

SEGMENTAÇÃO DE CLIENTES DE UMA MARCA DE VESTUÁRIO COM ALGORITMO DE APRENDIZADO¹

Cláudio Duarte Junior – claudio.duarte022@gmail.com

Iohanna Estenssoro – ioioestenssoro@gmail.com

Lucas da Silva Lima – lucasilvalima@gmail.com

Gustavo Rocha da Silva (Orientador) – gustavogrds@gmail.com

RESUMO

Com o passar do tempo, o uso de dados tornou-se extremamente relevante nas empresas, principalmente no setor de vestuário varejista possibilitando conhecer melhor o perfil de cada cliente. Em particular, a empresa Oriba desenvolveu o serviço Oriba em Casa para enviar peças de roupas sob remessa aos clientes. Na análise de dados a segmentação e clusterização são essenciais para definir as estratégias de relacionamento com o cliente. Para clusterizar os clientes e apresentar direcionamentos de atuação no contato com o consumidor, foram utilizados os *scores* definidos pelas variáveis de Recência, Frequência e Valor (RFV) dos dados colhidos no período de análise deste estudo. Em seguida estas informações foram inseridas no algoritmo não supervisionado *K-Means*. Por fim, este trabalho identifica os principais *clusters* nos quais a empresa deve direcionar a comunicação e ofertar o serviço de remessa.

Palavras-chave: Segmentação de Clientes, RFV, Varejo, K-Médias.

CUSTOMER SEGMENTATION OF A CLOTHING BRAND WITH A LEARNING ALGORITHM

ABSTRACT

Along the time, data usage has become extremely relevant in enterprises, especially in clothing retail enabling a better understanding of the profile of each customer. In particular, the company Oriba developed Oriba at Home service to ship garments under shipment to customers. The data segmentation and clustering are essential to define customer relationship strategies. In this study Recency, Frequency and Value (RFV) were the variables used to score client's data for the period analyzed. Using the score defined by RFV, the information was inputted on a non-supervised algorithm, K-Means, to cluster and present a guidance for customer relationship. Finally, this work identifies the main clusters in which the company must direct communication and offer the remittance service.

¹ Artigo do Trabalho de Conclusão de Curso, Graduação em Engenharia de Produção, EE, UPM, São Paulo, 2021.

Keywords: Customer Segmentation, RFV, Retail, K-Means.

1 INTRODUÇÃO

O comércio varejista é extremamente representativo na definição do PIB brasileiro. No entanto, durante a pandemia do Covid-19, foi decretado pelo Governo o fechamento temporário de todo o comércio “não-essencial” e a restrição da circulação de pessoas em locais fechados. Com isso, os segmentos varejistas de vestuários foram um dos mais afetados. Para possibilitar a reabertura do comércio, algumas medidas de segurança foram implementadas como, por exemplo, a obrigatoriedade do uso de máscara, a limitação do número de pessoas dentro de estabelecimentos, o fechamento de provadores e a redução do horário de funcionamento. Isto resultou no aumento de compras online e na queda do número de visitas às lojas físicas.

A Oriba Confecção e Comércio LTDA é uma entre as várias marcas varejistas de vestuário masculino que precisou se reinventar para manter seu padrão de atendimento e seguir todos os protocolos de segurança. Fundada em 2013, esta marca nasceu com o propósito de revolucionar a moda masculina básica nacional quebrando os padrões capitalistas.

A empresa produz peças básicas de alta qualidade a um preço justo, inspirando homens de todos os estilos a reduzirem as compras que seguem as tendências passageiras da moda. O maior apelo da loja é possibilitar mais *looks* com menos peças e assim reduzir o descarte de produtos no meio ambiente. Além da redução de descarte, a empresa se compromete em garantir que não existam más condições de trabalho nas linhas de produção (ORIBA, 2021).

Tendo em vista os ideais da marca, o conhecimento sobre o seu público e a preocupação com a saúde, em setembro de 2020 a Oriba desenvolveu uma nova experiência para os seus clientes das lojas físicas, chamada Oriba em Casa. Utilizando a base de dados de compras físicas, os vendedores, denominados como conselheiros, selecionam clientes que já consomem a marca, moram próximo às lojas físicas e possuem um ticket médio elevado.

Após esse levantamento, os conselheiros entram em contato via aplicativo de mensagens, ofertando o envio de uma mala com, em média, quinze peças de roupas escolhidas de acordo com o perfil, gosto, tamanho e demanda do cliente. Em até uma semana, as roupas são enviadas para a residência dele, permitindo-lhe experimentar as peças sem o uso de máscara, em um lugar mais confortável e sem pressa. No final desta experiência única, o consumidor escolhe quantas e quais peças deseja comprar e o restante é devolvido sem precisar sair de casa.

No comércio de varejo a área de experiência do cliente é de extrema relevância. Quando bem trabalhada pode ser usada na fidelização de clientes e na identificação de oportunidades de melhorias, aspirando maximizar os resultados financeiros. Essa área explora o desenvolvimento de produtos,

atendimento e ações de divulgação direcionadas aos consumidores de forma assertiva, para isso é essencial dispor de uma ferramenta que possibilite conhecer o perfil e os desejos do comprador.

Um dos maiores desafios, que qualquer empresa consolidada no mercado encontra, é o de como gerir de forma adequada as informações que são geradas durante os processos mais corriqueiros. O cliente é o foco da maioria das informações obtidas dentro de um negócio, porque é a partir da decisão dele, de adquirir um produto ou serviço, que é desencadeada a maioria dos processos produtivos. (MILANI *et al.*, 2015, p.170)

Gerenciar os dados de maneira efetiva a ponto de tratar os clientes particularmente é de suma importância. O processo de segmentação de dados é o protagonista para atingir um nível adequado de particularidade, porque esse processo consiste em agrupá-los de acordo com as semelhanças em variáveis definidas. Sabendo da importância de segmentar os usuários em diferentes categorias, outro desafio é estabelecer os critérios de divisão deles. Segmentação Geográfica, Demográfica, Psicográfica e Comportamental são algumas das opções mais utilizadas neste meio.

A Segmentação Comportamental é muito utilizada para identificar os mais diversos perfis em uma base de dados e avaliar a evolução comportamental de cada grupo. Consiste em utilizar dados com variáveis definidas, dividindo em segmentos o público em observação de acordo com os comportamentos que se assemelham em relação às variáveis.

Esse projeto avalia a efetividade em segmentar clientes de forma comportamental utilizando o método de Recência, Frequência e Valor (RFV). O RFV é utilizado para classificar os registros do banco de dados de acordo com os consumidores mais recentes, frequentes e rentáveis. Estas variáveis consistem em utilizar um valor de 1 a 5 para cada um dos critérios que dão nome à variável, onde 1 significa uma nota ruim e 5 uma nota excelente para cada um dos critérios. O produto esperado do método utilizado é a criação de perfis através do comportamento de compra.

Outro ponto a ser avaliado neste projeto é o algoritmo *K-means*. Dessa forma os dados são agrupados de maneira não supervisionada utilizando a distância entre centróides para formar grupos de acordo com as semelhanças dos dados vinculados às variáveis escolhidas. O resultado esperado é definir, de forma assertiva, grupos de clientes de acordo com seu perfil de compras.

Este trabalho avalia se a segmentação de clientes realizada após a aplicação do método RFV, realizada através do algoritmo não supervisionado *K-means* é capaz de fornecer insights para a escolha de clientes que aceitariam participar do novo método de vendas de roupas da marca Oriba.

2 REVISÃO DA LITERATURA

O objetivo de um sistema de segmentação é identificar grupos, nos quais são separados de forma que unam os clientes mais parecidos possíveis e ao mesmo tempo consigam distinguir os demais segmentos. Se esse agrupamento é bem projetado, os membros de um grupo vão possuir

interesses, atitudes e comportamentos semelhantes, respondendo analogamente aos elementos do mix de marketing, como preços, promoção e canal de vendas. Mediante os conhecimentos de segmentação, são formados roteiros estratégicos destinados a aproveitar as principais oportunidades de geração de lucro dentro de cada grupo exclusivo de clientes. Resultando no encurtamento dos ciclos de compra do cliente, aumentando gastos, construindo maior fidelidade, aprofundando as vendas de produtos cruzados, reduzindo custos de serviços e suportes. (SOUDAGAR, 2012)

As empresas capazes de saber que seus clientes diferem sobre questões de gostos e comportamentos, procuram formas de ter assertividade nas tomadas de decisão, tanto em relação a marketing quanto em relação a qualificar o relacionamento com seus consumidores, visando o sucesso de seu negócio. Para isso, é importante entender que a gestão de relacionamento com o cliente (*Customer Relationship Management – CRM*) está diretamente ligada ao método de RFV e torna-se uma ferramenta importante no auxílio para o aumento de lucratividade e identificação dos melhores usuários para que esses possam se tornar ainda mais importantes para as atividades da empresa (NGUYEN; MUTUM, 2012; ŠEBJAN; BOBEK; TOMINC, 2016). A estratégia de se possuir um CRM dentro da empresa é importante, pois a aproxima de seus clientes e dessa forma possibilita o gerenciamento de suas relações com eles. Existe uma forte correlação entre essa estratégia de negócio e a segmentação dos dados, pois quando há qualidade nas relações entre empresa e cliente, há também um maior entendimento sobre qual o padrão de comportamento do consumidor e o que o ele busca, trazendo informações importantes sobre ele no momento de classificá-lo dentro de algum grupo. O trabalho de Shim *et al.* (2012), por exemplo, trata de estratégias de CRM para um *shopping* online de pequeno porte e utiliza como base regras de associação e padrões sequenciais alcançados pela análise dos dados transacionais de compras. Os autores utilizam atributos como idade, frete, canal de registro, entre outros, além de RFV, para desenvolver um modelo classificador que conseguisse identificar os clientes VIPs e com isso, conseguir propor um conjunto de estratégias de *marketing* baseadas na identificação das regras e padrões alcançados pela análise.

Criado por Hughes (1994), RFV é um método de classificação de registros de um banco de dados de clientes que permite saber quem são os compradores mais recentes, os mais frequentes e os mais rentáveis, ou seja, é uma forma que permite a criação de perfis através do comportamento de compra de cada consumidor. Nos últimos anos, muitos pesquisadores consideram o modelo RFV para o desenvolvimento de previsão e classificação.

O método RFV permite a identificação dos diferentes tipos de consumidores, do histórico de compras e do relacionamento com a empresa, mensurando em que data a compra foi realizada, quantas vezes e quanto foi gasto em um determinado período, resultando no desenvolvimento de campanhas de *marketing* personalizadas (WEI *et al.*, 2010).

A maneira habitual para a segmentação do modelo RFV, é classificar os dados de cada cliente com base nos valores contidos em cada atributo R, F, V, isto é realizado pela divisão do conjunto de dados em cinco partes iguais. Os dados R são classificados em ordem crescente, do mais recente para o mais antigo; os dados F são classificados em ordem decrescente, maior frequência de compra para a menor; e os dados V também são classificados em ordem decrescente, do maior para o menor valor monetário. (CARVALHO, 2018). Com isso, considerando cada atributo, os 20% dos registros iniciais do topo do ordenamento é codificado como 5, o próximo segmento de 20% como 4, e assim por diante. Desta forma, todos os clientes podem ser apresentados por rótulos como 5-5-5,5-5-4,...,1-1-1; totalizando 125 possibilidades de classificação (5x5x5) (WEI *et al.*, 2013)

Em um estudo realizado por Hu & Yeh (2014) no setor de varejo utilizou-se análise de RFV somado ao algoritmo de clusterização *K-means*. O objetivo era definir um padrão de recência, frequência e valor e desenvolver um novo algoritmo para descobrir padrões RFV e aproximar conjuntos de clientes. Durante o estudo, os autores analisaram o comportamento de compra histórico dos clientes considerando esses padrões ao invés de avaliar valores de padrões do ponto de vista do cliente. Os resultados mostraram que a abordagem proposta foi eficiente e descobriu a maior parte dos padrões RFV do cliente.

Para um outro estudo realizado em uma companhia chinesa, You *et al.* (2015) além do método de RVF e *K-means*, também foi utilizada uma árvore de decisão. A ideia desse estudo foi propor um modelo para prever com precisão a quantidade mensal de fornecimentos usando a abordagem de recência, frequência e valor para agrupar os clientes em diferentes grupos e isso ajudou os gerentes a identificar características ocultas nas diferentes categorias de clientes (DOĞAN *et al.*, 2018). Além disso, esse modelo também ajudou a prever as estratégias de marketing que podem reduzir muito o estoque para cada categoria, e de acordo com os autores os resultados computacionais mostraram que o modelo aplicado supera os modelos de previsão mais conhecidos de *softwares* comerciais.

Em outro trabalho, Durson e Caber (2016), propuseram a segmentação de clientes de um hotel também utilizando o método RFV e *K-means*. Os autores obtiveram 8 *clusters* de clientes de acordo com o seu *score* RFV: leais, frequentes na temporada de verão, que compram em grande escala, frequentes na temporada de inverno, possuem alto potencial, potenciais de temporada de verão, potenciais na temporada de inverno e novos. Neste artigo os perfis dos clientes foram comparados com os novos segmentos.

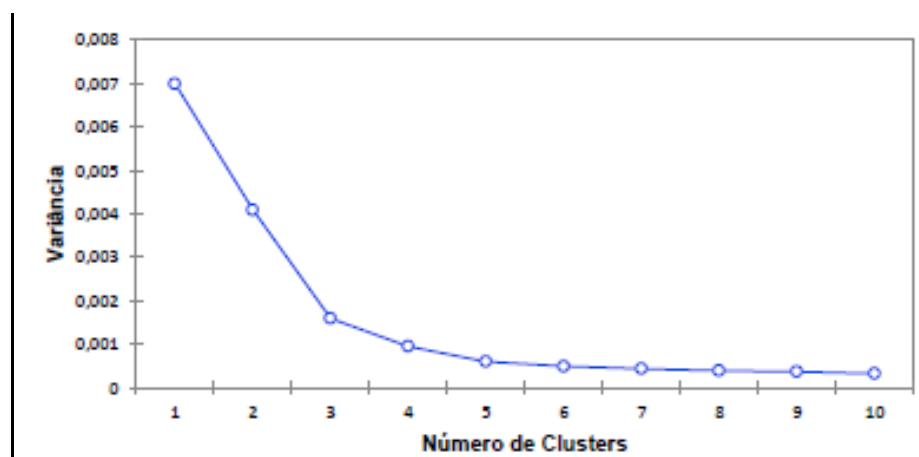
Sarvari (2016) desenvolveu um estudo na Turquia, utilizando dados de uma pizzaria pertencente a uma cadeia de restaurantes. Foram analisados os métodos RFV, *K-means* e associação de regras, com o objetivo de determinar a melhor abordagem para cada grupo de clientes, através de testes em diferentes cenários projetados. Com isso, chegou-se à conclusão de que uma abordagem adequada traz resultado se o segmento possuir uma semelhança forte entre si.

A partir do exposto acima, é possível notar que em grande parte das referências citadas a utilização do método RFV é aplicada em conjunto com o algoritmo *K-means*. Esse algoritmo é do tipo não supervisionado, ou seja, ele não trabalha com dados rotulados (SANTANA, 2017) mas utiliza apenas de sua própria inteligência na busca por padrões encontrados nos dados dos clientes, sem precisar de nenhuma instrução sobre como encontrar esses padrões. A principal característica do *K-means* é neste caso, atribuir os clientes em grupos ou *clusters* com dados que tenham similaridade entre si.

No artigo de Khajvand *et al.* (2011) foi realizada a clusterização de clientes com o algoritmo *K-means* em uma empresa de saúde e beleza com o intuito de identificar o valor de lealdade dos clientes para a implementação de um *marketing* mais direcionado e personalizado. Além da segmentação desses clientes, ter proporcionado estratégias de *marketing* e vendas mais efetivas que ajudaram na retenção de clientes também ajudou os tomadores de decisão a identificar segmentos de mercado com mais clareza (DOĞAN *et al.*, 2018).

Em um outro estudo, desta vez realizado em uma empresa multinacional americana de eletrodomésticos (DAPUNT, 2018) o autor utilizou o método *K-means* que por sua vez separou os clientes em quatro clusters também visando possibilitar direcionais estratégicos de *marketing* de acordo com as suas características. O autor desse estudo utilizou o método *Elbow* para definir a quantidade de grupos K. De acordo com Dapunt (2018, p.37) esse método valida e interpreta a consistência da análise do cluster examinando a porcentagem de variância em função do número de *clusters*. Para um melhor entendimento desse método a Figura 1, extraída desse mesmo estudo, exemplifica melhor o funcionamento do método *Elbow*. Pode-se notar pelo gráfico que a linha começa a se estabilizar a partir do *cluster* 4, que foi a quantidade definida pelo estudo.

Figura 1 - Método Elbow



Fonte: Luan de Oliveira Dapunt (2018)

De acordo com Carvalho (2018, p.127) “[...] um dos grandes desafios para utilização do algoritmo *K-means* é justamente encontrar o melhor número de K [...]”. Dado isso, é importante saber

que existem métodos disponíveis na literatura que auxiliam na definição do número K, além do exemplo apresentado acima temos os métodos: *Gap Statistic*, *Average Silhouette* e *Davies-Bouldin Index*, entre outros (ROJLERTJANYA, 2019). Apesar dessas dificuldades, não há solução totalmente adequada para encontrar o valor k verdadeiro porque cada método determina diferentes dimensões e variáveis (NAEEM & WUMAIER, 2018 *apud* ROJLERTJANYA, 2019, p. 17).

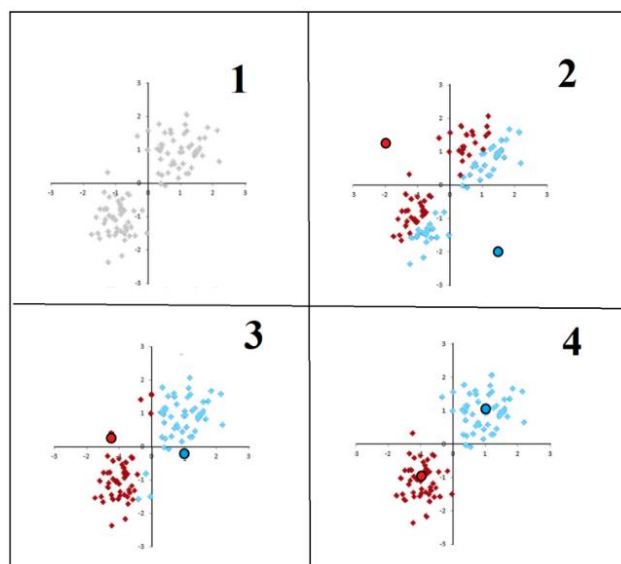
3 METODOLOGIA

Neste projeto foi aplicado o método RFV em conjunto com o algoritmo não-supervisionado *K-means*. O *K-means* é um algoritmo de clusterização relativamente simples devido a sua implementação e execução serem rápidas e tem sido amplamente utilizado na segmentação de clientes e no reconhecimento de padrões e recuperação de informações (Qin *et al*, 2010). Além disso possui um baixo custo computacional e isso permite que seja aplicado em grandes volumes de dados. (CARVALHO, 2015).

A aplicação do algoritmo *K-means* é realizada em 4 etapas. Na primeira etapa ocorre a definição aleatória do número de “k” centróides, ou seja, os pontos centrais dos clusters. Após isso é realizada a atribuição dos dados aos *clusters* através do cálculo da distância entre os pontos de dados a cada um dos centróides, esses registros são atribuídos aos centróides com a menor distância e dessa forma os dados estarão divididos de acordo com o número de centróides que foi determinado pelo argumento k. Visto que os dados já estão atribuídos aos seus respectivos *clusters* são calculados novos centróides utilizando a média de todos os pontos de cada *cluster*, e o valor médio será o novo centróide (CARVALHO, 2015). Por fim é realizada uma otimização do *K-means* através de repetições nas atribuições de dados aos *clusters* e os cálculos de novos centróides. As repetições são feitas até o momento em que os centróides dos *clusters* se tornam estáticos, ou seja, deixam de mudar as suas coordenadas.

Essas quatro etapas da aplicação do modelo *K-means* podem ser visualizadas na Figura 2, onde os círculos representam os centróides e os losangos representam os dados, pode-se perceber que os dados não se alteram, mas após a repetição das etapas 2 e 3 os centróides alteram de coordenadas conforme a alteração dos *clusters*.

Figura 2 - Etapas da clusterização do algoritmo *K-means*.



Fonte: Andras Kalmar (2012)

De acordo com Soudagar (2012, p. 18), apesar de ser um algoritmo de bastante praticidade em sua aplicação, o *K-means* possui algumas desvantagens. A cada execução os resultados não serão os mesmos já que os *clusters* finais dependem da primeira atribuição aleatória, além de ser difícil comparar a qualidade dos *clusters* produzidos e não ser apropriado caso esses sejam não-globulares, ou seja, diferentes da Figura 2 em que os dados aglomerados se assemelham ligeiramente a um formato circular.

Para o uso desse algoritmo no presente trabalho será necessário passar por alguns passos antes de iniciar as quatro etapas citadas anteriormente. Antes de definir o número K existe a necessidade seleção de características relacionadas a recência, frequência e valor, para isso cabe a pergunta: “Quais são os dados relacionados a recência, frequência e valor?”, por exemplo: a data de última compra de um cliente nos proporciona o *aging* (envelhecimento) do cliente através da diferença de dias entre essa data e uma data de referência pré-definida, nos dando um indicador de recência, ou seja, quanto maior o *aging* menor será o escore de recência desse cliente. Para frequência também poderá ser utilizado como dado principal as datas de transações do cliente, quanto maior o número de transações do cliente na loja maior será seu *score* de frequência. Por fim, o *score* de valor pode ser definido pelo ticket médio ou receita bruta gerada pelo cliente para a loja.

Após serem definidos quais dados representarão os três critérios do RFV haverá a etapa de extração de dados do banco de dados da empresa, porém os dados na grande maioria das vezes não são exportados na melhor forma de se trabalhar e por isso também será necessária uma etapa de tratativa de dados, para que eles fiquem imputáveis no algoritmo. Por fim, após a tratativa de dados, iniciam-se as quatro etapas de aplicação do algoritmo.

Ressaltamos ainda, que a saída do algoritmo *K-means* é uma etiqueta que atribui um grupo a um determinado conjunto de dados. A partir da saída do algoritmo, torna-se importante um processo

de estudo destes agrupamentos propostos. Esta etapa é realizada através de uma análise exploratória dos dados para cada um dos grupos criados. Desta forma, é possível obter uma segmentação que pode ser explorada através de finalidades de negócio.

4 RESULTADOS E DISCUSSÕES

Para as análises de dados deste estudo foram utilizadas três bases de dados, disponibilizadas pela Oriba. A *customers* que contém informações cadastrais dos clientes como id, nome, *e-mail*, telefone, cidade e estado; a *product sales* que possui dados sobre vendas de produtos especificando as características de cada produto, por exemplo, valor, tamanho, cor, uma breve descrição da peça de roupa e o canal (unidade da loja); e a base *sales* com dados referentes a vendas de produtos, sem informações relacionadas ao produto, mas com informações como a data de compra e o valor.

As bases foram disponibilizadas em um arquivo de banco de dados, então para conseguirmos juntá-las e filtrá-las de acordo com as informações necessárias utilizamos o banco SQLite. A partir disso foi possível identificar as compras registradas sem CPF, os clientes de fora da cidade de São Paulo e os dados que estavam duplicados, resultando em uma base de vendas no período entre setembro de 2020 e setembro de 2021.

Além das três bases de dados presentes no *dataset* citadas anteriormente, também tivemos acesso a uma base com os clientes que foram contatados para aderir a campanha do Oriba em Casa, onde foi possível cruzar os dados com as demais bases para obter informações sobre quais clientes aceitaram participar da campanha.

Através do Google Colab foi possível aplicar o algoritmo *K-means* para realização da clusterização, utilizando as linguagens de programação *python*, além da ferramenta Microsoft Excel que permitiu calcular os escores do RFV e os clusters onde cada cliente se encaixa. A partir destes resultados construímos os gráficos para análise.

Ao analisarmos a base de clientes do Oriba em Casa, deduzimos que muitos contatos foram feitos, porém apenas 12% foram convertidos, com isso, observamos o quão importante é a implantação de uma estratégia que seja mais assertiva na conversão do Oriba em Casa.

Figura 3 - Gráfico de Conversão Oriba em Casa



Fonte: Autoria Própria

Para realizar a criação do modelo RFV foram utilizadas as métricas:

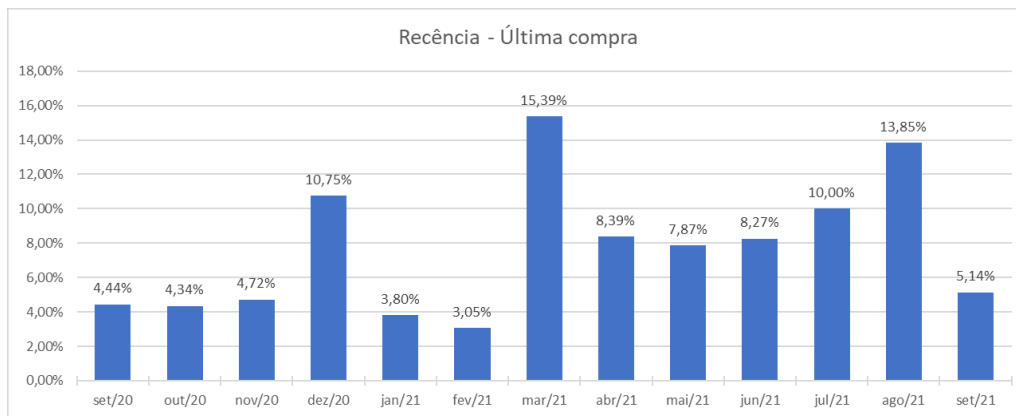
- Recência: diferença de dias entre a data de última compra até a data final do período analisado (15 setembro de 2021).
- Frequência: quantidade total de compras do cliente no período entre setembro de 2020 e setembro de 2021.
- Valor: valor total de compras que cada cliente realizou entre setembro de 2020 e setembro de 2021.

Cada métrica teve seus valores atribuídos de 1 a 5, sendo 1 o pior score e 5 o melhor.

Com os atributos já definidos buscamos realizar análises sobre cada métrica específica para entendermos a distribuição dos clientes da base, dentro de cada atributo e identificarmos possíveis padrões ou tendências nos gráficos.

A primeira métrica analisada foi a de recência, ao olharmos o gráfico da figura 4 podemos identificar que a maior parte dos clientes realizaram a sua última compra no mês de dezembro de 2020, mês em que há um aumento de vendas devido as festividades de final de ano. Em março de 2021, período em que antecedeu a segunda onda de covid, o consumo de grande parte da população subiu expressivamente. E em agosto de 2021, o gráfico indica a retomada de vendas após o período de *lockdown*.

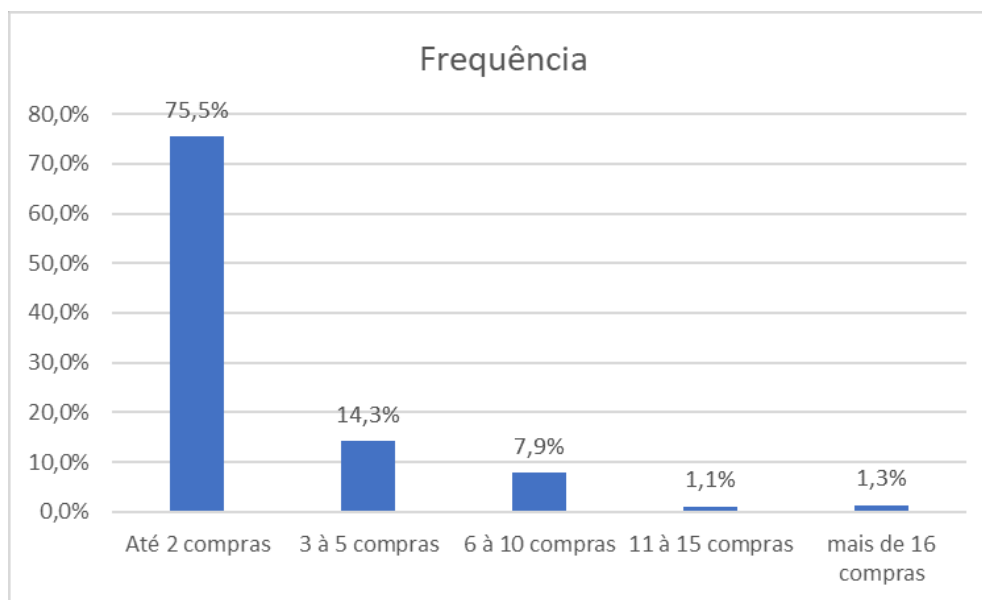
Figura 4 - Gráfico de distribuição do atributo Recência



Fonte: Autoria Própria

Em seguida foi estudado o atributo de frequência, onde ao observar o gráfico da figura 5 podemos apontar que mais de 75% dos clientes da base estudada realizaram até 2 compras no período de um ano. Isto mostra que a maior parte dos clientes não costuma comprar os produtos de forma recorrente, mas sim fazer compras de forma esporádica.

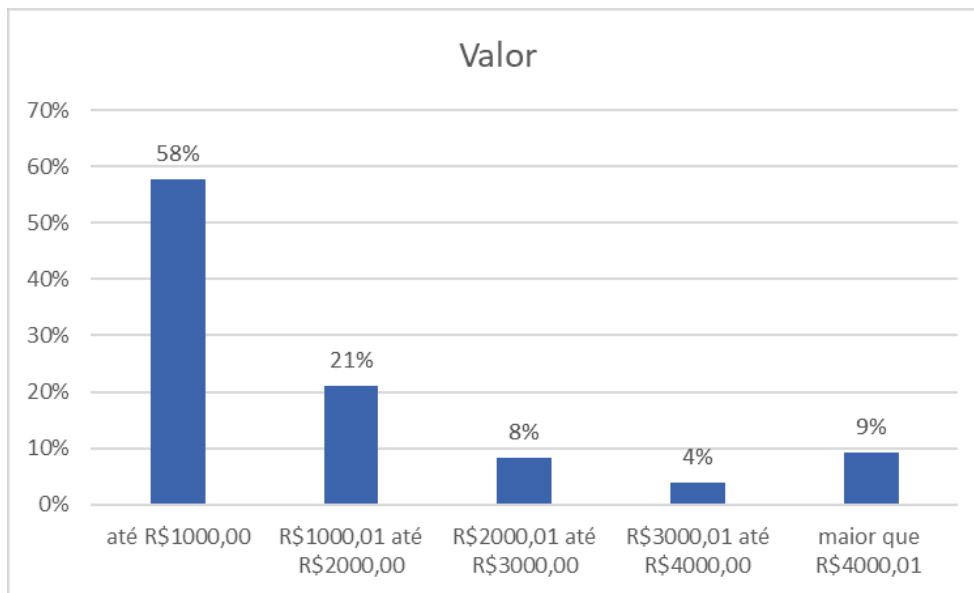
Figura 5 - Gráfico de distribuição do atributo Frequência



Fonte: Autoria Própria

Logo após foi averiguado o atributo valor, ao visualizarmos a figura 6, podemos dizer que aproximadamente 80% dos clientes gastaram até R\$2000,00.

Figura 6 - Gráfico de distribuição do atributo Valor

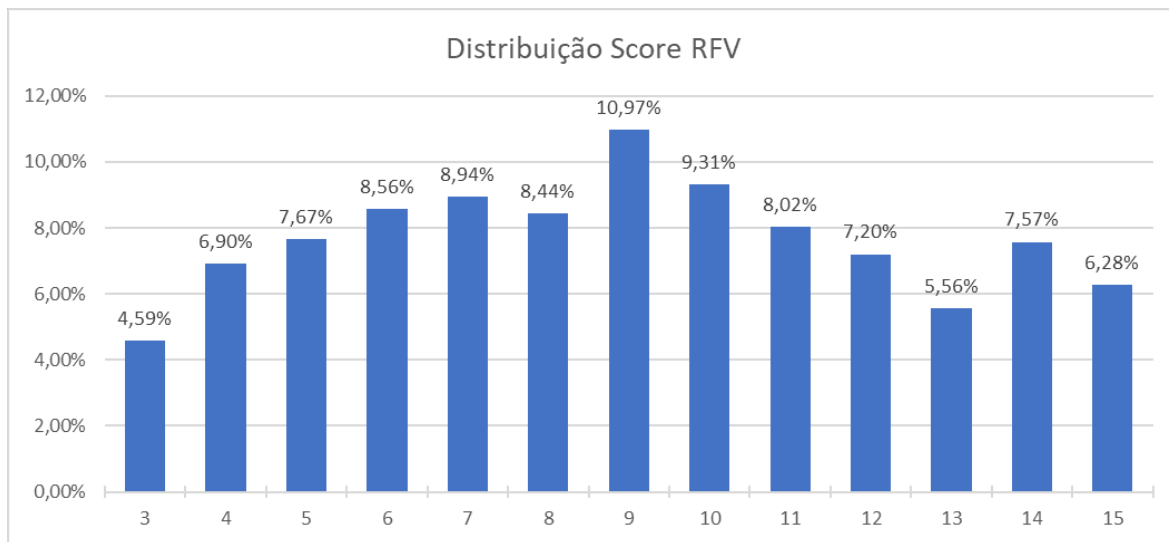


Fonte: Autoria Própria

Após analisarmos a base *sales* com os escores definidos, por meio dos atributos recência, frequência e valor separadamente, as hipóteses são de que grande parte dos clientes da Oriba realizaram sua última compra no final de 2020. Retornando apenas em março de 2021, período pré-fechamento das lojas físicas e em agosto 2021 com a reabertura das lojas e ação de dia dos pais. Além disso, o número de visitas e o ticket médio de compras nas lojas continuaram baixos, indicando que o mercado continua sentindo os impactos da pandemia e destacando a importância de inovar na forma de vender.

Depois de realizada a divisão dos clientes dentro de cada atributo RFV, numerando-os de 1 a 5 e atribuimos a cada cliente um *score* total, somando os escores específicos de cada produto. Assim cada cliente poderá ter um *score* total entre 3 e 15. No gráfico da figura 7, podemos identificar que não houve uma grande concentração de clientes dentro de um *score* específico, a distribuição dos clientes dentro de cada *score* ficou bem equilibrada, tendo a maior concentração de clientes com o *score* 9, considerado um *score* intermediário.

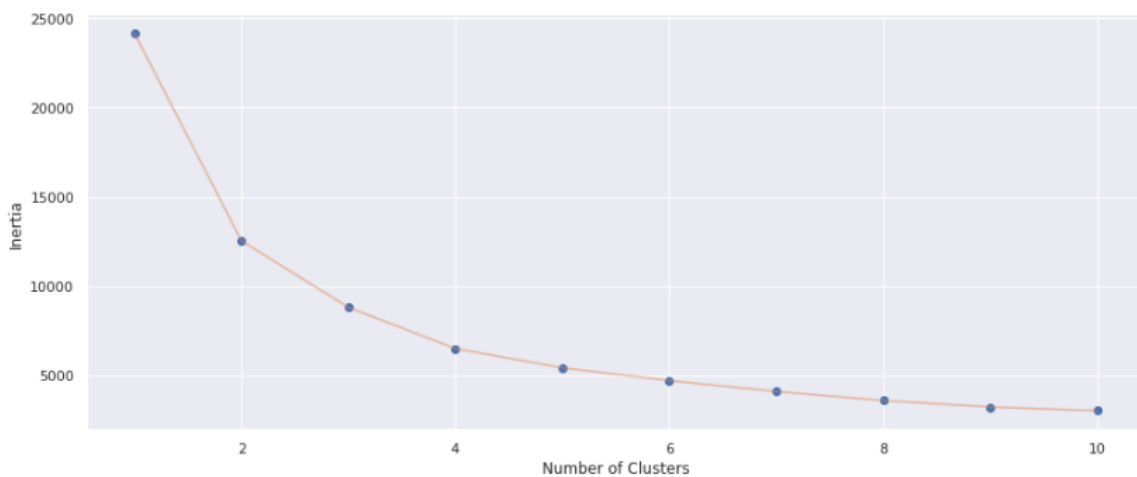
Figura 7 - Gráfico Distribuição Score RFV



Fonte: Autoria Própria

Após a consolidação da base *sales* e a aplicação da pontuação do RFV, optamos por utilizar o método do *Elbow*, mais conhecido como método do cotovelo, para determinar a quantidade de clusters ideal para o uso do *K-means*. Com os dados disponibilizados pela Oriba, identificamos que o ponto de inflexão corresponde a $k = 5$ e utilizamos este dado para definir qual cluster cada cliente pertence.

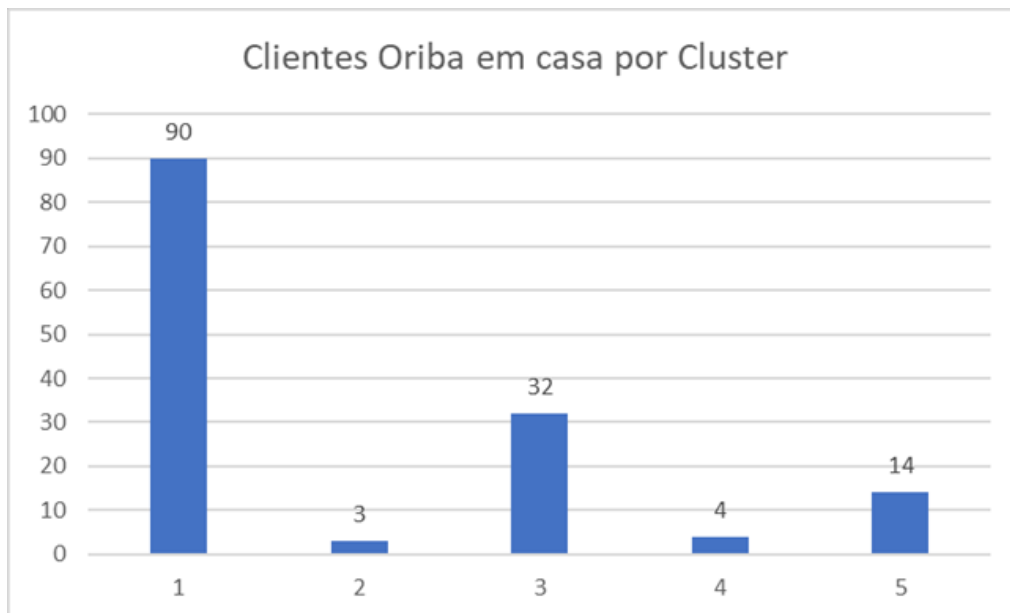
Figura 8 - Gráfico do Método *Elbow*



Fonte: Autoria Própria

Segundo a figura 3, apenas 12% dos contatos realizados para venda do Oriba em Casa foram revertidos, isso resulta em apenas 143 vendas concluídas em mais de 2000 contatos realizados. Devido ao baixo índice de conversão, observado na figura 9, podemos afirmar que o *cluster* que mais se destacou foi o 1, com cerca de 63% das vendas.

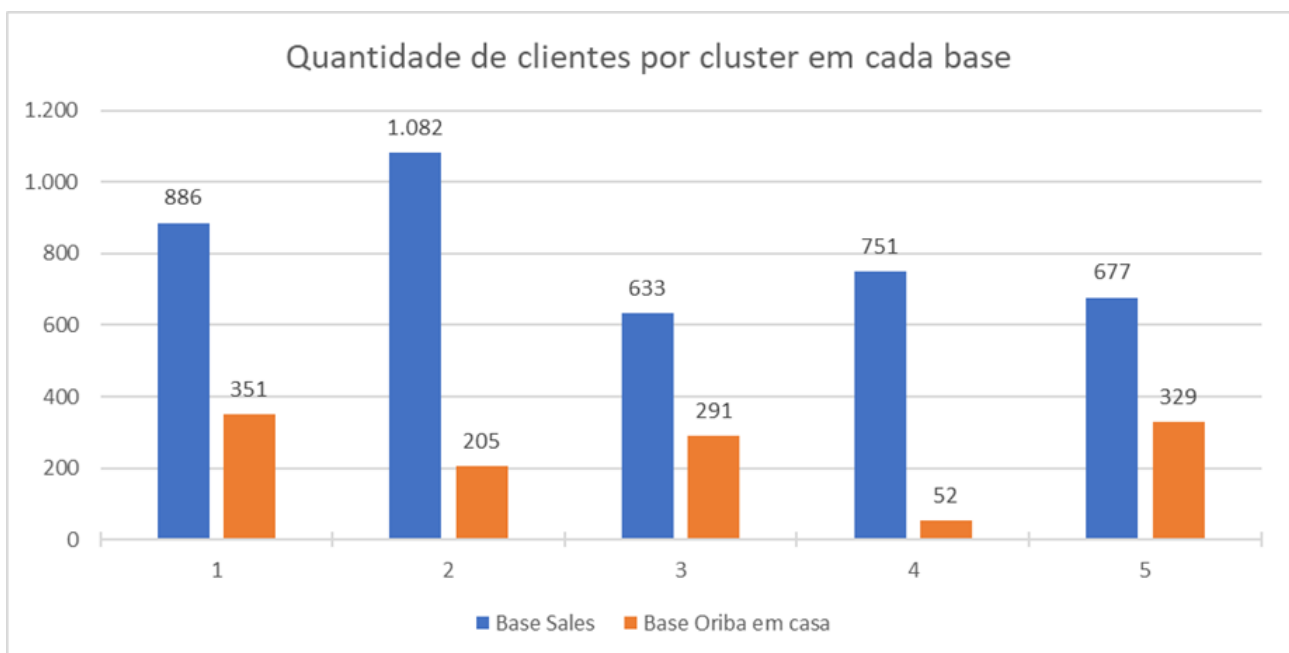
Figura 9 - Quantidade de clientes convertidos por *cluster*



Fonte: Aatoria Própria

Através da figura 10 é possível analisar que o *cluster* 1 tem uma das maiores parcelas de clientes da marca com o maior número de contatos feitos, porém esta quantidade de contatos realizados poderia ser maior, se houvesse um sistema de direcionamento para os conselheiros abordarem os clientes com maior recência, frequência e valor da loja, resultando em uma maior conversão do Oriba em Casa.

Figura 10 - Quantidade de clientes por *cluster*



Fonte: Aatoria Própria

A tabela da figura 11 nos traz um resumo dos dados citados anteriormente. Podemos identificar através dela a conversão por *cluster* sobre os clientes contatados e sobre a base total do período analisado. Na tabela também foi incluída a visão do percentual de clientes contatados sobre

o total de clientes por *cluster*, esse dado nos informa se a empresa está prospectando o melhor grupo de clientes a aceitar a campanha Oriba em casa.

Nessa mesma tabela podemos notar que 48,6% da base do *cluster 5* foi contatada, de maneira ineficiente, pois este grupo é formado por índices de *score* do RFV, que somam no máximo 5 pontos, ou seja é composta por clientes que possuem baixa frequência nas lojas físicas e baixo ticket médio. Também é possível visualizar que a taxa de conversão deste *cluster* foi extremamente baixa. Essa informação é de extrema relevância por evidenciar que o *cluster* com mais contatos garante maior reversão de clientes para o serviço ofertado.

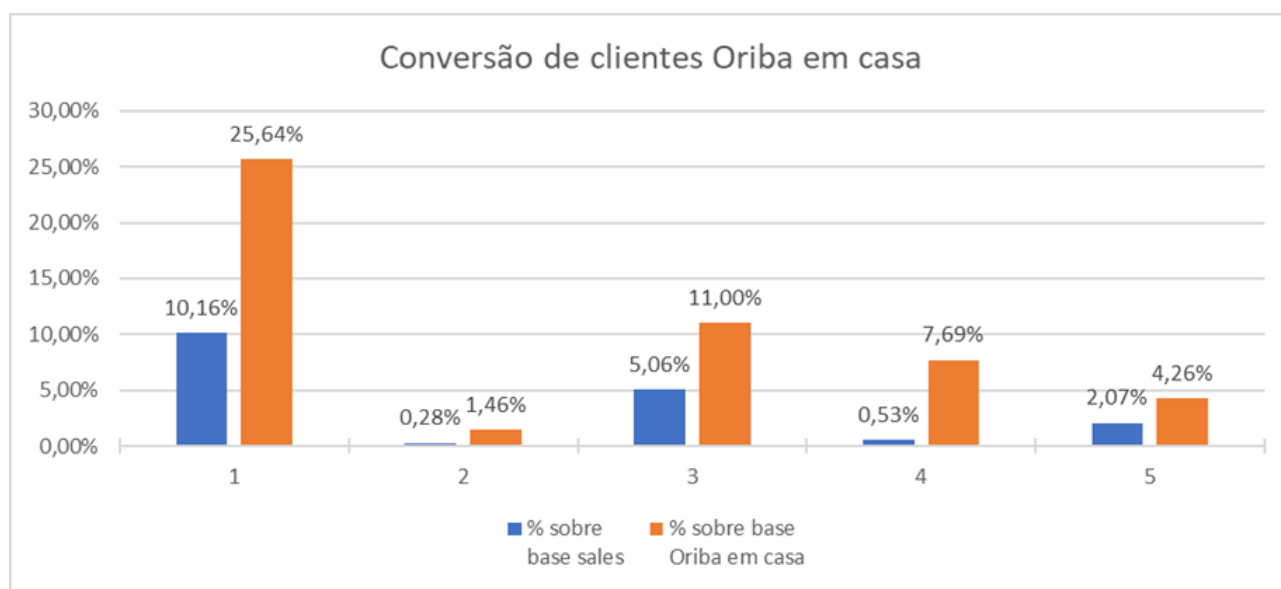
Figura 11 - Tabela resumo *clusters*

Cluster	Convertiu Oriba em Casa	Base Sales	Base Oriba em casa	% contatados por cluster	% sobre Base Sales	% sobre Base Oriba em casa
1	90	886	351	39,6%	10,16%	25,64%
2	3	1.082	205	18,9%	0,28%	1,46%
3	32	633	291	46,0%	5,06%	11,00%
4	4	751	52	6,9%	0,53%	7,69%
5	14	677	329	48,6%	2,07%	4,26%
Total	143	4.029	1.228	30,5%	3,55%	11,64%

Fonte: Aatoria Própria

Aprofundando o estudo dos clientes, podemos observar no grupo 1 que 25,64% dos clientes contatados foram convertidos no Oriba em Casa e que isso representa apenas 10,16% do *cluster 1* da base sales, ou seja, apesar do índice de conversão se destacar em relação aos demais *clusters*, continua sendo uma porcentagem baixa de contatos convertidos para o grupo que tem melhor taxa de conversão.

Figura 12 - Conversão de clientes Oriba em casa



Fonte: Aatoria Própria

Por fim, os dados apresentados na Figura 11 e na Figura 12 aparentemente demonstram que a utilização de *K-means* para clusterizar a base de dados dos clientes com as variáveis de RFV pode ser mais efetivo aproveitando resultados essenciais, por exemplo, o *cluster 5* por mais que a quantidade de contatos efetuados se assemelhe ao *cluster 1* possui uma taxa de conversão muito inferior ao *cluster 1*.

5 CONSIDERAÇÕES FINAIS

A principal proposta deste estudo foi a análise dos dados Oriba, por meio da aplicação do método RFV, para apresentar à empresa uma nova forma de enxergar seus clientes, e através disso tomar decisões estratégicas mais eficientes, não só para o seu modelo de negócio, mas também para proporcionar uma experiência ainda melhor aos seus clientes.

Entendemos que apenas com a utilização de 3 métricas é possível realizar uma segmentação de clientes eficiente que possa servir de suporte para as tomadas de decisões das lideranças da empresa. Apesar de ter poucas variáveis e termos poucos *insights* sobre o perfil do cliente existe uma conclusão importante sobre estratégias de contato, por exemplo, a diferença da taxa de conversão entre o *cluster 1* (25,64%) e o *cluster 5* (4,26%) na base de clientes contatados referente ao Oriba em Casa mesmo que a quantidade de clientes acionados por *cluster* se assemelham. Vale ressaltar que a aplicação desse modelo pode ser importante para a estratégia da empresa de aumentar o raio de entregas do serviço Oriba em Casa, pois dessa maneira o contato com o cliente para a conversão do serviço será mais efetiva.

REFERÊNCIAS

CARVALHO, Leandro da Silva. **UMA NOVA ABORDAGEM PARA A CRIAÇÃO DE PERFIS DE CLIENTES RENTÁVEIS UTILIZANDO MACHINE LEARNING EM AMBIENTE DE CLOUD COMPUTING**. 2018. 194 f. Tese (Doutorado) - Curso de Engenharia Civil, Coppe, Ufrj, Rio de Janeiro, 2018.

CARVALHO, Leandro da Silva. DESENVOLVIMENTO DE MODELOS RFM COM AJUDA DE PLATAFORMAS DE APRENDIZADO DE MÁQUINA NA NUVEM. In: CILAMCE, Não use números Romanos ou letras, use somente números Arábicos., 2015, Rio de Janeiro. **Proceedings of the XXXVI Iberian Latin-American Congress on Computational Methods in Engineering**. Rio de Janeiro: Ney Augusto Dumont, 2015. p. 22-25.

DAPUNT, Luan de Oliveira. **Uma abordagem de segmentação de clientes utilizando o método recency-frequency-monetary adaptado e o método Analytic Hierarchy Process**. 2018. 62 f. TCC (Graduação) - Curso de Engenharia de Produção Mecânica, Universidade Estadual Paulista Júlio de Mesquita Filho, Guaratinguetá, 2018.

DOĞAN, Onur; AYÇİN, Ejder; BULUT, Zeki Atıl. CUSTOMER SEGMENTATION BY USING RFM MODEL AND CLUSTERING METHODS: A CASE STUDY IN RETAIL INDUSTRY.

International Journal Of Contemporary Economics And Administrative Sciences. [S.I], p. 1-19. jan. 2018.

DURSUN, A. & CABER, M. (2016). **Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis.** *Tourism Management Perspectives*, 18, 153-160.

HU, Ya-Han; YEH, Tzu-Wei. Discovering valuable frequent patterns based on RFM analysis without customer identification information. **Knowledge-Based Systems**, [S.L.], v. 61, p. 76-88, maio 2014. Elsevier BV. <http://dx.doi.org/10.1016/j.knosys.2014.02.009>.

HUGHES, A., 1994. **Strategic database marketing.** Chicago, Ill.: Probus Pub. Co.

KHAJVAND, M., ZOLFAGHAR, K., ASHOORI, S. & ALIZADEH, S. (2011). **Estimating customer lifetime value based on RFM analysis of customer purchase behavior: case study.** *Procedia Computer Science*, 3, 57-63.

KALMAR, Andras. **Steps of the K-mean clustering algorithm.** 2012. Disponível em: https://www.researchgate.net/figure/Steps-of-the-K-mean-clustering_algorithm_fig5_321051036. Acesso em: 13 maio 2021.

MILANI, Aline Arruda; GASPAR, Victor; ALBRECH, Alfredo Junior Paiola; FAUSTO, Daiane Aparecida; MIGLIAVACCA, Rafaela Alenbrant. **Processo de gestão da carteira de clientes.** *Revista Ipecege*, [S.L.], v. 1, n. 2, p. 169-186, 27 mar. 2015. I-PECEGE. <http://dx.doi.org/10.22167/r.ipecege.2015.2.169>.

NAEEM, Sajid; WUMAIER, Aishan. **Study and Implementing K-mean Clustering Algorithm on English Text and Techniques to Find the Optimal Value of K.** *International Journal Of Computer Applications*, [S.L.], v. 182, n. 31, p. 7-14, 17 dez. 2018. Foundation of Computer Science. <http://dx.doi.org/10.5120/ijca2018918234>.

ORIBA (org.). **Nossa História.** Disponível em: <https://www.oriba.com.br/pages/quem-somos>. Acesso em: 19 maio 2021.

ROJLERTJANYA, Ponlacha. **CUSTOMER SEGMENTATION BASED ON THE RFM ANALYSIS MODEL USING K-MEANS CLUSTERING TECHNIQUE: a case of it solution and service provider in thailand.** 2019. 103 f. Dissertação (Mestrado) - Curso de Computer Engineering, Bangkok University, Bangkok, 2019.

SANTANA, Felipe. **Entenda o Algoritmo K-means e Saiba como Aplicar essa Técnica.** 2017. Disponível em: <https://minerandodados.com.br/entenda-o-algoritmo-k-means/#inicializacao>. Acesso em: 13 maio 2021.

SOUDAGAR, Rana. **Customer Segmentation and Strategy Definition in Segments: case study: an internet service provider in iran.** 2012. 80 f. Dissertação (Mestrado) - Curso de Administration, Department Of Business Administration, Technology And Social Sciences, Luleå University Of Technology, Luleå, 2012.

WEI, J., LIN, S. AND WU, H., 2010. **A review of the application of RFM model.** *African Journal of Business Management*, 4(19), pp.4199-4206.

WEI, J., LEE, M., CHEN, H. AND WU, H., 2013. **Customer relationship management in the hairdressing industry: An application of data mining techniques.** Expert Systems with Applications, 40(18), pp.7513-7518.

YOU, Zhen et al. **A decision-making framework for precision marketing.** Expert Systems With Applications, [S.L.], v. 42, n. 7, p. 3357-3367, maio 2015. Elsevier BV.
<http://dx.doi.org/10.1016/j.eswa.2014.12.022>.

AGRADECIMENTOS

A todos da empresa Oriba por nos receberem, disponibilizarem as informações detalhadas do serviço e atenderem eventuais dúvidas. Agradecemos principalmente ao Felipe Niwa e Juliana SPTO que foram essenciais para o desenvolvimento deste estudo.

Agradecemos também às nossas famílias por todo apoio e suporte durante todo o curso. E a todo corpo docente da Universidade Presbiteriana Mackenzie por nos prepararem com tamanha excelência para esse momento tão importante.