

**UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO E
SISTEMAS
NÍVEL MESTRADO**

FREDERICO ERNESTO EIDELWEIN

**DIMENSIONAMENTO DE ESTOQUES NA CADEIA DE SUPRIMENTOS A PARTIR
DA CLUSTERIZAÇÃO**

SÃO LEOPOLDO

2024

FREDERICO ERNESTO EIDELWEIN

**DIMENSIONAMENTO DE ESTOQUES NA CADEIA DE SUPRIMENTOS A PARTIR
DA CLUSTERIZAÇÃO**

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Engenharia de Produção e Sistemas, pelo Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Universidade do Vale do Rio dos Sinos (UNISINOS).

Orientadora: Prof. Dr. Maria Isabel Motta Morandi

Co-Orientador: Prof. Dr. Daniel Pacheco Lacerda

São Leopoldo

2024

E34d

Eidelwein, Frederico Ernesto.

Dimensionamento de estoques na cadeia de suprimentos a partir da clusterização / Frederico Ernesto Eidelwein. – 2024.
86 f. : il. ; 30 cm.

Dissertação (mestrado) – Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Engenharia de Produção e Sistemas, 2024.

“Orientadora: Prof. Dr. Maria Isabel Motta Morandi
Co-Orientador: Prof. Dr. Daniel Pacheco Lacerda”.

1. Dimensionamento de estoques. 2. Clusterização. 3. Cadeia de suprimentos. 4. Gestão logística. 5. Eficiência operacional.
I. Título.

CDU 658.5

Dados Internacionais de Catalogação na Publicação (CIP)
(Bibliotecária: Amanda Schuster – CRB 10/2517)

FREDERICO ERNESTO EIDELWEIN

**DIMENSIONAMENTO DE ESTOQUES NA CADEIA DE SUPRIMENTOS A PARTIR
DA CLUSTERIZAÇÃO**

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Engenharia de Produção e Sistemas pelo Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Universidade do Vale do Rio dos Sinos (UNISINOS).

Aprovado em 24 de Abril de 2024

BANCA EXAMINADORA

Fabio Antonio Sartori Piran – Unisinos

Rafael Kunst – Unisinos

Ricardo A. Cassel – UFRGS

Rafael Teixeira – College of charleston

AGRADECIMENTOS À CAPES

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

AGRADECIMENTOS

Gostaria de expressar minha sincera gratidão a todas as pessoas e instituições que contribuíram para a realização desta dissertação. Sem o apoio e a colaboração de cada um de vocês, este trabalho não teria sido possível.

Aos meus pais, Paulo e Doris, que sempre me apoiaram e incentivaram de todas as formas em meu caminho acadêmico e pessoal, suas palavras de encorajamento foram fundamentais para minha perseverança.

À minha noiva Larissa, que esteve ao meu lado em todos os momentos desses últimos 2 anos, e me apoiou incansavelmente. Sua presença foi um pilar fundamental durante essa jornada.

À minha orientadora, Prof. Dr. Maria Isabel Motta Morandi, e ao co-orientador, Prof. Dr. Daniel Pacheco Lacerda, pela orientação, paciência e valiosos insights. Suas sugestões e correções foram essenciais para o desenvolvimento deste trabalho. Aos demais professores, pelas correções e ensinamentos que me permitiram apresentar um melhor desempenho no meu processo de formação profissional.

Aos meus colegas de curso, pelo companheirismo e troca de experiências que me permitiram crescer não só como pessoa, mas também como formando.

À empresa onde este trabalho foi desenvolvido e aos meus colegas de trabalho, que se disponibilizaram para todas as minhas dúvidas e entrevistas necessárias. Sem eles, não seria possível a conclusão desta dissertação e meu crescimento como pessoa e profissional.

Aos amigos e familiares que me apoiaram e contribuíram para a realização deste trabalho, meu sincero agradecimento. Obrigado por compreenderem minha ausência enquanto eu me dedicava à pesquisa. Aos amigos, pela amizade incondicional e apoio demonstrado ao longo do tempo, em especial aos grupos M.F.C., Rotaract Club de Campo Bom, OFG.

RESUMO

O desenvolvimento de modelos utilizando dados disponíveis para as empresas é uma necessidade para manter que se mantenham competitivas no mercado global. Este trabalho aborda o dimensionamento e posicionamento de estoques na cadeia de suprimentos, utilizando a clusterização de produtos por suas características categóricas e histórico de vendas como base para projetar a alocação de novas coleções. Foi realizada uma consulta com especialistas para validação das variáveis disponíveis, atribuídas para produtos relacionados a coleções passadas, e coleção futura. Os produtos das coleções passadas foram então agrupados utilizando o algoritmo *k-modes* e seus desempenhos foram medidos para cada origem preferencial segundo o destino do pedido, conforme indicação de parceiro logístico. Utilizando os mesmos critérios, os produtos que compõe a coleção futura foram associados aos grupos criados, e com base no desempenho passado medido, seus volumes foram projetados para cada armazém. O estudo compara o resultado de três diferentes cenários, executado, potencial e do modelo. A pesquisa alcançou seus objetivos ao construir um modelo de agrupamento de produtos para coleções futuras, avaliar modelos de aprendizado de máquina não supervisionados aderentes ao problema, avaliar os impactos econômicos do modelo atual do contexto analisado, e avaliar os benefícios resultantes do emprego de modelos de aprendizagem de máquina não supervisionados no processo de posicionamento e dimensionamento de estoques. O modelo proposto define critérios para alocação de produtos e sugere a redução de custos operacionais e melhoria no serviço de logística, permitindo identificar fatores ocultos, como possível aumento no nível de serviço e satisfação do cliente.

Palavras-chave: dimensionamento de estoques; clusterização; cadeia de suprimentos; gestão logística; eficiência operacional.

ABSTRACT

The development of models using data available to companies is a necessity in order to remain competitive in the global market. This work addresses the sizing and positioning of inventories in the supply chain, using the clustering of products by their categorical characteristics and sales history as a basis for projecting the allocation of new collections. Experts were consulted to validate the available variables, assigned to products related to past collections and future collections. Products from past collections were then clustered using the k-modes algorithm and their performance was measured for each preferred origin according to the destination of the order, as indicated by the logistics partner. Using the same criteria, the products that compose the future collection were associated with the clusters created, and based on the measured past performance, their volumes were projected for each warehouse. The study compares the results of three different scenarios, executed, potential and model. The research achieved its objectives by creating a product clustering model for future collections, identifying relevant variables in the process and assessing the model's adherence to inventory sizing. The proposed model defines criteria for allocating products and suggests reducing operating costs and improving the logistics service, allowing hidden factors to be identified, such as a possible increase in the level of service and customer satisfaction.

Key-words: inventory sizing; clustering; supply chain; logistics management; operational efficiency.

SUMÁRIO

1 INTRODUÇÃO	10
1.1 Objeto e Problema de Pesquisa	11
1.2 Objetivo	14
1.2.1 Objetivo geral	15
1.2.2 Objetivos específicos.....	15
1.3 Justificativa	15
1.4 Delimitações	25
1.5 Estrutura do trabalho	25
2 FUNDAMENTAÇÃO TEÓRICA	27
2.1 Moda rápida e varejo	27
2.2 Cadeia de suprimento	29
2.2.1 Previsão de demanda.....	29
2.2.2 Posicionamento de estoque	33
2.3 Aprendizado de máquina e clusterização	34
2.3.1 K-modes	36
3 PROCEDIMENTOS METODOLÓGICOS	40
3.1 Delineamento da pesquisa	40
3.4 Método de trabalho	42
3.5 Coleta de dados	46
3.6 Análise de dados	48
4 ANÁLISE DOS RESULTADOS	51
4.1 Clusterização	51
4.1.1 Análise de Variáveis	51
4.1.2 Resultado de clusterização	53
4.2 Resultado da análise de distribuição de clusters	57
4.3 Resultado comparativo entre cenários	59
4.4 Avaliação de custo e tempo de trânsito	66
5 DISCUSSÃO DOS RESULTADOS	68
6 CONSIDERAÇÕES FINAIS	72
REFERÊNCIAS	75
APÊNDICE A – PROTOCOLO DA REVISÃO SISTEMÁTICA DA LITERATURA ...	81
APÊNDICE B – DICIONÁRIO DE DADOS	83

ANEXO A – RELAÇÃO DE CUSTO E TEMPO DE TRÂNSITO POR CENTRO LOGÍSTICO E ESTADO85

1 INTRODUÇÃO

No ano de 2022, cerca de 24 bilhões de pares de calçados foram produzidos no mundo, sendo que dos 5 principais países produtores, 4 estão situados no continente Asiático. O Brasil é o único país desta lista localizado na América do Sul. A produção de calçados brasileiros no mesmo ano foi de 886 milhões de pares, representando 3,7% da produção mundial, dos quais 142 milhões destinados à exportação. (ABICALÇADOS, 2024).

Segundo a Associação Brasileira das indústrias de Calçados - Abicalçados (2024), o Rio Grande do Sul é o segundo maior estado produtor do artigo no Brasil, com 192 milhões de pares produzidos em 2022. Este volume representa 21,7% da produção nacional, após o Ceará com 205 milhões. No entanto, o estado do Rio Grande do Sul é o maior exportador nacional de calçados, com um volume de 42,8 milhões de pares em 2022 (30% das exportações nacionais no ano), e com crescimento superior a 30% em comparação ao ano anterior, o que evidencia a importância do setor para a região.

A partir da década de 90, a produção de calçado seguiu a tendência de migração para os países asiáticos, em especial para a China, buscando maiores volumes de produção e preços mais competitivos (NEVES, 2018). Para se manter no mercado, as empresas que não fecharam ou migraram sua produção, readequaram suas estratégias de produção (NEVES, 2018).

A competitividade e a globalização forçaram as empresas a reverem e evoluírem em diversas funções organizacionais. Por consequência, a cadeia de suprimento deve ser vista pelas empresas como uma rede trabalhando com um mesmo propósito, para aumentar o valor para o seu cliente por meio da distribuição eficiente dos produtos (BATALHA; SILVA, 1999). Na indústria calçadista, a existência de elevado volume de componentes, e por consequência, partes envolvidas na cadeia, exige relação eficiente entre as empresas, que buscam desenvolver a rápida comunicação entre os elos com uma visão integrada da cadeia, trabalhando alinhadamente para entregar o melhor nível de serviço aos consumidores (NUNES et al., 2023).

A gestão da cadeia de suprimentos enfrenta desafios como giro reduzido de estoque e, conseqüentemente, elevado investimento, perda de venda por falta de produto em determinados locais e excesso do mesmo em outros, obsolescência do

estoque existente, altos custos de transporte etc. (COX III; SCHLEIER JR, 2013). Estes efeitos são, ainda, mais relevantes na indústria da moda, em que as tendências se transformam com uma frequência elevada. Desta forma setores como têxtil e calçadista buscam ter suas estratégias de planejamento e logística alinhadas e o mais acuradas possíveis, evitando despesas extras com envio, armazenagem e saldos de estações passadas.

No contexto específico deste estudo, os produtos são fabricados no Brasil e enviados aos Estados Unidos, acrescentando cerca de 2 semanas de tempo de trânsito ao lead time da cadeia de suprimentos. A atual estrutura conta com três armazéns no país de recebimento, que pertencem a uma subcontratada, e podem ser definidos como destino da mercadoria até o momento em que são despachados da fábrica, pois uma vez alocados em determinado centro logístico, o custo de movimentação entre armazéns seria elevado.

A extensão territorial dos Estados Unidos é ampla e, por consequência, aumenta consideravelmente o tempo de envio dos produtos após a compra online, fator levado em consideração pelos clientes no momento da compra. Existe também a possibilidade de compra de mais que um produto, e quando estes não estão localizados em um mesmo armazém, geram custos adicionais de movimentação e envio, além de desconforto ao consumidor quando recebe mais de um pacote.

Neste contexto a presente pesquisa tem como tema o posicionamento e dimensionamento dos estoques na rede logística existente. Nos capítulos seguintes deste trabalho serão apresentadas a fundamentação teórica para o trabalho, o método aplicado, assim como os resultados, discussões e conclusões dos autores em relação ao tema abordado.

1.1 Objeto e Problema de Pesquisa

A indústria de moda e calçadista apresenta desafios específicos em relação à gestão de estoques e posicionamento de produtos, devido à ampla variedade de modelos e tamanhos de calçados, à sazonalidade das demandas, ciclo de vida curto dos produtos e a necessidade de rápida adaptação às tendências de moda (STUKER, 2014). Os custos relacionados ao estoque correspondem de 15 a 20% do custo total dos produtos acabados, sendo que as empresas nos Estados Unidos

gastam aproximadamente US\$ 14 bilhões por ano com depreciação, obsolescência, impostos, seguros e armazenamento (TSAO et al., 2012).

A globalização da rede de ofertas e demanda agrega complexidade ao planejamento, tornando o desafio de previsão de demanda significativo e suscetível ao efeito chicote, levando a elevados estoques, ineficiências e falta de produtos. Para que as empresas varejistas se mantenham competitivas, a tomada de decisão deve ser orientada por dados para análise de demanda e oferta no planejamento operacional, aumentando o nível de serviço oferecido (GOPAL et al., 2022; PEREIRA; FRAZZON, 2021).

Para apoiar este processo de tomada de decisão baseado em dados, deve-se ampliar a pesquisa aplicada às áreas consideradas estratégicas para a empresa envolvendo dados para a geração de informação. Embora o estudo da capacidade da análise de dados tenha ganhado ênfase nos últimos anos, ainda existe carência no desenvolvimento de pesquisas relacionadas ao tema em operações da cadeia de suprimento e seus contextos (FOSSO WAMBA et al., 2018).

O acúmulo de dados aumenta exponencialmente, uma vez que mais fontes são consideradas, volume de transações e produtos cresce, e a quantidade de tecnologias de coleta são empregadas. Com isso, desafios como sanitização dos dados, privacidade, segurança da organização, armazenamento e acima de tudo capacidade analítica devem ser superados para o desenvolvimento da tomada de decisão eficiente na cadeia de suprimento e aumento de lucros na organização (GOPAL et al., 2022).

Estes desafios se traduzem em necessidade de adaptação por parte das empresas. Para isso é necessário investimento em capacitação para o desenvolvimento de pessoas qualificadas com habilidade de análise de dados (FOSSO WAMBA et al., 2018). As competências dos gestores na escolha das informações são também relacionadas como fator de influência no resultado das empresas (RAZAGHI; SHOKOUHYAR, 2021).

Há ainda preocupação quanto à qualidade do material coletado e da velocidade com que estes podem ser processados e transformados em informação, antes que se tornem obsoletos (FOSSO WAMBA et al., 2018). Para isso, mais que qualificação, existe a necessidade do interesse em desenvolvimento de estrutura e emprego de novas tecnologias, necessárias para a análise e manutenção de amplo volume de dados (GOPAL et al., 2022).

Outra importante característica da área de análise de dados, é a construção de informação utilizando diferentes fontes e de diferentes naturezas (JAHANI; JAIN; IVANOV, 2023). O setor de varejo de moda, mesmo apoiados em dados históricos, utiliza em sua maioria critérios qualitativos para traçar suas estratégias de dimensionamento de estoque e sortimento de produtos, uma vez que esta indústria tem significativas mudanças nas características de consumo com o passar do tempo (FISHER, 2009).

Existem oportunidades para a exploração dos dados coletados pelo varejo em seus diversos canais, principalmente online, e que, por sua vez, podem ser utilizados não somente para a previsão de demanda, mas para o posicionamento de seus produtos baseados em suas características e cruzadas com o perfil de consumo de determinadas regiões. Mesmo com sua importância identificada, parte da literatura ainda ignora fontes de dados não estruturados, perdendo potencial para suportar a tomada de decisão (JAHANI; JAIN; IVANOV, 2023).

Há questões importantes relacionadas ao *fast fashion* (moda rápida), que se caracteriza por mudanças rápidas e constantes para acompanhar as tendências atuais, reduzido volume de investimento em quantidades compradas e o curto ciclo de vida destes produtos (CHOI et al., 2014). Neste sentido, uma das principais questões a serem buscadas neste ramo é a redução dos tempos de produção (*lead-time*) e entrega (*delivery-time*), uma vez que estas são questões decisivas para o consumidor.

Em países de amplo território, como é o caso dos Estados Unidos, onde a operação deste estudo é desenvolvida, é observado outro fator importante, os perfis dos consumidores apresentam características distintas entre regiões, e há diferença do tempo de trânsito significativa dos diferentes centros logísticos até o cliente. Deste modo se faz importante que o sortimento de produtos seja o mais preciso possível, garantindo um nível de serviço satisfatório ao cliente, maior taxa de retorno, índice de satisfação e economia em taxas e despesas de operação logística. Atualmente o custo relacionado a logística representa entre 16% e 20% do faturamento, necessário para sustentar o negócio, porém com elevado potencial de redução por meio da racionalização dos processos e melhor planejamento de distribuição.

O tempo de entrega é também um fator cada vez mais importante para o consumidor. No mercado eletrônico americano, que em 2020 ultrapassou \$432

bilhões e projeta-se \$549 bilhões de faturamento em 2024, 56% dos clientes espera a entrega de seus produtos em até 1 dia, e destes, mais da metade declara esperar a entrega de 1 a 3 horas (KIM et al., 2023). No entanto, para suprir estas expectativas, é necessário apresentar uma rede de distribuição descentralizada, o que tradicionalmente se mostra inviável por questões relacionadas a investimento.

A rede que compõe a cadeia de suprimento é comumente ampla, sendo formada por empresas fornecedoras a jusante e a montante da empresa focal, podendo fornecer insumos à produção ou serviços, como beneficiamento e logística. A análise das informações ocultas no elevado volume de dados disponível para as empresas auxilia a interpretar o comportamento do consumidor, resultando melhorias de toda a rede de empresas e tornando-as mais lucrativas (FOSSO WAMBA et al., 2018; MAHESHWARI; GAUTAM; JAGGI, 2021).

No entanto, mesmo com o elevado potencial e viabilidade para o estudo dos dados e informações, oportunizando maiores lucros e novos negócios a empresa, é necessário que exista cooperação entre as empresas. Para o desenvolvimento desta estrutura, é necessário que exista compartilhamento de dados entre as partes envolvidas de maneira colaborativa, fator que se torna uma barreira em diversas situações para o desenvolvimento de informações complementares e de qualidade, cenário ideal de dados (JAHANI; JAIN; IVANOV, 2023; MAHESHWARI; GAUTAM; JAGGI, 2021).

Assim, existe uma busca constante, em um ambiente competitivo como o comércio eletrônico, para aproximar os produtos dos consumidores e regular ao máximo o dimensionamento dos estoques do consumo final da coleção, resultando em um giro de produtos que minimize o volume de descontos ao final dos períodos de venda, ampliando ganhos ao longo de toda a cadeia de suprimento. Este contexto desperta a seguinte questão de pesquisa: **como identificar padrões de agrupamento dos produtos para dimensionar os estoques na cadeia de suprimentos?**

1.2 Objetivo

Nesta seção são expostos o objetivo geral do trabalho, e os objetivos específicos, os quais guiarão esta pesquisa.

1.2.1 Objetivo geral

O objetivo geral do estudo é construir um modelo para dimensionamento de estoques, utilizando aprendizado de máquina, em uma empresa no contexto *fast fashion*.

1.2.2 Objetivos específicos

Os objetivos específicos para este trabalho são:

1. Avaliar os modelos de aprendizagem de máquina não supervisionada aderentes ao problema de dimensionamento de estoques na cadeia *fast fashion*;
2. Avaliar o impacto econômico do atual modelo de dimensionamento e posicionamento de estoques;
3. Avaliar os benefícios da utilização de modelos de aprendizagem de máquina não supervisionado para o posicionamento e dimensionamento de estoque no contexto *fast fashion*.

No tópico seguinte serão discutidas as questões que justificam a produção desta dissertação, tanto do ponto acadêmico quanto empresarial.

1.3 Justificativa

É importante que a pesquisa aponte as lacunas existentes na literatura para que fique clara sua relevância para o meio acadêmico e empresarial. Nesta seção são expostas as questões que justificam a realização deste estudo.

Para o embasamento teórico deste trabalho foi feita a Revisão Sistemática da Literatura (RSL) conforme proposto por Morandi e Camargo (2015) que contribuiu para a busca, análise e inclusão dos resultados relevantes para o tema central. Segundo Ermel et al (2022), o mapeamento da pesquisa existente auxilia o pesquisador a encontrar lacunas de conhecimento e a sua falta pode resultar em pesquisas sem relevância acadêmica. As etapas para a revisão sistemática realizada são apresentadas na Figura 1.

Figura 1 - Etapas para a Condução da RSL



Fonte: Morandi e Camargo (2015, p. 146)

Conforme mostrado na Figura 1, inicialmente deve-se definir a questão central do trabalho e elaboração do framework conceitual (MORANDI; CAMARGO, 2015). Para este trabalho, o tema central é o dimensionamento e posicionamento de estoque na cadeia de suprimento de uma empresa calçadista com operação no exterior pelo canal de e-commerce. Para Morandi e Camargo (2015), é necessária para a construção do framework, a definição do tipo de revisão sistemática, que neste estudo definiu-se como configurativa, em que por meio de procedimentos qualitativos e quantitativos buscam responder a uma questão de pesquisa ampla e construir teorias, abordando a revisão de estudo heterogêneos.

No segundo passo, Morandi e Camargo (2015) indicam a escolha da equipe de trabalho. Porém por esta se tratar de uma monografia, escrita por uma pessoa, considerou-se esta etapa como inválida.

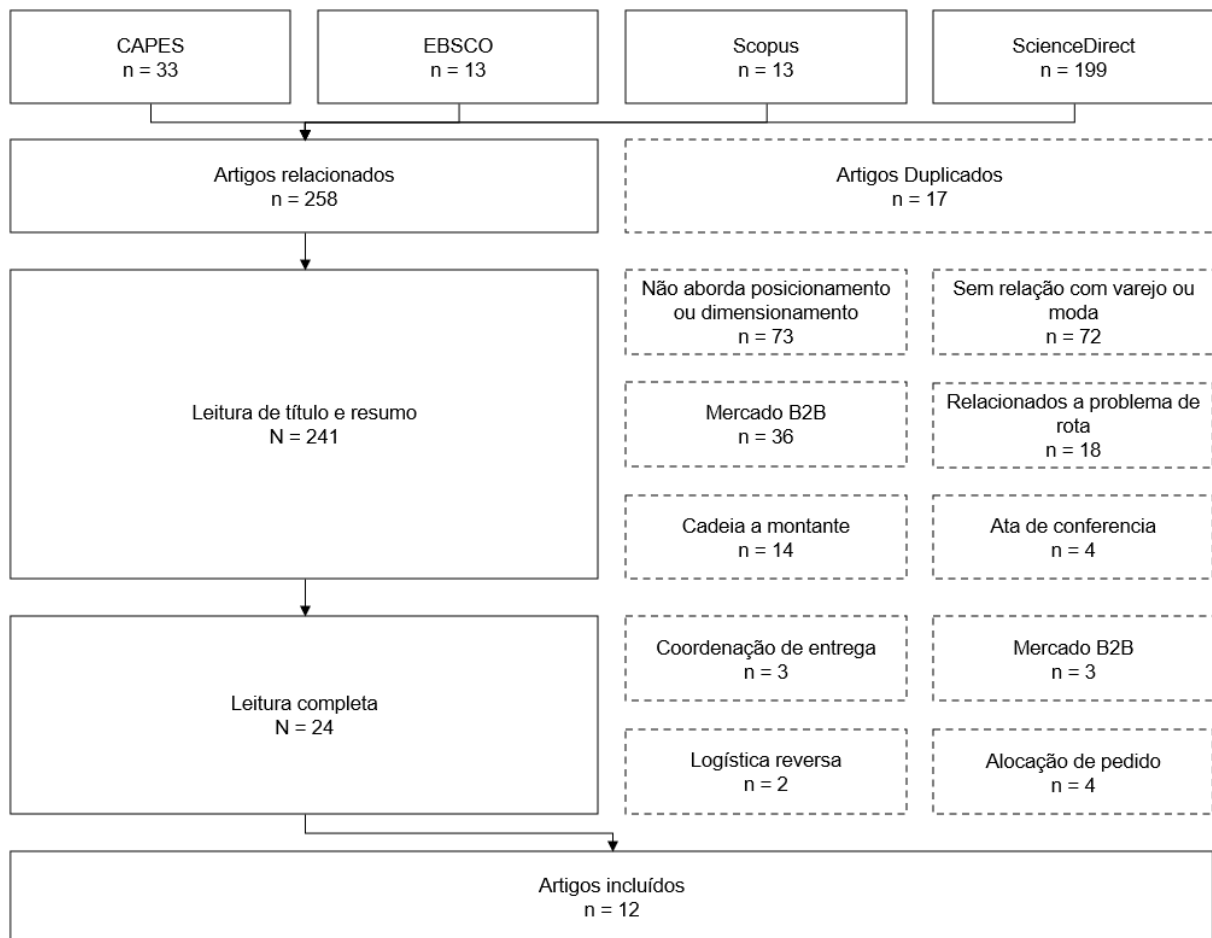
O terceiro passo é a estratégia de busca, no qual será especificado o que, onde e como será buscado (MORANDI; CAMARGO, 2015). Os termos de busca, os critérios de inclusão e exclusão e as fontes da busca estão especificadas no protocolo exposto no Apêndice A deste trabalho. Para esta pesquisa buscou-se artigos, monografias, dissertações, teses e livros com relação a cadeia de suprimento e *data analytics* em um contexto de varejo. As principais bases de dados utilizadas nesta pesquisa para embasar e justificar academicamente o estudo foram EBSCOHOST, CAPES, Scopus e Science Direct.

Após definir as bases de dados e realizar a pesquisa com as palavras chaves selecionadas, filtrou-se o idioma da pesquisa em inglês e português, respectivamente por sua relevância internacional, e por ser a língua do país onde o

trabalho foi desenvolvido, conforme apresentado no protocolo presente no Apêndice A. Não foram feitos filtros de ano de publicação nas bases pois entendeu-se que contribuições importantes para este poderiam ser excluídas.

Com a definição das bases de busca de bibliografia, a pesquisa foi elaborada conforme a quarta etapa da condução da RSL, utilizando publicações internacionais sem definição de ano de publicação. Na Figura 2 são apresentados os resultados.

Figura 2 – Fluxograma de resultados da seleção de trabalhos



FONTE: Elaborado pelo autor.

Após a RSL e análise dos trabalhos selecionados, pode ser evidenciado a relevância desta pesquisa para o meio acadêmico, uma vez que não foram identificados trabalhos que utilizem características dos produtos e perfil regional de compra para agrupamento o posicionamento dos estoques. A maior parte da atenção para o desenvolvimento de estudos na área de *Data Analytics* e *Supply Chain* está voltada para decisões de médio prazo, como alocação de pedido e gerenciamento de risco de suprimento, contudo, mesmo sendo uma das decisões

mais importantes, o planejamento de dimensionamento de remessas não recebeu atenção suficiente da pesquisa, sendo crucial para este tipo de decisão a utilização de feedback do consumidor (JAHANI; JAIN; IVANOV, 2023).

Na Tabela 1 são apresentadas as lacunas e oportunidades identificadas na literatura resultante do processo de revisão sistemática. Com isso a presente pesquisa se justifica por buscar preencher parte da carência de conhecimento identificado, como poucos estudos tratando do uso da análise de dados aplicado a casos reais na cadeia de suprimento, não somente discutindo tecnologias e propondo modelos, e a utilização das informações históricas capturadas das transações e contato do consumidor com o canal de venda para a melhor distribuição dos produtos pela cadeia de suprimento.

Tabela 1 – Resultado da RSL e suas lacunas.

Nº	Trabalho	Autor / Ano	Proposta	Limitações e sugestão trabalhos futuros	Lacuna
1	Capacitated location model with online demand pooling in a multi-channel supply chain	(LIU; ZHOU; ZHANG, 2010)	Apresenta um modelo de atribuição de demanda à armazéns regionais para atender pedidos online em uma cadeia multicanal, considerando riscos e custo de transporte do sistema de logística e estoque com o objetivo de reduzir os custos gerais do sistema de atribuição de demanda online.	É indicado para trabalhos futuros a formulação de restrições para o sistema e analisar se o modelo ainda se mantém aplicável e vantajoso, além disso, estudar cenários em que se aumenta o sortimento de produtos e a capacidade dos armazéns.	Não utiliza dos dados provenientes dos canais para a disposição dos produtos nos armazéns regionais em que a demanda é alocada.
2	A continuous approximation approach for the integrated facility-inventory allocation problem	(TSAO et al., 2012)	O trabalho resolve um problema de localização de instalações e alocação de estoques em uma rede de centros de distribuição e varejistas.	Sugere-se para trabalhos futuros o estudo de modelo de posicionamento com tomadores de decisão descentralizados.	Não propõe o sortimento dos produtos a serem alocados nos centros de distribuição, atribuindo consumidores a cada centro desprezando sua diferenciação de consumo.
3	Modeling logistics service providers in a non-cooperative supply chain	(SANTIBANEZ-GONZALEZ; DIABAT, 2016)	Apresenta um modelo matemático para o projeto de uma cadeia de suprimentos não cooperativa formada por quatro camadas composta por fabricantes, varejistas, clientes e provedores de serviços logísticos.	Estudar a percepção dos consumidores quanto ao nível de serviço dos provedores de serviços logísticos.	Utiliza o custo do serviço para avaliação do sistema com provedor logístico.

Nº	Trabalho	Autor / Ano	Proposta	Limitações e sugestão trabalhos futuros	Lacuna
4	Fuzzy inference systems and inventory allocation decisions: Exploring the impact of priority rules on total costs and service levels	(WANKE et al., 2017)	É proposto um modelo que, a partir do uso da lógica fuzzy, para apoio a decisão de nível e posicionamento de estoque considerando as variáveis de custo de estocagem e transporte, escassez e pedidos.	Pesquisas futuras são sugeridas para fundamentar a pesquisa utilizando dados reais e em maior volume, possibilitando a comparação entre o modelo proposto no artigo com outros modelos probabilísticos. A comparação entre o efeito do modelo nas variáveis de entrada e saída também devem ser avaliadas.	O sistema leva somente custos e nível de serviço em consideração para a definição do modelo.
5	Big data analytics in supply chain management between 2010 and 2016: Insights to industries	(TIWARI; WEE; DARYANTO, 2018)	O artigo analisa as produções bibliográficas publicadas entre os anos de 2010 e 2016 sobre o gerenciamento da cadeia de suprimento e big data, discutindo seus aspectos individuais e fornecendo insights para a indústria e varejo.	Os aspectos estudados devem ser analisados com dados reais para validar as proposições feitas no artigo.	A pesquisa não aplica ou estuda aplicações com dados e resultados reais, limitando-se a uma revisão da literatura existente.
6	Customer reviews for demand distribution and sales nowcasting: a big data approach	(SEE-TO; NGAI, 2018)	Utilização dos dados de avaliação de compras online para auxílio na previsão de demanda e gerenciamento de estoque.	Sugere-se para novas pesquisas a utilização de algoritmos, aprendizado de máquina, redes neurais e outros para previsão de demanda com base nos dados fornecidos pelo canal de venda. Devem ser usados dados de diferentes períodos para validação do modelo.	Mais dados referentes a trajetória do cliente pelo canal podem ser utilizados como fonte para análise.

Nº	Trabalho	Autor / Ano	Proposta	Limitações e sugestão trabalhos futuros	Lacuna
7	Autonomous Inventory and Capacity Management in an Omnichannel Retailing Scenario: A Review	(LINHARES; MACHADO, 2020)	A pesquisa busca entender como a cadeia de suprimento gerencia estoques e capacidades utilizando algoritmos de aprendizado de máquina para alcançar os objetivos do modelo omnicanal.	Indica-se para futuras pesquisas a aplicação do modelo em casos reais para comparar o desempenho do gerenciamento tradicional da cadeia de suprimento com as soluções de automação, e propor novos modelos integrando inteligência artificial.	Não aplica em modelos reais.
8	A data-driven approach to adaptive synchronization of demand and supply in omni-channel retail supply chains	(PEREIRA; FRAZZON, 2021)	O artigo busca por meio de dados criar uma abordagem para combinar previsão de demanda e aprendizado de máquina para otimizar a demanda da cadeia de suprimento em um ambiente de varejo omnicanal.	Devido à confidencialidade dos dados, informações relacionadas à custos e vendas online foram expostas genericamente. A inclusão de fatores de marketing, como preço e promoção, no aprendizado de máquina, para melhorar a precisão da previsão de demanda podem ser feitas.	Utiliza somente os dados de venda histórica para a previsão de demanda, ignorando dados de comportamento capturados dos canais e atributos dos produtos.
9	Challenges at the marketing–operations interface in omni-channel retail environments	(BIJMOLT et al., 2021)	Desenvolvimento de decisão quanto ao sortimento e estoque, distribuição e entrega e devoluções nas áreas de marketing e operações em uma estrutura de varejo omnicanal.	Estudar as dimensões abordadas pela pesquisa, integrando operações e marketing com o objetivo de maximizar ganhos para a empresa.	Não são abordados casos reais, a pesquisa apenas expõe a literatura existente em cada aspecto de seu escopo e aponta direções para pesquisas futuras.
10	Fast-moving consumer goods network design with pricing policy in an uncertain environment with correlated demands	(NASIRI; KALANTARI; KARIMI, 2021)	Neste artigo, é desenvolvido um modelo para determinar o preço ideal em diferentes canais de venda e distribuição, levando em consideração localização, alocação e controle de estoque de centros de distribuição.	Recomenda-se incluir à análise custos de manutenção de estoque, transporte, bem como custo da escassez de produto pela probabilidade de demanda.	Não utiliza para a definição de dimensionamento ou sortimento nos centros logísticos dados coletados em canais de venda.

Nº	Trabalho	Autor / Ano	Proposta	Limitações e sugestão trabalhos futuros	Lacuna
11	Impact of big data analytics on supply chain performance: an analysis of influencing factors	(GOPAL et al., 2022)	O artigo tem como proposta estudar o impacto dos diferentes métodos de aplicação de big data na cadeia de suprimento.	As sugestões estão voltadas no desenvolvimento do modelo, utilizando mais especialistas para a interpretação dos dados, e refino dos dados coletados.	O modelo não é testado em casos reais.
12	Network inventory deployment for responsive fulfillment	(KIM et al., 2023)	Compara o desempenho de uma rede de distribuição centralizada em um armazém, com uma rede descentralizada para atendimento de demanda por tempo de trânsito curto.	Estudar o desempenho do armazém por um período para determinar sua real eficiência, realizar pesquisa de custos de envio mais rápido, considerado fixo no estudo e independente de distância.	Estudar a alocação de sortimento dos produtos na rede descentralizada para viabilizar a operação.

FONTE: Elaborado pelo autor.

A partir das lacunas relacionadas, este trabalho pretende preencher lacunas relacionadas ao uso de atributos e características dos produtos vinculados ao desempenho histórico para criar um modelo de alocação e sortimento de centros logísticos. O estudo ainda aspira simular a aplicação do modelo em dados reais, esperando comparar seus resultados para medir sua aplicação.

Assim, como este estudo busca agregar conhecimento ainda não documentado na academia, é importante entregar valor à empresa na qual a operação é estudada. Neste caso a pesquisa foi inicialmente estimulada pela necessidade de desenvolvimento de um processo lógico e robusto para a disposição dos produtos nos armazéns logísticos.

A busca por um modelo para determinação do sortimento dos produtos e do volume a ser alocado em cada armazém parte da premissa de necessidade de melhor previsão do padrão de demanda futura. A partir da análise do comportamento do cliente, esta informação pode servir como meio para a melhoria das operações ligadas a cadeia de suprimento, uma vez que informações suficientes sejam coletadas, possibilita a compreensão de movimentos de demanda futura, sendo este um dos principais papéis da análise de dados na cadeia de suprimento (MAHESHWARI; GAUTAM; JAGGI, 2021).

Empresas líderes em seus segmentos, utilizam da análise do comportamento de seu consumidor em seus canais de venda para melhoria de suas operações. Exemplos são a Amazon e a Flipkart, empresas de comércio de produtos que utilizam dados de cliques, visualização e compras anteriores para auxiliar sua tomada de decisão e traçar padrões de comportamento.

A administração dos recursos relacionados à cadeia de suprimentos desempenha um papel crucial nos resultados de uma empresa, tanto de maneira negativa, sendo vista como um custo para o sistema, quanto de maneira positiva, encarada como uma vantagem competitiva da companhia em um ambiente de alta concorrência. Este ambiente, portanto, cria oportunidades para as organizações, que podem desenvolver e aprimorar sua capacidade de análise para obter melhor compreensão de seus valores, criando produtos, serviços e melhorando o seu nível de atendimento ao cliente (TIWARI; WEE; DARYANTO, 2018).

Dentre as dimensões que o posicionamento correto dos produtos pode auxiliar o resultado da empresa, está a oportunidade de redução de tempo de envio do produto para o cliente, questão cada vez mais importante no varejo. Existe no

caso estudado custos extras com envio em ocasiões em que a empresa tem mais de um produto a ser enviado para um mesmo pedido, neste caso há a possibilidade de os dois produtos não estarem disponíveis em um mesmo armazém, o que aumenta o custo de envio, pois dois valores de frete e movimentação serão gerados.

Uma das dificuldades mais significativas apontadas na literatura é a comunicação e compartilhamento de dados entre as empresas que compõem a cadeia de suprimento, que comumente é fragmentada, e para que a cadeia de suprimento seja eficiente, as partes precisam estar horizontalmente conectadas (JAHANI; JAIN; IVANOV, 2023; MAHESHWARI; GAUTAM; JAGGI, 2021). O caso estudado proporciona uma oportunidade de estudar um cenário próximo ao ideal em questões de posse dos dados necessários para a compreensão da cadeia, pois controla a maior parte do processo, desde sua produção, controle de estoques, distribuição e comércio dos produtos.

Além das questões relacionadas à falta de um critério estruturado para posicionamento dos produtos nos diferentes armazéns, o dimensionamento do volume de determinado produto nos armazéns é importante para sua disponibilidade no momento da compra, resultando na possível perda de uma venda. A venda perdida ainda pode ter outros impactos difíceis de serem mensurados, uma vez que o cliente pode entender isso como uma má experiência junto a empresa ou adquirir um produto similar de um concorrente.

A análise dos dados disponíveis, gerados de diferentes fontes pode oportunizar o aumento da efetividade da operação e lucro das empresas, aumentando a disponibilidade do produto para os clientes, uma vez que melhora a compreensão de movimentos futuros do mercado (FOSSO WAMBA et al., 2018; RAZAGHI; SHOKOUHYAR, 2021). Levar em consideração o comportamento do consumidor junto aos produtos se torna cada vez mais relevante, sendo sua capacidade de previsão fortemente ligada a inovação de produtos, fabricação e distribuição de demanda.

Em geral existe uma subutilização do amplo volume de dados gerados a respeito do negócio que poderia estar sendo trabalhado e transformado em força competitiva para a empresa. Evidencia-se a justificativa desta pesquisa por ampliar os estudos realizados analisando um caso real de uma empresa em um dos maiores e mais competitivos mercados consumidores do mundo, criando um modelo para utilizar os dados disponíveis para a empresa para determinar a quantidade e

sortimento dos produtos em seus armazéns, que se traduz em melhor experiência para o cliente, e por sua vez resultados positivos para a companhia.

1.4 Delimitações

Este trabalho se limitará a analisar o contexto atual da empresa, levando em consideração as variáveis relevantes para a criação do modelo que se propõe a representar de maneira simplificada a realidade. Desta maneira, a aplicação e avaliação de ganhos em um ambiente real não será testada, mas validada com os dados históricos fornecidos pela empresa em que este estudo se baseia.

Os dados coletados ainda considerarão somente o canal de venda online, não considerando vendas ao varejista terceiro ou destinadas aos pontos físicos próprios de venda, os quais seguem diretrizes diferentes previsão e análise, ou tem volumes reduzidos em comparação ao e-commerce. A modalidade de atacado também não será abordada neste modelo.

Quanto a origem do envio dos produtos enviados aos consumidores, não são abordados cenários em que novos centros logísticos são inseridos a cadeia. Lojas físicas não são consideradas para compor a estrutura de logística.

Este trabalho utilizará também como padrão as origens e destinos correspondentes indicadas por um fornecedor logístico parceiro. Sendo assim, novos estudos para determinar o centro logístico preferencial para cada área não serão realizados.

A divisão de tamanhos de calçado, produto comercializado pela empresa na qual este trabalho foi realizado, não será levada em consideração para a disponibilidade do produto, a fim de simplificar a análise dos dados. Uma vez que a distribuição do volume produzido acompanha a curva de consumo dos tamanhos, que seguem um comportamento padrão ao longo do tempo, entende-se que não é necessária a reavaliação no modelo.

1.5 Estrutura do trabalho

Este trabalho está organizado em seis capítulos, sendo o primeiro apresentando o contexto no qual foi desenvolvido e as questões norteadoras para a produção da pesquisa. Em seguida o referencial teórico, em que este estudo está

sustentado, é evidenciado, abordando os principais conceitos com relação ao objeto de estudo.

No terceiro capítulo o método utilizado para o desenvolvimento desta pesquisa é descrito detalhadamente, dando credibilidade a pesquisa. Este se divide em delineamento de pesquisa, método científico, de pesquisa e trabalho, coleta e análise de dados.

O quarto capítulo é composto pela análise dos resultados seguindo o método proposto no capítulo anterior, no qual são expostos os cenários resultantes e suas comparações de desempenho. No capítulo cinco os resultados são discutidos com a literatura explorada, e por fim, no sexto capítulo, as considerações finais do trabalho são expostas.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, apresenta-se a fundamentação teórica que sustenta esta pesquisa, abordando os tópicos: cadeia de suprimento, previsão de demanda, posicionamento de estoque, moda rápida e varejo, aprendizado de máquina e *clusterização*. Estes tópicos serão discutidos a partir da revisão da literatura, destacando os principais conceitos, modelos e aplicações relacionados ao tema da pesquisa.

2.1 Moda rápida e varejo

O conceito de varejo se caracteriza pela atividade de negócio que vende diretamente para o consumidor final, podendo a empresa ser fabricante, atacadista ou o próprio comerciante, sendo indiferente ao meio pelo qual o produto é vendido (STERN; EL-ANSARI; COUGHLAN, 1996). O contato com cliente na ponta da cadeia é o diferencial do varejo, uma vez que este detém de informações privilegiadas quanto as tendências e referências do mercado (ZAMBERLAN et al., 2009). Nos últimos anos o varejo digital atrai gradualmente um volume maior de consumidores para sua base, devido às facilidades de acesso e preços atraentes, e da mesma forma, os negócios têm interesse pelos dados de seus consumidores e preferências para conduzir as estratégias da empresa e prever seu comportamento (FALQUETO; CEZAR, 2022).

O varejo distingue seus produtos em dois tipos, produtos básicos e de moda. Produtos básicos tem demanda constante e ciclos de vida longo, o que torna seu gerenciamento mais fácil para a empresa, podendo ser usado simplesmente o histórico de vendas para determinação e manutenção de estoque. Produtos de moda tem maior variabilidade de demanda e ciclos de vida curto, o que dificulta sua gestão de demanda e estoque, e em geral são comprados somente antes do início do período de vendas com base em critérios subjetivos de especialistas (FISHER; RAJARAM, 2000).

O objetivo da empresa de varejo, assim como de outros modelos de negócio, é oferecer o produto certo, no local, tempo e preço competitivo, satisfazendo assim a demanda do consumidor (FISHER; RAJARAM, 2000). Porém, mesmo com as informações disponíveis para que as decisões corretas sejam tomadas, conduzindo

o negócio ao seu objetivo, muitas empresas têm problemas em alcançá-lo. Em seu estudo, Stuker (2014) aponta que os elementos que levam a este objetivo são a disponibilidade de dados precisos, a previsão da demanda, planejamento do estoque e velocidade da cadeia de suprimentos.

Outro fator importante para o varejo é o sortimento de produtos disponíveis, ou seja, o conjunto de produtos alocados em cada loja ou centro de distribuição. O objetivo de planejar o sortimento de produtos é maximizar a lucratividade do negócio, no entanto restrições de espaço físico e orçamento de compra devem ser observados (KÖK; FISHER; VAIDYANATHAN, 2008).

O planejamento deste mix precisa ser revisto periodicamente, uma vez que deve mudar de acordo com as estações de venda ou tendências da moda. Este planejamento é composto pela escolha da gama de produtos que será disponibilizada para o comércio, que por sua vez pode ser dividida em categorias e subcategorias. Estas escolhas definirão a amplitude, profundidade e volume de estoque de cada produto e categoria (KÖK; FISHER; VAIDYANATHAN, 2008).

Outra questão relacionada ao estoque é estratégia de precificação, que é um componente importante para o sucesso do varejo, pois não impacta somente no lucro final da operação, mas também na resposta que a cadeia de suprimento e fabricação deverá ter em relação a sua demanda, definindo o giro de produtos durante o período de vendas. Em produtos relacionados à moda, o giro das coleções é fundamental para seu sucesso, e para que isso aconteça a estratégia de precificação está vinculada ao volume de venda em determinados períodos do ciclo de vida dos produtos, determinando se haverá sobra ou não ao final da temporada (BITRAN; CALDENTEY, 2003).

No varejo, o segmento que trabalha mais próximo a estas questões relacionadas à moda é chamado de moda rápida (*fast fashion*). Este modelo de negócio se caracteriza por empresas que respondem às alterações de tendência de mercado rapidamente com produtos constantemente atualizados, quando estão em seu auge, com produtos de ciclo vida curto (CARO; MARTÍNEZ-DE-ALBÉNIZ, 2010).

Esta estratégia requer o desenvolvimento de duas características em suas partes envolvidas, alta capacidade de resposta por parte da cadeia de suprimento e produção, e apresentação de produtos contemporâneos (CACHON; SWINNEY, 2011). Estas características tornam o mercado da moda cada vez mais competitivo,

uma vez que as empresas buscam responder a qualquer alteração de tendência de maneira cada vez mais rápida, aumentando o volume de coleções das estações durante o ano.

2.2 Cadeia de suprimento

A cadeia de suprimento compreende um conjunto de processos que abrangem a circulação de materiais, informações e recursos, desde os fornecedores até os clientes. A gestão da cadeia visa coordenar e integrar eficientemente esses processos, buscando atender às necessidades e expectativas dos clientes e aumentar a competitividade da empresa, reduzindo custos e desperdícios.

2.2.1 Previsão de demanda

A indústria relacionada a moda tem como característica a imprecisão ou falta de dados históricos por presença de sazonalidade e mudanças de tendência. Desta forma a previsão de demanda se torna uma tarefa complexa, tendo produtos com ciclo muito curto e com interferência de fatores como meteorologia, economia, eventos promocionais e de marketing ou relacionados ao próprio produto, como preço e cores (THOMASSEY; FIORDALISO, 2006).

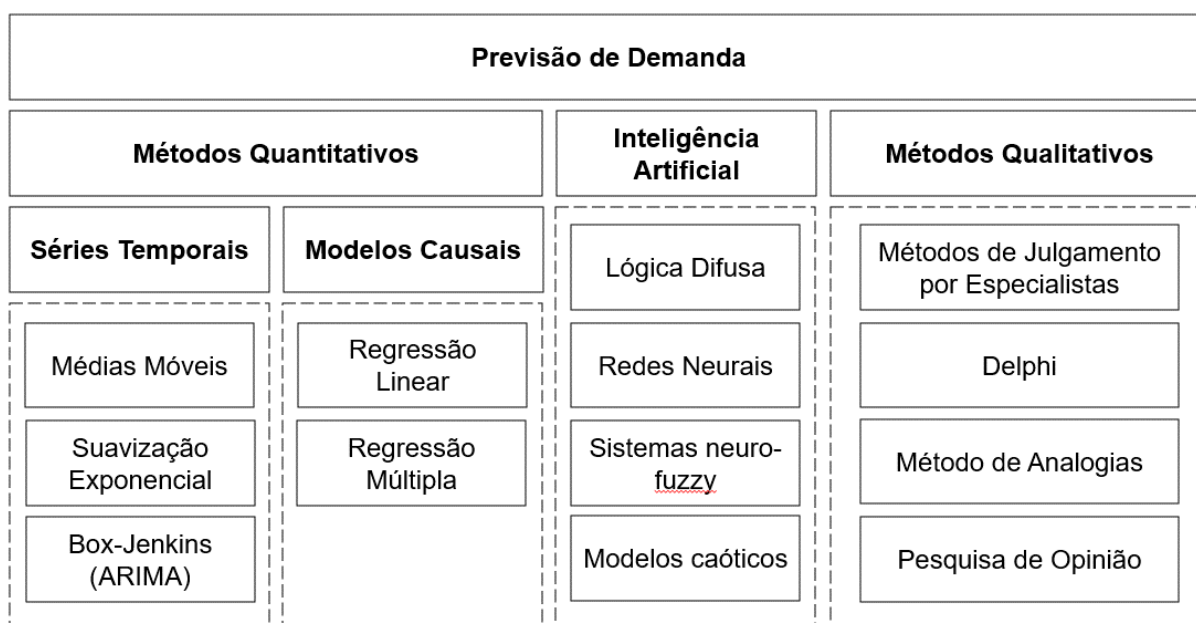
A previsão de demanda é relacionada ao pressuposto de que padrões e comportamentos dependentes anteriores se repetirão no futuro, ou serão antecipados pela tendência de moda, para mensurar grandezas futuras de interesse. A previsão envolve identificar quanto será comprado pelos clientes de um determinado produto em dadas condições de ambiente (BATISTA, 2017)

Segundo Thomassey (2010), a operação de reposição é uma das mais importantes na cadeia de suprimento relacionada a indústria da moda, e esta pode ser vista em duas partes, sendo a primeira o posicionamento inicial dos produtos nos canais de venda, em que as empresas se preparam para períodos de venda. E a segunda a reposição de produtos na cadeia durante o período de vendas, representando a capacidade de resposta rápida a demanda de mercado.

Os métodos de previsão de demanda podem também ser segregados em três grupos, quantitativos, qualitativos e inteligência artificial, conforme apresentado na Figura 3. Os métodos qualitativos são baseados em análises de especialistas e

pesquisa de mercado, que se traduzem em uma estimativa de demanda quando há poucos dados histórico ou padrão definido sobre o produto, o que é comum em lançamento de novos produtos ou tecnologias, situação recorrente da indústria de moda rápida (LÉLIS, 2012).

Figura 3 – Principais Modelos de Previsão de Demanda



FONTE: Adaptado de Stüker (2014) e Ackermann e Sellitto (2022)

Os principais métodos qualitativos são o Julgamento por Especialistas, Delphi, Métodos de Analogias e Pesquisas de Intenção e Opinião. Os dois primeiros utilizam o conhecimento de especialistas na área de análise para a definição de uma demanda, contudo, o método Delphi não exige reuniões deste time, apenas se baseia no preenchimento de formulários estruturados, que por sua vez se repetem até a redução na variação de resposta, que por fim são agrupadas e definidas por meio de moda ou mediana (STUKER, 2014).

O método de analogia é baseado em casos ou produtos similares para a definição de um padrão de comportamento histórico em ambientes de mesma característica. Por fim, a pesquisa de opinião busca por meio de pesquisa com uma amostra de população a definição de interesse em produtos ou serviços que posteriormente são traduzidos com métodos estatísticos e análises de especialistas (STUKER, 2014).

Modelos quantitativos, por sua vez, se baseiam em dados históricos para determinar projeções futuras utilizando modelos matemáticos. Estes modelos podem ser distintos em 3 grupos: Series temporais, modelos causais e inteligência artificial.

Nas últimas décadas diversos métodos para avaliar e obter uma previsão acurada foram estudados, sendo um dos mais comuns baseado em séries temporais, em que o comportamento passado se repetiria no futuro. Contudo, este método depende de amplo volume de dados históricos disponíveis, o que dificulta seu uso na indústria relacionada a moda (THOMASSEY, 2010).

Um dos modelos que pode ser considerado como séries temporais é o de média móvel. Este pode ser de dois tipos, média móvel simples ou ponderada. Na Média móvel simples a média de demanda do período mais recente é adotada como previsão para o período futuro, tendo a flutuação do resultado com o passar do tempo.

A média móvel ponderada diferencia-se pela aplicação de um peso distinto para cada dado histórico, sendo a soma deste peso aplicado igual a 1. Esta estratégia permite que o grau de importância de um período seja maior, podendo dar mais relevância a períodos mais recentes de análise ou reduzir o impacto de sazonalidade.

Modelos de suavização exponencial tem o objetivo de identificar padrões básicos de comportamento, observando os dados da mesma forma das series temporais, porém aplicam ponderação exponencial, dando mais importância aos valores com maior proximidade ao período de precisão (PELLEGRINI; FOGLIATTO, 2001). Os modelos mais conhecidos para suavização exponencial são: Suavização exponencial simples, linear de Holt e de Holt-Winters. Estes modelos se diferenciam pela possibilidade de inclusão de tendência e comportamento sazonal, e são comumente utilizados devido a sua simplicidade e facilidade de uso (MANCUSO, 2013). Apresentam também características similares à média móvel ponderada, porém sua variável de ajuste tem comportamento exponencial, formando um conjunto de previsão que decai a partir da data mais recente. Na previsão por meio da suavização exponencial simples, o resultado é obtido por meio do ajuste da previsão do período atual com um erro calculado, sendo utilizada quando o ambiente não apresenta tendências ou sazonalidade (LEMOS, 2006).

No modelo de Holt, são utilizadas duas constantes de suavização, uma para o cálculo de nível e outra para a tendência, que pode ser de crescimento ou

decrecimento da demanda, sendo que este desconsidera variações de sazonalidade. Para que o cálculo de demanda considere variações de sazonalidade, que se caracterizam pela repetição de ciclos, é utilizado o modelo de Holt-Winters, em que é acrescentado, de modo aditivo ou multiplicativo, na equação este padrão que se repete em intervalos conhecidos (PELLEGRINI; FOGLIATTO, 2001).

Outro modelo utilizado para a previsão de demanda é o ARIMA (Algoritmo Autoregressivo Integrated Moving Average), conhecido também por Box Jenkins, em homenagem aos estatísticos George Box e Gwilym Jenkins que tornaram o método popular. Neste modelo procura-se captar correlação seriada ou autocorrelação entre os valores da série temporal, e tem como ênfase a análise de propriedades estocásticas ou probabilísticas da série temporal (GUJARATI, 2004).

Há também os modelos causais, usados comumente quando existem dados e conhecimentos sobre o ambiente disponíveis para criar uma relação, e úteis quando fortes relações causais são esperadas, as direções destas relações são conhecidas, grandes alterações são previstas e as variáveis podem ser controladas ou calculadas com precisão (GREEN; ARMSTRONG, 2012). Estes modelos auxiliam em momentos como por exemplo o lançamento de novos produtos, ações de marketing ou significativas alterações de mercado.

Modelos causais podem ser fracionados em regressão linear simples ou múltipla. Na regressão linear simples, em que a relação de duas variáveis é observada, sendo uma independente e outra dependente e a múltipla, na qual mais de duas variáveis tem suas relações analisadas.

Por fim, a inteligência artificial, que ganha força com o desenvolvimento da computação, pode ser utilizada para a previsão de demanda em ambientes complexos, em que os dados não apresentam comportamento linear, por meio de aprendizado de máquina. As redes neurais (RNA) são eficientes na modelagem de problemas complexos e com dados insuficientes para sua compreensão, que por meio de aprendizado por exemplos pode se tornar uma ferramenta importante para a previsão de demanda (STUKER, 2014).

Existe a possibilidade de emprego de lógica difusa (Fuzzy logic), na qual, por meio de emprego de regras “se-então”, a ferramenta pode construir aspectos qualitativos e processo de raciocínio sem a utilização de modelos matemáticos (JANG, 1993). Entretanto, estes modelos podem não ser suficientes para a geração de resultados satisfatórios de previsão quando utilizados sozinhos, assim modelos

híbridos para a previsão de demanda, nos quais são utilizadas diversas fontes de dado e diferentes métodos para o mesmo problema, resultam em maior acurácia (ARMSTRONG, 2001).

Modelos neuro-fuzzy, por sua vez, se baseiam em uma combinação de lógica difusa e aprendizagem de redes neurais, e pode facilitar o desenvolvimento de métodos de combinação de informações e problemas complexos. Modelos caóticos, são também aplicados para problemas complexos e pode trabalhar com escassez de dados, onde a avaliação de variáveis revela padrões do sistema (Ackermann; Sellitto, 2022).

Atualmente, a análise de dados transformou a área da previsão de demanda, tornando este processo mais robusto e com alta taxa de atualização. Mas também é um desafio para as corporações, uma vez que o tráfego de dados e o volume de produtos é elevado, dificultando o processamento da informação simultaneamente. Outra questão a ser levada em consideração é a validade dos dados analisados, que se tornam obsoletos cada vez mais cedo (SEE-TO; NGAI, 2018).

2.2.2 Posicionamento de estoque

O estoque pode ser visto como uma estratégia para garantir a disponibilidade de um produto para suprir a demanda de um mercado (SANTOS et al., 2019). Custos de manutenção de estoque podem chegar a 40% do seu valor, que por sua vez pode representar até 60% dos ativos de uma empresa, evidenciando a importância de sua gestão eficiente (WANKE et al., 2017).

Quando uma rede de suprimento é desenhada, decisões a nível tático e operacional, como gerenciamento de estoque e distribuição de produtos e preços na cadeia de suprimento resumem seu desempenho (NASIRI; KALANTARI; KARIMI, 2021). A gestão dos estoques por sua vez, passa por 4 definições: onde localizar, quanto estocar, quando pedir e nível de segurança (WANKE, 1999). A alocação de estoques se resume no conjunto de decisões para a distribuição dos estoques dos fornecedores aos clientes, e quando estes não são suficientes, podem prejudicar o desempenho da operação e da organização (NASIRI; KALANTARI; KARIMI, 2021).

A questão de onde localizar os estoques para atender determinada demanda inicia por definir a estratégia entre centralizar ou descentralizar o estoque na cadeia de suprimento. Na centralização do estoque, os produtos ofertados são

posicionados em um único local, podendo proporcionar a empresa melhor gerenciamento e redução de custos com manutenção. Um estoque descentralizado consiste em distribuir os produtos em armazéns com diferentes localizações.

Um ponto positivo de descentralizar o estoque é a redução de prazos de envio, que se apresentam cada vez mais importantes para o consumidor no momento da compra. Outro aspecto importante relacionado a distribuição geográfica dos armazéns é a redução dos custos ligados a envio, decorrente do posicionamento estratégico dos mesmos (LIRA; BARBOSA; CAMERLENGO, 2021).

A principal desvantagem da distribuição descentralizada dos estoques está relacionada aos custos de manutenção dos níveis de produto e dos próprios galpões (LIRA; BARBOSA; CAMERLENGO, 2021). A complexidade gerada pela fragmentação dos volumes de estoque, e por sua vez necessidade de controles e métodos de distribuição, é outro complicador a ser levado em consideração, potencializando o desequilíbrio de distribuição.

O crescente volume e diversidade de produtos trabalhados pelo varejo traz complexidade para este sistema, que deve prever o consumo em cada nó da cadeia de suprimento, dificultando a companhia de alcançar o objetivo de oferecer o produto certo, na hora certa, para o cliente certo e pelo preço viável. Para isso, o uso de ERPs (planejamento de recursos empresariais) pelas empresas é comum, o que auxilia na logística interna, externa e controle de estoques e informações de produtos (GOPAL et al., 2022). No entanto, este tipo de sistema exige investimentos, o que limita seu uso por empresas de pequeno e médio porte.

O pressuposto de que no mercado da moda as decisões de produção inicial para suprir a demanda de futuras coleções devem ser tomadas com antecedência aumentam os riscos associados ao estoque, pois neste momento existem apenas informações limitadas e imprecisas disponíveis sobre a demanda. Com isso as decisões relacionadas ao posicionamento dos produtos são afetadas por estas restrições, reduzindo a capacidade do varejista de oferecer o produto certo, no local certo e no tempo certo (FISHER; RAJARAM, 2000).

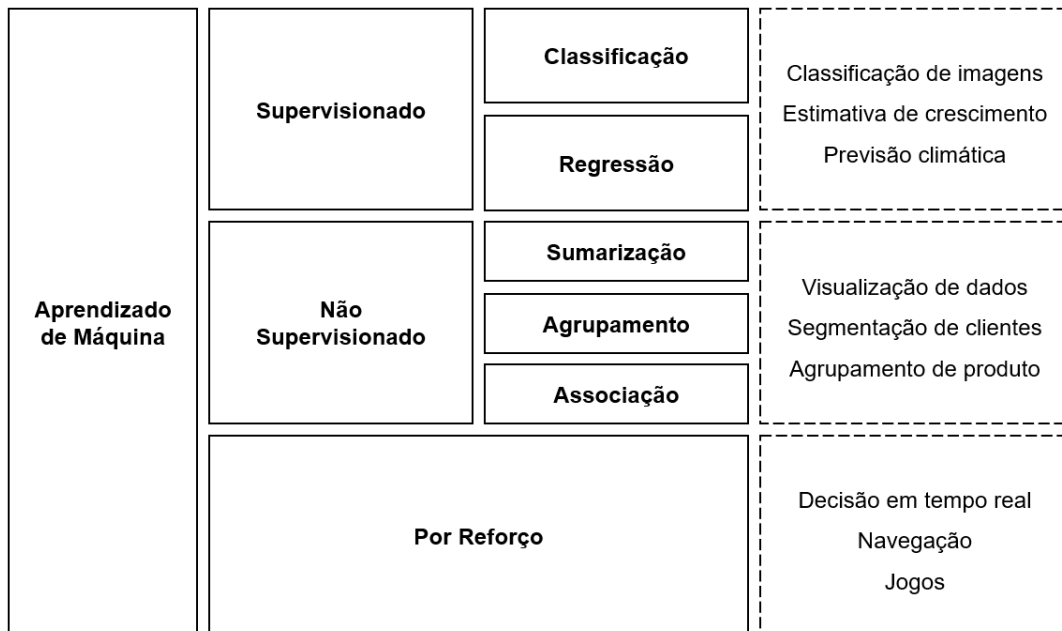
2.3 Aprendizado de máquina e clusterização

A clusterização ou agrupamento, é uma técnica que busca encontrar subgrupos homogêneos ou *clusters* em um conjunto de dados, e é comumente

utilizado na segmentação de clientes, análise de tendências e agrupamento de objetos. Esta técnica faz parte da área de estudo relacionada a mineração de dados, que pode ser definido como a descoberta de conhecimento e padrões em meio a um elevado volume de dados, podendo estes ser de diversas fontes (HAN; KAMBER; PEI, 2011).

Parte das técnicas que compõem a mineração de dados, como a clusterização, fazem parte da área de estudo conhecida como Aprendizado de Máquina, ou *Machine Learning* (WITTEN; FRANK, 2005). Esta por sua vez, se organiza em 3 classificações, conforme Figura 4: supervisionado, não supervisionado, e por reforço.

Figura 4 – Áreas do aprendizado de Máquina



FONTE: Adaptado de WANG et al. (2021)

O aprendizado de máquina supervisionado é utilizado quando existe a necessidade de o algoritmo aprender a partir de exemplos em uma base de dados estabelecida. Ou seja, a máquina é alimentada com rótulos em uma amostragem a partir de características dos dados (BERTOLINI et al., 2021).

A aprendizagem supervisionada é utilizada quando se deseja prever uma saída a partir de uma determinada entrada, sendo estas conhecidas e exibidas na base de treinamento, em que os rótulos são estabelecidos para cada grupo. Logo, o objetivo é prever precisamente os rótulos dos dados nos quais a saída é

desconhecida. Este modelo pode ser segregado em 2 tipos, regressão, quando o resultado é contínuo, e classificação quando categórico (WANG et al., 2021).

O aprendizado por máquina não supervisionado, por sua vez, é utilizado quando não há variável de resposta rotulada, portanto não há conhecimento prévio quanto aos dados de entrada. Este modelo busca descobrir padrões, relações ou classificações.

Neste tipo de aprendizado, os dados de treinamento são compostos apenas por exemplos, sem resultado associado. Esta operação é tipicamente classificada como agrupamento, porém pode ser estimativa de densidade ou redução de dimensionalidade (BERTOLINI et al., 2021).

Por fim, no aprendizado por reforço, o modelo recebe feedbacks, positivos ou negativos, a cada ação, resultando em uma solução personalizada ao ambiente ao qual responde. Desta forma o aprendizado por reforço aprende as ações ideais para sua tarefa, e é normalmente utilizado em modelos de tomada de decisão em tempo real e jogos, por exemplo (WANG et al., 2021).

2.3.1 K-modes

Dentre os diversos modelos de clusterização existentes, um dos mais conhecidos e utilizados é o *k-mean*, conhecido por sua eficiência em amplo volume de dados e simplicidade, sendo que este organiza cada objeto em grupos exclusivos (HAN; KAMBER; PEI, 2011). Neste método, é exigido como dado de entrada a quantidade de clusters como conhecimento prévio.

O método *k-mean* depende da distância Euclidiana e de k pontos, quantidade de clusters, que servirão como centroides, agregando por proximidade cada objeto a um centroide, sendo estes exclusivos (HAN; KAMBER; PEI, 2011). Por sua natureza, a técnica não pode utilizar variáveis categóricas para seu cálculo, uma vez que não possuem valor quantitativo.

Por sua estrutura simples e flexível, o método *k-mean* possibilita que sejam construídos modelos modificados ou variações do algoritmo original, conforme necessidade específica, conforme exemplos apresentados na Tabela 2. Uma das modificações possível é a troca de critério, resultando em modelos como *k-medoids*, *k-medians* e *k-modes* (AGGARWAL et al., 1999).

Tabela 2 – Variações do algoritmo k-means

Variação	Tipo de variável	Agrupamento	Particularidade
k-means	Numérica	Distância média	
k-medoids	Numérica	Distância média de um ponto real do conjunto de dados	Robusto a outliers e melhor interpretabilidade
k-medians	Numérica	Mediana dos dados	Facilita a busca por elementos mais representativos e é robusto a dados isolados
k-modes	Categórica	Moda dos atributos	Dados exclusivamente categóricos
k-prototypes	Numérica e Categórica	Atribuição de peso relativo pelo cálculo de custo	Combinação entre k-means e k-modes, que leva em consideração a distância Euclidiana e moda para definir o centroide

FONTE: Elaborado pelo autor.

O método *k-modes* altera o processo do algoritmo original, substituindo a função de distância Euclidiana pela junção das medidas de dissimilaridade, utilizando modas, ou valores mais frequentes, atualizados a cada interação (HUANG, 2009). A alteração permite que o modelo tenha como variáveis de entrada um banco de dados de objetos de características categóricas, e um número *k* representando o volume de *clusters* em que se deseja agrupar os objetos.

Segundo Huang (2009), o método *k-modes* utiliza os seguintes passos para agrupar um conjunto de dados:

1. Seleciona aleatoriamente o volume de objetos únicos como centro inicial do cluster;
2. Calcula a distância entre cada objeto de análise do conjunto até o centro de cada grupo, e atribui o objeto ao cluster de menor distância.
3. Seleciona uma nova moda para cada cluster e compara com a anterior. Caso sejam diferentes, retorna ao segundo passo, caso contrário, termina o processo.

Similar ao seu método original, o *k-modes* tem vantagens ao trabalhar com volumes amplos de dados, porém, é sensível a ruídos. Sendo assim, *outliers* podem causar alterações no processo e distorcer sua distribuição.

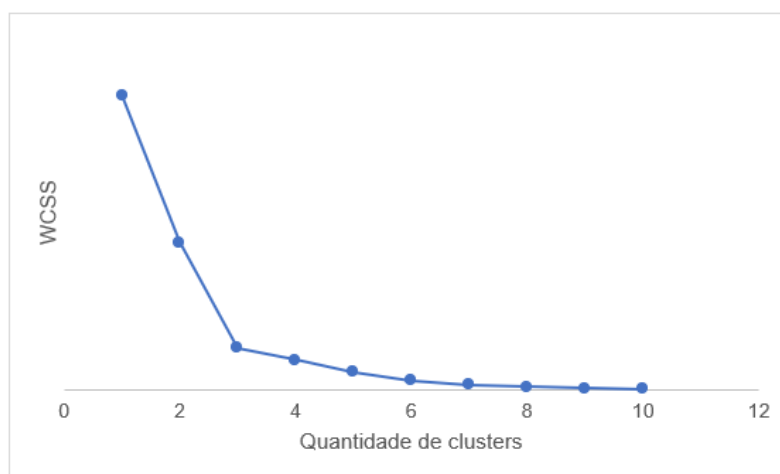
Um dos dados de entrada dos algoritmos como *k-modes* é a quantidade de clusters a serem criados para divisão do conjunto de dados. Além de ser uma

informação necessária para a criação do modelo, a decisão também impacta no resultado da análise, representando uma balança entre compressão e acurácia (HAN; KAMBER; PEI, 2011).

Um dos métodos mais comuns para encontrar a quantidade apropriada de clusters em determinado problema de agrupamento, é por meio do método do cotovelo (*Elbow method*). Este método consiste em calcular a função custo, que resulta soma dos quadrados intra-cluster (WCSS - *within-clusters sum-of-squares*), que mede a variância intra-cluster, para cada variação na quantidade de clusters e observar o momento em que a queda de WCSS não é mais significativa (LIU; DENG, 2021).

Sendo assim, o valor de WCSS para 1 cluster deve ser o maior possível, e quando a quantidade de clusters e de objetos no conjunto de dados se iguala, o valor de WCSS é zero (CUI, 2020). Logo, é possível representar esta curva de queda na variância interna, na qual observa-se que em determinado ponto, a queda no WCSS não é mais significativa, resultando na estabilização da adesão ao cluster, conforme apontado na Figura 5, na qual este ponto seria em 3 grupos.

Figura 5 – Exemplo gráfico para método do cotovelo

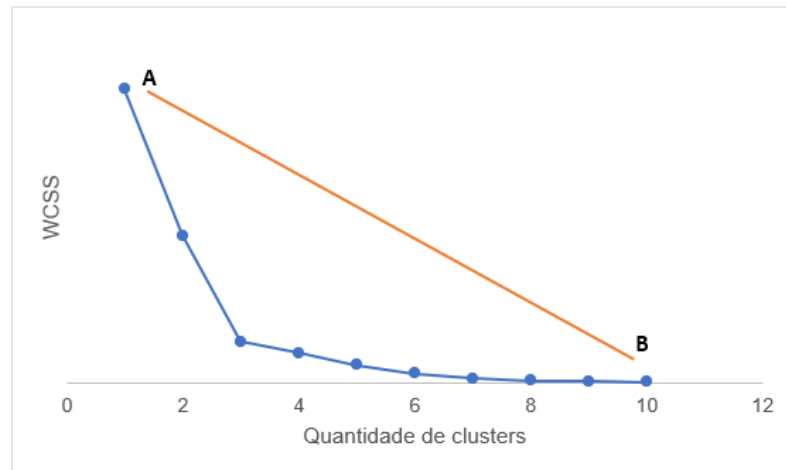


FONTE: Elaborado pelo autor

Como o método do cotovelo depende da interpretação, e em determinados casos o ponto de inflexão gerado pela variância não é explícito, esta análise se torna subjetiva (KANTARDZIC, 2011). Existe um método matemático para sugestão da quantidade apropriada de clusters em determinado problema, que consiste em calcular a maior distância entre a reta traçada entre os valores extremos de

quantidade de clusters calculado, como pode ser visto na Figura 6, representado pela reta A-B, em que novamente o ponto de inflexão seria em 3 clusters.

Figura 6 – Exemplo gráfico para método do cotovelo para medida de distância



FONTE: Elaborado pelo autor

Apesar da existência do cálculo matemático para auxiliar na leitura do gráfico e aplicação do método, indicar o equilíbrio entre maior homogeneidade dentro do cluster e a maior diferença entre clusters requer avaliação por parte do pesquisador. Este deve pesar o *tradeoff* entre acuracidade, e a compressão do conjunto de dados analisados.

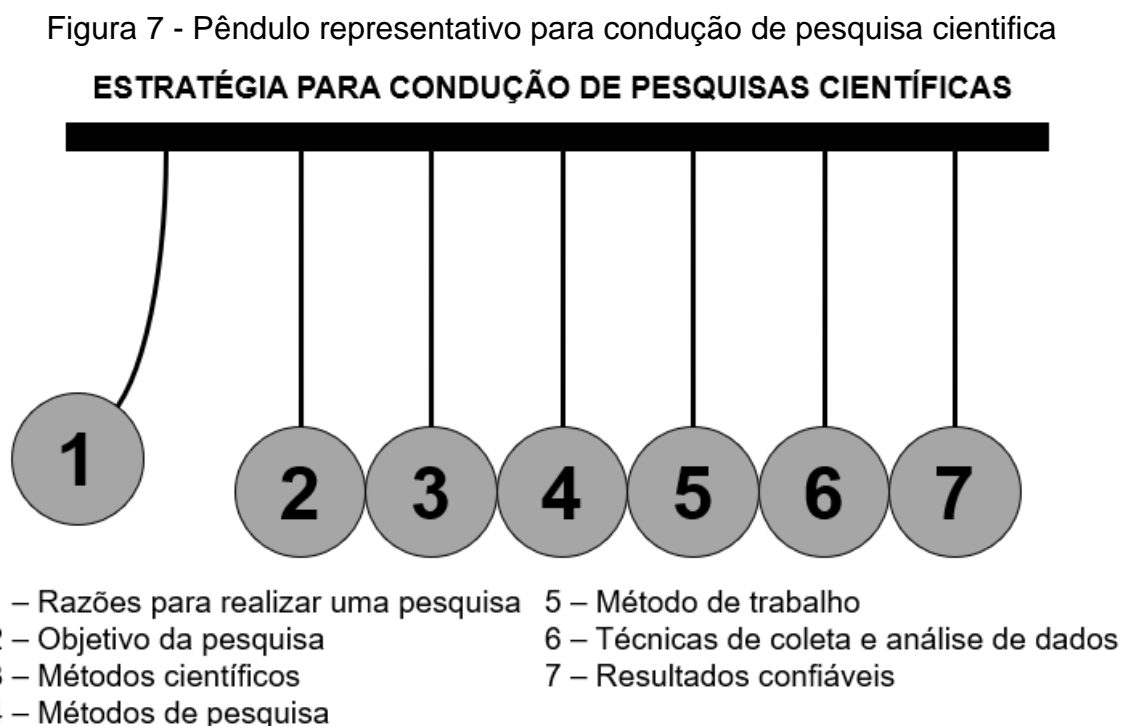
O próximo capítulo apresentará os métodos aplicados e passos tomados para a realização desta pesquisa com o objetivo de garantir replicabilidade e confiança nos resultados. Para isso o capítulo será dividido em delineamento da pesquisa, método de trabalho, coleta e análise de dados.

3 PROCEDIMENTOS METODOLÓGICOS

Este capítulo apresenta os procedimentos metodológicos utilizados. A primeira seção apresenta o delineamento da pesquisa, evidenciando o método científico e de pesquisa utilizados para o seguimento do estudo. Em seguida é apresentado o método de trabalho utilizado para a condução do estudo para alcançar os objetivos expostos no capítulo 1. Por fim, são descritos os métodos de coleta e análise de dados utilizados.

3.1 Delineamento da pesquisa

Dresch, Lacerda e Antunes Júnior (2015) afirmam que para gerar um resultado confiável, o pesquisador deve seguir uma estrutura de para condução de pesquisa, uma destas formas de estrutura é o pêndulo de Newton, formada por 7 decisões, conforme apresentado na Figura 7.



FONTE: Dresch, Lacerda e Antunes Júnior (2015, p. 16)

É importante ter consciência quanto ao alinhamento entre os elementos do pêndulo, pois uma vez que não seja feito, seus resultados podem estar comprometidos, dificultando o entendimento dos procedimentos e suas

contribuições. Dessa maneira, o pesquisador expõe os cuidados com o método, auxiliando na defesa de seus resultados (DRESCH; LACERDA; ANTUNES, 2015). Os passos 1 e 2 expostos na Figura 7 estão presentes no Capítulo 1 deste trabalho.

O método científico é “uma perspectiva ou premissa sobre como o conhecimento é construído” (DRESCH; LACERDA; ANTUNES, 2015). É importante que o contexto da pesquisa e o objetivo sejam levados em consideração para a decisão do método adotado.

Os métodos científicos podem ser classificados em indutivo, abduativo, dedutivo e hipotético-dedutivo. O método abduativo consiste em compreender um problema e criar uma teoria para explicá-lo, o dedutivo, por sua vez, se baseia no conhecimento científico para desenvolver elementos que explicam ou preveem um fenômeno, e o hipotético-dedutivo é baseado na construção de hipóteses a partir de conhecimento prévio (DRESCH; LACERDA; ANTUNES, 2015).

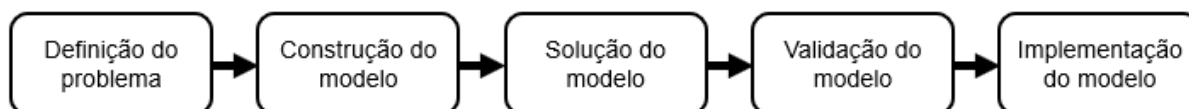
O método indutivo é caracterizado pela construção do conhecimento a partir da observação e proposição de uma teoria, em que o pesquisador faz conclusões a respeito do objeto de pesquisa (DRESCH; LACERDA; ANTUNES, 2015). Este método foi utilizado para a condução deste estudo, uma vez que a partir da pesquisa bibliográfica, utilizada como base construção do conhecimento base e das observações do pesquisador, utilizando as técnicas de coleta apresentadas nas seções seguintes, espera-se propor uma teoria que resultará em um método, que responder aos objetivos propostos no trabalho.

A formulação de estratégias para a tomada de decisão exige a utilização de modelos, que viabilizem a identificação dos problemas no contexto em que este será aplicado (MORABITO; PUREZA, 2018). Para que os objetivos deste trabalho sejam alcançados, a modelagem será utilizada como método de pesquisa, pois será criado um modelo de operação para a definição do posicionamento de produtos.

Para Morabito e Pureza (2018), a pesquisa axiomática normativa, busca compreender o comportamento de variáveis influenciadas pelo comportamento de outras dentro do modelo, buscando comparar as opções levantadas para que a melhor hipótese seja escolhida. Esta pesquisa será caracterizada desta forma, uma vez que se encaixa em seu contexto e objetivos, pois buscará por meio da análise do comportamento do consumidor e de seu perfil, definir o modelo de posicionamento dos produtos na cadeia, comparando com a realidade atual e com as hipóteses a serem levantadas.

O processo de proposto por Morabito e Pureza (2018), é apresentado na Figura 8. Conforme os autores, a primeira etapa é a definição do objeto, em que o contexto deve ser estudado e compreendido. Em seguida, as informações coletadas na fase anterior são avaliadas para a construção de um modelo, que objetiva sanar o problema proposto pela pesquisa. Na fase de validação, o modelo deve ser provado, buscando saber se este representa a realidade. O processo se encerra com a implementação do modelo proposto na prática.

Figura 8 - Processo de modelagem



FONTE: Morabito e Pureza (2018)

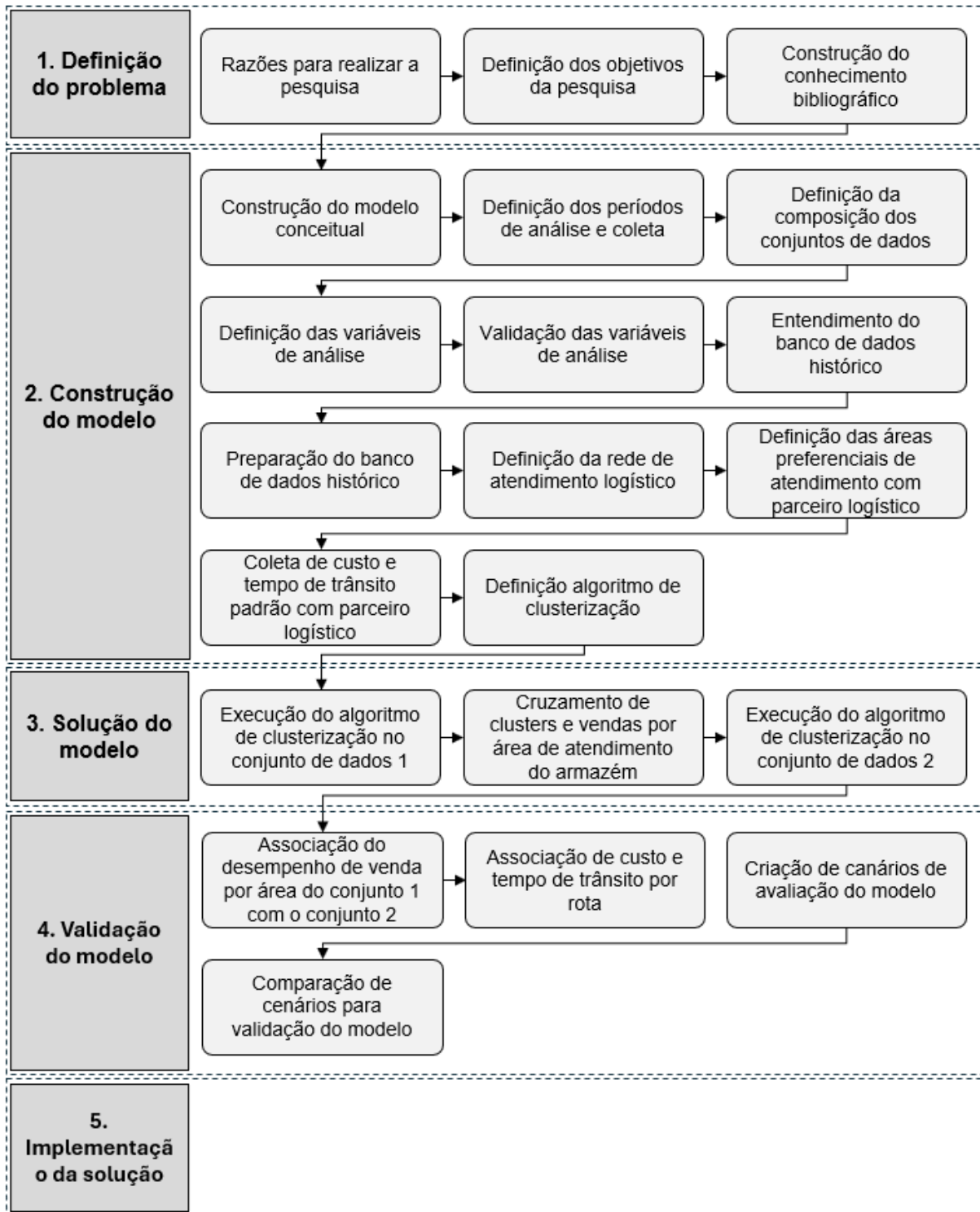
No capítulo seguinte o método de trabalho, utilizado para a construção deste estudo será exposto.

3.4 Método de trabalho

Para que a pesquisa alcance os objetivos propostos, segundo Dresch, Lacerda e Antunes Júnior (2015), o método de trabalho deve definir uma sequência de passos lógicos. Ainda segundo Dresch, Lacerda e Antunes Júnior (2015), para que a replicabilidade do estudo seja assegurada, este método deve estar estruturado e ser seguido corretamente, desta forma o resultado será mais claro e transparente, facilitando o reconhecimento por outros pesquisadores.

A figura 9 apresenta o método de trabalho utilizado neste estudo, que se divide em 5 fases, conforme proposto por Morabito e Pureza (2018). Como abordado no item 1.4, esta pesquisa avança até a fase 4, na qual a validação se dará pelos dados históricos. A fase 1 foi explorada no capítulo 1 deste trabalho conforme proposto por Dresch, Lacerda e Antunes Júnior (2015).

Figura 9 - Método de trabalho

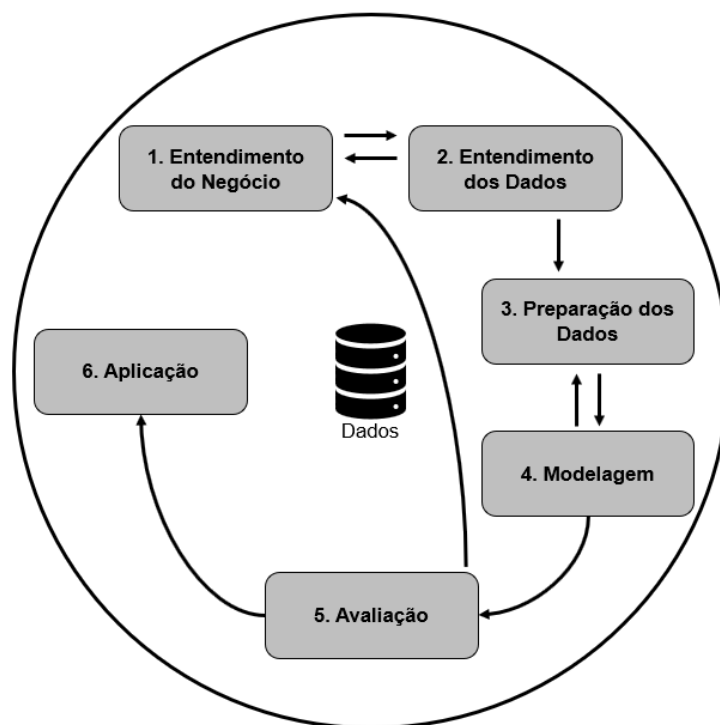


FONTE: Elaborado pelo autor

Nas Fases 2, 3 e 4 serão trabalhadas seguindo o método CRISP-DM (*CrossIndustry Standard Process for Data Mining*) conforme proposto por Shearer

(2000). Este método é composto por 6 etapas de tratamento, modelagem, validação e aplicação do método proposto, conforme apresentado na Figura 10.

Figura 10 - Método CRISP-DM



FONTE: Elaborado pelo autor com base em Shearer (2000)

Na primeira etapa, a de entendimento do negócio, conforme exposto no capítulo 1 deste trabalho, foi realizada a contextualização do ambiente, assim como a definição dos objetivos. Em seguida, no entendimento dos dados, foi realizada uma inspeção inicial dos dados disponíveis e sua qualidade para aplicação no trabalho.

Nesta etapa, foram definidas as coleções que farão parte desta análise, duas coleções servindo de base para a definição do comportamento do cluster referentes ao período de um ano, e a terceira, referente ao período de meio ano posterior às duas primeiras para teste de aplicação e validação. As datas que compõem cada coleção foram usadas para definir o período de coleta de desempenho dos produtos selecionados de cada coleção, e que serão detalhadas na seção de análise de dados.

Na terceira etapa, preparação dos dados, as bases selecionadas são tratadas para que possam compor o banco final a ser utilizado. Neste momento as variáveis

que serviram para compor o modelo de clusterização foram selecionadas conforme disponibilidade de dados, e validadas por especialistas escolhidos por sua relevância na operação, assim como os dados de histórico de venda para o período determinado foram também carregados e tratados para compor o conjunto de análise final.

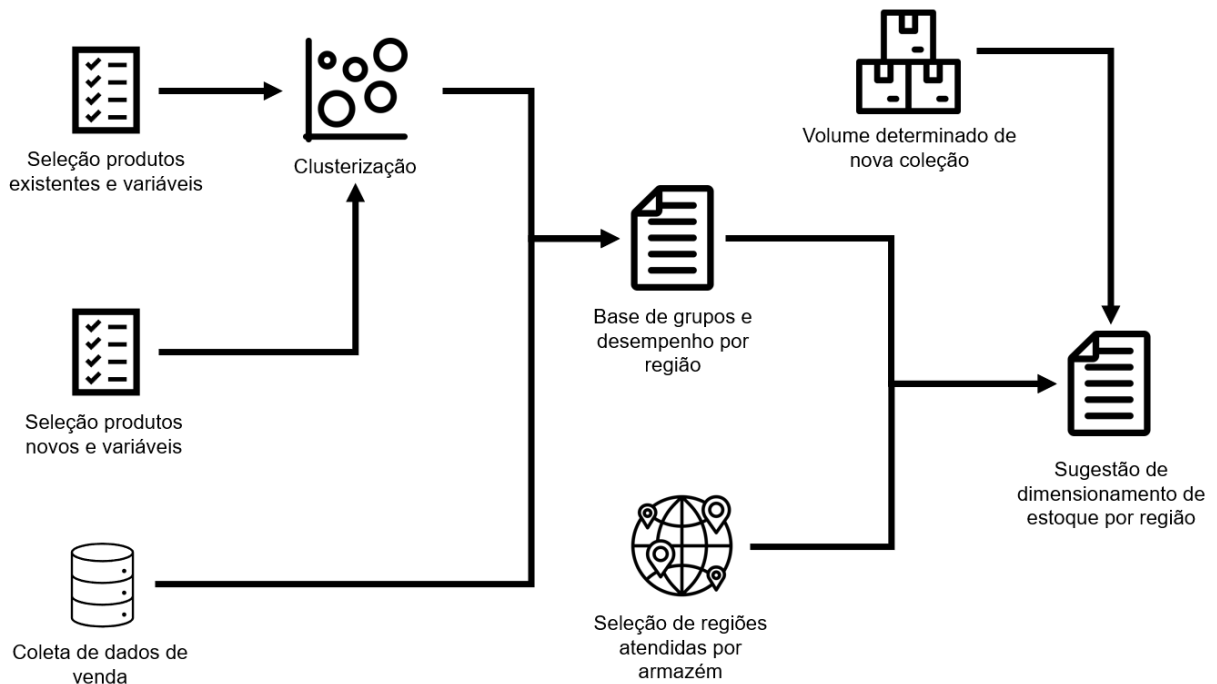
As regiões de atendimento preferencial de cada armazém foram organizadas seguindo a composição sugerida pelo prestador de serviço logístico parceiro da empresa na qual este trabalho foi desenvolvido, seguindo como base sua estrutura de entrega. Custo e tempo de trânsito foram também coletados dos documentos da empresa, conforme cotação vigente, para compor a análise de resultados e posterior comparação dos cenários.

Na modelagem, o algoritmo de clusterização *k-modes*, selecionado por sua característica de processar dados categóricos, foi aplicado inicialmente no conjunto relativo os produtos que compõem as duas coleções que servirão de base para o modelo. Utilizando como chave de conexão entre as bases o SKU dos produtos, os volumes vendidos para cada região de atendimento dos diferentes armazéns foram utilizados para fracionar a participação de cada cluster em cada armazém.

Por fim, compondo a fase 3, os parâmetros criados na clusterização do primeiro grupo de produtos foram utilizados no segundo, onde o algoritmo de clusterização foi aplicado. O resultado desta classificação é a alocação dos produtos conforme similaridade com o primeiro grupo de produtos.

Os percentuais de participação de venda por região de atendimento preferencial dos centros logísticos, definidos a partir da análise do desempenho comercial conhecido do primeiro grupo de produtos, foram utilizados para o dimensionamento dos estoques previstos entre os centros logísticos para cada cluster formado, em que os produtos da coleção futura foram alocados, segundo grupo. O método de trabalho proposto pode ser melhor compreendido a partir da leitura do fluxograma exposto na Figura 11.

Figura 11 – Fluxograma de trabalho



FONTE: Elaborado pelo autor

Na etapa de avaliação, os custos foram utilizados para compor três cenários, executado, previsto e modelo. O primeiro relativo ao desempenho da coleção de análise conforme executado no período real, o segundo, utilizado como potencial máximo de resultado, dividindo os volumes para o melhor atendimento da demanda conhecida no período, e o último, conforme proposto pelo modelo com base nas coleções passadas.

Por fim, na aplicação, os cenários simulados são comparados para avaliar o modelo proposto e definir sua eficiência de para alocação de sortimento e volume. O modelo é validado com dados reais históricos disponíveis, sendo a fase 4 do método exposto.

3.5 Coleta de dados

Nesta seção são expostos os métodos utilizados para a coleta de dados para a realização desta pesquisa. Segundo Dresch, Lacerda e Antunes (2015), as técnicas de coleta podem ser: documental, bibliográfica, entrevistas, grupos focais, questionários e observações diretas.

A técnica de coleta bibliográfica foi utilizada para embasar tecnicamente o pesquisador e encontrar o conhecimento existente acerca do tema deste trabalho. O mapeamento do conhecimento publicado condiciona o pesquisador a realizar novas descobertas (DRESCH; LACERDA; ANTUNES, 2015).

Para a execução desta técnica foram utilizadas as etapas propostas por Morandi e Camargo (2015), seguindo o protocolo da revisão sistemática da literatura exposto no Apêndice A, que recebeu a contribuição de dois especialistas relacionados da área de estudo. O resultado desta coleta está descrito na seção 1.3 deste documento.

A segunda técnica de coleta utilizada é a documental, na qual o pesquisador buscou o histórico de pedidos realizados no canal de e-commerce, armazenados na ferramenta de gerenciamento utilizada pela empresa, PostgreSQL, conforme exposto pelo dicionário de dados no Apêndice B deste trabalho. Estes dados foram acessados por meio do software Power BI, no qual a análise foi realizada.

Os períodos de interesse dos dados de venda são referentes às coleções de Spring 2022 e Fall 2022 para treinamento do modelo, abrangendo o período de 30/01/2022 a 30/07/2022, e de 31/07/2022 a 28/01/2023, respectivamente. E para análise e aplicação do modelo, a coleção de Spring 2023, relativo ao período de 29/01/2023 a 29/07/2023, sendo 26 semanas em cada coleção.

Foram utilizadas também as características dos produtos das coleções selecionadas. A escolha das características válidas para o trabalho foi realizada por avaliação de especialistas que representam as áreas chave do processo envolvidos na operação, conforme Tabela 3, e confirmadas empiricamente, comparando os dados históricos.

Tabela 3 – Especialistas avaliadores

Especialista	Cargo	Formação	Tempo na empresa	Critério de seleção
Especialista 1	Coordenador de Planejamento	Engenharia de Produção	1 ano	Conhecimento relacionado a estoques e distribuição do sortimento
Especialista 2	Gerente de Operações (Logística / Planejamento / Sourcing)	Administração	5 anos	
Especialista 3	Coordenador de <i>Marchandising</i>	Moda	1 ano	Conhecimento de mercado e tendências
Especialista 4	Gerente de <i>Merchandising</i>	Publicidade e Propaganda	5 anos	

FONTE: Elaborado pelo autor

Os SKUs e características dos produtos são armazenadas em tabelas de Excel e no sistema SAP, utilizado pela empresa. Após a análise, esta base validada também foi importada para análise no software Power BI.

No caso estudado, a empresa conta com parceiros fornecedores de serviços logísticos, que oferecem estudos de aprimoramento de entrega dentro do território de interesse com base na localização dos armazéns utilizados pela companhia e recursos disponíveis do prestador, definindo o melhor local de origem para determinados destinos. Esta informação fornecida foi utilizada para a definição das áreas atendidas por cada armazém.

3.6 Análise de dados

Esta seção apresenta as técnicas utilizadas para a análise dos dados coletados. A primeira etapa consiste em identificar as principais características dos produtos, para isso foi utilizado uma base de atributos de cada SKU, contendo informações como categoria, subcategoria, cor, material e faixa de preço.

Definidas as características de produto a serem controladas o pré-processamento e limpeza das bases acessadas foram feitos por meio de inspeção exploratória, buscando compreender os dados de interesse para a pesquisa. Uma vez identificados, estes foram utilizados para o modelo e separados do restante da consulta por meio de linguagem SQL (*Structured Query Language* - linguagem de consulta estruturada) no editor de Power Query, ferramenta disponível no software Power BI para transformação dos dados, no qual o restante da análise foi realizada. Desta forma não somente há maior clareza, mas também reduz o volume da consulta realizada ao banco, aumentando a performance do processo.

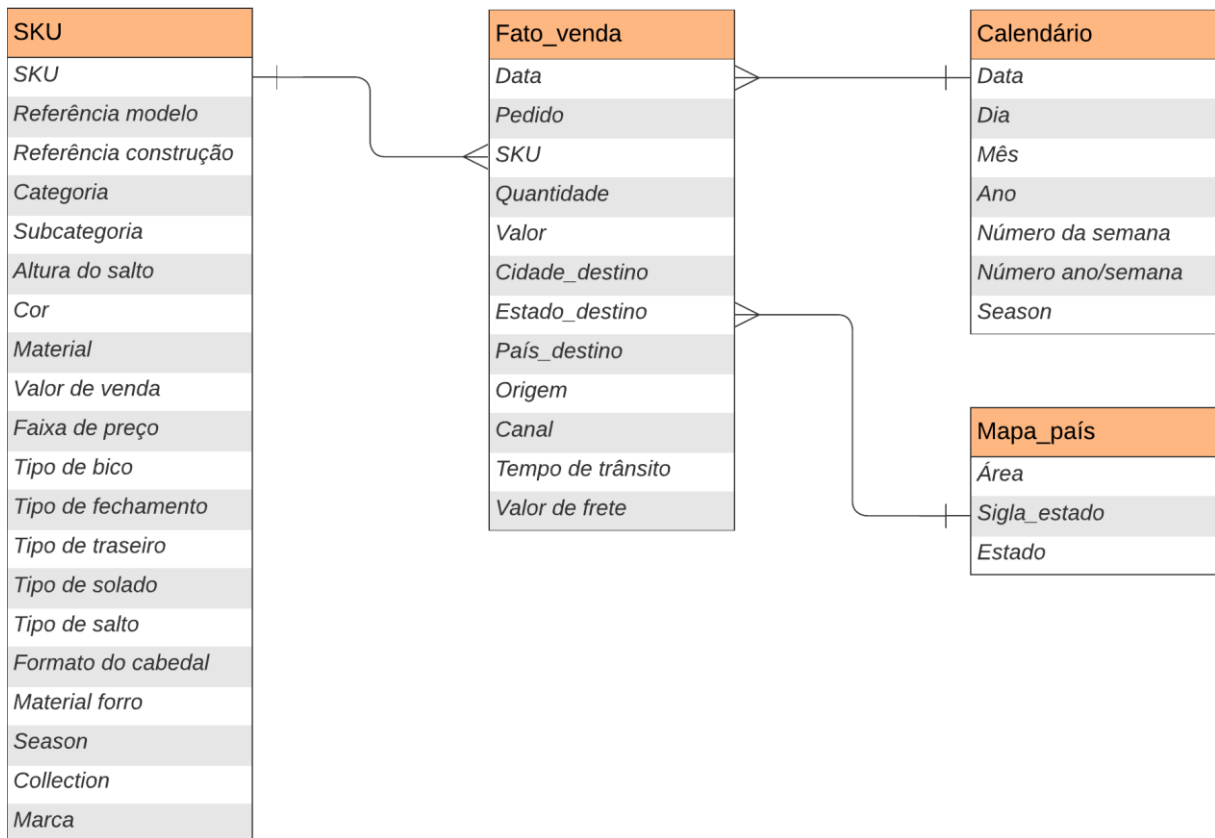
Com a base de produtos de treinamento e teste definidas, as mesmas foram carregadas para o Google Colab, serviço baseado em nuvem que permite ao usuário criar e executar códigos em Python. O Colab foi utilizado para avaliação do volume de clusters, por meio do teste do cotovelo, e posteriormente a criação dos clusters contendo SKUs.

Inicialmente a base de produtos de Spring 2022 e Fall 2022 foram carregadas para o modelo, em que o teste do cotovelo foi realizado para definição do volume de grupos e separação dos SKUs. Na sequência os produtos referentes a coleção de Spring 2023 foram carregados para que, utilizando os parâmetros criados na

seleção das coleções anteriores, os SKUs fossem alocados nos clusters da nova coleção. Por fim, os clusters foram atribuídos à base original e exportados.

Definidos os dados e bases de interesse na fase de transformação, estes foram relacionados no *software* Power BI por meio da chave SKU e data de cada produto e registro de venda, presente no arquivo contendo as características e nas bases selecionadas, conforme Figura 12. A partir disso foi criada uma consulta única de análise por meio de linguagem DAX (*Data Analysis Expressions* - Expressões de análise de dados).

Figura 12 – Diagrama de Entidades



FONTE: Elaborado pelo autor

Para criação do modelo base, utilizaram-se os dados de venda para cada região considerando os locais de atendimento em cada cluster. Este resultado foi aplicado como participação para cada grupo de produtos e armazém, criando um volume de distribuição dos produtos.

Os dados de venda e base de SKUs clusterizados foram cruzadas para construir 3 cenários para comparação entre os resultados das transações durante o

período da coleção de Spring 2023 nos produtos lançamento. No primeiro cenário, foi utilizado o desempenho real da empresa no período, consultado conforme histórico de desempenho documentado.

O segundo cenário é o considerado ideal conforme destino dos produtos comprados no e-commerce. A análise partiu dos dados de destino de cada venda, associando o armazém preferencial para seu atendimento.

Por fim, o terceiro cenário é a simulação do desempenho do processo, caso os produtos fossem enviados conforme alocação sugerida pelo modelo. Para isso, o volume inicial de cada produto foi alocado conforme sugestão de distribuição resultante do modelo, e o consumo deste volume foi feito preferenciando o armazém ideal para cada transação conforme disponibilidade. Os resultados desta análise são apresentados no capítulo seguinte.

4 ANÁLISE DOS RESULTADOS

Este capítulo apresenta os resultados obtidos com a aplicação dos métodos apresentados na seção anterior deste trabalho. A análise dos resultados será apresentada em três partes, sendo os resultados da clusterização dos produtos, da distribuição, e da avaliação e comparação dos cenários.

4.1 Clusterização

O método utilizado para clusterização neste estudo, *k-modes*, utiliza como variáveis de entrada o conjunto de dados dos objetos de análise e a quantidade de grupos que se deseja dividir os objetos. Nesta seção serão exploradas estas variáveis de entrada e sua avaliação, e os resultados da aplicação do método de clusterização.

4.1.1 Análise de Variáveis

Para análise e validação das variáveis foram consultados 4 especialistas, os quais tem contato direto com o produto, conhecimento de suas variáveis e processo da empresa, conforme listados no capítulo 3.6 deste trabalho. Para esta avaliação os especialistas foram consultados em grupo e sem formulário estruturado, tendo sido as variáveis relacionadas no banco de dados da empresa apresentadas, conforme Tabela 4.

Tabela 4 – Variáveis avaliadas

Variável	Incluído	Removido
Referência modelo		x
Referência construção		x
Categoria	x	
Subcategoria	x	
Altura do salto		x
Cor	x	
Material	x	
Valor de venda		x
Faixa de preço	x	
Tipo de bico	x	
Tipo de fechamento	x	
Tipo de traseiro	x	

Variável	Incluído	Removido
Tipo de solado	x	
Tipo de Salto		x
Descrição formato do cabedal		x
Material forro		x

FONTE: Elaborado pelo autor

Conforme apresentado na Tabela 4, 16 variáveis foram selecionadas para avaliação conjunta, dentre estas, 9 relacionadas para inclusão da composição do conjunto, e 7 removidas. Os motivos de remoção das variáveis apresentadas estão expostos na Tabela 5.

Tabela 5 – Variáveis removidas

Variável	Motivo remoção
Referência modelo	Volume excessivo de variantes
Referência construção	Variação entre modelos de mesma construção
Altura do salto	Informação resumida na categoria e subcategoria
Valor de venda	Informação resumida na faixa de preço
Tipo de salto	Dado não confiável
Descrição formato do cabedal	Dado não confiável
Material forro	Dado irrelevante

FONTE: Elaborado pelo autor

De acordo com o grupo de especialistas, no caso da referência do modelo, determinados SKUs podem ser sua única variação, impossibilitando agrupamentos consistentes. A referência da construção é um caso similar, porém, os especialistas julgam que as variantes entre construção podem não apresentar similaridade entre os produtos, podendo criar um ruído na análise ao não representar possíveis comportamentos de compra.

Nos casos da altura de salto e valor de venda, estas mesmas variáveis são utilizadas resumidamente nas análises da empresa, sendo representadas pela categoria e subcategoria no caso da altura de salto, e pela faixa de preço para o valor de venda do produto. Esta decisão influenciou ainda na escolha do algoritmo de clusterização, uma vez que foi utilizado o *k-modes* variação do *k-means* para variáveis categóricas. Caso variáveis contínuas fossem adicionadas ao modelo, variantes como a *k-prototype* poderia ser uma opção.

Os dados de tipo de salto e descrição do formato do cabedal são cadastrados pela equipe de cadastro e não passam por revisão da equipe de estilo, por este

motivo foi decidido que estas não seriam relacionadas. Por fim, o material do forro foi removido na análise por não significar relevância estética para o cliente, e muitas vezes a decisão de alteração desta característica é por questões financeiras.

As variáveis incluídas no trabalho foram organizadas e avaliadas quanto à qualidade do dado, verificando faltas de informação para possível correção. O resultado da composição está exposto na Tabela 6.

Tabela 6 – Composição de variáveis incluídas

Categoria	Subcategoria	Cor	Material	Faixa de preço
SANDAL	HIGH STILETTO	BLACK	LEATHER	79-98
BOOTIE	WEDGE	BROWN	MULTIMATERIAL	99-118
PUMP	MID STILETTO	SILVER	METALLIC	119-138
FLAT	FLAT	WHITE	SYNTHETIC	139-158
ESPADRILLE	MID BLOCK	BEIGE	VINYL	159-178
BOOT	BALLET	GOLD	NUBUCK	179-198
	HIGH BLOCK	ORANGE	SUEDE	'>=200
	PLATFORM	RED	METALLIC SYNTHETIC	
	FLATFORM	PINK	FABRIC	
	LOW STILETTO	PURPLE	PATENT	
	LOW BLOCK	BLUE	SNAKE	
	MULE	GREEN	PATENT SYNTHETIC	
	MAN TAILOR	MULTICOLOR	CROC	
		ANIMAL		
Bico	Fechamento	Traseiro	Tipo de solado	
BICO REDONDO	AMARRACAO	FECHADO	COURO	
BICO FINO	FIVELA	ABERTO	INJETADO	
BICO FOLHA	ELASTICO	SEM TRASEIRO	LAQUEADO	
BICO QUADRADO	ZIPER		TRATORADO	
	SEM FECHAMENTO			

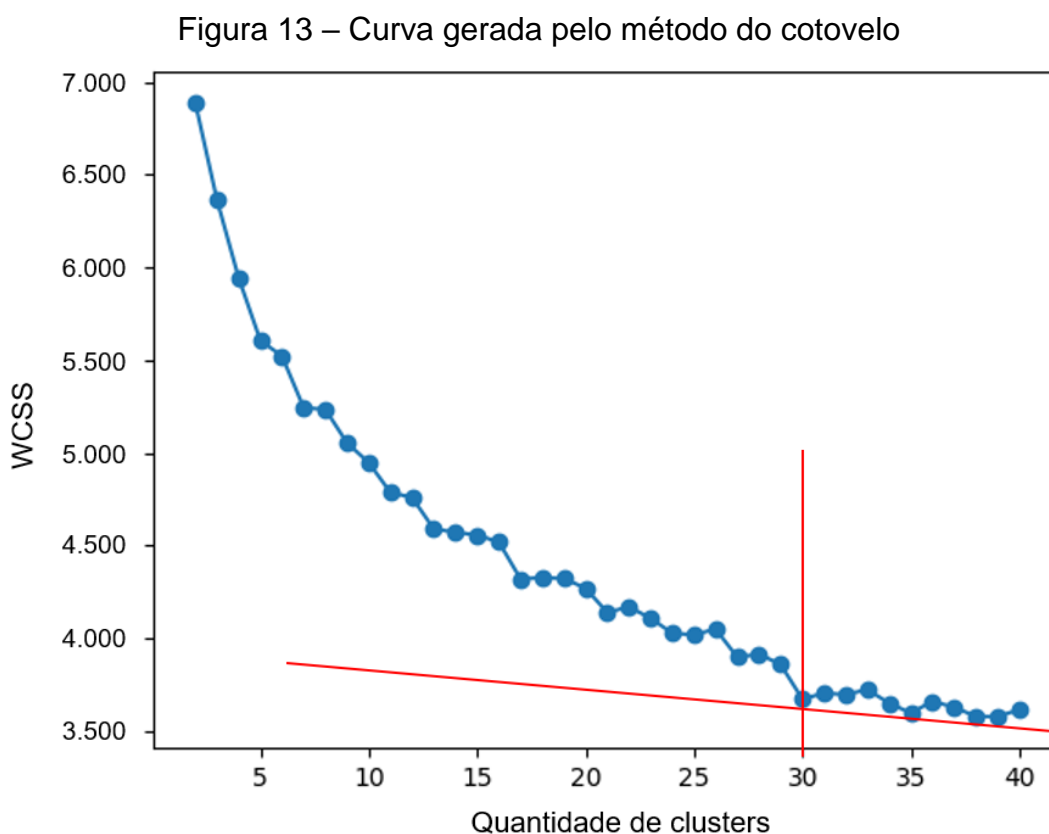
FONTE: Elaborado pelo autor

4.1.2 Resultado de clusterização

Antes da execução do algoritmo de clusterização, o método do cotovelo foi aplicado para compor os dados de entrada para o modelo. Conforme exposto no capítulo 3, o método do cotovelo tem como função apontar o ponto no qual existe o melhor balanço entre compressibilidade e acurácia por meio de um gráfico, no qual é

analisada a curva em busca do ponto de inflexão, ou seja, a queda da dispersão entre duas quantidades de clusters não é significativa.

A execução do método resultou na curva exposta na Figura 13. Para auxiliar na interpretação do método, o cálculo da distância entre a reta traçada pelo ponto de variância inicial e final foi realizado.



FONTE: Elaborado pelo autor

Esta técnica de análise apontou que a maior distância até a reta está na quantidade de 17 clusters. Porém, como o método depende da interpretação do pesquisador, um ponto de inflexão é percebido na quantidade de 30 grupos, no qual existe uma redução significativa na queda da dispersão e estabilização da curva, portanto, optou-se por utilizar a divisão do conjunto de dados em 30 grupos

O processo de clusterização foi realizado em duas etapas, inicialmente a tabela das coleções de Spring 2022 e Fall 2022 foram carregadas e o algoritmo executou a função de buscar as modas ou centroides, realizando o processo do algoritmo por completo, e vincular os rótulos a cada objeto do conjunto de dados.

O resultado da distribuição dos SKUs relativos à coleção de Spring 2022 e Fall 2022 nos clusters está exposto na Tabela 7. Concluído o processo de

clusterização das coleções que servirão de base para a distribuição dos SKUs de Spring 2023, foi realizada a análise de distribuição de vendas dos SKUs dispostos em cada cluster para formar a alocação ideal dos volumes de compra.

Tabela 7 – Distribuição de SKUs Spring 2022 e Fall 2022 nos clusters

Cluster	Quantidade de SKUs	% SKU
0	24	2,6%
1	56	6,0%
2	13	1,4%
3	84	9,0%
4	17	1,8%
5	49	5,2%
6	40	4,3%
7	30	3,2%
8	47	5,0%
9	18	1,9%
10	74	7,9%
11	78	8,3%
12	35	3,7%
13	50	5,3%
14	29	3,1%
15	27	2,9%
16	11	1,2%
17	30	3,2%
18	35	3,7%
19	28	3,0%
20	2	0,2%
21	22	2,3%
22	35	3,7%
23	16	1,7%
24	16	1,7%
25	16	1,7%
26	19	2,0%
27	9	1,0%
28	5	0,5%
29	22	2,3%
Total	937	100%

FONTE: Elaborado pelo autor

A segunda etapa do processo consiste em utilizar as mesmas modas ou centroides criados pelo algoritmo para clusterização de Spring 2022 e Fall 2022, e

rotular os SKUs da coleção de Spring 2023. Desta forma, a lógica de agrupamento utilizada para analisar as vendas de 2022 e criar a distribuição, é aplicada na coleção seguinte, resultando na seguinte distribuição de SKUs, conforme exposto na Tabela 8.

Tabela 8 - Distribuição de SKUs Spring 2023 nos clusters

Cluster	Quantidade de SKUs	% SKU
0	22	3,2%
1	50	7,3%
2	32	4,7%
3	48	7,0%
4	28	4,1%
5	58	8,5%
6	2	0,3%
7	31	4,5%
8	11	1,6%
9	22	3,2%
10	7	1,0%
11	58	8,5%
12	4	0,6%
13	-	0,0%
14	17	2,5%
15	14	2,0%
16	-	0,0%
17	34	5,0%
18	24	3,5%
19	18	2,6%
20	2	0,3%
21	5	0,7%
22	58	8,5%
23	4	0,6%
24	2	0,3%
25	45	6,6%
26	58	8,5%
27	9	1,3%
28	7	1,0%
29	16	2,3%
Total	686	100%

FONTE: Elaborado pelo autor

Conforme apresentado, para a nova coleção os clusters 13 e 16 não apresentaram produtos relacionados, assim como também é visto diferenças entre o

volume relativo de representantes de cada grupo. Este comportamento indica que há variação de estilos entre as coleções.

4.2 Resultado da análise de distribuição de clusters

A distribuição dos volumes de compra para as regiões atendidas por cada armazém dos produtos para a coleção de Spring 2023 foi realizada com base na performance de vendas das coleções anterior e local preferencial de origem para cada pedido. Conforme dados apresentados pela companhia logística parceira a relação de centro logístico preferencial e estado de destino do pedido está exposto na Tabela 9, na qual os centros localizados nos estados da Califórnia, Florida e New Jersey, são representados como C.L. CA, C.L. FL e C.L. NJ, respectivamente.

Tabela 9 – Estados de atendimento preferencial por centro logístico

C.L. CA		C.L. FL		C.L. NJ	
WA	Washington	AR	Arkansas	MN	Minnesota
DC	District of Columbia	LA	Louisiana	IA	Iowa
OR	Oregon	MS	Mississippi	MO	Missouri
ID	Idaho	AL	Alabama	WI	Wisconsin
MT	Montana	GA	Geórgia	IL	Illinois
WY	Wyoming	FL	Flórida	MI	Michigan
CA	Califórnia	OK	Oklahoma	IN	Indiana
NV	Nevada	TN	Tennessee	KY	Kentucky
UT	Utah	SC	Carolina do Sul	OH	Ohio
AZ	Arizona			WV	Virgínia Ocidental
CO	Colorado			ME	Maine
NM	Novo México			VT	Vermont
ND	Dakota do Norte			NY	Nova Iorque
SD	Dakota do Sul			PA	Pensilvânia
NE	Nebraska			MA	Massachusetts
KS	Kansas			NJ	Nova Jérsei
TX	Texas			VA	Virgínia
AK	Alasca			NC	Carolina do Norte
HI	Havaí			CT	Connecticut
				DE	Delaware
				MD	Maryland
				NH	Nova Hampshire
				RI	Rhode Island

FONTE: Elaborado pelo autor

O resultado das coleções passadas, por sua vez, foi cruzado com o local preferencial de envio para cada região do país. A Tabela 10 apresenta o resultado desta análise, que representa a distribuição prevista caso cada venda fosse atendida por seu centro logístico preferencial.

Tabela 10 – Distribuição prevista para coleções Spring 2022 e Fall 2022

Cluster	C.L. CA	C.L. FL	C.L. NJ
0	35,0%	22,3%	42,7%
1	37,5%	28,1%	34,5%
2	34,1%	27,8%	38,1%
3	33,0%	22,3%	44,7%
4	36,1%	25,9%	38,0%
5	38,9%	24,6%	36,5%
6	33,6%	14,2%	52,1%
7	36,4%	23,7%	39,9%
8	31,6%	29,5%	38,9%
9	33,6%	26,7%	39,7%
10	36,9%	19,5%	43,6%
11	30,4%	27,8%	41,8%
12	33,9%	19,0%	47,1%
14	29,4%	25,9%	44,7%
15	37,4%	19,9%	42,7%
17	31,2%	28,1%	40,8%
18	31,0%	26,8%	42,2%
19	29,5%	30,2%	40,3%
20	41,3%	21,0%	37,7%
21	36,2%	23,3%	40,5%
22	29,5%	27,1%	43,4%
23	30,6%	27,7%	41,7%
24	35,0%	29,1%	35,9%
25	31,8%	32,1%	36,1%
26	32,2%	23,7%	44,1%
27	38,8%	29,2%	31,9%
28	29,0%	21,2%	49,8%
29	28,7%	27,5%	43,8%
Total	33,6%	24,7%	41,7%

FONTE: Elaborado pelo autor

A partir da análise da tabela é possível identificar que o percentual de participação de envio por local preferencial varia de 41,3% a 28,7% para C.L. CA, de 32,1% a 14,2% para C.L. FL, e de 52,1% a 31,9% em C.L. NJ, sendo que em apenas 4 clusters o C.L. CA tem maior representação, e o restante em C.L. NJ.

Como resultado, do volume total comprado para a coleção de Spring 2023, 41,7% deve ser destinado ao C.L. NJ, 33,6% para C.L. CA, e 24,7% para C.L. FL.

4.3 Resultado comparativo entre cenários

Para a avaliação dos resultados do modelo em relação à distribuição da operação logística, serão analisados três cenários:

- Resultado executado;
- Resultado previsto;
- Resultado do modelo.

O resultado executado se refere aos pares enviados efetivos durante o período de fevereiro a julho de 2023, que se refere ao Spring 2023. Este contexto servirá como base de comparação para os resultados do modelo, uma vez que serviram de base para determinar a sua eficiência.

O segundo cenário, previsto, diz respeito ao resultado preferencial da operação, ou seja, caso todos os pedidos fossem enviados de seu centro logístico preferencial, no qual o resultado de relação entre tempo de trânsito e custo são os melhores. Junto ao resultado executado, o cenário previsto forma a baliza necessária para julgar o resultado da simulação de utilização do modelo.

Por fim, o resultado do modelo refere-se à simulação do uso do método proposto neste trabalho, no qual, com base no desempenho de determinado período anterior, e da clusterização dos produtos, são feitas as alocações do volume de compra para o próximo período. Os resultados para cada um dos cenários descritos são então comparados para compreender o impacto financeiro e operacional do modelo.

Para relacionar o resultado executado foram verificados os pedidos realizados no período de análise e identificados os dados de destino e origem do pedido, bem como o cluster dos SKUs. O resultado desta consulta para cada cluster criado está apontado na Tabela 11.

Tabela 11 – Resultado executado por cluster em Spring 2023

Cluster / Origem	Volume Absoluto				Volume Relativo		
	C.L. CA	C.L. FL	C.L. NJ	Total	C.L. CA	C.L. FL	C.L. NJ
0	135	137	672	944	14,3%	14,5%	71,2%
1	886	399	1.646	2.931	30,2%	13,6%	56,2%
2	465	361	1.496	2.322	20,0%	15,5%	64,4%
3	536	620	1.073	2.229	24,0%	27,8%	48,1%
4	297	208	407	912	32,6%	22,8%	44,6%
5	375	258	1.145	1.778	21,1%	14,5%	64,4%
6	56	1	6	63	88,9%	1,6%	9,5%
7	110	94	443	647	17,0%	14,5%	68,5%
8	47	35	577	659	7,1%	5,3%	87,6%
9	355	42	326	723	49,1%	5,8%	45,1%
10	359	58	268	685	52,4%	8,5%	39,1%
11	685	261	2.437	3.383	20,2%	7,7%	72,0%
12	0	0	5	5	0,0%	0,0%	100,0%
14	55	61	412	528	10,4%	11,6%	78,0%
15	99	15	259	373	26,5%	4,0%	69,4%
17	571	249	1.501	2.321	24,6%	10,7%	64,7%
18	580	699	1.308	2.587	22,4%	27,0%	50,6%
19	266	100	464	830	32,0%	12,0%	55,9%
20	19	4	26	49	38,8%	8,2%	53,1%
21	137	2	31	170	80,6%	1,2%	18,2%
22	929	545	2.145	3.619	25,7%	15,1%	59,3%
23	2	50	44	96	2,1%	52,1%	45,8%
24	0	0	61	61	0,0%	0,0%	100,0%
25	891	266	2.844	4.001	22,3%	6,6%	71,1%
26	1.111	733	2.718	4.562	24,4%	16,1%	59,6%
27	106	4	141	251	42,2%	1,6%	56,2%
28	22	10	56	88	25,0%	11,4%	63,6%
29	87	49	583	719	12,1%	6,8%	81,1%
TT Geral	9.181	5.261	23.094	37.536	24,50%	14,00%	61,50%

FONTES: Elaborado pelo autor

Para o cenário previsto, cada pedido deveria ser faturado desde sua origem com melhor relação entre custo e tempo de trânsito, considerando os dados da Tabela 9. A distribuição dos produtos para este contexto é apresentada na Tabela 12.

Tabela 12 – Resultado previsto por cluster em Spring 2023

Cluster / Origem	Volume Absoluto				Volume Relativo		
	C.L. CA	C.L. FL	C.L. NJ	Total	C.L. CA	C.L. FL	C.L. NJ
0	298	194	452	944	31,6%	20,6%	47,9%
1	947	724	1.260	2.931	32,3%	24,7%	43,0%
2	677	659	986	2.322	29,2%	28,4%	42,5%
3	695	495	1.039	2.229	31,2%	22,2%	46,6%
4	258	198	456	912	28,3%	21,7%	50,0%
5	552	424	802	1.778	31,0%	23,8%	45,1%
6	25	17	21	63	39,7%	27,0%	33,3%
7	185	143	319	647	28,6%	22,1%	49,3%
8	200	171	288	659	30,3%	25,9%	43,7%
9	243	150	330	723	33,6%	20,7%	45,6%
10	165	133	387	685	24,1%	19,4%	56,5%
11	957	945	1.481	3.383	28,3%	27,9%	43,8%
12	2	2	1	5	40,0%	40,0%	20,0%
14	149	148	231	528	28,2%	28,0%	43,8%
15	124	48	201	373	33,2%	12,9%	53,9%
17	825	589	907	2.321	35,5%	25,4%	39,1%
18	731	696	1.160	2.587	28,3%	26,9%	44,8%
19	206	271	353	830	24,8%	32,7%	42,5%
20	19	9	21	49	38,8%	18,4%	42,9%
21	40	44	86	170	23,5%	25,9%	50,6%
22	1.042	884	1.693	3.619	28,8%	24,4%	46,8%
23	29	17	50	96	30,2%	17,7%	52,1%
24	16	13	32	61	26,2%	21,3%	52,5%
25	1.173	1.073	1.755	4.001	29,3%	26,8%	43,9%
26	1.607	1.094	1.861	4.562	35,2%	24,0%	40,8%
27	62	77	112	251	24,7%	30,7%	44,6%
28	33	19	36	88	37,5%	21,6%	40,9%
29	264	172	283	719	36,7%	23,9%	39,4%
TT Geral	11.524	9.409	16.603	37.536	30,70%	25,10%	44,20%

FONTE: Elaborado pelo autor

Por fim, o cenário 3 representa os envios caso a distribuição sugerida pelo modelo de clusterização baseado no desempenho passado fosse seguido. Nesta situação, quando o volume vendido e nominalmente alocado em determinado centro logístico supera o volume de estoque inicialmente disposto, o próximo armazém com menor custo, segundo valores disponíveis no Anexo A, e estoque disponível é utilizado.

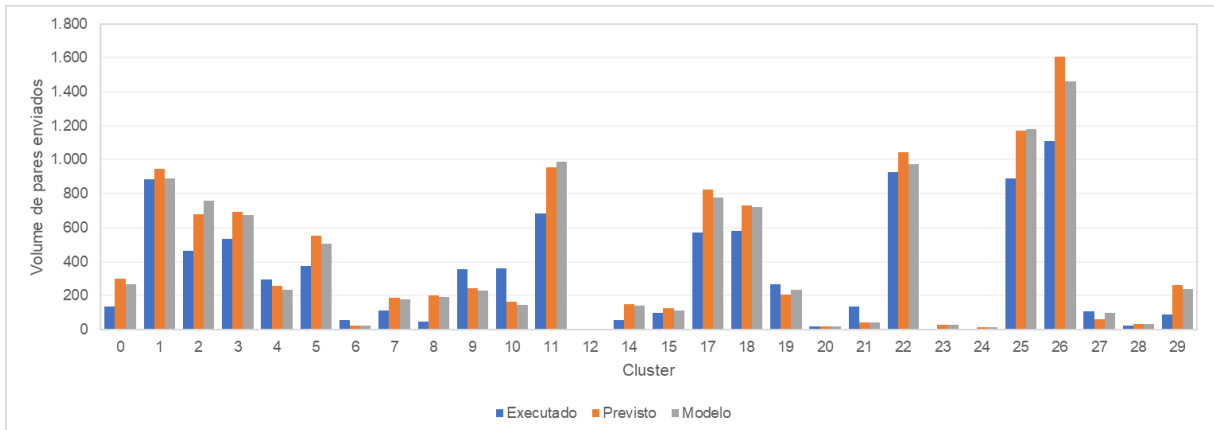
Tabela 13 – Resultado do modelo por cluster em Spring 2023

Cluster / Origem	Volume Absoluto				Volume Relativo		
	C.L. CA	C.L. FL	C.L. NJ	Total	C.L. CA	C.L. FL	C.L. NJ
0	268	209	467	944	28,4%	22,1%	49,5%
1	889	888	1.154	2.931	30,3%	30,3%	39,4%
2	757	627	938	2.322	32,6%	27,0%	40,4%
3	676	544	1.009	2.229	30,3%	24,4%	45,3%
4	236	269	407	912	25,8%	29,5%	44,6%
5	507	536	735	1.778	28,5%	30,2%	41,3%
6	21	12	30	63	33,3%	19,7%	47,0%
7	175	180	292	647	27,1%	27,8%	45,1%
8	191	195	273	659	29,0%	29,5%	41,5%
9	230	184	309	723	31,8%	25,4%	42,8%
10	146	196	343	685	21,3%	28,6%	50,1%
11	987	961	1.435	3.383	29,2%	28,4%	42,4%
12	2	2	1	5	40,0%	40,0%	20,0%
14	141	140	247	528	26,7%	26,5%	46,9%
15	110	85	178	373	29,5%	22,8%	47,7%
17	776	692	853	2.321	33,4%	29,8%	36,8%
18	721	765	1.100	2.587	27,9%	29,6%	42,5%
19	233	256	340	830	28,1%	30,9%	41,0%
20	19	10	20	49	38,8%	20,4%	40,8%
21	40	45	85	170	23,5%	26,5%	50,0%
22	973	1.068	1.577	3.619	26,9%	29,5%	43,6%
23	28	23	45	96	29,2%	23,6%	47,2%
24	13	21	27	61	21,9%	34,2%	43,8%
25	1.179	1.242	1.580	4.001	29,5%	31,0%	39,5%
26	1.461	1.076	2.025	4.562	32,0%	23,6%	44,4%
27	97	73	81	251	38,6%	29,1%	32,3%
28	32	21	35	88	36,3%	24,2%	39,6%
29	239	224	256	719	33,2%	31,1%	35,6%
TT Geral	11.148	10.544	15.844	37.536	29,70%	28,30%	42,00%

FONTE: Elaborado pelo autor

Os três cenários apresentados representam o mesmo volume, período e mix de vendas, alterando somente o centro logístico de origem do produto. As Figura 14, 13 e 14 apresentam de maneira visual o resultado absoluto enviado para cada cluster e canário em C.L. CA, C.L. FL e C.L. NJ.

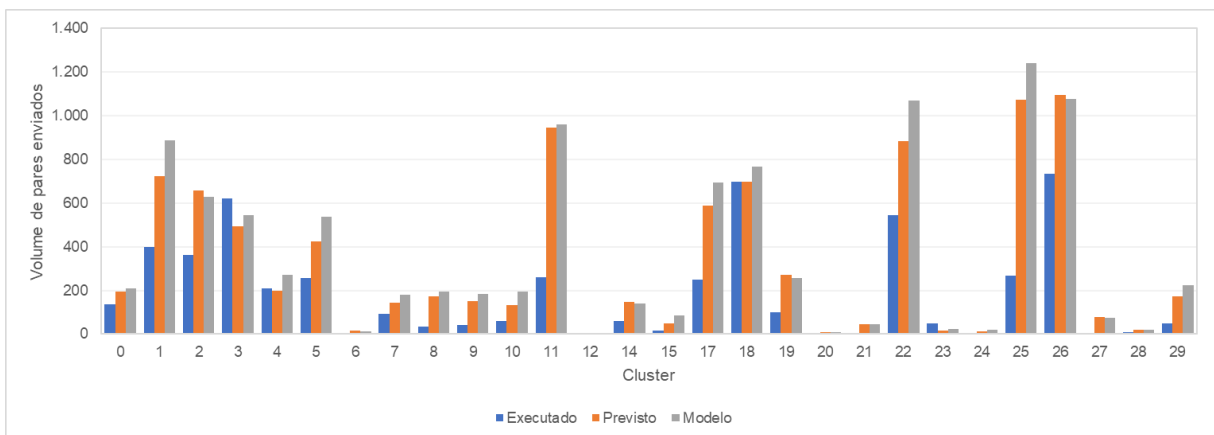
Figura 14 – Comparação do resultado dos cenários por cluster em C.L. CA



FONTE: Elaborado pelo autor

A partir da leitura do gráfico, é possível notar que o cenário simulado a partir da aplicação do modelo se mostra mais próximo do previsto na maior parte dos casos. Para o comparativo em C.L. CA, nota-se que o volume aumenta na maioria dos casos no cenário previsto em comparação ao executado, o que indica uma atual subutilização do armazém em relação a sua demanda.

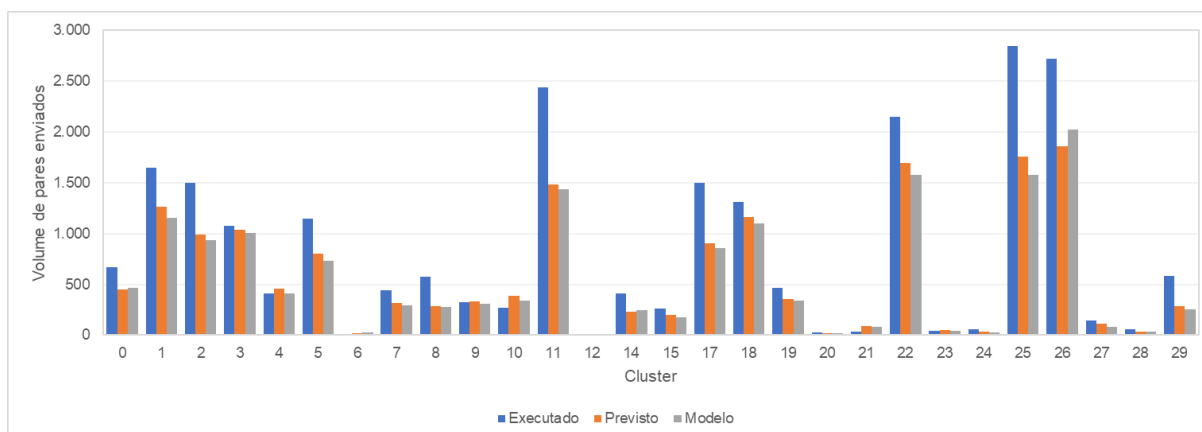
Figura 15 – Comparação do resultado dos cenários por cluster em C.L. FL



FONTE: Elaborado pelo autor

No caso do comparativo de cenários em C.L. FL, a leitura é similar a C.L. CA, porém, com maior discrepância entre os volumes utilizados no cenário executado, em que a subutilização do armazém é ainda maior. Neste caso, as diferenças nos clusters 1, 11, 17 e 25 chamam mais atenção.

Figura 16 – Comparação do resultado dos cenários por cluster em C.L. NJ



FONTE: Elaborado pelo autor

No comparativo entre cenários de C.L. NJ, ainda existe uma proximidade maior do resultado do modelo proposto com o previsto, porém, nota-se a inversão de utilização no cenário executado, em que o recurso foi claramente mais exigido em relação a sua demanda e aos outros centros logísticos. Neste caso, o melhor balanceamento seria certamente atingido entre as cargas de trabalho dos armazéns caso o modelo proposto fosse utilizado.

Para melhor entender os resultados de envios corretos entre os clusters, na Tabela 14 o cenário executado e o simulado a partir da aplicação do modelo são comparados quanto ao seu percentual de acerto em relação ao cenário previsto. O resultado mostra que em todos os casos o modelo apresenta um percentual de acerto maior que o cenário executado.

Em geral, não é vista uma relação entre os acertos e os clusters, em que, com exceção dos clusters 12 e 20, nos quais o volume é reduzido, resultando em maior sensibilidade a movimentos, todos se aproximam. Removendo os clusters 6, 12, 20, 23, 24 e 28, em que o volume de envios dos SKUs representantes é menor que 100 pares, a variação média absoluta para o percentual de acertos geral é de 3% para ambos os cenários.

Tabela 14 – Volume de envios da origem correta em Spring 23

Cluster	Acertos vs Nominal			
	Executado		Modelo	
	Absoluto	Relativo	Absoluto	Relativo
0	464	49%	868	92%
1	1216	41%	2665	91%
2	985	42%	2318	100%
3	944	42%	2163	97%
4	377	41%	802	88%
5	804	45%	1616	91%
6	27	43%	44	70%
7	324	50%	586	91%
8	281	43%	624	95%
9	285	39%	675	93%
10	314	46%	597	87%
11	1445	43%	3208	95%
12	1	20%	5	100%
14	227	43%	497	94%
15	181	49%	325	87%
17	1215	52%	2174	94%
18	1155	45%	2384	92%
19	357	43%	768	93%
20	35	71%	49	100%
21	58	34%	170	100%
22	1567	43%	3353	93%
23	41	43%	86	90%
24	32	52%	49	80%
25	1623	41%	3572	89%
26	2188	48%	4562	100%
27	99	39%	251	100%
28	32	36%	85	97%
29	332	46%	644	90%
Média		44%		92%

FONTE: Elaborado pelo autor

Para resumir o resultado apresentado nas tabelas anteriores, a Tabela 15 expõe o volume absoluto e relativo dos pedidos enviados de cada centro logístico, possibilitando a comparação de resultado entre cenários.

Tabela 15 – Resultado comparativo entre cenários

Cenário / Origem	Volume Absoluto				Volume Relativo		
	C.L. CA	C.L. FL	C.L. NJ	Total	C.L. CA	C.L. FL	C.L. NJ
Executado	9.181	5.261	23.094	37.536	24,5%	14,0%	61,5%
Previsto	11.524	9.409	16.603	37.536	30,7%	25,1%	44,2%
Modelo	11.148	10.544	15.844	37.536	29,7%	28,3%	42,0%

FONTE: Elaborado pelo autor

A partir da análise das tabelas é possível observar uma semelhança de distribuição tanto em cada cluster, mas com maior clareza do cenário previsto para o cenário projetado seguindo o modelo descrito neste estudo. Este é um indício de que há um ganho relacionado a utilização do modelo de agrupamento para previsão de comportamento de compra regional por perfil de produto.

Para melhor compreender os impactos da aplicação deste modelo é necessário realizar a comparação financeira relacionada a operação de envio do produto ao cliente, fator relevante para o sucesso de uma operação, e o tempo de trânsito até o endereço de destino, fator considerado pelo consumidor para compra de produtos em canais digitais. Para isso no tópico seguinte os resultados de custo e tempo de trânsito serão avaliados para cada cenário exposto.

4.4 Avaliação de custo e tempo de trânsito

Para melhor avaliar o resultado de custo e tempo de trânsito entre os cenários avaliados, os envios de cada centro logístico para cada destino foram relacionados com o custo unitário de envio e tempo de trânsito para cada relação de origem e destino conforme dados da empresa apresentados no Anexo A. A Tabela 16 resume estes valores para melhor comparação de resultado.

Tabela 16 – Resultado comparativo de custo e tempo de trânsito

	Custo TT [\$]	Variação	Custo médio [\$]	Variação	Tempo médio [Dias]	Variação
Executado	\$ 474.147	-	\$ 12,66	-	2,83	-
Previsto	\$ 390.334	-18%	\$ 10,42	-18%	1,54	-46%
Modelo	\$ 398.283	-16%	\$ 10,60	-16%	1,62	-43%

FONTE: Elaborado pelo autor

Os custos totais consideram a soma dos valores de envio de cada unidade dentro de seu respectivo contexto. O resultado executado considera valores reais da

operação, sendo este o valor máximo, que serve de referência para o cálculo do potencial de ganho existente.

Em oposição o resultado previsto considera os valores caso cada envio ocorresse desde sua origem prevista relacionada ao destino do pedido. Considera-se então este como o melhor resultado possível a ser alcançado com a operação.

Quanto ao custo, o resultado alcançado pela simulação da aplicação do modelo proposto para posicionamento e dimensionamento do estoque é de 16% menor que o executado, representando uma economia potencial de \$75.864 para a operação, sendo o equivalente a \$2,06 por item enviado. Este resultado representa ainda 2,03% superior ao valor previsto, indicando que o modelo apresenta relevante redução de custos e eficácia no processo.

Em relação ao tempo médio de trânsito, há um ganho de 43% comparado ao processo atual, próximo em 5,2% do valor previsto. Este ganho pode representar um potencial de satisfação do cliente com o serviço prestado, visto a importância desta medida. Esta variável pode ter benefícios intangíveis para o negócio, uma vez que a marca pode se beneficiar a longo prazo deste ganho de tempo de entrega.

5 DISCUSSÃO DOS RESULTADOS

Neste capítulo serão discutidos os resultados apresentados no capítulo anterior a partir da literatura disponível sobre o tema. O objetivo é explicitar claramente os resultados e contribuições deste trabalho.

É possível evidenciar que o modelo proposto traz resultados positivos para a cadeia de suprimento ao propor uma leitura das preferências de consumo regional vinculadas aos armazéns existentes da operação e propor a alocação de estoque de novos produtos para melhor atendimento da demanda existente. De maneira similar, outros trabalhos apresentam o agrupamento de clientes para traçar um perfil de consumo e ler as preferências de determinados grupos para previsão de demanda e alocação de estoque (BALAKRISHNAN et al., 2018; MURRAY; AGARD; BARAJAS, 2015; PEREIRA; FRAZZON, 2019; XUE; LIN, 2017).

A utilização e interpretação dos dados agrupados apresenta a possibilidade de previsão de comportamento viável para um grande conjunto de objetos, como o caso do contexto apresentado. Resultados similares foram encontrados por Murray et al. (2015), quando busca interpretar a demanda futura de um grupo amplo de consumidores.

Ao apresentar um modelo de agrupamento de produtos inéditos ao consumidor, em que não há relação com a base histórica de vendas, como é o caso do setor de moda *fast fashion*, análises de comportamento de produtos similares podem ser extraídas para a compreensão de movimentos futuros. Trabalhos similares de agrupamento na área da moda apresentam resultados mais genéricos, utilizando um volume reduzido de características para vincular os produtos com os clusters (FALLAH TEHRANI; AHRENS, 2017).

Este trabalho se diferencia ao relacionar um volume maior de atributos, com maior grau de similaridade entre os objetos de análise, os quais são validados por especialistas da empresa. O modelo ainda se concentra mais no produto do que no cliente, como outros trabalhos encontrados na área, apresentando um meio de posicionar novos lançamentos por sua característica física, e avaliar sua eficácia.

Neste contexto, esta pesquisa se beneficia pelas características compartilhadas entre os produtos da empresa, uma vez que são utilizados somente calçados, seus aspectos físicos são facilmente comparáveis, possibilitando a formação de um conjunto de dados robusto. O oposto pode ser observado em

trabalhos em que a clusterização foi realizada a partir de um conjunto de dados mais subjetivo, como o caso estudado por Holý et al. (2017), em que foram relatados problemas nas categorizações de produto, por utilizar meios subjetivos para definir as características dos produtos. Isso também se deve ao fato de que a comparação de produtos de diferentes naturezas aumenta a dificuldade de comparação de suas características com a percepção do consumidor.

Quanto à capacidade da atribuição da demanda ao armazém de melhor escolha para a operação, o trabalho expõe resultados positivos nos custos de operação da empresa, corroborado no trabalho de Berman et al. (2016) e Berling et al. (2023), que apresentam redução de custo de até 30%, sendo em média a redução de 9,7%. Neste estudo a redução de custo obtida na simulação do emprego do modelo proposto em dados históricos é de 16%, considerando somente a realocação dos novos modelos adquiridos pela empresa, sem levar em consideração reposições de produtos existentes.

Berman et al. (2016), salientam também que, em casos em que muitas instalações de envio operam simultaneamente, exista um *trade off* quanto ao armazém com menor custo de envio, escolha da empresa, e o armazém mais próximo, preferido pelo consumidor. No caso estudado, optou-se por considerar somente o envio desde os centros logísticos existentes, três espalhados por todo o território dos Estados Unidos, sem considerar a possibilidade de escolha de lojas, as quais contam com estoque reduzidos, mix de produtos específicos, e todas localizadas próximas aos centros logísticos, desta maneira, o local de menor custo e mais próximo, se tornam os mesmos.

Neste sentido, Bijmolt et al. (2021) reforçam que decisões importantes como a de possibilitar diferentes pontos de estoque, nos quais se utilizam lojas para envio de produtos adquiridos online, levam a consequências de escassez de estoque em pontos físicos de venda. Por outro lado, os autores chamam atenção para a possibilidade de incluir entregas ágeis para parte do sortimento a partir destes pontos de venda físicos, principalmente em produtos com estoques elevados. Por outro lado, Fisher et al. (2019) apontam para o benefício gerado por entregas rápidas em relação ao aumento de custos.

O modelo resulta em ganhos representativos para o tempo de trânsito médio até o cliente consumidor, com redução apresentada na simulação de 43% ou 1,21 dias a menos no caso. Este resultado pode representar um ganho para a

credibilidade da marca por melhorar a experiência do cliente através do aumento do nível de serviço oferecido, sendo pontualidade e tempo de entrega o aspecto mais importante ligado ao serviço de logística em uma operação de compras online (PEREIRA; FRAZZON, 2021).

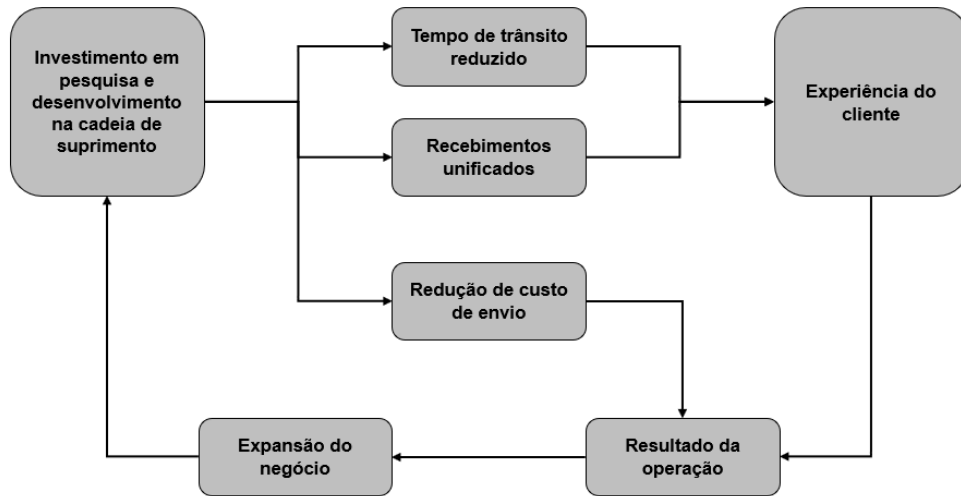
Relacionado ao tempo, a distância e a exposição dos produtos transportados aos riscos relacionados ao transporte, como avarias, intempéries e desvios tem sua probabilidade reduzida. Este ganho é relatado também no trabalho de Bijmolt et al. (2021), que relatam relação entre o tempo e distâncias extras percorridas como o aumento de avaria relatadas pelos consumidores nos produtos.

Outro fator oculto resultante de uma distribuição apropriada dos volumes de estoque é o envio completo de produtos a partir de um mesmo ponto, evitando entregas em diferentes momentos ao consumidor. O envio fragmentado pode resultar em uma experiência pior para o cliente, que prefere receber seus pedidos em um lote único (BIJMOLT et al., 2021).

Berman et al. (2016), descrevem que sua pesquisa resulta em um contrassenso em relação a tendência natural de distribuição de estoques, em que esta deve ser feita de maneira uniforme entre os postos de armazenagem para garantir a disponibilidade ideal em todo o território. Este estudo corrobora esta ideia ao apresentar uma distribuição para armazém fornecedores para áreas equivalentes em extensão, porém com padrões e volumes de consumo variáveis.

A partir dos resultados discutidos neste capítulo, é possível ainda construir relações entre os resultados potenciais para a empresa, o desenvolvimento de técnicas e modelos de aprendizagem de máquina relacionados à área de *Supply Chain* e a melhor experiência oferecida ao consumidor. A representação desta relação é representada pela Figura 17.

Figura 17 – Relação entre resultados e ganhos



FONTE: Elaborado pelo autor

A Figura 17 apresenta características similares ao ciclo de melhoria contínua, em que o desenvolvimento e melhorias no processo retornam como investimentos para novos desenvolvimentos. No capítulo seguintes as conclusões e considerações finais do trabalho, assim como sugestões para pesquisas futuras serão apresentadas.

6 CONSIDERAÇÕES FINAIS

A utilização de modelos com uso de *machine learning* está se difundindo pela indústria e resultando em avanços em todas as áreas, como na gestão da cadeia de suprimentos e operações. Investimentos por parte das empresas e avanços acadêmicos suportam este desenvolvimento.

Este capítulo aborda as considerações finais deste estudo, retomando os objetivos propostos, resultados e contribuições da pesquisa. Os temas deste trabalho foram discutidos entre a literatura existente e as práticas da empresa, aproximando as aplicações teóricas e contribuições práticas da operação de uma empresa calçadista com o objetivo principal de construir um modelo para dimensionar os estoques nos armazéns disponíveis da empresa utilizando aprendizagem de máquina.

O objetivo geral proposto foi atingido, uma vez que o modelo foi criado, simulando com dados reais e comparado com diferentes cenários. O modelo apresentou resultados positivos não apenas ao definir critérios para alocação dos produtos, mas também resultaria, caso aplicado, em redução de custos de operação com serviço de logística e melhora no tempo de trânsito do produto até o cliente.

Foram também propostos objetivos específicos, sendo estes avaliar os modelos de aprendizagem de máquina não supervisionados aderente ao caso, avaliar o impacto econômico do modelo atual, e avaliar os benefícios da utilização de modelos de aprendizagem de máquina não supervisionados para o posicionamento e dimensionamento de estoques. O primeiro objetivo foi alcançado ao comparar modelos de aprendizagem de máquina, e para o método proposto utilizar a clusterização por meio do algoritmo k-modes, que utiliza variáveis categóricas para compor o conjunto de elementos, porém diferentes variações do algoritmo podem ser empregadas para se adequar ao problema específico. O modelo de agrupamento é especialmente aderente ao contexto *fast fashion* pois por meio de produtos existentes, utilizar os mesmos critérios para associação de novos produtos aos grupos, criando uma relação de similaridade entre produtos com desempenho conhecido e novos lançamentos.

Os impactos econômicos do modelo atual exposto, utilizado pela empresa na qual este estudo se baseou para testar o modelo proposto, foram levantados a partir da comparação dos custos logísticos de entrega entre os cenários executado e

previsto. O impacto econômico no caso foi de 18%, resultado relevante frente ao montante apresentado, e considerando a conta logística como uma das mais relevantes para os custos de operação. Com isso, o segundo objetivo proposto foi também alcançado por esta pesquisa.

Por fim, os benefícios da utilização de modelos de aprendizagem de máquina não supervisionados apresenta benefícios ao processo em contexto *fast fashion* principalmente por proporcionar uma relação entre produtos com desempenho conhecido, e novos lançamentos, os quais ainda não foram apresentados ao mercado, resultando em um critério para o dimensionamento e posicionamento dos estoques ao longo da cadeia de suprimentos. Além disso, benefícios expostos na sessão de discussão dos resultados como consequência da redução de custos e *lead time* de entrega também podem ser atribuídos a utilização do aprendizado de máquina.

O modelo proposto ainda permite identificar uma possível melhor utilização dos recursos logísticos da empresa, oportunizando novos projetos de abertura de novas instalações a partir da justificativa de custos e aumento de nível de serviço oferecido. Índícios não mensurados de resultados relacionados a satisfação do cliente com a experiência oferecida no comércio eletrônico da marca com entregas mais rápidas e pedidos não fracionados também são associados ao trabalho desenvolvido.

Além das contribuições gerenciais da operação resultantes deste estudo, como a definição de um método para posicionamento dos estoques, este trabalho também apresentou contribuições acadêmicas. Neste sentido, a pesquisa promoveu avanços relacionados a utilização da técnica de clusterização adaptada para a associação de produtos inéditos no mercado com o histórico de vendas existente, e na utilização do modelo submetido a dados reais da operação, embasando futuras pesquisas na área.

As limitações apresentadas nesta pesquisa se relacionam a impossibilidade de testar o modelo na operação real, na qual suas fragilidades seriam evidenciadas oportunizando correções e avanços no modelo proposto. Questões relacionadas a expansão do escopo para práticas de *omnichannel*, integrando pontos físicos ao modelo operando como centros de distribuição, e atendimento do estoque destas lojas pelos armazéns também não foram abordadas no trabalho.

Para futuras pesquisas, além da expansão do modelo contemplando as limitações apresentadas, a interação com o canal de atacado poderia ser abordada, no qual os volumes transacionados são relevantes, e o canal se beneficia pela concentração de estoque para pedidos de reposição a partir do volume existente. Pesquisas agregando informações de adesão a campanhas de publicidade dos produtos, informações de solicitação de aviso de retorno de produtos ao estoque e visualização de produtos no site podem ser desenvolvidas.

REFERÊNCIAS

ABICALÇADOS. **Relatório setorial: Indústria de calçados do Brasil, 2024. Associação Brasileira das Indústrias de Calçados - Abicalçados** Novo Hamburgo Associação Brasileira das Indústrias de Calçados - Abicalçados, , 2024.

ACKERMANN, A. E. F.; SELBITTO, M. A. Métodos de previsão de demanda: uma revisão da literatura. **Inovar**, v. 32, n. 85, p. 83–99, 1 jan. 2022.

AGGARWAL, C. C. et al. **Fast algorithms for projected clustering**. Proceedings of the 1999 ACM SIGMOD international conference on Management of data. **Anais...** New York, NY, USA: ACM, jun. 1999.

ARMSTRONG, J. S. **Principles of Forecasting**. Boston, MA: Springer US, 2001. v. 30

BALAKRISHNAN, J. et al. Product recommendation algorithms in the age of omnichannel retailing – An intuitive clustering approach. **Computers & Industrial Engineering**, v. 115, p. 459–470, jan. 2018.

BATALHA, M. O.; SILVA, A. L. DA. Gestão de cadeias produtivas: novos aportes teóricos e empíricos. p. 249–266, 1999.

BATISTA, V. C. **PREVISÃO DE VENDAS: UM ESTUDO NO VAREJO DE VESTUÁRIO**. Porto Alegre: UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (UFRGS), 2017.

BERLING, P.; JOHANSSON, L.; MARKLUND, J. Controlling inventories in omni/multi-channel distribution systems with variable customer order-sizes. **Omega**, v. 114, p. 102745, 1 jan. 2023.

BERMAN, O.; KRASS, D.; MENEZES, M. B. C. Directed assignment vs. customer choice in location inventory models. **International Journal of Production Economics**, v. 179, p. 179–191, 1 set. 2016.

BERTOLINI, M. et al. Machine Learning for industrial applications: A comprehensive literature review. **Expert Systems with Applications**, v. 175, p. 114820, ago. 2021.

BIJMOLT, T. H. A. et al. Challenges at the marketing–operations interface in omni-channel retail environments. **Journal of Business Research**, v. 122, p. 864–874, 1 jan. 2021.

BITRAN, G.; CALDENTEY, R. An Overview of Pricing Models for Revenue Management. **Manufacturing & Service Operations Management**, v. 5, n. 3, p. 203–229, jul. 2003.

CACHON, G. P.; SWINNEY, R. The Value of Fast Fashion: Quick Response, Enhanced Design, and Strategic Consumer Behavior. **Management Science**, v. 57, n. 4, p. 778–795, abr. 2011.

CARO, F.; MARTÍNEZ-DE-ALBÉNIZ, V. The Impact of Quick Response in Inventory-Based Competition. **Manufacturing & Service Operations Management**, v. 12, n. 3, p. 409–429, jul. 2010.

CHOI, T.-M. et al. Fast fashion sales forecasting with limited data and time. **Decision Support Systems**, v. 59, n. 1, p. 84–92, mar. 2014.

COX III, J. F.; SCHLEIER JR, J. G. **Handbook da Teoria das Restrições**. Porto Alegre: BOOKMAN EDITORA LTDA, 2013.

CUI, M. Introduction to the K-Means Clustering Algorithm Based on the Elbow Method. **Geoscience and Remote Sensing**, v. 3, 2020.

DRESCH, A.; LACERDA, D. P.; ANTUNES, J. **DESIGN SCIENCE RESEARCH: Método de Pesquisa para Avanço da Ciência e Tecnologia**. Porto Alegre: Bookman ed., 2015.

ERMEL, A. P. C. et al. **Revisões da literatura: um método para a geração de conhecimento científico e tecnológico**. Rio de Janeiro: Folio Digital, 2022.

FALLAH TEHRANI, A.; AHRENS, D. Modified sequential k-means clustering by utilizing response: A case study for fashion products. **Expert Systems**, v. 34, n. 6, 20 dez. 2017.

FALQUETO, A. A.; CEZAR, L. C. SEGMENTAÇÃO VIA MACHINE LEARNING: PROPOSTA DE CLUSTERIZAÇÃO DE CONSUMIDORES DO E-COMMERCE DE UMA EMPRESA MULTINACIONAL DO VAREJO ESPORTIVO. **HOLOS**, v. 4, 28 dez. 2022.

FISHER, M. Rocket Science Retailing: The 2006 Philip McCord Morse Lecture. **Operations Research**, v. 57, n. 3, p. 527–540, jun. 2009.

FISHER, M. L.; GALLINO, S.; XU, J. J. The Value of Rapid Delivery in Omnichannel Retailing. **Journal of Marketing Research**, v. 56, n. 5, p. 732–748, 3 out. 2019.

FISHER, M.; RAJARAM, K. Accurate Retail Testing of Fashion Merchandise: Methodology and Application. **Marketing Science**, v. 19, n. 3, p. 266–278, ago. 2000.

FOSSO WAMBA, S. et al. Big data analytics in operations and supply chain management. **Annals of Operations Research**, v. 270, n. 1–2, p. 1–4, 20 nov. 2018.

GOPAL, P. R. C. et al. Impact of big data analytics on supply chain performance: an analysis of influencing factors. **Annals of Operations Research**, 27 maio 2022.

GREEN, K. C.; ARMSTRONG, J. S. Demand Forecasting: Evidence-Based Methods. **SSRN Electronic Journal**, 2012.

GUJARATI, D. N. **Basic Econometrics 4th Edition**. [s.l.: s.n.].

HAN, J.; KAMBER, M.; PEI, J. **Data Mining: Concepts and Techniques**. Waltham: Morgan Kaufmann, 2011. v. 3

HOLÝ, V.; SOKOL, O.; ČERNÝ, M. Clustering retail products based on customer behaviour. **Applied Soft Computing**, v. 60, p. 752–762, nov. 2017.

HUANG, J. Z. Clustering Categorical Data with k-Modes. Em: **Encyclopedia of Data Warehousing and Mining, Second Edition**. [s.l.] IGI Global, 2009. p. 246–250.

JAHANI, H.; JAIN, R.; IVANOV, D. Data science and big data analytics: a systematic review of methodologies used in the supply chain and logistics research. **Annals of Operations Research**, 11 jul. 2023.

JANG, J.-S. R. ANFIS: adaptive-network-based fuzzy inference system. **IEEE Transactions on Systems, Man, and Cybernetics**, v. 23, n. 3, p. 665–685, 1993.

KANTARDZIC, M. **Data Mining: Concepts, Models, Methods, and Algorithms**. 2. ed. [s.l.] John Wiley & Sons, 2011.

KIM, N. et al. Network inventory deployment for responsive fulfillment. **International Journal of Production Economics**, v. 255, p. 108664, 1 jan. 2023.

KÖK, A. G.; FISHER, M. L.; VAIDYANATHAN, R. Assortment Planning: Review of Literature and Industry Practice. Em: [s.l.: s.n.]. p. 99–153.

LÉLIS, E. C. **Administração da produção**. São Paulo: Editora Pearson, 2012.

LEMOS, F. DE O. **Metodologia para seleção de métodos de previsão de demanda**. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2006.

LIRA, D.; BARBOSA, A. C.; CAMERLENGO, L. A GESTÃO DE ESTOQUES NO MERCADO DE E-COMMERCE. **RECIMA21 - Revista Científica Multidisciplinar - ISSN 2675-6218**, v. 1, n. 1, p. e29741, 19 out. 2021.

LIU, F.; DENG, Y. Determine the Number of Unknown Targets in Open World Based on Elbow Method. **IEEE Transactions on Fuzzy Systems**, v. 29, n. 5, p. 986–995, maio 2021.

MAHESHWARI, S.; GAUTAM, P.; JAGGI, C. K. Role of Big Data Analytics in supply chain management: current trends and future perspectives. **International Journal of Production Research**, v. 59, n. 6, p. 1875–1900, 19 mar. 2021.

MANCUSO, A. C. B. **Uma investigação do desempenho de métodos de combinação de previsões: simulada e aplicada**. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2013.

MORABITO, R.; PUREZA, V. Modelagem e Simulação. Em: CAUCHICK-MIGUEL, P. A. (Ed.). **Metodologia de Pesquisa em Engenharia de Produção e Gestão de Operações**. 3. ed. Rio de Janeiro: Elsevier, 2018. p. 167–194.

MORANDI, M. I. W.; CAMARGO, L. F. R. **DESIGN SCIENCE RESEARCH: Método de Pesquisa para Avanço da Ciência e Tecnologia**. Porto Alegre: Bookman ed., 2015.

MURRAY, P. W.; AGARD, B.; BARAJAS, M. A. Forecasting Supply Chain Demand by Clustering Customers. **IFAC-PapersOnLine**, v. 48, n. 3, p. 1834–1839, 2015.

NASIRI, G. R.; KALANTARI, M.; KARIMI, B. Fast-moving consumer goods network design with pricing policy in an uncertain environment with correlated demands. **Computers & Industrial Engineering**, v. 153, p. 106997, 1 mar. 2021.

NEVES, M. M. Setor calçadista do Vale do Sinos e as estratégias de concorrência uma análise desde os anos 2000. v. 1, n. 1, 2018.

NUNES, P. DOS S. A. et al. Mapeamento e análise da cadeia de suprimentos: um estudo de caso na indústria calçadista. **Brazilian Journal of Production Engineering**, v. 9, n. 2, p. 99–112, 23 maio 2023.

PELLEGRINI, F. R.; FOGLIATTO, F. S. Passos para implantação de sistemas de previsão de demanda: técnicas e estudo de caso. **Production**, v. 11, n. 1, p. 43–64, jun. 2001.

PEREIRA, M. M.; FRAZZON, E. M. Towards a Predictive Approach for Omni-channel Retailing Supply Chains. **IFAC-PapersOnLine**, v. 52, n. 13, p. 844–850, 2019.

PEREIRA, M. M.; FRAZZON, E. M. A data-driven approach to adaptive synchronization of demand and supply in omni-channel retail supply chains. **International Journal of Information Management**, v. 57, p. 102165, 1 abr. 2021.

RAZAGHI, S.; SHOKOUHYAR, S. Impacts of big data analytics management capabilities and supply chain integration on global sourcing: a survey on firm performance. **Bottom Line**, v. 34, n. 2, p. 198–223, 2021.

SANTOS, B. F. et al. A IMPORTÂNCIA DA PREVISÃO DE VENDAS EM UM E-COMMERCE. **Dialogos em contabilidade teoria e pratica**, v. 7, 2019.

SEE-TO, E. W. K.; NGAI, E. W. T. Customer reviews for demand distribution and sales nowcasting: a big data approach. **Annals of Operations Research**, v. 270, n. 1–2, p. 415–431, 24 nov. 2018.

SHEARER, C. The CRISP-DM model: the new blueprint for data mining. **Journal of data warehousing**, v. 5, p. 13–22, 2000.

STERN, L. W.; EL-ANSARI, A.; COUGHLAN, A. **Marketing Channels**. 5. ed. [s.l.] Prentice Hall, 1996.

STUKER, T. A. **Modelo para atualização da previsão de demanda em cadeia de suprimentos de moda rápida na indústria calçadista**. São Leopoldo: Universidade do Vale do Rio dos Sinos, 2014.

THOMASSEY, S. Sales forecasts in clothing industry: The key success factor of the supply chain management. **International Journal of Production Economics**, v. 128, n. 2, p. 470–483, dez. 2010.

THOMASSEY, S.; FIORDALISO, A. A hybrid sales forecasting system based on clustering and decision trees. **Decision Support Systems**, v. 42, n. 1, p. 408–421, out. 2006.

TIWARI, S.; WEE, H. M.; DARYANTO, Y. Big data analytics in supply chain management between 2010 and 2016: Insights to industries. **Computers & Industrial Engineering**, v. 115, p. 319–330, 1 jan. 2018.

TSAO, Y. C. et al. A continuous approximation approach for the integrated facility-inventory allocation problem. **European Journal of Operational Research**, v. 222, n. 2, p. 216–228, 16 out. 2012.

WANG, S. et al. **Explainable AI for B5G/6G: Technical Aspects, Use Cases, and Research Challenges.** , mar. 2021.

WANKE, P. Formalizando uma política de estoques para a cadeia de suprimentos. **Revista Tecnológica**, 1999.

WANKE, P. et al. Fuzzy inference systems and inventory allocation decisions: Exploring the impact of priority rules on total costs and service levels. **Expert Systems with Applications**, v. 85, p. 182–193, 1 nov. 2017.

WITTEN, I. H.; FRANK, E. **Data Mining Practical Machine Learning Tools and Techniques.** 2. ed. San Francisco: Morgan Kaufmann, 2005.

XUE, H.; LIN, Y. **The omni-channel consumer segmentation method based on consumption data stream mining.** 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)(. **Anais...IEEE**, mar. 2017.

ZAMBERLAN, L. et al. O COMPORTAMENTO DO CONSUMIDOR NO VAREJO DE VESTUÁRIO: UM ESTUDO EXPLORATÓRIO DAS ATITUDES COM RELAÇÃO ÀS LOJAS. **Revista Administração**, p. 61–85, 2009.

APÊNDICE A – PROTOCOLO DA REVISÃO SISTEMÁTICA DA LITERATURA

ELEMENTO	ESCOLHA
Contexto	<p>A empresa em que este trabalho está sendo desenvolvido tem forte atuação e é consolidada no mercado de calçado brasileiro, e desde 2012 entrou para o mercado norte americano, mantendo sua produção e parte da operação e mão de obra no Brasil. Atualmente atua nas frentes B2B, mais recente, e B2C, com lojas físicas e e-commerce, em que este trabalho se concentra.</p> <p>A empresa produz os calçados no Brasil e exporta para os Estados Unidos, onde por meio de uma subcontratada, desenvolve sua operação com uma estrutura de 3 armazéns. A questão que motiva o desenvolvimento deste trabalho está relacionada com a assertividade de quais produtos e quantidade enviar para cada armazém, uma vez que pelas dimensões territoriais do país, o posicionamento estratégico dos produtos resulta em vantagem competitiva através de tempo de trânsito e custo de envio reduzido. Pretende-se com isso utilizar a relação dos dados gerados e capturados pelo canal de e-commerce durante a trajetória do cliente (visualização de produto e compra) com as características e região de localização do consumidor, para traçar um perfil das áreas atendidas preferencialmente por cada armazém, criando um modelo de dimensionamento e posicionamento dos estoques.</p>
Horizonte	Não definido.
Idiomas	Os termos de busca foram definidos na língua inglesa para alcançar uma amplitude maior de pesquisas, porém serão aceitos documentos em idioma Português e Inglês.
Objetivo da revisão	A revisão tem como objetivo central a proposta de um método de dimensionamento de estoques e posicionamento de produtos nas instalações existentes por meio da análise dos dados gerados pelos canais de venda.
Estratégia de revisão	(X) configurativa () agregativa
Crítérios de busca	<p>Crítérios de inclusão:</p> <ul style="list-style-type: none"> • Abordar o uso de <i>data analytics</i> para tomada de decisão na cadeia de suprimento. <p>Crítérios de exclusão:</p> <ul style="list-style-type: none"> • Trabalho que não atendam aos critérios de inclusão; • Estudo em duplicidade; • Trabalhos com foco em estudo de localização ótima para instalação de Centros de distribuição. • Falta de acesso ao documento completo.

ELEMENTO	ESCOLHA
Termos de busca	(model OR modeling OR simulation) AND "supply chain" AND (fashion OR retail) AND ("location-inventory" OR "inventory location" OR "stock positioning" OR "stock positioning" OR "stock sizing" OR "inventory sizing")
Fontes de busca	CAPES ScienceDirect Scopus Ebsco
Índices de busca	Título, palavra-chave e resumo.

FONTE: Desenvolvido com base em Morandi e Camargo (2015)

APÊNDICE B – DICIONÁRIO DE DADOS

Tabela	SKU		
Descrição	Tabela de SKUs e atributos		
Nome	Descrição	Tipo de dado	Restrição
SKU	Referência do SKU com 14 caracteres	Texto	Não nulo / 14 caracteres
Referência modelo	Referência do modelo representado pelos 10 primeiros dígitos do SKU	Texto	Não nulo / 10 caracteres
Referência construção	Referência da construção representada pelos 6 primeiros dígitos do SKU	Texto	Não nulo / 6 caracteres
Categoria	Nome da categoria	Texto	Não nulo
Subcategoria	Nome da subcategoria	Texto	Não nulo
Altura do salto	Altura do salto em centímetros	Numérico	Não nulo
Cor	Cor predominante do cabedal do modelo	Texto	Não nulo
Material	Material predominante do cabedal do modelo	Texto	Não nulo
Valor de venda	Valor de venda do produto	Numérico	Não nulo
Faixa de preço	Faixa de preço do modelo	Texto	Não nulo
Tipo de bico	Tipo de bico	Texto	Não nulo
Tipo de fechamento	Tipo de fechamento	Texto	Não nulo
Tipo de traseiro	Tipo de traseiro	Texto	Não nulo
Tipo de solado	Tipo de solado	Texto	Não nulo
Tipo de Salto	Tipo de Salto	Texto	Não nulo
Formato do cabedal	Descrição formato do cabedal	Texto	Não nulo
Material forro	Material predominante do forro interno do modelo	Texto	Não nulo
Season	Estação de lançamento do produto	Texto	Não nulo
Collection	Coleção de lançamento do produto	Texto	Não nulo
Marca	Marca do produto	Texto	Não nulo

Tabela	Calendário		
Descrição	Base de calendário		
Nome	Descrição	Tipo de dado	Restrição
Data	Data	Data	Não nulo
Dia	Dia	Numérico	Não nulo
Mês	Mês	Numérico	Não nulo
Ano	Ano	Numérico	Não nulo
Número da semana	Semana do ano conforme NRF (National Retail Federation)	Numérico	Não nulo
Número ano/semana	Ano e semana NRF	Numérico	Não nulo
Season	Estação comercial vigente	Texto	Não nulo

Tabela	Fato_venda		
Descrição	Tabela de informações relacionadas aos pedidos		
Nome	Descrição	Tipo de dado	Restrição
Data	Data	Data	Não nulo
Pedido	Número de referência do pedido	Numérico	Não nulo
SKU	Referência do SKU com 14 caracteres	Texto	Não nulo / 14 caracteres
Quantidade	Quantidade	Numérico	Não nulo
Valor	Valor pago	Numérico	Não nulo
Cidade_Destino	Cidade de destino	Texto	Não nulo
Estado_Destino	Estado do destino	Texto	Não nulo
País_Destino	País destino	Texto	Não nulo
Origem	Origem de envio	Texto	Não nulo
Canal	Canal de aquisição	Texto	Não nulo
Tempo de trânsito	Tempo de trânsito	Numérico	Não nulo
Valor de frete	Valor pago de frete	Numérico	Não nulo

Tabela	Mapa_país		
Descrição	Tabela de informações regionais		
Nome	Descrição	Tipo de dado	Restrição
Área	Área do estado	Texto	Não nulo
Sigla_estado	Sigla do estado	Texto	Não nulo
Estado	Nome do estado	Texto	Não nulo

**ANEXO A – RELAÇÃO DE CUSTO E TEMPO DE TRÂNSITO POR CENTRO
LOGÍSTICO E ESTADO**

	Destino	CUSTO [\$]			TEMPO DE TRÂNSITO [Dias]		
		C.L. CA	C.L. FL	C.L. NJ	C.L. CA	C.L. FL	C.L. NJ
AL	Alabama	13,6	11,9	13,1	4,5	1,8	2,7
AK	Alasca	70,3	48,7	87,8	6,6	7,0	8,5
AZ	Arizona	10,5	14,4	16,7	1,6	4,4	4,4
AR	Arkansas	13,4	12,8	14,1	4,0	2,2	2,9
CA	Califórnia	9,1	15,0	17,6	0,9	4,2	5,6
NC	Carolina do Norte	14,7	11,9	11,4	4,7	1,9	2,2
SC	Carolina do Sul	14,3	11,1	12,5	4,4	1,7	2,2
CO	Colorado	11,9	15,3	16,4	2,5	3,9	4,8
CT	Connecticut	15,9	13,4	9,8	4,7	2,5	1,1
ND	Dakota do Norte	16,6	15,1	15,7	4,3	3,8	4,2
SD	Dakota do Sul	18,0	11,5	17,8	4,1	4,0	3,6
DE	Delaware	13,7	10,2	9,3	4,7	1,9	1,3
DC	District of Columbia	14,4	10,0	10,5	4,5	1,5	1,2
FL	Flórida	14,1	9,1	13,3	4,7	0,8	3,2
GA	Geórgia	15,0	11,1	12,8	4,4	1,7	2,3
HI	Havaí	62,9	66,9	74,8	3,7	8,2	7,7
ID	Idaho	11,9	14,9	20,6	2,0	6,1	5,8
IL	Illinois	12,9	13,2	12,2	4,4	2,8	2,2
IN	Indiana	14,0	13,6	13,7	5,0	2,0	2,3
IA	Iowa	12,5	12,4	14,0	5,2	2,8	3,4
KS	Kansas	13,2	14,4	14,1	4,7	2,7	3,2
KY	Kentucky	14,1	12,9	12,3	4,7	2,2	2,2
LA	Louisiana	12,9	11,7	14,8	4,3	1,8	3,3
ME	Maine	16,0	10,8	11,7	5,5	3,6	2,1
MD	Maryland	14,4	10,8	10,2	4,7	1,7	1,1
MA	Massachusetts	14,1	12,9	10,2	5,8	2,8	1,2
MI	Michigan	14,5	12,3	11,2	4,9	2,7	2,1
MN	Minnesota	12,6	14,0	12,8	4,2	2,9	3,1
MS	Mississippi	14,0	12,8	15,3	3,1	1,7	2,8
MO	Missouri	14,4	14,1	14,4	4,1	2,8	2,6
MT	Montana	18,0	18,1	20,8	2,8	7,4	5,1
NE	Nebraska	12,6	15,7	15,7	4,8	2,7	3,3
NV	Nevada	9,4	14,0	17,9	1,0	4,5	4,7
NH	Nova Hampshire	16,3	12,4	11,6	5,5	3,2	1,5
NY	Nova Iorque	14,8	13,1	9,3	4,6	2,7	1,1
NJ	Nova Jérsei	14,2	12,2	9,3	4,7	2,6	1,2
NM	Novo México	12,1	12,4	16,6	2,4	3,2	5,1
OH	Ohio	13,4	12,3	11,2	4,5	1,9	2,1
OK	Oklahoma	13,0	14,4	15,2	2,7	2,8	3,3

OR	Oregon	11,4	13,0	16,4	3,1	5,1	6,4
PA	Pensilvânia	14,5	11,3	9,5	4,6	1,9	1,3
RI	Rhode Island	14,6	13,9	10,3	5,5	2,6	1,1
TN	Tennessee	14,0	11,3	12,9	4,2	1,8	2,4
TX	Texas	13,2	13,0	15,1	2,8	2,4	3,2
UT	Utah	11,6	15,1	17,1	4,4	5,8	6,1
VT	Vermont	19,1	14,1	11,2	6,1	3,0	1,4
VA	Virgínia	14,1	11,5	10,1	4,6	1,8	2,1
WV	Virgínia Ocidental	15,7	10,2	10,7	5,1	2,1	2,4
WA	Washington	11,4	13,7	17,2	2,3	5,0	6,6
WI	Wisconsin	13,3	13,7	12,4	4,6	2,4	2,8
WY	Wyoming	16,8	14,6	23,9	3,1	6,0	5,6