MEGI

MESTRADO
Estatística e Gestão de Informação

Customer Lifetime Value: A Framework for Application in the Insurance Industry

Building a Business Process to Generate and Maintain an Automatic Estimation Agent

Miguel Pio Leal Henriques

Research Project submited
as partial fulfilment for the Master Degree
in Statistics and Information Management

UNIVERSIDADE NOVA
ISEGI Estatística e Gestão de Informação

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

(This page was intentionally left in blank)

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

Customer Lifetime Value: A Framework for Application in the Insurance Industry

Building a Business Process to Generate and Maintain an Automatic Estimation Agent

Miguel Pio Leal Henriques

Research Project submited
as partial fulfilment for the Master Degree
in Statistics and Information Management

Supervisor: Fernando Bação, Phd

April, 2012

(This page was intentionally left in blank)

ACKNOWLEDMENTS

To the school, Universidade Nova de Lisboa - Instituto Superior de Estatística e Gestão de Informação and its faculty, for the excellence in teaching and contributing to the Nation's progress.

I would like to express gratitude to my Supervisor, Professor Fernando Bação, for supporting this project. His immense knowledge and sense of practicability were very helpful in the development of this work. Thanks for the fascinating and interactive classes, they were a real joy.

To my Colleagues and Senior Managers for revealing and discussing many *arts and trades* of this business: Dra. Ana Luísa Simões, Dra. Ana Ferreira, Dr. Francisco Ávila, Dr. João Nunes, Dr. João Maranga, Eng. Fernando Nunes, Dra. Teresa Barreira, Dr. António Marinho, Dr. Tiago Corrêa, Dr. Marcos Lopes, Dr. David Ribeiro and many others.

My sincere appreciation to my Company's Board, Dr. Antonio Varela Afonso, Dr. Eduardo Palhares and Eng. Lino Afonso, for supporting and financing this learning adventure. Working on one of the best Companies in the market is challenging, nevertheless the organizations is being a real school for me. A nod also goes to Mónica Silva for easing the logistics and red-tape between all parties.

My gratitude also goes to Dr. João Pedro Borges for its enthusiasm and challenging mind that re-energized my interest for many disciplines and for the interesting and demanding projects I've been challenged for, one of them is in the roots of this research work.

And last but not the least, to my wife for her infinite patience and kindness.

(This page was intentionally left in blank)

ABSTRACT

In recent years the topic of Customer *Lifetime Value (CLV)* or in its expanded version, *Customer Equity (CE)* has become popular as a strategic tool across several industries, in particular in retail and services. Although the core concepts of CLV modelling have been studied for several years and the mathematics that underpins the concept is well understood, the application to specific industries is not trivial. The complexities associated with the development of a CLV programme as a business process are not insignificant causing a myriad of obstacles to its implementation. This research project builds a framework to develop and implement the CLV concept as maintainable business process with the focus on the Insurance Industry, in particular for the nonlife line of business. Key concepts, as churn modelling, portfolio stationary premiums, fiscal policies and balance sheet information must be integrated into the CLV framework. In addition, *an automatic estimation machine (AEM)* is developed to standardize CLV calculations. The concept of AEM is important, given that CLV information "must be fit for purpose", when used in other business processes. The field work is carried out in a Portuguese *Bancassurance* Company which is part of an important Portuguese financial Group. Firstly this is done by investigating how to translate and apply the known CLV concepts into the insurance industry context. Secondly, a sensitivity study is done to establish the optimum parameters strategy. This is done by incorporating and comparing several *Datamining* concepts applied to churn prediction and customer base segmentation. Scenarios for balance sheet information usage and others actuarial concepts are analyzed to calibrate the Cash Flow component of the CLV framework. Thirdly, an Automatic Estimation Agent is defined for application to the current or the expanding firm portfolio, the advantages of using the SOA approach for deployment is also verified. Additionally a comparative impact study is done between two valuation views: the Premium/Cost driven versus the CLV driven. Finally a framework for a BPM is presented, not only for building the AEM but also for its maintenance according to an explicit performance threshold.

(This page was intentionally left in blank)

RESUMO

O tema do valor embebido do Cliente (Customer Lifetime Value ou CLV), ou na sua versão expandida, valoração patrimonial do Cliente (Customer Equity), adquiriu alguma relevância como ferramenta estratégica em várias indústrias, em particular na Distribuição e Serviços. Embora os principais conceitos subjacentes ao CLV tenham sido já desenvolvidos e a matemática financeira possa ser considerada trivial, a sua aplicação prática não o é. As complexidades associadas ao desenvolvimento de um programa de CLV, especialmente na forma de Processo de Negócio não são insignificantes, existindo uma miríade de obstáculos à sua implementação.

Este projecto de pesquisa desenvolve o enquadramento de adaptação, actividades e processos necessários para a aplicação do conceito à Industria de Seguros, especificamente para uma empresa que actue no Sector Não Vida. Conceitos-chave, como a modelação da erosão das carteiras, a estacionaridade dos prémios, as políticas fiscais e informação de balanço terão de ser integrados no âmbito do programa de modelação do valor embebido do Cliente. Um dos entregáveis será uma “máquina automática de estimação” do valor embebido, essa ferramenta servirá para padronizar os cálculos do CLV, para além disso é importante, dado que a informação do CLV será utilizada noutros processos de negócio, como por exemplo a distribuição ou vendas.

O trabalho de campo é realizado numa empresa de Seguros tipo Bancassurance pertença de um Grupo Financeiro Português relevante. O primeiro passo do trabalho será a compressão do conceito do CLV e como aplicá-lo aos Seguros. Em segundo lugar, será feito um estudo de sensibilidade para determinar a estratégia óptima de parâmetros através de aplicação de técnicas de modelação. Em terceiro lugar serão abordados alguns detalhes da máquina automática de estimação e a sua utilização do ponto de vista dos Serviços e Sistemas de Negócio (e.g. via SOA). Em paralelo será realizado um estudo de impacto comparativo entre as duas visões de avaliação do negócio: Rácio de Sinistralidade vs CLV. Por último será apresentado um desenho de processo para a manutenção continuada da utilização deste conceito no suporte ao negócio.

(This page was intentionally left in blank)

INDEX

TABLE INDEX

FIGURE INDEX

ABREVIATION LIST

Citation - Citation Databases

GS - Google Scholar

DMine – Datamining

DM - datamart

DW - Data Warehouse

ML – Machine Learning

FC – Financial Calculus

CE – Customer Equity

WoS - Web of Science

LoB – Line of Business

CoC – Cost of Capital

PV – Present Value

ACRONYMS LIST

AEM – Automatic Estimation Machine

BMS – Bonus-Malus System

BPM – Business Process Management

BPE – Business Process Engineering

BSC - Balanced Score Card

CLV/CLTV – Customer Lifetime Value

PLV – Policy Lifetime Value

PCLV – Past Customer Lifetime Value

FCLV – Future Customer Lifetime Value

SOA – Service Oriented Architecture

CRM – Customer Relationship Management

RFM – Recency, Frequency, Monetary

CLM – Customer LifeCycle Management

SEMMA – Sample, Explore, Model, Modify, Assess

CRISP-DM - Cross-Industry Standard Process for Data Mining

KPI – Key Performance Indicator

TERMS DEFINITION

Bancassurance – This term refers to the business model where a bank and an insurance company coordinate a partnership where typically the insurance company uses the bank sales channel in order to push insurance products. It is very popular in Europe especially in France where the term was coined.

Process – This term, in a business context, identifies the set of activities or actions carried out by one or several entities to achieve the completion of some product or service.

Master Process – A Master Process aggregates related processes that operate and concur to the same purpose.

Strategy Map - A strategy map is a diagram used to express the main goals of an organization or team. Usually it is a document associated with the balanced scorecard methodology and its objective is to facilitate the discussion and design of the organization's strategy. Usually they show: a) Just the key Strategic objectives in the form of geometric shapes; b) Causal relationships between the objectives using arrows that can be differentiated by the strength of effect.

Uniform Policy – A uniform policy is the draft policy Framework approved by the national regulator, (e.g. ISP – Instituto de Seguros de Portugal) that dictate the requirements that all insurance policies must contain. Which include the circumstances under which changes can be made to the policy, how the beneficiary can be changed; submission of proof of loss, reinstatement of the policy, etc.

1. INTRODUCTION

"The whole is different from the sum of its parts"
Aristotle in his work Metaphysics

Can a top manager have this holistic view when he looks to his Company balance sheet? How can he explain the results whenever they're good or bad? Modern commercial organizations obviously have interest in maximizing their returns but are also a complex set of interconnected processes and activities that form the *"playground"* where workers with different responsibilities operate their trades. Doing business is the main the purpose and certainly the main source of revenue. At some level of development and in specific contexts the majority of companies may be doing business with thousands or millions of customers. Are all customers equal? Of course not, but how different are them? **We may think:** all these customers are in some way **connected with the magic numbers** the top manager sees in his balance sheet. But how this reality emerges? What impact **each customer** has, and **will** have, in the **overall result**? Is there a process to connect the chasm between these apparently disconnected realities? Trying to cut through the fog and bring some degree of **reductionism** in measuring **elemental business relations and its embedded value** is the purpose of this work. Research has provided evidence of the relation between the customer's equity and the company value. All business is dependent on the behaviour of cash flows and it is demonstrated that every cash flows can be traced back to customers. Several authors proposed the concept of Customer Lifetime Value (CLV) as a good starting point to integrate the notion of the long term lifecycle into the actual firm's value. The assumption is that a value of a firm is the sum of the each customer's individual life time value. The **possibility to measure and manage** that information to take immediate decisions on a daily basis may have a profound impact in future results, even on the firm's ability to survive. This assumption is the basis for the rationale of this research.

1.1. Research Work Rationale

The motivation to develop this research work appeared, to the author, as real a business case in the corporation he is working in. The problem was presented in this fashion: "**We** recognise the importance of the CLV concept, as an insurance company operating in a complex and competitive market, it should be implemented since it may provide better financial results". The subject matter is considered strategic, not only by the board but also by several managing directors. It is recognised that the concept behind CLV would have impact in a number of business processes and offer in the medium to long term a real edge over the competition. One of the main focus ideas is the **potential future profitability** may have a real impact in **today's decisions**. Of course maximizing profitability should be the objective function, but the process of identifying and modulating value with an eye in each customer is accepted to be a stronger and resilient foundation for business expansion. During the preliminary investigations to set up an internal project it was noticeable that, although there was some literature about the CLV methodology and its principles, there wasn't evidence of a clear approach or solution for its usage, especially in the insurance context, in a **sustainable and practical way**. This situation reveals some degree of **knowledge gap** between what is wanted (using CLV) and what is the common understanding and practice in the industry. That **knowledge gap may be closed a little further** trough this work and **expand the knowledge** using a specific business case plus a engineering perspective and define a **project** with a more robust methodological and academic support.

1.2. Research Questions

The implementation of this project will address the following two research questions and related sub-questions:

1. **What is the understanding of CLV from the literature?**
   1.1. How can CLV be calculated?

1.2. What are the main critical factors that influence CLV?

1.3. How can the CLV be applied with insurance parameters and data?

1.4. What is the best CLV estimator for insurance?

1.5. How to measure CLV impact in the insurance business?

2. How can we integrate CLV into the Insurance Business Operations?

2.1. How to build an Automatic Estimation Machine?

2.2. How to expand the AEM effectiveness with SOA?

2.3. How to engineer a Business Process to support the CLV and AEM?

1.3. Research Objectives

There are two research objectives that can be satisfied from the research questions:

a) The first one is to understand the concept of Customer Lifetime Value (CLV) and how can it be, in the reasonable way, applied in the Insurance Industry grounded in quantitative studies that will be further detailed in the research design section.

b) The second is a practical and pragmatic one which is the definition of framework for a business process for the CLV application in a corporate environment in the context stated in the previous point. This framework must inform quantitatively the process owners about its continuous effectiveness.

1.4. Hypothesis

It is hypothesized that, for a certain business context and constraints, there is a practical approach to apply the CLV concept in the Insurance Industry, as a sustainable business process or framework, offering real competitive advantages for those who implement it.

1.5. Research Paradigms

The research paradigms used to achieve the research objectives and answer the research question are listed in the following table.

Table 1.1. *Research paradigms*

Paradigm	**Choice**
Philosophy	Pragmatism
Approach	Deductive
Strategy	Experiment
Choices	Multi Method
Time Horizon	Cross Sectional
Techniques and Procedures	**Applied Research via Project Development** => Research Methods + Formal Project Management => Literature Review, Hypothesis Formulation, Data Collection, Testing and Prototype Deployment

Source: Based on the research Onion hierarchy (Saunders, Lewies, Thornhill, 2007)

1.6. Overview of used Research Methods

The following table summarizes the research methods used in this research work. Further development will be given in the Methodology Chapter.

Table 1.2. *Research Methods used in the project*

Topic	**Research Method Used**
Relevance of the Topic	Qualitative questionnaires with key people
	Literature Review
Understanding the CLV concept, how can it be calculated and what dimensions influence it.	Literature Review

Expansion of CLV Concept into the Insurance Context	Literature Review
	Quantitative Data Collection
	Experimentation
	Focus Groups
Finding the Optimum Estimator	Quantitative Data Collection
	Experimentation
	Focus Groups
Business Integration	Experimentation + Prototyping
	Focus Groups

Source: The Author

1.7. Incorporation of Knowledge

From the knowledge management perspective during the development of the project there is incorporation of knowledge from several areas of science and technology. The relation between the several knowledge areas is shown in **Figure 1.1.**

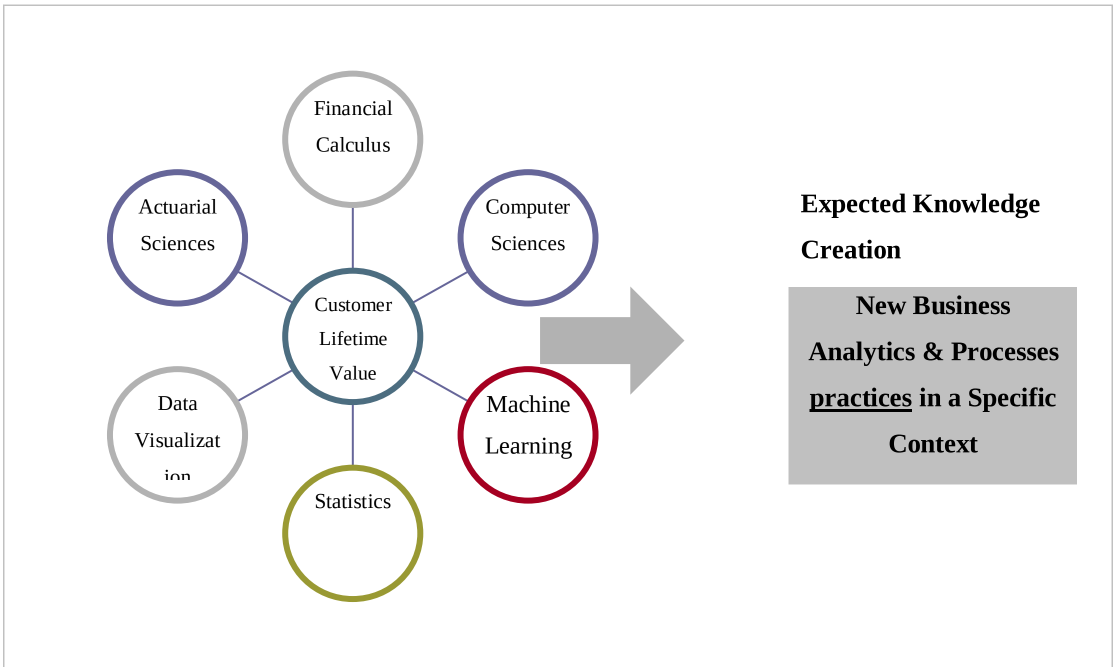


Figure 1.1. Knowledge Network

Source: The author

1.8. Field Work Location

The research work will be developed in a specific Portuguese Bancassurance Company integrated in a Portuguese financial Group.

1.9. The role of the researcher

The researcher is responsible for an Information Management Department on the Researched Company. Its role is supporting and developing BI and Analytical activities in the Company, crossing the gap between IT and the rest of the business. In this project, leaded by the CFO, the researcher:

- Investigated the concept of CLV;
- Investigated and applied datamining algorithms;
- Did some model sensitivity analysis;
- Did the programming and data management;
- Participated in the project meetings and discussions.

Other assumptions were discussed between the workgroup members and other business areas.

1.10. Assumptions

There is an important assumption that this research work does, which is the possibility of replication and applicability of a framework into a whole industry from the work developed in a very specific context. In the defence of this approach we must consider the following principles:

a) Being very regulated and supervised the Insurance Industry must obey a vast set of relatively uniform boundaries and constraints that extends beyond the local or national context. For example, all Portuguese insurance companies must be compliant with the future, European Union's Solvency II framework which enforces risk adjusted capital allocation under strict conditions and uniform market disclosure. This condition translates into uniform behaviours, product solutions and strategies across companies. Moreover, adding years of competition, fusions and acquisitions we may assume similar environments for the applicability of a CLV framework.

b) The theory of insurance is relatively uniform, not only at national level but also worldwide;

c) Whenever possible generalisation considerations will be made based on relevant scientific and technical literature;

1.11. Limitations and Delimitations

1.11.1. Delimitations

The research work has a clear delimitation since the CLV concept will only be studied for application on the Non-life line of business. The difference between the types of products, non-life vs. the life is such that the supervisory authorities require that the companies must specialized in only one of these two lines of business. If one Group want to explore life and non-life LOB's that is only permitted with legally and statutory independent companies, of course this separation extends into the assumptions about CLV calculation and usage.

1.11.2. Limitations

It must be clear that, although we have the assumption of a relatively uniform market it must be recognise that this argument and assumption is hypothetical and can only be tested by a wider research scope by including a larger number of firms in the qualitative and quantitative part of the experiments. That scope is in the present context difficult to achieve since the current researcher, as a worker in a specific Insurance Company, cannot be positioned as a neutral entity with equal research opportunities in different firms.

1.12. Conclusions about the basis of this work

This introductory chapter sets the stage for the research development, it must be stressed that although this work is presented as an academic work it is strongly rooted on a real project development. All research methods referred are subjected to the constraints, limitations and contingencies that affect all the stakeholders during their daily work routines. Pragmatism emerges from this necessity of combining practical knowledge with theory. Deduction comes from the necessity of applying general principles to specific cases in an experimental environment. Multi Methods are used to develop the research: meetings or focus groups are used to define strategies. Some correlational investigation (cross-sectional) is done around the known approaches developed by other investigators on this issue. The techniques used on the research are based on data collection, analysis, modelling and experimental design. This sum up the methods that are expected to make this project function and deliver what is expected: A framework for using CLV on Insurance Companies.

2. METHODOLOGY

2.1. Overview of Research Work Flow and Project Phases

As an overview the research project is divided in four phases governed by formal project management. Some, but not all, key topics are pointed out in Figure **2.1.**

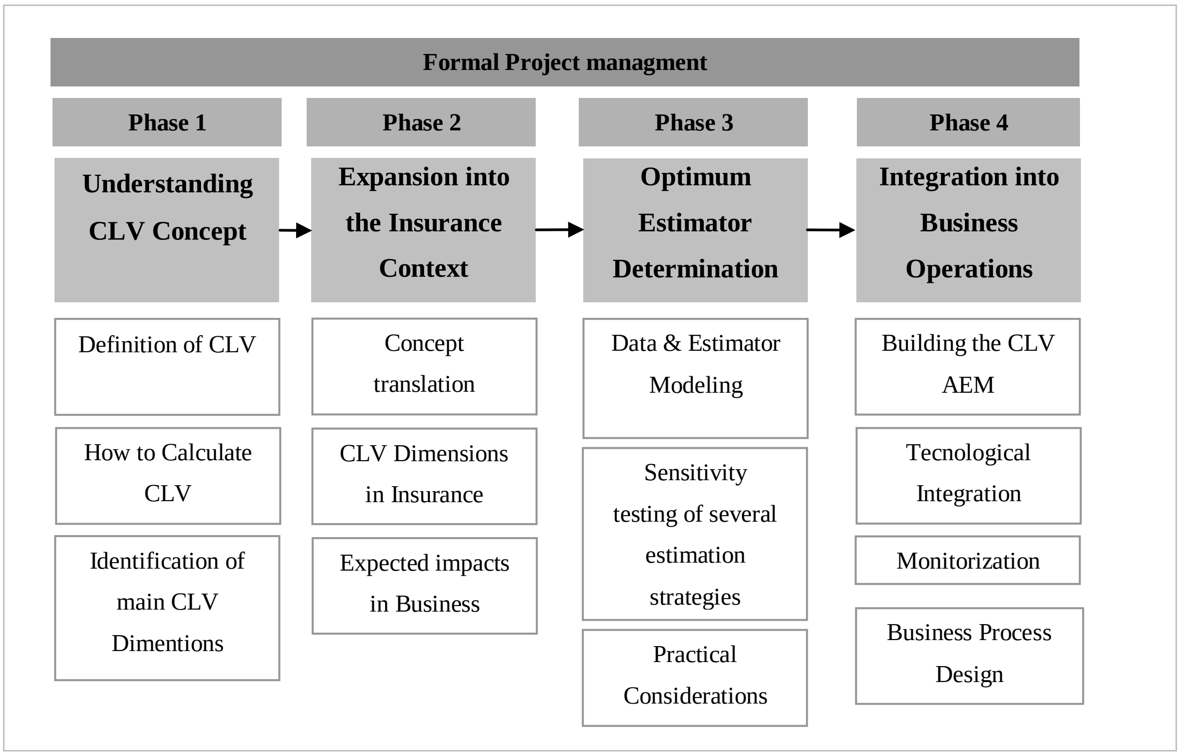


Figure 2.1. Research Work Flow

Source: The author

2.2. Phase 1 – Understanding CLV Concept

The first phase of the project is dedicated to build the theme theoretical foundation. This objective is achieved by **literature review** with emphasis in two areas:

Estimation Models and Application in commercial environments. For the estimation models there are key issues to check like what are the typical approaches used in CLV methodologies. In parallel some research about the expectations over the potential qualitative and quantitative impact on the business will be done. This is an important issue since the business will be the operator of the tools that will be created through this research project the deliverables must be accepted by people and integrated into the business operations. This research can be developed by using questionnaires issued to key managers.

2.3. Phase 2 - Expansion into the Insurance Context

The objective of this phase is translating the knowledge obtained during the first phase into the insurance context. This will be done mainly by experimentation and discussion using focus groups with technical personnel. During this phase two research problems must be addressed:

- Significant parameters must be identified;
- The most appropriate modelling approach must be selected.

These two problems define the **business frame and requirements** for CLV usage.

2.4. Phase 3 –Estimator Determination

With the business frame as the reference, the project enters in its wider phase where, through experimentation, the estimator (or collection of estimators) will be determined. To achieve that objective there are four main activities:

- Data collection & Data modelling;
- Application of algorithms for parameter estimation. It is implied the usage of a machine learning methodologies for testing several concurrent modelling techniques;
- Parameter Calibration;

- CLV model design and assessment.

2.5. Phase 4 - Integration into Business Operations

The final phase of the research work is prototyping a framework architecture which objective is supporting consistently a business process for CLV application. In figure **2.2** it is presented a simplified version of the framework.

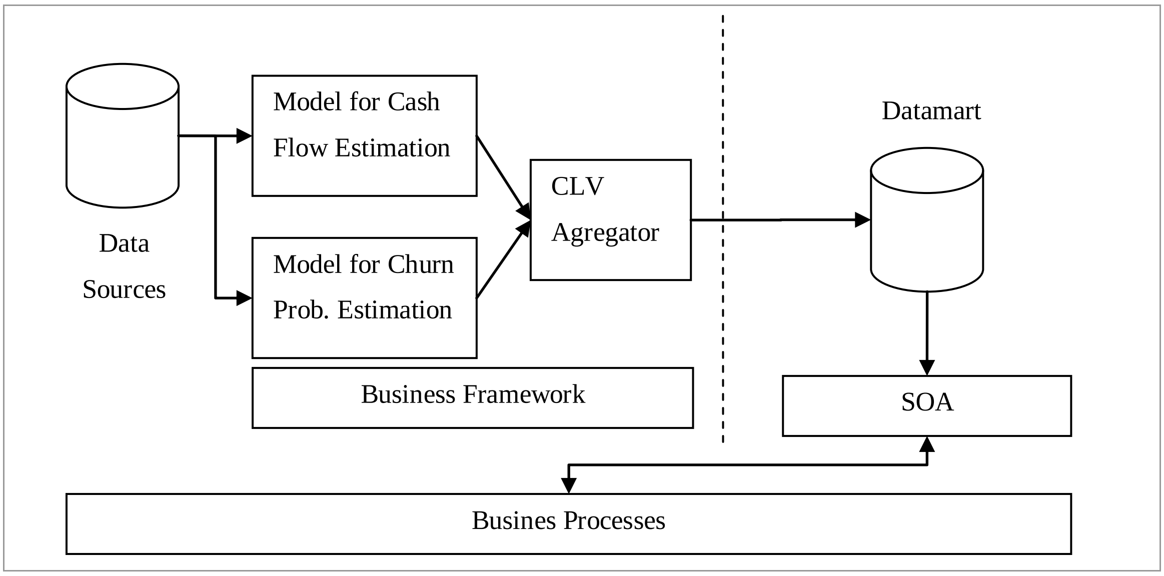


Figure 2.2. Automatic Estimation Machine Architecture - AEM

Source: The author

2.6. Conclusions on Methodology

Although the research methodology is grounded on literature review there is a great deal of experimental and engineering work evolved, especially on the third and fourth project phases. Formal project management is referred, but nothing is said about what kind of PM methodology will be used. Previous knowledge on the Organization indicates an Agile approach. Having this academic work, with the aforementioned phases, running along with the on situ project will bring together the best of both worlds.

3. UNDERSTANDING CUSTOMER LIFETIME VALUE

3.1. Introduction to Literature Review

Having the objective of identifying a framework to operate on a business, this research integrates conceptual and empirical work from several knowledge areas and backgrounds, namely: marketing, customer relationship management, actuarial sciences, financial calculus and machine learning.

To keep this review manageable it will be organized in sections **that reflect the research workflow and phases**. Starting by understanding the relevance of the research theme, Customer Lifetime Value, the research will then focus on the application of the concept on the insurance industry, in this phase insurance concepts will be researched for potential application in the CLV framework. In the third phase the focus will be directed onto the application of techniques and methods for estimating CLV. Lastly, the fourth phase will address the framework integration and interaction with other business processes.

To prepare the overview process the following section will illustrate the method of discovering relevant literature. Which is: with a bootstrap concept identify its elemental parts and then assemble an appropriate search equation for literature search in scientific databases like WoS. Some degree of triangulation with the insurance industry is also made to check the level of research maturity in that specific area.

3.2. Bootstrap concept

Taking into consideration the simplest mathematical definition of CLV, presented by Gupta, S. (2009) and several other authors, depicted in equation (1), we can identify the basic parameters that must understood to develop the research work.

$$CLV_i = \sum_{t=0}^{\infty} \frac{CF_t \times r(t)}{(1+d)^t} \quad (1)$$

Where for each customer **i**

- CLV is the Customer Lifetime Value
- **CF** is the net cash flow per period **(t)**
- **r** is the retention rate for period **(t)**
- **d** is the discount rate

From the equation (1) its clear that there are three fundamental concepts we must study to understand CLV which are by order of importance: **a)** retention rate over time (or churn); **b)** net cash flow over time; **c)** discount rate. Since discount rate is a collinear concept we're left with the first two parameters for further investigation. With that in mind the following concepts: **Insurance**; **Churn**, **CLV** will define the keywords and search equations as shown in the following sections.

3.3. Keywords

Customer profitability; Customer Lifetime Value; CLV; CLTV; Customer Equity; Churn; Customer Retention; Survival Analysis; Insurance; Bancassurance.

3.4. Search Equations

Table 3.1. *Search equations*

Search Set	Concept	Equation	# of Results
#7	(#4 or #5) and #3	(Churn Refined Or CLV Refined) and Insurance. We want the union of the intersection of the set "insurance" and "CLV" with "Churn" and "Insurance" so	10

		the expression must be "ored together". The reason is the need to investigate the phenomena of **Churn in insurance** independently of **CLV in insurance.**	
#6	#4 or #5	Churn Refined Or CLV Refined	501
#5	CLV Refined	TS=("lifetime*customer*value") or TS=(CLV) or TS=(LTV) or TS=(LCV) or TS=("customer equity") or TS=("Customer Life*time Value") or TS=("lifetime value") or TS=(CLV-models) or TS=("customer profitability") or TS=("CLTV") AND Language=(English) Refined by: Subject Areas=(BUSINESS OR MATHEMATICS, APPLIED OR MANAGEMENT OR COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE OR COMPUTER SCIENCE, INFORMATION SYSTEMS OR COMPUTER SCIENCE, THEORY & METHODS OR TELECOMMUNICATIONS OR COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS OR BUSINESS, FINANCE) AND Subject Areas=(BUSINESS OR MANAGEMENT OR COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE)	322
#4	Churn Refined	TS=(churn*) and (TS=(model*) or TS = ("survival analysis") or TS=(analysis)) AND Language=(English) Refined by: Subject Areas=(BUSINESS, FINANCE OR COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE OR OPERATIONS RESEARCH & MANAGEMENT SCIENCE OR MANAGEMENT OR COMPUTER SCIENCE, SOFTWARE ENGINEERING OR BUSINESS OR ECONOMICS OR COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS OR COMPUTER SCIENCE, CYBERNETICS OR MATHEMATICS, INTERDISCIPLINARY APPLICATIONS)	190
#3	Insurance	TS=(insurance) or TS=(Bancassurance)	48816
#2	Churn	TS=(churn*) and (TS=(model*) or TS = ("survival analysis") or TS=(analysis))	712
#1	CLV	TS=("lifetime*customer*value") or TS=(CLV) or TS=(LTV) or TS=(LCV) or TS=("customer equity") or TS=("Customer Life*time Value") or TS=("lifetime value") or TS=(CLV-models) or TS=("customer profitability") or TS=("CLTV")	2021

These search equations will define the core research documentation set. Further documentation *outside this boolean equation* will be added as needed.

3.5. Phase 1 – Understanding CLV

The objective of the following sections is the review of the central concepts behind CLV. It is not an in depth review of every method or approach that have been proposed in this area, the intention is to build a reasonable understanding of CLV adequate enough to support the project phases planned for this research work.

3.5.1. Brief definition of CLV

From common sense it's apparent that each individual Customer has an impact on the overall firm financial results as described on Glady, N., Baesens, B., & Croux, C. (2009b) work on modelling churn using customer lifetime value. Therefore the firm results are the emerging reality of all Customer contributions, each one different from each other and implying different economic valuations. Decision makers and researchers recognising the opportunity and importance of that idea for managerial purposes have shaped formulations to capture each contribution and operate over the customer base during its lifetime. Thus the denomination of this idea as CLV - Customer Lifetime Value, the **basic CLV model is defined as the total discounted lifetime cash flows generated by the customer**, some authors have summarized this concept giving us the following definition:

> CLV represents the **present value** of the **expected benefits** less the **costs of initializing**, **maintaining and developing the customer relationship** (Malthouse and Blattberg, 2005). Different academics explore the CLV concept (Dwyer, 1997; McDougall et al., 1997; Berger and Nasr, 1998; Jain and Singh, 2002; Mason 2003; Stahl et al., 2003), which can be summarized as **the sum of accumulated cash flows of a customer over their entire lifetime** with the organization. (De Oliveira Lima, E. (2009) Domain knowledge integration in data mining for churn and customer lifetime value modelling: new approaches and applications - Page 8)

Although in almost every serious business the past commercial records are useful and frequently are used to build traditional customer performance metrics, that information has a relative importance since it is used essentially to build what is called the Past Customer Value. This view can be distorted, mainly because it assumes the same Customer behaviour for the future, it doesn't incorporate other useful information about the long term potential. For example, an insurance customer had in the first relation year a loss, most likely this Customer is in the *red*, does that mean this is a bad Customer? It depends on what kind of products he has: he may have long lasting and profitable products making him a better or even a good customer in the long term. The contrary is also valid and this dichotomy is modelled by the CLV approach.

Contrary to the past Customer Value, CLV concept is essentially prospective, it looks into the future therefore it is **a more powerful concept and simultaneously a more useful tool for decision making since it works as an influencer over** the short term managerial decisions taking into account the long term effect of each Customer contribution. In the past Customer history lays the information needed to learn, estimate and adjust the parameters to calculate the future or potential value of Customers.

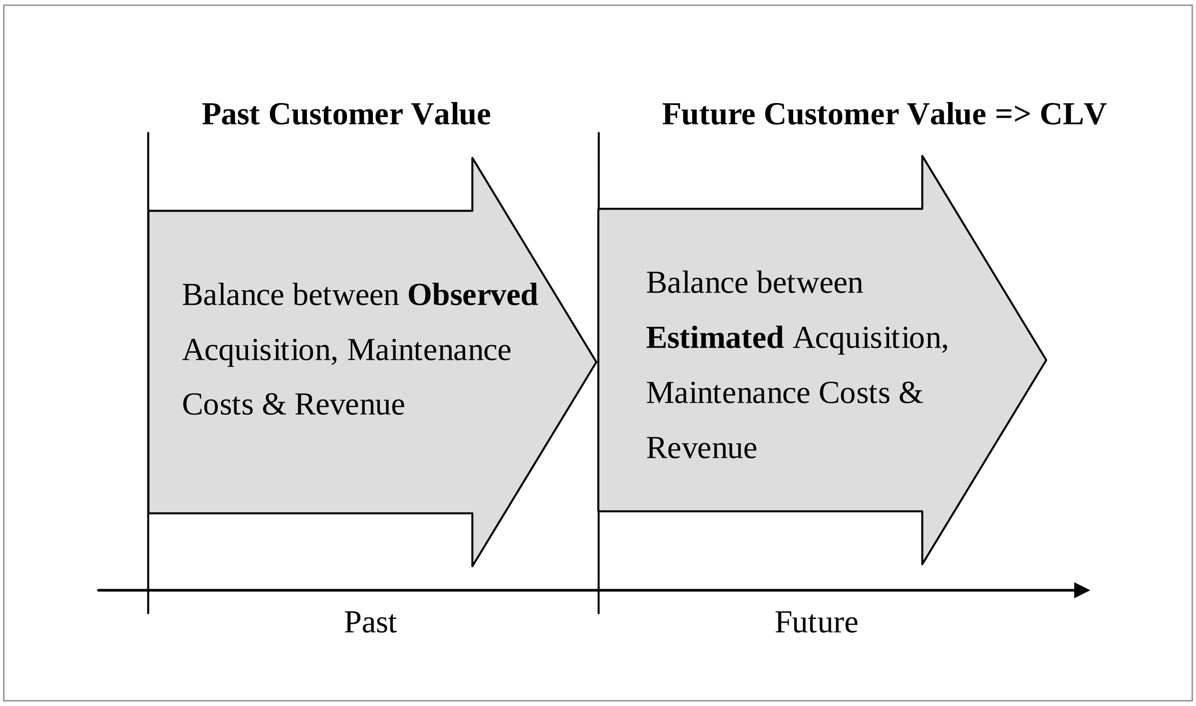


Figure 3.1. The Past and Future Customer Value

Source: The Author

For the future Customer value, since we may be considering large time spans the problem of estimating the probability of occurrence of those cash-flows is central. This implies not only the estimate of how **future costs and revenues will vary** but also the **estimation of the customer retention** probabilities as noted by Gupta, S. (2009). Another concept emerged from the research around CLV as described by Wiesel, T., Skiera, B. & Villanueva, J. (2008) on their work on Customer Equity. This concept connects the Customer Lifetime Value to the firm financial results which is a more extensive concept than CLV.

3.5.2. Customer Equity

Customer Equity (CE) is the approach where the connection between CLV and the firm value is taken into account. Customer's relations are seen as assets and customers are the drivers of revenues. In a more holistic view CLV can be integrated in a set of strategies and processes geared to generate economic value to the firm's stakeholders. Mathematically CE can be considered the **sum of the individual CLV for all firm customers** as shown in the next formula.

$$CE = \sum_{i=1}^{N} \sum_{t=0}^{\infty} \frac{CF_t \times r(t)}{(1+d)^t} \tag{2}$$

A more particular formula includes the discrimination of different lines of business, considering that each customer may have more than one of the firm's products.

$$CE = \sum_{i=1}^{N} \sum_{l=1}^{L} \sum_{t=0}^{\infty} \frac{CF_{lt} \times r_l(t)}{(1+d)^t} \tag{3}$$

Where

- **CE** is the Customer equity
- **CF** is the net cash flow per period **(t)** and line of business (**l)**
- **r** is the retention rate for period **(t)**
- **d** is the discount rate
- **N** is the total number of Customers
- **L** is the total number of Products

A firm with greater Customer Equity has a more profitable customer base in terms of whole lifetime, thus it is a characteristic of a **more valuable Organisation.**

3.5.3. The importance of CLV and Customer Equity CE

The importance of CLV is intrinsically connected with the commitment with the management of Customer Equity or in other word the Customer base. Given the importance to create value to the firm stakeholders all the processes related with the customer relation and its impact on value must be assessed. Managing Customer Equity, since not all customers are equal, entails the concept of Customer prioritization. As noted by Rust, R. T., Lemon, K. N. & Zeithaml, V. A. (2004) and Abele, Karin P. N. (2009), firms sponsoring CE management strategies develop some or all of the following activities:

a) Choosing the customers with greater CLV and nurturing them while "deselecting" the lower CLV customers as a strategy to increase profitability for the firm by:
    - having a finer and profitable pricing policy;
    - increasing investment only in key areas;

b) Since not all customers are equally profitable, CLV is a good metric to reallocate scarce resources within the organization by:
    - Create new products or services intended for special and more profitable segments;
    - Close services or abandon products;

c) As a side effect of CLV programmes is the ability for the organization to continually assess the customer development strategies;
d) Customer satisfaction is potentiated, thus having them as good market referees for the firm. This could have a good impact in acquisition costs;

These activities form the core of a master process that manages the Customer relations trough they're lifecycle, this is called Customer Lifecycle Management or CLM.

3.5.4. Relation between CLV/CE and Customer Lifecycle

Customer Lifecycle Management is a set of Processes with the objective of managing key moments in the relation between a firm and a Customer. Managers and Marketers usually make reference to the following events: **a)** Acquisition; **b)** Service; **c)** Expansion; **d)** Termination. From the interoperability between the CLV as a tool and the Customer Lifecycle Process emerges the opportunity to leverage the Customer Equity. In **Figure 3.2** it is depicted a version of this approach where 10 CLM events are referred.

Figure 3.2. Customer Lifecycle

Source: McKinsey & Company

At each CLM event the firm seeks to **maximize revenue** by targeting the more valuable and likely Customers to respond to tailored marketing activities. That can be done by:

1. Reducing costs on creating interest
2. Reducing costs for Acquisition
3. Increasing recurring revenue
4. Reducing the direct costs to serve
5. Increasing cross and Up-sell
6. Increasing renewal
7. Reducing cost of migration
8. Reducing churn
9. Limiting risk of bad debt
10. Reducing costs to win back

These concepts can be seen as a value chain, shown in **Figure 3.3.,** of processes which the ultimate aim is to maximize the value creation for its stakeholders: Customers, Employees and Shareholders.

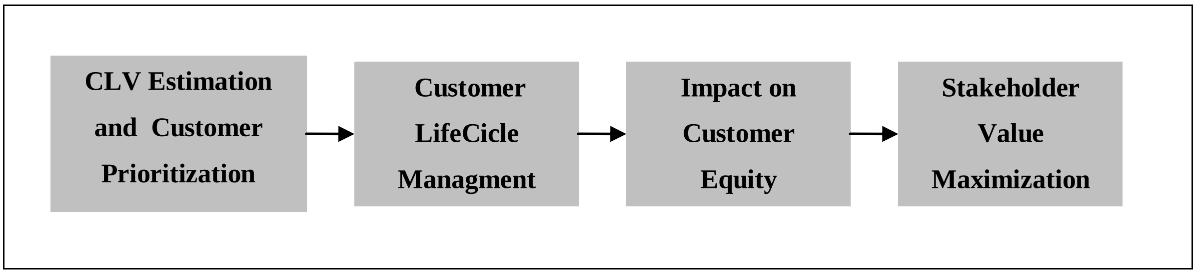


Figure 3.3. CLV Value Chain

Source: The Author

Having described the macro context of CLV and CE in the business arena, the following sections will focus on the different methodologies and approaches proposed by several authors for calculating CLV. The understanding of these approaches will be important for a more adequate methodology to develop Phase 2 and 3 of this research work.

3.5.5. Different Approaches to CLV/CE Modelling

Some researchers like Gupta, S., Hanssens, D., Kahn, W., Kumar, V., & Lin, N. (2006, Nov.), have done systematizations of methods for Customer Valuation and segmentation. Often more traditional methods are compared with CLV methodology therefore we have the following indicative taxonomy:

Traditional Methods

- RFM - Recency Frequency Monetary
- Past Customer Value

Newer Approaches

- CLV – Customer Live Time Value
- Aggregated Approaches
- Individual Approaches
- Mixed Approaches
- Probability Models

In this review only a very small number of methodologies will be described in more detail, for a more in depth study of these methodologies it is advisable to consult the authors referred in the beginning of this section.

3.5.5.1.The RFM model

This model is based in quintile binning of three variables **R** – Recency, how long the customer is registered, **F** – Frequency, how many times the customer made transaction in a period of time and **M** – Monetary, how much did he spent in that transactions. Typically this leads to a **125 Cluster Matrix**. The clusters with a better Frequency, Monetary and Recency characteristic have the most valuable Customers.

Although very simple and cost effective, RFM models tend to be more limited since the method only incorporates, by definition, three variables binned to a set of fixed clusters. The method only integrates the past performance and assumes coarsely equal performance for the future. There isn't also an individual economic or equity value attributed to each Customer, so there is some kind of "qualitative" feature in this kind of models.

3.5.5.2.Past Customer Value

In this method, PCV, it is assumed that the past Customer behaviour is an indicator of future Customer behaviour. The PCV is calculated considering the total sum of revenue of past transactions adjusted to present value. The PCV can be calculated with the following formula.

$$PCV_i = \sum_{t=0}^{T} CF_t \times (1+d)^t \qquad (4)$$

Where

- **PCV** is the Customer Past Value
- **CF** is the net cash flow per period **(t)**
- **d** is the discount rate for actualization
- **T** is the total number of time periods before the present

Like RFM this method has several limitations. It assumes similar economic performance and value in future transactions and the probability specific future events like desertion, cross-sell or up sell is not taken into consideration.

3.5.5.3.Advantages of CLV methods

Unlike traditional methods, CLV approaches take in account the probability of the customer being active in the future therefore having a potential to generate revenue. This analysis can be applied to the customer a whole entity or just to part of its product portfolio. The Customers with a higher probability of being active for a longer time, with positive revenue for the firm, will have a bigger present value. Given that the firm wants to maximise future profitability, the capability to prioritize the right customers is fundamental. Another aspect taken in account is the incorporation **of future acquisition and maintenance costs** in a view comparable **to an investment**. In this view, after an initial negative CLV associated with acquisition costs there will be a break even point where revenues will payoff the Customer variable and fixed costs, see **Figure 3.5.**

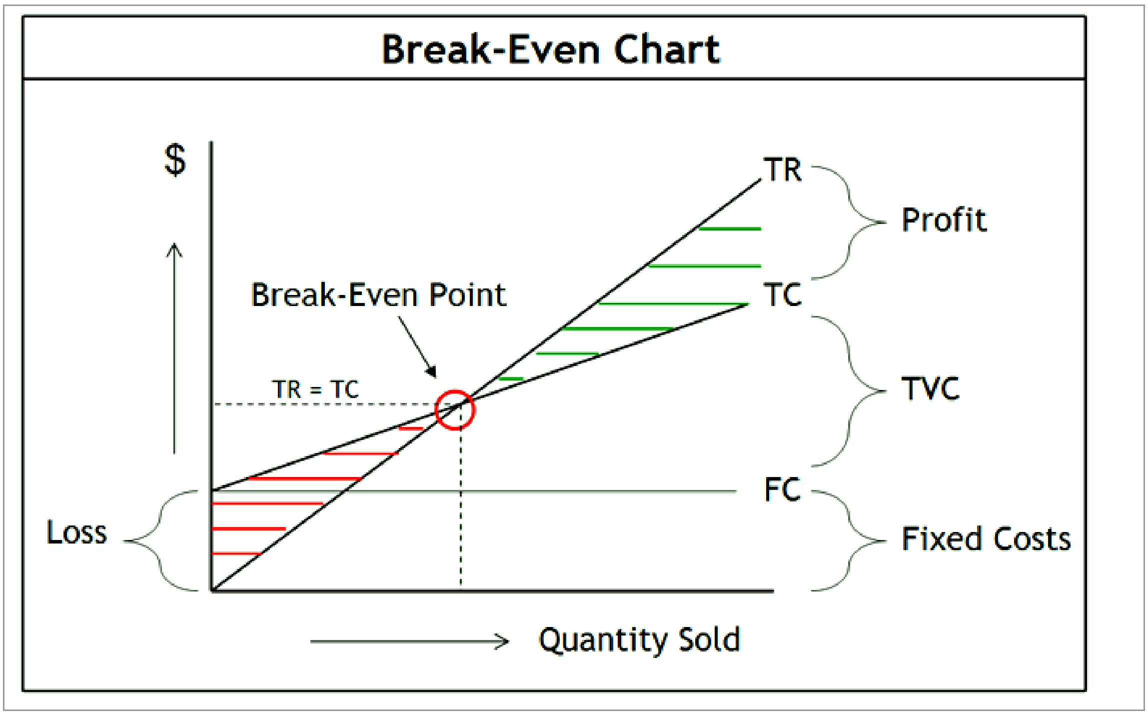


Figure 3.4. The Customer as an Investment.

Source: http://www.mrdashboard.com/Break-even_Analysis.html

- TR: Total Revenue which is calculated as number of products sold times unit price;

- TFC: Total Fixed Costs;

- TVC: Total Variable Cost which varies directly with transactions
- TR = Unit Price x Transactions
- TC = TFC + TVC
- P: Profit is calculated as total revenue minus total cost P = TR –TC

3.5.5.4.Aggregate CLV Approaches

In aggregate approaches the calculation of CLV e done with at the level of a segment or even considering the firm as a whole. The following formula states one aggregate approach to calculate the CLV for a customer in segments.

$$CLV_s = \left[\sum_{t=0}^{T} \frac{(GC - M)}{(1+d)^t} r^t \right] - A \qquad (5)$$

Where

- **GC** is the average gross contribution margin per customer
- **M** is the average marketing cost per customer
- **r** is the retention rate
- **T** is the number of periods of CLV estimation
- **t** is the period
- **A** is the average acquisition cost per customer

In this approach, **GC**, **M**, **A** and **r** are **aggregate quantities specific to the segment s** and averaged to the **number customers in that segment**, no further individual information is used, it is assumed that all customer in that segment will behave in a similar way. Some researchers proposed more complex aggregate approaches that take in account more factors embodied in concepts as return on acquisition, return on retention, return on up/cross-selling. In this case the probabilities for the referred events in each segment must be considered. Although being a finer model it suffers from the limitations of all the aggregate approaches, that is, individual Customer variations are not considered. Today in data driven organizations more customer data and information may be available that can be used to estimate with greater precision parameters like Customer or Product retention rate. These limitations lead us to another class of CLV approaches.

3.5.5.5.Individual CLV Approaches

Similar to aggregate approaches Individual approaches are based on the same mathematics but the principles of parameters estimation are different. The objective of individual approaches is to incorporate the maximum information specific to a customer in the calculations. Some proposed approaches state the individual CLV by the following expression

$$CLV_i = \sum_{n=t+1}^{t+x} P(Active)_{in} \times \frac{AGC_{it}}{(1+d)^n} - \sum_{n=1}^{x} M_{in} \times \left(\frac{1}{1+d}\right)^{n-1} - A_i \quad (6)$$

Where

- **i** is the Customer index
- **t** is the period for AGC is being calculated
- **n** is the number of periods beyond **t**
- **x** is the number of periods for CLV estimation
- **P(Active)** is the probability of the Customer **i** being active in period **t**
- **AGC** is the average gross contribution margin for Customer **i** in period **t**
- **d** is the discount rate for actualization

- **M** is the marketing cost per Customer **i**
- **A** is the average acquisition cost per Customer **i**

Contrary to the aggregate approaches the probability of a specific Customer being active P(Active) must be calculated individually rather than at segment or firm level. A variety of methods can be used to estimate this parameter. One proposed method is observing the past Customer behaviour and estimate future interactions, the following formula does that

$$P(Active) = (T / N)^n \quad (7)$$

Where

- **T** is the number of periods between Customer acquisition and the last cash flow record
- **N** is the number of periods between Customer Acquisition and the period for which P(Active) is to be calculated
- **n** is the number of Cash Flow records

This simple formula relates the total time frame since acquisition with the time until last interaction. If a customer was acquired 10 years ago and only did one cash flow event (purchase) in the first year the probability of being active in the end of the tenth year is only 10%. The formula is also very sensitive to the variation of the level of the activity. If the same customer did two cash-flow events in the same period the probability of being active in the same period is reduced tenfold. This specific formula for calculating P(Active) suggests an empiric calibration and adaptation approach using data from the firm and many variations of this formula can be obtained.

3.5.5.6.Mixed Strategies for CLV calculation

Since Aggregate or Individual Approaches have their own set of limitations and strong points, some authors: Fader, P. S., Hardie, B. G. S., & Jerath, K. (2007), Kumar, V., & George, M. (2007) and Wiesel, T., Skiera, B., & Villanueva, J. (2008) have proposed mixed strategies to maximize the applicability of a CLV framework in the real world. Depending on the specific business environment, each firm will have an optimal mix of approaches that may or not be used. **Figure 3.6** shows one approach to study and compare different approaches with dissimilar aggregation and assumptions.

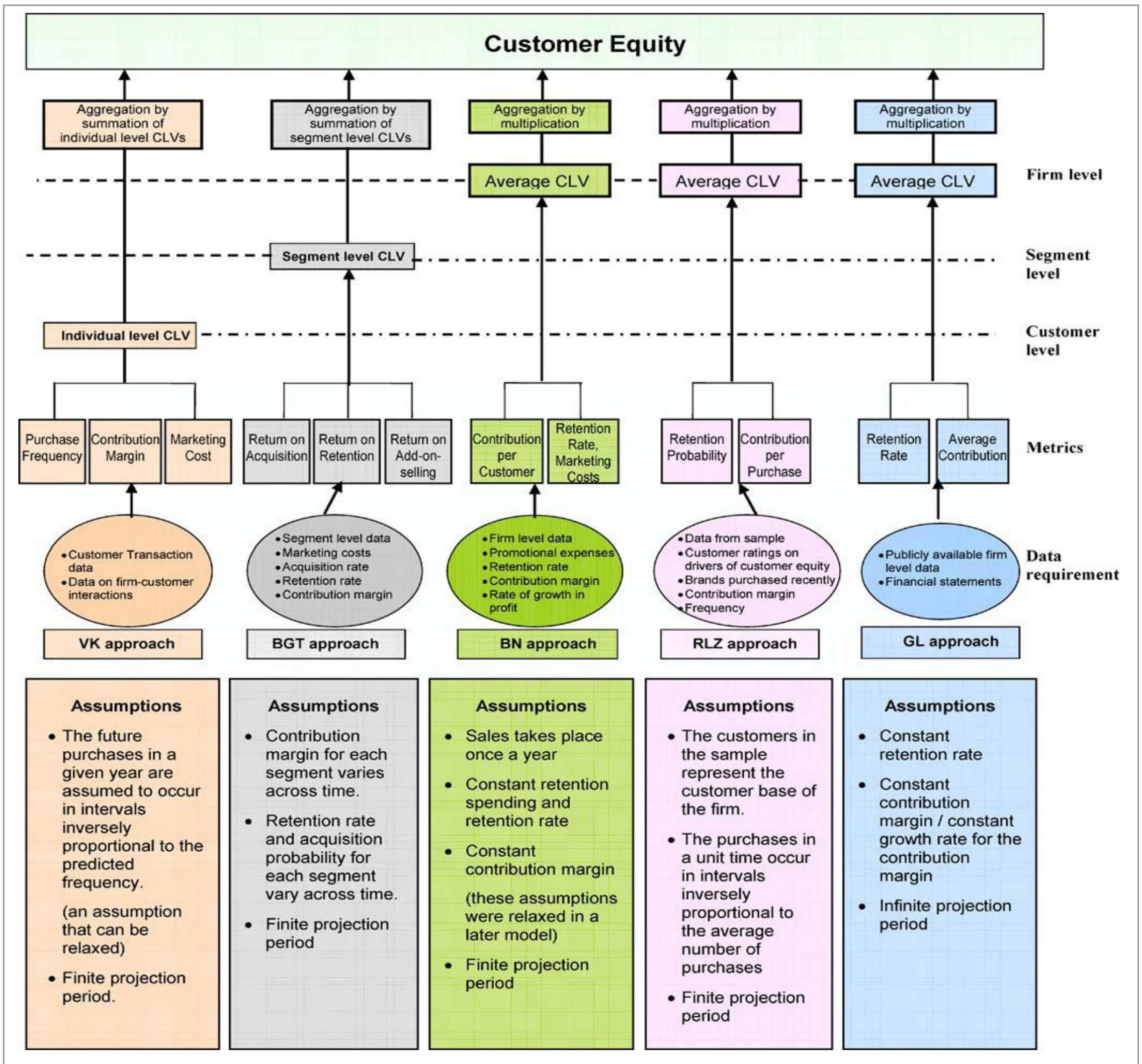


Figure 3.5. Mixed Strategies – Different concurrent approaches

Source: V. Kumar & Morris George

The availability and the level of granularity of data is key, for instance, if Customer data is not easily workable the individual CLV approaches are more difficult or impossible to use. In other scenarios, although customer level data is available, the intended CLV framework would only make sense if aggregate information should be integrated in the model. That happens with the integration of market data or firm level strategies. Other reason to use mixed approaches is to control the model complexity trying to achieve a goldilocks condition between cost, power and operation ability**.**

The following table summarizes, at the firm operations level, the differences and contexts between the different CLV calculation approaches.

Table 3.2. *Mixed Approaches*

	Individual	**Aggregate**
Method	CLV for every customer is calculated from customers' buying history and firm customer exchange characteristics. CLVs are then aggregated by summation to get the customer equity	Customer equity is computed from firm level measures like acquisition rate, average contribution margin or from average CLV of a sample of customers.
Challenging issues	• Longitudinal data on a large number of variables needed. • Does not incorporate competition in the CLV model • Differential treatment to customers on the basis of their CLV may lead to consumer backlash.	• Constant average margin and retention rate may not reflect the real life scenario. • The calculation of CLV over an infinite time period is over simplification • The components of customer equity (as per this approach) cannot be segregated completely • The use of average acquisition rates and retention rates for each segment as a surrogate for future retention probabilities ignores the impact of marketing actions. • Sample selection can play an important role in the accuracy of the metric. • Subjectivity of customer rating can affect the accuracy of the metric.
When and where to use?	When longitudinal database is available with the firm and the objectives are: • Formulation of customer level strategies • Formulation of firm level strategies • Financial valuation of a	When the firm does not have longitudinal database and the objectives are: • Formulation of firm level strategies • Financial valuation of a firm • Comparison of two firms in terms of their customer equity share or value to their shareholders

	firm.	
Industries	In industries where end-user transaction data is available and marketing investments can be customized	In industries where it is difficult to collect customer transaction data or to customize marketing investments

The authors of this study proposed a hybrid framework where the different approaches are used according to the limitations and strategies present in the specific business environment.

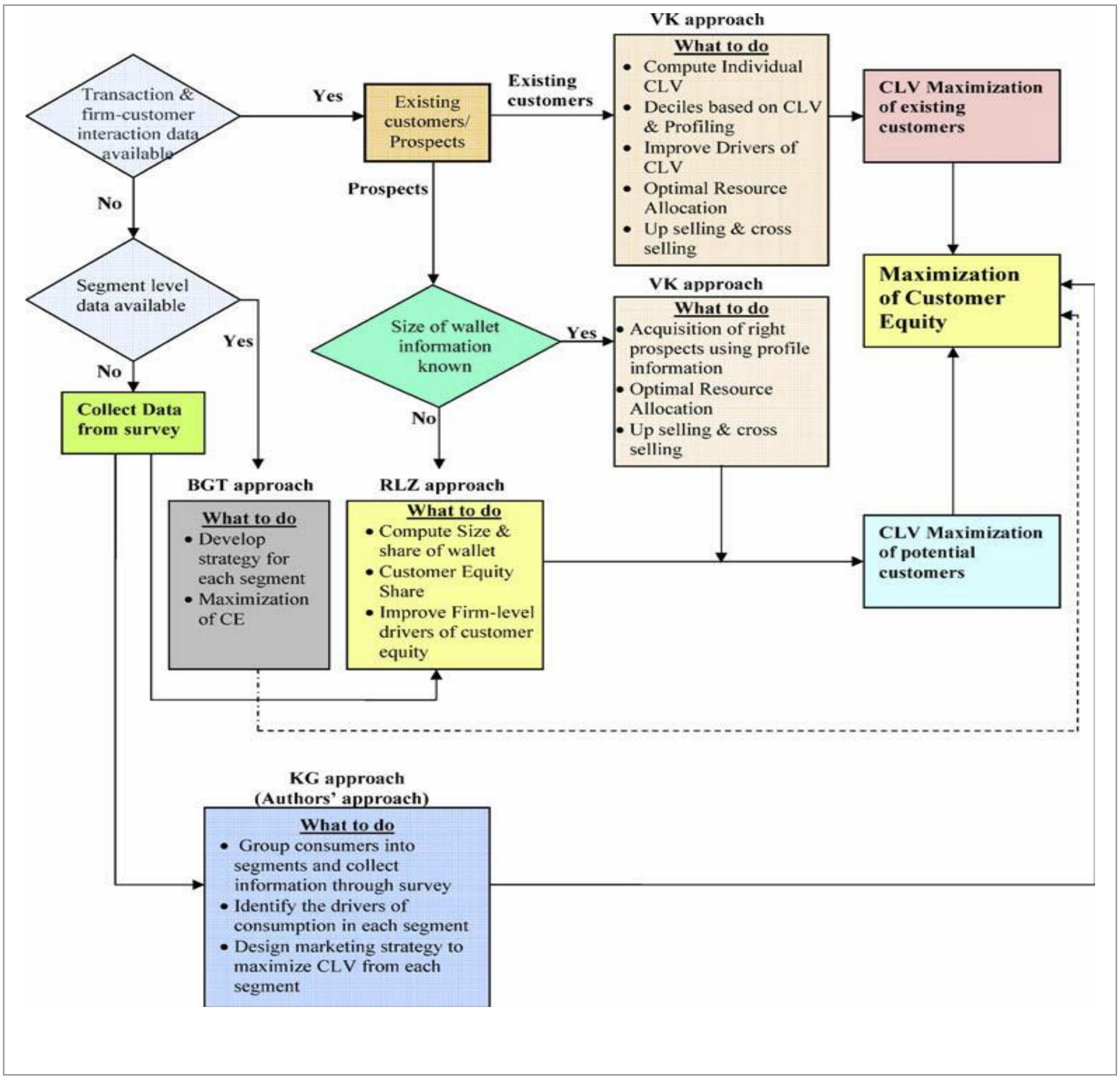


Figure 3.6. Mixed strategies for CLV/CE estimation – Proposed Hybrid Solution

Source: V. Kumar & Morris George

3.5.6. CLV Parameters

Taking into account the findings pointed in the Understanding CLV section, where the main elements where identified: Churn Ratio, Discount Rate and Cash-flow value and its variations. The challenge is identifying the optimum strategies to estimate these parameters.

3.5.6.1.Discount Rate

This parameter is one of the main factors for CLV calculation, as observed by Gupta, S. (2009), it projects future generated margins into their present value. This depends on the time preference over some kind of asset like money for instance. If someone or an organization does have a high time preference it means that he or she privileges having access to the asset in the short term, putting a higher cost or premium on having access to that asset latter in the future. In the case of a creditor or investor, the cost or premium, for lending an amount of capital and receiving that same amount of capital deferred in time defines a discount rate that compounds several features:

- Time value of money: External factors will change the future value of the present capital such as inflation;
- The Right to use that capital: The lender gives the privilege to the borrower to use the resources with the confidence and expectation of sharing partially the value/enjoyment created with that resources;
- Opportunity Cost: A measurement on the value of the next best alternative;
- Default Cost: The uncertainty of not receiving back the capital;

In the business context the discount rate depends greatly on the kind of operation being led by the firm, the source of its capital and the shareholders expectations. Thus the Cost of Capital of a specific firm will be a strong estimator for the discount rate.

Another reason to base the discount rate, for CLV purposes, on the cost of capital is when the Customer Equity approach is relevant. In this case the firm's value, Shareholder value and Customer Value are tightly connected.

For the Cost of Capital estimation several methods are proposed by economists and researchers as noted by De Oliveira Lima, E. (2009). Two methods are detailed in the following sections:

3.5.6.1.1 Weighted Average Cost of Capital

Broadly speaking, a company's investments are financed by either debt or equity. WACC is the average of the costs of these sources of financing, each of which is weighted by its respective use in the given situation. By taking a weighted average, we can see how much interest the company has to pay for its financing.

$$WACC = \frac{E}{V} \times Er + \frac{D}{V} \times Dr \times (1 - cT) \tag{8}$$

Where:

- E is the market value of the firm's equity;
- D is the market value of the firm's debt;
- V is the total financed capital (E + D);
- Er is market cost of equity;
- Dr is market cost of debt;
- cT is the corporate Tax;

A firm's WACC is the overall required return on the firm as a whole and it is often used internally by the company to determine the required or expected discount rate for cash flows. Usually WACC is used to determine the firm's Net Present Value.

3.5.6.1.2 Capital Asset Pricing Model

The Capital Asset Pricing Model or CAPM is based on the idea that an investor must be compensated by the sum of time value of money and risk. The time value of money is represented by the risk free rate and the risk part is supported by supplementary compensation, or Risk Premium, needed to the investor to accept that specific risk compared to the market risk. Thus the return on investment R_i is given by the sum of the risk free rate and the risk premium.

R_i = Risk Free Rate + Risk Premium

The risk-free rate usually is based on a risk free asset, for example 10-year government bond.

The risk premium depends on the systematic risk, risk which cannot be eliminated through diversification. Interest rates, recessions and wars are examples of systematic risks. There is also unsystematic Risk or "specific risk," this risk is specific to individual investment assets and can be diversified away as the investor adds different investments to the portfolio, it represents the component of an investment return that is not correlated with general market moves. Therefore a more formal model of CAPM is given by the following formula:

$$R_i = R_f + \beta \left| E(R_m) - R_f \right| \quad (9)$$

Where

- R_i is the required return on investment;
- R_f is the risk free rate;
- $E(R_m)-R_f$ is the expected market risk premium;
- $E(R_m)$ is the expected return on market portfolio;
- β is the total market variance explained by the investment;

When $\beta > 1$ the investment is riskier, has more fluctuations (meaning more specific risk) than the market, when $\beta < 1$ the investment is less risky and has less fluctuations than the market. When an investment is riskier the investor requires a higher return rate.

3.5.6.1.3 Discount Rate Analysis

Although there are several formal ways to estimate the discount rate for CLV programs there are no unique answers, it will depend on each specific organization's reality. What is the return required by investors, if applicable, what is the riskiness nature of the firm's operations. In each moment can we consider the assumptions about the risk environment as being true, does a risk free asset is really risk free. Ideally the annalists concerned with this issue should do sensitivity tests using different approaches, also the kind of products and their average lifecycle are relevant in the discount rate analysis. Long term products will have more sensitivity than short term products.

As an example the following figures show three scenarios where, for a fixed time frame of 50 periods, the effect of different discount rates is exposed, CF1: dr =0.05; CF2: dr = 0.075; CF3: dr = 0.1 for a unit of monetary value.

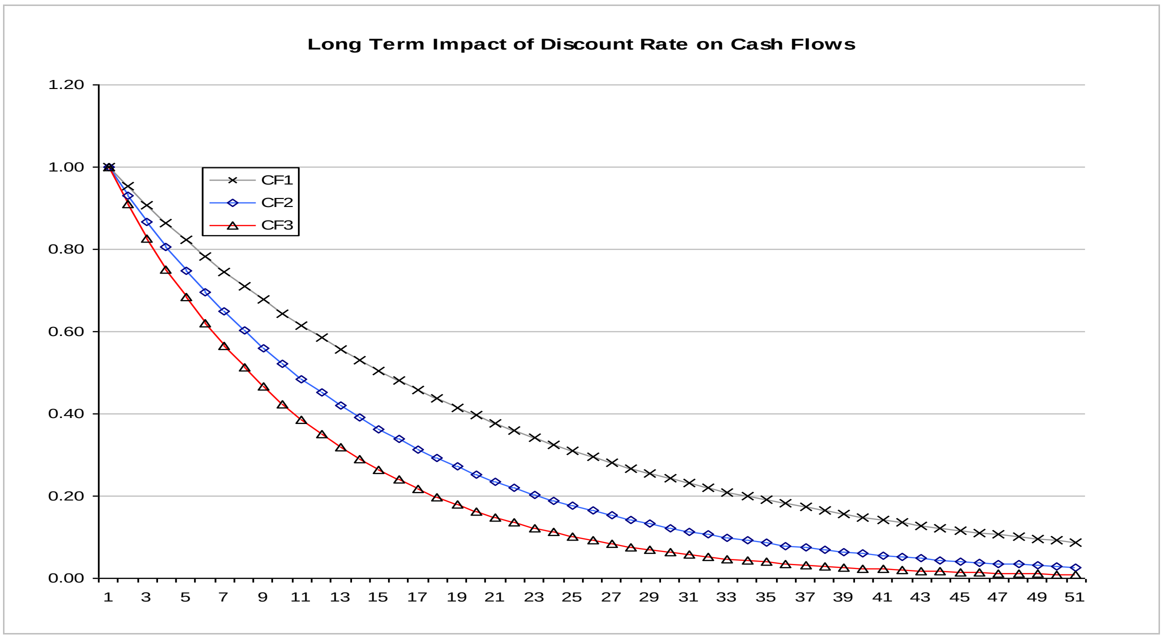


Figure 3.7. Discount Rate Long term Effect on Cash Flows

Source: The Author

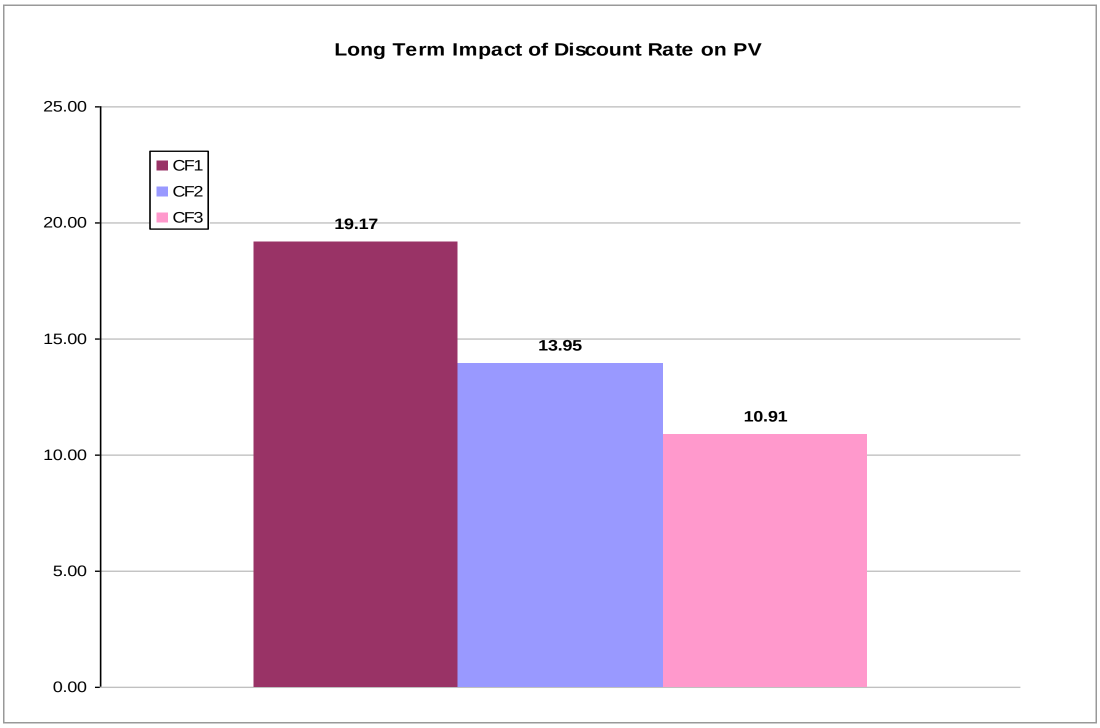


Figure 3.8. Long term Effect of Discount Rate on Present Value - Accrued

Source: The Author

In this specific case the doubling of the discount rate has an effect of reducing the present value in 43% during 50 periods. See annexe **A2** for further detail on this data.

3.5.6.2.Churn Ratio

Churn Ratio measures the rate customers end relations with a business therefore having a major impact in the generation of future cash flows. With this in mind it's clear that the estimation o this parameter is central in the context of CLV as noted by Gupta, S. (2009), Abele, Karin P. N. (2009), De Oliveira Lima, E. (2009) and many other authors. The approaches to estimate this parameter are further developed in the role of datamining section.

3.5.6.3.Cash Flow Variations

The cash flow estimation is not only determined by the churn behaviour, changes in customer product portfolio or expected variations in margin by internal or external reasons must be considered as was observed by Abele, Karin P. N. (2009*)* and Donkers, B., Verhoef, P. C., & de Jong, M. G. (2007). Some authors suggest that the CLV models must estimate those variations.

> As a result, the simple **profit regression model**, **which can deal with growth rates in profits, outperforms these models.** The model that has a good performance on all criteria is the Tobit II model, which combines a regression model for profit growth with a Probit model for customer retention, thereby capturing both dimensions of customer behavior. (Donkers, B., Verhoef, P. C., & de Jong, M. G. (2007) - Modelling CLV: A test of competing models in the insurance industry.Page 182)

3.5.7. The role of Datamining

As explained by Berry, Michael J. A. & Gordon Linoff (1997), datamining is the process of discovering by automatic or semiautomatic means significant and complex patterns in large quantities of data. The process can be described by the following steps: **a)** identifying sources of data; **b)** prepare data for analysis; **c)** build and train a computer model; **d)** apply the computer model to new data; **e)** test the validity of results. The problems solved by this datamining process fall in two major categories: The first is classification where the objective is deciding the degree of membership of an observations to a specific class (e.g. is red or is blue) and secondly estimation where the objective is to estimate a certain quantity (e.g. the temperature of a system). There is also other dimension in categorizing datamining methodologies: Supervised methods, where a target or dependent attribute is defined in advance and the algorithm tries to discover relations among independent attributes using *positive* and *negative* examples of the target variable and secondly unsupervised methods where the algorithm discovers the patterns without guided examples using all attributes present in data to convey to the analyst the underlying structure. The following table summarise some algorithms commonly used in datamining processes.

Table 3.3. Some *Datamining algorithms*

Algorithms	Problem Type	Methodology	Minus/Pluses
Neural Networks	Classification and estimation; Supervised	Simulation of the adaptive learning process by a network of artificial neurons.	**Pluses** • Applicable to many problems • Good results in complex problems • Widely used **Minuses** • Hard to explain results • Complex Data Preparation • Local Minima
Decision Trees	Classification and estimation; Supervised	Growing a readable tree like graph by choosing the most significant attributes and classes to explain the dependent attribute.	**Pluses** • Explainable results • Differentiate the importance of attributes • Able to a wide range of data types **Minuses** • Not appropriate for estimation problems • Only orthogonal solution spaces
Genetic Algorithms	Classification and estimation; Supervised	Simulation of the biologic evolutionary process (Selection, Crossover, mutation) to approximate a function that describes the dependent attribute.	**Pluses** • Explainable results • Easy to apply results • Able to a wide range of data types **Minuses** • Computationally expensive • No guaranty of optimally
Automatic Clustering Detection	Classification; Unsupervised	Grouping observations by measuring their relative distance to reference points and from each other.	**Pluses** • Easy to apply • Many Data types supported **Minuses** • Hard to explain results • Complex Data Preparation • Sensitivity to Initial parameters

All these algorithms use optimally criteria for determining the best possible solution. The supervised methods use, in parallel, a triplet of data sets for that matter: train set to adjust the model, validation set to control the adjustment process and test data set to measure the validity of the solution with data not used in the training process.

Datamining methodologies by their nature are in a good position to be used in CLV frameworks. The capability of discovering complex relationships on large datasets makes them powerful tools for estimating behaviours especially when individual CLV approaches are chosen, this power arises from the fact that the trained models can be applied to large sets of individual data records as policies or Customers. As referred in De Oliveira Lima, E. (2009) work, the following figure shows a proposed relation between the Datamining process and the CLV concept.

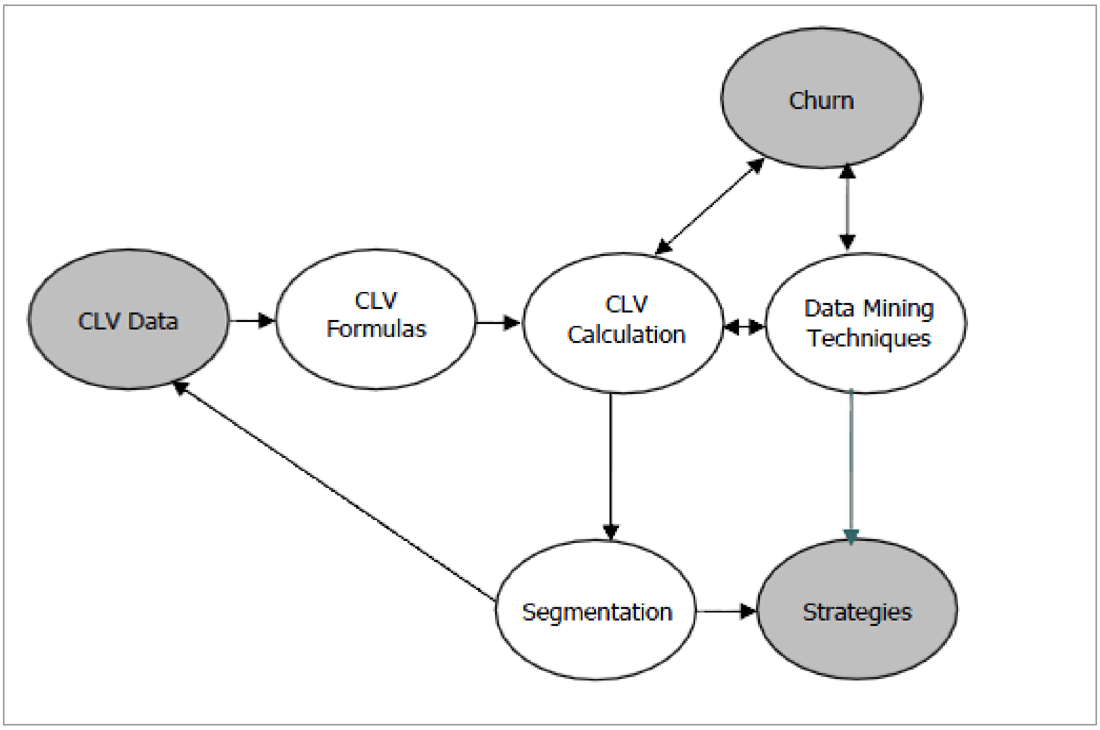


Figure 3.9. Role of Datamining in CLV frameworks

Source: Elen de Oliveira Lima

Recently several authors proposed hybrid methods to improve the effectiveness of datamining techniques to predict significant events, Lariviere, B., & Van den Poel, D. (2005) or Xie, Y. Y., Li, X., Ngai, E. W. T., & Ying, W. Y. (2009), proposed the usage of more sophisticated decision tree algorithms, like random forests and regression forests, to improve the effectiveness of churn and value prediction. The combination of neural networks and automatic clustering detection via Self Organizing Maps (SOM) was also proposed by Tsai, C. F., & Lu, Y. H. (2009) for the same purposes.

Outside of the mainstream of classic datamining techniques some authors propose de usage of more classical statistical methods to predict CLV dependent variables for example Quantile Regression by Benoit, D. F., & Van den Poel, D. (2009).

In the case of non-contractual relations between firms and their customers where future CF streams, due to non-reported dropouts or new transactions, are difficult to measure and therefore associate dependent variables, in these cases Negative Binomial Distribution Processes are preferred as stated in Glady’s, N., Baesens, B., & Croux’s, C. (2009a) paper on modified Pareto/NBD models for predicting CLV.

3.5.8. Hurdles in CLV/CE application

Although the CLV/CE approach appears very attractive, since it incorporates the long term impact of Customer profitability, its application has faced difficulties in implementation, not only by technological hurdles but also by organizational ones. Gneiser, M. S. (2010) noted that in some frameworks the implementation complexity appears to be, to the firm, an obstacle in demonstrating the real benefits of a CLV programme. Nevertheless several studies demonstrated that, during the last 40+ years, the slowly changing in firms marketing and operational paradigms, from product focused to a customer focused is increasing the interest and urgency in CLV methodologies.

> The growing economic importance of customer equity in customer-focused organizations has been investigated by research on customer relationship management and global investment management. A study showed that the investments in customer relationship management systems will double from 2% to 4% of revenue over the next three years (Jeffreys, 2006). This confirms the increasing role of customer equity in organizations. (Abele, Karin P. N. (2009) page 4)

3.5.9. Complexity versus performance

Since one of the objectives of this research work is the characterization of an operational CLV framework it is important to achieve a balance between complexity and practicability. Some studies focused on testing a large variety of modelling approaches as exemplified in Donkers, B., Verhoef, P. C., & de Jong, M. G. (2007) work on Modelling CLV: A test of competing models in the insurance industry where profit regression, RFM, bagging, multivariate choice models, multivariate duration models. The findings showed that although complex models could capture interesting features simple models performed well enough for the purpose.

> In general, management is reluctant to adopt complex models in cases where they perform very well (Verhoef et al. 2003). Thus, for the purpose of predicting individual CLV, the adoption of complex models by practitioners should not be expected. (Donkers, B., Verhoef, P. C., & de Jong, Page 182)

This complexity vs. efficacy aspect is important since implementation and maintenance costs could be significant having an impact in the acceptance on new ways of doing business operations. Building a new process that requires continuous and complex calibration, added to the uncertain results of innovative projects is a challenge in terms of change and project management. Several authors have studied the importance of business integration of Customer Management processes which is addressed in the next section.

3.5.10. Business Process Integration and Systems

Several authors have pointed the importance of having frameworks where a continuous and multidirectional process is maintained to keep the CLV estimates up to date and ready to be used by the organization. One example is the CPM (Customer Profitability Management system) proposed by Wang, H. F., & Hong, W. K. (2006). This framework is multi layered making reference to levels of strategic, tactical and operational activities. Other feature is the distinction between leading and reactive roles interactions between activities or processes. The following figure presents the outline of the framework.

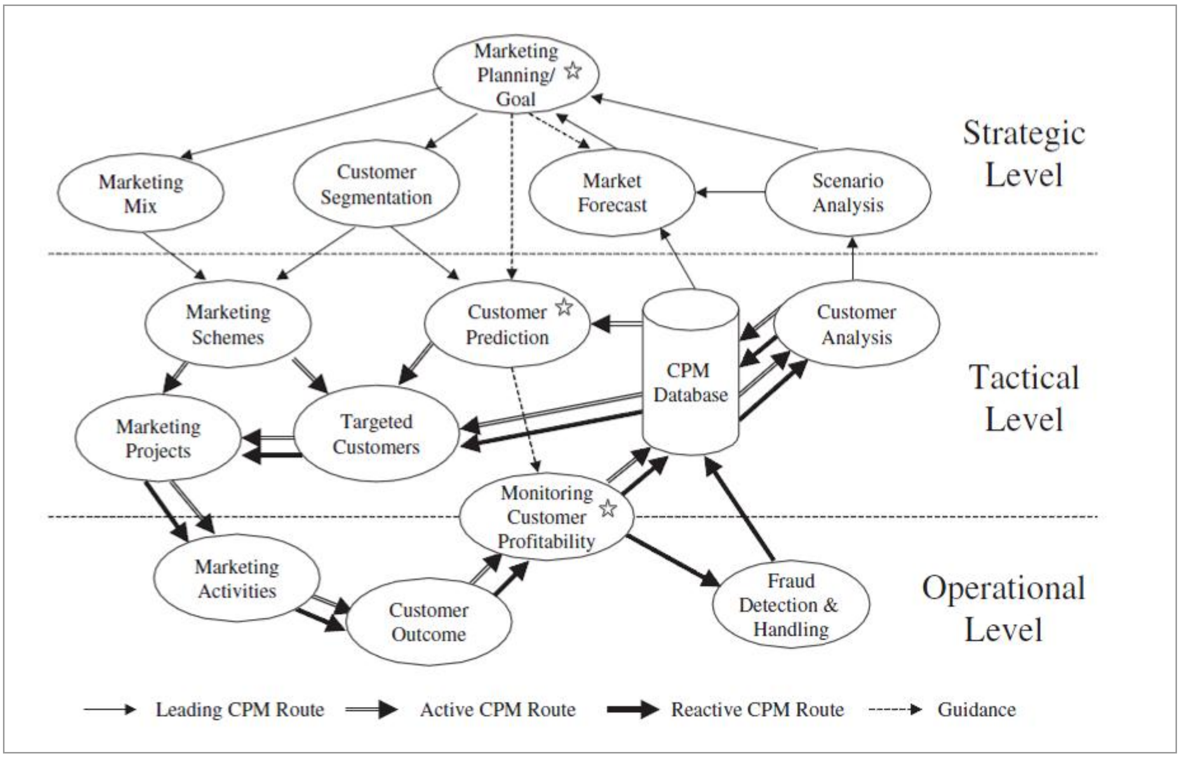


Figure 3.10. CPM - Framework
Source: Hsiao-Fan Wang, Wei-Kuo Hong

The leading role of marketing is highlighted since this function is central as a driver on the firm's strategic alignment process. A similar concept is presented in the CRM System Dynamics, described by Chan, S. L., Ip, W. H., & Cho, V. (2010). where CLV is continuously updated and correlated with the firm strategy and profitability metrics. These authors point out that some key processes must exist for the integration of CLV in the business operations:

a) Analytical Processes for estimating CLV parameters;
b) Automatic Classifiers for CLV calculation;
c) Strategic Planning processes for Marketing Activity Scheduling;
d) Data Management processes for Systems Integration and closing the loop;

With the above requirements in mind we must explore how information technologies are being used in customer centric processes. Studies have been done stressing the ever increasing need for integration between IT activities and other management operations like strategic planning and marketing activities as noted by Gneiser, M. S. (2010).

> Numerous studies in business practice illustrate that many CRM projects fail to achieve their objectives (Becker et al. 2009; Elmuti et al. 2009, p. 76). In particular, CRM projects have mainly focused on a system implementation without considering that this is only one component of a successful CRM strategy (Reinartz et al. 2004, p. 302; Richard et al. 2007, p. 423 f). Thus, not only a customer-centric design of the business model as well as the adjustment and alignment of processes, of application systems and of the infrastructure of the company have to be ensured. Rather, within a value-based management all CRM activities and decisions have to be consistently linked to the business goal of an enterprise, i.e. maximizing the long term shareholder value (Baueret al. 2006, p. 17; Rao and Bharadwaj 2008, p. 16) - Gneiser, M. S. (2010) Page 96

Customer Relationship Managing Systems (CRM's) have being implemented since the early 90's and by now the frameworks for those systems are relatively well known. In figure below the main components of a CRM system are presented.

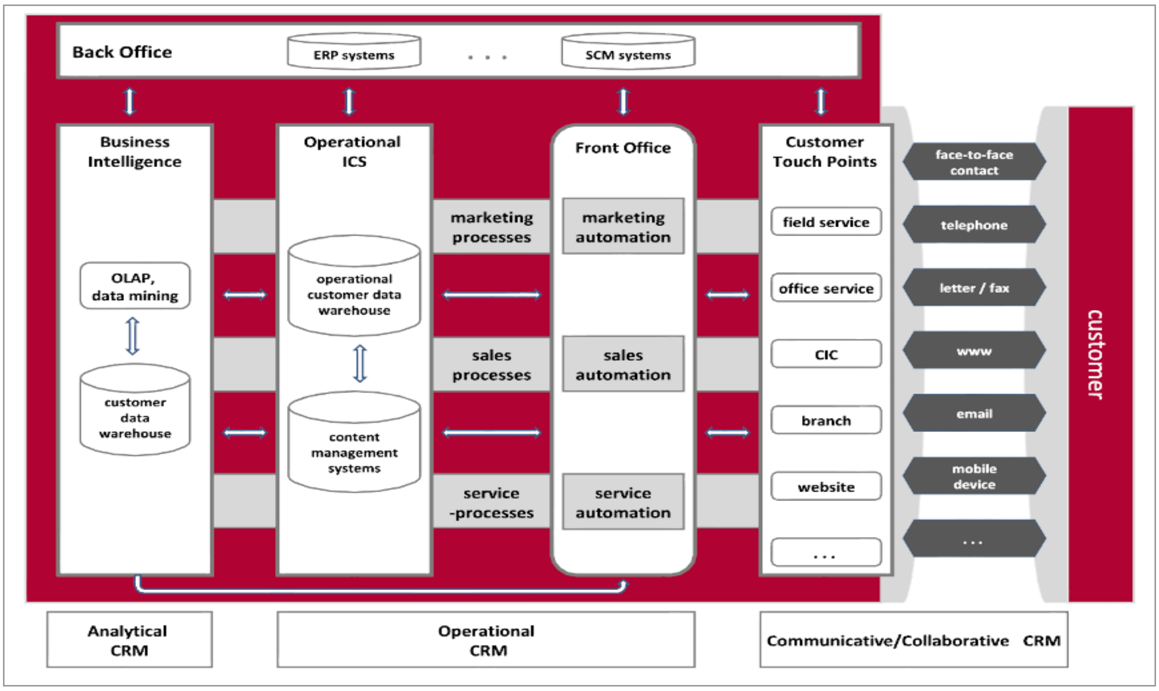


Figure 3.11. CRM Systems - Framework

Source: Martin S. Gneiser

A key distinction is made between the different kinds of CRM: Analytical, Operational and Collaborative. The CRM systems focus on customer operations rather than product operations and therefore creates a scaffolding to support an operational CLV framework since CRM and CLV share common processes described by Kumar, V., Venkatesan, R., Bohling, T., & Beckmann, D. (2008) in their work on Managing Customer Lifetime Value at IBM.

3.6. Literature Review Conclusions

This section closes the Literature review chapter which of course was aimed into understanding the main themes surrounding the CLV concept, the following issues where studied:

- Basic Definition;
- CLV/CE Components description;
- Strategies for CLV/CE frameworks;
- Some Modelling Techniques;
- Business Integration Architecture.

Of course not all aspects were covered or described in depth, that is impossible in the context of this research document, however some of those concerns will be referenced in the further investigation section where potential spinoff research can be identified.

The remainder of this document will be centred in the CLV Project that is being developed in the researched Company and will touch all the previous points.

4. EXPANSION INTO INSURANCE INDUSTRY

4.1. Focus Groups – The CLV Project Birth

As it was already stated this research work occurs in a specific Portuguese Bancassurance Company integrated in a broader financial Group. This process has been continuously improved since the Company has started its operations and, as it has matured, new methodologies have been introduced and recently the **Balanced Score Card** approach has been used to guide the strategic development process. With this methodology the strategic goals and their interactions where summarized into the Company's strategy map. See annexe **A1** for an example on this tool.

One of the useful advantages this tool brought was the quick review of the key areas where specific actions must be taken. The Company's strategic map follows the standard structure with the usual four levels: a) Stakeholders; b) Customers; c) Processes and d) Resources/Learning.

As in many other firms the financial perspective was a central aspect, not the *raison être,* but a very significant one since one of strategic objectives is to create value for the stakeholders. During the discussions the relation between the **Customer Perspective** and the **Financial Perspective** was clear. Since the primary objective it is to maximize the revenue for the stakeholders and simultaneously that objective is dependent of actions taken on the portfolio management.

The CLV approach was introduced in the context of this Strategic Development Process and was agreed that a strategic initiative should be launched aid the articulation of these two perspectives. To track this strategic initiative a workgroup was formed with the responsibility of developing adequate solutions.

4.2. The Importance of the Workgroup

It is vital in innovation or disruptive processes to involve the key individuals, especially if they have some kind of ownership of the business processes being affected. Since this project involves some complex issues the communication and understanding between the different firm cultures is very important for the later acceptance and effective utilization in operations. The project must include, early in its development, the opinions and sensitivities of important business areas, in this case: Marketing and Sales Department. They will put into practice the Customer prioritization accomplished through the CLV Framework. With this in mind the project must derived by a multi/trans disciplinary team.

4.3. Workgroup Structure

To assure the most adequate mix of knowledge and core competences for the project development the workgroup has the following structure.

Table 4.1. *Project Workgroup*

<table>
<tr><th colspan="3">Focus Group Context</th><th>Function</th><th>Expected contributions</th></tr>
<tr><td rowspan="4">Executive Committee</td><td rowspan="4">Extended Technical Team</td><td rowspan="4">Core Technical Team</td><td>Actuary</td><td>Specific product Parameters such as:
• Stationary Premium parameters
• Retention Rates</td></tr>
<tr><td>Financial Managing Director – (Leading Element)</td><td>Specific product information such as:
• Product Balance Sheets.
• Discount Ratios
Solution Evaluation and Approval</td></tr>
<tr><td>Insurance Technical Managing Director</td><td>Specific product information such as:
• Product Structure
• Product History
Solution Evaluation and Approval</td></tr>
<tr><td>Information Management</td><td>Business Intelligence
• Data Access and manipulation;
• Datamining tools operation;</td></tr>
</table>

				• Reporting; • Information Architecture
			Marketing Management Director	**Operational Issues such as:** • Communication with Sales Team • Integration in Marketing Actions **Solution Evaluation and Approval**
			Sales Managing Director	
			Other Technical Areas	**Operational Issues such as:** • IT Integration and Implementation • Quality Assurance
			Board	**Solution Evaluation and Approval**

The workgroup held regular meetings, managed technical work and communicated the results to the management team in the context of strategic meetings. The following sections will summarize the work developed during the research and project development.

In practice the project behind this strategic initiative is also a research project since it incorporates the following activities:

- Literature Review
- Focus Group Discussion
- Experimentation
- Prototyping & Engineering
- Business Process Management
- This Thesis

In the following sections the several topics researched during the development of this project will be addressed. Actual data will be used even though some limitations regarding the disclosure of sensitive data will be in place. The research will focus on each part of the CLV concept with the preoccupation of giving a satisfactory answer in the Insurance context but this activity it is also constrained to the researcher's Company project scope and ambitions.

4.4. Framework Development Strategy Summary

This table summarizes the main aspects of the strategy followed by the workgroup.

Table 4.2. Framework *Development strategy summary*

Function	**Strategy**
Customer Valuation Timeframe	Complete = Past Customer Value plus Future Customer Value
Customer Valuation Formula	Past and Future Customer values are separated. Three components are considered: • Future Customer Value: CLV / CE Concept • Past Customer Value : Discounted • Past Customer Value: Not Discounted
Past Costumer Value function in Framework	Informative only – Needed in the context of marketing operations. Not Used for extrapolation or to build assumptions about Customer Value.
CLV/CE function in Framework	Prospective and the main source of information for Customer Equity and prioritization on marketing operations.
CLV/CE Approach	Mixed: This approach was made during the analysis of the potential parameters to use in the model. While the objective is to maximize individualization it is impossible, or even not desirable (due to the principle of insurance mutualisation), to use individual information since some data are available only in large aggregates.
CLV/CE General Formula	$CLV_{pi} = \sum_{p=1}^{P} \sum_{t=1}^{T} P(Active)_{it} \times \frac{CF_{it}}{(1+dR)^{t}}$
CLV/CE Aggregate Components	For each Product/Segment • Average Cost of taxes; • Average Cost of reinsurance; • Average Distribution Margin; • Average Cash Flow Variations; • Average Claim Cost;
CLV/CE Individual Components	For Each Customer • Probability of Churn • Probability of Cross-Sell – Not Developed • Probability of Up Sell
Individual Components Estimation	By Datamining Methods
Discount rate	Based on CAPM CoC estimation
Time Period for discounted Cash flows	50 years

4.5. Data Used in the Project

4.5.1. Sample Size

The project used all the products in the portfolio but for the sake of simplicity, for this research document only the main products, representing **85%** of the portfolio where sampled. The research sample is made by a set of **412,636** policies belonging to **224.226** Customers. These observations are organized in five period snapshots taken between **2007** and **2010**, as shown in the following table. Each snapshot was taken in 31-Dec.

Table 4.3. Sample Structure

Year	Nº of Policies	Nº of Customers	Policies / Customers
2007	238,263	145,166	1.64
2008	253,133	154,111	1.64
2009	290,969	177,574	1.64
2010	302,880	177,448	1.70

4.5.2. Data Structures

Data were organized in two files: The **product maximal file and the customer maximal file.** The product maximal file is used to study and model events specific to each product and obtain the PLV part of the model, each record represents a policy while the customers maximal file is used to study and obtain the aggregate PLV or CLV part of the model. In the Customer file each record represents a different Customer, product details are lost and product metrics are aggregated or pivoted for each customer. In the annex section the structures of these files are presented. Annexe **A5** shows the data structure for the product maximal table.

4.5.3. Information subjects

The samples cover, for each policy or customer the following information subjects:

- Customer Information (Type, Gender, Literacy, Geography, Age, Marital status)
- Product Information (Type, Duration, Commercial Structure, Campaigns, special usage flags)
- Product Financial Transactions (Price, Costs and Revenue)

4.5.4. Product Structure

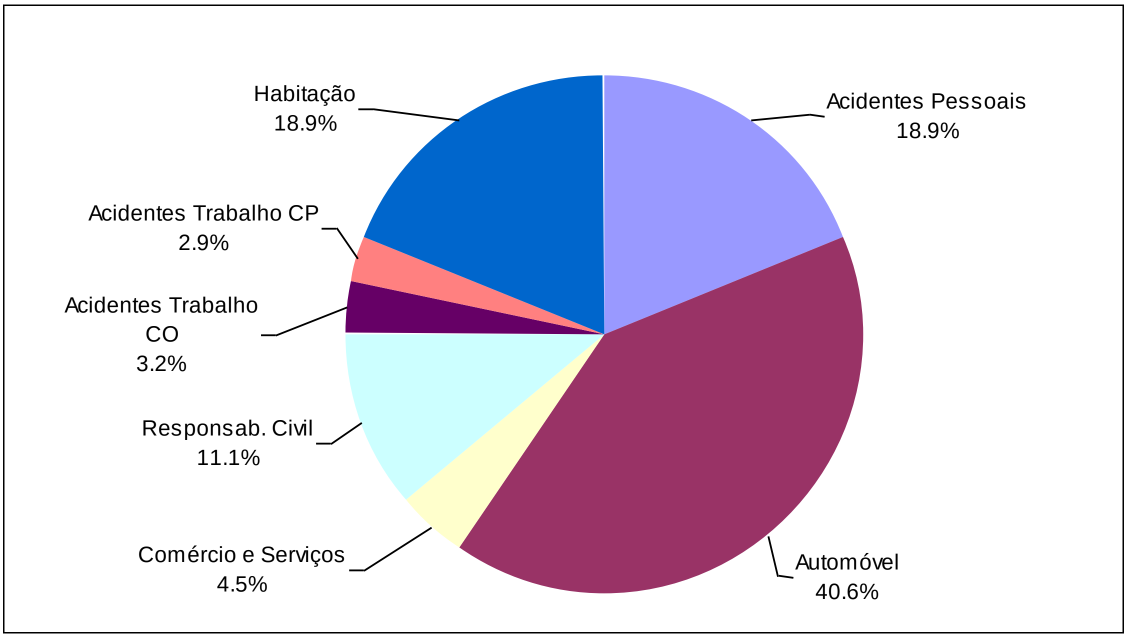


Figure 4.1. Sample Product Structure

Source: The author

For further information on sample statistics see annex **A6.**

4.6. Discount Rate for the Insurance Company

As was stated in the review phase one of the main aspects of CLV methodologies is the determination of appropriate discount rate for cash flows. Starting with the determination of this specific company cost of capital CoC where there is no need of incorporating other sources of financing than the premiums acquired by the operation. In this case the CoC can be approached by the CAPM where the discount rate can be correlated with expectations of investors to get hold of the Company's stock or participate in the distribution margin as an insurance intermediary.

For the determination of the CAPM components: **R_i = Risk Free Rate + Risk Premium** where the **risk free rate** can be based on the Germany Government Bonds with a maturity of 10 years, which in the time frame of the writing of this work[1] the yield had a monthly average of 1.5%.

For the Risk Premium, estimated by β x |E(Rm)-Rf| where |E(Rm)-Rf| is the expected market risk premium for a specific investment and β is the covariance between the earnings of the market portfolio and that specific investment.

For Insurance Companies with activity in Portugal there is no historical data for the β parameters therefore the recommendation is to use international market benchmarks where this parameter holds values between 0.8 and 1.

At last, the market premium risk |E(Rm)-Rf| for Portugal, according to a survey including 56 countries done by Fernandez, P., Aguirreamalloa and J., & Corre, has, in 2011, an average value of 6.5% (stddev = 1.7%; Q1=5.0%; Q3=7.2; max =14%; min = 4.5%) therefore:

$$\mathbf{R_i = 1.5\% + 0.9 \times |6.5 - 1.5\%| = 6\%.}$$

143

[1] During this period Portugal was under the external aid of IMF and EU. Portuguese 10Y Bonds have a yield of 12% due to the fact that those bonds are not perceived as risk free, depending on the development of the aid programme a "*haircut*" or even a credit event may occur. German bonds were a better free risk option.

For this parameter WACC was also proposed but was rapidly rejected since the Company's financing profile doesn't match the methodology. The main source of cash is the paid premiums. The current Company's debt is residual and there is no significant cost associated with firm's equity. So the CAPM based on external parameters was preferred.

4.7. Retention Rate

One of the advantages of the insurance business, in the context of a CLV programs, is its contractual nature embodied in the policy object. Each policy is a formal contract, ruled by general and specific legislation, which the Customer signs on. Contrary to non-contractual forms of business, like retail for instance, where the ending of a relation between the Customer and the Company is blurry needing some kind of recency methods for its estimation, as noted by Glady's, N., Baesens, B., & Croux's, C. (2009a), the termination of a policy type contract is a well acknowledge and formal event agreed between the parts to occur in a specific date. This comes into help to define the churn concept in the Insurance Industry.

Depending on the size of the Customer policy portfolio we may define two levels of churn: a) A contract type termination which we correlate with product retention rate b) whole portfolio termination which we correlate with customer retention rate. In the insurance industry, most policies types have annual terms with automatic renewal, which we call an annuity. The premium that guarantees the compensation for a well determined type of loss is paid in the beginning of each annuity. In modern legal insurance frameworks the termination of a contract can occur at any time during the annuity, in this case the Insurance Company reimburses the customer, in *prorata temporis* base, with the non-acquired premium for the remainder of the annuity, at that moment any Company's liability over a claim or loss is also terminated.

4.7.1. Aggregate Churn Rate

With this principle it is possible to define a portfolio retention rate by the simple ratio between the number of **policies in force** (Active) in one particular point in time and the **number of surviving** (not cancelled) policies in other point in time. For this effect the annual base is very convenient since a policy contract, in general, works by a stream of automatically renewed annuities or twelve month terms. In a rich and sufficiently large portfolio, the policies **starting term dates or termination dates** are determined by uncorrelated, independent and emerging processes depending largely on the Customer will and the Company's daily commercial affairs. Therefore we can assume statistical independence between the events behind the churn ratio as described before and accept that relation as sufficiently good as a random variable defined by the formula (11).

$$rR = Ne / Ni \tag{10}$$

Where

- rR is the retention rate;
- Ne is the number of policies in the end of the period;
- Ni is the number or policies in the beginning of the period;

Other immediately derived concept is the **erosion or Churn** rate rE, or desertion rate, depending if we are measuring policies or customers. This concept measures the rate at which the Company's loses policies or Customers and can be defined by the following formula:

$$rE = 1 - (Ne / Ni) \tag{11}$$

The usage of one of these concepts is a matter of convenience or adequacy in terms of communication or modelling, one is centred on survivability rate and the second on the erosion/churn rate of a portfolio.

4.7.2. Individual Churn Rate

While the aggregate churn rate is appropriate for usage for the whole company or segment, in the context of CLV a richer approach entails the estimation of the probability of a specific customer being active after a known time frame. The idea is to break down the retention rate in **multifactorial dimensions** that represent unique characteristics of the customer's as a person plus the contextual commercial relation with the Company.

Usually the relations between these dimensions are complex and the usage of statistical or datamining techniques come to our help. An individual retention rate model can be seen as a probabilistic model with the following general function:

$$rR_i = \Phi(Ne, Ni, \langle \arg vector \rangle) \tag{12}$$

Where

- rR_i is the individual retention rate for a customer.
- Ne is the number of policies in the end of the period;
- Ni is the number or policies in the beginning of the period;
- $\langle \arg vector \rangle$ is the set of dimensions needed by the algorithm to estimate the individual retention rates.

The Φ model can be determined by an analytic datamining/statistical process (e.g. SEMMA, CRISP-DM) and can be any combination of regression methods or supervised learning algorithms such as a neural network or a decision tree. For illustrative purposes the following figure shows the mechanism trough a very simple decision tree. Further information will be available in the section on datamining techniques.

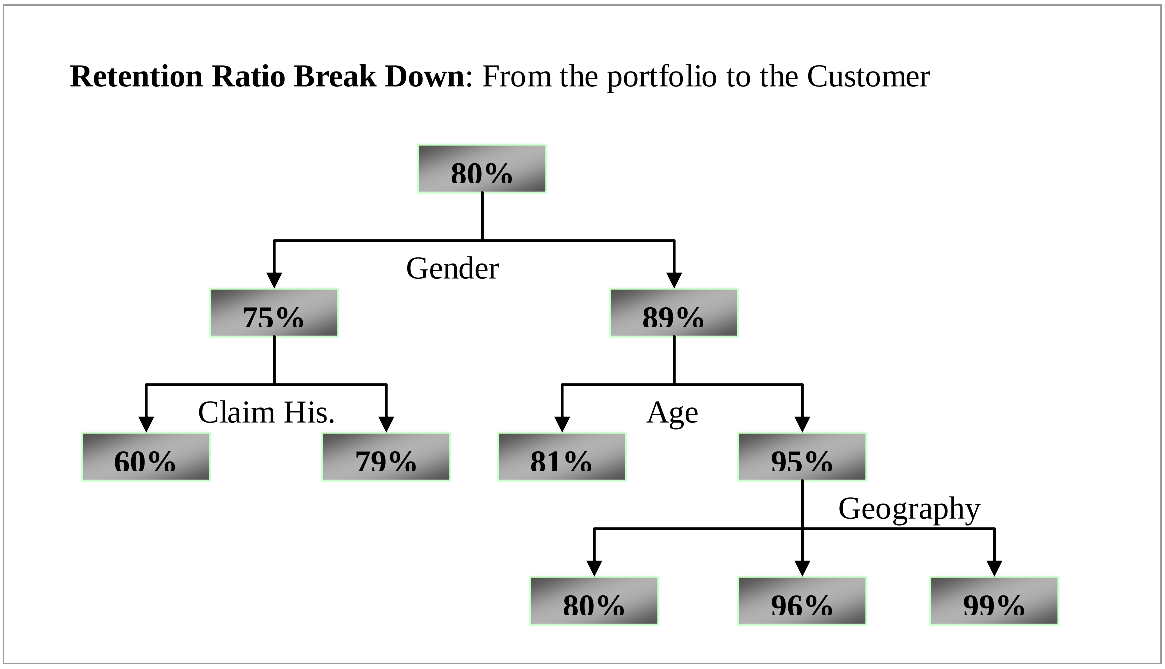


Figure 4.2. Individual Retention Rate Estimation Example.

Source: The author

4.7.3. Retention Rate and the Average life of a Policy

In aggregate terms there is a strong relation between de average life of a policy and the respective portfolio erosion/churn rate. As an example, if a portfolio has an erosion rate of 25% (annually) that means that in average 1 in 4 randomly chosen policies will be terminated in one year. In another view, it is expected that in average 1 randomly chosen policy will be terminated after 4 annuity years. Assuming the statistical independence of events we can write the following linear relation:

$$pL = 1/rE = \int_{t=1}^{\infty} Pactive(t)dt \quad (13)$$

Where:

- **pL** is the average policy Lifetime in a reference period;
- **rE** is the policy erosion rate in the same reference period;
- $P(active)$ is the probability of being active

An interesting aspect is that on a differential probability function for retention rate in order to time, its integration is equal to the average policy lifetime. The relation between the annual portfolio churn rate (or erosion) and average policy life is shown in the figure bellow.

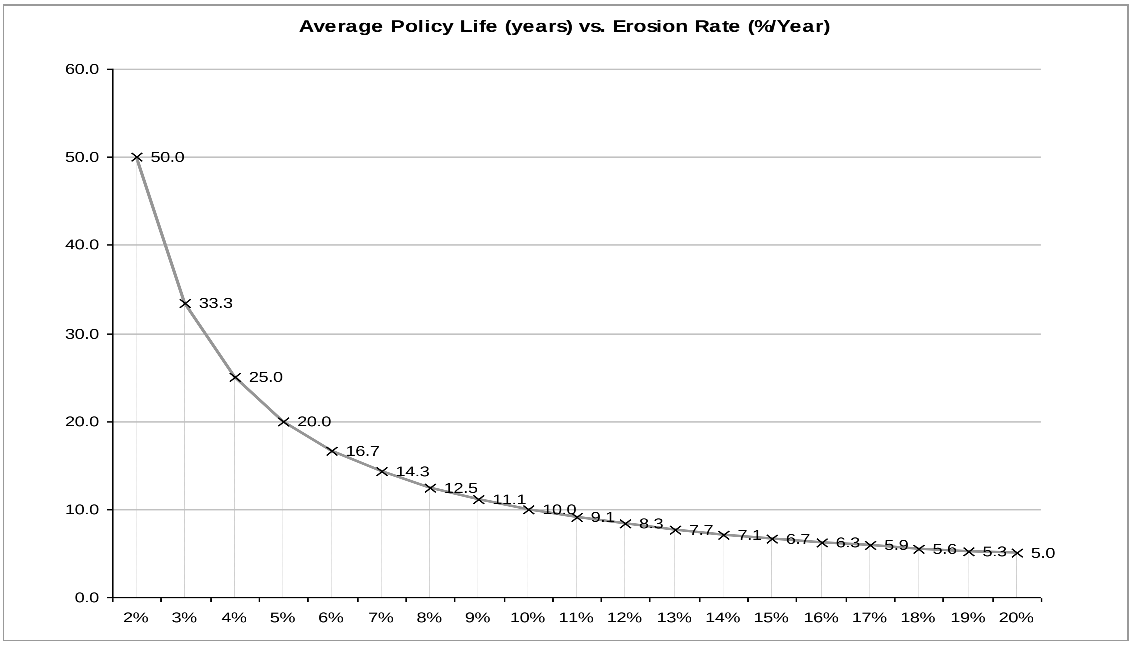


Figure 4.3. Average Policy Life vs. Erosion Rate

Source: The author

It is clear, trough this relation between retention rate and policy average life, that the erosion rate has a vital impact on the capacity of a portfolio to generate Cash Flows. With **2% of erosion rate**, in average, it is **expected a CF stream for 50 years**, contrasting with **only 5 years for an erosion rate of 20%.**

Equations (10) and (11) were adopted in the Company as the standard metric for churn. The process that led to this adoption preceded this project by at least three years. It directed the creation, by the researcher, of an information system that measures this, now accepted KPI, trough several dimensions: Month; Customer Type; Product; Geography, etc.

The adoption phases were:

- Internal Discussion – Technical Forum (informal: The Guild of Wise Men);
- Data Analysis – By the researcher;
- Prototyping - By the researcher;
- Final Solution Project Development – By the Researcher and IT Team;
- Rollout and maintenance by IT Team;

4.8. Present Value Function

Since we already defined, in the insurance context, how to calculate two of the most important parameters of PLV (retention rate and discount rate) it is useful to show how their values impact on the Present Value of a policy. For that effect a function of the Present Value for a policy was designed using the formula (1) as the kernel. The function is calibrated for a maximum period of T= **50 years** a time span large enough to cover almost all situations. The vector $\langle \text{var}\arg vector \rangle$ is null to eliminate any product/customer induced complexities and CF is constant and equal to 1 for all periods. The function output domain is in monetary units. The Discount rate dR and retention rates rR domains are calibrated using data from the author's researched Company portfolio. The general formulation for the PV function is shown below.

$$PV = \Psi(dR, \langle CF_t(rR, \langle \text{var}\arg vector \rangle = null) = 1 \rangle \ , T = 50) \quad (14)$$

The following graph is based in 5000 stochastic simulations of the function (14)

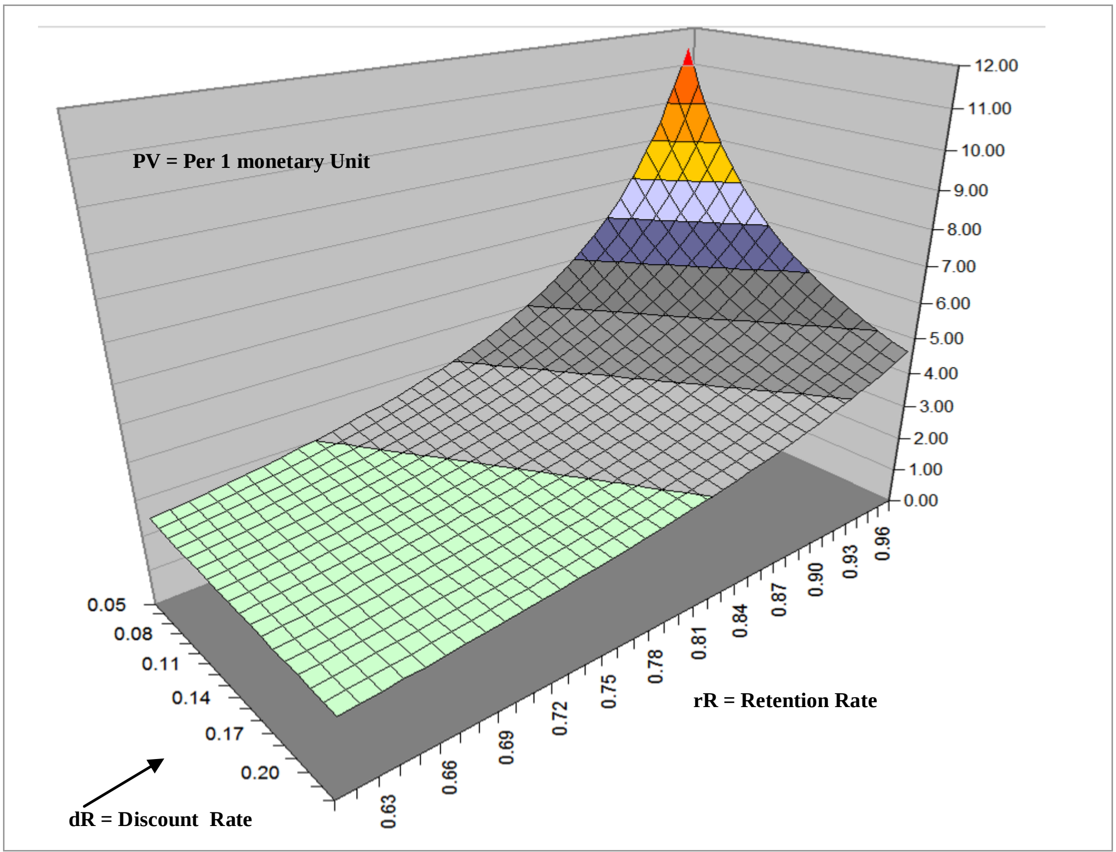


Figure 4.4. Present Value Function

Source: The author

Assuming a zero cost business operation this function translates each **present annuity policy premium** monetary unit into the respective discounted value in the same monetary unit.

For example for a policy with dR=0.05 and rR= 0.96, each € of premium actually has a present value of factored by 11.53, if the rR would be 0.80 the policy would have a present value factored by only 4.20, in other words, less 63%.

This simplification is only instrumental to show the effect of these two parameters. If the reality would be represented only by the surface formed by these 5000 simulations, as shown in the previous figure, the calculation of the PLV of any policy would be trivial as it would lay in any part of that surface: PLV = Ψ x (Annuity Premium) therefore CLV = Σ PLV and finally CE = Σ CLV.

The challenge of a CLV/CE programme is defining and calculating the $\langle CF_t(rR, \langle \text{var arg } vector \rangle) \rangle$ part of function (14), the used notation denotes the cash flow stream vector in which element is the **CF** in time **t**, the discount rate ***dR*** is assumed as being global.

This challenge has two components: **a)** Determining PV function dimensionality, higher dimensionality emerges for inclusion of **cash flow variations** along the policy life. **b)** Determining for **each policy where it lays in the multidimensional PV** surface, those issues will be treated in later sections.

Another key conclusion taken from this exercise is that not all Euros are equal, this is very important from the business point of view. Insurance companies, although having a prudential mechanism embedded in their business model, are used to reason and act, mainly outside their actuarial or financial areas, in terms of **present premiums**: **a)** Sales Objectives; **b)** Market Benchmarking, and not in **present value**. Making this difference more notorious and useful in the daily business operations would be transformational on how risk and value would be managed.

A more complete IT implementation of this function can be consulted in annexe **A13**, in annexe **A3** it is presented the table behind **Figure 4.4.**

4.9. A note on experimentation

To give some insight on the more practical aspects of this work, before the full CLV formula definition there was an experimental work developed and shared by the workgroup using essentially MS Excel to test ideas and potential parameters.

The work **started around formula (1)** and its expansion through a simple running table where cash cash-flows factors were explicitly calculated and affected only by a hypothetical discount rate. The final result was the much more general and abstract **formula 14**. In the next figure a more developed version of one of those tables is presented. The table runs for 50 years, the sum of the last column is the CLV. In the presented example a unitary premium is used since the resulting CLV (**a)** is a general linear factor that can be applied to any case. For example, if **a** = 0.5, then if a policy has a premium of 500€ its CLV will be 250€.

Ano	Prémio Anual (se estiver em Vigor)	% Desvalorização do Prémio	Factor de Crescimento Anual * Factor de Estacionaridade	Força do Factor de Estacionaridade (inv exp)	Margem Anual sobre Prémios, no Produto	Probabilidade de Apólice estar em Vigor no final do Ano	Factor de Actualização Anual (para o final do ano 0)	Valor actualizado dos Cash-Flows futuros esperados
Ano 0	1.000	0.0%	na	na		100.0%	1.00000	
Ano 1	1.000	0.0%	1.000	1.000	49.8%	88.8%	0.93120	0.4376
Ano 2	1.000	0.0%	1.000	0.980	49.8%	78.9%	0.86714	0.3619
Ano 3	1.000	0.0%	1.000	0.961	49.8%	70.1%	0.80748	0.2993
Ano 4	1.000	0.0%	1.000	0.942	49.8%	62.2%	0.75193	0.2476
Ano 5	1.000	0.0%	1.000	0.924	49.8%	55.3%	0.70020	0.2048
Ano 6	1.000	0.0%	1.000	0.906	49.8%	49.1%	0.65203	0.1694
Ano 7	1.000	0.0%	1.000	0.888	49.8%	43.6%	0.60717	0.1401
Ano 8	1.000	0.0%	1.000	0.871	49.8%	38.7%	0.56540	0.1159
Ano 9	1.000	0.0%	1.000	0.853	49.8%	34.4%	0.52650	0.0958
Ano 10	1.000	0.0%	1.000	0.837	49.8%	30.6%	0.49028	0.0793
Ano 11	1.000	0.0%	1.000	0.820	49.8%	27.1%	0.45655	0.0656

Figure 4.5. CF Table generated in Excel for testing purposes.

Source: The author

This schema, with the expanded running unitary Cash Flow stream (CF=1), provided a test bed where several considerations could be easily tested, the worksheet just ran the numbers, for example:

- What is the sensitivity of the discount rate on the actual value of an specific policy or even a portfolio;
- How to affect the CF factor. This parameter is now the most important one for the development of the rest of the model, several scenarios and assumptions were tested, for example:
    - What is the probability of a policy being active in the following period;
    - How does the Cash in and Out vary on each period, what parameters must be used?
    - It is possible to expand this approach for individual policies and Customers?

During this phase, trough the study of other research work on the matter and also the ideas discussed by the workgroup, the most adequate approach was defined. The next topic will detail how the CF parameter is organized.

4.10. The Cash Flow Parameter

This section addresses the Cash Flow (CF) parameter in the CLV formula in the context of this project. As other parameters in the CLV model the CF parameter has individual or aggregate sub components as depicted in the following summary table.

Table 4.4. Cash Flow Estimation Strategy

<table>
<tr><td rowspan="4">CF = CI (Cash In) – CO (Cash Out)</td><td rowspan="2">CI</td><td>Individual</td><td>• Premiums Known for each policy
• BMS – Is applied for Auto Insurance</td></tr>
<tr><td>Aggregate</td><td>• Premium variations</td></tr>
<tr><td rowspan="2">CO</td><td>Individual</td><td>• Not used</td></tr>
<tr><td>Aggregate</td><td>• Cost of taxes;
• Balance of Reinsurance;
• Distribution Cost;
• Claim Cost;</td></tr>
</table>

Of course all aggregate parts could be individualized using some kind of mathematic model as is done in this work for churn estimation. The following sections address some questions behind the CF parameter in the context of insurance.

4.10.1. Structural Cash Flow Variations

During product design, depending on product desired characteristics, the technical team may have to include some requirements for automatic premium variations. In some cases that requirements arise from the need of claims cost control trough an a posteriori rate making system or when the long term nature of contract recommends regular premium adjustments, in some cases these variations obey to legal requirements. In the following sections we're looking into some of factors affecting the Cash Flow stream with obvious impact in PLV and CLV.

4.10.1.1. Indexation and devaluation

Some products have premium variations enforced by legislation. In the case of home insurance the premium rises every term due to an indexation mechanism that depends on geography and property age. In auto insurance the premium, for the comprehensive insurance covers (loss or damage to the policy owner vehicle), lowers due to the vehicle devaluation. These premium variations were introduced by the legislator to reflect, with some fairness, the damaged or lost property market value.

In these cases, since the information is known for each policy, the CF variation can be integrated in the PLV/CLV formula. This can be done individually or in aggregate terms depending on the level of refinement and complexity the organization wants for its CLV model.

4.10.1.2. Bonus-Malus System in Auto Insurance

Most insurers around the world have introduced some form of merit-rating in automobile third party liability insurance. Such systems, penalizing at-fault accidents by premium surcharges and rewarding claim-free years by discounts, are called bonus-malus systems (BMS) in Europe and Asia. Those systems are based on a transition table where each entry corresponds to a bonus or malus level. See annexe **A4** for a BMS transition table example.

From the actuary point of view the policy transitions in such a system can be formally described stochastically by a Markov process which study is out of the scope of this work. Nevertheless, depending on the Company portfolio structure the expected effects of a BMS should be considered for CLV estimation purposes.

The researched Company has large Auto Insurance exposure therefore the BMS effects will be measured. The System has 20 levels numbered 1 to 20, the 7th level is the entry level for a policy where there is no surcharge or discount as depicted on the following figure.

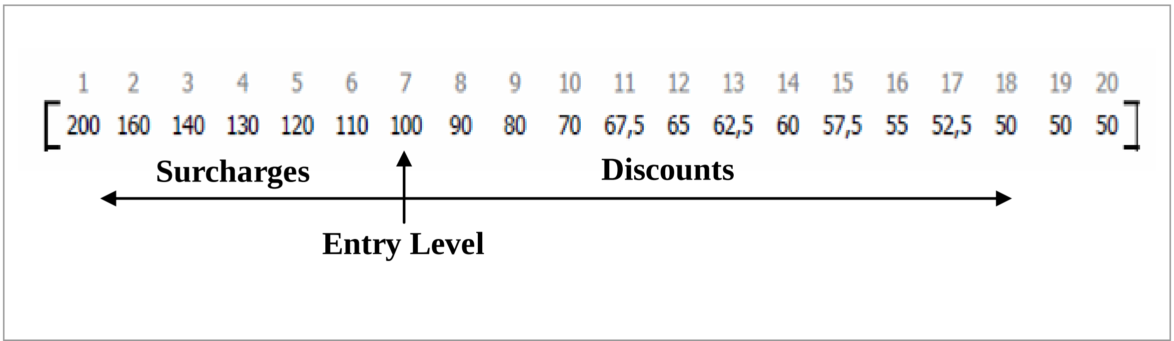


Figure 4.6. The BMS premium vector, Level vs. Premium multiplier

Source: The author

For levels higher than the entry one, 8 to 20, discounts can be attributed with a range between 10% and 50%. The maximum bonus can be achieved in 10 years if there are no faulty claims.

For each faulty claim there is a regression on the system depending on the current level and also on the number of claims on the term.

For lower levels, from 6 to 1 there are surcharges between 10% and 100%. Depending on which level the policy holder is, in average the system pulls back for the previous level after 4 year without faulty claims. An apparently obvious effect of the implementation of a BMS is a progressive decline of the observed average premium level, due to a concentration of policies in the high-discount levels. In the case of the researched Company, after applying the Markov process, the premium decay is shown in **Figure 4.6.**

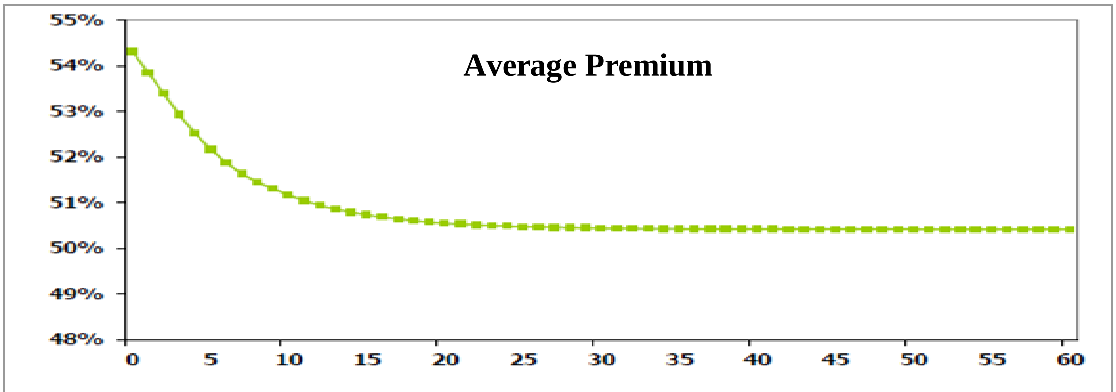


Figure 4.7. Evolution of the Average Premium in the BMS

Source: The author

The average premium level for which the system converges is called the stationary premium. In this case it corresponds to 50.42% of the entry premium. Loosing 50% of the cash-in on an essential product is a relevant effect to factor in the Cash Flow stream.

4.10.1.3. The Stationarity Function

As was pointed out in the previous section the estimation of the decay of average premium is relevant. There must be any kind of provision to include the BMS in the PV formula, the vector $\left\langle CF_t(rR, \left\langle \mathrm{var\,arg}\, vector \right\rangle \right\rangle$. This may be done by estimating the future term premiums factoring a decay factor correlated with the BMS. There are two approaches: **a)** a more formal that evolves the stochastic development of the Markov process associated with the Company's BMS. The following general function encapsulates the system

$$aP_{ti} = BMS(cL, \langle tM \rangle, cP) \qquad (15)$$

Where:

- aP_t is the average decayed premium factor in time **t**
- cL is the current BMS level;
- $\langle tM \rangle$ is the BMS transition table (see annexes);
- cP is the claim probability for policy **i**;
- **i** is the policy index;

Another approach **b)** is using an engineering proxy developed during this research work as described below. This approach has the following assumptions:

- The average premium shows some kind of exponential decay expressed in the empiric form and was determined by trial and error.

$$aP_{ti} = (1 - sP) + (a/t^b) \quad (16)$$

Where:

- $\mathbf{aP_t}$ is the average decayed premium factor in time **t**
- **sP** is the known relative stationary premium;
- **a** and **b** are form factors than can be determined by any means, including stochastic methods;
- **i** is the policy index;
- $\mathbf{aP_t}$ = 1 when sP =0; a=0; b=0);

- The product portfolio has a small faulty claim frequency

In the case of the function (16) parameters **a** and **b** where also determined by trial and error. Although it would be much more interesting using a Monte-Carlo method minimizing an error function since we know the form of the real one straight from the BMS.

The Auto Insurance claim distribution of the researched Company adjusts to a negative binominal distribution. In terms of stationarity, the BMS converges on the long term, to the following policy distribution on the BMS levels

BMS Level	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Distribution	0.00	0.00	0.00	0.00	0.01	0.02	0.04	0.07	0.10	0.25	0.23	0.22	0.22	0.68	0.65	0.86	0.81	5.05	4.78	85.98

Figure 4.8. The Company’s expected stationary distribution.

Source: The author

In the actual portfolio, 0.22% of policies have a malus condition, 0.55% are neutral and 99.23% have some kind of bonus. The average premium is 52.68% of the entry premium comparing with 50.42% in stationarity. These data means that there is a large concentration of policies in the higher discount level and due to a low claim frequency it is relatively easy to for the portfolio to converge to stationarity. The second assumption is met.

As an example, combining the formulas (6) and (16) we can define a CLV formula that integrates the BMS on an individual PLV/CLV approach the following formula.

$$CLV_i = \sum_{n=t+1}^{t+x} P(Active)_{in} \times \left((1-sP) + \left(a / (n+o_i)^b \right) \right) \times \frac{AGC_{it}}{(1+d)^n} - \sum_{n=1}^{x} M_{in} \times \left(\frac{1}{1+d} \right)^{n-1} - A_i \quad (17)$$

Where $\left((1-sP) + \left(a / (n+o_i)^b \right) \right)$ is the stationary function derived from (16) but with an extra term o_i that is an offset calibrated for each policy depending on its age on the portfolio. Older policies will have lower decays as it is assumed they're in higher levels on the BMS.

4.10.2. External Cash Flow Variations

4.10.2.1. Soft markets

The tem soft market refers to a market that has more potential sellers than buyers. A soft market can describe an entire industry, such as the Insurance Market. This type of market status is often referred to as a buyer's market as they have a greater power in determining the price of transactions. This situation can lead to a rapid price drop as sellers compete for the potential buyers. In recent years, due to the economic situation, the insurance market in Portugal is living a soft market status, where the whole non-life insurance market as shrunk. The large Insurance Companies, due to their dimension are more exposed having a greater negative impact on their portfolio. The erosion of the portfolio as the following mechanisms:

- Increased Churn;
- Decreasing Coverage on mandatory policies (e.g. Auto Insurance)

- Smaller Rates as companies try to recover competiveness;

4.10.2.2. Transparent Market

A transparent market is any market where all relevant information is fully and freely available to the public. This type of market operates in a higher degree of efficiency leading to smaller margins as firms compete for the available resources. The transparent market applies to the Insurance market depending on how many Companies operate and how mature the marketing processes are. In the case of Insurance Companies there is also a differential transparenceness depending on what kind of customers were dealing with or if the type of product is bases more or less in a uniform policy, for example: **a)** Corporate Customers are more pre-emptive in price negotiation; **b)** Worker's Compensation is a transparent product since its legal framework is complex uniform on the all companies, it's a mandatory policy for most open businesses and again, with this product, Corporate Customers tend to be very assertive looking for the best rate.

Understanding these phenomena is important since they affect the Cash Flow Streams, a good calibration of CLV models will reduce the potential distortions we can introduce by overlooking them. The next section presents an approach to incorporate the previous concerns in the CLV model.

4.10.3. Study of the Cash-In Aggregate Variations

The Study of the cash-in aggregate variations was done using the following method: Five portfolio snapshots or time-slices where sampled with one year apart. To eliminate **distortions induced by churn or new business** a "closed" portfolio was created by considering the common/surviving policies along these sequential snapshots. The variation on the annuity premium was then measured for each product as shown in the following in **Table 4.5**.

Table 4.5. Average Premium Variation (*"Closed"* Portfolio)

Average Premium		Year				Premium Variation				Average Var. MA3
Customer Type	Product	2007	2008	2009	2010	2007	2008	2009	2010	
Corporate	Acidentes Pessoais	196.8	200.0	200.4	201.4	n.a.	1.6%	0.2%	0.5%	0.8%
	Acidentes Trabalho CO	1,815.7	1,895.0	1,994.8	2,036.2	n.a.	4.4%	5.3%	2.1%	3.9%
	Acidentes Trabalho CP	264.2	278.3	294.7	301.7	n.a.	5.4%	5.9%	2.4%	4.5%
	Automóvel	417.4	387.5	368.3	353.4	n.a.	-7.2%	-5.0%	-4.0%	-5.4%
	Comércio e Serviços	297.5	302.6	307.8	311.6	n.a.	1.7%	1.7%	1.2%	1.6%
	Habitação	217.6	215.1	226.2	226.5	n.a.	-1.1%	5.2%	0.1%	1.4%
	Responsab. Civil	469.6	483.1	494.6	496.0	n.a.	2.9%	2.4%	0.3%	1.8%
Individual	Acidentes Pessoais	70.8	71.3	72.2	72.9	n.a.	0.7%	1.3%	0.9%	1.0%
	Acidentes Trabalho CO	565.1	600.5	600.8	619.9	n.a.	6.3%	0.0%	3.2%	3.2%
	Acidentes Trabalho CP	257.2	270.1	287.9	293.5	n.a.	5.0%	6.6%	2.0%	4.5%
	Automóvel	222.3	210.6	200.9	198.0	n.a.	-5.3%	-4.6%	-1.5%	-3.8%
	Comércio e Serviços	132.0	134.0	136.4	138.9	n.a.	1.5%	1.8%	1.8%	1.7%
	Habitação	101.7	100.3	105.7	105.5	n.a.	-1.4%	5.4%	-0.2%	1.3%
	Responsab. Civil	31.4	31.4	31.2	31.2	n.a.	-0.1%	-0.4%	-0.1%	-0.2%

The Cash-In variations where estimated by a three datum (Years) Moving Average MA3. Due to largely different business characteristics two Customer segments where studied: Individual and Corporate. The estimated values (excluding Auto Insurance due to the BMS) where used as aggregate CLV linear parameters, keeping their rate of variation along the CF Stream. These parameters should be regularly reviewed since they can change due to they're external nature.

4.10.4. Study of the Cash-Out Aggregate parameters

4.10.4.1. Claims

Contrary to premiums that are: **a)** guaranteed quantities for every regular in force policy; **b)** known in advance; **c)** paid at the beginning of the agreed exposure period; Claims have a stochastic nature. Usually in a healthy insurance portfolio their frequency shouldn't be very high although Claims must be analyzed in their dual nature: frequency and severity. The following figure shows the frequency in the researched insurance company.

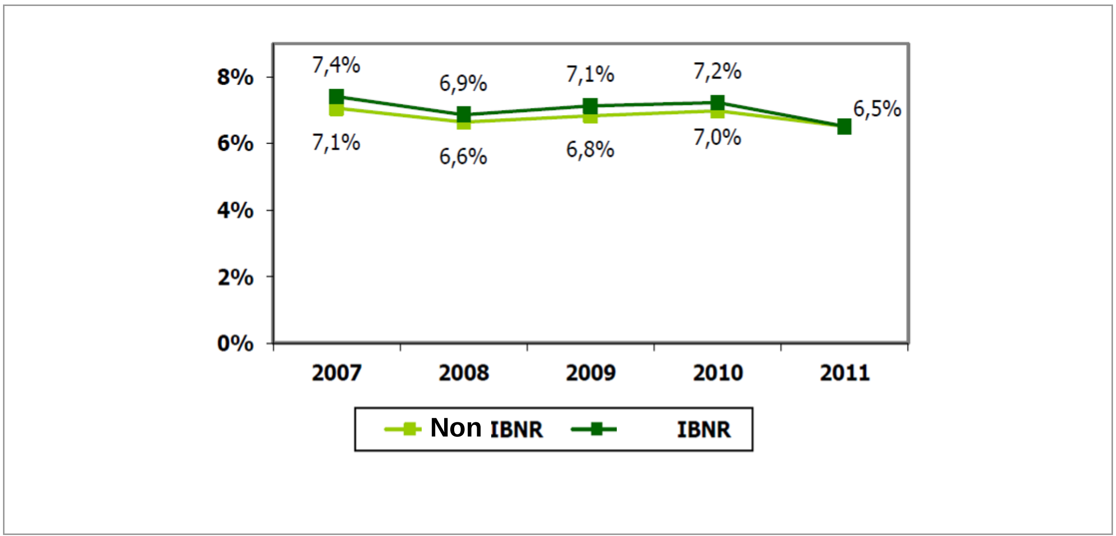


Figure 4.9. Global Claim Frequency

Source: The author

Building an individual stochastic claim sub model is a difficult job instead the usage of aggregated data from the Company's balance sheet was chosen.

4.10.4.2. Other aggregate Cash-Out parameters

Marketing, Taxes and Distribution Costs plus Reinsurance Balance where also obtained from the Company's balance sheets and are detailed by LoB / Product. The following section addresses how balance sheet can be used to parameterize the CLV model.

4.11. Balance Sheet Information Integration

By nature balance sheet data are aggregated quantities which dimensionality is usually limited to LoB, Channel or Subsidiary. Since we're applying an aggregate CLV approach there is the need to transpose the aggregate Balance sheet data into the desired level of granularity, in this case, policies. The method is to generate an appropriate Cash-in/Cash-out uniform relation which we can call the margin factor ***mF***. Our reference are the acquired commercial premiums, from with the proper cost element are deducted. Usually these factors are those related with the technical margin determination: Claims, Reinsurance Balance as shown in the following formula.

$$mF = 1 - (cR_1 + cR_2 ... + cR_n) \quad (18)$$

Each part $\boldsymbol{cR_n}$ is expressed by the ratio between acquired commercial premiums ***aP*** and the desired Cost Element. See annexe **A9** for further detail and example.

$$cR_n = aP / C_n \quad (19)$$

Depending on the CLV model strategy, more parameters may be considered, if the objective is to approximate the CLV model to a CE model, non technical information ought to be used such as salaries or services cost. In these cases an imputation strategy must be determined.

4.12. Product CLV Integration Strategy and CF Parameters

The following tables present the strategy used to integrate the Balance Sheet information per product. Most parameters have an aggregate nature minus the churn ratio and the specific CFSV for Auto Insurance.

Table 4.6. Product CLV Integration Strategy Matrix

Product	Churn Ratio	Upsell Ratio	CF Structural Variations	Claim Cost Ratio	Reinsurance Balance Ratio	Distribution Costs Ratio	Corporate Taxes Ratio
Automóvel	I	I	I	A	A	A	A
Acidentes Trabalho CP	I	I	A	A	A	A	A
Acidentes Trabalho CO	I	I	A	A	A	A	A
Habitação	I	I	A	A	A	A	A
Acidentes Pessoais	I	I	A	A	A	A	A
Comércio e	I	I	A	A	A	A	A

Serviços							
Responsab. Civil	I		A	A	A	A	A

A – Aggregate Approach; I – Individual Approach;

The actual calculated cR_n parameters are displayed bellow.

Table 4.7. Product Cash-Flows Parameters

Product	Churn Ratio	Upsell Ratio	CF Structural Variations	Claim Cost Ratio	Reinsurance Balance Ratio	Distribution Costs Ratio	Corporate Taxes Ratio
Automóvel	n.a.	n.a.	0 (Given by the SF)	0.6130	0.0870	0.0820	0.0050
Acidentes Trabalho CP	n.a.	n.a.	0.0450	0.9100	0.0300	0.0830	0.0050
Acidentes Trabalho CO	n.a.	n.a.	0.0380	0.6730	0.0300	0.0830	0.0050
Habitação	n.a.	n.a.	0.0150	0.2820	0.1580	0.2080	0.0023
Acidentes Pessoais	n.a.	n.a.	0.0090	0.3000	0.0150	0.2100	0.0023
Comércio e Serviços	n.a.	n.a.	0.0150	0.5330	0.1450	0.1500	0.0023
Responsab. Civil	n.a.	n.a.	0.0070	0.5560	0.1470	0.1600	0.0023

SF is the stationary function

4.13. The Project CLV Formula

The CLV formula used in this project is defined below. Due to space constraints and readability the formula is presented in four parts: **a)** CF part; **b)** discounted CF stream part; **c)** P(active) part and the **d)** Upsell part.

a) The Cash Flow part is defined by the formula

$$CF_{it} = cP_i \times [1-(cCr_i + rB_i + dC_i + cT_i)] \times (1+sV_{it}) \times \left((1-sP) + \left(a \big/ (n+o_{ti})^b \right) \right) \times (1+E_{it}(P(Upsell))) \qquad (20)$$

Where

- **i** is the policy index;
- **t** is the period;

- **cP** is the Commercial Premium;
- **cCr** is the Claim Cost Ratio;
- **rB** is the Reinsurance Balance Ratio;
- **dC** is the Distribution Cost Ratio;
- **cT** is the Corporate Cost Ratio;
- **sV** is the expected structural CF variation;
- **a** and **b** are form factors for the Stationary function decay;
- **o** is the Stationary function offset;
- **sP** is the Stationary factor if applicable (if not sP =0; a=0; b=0);
- **E_{it}(P(Upsell))** is the expected value for Upsell Ratio

b) Discounted CF part is defined by the formula

$$CLV_c = \sum_{i=1}^{P} \sum_{t=1}^{T} P(Active)_{it} \times \frac{CF_{it}}{(1+dR)^t} \qquad (21)$$

Where

- **P** is the number of Customer policies;
- **T** is the time frame for CF accumulation;
- **c** is the Customer index;

c) Probability of being active in period t

$$P(Active)_{it} = P(Active)_{i(t-1)} \times (1 - cR_i) \qquad (22)$$

Where

- cR_i is the churn rate for policy i;
- For **t**=0 => P(Active)=1

d) The Expected value for Upsell Ratio

$$E_{it}\left(P\left(Upsell\right)\right) = P\left(Upsell\right) \times uR_i \times cR_i \times n \qquad (23)$$

Where

- uR_i is the aggregate Upsell ratio
- cR_i is the churn rate for policy i;
- n is the number of expected Upsell events during the lifetime
- $P(Upsell)$ is the individual Upsell probability given by a DM model

Since Upsell may not be a statistically independent event, thus having equal probabilities of occurring in each period **t,** the **cR_i x n** term is introduced to normalize the probability P(Upsell), which carries the individualization characteristic, to 1 event in the average policy lifetime and adding a new parameter **n** which will be calibrated using any means of expert guidance. Having in consideration formulas (11) and (13), this done by dividing by the expected Upsell ratio by the expected policy lifetime which is equivalent to multiply by the expected churn rate. During this phase of the project **n** will be 1.

With no further do, this ends the discussion on the aggregate parts of the model. The P(Active) and E(P(Upsell)) are individual elements and will be addressed in the next section where they're treated by Datamining Modelling techniques and General model Application.

4.14. Data Mining Modelling

For the CLV model individual components Datamining techniques are used to estimate, for each policy, the probabilities of churn P(Active) and the expected value for Upsell E(P(Upsell)) . Supervised machine learning algorithms are adequate since the preconditions where met: **a)** The capability of defining adequate response variables; **b)** The existence of satisfactory independent variables; **c)** The existence of positive and negative examples;

4.14.1. General methodology

To accomplish the business objectives the CRISP-DM approach to datamining modelling is proposed since it covers all needed tasks for building worth solutions.

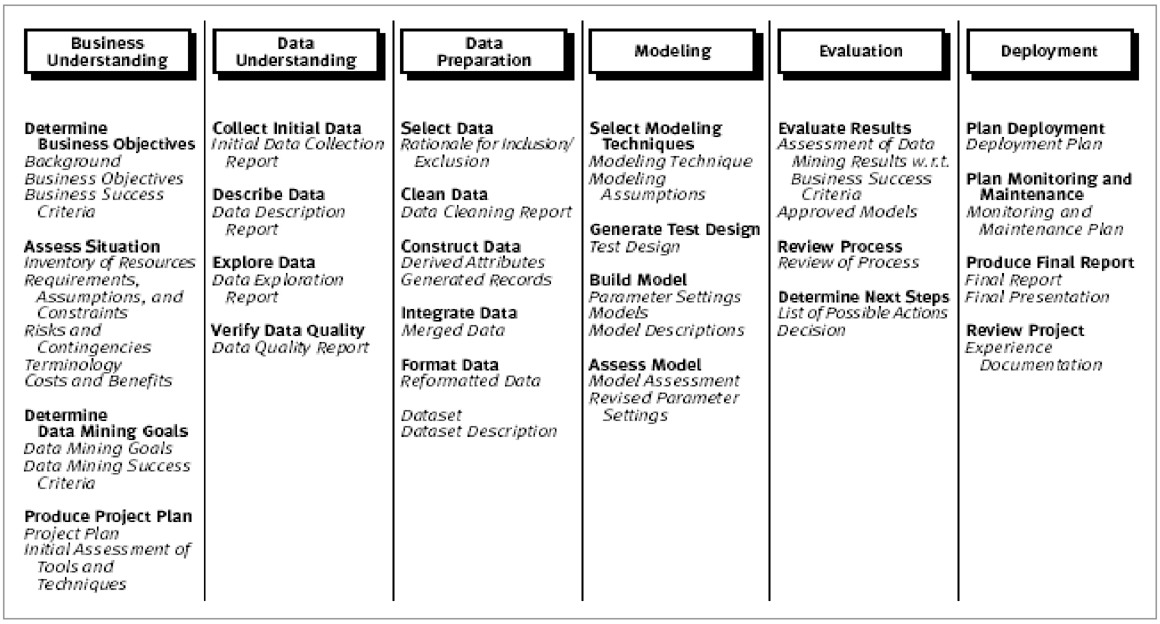


Figure 4.10. CRIP-DM Methodology

Source: CRISP-DM 1.0

For a more detailed version of this process see annexe **A12.**

4.14.2. Model Design

The business understanding, Data understanding and Data Preparation phases of the CRISP-DM process are summed in the following construct: Two time slots where used as reference, since the objective is not only to create datamining models for a specific period but also measure the stability of the models across time.

In first time slot, the Modelling phase, two portfolio snapshots are compared allowing for response variables creation and Model Calibration and Assessment.

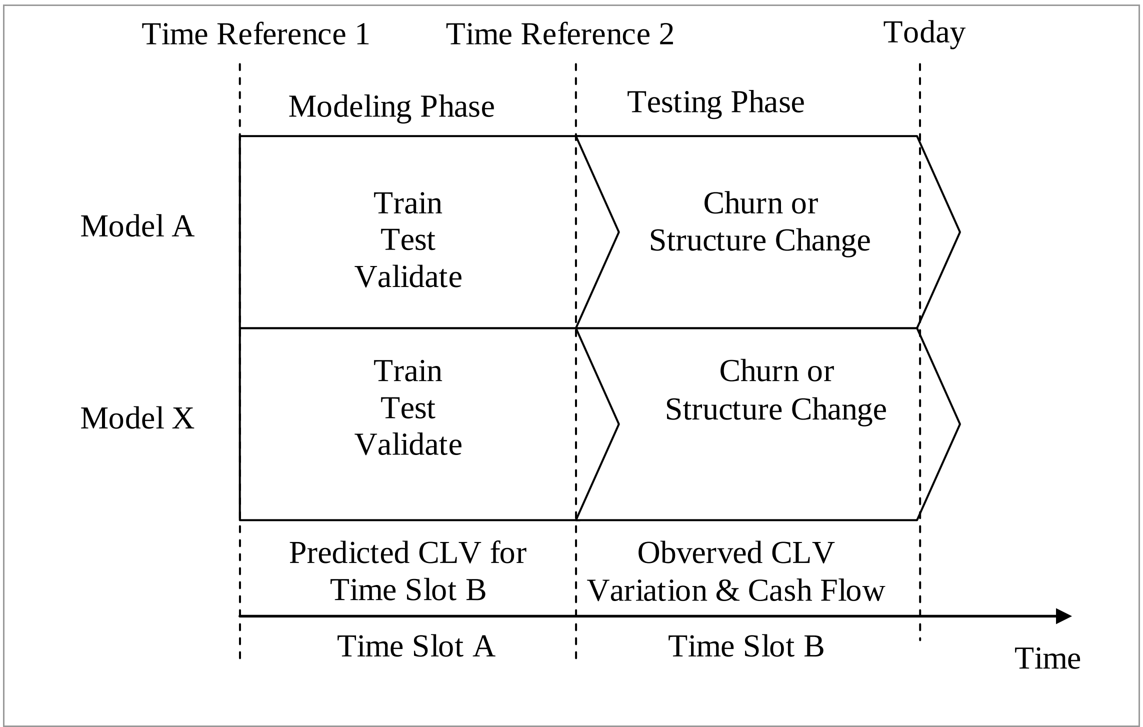


Figure 4.11. Model Design Construct

Source: The author

Time Slot B, or Testing Phase, is used to test the stability of a deployed model by using a different time slot. This procedure is proposed as an extension of the test sample used during the model calibration and assessment in Time Slot A, the difference is that the test data is now shifted in time allowing sensitivity testing for estimating the model expiration time. This is done by looking at the behaviour of the model's classification statistics.

4.14.3. Response Variables Definition

The response variables are defined by comparing, for each time slot, two portfolio snapshots. In this project, for convenience reasons, a 12 month interval was used, set to 31-Dec-XXXX.

Churn: Terminated policies in the second snapshot, when compared with the first one, are flagged with 1 defining a binary response variable. Configuring a data set this way permits the application of the churn ratio definition given by formula (12).

Upsell: Following the same comparison principle, policies that have a increase in premium above the defined threshold are flagged with 1.

4.14.4. Datamining Algorithms

Although several supervised machine learning algorithms are available, **decision trees were preferred** since they allow comprehensible results. The chosen Datamining tool offers an implementation of a relatively new decision-tree-based classification algorithm called SPRINT proposed by Shafer, J., Agrawal, R., & Mehta, M. (1996), which strong point is a highly efficient and parallelized design. The parameters available in the chosen too are set as following:

- Quality Measure: Information Gain Ratio;
- Over fitting Control: Pre-pruning by limiting the minimum number of records per node;
- Binary Nominal Splits: Forced

The independent variables and the minimum number records per node parameters were adjusted manually for each model configuration and for each studied product.

4.14.5. Datamining Model Workflow

A standard dataflow, depicted below, was defined and applied to all products for modelling churn.

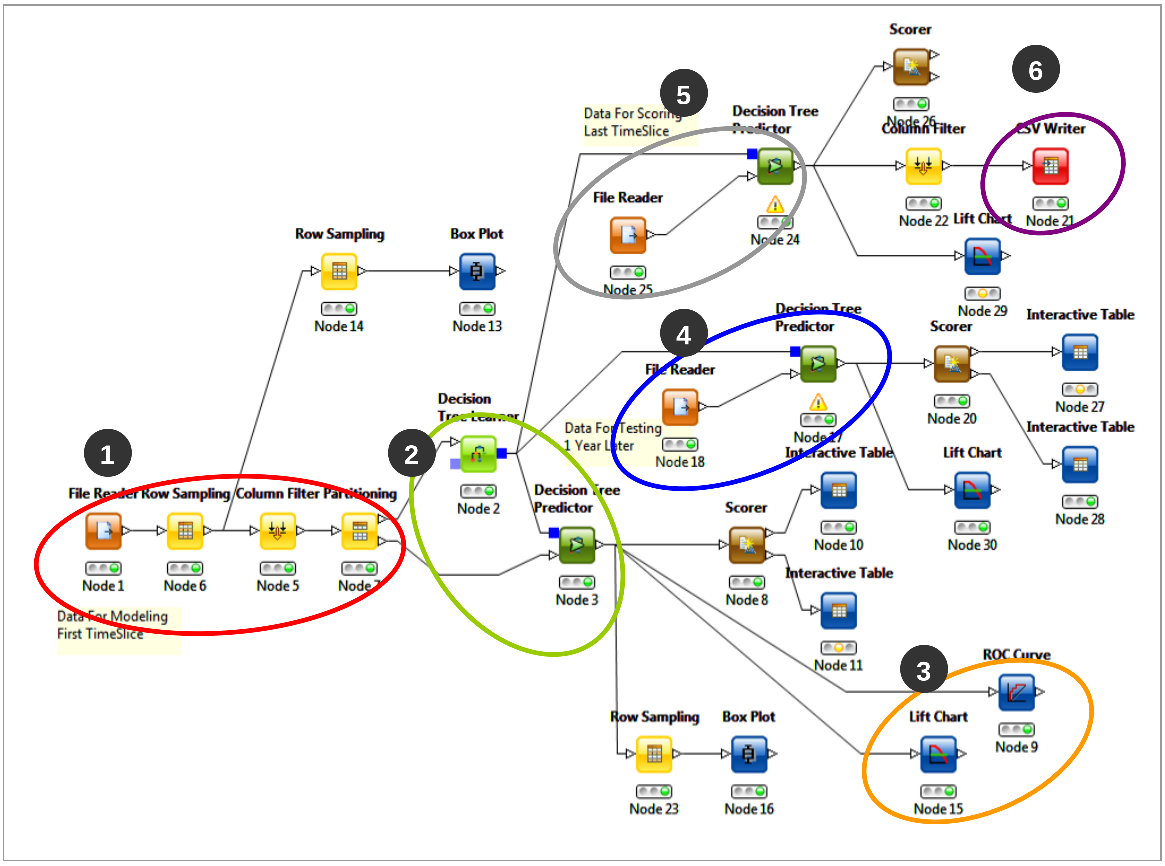


Figure 4.12. Standard DataMining Work Flow for the Project

Source: The author

1 – Input and Sampling Section

2 – Model Training Section => Modelling Phase

3 – Model Quality Statistics

4 – Model Quality Statistics Year + 1 = > Testing Phase

5 – Model Application for Reference Period

6 – Scored Results for integration on CLV Framework

4.14.6. Datamining Tool

KNIME or Konstanz Information Miner was used to develop this part of the project. Essentially is a workbench to define the flow of operations on data, typically as part of doing analytic work like joining, normalizing, cleaning and visualizing/building predictive models in a similar way to SPSS Clementine, SAS Enterprise Miner or FICO's Model Builder. KNIME is an open source platform. KNIME have pre-integrated R and WEKA, though the tool also includes core algorithms natively such as Support Vector Machines, Neural Networks and Decision Trees among others. In annexe **A7** an output example for decision tree is available.

During the project several tools were tried but were excluded due to several reasons, below there are some examples:

- SAS – Powerful but too Expensive;
- IBM/SPSS Clementine – Powerful but too Expensive;
- Statistica Miner – Step learning Curve to be used in this project;
- Rapidminer – Same as above, the Decision Tree algorithm is not very convenient for usage;
- Orange – Free with Interesting functionality but buggy and cannot deal with much data;
- Tanagra – Same as above;

In annexe A14 there are available some snapshots of early modelling experiments just for auto Insurance.

4.14.7. Datamining Model Results

Although the datamining models created were set for a binary classification problem, we're not interested in the final model decision (e.g. 1 or 0) but with the normalized class distribution. This information is obtained in each terminal tree node or leaf. This quantity is equivalent to the proportion of churn in each leaf. Having in mind that P(Active) = 1 - P(churn) , each policy classified in these leafs share the same probability of being active after a time lapse equal to time slot duration defined in the model design. The following table and graph summarizes the distribution of leaf churn probability for each product.

Table 4.8. Datamining Model Results for Churn

			Churn Levels Statistics			
Product	Nº of Levels	Policies	Average	Min	Max	StDev
Acidentes Pessoais	59	51,158	14.1%	6.3%	77.3%	9.0%
Acidentes Trabalho - CO	47	7,710	25.4%	6.5%	65.2%	9.0%
Acidentes Trabalho - CP	58	7,651	19.2%	3.5%	77.5%	10.7%
Automóvel	195	118,428	16.0%	0.0%	60.5%	6.4%
Comércio & Serviços	27	13,867	14.4%	2.4%	84.2%	6.6%
Habitação	80	66,925	7.1%	0.0%	58.8%	3.6%
Responsabilidade Civil	32	37,141	11.8%	0.2%	62.4%	4.8%

As was expected the probability classes relative frequency peak coincides with the average aggregate value for churn, as seen in appendix **A8**. One characteristic of the distributions is the long and shallow tail into the upper limit of P(churn) =1, this reveals some degree of model overfitting. This happens due to the lack of dimensionality of the reference data set and also the lack of finer controls provided by the datamining tool for tweaking the model training.

The usual model accuracy statistics, confusion matrix, sensitivity /specificity/recall must be seen only as instrumental since they give a quantitative and useful insight on model performance but their importance is lessen since what is valued in this problem is the information creation around the response variables, thus producing the desired individualization aspect needed for the selected CLV approach. For further details on the Model accuracy statistics see annexes **A10** and **A11**.

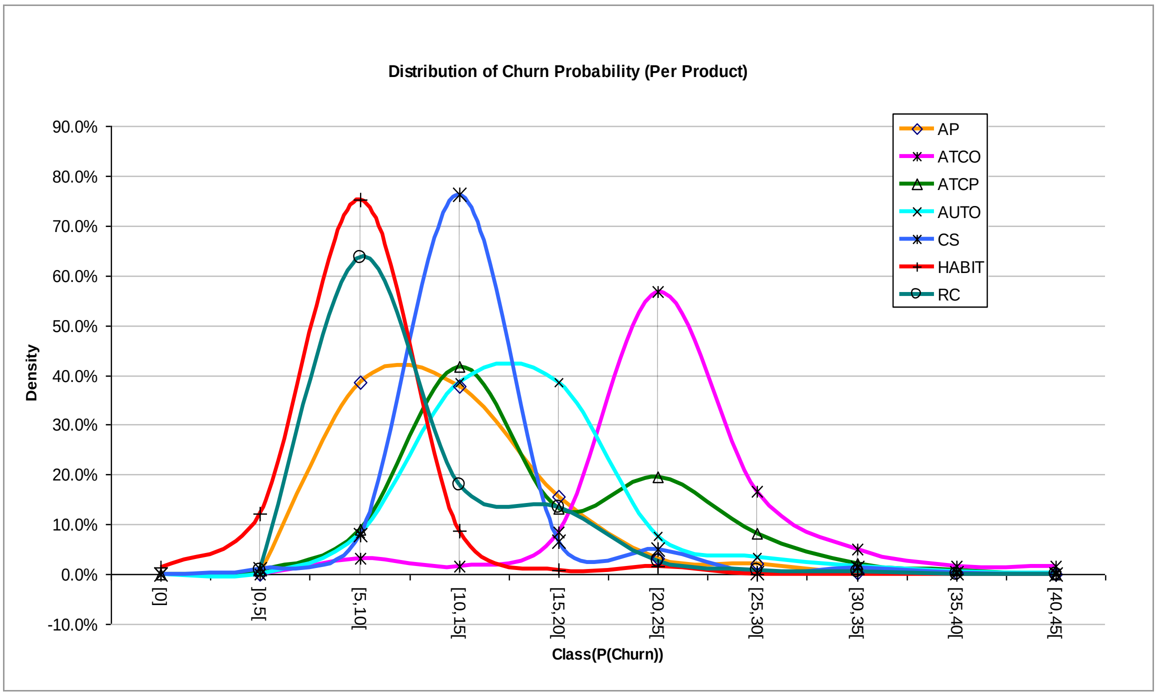


Figure 4.13. Distribution of Churn probability classes for each product

Source: The author

With this approach it was possible to increase significantly the probability variation for the target event (e.g. churn, Upsell, etc). For example, for the product Acidentes de Trabalho – CP (ATCP), we now have 58 different probability classes ranging from 0.03 to 0.775 instead of only the aggregate one of 0.19. The number of probability classes is limited by the dimensionality of the input data set and also by the desired generalization power of the datamining model.

Although it could be possible to avoid any over-fitting control during the training phase, therefore maximizing the number of probability classes available, the model would be a bad predictor for new business or unstable when rehashed in eventual CLV calculation updates further in time.

The following Chapter will demonstrate the impact on portfolio valuation using the assembled CLV model.

4.15. Conclusions on expansion into insurance industry

In this chapter some of the specifics of insurance business where studied and included on the CLV calculation:

- Discount Rate for the Insurance Company;
- Structural Cash Flow Variations (Internal/External) typical on insurance
  - Indexation / devaluation;
  - BMS;
- Measuring Retention Rate on policies;
- Technical Balance Sheet information integration
- Cash-In & Cash-Out

The strategy for modelling was also determined and a mixed approach was the easiest and richest and therefore was selected. **Equation (1)** was expanded including the intricacies mentioned above. This new resulting equation has terms that can be linked to the individual or customer parameters. For those Individual terms datamining methodologies where tested and used for predicting churn and Upsell probabilities.

With this mixed strategy it was possible achieve a great degree of individualization with a ratio of 2.71 meaning that for each CLV Level, in average, there are 2.71 Customers.

5. MODEL APPLICATION AND RESULTS

In this section we will analyse the results of applying the CLV model **defined by equations (20) to (23).** The model was applied to the portfolio's 2010 December snapshot, involving **302'880** policies from **177'448** distinct Customers. It is expected to observe a degree of correlation between aggregate premiums per customer and the corresponding CLV. The graph bellow depicts the structure of correlations found on the customer database.

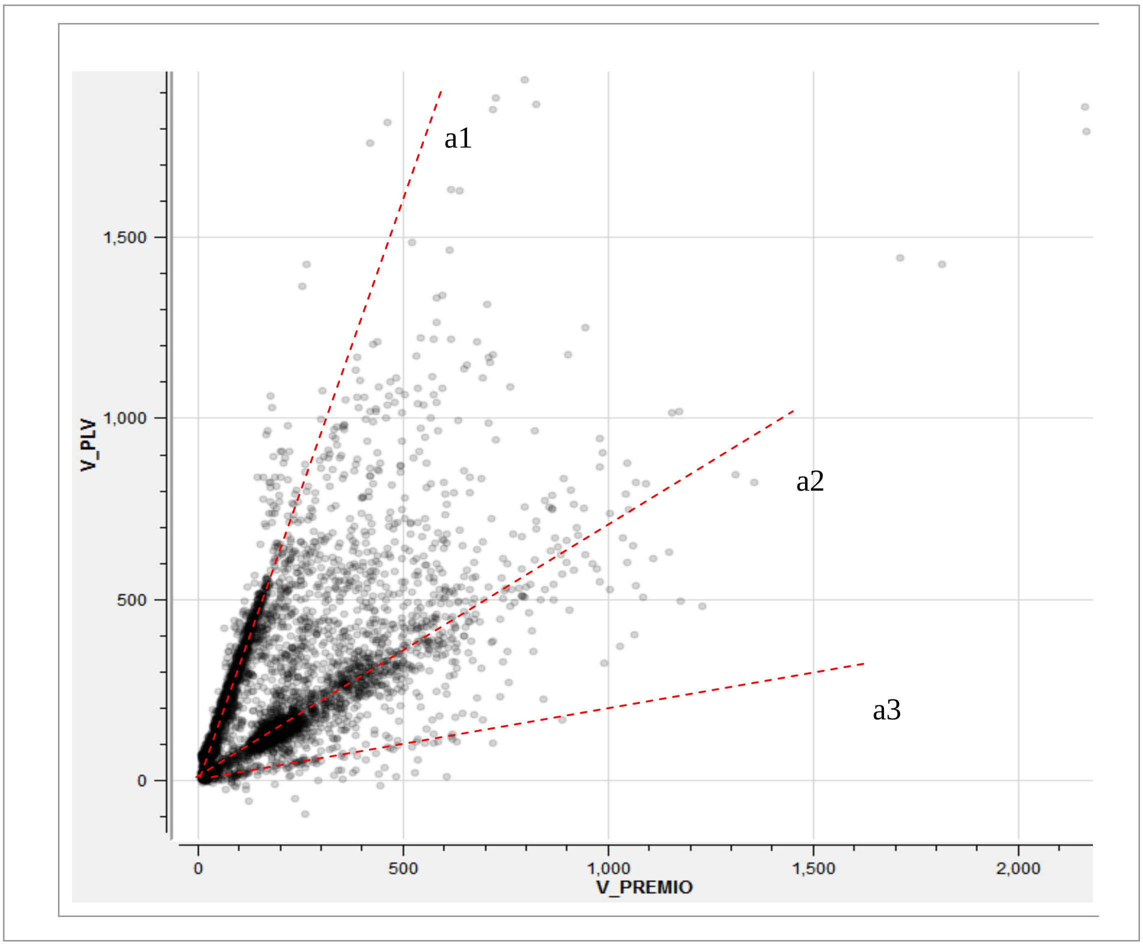


Figure 5.1. Scatter Plot – Premium vs. CLV

Source: The author

Three major correlation vectors can be identified corresponding to sets or clusters of customers with typical product ownership. Vector **a1** represents customers with mainly home insurance (Habitação*)*, **a2** is explained by customers mainly with auto insurance (Automóvel) and/or Liability (Responsabilidade Civil) and/or Business Workers Compensation (Acidentes de Trabalho – Conta Outrem), lastly **a3** has behind it most likely commercial multiple risks (Comércio e Serviços). The previous reading it's not the most important, what must be retained from the previous graph it's the richness of variability between the premiums view against the CLV concept. This interpretation is further explored in the next section on portfolio valuation.

5.1.1. Impact on valuation

Using the CLV model, the Company's in force portfolio is valuated in **63,224,350€** against **54,922,578€** given directly by the aggregate sum of annuities expressed in commercial premiums. Although these numbers are based on different concepts, premiums are what the customer pays in the present and CLV incorporates the margin generated during whole lifecycle, it is valid to use any them to evaluate the same portfolio. The difference is that the CLV concept is much more informative, it tells us where value (what is kept for the stakeholders after technical costs) is being generated or destroyed. These two visions are summarized in the table bellow.

Table 5.1. Two Visions on Value

Product	Nº of policies	Comercial Premium	Share of Value	Lifetime Value	Share of Value	Ratio
		A		B		B/A
AP	51,158	3,386,360	6.2%	8,790,408	13.9%	2.60
ATCO	7,710	9,610,741	17.5%	8,983,754	14.2%	0.93
ATCP	7,651	2,059,785	3.8%	-333,763	-0.5%	-0.16
AUTO	118,428	27,388,953	49.9%	19,505,484	30.9%	0.71
CS	13,867	3,114,785	5.7%	619,955	1.0%	0.20
HABIT	66,925	7,482,493	13.6%	24,304,155	38.4%	3.25
RC	37,141	1,879,461	3.4%	1,354,357	2.1%	0.72
Total	**302,880**	**54,922,578**	**100.0%**	**63,224,350**	**100.0%**	**1.15**

The share of value between products is completely different in these two scenarios, the products HABIT (Multirriscos Habitação) and ATCP *(Acidentes de Trabalho Conta Própria*) have the most extreme variations in share of value, the first one climbs from 13.6% to 38.4% and the second one from **3.8%** goes to **-0.5%,** destroying value.

Another interesting reading of this data is the ratio between PLV and commercial premiums. This ratio indicates the value, in margin, generated by each monetary unit in premiums. For example, **each Euro** of premiums in *Multiriscos Habitação* will be generated, in average, **3.25 Euros** of margin during the contract duration. The following figure shows the PLV model behaviour for each product.

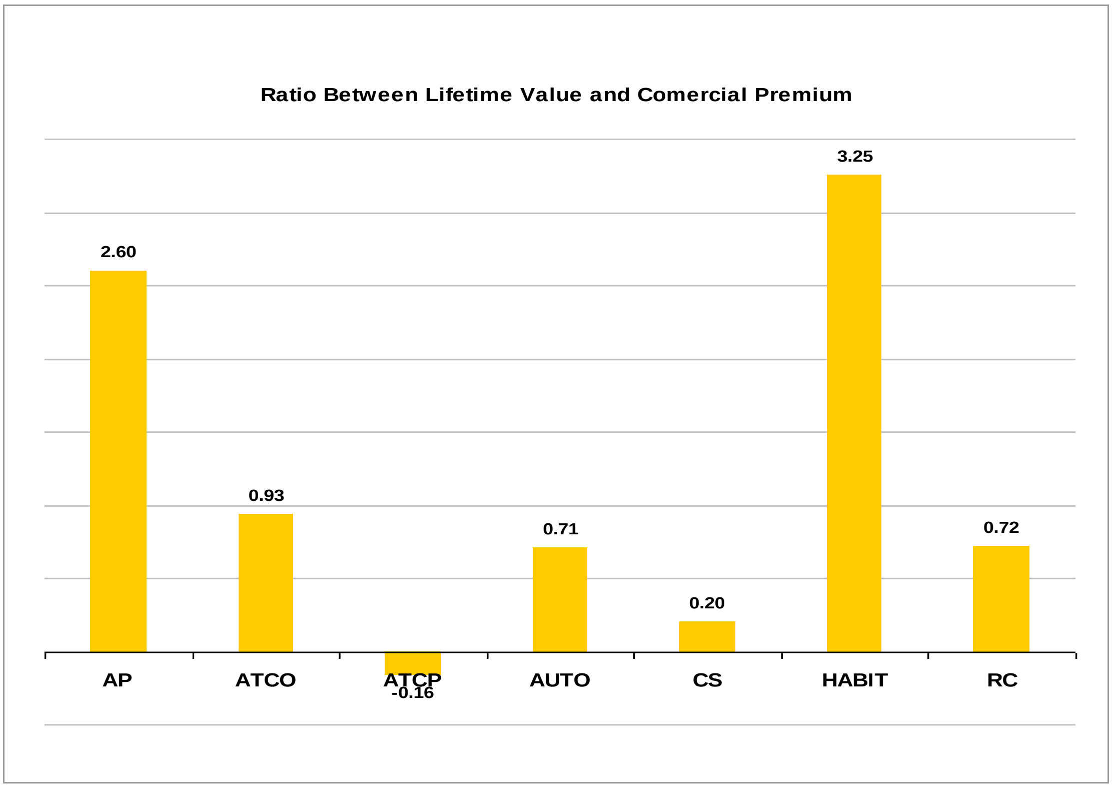


Figure 5.2. Leverage between premiums and PLV

Source: The author

This new view on value is very important for decision making functions such as board of directors deciding about the firm’s marketing and commercial strategies. In this case, in a simplistic scrutiny, if the objective is maximizing results the focus should be putted on home insurance and personal accidents, which is not obvious when looking only to premiums. This capability given by the CLV model will useful in Customer segmentation as noticed in the next section.

5.1.2. Impact on Segmentation

One of the capabilities potentiated by using CLV is the ability to segment the Customer base. Bellow it is possible to observe the cumulative distribution of customers by CLV bin. The x axis is ordered from left to right in descending order of CLV value.

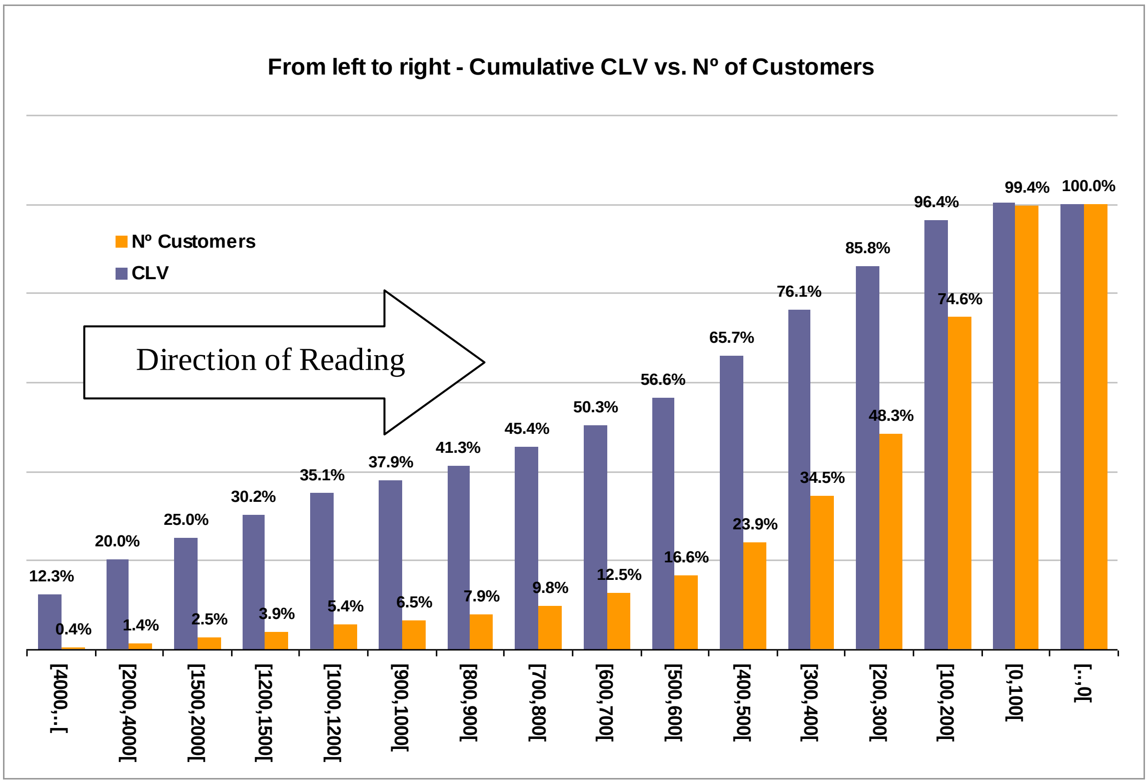


Figure 5.3. Distribution between Customers and Value

Source: The author

As an example of reading the above graph, with only 2.5% of the most valuable customers we get 25% of portfolio aggregate value. This relation reveals a great potential to build and develop strategies around segments based on CLV.

5.1.3. Impact on the Distribution Process

The Distribution Process covers all activities related with sales, including performance assessment, which is being done using the premiums metric. In this section we will analyse the impact on assessment when changing metrics between lifetime value and premiums. The Distribution network has 800 branches organized in 89 regional offices. As usual the assessment process has rakings for branches, offices, regions and sales persons for new business and for in force portfolio. The following statistics shows, at the regional office level, the impact on ranking position when CLV is used on portfolio assessment, 93% of the offices change rank.

Table 5.2. Impact on portfolio value rank when using CLV

Statistic	Value
Rank Position Change Nº	83.00
Average Change	3.60
Max Change	16.00
Min Change	-11.00
StdDev of Change	3.03

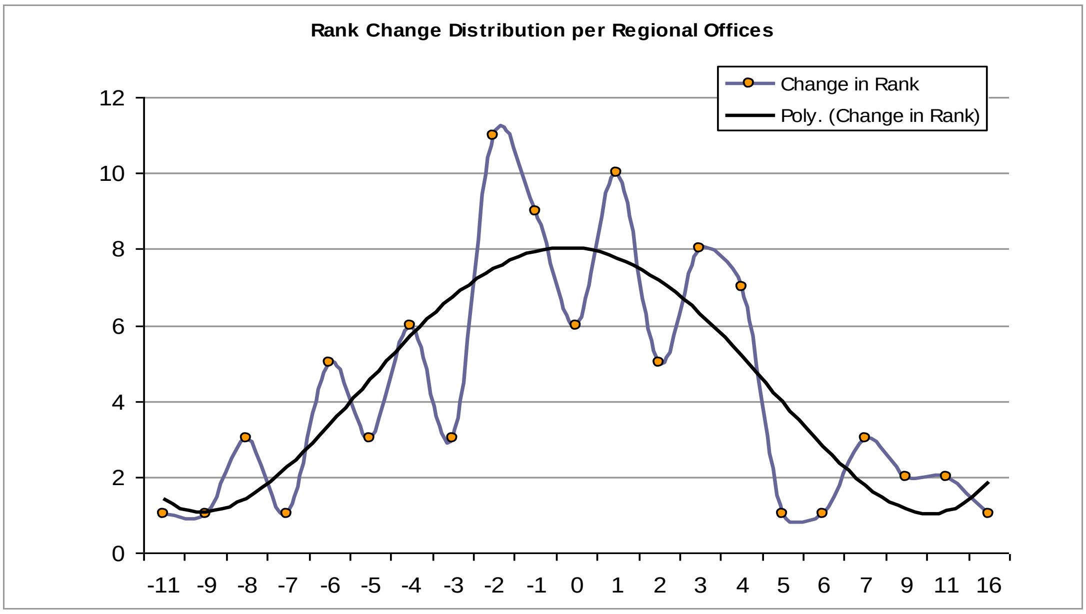


Figure 5.4. Distribution on portfolio value rank when using CLV

Source: The author

5.2. Integration into Business Operations

Since one of the objectives of this research work is not only to provide a technical framework for estimating the CLV in the Insurance industry but also an extended framework for integration into other business processes. In this context three levels of framework design are covered: Strategic, Tactical and Operational;

5.2.1. Strategic Level

At this level the CLV framework interfaces with the strategic development and marketing processes. It is used in deciding the big picture, for example: focus more on one kind of customers than others; design new products; focus the sales force on specific products. In practice, helping in the design of the firm's marketing and business plans. This is done by using segmentation information and what-if analysis.

5.2.2. Tactical Level

At this level the CLV framework will interface with the distribution process on campaign support trough leads selection, customer prioritization and sales statistics based also on the CLV metric.

5.2.3. Operational Level - Usage

At this level the CLV framework will be a support process, it will CLV data to other systems, preferably by SOA busses. One usage example is having CLV information on product underwriting simulations for discount decisions.

5.2.4. Operational Level – Infrastructure (AEM)

To support all these process layers the CLV framework is organized in a set of activities that can be automated by following the workflow depicted below. This workflow defines the AEM or Automatic Estimation Machine. This set of procedures and activities can be iterated in a continuously improving process witch general for is shown below.

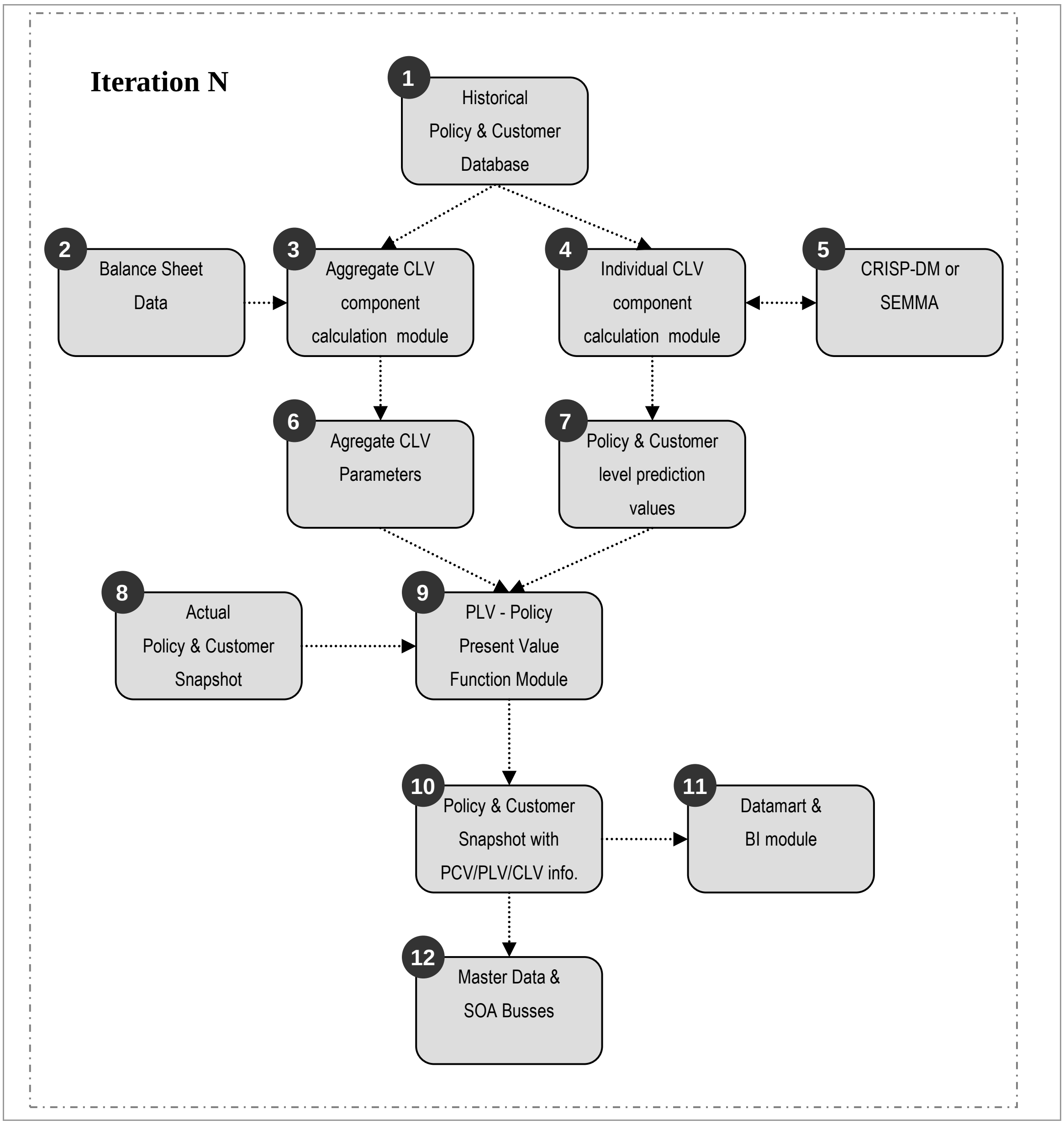


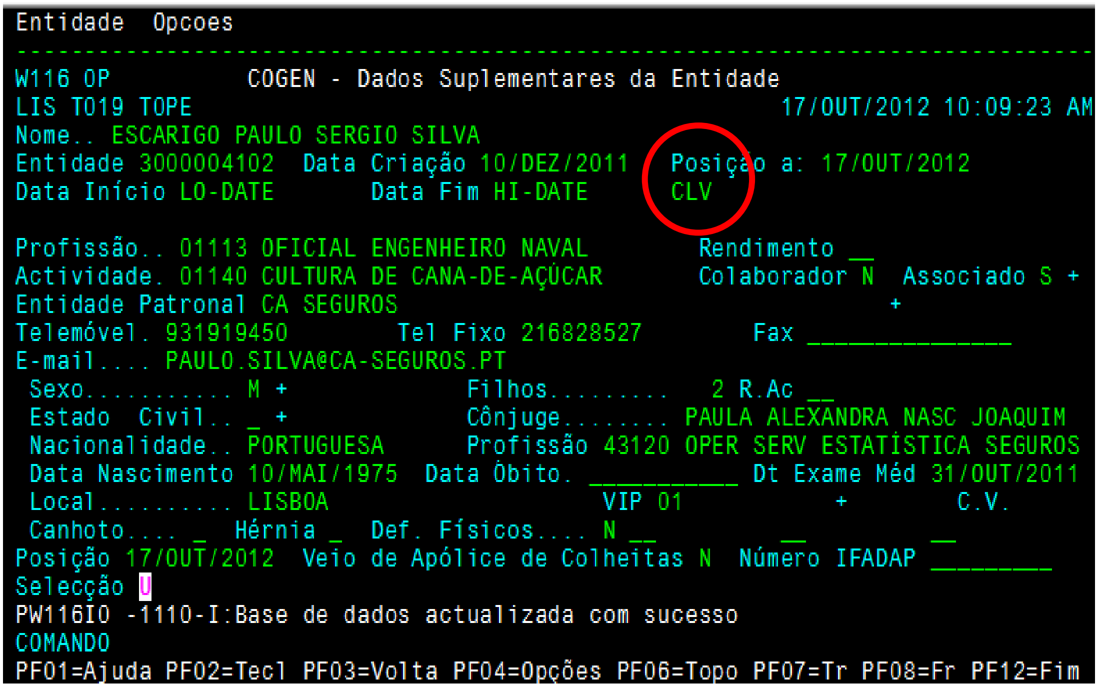
Figure 5.5. Automatic Estimation Agent Architecture

Source: The author

1. **Historical Policy & Customer Database –** Is the Datamart that Automatically Stages out detailed data.
2. **Balance Sheet Data** – Is Automatically Staged out from the Accounting System;
3. **Aggregate CLV Component Calculation Module –** Where aggregate assumptions and calibration are automatically determined;
4. **Individual CLV Component Calculation Module –** Where calibrated Datamining models are applied;
5. **CRISP-DM or SEMMA –** where Datamining models are trained and calibrated. Some human input may be needed;
6. **Aggregate CLV Parameters –** These parameters are automatically loaded to the CLV model;
7. **Policy & Customer level prediction values –** These values are automatically loaded to the CLV model;
8. **Actual Policy & Customer Snapshot –** The desired portfolio snapshot is ready to be loaded to the CLV model;
9. **PLV – Policy Present Value Function Module –** This module calculates the PLV/CLV for each policy and Customer;
10. **Policy & Customer Snapshot with PCV/PLV/CLV info. –** This information is automatically loaded into the Datamart;
11. **Datamart & BI module –** This module provides data for marketing strategies and analysis;
12. **Datamart & SOA Busses –** By using SOA approaches the CLV data can be automatically integrated in other systems;

5.2.5. Master Data vs. Datamart

Although the CLV information could be **solely** treated in the informational systems realm, on a Datamart, its potential is maximized when used in the operational master data as shown in the figure below with simulated data.



```
Entidade  Opcoes
---------------------------------------------------------------------------------
W116 OP          COGEN - Dados Suplementares da Entidade
LIS T019 TOPE                                        17/OUT/2012 10:09:23 AM
Nome.. ESCARIGO PAULO SERGIO SILVA
Entidade 3000004102  Data Criação 10/DEZ/2011   Posição a: 17/OUT/2012
Data Início LO-DATE       Data Fim HI-DATE      CLV

Profissão.. 01113 OFICIAL ENGENHEIRO NAVAL        Rendimento __
Actividade. 01140 CULTURA DE CANA-DE-AÇÚCAR       Colaborador N  Associado S +
Entidade Patronal CA SEGUROS                                     +
Telemóvel. 931919450       Tel Fixo 216828527        Fax _______________
E-mail.... PAULO.SILVA@CA-SEGUROS.PT
 Sexo........... M +            Filhos......... 2 R.Ac __
 Estado  Civil.. _ +            Cônjuge........ PAULA ALEXANDRA NASC JOAQUIM
 Nacionalidade.. PORTUGUESA     Profissão 43120 OPER SERV ESTATÍSTICA SEGUROS
 Data Nascimento 10/MAI/1975  Data Óbito. ___________ Dt Exame Méd 31/OUT/2011
 Local.......... LISBOA                   VIP 01        +         C.V.
 Canhoto.... _  Hérnia _  Def. Físicos.... N __          __         __
Posição 17/OUT/2012  Veio de Apólice de Colheitas N  Número IFADAP _________
Selecção U
PW116I0 -1110-I:Base de dados actualizada com sucesso
COMANDO
PF01=Ajuda PF02=Tecl PF03=Volta PF04=Opções PF06=Topo PF07=Tr PF08=Fr PF12=Fim
```

Figure 5.6. CLV on Entity Sub-System on Company's Operational System

Source: The author

Depending on the Company's strategy on Services, CLV based information may be present on daily operations. In this scenario the master data is more appropriate for Information propagation; if SOA Busses are being built around the main Operational System other systems will easily interface with it. This way is simpler to propagate the CLV based information for usage on Simulators, Web Portals, Claim Management System, Contact Centres, Billing, etc.

This also isolates the informational system from operations.

5.2.6. Communication and Change Management Issues

One of the issues found during the discussions on CLV usage is how to communicate the concept with other departments. Since people are used to link the value of something by its facial value it is difficult to change their thinking and decision making based on a different valuation method. Phrases like the following one are typical:

> You're asking me to tell my "troops" that their going to be judged not by what they sell but by that number called CLV, I understand the idea but its not clear to me and for them, how you do your calculations its like saying that 5 = 2. That is unnatural twist, this not going to work.

Some workarounds may be arranged to ease the concept introduction.

5.2.1. Naïve Model

A simplified Naïve model can be used to bias the sales and portfolio valuation, this model should obey to the following principles:

- Work on the positive side, by bonus;
- Be simple to understand and communicate;
- Keep a correlation with the CLV leverage as shown in previous chapters;

As an example for the products studied in this work we can build this conversion table between premium and adjusted premium.

Table 5.3. Surrogate Multipliers for CLV

Product	CLV Leverage	Multipliers
AP	2.60	X 3
ATCO	0.93	X 1
ATCP	-0.16	X 1
AUTO	0.71	X 1
CS	0.20	X 1
HABIT	3.25	X 4
RC	0.72	X 1

5.2.2. Quantile Ranking

Other form to make operational the CLV is not giving a meaning to its financial value but use it for ranking customers, using their absolute rank position to build segments with different service approaches.

The CLV can be ranked in quantiles, with variable or fixed resolution. In the example below a customer set is ranked by CLV decile, the first decile is further refined with percentile resolution. This way anyone using this rank can locate the customer concerning its relative embedded value, for example a call centre operator can easily adjust the service profile according, if he or she is talking with someone belonging to the first percentile (the most valuable customers) or one not so valuable belonging to percentile 60.

Table 5.4. Quantile Ranking of CLV

Customer Type	Quantile Resolution	Quantile	Nº of Costumers	CLV (€)	W-Weight	W - Accumulated
Corporate	Percentile	1	160	10,486,962	30%	30%
		2	160	2,643,460	7%	37%
		3	160	1,857,715	5%	42%
		4	160	1,449,118	4%	47%
		5	160	1,176,598	3%	50%
		6	160	996,886	3%	53%
		7	160	888,866	3%	55%
		8	160	788,054	2%	58%
		9	160	714,799	2%	60%
		10	160	655,267	2%	61%
	Decile	20	1,600	4,685,745	13%	75%
		30	1,600	2,867,090	8%	83%
		40	1,600	1,949,092	6%	88%
		50	1,600	1,388,613	4%	92%
		60	1,600	991,291	3%	95%
		70	1,600	736,389	2%	97%
		80	1,600	498,725	1%	99%
		90	1,600	339,716	1%	100%
		100	1,600	155,171	0%	100%
Total			**16,000**	**35,269,556**	**100%**	

5.2.3. Service and Process matching

Other challenge is the identification of different regimes where and how CLV may be used. In some scenarios the actual CLV value is useful for a pure customer financial valuation, in other cases where the customer services are in stake a pure CLV usage may be impractical. The Organization should design a case by case setting list where the most adequate usage is explained. This almost encompasses the definition of a CLV usage policy. For example, the following table exemplifies

Table 5.5. Case Setting Policy

Process Context	Scenario	Usage
Marketing	Customer Lead identification	Selection by Quantile Ranking with in conjunction with other desired or mandatory criteria.
	Communication with the customer	Information is Private To the Company. Customers aren't informed of their CLV status or its impact on the service levels they're subjected to.
Underwriting	Portfolio Adjusting	…
	New Business Analysis	…
		…

5.3. Conclusions on Model Results and Business Impact

After the model application the results were in line with what was expected, in other words, the pretended goals were achieved.

Firstly the model generated a large degree of Customer individualization that was achieved by the usage of the individual components on the mixed approach methodology.

Secondly, the model offered a different view on the portfolio value. The share of value between products and inherently between Customers is changed. Some products seen as valuable now are value destructors or less important on that matter. In this line of reasoning, one new metric can be defined after the model application. This is the **CLV leverage** witch is the ratio between Premiums and CLV.

This new method of business valuation also brings new challenges for its usage, for example: Customer segments may be arranged in different manners, the usual sales rakings may be completely different. If the objective is really operating the business with this metric it will inevitably generate communication and change management issues. These challenges may be mitigated by using simplified models or other similar strategies. All the involved stakeholders must agree in some degree with this new approach, nevertheless a strong leadership is needed for helping to enforce the change and the definition of a usage policy may be very advisable.

For an effective CLV program implementation, the systems and supporting processes are vital. One of the objectives of this work is also providing a framework for the continuous support and tuning of the CLV parameters.

Although what is provided in this chapter is very general it reflects the real workflow followed by one Organization. In this realm it should be stressed that the CLV, to be used in whole gamut of processes, should be also part of the master data for ease of data communication and automation.

This assumption depends on he Organization’s systems maturity and especially on its Marketing and Services Strategy.

6. RESEARCH CONCLUSIONS

Organizations, especially financial institutions in the current economic environment, are facing some challenges regarding how to **keep profitable and sustainable**. The interest in using metrics as Customer Lifetime Value for managing Customer Lifecycle can be a powerful approach to meet those challenges. One of the key points of this approach is the ability of influence the decision making process with a deeper insight on the long term impact on margin and value creation.

This research work appears as challenge in the context of a real business problem: How to apply the CLV methodology on an insurance Company.

Although there is plenty of literature about the general concept and sometimes about some unfathomable and complex mathematical methodologies, not much was found (by the researcher) regarding the practical aspects on implementing a CLV framework in the Insurance Industry.

The first part of the work was focused on learning the CLV concept and formulation. Some classical valuation methodologies where investigated like Recency Frequency Methods or Past Customer Value, these where compared with the more modern CLV methodologies, where specific information can be used to predict the probability of customer being active in the future while generating revenue. This probabilistic view, looking into the future, is very important since it gives us a prospective measure of the potential Customer value, nothing can be done about the past but you can make better decisions about the future.

Other important dimension learnt during the literature review is the difference between individual (customer centric) and aggregate (domain averaged) CLV calculation methodologies. The most interesting are the mixed approaches where, depending on the level of data granularity, the individual and aggregate approaches are combined in a single model. Mixed approaches maximize the firm’s information usage while maintaining the ability of giving a different, but fact based value, for each customer. These mixed approaches were used for the framework design and project development.

The second part of the work is centred on the Company's framework and project development. The first challenge was translating the CLV parameters usage for insurance. Three main parameters were identified: discount rate; retention rate and cash flow variations. For each one there are specificities facilitating or not their usage. For example, due to the contractual nature of the business, it is easy to define the churn event. In the case of the discount rate, we're in the presence of more specific problem, since the solution it's determined by the firm's financing strategy and structure: Does the firm needs to be financed through any kind of market asset or not, are there any shareholders, stakeholder with expectations on their capital investments. Since the cost of capital can be used to set the discount rate the most appropriate CoC model must be used, WACC and CAPM where studied. For this project the CAPM was used since it uses external information and the WACC is not suitable for the researched firm financing scheme.

More complex was the estimation of the Cash Flows streams. Premiums are affected by a series of internal and external, non-random, factors that should be accounted for if we want to build more solid CLV model. For example this project includes in its modelling scheme the BMS (bonus malus system) for auto insurance. This aposteriori ratemaking model does have a profound impact on cash flows, reducing the premium in 50% in a predictable way, depending on each policy lifecycle state. Cash Flows are defined by Cash in and Cash out parameters obtained from balance sheet data, a method for using this information was presented, this method combines claim, taxes and distribution costs plus reinsurance balance.

To use the mixed CLV approach, datamining models were developed for estimating churn and upselling events, the estimated values were then combined with the previously mentioned balance sheet data. With all these tools every policy was valuated using a specifically design present value function that implements the CLV model according to the researched firm specification. Two views on value where compared, the classical one based on premiums and the CLV one, based on potential margin. The impact was analysed on aggregate product and on the distribution process. Using CLV valuation the panorama changes significantly, some products like Personal

Accidents or Home Insurance become the most important ones regarding value. The home insurance product achieves a ration superior by factor of 3.25 when considering ebbed value versus premium. In the distribution process there is a very significant impact almost all branches changes their posing on the ranking. This is a motive of concern since it opens a new challenge for the internal marketing for adoption of this this framework especially to the commercial division of the Insurance Company not to mention the Bank.

Finally a framework for a set of processes (analytical, IT and business) was presented in the form of an AEM (Automatic Estimation Machine), the objective is to gradually automate the whole framework in the most efficient way.

Independently of these new challenges opened by this project it is expected that this research work will be useful for managers and technical personnel since it provides an empirical experimentation on the applicability of CLV.

With this in mind we can say that the **research questions presented in the introduction of this work were answered.**

During the initial meetings with the managing directors, where this CLV subject was being discussed, it was recognized that this approach, if properly developed and exploited, it is expected to bring a positive impact in the following business processes or activities:

a) Strategic Management
b) Customer Segmentation
c) Product Design and Pricing
d) Commercial Incentive Programmes
e) Underwriting
f) Results Distribution

At the time of closing this document, the implementation and development of the CLV Project within the researched Company continues, it is scheduled to be finished and operational during 2012/2013.

7. FURTHER INVESTIGATION

This section designates some questions that weren't answered or developed during this research project and might or should be addressed in the future.

7.1. Data Setting and CLV Model

- In this project, the most granular data is the policy, but since some events occur in only in specific policy covers, investigation concerning the adaptation of the data model to calculate PLV at coverage level would improve the model precision;
- In the case of Auto Insurance, the BMS mechanism dominates the CF variations, there is space for model improvement by somehow including non-structural CF variations for this product;
- This project only included partial balance sheet information. It uses a simplified technical margin model. More balance sheet information can be used, approximating the CLV model to a CE model. But since that information is only provided in large aggregates some clever imputation method is needed.

7.2. Datamining Models

- In the individual approach model parts based on datamining models are estimating only event occurrence probabilities. These models can be extended to estimate also the magnitude (in Euros) of those events, for example E(Upsell) = P(Upsell) x Mag(Upsell) vs. E(Upsell) = P(Upsell) x AggregateUpSellValue;

- The individual approaches can be extended by incorporating models for events like Cross Sell. The current project only addressed the Upsell and Churn events.
- How to extend de Customer Equity concept beyond the current portfolio? Since the Insurance Company is operating in a Bancassurance environment, many Bank customers may not have any kind of insurance products, discovering patterns trough Datamining Models for new customers and its potential value could add a new dimension to a CLV programme.

7.3. Business Processes

- The effectiveness of the framework must be assessed by its usage. This is key especially when developing marketing strategies or on the distribution process assessment and objectives setting. Does the sales performance reflects the new strategies designed from a CLV framework? How the portfolio embedded value is changing?
- The **change management is also an issue deserving its own research line**. The best models and techniques have no value if the organization cannot make them work as valuable tools for the people behind the business processes. Analysing the factors behind a successful CLV implementation from the business and people viewpoint is so important as the mathematical ideas behind it.

REFERENCES

Abele, Karin P. N. (2009*).* ‘Customer equity: dimensions and realisation process’, De Montfort University Open Research Arch, Leicester

Benoit, D. F., & Van den Poel, D. (2009). Benefits of quantile regression for the analysis of customer lifetime value in a contractual setting: An application in financial services. *Expert Systems with Applications, 36(7),* 10475-10484. doi: 10.1016/j.eswa.2009.01.031

Berry, Michael J. A. & Gordon Linoff (1997). Data Mining Techniques For Marketing, Sales, and Customer Support. Ed. Robert Elliott. N.p.: Jonh Wiley & Sons, Inc, 1997.

Chan, S. L., Ip, W. H., & Cho, V. (2010). A model for predicting customer value from perspectives of product attractiveness and marketing strategy. *Expert Systems with Applications,* 37(2), 1207-1215. doi: 10.1016/j.eswa.2009.06.030

De Oliveira Lima, E. (2009). Domain knowledge integration in data mining for churn and customer lifetime value modelling: new approaches and applications. Retrieved from http://eprints.soton.ac.uk/65692/

Donkers, B., Verhoef, P. C., & de Jong, M. G. (2007). Modeling CLV: A test of competing models in the insurance industry. [Article]. *Qme-Quantitative Marketing and Economics, 5*(2), 163-190. doi: 10.1007/s11129-006-9016-y

Fader, P. S., Hardie, B. G. S., & Jerath, K. (2007). Estimating CLV using aggregated data: The Tuscan lifestyles case revisited. *Journal of Interactive Marketing, 21*(3), 55-71. doi: 10.1002/dir.20085

Fernandez, P., Aguirreamalloa, J., & Corre, L. (2011). Market Risk Premium used in 56 Countries in 2011: A survey with 6014 answers. Working Paper. CIIF: IESE - Business School University of Navarra.

Glady, N., Baesens, B., & Croux, C. (2009a). A modified Pareto/NBD approach for predicting customer lifetime value. *Expert Systems with Applications, 36*(2), 2062-2071. doi: 10.1016/j.eswa.2007.12.049

Glady, N., Baesens, B., & Croux, C. (2009b). Modeling churn using customer lifetime value. *European Journal of Operational Research, 197*(1), 402-411. doi: 10.1016/j.ejor.2008.06.027

Gneiser, M. S. (2010). Value-Based CRM The Interaction of the Triad of Marketing, Financial Management, and IT. *Business & Information Systems Engineering, 2*(2), 95-103. doi: 10.1007/s12599-010-0095-7

Gupta, S. (2009). Customer-Based Valuation. *Journal of Interactive Marketing, 23*(2), 169-178. doi: 10.1016/j.intmar.2009.02.006

Gupta, S., Hanssens, D., Kahn, W., Kumar, V., & Lin, N. (2006, Nov.). Modeling Customer Lifetime Value. Journal of Service Research, Volume 9(No. 2), 139-155. Retrieved n.d., from (DOI: 10.1177/1094670506293810).

Homburg, C., Droll, M., & Totzek, D. (2008). Customer prioritization: Does it pay off, and how should it be implemented? *Journal of Marketing, 72*(5), 110-130.

Kumar, V., & George, M. (2007). Measuring and maximizing customer equity: a critical analysis. *Journal of the Academy of Marketing Science, 35*(2), 157-171. doi: 10.1007/s11747-007-0028-2

Kumar, V., Venkatesan, R., Bohling, T., & Beckmann, D. (2008). Practice prize report - The power of CLV: Managing customer lifetime value at IBM. *Marketing Science, 27*(4), 585-599. doi: 10.1287/mksc.1070.0319

Kumar, V. (n.d.). Customer Lifetime Value – The Path to Profitability, Foundations and Trends in Marketing. (Vol. 2). New York: John Wiley & Sons..

Lariviere, B., & Van den Poel, D. (2005). Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Systems with Applications, 29*(2), 472-484. doi: 10.1016/j.eswa.2005.04.043

Rust, R. T., Lemon, K. N., & Zeithaml, V. A. (2004). Return on marketing: Using customer equity to focus marketing strategy. *Journal of Marketing, 68*(1), 109-127.

Shafer, J., Agrawal, R., & Mehta, M. (1996). SPRINT: A scalable parallel classifier for data mining. *Proc. of the 22nd Int'l Conference on Very Large Databases*. (p. 544-555). Morgan Kaufmann.

Tsai, C. F., & Lu, Y. H. (2009). Customer churn prediction by hybrid neural networks. Expert Systems with Applications, 36(10), 12547-12553. doi: 10.1016/j.eswa.2009.05.032

Wang, H. F., & Hong, W. K. (2006). Managing customer profitability in a competitive market by continuous data mining. Industrial Marketing Management, 35(6), 715-723. doi: 10.1016/j.indmarman.2005.06.005

Wiesel, T., Skiera, B., & Villanueva, J. (2008). Customer equity: An integral part of financial reporting. Journal of Marketing, 72(2), 1-14.

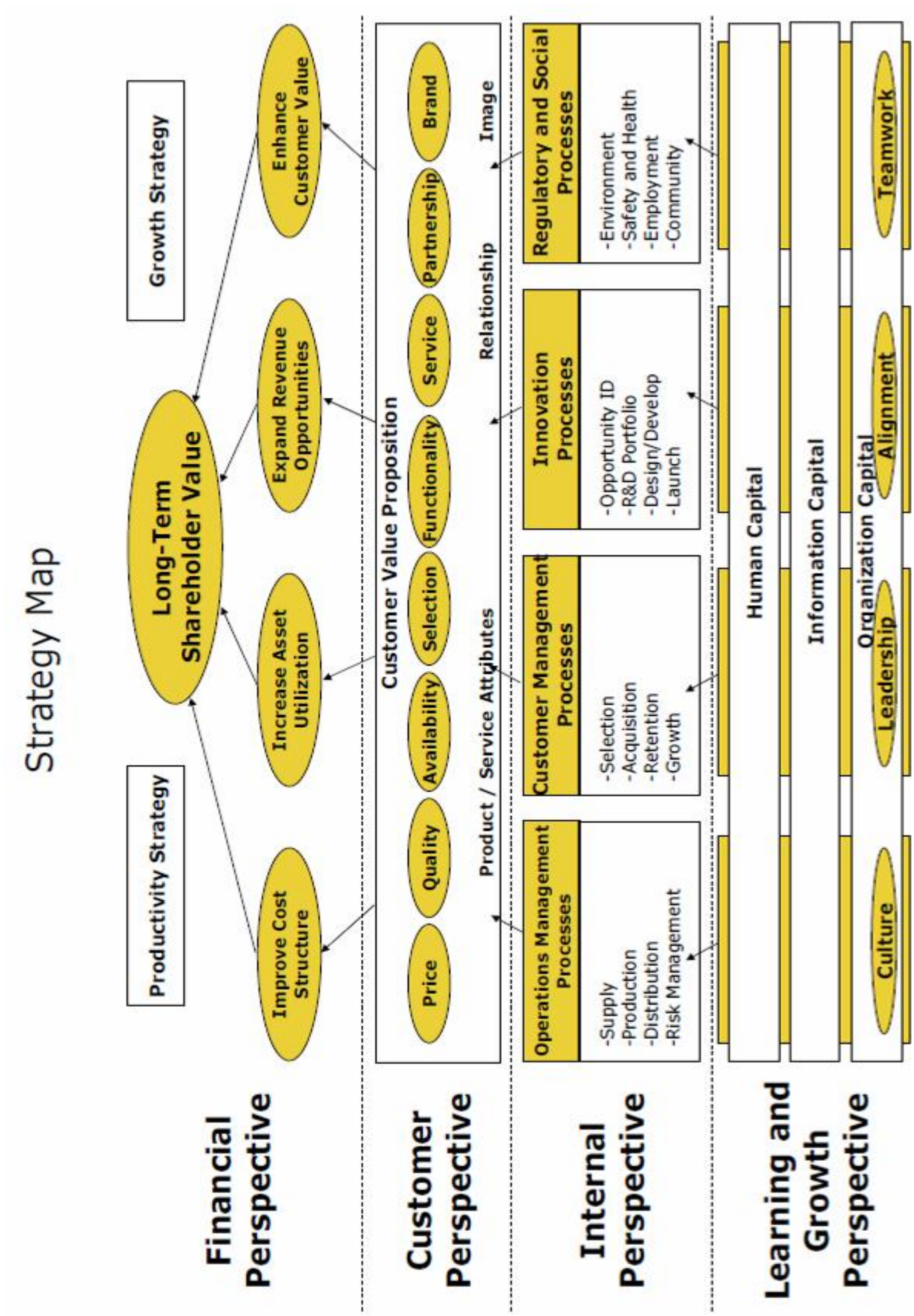
Xie, Y. Y., Li, X., Ngai, E. W. T., & Ying, W. Y. (2009). Customer churn prediction using improved balanced random forests. Expert Systems with Applications, 36(3), 5445-5449. doi: 10.1016/j.eswa.2008.06.121

(This page was intentionally left in blank)

ANNEXES

(This page was intentionally left in blank)

A1 - EXAMPLE OF A STRATEGY MAP



A2 - DISCOUNT RATE SENSITIVITY ON CASH FLOW

		CF1		CF2		CF3
Discount Rate		0.05		0.075		0.1
Retention Rate		1		1		1

Period (t)	Prob1(active)	CF1	Prob2(active)	CF2	Prob3(active)	CF3
0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1	1.0000	0.9524	1.0000	0.9302	1.0000	0.9091
2	1.0000	0.9070	1.0000	0.8653	1.0000	0.8264
3	1.0000	0.8638	1.0000	0.8050	1.0000	0.7513
4	1.0000	0.8227	1.0000	0.7488	1.0000	0.6830
5	1.0000	0.7835	1.0000	0.6966	1.0000	0.6209
6	1.0000	0.7462	1.0000	0.6480	1.0000	0.5645
7	1.0000	0.7107	1.0000	0.6028	1.0000	0.5132
8	1.0000	0.6768	1.0000	0.5607	1.0000	0.4665
9	1.0000	0.6446	1.0000	0.5216	1.0000	0.4241
10	1.0000	0.6139	1.0000	0.4852	1.0000	0.3855
11	1.0000	0.5847	1.0000	0.4513	1.0000	0.3505
12	1.0000	0.5568	1.0000	0.4199	1.0000	0.3186
13	1.0000	0.5303	1.0000	0.3906	1.0000	0.2897
14	1.0000	0.5051	1.0000	0.3633	1.0000	0.2633
15	1.0000	0.4810	1.0000	0.3380	1.0000	0.2394
16	1.0000	0.4581	1.0000	0.3144	1.0000	0.2176
17	1.0000	0.4363	1.0000	0.2925	1.0000	0.1978
18	1.0000	0.4155	1.0000	0.2720	1.0000	0.1799
19	1.0000	0.3957	1.0000	0.2531	1.0000	0.1635
20	1.0000	0.3769	1.0000	0.2354	1.0000	0.1486
21	1.0000	0.3589	1.0000	0.2190	1.0000	0.1351
22	1.0000	0.3418	1.0000	0.2037	1.0000	0.1228
23	1.0000	0.3256	1.0000	0.1895	1.0000	0.1117
24	1.0000	0.3101	1.0000	0.1763	1.0000	0.1015
25	1.0000	0.2953	1.0000	0.1640	1.0000	0.0923
26	1.0000	0.2812	1.0000	0.1525	1.0000	0.0839
27	1.0000	0.2678	1.0000	0.1419	1.0000	0.0763
28	1.0000	0.2551	1.0000	0.1320	1.0000	0.0693
29	1.0000	0.2429	1.0000	0.1228	1.0000	0.0630
30	1.0000	0.2314	1.0000	0.1142	1.0000	0.0573
31	1.0000	0.2204	1.0000	0.1063	1.0000	0.0521
32	1.0000	0.2099	1.0000	0.0988	1.0000	0.0474
33	1.0000	0.1999	1.0000	0.0919	1.0000	0.0431
34	1.0000	0.1904	1.0000	0.0855	1.0000	0.0391
35	1.0000	0.1813	1.0000	0.0796	1.0000	0.0356
36	1.0000	0.1727	1.0000	0.0740	1.0000	0.0323
37	1.0000	0.1644	1.0000	0.0688	1.0000	0.0294
38	1.0000	0.1566	1.0000	0.0640	1.0000	0.0267
39	1.0000	0.1491	1.0000	0.0596	1.0000	0.0243
40	1.0000	0.1420	1.0000	0.0554	1.0000	0.0221
41	1.0000	0.1353	1.0000	0.0516	1.0000	0.0201
42	1.0000	0.1288	1.0000	0.0480	1.0000	0.0183
43	1.0000	0.1227	1.0000	0.0446	1.0000	0.0166
44	1.0000	0.1169	1.0000	0.0415	1.0000	0.0151
45	1.0000	0.1113	1.0000	0.0386	1.0000	0.0137
46	1.0000	0.1060	1.0000	0.0359	1.0000	0.0125
47	1.0000	0.1009	1.0000	0.0334	1.0000	0.0113
48	1.0000	0.0961	1.0000	0.0311	1.0000	0.0103
49	1.0000	0.0916	1.0000	0.0289	1.0000	0.0094
50	1.0000	0.0872	1.0000	0.0269	1.0000	0.0085
		19.1687		**13.9479**		**10.9063**

A3 - PRESENT VALUE FUNCTION REFERENCE TABLE

Per monetary unit calibrated for 50 CF stream of years

		dr																	
eR	rr	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20	0.21	0.22
4%	0.96	11.53	10.53	9.68	8.98	8.37	7.85	7.39	7.00	6.65	6.33	6.05	5.80	5.57	5.36	5.17	5.00	4.84	4.69
5%	0.95	10.43	9.60	8.89	8.29	7.78	7.33	6.93	6.59	6.28	6.00	5.75	5.52	5.32	5.13	4.96	4.80	4.65	4.52
6%	0.94	9.51	8.81	8.22	7.71	7.26	6.87	6.53	6.22	5.95	5.70	5.48	5.27	5.09	4.92	4.76	4.62	4.48	4.36
7%	0.93	8.73	8.14	7.64	7.20	6.81	6.47	6.17	5.89	5.65	5.43	5.23	5.04	4.87	4.72	4.58	4.44	4.32	4.21
8%	0.92	8.07	7.57	7.13	6.75	6.41	6.11	5.84	5.60	5.38	5.18	5.00	4.83	4.68	4.54	4.41	4.29	4.17	4.07
9%	0.91	7.49	7.06	6.69	6.35	6.05	5.79	5.55	5.33	5.14	4.96	4.79	4.64	4.50	4.37	4.25	4.14	4.03	3.94
10%	0.90	7.00	6.62	6.29	6.00	5.74	5.50	5.29	5.09	4.91	4.75	4.60	4.46	4.33	4.21	4.10	4.00	3.90	3.81
11%	0.89	6.56	6.23	5.94	5.68	5.45	5.24	5.05	4.87	4.71	4.56	4.42	4.30	4.18	4.07	3.97	3.87	3.78	3.70
12%	0.88	6.18	5.89	5.63	5.40	5.19	5.00	4.83	4.67	4.52	4.38	4.26	4.14	4.03	3.93	3.84	3.75	3.67	3.59
13%	0.87	5.83	5.58	5.35	5.14	4.95	4.78	4.62	4.48	4.35	4.22	4.11	4.00	3.90	3.81	3.72	3.64	3.56	3.49
14%	0.86	5.53	5.30	5.10	4.91	4.74	4.58	4.44	4.31	4.19	4.07	3.97	3.87	3.77	3.69	3.61	3.53	3.46	3.39
15%	0.85	5.25	5.05	4.86	4.70	4.54	4.40	4.27	4.15	4.04	3.93	3.83	3.74	3.66	3.58	3.50	3.43	3.36	3.30
16%	0.84	5.00	4.82	4.65	4.50	4.36	4.23	4.11	4.00	3.90	3.80	3.71	3.62	3.55	3.47	3.40	3.33	3.27	3.21
17%	0.83	4.77	4.61	4.46	4.32	4.19	4.07	3.96	3.86	3.77	3.68	3.59	3.52	3.44	3.37	3.31	3.24	3.18	3.13
18%	0.82	4.57	4.42	4.28	4.15	4.04	3.93	3.83	3.73	3.65	3.56	3.48	3.41	3.34	3.28	3.22	3.16	3.10	3.05
19%	0.81	4.37	4.24	4.12	4.00	3.89	3.79	3.70	3.61	3.53	3.45	3.38	3.31	3.25	3.19	3.13	3.08	3.03	2.98
20%	0.80	4.20	4.08	3.96	3.86	3.76	3.67	3.58	3.50	3.42	3.35	3.29	3.22	3.16	3.11	3.05	3.00	2.95	2.90
21%	0.79	4.04	3.93	3.82	3.72	3.63	3.55	3.47	3.39	3.32	3.26	3.19	3.14	3.08	3.03	2.98	2.93	2.88	2.84
22%	0.78	3.89	3.79	3.69	3.60	3.52	3.44	3.36	3.29	3.23	3.17	3.11	3.05	3.00	2.95	2.90	2.86	2.81	2.77
23%	0.77	3.75	3.66	3.57	3.48	3.41	3.33	3.26	3.20	3.14	3.08	3.03	2.97	2.92	2.88	2.83	2.79	2.75	2.71
24%	0.76	3.62	3.53	3.45	3.37	3.30	3.24	3.17	3.11	3.05	3.00	2.95	2.90	2.85	2.81	2.77	2.73	2.69	2.65
25%	0.75	3.50	3.42	3.34	3.27	3.21	3.14	3.08	3.03	2.97	2.92	2.87	2.83	2.79	2.74	2.70	2.67	2.63	2.60
26%	0.74	3.39	3.31	3.24	3.18	3.11	3.06	3.00	2.95	2.90	2.85	2.80	2.76	2.72	2.68	2.64	2.61	2.57	2.54
27%	0.73	3.28	3.21	3.15	3.09	3.03	2.97	2.92	2.87	2.83	2.78	2.74	2.70	2.66	2.62	2.59	2.55	2.52	2.49
28%	0.72	3.18	3.12	3.06	3.00	2.95	2.89	2.85	2.80	2.76	2.71	2.67	2.64	2.60	2.57	2.53	2.50	2.47	2.44
29%	0.71	3.09	3.03	2.97	2.92	2.87	2.82	2.77	2.73	2.69	2.65	2.61	2.58	2.54	2.51	2.48	2.45	2.42	2.39
30%	0.70	3.00	2.94	2.89	2.84	2.79	2.75	2.71	2.67	2.63	2.59	2.56	2.52	2.49	2.46	2.43	2.40	2.37	2.35
31%	0.69	2.92	2.86	2.82	2.77	2.72	2.68	2.64	2.60	2.57	2.53	2.50	2.47	2.44	2.41	2.38	2.35	2.33	2.30
32%	0.68	2.84	2.79	2.74	2.70	2.66	2.62	2.58	2.55	2.51	2.48	2.45	2.42	2.39	2.36	2.33	2.31	2.28	2.26
33%	0.67	2.76	2.72	2.68	2.63	2.60	2.56	2.52	2.49	2.46	2.43	2.40	2.37	2.34	2.31	2.29	2.26	2.24	2.22
34%	0.66	2.69	2.65	2.61	2.57	2.53	2.50	2.47	2.43	2.40	2.38	2.35	2.32	2.29	2.27	2.25	2.22	2.20	2.18
35%	0.65	2.62	2.59	2.55	2.51	2.48	2.44	2.41	2.38	2.35	2.33	2.30	2.27	2.25	2.23	2.20	2.18	2.16	2.14
36%	0.64	2.56	2.52	2.49	2.45	2.42	2.39	2.36	2.33	2.31	2.28	2.25	2.23	2.21	2.19	2.16	2.14	2.12	2.10
37%	0.63	2.50	2.47	2.43	2.40	2.37	2.34	2.31	2.29	2.26	2.24	2.21	2.19	2.17	2.15	2.12	2.11	2.09	2.07
38%	0.62	2.44	2.41	2.38	2.35	2.32	2.29	2.27	2.24	2.22	2.19	2.17	2.15	2.13	2.11	2.09	2.07	2.05	2.03
39%	0.61	2.39	2.36	2.33	2.30	2.27	2.24	2.22	2.20	2.17	2.15	2.13	2.11	2.09	2.07	2.05	2.03	2.02	2.00

eR = erosion rate
rR= retention rate
dR = discount rate

A4 - BMS TRANSITION MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	{1,...}	{0}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	{1,...}	-	{0}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	{1,...}	-	-	-	{0}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	{1,...}	-	-	-	-	{0}	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	{2,...}	{1}	-	-	-	-	{0}	-	-	-	-	-	-	-	-	-	-	-	-	-
6	{2,...}	-	{1}	-	-	-	-	{0}	-	-	-	-	-	-	-	-	-	-	-	-
7	{2,...}	-	-	{1}	-	-	-	{0}	-	-	-	-	-	-	-	-	-	-	-	-
8	{3,...}	{2}	-	-	{1}	-	-	-	-	{0}	-	-	-	-	-	-	-	-	-	-
9	{3,...}	-	{2}	-	-	{1}	-	-	-	{0}	-	-	-	-	-	-	-	-	-	-
10	{3,...}	-	-	{2}	-	-	{1}	-	-	-	{0}	-	-	-	-	-	-	-	-	-
11	{3,...}	-	-	{2}	-	-	{1}	-	-	-	-	{0}	-	-	-	-	-	-	-	-
12	{4,...}	{3}	-	-	{2}	-	-	{1}	-	-	-	-	{0}	-	-	-	-	-	-	-
13	{4,...}	-	{3}	-	{2}	-	-	{1}	-	-	-	-	-	{0}	-	-	-	-	-	-
14	{4,...}	-	{3}	-	-	{2}	-	-	{1}	-	-	-	-	-	{0}	-	-	-	-	-
15	{4,...}	-	{3}	-	-	{2}	-	-	{1}	-	-	-	-	-	-	{0}	-	-	-	-
16	{4,...}	-	-	{3}	-	-	{2}	-	-	{1}	-	-	-	-	-	-	{0}	-	-	-
17	{4,...}	-	-	{3}	-	-	{2}	-	-	{1}	-	-	-	-	-	-	-	{0}	-	-
18	-	-	{4,...}	-	-	{3}	-	-	{2}	-	-	-	-	{1}	-	-	-	-	{0}	-
19	-	-	-	{4,...}	-	-	{3}	-	-	{2}	-	-	-	-	-	{1}	-	-	-	{0}
20	-	-	-	-	-	{4,...}	-	-	{3}	-	-	-	-	{2}	-	-	-	{1}	-	{0}

Bonus-Malus systems can be considered as homogeneous Markov chains with a finite state space of bonus malus classes. The classes are ordered such that the corresponding premiums are decreasing. The first class is sometimes called super malus class and the last class super bonus class.

Frequently a Poisson distribution is used to model the transition probabilities within a bonus-malus system. To be more specific, the Poisson distribution describes the number of claims for an individual and the transition probabilities are determined from this claim frequency distribution.

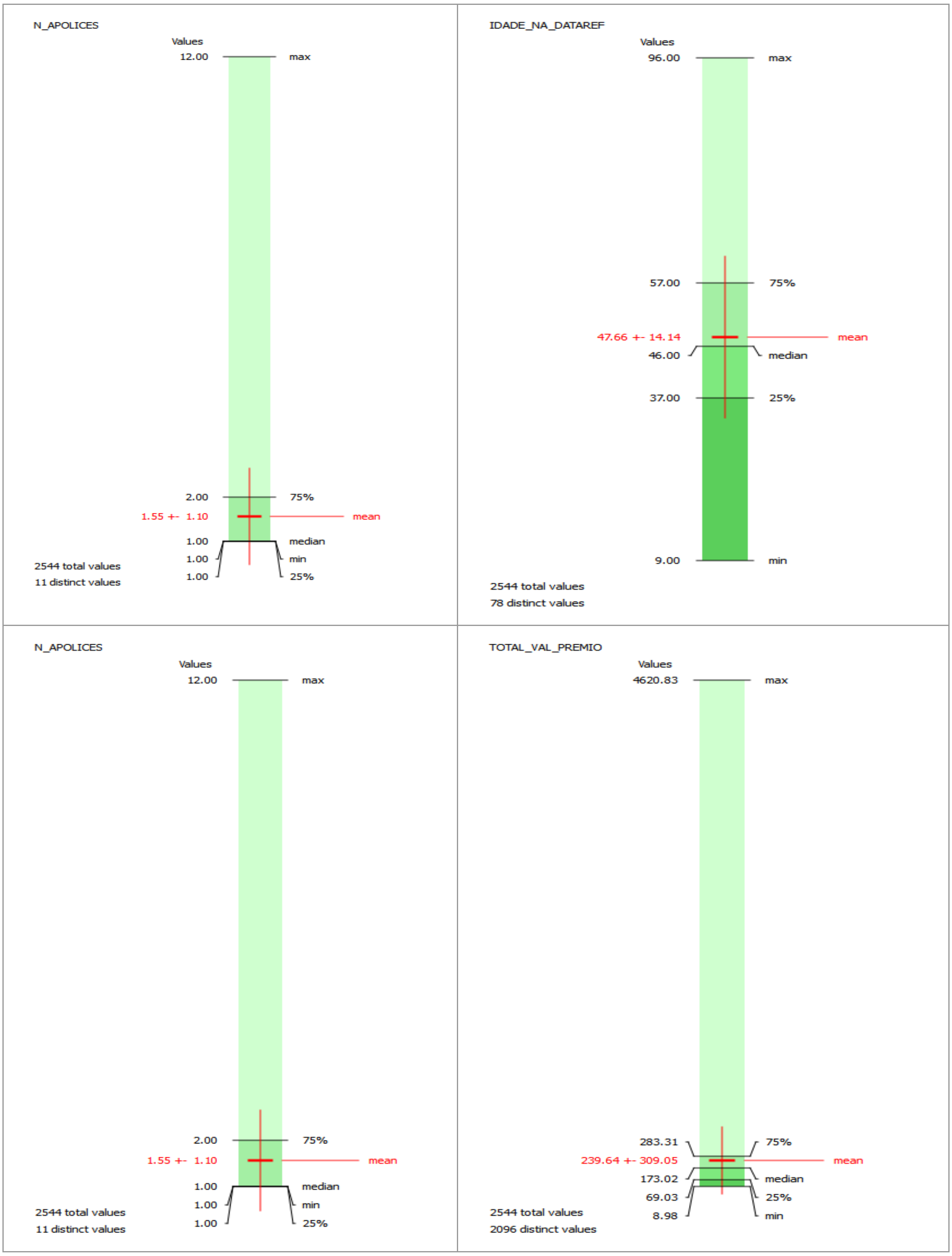
.

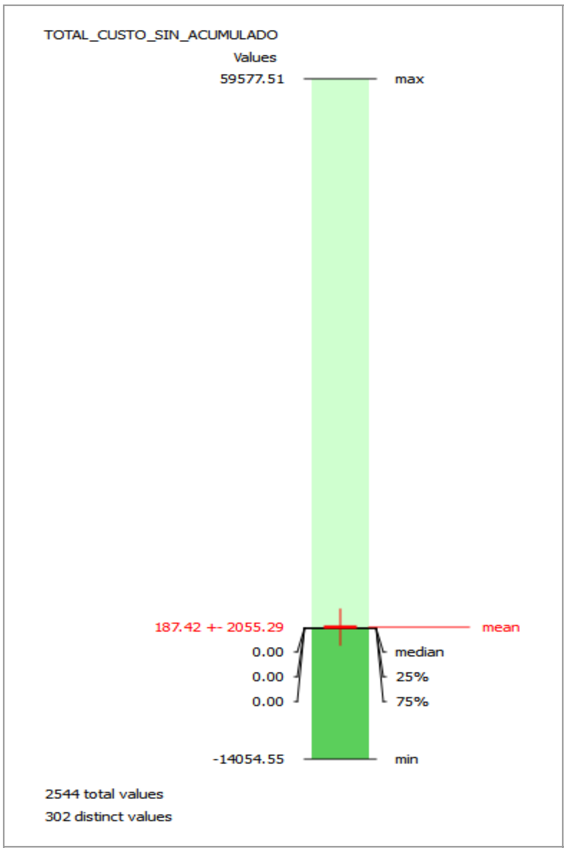
A5 - PRODUCT MAXIMAL TABLE

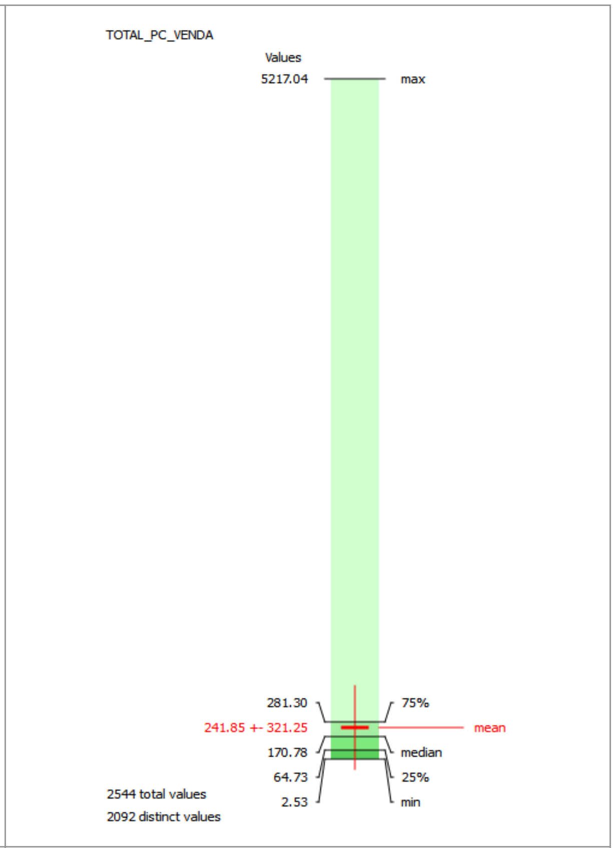
```
CREATE TABLE [dbo].[MAXIMAL_PRODUTO]
(
      [Ano]                          [int] NULL,
      [DUR_ANOS]                     [int] NULL,
      [DUR_DIAS]                     [int] NULL,
      [IDADE_NA_DATAREF]             [int] NULL,
      [FLAG_COM]                     [int] NOT NULL,
      [FLAG_TELEMOVEL]               [int] NOT NULL,
      [CLIENTE]                      [varchar](10) NULL,
      [COD_APOLICE]                  [nvarchar](255) NULL,
      [COD_PRODUTO]                  [nvarchar](255) NULL,
      [DSC_AREA_PRODUTO]             [nvarchar](255) NULL,
      [VAL_PREMIO]                   [decimal](18, 2) NULL,
      [COD_TIPO_FRACCIONAMENTO]      [nvarchar](255) NULL,
      [COD_TIPO_CLIENTE]             [nvarchar](255) NULL,
      [IND_COLABORADOR]              [nvarchar](255) NULL,
      [delegado]                     [nvarchar](50) NULL,
      [canal]                        [nvarchar](15) NULL,
      [sexo]                         [nvarchar](1) NULL,
      [estado_civil]                 [varchar](1) NULL,
      [premio_comercial_venda]       [decimal](18, 2) NULL,
      [flag_campanha]                [int] NOT NULL,
      [custo_sinistros_acumulados]   [decimal](18, 2) NULL,
      [custo_sinistros_ano]          [decimal](18, 2) NULL,
      [premios_acumulados]           [decimal](18, 2) NULL,
      [premios_ano]                  [decimal](18, 2) NULL,
      [flag_tem_sinistros]           [int] NOT NULL,
      [flag_anulada]                 [int] NOT NULL
)
```

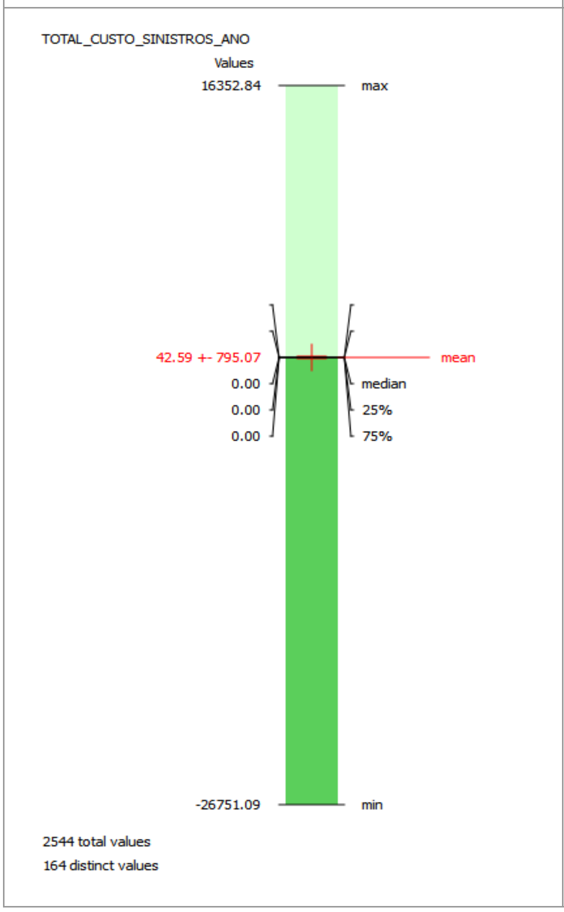
A6 - CUSTOMER DESCRIPTIVE STATISTICS

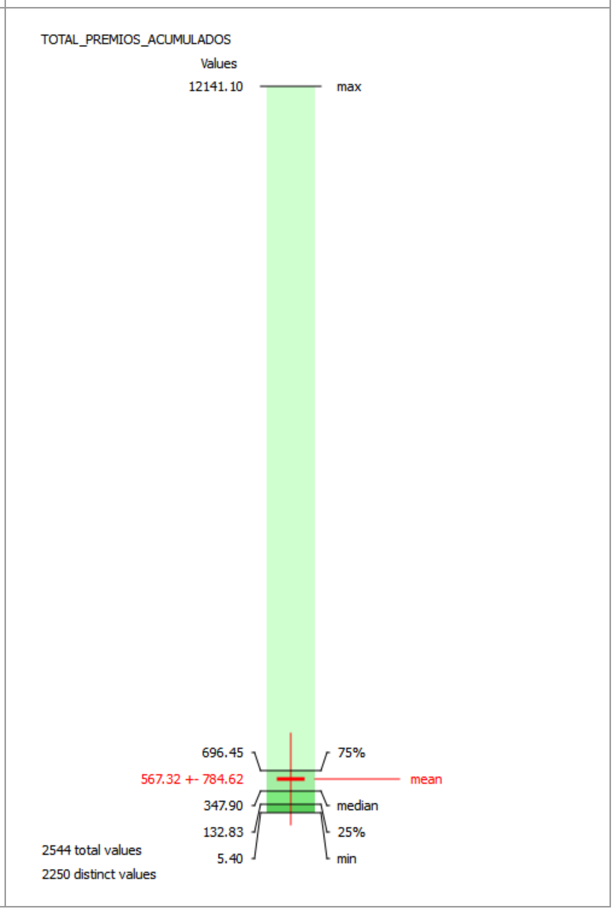
Reference year 2007 - 2544 Observations Random Sampled

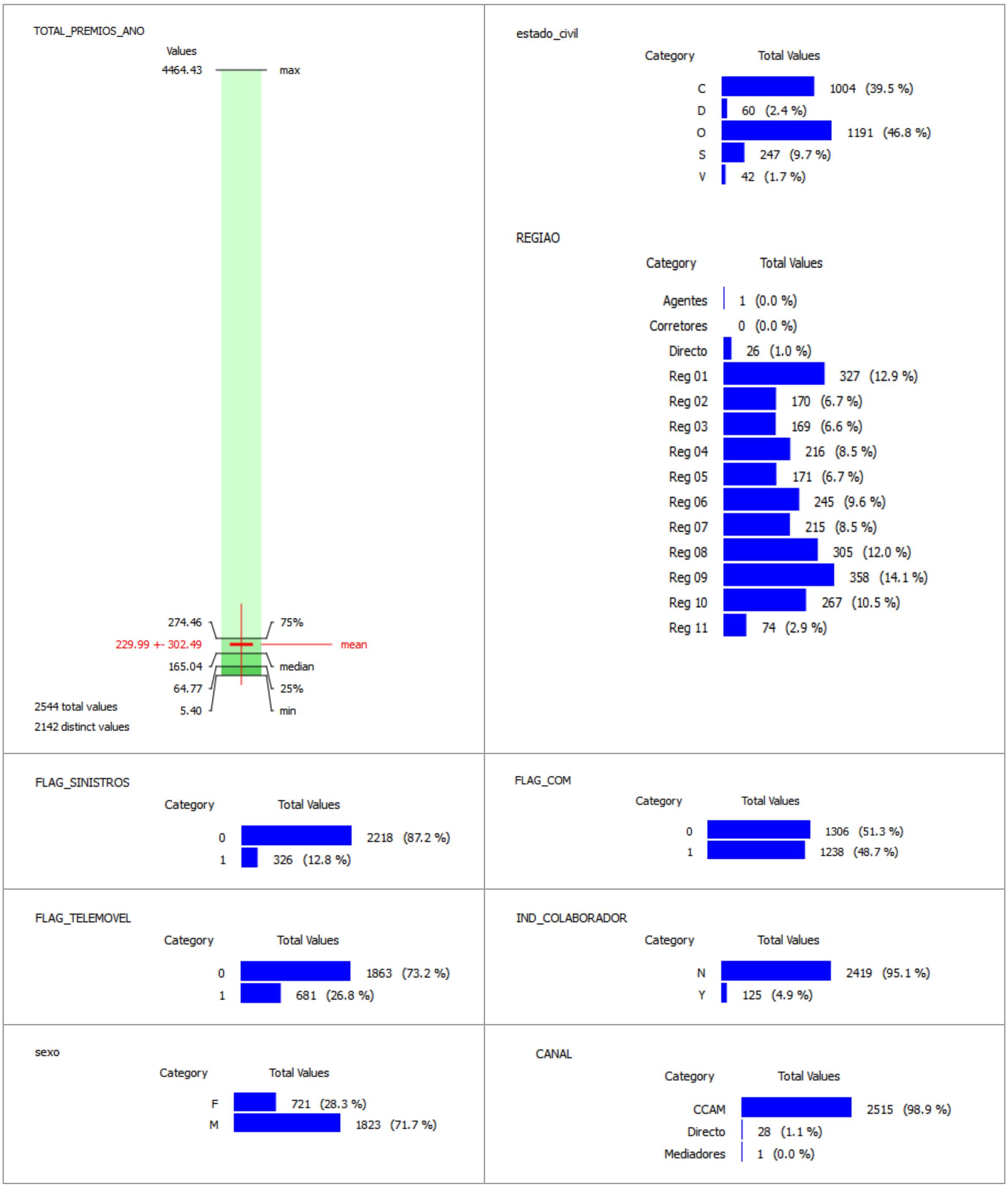



TOTAL_CUSTO_SIN_ACUMULADO
Values
59577.51
max
187.42 +- 2055.29
mean
0.00
median
0.00
25%
0.00
75%
-14054.55
min
2544 total values
302 distinct values


TOTAL_PC_VENDA
Values
5217.04
max
281.30
75%
241.85 +- 321.25
mean
170.78
median
64.73
25%
2.53
min
2544 total values
2092 distinct values

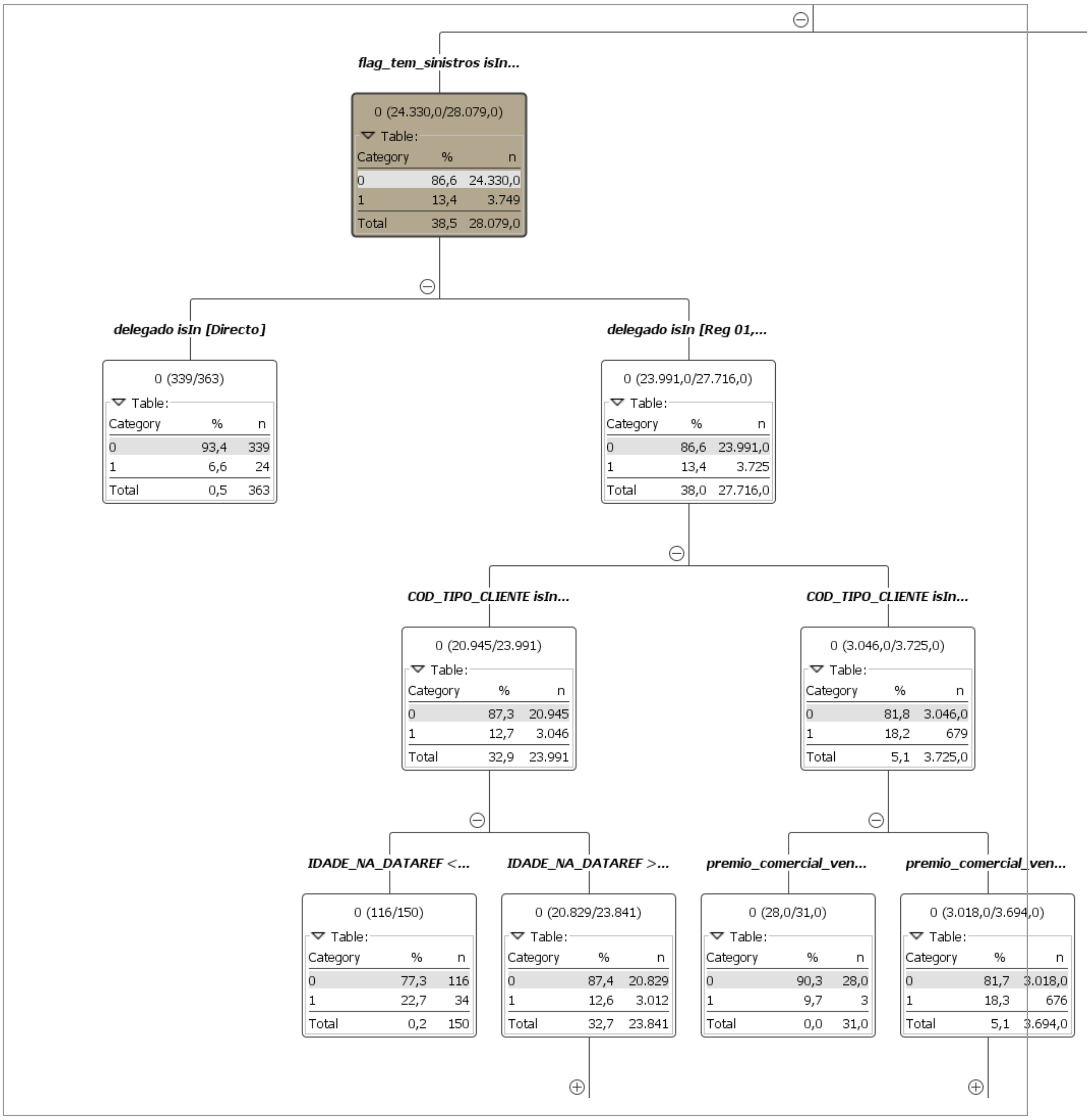

TOTAL_CUSTO_SINISTROS_ANO
Values
16352.84
max
42.59 +- 795.07
mean
0.00
median
0.00
25%
0.00
75%
-26751.09
min
2544 total values
164 distinct values


TOTAL_PREMIOS_ACUMULADOS
Values
12141.10
max
696.45
75%
567.32 +- 784.62
mean
347.90
median
132.83
25%
5.40
min
2544 total values
2250 distinct values


TOTAL_PREMIOS_ANO
Values
4464.43 max
274.46 75%
229.99 +- 302.49 mean
165.04 median
64.77 25%
5.40 min
2544 total values
2142 distinct values
estado_civil
Category Total Values
C 1004 (39.5 %)
D 60 (2.4 %)
O 1191 (46.8 %)
S 247 (9.7 %)
V 42 (1.7 %)
REGIAO
Category Total Values
Agentes 1 (0.0 %)
Corretores 0 (0.0 %)
Directo 26 (1.0 %)
Reg 01 327 (12.9 %)
Reg 02 170 (6.7 %)
Reg 03 169 (6.6 %)
Reg 04 216 (8.5 %)
Reg 05 171 (6.7 %)
Reg 06 245 (9.6 %)
Reg 07 215 (8.5 %)
Reg 08 305 (12.0 %)
Reg 09 358 (14.1 %)
Reg 10 267 (10.5 %)
Reg 11 74 (2.9 %)
FLAG_SINISTROS
Category Total Values
0 2218 (87.2 %)
1 326 (12.8 %)
FLAG_COM
Category Total Values
0 1306 (51.3 %)
1 1238 (48.7 %)
FLAG_TELEMOVEL
Category Total Values
0 1863 (73.2 %)
1 681 (26.8 %)
IND_COLABORADOR
Category Total Values
N 2419 (95.1 %)
Y 125 (4.9 %)
sexo
Category Total Values
F 721 (28.3 %)
M 1823 (71.7 %)
CANAL
Category Total Values
CCAM 2515 (98.9 %)
Directo 28 (1.1 %)
Mediadores 1 (0.0 %)

A7 - EXAMPLE OF DECISION TREE

Partial decision tree model for the Auto Insurance. In each node the percentage of the Class 1 is equivalent of churn probability with for the policies classified in that node.



A8 - COMPANY'S AGGREGATE CHURN RATE

Year	Product	Inicial portfolio	Surfivin g Porfolio	Retention Rate	Churn Rate	Average Policy Life (Years)
2007	Acidentes Pessoais	33,265	28,100	84.47%	15.53%	6.4
	Acidentes Trabalho CO	7,463	5,599	75.02%	24.98%	4.0
	Acidentes Trabalho CP	7,392	5,928	80.19%	19.81%	5.0
	Automóvel	104,181	87,031	83.54%	16.46%	6.1
	Comércio e Serviços	12,024	10,276	85.46%	14.54%	6.9
	Habitação	53,839	49,957	92.79%	7.21%	13.9
	Responsab. Civil	20,099	17,453	86.84%	13.16%	7.6
2007 Total		**238,263**	**204,344**	**85.76%**	**14.24%**	**7.0**
2008	Acidentes Pessoais	36,892	31,246	84.70%	15.30%	6.5
	Acidentes Trabalho CO	7,517	5,710	75.96%	24.04%	4.2
	Acidentes Trabalho CP	7,461	6,016	80.63%	19.37%	5.2
	Automóvel	108,057	91,644	84.81%	15.19%	6.6
	Comércio e Serviços	12,607	11,011	87.34%	12.66%	7.9
	Habitação	57,987	54,492	93.97%	6.03%	16.6
	Responsab. Civil	22,612	19,583	86.60%	13.40%	7.5
2008 Total		**253,133**	**219,702**	**86.79%**	**13.21%**	**7.6**
2009	Acidentes Pessoais	61,722	45,693	74.03%	25.97%	3.9
	Acidentes Trabalho CO	7,657	5,991	78.24%	21.76%	4.6
	Acidentes Trabalho CP	7,583	6,215	81.96%	18.04%	5.5
	Automóvel	113,909	98,027	86.06%	13.94%	7.2
	Comércio e Serviços	13,289	11,697	88.02%	11.98%	8.3
	Habitação	63,157	59,419	94.08%	5.92%	16.9
	Responsab. Civil	23,652	20,464	86.52%	13.48%	7.4
2009 Total		**290,969**	**247,506**	**85.06%**	**14.94%**	**6.7**
2010	Acidentes Pessoais	51,158	42,555	83.18%	16.82%	5.9
	Acidentes Trabalho CO	7,710	5,961	77.32%	22.68%	4.4
	Acidentes Trabalho CP	7,651	6,212	81.19%	18.81%	5.3
	Automóvel	118,428	101,488	85.70%	14.30%	7.0
	Comércio e Serviços	13,867	12,177	87.81%	12.19%	8.2
	Habitação	66,925	63,410	94.75%	5.25%	19.0
	Responsab. Civil	37,141	29,928	80.58%	19.42%	5.1
2010 Total		**302,880**	**261,731**	**86.41%**	**13.59%**	**7.4**

A9 - EXAMPLE OF PRODUCT BALANCE SHEET

Produto: Acidentes Pessoais

	Real	Orçamento	Desv. Abs.	Desv. %
(1) Prémios Adquiridos				
Prémios Brutos Emitidos				
Var. Provisão PNA				
(2) Custos com Sinistros				
Montantes Pagos				
Var. Provisão Sinistros				
(3) Margem Após Sinistros = (1)+(2)				
(4) Saldo de Resseguro				
Prémios de Resseguro Cedido				
Var. Provisão PNA, parte Resseguro				
Sinistros Pagos, parte Resseguro				
Var. Prov. Sinistros, parte resseguradores				
Comissões de Resseguro				
(5) Outros Proveitos e Custos Técnicos				
Var. Provisão Riscos em Curso				
Var. Prov. Desvios Sinistralidade				
Impostos e Taxas sobre Seguros				
Outros Custos e Proveitos Técnicos				
(6) Margem Técnica = (3)+(4)+(5)				
(7) Custos Comerciais				
Remunerações Mediação CCAM				
Var. Custos Aquisição Diferidos				
(8) Margem Distribuição = (6)+(7)				
Prémios Adquiridos / PBE				
Custos com Sinistros / Prémios Adquiridos				
Prémios Resseguro / PBE				
Saldo Resseguro / Prémios Adquiridos				
Margem Técnica / Prémios Adquiridos				
Remun Mediação / Prémios Comerciais				
Margem Distribuição / Prémios Adquiridos				

cR_n (Prémios Brutos Emitidos through Saldo de Resseguro)

A10 - MODEL ACURACY STATISTICS

The cutoff value for the model classification decision is determined by the average product churn rate.

Churn MODEL

RC

Cutoff= 0.118

RC		Real 1	Real 0	
Predicted	1	950	11,307	12,257
	0	2,565	52,103	54,668
		3,515	63,410	66,925

Sensitivity 0.270270
Specivity 0.821684
Accuracy 0.792723
Recall 0.270270
Precision 0.017907
f-statistic 0.033588

AP

Cutoff= 0.14

AP		Real 1	Real 0	
Predicted	1	2,296	9,813	12,109
	0	6,307	32,742	39,049
		8,603	42,555	51,158

Sensitivity 0.266884
Specivity 0.769404
Accuracy 0.684898
Recall 0.266884
Precision 0.065529
f-statistic 0.105222

ATCO

Cutoff= 0.254

ATCO		Real 1	Real 0	
Predicted	1	615	1,712	2,327
	0	1,134	4,249	5,383
		1,749	5,961	7,710

Sensitivity 0.351630
Specivity 0.712800
Accuracy 0.630869
Recall 0.351630
Precision 0.126439
f-statistic 0.185997

ATCP

Cutoff= 0.192

ATCP		Real 1	Real 0	
Predicted	1	700	2,605	3,305
	0	739	3,607	4,346
		1,439	6,212	7,651

Sensitivity 0.486449
Specivity 0.580650
Accuracy 0.562933
Recall 0.486449
Precision 0.162526
f-statistic 0.243648

AUTO

Cutoff= 0.160

AUTO		Real 1	Real 0	
Predicted	1	10,817	53,278	64,095
	0	6,123	48,210	54,333
		16,940	101,488	118,428

Sensitivity 0.638548
Specivity 0.475032
Accuracy 0.498421
Recall 0.638548
Precision 0.183255
f-statistic 0.284782

CS

Cutoff= 0.144

CS		Real 1	Real 0	
Predicted	1	296	1,746	2,042
	0	1,394	10,431	11,825
		1,690	12,177	13,867

Sensitivity 0.175148
Specivity 0.856615
Accuracy 0.773563
Recall 0.175148
Precision 0.027594
f-statistic 0.047677

HABIT	
Cutoff= 0.071	

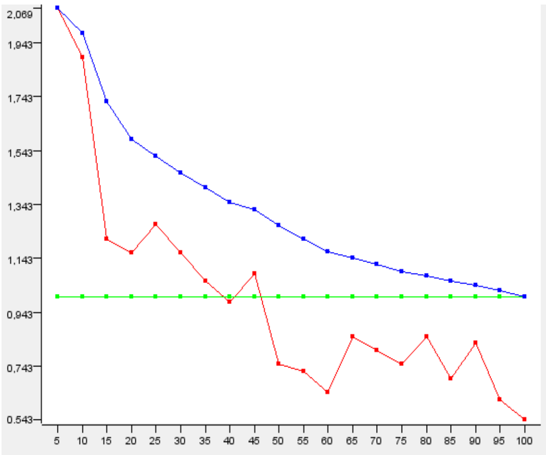
HABIT		Real		
		1	0	
Predicted	1	950	11,307	12,257
	0	2,565	52,103	54,668
		3,515	63,410	66,925

Sensitivity	0.27027
Specivity	0.82168
Accuracy	0.79272
Recall	0.27027
Precision	0.01791
f-statistic	0.03359

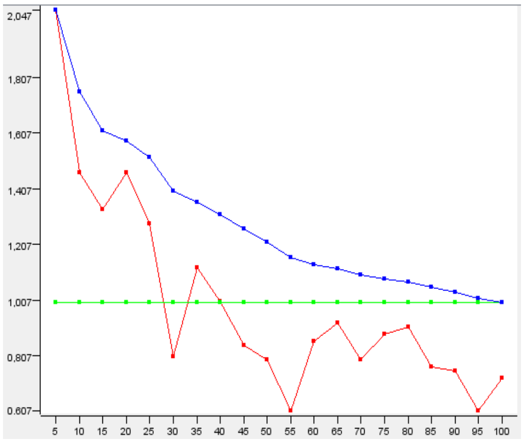
A11 - MODEL LIFT CHARTS

CHURN MODEL

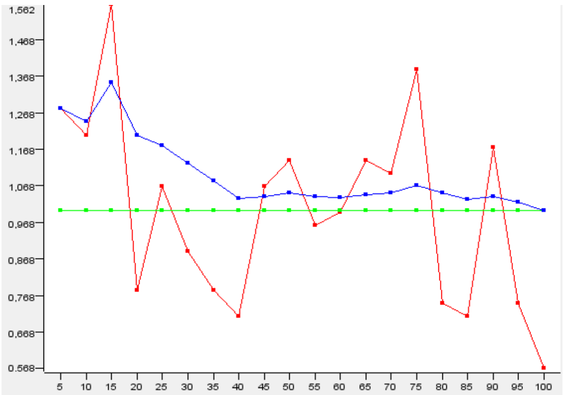
RC



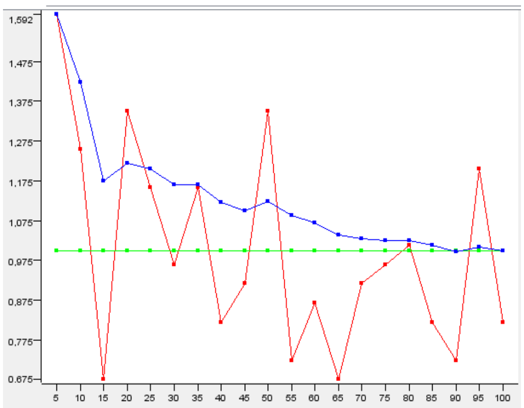
AP



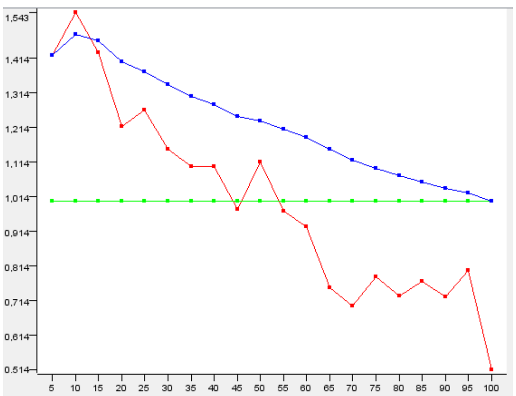
ATCO



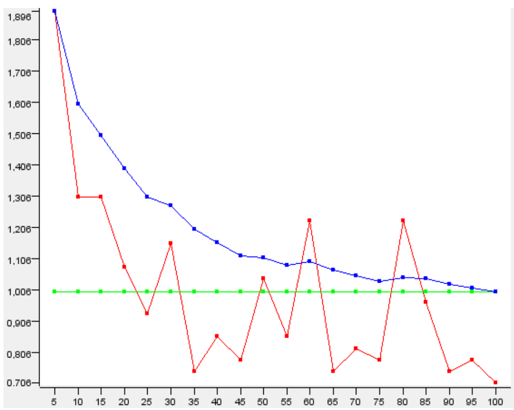
ATCP



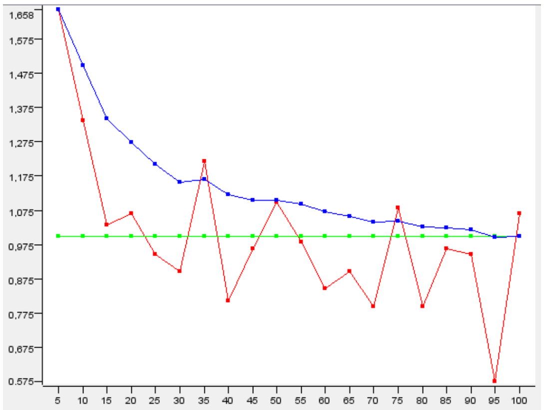
AUTO



CS

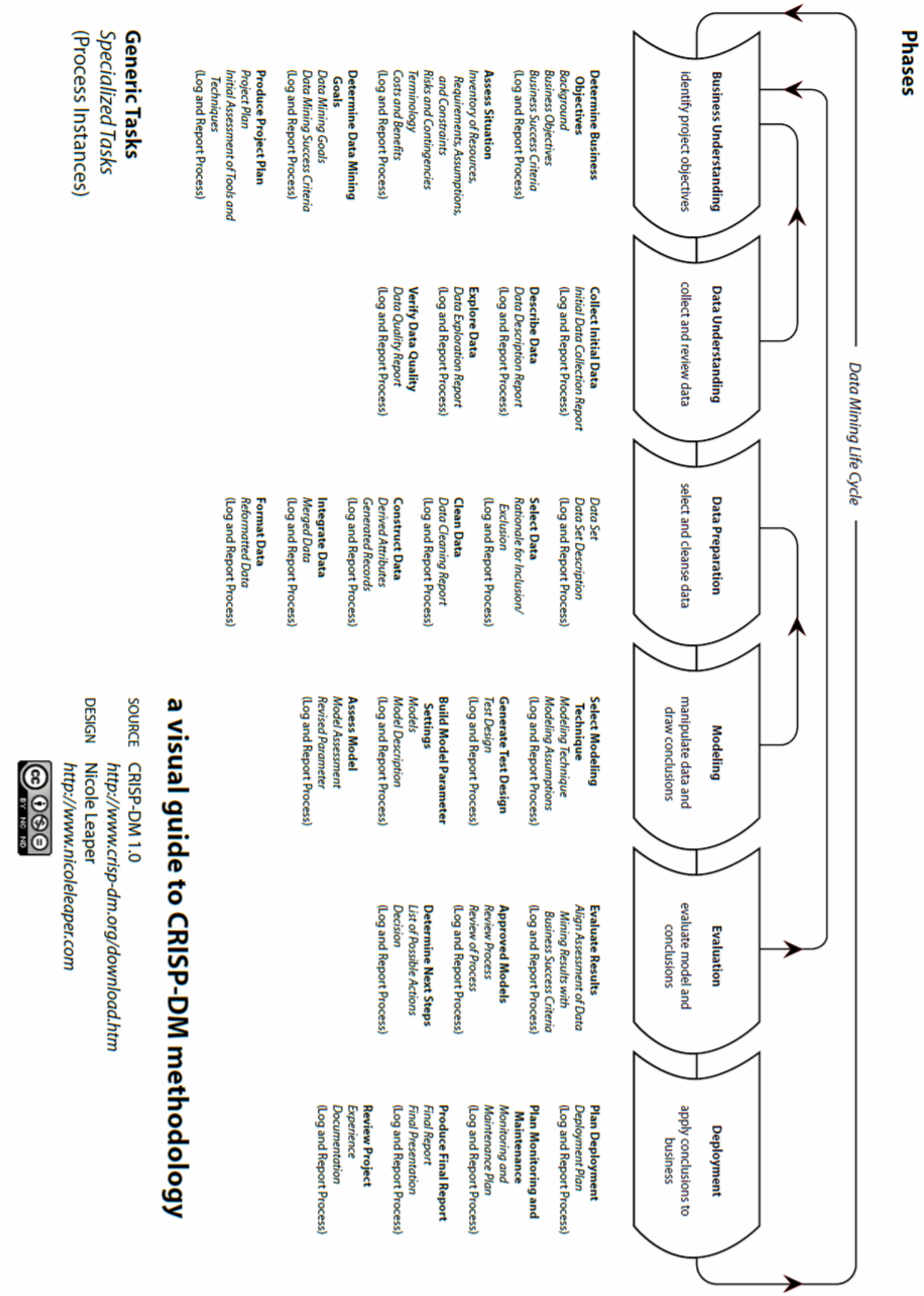


HABIT



- Green Line – Base Line
- Blue Line – Cumulative Lift
- Red Line – Lift

A12 - THE CRISP-DM PROCESS



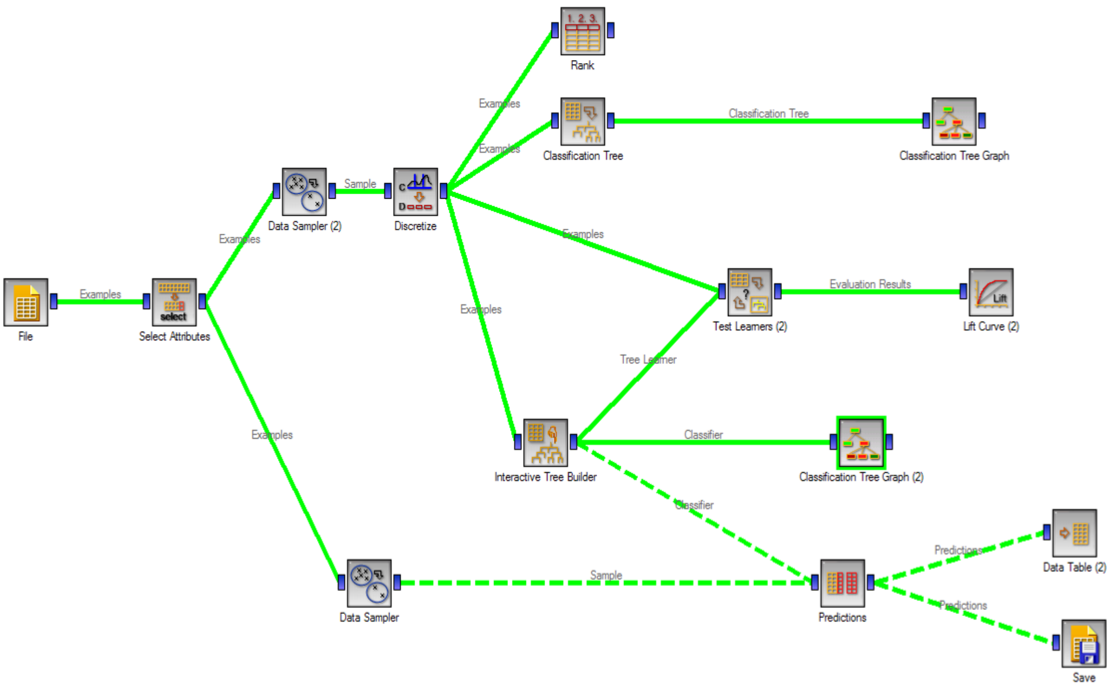
A13 - THE PLV FUNCTION IMPLEMENTATION

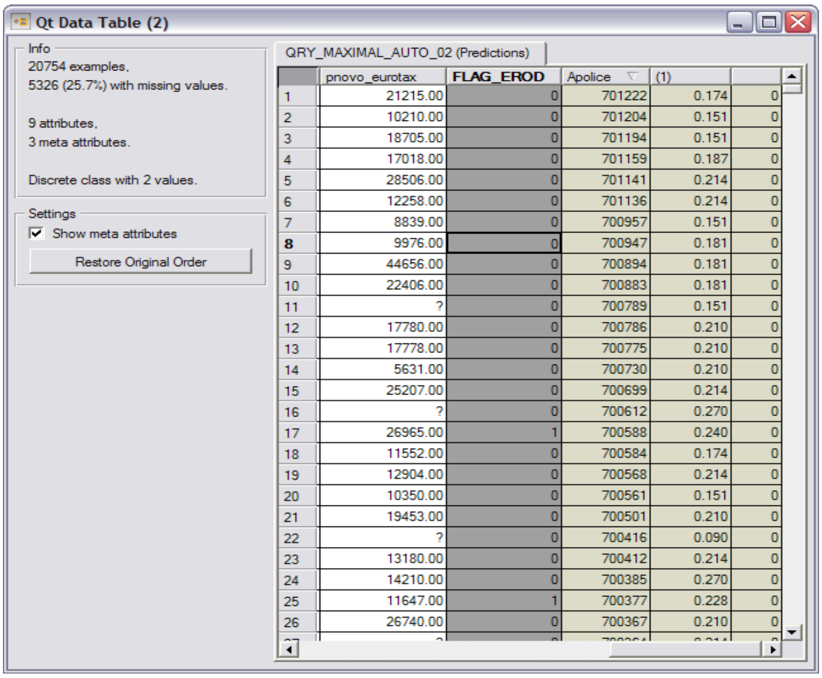
The used programming language is T-SQL 2008

```
CREATE FUNCTION dbo.PLV (@cashin float,@mF as float,@discount_rate
float,@retention_rate as float,@strucVar as float, @stationarity_rate
as float, @stationarity_offset as int,@stationarity_alfa as
float,@stationarity_beta as float, @EPUpsell as float)
RETURNS float
AS
BEGIN

      DECLARE @CF                      as float
      DECLARE @acum_CF                 as float
      DECLARE @prob_active             as float
      Declare @stationary_term         as float
      Declare @cashin_adjusted         as float

      DECLARE @period                  as int
      DECLARE @periods                 as int


      set @periods = 50
      set @period  = 0
      set @acum_CF  = 0
      set @prob_active = 1
      set @stationary_term = 1
      set @cashin_adjusted = @cashin

      while @period<@periods
      begin

            if (@period>0 and  @stationarity_rate>0)
                  set @stationary_term = (1-@stationarity_rate) +
(@stationarity_beta / power(@period +
@stationarity_offset,@stationarity_alfa))

            set @CF  = @cashin_adjusted * @mF * @prob_active * (1 +
@EPUpsell) * @stationary_term / power((1 + @discount_rate),@period)

            set @acum_CF = @acum_CF + @CF
            set @prob_active= @prob_active * (1-@retention_rate)
            set @cashin_adjusted = @cashin_adjusted * (1+@strucVar)


            set @period = @period + 1
      end

      RETURN (@acum_CF)

END
```

A14 – EARLY EXPERIMENTS WITH DM MODELS

Orange Dataflow for Auto Insurance churn model – Decision Tree





Qt Data Table (2)

Info
20754 examples,
5326 (25.7%) with missing values.

9 attributes,
3 meta attributes.

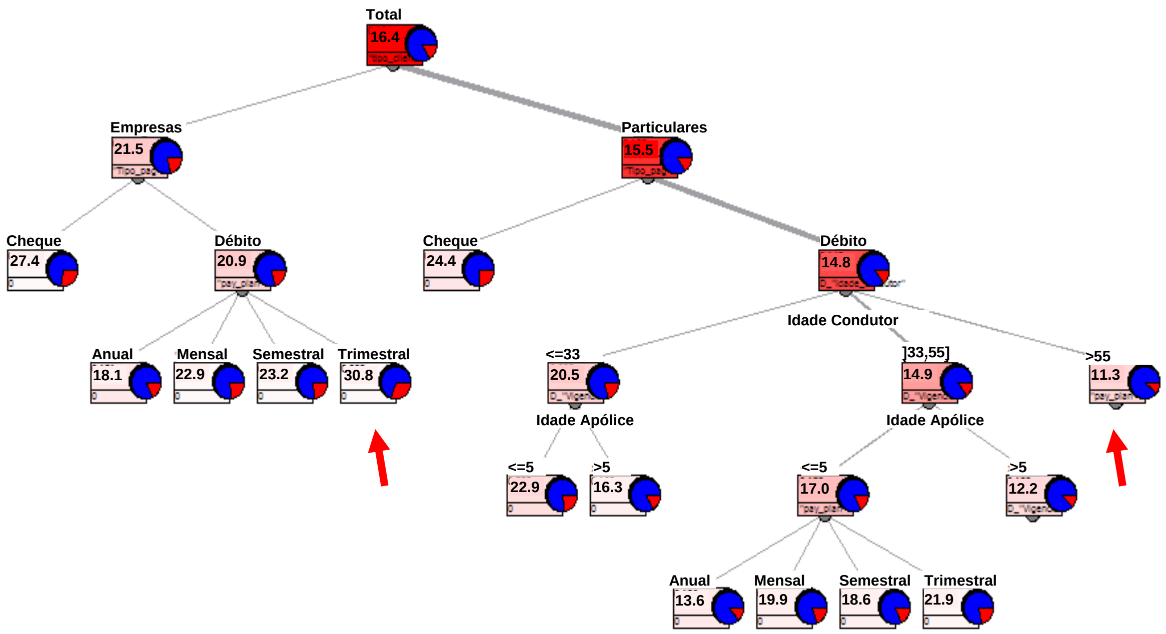
Discrete class with 2 values.

Settings
Show meta attributes
Restore Original Order

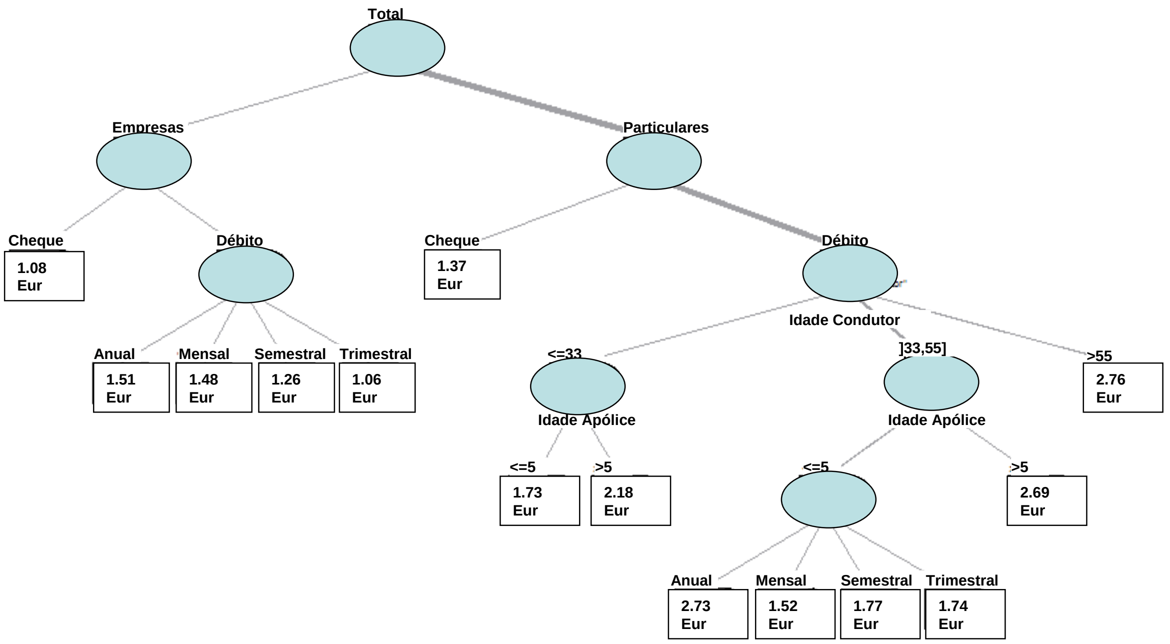
QRY_MAXIMAL_AUTO_02 (Predictions)

	pnovo_eurotax	FLAG_EROD	Apolice	(1)	
1	21215.00	0	701222	0.174	0
2	10210.00	0	701204	0.151	0
3	18705.00	0	701194	0.151	0
4	17018.00	0	701159	0.187	0
5	28506.00	0	701141	0.214	0
6	12258.00	0	701136	0.214	0
7	8839.00	0	700957	0.151	0
8	9976.00	0	700947	0.181	0
9	44656.00	0	700894	0.181	0
10	22406.00	0	700883	0.181	0
11	?	0	700789	0.151	0
12	17780.00	0	700786	0.210	0
13	17778.00	0	700775	0.210	0
14	5631.00	0	700730	0.210	0
15	25207.00	0	700699	0.214	0
16	?	0	700612	0.270	0
17	26965.00	1	700588	0.240	0
18	11552.00	0	700584	0.174	0
19	12904.00	0	700568	0.214	0
20	10350.00	0	700561	0.151	0
21	19453.00	0	700501	0.210	0
22	?	0	700416	0.090	0
23	13180.00	0	700412	0.214	0
24	14210.00	0	700385	0.270	0
25	11647.00	1	700377	0.228	0
26	26740.00	0	700367	0.210	0

Churn Decision Tree for Auto Insurance



Shadow CLV Decision Tree for Auto Insurance



(This page was intentionally left in blank)