Tunisian Republic
Ministry of Higher Education and Scientific Research
Carthage University - Engineering School of Statistics and Information Analysis

Graduation Project presented for the obtention of

National Engineering Diploma in Statistics and Information Analysis

Submitted by

Cyrine KOUKI

Customer life time value modeling using probabilistic models and machine learning

Defended on 22/06/2024 in front of the committee composed of:

Dr. Ines Abdeljaoued-Tej	President
Dr. Farouk Mhamdi	Examiner
Dr. Hichem Rammeh	Supervisor
M. Abdelkader Zouabi	Supervisor

A Graduation Project made at

Orange Tunisia

Dedication

To my dearest parents, Mokhtar and Rawdha Kouki, my twin sisters, Samar and Raafa, your unwavering support and love has fuelled my academic journey. Your belief in me and your advice have been my guiding lights. This achievement is a testament to your presence.

With the deepest gratitude to all of my teachers and professors for their dedication, commitment and expertise.

Thanks

I would like to thank the Orange Tunisia team for the excellent internship environment and my special thanks go to Mr Abdelkader ZOUABI for his advice throughout the internship.

I would like to thank my academic supervisor Dr. Hichem RAMMEH for his advice throughout the internship.

Thank you to Dr. Inès ABDELJAOUED TEJ for the time she devoted to me throughout the internship. Her ongoing advice and guidance allowed me to profoundly improve my work and the internship report.

I would like to thank the ESSAI academic team for the quality of the teaching they provided throughout my academic career.

I cannot end without thanking my family for their support throughout the internship. THANK YOU!

Abstract

In today's rapidly evolving business landscape, understanding customers and fostering strong relationships with them is crucial for long-term success. This study focuses on an innovative approach to Customer Lifetime Value (CLV) estimation, which combines survival analysis and revenue prediction. By doing so, we can accurately forecast the profit a customer will generate for a company over time. This not only provides insights into future profits but also helps identify the key variables and metrics that influence customer value, enabling companies to implement strategies to improve customer retention.

Additionally, we performed RFM (Recency, Frequency, Monetary) analysis to gain a deeper understanding of customer behavior and spending patterns. This classification helps segment customers based on their purchasing habits, providing valuable information to tailor marketing efforts and improve customer engagement. Together, these methods offer a comprehensive view of customer value, helping businesses make informed decisions and drive sustainable growth

Keywords— Machine learning, Prediction, Modeling, Customer lifetime value, Cox model, Gamma Gamma Model,XGBoost,Customer classification,Customer behaviour understanding

Résumé

Dans le monde des affaires d'aujourd'hui, en constante une évolution, comprendre les clients et établir des relations solides avec eux est essentiel pour réussir à long terme. Cette étude combine l'analyse de survie et la prédiction des revenus pour estimer combien de profit un client rapportera à une entreprise au fil du temps (CLV). Cela ne nous fournit pas seulement des informations sur les profits futurs, mais nous aide également à comprendre comment certaines variables et métriques les influencent, nous permettant ainsi de prendre des mesures pour fidéliser davantage de clients.

Nous avons également réalisé une classification RFM (Récence, Fréquence, Monétaire) pour comprendre les pratiques et les dépenses des clients. Cette double approche offre non seulement une vision claire des profits futurs mais permet aussi de mieux comprendre et influencer les comportements clients, ce qui est crucial pour la croissance durable d'une entreprise.

Mots clés— Apprentissage automatique, Prédiction, Modélisation, Pré-traitement des données, Cox model, Gamma Gamma Model,XGBoost

Contents

List of Figures

List of Tables

Introduction

One of the most significant elements of every successful company are customers as they represent the company's origin of its earnings, throughout the years companies became more and more competitive to provide its clients a better personalized experience.

The key metric to help the companies manage their customers strategy is customer life time value also known as CLV [13]. It is a key that measures how much benefit a client can bring to a business over the duration of its entire period of their relationship. Based on this measurement, the company makes decisions on how much money it should spend in advertising campaign, to retain a customer or attract a new one. CLV is an important indicator in the marketing process [7].

Within this framework, this end of study project uses passed data obtained from Orange telecommunication company's clients to model the CLV. We use different survival analysis methods to estimate the survival probability by taking the churn [11] of a costumer as the event of interest. We estimated the revenue and used the two estimated quantities to obtain the CLV.

We were provided a total of five customer data four of each consisting of approximately 200,927 customers and containing 21 features. The customers are consistent across all datasets, meaning that the same set of customers is present in each dataset. These datasets cover a time span of 24 months, 6 months each, starting from January 2022 and ending in January 2024. In this internship study, data engineering occupies the most important contribution and consumes a lot of time.

In Chapter 1, we introduce Orange Tunisia particularly focusing on the Customer Value Management (CVM) department. The chapter highlightes the importance of customer retention and the calculation of Customer Lifetime Value (CLV) for optimizing customer relationship management [12]. We provide an overview of the CRISP-DM methodology [14] used in this project. Additionally, we introduce common CLV models emphasizing on the limitations of Orange's detection models that only focus on churn prediction and mention the necessity to develop specific models that incorporate the unique characteristics of new customers using survival models to better estimate CLV and enhance retention strategies.

In chapter 2 we present the concept of Customer Life Time Value (CLV) and the formula of calculating it. The RFM modeling is used for two purposes: labelling customer segments and evaluating the CLV. The second techniques to (robustly) assess the CLV is to combine survival models for the survival probability estimation and generalized linear model for the revenue estimation. For this purpose we present the most known survival models: Kaplan-meier survival model, the Cox proportional hazard model and the Accelerated Failure Time (AFT) model [18].

In chapter 3, the project will begin by segmenting customers based on their Recency, Frequency, and Monetary (RFM) scores to better understand their value and engagement. This will also be useful for the Gamma Gamma model.

In chapter 4, we will use two separate approaches XGBoost classifiers and regressors to determine both the probability of being ”alive” (non-churner) and the revenue generated per month to estimate the CLV for a short-term period (3 months) and for the second approach we will calculate the CLV using two advanced probabilistic models: the Beta Geometric model and the Gamma Gamma model [6]. We perform a temporal split of the data into calibration and holdout sets for model training and validation, aiming to reveal insights into customer purchase behaviors and survival probabilities. Overall, this comprehensive approach will demonstrate the effectiveness of various models in predicting in CLV assessment.

In Chapter 5, we use survival analysis to estimate the probability of churn across different segments. Kaplan-Meier curves and the Cox model were employed to analyze survival rates and churn risk factors. We also conducted an Accelerated Failure Time (AFT) model to further understand the impact of various factors on survival time. We performed a Linear Regression analysis on the logarithm of customer revenue, accounting for customer heterogeneity. To refine our revenue estimation, we employed the XGBoost algorithm. We analyzed feature importance to understand the key predictors of revenue.

Chapter 1

From churn to CLV in Orange Tunisia

Customer life time value might seem slightly challenging to grasp that why we first need to understand the factors influencing CLV and the effective models to predict it. Furthermore, we introduce the CRISP-DM methodology, detailing its phases and how it guided the CLV prediction process.

1.1 Orange Tunisia

Orange is a global telecommunications company, operating in 26 countries around the world. As of December 31, 2023, it has a total number of 298 million individuals globally. In terms of financial performance, the company generated €44.1 billion in revenue.

Moreover Orange Tunisia, this end of study project host, is resulting from an alliance between Orange (49%) and Investec (51%). As the second private telecommunication operator to secure a mobile phone license in Tunisia, it serves more than 4 million customers across almost the entire territory with a capital of 31,335,600 Tunisian Dinars.

Orange's Main Professions and organizational chart

The project took place within the CVM (Customer Value Management) department of the Marketing division. The CVM team is responsible for customer value management. Their aim is to enhance customer satisfaction, strengthen engagement, and maximize their value throughout their journey with Orange.
To achieve this, the CVM department employs various customer relationship management techniques, including data analysis, customer segmentation, offer personalization, marketing campaign management, and loyalty program management. Orange encompasses a variety of other essential roles for its smooth operation and development. These roles fall into several key areas:

1. **Technologies & Innovation:** Covering research, design, information and communication technology, deployment and operation, data management, and security.

2. **Client:** Involving marketing, client coordination, sales, and customer relations, especially in the enterprise segment.

3. **Support Functions:** Including areas such as finance, performance, strategy, human resources, communication, real estate and occupancy services, as well as protection.

4. **Accompaniment (to the company's evolution):** Including management, project management, animation, and processes.

Each of these roles plays a crucial part in the overall success of Orange. The Technologies & Innovation teams focus on innovative solutions and infrastructure management.
The Client teams are responsible for promoting offers, coordinating sales activities, and managing customer relations. Support Functions aid the company's operations, from financial management to communication, human resources, real estate, and beyond.
Accompanying roles ensure the effective implementation of changes necessary for Orange's continued growth and success. These encompass management, project management, animation and process optimisation. In essence, each profession within Orange plays a pivotal role in steering the company towards continued prosperity and performance in the dynamic telecommunications industry, see Figure 1.2.

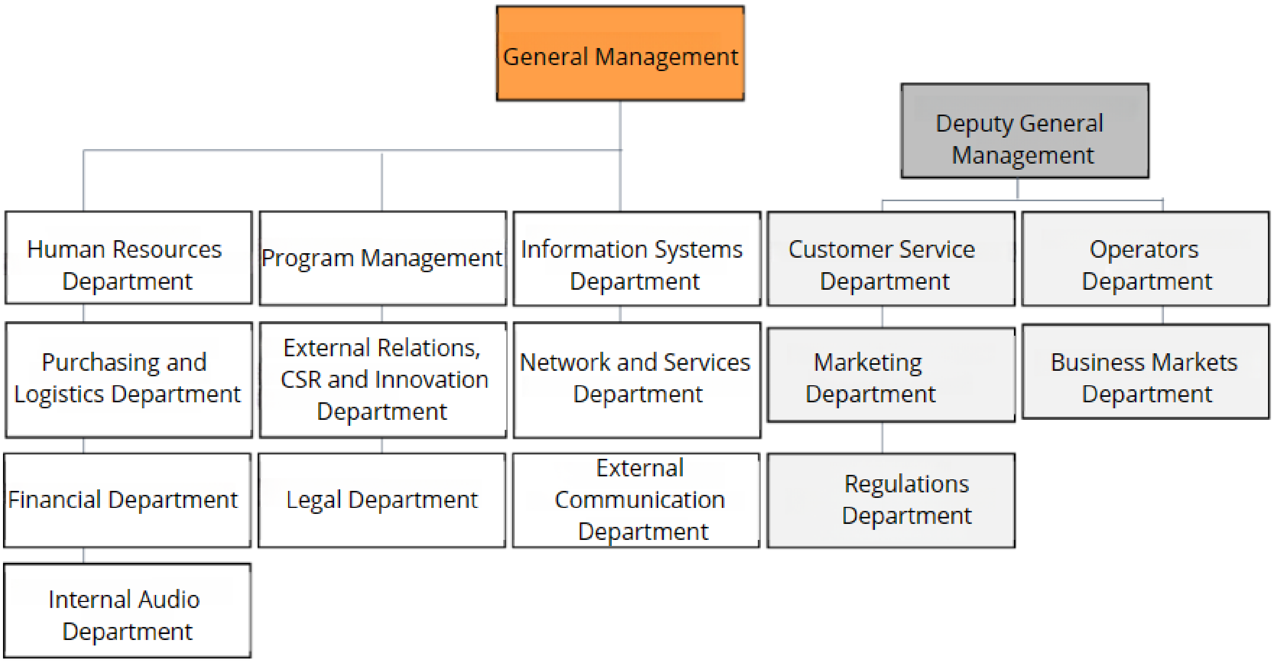

General Management
Deputy General Management
Human Resources Department
Program Management
Information Systems Department
Customer Service Department
Operators Department
Purchasing and Logistics Department
External Relations, CSR and Innovation Department
Network and Services Department
Marketing Department
Business Markets Department
Financial Department
Legal Department
External Communication Department
Regulations Department
Internal Audio Department

Figure 1.2: Organizational chart of the company

1.2 Problematic, Objective and Methodology

Problematic Customer retention is a top priority for Orange, and understanding customer lifetime value (CLV) is essential for optimizing customer relationship management (CRM) [12]. By calculating CLV, Orange can identify its most profitable customers and implement targeted marketing strategies to retain them in the long term. In this context, Orange's marketing department wants to leverage customer data to analyze their lifetime value and optimize marketing actions accordingly.

In the following, we will explain in detail the steps to be followed for the prediction of CLV, in order to identify and retain high-value customers. By analyzing customer behavior and purchasing patterns, we want to develop predictive models to estimate the future revenue a customer is likely to generate for the company.

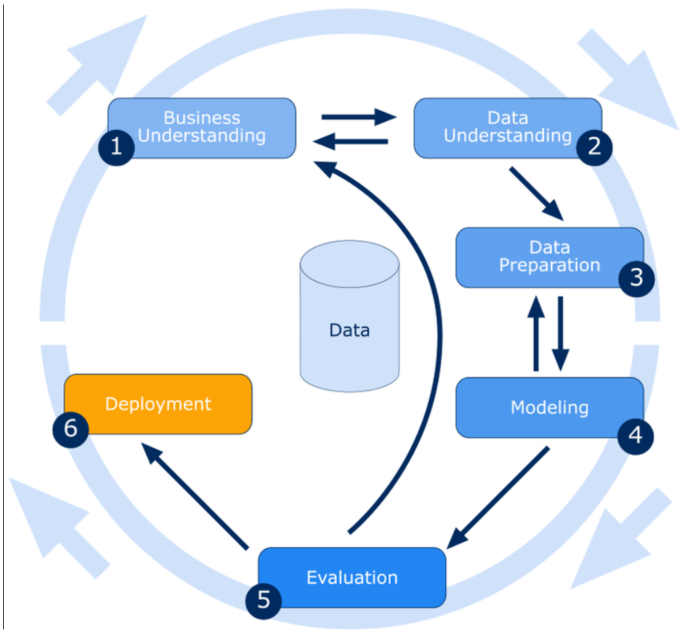


Figure 1.3: Cross-Industry Standard Process for Data Mining (CRISP-DM)

CRISP-DM, known as the Cross-Industry Standard Process for Data Mining, provides a structured roadmap for managing data mining and data science projects [14]. It ensures a systematic approach while allowing flexibility for adaptation to specific project needs. This methodology comprises six phases (Figure 1.3):

- Business Understanding: Establish clear project objectives and requirements from a business perspective, developing a detailed project plan.
- Data Understanding: Collect and comprehend data to determine its suitability for project goals, exploring data distributions and relationships.
- Data Preparation: Get the data ready for modeling by selecting relevant subsets, cleaning, and formatting it to suit modeling tools.
- Modeling: Apply various modeling techniques to build models that meet project objectives, evaluating and refining them as needed.
- Evaluation: Thoroughly assess models to ensure they meet business objectives, reviewing the modeling process and deciding on deployment.
- Deployment: Implement models into production to deliver business benefits, developing a deployment plan and monitoring model performance.

The advantage of CRISP-DM lies in its non-linear process, allowing for iteration and revisiting previous steps as needed. This flexibility enables adaptation to various projects and evolving requirements, ensuring successful outcomes.

Therefore to achieve accurate customer lifetime value (CLV) prediction, we followed this proccess : First, we gathered and prepared customer data relevant to CLV, containing activity history, engagement metrics, churn indicators, and timestamps. Next, we delved into exploratory data analysis (EDA) to understand the distribution of variables, identify patterns and relationships between them and address any outliers or missing values. This analysis lays the groundwork for feature engineering, where we created new and more predictive features from the existing data. Following that, with a clear understanding of the data and the CLV prediction objective, we fitted the following models :

- Pareto Model: Assumes a power-law distribution between customer value and frequency, suitable for identifying top-tier customers [8].
- Beta-Geometric Model: Accounts for customer acquisition and retention rates, useful for predicting long-term CLV [1].
- Gamma-Gamma Model: Incorporates customer purchase frequency and average transaction value, effective for modeling customer spending behavior [6].
- Machine Learning Models (e.g., XGBoost): Employ machine learning algorithms to capture complex non-linear relationships in the data and improve prediction accuracy [3].
- Survival Models (e.g., Cox Proportional Hazards, AFT, Kaplan-Meier): Utilize survival analysis techniques to model the time-to-event (churn) data and estimate customer retention probabilities [10].

After training each selected model on the prepared data, we evaluated its performance using metrics like c-index, R-squared, and root mean squared error [16]. This allowed us to choose the best-performing model for predicting CLV.

By modeling the CLV, Orange can obtain a range of strategic advantages. First, it allows the company to identify the high-value customers, enabling targeted marketing efforts to retain these crucial segments. By tailoring campaigns to specific customer groups based on their predicted CLV, Orange can maximize campaign effectiveness and return on investment (ROI). Furthermore, proactive identification of churn-risk customers becomes possible, allowing for the implementation of targeted retention strategies. This not only reduces churn rates but also increases overall customer lifetime value. Additionally, CLV insights can refine customer segmentation, paving the way for more personalized and effective customer engagement strategies. Ultimately, by focusing on high-value customers and optimizing marketing efforts, Orange can achieve significant revenue growth and enhance overall profitability.
Now that the context is set, let's move on to the key questions and challenges to address:
- What factors influence customer lifetime value (CLV)?
- How to build an effective model to predict and optimize the CLV of Orange customers?

1.3 What about CLV Modeling?

As mentioned above, Customer life time value CLV is a metric used in marketing to estimate the total revenue a business can expect from a customer over the entire course of their relationship with the company. Calculating CLV might seem slightly challenging to grasp that's why we need first to understand a few key customer behaviors.

- Customer Lifespan: This metric estimates how long, on average, a customer remains actively engaged with the business. It is calculated by dividing the total number of active customer years by the total number of customers within that period.
- Purchase Frequency: This estimates how Often Do They Buy. To calculate this, simply track the number of times a customer makes a purchase within a specific time-frame.
- Average Purchase Value: This metric reveals the average amount a customer spends on a typical purchase or engagement. Divide the total customer spending over a period by the number of purchases made during that time.
- Average Customer Value: This element combines purchase frequency and average purchase value. It essentially tells you how much a customer contributes to your business, on average, during their active customer period.

1.3.1 Key Components of CLV

To calculate CLV, several key factors come into play: Revenue per Customer, Customer Acquisition and Retention Cost, Customer Retention Rate, Profit Margin, and Discount Rate. These different components are described below:

- Revenue per Customer: The total income a business generates from a single customer over a specific period.
- Customer Acquisition and Retention Cost: The expense incurred by a business to acquire and retain a customer. This includes marketing costs, sales commissions, and any other expenses associated with the customer.
- Probability of survival of the customer: Its the probability that an individual did not leave the company until a specified time
- Profit Margin per customer: The percentage of revenue per customer that remains as profit after deducting all business costs.
- Discount Rate: A factor used to account for the time value of money. Future earnings are worth less than present earnings, so the discount rate adjusts future revenue streams to their present value.

By understanding these individual pieces, we can then use a Equation (2.1) to calculate the total CLV, providing a clearer picture of the revenue a customer can generate throughout their relationship with the company.

1.3.2 Common CLV Models

There are various CLV models, each with its strengths and weaknesses .

- Historical CLV: This is a simple model that relies on past data to calculate the total revenue a customer has generated for the business over a specific period. While easy to implement, it doesn't predict future customer value.
- Predictive CLV: These models leverage statistical methods and machine learning to forecast future revenue from a customer . Here are some common approaches:
  - Cohort Analysis: Groups customers by their acquisition date and tracks their purchase behavior over time. This allows for comparisons between customer groups acquired at different points.
  - Probabilistic Models: These models predict future behavior based on probabilities. Examples include the Pareto/NBD (Negative Binomial Distribution) model and the BG/NBD (Beta Geometric/NBD) model. Both account for customer churn and variations in purchasing frequency among customers.
  - Machine Learning Models: These models use algorithms like regression, decision trees, and neural networks to predict CLV based on historical data and customer characteristics.

Methods used in Orange
Orange currently employs various detection models that primarily focus on churn prediction using Logistic Regression. However, these models do not account for customer lifetime value (CLV) or use survival models, which can limit their effectiveness. By concentrating solely on churn prediction, Orange might not efficiently detect risks associated with new customers, who have different profiles and histories compared to existing customers.

1.4 Data description and feature engineering

We were provided with a total of five datasets four of each consisting of approximately 200,927 customers and containing 121 features.
It's important to note that the customers are consistent across all datasets, meaning that the same set of cus- tomers is present in each dataset. These datasets cover a time span of 24 months, 6 months each, starting from January 2022 and ending in January 2024. The time dimension is represented in the columns for each dataset there are 6 columns representing each feature, for example the first column v1_jan22 represents the feature v1 for the month of January of 2022. The fifth dataset has 200,927 and one column presenting the life time which refers to seniority of the customer in the company. Table 1.1 details the type and meaning of every variable.

Table 1.1: Description of Variables

Variable	Type	Description
msisdn	String	Mobile Station International Subscriber Directory Number
v1	Categorical	1 = the client churned before and returned within that one
v2	Categorical	1 = the client did not make any activity for the whole month
v3	Categorical	1 = The client uses the application My Orange
v4	Numeric	Incoming calls duration in minutes
v5	Numeric	Dialed calls duration in minutes
v66	Numeric	Number of incoming contacts
v77	Numeric	Number of outgoing contacts
v88	Numeric	Total data usage in kilobyte
v22	Numeric	Total top-up week 1 in TND
v23	Numeric	Total top-up week 2 in TND
v24	Numeric	Total top-up week 3 in TND
v25	Numeric	Total top-up week 4 in TND
v26	Numeric	Total top-up remaining in TND
v27	Numeric	Frequency of top-up week 1
v28	Numeric	Frequency of top-up week 2
v29	Numeric	Frequency of top-up week 3
v30	Numeric	Frequency of top-up week 4
v31	Numeric	Number of top-ups remaining
v32	Numeric	Total revenue in TND
v33	Numeric	Total voice revenue in TND
life_time	Numeric	Seniority of customer in the Orange company in months

For feature engineering we had to change the format of our data in order to manipulate it effectively. first, we separated the time dimension from the features by creating a single column representing the date. Then, we merged all datasets based on lines and sorted them by MSISDN and date. This resulted in a new dataset containing 5 million entries and 21 columns.
Secondly, troughout the entire steps of the project we always needed metrics to calculate. Here is a breakdown of all the metrics we have needed along the way, see Table 1.2.

Table 1.2: Description of the new features

Feature	Type	Description
`actualactivation`	Numeric	Number of months of activation during the customer survival
`activation`	Numeric	Number of non churn
`Days_Difference`	Numeric	Total period of survival in days before churning within the study period
`churn`	Categorical	Binary indicator of churn
`nbchurn2`	Categorical	Classes of churning rate
`nombre_recharge_value_sum`	Numeric	Total number of top-ups during the customer survival
`numberofchurn`	Numeric	Number of times a customer has churned
`Recency`	Numeric	Measures how recently a customer made a purchase in days
`sales_value_last_2weeks`	Numeric	Amount of revenue generated in the last two weeks in TND
`sales_value_sum`	Numeric	Total revenue generated during the customer survival
`v31_value_sum`	Numeric	Number of top-ups remaining during the customer survival
`v33_value_sum`	Numeric	Total voice revenue in TND during the customer survival
`v4_value_sum`	Numeric	Incoming calls duration in minutes during the customer survival
`v6`	Numeric	Number of incoming calls during the customer survival
`v7`	Numeric	Total number of outgoing calls
`v8`	Numeric	Total data usage in kilo octet

All of these features were added to a new dataset called **data1** consisting of 200 thousand entries and 16 columns.

1.5 Conclusion

It is essential to develop a specific model that incorporates the unique characteristics of new customers and utilizes survival models to estimate CLV. This approach would better identify the specific risk signals associated with these customers and enhance Orange's ability to anticipate potential issues. By leveraging survival models, Orange could implement more tailored retention and satisfaction strategies, thereby addressing the current gap and optimizing the management of risks related to new customers.

Chapter 2

Theoretical aspects

After understanding the general concept of CLV it is time to explore the theoretical aspects of it. We will delve into the critical parameters that influence CLV (survival probability, revenue generation, discounting rate...). Additionally, we will discuss various survival models used to estimate survival probability (Kaplan-Meier, Cox proportional hazards, and Accelerated Failure Time models...). Further, we will examine different modeling techniques for predicting CLV, such as the Pareto/NBD, BG/NBD, and Gamma-Gamma models, highlighting their assumptions, mathematical formulations, and practical applications in customer value management.

2.1 Customer Life Time value

As stated before, the CLV is the discounted (present) total Expected Revenue generated by a given customer for a period of time[1]. Three important parameters are to be considered: The survival probability until a period t, the revenue generated, and the discounting rate, δ. Formally, we can define the CLV as follows:

$$\text{CLV}(T) = \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{P_{it}\, R_{it}}{(1+\delta)^t}\ , \tag{2.1}$$

where N is the total number of customers, P_{it} is the probability of survival of the customer i until period t, R_{it} is the revenue generated by the customer i for the preriod t, and δ the discounting rate, usually a given parameter (i.e. the policy rate).

Hence, to do the assessment of the CLV we need to address two major modeling:

- The survivals model to predict the survival probability.
- The generalized models to predict the revenue.

In case where we assume that the revenue is homogeneous over time and customers ($R_{it} = R_0$), and the survival probability is homogeneous for the customers ($P_{it} = P_t$), the CLV(T) is resumed to:

$$\text{CLV}(T) = N \cdot R_0 \sum_{t=1}^{T} \frac{P_t}{(1+\delta)^t} \tag{2.2}$$

[1] It's worth noting that the expected net revenue should be taken into account, whenever the total cost of Acquisition and Retention of a customer: for a period of time t, is provided. Otherwise, we can use the total operating cost as a proxy.

Equation (2.2) states that, under the assumption of homogeneity between customers, the CLV for the company is equal to the total revenue generated at the study period ($N \cdot R_0$) times the total discounted present value of the survival probability.

2.2 Survival models

Now to estimate the survival probability of a customer we will use three suvival models (KM, Cox and AFT). Kaplan-Meier (KM) estimates event probability at a given time from time-to-event data. whereas Cox estimates the hazard function considering explanatory variables' effects. Accelerated Failure Time model (AFT) focuses on survival times, assuming they depend on explanatory variables. We'll delve deeper into each model in the following section.

2.2.1 Kaplan-Meier: Non parametric survival curve

Kaplan-Meier Survival curve is the a non parametric curve where for each period the survival probability is defined as the number of individuals alive (in our case active) divided by the number of individuals at risk [15]. More formally, let's consider the duration ordered in ascending order: $t_1 < t_2 < \cdots < T_{\max}$. The estimator of the survival probability at time t_i is defined as:

$$\begin{aligned} \hat{S}\left(t_i\right) &= \prod_{j=1}^{i} \hat{P}\left(T > t_j | T \geq t_j\right) && (2.3) \\ &= \prod_{j=1}^{i} \left(\frac{N\left(t_j\right) - D\left(t_j\right)}{N\left(t_j\right)}\right), && (2.4) \end{aligned}$$

where:

- $N\left(t_j\right)$ is the number of individuals at risk just before t_j;
- $D\left(t_j\right)$ is the number of events (i.e. churns) occurring at time t_j;
- $S(t) = 1$ if $t < t_1$.

The log-rank test: To compare between categories of individuals (clients), the log-rank test is a commonly used statistic test. For that, let us consider G groups of individuals. For each period we observe the following numbers:

- $N_g\left(t_j\right)$ the number of individuals at risk before t_j of the group g, $g = 1, 2, \cdots, G$.
- $D_g\left(t_j\right)$ the total number of events (i.e. Churn.) at date t_j in the group g, $(g = 1, 2, \cdots G)$.
- Under the assumption of independence between groups, the number of estimated events at date t_j is given by:

$$e_g\left(t_j\right) = \frac{N_g\left(t_i\right)}{\sum_{k=1}^{G} N_k\left(t_i\right)} \times \sum_{k=1}^{G} D_k\left(t_i\right). \qquad (2.5)$$

The statistics of comparison between all the G groups is given by the log-rang statistics:

$$\log - \text{rank} = \sum_{g=1}^{G} \frac{\left(O_g - E_g\right)^2}{E_g} \longrightarrow \chi^2(G-1) \qquad (2.6)$$

where O_g total number of events in the groups g and $E_g = \sum_{t_i} e_g\left(t_i\right)$.

2.2.2 Cox's proportional hazards model (Cox regression)

We are interested in finding out how long it will take for an event, like churn in our case, to happen. The Cox proportional model is a parametric model that estimates the relation between length of survival and a single or multiple predictor variable(s) [9]. It is already integrated in almost data software packages and allows researchers and analysts to use it to perform survival analysis. Mathematically, the Cox model can be stated as:

$$h(t|x) \quad = \quad h_0(t) \times \exp(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p), \tag{2.7}$$

where:

- $h(t|x)$ is the hazard function at time t for an individual with covariate values x.
- $h_0(t)$ is the baseline hazard function, which represents the hazard at time t for an individual with all covariates set to zero.
- $\beta_1, \beta_2, \ldots, \beta_p$ are the regression coefficients which represent the effects of predictor variables $x_1, x_2, \ldots, x_p$.

The proportionality property of the Cox Model allows analyst to easily evaluate the hazard ratios (HR) between groups or between individuals. Consider two individual with different covariate values, denoted by x_1 and x_2. Their, respective, hazard functions can be expressed as:

$$h(t, x_1) = h_0(t)\exp(\beta x_1)$$
$$h(t, x_2) = h_0(t)\exp(\beta x_2)$$

The core principle of proportionality lies in the fact that the ratio of these hazard functions remains constant across time t:

$$\begin{aligned}\frac{h(t, x_1)}{h(t, x_2)} &= \frac{h_0(t)\exp(\beta x_1)}{h_0(t)\exp(\beta x_2)} \\ &= \exp(\beta x_1)\exp(-\beta x_2) \\ &= \exp(\beta(x_1 - x_2))\end{aligned}$$

This ratio solely depends on the difference between the covariate values $(x_1 - x_2)$ and the coefficient β, independent of the specific time point t. This signifies that the influence of a covariate on the hazard ratio remains constant throughout the entire follow-up period. The Cox model uses partial probability to estimate regression coefficients (β), allowing hazard ratios to be modeled directly without specifying the hazard ratio. This makes Cox regression particularly useful when the hazard hypothesis is relevant but the underlying hazard function is unknown or difficult to model.

2.2.3 Accelerated Failure Time (AFT) Model

The AFT model relates the survival time T of a subject to a linear function of covariates $\mathbf{x}_i$ and an error term ε_i (see [17]). It assumes a specific parametric distribution for the survival time, such as Weibull or Log-normal. This allows for direct modeling of the survival time itself. The log-normal ,odel is defined as follows:

$$\log(T_i) = \beta_0 + \boldsymbol{\beta}^T \mathbf{x}_i + \varepsilon_i \tag{2.8}$$

where:

- T_i is the survival time for subject i.
- β_0 is the intercept term.
- $\boldsymbol{\beta}$ is the vector of regression coefficients for covariates.
- $\mathbf{x}_i$: Vector of covariate values for subject i.
- ε_i: Error term following a normal distribution.

The AFT models estimate the effect of covariates on the logarithm of the survival time. This translates to a multiplicative effect on the actual survival time. A one-unit **increase** in a covariate with a positive coefficient β_j leads to a proportional **increase** in the survival time (assuming positive survival times)[2]. Obviously, survival data are characterised by the censorship property. Indeed, taking into account censored data, the likelihood function is defined by:

$$f(t_1, t_2, \cdots, t_N) = \prod_{uncensored} f(t_i|x_i, \theta) \cdot \prod_{censored} S(t_i|x_i, \theta)$$

where $f()$ and $S()$ are respectively the density and the survival functions. In the case of log-normal model the likelihood is defined by:

$$f(t_1, t_2, \cdots, t_N) = \prod_{uncensored} \phi(t_i|x_i, \theta) \cdot \prod_{censored} (1 - \Phi(t_i|x_i, \theta)). \tag{2.9}$$

where $\phi()$ stands for the normal density for the event cases and $\Phi()$ stands for the distribution (cumulative probability) for the censored cases. For the AFT model we have to be careful when we interpret the effect of variable on the survival time. For continuous explanatory variable, The coefficient gives an estimate effect of an extra-unit of this variable on the expected logarithm of the survival time. Hence comparison between individuals and log ratio of the expected survival times:

$$E\left(log(t_1) - log(t_2)\right) = E\left(log(t_1|t_2) = \beta^T(x_1 - x_2).\right. \tag{2.10}$$

In the case of categorical variable the exponential of coefficients can be interpreted directly as ratio of the geometric means between a group and the reference group. Let's take the example of a bivariate categorical variable ($Group_1$ and $Group_2$), the model is expressed as follows:

$$\log(T_i) = \beta_0 + \boldsymbol{\beta}^T \mathbf{x}_i + \theta\ Group_2 + \varepsilon_i,$$

and

$$E(\log(T_i|Group_2) - E(\log(T_i|Group_1) = \theta$$

and $\exp(\theta)$ is the ratio of the geometric mean of the survival time between $Group_1$ and $Group_2$.

[2] Different AFT models correspond to different underlying survival time distributions, each with its own interpretation of the coefficients. It depends on the assumption of the error term ε.

2.2.4 Comparison of AFT and Cox Models

The Accelerated Failure Time (AFT) model and the Cox proportional hazards model are tools for analyzing survival data but they differ in their approach and underlying assumptions. Here, we see their key features and compare them

Choosing the right model for survival analysis is crucial. In order to state between the two models, Cox model and AFT, the Table 2.1 provides a brief comparison[3].

Table 2.1: Comparing AFT and COX Models

Feature	**AFT Model**	**Cox Model**
`Outcome`	Survival Time	Hazard Function
`Parametric Assumption`	Yes (specific distribution)	No
`Proportionality Assumption`	Not required	Required
`Interpretation`	Multiplicative effect on survival time	Hazard Ratio
`Application`	time to event prediction, understanding covariate effects on survival time	Identifying factors influencing event risk

AFT models directly analyze survival time, assuming a specific distribution, while Cox PH models focus on the hazard function, offering more flexibility. AFT results show how covariates impact survival time, while Cox PH results indicate how they influence event risk.

2.3 BG/NBD Model: A Computationally Efficient Alternative

The Beta Geometric/Negative Binomial Bistribution (BG/NBD) model, introduced by Fader, Hardie, and Lee (2005), offers an alternative to the Pareto/NBD model. While capturing purchase behavior with similar principles, the BG/NBD model utilizes different mathematical structures, leading to computational advantages.

Core Assumptions:

Purchase Rate: Customer purchase rate follows a Beta distribution with shape parameters α and β. The beta distribution allows for flexible modeling of purchase rates, ranging from low to high frequencies. The pobability density function for the purchase rate (λ) is:

$$f(\lambda|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \lambda^{\alpha-1}(1-\lambda)^{\beta-1}, \quad 0 < \lambda < 1 \tag{2.11}$$

[3] For a detailed comparison between AFT and Cox model refer to [18]

Customer Churn: Customer Churn: Similar to the Pareto/NBD model, customer churn follows a **negative binomial distribution** with parameters r and p.
While both models provide valuable tools for CLV analysis, the BG/NBD model's computational efficiency can be advantageous in situations with large datasets or real-time applications.

2.4 XGBoost Regressor and XGBoost Classifier

Following the origin author [2], the Extreme Gradient Boosting known as XGBoost builds a predictive model by combining the predictions of multiple individual models in an iterative manner. It can be stated as follows. For a given data set with n examples and m features

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\} \left(|\mathcal{D}| = n, \mathbf{x}_i \in \mathbb{R}^m, y_i \in \mathbb{R}\right) \tag{2.12}$$

a tree ensemble model uses K additive functions to predict the output.

$$\hat{y}_i = \phi\left(\mathbf{x}_i\right) = \sum_{k=1}^{K} f_k\left(\mathbf{x}_i\right), \quad f_k \in \mathcal{F},$$

where

- $\mathcal{F} = \left\{f(\mathbf{x}) = w_{q(\mathbf{x})}\right\} \left(q : \mathbb{R}^m \to T, w \in \mathbb{R}^T\right)$ is the space of regression trees.
- q represents the structure of each tree that maps an example to the corresponding leaf index.
- T is the number of leaves in the tree.
- f_k corresponds to an independent tree structure q and leaf weights w. Each regression tree contains a continuous score on each of the leaf, we use w_i to represent score on i-th leaf.

To find optimal prediction of the target variable, we minimize the following regularized objective.

$$\mathcal{L}(\phi) = \sum_i l\left(\hat{y}_i, y_i\right) + \sum_k \Omega\left(f_k\right)$$
$$\text{where } \Omega(f) = \gamma T + \frac{1}{2}\lambda\|w\|^2$$

Here l is a differentiable convex loss function that measures the difference between the prediction $\hat{y}_i$ and the target y_i. The second term Ω penalizes the complexity of the model (i.e., the regression tree functions). The additional regularization (λ) term helps to smooth the final learnt weights to avoid over-fitting. Two concepts make XGBoost the most used algorithm : Bagging and Boosting. Bagging helps reduce the variance of the predictors by aggregating the results of training the dataset over the trees. The final result corresponds to the aggregation (i.e. average). The Boosting is the fact that predictor are constructed sequentially. Indeed, each predictor takes the previous predictor and reduce the error terms. Notice that two version of XGBoost are used : XGBosst regressor for the continuous outcome variables and XGBoost classifier for the categorical outcome variables. And the main metrics used are the Root Mean Square Error (RMSE) or the Mean Absolute Error (MAE), Mean Absolute Percent Error (MAPE), Negative Log-Likelikhood (logloss) for XGBoost regressors and the Error (in binary classifier), MError (Multi classifier), Area Under Curve (AUC), the precision (Prec), the Mean Average Precion (MAP) for the XGBoost Classifier.

2.5 Gamma Gamma Model

Following the leading idea of the purchase frequency models (Pareto/NBD & BG/NBD), the gamma-gamma model incorporates timing (between purchases) for a more comprehensive CLV analysis. It uses beta and gamma distributions to capture heterogeneity in both frequency and timing patterns. It leads leading to potentially more accurate CLV predictions. This is particularly valuable for customers with irregular purchase behavior. Nevertheless. the gamma-gamma model is computationally more complex to implement.

2.5.1 Model Assumptions

Transaction Value: The model assumes individual transaction values (z_i) follow a gamma distribution with parameters p (shape) and ν (scale). This means the distribution of transaction amounts can be skewed and positive-valued.

$$z_i \sim \text{gamma}(p, \nu) \tag{2.13}$$

Here, $E(z_i|p,\nu) = \zeta = p|\nu$ represents the expected value (average) of a single transaction.
Number of Transactions: The model further assumes that the number of transactions a customer makes (x) is independent of the transaction values.
Total Spend: Based on the properties of the gamma distribution, the total spend across x transactions ($\sum z_i$) can be modeled as another gamma distribution with adjusted parameters px and ν. This reflects the idea that customers with more transactions tend to have higher total spend.

$$\sum z_i \sim \text{gamma}(px, \nu) \tag{2.14}$$

Heterogeneity in Spending: The model acknowledges that customers exhibit varying spending patterns. To capture this, it introduces another gamma distribution to model the variation in the average transaction value (ζ) across customers. This distribution has parameters q (shape) and γ (scale).

$$\nu \sim \text{gamma}(q, \gamma) \tag{2.15}$$

2.5.2 Combining the Elements: The Final Model

By combining these elements, we arrive at the core expression for the Gamma-Gamma model, representing the conditional expectation (`average`) expenditure for a customer who has made x transactions:

$$E(z|x,p,q,\gamma,\nu) \;=\; \left(\frac{\gamma + xE(z)}{\nu}\right)^{p+q} \cdot \exp\left(-\nu(\gamma + xE(z))\right) \cdot \frac{\Gamma(px+q)}{\Gamma(p)\cdot\Gamma(q)} \tag{2.16}$$

$\Gamma(.)$ represents the Gamma function, essential for calculations in the model [4].

2.6 Conclusion

This chapter explored the diverse landscape of Customer Lifetime Value (CLV) prediction models. We delved into the key concepts, benefits, and drawbacks of various approaches, along with their underlying mathematical foundations.

Survival analysis models, like the Kaplan-Meier estimator, Cox proportional hazards, and Accelerated Failure Time (AFT) models, provided a solid foundation for understanding customer churn (customer loss) and its timing. These models shed light on critical factors influencing customer longevity.

Machine learning, like XGBoost, a tree-based model, demonstrated its power in handling complex relationships and large datasets. XGBoost's strength lies in its ability to capture non-linear effects on CLV, leading to potentially more accurate predictions.

Probabilistic models, such as the Gamma-Gamma and Beta Geometric (BG/NBD) models, offered a structured framework for analyzing customer purchase behavior patterns, considering factors like purchase frequency and monetary value.

By understanding the strengths and weaknesses of these diverse approaches, businesses can leverage CLV prediction models to make informed decisions regarding customer acquisition, retention, and resource allocation strategies

Chapter 3

Customer segmentation via RFM classification

In the telecom industry, RFM (Recency, Frequency, Monetary) classification is used to segment customers based on their interaction and spending behavior. Recency measures how recently a customer has used Telecom services, with recent users being more engaged. Frequency tracks how often customers use services like calls, texts, and data, indicating their engagement level. Monetary assesses the amount spent on Telecom services, identifying high-value customers. By analyzing these metrics, Telecom companies can categorize customers into different segments and tailor marketing strategies, offers, and retention efforts to enhance customer satisfaction and maximize revenue.
In order to do this classification, we undertake the following steps:

1. We evaluate the necessary RFM metrics: The Recency, The Frequency and the Amount of Money.

2. We normalize the calculated metrics to a common scale.

3. We sort in descending order the Frequency and Monetary values and in ascending order The Recency.

4. To build the RFM segments, we did two different treatments:

   - First, we created an overall RFM score by applying different weights to these metrics (see below), and we ranked customers by their RFM scores to identify and target the most valuable segments.
   - Second, we binned the Recency, Frequency, and Monetary values into quantiles. Each customer received a score from 1 to 5 for recency, frequency, and monetary value based on the quantile binning.

3.1 First RFM classification: Weighted sum score

The RFM Classification gives more importance to higher spenders and frequent buyers by allowing high weights. And, It accords lower importance to the recency variable with lower weight. The RFM score is a weighted sum of the three variables : monetary, frequency and recency. It's defined as follows:

$$\texttt{RFM\_Score} = 0.13 \times \text{Recency} + 0.29 \times \text{Frequency} + 0.58 \times \text{Monetary}. \tag{3.1}$$

The assignment (or labelling) of the customer segments (or clusters) is based on the following criteria:

1. **Top Customers:** `RFM_Score` greater than 4.5

2. **High Value Customer:** `RFM_Score` between 4 and 4.5

3. **Medium Value Customer:** `RFM_Score` between 3 and 4

4. **Low Value Customers:** `RFM_Score` between 1.6 and 3

5. **Lost Customers:** `RFM_Score` less than or equal to 1.6

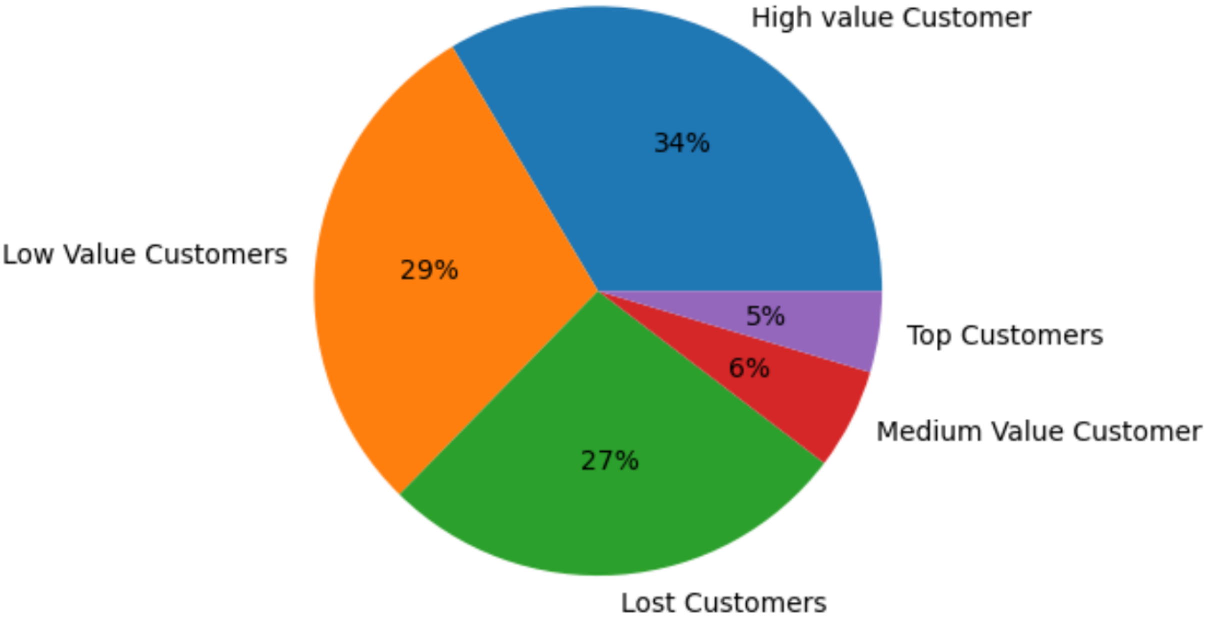


Figure 3.1: Distribution of segments

As shown in Figure 3.1, high value customers represent 34% of the total customers followed by the low value customers (29%). The segment of ”Lost customers” covers 27% of the customers. Medium customers and Top customers segments represent, respectively, 6% and 5%.

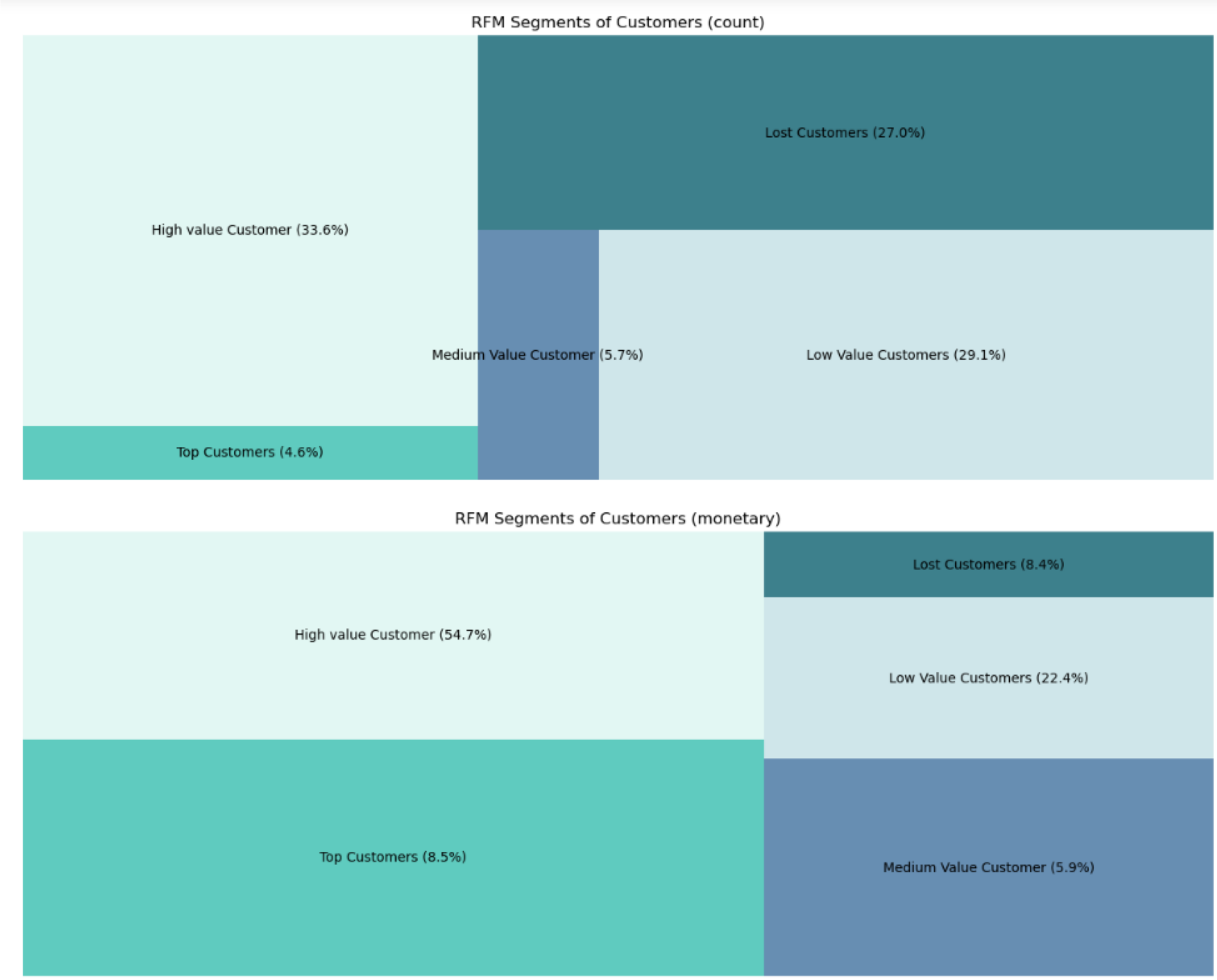


Figure 3.2: Customer Lifetime Value: Distribution by Segments (Count & Monetary)

The heatmap of the Monetary and Count (see figure 3.2) shows that ”High value Customers” represents 33% of the studied population and generates 54.7% of the total revenue and could be classified as the most important segment monetary wise. And the company should allow high importance to. ”Low value customers” accounts for 29% of population and accounts for 22.4% of the total generated revenue. ”Top customers” occupy the third place with 4.6% of the customers and 8.5% of the total revenue. ”Lost customers” represent 27% of the population and 8.4% of the revenue. ”Medium value customers” occupies the last position with 5.7% of the population and 5.9% of the generated total revenue.

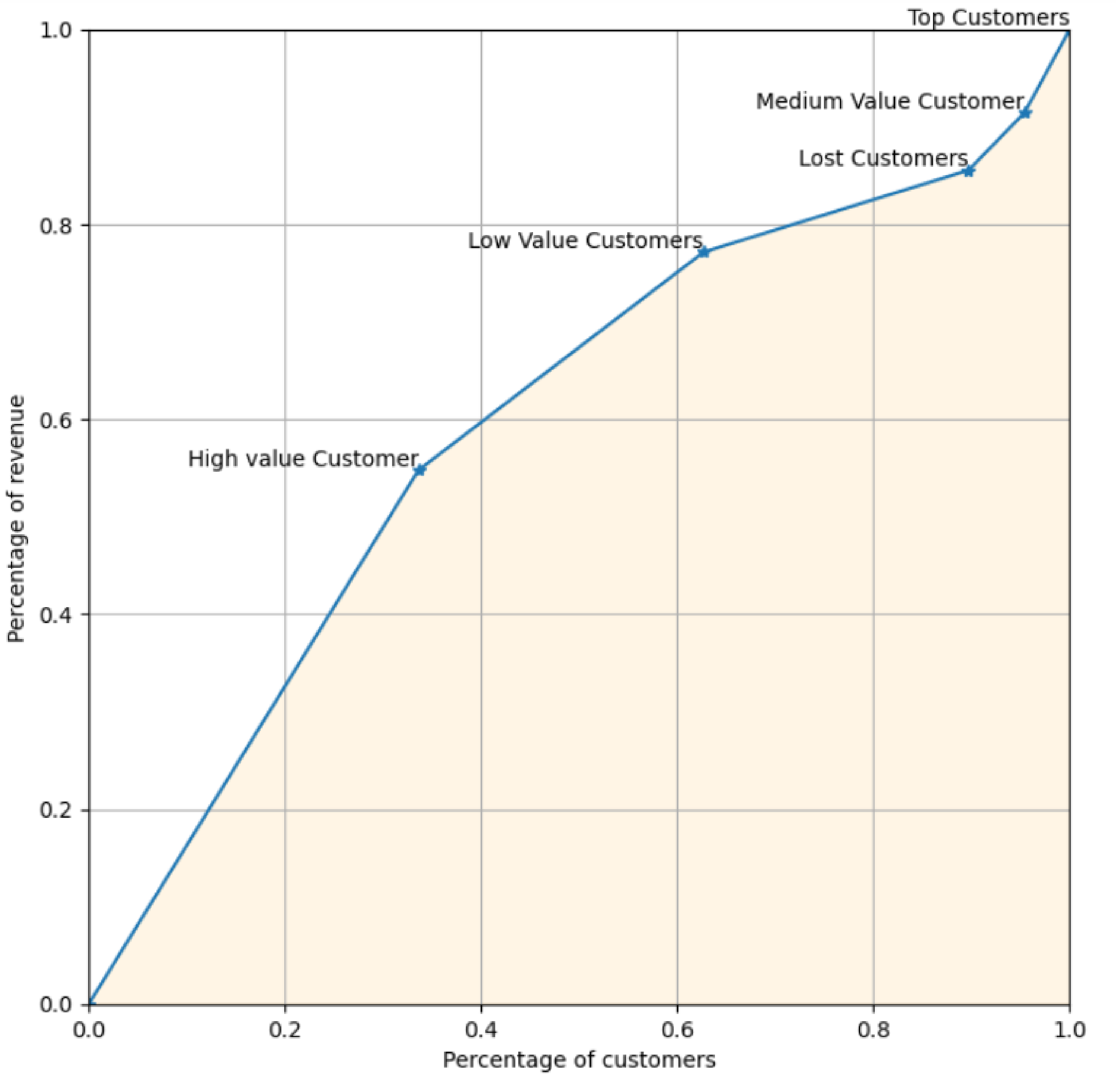


Figure 3.3: Segment contribution to revenue

Figure 3.3 is the representation of the cumulative population vs cumulative contribution to the revenue exhibits a well know concept of Pareto distribution 80:20; usually admitted in marketing studies. This representation show that we are able to say that the two segments ”High value customers” and ”Low value customers” accounts 62% of the customers and generates about 80% of the revenue. Adding ”top customers” and ”Medium value customers” shows that we can reach 91.6% of the total revenue. ”Lost customers” which represent 27% of the customers but with 8.4% of potential additional revenue. We recommend that company should concentrate on the previous announced segments and allow importance to the ”Lost customers” if the potential additional revenue far exceeds the invested cost.

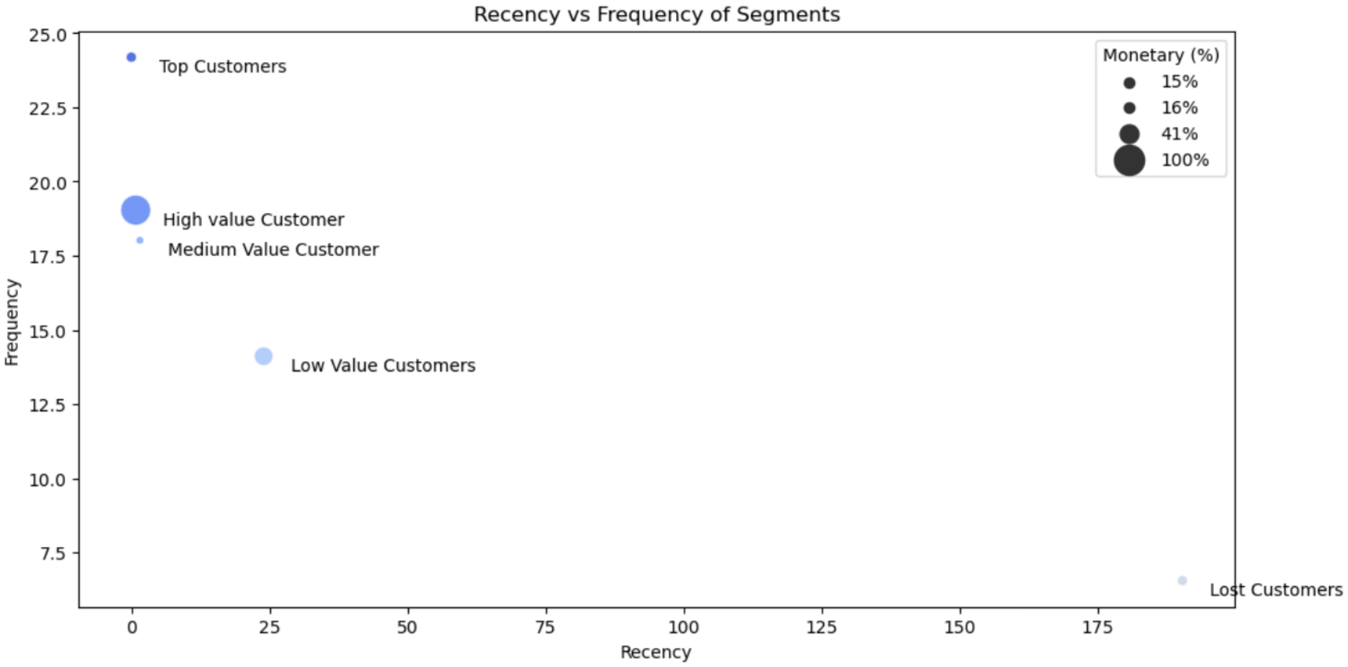


Figure 3.4: Behaviour and Spending analysis

The scatter plot in Figure 3.4 shows that Top Customers are located at the top left of the plot, indicating high frequency and low recency values, meaning they engage frequently and have made recent activities. Their larger bubble size shows significant monetary contributions. High Value Customers, positioned near the top with moderate recency and high frequency, engage frequently but not as recently as top customers, and also have high monetary contributions. Medium Value Customers are similar to high value customers but with slightly less recency and frequency, contributing a moderate amount monetarily. Low Value Customers, positioned in the middle with moderate recency and lower frequency, engage occasionally but not frequently, with low monetary contributions as shown by smaller bubble size. Lost Customers, located at the bottom right with high recency and very low frequency, haven't made a engagement in a long time and rarely did even when they were active, contributing very little monetarily.

Considering the evidence presented, we can ultimately conclude:

Top Customers: These are the most valuable customers with recent and frequent service engagement.

High Value Customers: Similar to Top Customers with high monetary value , but with less frequent service engagement.

Medium Value Customers: Moderate monetary value and moderate service engagement frequency.

Low Value Customers: Low monetary value and infrequent service engagement.

Lost Customers: No service engagement in over a year.

While this approach helps us understand customers through their spending it doesn't really help us grasp their practices so we may fail understanding their behaviour like loyalty and engagement frequency .

3.2 Second RFM classification

For the classification procedure we binned the Recency, Frequency, and Monetary values into quantiles, labeling Recency bins as [5, 4, 3, 2, 1] setting 5 for the most recent activities and 1 for the least recent, and Frequency and Monetary bins as [1, 2, 3, 4, 5] where we assigned 5 for the highest values (fifth quantile) and 1 for the lowest (first quantile). This method was then applied to assign each customer a `r_score`, `f_score`, and `m_score` from 1 to 5 based on their respective quantiles.

1. **Champions:** Customers with `r_score of 5`, `m_score` and `f_score` between 4 and 5
2. **Loyal Customers**: Customers with `r_score` between 3 and 4, **m_score** and `f_score` between 4 and 5.
3. **Potential Loyalists:** Customers with `r_score` between 4 and 5, `m_score` and `f_score` between 2 and 3.
4. **New Customers:** Customers with `r_score of 5`, `m_score` and `f_score of 1`.
5. **Promising:** Customers with `r_score of 4`, `m_score` and `f_score` of 1.
6. **Needing Attention:** Customers with `r_score of 3`, `m_score` and `f_score` of 3.
7. **About to Sleep:** Customers with `r_score` of 3, `m_score` and `f_score` between 1 and 2.

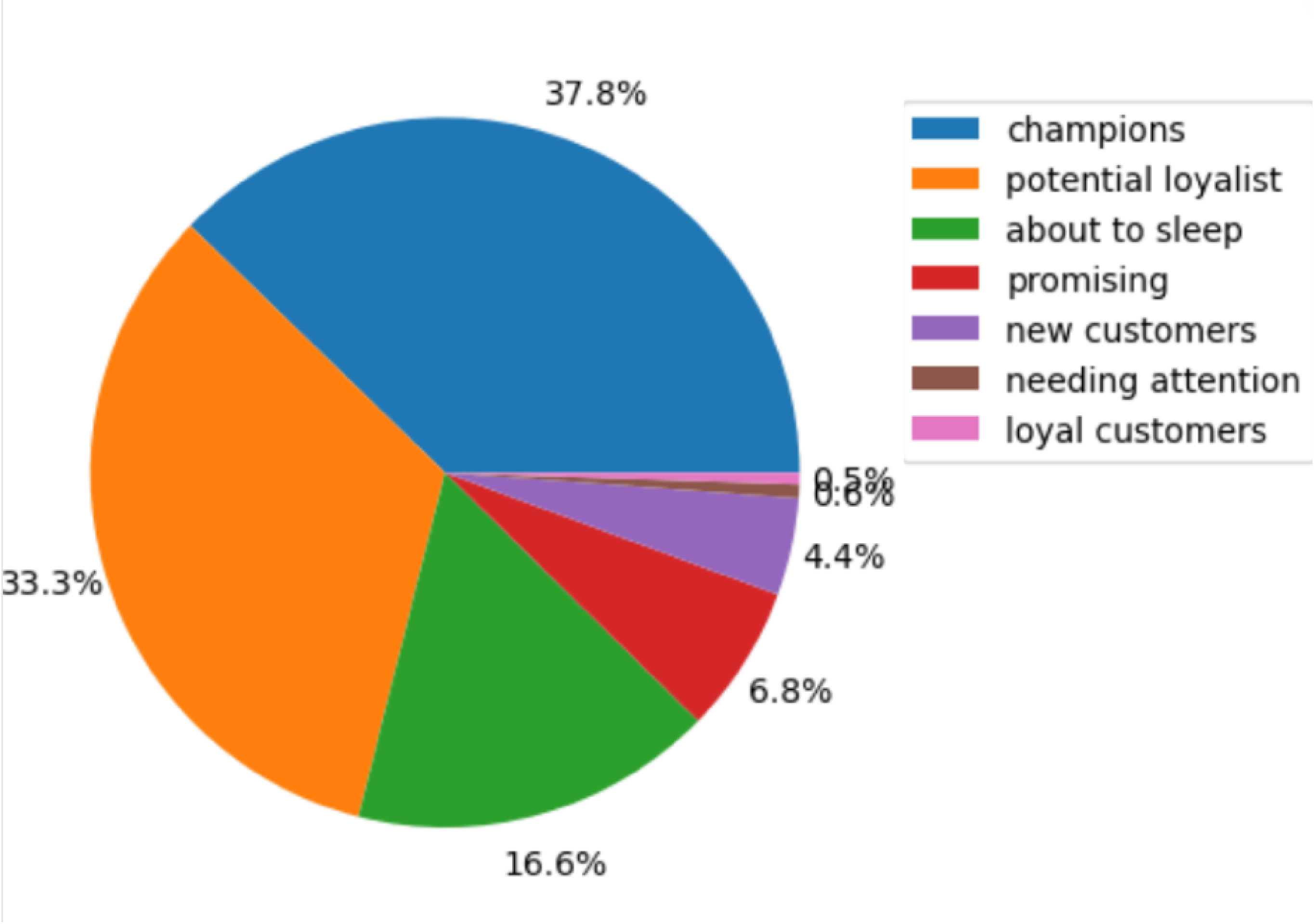


Figure 3.5: Distribution of segments

As shown in Figure 3.5, champions and potential loyalists cover most of the population 37.8% and 33.3% respectively and then comes about to sleep and promising customers 16.6% and 6.8% respectively and finally loyal , needing attention and new customers collectively comprise a smaller segment (around 6%).

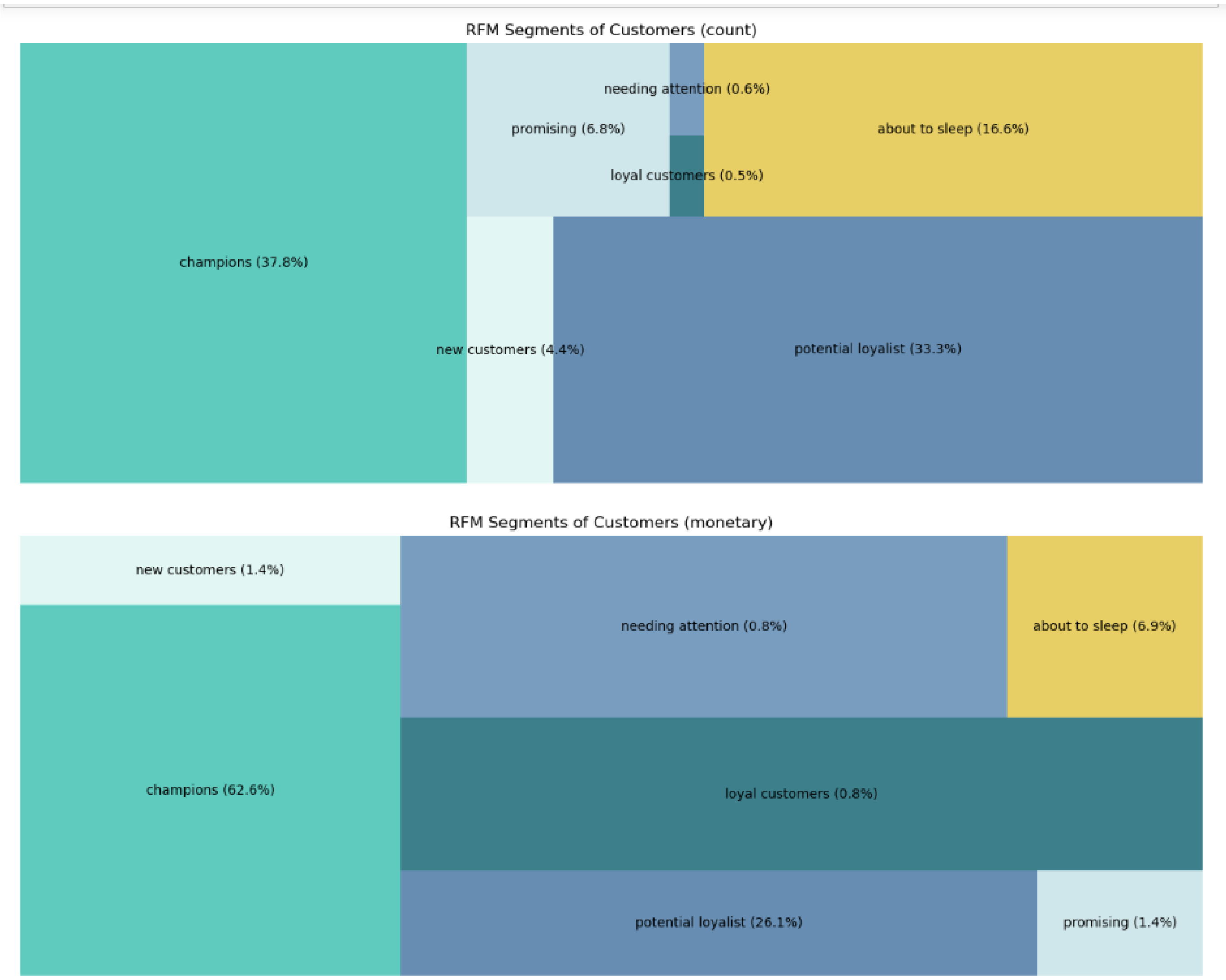


Figure 3.6: Customer Lifetime Value: Distribution by Segments (Count & Monetary)

Figure 3.6 shows that ”Champions” are crucial for the company at both customer count and monetary value, highlighting their importance. ”Potential Loyalists” form a significant segment of the customer base and contribute substantially to revenue. This suggest that additional efforts to convert them to ”Champions” could be beneficial. The ”About to Sleep” constitute an important group in counts but with a low monetary contribution, leading to a need for re-engagement strategies. ”Loyal Customers”; though a small segment in counts, hold high monetary value, making them a valuable group to allow importance to.

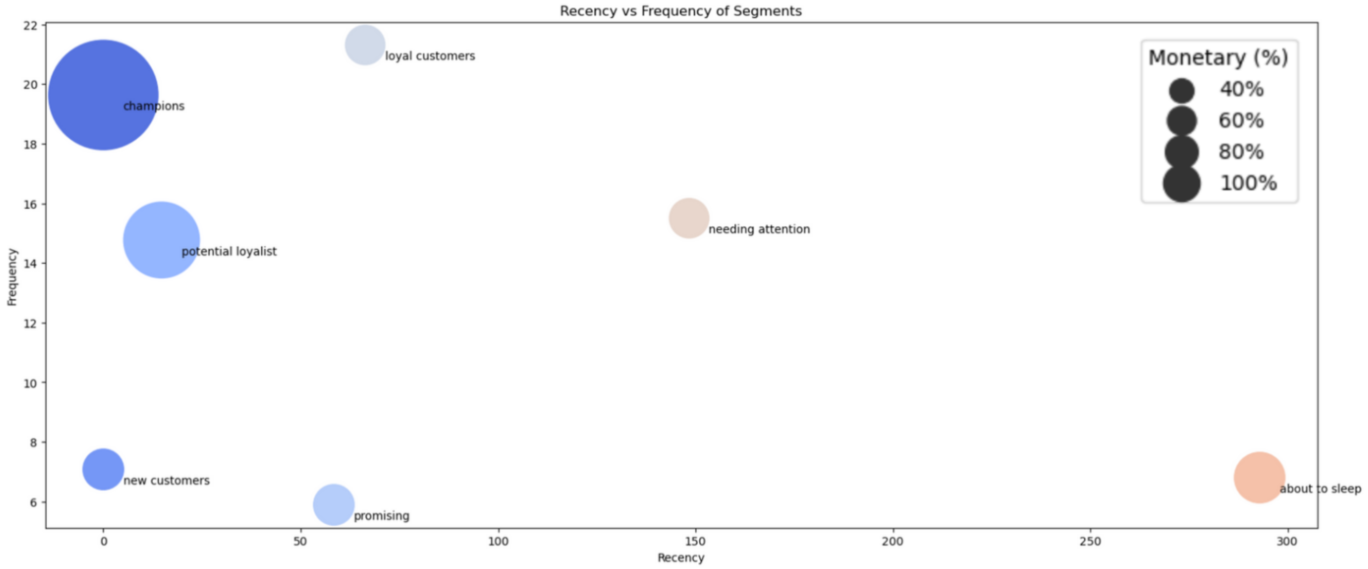


Figure 3.7: Behaviour and Spending analysis

Figure 3.7 shows RFM analysis of each segment. ”Champions” are frequent customers with the most recent activities. On the other hand, ”Loyal Customers” have high frequency but slightly longer recency than the champions. This suggests that we can work on them to elevate them to champion status.

As for those ”Needing attention”, they exhibit better frequency than potential loyalists but have significantly longer recency. We can either improve their recency to move them toward potential loyalist status or enhance their frequency to bring them closer to being loyal customers.

”New”, ”Promising”, and ”About-to-sleep” customers have low frequency. However, the ”New” and ”promising” ones have impressive recency, indicating that they could be very valuable in the future. Finally, the ”About-to-sleep” customers are concerning and should be targeted as they may become future churners.

Based on the previous analysis, we can summarize the following:

- **Champions:** Customers who have made an engagement very recently, do so frequently, and have high spending, indicating high value and strong loyalty.
- **Loyal Customers:** Regular buyers who spend a substantial amount, showing consistent engagement and responsiveness to promotional offers.
- **Potential Loyalists:** Newer customers who have made multiple engagements recently and spent a significant amount, indicating potential for future loyalty.
- **New Customers:** Individuals who have made their first engagement recently but have not yet shown frequency in their buying behavior.
- **Promising:** Customers who have made recent engagement s but have lower spending, showing potential for increased engagement.

- **Needing Attention:** Customers with above-average recent activity, frequency, and spending, who might not have engagement d very recently, suggesting they require re-engagement efforts.
- **About To Sleep:** Customers with below-average recency, frequency, and spending, indicating they are at risk of becoming inactive without prompt reactivation.

3.3 Conclusion

Although we can assign different weights to the data we wish to prioritise, the main objective of RFM is to observe and understand consumer behaviour in addition to spending. This chapter has helped us to understand that ”High Value Customers” represent 27% of the company’s clientele and generate more than 50% of its revenue. The most valuable customers are those with high frequency and very low recency, meaning they engage with the company frequently and very recently.

This study also shows what characteristics we can focus on to move a customer from one segment to another, as we saw in Figure 3.7 that we can target the customers that need attention, either by improving their recency to move them towards potential loyalist status, or by improving their frequency to move them closer to being loyal customers.

Chapter 4

Three-Month CLV Estimation: ML versus Lifetimes

In this section we used two seperate approaches to predict the clv first we employed XGBoost classifier and regressor to calculate both the probability of being ”alive” (active / non churner) in the next 3 months and the amount of money generated for each month,as for the second approach we used the beta geometric model and Gamma Gamma Model of the package Lifetimes. For the two approaches we made both temporal and random split for this modelling. Meaning that for the temporal split we cut off the last 3 months of data and we used it as a target data. We used the rest for training that we also randomly spit into train and test.

4.1 CLV Prediction Using XGBoost for a short period (3 months)

First we started off by building and setting the features: recency, sales last two weeks, life time, number of historical churn, frequency of activation, number of non churn, the average sales, the sum sales.

4.1.1 Predicting the revenue for the next 3 months

We used the grid search to get the best set of hyperparmeters: learning rate of 0.05, maximum depth of 5 and 100 estimators. This resulted as shown in Table 4.1 R-squared of 0.58 and a root mean squared error of x[1] a value less than half of the standard deviation of the real revenue.

Metrics	Train	Test
Mean Squared Error[1]	x	y
R-squared	0.58	0.56

Table 4.1: XGBoost Model evaluation

The model generalizes well to unseen data.

[1] For confidontial reason we are not able to state the real value of RMSE as the reader can guess the range of std value of the revenue

In order to minimise the root mean squared error we applied the logarithm of the real revenue before training the model. We obtained a very low RMSE of 0.811 and an R-squared of 0.77. For the assessment of the revenue we apply the exponential function of the predicted variables times the exponential of half the estimated variance[2]. Afterwards, we explored the feature importance for this prediction

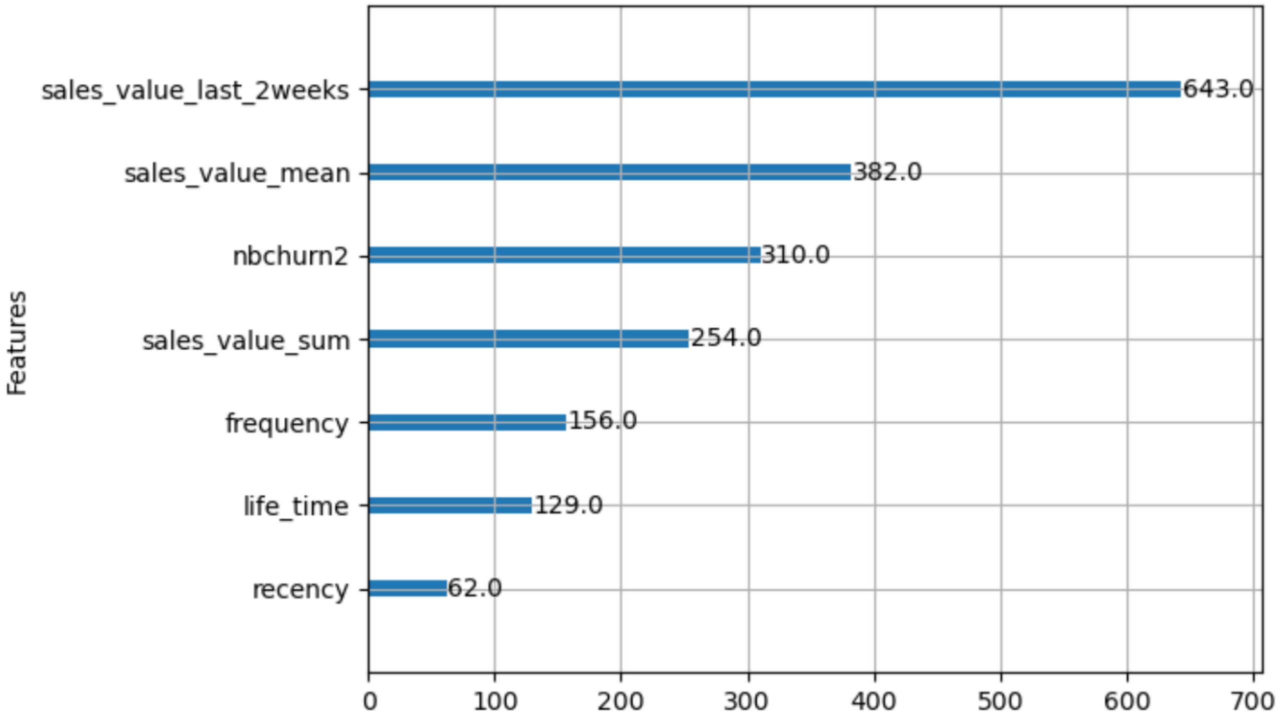


Figure 4.1: Feature importance for the revenue prediction

The results obtained show that the `sales_value_last_2weeks` is the most important feature to predict the revenue (see Figure 4.1).
The variables `sales_value_mean` and `nbchurn` and `sales_value_sum` occupy respectively the three following position.

4.1.2 Predicting the probability to be alive (active)

Finally we moved to predicting the probability to be alive. For this matter we used the XGboost classifier with the help of grid search that determined that the best parameters are a learning rate of 0.1, maximum depth of 7 and 300 estimators. Resulting in an RMSE 0.21 of and a R-squared of 0.6. Let's now explore the feature importance for the probability to be alive by observing Figure 4.2.

[2] When we adjust a lognormal model with mean μ and variance σ^2, the expectation of the variable (exponential of the logarithm) is equal $exp(\mu + \sigma^2/2)$

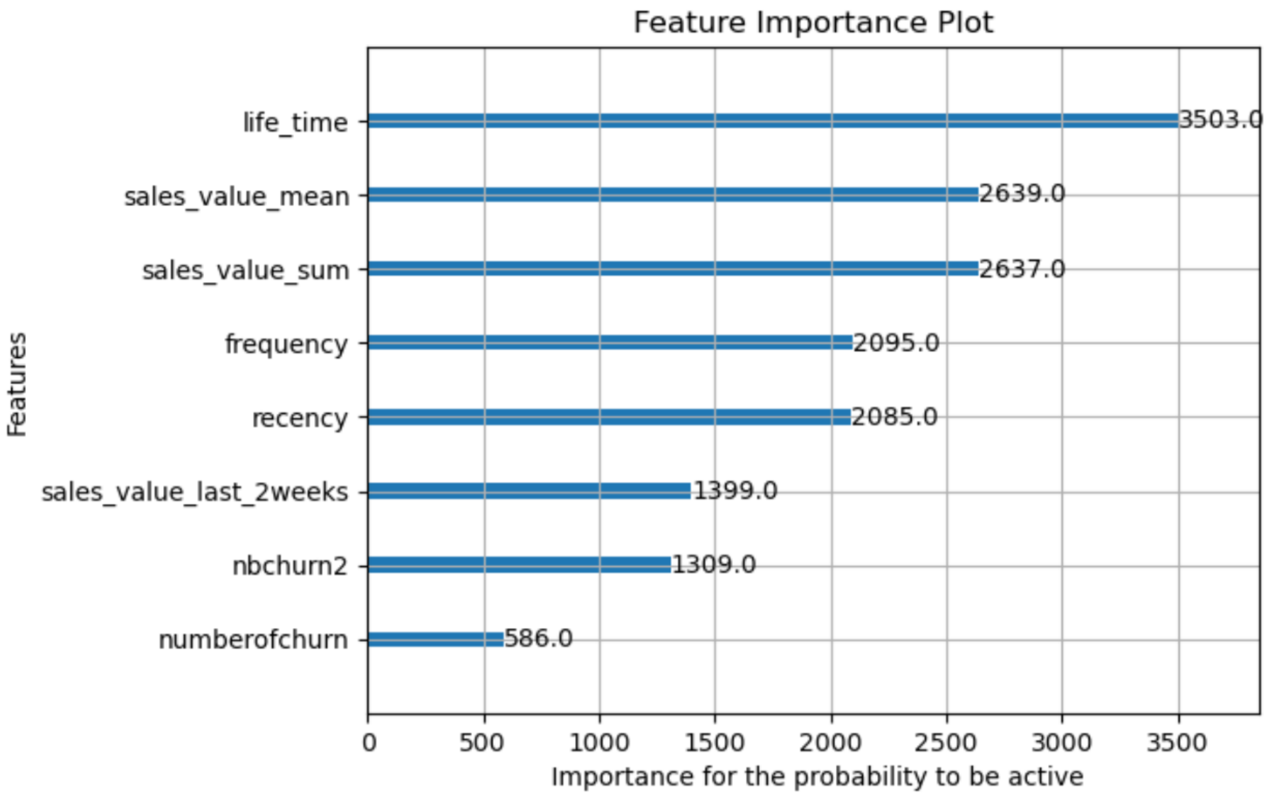


Figure 4.2: Feature importance of the probability to be alive

The results obtained show that the `life_time` is the most important feature to predict the survival probability (see Figure 4.2). The variables `sales_value_mean` and `sales_value_sum` are equivalent in terms of importance to predict the survival probability. At the same time, recency and frequency contribute the same way in the prediction of the survival probability.

4.2 Beta Geometric, Gamma Gamma Model

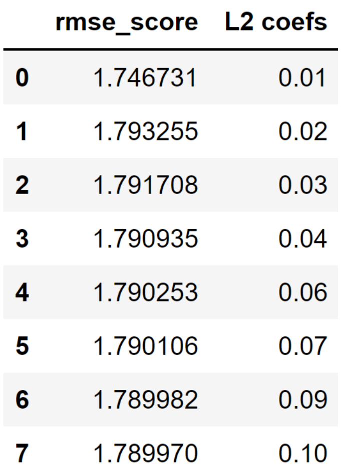
For this section we plan to utilize the **lifetime** package, designed primarily for CLV prediction. This package provides access to several advanced modeling techniques: the Beta geometric model, a probabilistic model that forecasts the likelihood of a customer engaging with the company again, and the Gamma Gamma model also a probabilitic model , which estimates the average transaction value for each customer. We'll only require the following metrics: **Recency**, **Frequency**, **Monetary** and **T** for Tenure defined as follow [5]:

- Frequency indicates how many times a customer has bought something. This means it's the total number of purchases minus one.
- T stands for the customer's age, measured in the time units selected (monthly in our case).
- Recency measures the time span from when a customer made their first purchase to their most recent one.
- Monetary_value is the average worth of a customer's purchases. It's calculated by adding up all the purchases a customer has made and dividing that total by the number of purchases.

We can either calculate these metrics ourselves or use the predefined function of the lifetime package called summary_data_from_transaction_data which requires only three inputs **customer Id** (`msisdn`), a **timespan** (`month_year` in our case) and **the engagement value** (`sales_value_sum` in our case). This function organised and computed our data to give us the metrics needed .

Let's begin by putting into action the Beta Geo Fitter. However, before we do that, we must remove any rows that have a monetary value of 0 or less, or we'll encounter an error while dealing with the gamma gamma model later on. As previously mentioned, the Beta Geo model is designed to predict the likelihood of a customer being alive in the upcoming period and also estimate the number of future purchases a customer is likely to make.For these predictions we need as an input the recency and frequency. As usual for this prediction task we preceded with a temporal split using the predefined function of the lifetime package called **calibration_and_holdout_data** that splits our data into calibration and holdout. During the calibration phase, the transactions are utilized for model training, while the transactions that happened during the observation phase (referred to as "holdout" transactions) are employed for model validation.

The L2 regularization is employed by the BG/NBD model. We will use "Grid Search" to determine the optimal L2 coefficient.



	rmse_score	L2 coefs
0	1.746731	0.01
1	1.793255	0.02
2	1.791708	0.03
3	1.790935	0.04
4	1.790253	0.06
5	1.790106	0.07
6	1.789982	0.09
7	1.789970	0.10

Figure 4.3: RMSE score by L2 coefficient

The best L2 value is 0.01 according to Figure 4.3.

What makes this package intriguing is that it can also show how the regressors (specifically, frequency and recency in the beta geometric scenario) impact the prediction.

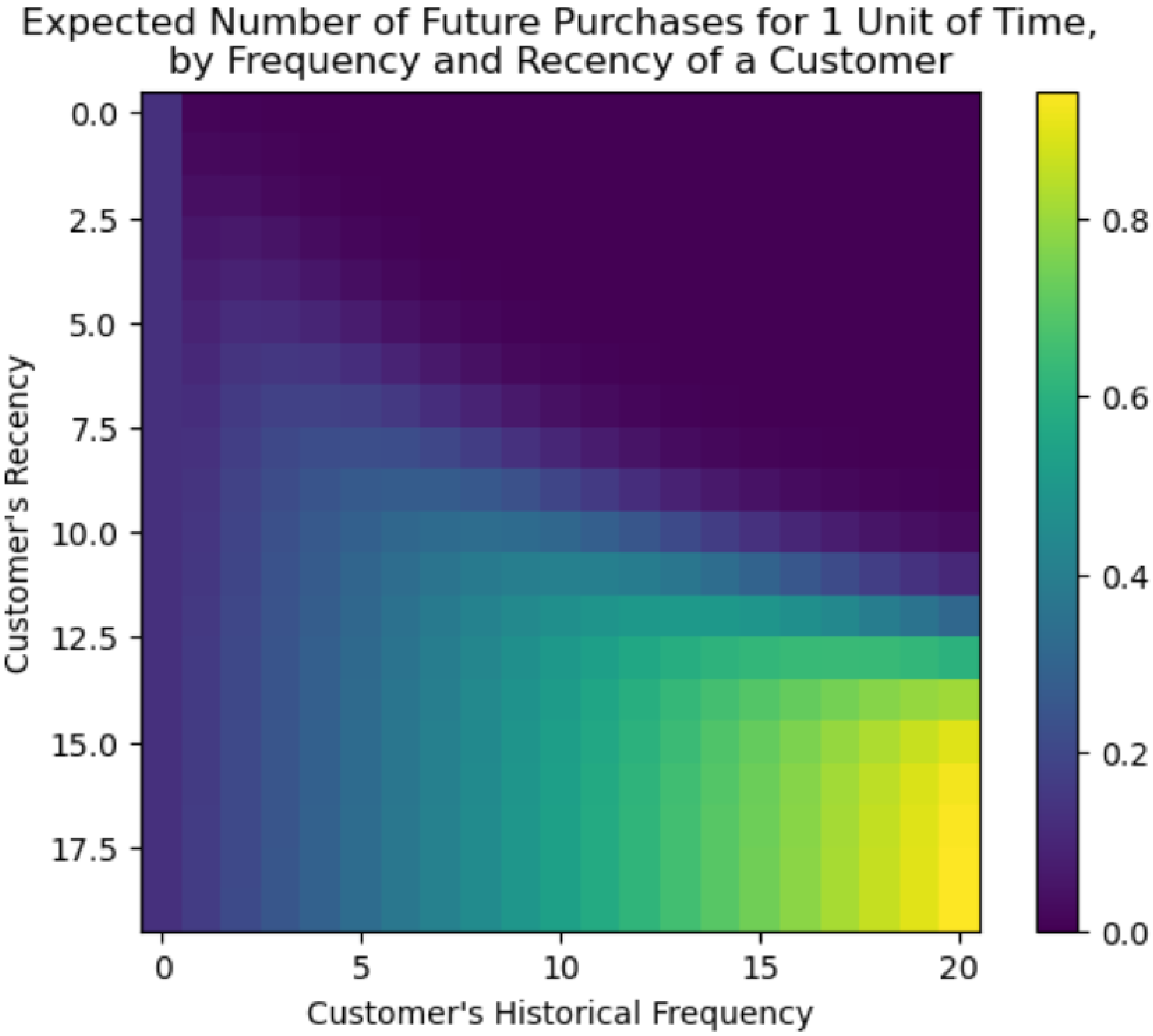


Figure 4.4: Expected number of future purchases for 1 unit of time by recency and frequency

It's evident that a customer qualifies as high value one if they've made 20 purchases, with their most recent transaction occurring between 15 to 17 months ago. Conversely, customers at the very bottom-right of the chart in Figure 4.4 are considered your most loyal, having made 20 purchases. Those at the top-right, on the other hand, are your most transient, having made numerous purchases in a short span and not being seen for months.

Additionally, there's this intriguing "tail" around the number 5 to 20. This signifies customers who make purchases less frequently but have been observed recently, suggesting they could potentially make another purchase. Whether these customers are still active or have simply paused their buying, remains uncertain.

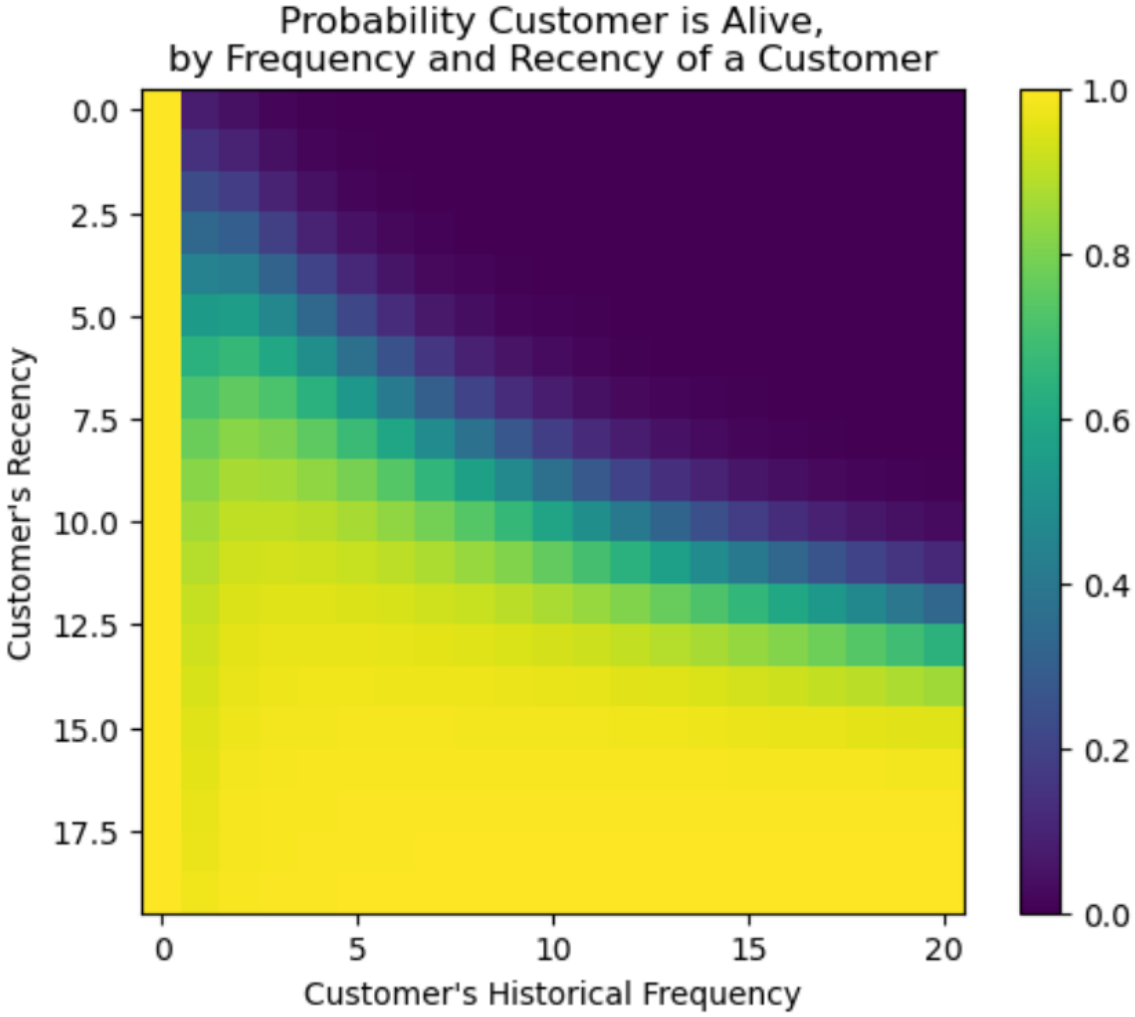


Figure 4.5: Probability to be alive by recency and frequency

Figure 4.5 shows how recency and frequency influence the chance of survival calculated by the model. The highest-value customers are those presented by the yellow area they have had multiple interactions with the company, and there's a noticeable difference between their first purchase and their latest one.

The RMSE of this prediction is 1.8 and the R-squared is 0.27. This indicates that the prediction is awful.

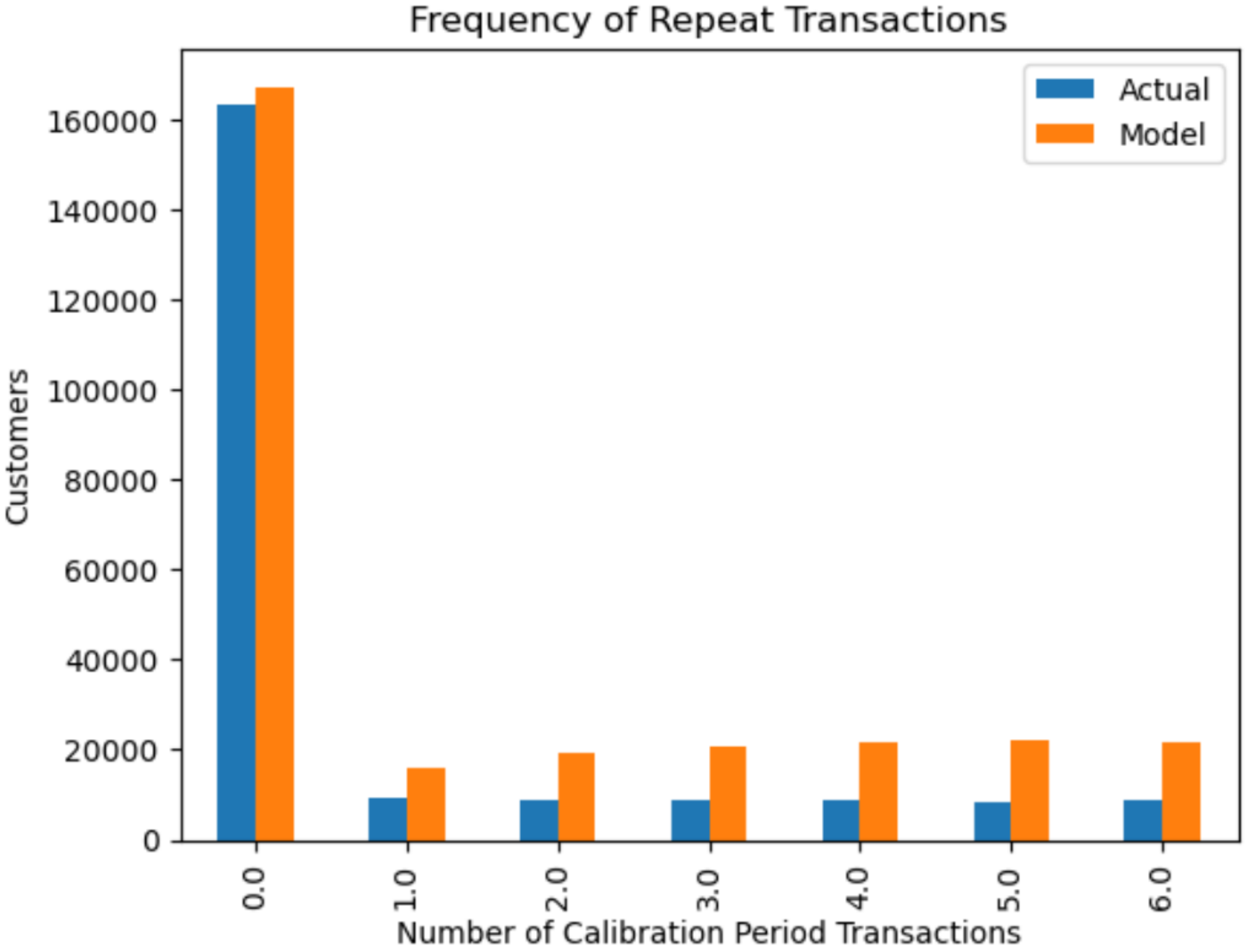


Figure 4.6: Comparison of Actual vs. Modeled Repeat Transaction Frequency

Figure 4.6 indicates that the model's predictions closely match the real data, suggesting it accurately reflects the underlying process.

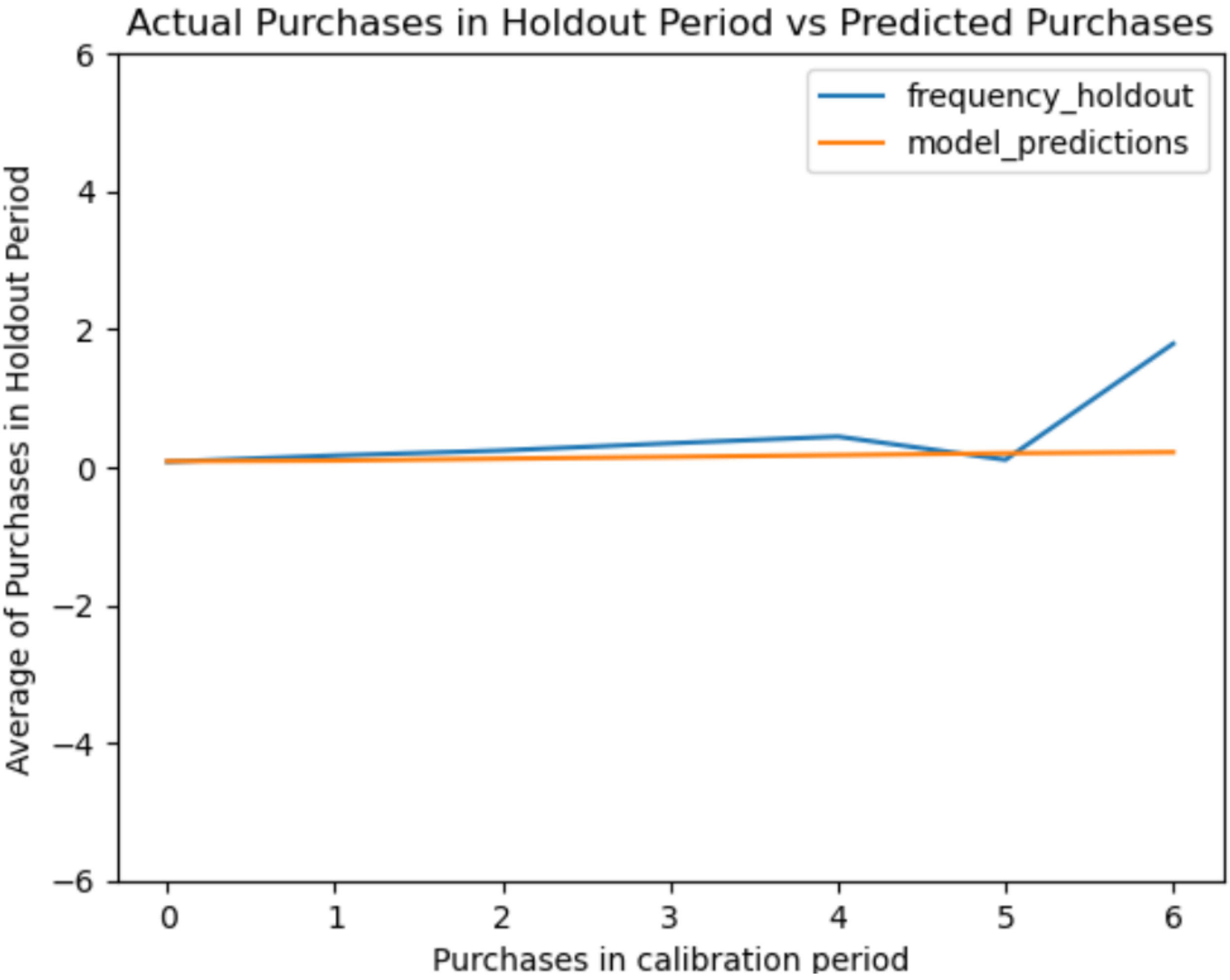


Figure 4.7: Comparison of Actual vs. Predicted Purchases in Holdout Period Based on Calibration Period Purchases

Both the actual (blue line) and predicted (orange line) purchases generally increase as the number of purchases in the calibration period increases. This suggests that customers who purchase more in the calibration period tend to also purchase more in the holdout period.While both lines show a similar increasing trend, there is a noticeable divergence between the predicted and actual values. In summary, the graph indicates that while the model captures the general trend that higher purchase frequency in the calibration period leads to higher purchase frequency in the holdout period, it tends to overestimate the actual number of purchases across all segments

Finally the Gamma-Gamma model was fitted, allowing us to forecast each customer's spending for the next three months. We implemented the function **customer__lifetime__value**, by using the predetermined beta geometric model, the prediction is also based on the frequency, recency , total time observed (T), and monetary value. the function takes also a discount rate δ defined above 1.3.1. However the RMSE is very high (38) and the R-squared is extremely low: 0.15.

4.3 Conclution:

Although the Lifetime and XGBoost models are both powerful tools, in our case the XGBoost classifier and regressor approach produced better prediction quality. Furthermore, unlike lifetime prediction models that can only indicate the impact of frequency and recency metrics, the XGBoost model can tell us how our factors affect the likelihood of survival or revenue forecast.

Chapter 5

CLV estimation using survival analysis and generalized models

This analysis presents a comprehensive examination of customer tenure using both the Kaplan-Meier survival curves and the Cox Proportional Hazards model . The objective is to understand the factors influencing customer retention and the survival probabilities of different customer segments based on historical churn behavior. We used the Kaplan-Meier curves to show survival probability over time and the Cox proportional hazards models to analyze how factors influence the risk of an event.

5.1 Kaplan-Meier Survival Curves

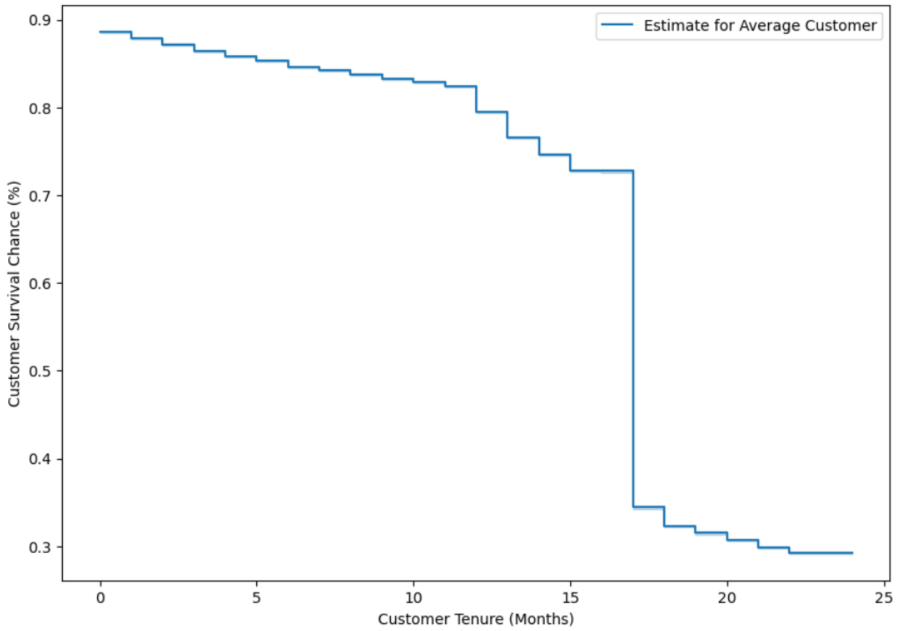


Figure 5.1: Kaplan-Meier Survival Curve for Customer Tenure

Figure 5.1 detailed the Kaplan-Meier survival curve. It depicts a clear pattern of customer churn. Initially, churn is low with a high survival rate (near 100%) for the first few months. This is followed by a period of steady decline as churn increases consistently. However, around the 17-month mark, a more pronounced churn event leads to a sharper decline, suggesting a critical period for customer churn where intervention strategies might be necessary to improve retention. Finally, by 20 months, the churn levels off, leaving a remaining customer base of about 30%. This indicates that a substantial portion of customers tend to churn within the first two years.

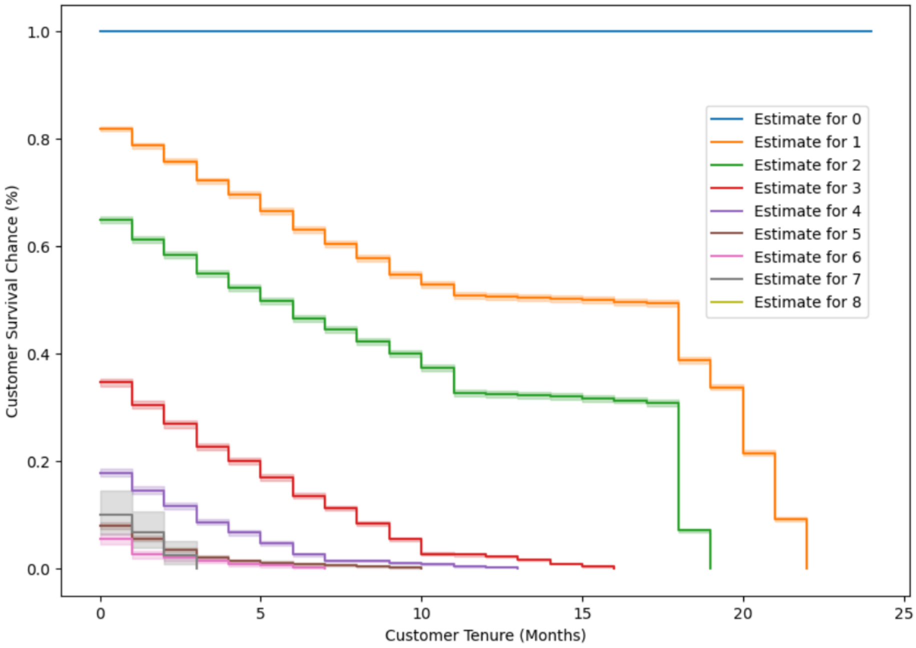


Figure 5.2: Kaplan-Meier Survival Curves by Historical Churn Frequency

Figure 5.2 reveals a clear correlation between churn history and future churn likelihood. Customers who never churned (blue curve) maintain a very high survival rate, suggesting minimal churn risk. However, churn history progressively worsens the situation. For customers who churned once (orange curve), the survival probability shows a moderate decline, indicating a higher churn rate compared to those who never churned. This effect becomes more pronounced for customers with two churns (green curve), where the survival probability declines more steeply.

The trend continues for customers with three or more churns (red curve and beyond). These curves plummet rapidly, indicating very high churn rates.The curves of customers with extremely high churn history (4 or more churns) are very overlapped show a negligible chance of survival and highlight the significant impact of past churn behavior on future customer retention.

This information led us to categorize the customers based on their number of churns. We created four groups: the first for people who churned once, the second for those who churned twice, the third group for those who churned three times, and the last group for customers who churned four or more times.

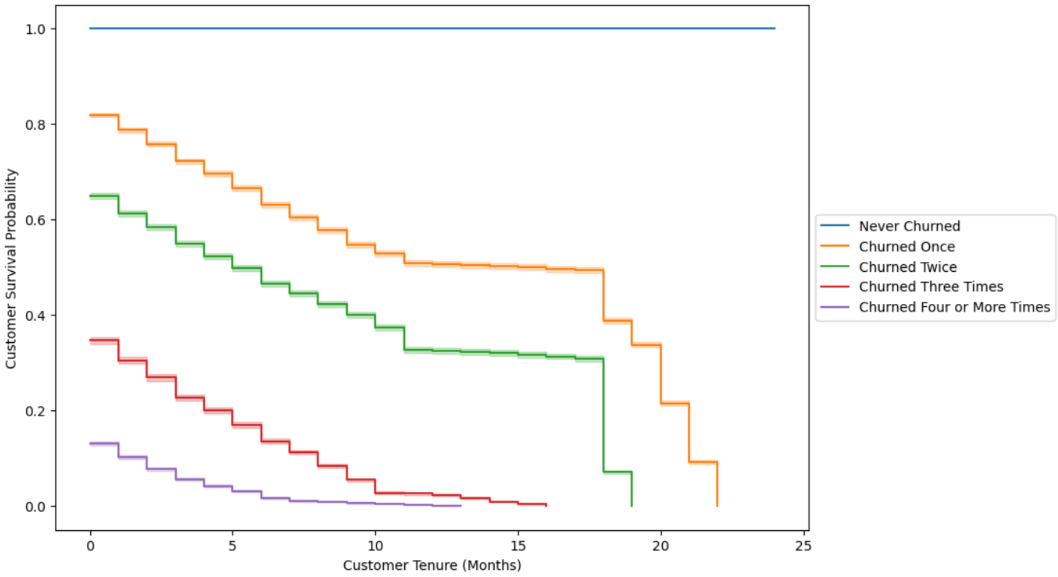
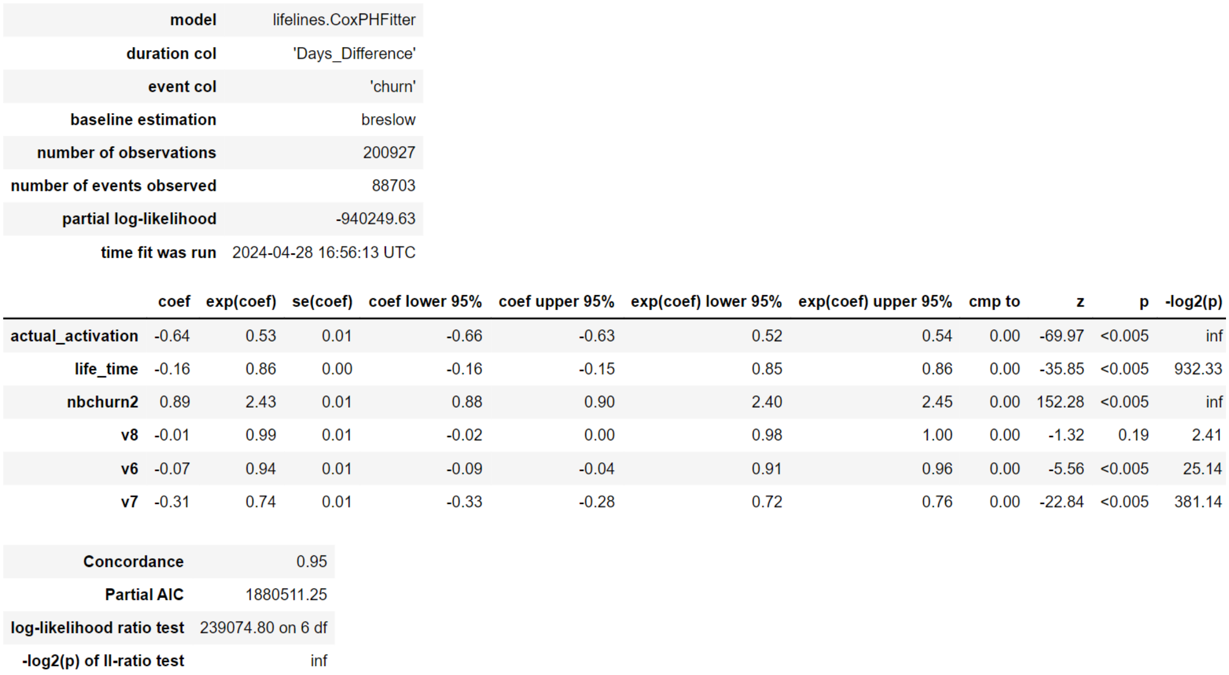


Figure 5.3: Kaplan-Meier Survival Curves by Historical Churn Frequency

Figure 5.3 provides a clearer visual representation of how historical churn frequency affects customer retention. Customers with no previous churns have the highest retention rates, while those with multiple churns have significantly lower survival probabilities.
This can be considered as an other customer classification.

5.2 Cox Proportional Hazards

For the purpose of adjusting the Cox-proportional hazard model we proceed by splitting the database into a **train** for training the model and **test** for testing it. Figure 5.4a reveals several factors influencing customer churn. the frequency (`actual_activation`) has a negative and significant impact on the churn risk. Indeed, an additional activation (re-charge) would reduce churn risk by 47%. Similarly, longer customer tenure (`life_time`) proves beneficial, with each unit increase lowering churn risk by 14%. Conversely, past churn behavior (`nbchurn2`) is significantly correlated with high churn risk. Indeed, an additional increase in the previous churn behavious would lead to more than doubling the likelihood of churn. In the other hand, the variable ”number of outgoing calls” (`v7_value_sum`) are associated with a 26% reduction in churn risk, while the ”number incoming calls” (`v6_value_sum`) has a modest protective effect. Thea variable ”total data usage” (`v8_value_sum`) shows a neglected impact on churn.

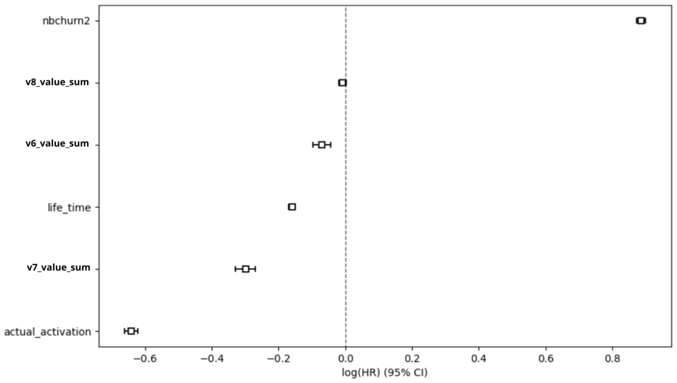


model	lifelines.CoxPHFitter
duration col	'Days_Difference'
event col	'churn'
baseline estimation	breslow
number of observations	200927
number of events observed	88703
partial log-likelihood	-940249.63
time fit was run	2024-04-28 16:56:13 UTC

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
actual_activation	-0.64	0.53	0.01	-0.66	-0.63	0.52	0.54	0.00	-69.97	<0.005	inf
life_time	-0.16	0.86	0.00	-0.16	-0.15	0.85	0.86	0.00	-35.85	<0.005	932.33
nbchurn2	0.89	2.43	0.01	0.88	0.90	2.40	2.45	0.00	152.28	<0.005	inf
v8	-0.01	0.99	0.01	-0.02	0.00	0.98	1.00	0.00	-1.32	0.19	2.41
v6	-0.07	0.94	0.01	-0.09	-0.04	0.91	0.96	0.00	-5.56	<0.005	25.14
v7	-0.31	0.74	0.01	-0.33	-0.28	0.72	0.76	0.00	-22.84	<0.005	381.14

Concordance	0.95
Partial AIC	1880511.25
log-likelihood ratio test	239074.80 on 6 df
-log2(p) of ll-ratio test	inf

(a) Model summary



(b) Forest plot of the log hazard ratios

Figure 5.4: Survival Analysis of Customer Churn Using Cox Proportional Hazards Model

On the other hand according to Figure 5.4b actual_activation, longer customer tenure (`life_time`), higher values of incoming and outgoing contacts (`v6_value_sum` and `v7_value_sum`), all reduce churn risk with statistically significant effects. In contrast, past churn behavior (`nbchurn2`) increases churn risk, while data usage (`v8_value_sum`) has minimal impact.

The high concordance index (0.95) and the highly significant log-likelihood ratio test demonstrate that the CoxPH model is effective in predicting customer churn. The statistically significant coefficients suggest that the included variables (except `v8`) are meaningful predictors of churn.

However, while this model performs well, testing on new data is essential. After testing it on new customers, we have obtained a closer c-index of about 0.955.

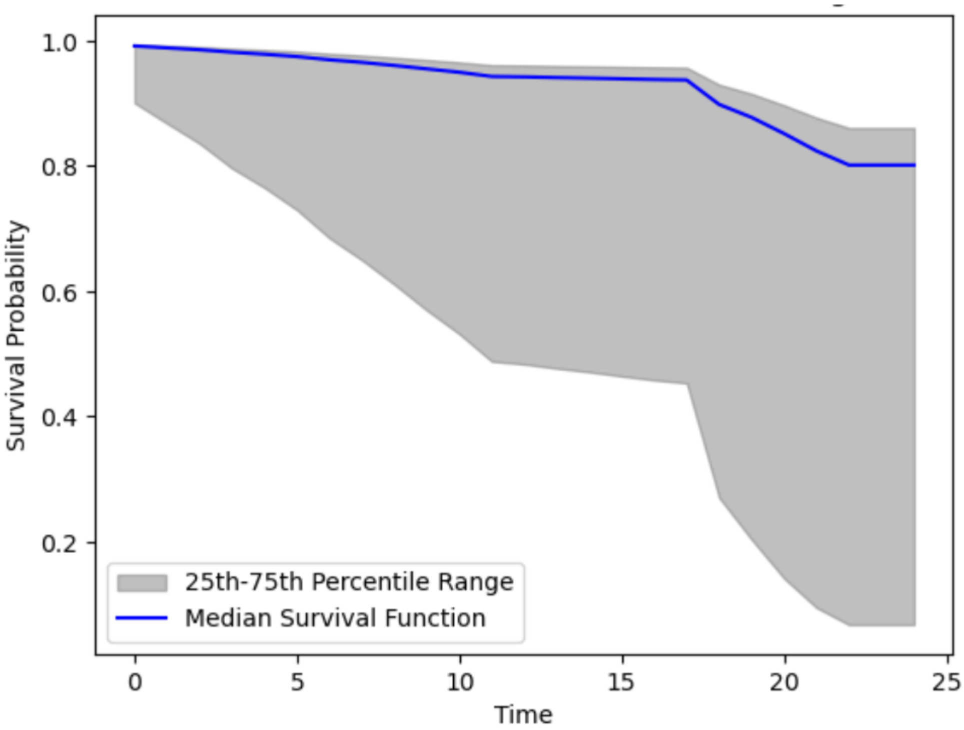


Figure 5.5: Median Survival Function with Interquartile Range

Figure 5.5 shows the median survival function , along with the interquartile range (25th-75th percentile) shaded in gray.The blue line represents the median survival probability over time. It indicates the point at which 50% of the customers are expected to have churned by a specific time. Initially it starts from 1, meaning no customers have churned then it starts to decrease as the time progresses. At the end of the study period (month 24) 50% of the customers have about 85% probability to survive.
The gray area represents the interquartile range (IQR), which covers the 25th to 75th percentile of the survival probabilities. The IQR provides a measure of the variability or spread in the survival probabilities. A wider gray area indicates more variability, while a narrower area indicates less variability. In our case, the survival probability remains initially high with a narrow IQR, indicating that most customers are retained in the early stages. However by around 10-15 months, there is a noticeable decline in the median survival probability, and the IQR begins to widen. At the end of study period (month 20) the IQR shows significant variability, suggesting that customer churn becomes more unpredictable during this period. This results sign of heterogeneity between customers.

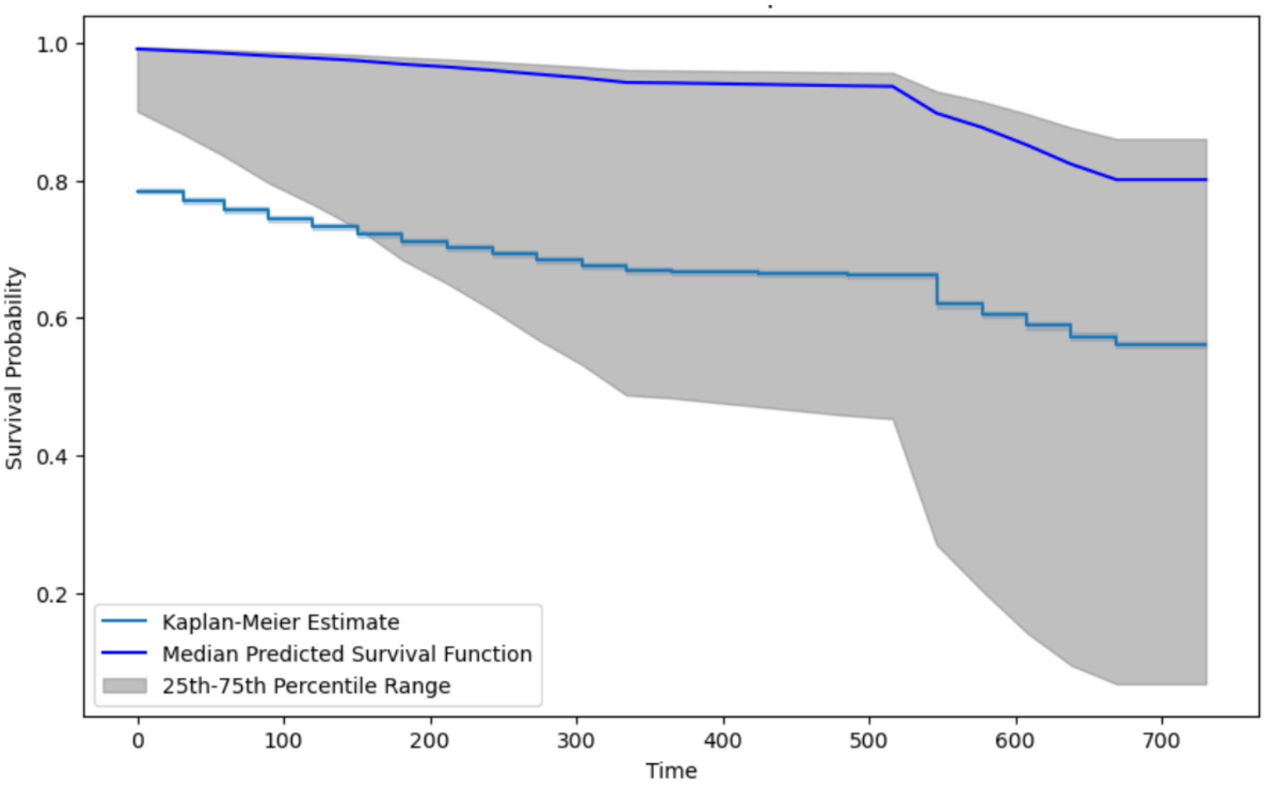
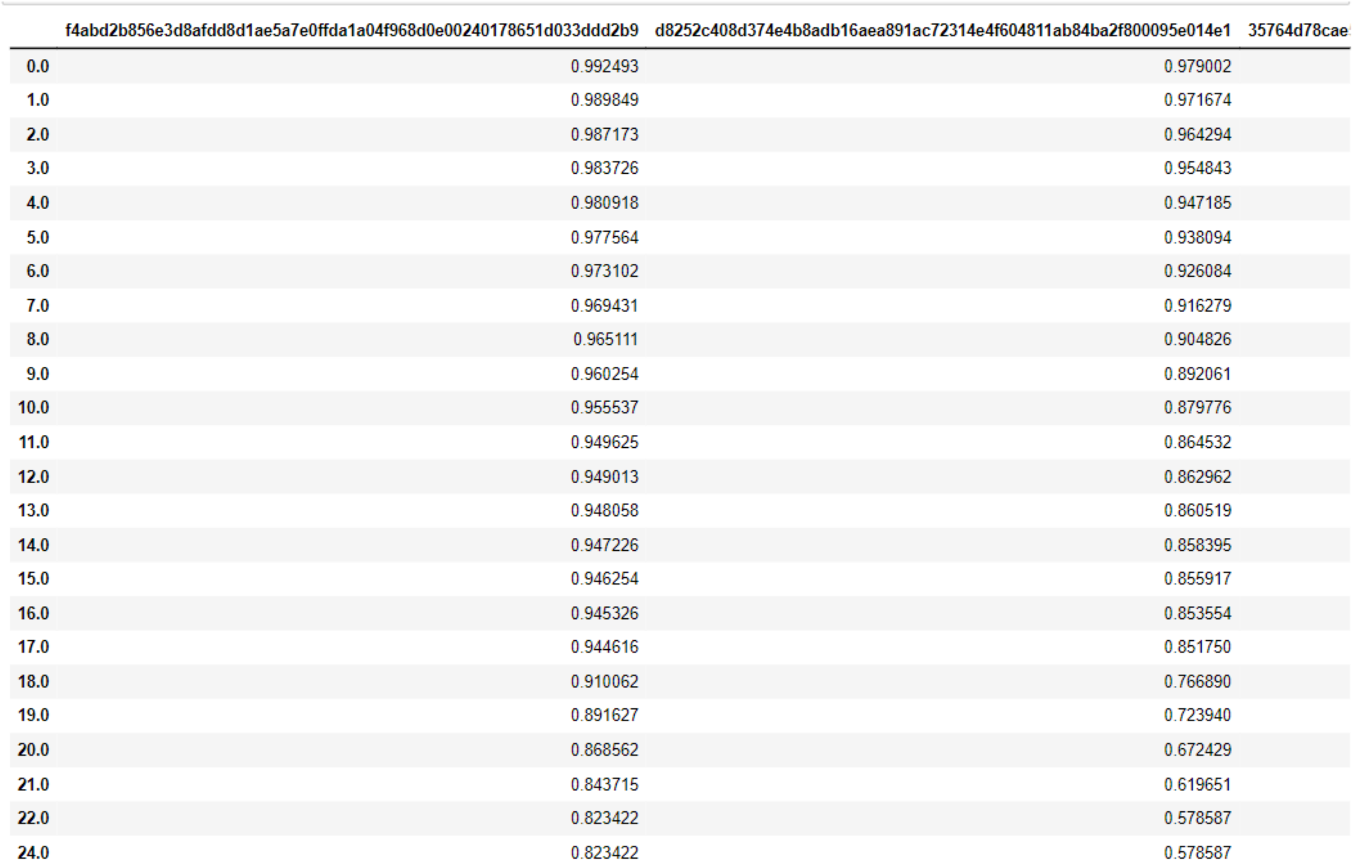


Figure 5.6: Comparison of Kaplan-Meier Estimate and Median Predicted Survival Function

Figure 5.6 compares the Kaplan-Meier survival estimate with the median predicted survival function, incorporating the interquartile range. Let's notice that at the beginning period, we have already 20% of customers are churner. This is at the origin of the difference between kaplan-meier and median predicted survival function of Cox. We notice also that this difference almost remains for the hole period. Gradually, the Kaplan-Meier estimate shows a decrease. As parametric survival function, Cox model leads to smooth median predicted survival function that decline slower and exhibits a continuous prediction. In the other hand; The Kaplan-Meier, a non-parametric function, leads to non-continuous survival function. Nevertheless, it remains inside the IQR median survival survival function.

As the Cox proportional survival model will be used to evaluate the CLV for the each customer, we have to forecast the survival probability of each customer and for each period ($t = 0, \cdots, 24$).

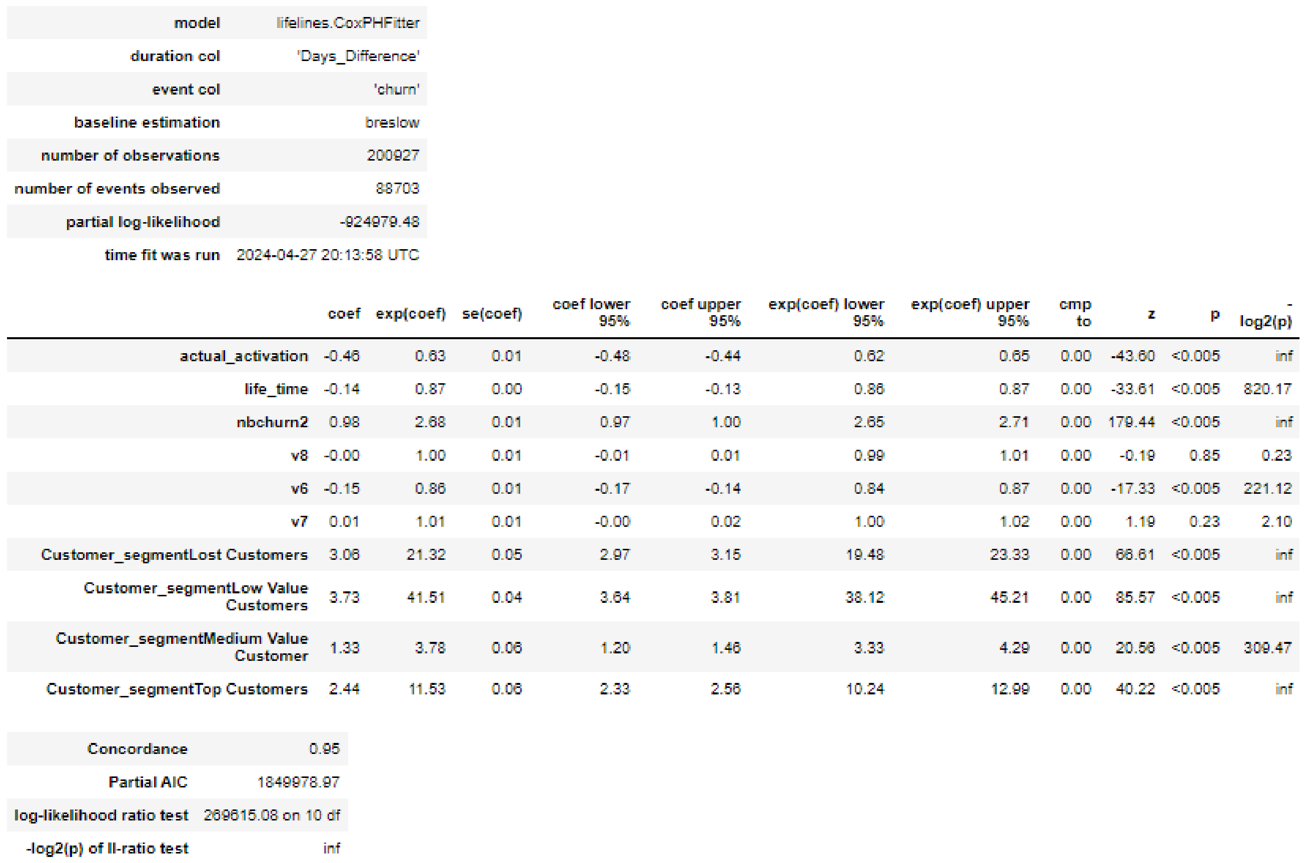


	f4abd2b856e3d8afdd8d1ae5a7e0ffda1a04f968d0e00240178651d033ddd2b9	d8252c408d374e4b8adb16aea891ac72314e4f604811ab84ba2f800095e014e1	35764d78cae
0.0	0.992493	0.979002	
1.0	0.989849	0.971674	
2.0	0.987173	0.964294	
3.0	0.983726	0.954843	
4.0	0.980918	0.947185	
5.0	0.977564	0.938094	
6.0	0.973102	0.926084	
7.0	0.969431	0.916279	
8.0	0.965111	0.904826	
9.0	0.960254	0.892061	
10.0	0.955537	0.879776	
11.0	0.949625	0.864532	
12.0	0.949013	0.862962	
13.0	0.948058	0.860519	
14.0	0.947226	0.858395	
15.0	0.946254	0.855917	
16.0	0.945326	0.853554	
17.0	0.944616	0.851750	
18.0	0.910062	0.766890	
19.0	0.891627	0.723940	
20.0	0.868562	0.672429	
21.0	0.843715	0.619651	
22.0	0.823422	0.578587	
24.0	0.823422	0.578587	

Figure 5.7: Survival function

The Figure 5.7 gives an example for two customers. Of course, the survival probability could be evaluated for each segment of customers identified in the RFM classification. And later on; we can identify the contribution of each segment to the formation of the CLV. This leads to re-estimate the Cox proportional model by adding the label segments as explanatory variables.

The Figure 5.8 presents the results of the augmented model, where the "Lost customers" are the reference point for this analysis. The results conduct to the fact that "Low Value Customers" have a significantly high risk of churn, with a hazard ratio of 21.32, compared to the "Lost Customers". This means that they are over 21 times more likely to churn than those already lost. "Medium Value Customers" have a lower have a higher risk, than the "Lost customers", but with a lesser extent, with a hazard ratio of 3.78. More surprising, "Top Customers" have high hazard ratio of 11.53, compared to the "Lost Customers". This indicates that the company should allow high importance to these customers and act to reduce their churn probability.



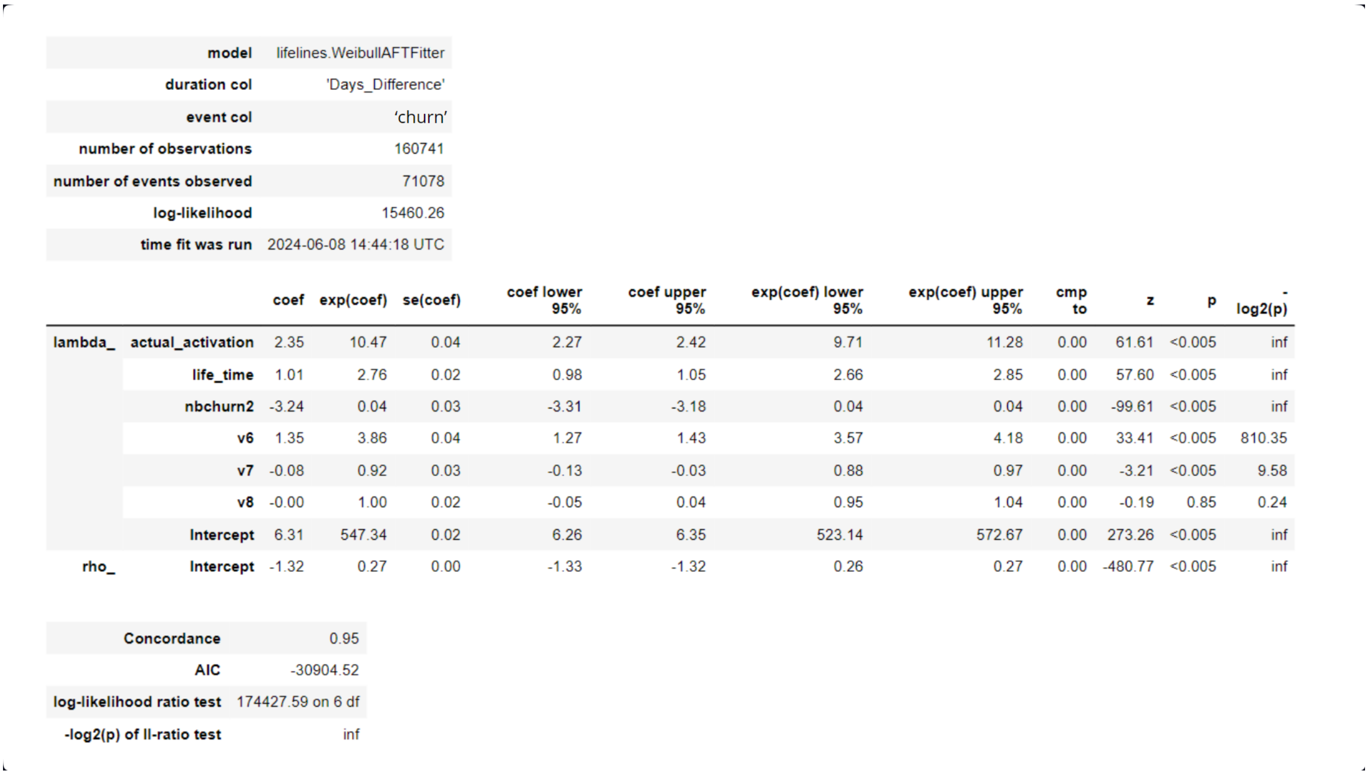
model	lifelines.CoxPHFitter
duration col	'Days_Difference'
event col	'churn'
baseline estimation	breslow
number of observations	200927
number of events observed	88703
partial log-likelihood	-924979.48
time fit was run	2024-04-27 20:13:58 UTC

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
actual_activation	-0.46	0.63	0.01	-0.48	-0.44	0.62	0.65	0.00	-43.60	<0.005	inf
life_time	-0.14	0.87	0.00	-0.15	-0.13	0.86	0.87	0.00	-33.61	<0.005	820.17
nbchurn2	0.98	2.68	0.01	0.97	1.00	2.65	2.71	0.00	179.44	<0.005	inf
v8	-0.00	1.00	0.01	-0.01	0.01	0.99	1.01	0.00	-0.19	0.85	0.23
v6	-0.15	0.86	0.01	-0.17	-0.14	0.84	0.87	0.00	-17.33	<0.005	221.12
v7	0.01	1.01	0.01	-0.00	0.02	1.00	1.02	0.00	1.19	0.23	2.10
Customer_segmentLost Customers	3.06	21.32	0.05	2.97	3.15	19.48	23.33	0.00	66.61	<0.005	inf
Customer_segmentLow Value Customers	3.73	41.51	0.04	3.64	3.81	38.12	45.21	0.00	85.57	<0.005	inf
Customer_segmentMedium Value Customer	1.33	3.78	0.06	1.20	1.46	3.33	4.29	0.00	20.56	<0.005	309.47
Customer_segmentTop Customers	2.44	11.53	0.06	2.33	2.56	10.24	12.99	0.00	40.22	<0.005	inf

Concordance	0.95
Partial AIC	1849978.97
log-likelihood ratio test	269615.08 on 10 df
-log2(p) of ll-ratio test	inf

Figure 5.8: Cox proportional hazard segments

5.3 Accelerated Failure Time: AFT



model	lifelines.WeibullAFTFitter
duration col	'Days_Difference'
event col	'churn'
number of observations	160741
number of events observed	71078
log-likelihood	15460.26
time fit was run	2024-06-08 14:44:18 UTC

		coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
lambda_	**actual_activation**	2.35	10.47	0.04	2.27	2.42	9.71	11.28	0.00	61.61	<0.005	inf
	life_time	1.01	2.76	0.02	0.98	1.05	2.66	2.85	0.00	57.60	<0.005	inf
	nbchurn2	-3.24	0.04	0.03	-3.31	-3.18	0.04	0.04	0.00	-99.61	<0.005	inf
	v6	1.35	3.86	0.04	1.27	1.43	3.57	4.18	0.00	33.41	<0.005	810.35
	v7	-0.08	0.92	0.03	-0.13	-0.03	0.88	0.97	0.00	-3.21	<0.005	9.58
	v8	-0.00	1.00	0.02	-0.05	0.04	0.95	1.04	0.00	-0.19	0.85	0.24
	Intercept	6.31	547.34	0.02	6.26	6.35	523.14	572.67	0.00	273.26	<0.005	inf
rho_	**Intercept**	-1.32	0.27	0.00	-1.33	-1.32	0.26	0.27	0.00	-480.77	<0.005	inf

Concordance	0.95
AIC	-30904.52
log-likelihood ratio test	174427.59 on 6 df
-log2(p) of ll-ratio test	inf

Figure 5.9: Survival Analysis using Weibull AFT Model on Customer Churn Data

The analysis shown in Figure 5.9 reveals several factors influencing customer churn. Higher `actual_activation` (frequency) (coef: 2.97, exp(coef): 19.47) and `life_time` (customer seniority) (coef: 1.38, exp(coef): 3.98) significantly increase survival time, indicating reduced churn probability for more engaged and longstanding customers. Conversely, `nbchurn2` (coef: -4.12, exp(coef): 0.02) had a strong negative effect, reducing survival time by 98%, signifying a substantial increase in churn likelihood. While `v6` (number of incoming calls) (coef: 1.77, exp(coef): 5.87) also exhibited a positive effect, increasing survival time, `v7` (number of outgoing calls) (coef: -0.12, exp(coef): 0.88) had a minor negative impact, suggesting a slight rise in churn probability. Notably, `v8` (Data usage) (coef: -0.01, exp(coef): 0.99) displayed a negligible effect (1% decrease in survival time) and was statistically insignificant (p-value: 0.79). Additionally, the significant and positive intercept (Lambda) indicates a baseline association with survival time, while the Rho value (Shape parameter) sheds light on the hazard function's behavior over time.

By contemplating Figures 5.4a and 5.9 we can set the differences and similarities of the prediction. Both models don't predict significantly the coefficient of `v8` (Data usage), they also give the same roles to the number of historical churn, actual activation (frequency), `life_time` and `v6` (number of incoming contacts) on the survival of the client in the company for the aft (on the hazard of leaving the company for the Cox), added to that they have high c-index reflecting the good quality of the prediction. However, the cox prediction state that the number of outgoing calls (`v7`) minimise the hazard to leave the company and the aft says the contrary.

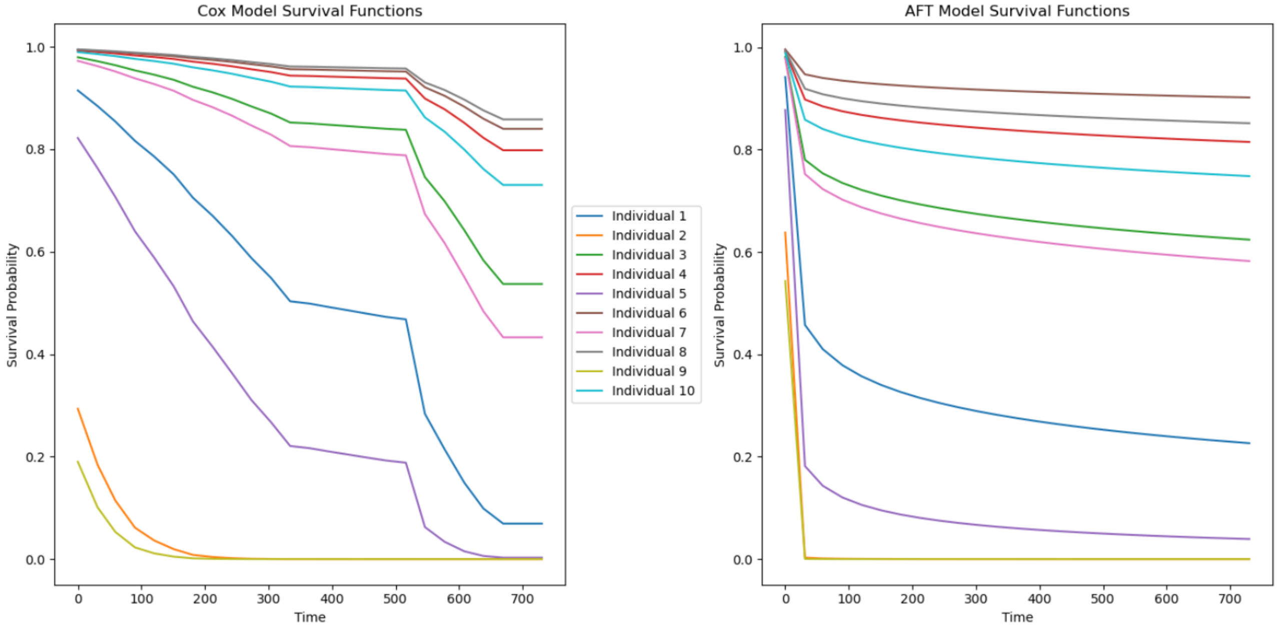


Figure 5.10: Comparison of Survival Functions Using Cox and AFT Model of ten random individual

Figure 5.10 shows a comparison between the survival function calculated by the cox and the AFT model. In the Cox PH model, significant variation among individuals is evident. Individual 9 (yellow) experiences a rapid decline, reaching a survival probability of nearly 0% by around 5 months. Similarly, Individual 2 (orange) and Individual 7 (brown) show steep declines, dropping to near 0% within 7 months and 10 units, respectively. Individual 5 (purple) also exhibits a notable decline, reaching about 20% survival probability by 15 months. Conversely, Individual 8 (gray) maintains a higher survival probability, staying above 70% at 25 units, indicating a lower risk compared to others.

On the right, the AFT model shows a more gradual and uniform decline in survival probabilities. Individual 9 (yellow), consistent with the Cox PH model, still shows a rapid decline, with survival probability close to 0% by around 5 units. Other individuals, such as Individual 8 (gray) and Individual 10 (cyan), exhibit a slower decline. By 25 months, Individual 8 maintains a survival probability around 70%, while Individual 10 remains above 50%.

Comparing the two models, the Cox PH model emphasizes sharper differences in risk levels among individuals, with more abrupt declines for high-risk individuals. For example, Individual 1 (blue) drops to about 10% survival probability by 20 monthsin the Cox PH model, whereas in the AFT model, it declines more gradually to approximately 30% within the same timeframe. The AFT model suggests a smoother and more consistent decline in survival probabilities across the cohort.

Overall, this comparison underscores the distinct perspectives each model provides. The Cox PH model emphasizes individual risk variations sharply, whereas the AFT model offers a more homogenized view of survival probabilities over time

5.4 Predicting the revenue: Linear Regression, XGBoost

5.4.1 Correlation study

In order to forecast the revenue for each customers, we adjust the revenue on the charateristics of the customers. Figure 5.11 presents the correlation coefficients between various customer metrics, including activation measures, sales values, and customer lifetime. The values range from -1 to 1, where -1 indicates a perfect negative correlation, 1 indicates a perfect positive correlation, and 0 indicates no correlation. The results indicates that there are high positive correlations between `Days_difference`, `activation` and `actual_activation` and `life_time`. The variable `nubchurn2` is negatively correlated with the variables cited below.

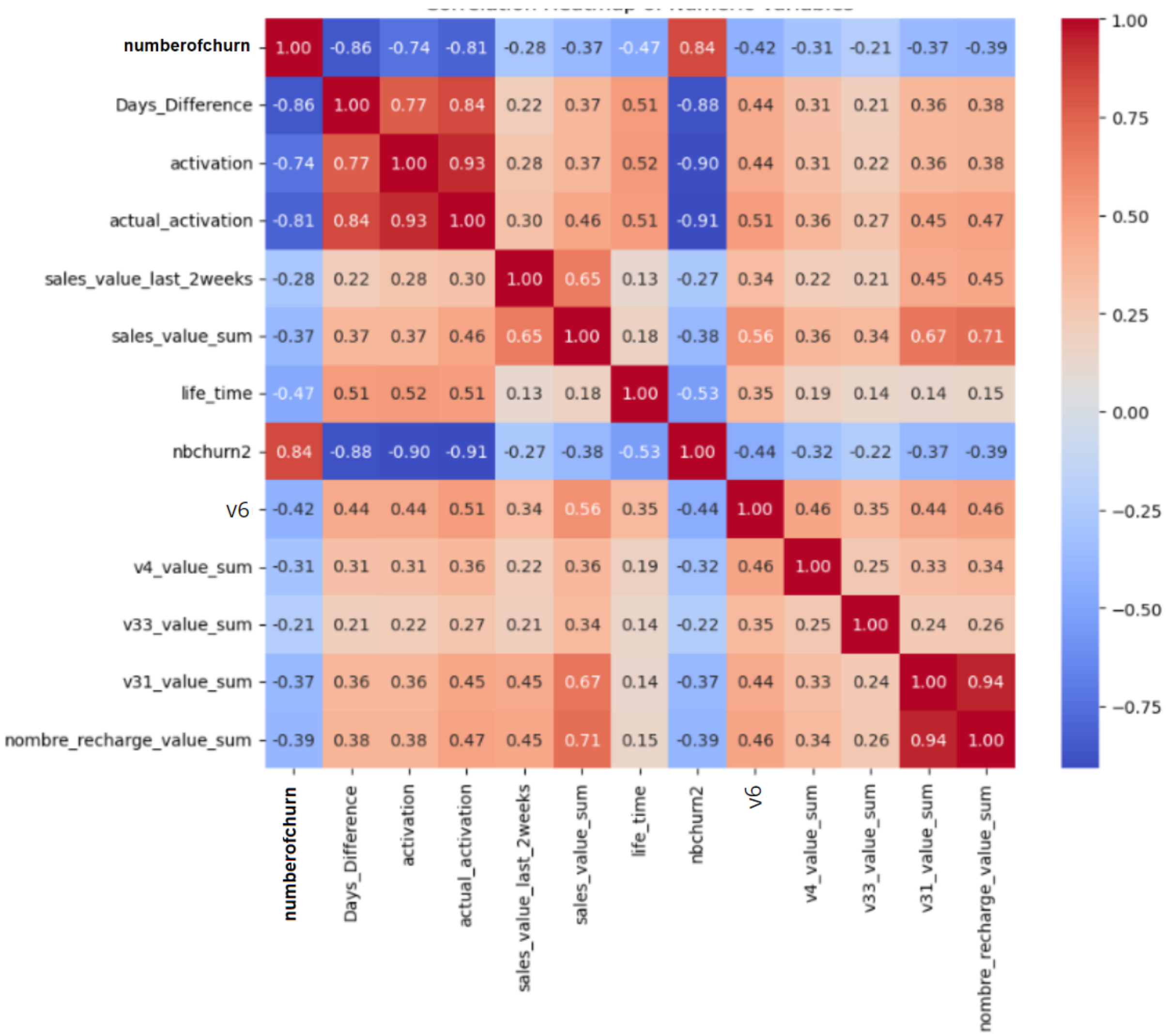


Figure 5.11: Feature correlation

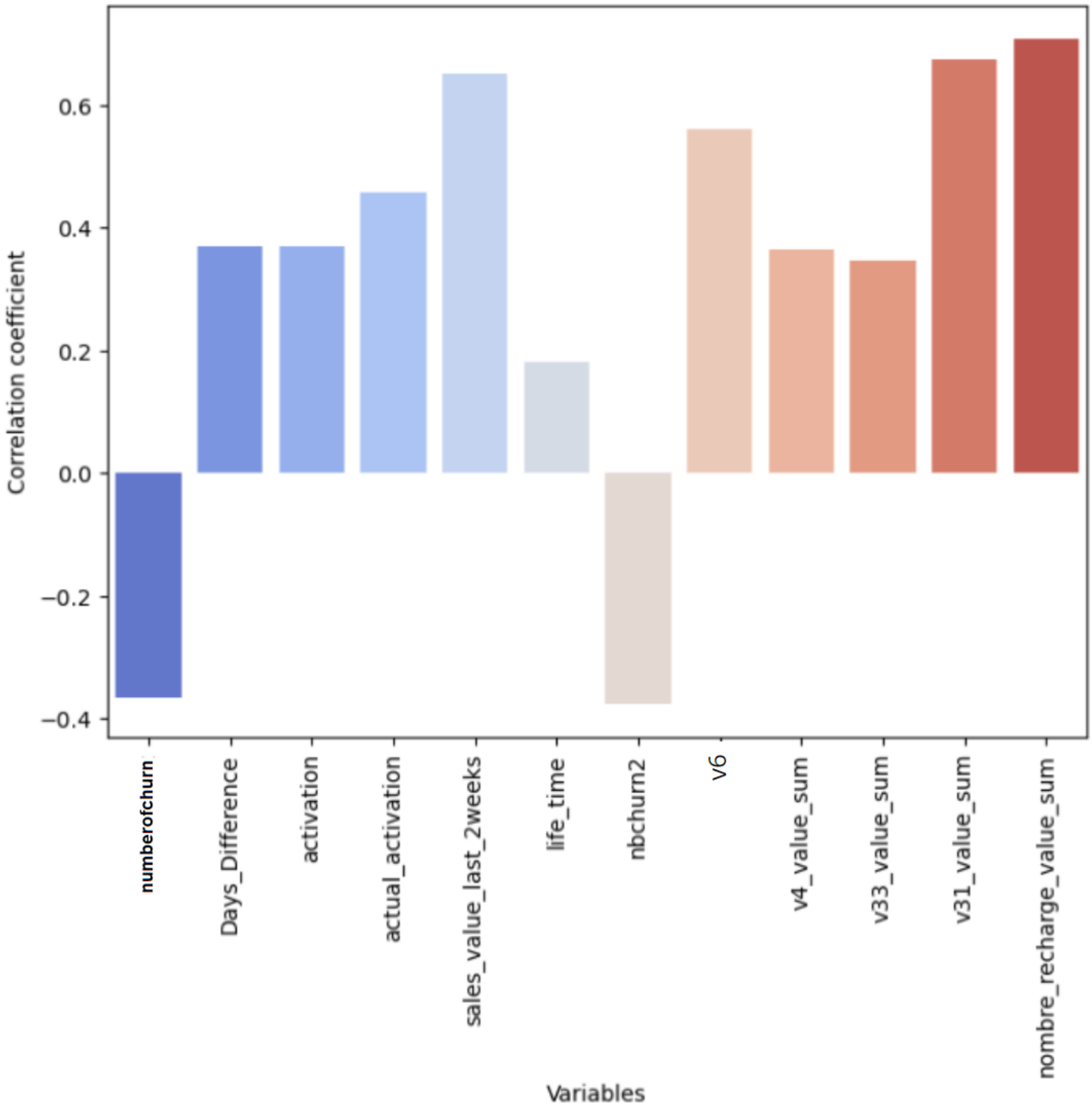
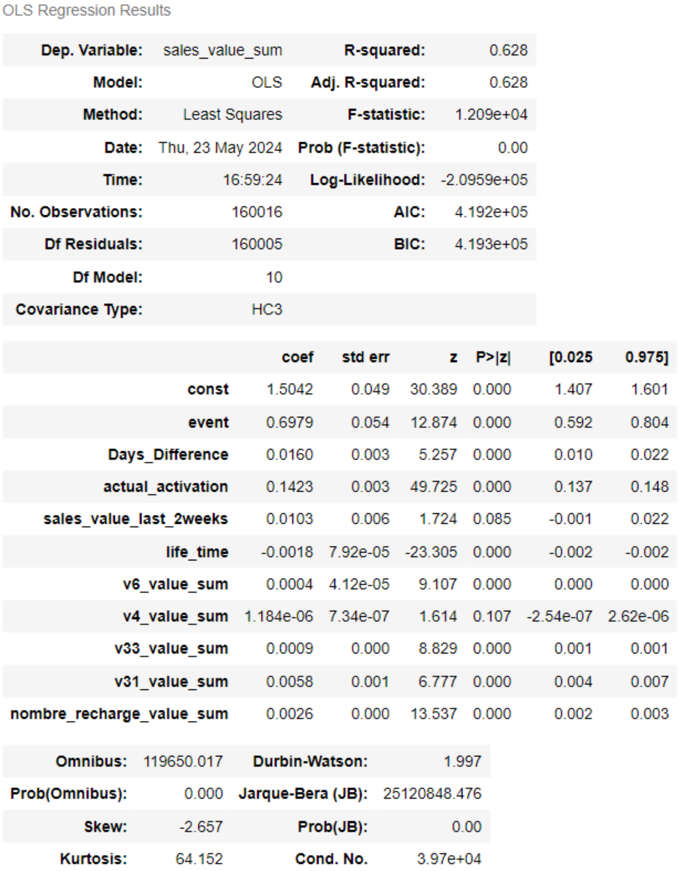


Figure 5.12: Correlation with target

The bivariate correlation coefficients with the target, the revenue, are sullarized in Figure 5.12. The variable `nombre_recharge_value_sum` is the most positive correlated variable with the revenue followed by the variables `Sales_value_last_2weeks`, `nuber_incoming_contacts`, `actual_activation` and the `activation` and the `Days_difference`. The number of historical churns has a negative correlation with the revenue. The survival time before churning (`Days_Difference`), activation, actual activation, total incoming call duration (`v4_value_sum`), and voice revenue have a moderate positive correlation with revenue. Finally, life time of the customer in the company has surprisingly very low positive correlation with revenue.

5.4.2 Linear Regression

The result of the Linear Regression of the logarithm, of the revenue on the covariates is depicted in Figure 5.13. The model includes various predictors to understand their impact on the sales value. Despite the heterogeneity of the customers, The model exhibits a good fit, with R squared of 0.62. Customer churn status, the survival time of customers (`Days_Difference`), and higher past activation levels (`actual_activation`) are all significantly linked to the logarithm of the revenue.



OLS Regression Results

Dep. Variable:	sales_value_sum	**R-squared:**	0.628
Model:	OLS	**Adj. R-squared:**	0.628
Method:	Least Squares	**F-statistic:**	1.209e+04
Date:	Thu, 23 May 2024	**Prob (F-statistic):**	0.00
Time:	16:59:24	**Log-Likelihood:**	-2.0959e+05
No. Observations:	160016	**AIC:**	4.192e+05
Df Residuals:	160005	**BIC:**	4.193e+05
Df Model:	10		
Covariance Type:	HC3		

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	1.5042	0.049	30.389	0.000	1.407	1.601
event	0.6979	0.054	12.874	0.000	0.592	0.804
Days_Difference	0.0160	0.003	5.257	0.000	0.010	0.022
actual_activation	0.1423	0.003	49.725	0.000	0.137	0.148
sales_value_last_2weeks	0.0103	0.006	1.724	0.085	-0.001	0.022
life_time	-0.0018	7.92e-05	-23.305	0.000	-0.002	-0.002
v6_value_sum	0.0004	4.12e-05	9.107	0.000	0.000	0.000
v4_value_sum	1.184e-06	7.34e-07	1.614	0.107	-2.54e-07	2.62e-06
v33_value_sum	0.0009	0.000	8.829	0.000	0.001	0.001
v31_value_sum	0.0058	0.001	6.777	0.000	0.004	0.007
nombre_recharge_value_sum	0.0026	0.000	13.537	0.000	0.002	0.003

Omnibus:	119650.017	**Durbin-Watson:**	1.997
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	25120848.476
Skew:	-2.657	**Prob(JB):**	0.00
Kurtosis:	64.152	**Cond. No.**	3.97e+04

Figure 5.13: Linear Regression model results

An additional increase of the `Days_difference` (expressed in months contribute to an increase of the revenue by 1.9%. An additional month of life_time is associated to an increase of the revenue by 0.16%. In the other hand, an additional recharge leads to an increase of the revenue by 0.26%. The other variables, including various value sums (`v31`, `v33`, `"nombre_recharge"`, etc.), are also associated to a positive impact on sales value (the revenue).

In order to check for a eventual overfit of the model we summarized the metrics for both train and test datasets. According to Table 5.1, the model has a mean squared error of 0.8 on the test data and 0.77 on the training data. This suggests that the model generalizes well to unseen data. The mean absolute error is 0.55 for test data and 0.63 for training data. Finally, the R-squared value is 0.62 for test data and and 0.63 for training data.

Metrics	Test	Train
Mean Squared Error	0.8	0.77
Root Mean Squared Error	0.89	0.87
Mean Absolute Error	0.55	0.63
R-squared	0.62	0.63

Table 5.1: Linear Regression Model evaluation

5.4.3 XGBoost

In order to best perform our model we use the algorithm XGBoost. The central idea behind this algorithm is to achieve an even higher R-squared value; in presence of q dataset characterised by a pronounced heterogeneity. This objective is achieved through a grid search, which identified the optimal hyper-parameters: 205 estimators, 5 maximum depths, and a learning rate of 0.1. This resulted in a very promising training R-squared of 0.9 instead of 0.6 obtained by Linear Regression. Table 5.2 shows that the train and test datasets metrics are equivalent. This is an argument that improve unseen dataset performance of the trained model.

Metrics	Train	Test
Mean Squared Error	0.51	0.49
R-squared	0.91	0.88

Table 5.2: XGBoost Model evaluation

Figure 5.14 show that the variable `v33_value_sum` is the most important feature followed by the variable `nombre_recharge_value_sum` and the variable `v6_value_sum`. The variable `life_time` arrives at the 4th rank and the variables `v4_value_sum` and `sales_value_las_2weeks` occupy respectively the 5th and 6th positions.

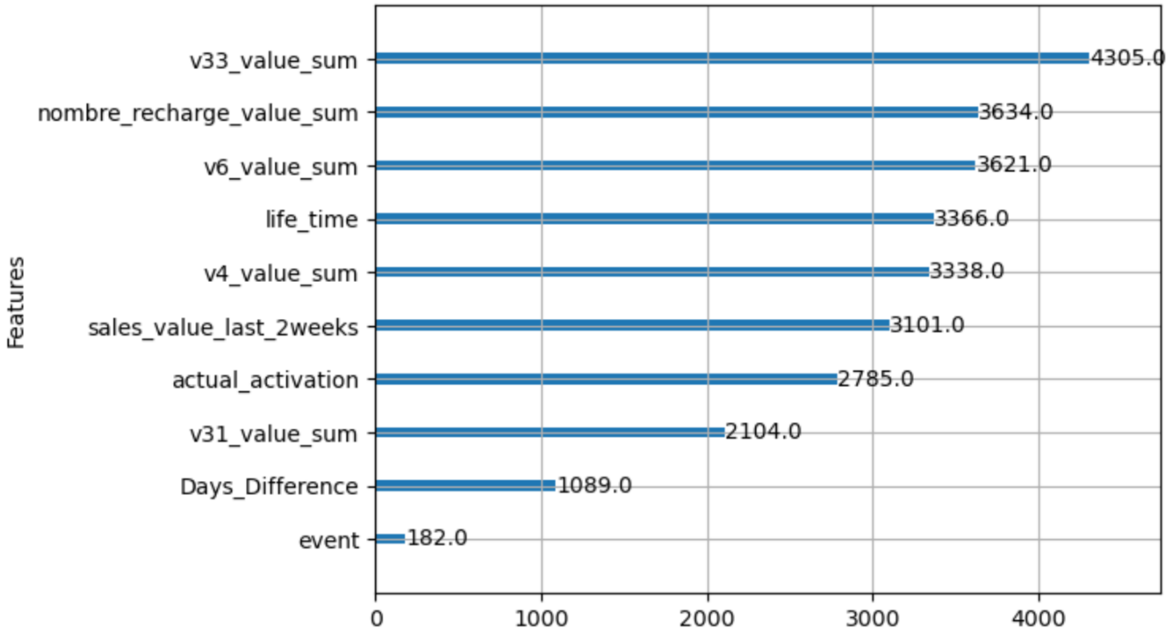
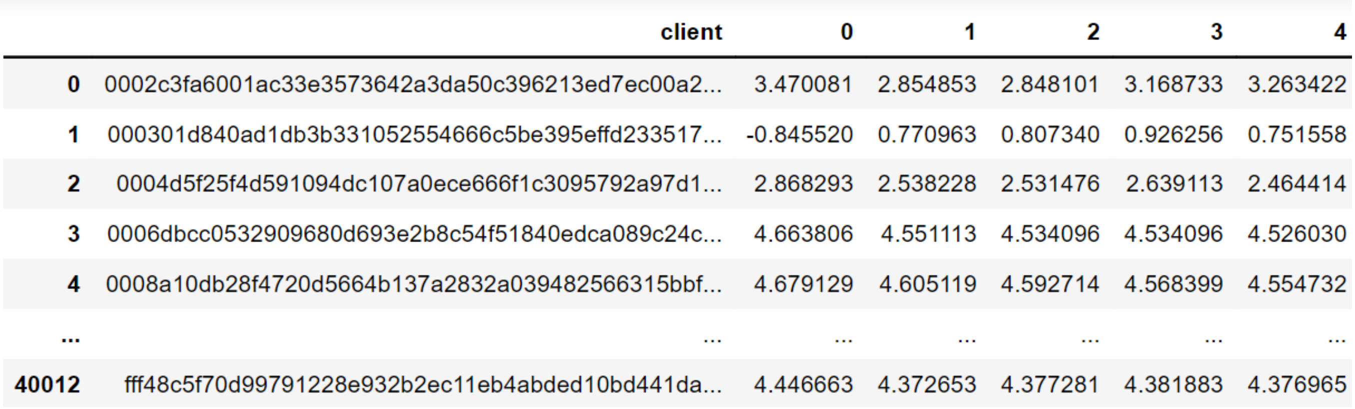


Figure 5.14: Feature importance of the XGBoost model

As in the Linear Regression modeling, the XGboost model can be used to obtain forecasts of the revenue values for different `Days_Differnce` values, other things are maintained fixed. Here is an example of what we have obtained (see Figure 5.15).



	client	0	1	2	3	4
0	0002c3fa6001ac33e3573642a3da50c396213ed7ec00a2...	3.470081	2.854853	2.848101	3.168733	3.263422
1	000301d840ad1db3b331052554666c5be395effd233517...	-0.845520	0.770963	0.807340	0.926256	0.751558
2	0004d5f25f4d591094dc107a0ece666f1c3095792a97d1...	2.868293	2.538228	2.531476	2.639113	2.464414
3	0006dbcc0532909680d693e2b8c54f51840edca089c24c...	4.663806	4.551113	4.534096	4.534096	4.526030
4	0008a10db28f4720d5664b137a2832a039482566315bbf...	4.679129	4.605119	4.592714	4.568399	4.554732
...	...	...	...	...	...	...
40012	fff48c5f70d99791228e932b2ec11eb4abded10bd441da...	4.446663	4.372653	4.377281	4.381883	4.376965

Figure 5.15: Revenue by month and customer

Then we combine the survival probabilities given by the cox model and the forecasted values of the revenue we are able to calculate the CLV[1] using Equation (2.1).

[1] We set δ to 0.08/12

5.5 Conclusion

The AFT and Cox models are powerful survival analysis tools that can help determine the influence of various features on the likelihood of customer retention. Our analysis revealed that the number of incoming contacts, customer seniority with the company, and engagement frequency all increase the likelihood of customer ”survival”.

Conversely, a high historical number of churn events significantly decreases the chance of staying. Additionally, we implemented two revenue prediction models: Linear Regression and XGBoost Regressor, achieving decent evaluation metrics with R-squared values of 0.6 and 0.9, respectively.

Finally, we combined the survival models and the revenue prediction models to calculate customer lifetime value (CLV).

Conclusion

The main objective of this end-of-studies report was the assessment of the Customer Lifetime Value (CLV) for the customers of Orange - Tunisia. To that, we combined two major modeling tasks: the survival model to predict the survival probability and the generalized models to predict the revenue.

We initiated our customer segmentation process by analyzing their recency, frequency, and monetary (RFM) scores to better assess their value and engagement. We performed two classifications for this purpose.

In the first classification, we divided customers into five segments based on specific RFM score ranges. The segments were: ”high value customers” (34% of the total), ”low value customers” (29%), and ”lost customers” (27%), with smaller segments comprising top customers and medium value customers.Although ”high value customers” contributed a substantial portion of the revenue, followed by ”top customers” and ”low value customers”. ”Lost customers” represented a notable segment in terms of count.

The second RFM classification involved binning recency, frequency, and monetary values into quantiles to create more granular segments. The analysis revealed that ”Champions” and ”Potential Loyalists” constituted a significant portion of the customer base.

Our journey began by estimating the CLV for short-term period (3 months), we used XGBoost classifiers and regressors to determine both the probability of being ”alive” (nonchurner) and the revenue generated per month. The best hyperparameters (learning rate: 0.05, maximum depth: 5, 100 estimators) resulted in an R-squared of 0.58 and a satisfactory RMSE, indicating good generalization to unseen data.

To further refine the model, we applied a logarithmic transformation to the revenue before training, achieving an RMSE of 0.811 and an R-squared of 0.77. The most important feature for revenue prediction was the amount of revenue generated in the last two weeks (of the period of the study) in TND, followed by the total revenue generated during the customer survival. For predicting the probability of being active, we used an XGBoost classifier with optimal parameters (learning rate: 0.1, maximum depth: 7, 300 estimators), resulting in an RMSE of 0.21 and an R-squared of 0.6. The most crucial predictor for survival probability was life_time, with the mean revenue generated during the customer survival, recency, and frequency also playing significant roles.

We also calculated the CLV by employing two advanced probabilistic models: the Beta Geometric model and the Gamma Gamma model. These models require key customer metrics: recency, frequency, monetary value, and tenure (T). The Beta Geometric model forecasted the likelihood of future customer engagement, while the Gamma Gamma model estimated the average transaction value. We applied the Beta Geometric model and using recency and frequency as inputs, we performed a temporal split of the data into calibration and holdout sets for model training and validation. The Beta Geometric model revealed insights into customer purchase behaviors and survival probabilities, highlighting high-value customers and identifying loyal versus transient customers. The model's accuracy was validated through RMSE and R-squared values. Overall, the analysis demonstrated the effectiveness of these models in predicting customer behavior and aiding in CLV assessment.

Secondly, we used survival analysis to estimate the probability of churn across different estimated segments. Kaplan-Meier curves showed a high initial survival rate, a steady decline in survival probability, a critical churn period around the 17-month mark, and stabilization at about 30% by 20 months. Survival curves by churn history indicated that customers with no past churn maintain high survival rates, while those with multiple churns have rapid declines. The Cox model revealed that increased customer activation frequency and longer tenure significantly reduce churn risk, while past churn behavior increases it, outgoing calls also reduce churn risk, but data usage has minimal impact on it. The model demonstrated high predictive accuracy. Analysis of survival probabilities across segments highlighted that "Low Value Customers" and "Top Customers" showed surprisingly high churn risks. These insights underscore the need for targeted retention strategies, particularly for high-value segments and customers with a churn history, to enhance retention and maximize CLV.

Then, we conducted an Accelerated Failure Time (AFT) model we found that higher customer activation frequency and longer tenure significantly increase survival time, indicating lower churn risk, while historical churn behavior strongly decreases survival time, highlighting higher churn risk. Incoming call frequency also positively impacted survival time, while outgoing calls had a minor negative impact, and data usage showed negligible influence.

After that, we performed a linear regression analysis on the logarithm of customer revenue. The model, which accounted for customer heterogeneity, showed a good fit with an R-squared value of 0.62. Significant predictors included customer churn status, survival time, and past activation levels. Specifically, an additional month of survival time contributed to a 1.9% increase in revenue, an extra month of customer life time increased revenue by 0.16%, and an additional recharge led to a 0.26% increase in revenue. On the test data the model had a mean squared error of 0.8. The mean absolute error was 0.55 and The R-squared values were 0.62.

We employed the XGBoost algorithm to estimate the revenue and through a grid search, we identified optimal hyperparameters: 205 estimators, a maximum depth of 5, and a learning rate of 0.1. This optimization yielded an impressive training R-squared of 0.91 compared to the 0.62 obtained from linear regression.

Feature importance analysis revealed that Total voice revenue in TND was the most significant predictor, followed by Total number of top-ups during the customer survival. By combining survival probabilities from the Cox model with the forecasted revenue values, we calculated the CLV.

Improving a Customer Lifetime Value (CLV) prediction project for a telecommunication company hinges significantly on the availability of comprehensive and detailed data. With enhanced access to **customer demographics** such as **age**, **gender**, **location**, and **income level**, we can better understand and segment our customer base.

Incorporating customer behavior data **browsing history on the company's website or app**, engagement level with **customer support**, and **feedback** can refine our predictions by highlighting individual preferences and **satisfaction levels**.

Enhanced marketing and engagement data, such as **responsiveness to promotions** can help tailor marketing efforts to maximize returns. **Financial metrics** like **gross margin**, **customer acquisition cost**, and **return on investment** provide insights into the profitability of each customer, aiding in better resource allocation.

Bibliography

[1] Guilherme Brandelli Bucco. *Development of a stochastic model to estimate customer value*. PhD thesis, Universidade Federal Do Rio Grande Do Sul, Escola Da Administrão, 2019.

[2] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[3] Tianqi Chen, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu Cho, Kailong Chen, Rory Mitchell, Ignacio Cano, Tianyi Zhou, et al. Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1(4):1–4, 2015.

[4] Richard Colombo and Weina Jiang. A stochastic rfm model. *Journal of Interactive Marketing*, 13(3):2–12, 1999.

[5] Cameron Davidson-Pilon. lifetimes documentation. https://readthedocs.org/projects/lifetimes/downloads/pdf/latest/, Jul 06, 2020. [Accessed 08-06-2024].

[6] Peter S. Fader and Bruce G. S. Hardie. The gamma-gamma model of monetary value. https://www.brucehardie.com/notes/025/gamma_gamma.pdf, 2013. [Accessed 28-03-2024].

[7] Peter S Fader, Bruce GS Hardie, and Ka Lok Lee. Rfm and clv: Using iso-value curves for customer base analysis. *Journal of marketing research*, 42(4):415–430, 2005.

[8] Nicolas Glady, Bart Baesens a b, and Christophe Croux a. A modified pareto/nbd approach for predicting customer lifetime value. *Expert Systems with Application*, 36(2), 2009.

[9] Frank E Harrell, Jr and Frank E Harrell. Cox proportional hazards regression model. *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis*, pages 475–519, 2015.

[10] David W Hosmer Jr, Stanley Lemeshow, and Susanne May. *Applied survival analysis: regression modeling of time-to-event data*, volume 618. John Wiley & Sons, 2008.

[11] Bingquan Huang, Mohand Tahar Kechadi, and Brian Buckley. Customer churn prediction in telecommunications. *Expert Systems with Applications*, 39(1):1414–1425, 2012.

[12] Dimple Kaul. Customer relationship management (crm), customer satisfaction and customer lifetime value in retail. *Review of Professional Management-A Journal of New Delhi Institute of Management*, pages 55–60, 2017.

[13] Katherine N Lemon and Tanya Mark. Customer lifetime value as the basis of customer segmentation: Issues and challenges. *Journal of Relationship Marketing*, 5(2-3):55–69, 2006.

[14] Serhat Peker and Özge Kart. Transactional data-based customer segmentation applying crisp-dm methodology: A systematic review. *Journal of Data, Information and Management*, 5(1):1–21, 2023.

[15] Jason T Rich, J Gail Neely, Randal C Paniello, Courtney CJ Voelker, Brian Nussenbaum, and Eric W Wang. A practical guide to understanding kaplan-meier curves. *Otolaryngology—Head and Neck Surgery*, 143(3):331–336, 2010.

[16] Amir Sorayaie Azar, Samin Babaei Rikan, Amin Naemi, Jamshid Bagherzadeh Mohasefi, Habibollah Pirnejad, Matin Bagherzadeh Mohasefi, and Uffe Kock Wiil. Application of machine learning techniques for predicting survival in ovarian cancer. *BMC medical informatics and decision making*, 22(1):345, 2022.

[17] Lee-Jen Wei. The accelerated failure time model: a useful alternative to the cox regression model in survival analysis. *Statistics in medicine*, 11(14-15):1871–1879, 1992.

[18] Ali ZARE. A comparison between accelerated failure-time and cox proportional hazard models in analyzing the survival of gastric cancer patients. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4645729/, 2015.