



Ana Clara Aragão Fernandes

**Segmentação e o Modelo RFM no Varejo Brasileiro:
Uma Análise com Bases de Dados Transacionais do Varejo de
Vestuário**

Dissertação de Mestrado

Dissertação apresentada ao Programa de Pós-graduação em Administração de Empresas da PUC-Rio como requisito parcial para obtenção do título de Mestre em Administração de Empresas

Orientador: Jorge Brantes Ferreira

Rio de Janeiro, 18 de abril de 2022.



Ana Clara Aragão Fernandes

**Segmentação e o Modelo RFM no Varejo Brasileiro: Uma
Análise com Bases de Dados Transacionais do Varejo de Vestuário**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Administração de Empresas do Departamento de Administração da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Jorge Brantes Ferreira

Orientador

Departamento de Administração – PUC-Rio

Angela Maria Cavalcanti da Rocha

Departamento de Administração - PUC-Rio

Angilberto Sabino de Freitas

Departamento de Engenharia de Produção - UEZO

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Ana Clara Aragão Fernandes

A autora conclui o mestrado em Administração de Empresas pelo Programa de Pós-graduação em Administração de Empresas da PUC-Rio em 2020. Formada em estatística pela Escola Nacional de Ciências Estatísticas (ENCE-IBGE) em dezembro de 2015, com experiência no mercado de telecomunicações, com ênfase no entendimento do comportamento do consumidor e marketing analítico.

Ficha Catalográfica

Fernandes, Ana Clara Aragão

Segmentação e o Modelo RFM no Varejo Brasileiro: Uma Análise com Bases de Dados Transacionais do Varejo de Vestuário/ Ana Clara Aragão Fernandes ; orientador: Jorge Brantes Ferreira. – 2022.

66 f. : il. color. ; 30 cm

Dissertação (mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Administração, 2022.

Inclui bibliografia

1. Administração – Teses. 2. RFM. 3. Modelos de classificação. 4. CRM. 5. COVID-19. 6. Cluster. I. Ferreira, Jorge Brantes. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Administração. III. Título.

CDD: 658

Ao meu filho, Luiz Philippe, pedaço de mim forjado em muita luta.
Você, meu filho que me desafia diariamente a te compreender.
Você, que aos olhos dos outros é apenas mais um menino autista.
Você, que aos meus olhos é o maior guerreiro que eu conheço.
Muito pequeno, lutou bravamente pela sua vida. E venceu.
E continua vencendo todos os dias.

Você que muitas vezes não compreende a minha ausência.
Você que me fez querer desistir e desistir de desistir inúmeras vezes.
Luizinho, meu filho, aprendi com você a não desistir.

Nosso caminho é de muita luta.
E nós lutamos juntos.

Agradecimentos

Ao meu orientador, Jorge Brantes, pelo estímulo e parceria para a realização deste trabalho.

Aos meu filho, Luiz Philippe, pela companhia, atenção e carinho de todas as horas.

Aos meus amigos, Nathalia Barbieri em especial, por todo apoio, paciência e compreensão.

Aos meus colegas da PUC-Rio.

A todos os professores e funcionários do IAG pelos ensinamentos e apoio.

A todos os amigos e familiares que de uma forma ou de outra me deram força ou me ajudaram nesses tempos turbulentos.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Resumo

Fernandes, Ana Clara Aragão; Ferreira, Jorge Brantes. **Segmentação e o Modelo RFM no Varejo Brasileiro: Uma Análise com Bases de Dados Transacionais do Varejo de Vestuário.** Rio de Janeiro, 2022. Número de páginas 66p. Dissertação de Mestrado - Departamento de Administração, Pontifícia Universidade Católica do Rio de Janeiro.

A pandemia de Covid-19 alterou o comportamento do consumidor no varejo a nível mundial. Este trabalho apresenta uma análise longitudinal do comportamento do consumidor ao longo entre 2018 e 2021, possibilitando, dessa forma, a comparação entre o comportamento do consumidor pré e pós pandemia de covid-19 em uma loja do varejo brasileiro. Para realizar essa análise, o modelo RFM é aplicado a partir de métodos de inteligência artificial para a análise de grandes volumes de dados transacionais com o objetivo de classificar os clientes de acordo com os seus comportamentos de consumo. Para o caso apresentado foram identificados 5 segmentos de consumo distintos e de grande utilidade para a gestão de CRM da empresa.

Palavras- chave

RFM; Modelos de Classificação; CRM; COVID-19; Cluster

Abstract

Fernandes, Ana Clara Aragão; Ferreira, Jorge Brantes. (Advisor) **The RFM Model: The Impact of Data Science on Model Applicability Development, strategies and applications in the Brazilian retail market.** Rio de Janeiro, 2022. Número de páginas 66p. MSc. Dissertation – Departamento de Administração, Pontifícia Universidade Católica do Rio de Janeiro.

The Covid-19 pandemic has changed consumer behavior in retail worldwide. This work presents a longitudinal analysis of consumer behavior between 2018 and 2021, thus making it possible to compare consumer behavior before and after the covid-19 pandemic in a Brazilian retail store. To perform this analysis, the RFM model is applied using artificial intelligence methods to analyze large volumes of transactional data in order to classify customers according to their consumption behaviors. For the case presented, 5 distinct and very useful consumer segments were identified for the company's CRM management.

Keywords

CRM; RFM; Classification Models; CRM; COVID-19; Cluster

Sumário

1. O problema	1
1.1 Introdução	2
1.2 Objetivo	4
1.3 Relevância	5
1.4 Delimitação	6
2. Referencial teórico	7
2.1 A história é contada a partir de dados	7
2.2 O marketing baseado em dados	11
2.3 CRM baseado em dados	14
2.4 CRM e Segmentação no Varejo	16
2.5 O modelo RFM: passado, presente e futuro	18
2.6 Modelo RFM no Varejo: uma análise exploratória	21
2.7 O impacto da pandemia de Covid-19 no comportamento do consumidor	26
3. Metodologia	28
3.1 Design da Pesquisa	28
3.2 Universo e amostra	29
3.3 Procedimentos de Coleta de Dados	29
3.4 Tratamento dos dados	30
3.5 K-Means e modelos de classificação	33
3.6 Limitações do estudo	34
4. Análise e Interpretação dos Resultados	35
4.1 Resultados do modelo	35
4.2 Análise longitudinal	38
4.2.1 Recência	39
4.2.2 Frequência	41
4.2.3 Valor	42
4.2.4 Volume	42

4.2.5 Clientes Promocionais	43
4.3 Aplicações práticas	44
5 Conclusão	45
5.1 Implicações Teóricas e Práticas	47
5.2 Sugestões e recomendações para novas pesquisas	48
6. Referências Bibliográficas	49

Lista de figuras

Figura 1 – Crescimento da produção científica de Inteligência artificial aplicada ao marketing versus capacidade global de armazenamento de dados.....	9
Figura 2 – Evolução relacionamento com o cliente versus capacidade produtiva adaptado de Klingenberg et al. (2021)	12
Figura 3 – Quantidade de artigos sobre o modelo RFM publicados anualmente	20
Figura 4 – Número de clusters	35
Figura 5 – Gênero por segmento calculado	37
Figura 6 – Canal preferencial por segmento	37
Figura 7 – Análise Longitudinal da Recência	39
Figura 8 – Análise longitudinal por segmento – Recência	40
Figura 9 – Análise longitudinal – Frequência.....	41
Figura 10 – Análise longitudinal – Valor médio por cliente	42
Figura 11 – Análise Longitudinal – Quantidade de itens.....	43
Figura 12 – Distribuição de clientes promocionais por segmento	43

Lista de tabelas

Tabela 1: Trabalhos publicados	21
Tabela 2: Variáveis de segmentação.....	32
Tabela 3: Clusters Estimados e Principais Indicadores de Segmentação.....	36
Tabela 4: Períodos de referência.....	38

1. O problema

O mercado de varejo como um todo vem passando por um processo de grande transformação. Um dos principais gatilhos dessa transformação é o desenvolvimento tecnológico. Atrélado a esse desenvolvimento, a pandemia de covid-19 acelerou a transformação digital desse segmento.

Vivemos em um mundo digital em constante transformação. O desenvolvimento tecnológico nunca foi tão acelerado. A tecnologia altera a forma que interagimos com o mundo em todos os aspectos. Com as empresas de varejo, não seria diferente.

Numa visão estratégica, as empresas do século XXI estão inseridas num contexto de volatilidade, incerteza, complexidade e ambiguidade. Entre as consequências desse contexto de incerteza e rápida transformação digital, a maneira de fazer negócios mudou e, com ela, o relacionamento entre empresas e clientes.

A revolução tecnológica, o avanço das tecnologias de comunicação, a invenção do smartphone e o surgimento das redes sociais foram catalisadores de uma revolução na forma que se faz o marketing. Junto com o avanço das tecnologias de comunicação, o mundo também avançou na sua capacidade de armazenar dados. Em outras palavras, o desenvolvimento tecnológico nos permite coletar, armazenar e analisar grandes volumes de dados, em alta velocidade a baixo custo. Esse processo de desenvolvimento tecnológico ficou conhecido como 4ª revolução Industrial e possibilitou o desenvolvimento de um novo campo: a ciência de dados.

A ciência de dados surge como uma interseção das áreas de Estatística e da ciência da computação. Desta forma, a ciência de dados utiliza métodos estatísticos computacionalmente intensivos para encontrar padrões e realizar previsões a partir dos grandes volumes de dados armazenados. A ciência de dados pode ser aplicada em todas as áreas de conhecimento, inclusive dentro de empresas, e em todas as áreas que lidem, de alguma forma, com análise de dados. Dentre as possibilidades de aplicação da ciência de dados nas empresas, destaca-se o marketing e a aplicabilidade da ciência de dados em estudos de comportamento de consumo, como o que será apresentado neste trabalho.

Sob a ótica da ciência de dados, o marketing busca encontrar padrões de comportamento do consumidor, a partir da análise de dados das interações anteriores dos clientes com a marca em questão. A partir dos padrões observados computacionalmente, o profissional de marketing pode desenvolver e direcionar estratégias específicas para cada grupo de consumidores, permitindo uma gestão de relacionamento (CRM) com o cliente mais efetiva.

Este trabalho se propõe a, sob a ótica da ciência de dados, apresentar e discutir a evolução dos estudos de segmentação de clientes no varejo. Além disso, optou-se por testar e analisar, com o uso de bases de dado reais de varejo, a aplicação de um dos modelos de classificação e diferenciação de clientes mais usados na indústria varejista: o modelo RFM – Recência, Frequência e Valor Monetário (Bult e Wansbeek, 1995). Este modelo ainda é efetivamente usado para entender e analisar as características do comportamento do cliente (Fader, Hardie e Lee, 2005; Peker, Kocyigit e Eren, 2017; Prathima *et al.*, 2021; Sivaguru e Punniyamorthy, 2021; Yoseph *et al.*, 2020) e também é bastante conhecido e continua sendo um dos métodos mais utilizados para análise e segmentação do cliente devido à sua simplicidade e aplicabilidade (Bult e Wansbeek, 1995). Além disso, o modelo RFM e suas variações, em sua maioria, levam em consideração apenas os dados transacionais das próprias empresas, evitando, desta forma, qualquer questão relacionada à privacidade digital do consumidor ao mesmo tempo que permite às empresas obter insights de comportamento de seus clientes de maneira efetiva, capaz de gerar valor para o negócio.

1.1 Introdução

A partir do desenvolvimento tecnológico de coleta, armazenamento e análise de dados, a importância e a intensidade do relacionamento da empresa com os seus clientes cresceu exponencialmente. Desse modo, foram desenvolvidas técnicas e sistemas de gerenciamento de relacionamento com clientes, conhecidas como *Customer Relationship Management* (CRM), que compõem um grupo de sistemas de informações que permite às organizações entrar em contato com clientes e coletar, armazenar e analisar dados destes clientes para fornecer uma visão abrangente dos mesmos, com o objetivo de medir o valor dos clientes e de suas interações com as empresas (Khodakarami e Chan, 2014).

Um CRM efetivo é capaz de obter uma visão do comportamento e valor dos clientes usando elementos-chave: pessoas, processos e tecnologia. Além disso, os resultados de CRM da companhia devem ser adotados e utilizados efetivamente por seus gestores (Gupta, Aggarwal e Rani, 2016). Dessa forma, a gestão desse conhecimento tem sido vista como um ativo de interesse para as empresas, que perceberam que isso pode contribuir na obtenção de vantagem competitiva. Essa percepção implica na necessidade de utilização dessas ferramentas analíticas de CRM. (García-Murillo e Annabi, 2002; Krafft, Sajtos e Haenlein, 2020; Tavakoli *et al.*, 2018).

Os crescentes recursos das ferramentas e técnicas de análise de negócios e a tomada de decisões orientada por dados tornaram-se ponto focal da estratégia em muitas empresas, especialmente no varejo. Os varejistas há muito reconhecem que a tomada de decisões orientada por dados pode melhorar a qualidade da decisão. Dessa forma, os clientes são considerados o ativo mais importante de uma organização (Chorianopoulos, 2016; Gupta e Lehmann, 2003; Kumar, 2018). De acordo com Cil (2012), o sucesso de uma empresa de varejo é influenciado pela sua rápida capacidade de resposta e sua habilidade de compreender os comportamentos dos seus clientes.

A evolução tecnológica e o barateamento do armazenamento de dados contribuem significativamente para a ampliação do uso de técnicas analíticas e computacionalmente intensivas com o objetivo de buscar novos conhecimentos e insights do comportamento do cliente através da análise de dados coletados a partir das interações do cliente com a marca. (Chen, Chiang e Storey, 2012; Klingenberg, Borges e Antunes, 2021; Madsen, 2019)

Vista como resultado de atividades de CRM, a satisfação do cliente é relacionada positivamente com o valor financeiro da companhia (Garside, 1998; Verhoef *et al.*, 2010). Ao considerarmos que quanto maior o conhecimento da empresa sobre o seu cliente, maior será a sua capacidade de condução de ações de CRM personalizadas e, portanto, maior será o impacto positivo no seu valor financeiro. Ainda, ao assumirmos que a principal forma de conhecer o comportamento do cliente é a partir da análise de dados, quanto maior for a capacidade de análise da empresa, maior será a sua capacidade de criação de valor monetário (Dangi e Malik, 2017; Shah e Murthi, 2021).

Nos mais diversos segmentos de varejo, um dos modelos de classificação de clientes mais utilizados é o modelo RFM (Recência – Frequência – valor Monetário), que se destaca pela sua simplicidade e interpretatividade dos resultados obtidos, o que torna a adoção e utilização efetiva destes segmentos mais rápida pelos gestores, tornando o CRM da empresa mais efetivo.

Este trabalho apresenta um modelo de segmentação de clientes a partir do histórico de consumo do mesmo. Durante o período de três anos, o modelo foi replicado mensalmente, a fim de capturar mudanças de comportamento desses clientes ao longo do tempo, com o objetivo de avaliar a estabilidade longitudinal do modelo proposto. Considerando que o período de análise engloba os anos de 2019 a 2021 e, portanto o período pré pandemia e os dois primeiros anos da Pandemia de Covid-19, a análise proposta permite a geração de insights e conclusões acerca das alterações do perfil de consumo dos clientes durante a pandemia.

1.2 Objetivo

Este trabalho tem como objetivo discutir o impacto do uso intensivo de dados no relacionamento com o cliente e como a pandemia de COVID-19 influenciou o comportamento do consumidor no varejo. Para atingir tal objetivo, analisa-se um conjunto de dados de uma marca de lojas de varejo de moda a partir do desenvolvimento de um modelo estatístico de segmentação de clientes fundamentado no modelo RFM (Bult e Wansbeek, 1995; Colombo e Jiang, 1999; Peker, Kocyigit e Erhan Eren, 2018)

Neste contexto, o trabalho apresenta um caso real de uma empresa de varejo de moda, onde desenvolveu-se um modelo de segmentação de clientes a partir da base de dados transacionais desta empresa.

Considerando a pandemia de COVID-19, que potencializou/ acelerou a transformação digital do varejo e do comportamento do consumidor, este trabalho apresenta uma análise longitudinal do perfil dos clientes da marca analisada. Essa análise longitudinal permite observar e quantificar, na prática, a mudança do comportamento do consumidor nos segmentos propostos.

Para atingir esse objetivo proposto, esse estudo prevê, como objetivos intermediários a serem alcançados, que são listados a seguir

- Busca extensiva de outros trabalhos que aplicam o modelo RFM em indústrias de varejo e análise das aplicações realizadas nesses trabalhos
- Aplicação de técnicas computacionalmente intensivas, especificamente no campo dos modelos de classificação, utilizando como parâmetro de entrada para treinamento do modelo matemático de classificação, entre outras variáveis, o modelo RFM.
- Análise longitudinal dos segmentos de clientes de uma empresa de varejo de moda com base na classificação obtida pelo modelo proposto
- Análise da estabilidade do modelo ao longo do período analisado.

Em termos gerenciais, o modelo proposto pode ser utilizado como base para melhoria na gestão de relacionamento com o cliente e desenvolvimento de campanhas de marketing. Ao realizar a segmentação dos clientes da marca a partir de dados transacionais, o modelo proposto permite aos profissionais de marketing o desenvolvimento de comunicações e promoções de marketing desenvolvidas exclusivamente para cada grupo de clientes. Por utilizar apenas dados transacionais na modelagem de dados, a reprodutibilidade do modelo é razoavelmente simples, por não necessitar de grandes cruzamentos de tabelas de dados, muitas vezes disponíveis internamente nas empresas em ambientes distintos e de integração complexa.

1.3 Relevância

De acordo com KUMAR, RAMACHANDRAN, *et al.*, (2021), o futuro do marketing depende dos esforços das empresas em adquirir um entendimento holístico de seus clientes, assim como suas necessidades e comportamentos. Desta forma, este trabalho traz a aplicação de modelos de inteligência artificial em duas empresas brasileiras de varejo com o objetivo de classificar seus clientes a partir do consumo realizado por eles.

Sabe-se, ainda, que as pesquisas de Inteligência artificial aplicadas ao marketing realizadas até hoje concentram-se em 3 países: Estados Unidos, Europa e oeste da Ásia (Mustak *et al.*, 2021). A análise de dados de empresas brasileiras torna-se relevante no contexto do desenvolvimento estratégico e tecnológico das empresas latino-americanas e de países em desenvolvimento.

Além disso, o acesso às bases de dados reais de empresas é um desafio no contexto de pesquisas acadêmicas, uma vez que tais bases de dados devem possuir grandes quantidades de registros para serem analisadas e muitas variáveis referentes a cada um destes registros, mantendo a confidencialidade dos dados das empresas.

As informações que esse estudo pretende produzir podem se mostrar de interesse para profissionais de marketing pois fornece base teórica para a aplicação de métodos computacionalmente intensivos a partir de modelos teóricos de marketing, ampliando a aplicabilidade desses modelos, no cenário tecnológico atual.

Os resultados a serem alcançados também poderão ser úteis para gestores posto que o modelo resultante desta análise é extremamente simples e seus outputs são facilmente interpretáveis para além do linguajar técnico da ciência de dados.

A análise proposta possibilita ampliar horizontes para a aplicação de modelos de marketing para decisões baseadas em dados, tanto no meio acadêmico quanto gerencial.

1.4 Delimitação

Este estudo aborda o impacto da utilização de dados transacionais pelas empresas sob o ponto de vista do relacionamento com o cliente e como o uso desses dados afeta o valor do cliente ao longo do tempo. Sendo assim, somente dados transacionais que foram disponibilizados serão avaliados. Embora relevante, não se pretende tratar da coleta de dados não transacionais pela empresa, como, por exemplo, dados de redes sociais dos clientes, e nem analisar o impacto desses dados nos modelos de classificação dos clientes, pois esta perspectiva necessitaria de dados que não foram disponibilizados pela empresa neste trabalho.

2. Referencial teórico

Neste capítulo são discutidos aspectos teóricos e estudos relacionados com a análise de comportamento do cliente. Esta seção está dividida em 5 partes: a primeira delas aborda as questões relacionadas à história do desenvolvimento da ciência de dados. Em seguida, o uso de dados nas pesquisas na área de marketing. A seguir, a discussão é direcionada ao uso de dados na gestão do relacionamento com o cliente, contextualizando o problema de pesquisa.

A quarta sessão aborda o modelo RFM (Recência – Frequência – Valor Monetário), em seus aspectos teóricos, gerenciais e metodológicos ao longo dos anos, desde o seu desenvolvimento inicial, analisando as implicações, limitações e consequências do uso do modelo nos mais diversos mercados.

Finalmente, na quinta sessão do capítulo, serão apresentadas as ferramentas utilizadas nas análises e os modelos computacionais que serão utilizados na análise de dados proposta.

2.1 A história é contada a partir de dados

Recentemente, a expressão “*Os dados são o petróleo do século XXI*” foi cunhada por Clive Humby, um matemático londrino especializado em ciência de dados (BAKER TILLY, 2021). Essa expressão se tornou, então, um mantra que vem sendo repetido exaustivamente.

Porém, assim como o petróleo, os dados precisam de tratamento e análises, através de ferramentas específicas. O crescimento exponencial da nossa capacidade de coleta e armazenamento de dados, traz novos desafios computacionais para o processamento de dados. Porém, ao contrário do petróleo, que é um recurso finito, os dados são considerados um recurso imaterial, um bem intangível, ou seja, nunca serão extintos, pois os dados são o ingrediente principal do que nos torna humanos: o conhecimento e a nossa capacidade de transmiti-lo a outros indivíduos.

Enquanto humanos, saímos da pré-história quando inventamos a escrita e passamos a coletar dados. Inventamos o dinheiro e o comércio e aprendemos a

realizar registros financeiros. Evoluímos da transmissão oral de conhecimento entre gerações para a escrita e, posteriormente, a imprensa foi desenvolvida.

A coleta e o armazenamento de dados são inerentes ao ser humano: somos seres capazes de aprender, armazenar a informação e transmiti-las a outros indivíduos. Ao longo da história, transformamos a maneira como coletamos os dados e os transmitimos aos nossos semelhantes e inventamos maneiras de tornar essa transmissão mais eficiente e menos dependente da nossa própria memória: livros, registros, cadernos, fotos, vídeos, diários, agendas e, mais recentemente, nossos computadores e smartphones nos permitem armazenar dados em volume e velocidade nunca vistos.

Foi a partir da terceira revolução industrial, a partir da segunda metade do século XX, com a criação dos primeiros computadores que a capacidade de coleta e armazenamento de dados começou a se expandir exponencialmente.

No início do século XXI, teve início a quarta revolução industrial, baseada no desenvolvimento de tecnologias digitais e na coleta e armazenamento de dados, que entram em todas as esferas das atividades humanas, possibilitando a digitalização da economia. Esse processo permite aumentar a eficácia dos processos gerenciais e produtivos, inclusive no desenvolvimento do marketing.(Klingenberg, Borges e Antunes, 2021)

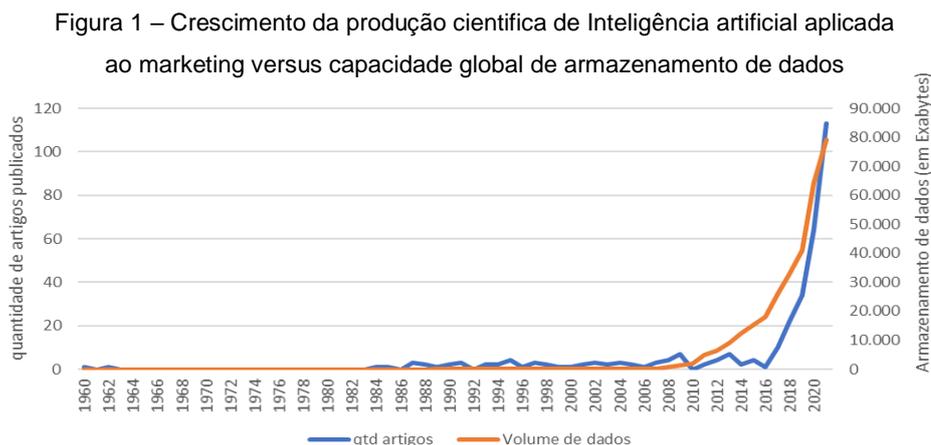
Esse avanço tecnológico produziu mudanças estruturais na estratégia das empresas e modificou paradigmas de negócio (Klingenberg, Borges e Antunes, 2021). Esses avanços, como por exemplo o uso de computadores pessoais, o desenvolvimento e ampliação da internet, o surgimento dos smartphones e, finalmente, a computação na nuvem, ampliaram o potencial de conhecimento das organizações em relação às necessidades dos consumidores (Kumar *et al.*, 2019). Além disso, as empresas em um ambiente baseado em dados criam, disseminam e usam o conhecimento obtido como uma fonte-chave de vantagem competitiva (McEvily e Chakravarthy, 2002).

O uso de tecnologias orientadas por dados permite que as empresas coletem e acessem dados de maneira mais rápida e granular de seus consumidores finais (Kumar, Ramachandran e Kumar, 2021). No contexto da aplicação dessas tecnologias ao CRM, as empresas tornam-se capazes de obter novos insights e melhores respostas às necessidades e desejos de seus clientes, levando a um maior

índice de satisfação dos mesmos e, portanto, entregando mais valor à companhia (Klingenberg, Borges e Antunes, 2021).

As pesquisas na área de inteligência artificial receberam, ente 2019 e 2021, em torno de 5 bilhões de dólares em investimentos de capital de risco. Assim, estão sendo preparadas para exercer efeitos transformadores nos mercados e nas estratégias de marketing ao redor do planeta (Mustak *et al.*, 2021; Rangaswamy *et al.*, 2020). Apesar deste enorme potencial, o marketing ainda possui um gap no entendimento de como as tecnologias de IA podem ser aplicadas atualmente e como poderão evoluir no futuro (Mustak *et al.*, 2021).

Ano a ano, conforme pode ser observado na figura 1, a capacidade de armazenamento e análise de dados cresce exponencialmente. Nunca na história capturamos tantos dados – e nem fomos capazes de processar esse volume de informação de maneira efetiva. Com o rápido desenvolvimento tecnológico, esse processo passou a ser cada vez mais acelerado. Segundo projeções realizadas pelo site Statista, em 2024 o mundo produzirá 149.000 exabytes de dados. A título de comparação, no ano de 2010 foram coletados e armazenados 2000 exabytes de dados ao redor do planeta. Para trazer a noção de ordem de grandeza, um exabyte equivale a um milhão de terabytes. O HD de um notebook em 2021 possui, em média, 1 terabyte de espaço de armazenamento.



Considerando que o primeiro artigo publicado sobre inteligência artificial no marketing ocorreu em 1960 (Head, 1960) e o segundo em 1962 (Goeldner, 1962), o número de artigos publicados sob esse tema permaneceu abaixo de dois dígitos até 2015. Porém, a partir de 2016, há um rápido crescimento em publicações sobre

o tema, demonstrando um intenso crescimento no interesse de pesquisadores nesse campo, em paralelo ao desenvolvimento computacional e tecnológico da IA no mundo e do seu crescente uso em estratégias de marketing (Mustak *et al.*, 2021). Assim, a inteligência artificial (IA) vem ganhando importância nos estudos de marketing.

O uso destas novas tecnologias permite aos profissionais de marketing o desenvolvimento de aplicações baseadas em dados com o objetivo de gerenciar o relacionamento com o cliente (CRM) e guiar a empresa no aproveitamento das oportunidades de negócio, atendendo às necessidades e requisitos de seus clientes. (Taleb, Salahat e Ali, 2020). O uso das tecnologias de Inteligência artificial (IA) permite o exame detalhado e em intervalos de tempo muito curtos de conjuntos de dados massivos, com o objetivo de revelar padrões ocultos, correlações desconhecidas, tendências de mercado e preferências e necessidades dos clientes (White, 2015).

Entre as diversas aplicações possíveis, o uso de grandes volumes de dados permite as mais diversas análises, identificando e potencializando estratégias de negócios com base no perfil de comportamento do cliente. Por outro lado, aumenta a complexidade destas operações. Um exemplo é a necessidade de uso de tecnologias de armazenamento e processamento de dados capazes de lidar de forma eficiente com o enorme volume de dados coletados. Entre as principais vantagens do uso de grande volume de dados, temos a possibilidade de analisar as transações dos clientes em curto intervalo de tempo, de modo a permitir um melhor entendimento dos comportamentos de compra destes clientes, possibilitando a análise estratégica desses comportamentos e utilizando esses resultados para promover a melhoria (ou adequação) de produtos e serviços (Kunz *et al.*, 2017; Taleb, Salahat e Ali, 2020).

As ferramentas computacionais do campo da IA vem sendo desenvolvidas em ritmo acelerado. Algumas destas técnicas utilizam métodos estatísticos computacionalmente intensivos, como as análises de regressão, clusterização e análise de componentes principais. Outras técnicas são mais recentes, baseadas em conceitos computacionais mais recentes, como árvores de decisão e redes neurais artificiais, e também demandam forte utilização de ferramentas computacionais ainda mais robustas.

Cabe ressaltar que muitas das técnicas estatísticas computacionalmente intensivas citadas foram incluídas no escopo da inteligência artificial, como é o caso dos modelos estatísticos de clusterização, notadamente o algoritmo ‘k-means’, que é ao mesmo tempo um método estatístico computacionalmente intensivo e é classificado como um modelo de classificação de aprendizado de máquina. Em outras palavras, as técnicas estatísticas computacionalmente intensivas correspondem a um subconjunto dos modelos de inteligência artificial.

2.20 marketing baseado em dados

Embora encontrem-se registros de práticas que podem ser relacionadas ao marketing até 7.000 a.C., o pensamento de marketing como uma disciplina distinta surgiu a partir do desenvolvimento da economia no século XX. À medida que este campo de pesquisa se desenvolveu, nos primeiros três quartos do século o objeto de estudo era, principalmente as transações e trocas de mercadorias. Nas últimas 3 décadas do século XX, o marketing como campo de estudo voltou-se à analisar os relacionamentos entre as empresas e seus clientes (Sheth e Parvatiyar, 1995) ao mesmo tempo em que o desenvolvimento tecnológico e computacional crescia exponencialmente, conforme descrito na sessão anterior.

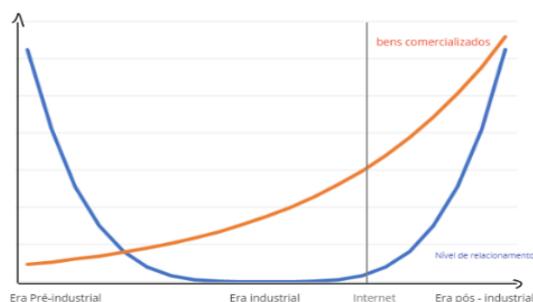
O desenvolvimento computacional e tecnológico marcou o início da era pós-industrial na sociedade. A revolução tecnológica está modificando as atividades das instituições de marketing. O atual desenvolvimento e introdução de sofisticados sistemas informacionais em nossa sociedade está tornando mais fácil para os consumidores interagirem diretamente com os produtores. Os produtores também estão se tornando mais conhecedores de seus consumidores, mantendo e acessando sofisticados bancos de dados capazes de capturar informações relacionadas a cada interação com consumidores individuais, a um custo muito baixo. Dessa forma, a tecnologia permite a individualização da prática de marketing para um grande volume de clientes. (Sheth e Parvatiyar, 1995).

Desta forma, a orientação para o relacionamento com o cliente no contexto do marketing está de volta, desta vez em um novo patamar de capacidade produtiva da sociedade, conforme é observado na figura 2. Na era pré-industrial, o relacionamento com o cliente era individualizado e limitado ao número de relacionamentos do comerciante com a sociedade. A industrialização levou a uma

inversão na relação entre oferta e demanda, quando devido aos esforços de produção em massa os produtores criaram excesso de oferta de bens e serviços, levando ao desenvolvimento do marketing de massa, deixando de lado a individualização deste contato com o consumidor.

Desta forma, este desenvolvimento tecnológico causou impactos nas estratégias de marketing. A tecnologia ampliou a capacidade das empresas de se relacionarem com os seus clientes: o surgimento das grandes mídias e de outras tecnologias personificadas de comunicação de marketing, como os catálogos tiveram grande impacto nas estratégias de marketing. (Rust, 2020).

Figura 2 - Evolução relacionamento com o cliente versus capacidade produtiva adaptado de Klingenberg et al. (2021)



Fonte: elaboração própria.

A grande mudança seguinte foi o surgimento da internet. Em particular, a Internet e outras tecnologias de informação modernas resultaram em relacionamentos mais profundos com os clientes, facilitando um CRM mais eficaz (Winer, 2001) e expansão do setor de serviços (Rust e Huang, 2014). Ou seja, uma visão de marketing verdadeiramente centrada no cliente é o resultado natural da tecnologia, baseada no valor do cliente (Rust *et al.*, 2004)

Na era pós-industrial, podemos dividir o marketing em duas fases: o marketing de massa, anterior à “Revolução dos dados”, e o marketing baseado em dados.

Na primeira fase, o marketing baseou-se nas premissas dos 4 P’s: Produto, Preço, Promoção e Praça. Era necessário mostrar o produto, definir um preço único, realizar ações promocionais e convencer o cliente, do qual não havia muitas informações, a ir até o ponto de venda para adquirir o produto ou serviço. A comunicação das ofertas e as propagandas não poderiam ser direcionadas a um

público específico pois não havia meios de obter micro segmentos de clientes e esse tipo de ação seria extremamente caro para a empresa. (Haley, 1968)

Desde então, os profissionais de marketing concluíram que consumidores diferentes possuem comportamentos diferentes e, por isso, reagem de maneira diferente às comunicações de marketing. (Haley, 1968, 1988) A partir desta percepção, estes profissionais começaram a buscar meios de diferenciar o contato com os consumidores, através da segmentação de mercados. Historicamente, talvez o primeiro tipo de segmentação a existir foi a segmentação geográfica. Os pequenos fabricantes, que desejavam limitar seus investimentos, ou cujos canais de distribuição não eram grandes o suficiente para cobrir todo o país, segmentaram os seus mercados, com efeito, vendendo seus produtos apenas em algumas áreas do país. (Haley, 1968)

Com a evolução da tecnologia e, principalmente, da internet, o marketing passou por uma grande transformação: comunicações direcionadas ao cliente específico, enviadas por e-mail e a baixíssimo custo. Os primeiros e-mails com conteúdo de marketing personalizado foram disparados ainda na década de 90: essa personalização se resumia ao título do e-mail contendo o nome do cliente e o conteúdo propriamente dito ainda era basicamente o mesmo para todos os clientes. (Dangi e Malik, 2017)

Com a crescente digitalização dos meios de comunicação das últimas décadas, desenvolveram-se novas capacidades de marketing baseadas em análises de grandes volumes de dados. Ano após ano, essas capacidades evoluem exponencialmente – tornando a gestão do relacionamento com o cliente cada vez mais segmentada e assertiva para gerar novos insights e direcionamentos estratégicos aos profissionais de marketing. (Dangi e Malik, 2017)

Mais recentemente, com o crescimento exponencial da nossa capacidade tecnológica e o barateamento dos custos de armazenamento de dados, nunca antes fomos capazes de coletar, armazenar e interpretar dados dos clientes em tão pouco tempo e em volume tão grande. Isso permite ao profissional de marketing ampliar o horizonte de planejamento estratégico do foco do produto para o foco no cliente. Essa ampliação de escopo das estratégias de marketing é o ponto de partida para estudos de relacionamento com o cliente como o conhecemos nos dias de hoje.

O uso de dados como espinha dorsal da gestão de relacionamento com os clientes (do inglês *Customer Relationship Management*, CRM) vem se tornando

cada vez mais frequente. Deste modo, evoluímos do modelo de gestão de marketing personalizado baseado no produto para um modelo centrado na experiência do cliente. As decisões baseadas em dados são parte fundamental do campo de CRM nas empresas. Segundo Khodakarami e Chan (2014), o termo *Customer Relationship Management* (CRM) refere-se ao conjunto de técnicas e sistemas de gerenciamento de clientes que permite às organizações o contato direto [e personalizado] com o cliente, assim como a coleta e o armazenamento desses dados.

Desse modo, um CRM é dito bem-sucedido se é capaz de obter uma visão do comportamento e valor dos clientes usando elementos-chave: pessoas, processos e tecnologia. Além disso, os resultados de CRM da companhia devem ser adotados e utilizados efetivamente por seus gestores (Gupta, Aggarwal e Rani, 2016). Deste modo, pode-se definir CRM como um conjunto de estratégias e práticas que ajudam as empresas a gerenciar e analisar as interações com os clientes em relação aos produtos e serviços da empresa para ajudar a satisfazer seus clientes, além de desenvolver a sua base de clientes. (Taleb, Salahat e Ali, 2020).

Dentre todos os aspectos que envolvem o desenvolvimento e a manutenção de um CRM efetivo, o armazenamento de todas as informações relacionadas ao consumidor e a manutenção desses dados para análises posteriores é um dos principais desafios (Khodakarami e Chan, 2014). Através desses dados coletados, a demanda do cliente é analisada com o objetivo de extrair informações úteis à tomada de decisão estratégica da empresa sobre como deve se relacionar com os seus clientes de maneira a maximizar o valor de cada cliente.

2.3 CRM baseado em dados

Conforme discutido acima, a gestão do relacionamento com o cliente foi transformada pela tecnologia. Deste modo, o CRM passou a ser definido como a prática de analisar e utilizar bancos de dados de marketing e alavancar tecnologias de comunicação para determinar práticas e métodos corporativos que maximizarão o valor da vida útil de cada cliente (Verhoef *et al.*, 2010). Assim, o CRM é uma abordagem estratégica que tem por objetivo criar valor para a companhia, através do desenvolvimento de relacionamentos adequados com os principais clientes e segmentos de clientes. O CRM une o potencial das estratégias de marketing de relacionamento e da tecnologia da informação para criar relacionamentos lucrativos

e de longo prazo com os clientes e outras partes interessadas, oferecendo oportunidades baseadas no uso de dados, obtendo informações para entender os clientes e criar valor com eles. (Payne e Frow, 2005).

Na prática, o CRM implica na realização de análises de dados dos clientes, através de bancos de dados de marketing, aproveitando a disponibilidade de tecnologias de comunicação para determinar práticas e métodos corporativos que maximizem o valor de vida de cada cliente para a empresa (Kumar *et al.*, 2012). Com o avanço tecnológico recente, as análises de CRM passaram a contar com o uso de técnicas oriundas da ciência de dados, potencializando a capacidade das empresas em identificar clientes e segmentos de clientes, prever comportamentos específicos nesses grupos e, a partir daí, direcionar suas comunicações e ações de marketing de forma a criar valor para a companhia.

Historicamente, as iniciativas baseadas em CRM tem por objetivo atrair e reter a lealdade do cliente e manter (e ampliar) a vantagem competitiva (Smith, 2010). Do ponto de vista estratégico, as empresas devem buscar e manter clientes valiosos. Isso requer agregar valor a cada interação com o cliente. Desse modo, o CRM se faz necessário para a existência de um negócio (Das e Mishra, 2018).

Uma estratégia de CRM pode ser caracterizada em três etapas: primeiro, a aquisição de dados do cliente; segundo, a segmentação de clientes em grupos homogêneos; e, finalmente, a aplicação de estratégias diferenciais de gerenciamento de relacionamento para cada segmento identificado e, em algumas circunstâncias, para cada cliente (Nairn e Bottomley, 2003).

O tema central do CRM gira em torno das relações individuais entre consumidor e vendedor. Essas relações são consideradas de natureza longitudinal, em que ambas as partes se beneficiam do estabelecimento dessa relação simbiótica. Do ponto de vista estratégico, o conceito de CRM pode ser visto como uma cultura ou valor organizacional distinto que coloca a relação entre consumidor e vendedor no centro operacional da empresa (Sin, Tse e Yim, 2005). Dessa forma, o CRM é considerado uma ferramenta útil para as empresas na manutenção de relacionamentos valiosos e duradouros com os seus clientes (Gupta, Aggarwal e Rani, 2016). Em outras palavras, a estratégia de CRM baseado em dados tem como questão central a escolha de quais clientes, com base nos dados disponíveis, a empresa deve atingir com uma determinada oferta (Colombo e Jiang, 1999).

2.4 CRM e Segmentação no Varejo

Criar uma experiência superior para o cliente parece ser um dos objetivos centrais dos ambientes de varejo atuais. Varejistas de todo o mundo adotaram o conceito de gerenciamento da experiência do cliente, com muitos incorporando a noção em suas declarações de missão (Verhoef *et al.*, 2009).

Uma das formas mais comuns de buscar essa experiência superior do cliente é através da segmentação do mercado. Através da segmentação, é possível identificar grupos de clientes com perfis similares e, a partir desses grupos, personalizar a experiência individual do cliente em função do seu histórico de comportamento de consumo. Existem vários benefícios da segmentação de mercado. Por exemplo, a segmentação de mercado fornece uma capacidade aprimorada de identificar e explorar novas oportunidades de mercado para benefício comercial, e é uma área importante para o desenvolvimento estratégico de produtos, sendo útil para medir o nível de satisfação do cliente, assim como para entender e caracterizar o comportamento dos clientes (Fonseca e Cardoso, 2007). Ao segmentar um mercado através de desejos e necessidades semelhantes, entender esses segmentos, o varejista cria um conceito de marca que seja significativo para esse segmento e usa o conceito para criar uma experiência de compra exclusiva para cada segmento (Verhoef *et al.*, 2010).

Dessa forma, ao segmentar um mercado e identificar perfis de comportamento do cliente, generalizamos o comportamento desses clientes dentro dos grupos com o objetivo de personalizar os contatos com o cliente de cada grupo. Ou seja, para personalizar o relacionamento com o cliente, as ferramentas da ciência de dados nos levam a criar generalizações de comportamento para grupos menores de clientes. (Verhoef *et al.*, 2010)

O ponto de partida do modelo conceitual de CRM é o reconhecimento de que o ambiente de varejo é, atualmente, repleto de dados, em volume esmagador, tanto a nível do cliente individual, ao nível de produto e ao nível da loja (Verhoef *et al.*, 2010). Com dados oriundos de diversas fontes informacionais, desde o leitor de códigos de barra do ponto de venda, passando por dados do cliente e dados de transações financeiras, até dados de fornecedores, a integração informacional é um verdadeiro desafio para as empresas varejistas. Uma vez que esses dados estejam corretamente armazenados e integrados, a próxima etapa no processo de CRM é a

transformação dos dados em informações gerenciais úteis à tomada de decisão estratégica.

As decisões de CRM baseiam-se, principalmente, em criar contatos relevantes entre o consumidor e a marca, selecionar clientes apropriados (de acordo com o perfil identificado) para receber esses contatos e entregá-los no momento certo ao cliente (Verhoef *et al.*, 2010). Em outras palavras, o desafio da decisão de CRM consiste em entregar a oferta certa, para o cliente certo, no momento correto, de maneira a melhorar o seu relacionamento com a marca em questão.

Finalmente, de acordo com Gupta (2004), não podemos deixar de lado o fato dos resultados da estratégia de CRM afetarem o valor da empresa. Ou seja, ao melhorar a relação do cliente com a marca, espera-se que essa melhoria seja refletida no valor do cliente para a companhia. Dessa forma, criar uma experiência superior ao cliente parece ser um dos objetivos centrais nos ambientes de varejo atuais.

Os resultados das ações de marketing tomados a partir das ações estratégicas de segmentação de clientes podem ser medidos de diversas maneiras. Esse impacto pode ser de dois tipos: tático ou estratégico (Payne e Frow, 2005; Verhoef *et al.*, 2010).

No nível estratégico, a ação de CRM é abordada de maneira ampla, com o objetivo de gerenciar relacionamentos com os clientes, a fim de criar valor para o acionista. Nesse nível, os resultados possíveis das ações de CRM podem ser: desenvolvimento ou ampliação de vantagem competitiva sustentável, melhoria da eficiência de canal de relacionamento da empresa com o cliente, entre outros. Esses resultados levam a retornos financeiros para a empresa (Verhoef *et al.*, 2010). Já a nível tático, um exemplo seria a implementação de um projeto de solução tecnológica, que possibilite, por exemplo, uma melhoria na satisfação do cliente, aquisição de novos clientes e retenção de clientes antigos, levando a uma maximização do valor de vida do cliente (Verhoef *et al.*, 2010).

Para implementar CRM baseado em dados de maneira eficaz, foram desenvolvidas diversas ferramentas a partir da aplicação de tecnologias de análise de grandes volumes de dados. Dentre essas tecnologias, destacam-se os modelos de inteligência artificial (IA) e machine learning (ML).

2.5 O modelo RFM: passado, presente e futuro

Logo após o desenvolvimento do marketing enquanto escola de pensamento (Alderson e Cox, 1948), surgiram os primeiros estudos que indicam a necessidade de segmentar os clientes como alternativa para as estratégias de marketing. Smith (1956) questiona a aplicabilidade da teoria da concorrência perfeita em marketing, ao observar que tal teoria pressupõe homogeneidade entre os componentes tanto do lado da demanda quanto do lado da oferta do mercado. Para fundamentar seu questionamento, Smith discute a não-homogeneidade da demanda do consumidor por produtos, trazendo luz à discussão da importância de se segmentar clientes como uma das principais estratégias de marketing.

Pouco tempo depois, em 1960, Head (1960) amplia a discussão iniciada por Smith, incluindo, pela primeira vez, a importância do uso de dados na segmentação de clientes a partir do desenvolvimento de ferramentas que possibilitam a automação da coleta e análise de dados. Em 1968, Haley (1968) aprofunda a discussão sobre a importância da segmentação de clientes em marketing, afirmando que a segmentação de clientes baseia-se na demanda de mercado e, dessa forma, representa um ajuste racional e mais preciso do produto e do esforço de marketing aos requisitos do consumidor ou usuário. Propõe ainda que a segmentação de marketing busque identificar quais benefícios [ou características] cada grupo de clientes busca em um produto.

Punj e Stewart, em 1983, apresentam uma revisão dos artigos publicados, até então, sobre clusterização em marketing, apresentando ainda aplicações e sugestões de análise. A barreira tecnológica para processamento de dados fica evidente quando o autor conclui que “clusterizar todas as observações pode não ser uma boa prática”. Além disso, estes autores chamam a atenção para a importância de se atentar à qualidade das variáveis selecionadas para o método de clusterização, pois a escolha inadequada das variáveis pode distorcer o resultado, tornando a segmentação obtida inútil.

Vinte anos após o primeiro artigo, Haley (1988) revisita a teoria da “segmentação por benefícios” e fornece diretrizes para o uso adequado e sugere orientações para melhorias adicionais no método. Para tal, orienta o uso de informações geográficas, demográficas, comportamentais e de volume de uso, além de variáveis de estilo de vida, benefícios percebidos, valores e crenças dos

indivíduos e predisposição ao consumo da marca. Apesar da aplicabilidade e assertividade desse tipo de segmentação, ela esbarra, ainda hoje, na dificuldade de uma coleta precisa de dados para obter uma segmentação de alto valor para as empresas.

Piercy and Morgan (1993) examinam a relevância da segmentação de mercado como estratégia de marketing e sugerem uma perspectiva mais concentrada no contexto organizacional e ambiental da segmentação, com o objetivo de fornecer tanto uma maior compreensão do escopo potencial da segmentação quanto uma visão mais eficaz na implementação de estratégias baseadas em segmentos. Propõem ainda que sejam realizadas segmentações distintas para cada propósito da organização, sejam eles estratégicos, gerenciais ou operacionais.

Em 1995, quarenta e sete anos após o artigo seminal de Alderson and Cox, Bult e Wansbeek (1995) discutem, pela primeira vez, o modelo RFM (Recência, Frequência e Valor Monetário) a partir dos seus aspectos analíticos e estatísticos, permitindo a segmentação de campanhas de marketing direto, enviadas por correio para os clientes. Dadas as baixas taxas de resposta neste setor (geralmente em torno de 2%), esses modelos foram desenvolvidos para direcionar programas de marketing em clientes específicos com o objetivo de melhorar as taxas de resposta. Antes desses modelos, as empresas normalmente usavam perfis demográficos de clientes para fins de segmentação. No entanto, a pesquisa sugere fortemente que as compras passadas dos consumidores são melhores preditores de seu comportamento de compra futuro do que os dados demográficos (Gupta *et al.*, 2006).

O modelo RFM baseia-se em três variáveis de consumo dos clientes: a Recência, que indica o tempo passado desde a última interação do cliente com a marca, a Frequência, que indica o número de interações em determinado período e, finalmente o Valor Monetário gasto por cada cliente em determinado período. O modelo RFM de aplicação mais simples divide os clientes em 5 classes, para cada variável, com base nos quintis das distribuições observadas para cada variável. Aplicada desta maneira, a análise RFM foi, muito provavelmente, o primeiro “modelo preditivo” a utilizar dados em marketing. Outra possibilidade é a atribuição de pesos distintos para cada variável com a posterior criação de um score RFM para cada cliente (Blattberg, Kim e Neslin, 2008; Gupta *et al.*, 2006; Hughes, 1996).

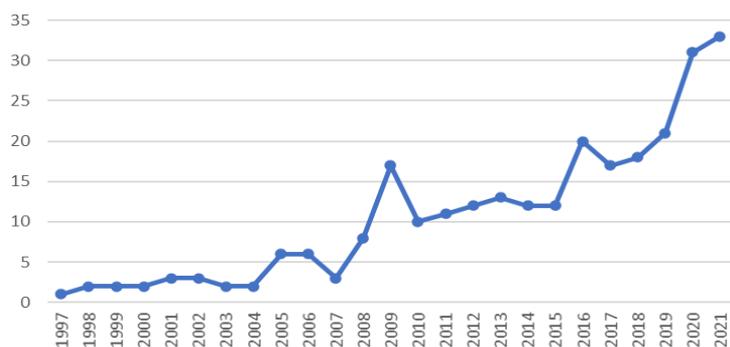
Desenvolvido para lidar com comunicações de marketing direto, antes mesmo da popularização da internet, o modelo RFM adaptou-se muito bem às demandas de CRM. (Anitha e Patil, 2020; Chiang, 2017; Kabasakal, 2020; Rust e Verhoef, 2005; Yoseph *et al.*, 2020)

Dessa forma, o modelo RFM ainda é um dos métodos mais utilizados para segmentar clientes no setor de varejo. Com o desenvolvimento acelerado das tecnologias de inteligência artificial, torna-se computacionalmente viável o uso do modelo RFM dentro destas aplicações, elevando a capacidade das empresas em segmentar seus clientes. Ao longo do tempo e com a ampliação da capacidade tecnológica de análise de dados, o uso deste modelo vem sendo aprimorado e reinventado. (Carrasco *et al.*, 2019)

Esse modelo fornece uma análise eficaz para os tomadores de decisão a fim de atingir seus clientes e desenvolver estratégias de marketing adequadas de acordo com seus comportamentos anteriores (Martínez, R G *et al.*, 2021; Piskunova e Klochko, 2020; Sivaguru e Punniyamoorthy, 2021; Tavakoli *et al.*, 2018). O método ajuda a diferenciar os clientes da empresa a partir da análise de dados estritamente transacionais, possibilitando à empresa que busque a manutenção dos melhores relacionamentos (Anitha e Patil, 2020; Behera *et al.*, 2020; Chorianopoulos, 2015). Para tal, o método classifica os clientes em diferentes segmentos com base em sua taxa de recência, frequência e valor monetário. A ideia básica que sustenta essa análise é que clientes com melhores RFM comprarão continuamente (Blattberg, Kim e Neslin, 2008; Kabasakal, 2020; Peker, Kocyigit e Erhan Eren, 2018). Este modelo mostra que diferentes segmentos de clientes exigem um mix de marketing diferente, ou seja, os insights derivados da análise de RFM são posteriormente usados para apoiar o desenvolvimento e o planejamento da estratégia de marketing da empresa. Um dos maiores benefícios do modelo RFM é a interpretabilidade dos indicadores gerados. (Sivaguru e Punniyamoorthy, 2021)

Devido a estas características, o número de estudos envolvendo o modelo RFM vem se ampliando. Uma busca de artigos através da plataforma Scopus, usando as palavras chaves “RFM and Marketing” retornou 270 artigos publicados desde 1998 com o tema.

Figura 3 – Quantidade de artigos sobre o modelo RFM publicados anualmente



Fonte: elaboração própria.

Como pode ser visto na Figura 3, a partir de 2017 há um crescimento exponencial no número de artigos publicados sob o tema, período consistente com o aumento da capacidade de armazenamento e análise de dados pelas empresas. Ao mesmo tempo, o desenvolvimento tecnológico também permitiu o crescimento da complexidade das análises realizadas pelas empresas.

2.6 Modelo RFM no Varejo: uma análise exploratória

Uma das principais aplicações do modelo RFM ocorre no varejo, nos seus mais diversos segmentos: comércio, cafeterias, bancos, telecomunicações, entre outros. Desta forma, esta sessão apresentará uma série de casos aplicados do modelo RFM em empresas de varejo, com o objetivo de exemplificar a utilização deste modelo em diversos cenários ao redor do mundo.

Uma das aplicações mais comuns do modelo RFM é o uso de suas variáveis formadoras como variáveis de entrada para modelos de segmentação de clientes baseados em inteligência artificial. Dentro desta classe de modelos, o maior destaque é dado ao método de classificação denominado *k-means*. Para maiores detalhes teóricos do método, consultar Hair et al. (2014).

A Tabela 1 a seguir apresenta os artigos publicados nos últimos anos, com aplicações de dados reais de diversos segmentos de varejo, que de alguma forma utilizam as variáveis do RFM como variáveis de entrada para a aplicação de modelos computacionalmente intensivos de análise de grandes volumes de dados.

Tabela 1: Trabalhos publicados

Referência	Segmento de mercado	Utilização RFM
(Li et al., 2017)	não especificado	K-means
(Peker, Kocyigit e Eren, 2017)	supermercados	K-Means

(Martínez <i>et al.</i> , 2019)	loja de móveis	K-means
(Yoseph e Heikkila, 2019)	moda	K-means
(Zhang <i>et al.</i> , 2020)	não especificado	modelos de classe latente
(Ahn, Kim e Lee, 2020)	bancos	modelos de previsão
(Hsu e Huang, 2020)	supermercados	K-means
(Anitha e Patil, 2020)	não especificado	K-means
(Yoseph <i>et al.</i> , 2020)	não especificado	K-means
(George e Gaikwad, 2020)	e-commerce	K-means
(Lin <i>et al.</i> , 2021)	supermercados	K-means, fuzzy C-means, mapas auto-organizados e análise discriminante
(Kit e Azmi, 2021)	e-commerce	K-means
(Martínez, R G <i>et al.</i> , 2021)	moda	AHP (modelos hierárquicos)
(Rahim <i>et al.</i> , 2021)	não especificado	Redes neurais e árvores de decisão
(Zhao e Li, 2021)	e-commerce	K-means
(Seetha e Thakar, 2020)	não especificado	K-means
(Carrasco <i>et al.</i> , 2019)	não especificado	K-means
(DehghaniZadeh, Fathian e Gholamian, 2018)	passagens aéreas	AHP (modelos hierárquicos)
(Tavakoli <i>et al.</i> , 2018)	e-commerce	K-means
(Haghighatnia, Abdolvand e Rajae Harandi, 2018)	não especificado	K-means

Fonte: elaboração própria.

No ambiente competitivo atual, a análise de cluster como ferramenta de segmentação de dados é um método bastante popular de busca de conhecimento comercial nos mais diversos mercados (Peker, Kocyigit e Eren, 2017). Um dos principais objetivos da ciência de dados em varejo é extrair *insights* acionáveis e entender, com base nos dados minerados, o comportamento do cliente (Boratto et al., 2018; Karimi, 2019; Motiwalla et al., 2019).

A análise do comportamento do consumidor é vital para que os varejistas hoje em dia compreendam melhor e atendam às necessidades dos clientes (Chang et al. 2014; Griva et al. 2018; Zhou et al. 2020). Como ferramenta mais utilizada nas publicações recentes descritas na tabela 1, o método k-means é o mais frequentemente utilizado para este tipo de análise.

Todos os artigos citados na tabela 1 utilizam, ao menos em parte, o modelo RFM para definir as variáveis de entrada dos modelos propostos. Ao discutir os modelos de segmentação de clientes com base em algoritmos de classificação, o pesquisador toma para si a escolha das variáveis que determinarão o modelo final.

Deste modo, a escolha das variáveis que serão incluídas no modelo de segmentação de clientes tem impacto significativo nos resultados obtidos.

Ao optar pela utilização de variáveis bastante fundamentadas na literatura de marketing, como é o caso do modelo RFM, os pesquisadores buscam garantir a explicabilidade das variáveis selecionadas para o modelo, tornando a utilização desses resultados mais simples.

Diversos autores aplicam conjuntamente o modelo RFM no campo de estudos de aprendizado de máquina para obter modelos de segmentação de clientes baseados em dados transacionais. Tavakoli et al. (2018) reforça a simplicidade e efetividade do uso do modelo RFM na segmentação de clientes de varejo, trazendo uma aplicação com dados de uma grande rede de e-commerce do Irã. Zhang et al. (2020) o faz através da análise latente probabilística, que é amplamente utilizada na posterior clusterização dos clientes. A utilização do modelo de classe latente permite o agrupamento de clientes em classes latentes e o cálculo das probabilidades de atribuição de cada cliente a cada classe. Neste artigo, os autores propõem um novo modelo de classe latente para a análise de RFM com base nos dados do histórico de compras. O modelo proposto permite decidir a pontuação de RFM e segmentar clientes automaticamente, através de técnicas de análise de cluster. Rahim et al. (2021) aplica o modelo de recência, frequência e monetário (RFM) e técnicas de inteligência artificial e modelagem de dados (Redes Neurais Artificiais do tipo MLP, Support vector machine (SVM) e árvores de decisão) para detecção de padrões de comportamento de um cliente. Já Ahn et al. (2020) aplica o modelo RFM na indústria de corretagem de valores e bancos de investimento, utilizando grande volume de dados, confirmando a aplicabilidade do modelo RFM, em conjunto com modelos de análise de grandes volumes de dados, é capaz de gerar valor para o setor.

(Li *et al.*, 2017) discute a importância dos dados de varejo para empresas orientadas pelo cliente, especialmente na geração de insights de negócio. Apresenta o modelo KID, que é um modelo cognitivo de dados para conhecimento, pode ser aplicado a negócios de varejo para dar suporte à análise de dados de varejo. Para demonstrar a aplicação proposta, o artigo apresenta um estudo de caso de avaliação de valor para o cliente através dos algoritmos RFM e K-means. Três anos depois, Anitha & Patil (2020) apresentam um artigo de aplicação do modelo RFM através

da análise de cluster utilizando dados de uma empresa varejista em streaming, ou seja, dados coletados em tempo real.

Martínez, Carrasco, Sanchez-Figueroa, & Gavilan (2021) apresentam um método de segmentação de clientes baseado no modelo linguístico fuzzy e na definição de hierarquias de produtos, utilizando o modelo RFM como entrada do modelo proposto e, dessa forma, fornecendo uma interpretabilidade linguística ao modelo de classificação proposto, permitindo a compreensão estratégica do resultado. O modelo proposto, denominado pelos autores de “modelo linguístico fuzzy de 2 tuplas RFM” foi desenvolvido a partir dos modelos de hierarquia de produtos (AHP), revela-se uma ferramenta útil para incluir critérios de negócios, catálogos de produtos e insights de clientes na definição de estratégias comerciais.

Lin et al. (2021) usa o modelo RFM para converter dados de transações do cliente em valores de entrada e, em seguida, agrupa esses clientes usando métodos de Inteligência artificial, combinando os modelos K-means, fuzzy C-means e mapas auto-organizados. Em seguida, usando a análise discriminante, escolhe-se o método de maior capacidade de discriminação de clientes e, a partir dele, é calculada uma score de pontuação para o CLV do cliente.

O modelo RFM é associado a métodos computacionalmente intensivos não apenas para segmentar clientes, como para identificar cestas de produtos específicas para os grupos de clientes. Hsu e Huang (2020), por exemplo, utilizam o modelo RFM como preditor do comportamento do cliente na identificação de produtos críticos para cada perfil de consumo. Desse modo, amplia os horizontes de aplicabilidade do modelo RFM na caracterização do perfil de consumo dos mesmos. Já George e Gaikwad (2020) utilizam o modelo RFM para segmentar os clientes e, a partir dos grupos encontrados, realizar previsões de fluxo de compras desses clientes, através de análise de series temporais. Dessa forma, amplia ainda mais as possibilidades de aplicação do modelo RFM, obtendo ainda mais clareza na identificação dos padrões de consumo dos clientes analisados.

Yoseph & Heikkila (2019) combinam os conceitos de RFM e CLV (Customer Lifetime Value) para segmentar clientes de um varejista de roupas e acessórios de moda sediada no Kuwait. Para atingir tal objetivo, aplicou-se um modelo de regressão com o objetivo de analisar o comportamento de compra do cliente. Na etapa seguinte, foi utilizado, entre outros métodos, a aplicação do algoritmo K-means para segmentar os clientes da varejista analisada, concluindo que a análise

RFM é mais efetiva que a aplicação do CLV na detecção de segmentos de mercado relevantes e estratégias de marketing adequadas para a indústria.

Ao mesmo tempo, ao unir o modelo RFM aos modelos de aprendizado de máquina, novas variáveis passaram a ser incluídas no processo de modelagem, dentre as quais se destaca o modelo LRFM(P). Peker et al. (2017) e Martínez et al. (2019) propõem a inclusão da dimensão “duração do relacionamento do cliente com a empresa” no modelo RFM. Peker et al. (2017) inclui ainda a dimensão “periodicidade” para a classificação desses cliente, num modelo denominado LRFMP (Length, Recency, Frequency, Monetary and Periodicity), e aplica o modelo proposto na classificação de clientes no setor de varejo alimentar, identificando diferentes segmentos de clientes nesta indústria. Martínez et al. (2019) e Seetha e Thakar, (2020) discutem o impacto da dimensão “duração do relacionamento” no que diz respeito à fidelização de clientes. O primeiro propõe a aplicação de um modelo de lógica fuzzy como solução para a identificação dos perfis dos clientes e aplica seus achados em uma base de dados de uma varejista de roupas de comércio eletrônico. O último, aplica a variável duração do relacionamento através do algoritmo k-means.

Mais recentemente, em 2021, Kit & Azmi examinam a utilidade do modelo LRFMP utilizando dados de uma indústria de varejo online sediada na Malásia, incluindo na análise os resultados do índice de silhueta e a automação da definição do número de clusters proposto no modelo. No caso da Malásia, as variáveis incluídas no modelo não foram consideradas significativas, e a hipótese do efeito do grande volume de clientes sem recorrência de compra seja a responsável pela ausência dos efeitos de periodicidade e comprimento do relacionamento do cliente na base utilizada nos estudos.

O modelo RFM também pode ser acrescido de variáveis específicas do segmento que está sendo analisado. Por exemplo, DehghaniZadeh et al. (2018) inclui a variável “distância percorrida” ao segmentar clientes de uma companhia aérea usando modelos hierárquicos de segmentação.

O modelo RFM associado a algoritmos de Inteligência artificial também é explorado no ambiente de e-commerce. Zhao & Li (2021) e Rahim et al. (2021) fazem tal análise. O primeiro aplica o método k-means++, utilizando ponderação por pesos para segmentar clientes online. O segundo, inclui as dimensões de periodicidade e duração do relacionamento do cliente com a marca.

Uma das aplicações do modelo RFM associado a algoritmos de segmentação é no gerenciamento de campanhas de marketing, conforme pode ser visto em Carrasco et al. (2019), que apresenta um esquema de gerenciamento de campanha relacional integrado baseado no processo analítico RFM que incorpora modelos de Inteligência artificial para obter uma maior precisão e uma melhor interpretabilidade linguística dos resultados do modelo RFM e a posterior segmentação, a fim de desenvolver um plano de campanha mais eficaz. Outra aplicação possível, demonstrada por Haghghatnia et al. (2018) é na formulação do modelo RFM na utilização de descontos por parte dos clientes para analisar o seu comportamento. Neste trabalho, Haghghatnia calcula as dimensões RFM com base nos descontos concedidos aos clientes e os segmenta, identificando o impacto dos descontos tanto no comportamento dos clientes como na lucratividade da empresa.

2.7 O impacto da pandemia de Covid-19 no comportamento do consumidor

Um dos eventos mais dramáticos da última década é a pandemia de covid-19. Desde que o vírus se espalhou rapidamente ao redor do mundo, intervenções governamentais se alastraram muito rapidamente ao redor do mundo. Intervenções desta magnitude nos comportamentos da população não eram observadas desde o fim da segunda guerra mundial (Verhoef, Noordhoff e Sloot, 2022).

O vírus COVID-19 mudou a cultura de consumo das pessoas. Em particular, a COVID-19 exigiu mudanças contínuas nos principais mercados onde as pessoas passaram a comprar produtos de consumo offline para online. O medo da infecção e as políticas de distanciamento social levaram os consumidores a ficar dentro de casa e a reduzir o contato pessoal. Além disso, como os estilos de vida mudaram rapidamente para os modos internos, as compras offline diminuíram e os canais de distribuição mudaram à medida que o mercado de consumo online não presencial cresceu (Tran, 2021; Truong e Truong, 2022; Verhoef, Noordhoff e Sloot, 2022).

Verhoef, Noordhoff e Sloot (2022) propõe um framework teórico para analisar os efeitos da pandemia de COVID-19 no varejo, a partir do pressuposto de mudança nas necessidades de consumo e de comportamento desses consumidores. Verhoef et al. (2022) defende que a transformação digital e tecnológica no momento da pandemia acelerou a migração do consumo offline para o online, independente da idade (Basu e Swaminathan, 2021).

Da perspectiva do mercado de varejo, Verhoef, Noordhoff e Sloot (2022), no framework proposto, reforçam a necessidade de respostas táticas e estratégicas a essa mudança de perspectiva do consumidor.

3. Metodologia

A partir de registros transacionais de uma empresa de varejo brasileira foram obtidos indicadores do modelo de Recência, frequência e valor para cada cliente. A partir destes indicadores, foi aplicado um modelo computacionalmente intensivo para segmentar esses clientes com base em seus perfis de consumo na marca. A partir dos resultados obtidos mês a mês, foi realizada uma análise longitudinal do comportamento de consumo dos clientes nesta marca.

3.1 Design da Pesquisa

Para este trabalho foi utilizado um conjunto de dados que tem origem em uma marca de vestuário brasileira, que possui diversas lojas espalhadas por todo o Brasil

O conjunto de dados disponibilizado pela empresa inclui as variáveis de recência, frequência e valor a nível de cliente. Os dados foram extraídos dos registros informacionais de vendas da rede varejista e agregado ao nível de cliente. Essa base de dados engloba todos os clientes cadastrados da rede no período analisado. Não foi realizada amostragem nos dados (foi utilizada a totalidade dos clientes da empresa nos períodos avaliados). O objetivo é identificar, através de modelos preditivos e técnicas de inteligência artificial, segmentos de clientes em função do perfil e comportamento de consumo individuais, para cada segmento de mercado estudado, permitindo que a empresa desenvolva estratégias de marketing com base no comportamento de seus clientes.

O trabalho busca segmentar clientes com base nas variáveis de Recência, Frequência e Valor dos mesmos, a partir de dados transacionais disponibilizados pela empresa. A partir dos segmentos obtidos, serão analisados os perfis sociodemográficos destes clientes, assim como a análise longitudinal dos perfis obtidos. Essas análises têm como objetivo a obtenção de insights de negócio e planejamento estratégico de ações de marketing.

A marca de roupas disponibilizou, além dos dados dos clientes, resultados de duas campanhas de marketing realizada com base nos segmentos obtidos para este

trabalho, possibilitando ainda a realização de uma análise de resultados de campanha de marketing.

3.2 Universo e amostra

A marca de roupas possui aproximadamente 120 mil clientes cadastrados com compras no período de janeiro de 2018 a dezembro de 2021. Os dados foram disponibilizados em sua totalidade, não havendo a necessidade de obtenção de amostragem.

Optou-se por trabalhar com a análise longitudinal de forma a englobar o período de 1 ano antes da pandemia até os dias atuais, possibilitando, desta forma, a avaliação da estabilidade do modelo num contexto de reflexos da recessão econômica provocada pela COVID-19 em itens não essenciais, como é o caso da indústria de vestuário.

3.3 Procedimentos de Coleta de Dados

Os dados transacionais foram disponibilizados por uma empresa de varejo de moda brasileira, com operação nacional, referente ao período de janeiro de 2019 a dezembro de 2021. Inicialmente, os dados disponibilizados estavam ao nível de produto, com a discriminação detalhada de cada item adquirido por cada cliente no período analisado.

Por questões de confidencialidade dos dados, optou-se por agregar os dados ao nível de cupons e, posteriormente, ao nível de cliente.

A base transacional foi reduzida para englobar as seguintes variáveis:

- Código do cliente: Variável de identificação
- Recência: número de dias desde a última compra do cliente
- Frequência: número de compras realizadas pelo cliente nas lojas da marca, físicas ou online, nos 12 meses anteriores à data da análise
- Valor: Total, em Reais, gasto pelo cliente nas lojas da marca nos 12 meses anteriores à data da análise
- Volume: Quantidade de itens adquiridos pelo cliente nas lojas da marca nos 12 meses anteriores à data da análise
- Indicador de cliente promocional: igual a 1 se o cliente comprou mais da metade dos itens adquiridos com pelo menos 50% de desconto

Para fins de análise de resultados, a empresa disponibilizou uma base tratada separadamente com dados demográficos dos clientes após a segmentação, assim como tabelas de resultado gerencial de duas campanhas de marketing realizadas com base na segmentação realizada. Ou seja, após o desenvolvimento do modelo, foi feito o cruzamento dos segmentos dos clientes com outros dados (gênero, canal preferencial do cliente, idade e região de moradia) gerando o relatório que foi disponibilizado.

Além disso, testou-se os resultados do modelo de segmentação através da seleção de uma amostra de clientes da loja e envio de comunicações de marketing à época da campanha da Black Friday de 2021. Os resultados de conversão neste cenário serão apresentados, mostrando a relevância dos segmentos obtidos para o planejamento estratégico da marca em questão.

3.4 Tratamento dos dados

Para a análise dos dados disponibilizados, foi utilizado o serviço Databricks, que é uma plataforma de computação na nuvem para gerenciamento e armazenamento de grandes volumes de dados e que permite a programação dos modelos nas linguagens Python.

Os dados transacionais da marca de roupas analisada foram obtidos a partir do sistema de vendas da loja. A identificação dos clientes se dá por meio de uma identificação interna, sem relação com dados externos e, portanto, a identidade dos clientes foi preservada. Os dados utilizados neste caso são disponibilizados na plataforma em d+1. Ou seja, 1 dia após a visita do cliente na loja a plataforma é capaz de atualizar a base de dados, recalculando o segmento do cliente diariamente. Porém, por questões de negócio, apesar dos dados serem disponibilizados diariamente, os segmentos são atualizados mensalmente, sempre no primeiro dia do mês. A atualização mensal dos dados é realizada de forma automática e os indicadores calculados são incluídos em painéis gerenciais de CRM e posteriormente utilizados na definição das estratégias de comunicação com os clientes da marca, assim como na apuração dos resultados das campanhas do mês anterior.

A partir dos dados transacionais ao nível de produto, foram derivadas as variáveis de interesse para o modelo de segmentação: Recência, Frequência e Valor Monetário.

Todos os indicadores consideram o período dos 12 meses anteriores ao período ao qual as tabelas de resultado se apresentam. O período de 12 meses foi escolhido devido a correspondência com o ciclo de negócio, que também é de 12 meses. Por exemplo, os dados de segmentação de janeiro de 2021 foram obtidos com base nas compras realizadas pelos clientes entre janeiro e dezembro de 2020. Para o mês de fevereiro de 2021, as variáveis foram derivadas a partir das vendas realizadas entre fevereiro de 2020 e janeiro de 2021. Ou seja, os indicadores sempre contemplam os dados transacionais dos últimos 12 meses em relação ao período correspondente da análise.

As variáveis do modelo RFM foram assim definidas:

- *Recência* = número de dias desde a última compra na marca
- *Frequência* =
número de compras na marca nos últimos 12 meses
- *Valor Monetário* =
total gasto pelo cliente nos últimos 12 meses (R\$)

Além das tradicionais variáveis do modelo RFM, foram incluídas duas novas variáveis no modelo: o volume de peças adquiridas e uma variável indicadora para identificar clientes que compram prioritariamente itens promocionais da marca.

A variável volume foi incluída após observar que a amplitude de preços dos produtos da marca poderia gerar inconsistências na definição dos perfis de consumo. Considerando que os preços dos itens da marca encontram-se no intervalo entre R\$29,00 e R\$6.499,99, com preço médio em torno de R\$600 reais, a inclusão do volume de peças adquiridas mostrou-se, empiricamente, significativa para a capacidade de diferenciação de perfis do modelo.

Para exemplificar este impacto, observamos o caso hipotético a seguir: a cliente A foi na loja 3 vezes no último ano e gastou no período R\$2000,00. A cliente B teve a mesma frequência e o mesmo gasto no período, e ambas visitaram a loja recentemente. Porém, a cliente A comprou apenas 3 itens e a cliente B adquiriu 8 itens.

A variável indicadora para a propensão de compras de itens em promoção foi definida de acordo com as regras de negócio da empresa, com base no percentual de itens com mais de 40% de desconto comprados por cada cliente. Ou seja:

$$promo = \begin{cases} 1, & \text{se os produtos tiveram pelo menos 40\% de desconto} \\ 0, & \text{caso contrário} \end{cases}$$

Retomando o caso das clientes A e B, e acrescentando a informação que a cliente A não adquiriu nenhum item com mais de 40% de desconto e que metade dos itens da cliente B tiveram mais de 40% de desconto, fica clara a diferença no perfil de consumo destas clientes para além das 3 variáveis principais RFM usadas no modelo. Ainda que tenha ocorrido a necessidade de acrescentar mais duas variáveis ao modelo RFM clássico, ambas as variáveis também são obtidas a partir de dados transacionais e, por isso não aumentam significativamente a complexidade da obtenção dos dados.

Clientes que completaram 1 ano sem realizar nenhuma compra na marca foram excluídos do modelo de segmentação e tratados como “churn”, ou seja, clientes que abandonaram a marca.

Além disso, para este modelo, por questões de negócio, as variáveis foram calculadas desconsiderando-se o canal da venda. Ou seja, todas as variáveis do modelo contemplam tanto as compras nas lojas físicas como no site da marca. Além disso, cupons exclusivamente de trocas são desconsiderados no cálculo dos indicadores, assim como foram excluídos da análise as compras realizadas por funcionários da rede de lojas.

Após o cálculo das cinco variáveis descritas para cada um dos clientes da marca (ilustradas na Tabela 2), foi aplicada a segmentação de clientes a partir do cálculo das distâncias entre as variáveis, denominado ‘k-means’.

Tabela 2: Variáveis de segmentação

VARIÁVEL	DESCRIÇÃO
RECÊNCIA	Número de dias desde a última compra na data da segmentação
FREQUÊNCIA	Número de tickets distintos nos 12 meses anteriores à segmentação
VALOR	Valor Total gasto pelo cliente nos 12 meses anteriores
VOLUME	Quantidade de itens adquiridos pelo cliente nos últimos 12 meses
CLIENTE PROMOCIONAL	Variável indicadora. Recebe valor 1 se o cliente for classificado como cliente promocional e 0 caso contrário

Fonte: elaboração própria.

Observou-se, empiricamente, que o cálculo do z-score de cada variável, exceto a indicação de compra de itens em promoção, gerou uma melhora significativa da capacidade de classificação do modelo.

O z-score é calculado subtraindo-se a média da variável e dividindo-se o resultado pelo desvio-padrão da variável original:

$$Z_x = \frac{X - E[X]}{\sigma_x}$$

Onde Z_x é o z-score calculado, X é o valor observado da variável para o cliente, $E[X]$ é a média dos valores da variável X e σ_x é o desvio-padrão da variável (Bussab e Morettin, 2010).

Desta forma, aplicou-se o modelo k-means às seguintes variáveis derivadas: Z-Recência (z-score da variável recência), Z-Frequência (z-score da variável Frequência), Z-Valor (z-score da variável Valor Monetário), Z-Volume (z-score da variável volume e Indicador de cliente promocional (variável dummie, do tipo 0-1).

Também foram disponibilizados pela marca dados de campanhas de marketing segmentadas com base nos segmentos obtidos, assim como a análise demográfica do perfil destes clientes.

3.5 K-Means e modelos de classificação

A análise de cluster é um conjunto de técnicas multivariadas cujo objetivo principal é agrupar objetos com base nas características que eles possuem, com base em características pré-determinadas dos objetos que desejamos classificar. Neste trabalho, o nosso objetivo é classificar os clientes da rede varejista com base em variáveis de consumo. Para tal, foi escolhido o modelo RFM como base de segmentação.

Os segmentos resultantes devem apresentar alta homogeneidade interna (dentro do cluster) e alta heterogeneidade externa (entre-cluster). Assim, espera-se que os clientes que estejam classificados num mesmo cluster tenham comportamentos parecidos entre si e que também sejam significativamente diferentes em relação aos indivíduos dos demais clusters.

A análise de cluster é a única técnica multivariada que não estima a variável empiricamente, mas usa as variáveis conforme especificado pelo pesquisador. O

foco deste tipo de análise está na comparação de objetos com base nas variáveis, não na estimativa de uma variável derivada. Essa distinção torna a definição da variável um passo crítico na análise de cluster.

Um dos principais papéis da análise de cluster é a redução de um conjunto de dados. Por exemplo, um pesquisador pode se deparar com um grande número de observações que não têm sentido, a menos que sejam classificadas em grupos gerenciáveis. A análise de cluster pode realizar esse procedimento de redução de dados objetivamente, transformando grandes conjuntos de dados em segmentos gerenciáveis de informação. Dessa forma, o pesquisador fornece uma descrição mais concisa e compreensível das observações, com perda mínima de informações.

O objetivo principal da análise de cluster é definir a estrutura dos dados colocando as observações mais semelhantes em grupos. O método escolhido para este trabalho é denominado “k-means”, um método computacionalmente intensivo de classificação baseado no cálculo de distâncias entre as observações.

Maiores detalhes sobre o passo a passo do método podem ser encontrados em Hair *et al.* (2014).

3.6 Limitações do estudo

O estudo limita-se a apresentar um estudo de caso de uma marca de roupas brasileira, propondo um modelo de segmentação com base exclusivamente em dados transacionais dos clientes desta marca. Assim, outras dimensões do comportamento do consumidor não foram analisadas, como, por exemplo, estilo de vestimenta preferida (modelo, cor, tamanho, tipo de tecido, etc.), ou mesmo a necessidade de uso (festas, trabalho, prática de esportes, etc..).

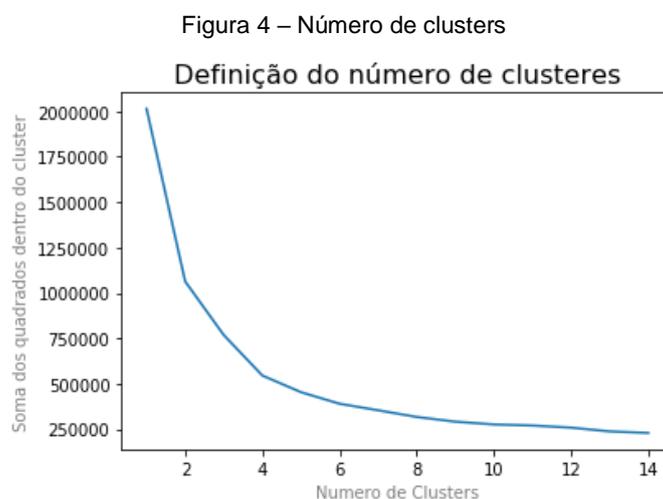
Além disso, a identificação dos clientes se dá por códigos internos e os controles internos de cadastro e identificação dos clientes não possui validação de cadastro único. Portanto, podem existir clientes com mais de um número de cadastro 3, sem maiores informações de vinculação dos cadastros, esses casos foram considerados como independentes na análise.

4. Análise e Interpretação dos Resultados

Este capítulo foi organizado em 3 partes: a primeira apresenta uma análise geral dos resultados do modelo de classificação usando RFM, a segunda etapa traz a análise longitudinal dos modelos e, finalmente, a última seção apresenta resultados obtidos na gestão de relacionamento com o cliente a partir do modelo apresentado

4.1 Resultados do modelo

Para a definição do número ideal de clusters, utilizou-se o método recursivo comumente conhecido como “método do cotovelo”, que indica graficamente o melhor ajuste em função do número de segmentos solicitado. A análise gráfica é apresentada na figura 4.



Fonte: elaboração própria.

Assim, o modelo de classificação aplicado foi capaz de identificar 5 grandes segmentos de clientes na marca, com score de *silhouette* igual a 0,9, indicando um ótimo ajuste do modelo. Deste modo, atribuíram-se as seguintes nomenclaturas dos clusters: No segmento A, temos os clientes de altíssimo valor. No segmento B, temos clientes Assíduos. No segmento C, os clientes potenciais. No segmento D, os clientes eventuais. Finalmente, no segmento E, os clientes de alto risco de abandonar a marca.

A tabela 3 traz os principais indicadores do modelo de classificação:

Tabela 3: Clusters Estimados e Principais Indicadores de Segmentação

Segmento	Recência Média	Frequência Média	Ticket Médio	Volume médio	Quantidade de clientes	Quantidade de clientes Promo	Representatividade das vendas
A Altíssimo Valor	72	11	1838,00	43	1207	155	13%
B Clientes Assíduos	81	10,1	448	26	2191	608	8%
C Clientes Potenciais	113	3,7	810	11	13384	3425	28%
D Clientes Eventuais	70	1,4	419	3	43475	3276	23%
E Clientes Churn	259	1,3	415	3	60042	11972	29%

Fonte: elaboração própria.

Conforme pode ser visto na tabela, o segmento A (altíssimo valor) apresenta os clientes de maior valor para a marca: Esses clientes visitam frequentemente as lojas da marca, possuem ticket médio acima de R\$1800,00 por visita e no período analisado adquiriram em média 42 peças da marca. Ou seja, a cada visita esses clientes compram, em média, quase 4 peças da coleção disponível naquele momento. Esses clientes representam aproximadamente 1% da base ativa de clientes da marca e apenas 13% deles são classificados como clientes promocionais. Como consequência, esta pequena parcela dos clientes da marca é responsável por 13% do faturamento da marca no período analisado.

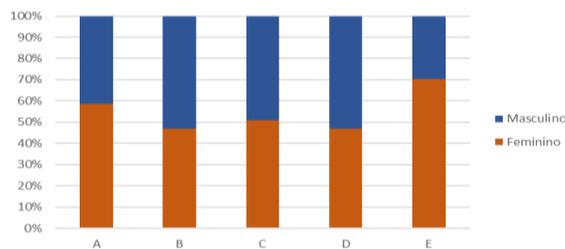
Já o segmento B (clientes assíduos) possui uma frequência de visitas pouco menor que os clientes de altíssimo valor (A), porém com um gasto médio bastante inferior e, conseqüentemente, o volume de peças adquiridas, assim como o tamanho da cesta de produtos é significativamente menor. Neste grupo, temos o dobro de base ativa em relação aos clientes de altíssimo valor (segmento A) mas, ainda assim, representando uma pequena fatia da base ativa de clientes, que é responsável por 8% do faturamento da marca.

Os clientes classificados no segmento C (clientes potenciais) são aqueles com maior potencial de crescimento na marca. A frequência média nas lojas deste grupo é relativamente baixa, porém o ticket médio destes clientes é bastante elevado, quase o dobro do gasto médio dos clientes assíduos (segmento B). Em termos de volume de clientes, este também é um segmento maior que os grupos anteriores e representa 11% da base ativa de clientes.

Os clientes dos segmentos D (clientes eventuais) e E (clientes de elevado risco de churn) possuem frequência baixa, volume de peças baixo, maior concentração de clientes promocionais e ticket médio também baixo. Diferem-se na dimensão da Recência: os clientes eventuais possuem recência baixa e os clientes em risco de churn, Recência alta. Ou seja, estes últimos estão a longo período sem visitar as lojas da marca e por isso apresentam alto risco de churn, ou seja, de abandonarem a marca em um período de tempo curto.

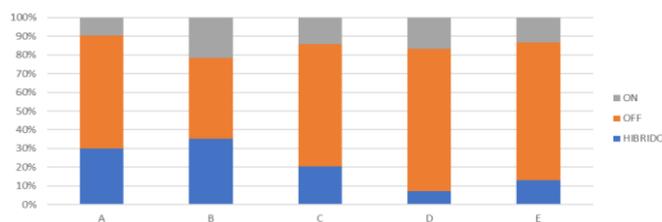
A análise demográfica dos segmentos também traz resultados interessantes: No quesito gênero (Figura 5), observa-se maior concentração de mulheres nos segmentos mais extremos. Em relação ao canal preferencial de compras (Figura 6), os clientes de altíssimo valor (segmento A) e os clientes Assíduos (segmento B) apresentam maior concentração de clientes nos canais online. Em relação à faixa de idade e região de moradia, não houve diferenças significativas entre os segmentos.

Figura 5 – Gênero por segmento calculado



Fonte: elaboração própria.

Figura 6 – Canal preferencial por segmento



Fonte: elaboração própria.

4.2 Análise longitudinal

A análise longitudinal do modelo apresentado foi desenvolvida da seguinte forma: utilizando-se o conceito de data de referência, replicou-se o modelo considerando o recorte temporal em referência a períodos passados. Todos os modelos foram desenvolvidos ao longo do ano de 2021, mesmo aqueles que se referem aos anos de 2019 e 2020.

Durante o desenvolvimento do modelo, e com base no comportamento sazonal do segmento de mercado estudado, estabeleceu-se como sendo o período ideal de análise os dados transacionais dos últimos 12 meses em relação ao período de segmentação.

Desta forma, a classificação de clientes referente ao mês de janeiro de 2019 leva em consideração os dados transacionais dos 12 meses imediatamente anteriores à data de referência de 1º de janeiro de 2019, ou seja, os dados do período compreendido entre 1º de janeiro de 2018 e 31 de dezembro de 2018. A tabela completa pode ser observada na Tabela .

Tabela 4: Períodos de referência

ANO_MÊS	DATA DE REFERÊNCIA	DATA INÍCIO	DATA FINAL
201901	01/01/2019	01/01/2018	31/12/2018
201902	01/02/2019	01/02/2018	31/01/2019
201903	01/03/2019	01/03/2018	28/02/2019
201904	01/04/2019	01/04/2018	31/03/2019
201905	01/05/2019	01/05/2018	30/04/2019
201906	01/06/2019	01/06/2018	31/05/2019
201907	01/07/2019	01/07/2018	30/06/2019
201908	01/08/2019	01/08/2018	31/07/2019
201909	01/09/2019	01/09/2018	31/08/2019
201910	01/10/2019	01/10/2018	30/09/2019
201911	01/11/2019	01/11/2018	31/10/2019
201912	01/12/2019	01/12/2018	30/11/2019
202001	01/01/2020	01/01/2019	31/12/2019
202002	01/02/2020	01/02/2019	31/01/2020
202003	01/03/2020	01/03/2019	29/02/2020
202004	01/04/2020	01/04/2019	31/03/2020
202005	01/05/2020	01/05/2019	30/04/2020
202006	01/06/2020	01/06/2019	31/05/2020
202007	01/07/2020	01/07/2019	30/06/2020
202008	01/08/2020	01/08/2019	31/07/2020
202009	01/09/2020	01/09/2019	31/08/2020

202010	01/10/2020	01/10/2019	30/09/2020
202011	01/11/2020	01/11/2019	31/10/2020
202012	01/12/2020	01/12/2019	30/11/2020
202101	01/01/2021	01/01/2020	31/12/2020
202102	01/02/2021	01/02/2020	31/01/2021
202103	01/03/2021	01/03/2020	28/02/2021
202104	01/04/2021	01/04/2020	31/03/2021
202105	01/05/2021	01/05/2020	30/04/2021
202106	01/06/2021	01/06/2020	31/05/2021
202107	01/07/2021	01/07/2020	30/06/2021
202108	01/08/2021	01/08/2020	31/07/2021
202109	01/09/2021	01/09/2020	31/08/2021
202110	01/10/2021	01/10/2020	30/09/2021
202111	01/11/2021	01/11/2020	31/10/2021
202112	01/12/2021	01/12/2020	30/11/2021

Fonte: elaboração própria.

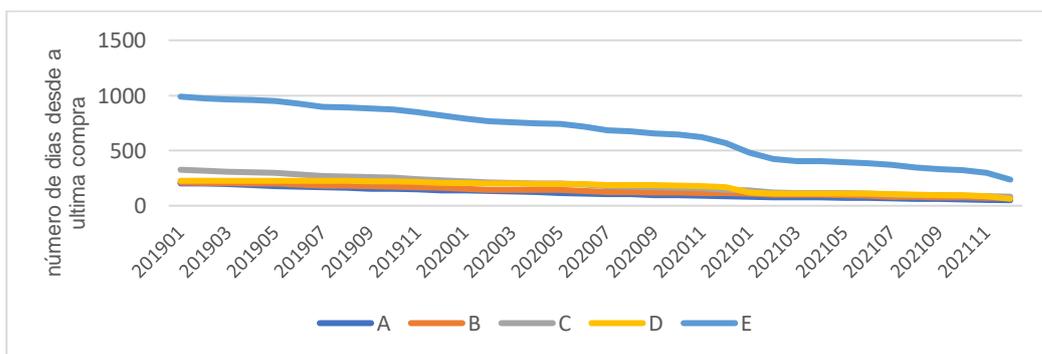
A partir dos dados transacionais referentes ao período especificado, replicou-se, mês a mês, o modelo de classificação. A seguir serão apresentadas a análise longitudinal do comportamento dos segmentos.

Conforme será observado nas análises a seguir, o comportamento dos clusters é influenciado por variações no posicionamento estratégico da marca, assim como por questões externas de grande impacto, como a pandemia de COVID-19.

4.2.1 Recência

Ao contrário das demais variáveis, a recência é inversamente proporcional à qualidade do relacionamento do cliente com a marca (Figura 7). Ou seja, quanto menor for o número de dias desde a última compra daquele cliente na marca, maior será o impacto positivo desse indicador no modelo de segmentação.

Figura 7 – Análise Longitudinal da Recência

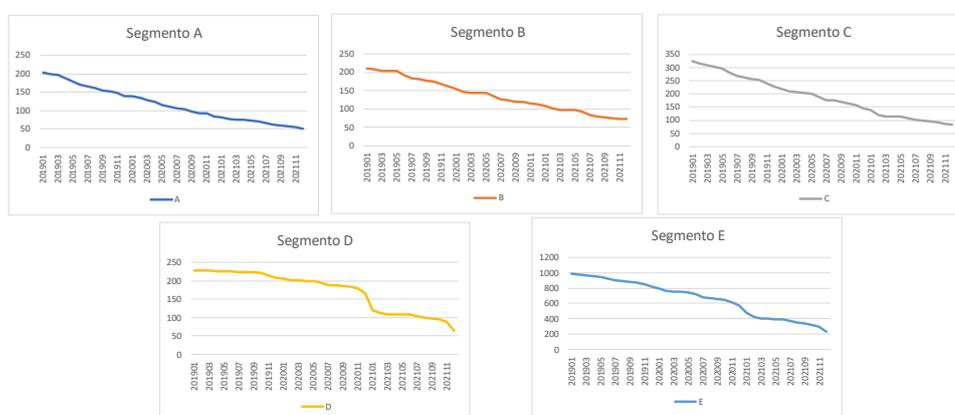


Fonte: elaboração própria.

Desse modo, o segmento E (risco de churn) apresenta os piores indicadores, porém ao longo do tempo e, principalmente após a implementação do modelo com ações de CRM baseadas nesta segmentação, este indicador decaiu significativamente. Esse comportamento do segmento E (risco de churn) permite inferir, em termos de negócio, que o modelo de classificação foi capaz de identificar os clientes com maior potencial de abandonar a marca em análise.

Analisando separadamente cada segmento (Figura 8), é possível observar um suave aumento no indicador de recência no período de 202003, que coincide com o início da pandemia de COVID-19 no mundo e consequente fechamento dos canais de venda físicos da marca. No caso dos clientes de altíssimo valor e assíduos (segmentos A e B), observou-se também que o aumento do indicador de recência foi pontual e que, após um pequeno período de ajuste na comunicação e reabertura parcial das lojas, estes clientes retornaram ao padrão anterior em relação ao indicador de recência. Os demais segmentos apresentaram comportamento semelhante, porém o período para o retorno ao nível pré-pandemia foi maior.

Figura 8 – Análise longitudinal por segmento – Recência



Fonte: elaboração própria.

Com o início efetivo da aplicação do modelo de segmentação na definição das estratégias de CRM da marca, observa-se ainda uma suave redução do indicador nos segmentos C (clientes potenciais) e D (clientes eventuais).

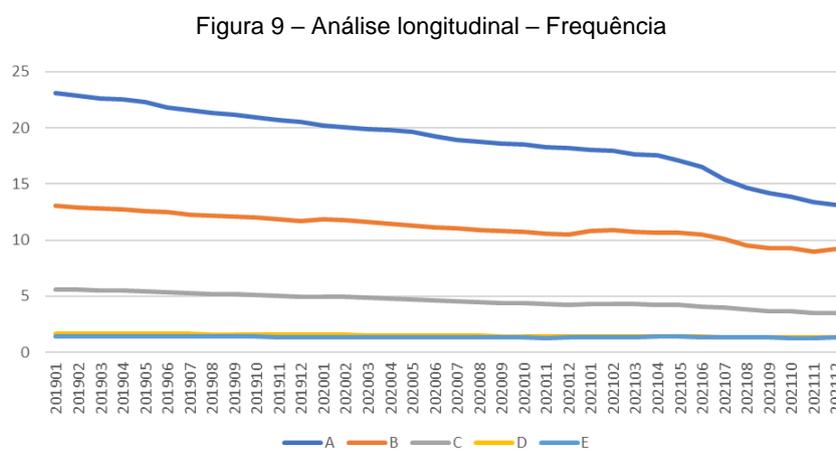
Historicamente, o mês de março para esta marca em questão é bastante importante pois é quando a marca realiza grandes queimas de estoque com descontos de até 80% no valor das peças. No ano de 2021, após a implementação da segmentação apresentada e sua utilização na gestão estratégica de marketing e

de comunicação com os clientes com base nos segmentos apresentados observa-se uma grande redução na recência dos clientes destes segmentos. Conforme será analisado mais adiante neste trabalho, este achado corrobora os indicadores de volumetria de clientes classificados como promocionais nos segmentos D (eventuais) e E (risco de churn).

4.2.2 Frequência

Conforme explicitado anteriormente, o indicador de frequência considera o número de compras que cada cliente realizou no período de um ano anterior à data da segmentação. Desta forma, os valores aqui apresentados (Figura 9) representam o número médio de visitas dos clientes de determinado segmento no período de um ano nas lojas da marca.

Assim, este indicador é diretamente proporcional à qualidade do relacionamento do cliente com a marca. Ou seja, quanto maior a frequência destes clientes na loja, mais próximo se torna o seu relacionamento com a marca.



Fonte: elaboração própria.

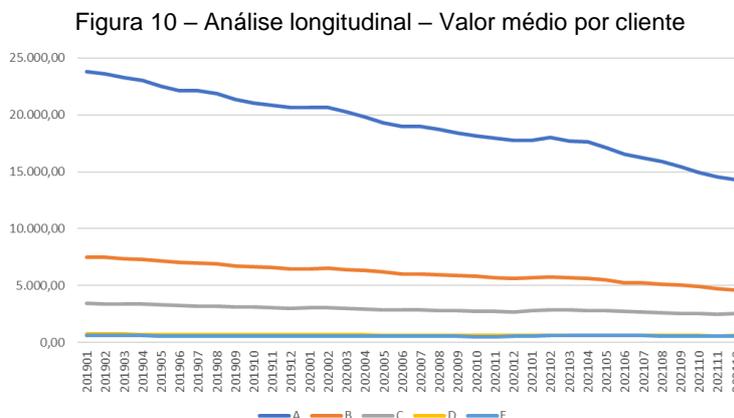
A análise desta dimensão revela uma tendência de queda no número de visitas média dos clientes no período de análise especialmente nos segmentos de altíssimo valor (A) e assíduos (B). No caso dos clientes assíduos (B), observa-se uma melhora do relacionamento com o cliente no período de 202110.

Ao observar apenas a dimensão frequência dos clientes eventuais(D) e em risco de churn(E) não apresentam diferenças de comportamento entre si. Em relação à frequência, os clientes potenciais (grupo C) mostram-se mais estável ao longo do período analisado, porém com leve tendência de queda nos meses finais da análise.

4.2.3 Valor

A observação do gasto médio do cliente por período de análise é, sozinha, um dos principais indicadores de relacionamento de um cliente com determinada marca ou produto. Neste modelo de classificação multivariado, ela é apenas mais uma dimensão que agrega valor e, sem dúvidas é a dimensão de maior interesse estratégico para a melhoria dos resultados de uma empresa. Na empresa em questão, não é diferente e o comportamento dos clientes, principalmente no segmento A, reflete o momento estratégico da empresa, que perdeu bastante espaço de mercado no período analisado devido à entrada de novos concorrentes no mercado e, por isso, se viu obrigada a rever sua estratégia de precificação de produtos, assim como a variedade de peças produzidas, reduzindo significativamente o preço médio de suas peças com o objetivo de manter-se competitiva no mercado.

Conforme pode ser observado na figura 10, esta mudança de posicionamento estratégico refletiu fortemente nos clientes de altíssimo valor (cluster A) que reduziu o seu gasto médio na marca.

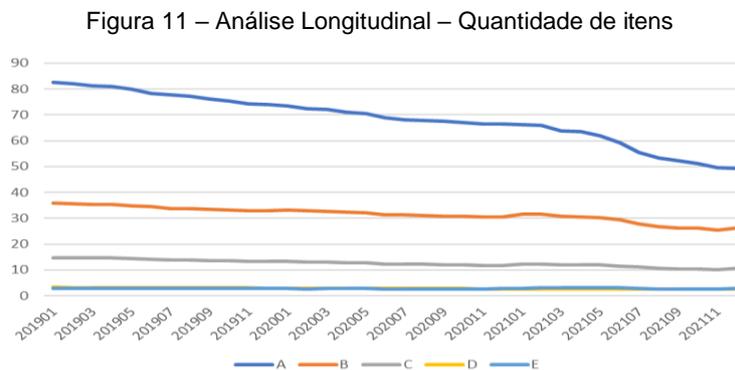


Fonte: elaboração própria.

4.2.4 Volume

Conjuntamente ao valor gasto, incluiu-se na análise a quantidade de peças adquiridas pelos clientes no período de análise (Figura 11). Conforme dito anteriormente, esta variável foi incluída no modelo devido à grande variância no preço dos itens da loja. Conforme discutido no tópico anterior em relação às alterações nas estratégias de precificação e variedade de itens nas coleções

comercializadas, a análise da quantidade de itens reflete esta alteração, especialmente para os clientes de altíssimo valor (grupo A).



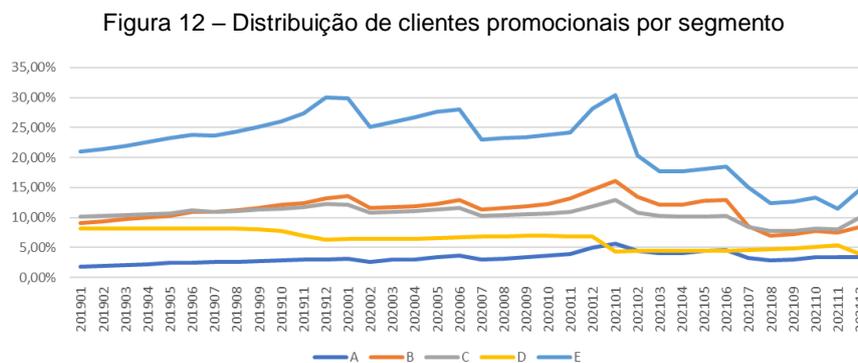
Fonte: elaboração própria.

4.2.5 Clientes Promocionais

A definição de cliente promocional, apresentada anteriormente, nos leva a compreender o indicador de percentual de clientes promocionais num determinado segmento como um fator de propensão à impactos de ações promocionais.

Ou seja, não necessariamente o segmento que apresenta maior concentração de clientes promocionais indica um relacionamento ruim com a marca. Tudo vai depender do objetivo da ação de CRM estudada. Ou seja, se existe uma promoção a ser lançada, pode ser mais interessante impactar os clientes classificados no grupo E em relação aos clientes do grupo A ou B, por exemplo. Da mesma, forma, deve-se evitar o contato com os clientes classificados no grupo A para ações de descontos agressivos.

Conforme visto anteriormente, os clientes que consomem menos da marca (grupos D e E) também concentram a maior parte dos clientes promocionais da marca (Figura 12).



Fonte: elaboração própria.

4.3 Aplicações práticas

Este modelo foi construído com o objetivo de direcionar a definição de estratégias de CRM de uma empresa de moda. Ao longo da análise longitudinal, algumas destas aplicações foram citadas como exemplo. Agora, vamos analisar conjuntamente o modelo de classificação a partir de todas as dimensões consideradas no desenvolvimento do modelo.

A análise longitudinal do modelo de classificação garante a estabilidade do modelo de classificação de clientes e, portanto, fundamenta a automação do modelo pela empresa e a manutenção, ao longo do tempo, da validade da classificação realizada.

Assim, a área de CRM torna-se capaz de fundamentar a base da sua estratégia de segmentação de clientes para as mais diversas ações de marketing direto com cada cliente da marca.

Com base nos resultados do modelo de classificação apresentado, a empresa tornou-se capaz de definir grandes estratégias de comunicação e relacionamento com seus clientes com base nos dados estritamente transacionais dos seus clientes.

Além disso, considerando o cenário imposto pela pandemia de covid-19, as variações no comportamento de consumo destes clientes mostraram-se fundamentais para guiar a adoção de novas estratégias de marketing no relacionamento com os clientes em questão.

As aplicações práticas deste modelo na definição de estratégias de CRM que foram enumeradas neste trabalho são apenas algumas poucas possibilidades de uso.

Este modelo de classificação pode, ainda, ser cruzado com outros modelos de comportamento de consumo para gerar insights ainda mais poderosos para o negócio. Conforme já mencionado neste trabalho, o modelo de classificação utilizou apenas variáveis transacionais dos clientes. Outros modelos, que visem outras dimensões de consumo podem ser construídos e estes resultados cruzados com este modelo, como por exemplo, o desenvolvimento de modelos de classificação de estilo de vestuário, necessidades do cliente, momento de vida e outras dimensões que a empresa julgar importantes para a definição estratégica de comunicação personalizada com seus clientes.

5. Conclusão

Este trabalho discute o impacto da pandemia de covid-19 em uma marca de roupas do varejo brasileiro. Para tanto, apresenta um estudo de caso de uma marca de varejo de moda brasileira, analisando a aplicação do modelo RFM através do uso de modelos de aprendizado de máquina e a estabilidade deste modelo de segmentação de clientes como ferramenta de CRM.

Este estudo destaca-se, além da utilização de dados reais de uma marca brasileira, pela amplitude do período de dados disponibilizado: foram utilizados dados transacionais, coletados ao longo de 4 anos – englobando tanto o período pré-pandemia quanto o período pandêmico.

Esse trabalho pretendeu investigar a estabilidade do comportamento de consumo dos clientes da marca, com objetivo de segmentar clientes para fins de CRM. O foco do estudo recaiu sobre a identificação desses segmentos e a avaliação do comportamento dos segmentos identificados ao longo do período analisado. Essa questão se mostra estrategicamente importante na medida que fundamenta a definição de estratégias de comunicação não apenas no curto prazo, mas também possibilitando à empresa a utilização do modelo proposto no planejamento de CRM a médio prazo, dada a estabilidade do comportamento dos perfis de cliente no período analisado.

Para atingir os objetivos pretendidos, este trabalho analisou o impacto direto do desenvolvimento da análise de dados proporcionado pelo desenvolvimento de novas tecnologias de processamento da informação, relacionando-o com o desenvolvimento das aplicações de dados em marketing.

Os dados foram disponibilizados pela empresa em questão e é composto pela totalidade dos registros transacionais da marca entre janeiro de 2018 e dezembro de 2021. Estes dados foram então reduzidos ao nível de cliente, trazendo as variáveis de recência, frequência, valor e volume de itens adquiridos mês a mês, ao longo de todo o período analisado e então os clientes foram classificados em cada um dos 5 segmentos definidos na análise.

Dentre os principais resultados observados, a capacidade de discriminação dos clientes em razão dos seus perfis de consumo fica evidente, assim como o comportamento estável ao longo do tempo dos perfis identificados possibilita análises comparativas entre os resultados de CRM da empresa.

O segmento A contém os maiores clientes da marca, representando aproximadamente 1% da base ativa e 13% da receita anual da empresa. São denominados “clientes super especiais” e são impactados por ações exclusivas e, muitas vezes, entregues diretamente no endereço do cliente. Nas lojas, esses clientes recebem atendimento diferenciado e personalizado, de acordo com o perfil de cada um desses clientes, tendo como principal estratégia de CRM o encantamento desses clientes com a marca e seus produtos.

O Segmento B também apresenta os clientes considerados assíduos para a marca. Em termos estratégicos, foram denominados “clientes especiais”. Representam 2% da base ativa e 8% do faturamento da marca. Apresentam consumo bastante elevado, mas em média 50% menor em relação ao grupo A. Para este grupo, as principais estratégias de CRM passam por gerar encantamento com a marca e seus produtos e ainda um foco maior na potencialização do consumo destes clientes. Para tal, podem ser desenvolvidas ações de exibição de produtos, convites para os grandes desfiles da marca, entre outros.

O segmento C é o segmento de maior potencial de crescimento da marca. É responsável por 11% da base ativa e 28% do faturamento da marca, apresenta clientes com elevado gasto em dinheiro, porém com frequência relativamente baixa. Este cliente é o que possui o maior potencial de crescimento da marca e, como estratégia de CRM, este cliente deve ser estimulado a engajar nas ações da marca, aumentando assim a sua frequência e maximizando o número de visitas destes clientes às lojas da marca.

Por exemplo, em uma campanha hipotética de distribuição de brindes, os clientes classificados nos grupos A e B podem ser considerados elegíveis a receber o brinde em suas residências. Para os clientes do grupo C, a estratégia de convidá-los a comparecer fisicamente a uma loja para retirar o brinde pode ser mais interessante para o objetivo de maximizar a ida desses clientes à loja.

Os segmentos D e E são formados pelos consumidores eventuais das marcas, que realizam em média 1 a 2 compras por ano na marca. Esses grupos diferenciam-se, principalmente, pela recência. Ou seja, o modelo de classificação de clientes

apresentado também foi capaz de identificar os clientes prestes a serem perdidos pela marca. Dentre as estratégias possíveis para estes grupos, pode-se optar por oferecer descontos, com o objetivo de atrair esses clientes para a marca, potencializando esses contatos e, eventualmente, convertendo-os em aumento da frequência e da quantidade de itens que esses clientes compram na marca.

Dentre as principais questões abordadas, destaca-se a permanência do modelo RFM como uma importante ferramenta de classificação de clientes no varejo, salientando-se o impacto do uso de técnicas computacionalmente intensivas na sua aplicação prática e o quanto essas técnicas permitem o desenvolvimento de estratégias de CRM efetivas. Também se discutiu a perspectiva longitudinal do comportamento dos grupos segmentados. Por outro lado, a investigação apontou que questões de posicionamento estratégico global da companhia podem se tornar evidentes neste tipo de análise e não devem ser deixadas de lado ao analisarmos longitudinalmente esses resultados.

5.1 Implicações Teóricas e Práticas

Este trabalho reafirma a importância da fundamentação teórica de variáveis em modelos de segmentação. Os dados sobre o comportamento dos clientes serão cada vez mais abundantes. Cabe ao pesquisador a responsabilidade da escolha das variáveis que serão efetivamente utilizadas nos modelos de classificação.

Nem sempre maior quantidade de variáveis vai, necessariamente, levar a uma melhor segmentação. O modelo apresentado, de comportamento de consumo, utilizou apenas as variáveis do modelo RFM, acrescidas de apenas mais duas variáveis, também definidas com base nos dados transacionais e, ainda assim, foi capaz de obter um modelo de classificação bastante efetivo.

Em termos teóricos, este trabalho reforça a alta capacidade discriminante do modelo RFM quando aplicado através de métodos computacionalmente intensivos, além de demonstrar grande capacidade de identificação de clientes de alto valor e churn para as empresas.

Por ser um modelo de fácil construção e processado de maneira relativamente rápida, apesar de intensivo, a metodologia proposta neste trabalho pode ser replicada nos mais diversos segmentos de varejo.

5.2 Sugestões e recomendações para novas pesquisas

Como desdobramentos futuros, essa linha de pesquisa pode ser estendida através da inclusão de novas variáveis no modelo de segmentação de consumo, como a inclusão da dimensão tempo de relacionamento e periodicidade das visitas, conforme proposto por Peker, Kocyigit e Eren (2017). Também pode ser incluída no modelo a dimensão margem de lucro, que é proporcional ao valor do CLV (Customer Lifetime Value) (Gupta, Aggarwal e Rani, 2016) como dimensão de segmentação.

A segmentação de clientes pelo perfil de consumo na marca pode, ainda, ser considerada um ponto de partida para o desenvolvimento de outras segmentações para as mais diversas dimensões e perfis dos clientes como, por exemplo, perfil de uso das roupas da marca, ou mesmo estilo de moda adotado pelos indivíduos. Analisados conjuntamente, essas dimensões podem permitir o desenvolvimento de personas de marketing.

6. Referências Bibliográficas

AHN, Y.; KIM, D.; LEE, D.-J. Customer attrition analysis in the securities industry: a large-scale field study in Korea. **International Journal of Bank Marketing**, v. 38, n. 3, p. 561–577, 2020.

ALDERSON, W.; COX, R. Towards a Theory of Marketing. **Journal of Marketing**, v. 13, n. 2, p. 137–152, 6 out. 1948.

ANITHA, P.; PATIL, M. M. RFM model for customer purchase behavior using K-Means algorithm. **Journal of King Saud University - Computer and Information Sciences**, 2020.

BAKER TILLY BR. **Dados: o novo petróleo do mundo e combustível para o futuro**. Disponível em: <<https://bakertillybr.com.br/dados-novo-petroleo/>>. Acesso em: 14 nov. 2021.

BASU, M.; SWAMINATHAN, V. Consuming in a crisis: pandemic consumption across consumer segments and implications for brands. **Journal of Product and Brand Management**, n. September, 2021.

BEHERA, R. K. *et al.* Personalized digital marketing recommender engine. **Journal of Retailing and Consumer Services**, v. 53, n. September 2018, p. 101799, 2020.

BLATTBERG, R. C.; KIM, B.-D.; NESLIN, S. A. RFM Analysis. *In: Database marketing*. [s.l.: s.n.]. p. 323–337.

BULT, J. R.; WANSBEEK, T. Optimal Selection for Direct Mail. **Marketing Science**, v. 14, n. 4, p. 378–394, 1995.

BUSSAB, W. DE O.; MORETTIN, P. A. **Estatística Básica**. 6. ed. São Paulo: Saraiva, 2010.

CARRASCO, R. A. *et al.* A Fuzzy Linguistic RFM Model Applied to Campaign Management. **International Journal of Interactive Multimedia and Artificial Intelligence**, v. 5, n. 4, p. 21, 2019.

CHEN, H.; CHIANG, R. H. L.; STOREY, V. C. Business intelligence and analytics: From big data to big impact. **MIS Quarterly: Management Information Systems**, v. 36, n. 4, p. 1165–1188, 2012.

CHIANG, W. Y. W.-Y. Discovering customer value for marketing systems: an empirical case study. **International Journal of Production Research**, v. 55, n. 17, p. 5157–5167, 2017.

CHORIANOPOULOS, A. Value segmentation and cross-selling in retail. *In:*

Effective CRM Using Predictive Analytics. [s.l: s.n.]. p. 267–331.

____. **An overview of data mining: The applications, the methodology, the algorithms, and the data.** [s.l: s.n.].

CIL, I. Consumption universes based supermarket layout through association rule mining and multidimensional scaling. **Expert Systems with Applications**, v. 39, n. 10, p. 8611–8625, ago. 2012.

COLOMBO, R.; JIANG, W. **A STOCHASTIC RFM MODEL** *Journal of Interactive Marketing*. [s.l: s.n.].

DANGI, H.; MALIK, A. **Personalisation in marketing: an exploratory study** *Int. J. Internet Marketing and Advertising*. [s.l: s.n.].

DAS, S.; MISHRA, M. Business governance and society: Analyzing shifts, conflicts, and challenges. **Business Governance and Society: Analyzing Shifts, Conflicts, and Challenges**, p. 43–54, 2018.

DEGHANIZADEH, M. R.; FATHIAN, M.; GHOLAMIAN, M. R. LDcFR: A new model to determine value of airline passengers. **Tourism and Hospitality Research**, v. 18, n. 3, p. 357–366, 1 jul. 2018.

FADER, P. S.; HARDIE, B. G. S.; LEE, K. L. **RFM and CLV: Using iso-value curves for customer base analysis** *Journal of Marketing Research*, nov. 2005.

FONSECA, J. R. S.; CARDOSO, M. G. M. S. Supermarket customers segments stability. **Journal of Targeting, Measurement and Analysis for Marketing**, v. 15, n. 4, p. 210–221, set. 2007.

GARCÍA-MURILLO, M.; ANNABI, H. Customer knowledge management. **Journal of the Operational Research Society**, v. 53, n. 8, p. 875–884, 2002.

GARSDIE, J. Customer Satisfaction and Quality. *In: Plan to Win*. [s.l: s.n.]. v. 68p. 199–224.

GEORGE, J. P.; GAIKWAD, S. M. Time series forecasting for understanding potential buyer behavior with ecommerce. **Indian Journal of Ecology**, v. 47, p. 141–143, 2020.

GOELDNER, C. R. Automation in Marketing. **Journal of Marketing**, v. 26, n. 1, p. 53, 1962.

GUPTA, G.; AGGARWAL, H.; RANI, R. Segmentation of retail customers based on cluster analysis in building successful CRM. **Int. J. Business Information Systems**, v. 23, n. 2, p. 212–228, 2016.

GUPTA, S. *et al.* Valuing customers. **BT Technology Journal**, v. 23, n. 3, p. 24–29, 16 fev. 2005.

_____. **Modeling customer lifetime value** *Journal of Service Research*, 29 nov. 2006. Disponível em: <<http://journals.sagepub.com/doi/10.1177/1094670506293810>>. Acesso em: 7 jan. 2021

GUPTA, S.; LEHMANN, D. R. Customers as assets. **Journal of Interactive Marketing**, v. 17, n. 1, p. 9–24, 2003.

HAGHIGHATNIA, S.; ABDOLVAND, N.; RAJAE HARANDI, S. Evaluating discounts as a dimension of customer behavior analysis. **Journal of Marketing Communications**, v. 24, n. 4, p. 321–336, 2018.

HAIR, J. F. *et al.* **Multivariate Data Analysis**. 7. ed. [s.l.] Pearson Education, 2014.

HALEY, R. I. **Benefit Segmentation: A Decision-oriented Research Tool** *Journal of Marketing*. [s.l: s.n.].

_____. BENEFIT SEGMENTATION — 20 YEARS LATER. **THE JOURNAL OF CONSUMER MARKETING**, p. 5–13, 1988.

HEAD, G. W. What Does Automation Mean to the Marketing Man? **Journal of Marketing**, v. 24, n. 4, p. 35, 1960.

HSU, P. Y. P.-Y.; HUANG, C.-W. C. W. IECT: A methodology for identifying critical products using purchase transactions. **Applied Soft Computing Journal**, v. 94, p. 106420, 2020.

HUGHES, A. M. Boosting response with RFM. **Marketing Tools**, v. 3, n. 3, p. 4, maio 1996.

KABASAKAL, I. Customer Segmentation Based On Recency Frequency Monetary Model : A Case Study in E-Retailing 1 Güncellik Sıklık Parasallık Modeline Dayalı Müşteri Bölümlendirme : E- Perakende Sektöründe Bir Uygulama. p. 47–56, 2020.

KHODAKARAMI, F.; CHAN, Y. E. Exploring the role of customer relationship management (CRM) systems in customer knowledge creation. **Information and Management**, v. 51, n. 1, p. 27–42, 2014.

KIT, T. C.; AZMI, N. F. M. Customer Profiling for Malaysia Online Retail Industry using K-Means Clustering and RM Model. **International Journal of Advanced Computer Science and Applications**, v. 12, n. 1, p. 106–113, 2021.

KLINGENBERG, C. O.; BORGES, M. A. V.; ANTUNES, J. A. V. Industry 4.0 as a data-driven paradigm: a systematic literature review on technologies. **Journal of Manufacturing Technology Management**, v. 32, n. 3, p. 570–592, 2021.

KRAFFT, M.; SAJTOS, L.; HAENLEIN, M. Challenges and Opportunities for Marketing Scholars in Times of the Fourth Industrial Revolution. **Journal of**

Interactive Marketing, v. 51, p. 1–8, 2020.

KUMAR, V. *et al.* Strategic Customer Relationship Management Today. **Customer Relationship Management**, v. 5, n. 1998, p. 3–20, 2012.

____. A theory of customer valuation: Concepts, metrics, strategy, and implementation. **Journal of Marketing**, v. 82, n. 1, p. 1–19, 2018.

____. Understanding the role of artificial intelligence in personalized engagement marketing. **California Management Review**, v. 61, n. 4, p. 135–155, 2019.

KUMAR, V.; RAMACHANDRAN, D.; KUMAR, B. Influence of new-age technologies on marketing: A research agenda. **Journal of Business Research**, v. 125, n. October 2019, p. 864–877, 2021.

KUNZ, W. *et al.* Customer engagement in a Big Data world. **Journal of Services Marketing**, v. 31, n. 2, p. 161–171, 2017.

LI, J. *et al.* **A Rule-Based Knowledge Discovery Engine Embedded Semantic Graph Knowledge Repository for Retail Business** 4th International Conference on Advanced Cloud and Big Data, CBD 2016. **Anais...** Faculty of Computer and Information Sciences, Hosei University, Tokyo, Japan: Institute of Electrical and Electronics Engineers Inc., 2017 Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85013155815&doi=10.1109%2FCBD.2016.024&partnerID=40&md5=fdb162b803f02f5d224ce9dabe572b06>>

LIN, R. *et al.* Applied big data analysis to build customer product recommendation model. **Sustainability (Switzerland)**, v. 13, n. 9, 2021.

MADSEN, D. Ø. The emergence and rise of industry 4.0 viewed through the lens of management fashion theory. **Administrative Sciences**, v. 9, n. 3, 2019.

MARTÍNEZ, R. G. *et al.* A comparison between Fuzzy Linguistic RFM Model and traditional RFM model applied to Campaign Management. Case study of retail business. **Procedia Computer Science**, v. 162, n. Itqm, p. 281–289, 2019.

MARTÍNEZ, R G *et al.* New fuzzy RFMD model. A method to identify new customer profiles due to an increase in online activity. Case study of a retail business . **Estudios de Economía Aplicada**, v. 39, n. 8, 2021.

MARTÍNEZ, ROCÍO G. *et al.* An rfm model customizable to product catalogues and marketing criteria using fuzzy linguistic models: Case study of a retail business. **Mathematics**, v. 9, n. 16, 2021.

MCEVILY, S. K.; CHAKRAVARTHY, B. The persistence of knowledge-based advantage: An empirical test for product performance and technological knowledge. **Strategic Management Journal**, v. 23, n. 4, p.

285–305, 2002.

MUSTAK, M. *et al.* Artificial intelligence in marketing: Topic modeling, scientometric analysis, and research agenda. **Journal of Business Research**, v. 124, n. January 2020, p. 389–404, 2021.

NAIRN, A.; BOTTOMLEY, P. Something approaching science? Cluster analysis procedures in the CRM era. **International Journal of Market Research**, v. 45, n. 2, p. 241- 261+269, 2003.

PAYNE, A.; FROW, P. A strategic framework for customer relationship management. **Journal of Marketing**, v. 69, n. 4, p. 167–176, 2005.

PEKER, S.; KOCYIGIT, A.; EREN, P. E. E. LRFMP model for customer segmentation in the grocery retail industry: a case study. **Marketing Intelligence and Planning**, v. 35, n. 4, p. 544–559, 2017.

PEKER, S.; KOCYIGIT, A.; ERHAN EREN, P. **A methodology for product segmentation using sale transactions** (V. B. et al., Eds.)2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2018 - Proceedings. **Anais...**Department of Software Engineering, Atilim University, Ankara, Turkey: Institute of Electrical and Electronics Engineers Inc., 2018Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85050189891&doi=10.23919%2FMIPRO.2018.8400226&partnerID=40&md5=5f7f1d112cc2c879fe17acdb2a8968f2>>

PIERCY, N. F.; MORGAN, N. A. Strategic and operational market segmentation: A managerial analysis. **Journal of Strategic Marketing**, v. 1, n. 2, p. 123–140, 1993.

PISKUNOVA, O.; KLOCHKO, R. **Classification of e-commerce customers based on Data Science techniques** (G. L. et al., Eds.)Workshop on the 12th International Scientific Practical Conference Modern Problems of Social and Economic Systems Modelling, MPSESM-W 2020. **Anais...**Kyiv National Economic University named after Vadym Hetman, Kyiv, Ukraine: CEUR-WS, 2020Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85095679985&partnerID=40&md5=7be8a9cc901b304ab290d2e25f4c2e86>>

PRATHIMA, J. *et al.* Finding the Most Profitable customer using CLV. v. 9, n. 7, p. 47–50, 2021.

PUNJ, G.; STEWART, D. W. Cluster Analysis in Marketing Research: Review and Suggestions for Application. **Journal of Marketing Research**, v. 20, n. 2, p. 134–148, 1983.

RAHIM, M. A. *et al.* RFM-based repurchase behavior for customer classification and segmentation. **Journal of Retailing and Consumer Services**, v. 61, 2021.

RANGASWAMY, A. *et al.* The Role of Marketing in Digital Business Platforms. **Journal of Interactive Marketing**, v. 51, p. 72–90, 2020.

RUST, R. T. *et al.* Return on Marketing: Using Customer Equity to Focus Marketing Strategy. **Journal of Marketing**, v. 68, n. 1, p. 109–127, 10 jan. 2004.

_____. The future of marketing. **International Journal of Research in Marketing**, v. 37, n. 1, p. 15–26, 2020.

RUST, R. T.; HUANG, M. H. The service revolution and the transformation of marketing science. **Marketing Science**, v. 33, n. 2, p. 206–221, 2014.

RUST, R. T.; VERHOEF, P. C. Optimizing the marketing interventions mix in intermediate-term CRM. **Marketing Science**, v. 24, n. 3, p. 477–489, jun. 2005.

SEETHA, A.; THAKAR, U. Finding Customer Loyalty Based on Weighted RFMD Clustering Model. **SSRN Electronic Journal**, p. 1–10, 2020.

SHAH, D.; MURTHI, B. P. S. Marketing in a data-driven digital world: Implications for the role and scope of marketing. **Journal of Business Research**, v. 125, n. July 2020, p. 772–779, 2021.

SHETH, J. N.; PARVATIYAR, A. The Evolution of Relationship Marketing. **International Business Review**, v. 4, n. 4, p. 397–418, 1995.

SIN, L. Y. M.; TSE, A. C. B.; YIM, F. H. K. CRM: Conceptualization and scale development. **European Journal of Marketing**, v. 39, n. 11–12, p. 1264–1290, 2005.

SIVAGURU, M.; PUNNIYAMOORTHY, M. Performance-enhanced rough k-means clustering algorithm. **Soft Computing**, v. 25, n. 2, p. 1595–1616, 2021.

SMITH, A. D. Retail-based loyalty card programmes and CRM concepts: An empirical study. **International Journal of Innovation and Learning**, v. 7, n. 3, p. 303–330, 2010.

SMITH, W. R. Product Differentiation and Market Segmentation as Alternative Marketing Strategies. **Journal of Marketing**, v. 21, n. 1, p. 3–8, 1956.

TALEB, N.; SALAHAT, M.; ALI, L. Impacts of Big-Data Technologies in Enhancing CRM Performance. **2020 6th IEEE International Conference on Information Management, ICIM 2020**, p. 257–263, 2020.

TAVAKOLI, M. *et al.* Customer Segmentation and Strategy Development Based on User Behavior Analysis, RFM Model and Data Mining Techniques: A Case Study. **Proceedings - 2018 IEEE 15th International Conference on e-Business Engineering, ICEBE 2018**, p. 119–126, 2018.

TRAN, L. T. T. Managing the effectiveness of e-commerce platforms in a pandemic. **Journal of Retailing and Consumer Services**, v. 58, n. April 2020, p. 102287, 2021.

TRUONG, D.; TRUONG, M. D. How do customers change their purchasing behaviors during the COVID-19 pandemic? **Journal of Retailing and Consumer Services**, v. 67, n. September 2021, p. 102963, 2022.

VERHOEF, P. C. *et al.* Customer Experience Creation: Determinants, Dynamics and Management Strategies. **Journal of Retailing**, v. 85, n. 1, p. 31–41, mar. 2009.

_____. CRM in data-rich multichannel retailing environments: A review and future research directions. **Journal of Interactive Marketing**, v. 24, n. 2, p. 121–137, maio 2010.

VERHOEF, P. C.; NOORDHOFF, C. S.; SLOOT, L. Reflections and predictions on effects of COVID-19 pandemic on retailing. **Journal of Service Management**, 2022.

WHITE, T. Haddop - The Definitive Guide. 2015.

WINER, R. S. A framework for customer relationship management. **California Management Review**, v. 43, n. 4, p. 89–105, 2001.

YOSEPH, F. *et al.* The impact of big data market segmentation using data mining and clustering techniques. **Journal of Intelligent and Fuzzy Systems**, v. 38, n. 5, p. 6159–6173, 2020.

YOSEPH, F.; HEIKKILA, M. **Segmenting retail customers with an enhanced RFM and a hybrid regression/clustering method** (H. D., B. M.R., & R. P.K., Eds.)2018 International Conference on Machine Learning and Data Engineering, iCMLDE 2018. **Anais...**Abo Akademi University, Turku, Finland: Institute of Electrical and Electronics Engineers Inc., 2019Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85062223282&doi=10.1109%2FiCMLDE.2018.00029&partnerID=40&md5=e4e911e1660a1dcd999774d794a0d889>>

ZHANG, Q. *et al.* Analysis of purchase history data based on a new latent class model for RFM analysis. **Industrial Engineering and Management Systems**, v. 19, n. 2, p. 476–483, 2020.

ZHAO, R.; LI, C. **Research on E-commerce Customer Segmentation Based on RFAC Model**2021 IEEE International Conference on Power, Intelligent Computing and Systems, ICPICS 2021. **Anais...**IEEE, 2021