two controls for helping the user's experience and to present results based on parameters well-defined. The first control, "Control Sum of Parameters Entity Group" guarantees that the sum of the parameters chosen in the Entity Group is 1. So, if, for example, the user would change the parameter "Opening Purpose" for 0,25 the control would present a red value with a value of 0,05, which means, the sum exceeds 0,05 the supposed total defined in the literature review of the model, which is 1. The second control follows the same logic as the previous one, it guarantees that the sum of the macro parameters chosen is 1, and it also follows the same logic if a value higher than 1 is the result of the sum.

The only parameter that follows a different method of creation is the "Sum of Scores * Micro Parameter of the 4 Risk Factors" since it has 3 different possible values, 20, 40 and 100.

As was mentioned in the demonstration chapter, it was impossible to present and utilize in this master thesis data regarding the transactional history of each client. Considering this group in the calculation of the risk was considered an average score for all the clients. This value of the average score means that the sum of the scores multiplied by the respective micro-parameter is 20, 40 or 100. The four risk factors were assumed as an example in the case of having an average for the four of them.

This assumption can influence the accuracy of the model, but as it was mentioned, the solution was developed to receive real data and calculate the risk with the respective scores for each client.

The user can choose one of the 3 values to apply in the calculation of risk. The results presented took into consideration a conservative vision where the average score for all the clients is 40 points. Since the macro-parameter defined for the behavioural group is 60%, this decision considerably impacts the results. However, it is expected that the transactional history of the clients impacts their risk because they have a relationship with an institution older than 6 months. Since the bank already knows the KYC information about the client and their information already influenced the risk during the first months (onboarding phase), it is normal to follow a logic of analysing and give more importance to the transactional behaviour of the client to understand if the risk is aligned with the characteristics of him or if he has a suspicious history of transactions that do not match with his characteristics.

In this study, the distribution followed was based on the best practices considered in the market and should be adapted to the reality of the client in order to answer his necessities in a complete way. The objective is to present this report and change the parameters for analysing the results performed in real-time in terms of distributions and the risk calculated for each client.

The following dashboards were developed based on the requirements of presenting the distribution of scores for the Baseline vision to understand the impact of adding a behavioural group to the model. It was considered relevant to divide the high-risk clients into a specific dashboard since this group of clients requires a detailed analysis and can represent a problem of Money Laundering or illegal activities to a bank. As it is possible to analyse, it was created a line graph for each level of risk, and it presents the distribution of the total of clients per score. In the cards performed it is possible to understand what the average of scores per model and type of client in each level of risk is. General information is also presented in the cards on the right where it is mentioned the total number of clients per risk level and the respective percentage in the total number of clients. The zoom slicer added in the visualization was to analyse in a more detailed way the total number of clients since the scale has a very high maximum value (5k).

Figure 3.22 – Dashboard Medium and Low Risk Clients

Regarding the High-Risk clients, as it was mentioned, a specific dashboard was created with some statistics in order to verify the impact of the strategic rules defined. Based on this group of clients it is possible to understand whether the number of clients is aligned with the expectations of a bank regarding the generation of alerts and the necessary diligence that they need to apply. It presents the number of clients regarding each vision created and it includes a variation without the application of strategic rules, which means, if there was not any heuristic rule created how many clients would be High-Risk. This analysis can help to understand if the scale of levels of risk is well defined or if some adjustments should be made in order to reduce or increase the number of clients at this level. As it is possible to observe, the strategic rules have a huge impact on the final risk of most of the clients, and just 19 clients would be High-Risk without any transformation of their score. By analysing these visualizations, the banks can evaluate the impact of creating or removing a strategic rule and analyse in the results if that choice had a several impact.

Figure 3.23 – Dashboard High-Risk Clients

One of the requirements was to present statistics about the distribution of the average scores per each group of risk factors and to present what are the risk factors that contribute more to the final risk.

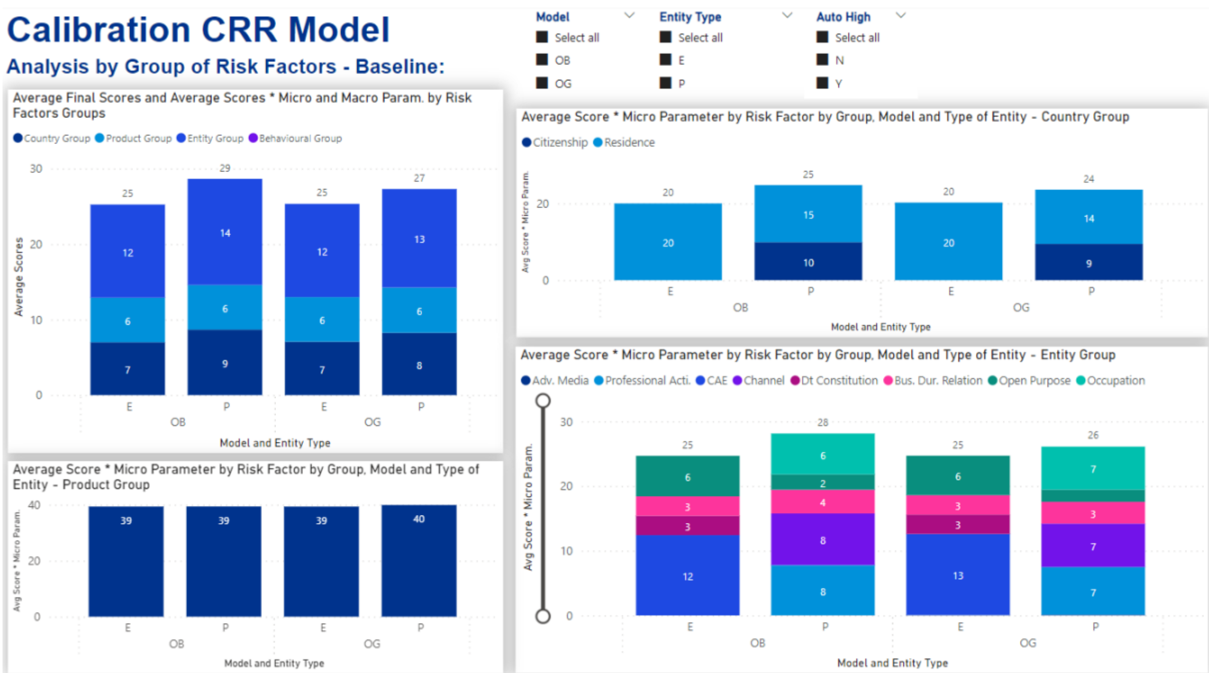


Figure 3.24 – Dashboard Exploratory Analysis Baseline

In Figure 3.24, it is presented how the dashboard "Exploratory Analysis – Baseline" was prototyped. This page was created based on the calculations and results obtained for the baseline vision, and the attributes used are calculated columns as mentioned in the construction of this vision. To help the user access and analyse the information, it followed a structure of starting with high-level analysis and then going through each group of risk factors. The first graph in the top left corner presents the general view of the scores and how each group contributes to those values. It is divided into the model and the entity type. Since in this vision, it was not considered the behavioural group, weight was 0, and there is no part of the bar related to that attribute. Each of the values presented in each bar represents the average of results obtained in each group of risk multiplied by the macro-parameter defined.

The same creation method was followed for the other graphs, it just changed the attributes considered since each graph represents one group of risk factors. The complexity of the entity group in terms of the number of risk factors impacts the interpretability and capacity of analysis of its graph. However, a zoom slicer was used for navigating the visualization and understanding the impact of all the attributes.

There are also slicers at the top of the page, which allow the user to see the results regarding the type of entity, model or remotion of Auto-High clients. Even though there is already a separation in the graphs, these slicers make the analysis of the graph easier.

The next figure (Figure 3.25) presents a similar dashboard compared with the one presented above and it contains the results of the Variation vision. It has the same graphs, but for calculating the values created, also new measures were performed with the average of the scores per risk factor and as a higher-level analysis, the average of scores per groups.

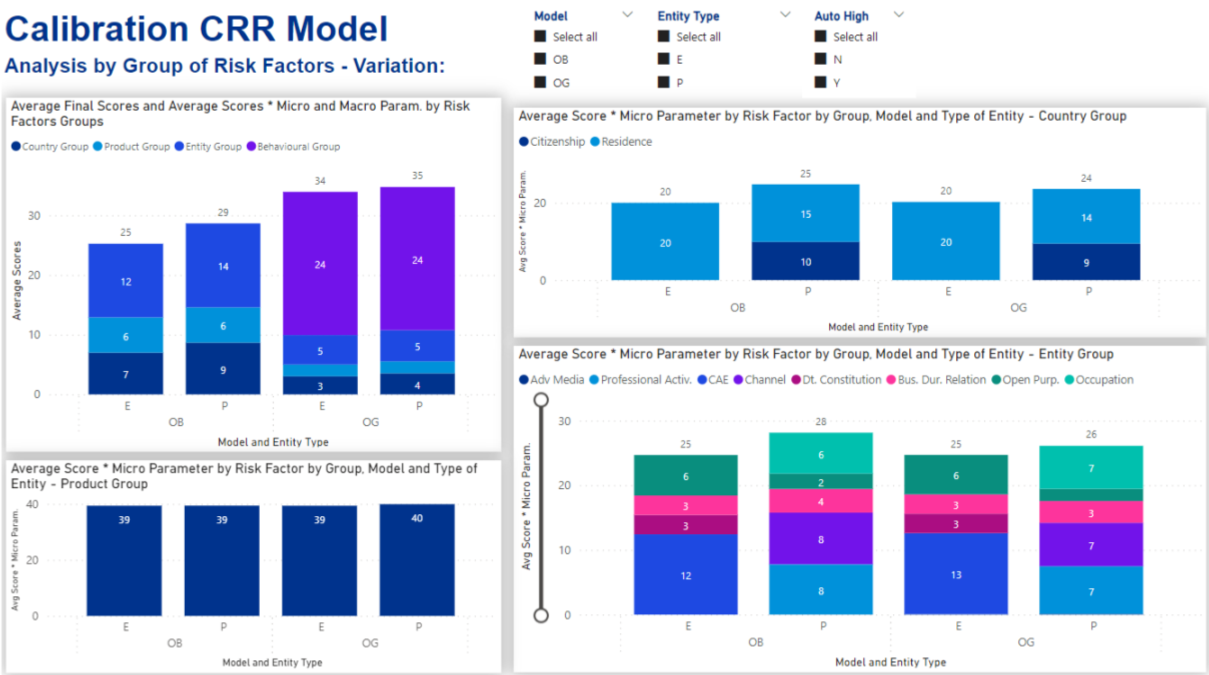


Figure 3.25 – Dashboard Exploratory Analysis Variation

3.6. Results and Evaluation

Concerning the requirements defined at the beginning of this study, there was an alignment of the expectations before starting to prototype the solution. By developing this tool, it was possible to understand through the demonstration chapter the powerfulness of the solution and how it can improve the presentation of results performed by a CRR Model. Additionally, while the final version of the Power BI dashboards was being developed, some discussions and meetings with specialists of the area (AML/CDD and DV) were placed for sharing the results achieved. The feedback received was positive, and the overall evaluation focused on the impressive capacity of the tool on presenting real-time analysis regarding adjustments and changes performed during these discussions.

Even though this master thesis did not use real data for presenting the results, the capacity to answer all the analyses initially defined improved the credibility of the solution developed. It is important to mention that the purpose of this study was also to develop an adaptive tool that could be personalized regarding the necessities and characteristics of each client, and, mainly, a tool that can take as input different risk factors and can be adapted to each CRR Model created based on weighted models.

The easy access to Power BI also helps the visualization of data, where it is possible to synthesize the information on each page of the report and analyse it from a detailed perspective with the application of the slicers on specific topics. However, the exploratory results for both visions have some detailed information, so it can be difficult to analyze the impacts in the graph and understand the values presented. It was a topic identified as relevant to improve in future projects.

Concerning the demonstration problem, the main topic was to evaluate the relevance of changing the macro parameters of the vision Variation, in order to add the impact of the behavioural group in the risk calculation. Again, Power BI had a significant impact on the effectiveness, accuracy, and efficiency of the results obtained, since the changes applied were quickly performed and the results were successfully updated in real-time. For the business world, specifically financial institutions and consulting companies, where several projects regarding CRR Models are developed, now it is possible to add this tool to presentations or discussions with clients, and if some adjustments are needed in each weight of the risk factors, it is possible to present its impact without being dependent on other executions or manual changes that can take time and resources.

As will be mentioned in the limitations, some challenges were faced during the process of developing this solution. However, it is now possible to measure and analyze the final results of this solution, and, most importantly, evaluate the impact of Power BI in the calibration process of CRR Models.

The results achieved in the demonstration show how Power BI can easily lead with a universe of 40.000 clients and their information by accessing the dataset in a quick way. For creating the variation vision, it is needed to create several measures, that influence the time spent performing the visualizations utilized in each dashboard. For evaluating the performance of some dashboards, the Performance Analyzer by Power BI was taken into consideration. Performance Analyzer inspects and displays the time required to update or refresh all visuals that user interactions initiate (Decoding the Most Unused Feature of Power BI – the Performance Analyzer, n.d.).

With this evaluation tool, it was possible to analyze how much time the visualizations take to update their results regarding the parameters chosen in the slicers of the pages presented before. With this, it is possible to present if, for example, in a meeting, it would be necessary to spend too much time updating the results if some changes were made. For the baseline vision, the results were significantly faster since they were not dependent on the parameters created.

The results achieved for the dashboard "Resume" can be analyzed in the figure below:

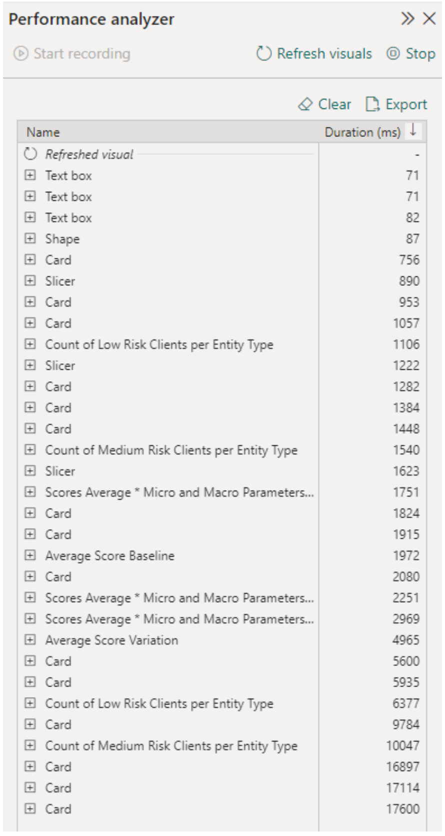


Figure 3.26 – Performance Analyzer of the Resume Dashboard

As it is possible to see from Figure 3.26, the maximum time that a visual took to refresh its results was 17600 ms, which means, 17.6 seconds. Since the dashboard "Resume" has the main numbers and the visualizations presented use information from the most complex measures, it was just analyzed the performance of the visuals on this page. The results obtained show that the visualizations will not take too much time when, for example, it is decided to choose new parameters in the respective pages for analyzing the results in this dashboard. These results also help to understand which visualizations are taking more time. It was expected that the cards registered the maximum time since they are the ones related to the distributions of the final risk per each level of risk. It was predictable that this behaviour would occur because these cards perform statistics regarding the last step of calculations, the final score. The final score is dependent on some other measures, which also use several results from others in their calculation. In the end, it is expected that the dimension of the dataset used as input also impacts these results. However, it is normal that the software takes some time to perform all the results since the number of measures created is also complex, with a lot of dependencies between each one of them.

Analysing the intervals of time also shows that changing one or all the parameters does not impact the performance of the report.

In terms of communication, as it is the last chapter of the methodology used, this report can also be useful in different areas since it is based on presenting results performed by a weighted model, so it can be also adapted to different industries and consider different risk factors for new studies. Since it was developed to be used as a tool, it was important to have the validation and recognition of the work that was done. In this specific part, after presenting the final solution to specialists of the area regarding the CRR Models, it was possible to validate the logic besides the results achieved. Lastly, the outcome received was that the solution was correct and complete to be used in future studies or in real projects.

4 LIMITATIONS

The main limitation of this study is the data used for performing the visualizations. Even though the data was in the right format, all the parametrizations were given without business context, except the score of each country that was analysed in the list of the UE. The source of data is not possible to present in this study regarding the confidentiality terms adopted by the company in question.

The initial solution defined to developed in Power BI had the objective of taking into consideration the transactional history of each client and the dynamic definition of intervals of values as it was created for the parameters, for attributing the respective scores. It was not possible to present the dashboards for those concepts since there was no access to real data and it was not generated randomly information for transactional variables.

Based on the dashboards developed, as mentioned, several measures were created in order to answer all the requirements and develop a complete solution. Some challenges were faced when Power BI did not charge in the visualizations because of its memory limit. This limitation was just faced when it was tried to create measures for a specific group of clients (high-risk clients) since it was based on defining measures that depended on other measures and on the parameters chosen and the time to charge the results was too long.

Lastly, in the figures presented, some names of some variables are in Portuguese because it was the original language considered when this study started. However, as it is possible to visualize in the front-end of the solution, everything is in the language of this document and does not impact the user experience when analyzing the report.

5 CONCLUSIONS AND FUTURE WORKS

The main objective of this master's thesis was to develop a tool that could impact the calibration process of CRR Models and improve the decision-making processes of parameters when presenting final distributions. Plus, it was also an objective to perform a report that could easily present results regarding the characteristics of the clients and could also be used as a supportive tool for the decision-making process of the weights of each risk factor. The first conclusion achieved was that it was possible to create a report in Power BI that answered the points mentioned.

The methodology applied helped the process of developing the dashboards since several changes and adjustments were made regarding the necessities that occurred and the requirements defined. Due to the challenge of creating a tool that could impact this calibration process, it was necessary to invest time in learning technical concepts for replicating the logic of the mathematical model with the possibility of dynamic changes in the parameters. Also, some meetings were taken with specialists of the areas in order to evaluate what could be necessary to adapt and validate the results achieved. After all these steps, it was possible to successfully create a report that answered the requirements defined with accurate and correct results.

The demonstration phase presented in this study was very important in order to evaluate the performance and impact of the tool created. In the future, regarding each client's transactional profile, it will be possible to demonstrate an extension of the impact of this tool, since the dashboards were successfully tested for a sample of information created manually. The power of developing a dynamic tool is based on the capacity to analyse in real time the impact of applying changes to the parameters and evaluate its impact. In Power BI, the visualizations easily produced the results needed, improving and creating a better user experience. Plus, the controls provided were also created for anyone familiar with Power BI to be able to analyze the report and simulate results.

Nevertheless, there are further works that can improve the accuracy of the results, like applying machine learning techniques to the definition of the parameters based on the relevance of the variables used in the mathematical formula. Even though this model was simple and did not have an impact on using it in this tool, it is important to have an optimized and efficient model with the right variables and techniques for optimizing the weights. Plus, applying those techniques can give the possibility of changing the visions presented in the Power BI, by focusing on some specific points, like different parametrizations (scores) considered in each vision, or a detailed analysis for a specific group of clients.

Concluding, there can be several adjustments to this report in order to fulfil the necessities of each situation. However, it is possible to confirm that Power BI can be used on the calibration process of Customer Risk Rating Models since all the dashboards presented were focused on providing supportive results that can improve the decision process of parameters in the universe of weighted models.

BIBLIOGRAPHICAL REFERENCES

AdvisoryHQ, B. (n.d.). AML KYC Risk Rating Assessment Template, Methodology, Rating Matrix – Download Template – AdvisoryHQ. https://www.advisoryhq.com/articles/aml-risk-assessment-template-and-sample/

Alessa. (2023, August 28). Elements of customer risk: profiles and relationships. Alessa. https://alessa.com/blog/elements-of-customer-risk-profile/

Anti-Money-Laundering.eu. (2023, November 28). Customer Risk Rating Methodology (CRRM) - Anti-Money laundering. Anti-Money Laundering. https://anti-money-laundering.eu/customer-risk-rating-methodology/

Brush, K., & Burns, E. (2022, December 8). Data Visualization. Business Analytics. https://www.techtarget.com/searchbusinessanalytics/definition/data-visualization

Cameron, N. (2021, November 24). Anti-Money Laundering/Financing Terrorism Compliance Risk Rating. KPMG. https://kpmg.com/ae/en/home/insights/2021/11/anti-money-laundering-financing-terrorism-compliance-risk-rating.html

Cameron, S. (2023, April 20). Customer Risk Assessment: What you Need to Know. ComplyAdvantage. https://complyadvantage.com/insights/customer-risk-assessment/

Cams, C. D. L. (2024, February 21). The Importance of Customer Risk Assessment and Risk Rating in Financial Institutions. https://www.linkedin.com/pulse/importance-customer-risk-assessment-rating-financial-dike-llm-cams-73c2f?trk=public_post_main-feed-card_feed-article-content

Central Bank of Ireland. (2021, June 23). Anti-Money Laundering and Countering the Financing of Terrorism Guidelines for the Financial Sector. https://www.centralbank.ie/docs/default-source/regulation/amld-/guidance/anti-money-laundering-and-countering-the-financing-of-terrorism-guidelines-for-the-financial-sector.pdf

City &Talent. (2024, January 20). Decoding the most unused feature of Power BI – the Performance Analyzer. City &Talent. https://cityandtalent.com/business-intelligence/decoding-the-most-unused-feature-of-power-bi-the-performance-analyzer/

De Koker, L. (2006). Money laundering control and suppression of financing of terrorism: some thoughts on the impact of customer due diligence measures on financial exclusion. Journal of Financial Crime, vol 13(1). Emerald. pp. 26-50.

Eppler, M. J. (2023, February 2). Data DESIGN: Six "must-know" Data Visualization Principles for Everyone. https://www.linkedin.com/pulse/data-design-six-must-know-visualization-principles-everyone-eppler

FATF. (2012-2023). International Standards on Combating Money Laundering and the Financing of Terrorism & Proliferation. FATF. www.fatf-gafi.org/en/publications/Fatfrecommendations/Fatf-recommendations.html

FATF Annual Report 2005-2006. (n.d.). FATF. https://www.fatf-gafi.org/en/publications/Fatfgeneral/Fatfannualreport2005-2006.html

Favarel-Garrigues, G., Godefroy, T., & Lascoumes, P. (2009). Combating Money Laundering : Sentinels and tools at work. https://hal.science/hal-00835957

Financial Crime Academy (2023, November 6). The History Of Money Laundering: Know The Important Origin Of Money Laundering. Financial Crime Academy. https://financialcrimeacademy.org/the-history-of-money-laundering-2/

Hammond, T. (2023, November 22). 6 most useful data visualization principles for analysts. ThoughtSpot. https://www.thoughtspot.com/data-trends/data-visualization/data-visualization-principles

High risk third countries and the international context content of anti-money laundering and countering the financing of terrorism. (n.d.). Finance. https://finance.ec.europa.eu/financial-crime/high-risk-third-countries-and-international-context-content-anti-money-laundering-and-countering_en

History of Anti-Money Laundering Laws | FinCEN.gov. (n.d.). https://www.fincen.gov/history-anti-money-laundering-laws

Ibitola, J. (2023, October 12). Customer Risk Profiling: A Key to AML Compliance. Flagright. https://www.flagright.com/post/customer-risk-profiling-a-key-to-aml-compliance-2

Iceland: Financial Sector Assessment Program-Technical Note on Anti-Money Laundering/Combating the Financing of Terrorism. (2023, July 28). IMF eLibrary. https://doi.org/10.5089/9798400248238.002.A001

Mathias, E. (2023, April 21). Leveraging Anti-money Laundering Measures to Improve Tax Compliance and Help Mobilize Domestic Revenues. IMF eLibrary. https://doi.org/10.5089/9798400240409.001.A001

McKowen, K. (2023, September 7). How do you successfully operationalize your client risk rating model? EY – Switzerland. https://www.ey.com/en_ch/disrupting-financial-crime/how-do-you-successfully-operationalize-your-client-risk-rating-model

Mikkelsen, D., Pravdic, A., & Richardson, B. (2019, September 16). Flushing out the money launderers with better customer risk-rating models. McKinsey & Company. https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/flushing-out-the-money-launderers-with-better-customer-risk-rating-models

Oracle® Financial Services Know Your Customer Risk Assessment Guide. (n.d.). Oracle. https://docs.oracle.com/cd/E91253_01/PDF/8.0.4.0.0/KYC_8.0.4_Risk_Assessment_Guide.pdf

Parida, D. K., & Kumar, D. P. (2020, July 15). A Framework to Score the Risk Associated with Suspicious Money Laundering Activity and Social Media Profile. Indian Journal of Finance and Banking, 4(2), 1–10. https://doi.org/10.46281/ijfb.v4i2.662

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2008). A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems, 24(3), 45-77.

Ratanachu-Ek, N. (2019, May 16). Using optimization to define weights and parameters for customer risk rating models - The SAS Data Science Blog. The SAS Data Science Blog. https://blogs.sas.com/content/subconsciousmusings/2019/05/16/using-optimization-to-define-weights-and-parameters-for-customer-risk-rating-models/

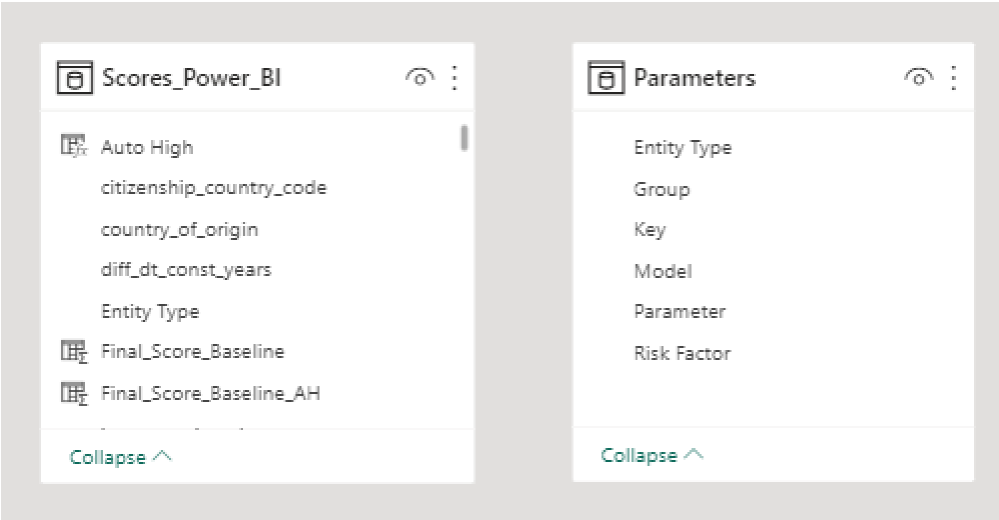
Rivera, E., West, J., & Suplee, C. (2015). Addressing AML Regulatory Pressures by Creating Customer Risk Rating Models with Ordinal Logistic Regression. SAS Institute Inc. https://support.sas.com/resources/papers/proceedings15/SAS2603-2015.pdf

Santosh, E. & Mohanty, L. (2011). The Growing Threat of Money Laundering. Capgemini. https://www.capgemini.com/wp-content/uploads/2017/07/The_Growing_Threat_of_Money_Laundering.pdf

Zimiles, E. S. (2004, July 1). KPMG Survey: Banks accept more costly money-laundering laws, expect heightened cooperation with regulators. Journal of Investment Compliance, 5(3), 26–30. https://doi.org/10.1108/15285810410737512

APPENDIX

Model View of the Report

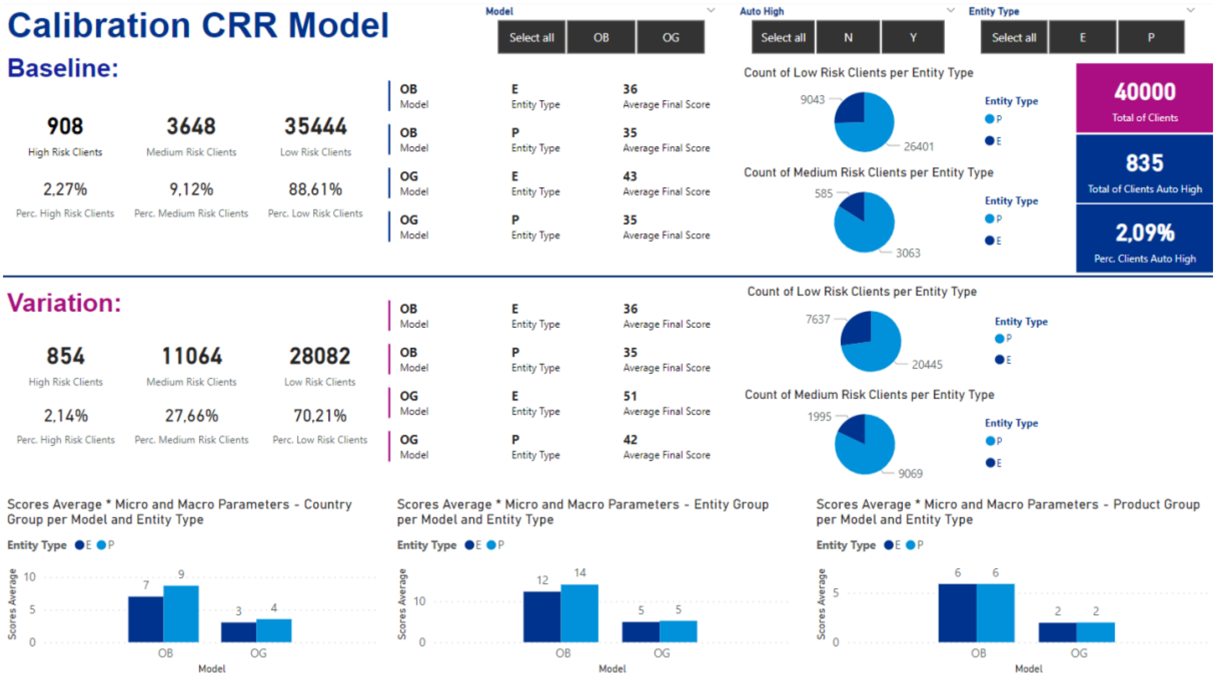


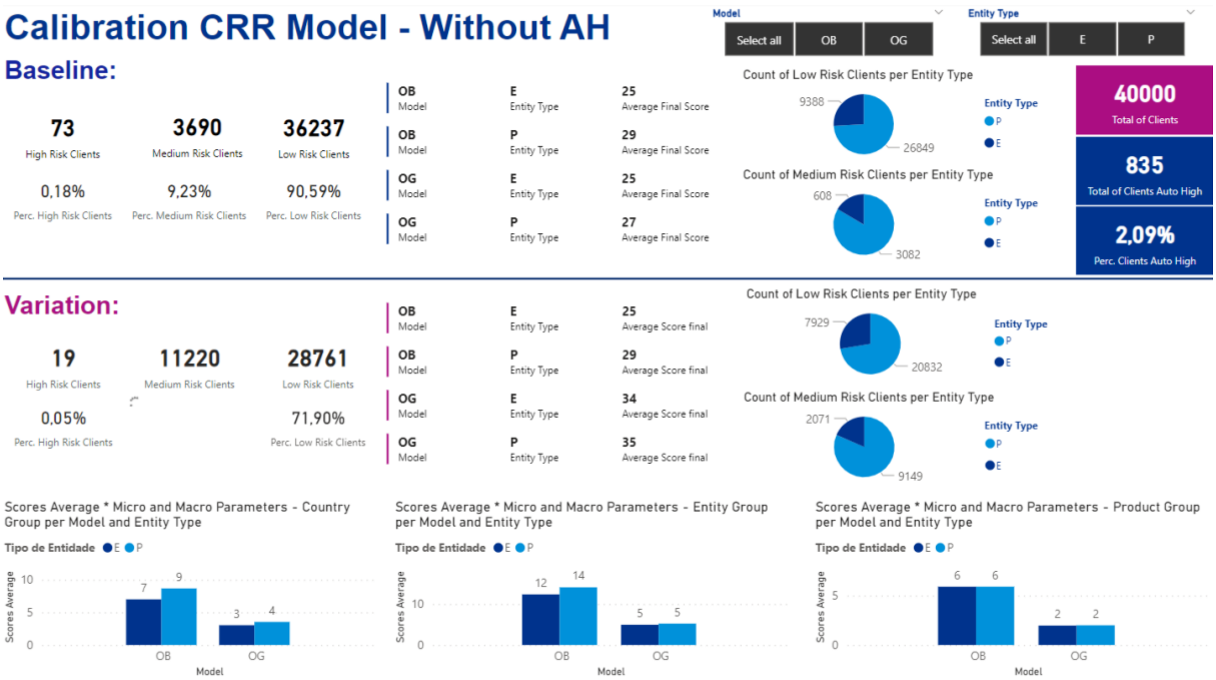
Parameters Baseline

Entity Type	Model	Group	Risk Factor	Parameter
E	Onboarding	Entidade	Data_Constituição	0,15
E	Onboarding	Entidade	Dur_Rel_Neg	0,15
E	Onboarding	Entidade	Flag_Adv_Media	0,15
E	Onboarding	Entidade	Prop_Rel_Neg	0,20
E	Onboarding	Entidade	CAE	0,35
E	Onboarding	Entidade	Macro_Ponderador	0,50
E	Onboarding	País	Macro_Ponderador	0,35
E	Onboarding	País	Residência	1,00
E	Onboarding	Produto	Macro_Ponderador	0,15
E	Onboarding	Produto	Produto	1,00
E	Ongoing	Entidade	Data_Constituição	0,15
E	Ongoing	Entidade	Dur_Rel_Neg	0,15
E	Ongoing	Entidade	Flag_Adv_Media	0,15
E	Ongoing	Entidade	Prop_Rel_Neg	0,20
E	Ongoing	Entidade	CAE	0,35
E	Ongoing	Entidade	Macro_Ponderador	0,50
E	Ongoing	País	Macro_Ponderador	0,35
E	Ongoing	País	Residência	1,00
E	Ongoing	Produto	Macro_Ponderador	0,15
E	Ongoing	Produto	Produto	1,00

Entity Type	Model	Group	Risk Factor	Parameter
P	Onboarding	Entidade	Prop_Rel_Neg	0,05
P	Onboarding	Entidade	CAE	0,10
P	Onboarding	Entidade	Atividade_Profissional	0,15
P	Onboarding	Entidade	Canal_Contratação	0,15
P	Onboarding	Entidade	Dur_Rel_Neg	0,15
P	Onboarding	Entidade	Flag_Adv_Media	0,15
P	Onboarding	Entidade	Profissão	0,25
P	Onboarding	Entidade	Macro_Ponderador	0,50
P	Onboarding	País	Macro_Ponderador	0,35
P	Onboarding	País	Nacionalidade	0,35
P	Onboarding	País	Residência	0,65
P	Onboarding	Produto	Macro_Ponderador	0,15
P	Onboarding	Produto	Produto	1,00
P	Ongoing	Entidade	Prop_Rel_Neg	0,05
P	Ongoing	Entidade	CAE	0,10
P	Ongoing	Entidade	Atividade_Profissional	0,15
P	Ongoing	Entidade	Canal_Contratação	0,15
P	Ongoing	Entidade	Dur_Rel_Neg	0,15
P	Ongoing	Entidade	Flag_Adv_Media	0,15
P	Ongoing	Entidade	Profissão	0,25
P	Ongoing	Entidade	Macro_Ponderador	0,50
P	Ongoing	País	Macro_Ponderador	0,35
P	Ongoing	País	Nacionalidade	0,35
P	Ongoing	País	Residência	0,65
P	Ongoing	Produto	Macro_Ponderador	0,15
P	Ongoing	Produto	Produto	1,00

Dashboards Developed





Calibration CRR Model - Definition of Parameters - Variation (1/2)

Individuals Onboarding:

Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
Entity Group Macro Parameter	0,50	Professional Activity	0,15	CAE	0,10	Product Group Macro Parameter	0,15
Channel	0,15	Occupation	0,25	Opening Purpose	0,05		
Flag Adverse Media	0,15	Business Duration Relation	0,15				
Country Group Macro Parameter	0,35	Residence	0,65	Citizenship Parameter	0,35		

Controls

0,00
Control Sum of Parameters Entity Group

0,00
Control Sum of Parameters Groups

Individuals Ongoing:

Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
Entity Group Macro Parameter	0,20	Profissional Activity	0,15	CAE	0,10	Product Group Macro Parameter	0,05
Channel	0,15	Occupation	0,25	Opening Purpose	0,05		
Flag Adverse Media	0,15	Business Duration Relation	0,15				
Country Group Macro Parameter	0,15	Residence	0,65	Citizenship Parameter	0,35		
Behavioural Group Macro Parameter	0,60	Sum of Scores * Micro Parameters of the 4 Risk Factors	40				

Controls

0,00
Control Sum of Parameters Entity Group

0,00
Control Sum of Parameters Groups

Calibration CRR Model - Definition of Parameters - Variation (2/2)

Entities Onboarding:

Parameter	Value	Parameter	Value	Parameter	Value
Entity Group Macro Parameter	0,50	Constitution Date	0,15	Country Group Macro Parameter	0,35
Business Duration Relation	0,15	Flag Adverse Media	0,15	Product Group Macro Parameter	0,15
Opening Purpose	0,20	Channel	0,00		
CAE	0,35				

Controls

0,00
Control Sum of Parameters Entity Group

0,00
Control Sum of Parameters Groups

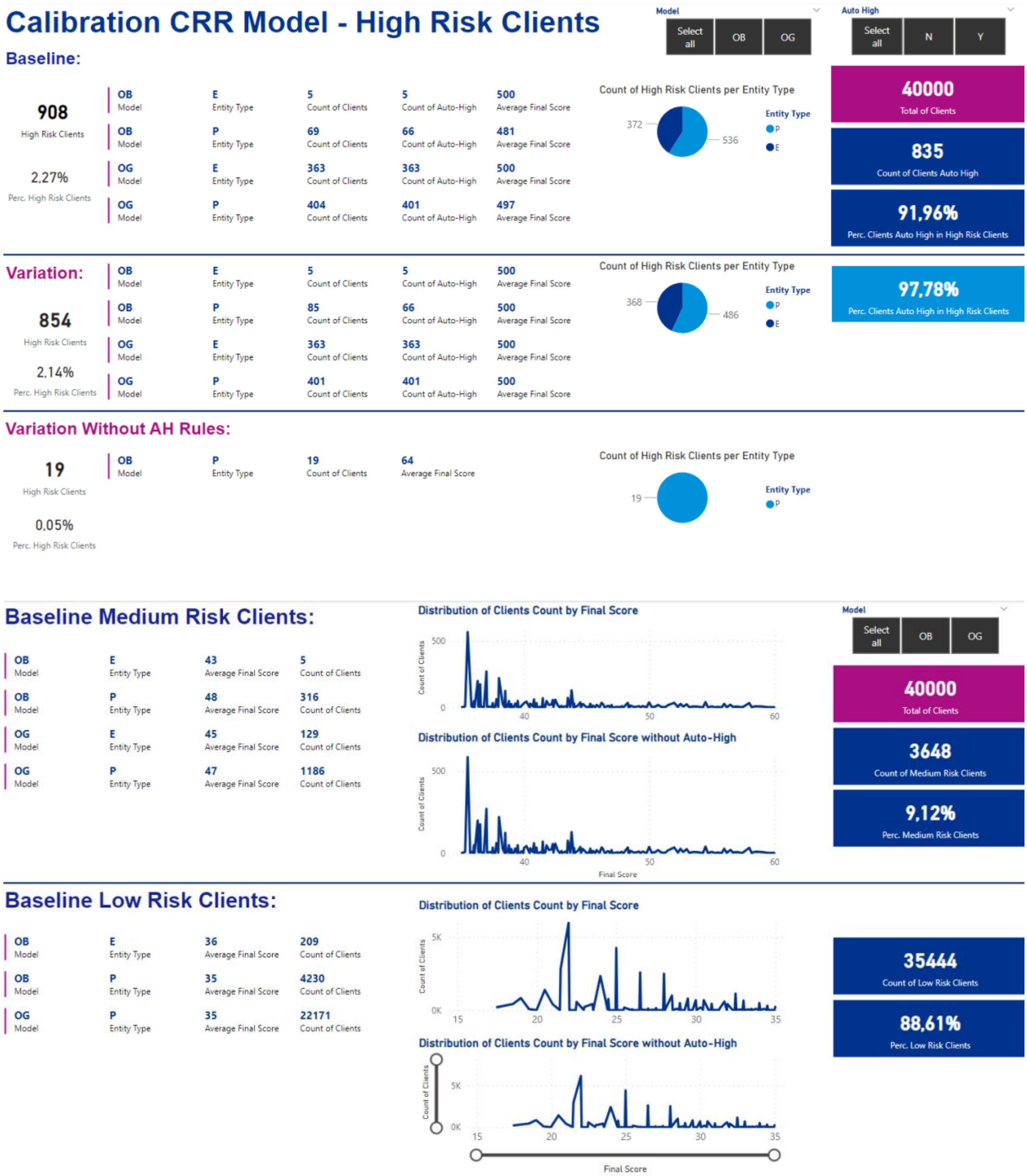
Entities Ongoing:

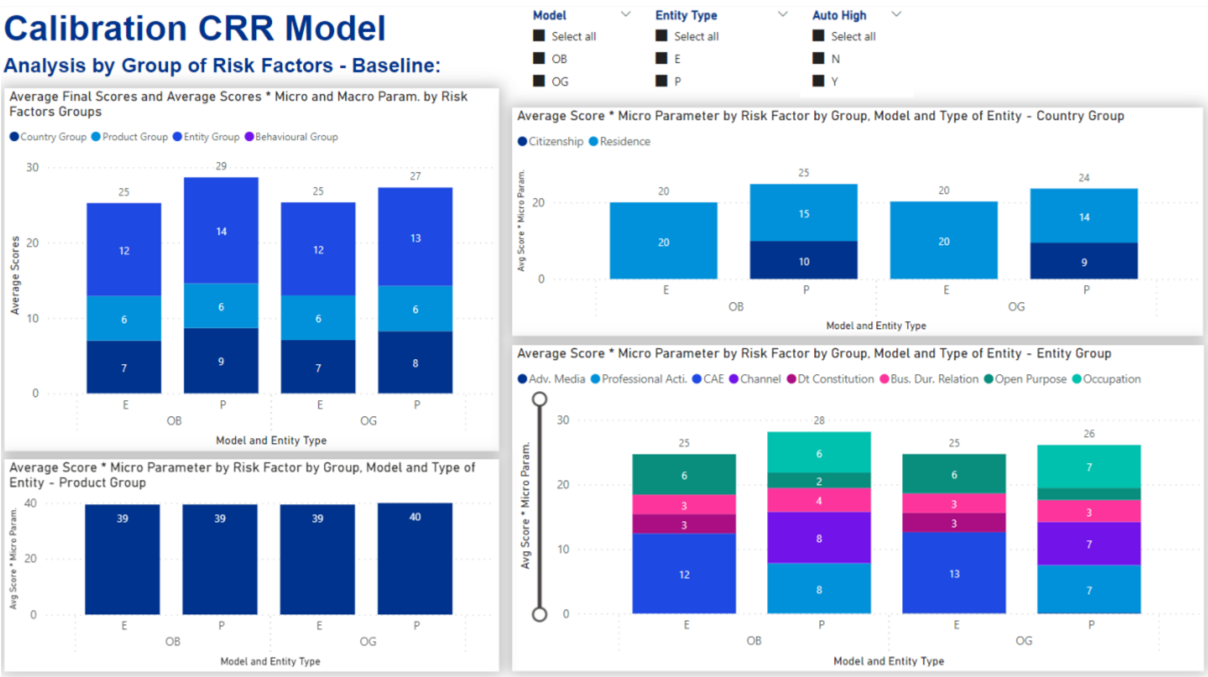
Parameter	Value	Parameter	Value	Parameter	Value
Entity Group Macro Parameter	0,20	Constitution Date	0,15	Country Group Macro Parameter	0,15
Business Duration Relation	0,15	Flag Adverse Media	0,15	Product Group Macro Parameter	0,05
Opening Purpose	0,20	Channel	0,00	Behavioural Group Macro Parameter	0,60
CAE	0,35			Sum of Scores * Micro Parameters of the 4 Risk Factors	40

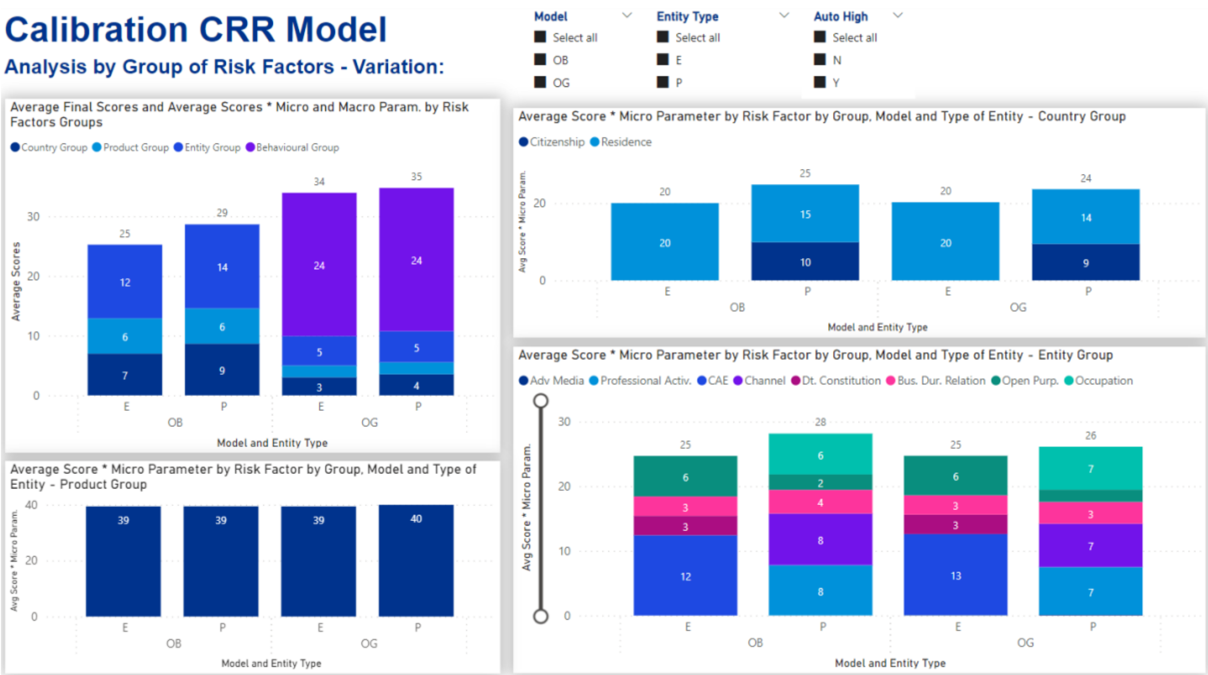
Controlos

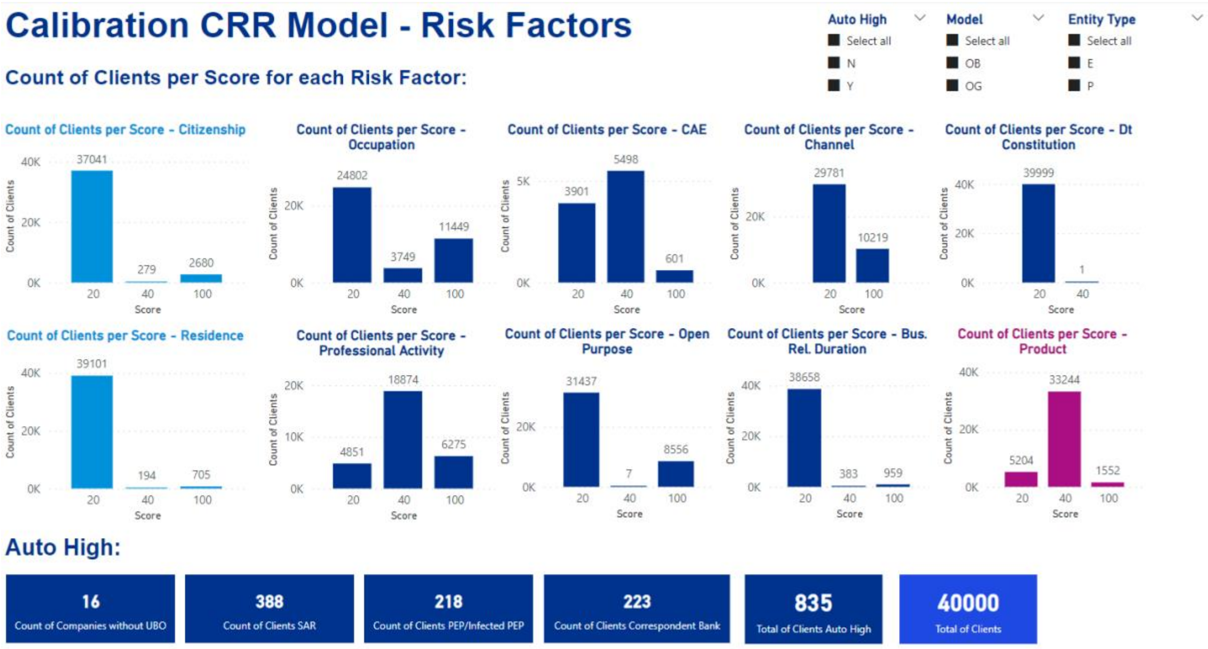
0,00
Control Sum of Parameters Entity Group

0,00
Control Sum of Parameters Groups


Calibration CRR Model - High Risk Clients
Model
Select all
OB
OG
Auto High
Select all
N
Y
Baseline:
908
High Risk Clients
2,27%
Perc. High Risk Clients
OB Model | E Entity Type | 5 Count of Clients | 5 Count of Auto-High | 500 Average Final Score
OB Model | P Entity Type | 69 Count of Clients | 66 Count of Auto-High | 481 Average Final Score
OG Model | E Entity Type | 363 Count of Clients | 363 Count of Auto-High | 500 Average Final Score
OG Model | P Entity Type | 404 Count of Clients | 401 Count of Auto-High | 497 Average Final Score
Count of High Risk Clients per Entity Type
372
536
Entity Type
P
E
40000
Total of Clients
835
Count of Clients Auto High
91,96%
Perc. Clients Auto High in High Risk Clients
Variation:
854
High Risk Clients
2,14%
Perc. High Risk Clients
OB Model | E Entity Type | 5 Count of Clients | 5 Count of Auto-High | 500 Average Final Score
OB Model | P Entity Type | 85 Count of Clients | 66 Count of Auto-High | 500 Average Final Score
OG Model | E Entity Type | 363 Count of Clients | 363 Count of Auto-High | 500 Average Final Score
OG Model | P Entity Type | 401 Count of Clients | 401 Count of Auto-High | 500 Average Final Score
Count of High Risk Clients per Entity Type
368
486
Entity Type
P
E
97,78%
Perc. Clients Auto High in High Risk Clients
Variation Without AH Rules:
19
High Risk Clients
0,05%
Perc. High Risk Clients
OB Model | P Entity Type | 19 Count of Clients | 64 Average Final Score
Count of High Risk Clients per Entity Type
19
Entity Type
P
Baseline Medium Risk Clients:
OB Model | E Entity Type | 43 Average Final Score | 5 Count of Clients
OB Model | P Entity Type | 48 Average Final Score | 316 Count of Clients
OG Model | E Entity Type | 45 Average Final Score | 129 Count of Clients
OG Model | P Entity Type | 47 Average Final Score | 1186 Count of Clients
Distribution of Clients Count by Final Score
Count of Clients
500
0
40
50
60
Distribution of Clients Count by Final Score without Auto-High
Count of Clients
500
0
40
50
60
Final Score
Model
Select all
OB
OG
40000
Total of Clients
3648
Count of Medium Risk Clients
9,12%
Perc. Medium Risk Clients
Baseline Low Risk Clients:
OB Model | E Entity Type | 36 Average Final Score | 209 Count of Clients
OB Model | P Entity Type | 35 Average Final Score | 4230 Count of Clients
OG Model | P Entity Type | 35 Average Final Score | 22171 Count of Clients
Distribution of Clients Count by Final Score
Count of Clients
5K
0K
15
20
25
30
35
Distribution of Clients Count by Final Score without Auto-High
Count of Clients
5K
0K
15
20
25
30
35
Final Score
35444
Count of Low Risk Clients
88,61%
Perc. Low Risk Clients


Calibration CRR Model
Analysis by Group of Risk Factors - Baseline:
Model
Select all
OB
OG
Entity Type
Select all
E
P
Auto High
Select all
N
Y
Average Final Scores and Average Scores * Micro and Macro Param. by Risk Factors Groups
Country Group
Product Group
Entity Group
Behavioural Group
Average Scores
Model and Entity Type
Average Score * Micro Parameter by Risk Factor by Group, Model and Type of Entity - Product Group
Avg Score * Micro Param.
Average Score * Micro Parameter by Risk Factor by Group, Model and Type of Entity - Country Group
Citizenship
Residence
Average Score * Micro Parameter by Risk Factor by Group, Model and Type of Entity - Entity Group
Adv. Media
Professional Acti.
CAE
Channel
Dt Constitution
Bus. Dur. Relation
Open Purpose
Occupation


Calibration CRR Model
Analysis by Group of Risk Factors - Variation:
Model
Select all
OB
OG
Entity Type
Select all
E
P
Auto High
Select all
N
Y
Average Final Scores and Average Scores * Micro and Macro Param. by Risk Factors Groups
Country Group
Product Group
Entity Group
Behavioural Group
Average Scores
Model and Entity Type
Average Score * Micro Parameter by Risk Factor by Group, Model and Type of Entity - Product Group
Avg Score * Micro Param.
Average Score * Micro Parameter by Risk Factor by Group, Model and Type of Entity - Country Group
Citizenship
Residence
Average Score * Micro Parameter by Risk Factor by Group, Model and Type of Entity - Entity Group
Adv Media
Professional Activ.
CAE
Channel
Dt. Constitution
Bus. Dur. Relation
Open Purp.
Occupation



Calibration CRR Model - Parameters Baseline

Entity Type	Model	Group	Risk Factor	Parameter
E	Onboarding	Entidade	Data_Constituição	0,15
E	Onboarding	Entidade	Dur_Rel_Neg	0,15
E	Onboarding	Entidade	Flag_Adv_Media	0,15
E	Onboarding	Entidade	Prop_Rel_Neg	0,20
E	Onboarding	Entidade	CAE	0,35
E	Onboarding	Entidade	Macro_Ponderador	0,50
E	Onboarding	País	Macro_Ponderador	0,35
E	Onboarding	País	Residência	1,00
E	Onboarding	Produto	Macro_Ponderador	0,15
E	Onboarding	Produto	Produto	1,00
E	Ongoing	Entidade	Data_Constituição	0,15
E	Ongoing	Entidade	Dur_Rel_Neg	0,15
E	Ongoing	Entidade	Flag_Adv_Media	0,15
E	Ongoing	Entidade	Prop_Rel_Neg	0,20
E	Ongoing	Entidade	CAE	0,35
E	Ongoing	Entidade	Macro_Ponderador	0,50
E	Ongoing	País	Macro_Ponderador	0,35
E	Ongoing	País	Residência	1,00
E	Ongoing	Produto	Macro_Ponderador	0,15
E	Ongoing	Produto	Produto	1,00

Entity Type	Model	Group	Risk Factor	Parameter
P	Onboarding	Entidade	Prop_Rel_Neg	0,05
P	Onboarding	Entidade	CAE	0,10
P	Onboarding	Entidade	Atividade_Profissional	0,15
P	Onboarding	Entidade	Canal_Contratação	0,15
P	Onboarding	Entidade	Dur_Rel_Neg	0,15
P	Onboarding	Entidade	Flag_Adv_Media	0,15
P	Onboarding	Entidade	Profissão	0,25
P	Onboarding	Entidade	Macro_Ponderador	0,50
P	Onboarding	País	Macro_Ponderador	0,35
P	Onboarding	País	Nacionalidade	0,35
P	Onboarding	País	Residência	0,65
P	Onboarding	Produto	Macro_Ponderador	0,15
P	Onboarding	Produto	Produto	1,00
P	Ongoing	Entidade	Prop_Rel_Neg	0,05
P	Ongoing	Entidade	CAE	0,10
P	Ongoing	Entidade	Atividade_Profissional	0,15
P	Ongoing	Entidade	Canal_Contratação	0,15
P	Ongoing	Entidade	Dur_Rel_Neg	0,15
P	Ongoing	Entidade	Flag_Adv_Media	0,15
P	Ongoing	Entidade	Profissão	0,25
P	Ongoing	Entidade	Macro_Ponderador	0,50
P	Ongoing	País	Macro_Ponderador	0,35
P	Ongoing	País	Nacionalidade	0,35
P	Ongoing	País	Residência	0,65
P	Ongoing	Produto	Macro_Ponderador	0,15
P	Ongoing	Produto	Produto	1,00

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa