

**UNIVERSIDADE DE SÃO PAULO**

Instituto de Ciências Matemáticas e de Computação

## Fidelização e Retenção de Clientes em Modelos de Negócios Não Contratuais: Estimando a Probabilidade de Churn

**Vitor Wilgner Buzinaro**

Dissertação de Mestrado do Programa de Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria.

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: \_\_\_\_\_

**Vitor Wilgner Buzinaro**

## Fidelização e Retenção de Clientes em Modelos de Negócios Não Contratuais: Estimando a Probabilidade de Churn

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria. *VERSÃO REVISADA*.

Área de Concentração: Matemática, Estatística e Computação

Orientadora: Profa. Dra. Gleici da Silva Castro Perdoná

**USP – São Carlos**  
**Março de 2024**

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi  
e Seção Técnica de Informática, ICMC/USP,  
com os dados inseridos pelo(a) autor(a)

B992f Buzinaro, Vitor  
Fidelização e Retenção de Clientes em Modelos de  
Negócios Não Contratuais: Estimando a Probabilidade  
de Churn / Vitor Buzinaro; orientadora Gleici da  
Silva Castro Perdoná. -- São Carlos, 2024.  
66 p.

Dissertação (Mestrado - Programa de Pós-Graduação  
em Mestrado Profissional em Matemática, Estatística  
e Computação Aplicadas à Indústria) -- Instituto de  
Ciências Matemáticas e de Computação, Universidade  
de São Paulo, 2024.

1. Transformação Digital. 2. Negócios não  
contratuais. 3. Classificação de churn. 4. Análise  
de sobrevivência. I. da Silva Castro Perdoná,  
Gleici , orient. II. Título.

**Vitor Wilgner Buzinaro**

**Customer Loyalty and Retention in Non-Contractual  
Business Models: Estimating Churn Probability**

Master dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC- USP, in partial fulfillment of the requirements for the degree of the Master – Professional Masters in Mathematics, Statistics and Computing Applied to Industry. *FINAL VERSION.*

Concentration Area: Mathematics, Statistics and Computing

Advisor: Profa. Dra. Gleici da Silva Castro Perdoná

**USP – São Carlos  
March 2024**

*“N3o desistir 3 indispens3vel.”*

*Stephen Hawking*

## RESUMO

BUZINARO, V. **Fidelização e Retenção de Clientes em Modelos de Negócios Não Contratuais: Estimando a Probabilidade de Churn**. 2024. 66p. Dissertação (Mestrado – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

A transformação digital no varejo tem causado uma mudança profunda na forma como as empresas conduzem seus negócios, com a integração de tecnologias avançadas, como análise de dados e inteligência artificial, impactando a interação com os clientes. O *e-commerce* expandiu as fronteiras do varejo, oferecendo conveniência e acesso a uma variedade maior de produtos. A transformação digital tornou-se uma necessidade para empresas que desejam permanecer competitivas e atender às crescentes demandas dos consumidores por uma experiência de compra eficiente e personalizada.

Dentro desse contexto, identificar a saúde do cliente ao longo de seu ciclo de vida é essencial, uma vez que a desistência pode ocorrer de forma imprevisível. O modelo BG/NBD (*Beta-geometric Negative Binomial Distribution*) foi usado para prever a probabilidade de *churn* (desistência) dos clientes, com resultados significativos e boa capacidade preditiva. A análise de Kaplan-Meier foi usada para examinar o comportamento de compra e informações demográficas, destacando que clientes com cestas de produtos maiores tendem a permanecer mais tempo, enquanto aqueles com compras de alto valor têm relacionamentos menos duradouros.

Em resumo, o modelo BG/NBD foi eficaz na classificação de clientes ativos e inativos, e a análise de Kaplan-Meier forneceu informações valiosas sobre retenção e fidelização de clientes, capacitando a empresa a tomar decisões estratégicas com recursos da transformação digital.

**Palavras-chave:** Transformação Digital. Negócios não contratuais. Classificação de *churn*. Análise de sobrevivência.

## ABSTRACT

BUZINARO, V. **Customer Loyalty and Retention in Non-Contractual Business Models: Estimating Churn Probability**. 2024. 66p. Dissertation (Master's Degree – Professional Master's in Applied Mathematics, Statistics, and Computer Science for Industry) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

The digital transformation in retail has caused a profound change in how companies conduct their businesses, with the integration of advanced technologies such as data analysis and artificial intelligence impacting customer interactions. E-commerce has expanded the boundaries of retail, offering convenience and access to a wider range of products. Digital transformation has become a necessity for companies that want to remain competitive and meet the growing demands of consumers for an efficient and personalized shopping experience.

Within this context, identifying the customer's health throughout their lifecycle is essential, as churn can occur unpredictably. The BG/NBD model was used to predict the probability of customer churn, with significant results and good predictive capability. Kaplan-Meier analysis was used to examine purchasing behavior and demographic information, highlighting that customers with larger product baskets tend to stay longer, while those with high-value purchases have shorter-lasting relationships.

In summary, the BG/NBD model was effective in classifying active and inactive customers, and Kaplan-Meier analysis provided valuable information on customer retention and loyalty, empowering the company to make data-driven strategic decisions based on digital transformation capabilities.

**Keywords:** Digital transformation. Non contractual business settings. Churn classification. Survival analysis.

## LISTA DE FIGURAS

Figura 1 – Exemplo aplicado da função densidade probabilidade . . . . .	30
Figura 2 – Exemplo aplicado da Distribuição Gama . . . . .	31
Figura 3 – Exemplo aplicado da Distribuição Geométrica . . . . .	33
Figura 4 – Exemplo aplicado da Distribuição Beta . . . . .	34
Figura 5 – Transações no tempo . . . . .	34
Figura 6 – Probabilidade da primeira transação . . . . .	35
Figura 7 – Probabilidade da segunda transação . . . . .	35
Figura 8 – Matriz de Confusão . . . . .	40
Figura 9 – Estrutura da Análise Exploratória . . . . .	44
Figura 10 – Faturamento Mensal . . . . .	45
Figura 11 – Segmentação de Produtos Sazonais . . . . .	46
Figura 12 – Box plot e distribuição RFM . . . . .	46
Figura 13 – Bases de Treino e Teste . . . . .	48
Figura 14 – Curva ROC . . . . .	49
Figura 15 – Matriz de Confusão BG/NBD . . . . .	50
Figura 16 – Curva Empírica de Sobrevivência Kaplan-Meier . . . . .	52
Figura 17 – Curva de Sobrevivência Kaplan-Meier para Itens Distintos . . . . .	54
Figura 18 – Curva de Sobrevivência Kaplan-Meier para Ticket Médio . . . . .	55
Figura 19 – Curva de Sobrevivência Kaplan-Meier para Preço Médio . . . . .	56
Figura 20 – Curva de Sobrevivência Kaplan-Meier para Sazonalidade . . . . .	58



## LISTA DE TABELAS

Tabela 1 – Exemplo da base de dados Online Retail II . . . . .	28
Tabela 2 – Medidas de Resumo . . . . .	47
Tabela 3 – Ajuste do Modelo BG/NBD . . . . .	49
Tabela 4 – Métricas Classificação obtidas do ajuste do modelo BG/NBD . . . . .	51
Tabela 5 – Score Ticket Médio . . . . .	55
Tabela 6 – Preço Médio de Produtos por Sazonalidade . . . . .	57

## LISTA DE ABREVIATURAS E SIGLAS

AUC	Area Under The Curve
CDF	Função de Distribuição Acumulada
BG/BB	Beta-geometric Beta-binomial Model
BG/NBD	Beta-Geometric Negative Binomial Distribution Model
BTYD	Buy Till You Die
CLV	Customer Lifetime Value
CRM	Customer Relationship Management
E	Especificidade
FN	Falso Negativo
FP	Falso Positivo
IA	Inteligência Artificial
KM	Kaplan-Meier
MLE	Máxima Verossimilhança
MBG/NBD	Modified Beta-geometric Negative Binomial Distribution Model
NBD	Negative Binomial Distribution Model
Pareto/GGG	Pareto Gamma Gamma Gamma Model
Pareto/NBD	Pareto Negative Binomial Distribution Model
PDF	Função de Densidade de Probabilidade
PMF	Função de Massa de Probabilidade
RFM	Recência, Frequencia e Valor Monetário
ROC	Receiver Operating Characteristic
RPA	Robotic Process Automation
S	Sensibilidade
VN	Verdadeiro Negativo

VP	Verdadeiro Positivo
VPN	Valor Predito Negativo
VPP	Valor Predito Positivo

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>14</b>
<b>1.1</b>	<b>Objetivo</b>	<b>15</b>
1.1.1	Geral	15
1.1.2	Específicos	16
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>17</b>
<b>2.1</b>	<b>Impacto da Transformação Digital nas Empresas e Comportamento do Cliente: Criando um Ecossistema Digital</b>	<b>17</b>
<b>2.2</b>	<b>Evolução dos Modelos <i>Buy Till You Die</i></b>	<b>22</b>
<b>2.3</b>	<b>Análise de Sobrevivência na Investigação de <i>Customer Churn</i></b>	<b>25</b>
<b>3</b>	<b>METODOLOGIA</b>	<b>27</b>
3.0.1	Base de Dados	27
3.0.1.1	Estrutura da base de dados	27
<b>3.1</b>	<b>O Modelo BG/NBD</b>	<b>28</b>
3.1.1	Processo de Poisson para modelar transações e distribuição exponencial para modelar o tempo entre compras	29
3.1.2	Distribuição Gama para variação no comportamento de compra em uma população.	30
3.1.3	O Processo de Desistência	32
3.1.3.1	Distribuição Beta para descrever a variação na probabilidade de desativação	33
3.1.4	Implementação do Modelo e Estimação de Parâmetros	34
3.1.5	Função de Probabilidade de “Estar Vivo”	37
3.1.6	Limitações do Modelo	38
3.1.7	Avaliação do Modelo	39
3.1.7.1	Matriz de Confusão	39
3.1.7.2	Curva ROC e área sobre a curva (AUC)	41
3.1.7.3	Kolmogorov-Smirnov	41
<b>3.2</b>	<b>Análise de Sobrevivência</b>	<b>42</b>
3.2.1	Kaplan-Meier	43
<b>3.3</b>	<b>Estruturação do Código em Python</b>	<b>43</b>
<b>4</b>	<b>RESULTADOS</b>	<b>44</b>
<b>4.1</b>	<b>Análise Exploratória</b>	<b>44</b>
<b>4.2</b>	<b>Modelo BG/NBD</b>	<b>48</b>
<b>4.3</b>	<b>Análise de Sobrevivência Kaplan-Meier</b>	<b>52</b>

<b>5</b>	<b>CONCLUSÃO</b> . . . . .	<b>59</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>61</b>

## 1 INTRODUÇÃO

Em um mundo extremamente competitivo e com mercados saturados, para otimizar investimentos, atenuar custos, maximizando lucro e performance é preciso ter conhecimento e informações concretas para tomada de decisões assertivas. Com o avanço da tecnologia, uma abundância de registros são gerados, coletados e armazenados diariamente, dando vida ao conceito *Big Data*. Decorrente do grande volume de informações, se faz necessário uma análise microeconômica minuciosa e crítica com ênfase nas características da demanda, a fim de extrair os componentes essenciais para o processo decisório. Esta análise tem por objetivo contribuir para a formação das estratégias corporativas, que buscam fomentar a competitividade das empresas, desvendar padrões e tendências do comportamento do consumidor e do mercado (SHAW; STONE, 1990).

No que tange o setor de *marketing* digital atual, a construção de uma visão analítica por meio de diferentes indicadores é fundamental para garantir desenvolvimento em um ambiente competitivo. Dentre eles, se pode citar *customer churn* (NESLIN *et al.*, 2006).

A perda de clientes, conhecida como *customer churn* é um desafio crítico para empresas de diversas áreas. A gestão de relacionamento com o cliente voltada para vendas, se destaca por cultivar e manter os dados dos clientes. Desta forma, tornou-se uma das ferramentas mais importantes para combater e mitigar a desativação de consumidores (GANESH; ARNOLD; REYNOLDS, 2000). Sendo assim, cabe ao *customer relationship management* (CRM) explorar a lealdade dos clientes, criando uma experiência prazerosa e satisfatória por meio dos dados já adquiridos. O conhecimento das preferências dos usuários agrega vantagem competitiva à empresa, pois a conversão desses clientes se apresenta de forma menos custosa para a empresa em comparação com a prospecção de uma nova clientela (ARSLAN, 2020).

É preciso reter bons clientes evitando que estes se transformem em usuários inativos, esse movimento pode resultar em consequências financeiras significativas. Desta forma, nem todo *customer churn* é contratual, ou seja, nem sempre ocorre no final de um contrato formal e sim quando os clientes simplesmente abandonam uma empresa sem aviso prévio. Portanto, é fundamental desenvolver estratégias eficazes de detecção para evitar ou minimizar estas perdas. A detecção de *customer churn* não contratual é um problema complexo de análise de dados. Envolve a identificação de sinais ou indicadores de que um consumidor está prestes a abandonar uma empresa (XIAHOU; HARADA, 2022). A transformação digital oferece uma série de oportunidades para a detecção e prevenção dessa desativação aleatória. A coleta de dados em tempo real por meio de várias fontes, como interações do cliente em canais digitais, mídias sociais e dispositivos móveis, permite que as empresas tenham uma visão mais completa do comportamento dos clientes. Esses

dados podem ser analisados usando técnicas avançadas de análise de dados e algoritmos de aprendizado de máquina para identificar padrões e indicadores. Existem várias abordagens para tratar esse desafio. Uma estratégia comumente utilizada é a análise preditiva, em que modelos estatísticos são construídos com base em dados históricos para prever a probabilidade de *customer churn*.(AHN *et al.*, 2020).

A personalização e a segmentação de *marketing* são outras áreas-chave em que a transformação digital pode desempenhar um papel importante na redução de usuários inativos em um cenário não contratual. Por meio do uso de sistemas de gerenciamento de relacionamento e ferramentas de automação de *marketing*, podemos conduzir uma análise mais profunda de covariáveis em relação ao evento de *churn* (JAHROMI; STAKHOVYCH; EWING, 2014). Conectando com as modelagens de ciência de dados, as empresas podem identificar os clientes com maior probabilidade de abandonar seus produtos ou serviços. Isso permite que sejam implementadas ações proativas para retenção, como ofertas personalizadas, melhorias no atendimento ao cliente e programas de fidelidade. Além disso, através de canais digitais, como *e-mails*, aplicativos móveis e *chatbots*, as empresas podem se envolver com os clientes, fornecer suporte e obter *feedback* em tempo real. Essa comunicação contínua e bidirecional fortalece o relacionamento com os consumidores e a resolver problemas ou preocupações antes que levem-os a inatividade (PIEPPONEN *et al.*, 2022).

Dado este contexto, a presente pesquisa busca analisar e prever o *churn* de clientes baseando-se no comportamento transacional com o método BG/NBD (*Beta-geometric Negative Binomial Distribution*), que leva este nome por sua composição ter como base a distribuição binomial negativa. Além de explorar as covariáveis que auxiliam na detecção desse evento com análise de sobrevivência. Para tanto, é necessário identificar e extrair variáveis relevantes ao negócio para a construção dos modelos ao nível dos consumidores. Desta forma, iremos realizar a predição da propensão dos clientes serem *churners*, com informações de recência, frequência e a contribuição dos atributos comportamentais e demográficos para identificar o tempo médio para o fim do relacionamento entre cliente e empresa.

## 1.1 Objetivo

### 1.1.1 Geral

Estimar a probabilidade de um cliente terminar o relacionamento com uma empresa e analisar quais atributos demográficos e comportamentais estão associados ao tempo médio até o término. Assim, possibilitando criar ações de retenção para prolongar o vínculo por mais ciclos de consumo, por meio de estímulos personalizados à compra.

### 1.1.2 Específicos

São objetivos específicos deste trabalho:

- Tratar e adequar os dados para implementação dos modelos com análise exploratória, permitindo traçar o perfil dos consumidores.
- Prever e classificar grupos de clientes que estão propensos a abandonar o relacionamento com a empresa com o método BG/NBD.
- Identificar e definir quais as variáveis relevantes para a compreensão de *customer churn* com análise de sobrevivência.
- Sugerir iniciativas para conter a saída de clientes, com a finalidade de retê-los por mais ciclos de consumo, perpetuando sua fidelidade.



## 2 REFERENCIAL TEÓRICO

O presente capítulo irá se concentrar na construção do arcabouço bibliográfico que dará suporte a este trabalho. Seu objetivo é constituir de forma centralizada a consolidação do conhecimento necessário no processo elaborativo desta dissertação. Em primeiro lugar, aprofundaremos os efeitos da transformação digital nas empresas e no comportamento dos clientes, além de como essas mudanças se conectam para criar um ecossistema digital. Exploraremos como os avanços na tecnologia, no surgimento de plataformas digitais e nas expectativas dos consumidores, em constante evolução, têm catalisado uma mudança de paradigma no cenário empresarial. Em seguida, a apresentação abordará como a ciência de dados tem contribuído para entender a classificação entre clientes ativos e inativos (*Churn*), com foco em modelos probabilísticos da linhagem *Buy Till You Die* (BYTD). Por fim, será apresentado como os métodos de análise de sobrevivência contribuem para medir a duração esperada do relacionamento dos clientes com as empresas.

### 2.1 Impacto da Transformação Digital nas Empresas e Comportamento do Cliente: Criando um Ecossistema Digital

A transformação digital é impulsionada por vários fatores e pode ser definida em diversas configurações. Sendo assim, estudiosos têm oferecido diferentes definições para capturar a essência da transformação digital. Por exemplo, Parviainen *et al.* (2017) a define como "mudanças nas formas de trabalho, funções e ofertas comerciais causadas pela adoção de tecnologias digitais em uma organização, ou no ambiente operacional da organização". Da mesma forma, Hinings, Gegenhuber e Greenwood (2018) descrevem a transformação digital como "o efeito combinado de várias inovações digitais que trazem novos atores, estruturas, práticas, valores e crenças que mudam, ameaçam, substituem ou complementam as regras existentes dentro de organizações, ecossistemas, indústrias ou campos". Já Ven *et al.* (2021) sintetizam essa transformação como "a combinação intencional e abrangente de tecnologias digitais e competências nas operações e cultura de uma organização, a fim de otimizar o desempenho e alcançar objetivos estratégicos". Portanto, ao aproveitar as tecnologias digitais, as organizações podem responder rapidamente às mudanças de mercado, demandas dos clientes e tendências emergentes. Práticas ágeis, como desenvolvimento iterativo e implantação contínua, fomentam a inovação, experimentação, além de reduzir o tempo para lançamento no mercado (RODRÍGUEZ *et al.*, 2019).

A literatura recente enfatiza a natureza multidimensional da transformação digital ao identificar várias dimensões que contribuem para uma compreensão abrangente desse conceito. Por exemplo, Lacity e Willcocks (2014) delineiam três dimensões: tecnológica, organizacional e estratégica. A dimensão tecnológica envolve a adoção de ferramentas e

---

tecnologias digitais. A dimensão organizacional refere-se a mudanças na estrutura organizacional, processos e cultura para apoiar as iniciativas digitais. A dimensão estratégica concentra-se em alinhar os esforços de transformação digital com os objetivos estratégicos gerais da organização. Conseqüentemente, a transformação digital tem um impacto significativo em vários fluxos de valor dentro das empresas. Ao adotar tecnologias digitais, as organizações podem automatizar e otimizar operações, resultando em maior eficiência e economia de custos (BHARADWAJ *et al.*, 2013).

Ter uma visão clara e estratégica, alinhada aos objetivos do negócio é fundamental. A falta de uma estratégia digital abrangente pode prejudicar o progresso das iniciativas de transformação (KANE *et al.*, 2015). As organizações precisam de um roteiro e visão claros para orientar seus esforços. Sem uma estratégia bem definida, as empresas podem enfrentar iniciativas digitais desconexas, falta de alinhamento com os objetivos de negócio e dificuldade em mensurar o sucesso dos projetos de transformação digital (BRESCIANI *et al.*, 2021). Desenvolver e implementar uma estratégia digital robusta é essencial para superar esses desafios.

Essencialmente a transformação digital não está apenas em utilizar tecnologias mais avançadas, mas também em adotar um modelo estratégico onde o cliente está no centro das decisões. Desta forma, é possível compreender as necessidades dos clientes e desenvolver experiências personalizadas de acordo com suas preferências por meio dos dados, criando oportunidades para aumentar a vantagem competitiva em relação à concorrência (BERMAN, 2012). As empresas também estão reconhecendo a necessidade de uma abordagem centrada no cliente, pois leva a uma melhoria na satisfação, na fidelidade e na lucratividade a longo prazo, porém envolve repensar os processos de negócio, a cultura e os sistemas (KHANOM, 2023). Isso resulta em modificações fundamentais e estruturais na forma como uma empresa opera e gera valor para seus colaboradores e clientes. Essas modificações estão atreladas as constantes mudanças nas expectativas dos clientes, que geram demanda por experiências digitais fluídas moldando as iniciativas da transformação digital (KHANOM, 2023). Assim, os clientes esperam conveniência, tempos de resposta rápidos e soluções sob medida, levando as empresas a adotarem tecnologias digitais para atender estas necessidades em evolução (AKTER; D'AMBRA; RAY, 2013).

Colocar o cliente como centro da estratégia auxilia no rompimento da fronteira entre as experiências online e *offline*, dando origem ao conceito de jornada *omnichannel* do cliente. Gerea, Gonzalez-Lopez e Herskovic (2021) destacam o papel de uma abordagem *omnichannel* na transformação digital, permitindo que os clientes interajam com uma marca por meio de vários canais (por exemplo, *site*, aplicativo móvel, mídias sociais), mantendo o contexto e a continuidade. Uma estratégia *omnichannel* eficaz aprimora o envolvimento do cliente. Os clientes agora esperam uma experiência integrada em vários pontos de contato, seja navegando online, visitando uma loja física ou interagindo com o atendimento ao

---

cliente (VERHOEF; KANNAN; INMAN, 2015). O impacto da transformação digital no engajamento do cliente foi analisado por Hoyer *et al.* (2020), que abordaram a importância das interações e experiências que estes têm com uma marca por meio de pontos de contatos digitais, como *sites*, aplicativos móveis e *chatbots*. As organizações devem garantir consistência na mensagem, identidade visual e atendimento ao cliente em todos esses canais, a fim de proporcionar uma experiência coesa e satisfatória (LEMON; VERHOEF, 2016). A jornada do cliente não se limita mais às interações tradicionais em lojas físicas, mas se expandiu para englobar os pontos de contato digitais. Conforme afirmado por Bolton *et al.* (2018), iniciativas bem-sucedidas da transformação digital devem alinhar os investimentos em tecnologia com as expectativas e preferências dos clientes.

As organizações devem desenvolver um plano de ação para a transformação digital, definindo objetivos específicos, marcos e métricas de desempenho. A colaboração com *stakeholders* externos, incluindo fornecedores de tecnologia, *startups* e parceiros do setor, podem acelerar as iniciativas de transformação digital (WESTERMAN; BONNET; MCAFEE, 2015). Os autores Horlach *et al.* (2017) analisaram a aplicação de metodologias ágeis, como Agile, Scrum e DevOps, na transformação digital. Eles enfatizam a importância da colaboração interfuncional, desenvolvimento iterativo e melhoria contínua para aumentar a agilidade e a flexibilidade para construção de um ambiente orientado ao cliente. Da mesma forma, Loonam *et al.* (2018) destacam a importância da flexibilidade organizacional na transformação digital, permitindo que as organizações adotem novas tecnologias, processos e práticas centradas no cliente. Berman (2012) apontam que as organizações que adotam tecnologias digitais podem se diferenciar dos concorrentes oferecendo propostas personalizadas, melhorando as experiências dos clientes e otimizando seus projetos. A adaptabilidade é crucial para desenvolver rapidamente soluções para as preferências que são construídas e sobrepostas diariamente pela dinâmica do mercado (TERHO *et al.*, 2022).

Melhorar a eficiência operacional é um impulsionador fundamental da transformação digital, por conta da competição cada vez mais acirrada em diversas indústrias. As empresas não enfrentam somente os concorrentes tradicionais, mas também *startups* ágeis e empresas nativas digitais. Para se manterem relevantes e competitivas, as organizações reconhecem a necessidade de aproveitar as tecnologias digitais para diferenciar, melhorar a eficiência, criando modelos de negócios inovadores (TEECE, 2018). Portanto, as organizações buscam moldar seus processos, automatizar tarefas manuais e aproveitar a análise de dados para tomar decisões permitindo que as empresas otimizem a alocação de recursos, reduzam custos e aumentem a produtividade, oferecendo um forte incentivo para iniciativas de transformação digital (TRAN-DANG; KIM, 2021). A automação, a otimização de processos e a tomada de decisões orientada por dados permitem que as organizações reduzam erros manuais e alcancem economias de custo com tecnologias de processos robóticos (RPA) e monitoramento em tempo real, além de tempos de resposta mais rápidos (BU *et al.*, 2022). A

---

computação em nuvem permite infraestrutura escalável, alocação de recursos sob demanda e maior flexibilidade (BORANGIU *et al.*, 2019). Além disso, as tecnologias impulsionadas por inteligência artificial capacitam as organizações a extraírem insights valiosos, facilitando a tomada de decisões orientada por dados (KITSIOS; KAMARIOTOU, 2021). No estudo de Trawnih *et al.* (2022) revelaram que o impacto da inteligência artificial na personalização melhorou a experiência dos clientes, aumentou o engajamento e impulsionou as taxas de conversão. Esse achado destaca o potencial transformador da inteligência artificial na criação de jornadas do cliente personalizadas e sem interrupções.

Irawan, Susanto e Juliadi (2020) argumentam que a transformação digital também revolucionou a aquisição de clientes ao aproveitar os dados e *insights* dos clientes. As organizações agora podem personalizar suas ofertas, adaptando e otimizando seus produtos, serviços e mensagens de *marketing* para que os processos de nutrição de *leads* atendam às preferências e necessidades individuais desses membros que acabaram de se conectar com a empresa, melhorando as taxas de conversão (BLEIER; KEYSER; VERLEYE, 2018). Técnicas de personalização, como sistemas de recomendação e publicidade direcionada, tornaram-se parte integrante da melhoria da experiência e da promoção da fidelidade do cliente (LEMON; VERHOEF, 2016).

A transformação digital bem sucedida aprimora a eficiência operacional, as experiências dos clientes e fomenta a inovação, até mesmo em empresas de pequeno porte (CHEN *et al.*, 2021). Para tal, a implementação eficaz requer uma cultura organizacional de apoio e práticas eficazes de gestão da mudança, como foi apontado por Jaradat *et al.* (2022) em seu estudo de adoção de sistemas de inteligência de negócios no processo de tomada de decisão. As empresas que adotam a mentalidade digital em primeiro lugar, promovem habilidades digitais, estabelecem uma forte liderança (KURTI; HAFTOR, 2015). Para que isso ocorra, as organizações devem promover uma cultura de inovação, colaboração e aprendizado contínuo para adotar as tecnologias digitais (KANE *et al.*, 2015). As estratégias de gestão com comunicação clara, engajamento dos funcionários e programas de treinamento, são essenciais para garantir uma transição suave e minimizar a resistência à mudança (VARSHNEY, 2020). Schneider e Kokshagina (2021) argumentam que fomentar uma cultura de inovação, aprendizado e experimentação é essencial para navegar com sucesso na transformação digital, promover o envolvimento e construir relacionamentos de longo prazo.

Um dos principais desafios na transformação digital é a resistência à mudança dentro das organizações. Pesquisas sugerem que os funcionários podem relutar em adotar novas tecnologias, modificar fluxos de trabalho estabelecidos em adotar formas de trabalho digital (ALMATRODI; LI; ALOJAIL, 2023). Superar essa resistência requer estratégias eficazes de gestão, comunicação e a criação de uma cultura que incentive a abertura e a inovação. As organizações também precisam lidar com a resistência à mudança por parte

---

de outros *stakeholders*, além de gerenciar o impacto nos cargos e habilidades profissionais (MARTÍNEZ-PELÁEZ *et al.*, 2023).

Outro desafio da transformação digital é o processo de integração de diversas tecnologias e sistemas em diferentes unidades de negócio. Isso pode ser complexo e exigir recursos, especialmente com o surgimento de tecnologias disruptivas que impulsionam a transformação digital, como a Internet das Coisas, que criou oportunidades para as empresas aprimorarem suas operações, inovarem seus produtos e serviços e obterem vantagens competitivas (WIMELIUS *et al.*, 2021). O potencial transformador dessas tecnologias motiva as empresas a embarcarem em iniciativas de transformação digital. As organizações se esforçam para alinhar suas operações a esses requisitos, impulsionando a integração de soluções digitais em seus processos e sistemas (VOLBERDA *et al.*, 2021).

Uma equipe proficiente com as habilidades e capacidades adequadas para aproveitar as novas tecnologias de forma eficaz é essencial para o processo de digitalização (COLBERT; YEE; GEORGE, 2016). No entanto, as organizações frequentemente enfrentam lacunas de talento em termos de habilidades e conhecimentos digitais. Contratar e reter funcionários com as competências digitais necessárias, aprimorar as habilidades da força de trabalho existente e promover uma cultura de aprendizado são desafios cruciais na transformação digital (GOULART; LIBONI; CEZARINO, 2022). Além disso, a transferência de conhecimento é um ponto crítico para as empresas, que frequentemente lidam com o desafio de sistemas legados e infraestrutura desatualizada que dificultam os esforços de transformação digital. Sistemas legados podem não ser compatíveis com as novas tecnologias ou exigir modificações significativas para integrar soluções digitais. Modernizar e integrar sistemas legados pode ser complexo, demorado e custoso, representando desafios significativos para as organizações (MANCHA; SHANKARANARAYANAN, 2021).

O comportamento do cliente empoderado pela acessibilidade à informação passou por mudanças significativas, permitindo que eles tomem decisões mais informadas (LIANG; TURBAN, 2011). A capacidade de interagir com as empresas em tempo real também influenciou as expectativas dos clientes por respostas imediatas e serviços oportunos (CHOUDHURY; HARRIGAN, 2014). Os clientes agora esperam gratificação instantânea (HOFFMAN; NOVAK, 2018). O surgimento das redes sociais, das comunidades e serviços *online* também transformou a forma como os clientes descobrem, avaliam, compartilham informações e interagem sobre produtos e serviços com interfaces de inteligência artificial (IA) (KAPLAN; HAENLEIN, 2019). As plataformas de mídia social se tornaram canais influentes para os clientes expressarem suas opiniões, compartilharem experiências e interagirem com as marcas (HUANG; BENYOUCEF, 2013). Glucksman (2017) estudou o papel das plataformas de mídia social, do *marketing* de influência e do conteúdo personalizado na criação de conscientização de marca. Eles enfatizam que a transformação digital permite que as organizações alcancem um público mais amplo, personalizem a

mensagem e engajem os clientes de forma eficaz durante a fase de conscientização. As empresas precisam monitorar ativamente e responder aos *feedbacks* solicitados nas mídias sociais para manter a reputação da marca e construir a confiança do cliente (CHEN; FAY; WANG, 2011).

Apesar de seus benefícios, a transformação digital na jornada e empoderamento do cliente também apresenta vários desafios ligados a segurança das informações (KIM; KWON, 2019). No trabalho de Muralidhara (2017) são discutidos os desafios relacionados à privacidade e segurança dos dados, enfatizando a necessidade das organizações estabelecerem medidas robustas de proteção de dados e estarem em conformidade com regulamentações. Eles também destacam o desafio de gerenciar as expectativas dos clientes no mundo digital e a importância de manter a confiança. À medida que coleta e utiliza grandes quantidades de dados do cliente, considerações éticas e preocupações com a privacidade se tornaram proeminentes. Os clientes estão cada vez mais cientes das questões de privacidade de dados e exigem transparência e controle sobre suas informações pessoais (GRAEFF; HARMON, 2002). As organizações devem aderir às regulamentações de proteção de dados, garantir o armazenamento e transmissão seguros e obter consentimento explícito dos clientes em relação ao uso dos dados (MARTIN; MURPHY, 2017)

Outro desafio é a fase pós-compra, que envolve atividades como cumprimento do pedido, entrega e suporte pós-venda. Gong (2023) analisam como a transformação digital melhorou as atividades pós-compra por meio de tecnologias como automação da cadeia de suprimentos, rastreamento em tempo real e comunicações personalizadas. Eles destacam a importância de criar uma experiência pós-compra encantadora para aumentar a fidelidade do cliente.

À medida que observamos o processo evolutivo da transformação digital nas empresas e seu impacto no comportamento do cliente, torna-se claro que essa interseção é fundamental para o desenvolvimento de políticas de retenção e reativação eficazes. A contínua adaptação das organizações às mudanças tecnológicas e às demandas do mercado permite uma compreensão mais profunda das necessidades e preferências. Desta forma, serão exploradas algumas políticas resultantes desse processo no capítulo de resultados, tais como programas de fidelidade dinâmicos que se ajustam às preferências individuais dos clientes, estratégias de marketing personalizadas baseadas em dados comportamentais, iniciativas de reativação de clientes através de abordagens multicanais e investimentos em experiências digitais aprimoradas para aumentar o engajamento do cliente.

## 2.2 Evolução dos Modelos *Buy Till You Die*

Nos últimos 60 anos os modelos *Buy Till You Die* (BTYD) têm desempenhado um papel significativo na compreensão e previsão do comportamento de compra dos clientes. Esses modelos fornecem *insights* valiosos sobre o valor vitalício do cliente, dinâmica de

compras repetidas e segmentação de clientes. Ao longo dos anos, diversas variações e extensões dos modelos BTYD originais foram propostas, aprimorando sua precisão e aplicabilidade (MCCARTHY; WADSWORTH, 2014). Nesta seção, revisamos a evolução dos modelos BTYD e suas contribuições para o campo da análise de clientes.

O início da linhagem dos modelos BTYD se deu a partir do modelo *Negative Binomial Distribution* (NBD) que sinteticamente pode ser sumarizado como um modelo estatístico utilizado para analisar a frequência de eventos discretos, como compras, ao longo do tempo. O NBD descreve a probabilidade de observar um número específico de eventos até que um número predeterminado de sucessos seja alcançado. Esse modelo permite compreender as variações individuais nas taxas de compra e interação, sendo uma ferramenta valiosa para prever o comportamento futuro dos consumidores (EHRENBERG, 1959). Em termos gerais, o modelo NBD considera a aleatoriedade nas taxas de compra ou interação de um indivíduo ao longo do tempo, levando em conta a variação individual. O modelo é útil para entender a frequência de compras de diferentes consumidores e para prever quantas compras futuras podem ser esperadas de um determinado grupo, porém não leva em consideração a desativação dos clientes (EHRENBERG, 1959).

Com base no modelo NBD, Schmittlein, Morrison e Colombo (1987) introduziram o modelo *Pareto Negative Binomial Distribution* (Pareto/NBD) em 1987. Este modelo é uma estrutura estatística que combina elementos da distribuição de Pareto, para descrever a distribuição de clientes ativos, com o NBD, que captura a aleatoriedade nas taxas de compras de um cliente. O modelo de Pareto/NBD proporciona uma abordagem única para entender a fidelidade e a retenção de clientes, considerando tanto o número de compras realizadas quanto o tempo entre essas compras. Como afirmado por Schmittlein, Morrison e Colombo (1987), essa estrutura estatística oferece uma visão abrangente das interações dos clientes com uma empresa, permitindo previsões mais precisas das compras futuras e auxiliando nas estratégias de *marketing* direcionadas, por adicionar o processo de desistência de forma heterogênea, sendo considerado um dos principais modelos de sua classe.

Fader, Hardie e Lee (2005a) propuseram o modelo *Beta Geometric/Negative Binomial Distribution* (BG/NBD) em 2005, que expandiu o *framework* BTYD incorporando a distribuição Beta para modelar a heterogeneidade dos clientes nas probabilidades de compra. Esse modelo capturou tanto a rotatividade de clientes quanto a dinâmica de compras repetidas com uma abordagem matemática mais simples em relação ao modelo Pareto/NBD, tornando-se uma ferramenta valiosa e pela otimização de custo computacional.

O modelo *Modified Beta-geometric Negative Binomial Distribution* (MBG/NBD), proposto por Batislam, Denizel e Filiztekin (2007), e posteriormente desenvolvido por (HOPPE; WAGNER, 2007) no mesmo ano, estendeu o modelo NBD incorporando a desis-

tência do cliente e períodos de inatividade. Esse modelo proporcionou uma representação mais abrangente do comportamento do cliente, considerando tanto os padrões de compras repetidas quanto a rotatividade de clientes, permitindo adicionar clientes inativos.

Em 2010, Fader, Hardie e Shang (2010) propuseram o modelo *Beta-geometric Beta-binomial* (BG/BB), que estendeu o modelo Beta Geométrico incorporando uma distribuição beta para modelar a heterogeneidade nas probabilidades de compra. Esse modelo pressupõe que todo cliente sem uma transação repetida ainda não tenha terminado o relacionamento, independentemente do tempo decorrido de inatividade. Desta forma, o modelo capturou a rotatividade de clientes e a dinâmica de compras repetidas, muito demandadas em estratégias de retenção de clientes e campanhas de *marketing*.

Por fim, em 2016, Platzner e Reutterer (2016) introduziram o modelo *Pareto Gamma Gamma Gamma* (Pareto/GGG). Este modelo incorporou a distribuição Gama Generalizada para capturar a heterogeneidade na frequência de compra dos clientes, ajustando a variação no grau de regularidade dentro dos momentos das transações. Forneceu um *framework* flexível para modelar o comportamento do cliente e permitiu uma segmentação precisa dos clientes e previsões.

Em geral, a evolução dos modelos BTYD contribuiu para uma compreensão mais profunda do comportamento de compra dos clientes e ampliou a forma com que as empresas tomam decisões baseadas em dados. Esses modelos incorporaram a heterogeneidade, efeitos específicos e componentes adicionais para capturar as complexidades do comportamento do cliente. Os avanços nos modelos BTYD melhoraram sua precisão, capacidade preditiva e aplicabilidade em várias indústrias (MCCARTHY; WADSWORTH, 2014).

Para o desenvolvimento deste trabalho, o modelo BG/NBD foi escolhido devido à sua capacidade de lidar com características específicas do comportamento do cliente, como compras repetidas e inatividade entre compras. Ele é particularmente adequado para empresas com uma base de clientes relativamente grande e heterogênea, onde os padrões de compra podem variar significativamente entre os indivíduos. O BG/NBD também é flexível o suficiente para acomodar diferentes cenários de retenção de clientes, incluindo clientes que podem fazer compras esporádicas ao longo do tempo. Além disso, o modelo oferece interpretabilidade, permitindo que as empresas compreendam melhor os padrões de compra de seus clientes e identifiquem estratégias eficazes de *marketing* e retenção. Sua capacidade de prever o tempo até a próxima compra e a probabilidade de churn fazem do BG/NBD uma escolha valiosa. Ademais, é um modelo amplamente utilizado e com sólida literatura acerca de sua implementação e consistência de resultados.



### 2.3 Análise de Sobrevivência na Investigação de *Customer Churn*

A perda de clientes representa um desafio significativo para empresas em várias indústrias. Compreender os fatores que contribuem para a perda de clientes e prever com precisão eventos de perda é essencial para desenvolver estratégias eficazes de retenção. Técnicas de análise de sobrevivência fornecem informações valiosas ao considerar o tempo médio até a perda como um evento de interesse. Este capítulo explora a aplicação de técnicas de análise de sobrevivência no contexto da análise de perda de clientes.

O estimador de Kaplan-Meier, introduzido por Kaplan e Meier (1958), é amplamente utilizado na análise de perda de clientes. Ele permite às empresas estimar a probabilidade de sobrevivência do cliente (ou retenção) ao longo do tempo, levando em conta a censura. Por exemplo, Masarifoglu e Buyuklu (2019) aplicaram o estimador de Kaplan-Meier para analisar a perda de clientes na indústria de telecomunicações, fornecendo *insights* valiosos sobre os padrões de retenção de clientes. Além disso, Larivière e Poel (2004) utilizaram o estimador para examinar a perda de clientes no setor bancário, iluminando as probabilidades de sobrevivência de diferentes segmentos de clientes.

O teste de *log-rank*, proposto por Mantel (1966), é comumente utilizado para comparar padrões de perda de clientes entre diferentes grupos de clientes. Ele ajuda a determinar se existem diferenças significativas no comportamento de perda de clientes com base em características específicas. Por exemplo, Ochola (2019) aplicaram o teste de *log-rank* para comparar as taxas de perda de clientes entre clientes de alto e baixo valor na indústria de comércio eletrônico, revelando padrões distintos de perda entre esses segmentos. Em outro estudo, Toccu, Fasso *et al.* (2013) utilizaram o teste de *log-rank* para examinar diferenças de perda de clientes entre diferentes segmentos pautados em regiões geográficas no setor de telecomunicações.

O modelo de riscos proporcionais de Cox, introduzido por Cox Cox (1972), é uma ferramenta poderosa para a previsão de perda de clientes, incorporando múltiplas covariáveis. Vários estudos têm utilizado esse modelo para entender o impacto de diversos fatores na perda de clientes. Por exemplo, Periáñez *et al.* (2016) utilizaram o modelo de Cox para analisar a perda de clientes na indústria de jogos *online* e identificaram preditores significativos de perda, como atividade de jogos e interação social. Em outro estudo, Chen *et al.* (2022) aplicaram o modelo de Cox para prever a perda de clientes no setor de seguros, incorporando dados demográficos dos clientes, características das apólices e histórico de sinistros como covariáveis.

Além das técnicas fundamentais, métodos avançados de análise de sobrevivência têm sido utilizados na investigação de perda de clientes. Por exemplo, Raja *et al.* (2021) utilizaram modelos de fragilidade para examinar a perda de clientes na indústria de telecomunicações móveis, levando em conta a heterogeneidade não observada entre os

clientes. Além disso, Mena *et al.* (2023) aplicaram modelos de covariáveis variantes no tempo para estudar a perda de clientes no setor de comércio eletrônico, capturando a natureza dinâmica do comportamento do cliente e seu impacto no risco de perda.

Dado este contexto referente os métodos de análise de sobrevivência, o método (KM) Kaplan-Meier é uma escolha sólida para analisar churn e identificar marcos de retenção e reativação de clientes. Ao aplicar o método KM a dados de churn, podemos estimar a probabilidade de um cliente permanecer ativo ao longo do tempo, fornecendo insights valiosos sobre a taxa de retenção ao longo do ciclo de vida do cliente. Além disso, o Kaplan-Meier é robusto em lidar com dados censurados, o que é comum em análises de churn, onde nem todos os clientes saem do serviço durante o período de observação. Isso permite que o método capture com precisão a dinâmica do churn, mesmo quando há clientes que permanecem ativos além do período de observação. Ao identificar os pontos de inflexão na curva de sobrevivência estimada pelo Kaplan-Meier, podemos destacar marcos importantes de retenção e reativação de clientes, permitindo que as empresas identifiquem estratégias eficazes para reduzir o churn e maximizar o valor do cliente ao longo do tempo, como iremos pontuar na seção de resultados.

### 3 METODOLOGIA

Neste capítulo apresentaremos a metodologia adotada neste trabalho, discutiremos a base de dados utilizada, descrevendo suas características e a forma como foi coletada e organizada para a análise. Por fim, abordaremos o modelo BG/NBD amplamente utilizado no campo da pesquisa de retenção de clientes (BTYD), análise de sobrevivência Kaplan-Meier e as métricas de avaliação dos modelos.

#### 3.0.1 Base de Dados

O conjunto de dados analisado neste trabalho foi obtido a partir do repositório da Universidade da Califórnia Irvine *Machine Learning Repository*, intitulado por *Online Retail II*<sup>1</sup>. Este conjunto de dados é composto por todas as transações que ocorreram em uma empresa de varejo *online* puramente digital sediada no Reino Unido. Estas transações ocorreram no período entre 01 de dezembro de 2009 e 09 de dezembro de 2011. A empresa em questão se especializa na venda de produtos de presentes variados adequados para todas as ocasiões. É importante notar que muitos dos clientes dessa empresa são atacadistas, o que significa que eles compram produtos em grande quantidade para revenda. Portanto, vale destacar que é uma base pautada em modelo de negócio não contratual, ou seja, não há acordos ou contratos vinculativos que estabeleçam relações formais ou obrigações legais entre as partes envolvidas. Em outras palavras, as transações comerciais ocorrem de maneira informal e sem a necessidade de acordos legais ou contratos formalizados, o que implica em uma desistência aleatória. Isso faz com que o tema de classificação de *churn* ganhe relevância para esse tipo de negócio.

##### 3.0.1.1 Estrutura da base de dados

A base contém informações transacionais em nível de nota fiscal, conta com detalhes de registro de cada item, como preço, quantidade, país da transação, data de emissão da nota. Esses dados podem ser úteis para entender os padrões de compra por região, sazonalidade, preferência e comportamento de compra dos clientes, a fim de auxiliar no desenvolvimento de métodos para detecção das inativações. Exemplo da base de dados é apresentada conforme a Tabela 1.

Sendp cada variável:

1. InvoiceNo: Número de identificação exclusivo para cada fatura ou transação.
2. StockCode: Código de identificação único para cada produto ou item vendido.

---

<sup>1</sup> <http://archive.ics.uci.edu/dataset/502/online+retail+ii>

Tabela 1 – Exemplo da base de dados Online Retail II

Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	01/12/2009 07:45	6.95	13085.0	United Kingdom
489434	79323P	PINK CHERRY LIGHTS	12	01/12/2009 07:45	6.75	13085.0	United Kingdom
489434	79323W	WHITE CHERRY LIGHTS	12	01/12/2009 07:45	6.75	13085.0	United Kingdom
489434	22041	RECORD FRAME 7*SINGLE SIZE	48	01/12/2009 07:45	2.1	13085.0	United Kingdom
...	...	..	...	...	...	...	...

**Fonte:** Elaborada pelo autor.

3. Description: Descrição do produto ou item.
4. Quantity: Quantidade de produtos vendidos em cada transação.
5. InvoiceDate: Data e hora da transação.
6. UnitPrice: Preço unitário de cada produto em libras.
7. CustomerID: Número de identificação único para cada cliente.
8. Country: País onde a transação foi registrada.

A base de dados em questão contém um total de 5942 clientes e 53628 faturas. Ao longo dessas transações, um total de 10.608.492 produtos foram vendidos, refletindo a diversidade e o volume das operações comerciais registradas. A receita total gerada alcança a expressiva marca de 19.287.250 libras esterlinas, denotando a relevância econômica das atividades comerciais documentadas. Além disso, a base de dados destaca-se pela sua riqueza em diversidade, com um total de 4070 produtos únicos registrados, evidenciando a amplitude do catálogo disponibilizado aos clientes.

### 3.1 O Modelo BG/NBD

Neste seção, fornecemos uma metodologia detalhada para analisar o comportamento de compra dos clientes usando o modelo BG/NBD, conforme proposto por Fader, Hardie e Lee (2005a). Explicamos os princípios subjacentes do modelo, seus principais componentes e a estimação dos parâmetros do modelo.

As informações obtidas ao aplicar o modelo BG/NBD podem ser aproveitadas em várias aplicações empresariais. Por exemplo, as empresas podem usar o modelo para prever o comportamento futuro do cliente, estimar o valor vitalício do cliente ou *customer lifetime value* (CLV) e probabilidade condicional em estar vivo. O CLV é uma métrica crucial que quantifica a receita esperada gerada por um cliente ao longo de todo o relacionamento com a empresa. Ao estimar o CLV, as empresas podem tomar decisões sobre estratégias de aquisição e retenção de clientes, alocação de recursos e estratégias de precificação.

Além disso, o modelo pode ajudar a identificar clientes de alto valor e direcioná-los com esforços de *marketing* personalizados. Ao entender os fatores que influenciam

o comportamento de compra do cliente, as empresas podem projetar campanhas de *marketing* eficazes, programas de fidelidade e recomendações personalizadas para melhorar a satisfação e a fidelidade do cliente.

O modelo BG/NBD também pode ser usado para avaliar o impacto de intervenções ao incorporar atividades de *marketing* ao modelo, como ofertas promocionais, assim, as empresas podem avaliar a eficácia dessas intervenções em termos de geração de receita. Estas informações podem orientar futuros investimentos em *marketing* e otimizar as estratégias. O modelo BG/NBD oferece uma estrutura poderosa para analisar os padrões de compra dos clientes e prever compras futuras. Ao combinar dois processos principais, o processo de transação e o processo de desistência, o modelo fornece informações sobre o valor vitalício e a rotatividade de clientes.

Desta forma, o método probabilístico BG/NBD assume que as transações dos clientes seguem um processo de Poisson, e a desistência do cliente segue uma distribuição exponencial.

### 3.1.1 Processo de Poisson para modelar transações e distribuição exponencial para modelar o tempo entre compras

Primeiramente, vamos focar nos comportamentos de compras repetidas de clientes ativos. Podemos assumir que, enquanto um cliente permanece ativo, suas transações seguem um processo de Poisson com uma taxa constante de compras  $\lambda$ , note que esta é uma limitação, pois em um cenário real a frequência de compra dos clientes oscila ao longo do tempo. Com essa suposição, podemos modelar o tempo até a próxima compra  $\Delta t$  como uma distribuição exponencial parametrizada por  $\lambda$ . A função de densidade de probabilidade (PDF) dessa distribuição é a seguinte:

$$f(\Delta t) = \lambda e^{-\lambda \Delta t} \quad (3.1)$$

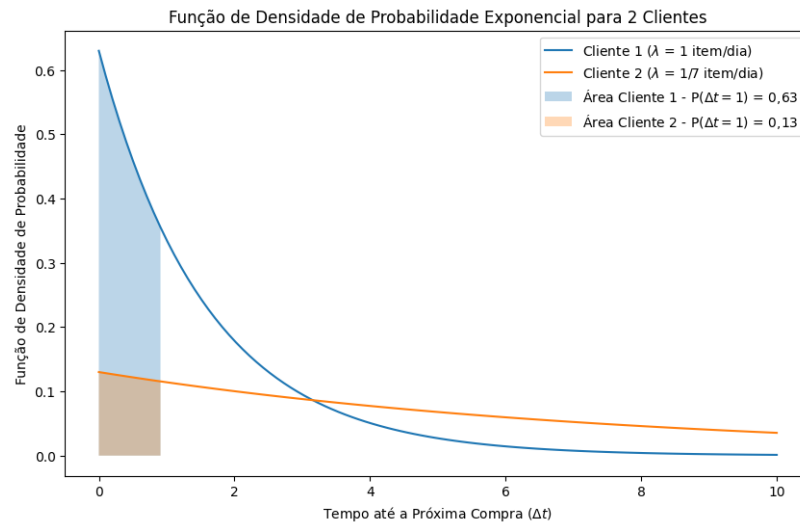
em que:

- $\Delta t$ : Tempo até a próxima compra;
- $\lambda$ : Taxa de compras;
- $e$ : Número de Euler.

Cada cliente ativo terá sua própria distribuição exponencial que podemos usar para prever a probabilidade do tempo da próxima compra (ver Figura 1).

O gráfico mostra a PDF de duas distribuições exponenciais associadas a dois clientes. O primeiro cliente (a curva azul) normalmente compra um item todos os dias (sua

Figura 1 – Exemplo aplicado da função densidade probabilidade



**Fonte:** Elaborada pelo autor.

taxa de compra  $\lambda$  é de 1 item/dia). A probabilidade de que sua próxima compra ocorra dentro do próximo dia é dada pela área sob a curva azul entre 0 e 1, sendo 0,63. O segundo cliente compra um item a cada semana (sua  $\lambda$  é de 1/7 item/dia). Vemos que é muito menos provável que sua próxima compra ocorra no próximo dia, pois a probabilidade é 0,13.

### 3.1.2 Distribuição Gama para variação no comportamento de compra em uma população.

É útil pensar que todos esses clientes com suas diferentes  $\lambda$ 's contribuem para uma distribuição lambda em toda a empresa. Nossa tarefa agora é modelar essa distribuição  $\lambda$ , para flexibilizar a suposição que a distribuição Poisson exige quanto a taxa de compra constante ao longo do tempo do cliente. Ao fazer isso, precisaremos atender aos seguintes requisitos:

- A distribuição deve preferencialmente ser bem estudada.
- Uma vez que  $\lambda$  só pode ter valores em números reais positivos, a distribuição escolhida deve ter apenas valores positivos.
- A distribuição precisa ser flexível o suficiente para modelar diferentes bases de clientes com diferentes comportamentos de compra.

A distribuição Gama atende a todos esses requisitos; é a que é usada no BG/NBD para modelar  $\lambda$ . Ela é parametrizada pelo parâmetro de forma  $r$  e pelo parâmetro de escala  $\alpha$ . Diferentes combinações desses dois parâmetros resultam na distribuição Gama assumindo formas distintas. Portanto, a PDF da distribuição:

$$f(\lambda|r, \alpha) = \frac{\alpha^r \lambda^{r-1} e^{-\lambda\alpha}}{\Gamma(r)}, \quad \lambda > 0 \quad (3.2)$$

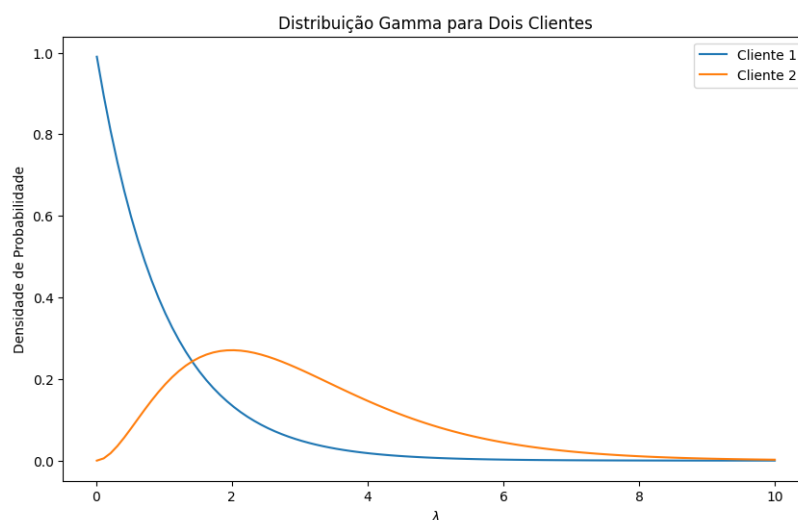
onde:

- $\alpha$ : Parâmetro de escala da distribuição Gama. Controla a amplitude e a média da taxa de compra;
- $r$ : Parâmetro de forma da distribuição Gama. Influencia a forma da distribuição, ajustando-a a diferentes padrões de taxa de compra.
- $\lambda$ : Taxa de compra que estamos modelando;
- $\Gamma(r)$ : Função Gama, usada para normalização. Mantém a integral da PDF adequada;
- $e$ : Número de Euler.

Heterogeneidade em  $\lambda$  segue uma distribuição Gama, evidenciando que ao longo da atividade do cliente dentro da empresa ele realiza transações que são distribuídas aleatoriamente.

É importante observar que esta distribuição Gama não é apenas um jargão teórico. Na verdade, uma distribuição Gama específica é uma descrição quantitativa do comportamento de compra coletiva de uma base de clientes específica, o que tem implicações comerciais, observe na Figura 2.

Figura 2 – Exemplo aplicado da Distribuição Gama



**Fonte:** Elaborada pelo autor.

Por exemplo, a linha azul mostra uma distribuição Gama inclinada para baixo e inclinada para a esquerda, resultante ao definir lambda e alfa iguais a 1. Se esta distribuição

correspondesse à minha base de clientes, a forte assimetria à esquerda indica que a maioria dos clientes tem uma taxa de compra próxima de zero. Ou seja, eles quase não compram nenhum item. Outra distribuição Gama é mostrada em laranja. Essa é uma distribuição mais saudável na qual o  $\lambda$  atinge o pico em torno de 2, o que significa que uma parte considerável da população compra dois itens por dia.

À medida que aumentamos a flexibilidade na modelagem da taxa de compra com a distribuição gama, a distribuição de Poisson pode não ser suficiente para capturar a variabilidade nos dados. Nesse ponto, nos aproximamos de uma distribuição que se assemelha à distribuição binomial negativa. A distribuição binomial negativa é mais adequada para modelar processos em que o número de eventos (compras) até que um certo número de falhas (ou churn) seja alcançado é relevante.

### 3.1.3 O Processo de Desistência

Agora vamos lidar com o processo de desativação dos clientes. Como mencionado anteriormente, após cada compra, os clientes terão uma probabilidade  $p$  atribuída para a ruptura do relacionamento com a empresa. Consequentemente, a transação após a qual um cliente se desativa é distribuída de acordo com a distribuição geométrica. A função de massa de probabilidade (PMF) dessa distribuição discreta é:

$$P(\text{número de compras até a desativação} = x) = p(1 - p)^{x-1}, \quad x = 1, 2, 3, \dots \quad (3.3)$$

em que:

- $p$ : Probabilidade de desativação após cada compra;
- $1 - p$ : Probabilidade de não desativação após cada compra;
- $x$ : Número de transações até a desativação.

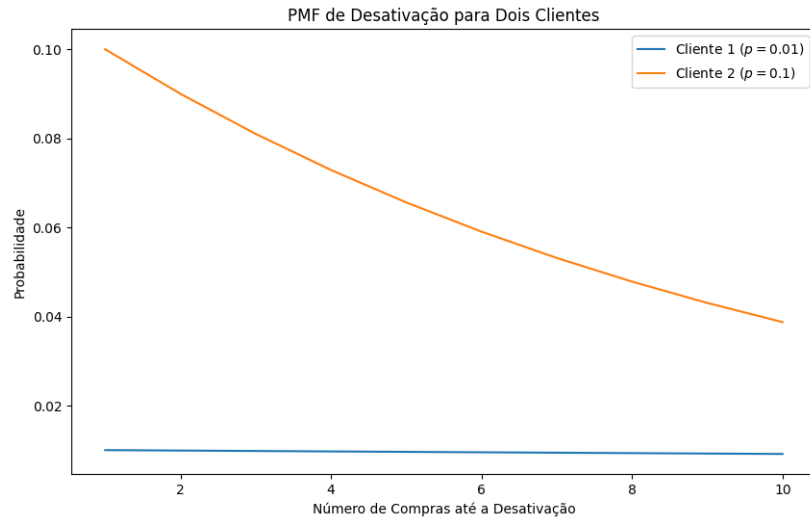
Essa PMF indica se um cliente se desativa somente após a  $x$  transação, eles devem ter sobrevivido às  $x - 1$  transações anteriores. Cada uma dessas sobrevivências carrega a probabilidade  $(1 - p)$ , daí o componente  $(1 - p)^{x-1}$ .

Assim, o processo de desativação de um cliente no modelo BG/NBD utiliza a probabilidade de desativação após cada compra e a distribuição geométrica para capturar o comportamento de desativação ao longo do tempo. Veja o exemplo na Figura 3 que compara dois clientes com  $p = 0.01$  e  $p = 0.1$ .

Por definição, um cliente deve realizar pelo menos uma transação antes de se desativar. Quanto maior o valor de  $p$ , mais provável é que a desativação ocorra mais cedo. Podemos observar que o cliente com  $p = 0.01$  (linha azul) tem uma probabilidade muito menor de se desativar cedo em comparação com o cliente com  $p = 0.1$  (linha laranja).



Figura 3 – Exemplo aplicado da Distribuição Geométrica



**Fonte:** Elaborada pelo autor.

### 3.1.3.1 Distribuição Beta para descrever a variação na probabilidade de desativação

Assim como  $\lambda$ , é útil considerar que uma população de clientes está associada a uma distribuição de  $p$ . No entanto, desta vez, não podemos usar a distribuição Gama, que não tem limite superior. Precisaremos de outra distribuição que seja igualmente flexível, mas cujo intervalo de valores seja de 0 a 1 (porque  $p$  pode variar apenas entre 0 e 1).

A distribuição Beta atende às nossas necessidades. A PDF da distribuição Beta:

$$f(p|a, b) = \frac{p^{a-1}(1-p)^{b-1}}{B(a, b)}, \quad 0 \leq p \leq 1, \quad (3.4)$$

em que:

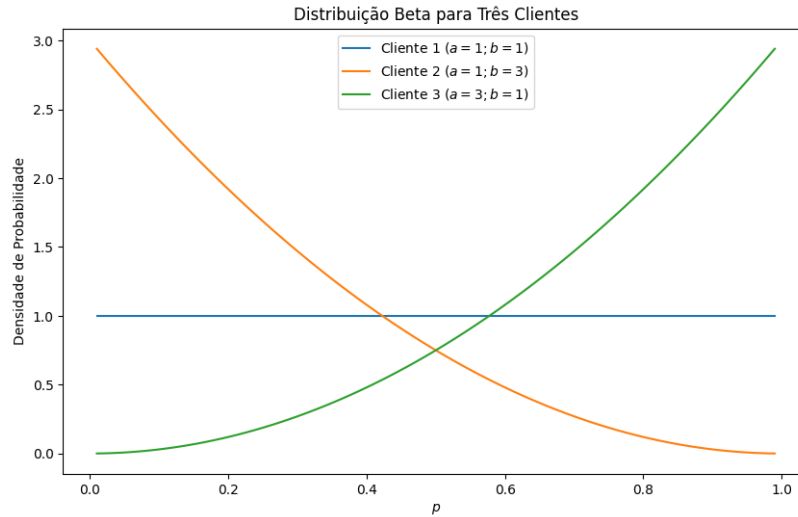
- $a$ : Parâmetro de forma da distribuição Beta. Controla a forma da distribuição e influencia a probabilidade de desativação do cliente.
- $b$ : Parâmetro de forma da distribuição Beta. Também controla a forma da distribuição e influencia a probabilidade de desativação do cliente.

A distribuição Beta é usada para modelar a variação na probabilidade de desativação  $p$  entre diferentes clientes. Ela é uma escolha adequada, pois seus valores estão no intervalo de 0 a 1, que é o intervalo permitido para probabilidades. A função  $B(a, b)$  é a função Beta, que atua como um normalizador para garantir que a PDF esteja corretamente normalizada.

Esses parâmetros  $a$  e  $b$  permitem ajustar a forma da distribuição de probabilidade da variável  $p$ , ou seja, a probabilidade de um cliente desativar após cada compra. Diferentes

combinações de  $a$  e  $b$  resultam em diferentes formas de distribuição, capturando a variação nas taxas de desativação entre os clientes. Podemos ver as implicações de diferentes parametrizações na função beta por meio de exemplo gráfico, como na figura Figura 4.

Figura 4 – Exemplo aplicado da Distribuição Beta



**Fonte:** Elaborada pelo autor.

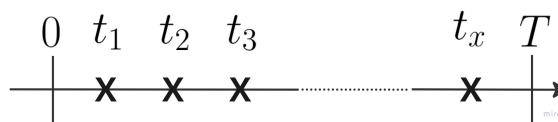
Assim como a distribuição Gama, essa distribuição Beta também carrega implicações comerciais. Uma distribuição Beta assimétrica à esquerda, que concentra a maior parte de sua probabilidade próxima a 0, sugere que a maioria dos clientes tem  $p$  baixo e não tem probabilidade alta de desativar cedo.

### 3.1.4 Implementação do Modelo e Estimação de Parâmetros

Após examinar todas as distribuições com as quais descrevemos quantitativamente o comportamento de nossos clientes, como então obtemos os melhores parâmetros de máxima verossimilhança (MLE), que maximizam a probabilidade de que o modelo tenha produzido os dados que foram efetivamente observados.

Suponha que estejamos atualmente no tempo  $T$  e estamos olhando para trás nas transações históricas de um cliente específico que possui uma taxa de compra  $\lambda$ . Ele fez sua primeira transação em  $t_1$  e a última em  $t_x$ . Esses pontos, representados em uma linha do tempo são representados na Figura 5.

Figura 5 – Transações no tempo

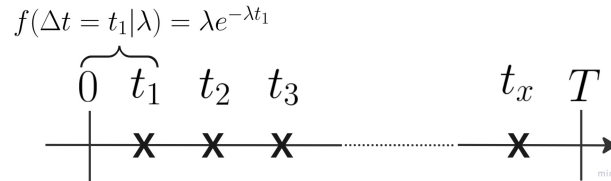


**Fonte:** Elaborada pelo autor.

Podemos derivar a função de verossimilhança individual dessa pessoa seguindo os passos abaixo:

- A probabilidade de a primeira transação ocorrer em  $t_1$  é descrita usando a distribuição exponencial que explicamos anteriormente (ver Figura 6).

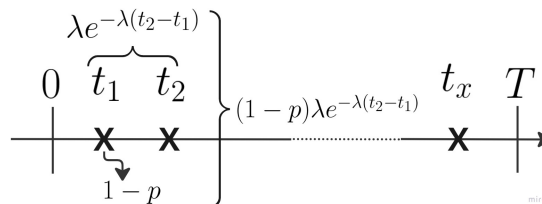
Figura 6 – Probabilidade da primeira transação



**Fonte:** Elaborada pelo autor.

- A probabilidade da segunda transação ocorrer em  $t_2$  é a probabilidade de o cliente permanecer ativo após  $t_1(1 - p)$  multiplicada pelo componente de verossimilhança exponencial padrão (ver Figura 7).

Figura 7 – Probabilidade da segunda transação



**Fonte:** Elaborada pelo autor.

- Esse padrão de verossimilhança é repetido para cada transação subsequente, ou seja, a verossimilhança da transação  $x$  ocorrendo em  $t_x$  é definida na função abaixo.

$$(1 - p)\lambda e^{-\lambda(t_x - t_{x-1})}. \quad (3.5)$$

Agora, vamos analisar o que acontece após a última transação em  $t_x$ . Não observamos nenhuma transação entre  $t_x$  e  $T$ , e essa ausência pode ser devido a um dos seguintes cenários:

- O cliente foi desativado após sua última transação em  $t_x$ . Como sabemos, a probabilidade disso acontecer é  $p$ .

- O cliente permanece ativos, mas não fez nenhuma transação durante esse intervalo. A probabilidade disso acontecer é:

$$(1 - p)e^{-\lambda(T-t_{x-1})} \quad (3.6)$$

A probabilidade de observar o padrão de transações que observamos é simplesmente a multiplicação de todas as verossimilhanças para as transações anteriores, multiplicada pela soma das verossimilhanças dos dois cenários:

$$\begin{aligned} L(\lambda, p|t_1, t_2, \dots, t_x, T) &= \lambda e^{-\lambda t_1} (1 - p) \lambda e^{-\lambda(t_2-t_1)} \dots (1 - p) \lambda e^{-\lambda(t_x-t_{x-1})} \\ &\quad \times \{p + (1 - p)e^{-\lambda(T-t_{x-1})}\} \\ &= p(1 - p)^{x-1} \lambda^x e^{-\lambda t_x} + (1 - p)^x \lambda^x e^{-\lambda T} \end{aligned} \quad (3.7)$$

A fórmula de verossimilhança definida acima é aplicável a clientes que fizeram algumas compras no período de observação, uma vez que assumimos que todos os clientes estavam ativos desde o início, a probabilidade de um cliente não fazer nenhuma compra entre o tempo  $[0, T]$  é a função de sobrevivência da exponencial padrão dada por:

$$L(\lambda|X = 0, T) = e^{-\lambda T}. \quad (3.8)$$

Por último, combinando as duas fórmulas de verossimilhança (3.7) e (3.8), obtemos uma fórmula generalizada para cada cliente, independentemente do número de transações que eles fizeram (ou da ausência delas):

$$L(\lambda, p|X = x, T) = (1 - p)^x \lambda^x e^{-\lambda T} + \delta_{x>0} p(1 - p)^{x-1} \lambda^x e^{-\lambda t_x} \quad (3.9)$$

em que  $\delta_{x>0} = 1$  se  $x > 0$ , se não 0.

Podemos então, testar de modo computacional, diferentes valores de  $p$  e  $\lambda$  e escolher uma combinação  $(p, \lambda)$  que maximize essa verossimilhança. Esses serão os valores dos parâmetros que melhor descrevem o comportamento de compra e a probabilidade de desativação desse indivíduo específico.

Observe que essa função de verossimilhança ao nível individual envolve apenas três variáveis desconhecidas que precisam ser fornecidas pelos dados:

- $x$ : o número de transações repetidas. Isso também é chamado de frequência;

- $t_x$ : a idade do cliente no momento de sua última transação. Isso também é chamado de recência;
- $T$ : a idade do cliente no momento da análise.

Curiosamente, o tempo exato das transações anteriores ( $t_1$  a  $t_{x-1}$ ) não é necessário.

Um conjunto de dados cujas linhas correspondem a diferentes IDs de clientes e cujas colunas indicam o  $x$ ,  $t_x$  e  $T$  de cada cliente é chamado de formato RFM (recência, frequência e valor monetário) (MIGLAUTSCH, 2002).

### 3.1.5 Função de Probabilidade de “Estar Vivo”

A probabilidade de “estar vivo” no modelo BG/NBD é uma medida estatística que avalia a probabilidade de um cliente ainda estar ativo (não ter se desativado) até o momento de análise, dado o seu histórico de compras (FADER; HARDIE; LEE, 2005a; FADER; HARDIE; LEE, 2005b).

A fórmula para a probabilidade de “estar vivo” no modelo BG/NBD envolve parâmetros específicos do modelo, bem como informações do cliente, como a frequência de compras  $x$ , a recência da última compra  $t_x$  e a idade do cliente  $T$  no momento da análise. A fórmula geral é a seguinte:

$$P(\text{Estar Vivo}|x, t_x, T) = \frac{1}{1 + \frac{a}{b+x-1} \cdot \frac{1-(1+\lambda)(1-p)^x}{1-(1+\lambda)(1-p)^T}} \quad (3.10)$$

onde:

- $a$  e  $b$  são parâmetros da distribuição Beta que modelam a probabilidade de desativação  $p$  entre os clientes;
- $\lambda$  é a taxa média de compra entre os clientes ativos;
- $p$  é a probabilidade média de desativação entre os clientes;
- $x$  é a frequência de compras (o número de transações repetidas);
- $t_x$  é a recência da última compra (idade do cliente na última transação);
- $T$  é a idade do cliente no momento da análise.

Note que ao empregarmos a distribuição Gama para modelar o tempo entre compras em processos de Poisson, onde as taxas de compra  $\lambda$  ocorrem de forma independente em intervalos de tempo fixos. Ao combinar a distribuição Poisson com a distribuição gama, obtemos a distribuição binomial negativa, que é uma generalização mais flexível

da distribuição Poisson. Enquanto a distribuição Poisson assume uma taxa constante de ocorrência de compras ao longo do tempo, a distribuição binomial negativa permite que essa taxa varie, tornando-a mais adequada para modelar situações onde a taxa de compras pode ser variável ou incerta. Essa flexibilidade adicionada pode ser crucial em contextos como a modelagem de comportamento de clientes, onde as taxas de compra ou desativação podem variar significativamente entre diferentes clientes ou ao longo do tempo.

Para o objetivo deste trabalho, iremos utilizar a implementação da função probabilidade de “estar vivo” simplificada, disponível no pacote `lifetimes`, que segue a abordagem apresentada por Hardie (2007) demonstrada abaixo:

$$P(\text{Estar vivo}|x, t_x, T, r, \alpha, a, b) = (r + x) \ln \left( \frac{\alpha + T}{\alpha + t_x} \right) + \ln \left( \frac{a}{b + \max(x, 1) - 1} \right). \quad (3.11)$$

Essa função auxiliará na construção de uma variável de identificação de *churn* da base de clientes de acordo com seu comportamento de compra.

### 3.1.6 Limitações do Modelo

O modelo BG/NBD baseia-se em diversos pressupostos, que podem limitar sua aplicabilidade em determinados cenários. Alguns pressupostos-chave incluem:

- **Pressuposto de estacionariedade:** O modelo assume que o comportamento do cliente permanece constante ao longo do tempo, com taxas de transação e taxas de abandono independentes do período de tempo em consideração. Na realidade, o comportamento do cliente pode mudar ao longo do tempo devido a diversos fatores, como tendências de mercado, ofertas de produtos ou eventos externos.
- **Pressuposto de independência:** O modelo assume que as transações dos clientes e os eventos de abandono são independentes entre si, ou seja, o comportamento de transação de um cliente não afeta sua probabilidade de abandono. No entanto, na prática, esses eventos podem estar correlacionados, e o comportamento do cliente pode influenciar sua propensão a *churn*.
- **Pressuposto de homogeneidade:** O modelo pressupõe que todos os clientes seguem o mesmo processo subjacente com os mesmos valores de parâmetros. No entanto, o comportamento do cliente pode variar significativamente entre diferentes segmentos, demografias ou grupos de clientes. Não considerar a heterogeneidade pode levar a resultados tendenciosos.

A qualidade e disponibilidade dos dados também podem impor limitações à aplicação do modelo BG/NBD. Algumas limitações comuns dos dados incluem:

- **Dados censurados:** O modelo pressupõe informações completas sobre o histórico de transações do cliente e os momentos de abandono. No entanto, na prática, os dados podem ser censurados, o que significa que temos apenas informações parciais, especialmente para clientes que ainda estão ativos ou que ainda não experimentaram eventos de abandono. Dados censurados podem afetar a precisão das estimativas dos parâmetros e das previsões subsequentes.
- **Duração limitada dos dados:** O desempenho do modelo pode ser afetado pela duração dos dados observados. A precisão dos parâmetros estimados e das previsões pode diminuir quando a duração dos dados é relativamente curta, pois o modelo requer informações históricas suficientes para capturar com precisão os padrões de comportamento do cliente a longo prazo.
- **Qualidade e consistência dos dados:** A precisão e confiabilidade do modelo dependem muito da qualidade e consistência dos dados de entrada. Dados incompletos, errôneos ou inconsistentes podem levar a estimativas tendenciosas dos parâmetros e previsões pouco confiáveis. Técnicas de limpeza e pré-processamento de dados devem ser empregadas para mitigar essas limitações.

### 3.1.7 Avaliação do Modelo

Para a avaliação do modelo, de forma geral, conforme Müller e Guido (2016) a base de dados é dividida nas seguintes proporções: 70% da base constitui a amostra de desenvolvimento/treinamento, a qual é utilizada para fazer toda a etapa de seleção de variáveis e ajuste dos modelos, enquanto que os 30% restantes compõem a chamada amostra de teste/validação. Essa metodologia de partição do banco de dados é conhecida como *holdout*, cujo resultado obtido na amostra de treinamento é confrontado com novos dados, não utilizados anteriormente para treinar o modelo, para verificar a eficácia do modelo na previsão do resultado de interesse. Para essa validação, são utilizadas diversas métricas.

#### 3.1.7.1 Matriz de Confusão

Para realizar a validação da predição binária de *churn*, utilizaremos a matriz de confusão, uma ferramenta que permite a visualização do desempenho do algoritmo (ver Figura 15).

Os componentes da matriz de confusão são:

- **Verdadeiro Positivo (VP):** São os casos que foram corretamente previstos como inativos pelo modelo.
- **Falso Positivo (FP):** São os casos que foram erroneamente previstos como inativos pelo modelo, quando na verdade são ativos.

Figura 8 – Matriz de Confusão

		Predito	
		Ativo	Inativo
Real	Ativo	Verdadeiro Negativo (VN)	Falso Positivo (FP)
	Inativo	Falso Negativo (FN)	Verdadeiro Positivo (VP)

**Fonte:** Elaborada pelo autor.

- **Falso Negativo (FN):** São os casos que foram erroneamente previstos como ativos pelo modelo, quando na verdade são inativos.
- **Verdadeiro Negativo (VN):** São os casos que foram corretamente previstos como ativos pelo modelo.

As principais medidas utilizadas para o balizamento da eficiência do modelo na predição dos resultados são apresentadas a seguir:

- **Sensibilidade (Recall ou S):** Sensibilidade mede a proporção de verdadeiros positivos em relação a todos os exemplos positivos reais. É a capacidade do modelo de identificar positivos reais.

$$S = \frac{VP}{VP + FN}$$

- **Especificidade (E):** Especificidade mede a proporção de verdadeiros negativos em relação a todos os exemplos negativos reais. É a capacidade do modelo de identificar negativos reais.

$$E = \frac{VN}{VN + FP}$$

- **Valor Preditivo Positivo (VPP ou Precisão):** Valor Preditivo Positivo mede a proporção de verdadeiros positivos em relação a todos os exemplos preditos como positivos. É a precisão do modelo para classificar positivos.

$$VPP = \frac{VP}{VP + FP}$$



- **Valor Preditivo Negativo (VPN):** Valor Preditivo Negativo mede a proporção de verdadeiros negativos em relação a todos os exemplos preditos como negativos.

$$VPN = \frac{VN}{VN + FN}$$

- **Acurácia (Accuracy):** Acurácia mede a proporção de predições corretas em relação a todas as predições.

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN}$$

- **F1-Score:** O F1-Score é uma métrica harmônica entre a especificidade e sensibilidade. É útil quando você deseja equilibrar precisão e recall.

$$F1-Score = \frac{2 \cdot S \cdot E}{S + E}$$

### 3.1.7.2 Curva ROC e área sobre a curva (AUC)

Para avaliar o desempenho do modelo BG/NBD, pode-se utilizar a curva *receiver operating characteristics* (ROC). A curva ROC é uma representação gráfica do desempenho de classificação do modelo, ilustrando a relação entre a taxa de verdadeiros positivos (sensibilidade) e a taxa de falsos positivos (1 - especificidade) em vários limiares de classificação.

No contexto da estimativa da probabilidade condicional de estar vivo, o modelo BG/NBD pode ser usado para classificar os clientes em diferentes faixas com base na sua probabilidade de estar vivo. Por exemplo, os clientes podem ser classificados como clientes de alto valor, médio valor ou baixo valor. A curva ROC nos permite avaliar a capacidade do modelo de distinguir entre essas probabilidades.

A *area under curve* (AUC) foi calculada para quantificar a precisão preditiva do modelo. O valor da AUC varia de 0 a 1, onde um valor mais alto indica melhor desempenho preditivo. Uma AUC de 0,5 sugere que o modelo não é melhor do que um palpite aleatório, enquanto uma AUC de 1,0 indica uma precisão de classificação perfeita.

Para construir a curva ROC e calcular a AUC, os valores de probabilidade pelo modelo BG/NBD foram comparados aos valores reais de *churn* em um conjunto de dados de validação. Ao comparar o *churn* previsto com o *churn* real, a curva ROC pode ser gerada.

### 3.1.7.3 Kolmogorov-Smirnov

Este método avalia se duas amostras são provenientes da mesma distribuição de probabilidade contínua. A estatística de teste para o teste de Kolmogorov-Smirnov é

---

calculada pela diferença entre as funções de distribuição acumulada (CDF) empíricas das duas amostras:

$$D = \max |F_1(x) - F_2(x)|$$

onde:

- $F_1(x)$  é a CDF empírica da primeira amostra;
- $F_2(x)$  é a CDF empírica da segunda amostra;
- $D$  é a estatística de teste de Kolmogorov-Smirnov.

O teste de hipóteses associado ao método Kolmogorov-Smirnov é definido como:

- **Hipótese nula ( $H_0$ ):** As duas amostras são provenientes da mesma distribuição;
- **Hipótese alternativa ( $H_1$ ):** As duas amostras não são provenientes da mesma distribuição.

Se o valor  $D$  calculado for grande e o valor-p associado for menor que um certo nível de significância (geralmente ( $p \leq 0.05$ )), então a hipótese nula é rejeitada, o que sugere que as duas amostras provavelmente não são provenientes da mesma distribuição.

### 3.2 Análise de Sobrevida

Nesta seção de metodologia, discutiremos a implementação do modelo Kaplan-Meier (KM) no contexto de fidelização e retenção de clientes. Este é um modelo estatístico amplamente utilizado para analisar dados de tempo até o evento, em que o objetivo é compreender a probabilidade de sobrevida em vários intervalos de tempo e ilustrar graficamente a sobrevida ao longo do tempo, ou seja, representar a probabilidade de um evento ocorrer em um determinado momento, dado que o indivíduo não tenha encerrado o relacionamento com a empresa até aquele momento.

Em nossa aplicação empresarial, nosso objetivo é investigar os fatores que influenciam a rotatividade de clientes. Para tal, precisamos de dados sobre as características dos clientes, padrões de uso do serviço e o tempo até o evento de rotatividade ou censura. O processo de coleta e transformação de dados envolve a extração de variáveis relevantes, como dados demográficos do cliente e detalhes do comportamento de compra.

### 3.2.1 Kaplan-Meier

O modelo de Kaplan-Meier é uma técnica usada para estimar a função de sobrevivência a partir de informações sobre o tempo até o evento de interesse, como o fim do relacionamento entre cliente e empresa. A função de sobrevivência estimada, denotada por  $\hat{S}(t)$ , é a probabilidade de um indivíduo estar ativo além de um tempo  $t$  (KAPLAN; MEIER, 1958). Para calcular  $\hat{S}(t)$ , usamos a seguinte fórmula:

$$\hat{S}(t) = \prod_{i:t_i \leq t} \left(1 - \frac{d_i}{n_i}\right) \quad (3.12)$$

onde:

- $\hat{S}(t)$ : é a estimativa da função de sobrevivência em um tempo  $t$ ;
- $t_i$ : são os tempos de eventos observados;
- $d_i$ : é o número de eventos ocorridos no tempo  $t_i$ ;
- $n_i$ : é o número de indivíduos sob risco no tempo  $t_i$ , ou seja, aqueles que ainda não sofreram o evento ou foram censurados até o tempo  $t_i$ .

A curva de Kaplan-Meier é uma representação gráfica da função de sobrevivência estimada  $\hat{S}(t)$ . Ela é uma série de passos decrescentes que representam as estimativas da probabilidade de sobrevivência em intervalos de tempo discretos.

O modelo de Kaplan-Meier leva em consideração a censura de dados, o que significa que ele não exige que todos os indivíduos tenham alcançado o evento de interesse até o final do estudo. Ele calcula as estimativas da função de sobrevivência com base nas informações disponíveis e no número de eventos ocorridos.

## 3.3 Estruturação do Código em Python

A estruturação das bases para análise exploratória, tratamento dos dados, execução e ajustar os modelos apresentados anteriormente, foi escolhido o programa Visual Studio Code em um ambiente de integração em linguagem Python. As buscas por códigos construídos anteriormente teve sucesso no sentido de encontrar os pacotes `lifelines`<sup>2</sup> versão 0.27.7 e `lifetimes`<sup>3</sup> versão 0.11.2. Ambas as bibliotecas são gratuitas e disponíveis ao público e possuem comunidade ativa e contam com exemplos práticos e orientativos de utilização e construção de análises.

<sup>2</sup> <https://lifelines.readthedocs.io/en/latest/fitters/regression/CoxPHFitter.html>

<sup>3</sup> <https://lifetimes.readthedocs.io/en/latest/lifetimes.fitters.html>

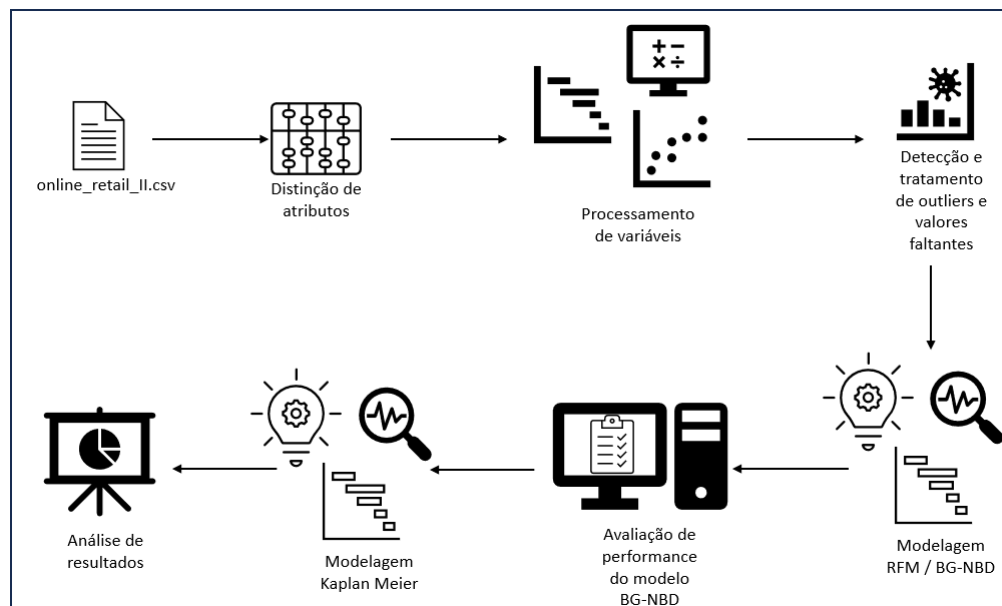
## 4 RESULTADOS

Os resultados são apresentados em 2 partes; a primeira incorpora a análise exploratória do comportamento de compra e resultados da implementação do modelo BG/NBD; a segunda foi construída a partir da análise exploratória das bases demográfica e comportamental utilizando técnicas de análise de sobrevivência.

### 4.1 Análise Exploratória

A análise exploratória de dados para modelagem BG/NBD e Kaplan-Meier (KM) foi construída a partir da base transacional, que foi formatada em uma sequência de etapas fundamentais na avaliação e compreensão do conjunto de dados. Este fluxo permitiu examinar e visualizar os dados de forma a identificar padrões e tendências, o que suportou os *inputs* dos modelos BG/NBD e KM. A estrutura pode ser observada na Figura 9.

Figura 9 – Estrutura da Análise Exploratória



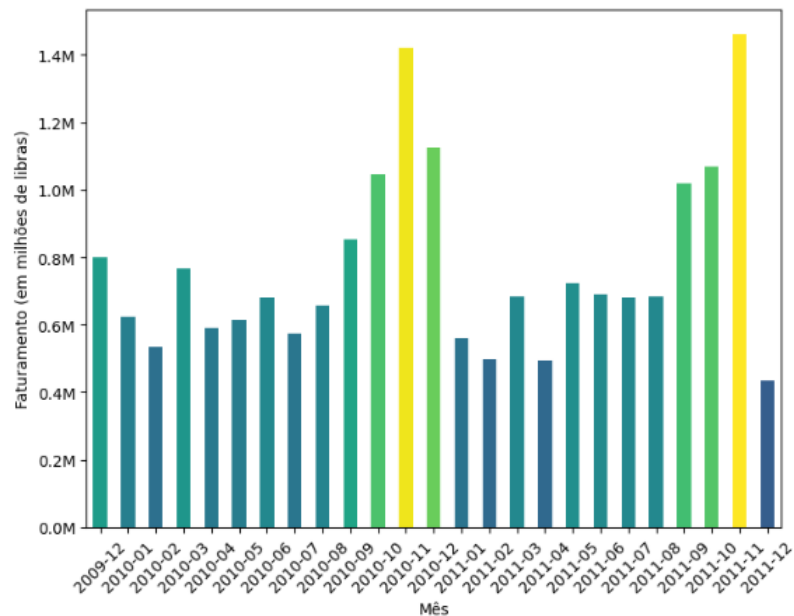
**Fonte:** Elaborada pelo autor.

Inicialmente, os atributos foram carregados e categorizados de acordo com sua natureza, com o propósito de assegurar coerência na medida e na formatação durante as etapas subsequentes de transformação e modelagem. Em um segundo momento, procedeu-se à realização de verificações de consistência nos dados, tais como a identificação de valores faltantes ou nulos.

Na sequência examinamos padrões de sazonalidade, por ser um elemento de suma importância ao avaliar o comportamento de compra e sua relação com o churn. Em primeiro

momento apresentamos o faturamento mensal na Figura 10.

Figura 10 – Faturamento Mensal

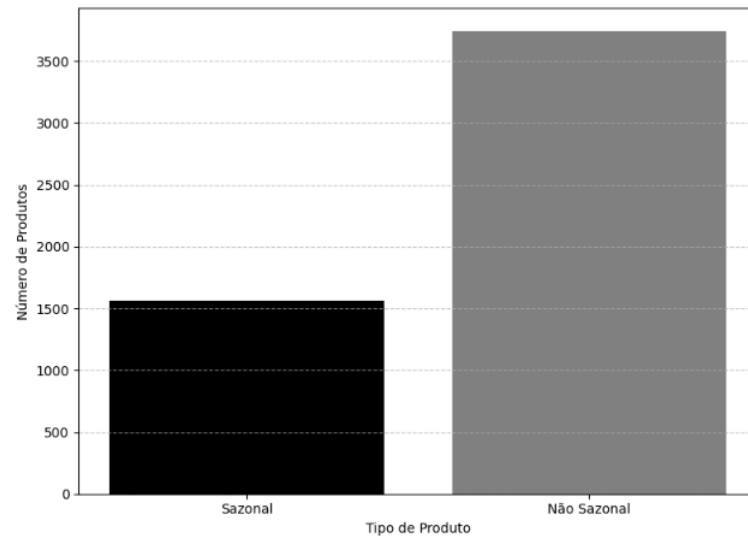


**Fonte:** Elaborada pelo autor.

Observe que há um padrão de comportamento de compras com flutuações recorrentes ao longo dos meses com maior volume no último trimestre do ano. Isso sugere que a base de produtos é influenciada por fatores como estações do ano, feriados, eventos sazonais, entre outros. No contexto do comportamento de compra, entender como os padrões sazonais afetam as decisões de compra dos clientes é crucial para identificar tendências, prever demandas futuras e antecipar possíveis períodos de *churn*. Para compararmos o impacto de produtos sazonais em relação a desativação de clientes, segmentamos os produtos entre sazonal e não sazonal, a partir da decomposição sazonal da soma da quantidade vendida ao longo dos meses de cada produto. O produto foi considerado sazonal se pelo menos 100% da sua variação é explicada pela componente sazonal. Para decompor a série temporal de cada produto e extrair a componente sazonal, utilizamos o pacote `statsmodels.tsa.seasonal`, o resultado da classificação os produtos poder ser vista na Figura 11

Após análise da distribuição de produtos sazonais e não sazonais, observamos que do total de produtos distintos, 3.740 são identificados como não sazonais, enquanto 1.565 são categorizados como sazonais. Esta distinção é crucial, pois os produtos sazonais, por natureza, podem não requerer uma renovação frequente, resultando em um perfil de compra diferenciado. A não fidelidade do consumidor pode estar diretamente relacionada a essa diferenciação, influenciando padrões de compra e comportamentos de mercado. Essas descobertas nos motivam a explorar mais a fundo os resultados obtidos através da análise de Kaplan Meier, buscando compreender melhor os padrões de sobrevivência e fidelidade dos consumidores em relação aos diferentes tipos de produtos.

Figura 11 – Segmentação de Produtos Sazonais

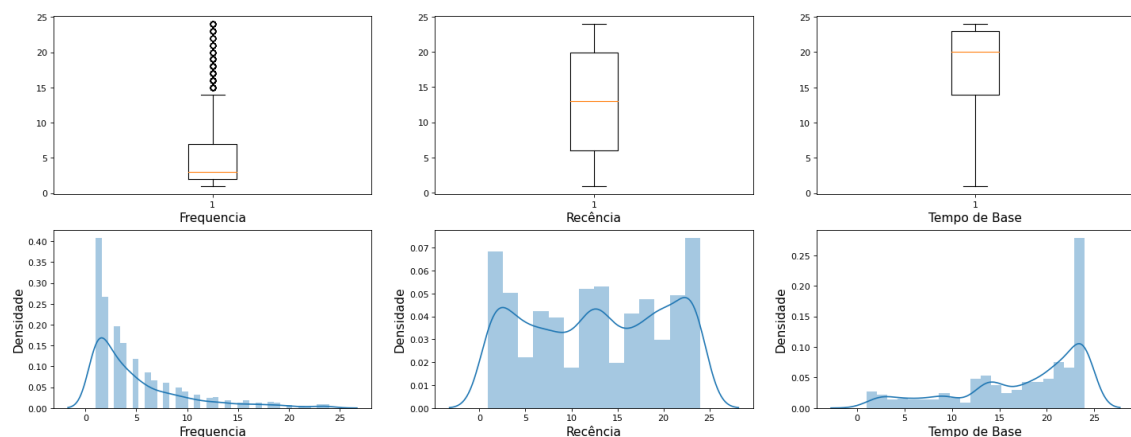


**Fonte:** Elaborada pelo autor.

Posteriormente criamos a estrutura de dados RFM (recência, frequência e valor monetário) com partição mensal (MIGLAUTSCH, 2002). Esta é uma técnica famosa para segmentação de clientes e avaliação do comportamento de compra dos clientes (HAIYING; YU, 2010). Para tal atividade, utilizou-se a função *summary data from transaction data* do módulo *utils*<sup>1</sup> do pacote *lifetimes*, em que o papel é transformar dados de transações em um formato resumido mais adequado para a modelagem de comportamento do cliente usando modelos de recência e frequência como o BG/NBD.

A partir do *dataset* RFM, utilizamos a técnica de análise univariada para examinar a distribuição com histogramas e gráficos de barras, como apresentados na Figura 12.

Figura 12 – Box plot e distribuição RFM



**Fonte:** Elaborada pelo autor.

<sup>1</sup> <https://lifetimes.readthedocs.io/en/latest/lifetimes.html>

Medidas resumo como tendências centrais (média, mediana e moda), dispersão (variância e desvio padrão) são apresentados na Tabela 2.

Tabela 2 – Medidas de Resumo

Métricas	Frequência	Recência	Tempo de Base
Média	5	12,7	17,5
Mediana	3	13	20
Moda	1	23	24
Variância	23,3	54,7	43,1
Desvio Padrão	4,8	7,4	6,6

Fonte: Elaborada pelo autor.

Notavelmente, a média de compras por cliente é de aproximadamente 5, com uma mediana de 3, indicando uma distribuição enviesada para a direita, conforme ilustrado pelo boxplot da frequência. Esse gráfico também revela a presença de outliers significativos acima do valor 15, apontando para uma parcela de clientes com um volume excepcionalmente alto de transações. Em contraste, tanto a recência quanto o tempo de base demonstram distribuições mais homogêneas, sem a presença de outliers. O histograma da recência revela três picos distintos, sugerindo diferentes padrões de comportamento entre os clientes, com agrupamentos notáveis nos períodos de 1 a 4 meses, 13 a 14 meses e 23 a 24 meses. Por sua vez, o histograma do tempo de base apresenta uma assimetria à esquerda, com um pico concentrado entre 23 e 25 meses, indicando uma maior estabilidade ou fidelidade por parte de uma parte da base de clientes.

Em seguida separou-se os dados entre treino e teste de forma randômica com o auxílio do módulo *train test split*<sup>2</sup> do pacote **SKlearn**. Essa etapa contou a a criação de uma variável para a classificação de *churn*, utilizando uma abordagem que considerou o período de inatividade dos clientes. Para isso, foi estabelecida uma regra que envolveu a criação de uma variável simulada de dias inativos, calculada como o intervalo de tempo entre a data da última compra de um cliente e o último dia registrado na base de dados. A média do intervalo de tempo entre a data da última compra de um cliente ( $D_{\text{última compra}}$ ) e o último dia registrado na base de dados ( $D_{\text{último dia}}$ ) foi calculada como critério para classificar o *churn*, e pode ser representada pela seguinte fórmula:

$$\text{Média do intervalo} = \frac{1}{n} \sum_{i=1}^n (D_{\text{último dia}} - D_{\text{última compra},i}) \quad (4.1)$$

em que:

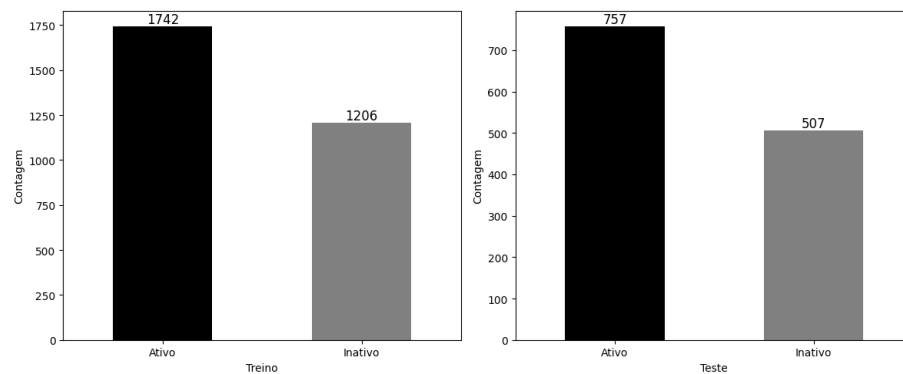
- $n$ : é o número total de clientes na base de dados.;
- $D_{\text{última compra},i}$ : é a data da última compra do cliente  $i$ ;

<sup>2</sup> [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.train\\_test\\_split.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html)

- $D_{\text{último dia}}$  é o último dia registrado na base de dados.

Estabelecendo um limiar de 87 dias. Assim, os clientes cuja última transação ocorreu há mais de 87 dias em relação à data mais recente da base foram classificados como *churn*, enquanto os demais foram considerados como ativos. Na sequência, foi conduzida a partição da base seguindo a estratégia proposta por Müller e Guido (2016), do total 4212 clientes de 70% para treino com 2948 clientes e 30% para teste com 1.264 clientes (ver Figura 13).

Figura 13 – Bases de Treino e Teste



**Fonte:** Elaborada pelo autor.

É importante observar que há um desbalanceamento entre as duas classes, a distribuição desbalanceada dos dados reflete fielmente a distribuição real do problema que estamos tentando modelar. Tentar artificialmente balancear a base pode introduzir distorções que não representam adequadamente o mundo real. Outro aspecto a considerar é que, em alguns casos, as classes minoritárias podem ser de interesse especial e, portanto, é importante que o modelo seja capaz de aprender com elas, mesmo que sejam menos frequentes, como é esperado em casos de análise de *churn*. Por fim, cabe destacar que um rebalanceamento da base pode impactar nos resultados de predição e validação do modelo.

## 4.2 Modelo BG/NBD

O próximo passo foi o ajuste do modelo BG/NBD ao conjunto de dados de treino. O modelo foi ajustado com base no módulo BetaGeoFitter<sup>3</sup> do pacote `lifetimes`, conforme Tabela 3.

A análise dos resultados obtidos a partir do ajuste do modelo BG/NBD apresentados na Tabela 3 revela importantes insights sobre o comportamento de clientes em determinado contexto de negócio. Os parâmetros estimados, representados pelos coeficientes  $r$ ,  $\alpha$ ,  $a$  e  $b$ , descrevem características fundamentais do processo de compra e retenção de clientes. O

<sup>3</sup> <https://lifetimes.readthedocs.io/en/latest/lifetimes.fitters.html>



Tabela 3 – Ajuste do Modelo BG/NBD

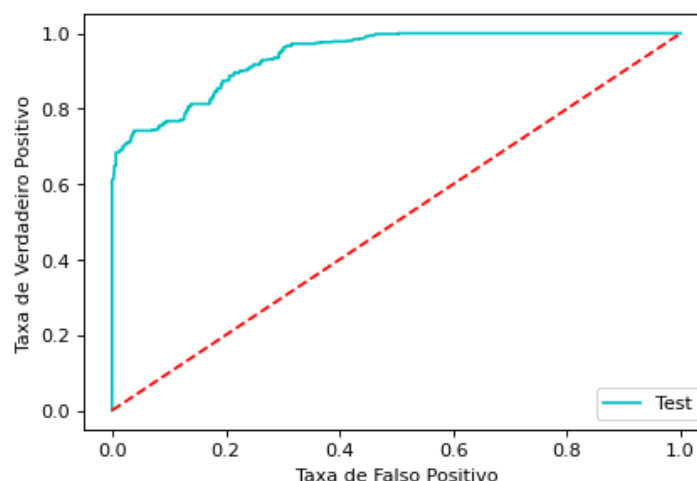
Parâmetros	Coef	se(coef)	Limite Inf 95%	Limite Sup 95%
$r$	3,43	0,16	3,11	3,75
$\alpha$	10,43	0,50	9,44	11,42
$a$	0,09	0,01	0,07	0,12
$b$	0,82	0,13	0,56	1,08

**Fonte:** Elaborada pelo autor.

coeficiente  $r$ , que indica a taxa de recorrência de compras, foi estimado em 3,43 com um erro padrão de 0,16, sugerindo uma alta frequência de retorno dos clientes. O parâmetro  $\alpha$ , que representa a taxa de ativação, foi estimado em 10,43 com um erro padrão de 0,50, indicando uma tendência favorável à ativação de novos clientes. Os parâmetros  $a$  e  $b$ , que modelam a distribuição da probabilidade de abandono, foram estimados em 0,09 e 0,82, respectivamente, com erros padrão de 0,01 e 0,13, revelando um comportamento de retenção estável, mas com uma taxa moderada de abandono. Os intervalos de confiança de 95% para os coeficientes indicam a precisão das estimativas. Esses resultados oferecem uma base sólida para a compreensão e previsão do comportamento futuro dos clientes, auxiliando na formulação de estratégias de marketing e gestão de relacionamento.

Após a realização do ajuste do modelo BG/NBD, empregou-se a função de probabilidade de “estar vivo” para os usuários contidos na base de teste, resultando na obtenção de probabilidades individualizadas. Subsequentemente, procedeu-se ao cálculo da diferença entre um e as probabilidades previstas, visando a determinação da probabilidade de *churn*. Para avaliar o desempenho preditivo do modelo, foi utilizada a curva ROC (ver Figura 14).

Figura 14 – Curva ROC

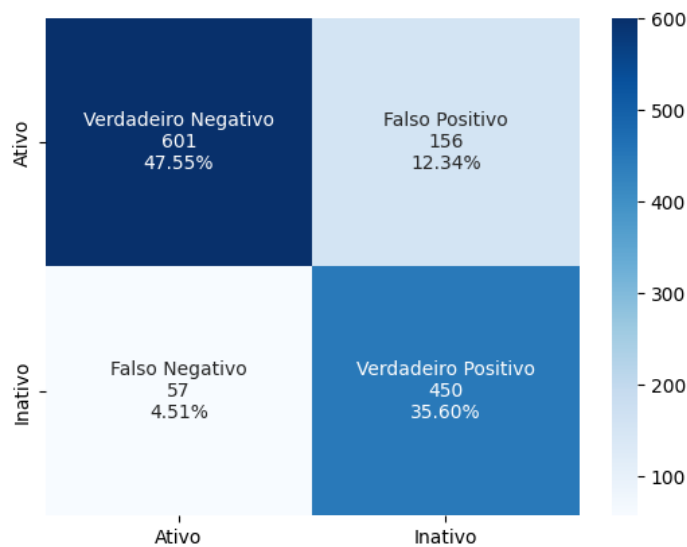


**Fonte:** Elaborada pelo autor.

A AUC apresentou o valor de 94% ao ser comparada com a classificação simulada de *churn* do *dataset* de teste. Enquanto que o teste de Kolmogorov-Smirnov foi de 0,6,

indicando a máxima diferença absoluta entre as funções de distribuição acumulada (CDF) empíricas das duas classes. Quanto maior esse valor, maior é a divergência entre as distribuições das duas classes em relação à característica em questão e o  $p$  – *valor* muito próximo de 0, portanto o modelo é relevante em separar os grupos de *churn* e não *churn*. Vale ressaltar que a correspondência de clientes da base de teste para a validação das predições foi de 1264. Ademais, para fins práticos, precisamos converter as probabilidades preditas em variáveis binárias. Para tal conversão, é necessário utilizar um ponto de corte para delimitar usuários ativos e inativos. A utilização de um critério de corte para transformar probabilidades contínuas em uma classificação binária é uma prática comum em diversos contextos analíticos. No presente estudo, empregou-se tal abordagem com o intuito de classificar os clientes como ativos ou inativos. A escolha do ponto de corte, neste caso, foi conduzida de forma empírica, visando maximizar a acurácia do modelo. Conseqüentemente, foi estabelecido um limiar de 7.5% para as probabilidades preditas. Ou seja, os clientes cuja probabilidade de churn excedesse esse valor foram classificados como inativos, enquanto os demais foram considerados ativos. Esta metodologia proporciona uma maneira objetiva de interpretar as probabilidades contínuas geradas pelo modelo, permitindo a construção da matriz de confusão apresentada na Figura 15.

Figura 15 – Matriz de Confusão BG/NBD



**Fonte:** Elaborada pelo autor.

A análise dos resultados da matriz de confusão proporciona uma compreensão detalhada do desempenho do modelo de classificação utilizado neste estudo. Dentre os quatro quadrantes da matriz, destaca-se a alta proporção de Verdadeiros Negativos, representando 47.55% do total de clientes. Essa categoria indica os casos em que o modelo classificou corretamente os clientes como ativos, alinhando-se com a realidade observada. No entanto, é importante observar que houve uma parcela de Falsos Negativos,

correspondendo a 4.51% dos clientes, o que indica uma falha do modelo em identificar adequadamente alguns clientes ativos. Este aspecto merece atenção, pois pode implicar em oportunidades perdidas de retenção de clientes e maximização do valor do cliente. Além disso, os Falsos Positivos representam 12.34% do total, sugerindo que o modelo identificou erroneamente alguns clientes como ativos quando, na verdade, eram inativos. Esta situação pode acarretar em custos desnecessários e ineficiência nas estratégias de marketing e retenção de clientes. Por fim, os Verdadeiros Positivos, que representam 35.60% do total, indicam a capacidade do modelo em identificar corretamente os clientes inativos. Embora essa taxa seja significativa, é crucial considerar os resultados globais da matriz de confusão para uma avaliação abrangente do desempenho do modelo e a formulação de estratégias apropriadas para mitigar as limitações identificadas. A partir das métricas obtidas pela matriz de confusão, podemos derivar outras métricas de classificação, exibidas na Tabela 4.

Tabela 4 – Métricas Classificação obtidas do ajuste do modelo BG/NBD

<b>Precisão</b>	0,74
<b>Sensibilidade</b>	0,89
<b>Especificidade</b>	0,79
<b>F1-Score</b>	0,81
<b>Acurácia</b>	0,83

**Fonte:** Elaborada pelo autor.

Os resultados da avaliação do modelo de classificação revelam um desempenho robusto em diversas métricas essenciais. A especificidade do modelo, que quantifica sua habilidade de identificar corretamente os Verdadeiros Negativos (VN) ou usuários ativos, alcançou aproximadamente 0,79, evidenciando sua eficácia na classificação desses casos. Por sua vez, a sensibilidade, que avalia a capacidade do modelo de detectar os Verdadeiros Positivos (VP) ou usuários inativos, obteve um valor de 0,89, destacando sua capacidade de identificar esses clientes com precisão. A precisão, que reflete a proporção de previsões positivas corretas em relação ao total de previsões positivas, atingiu aproximadamente 0,74, sinalizando um equilíbrio satisfatório entre previsões precisas. A acurácia global do modelo, que considera todas as previsões corretas, foi de cerca de 0,83, demonstrando sua precisão geral. Além disso, o F1-Score, uma métrica que combina precisão e sensibilidade, alcançou 0,81, enfatizando um equilíbrio sólido entre previsões corretas e a capacidade de recuperar todas as instâncias positivas. Essas métricas refletem a capacidade do modelo em realizar previsões precisas e abrangentes, fornecendo insights valiosos para a tomada de decisões estratégicas.

Em resumo, estes resultados coletivos indicam um desempenho confiável do modelo de classificação na tarefa de distinguir usuários ativos e inativos. Desta forma, o método pode ser utilizado para construir uma base a nível cliente com a classificação estimada de

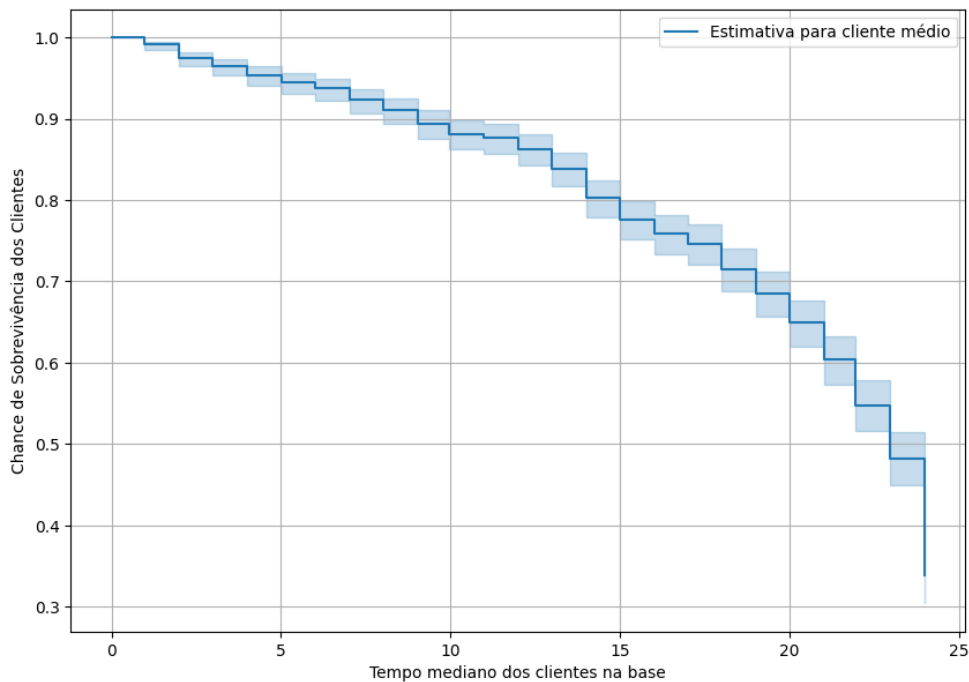
*churn*. Na sequência iremos estimar a probabilidade de desativação em relação tempo dos clientes na base, utilizando agora uma base classificada entre *churn* ou não *churn*.

### 4.3 Análise de Sobrevivência Kaplan-Meier

Com a base de teste agora classificada com as predições do modelo e convertidas em valor binário, podemos utilizar o método estatístico Kaplan-Meier (KM) para estimar a função de sobrevivência e identificar quais os eventos que contribuem para desativação da base de clientes ao longo do tempo.

Inicialmente, a curva de sobrevivência foi estimada para as variáveis *churn* e tempo de base. Desta forma, obtivemos a probabilidade de um cliente sobreviver na base (ver Figura 16).

Figura 16 – Curva Empírica de Sobrevivência Kaplan-Meier



**Fonte:** Elaborada pelo autor.

Com base nos resultados obtidos através das estimativas de KM, torna-se evidente que, aproximadamente no nono mês, os clientes apresentam uma taxa de sobrevivência em torno de 90%. Este ponto pode ser identificado como o primeiro marco significativo no ciclo de vida do cliente, no qual a empresa pode considerar a antecipação da comemoração do primeiro ano de relacionamento como uma estratégia viável. Tal antecipação poderia ser realizada mediante a implementação de uma campanha que ofereça benefícios ao cliente, com o objetivo de manter a frequência de compra.

Outro aspecto digno de destaque ocorre no décimo quinto mês, momento em que essa probabilidade de sobrevivência declina para aproximadamente 80%. Este estágio

---

marca uma oportunidade para a empresa integrar pesquisas de satisfação à jornada do cliente. Tais pesquisas, constituídas por formulários concisos e questões objetivas, visam identificar as causas subjacentes aos pontos de atrito no relacionamento, os quais podem estar relacionados à logística, qualidade, preço ou falhas técnicas. Uma vez identificada a raiz do problema, soluções personalizadas podem ser implementadas com o intuito de recuperar esses clientes.

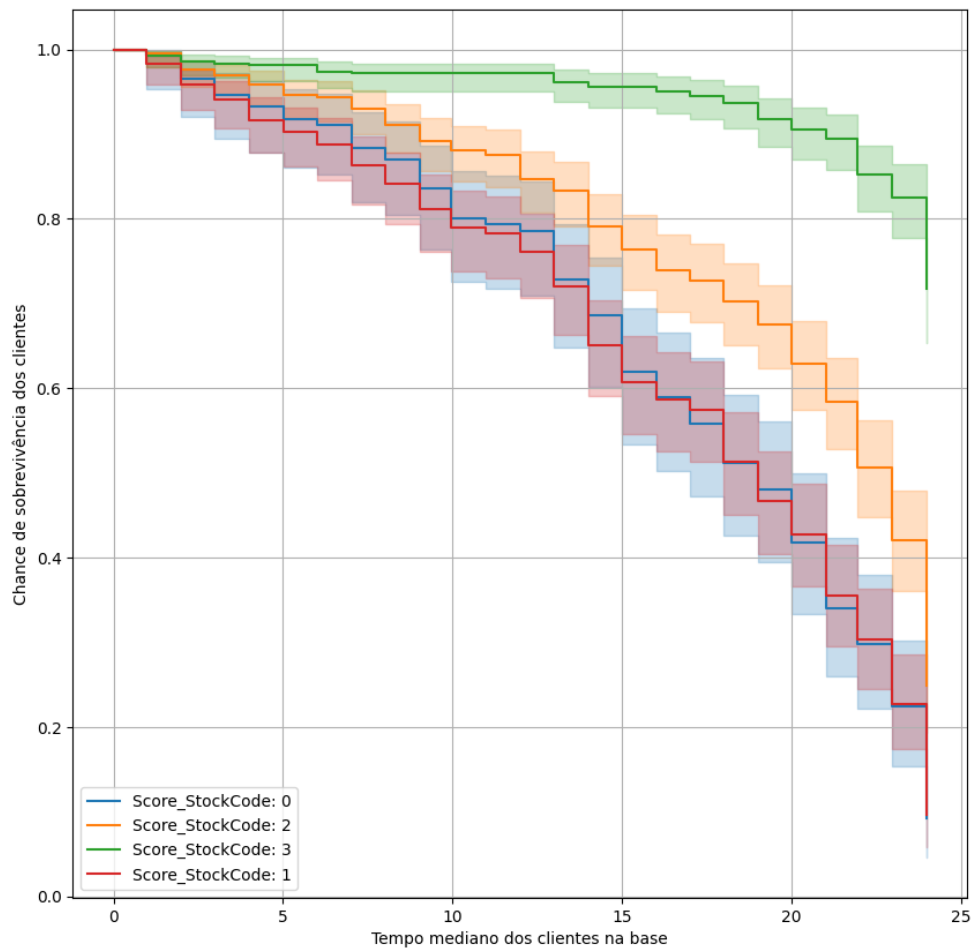
Os períodos subsequentes revelam uma diminuição mais acentuada na probabilidade de sobrevivência, sugerindo uma saturação na base de clientes. Entre o décimo quinto e o vigésimo mês, é fundamental que a empresa concentre seus esforços na compreensão das razões que aceleram o processo de desativação, pois a partir deste ponto as estratégias devem se voltar para a reativação, e não mais para a retenção.

O período mais crítico se situa entre o vigésimo e o vigésimo quinto mês, quando, em média, os clientes atingem uma probabilidade de sobrevivência de 50% na base, indicando que a maioria dos clientes agora se encontra no grupo de inativos. Neste contexto, a empresa deve encarar este marco como um ponto de partida para a aquisição de novos clientes. Deve-se considerar que o custo e o risco de reativação podem não compensar financeiramente, além do fato de que este grupo pode ser propenso a se tornar detratores da marca, o que poderia impactar negativamente na aquisição de novos clientes. Ademais, é importante ressaltar que parte dessa base de clientes pode ter sido desativada devido à falência, eliminando qualquer possibilidade de retorno.

Adicionalmente, incluímos a variável *Score\_StockCode*, obtida utilizando os quartis, onde cada escore representa um quartil, ( $0 = < Q1$ ,  $1 = Q1-Q2$ ,  $2 = Q2-Q3$ ,  $3 = > Q3$ ) para o número de itens distintos por cliente. Isso possibilita verificar se clientes com uma cesta de produtos maior contribui para distinção nas probabilidades de desligamento, ver Figura 17.

Em termos gerais, a variável *Score\_StockCode* revela que grupos com uma maior distribuição de produtos distintos impacta de forma positiva na retenção de clientes. Esse comportamento pode ser explorado a partir de estratégias de penetração de novas categorias, como *cross sell* e *upsell*. Abordagens como produtos sugeridos ou similares, “quem viu, comprou também” baseadas em algoritmos de recomendação que analisam o comportamento do usuário, como histórico de compras, produtos visualizados e itens adicionados ao carrinho, para sugerir produtos adicionais que possam ser personalizados para o interesse do cliente. Essas estratégias são eficazes para aumentar as vendas, melhorar a experiência do cliente e aumentar o envolvimento no site, proporcionando aos usuários uma experiência de compra mais relevante. Outros aspectos a serem analisados é qualidade e preço dos produtos para clientes que possuem uma cesta de produtos menor, compreender se atacar essas variáveis auxiliam em uma maior taxa de retenção pode ser uma estratégia válida também. Portanto, o tamanho da cesta fornece informações relevantes para determinar se um grupo é mais propenso à inativação do que outro e nos

Figura 17 – Curva de Sobrevivência Kaplan-Meier para Itens Distintos



**Fonte:** Elaborada pelo autor.

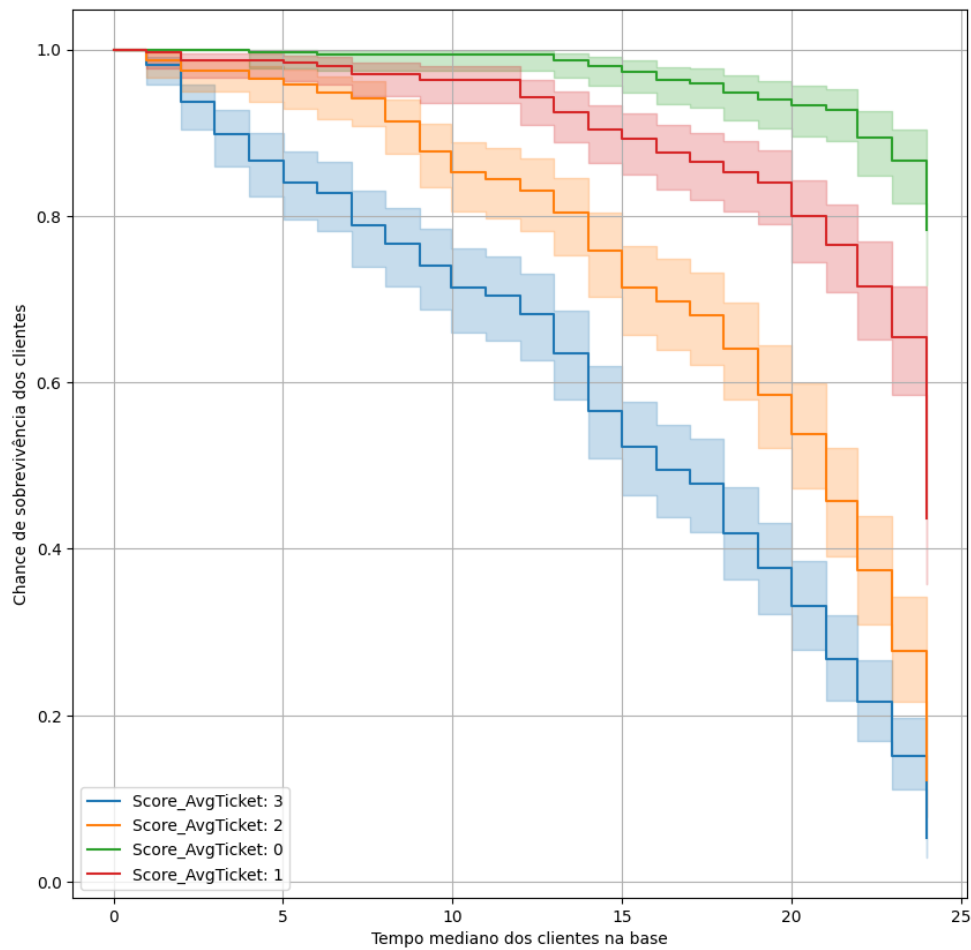
oferece oportunidades de melhoria e implementação de novas funcionalidades para mitigar o risco de *churn*.

Prosseguindo para uma variável de poder de compra, que corresponde ao grupo ou escore de *ticket* médio ao qual o cliente pertence, em outras palavras, é o gasto médio por compra segmentado por quartil, as curvas de sobrevivência estimadas são mostradas na Figura 18.

Dentre as 4 segmentações de *ticket* médio geradas, a que mais se destaca, em termos de fidelização de clientes, é a 0, ou seja, a chance de sobrevivência dos clientes com menor *ticket* médio é mais alta do que os grupos com um gasto por compra mais elevado. Precisamos nos questionar se o impacto desse comportamento é positivo para a empresa.

A Tabela 5 mostra a frequência média por grupo, o grupo de escore 0 possui frequência de 10 compras, enquanto que o *ticket* médio é de 70 centavos de moeda local. Ao compararmos com os demais grupos, se fizermos a multiplicação da frequência média pelo *ticket* médio, o grupo 0 é o que está entregando o menor valor para a empresa, ou

Figura 18 – Curva de Sobrevivência Kaplan-Meier para Ticket Médio



**Fonte:** Elaborada pelo autor.

seja, apesar de serem clientes com maior recorrência, a participação no faturamento da empresa é o de menor relevância.

Tabela 5 – Score Ticket Médio

Score Ticket Médio	Ticket Médio (Moeda Local)	Frequência Média
0	0,70	10
1	3,65	5
2	7,88	3
3	29,30	2

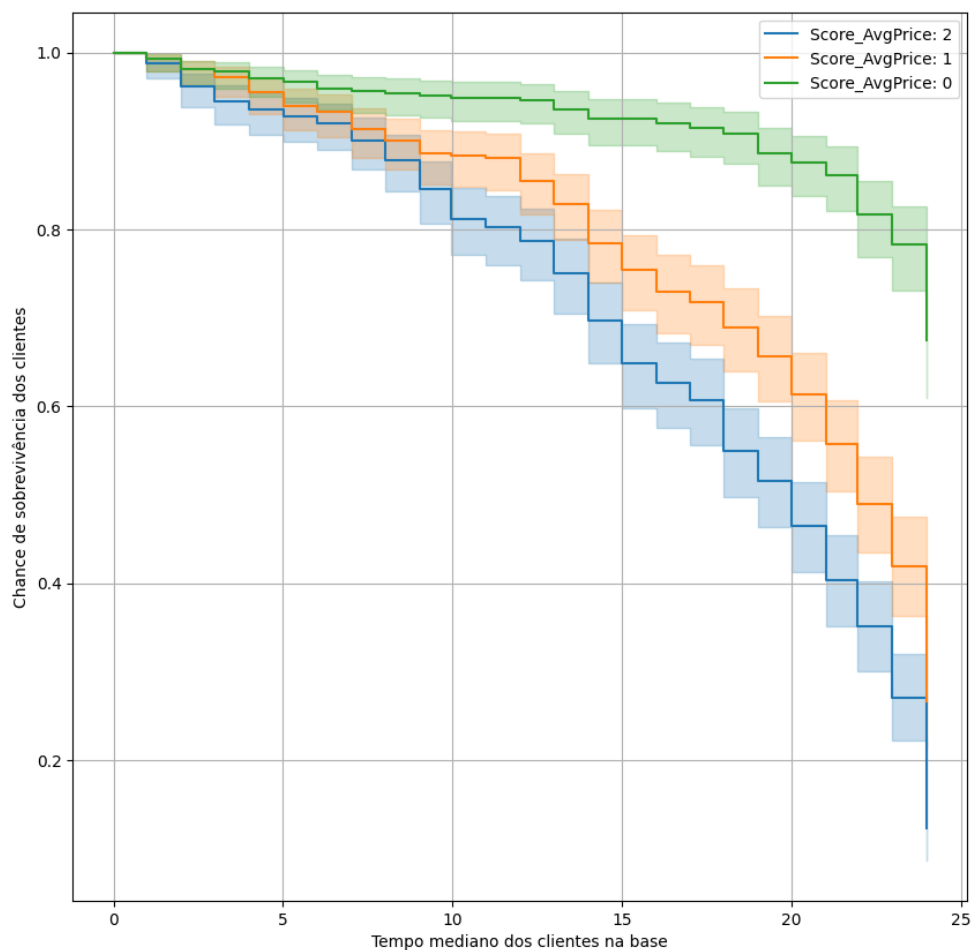
**Fonte:** Elaborada pelo autor.

É preciso entender qual o público alvo do negócio, focar em estratégias de aumento de frequência dos outros grupos, pois entregam maior valor por compra e ao mesmo tempo, incentivar o grupo 0 a adicionar mais valor em suas compras. Outro ponto importante é entender o ciclo de compra de clientes com perfis diferentes, pois possuem estruturas para armazenamento e controle de estoque. Esse tipo de análise permite construir programas de fidelização personalizados com características diferentes, aumento do engajamento e

rentabilidade do negócio. Por exemplo, um programa de fidelidade que oferece benefícios como frete grátis ou cupom de desconto uma vez ao mês, pode auxiliar o aumento da frequência de clientes da categoria 3. Por outro lado, criar critérios de valor mínimo de compra para ativar benefícios, também pode ser relevante como estímulo para compras de maior valor para o grupo 0.

O comportamento de compra dps clientes além de estar associado ao valor da cesta de produtos, também possui relação com o preço médio dos produtos adicionados na cesta que é apresentada na Figura 19.

Figura 19 – Curva de Sobrevivência Kaplan-Meier para Preço Médio



**Fonte:** Elaborada pelo autor.

A variável *Score\_AvgPrice* foi calculada utilizando os tercís, onde cada escore representa um tercil, ( $0 \leq T1$ ,  $1 = T1-T2$ ,  $3 \Rightarrow T2$ ) para o preço médio dos produtos na cesta. Clientes que compram produtos de menor valor estão associados a um maior grau de fidelidade em relação aos clientes que compram produtos mais caros. Portanto, ao relacionarmos estes resultados às análises de itens distintos e *ticket* médio podemos inferir que a empresa deve dedicar tempo e esforço na criação de iniciativas de retenção para o segmento de clientes que compram produtos mais caros, com baixa frequência de compra e



por fim, que possuem uma cesta de produtos pouco diversa. Por se tratar de uma empresa de varejo online com uma ampla variedade de produtos, o efeito da sazonalidade pode explicar o comportamento desse segmento de clientes. Para tanto, precisamos identificar clientes em que a cesta de produtos seja majoritariamente formada por produtos sazonais e calcular o preço médio dos produtos na cesta, apresentados na Tabela 6.

Tabela 6 – Preço Médio de Produtos por Sazonalidade

<b>Sazonal</b>	2,95 libras
<b>Não Sazonal</b>	1,95 libras

**Fonte:** Elaborada pelo autor.

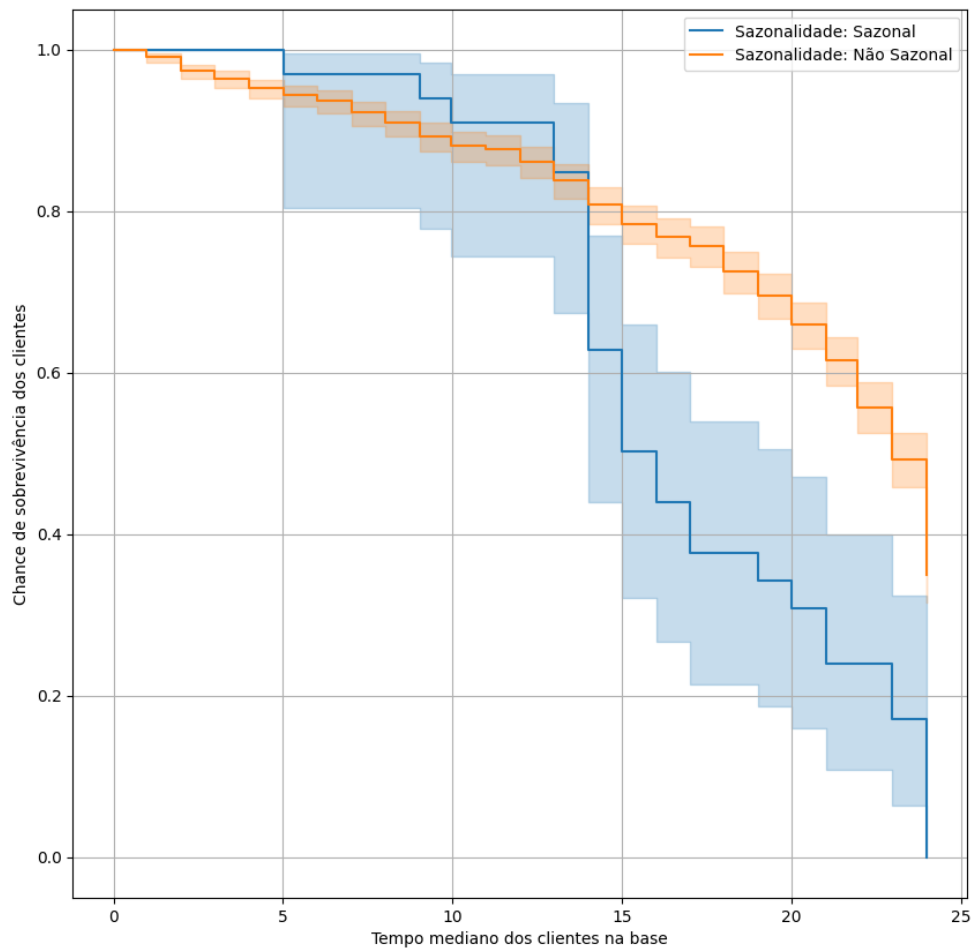
Produtos sazonais possuem um preço médio mais alto em relação aos produtos não sazonais. Estes resultados corroboram com a análise feita anteriormente sobre o segmento de clientes com maior probabilidade de *churn*. São clientes que fazem compras em períodos específicos com volume monetário mais alto em uma gama de itens atreladas a certos momentos do ano.

Contextualizar a taxa de sobrevivência com a curva KM para clientes que possuem cestas amplamente sazonal contribui para entendimento do comportamento de clientes sazonais, que é exibida na Figura 20.

Observe que a taxa de decaimento dos clientes com a cesta majoritariamente sazonal é expressivamente superior ao dos clientes com produtos não sazonais, logo após um ano de relacionamento, por volta do decimo quinto mês, a expectativa de sobrevivência é de 50%. Comercial e estrategicamente, isso implica em uma segmentação de clientes que dependem de estímulos pontuais e assertivos. Geralmente esse perfil de usuário tende a ser mais inquisitivo e volátil a preços, sendo assim, comunicações voltadas a precificação do item auxiliam no processo decisório de compra destes consumidores. Outro aspecto a ser levado em consideração é o giro de estoque dos produtos sazonais, realizar previsão de demanda como ferramenta é de extrema importância, evitando problemas de gestão de inventário. Ademais, a análise permite afirmar que a sazonalidade de produtos por natureza não possui renovação de compras e fidelidade de clientes.

Em suma, para maximizar a retenção de clientes é essencial adotar um modelo estratégico dinâmico e personalizado, que atue em diferentes pontos da curva de sobrevivência, pois o comportamento dos clientes oscila conforme o tempo na base aumenta. Isso inclui a oferta de descontos personalizados, a introdução de novos produtos ou atualizações alinhadas às necessidades, bem como a implementação de técnicas de cross-selling e up-selling para incentivar compras adicionais. Fortalecer os programas de fidelidade com recompensas relevantes, como ofertas exclusivas, são igualmente importantes. Além disso, a comunicação personalizada e um serviço excepcional ao cliente desempenham um papel crucial nesse processo, demonstrando cuidado e criando um ambiente favorável para a

Figura 20 – Curva de Sobrevivência Kaplan-Meier para Sazonalidade



**Fonte:** Elaborada pelo autor.

manutenção da fidelidade dos consumidores. Essas estratégias estruturadas visam não apenas reter os clientes, mas também impulsionar o crescimento sustentável de modelos de negócios não contratuais em um cenário digital.

## 5 CONCLUSÃO

A transformação digital no varejo tem se revelado uma mudança profunda e impactante na forma como os negócios são conduzidos. A integração de tecnologias avançadas, como o uso de análise de dados, automação, inteligência artificial e experiências de compra *online*, está revolucionando a maneira como as empresas se envolvem com seus clientes. O *e-commerce*, por exemplo, tem ampliado as fronteiras do varejo, proporcionando aos consumidores conveniência e acesso a uma gama mais ampla de produtos. Além disso, a transformação digital não é mais uma opção, mas uma necessidade para as empresas que desejam permanecer competitivas, adaptando-se às crescentes demandas dos consumidores por uma experiência de compra mais eficiente, conveniente e personalizada.

Dentro de um ambiente não contratual, no qual o varejo está inserido, identificar a jornada do cliente ao longo do ciclo de vida, se faz de extrema importância, por se tratar de um processo de desistência aleatória. Saber quais clientes estão mais propensos a terminar a relação com a empresa, auxilia na remediação, prevenção e retenção de forma estratégica, sem comprometer financeiramente as operações. Neste sentido, aplicamos o modelo BG/NBD ao conjunto de dados, junto a função probabilidade de “estar vivo” e ao confrontar com uma variável de *churn* simulada, apresentou resultados relevantes e com boa capacidade preditiva de ambas as classes ativos e inativos, acurácia de 83% e AUC de 94%. Desta forma, mostrou-se que esta metodologia é uma boa alternativa para a estimação da classificação de um cliente.

Implementada a variável binária de classificação de *churn* gerada pelo modelo BG/NBD, o próximo passo foi estimar as curvas de sobrevivência de Kaplan-Meier para variáveis comportamentais de compra. Esta análise nos permitiu extrair informações importantes a respeito da base. Clientes com uma cesta de produtos maior, tendem a perpetuar mais na base. Enquanto que cliente com compras de ticket médio mais altos, tendem a ter seus relacionamentos menos duradouros, ou seja, ter uma relação perene, não necessariamente indica que são os melhores clientes, é preciso investigar o valor monetário que estão entregando ao longo de sua jornada. Em relação a variável demográfica, por se tratar de uma informação geográfica, a nível país, os resultados foram pouco significativos para a análise desta base de dados.

Considerando o avanço contínuo da transformação digital nas empresas e seu impacto nas preferências e comportamento dos clientes, fica evidente a importância dessa convergência para o desenvolvimento de políticas eficazes de retenção e reativação. A adaptação constante das organizações às mudanças tecnológicas e às demandas do mercado proporciona uma compreensão mais aprofundada das necessidades dos clientes. Após análises detalhadas, observa-se que o modelo BG/NBD se mostrou estatisticamente

relevante na diferenciação entre usuários ativos e inativos, enquanto a análise Kaplan-Meier ofereceu *insights* valiosos sobre os desafios de retenção e fidelização de clientes em um contexto empresarial digital. Esse contraste entre a eficácia dos modelos estatísticos e a realidade da possível perda de clientes em meio à transformação digital capacita as empresas a tomarem decisões estratégicas embasadas em dados sólidos. Assim, compreendendo como a transformação digital influencia o comportamento do cliente, as empresas podem aproveitar análises estatísticas como o modelo BG/NBD e a análise Kaplan-Meier para identificar padrões de comportamento e implementar estratégias mais eficientes de retenção.

Por fim, destaca-se que avanços podem ser feitos para melhorar a capacidade preditiva do modelo, como utilização de técnicas de aprendizado de máquina, escrever a probabilidade  $p$  do modelo BG/NBD em função das covariáveis comportamentais, empregar processamento de linguagem natural (PLN) na base para extrair categorias das descrições dos produtos, explorar bases com maior número de variáveis a serem analisadas no modelo Kaplan-Meier e/ou utilização de modelos de regressão de análise de sobrevivência.

## REFERÊNCIAS

- AHN, J. *et al.* A survey on churn analysis in various business domains. **IEEE Access**, IEEE, v. 8, p. 220816–220839, 2020.
- AKTER, S.; D’AMBRA, J.; RAY, P. Development and validation of an instrument to measure user perceived service quality of mhealth. **Information & Management**, Elsevier, v. 50, n. 4, p. 181–195, 2013.
- ALMATRODI, I.; LI, F.; ALOJAIL, M. Organizational resistance to automation success: How status quo bias influences organizational resistance to an automated workflow system in a public organization. **Systems**, MDPI, v. 11, n. 4, p. 191, 2023.
- ARSLAN, I. K. The importance of creating customer loyalty in achieving sustainable competitive advantage. **Eurasian Journal of Business and Management**, Eurasian Publications, v. 8, n. 1, p. 11–20, 2020.
- BATISLAM, E. P.; DENIZEL, M.; FILIZTEKIN, A. A customer lifetime value framework for customer selection and resource allocation strategy. **Journal of Marketing**, v. 71, n. 3, p. 19–36, 2007.
- BERMAN, S. J. Digital transformation: opportunities to create new business models. **Strategy & leadership**, Emerald Group Publishing Limited, v. 40, n. 2, p. 16–24, 2012.
- BHARADWAJ, A. *et al.* Digital business strategy: toward a next generation of insights. **MIS quarterly**, JSTOR, p. 471–482, 2013.
- BLEIER, A.; KEYSER, A. D.; VERLEYE, K. Customer engagement through personalization and customization. **Customer engagement marketing**, Springer, p. 75–94, 2018.
- BOLTON, R. N. *et al.* Customer experience challenges: bringing together digital, physical and social realms. **Journal of service management**, Emerald Publishing Limited, v. 29, n. 5, p. 776–808, 2018.
- BORANGIU, T. *et al.* **Digital transformation of manufacturing through cloud services and resource virtualization**. [*S.l.: s.n.*]: Elsevier, 2019. 150–162 p.
- BRESCIANI, S. *et al.* Building a digital transformation strategy. *In: Digital transformation management for agile organizations: A compass to sail the digital world*. [*S.l.: s.n.*]: Emerald Publishing Limited, 2021. p. 5–27.
- BU, S. *et al.* Robotic process automation: A new enabler for digital transformation and operational excellence. **Business Communication Research and Practice**, Korean Association for Business Communication, v. 5, n. 1, p. 29–35, 2022.
- CHEN, C.-L. *et al.* Role of government to enhance digital transformation in small service business. **Sustainability**, MDPI, v. 13, n. 3, p. 1028, 2021.
- CHEN, Y.; FAY, S.; WANG, Q. The role of marketing in social media: How online consumer reviews evolve. **Journal of interactive marketing**, SAGE Publications Sage CA: Los Angeles, CA, v. 25, n. 2, p. 85–94, 2011.

- CHEN, Y. *et al.* Implementation of penalized survival models in churn prediction of vehicle insurance. **Journal of Business Research**, Elsevier, v. 153, p. 162–171, 2022.
- CHOUDHURY, M.; HARRIGAN, P. Crm to social crm: The integration of new technologies into customer relationship management. **Journal of Strategic Marketing**, Taylor & Francis, v. 22, n. 2, p. 149–176, 2014.
- COLBERT, A.; YEE, N.; GEORGE, G. **The digital workforce and the workplace of the future**. [*S.l.: s.n.*]: Academy of Management Briarcliff Manor, NY, 2016. 731–739 p.
- COX, D. R. Regression models and life-tables. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 34, n. 2, p. 187–202, 1972.
- EHRENBERG, A. S. C. The pattern of consumer purchases. **Journal of the Royal Statistical Society. Series A (General)**, v. 121, n. 1, p. 38–64, 1959.
- FADER, P. S.; HARDIE, B. G.; LEE, K. L. "counting your customers"the easy way: An alternative to the pareto/nbd model. **Marketing Science**, v. 24, n. 2, p. 275–284, 2005.
- FADER, P. S.; HARDIE, B. G.; LEE, K. L. Rfm and clv: Using iso-value curves for customer base analysis. **Journal of Marketing Research**, Wiley Online Library, v. 42, n. 4, p. 415–430, 2005.
- FADER, P. S.; HARDIE, B. G.; SHANG, J. Customer-base analysis in a discrete-time noncontractual setting. **Marketing Science**, v. 29, n. 6, p. 1086–1108, 2010.
- GANESH, J.; ARNOLD, M. J.; REYNOLDS, K. E. Understanding the customer base of service providers: An examination of the differences between switchers and stayers. **Journal of Marketing**, v. 64, n. 3, p. 65–87, 2000.
- GEREA, C.; GONZALEZ-LOPEZ, F.; HERSKOVIC, V. Omnichannel customer experience and management: An integrative review and research agenda. **Sustainability**, MDPI, v. 13, n. 5, p. 2824, 2021.
- GLUCKSMAN, M. The rise of social media influencer marketing on lifestyle branding: A case study of lucie fink. **Elon Journal of undergraduate research in communications**, v. 8, n. 2, p. 77–87, 2017.
- GONG, S. Digital transformation of supply chain management in retail and e-commerce. **International Journal of Retail & Distribution Management**, Emerald Publishing Limited, 2023.
- GOULART, V. G.; LIBONI, L. B.; CEZARINO, L. O. Balancing skills in the digital transformation era: The future of jobs and the role of higher education. **Industry and Higher Education**, SAGE Publications Sage UK: London, England, v. 36, n. 2, p. 118–127, 2022.
- GRAEFF, T. R.; HARMON, S. Collecting and using personal data: consumers' awareness and concerns. **Journal of consumer marketing**, MCB UP Ltd, v. 19, n. 4, p. 302–318, 2002.

- 
- HAIYING, M.; YU, G. Customer segmentation study of college students based on the rfm. *In: IEEE. 2010 International Conference on E-Business and E-Government. [S.l.: s.n.]*, 2010. p. 3860–3863.
- HARDIE, B. G. Computing the probability alive for the bg/nbd model. 2007. Disponível em: [http://www.brucehardie.com/notes/021/palive\\_for\\_BGNBD.pdf](http://www.brucehardie.com/notes/021/palive_for_BGNBD.pdf).
- HININGS, B.; GEGENHUBER, T.; GREENWOOD, R. Digital innovation and transformation: An institutional perspective. **Information and Organization**, Elsevier, v. 28, n. 1, p. 52–61, 2018.
- HOFFMAN, D. L.; NOVAK, T. P. Consumer and object experience in the internet of things: An assemblage theory approach. **Journal of Consumer Research**, Oxford University Press, v. 44, n. 6, p. 1178–1204, 2018.
- HOPPE, D.; WAGNER, U. Modeling multiple customer purchases and customer base analysis on aggregate sales data. **Journal of Applied Econometrics**, v. 22, n. 7, p. 1179–1204, 2007.
- HORLACH, B. *et al.* Increasing the agility of it delivery: five types of bimodal it organization. 2017.
- HOYER, W. D. *et al.* Transforming the customer experience through new technologies. **Journal of interactive marketing**, SAGE Publications Sage CA: Los Angeles, CA, v. 51, n. 1, p. 57–71, 2020.
- HUANG, Z.; BENYOUCEF, M. From e-commerce to social commerce: A close look at design features. **Electronic Commerce Research and Applications**, Elsevier, v. 12, n. 4, p. 246–259, 2013.
- IRAWAN, C.; SUSANTO, B.; JULIADI, R. Implementation of digital marketing at printing company to increase customer acquisition and customer retention (case study: Citra kreasindo mandiri company). *In: Conference Series. [S.l.: s.n.]*, 2020. v. 3, n. 1, p. 619–639.
- JAHROMI, A. T.; STAKHOVYCH, S.; EWING, M. Managing b2b customer churn, retention and profitability. **Industrial Marketing Management**, Elsevier, v. 43, n. 7, p. 1258–1268, 2014.
- JARADAT, Z. *et al.* Factors influencing business intelligence adoption: evidence from jordan. **Journal of Decision Systems**, Taylor & Francis, p. 1–21, 2022.
- KANE, G. C. *et al.* Strategy, not technology, drives digital transformation. **MIT Sloan Management Review**, v. 56, n. 1, p. 19–21, 2015.
- KAPLAN, A. M.; HAENLEIN, M. Siri, siri, in my hand: Who's the fairest in the land? on the interpretations, illustrations, and implications of artificial intelligence. **Business Horizons**, Elsevier, v. 62, n. 1, p. 15–25, 2019.
- KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. **Journal of the American statistical association**, Taylor & Francis, v. 53, n. 282, p. 457–481, 1958.

- 
- KHANOM, M. T. Business strategies in the age of digital transformation. **Journal of Business**, v. 8, n. 01, p. 28–35, 2023.
- KIM, S. H.; KWON, J. How do ehRs and a meaningful use initiative affect breaches of patient information? **Information Systems Research**, INFORMS, v. 30, n. 4, p. 1184–1202, 2019.
- KITSIOS, F.; KAMARIOTOU, M. Artificial intelligence and business strategy towards digital transformation: A research agenda. **Sustainability**, MDPI, v. 13, n. 4, p. 2025, 2021.
- KURTI, E.; HAFTOR, D. Barriers and enablers of digital business model transformation. *In*: ACADEMIC CONFERENCES INTERNATIONAL LIMITED. **The European Conference on Information Systems Management**. [S.l.: s.n.], 2015. p. 262.
- LACITY, M. C.; WILLCOCKS, L. P. Nine practices for best-in-class bpo performance. **MIS quarterly executive**, v. 13, n. 3, 2014.
- LARIVIÈRE, B.; POEL, D. Van den. Investigating the role of product features in preventing customer churn, by using survival analysis and choice modeling: The case of financial services. **Expert Systems with Applications**, Elsevier, v. 27, n. 2, p. 277–285, 2004.
- LEMON, K. N.; VERHOEF, P. C. Understanding customer experience throughout the customer journey. **Journal of Marketing**, Sage Publications, v. 80, n. 6, p. 69–96, 2016.
- LIANG, H.; TURBAN, E. Introduction to the special issue social commerce: A research framework for social commerce. **International Journal of Electronic Commerce**, Taylor & Francis, v. 16, n. 2, p. 5–14, 2011.
- LOONAM, J. *et al.* Towards digital transformation: Lessons learned from traditional organizations. **Strategic Change**, Wiley Online Library, v. 27, n. 2, p. 101–109, 2018.
- MANCHA, R.; SHANKARANARAYANAN, G. Making a digital innovator: antecedents of innovativeness with digital technologies. **Information Technology & People**, Emerald Publishing Limited, v. 34, n. 1, p. 318–335, 2021.
- MANTEL, N. Evaluation of survival data and two new rank order statistics arising in its consideration. **Cancer chemotherapy reports**, v. 50, n. 3, p. 163–170, 1966.
- MARTIN, K. D.; MURPHY, P. E. The role of data privacy in marketing. **Journal of the Academy of Marketing Science**, Springer, v. 45, n. 2, p. 135–155, 2017.
- MARTÍNEZ-PELÁEZ, R. *et al.* Role of digital transformation for achieving sustainability: Mediated role of stakeholders, key capabilities, and technology. **Sustainability**, MDPI, v. 15, n. 14, p. 11221, 2023.
- MASARIFOGLU, M.; BUYUKLU, A. H. Applying survival analysis to telecom churn data. **American Journal of Theoretical and Applied Statistics**, v. 8, n. 6, p. 261–275, 2019.
- MCCARTHY, D.; WADSWORTH, E. Buy’til you die—a walkthrough. **Buy ‘Til You Die—A Walkthrough**, 2014.



- MENA, G. *et al.* Exploiting time-varying rfm measures for customer churn prediction with deep neural networks. **Annals of Operations Research**, Springer, p. 1–23, 2023.
- MIGLAUTSCH, J. Application of rfm principles: What to do with 1–1–1 customers? **Journal of Database Marketing & Customer Strategy Management**, Springer, v. 9, p. 319–324, 2002.
- MÜLLER, A. C.; GUIDO, S. **Introduction to machine learning with Python: a guide for data scientists**. [S.l.: s.n.]: "O'Reilly Media, Inc.", 2016.
- MURALIDHARA, P. The evolution of cloud computing security: Addressing emerging threats. **INTERNATIONAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY**, v. 1, n. 4, p. 1–33, 2017.
- NESLIN, S. A. *et al.* Defection detection: Measuring and understanding the predictive accuracy of customer churn models. **Journal of marketing research**, SAGE Publications Sage CA: Los Angeles, CA, v. 43, n. 2, p. 204–211, 2006.
- OCHOLA, M. O. **Attrition Modelling for Online Media Users by Cox Proportional Hazards**. 2019. Tese (Doutorado) — University of Nairobi, 2019.
- PARVIAINEN, P. *et al.* Tackling the digitalization challenge: how to benefit from digitalization in practice. **International journal of information systems and project management**, v. 5, n. 1, p. 63–77, 2017.
- PERIÁÑEZ, Á. *et al.* Churn prediction in mobile social games: Towards a complete assessment using survival ensembles. *In*: IEEE. **2016 IEEE international conference on data science and advanced analytics (DSAA)**. [S.l.: s.n.], 2016. p. 564–573.
- PIEPPONEN, A. *et al.* Digital transformation of the value proposition: A single case study in the media industry. **Journal of Business Research**, v. 150, p. 311–325, 2022. ISSN 0148-2963. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0148296322004404>.
- PLATZER, M.; REUTTERER, T. Incorporating heterogeneity of consumer preferences in customer-base analysis. **International Journal of Research in Marketing**, v. 33, n. 1, p. 186–207, 2016.
- RAJA, J. B. *et al.* A systemic method of nesting multiple classifiers using ensemble techniques for telecom churn prediction. *In*: SPRINGER. **Evolutionary Computing and Mobile Sustainable Networks: Proceedings of ICECMSN 2020**. [S.l.: s.n.], 2021. p. 13–25.
- RODRÍGUEZ, P. *et al.* Advances in using agile and lean processes for software development. *In*: **Advances in computers**. [S.l.: s.n.]: Elsevier, 2019. v. 113, p. 135–224.
- SCHMITTLEIN, D. C.; MORRISON, D. G.; COLOMBO, R. Counting your customers: Who are they and what will they do next? **Management Science**, v. 33, n. 1, p. 1–24, 1987.
- SCHNEIDER, S.; KOKSHAGINA, O. Digital transformation: What we have learned (thus far) and what is next. **Creativity and innovation management**, Wiley Online Library, v. 30, n. 2, p. 384–411, 2021.

- SHAW, R.; STONE, M. **Database Marketing: Strategy and Implementation**. Wiley, 1990. ISBN 9780471623458. Disponível em: <https://books.google.com.br/books?id=AieMQgAACAAJ>.
- TEECE, D. J. Business models and dynamic capabilities. **Long range planning**, Elsevier, v. 51, n. 1, p. 40–49, 2018.
- TERHO, H. *et al.* Digital content marketing in business markets: Activities, consequences, and contingencies along the customer journey. **Industrial Marketing Management**, Elsevier, v. 105, p. 294–310, 2022.
- TOCCU, M.; FASSO, A. *et al.* Statistical analysis of customer lifetime value: a case study on telecommunications data using sas. **Statistica & Applicazioni**, Vita & Pensiero, v. 11, n. 2, p. 177–192, 2013.
- TRAN-DANG, H.; KIM, D.-S. The physical internet in the era of digital transformation: perspectives and open issues. **IEEE Access**, IEEE, v. 9, p. 164613–164631, 2021.
- TRAWNIH, A. *et al.* Understanding artificial intelligence experience: A customer perspective. **International Journal of Data and Network Science**, v. 6, n. 4, p. 1471–1484, 2022.
- VARSHNEY, D. Digital transformation and creation of an agile workforce: Exploring company initiatives and employee attitudes. *In*: **Contemporary global issues in human resource management**. [*S.l.: s.n.*]: Emerald Publishing Limited, 2020. p. 89–105.
- VEN, A. H. Van de *et al.* **The Innovation Journey**. [*S.l.: s.n.*]: Oxford University Press, 2021.
- VERHOEF, P. C.; KANNAN, P. K.; INMAN, J. J. From multi-channel retailing to omni-channel retailing: Introduction to the special issue on multi-channel retailing. **Journal of Retailing**, v. 91, n. 2, p. 174–181, 2015.
- VOLBERDA, H. W. *et al.* Strategizing in a digital world: Overcoming cognitive barriers, reconfiguring routines and introducing new organizational forms. **Long Range Planning**, Elsevier, v. 54, n. 5, p. 102110, 2021.
- WESTERMAN, G.; BONNET, D.; MCAFEE, A. **Leading Digital: Turning Technology into Business Transformation**. [*S.l.: s.n.*]: Harvard Business Review Press, 2015.
- WIMELIUS, H. *et al.* A paradoxical perspective on technology renewal in digital transformation. **Information systems journal**, Wiley Online Library, v. 31, n. 1, p. 198–225, 2021.
- XIAHOU, X.; HARADA, Y. B2c e-commerce customer churn prediction based on k-means and svm. **Journal of Theoretical and Applied Electronic Commerce Research**, MDPI, v. 17, n. 2, p. 458–475, 2022.