

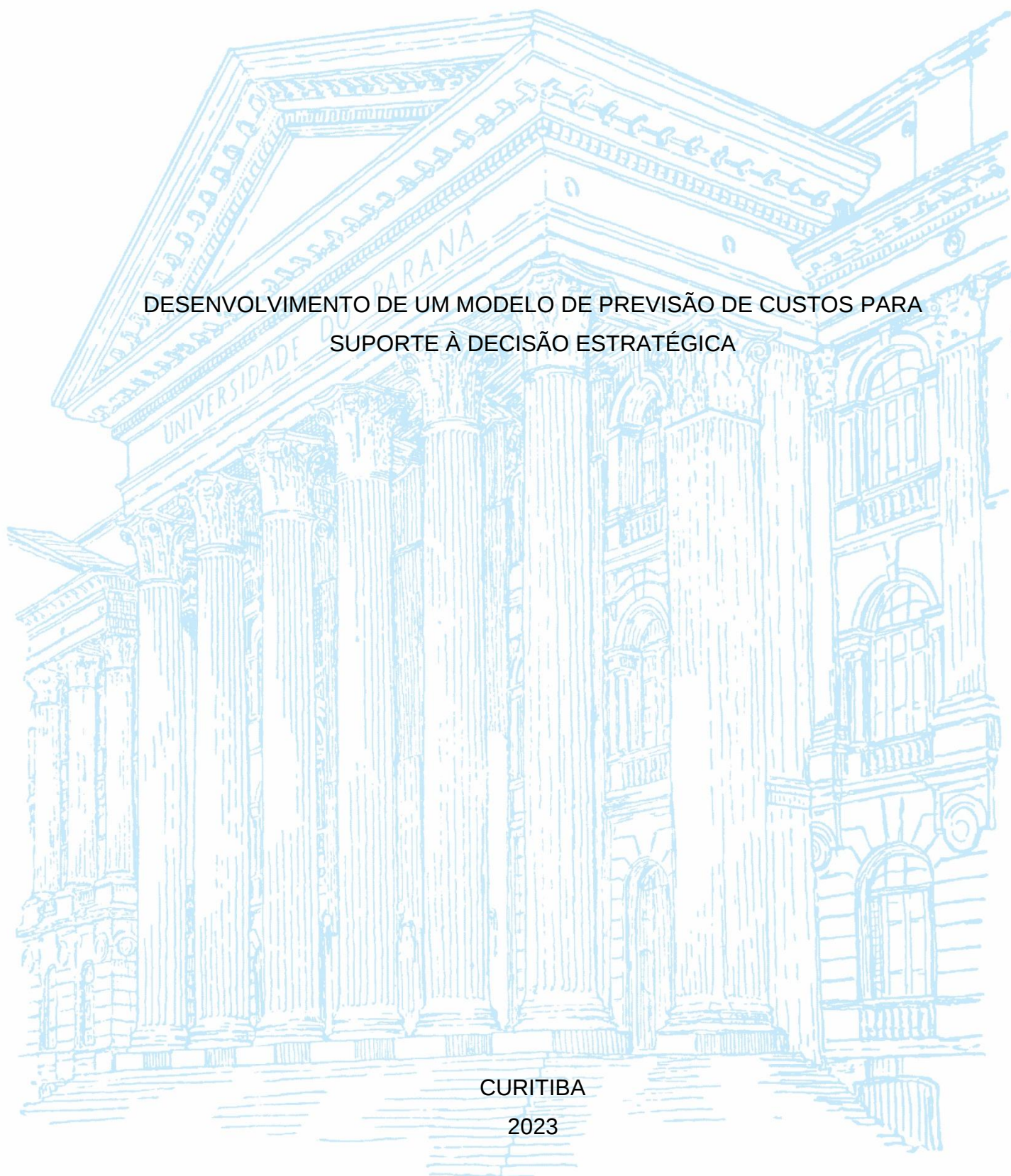
UNIVERSIDADE FEDERAL DO PARANÁ

AMANDA GLUS SCHARNOSKI

DESENVOLVIMENTO DE UM MODELO DE PREVISÃO DE CUSTOS PARA  
SUPORTE À DECISÃO ESTRATÉGICA

CURITIBA

2023



AMANDA GLUS SCHARNOSKI

DESENVOLVIMENTO DE UM MODELO DE PREVISÃO DE CUSTOS PARA  
SUPORTE À DECISÃO ESTRATÉGICA

TCC apresentado ao curso de Graduação em Engenharia de Produção, Setor de Tecnologia, Universidade Federal do Paraná, como requisito à obtenção do título de Bacharel em Engenharia de Produção.

Orientadora: Prof.<sup>a</sup> Dra. Silvana Pereira Detro

CURITIBA

2023

## RESUMO

O constante desafio de competição empresarial no mercado global pode ser significativamente intensificado diante das volatilidades nos preços de matérias primas, especialmente no setor de *commodities*, como o petróleo e seus derivados. Nesse cenário, a previsão de custos emerge no contexto do *Business Intelligence* como uma fusão poderosa de dados e inteligência analítica, permitindo uma gestão eficaz de custos e consequente rentabilidade a longo prazo. O presente trabalho propõe o desenvolvimento de um modelo de previsão de custos de modo a oferecer suporte à uma empresa petrolífera para a tomada de decisões estratégicas. Destaca-se nesse processo a importância da combinação de métodos estatísticos, dados de mercado e intuição humana para previsões mais robustas. Para tanto, utilizaram-se ferramentas como o Alteryx e Tableau. A metodologia CRISP-DM selecionada proporcionou uma abordagem sistemática e direcionamento para o projeto, resultando em um erro de previsão percentual médio de 3,2% para um intervalo de três meses no futuro. A implementação do modelo gerou vantagens imediatas para a organização no que se refere à possibilidade de adoção de estratégias de mitigação de custos informadas e resposta ágil às mudanças de mercado. Propostas futuras incluem análises quantitativas dos benefícios capturados e aprimoramento do modelo por meio do estudo detalhado dos demais componentes de custo.

Palavras-chave: Previsão de custos. Inteligência de Negócios. CRISP-DM. Cadeia de Suprimento.

## **ABSTRACT**

The constant challenge of business competition in the global market can be significantly intensified given the volatility in the prices of raw materials, especially in the commodities sector, such as oil and its derivatives. In this scenario, cost forecasting emerges in the context of Business Intelligence as a powerful fusion of data and analytical intelligence, allowing effective cost management and consequent long-term profitability. The present work proposes the development of a cost forecasting model to support an oil company in strategic decision-making. The importance of combining statistical methods, market data, and human intuition for more robust predictions is highlighted in this process. Tools such as Alteryx and Tableau were employed for this purpose. The chosen CRISP-DM methodology provided a systematic approach and guidance for the project, resulting in an average percent prediction error of 3.2% for a three-month forecast lag. The implementation of the model immediately benefited the organization in terms of the ability to adopt informed cost mitigation strategies and respond promptly to market changes. Future proposals include quantitative analyses of captured benefits and improvement of the model through a detailed study of the other cost components.

**Keywords:** Cost forecasting. Business Intelligence. CRISP-DM. Supply Chain.

## LISTA DE FIGURAS

|                                                                                                        |    |
|--------------------------------------------------------------------------------------------------------|----|
| FIGURA 1 - Elementos da Cadeia de Suprimentos .....                                                    | 14 |
| FIGURA 2 - Estrutura da Cadeia de Suprimentos.....                                                     | 15 |
| FIGURA 3 - Técnicas de Análise de Dados .....                                                          | 18 |
| FIGURA 4 - Estrutura de uma Cadeia de Suprimentos orientada por Dados .....                            | 21 |
| FIGURA 5 - Exemplo de Arquitetura de BI .....                                                          | 23 |
| FIGURA 6 - Fases da metodologia CRISP-DM para a mineração de dados .....                               | 26 |
| FIGURA 7 - Visão geral das tarefas do CRISP-DM e seus resultados .....                                 | 29 |
| FIGURA 8 - Estrutura de Custos da empresa .....                                                        | 31 |
| FIGURA 9 - Divisão do Custo dos Produtos Vendidos na empresa.....                                      | 32 |
| FIGURA 10 - Consumo dos Grupos de Óleos Básicos na empresa .....                                       | 33 |
| FIGURA 11 - <i>Workflow</i> em Alteryx para calcular as projeções dos óleos básicos ....               | 38 |
| FIGURA 12 - <i>Workflow</i> em Alteryx dos modelos ARIMA e ETS.....                                    | 40 |
| FIGURA 13 - Agregação dos valores MAPE por intervalo de precisão .....                                 | 46 |
| FIGURA 14 - <i>Workflow</i> em Alteryx para calcular as projeções de custos dos<br>lubrificantes ..... | 48 |
| FIGURA 15 - Métricas MAPE de avaliação do modelo .....                                                 | 50 |
| FIGURA 16 - Integração e formação de custo na Cadeia de Suprimentos .....                              | 53 |
| FIGURA 17 - Evolução e tendência do preço dos grupos de óleos básicos .....                            | 54 |
| FIGURA 18 - Análise mensal dos componentes de custo dos lubrificantes.....                             | 55 |
| FIGURA 19 - Impacto médio dos componentes de custo no produto final .....                              | 57 |
| FIGURA 20 - Impacto médio acumulado dos componentes de custo no produto final<br>.....                 | 58 |

## LISTA DE GRÁFICOS

|                                                                                      |    |
|--------------------------------------------------------------------------------------|----|
| GRÁFICO 1 - Resultados MAPE obtidos por séries temporais em janeiro de 2022 .        | 43 |
| GRÁFICO 2 - Resultados MAPE obtidos por dados de mercado em janeiro de 2022<br>..... | 44 |
| GRÁFICO 3 - Resultados MAPE obtidos por séries temporais em outubro de 2022          | 45 |
| GRÁFICO 4 - Resultados MAPE obtidos por dados de mercado em outubro de 2022<br>..... | 45 |
| GRÁFICO 5 - Relação entre o custo dos óleos básicos e o custo dos lubrificantes .    | 55 |

## LISTA DE QUADROS

|                                                                             |    |
|-----------------------------------------------------------------------------|----|
| QUADRO 1 - Dados gerados nos estágios da Cadeia de Suprimento.....          | 22 |
| QUADRO 2 - Colunas extraídas do SAP HANA como <i>input</i> do modelo .....  | 35 |
| QUADRO 3 - Colunas extraídas do Snowflake como <i>input</i> do modelo ..... | 36 |

## **LISTA DE TABELAS**

|                                                                         |    |
|-------------------------------------------------------------------------|----|
| TABELA 1 - Comparação dos resultados MAPE obtidos por cada modelo ..... | 42 |
| TABELA 2 - Resultados MAPE obtidos em janeiro de 2022 .....             | 43 |
| TABELA 3 - Resultados MAPE obtidos em outubro de 2022 .....             | 44 |

## SUMÁRIO

|                                            |           |
|--------------------------------------------|-----------|
| <b>1 INTRODUÇÃO .....</b>                  | <b>10</b> |
| 1.1 OBJETIVOS .....                        | 11        |
| 1.2 ESTRUTURA DO DOCUMENTO .....           | 11        |
| <b>2 REVISÃO DE LITERATURA .....</b>       | <b>12</b> |
| 2.1 ÓLEOS LUBRIFICANTES E PETRÓLEO .....   | 12        |
| 2.2 CADEIA DE SUPRIMENTO .....             | 13        |
| 2.3 ANÁLISE DE DADOS.....                  | 16        |
| 2.4 INTELIGÊNCIA DE NEGÓCIOS .....         | 22        |
| <b>3 METODOLOGIA DA PESQUISA.....</b>      | <b>25</b> |
| <b>4 APRESENTAÇÃO DOS RESULTADOS .....</b> | <b>30</b> |
| 4.1 COMPREENSÃO DA EMPRESA ESTUDADA .....  | 30        |
| 4.2 COMPREENSÃO DOS DADOS.....             | 33        |
| 4.3 PREPARAÇÃO DOS DADOS .....             | 36        |
| 4.4 MODELAGEM .....                        | 37        |
| 4.5 AVALIAÇÃO .....                        | 49        |
| 4.6 IMPLEMENTAÇÃO .....                    | 51        |
| <b>5 CONSIDERAÇÕES FINAIS .....</b>        | <b>60</b> |
| <b>REFERÊNCIAS.....</b>                    | <b>61</b> |

## 1 INTRODUÇÃO

Na situação econômica atual, as empresas enfrentam uma competição intensa pelo mercado consumidor. A volatilidade nos preços das *commodities* que marcou o cenário global nos últimos anos, especialmente no setor de petróleo e seus derivados, intensificou esse ambiente desafiador. Diversos setores, incluindo a indústria de lubrificantes, enfrentam a constante dificuldade de gerenciar os custos de suas matérias-primas, buscando equilibrar a qualidade dos produtos ofertados e a lucratividade do negócio. Dessa forma, o sucesso ou fracasso de uma organização está intrinsicamente ligado às estratégias eficazes de gestão de custos e adaptabilidade (PORTER, 1997).

Para Mahanty (2014), custo é definido como o valor monetário dos recursos utilizados para produzir um bem ou serviço. Eles são elementos cruciais que afetam a rentabilidade e a viabilidade a longo prazo de qualquer negócio. Nesse cenário, a previsão de custos emerge no contexto do *Business Intelligence* (BI) como uma fusão poderosa de dados e inteligência analítica. Utilizando técnicas de análise preditiva, como ARIMA e ETS, as organizações podem extrair *insights* valiosos das informações históricas para antecipar tendências de custos (DAVENPORT e HARRIS, 2017).

Embora métodos estatísticos sejam fundamentais, a interação entre a objetividade por eles proporcionada e a intuição e experiência humana leva a previsões mais robustas, fundamentadas tanto em fatores quantitativos quanto em qualitativos. Segundo Armstrong (2001), a combinação de diferentes tipos de previsões e conhecimentos, como dados históricos, econômicos e de especialistas, pode melhorar substancialmente a acurácia das previsões. Em 30 comparações empíricas por ele realizadas, a redução do erro prévio para previsões combinadas com pesos iguais foi, em média, de cerca de 12,5%, variando de 3 a 24%.

Este trabalho não apenas contribui para o campo da previsão de custos de matérias-primas e produtos finais, mas também destaca a importância da colaboração entre métodos estatísticos, dados de mercado externos e intuição humana no processo de tomada de decisões empresariais. Ao oferecer uma abordagem holística e interdisciplinar para a gestão de custos e estratégias de precificação, este estudo demonstra como a análise de dados pode transformar desafios em oportunidades, proporcionando uma base sólida para decisões estratégicas no dinâmico mercado de *commodities*.

## 1.1 OBJETIVOS

Esse trabalho tem como objetivo principal estruturar e implementar um modelo de previsão de custos de modo a oferecer suporte à empresa para a tomada de decisões estratégicas.

Para alcançar os resultados esperados, faz-se necessário o mapeamento e a realização dos seguintes objetivos específicos:

- Identificar o problema da empresa e definir o modelo ideal para o contexto;
- Estruturar e implementar o modelo de previsão de custos;
- Comparar e avaliar a performance do modelo;
- Criar *dashboards* para a visualização dos dados;
- Extrair informações úteis dos dados para suportar a tomada de decisões.

## 1.2 ESTRUTURA DO DOCUMENTO

Este documento está dividido em cinco capítulos. O primeiro e presente capítulo contém a introdução ao trabalho proposto, contextualizando o mesmo em termos de benefícios e objetivos a serem alcançados.

O segundo capítulo apresenta o referencial teórico consultado para a definição da proposta de pesquisa, apresentando conceitos fundamentais sobre Lubrificantes, Óleos Básicos, Cadeia de Suprimentos, Análise de Dados e Inteligência de Negócios. O terceiro capítulo descreve a metodologia de pesquisa utilizada no trabalho e a descrição das ferramentas utilizadas em cada etapa. No quarto capítulo, são registrados os detalhamentos técnicos e resultados obtidos com a implementação do modelo. Por fim, no quinto capítulo são apresentadas as considerações finais da autora, bem como conclusões e sugestões para trabalhos futuros.

## 2 REVISÃO DE LITERATURA

Este capítulo apresenta uma revisão de literatura sobre os principais conceitos relacionados à pesquisa, como Óleos lubrificantes, Cadeia de Suprimento, Análise de Dados, Previsão de Custos e Inteligência de Negócios, proporcionando um embasamento teórico necessário para a compreensão do desenvolvimento do modelo de análise de dados que será abordado posteriormente.

### 2.1 ÓLEOS LUBRIFICANTES E PETRÓLEO

Os lubrificantes são formulados pela mistura de óleos básicos e aditivos para atender a uma série de especificações de desempenho, que se relacionam às propriedades químicas e físicas do óleo formulado. Evidentemente, o desempenho do lubrificante é determinado pelos materiais utilizados em sua formulação. Os lubrificantes possuem várias funções, mas a principal é fornecer uma camada fluida para separar superfícies móveis, além de remover calor e partículas de desgaste, minimizando o atrito (FOX et al., 2010).

A necessidade de adicionar aditivos aos óleos básicos surgiu devido ao avanço tecnológico dos equipamentos, que passaram a exigir uma lubrificação mais avançada, incapaz de ser atingida pelo óleo mineral puro. A incorporação de aditivos nos lubrificantes tem o propósito de conferir características essenciais, como dispersão, detergência, resistência ao desgaste, proteção antioxidante, anticorrosiva, antiespumante, ajuste de viscosidade, emulsificação, redução do ponto de fluidez, adesividade, entre outras. A quantidade de aditivos recomendada pelos fornecedores varia, em média, de 0,5% a 28% em volume (SILVEIRA et al., 2006).

A maioria dos óleos básicos de lubrificantes é produzida por meio do refinamento do petróleo bruto. O petróleo desempenha um papel crucial na economia global, representando cerca de dois terços do consumo de energia mundial. Seu preço é determinado principalmente pela oferta e demanda, mas é fortemente influenciado por eventos irregulares como condições climáticas, estoques, crescimento econômico e fatores políticos. Devido à complexidade e volatilidade desse mercado, a previsão do preço do petróleo é desafiadora, apesar de sua importância significativa (XIE et al., 2006).

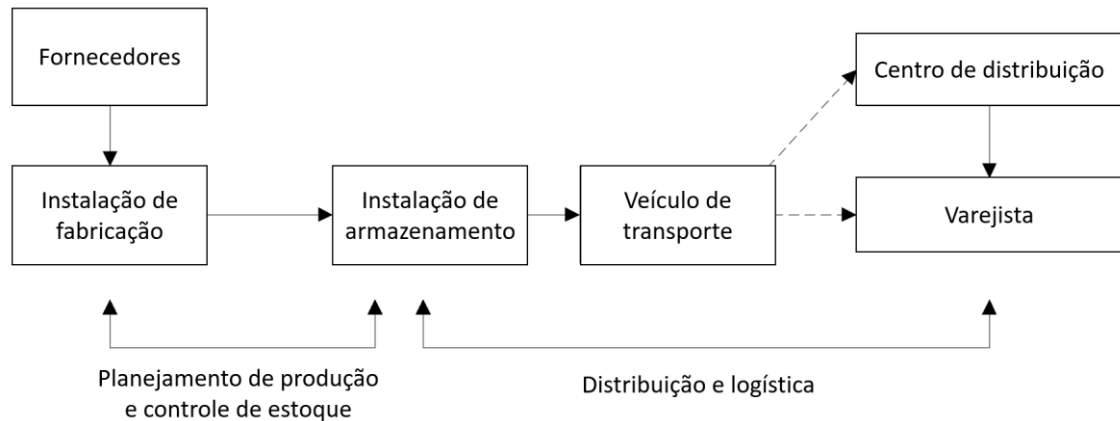
O American Petroleum Institute (API) (2020) categoriza os óleos básicos de lubrificantes em grupos numerados, inicialmente definidos como um meio para desenvolver diretrizes de intercâmbio em óleos de motor. Essas categorias são amplamente utilizadas nas especificações de lubrificantes em todo o mundo. São elas:

- a) Grupo I: Óleos básicos refinados a partir de petróleo bruto por meio de processos de solvente ou ácido. São óleos básicos de qualidade mais baixa, geralmente utilizados em aplicações simples;
- b) Grupo II: Óleos básicos refinados a partir de petróleo bruto por meio de processos de hidro craqueamento. São óleos básicos de qualidade intermediária, com melhores propriedades do que os do Grupo I;
- c) Grupo III: Óleos básicos refinados por hidro craqueamento, mas com um processo mais rigoroso, resultando em óleos básicos de alta qualidade. Alguns óleos básicos do Grupo III são quase equivalentes aos óleos sintéticos;
- d) Grupo IV: Óleos básicos totalmente sintéticos, conhecidos como polialfaolefinas (PAOs). São produzidos quimicamente e oferecem alto desempenho em uma ampla faixa de temperaturas;
- e) Grupo V: Compreende todos os óleos básicos que não se enquadram nas categorias anteriores, incluindo óleos sintéticos não PAO, como por exemplo ésteres, e óleos básicos de origem não convencional, como óleos vegetais.

## 2.2 CADEIA DE SUPRIMENTO

De acordo com o Council of Supply Chain Management Professionals (CSCMP, 2023), a cadeia de suprimentos engloba todas as atividades envolvidas no fluxo e na transformação de mercadorias, desde a matéria-prima até o consumidor final, bem como os fluxos de informações associados e a cooperação e coordenação entre fornecedores, fabricantes, distribuidores, varejistas e clientes finais. Beamon (1998) propõe que, em seu nível mais elevado, a cadeia de suprimentos é composta por dois processos fundamentais e interligados, que fornecem a estrutura básica para a movimentação e conversão de matérias-primas em produtos finais: (1) o Processo de Planejamento da Produção e Controle de Inventário, e (2) o Processo de Distribuição e Logística, ilustrados na Figura 1:

FIGURA 1 - Elementos da Cadeia de Suprimentos



FONTE: Adaptado de Beamon, 1998

O Gerenciamento da Cadeia de Suprimentos ou *Supply Chain Management* (SCM) é um conjunto de decisões e atividades sincronizadas, utilizadas para integrar de forma eficaz os fornecedores, fabricantes, transportadores, armazéns, varejistas e clientes, de modo que o produto ou serviço certo seja distribuído nas quantidades adequadas, nos locais apropriados e no momento correto, a fim de minimizar os custos em toda a cadeia e ao mesmo tempo satisfazer os requisitos de nível de serviço ao cliente (MISRA et al., 2010).

Para Mentzer et al. (2001), as funções comerciais tradicionais de *marketing*, vendas, pesquisa e desenvolvimento, previsão, produção, aquisição, logística, tecnologia da informação, finanças e atendimento ao usuário gerenciam e concretizam os fluxos que fornecem valor ao cliente e resultam em vantagens competitivas. Os fluxos, por sua vez, são direcionais e concernem aos produtos, serviços, recursos financeiros, informações, demanda e previsões (Figura 2).

FIGURA 2 - Estrutura da Cadeia de Suprimentos



FONTE: Adaptado de Mentzer et al., 2001

Os custos relacionados à cadeia de suprimentos afetam diretamente o Custo dos Produtos Vendidos, em inglês, *Cost of Goods Sold* (COGS). Em finanças e economia, COGS é uma métrica financeira que representa os custos diretamente associados à produção ou aquisição dos bens que uma empresa vende durante um determinado período de tempo. Esses custos incluem o custo dos materiais, mão de obra direta e os custos de produção indiretos associados à fabricação ou aquisição dos produtos. O COGS é subtraído das receitas de vendas para calcular o lucro bruto de uma empresa, sendo uma medida fundamental para o cálculo da rentabilidade (MOHEB-ALIZADEH e HANDFIELD, 2018).

A SCM evoluiu significativamente ao longo das décadas como consequência da pressão da intensa competição global e tecnologia da informação (TAN, 2002), tornando-se um elemento consolidado na educação em gestão e um campo profissional próprio (NEW e WESTBROOK, 2004). Observou-se a passagem de uma abordagem tradicional de competição entre fornecedores e clientes para um sistema de compartilhamento de informações, sincronização de processos e tomada de decisões em conjunto entre os membros da cadeia, objetivando torná-las ágeis, flexíveis e responsivas (LAMBERT, 1998). Para Autry e Golicic (2010), é inquestionável a existência de uma relação proativa entre comprador e fornecedor, além da integração dos demais elos da cadeia de suprimentos.

O foco primordial de redução de gastos da *Supply Chain* também foi redirecionado para a eficiência de custos, confiabilidade e previsibilidade do futuro, mudança influída e suportada pelo surgimento de inúmeros sistemas de gerenciamento, como Sistemas de Planejamento de Requisitos de Materiais (MRP), Sistemas de Planejamento de Recursos Empresariais (ERP), Sistemas de Gerenciamento de Transporte (TMS) e Sistemas de Gerenciamento de Armazém (WMS), projetados para facilitar a gestão eficiente de todas as atividades envolvidas na cadeia de suprimentos (CHANDRA e KUMAR, 2000).

Historicamente, muitas cadeias de suprimentos são otimizadas apenas para eficiência ao invés de resiliência, o que as torna vulneráveis a falhas em cascata quando expostas a ameaças sistêmicas. A resiliência de uma SC refere-se à sua capacidade de resistir e se recuperar de interrupções, choques ou eventos inesperados, mantendo suas funções essenciais e atendendo às demandas dos clientes (TANG, 2006). A magnitude sem precedentes da pandemia do Covid-19 evidenciou essa fragilidade nas cadeias de suprimentos globais, que sofreram impacto substancial, apresentando interrupções generalizadas na produção, distribuição e logística, além de uma mudança significativa nos padrões de consumo (GOLAN et al., 2020).

Murino et al. (2011) levantam diversas estratégias que objetivam aumentar a resiliência de uma SC, como a busca de uma diversificação de suas fontes de suprimento, redução da dependência de um único fornecedor ou país e melhorias na capacidade de resposta a interrupções. A gestão de riscos, a flexibilidade operacional e a implementação de estratégias de contingência, como a capacidade de realocar recursos e alterar fluxos de trabalho, se tornaram fatores críticos e uma tendência emergente nessa esfera (GIUNIPERO e ELTANTAWY, 2004).

## 2.3 ANÁLISE DE DADOS

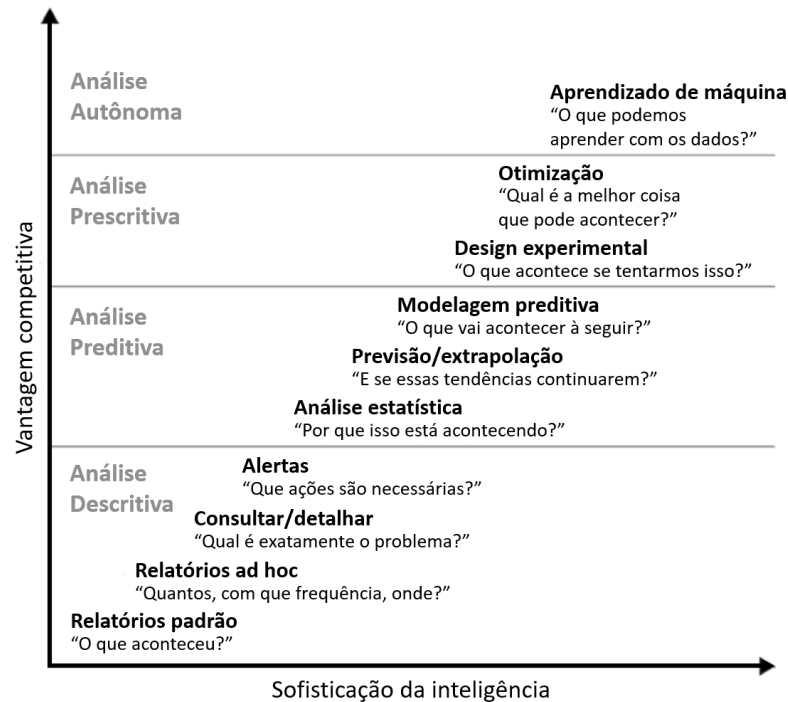
A necessidade de entender e interpretar informações existe desde os primórdios da humanidade, mas a solidificação sistemática da Análise de Dados como uma disciplina formal ocorreu na década de 1960, com o desenvolvimento de técnicas estatísticas mais avançadas e o aumento do poder de processamento dos computadores (HAND, 2004).

A Análise de Dados é uma área multidisciplinar que envolve a aplicação de técnicas estatísticas, matemáticas e computacionais. Para Abdul-Jabbar (2022), ela consiste no processo de inspeção de dados brutos visando a extração de informações relevantes, geração de conhecimento e apoio a tomada de decisões. A análise de dados é um campo abrangente, dinâmico e complexo devido à variedade de tipos de dados e ao seu crescimento exponencial. Além disso, o propósito da análise diverge de acordo com o tipo de aplicação requerida (RAJARAMAN, 2016).

Segundo Davenport e Harris (2017), as técnicas de análise podem ser categorizadas em quatro tipos principais: descritivas, preditivas, prescritivas e autônomas, sendo que o potencial de vantagem competitiva aumenta com a utilização de análises mais sofisticadas. A análise descritiva fornece um resumo dos dados históricos disponíveis de diversas fontes, geralmente por meio de estatísticas descritivas, gráficos e visualizações, buscando identificar padrões e tendências. A análise preditiva visa fazer previsões sobre eventos futuros com base nos dados históricos disponíveis, utilizando técnicas estatísticas, modelos matemáticos e algoritmos de aprendizado de máquina. A análise prescritiva, por meio de técnicas avançadas de modelagem, simulações e algoritmos de otimização, gera recomendações e ações para otimizar os resultados desejados, identificando a melhor solução possível entre várias disponíveis. A análise autônoma refere-se à capacidade de um sistema ou algoritmo de realizar análises e tomar decisões de forma automática e independente, sem intervenção humana direta. Essa abordagem envolve o uso de inteligência artificial, aprendizado de máquina e algoritmos avançados para automatizar processos analíticos complexos, sendo especialmente útil em situações em que grandes volumes de dados estão envolvidos ou quando as decisões precisam ser tomadas em tempo real (DAVENPORT e HARRIS, 2017).

Conforme ilustrado na Figura 3, cada uma dessas abordagens levanta uma série de questões sobre as atividades comerciais de uma organização.

FIGURA 3 - Técnicas de Análise de Dados



FONTE: Adaptado de Davenport e Harris, 2017

Um termo importante quando se tratando de técnicas preditivas é a série temporal. Uma série temporal é frequentemente o resultado da observação de um processo subjacente, cujos valores são coletados a partir de medições feitas em instantes de tempo uniformemente espaçados, de acordo com uma determinada taxa de amostragem. Uma série temporal pode, assim, ser definida como uma coleção de observações feitas sequencialmente ao longo do tempo. A série pode ser univariada, analisando uma única variável, ou multivariada, quando abrange simultaneamente múltiplas dimensões dentro do mesmo intervalo de tempo (ESLING e AGON, 2012).

A previsão de séries temporais é uma área importante na qual observações passadas da mesma variável são coletadas e analisadas para desenvolver um modelo que descreva seu relacionamento subjacente. O modelo é então usado para extrapolar a série temporal para o futuro. Esta abordagem de modelagem é particularmente útil quando há pouco conhecimento disponível sobre o processo gerador de dados ou quando não existe um modelo explicativo satisfatório que relacione a variável de previsão a outras variáveis explicativas. Muito esforço tem sido dedicado ao longo das últimas décadas para o desenvolvimento e aprimoramento de modelos de previsão de séries temporais (ZHANG, 2003).

Em uma série de artigos e um livro subsequente, Box e Jenkins (1970) apresentaram e ilustraram uma metodologia para construir modelos de previsão autorregressivos integrados de médias móveis, do inglês, *Autoregressive Integrated Moving Average* (ARIMA). O trabalho deles teve um impacto enorme na teoria e prática da análise e previsão modernas de séries temporais (GOOIJER e HYNDMAN, 2006). Nesse modelo, o valor futuro de uma variável é assumido como uma função linear de várias observações passadas e erros aleatórios. Ou seja, o processo subjacente que gera a série temporal tem a seguinte forma (Equação 1):

$$y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (1)$$

Onde  $y_t$  e  $\varepsilon_t$  são o valor real e o erro aleatório no período  $t$ , respectivamente;  $\phi_i (i = 1, 2, \dots, p)$  e  $\theta_j (j = 1, 2, \dots, q)$  são parâmetros do modelo,  $p$  e  $q$  são números inteiros e frequentemente referidos como ordens do modelo. Os erros aleatórios,  $\varepsilon_t$ , são assumidos como independentes e identicamente distribuídos, com uma média de zero e uma variância constante de  $\sigma^2$ .

Enquanto o modelo ARIMA descreve a autocorrelação nos dados, o modelo de Suavização Exponencial, do inglês *Exponential Smoothing Model* (ETS) se baseia em descrever a tendência nos dados, como sugerido por Holt (1957) e Winters (1960) (SHRIVASTRI et al., 2022). A ideia por trás do modelo é que as previsões calculadas são a média ponderada das observações passadas, e o peso decai exponencialmente com o tempo, ou seja, as observações atuais têm pesos maiores em comparação com as anteriores. Isso gera previsões mais precisas em comparação com métodos de média simples (JAIN e MALLICK, 2017).

Com o advento da tecnologia moderna, grandes volumes de dados estão sendo gerados em alta velocidade. Nesse cenário, é substancial reconhecer o papel dos Sistemas de Informação (SI) e das Tecnologias da Informação (TI), que fornecem diferentes ferramentas e técnicas para o acesso a grandes quantidades de dados gerados por meio de plataformas eletrônicas (AYDINER et al., 2019). Por outro lado, também emergiram adversidades, como a denominada “Sobrecarga de informações”, na qual indivíduos ou organizações são expostos a uma quantidade excessiva de informações que excede sua capacidade de processá-las e absorvê-las de forma

eficaz, resultando em dificuldades para extrair *insights* relevantes (EPPLER e MENGIS, 2004).

Frequentemente os termos "*Big Data*" e "*Big Data Analytics*" são usados de forma intercambiável, refletindo a opinião comum de que *Big Data* não se refere apenas ao problema da sobrecarga de informações, mas também aos instrumentos analíticos utilizados para gerenciar o fluxo de dados e transformá-lo em fontes produtivas e utilizáveis (MALTBY, 2011). Dessa forma, atribuindo importância significativa a outros atributos além de volume, Russom et al. (2011) definem *Big Data* em termos dos "Três V's": 'Volume', 'Velocidade' e 'Variedade'. Embora essa definição seja relativa e varie de acordo com diversos fatores, 'Volume' se refere à magnitude dos dados, que podem consumir uma grande quantidade de armazenamento. 'Velocidade' é a frequência de geração ou entrega de dados, bem como à rapidez com que eles devem ser analisados e utilizados. 'Variedade' destaca que os dados são gerados a partir de uma grande diversidade de fontes e em inúmeros formatos, contendo campos multidimensionais estruturados, semiestruturados e não estruturados (RUSSOM et al., 2011). Posteriormente, alguns autores adicionaram a essa definição mais um 'V', representando 'Valor', a fim de enfatizar a importância de extrair benefícios da análise dos grandes volumes de dados disponíveis (GANTZ e REINSEL, 2012).

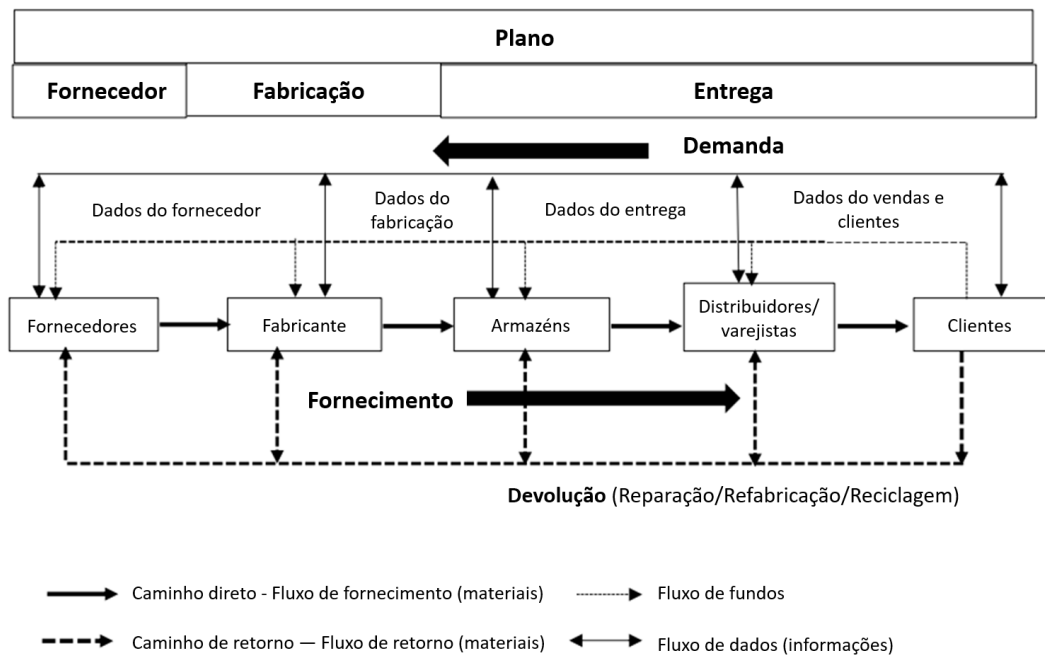
É evidente que o conceito de *Big Data* não é novo, mas sim a capacidade de aproveitar analiticamente seu potencial por meio de novas tecnologias, que, ao contrário dos sistemas tradicionais, são capazes de lidar instantaneamente com enormes volumes de dados (GANDOMI e HAIDER, 2015). White (2012) também sugeriu que uma quinta dimensão, 'Veracidade', fosse adicionada. Essa refere-se à qualidade e ao nível de confiança das fontes, e é abordada por meio de ferramentas e análises desenvolvidas para o gerenciamento e mineração de dados incertos.

Dentre suas incontáveis aplicações, historicamente, a análise de dados desempenhou um papel significativo na SCM ao reduzir os custos operacionais, tornando-se cada vez mais essencial com o surgimento de sistemas de planejamento de recursos empresariais (ERP) e, mais recentemente, com aplicações de *Big Data* (SOUZA, 2014). Como disciplina, a Análise da Cadeia de Suprimentos ou *Supply Chain Analytics* (SCA) refere-se à aplicação de técnicas analíticas e tecnologias para entender, otimizar e tomar decisões estratégicas relacionadas às operações de uma

cadeia de suprimentos, envolvendo o uso de dados e informações coletadas ao longo da cadeia (TRKMAN et al., 2010).

Na Figura 4, Biswas e Sen (2016) retratam uma estrutura típica de uma cadeia de suprimentos orientada por dados, na qual esses são classificados em categorias com base no processo e na natureza de seu uso.

FIGURA 4 - Estrutura de uma Cadeia de Suprimentos orientada por Dados



FONTE: Adaptado de Biswas e Sem, 2016

O Quadro 1 apresenta exemplos de diferentes tipos de dados que podem ser gerados em cada um desses estágios:

QUADRO 1 - Dados gerados nos estágios da Cadeia de Suprimento

| <b>Estágio</b>                       | <b>Dados gerados</b>                                                                                                                                                                                                                                                                                                                                                       |
|--------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Fornecedor                           | Inventário, Prazo de entrega, Especificações técnicas, Qualidade, Custo das matérias-primas, Termos de pagamento, Variações de preço, Desempenho, Pontualidade nas entregas, Conformidade com requisitos contratuais, Rastreabilidade                                                                                                                                      |
| Fabricante                           | Quantidade produzida, Tempo de ciclo, Eficiência da produção, Qualidade, Registros de manutenção, Planejamento da produção, Previsão de demanda, Programação da produção, Alocação de recursos, Custo de mão de obra, Custo de energia, Custo de manutenção, Custo de transporte, Eficiência Logística, Tempo de transporte, Rotas de distribuição                         |
| Distribuidor, Atacadista e Varejista | Inventário, Datas de recebimento, Datas de envio, Movimentação de produtos, Volume de vendas, Pedidos atendidos, Devoluções de mercadorias, Desempenho Logístico, Tempo de entrega, Taxa de rotação de estoque, Perdas de estoque, Tempo de processamento de pedidos, Capacidade de armazenamento, Margem de lucro, Ticket médio, Taxa de conversão, Análise de tendências |
| Cliente                              | Comportamento de compras, Histórico de compras, Avaliações de produtos, Dados demográficos, Informações de contato dos clientes                                                                                                                                                                                                                                            |

FONTE: A autora, 2023

## 2.4 INTELIGÊNCIA DE NEGÓCIOS

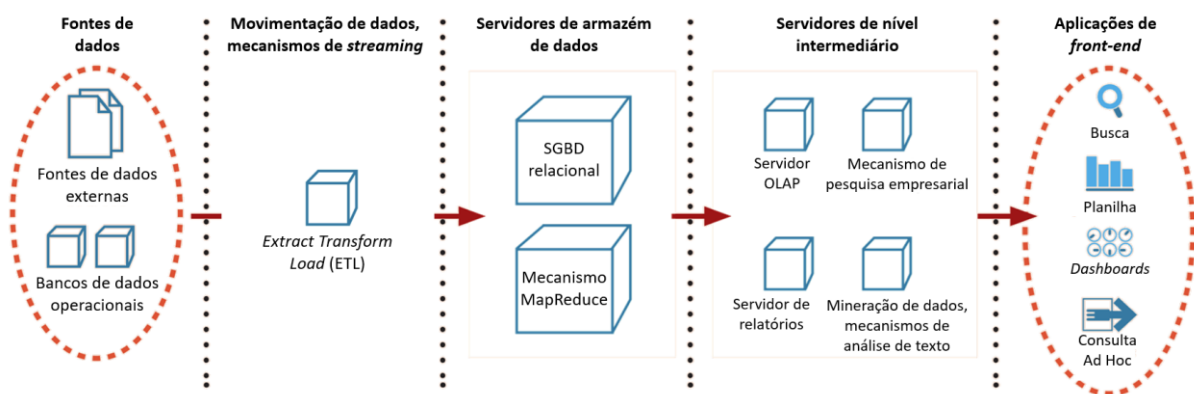
Reconhecendo que a quantidade de informações geradas pelos negócios era grande demais para ser analisada pelos humanos de maneira eficiente, Luhn (1958) propôs um sistema que usaria computadores para automatizar o processo de análise de dados e tomada de decisões, estabelecendo as bases para o desenvolvimento de ferramentas e técnicas modernas de *Business Intelligence* (BI). No entanto, o termo só foi cunhado em meados dos anos 80, quando a tecnologia se tornou mais sofisticada e começou a ser mais amplamente adotada.

Sua popularização é frequentemente creditada a Dresner (1989), que considerava BI um conjunto de conceitos e métodos que visam melhorar a tomada de decisões de negócio, usando sistemas de suporte baseados em fatos. Assim como essa, a grande maioria das definições de Inteligência de Negócios na literatura possuem enfoque comum em sua capacidade de geração de informações significativas e úteis para apoiar processos decisórios. Contudo, alguns autores abordam em termos mais técnicos. Para Turban et al. (2009), o BI é a integração de múltiplos sistemas para providenciar acesso a qualquer dado de maneira fácil e rápida, possivelmente em tempo real, bem como conduzir manipulações e análises apropriadas. Dessa forma, não é uma ferramenta de caráter simplesmente informativo, mas sim uma solução estratégica para as companhias (TURBAN et al., 2009).

O estudo de Sahay e Ranjan (2008) ressaltou a importância da Inteligência de Negócios para melhorar a eficiência e eficácia de uma SCM orientada por métricas e baseada em fatos. Para esse fim, diferentes tipos de *software* e elementos são incluídos no processo. Os *softwares* podem ser classificados de acordo com as três funções da SCM: SCP (*Supply Chain Planning*), que planeja a produção, a distribuição, o transporte e realiza previsões; SCE (*Supply Chain Execution*), responsável por integrar os dados relacionados à gestão das atividades operacionais da cadeia de suprimentos; e SCEM (*Supply Chain Event Management*), que gerencia eventos na cadeia de suprimentos (LANGLOIS e CHAUVEL, 2017).

Uma arquitetura eficiente de BI, ilustrada na Figura 5, implica em uma rede dinâmica de comunicação organizacional interconectada (CHAUDHURI et al., 2011). Os dados são provenientes de diversas fontes, como bancos de dados operacionais de vários departamentos internos e fornecedores externos. Por meio de ETL (Extrair, Transformar e Carregar), as informações coletadas são carregadas em um repositório chamado *Data Warehouse*. Inmon (2005) enfatiza o valor do *Data Warehousing* como base sólida para a implementação eficaz de soluções de BI, uma vez que permite uma visão única e consistente dos dados, facilitando a análise e o relatório de informações.

FIGURA 5 - Exemplo de Arquitetura de BI



FONTE: Adaptado de Chaudhuri et al., 2011

Pelo fato de apresentarem qualidade variável e usarem representações, códigos e formatos inconsistentes, no processo do ETL ocorre a integração, limpeza e padronização dos dados. Para tarefas de BI em tempo quase real, existem motores especializados chamados de *Complex Event Processing* (CEP) (CHAUDHURI et al., 2011). Daniel et al. (2008) contribuiu significativamente na conscientização e defesa

da precisão, integridade, consistência e relevância dos dados utilizados nas atividades de BI, argumentando que uma qualidade inadequada de informações pode levar a conclusões errôneas, decisões ruins e perda de oportunidades de negócios.

Os servidores de *Data Warehouse* são complementados por um conjunto de servidores intermediários que fornecem funcionalidades especializadas para diferentes cenários de BI. Os servidores de processamento analítico online (OLAP) são um exemplo, expondo de forma eficiente a visão multidimensional dos dados para aplicativos ou usuários e permitindo operações como filtragem, agregação e pivoteamento (CHAUDHURI et al., 2011).

Por fim, realiza-se o que pode ser considerado o resultado de um sistema BI: a criação de relatórios e visualizações, gerando informações que podem ser lidas de maneira eficiente e permitindo a conexão dos dados analisados às estratégias de negócio. Nessa etapa, diferentes métodos de visualização podem ser adotados. Por ser uma visão rápida e de fácil visualização, o *Dashboard* é fundamental em um sistema BI, consistindo em um agrupado de visões analíticas relacionadas, como tabelas, indicadores, gráficos, planilhas e outros insumos (SILVA et al., 2016).

Quando uma empresa opta por adotar a BI em sua cadeia de suprimentos, é evidente a necessidade de estratégias de governança de dados, com o estabelecimento de políticas que garantam a qualidade, integridade e confiabilidade das informações. O BI também deve estar alinhado com as metas organizacionais das empresas e ser projetado para fornecer elementos relevantes e acionáveis para os usuários finais, que assim como os demais *stakeholders*, devem ser envolvidos em seu processo de implementação (IMHOFF et al., 2001).

O campo do BI é notavelmente multidisciplinar e tem uma natureza de constante evolução à medida que novas tecnologias, metodologias e abordagens são introduzidas. De forma gradual, a Inteligência de Negócio passou a ser fundamental para o sucesso empresarial, sendo as empresas que a utilizam de maneira eficaz como base de tomada de decisões mais bem-sucedidas do que aquelas que não o fazem (DAVENPORT e HARRIS, 2017).

### 3 METODOLOGIA DA PESQUISA

Até o presente ponto, foi realizada uma pesquisa secundária da literatura sem delimitações temporais, abrangendo trabalhos desde o ano de 1958, representativo do surgimento dos conceitos em estudo, até 2022, para que fossem levadas em considerações atualidades da área, como o impacto do Covid-19 nas cadeias de suprimentos globais. A busca por artigos científicos se deu na língua inglesa, devido à maior disponibilidade, cobertura global, relevância para a comunidade científica e maior chance de citações e impacto no meio acadêmico (LILLIS e CURRY, 2010). A revisão foi conduzida usando principalmente os descritores “*Lubricants*”, “*Base Oils*”, “*Supply Chain*”, “*Data Analytics*”, “*Cost Forecasting*” e “*Business Intelligence*” nas seguintes bases de dados: Google Scholar, Elsevier (SCOPUS), Emerald, Research Rabbit e Science Direct. Esses termos, bem como outros relacionados, foram combinados de várias maneiras para buscar publicações pertinentes que pudessem servir de fundamento para a continuidade da pesquisa.

O artigo possui natureza aplicada e caráter exploratório. No Manual Frascati (2002), a pesquisa aplicada é definida como uma investigação realizada para adquirir novos conhecimentos, dirigida principalmente a um objetivo ou finalidade prática específica. Ela é conduzida para desenvolver possíveis usos para o conhecimento disponível a partir da pesquisa básica.

O procedimento técnico aplicado foi o estudo de caso, uma modalidade de pesquisa que tem como objetivo o aprofundamento de uma unidade específica em seu contexto, como um indivíduo, grupo ou instituição. Ao contrário de buscar representatividade ou generalização, o estudo de caso prioriza a profundidade e a riqueza das informações obtidas, oferecendo uma visão integrada do fenômeno investigado (LUDKE e ANDRÉ, 1986).

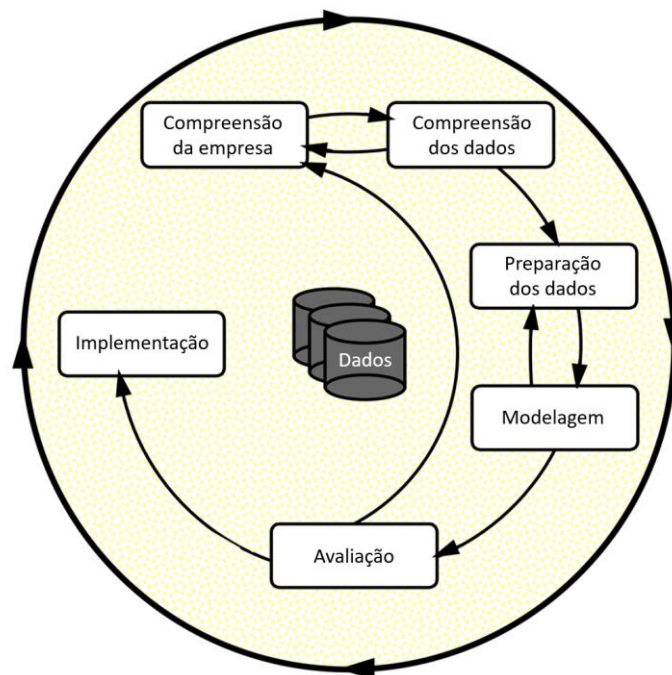
Quanto a abordagem dos dados, o trabalho é quantitativo. Para Creswell (2017), essa abordagem é utilizada para identificar fatores que influenciam resultados e avaliar a eficácia de intervenções, envolvendo a formulação de hipóteses estreitas e a coleta de dados antes e depois de tratamentos experimentais. Essa abordagem envolve a utilização de métodos matemáticos ou estatísticos para analisar as informações coletadas.

Segundo Gandomi e Haider (2015), os dados não possuem valor intrínseco. Seu potencial valor é alcançado apenas quando são utilizados para orientar a tomada

de decisões. Para isso, as organizações necessitam de processos eficientes que transformem grandes volumes de dados diversificados e em constante movimento em *insights* significativos. Dessa forma, é de suma importância estabelecer uma metodologia científica para auxiliar no direcionamento e desenvolvimento da pesquisa, permitindo o alcance de todos os objetivos estipulados. O presente estudo foi realizado em conformidade com a metodologia CRISP-DM, sigla para o inglês *Cross Industry Standard Process for Data Mining*, que em português significa Processo Padrão de Vários Segmentos de Mercados para Mineração de Dados.

Em 1999, foi introduzida a primeira versão do CRISP-DM. Esta metodologia foi concebida para catalogar e orientar as etapas mais comuns em projetos de mineração de dados e rapidamente se tornou o padrão amplamente aceito na área (MARTÍNEZ-PLUMED, 2019). O ciclo de vida de um projeto de mineração de dados é dividido em seis fases, representado na Figura 6, cuja sequência não é estrita. As setas indicam apenas as dependências mais importantes e frequentes entre as etapas, mas em um projeto específico, isso depende dos resultados obtidos anteriormente.

FIGURA 6 - Fases da metodologia CRISP-DM para a mineração de dados



FONTE: Adaptado de Wirth e Hipp, 2000

O círculo externo simboliza a natureza cíclica da própria mineração de dados, que não estará concluída uma vez que uma solução for implantada. Essa implementação pode gerar novas perguntas, que por sua vez podem gerar processos subsequentes que se beneficiarão das lições aprendidas das etapas anteriores. A seguir, apresenta-se uma explicação de cada etapa do processo (WIRTH e HIPPE, 2000):

- a) *Business Understanding* ou Compreensão da Empresa: Esta fase inicial concentra-se em compreender os objetivos e requisitos do projeto do ponto de vista empresarial, para em seguida transformar esse conhecimento em uma definição do problema de mineração de dados e em um plano preliminar para alcançar os objetivos. Neste estudo, o objetivo definido foi a implementação de um modelo de previsão de custos. Para finalizar a etapa, a situação da empresa foi avaliada para obter uma visão geral dos recursos disponíveis e necessários para o projeto;
- b) *Data Understanding* ou Compreensão dos Dados: A fase começa com uma coleta inicial de dados e segue com atividades para se familiarizar com eles, descrevê-los, identificar problemas de qualidade, obter primeiras percepções ou encontrar subconjuntos interessantes para formular hipóteses sobre informações ocultas. Os principais dados utilizados no trabalho foram os valores históricos dos produtos, fornecidos pela própria empresa, e as projeções de custos de especialistas do mercado. Existe uma ligação estreita entre a Compreensão do Negócio e a Compreensão dos Dados, uma vez que a formulação do problema de mineração de dados e o plano do projeto exigem pelo menos algum entendimento dos dados disponíveis;
- c) *Data Preparation* ou Preparação dos Dados: Abrange todas as atividades necessárias para construir o conjunto de dados final, que serão inseridos nas ferramentas de modelagem, a partir dos dados brutos iniciais. As tarefas de preparação de dados provavelmente serão realizadas várias vezes e não possuem uma ordem específica. Elas incluem seleção de tabelas, registros e atributos, limpeza de dados, criação de novos atributos e transformação de dados para as ferramentas de modelagem;
- d) *Modeling* ou Modelagem: Nesta fase, várias técnicas de modelagem são selecionadas e aplicadas, e seus parâmetros são definidos. Geralmente

existem várias técnicas para o mesmo tipo de problema de mineração de dados. No presente estudo, foram realizadas previsões utilizando o ARIMA e ETS, modelos de séries temporais, e as projeções de dados de mercado. Existe uma ligação estreita entre Preparação de Dados e Modelagem, na qual, muitas vezes, problemas nos dados são percebidos durante a modelagem ou surgem ideias para construir novos dados;

- e) *Evaluation* ou Avaliação: Antes de prosseguir para a implementação final, é importante avaliar mais minuciosamente o modelo e revisar as etapas executadas para construí-lo, garantindo que ele atenda adequadamente aos objetivos comerciais. Neste trabalho, foram realizados testes para comparar o erro médio percentual apresentado por cada modelo de previsão implementado. No final desta fase, uma decisão sobre o uso dos resultados da mineração de dados deve ser tomada;
- f) *Deployment* ou Implementação: Nesta fase, normalmente o conhecimento adquirido precisará ser organizado e apresentado de uma forma que o cliente possa utilizá-lo. Dependendo dos requisitos, essa fase pode ser tão simples quanto gerar um relatório ou tão complexa quanto implementar um processo repetível de mineração de dados. Nesta pesquisa, a etapa de implementação englobou a integração dos resultados do modelo na arquitetura atual da empresa e consequente elaboração de *dashboards* para o reportar as informações geradas.

A Figura 7 a seguir representa uma visão geral das etapas do CRISP-DM e dos resultados individuais esperados em cada uma delas:

FIGURA 7 - Visão geral das tarefas do CRISP-DM e seus resultados

| Compreendimento da Empresa                                                                                                   | Compreendimento dos Dados                                                                                                                                                                                                                                                                 | Preparação dos Dados                                                                                                                                                                                                                                                                                                                                                                    | Modelagem                                                                                                                                                                                                                                                                                                                                                   | Avaliação                                                                                                                                                                                                                                                                   | Implementação                                                                                                                                                                                                                                                                                  |
|------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Determinar os objetivos de negócios</b><br><i>Background</i><br>Objetivos de negócios<br>Critérios de sucesso empresarial | <b>Coletar dados iniciais</b><br>Relatório Inicial de Coleta de Dados<br><br><b>Descrever dados</b><br>Relatório de descrição de dados<br><br><b>Explorar dados</b><br>Relatório de exploração de dados<br><br><b>Verificar a qualidade dos dados</b><br>Relatório de qualidade dos dados | <b>Conjunto de dados</b><br>Descrição do conjunto de dados<br><br><b>Selecionar os dados</b><br>Justificativa para inclusão/exclusão<br><br><b>Limpar os dados</b><br>Relatório de limpeza dos dados<br><br><b>Construir os dados</b><br>Atributos derivados<br>Registros gerados<br><br><b>Integrar os dados</b><br>Dados mesclados<br><br><b>Formatar dados</b><br>Dados reformatados | <b>Selecionar a técnica de modelagem</b><br>Técnica de modelagem<br>Suposições de modelagem<br><br><b>Gerar <i>design</i> de teste</b><br>Projeto de teste<br><br><b>Construir o modelo</b><br>Configurações de parâmetros<br>Modelos<br>Descrição do modelo<br><br><b>Avaliar o modelo</b><br>Avaliação do modelo<br>Configurações de parâmetros revisadas | <b>Avaliar os resultados</b><br>Avaliação de resultados de mineração de dados, sucessp empresarial, critérios<br>Modelos aprovados<br><br><b>Processo de revisão</b><br>Revisão do processo<br><br><b>Determinar próximas etapas</b><br>Lista de possíveis ações<br>Decisão | <b>Planejar a implementação</b><br>Plano de preparação<br><br><b>Planejar o monitoramento e manutenção</b><br>Plano de monitoramento e manutenção<br><br><b>Produzir relatório final</b><br>Relatório final<br>Apresentação final<br><br><b>Revisar projeto</b><br>Documentação de experiência |

FONTE: Adaptado de Wirth e Hipp, 2000

## 4 APRESENTAÇÃO DOS RESULTADOS

O presente capítulo traz informações referentes à metodologia CRISP-DM aplicada ao estudo. O projeto foi desenvolvido em uma empresa multinacional do ramo de petróleo e gás, presente em seis continentes e com atuação em diferentes linhas, como: exploração e produção, distribuição, químicos, gás natural e comercialização de energia. Atualmente, a organização conta com mais de 2.000 funcionários no Brasil e 70.000 no mundo. A análise ocorreu na interface entre os setores de cadeia de suprimentos e comercial da empresa.

Por questões legais de confidencialidade de informações da empresa em estudo, foram criados dados sintéticos que possuam características semelhantes aos originais, mas que não exponham informações privadas. Esses dados podem ser gerados por meio de técnicas de síntese, preservando padrões e distribuições observados originalmente (NIKOLENKO, 2021).

### 4.1 COMPREENSÃO DA EMPRESA ESTUDADA

Um problema frequente pelo qual a companhia em estudo passa é a incapacidade de antecipar, de forma eficaz, os custos de seus produtos. A falta de previsão de custos em uma empresa multinacional representa um desafio significativo que pode impactar profundamente sua eficiência operacional e competitividade no mercado global. Quando uma organização não consegue antecipar variações nos custos de produção, matéria-prima ou logística, ela enfrenta dificuldades em tomar decisões estratégicas bem fundamentadas. Essa falta de visibilidade pode resultar em margens de lucro imprevisíveis, negociações desvantajosas, problemas de gestão de estoque e até mesmo perda de clientes devido a preços instáveis. Além disso, a incapacidade de prever custos pode levar a riscos financeiros significativos e a uma resposta reativa, em vez de proativa, a mudanças no mercado. Portanto, é crucial para empresas multinacionais implementarem ferramentas eficazes de previsão de custos para garantir estabilidade, competitividade e sucesso a longo prazo em um ambiente de negócios dinâmico e desafiador.

O processo de criação da solução de *Business Intelligence* teve início com encontros realizados junto aos principais tomadores de decisões das áreas a qual o projeto estava relacionado, bem como com o departamento técnico da empresa. O

objetivo era compreender os requisitos do modelo de BI e como ele poderia ser integrado à infraestrutura técnica já em uso na organização. A organização reportou que, atualmente, a maioria de suas unidades de operação já estão realizando cálculos para a previsão de custos, porém de forma manual, tornando o processo demorado e custoso, com inúmeras limitações e sem padronização a nível global.

Diante dos fatos, por meio deste projeto, constatou-se que a empresa almeja a aprimoração de seu sistema de BI vigente, através do desenvolvimento de um protótipo de modelo de previsão de custos, que seja automatizado e padronizado a nível global, para o suporte a tomada de decisões. O intervalo de interesse estabelecido é de três meses no futuro, ou seja, anseia-se obter dados de projeção dos custos dos produtos para o próximo trimestre em relação ao momento de previsão.

Faz-se essencial compreender a estrutura de custos da organização, visto que essa será a base para o cálculo das previsões. Atualmente na companhia, o Custo dos Produtos Vendidos (CPV) para lubrificantes é dividido em 16 principais componentes, agrupados em quatro grupos, conforme apresentado na Figura 8.

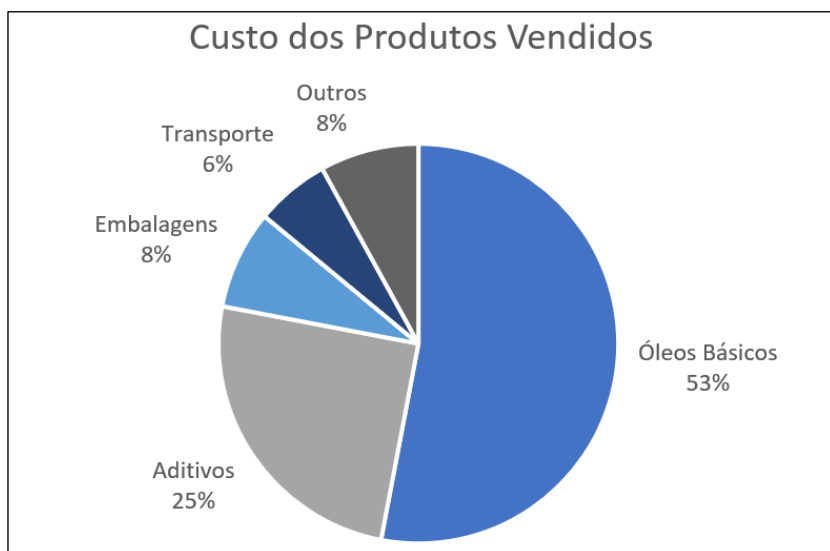
FIGURA 8 - Estrutura de Custos da empresa

| Componente de Custo | Nome do Componente de Custo              | Grupo de Custo             |
|---------------------|------------------------------------------|----------------------------|
| 001                 | Óleos básicos                            | 01 (Materiais)             |
| 002                 | Óleos residuais                          |                            |
| 003                 | Aditivos                                 |                            |
| 004                 | Materiais de embalagens                  |                            |
| 005                 | Produtos adquiridos                      |                            |
| 006                 | Paletes                                  |                            |
| 007                 | Transportes intermediários COGS          |                            |
| 008                 | Processos de mistura                     | 02 (Produção)              |
| 009                 | Processos de embalagem                   |                            |
| 010                 | Processos terceirizados/Outros processos |                            |
| 011                 | Laboratórios                             |                            |
| 012                 | Manuseio de materias a granel            | 03 (Manuseio)              |
| 013                 | Manuseio de materiais embalados          |                            |
| 014                 | Manuseio terceirizado                    |                            |
| 015                 | Transportes intermediários OPEX          |                            |
| 016                 | Margem do comerciante                    | 04 (Margem do Comerciante) |

FONTE: A autora, 2023

De uma forma geral, os componentes que mais impactam o custo final dos lubrificantes são óleos básicos, aditivos, materiais de embalagens e transportes. A relação entre eles está representada na Figura 9.

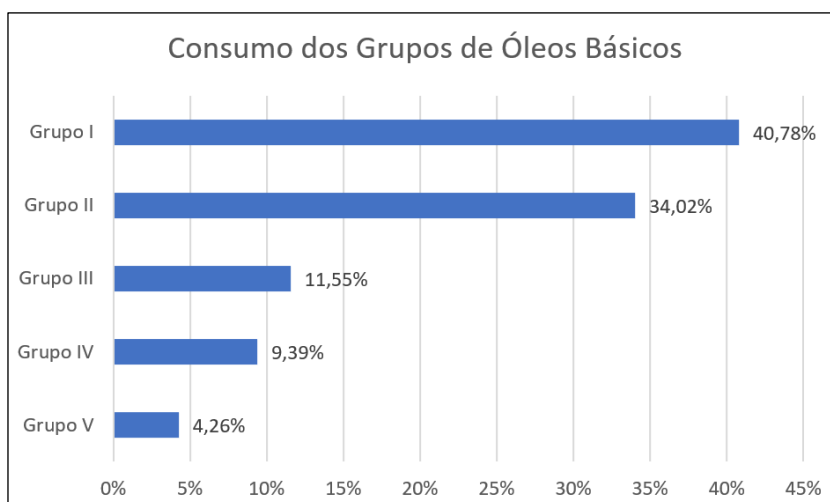
FIGURA 9 - Divisão do Custo dos Produtos Vendidos na empresa



FONTE: A autora, 2023

Uma vez que os óleos básicos são um dos principais componentes dos lubrificantes, existe uma forte correlação entre seus custos. Mudanças nos preços dos óleos básicos geralmente afetam diretamente os custos dos lubrificantes. Dentre os óleos básicos, os Grupo I e II são os mais consumidos pela empresa (Figura 10).

FIGURA 10 - Consumo dos Grupos de Óleos Básicos na empresa



FONTE: A autora, 2023

Pelo fato dos óleos básicos de Grupo I e II serem produzidos a partir do petróleo bruto e envolverem processos de refino relativamente simples e bem estabelecidos, eles tendem a ter especificações mais padronizadas em comparação com óleos de grupos mais especializados. Isso reduz a variabilidade nos custos associados às variações nas especificações.

Embora as *commodities* estejam sujeitas a flutuações de preço devido a eventos externos, como mudanças nas políticas governamentais ou condições climáticas extremas, óleos básicos de Grupo I e II são menos sensíveis a esses fatores, além de seguirem os padrões de mercado de uma forma geral. Quanto ao seu uso, eles estão intrinsecamente ligados a indústrias de base, como a automotiva e industrial, que possuem padrões de consumo consistentes. Todos esses fatores, juntamente com a disponibilidade e estabilidade nas fontes de petróleo, contribuem para a previsibilidade da demanda e, conseqüentemente, de seus preços. Em vista disso, os óleos básicos de Grupo I e II foram selecionados como principal fator base para as previsões de custos dos lubrificantes.

#### 4.2 COMPREENSÃO DOS DADOS

Foi feita uma relação das ferramentas que seriam necessárias no processo de implementação do modelo, escolhidas por já estarem disponíveis no ambiente pesquisado e pela facilidade de comunicação e integração entre elas. São elas:

- SAP HANA (*High-performance Analytic Appliance*): Um sistema de gerenciamento de banco de dados (DBMS) relacional e em memória desenvolvido pela SAP. Ele oferece uma plataforma de computação e armazenamento de dados em tempo real, permitindo o processamento rápido e eficiente de grandes volumes de dados. O SAP HANA é projetado para acelerar análises complexas e consultas, oferecendo desempenho superior em comparação com sistemas de banco de dados tradicionais. Além disso, ele integra capacidades de análise preditiva e *machine learning*, fornecendo uma solução abrangente para análise de dados e suporte à tomada de decisões em tempo real nas organizações;
- Snowflake Data Cloud: Uma plataforma de armazenamento de dados em nuvem que oferece um ambiente altamente flexível e escalável para armazenamento, processamento e análise de dados. Ele permite que as organizações armazenem e compartilhem dados de forma segura e eficiente em uma arquitetura de dados totalmente na nuvem. O Snowflake utiliza uma estrutura de dados baseada em armazenamento separado de computação, permitindo que os usuários dimensionem recursos conforme necessário, pagando apenas pelo que utilizam. Além disso, oferece recursos avançados de compartilhamento de dados entre organizações, facilitando a colaboração e a análise de dados em larga escala em tempo real;
- SAP HANA Studio: Uma ferramenta de desenvolvimento integrada fornecida pela SAP para trabalhar com o SAP HANA. O SAP HANA Studio permite aos desenvolvedores criar, gerenciar e otimizar aplicativos que utilizam o SAP HANA. Ele oferece uma interface gráfica para desenvolver modelos de dados, criar procedimentos armazenados, gerenciar segurança, monitorar o desempenho e administrar o banco de dados SAP HANA;
- Alteryx Designer: Uma plataforma avançada de análise de dados que oferece uma ampla gama de ferramentas para preparação, análise e visualização de dados. Ele permite aos usuários criar fluxos de trabalho intuitivos e automatizados para limpar, transformar e analisar dados complexos de diversas fontes. O Alteryx Designer é conhecido por sua capacidade de simplificar tarefas complexas de análise de dados, unindo

dados de diferentes fontes, aplicando transformações e facilitando a criação de modelos preditivos, tudo em uma interface de usuário amigável;

- Tableau: Uma ferramenta de visualização de dados e BI que permite aos usuários criar visualizações interativas e painéis dinâmicos a partir de conjuntos de dados complexos. Ele oferece uma interface amigável que facilita a criação de gráficos, tabelas e *dashboards* interativos, permitindo explorar e entender dados de maneira intuitiva.

Tendo os sistemas que serão utilizados na execução do projeto definidos, iniciou-se um estudo dos dados buscando uma familiarização com esses. Inicialmente foram levantados os dados disponíveis necessários do SAP HANA, referentes aos componentes de custo históricos, a nível de material e planta (Quadro 2):

QUADRO 2 - Colunas extraídas do SAP HANA como *input* do modelo

| Campo                      | Tipo   | Significado                                        |
|----------------------------|--------|----------------------------------------------------|
| PLANT                      | String | Código da planta                                   |
| MATERIAL_KEY               | String | Código do material                                 |
| CD_YEAR_MONTH              | Data   | Data                                               |
| PLANT_NAME                 | String | Descrição da planta                                |
| MATERIAL_DESC              | String | Descrição do material                              |
| LS_GLOBAL_PROD_FAMILY_DESC | String | Família do produto                                 |
| CD_MATERIAL_GROUP          | String | Grupo do produto                                   |
| SOURCE_SYSTEM              | String | Região Global do sistema em que                    |
| CM_USD_RATE_AVG_CM         | Float  | Taxa de câmbio da moeda local para USD             |
| CM_COST_GOOD_SOLD_REF_COST | Float  | Valor total na moeda local                         |
| CM_COST_GOOD_SOLD_LITER    | Float  | Valor unitário do litro na moeda local             |
| KST001                     | Float  | Custo dos Oléos básicos                            |
| KST002                     | Float  | Custo dos Oléos residuais                          |
| KST003                     | Float  | Custo dos Aditivos                                 |
| KST004                     | Float  | Custo dos Materiais de embalagens                  |
| KST005                     | Float  | Custo dos Produtos adquiridos                      |
| KST006                     | Float  | Custo dos Paletes                                  |
| KST007                     | Float  | Custo de Transportes intermediários COGS           |
| KST008                     | Float  | Custo dos Processos de mistura                     |
| KST009                     | Float  | Custo dos Processos de embalagem                   |
| KST010                     | Float  | Custo dos Processos terceirizados/Outros processos |
| KST011                     | Float  | Custo dos Laboratórios                             |
| KST012                     | Float  | Custo de Manuseio de materias a granel             |
| KST013                     | Float  | Custo de Manuseio de materiais embalados           |
| KST014                     | Float  | Custo de Manuseio terceirizado                     |
| KST015                     | Float  | Custo de Transportes intermediários OPEX           |
| KST016                     | Float  | Custo da Margem do comerciante                     |

FONTE: A autora, 2023

Além das informações internas da empresa, também foram utilizados dados referentes ao mercado internacional de óleos básicos, publicados pela Argus. Desenvolvida pela ARGUS MEDIA, a Argus é uma plataforma global de *software* usada para monitoramento, gestão e análise de riscos no setor de energia e *commodities*. Os relatórios produzidos pela Argus de preços e previsão de custos de óleos básicos são replicados em um repositório no Snowflake. O Quadro 3 descreve as colunas de interesse que serão utilizadas.

QUADRO 3 - Colunas extraídas do Snowflake como *input* do modelo

| <b>Campo</b>           | <b>Tipo</b> | <b>Significado</b>                    |
|------------------------|-------------|---------------------------------------|
| CODE                   | String      | Código do produto                     |
| CODE_DISPLAY_NAME      | String      | Nome do produto                       |
| CONTRACT_MONTH         | Float       | Mês ao qual a previsão de refere      |
| CONTRACT_YEAR          | Float       | Ano ao qual a previsão se refere      |
| DELIVERY_MODE          | String      | Modo de entrega                       |
| FILE_NAME              | String      | Nome da fonte dos dados               |
| FREQUENCY              | String      | Frequência de postagem dos relatórios |
| OPR_DATE               | Data        | Data em que os dados foram publicados |
| PRICE_TYPE_DESCRIPTION | String      | Descrição do tipo de preço            |
| UNIT                   | String      | Unidade de medida dos dados           |
| VALUE                  | Float       | Valor numérico                        |

FONTE: A autora, 2023

#### 4.3 PREPARAÇÃO DOS DADOS

O tratamento de uma base de dados de preços históricos é essencial para garantir a precisão e confiabilidade de modelos de previsão de custos. Dados limpos, livres de erros e *outliers*, são fundamentais para evitar viés, distorção e para melhorar a eficiência computacional do modelo, facilitando análises exploratórias, preparação para modelagem e interpretação dos resultados. A confiabilidade das previsões aumenta, levando a decisões mais informadas e eficazes nos negócios e simplificando futuras atualizações, mantendo o modelo relevante ao longo do tempo.

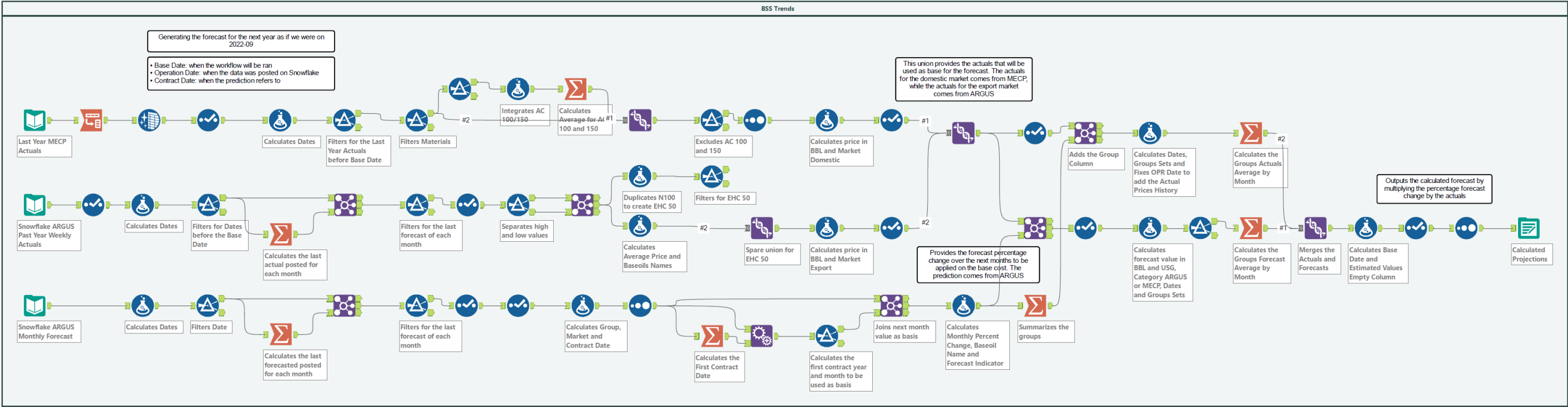
Nesse ponto da arquitetura, tanto no SAP HANA quanto no Snowflake, o processo de ETL já foi executado e os dados estão integrados e consolidados em *Data Warehouses*, armazenados em tabelas, já tendo passado, portanto, por uma etapa inicial de pré-processamento. Entretanto, a parte inicial do *workflow* desenvolvido realiza mais algumas operações de tratamento. Dentre elas, estão:

- Remoção de dados duplicados ou inconsistentes, como valores de preço negativos;
- Tratamento de valores ausentes ou nulos utilizando o *forward filling*, uma técnica para substituir esses valores pelo último valor observado conhecido;
- Conversão das unidades de medida dos dados;
- Padronização dos códigos e nomes de produtos para a nomenclatura utilizada na empresa;
- Agregação dos custos dos produtos ao nível de granularidade mensal.

#### 4.4 MODELAGEM

A Argus é especializada em fornecer informações detalhadas sobre mercados de *commodities* específicos, incluindo fatores relevantes que afetam seus preços, como eventos geopolíticos, regulamentações governamentais ou mudanças na demanda global. Isso pode fornecer *insights* valiosos para previsões mais precisas e informadas dos óleos básicos. Desta forma, optou-se por utilizar como base do modelo de previsão os dados publicados mensalmente pela Argus. O modelo foi desenvolvido em Alteryx e dividido em duas partes. A primeira delas realiza os cálculos das projeções de custo dos óleos básicos de Grupo I e II, tanto para o mercado doméstico quanto para o internacional (Figura 11):

FIGURA 11 - Workflow em Alteryx para calcular as projeções dos óleos básicos



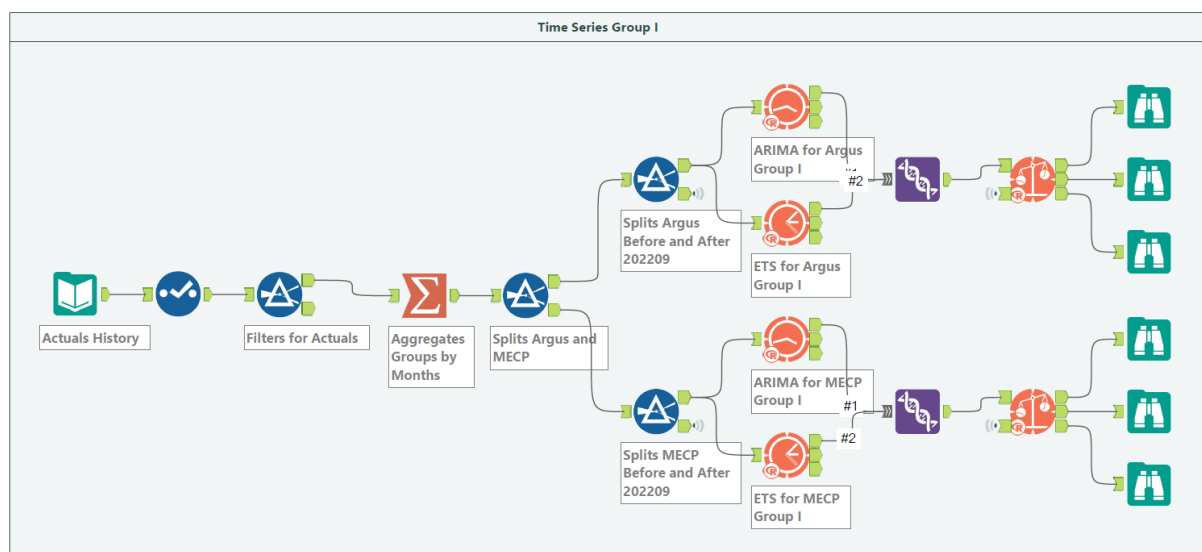
FONTE: A autora, 2023

O *workflow* da Figura 11 é alimentado com os valores históricos dos óleos básicos nos mercados doméstico e internacional e as previsões de custos da Argus. Alternativamente a utilizar os valores absolutos projetados pela ARGUS MEDIA, é calculada a variação percentual desses, levando em consideração a proporção de aumento ou decréscimo mensal entre os custos dos produtos ao longo do próximo ano. Essa estratégia permite capturar as previsões de comportamento relativo de mercado especuladas pelos especialistas. Dessa forma, as previsões percentuais foram aplicadas sobre os custos mais recentes reportados pela empresa em estudo, gerando uma previsão de cada óleo básico para o próximo ano. Por fim, o *workflow* agrega esses valores de acordo com a sua classificação de grupo por meio de uma média simples.

Para qualquer projeto de desenvolvimento de um modelo de dados, é fundamental comparar diferentes técnicas e abordagens para determinar qual delas apresenta melhor desempenho em termos de precisão. É crucial escolher o modelo que mais se adequa ao conjunto de dados em análise, levando em consideração seus padrões e comportamentos singulares.

Como abordado anteriormente, técnicas de séries temporais são eficazes na identificação de sazonalidades, permitindo ajustar automaticamente as previsões quando ocorrem mudanças nas tendências ou padrões dos dados. Para avaliar a precisão do modelo, foram então realizadas previsões dos custos utilizando também os modelos ARIMA e ETS, conforme o *workflow* da Figura 12:

FIGURA 12 - Workflow em Alteryx dos modelos ARIMA e ETS



FONTE: A autora, 2023

Esses modelos foram selecionados por serem os mais comumente utilizados para a previsão de séries temporais na literatura, além de sua relativa facilidade de implementação, especialmente quando comparados a abordagens mais complexas de aprendizado de máquina. Isso se mostrou vantajoso na situação em estudo, em que a simplicidade e a rapidez de implementação são prioridades.

No Alteryx, a ferramenta ARIMA disponível estima um modelo de previsão de série temporal, seja ele univariado ou multivariado, utilizando um método de média móvel integrada autorregressiva. O usuário pode especificar diretamente os parâmetros subjacentes do modelo, ou, como foi o caso deste trabalho, optar por uma abordagem automatizada do *software* que desenvolve um modelo baseado em critérios estatísticos.

Já a ferramenta ETS visa estimar um modelo univariado de previsão de série temporal através de um método de suavização exponencial. Essa abordagem de previsão se baseia em uma média ponderada de observações passadas, com os pesos diminuindo de tamanho para valores passados mais distantes, seguindo uma função de decréscimo exponencial. Essa funcionalidade é capaz de levar em conta três componentes da série temporal: nível, tendência e sazonalidade. Da mesma forma que o ARIMA, a ferramenta pode usar métodos totalmente automatizados para modelar os três componentes da melhor maneira com base em critérios estatísticos, ou os mesmos podem ser especificados manualmente.

Como conjunto de treinamento para os modelos foram utilizados dados históricos de custo dos óleos básicos de Grupo I e II, tanto para o mercado internacional quanto para o doméstico, a partir de janeiro de 2014. Uma vez que o ambiente em que o modelo está sendo aplicado passa por mudanças significativas, como sazonalidades, eventos especiais, e mudanças nas condições econômicas, faz-se necessário testá-lo em variadas condições para garantir uma análise mais apurada de seu comportamento. Dessa forma, foram realizados testes de previsões baseados em 13 meses consecutivos: de novembro de 2021 a novembro de 2022, totalizando mais de 50 diferentes cenários de simulação.

Para avaliar seus desempenhos, utilizaram-se conjuntos de teste de dados históricos não vistos pelos modelos. Como critério de sucesso para o projeto, adotou-se a métrica de *Mean Absolute Percentage Error* (MAPE), ou Erro Percentual Absoluto Médio, utilizada para avaliar a precisão de um modelo de previsão em termos percentuais. Ela calcula a média das porcentagens das diferenças absolutas entre as previsões do modelo e os valores reais, conforme a Equação 2.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{X_t - \hat{X}_t}{X_t} \times 100 \quad (2)$$

Onde:

- $X_t$  representa o valor real da observação  $t$ ;
- $\hat{X}_t$  representa o valor previsto pelo modelo para a observação  $t$ ;
- $n$  é o número total de observações.

Foi calculado o MAPE dos três modelos nos diferentes cenários para um intervalo de previsão de 1 mês, 3 meses e 11 meses no futuro. Os resultados estão sintetizados na Tabela 1.

TABELA 1 - Comparação dos resultados MAPE obtidos por cada modelo

| Data Base | Mercado       | Grupo    | 1 mês |       |       | 3 meses |       |       | 11 meses |       |        |
|-----------|---------------|----------|-------|-------|-------|---------|-------|-------|----------|-------|--------|
|           |               |          | ARGUS | ARIMA | ETS   | ARGUS   | ARIMA | ETS   | ARGUS    | ARIMA | ETS    |
| 2021-11   | Doméstico     | Grupo I  | 5,26  | 0,49  | 1,42  | 9,36    | 0,95  | 2,19  | 22,91    | 18,76 | 10,78  |
|           |               | Grupo II | 7,45  | 2,10  | 0,67  | 12,27   | 4,42  | 0,55  | 29,68    | 7,98  | 15,88  |
|           | Internacional | Grupo I  | 2,20  | 6,20  | 4,60  | 4,24    | 10,04 | 3,33  | 17,58    | 60,32 | 30,74  |
|           |               | Grupo II | 1,54  | 0,75  | 2,31  | 15,56   | 7,25  | 16,21 | 24,21    | 63,34 | 29,03  |
| 2021-12   | Doméstico     | Grupo I  | 4,91  | 0,47  | 0,54  | 7,78    | 2,94  | 0,51  | 22,40    | 23,10 | 14,68  |
|           |               | Grupo II | 6,14  | 1,12  | 0,49  | 11,12   | 1,59  | 1,57  | 28,91    | 13,91 | 17,15  |
|           | Internacional | Grupo I  | 4,09  | 3,34  | 5,90  | 9,64    | 4,80  | 9,78  | 17,92    | 49,64 | 22,75  |
|           |               | Grupo II | 16,15 | 13,10 | 15,14 | 21,04   | 11,22 | 16,13 | 25,66    | 72,92 | 33,83  |
| 2022-01   | Doméstico     | Grupo I  | 5,62  | 2,42  | 1,25  | 9,14    | 6,23  | 3,62  | 24,70    | 25,49 | 17,92  |
|           |               | Grupo II | 4,95  | 0,62  | 0,05  | 11,29   | 4,93  | 6,15  | 27,81    | 14,81 | 20,17  |
|           | Internacional | Grupo I  | 4,56  | 1,19  | 0,09  | 11,34   | 21,24 | 14,42 | 22,20    | 50,73 | 38,24  |
|           |               | Grupo II | 9,51  | 8,11  | 8,92  | 16,64   | 36,89 | 31,39 | 28,06    | 71,86 | 66,73  |
| 2022-02   | Doméstico     | Grupo I  | 3,31  | 0,62  | 0,24  | 7,87    | 4,79  | 5,63  | 22,92    | 12,89 | 16,22  |
|           |               | Grupo II | 6,15  | 3,63  | 4,33  | 9,98    | 7,60  | 9,25  | 27,69    | 16,82 | 22,12  |
|           | Internacional | Grupo I  | 1,27  | 11,59 | 5,06  | 18,99   | 42,66 | 28,97 | 26,05    | 56,85 | 42,19  |
|           |               | Grupo II | 2,78  | 13,89 | 9,17  | 24,40   | 48,28 | 37,68 | 32,00    | 60,12 | 53,78  |
| 2022-03   | Doméstico     | Grupo I  | 3,75  | 5,53  | 5,89  | 9,80    | 11,46 | 11,92 | 21,88    | 17,07 | 17,78  |
|           |               | Grupo II | 8,67  | 7,32  | 8,42  | 10,79   | 7,49  | 9,28  | 25,09    | 10,07 | 15,55  |
|           | Internacional | Grupo I  | 26,93 | 39,41 | 32,07 | 29,38   | 42,66 | 35,27 | 26,38    | 32,70 | 30,90  |
|           |               | Grupo II | 34,50 | 47,47 | 39,45 | 37,25   | 50,26 | 43,17 | 31,66    | 38,12 | 36,84  |
| 2022-04   | Doméstico     | Grupo I  | 6,94  | 3,05  | 3,27  | 14,16   | 6,86  | 7,33  | 20,24    | 7,70  | 7,37   |
|           |               | Grupo II | 4,73  | 11,36 | 11,00 | 13,28   | 7,66  | 7,45  | 19,32    | 11,34 | 8,74   |
|           | Internacional | Grupo I  | 4,59  | 16,56 | 21,67 | 10,21   | 20,00 | 32,15 | 9,90     | 36,39 | 106,93 |
|           |               | Grupo II | 6,87  | 13,68 | 4,10  | 13,86   | 12,11 | 10,29 | 15,53    | 34,60 | 22,02  |
| 2022-05   | Doméstico     | Grupo I  | 8,32  | 2,97  | 3,48  | 13,31   | 4,05  | 4,26  | 14,75    | 11,41 | 10,93  |
|           |               | Grupo II | 16,70 | 16,87 | 15,84 | 22,46   | 20,96 | 20,42 | 23,39    | 12,71 | 15,53  |
|           | Internacional | Grupo I  | 4,81  | 2,98  | 2,55  | 9,58    | 6,63  | 5,20  | 12,72    | 8,96  | 27,47  |
|           |               | Grupo II | 6,45  | 6,43  | 6,39  | 11,97   | 11,01 | 10,35 | 20,18    | 13,49 | 29,17  |
| 2022-06   | Doméstico     | Grupo I  | 6,16  | 2,32  | 0,11  | 9,04    | 8,38  | 2,69  | 6,88     | 46,19 | 21,85  |
|           |               | Grupo II | 9,65  | 4,92  | 5,66  | 11,15   | 2,85  | 3,49  | 8,90     | 26,91 | 13,58  |
|           | Internacional | Grupo I  | 7,29  | 4,43  | 3,33  | 7,01    | 6,45  | 7,89  | 20,22    | 15,40 | 42,17  |
|           |               | Grupo II | 8,92  | 7,96  | 7,02  | 7,69    | 5,87  | 7,87  | 28,87    | 23,56 | 40,60  |
| 2022-07   | Doméstico     | Grupo I  | 3,42  | 3,91  | 2,03  | 3,87    | 11,17 | 6,66  | 5,81     | 43,80 | 25,71  |
|           |               | Grupo II | 1,18  | 4,82  | 3,55  | 1,75    | 11,27 | 8,53  | 4,59     | 41,03 | 26,10  |
|           | Internacional | Grupo I  | 0,34  | 3,91  | 6,59  | 14,19   | 19,81 | 23,37 | 32,21    | 28,95 | 65,75  |
|           |               | Grupo II | 0,69  | 4,45  | 1,45  | 14,54   | 20,07 | 17,69 | 45,10    | 42,36 | 56,97  |
| 2022-08   | Doméstico     | Grupo I  | 5,04  | 3,34  | 2,44  | 7,02    | 9,07  | 6,67  | 5,95     | 36,55 | 23,05  |
|           |               | Grupo II | 5,52  | 2,41  | 2,11  | 6,45    | 9,90  | 8,18  | 6,45     | 36,13 | 23,78  |
|           | Internacional | Grupo I  | 10,38 | 14,51 | 15,59 | 15,83   | 21,72 | 26,44 | 25,13    | 20,63 | 38,70  |
|           |               | Grupo II | 10,80 | 14,24 | 17,59 | 18,88   | 24,55 | 32,65 | 39,25    | 35,40 | 63,19  |
| 2022-09   | Doméstico     | Grupo I  | 2,43  | 3,51  | 4,20  | 7,58    | 4,30  | 5,22  | 6,21     | 22,99 | 17,72  |
|           |               | Grupo II | 0,46  | 8,61  | 7,84  | 6,60    | 11,25 | 8,87  | 7,59     | 36,61 | 23,69  |
|           | Internacional | Grupo I  | 5,65  | 0,10  | 3,33  | 5,94    | 5,37  | 10,48 | 11,31    | 16,51 | 31,39  |
|           |               | Grupo II | 4,14  | 0,50  | 10,94 | 10,20   | 1,82  | 24,37 | 23,00    | 7,65  | 44,58  |
| 2022-10   | Doméstico     | Grupo I  | 6,89  | 2,46  | 2,23  | 8,48    | 2,30  | 2,60  | 5,76     | 5,40  | 5,00   |
|           |               | Grupo II | 7,19  | 0,52  | 0,30  | 11,54   | 2,66  | 1,61  | 6,74     | 22,04 | 12,46  |
|           | Internacional | Grupo I  | 1,70  | 5,87  | 1,49  | 2,31    | 10,74 | 8,78  | 10,55    | 19,38 | 21,24  |
|           |               | Grupo II | 9,04  | 2,92  | 14,92 | 10,93   | 4,40  | 20,84 | 23,34    | 8,39  | 33,81  |
| 2022-11   | Doméstico     | Grupo I  | 2,42  | 1,43  | 1,71  | 1,16    | 6,35  | 7,05  | 4,33     | 8,90  | 12,97  |
|           |               | Grupo II | 5,86  | 1,26  | 0,20  | 4,91    | 8,12  | 4,80  | 5,87     | 26,99 | 14,70  |
|           | Internacional | Grupo I  | 0,96  | 1,49  | 3,49  | 1,35    | 5,01  | 5,63  | 8,81     | 9,68  | 10,41  |
|           |               | Grupo II | 2,07  | 4,69  | 5,64  | 5,39    | 14,25 | 9,37  | 17,03    | 12,99 | 17,38  |

FONTE: A autora, 2023

Percebe-se uma variância elevada nos resultados obtidos. Nas previsões de 11 meses, por exemplo, os índices MAPE variaram de 4,33% até 106,93%. Para melhor compreender essa diferença, foram selecionados dois desses casos para que uma análise mais aprofundada fosse desenvolvida, devido à sua ampla disparidade de resultados. O primeiro é a previsão realizada em janeiro de 2022 para o Grupo II de óleos básicos do mercado internacional. Os resultados estão reapresentados na Tabela 2 a seguir:

TABELA 2 - Resultados MAPE obtidos em janeiro de 2022

| Data Base | Mercado       | Grupo    | 11 meses |       |       |
|-----------|---------------|----------|----------|-------|-------|
|           |               |          | ARGUS    | ARIMA | ETS   |
| 2022-01   | Internacional | Grupo II | 28,06    | 71,86 | 66,73 |

FONTE: A autora, 2023

Os erros significativos observados pelos modelos de séries temporais nesse cenário ficam evidentes por meio de uma análise gráfica de comparação entre as previsões mensais geradas pelos modelos e os valores de fato realizados (Gráfico 1):

GRÁFICO 1 - Resultados MAPE obtidos por séries temporais em janeiro de 2022



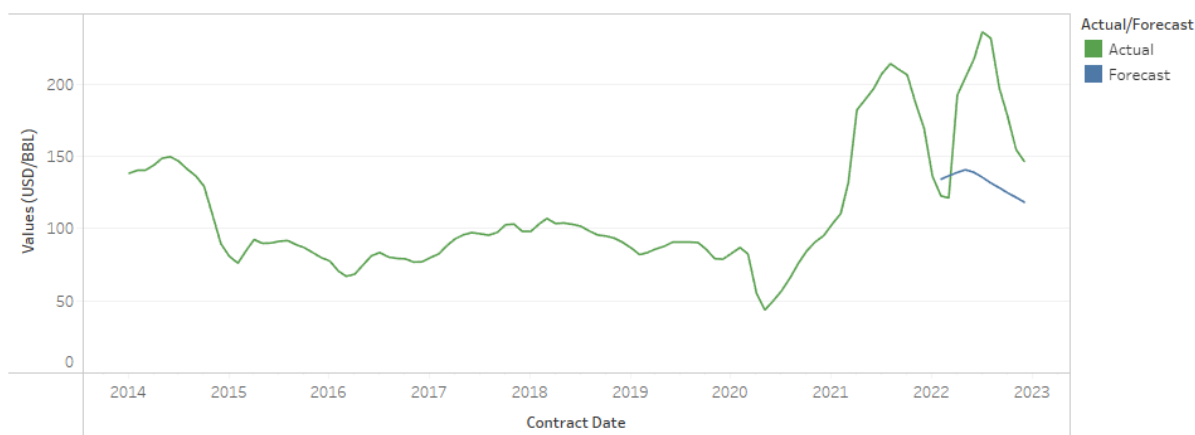
FONTE: A autora, 2023

Observa-se no Gráfico 1 que o período consequente ao mês de janeiro de 2022, selecionado como base para o cálculo das previsões, foi marcado por um aumento brusco dos valores dos óleos básicos no mercado doméstico, dissemelhante a tendência histórica em vigor nos meses anteriores. Compreende-se que esse foi o principal motivo da elevada porcentagem de erro obtida.

Em geral, os modelos de previsão baseados em séries temporais assumem estacionariedade nos dados, isto é, que as suas propriedades estatísticas não mudam ao longo do tempo. Como essa não é a realidade do estudo em questão, a precisão das previsões é prejudicada. Essa dependência excessiva de valores históricos resulta na incapacidade dos modelos de lidar com mudanças abruptas nos dados. Até mesmo cenários de padrões sazonais podem comprometer os resultados, caso esses ciclos sejam muito complexos.

Similarmente, o Gráfico 2 compara as previsões realizadas pelo modelo baseado nos dados de mercado com os dados realizados:

GRÁFICO 2 - Resultados MAPE obtidos por dados de mercado em janeiro de 2022



FONTE: A autora, 2023

Apesar de não apresentar um desempenho totalmente satisfatório, as previsões obtidas no Gráfico 2 seguiram, mesmo que em uma proporção menor, o comportamento de aumento e queda nos custos dos óleos básicos observadas nos próximos meses.

O segundo caso selecionado para análise foi referente à previsão de outubro de 2022, para o Grupo I do mercado doméstico, retratados na Tabela 3 e Gráfico 3:

TABELA 3 - Resultados MAPE obtidos em outubro de 2022

| Data Base | Mercado   | Grupo   | 11 meses |       |      |
|-----------|-----------|---------|----------|-------|------|
|           |           |         | ARGUS    | ARIMA | ETS  |
| 2022-10   | Doméstico | Grupo I | 5,76     | 5,40  | 5,00 |

FONTE: A autora, 2023

GRÁFICO 3 - Resultados MAPE obtidos por séries temporais em outubro de 2022



FONTE: A autora, 2023

No Gráfico 3, constata-se que os meses anteriores ao período em análise já haviam apresentado uma queda nos custos dos óleos básicos, que perdurou pelos meses seguintes. Essa constância de comportamento resultou em erros percentuais baixos. O melhor ajuste das previsões nesse cenário atesta a dependência e limitação dos modelos de séries temporais aos dados históricos, mostrando-se mais eficazes para capturar tendências lineares ou padrões de menor volatilidade.

De maneira análoga, o modelo baseado na Argus gerou previsões apuradas, como ilustradas no Gráfico 4:

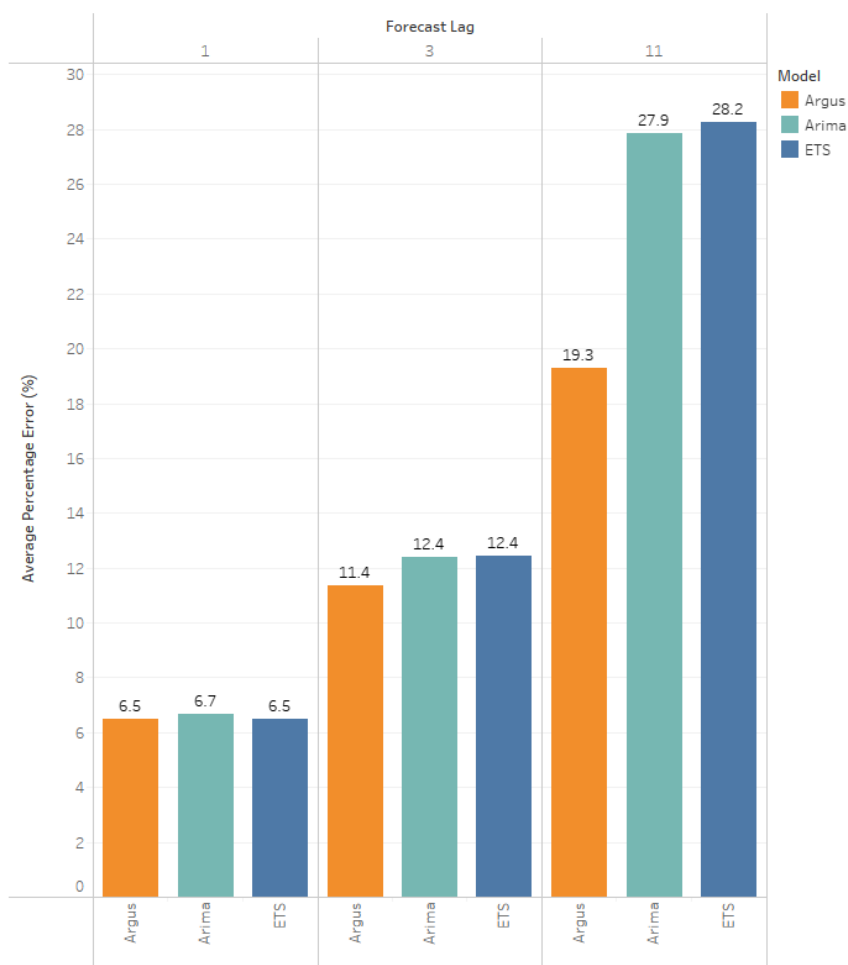
GRÁFICO 4 - Resultados MAPE obtidos por dados de mercado em outubro de 2022



FONTE: A autora, 2023

Frente às diversas métricas calculadas até esse ponto, criou-se uma visualização com o intuito de facilitar a seleção do modelo que apresentou o menor erro percentual médio (Figura 13):

FIGURA 13 - Agregação dos valores MAPE por intervalo de precisão



FONTE: A autora, 2023

A Figura 13 sintetiza as médias MAPE obtidas por cada modelo, agregadas a nível de intervalo de previsão. Constata-se que o modelo construído com base nos dados da Argus apresentou os menores erros percentuais, principalmente a longo prazo, com uma diferença de quase 10% em relação aos modelos de séries temporais.

O que torna os dados de mercado eficazes preditores de custos é a sua capacidade de considerar fatores externos que podem influenciar os preços do petróleo, como eventos geopolíticos imprevistos ou mudanças nas políticas comerciais. Sua atualização em tempo real permite que as empresas reajam rapidamente às mudanças no mercado global. A amplitude e variedade desses dados fornecem uma compreensão aprofundada do comportamento do mercado, algo que os modelos ARIMA, por sua natureza restrita a dados históricos, não conseguem capturar completamente. Ademais, no caso das séries temporais, a precisão dos resultados pode ser afetada por dados ausentes, *outliers* ou outros problemas de

qualidade. Em visto disso, o modelo fundamentado na Argus inicialmente proposto e desenvolvido foi selecionado como base para a continuação do projeto.

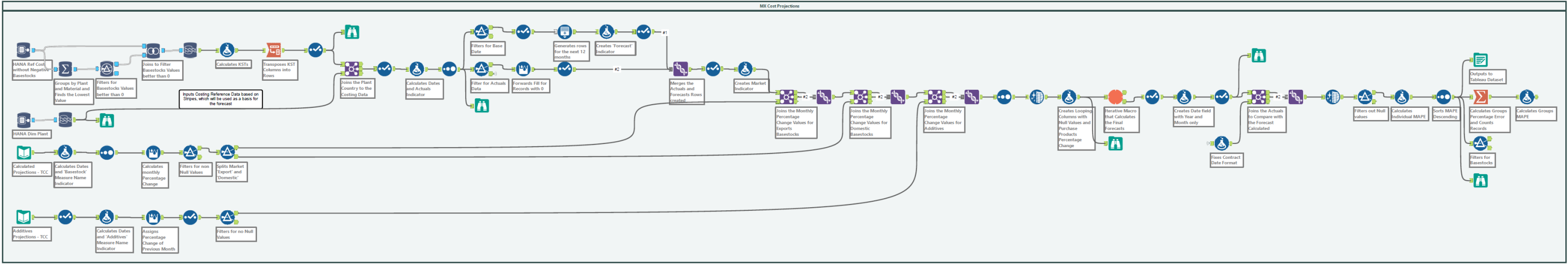
O *output* desse primeiro *workflow* é então publicado no *SharePoint*, uma plataforma de colaboração e gerenciamento de conteúdo desenvolvida pela *Microsoft*, para que os usuários possam ter acesso fácil ao arquivo. Ele gera uma previsão de custo dos óleos básicos para os próximos onze meses com base em fatores quantitativos, o que inclui dados numéricos e estatísticos. No entanto, reconhece-se que as influências qualitativas também desempenham um papel crucial nas previsões precisas. Para abordar essa lacuna, o processo permite que os usuários intervenham manualmente no meio do caminho, sobrepondo a previsão calculada com um valor estimado.

A inclusão de fatores qualitativos é fundamental, pois suas influências não podem ser totalmente capturadas por dados puramente quantitativos. As notícias, por exemplo, frequentemente contêm informações sobre eventos iminentes ou mudanças no mercado que não foram refletidos nos dados quantitativos utilizados inicialmente e que podem ter um impacto substancial nos custos dos materiais. A capacidade de incorporar fatores qualitativos permite uma resposta mais ágil às mudanças no ambiente de negócios. Isso é especialmente útil quando eventos imprevistos, como crises econômicas, flutuações nos preços das *commodities* ou anúncios governamentais, afetam os custos e as operações das empresas.

Ao permitir que os usuários adicionem esses fatores qualitativos às previsões quantitativas, o processo se torna mais adaptável e capaz de antecipar situações variadas, proporcionando uma visão mais holística das futuras tendências e desafios. Isso permite a adaptação das previsões para atender à singularidade de cada caso. Por fim, também possibilita a simulação de cenários, permitindo a análise do comportamento dos dados em diferentes situações hipotéticas e a compreensão de possíveis impactos de decisões e estratégias.

Nos casos nos quais o usuário não adicionar estimações de custo, serão utilizados os próprios valores de previsão calculados pelo modelo. O arquivo contendo essas informações é utilizado como um dos *inputs* do segundo *workflow* desenvolvido (Figura 14):

FIGURA 14 - Workflow em Alteryx para calcular as projeções de custos dos lubrificantes



FONTE: A autora, 2023

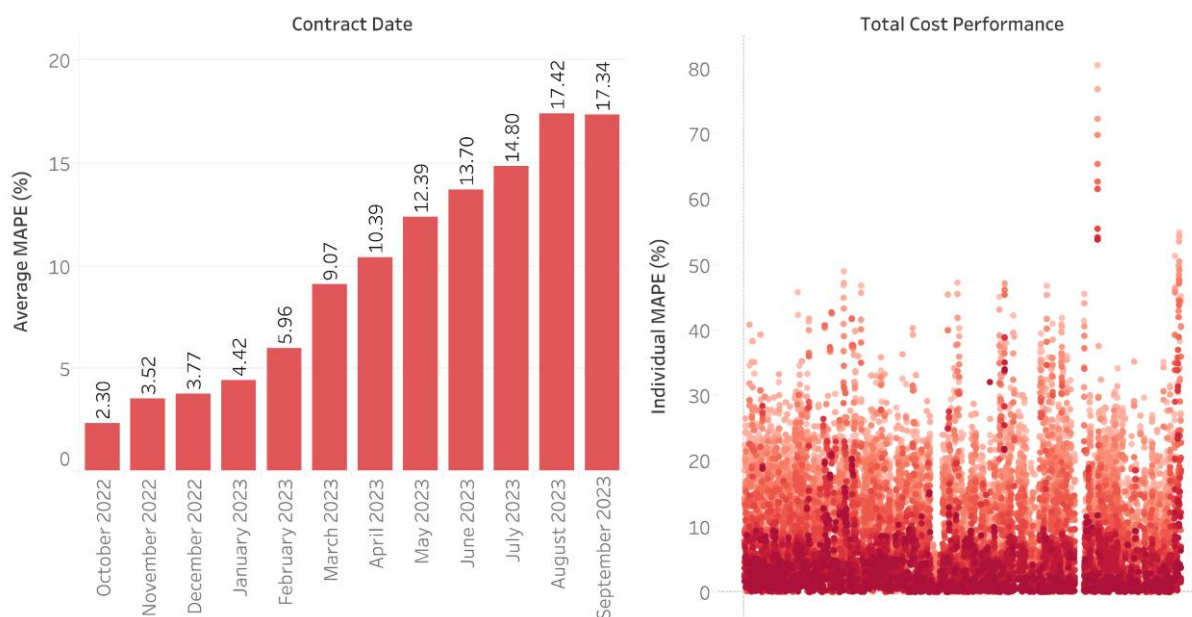
O objetivo é estimar o custo dos produtos finais a nível de seus componentes, de acordo com a estrutura apresentada previamente. Portanto, a previsão do custo do lubrificante como um todo será a soma das previsões individuais de cada um de seus componentes.

A previsão dos óleos básicos, o componente foco deste estudo, leva em conta dois principais elementos. O primeiro deles é obtido por meio da aplicação da variação percentual estimada referente ao grupo do produto em análise sobre seu custo mais recente. O segundo é o desvio padrão dos custos históricos do componente nos seis meses anteriores ao momento de previsão. Em seguida, são atribuídos pesos a esses valores, estabelecidos por meio de testes, com o auxílio de técnicas de regressão linear. Por meio de um *looping*, esse processo se repete, de forma que o resultado de uma previsão é imediatamente utilizado como entrada para a próxima previsão. Nessa primeira versão do modelo, os custos dos demais componentes foram calculados baseado apenas no elemento de fator histórico dos dados.

#### 4.5 AVALIAÇÃO

Para qualquer modelo de dados, a etapa de avaliação é fundamental para determinar a eficácia do modelo desenvolvido. Para avaliar o desempenho das previsões dos custos finais dos lubrificantes, envolvendo todos os componentes de custos, adotou-se a mesma métrica MAPE aplicada previamente. Foram comparados os custos de 1386 diferentes materiais em 12 plantas da organização, utilizando como intervalo de teste os meses entre outubro de 2022 e setembro de 2023. Os resultados do teste estão sintetizados nos gráficos apresentados na Figura 15.

FIGURA 15 - Métricas MAPE de avaliação do modelo



FONTE: A autora, 2023

No gráfico de barras à esquerda, observa-se que, à medida que o horizonte de previsão aumenta, o MAPE tende a aumentar. Isso ocorre, pois, padrões de curto prazo geralmente são mais fáceis e menos complexos de prever do que padrões de longo prazo, uma vez que esses últimos estão sujeitos a uma maior variabilidade devido ao impacto de fatores inesperados e imprevisíveis. Existem diversas outras razões que podem gerar esse resultado, como pequenos desvios nas previsões iniciais que podem se propagar e levar a grandes discrepâncias nas previsões a longo prazo.

O *workflow* apresentou um MAPE final de 3,2% para o próximo trimestre de previsão, e um MAPE de 9,5% para o próximo ano. Esse erro percentual foi aferido com base nas projeções calculadas pelo modelo, mas pode ser reduzido substancialmente considerando a possibilidade de sobreposição dessas por estimativas manuais como *input* do modelo.

O gráfico à direita na Figura 15 ilustra os mesmos dados, porém em uma granularidade a nível de material. Os pontos estão coloridos de acordo com o horizonte de previsão, ou seja, o número de períodos de tempo no futuro para os quais as previsões estão sendo feitas. Desta forma, as previsões de custos para o próximo mês recebem a cor mais escura, e as previsões mais distantes, uma cor mais clara. Percebe-se que os dados apresentam uma tendência de dispersão em relação ao eixo

X à medida em que esse horizonte de previsão aumenta, simbolizando um MAPE mais elevado.

A avaliação de um bom MAPE pode variar dependendo do contexto específico da aplicação. No caso do presente estudo, as expectativas iniciais dos *stakeholders* envolvidos foram atendidas.

Quanto à performance do modelo, o primeiro *workflow* possui um tempo de processamento total médio de 29 segundos, enquanto o segundo requer aproximadamente 8 minutos e 26 segundos. A redução de tempo proporcionada foi extremamente significativa, considerando que esse processo era realizado manualmente ao longo de todo o mês.

#### 4.6 IMPLEMENTAÇÃO

A previsão de custos é essencial para o planejamento financeiro e estratégico de uma organização, preparando-a para potenciais despesas e auxiliando na tomada de decisões informadas, como, por exemplo, a negociação com fornecedores. Em empresas de grande porte como a de estudo, os contratos de custos com fornecedores são documentos fundamentais que estabelecem acordos detalhados entre a companhia e seus fornecedores. Frequentemente são negociados contratos de custos com fornecedores em larga escala, visando garantir preços competitivos, prazos de pagamento favoráveis e condições que atendam às necessidades da empresa. Uma vez que um contrato é estabelecido para determinado produto e o custo dele está previamente determinado entre as partes, não se faz mais necessária a previsão desse. Dessa forma, até que o contrato se encerre, é de maior valor para a empresa exibir os custos acordados com o fornecedor, em oposição às previsões calculadas pelo modelo.

Como forma de adicionar a possibilidade de inclusão dos valores acordados com fornecedores, foi criada uma tabela para ser alimentada pelo departamento de compras da empresa. Cada registro no arquivo define alterações de preços nas seguintes esferas, ou em uma combinação delas: fornecedor, componente, grupo do componente e região da planta. Vários registros podem afetar o mesmo produto. Por exemplo, um fornecedor pode anunciar um aumento de preço amplo de 10 USD/BBL, bem como um aumento de 15 USD/BBL para um componente específico. Todas as alterações de preço anunciadas no mesmo mês para o mesmo componente serão

aplicadas cumulativamente. Elas podem ser tanto valores absolutos quanto valores relativos, mensurados em USD/BBL ou em porcentagens. Quando há várias alterações de preços para um componente para o mesmo mês, as alterações percentuais são somadas primeiro e aplicadas ao preço base, e em seguida, as alterações em USD/BBL são adicionadas.

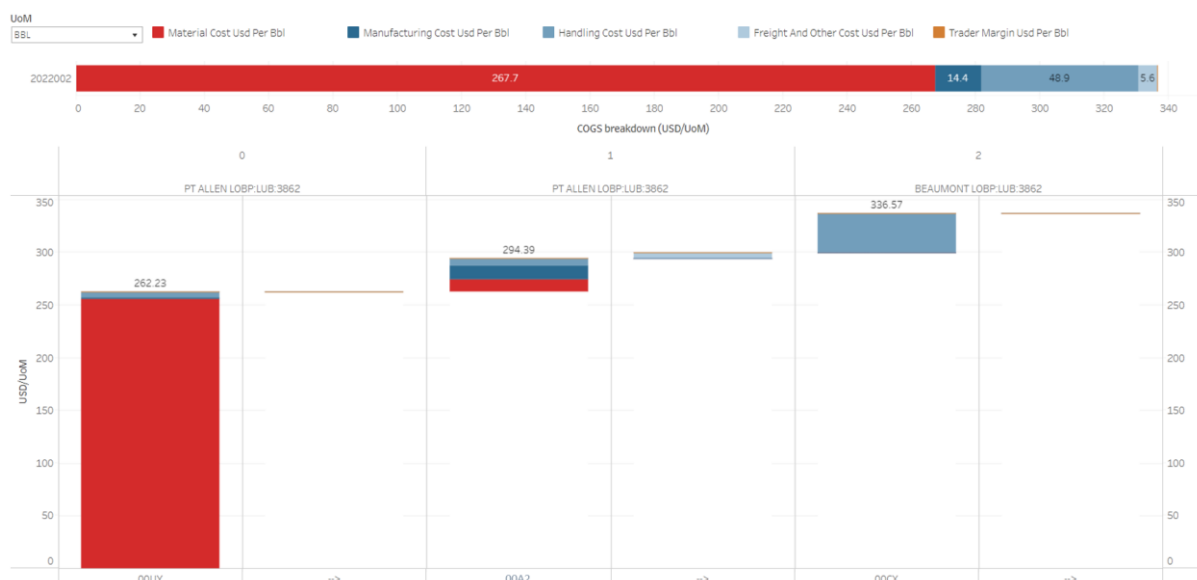
Usualmente, a planta da qual os materiais são vendidos para o cliente final não é a planta na qual eles foram fabricados. No cenário atual de globalização logística, os materiais podem ser produzidos em um local e serem posteriormente movidos diversas vezes para diferentes regiões, além da possibilidade de serem reembalados e renomeados nesse processo. Às vezes, o movimento pode até ser entre continentes. No caso do arquivo alimentado pelo departamento de compras, as alterações de preços são referentes às unidades de produção, ou seja, o início da cadeia de suprimentos. A alteração de preços nessas plantas iniciais pode desencadear uma série de ajustes e negociações ao longo da cadeia, afetando os valores nas plantas intermediárias e, eventualmente, nos preços finais para o consumidor. Isto posto, é de suma importância ter conhecimento do caminho percorrido pelo material desde sua origem, para assim poder calcular a propagação de possíveis mudanças de preços no restante da cadeia.

A empresa em estudo não possuía tal informação disponível, trabalhando somente com transações isoladas entre plantas. À vista disso, foi necessária a criação de um modelo de integração no SAP HANA que fornecesse detalhes sobre toda a cadeia de suprimentos. O principal objetivo é descobrir, para cada material vendido em um determinado local em um determinado período, qual sua planta, material e fornecedor de origem, assim como calcular os custos de cada etapa da cadeia, bem como do todo.

Utilizando como base as transações de vendas, estoque nas unidades de armazenamento, e dados da programação e planejamento da produção e logística, o modelo opera por meio de um *looping*, que busca determinar a planta e material de origem em cada etapa em análise, começando pela planta na qual a venda foi realizada. Ao final de cada iteração, os dados são reorganizados, e os campos identificados referentes a procedência do produto assumem a posição dos campos de destino, cuja origem deseja-se determinar. Esse processo continua até que não haja nenhum registro de fornecimento correspondente encontrado e, logo, assume-se que foram encontrados o local e material de origem almejados.

Foi criada uma *dashboard* no Tableau para apresentar visualmente a integração proporcionada pelo modelo (Figura 16):

FIGURA 16 - Integração e formação de custo na Cadeia de Suprimentos



FONTE: A autora, 2023

Por meio de um gráfico em cascata, é representada a formação de custos ao longo da cadeia de suprimentos, divididos em cinco principais categorias, similares às abordadas anteriormente: Material, Produção, Manuseio, Margem do Comerciante e Frete e outros custos. No exemplo da Figura 16, o produto em análise foi produzido em uma planta e posteriormente movido para outra para ser embalado. Isso pode ser constatado uma vez que na segunda etapa da cadeia houve majoritariamente custos com produção e manuseio, em contradição à um pequeno custo de materiais, referente ao valor das embalagens.

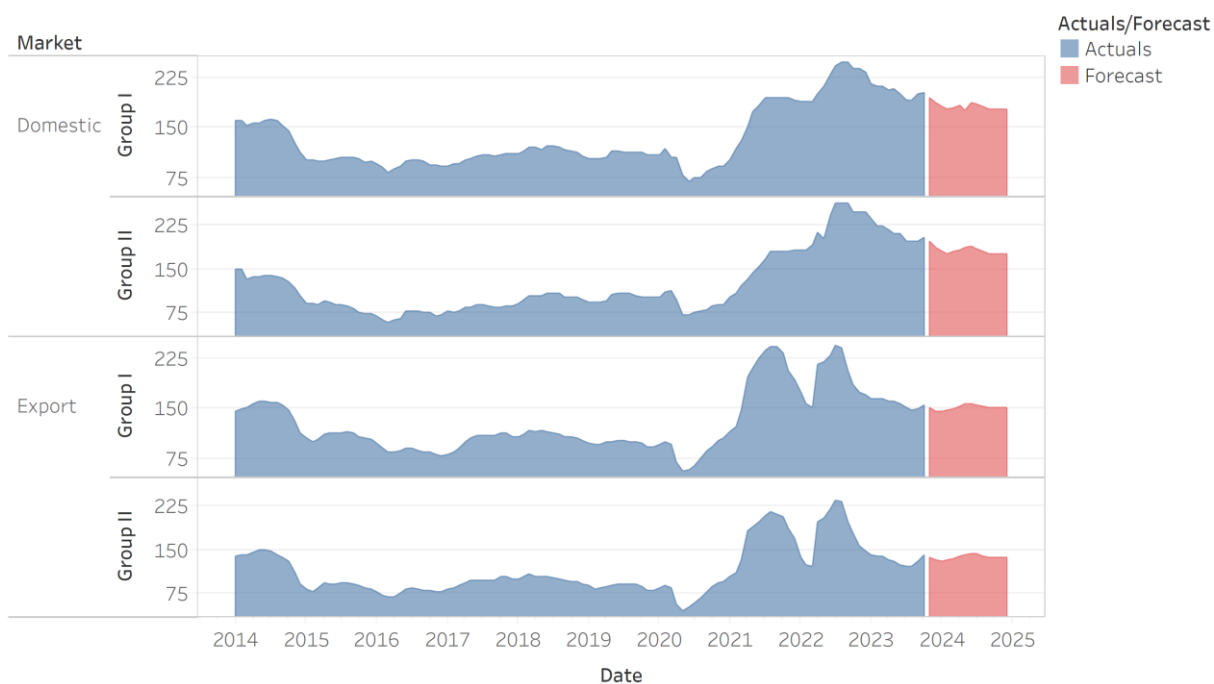
O usuário pode interagir com a *dashboard* por meio de filtros de Planta, Produto Final ou Intermediário, Matéria Prima, Período Fiscal, Região e número de etapas da cadeia de suprimentos. Mais detalhes são obtidos ao se passar o *mouse* por cima do gráfico.

Com o objetivo de consolidar todos os dados calculados e disponibilizados até este ponto, foi desenvolvido um *workflow* final no SAP HANA, construído em cima dos dados históricos de custos dos produtos e unindo o relatório de previsão de custos gerado pelo Alteryx, a planilha atualizada pelo setor de compras com os anúncios de

mudanças de preços dos fornecedores e o modelo de integração da cadeia de suprimentos.

O resultado calculado alimenta diretamente uma *dashboard* no Tableau, proporcionando uma representação visual dos dados e facilitando a interpretação das informações. Na Figura 17, o gráfico permite uma análise da evolução e tendência do preço dos grupos de óleos básicos nos mercados internacional e doméstico:

FIGURA 17 - Evolução e tendência do preço dos grupos de óleos básicos



FONTE: A autora, 2023

Para uma análise mais minuciosa, o usuário pode recorrer à uma tabela contendo planta, material, componente de custo e valores mensais (Figura 18):

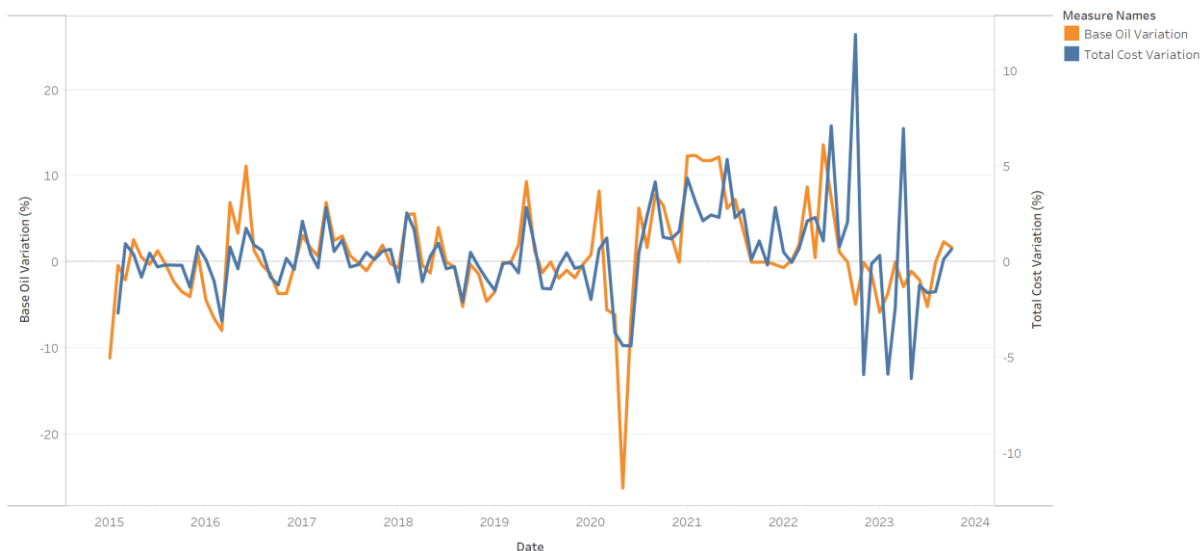
FIGURA 18 - Análise mensal dos componentes de custo dos lubrificantes

| Plant | Material Num and..             | Cost Components             | Actuals   |             |                | Forecast     |               |               |              |               |            |            |
|-------|--------------------------------|-----------------------------|-----------|-------------|----------------|--------------|---------------|---------------|--------------|---------------|------------|------------|
|       |                                |                             | July 2022 | August 2022 | September 2022 | October 2022 | November 2022 | December 2022 | January 2023 | February 2023 | March 2023 | April 2023 |
| 00A9  | 100000 - M-DTE OIL HVYMED BULK | 1. Basestocks               | 243.35    | 263.75      | 263.38         | 266.12       | 253.21        | 238.69        | 224.07       | 213.42        | 209.00     | 211.45     |
|       |                                | 3. Additives/Synth Base     | 12.89     | 13.21       | 14.30          | 14.86        | 15.21         | 15.57         | 15.91        | 15.88         | 15.82      | 15.74      |
|       |                                | 7. Intermed Transp COGS     | 0.26      | 0.26        | 0.30           | 0.30         | 0.30          | 0.30          | 0.30         | 0.31          | 0.31       | 0.31       |
|       |                                | 8. Blending                 | 2.93      | 2.93        | 2.93           | 2.93         | 2.94          | 2.94          | 2.94         | 2.94          | 2.94       | 2.94       |
|       |                                | 11. Lab                     | 3.69      | 3.69        | 3.69           | 3.69         | 3.69          | 3.70          | 3.70         | 3.70          | 3.70       | 3.70       |
|       |                                | 13. Bulk Material Handling  | 4.72      | 4.36        | 4.72           | 4.74         | 4.76          | 4.77          | 4.79         | 4.80          | 4.80       | 4.81       |
|       |                                | Total Cost                  | 267.94    | 287.97      | 289.42         | 292.23       | 294.82        | 297.15        | 298.11       | 298.49        | 298.82     | 299.05     |
| 00CX  | 100000 - M-DTE OIL HVYMED BULK | 1. Basestocks               | 239.16    | 255.66      | 270.41         | 272.58       | 258.85        | 243.67        | 228.91       | 217.96        | 213.50     | 216.02     |
|       |                                | 2. Slop                     | 0.13      | 0.13        | 0.13           | 0.14         | 0.14          | 0.14          | 0.14         | 0.14          | 0.14       | 0.14       |
|       |                                | 3. Additives/Synth Base     | 12.89     | 13.15       | 14.38          | 14.90        | 15.22         | 15.55         | 15.90        | 15.87         | 15.81      | 15.73      |
|       |                                | 8. Blending                 | 2.13      | 2.13        | 2.13           | 2.14         | 2.14          | 2.15          | 2.15         | 2.15          | 2.15       | 2.15       |
|       |                                | 11. Lab                     | 2.00      | 2.00        | 2.00           | 2.00         | 2.00          | 2.00          | 2.00         | 2.00          | 2.00       | 2.00       |
|       |                                | 13. Bulk Material Handling  | 7.10      | 4.90        | 7.10           | 7.21         | 7.32          | 7.43          | 7.52         | 7.61          | 7.62       | 7.64       |
|       |                                | 14. Packed Material Handl.. | 0.22      | 0.22        | 0.22           | 0.22         | 0.22          | 0.22          | 0.22         | 0.22          | 0.22       | 0.22       |
| 00FG  | 100000 - M-DTE OIL HVYMED BULK | 16. 3P Handling             | 5.54      | 5.54        | 5.54           | 5.82         | 6.09          | 6.30          | 6.33         | 6.36          | 6.39       | 6.41       |
|       |                                | Total Cost                  | 269.22    | 277.97      | 301.96         | 304.38       | 306.77        | 309.11        | 310.65       | 311.75        | 312.10     | 312.37     |
|       |                                | 1. Basestocks               | 239.16    | 255.66      | 270.41         | 272.58       | 258.85        | 243.67        | 228.91       | 217.96        | 213.50     | 216.02     |
|       |                                | 2. Slop                     | 0.13      | 0.13        | 0.13           | 0.14         | 0.14          | 0.14          | 0.14         | 0.14          | 0.14       | 0.14       |
|       |                                | 3. Additives/Synth Base     | 12.89     | 13.15       | 14.38          | 14.90        | 15.22         | 15.55         | 15.90        | 15.87         | 15.81      | 15.73      |
|       |                                | 7. Intermed Transp COGS     | 8.12      | 8.08        | 8.22           | 8.23         | 8.23          | 8.24          | 8.25         | 8.25          | 8.25       | 8.25       |
|       |                                | 8. Blending                 | 2.13      | 2.13        | 2.13           | 2.14         | 2.14          | 2.15          | 2.15         | 2.15          | 2.15       | 2.15       |
|       |                                | 11. Lab                     | 2.00      | 2.00        | 2.00           | 2.00         | 2.00          | 2.00          | 2.00         | 2.00          | 2.00       | 2.00       |
|       |                                | 13. Bulk Material Handling  | 13.43     | 11.23       | 10.23          | 10.40        | 10.51         | 10.61         | 10.72        | 10.76         | 10.77      | 10.79      |
|       |                                | 14. Packed Material Handl.. | 0.22      | 0.22        | 0.22           | 0.22         | 0.22          | 0.22          | 0.22         | 0.22          | 0.22       | 0.22       |
|       |                                | 15. Intermed Transp OPEX    | 2.00      | 2.00        | 2.00           | 2.04         | 2.04          | 2.04          | 2.04         | 2.04          | 2.04       | 2.04       |
|       |                                | 16. 3P Handling             | 5.54      | 5.54        | 5.54           | 5.82         | 6.09          | 6.30          | 6.33         | 6.36          | 6.39       | 6.41       |
|       |                                | Total Cost                  | 285.62    | 294.38      | 315.27         | 317.54       | 319.80        | 322.00        | 323.39       | 324.36        | 324.69     | 324.94     |

FONTE: A autora, 2023

O Gráfico 5, por sua vez, ilustra a relação entre a variação percentual mensal do preço dos óleos básicos e a variação percentual mensal do custo total dos lubrificantes finalizados.

GRÁFICO 5 - Relação entre o custo dos óleos básicos e o custo dos lubrificantes



FONTE: A autora, 2023

Detecta-se um padrão similar entre a tendência das duas linhas, mesmo que em diferentes escalas, fruto do impacto direto que a variação do preço da matéria prima possui no custo do produto final. O gráfico permite facilmente identificar essas relações. Até 2020, por exemplo, as variações nos custos da matéria prima representavam, em geral, um impacto de 50% nos custos finais. Dessa forma, um aumento de 20% no valor dos óleos básicos resultava em um aumento de 10% no custo dos lubrificantes.

Percebe-se que, em alguns casos, as flutuações nos preços das matérias-primas não são imediatamente refletidas no custo final dos produtos, sendo percebidas nos períodos posteriores a mudança. Isso ocorre por inúmeros motivos. Primeiramente, o estabelecimento de contratos de compra a prazo com fornecedores pode garantir preços fixos por um período específico, independentemente das mudanças no mercado. Além disso, as empresas costumam manter estoques significativos de matérias-primas. Os custos desses estoques existentes não são afetados pelas mudanças nos preços até que precisem ser renovados a preços mais altos.

No caso da empresa em estudo, mudanças nos preços das matérias-primas podem levar tempo para se propagar através das inúmeras etapas da cadeia de suprimentos, nas quais contratos, transporte internacional e negociações com intermediários podem atrasar a adaptação aos novos preços.

Em uma forma de sintetizar o impacto de cada componente de custo sobre o produto final, foi criada uma tabela contendo a variação média mensal, em unidade de custo por unidade de volume, de cada um desses componentes (Figura 19).

FIGURA 19 - Impacto médio dos componentes de custo no produto final

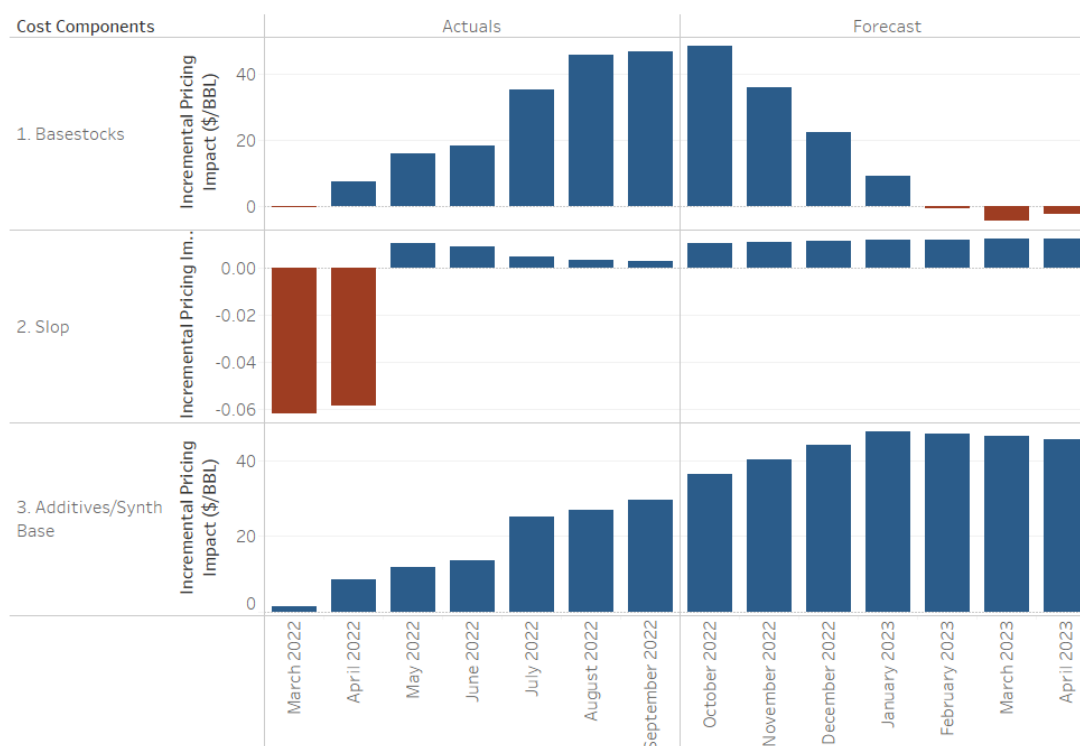
| Cost Components                     |                              | Actuals    |            |          |           |           |             |                | Forecast     |               |               |
|-------------------------------------|------------------------------|------------|------------|----------|-----------|-----------|-------------|----------------|--------------|---------------|---------------|
|                                     |                              | March 2022 | April 2022 | May 2022 | June 2022 | July 2022 | August 2022 | September 2022 | October 2022 | November 2022 | December 2022 |
| Pricing Impact (\$/BBL)             | 1. Basestocks                | (0.10)     | 7.63       | 8.42     | 2.25      | 17.00     | 10.52       | 0.97           | 1.76         | (12.44)       | (13.80)       |
|                                     | 2. Slop                      | (0.06)     | 0.00       | 0.07     | 0.00      | 0.00      | 0.00        | 0.00           | 0.01         | 0.00          | 0.00          |
|                                     | 3. Additives/Synth Base      | 1.33       | 7.03       | 3.28     | 1.87      | 11.51     | 1.84        | 2.76           | 6.56         | 3.89          | 3.89          |
|                                     | 4. Packaging Components      | 0.07       | (0.87)     | (0.92)   | (0.09)    | 0.21      | (1.83)      | 1.91           | 0.13         | 0.10          | 0.10          |
|                                     | 6. Pallets                   | 0.00       | 0.00       | 0.88     | 0.00      | 2.23      | (2.23)      | 0.01           | 0.12         | 0.09          | 0.08          |
|                                     | 7. Intermed Transp COGS      | 0.00       | (0.02)     | 0.26     | 0.07      | 0.27      | (0.42)      | 0.61           | 0.07         | 0.06          | 0.06          |
|                                     | 8. Blending                  | 0.69       | (1.97)     | (0.01)   | (0.02)    | 0.18      | (0.03)      | 0.00           | 0.05         | 0.05          | 0.04          |
|                                     | 9. Packaging                 | (0.04)     | (1.66)     | (0.09)   | (0.08)    | 0.15      | (0.02)      | (0.02)         | 0.05         | 0.05          | 0.03          |
|                                     | 11. Lab                      | 0.69       | (0.13)     | 0.01     | (0.03)    | (0.15)    | (0.01)      | 0.00           | 0.02         | 0.02          | 0.01          |
|                                     | 13. Bulk Material Handling   | (0.08)     | (0.11)     | (0.25)   | 0.14      | 0.19      | (1.19)      | 0.65           | 0.13         | 0.08          | 0.08          |
|                                     | 14. Packed Material Handling | (0.01)     | (0.02)     | (1.24)   | 0.02      | 0.09      | (0.33)      | 0.30           | 0.07         | 0.03          | 0.02          |
|                                     | 15. Intermed Transp OPEX     | (0.04)     | 0.00       | (0.12)   | 0.00      | (0.01)    | (0.01)      | 0.04           | 0.01         | 0.00          | 0.00          |
|                                     | 16. 3P Handling              | (0.02)     | 0.00       | (0.02)   | 0.01      | 0.17      | 0.01        | 0.10           | 0.02         | 0.02          | 0.01          |
|                                     | Total Cost                   | 2.41       | 9.81       | 10.11    | 4.10      | 31.84     | 4.04        | 9.53           | 2.85         | 2.64          | 2.23          |
| Incremental Pricing Impact (\$/BBL) | 1. Basestocks                | (0.10)     | 7.53       | 15.95    | 18.21     | 35.20     | 45.72       | 46.69          | 48.45        | 36.01         | 22.21         |
|                                     | 2. Slop                      | (0.06)     | (0.06)     | 0.01     | 0.01      | 0.01      | 0.00        | 0.00           | 0.01         | 0.01          | 0.01          |
|                                     | 3. Additives/Synth Base      | 1.33       | 8.36       | 11.63    | 13.50     | 25.01     | 26.85       | 29.61          | 36.18        | 40.07         | 43.95         |
|                                     | 4. Packaging Components      | 0.07       | (0.80)     | (1.72)   | (1.81)    | (1.60)    | (3.43)      | (1.51)         | (1.38)       | (1.28)        | (1.18)        |
|                                     | 6. Pallets                   | 0.00       | 0.01       | 0.89     | 0.89      | 3.12      | 0.90        | 0.90           | 1.02         | 1.11          | 1.19          |
|                                     | 7. Intermed Transp COGS      | 0.00       | (0.02)     | 0.24     | 0.31      | 0.58      | 0.16        | 0.77           | 0.84         | 0.91          | 0.97          |
|                                     | 8. Blending                  | 0.69       | (1.28)     | (1.29)   | (1.31)    | (1.13)    | (1.16)      | (1.16)         | (1.11)       | (1.07)        | (1.03)        |
|                                     | 9. Packaging                 | (0.04)     | (1.70)     | (1.78)   | (1.87)    | (1.71)    | (1.73)      | (1.76)         | (1.70)       | (1.66)        | (1.62)        |
|                                     | 11. Lab                      | 0.69       | 0.56       | 0.57     | 0.54      | 0.39      | 0.39        | 0.38           | 0.40         | 0.42          | 0.43          |
|                                     | 13. Bulk Material Handling   | (0.08)     | (0.19)     | (0.44)   | (0.30)    | (0.11)    | (1.30)      | (0.65)         | (0.52)       | (0.44)        | (0.36)        |
|                                     | 14. Packed Material Handling | (0.01)     | (0.03)     | (1.27)   | (1.25)    | (1.16)    | (1.49)      | (1.19)         | (1.12)       | (1.10)        | (1.07)        |
|                                     | 15. Intermed Transp OPEX     | (0.04)     | (0.05)     | (0.17)   | (0.17)    | (0.18)    | (0.18)      | (0.14)         | (0.13)       | (0.13)        | (0.13)        |
|                                     | 16. 3P Handling              | (0.02)     | (0.02)     | (0.04)   | (0.03)    | 0.13      | 0.14        | 0.24           | 0.26         | 0.28          | 0.29          |
|                                     | Total Cost                   | 2.41       | 12.22      | 22.33    | 26.43     | 58.27     | 62.32       | 71.84          | 74.69        | 77.33         | 79.56         |

FONTE: A autora, 2023

Dessa forma, permite-se, para os meses passados, a análise da incidência que cada componente teve no custo final do produto, assim como a projeção do impacto esperado para períodos futuros. No caso capturado na Figura 19, o usuário teria a sua disposição a previsão de que, para outubro de 2022, os lubrificantes apresentariam um aumento médio de 2.85 USD/BBL em relação ao mês anterior, sendo 1.76 USD/BBL desse valor proveniente dos custos dos óleos básicos. A tabela também calcula o impacto acumulado dessas mudanças.

Essa mesma informação é exibida na forma de gráfico na Figura 20:

FIGURA 20 - Impacto médio acumulado dos componentes de custo no produto final



FONTE: A autora, 2023

Tendo as etapas de Modelagem, Avaliação e Implementação do modelo finalizadas, redigiu-se um relatório para a apresentação dos resultados e *insights* iniciais para os gestores da empresa, além de um treinamento sobre como utilizar a nova ferramenta desenvolvida e boas práticas.

A implementação desse modelo ofereceu diversas vantagens estratégicas à organização. Ao analisar e antecipar as flutuações nos preços das matérias-primas, a companhia poderá tomar decisões mais informadas em relação à suas negociações com fornecedores, precificações e consequentes margens de lucro. Ademais, esse modelo permite uma gestão proativa de riscos, permitindo a adoção de estratégias para mitigar os impactos de aumentos de custos, como a busca por novos fornecedores, a otimização dos níveis de estoque, ou a utilização de formulações alternativas para os mesmos produtos.

A antecipação desses custos e receitas por meio de simulações permite aprimorar o planejamento orçamentário, resultando em uma gestão financeira mais eficiente. Esse conhecimento também permite uma resposta rápida a mudanças abruptas no mercado, mantendo a competitividade.

Uma gestão eficaz de preços não apenas impulsiona a eficiência operacional da companhia, mas também melhora a satisfação dos clientes, garantindo preços justos e competitivos e uma maior transparência com esses. Uma vez que a projeção de custos foi calculada, a empresa poderá informar antecipadamente seus compradores sobre os impactos que as alterações de preços previstas terão em seus produtos. Isso lhes dará tempo para replanejar e aumentará a probabilidade de fidelização. Em última análise, a utilização de modelos de previsão e simulação também impulsiona a inovação e a competitividade no mercado, sendo uma estratégia essencial para empresas modernas.

O modelo será utilizado e ficará sob a responsabilidade dos setores de Vendas, Compras e Otimização de Produtos da empresa, que foram devidamente instruídos quanto à sua manutenção e monitoramento.

Embora tenha apresentado desempenho satisfatório, o modelo pode ainda ser incrementado em um novo ciclo das etapas da metodologia CRISP-DM. A aplicação e resultados desse protótipo inicial do projeto despertaram o interesse dos tomadores de decisão da empresa em realizar o mesmo estudo para a previsão dos demais componentes de custo dos produtos, como aditivos, transporte e armazenamento. Sugere-se também a utilização do cálculo de projeção individual dos óleos básicos como base do modelo, alternativamente à média dos grupos a qual pertencem. Os aperfeiçoamentos e a incorporação de novos cálculos e visualizações permitirão uma maior precisão quanto à previsão dos custos finais dos lubrificantes da empresa.

## 5 CONSIDERAÇÕES FINAIS

Este trabalho teve como principal objetivo estruturar e implementar um modelo de previsão de custos de modo a oferecer suporte à empresa para a tomada de decisões estratégicas. O referencial teórico redigido permitiu a aquisição de uma bagagem de conhecimentos suficiente para a condução do projeto no setor de Cadeia de Suprimento de uma petrolífera multinacional.

Mediante a metodologia CRISP-DM, a implementação do estudo foi conduzida de forma sistemática, proporcionando uma base robusta para a obtenção de resultados significativos e a geração de valor tangível para a empresa. Destacaram-se como ferramentas principais utilizadas o Alteryx e o Tableau, selecionadas pela facilidade de utilização e capacidade de integração com as demais aplicações da empresa. O erro percentual absoluto obtido pelo modelo foi satisfatório, com um valor médio de 3,2% para um intervalo de previsão de três meses futuros.

De imediato, pode-se considerar que o projeto gerou vantagens significativas para a organização no que se refere à possibilidade de adoção de estratégias fundamentadas para mitigar os impactos de aumentos de custos frente às previsões geradas. Tais estratégias incluem negociações informadas com fornecedores, mudanças nas precificações, otimização de níveis de estoque e utilização de formulações alternativas para os produtos. Esse conhecimento permite uma resposta antecipada às mudanças abruptas no mercado, mantendo a competitividade frente a concorrência.

Como trabalhos futuros propõe-se realizar uma análise quantitativa dos benefícios capturados pela empresa após a implementação do modelo. Ademais, como um segundo ciclo do CRISP-DM, sugere-se a aplicação de novos testes e cálculos inerentes aos demais componentes de custo, como aditivos, transporte e armazenamento, que possivelmente aumentem a precisão das previsões de custo dos produtos finais.

## REFERÊNCIAS

PORTER, Michael E. Competitive strategy. **Measuring business excellence**, v. 1, n. 2, p. 12-17, 1997.

MAHANTY, Aroop K. **Intermediate microeconomics with applications**. Academic Press, 2014.

ARMSTRONG, J. Scott. **Combining forecasts**. Springer US, 2001.

LAMBERT, Douglas M.; COOPER, Martha C.; PAGH, Janus D. Supply chain management: implementation issues and research opportunities. **The international journal of logistics management**, v. 9, n. 2, p. 1-20, 1998.

GOLAN, Maureen S.; JERNEGAN, Laura H.; LINKOV, Igor. Trends and applications of resilience analytics in supply chain modeling: systematic literature review in the context of the COVID-19 pandemic. **Environment Systems and Decisions**, v. 40, n. 2, p. 222-243, 2020.

MENTZER, John T. et al. Defining supply chain management. **Journal of Business logistics**, v. 22, n. 2, p. 1-25, 2001.

MURINO, Teresa; ROMANO, Elpidio; SANTILLO, Liberatina C. Supply chain performance sustainability through resilience function. In: **Proceedings of the 2011 Winter Simulation Conference (WSC)**. IEEE, 2011. p. 1600-1611

TANG, Christopher S. Perspectives in supply chain risk management. **International journal of production economics**, v. 103, n. 2, p. 451-488, 2006.

GIUNIPERO, Larry C.; ALY ELTANTAWY, Reham. Securing the upstream supply chain: a risk management approach. **International journal of physical distribution & logistics management**, v. 34, n. 9, p. 698-713, 2004.

BEAMON, Benita M. Supply chain design and analysis:: Models and methods. **International journal of production economics**, v. 55, n. 3, p. 281-294, 1998.

NEW, Steve; WESTBROOK, Roy (Ed.). **Understanding supply chains: concepts, critiques, and futures**. OUP Oxford, 2004.

MISRA, Vikas; KHAN, M. I.; SINGH, U. K. Supply chain management systems: architecture, design and vision. **Journal of Strategic Innovation and Sustainability**, v. 6, n. 4, p. 96-101, 2010.

TAN, Keah Choon. Supply chain management: practices, concerns, and performance issues. **Journal of Supply Chain Management**, v. 38, n. 4, p. 42-53, 2002.

CHANDRA, Charu; KUMAR, Sameer. Supply chain management in theory and practice: a passing fad or a fundamental change?. **Industrial Management & Data Systems**, 2000.

AUTRY, Chad W.; GOLICIC, Susan L. Evaluating buyer–supplier relationship–performance spirals: A longitudinal study. **Journal of operations management**, v. 28, n. 2, p. 87-100, 2010.

COUNCIL OF SUPPLY CHAIN MANAGEMENT PROFESSIONALS (CSCMP). **SCM definitions and glossary of terms**. Disponível em: <<https://cscmp.org/CSCMP>>. Acesso em: 28 abr. 2023.

SOUZA, Gilvan C. Supply chain analytics. **Business Horizons**, v. 57, n. 5, p. 595-605, 2014.

HAND, David J. Strength in diversity: the advance of data analysis. In: **Knowledge Discovery in Databases: PKDD 2004: 8th European Conference on Principles and Practice of Knowledge Discovery in Databases, Pisa, Italy, September 20-24, 2004. Proceedings 8**. Springer Berlin Heidelberg, 2004. p. 18-26.

ABDUL-JABBAR, Safa S. et al. Data Analytics and Techniques. **ARO-THE SCIENTIFIC JOURNAL OF KOYA UNIVERSITY**, v. 10, n. 2, p. 45-55, 2022.

RAJARAMAN, V. Big data analytics. **Resonance**, v. 21, p. 695-716, 2016.

AYDINER, Arafat Salih et al. Business analytics and firm performance: The mediating role of business process performance. **Journal of business research**, v. 96, p. 228-237, 2019.

EPPLER, Martin J.; MENGIS, Jeanne. The Concept of Information Overload-A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines (2004) *The Information Society: An International Journal*, 20 (5), 2004, pp. 1–20. **Kommunikationsmanagement im Wandel: Beiträge aus 10 Jahren= mcm institute**, p. 271-305, 2008.

MALTBY, Dylan. Big data analytics. In: **74th Annual Meeting of the Association for Information Science and Technology (ASIST)**. 2011. p. 1-6.

DAVENPORT, Thomas; HARRIS, Jeanne. **Competing on analytics: Updated, with a new introduction: The new science of winning**. Harvard Business Press, 2017.s

RUSSOM, Philip et al. Big data analytics. **TDWI best practices report, fourth quarter**, v. 19, n. 4, p. 1-34, 2011.

GANTZ, John; REINSEL, David. The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. **IDC iView: IDC Analyze the future**, v. 2007, n. 2012, p. 1-16, 2012.

WHITE, Martin. Digital workplaces: Vision and reality. **Business information review**, v. 29, n. 4, p. 205-214, 2012.

GANDOMI, Amir; HAIDER, Murtaza. Beyond the hype: Big data concepts, methods, and analytics. **International journal of information management**, v. 35, n. 2, p. 137-144, 2015.

TRKMAN, Peter et al. The impact of business analytics on supply chain performance. **Decision Support Systems**, v. 49, n. 3, p. 318-327, 2010.

BISWAS, Sanjib; SEN, Jaydip. A proposed framework of next generation supply chain management using big data analytics. In: **Proceedings of National Conference on Emerging Trends in Business and Management: Issues and Challenges**. 2016.

LUHN, Hans Peter. A business intelligence system. **IBM Journal of research and development**, v. 2, n. 4, p. 314-319, 1958.

DRESNER, Howard. Business intelligence. **Gartner Inc**, 1989.

SAHAY, B. S.; RANJAN, Jayanthi. Real time business intelligence in supply chain analytics. **Information Management & Computer Security**, v. 16, n. 1, p. 28-48, 2008.

TURBAN, Efraim et al. **Business intelligence: um enfoque gerencial para a inteligência do negócio**. Bookman Editora, 2009.

INMON, William H. **Building the data warehouse**. John wiley & sons, 2005.

IMHOFF, Claudia; LOFTIS, Lisa; GEIGER, Jonathan G. **Building the customer-centric enterprise: Data warehousing techniques for supporting customer relationship management**. Wiley, 2001.

LANGLOIS, Audrey; CHAUVEL, Benjamin. The impact of supply chain management on business intelligence. **Journal of Intelligence Studies in Business**, v. 7, n. 2, 2017.

CHAUDHURI, Surajit; DAYAL, Umeshwar; NARASAYYA, Vivek. An overview of business intelligence technology. **Communications of the ACM**, v. 54, n. 8, p. 88-98, 2011.

DA SILVA, Denilson et al. INTELIGÊNCIA DE NEGÓCIO. **Maiêutica-Tecnologias da Informação**, v. 1, n. 01, 2016.

DANIEL, Florian et al. Managing Data Quality in Business Intelligence Applications. In: **QDB/MUD**. 2008. p. 133-143.

LILLIS, Theresa M.; CURRY, Mary Jane. **Academic writing in global context**. London: Routledge, 2010.

NIKOLENKO, Sergey I. **Synthetic data for deep learning**. Springer Nature, 2021.

OECD. **Frascati Manual: Proposed Standard Practice for Surveys of Measurement of Research and Experimental Development**. OECD, Paris, 2002.

CRESWELL, John W.; CRESWELL, J. David. **Research design: Qualitative, quantitative, and mixed methods approaches**. Sage publications, 2017.

MARTÍNEZ-PLUMED, Fernando et al. CRISP-DM twenty years later: From data mining processes to data science trajectories. **IEEE Transactions on Knowledge and Data Engineering**, v. 33, n. 8, p. 3048-3061, 2019.

WIRTH, Rüdiger; HIPPE, Jochen. CRISP-DM: Towards a standard process model for data mining. In: **Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining**. 2000. p. 29-39.

XIE, Wen et al. A new method for crude oil price forecasting based on support vector machines. In: **Computational Science-ICCS 2006: 6th International Conference, Reading, UK, May 28-31, 2006, Proceedings, Part IV 6**. Springer Berlin Heidelberg, 2006. p. 444-451.

SILVEIRA, Eva Lúcia Cardoso et al. Determinação de contaminantes em óleos lubrificantes usados e em esgotos contaminados por esses lubrificantes. **Química Nova**, v. 29, p. 1193-1197, 2006.

FOX, Malcolm F. **Chemistry and technology of lubricants**. Dordrecht: Springer, 2010.

AMERICAN PETROLEUM INSTITUTE (API). **Oil categories**. Disponível em: <[API | Categorias de óleo](#)>. Acesso em: 13 abr. 2023.

DE GOOIJER, Jan G.; HYNDMAN, Rob J. 25 years of time series forecasting. **International journal of forecasting**, v. 22, n. 3, p. 443-473, 2006.

ZHANG, G. Peter. Time series forecasting using a hybrid ARIMA and neural network model. **Neurocomputing**, v. 50, p. 159-175, 2003.

ESLING, Philippe; AGON, Carlos. Time-series data mining. **ACM Computing Surveys (CSUR)**, v. 45, n. 1, p. 1-34, 2012.

SHRIVASTRI, Shweta et al. A Comparative Study between (ARIMA-ETS) Models to Forecast Wheat Production and its Importance's in Nutritional Security. **J. Agric., Biol. Appl. Stat**, v. 1, n. 1, p. 25-37, 2022.

JAIN, Garima; MALLICK, Bhawna. A study of time series models ARIMA and ETS. **Available at SSRN 2898968**, 2017.

MOHEB-ALIZADEH, Hadi; HANDFIELD, Robert. The Impact of raw materials price volatility on cost of goods sold (COGS) for product manufacturing. **IEEE Transactions on Engineering Management**, v. 65, n. 3, p. 460-473, 2018.

LUDKE, Menga; ANDRÉ, Marli. Pesquisa em educação: abordagens qualitativas. **Em Aberto**, v. 5, n. 31, 1986.