LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA

MASTER OF SCIENCE
MATHEMATICAL FINANCE

MASTERS FINAL WORK
INTERNSHIP REPORT

MODELING FINANCIAL NEEDS OF BANK CLIENTS THROUGHOUT THEIR LIFE CYCLES

RICARDO FANHA VICENTE SOARES

OCTOBER - 2018

LISBON SCHOOL OF ECONOMICS & MANAGEMENT
UNIVERSIDADE DE LISBOA

MASTER OF SCIENCE
MATHEMATICAL FINANCE

MASTERS FINAL WORK
INTERNSHIP REPORT

MODELING FINANCIAL NEEDS OF BANK CLIENTS THROUGHOUT THEIR LIFE CYCLES

RICARDO FANHA VICENTE SOARES

SUPERVISION:
PEDRO LOUREIRO
ONOFRE ALVES SIMÕES

OCTOBER – 2018

Acknowledgements

I would like to express my deepest gratitude to Professor Onofre Simões and Dr. Pedro Loureiro, for the guidance, the support and the good words, the patience and out-of-hours work done towards this project. For every lesson, advice and problem solving, I thank you.

To Gonçalo Traquina, Frederico Silva and António Leitão, for the precious advice in each stage of the project and for every tool provided, which enabled every analysis performed throughout this work, I thank you.

To Miguel Caiado, for your time dedicated to this work, your aid, motivation and optimism, I thank you.

To KPMG and the M&RC department, I'm grateful for the excellent work environment set to an intern.

To my Master's colleagues and friends, for the kind words and the support given, I thank you.

To everyone who answered my survey, I thank you.

To Mihaela. Without you, this work would be impossible. I deeply thank you, for being there for me at all times, for keeping me cheerful and for helping me overcome every obstacle.

And a special gratitude to my parents, Cristina and António, and my grandmother, Teresa, because I owe it all to you.

Abstract

The constant change that has been noticeable on a worldwide level has impacted the economical, geopolitical and social frameworks, with direct repercussions across every sector. For example, the financial sector is compelled to rethink strategies and to adapt its procedures in order to be up to date with respect to the current technological and social disruption, thus creating new business methods and opportunities.

Currently, the financial institutions need new approaches to retain and to capture clients, such as knowing them and anticipating their needs. It is in this context that the concept of financial necessities arises, a subject that needs academic and professional investigation.

With this work, it is intended to describe the different stages of the human life cycle, according to data collected during the project, and to provide an analysis on how its components are related to the decision process of acquiring financial products and services.

It is also expected to provide a framework to the life cycle – financial necessities relationship and to develop a statistical model that predicts these needs throughout the life cycle and optimizes the propensity to consumption of financial products and services.

The analysis carried out in this work discloses relevant insights, such as the existence of correlation between typical life events (e.g. marriage or childbirth) with the acquisition of financial products, a detailed framework of the life cycle - financial needs relationship and a predictive model which outputs the next best offer for clients.

Key Words: Life Cycle, Financial Necessities, Cluster Analysis, Alteryx

Resumo

A constante mudança que se tem feito notar a nível global, gerou um enorme impacto nas grandes envolventes económicas, geopolíticas e sociais, fazendo-se sentir por todos os sectores. Nomeadamente, o sector financeiro tem sido obrigado a repensar estratégias e a adaptar abordagens para conseguir acompanhar toda a disrupção tecnológica e social que se tem sentido, criando e aproveitando novas oportunidades de negócio.

Atualmente, as instituições financeiras precisam de novas estratégias para reter os clientes atuais e capturar novos clientes, tal como a antecipação das suas necessidades. É neste contexto que surge o tema das necessidades financeiras de cada pessoa, um tema que carece de uma abordagem mais prática e realista, bem como de um enquadramento técnico adaptado ao tempo presente.

Com este trabalho é ambicionado descrever as diferentes fases do ciclo de vida de uma pessoa, com base em dados recolhidos, e analisar de que modo é que os seus componentes têm influência na decisão do consumidor, relativamente à aquisição de produtos e serviços financeiros.

É também pretendida a elaboração de um enquadramento da relação ciclo de vida – necessidades financeiras e o desenvolvimento de um modelo que projete estas necessidades ao longo do ciclo e que otimize a propensão ao consumo, mediante a posição de cada pessoa no ciclo de vida.

Esta análise revela pontos bastante relevantes, tal como a existência de correlação entre a ocorrência de certos eventos (como, por exemplo, casamento, nascimento de um filho) e a aquisição de produtos financeiros, um esquema detalhado da relação entre o ciclo de vida e as necessidades financeiras e ainda um modelo que prevê a melhor oferta de produtos e serviços financeiros para cada cliente.

Palavras-chave: Necessidades Financeiras, Ciclo de Vida, Análise de Clusters, Alteryx

Table of Contents

Table of figures

List of Abbreviations

P&S – Products and Services

1. Introduction

The present report resulted from a five months curricular internship at KPMG-Advisory, within the Management & Risk Consulting department, as part of the conclusion of ISEG's master program in Mathematical Finance.

Apart from applying concepts learned throughout the academic part of the masters, the internship also allowed for learning new notions and approaches to topics addressed there. Most importantly, during the internship it was possible to be where theory meets practice and to see how what had been learned was used to create value within the financial system, particularly how statistical and econometric theory could be used for gaining insights and building models to optimize clients' financial consumption and improve financial institutions' performance.

The life cycle of a regular person is correlated with his/her own financial life cycle, where the typical life events, such as marriage or retirement, trigger needs to purchase financial products and services. This empiric perspective has increasingly gained importance within the financial business area, although lacking models and concrete tools that assure statistical significance, allowing the estimation of propensity to consumption regarding the relationship life cycle - financial products and services (P&S).

Hence, the scope of this project is to develop a life cycle framework and to build a model that optimizes the propensity to consumption of financial products and services throughout the course of life.

2. Life Cycle and Financial Needs

2.1 Motivation

In line with the insights of Sikander Sattar, KPMG Portugal CEO, in KVISION (Jan 2018), it is easily perceivable that the world is changing, fast and unpredictably. The economic, geopolitical and social pictures exhibit strong shifts and disruptions. These changes affect all economic sectors worldwide and the financial sector is not an exception.

Under this frame, financial players have to adapt to the new circumstances and must be leaders of innovation. The markets don't behave as they did 50, 40 or 20 years ago. Now, financial institutions, especially banks, have to create business, that is, provide new services to capture new clients (or to retain their own). Essentially, banks need to embrace change. Financial institutions now think of new ways to be in touch with their clients and one of the approaches chosen is to predict each of their clients' needs. In order to obtain more accurate predictions, they must know each client to the maximum extent possible. And this implies knowing their necessities, immediately arising the question "What factors or events generate these necessities?"

In the course of life, human beings experience several events of all sorts: cultural, physiological, random, professional, etc. Some events, such as Marriage, Childbirth or First Job, are situational or occasional and some events are inevitable, occurring in certain periods or stages of life, due to the combination of some demographic, geographical and financial factors, among others, depending on multiple circumstances and aspects of an individual's lifetime.

Events generate/trigger a set of needs, specifically financial, such as the need for saving, safeguarding, earn/growing and borrowing, that are satisfied by financial products and services. Accordingly, throughout their life cycles, people rely on financial Products and Services to answer needs and overcome difficulties, imposed by the occurrence of events, therefore becoming increasingly prominent the role of financial products and of financial institutions in one's life.

Being a progressively relevant topic, the Life Cycle - Financial Needs relationship is a subject with increasingly relevance within KPMG and its line of business.

Its business approach lacks not only a complete and realistic framework but also a robust statistical model that estimates a customer's propensity to consume financial P&S, based on the customer's life and financial life cycle, thus motivating this project.

The main purpose of this work is to find which drivers generate clients' necessities for financial products and provide this topic with a complete framework and a robust statistical model that can be used for the prediction of such needs.

2.2 Theoretical Background

There are numerous studies on the Human life cycle and its stages, by different kinds of researchers, mainly biologists and sociologists, such as B. Bogin & B. Smith's (1996), writing on the evolution of the life stages throughout history and Erikson, E. (1998), developing a detailed theory of the psychosocial life stages of human development, describing eight life-cycle stages, from birth to death. Later, Slater (2003) develops the Erikson argument, with special focus on the adult life stages. Also, the Dr Thomas Armstrong study in 2008 (The 12 Stages of Life, AILHD), provided advances in this field.

On the other hand, the financial life cycle subject lacks research and investigation, therefore being a somewhat difficult field to analyze. Nevertheless, in this section, the studies and theories that this work is developed upon will be presented.

2.2.1 Life Cycle

According to B. Bogin & B. Smith (1996), the Human life cycle is best described by five stages: Infant, Child, Juvenile, Adolescent and Adult, with factors as the change in human growth rates, meaning that hum an beings have different rates of development during their lifespan which at some extent are related with the changes in the trophic and reproductive behaviour; that is, the differences in the reproductive system that are associated with each life stage.

The scope of the present work regards only the Adult stage, as it will analyze the financial behaviour of the adult population during the Adult stage. As this stage is too broad, it is necessary to split it into four sub-stages, considering a number of different characteristics from the ones stated by B. Bogin & B. Smith (1996) (biological and physiological), such as demographic aspects (marriage, income, number of children, age and professional activity). Therefore, the four sub-stages proposed are, respectively, Young Adults (Early Adulthood), Adults, Midlife and Mature Adults.

The Young Adults life stage comprehend people from 18 to 24 years old who started working in the last five years, that is, who are in the very beginning of their professional life/career and are on the process of separation from their parents' house/guard/allowance, becoming independent and accumulating responsibilities.

People between 25 and 35 years old, who already have an ongoing career, are placed in the Adults stage. Adults are generally independent, with priorities different from those of the Young Adults, more focused on the career, finding a mate and building a family.

Midlife includes people from 35 to 54 years old who are in a later phase, career wise, and whose priorities are more oriented towards growing the family and planning the future.

Lastly, the Mature Adults go from 55 to 67 years old and are on the verge of retirement (or newly retired).

2.2.2 Financial Life Cycle

According to the KPMG Database, the financial life cycle of a human being is partitioned into four stages. Financial Dependency, Financial Independency pre-family, Financial Independency with dependents and Financial Independency pre-retirement & retirement.

The KPMG Database is an extensive collection of data gathered during the firm's activity, that is, a database that contains information associated with KPMG

projects of every kind, from multiple lines of business regarding different firms and people.

Anyone who relies on other person (usually parents) for sustenance belongs to the Financial Dependency Stage.

The second stage includes people who already work and have already moved out of the parent's house, providing for themselves, before marrying and starting a family.

Financial Independency with dependents include people who already started a family of their own, while working and providing for themselves and their dependents.

In the last stage of the financial life cycle we find people who are nearly retiring (or already) retired, whose children are now financially independent.

2.2.3 Financial Products & Services and Financial Needs

The financial products and services sold by most Portuguese banks fall into five major categories: Basic P&S, Credit, Investment, Insurance and Savings. Financial P&S exist to answer a person's financial needs. These needs are classified, according to the KPMG database, as Grow, Safeguard, Spend, Earn and Borrow. As a disclaimer, this approach is adopted due to the almost non-existence of research in this particular subject – different types of financial necessities – thus becoming interesting to test against the data collected when elaborating the project.

During their lifespan, people often need to protect themselves against threats and risks, that is, they need to safeguard themselves or their relatives against certain events. One way to safeguard is to buy insurance.

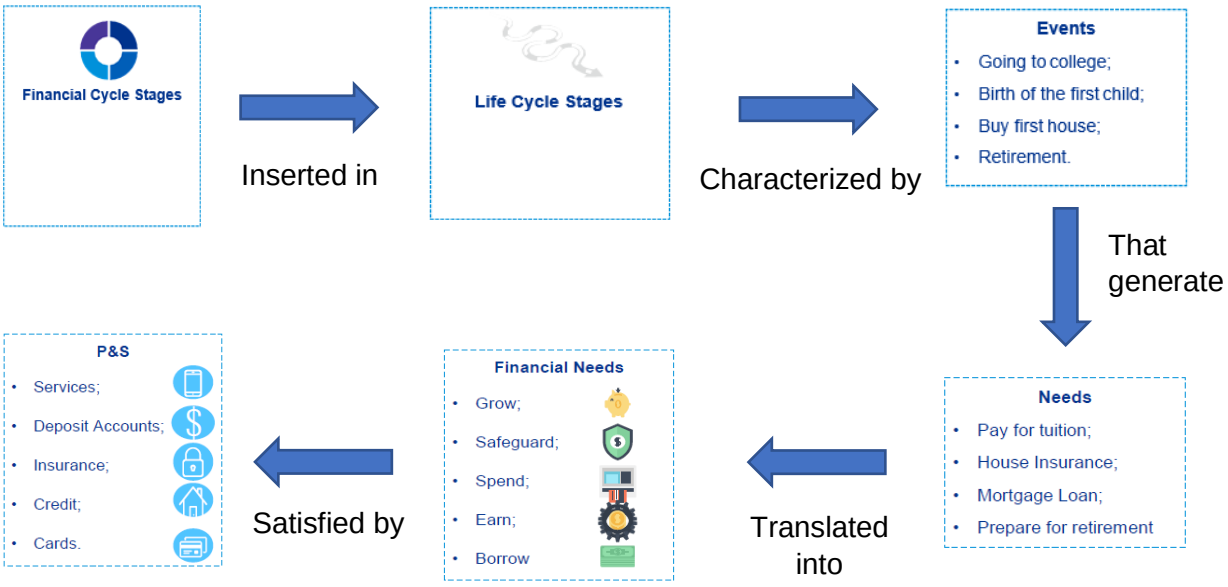
Also, there are moments in life where people lack liquidity, that is, have needs but do not own sufficient resources to satisfy those needs, such as buying a house for their family. In these cases, people generally resort to financial institutions, more specifically, to credit, thus incurring in debt.

Grow is, most times, a constant need throughout the life cycle, leading people to acquire one, or several, savings products, such as savings accounts. Although being a permanent necessity, ideally, there can be times where it isn't possible to save.

The Earn financial need has a risk component associated. Alike Growing, people need to Earn their entire life, therefore purchasing investment products and services from banks, such as participation in funds and investment management, a service that invests on behalf of the client.

2.3 The Life Cycle – Financial Needs Framework

Figure 1 – The KPMG life cycle – financial needs framework



Source: KPMG Database

According to the KPMG Database, the financial life cycle is somewhat inserted in the (general) life cycle, as the age and the occurrence of certain events define the limits of each stage, therefore being considered as a part of the Life Cycle. The life cycle stages are characterized by the occurrence of some stage-specific events, such as getting the first job – usually occurs between the adolescence

phase and the young adult phase, also during the latter, or having children – most common in the Adult and Midlife phases.

Most events throughout life generate needs, namely financial needs. Buying a house very often generates, or has generated the need for borrowing, therefore contracting debt from the bank. The birth of a child generates the need to Safeguard and to Grow, therefore buying insurance and a savings account.

Hence, financial P&S are present throughout life, in every stage of the life cycle.

In the next chapters, the theory presented and the veracity of the previous arguments will be tested against a set of data collected by the author, resorting to multiple statistical procedures and softwares.

3. Survey

In order to study the relationship Life Cycle – Financial Needs and to develop the “Best Offer” Model, it is necessary to gather information on the individual’s demographics, their financial P&S consumption behavior and also the moment in life that certain events occurred, during the life cycle. To obtain this kind of information, a survey was designed and put into practice.

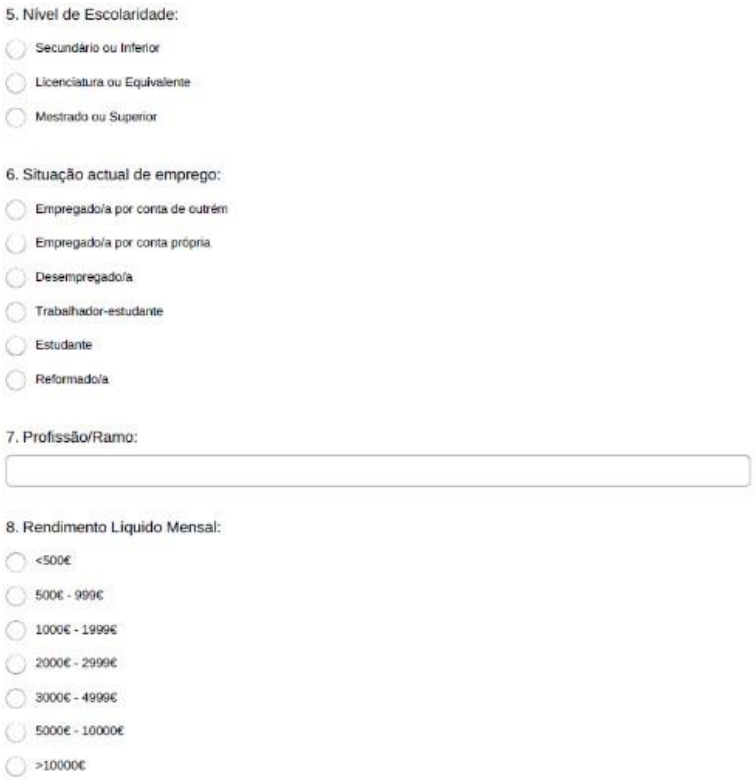
As a start, it was established that the targeted audience would be the active and retired Portuguese population. Students and younger individuals are not included in this work’ scope, since they do not have an established financial consumption behavior.

3.1 The Survey

The survey, hereafter referred to as "survey” (see Appendix A) was built and carried out in the Online Survey Platform “SurveyMonkey”, an online survey tool and it is segmented into three sections of questions: Demographic, Financial P&S consumption and Life Events. The language is Portuguese, with some translation notes subtitled.

The Demographic section (see Figure 2) contains questions on the respondents’ demographics, such as age, activity, and number of children, among others, and also questions on personal budgeting.

Figure 2 – Example of the survey's demographic questions

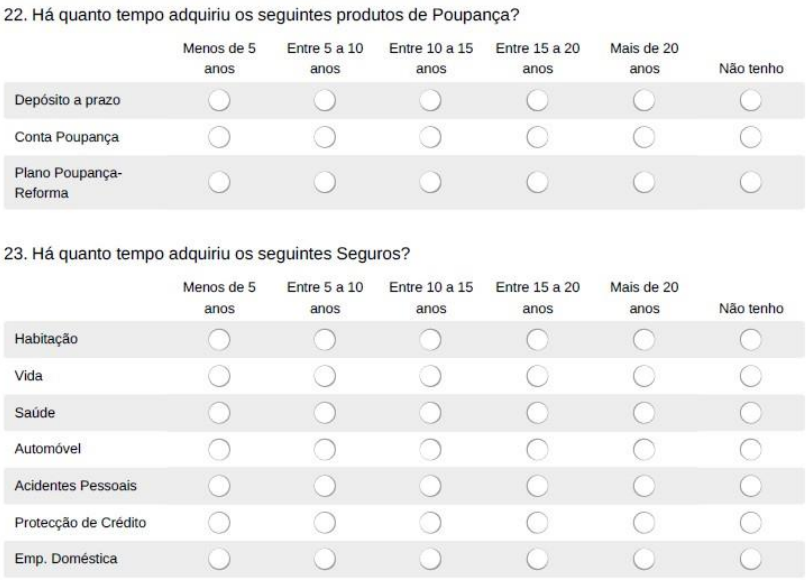


Source: Survey

Translation notes: 5. Education level; 6. Current employment situation; 7. Business/sector; 8. Monthly net income.

The Financial P&S Consumption section asks "when a certain product was acquired", where the respondent chooses from a set of answers going from "in the last 5 years" or "5 to 10 years ago" to "more than 20 years ago". There is also an "I don't have this product" option. The questions are divided into five categories, which are the same used for the P&S: Basic P&S, Credit, Investment, Insurance and Savings (see Figure 3 for an illustration).

Figure 3 – Example of the P&S survey questions



22. Há quanto tempo adquiriu os seguintes produtos de Poupança?

	Menos de 5 anos	Entre 5 a 10 anos	Entre 10 a 15 anos	Entre 15 a 20 anos	Mais de 20 anos	Não tenho
Depósito a prazo	○	○	○	○	○	○
Conta Poupança	○	○	○	○	○	○
Plano Poupança-Reforma	○	○	○	○	○	○

23. Há quanto tempo adquiriu os seguintes Seguros?

	Menos de 5 anos	Entre 5 a 10 anos	Entre 10 a 15 anos	Entre 15 a 20 anos	Mais de 20 anos	Não tenho
Habitação	○	○	○	○	○	○
Vida	○	○	○	○	○	○
Saúde	○	○	○	○	○	○
Automóvel	○	○	○	○	○	○
Acidentes Pessoais	○	○	○	○	○	○
Protecção de Crédito	○	○	○	○	○	○
Emp. Doméstica	○	○	○	○	○	○

Source: Survey

Translation notes: Timeline - How long have you purchased insurance products: House; Life; Health; Car; Personal accident insurance; Credit protection insurance; Domestic employees' insurance.

Lastly, the Events section (Figure 4) includes just the one question – When certain events have occurred in the respondent's life, with the same type of answer options used in the previous section.

Figure 4 – The events survey question

24. Em que momento da sua vida se deram os seguintes acontecimentos?

	Nos últimos 5 anos	Há 5 a 10 anos	Há 10 a 15 anos	Há 15 a 20 anos	Há mais de 20 anos	N/A
Entrada na Faculdade	○	○	○	○	○	○
Saída de casa dos pais	○	○	○	○	○	○
Primeiro Emprego	○	○	○	○	○	○
Casamento	○	○	○	○	○	○
Compra de casa	○	○	○	○	○	○
Nascimento do primeiro filho	○	○	○	○	○	○
Formação Profissional (eg: MBA, cursos executivos, etc)	○	○	○	○	○	○
Emigração	○	○	○	○	○	○
Criação/Expansão de negócio próprio	○	○	○	○	○	○
Doença	○	○	○	○	○	○
Divórcio	○	○	○	○	○	○
Reforma	○	○	○	○	○	○

Source: Survey

Translation notes: Timeline – Admission to University; Move out of parent's house; First job; Marriage; First House; Birth of the first child; Professional Training; Emigration; Own business; Disease; Divorce.

3.2 Goals and Limitations

Before launching the survey, the desirable characteristics of the database were defined.

In order to produce statistically significant insights on the relationship Life Cycle - Financial Needs and capture consumption patterns for the development of a good model, the database should:

- Have at least 500 to 1000 proper observations, or answers to the survey;
- Have demographically heterogeneous observations.

Ideally, the survey should have answers from different regions, by people of all ages, with different professions and lifestyles, in order to capture different patterns of financial products consumption.

Facing these objectives, there were several limitations that this survey, and therefore the project, were subject to.

Since this is a very early approach to the Life Cycle – Financial Needs relationship by the firm, there was an attempt to control the source of the data. The survey was limited to KPMG employees, ISEG alumni, friends and family, and LinkedIn connections.

Time was a constraint as well. Notwithstanding tight deadlines, the survey would benefit from being public for a longer period of time, in order to capture a greater number of answers.

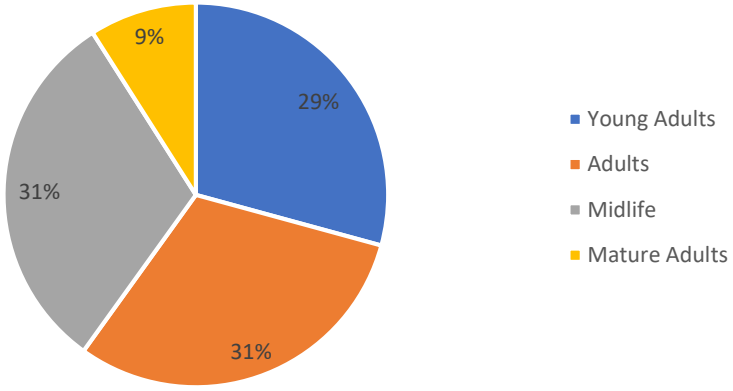
These limitations shadowed a few objectives and had a footprint on the results of the project.

3.3 The Database

The survey went public on May 14, 2018 and was closed three weeks later on June 4, with a total of 342 answers. Of the total respondents, 52 were students, thus removed from the database, as students were out of the project's scope. From the 290 non-student answers, 13 observations were removed due to incompletion of the questionnaire (had only 10 questions answered), arriving to 277 suitable observations, this being the final number of observations.

The dataset is composed of 81 Young Adults, 85 Adults, 86 “Midlifers” and 25 Mature Adults, as illustrated in Figure 5.

Figure 5 – Database composition



Source: Survey

The main features of the sample are:

- 93% of the Young Adults started working in the last five years. 80% acquired a bank account when they were between 15 and 20 years old and half of them don’t have a credit card. Of those who have, 74% acquired during the last five years.
- Almost 60% don’t have any savings product purchased from the bank, while the remaining 40% that have bought one during the last five years. Still, Young Adults don’t spend much on financial products.

The idea of transition between stages arises from the description above, in which most products were acquired recently, and an important event occurred in the last five years of almost every Young Adult.

- It is possible to see a few differences in the Adult population picture. 75% already have a credit card and 25% have house loans (almost a 25% increase from the previous stage). From these 25% who have house loans, that is, 20 respondents, 15 bought the house in the last five years, where 11 of those 15 respondents also married during the last five years.

- Only 15% have investments P&S and 77% have purchased savings products, of which 50% have been acquired during the last five years.

In terms of events, although existing more married people than the stage before, the most common event between Adults is the first job.

- For Midlife, the consumption behaviour is certainly different from the previous stages. 90% have acquired credit cards, 92% are married and 90% own a house, where 72% of the respondents have house loans.
- More than 30% have investment services purchased on the bank (the double of the previous stage) and 70% have savings. Regarding insurance P&S, this is the stage that acquires more insurance services, with 77% having life insurance and 80% having car insurance.

Hence, it is possible to notice a very different consumption pattern and, therefore, financial needs, with respect to the other life stages.

- Finally, the Mature Adults show a distinct consumption pattern. While 40% have investments, less than 50% have house loans. In terms of savings, half of the mature adults have a savings account and 92% have car insurance. The most common event is retirement.

4. Modelling Life Cycle Financial Needs

To answer the research question, the procedure is divided into three parts: i) Data Preparation; ii) Life Cycle-Financial Needs relationship research; iii) “Best Offer” Model

4.1 Data preparation

Data Preparation is the process of collecting, cleaning, and consolidating data into one file or data table, primarily for use in analysis, being that the observations are composed of the answers to the survey.

4.1.1 Data cleansing

After the data has been collected, the first step on preparation is the data cleansing. To cleanse the data, firstly was needed to extract the survey answers

to Excel. Then, answers were organized, clearing unnecessary white spaces and non-relevant information/data. Secondly, the observations that did not prove to be relevant were removed. Afterwards, with the help of histograms and scatterplots, the outliers were identified. The inconsistent outliers, such as absurd observations (i.e., people with 25 years married for 20 years), were removed. The final process of cleansing the data was dealing with missing data. Two methods were carried out – Missing Data Deletion and Missing Data Imputation.

The Missing Data Deletion implies the deletion of unusable observations, such as observations with too many missing values. There were 13 observations within the dataset composed of mostly null answers. Those observations were removed from the data.

The Missing Data Imputation is a method for substituting null values with other values. There are single and multiple imputation methods. Single imputation processes include the mean imputation, median imputation and the regression imputation. The method used in this work was the Mean Imputation.

The mean imputation is a process aiming "[…] to replace each missing value with the mean of the observed values for that variable", (Gelman, 2007).

After arriving at the final number of observations fit to start the analysis (277), there were still missing values, although of a not relevant size. Thus, each missing value was imputed by the mean value for that variable within the correspondent life stage.

4.1.2 Transforming and blending data

The second step is to transform and blend the data. In order to research the Life Cycle – Financial Needs topic and to develop the "Best Offer" Model, the data must be restructured, that is, reshaped to another data type – Data can be of different types, Categorical or Numerical (Data Types in Statistics, Towards Data Science). Categorical data includes nominal (generally data in form of text) and ordinal data (values representing discrete and ordered units without mathematical significance, such as ranks, qualifying statements resorting to numbers). Numerical data can be discrete (number of cards within a deck) or

continuous (measurements, such as height or weight). The survey data is composed mostly of nominal data and a few procedures carried out in the next chapters, required numerical or ordinal data, thus resorting to the transformation of the dataset.

The transformation and blending operations performed throughout this work are the conversion of the nominal data into numerical and categorical data and the creation of new variables from the answers to the survey.

Three procedures were applied to transform data:

Answers of Yes/No questions were assigned 1 or 0 and answers composed of ranges were assigned their range midpoint. The answers for the products and event questions were assigned their midpoint (less than 5 years were appointed with 2.5, 5-10 years were appointed 7.5), but 0 if the answer was negative (or N/A). The answers for the question "*Se tem cartão de crédito, com que regularidade o utiliza*" (If you have a credit card, how often do you use it?) were transformed into ranks, in which 1 was the first and 5 was the last answer. Finally, for questions with more than 2 answers, the latter were aggregated logically (*eg* if the inquired had a bachelor or superior, than 1, if not, 0). The full index is in the Appendix B.

Regarding, the creation of new variables within the dataset, these resulted from operations between existing ones, such as subtracting the age of the respondent's dependents to the age of the respondent, which yielded the new variable "Age (of the respondent) when the dependant was born".

4.2 Life Cycle and Financial Needs

For the second part of the methodology, to study the relationship between the Human being cycle of Life and the individual's financial needs, data was explored with the aid of Alteryx tools, a Data Analytics software, which produces statistical summaries of the data, correlation matrices, histograms, scatterplots, and frequency tables.

The Alteryx tools used to produce the next insights were "Field Summary", "Pearson Correlation", "Scatterplot" and "Frequency Table".

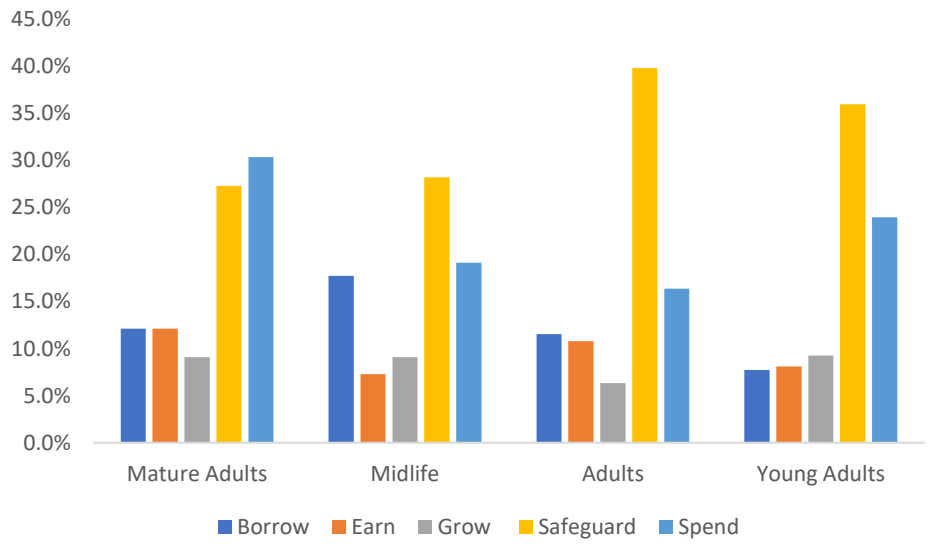
The Field Summary tool analyzes data and creates a summary report containing descriptive statistics of the data, such as mean, median and standard deviation. The Pearson Correlation tool calculates Pearson Correlation coefficient, thus measuring the linear dependence between two variables as well as the covariance.

«The purpose of correlation analysis is to measure and interpret the strength of a linear or nonlinear relationship between two continuous variables (…) For example, when the value of the predictor is manipulated (increased or decreased) by a fixed amount, the outcome variable changes proportionally (linearly)», (Silverman, Tuncali and Zou, 2003).

The Frequency Table tool produces a frequency analysis for selected fields.

As stated previously, people in different life stages have different needs, in particular different financial needs. The dataset reflects this idea, as every result and output of this chapter is based on the dataset. The graph below was originated with the aid of the previous tools and it shows the importance of each class of financial needs for every life stage:

Figure 6 – Financial needs per life stage



Source: Survey

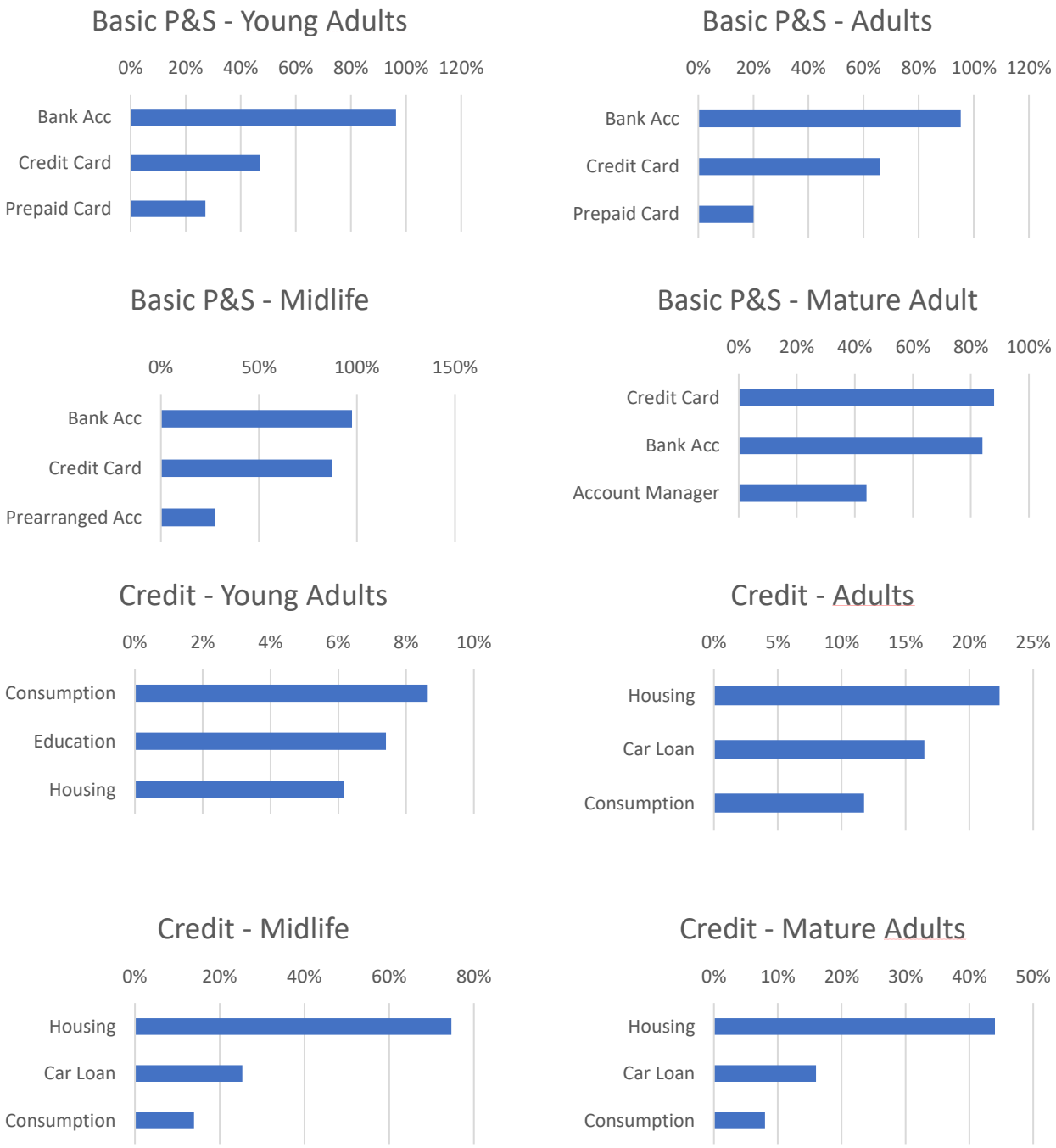
These new variables (Borrow, Earn, Grow, Spend and Safeguard) were calculated assigning a financial product to the correspondent need that it satisfies.

Figure 6 expresses that the Young Adults show more concerns on spending and safeguarding, versus the other needs, being small the percentage of those who have needs of borrowing and earning. This confirms the theory that young adults start having more responsibilities (start to support themselves), but also have more money available, since the first job is within this life stage, for most cases (see the next figure). It is also in this life stage that the values for “Grow” are higher (9.3%), meaning that young adults worry about saving and worry about the future. The “Borrow” necessity gains relevance in the Adult stage, being even higher during Midlife, corresponding, to credit services for buying houses, cars or goods, among others, possibly due to the occurrence of events such as marriage or childbirth.

Conversely, Mature Adults do not need to safeguard as much as people in the younger stages. Instead, they acquire products and services related to spending and investing.

The figures below show the three products most owned within life stages, for basic financial products and credit P&S.

Figure 7 – Basic and credit P&S owned by the respondents, in %



Source: Survey

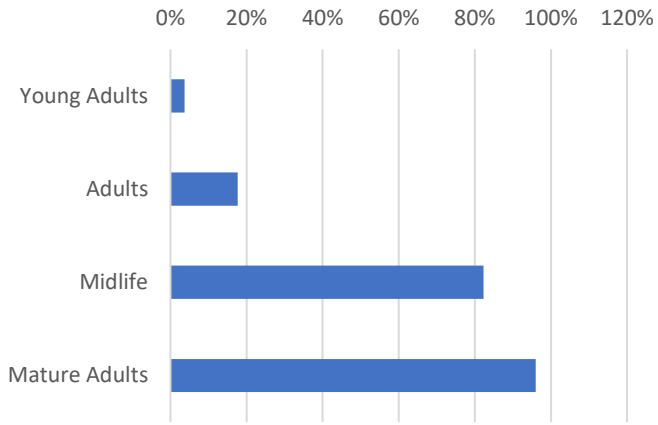
Credit cards and standard bank accounts are the most common products sold by banks, and this fact is confirmed by Figure 7. While most Young Adults and Adults need prepaid cards, for “Midlifers”, the need is absolutely different. People within this stage need to purchase bank accounts that have specific features, such as

the possibility of salary advances. For Mature Adults, the Figure confirms the theory. Their daily expenses are reduced, since their children start living on their own and loan charges start to decrease, therefore having more money available. While other needs cease to exist, with more money available, other needs arise, such as wealth management. Figure 7 illustrates this argument, with more that 40% of Mature Adults having an account manager.

Figure 7 also confirms the reasoning presented previously. It is possible to capture the behaviour towards the credit services, especially when it comes to borrowing money to buy a house. Young Adults don't resort to credit, with only 9% having borrowed for consumption. In the Adult phase, the housing loans start to have significance, with 22% having incurred in debt to buy a house. But is the transition into the midlife stage that is critical, regarding housing loans. 75% of "Midlifers" have already purchased housing loans. Accordingly, the difference between married Adults and married "Midlifers" on Figure 8 is similar to the housing loans numbers on Figure 7.

To corroborate the fact mentioned above, the numbers of Mature Adults who have loans is significantly lower than "Midlifers", thus having lower expenditures, at least on this level.

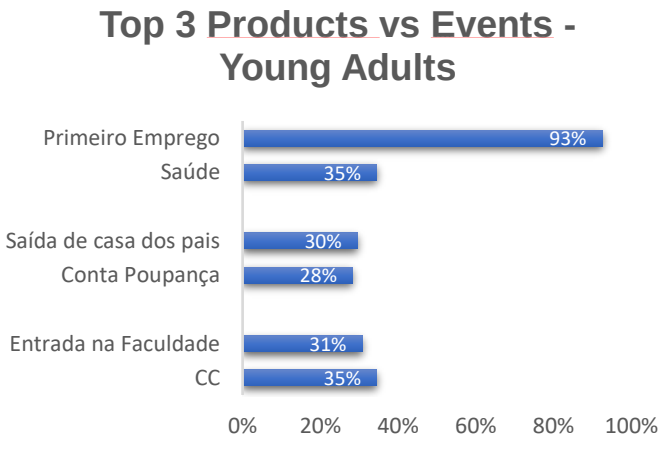
Figure 8 – Married respondents, per life stage, in %

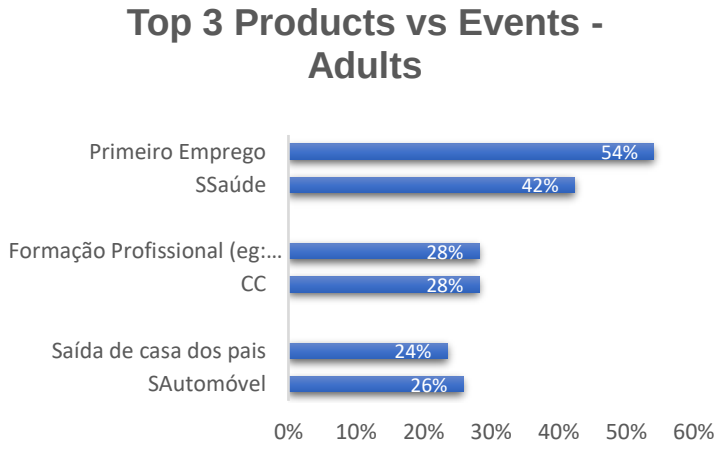


Source: Survey

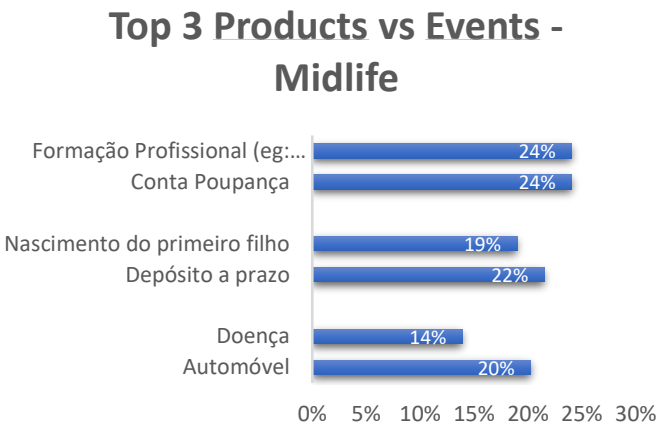
Figure 9 aids in this analysis and provides insights on the relationship Life Cycle – Financial Needs.

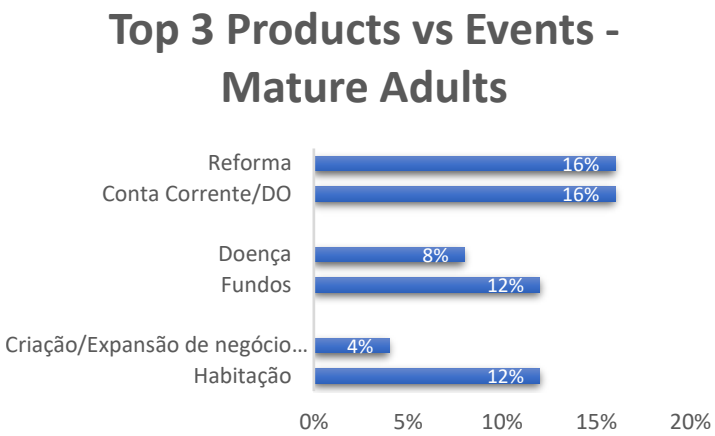
Figure 9 – Top 3 products and top 3 events, for each life stage, in %





Translation notes: Young Adults: First job/Health insurance; Move out of parents' house/Savings Account; Admission to University/Credit Card. Adults: First job/Health insurance; Training/Credit Card; Move out of parents' house/Car insurance.





Source: Survey

Translation notes: Midlife: Training/Savings Account; Birth of first child/Term deposit; Disease/Car insurance.
Mature adults: Retirement/Checking account; Disease/Investment Fund; Own

Figure 9 compares the events that happened in the last five years in each respondent's life, and, cumulatively, the latest events of each life stage. Also, the three most bought P&S during the last five years.

For Young Adults, after getting their first job they bought, a savings account and a health insurance (maybe part of the job contract). But for Midlife, the picture changes. Since 24% studied to achieve higher education levels and 19% had a child, the products they bought were majorly savings products. For Mature Adults, the latest events are retirement, generating needs of investing their life savings and updating their banking account to a most appropriate one. The bar graph illustrates this argument.

It is easy to understand that, in fact, there are life stages, during everyone's life cycle. Stages that depend on multiple factors and variables. Even though people with the same age, the same income or with the same number of children might not belong in the same life stage, their life follow patterns, in the form of events, decisions or consumptions. With this study, it is possible to conclude that there are correlations (see Appendix C) within people's behaviours, especially between life events, needs and financial needs, and decisions of consumption.

The next step is to model and to predict which financial products and services will a certain person need, based on his/her life stage, life events, demographics and his own subscriptions of financial P&S.

4.3 The "Best Offer" Model

The purpose of the "Best Offer" Model is to predict the short-term financial needs of a bank costumer, taking into account the client's profile and financial life cycle stage, information on his/her demographics and past events and also his/her client profile, i.e. which P&S he/she already owns.

The model was developed by capturing different consumer behaviours within the dataset, with resource to statistical procedures of Cluster Analysis. There are several algorithms and methods of clustering data, such as mean-shift, Gaussian Mixture Models and K-Means. Although these methods were fit to the scope of this paper, the K-Means was chosen mainly due to its simplicity and for being optimized in the Alteryx software.

The k-means clustering aims to partition the sample (*n*) into *k* clusters, in order to minimize the within-cluster sum of squares, i.e. the variance. The objective is to find

$$\arg\min_{\boldsymbol{S}} \sum_{i=1}^{k} \sum_{x \in S_i} ||x - \mu_i||^2 = arg\min_{\boldsymbol{S}} \sum_{i=1}^{k} |S_i| \, Var \, S_i$$

Where μ_i is the mean of points in S_i.

The cluster analysis identified different consumer clusters and, as referenced before, it was performed with the k-means clustering method, in Alteryx.

"(…) The main idea behind these techniques is the minimization of a certain criterion function usually taken up as a function of the deviations between all patterns from their respective cluster centers. Usually, the minimization of such criterion function is sought utilizing an iterative scheme which starts with an arbitrary chosen initial cluster configuration of the data, then alters the cluster

membership in an iterative manner to obtain a better configuration. (…) A K-Means algorithm alternates between two major steps until a stopping criterion is satisfied."

In Selim & Ismail (1984)

The method for the model development is comprised of several steps:

The first issue addressed was the variable reduction. The dataset includes 80 original variables – the survey questions - plus 25 new Yes/No variables, whether the respondent have/have had (1) or doesn't have (0) a determined financial P&S. After performing tests utilizing every variable in the dataset, results returned very poor, implying that 105 variables for 277 observations was too much. Therefore, a selection of variables was required.

To reduce the number of variables and, at the same time, input the most relevant ones, regressions would be run on the newly created variables, to conclude which old ones would be related with the dependent variable (the dependent variables of these regressions were the newly created variables) and therefore, be selected to the cluster analysis. Since the dependent variable is dichotomous, the choice was between the logit and the probit regressions. With the aid of the Alteryx Distribution Analysis tool, a tool that determines which distributions better fit the data, it was possible to conclude that the data does not follow a normal distribution.

Since the probit regression model assumes the underlying distribution is normal (Probit Regression, IDRE), the logistic model was chosen to carry out the regressions. Logistic regression sometimes called the logistic model or logit model, analyzes the relationship between multiple independent variables and a categorical dependent variable, and estimates the probability of occurrence of an event by fitting data to a logistic curve. Binary logistic regression is typically used when the dependent variable is dichotomous and the independent variables are either continuous or categorical (Park, 2013). The logit equation is as follows:

$$\rho_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1,i} + \beta_k x_{k,i})}}$$

The tests were run to every new variable and the three most relevant to each P&S were chosen to perform the clustering analysis. Two examples can be seen in Figures 10 and 11. For instance, on Figure 10, it is possible to see that the three most relevant variable are the "New_Fundos_de_Investimento" (Participation in investment funds), "Género" (Gender) and "Número de filhos" (Number of children).

Figure 10 – Logit regression on "Housing loans"

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-4.0722858	0.5997851	-6.790	1.12e-11***
Género.	1.1444575	0.3888309	2.943	0.00325**
Rendimento.Líquido.Mensal.	0.0003492	0.0001233	2.831	0.00464**
Número.de.Filhos.	-0.7856046	0.2690081	-2.920	0.0035**
Saída_de_casa_dos_pais	-0.1062552	0.0464747	-2.286	0.02224*
Primeiro_Emprego	0.1228771	0.0494126	2.487	0.01289*
Criação_Expansão_de_negócio_próprio	0.0679670	0.0336331	2.021	0.0433*
New_Cartão_de_Crédito1	0.6646669	0.4295993	1.547	0.12182
New_Gestor_património_financeiro1	0.7193829	0.4110459	1.750	0.0801.
New_Fundos_de_Investimento__Obrigações__CFD_s1	1.1369313	0.3730815	3.047	0.00231**
New_Consumo1	-0.8215007	0.5070299	-1.620	0.10518
New_Depósito_a_prazo1	0.6437175	0.3770644	1.707	0.08779.
New_Plano_Poupança_Reforma1	0.6052865	0.3937899	1.537	0.12427

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial taken to be 1)

Null deviance: 315.32 on 276 degrees of freedom
Residual deviance: 230.88 on 264 degrees of freedom
McFadden R-Squared: 0.2678, Akaike Information Criterion 256.9
Number of Fisher Scoring iterations: 5

Source: Survey

Figure 11 - Logit regression on the answer "Investments in stocks"

Source: Survey

Min	1Q	Median	3Q	Max
-2.103	-0.920	-0.348	0.690	2.381

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.77351	0.3435	-8.074	6.80e-16***
Estado.civil.	1.05619	0.3535	2.988	0.00281**
Nascimento_do_primeiro_filho	0.05222	0.0194	2.691	0.00712**
New_Plano_Poupança_Reforma1	0.50552	0.3426	1.475	0.1401
New_Vida1	2.13330	0.3458	6.169	6.87e-10***

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial taken to be 1)

Null deviance: 365.6 on 276 degrees of freedom
Residual deviance: 257.29 on 272 degrees of freedom
McFadden R-Squared: 0.2963, Akaike Information Criterion 267.3

The first approach to the clustering analysis is the diagnosis. The Alteryx Diagnosis tool assesses the appropriate number of clusters to specify, given the data and the clustering method – K-Means is the method chosen in this project (K-Centroids Diagnostics Tool, Alteryx Documentation). The user sets a range of desirable clusters (within the scope of this work, the range was set between three and six clusters). The tool produces qualitative information for each number of clusters, through two indices, the Adjusted Rand and the Calinski-Harabasz. These indices serve as indicators for cluster quality. The Adjusted Rand Indice according to Santos (2009).

“The Adjusted Rand Index (ARI) is frequently used in cluster validation since it is a measure of agreement between two partitions: one given by the clustering process and the other defined by external criteria.”

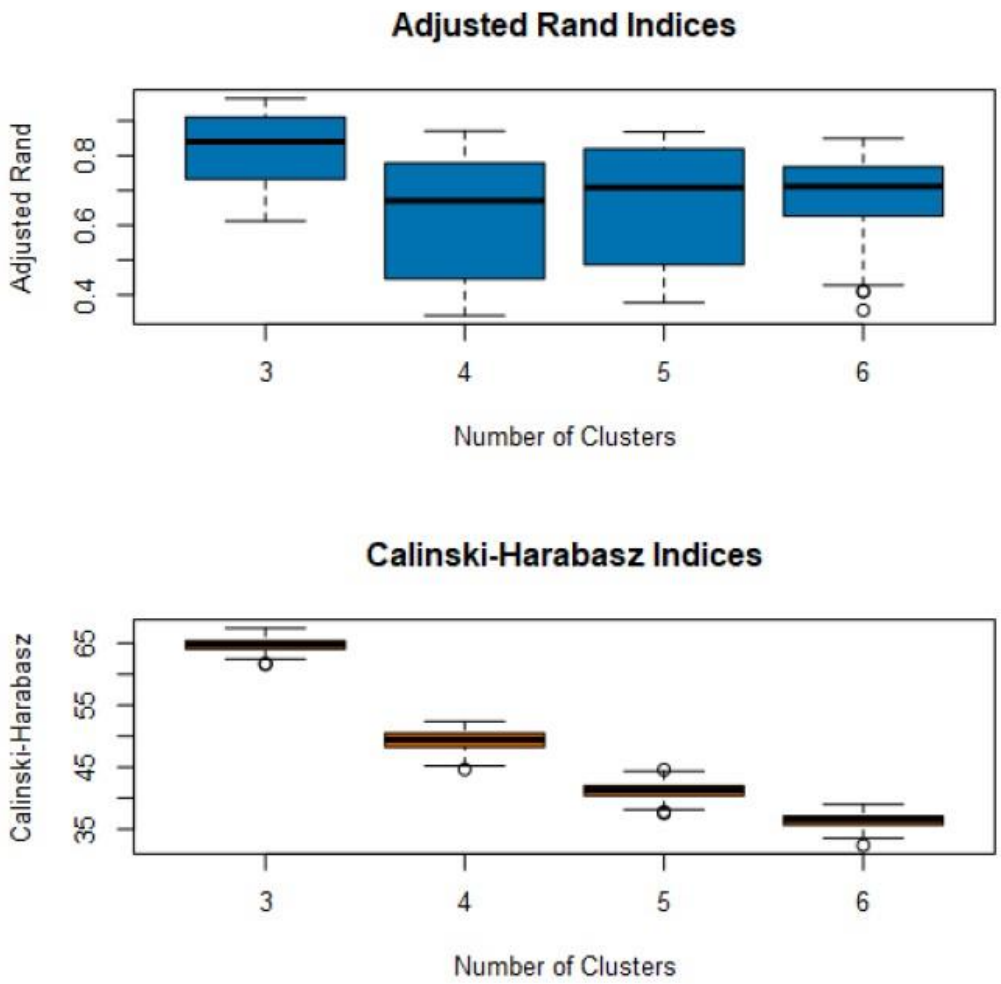
Further, Yeung, Ka Yee. and. Ruzzo, Walter L. (2001), explain that

“The Rand index lies between 0 and 1. When the two partitions agree perfectly, the Rand index is 1.”

The Calinski-Harabasz, according to Liu, Y., Li, Z., Xiong, H., Gao, X., & Wu, J. (2010, December) "[…] evaluates the cluster validity based on the average between- and within cluster sum of squares."

For both indices, the greater their average value and the lower the interquartile range, the best the number of clusters fit the data (Desgraupes, 2013). As stated in Figure 12, the best number of clusters for the dataset was three.

Figure 12 – Partition indicator: cluster analysis

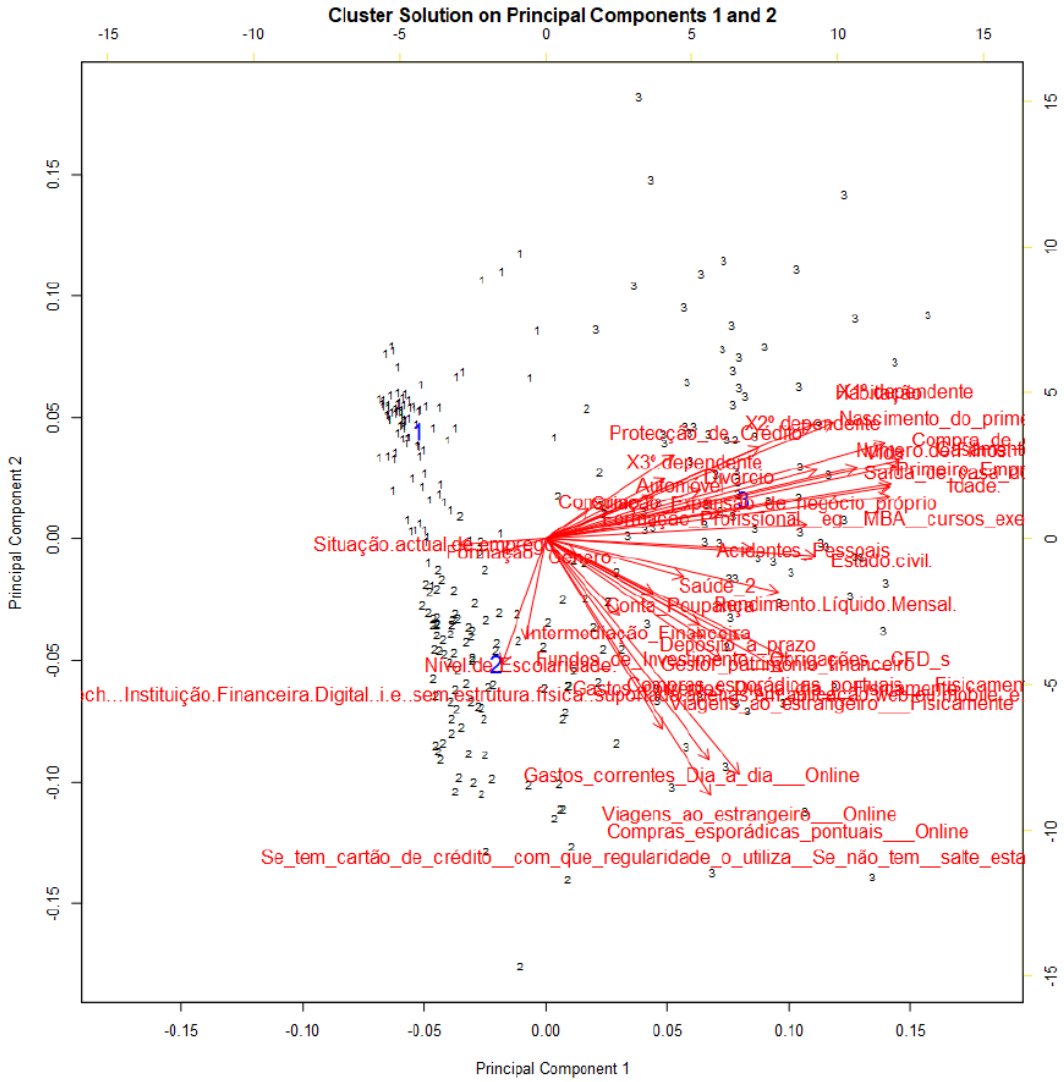


Source: Survey

In both indices, the greatest value observed had respect to three clusters, meaning that a partition of the data in three implied that there were different patterns with significant differences between them. Figure 12 was the tentative that exhibited the best values (See Annex D for one of the failed tests).

After selecting the appropriate number of clusters, the K-Means algorithm was run and the three clusters obtained – Figure 13.

Figure 13 – Cluster solution on principal components 1 and 2



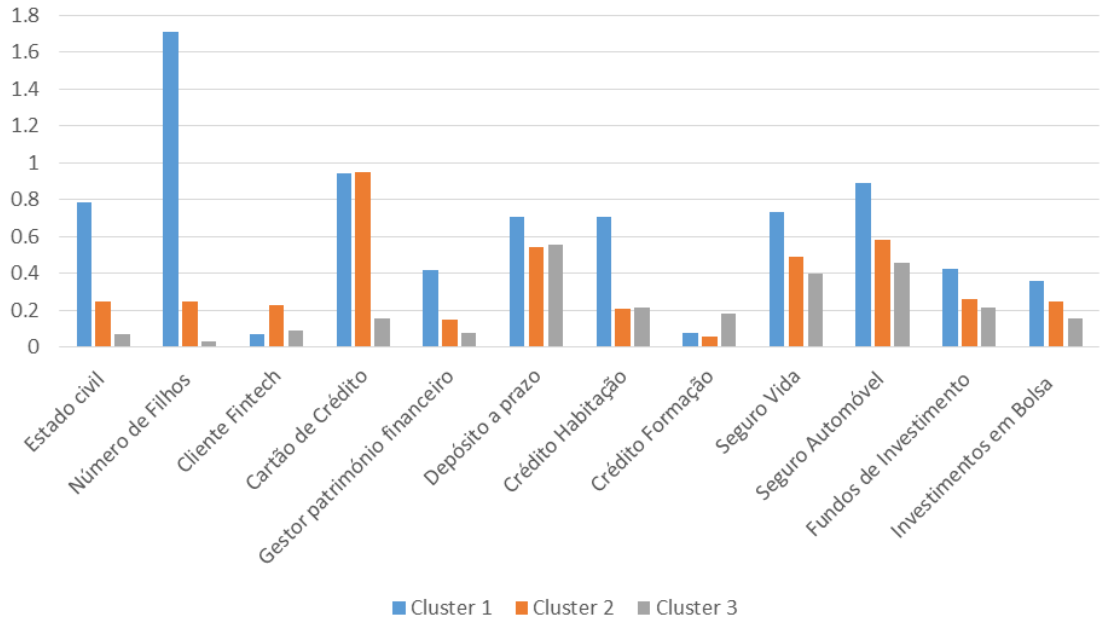
Source: Survey

Regarding Figure 13, it is possible to notice the three clusters. Cluster 1 and Cluster 2 are significantly condensed, which is very positive, and also well separated. Cluster 2 shows a bigger level of dispersion than Cluster 1, although still relatively small. Cluster 3 is quite dispersed throughout the data, although very separated from Clusters 1 and 2, meaning that there really are different consumption and demographic patterns among the database. Cluster 1 and 2 depict clear patterns, well concentrated, while cluster three doesn't portray a clear, well defined pattern, rather consumption behaviours different from most of the observations, including a few outliers.

The final step before the model development was appending the clusters to the database, that is, each inquire was allocated to the correspondent cluster, performed with the aid of the Alteryx Append Cluster tool (the tool appends the observation with the respective cluster, embedding into the dataset).

The three clusters obtained show different demographic characteristics and different consumption patterns. Cluster 1 include 89 observations, while 2 and 3 have 100 and 88, respectively.

Figure 14 – Clusters characteristics



Source: Survey

Cluster 1 is 80% married, with more than one child. Uses the credit card actively, as Cluster 2, and 40% have portfolio managers. Almost 100% have a car (Seguro Automóvel), as opposed to Clusters 2 and 3, with 60% and 40% having a vehicle, respectively. Regarding investments, Cluster 1 invests more, although the difference is not relevant. Cluster 1 exhibit a strong consumption behavior, with more P&S subscriptions than Clusters 2, which also has higher values than Cluster 3.

Regarding the model *per se,* it comprises three phases:

Phase One: The Score model - in which specific information of an individual is inserted, such as marital status, number of children, events etc, and through multipliers associated with each input, the model places the individual into the matching cluster.

Phase Two: The Crossmatch – the model crossmatches the cluster typical P&S wallet with the individual's own products, resorting to excel formulas, flagging products or services that the client doesn't have.

Phase Three: The Output - the model crossmatches the cluster average age of acquisition of such products (new variable calculated previously) with the age of the individual. The model outputs the flagged P&S with average age of acquisition is within the interval "age of the individual - 5 < age of the individual + 5", being so a probable future need for the client, thus optimizing his/her propensity for consumption.

With this model, financial institutions can better understand and, most importantly, can anticipate their clients' needs. This knowledge will allow them to personalize the offer of P&S to each clients in a more adequate way, thus improving their customer service and their relationships with customers.

5. Results and Conclusions

The present work addresses an analysis of the behaviour of the consumer of financial P&S, an issue with increasingly importance in the financial industry, particularly within the banking sector.

The purpose of the work is therefore the conceptualization of the aspects of one's life related to the need to purchase financial P&S, and is also the development of a predictive model which aids in the estimation of a person's future financial necessities (and in the optimization of the propensity to consume financial products and services). This project's scope is in accord with the KPMG Advisory guidelines and business model.

Regarding the development of the project, several tools were employed:

The online survey tool, SurveyMonkey, whose purpose was to obtain a solid database, providing the foundation for all analysis procedures. The MS Office tools, Excel, PowerPoint and Word, provided with the interface for the entire project workflow, supporting the creation of tables, graphics and figures exhibited throughout this text, as well as enabling the design and the implementation of the model. Lastly, the employment of the Alteryx software, in which all statistical operations and computation presented in this project were performed.

The survey consists of 24 questions, placed under three sections - demographic, financial and event questions.

The objective of the survey was to build a broad database, to avoid hindering the development and possible outcomes associated with the research. Only seven questions, out of the 24, were not considered relevant to the project, mostly due to the time constraints set at the beginning. The remaining questions provided quality data, therefore enabling further work.

As a side note, it is worth to mention that the removed questions could potentially complement and further the analysis to the extent of arriving to new insights and conclusions regarding the life cycle - financial needs framework.

The answers to the survey originated a database with a total of 277 observations and 62 variables (answers to the questions). The database provided support in every stage of the project, although exhibiting a few frailties, due to the small number of observations and the homogeneity of the data. Although being an early and preliminary study, a greater number of observations and a wider distribution of the survey could improve the outcomes both in quality and quantity, especially regarding the cluster analysis.

The life stage hypothesis based on the KPMG analysis was verified within the dataset. It is possible to separate the observations in four life stages, with different demographic characteristics, products bought and different needs; especially concerning the "Young Adult" stage; although being not so common within the life stage subject, it currently is a realistic concept, at least within the dataset. The difference between Adult and Child is not so clear nowadays, as young people generally study until their mid twenties, thus deserving its own stage – the “Young Adult”, as they have a specific set of needs and a particular financial behaviour.
Another important result is the KPMG financial needs study that also is verified within the dataset, in which one of its main conclusions is the relationship between the life cycle and the financial necessities, that is, the needs are different in each life stage, therefore the importance of modelling this relationship.

As for the development of the model, it started with the data transformation, in order to prepare the data for the cluster analysis. Then, due to the sensitivity of the K-Means algorithm (general to all data partition algorithms) in relation to the Variables/Observations ratio, a reduction of variables was made - the most relevant ones were chosen, taking into account financial products, using logit regressions, with these products as dependent variables.

After analyzing clusters with the chosen variables, the optimal number was in 3 partitions, with different consumption behaviours of financial products. For the purposes of the model, clusters were analyzed, in demographic terms and events that have already witnessed and chosen the main differentiating factors of each cluster. These differentiating factors constituted the inputs section of the model -

that is, it is based on these variables, multiplied by weights calculated for each cluster, that the model allocates each person to a cluster.

Finally, the current subscriptions of each person are compared to the common subscriptions of the respective cluster, and the template selects the most common cluster products that the client does not have, revealing a product that the customer is likely to need.

In this way, it is possible to state that, in view of this preliminary study, there are still several issues within the topic to be explored, such as the introduction of new factors that may influence consumer decision-making, such as risk attitudes. Another future development concerns the size of the database. A considerable size of the database, with more heterogeneous data, would improve the results obtained.

References

American Institute for Learning and Human Development. *The 12 Stages of Life.* [online] Available at: http://www.institute4learning.com/resources/articles/the-12-stages-of-life/ [Accessed 29 Mar 2018]

Alteryx Community. [online] Available at: https://community.alteryx.com/ [Accessed 15 May 2018]

Alteryx Documentation. *K-Centroids Diagnostics Tool.* [online] Available at: https://help.alteryx.com/current/K-Centroids_Diagnostics.htm/ [Accessed 20 Jun 2018]

Alteryx. *The Leading Platform for Self-Service Data Analytics.* [online] Available at: https://www.youtube.com/user/alteryx/ [Accessed 24 May 2018]

Armstrong, T. (2007), – The human odyssey. *Navigating the twelve stages of life.* New York, Sterling Publishing Co., Inc.

Bogin, B., & Smith, B. H. (1996), – Evolution of the human life cycle, || *American Journal of Human Biology: The Official Journal of the Human Biology Association*, *8*(6), 703-716.

Callanan, T. P., Guo, L. & Jeske, D. R. (2011), – Identification of key drivers of Net Promoter Score using a statistical classification model, || in *Efficient Decision Support Systems-Practice and Challenges From Current to Future*. InTech.

Desgraupes, B. (2013), – Clustering indices. *University of Paris Ouest-Lab Modal'X*, *1*, 34.

Embrechts, M. & Santos, J. M. (2009, September), – On the use of the adjusted rand index as a metric for evaluating supervised classification, || in *International Conference on Artificial Neural Networks* (pp. 175-184). Springer, Berlin, Heidelberg.

Gao, X., Li, Z., Liu, Y., Wu, J & Xiong, H. (2010, December), – Understanding of internal clustering validation measures, || in *Data Mining (ICDM), 2010 IEEE 10th International Conference on* (pp. 911-916). IEEE.

Gelman, A., & Hill, J. (2007), – *Data analysis using regression and multilevel/hierarchical models,* chapter 25. Cambridge university press.

Institute for Digital Research and Education. *Probit Regression* | *Stata Data Analysis Examples.* [online] Available at: https://stats.idre.ucla.edu/stata/dae/probit-regression/ [Accessed 16 May 2018]

Ismail, M. A. & Selim, S. Z. (1984), – K-means-type algorithms: A generalized convergence theorem and characterization of local optimality, || *IEEE Transactions on pattern analysis and machine intelligence*, (1), 81-87.

MathWorks. *Calinski Harabasz Evaluation.* [online] Available at: http://www.mathworks.com/help/stats/clustering.evaluation.calinskiharabaszevaluation-class.html/ [Accessed 14 Jun 2018]

Park, H. (2013), – An introduction to logistic regression: basic concepts to interpretation, || in *Journal of Korean Academy, 43(2),154-164*.

Ruzzo, W. L. & Yeung, K. Y. (2001), – Details of the adjusted rand index and clustering algorithms, supplement to the paper an empirical study on principal component analysis for clustering gene expression data. *Bioinformatics*, *17*(9), 763-774.

Slater, C. L. (2003), – Generativity versus stagnation: An elaboration of Erikson's adult stage of human development, || *Journal of Adult Development*, *10*(1), 53-65.

Towards Data Science. *Data Types in Statistics.* [online] Available at: https://towardsdatascience.com/data-types-in-statistics-347e152e8bee/ [Accessed 23 Apr 2018]

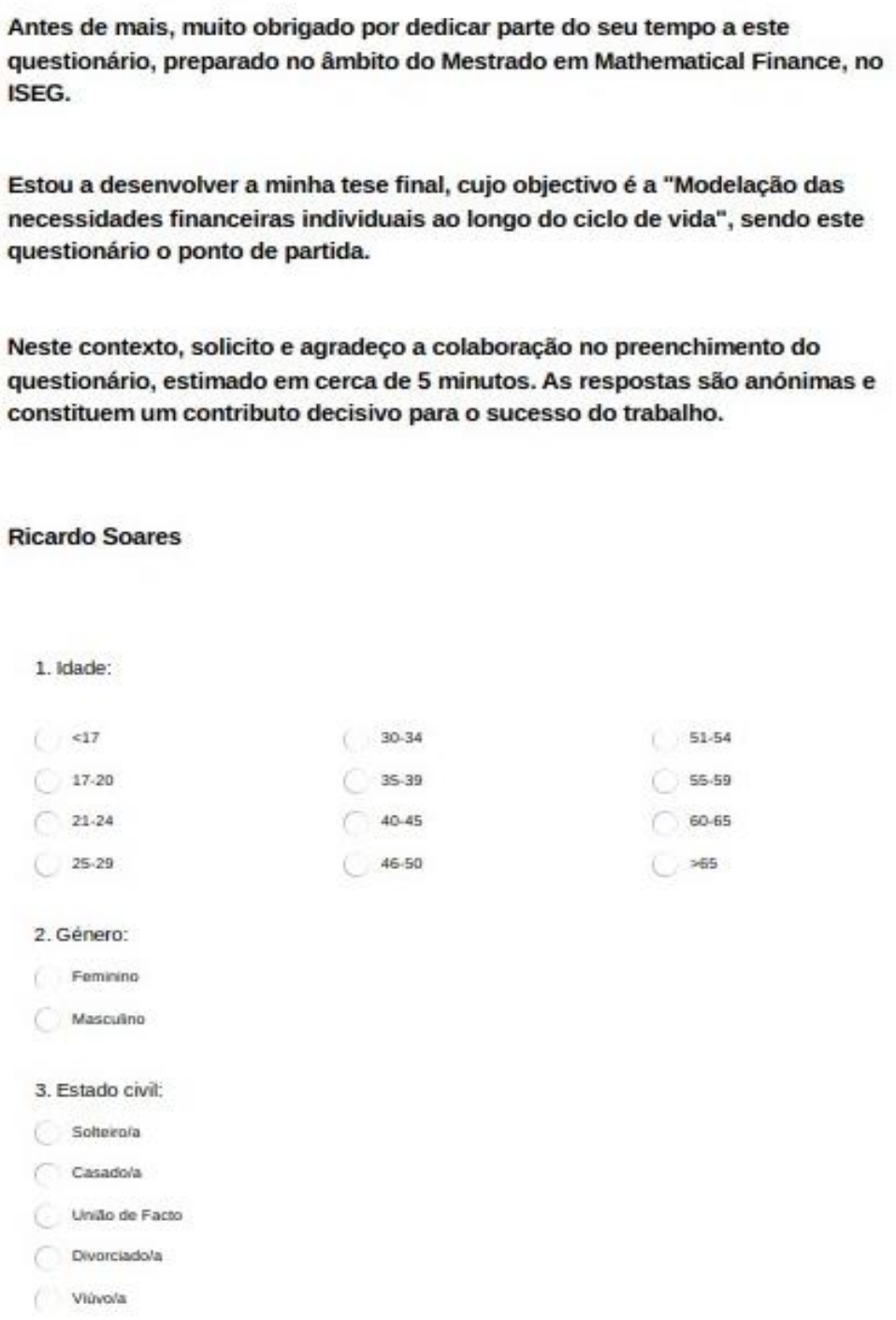
Silverman, S.G., Tuncali, K. & Zou, K.H. (2003), – Correlation and simple linear regression. *Radiology, 227 (3),* 617-628.

.

Appendix

Appendix A. Survey

Figure 15 - Survey



Antes de mais, muito obrigado por dedicar parte do seu tempo a este questionário, preparado no âmbito do Mestrado em Mathematical Finance, no ISEG.

Estou a desenvolver a minha tese final, cujo objectivo é a "Modelação das necessidades financeiras individuais ao longo do ciclo de vida", sendo este questionário o ponto de partida.

Neste contexto, solicito e agradeço a colaboração no preenchimento do questionário, estimado em cerca de 5 minutos. As respostas são anónimas e constituem um contributo decisivo para o sucesso do trabalho.

Ricardo Soares

1. Idade:

<17	30-34	51-54
17-20	35-39	55-59
21-24	40-45	60-65
25-29	46-50	>65

2. Género:

- Feminino
- Masculino

3. Estado civil:

- Solteiro/a
- Casado/a
- União de Facto
- Divorciado/a
- Viúvo/a

4. Nacionalidade:

- ( ) Portuguesa
- ( ) Outra - Qual?

5. Nível de Escolaridade:

- ( ) Secundário ou Inferior
- ( ) Licenciatura ou Equivalente
- ( ) Mestrado ou Superior

6. Situação actual de emprego:

- ( ) Empregado/a por conta de outrém
- ( ) Empregado/a por conta própria
- ( ) Desempregado/a
- ( ) Trabalhador-estudante
- ( ) Estudante
- ( ) Reformado/a

7. Profissão/Ramo:

8. Rendimento Líquido Mensal:

- ( ) <500€
- ( ) 500€ - 999€
- ( ) 1000€ - 1999€
- ( ) 2000€ - 2999€
- ( ) 3000€ - 4999€
- ( ) 5000€ - 10000€
- ( ) >10000€

9. Residência (Distrito):

2

10. Número de Filhos:

- 0
- 1
- 2
- 3
- 4 ou mais

11. Número de dependentes:

- 0
- 1
- 2
- 3
- 4 ou mais

12. Idade dos dependentes:

	0-5	6-10	11-16	17-21	22 ou mais
1º dependente					
2º dependente					
3º dependente					
4º dependente					

13. De quantas instituições bancárias era/é cliente?

	0	1	2	3	4 ou mais
Há 5 anos:					
Actualmente					

14. É cliente de alguma instituição bancária internacional? (Fora de Portugal)

- Sim
- Não

3

15. É cliente de alguma FinTech? (Instituição Financeira Digital)

(i.e. sem estrutura física, suportada apenas em aplicação web ou mobile, e.g.: Revolut)

- Sim
- Não

Se sim, qual?

16. Hierarquize consoante o peso que tem no seu orçamento:

(Sendo 1 o que tem maior peso e 7 o que tem menor)

	1	2	3	4	5	6	7
Casa e alimentação							
Transporte							
Educação							
Saúde							
Lazer/Entretenimento/Hobbies							
Pagamento de Dívida não-corrente/Outras dívidas							
Poupança/Investimento							

17. Há quanto tempo adquiriu os seguintes Produtos Financeiros?

	Menos de 5 anos	Entre 5 a 10 anos	Entre 10 a 15 anos	Entre 15 a 20 anos	Mais de 20 anos	Não tenho
Conta Corrente/DO						
Cartão de Crédito						
Cartão Pré-Pago						
Solução MultiProduto						
Conta Antecipação de ordenado						
Gestor património financeiro						

4

18. Há quanto tempo tem os seguintes Investimentos/Participações?

	Menos de 5 anos	Entre 5 a 10 anos	Entre 10 a 15 anos	Entre 15 a 20 anos	Mais de 20 anos	Não tenho
Fundos de Investimento, Obrigações, CFD's	○	○	○	○	○	○
Bolsa (Acções, ETF's, Futures, Certificados)	○	○	○	○	○	○
Produtos Estruturados	○	○	○	○	○	○
Conta Moeda Estrangeira	○	○	○	○	○	○
Intermediação Financeira	○	○	○	○	○	○

19. Se tem cartão de crédito, com que regularidade o utiliza?

(Se não tem, salte esta e a próxima pergunta)

- ○ Diariamente
- ○ Semanalmente
- ○ Mensalmente
- ○ Semestralmente
- ○ Anualmente

20. E em quê?

	Fisicamente	Online
Gastos correntes/Dia-a-dia		
Compras esporádicas/pontuais		
Viagens ao estrangeiro		

Outro (especifique)

21. Há quanto tempo contraiu os seguintes empréstimos/serviços de Crédito?

	Menos de 5 anos	Entre 5 a 10 anos	Entre 10 a 15 anos	Entre 15 a 20 anos	Mais de 20 anos	Não tenho
Habitação	○	○	○	○	○	○
Automóvel	○	○	○	○	○	○
Consumo	○	○	○	○	○	○
Formação	○	○	○	○	○	○

5

22. Há quanto tempo adquiriu os seguintes produtos de Poupança?

	Menos de 5 anos	Entre 5 a 10 anos	Entre 10 a 15 anos	Entre 15 a 20 anos	Mais de 20 anos	Não tenho
Depósito a prazo						
Conta Poupança						
Plano Poupança-Reforma						

23. Há quanto tempo adquiriu os seguintes Seguros?

	Menos de 5 anos	Entre 5 a 10 anos	Entre 10 a 15 anos	Entre 15 a 20 anos	Mais de 20 anos	Não tenho
Habitação						
Vida						
Saúde						
Automóvel						
Acidentes Pessoais						
Protecção de Crédito						
Emp. Doméstica						

24. Em que momento da sua vida se deram os seguintes acontecimentos?

	Nos últimos 5 anos	Há 5 a 10 anos	Há 10 a 15 anos	Há 15 a 20 anos	Há mais de 20 anos	N/A
Entrada na Faculdade						
Saída de casa dos pais						
Primeiro Emprego						
Casamento						
Compra de casa						
Nascimento do primeiro filho						
Formação Profissional (eg: MBA, cursos executivos, etc)						
Emigração						
Criação/Expansão de negócio próprio						
Doença						
Divórcio						
Reforma						

Source: Author 6

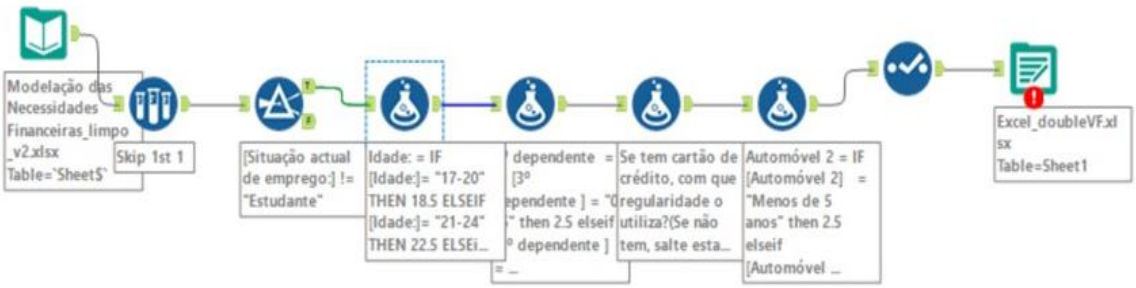
Appendix B. Data type transformation

Figure 17 – New data type of the survey's answers

Question number	New Type	Classification
1, 8, 12, 17, 18, 21, 22, 23, 24	Numerical	MidPoint
2, 3, 4, 5, 6, 9, 14, 15, 20	Numerical	1/0
10, 11, 13, 16, 19	Categorical	Ranking

Source: Author

Figure 18 – Alteryx workflow for changing the data type



Source: Author

Appendix C. Correlation Table

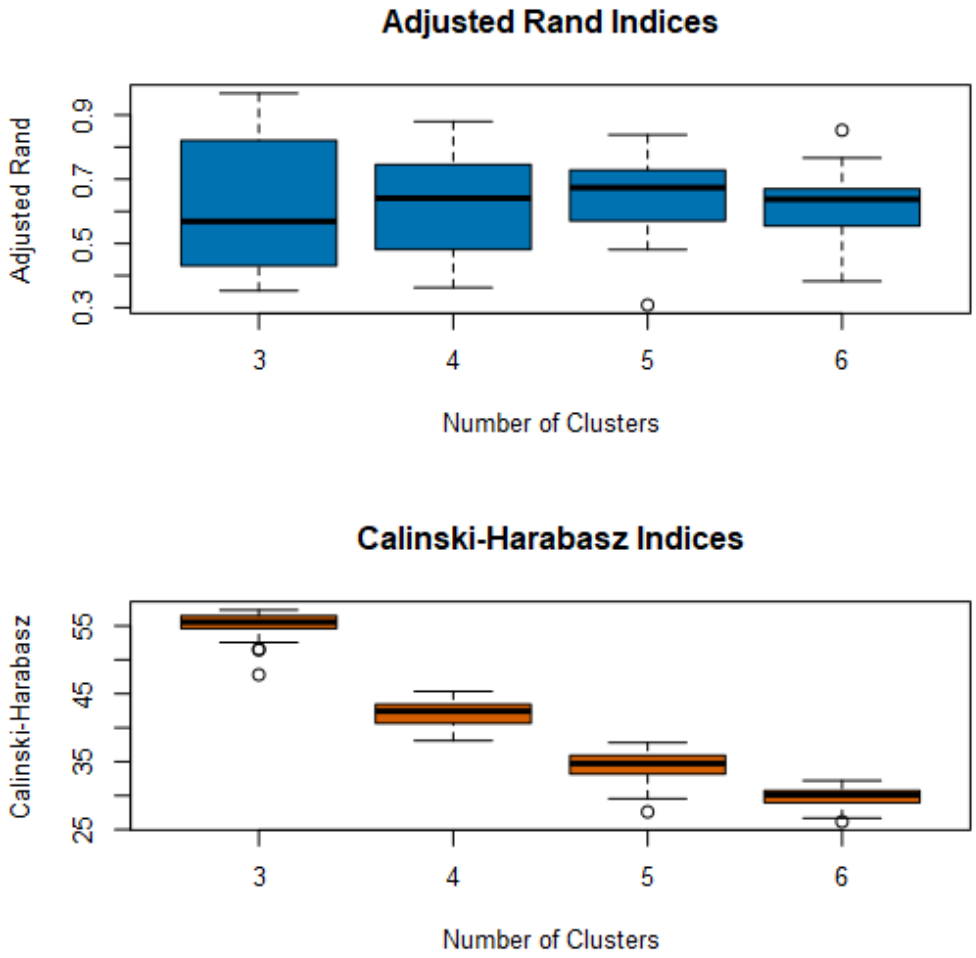
Figure 20 – Correlation table

Variável	Conta Corrente	Cartão de Crédito	Cartão Pré-Pago	Solução MultiProduto	Conta Antecipação de ordenado	Gestor património financeiro	Fundos de Investimento	Bolsa	Produtos Estruturados	Conta Moeda Estrangeira	Intermediação Financeira	Regularidade no uso de cartão de crédito	Gastos correntes Dia a dia - Fisicamente	Gastos correntes Dia a dia - Online	Compras esporádicas pontuais - Fisicamente	Compras esporádicas pontuais - Online	Viagens ao estrangeiro - Fisicamente	Viagens ao estrangeiro - Online	Crédito Habitação	Crédito Automóvel	Crédito Consumo	Crédito Formação	Depósito a prazo	Conta Poupança	Plano Poupança Reforma	Seguro Habitação	Seguro Vida	Seguro Saúde	Seguro Automóvel	Seguro Acidentes Pessoais	Seguro Proteção de Crédito	Seguro Emp Doméstica
Entrada na Faculdade	0,55658	0,69818	0,0728	0,305817678	0,300266267	0,348291	0,350234989	0,3563	0,274777093	-0,0395029	0,145186744	0,33105615	0,23798254	0,17473	0,40079493	0,34360133	0,39572015	0,2598657	0,5224632	0,1640832	0,0395429	-0,111414	0,321265	0,2224103	0,4727259	0,6435287	0,545	0,2439	0,5279394	0,324564	0,2327414	0,36204
Saída de casa dos pais	0,53481	0,69477	0,1482	0,274592971	0,337264983	0,3895841	0,298300036	0,32083	0,216151374	0,08586711	0,148439623	0,28188985	0,27619024	0,124662	0,46009443	0,30673359	0,38664627	0,2473325	0,621651	0,1889516	0,0981031	-0,113985	0,327636	0,1631334	0,4725007	0,7754391	0,57	0,2241	0,6994161	0,376794	0,2472568	0,3695726
Primeiro Emprego	0,55833	0,63905	0,0803	0,287732759	0,393097675	0,3803517	0,31927804	0,38201	0,261824391	0,00896377	0,13312979	0,26279401	0,25949672	0,159292	0,43604669	0,29598654	0,4170452	0,2222417	0,622147	0,2118273	0,013541	-0,139466	0,330098	0,1745128	0,5131964	0,7884107	0,584	0,2255	0,7296184	0,373704	0,2358939	0,3483219
Casamento	0,38662	0,67243	0,057	0,288505119	0,475890192	0,4473851	0,290875324	0,325	0,269055893	-0,0035689	0,180743823	0,22902571	0,2350413	0,172819	0,39971334	0,33697466	0,39486891	0,2270537	0,6628017	0,2458364	0,0660833	-0,122238	0,232955	0,1669305	0,4209129	0,7767537	0,575	0,2388	0,6599053	0,447436	0,2964143	0,4055228
Compra de casa	0,47741	0,70836	0,1647	0,308287323	0,434841111	0,4160455	0,332056557	0,38602	0,289370411	-0,0363316	0,168444466	0,22817658	0,24739398	0,121124	0,39196338	0,27599908	0,39548066	0,2048653	0,6365039	0,2234619	0,0366663	-0,131919	0,347098	0,1968646	0,518517	0,863632	0,615	0,2212	0,6946838	0,400668	0,2631912	0,418929
Nascimento do primeiro filho	0,36206	0,61865	0,0611	0,270897729	0,432435356	0,3499586	0,240179622	0,30404	0,208523807	-0,0099949	0,203828027	0,18181809	0,23010677	0,149552	0,34444914	0,262906	0,36448952	0,20824	0,6094401	0,2517098	0,0639837	-0,120308	0,205649	0,1293433	0,4164696	0,7614463	0,534	0,1969	0,6189328	0,388427	0,2969513	0,4666323
Formação Profissional	0,41122	0,5339	0,1187	0,224809717	0,250945455	0,450034	0,388339917	0,37723	0,31530855	0,02238002	0,246557346	0,17623111	0,14062215	0,030262	0,30744315	0,18609485	0,31514109	0,1910655	0,3917682	0,1811314	-0,003733	-0,061671	0,275622	0,1011873	0,3528981	0,5648018	0,455	0,1741	0,5404895	0,31729	0,072343	0,4583334
Emigração	0,07218	0,0523	0,0692	0,01908322	0,149478149	0,1216412	0,074526206	-0,0579	-0,053261	0,25595571	-0,050794284	-0,00704876	0,08567093	0,037775	0,03237905	-0,041834	0,02952163	0,0798954	0,0563412	-0,0527816	0,0105526	0,1053216	0,025142	-0,0501276	0,0341445	0,0226301	0,034	0,0424	0,0656181	0,037106	-0,053342	0,1021658
Criação de negócio	0,10928	0,33076	0,032	0,135504172	-0,00159824	0,1504754	0,033266573	0,14441	0,040774912	0,11789583	0,036042623	0,18012317	0,25974443	0,150317	0,20505132	0,14915579	0,20973915	0,1004199	0,2640763	0,1959084	0,1781512	0,1079711	0,132466	0,1429043	0,2125715	0,2623113	0,313	0,231	0,2936522	0,347716	0,2146177	0,3540823
Doença	0,11249	0,0432	0,0476	-0,061541685	0,051050138	0,111645	0,097818349	0,02937	0,047770996	0,03099826	0,16163086	-0,05834115	0,03265877	-0,01246	0,05252999	-0,0812363	0,00097412	-0,0107093	0,1163201	-0,0049989	-0,024011	0,0626719	0,049851	-0,071957	0,0588651	0,2198478	0,166	0,0672	0,2094717	0,02715	-0,048006	0,0277204
Divórcio	0,17681	0,31803	-0,0017	-0,030177002	0,186233196	0,1652541	0,002075763	0,078	0,089980809	-0,0220819	-0,040906624	0,12565945	0,13804446	0,058693	0,19519369	0,12768191	0,18329635	0,0414396	0,2127345	0,0837895	-0,025958	-0,064322	0,108269	0,0577237	0,0491498	0,3090069	0,287	0,0298	0,3601247	-0,002904	0,1469521	0,1301142
Reforma	0,00861	0,05633	-0,0055	0,010607043	0,055448854	0,0739847	0,025642312	0,0406	-0,008906464	0,02222051	0,02894424	-0,02516986	0,176103	0,14384	-0,0249441	0,07999557	0,00449286	0,0981016	0,1252299	-0,0010379	-0,037602	-0,031407	0,039477	-0,0694175	-0,000938	0,1614272	-0,002	-0,032	0,0871492	0,002676	-0,031849	0,0497168

Source: Author

Appendix D. Cluster Analysis

Figure 21 – Partition indicator: preliminary cluster analysis – bad attempt



Source: Author