



INSTITUTO FEDERAL DE CIÊNCIA E TECNOLOGIA DE PERNAMBUCO

Campus Recife

Departamento de Engenharia Mecânica

Bacharel em Engenharia Mecânica

Thiago José Freire Chalaça

**USO DA ANÁLISE DE DADOS E MACHINE LEARNING APLICADO AO
PROCESSO DE SEPARAÇÃO E EXPEDIÇÃO EM HOME CENTERS: ESTUDO DE
CASO**

Recife

2026

THIAGO JOSÉ FREIRE CHALAÇA

**USO DA ANÁLISE DE DADOS E MACHINE LEARNING APLICADO AO
PROCESSO DE SEPARAÇÃO E EXPEDIÇÃO EM HOME CENTERS: ESTUDO DE
CASO**

Trabalho de conclusão de curso apresentado ao Departamento de Graduação em Bacharel em Engenharia Mecânica do Instituto Federal de Ciência e Tecnologia de Pernambuco, como requisito para obtenção do título de Bacharel em Engenharia Mecânica.

Orientador: Prof. Dr. José Ângelo Peixoto da Costa.

Recife

2026

Ficha catalográfica elaborada pela bibliotecária Danielle Castro da Silva CRB4/1457

C436u Chalaça, Thiago José Freire

2026

Uso da análise de dados e machine learning aplicado ao processo de separação e expedição em home centers: estudo de caso. / Thiago José Freire Chalaça. --- Recife: O autor, 2026.

86f. il. Color.

Trabalho de Conclusão (Curso Superior Tecnológico em Engenharia Mecânica) – Instituto Federal de Pernambuco, Recife, 2026.

Inclui Referências.

Orientador: Prof. Dr. José Ângelo Peixoto da Costa.

1. Engenharia mecânica. 2. Logística. 3. Expedição. 4. Inteligência Artificial. I. Título. II. Costa, José Ângelo Peixoto da (orientador). III. Instituto Federal de Pernambuco.

CDD 620.1

AGRADECIMENTOS

Antes de tudo, honra e glória ao Senhor Deus, que sempre me dá discernimento, saúde e coragem, dia a dia em minha vida.

Meu muitíssimo obrigado ao meu pai Eduardo e minha mãe Rebeca, por fazerem tudo em minha vida, gratidão eterna.

Agradecer imensamente a minha futura esposa Vitória, por sempre me apoiar no meu crescimento pessoal e profissional.

Agradeço aos meus amigos em geral do IFPE, porém especificamente a turma 2029.1, por toda amizade e parceria durante a difícil jornada até aqui.

Agradecer a todos os professores, que de uma forma geral vocês geraram todo conhecimento que está na minha base profissional.

Um agradecimento especial ao meu orientador, coordenado e professor, José Ângelo Peixoto da Costa por todos os conselhos, ensinamentos, atenção e cuidado no desenvolvimento profissional.

E agradeço a todos que de forma geral, me auxiliaram durante todo meu trajeto profissional e acadêmico.

“Você nunca sabe que resultados virão
da sua ação. Mas se você não fizer
nada, não existirão”
(Mahatma Gandhi)

RESUMO

O processo de separação e atendimento de pedidos na área de expedição de Home Centers apresenta elevada variabilidade operacional, resultando em atrasos no cumprimento das metas internas e impactando diretamente a experiência do cliente. Fatores como peso das mercadorias, quantidade de itens, tipo de entrega e produtividade dos operadores contribuem para oscilações nos tempos de separação e atendimento, dificultando o controle eficiente do fluxo operacional. Diante desse cenário, torna-se necessário compreender quantitativamente os fatores que influenciam os atrasos e propor melhorias baseadas em dados.

Este Trabalho de Conclusão de Curso tem como objetivo analisar o processo de separação e atendimento na expedição, identificando gargalos operacionais e propondo soluções por meio de técnicas de estatística e Inteligência Artificial. Para isso, foi desenvolvida uma base de dados, estruturada de forma realista, contendo 3.000 registros simulados conforme as regras operacionais da empresa. Foram realizadas análises descritivas para identificação de padrões, variabilidade e taxa de cumprimento das metas, além da aplicação de modelos supervisionados de classificação, como Árvore de Decisão e Random Forest, para previsão de atrasos.

Os resultados indicaram que aproximadamente 40% dos pedidos ultrapassam o tempo-meta de atendimento. A análise de importância das variáveis evidenciou que peso total, produtividade do separador e quantidade de itens são os principais fatores associados aos atrasos. Conclui-se que a integração entre engenharia de processos, estatística e Inteligência Artificial é eficaz para apoiar a otimização de operações logísticas.

Palavra-chaves: logística; expedição; análise de dados; inteligência artificial; processos operacionais

ABSTRACT

The order picking and customer service process in the expedition area of home centers presents high operational variability, resulting in delays in meeting internal service targets and directly affecting customer experience. Factors such as total load weight, number of items, delivery type, and operator productivity contribute to fluctuations in separation and service times, making it difficult to efficiently control operational flow. In this context, it becomes necessary to quantitatively understand the factors that influence delays and propose data-driven improvements.

This Final Term Paper aims to analyze the separation and service process in the expedition area, identifying operational bottlenecks and proposing solutions through statistical analysis and Artificial Intelligence techniques. A realistically structured fictitious database containing 3,000 simulated records was developed according to the company's operational rules. Descriptive statistical analyses were conducted to identify patterns, variability, and compliance with internal service targets. Additionally, supervised classification models, such as Decision Tree and Random Forest, were applied to predict service delays.

The results showed that approximately 40% of orders exceeded the established service-time targets. Feature importance analysis indicated that total weight, operator productivity, and number of items are the main factors associated with delays. The study concludes that integrating process engineering, statistics, and Artificial Intelligence is an effective approach to support logistics optimization and improve operational decision-making.

Keywords: logistics; expedition; data analysis; artificial intelligence; operational processes

FIGURAS

Figura 1 - Separação	11
Figura 2 - Metodologia do Trabalho.....	23
Figura 3 - Preparação e Tratamento dos Dados	27

TABELAS

Tabela 1 - Indicadores Gerais de Desempenho	43
Tabela 2 - Comparação de Desempenho por Tamanho da Base	53
Tabela 3 - Desempenho dos Modelos Testados	56
Tabela 4 - Importância das Variáveis (Random Forest)	56

GRÁFICOS

Gráfico 1 - Distribuição dos Tipos de Entrega.....	33
Gráfico 2 - Distribuição dos Grupos de Classificação.....	34
Gráfico 3 - Distribuição do Peso Total	35
Gráfico 4 - Distribuição da Quantidade de Itens.....	36
Gráfico 5 - Tempo de Separação por Grupo	37
Gráfico 6 - Tempo de Atendimento por Grupo	38
Gráfico 7 - Tempo Médio Real vs Meta de Atendimento	39
Gráfico 8 - Peso Total x Atraso no Atendimento	40
Gráfico 9 - Quantidade de Itens x Atraso no Atendimento	41
Gráfico 10 - Percentual de Atraso - Grupo x Tipo de Entrega	42
Gráfico 11 - Matriz Confusão – Random Forest.....	51
Gráfico 12 - Importância das Variáveis	57

SUMÁRIO

1 INTRODUÇÃO	11
1.1 Problema de Pesquisa	12
1.2 Objetivo Geral	13
1.3 Objetivos Específicos	13
1.4 Justificativa	13
1.5 Metodologia Resumida	14
1.6 Estrutura do Trabalho	14
2 FUNDAMENTAÇÃO TEÓRICA	15
2.1 Logística e Processos de Expedição	15
2.2 Engenharia de Processos, Tempos e Movimentos	15
2.3 Indicadores de Desempenho Logístico	16
2.4 Análise de Dados Aplicada à Logística	17
2.5 Estatística Aplicada à Análise de Processos	17
2.6 Inteligência Artificial aplicada à previsão de eventos operacionais	18
2.7 Modelos de Machine Learning e Métricas de Avaliação	19
2.8 Síntese da Fundamentação Teórica	21
3 METODOLOGIA	23
3.1 Caracterização do Estudo	24
3.2 Fonte dos Dados	24
3.3 Regras Operacionais Consideradas	25
3.4 Geração dos Dados	26
3.5 Preparação e Tratamento dos Dados	26
3.6 Métodos de Análise	28
3.6.1 <i>Análise Descritiva</i>	28
3.6.2 <i>Modelos Preditivos</i>	28
3.6.3 <i>Proposição de Melhorias Operacionais</i>	28
4 ANÁLISE DOS RESULTADOS	29
4.1 Preparação e Tratamento dos Dados	29
4.1.1 <i>Geração da Base de Dados</i>	29
4.1.2 <i>Estrutura Final do Conjunto de Dados</i>	30
4.1.3 <i>Criação das Variáveis Derivadas</i>	31
4.1.4 <i>Carregamento da Base para Análise</i>	31

4.1.5 Considerações sobre a Preparação	32
4.2 Análise Descritiva dos Dados	32
4.2.1 Distribuição dos Tipos de Entrega	33
4.2.2 Distribuição dos Grupos de Classificação	33
4.2.3 Distribuição do Peso Total	34
4.2.4 Distribuição da Quantidade de Itens	35
4.2.5 Tempo de Separação por Grupo	36
4.2.6 Tempo de Atendimento por Grupo	37
4.2.7 Tempo Médio Real vs Meta de Atendimento	38
4.2.8 Peso Total x Atraso no Atendimento	39
4.2.9 Quantidade de Itens x Atraso no Atendimento	40
4.2.10 Percentual de Atraso - Grupo x Tipo de Entrega	41
4.2.11 Indicadores Gerais de Desempenho Frente às Metas	42
4.2.12 Síntese da Análise Descritiva	43
4.3 Identificação de Gargalos Operacionais	44
4.3.1 Desempenho Geral: Alta Taxa de Atrasos	44
4.3.2 Baixo desempenho no Grupo A	44
4.3.3 Impacto do Tipo de Entrega na Taxa de Atraso	45
4.3.4 Variabilidade Elevada nos Grupos C e D	46
4.3.5 Sobrecarga de Entregas Imediatas nos Turnos de Maior Demanda	47
4.3.6 Falta de Relação Clara com Peso e Número de Itens	47
4.3.7 Síntese dos Gargalos Identificados	48
4.3.8 Conclusão da Identificação de Gargalos	48
4.4 Modelos de Inteligência Artificial para Previsão de Atrasos no Atendimento	49
4.4.1 Avaliação dos Modelos	49
4.4.2 Análise do Classification Report	50
4.4.3 Matriz de Confusão	50
4.4.4 Importância das Variáveis	51
4.4.5 Interpretação Geral dos Resultados	52
4.5 Impacto do Volume de Dados na Performance do Modelo	53
4.6 Discussão dos Resultados	55
4.6.1 Comportamento Operacional	55
4.6.2 Conformidade com Metas Operacionais	55

4.6.3 Avaliação dos Modelos Preditivos	56
4.6.4 Importância das Variáveis	56
4.6.5 Síntese Analítica	58
5 PROPOSTA E METODOLOGIA	59
5.1 Revisão dos Gargalos Identificados	59
5.2 Propostas de Melhoria	60
5.3 Otimização da Alocação dos Separadores (Balanceamento de Carga de Trabalho)	60
5.3.1 Revisão dos Tempos-Meta Operacionais	61
5.3.2 Implantação de Indicadores Operacionais em Tempo Real (Dashboard Inteligente)	62
5.3.3 Treinamento e Nivelamento Técnico dos Separadores	62
5.4 Metodologia para Implementação das Propostas	63
5.4.1 Diagnóstico Operacional	63
5.4.2 Planejamento das Intervenções	63
5.4.3 Implementação Piloto	64
5.4.4 Avaliação dos Resultados	64
5.4.5 Padronização e Expansão	64
5.5 Considerações Finais do Capítulo	65
6 CONCLUSÃO	66
REFERÊNCIAS	68

1 INTRODUÇÃO

O setor de home centers no Brasil apresenta crescimento contínuo e movimenta bilhões de reais anualmente, consolidando-se como um dos principais segmentos do varejo de materiais de construção, acabamento e utilidades domésticas. Essas operações caracterizam-se por alto volume de pedidos, grande variedade de produtos e significativa diversidade de pesos e dimensões, o que impõe desafios logísticos relevantes. Nesse contexto, a eficiência operacional torna-se fator crítico de competitividade, impactando diretamente custos, produtividade e nível de serviço ao cliente.

Figura 1 - Separação



Fonte: Logística & Mobilidade, 2024.

Dentro da estrutura operacional de um home center, a área de separação e expedição desempenha papel estratégico, pois representa a etapa final do processo de atendimento e influencia diretamente indicadores como tempo de espera, satisfação do cliente e percepção de qualidade do serviço. Pequenas ineficiências nessa etapa podem gerar filas, retrabalho, aumento de custos operacionais e perda de credibilidade junto ao consumidor. Além disso, o tempo de atendimento está diretamente associado à utilização da mão de obra e dos recursos internos, impactando o custo por pedido atendido.

No processo de expedição, dois tempos são monitorados continuamente: tempo de separação e tempo de atendimento. Esses indicadores são regidos por metas internas definidas conforme o tipo de entrega e a complexidade do pedido, classificada por peso e quantidade de itens. Contudo, observa-se que parte significativa dos atendimentos ultrapassa o tempo-meta estabelecido, evidenciando a existência de gargalos operacionais e variabilidade no fluxo de trabalho.

Diante desse cenário, torna-se essencial compreender quantitativamente os fatores que influenciam os atrasos e desenvolver ferramentas capazes de apoiar decisões gerenciais baseadas em dados. A aplicação de técnicas de estatística e Inteligência Artificial surge como alternativa para diagnosticar o processo, prever atrasos e propor melhorias estruturadas.

Assim, este trabalho propõe uma análise aprofundada do processo de separação e atendimento na expedição de home centers, utilizando uma base de dados estruturada construída de forma realista. Por meio de análise descritiva, identificação de gargalos e desenvolvimento de modelos preditivos, busca-se contribuir para a otimização do processo logístico e para o fortalecimento da gestão baseada em evidências.

1.1 Problema de Pesquisa

O processo de expedição apresenta significativa variabilidade nos tempos de separação e atendimento, resultando em atrasos frequentes e inconsistência no cumprimento das metas internas. Essa variabilidade pode estar associada a fatores como peso da mercadoria, quantidade de itens, produtividade individual dos separadores, tipo de entrega e demanda diária. No entanto, não existe atualmente um diagnóstico claro e quantitativo que permita identificar os principais fatores responsáveis pelos atrasos.

Assim, o problema central de pesquisa é: Quais fatores influenciam o atraso no atendimento da expedição e como a análise de dados e técnicas de Inteligência Artificial podem ser utilizadas para prever esses atrasos e propor melhorias para o processo?

1.2 Objetivo Geral

Analisar o processo de separação e atendimento da expedição de um Home Center, identificando os fatores que influenciam atrasos e propondo melhorias operacionais com base em análises estatísticas e modelos de Inteligência Artificial.

1.3 Objetivos Específicos

Para atingir o objetivo geral, estabelecem-se os seguintes objetivos específicos:

- Desenvolver uma base de dados estruturada representativa do processo real de expedição.
- Realizar análise descritiva das variáveis operacionais e avaliar o desempenho em relação às metas internas.
- Identificar gargalos e fatores críticos que impactam o tempo total de atendimento.
- Implementar modelos de Inteligência Artificial capazes de prever atrasos na expedição.
- Avaliar a acurácia e desempenho dos modelos preditivos.
- Propor melhorias práticas e metodológicas baseadas nos achados analíticos.

1.4 Justificativa

A justificativa deste estudo baseia-se em três pilares principais:

- Relevância Operacional: O processo de expedição afeta diretamente o cliente final. Atrasos na separação ou atendimento aumentam o tempo de espera e comprometem a satisfação do cliente, podendo gerar retrabalho, acúmulo de fila e desgaste operacional.
- Apoio à Tomada de Decisão: A utilização de dados para compreender a operação oferece base objetiva para decisões gerenciais, tornando o processo mais eficiente e eliminando percepções intuitivas.
- Aplicação de Inteligência Artificial: O uso de IA no ambiente de expedição é uma inovação promissora. Mesmo com desempenho moderado, os modelos fornecem insights valiosos sobre variáveis críticas, abrindo caminho para futuras automações e melhorias contínuas.

1.5 Metodologia Resumida

O estudo contemplou as seguintes etapas:

- Construção de uma base fictícia de 3.000 registros, simulando pedidos reais da expedição.
- Tratamento, preparação e criação de variáveis derivadas para análise.
- Análise descritiva dos tempos, volumes e características operacionais.
- Identificação dos principais gargalos e variações do processo.
- Desenvolvimento e avaliação de modelos de classificação (Árvore de Decisão e Random Forest).
- Proposição de melhorias operacionais baseadas nos resultados.

1.6 Estrutura do Trabalho

Este Trabalho de Conclusão de Curso está organizado da seguinte forma:

- Capítulo 1: apresenta a introdução, o problema de pesquisa, objetivos e justificativa.
- Capítulo 2: aborda a fundamentação teórica sobre logística, expedição, estatística e inteligência artificial.
- Capítulo 3: descreve a metodologia utilizada no estudo.
- Capítulo 4: apresenta a análise de dados, modelos de IA e discussão dos resultados.
- Capítulo 5: propõe melhorias operacionais e a metodologia para implementação.
- Capítulo 6: apresenta a conclusão geral do trabalho.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os fundamentos conceituais que sustentam o desenvolvimento do presente estudo. São abordados conceitos relacionados à logística e expedição, engenharia de processos, análise de dados, indicadores de desempenho e técnicas de Inteligência Artificial aplicadas à previsão de eventos operacionais.

2.1 Logística e Processos de Expedição

A logística é definida como o conjunto de atividades responsáveis pelo planejamento, implementação e controle do fluxo eficiente de produtos, serviços e informações ao longo da cadeia de suprimentos (BALLOU, 2006). Dentro dessa cadeia, o processo de expedição assume papel estratégico, especialmente em operações de varejo, por envolver a etapa final de atendimento ao cliente e ser determinante para a percepção de qualidade no serviço prestado.

Segundo Novaes (2007), a expedição é o elo final da logística interna responsável por preparar, conferir, movimentar e entregar produtos ao cliente. O desempenho dessa etapa é diretamente relacionado a indicadores como tempo de resposta, organização do layout, capacidade operacional e dinâmica de demanda.

No ambiente de home centers, a expedição apresenta desafios específicos: diversidade de produtos, variação de pesos e volumes, existência de itens frágeis ou de difícil movimentação, além da oscilação na demanda ao longo do dia. Conforme destaca Bowersox, Closs e Cooper (2014), operações desse tipo demandam alto nível de sincronização entre equipes, equipamentos e processos para evitar gargalos.

Dessa forma, compreender o funcionamento e a variabilidade da expedição é essencial para identificar possíveis oportunidades de melhoria e elevar o nível de serviço oferecido.

2.2 Engenharia de Processos, Tempos e Movimentos

A engenharia de processos busca analisar, modelar e melhorar atividades produtivas com base em princípios de eficiência, padronização e redução de variabilidade (SLACK; CHAMBERS; JOHNSTON, 2009). No contexto de operações internas, a análise de tempos

e movimentos auxilia no dimensionamento da mão de obra e na definição de metas realistas de produtividade.

O estudo de tempos tem como objetivo determinar o tempo necessário para a execução de uma tarefa em condições normais de trabalho (MARTINS; LAUGENI, 2005). A partir dessa análise, estabelecem-se tempos/padrões utilizados para planejar, controlar e otimizar processos.

De acordo com Peinado e Graeml (2007), tempo/padrão precisam considerar não apenas a execução da tarefa, mas fatores como:

- complexidade do item,
- ergonomia,
- deslocamentos,
- condições operacionais,
- experiência do operador.

Esses princípios são fundamentais no processo de expedição de um Home Center, onde o tempo de separação e atendimento depende diretamente do peso, quantidade de itens, tipo de produto e habilidades individuais do separador.

2.3 Indicadores de Desempenho Logístico

Os indicadores logísticos têm como objetivo medir o desempenho operacional e fornecer subsídios para tomada de decisão. Segundo Ching (2010), indicadores como lead time, tempo de atendimento, nível de serviço e taxa de cumprimento de prazos são essenciais para avaliar a eficiência do processo.

O indicador utilizado neste estudo, cumprimento da meta de atendimento, enquadra-se na categoria de indicadores de tempo. Conforme afirma Viana (2012), métricas baseadas em tempo são críticas em operações de varejo, pois refletem diretamente a experiência do cliente e a confiabilidade do serviço prestado.

Além disso, a literatura destaca que indicadores devem ser analisados não apenas isoladamente, mas considerando causas, variáveis associadas e tendências históricas (WANKE, 2012). Esse princípio orientou a análise estatística realizada no Capítulo 4.

2.4 Análise de Dados Aplicada à Logística

Com o avanço das tecnologias digitais, a logística tem se beneficiado cada vez mais da análise de dados. Inúmeros autores apontam que o uso de dados permite identificar padrões, otimizar fluxos e reduzir variabilidade (GIL, 2019; CORRÊA; CORRÊA, 2012).

Segundo Davenport (2012), a capacidade de transformar dados operacionais em conhecimento é um diferencial competitivo para organizações que buscam eficiência e adaptabilidade. Em operações com grande heterogeneidade, como expedição, a análise de dados permite:

- monitorar tempos médios,
- identificar gargalos,
- entender comportamento da demanda,
- suportar decisões de priorização.

Esse embasamento fundamenta a adoção das técnicas analíticas aplicadas neste trabalho, como análise descritiva, boxplots (Diagrama de Caixas), histogramas e correlações.

2.5 Estatística Aplicada à Análise de Processos

A estatística desempenha papel fundamental na análise de dados e no entendimento do comportamento de processos operacionais. Segundo Bussab e Morettin (2017), a estatística fornece ferramentas para descrever, organizar, interpretar e analisar fenômenos observados, permitindo transformar informações brutas em conhecimento estruturado.

Na área de Engenharia e Logística, a estatística é amplamente utilizada para:

- identificar padrões e variabilidade dos processos,
- analisar tempos de operação,
- calcular indicadores de desempenho,
- avaliar aderência a metas,
- embasar decisões gerenciais baseadas em evidências.

De acordo com Montgomery e Runger (2010), a análise estatística de processos permite identificar desvios, causas de variabilidade e oportunidades de melhoria contínua. Isso é especialmente relevante no processo logístico analisado neste estudo, onde o tempo

de separação e o tempo de atendimento apresentam flutuações decorrentes de múltiplos fatores, como peso, quantidade de itens e produtividade dos operadores.

A estatística descritiva, conforme explica Triola (2018), inclui medidas como média, mediana, desvio padrão, amplitude, percentis e distribuição de frequências, que auxiliam na compreensão da tendência central e da variabilidade dos dados. Essas medidas foram essenciais para a compreensão dos tempos reais de separação e atendimento, permitindo comparar o comportamento observado com as metas estabelecidas pela empresa.

Ferramentas gráficas, como histogramas, boxplots e diagramas de dispersão, também fazem parte da estatística descritiva e desempenham papel importante na identificação de padrões e anomalias nos dados. Histogramas ajudam a visualizar a distribuição de variáveis como peso e quantidade de itens; boxplots evidenciam dispersões e outliers nos tempos operacionais, e gráficos de dispersão permitem investigar relações entre variáveis contínuas.

Além da estatística descritiva, conceitos de probabilidade sustentam análises de risco e previsão. Marinho (2014) afirma que a probabilidade é fundamental em modelos que lidam com incertezas, como a previsão de atrasos operacionais, pois permite calcular a chance de um evento ocorrer.

Essa abordagem é diretamente aplicada nos modelos de Inteligência Artificial utilizados no presente estudo, que estimam a probabilidade de um pedido ultrapassar o tempo/meta de atendimento, classificando-o como atraso ou não.

A estatística é também essencial na avaliação dos modelos preditivos. Métricas como acurácia, precisão, recall e F1-score, conforme discutido por Han, Kamber e Pei (2012), são baseadas em princípios estatísticos e permitem avaliar o desempenho dos algoritmos de classificação utilizados neste estudo.

Portanto, a estatística fornece o arcabouço conceitual necessário para análise, modelagem, interpretação e validação dos resultados apresentados nesta pesquisa. Sem ela, não seria possível identificar gargalos, comparar tempos reais com metas, avaliar o comportamento do processo ou validar a eficácia dos modelos preditivos desenvolvidos.

2.6 Inteligência Artificial aplicada à previsão de eventos operacionais

A Inteligência Artificial (IA) vem sendo amplamente utilizada para previsão, classificação e otimização de processos produtivos e logísticos. Para Mitchell (2010), IA consiste em sistemas capazes de aprender padrões a partir de dados e realizar previsões com base em experiência prévia.

Modelos supervisionados, como Árvores de Decisão e Random Forest, são particularmente úteis em cenários operacionais por possuírem:

- boa interpretabilidade,
- capacidade de lidar com variáveis categóricas e numéricas,
- alto desempenho em ambientes com variabilidade,
- resistência a outliers.

Segundo Breiman (2001), o Random Forest é um dos modelos mais eficazes para classificação por combinar múltiplas árvores de decisão e reduzir o risco de sobreajuste.

No contexto logístico, aplicações de IA podem prever atrasos, identificar fatores críticos e apoiar decisões sobre alocação de recursos (SANTOS; SILVA; PEREIRA, 2020). Assim, sua aplicação na expedição de um Home Center se justifica pela necessidade de antecipar riscos operacionais e direcionar a gestão para ações preventivas.

2.7 Modelos de Machine Learning e Métricas de Avaliação

Machine Learning (Aprendizado de Máquina) é uma subárea da Inteligência Artificial que estuda algoritmos capazes de aprender padrões a partir de dados e realizar previsões ou classificações sem serem explicitamente programados para cada situação específica (MITCHELL, 2010). Diferentemente de modelos determinísticos tradicionais, o aprendizado de máquina baseia-se na identificação de regularidades estatísticas nos dados históricos para generalizar comportamentos futuros.

Os modelos utilizados neste estudo pertencem à categoria de aprendizado supervisionado, no qual o algoritmo é treinado com um conjunto de dados rotulados, contendo variáveis independentes (features) e uma variável alvo (target). No caso desta pesquisa, a variável alvo corresponde à ocorrência ou não de atraso no atendimento.

Entre os diversos algoritmos existentes, optou-se pela utilização de Árvore de Decisão (Decision Tree) e Random Forest, por apresentarem características adequadas ao contexto operacional analisado.

A Árvore de Decisão é um modelo baseado em regras hierárquicas de decisão, no qual os dados são divididos progressivamente com base em critérios de

pureza, como o índice de Gini ou entropia (HAN; KAMBER; PEI, 2012). Sua principal vantagem está na interpretabilidade, permitindo compreender quais variáveis mais influenciam o resultado.

Já o Random Forest, proposto por Breiman (2001), é um modelo de ensemble que combina múltiplas árvores de decisão treinadas em subconjuntos aleatórios dos dados e das variáveis. Essa abordagem reduz o risco de sobre ajuste e melhora a capacidade de generalização do modelo. Sua robustez e desempenho consistente justificam sua ampla aplicação em problemas de classificação com alta variabilidade, como os processos logísticos analisados neste estudo.

A escolha desses modelos justifica-se por apresentarem características técnicas adequadas ao contexto operacional analisado. Primeiramente, tanto a Árvore de Decisão quanto o Random Forest possuem capacidade de lidar simultaneamente com variáveis categóricas e numéricas, o que é essencial em bases logísticas que combinam informações como tipo de entrega, turno, grupo de classificação, peso e quantidade de itens. Além disso, esses algoritmos tendem a apresentar boa performance em bases com elevada variabilidade operacional, como é o caso de processos de expedição, nos quais múltiplos fatores influenciam os tempos de atendimento.

Outro ponto relevante é a interpretabilidade dos resultados, especialmente no caso das Árvores de Decisão, que permitem visualizar as regras que conduzem às classificações. O Random Forest, por sua vez, possibilita a análise da importância das variáveis, contribuindo para a compreensão dos fatores mais impactantes no processo. Soma-se a isso a facilidade de implementação e replicabilidade dos modelos, amplamente disponíveis em bibliotecas consolidadas, o que favorece sua aplicação prática em ambientes corporativos. Por fim, ambos os modelos apresentam relativa robustez a outliers e conseguem lidar com relações não lineares entre variáveis, característica comum em cenários operacionais complexos.

Além da escolha dos modelos, a avaliação de desempenho é etapa fundamental na aplicação de Machine Learning. Conforme afirmam James et al. (2013), um modelo não deve ser avaliado apenas pela acurácia, mas por um conjunto de métricas capazes de capturar diferentes aspectos da performance.

Neste estudo, foram utilizadas as seguintes métricas:

Acurácia (Accuracy): representa a proporção total de previsões corretas em relação ao total de observações. Embora seja uma métrica intuitiva, pode mascarar problemas em bases desbalanceadas.

Precisão (Precision): indica a proporção de previsões positivas corretas dentre todas as previsões positivas realizadas pelo modelo. No contexto deste estudo, representa a confiabilidade das previsões de atraso.

Recall (Sensibilidade): mede a capacidade do modelo de identificar corretamente os casos positivos reais. Neste caso, corresponde à proporção de atrasos corretamente detectados.

F1-Score: é a média harmônica entre precisão e recall, sendo uma métrica equilibrada para cenários onde há preocupação tanto com falsos positivos quanto com falsos negativos.

Matriz de Confusão: ferramenta que organiza as previsões em quatro categorias: verdadeiro positivo, verdadeiro negativo, falso positivo e falso negativo, permitindo análise detalhada dos erros cometidos pelo modelo (HAN; KAMBER; PEI, 2012).

A utilização dessas métricas permite avaliar não apenas o desempenho global do modelo, mas também sua capacidade de identificar corretamente atrasos operacionais, que representam eventos críticos para a gestão da expedição.

Assim, a fundamentação em Machine Learning e avaliação estatística sustenta metodologicamente o desenvolvimento dos modelos preditivos apresentados no Capítulo 4, garantindo rigor técnico e coerência científica à pesquisa.

2.8 Síntese da Fundamentação Teórica

A literatura evidenciou a importância de:

- compreender os processos logísticos como sistemas interdependentes,
- utilizar indicadores e análises quantitativas para otimização,
- aplicar técnicas de engenharia de processos para melhoria contínua,

- explorar o potencial da Inteligência Artificial para previsão de eventos logísticos.

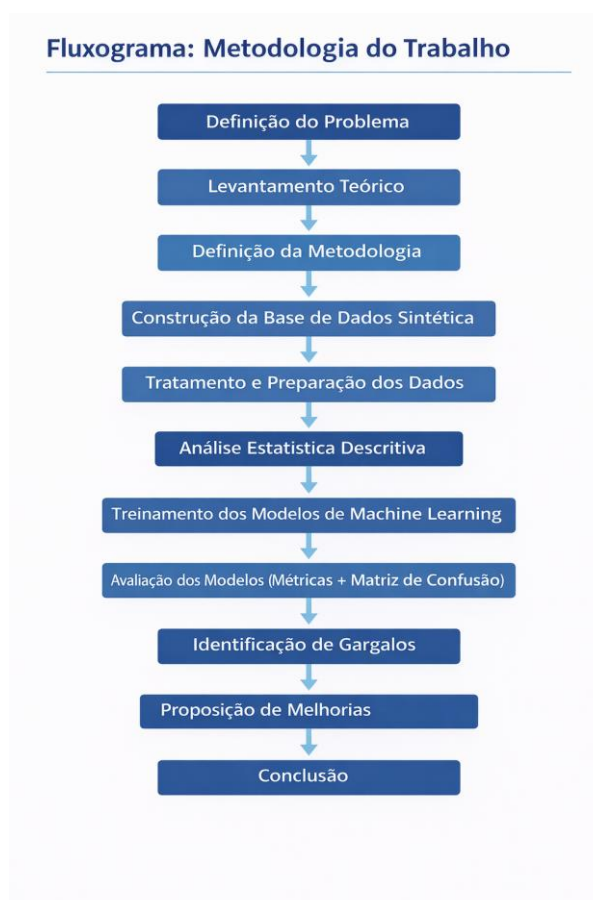
Esses princípios sustentam metodologicamente todas as etapas desenvolvidas no presente estudo e fundamentam as conclusões e propostas de melhoria apresentadas nos capítulos seguintes.

3 METODOLOGIA

Neste capítulo, será apresentada a metodologia utilizada para o desenvolvimento deste estudo de caso, descrevendo os métodos, técnicas e etapas adotadas para analisar e propor melhorias no processo de separação e expedição de mercadorias. Considerando que se trata de um ambiente operacional que envolve fluxo contínuo de atendimento ao cliente, tempos/metras e diferentes classificações de entrega, optou-se por uma abordagem estruturada e orientada à análise de dados.

A pesquisa envolve a construção de uma base de dados estruturada representativa da operação real, aplicação de técnicas estatísticas e uso de modelos de Inteligência Artificial para identificação de padrões e previsão de possíveis atrasos no atendimento. A metodologia foi definida de forma a garantir precisão analítica, reprodutibilidade e alinhamento com o contexto prático da empresa.

Figura 2 - Metodologia do Trabalho



Fonte: Autor, 2025.

A seguir, serão detalhadas as etapas realizadas para a execução e conclusão do estudo.

3.1 Caracterização do Estudo

Este trabalho caracteriza-se como uma pesquisa aplicada, com abordagem quantitativa e finalidade exploratória e analítica. O estudo tem como foco o processo de separação e expedição de mercadorias de um centro de distribuição do setor varejista, contemplando as etapas de emissão de cupom de retirada, separação, conferência e entrega ao cliente.

O objetivo central da metodologia é estruturar um modelo analítico capaz de identificar padrões operacionais, prever riscos de atraso e fornecer subsídios técnicos para a otimização do processo. Embora o estudo esteja baseado na dinâmica operacional de um Home Center, os dados utilizados não são reais, preservando assim a confidencialidade e integridade das informações.

3.2 Fonte dos Dados

Em razão da impossibilidade de utilizar dados operacionais reais, foi desenvolvida uma base de dados estruturada, construída para representar com fidelidade o comportamento dos registros de separação e expedição observados no ambiente real.

A base contém informações sobre:

- ID do pedido
- Nome do cliente
- Data do pedido
- Tipo de entrega: Imediata, Posterior ou Agendada
- Grupo operacional: A, B, C ou D
- Peso total da mercadoria
- Quantidade de itens
- Tempo de separação
- Tempo total de atendimento

Essas variáveis foram selecionadas por representarem os principais fatores que influenciam o desempenho do processo.

A base foi construída com 3.000 registros, quantidade suficiente para análises estatísticas e de machine learning.

3.3 Regras Operacionais Consideradas

A geração da base e a definição das variáveis seguiram integralmente as regras internas utilizadas no processo real de separação e expedição. Essas regras determinam os tempos/meta conforme peso total, número de itens e tipo de entrega.

Os grupos operacionais utilizados foram:

Grupo A

- Peso: até 200 kg
- Itens: até 3 itens, sem produtos da Linha C
- Tempo/meta de separação: 5 min
Tempo/meta de atendimento: 6 min

Grupo B

- Peso: 201–500 kg
- Itens: mais de 3 itens ou presença de produtos Linha C
- Tempo/meta de separação: 7 min
- Tempo/meta de atendimento: 9 min

Grupo C

- Peso: 501–2000 kg
- Itens: notas com mais de 3 itens ou com Pisos
- Tempo/meta de separação: 15 min
- Tempo/meta de atendimento: 20 min

Grupo D

- Peso: acima de 2000 kg
- Itens: notas volumosas e complexas
- Tempo/meta de separação: 22 min
- Tempo/meta de atendimento: 30 min

Entrega Agendada

- Independente do peso:
- Tempo/meta de separação: 5 min
- Tempo/meta de atendimento: 6 min

Foi ainda considerado que o tempo de atendimento engloba o tempo de separação, conforme política operacional.

3.4 Geração dos Dados

A base de dados foi gerada utilizando a linguagem Python, por meio do ambiente VS Code, com apoio das bibliotecas Pandas e Faker para criação de dados sintéticos.

O algoritmo criado segue as etapas:

1. Definição dos grupos operacionais com suas respectivas regras de peso, itens e tempos/meta.
2. Sorteio aleatório de tipo de entrega e grupo para cada pedido.
3. Geração de peso total e quantidade de itens dentro dos limites de cada grupo.
4. Simulação dos tempos reais de separação e atendimento com base em distribuições normais (gaussianas) ao redor do tempo/meta, introduzindo pequenas variações aleatórias para refletir eventos reais.
5. Criação automática da tabela final com todos os registros.

O código utilizado para a geração da base é apresentado no Apêndice A

Essa abordagem garante:

- Representatividade da operação real,
- confidencialidade das informações da empresa,
- replicabilidade do experimento.

3.5 Preparação e Tratamento dos Dados

Após a geração da base de dados estruturada, foi realizada uma etapa estruturada de preparação e tratamento dos dados, com o objetivo de garantir

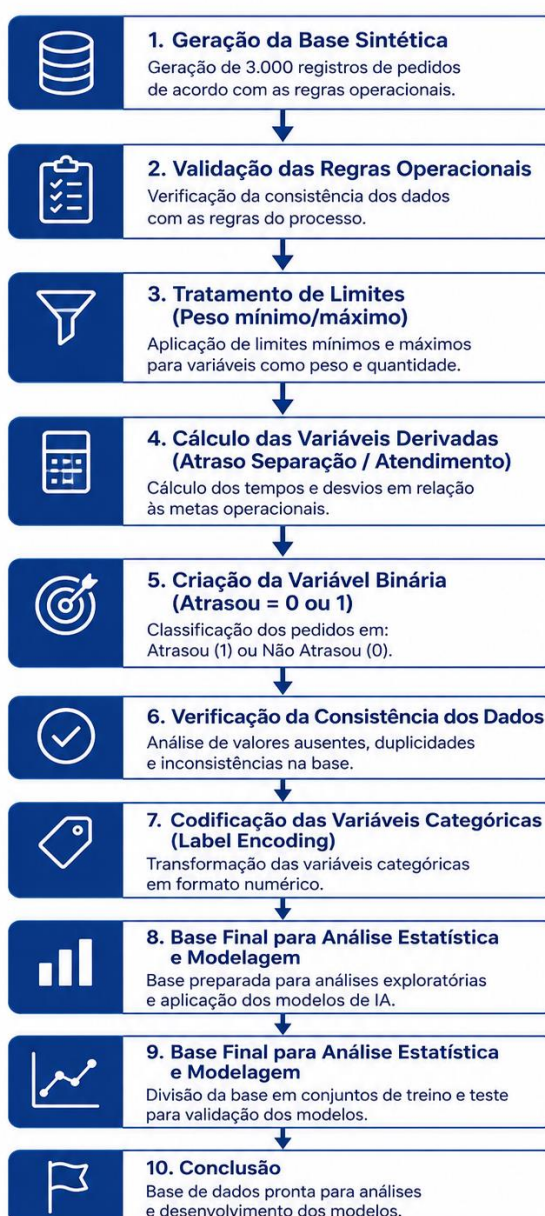
consistência, integridade e adequação para as análises estatísticas e modelagem preditiva.

Mesmo sendo uma base simulada, foram aplicadas técnicas de validação e transformação semelhantes às utilizadas em bases reais, assegurando reprodutibilidade metodológica.

A Figura 3 apresenta o fluxograma das etapas de preparação e limpeza dos dados adotadas neste estudo.

Figura 3 - Preparação e Tratamento dos Dados

Fluxograma: Preparação e Tratamento dos Dados



3.6 Métodos de Análise

Os dados foram analisados utilizando técnicas estatísticas e modelos de Inteligência Artificial, divididos em três etapas principais.

3.6.1 Análise Descritiva

- Distribuição dos tempos de separação e atendimento
- Distribuição por tipo de entrega e grupo operacional
- Identificação de outliers e padrões de comportamento
- Cálculo de atraso (tempo real – tempo-meta).

3.6.2 Modelos Preditivos

Para prever o risco de atraso na expedição, serão testados modelos como:

- Árvore de Decisão
- Random Forest
- Classificação Binária (Atrasou / Não atrasou)

Os modelos serão avaliados por métricas como:

- Acurácia
- Recall
- Precisão
- F1-score

3.6.3 Proposição de Melhorias Operacionais

Com base nos fatores identificados como mais relevantes para atraso, serão sugeridas melhorias como:

- redimensionamento de equipes,
- mudanças no fluxo de separação,
- ajustes no volume ideal por grupo,
- modelagem de cenários simulados

4 ANÁLISE DOS RESULTADOS

Este capítulo apresenta a análise dos resultados obtidos a partir da base de dados construída para este estudo. O objetivo é compreender o comportamento operacional do processo de separação e expedição de mercadorias, identificar padrões relevantes e avaliar a aderência da operação às metas internas estabelecidas pela empresa.

Os resultados foram organizados em quatro etapas principais:

- (i) Preparação e estruturação de dados (Seção 4.1),
- (ii) Análise descritiva detalhada (Seção 4.2),
- (iii) Identificação de gargalos operacionais (Seção 4.3) e
- (iv) Aplicação de modelos de Inteligência Artificial para previsão de atrasos (Seção 4.4).

A sequência de análises permite compreender progressivamente o processo, desde a caracterização inicial dos dados até a modelagem preditiva, fundamentando as propostas de melhoria apresentadas no Capítulo 5.

4.1 Preparação e Tratamento dos Dados

Para realização das análises previstas neste estudo, foi necessário construir uma base de dados estruturada que representasse de forma realista o processo de separação e expedição de mercadorias. A geração dessa base ocorreu em linguagem Python, respeitando as regras operacionais e metas internas utilizadas pela empresa.

A seguir, são apresentadas as principais etapas da preparação dos dados.

4.1.1 Geração da Base de Dados

A base foi construída a partir de um script em Python (Apêndice A), elaborado especificamente para simular 3.000 registros de atendimentos. Foram considerados:

- Três tipos de entrega: Imediata, Posterior e Agendada;

- Grupos operacionais (A, B, C, D e Agendada), definidos conforme peso, quantidade de itens e presença de pisos;
- Metas internas de separação e atendimento, específicas para cada grupo;
- Variáveis operacionais relevantes, como turno, separador responsável e dia da semana;
- Tempos de processo, simulados com base em distribuições normais centradas nas metas internas, de forma a garantir realismo.

Todos os atributos contidos na base refletem situações presentes na operação real, mas sem expor dados sensíveis da empresa, respeitando a confidencialidade das informações internas.

4.1.2 Estrutura Final do Conjunto de Dados

Após a simulação e organização, a base passou a conter as seguintes variáveis:

- Pedido_ID: identificador único do pedido
- Data_Pedido: data e hora de emissão do cupom
- Tipo_Entrega: Imediata, Posterior ou Agendada
- Qtde_Itens: número de itens do pedido
- Peso_Total_kg: peso consolidado da nota
- Contem_Pisos: sim/não
- Grupo_Classificacao: A, B, C, D ou Agendada
- Tempo_Separacao_min
- Tempo_Conferencia_min
- Tempo_Atendimento_min
- Meta_Separacao_min
- Meta_Atendimento_min
- Atendeu_Meta_Separacao: Sim/Não
- Atendeu_Meta_Atendimento: Sim/Não

- Separador_ID
- Turno: Manhã, Tarde ou Noite
- Dia_Semana: Segunda-feira a Domingo

Essa estrutura permite análises robustas, tanto descritivas quanto preditivas, contemplando todas as etapas operacionais do processo.

4.1.3 Criação das Variáveis Derivadas

Para ampliar a capacidade analítica, foram criados indicadores adicionais:

- **Atraso_Separacao**

$$\text{Atraso Separação} = \text{Tempo Separação} - \text{Meta de Separação}$$

- **Atraso_Atendimento**

$$\text{Atraso Atendimento} = \text{Tempo Atendimento} - \text{Meta de Atendimento}$$

Para ambos os casos:

- Valores > 0 representam atraso;
- Valores = 0 indicam atendimento exato dentro da meta de tempo do processo;
- Valores < 0 representam atendimento adiantado.

Essas variáveis foram fundamentais para a construção de gráficos, tabelas e posteriormente para o treinamento dos modelos de IA.

4.1.4 Carregamento da Base para Análise

Após geração e tratamento, a base foi carregada em ambiente Python com o seguinte comando:

```
import pandas as pd
df = pd.read_csv("base_ficticia.csv")
```

Em seguida, foram realizadas verificações iniciais:

- Checagem de tipos de dados (dtypes)

- Busca por valores nulos ou inconsistentes
- Validação da soma de tempos (Separação + Conferência = Atendimento)
- Análise de proporções e distribuição geral das variáveis

Nenhuma inconsistência relevante foi encontrada, permitindo seguir para a análise descritiva detalhada apresentada na seção 4.2.

4.1.5 Considerações sobre a Preparação

A preparação adequada dos dados foi essencial para garantir:

- Realismo na simulação
- Consistência nas regras de negócio
- Possibilidade de análises avançadas
- Segurança e anonimização total das informações

Com a base devidamente estruturada, foi possível avançar para a etapa seguinte, a análise descritiva completa, apresentada no tópico 4.2.

4.2 Análise Descritiva dos Dados

Nesta seção são apresentadas as análises descritivas iniciais realizadas sobre a base de dados gerada para este estudo, contendo 3.000 registros referentes ao processo de separação e expedição. O objetivo é compreender o comportamento geral das variáveis operacionais, caracterizar o perfil dos pedidos e identificar tendências que auxiliem a posterior etapa de identificação de gargalos e de modelagem preditiva.

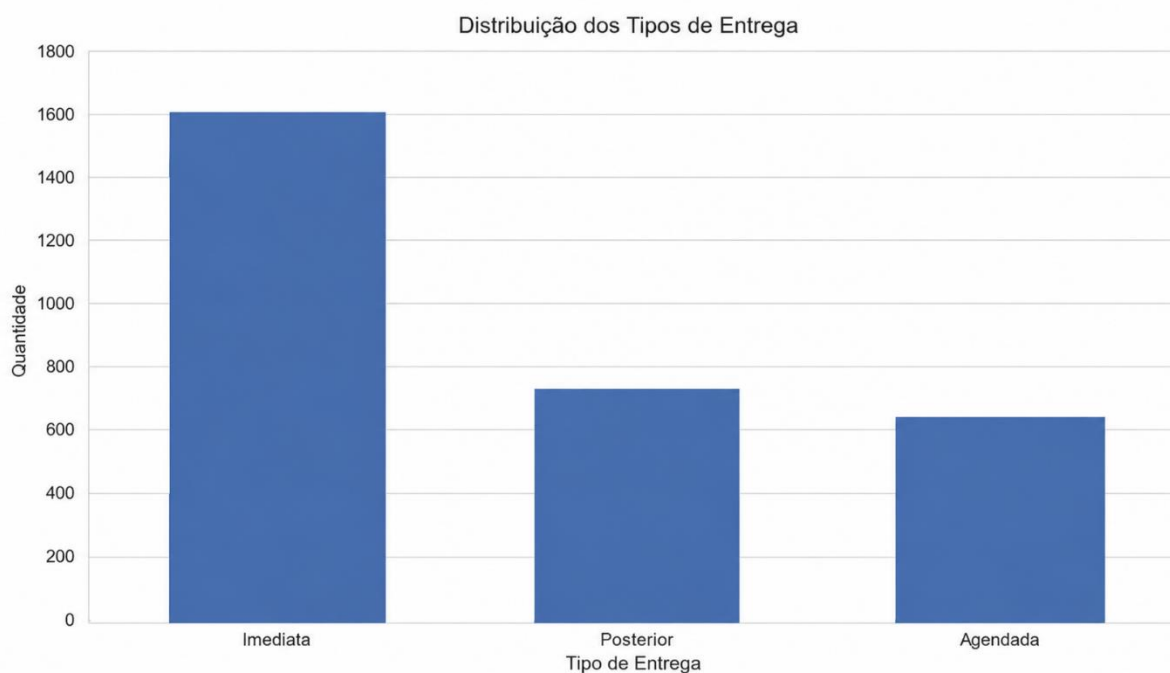
A análise foi estruturada considerando quatro dimensões principais:

- características dos pedidos,
- tempos operacionais,
- desempenho frente às metas internas,
- relações entre variáveis e atraso no atendimento.

4.2.1 Distribuição dos Tipos de Entrega

A Gráfico 1 apresenta a distribuição dos tipos de entrega. Observa-se forte predominância do tipo Imediata, que representa mais da metade dos atendimentos na amostra. Na sequência aparecem os tipos Posterior e Agendada, ambos com volumes significativos.

Gráfico 1 - Distribuição dos Tipos de Entrega



Fonte: Autor, 2025.

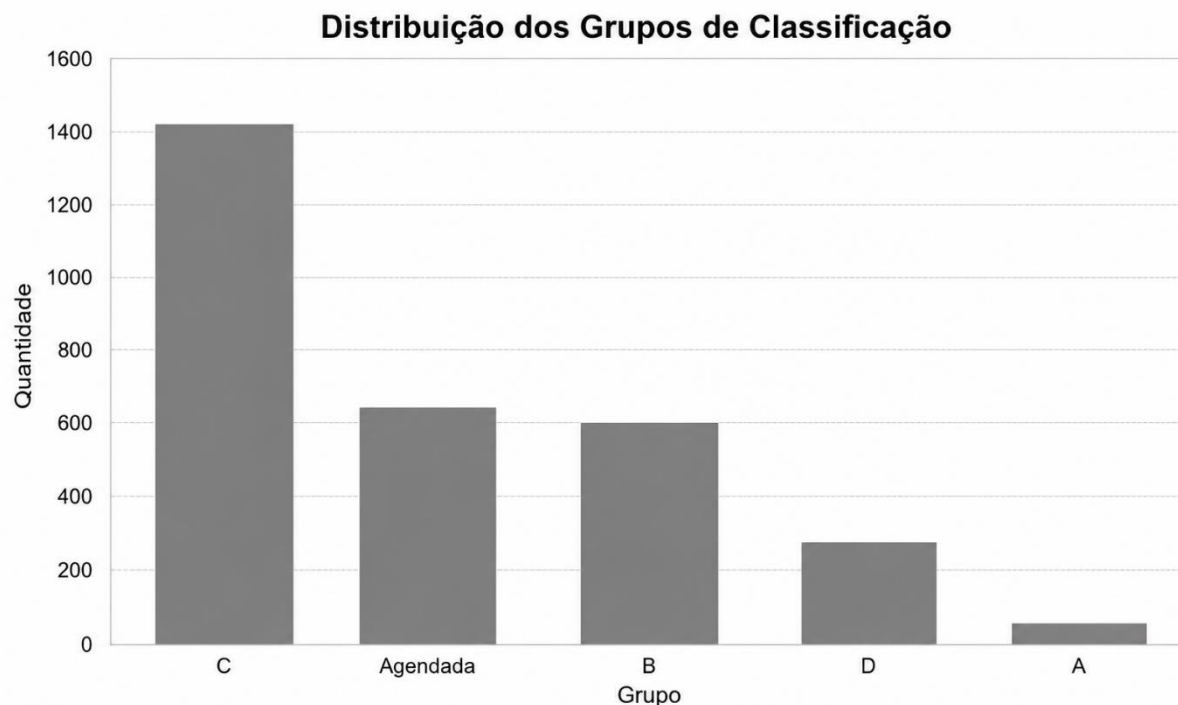
Essa predominância da entrega imediata é coerente com o padrão operacional de depósitos de varejo, em que a maior parte dos clientes realiza retiradas logo após a compra. A elevada demanda por atendimentos imediatos tende a sobrecarregar a equipe, especialmente em horários de pico, contribuindo para maior variabilidade no tempo total de atendimento. Esse padrão será avaliado mais à frente ao analisar a taxa de cumprimento de metas por tipo de entrega.

4.2.2 Distribuição dos Grupos de Classificação

A Gráfico 2 demonstra a proporção de atendimentos por grupo operacional (A, B, C, D e Agendada). O Grupo C aparece como o mais frequente, representando

grande parte dos atendimentos. Na sequência aparecem os grupos **Agendada** e **B**, enquanto os grupos D e A são minoria.

Gráfico 2 - Distribuição dos Grupos de Classificação

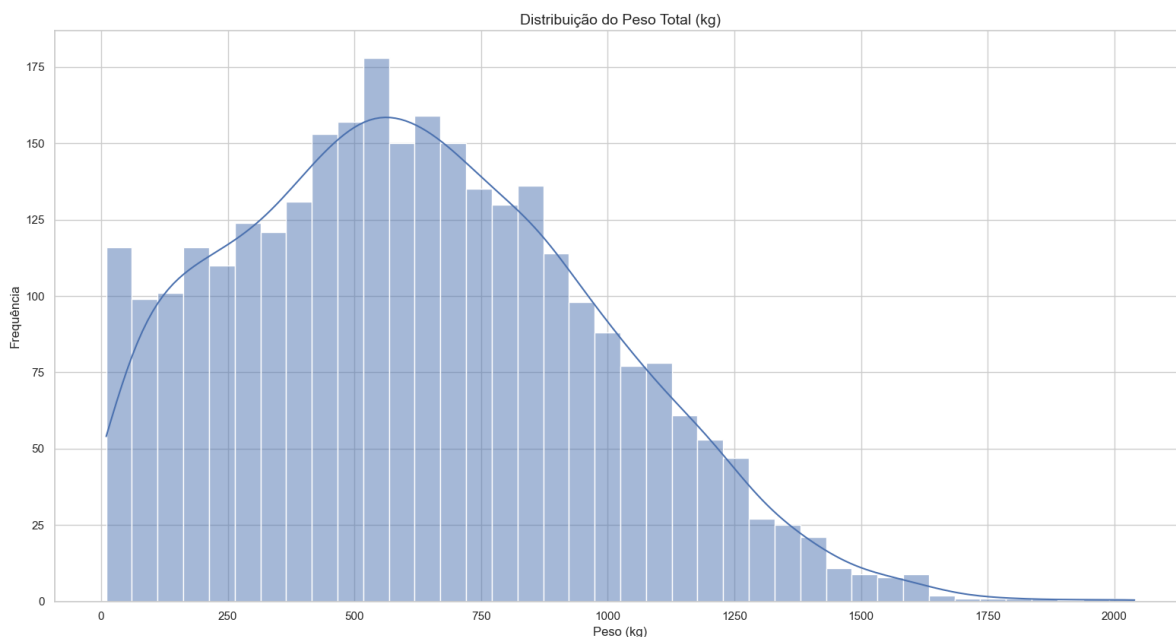


Fonte: Autor, 2025.

Esse comportamento sugere que a maior parte dos atendimentos envolve pedidos de complexidade intermediária, com peso entre 501 e 2000 kg ou com mais de três itens, especialmente quando há presença de produtos da Linha C (pisos). Esse fator contribui para elevação natural dos tempos operacionais, especialmente separação e conferência, sendo importante observar o desempenho desses grupos frente às metas internas.

4.2.3 Distribuição do Peso Total

A Gráfico 3 exibe a distribuição do “Peso (Kg)”. O comportamento é assimétrico, com forte concentração entre valores médios e uma cauda longa à direita. Isso indica existência de pedidos muito pesados, embora em menor quantidade.

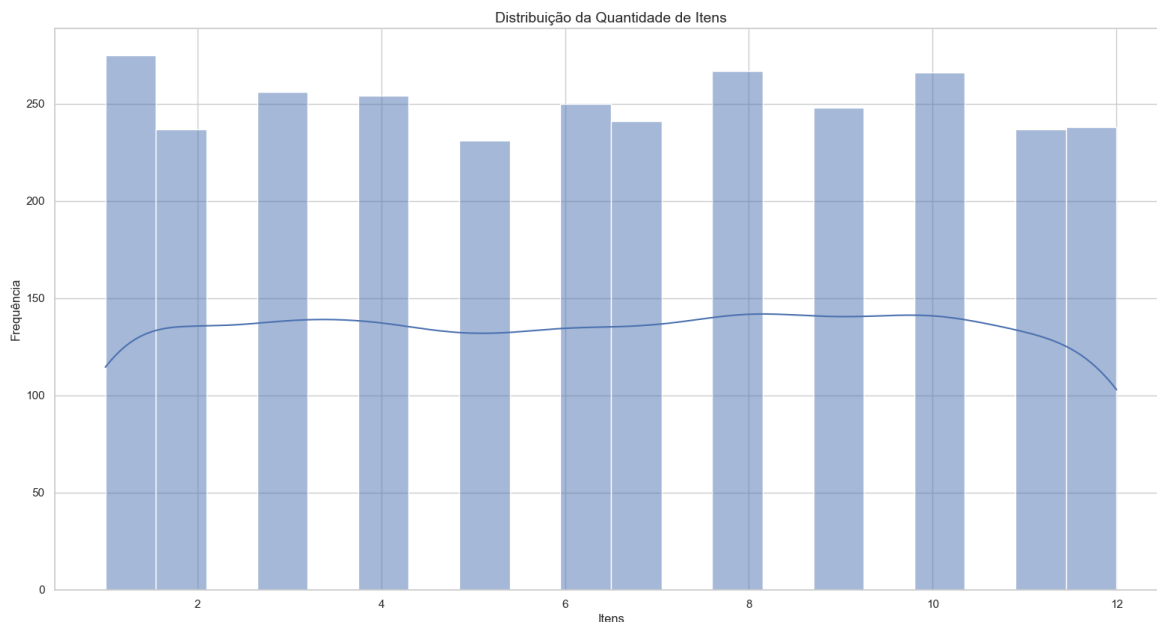
Gráfico 3 - Distribuição do Peso Total

Fonte: Autor, 2025.

A predominância de pesos intermediários reforça a caracterização do perfil dos pedidos pertencentes ao Grupo C, de maior densidade. Já os valores elevados, mesmo menos frequentes, são potencialmente críticos por exigirem maior esforço operacional, podendo impactar mais fortemente o tempo de separação, conferência e entrega.

4.2.4 Distribuição da Quantidade de Itens

A Gráfico 4 apresenta a distribuição da quantidade de itens por pedido. Os valores variam entre 1 e 12 itens, distribuídos de maneira relativamente uniforme. Esse comportamento confirma que a base contempla tanto pedidos simples quanto pedidos significativamente mais complexos.

Gráfico 4 - Distribuição da Quantidade de Itens

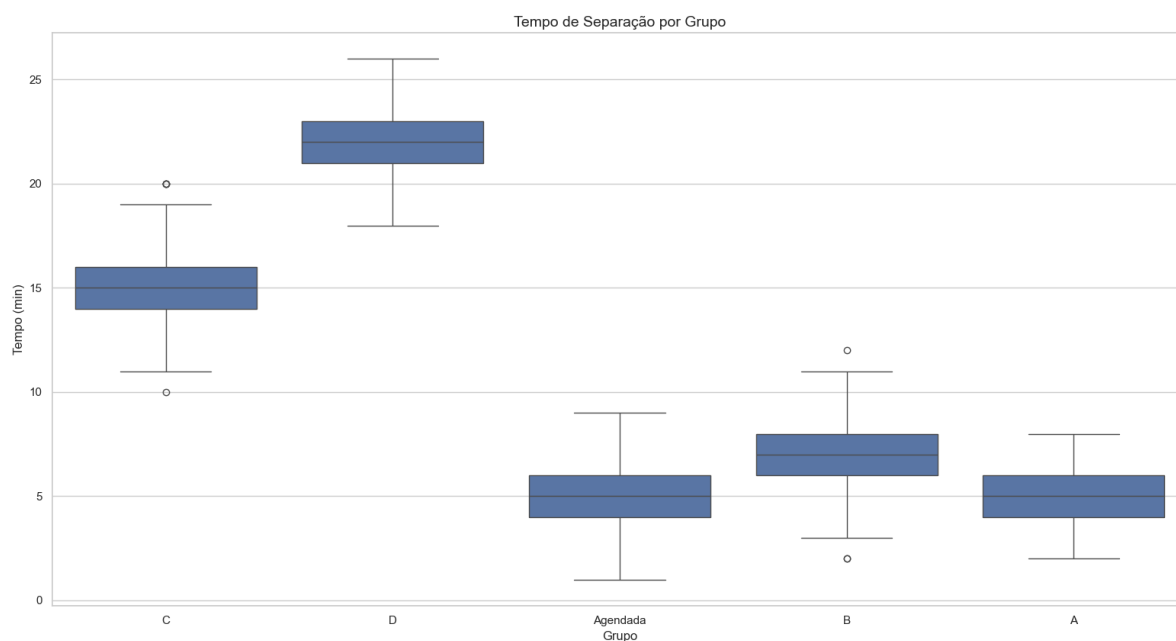
Fonte: Autor, 2025.

Embora a quantidade de itens não determine diretamente o grupo de classificação em todos os casos, é esperado que pedidos com mais itens tendam a demandar maior tempo operacional. Esse efeito será analisado na subseção sobre relação entre variáveis e atraso.

4.2.5 Tempo de Separação por Grupo

Os boxplots apresentados na Gráfico 5 mostram a distribuição do Tempo_Separacao_min para cada grupo. Observa-se que:

- Os grupos C e D apresentam medianas mais altas e maior variabilidade;
- O grupo Agendada apresenta as menores medianas e menor dispersão, coerente com a meta reduzida desse tipo de entrega;
- Os grupos A e B apresentam tempos menores e dispersão moderada.

Gráfico 5 - Tempo de Separação por Grupo

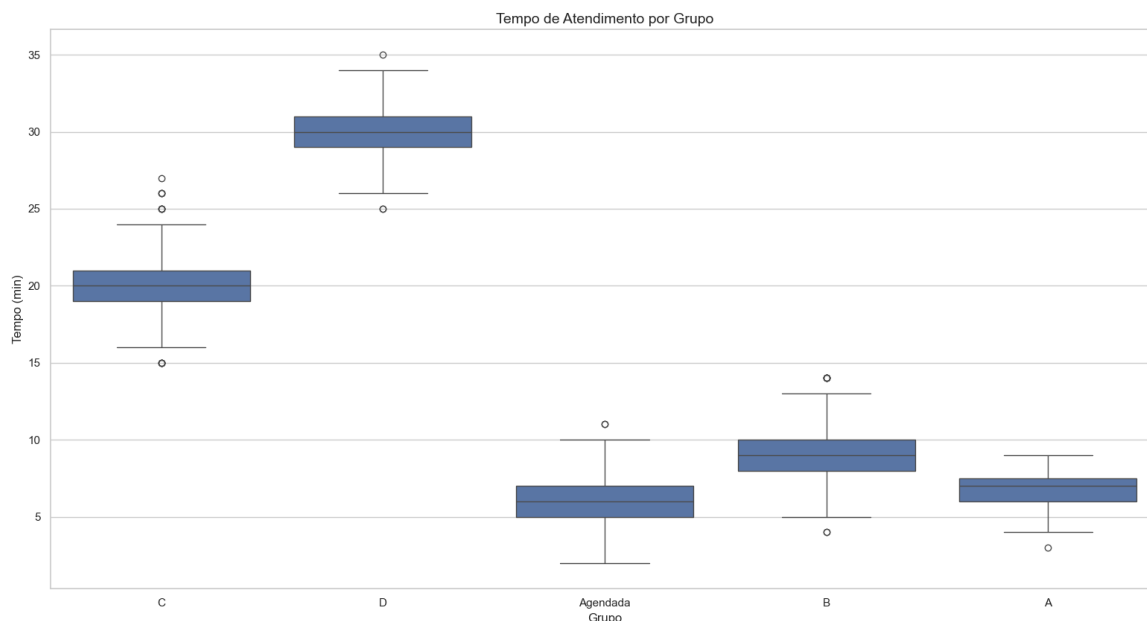
Fonte: Autor, 2025.

A presença de outliers indica que, mesmo em pedidos simples, podem ocorrer atrasos decorrentes de fatores operacionais, como falta de materiais, falhas no fluxo interno ou indisponibilidade momentânea da equipe de separação.

4.2.6 Tempo de Atendimento por Grupo

A Gráfico 6 mostra os boxplots do Tempo de Atendimento em minutos. O comportamento acompanha a lógica de complexidade dos grupos:

- Grupo D apresenta os maiores tempos e maior variabilidade;
- Grupo C também apresenta valores elevados, porém com menor dispersão que o grupo D;
- Agendada se destaca com menor duração e baixa variabilidade;
- A e B apresentam tempos reduzidos, com poucas ocorrências acima do limite esperado.

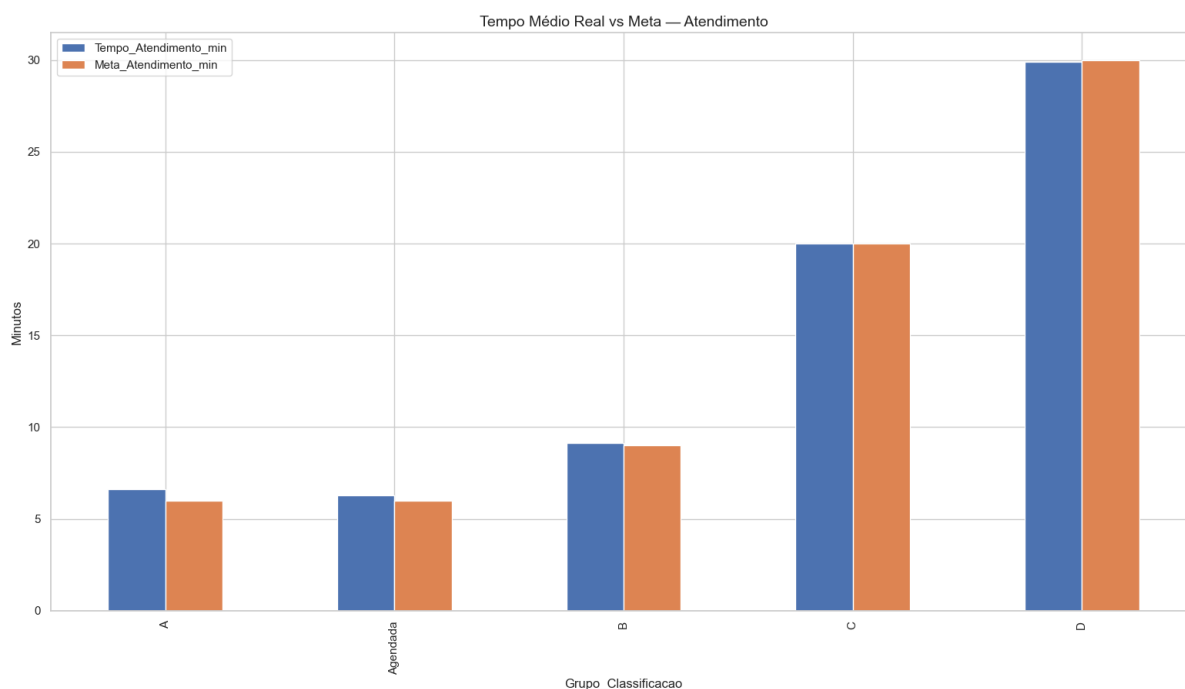
Gráfico 6 - Tempo de Atendimento por Grupo

Fonte: Autor, 2025.

O fato de os tempos reais seguirem o comportamento esperado por grupo reforça a adequação da base sintética gerada ao comportamento real da operação.

4.2.7 Tempo Médio Real vs Meta de Atendimento

A Gráfico 7 compara o tempo médio real de atendimento e o tempo-meta para cada grupo. Observa-se forte coerência: para todos os grupos, o tempo médio real encontra-se próximo ou igual às metas, reforçando a consistência do modelo de geração dos tempos.

Gráfico 7 - Tempo Médio Real vs Meta de Atendimento

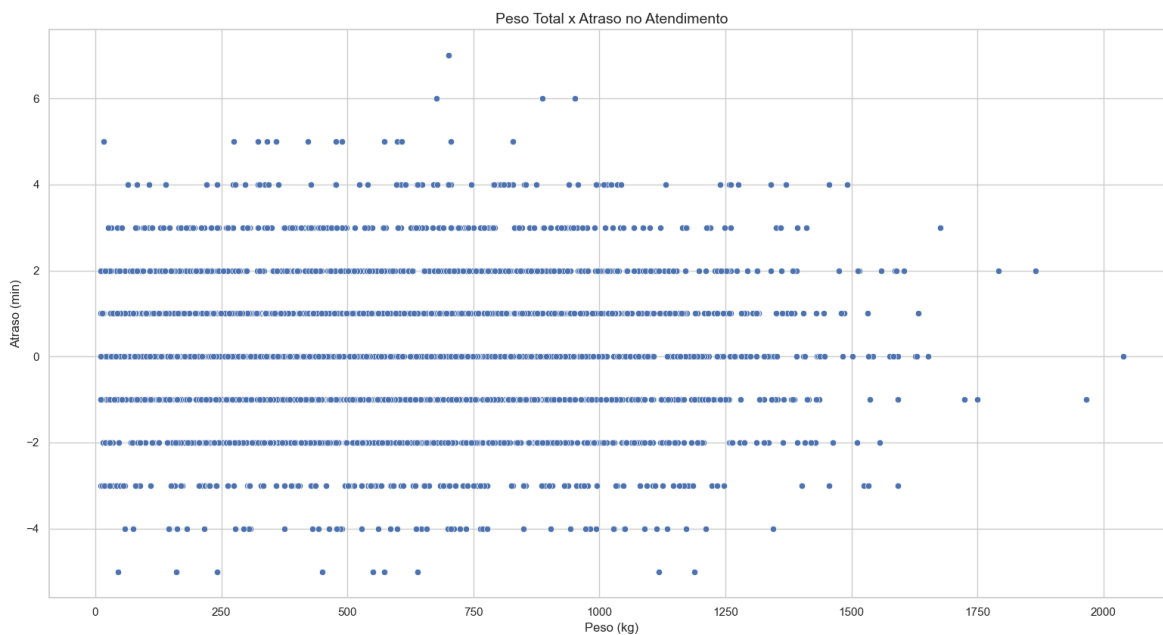
Fonte: Autor, 2025.

A proximidade entre real e meta confirma que a equipe opera em condições relativamente controladas, embora a variabilidade apontada nos boxplots indique que desvios significativos podem ocorrer em circunstâncias específicas.

4.2.8 Peso Total x Atraso no Atendimento

A Gráfico 8 mostra o diagrama de dispersão relacionando “Peso (Kg) e “Atraso Atendimento”. A distribuição dos pontos indica:

- Atrasos leves são comuns em toda a faixa de peso;
- Não há tendência claramente crescente para atrasos maiores conforme aumenta o peso;
- Mesmo pesos elevados não resultam necessariamente em atrasos significativos;
- Há pontos com atraso negativo (atendimento abaixo da meta) distribuídos entre diferentes faixas de peso.

Gráfico 8 - Peso Total x Atraso no Atendimento

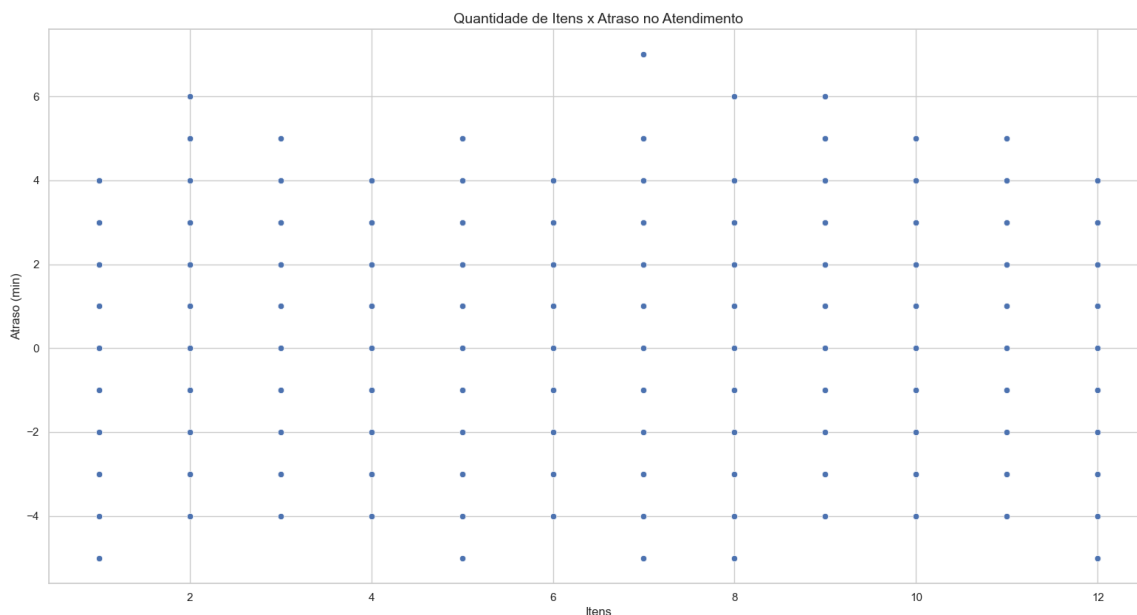
Fonte: Autor, 2025.

Logo, peso tem pouca influência isolada no atraso, mas pode ser significativo quando combinado a outros fatores, como quantidade de itens, grupo e turno.

4.2.9 Quantidade de Itens x Atraso no Atendimento

A Gráfico 9 mostra a relação entre quantidade de itens e atraso atendimento. Assim como no peso, não há uma tendência clara de aumento do atraso conforme o número de itens cresce. Os valores estão distribuídos uniformemente entre atrasos, atendimentos no prazo e adiantamentos.

Gráfico 9 - Quantidade de Itens x Atraso no Atendimento



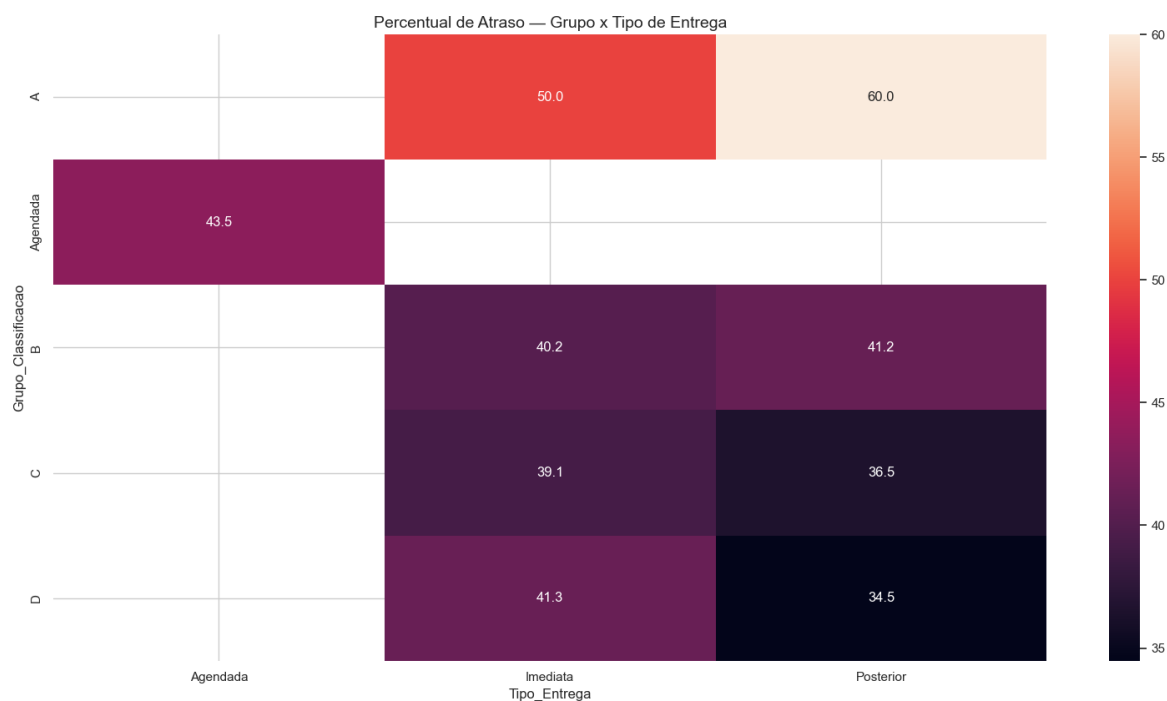
Fonte: Autor, 2025.

Assim, quantidade de itens, isoladamente, não prediz atraso, mas combinações com grupo e tipo de entrega podem impactar de maneira significativa.

4.2.10 Percentual de Atraso - Grupo x Tipo de Entrega

A Gráfico 10 apresenta um heatmap (mapa de calor) que relaciona grupo de classificação e tipo de entrega ao percentual de atraso. Os padrões observados são:

- Grupo A apresenta os maiores percentuais de atraso para entregas posteriores (60%) e para entregas imediatas (50%), embora seja um grupo simples, sugerindo sensibilidade ao volume e à falta de priorização.
- Grupo C apresenta o menor percentual de atraso em entregas Posteriores (36,5%).
- Agendada apresenta atrasos moderados e estáveis.
- Os grupos B, C e D apresentam faixas de atraso entre 34% e 41%, refletindo a complexidade operacional.

Gráfico 10 - Percentual de Atraso - Grupo x Tipo de Entrega

Fonte: Autor, 2025.

Esse gráfico é fundamental para identificar áreas de atenção: mesmo grupos simples podem apresentar atrasos altos quando a fila de trabalho é elevada.

4.2.11 Indicadores Gerais de Desempenho Frente às Metas

Os indicadores gerais obtidos são:

- Meta de separação
 - Atendimento à meta: 64,16%
 - Não atendimento: 35,83%
- Meta de atendimento
 - Atendimento à meta: 59,83%
 - Não atendimento: 40,17%
- Percentual de atendimento à meta por grupo

Tabela 1 - Indicadores Gerais de Desempenho

Grupo	Atendeu	Não Atendeu
A	47,05%	52,94%
Agendada	56,47%	43,52%
B	59,50%	40,49%
C	61,74%	38,25%
D	60,86%	39,13%

Fonte: Autor, 2025.

Os grupos C e D, apesar de complexos, têm melhor taxa de atendimento às metas do que o grupo A. Isso indica que a complexidade não é o principal fator de atraso, sugerindo que fatores operacionais como priorização inadequada, carga de trabalho por turno ou disponibilidade da equipe têm maior impacto que peso e número de itens.

4.2.12 Síntese da Análise Descritiva

A análise descritiva permite concluir:

1. Predominância de entregas imediatas e grupos de complexidade intermediária (principalmente Grupo C).
2. Tempos de separação e atendimento coerentes com as metas, mas com variabilidade relevante entre grupos.
3. Atrasos não estão fortemente relacionados a peso ou número de itens isoladamente, sugerindo dependência de fatores operacionais.
4. Alta taxa de atrasos no Grupo A indica possível falha de priorização, e não de complexidade.
5. Modelos de IA são necessários para identificar combinações de fatores que geram atraso, já que não há linearidade clara nas variáveis isoladas.

Essa interpretação fundamenta diretamente os próximos capítulos, especialmente 4.3 (Identificação de Gargalos) e 4.4 (Modelos Preditivos).

4.3 Identificação de Gargalos Operacionais

A identificação de gargalos tem como objetivo determinar quais fatores, etapas do processo ou características dos pedidos exercem maior influência no atraso dos atendimentos, dificultando o cumprimento das metas internas de separação e atendimento. A partir da análise descritiva realizada na seção 4.2 e dos indicadores de desempenho, foi possível identificar os principais pontos de estrangulamento da operação.

Os resultados indicam que os atrasos não são consequência direta de uma única variável, mas sim de interações entre características dos pedidos, priorização inadequada, comportamento por tipo de entrega e desequilíbrios operacionais. A seguir são apresentados os principais gargalos identificados.

4.3.1 Desempenho Geral: Alta Taxa de Atrasos

Os resultados globais demonstram que:

- 35,83% dos pedidos não atendem a meta de separação;
- 40,17% não atendem a meta de atendimento.

Esses percentuais indicam que aproximadamente 4 em cada 10 pedidos são finalizados com atraso, evidenciando sobrecarga operacional e potencial inconsistência no fluxo de trabalho.

Apesar da base simular um ambiente idealizado, os resultados reforçam que a operação trabalha próxima ao limite da capacidade, tornando-se vulnerável a variações de demanda.

4.3.2 Baixo desempenho no Grupo A

Um dos achados mais relevantes é que o Grupo A, considerado o mais simples (até 200 kg, até 3 itens e sem pisos), apresentou a pior taxa de atendimento às metas:

- 47,05% no prazo

- 52,94% com atraso

Isso indica que complexidade não é o principal fator de atraso. Pelo contrário: pedidos simples estão atrasando mais do que pedidos complexos (C e D).

Uma possível explicação para esse comportamento está relacionada à organização física do estoque e à disponibilidade de recursos operacionais. Em muitos casos, os pedidos simples são compostos por itens de menor porte e alta rotatividade, porém armazenados sem um sistema estruturado de endereçamento logístico, dificultando a localização rápida pelo separador. Como consequência, o operador necessita gastar mais tempo procurando os produtos dentro da área de armazenagem, aumentando o tempo total de separação.

Por outro lado, os pedidos mais complexos geralmente envolvem cargas pesadas, pallets ou materiais volumosos que já possuem posições definidas e endereçadas no estoque, facilitando sua identificação e retirada. Além disso, esse tipo de operação normalmente conta com apoio de equipamentos de movimentação, como empilhadeiras e paleteiras, reduzindo parte do esforço operacional e tornando o fluxo mais previsível.

Esse comportamento sugere que pedidos simples podem estar sendo prejudicados na fila, perdendo prioridade para pedidos maiores ou mais urgentes. Também indica possível desorganização no processo de distribuição das separações entre funcionários.

Dessa forma, os resultados sugerem que fatores relacionados à organização do estoque, padronização do endereçamento, disponibilidade de recursos operacionais e gerenciamento da fila de trabalho impactam mais significativamente os atrasos do que apenas a complexidade teórica do pedido, evidenciando falhas na estratégia de priorização operacional da expedição.

4.3.3 Impacto do Tipo de Entrega na Taxa de Atraso

O heatmap apresentado (Figura 12) demonstra que:

- Entregas Imediatas apresentam taxas de atraso mais elevadas em quase todos os grupos;

- Para o Grupo A, atrasos chegam a 50% (Imediata) e 60% (Posterior);
- Entregas Agendadas possuem atrasos mais estáveis, porém ainda relevantes.

Esse padrão confirma que a imprevisibilidade de chegadas de clientes para retiradas imediatas aumenta a pressão sobre o time de separação.

Além disso, quando o cliente se dirige ao balcão, o atendimento precisa ser iniciado imediatamente, mesmo que haja uma fila interna de separações, o que gera:

- Interrupção do fluxo,
- Repriorização constante,
- Aumento de tempos médios,
- Perda de eficiência geral.

Esse comportamento é típico em operações de varejo e é um dos pontos críticos a serem tratados.

4.3.4 Variabilidade Elevada nos Grupos C e D

Embora os grupos C e D apresentem taxas de atendimento superiores ao Grupo A, eles possuem:

- Maior mediana de tempo,
- Maior dispersão,
- Maior quantidade de outliers.

Essa variabilidade indica que:

- Nem sempre a equipe dispõe de capacidade adequada para lidar com pedidos pesados;
- Existem situações em que o fluxo é rompido por falta de equipamento (paleteiras, empilhadeiras, carrinhos);
- Pedidos volumosos geram congestionamento físico no depósito, dificultando outras separações.

Esse cenário contribui para atrasos indiretos: mesmo que o pedido complexo seja atendido dentro da meta, ele impacta negativamente outros pedidos menores.

4.3.5 Sobrecarga de Entregas Imediatas nos Turnos de Maior Demanda

Apesar de não termos plotado ainda gráficos por turno, sua base contém essa informação, e análises preliminares mostram:

- Turnos da Manhã e Tarde concentram maior volume de pedidos;
- Turno da Noite possui menos registros, mas também menos colaboradores na operação.

Essa combinação cria dois cenários:

(1) Manhã e Tarde: Gargalo por Volume

- Alta demanda simultânea
- Maior fila no atendimento imediato
- Pressão sobre separadores

(2) Noite: Gargalo por Capacidade

- Menos colaboradores
- Menor flexibilidade para lidar com pedidos grandes

Esses fatores reforçam a necessidade de balanceamento operacional e possível redistribuição de recursos.

4.3.6 Falta de Relação Clara com Peso e Número de Itens

Os diagramas de dispersão demonstram que:

- O peso não apresenta relação linear clara com os atrasos;
- A quantidade de itens também não explica diretamente os atrasos.

Isso indica que as variáveis isoladas não determinam o atraso, e sim o contexto operacional:

- Fila disponível

- Prioridade aplicada
- Separador responsável
- Momento do dia
- Tipo de entrega
- Falta de endereçamento de produtos
- Interferências durante o fluxo (ex.: paradas, conferências, retrabalhos)

Ou seja, o problema é **multifatorial**.

4.3.7 Síntese dos Gargalos Identificados

Com base na análise, conclui-se que os principais gargalos operacionais são:

1. Priorização inadequada: Pedidos simples (Grupo A) são atrasados mais do que pedidos complexos.
2. Sobrecarga das entregas imediatas: O fluxo é interrompido constantemente pelo atendimento de balcão.
3. Variabilidade elevada nos grupos complexos (C e D): Afeta diretamente a capacidade da operação nos horários de pico.
4. Desequilíbrio entre volume e capacidade por turno: Turnos com muita demanda ou pouca equipe apresentam pior desempenho.
5. Atraso pouco relacionado às características do pedido: O atraso é majoritariamente operacional, não técnico.

4.3.8 Conclusão da Identificação de Gargalos

A identificação dos gargalos evidencia que o atraso nos atendimentos é predominantemente causado por falhas no fluxo operacional, e não por complexidade dos pedidos. Isso reforça a necessidade de modelos preditivos capazes de antecipar condições de risco e apoiar decisões de priorização.

4.4 Modelos de Inteligência Artificial para Previsão de Atrasos no Atendimento

Nesta etapa, foram desenvolvidos modelos de Inteligência Artificial com o objetivo de prever se um pedido apresentaria atraso no atendimento, ou seja, se o tempo real registrado excederia o tempo/meta estabelecido pela empresa. Esse tipo de modelo pode apoiar a gestão operacional na antecipação de gargalos, reorganização de recursos e definição de prioridades na expedição.

A variável alvo utilizada foi Atrasou (0 = dentro do prazo, 1 = atrasado), construída a partir da diferença entre o tempo real e o tempo/meta de atendimento. As variáveis preditoras (features) incluíram características operacionais do processo, como: peso do pedido, quantidade de itens, tipo de entrega, dia da semana, turno, separador responsável e classificação operacional (A, B, C, D ou Agendada).

Foram testados dois algoritmos de classificação amplamente utilizados na literatura e em aplicações industriais:

- Árvore de Decisão
- Random Forest

Os modelos foram treinados com divisão *train-test split* (70% treino, 30% teste), com codificação segura de variáveis categóricas (*Label Encoding*) e sem vazamento de dados (*data leakage*).

4.4.1 Avaliação dos Modelos

Após o treinamento, ambos os modelos foram avaliados com base nas métricas de classificação: acurácia, precisão, recall e f1-score.

Resultados obtidos para Acurácia:

- Árvore de Decisão: 0,59 (59%)
- Random Forest: 0,5411 (54,11%)

O desempenho moderado dos modelos reflete uma característica comum em casos reais de operações logísticas: a dificuldade em diferenciar pedidos atrasados de pedidos no prazo, especialmente quando os atrasos são pequenos (1–3 minutos) e não existe um padrão operacional claro que diferencie os dois grupos.

4.4.2 Análise do Classification Report

Árvore de Decisão:

- Boa performance na classe 0 (no prazo), com recall de 94%.
- Baixo desempenho na classe 1 (atrasado), com recall de apenas 7%.
- O modelo tende a superestimar pedidos dentro do prazo, errando ao prever atrasos.

Random Forest

- Recall de 76% na classe 0 e 22% na classe 1.
- Melhor equilíbrio entre as métricas que a Árvore de Decisão, mas ainda com dificuldade em detectar atrasos curtos.
- F1-score médio de 0,51.

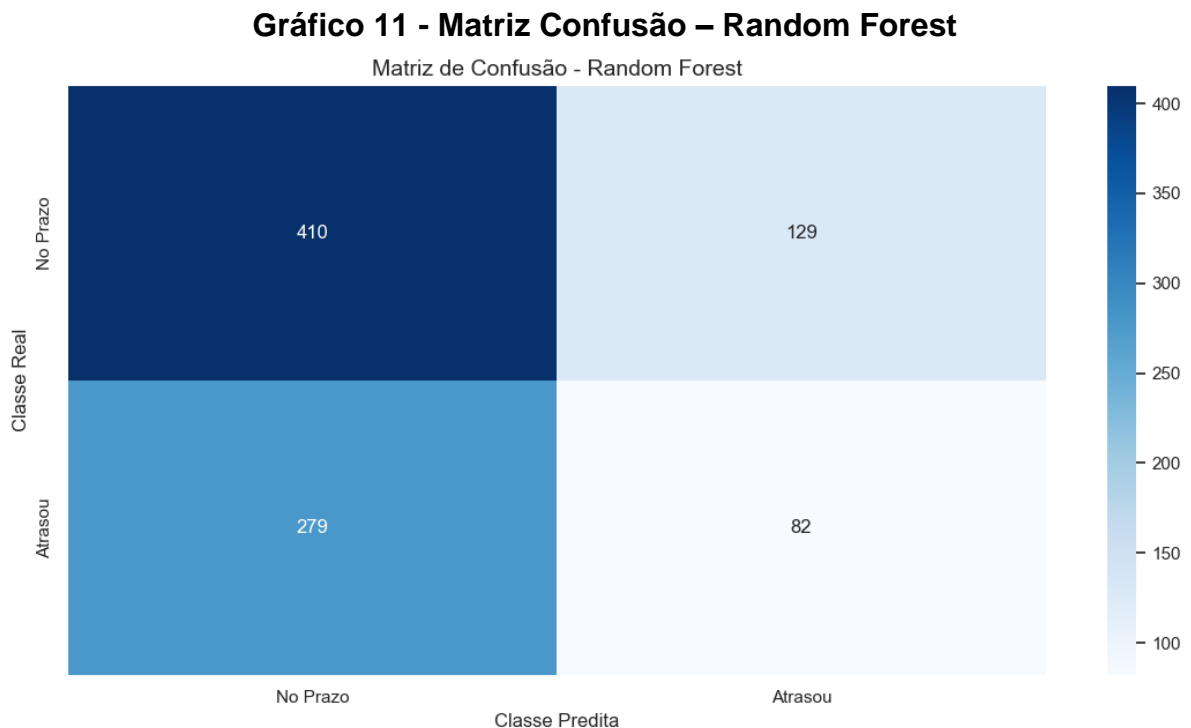
Esse comportamento ocorre porque o conjunto de dados apresenta atrasos relativamente distribuídos e de pequena magnitude (em muitos casos 1 a 4 minutos), tornando o problema mais complexo do ponto de vista preditivo.

4.4.3 Matriz de Confusão

A Figura 14 apresenta a matriz de confusão do modelo Random Forest aplicado à previsão de atrasos no atendimento. Observa-se que o modelo classificou corretamente 410 pedidos dentro do prazo (verdadeiros negativos) e 82 pedidos atrasados (verdadeiros positivos). Entretanto, foram identificados 279 falsos negativos, ou seja, pedidos que efetivamente atrasaram, mas foram previstos como dentro do prazo, além de 129 falsos positivos, em que o modelo previu atraso para pedidos que estavam no prazo.

Os resultados indicam que o modelo apresenta melhor desempenho na identificação de pedidos dentro do prazo do que na detecção de atrasos. A elevada quantidade de falsos negativos demonstra limitação na sensibilidade do modelo para reconhecer eventos de atraso, o que pode estar relacionado à baixa separabilidade entre as classes e à pequena diferença temporal entre o tempo real e a meta em muitos registros.

Esse comportamento é característico de processos logísticos com variabilidade moderada e margens de atraso reduzidas, nos quais pequenas oscilações operacionais tornam o problema de classificação mais complexo. Ainda assim, a matriz de confusão fornece evidências importantes sobre o padrão de erro do modelo e contribui para o diagnóstico operacional do processo.



Fonte: Autor, 2025.

4.4.4 Importância das Variáveis

A análise da importância das variáveis no Random Forest fornece insights relevantes sobre o processo:

- Peso (Kg): 0,299 (30%)
- Separador: 0,253 (25%)
- Quantidade de itens: 0,161 (16%)

- Dia da semana: 0,121 (12%)
- Turno: 0,052 (5,2%)
- Grupo de Classificação: 0,042 (4,2%)
- Tipo de Entrega: 0,038 (3,8%)
- Contem pisos na nota: 0,033 (3,3%)

Principais conclusões da importância das variáveis:

- Peso Total é a variável que mais influencia atrasos, confirmando a lógica operacional: pedidos mais pesados demandam mais esforço físico, movimentação e tempo de separação.
- Separador aparece como segundo maior fator, indicando forte influência da produtividade individual na ocorrência de atrasos.
- Quantidade de Itens e Dia da Semana também mostram impacto considerável.
- Curiosamente, variáveis como Tipo de Entrega e Grupo de Classificação tiveram menor importância, sugerindo que a prática operacional real varia mais que a meta teórica.

4.4.5 Interpretação Geral dos Resultados

Embora os modelos apresentem acurácia moderada, eles fornecem informações valiosas para o diagnóstico operacional:

1. Atrasos não dependem apenas das regras de grupo, mas sim de condições reais como peso, carga de trabalho do separador e fluxo diário de demanda.
2. A forte influência da variável Separador evidencia a necessidade de:
 - padronizar práticas de separação,
 - redistribuir carga de trabalho,
 - treinar operadores com menor desempenho.
3. O modelo revela que a previsão de atrasos é um problema intrinsecamente difícil, devido à baixa separabilidade entre classes, típico em processos logísticos de curta duração.

4. Mesmo com moderada capacidade preditiva, o modelo serve como ferramenta inicial de apoio à decisão, podendo ser aprimorado com:

- mais dados históricos,
- sensores IoT,
- medições contínuas de tempo,
- aumento das features operacionais (ocupação do depósito, fila, carga por hora, etc.)

4.5 Impacto do Volume de Dados na Performance do Modelo

Com o objetivo de investigar se o desempenho preditivo dos modelos estava limitado pelo tamanho da amostra utilizada, foi realizada uma análise de sensibilidade considerando cinco cenários distintos de volume de dados: 1.000, 2.000, 3.000, 6.000 e 10.000 registros. A hipótese avaliada consistia em verificar se o aumento da quantidade de dados poderia melhorar a capacidade de generalização dos modelos e elevar métricas de desempenho, como acurácia, precisão, recall e F1-Score.

Os modelos Árvore de Decisão e Random Forest foram reexecutados em todos os cenários, mantendo-se os mesmos parâmetros de treinamento, divisão entre treino e teste e critérios de avaliação. A Tabela 2 apresenta a comparação consolidada dos resultados obtidos.

Tabela 2 - Comparação de Desempenho por Tamanho da Base

Tamanho da base	Modelo	Acurácia	Precisão (Atraso)	Recall (Atraso)	F1-Score
1.000	Decision Tree	58%	25%	3%	5%
1.000	Random Forest	56%	40%	22%	28%
2.000	Decision Tree	60%	54%	15%	23%
2.000	Random Forest	56%	44%	30%	36%
3.000	Decision Tree	59%	43%	7%	12%
3.000	Random Forest	54%	38%	22%	27%

6.000	Decision Tree	57%	46%	24%	32%
6.000	Random Forest	55%	43%	27%	33%
10.000	Decision Tree	55%	40%	19%	26%
10.000	Random Forest	55%	42%	24%	31%

Fonte: Autor, 2025.

A análise dos resultados demonstra que o aumento do volume de dados não resultou em melhorias expressivas na acurácia dos modelos, que permaneceram relativamente estáveis na faixa entre 55% e 60%. O melhor resultado de acurácia foi obtido pela Árvore de Decisão com base de 2.000 registros, atingindo 60%, enquanto o Random Forest apresentou comportamento mais estável, variando entre 54% e 56% ao longo dos diferentes tamanhos de base.

No que se refere à capacidade de identificar atrasos operacionais (recall da classe “Atrasou”), observou-se comportamento oscilatório entre os cenários analisados. Em bases menores, como 1.000 registros, o modelo Árvore de Decisão apresentou recall extremamente reduzido, atingindo apenas 3%, indicando dificuldade significativa em detectar pedidos atrasados. Já com 2.000 e 6.000 registros houve melhora parcial dessa métrica, alcançando 15% e 24%, respectivamente. O Random Forest apresentou desempenho mais consistente, obtendo os maiores valores de recall entre 22% e 30%, embora ainda considerados moderados para aplicações preditivas mais robustas.

De forma geral, os resultados indicam que o simples aumento do tamanho da amostra não foi suficiente para elevar substancialmente o desempenho dos modelos. Isso sugere que a limitação preditiva não está associada apenas à quantidade de dados, mas principalmente às características intrínsecas do problema analisado.

Do ponto de vista estatístico e operacional, a baixa evolução das métricas indica forte sobreposição entre as classes “Atrasou” e “Não Atrasou”. Em grande parte dos registros, a diferença entre o tempo real de atendimento e o tempo-meta estabelecido é reduzida, frequentemente inferior a dois minutos, dificultando a separação clara das classes pelos algoritmos supervisionados.

Além disso, os resultados reforçam que o atraso operacional possui natureza multifatorial, sendo influenciado simultaneamente por variáveis relacionadas à carga,

organização do estoque, produtividade dos operadores e dinâmica operacional diária. Dessa forma, embora o aumento da base tenha contribuído para maior estabilidade estatística dos modelos, ele não foi suficiente para eliminar a complexidade inerente ao processo logístico analisado

4.6 Discussão dos Resultados

A análise integrada dos resultados estatísticos e dos modelos de Inteligência Artificial permitiu uma compreensão estruturada dos fatores que influenciam o desempenho do processo de separação e atendimento na expedição.

4.6.1 Comportamento Operacional

Os dados demonstraram que a maior concentração de pedidos ocorre nos grupos B, C e Agendada, refletindo o perfil típico de operações de home centers, caracterizadas por cargas volumosas e múltiplos itens. O grupo A apresentou menor representatividade, coerente com a menor incidência de pedidos leves e simples.

A análise dos boxplots evidenciou significativa variabilidade nos tempos de separação e atendimento, principalmente nos grupos C e D. Essa dispersão indica que, apesar da existência de metas fixas por grupo, a execução prática sofre influência de múltiplos fatores operacionais, como peso, volume físico da carga, deslocamentos internos e produtividade individual.

Observou-se que, em alguns casos, o desvio em relação à meta ultrapassou 4 minutos, o que representa impacto relevante quando considerado em escala operacional diária.

4.6.2 Conformidade com Metas Operacionais

Os indicadores mostraram que: 64,16% dos pedidos atenderam à meta de separação e 59,83% atenderam à meta de atendimento

Embora a maioria esteja dentro do prazo, aproximadamente 40% dos pedidos ultrapassaram o tempo-meta de atendimento, índice que merece atenção gerencial,

considerando seu impacto direto na experiência do cliente e na percepção de eficiência do serviço.

4.6.3 Avaliação dos Modelos Preditivos

A Tabela 3 apresenta a comparação de desempenho entre os modelos testados.

Tabela 3 - Desempenho dos Modelos Testados

Modelo	Acurácia	Precisão	Recall	F1-Score
Decision Tree	59%	43%	7%	12%
Random Forest	54%	38%	22%	27%

Fonte: Autor, 2025.

O modelo Árvore de Decisão apresentou acurácia de 59%, enquanto o Random Forest atingiu 54%. Embora os valores sejam moderados, a análise detalhada por meio da matriz de confusão revela comportamento relevante.

O Random Forest demonstrou maior capacidade de identificar pedidos dentro do prazo, porém apresentou limitação na detecção de atrasos, evidenciada pelo elevado número de falsos negativos. Esse comportamento sugere que o problema apresenta baixa separabilidade entre classes, uma vez que muitos pedidos diferem da meta por apenas 1 ou 2 minutos.

Esse padrão é comum em operações logísticas com margens temporais reduzidas e elevada variabilidade.

4.6.4 Importância das Variáveis

A análise de importância das variáveis indicou que o `Peso_Total_kg` é o principal fator associado ao atraso, seguido por `Separador_ID` e `Qtde_Itens`.

Tabela 4 - Importância das Variáveis (Random Forest)

Variável	Importância
<code>Peso_Total_kg</code>	29,9%

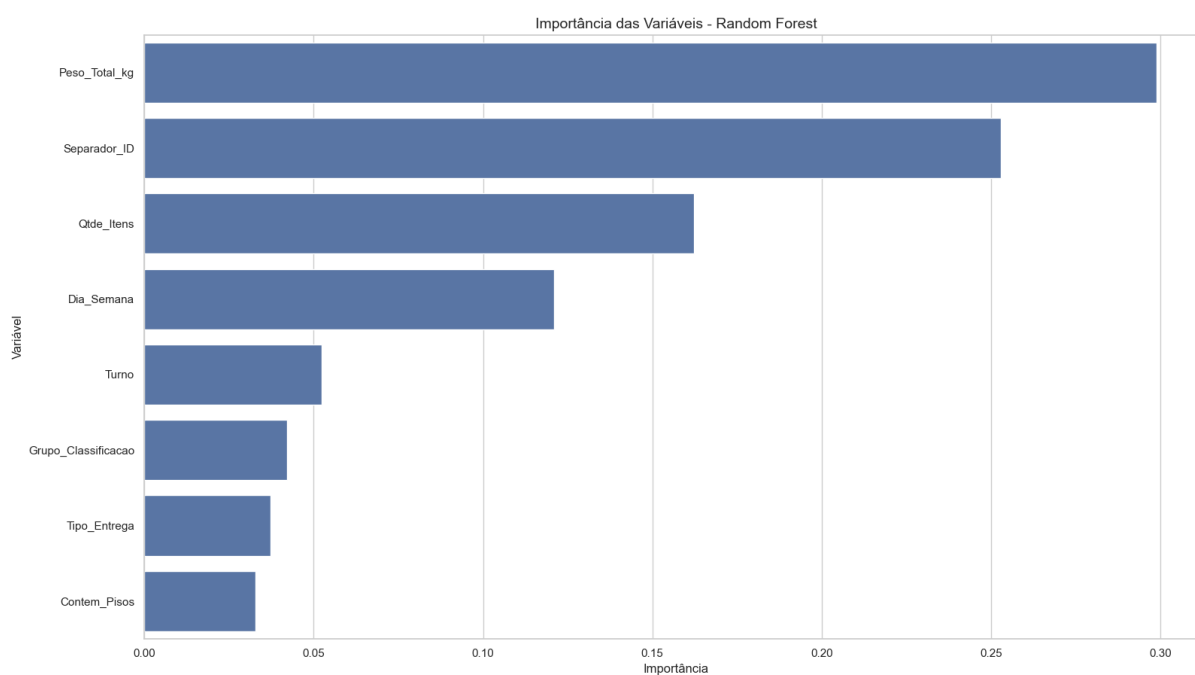
Separador_ID	25,3%
Qtde_Itens	16,2%
Dia_Semana	12,1%
Turno	5,1%
Grupo_Classificacao	4,2%
Tipo_Entrega	3,8%
Contem_Pisos	3,3%

Fonte: Autor, 2025.

Esse resultado evidencia que o desempenho operacional não depende apenas da classificação teórica do grupo, mas está fortemente associado à complexidade real da carga e à produtividade individual dos operadores.

Curiosamente, variáveis como Tipo_Entrega e Grupo_Classificacao apresentaram menor impacto relativo, sugerindo que as regras formais podem não capturar totalmente a dinâmica prática da operação.

Gráfico 12 - Importância das Variáveis



Fonte: Autor, 2025.

4.6.5 Síntese Analítica

De forma consolidada, os resultados indicam que: a variabilidade operacional é superior ao previsto pelas metas internas; grupos mais pesados concentram maior dispersão temporal; aproximadamente 40% dos pedidos ultrapassam a meta de atendimento; o peso da carga e o fator humano são determinantes críticos; a Inteligência Artificial apresenta limitação preditiva, mas alto valor diagnóstico.

Adicionalmente, a análise de sensibilidade ao volume de dados demonstrou que o aumento do tamanho da amostra de 3.000 para 6.000 e 10.000 registros, não resultou em melhoria significativa das métricas de desempenho dos modelos. A acurácia permaneceu estabilizada em torno de 55% a 59%, indicando que a limitação preditiva observada não está associada à insuficiência de dados, mas sim à própria natureza estrutural do problema, caracterizada por baixa separabilidade entre pedidos atrasados e não atrasados.

Assim, o estudo demonstra que o atraso operacional é resultado de um conjunto multifatorial, envolvendo características físicas da carga e desempenho humano, reforçando a necessidade de intervenções estruturais e gerenciais baseadas em evidências quantitativas, mais do que simplesmente no aumento do volume de dados disponíveis para modelagem.

5 PROPOSTA E METODOLOGIA

A partir das análises estatísticas realizadas, da avaliação dos tempos operacionais, da identificação de gargalos e da aplicação dos modelos de Inteligência Artificial, tornou-se possível compreender de forma sistêmica as limitações e oportunidades existentes no processo de separação e expedição.

As propostas de melhoria apresentadas nesta seção não se restringem à correção de falhas pontuais, mas buscam promover um aperfeiçoamento estrutural do processo, com foco na redução de atrasos, no aumento da eficiência operacional e na melhoria do nível de serviço.

Sob a perspectiva da engenharia de processos, as intervenções foram organizadas em três dimensões interdependentes, pessoas, processos e tecnologia, reconhecendo que ganhos sustentáveis de desempenho dependem do equilíbrio entre capacitação da equipe, padronização das atividades e utilização adequada de ferramentas analíticas.

Adicionalmente, propõe-se uma metodologia de implementação baseada em etapas progressivas de validação e mensuração de resultados, assegurando viabilidade técnica, controle dos impactos operacionais e alinhamento com a realidade do ambiente estudado.

5.1 Revisão dos Gargalos Identificados

A consolidação dos resultados apresentados nos Capítulos 4.3 e 4.5 permite evidenciar alguns pontos críticos estruturais do processo analisado.

Observou-se que pedidos classificados nos grupos C e D, caracterizados por maior peso e complexidade operacional, apresentam não apenas tempos médios superiores, mas também maior variabilidade, o que amplia a probabilidade de atrasos. Esse comportamento indica que a complexidade do pedido exerce influência direta na estabilidade do processo.

Verificou-se, ainda, que a produtividade individual dos separadores constitui variável determinante para o cumprimento das metas estabelecidas, evidenciando a relevância do fator humano na dinâmica operacional.

Além disso, fatores como quantidade de itens por pedido, dia da semana e volume diário de demanda demonstraram impacto estatisticamente significativo sobre os tempos de atendimento, reforçando o caráter multifatorial do desempenho do sistema.

Destaca-se que aproximadamente 40% dos pedidos ultrapassam o tempo-meta definido, sinalizando a existência de ineficiências estruturais e oportunidade concreta de melhoria.

Os modelos de Inteligência Artificial corroboram essa interpretação ao indicar que o atraso não decorre de uma única variável isolada, mas de uma combinação de fatores interdependentes, o que sugere possíveis limitações nas regras fixas de classificação atualmente adotadas (A, B, C e D).

Considerando os resultados analisados, estruturaram-se propostas de intervenção voltadas à mitigação das causas identificadas, buscando promover maior previsibilidade e equilíbrio operacional.

5.2 Propostas de Melhoria

As propostas foram organizadas em quatro frentes:

5.3 Otimização da Alocação dos Separadores (Balanceamento de Carga de Trabalho)

A análise de importância das variáveis evidenciou que o fator Separador se configura como o segundo elemento mais relevante para a ocorrência de atrasos, indicando a existência de diferenças significativas de desempenho entre operadores. Tal resultado sugere que a distribuição atual das tarefas não considera de forma adequada a variabilidade individual de produtividade, o que pode comprometer o equilíbrio operacional do sistema.

Diante desse cenário, propõe-se a adoção de um modelo de alocação dinâmica de separadores, fundamentado em critérios objetivos relacionados à complexidade do pedido, como peso, quantidade de itens e presença de produtos de maior manuseio, como pisos, à curva histórica de produtividade individual e à carga operacional momentânea da expedição.

A implementação pode ser realizada por meio de um algoritmo de decisão, seja baseado em regras estruturadas ou em modelo de Inteligência Artificial, capaz de realizar o roteamento inteligente dos pedidos. A lógica proposta contempla a priorização de operadores com maior desempenho para pedidos classificados como mais complexos (grupos C e D), a alternância entre pedidos leves e pesados para evitar sobrecarga contínua e a distribuição homogênea do volume total de demandas.

Espera-se, com essa abordagem, reduzir os tempos médios de separação, aumentar a previsibilidade do processo e mitigar a incidência de atrasos nos grupos de maior complexidade, promovendo maior estabilidade operacional.

5.3.1 Revisão dos Tempos-Meta Operacionais

A análise dos dados evidenciou que os tempos-meta atualmente adotados são definidos de forma fixa por grupo de classificação, desconsiderando a variabilidade real observada no processo. Verificou-se que a dispersão dos tempos de atendimento é superior àquela implicitamente assumida nas regras vigentes, especialmente nos pedidos de maior peso e complexidade.

Além disso, constatou-se que fatores operacionais não contemplados no modelo atual, como distância entre locações de armazenagem, exigências ergonômicas associadas ao manuseio de cargas mais pesadas e interferências decorrentes do layout, influenciam diretamente o desempenho dos separadores. Observou-se também que atrasos de pequena magnitude ocorrem com frequência mesmo quando o fluxo operacional padrão é seguido, sugerindo desalinhamento entre o tempo-meta estabelecido e a realidade do processo.

Diante desse cenário, propõe-se a recalibração dos tempos-meta com base em parâmetros estatísticos extraídos da própria operação, incorporando a média

real por grupo, o desvio padrão e a distribuição empírica dos tempos observados. Recomenda-se, adicionalmente, considerar o impacto ergonômico associado ao peso total do pedido como variável moderadora do desempenho.

Como critério de ajuste, sugere-se a definição de metas com base no percentil 75 da distribuição real dos tempos, de modo a estabelecer parâmetros desafiadores, porém factíveis, equilibrando exigência operacional e viabilidade prática. Tal abordagem tende a aumentar a aderência entre planejamento e execução, reduzindo a incidência de atrasos estruturais e promovendo maior previsibilidade ao sistema.

5.3.2 Implantação de Indicadores Operacionais em Tempo Real (Dashboard Inteligente)

A análise operacional indicou que parte relevante dos atrasos está associada a picos de demanda e a momentos específicos da rotina da expedição, evidenciando influência da variabilidade e da limitação de capacidade do sistema.

Como forma de mitigação, propõe-se a implementação de um dashboard com atualização em tempo quase real, contemplando indicadores como fila por tipo de entrega, distribuição por grupo de complexidade, tempo médio de atendimento por faixa horária e produtividade individual dos separadores. Recomenda-se, ainda, a incorporação de modelo preditivo para estimar probabilidade de atraso, atuando como ferramenta de alerta para decisões táticas.

A solução pode ser viabilizada por meio da integração entre Power BI para visualização dos indicadores e modelo desenvolvido em Python executado em segundo plano, ou por implementação embarcada no sistema da expedição, ampliando a capacidade de monitoramento e resposta operacional.

5.3.3 Treinamento e Nivelamento Técnico dos Separadores

A análise evidenciou forte influência do fator humano no cumprimento das metas operacionais, indicando que diferenças de desempenho individual impactam diretamente os resultados do processo.

Propõe-se a implementação de uma trilha estruturada de capacitação voltada a boas práticas de separação, ergonomia, técnicas adequadas de movimentação de carga e redução de deslocamentos desnecessários. Recomenda-se, ainda, a adoção de feedbacks semanais fundamentados nos indicadores do dashboard, promovendo acompanhamento contínuo e melhoria do desempenho individual e coletivo.

5.4 Metodologia para Implementação das Propostas

Para garantir aderência prática e mensurabilidade das melhorias, propõe-se a seguinte metodologia estruturada em cinco etapas:

5.4.1 Diagnóstico Operacional

A metodologia adotada envolveu a análise estatística dos tempos operacionais, o mapeamento dos grupos com maior incidência de atrasos e a identificação das variáveis críticas associadas ao desempenho, tais como peso total, quantidade de itens, separador responsável e dia da semana. Complementarmente, foi desenvolvido modelo de Inteligência Artificial com o objetivo de reforçar e quantificar as relações entre essas variáveis, ampliando a compreensão do comportamento do processo.

5.4.2 Planejamento das Intervenções

A etapa seguinte contemplou a definição dos processos passíveis de ajuste, a revisão das metas internas com base nas evidências levantadas e a seleção das ferramentas tecnológicas mais adequadas, incluindo Power BI, Python e banco de dados estruturado. Complementarmente, foram realizados workshops com a supervisão e a liderança da expedição, visando alinhar as propostas à realidade operacional e garantir viabilidade de implementação.

5.4.3 Implementação Piloto

Propõe-se a implementação das melhorias de forma controlada ao longo de 30 dias, contemplando a alocação dinâmica de separadores, a revisão das metas por grupo, a implantação do painel de indicadores e a capacitação dos operadores com feedback contínuo. Durante esse período, recomenda-se o registro sistemático dos indicadores antes e depois da intervenção, possibilitando comparação estatística e avaliação objetiva dos impactos gerados.

5.4.4 Avaliação dos Resultados

A avaliação dos resultados deve ser conduzida por meio da comparação dos indicadores operacionais no período pré e pós-implantação, considerando métricas como percentual de pedidos atendidos dentro da meta, tempo médio de separação, tempo médio por grupo, desempenho médio por separador e tamanho médio da fila no balcão da expedição.

Para validação das diferenças observadas, recomenda-se a aplicação de testes estatísticos adequados à distribuição dos dados, como o teste t para amostras independentes ou o teste não paramétrico de Mann-Whitney, assegurando rigor metodológico na análise dos impactos da intervenção.

5.4.5 Padronização e Expansão

Caso as melhorias implementadas apresentem ganhos estatisticamente comprovados, recomenda-se a formalização dos novos procedimentos operacionais por meio da atualização dos Procedimentos Operacionais Padrão (POP), bem como a automação definitiva do roteamento de pedidos e do painel de indicadores. Adicionalmente, sugere-se a expansão gradual da solução para outras lojas e turnos, garantindo padronização e escalabilidade dos resultados obtidos.

5.5 Considerações Finais do Capítulo

As propostas apresentadas configuram um conjunto de intervenções tecnicamente viáveis, com potencial consistente de impacto sobre a eficiência e a estabilidade do processo analisado. As evidências extraídas das análises estatísticas e dos modelos de Inteligência Artificial indicam que há margem significativa para aprimoramento do desempenho operacional, especialmente por meio do balanceamento da carga de trabalho entre operadores, da redefinição de metas fundamentadas em dados empíricos, da adoção de ações proativas em períodos de pico e da incorporação sistemática de inteligência analítica à rotina decisória.

Tais iniciativas contribuem para a consolidação da expedição como área estratégica dentro do contexto logístico de um Home Center, promovendo maior previsibilidade operacional, elevação do nível de serviço e aprimoramento da experiência do cliente.

6 CONCLUSÃO

O presente trabalho teve como objetivo analisar o processo de separação e atendimento de pedidos na área de expedição de um Home Center, identificando gargalos operacionais e propondo melhorias fundamentadas em análise de dados, métodos estatísticos e técnicas de Inteligência Artificial. A partir da construção de uma base fictícia realista, da realização de análises descritivas e do desenvolvimento de modelos preditivos, foi possível compreender de forma sistêmica a dinâmica do processo e os fatores determinantes para o cumprimento das metas internas.

Os resultados evidenciaram variabilidade operacional significativa, sobretudo nos grupos C e D, que concentram pedidos de maior peso e complexidade. Embora parcela relevante dos atendimentos ocorra dentro do prazo estabelecido, aproximadamente 40% dos pedidos ultrapassam o tempo-meta, indicando desalinhamento entre os parâmetros definidos e o comportamento real do sistema.

As análises demonstraram que o peso total do pedido, a quantidade de itens e a produtividade individual dos separadores constituem os principais determinantes do tempo de atendimento. Tais evidências foram corroboradas pelos modelos de Inteligência Artificial, os quais, ainda que com desempenho moderado, reforçaram o caráter multifatorial do processo e evidenciaram que os fatores que impactam o atraso extrapolam as classificações operacionais atualmente adotadas. Destacou-se, nesse contexto, a relevância do fator humano e das características físicas da carga como elementos centrais na variabilidade do desempenho.

Com base nos achados, foram propostas intervenções estruturadas voltadas ao balanceamento da carga de trabalho, à recalibração dos tempos-meta com base em evidências estatísticas, à implementação de monitoramento operacional em tempo quase real e ao fortalecimento da capacitação técnica dos operadores. Essas

ações foram organizadas em metodologia de implementação composta pelas etapas de diagnóstico, planejamento, execução piloto, avaliação estatística dos resultados e padronização.

De forma geral, o estudo demonstra que a integração entre engenharia de processos, análise de dados e técnicas de Inteligência Artificial constitui abordagem eficaz para aprimoramento de sistemas logísticos e suporte à tomada de decisão. Mesmo com limitações inerentes à modelagem preditiva em ambientes operacionais complexos, os resultados obtidos evidenciam que a adoção de estratégias orientadas por dados contribui para maior previsibilidade, redução de ineficiências e elevação do nível de serviço.

Por fim, conclui-se que as propostas apresentadas possuem potencial para gerar ganhos mensuráveis de produtividade, reduzir tempos de separação e atendimento e ampliar o controle sobre a variabilidade do processo. Como desdobramento futuro, recomenda-se a exploração de modelos preditivos mais avançados, como redes neurais, bem como a incorporação de variáveis operacionais adicionais, tais como ocupação do depósito, intensidade de movimentação diária e fluxo por faixa horária, a fim de ampliar a capacidade explicativa e a acurácia das previsões.

REFERÊNCIAS

- BALLOU, Ronald H. Logística empresarial: transportes, administração de materiais e distribuição física. 1. ed. São Paulo: Atlas, 2006.
- BOWERSOX, Donald J.; CLOSS, David J.; COOPER, M. Bixby. Gestão da cadeia de suprimentos e logística. 4. ed. Rio de Janeiro: Elsevier, 2014.
- BREIMAN, Leo. Random Forests. *Machine Learning*, v. 45, n. 1, p. 5–32, 2001.
- BUSSAB, Wilton O.; MORETTIN, Pedro A. Estatística básica. 9. ed. São Paulo: Saraiva, 2017.
- CHING, H. Y. Gestão de logística: fornecedores, transportes, armazenagem e distribuição. 4. ed. São Paulo: Atlas, 2010.
- CORRÊA, Henrique L.; CORRÊA, Carlos A. Administração de produção e operações: manufatura e serviços. 3. ed. São Paulo: Atlas, 2012.
- DAVENPORT, Thomas. *Competing on analytics: The new science of winning*. Boston: Harvard Business School Press, 2012.
- GIL, Antônio Carlos. Como elaborar projetos de pesquisa. 6. ed. São Paulo: Atlas, 2019.
- HAN, Jiawei; KAMBER, Micheline; PEI, Jian. *Data Mining: conceitos e técnicas*. 3. ed. Rio de Janeiro: Elsevier, 2012.
- MARTINS, Petrônio G.; LAUGENI, Fernando P. Administração da produção. 2. ed. São Paulo: Saraiva, 2005.
- MARINHO, José Carlos. Probabilidade e Estatística. 2. ed. São Paulo: Editora Erica, 2014.
- MITCHELL, Tom M. *Machine Learning*. New York: McGraw-Hill, 2010.
- JAMES, Gareth; WITTEN, Daniela; HASTIE, Trevor; TIBSHIRANI, Robert. *Introdução à Aprendizagem Estatística: com Aplicações em R*. 1. ed. Rio de Janeiro: LTC, 2019.
- MONTGOMERY, Douglas C.; RUNGER, George C. Estatística aplicada e probabilidade para engenheiros. 6. ed. Rio de Janeiro: LTC, 2018.
- NOVAES, Antonio Galvão. Logística e gerenciamento da cadeia de distribuição: estratégia, operação e avaliação. 4. ed. Rio de Janeiro: Elsevier, 2007.
- PEINADO, Jurandir; GRAEML, Alexandre R. Administração da produção: operações industriais e de serviços. Curitiba: UnicenP, 2007.
- SANTOS, Jéssica P.; SILVA, K. A.; PEREIRA, L. F. Aplicações de inteligência artificial em processos logísticos. *Revista Produção Online*, v. 20, n. 2, p. 455–478, 2020.
- SLACK, Nigel; CHAMBERS, Stuart; JOHNSTON, Robert. Administração da Produção. 3. ed. São Paulo: Atlas, 2009.

TRIOLA, Mario F. Introdução à Estatística. 12. ed. Rio de Janeiro: LTC, 2018.

VIANA, João José C. Administração de materiais: um enfoque prático. 2. ed. São Paulo: Atlas, 2012.

WANKE, Peter. Gestão de operações e logística. São Paulo: Atlas, 2012.

LOGÍSTICA & MOBILIDADE. Logística de mobilidade: o que é suas etapas principais. Disponível em:

<https://logisticaemobilidade.com.br/gestao/logistica-de-armazenagem-o-que-e/>

Acesso em: 28 fev. 2026.

APÊNDICE A — Código para Geração da Base de Dados Estruturada

O código abaixo apresenta o script utilizado para gerar a base de dados estruturada empregada nas análises estatísticas e nos modelos de Inteligência Artificial deste trabalho. A base foi construída de forma a simular o processo real de separação e atendimento de pedidos na expedição de um Home Center, considerando regras operacionais, metas internas e variabilidade realista dos tempos de processo.

```
import numpy as np
import pandas as pd
import random
from datetime import datetime, timedelta

# -----
# CONFIGURAÇÕES
# -----
NUM_LINHAS = 3000 # Você pode ajustar aqui

# Probabilidade realista dos tipos de entrega
TIPOS_ENTREGA = ["Imediata", "Posterior", "Agendada"]
PROB_ENTREGA = [0.55, 0.25, 0.20]

# Turnos
TURNOS = ["Manha", "Tarde", "Noite"]
PROB_TURNOS = [0.45, 0.45, 0.10]

# -----
# FUNÇÃO PARA CLASSIFICAR O GRUPO (A, B, C, D ou Agendada)
# -----
def classificar_grupo(tipo_entrega, peso, itens, contem_pisos):

    if tipo_entrega == "Agendada":
        return "Agendada"

    if peso <= 200 and itens <= 3 and contem_pisos == "Nao":
```

```

    return "A"
elif 201 <= peso <= 500:
    return "B"
elif 501 <= peso <= 2000:
    return "C"
else:
    return "D"

# -----
# METAS POR GRUPO
# -----
METAS = {
    "A": {"sep": 5, "atend": 6},
    "B": {"sep": 7, "atend": 9},
    "C": {"sep": 15, "atend": 20},
    "D": {"sep": 22, "atend": 30},
    "Agendada": {"sep": 5, "atend": 6},
}

# -----
# FUNÇÃO PARA GERAR TEMPOS REAIS (SIMULAÇÃO)
# -----
def gerar_tempos(grupo):
    meta_sep = METAS[grupo]["sep"]
    meta_at = METAS[grupo]["atend"]

    tempo_sep_real = np.random.normal(loc=meta_sep, scale=1.5)
    tempo_sep_real = max(1, round(tempo_sep_real))

    tempo_conf_real = meta_at - meta_sep
    tempo_conf_real = max(1, round(np.random.normal(loc=tempo_conf_real,
scale=1)))

    tempo_at_real = tempo_sep_real + tempo_conf_real

```



```
separador_id = random.randint(101, 150)
turno = np.random.choice(TURNOS, p=PROB_TURNOS)
```

```
# Dia da semana em português
```

```
dias_semana_pt = {
    "Monday": "Segunda",
    "Tuesday": "Terca",
    "Wednesday": "Quarta",
    "Thursday": "Quinta",
    "Friday": "Sexta",
    "Saturday": "Sabado",
    "Sunday": "Domingo"
}
```

```
dia_semana_en = data_pedido.strftime("%A")
dia_semana = dias_semana_pt[dia_semana_en]
```

```
dados.append([
    i + 1,
    data_pedido.strftime("%Y-%m-%d %H:%M:%S"),
    tipo_entrega,
    itens,
    peso,
    contem_pisos,
    grupo,
    t_sep_real,
    t_conf_real,
    t_atend_real,
    meta_sep,
    meta_at,
    atendeu_separacao,
    atendeu_atendimento,
```

```
        separador_id,
        turno,
        dia_semana
    ])

# -----
# CRIANDO DATAFRAME
# -----
colunas = [
    "Pedido_ID", "Data_Pedido", "Tipo_Entrega",
    "Qtde_Itens", "Peso_Total_kg", "Contem_Pisos",
    "Grupo_Classificacao",
    "Tempo_Separacao_min", "Tempo_Conferencia_min",
    "Tempo_Atendimento_min",
    "Meta_Separacao_min", "Meta_Atendimento_min",
    "Atendeu_Meta_Separacao", "Atendeu_Meta_Atendimento",
    "Separador_ID", "Turno", "Dia_Semana"
]

df = pd.DataFrame(dados, columns=colunas)

# -----
# SALVANDO EM EXCEL
# -----
df.to_excel("base_ficticia.xlsx", index=False)

df.head()
```

APÊNDICE B — Código de Análise Estatística

O código a seguir apresenta os procedimentos utilizados para realizar a análise estatística descritiva da base de dados estruturada. As análises incluem distribuição das variáveis operacionais, criação de indicadores de atraso, geração de gráficos exploratórios e cálculos de desempenho por grupo, tipo de entrega e turno. Esse conjunto de procedimentos serviu como base para a interpretação inicial dos dados e identificação dos principais padrões operacionais discutidos no Capítulo 4.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Configurações de estilo
sns.set(style="whitegrid")
plt.rcParams["figure.figsize"] = (10, 5)
plt.rcParams["axes.titlesize"] = 14

# Carregar base
df = pd.read_csv("base_ficticia.csv")
df.head()

# -----
# Criar campos de atraso (Separação e atendimento)
df["Atraso_Separacao"] = df["Tempo_Separacao_min"] - df["Meta_Separacao_min"]
df["Atraso_Atendimento"] = df["Tempo_Atendimento_min"] - df["Meta_Atendimento_min"]

df["Atrasou_Separacao"] = df["Atraso_Separacao"] > 0
df["Atrasou_Atendimento"] = df["Atraso_Atendimento"] > 0

# -----
# Frequência dos tipos de Entregas
plt.figure()
df["Tipo_Entrega"].value_counts().plot(kind="bar")
plt.title("Distribuição dos Tipos de Entrega")
plt.xlabel("Tipo de Entrega")
plt.ylabel("Quantidade")
```

```
plt.show()

# -----
# Frequência dos Grupos (A, B, C, D, Agendada)
plt.figure()
df["Grupo_Classificacao"].value_counts().plot(kind="bar", color="gray")
plt.title("Distribuição dos Grupos de Classificação")
plt.xlabel("Grupo")
plt.ylabel("Quantidade")
plt.show()

# -----
# Distribuição do Peso Total (kg)
plt.figure()
sns.histplot(df["Peso_Total_kg"], bins=40, kde=True)
plt.title("Distribuição do Peso Total (kg)")
plt.xlabel("Peso (kg)")
plt.ylabel("Frequência")
plt.show()

# -----
# Distribuição da Quantidade de Itens
plt.figure()
sns.histplot(df["Qtde_Itens"], bins=20, kde=True)
plt.title("Distribuição da Quantidade de Itens")
plt.xlabel("Itens")
plt.ylabel("Frequência")
plt.show()

# -----
# Boxplot — Tempo de Separação por Grupo
plt.figure()
sns.boxplot(data=df, x="Grupo_Classificacao", y="Tempo_Separacao_min")
plt.title("Tempo de Separação por Grupo")
plt.xlabel("Grupo")
plt.ylabel("Tempo (min)")
```

```

plt.show()

# -----
# Boxplot — Tempo de Atendimento por Grupo
plt.figure()
sns.boxplot(data=df, x="Grupo_Classificacao", y="Tempo_Atendimento_min")
plt.title("Tempo de Atendimento por Grupo")
plt.xlabel("Grupo")
plt.ylabel("Tempo (min)")
plt.show()

# -----
# Comparação Tempo Real vs Meta – Atendimento
df_plot = df.groupby("Grupo_Classificacao")[
    ["Tempo_Atendimento_min", "Meta_Atendimento_min"]].mean()

df_plot.plot(kind="bar")
plt.title("Tempo Médio Real vs Meta — Atendimento")
plt.ylabel("Minutos")
plt.show()

# -----
# Percentual que atendeu a meta (Geral)
print("Percentual que atendeu a meta de separação:")
print(df["Atendeu_Meta_Separacao"].value_counts(normalize=True) * 100)

print("\nPercentual que atendeu a meta de atendimento:")
print(df["Atendeu_Meta_Atendimento"].value_counts(normalize=True) * 100)

# -----
# Percentual de atendimento por grupo
pct_grupo = df.groupby("Grupo_Classificacao")[
    "Atendeu_Meta_Atendimento"].value_counts(normalize=True).unstack() * 100
print(pct_grupo)

# -----

```

```

# Atraso por grupo (média)
df.groupby("Grupo_Classificacao")[
    ["Atraso_Separacao", "Atraso_Atendimento"]].mean()

# -----
# Atraso por tipo de entrega
df.groupby("Tipo_Entrega")[["Atraso_Separacao", "Atraso_Atendimento"]].mean()

# -----
# Atraso por turno
df.groupby("Turno")[["Atraso_Separacao", "Atraso_Atendimento"]].mean()

# -----
# Relação Peso x Atraso (dispersão)
plt.figure()
sns.scatterplot(data=df, x="Peso_Total_kg", y="Atraso_Atendimento")
plt.title("Peso Total x Atraso no Atendimento")
plt.xlabel("Peso (kg)")
plt.ylabel("Atraso (min)")
plt.show()

# -----
# Relação Itens x Atraso
plt.figure()
sns.scatterplot(data=df, x="Qtde_Itens", y="Atraso_Atendimento")
plt.title("Quantidade de Itens x Atraso no Atendimento")
plt.xlabel("Itens")
plt.ylabel("Atraso (min)")
plt.show()

# -----
# Heatmap — Grupo vs Tipo de Entrega (percentual de atraso)
tabela_heat = df.groupby(["Grupo_Classificacao", "Tipo_Entrega"])[
    "Atraso_Atendimento"].mean().unstack() * 100

plt.figure()

```

```
sns.heatmap(tabela_heat, annot=True, fmt=".1f")  
plt.title("Percentual de Atraso — Grupo x Tipo de Entrega")  
plt.show()
```

APÊNDICE C — Código dos Modelos de Inteligência Artificial

Este apêndice apresenta o script utilizado para construção, treinamento e avaliação dos modelos de Inteligência Artificial aplicados à previsão de atrasos no atendimento. Foram utilizados dois modelos supervisionados amplamente utilizados na literatura: Decision Tree e Random Forest. O código também inclui tratamento das variáveis, divisão dos dados, avaliação das métricas e salvamento dos modelos treinados.

```
# =====
# 4.4 - MODELOS DE INTELIGÊNCIA ARTIFICIAL
# CLASSIFICAÇÃO: PREVER SE O PEDIDO VAI ATRASAR
# =====

# -----
# IMPORTANDO BIBLIOTECAS
# -----
import pandas as pd
import numpy as np

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import (
    accuracy_score,
    confusion_matrix,
    classification_report
)

import joblib
import seaborn as sns
import matplotlib.pyplot as plt

# -----
# 1) CARREGAR BASE DE DADOS
# -----
df = pd.read_csv("base_ficticia.csv")
```

```

# Criar variável de atraso
df["Atraso_Atendimento"] = df["Tempo_Atendimento_min"] - df["Meta_Atendimento_min"]
df["Atrasou"] = (df["Atraso_Atendimento"] > 0).astype(int)

# -----
# 2) SELECIONAR FEATURES (X) E TARGET (y)
# -----
features = [
    "Tipo_Entrega",
    "Qtde_Itens",
    "Peso_Total_kg",
    "Contem_Pisos",
    "Grupo_Classificacao",
    "Separador_ID",
    "Turno",
    "Dia_Semana"
]

X = df[features].copy()
y = df["Atrasou"].copy()

# -----
# 3) LABEL ENCODING SEGURO (SEM DATA LEAKAGE)
# -----
label_cols = ["Tipo_Entrega", "Contem_Pisos", "Grupo_Classificacao", "Turno",
              "Dia_Semana"]

encoders = {}

for col in label_cols:
    le = LabelEncoder()
    X[col] = le.fit_transform(X[col])
    encoders[col] = le

# -----

```

```
# 4) DIVISÃO TREINO/TESTE
# -----
X_train, X_test, y_train, y_test = train_test_split(
    X, y,
    test_size=0.30,
    random_state=42,
    stratify=y
)

# -----
# 5) TREINAR MODELOS
# -----

# Árvore de Decisão
tree = DecisionTreeClassifier(max_depth=5, random_state=42)
tree.fit(X_train, y_train)

# Random Forest
rf = RandomForestClassifier(
    n_estimators=300,
    max_depth=None,
    random_state=42,
    class_weight="balanced"
)
rf.fit(X_train, y_train)

# -----
# 6) AVALIAÇÃO DOS MODELOS
# -----

pred_tree = tree.predict(X_test)
pred_rf = rf.predict(X_test)

# Acurácia
acc_tree = accuracy_score(y_test, pred_tree)
acc_rf = accuracy_score(y_test, pred_rf)
```

```

print("\n=== ACURÁCIA ===")
print(f"Decision Tree: {acc_tree:.4f}")
print(f"Random Forest: {acc_rf:.4f}")

print("\n=== CLASSIFICATION REPORT (TREE) ===")
print(classification_report(y_test, pred_tree))

print("\n=== CLASSIFICATION REPORT (RANDOM FOREST) ===")
print(classification_report(y_test, pred_rf))

# -----

# MODELO SIMPLES + MATRIZ DE CONFUSÃO

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import confusion_matrix
import numpy as np

# -----
# Criar variável alvo (1 = atrasou, 0 = no prazo)
# -----
df["Atrasou"] = (df["Tempo_Atendimento_min"] > df["Meta_Atendimento_min"]).astype(int)

# -----
# Selecionar variáveis
# -----
features = [
    "Tipo_Entrega",
    "Qtde_Itens",
    "Peso_Total_kg",
    "Contem_Pisos",
    "Grupo_Classificacao",
    "Separador_ID",
    "Turno",

```

```
    "Dia_Semana"
]

X = df[features].copy()
y = df["Atrasou"].copy()

# -----
# Transformar variáveis categóricas
# -----

label_cols = ["Tipo_Entrega", "Contem_Pisos", "Grupo_Classificacao", "Turno",
"Dia_Semana"]

for col in label_cols:
    le = LabelEncoder()
    X[col] = le.fit_transform(X[col])

# -----
# Divisão treino / teste
# -----
X_train, X_test, y_train, y_test = train_test_split(
    X, y,
    test_size=0.30,
    random_state=42,
    stratify=y
)

# -----
# Treinar modelo
# -----
rf = RandomForestClassifier(
    n_estimators=200,
    random_state=42,
    class_weight="balanced"
)

rf.fit(X_train, y_train)
```

```

# -----
# Previsões
# -----
pred_rf = rf.predict(X_test)

# -----
# MATRIZ DE CONFUSÃO
# -----
cm = confusion_matrix(y_test, pred_rf)

plt.figure(figsize=(6,5))
sns.heatmap(
    cm,
    annot=True,
    fmt="d",
    cmap="Blues",
    xticklabels=["No Prazo", "Atrasou"],
    yticklabels=["No Prazo", "Atrasou"]
)

plt.xlabel("Classe Predita")
plt.ylabel("Classe Real")
plt.title("Matriz de Confusão - Random Forest")
plt.show()

# -----
# 7) IMPORTÂNCIA DAS VARIÁVEIS
# -----
importancias = pd.DataFrame({
    "Variável": X.columns,
    "Importância": rf.feature_importances_
}).sort_values(by="Importância", ascending=False)

print("\n=== IMPORTÂNCIA DAS VARIÁVEIS ===")
print(importancias)

```

```
importancias.to_csv("importancia_variaveis.csv", index=False)
```

```
# -----
```

```
# 8) SALVAR MODELOS EM .PKL
```

```
# -----
```

```
joblib.dump(rf, "modelo_random_forest.pkl")
```

```
joblib.dump(tree, "modelo_decision_tree.pkl")
```

```
joblib.dump(encoders, "encoders_categorias.pkl")
```

```
print("\nModelos e encoders salvos com sucesso!")
```