



UNIVERSIDADE DE BRASÍLIA
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, CONTABILIDADE E
GESTÃO DE POLÍTICAS PÚBLICAS - FACE
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO – PPGA

THAÍS FERREIRA LOPES OLIVEIRA

RECLAMAÇÕES E JUDICIALIZAÇÃO POR CONSUMIDORES DO
TRANSPORTE AÉREO BRASILEIRO: UM DIAGNÓSTICO COM
APRENDIZADO DE MÁQUINA

BRASÍLIA
2025

THAÍS FERREIRA LOPES OLIVEIRA

**RECLAMAÇÕES E JUDICIALIZAÇÃO POR CONSUMIDORES DO
TRANSPORTE AÉREO BRASILEIRO: UM DIAGNÓSTICO COM
APRENDIZADO DE MÁQUINA**

Dissertação apresentada como requisito para a obtenção do título de Mestre no Programa de Pós-Graduação em Administração da Universidade de Brasília.

Área de concentração: Finanças e Métodos Quantitativos

Orientador: Prof. Dr. Victor Rafael Rezende Celestino

**BRASÍLIA
2025**

THAÍS FERREIRA LOPES OLIVEIRA

RECLAMAÇÕES E JUDICIALIZAÇÃO POR CONSUMIDORES DO TRANSPORTE
AÉREO BRASILEIRO: UM DIAGNÓSTICO COM APRENDIZADO DE MÁQUINA

Dissertação apresentada como requisito para a obtenção do título de Mestre no Programa de Pós-Graduação em Administração da Universidade de Brasília.

Área de concentração: Finanças e Métodos Quantitativos

Data de defesa: 22/01/2025.

BANCA EXAMINADORA

Prof. Dr. Victor Rafael Rezende Celestino
Orientador

Profa. Dra. Eluiza Alberto de Moraes Watanabe
Examinadora interna (PPGA/UnB)

Prof. Dr. Marcelo Xavier Guterres
Examinador externo (ITA)

Dedico este trabalho a todos que se dedicam ao ensino e à pesquisa, acreditando no poder transformador da educação e da ciência.

AGRADECIMENTOS

Primeiramente, agradeço a Deus por me conceder o dom da vida e por me guiar a viver cada etapa no tempo certo. Todo o percurso do mestrado foi repleto de desafios, tanto no âmbito acadêmico quanto profissional e pessoal, mas foi Ele quem me sustentou e me possibilitou concluir mais essa jornada.

Aos meus pais e irmãos, Luiz, Georgina, Rayanne e Thales, que, apesar de não entenderem profundamente o tema, me incentivaram e me apoiaram durante todo o mestrado. E ao meu marido, Gustavo, que esteve ao meu lado de maneira mais próxima na construção dessa pesquisa, ouvindo-me nos momentos de desânimo, me dando forças e ajudando-me a enxergar o meu potencial.

Aos meus amigos, que estiveram comigo em todos os momentos, sendo fundamentais para os momentos de descanso e para torcer pelo meu sucesso.

Aos meus colegas de turma e de projeto de pesquisa, que enriqueceram minha experiência com a troca constante de conhecimentos sobre o tema.

A todos os professores que, ao longo desses dois anos, contribuíram para minha formação profissional e pessoal. Em especial, ao professor Victor Rafael, que me orientou na conclusão deste trabalho.

“Que a tua vida não seja uma vida estéril. - Sê útil. - Deixa rasto.”
(Caminho, 1 - São Josemaria Escrivá)

RESUMO

O setor de transporte aéreo é essencial para o desenvolvimento econômico, impulsionando o emprego, o turismo e a conectividade global. No entanto, a pandemia de COVID-19 em 2020 trouxe desafios significativos, resultando em uma redução no número de voos e passageiros. Em contrapartida, houve um aumento nas reclamações de consumidores e litígios judiciais contra companhias aéreas no Brasil entre 2019 e 2023. Diante disso, este estudo investigou as características temporais, geográficas, por empresas e motivos das reclamações e processos judiciais contra seis companhias aéreas que operam no Brasil, analisando a existência de correlação entre os motivos de reclamações e processos e o efeito no desempenho dos modelos de aprendizado de máquina elaborados em um trabalho anterior, após a padronização dos motivos e causas dos processos. Os resultados mostraram que, enquanto as reclamações foram predominantemente motivadas por questões contratuais, os processos judiciais tiveram como principal causa problemas operacionais ao longo do tempo. Além disso, foi identificada uma forte correlação positiva entre reclamações e processos judiciais relacionados a bagagem, contrato e problemas operacionais. Em relação aos modelos de aprendizado de máquina, observou-se uma pequena queda no desempenho após a padronização, porém os modelos continuaram com métricas altas, com destaque para o modelo *Random Forest*. Dessa forma, este estudo contribui para o entendimento das dinâmicas entre reclamações de consumidores e litígios judiciais no setor aéreo brasileiro, ampliando a literatura sobre o tema, além de fornecer um recurso valioso para as companhias desenvolverem estratégias eficazes de resolução de problemas.

Palavras-chave: consumidor, reclamações, judicialização, transporte aéreo, aprendizado de máquina.

ABSTRACT

The air transport sector is essential for economic development, driving employment, tourism, and global connectivity. However, the COVID-19 pandemic in 2020 brought significant challenges, leading to a reduction in the number of flights and passengers. Conversely, there was an increase in consumer complaints and legal disputes against airlines in Brazil between 2019 and 2023. Given this, this study investigated the temporal, geographical, company-specific, and reason-based characteristics of complaints and lawsuits against six airlines operating in Brazil, analyzing the correlation between the reasons for complaints and lawsuits and the effect on the performance of machine learning models developed in a previous study after standardizing the reasons and causes of lawsuits. The results showed that while complaints were predominantly motivated by contractual issues, lawsuits were primarily caused by operational problems over time. Additionally, a strong positive correlation was identified between complaints and lawsuits related to baggage, contracts, and operational issues. Regarding the machine learning models, a slight drop in performance was observed after standardization; however, the models still maintained high metrics, with the *Random Forest* model standing out. Thus, this study contributes to the understanding of the dynamics between consumer complaints and legal disputes in the Brazilian airline sector, expanding the literature on the topic while also providing a valuable resource for airlines to develop effective problem-resolution strategies.

Key-words: consumer, complaints, judicialization, air transport, machine learning.

LISTA DE ILUSTRAÇÕES

Figura 1 – Gastos em milhões (R\$MM) com assistência a passageiros, indenizações extrajudiciais e condenações judiciais	35
Figura 2 – Evolução anual do número de passageiros pagos - mercados doméstico e internacional	45
Figura 3 – Evolução da oferta (ASK), da demanda (RPK) e do aproveitamento (LF) - mercados doméstico e internacional	46
Figura 4 – Histogramas e QQ-plots das Variáveis	62
Figura 5 – Quantidade de Reclamações por Ano de Abertura	76
Figura 6 – Quantidade de Reclamações por Região	78
Figura 7 – Quantidade de Reclamações por Região e Ano	79
Figura 8 – Quantidade de Reclamações por Empresa Aérea	81
Figura 9 – Quantidade de Reclamações por Empresa Aérea e Ano	82
Figura 10 – Quantidade de Reclamações por Empresa Aérea e Região	83
Figura 11 – Quantidade de Reclamações por Categoria de Motivo	84
Figura 12 – Quantidade de Reclamações por Categoria de Motivo e Região	85
Figura 13 – Quantidade de Reclamações por Categoria de Motivo e Empresa Aérea	86
Figura 14 – Quantidade de Processos Judiciais por Ano de Abertura	93
Figura 15 – Intervalos de Tempo entre a Data do Voo e a Data de Distribuição	95
Figura 16 – Quantidade de Processos Judiciais por Região	96
Figura 17 – Quantidade de Processos Judiciais por Região e Ano	97
Figura 18 – Quantidade de Processos Judiciais por Empresa Aérea	99
Figura 19 – Quantidade de Processos Judiciais por Empresa Aérea e Ano	100
Figura 20 – Quantidade de Processos Judiciais por Empresa Aérea e Região	101
Figura 21 – Quantidade de Processos Judiciais por Categoria de Motivo	102
Figura 22 – Quantidade de Processos Judiciais por Categoria de Motivo e Região	103
Figura 23 – Quantidade de Processos Judiciais por Categoria de Motivo e Empresa Aérea	104
Figura 24 – Quantidade de Processos Judiciais por Categoria de Causa	106
Figura 25 – Quantidade de Processos Judiciais por Categoria de Causa e Região	108
Figura 26 – Quantidade de Processos Judiciais por Categoria de Causa e Empresa Aérea	109
Figura 27 – Valor Máximo do Pedido Inicial dos Processos Judiciais por Região e Ano	112
Figura 28 – Valor Máximo do Pedido Inicial dos Processos Judiciais por Categoria de Motivo e Ano	113
Figura 29 – Valor Máximo do Dano Moral dos Processos Judiciais por Região e Ano	114

Figura 30 – Valor Máximo do Dano Moral dos Processos Judiciais por Categoria de Motivo e Ano	115
Figura 31 – Valor Máximo do Dano Material dos Processos Judiciais por Região e Ano	116
Figura 32 – Valor Máximo do Dano Material dos Processos Judiciais por Categoria de Motivo e Ano	117
Figura 33 – Quantidade de Processos Judiciais por Situação Processual e Sentença e Ano	118
Figura 34 – Valor Máximo do Total de Condenação dos Processos Judiciais por Região e Ano	120
Figura 35 – Valor Máximo do Total de Condenação dos Processos Judiciais por Categoria de Motivo e Causa	121
Figura 36 – Valor Máximo do Total de Condenação dos Processos Judiciais por Jurisdição e Situação Processual e Sentença	122
Figura 37 – Valor Máximo Pago dos Processos Judiciais por Região e Ano	123
Figura 38 – Valor Máximo Pago dos Processos Judiciais por Categoria de Motivo e Causa	124
Figura 39 – Valor Máximo Pago dos Processos Judiciais por Jurisdição e Situação Processual e Sentença	125
Figura 40 – Matriz de Correlação entre as Quantidades dos Motivos de Reclamações e Processos	126
Figura 41 – Importância das variáveis no Modelo NB	135
Figura 42 – Importância das variáveis no Modelo SVM Linear	136
Figura 43 – Importância das variáveis no Modelo RF	138
Figura 44 – Subcampos da Inteligência Artificial	148
Figura 45 – Fluxograma do funcionamento do Consumidor.gov.br e atuação da ANAC ¹⁵¹	

LISTA DE TABELAS

Tabela 1 – Variáveis das bases de dados do Consumidor.gov.br	47
Tabela 2 – Variáveis das bases de dados de judicialização	49
Tabela 3 – Resultados do Teste de Shapiro-Wilk	61
Tabela 4 – Intervalos das variáveis de indenização em reais (R\$)	65
Tabela 5 – Variáveis	66
Tabela 6 – Matriz de Confusão	71
Tabela 7 – Fórmulas das métricas de avaliação de classificação.	71
Tabela 8 – Reclamações: Estatísticas Descritivas - Variáveis Categóricas	74
Tabela 9 – Reclamações: Estatísticas Descritivas - Variáveis Numéricas	74
Tabela 10 – Índice de Reclamações por 100 mil Passageiros por Ano	77
Tabela 11 – Quantidade de Reclamações por Mês e Ano de Abertura	77
Tabela 12 – Quantidade de Reclamações por UF e Ano	80
Tabela 13 – Quantidade de Reclamações por Categoria de Motivo e Ano	84
Tabela 14 – Top 5 das Reclamações por Código ANAC e Ano	88
Tabela 15 – Judicialização: Estatísticas Descritivas - Variáveis Categóricas	91
Tabela 16 – Judicialização: Estatísticas Descritivas - Variáveis Numéricas	91
Tabela 17 – Índice de Processos Judiciais e Reclamações por 100 mil Passageiros por Ano	93
Tabela 18 – Quantidade de Processos Judiciais por Mês e Ano de Abertura	94
Tabela 19 – Quantidade de Processos Judiciais por UF e Ano	98
Tabela 20 – Quantidade de Processos Judiciais por Categoria de Motivo e Ano	102
Tabela 21 – Top 5 dos Processos Judiciais por Objeto, Subobjeto e Ano	105
Tabela 22 – Quantidade de Processos Judiciais por Categoria de Causa e Ano	107
Tabela 23 – Top 5 dos Processos Judiciais por Causa e Ano	110
Tabela 24 – Judicialização: Estatísticas Descritivas - Valores de Condenação e Paga- mento	119
Tabela 25 – Desempenho global dos modelos de aprendizado de máquina	128
Tabela 26 – Comparação do desempenho dos modelos de aprendizado de máquina	129
Tabela 27 – Matriz de Confusão - Naive Bayes	130
Tabela 28 – Matriz de Confusão - SVM Linear	130
Tabela 29 – Matriz de Confusão - Random Forest	131
Tabela 30 – Métricas da classe Baixo	131
Tabela 31 – Métricas da classe Médio	131
Tabela 32 – Métricas da classe Alto	131
Tabela 33 – Correlação entre Variáveis	134

LISTA DE QUADROS

Quadro 1 – Quadro resumo dos trabalhos de aplicações do aprendizado de máquina nas questões de consumidores do transporte aéreo	27
Quadro 2 – Quadro resumo dos trabalhos de aplicações do aprendizado de máquina nas questões de consumidores do transporte aéreo e pandemia de COVID-19	31
Quadro 3 – Quadro resumo dos trabalhos de aplicações do aprendizado de máquina nas questões de judicialização do transporte aéreo	41
Quadro 4 – Técnicas de análise de dados	51
Quadro 5 – Variáveis comuns nas duas bases de dados	52
Quadro 6 – Lista de palavras-chave por categoria para agrupamento dos motivos .	55
Quadro 7 – Lista de palavras-chave por categoria para agrupamento das causas . .	56
Quadro 8 – Tipos de dados e quantidade de valores ausentes da base final de judicialização	57
Quadro 9 – Tipos de dados e quantidade de valores ausentes da base final de reclamações	59
Quadro 10 – Taxonomia de IA	146

LISTA DE ABREVIATURAS E SIGLAS

ABEAR	Associação Brasileira das Empresas Aéreas
AED	Análise Exploratória de Dados
AI	<i>Artificial Intelligence</i>
AMB	Associação dos Magistrados Brasileiros
ALTA	Associação Latino-Americana e do Caribe de Transporte Aéreo
ANAC	Agência Nacional de Aviação Civil
art.	Artigo
ASK	<i>Available Seats-Kilometers</i>
CBA	Código Brasileiro de Aeronáutica
CDC	Código de Defesa do Consumidor
CNJ	Conselho Nacional de Justiça
DL	<i>Deep Learning</i>
DS	<i>Design Science</i>
DS	<i>Design Science Research</i>
FGV-SP	Fundação Getulio Vargas de São Paulo
IA	Inteligência Artificial
IATA	<i>International Air Transport Association</i>
IBAER	Instituto Brasileiro de Direito Aeronáutico
JEC/UFSC	Juizado Especial Cível da Universidade Federal de Santa Catarina
JURCAIB	Junta de Representantes das Companhias Aéreas Internacionais do Brasil
LDA	<i>Latent Dirichlet Allocation</i>
LF	<i>Load Factor</i>
LLM	<i>Large Language Models</i>

LSA	<i>Latent Semantic Analysis</i>
ML	<i>Machine Learning</i>
MLP	<i>Multiple Layer Perceptron</i>
MOSTA	<i>Multi-view Overlapping cluSTering of legAl judgments</i>
OAB	Ordem dos Advogados do Brasil
OE	Objetivo Específico
PIB	Produto Interno Bruto
PCA	<i>Principal Component Analysis</i>
PUCRS	Pontifícia Universidade Católica do Rio Grande do Sul
RF	<i>Random Forest</i>
RNA	Redes Neurais Artificiais
RPK	<i>Revenue Passenger-Kilometres</i>
Senacon	Secretaria Nacional do Consumidor
STM	<i>Structural Topic Model</i>
SVM	<i>Support Vector Machine</i>
UnB	Universidade de Brasília

SUMÁRIO

1	INTRODUÇÃO	16
1.1	Formulação do problema	17
1.2	Pergunta de pesquisa	19
1.3	Objetivos de pesquisa	20
1.3.1	<i>Objetivo geral</i>	20
1.3.2	<i>Objetivos específicos</i>	20
1.4	Justificativa da pesquisa	20
1.5	Estrutura do trabalho	21
2	REFERENCIAL TEÓRICO	22
2.1	Reclamações contra empresas de transporte aéreo	22
2.1.1	<i>Aplicação do aprendizado de máquina em reclamações e avaliações de clientes de transporte aéreo</i>	23
2.1.1.1	<i>Reclamações e avaliações de clientes do transporte aéreo no contexto da pandemia de COVID-19</i>	28
2.2	Judicialização do transporte aéreo	32
2.2.1	<i>Contextualização da judicialização do transporte aéreo no Brasil</i>	32
2.2.2	<i>Aplicação do aprendizado de máquina na judicialização do transporte aéreo</i>	36
3	MÉTODOS E TÉCNICAS DE PESQUISA	42
3.1	Tipologia e descrição geral dos métodos de pesquisa	42
3.1.1	<i>Design Science Research</i>	42
3.2	Procedimentos técnicos para a realização da pesquisa	43
3.3	Caracterização do objeto de estudo	44
3.4	População e amostra da pesquisa	46
3.5	Procedimento de coleta de dados	47
3.5.1	<i>Dados de reclamações de consumidores de companhias aéreas</i>	47
3.5.2	<i>Dados de judicialização do transporte aéreo</i>	49
3.6	Procedimentos de análise de dados	51
3.6.1	<i>Análise exploratória de dados</i>	51
3.6.1.1	<i>Coleta e organização de dados</i>	51
3.6.1.2	<i>Tratamento de dados</i>	53
3.6.1.2.1	<i>Tratamento dos dados de judicialização</i>	53
3.6.1.2.2	<i>Tratamento dos dados do Consumidor.gov.br</i>	58
3.6.1.3	<i>Análise de dados</i>	59
3.6.1.4	<i>Apresentação e interpretação de dados</i>	60
3.6.2	<i>Análise de correlação</i>	60

3.6.3	<i>Desenvolvimento dos modelos de aprendizado de máquina . . .</i>	63
3.6.3.1	<i>Seleção das variáveis para os modelos</i>	64
3.6.3.2	<i>Preparação dos dados para os modelos</i>	67
3.6.3.3	<i>Algoritmos de aprendizado de máquina</i>	68
3.6.3.4	<i>Naive Bayes Classifier</i>	69
3.6.3.5	<i>Support Vector Machine</i>	69
3.6.3.6	<i>Random Forest Classifier</i>	70
3.6.3.7	<i>Avaliação das métricas de desempenho dos modelos</i>	70
3.6.3.8	<i>Importância das variáveis em cada modelo</i>	72
4	RESULTADOS E DISCUSSÃO	73
4.1	Análise Exploratória dos Dados	73
4.1.1	<i>Dados de reclamações de consumidores de companhias aéreas</i>	73
4.1.1.1	<i>Visão por Ano e Mês</i>	75
4.1.1.2	<i>Visão por Região e UF</i>	78
4.1.1.3	<i>Visão por Empresa Aérea</i>	80
4.1.1.4	<i>Motivos das reclamações</i>	83
4.1.2	<i>Dados de judicialização do transporte aéreo</i>	90
4.1.2.1	<i>Visão por Ano e Mês</i>	92
4.1.2.2	<i>Visão por Região e UF</i>	96
4.1.2.3	<i>Visão por Empresa Aérea</i>	98
4.1.2.4	<i>Motivos dos processos judiciais</i>	101
4.1.2.5	<i>Causas dos processos judiciais</i>	106
4.1.2.6	<i>Indenizações dos processos judiciais</i>	111
4.2	Correlação entre as reclamações e os processos judiciais contra companhias aéreas	126
4.3	Resultados dos modelos de aprendizado de máquina para previsão do valor total de condenação	128
4.3.1	<i>Matrizes de confusão</i>	130
4.3.2	<i>Importância das variáveis</i>	133
4.4	Discussão dos resultados	139
5	CONSIDERAÇÕES FINAIS	142
APÊNDICE A	INTELIGÊNCIA ARTIFICIAL	146
A.0.1	<i>Aprendizado de Máquina</i>	147
APÊNDICE B	CONSUMIDOR.GOV.BR	150
B.1	Adesão do Consumidor.gov.br pela ANAC	151
	REFERÊNCIAS	153

1 INTRODUÇÃO

O setor de transporte aéreo é estratégico para o desenvolvimento e crescimento econômico de um país, visto que ele cria novas oportunidades de emprego, aumenta o número de turistas, transfere mercadorias através de carga aérea de forma mais rápida, melhora as oportunidades de negócios internacionais e o investimento estrangeiro direto (Ali; Bakhsh; Yasin, 2023). Segundo a *International Air Transport Association (IATA)*, 1,1% do PIB brasileiro vem do transporte aéreo e dos turistas que chegam por via aérea. Além disso, 839 mil empregos são gerados por esse setor e pelos turistas que chegam ao país, o que reforça a sua importância (IATA, 2019).

No entanto, o setor de transporte aéreo enfrentou um grande desafio em 2020: a pandemia de COVID-19. Devido às regras de distanciamento social, determinações de *lockdown* e fechamento de fronteiras pelos países, esse setor foi um dos mais afetados, junto com aqueles relacionados a turismo, viagens e hotelaria (Borges; Alvarenga, 2023). No Brasil, o mercado aéreo apresentou uma expressiva redução do número de decolagens e de passageiros transportados em 2020. Segundo dados da Agência Nacional de Aviação Civil (ANAC), foram realizados 465 mil voos regulares e não-regulares em 2020, somando-se os mercados doméstico e internacional, o que representa uma queda de 51% com relação a 2019. Além disso, o número de passageiros transportados apresentou a primeira queda desde 2016, com 52 milhões de passageiros domésticos e internacionais, representando uma redução de 56% sobre o valor de 2019 (ANAC, 2021a).

Nos anos seguintes, o mercado aéreo brasileiro começou a apresentar uma recuperação de suas operações, mas sem atingir os níveis de 2019. Em 2021, foram realizados cerca de 600 mil voos regulares e não-regulares e aproximadamente 67 milhões de passageiros domésticos e internacionais foram transportados, o que representa uma redução de 37% e 43%, respectivamente, em comparação a 2019 (ANAC, 2022a). Em 2022, essa redução foi de 12% e 18%, respectivamente, em comparação a 2019, sendo realizados cerca de 831 mil voos com aproximadamente 98 milhões de passageiros transportados (ANAC, 2023a). Já em 2023, foram realizados mais de 911 mil voos regulares e não regulares, representando uma redução de 3% em relação ao nível de 2019. Além disso, foram transportados aproximadamente 112,7 milhões de passageiros domésticos e internacionais, o que representa uma queda de 5% em comparação a 2019 (ANAC, 2024a).

Embora tenha havido uma redução no número de decolagens e no transporte de passageiros, a quantidade de reclamações de consumidores contra empresas aéreas aumentou após 2019. Segundo dados do Consumidor.gov.br¹, foram registradas 42.028

¹ Plataforma utilizada pela ANAC para monitorar as principais dificuldades enfrentadas pelos consumidores e, assim, subsidiar a regulação e a fiscalização do setor. Os dados foram retirados do painel

reclamações em 2019, seguidas por 62.399 em 2020, 95.850 em 2021, 107.204 em 2022 e 86.241 em 2023. Percebe-se que o número de reclamações registradas cresceu 48% em 2020, 54% em 2021 e 12% em 2022, apresentando uma queda de 20% em 2023, porém com um valor absoluto 105% maior em comparação a 2019.

Ao analisar o índice de reclamações registradas por grupo de 100 mil passageiros pagos transportados, percebe-se que o índice médio, considerando empresas brasileiras e estrangeiras, saiu de 35,83 em 2019 para o pico de 142,42 em 2021, um aumento de 297%. O tema mais reclamado pelos consumidores entre 2019 e 2023 foi o reembolso, bastante influenciado pelas medidas emergenciais adotadas diante da pandemia de COVID-19 (ANAC, 2023b).

Além desse cenário de várias reclamações, o número de litígios com empresas aéreas é bastante expressivo no Brasil. Fato que gera questionamentos do porquê dessa grande quantidade de processos judiciais, visto que o sistema de transporte aéreo brasileiro possui bons índices de eficiência em sua operação (ABEAR, 2023). Entre as razões para essa situação, os autores apontam um incentivo à judicialização no país, devido ao acesso fácil e gratuito à Justiça, às decisões judiciais e administrativas que ampliam os deveres das companhias aéreas para além do que está regulamentado, à falta de aplicação de tratados internacionais, à concessão de indenização por danos morais de forma punitiva, entre outros (Starling; Villa, 2022; Scaranello; Klarmann; Silva, 2022). Essa judicialização resulta em diversos custos para as companhias aéreas, que acabam sendo transferidos ao consumidor, principalmente através do aumento no valor das passagens aéreas. Além disso, as companhias aéreas optam por restringir a malha aérea como forma de minimizar a ocorrência de novos litígios.

1.1 Formulação do problema

Diante do crescente número de reclamações e litígios judiciais envolvendo companhias aéreas no Brasil, é essencial compreender as falhas nos serviços que contribuem para esse cenário, isto é, quais aspectos do serviço não estão atendendo às expectativas dos clientes em termos de qualidade e desempenho. Esse desafio é ainda mais relevante considerando as particularidades do setor aéreo, que é vulnerável a falhas devido à grande quantidade de provedores envolvidos, ao elevado número de passageiros e à natureza do serviço, que depende diretamente das interações humanas (Atalik, 2007). Nesse contexto, identificar pontos críticos nas operações das companhias aéreas é fundamental para aprimorar a experiência do consumidor e reduzir a judicialização das reclamações.

Com esse objetivo, estudos anteriores utilizaram o aprendizado de máquina como ferramenta para ajudar gestores a compreender melhor os motivos e as causas por trás

interativo Reclamações de passageiros, disponibilizado pela ANAC, disponível neste link.

das reclamações de clientes e dos processos judiciais contra companhias aéreas. Pesquisas como as de Sezgen, Mason e Mayer (2019), Korfiatis *et al.* (2019), Kwon *et al.* (2021) e Song *et al.* (2024) aplicaram técnicas como modelagem de tópicos, análise de sentimentos e modelos preditivos em dados de avaliações on-line, a fim de identificar os aspectos mais relevantes da experiência dos clientes e os fatores que geram satisfação ou insatisfação. Esses estudos apontaram que fatores como o atendimento ao cliente, os serviços de bordo e o custo-benefício são cruciais para a satisfação dos passageiros, enquanto problemas como atrasos, interrupções de voo, questões com bagagem, reembolsos e cancelamentos são as principais causas de insatisfação.

Além disso, trabalhos como os de Rita, Moro e Cavalcanti (2022), Çalli e Çalli (2023) e Kim, Lim e Ha (2024) investigaram as mudanças no comportamento dos consumidores nos períodos pré, durante e pós-pandemia de COVID-19. Esses estudos identificaram o surgimento de novos temas na pandemia e um aumento significativo das queixas relacionadas a cancelamentos e reembolsos, resultado das adaptações forçadas no setor aéreo devido às medidas de segurança e restrições sanitárias.

No entanto, uma lacuna nas pesquisas existentes é a dificuldade de classificar reclamações que envolvem múltiplos tópicos, como destacado por Çalli e Çalli (2023). A maioria dos estudos analisou dados não estruturados, como avaliações on-line, o que dificultou uma categorização precisa das queixas. O presente estudo, por sua vez, propõe uma análise utilizando dados estruturados, permitindo uma classificação mais detalhada das reclamações, incluindo aquelas que abrangem múltiplos problemas simultaneamente.

Além disso, grande parte das pesquisas anteriores se concentrou na análise retrospectiva das reclamações, sem abordar suas possíveis implicações futuras, especialmente no que diz respeito à evolução para litígios judiciais. Este estudo examina ambos os temas, analisando seu comportamento ao longo do período estudado e investigando a correlação entre os motivos das reclamações e dos processos judiciais. Ao explorar essa relação, busca-se identificar quais aspectos do serviço estão mais frequentemente associados a litígios e como esses problemas são tratados no contexto jurídico.

No contexto da judicialização, além de identificar os motivos com maior potencial para gerar litígios, também é essencial reconhecer as variáveis mais relevantes para a definição do valor das indenizações judiciais, a fim de mitigar os impactos financeiros e aprimorar a gestão dos litígios. O trabalho de Torres, Guterres e Celestino (2023), por exemplo, comparou três modelos de aprendizado de máquina (*Random Forest*, *Support Vector Machines* e *Naive Bayes*) para prever o valor das indenizações, demonstrando que o modelo *Random Forest* apresentou o melhor desempenho. Contudo, os autores destacaram uma limitação relacionada à falta de padronização na classificação dos motivos e causas dos processos judiciais.

Os motivos dos processos judiciais correspondem aquilo que motivou o(s) autor(es)

a iniciar o processo. Já as causas dos processos judiciais correspondem à causa alegada pela empresa ré, isto é, a defesa ou justificativa alegada sobre a reclamação feita contra ela. Não existe uma padronização para os termos que devem ser usados para descrever esses dois campos nos processos judiciais, o que resulta no uso de termos diferentes para descrever um mesmo problema. Em relação ao motivo, por exemplo, processos relacionados ao cancelamento ou atraso de voos podem ser descritos de diversas formas, como: "Atraso ou Cancelamento de Voo", "Cancelamento de Voo", "Atraso de Voo", "CANCELAMENTO", "ATRASSO", "Cambios y cancelaciones", entre outros. Dessa forma, a limitação mencionada por Torres, Guterres e Celestino (2023) é que cada uma dessas descrições podem ter sido interpretadas pelos modelos de aprendizado de máquina como eventos distintos, o que provavelmente impactou o desempenho dos algoritmos.

Essa diversidade nas classificações dos motivos e causas dos processos judiciais representa um desafio importante, já que os casos são iniciados em diferentes instâncias da justiça e envolvem empresas com abordagens distintas para a gestão das queixas. Estudos anteriores, como os de Sabo *et al.* (2022) e Martino, Pio e Ceci (2023), já ressaltaram a complexidade dos dados jurídicos e a necessidade da participação de especialistas na definição dos rótulos de classificação de grupos. Para superar essa limitação, este estudo propõe uma padronização na classificação dos motivos e causas, ou seja, uma abordagem sistemática para reclassificar reclamações e processos que tratam de um mesmo tema com um termo padrão. No caso dos termos descritos anteriormente, a padronização faria com que os motivos de todos os processos fossem reclassificados para 'cancelamento ou atraso de voo'. Isso pode melhorar a precisão dos modelos preditivos e, conseqüentemente, das estimativas de indenização, contribuindo para um gerenciamento mais eficiente dos litígios no setor aéreo.

1.2 Pergunta de pesquisa

Diante da problemática apresentada, formulou-se as seguintes perguntas de pesquisa:

- Quais são os principais motivos de reclamações de consumidores e processos judiciais contra empresas de transporte aéreo no Brasil?
- Existe correlação entre os principais motivos de reclamações de consumidores e de processos judiciais contra empresas de transporte aéreo no Brasil?
- Existe melhora no desempenho dos modelos de aprendizado de máquina de previsão do valor de indenização, quando é aplicada a padronização da classificação dos motivos e causas de processos judiciais?

1.3 Objetivos de pesquisa

1.3.1 *Objetivo geral*

A presente pesquisa tem como objetivo geral analisar as reclamações de consumidores e os processos judiciais contra empresas de transporte aéreo no Brasil, além de examinar a correlação entre os motivos de reclamações e processos judiciais e aplicar aprendizado de máquina para prever o valor de indenizações.

1.3.2 *Objetivos específicos*

São objetivos específicos (OE) da presente pesquisa:

- Analisar as características temporais, geográficas, por empresas e motivos de reclamações de consumidores e processos judiciais contra companhias aéreas (OE 1);
- Identificar a correlação entre os motivos de reclamação de consumidores e de processos judiciais contra companhias aéreas (OE 2);
- Padronizar a classificação de motivos e causas dos processos judiciais e avaliar os efeitos no desempenho dos modelos de aprendizado de máquina, elaborados por Torres, Guterres e Celestino (2023), para previsão do valor de indenização (OE 3).

1.4 Justificativa da pesquisa

O transporte aéreo brasileiro enfrenta um cenário desafiador, marcado pelo aumento excessivo de reclamações de consumidores e pela crescente judicialização dos conflitos com as companhias aéreas. Esse quadro impacta negativamente os resultados financeiros das empresas do setor, gerando consequências como a inviabilidade de algumas operações, o aumento do preço das passagens e a redução da disponibilidade de malha aérea em determinadas regiões. Além disso, contribui para a criação de um ambiente de insegurança jurídica, afetando a confiança dos consumidores e das empresas. Diante desse contexto, é fundamental estudar esses temas, a fim de gerar conhecimento que possa auxiliar tanto as companhias aéreas quanto as esferas públicas responsáveis a adotarem medidas mitigadoras, contribuindo para a melhoria do serviço e a redução da judicialização.

Apesar da relevância do tema, existem poucos estudos que tratam especificamente dos fatores que influenciam as reclamações e os litígios judiciais no setor aéreo brasileiro, especialmente em uma análise integrada desses dois aspectos. Este trabalho, portanto, busca preencher essa lacuna ao investigar as motivações e causas das reclamações de consumidores e dos processos judiciais contra as companhias aéreas, com especial atenção

às mudanças no cenário pré, durante e pós-pandemia de COVID-19. Além disso, analisa a possibilidade de correlação entre as razões dessas reclamações e os processos judiciais. A compreensão dessa relação pode ser útil para as companhias aéreas ao identificarem padrões de reclamações frequentemente associados a disputas legais, auxiliando no desenvolvimento de estratégias para lidar com esses problemas.

Adicionalmente, a padronização na classificação dos motivos das reclamações e dos processos judiciais representa uma contribuição importante deste estudo, visto que a diversidade na forma como os problemas são classificados nos litígios pode comprometer a precisão das análises e das previsões sobre os valores das indenizações. A uniformização dessas categorias ajudará a identificar com maior clareza os fatores que mais frequentemente levam à judicialização e influenciam o valor das indenizações.

A execução desta dissertação se justifica, portanto, pela sua contribuição à literatura acadêmica, ao trazer uma análise profunda de um tema de grande relevância para o setor aéreo brasileiro. Além disso, este estudo amplia a compreensão sobre a aplicação da inteligência artificial na gestão de reclamações e litígios judiciais no setor aéreo, o que pode favorecer o desenvolvimento de ferramentas mais eficazes para identificar e corrigir falhas nos serviços prestados pelas companhias aéreas e, conseqüentemente, na melhoria da experiência do consumidor.

1.5 Estrutura do trabalho

Este trabalho está estruturado em cinco seções. A primeira corresponde à introdução. A segunda aborda o referencial teórico que fundamenta a pesquisa. A terceira descreve os métodos e técnicas de pesquisa utilizados. A quarta apresenta os resultados obtidos e a discussão correspondente. Por fim, a quinta seção contempla as considerações finais, as limitações do estudo e sugestões para pesquisas futuras. Ao final, são disponibilizadas as referências e os apêndices desta dissertação.

2 REFERENCIAL TEÓRICO

Esta seção do trabalho apresenta as bases conceituais que fundamentam a pesquisa, essenciais para as discussões subsequentes, sendo elas: i) reclamações contra empresas de transporte aéreo; e ii) judicialização do transporte aéreo.

2.1 Reclamações contra empresas de transporte aéreo

O comportamento de reclamação do consumidor refere-se ao conjunto de respostas comportamentais e não comportamentais expressas pelos consumidores. Esse comportamento envolve a comunicação de percepções negativas relacionadas a um episódio de consumo, motivada pela insatisfação com a experiência vivida (Rogers; Williams, 1990). Essa insatisfação geralmente decorre de uma falha no serviço, ou seja, quando a qualidade do serviço não atende às expectativas do cliente (Chahal; Devi, 2015 *apud* Park; Park, 2016). Tais falhas são uma das principais causas de desgaste entre as empresas e seus clientes, uma vez que afeta negativamente as percepções de qualidade do serviço e as intenções de compra, além de resultar em uma redução da satisfação do cliente e gerar boca a boca negativo, caso não seja adequadamente tratada (Lin; Wang; Chang, 2011 *apud* Park; Park, 2016).

O setor aéreo, devido à sua expansão e às características inerentes ao serviço, é bastante propenso a enfrentar experiências de falha (Park; Park, 2016). Até mesmo as melhores companhias aéreas enfrentam ocasionalmente falhas no serviço, como *overbooking* ou atrasos em voos. Tais incidentes podem representar custos significativos para as empresas, já que os clientes frequentemente optam por mudar para outros provedores de serviços após experiências insatisfatórias (Chang; Chang, 2010). Considerando que a competição entre as companhias aéreas aumentou significativamente, estratégias de diferenciação através da melhoria da qualidade do serviço tornam-se essenciais para a sobrevivência no mercado (Kim; Lim; Ha, 2024). Em vista disso, quando a eliminação ou prevenção das falhas no serviço não for possível, as companhias aéreas precisam buscar soluções para lidar com elas de forma eficaz (Chang; Chang, 2010; Park; Park, 2016).

Segundo Song *et al.* (2024), para melhorar a qualidade do serviço, as companhias aéreas precisam examinar cada etapa que o cliente vivencia ao longo da cadeia de serviço, visto que essas atividades têm impacto direto na qualidade do serviço e no resultado final de satisfação. Além disso, as companhias aéreas precisam entender de forma precisa os desejos e as necessidades de seus clientes, a fim de atender às suas expectativas e aprimorar seu desempenho no mercado (Tsafarakis; Kokotas; Pantouvakis, 2018 *apud* Kim; Lim; Ha, 2024). Entretanto, conforme Kim, Lim e Ha (2024), as companhias aéreas têm

dificuldade em satisfazer as expectativas dos clientes devido à variação nas necessidades dos passageiros e nas expectativas quanto à qualidade do serviço.

2.1.1 Aplicação do aprendizado de máquina em reclamações e avaliações de clientes de transporte aéreo

Tendo em vista as dificuldades em compreender melhor as variações nas necessidades e expectativas dos passageiros, a aplicação de aprendizado de máquina tem sido explorada como uma ferramenta para auxiliar gestores e empresas a entender esse fenômeno. Estudos como os de Sezgen, Mason e Mayer (2019), Korfiatis *et al.* (2019) e Kwon *et al.* (2021), por exemplo, utilizaram técnicas de aprendizado de máquina para analisar avaliações on-line de companhias aéreas, com o objetivo de identificar os aspectos mais relevantes da experiência dos clientes e os fatores que geram satisfação ou insatisfação.

Na análise de Sezgen, Mason e Mayer (2019), o foco foi identificar quais atributos de serviço levam à satisfação e à insatisfação dos passageiros com base em diferentes modelos de negócios e classes de serviço das companhias aéreas. Para isso, os autores analisaram mais de 5 mil avaliações em inglês de passageiros internacionais de 50 companhias aéreas, coletadas no site TripAdvisor, de janeiro a dezembro de 2017. Na análise, utilizaram a técnica *Latent Semantic Analysis* (LSA)¹, que é uma abordagem de processamento de linguagem natural. Os resultados demonstraram que os principais fatores que impactam as avaliações positivas dos passageiros foram:

- **Passageiros de cabine econômica:** 1. Funcionários simpáticos e prestativos; 2. Experiência e atendimento ao cliente sem complicações; 3. Assentos confortáveis e espaço para as pernas;
- **Passageiros de cabine *premium*:** 1. Valor (custo-benefício valor pago x valor da experiência); 2. Funcionários simpáticos e prestativos; 3. Alimentação e bebidas; 4. Serviço de bordo;
- **Passageiros de companhias aéreas de baixo custo:** 1. Preço baixo 2. Amigabilidade e cortesia dos funcionários; 3. Bom serviço da tripulação da cabine.

Por outro lado, os principais fatores que impactam as avaliações negativas dos passageiros foram:

- **Passageiros de cabine econômica:** 1. Assentos desconfortáveis e pouco espaço para as pernas; 2. Bagagens e interrupções de voo; 3. Falta de profissionalismo dos funcionários; 4. Mau atendimento e alimentos e bebidas;

¹ LSA é uma técnica de modelagem de tópicos que examina as associações entre uma enorme coleção de arquivos de texto e as palavras/frases incluídas, capturando a relação semântica entre termos e documentos (Gupta; Chatterjee; Gupta, 2022).

- **Passageiros de cabine *premium*:** 1. Falta de profissionalismo dos funcionários; 2. Assentos desconfortáveis e aeronaves antigas; 3. Bagagens e interrupções de voo;
- **Passageiros de companhias áreas de baixo custo:** 1. Assentos desconfortáveis e pouco espaço para as pernas; 2. Interrupções de voo; 3. Prestação consistente de serviços ruins; 4. Falta assistência ao cliente; 5. Taxas extras ou ocultas.

Assim, Sezgen, Mason e Mayer (2019) concluíram que existem diferenças que impulsionam a satisfação e a insatisfação dos passageiros, dependendo da classe de viagem aérea adquirida e se a companhia aérea é de baixo custo ou de serviço completo.

Korfiatis *et al.* (2019) também utilizaram dados de avaliações de passageiros publicadas no TripAdvisor. Os autores coletaram mais de 557 mil avaliações, abrangendo diversas companhias aéreas e informações detalhadas, como data do voo, nome da companhia aérea, rota (aeroporto de partida e destino), classe de cabine (primeira classe, classe executiva, econômica premium e classe econômica) e nível de contribuição dos revisores para a plataforma (calculado a partir do número de postagens de avaliação). A partir dessas avaliações, eles aplicaram técnicas de mineração de texto, utilizando a modelagem de tópicos *Structural Topic Model* (STM)², combinada com análises estatísticas, para extrair dados significativos e explorar relações entre as dimensões da qualidade do serviço.

Na modelagem de tópicos, Korfiatis *et al.* (2019) identificaram que o atendimento ao cliente é o principal fator associado ao aumento da satisfação, com destaque para o tópico 'elogios à equipe'. A relação custo-benefício também tem um impacto positivo significativo. Por outro lado, atrasos, críticas à equipe e problemas com reembolsos/cancelamentos são fatores críticos que geram insatisfação, especialmente quando não tratados adequadamente. Os autores também realizaram análises temporais para observar como as dimensões da qualidade do serviço evoluem ao longo do tempo, especialmente durante períodos de alta demanda, como férias de verão e feriados. Os resultados mostraram que tópicos como 'atrasos' se tornam mais prevalentes em períodos de alta demanda, como o verão e o Natal, refletindo condições específicas do mercado.

Por fim, Korfiatis *et al.* (2019) utilizaram os dados para mapear a competição entre companhias aéreas, demonstrando como diferentes empresas se posicionam em relação às dimensões de qualidade do serviço mais valorizadas pelos passageiros. A análise revelou que, dependendo do segmento de mercado (intra-europeu ou intercontinental), as companhias aéreas adotam estratégias distintas, ajustando-se às expectativas dos consumidores. As empresas de baixo custo, como EasyJet e Ryanair, se destacaram pela sua ênfase na economia, enquanto companhias *premium*, como Singapore Airlines e Emirates, priorizaram um serviço de maior qualidade, com ênfase no conforto e no atendimento ao cliente.

² O STM é uma extensão probabilística da LDA que permite incorporar variáveis de nível de documento, como notas ou metadados, para melhorar a inferência dos tópicos presentes nos textos (Roberts *et al.*, 2014).

Já Kwon *et al.* (2021) investigaram os fatores que geram satisfação ou insatisfação nos clientes com experiência no uso de companhias aéreas na Ásia. Para isso, os autores coletaram mais de 14 mil avaliações de clientes sobre 27 companhias aéreas asiáticas, publicadas no site Skytrax (airlinequality.com). Eles utilizaram técnicas de mineração de texto, incluindo modelagem de tópicos *Latent Dirichlet Allocation* (LDA)³ e análise de sentimentos⁴, para extrair e analisar dados significativos das avaliações.

Na modelagem de tópicos, Kwon *et al.* (2021) identificaram seis tópicos principais que influenciam a experiência do cliente e suas intenções de compra ao escolher a companhia aérea, sendo eles: refeições a bordo (relacionado à qualidade das refeições), entretenimento (opções de entretenimento durante o voo), classe de assento (diferenciação entre classes de assento como econômica e executiva), conforto do assento (focado no conforto e espaço dos assentos), Singapore Airlines (referências específicas à companhia aérea) e serviço da equipe (qualidade do serviço prestado pela tripulação). Já na análise de sentimento, os resultados destacaram palavras negativas como *'poor'*, *'bad'*, *'problem'*, *'difficult'* e *'delayed'*, associadas a problemas como atrasos e falhas no serviço. Por outro lado, palavras positivas como *'good'*, *'great'*, *'fantastic'*, *'excellent'* e *'amazing'* indicaram satisfação, especialmente em aspectos como refeições a bordo e limpeza. A palavra *"smiling"* foi destacada como um fator positivo, sugerindo que o sorriso da tripulação impacta favoravelmente a experiência do cliente.

Song *et al.* (2024) fizeram uma aplicação diferente das técnicas de aprendizado de máquina. Os autores buscaram prever a satisfação dos passageiros com o serviço de bordo após atrasos nos horários de partida e de chegada dos voos, utilizando um conjunto de dados de satisfação de passageiros de uma companhia aérea disponibilizado no Kaggle⁵. Para isso, primeiramente, eles realizaram uma análise da interação entre serviços individuais de bordo e características dos passageiros usando o coeficiente de correlação de Pearson e métodos de agrupamento (*Principal Component Analysis* - PCA e *K-means clustering*). Com base nas interações identificadas, desenvolveram um modelo de previsão de satisfação com Wide & Deep⁶, que é um algoritmo de aprendizagem de máquina que combina o modelo linear generalizado e a rede neural profunda (Song *et al.*, 2024). Os

³ LDA é uma técnica de modelagem de tópicos em que o modelo assume que cada documento é uma mistura de vários tópicos e que cada palavra em um documento é atribuída a um desses tópicos. Dessa forma, o LDA tenta estimar os tópicos subjacentes e as distribuições de palavras para cada tópico, analisando a frequência com que as palavras ocorrem juntas nos documentos (Blei; Ng; Jordan, 2003).

⁴ A análise de sentimento refere-se a uma técnica que classifica ou quantifica emoções em texto e as transforma em informações objetivas (Jang; Park; Kim, 2015).

⁵ Kaggle é uma plataforma de competição de ciência de dados e uma comunidade online de cientistas de dados e profissionais de aprendizado de máquina da Google LLC. Além disso, diversos conjuntos de dados são disponibilizados no site: <https://www.kaggle.com/>.

⁶ O algoritmo Wide & Deep foi proposto pelo Google em 2016 e é aplicado à área de sistemas de recomendação (Cheng *et al.*, 2016 *apud* Song *et al.*, 2024). Ele utiliza da interação de recursos para explorar a relação não linear entre recursos e melhorar a precisão do modelo (Tompson *et al.*, 2014 *apud* Song *et al.*, 2024).

autores compararam o desempenho do modelo Wide & Deep com outros cinco modelos de aprendizado de máquina (*Multiple Layer Perceptron* - MLP, *Support Vector Machine* - SVM, Árvore de decisão, Regressão logística e *Random Forest* - RF), utilizando métricas de acurácia, precisão, *recall*, *F1-Score* e curva AUC. O modelo Wide & Deep foi o que apresentou melhor desempenho, seguido pelo *Random Forest*.

Song *et al.* (2024) também fizeram uma análise de importância dos recursos utilizando o algoritmo DeepLIFT⁷, a fim de interpretar o modelo Wide & Deep e desvendar os recursos mais importantes que afetam a satisfação dos passageiros. Os resultados mostraram que, para os passageiros que continuam a viajar em voos atrasados, o serviço Wi-Fi a bordo é o recurso mais importante em termos de satisfação final do passageiro, seguido pela distância do voo e pelo tipo de viagem. As classificações do serviço de check-in e manuseio de bagagem também têm impacto significativo na previsão do modelo de satisfação.

Os trabalhos apresentados demonstram como esse tema vêm sendo estudado. O Quadro 1 apresenta um resumo dos objetivos e principais resultados e conclusões de cada estudo.

⁷ O algoritmo DeepLIFT é um método amplamente usado para explicar modelos de aprendizagem profunda, oferecendo interpretação global e local. Ele avalia a importância do recurso em todas as amostras calculando o grau de contribuição de cada recurso para a saída do modelo. (Song *et al.*, 2024).

Quadro 1 – Quadro resumo dos trabalhos de aplicações de aprendizado de máquina nas questões de consumidores do transporte aéreo

Autores	Ano	Periódico	Objetivo	Principais Resultados/Conclusões
Sezgen, Mason e Mayer	2019	Journal of Air Transport Management	Identificar quais atributos de serviço levam à satisfação e à insatisfação dos passageiros com base em diferentes modelos de negócios e classes de serviço das companhias aéreas.	A satisfação de passageiros de cabine econômica é impactada positivamente por funcionários simpáticos e prestativos, enquanto os de cabine premium priorizam o valor da experiência e os de companhias áreas de baixo custo, o preço baixo. Por outro lado, o que mais impacta negativamente a satisfação tanto dos passageiros de cabine econômica como de companhias aéreas de baixo custo são os assentos desconfortáveis e pouco espaço para as pernas, enquanto os de cabine premium focam na falta de profissionalismo dos funcionários.
Konfiatis et al.	2019	Expert Systems with Applications	Explorar relações entre as dimensões da qualidade do serviço com base em avaliações de passageiros e mapear a competição entre companhias aéreas.	Atendimento ao cliente e custo-benefício são fatores críticos para satisfação, enquanto atrasos, críticas à equipe e problemas com reembolsos/cancelamentos são fatores críticos para insatisfação. Tópicos sobre atrasos aumentam em períodos de alta temporada. As empresas de baixo custo enfatizam a economia, enquanto as premium priorizam conforto e serviço de qualidade.
Kwon et al.	2021	Information	Investigar fatores de satisfação e insatisfação em clientes de companhias aéreas asiáticas, com foco nas experiências de voo.	Seis tópicos principais impactam a experiência do cliente: refeições a bordo, entretenimento, classe de assento, conforto do assento, Singapore Airlines e serviço da equipe. A análise de sentimentos revela uma ligação entre palavras negativas e problemas como atrasos e falhas no serviço. Por outro lado, palavras positivas estão relacionadas a satisfação com refeições a bordo, limpeza e simpatia da tripulação.
Song et al.	2024	Journal of Air Transport Management	Prever a satisfação dos passageiros com o serviço de bordo após atrasos nos horários de partida e de chegada dos voos.	O serviço Wi-Fi a bordo é o recurso mais importante em termos de satisfação final do passageiro, seguido pela distância do voo e pelo tipo de viagem.

Fonte: Elaborado pela autora.

2.1.1.1 Reclamações e avaliações de clientes do transporte aéreo no contexto da pandemia de COVID-19

Alguns trabalhos também foram elaborados considerando o contexto da pandemia de COVID-19, que foi uma situação que causou mudanças nas necessidades dos passageiros. Rita, Moro e Cavalcanti (2022), por exemplo, realizaram uma pesquisa com o objetivo de avaliar como as companhias aéreas lidaram com os problemas de cancelamentos e insatisfação enfrentados pelos clientes durante o período de incerteza causado pela pandemia de COVID-19, comparando os resultados obtidos com as informações divulgadas pela mídia. O estudo analisou 885 avaliações de clientes publicadas no TripAdvisor, relacionadas às 10 maiores companhias aéreas do mundo, considerando o número de passageiros em 2019. As avaliações abrangeram o período crítico de março a maio de 2020, quando as restrições de viagens foram amplamente implementadas e gradualmente flexibilizadas. Para a análise, os autores utilizaram a ferramenta *Valence Aware Dictionary for sEntiment Reasoning* (VADER), que é um algoritmo baseado em Python para análise de sentimentos, classificando as avaliações como positivas, negativas ou neutras.

Ao todo, Rita, Moro e Cavalcanti (2022) testaram 5 hipóteses, alcançando os seguintes resultados:

- **H1 - As companhias aéreas não estão sendo capazes de lidar adequadamente com cancelamentos e atrasos de voos:** A hipótese foi confirmada, com a análise das avaliações demonstrando insatisfação generalizada dos clientes em relação às políticas de reembolso e aos processos adotados pelas companhias durante a pandemia. As críticas refletiram a má gestão e a incapacidade de atender às expectativas dos passageiros;
- **H2 - A pandemia impactou negativamente a satisfação média do cliente das companhias aéreas:** Os resultados indicaram uma predominância de avaliações negativas, evidenciando o impacto adverso da pandemia na satisfação dos passageiros. Questões relacionadas a reembolsos e serviços online foram as mais criticadas, reforçando a ideia de que o período afetou significativamente a percepção dos clientes;
- **H3a - O valor da marca das companhias aéreas está sendo afetado negativamente pela pandemia:** A hipótese foi validada ao se observar que elementos como imagem e consistência das marcas foram prejudicados. A gestão inadequada da crise diminuiu a confiança e a percepção geral das companhias aéreas pelos clientes;
- **H3b - O atendimento ao cliente das companhias aéreas durante a pandemia está movendo os clientes para a concorrência:** Os dados mostraram que experiências negativas levaram os clientes a buscar alternativas em outras empresas,

evidenciando uma queda na lealdade à marca e um aumento da vulnerabilidade competitiva das companhias;

- **H4 - Os clientes tendem a escrever comentários mais extremos do que moderados:** A análise revelou uma prevalência significativa de avaliações extremamente positivas ou negativas, confirmando que os clientes expressam suas experiências de forma polarizada, especialmente em situações de insatisfação ou alta satisfação.

Por outro lado, o trabalho de Çallı e Çallı (2023) teve como objetivo entender as diferenças nas reclamações de consumidores entre companhias aéreas de baixo custo e serviço completo durante o início da pandemia de COVID-19, além de avaliar as variações dessas reclamações em relação ao período anterior à pandemia. Os autores utilizaram o algoritmo de LDA para categorizar 10.594 reclamações registradas contra a Pegasus Airlines e a Turkish Airlines⁸ no site *sikayetvar.com*, entre 22 de março de 2020 e 28 de fevereiro de 2021.

Em relação à primeira pergunta de pesquisa, os autores identificaram que ambas empresas tiveram tópicos de reclamação similares, como reembolso, bilhetes em aberto, bagagem, serviço ao cliente, membros da família, distanciamento social, assentos e website. No entanto, a Turkish Airlines apresentou mais reclamações relacionadas ao programa de milhas, assentos de classe executiva, equipe de check-in e conflitos de check-in online/offline. Por outro lado, a Pegasus Airlines teve mais reclamações notáveis sobre a franquia de bagagem de mão, um call center que cobra por minuto e questões de viagem de membros da família relacionados à idade. As reclamações mais comuns contra a Turkish Airlines foram sobre reembolsos, cancelamentos, serviço ao cliente e website, enquanto para a Pegasus, os tópicos mais frequentes foram reembolsos e bilhetes em aberto.

A respeito da segunda pergunta, Çallı e Çallı (2023) identificaram que, no início da pandemia de COVID-19, surgiram novos fatores que não estavam presentes em estudos anteriores. O mais evidente foi a questão dos bilhetes em aberto, devido às proibições de voos e a falta de informações dos clientes sobre as regulamentações. Além disso, fatores relacionados à família, especialmente reclamações sobre regulamentos de viagem para menores de 20 anos e maiores de 65, também foram tópicos críticos de reclamação para ambas as companhias aéreas. Outros subtemas, como distanciamento social, assentos vazios para segurança sanitária, voos com capacidade total e uso de máscaras, também surgiram, diferenciando-se do período pré-pandemia.

Os autores também perceberam que fatores comuns antes da pandemia, como equipe de cabine, limpeza, refeições e entretenimento, não foram mencionados como reclamações pelos passageiros durante a pandemia. Isso sugere que, no início da pandemia, os passageiros

⁸ Para o estudo, considerou-se a Turkish Airlines como opção de companhia aérea com serviço completo e a Pegasus Airlines, de serviço de baixo custo.

ou ignoraram esses aspectos ou não tiveram experiências negativas relacionadas a eles. No entanto, questões relacionadas a assentos e bagagens mantiveram-se semelhantes em termos de satisfação dos consumidores em ambos os períodos. Reclamações sobre atrasos, que eram típicas antes da pandemia, foram mais frequentes entre clientes de companhias de baixo custo durante a pandemia, enquanto queixas relacionadas ao check-in foram mais comuns em companhias de serviço completo. Além disso, reclamações sobre reembolsos e problemas relacionados surgiram durante a pandemia para ambas as companhias. Por fim, queixas relacionadas a serviços ao cliente e websites mostraram-se semelhantes aos estudos realizados antes da COVID-19.

Já Kim, Lim e Ha (2024) realizaram uma pesquisa com o objetivo de comparar as mudanças nas percepções dos passageiros sobre os serviços aéreos antes e durante a pandemia de COVID-19. Nela, foram analisadas avaliações de clientes do site TripAdvisor a respeito de 57 das 100 companhias aéreas que a Skytrax seleciona a cada ano como qualificadas, excluindo aquelas com poucas avaliações em inglês. O período considerado foi de 2018 a dezembro de 2021. Os autores utilizaram modelagem de tópicos e análise de sentimentos para conduzir a análise comparativa. Primeiramente, os tópicos influentes foram extraídos das avaliações usando o modelo LDA, sendo distribuídos em seguida nas cinco dimensões do modelo SERVQUAL⁹ (tangível, confiabilidade, capacidade de resposta, empatia e segurança), a fim de facilitar a comparação antes e durante a COVID-19. Em seguida, as avaliações foram classificadas como negativas ou positivas com base na pontuação da avaliação e um modelo de classificação para revisões foi criado usando modelos de regressão logística (que é algoritmo de aprendizado de máquina supervisionado), tendo a sua precisão avaliada por meio de uma matriz de confusão. Por fim, as palavras-chave negativas foram extraídas usando análise de regressão logística para identificar com quais partes dos serviços aéreos os passageiros estavam insatisfeitos.

Como resultado, Kim, Lim e Ha (2024) concluíram que, antes da COVID-19, os passageiros estavam mais preocupados com as dimensões tangíveis e de confiabilidade, sendo os temas principais das avaliações: pontualidade, qualidade dos colaboradores e serviços de bordo. Por outro lado, durante a COVID-19, houve uma mudança de foco para os serviços terrestres e a capacidade de resposta, refletido no aumento das menções sobre serviço de reembolso, serviços a bordo, alterações e reservas, adaptação a mudanças de horários, procedimentos de check-in e qualidade dos colaboradores, especialmente em garantir o cumprimento das regras de quarentena e uso de máscaras.

Os trabalhos apresentados demonstram como esse tema vêm sendo estudado. O Quadro 2 apresenta um resumo dos objetivos e principais resultados e conclusões de cada estudo.

⁹ O modelo SERVQUAL foi proposto por Parasuraman, Zeithaml e Berry (1988) e trata-se de uma escala utilizada para medir a qualidade do serviço dividindo os serviços ao cliente em cinco dimensões, que oferecem critérios óbvios (Kim; Lim; Ha, 2024).

Quadro 2 – Quadro resumo dos trabalhos de aplicações de aprendizado de máquina nas questões de consumidores do transporte aéreo e pandemia de COVID-19

Autores	Ano	Periódico	Objetivo	Principais Resultados/Conclusões
Rita, Moro e Cavalcanti	2022	Journal of Air Transport Management	Avaliar como as companhias aéreas lidaram com cancelamentos e insatisfação dos clientes durante a pandemia.	O estudo confirmou que as companhias aéreas falharam na gestão de cancelamentos e reembolsos durante a pandemia, o que prejudicou a satisfação dos passageiros e a imagem das marcas. Isso resultou em queda na lealdade e aumento da migração para concorrentes.
Çalh e Çalh	2023	Transportation Research Record	Estudar as mudanças nas reclamações de consumidores entre companhias aéreas de baixo custo e serviço completo no início da pandemia de COVID-19.	Durante a pandemia, surgiram novas reclamações, como bilhetes em aberto, questões de família e distanciamento social. As reclamações de reembolso e serviço ao cliente aumentaram.
Kim, Lim e Ha	2024	Journal of Air Transport Management	Comparar as mudanças nas percepções dos passageiros sobre os serviços aéreos antes e durante a pandemia de COVID-19.	Antes da COVID-19, os passageiros priorizavam pontualidade, qualidade dos colaboradores e serviços de bordo. Durante a COVID-19, o foco mudou para serviço de reembolso, serviços a bordo, alterações e reservas, adaptação a mudanças de horários, procedimentos de check-in e qualidade dos colaboradores.

Fonte: Elaborado pela autora.

2.2 Judicialização do transporte aéreo

2.2.1 Contextualização da judicialização do transporte aéreo no Brasil

O número de litígios com empresas de transporte aéreo é bastante expressivo no Brasil. Em 2021, dados do Instituto Brasileiro de Direito Aeronáutico (IBAER) apresentaram que 98,5% das ações cíveis no mundo contra companhias aéreas estavam concentradas no Brasil (CNJ, 2021), o que gera um questionamento do porquê dessa quantidade de processos judiciais. Num primeiro momento, essa alta judicialização poderia ser relacionada a uma prestação de serviço de baixa qualidade. Porém, as empresas aéreas brasileiras são reconhecidas mundialmente pelos altos índices de eficiência na operação aérea, especialmente na pontualidade e regularidade em comparação com países como os Estados Unidos, que possui um dos mercados mais maduros e é comparável ao mercado brasileiro pelas dimensões continentais (ABEAR, 2023). Isso coloca em evidência a boa organização operacional do sistema de transporte aéreo brasileiro e retorna a dúvida da justificativa dessa quantidade de processos judiciais contra empresas aéreas.

Segundo Torres (2022, p. 13), um estudo da IATA mostrou que "há incentivos a judicialização no país, resumidos em falta de aplicação de tratados internacionais e concessão de indenização por danos morais de forma punitiva e sem comprovação do dano causado". Essa questão é corroborada por Starling e Villa (2022), que em uma análise destacaram que grande parte das ações judiciais contra companhias aéreas tem por objeto unicamente danos morais, o que significa dizer que tais ações consideram que um eventual atraso de voo, por exemplo, seria um descontentamento grave o suficiente para gerar uma lesão ao direito da personalidade. No entanto, tal falha do serviço poderia dar causa apenas a um dano de ordem material. Assim, as autoras apontaram o que diz a legislação específica do transporte aéreo nacional (Código Brasileiro de Aeronáutica e a Resolução nº 400/2016 da ANAC) e internacional (Convenção de Montreal) a respeito do dano moral, que pode ser resumido da seguinte forma:

- De acordo com o artigo 251-A do Código Brasileiro de Aeronáutica (CBA, Lei nº 7.565/1986), o dano moral resultante de uma falha na prestação de serviços de transporte aéreo só pode ser considerado quando houver evidência de que o passageiro sofreu prejuízo devido à ação ou omissão da companhia aérea. Ressalta-se que esse artigo foi incluído em 2020, no início da pandemia do COVID-19, até então o Judiciário, nas primeiras e segundas instâncias, tinha como regra o entendimento de dano presumido (*in re ipsa*), ou seja, a narrativa do passageiro sobre a falha na prestação do serviço aéreo já era suficiente para configurar o dano moral, cabendo ao magistrado determinar o valor da indenização a ser pago ao passageiro;

- Por outro lado, a Resolução nº 400/2016 da ANAC "contempla apenas a prestação de assistência material diante de problemas durante a execução do contrato de transporte aéreo, não havendo previsão de compensação por prejuízos morais eventualmente sofridos pelo passageiro"(Starling; Villa, 2022);
- A Convenção de Montreal também não prevê indenização por dano moral, mas há possibilidade para dano presumido. De acordo com o tratado internacional (capítulo III, artigos 17 a 19), o transportador deve indenizar o passageiro, familiares (no caso de morte) ou o dono da carga, no caso de ter sido efetivamente comprovado o prejuízo sofrido, porém o artigo 19 faz uma ressalva de que ele não será responsabilizado se puder provar que tomou todas as medidas necessárias para prevenir o dano ou se a prevenção do mesmo era impossível;
- Em relação a aplicabilidade ou não do Código de Defesa do Consumidor (CDC), as autoras afirmam que é um equívoco ter como premissa o fato de que a relação entre companhia aérea e passageiro é uma relação de consumo e, portanto, regulada pelo CDC, tendo em vista que há legislação específica sobre o tema. Assim, após embates sobre o tema salientando que a legislação correta a ser aplicada nos processos que discutem as relações de transporte aéreo internacional é a Convenção de Montreal, o Supremo Tribunal Federal apreciou a questão e proferiu decisão, no RE 636.331/RJ, publicando o Tema 210, que dispõe:

“Nos termos do art. 178 da Constituição da República, as normas e os tratados internacionais limitadores da responsabilidade das transportadoras aéreas de passageiros, especialmente as Convenções de Varsóvia e Montreal, têm prevalência em relação ao Código de Defesa do Consumidor”.

De forma complementar, Scaranello, Klarmann e Silva (2022) apresentam que a Resolução nº 400/2016 da ANAC regulou as condições gerais do transporte aéreo e trouxe disposições sobre os deveres das companhias aéreas, inclusive em relação às matérias que os consumidores mais demandam o Poder Judiciário, como alterações nos horários dos voos (artigo 12), atrasos e cancelamentos de voos (artigos 20, 21, 26 e 27), extravio de bagagens (artigos 32 e 33), preterição de embarque (artigo 24) e no-show (artigo 19). Entretanto, apesar dessa regulação, é comum ocorrer decisões judiciais e administrativas ampliando os deveres das companhias aéreas para além do que determina a norma que regulamenta o setor, assim como concedendo indenizações por danos morais em situações já reguladas. Conforme as autoras, isso gera estímulos à judicialização excessiva das companhias aéreas, pois faz com que cada vez mais consumidores busquem algum tipo de reparação perante o Poder Judiciário.

Silveira (2021), em sua dissertação, levantou dois estudos relacionados à tomada de decisão sobre o ajuizamento de uma ação, que apesar de não terem o transporte aéreo como objeto, contribuem com dados relevantes sobre a litigiosidade no Brasil e ratificam

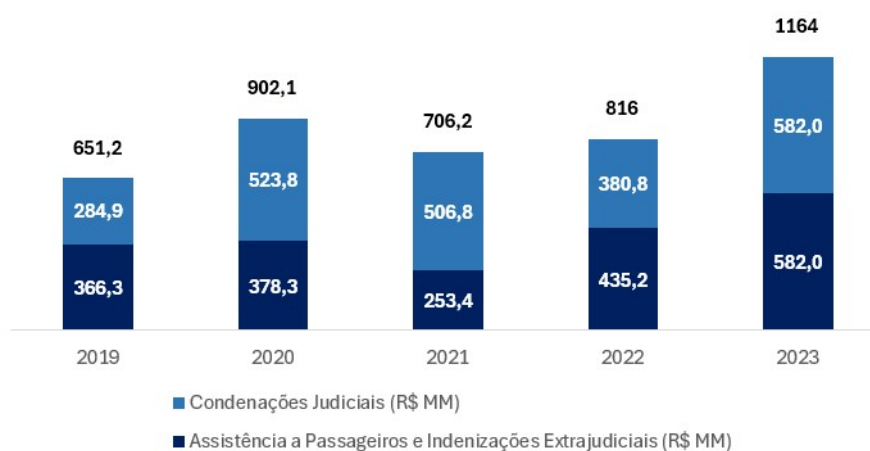
as informações apresentadas acima. O primeiro estudo trata-se da série "O Judiciário destrinchado pelo *Big Data*" (Salama; Carlotti; Yeung, 2018), desenvolvido pela Fundação Getúlio Vargas de São Paulo (FGV-SP), que teve como segundo relatório o tema "Quando litigar vale mais a pena que fazer acordo". Apesar de ter tido uma abordagem direcionada aos processos trabalhistas, essa pesquisa trouxe respostas ao questionamento sobre o que estimula o litígio. Após a análise de 130 mil processos trabalhistas, os resultados da pesquisa mostraram que em 89% dos casos os pedidos dos reclamantes foram atendidos total ou parcialmente, sugerindo uma previsibilidade do autor sobre o resultado da demanda. Além disso, a pesquisa indicou que o baixo custo para litigar, devido à gratuidade judiciária e a alta oferta de advogados, é um incentivo significativo para o litígio, reduzindo o interesse em acordos.

O segundo estudo teve como área temática "Diagnóstico sobre as causas do progressivo aumento das demandas judiciais cíveis no Brasil" (PUCRS, 2011) e foi realizado pela Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), atendendo à demanda do Conselho Nacional de Justiça (CNJ). A pesquisa ressaltou a racionalidade dos agentes na tomada de decisão e identificou quatro motivos principais para o litígio, sendo eles: i) a ausência de custos e riscos associados à litigância; ii) a busca por ganho financeiro por parte do autor; iii) o uso do judiciário para retardar uma responsabilidade; e iv) a percepção de prejuízo moral, físico ou financeiro da parte demandante.

À vista disso, percebe-se que foi estabelecida uma cultura do litígio no país ("indústria da indenização") por diversos motivos como: o acesso ao Judiciário de forma ampla e gratuita; a atuação do Judiciário de forma mais protetiva ao consumidor, muitas vezes impondo obrigações para além daquelas previstas na regulamentação ou até aplicando as legislações existentes sem coerência e unidade; o histórico de concessão de indenização por danos morais de forma punitiva; entre outros.

Essa cultura do litígio, caracterizada pela busca constante por indenizações, tem repercussões diretas sobre os custos operacionais das empresas aéreas. De acordo com os Painéis de Indicadores do Transporte Aéreo da ANAC dos anos de 2019 a 2023, os custos e despesas dos serviços aéreos totalizaram R\$40,7 bilhões, R\$29,1 bilhões, R\$36,2 bilhões, R\$54,4 bilhões e R\$58,2 bilhões, respectivamente. Desse total, os gastos com assistência a passageiros, indenizações extrajudiciais e condenações judiciais representaram 1,6%, 3,1%, 2,1%, 1,5% e 2% respectivamente a cada ano, o que equivale aproximadamente aos valores representados na Figura 1. Em termos absolutos, os os custos com essas despesas alcançaram, por exemplo, mais de R\$1 bilhão em 2023, refletindo o peso da judicialização nas finanças do setor.

Figura 1 – Gastos em milhões (R\$MM) com assistência a passageiros, indenizações extrajudiciais e condenações judiciais



Fonte: Elaborada pela autora com dados da ANAC (2021c, 2021d, 2022b, 2023c, 2024b).

Observa-se na Figura 1 que 2020 apresentou um crescimento de 39% das despesas decorrentes de indenizações em processos judiciais ou administrativos em comparação a 2019. Embora tenha ocorrido uma queda de 16% em 2021, os valores voltaram a crescer nos dois anos subsequentes, com aumentos de 7% em 2022 e de 43% em 2023. Esses gastos com indenizações se aproximaram dos valores das tarifas aeroportuárias, que em 2019 e 2020 representaram 2,8% do total de despesas das empresas, 2,4% em 2021, 2,3% em 2022 e 2% em 2023. É importante observar que as tarifas aeroportuárias brasileiras são algumas das mais altas do mundo, evidenciando o alto custo das indenizações para o setor (Scaranello; Klarmann; Silva, 2022).

Além do impacto financeiro às empresas aéreas, essas altas despesas com indenizações em processos judiciais ou administrativos afeta também os consumidores, pois os custos são repassados aos preços das passagens. Além disso, o volume de processos já causa impacto na malha aérea das empresas. Sob a justificativa de enfrentar um alto número de ações judiciais no Norte do país, as companhias aéreas têm reduzido ou mesmo cortado a oferta de voos entre cidades da região, que já é a menos atendida pelo transporte aéreo no Brasil (Ferraz, 2024).

A Gol e a Azul, por exemplo, reduziram voos em Rondônia com o argumento de um volume desproporcional de ações no estado, visto que só a Azul foi alvo de 15 mil processos entre janeiro de 2022 e junho de 2023 (Barbosa, 2024). O caso de Rondônia chama atenção, visto que o Estado apresentou a passagem aérea mais cara do país, além disso a redução de voos gerou a apresentação de uma ação civil pública por parte da prefeitura de Porto Velho contra a Gol e a Azul em agosto de 2023 e uma denúncia, enviada em dezembro de 2023, pela seccional rondoniense da OAB à Senacon do Ministério da Justiça, que também engloba a Latam. Esse assunto também mobiliza outros Estados da região, como Acre e

Roraima, igualmente afetados (Ferraz, 2024).

Assim, fica evidente que os efeitos da judicialização do transporte aéreo vão além das empresas, afetando também os consumidores e a própria dinâmica do mercado. A cultura do litígio e as altas indenizações impacta diretamente os resultados das empresas do setor, comprometendo a viabilidade da operação. Além disso, ela gera preocupações sobre a sustentabilidade do setor, desestimula investimentos e inibe a entrada de novas empresas no Brasil (Botelho; Costa, 2023).

2.2.2 Aplicação do aprendizado de máquina na judicialização do transporte aéreo

A busca por literatura sobre esse tema foi feita usando Scopus e Web of Science, além de algumas ferramentas de inteligência artificial de busca de pesquisa acadêmica, como o Elicit¹⁰ e o Research Rabbit¹¹. Foram encontrados poucos trabalhos sobre a aplicação de aprendizado de máquina em casos de judicialização no transporte aéreo, como identificado também por Pont *et al.* (2023). Assim, também foram consideradas publicações que tratam do setor judiciário de forma geral.

Os trabalhos encontrados podem ser divididos em dois grupos. O primeiro grupo são trabalhos que pesquisaram a aplicação de abordagens de agrupamento ou *clustering*, com o objetivo de agrupar os processos judiciais em grupos homogêneos (*clusters*), a fim de possibilitar uma análise com maior clareza do conjunto de dados (Sicsú; Samartini; Barth, 2023). Raghuvier *et al.* (2012) utilizaram *Latent Dirichlet Allocation* (LDA) para propor uma agrupamento com base em tópicos. Raghav *et al.* (2015) aplicaram K-means para testar três agrupamentos utilizando citações e links sobre outros julgamentos. Zhang e Zhou (2019) usaram um algoritmo de vetor de parágrafo (Doc2Vec) para atualizar os resultados do agrupamento de documentos jurídicos, comparando a similaridade do texto sem reimplementar o agrupamento. Martino, Pio e Ceci (2023) propuseram um novo método para identificar grupos de julgamentos jurídicos de acordo com suas características, chamado *Multi-view Overlapping cluSTering of legAl judgments* (MOSTA), que é capaz de identificar uma representação fundida que considera tanto o conteúdo textual das sentenças judiciais quanto os atos jurídicos que elas citam, levando em consideração a granularidade das citações. Além desses, os trabalhos de Sabo *et al.* (2022) e Lima e Costa (2022) estudaram o tema considerando dados do contexto brasileiro.

Sabo *et al.* (2022) buscaram agrupar julgamentos judiciais do Juizado Especial Cível localizado na Universidade Federal de Santa Catarina (JEC/UFSC). Foram utilizados dados de 665 sentenças judiciais proferidas entre 1º de janeiro de 2014 e 3 de junho de 2019, nas quais consumidores reivindicavam indenizações morais e materiais às companhias aéreas

¹⁰ <https://elicit.com/>

¹¹ <https://researchrabbitapp.com/>

por falhas no serviço. Os autores aplicaram quatro algoritmos de *clustering*: Hierárquico e Lingo (*clustering* suave), *K-means* e Propagação de Afinidade (*clustering* rígido). A avaliação dos resultados considerou quatro critérios: entropia e pureza; capacidade do algoritmo em fornecer rótulos; avaliação de perito jurídico; e complexidade experimental. Como resultado, os autores identificaram que o *clustering* hierárquico demonstrou ser a abordagem mais vantajosa, porém a tarefa de fornecer os rótulos cabia exclusivamente ao especialista. Por outro lado, o algoritmo lingo demonstrou a capacidade de gerar rótulos, permitindo ao especialista apenas identificar aqueles que não deveriam ser uma variável do julgamento. De forma geral, os autores concluíram que todas as abordagens testadas permitiram identificar padrões, porém com o apoio de especialistas jurídicos, demonstrando uma limitação das técnicas em explicar um evento jurídico.

Lima e Costa (2022) realizaram uma avaliação de seis diferentes abordagens para o agrupamento, também utilizando bases de dados de documentos jurídicos brasileiros. Eles testaram os algoritmos *K-means*, *Mini Batch K-Means* e HDBSCAN em diferentes hiperparâmetros e em conjunto ou não com o Mapa de Kohonen como técnica de pré-clustering. Além disso, o trabalho também propôs uma nova estrutura orientada para o Processamento de Linguagem Natural (PLN) para a avaliação específica de clusters textuais. Os resultados obtidos demonstraram que o *K-means* e o *Mini Batch K-Means* são as melhores escolhas, e o uso do mapa de Kohonen pode aumentar o desempenho geral do *clustering*.

O segundo grupo de trabalhos pesquisaram abordagens de previsão, que envolve a aplicação de modelos estatísticos ou de aprendizado de máquina para prever valores futuros com base em dados históricos, como a previsão de indenizações.

Pont *et al.* (2023) investigaram a aplicação de várias técnicas de PLN e aprendizado de máquina em um *pipeline* para regressão em texto, com o objetivo de prever o valor de compensação por danos imateriais (isto é, danos morais), a partir de processos judiciais em que os consumidores tiveram problemas com companhias aéreas e receberam esse tipo de indenização. O conjunto de dados utilizado era composto por 928 sentenças proferidas entre fevereiro de 2011 e setembro de 2020 no Juizado Especial Estadual localizado na Universidade Federal de Santa Catarina. Para preparar os dados, foi necessária algumas operações de pré-processamento, como tokenização, normalização, filtragem e outras. Ao todo, o conjunto de dados continha um vocabulário de 16.924 palavras, 712.057 tokens no total e uma média de 758 tokens por documento (após a etapa de pré-processamento). Os rótulos (valores de compensação) variaram de R\$304 a R\$25.000, com média de R\$6.344 e desvio padrão de R\$3.471. As variáveis utilizadas pelos autores para o problema de previsão foram: data do julgamento, juiz, tipo de juiz, perda permanente de bagagem, bagagem adulterada, perda temporária de bagagem, intervalo de perda, cancelamento de voo, atraso de voo, intervalo de atraso, condições meteorológicas adversas, culpa do

consumidor, *overbooking*, não comparecimento, direito ao arrependimento e ao pedido de reembolso e *downgrade*.

O trabalho de Pont *et al.* (2023) foi dividido em duas partes. Na primeira, eles avaliaram dois *pipelines* (*pipelines* de linha de base e *pipelines* completos), sendo o primeiro mais simples e o segundo baseado no primeiro, mas com algumas melhorias, chamadas de ajustes (extração de N-gramas, seleção de recursos, adição de AELE, remoção de *outliers*, validação cruzada e prevenção de sobreajuste). Ao todo, os autores aplicaram 11 técnicas de aprendizado de máquina para regressão (*AdaBoost*, *Bagging*, *Decision tree*, *Neural network*, *Elastic net*, *Ensemble voting*, *Gradient boosting*, *Random forest*, *Ridge*, *SVM*, *XGBoosting*) e utilizaram três métricas para avaliar a qualidade da previsão (*Mean Absolute Error* - MAE, *Root Mean Square Error* - RMSE e o coeficiente de determinação - R^2)¹². Na segunda parte, os autores avaliaram ainda mais os ajustes propostos, testando 80 combinações distintas de *pipelines*. As principais conclusões dos autores foram:

- Na primeira parte, o *pipeline* completo alcançou resultados superiores comparado ao *pipeline* de linha de base. A melhor técnica nos experimentos foi o *Ensemble voting* com R^2 de 0,74, RMSE de 1.750 e MAE de 915. Por outro lado, o SVM foi o que teve pior desempenho, por estar subajustado;
- Na segunda parte, algumas combinações de ajustes alcançaram melhor qualidade de previsão do que o *pipeline* completo. O melhor *pipeline* foi o com seleção de recursos, extração de N-gramas, prevenção de sobreajuste e remoção de *outliers* (todos os conjuntos de dados) ativados, enquanto remoção de *outliers* (conjunto de treinamentos), adição de AELE e validação cruzada estavam desativados. A melhor técnica foi novamente o *Ensemble voting* com R^2 de 0,78, RMSE de 1.586 e MAE de 803. E a pior técnica, continuou sendo o SVM.
- Ao testar o impacto de cada ajuste no desempenho, os autores identificaram que a extração de N-gramas e a adição de AELE produzem ganhos na qualidade da predição, mas aumentam ligeiramente o tempo de execução. Por outro lado, a seleção de recursos e a prevenção de sobreajuste, reduz o tempo de execução consideravelmente. Assim, a adoção desses quatro ajustes nos *pipeline* implica em ganhos em termos de qualidade de previsão e tempo de execução. Dessa forma, considerando a qualidade da previsão e o tempo de execução, *pipeline* foi o com os ajustes seleção de recursos, extração de N-gramas, adição de AELE, prevenção de sobreajuste e remoção de

¹² Essas são algumas métricas comuns para problemas de previsão, em que: MAE é o erro médio absoluto e representa a média da diferença absoluta entre os valores reais e os previstos por um modelo; o RMSE é raiz do erro quadrático médio e representa o erro médio das diferenças quadradas entre os valores previstos e reais; e o coeficiente de determinação é um número entre 0 e 1 que mede o quão bem um modelo estatístico prevê um resultado.

outliers (todos os conjuntos de dados) ativados e os restantes desativados. E o melhor modelo foi o *Ensemble voting* com R^2 de 0,76, RMSE de 1.683 e MAE de 866.

- A motivação da pesquisa foi a hipótese de que, ao informar às partes o valor estimado de reparação de danos morais na fase inicial de uma ação judicial, poderia incentivar um acordo entre eles. E de acordo com a avaliação dos resultados por um perito jurídico, as previsões são úteis no ambiente jurídico e podem encorajar as partes envolvidas a um acordo. Tendo em vista que o erro MAE dos melhores *pipelines* foi inferior a mil, o perito acredita que esse valor é quase irrelevante no contexto da compensação legal, ou seja, seria bem aceitável abrir mão de aproximadamente mil reais da indenização em audiências de conciliação, proporcionando o aumento de acordos.

Torres, Guterres e Celestino (2023) procuraram destacar os principais fatores que levam os clientes a tomar medidas judiciais contra companhias aéreas e mostrar o que mais afeta os valores de indenizações, além de identificar as melhores técnicas de aprendizado de máquina para prever os valores indenizados. O estudo foi baseado em dados de ações judiciais ajuizadas em cidades brasileiras, no período de 2016 a 2021. As variáveis utilizadas pelos autores foram: companhia aérea, ano, temporada, região, dano moral, dano material, motivo do processo, causa alegada pela empresa, decisão e valor pago de indenização. Para equilibrar a frequência, os valores de dano moral, dano material e valor pago de indenização foram divididos em três categorias (baixo, médio e alto)¹³, tornando-se variáveis categóricas. Tendo em vista que o rótulo a ser previsto era o valor pago de indenização, diferente de Pont *et al.* (2023), esse trabalho resolveu um problema de classificação.

Assim, Torres, Guterres e Celestino (2023) compararam o desempenho da Regressão Logística Multinomial e de três modelos de aprendizado de máquina (*Random Forest*, SVM e *Naive Bayes*), utilizando as seguintes métricas para avaliar o desempenho: precisão, área sob a curva ROC e matriz de confusão. Entre os modelos analisados, o *Random Forest* teve o melhor desempenho em relação aos modelos de aprendizagem de máquina, seguido pelo SVM e o *Naive Bayes*. Seus valores foram semelhantes à Regressão Logística Multinomial, que se mostrou importante na classificação de conjuntos de dados categóricos. Contudo, apesar da melhor precisão (83,1%), o tempo de processamento do *Random Forest* foi muito elevado, o que pode ser uma desvantagem se o poder computacional for limitado. Assim, a Regressão Logística Multinomial apresentou a melhor relação custo-benefício e forneceu inferências importantes sobre a relação entre as variáveis, como:

- As companhias aéreas compensam de maneira diferente e a época em que ocorre as falhas no serviço não influencia nos valores da indenização;

¹³ Os autores consideraram Baixo \leq R\$1000,00, R\$1000,00 $<$ Médio \leq R\$5000,00 e Alto $>$ R\$5000,00.

- Em 2020, apesar da alta demanda judicial, houve o menor valor de indenização, possivelmente devido à pandemia de COVID-19;
- A região Norte apresentou um comportamento diferenciado do restante do Brasil, muito influenciado por Rondônia, que possui as maiores indenizações devido principalmente ao veredicto do juiz, visto que os motivos e causas das falhas de serviço no Estado são semelhantes aos demais;
- O dano moral é o principal determinante da indenização final;
- Atrasos de voos, cancelamentos e falhas das companhias aéreas têm um efeito negativo nas indenizações;
- Quando parte do pedido de indenização é rejeitada (processo improcedente), a indenização paga pelas companhias aéreas tende a ser menor, como esperado.

Os trabalhos apresentados demonstram como esse tema vêm sendo estudado. O Quadro 3 apresenta um resumo dos objetivos e principais resultados e conclusões de cada estudo.

Quadro 3 – Quadro resumo dos trabalhos de aplicações de aprendizado de máquina nas questões de judicialização do transporte aéreo

Autores	Ano	Periódico	Objetivo	Principais Resultados/Conclusões
Sabo et al.	2022	Artificial Intelligence and Law	Agrupar julgamentos judiciais de consumidores que reivindicaram indenizações de danos morais e materiais às companhias aéreas por falhas no serviço.	Clustering Hierárquico demonstrou ser a abordagem mais vantajosa.
Pont. et al.	2023	PeerJ Computer Science	Investigar técnicas de regressão de texto para prever o valor da compensação de julgamentos legais em que os consumidores tiveram problemas com companhias aéreas e reivindicam danos morais.	O <i>pipeline</i> completo superou o de linha de base, mas algumas combinações de ajustes obtiveram melhor qualidade de previsão. Os ajustes de extração de N-gramas, adição de AELE, seleção de recursos e prevenção de sobreajuste resultaram em melhorias tanto na qualidade das previsões quanto no tempo de execução. A técnica <i>Ensemble voting</i> apresentou o melhor desempenho. E segundo um perito jurídico, as previsões de danos morais são úteis no ambiente jurídico e podem incentivar acordos, pois a diferença seria aceitável em audiências de conciliação.
Torres, Gu- terres e Ce- lestino	2023	Frontiers in Future Transportation	Destacar os principais fatores que levam os clientes a tomar medidas judiciais contra companhias aéreas e mostrar o que mais afeta os valores de indenizações, além de identificar as melhores técnicas de aprendizado de máquina para prever os valores indenizados.	O <i>Random Forest</i> teve o melhor desempenho em relação aos modelos de aprendizagem de máquina, mas a Regressão Logística Multinomial apresentou a melhor relação custo-benefício. Atrasos de voos, cancelamentos e falhas das companhias aéreas têm um efeito negativo nas indenizações. A indenização acima da média é uma tendência em alguns estados, sendo o dano moral concedido aos clientes a principal causa de maiores indenizações.

Fonte: Elaborado pela autora.

3 MÉTODOS E TÉCNICAS DE PESQUISA

3.1 Tipologia e descrição geral dos métodos de pesquisa

3.1.1 *Design Science Research*

O método que foi utilizado para o desenvolvimento deste trabalho é o *Design Science Research* (DSR). Esse método é derivado do paradigma da *Design Science* (DS), ciência que procura desenvolver e projetar soluções visando melhorar sistemas existentes, resolver problemas ou, ainda, criar novos artefatos que contribuam para uma atuação humana mais eficiente seja nas organizações ou na sociedade (Dresch; Lacerda; Junior, 2015).

A DSR gera um conhecimento de natureza prescritiva com o objetivo de contribuir na resolução de problemas reais (Aken; Berends; Bij, 2012 *apud* Cauchick-Miguel, 2019). Ressalta-se que as pesquisas que resultam em uma prescrição têm uma aplicabilidade simplificada, o que facilita sua adoção pelos profissionais nas organizações e pode contribuir para o reconhecimento da relevância dessas pesquisas na prática (Dresch; Lacerda; Junior, 2015).

Dessa forma, a DSR procura, a partir da compreensão do problema, criar e avaliar artefatos que possam modificar situações, melhorando ou tornando-as mais desejáveis. Uma característica fundamental desse tipo de pesquisa é a busca da solução satisfatória para a situação, mesmo que não seja necessariamente a solução ótima. Todavia, deve ser possível generalizar as soluções desenvolvidas para uma determinada classe de problemas, permitindo que outros pesquisadores e profissionais, em diferentes situações, também possam aplicar o conhecimento gerado (Dresch; Lacerda; Junior, 2015). Neste trabalho, procura-se a solução para o problema relacionado ao número elevado de reclamações de consumidores e de processos judiciais contra empresas de transporte aéreo no Brasil. Os resultados gerados poderão ser aplicados nesse setor, assim como generalizados para outras organizações ou áreas que lidem com problemas similares.

Por fim, define-se artefato como "a organização dos componentes do ambiente interno para atingir objetivos em um determinado ambiente externo"(Dresch; Lacerda; Junior, 2015, p. 107). Segundo Cauchick-Miguel (2019), os artefatos podem ser classificados em constructos, modelos, métodos, instanciações ou *design propositions*, sendo que este trabalho foca na construção de um modelo, aplicando aprendizado de máquina. O modelo é um "conjunto de elementos e relações que representam a estrutura geral da realidade"(Cauchick-Miguel, 2019, p. 112). Na DSR, a preocupação principal em relação aos modelos está na sua utilidade e não na sua aderência a realidade. Assim, o modelo pode

ser impreciso sobre os detalhes da realidade, mas é essencial que ele seja capaz de capturar a estrutura geral da realidade para garantir sua utilidade (Dresch; Lacerda; Junior, 2015).

3.2 Procedimentos técnicos para a realização da pesquisa

Para a realização desta pesquisa foi utilizado o modelo de DSR proposto por Peffers *et al.* (2007), que segundo Brocke, Hevner e Maedche (2020) é amplamente referenciado. Ele é composto por seis etapas, sendo elas: i) identificação e motivação do problema, ii) definição dos objetivos de uma solução, iii) *design* e desenvolvimento, iv) demonstração, v) avaliação e vi) comunicação.

A primeira etapa refere-se à atividade de definir um problema específico de pesquisa e justificar o valor de uma solução (Brocke; Hevner; Maedche, 2020). Como evidenciado anteriormente na introdução e no referencial teórico, o problema estudado neste trabalho encontra-se na quantidade de reclamações e judicialização do transporte aéreo no Brasil, o que gera diversos impactos para as companhias aéreas e, como consequência, aos próprios consumidores. Dessa forma, uma solução justifica-se visto que um diagnóstico e modelo desses temas contribuirá ao gerar informações capazes de auxiliar as empresas a prever tais ocorrências e atuar de forma a mitigar possíveis danos.

Na segunda etapa, o objetivo da solução é inferido a partir da definição do problema e do conhecimento do que é possível e viável (Brocke; Hevner; Maedche, 2020). Dessa forma, com base no problema delineado e no que foi levantado de aplicações anteriores, foram definidos os objetivos específicos deste trabalho, que incluem: identificar os principais motivos e causas de reclamações de consumidores e processos judiciais contra companhias aéreas, analisando diferenças entre o período pré, durante e pós pandemia do COVID-19 (OE 1); identificar a correlação entre os motivos de reclamação de consumidores e de processos judiciais contra companhias aéreas (OE 2); e padronizar a classificação de motivos e causas dos processos judiciais, a fim de avaliar os efeitos no desempenho dos modelos de aprendizado de máquina de previsão do valor de indenização (OE 3).

A terceira etapa inclui a criação de um artefato em si (Brocke; Hevner; Maedche, 2020). O artefato deste trabalho consiste no método de padronização das classificações de motivos e causas e no modelo baseado em aprendizado de máquina para prever o valor total de condenação.

Na quarta e quinta etapa, deve-se demonstrar o uso do artefato para resolver uma ou mais instâncias do problema e medir quão bem o artefato suporta uma solução para o problema (Brocke; Hevner; Maedche, 2020). Nesse sentido, foram apresentados os resultados dos modelos desenvolvidos e avaliadas as métricas de desempenho alcançadas, verificando se a padronização dos motivos e causas dos processos contribuiu para melhorar o desempenho dos modelos na previsão do valor total de condenação.

Por fim, na sexta etapa, os aspectos do problema e do artefato projetado são comunicados às partes interessadas relevantes (Brocke; Hevner; Maedche, 2020). A apresentação e publicação deste trabalho servirá para essa finalidade.

3.3 Caracterização do objeto de estudo

O objeto de estudo deste trabalho é o setor de transporte aéreo no Brasil. Tal setor é essencial para o desenvolvimento econômico e social de um país, visto que possibilita conexões rápidas, facilita o deslocamento de pessoas e bens, permite o fluxo de agentes de negócios e impulsiona as atividades comerciais e o turismo (ONTL, 2022). Dessa forma, o setor aéreo é de extrema importância para a economia. Em 2022 no Brasil, ele representou R\$14,2 bilhões para o Produto Interno Bruto (PIB), de forma direta (aéreas), além de contribuir com a geração de 109,6 mil empregos de forma indireta (aeroportos, Esatas, entre outros) e recolher R\$5,1 bilhões de tributos (ABEAR, 2024).

Segundo Oliveira (2009), esse setor é composto por uma multiplicidade de produtos e fornecedores, que podem ser configurados nos seguintes grupos:

"Quem transporta: empresas aéreas regulares, empresas aéreas de fretamento, empresas de táxi aéreo, etc. (operadoras de aviação comercial), pessoas físicas (aviação geral, com aeronaves de menor porte), instituições militares (aviação militar).

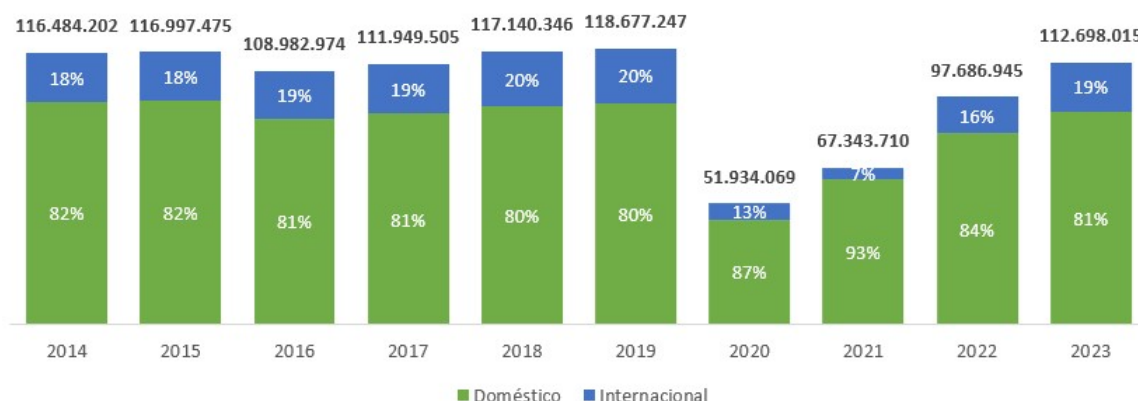
O que transporta: passageiros, carga e correio.

De onde para onde: doméstico (origem e destino no País), regional (origem e destino no País, mas envolve não ligações principais ou "tronco"), internacional (origem ou destino fora do País)." (Oliveira, 2009, p. 31)

Assim, esse trabalho foca no transporte aéreo de passageiros, principalmente nas empresas aéreas regulares de passageiros que atuam no mercado doméstico e internacional no Brasil. Esse setor apresentou um crescimento surpreendente ao longo das últimas décadas. Um exemplo disso é a evolução do número de passageiros pagos, conforme demonstra a Figura 2.

Na Figura 2, nota-se um aumento na quantidade de passageiros pagantes transportados no Brasil até 2015, um crescimento que vinha sendo constante desde 2010. Em 2016, a recessão interna do país gerou uma queda, principalmente em relação aos passageiros domésticos. A partir de 2017 até 2019, o mercado recuperou o volume de passageiros, apresentando crescimento no volume doméstico e internacional. A expressiva retração a partir de 2020 foi decorrente da pandemia de COVID-19, que impôs restrições por diversos órgãos, impactando principalmente o segmento internacional. Apesar disso, a partir do ano de 2021, é possível identificar um aumento crescente no volume de passageiros (ONTL, 2022).

Figura 2 – Evolução anual do número de passageiros pagos - mercados doméstico e internacional



Fonte: Elaborado pela autora com dados da ANAC (2024a).

Além do número de passageiros, outras duas métricas demonstram o crescimento do setor, sendo elas:

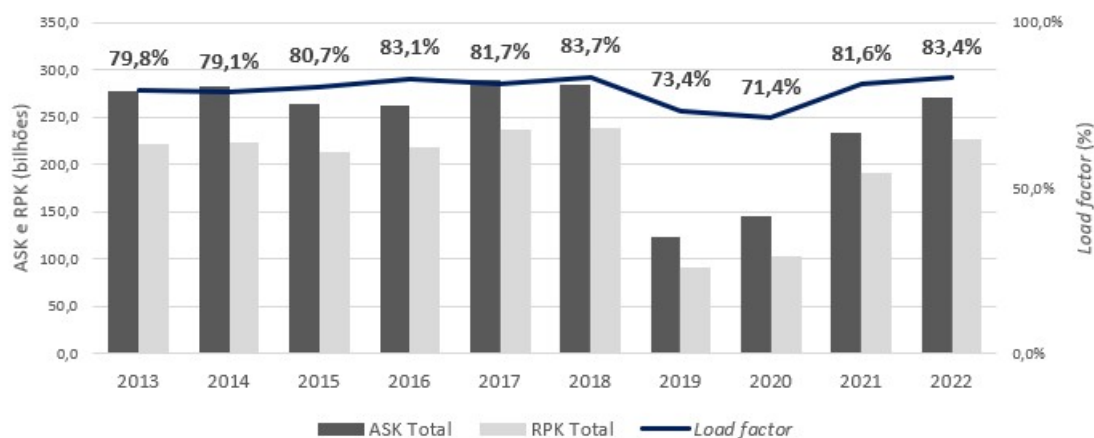
- **ASK:** que significa *available seats-kilometers* ou assentos-quilômetros oferecidos, e é a medida da oferta, mensurando a capacidade produtiva de transporte das companhias aéreas;
- **RPK:** que significa *revenue passenger-kilometres* ou passageiros-quilômetros pagos transportados, e é a medida da demanda, sendo um indicador de tráfego e fluxo de passageiros.

A fração entre o RPK e o ASK resulta no indicador fator de aproveitamento (*load factor*) ou taxa de ocupação das aeronaves, que é usualmente associado à eficiência das empresas aéreas (Godoi, 2023). A Figura 3 apresenta os valores desses indicadores nos últimos anos.

Observa-se na Figura 3 que o mercado brasileiro de transporte aéreo de passageiros apresentou taxas de crescimento significativas, com ocupação média das aeronaves alcançando mais de 80%. Isso permite concluir que a demanda vem apresentando um crescimento maior que a oferta, resultando em maior aproveitamento nos voos (Gomes; Fonseca, 2014).

Por fim, sabe-se que muitas empresas atuam nesse setor de transporte de passageiros, porém a atuação de algumas é mais relevante no Brasil. Conforme o Painel de Indicadores do Transporte Aéreo de 2023, a demanda doméstica foi atendida, principalmente, pelas empresas Latam (37,8%), Gol (33,3%) e Azul (28,4%), sendo os 0,5% restantes atendidos por outras. Enquanto a demanda internacional foi atendida pela Latam (18,1%), TAP (10,6%), Azul (6,2%), American Airlines (5,2%), Air France (5,1%) e demais estrangeiras (54,8%) (ANAC, 2024b).

Figura 3 – Evolução da oferta (ASK), da demanda (RPK) e do aproveitamento (LF) - mercados doméstico e internacional



Fonte: Elaborado pela autora com dados da ANAC (2024a).

3.4 População e amostra da pesquisa

A população deste trabalho foi composta por consumidores de companhias aéreas nacionais e estrangeiras que operam no Brasil. A amostra foi formada por consumidores de 16 estados brasileiros, distribuídos nas seguintes regiões: Norte (AM, RO e TO), Nordeste (BA, CE e PE), Centro-Oeste (DF, GO, MS e MT), Sudeste (ES, MG, RJ e SP) e Sul (PR e SC). Esses consumidores registraram reclamações na Plataforma Consumidor.gov.br e entraram com processos judiciais contra seis empresas: três nacionais (Azul Linhas Aéreas, Gol Linhas Aéreas e Latam Airlines) e três internacionais (United Airlines, Copa Airlines e TAP Air Portugal). A amostra incluiu 316.799 reclamações e 555.546 processos judiciais, referentes ao período de 2019 a 2023.

A delimitação das Unidades Federativas, das empresas e do período analisado foi determinada pela disponibilidade de dados fornecidos pelas próprias companhias, que participaram do projeto de pesquisa "Diagnóstico sobre a Judicialização do Transporte Aéreo no Brasil: Uma Aplicação de Aprendizado de Máquina"¹, que serviu como fonte de dados para este estudo. Dessa forma, a amostra foi não probabilística, obtida por conveniência.

¹ Projeto de pesquisa iniciado em 05 de dezembro de 2023, com vigência até 05 de dezembro de 2024, decorrente de uma parceria da Associação Brasileira das Empresas Aéreas (ABEAR), da Associação Internacional de Transportes Aéreos (IATA), da Associação Latino-Americana e do Caribe de Transporte Aéreo (ALTA), da Junta de Representantes das Companhias Aéreas Internacionais do Brasil (JURCAIB), da Associação dos Magistrados Brasileiros (AMB) e da Universidade de Brasília (UnB).

3.5 Procedimento de coleta de dados

Os dados coletados para a realização deste trabalho foram derivados de análise documental e de fontes secundárias. Duas fontes de dados foram utilizadas, sendo uma relacionada a cada tema: reclamações de consumidores de companhias aéreas e judicialização do transporte aéreo.

3.5.1 *Dados de reclamações de consumidores de companhias aéreas*

Os dados relacionados às reclamações de consumidores de companhias aéreas foram extraídos das bases de dados do Consumidor.gov.br, disponibilizadas pela ANAC². O Consumidor.gov.br é uma plataforma pública que facilita a resolução de conflitos entre consumidores e empresas, sem substituir os canais tradicionais de atendimento, como Procons e Juizados Especiais Cíveis. Desde 2019, essa plataforma foi adotada pela ANAC como sistema eletrônico de atendimento de reclamações contra empresas aéreas, obrigando-as a aderirem e estarem ativas na plataforma. Assim, embora ela não seja o único canal disponível para os consumidores registrarem reclamações, é o canal adotado e fiscalizado pela agência reguladora do setor aéreo. O Apêndice B apresenta mais informações sobre o Consumidor.gov.br.

As bases de dados coletadas continham 28 variáveis que apresentam informações das reclamações registradas na plataforma, assim como sobre o atendimento prestado pelas empresas. Foram extraídas as bases de dados de reclamações referentes aos anos de 2019 a 2023, abrangendo assim o mesmo período temporal dos dados de judicialização. No total, as bases apresentavam 432.767 registros. A Tabela 1 apresenta a descrição de cada uma dessas variáveis. É importante ressaltar que a classificação das reclamações utilizadas no Consumidor.gov.br (variáveis Área, Assunto, Grupo Problema e Problema), por ser mais genérica, ainda não reflete por completo as peculiaridades do setor de aviação civil. Por isso, a ANAC elaborou e utiliza uma classificação própria (variável Código classificador ANAC), conforme as especificações elencadas e exemplificadas na Portaria nº 764/SAS, de 14 de março de 2019.

Tabela 1 – Variáveis das bases de dados do Consumidor.gov.br

Variável	Descrição
Gestor	Nome da entidade de defesa do consumidor, responsável pela gestão da reclamação do consumidor.
Região	Sigla da região geográfica do consumidor.

² <https://www.anac.gov.br/aceso-a-informacao/dados-abertos/areas-de-atuacao/voos-e-operacoes-aereas/dados-do-consumidor-gov>

Variável	Descrição
UF	Sigla do estado do consumidor.
Cidade	Município do consumidor.
Ano Abertura	Ano de abertura da reclamação pelo consumidor.
Mês Abertura	Número do mês de abertura da reclamação pelo consumidor.
Data Abertura	Data de abertura da reclamação pelo consumidor.
Data e Hora Resposta	Data e hora da resposta da reclamação pela empresa.
Data e Hora Análise	Data e hora da análise da recusa da empresa pelo Gestor da reclamação.
Data e Hora Recusa	Data e hora do registro da recusa pela empresa e envio da reclamação para análise do Gestor.
Data Finalização	Data de finalização da reclamação.
Prazo Resposta	Data limite para resposta da empresa. Caso a reclamação tenha sido recusada pela empresa e encaminhada para análise do Gestor, o prazo se altera, considerando o tempo que a reclamação tenha ficado em análise pelo gestor.
Prazo Análise Gestor (em dias)	Número de dias levados para a análise da recusa da empresa pelo Órgão Gestor da reclamação. Diferença, em dias, entre a Data Análise e a Data Recusa.
Tempo Resposta (em dias)	Número de dias para a resposta da reclamação, entre a Data de Resposta e a Data de Abertura, desconsiderando o tempo que a reclamação tenha ficado em análise pelo Gestor (se for o caso).
Nome Fantasia	Nome pelo qual a empresa reclamada é conhecida no mercado.
Área	Área à qual pertence o assunto objeto da reclamação.
Assunto	Assunto objeto da reclamação.
Grupo Problema	Agrupamento do qual faz parte o problema classificado na reclamação.
Problema	Descrição do problema objeto da reclamação.
Forma Contrato	Descrição do meio utilizado para contratação/aquisição do produto ou serviço reclamado.
Procurou Empresa	Sigla da resposta do consumidor à pergunta: "Procurou a empresa para solucionar o problema?" (S = Sim e N = Não).
Respondida	Sigla que indica se a empresa respondeu à reclamação ou não (S = Sim e N = Não).

Variável	Descrição
Situação	Situação atual da reclamação no sistema (Finalizada não avaliada e Finalizada avaliada).
Avaliação Reclamação	Classificação atribuída pelo consumidor sobre o desfecho da reclamação (Não Avaliada, Não Resolvida, Resolvida).
Nota do Consumidor	Número da nota de 1 a 5 atribuída pelo consumidor ao atendimento da empresa.
Edição de Conteúdo	Sigla que indica se houve ou não edição de conteúdo da reclamação pelo Órgão Gestor (S = Sim e N = Não).
Interação do Gestor	Sigla que indica se houve ou não interação do Gestor com a empresa ou com o consumidor ao longo da reclamação (S = Sim e N = Não).
Código classificador ANAC	Código classificador informado pela empresa aérea ao término do texto da resposta final ao consumidor, conforme as especificações elencadas e exemplificadas na Portaria nº 764/SAS, de 14 de março de 2019.

Fonte: ANAC (2021b).

3.5.2 Dados de judicialização do transporte aéreo

Os dados relativos à judicialização do transporte aéreo foram obtidos a partir do projeto de pesquisa 'Diagnóstico sobre a Judicialização do Transporte Aéreo no Brasil: Uma aplicação de Aprendizado de Máquina', da qual a autora também foi pesquisadora. Dessa forma, foi obtida a permissão do uso dos dados no presente trabalho como continuação dos desdobramentos da parceria de pesquisa. As bases de dados deste projeto foram disponibilizadas em planilhas Excel pelas próprias companhias aéreas (Azul Linhas Aéreas, Gol Linhas Aéreas, Latam Airlines, United Airlines, Copa Airlines e TAP Air Portugal).

As bases de dados coletadas continham 24 variáveis que apresentam informações dos processos judiciais movidos contra essas empresas aéreas no período de 2019 a 2023, considerando apenas 16 estados brasileiros (AM, BA, CE, DF, ES, GO, MG, MS, MT, PE, PR, RJ, RO, SC, SP e TO). No total, as bases apresentavam 555.625 registros. A Tabela 2 apresenta a descrição de cada uma dessas variáveis.

Tabela 2 – Variáveis das bases de dados de judicialização

Variáveis	Definição
empresa_re	Nome da empresa ré no processo.
cnpj_re	CNPJ da empresa ré no processo.
numero_processo	Número do processo judicial.

Variáveis	Definição
numero_processo_regex	Número do processo judicial sem formatação, apenas números.
num_proc_ano	Ano de distribuição do processo.
data_distribuiçao	Data de distribuição do processo.
UF	Unidade da Federação onde o processo foi distribuído.
jurisdicao	Classificação do tribunal onde tramita o processo.
n_autores	Quantidade de autores do processo.
data_voo	Data em que o voo ocorreu.
operacao	Tipo de operação da empresa ré, isto é, informa se o voo foi doméstico ou internacional.
obj_principal	Objeto principal do processo, isto é, a motivação do(s) autor(es).
subobj_principal	Subobjeto principal do processo, complementando o objeto principal.
causa_alegada	Causa alegada pela empresa ré, isto é, a defesa ou justificativa sobre a reclamação feita contra ela.
valor_pedido_inicial	Valor do pedido inicial (R\$) de indenizações do processo.
valor_dano_moral	Valor do dano moral (R\$) pedido no processo, isto é, prejuízos não materiais, como sofrimento emocional ou psicológico.
valor_dano_material	Valor do dano material (R\$) pedido no processo, isto é, perdas financeiras diretas.
valor_total_condenacao	Valor total da condenação (R\$) do processo, isto é, o valor total que a parte ré deve pagar ao(s) autor(es) ao final do processo, somando todas as indenizações.
valor_pago	Valor de indenização efetivamente pago (R\$) ao(s) autor(es) do processo.
situacao_processual	Situação processual do processo, isto é, status atual do processo no sistema judicial, indicando em que fase ou estágio o processo se encontra.
resultado_decisao_sentenca	Resultado da decisão proferida na sentença do processo. Por exemplo, se foi favorável ou desfavorável ao(s) autor(es) ou réu.
instancia	Instância da sentença do processo, isto é, nível do judiciário em que a sentença ou decisão foi proferida.
data_sentenca	Data em que a sentença foi proferida pelo juiz.

Variáveis	Definição
data_acordao	Data em que o acórdão foi publicado, que é a decisão proferida por um tribunal de instância superior em revisão de uma sentença anterior.

Fonte: Elaborado pela autora.

3.6 Procedimentos de análise de dados

A análise dos dados coletados foi conduzida por meio do Jupyter Notebook, utilizando a linguagem de programação Python e suas bibliotecas. Todos os arquivos estão disponíveis em uma pasta no GitHub e podem ser acessados por meio deste link. O Quadro 4 sistematiza, por objetivo específico, as técnicas de análise de dados que serão aplicadas.

Quadro 4 – Técnicas de análise de dados

OE	Técnica de análise de dados
1 e 3	Análise exploratória de dados
2	Análise de correlação
3	Desenvolvimento e avaliação dos modelos de aprendizado de máquina

Fonte: Elaborado pela autora.

3.6.1 Análise exploratória de dados

A análise exploratória de dados (AED) tem como objetivo verificar os dados para qualquer aplicação estatística. Dessa forma, obtêm-se entendimento sobre os dados coletados e, principalmente, sobre as relações existentes entre as variáveis. Essa análise geralmente é dividida nas seguintes etapas: coleta, organização, tratamento, análise, apresentação e interpretação dos dados (Ferreira *et al.*, 2021). Para realizar essa análise, foram utilizadas as bibliotecas Pandas, Numpy, Matplotlib e Seaborn do Python. Os arquivos com os códigos em Python podem ser acessados por meio deste link.

3.6.1.1 Coleta e organização de dados

Na etapa de coleta de dados, foram definidos os métodos de coleta, conforme apresentado na seção 3.5. A organização dos dados, por sua vez, consistiu em estabelecer a maneira de extrair informações relevantes dos dados coletados (Ferreira *et al.*, 2021). Assim, foram analisadas as variáveis disponíveis em ambas as bases de dados e suas descrições, com o objetivo de identificar aquelas úteis para os objetivos deste estudo.

Quanto aos dados do Consumidor.gov.br, as seguintes variáveis foram excluídas da análise:

- **Cidade:** porque a análise concentrou-se nas regiões e Unidades Federativas (UF);
- **Data Abertura, Data e Hora Resposta, Data e Hora Análise, Data e Hora Recusa, Hora Finalização, Prazo Resposta, Prazo Análise Gestor (em dias), Edição de Conteúdo e Interação do Gestor:** por serem informações específicas da tratativa da reclamação, visto que a análise foi direcionada ao ano e mês de abertura das reclamações (Ano Abertura e Mês Abertura);
- **Área, Assunto, Grupo Problema e Problema:** por serem informações muito genéricas que não refletem as peculiaridades do setor de aviação civil. No lugar delas, a variável Código classificador ANAC foi utilizada para analisar os motivos das reclamações.

Em relação aos dados de judicialização, foram desconsideradas as seguintes variáveis:

- **numero_processo e numero_processo_regex:** para preservar a confidencialidade dessas informações sensíveis;
- **cnpj_re:** porque foi considerado o nome da empresa (empresa_re) para a identificação das companhias aéreas;
- **data_sentenca e data_acordao:** por serem informações específicas da tramitação do processo judicial, visto que a análise foi direcionada à data de início do processo (data_distribuiçao);
- **situacao_processual e resultado_decisao_sentenca:** porque as informações dessas variáveis foram agrupadas em uma nova variável denominada situacao_processual_sentenca.

Além disso, foram identificadas as variáveis comuns nas duas bases de dados que seriam úteis para a análise de correlação. O Quadro 5 apresenta as variáveis correspondentes em cada base.

Quadro 5 – Variáveis comuns nas duas bases de dados

Variáveis do Consumidor.gov.br	Variáveis da judicialização
UF	UF
Ano Abertura	num_proc_ano
Mês Abertura	data_distribuiçao
Nome Fantasia	empresa_re
Código classificador ANAC	obj_principal
	subobj_principal

Fonte: Elaborado pela autora.

3.6.1.2 Tratamento de dados

A etapa de tratamento de dados teve como objetivo melhorar e adequar o conjunto de dados obtidos (Ferreira *et al.*, 2021). Para isso, as bases de dados foram tratadas e ajustadas, conforme as necessidades observadas.

Considerando a diversidade de termos presentes nas descrições de algumas variáveis, foram construídos dicionários de dados para padronizar os valores em categorias semelhantes, com o objetivo de uniformizar os dados e possibilitar uma análise mais eficaz. Todos os tratamentos realizados estão detalhados nos tópicos 3.6.1.2.1 e 3.6.1.2.2 a seguir.

3.6.1.2.1 Tratamento dos dados de judicialização

O tratamento dos dados de judicialização envolveu um pré processamento³ com o objetivo de padronizar os nomes das colunas e das categorias das variáveis, além de agrupar todas as bases disponibilizadas pelas empresas aéreas em um arquivo único.

Primeiramente, foram definidas diversas funções para a reclassificação dos dados em cada coluna da base de dados. Na coluna 'empresa_re', os nomes das empresas aéreas foram reclassificados apenas para Azul, Copa, Gol, Latam, Tap e United. Na coluna 'cnpj_re', os campos vazios foram substituídos por 'Não informado' e o tipo da coluna foi alterado para *string*, que é um tipo de dado em programação que representa uma sequência de caracteres.

As colunas de data ('data_distribicao', 'data_voo', 'data_setenca', 'data_acordao') também tiveram o tipo alterado para *string* e os campos vazios foram substituídos por uma *string* vazia (' '). Nessas colunas, também foi aplicada uma etapa para extrair apenas os dez primeiros caracteres, a fim de considerar apenas a parte da data em registros que tenham sido informados data e hora (formato *timestamp*), e as datas foram formatadas no padrão ano, mês e dia (AAAA-MM-DD). Na coluna 'UF', os nomes das Unidades Federativas foram reclassificados apenas para as siglas correspondentes (AC, AL, AP, AM, BA, CE, DF, ES, GO, MA, MS, MT, MG, PA, PB, PR, PE, PI, RJ, RN, RS, RO, RR, SC, SP, SE e TO) e os campos vazios foram substituídos por 'Não informado'.

Os valores da coluna 'jurisdicao' foram reclassificados em quatro categorias: 'Juizado Especial Cível', 'Vara Cível', 'Não informado' e 'Outros', que abrange registros não contemplados nos grupos anteriores⁴. Já na coluna 'n_autores', os valores nulos, vazios ou inválidos foram substituídos por zero. Nos registros em que os nomes dos autores foram informados, os nomes foram separados por barras, calculando o número de autores com base na contagem de barras com a adição de 1. Por fim, o tipo da coluna foi alterado para

³ Arquivo 'AED_pre_processamento_dados_judicializacao_vfinal.ipynb' disponibilizado no link.

⁴ O dicionário completo da jurisdicao pode ser acessado por meio deste link.

inteiro, que é um tipo de dado numérico que representa números inteiros, ou seja, números sem parte decimal.

Os valores da coluna 'operacao' deveriam informar se o voo foi Doméstico ou Internacional. Entretanto, em algumas bases recebidas esse campo informava o código IATA dos aeroportos. Dessa forma, esses campos foram reclassificados em quatro categorias: 'Doméstico' (quando possuíam apenas siglas de aeroportos nacionais), 'Internacional' (quando possuíam alguma sigla de aeroporto internacional), 'Não informado' e 'Outros'⁵.

Nas colunas de valores de indenização ('valor_pedido_inicial', 'valor_dano_moral', 'valor_dano_material', 'valor_total_condenacao' e 'valor_pago'), valores inválidos ou ausentes foram substituídos por zero, os espaços em branco, símbolos monetários e traços foram removidos, e o formato dos números decimais foi ajustado. Além disso, foram tratadas somas de valores representados como *strings*. Por fim, os valores foram convertidos para o tipo numérico *float*, que é um tipo de dado numérico que representa números reais, ou seja, números que podem ter parte inteira e parte decimal.

Os dados das colunas 'situacao_processual' e 'resultado_decisao_sentenca' foram agrupados em uma nova coluna chamada 'situacao_processual_sentenca'. Esses dados foram analisados em conjunto e reclassificados em seis categorias: 'Acordo', 'Condenação', 'Decisão favorável à aérea (improcedência)', 'Extinção sem julgamento do mérito', 'Não informado' e 'Outros'⁶. Quanto aos valores da coluna 'instancia', foram reclassificados em quatro categorias: '1ª Instância', '2ª Instância', '3ª Instância' e 'Não informado'.

Para a reclassificação dos valores das colunas a respeito dos motivos e das causas dos processos judiciais ('obj_principal', 'subobj_principal' e 'causa_alegada'), foi aplicada uma técnica de agrupamento desenvolvida no projeto de pesquisa 'Diagnóstico sobre a Judicialização do Transporte Aéreo no Brasil: Uma Aplicação de Aprendizado de Máquina', do qual a autora deste trabalho participou. A técnica proposta adotou critérios de inclusão e exclusão baseados em palavras-chave, semelhantes aos utilizados na análise de conteúdo e em revisões sistemáticas da literatura, com o intuito de agrupar os processos judiciais relacionados ao mesmo tema. Assim, os pesquisadores estabeleceram novas categorias de motivos e causas, além de definir as palavras-chave para a inclusão dos processos em cada grupo. A quantidade de categorias foi definida com o auxílio de duas advogadas e pesquisadoras que atuam em processos judiciais contra empresas aéreas. Assim, foram propostas quatro categorias para os motivos e sete para as causas, de modo a separar adequadamente os processos em grupos representativos do problema, sem gerar uma segmentação excessiva que prejudicasse a análise.

Em relação aos motivos, os processos foram reclassificados em quatro categorias: 'Problemas Operacionais', 'Bagagem', 'Contrato' e 'Outros', em que ficam aqueles que não

⁵ O dicionário completo da operacao pode ser acessado por meio deste link.

⁶ O dicionário completo da situacao_processual_sentenca pode ser acessado por meio deste link.

se enquadram nas anteriores. Para fazer essa reclassificação, foram utilizadas as variáveis 'obj_principal' e 'subobj_principal', de forma conjunta. Dessa forma, foi feita a análise do texto presente nessas duas variáveis, verificando a presença ou não das palavras-chave definidas. Por exemplo, caso o texto do objeto e subobjeto apresentassem a palavra-chave 'acidente', o processo judicial receberia a classe 'problemas operacionais'. Ressalta-se que o processo judicial pode ter recebido mais de uma classe, tendo em vista que é comum os processos tratarem de diversos assuntos. A Tabela 6 apresenta as categorias e as listas de palavras-chave definidas para fazer a reclassificação dos motivos.

Quadro 6 – Lista de palavras-chave por categoria para agrupamento dos motivos

Categorias	Palavras-chave
PROBLEMAS OPERACIONAIS	'acidente', 'aeroporto', 'aeronave', 'alteração', 'assento', 'assiatance', 'assistência', 'atendimento', 'atr.', 'atraso', 'avião', 'cadeira de rodas', 'cambio', 'canc.', 'cancel', 'catering', 'check in', 'conex', 'demora', 'denied', 'delay', 'desembarque', 'downgrade', 'downgrade', 'embarque', 'espaço', 'falha', 'flight', 'fretamento', 'incidente', 'malha', 'manutenção', 'menor desacompanhado', 'overbooking', 'dano a passageiro', 'poltrona', 'pontualidade', 'problemas operacionais', 'refeição', 'remarcação', 'segurança', 'service', 'serviço', 'schedule', 'sobreventa', 'solicitação especial', 'tripulação', 'voo'
BAGAGEM	'anima', 'bagagem', 'baggage', 'carg', 'equipaje', 'extravio'
CONTRATO	'bilhete', 'boleto', 'cadastro', 'call center', 'clube', 'cobrança', 'contrato', 'corona', 'covid', 'diverg', 'document', 'economy', 'fee', 'fidelidade', 'fraud', 'flyer', 'milhas', 'multa', 'multiplus', 'show', 'pass', 'plusgrade', 'políticas', 'pontuação', 'prescrição', 'producto', 'programa', 'promoç', 'propaganda', 'reembolso', 'regra', 'reimbursement', 'requisitos de viagem', 'reserva', 'seguro', 'site', 'tarifa', 'taxa', 'upgrade', 'prorrogação de voucher', 'viag'

Fonte: Elaborado pela autora.

A mesma lógica foi aplicada para a reclassificação dos dados de causa, sendo definidas sete categorias: 'Aérea', 'Cliente', 'Terceiros', 'Casos Fortuitos', 'Força Maior', 'Não informado' e 'Outros'. Para fazer essa reclassificação, foram utilizadas as variáveis 'obj_principal', 'subobj_principal' e 'causa_alegada', de forma conjunta. A Tabela 7 apresenta as categorias e as listas de palavras-chave definidas para fazer a reclassificação de causas.

Quadro 7 – Lista de palavras-chave por categoria para agrupamento das causas

Categorias	Palavras-chave
AÉREA	'abastecimento', 'aeronave', 'ajuda técnica', 'bagagem', 'cancelamento indevido', 'carga', 'cargos', 'cco', 'clube tudoazul', 'cobrança equivocada', 'conexão', 'culpa tam', 'divergência de informaç', 'empresa', 'entrega', 'falha', 'impedimento', 'malha', 'manutenção', 'oferta não cumprida', 'operacional', 'overbooking', 'overload', 'pontuação não creditada', 'planejamento', 'reembolso', 'regra', 'reserva', 'segurança', 'serviço', 'transporte terrestre', 'tripula', 'voo'
CLIENTE	'ausência de responsabilidade', 'autor', 'bagagem de mão', 'bagagem entregue dentro do prazo legal', 'comprovação', 'cliente', 'compra não reconhecida', 'direito de arrependimento', 'documentação', 'exclusiva', 'inconformidade', 'não verificada ocorr', 'no show', 'pax', 'pontos expirados', 'prescrição'
TERCEIROS	'aeroporto', 'agência', 'externo', 'fraude', 'ilícito', 'imigração', 'terceiro', 'tráfego'
CASOS FORTUITOS	'fortuito'
FORÇA MAIOR	'condições climáticas', 'corona', 'covid', 'força maior', 'force majeure', 'tempo', 'meteorologia', 'weather'
NÃO INFORMADO	'não aplicável', 'não informado', 'informação inexistente', 'pi não apurado', 'subsídios negativos'

Fonte: Elaborado pela autora.

Por fim, foram adicionadas onze variáveis booleanas à base de dados de judicialização, sendo quatro relacionadas aos motivos e sete às causas. Nessas variáveis, o valor *True* ou 1 indica que o processo pertence ao grupo correspondente, enquanto *False* ou 0, indica que ele não pertence.

Após todo esse pré processamento, as colunas desnecessárias foram excluídas, sendo elas: 'numero_processo', 'cnpj_re', 'situacao_processual', 'resultado_decisao_sentenca', 'data_sentenca', 'data_acordao', 'numero_processo_regex'. As linhas em que o nome da empresa ré estava como 'Não informado' também foram excluídas. As colunas 'mes_distribuicao' e 'regiao' foram criadas a partir das colunas 'data_distribuicao' e 'UF', respectivamente. Alguns valores negativos que existiam nas colunas 'valor_pedido_inicial' e 'valor_pago' foram substituídos por zero, tendo em vista que devem ter sido erros de registro. E optou-se por agrupar os dados classificados como 'Não informado' na categoria 'Outros', para as variáveis 'operacao', 'jurisdicao' e 'situacao_processual_sentenca'⁷.

Por fim, foi verificada a presença de dados faltantes. Apenas as colunas 'data_distribuicao' (110), 'data_voo' (69395), 'obj_principal' (66445), 'suobj_principal' (208205) e 'causa_alegada' (110120) apresentaram dados ausentes. Como esses casos foram

⁷ Arquivo 'AED_dados_judicializacao_vfinal.ipynb' disponibilizado no link.

classificados como 'Não informado' durante o processo de reclassificação, optou-se por não eliminar os registros com dados faltantes, considerando que isso não afetará a análise principal. Assim, após as etapas de tratamento, a base de dados de judicialização foi reduzida para 555.546 linhas e 31 colunas. O Quadro 8 apresenta os tipos de dados e a quantidade de valores ausentes de cada variável da base de judicialização final.

Quadro 8 – Tipos de dados e quantidade de valores ausentes da base final de judicialização

Coluna	Tipo de Dado	Qtd de Valores Ausentes
empresa_re	object	0
num_proc_ano	int64	0
data_distribuicao	object	110
UF	object	0
operacao	object	0
jurisdicao	object	0
instancia	object	0
situacao_processual_sentenca	object	0
n_autores	int64	0
data_voo	object	69395
valor_pedido_inicial	float64	0
valor_dano_moral	float64	0
valor_dano_material	float64	0
valor_total_condenacao	float64	0
valor_pago	float64	0
obj_problemas_operacionais	bool	0
obj_bagagem	bool	0
obj_contrato	bool	0
obj_outros	bool	0
causa_nao_informado	bool	0
causa_forca_maior	bool	0
causa_fortuitos	bool	0
causa_terceiros	bool	0
causa_cliente	bool	0
causa_aerea	bool	0
causa_outros	bool	0
obj_principal	object	66445
subobj_principal	object	208205
causa_alegada	object	110120
mes_distribuicao	int32	0
regiao	object	0

Fonte: Elaborado pela autora.

3.6.1.2.2 Tratamento dos dados do Consumidor.gov.br

O tratamento dos dados provenientes do Consumidor.gov.br⁸ envolveu algumas etapas iniciais a fim de igualar a estrutura da base de dados com a da base de judicialização. Essas etapas incluíram a seleção das reclamações das mesmas companhias aéreas (Azul Linhas Aéreas, Gol Linhas Aéreas, Latam Airlines, United Airlines, Copa Airlines e TAP Air Portugal) e estados brasileiros (AM, BA, CE, DF, ES, GO, MG, MS, MT, PE, PR, RJ, RO, SC, SP e TO) dos processos judiciais, garantindo a correspondência entre as duas bases. O objetivo foi alinhar as informações, permitindo uma análise comparativa mais consistente entre as reclamações registradas na plataforma Consumidor.gov.br e os processos judiciais relacionados ao transporte aéreo.

Em relação as padronizações, os nomes das empresas aéreas foram reclassificados apenas para Azul, Copa, Gol, Latam, Tap e United. Além disso, os motivos das reclamações foram reclassificados utilizando a mesma lógica da técnica de agrupamento do projeto de pesquisa "Diagnóstico sobre a Judicialização do Transporte Aéreo no Brasil: Uma Aplicação de Aprendizado de Máquina", apresentada no tópico 3.6.1.2.1. Dessa forma, os seguintes passos foram realizados:

- Primeiramente, foram levantadas as descrições dos códigos classificadores da ANAC, constantes na tabela do Anexo II da Portaria nº 764/SAS, de 14 de março de 2019⁹;
- A partir da análise dos campos Tema, Subtema e Assunto, cada código foi reclassificado em uma das quatro categorias: Problemas Operacionais, Bagagem, Contrato e Outros¹⁰;
- Em seguida, foi definida uma função em Python para selecionar apenas os códigos da coluna 'Código Classificador ANAC'. Esses códigos foram então adicionados em uma nova coluna chamada 'codigo_anac'. Nos casos em que a reclamação possuía mais de um código, eles foram separados por uma vírgula na nova coluna;
- Uma nova função em Python foi criada para reclassificar os motivos das reclamações. Essa função analisava os códigos da coluna 'codigo_anac' e retornava a categoria a qual eles pertenciam. Como resultado, foram adicionadas quatro variáveis booleanas à base de dados de reclamações. Nessas variáveis, o valor '*True*' ou 1 indica que o processo pertence ao grupo correspondente, enquanto '*False*' ou 0 indica que ele não pertence.

⁸ Arquivo 'AED_dados_consumidor_gov_vfinal.ipynb' disponibilizado no link.

⁹ <https://www.anac.gov.br/assuntos/legislacao/legislacao-1/portarias/2019/portaria-no-0764-sia-11-03-2019>

¹⁰ O dicionário completo dos códigos classificadores da ANAC pode ser acessado por meio deste link

Por fim, as colunas não relevantes para a análise foram excluídas, sendo elas: 'Cidade', 'Data Abertura', 'Data e Hora Resposta', 'Data e Hora Análise', 'Data e Hora Recusa', 'Data Finalização', 'Prazo Resposta', 'Prazo Análise Gestor (em dias)', 'Nome Fantasia', 'Área', 'Assunto', 'Grupo Problema', 'Problema', 'Edição de Conteúdo', 'Interação do Gestor', 'Código Classificador ANAC'. Em seguida, foi verificada a presença de dados faltantes. As colunas 'Tempo Resposta (em dias)', 'Avaliação Reclamação' e 'Nota do Consumidor' apresentaram 212, 2 e 130.118 valores ausentes, respectivamente. Como essas variáveis são complementares e não afetam diretamente a análise principal, optou-se por manter os registros com dados ausentes. Assim, após as etapas de tratamento, a base de dados de reclamações foi reduzida para 316.799 linhas e 18 colunas. O Quadro 9 apresenta os tipos de dados e quantidade de valores ausentes de cada variável da base de reclamações final.

Quadro 9 – Tipos de dados e quantidade de valores ausentes da base final de reclamações

Coluna	Tipo de Dado	Qtd de Valores Ausentes
Gestor	object	0
Região	object	0
UF	object	0
Ano Abertura	int64	0
Mês Abertura	int64	0
Tempo Resposta (em dias)	float64	212
Forma Contrato	object	0
Procurou Empresa	object	0
Respondida	object	0
Situação	object	0
Avaliação Reclamação	object	2
Nota do Consumidor	float64	130118
empresa_aerea	object	0
codigo_anac	object	0
obj_problemas_operacionais	bool	0
obj_bagagem	bool	0
obj_contrato	bool	0
obj_outros	bool	0

Fonte: Elaborado pela autora.

3.6.1.3 Análise de dados

Na etapa de análise de dados, as bases de dados foram analisadas de forma separada, por meio de dois tipos de análises, sendo elas:

- **Análise univariada:** que consistiu em avaliar cada uma das variáveis isoladamente, a fim de analisar a ocorrência de seus possíveis valores e definir a frequência (Ferreira *et*

al., 2021). Por exemplo, frequência de reclamações e processos judiciais por motivação e causa;

- **Análise multivariada:** consistiu em estabelecer relações entre duas ou mais variáveis (Ferreira *et al.*, 2021). Por exemplo, a quantidade de reclamações ou processos judiciais por região e a companhia aérea.

3.6.1.4 Apresentação e interpretação de dados

Por fim, os dados foram apresentados, por meio de tabelas e gráficos, e interpretados, a fim de responder quais são os principais motivos e causas de reclamações de consumidores e processos judiciais contra empresas de transporte aéreo no Brasil e se houve alguma mudança observada no período pré, durante e pós pandemia de COVID-19.

Nesse contexto, o ano de 2019 foi considerado como o período pré-pandemia, enquanto os anos de 2020 e 2021 foram classificados como o período durante a pandemia, tendo em vista que 2020 foi o ano de início da disseminação global da doença, e 2021 ainda foi marcado por novas ondas de contágio e a manutenção de medidas de emergenciais. De acordo com a ANAC (2023b), essas medidas incluíram a Lei nº 14.034/2020, posteriormente alterada pela Lei nº 14.174/2021, que flexibilizava regras para alteração de passagens, cancelamento, reembolso e crédito, aplicando-se a situações ocorridas até o final de 2021. Já os anos de 2022 e 2023 foram considerados como o período pós-pandemia, refletindo a transição para a normalidade operacional e regulatória no setor de transporte aéreo.

3.6.2 Análise de correlação

Correlações ocorrem quando dois eventos, que não necessariamente estão relacionados como causa e efeito, tendem a ocorrer simultaneamente (Ferreira *et al.*, 2021). Dessa forma, este trabalho teve como objetivo analisar se existe correlação entre os motivos de reclamação e processos judiciais.

Para essa análise, inicialmente foi realizada a consolidação dos dados de reclamações e processos judiciais. Para isso, foram contabilizadas as quantidades de reclamações e processos para cada tipo de motivo ('obj_problemas_operacionais', 'obj_bagagem', 'obj_contrato' e 'obj_outros'), por meio do agrupamento das informações por região, ano, mês e empresa aérea. Posteriormente, as duas bases de dados foram integradas, considerando as colunas em comum, resultando em um *dataframe* com 14 colunas e 1638 linhas.

Em seguida, a distribuição dos dados foi analisada por meio de histogramas, QQ-plot e teste de Shapiro-Wilk, utilizando as bibliotecas *pandas*, *matplotlib*, *seaborn*, *scipy.stats* do Python. Os histogramas são representações gráficas que exibem a frequência de ocorrência dos dados dentro de intervalos específicos, permitindo visualizar a forma da

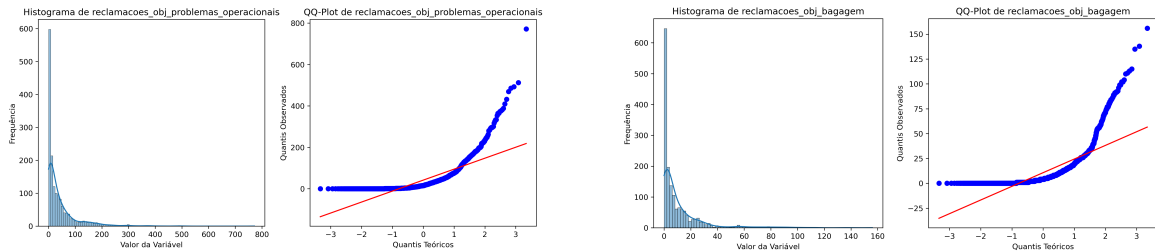
distribuição, incluindo características como simetria, assimetria e a presença de múltiplos picos. O QQ-plot (*Quantile-Quantile Plot*) compara os quantis da distribuição empírica dos dados com os quantis de uma distribuição teórica, geralmente a normal; quando os pontos do gráfico se alinham a uma linha reta, isso sugere que os dados seguem a distribuição esperada. Por fim, o teste de Shapiro-Wilk é um teste estatístico utilizado para avaliar a normalidade dos dados, verificando se eles se ajustam a uma distribuição normal. Se o valor de p obtido for inferior a um nível de significância predefinido (como 0,05), rejeita-se a hipótese de normalidade. A Tabela 3 e a Figura 4 apresenta os resultados do teste de Shapiro-Wilk, os histogramas e os QQ-plots obtidos.

Tabela 3 – Resultados do Teste de Shapiro-Wilk

Variável	Estatística	p-valor	Normalidade
reclamacoes_obj_problemas_operacionais	0,6283	1,33E-50	Não
reclamacoes_obj_bagagem	0,6068	1,57E-51	Não
reclamacoes_obj_contrato	0,5796	1,20E-52	Não
reclamacoes_obj_outros	0,5262	1,13E-54	Não
processos_obj_problemas_operacionais	0,7316	2,10E-45	Não
processos_obj_bagagem	0,6175	4,50E-51	Não
processos_obj_contrato	0,7037	5,94E-47	Não
processos_obj_outros	0,1994	5,07E-64	Não

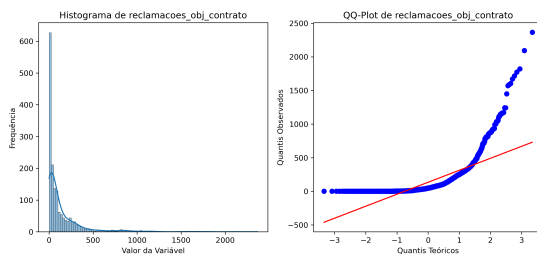
Fonte: Elaborado pela autora.

Figura 4 – Histogramas e QQ-plots das Variáveis

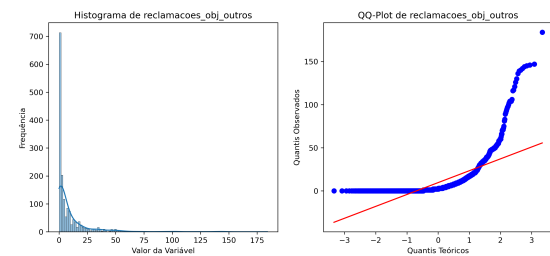


(a) Reclamações - Problemas Operacionais

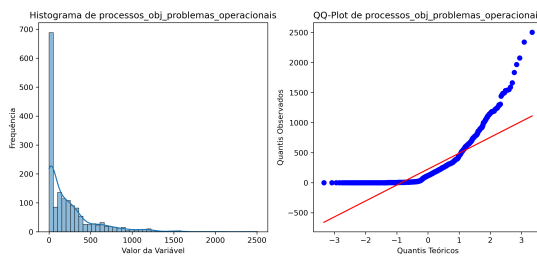
(b) Reclamações - Bagagem



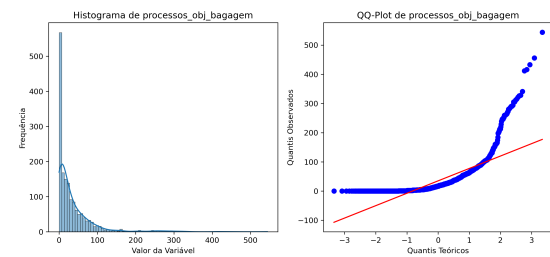
(c) Reclamações - Contrato



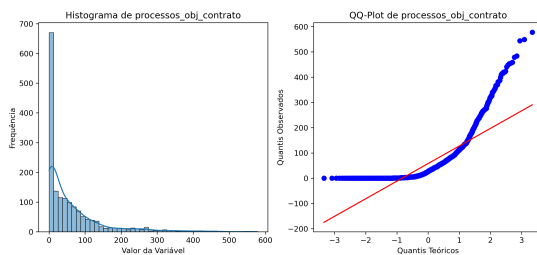
(d) Reclamações - Outros



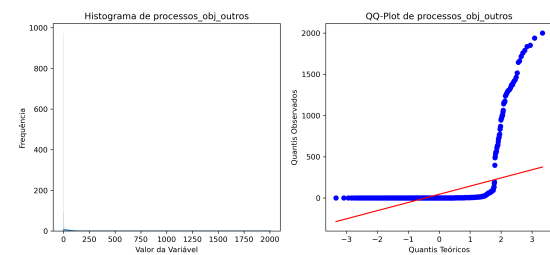
(e) Processos - Problemas Operacionais



(f) Processos - Bagagem



(g) Processos - Contrato



(h) Processos - Outros

Fonte: Elaborada pela autora.

Com base nos resultados apresentados na Tabela 3, todas as variáveis analisadas apresentaram valores de p-valor extremamente baixos (menores que 0,05), indicando que as distribuições dessas variáveis não seguem uma distribuição normal. Na Figura 4, os histogramas e QQ-plots também confirmaram esses resultados. Nos histogramas, é possível observar que as distribuições das variáveis não apresentam a forma característica de uma distribuição normal, sendo frequentemente assimétricas de forma inclinada para a

direita (distribuições assimétricas positivas). Os QQ-plots, por sua vez, mostram que os pontos não seguem uma linha reta, o que reforça a conclusão de que os dados não seguem uma distribuição normal. Devido à ausência de normalidade, optou-se pelo coeficiente de correlação de Spearman, uma abordagem não paramétrica recomendada como alternativa ao coeficiente de Pearson quando as variáveis quantitativas não seguem uma distribuição normal (Bauer, 2007).

O coeficiente de correlação de Spearman (ρ) é uma medida estatística baseada em postos. No caso de variáveis quantitativas, a correlação entre X e y pode ser calculada da mesma forma que o coeficiente de Pearson, mas substituindo os valores das duas variáveis pelos seus postos (Bauer, 2007). Dessa forma, esse coeficiente quantifica a direção e a força da relação entre duas variáveis.

Esse coeficiente varia de -1 a 1, sendo que o sinal indica a direção da correlação (negativa ou positiva) e o valor indica a força. Assim, quando $\rho = 1$, há uma correlação positiva perfeita, indicando que as amostras em X e Y estão positivamente relacionadas de maneira ascendente (ou seja, quando X aumenta, Y também aumenta). Por outro lado, quando $\rho = -1$, há uma correlação negativa perfeita, indicando que as amostras em X e Y estão negativamente relacionadas de forma descendente (ou seja, quando X aumenta, Y diminui). Já um coeficiente de correlação de 0 significa que não há um relacionamento entre X e Y (Ferreira *et al.*, 2021). Em relação à força, quanto mais perto de 1 mais forte é o nível de associação entre as variáveis, independente do sinal. Quanto mais perto de zero, menor é o nível de associação (Paranhos *et al.*, 2014).

Calculou-se o coeficiente de correlação de Spearman (ρ) entre as variáveis relacionadas aos motivos, utilizando a função `.corr(method='spearman')` da biblioteca Pandas no Python. O arquivo com o código em Python pode ser acessados por meio deste link.

3.6.3 *Desenvolvimento dos modelos de aprendizado de máquina*

O modelo é o resultado obtido ao aplicar um algoritmo a uma base de dados. O algoritmo, por sua vez, é um conjunto de procedimentos executados em sequência, com o objetivo de transformar um conjunto de dados de entrada em um ou mais valores de saída (Sicsú; Samartini; Barth, 2023). Neste trabalho, foram utilizados algoritmos supervisionados de classificação, ou seja, algoritmos que visam prever em qual categoria uma nova observação deve ser classificada.

Este estudo reaplicou os três algoritmos de aprendizado de máquina (*Naive Bayes Classifier*, *Support Vector Machine* e *Random Forest Classifier*) utilizados por Torres, Guterres e Celestino (2023), mas com algumas diferenças. A primeira foi a utilização da técnica detalhada no tópico 3.6.1.2.1, para a padronização dos motivos dos processos judiciais (objeto e subobjeto principal) e das causas alegadas pelas companhias aéreas.

O objetivo era superar a limitação identificada pelos autores, de que as diferenças na classificação dos motivos e das causas dos processos judiciais por cada companhia aérea poderiam ter influenciado o desempenho dos algoritmos, já que eventos semelhantes poderiam ter sido interpretados de formas distintas na criação do banco de dados. A segunda diferença foi a inclusão de novas variáveis preditoras, como: operação, jurisdição, instância, número de autores e valor do pedido inicial.

O objetivo principal dos modelos desenvolvidos foi prever o valor total de condenação, e por isso, foram utilizados apenas os dados de processos judiciais que tinham a situação processual e a sentença igual a "Condenação" ou "Acordo". Para o desenvolvimento dos modelos, foi necessário realizar a seleção das variáveis relevantes, a preparação dos dados e a definição dos algoritmos e hiperparâmetros. Além disso, métricas apropriadas foram aplicadas para avaliar o desempenho dos modelos. Para realizar essas etapas, foram utilizadas as bibliotecas Pandas, Numpy, Matplotlib, Seaborn, Time, Scikit-learn e Imbalanced-learn do Python. Os arquivos com os códigos em Python podem ser acessados por meio deste link.

3.6.3.1 Seleção das variáveis para os modelos

As variáveis selecionadas foram:

- **Companhia aérea:** por fornecer entendimento sobre a frequência de processos judiciais que cada empresa recebe dos passageiros, o que pode influenciar a decisão final dependendo das reivindicações;
- **Ano e mês de distribuição do processo:** por permitir que as companhias aéreas entendam a proporção de ações judiciais ao longo do período e avaliem se o comportamento dos valores de indenização muda nos meses de alta e baixa temporada ou devido a algum fator externo, como a pandemia de COVID-19;
- **Região:** por ser uma variável essencial para compreender o comportamento das ações judiciais em todo o Brasil, onde há uma vasta extensão territorial com regiões tanto altamente desenvolvidas quanto subdesenvolvidas, assim conhecer as áreas com as maiores concentrações de processos auxilia as companhias aéreas a agirem de maneira eficaz nessas localidades;
- **Operação:** por indicar se os processos judiciais são mais relacionados a voos nacionais ou internacionais;
- **Jurisdição:** por indicar a classificação do tribunal onde tramitou o processo;
- **Instância:** por indicar em qual instância o processo se encontra;

- **Situação processual e sentença:** por mostrar o resultado do processo e se o pedido do consumidor foi considerado após o encerramento do processo;
- **Número de autores:** por mostrar a quantidade de autores envolvidos no processo;
- **Valor do pedido inicial, valor do dano moral e valor do dano material:** por fornecerem informações sobre os tipos de indenizações que os passageiros solicitaram, seja com maior incidência de danos patrimoniais, que visam compensar perdas materiais, ou danos morais, que são mais proeminentes nas decisões judiciais;
- **Valor total da condenação:** por corresponder ao valor de indenização pago pelas companhias aéreas aos seus clientes em razão das reclamações apresentadas no processo. Essa variável engloba tanto ações judiciais em que o juiz decidiu o valor final quanto acordos firmados entre as partes;
- **Motivos e causas:** por fornecerem dados sobre os principais motivos dos processos judiciais e a origem dos problemas que eles mencionaram.

Com as variáveis selecionadas, as variáveis de valor do pedido inicial, valor do dano moral, valor do dano material e valor total de condenação foram categorizadas em três intervalos: Baixo, Médio e Alto¹¹. Para definir esses intervalos, foram excluídos os valores acima de R\$1 milhão e os valores repetidos, já que, em alguns casos, muitos registros tinham valor igual a zero, fazendo com que vários percentis também fossem zero. Com base nos valores únicos de cada variável, foram calculados os percentis 25 e 75 para definir os intervalos:

- Baixo: valores até o percentil 25;
- Médio: valores entre o percentil 25 e o 75;
- Alto: valores acima do percentil 75.

A Tabela 4 apresenta os valores dos intervalos para cada variável.

Tabela 4 – Intervalos das variáveis de indenização em reais (R\$)

Variável	Baixo	Médio	Alto
Valor do pedido inicial	$\leq 10.332,64$	$> 10.332,64$ e $\leq 22.281,32$	$> 22.281,32$
Valor do dano moral	$\leq 1.520,93$	$> 1.520,93$ e $\leq 5.970,86$	$> 5.970,86$
Valor do dano material	$\leq 450,60$	$> 450,60$ e $\leq 2.638,10$	$> 2.638,10$
Valor total de condenação	$\leq 1.729,02$	$> 1.729,02$ e $\leq 7.236,81$	$> 7.236,81$

Fonte: Elaborado pela autora.

¹¹ Torres, Guterres e Celestino (2023) definiram os seguintes intervalos: Baixo \leq R\$1.000,00; R\$1.000,00 < Médio \leq R\$5.000,00; Alto $>$ R\$5.000,00. No entanto, ao reaplicar essa divisão, as classes ficaram mal distribuídas. Assim, optou-se por definir os intervalos com base nos percentis dos dados de cada variável.

A Tabela 5 apresenta um resumo final das variáveis utilizadas, suas definições e categorias.

Tabela 5 – Variáveis

Variável	Definição	Categorias
empresa_re	Nome da empresa ré no processo.	Azul Copa Gol Latam Tap United
num_proc_ano	Ano de distribuição do processo.	2019 2020 2021 2022 2023
mes_distribuicao	Mês de distribuição do processo.	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
regiao	Região em que a ação foi iniciada.	Centro-Oeste Nordeste Sul Norte Sudeste
operacao	Tipo de operação.	Doméstico Internacional Outros
jurisdicao	Tribunal onde o processo tramitou.	Juizado Especial Cível Vara Cível Outros
instancia	Instância da sentença do processo.	1ª Instância 2ª Instância 3ª Instância Não informado
situacao_processual_sentenca	Julgamento do processo.	Condenação Acordo
n_autores	Quantidade de autores do processo.	Não se aplica
valor_pedido_inicial	Valor inicial pedido pelo(s) autor(es).	Baixo Médio Alto
valor_dano_moral	Pedido de compensação por dano moral.	Baixo Médio Alto
valor_dano_material	Pedido de compensação por dano material.	Baixo Médio Alto
valor_total_condenacao	Valor total da condenação.	Baixo Médio Alto
obj_problemas_operacionais, obj_bagagem, obj_contrato, obj_outros	Motivo apresentado pelo passageiro para ação judicial.	Problemas operacionais Bagagem Contrato Outros
causa_aerea, causa_cliente, causa_terceiros, causa_fortuitos, causa_forca_maior, causa_ao_informado, causa_outros	Fator informado pela empresa para o problema.	Aérea Cliente Terceiros Fortuitos Força maior Não informado Outros

Fonte: Elaborado pela autora.

3.6.3.2 Preparação dos dados para os modelos

Para preparar os dados para os modelos, primeiramente foi feita a divisão dos dados de entrada (X) e os dados de saída (y), excluindo a coluna de valor total de condenação do conjunto de dados principal para formar o X, e mantendo-a no y. Em seguida, tendo em vista que a maior parte do conjunto de dados era formado por variáveis categóricas, aplicou-se *one-hot encoding* nas variáveis do X e *label encoding* no y.

O *one-hot encoding*¹² cria uma nova coluna para cada valor único de uma coluna existente, indicando com 1 quando o valor é afirmativo e com 0 quando negativo, ele é recomendado quando se trata de variáveis nominais. Por outro lado, o *label encoding*¹³ designa um valor inteiro (0, 1, 2, ...) para cada valor único da coluna existente, sendo recomendando quando se trata de variáveis ordinais.

Nas variáveis numéricas de X, aplicou-se o *MinMaxScaler*¹⁴ para escalar os valores em um intervalo de 0 a 1. Dessa forma, os valores foram ajustados para que o valor mínimo de cada variável fosse 0 e o máximo fosse 1, preservando a relação entre os valores originais. Essa normalização garante que os dados numéricos estejam na mesma escala, o que pode melhorar o desempenho dos algoritmos de machine learning.

Por fim, a base de dados foi dividida duas partes, uma para treino e outra teste. Para isso, utilizou-se a função *train_test_split*¹⁵, com os seguintes parâmetros: `test_size=0.2`, `stratify=y`, `random_state=42`. Isso quer dizer que:

- **test_size=0.2:** 20% dos dados foram reservados para o conjunto de teste e 80% para o de treino. Essa divisão permite a avaliação do modelo em dados não vistos durante o treinamento, garantindo uma melhor estimativa de seu desempenho em dados reais;
- **stratify=y:** a divisão foi estratificada com base na variável y, ou seja, as proporções relativas de diferentes classes em y são mantidas nos conjuntos de treino e teste;
- **random_state=42:** garante a reprodutibilidade da divisão, ou seja, que a divisão será sempre a mesma se o código for executado várias vezes.

Ressalta-se que a base de dados utilizada é uma base desbalanceada, isto é, a quantidade das categorias baixo, médio e alto de valor total de condenação não são iguais¹⁶, existindo uma menor incidência de uma classe (classe minoritária) em relação as outras (classes majoritárias). Essa distribuição desigual das classes pode levar a modelos

¹² <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html>

¹³ <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html>

¹⁴ <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>

¹⁵ https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html

¹⁶ Quantidade de registros por classe: 103018 Baixo, 77477 Médio e 26799 Alto.

enviesados e impactar a capacidade de generalização do modelo. Dessa forma, foi aplicada uma técnica de subamostragem (*undersampling*), que consiste em reduzir os exemplos da classe majoritária. Essa técnica foi aplicada apenas nos dados de treino do modelo, preservando a integridade dos dados de teste, a fim de possibilitar a avaliação de como o modelo se comportaria em situações reais refletidas pela base de dados desbalanceada. Ou seja, ajustar a distribuição das classes na base de teste artificialmente poderia levar a conclusões falsas e otimistas sobre o modelo.

Com base nisso, a técnica utilizada nos dados de treino foi o *RandomUnderSampler*¹⁷. O parâmetro `sampling_strategy='auto'` define que a amostragem será feita de forma a reduzir as classes majoritárias até o tamanho da classe minoritária, enquanto o parâmetro `random_state=42` garante a reprodutibilidade do processo, fixando a semente do gerador de números aleatórios.

3.6.3.3 Algoritmos de aprendizado de máquina

Após as etapas de pré-processamento, os três modelos de aprendizado de máquina foram executados, utilizando os algoritmos e melhores hiperparâmetros identificados por Torres, Guterres e Celestino (2023), sendo eles:

- **Naive Bayes Classifier (NB):** algoritmo *MultinomialNB*¹⁸ com alpha igual a 8;
- **Support Vector Machine (SVM):** algoritmo *LinearSVC*¹⁹ com valor de penalidade (penalty) igual a 12, valor de perda (*loss*) igual a *hinge* e valor de regularização (C) igual a 2;
- **Random Forest Classifier (RF):** algoritmo *RandomForestClassifier*²⁰ com o número de árvores (*n_estimators*) igual 1.000, medida de qualidade (*criterion*) igual a *gini*, profundidade máxima da árvore (*max_depth*) igual a 12 e número mínimo de divisões de um nó (*min_samples_split*) igual a 4.

Tendo em vista que os dados deste trabalho são multiclasse, foi necessária uma modificação interna nos algoritmos que permite dividir a classificação multiclasse em várias classificações binárias. Dessa forma, foi aplicada a abordagem *one vs one* ("ovo"), que treina um classificador binário para cada par de classes possíveis no conjunto de dados. Isso foi utilizado para obter a área sob a curva ROC de cada modelo.

¹⁷ https://imbalanced-learn.org/stable/references/generated/imblearn.under_sampling.RandomUnderSampler.html

¹⁸ https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html

¹⁹ <https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html>

²⁰ <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>

3.6.3.4 Naive Bayes Classifier

O *Naive Bayes Classifier* (NB) é um algoritmo supervisionado fundamentado no raciocínio bayesiano. Ele é utilizado em tarefas de aprendizado em que cada instância x é descrita por um conjunto de atributos (Ex.: companhia aérea, ano, época, região, valor do dano moral e material etc.) e em que a classe alvo pode assumir qualquer valor de um conjunto V (Ex.: baixo, médio e alto). Dessa forma, o NB classifica uma instância de dados ao atribuir o valor mais provável de y para o alvo, considerando os valores de atributos (Mitchell; Mitchell, 1997).

O algoritmo Naive Bayes opera em várias etapas para realizar essa classificação de instâncias. Primeiramente, calcula as probabilidades *a priori* de cada classe com base nos dados de treinamento. Em seguida, estima as probabilidades condicionais de cada valor de atributo para cada classe. Utilizando o Teorema de Bayes e assumindo independência condicional entre os atributos, calcula a probabilidade *a posteriori* de cada classe para a instância em teste. A classe com a maior probabilidade *a posteriori* é então escolhida como a predição para a instância.

3.6.3.5 Support Vector Machine

A *Support Vector Machine* (SVM) é uma técnica que ajuda na seleção de um classificador adequado para um conjunto de dados específico (Faceli *et al.*, 2021). Assim, seu objetivo é encontrar uma fronteira que separe as possíveis classes no conjunto de dados e classificar o maior número possível de exemplos, maximizando a distância da fronteira aos pontos mais próximos desta (Géron, 2019).

O funcionamento do SVM pode ser descrito em termos de otimização. O algoritmo busca um hiperplano que maximize a margem entre as classes, ou seja, a distância entre o hiperplano e os pontos mais próximos de cada classe, conhecidos como vetores de suporte. Isso permite uma generalização robusta, já que o SVM foca nos exemplos que são mais difíceis de classificar (os vetores de suporte). Em situações em que os dados não são linearmente separáveis, o SVM utiliza o truque do núcleo (kernel), que mapeia os dados para um espaço de maior dimensionalidade, onde as classes podem ser separadas linearmente. (Géron, 2019; Faceli *et al.*, 2021).

Ressalta-se que como o algoritmo aplicado foi o *LinearSVC*, que não possui um método nativo para calcular as probabilidades previstas de cada classe, utilizou-se um calibrador de modelos para classificações multiclasse chamado *CalibratedClassifierCV*. Com isso, foi possível obter o valor da área sob a curva ROC.

3.6.3.6 *Random Forest Classifier*

O *Random Forest* (RF) é um algoritmo *ensemble*, isto é, que combina o resultado de múltiplos modelos em busca de produzir um melhor modelo preditivo. Assim, ele combina várias árvores de decisão, treinadas com atributos e conjunto de dados distintos, e utiliza o método *bagging* para unir o resultado de todas as árvores geradas e identificar a classe final por meio de votação (Breiman, 2001). Ressalta-se que árvores de decisão são um tipo de estrutura de dados que possui nós internos, ramos e nós folhas. Em cada nó interno ocorre a verificação de um dos atributos. Os ramos representam os possíveis resultados da verificação. E os nós folhas representam as classes a serem atribuídas (Castro; Ferrari, 2016 *apud* Rocha, 2020).

De forma geral, as principais etapas do algoritmo são: i) um conjunto de n amostras é criado utilizando *bootstrapping*, isto é, são selecionadas amostras aleatórias com reposição (ou seja, a mesma linha pode ser escolhida mais de uma vez); ii) uma árvore de decisão é construída por amostra, resultando na predição de uma classe; iii) as classes determinadas por cada árvore são computadas como um voto, sendo que a classe com maior quantidade de votos se torna a classe predita do modelo (Freitas, 2018; Torres, 2022).

Devido à aleatoriedade das amostras de dados e das variáveis consideradas na tomada de decisão em cada nó das árvores de decisão, o RF reduz problemas de *overfitting*, isto é, um sobreajuste aos dados de treinamento, o que prejudica a generalização do modelo (Breiman, 2001). Além disso, o RF é capaz de descobrir relações não lineares, semelhante às redes neurais artificiais, porém adota uma abordagem de aprendizagem diferente (Rajendran; Srinivas; Grimshaw, 2021).

3.6.3.7 *Avaliação das métricas de desempenho dos modelos*

Para avaliar o desempenho de cada modelo foram utilizadas as métricas avaliadas por Torres, Guterres e Celestino (2023) para fins de comparação, sendo elas: matriz de confusão, área sob a curva ROC, acurácia e tempo de processamento do modelo em minutos. Adicionalmente, também calculou-se três métricas comumente usadas para problemas de classificação multiclasse²¹: precisão, *recall* e *F1-score* (Bradley; Rajendran, 2021 *apud* Rajendran; Srinivas; Grimshaw, 2021).

A matriz de confusão é uma representação tabular do desempenho de um modelo de classificação no conjunto de teste, mostrando o número de previsões corretas e incorretas para cada categoria (Ahmad *et al.*, 2020). A Tabela 6 apresenta um exemplo dessa matriz, em que as linhas representam as categorias reais e as colunas as previstas, resultando em quatro parâmetros:

²¹ https://scikit-learn.org/stable/modules/model_evaluation.html#classification-metrics

- **verdadeiro positivo (TP):** casos em que o modelo previu corretamente a classe positiva (verdadeira);
- **falso positivo (FP):** casos em que o modelo previu incorretamente a classe positiva (falso);
- **falso negativo (FN):** casos em que o modelo previu incorretamente a classe negativa (falso);
- **verdadeiro negativo (TN):** casos em que o modelo previu corretamente a classe negativa (verdadeiro).

Tabela 6 – Matriz de Confusão

	Previsto Positivo	Previsto Negativo
Real Positivo	TP	FN
Real Negativo	FP	TN

Fonte: Elaborado pela autora.

A área sob a curva ROC (AUC-ROC) mostra se o classificador funciona melhor do que uma escolha aleatória. Para isso, as taxas de verdadeiros-positivo e falsos-positivo são relacionadas, o que permite saber se o classificador consegue diferenciar bem as classes (Murphy, 2012). Quanto maior a AUC-ROC, melhor o modelo é em distinguir as classes.

A acurácia indica uma performance geral do modelo. Ela é definida como a taxa total de acerto do algoritmo, ou seja, calcula dentre todas as classificações, quantas ele classificou corretamente, considerando tanto as verdadeiras como as falsas (Ahmad *et al.*, 2020). Já a precisão mostra a proporção de verdadeiros positivos em relação a todos os classificados como positivos. Por outro lado, o *recall* mostra quantas observações positivas foram classificadas corretamente em relação ao total de observações positivas. Por fim, o *F1-score* é uma média harmônica entre a precisão e o *recall*, sendo utilizada para comprar dois modelos quando estas métricas são importantes (Sicsú; Samartini; Barth, 2023). Todas essas medidas variam de 0 a 1, sendo que quando maior o valor, melhor o desempenho de classificação. A Tabela 7 apresenta as fórmulas de cada métrica.

Tabela 7 – Fórmulas das métricas de avaliação de classificação.

Métrica	Fórmula
Acurácia	$\frac{TP+TN}{TP+TN+FP+FN}$
Precisão	$\frac{TP}{TP+FP}$
<i>Recall</i>	$\frac{TP}{TP+FN}$
<i>F1-score</i>	$2 \cdot \frac{\text{Precisão} \cdot \text{Recall}}{\text{Precisão} + \text{Recall}}$

Fonte: Elaborado pela autora.

3.6.3.8 Importância das variáveis em cada modelo

Para avaliar as variáveis mais importantes para cada modelo, foi aplicada o *permutation importance*. Essa técnica mede o impacto de cada variável no desempenho do modelo, através de uma alteração aleatória nos seus valores. Quando uma variável é permutada (ou seja, seus valores são embaralhados), o modelo perde parte de sua capacidade de previsão, o que reflete a importância dessa variável. Dessa forma, quanto maior a queda na performance do modelo após a permutação de uma variável, maior é a sua importância. Assim, para calcular a importância das variáveis, a função *permutation_importance*²² da biblioteca Scikit-learn do Python foi utilizada, repetindo o processo 10 vezes para garantir a robustez dos resultados. Em seguida, foi gerado um gráfico com as variáveis ordenadas de acordo com a sua importância, permitindo identificar as que mais contribuem para o modelo.

De forma complementar, também foi considerada a multicolinearidade e os efeitos específicos dos modelos para analisar os resultados de importância das variáveis. A multicolinearidade ocorre quando duas ou mais variáveis independentes são altamente correlacionadas. Quando isso ocorre, as pontuações de importância das variáveis correlacionadas podem ser diluídas, já que a permutação de uma variável pode afetar a previsão do modelo de forma similar à permutação de outra variável correlacionada. Dessa forma, a correlação entre as variáveis foi calculada por meio do coeficiente de Pearson, já apresentado na seção 3.6.2.

Em relação aos efeitos específicos dos modelos, ressalta-se que diferentes tipos de modelos podem atribuir importâncias distintas às variáveis, mesmo que as variáveis sejam as mesmas. Por exemplo, no *Naive Bayes*, que assume a independência entre os recursos, a importância de permutação pode se comportar de maneira diferente se as variáveis forem fortemente correlacionadas, já que o modelo não leva em consideração essas correlações. Por outro lado, no SVM, os kernels não lineares são capazes de capturar padrões mais complexos, o que pode levar a uma classificação diferente da importância dos recursos em comparação com modelos lineares. Já em modelos como *Random Forest*, que lida bem com interações e correlações entre os recursos, a distribuição da importância pode ser distinta, refletindo a capacidade do modelo de capturar essas interações de forma mais robusta, em contraste com modelos mais simples. Tais características foram consideradas na análise dos resultados.

²² https://scikit-learn.org/stable/modules/generated/sklearn.inspection.permutation_importance.html

4 RESULTADOS E DISCUSSÃO

4.1 Análise Exploratória dos Dados

4.1.1 *Dados de reclamações de consumidores de companhias aéreas*

Este tópico apresenta os resultados da análise exploratória das 316.799 reclamações analisadas. Inicialmente, foram calculadas as estatísticas descritivas das variáveis. Para as variáveis categóricas, foram obtidas medidas como: número total de valores não ausentes (*count*); número de categorias ou valores distintos na coluna (*unique*); o valor mais comum ou mais frequente na coluna (*top*); e o número de vezes que o valor mais comum aparece (*freq*). Para as variáveis numéricas, foram obtidas medidas como: número total de valores não ausentes (*count*); média dos valores (*mean*); desvio padrão, que indica a dispersão dos dados em relação à média (*std*); valor mínimo na coluna (*min*); primeiro quartil (Q1), que é o valor abaixo do qual 25% dos dados se encontram (25%); mediana, ou segundo quartil, que divide os dados ao meio (50%); terceiro quartil (Q3), que é o valor abaixo do qual 75% dos dados se encontram (75%); e valor máximo na coluna (*max*). As Tabelas 8 e 9 apresentam os resultados de cada uma dessas medidas.

Tabela 8 – Reclamações: Estatísticas Descritivas - Variáveis Categóricas

Variáveis	count	unique	top	freq
Gestor	316799	18	Fundação de Proteção e Defesa do Consumidor	78827
Região	316799	5	SE	164550
UF	316799	16	SP	78929
Forma Contrato	316799	9	Internet	274536
Procurou Empresa	316799	2	S	292622
Respondida	316799	2	S	316587
Situação	316799	2	Finalizada avaliada	186673
Avaliação Reclamação	316797	3	Não Avaliada	130126
empresa_aerea	316799	6	Latam	130825
codigo_anac	316799	5470	5300	30692

Fonte: Elaborado pela autora.

Tabela 9 – Reclamações: Estatísticas Descritivas - Variáveis Numéricas

Variáveis	count	mean	std	min	25%	50%	75%	max
Ano Abertura	316799	2021,38	1,25	2019	2020	2021	2022	2023
Mês Abertura	316799	6,55	3,56	1	3	7	10	12
Tempo Resposta (em dias)	316587	4,91	3,27	0	2	5	7	16
Nota do Consumidor	186681	3,30	1,75	1	1	4	5	5

Fonte: Elaborado pela autora.

As estatísticas das Tabelas 8 e 9 apresentam uma visão geral sobre o comportamento e a distribuição dos dados relacionados às reclamações. Em relação as variáveis associadas ao processo de tratativa das reclamações, foram observados os seguintes pontos:

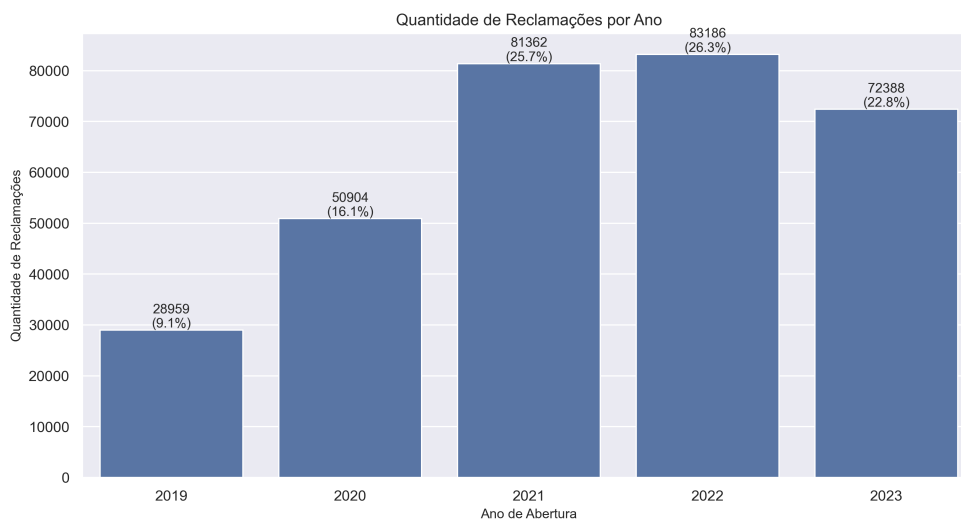
- **Gestor:** a Fundação de Proteção e Defesa do Consumidor foi a principal entidade responsável pela gestão das reclamações contra empresas aéreas (24,88%), seguida pela Secretaria Nacional do Consumidor (14,62%) e o Programa Estadual de Proteção e Defesa do Consumidor (13,88%);
- **Tempo Resposta (em dias):** o tempo médio de resposta é de 4,91 dias;
- **Forma Contrato:** a principal forma de contratação utilizada pelos consumidores foi a internet (86,65%);
- **Procurou Empresa:** 92,36% dos consumidores afirmaram ter procurado a empresa aérea para solucionar o problema;
- **Respondida:** 99,93% das reclamações foram respondidas pelas empresas;
- **Situação:** no sistema consta que 58,92% das reclamações foram finalizadas e avaliadas, enquanto 41,07% foram finalizadas, mas não avaliadas;
- **Avaliação Reclamação:** 41,07% dos consumidores não avaliaram o desfecho da reclamação. Entre os que avaliaram: 37,99% consideraram a reclamação como resolvida, enquanto 20,92% a classificaram como não resolvida;
- **Nota do Consumidor:** esta variável apresentou a maior quantidade de dados ausentes (130.118 registros), o que indica que muitos consumidores não avaliaram a empresa ao final do processo. Entre os consumidores que responderam, 42,82% atribuíram a nota 5 para o atendimento, seguidos por 31,45% que deram a nota 1. A nota média foi de 3,3.

Os subtópicos a seguir oferecem uma análise mais detalhada das variáveis de período, localização, empresa e motivo das reclamações.

4.1.1.1 Visão por Ano e Mês

A Figura 5 apresenta a quantidade de reclamações e o percentual relativo ao total por ano.

Figura 5 – Quantidade de Reclamações por Ano de Abertura



Fonte: Elaborada pela autora.

Observa-se na Figura 5 que 2021 e 2022 foram os anos com as maiores quantidades de reclamações registradas, representando 25,7% e 26,3% do total, respectivamente. Considerando que a plataforma Consumidor.gov.br foi adotada pela ANAC para reclamações contra empresas aéreas apenas no segundo trimestre de 2019, é esperado que o número de reclamações tenha sido menor nesse ano, já que foi um período de adaptação, em que os consumidores estavam se familiarizando com a plataforma.

De 2019 a 2022, o número de reclamações cresceu continuamente, mas houve uma redução de 12,98% em 2023 em comparação ao ano anterior. No entanto, é importante destacar que, devido à pandemia de COVID-19, houve uma queda no número de passageiros transportados nesses anos, e até o final de 2023, o volume de passageiros ainda não havia retornado aos níveis de 2019 (ANAC, 2024a). Nesse contexto, o índice de reclamações registradas por grupo de 100 mil passageiros pagos transportados - ou seja, a proporção do total de reclamações registradas pelo total de passageiros - permite uma análise mais precisa do aumento de reclamações durante o período. Os valores desse índice estão apresentados na Tabela 10.

Tabela 10 – Índice de Reclamações por 100 mil Passageiros por Ano

Ano	Total de Passageiros Pagos Transportados	Total de Reclamações	Índice de Reclamações por 100 mil Passageiros
2019	118.677.247	28.959	24,4
2020	51.934.069	50.904	98,0
2021	67.343.710	81.362	120,8
2022	97.686.945	83.186	85,2
2023	112.698.015	72.388	64,2

Fonte: Elaborado pela autora.

Na Tabela 10, é possível observar que 2020 e 2021 foram os anos com os maiores índices de reclamações, provavelmente influenciados pelas medidas tomadas devido à pandemia de COVID-19. Em 2021, por exemplo, o índice foi de 120,8 reclamações a cada 100 mil passageiros, o que representa um aumento de 385% em comparação a 2019, mesmo com uma redução de 43% no número total de passageiros pagos transportados. Quanto ao comportamento das reclamações ao longo dos meses, a Tabela 11 apresenta a quantidade de reclamações por mês e ano.

Tabela 11 – Quantidade de Reclamações por Mês e Ano de Abertura

Mês de Abertura	2019	2020	2021	2022	2023	Total	Percentual do Total
Janeiro	2003	2969	6403	12304	6422	30101	9,50%
Fevereiro	1793	2415	5696	8510	5154	23568	7,44%
Março	1882	9618	6977	8707	5875	33059	10,44%
Abril	2135	4077	5083	6471	5334	23100	7,29%
Mai	2084	3336	5151	6081	5568	22220	7,01%
Junho	2413	3217	5209	6281	5537	22657	7,15%
Julho	2791	3228	5512	6079	5729	23339	7,37%
Agosto	2352	2987	5894	6783	6518	24534	7,74%
Setembro	2478	4043	7905	5597	6296	26319	8,31%
Outubro	3147	4543	9304	5687	8210	30891	9,75%
Novembro	3136	5159	9530	5356	6437	29618	9,35%
Dezembro	2745	5312	8698	5330	5308	27393	8,65%
Total	28959	50904	81362	83186	72388	316799	100%

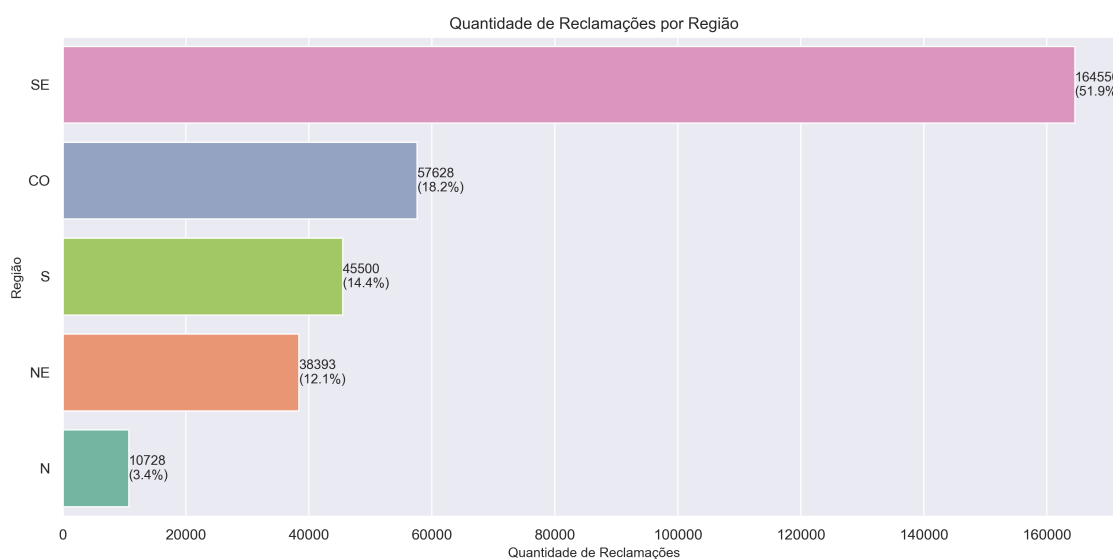
Fonte: Elaborado pela autora.

Na Tabela 11, é possível perceber que, de forma geral, as reclamações estão distribuídas de forma relativamente equilibrada ao longo dos meses. O mês com maior percentual foi março (10,44%), seguido por outubro (9,75%), janeiro (9,50%), novembro (9,35%) e dezembro (8,65%). Vale ressaltar que janeiro e dezembro são considerados meses de alta temporada devido às férias escolares. Por outro lado, os outros meses incluem feriados importantes, como Carnaval, Páscoa, Dia das Crianças, Proclamação da República e Dia da Consciência Negra, eventos que frequentemente provocam um aumento significativo no número de viagens e, conseqüentemente, nas reclamações.

4.1.1.2 Visão por Região e UF

A Figura 6 apresenta a quantidade de reclamações e o percentual relativo ao total por região. Ressalta-se que, apenas 16 estados brasileiros foram incluídos nos dados dessa pesquisa. Dessa forma, cada região considera as seguintes UFs: Norte (AM, RO e TO), Nordeste (BA, CE e PE), Centro-Oeste (DF, GO, MS e MT), Sudeste (ES, MG, RJ e SP) e Sul (PR e SC).

Figura 6 – Quantidade de Reclamações por Região



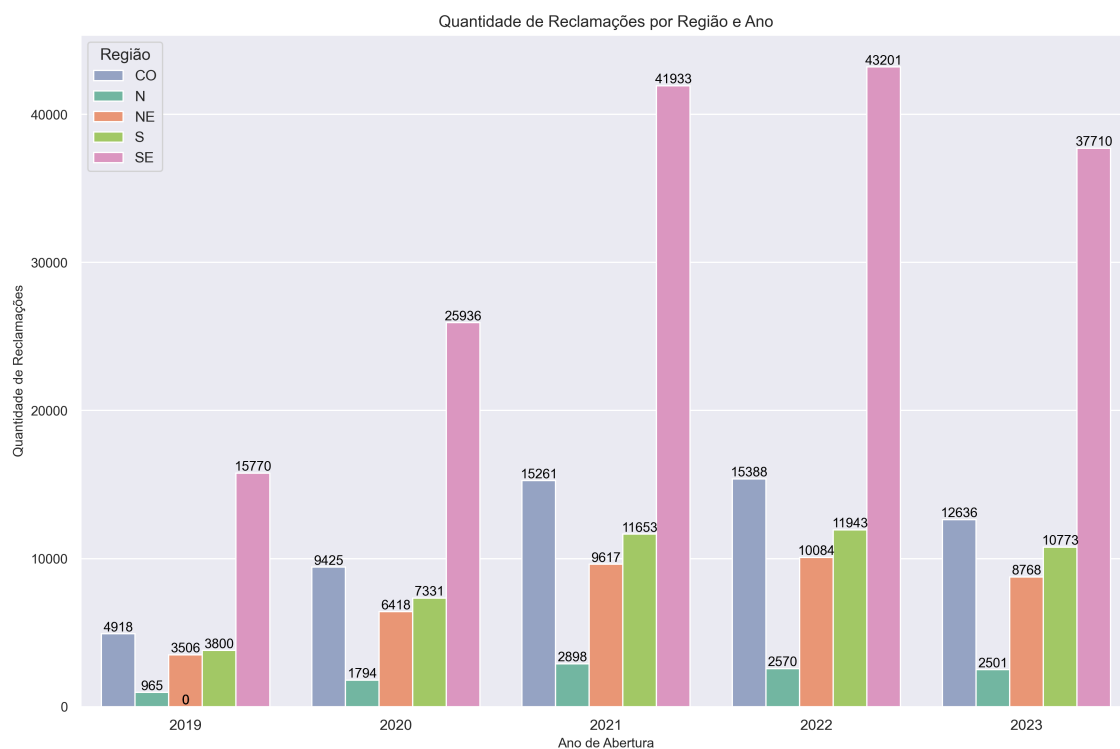
Fonte: Elaborada pela autora.

Na Figura 5, observa-se que o Sudeste é a região com maior concentração de reclamações (51,9%), seguido pelo Centro-Oeste (18,2%) e pelo Sul (14,4%). Segundo o ANAC (2024a)¹, 50,1% dos passageiros pagos no mercado doméstico eram do Sudeste, sendo esta região também responsável pela maior parte das decolagens (50,8%). Isso justifica a maior quantidade de reclamações registradas no Sudeste. Por outro lado, embora o Nordeste seja a segunda região com o maior número de passageiros e decolagens, sua quantidade de reclamações ficou abaixo da do Centro-Oeste e do Sul, que ocupam a quarta e a terceira posição, respectivamente, nesses índices. Uma hipótese é que isso esteja relacionado ao perfil dos consumidores dessas regiões. A Figura 11 apresenta a quantidade de reclamações por região e ano.

¹ Participação no número de passageiros pagos por região - 2023 - mercado doméstico: Sudeste (50,1%), Nordeste (19,3%), Sul (12,9%), Centro-Oeste (12,3%) e Norte (5,4%).

Participação no número de decolagens por região - 2023 - mercado doméstico: Sudeste (50,8%), Nordeste (18,7%), Sul (13,2%), Centro-Oeste (11,2%) e Norte (6,1%).

Figura 7 – Quantidade de Reclamações por Região e Ano



Fonte: Elaborada pela autora.

Na Figura 7, observa-se que as regiões apresentaram um crescimento contínuo nas reclamações até 2022, seguido por uma redução em 2023. A única exceção foi a região Norte, que registrou uma diminuição na quantidade de reclamações em 2022 e 2023. A Tabela 12 apresenta a quantidade de reclamações por UF e ano.

Tabela 12 – Quantidade de Reclamações por UF e Ano

UF	2019	2020	2021	2022	2023	Total	Percentual do Total
SP	8593	12925	20091	20515	16805	78929	24,91%
MG	3395	6501	10851	12018	11229	43994	13,89%
DF	3178	5873	9371	9024	6874	34320	10,83%
RJ	2831	4845	8227	7878	6956	30737	9,70%
PR	2390	4394	6855	6803	6107	26549	8,38%
SC	1410	2937	4798	5140	4666	18951	5,98%
BA	1441	2489	4091	4280	3748	16049	5,07%
CE	1291	2161	2951	3281	2706	12390	3,91%
GO	894	1712	3002	3098	2750	11456	3,62%
ES	951	1665	2764	2790	2720	10890	3,44%
PE	774	1768	2575	2523	2314	9954	3,14%
MT	508	1137	1845	2033	1816	7339	2,32%
AM	467	923	1581	1270	1155	5396	1,70%
MS	338	703	1043	1233	1196	4513	1,42%
RO	305	527	872	723	769	3196	1,01%
TO	193	344	445	577	577	2136	0,67%
Total	28959	50904	81362	83186	72388	316799	100%

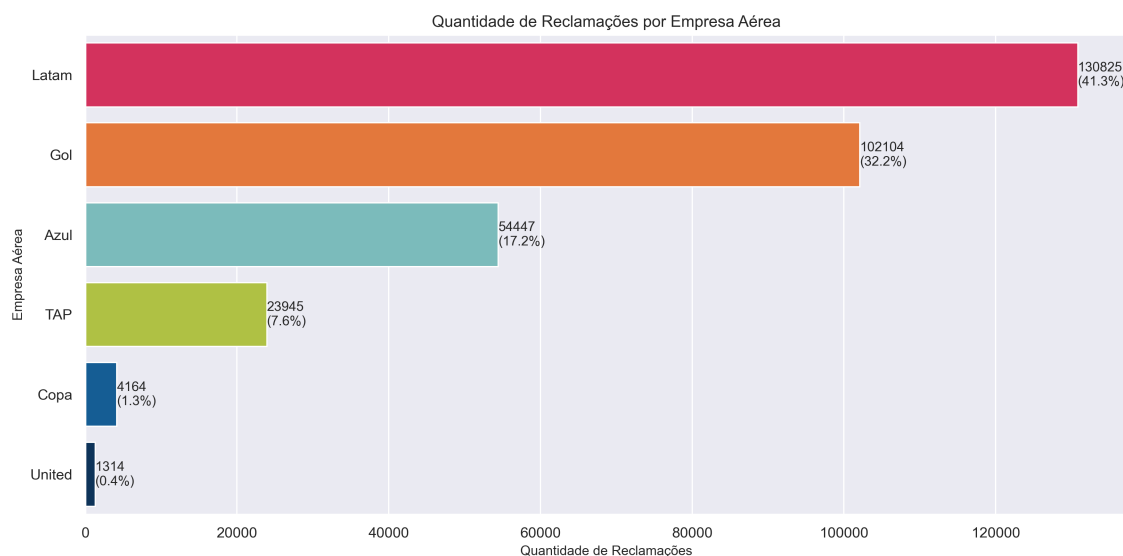
Fonte: Elaborado pela autora.

Na Tabela 12, observa-se que entre as quatro Unidades Federativas com maior número de reclamações, três pertencem ao Sudeste (SP, MG e RJ), enquanto a quarta é do Centro-Oeste (DF). Ao comparar os dados do DF com os da Figura 11, observa-se que o DF foi o responsável pela maior parte das reclamações na região Centro-Oeste ao longo dos anos. Esse destaque, apesar da população do DF ser menor em comparação aos outros estados, pode estar relacionado ao fluxo de passageiros no Aeroporto de Brasília, em função da localização no centro do país. Esse fenômeno é particularmente notável em comparação aos estados do Nordeste, que, embora tenham uma população maior, apresentam um número inferior de reclamações.

4.1.1.3 Visão por Empresa Aérea

A Figura 8 apresenta a quantidade de reclamações e o percentual relativo ao total por empresa aérea.

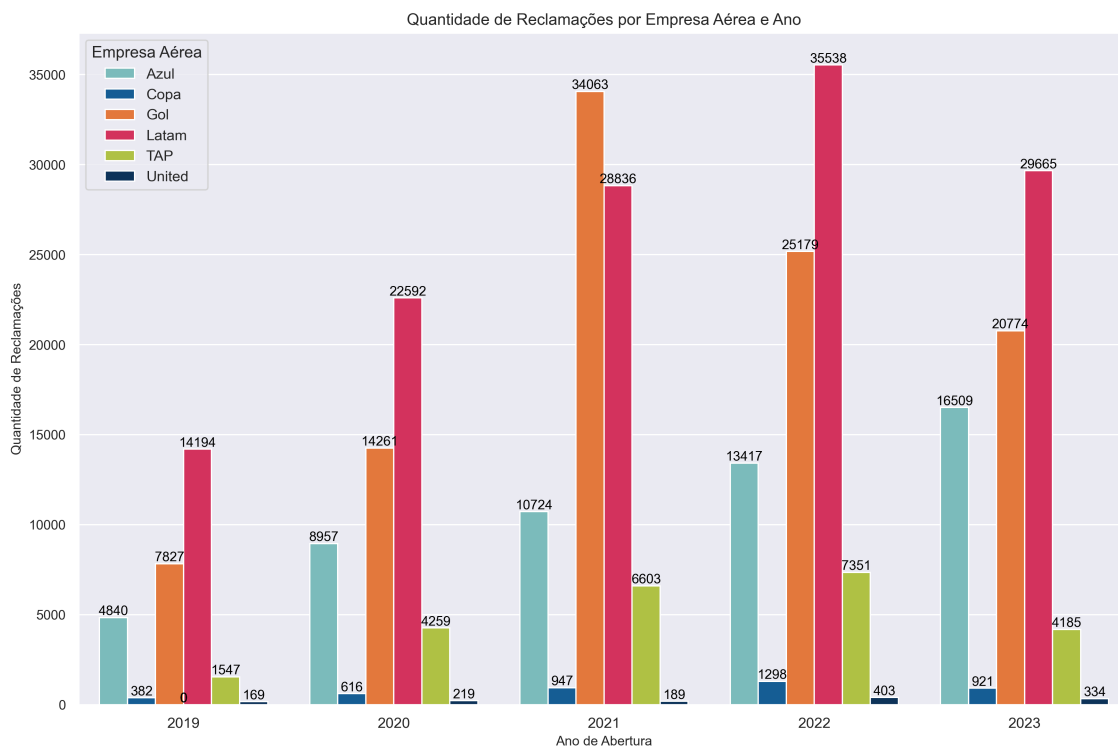
Figura 8 – Quantidade de Reclamações por Empresa Aérea



Fonte: Elaborada pela autora.

Na Figura 8, observa-se que a maioria das reclamações está relacionada às empresas nacionais, com destaque para a Latam (41,3%), seguida pela Gol (32,2%) e pela Azul (17,2%). Entre as internacionais, a principal foi a TAP (7,6%), seguida pela Copa (1,3%) e a United (0,4%). A Figura 9 apresenta a quantidade de reclamações por empresa aérea e ano.

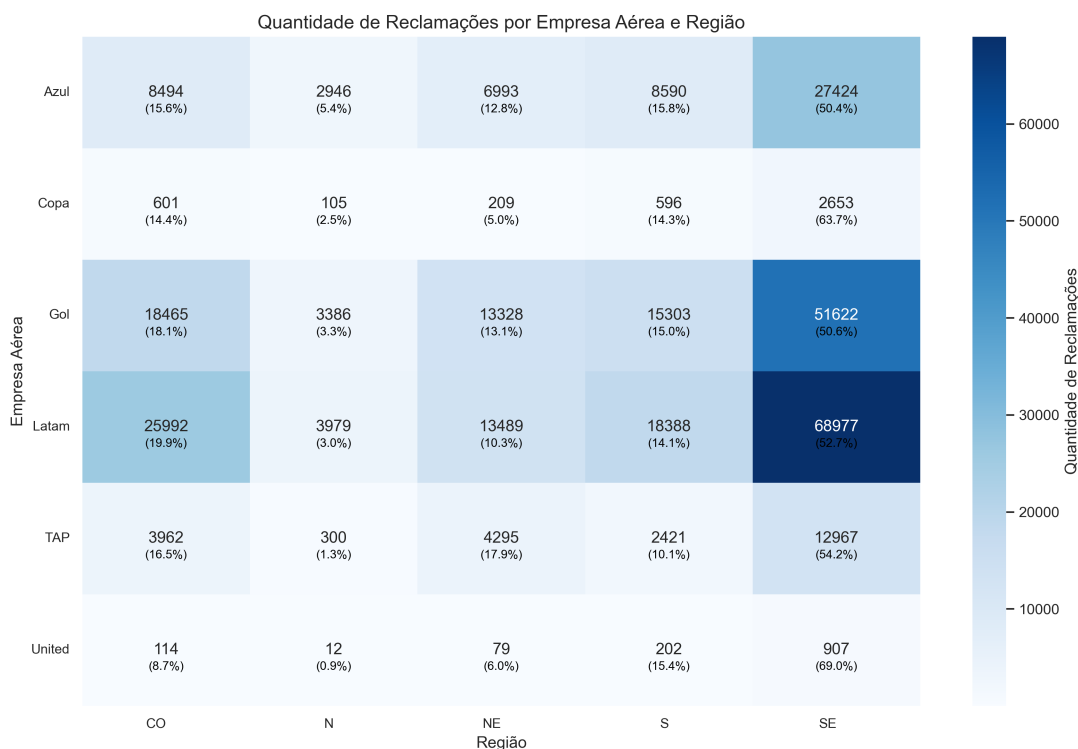
Figura 9 – Quantidade de Reclamações por Empresa Aérea e Ano



Fonte: Elaborada pela autora.

Na Figura 9, observa-se que a maioria das empresas apresentou um crescimento constante nas reclamações até 2022, seguido de uma queda acentuada em 2023. Entre as companhias nacionais, a Latam e a Gol se destacam, apresentando volumes significativamente superiores de reclamações ao longo dos anos. A Azul, por outro lado, apesar de ter mostrado um crescimento contínuo nas reclamações a cada ano, ainda se mantém na posição de empresa menos reclamada entre as nacionais. No caso das empresas internacionais, a TAP se destaca como a mais reclamada no período analisado. Vale ressaltar que a diferença no número de reclamações entre as empresas nacionais e internacionais também é influenciada pelo volume de voos, uma vez que as companhias nacionais realizam um número maior de decolagens e transportam mais passageiros, o que contribui para um aumento proporcional nas reclamações. Considerando que essa variação também ocorre por região, a Figura 10 apresenta um gráfico de calor da quantidade de reclamações por empresa e região.

Figura 10 – Quantidade de Reclamações por Empresa Aérea e Região



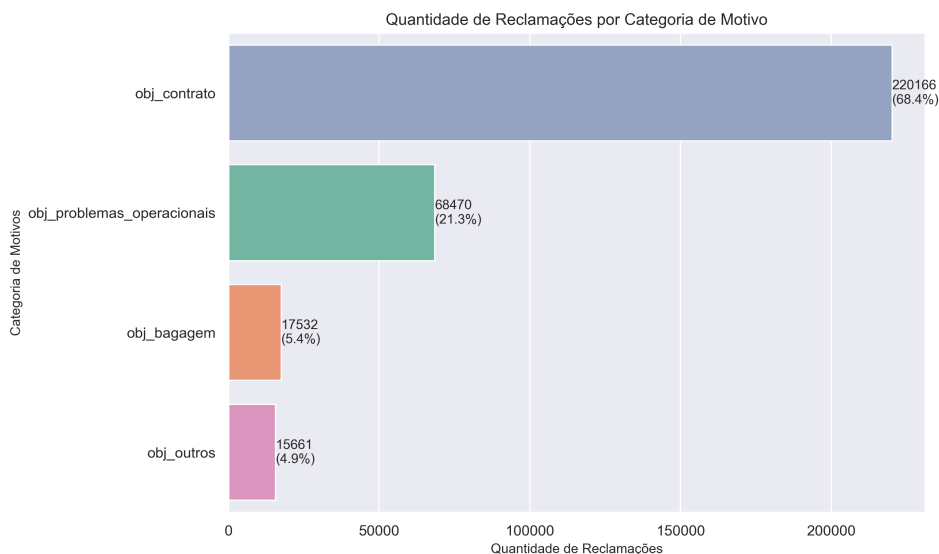
Fonte: Elaborada pela autora.

Na Figura 10, como era esperado, observa-se que a maioria das reclamações das empresas está concentrada nas regiões Sudeste, Centro-Oeste e Sul. Além disso, um destaque interessante é a TAP, que tem o Nordeste como a segunda região com o maior número de reclamações.

4.1.1.4 Motivos das reclamações

A Figura 11 apresenta a quantidade de reclamações e o percentual relativo ao total por categoria de motivo e a Tabela 13 apresenta a quantidade de reclamações por categoria de motivo e ano. Ressalta-se que, como uma reclamação pode ser enquadrada em mais de uma categoria, a contagem do total de reclamações por categoria pode superar o número de reclamações registradas.

Figura 11 – Quantidade de Reclamações por Categoria de Motivo



Fonte: Elaborada pela autora.

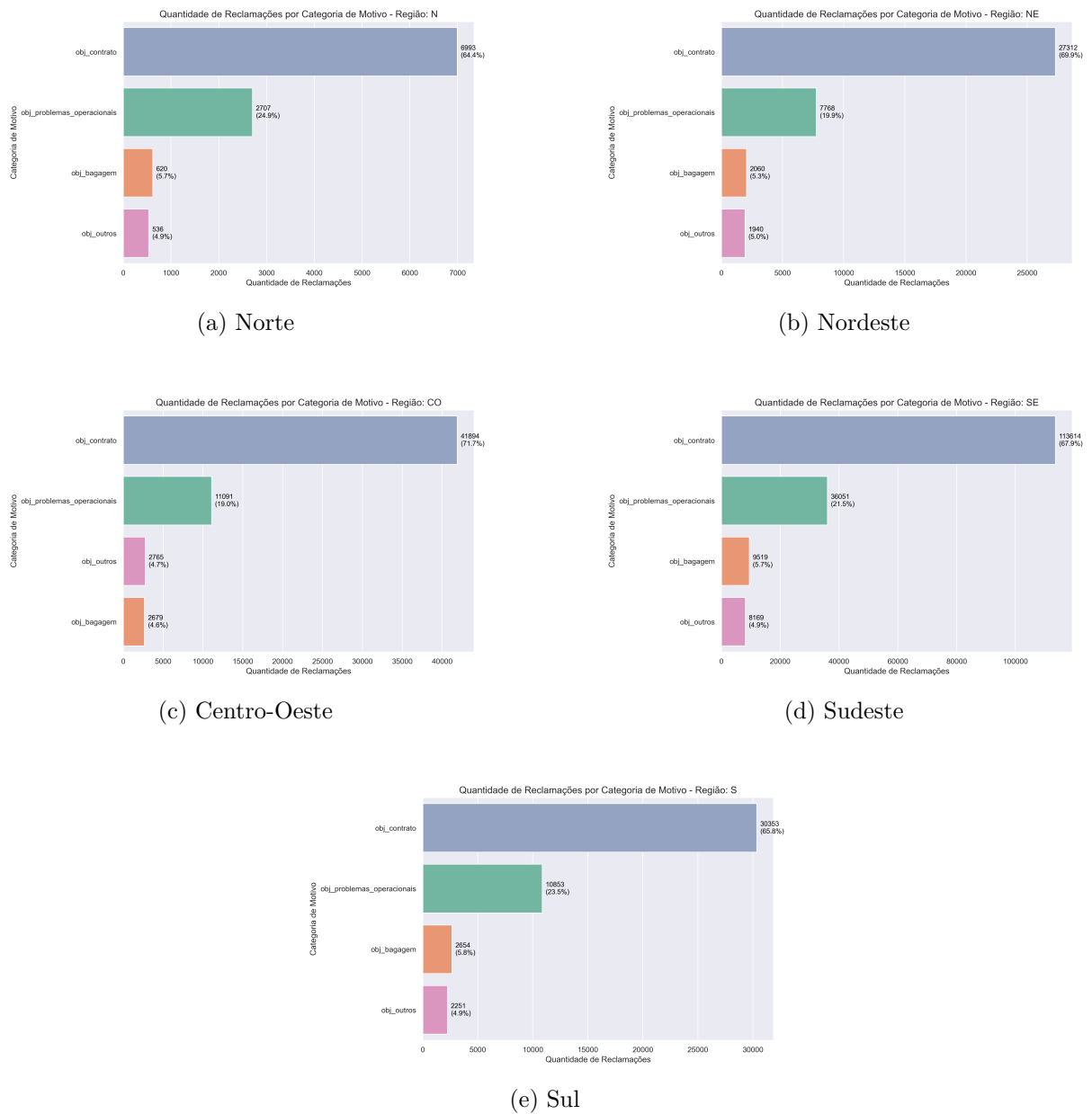
Tabela 13 – Quantidade de Reclamações por Categoria de Motivo e Ano

Categoria	2019	2020	2021	2022	2023	Total	Percentual do Total
obj_contrato	16241	38378	64903	60023	40621	220166	69,49%
obj_problemas_operacionais	9544	9870	12097	15359	21600	68470	21,61%
obj_bagagem	2710	1447	2078	5026	6271	17532	5,53%
obj_outros	988	1891	3767	4483	4532	15661	4,94%
Total	29483	51586	82845	84891	73024	316799	100%

Fonte: Elaborado pela autora.

Na Figura 11 e na Tabela 13, observa-se que a maior parte das reclamações registradas entre 2019 e 2023 está relacionada a contratos, representando 68,4% do total. Em segundo lugar, estão os problemas operacionais, com 21,3%. As categorias de bagagem e outros motivos apresentaram números bem próximos, com 5,4% e 4,9%, respectivamente. A Figura 12 apresenta a quantidade de reclamações por categoria de motivo e região.

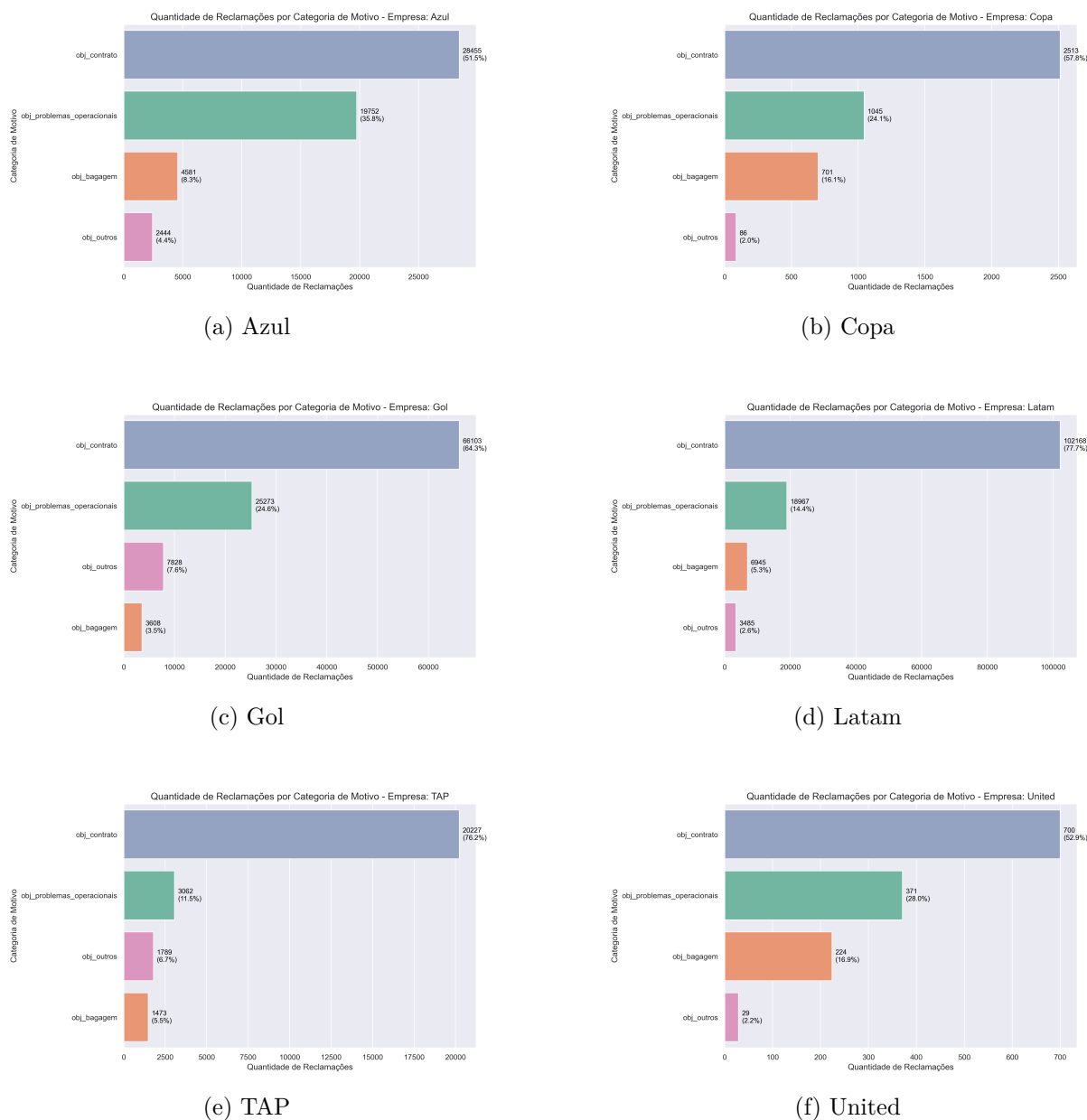
Figura 12 – Quantidade de Reclamações por Categoria de Motivo e Região



Fonte: Elaborada pela autora.

Na Figura 12, observa-se que a ordem das categorias de motivos das reclamações se manteve consistente em todas as regiões. A categoria contratos se destaca como principal fonte de reclamações, com percentuais variando entre 64% e 72%. A Figura 13 apresenta a quantidade de reclamações por categoria de motivo e por empresa aérea.

Figura 13 – Quantidade de Reclamações por Categoria de Motivo e Empresa Aérea



Fonte: Elaborada pela autora.

Na Figura 12, observa-se que a ordem de categorias de motivos também se manteve para todas as empresas. O principal motivo de reclamação, em todas as empresas, é a categoria de contratos, com percentuais variando entre 52% e 77%. A categoria de problemas operacionais também se destaca, especialmente na Azul, que apresenta o maior percentual entre as empresas, com 35,76% das reclamações relacionadas a esse motivo. Este dado é interessante, pois a Azul possui uma proporção maior de queixas relacionadas a problemas operacionais, em comparação com as demais empresas, como a Latam, que, apesar de ser a empresa aérea com mais reclamações, apresenta apenas 14,42% nesse mesmo quesito.

Por fim, considerando que cada categoria de motivo inclui diversos códigos ANAC, a Tabela 14 apresenta o detalhamento dos cinco códigos que mais receberam reclamações por ano.

Tabela 14 – Top 5 das Reclamações por Código ANAC e Ano

Ano Abertura	Código ANAC	Categoria de Motivo	Tema	Subtema	Assunto	Qtd de Reclamações	Percentual do Total do Ano
2019	1400	Contrato	Alteração pelo passageiro	Multas contratuais	-	2482	8,57%
	2225	Problemas operacionais	Alteração pela empresa aérea	Alternativas em alterações ou cancelamentos programados (reacomodação, reembolso integral ou outra modalidade de transporte)	Discorda das alternativas de reacomodação	2380	8,22%
	8000	Contrato	Programas de fidelidade	-	-	1638	5,66%
	0110	Contrato	Oferta e compra	Informações durante a compra da passagem	Regras e multas para no-show, cancelamento, remarcação e reembolso	1361	4,70%
	3325	Problemas operacionais	Execução do voo	Alternativas em atrasos, cancelamentos, interrupção do serviço ou preterição (reacomodação, reembolso integral ou outra modalidade de transporte)	Discorda das alternativas de reacomodação	1282	4,43%
2020	1400	Contrato	Alteração pelo passageiro	Multas contratuais	-	9069	17,82%
	0110	Contrato	Oferta e compra	Informações durante a compra da passagem	Regras e multas para no-show, cancelamento, remarcação e reembolso	5714	11,23%
	2225	Problemas operacionais	Alteração pela empresa aérea	Alternativas em alterações ou cancelamentos programados (reacomodação, reembolso integral ou outra modalidade de transporte)	Discorda das alternativas de reacomodação	4358	8,56%
	5300	Contrato	Reembolso	Prazo de reembolso	-	3553	6,98%
	5105	Contrato	Reembolso	Valor do reembolso	Reembolso decorrente de alteração da passagem pelo passageiro	2744	5,39%
2021	0110	Contrato	Oferta e compra	Informações durante a compra da passagem	Regras e multas para no-show, cancelamento, remarcação e reembolso	15062	18,51%
	5300	Contrato	Reembolso	Prazo de reembolso	-	10793	13,27%
	8000	Contrato	Programas de fidelidade	-	-	5721	7,03%
	2225	Problemas operacionais	Alteração pela empresa aérea	Alternativas em alterações ou cancelamentos programados (reacomodação, reembolso integral ou outra modalidade de transporte)	Discorda das alternativas de reacomodação	5296	6,51%
	1815	Contrato	Alteração pelo passageiro	Funcionamento dos canais para alterações contratuais	Dificuldade em remarcar o voo	3209	3,94%
2022	8000	Contrato	Programas de fidelidade	-	-	10456	12,57%
	5300	Contrato	Reembolso	Prazo de reembolso	-	9822	11,81%
	0110	Contrato	Oferta e compra	Informações durante a compra da passagem	Regras e multas para no-show, cancelamento, remarcação e reembolso	6333	7,61%
	2225	Problemas operacionais	Alteração pela empresa aérea	Alternativas em alterações ou cancelamentos programados (reacomodação, reembolso integral ou outra modalidade de transporte)	Discorda das alternativas de reacomodação	6246	7,51%
	1400	Contrato	Alteração pelo passageiro	Multas contratuais	-	4666	5,61%
2023	2225	Problemas operacionais	Alteração pela empresa aérea	Alternativas em alterações ou cancelamentos programados (reacomodação, reembolso integral ou outra modalidade de transporte)	Discorda das alternativas de reacomodação	8952	12,37%
	8000	Contrato	Programas de fidelidade	-	-	7274	10,05%
	5300	Contrato	Reembolso	Prazo de reembolso	-	5887	8,13%
	1400	Contrato	Alteração pelo passageiro	Multas contratuais	-	5836	8,06%
	9900	Outros	Outros	Outros	-	4467	6,17%

Fonte: Elaborado pela autora.

Na Tabela 14, observa-se que, em 2019, a maioria das reclamações foi devido a multas contratuais de alterações feitas pelo passageiro (8,57%). Em segundo lugar, ficaram as reclamações devido a discordância das alternativas de acomodação de alterações feitas pela empresa aérea (8,21%). As demais reclamações foram a respeito de programas de fidelidade (5,65%), regras e multas para no-show, cancelamento, remarcação e reembolso no momento da oferta e compra (4,69%) e discordância das alternativas de acomodação durante a execução do voo (4,42%).

Em 2020, as multas contratuais de alterações feitas pelo passageiro (17,81%) continuaram em primeiro lugar, seguidas pelas reclamações sobre regras e multas para no-show, cancelamento, remarcação e reembolso no momento da oferta e compra (11,22%) e pela discordância das alternativas de acomodação de alterações feitas pela empresa aérea (8,56%). As duas últimas posições foram relacionadas a problemas com reembolso, referentes a prazo (6,97%) e valor decorrente de alteração da passagem pelo passageiro (5,39%).

Em 2021, as reclamações foram predominantemente relacionadas regras e multas para no-show, cancelamento, remarcação e reembolso no momento da oferta e compra (18,51%). Em segundo lugar, figuraram as reclamações sobre o prazo de reembolso (13,27%), seguidas pelas questões envolvendo programas de fidelidade (7,03%). A discordância das alternativas de acomodação de alterações feitas pela empresa aérea permaneceu relevante (6,51%). Outra reclamação significativa foi a dificuldade em remarcar o voo, relacionada às alterações feitas pelo passageiro (3,94%).

Em 2022, o padrão de reclamações continuou a refletir questões contratuais e de reembolso. As reclamações sobre programas de fidelidade lideraram (12,57%), seguidas de perto pelas reclamações sobre o prazo de reembolso (11,81%). Também houve um número considerável de reclamações sobre regras e multas para no-show, cancelamento, remarcação e reembolso no momento da oferta e compra (7,61%). Outros temas relevantes em 2022 foram a discordância das alternativas de acomodação de alterações feitas pela empresa aérea (7,51%) e as multas contratuais de alterações feitas pelo passageiro (5,61%).

Já em 2023, as reclamações sobre a discordância das alternativas de acomodação de alterações feitas pela empresa aérea lideraram (12,37%). O segundo lugar foi ocupado novamente pelos programas de fidelidade (10,05%), seguidos de perto pelas reclamações sobre o prazo de reembolso (8,13%) e pelas multas contratuais de alterações feitas pelo passageiro (8,06%). Um novo tema que emergiu em 2023 foi a categoria "Outros", que representou 6,17% das reclamações, abrangendo casos diversos não classificados nas categorias anteriores.

Essa sequência de temas ao longo dos anos demonstra uma continuidade de questões centradas nas alterações de passagens e reembolsos, com uma ênfase crescente em problemas relativos aos programas de fidelidade e à acomodação dos passageiros em casos de

cancelamento ou alteração de voos. Esses resultados corroboram com os achados de Korfiatis *et al.* (2019), Kwon *et al.* (2021), Çallı e Çallı (2023) e Kim, Lim e Ha (2024), destacando a insatisfação com reembolsos e cancelamentos como uma questão recorrente.

A respeito das reclamações sobre reembolso, de acordo com ANAC (2023b), elas foram incrementadas pelo contexto excepcional da pandemia de COVID-19, visto que ela fez com que medidas emergenciais para a aviação civil fossem adotadas, como a Lei nº 14.034/2020 alterada pela Lei nº 14.174/2021. Essa lei flexibilizava as regras para alteração de passagens, cancelamento, reembolso e crédito e se aplicou a situações ocorridas até o final do ano de 2021, o que gerou reflexos ainda no ano de 2022. Percebe-se que, de fato, em 2020 e 2021, os problemas relacionados a reembolsos foram mais predominantes, o que corrobora os achados de Çallı e Çallı (2023) e Kim, Lim e Ha (2024), que também observaram um aumento nas queixas sobre reembolsos e problemas correlatos durante a COVID-19.

4.1.2 Dados de judicialização do transporte aéreo

Este tópico apresenta os resultados da análise exploratória dos 555.546 processos judiciais analisados. As Tabelas 15 e 16 apresentam as estatísticas descritivas das variáveis.

Tabela 15 – Judicialização: Estatísticas Descritivas - Variáveis Categóricas

Variáveis	count	unique	top	freq
empresa_re	555546	6	Azul	206654
data_distribuicao	555436	2283	01/09/2023	2801
UF	555546	16	SP	100690
operacao	555546	3	Doméstico	365341
jurisdicao	555546	3	Juizado Especial Cível	412065
instancia	555546	4	1ª Instância	322840
situacao_processual_sentenca	555546	5	Outros	304091
data_voo	486151	4407	-	58618
obj_principal	489101	792	Atraso ou Cancelamento de Voo	80710
subobj_principal	347341	849	***	142151
causa_alegada	445426	180	Culpa Tam	110837
regiao	555546	5	SE	244507

Fonte: Elaborado pela autora.

Tabela 16 – Judicialização: Estatísticas Descritivas - Variáveis Numéricas

Variáveis	count	mean	std	min	25%	50%	75%	max
num_proc_ano	555546	2021,30	1,49	2019	2020	2022	2023	2023
n_autores	555546	1,08	0,79	0	1	1	1	61
valor_pedido_inicial	555546	17809,65	118811,83	0,00	10000,00	15000,00	20727,79	61791950,00
valor_dano_moral	555546	1856,66	4047,84	0,00	0,00	0,00	3000,00	772819,84
valor_dano_material	555546	381,81	13627,44	0,00	0,00	0,00	1000,00	10000000,00
valor_total_condenacao	555546	1506,39	4980,28	0,00	0,00	0,00	1000,00	1188888,00
valor_pago	555546	2127,99	12197,07	0,00	0,00	0,00	3000,00	7889160,00
mes_distribuicao	555546	6,42	3,49	0	3	6	10	12

Fonte: Elaborado pela autora.

As estatísticas das Tabelas 15 e 16 apresentam uma visão geral sobre o comportamento e a distribuição dos dados relacionados aos processos judiciais. Em relação as variáveis associadas ao processo de tratativa dos processos judiciais, foram observados os seguintes pontos:

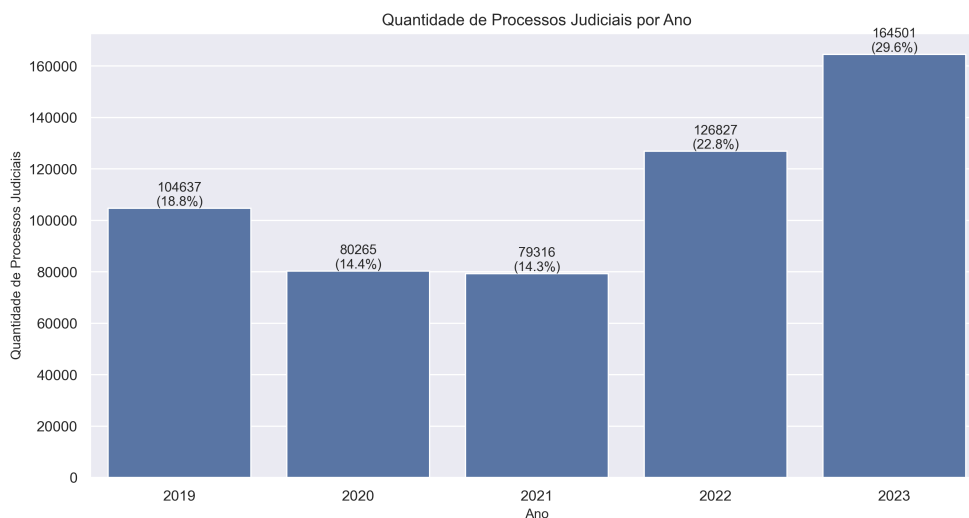
- **Operação (operacao):** 65,76% dos processos foram relacionados a voos domésticos, 18,22% a voos internacionais e 16,01% não foram informados;
- **Jurisdição (jurisdicao):** a maioria dos processos tramitaram em Juizados Especiais Cíveis (74,17%), que têm competência para conciliar, processar e julgar as causas cíveis de menor complexidade, incluindo causas cujo valor não exceda quarenta vezes o salário mínimo. Os demais processos tramitaram em Varas Cíveis (24,50%) e Outros (1,31%), categoria que abrange processos nos quais não foi possível identificar a jurisdição;
- **Situação processual e sentença (situacao_processual_sentenca):** a maioria dos processos teve sua situação ou sentença não informada, sendo classificados na categoria Outros (54,73%). Entre os demais, 22,68% terminaram em condenação, 14,62% em acordo, 6,99% tiveram decisão favorável à aérea (improcedência) e 0,95% foram extintos sem julgamento de mérito;
- **Instância (instanci):** a maioria das sentenças ou decisões foram proferidas na 1ª Instância (58,11%);
- **Número de autores (n_autores):** a maioria dos processos teve apenas 1 autor (66,91%). Tendo em vista isso e que havia muitos registros sem essa informação (15,63% foram classificados como zero), a média de autores ficou em 1,08. Vale ressaltar que houve um processo com 61 autores (máximo).

Os subtópicos a seguir oferecem uma análise mais detalhada das variáveis de período, localização, empresa, motivo, causa e danos dos processos judiciais.

4.1.2.1 Visão por Ano e Mês

A Figura 14 apresenta a quantidade de processos judiciais e o percentual relativo ao total por ano.

Figura 14 – Quantidade de Processos Judiciais por Ano de Abertura



Fonte: Elaborada pela autora.

Observa-se na Figura 14 que, ao contrário das reclamações, que apresentaram um crescimento contínuo de 2019 a 2022, a quantidade de processos judiciais reduziu em 2020 e 2021. Apesar disso, em todos os anos, a quantidade de processos superou a de reclamações, com exceção de 2021, quando foram registradas 81.362 reclamações. A partir de 2022, o cenário mudou significativamente, com um crescimento expressivo de 59,9% no número de processos judiciais. Em 2023, o crescimento continuou, embora em um ritmo menor, com um aumento de 29,7% em comparação a 2022, totalizando 164.501 processos.

Tendo em vista a queda no número de passageiros transportados nesses anos, a Tabela 17 apresenta o índice de processos judiciais por 100 mil passageiros pagos transportados, em comparação ao mesmo índice de reclamações.

Tabela 17 – Índice de Processos Judiciais e Reclamações por 100 mil Passageiros por Ano

Ano	Total de Passageiros Pagos Transportados	Total de Processos Judiciais	Índice de Processos Judiciais por 100 mil Passageiros	Índice de Reclamações por 100 mil Passageiros
2019	118.677.247	104.637	88,2	24,4
2020	51.934.069	80.265	154,6	98,0
2021	67.343.710	79.316	117,8	120,8
2022	97.686.945	126.827	129,8	85,2
2023	112.698.015	164.501	146,0	64,2

Fonte: Elaborado pela autora.

Na Tabela 17, observa-se que, em todos os anos analisados, o índice de processos judiciais foi superior ao de reclamações, exceto em 2021. Isso revela um alto grau de judicialização no setor aéreo, indicando que um número maior de consumidores opta por

buscar a justiça para resolver seus problemas em vez de tentar solucioná-los diretamente com as empresas por meio de reclamações. O ano de 2020 apresentou o maior índice de processos judiciais, com 154,6 por 100 mil passageiros. No entanto, em 2023, esse índice se aproximou significativamente, alcançando 146. Vale destacar que o número de passageiros em 2023 foi mais do que o dobro do registrado em 2020, o que reforça a tendência crescente de judicialização no setor. Quanto ao comportamento das reclamações ao longo dos meses, a Tabela 18 apresenta a quantidade de processos judiciais por mês e ano.

Tabela 18 – Quantidade de Processos Judiciais por Mês e Ano de Abertura

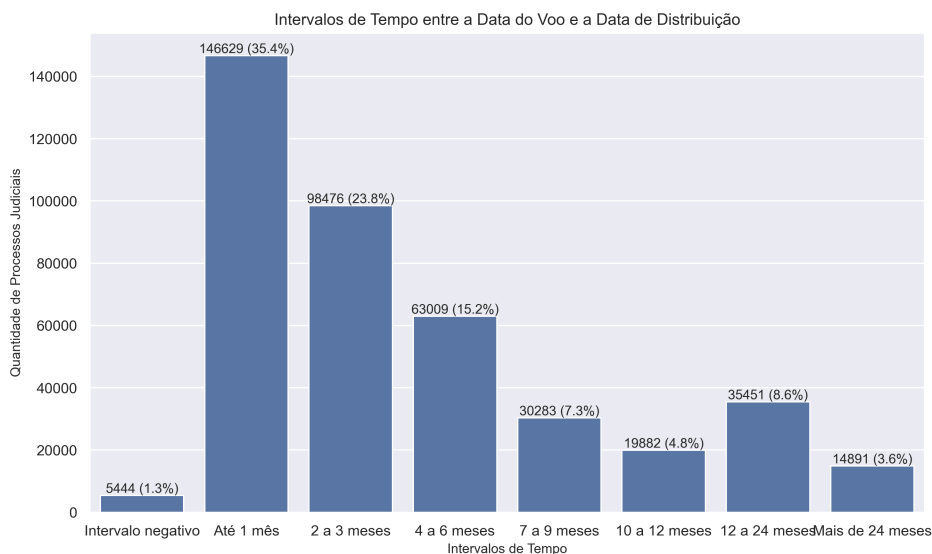
Mês de Abertura	2019	2020	2021	2022	2023	Total	Percentual do Total
Janeiro	6993	10534	5618	10102	11336	44583	8,03%
Fevereiro	8140	10925	6612	12562	12440	50679	9,12%
Março	8646	11174	8006	12645	16330	56801	10,22%
Abril	8545	5672	6296	9888	12025	42426	7,64%
Maio	9015	5481	6004	10762	13906	45168	8,13%
Junho	6884	5323	6017	9976	12884	41084	7,40%
Julho	8789	5752	6346	10006	13353	44246	7,96%
Agosto	9439	4549	6307	11342	12416	44053	7,93%
Setembro	9155	4665	6064	10169	15565	45618	8,21%
Outubro	10265	5210	6477	10308	14999	47259	8,51%
Novembro	9832	5481	7819	9886	15915	48933	8,81%
Dezembro	8887	5489	7729	8393	13198	43696	7,87%
Não informado	47	10	21	788	134	1000	0,18%
Total	104637	80265	79316	126827	164501	555546	100%

Fonte: Elaborado pela autora.

Na Tabela 18, observa-se que os processos judiciais estão distribuídos de maneira relativamente equilibrada ao longo dos meses. Assim como nas reclamações, o mês com o maior percentual de processos foi março (10,22%), seguido por fevereiro (9,12%), novembro (8,81%), outubro (8,51%), setembro (8,21%), maio (8,13%) e janeiro (8,03%). Em termos gerais, a distribuição dos processos segue um padrão similar ao das reclamações, com maior concentração nos meses que coincidem com férias escolares e feriados importantes.

De forma complementar, a Figura 15 apresenta a quantidade de processos judiciais e o percentual relativo ao total por intervalo de tempo em meses entre a data do voo e a data de distribuição, isto é, quando o processo judicial foi iniciado.

Figura 15 – Intervalos de Tempo entre a Data do Voo e a Data de Distribuição



Fonte: Elaborada pela autora.

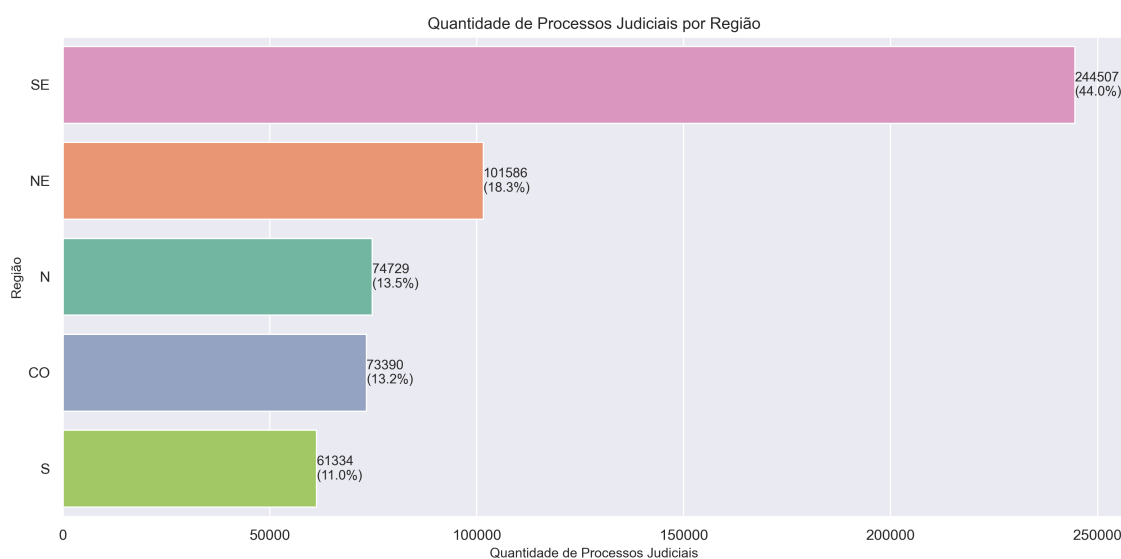
Na Figura 15, observa-se que o intervalo de tempo entre a data do voo e a data de distribuição dos processos judiciais apresenta uma variação considerável. A maior parte dos processos está concentrada nos intervalos de 1 a 3 meses, seguida pela faixa de 4 a 6 meses. O menor intervalo corresponde ao valor negativo, que refere-se a casos em que a data de distribuição informada era anterior à data do voo, sugerindo possíveis erros de registro. A categoria 'Até 1 mês' se destaca com uma quantidade significativa de processos (35,4%), o que indica que, em muitos casos, o consumidor busca a justiça em um curto período após o voo. Por outro lado, a presença de processos com intervalos de 12 a 24 meses (8,6%) e mais de 24 meses (3,6%) chama a atenção, refletindo uma possível demora por parte do consumidor em buscar uma solução. Algumas hipóteses para essa demora incluem a tentativa de resolver o problema por outros meios antes de recorrer à justiça, sem obter um resultado satisfatório, ou a possível influência de terceiros, como empresas especializadas, chamadas *civic techs*². Contudo, a confirmação dessas hipóteses exigiriam análises aprofundadas desses processos, caso a caso, em busca de explicações, o que não está no escopo deste trabalho.

² *Civic techs* são empresas que utilizam de recursos tecnológicos e de redes sociais para captar passageiros com alguma insatisfação com o serviço prestado e os estimula a buscar a justiça como alternativa de reparação, às vezes sem ao menos recorrer primeiramente a própria companhia aérea para manifestar a insatisfação. Essas empresas muitas vezes entram com ações de danos morais, o que contradiz as legislações vigentes. Ademais, impedem a resolução do problema de forma administrativa, tendo em vista compensações judiciais mais lucrativas, o que sobrecarrega o sistema judiciário e gera altos custos às empresas aéreas (Pedroso, 2023).

4.1.2.2 Visão por Região e UF

A Figura 16 apresenta a quantidade de processos judiciais e o percentual relativo ao total por região. Ressalta-se que, apenas os 16 estados brasileiros foram incluídos nos dados dessa pesquisa. Dessa forma, cada região considera as seguintes UFs: Norte (AM, RO e TO), Nordeste (BA, CE e PE), Centro-Oeste (DF, GO, MS e MT), Sudeste (ES, MG, RJ e SP) e Sul (PR e SC).

Figura 16 – Quantidade de Processos Judiciais por Região



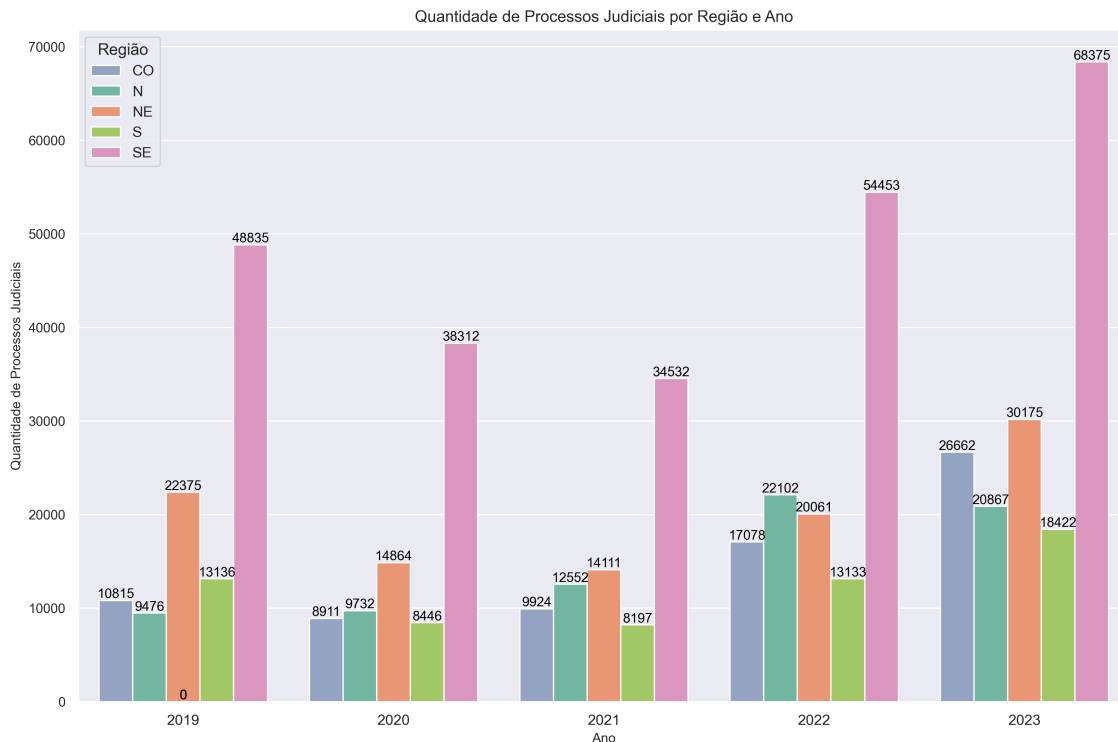
Fonte: Elaborada pela autora.

Na Figura 14, observa-se que o Sudeste é a região com a maior concentração de processos judiciais, com 44%, seguido pelo Nordeste (18,3%), Norte (13,5%), Centro-Oeste (13,2%) e Sul (11%). Considerando que o Sudeste e o Nordeste são as regiões com o maior número de passageiros e decolagens³, é compreensível que essas regiões apresentem também o maior número de processos. No entanto, chama a atenção o caso da região Norte, que ocupa a terceira posição, com 13,5% dos processos, apesar de ser a região com a menor participação no número de passageiros e de decolagens. Esse dado sugere que, embora a quantidade de voos e passageiros seja significativamente menor, a judicialização na região Norte é proporcionalmente alta, o que pode indicar problemas locais específicos no setor aéreo, os quais podem estar gerando insatisfação entre os consumidores e, conseqüentemente, mais disputas judiciais. Além disso, os dados corroboram as informações apresentadas no referencial, que indicam que as companhias aéreas têm reduzido ou até cortado a oferta de

³ Participação no número de passageiros pagos por região - 2023 - mercado doméstico: Sudeste (50,1%), Nordeste (19,3%), Sul (12,9%), Centro-Oeste (12,3%) e Norte (5,4%) (ANAC, 2024a).
Participação no número de decolagens por região - 2023 - mercado doméstico: Sudeste (50,8%), Nordeste (18,7%), Sul (13,2%), Centro-Oeste (11,2%) e Norte (6,1%) (ANAC, 2024a).

voos entre cidades do Norte do país devido ao elevado número de ações judiciais na região. A Figura 11 apresenta a quantidade de processos judiciais por região e ano.

Figura 17 – Quantidade de Processos Judiciais por Região e Ano



Fonte: Elaborada pela autora.

Na Figura 17, observa-se que o Centro-Oeste e o Norte apresentaram aumentos expressivos nos processos judiciais entre 2021 e 2023. O Centro-Oeste registrou um crescimento de 72,1% em 2022 e 56,1% em 2023. Enquanto o Norte, embora tenha registrado uma queda de 5,6% em 2023, teve um aumento de 76,1% em 2022. As demais regiões, embora tenham apresentado uma diminuição nos processos em 2020 e 2021, também registraram aumentos nos anos subsequentes. O Nordeste teve um crescimento de 42,1% em 2022 e 50,4% em 2023. No Sul, os crescimentos foram de 60,2% em 2022 e 40,2% em 2023. Por fim, o Sudeste registrou aumentos de 57,6% em 2022 e 25,6% em 2023. A Tabela 19 apresenta a quantidade de processos judiciais por UF e ano.

Tabela 19 – Quantidade de Processos Judiciais por UF e Ano

UF	2019	2020	2021	2022	2023	Total	Percentual do Total
SP	19305	16832	13734	22596	28223	100690	18,12%
RJ	14677	10416	10862	15519	18716	70190	12,63%
BA	12443	8486	8762	12208	18261	60160	10,83%
MG	11756	9006	7775	12972	17405	58914	10,60%
RO	6830	6765	8735	16633	12474	51437	9,26%
PR	9268	5708	5693	8862	11267	40798	7,34%
MT	2840	2882	3081	7063	12984	28850	5,19%
PE	7390	4429	3646	5197	7204	27866	5,02%
SC	3868	2738	2504	4271	7155	20536	3,70%
AM	2007	2492	3278	4755	7306	19838	3,57%
GO	2727	2384	2845	4220	6218	18394	3,31%
DF	3505	2401	2559	3438	4094	15997	2,88%
ES	3097	2058	2161	3366	4031	14713	2,65%
CE	2542	1949	1703	2656	4710	13560	2,44%
MS	1743	1244	1439	2357	3366	10149	1,83%
TO	639	475	539	714	1087	3454	0,62%
Total	104637	80265	79316	126827	164501	555546	100%

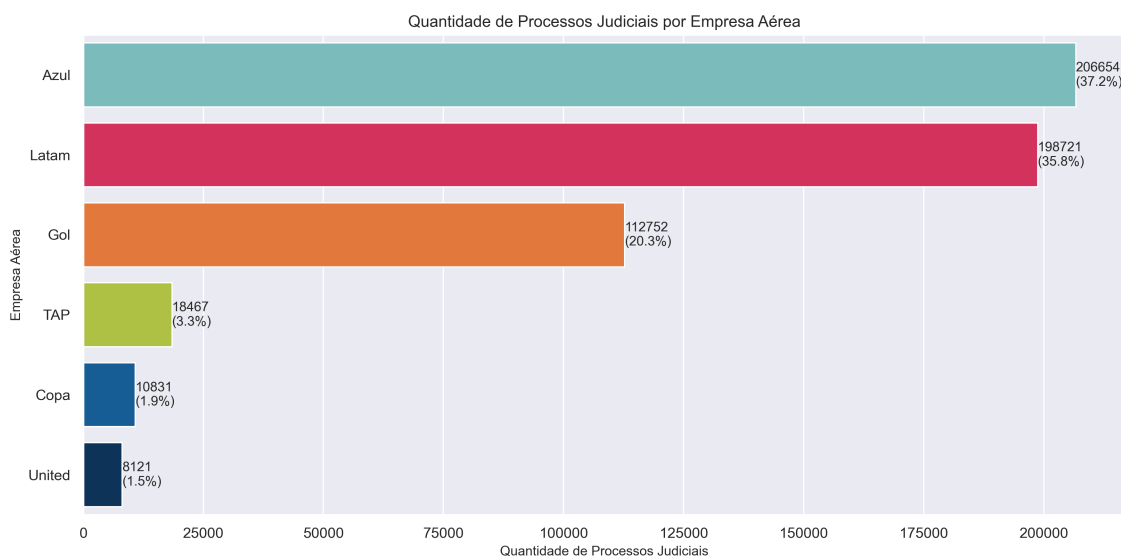
Fonte: Elaborado pela autora.

Na Tabela 19, observa-se que, entre as seis Unidades Federativas com maior número de processos judiciais, três pertencem ao Sudeste (SP, RJ e MG), enquanto as demais são Bahia, Rondônia e Paraná. Em relação a Rondônia, a UF apresentou um aumento considerável de processos a partir de 2021, destacando-se significativamente em comparação com outras UF's da região Norte, como Amazonas e Tocantins. O número de processos também cresceu substancialmente em estados como Pernambuco e Mato Grosso. Apesar desse crescimento, algumas Unidades Federativas, como o Distrito Federal, que registrou um alto volume de reclamações, apresentaram números mais baixos de processos, com participação inferior a 3% do total.

4.1.2.3 Visão por Empresa Aérea

A Figura 18 apresenta a quantidade de reclamações e o percentual relativo ao total por empresa aérea.

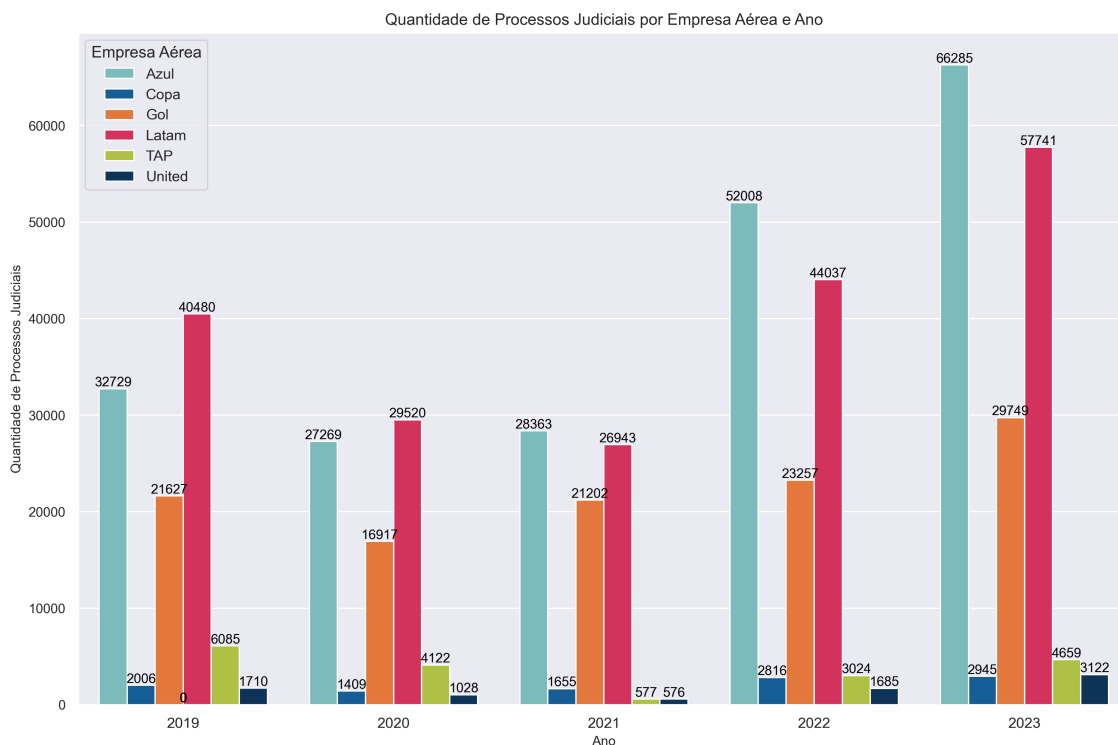
Figura 18 – Quantidade de Processos Judiciais por Empresa Aérea



Fonte: Elaborada pela autora.

Na Figura 18, observa-se que a maioria dos processos está relacionada às empresas nacionais, com destaque para a Azul (37,2%), seguida pela Latam (35,8%) e pela Gol (20,3%). Entre as empresas internacionais, a principal foi a TAP (3,3%), seguida pela Copa (1,9%) e pela United (1,5%). Comparando com o ranking por quantidade de reclamações, as empresas internacionais mantiveram a mesma ordem, enquanto nas empresas nacionais, a Azul, que possui o menor número de reclamações, foi a que teve o maior volume de processos. Destaca-se, entretanto, que as companhias aéreas adotaram critérios distintos para informar os processos em suas bases de dados. A Gol, por exemplo, excluiu os processos relacionados ao seu programa de fidelidade, enquanto as demais incluíram esses processos. Isso pode explicar por que a Gol apresentou um volume menor em comparação à Azul e à Latam. A Figura 19 apresenta a quantidade de reclamações por empresa aérea e ano.

Figura 19 – Quantidade de Processos Judiciais por Empresa Aérea e Ano



Fonte: Elaborada pela autora.

Na Figura 19, observa-se que, em 2023, foi registrado o maior número de processos judiciais para todas as empresas. A Azul, a Copa e a Gol apresentaram uma queda no número de processos em 2020, mas tiveram um crescimento constante a partir de 2021. Por outro lado, a Latam, a TAP e a United reduziram o volume de processos em 2020 e 2021, mas voltaram a crescer a partir de 2022. De forma geral, as empresas apresentaram aumentos significativos de 2021 para 2022: a TAP teve um aumento de 424,09%, a United de 192,53%, a Latam de 63,45%, a Azul de 83,36% e a Copa de 70,15%. A Gol foi a única empresa em que o crescimento ao longo dos anos foi mais gradual.

Assim como nas reclamações, a diferença no número de processos judiciais entre as empresas nacionais e internacionais também é influenciada pelo volume de voos, já que as companhias nacionais realizam mais decolagens e transportam mais passageiros. Considerando que essa variação também ocorre por região, a Figura 20 apresenta um gráfico de calor da quantidade de processos por empresa e região.

Figura 20 – Quantidade de Processos Judiciais por Empresa Aérea e Região



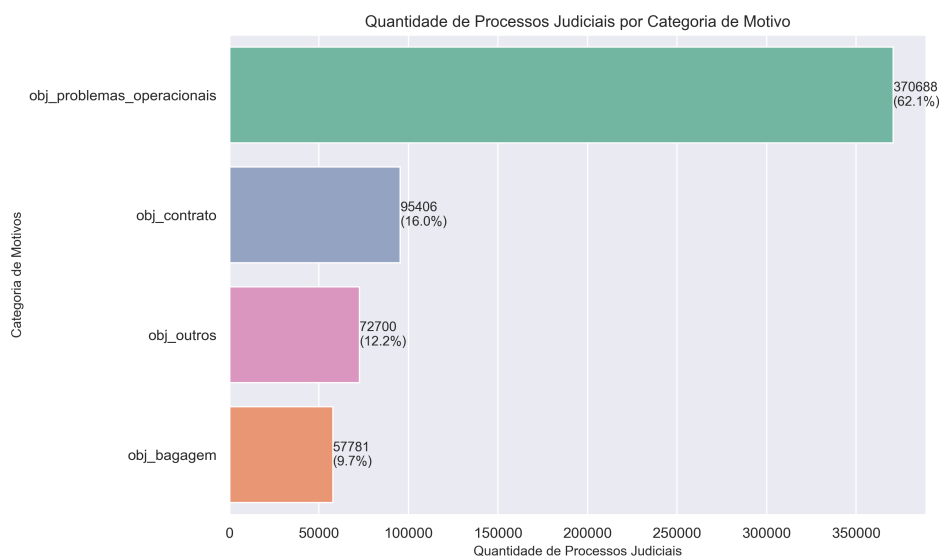
Fonte: Elaborada pela autora.

Na Figura 20, observa-se que a maioria dos processos judiciais contra as empresas está concentrada na região Sudeste, seguido pelo Nordeste. No caso da Azul, embora a maior parte dos processos esteja concentrada no Sudeste (34,4%), observa-se uma distribuição relativamente equilibrada entre as outras regiões: Norte (19,8%), Nordeste (19,5%) e Centro-Oeste (16,1%). Esse padrão sugere um descontentamento abrangente dos consumidores com os serviços da empresa, indicando que os problemas não são específicos de uma região.

4.1.2.4 Motivos dos processos judiciais

Os motivos dos processos judiciais corresponde ao objetivo e subobjeto principal, isto é, aquilo que motivou o(s) autor(es) a iniciar o processo. A Figura 21 apresenta a quantidade de processos judiciais e o percentual relativo ao total por categoria de motivo e a Tabela 20 apresenta a quantidade de processos judiciais por categoria de motivo e ano. Ressalta-se que, como um processo pode ser enquadrado em mais de uma categoria, a contagem do total de processos por categoria pode superar o número de processos registrados.

Figura 21 – Quantidade de Processos Judiciais por Categoria de Motivo



Fonte: Elaborada pela autora.

Tabela 20 – Quantidade de Processos Judiciais por Categoria de Motivo e Ano

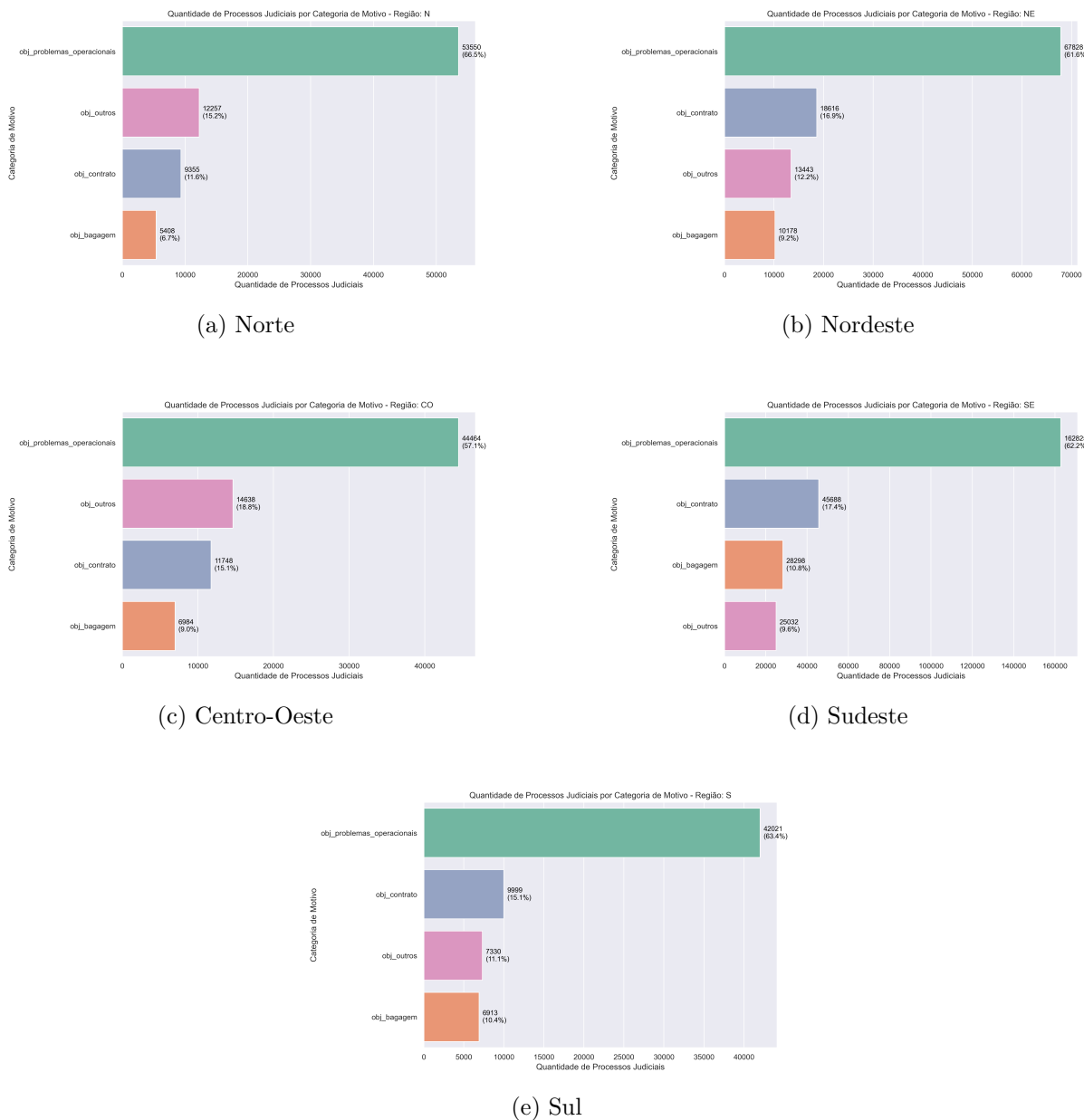
Categoria	2019	2020	2021	2022	2023	Total	Percentual do Total
obj_problemas_operacionais	84318	61279	58981	94700	71410	370688	62,13%
obj_contrato	12344	16064	24206	27342	15450	95406	15,99%
obj_outros	1029	879	823	1360	68609	72700	12,18%
obj_bagagem	12008	8188	6741	16700	14144	57781	9,68%
Total	109699	86410	90751	140102	169613	596575	100%

Fonte: Elaborado pela autora.

Na Figura 21 e na Tabela 20, observa-se que a maior parte dos processos judiciais registrados entre 2019 e 2023 está relacionada a problemas operacionais, representando 62,13% do total. Em segundo lugar, aparecem os processos de contrato, com 15,99% do total. A categoria Outros ficou em terceiro lugar (12,18%), devido a uma grande quantidade de registros com dados faltantes em 2023. Por fim, os processos relacionados à bagagem ocuparam o quarto lugar, representando 9,68% do total.

Quando comparados aos dados de reclamações, apresentados na Figura 11 e na Tabela 13, observa-se uma inversão na participação das categorias de problemas operacionais e contrato. Enquanto nos processos judiciais os problemas operacionais predominam, nas reclamações, os problemas relacionados a contratos lideram, com 68,4% do total. Essas diferenças podem refletir variações nas formas de resolução de conflitos pelos consumidores, com uma maior concentração de processos judiciais envolvendo problemas operacionais e um volume mais alto de reclamações relacionadas a questões contratuais. A Figura 22 apresenta a quantidade de processos judiciais por categoria de motivo e região.

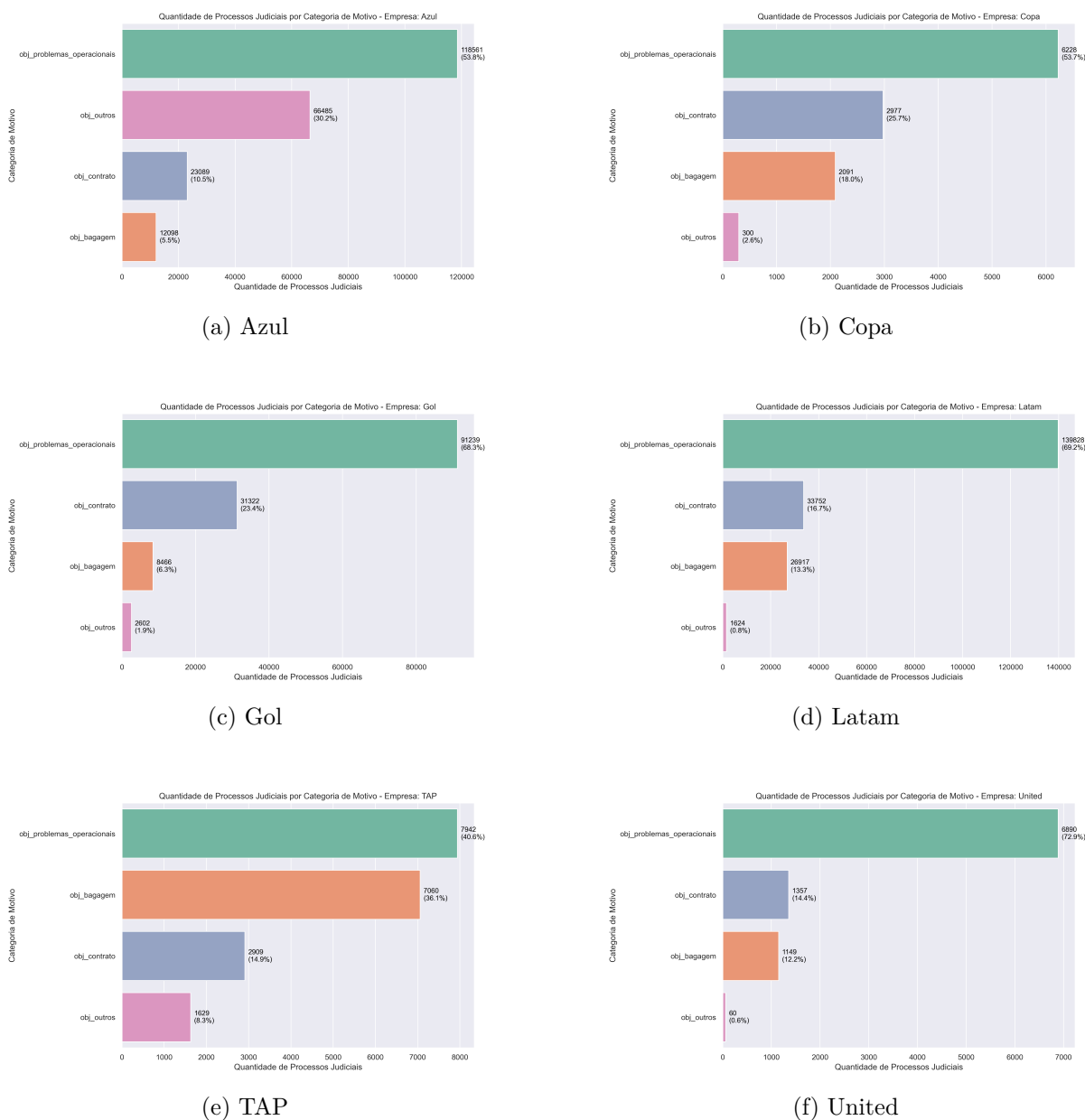
Figura 22 – Quantidade de Processos Judiciais por Categoria de Motivo e Região



Fonte: Elaborada pela autora.

Na Figura 22, observa-se que a ordem das categorias de motivos dos processos judiciais se manteve parecida em todas as regiões, com variações apenas na posição da categoria Outros, que se destacou mais em regiões como Norte e Centro-Oeste, devido à maior quantidade de dados faltantes. Em todas as regiões, a categoria problemas operacionais foi predominante, com percentuais variando entre 57,1% e 66,5%. A Figura 23 apresenta a quantidade de processos judiciais por categoria de motivo e por empresa aérea.

Figura 23 – Quantidade de Processos Judiciais por Categoria de Motivo e Empresa Aérea



Fonte: Elaborada pela autora.

Na Figura 23, observa-se que a ordem das categorias de motivos dos processos judiciais se manteve consistente para todas as empresas, com a sequência de problemas operacionais, contrato e bagagem, com exceção das empresas Azul e TAP. Na Azul, observa-se uma maior concentração na categoria Outros, que ocupou o segundo lugar com 30,2% do total. Esse valor indica que a maioria dos dados faltantes foram relativos aos dados fornecidos por essa empresa. Já na TAP, a categoria Bagagem ficou em segundo lugar, com 36,1% do total, uma proporção bem superior à das demais empresas, o que sugere que essa companhia possui um número consideravelmente maior de processos relacionados a problemas com bagagem.

Por fim, considerando que cada categoria de motivo inclui diversos objetos e subobjetos, a Tabela 21 apresenta o detalhamento dos cinco principais objetos e subobjetos dos processos judiciais por ano.

Tabela 21 – Top 5 dos Processos Judiciais por Objeto, Subobjeto e Ano

Ano	Categoria do Motivo	Objeto Principal	Subobjeto Principal	Qtd de Processos Judiciais	Percentual do Total do Ano
2019	Problemas operacionais	CANCELAMENTO DE VOO	***	13418	18,45%
	Problemas operacionais	Atraso de avião	***	9185	12,63%
	Problemas operacionais	ALTERAÇÃO DE VOO	***	4076	5,60%
	Problemas operacionais	ATRASSO	ATR. V. - TRÁFEGO AÉ-REO	3864	5,31%
	Bagagem	BAGAGEM	Extravio Temporário	3692	5,08%
2020	Problemas operacionais	CANCELAMENTO DE VOO	***	10653	19,17%
	Problemas operacionais	Atraso de avião	***	5431	9,77%
	Bagagem	BAGAGEM	Extravio Temporário	2221	4,00%
	Problemas operacionais	ALTERAÇÃO DE VOO	***	1943	3,50%
	Contrato	REEMBOLSO	Não Realizado	1862	3,35%
2021	Problemas operacionais	CANCELAMENTO DE VOO	***	8708	16,33%
	Problemas operacionais e Contrato	CANCELAMENTO	COVID-19	5537	10,38%
	Problemas operacionais	ALTERAÇÃO DE VOO	***	3840	7,20%
	Contrato	REEMBOLSO	Não Realizado	2984	5,59%
	Problemas operacionais	Atraso de avião	***	2679	5,02%
2022	Problemas operacionais	CANCELAMENTO DE VOO	***	13323	16,97%
	Problemas operacionais	Atraso de avião	***	6060	7,72%
	Contrato	REEMBOLSO	Não Realizado	4986	6,35%
	Problemas operacionais	ALTERAÇÃO DE VOO	***	4164	5,30%
	Bagagem	BAGAGEM	Extravio Temporário	3837	4,89%
2023	Problemas operacionais	CANCELAMENTO DE VOO	***	17294	19,87%
	Problemas operacionais	Atraso de voo	***	13862	15,93%
	Bagagem	BAGAGEM	Extravio Temporário	5423	6,23%
	Problemas operacionais	ATRASSO	ATR. V. - TRÁFEGO AÉ-REO	5278	6,07%
	Problemas operacionais	ALTERAÇÃO DE VOO	***	5263	6,05%

Fonte: Elaborado pela autora.

Na Tabela 21, observa-se que o principal motivo dos processos judiciais em todos os anos analisados foi o cancelamento de voo. Outros motivos recorrentes foram atraso de voo, alteração de voo e extravio temporário de bagagem, que apareceram em todos os anos, com exceção de 2021, quando o tema bagagem não figurou entre os cinco primeiros.

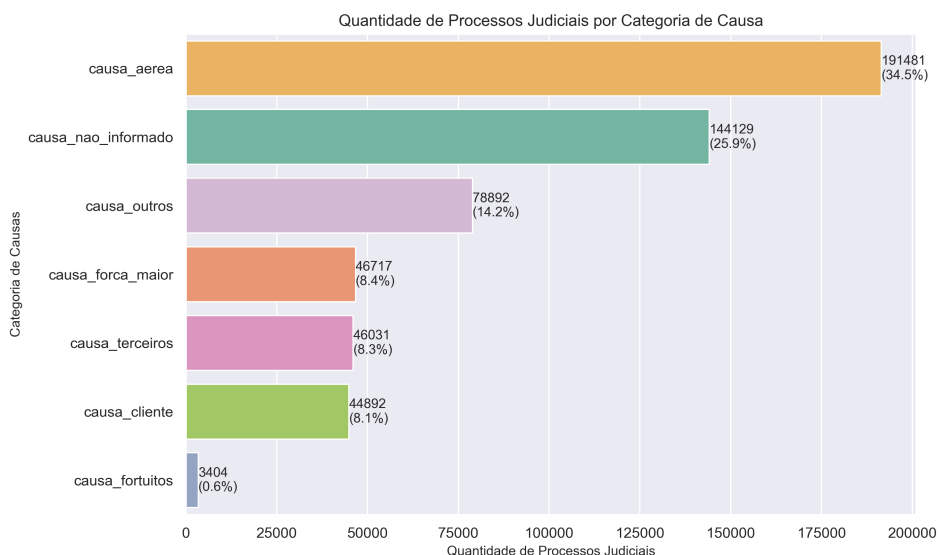
Um dado relevante refere-se aos processos relacionados a reembolso entre 2020 e

2022, período marcado pela pandemia de COVID-19. Em 2021, surgiu também um novo tema: o cancelamento de voo devido à COVID-19, com 5.537 processos (10,38% do total). Como apresentado anteriormente, as medidas emergenciais implementadas nesse período flexibilizaram as regras de alteração de passagens, cancelamentos, reembolsos e créditos para situações ocorridas até o final de 2021, o que gerou reflexos até 2022. Dessa forma, conclui-se que esse contexto influenciou o aumento desses temas tanto nas reclamações quanto nos processos judiciais durante esse período, corroborando os resultados de Çallı e Çallı (2023) e Kim, Lim e Ha (2024).

4.1.2.5 Causas dos processos judiciais

As causas dos processos judiciais correspondem à causa alegada pela empresa ré, isto é, a defesa ou justificativa alegada sobre a reclamação feita contra ela. A Figura 24 apresenta a quantidade de processos judiciais e o percentual relativo ao total por categoria de motivo e a Tabela 22 apresenta a quantidade de processos judiciais por categoria de motivo e ano. Ressalta-se que, como um processo pode ser enquadrado em mais de uma categoria, a contagem do total de processos por categoria pode superar o número de processos registrados.

Figura 24 – Quantidade de Processos Judiciais por Categoria de Causa



Fonte: Elaborada pela autora.

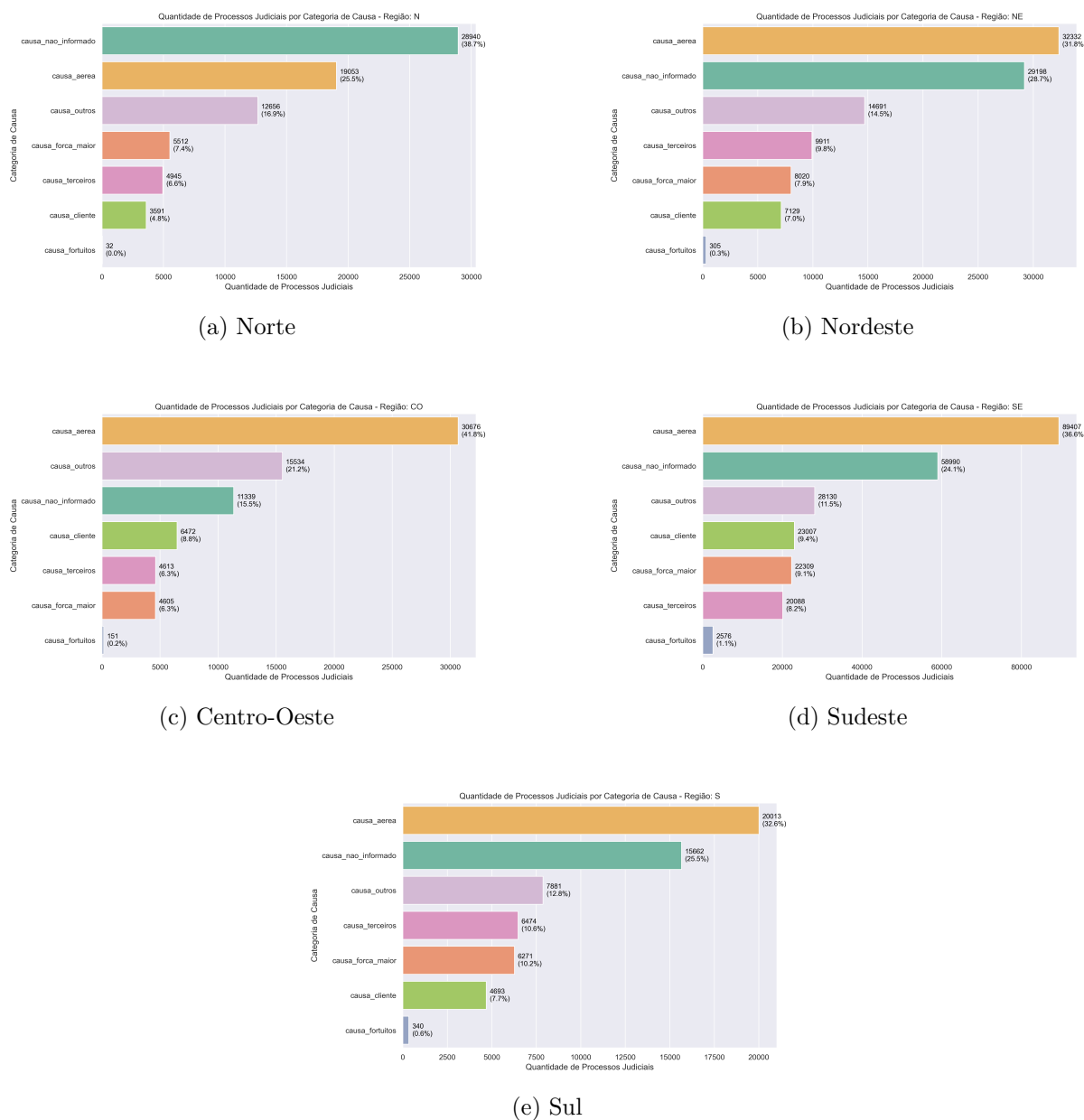
Tabela 22 – Quantidade de Processos Judiciais por Categoria de Causa e Ano

Categoria	2019	2020	2021	2022	2023	Total	Percentual do Total
causa_aerea	41819	26508	24833	47254	51067	191481	34,46%
causa_nao_informado	33989	29600	29216	46224	5100	144129	25,94%
causa_outros	2420	1501	865	2289	71817	78892	14,20%
causa_forca_maior	5424	8958	11870	11029	9436	46717	8,40%
causa_terceiros	10628	5917	4539	7935	17012	46031	8,28%
causa_cliente	9005	7124	7741	10953	10069	44892	8,08%
causa_fortuitos	1352	657	252	1143	0	3404	0,61%
Total	104637	80265	79316	126827	164501	555546	100%

Fonte: Elaborado pela autora.

Na Figura 24 e na Tabela 22, observa-se que a maior parte dos processos judiciais registrados entre 2019 e 2023 teve como causa a própria empresa aérea, representando 34,46% do total. Entre os problemas incluídos nessa categoria estão questões relacionadas a abastecimento, aeronave, tripulação, conexão, bagagem, cancelamento indevido, entre outros. Em segundo lugar, estão os processos com causa não informada, com 25,94% do total, devido ao preenchimento incompleto desse campo pelas empresas. A categoria Outros ocupa o terceiro lugar (14,20%), em grande parte devido a uma quantidade significativa de registros com dados faltantes em 2023. Em seguida, as causas por força maior, terceiros e clientes aparecem em sequência, variando de 8,40% a 8,08%. Por fim, a causa por fortuitos ocupa a última posição, com apenas 0,61%. A Figura 25 apresenta a quantidade de processos judiciais por categoria de causa e região.

Figura 25 – Quantidade de Processos Judiciais por Categoria de Causa e Região

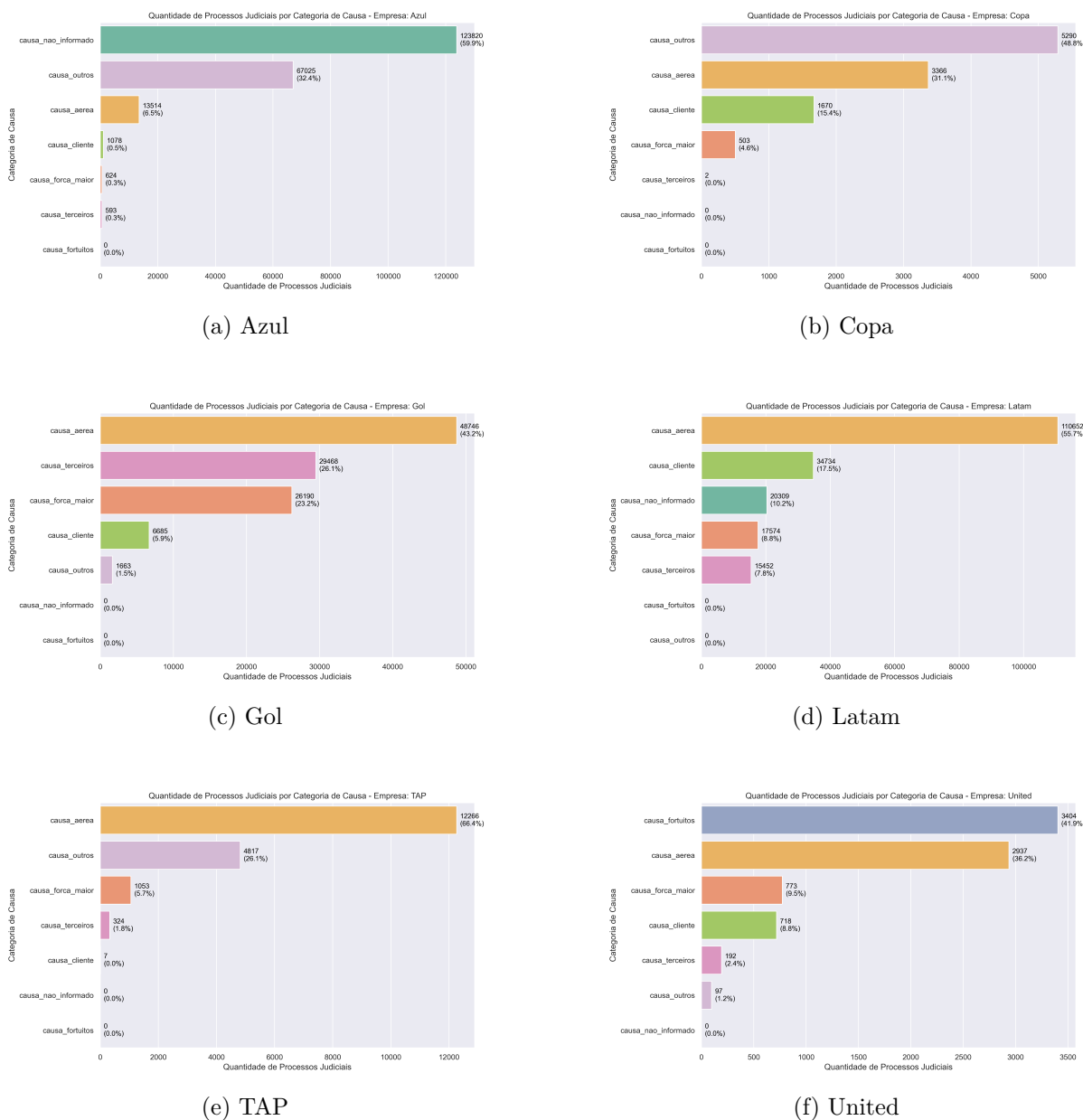


Fonte: Elaborada pela autora.

Na Figura 25, observa-se que, em todas as regiões, a causa aérea foi a principal, com exceção do Norte, onde a categoria não informado se destacou devido à maior quantidade de dados faltantes. As categorias não informado e outros também mantiveram-se significativas nas demais regiões, ocupando, geralmente, o segundo e terceiro lugares, o que reflete a incompletude das informações fornecidas pelas empresas. As causas relacionadas a força maior, terceiros, clientes e fortuitos apareceram em quantidades menores, com a posição dessas categorias variando entre as regiões. A segunda maior categoria, considerando os processos em que a causa é informada, foi cliente no Centro-Oeste e no Sudeste, terceiros no Nordeste e no Sul, e força maior no Norte. A Figura 26 apresenta a quantidade de

processos judiciais por categoria de motivo e por empresa aérea.

Figura 26 – Quantidade de Processos Judiciais por Categoria de Causa e Empresa Aérea



Fonte: Elaborada pela autora.

Analisando os dados apresentados na Figura 26, é possível observar diferenças entre as companhias. Na Azul, por exemplo, a principal categoria registrada foi a causa não informado (59,9%), seguida pela outros (32,4%) e aérea (6,5%). Já a Gol apresenta uma distribuição diferente, com a causa aérea liderando (43,2%), seguida por terceiros (26,1%) e força maior (23,2%). Para a Latam, a maior causa é também aérea (55,7%), seguida por cliente (17,5%) e não informada (10,2%). A Copa também apresentou uma alta quantidade de ocorrências atribuídas à categoria outros (48,8%), seguido pela causa aérea (31,1%) e cliente (15,4%). A TAP apresentou a causa aérea como a mais frequente

(66,4%), seguida por outros (26,1%) e força maior (5,7%). A United, por sua vez, teve como principal causa fortuitos (41,9%), seguida de aérea (36,2%) e força maior (9,5%). De modo geral, a principal causa mais comum entre as empresas é a própria aérea, que representa a maior parte das ocorrências, especialmente para Gol, Latam e TAP.

Por fim, considerando que cada categoria de causa incluía diversas descrições, a Tabela 23 apresenta o detalhamento das cinco principais causas alegadas dos processos judiciais por ano. Ressalta-se que, para a classificação das categorias foram analisadas a causa alegada em conjunto com o objeto e subobjeto principal, devido a presença de informações incompletas.

Tabela 23 – Top 5 dos Processos Judiciais por Causa e Ano

Ano	Categoria da Causa	Objeto Principal	Subobjeto Principal	Causa Alegada	Qtd de Processos Judiciais	Percentual do Total do Ano
2019	Aérea	CANCELAMENTO DE VOO	***	Culpa Tam	8628	13,83%
	Aérea	Atraso de avião	***	Culpa Tam	7311	11,72%
	Terceiros	ATRASSO	ATR. V. - TRÁFEGO AÉREO	-	3864	6,19%
	Aérea	ALTERAÇÃO DE VOO	***	Culpa Tam	2868	4,60%
2020	Terceiros	CANCELAMENTO	CANC. V. - TRÁFEGO AÉREO	-	2784	4,46%
	Aérea	CANCELAMENTO DE VOO	***	Culpa Tam	5326	11,05%
	Aérea	Atraso de avião	***	Culpa Tam	4030	8,36%
	Força maior	CANCELAMENTO DE VOO	***	Caso fortuito/força Maior	2712	5,63%
2021	Força maior	CANCELAMENTO	COVID-19	-	1603	3,33%
	Aérea	BAGAGEM	Extravio Temporário	Culpa Tam	1589	3,30%
	Força maior	CANCELAMENTO	COVID-19	-	5537	11,16%
	Aérea	CANCELAMENTO DE VOO	***	Culpa Tam	4810	9,69%
	Aérea	ALTERAÇÃO DE VOO	***	Culpa Tam	2906	5,86%
	Aérea	CANCELAMENTO	CANC. V. - MALHA AÉREA	-	2565	5,17%
	Aérea	Atraso de avião	***	Culpa Tam	2088	4,21%
2022	Aérea	CANCELAMENTO DE VOO	***	Culpa Tam	8729	12,35%
	Aérea	Atraso de avião	***	Culpa Tam	4877	6,90%
	Aérea	ALTERAÇÃO DE VOO	***	Culpa Tam	3165	4,48%
	Força maior	CANCELAMENTO	COVID-19	-	3034	4,29%
	Aérea	BAGAGEM	Extravio Temporário	Culpa Tam	2976	4,21%
	2023	Aérea	Atraso de voo	***	Culpa Tam	10680
Aérea		CANCELAMENTO DE VOO	***	Culpa Tam	9963	12,07%
Terceiros		ATRASSO	ATR. V. - TRÁFEGO AÉREO	-	5278	6,40%
Aérea		ALTERAÇÃO DE VOO	***	Culpa Tam	4018	4,87%
Força maior		CANCELAMENTO DE VOO	***	Caso fortuito/força Maior	3849	4,66%

Fonte: Elaborado pela autora.

Observa-se na Tabela 23 que, ao longo dos cinco anos analisados, as causas alegadas variaram entre 'Culpa Tam', 'Caso fortuito/força Maior' e '-'. A principal causa, em todos os anos, foi a responsabilidade da própria empresa aérea, associada a cancelamentos ou atrasos de voo, com exceção de 2021, quando o cancelamento foi atribuído à pandemia de

COVID-19. As demais categorias observadas foram força maior e terceiros. No caso de força maior, alguns processos estavam relacionados à pandemia de COVID-19 em 2020 e 2021, enquanto outros foram classificados como tal pela própria empresa. Já os casos de terceiros referem-se a situações em que a culpa pelo atraso foi atribuída ao tráfego aéreo.

4.1.2.6 Indenizações dos processos judiciais

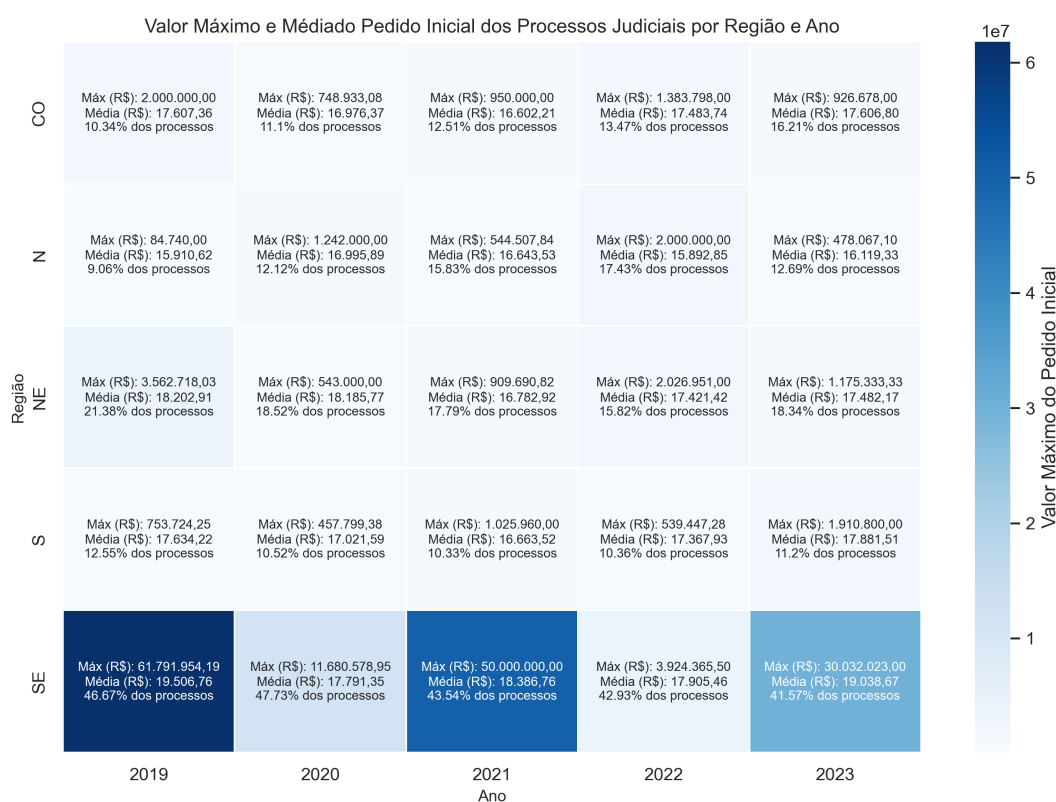
Em relação às indenizações dos processos judiciais, existem algumas variáveis a serem analisadas:

- **valor_pedido_inicial:** valor do pedido inicial (R\$) de indenizações do processo;
- **valor_dano_moral:** valor do dano moral (R\$) pedido no processo, isto é, prejuízos não materiais, como sofrimento emocional ou psicológico;
- **valor_dano_material:** valor do dano material (R\$) pedido no processo, isto é, perdas financeiras diretas;
- **valor_total_condenacao:** valor total da condenação (R\$) do processo, isto é, o valor total que a parte ré deve pagar ao(s) autor(es) ao final do processo, somando todas as indenizações;
- **valor_pago:** valor de indenização efetivamente pago (R\$) ao(s) autor(es) do processo.

Tendo em vista que os dados em branco foram substituídos por zero em todas as colunas relacionadas à indenização, o valor mínimo para todas as variáveis ficou igual a zero na Tabela 16, que apresenta as estatísticas descritivas das variáveis numéricas da base de judicialização. No entanto, ao calcular os valores mínimos posteriormente, desconsiderando todas as linhas com valores zero, o valor mínimo registrado foi de R\$ 0,01 para todas as variáveis. Com base nisso, nas análises a seguir, foram apresentadas a média, o desvio-padrão e o valor máximo de cada variável de indenização. Ressalta-se, ainda, que não foram apresentados os valores por empresa aérea, a fim de preservar a confidencialidade das informações.

A respeito do valor do pedido inicial dos processos, na Tabela 16, observa-se que a média geral foi de R\$ 17.809,65, com desvio padrão de R\$ 118.811,83. O valor máximo registrado foi de R\$ 61.791.950,00. A Figura 27 apresenta o valor máximo e a média do pedido inicial, além do percentual de processos judiciais por região e ano.

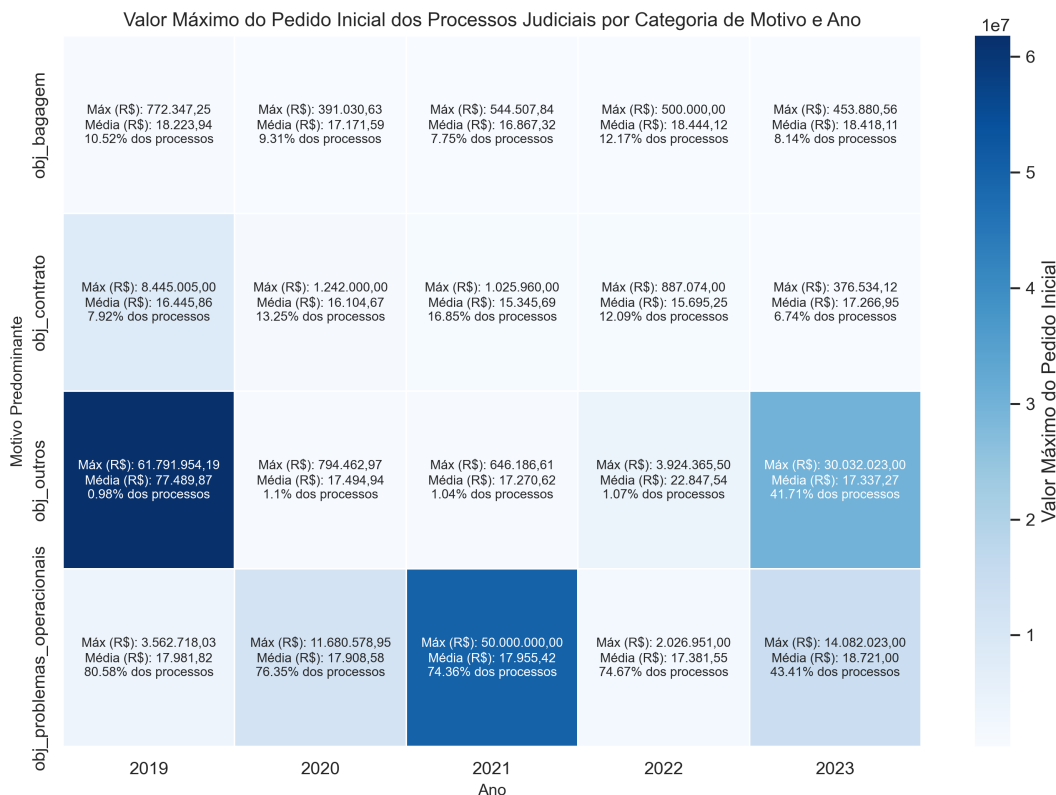
Figura 27 – Valor Máximo do Pedido Inicial dos Processos Judiciais por Região e Ano



Fonte: Elaborada pela autora.

Na Figura 27, percebe-se que os processos com maiores valores de pedido inicial ocorreram no Sudeste, alcançando R\$ 61.791.954,19 em 2019, R\$ 50.000.000,00 em 2021 e R\$ 30.032.023,00 em 2023. O Sudeste também se destacou como a região com as médias mais altas ao longo dos anos, variando de R\$ 17 a R\$ 19 mil. As demais regiões apresentaram variações menos expressivas, com picos isolados. No Centro-Oeste e no Norte, os picos ocorreram em 2021, alcançando o máximo de R\$ 1.383.798,00 e R\$ 2.000.000,00, respectivamente. O Nordeste mostrou uma redução significativa nos valores máximos após 2019, quando alcançou R\$ 3.562.718,03. Por fim, no Sul, o pico ocorreu em 2023, com um máximo de R\$ 1.910.800,00. A Figura 28 detalha o valor máximo e a média do pedido inicial, além do percentual de processos judiciais por categoria de motivo e ano.

Figura 28 – Valor Máximo do Pedido Inicial dos Processos Judiciais por Categoria de Motivo e Ano

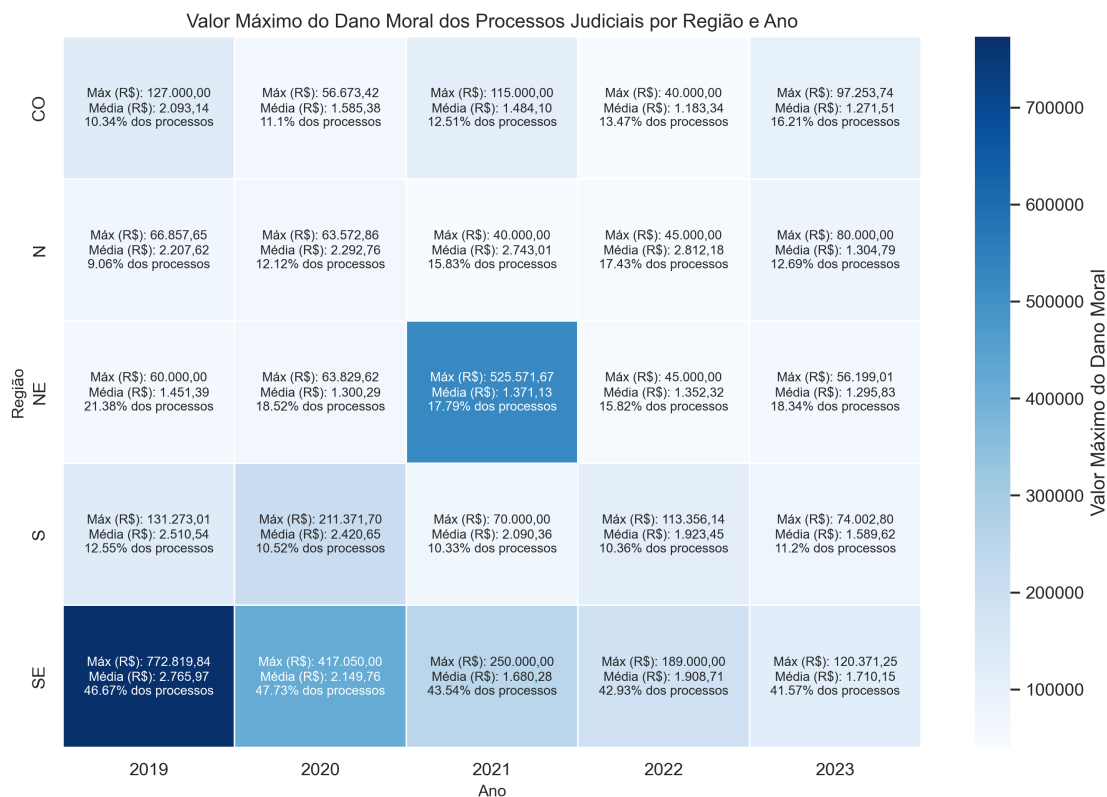


Fonte: Elaborada pela autora.

Conforme a Figura 28, em 2019 e 2023, os processos com os maiores valores de pedido inicial no Sudeste foram classificados na categoria outros, devido à ausência de informações específicas sobre o motivo. Já o processo de 2021 foi relacionado a problemas operacionais. No que se refere às médias, a categoria bagagem apresentou os valores mais altos ao longo dos anos, variando entre R\$ 16 e R\$ 18 mil, apesar de registrar o menor percentual de processos. A categoria outros também apresentou médias elevadas, influenciadas pelos altos valores dos poucos processos enquadrados nessa classificação. Em contrapartida, a categoria contrato apresentou as médias mais baixas, variando entre R\$ 15 e R\$ 17 mil.

Sobre o valor do dano moral nos processos, conforme a Tabela 16, a média geral foi de R\$ 1.856,66, com desvio padrão de R\$ 4.047,84. O valor máximo atingiu R\$ 772.819,84. A Figura 29 apresenta o valor máximo, a média do dano moral e o percentual de processos judiciais por região e ano.

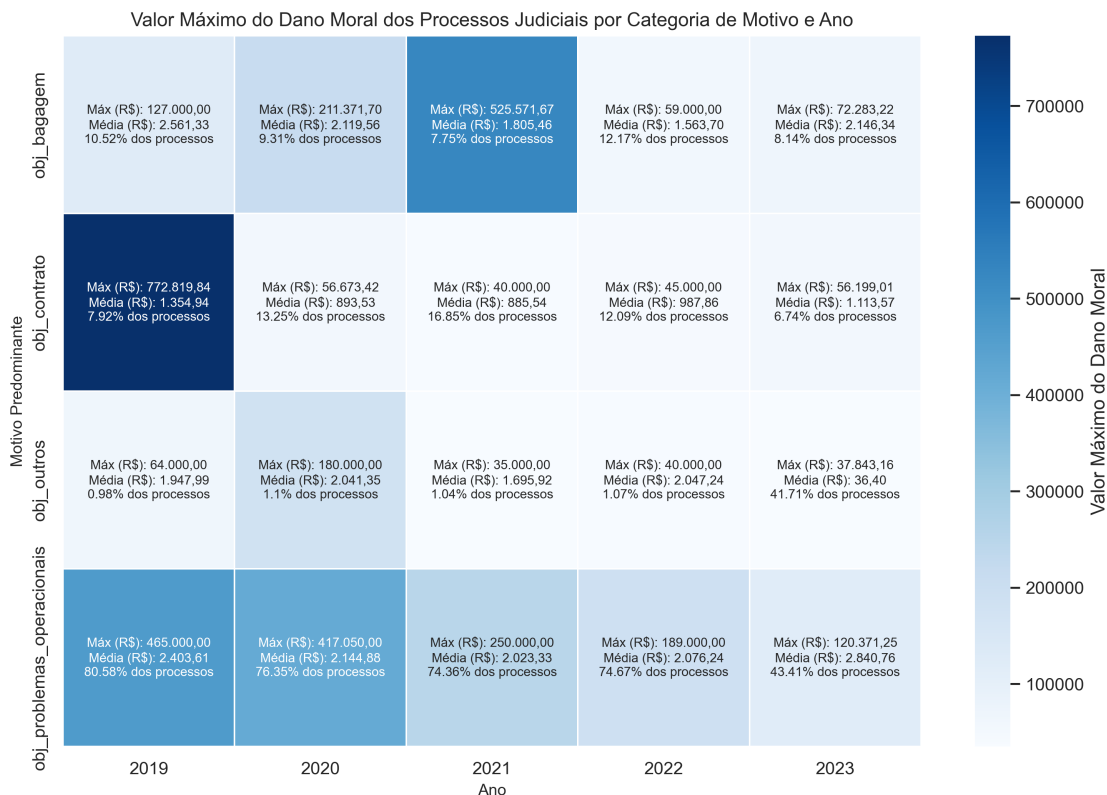
Figura 29 – Valor Máximo do Dano Moral dos Processos Judiciais por Região e Ano



Fonte: Elaborada pela autora. Nota: Esses valores máximos correspondem aos solicitados nos processos e não necessariamente aos decididos em condenação.

De acordo com a Figura 29, os processos com os maiores valores de dano moral ocorreram nas regiões Sudeste e Nordeste, alcançando R\$ 772.819,84 em 2019, R\$ 417.050,00 em 2020 e R\$ 525.571,67 em 2021. Quanto às médias, as regiões Norte e Sul destacaram-se com valores superiores a R\$ 2 mil entre 2019 e 2022. A Figura 30 apresenta o valor máximo, a média do dano moral e o percentual de processos judiciais por categoria de motivo e ano.

Figura 30 – Valor Máximo do Dano Moral dos Processos Judiciais por Categoria de Motivo e Ano

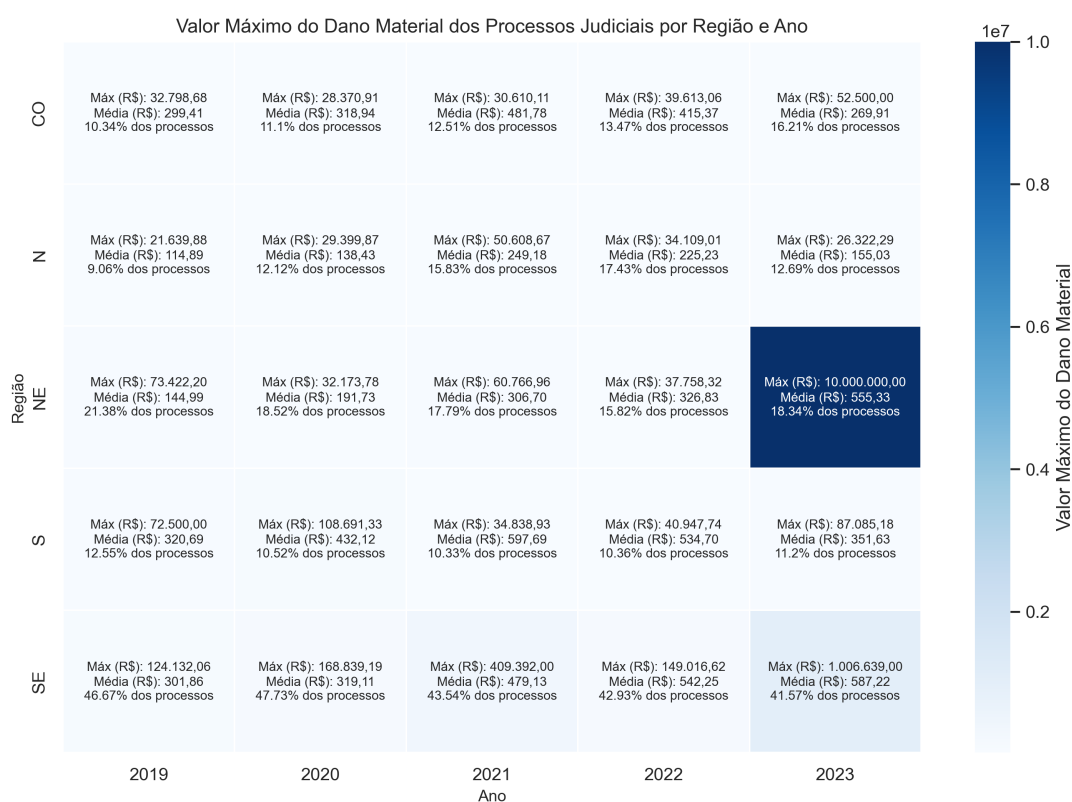


Fonte: Elaborada pela autora. Nota: Esses valores máximos correspondem aos solicitados nos processos e não necessariamente aos decididos em condenação.

Na Figura 30, observa-se que o processo com o maior valor de dano moral (R\$ 772.819,84), ocorrido em 2019, teve como motivo a categoria contrato. Outros processos relevantes destacaram-se por problemas operacionais (R\$ 465.000,00 em 2019 e R\$ 417.050,00 em 2020) e bagagem (R\$ 525.571,67 em 2021). Em relação às médias, a categoria problemas operacionais apresentou os valores mais altos, variando entre R\$ 2.023,33 e R\$ 2.840,76. Já a categoria contrato, apesar de conter o maior valor de dano moral, teve as médias mais baixas, variando entre R\$ 885,54 e R\$ 1.113,57.

Em relação ao dano material nos processos, conforme os dados da Tabela 16, a média foi de R\$381,81, com desvio padrão de R\$13.627,44. O valor máximo registrado foi de R\$10.000.000,00. A Figura 31 apresenta o valor máximo, a média do dano material e o percentual de processos judiciais por região e ano.

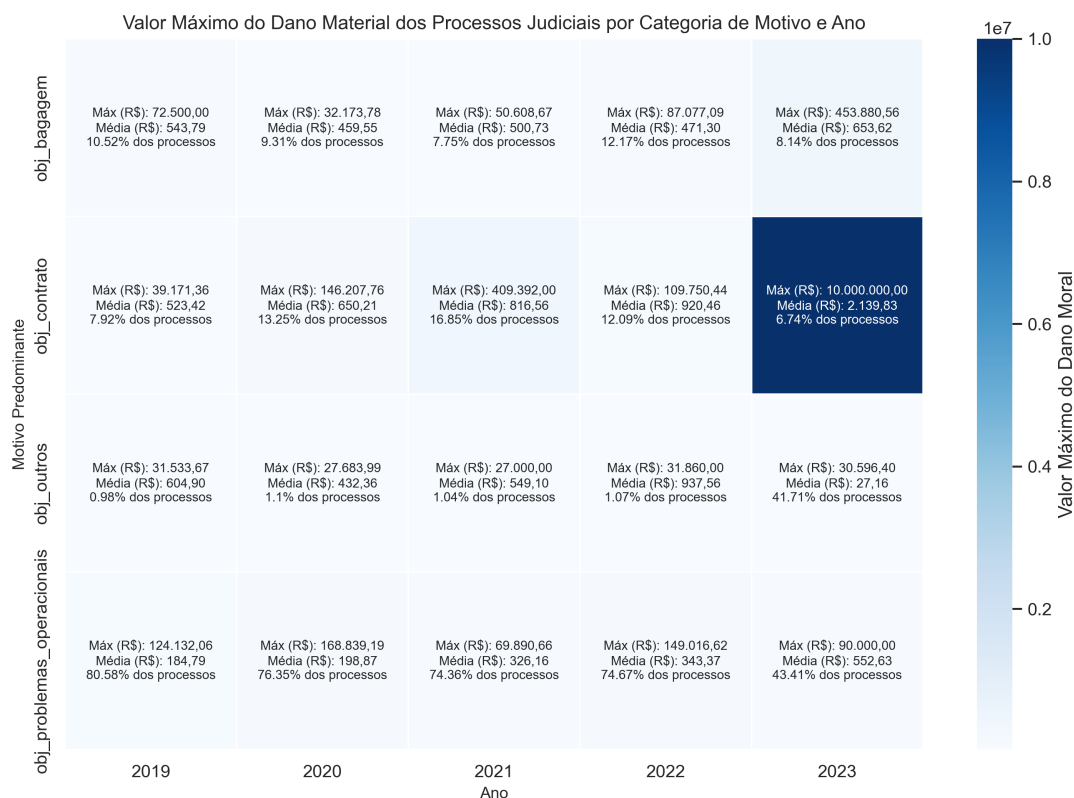
Figura 31 – Valor Máximo do Dano Material dos Processos Judiciais por Região e Ano



Fonte: Elaborada pela autora. Nota: Esses valores máximos correspondem aos solicitados nos processos e não necessariamente aos decididos em condenação.

Na Figura 31, observa-se que o processo com maior valor de dano material (R\$ 10.000.000,00) ocorreu na região Nordeste em 2023. O segundo maior valor foi de R\$ 1.006.639,00 e ocorreu no Sudeste, também em 2023. Embora esses valores sejam pontualmente elevados, a análise geral evidencia que os valores de danos materiais tendem a ser menores que os de dano moral. Em relação às médias, o Sul e o Sudeste destacaram-se como as regiões com os maiores valores médios ao longo dos anos. Por outro lado, o Norte foi a região com as menores médias registradas, variando de R\$ 114,89 a R\$ 155,03. A Figura 32 apresenta o valor máximo, a média do dano material e o percentual de processos judiciais por categoria de motivo e ano.

Figura 32 – Valor Máximo do Dano Material dos Processos Judiciais por Categoria de Motivo e Ano



Fonte: Elaborada pela autora. Nota: Esses valores máximos correspondem aos solicitados nos processos e não necessariamente aos decididos em condenação.

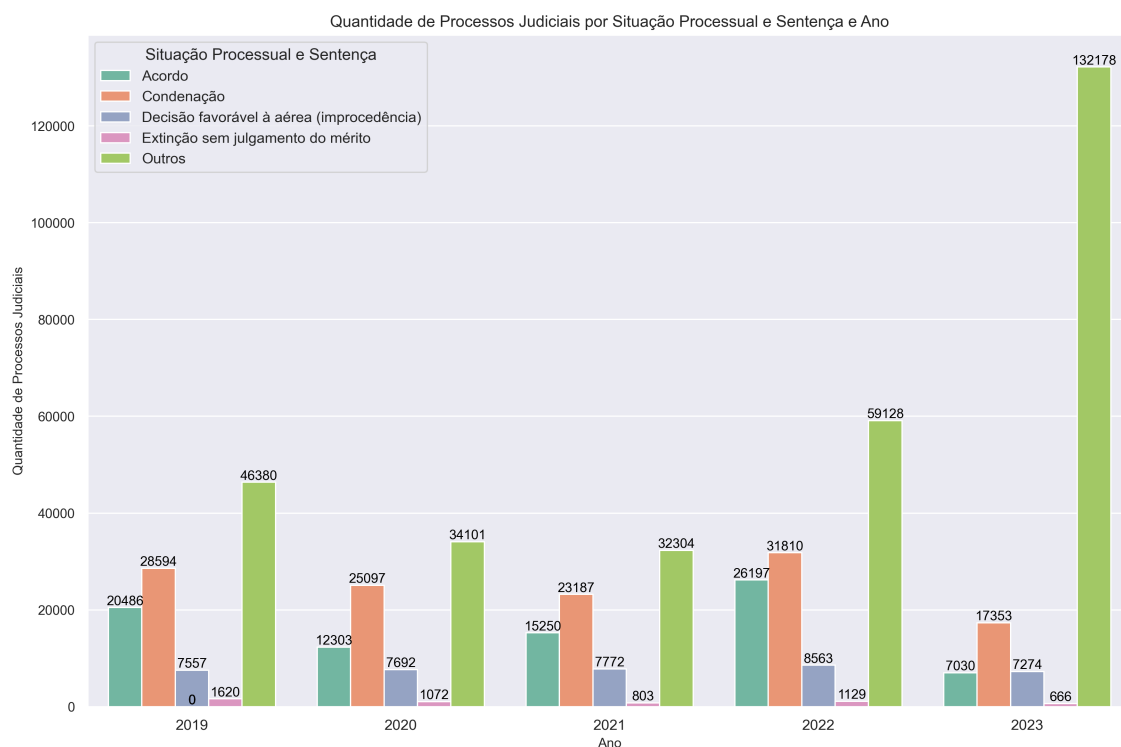
Na Figura 32, observa-se que o processo com o maior valor de dano material (R\$ 10.000.000,00), registrado em 2023, teve como motivo predominante a categoria contrato. Outros processos relevantes destacaram-se na categoria problemas operacionais, com valores máximos de R\$ 453.880,56 em 2023 e R\$ 168.839,19 em 2020. A categoria bagagem também apresentou valores expressivos, com um máximo de R\$ 87.077,09 em 2022. Em relação às médias, a categoria problemas operacionais apresentou os valores mais altos ao longo dos anos, variando entre R\$ 184,79 em 2019 e R\$ 552,63 em 2023. Já a categoria contrato, embora tenha registrado o maior valor isolado de dano material, apresentou médias mais moderadas, variando entre R\$ 523,42 e R\$ 2.139,83. Por outro lado, a categoria outros apresentou as médias mais baixas, com um valor de apenas R\$ 27,16 em 2023, provavelmente devido a falta de informações nos processos incluídos nessa categoria.

É interessante observar que os valores apurados do dano moral e do dano material, não tem qualquer vinculação necessariamente com valor do serviço prestado pelas companhias aéreas. Conforme o (ANAC, 2024b), o valor da tarifa média de um bilhete para

voos domésticos foi de R\$638,87 em 2023, sendo que a média apurada, em processos em que a situação processual e sentença foi condenação, a média apurada de danos morais e materiais para voos nacionais, no mesmo ano, foi de R\$4.229,88 e R\$924,20, respectivamente. Percebe-se que o valor do dano moral, principalmente, trata-se de um valor seis vezes superior (562,08%) ao valor médio do serviço prestado. Já o valor do dano material, que é relacionado a perdas financeiras diretas, é 44,66% maior que o valor médio da tarifa.

Antes de analisar os valores de total da condenação e pago, é importante destacar que, para que esses valores sejam aplicáveis, o processo precisa ter sido finalizado com decisão favorável ao(s) autor(es). Ou seja, o juiz ou tribunal responsável pelo caso deve ter decidido que a parte ré (empresa aérea) é responsável por algum tipo de obrigação ou reparação em favor da parte autora (consumidor). Dessa forma, podem ser consideradas apenas determinadas situações processuais e sentenças, como: condenação, acordo e outros, que englobam processos nos quais esses dados não foram informados ou que ainda não possuem decisão definitiva. A Figura 33 ilustra a quantidade de processos judiciais por situações processuais e sentença ao longo dos anos, evidenciando a evolução dos processos com decisões favoráveis e aquelas que ainda se encontram pendentes ou sem resolução final.

Figura 33 – Quantidade de Processos Judiciais por Situação Processual e Sentença e Ano



Fonte: Elaborada pela autora.

Na Figura 33 observa-se que a maioria dos processos judiciais está concentrada na categoria 'Outros', que inclui os casos em que esse dado não foi informado ou que

ainda não apresentam uma decisão definitiva. A segunda maior categoria é a 'Condenação', que, embora tenha apresentado uma queda de 45,45% em 2023, ainda representa uma parte significativa dos processos, refletindo os casos em que a decisão judicial foi favorável ao autor. Em seguida, a categoria 'Acordo' também se destaca, com um crescimento expressivo em 2022 (71,78%), mas uma queda significativa em 2023 (-73,16%), o que pode indicar mudanças nos padrões de resolução dos casos. Por outro lado, a categoria 'Decisão favorável à aérea (improcedência)' se manteve constante ao longo dos anos, indicando a continuidade de processos que resultam em favor das empresas aéreas. A categoria 'Extinção sem julgamento do mérito' apresentou um comportamento mais volátil, com um aumento em 2022 (40,59%), mas uma queda em 2023 (-41%). Essas variações refletem a complexidade e a diversidade nas resoluções dos processos judiciais ao longo dos anos.

Com base nisso, destaca-se que, embora as empresas tenham informado valores de total da condenação e pago para processos classificados como 'Decisão favorável à aérea (improcedência)' ou 'Extinção sem julgamento do mérito', esses valores foram excluídos das análises a seguir. Isso ocorre porque tais situações não comportam a existência de valores de condenação⁴. A presença desses dados sugere um possível erro no registro de informações pelas empresas aéreas. Dessa forma, as estatísticas descritivas das variáveis `valor_total_condenacao` e `valor_pago` foram recalculadas desconsiderando os processos classificados como 'Decisão favorável à aérea (improcedência)' ou 'Extinção sem julgamento do mérito'. Os resultados obtidos estão na Tabela 24.

Tabela 24 – Judicialização: Estatísticas Descritivas - Valores de Condenação e Pagamento

Estatísticas	valor_total_condenacao	valor_pago
count	511398	511398
mean	1625,98	2307,23
std	5163,23	12695,24
min	0,00	0,00
25%	0,00	0,00
50%	0,00	0,00
75%	1751,00	3228,05
max	1188888,00	7889160,00

Fonte: Elaborado pela autora.

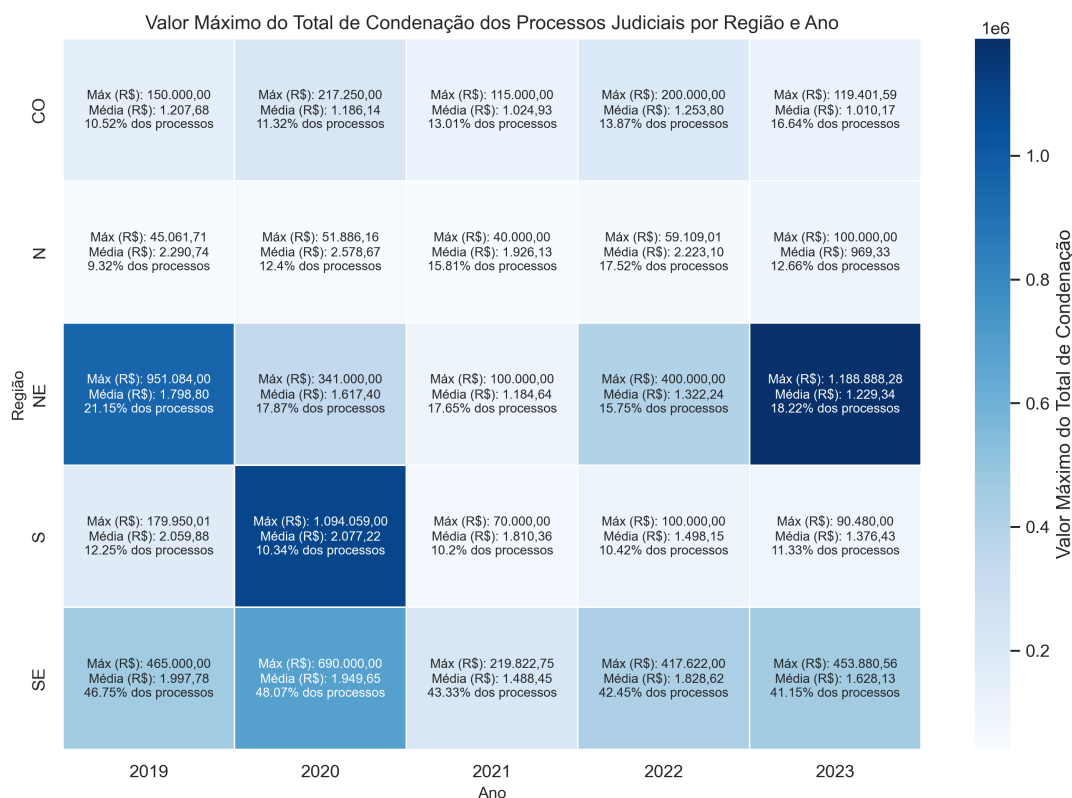
No que tange ao valor total da condenação, a Tabela 24 revela que a média foi de R\$ 1.625,98, com desvio padrão de R\$ 5.163,23⁵. O valor máximo foi de R\$1.188.888,00.

⁴ Na decisão favorável à aérea (improcedência), o juiz julga o mérito e decide que o autor está errado. Já na extinção sem julgamento do mérito, o juiz determina que não analisará o mérito da ação (ou seja, não julgará se o autor está certo ou errado) devido a alguma falha processual, como um problema constitucional ou formal que impede o prosseguimento da ação. Por esses motivos, processos nessas situações não deveriam apresentar valores de condenação.

⁵ Na Tabela 16, que considerou todos os processos, a média foi de R\$ 1.506,39, com desvio padrão de R\$ 4.980,28. Assim, as medidas tiveram um aumento de R\$ 116,59 e R\$ 182,95, respectivamente.

A Figura 34 apresenta o valor máximo, a média do total da condenação e o percentual de processos judiciais por região e ano.

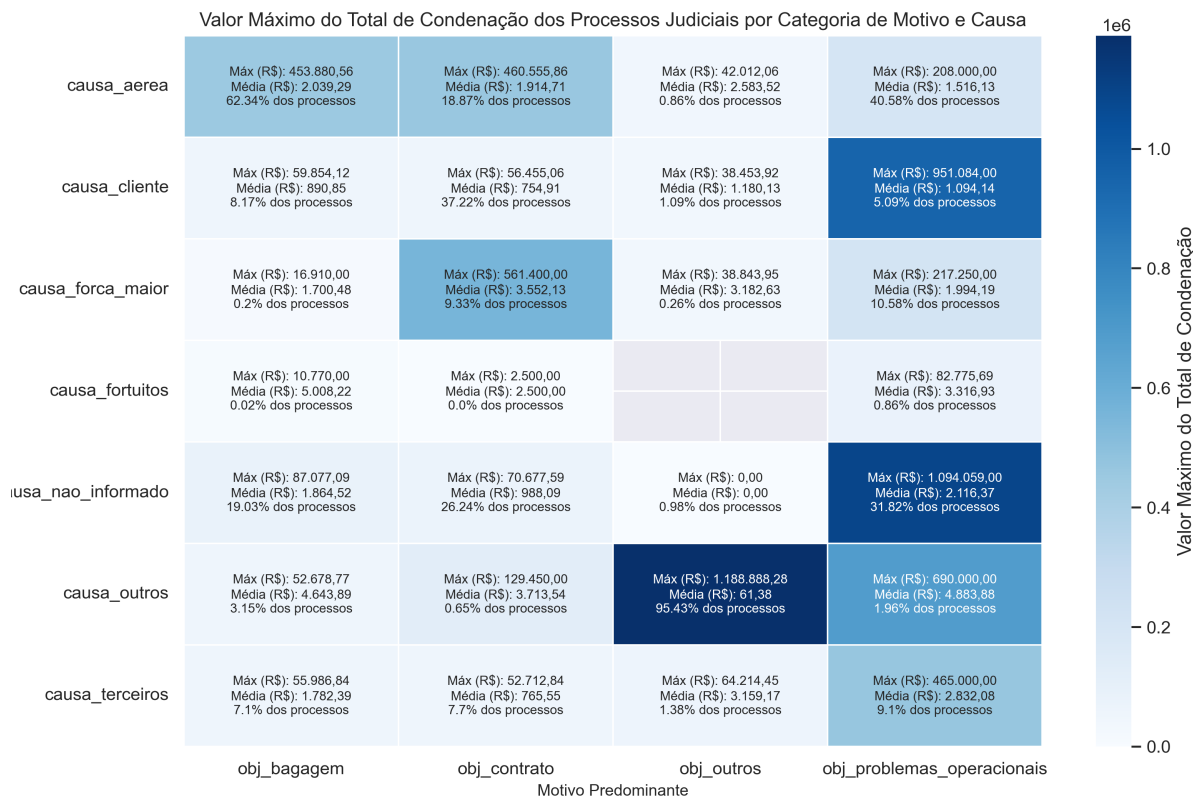
Figura 34 – Valor Máximo do Total de Condenação dos Processos Judiciais por Região e Ano



Fonte: Elaborada pela autora.

De acordo com a Figura 34, os processos com os maiores valores totais de condenação ocorreram no Nordeste (R\$ 951.084,00 em 2019 e R\$ 1.188.888,28 em 2023) e Sul (R\$ 1.094.059,00 em 2020). O Sudeste, apesar de não ter apresentado os valores máximos mais altos, foi a região com maior constância em valores altos, variando de R\$ 219.822,75 a R\$ 690.000,00. Em relação às médias, o Norte, apesar de apresentar valores máximos menores de condenação, alcançou médias acima de R\$ 2 mil nos anos de 2019, 2020 e 2022. O Sudeste destacou-se também nas médias, com valores consistentes e elevados, refletindo sua maior concentração de processos com condenações mais significativas. Por outro lado, o Centro-Oeste apresentou as menores médias entre todas as regiões, com valores entre R\$ 1.010,17 e R\$ 1.207,68. A Figura 35 apresenta o valor máximo, a média do total da condenação e o percentual de processos judiciais, organizados por motivo e causa alegada pela empresa, a fim de proporcionar o entendimento mais completo dos processos.

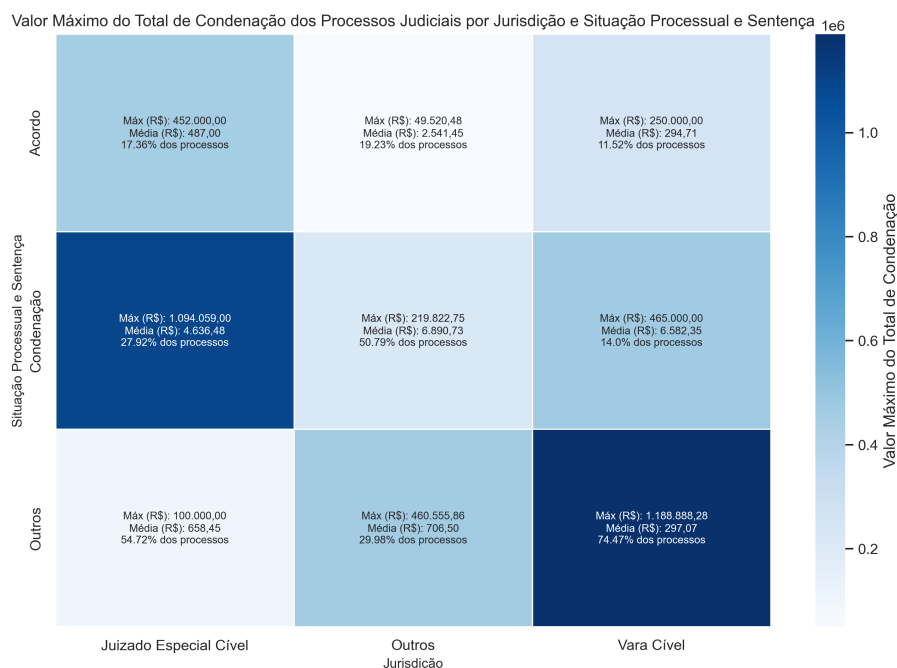
Figura 35 – Valor Máximo do Total de Condenação dos Processos Judiciais por Categoria de Motivo e Causa



Fonte: Elaborada pela autora.

Na Figura 35, observa-se que o processo com o maior valor total de condenação (R\$ 1.188.888,28) não tinha informações sobre o motivo e a causa, sendo classificado na categoria outros nessas duas características. Apesar disso, a média foi extremamente baixa (R\$ 61,38), indicando casos isolados de alto valor. Na categoria de motivo de problemas operacionais, as causas não informada, cliente, outros e terceiros apresentaram valores mais altos, sendo a causa não informada a que obteve o maior valor (R\$ 1.094.059,00), com uma média consistente de R\$ 2.116,37. Já na categoria de motivo bagagem, a causa aérea destacou-se com o maior valor máximo (R\$ 453.880,56) e a maior frequência de processos (62,34%), com média de R\$ 2.039,29. Em contrapartida, na categoria de motivo contrato, a causa força maior apresentou o maior valor máximo (R\$ 561.400,00) e média de de R\$ 3.552,13, indicando impacto financeiro elevado, embora menos frequente. A Figura 36 apresenta o valor máximo, a média do total da condenação e o percentual de processos judiciais, organizados por jurisdição e resultado processual e sentença.

Figura 36 – Valor Máximo do Total de Condenação dos Processos Judiciais por Jurisdição e Situação Processual e Sentença



Fonte: Elaborada pela autora.

Observa-se na Figura 36 que o processo com o valor mais alto de condenação (R\$ 1.188.888,28) ocorreu em uma Vara Cível, sendo classificado na categoria outros, por falta de informações detalhadas sobre a situação processual e sentença. No Juizado Especial Cível, destaca-se que o valor máximo de condenação foi de R\$ 1.094.059,00, enquanto o maior valor de um acordo foi de R\$ 452.000,00. Esses valores chamam a atenção, pois estão significativamente acima do limite estabelecido para os juizados especiais em 2023, que era de R\$52.800,00⁶. A discrepância revela que, apesar da competência dos Juizados Especiais Cíveis ser voltada para causas de menor complexidade e valores reduzidos, processos de maior impacto financeiro têm sido registrados, sendo necessário uma análise mais aprofundada desses processos em busca de explicações.

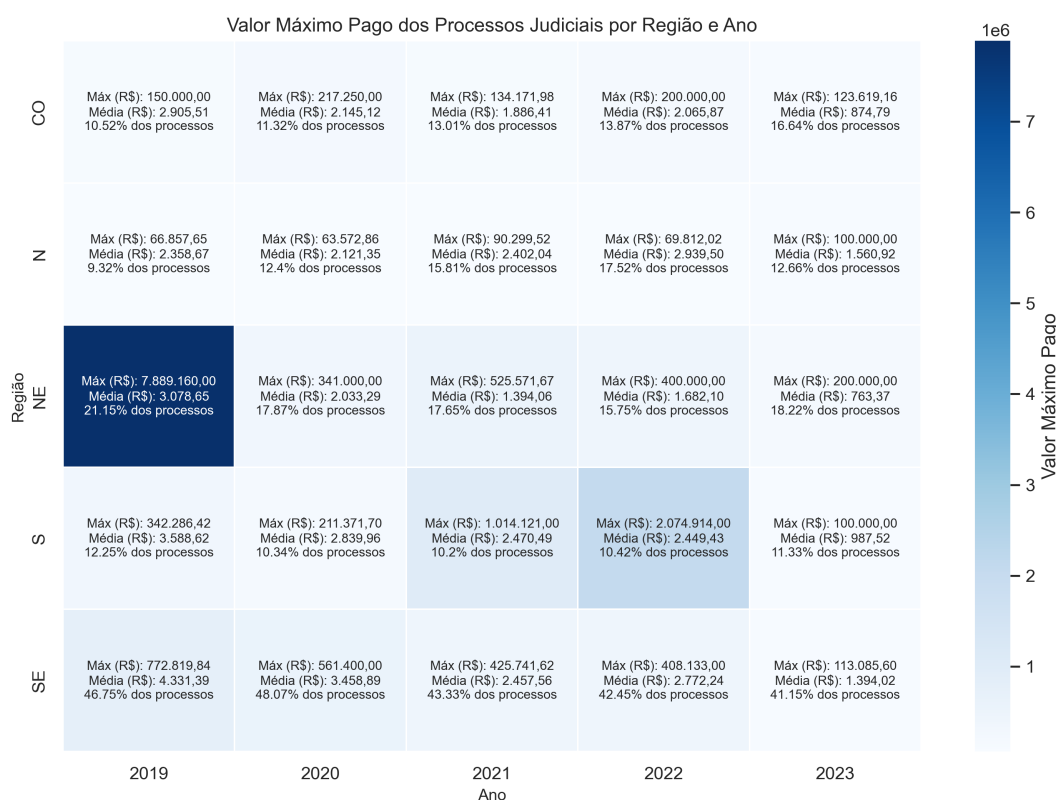
Além disso, na categoria de condenação, os processos nas Varas Cíveis apresentaram uma média de R\$6.582,35, enquanto no Juizado Especial Cível, a média foi de R\$4.636,48, indicando que os valores médios são menores nos juizados, mas ainda relevantes. Por fim, os acordos demonstram valores medianos entre as categorias analisadas, reforçando o papel conciliador dos Juizados Especiais e das Varas Cíveis, mas com uma frequência menor em relação aos processos de condenação.

Quanto ao valor pago nos processos, a média foi de R\$ 2.307,23, com desvio padrão

⁶ O Juizado Especial Cível têm competência para conciliar, processar e julgar as causas cíveis de menor complexidade, incluindo causas cujo valor não exceda quarenta vezes o salário mínimo. Considerando que o salário mínimo no Brasil em 2023 era de R\$ 1.320,00, multiplicado por quarenta o valor limite seria de R\$52.800,00.

de R\$ 12.695,24, conforme os dados da Tabela 24⁷. O valor máximo foi de R\$7.889.160,00. A Figura 37 apresenta o valor máximo, a média do valor pago e o percentual de processos judiciais, organizados por região e ano.

Figura 37 – Valor Máximo Pago dos Processos Judiciais por Região e Ano



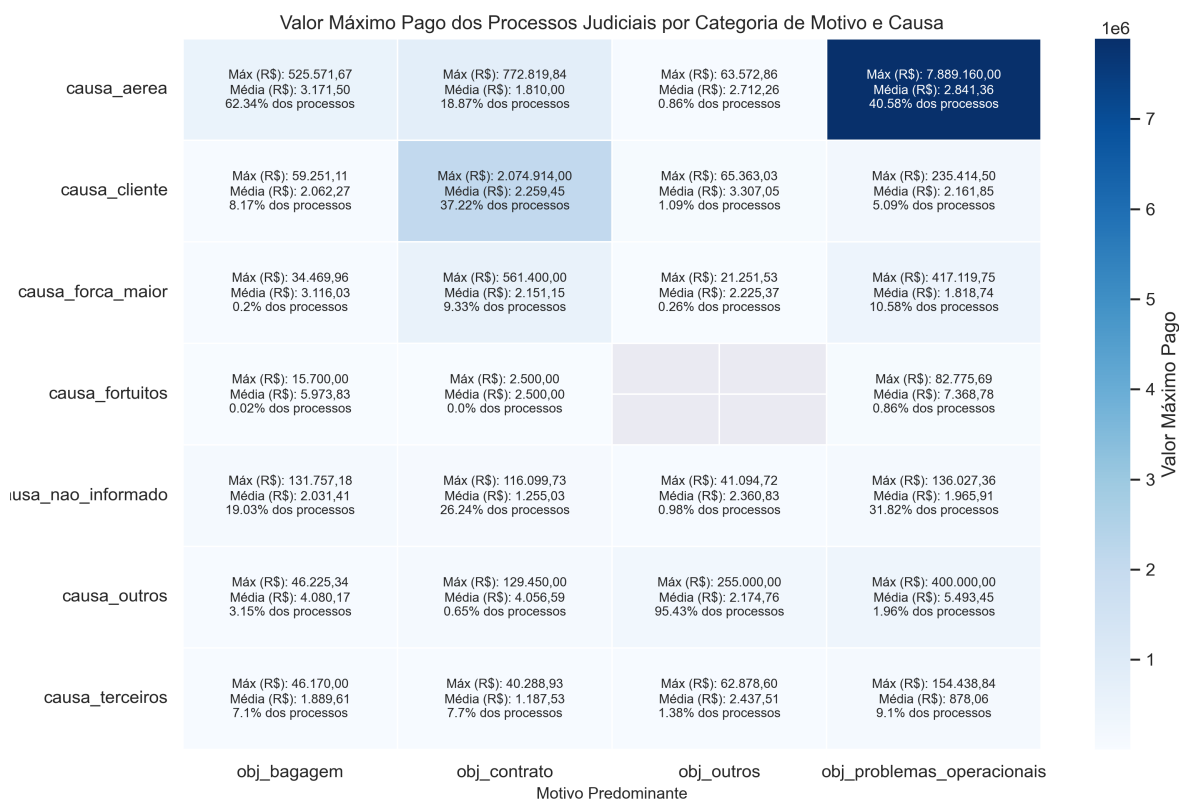
Fonte: Elaborada pela autora.

De acordo com a Figura 37, o processo com maior valor pago (R\$ 7.889.160,00) ocorreu na região Nordeste em 2019. Além desse, outros dois processos que ocorreram no Sul se destacaram, com os valores de R\$ 1.014.121,00 em 2021 e R\$ 2.074.914,00 em 2022. Em relação às médias, observa-se que a região Sudeste apresentou os valores mais altos na maior parte dos anos analisados, variando de R\$ 1.394,02 a R\$ 4.331,39, o que reflete uma maior concentração de processos com condenações significativas na região. Por outro lado, a região Nordeste, embora tenha registrado o maior valor pago em um único processo, apresentou médias relativamente mais baixas em outros anos, variando de R\$763,37 a R\$3.078,65. Essa oscilação sugere que os valores extremos na região não refletem uma tendência geral, mas sim casos pontuais de maior impacto financeiro. A

⁷ Na Tabela 16, que considerou todos os processos, a média foi de R\$ 2.127,99, com desvio padrão de R\$ 12.197,07. Assim, os medidas tiveram um aumento de R\$ 179,24 e R\$ 498,17, respectivamente.

Figura 38 apresenta o valor máximo, a média do valor pago e o percentual de processos judiciais, organizados por motivo e causa alegada pela empresa.

Figura 38 – Valor Máximo Pago dos Processos Judiciais por Categoria de Motivo e Causa



Fonte: Elaborada pela autora.

Na Figura 38, observa-se que o processo com maior valor pago (R\$ 7.889.160,00) foi motivado por problemas operacionais, sendo a causa atribuída à própria empresa aérea. Ainda dentro da categoria de problemas operacionais, destacam-se processos com valores pagos superiores a R\$ 400 mil, em que a causa foi atribuída a força maior (R\$ 417.119,75) e outros (R\$ 400.000,00), evidenciando que situações adversas ou ambíguas também podem gerar impactos financeiros expressivos para as companhias.

Na categoria de contrato, merecem destaque os valores máximos pagos de R\$ 2.074.914,00, atribuídos ao cliente, R\$ 772.819,84, com causa pela própria empresa aérea, e R\$ 561.400,00, relacionados a força maior. Esses resultados reforçam a relevância das disputas contratuais no contexto judicial, onde o descumprimento de obrigações ou divergências interpretativas podem acarretar valores substanciais.

No contexto de processos relacionados à bagagem, evidencia-se um valor máximo de R\$ 525.571,67, novamente com a causa atribuída à empresa aérea, acompanhado de uma

média de R\$ 3.171,50. Esse dado sugere que, embora os processos relacionados à bagagem possam parecer de menor complexidade, eles ainda têm o potencial de atingir valores significativos, provavelmente devido à recorrência ou à gravidade dos danos relatados.

De forma geral, a análise evidencia que os problemas operacionais e contratuais são os principais responsáveis pelos valores mais altos pagos em processos judiciais, especialmente quando a causa é atribuída à empresa aérea. As médias também refletem essa tendência, com valores que variam entre categorias e causas, mas apontam para uma concentração de maiores impactos financeiros em situações que envolvem diretamente as responsabilidades operacionais e contratuais das companhias aéreas. A Figura 39 apresenta o valor máximo, a média do valor pago e o percentual de processos judiciais, organizados por jurisdição e resultado processual e sentença.

Figura 39 – Valor Máximo Pago dos Processos Judiciais por Jurisdição e Situação Processual e Sentença



Fonte: Elaborada pela autora.

Na Figura 39, observa-se que o maior valor pago (R\$ 7.889.160,00) ocorreu em um processo de condenação no Juizado Especial Cível. Destaca-se que foi um valor bem acima do teto para os juizados especiais. Ainda no Juizado, os valores máximos de acordo e outros também foram expressivos, sendo superiores a R\$400 mil. Esse padrão reflete que, embora os valores médios sejam menores em comparação à Vara Cível, o impacto financeiro ainda é considerável.

Na Vara Cível, o valor mais alto (R\$ 772.819,84) não tinha informações a respeito da situação processual e sentença, sendo classificado na categoria outros. Nos casos em que

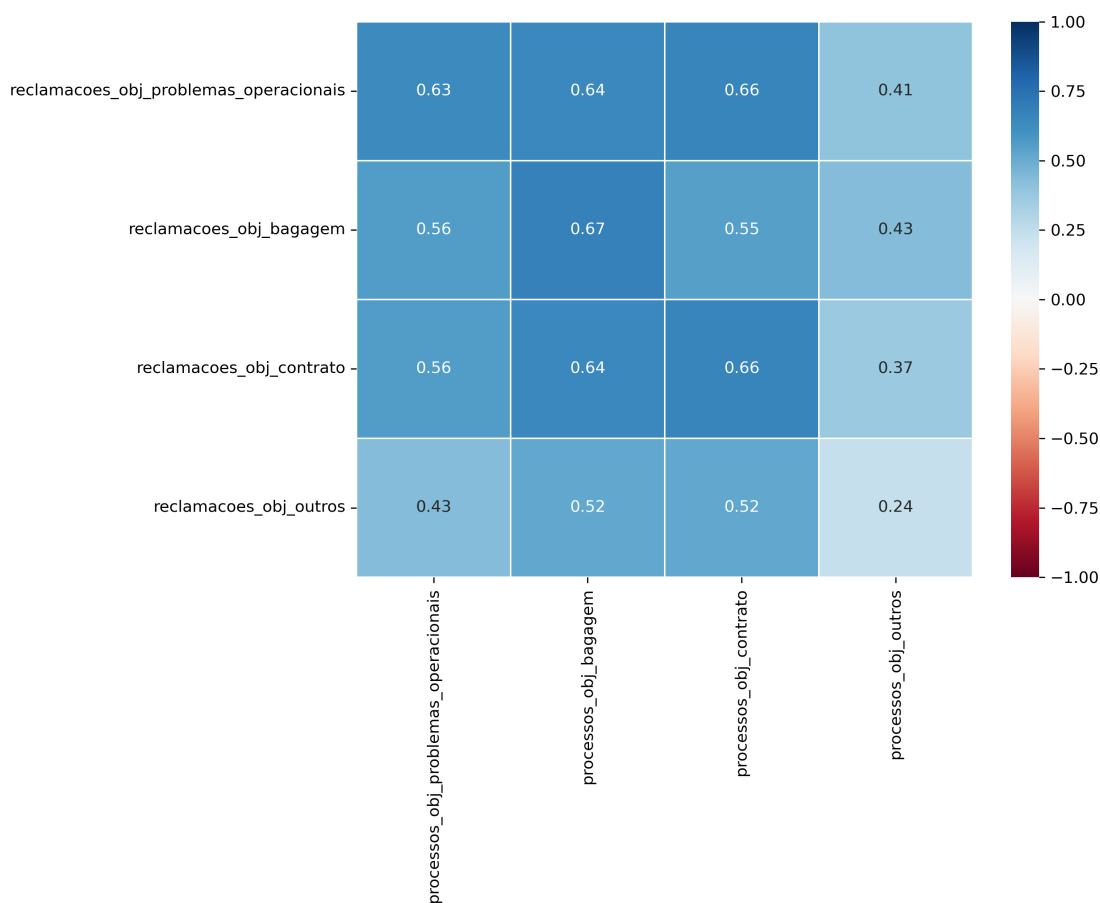
a jurisdição foi classificada como outros, também tiveram valores significativos na categoria de situação processual e sentença outros (R\$ 596.503,17) e condenação (R\$ 425.741,62).

Dessa forma, a análise das indenizações dos processos judiciais revela a complexidade e diversidade dos litígios enfrentados pelas companhias aéreas, tanto em termos de valores pleiteados quanto de decisões judiciais. Os dados evidenciam diferenças regionais e por categoria de motivo, além de destacar valores máximos pontualmente elevados que, em muitos casos, extrapolam os padrões médios do setor. Apesar de os valores pagos geralmente serem inferiores aos solicitados, as condenações e acordos refletem a importância de compreender e gerenciar adequadamente os fatores que levam às judicializações, visando reduzir custos e melhorar a experiência dos consumidores.

4.2 Correlação entre as reclamações e os processos judiciais contra companhias aéreas

Essa seção apresenta os resultados das correlações entre os motivos de reclamação e processos judiciais. A Figura 40 apresenta um gráfico de calor das correlações identificadas.

Figura 40 – Matriz de Correlação entre as Quantidades dos Motivos de Reclamações e Processos



Fonte: Elaborada pela autora.

Observa-se na matriz de correlação apresentada que as correlações mais altas foram entre as reclamações e os processos judiciais relacionados a bagagem (0,67) e a contrato (0,66), indicando uma forte relação positiva. Isso sugere que, à medida que aumentam as reclamações relacionadas a bagagem, como extravios, avarias ou problemas com bagagem de mão, também ocorre um aumento nos processos judiciais sobre esses temas. Da mesma forma, reclamações sobre problemas contratuais, como questões de oferta e compra de passagens, reembolsos, taxas, multas e programas de fidelidade, apresentam forte associação com processos judiciais da mesma categoria.

Outro aspecto relevante é a correlação entre reclamações de problemas operacionais e processos judiciais correlatos (0,63), evidenciando que questões como atrasos e cancelamentos de voos frequentemente resultam em litígios.

Além dessas, também foram observadas outras correlações significativas:

- Entre reclamações sobre problemas operacionais e processos relacionados a contrato (0,66) e bagagem (0,64);
- Entre reclamações sobre bagagem e processos relacionados a problemas operacionais (0,56) e contrato (0,55);
- Entre reclamações sobre contrato e processos relacionados a bagagem (0,64) e a problemas operacionais (0,56).

Esses valores indicam que problemas em uma área podem estar relacionados a litígios em outras, sugerindo uma interdependência entre diferentes tipos de reclamações e processos judiciais.

Por outro lado, as correlações envolvendo reclamações e processos classificados como 'outros' foram mais baixas, variando entre 0,24 e 0,52, indicando uma relação mais fraca entre esses tipos de reclamações e processos judiciais e os demais temas.

De forma geral, observa-se que as variáveis apresentam associações fortes ou moderadas, especialmente nos casos de bagagem, contrato e problemas operacionais. Isso reforça a importância de medidas preventivas para mitigar reclamações nessas áreas, reduzindo potenciais litígios. Entretanto, vale destacar que a correlação não implica causalidade, ou seja, uma redução nas reclamações não significa, necessariamente, uma diminuição dos processos judiciais. Ainda assim, esses dados podem ser utilizados pelas empresas para aprimorar o atendimento e minimizar os principais pontos de insatisfação dos clientes.

4.3 Resultados dos modelos de aprendizado de máquina para previsão do valor total de condenação

Essa seção apresenta os resultados dos modelos de aprendizado de máquina para prever o valor total de condenação. Como mencionado anteriormente, o valor total de condenação é relevante apenas para processos finalizados com decisão favorável ao(s) autor(es), ou seja, processos cuja situação processual e sentença sejam classificadas como condenação, acordo ou outros. Contudo, a categoria 'Outros' incluía 222.072 registros com dados não informados e 82.019, sem decisão definitiva. Dessa forma, para evitar viés nos modelos, optou-se por utilizar apenas os dados de processos com 'Condenação' (126.041 registros) ou 'Acordo' (81.266 registros). Também foram excluídos 13 registros dos processos que possuíam valores acima de R\$1 milhão para valor do pedido inicial, valor do dano moral, valor do dano material ou valor total de condenação. Assim, os modelos foram treinados e testados com um total de 207.294 registros de processos judiciais.

A Tabela 25 apresenta os resultados das métricas globais de desempenho de cada modelo, sendo que os melhores valores estão em negrito. Ressalta-se que para a precisão, o *recall* e o *F1-score* foi utilizado o parâmetro *average='weighted'*, tendo em vista o problema multiclasse. Esse parâmetro calcula a média dessas métricas, ponderando cada classe pela sua distribuição no conjunto de dados. Isso significa que classes com mais amostras têm um impacto maior na métrica final do que classes com menos amostras, refletindo melhor o desempenho geral do modelo em um problema multiclasse.

Tabela 25 – Desempenho global dos modelos de aprendizado de máquina

Métricas	NB	SVM Linear	RF
Acurácia	0,7228	0,7834	0,8192
Precisão	0,7438	0,8180	0,8539
<i>Recall</i>	0,7228	0,7834	0,8192
<i>F1-score</i>	0,7239	0,7925	0,8275
AUC-ROC	0,9026	0,9016	0,9463
Tempo de processamento (min)	0,1617	0,2404	12,6503

Fonte: Elaborado pela autora.

Observa-se na Tabela 25, que o RF se destacou com os melhores resultados para todas as métricas entre os três modelos, o que está de acordo com o trabalho de Torres, Guterres e Celestino (2023) e outros autores que eles usaram como referência (Ting *et al.*, 2011; Tsangaratos; Iliá, 2016; Lei *et al.*, 2017). Em relação aos tempos de processamento, o NB foi o que processou em menor tempo devido à sua simplicidade em comparação com as outras técnicas, seguido pelo SVM e o RF.

Analisando os resultados do RF, percebe-se que ele apresentou uma acurácia de 0,8192, o que significa que 81,92% das previsões feitas pelo modelo estão corretas em relação ao total de previsões realizadas. Em outras palavras, o modelo está classificando

corretamente a maioria das instâncias. Em relação à precisão dessas predições, ou seja, a proporção de predições positivas corretas em relação ao total de predições positivas, o valor foi de 0,8539. Isso significa que o modelo acertou 85,39% das predições positivas realizadas, indicando um bom desempenho na redução de falsos positivos. O *recall* foi de 0,8192, o que indica que o modelo capturou corretamente 81,92% das instâncias positivas existentes no conjunto de dados. Essa métrica é especialmente relevante em cenários onde é importante identificar todas as instâncias positivas, que no caso desse problema seria a identificação principalmente dos processos com valor pago alto. O *F1-score*, que combina precisão e *recall* em uma única métrica harmonizada, foi de 0,8275. Esse valor reflete um bom equilíbrio entre a capacidade do modelo de identificar corretamente as instâncias positivas (*recall*) e evitar falsos positivos (precisão), indicando um desempenho geral robusto. Por fim, a métrica AUC-ROC, que mede a capacidade do modelo de distinguir entre classes, apresentou um valor de 0,9463. Isso demonstra que o modelo tem uma excelente capacidade discriminativa, sendo capaz de separar as classes de forma eficiente mesmo no cenário com dados desbalanceados.

Este trabalho teve como objetivo superar a limitação relacionada às diferenças na classificação dos motivos e causas dos processos judiciais entre as companhias aéreas, conforme identificado por Torres, Guterres e Celestino (2023). Para isso, foi realizada a padronização dessas classificações, isto é, a reclassificação dos motivos em quatro categorias ('Problemas Operacionais', 'Bagagem', 'Contrato' e 'Outros') e das causas em sete categorias ('Aérea', 'Cliente', 'Terceiros', 'Casos Fortuitos', 'Força Maior', 'Não informado' e 'Outros'). Em seguida, os mesmos modelos foram executados, com o intuito de avaliar o desempenho após essa etapa. A Tabela 26 apresenta uma comparação entre os desempenhos obtidos por Torres, Guterres e Celestino (2023) e os resultados alcançados neste estudo, permitindo analisar se houve impacto nos desempenhos dos modelos.

Tabela 26 – Comparação do desempenho dos modelos de aprendizado de máquina

Métricas	Torres, Guterres e Celestino (2023)			Este trabalho			Diferença de desempenho		
	NB	SVM Linear	RF	NB	SVM Linear	RF	NB	SVM Linear	RF
Acurácia	0,766	0,809	0,831	0,7228	0,7834	0,8192	-5,64%	-3,16%	-1,42%
AUC-ROC	0,914	0,879	0,957	0,9026	0,9016	0,9463	-1,25%	2,57%	-1,12%
Tempo de processamento (min)	0,58	32,19	256,13	0,16	0,24	12,65	-72,41%	-99,25%	-95,06%

Fonte: Elaborado pela autora.

A análise apresentada na Tabela 26 evidencia uma queda no desempenho dos modelos após a padronização das classificações de motivos e causas. O modelo NB apresentou uma redução de -5,64% na acurácia e de -1,25% na AUC-ROC, indicando uma ligeira perda de desempenho em termos de precisão e discriminação entre classes. O SVM Linear também apresentou uma queda de -3,16% na acurácia, embora tenha mostrado uma melhora de 2,57% na AUC-ROC, sugerindo que a padronização impactou positivamente sua capacidade de distinguir classes, apesar de uma pequena redução na acurácia geral.

Por outro lado, o modelo RF apresentou reduções de -1,42% na acurácia e -1,12% na AUC-ROC. Quanto aos tempos de processamento, observou-se uma redução significativa para o SVM Linear (-99,25%) e o RF (-95,06%). Isso se deve principalmente à ausência da aplicação do *GridSearchCV* neste trabalho, uma vez que foram utilizados os mesmos hiperparâmetros definidos por Torres, Guterres e Celestino (2023).

Apesar das reduções observadas no desempenho das métricas, os valores permaneceram elevados, o que demonstra a robustez dos modelos desenvolvidos neste trabalho. É importante ressaltar que os modelos propostos consideraram uma quantidade maior de dados e companhias aéreas em comparação ao estudo de Torres, Guterres e Celestino (2023), que avaliou apenas duas empresas do setor. Além disso, foram incorporadas mais variáveis ao processo de modelagem (operacao, jurisdicao, instancia, n_autores, valor_pedido_inicial), o que aumentou a complexidade da análise. Dessa forma, o fato de os resultados se manterem próximos aos encontrados no trabalho anterior evidencia que a inclusão das novas variáveis, juntamente com a padronização das classificações, foi uma estratégia eficaz para aprimorar a precisão dos modelos e otimizar a eficiência computacional, especialmente em cenários mais complexos. Isso reforça a viabilidade da abordagem adotada neste trabalho para aplicações práticas em larga escala.

4.3.1 Matrizes de confusão

Além das métricas globais apresentadas, também foram calculadas as matrizes de confusão de cada modelo. Por meio delas, é possível avaliar o desempenho de um classificador por meio da análise de quais classes foram classificadas corretamente. As Tabelas 27, 28 e 29 apresentam as matrizes de confusão dos três modelos.

Tabela 27 – Matriz de Confusão - Naive Bayes

	Predito: Baixo	Predito: Médio	Predito: Alto
Real: Baixo	4102	620	638
Real: Médio	1020	17205	2379
Real: Alto	3579	3255	8661

Fonte: Elaborado pela autora.

Tabela 28 – Matriz de Confusão - SVM Linear

	Predito: Baixo	Predito: Médio	Predito: Alto
Real: Baixo	4531	147	682
Real: Médio	831	17276	2497
Real: Alto	3838	985	10672

Fonte: Elaborado pela autora.

Tabela 29 – Matriz de Confusão - Random Forest

	Predito: Baixo	Predito: Médio	Predito: Alto
Real: Baixo	4781	44	535
Real: Médio	873	18143	1588
Real: Alto	3608	846	11041

Fonte: Elaborado pela autora.

A análise das matrizes de confusão apresentadas nas Tabelas 27, 28 e 29 revela diferenças significativas no desempenho dos modelos de aprendizado de máquina avaliados. De maneira geral, todos os modelos apresentaram mais erros na classificação da classe Médio. O Naive Bayes, por exemplo, apresentou elevados índices de confusão, classificando muitos processos com valor pago Médio como Baixo (1.020 erros) ou Alto (2.379 erros). O SVM Linear demonstrou melhorias no desempenho geral, com maior precisão na classificação da classe Médio, mas ainda apresentou confusões frequentes entre as classes Alto e Médio (2.497 erros). O Random Forest, embora tenha cometido 1.588 erros na distinção entre as classes Alto e Médio, destacou-se como o modelo mais eficiente, apresentando o menor número de erros em todas as classes, especialmente na classificação da classe Alto.

A partir dos valores apresentados nas Tabelas 27, 28 e 29 foi possível calcular as métricas precisão, *recall* e *F1-score* para cada classe, conforme apresentado nas Tabelas 30, 31 e 32.

Tabela 30 – Métricas da classe Baixo

Métricas	NB	SVM Linear	RF
Precisão	47,14%	49,25%	51,62%
<i>Recall</i>	76,53%	84,53%	89,20%
<i>F1-score</i>	58,35%	62,24%	65,39%

Fonte: Elaborado pela autora.

Tabela 31 – Métricas da classe Médio

Métricas	NB	SVM Linear	RF
Precisão	81,62%	93,85%	95,32%
<i>Recall</i>	83,50%	83,85%	88,06%
<i>F1-score</i>	82,55%	88,57%	91,55%

Fonte: Elaborado pela autora.

Tabela 32 – Métricas da classe Alto

Métricas	NB	SVM Linear	RF
Precisão	74,17%	77,05%	83,87%
<i>Recall</i>	55,90%	68,87%	71,26%
<i>F1-score</i>	63,75%	72,73%	77,05%

Fonte: Elaborado pela autora.

Conforme os dados apresentados nas Tabelas 30, 31 e 32, observa-se os seguintes resultados:

- **Precisão:** A precisão indica a proporção de previsões corretas em relação ao total de previsões feitas para uma classe específica. Ou seja, avalia a qualidade das previsões positivas para cada classe. Os resultados foram:
 - **Classe Baixo:** O melhor modelo foi o RF, com 51,62% de precisão, seguido pelo SVM Linear (49,25%) e o NB (47,14%).
 - **Classe Médio:** O RF novamente foi o destaque, alcançando 95,32%, seguido pelo SVM Linear (93,85%) e o NB (81,62%).
 - **Classe Alto:** O RF manteve o melhor desempenho com 83,87%, seguido pelo SVM Linear (77,05%) e o NB (74,17%).
- **Recall:** O *recall* mede a proporção de instâncias reais de uma classe que foram corretamente identificadas pelo modelo. Ele avalia o quanto o modelo conseguiu encontrar de todas as ocorrências reais de uma classe. Os resultados foram:
 - **Classe Baixo:** O RF teve o melhor *recall* com 89,20%, seguido pelo SVM Linear (84,53%) e o NB (76,53%).
 - **Classe Médio:** O RF liderou novamente com 86,06%, seguido pelo SVM Linear (83,85%) e o NB (83,50%).
 - **Classe Alto:** O RF apresentou o maior *recall* com 71,26%, enquanto o SVM Linear (68,87%) e o NB (55,90%) ficaram em posições inferiores.
- **F1-score:** O *F1-score* é a média harmônica entre precisão e *recall*, sendo uma métrica útil para equilibrar esses dois aspectos, especialmente quando há desequilíbrio de classes. Os resultados foram:
 - **Classe Baixo:** O RF obteve o melhor *F1-score* com 65,39%, seguido pelo SVM Linear com 62,24% e o NB com 58,35%.
 - **Classe Médio:** O RF novamente destacou-se com 91,55%, seguido pelo SVM Linear com 88,57% e o NB com 82,55%.
 - **Classe Alto:** O RF alcançou o maior *F1-score* com 77,05%, com o SVM Linear (72,73%) e o NB (63,75%) logo atrás.

Os resultados indicam que o modelo Random Forest (RF) apresentou o melhor desempenho geral em todas as métricas e classes, seguido pelo SVM Linear e o NB. Na classe Médio, o RF destacou-se ao alcançar os maiores valores de precisão (95,32%), *recall* (86,06%) e *F1-score* (91,55%). Para a classe Alto, o RF manteve-se como o modelo mais

eficaz, com 68,11% de precisão, 77,98% de *recall* e 77,98% de *F1-score*. Na classe Baixo, o RF também liderou com 83,87% de precisão, 71,26% de *recall* e 77,05% de *F1-score*.

Quando comparados de maneira geral, o RF se destaca como o modelo mais consistente, apresentando um equilíbrio entre precisão, *recall* e *F1-score* em todas as classes. O SVM Linear mostrou-se competitivo, especialmente nas classes Médio e Alto, mas teve desempenho inferior ao RF em todas as métricas. O NB, embora apresente simplicidade e velocidade computacional como vantagens, mostrou-se menos eficaz, especialmente em cenários que exigem maior discriminação entre classes. Esses resultados reforçam a escolha do RF como o modelo mais robusto e confiável para o conjunto de dados analisado.

4.3.2 Importância das variáveis

Esse tópico apresenta as variáveis mais importantes de cada modelo identificadas por meio da técnica *permutation importance*. Porém, antes de analisar os resultados obtidos, é necessário observar as correlações entre as variáveis, tendo em vista que a multicolinearidade pode afetar os resultados de importância. Como apresentado anteriormente, a multicolinearidade ocorre quando duas ou mais variáveis independentes são altamente correlacionadas. Quando isso ocorre, as pontuações de importância das variáveis correlacionadas podem ser diluídas, já que a permutação de uma variável pode afetar a previsão do modelo de forma similar à permutação de outra variável correlacionada. Dessa forma, a Tabela 33 apresenta as correlações identificadas que são maiores que 0,5 e menores que -0,5.

Tabela 33 – Correlação entre Variáveis

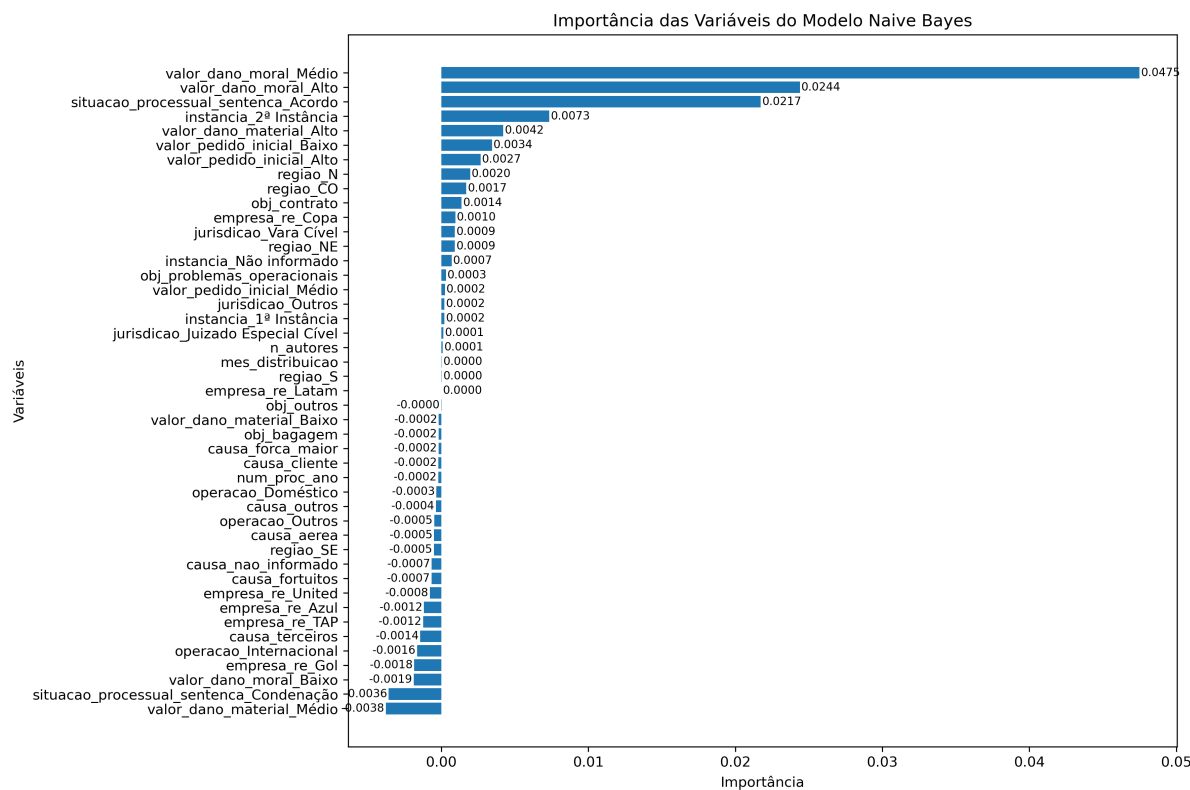
Variável 1	Variável 2	Correlação
situacao_processual_sentenca_Condenação	situacao_processual_sentenca_Acordo	1
jurisdicao_Vara Cível	jurisdicao_Juizado Especial Cível	0,929828
causa_nao_informado	empresa_re_Azul	0,910104
instancia_Não informado	instancia_1ª Instância	0,835570
valor_dano_material_Médio	valor_dano_material_Baixo	0,798854
empresa_re_Gol	empresa_re_Azul	0,784549
valor_dano_moral_Médio	valor_dano_moral_Baixo	0,775826
operacao_Outros	empresa_re_Gol	0,761142
causa_nao_informado	empresa_re_Gol	0,714021
causa_nao_informado	instancia_Não informado	0,707243
operacao_Doméstico	empresa_re_Azul	0,697312
operacao_Outros	operacao_Doméstico	0,693410
causa_fortuitos	empresa_re_United	0,667584
instancia_Não informado	empresa_re_Azul	0,658808
valor_pedido_inicial_Médio	valor_pedido_inicial_Baixo	0,656068
causa_nao_informado	operacao_Doméstico	0,626000
obj_bagagem	obj_problemas_operacionais	0,606302
causa_nao_informado	instancia_1ª Instância	0,595926
operacao_Outros	empresa_re_Azul	0,594983
instancia_Não informado	empresa_re_Gol	0,583793
instancia_1ª Instância	empresa_re_Azul	0,562528
causa_aerea	causa_nao_informado	0,554392
valor_dano_material_Baixo	valor_dano_material_Alto	0,549135
causa_nao_informado	operacao_Outros	0,542085
operacao_Internacional	empresa_re_TAP	0,541835
instancia_1ª Instância	empresa_re_Gol	0,515074
operacao_Internacional	operacao_Doméstico	0,514533
valor_dano_moral_Baixo	valor_dano_moral_Alto	0,512360

Fonte: Elaborado pela autora.

Na Tabela 33, percebe-se que as correlações são todas positivas, isto é, as variáveis aumentam ou diminuem juntas. Além disso, foram identificadas algumas correlações fortes, que são aquelas acima de 0,7 ou abaixo de -0,7, como: 'situacao_processual_sentenca_Condenação' e 'situacao_processual_sentenca_Acordo' que apresentaram uma correlação positiva perfeita (1), o que significa que elas aumentam ou diminuem juntas de forma linear; 'jurisdicao_Vara Cível' e 'jurisdicao_Juizado Especial Cível' (0,929828); 'causa_nao_informado' e 'empresa_re_Azul' (0,910104); 'instancia_Não informado' e 'instancia_1ª Instância' (0,835570). As correlações entre 'valor_dano_material_Médio' e 'valor_dano_material_Baixo' (0,798854), 'empresa_re_Gol' e 'empresa_re_Azul' (0,784549), 'valor_dano_moral_Médio' e 'valor_dano_moral_Baixo' (0,775826) e 'operacao_Outros' e 'empresa_re_Gol' (0,761142), 'causa_nao_informado' e 'empresa_re_Gol' (0,714021) e 'causa_nao_informado' e 'instancia_Não informado' (0,707243) também se destacaram. Esses altos valores indicam que as variáveis estão fortemente relacionadas, o que pode sugerir que ambas influenciam o modelo de forma similar.

Tendo como base as correlações identificadas, as Figuras 41, 42 e 43 apresentam os resultados da *permutation importance* e fornecem uma visão detalhada das variáveis que mais influenciaram a previsão do valor total de condenação em cada modelo.

Figura 41 – Importância das variáveis no Modelo NB



Conforme a Figura 41, as cinco variáveis com maior importância no modelo NB foram: `valor_dano_moral_Médio` (0,0475), `valor_dano_moral_Alto` (0,0244), `situacao_processual_sentenca_Acordo` (0,0217), `instancia_2ª Instância` (0,0073) e `valor_dano_material_Alto` (0,0042). Esses resultados indicam que variáveis relacionadas a valores de danos morais e materiais desempenham um papel central no desempenho do modelo, evidenciando a relevância dos valores de indenização nos processos analisados. O destaque para '`valor_dano_moral_Médio`', '`valor_dano_moral_Alto`' e '`valor_dano_material_Alto`' sugere que disputas envolvendo essas faixas de valores possuem características específicas que influenciam fortemente a classificação do valor total de condenação.

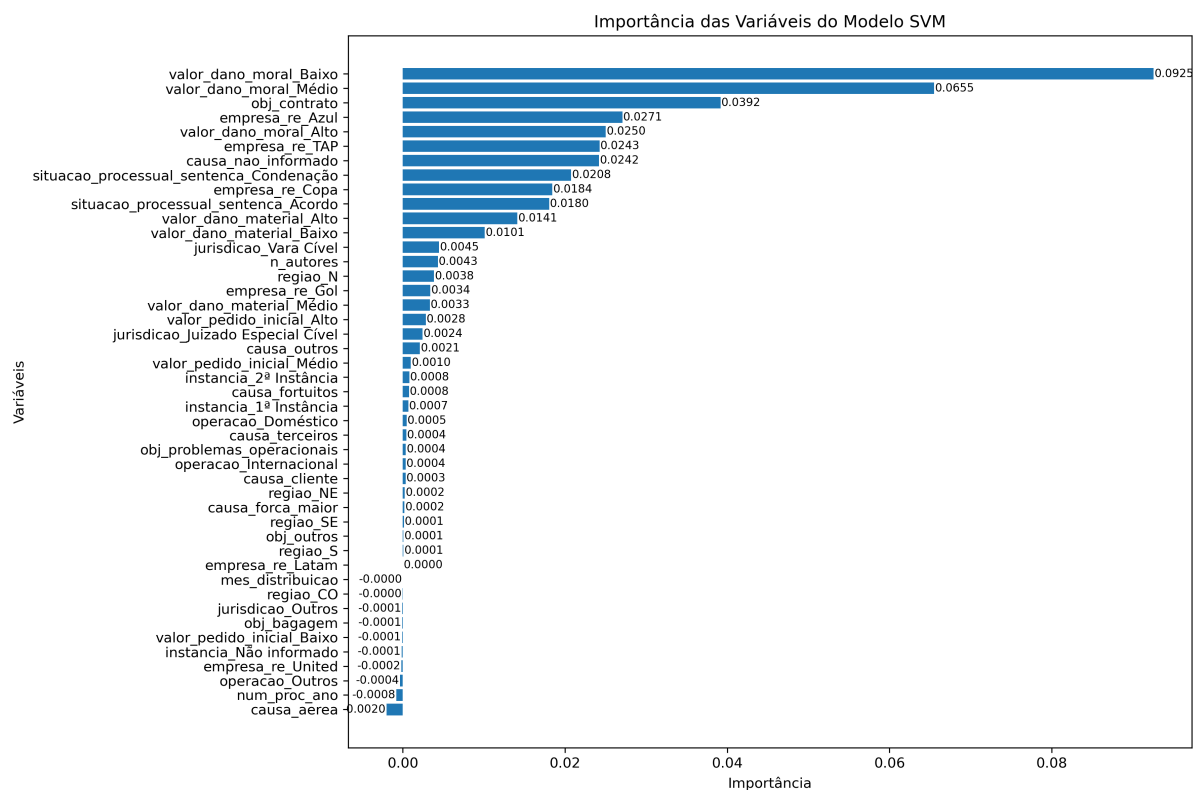
A variável '`situacao_processual_sentenca_Acordo`' aponta para a relevância de processos resolvidos por acordo, indicando que tais casos apresentam padrões processuais distintos que o modelo consegue identificar com maior precisão. Apesar da correlação perfeita (1) entre '`situacao_processual_sentenca_Acordo`' e '`situacao_processual_sentenca_Condenação`', é interessante observar que a importância atribuída a esta última foi

negativa (-0,0036). Isso pode refletir alguma redundância ou menor utilidade dessa variável no contexto do modelo. A inclusão de 'instancia_2ª Instância' entre as mais importantes demonstra a influência de casos analisados em grau de recurso, que frequentemente apresentam maior complexidade jurídica.

Além disso, variáveis regionais, como 'regiao_N' (0,0020) e 'regiao_CO' (0,0017), mostram uma importância moderada, sugerindo que diferenças geográficas podem ter algum impacto nos resultados dos processos. Por outro lado, variáveis relacionadas a operações, como 'operacao_Internacional' (-0,0016) e 'operacao_Doméstico' (-0,0003), e algumas categorias de empresas, como 'empresa_re_TAP' (-0,0012) e 'empresa_re_United' (-0,0008), apresentaram impacto negativo ou próximo de zero. Isso sugere que essas características não fornecem informações relevantes ou até confundem o modelo. A baixa importância de variáveis como 'causa_cliente' (-0,0002) e 'causa_forca_maior' (-0,0002) reforça que algumas categorias de causas processuais não têm peso significativo na predição.

Em síntese, o modelo demonstrou maior sensibilidade a variáveis financeiras e processuais, enquanto outros fatores, como operações e características regionais ou específicas de causas, tiveram menor impacto ou não foram discriminantes relevantes.

Figura 42 – Importância das variáveis no Modelo SVM Linear



Na Figura 42, observa-se que as cinco variáveis mais importantes no modelo SVM Linear foram: valor_dano_moral_Baixo (0,0925), valor_dano_moral_Médio (0,0655),

obj_contrato (0,0392), empresa_re_Azul (0,0271) e valor_dano_moral_Alto (0,0250). Nota-se uma consistência nas variáveis relacionadas a danos morais, especialmente no que se refere a 'valor_dano_moral_Médio' e 'valor_dano_moral_Alto', que apareceram nas listas de variáveis mais importantes no NB e SVM Linear. Isso indica que, independentemente do modelo, os valores atribuídos aos danos morais desempenham um papel central na predição do valor total de condenação.

A variável 'obj_contrato' também se destacou, o que indica que, para o modelo SVM, processos motivados por problemas contratuais tiveram influência na determinação do valor de condenação. A presença das variáveis 'empresa_re_Azul', 'empresa_re_TAP' (0,0243), 'empresa_re_Gol' (0,0034) e 'empresa_re_Copa' (0,0208) entre as mais importantes indica que o nome da empresa aérea influenciou na determinação do valor de condenação neste modelo. Essa influência pode estar associada ao histórico de litígios ou práticas específicas de cada empresa.

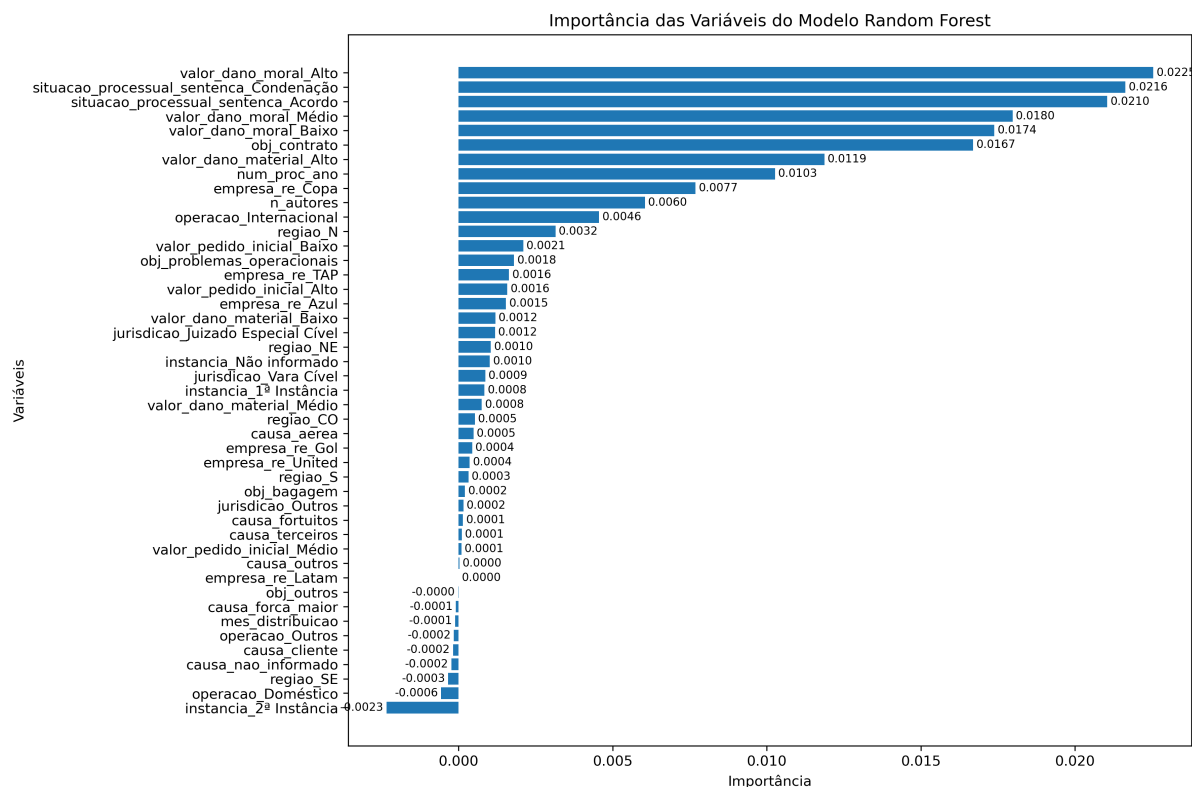
Diferente do NB, no modelo SVM, as variáveis 'situacao_processual_sentenca_Condenação' (0,0184) e 'situacao_processual_sentenca_Acordo' (0,0180) apresentaram uma importância mais equilibrada, sugerindo que o SVM é sensível às diferenças entre sentenças de acordo e condenação, possivelmente refletindo diferenças nos valores de indenização ou nas características dos casos.

Além disso, a relevância da variável 'causa_ao_informado' (0,0243) sugere que a falta de especificação da causa do processo pode ter um impacto significativo nos resultados, possivelmente indicando padrões processuais ou decisões judiciais em casos com dados incompletos.

Adicionalmente, as variáveis regionais não tiveram tanto destaque, sendo 'regiao_N' (0,0038) a mais significativa nesse grupo. Isso pode indicar que diferenças geográficas têm um impacto menor nos resultados quando comparadas a outros fatores, como o tipo de dano ou a natureza do processo. Outras variáveis de menor impacto, mas ainda relevantes, incluem 'valor_dano_material_Alto' (0,0141) e 'valor_dano_material_Baixo' (0,0123), que destacam a importância das variáveis financeiras na predição do modelo.

Em síntese, esses resultados sugerem que o modelo SVM Linear foi mais sensível a variáveis relacionadas a danos financeiros e características específicas das empresas, enquanto fatores processuais e regionais tiveram um impacto mais modesto na predição do valor de condenação.

Figura 43 – Importância das variáveis no Modelo RF



Na Figura 43, observa-se que as cinco variáveis mais importantes no modelo RF foram: 'valor_dano_moral_Alto' (0,0225), 'situacao_processual_sentenca_Condenação' (0,0216), 'situacao_processual_sentenca_Acordo' (0,0210), 'valor_dano_moral_Médio' (0,0180) e 'valor_dano_moral_Baixo' (0,0174). Assim como nos modelos anteriores, as variáveis relacionadas a danos morais continuam a desempenhar um papel central na predição do valor total de condenação, reforçando a importância dessas variáveis no contexto jurídico analisado.

A importância das variáveis 'situacao_processual_sentenca_Condenação' e 'situacao_processual_sentenca_Acordo' no RF sugere que o modelo é sensível ao tipo de decisão judicial, refletindo as diferenças entre sentenças de condenação e acordos, que possivelmente influenciam significativamente o valor das indenizações.

A variável 'obj_contrato' (0,0167) também foi destacada, indicando que questões contratuais continuam sendo relevantes na determinação do valor total de condenação. Adicionalmente, a variável 'valor_dano_material_Alto' (0,0119) aparece como uma das mais importantes, evidenciando a relevância de danos materiais no modelo RF.

Variáveis relacionadas a características das empresas, como 'empresa_re_Copa' (0,0077) e 'empresa_re_TAP' (0,0016), também aparecem, embora com menor importância, sugerindo que, mesmo com a presença de variáveis de danos financeiros e processuais, a

identidade da empresa aérea ainda tem influência nos resultados.

Outro ponto de destaque é a inclusão de variáveis operacionais e regionais, como 'num_proc_ano' (0,0103) e 'regiao_N' (0,0032), que indicam que aspectos relacionados ao ano e à localização geográfica também desempenham papéis na predição do modelo, embora de forma menos significativa.

Por fim, observa-se uma menor importância de variáveis como 'empresa_re_Gol' (0,0004) e 'causa_aerea' (0,0005), sugerindo que, neste modelo, essas variáveis não são tão preditivas quanto os fatores financeiros e processuais mais diretamente relacionados ao resultado do processo.

Em síntese, o modelo RF reforça a centralidade dos danos financeiros e processuais na predição do valor de condenação, enquanto as variáveis relacionadas a características das empresas e a localização geográfica desempenham um papel secundário, mas ainda relevante.

Em relação às variáveis mais importantes apresentadas para cada modelo, percebe-se que há uma convergência na relevância de algumas variáveis, especialmente aquelas relacionadas a danos morais e materiais. No modelo *Naive Bayes*, destaca-se a preponderância das variáveis financeiras e processuais, evidenciando uma sensibilidade maior a fatores diretamente ligados ao valor das condenações, embora a simplicidade do modelo possa limitar sua capacidade de capturar interações mais complexas. No caso do SVM, observa-se uma diversificação na importância das variáveis, incluindo tanto fatores financeiros quanto características específicas das empresas, refletindo a habilidade do modelo em identificar padrões mais sutis e interações não lineares. Por outro lado, o *Random Forest*, com sua capacidade intrínseca de lidar com multicolinearidade e interações complexas, apresenta uma distribuição de importância mais equilibrada, destacando tanto variáveis financeiras quanto processuais, e incorporando de maneira mais efetiva as nuances das interações entre as variáveis no contexto do problema.

4.4 Discussão dos resultados

Este estudo analisou as reclamações de consumidores e os processos judiciais contra empresas de transporte aéreo no Brasil. A análise temporal revelou dinâmicas distintas entre reclamações e processos judiciais. Enquanto o número de reclamações apresentou um crescimento contínuo entre 2019 e 2022, seguido por uma redução em 2023, os processos judiciais tiveram um comportamento diferente, com queda nos anos de 2020 e 2021, e aumento expressivo a partir de 2022. Esses padrões sugerem que fatores externos, como a pandemia da COVID-19, influenciaram de maneira significativa a relação entre consumidores e empresas aéreas, resultando em variações na frequência e intensidade das disputas registradas.

Ainda em relação ao tempo, observou-se que março apresentou o maior percentual de reclamações e processos judiciais. Outros meses com altos registros de reclamações foram outubro, janeiro, novembro e dezembro, enquanto os processos se concentraram também em fevereiro, novembro, outubro, setembro, maio e janeiro. Esses meses coincidem, em grande parte, com períodos de alta demanda no setor aéreo, como férias escolares e feriados prolongados, indicando uma relação entre o aumento do fluxo de passageiros e o volume de disputas registradas. Tal achado corrobora os resultados de Korfiatis *et al.* (2019), que analisaram as evoluções das dimensões de serviço ao longo do tempo, especialmente durante períodos de alta demanda, como férias de verão e feriados, e identificaram que há um aumento de reclamações em períodos de alta demanda, como o verão e o Natal.

Do ponto de vista geográfico, os dados indicam que a região Sudeste concentrou a maior parte dos registros, tanto de reclamações quanto de processos judiciais, o que pode ser explicado pelo maior volume de passageiros e operações aéreas na região. No entanto, um dado que merece destaque é a posição da região Norte, que, apesar de ter um volume de voos inferior ao das demais regiões, apresentou um índice de judicialização elevado. Esse achado corrobora com Barbosa (2024) e (Ferraz, 2024), que apontaram para uma relação entre o alto índice de judicialização e a reconfiguração da malha aérea, especialmente na região Norte.

Ao avaliar as empresas individualmente, verificou-se que a TAP foi a companhia internacional com o maior número de reclamações e processos, enquanto entre as nacionais, a Latam liderou em reclamações e a Azul em processos. A análise também apontou um possível viés nos dados da Gol, que não incluiu processos relacionados ao seu programa de fidelidade, ao contrário de outras empresas. Essa particularidade pode ter contribuído para um volume de processos aparentemente menor em comparação às demais companhias, evidenciando a necessidade de considerar diferenças nos critérios de registro de dados ao interpretar os resultados.

A análise dos motivos de reclamações e processos judiciais trouxe informações sobre os principais focos de insatisfação dos consumidores. Enquanto as reclamações foram predominantemente motivadas por questões contratuais, os processos judiciais tiveram como principal causa problemas operacionais. Observou-se uma continuidade nas reclamações de questões relacionadas a alterações de passagens e reembolsos, com uma ênfase crescente em problemas relativos aos programas de fidelidade e à acomodação de passageiros em casos de cancelamento ou mudança de voos. Nos processos judiciais, o cancelamento de voo foi o principal motivo em todos os anos analisados, seguido por atraso de voo, alteração de voo e extravio temporário de bagagem. Esses achados corroboram os resultados de Korfiatis *et al.* (2019), Kwon *et al.* (2021), Çallı e Çallı (2023) e Kim, Lim e Ha (2024), que destacam a insatisfação com reembolsos e cancelamentos como um problema recorrente no setor.

O impacto da pandemia de COVID-19 intensificou as reclamações sobre reembolsos, conforme apontado por ANAC (2023b). As medidas emergenciais para a aviação civil, estabelecidas pela Lei nº 14.034/2020 e alteradas pela Lei nº 14.174/2021, flexibilizaram as regras para alteração de passagens, cancelamento, reembolso e concessão de crédito até o final de 2021, com efeitos que se estenderam até 2022. Esse contexto levou a um aumento expressivo das queixas relacionadas a reembolsos nesse período, corroborando os achados de Çallı e Çallı (2023) e Kim, Lim e Ha (2024), que também identificaram um crescimento nas reclamações sobre reembolsos e questões correlatas durante a pandemia.

A análise de correlação revelou uma forte relação positiva entre reclamações e processos judiciais dentro das mesmas categorias, especialmente em relação a bagagem (0,67), contrato (0,66) e problemas operacionais (0,63). Além das correlações dentro de cada grupo, também foram observadas associações significativas entre reclamações e processos judiciais de diferentes categorias. Destacam-se as conexões entre reclamações sobre problemas operacionais e processos relacionados a contrato (0,66) e bagagem (0,64), assim como entre reclamações sobre bagagem e processos contratuais (0,55). Esses achados sugerem uma interdependência entre diferentes tipos de insatisfação dos consumidores, indicando que problemas em uma área podem desencadear disputas jurídicas em outra. Por outro lado, as correlações envolvendo reclamações e processos classificados como "outros" foram consideravelmente menores, variando entre 0,24 e 0,52. Isso indica uma relação mais fraca entre essas reclamações e os processos judiciais, possivelmente devido à diversidade e menor frequência de temas incluídos nessa categoria.

Por fim, os resultados do modelo de aprendizado de máquina demonstraram que o *Random Forest* apresentou o melhor desempenho em todas as métricas analisadas, corroborando os achados de Torres, Guterres e Celestino (2023) e dos autores por eles referenciados (Ting *et al.*, 2011; Tsangaratos; Ilia, 2016; Lei *et al.*, 2017). Além disso, a inclusão das novas variáveis, aliada à padronização das classificações de motivos e causas dos processos judiciais, mostrou-se uma estratégia eficaz para aprimorar a precisão dos modelos e otimizar a eficiência computacional. Essa abordagem permitiu superar a inconsistência na categorização dos processos entre companhias aéreas e demonstrou potencial para aplicações em cenários mais complexos, contribuindo para uma análise mais estruturada e padronizada das disputas judiciais no setor aéreo.

5 CONSIDERAÇÕES FINAIS

As companhias aéreas no Brasil enfrentam um volume significativo de reclamações e processos judiciais, impactando negativamente seus resultados financeiros e gerando consequências para os consumidores, como o aumento do preço das passagens e a redução da disponibilidade de malha aérea em determinadas regiões. Este estudo teve como objetivo analisar as reclamações de consumidores e os processos judiciais contra empresas de transporte aéreo no Brasil, investigando diferenças entre os períodos pré, durante e pós-pandemia da COVID-19, bem como a correlação entre essas variáveis. Além disso, foi avaliado o impacto da padronização da classificação dos motivos e causas dos processos judiciais no desempenho de modelos de aprendizado de máquina para previsão do valor das indenizações judiciais.

O primeiro objetivo específico foi analisar as características de período, localização, empresa, motivos e causas das reclamações e processos judiciais. A análise revelou um crescimento contínuo das reclamações de 2019 a 2022, com redução em 2023, enquanto os processos judiciais diminuíram em 2020 e 2021, mas cresceram significativamente a partir de 2022. O Sudeste liderou em registros tanto de reclamações como de processos judiciais. No entanto, a Região Norte destacou-se assumindo a terceira posição em quantidade de processos judiciais, apesar de ter menor volume de voos e passageiros, o que confirmou uma judicialização proporcionalmente alta na região. Entre as empresas, a TAP foi a companhia internacional com mais reclamações e processos, enquanto entre as nacionais, a Latam liderou em reclamações e a Azul em processos. Vale destacar que a Gol não incluiu em suas bases de dados os processos relacionados ao seu programa de fidelidade, ao contrário das demais, que os incluíram. Isso pode explicar o volume menor de processos apresentado pela Gol em comparação à Azul e à Latam.

Quanto aos motivos, os problemas operacionais predominaram nos processos judiciais, enquanto questões contratuais lideraram as reclamações. Ademais, observou-se que há temas comuns ao longo de todos os anos, como: alterações de passagens, reembolsos, cancelamento de voo, atraso de voo, alteração de voo e extravio temporário de bagagem. No entanto, problemas com reembolsos tiveram um aumento em 2020 e 2021, durante a pandemia de COVID-19. A respeito das causas, a principal é a própria companhia aérea. E observou-se que, de forma geral, os pedidos de danos morais são superiores aos materiais.

O segundo objetivo foi investigar a correlação entre as reclamações e os processos judiciais, buscando identificar se existe uma correspondência entre as questões mais frequentemente reportadas pelos consumidores e os motivos que levam a ações judiciais. Os resultados indicaram uma forte relação positiva entre reclamações e processos judiciais dentro das mesmas categorias, especialmente em relação a bagagem, contrato e

problemas operacionais.

O terceiro e último objetivo foi avaliar o impacto da padronização da classificação dos motivos e causas das reclamações e processos no desempenho dos modelos de aprendizado de máquina elaborados por Torres, Guterres e Celestino (2023), visando aprimorar a previsão do valor das indenizações judiciais. A análise demonstrou que os modelos apresentaram uma pequena queda no desempenho após a padronização. Porém, apesar das reduções observadas no desempenho das métricas, os valores permaneceram elevados, o que demonstra a robustez dos modelos desenvolvidos neste trabalho, principalmente do modelo *Random Forest*, que apresentou o melhor desempenho. De forma complementar, a análise da importância das variáveis indicou que o dano moral tem forte influência nas decisões judiciais de indenizações em processos contra companhias aéreas.

De forma geral, acredita-se que a incorporação de mais variáveis e dados de mais companhias tornou os modelos mais complexos, sendo necessário um novo ajuste nos hiperparâmetros, o que pode ser explorado em trabalhos futuros. Apesar disso, o fato de os resultados se manterem próximos aos encontrados no trabalho Torres, Guterres e Celestino (2023) evidencia que a inclusão das novas variáveis, juntamente com a padronização das classificações, foi uma estratégia eficaz para aprimorar a precisão dos modelos e otimizar a eficiência computacional, especialmente em cenários mais complexos. Isso reforça a viabilidade da abordagem adotada neste trabalho para aplicações práticas em larga escala.

Dessa forma, este estudo contribui significativamente para a compreensão das dinâmicas entre reclamações de consumidores e processos judiciais no setor aéreo brasileiro. A padronização da categorização dos motivos e causas permitiu uma visão mais estruturada dos principais fatores de insatisfação dos consumidores, possibilitando comparações mais precisas ao longo do tempo e entre diferentes companhias. Além disso, a aplicação de modelos preditivos, como o *Random Forest*, demonstrou o potencial da inteligência artificial para prever valores de indenizações e otimizar a tomada de decisões jurídicas e operacionais. Esse conhecimento pode auxiliar as empresas na previsão de custos futuros, na criação de estratégias para mitigar riscos e na proposição de acordos mais eficientes. As análises geradas também permitem às companhias identificar padrões recorrentes de reclamações e processos, possibilitando a implementação de medidas preventivas e aprimorando a qualidade do atendimento ao cliente. Para o setor aéreo como um todo, a abordagem adotada neste estudo pode servir como referência para o desenvolvimento de políticas regulatórias mais alinhadas às demandas dos consumidores, promovendo maior transparência, eficiência operacional e confiança no mercado.

Entre as limitações, destaca-se a grande quantidade de dados faltantes nas bases de judicialização, assim como a disponibilização de dados apenas de 16 estados brasileiros, o que limitou a análise. Futuras melhorias nos dados podem aprimorar a análise e o desempenho dos modelos de aprendizado de máquina. Para estudos futuros, além do

teste de novos hiperparâmetros para os modelos, sugere-se também investigar o perfil dos consumidores por região e Unidade Federativa, utilizando dados demográficos e pesquisas de comportamento de consumo, a fim de entender o porquê dos consumidores de algumas regiões reclamarem mais, enquanto os de outras optam pela judicialização. Além disso, sugere-se explorar detalhadamente os processos abertos após um longo período da data do voo, buscando entender as motivações dos consumidores em buscar a justiça depois de tanto tempo e os resultados de indenizações desses processos.

Por fim, depreende-se que todos os objetivos foram alcançados e que os resultados alcançados reforçam a importância de uma análise detalhada das reclamações e processos judiciais para melhorar a transparência e a confiança dos consumidores, além de oferecer ferramentas para minimizar conflitos e aprimorar os serviços prestados.

Apêndices

APÊNDICE A – INTELIGÊNCIA ARTIFICIAL

De forma geral, a Inteligência Artificial (IA) compreende qualquer técnica que permite aos computadores imitar o comportamento humano e reproduzir ou superar a tomada de decisões humanas para resolver tarefas complexas de forma independente ou com intervenção humana mínima (Pearson, 2021 *apud* Janiesch; Zschech; Heinrich, 2021). Segundo Kaplan e Haenlein (2019), IA pode ser definida como a capacidade de um sistema interpretar dados externos corretamente, aprender com esses dados e usar esses aprendizados para atingir objetivos e tarefas específicas por meio de adaptação flexível.

Ao longo do tempo, o campo de IA apresentou diversos avanços e deu origem a outros subcampos. No entanto, tendo em vista a rápida evolução da IA e as diversas perspectivas a partir das quais ela é considerada, não existe uma taxonomia de IA comumente acordada entre as comunidades de investigação, literatura ou relatórios (Samoili *et al.*, 2021). A partir disso, uma pesquisa realizada por Samoili *et al.* (2021), com o objetivo de propor uma definição operacional de IA, utilizou um método iterativo que fornece uma taxonomia concisa e lista de palavras-chave que caracterizam os domínios centrais da área de pesquisa em IA atualmente, a fim de auxiliar seu monitoramento. O Quadro 10 apresenta os domínios e subdomínios propostos pelos autores.

Quadro 10 – Taxonomia de IA

Taxonomia de IA		
Domínio de IA	Subdomínio de IA	
Essenciais	Raciocínio	Representação do conhecimento
		Raciocínio automatizado
		Raciocínio de bom senso
	Planejamento	Planejamento e agendamento
		Procurando (<i>Searching</i>)
		Otimização
	Aprendizagem	Aprendizado de máquina
	Comunicação	Processamento de linguagem natural
Percepção	Visão computacional	
	Processamento de áudio	
Transversal	Integração e Interação	Sistemas multiagentes
		Robótica e automação
		Veículos conectados e automatizados
	Serviços	Serviços de IA
	Ética e Filosofia	Ética da IA
Filosofia da IA		

Fonte: Elaborada pela autora com dados da Samoili *et al.* (2021).

Observa-se no Quadro 10, que os domínios estão divididos em essenciais e transversais. Os essenciais são os objetivos fundamentais da IA, enquanto os transversais são questões comuns a todos os domínios essenciais. Apesar disso, ressalta-se que os domínios e subdomínios sugeridos são subconjuntos relacionados e não separados. Isso ocorre devido a natureza do campo de IA que abrange aplicações interligadas e avanços teóricos com limites confusos (Samoili *et al.*, 2021). Assim, a taxonomia proposta pelos autores não pretendia ser uma classificação rígida ou exaustiva, mas sim uma coleção abrangente de áreas que representam a IA.

Entre os domínios e subdomínios de IA apresentados no Quadro 10, a aprendizagem e o aprendizado de máquina destacam-se para o objetivo deste trabalho. A aprendizagem refere-se à capacidade dos sistemas aprenderem, decidirem, preverem, adaptarem-se e reagirem automaticamente às mudanças, melhorando a partir da experiência, sem serem explicitamente programados (Samoili *et al.*, 2021). E o aprendizado de máquina está detalhado no subtópico a seguir.

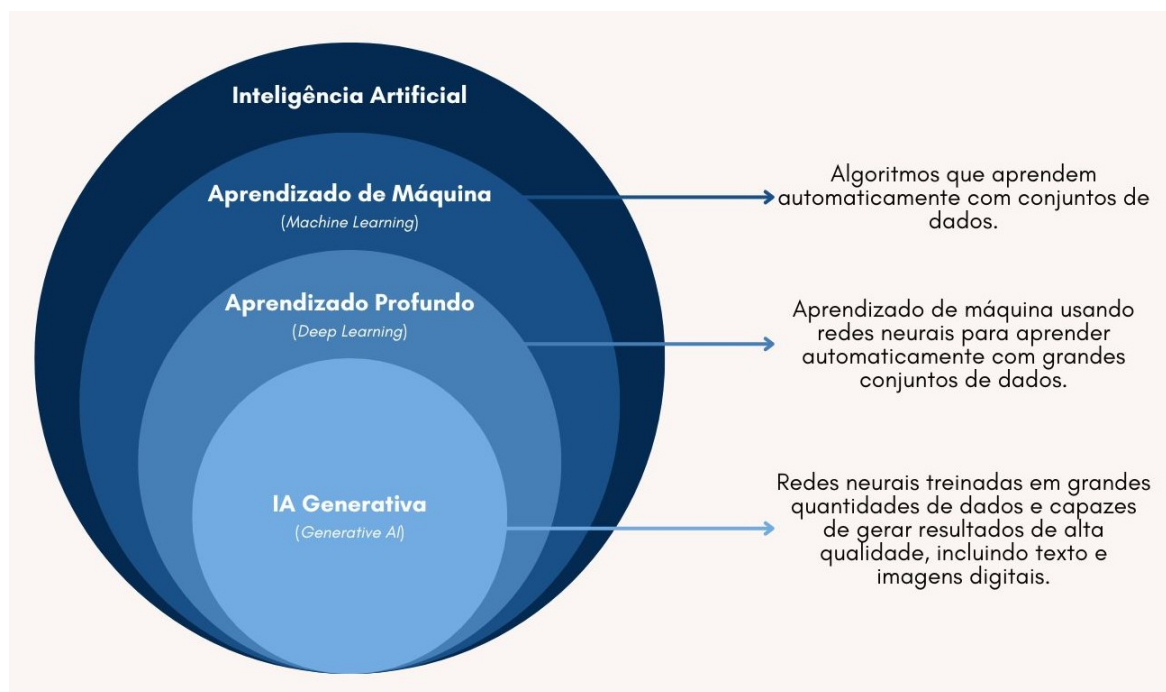
A.0.1 *Aprendizado de Máquina*

O aprendizado de máquina (Machine Learning - ML, em inglês) é um subcampo da IA que estuda a capacidade de resolução de problemas por meio de grande volume de dados. Para alcançar esse objetivo, ele se dedica a treinar algoritmos capazes de determinar regras ou parâmetros que estabelecem uma relação entre os dados de entrada (conhecidos como atributos preditivos) e os dados de saída (chamados de atributo alvo), viabilizando a execução de diversas tarefas, tais como classificação, previsão e agrupamento de dados (Lenz *et al.*, 2020).

Existem quatro categorias principais de aprendizado, sendo eles: supervisionado, não supervisionado, semisupervisionado e por reforço (Faceli *et al.*, 2021). A principal diferença entre essas categorias reside no tipo de dados empregados em cada uma. O aprendizado supervisionado faz uso de conjuntos de dados rotulados para treinar os algoritmos, ao passo que o aprendizado não supervisionado utiliza dados não rotulados. Os algoritmos de aprendizado semisupervisionado são capazes de lidar com uma grande quantidade de dados não rotulados e parcialmente rotulados, geralmente constituindo uma combinação dos algoritmos das duas categorias anteriores. Por fim, no aprendizado por reforço, um agente aprende de forma autônoma através da observação do ambiente, seleção e execução de ações, recebendo recompensas positivas ou penalidades com base em suas escolhas (Géron, 2019).

O aprendizado de máquina é um dos principais subdomínios da IA e assim como ela, evoluiu bastante nos últimos anos dando origem a outros subcampos, como o aprendizado profundo (Deep Learning - DL, em inglês) e a IA generativa (Generative AI, em inglês). A Figura 44 apresenta a relação entre todos esses campos de estudo.

Figura 44 – Subcampos da Inteligência Artificial



Fonte: Adaptado de HMG (2024).

Como observado na Figura 44, a IA generativa é um subcampo do aprendizado profundo que, por sua vez, é um subcampo do aprendizado de máquina. Uma característica comum desses dois subcampos é o uso de redes neurais, que são algoritmos de aprendizado de máquina inspirados em redes neurais biológicas. As redes neurais são compostas por neurônios artificiais interconectados, funcionando de forma semelhante às sinapses no cérebro. Cada conexão entre neurônios transmite sinais cuja intensidade é ajustada por um peso durante o aprendizado. Os sinais são processados pelos neurônios subsequentes se ultrapassarem um limite determinado por uma função de ativação. Geralmente, os neurônios são organizados em várias camadas, sendo que uma camada de entrada recebe os dados iniciais e uma camada de saída gera o resultado final, podendo existir várias camadas ocultas no meio. O número de camadas, neurônios e outras propriedades, como taxa de aprendizagem e função de ativação, são hiperparâmetros do modelo, definidos manualmente ou determinados por meio de uma rotina de otimização (Janiesch; Zschech; Heinrich, 2021).

O aprendizado profundo tem como base redes neurais profundas, ou seja, redes neurais que possuem mais de uma camada oculta, organizadas em arquiteturas de redes profundamente aninhadas e com neurônios avançados. Isso significa que eles podem usar operações avançadas, como convoluções, ou múltiplas ativações em um neurônio, no lugar de apenas uma simples função de ativação. Essas características permitem que redes neurais profundas sejam alimentadas com dados de entrada brutos e descubram automaticamente uma representação necessária para a tarefa de aprendizado correspondente (Janiesch;

Zschech; Heinrich, 2021).

A IA Generativa refere-se a IA capaz de produzir novos conteúdos, como imagens, textos ou músicas. Ela funciona usando grandes quantidades de dados, na maioria das vezes colhidos na Internet, para treinar um modelo nos padrões e na estrutura subjacentes desses dados. Após várias iterações de treinamento, envolvendo tanto máquinas quanto humanos, o modelo é capaz de gerar novos conteúdos, semelhantes aos utilizados no treinamento. O lançamento do ChatGPT em novembro de 2022 aumentou a consciência pública e o interesse nessa tecnologia, assim como iniciou uma aceleração no mercado de produtos de IA Generativa utilizáveis, como o Claude, Bard, Bedrock e Dall-E. No entanto, esse subcampo não é novo e têm sido objeto de pesquisas na última década (HMG, 2024).

A IA Generativa é composta por modelos de fundação, Large Language Models (LLM) e interfaces de usuários. Os modelos de fundação são grandes redes neurais treinadas em conjuntos de dados extremamente grandes para produzir respostas que se assemelham a esses conjuntos de dados. Esses modelos podem não ser baseados em linguagem e podem ser treinados em dados não textuais. Enquanto isso, os LLM são modelos de fundação treinados especificamente com dados de texto e linguagem natural para gerar resultados baseados em texto de alta qualidade. Por fim, as interfaces de usuários para modelos de fundação e LLMs são formas fáceis de usar para pessoas sem experiência técnica, como o ChatGPT e o Bard (HMG, 2024).

APÊNDICE B – CONSUMIDOR.GOV.BR

O Consumidor.gov.br é um serviço público e gratuito que viabiliza a comunicação direta entre consumidores e empresas para solução alternativa de conflitos de consumo pela internet. A Secretaria Nacional do Consumidor (Senacon) do Ministério da Justiça é a responsável pela gestão e disponibilização dessa plataforma, assim como pela articulação da cooperação técnica com outros órgãos e entidades do Sistema Nacional de Defesa do Consumidor que apoiam e atuam na consecução dos objetivos do serviço (Consumidor.gov.br, 2024b).

Os principais objetivos do Consumidor.gov.br são:

- Ampliar o atendimento aos consumidores;
- Incentivar a competitividade pela melhoria da qualidade de produtos, serviços e do relacionamento entre consumidores e empresas;
- Aprimorar as políticas de prevenção de condutas que violem os direitos do consumidor; e
- Fortalecer a promoção da transparência nas relações de consumo.

Nesse sentido, a intenção dessa plataforma é ampliar o acesso dos consumidores à solução de conflitos de consumo não resolvidos pelos canais de atendimento das empresas. No entanto, é importante destacar que o Consumidor.gov.br não constitui um processo administrativo e não substitui o atendimento tradicional oferecido pelos Órgãos de Defesa do Consumidor. Portanto, os consumidores podem utilizar esse serviço sem prejudicar o acesso aos canais de atendimento tradicionais providos pelo Estado, como os Procons Estaduais e Municipais, as Defensorias Públicas, o Ministério Público e os Juizados Especiais Cíveis (Consumidor.gov.br, 2024a).

Em relação a seu funcionamento, destaca-se que só é possível fazer reclamações de empresas cadastradas no site, visto que a participação das empresas no Consumidor.gov.br é voluntária e está sujeita à adesão formal ao serviço. Para isso, as empresas devem assinar um termo comprometendo-se a conhecer, analisar e empregar todos os esforços disponíveis para resolver os problemas apresentados. Por outro lado, os consumidores devem identificar-se adequadamente e comprometer-se a fornecer todos os dados e informações relativas à reclamação relatada (Consumidor.gov.br, 2024a).

Entre os benefícios proporcionados às empresas que aderem ao Consumidor.gov.br está o acesso direto e gratuito a um sistema que facilita o recebimento e tratamento das reclamações de seus consumidores, especialmente aquelas que não foram resolvidas

pelos canais tradicionais de atendimento, evitando assim litígios administrativos e/ou judiciais. Adicionalmente, a empresa obtém uma oportunidade adicional para demonstrar seu compromisso com a resolução de conflitos de consumo, por meio dos rankings publicados na plataforma. Por fim, as informações contidas na base de dados possibilitam uma análise detalhada dos problemas relacionados aos seus produtos e serviços, permitindo a adoção de medidas de melhoria e ajuste (Consumidor.gov.br, 2024a).

B.1 Adesão do Consumidor.gov.br pela ANAC

A ANAC foi a primeira agência reguladora federal a aderir à plataforma Consumidor.gov.br. Essa adesão ocorreu em dezembro de 2016, por meio do Acordo de Cooperação Técnica firmado com a Senacon do Ministério da Justiça. A partir do segundo trimestre de 2019, o Consumidor.gov.br tornou-se o sistema eletrônico de atendimento adotado pela ANAC para reclamações de consumidores contra empresas aéreas que operem serviços de transporte aéreo regular de passageiros, doméstico e internacional, no Brasil. Assim, em razão do previsto no artigo 39 da Resolução ANAC nº 400/2016, as empresas foram obrigadas a aderir e estar ativas no Consumidor.gov.br, além de responder as reclamações registradas na plataforma no prazo estabelecido. A Figura 45 apresenta o fluxograma de funcionamento da plataforma e a atuação da ANAC.

Figura 45 – Fluxograma do funcionamento do Consumidor.gov.br e atuação da ANAC



Fonte: ANAC (2023d).

Como apresentado na Figura 45, a ANAC monitora as reclamações registradas

pelos passageiros no Consumidor.gov.br e avalia, em âmbito coletivo, os serviços prestados pelas empresas do transporte aéreo. O objetivo é identificar as principais dificuldades enfrentadas pelos consumidores, bem como as tratativas fornecidas pelas empresas aéreas e, assim, subsidiar as ações de regulação e fiscalização do setor. Dessa forma, a ANAC mantém um painel interativo e publica, trimestralmente e anualmente, o Boletim de Monitoramento do Consumidor.gov.br, com dados dos principais indicadores das empresas aéreas relacionados a quantidade de reclamações e desempenho na solução de problemas na plataforma.

REFERÊNCIAS

ABEAR. **Panorama 2022 – O setor aéreo em dados e análises**. 2023. Disponível em: <<https://www.abear.com.br/wp-content/uploads/2023/12/Panorama2022-vf.pdf>>. 17, 32

ABEAR. **Cenário Setor Aéreo - 05 de março de 2024**. 2024. Disponível em: <<https://legis.senado.leg.br/sdleg-getter/documento/download/22408183-36e7-4458-8087-45d9a8833872>>. 44

AHMAD, I. *et al.* Fake news detection using machine learning ensemble methods. **Complexity**, v. 2020, n. 1, p. 8885861, 2020. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1155/2020/8885861>>. 70, 71

AKEN, J. E. V.; BERENDS, H.; BIJ, H. V. D. **Problem solving in organizations**. [S.l.]: Cambridge university press, 2012. 42

ALI, R.; BAKHSH, K.; YASIN, M. A. Causal nexus between air transportation and economic growth in brics countries. **Journal of Air Transport Management**, v. 107, p. 102335, 2023. ISSN 0969-6997. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0969699722001545>>. 16

ANAC. **Anuário do Transporte Aéreo - 2020**. 2021. Disponível em: <<https://www.gov.br/anac/pt-br/assuntos/dados-e-estatisticas/mercado-do-transporte-aereo/panorama-do-mercado/anuario-transporte-aereo/2020.zip>>. 16

ANAC. **Metadados do conjunto de dados: Dados do consumidor.gov**. 2021. Acesso em: 3 de novembro de 2024. Disponível em: <<https://www.anac.gov.br/acesso-a-informacao/dados-abertos/areas-de-atuacao/voos-e-operacoes-aereas/dados-do-consumidor-gov/52-dados-do-consumidor-gov>>. 49

ANAC. **Painel de Indicadores do Transporte Aéreo 2019**. 2021. Disponível em: <<https://www.gov.br/anac/pt-br/assuntos/dados-e-estatisticas/mercado-do-transporte-aereo/painel-de-indicadores-do-transporte-aereo/painel-de-indicadores-do-transporte-aereo-2019>>. 35

ANAC. **Painel de Indicadores do Transporte Aéreo 2020**. 2021. Disponível em: <<https://www.gov.br/anac/pt-br/assuntos/dados-e-estatisticas/mercado-do-transporte-aereo/painel-de-indicadores-do-transporte-aereo/painel-de-indicadores-do-transporte-aereo-2020>>. 35

ANAC. **Anuário do Transporte Aéreo - 2021**. 2022. Disponível em: <<https://www.gov.br/anac/pt-br/assuntos/dados-e-estatisticas/mercado-do-transporte-aereo/panorama-do-mercado/anuario-transporte-aereo/2021.zip>>. 16

ANAC. **Painel de Indicadores do Transporte Aéreo 2021**. 2022. Disponível em: <<https://www.gov.br/anac/pt-br/assuntos/dados-e-estatisticas/mercado-do-transporte-aereo/painel-de-indicadores-do-transporte-aereo/painel-de-indicadores-do-transporte-aereo-2021>>. 35

ANAC. **Anuário do Transporte Aéreo - 2022**. 2023. Disponível em: <<https://www.gov.br/anac/pt-br/assuntos/dados-e-estatisticas/mercado-do-transporte-aereo/panorama-do-mercado/anuario-transporte-aereo/2022.zip>>. 16

ANAC. **Boletim ANAC Consumidor 2022**. 2023. Disponível em: <https://www.gov.br/anac/pt-br/canais_atendimento/consumidor/boletins/2022/anual-2022/view>. 17, 60, 90, 141

ANAC. **Painel de Indicadores do Transporte Aéreo 2022**. 2023. Disponível em: <<https://www.gov.br/anac/pt-br/assuntos/dados-e-estatisticas/mercado-do-transporte-aereo/painel-de-indicadores-do-transporte-aereo/painel-de-indicadores-do-transporte-aereo-2022>>. 35

ANAC. **Reclamação de passageiro contra empresas aéreas**. 2023. Acesso em: 2 de março de 2024. Disponível em: <https://www.gov.br/anac/pt-br/canais_atendimento/consumidor>. 151

ANAC. **Anuário do Transporte Aéreo - 2023**. 2024. Disponível em: <https://www.gov.br/anac/pt-br/assuntos/dados-e-estatisticas/mercado-do-transporte-aereo/panorama-do-mercado/anuario-transporte-aereo/anuario_2023.zip>. 16, 45, 46, 76, 78, 96

ANAC. **Painel de Indicadores do Transporte Aéreo 2023**. 2024. Disponível em: <<https://www.gov.br/anac/pt-br/assuntos/dados-e-estatisticas/mercado-do-transporte-aereo/painel-de-indicadores-do-transporte-aereo/painel-de-indicadores-do-transporte-aereo-2023>>. 35, 45, 117

ATALIK, Ö. Customer complaints about airline service: a preliminary study of turkish frequent flyers. **Management Research News**, Emerald Group Publishing Limited, v. 30, n. 6, p. 409–419, 2007. 17

BARBOSA, M. **Aéreas reservam R\$1 bilhão por ano para honrar decisões judiciais**. 2024. Disponível em: <<https://oglobo.globo.com/economia/negocios/noticia/2024/02/17/aereas-reservam-r-1-bilhao-por-ano-para-honrar-decisoes-judiciais.ghtml>>. 35, 140

BAUER, L. Dissertação de Mestrado, **Estimação do coeficiente de correlação de Spearman ponderado**. Porto Alegre, Rio Grande do Sul: [s.n.], 2007. Disponível em: <<https://lume.ufrgs.br/handle/10183/11499>>. 63

BLEI, D. M.; NG, A. Y.; JORDAN, M. I. Latent dirichlet allocation. **Journal of machine Learning research**, v. 3, n. Jan, p. 993–1022, 2003. Disponível em: <<https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf?ref=http://githubhelp.com>>. 25

BORGES, L. C.; ALVARENGA, S. Turismo aéreo e crise: os desafios enfrentados pela aviação comercial durante covid-19. **GEOFRONTER**, v. 9, n. 1, out. 2023. Disponível em: <<https://periodicosonline.uems.br/index.php/GEOF/article/view/8204>>. 16

BOTELHO, J. R.; COSTA, P. **Excesso de judicialização na aviação comercial: os danos a um setor essencial**. 2023. Disponível em: <<https://www.jota.info/coberturas-especiais/aviacao-desafios-da-retomada/excesso-de-judicializacao-na-aviacao-comercial-os-danos-a-um-setor-essencial-29112023#:~:text=Excesso%20de%20judicializa%C3%A7%C3%A3o%20na%20avia%C3%A7%C3%A3o>>

20comercial%3A%20os%20danos%20a%20um%20setor%20essencial,-Reflex%C3%B5es%20sobre%20as&text=Em%202019%2C%20a%20Terceira%20Turma,automaticamente%20em%20dano%20moral%20presumido.> 36

BRADLEY, J.; RAJENDRAN, S. Increasing adoption rates at animal shelters: A two-phase approach to predict length of stay and optimal shelter allocation. **BMC Veterinary Research**, Springer, v. 17, p. 1–16, 2021. Disponível em: <<https://link.springer.com/article/10.1186/s12917-020-02728-2>>. 70

BREIMAN, L. Random forests. **Machine learning**, Springer, v. 45, p. 5–32, 2001. Disponível em: <<https://link.springer.com/article/10.1023/a:1010933404324>>. 70

BROCKE, J. V.; HEVNER, A. R.; MAEDCHE, A. **Design science research: cases**. [S.l.]: Springer, 2020. 43, 44

ÇALLI, L.; ÇALLI, F. Understanding airline passengers during covid-19 outbreak to improve service quality: topic modeling approach to complaints with latent dirichlet allocation algorithm. **Transportation Research Record**, SAGE Publications Sage CA: Los Angeles, CA, v. 2677, n. 4, p. 656–673, 2023. 18, 29, 90, 106, 140, 141

CASTRO, L. de; FERRARI, D. **Introdução à Mineração de Dados: Conceitos Básicos**. [S.l.: s.n.], 2016. 70

CAUCHICK-MIGUEL, P. A. **Metodologia científica para engenharia**. [S.l.]: Elsevier Brasil, 2019. 42

CHAHAL, H.; DEVI, P. Consumer attitude towards service failure and recovery in higher education. **Quality Assurance in Education**, Emerald Group Publishing Limited, v. 23, n. 1, p. 67–85, 2015. Disponível em: <<https://www.emerald.com/insight/content/doi/10.1108/QAE-07-2013-0029/full/html>>. 22

CHANG, Y.-W.; CHANG, Y.-H. Does service recovery affect satisfaction and customer loyalty? an empirical study of airline services. **Journal of Air Transport Management**, v. 16, n. 6, p. 340–342, 2010. ISSN 0969-6997. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0969699710000591>>. 22

CHENG, H.-T. *et al.* Wide & deep learning for recommender systems. In: **Proceedings of the 1st Workshop on Deep Learning for Recommender Systems**. New York, NY, USA: Association for Computing Machinery, 2016. (DLRS 2016), p. 7–10. ISBN 9781450347952. Disponível em: <<https://doi.org/10.1145/2988450.2988454>>. 25

CNJ. **Cartilha do Transporte Aéreo**. 2021. Disponível em: <https://www.cnj.jus.br/wp-content/uploads/2021/05/cartilha-transporte-aereo-CNJ_2021-05-20_V10.pdf>. 32

CONSUMIDOR.GOV.BR. **Conheça o Consumidor.gov.br**. 2024. Acesso em: 2 de março de 2024. Disponível em: <<https://www.consumidor.gov.br/pages/conteudo/publico/1>>. 150, 151

CONSUMIDOR.GOV.BR. **Sobre o serviço**. 2024. Acesso em: 2 de março de 2024. Disponível em: <<https://www.consumidor.gov.br/pages/conteudo/sobre-servico>>. 150

DRESCH, A.; LACERDA, D. P.; JUNIOR, J. A. V. A. **Design science research: método de pesquisa para avanço da ciência e tecnologia**. [S.l.]: Bookman Editora, 2015. 42, 43

- FACELI, K. *et al.* **Inteligência artificial: uma abordagem de aprendizado de máquina.** [S.l.: s.n.], 2021. 69, 147
- FERRAZ, L. **Em resposta a processos judiciais, aéreas cortam oferta de voos no Norte do Brasil.** 2024. Disponível em: <<https://valor.globo.com/empresas/noticia/2024/02/14/em-resposta-a-processos-judiciais-aereas-cortam-oferta-de-voos-no-norte-do-brasil.ghml>>. 35, 36, 140
- FERREIRA, R. *et al.* Preparação e análise exploratória de dados. **Porto Alegre: Sagah,** p. 13–21, 2021. 51, 53, 60, 63
- FREITAS, C. C. G. de. Dissertação de Mestrado em Economia, **Demanda por seguro de automóvel no Rio de Janeiro.** Rio de Janeiro, Rio de Janeiro: [s.n.], 2018. Disponível em: <<https://repositorio.fgv.br/items/ee5b7cea-d645-46b5-9b36-aebf4982d1a8>>. 70
- GÉRON, A. **Mãos à obra: aprendizado de máquina com Scikit-Learn & TensorFlow.** [S.l.]: Alta Books, 2019. 69, 147
- GODOI, F. M. R. de. Desempenho das empresas aéreas brasileiras em aeroportos coordenados e não coordenados. **Revista Debates em Economia Aplicada – REDEA,** v. 2, n. 3, fev. 2023. Disponível em: <<https://www.portaldeperiodicos.idp.edu.br/redea/article/view/6998>>. 45
- GOMES, S. B. V.; FONSECA, P. V. d. R. Análise econômico-operacional do setor de transporte aéreo: indicadores básicos. **BNDS Setorial,** v. 40, p. 131–162, 2014. Disponível em: <<https://web.bndes.gov.br/bib/jspui/handle/1408/2861>>. 45
- GUPTA, I.; CHATTERJEE, I.; GUPTA, N. Latent semantic analysis based real-world application of topic modeling: A review study. In: **2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS).** [s.n.], 2022. p. 1142–1149. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/9742848/references#references>>. 23
- HMG. **Generative AI framework for HM Government.** 2024. Disponível em: <https://assets.publishing.service.gov.uk/media/65c3b5d628a4a00012d2ba5c/6.8558_CO_Generative_AI_Framework_Report_v7_WEB.pdf>. 148, 149
- IATA. **O valor do Transporte Aéreo no Brasil.** 2019. Disponível em: <<https://www.iata.org/contentassets/bc041f5b6b96476a80db109f220f8904/brazil-o-valor-do-transporte-aereo.pdf>>. 16
- JANG, K.; PARK, S.; KIM, W.-J. Automatic construction of a negative/positive corpus and emotional classification using the internet emotional sign. **Journal of KIISE,** Korean Institute of Information Scientists and Engineers, v. 42, n. 4, p. 512–521, 2015. 25
- JANIESCH, C.; ZSCHECH, P.; HEINRICH, K. Machine learning and deep learning. **Electronic Markets,** Springer, v. 31, n. 3, p. 685–695, 2021. Disponível em: <<https://link.springer.com/article/10.1007/s12525-021-00475-2>>. 146, 148, 149
- KAPLAN, A.; HAENLEIN, M. Siri, siri, in my hand: Who’s the fairest in the land? on the interpretations, illustrations, and implications of artificial intelligence. **Business Horizons,** v. 62, n. 1, p. 15–25, 2019. ISSN 0007-6813. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0007681318301393>>. 146

- KIM, D.; LIM, C.; HA, H.-K. Comparative analysis of changes in passenger's perception for airline companies' service quality before and during covid-19 using topic modeling. **Journal of Air Transport Management**, v. 115, p. 102542, 2024. ISSN 0969-6997. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0969699724000073>>. 18, 22, 30, 90, 106, 140, 141
- KORFIATIS, N. *et al.* Measuring service quality from unstructured data: A topic modeling application on airline passengers' online reviews. **Expert Systems with Applications**, Elsevier, v. 116, p. 472–486, 2019. 18, 23, 24, 90, 140
- KWON, H.-J. *et al.* Topic modeling and sentiment analysis of online review for airlines. **Information**, MDPI, v. 12, n. 2, p. 78, 2021. 18, 23, 25, 90, 140
- LEI, M. *et al.* Automatically classify chinese judgment documents utilizing machine learning algorithms. In: SPRINGER. **Database Systems for Advanced Applications: DASFAA 2017 International Workshops: BDMS, BDQM, SeCoP, and DMMOOC, Suzhou, China, March 27-30, 2017, Proceedings 22**. [S.l.], 2017. p. 3–17. 128, 141
- LENZ, M. *et al.* **Fundamentos de aprendizagem de máquina**. [S.l.: s.n.], 2020. 147
- LIMA, J. P.; COSTA, J. A. Comparing clustering techniques on brazilian legal document datasets. In: SPRINGER. **International Conference on Hybrid Artificial Intelligence Systems**. [S.l.], 2022. p. 98–110. 36, 37
- LIN, H.-H.; WANG, Y.-S.; CHANG, L.-K. Consumer responses to online retailer's service recovery after a service failure: A perspective of justice theory. **Managing Service Quality: An International Journal**, Emerald Group Publishing Limited, v. 21, n. 5, p. 511–534, 2011. Disponível em: <<https://www.emerald.com/insight/content/doi/10.1108/09604521111159807/full/html>>. 22
- MARTINO, G. D.; PIO, G.; CECI, M. Multi-view overlapping clustering for the identification of the subject matter of legal judgments. **Information Sciences**, Elsevier, v. 638, p. 118956, 2023. 19, 36
- MITCHELL, T. M.; MITCHELL, T. M. **Machine learning**. [S.l.]: McGraw-hill New York, 1997. v. 1. 69
- MURPHY, K. P. **Machine learning: a probabilistic perspective**. [S.l.]: MIT press, 2012. 71
- OLIVEIRA, A. **Transporte aéreo: economia e políticas públicas**. [S.l.]: Pezco Editora Desenvolvimento, 2009. 44
- ONTL. **Boletim de Logística - A importância do transporte aéreo para o Brasil**. 2022. Disponível em: <<https://ontl.infrasa.gov.br/wp-content/uploads/2022/02/Setor-Aereo-Brasileiro-v3.pdf>>. 44
- PARANHOS, R. *et al.* Desvendando os mistérios do coeficiente de correlação de pearson: o retorno. **Leviathan (São Paulo)**, n. 8, p. 66–95, 2014. 63

- PARASURAMAN, A.; ZEITHAML, V. A.; BERRY, L. L. Servqual: A multiple-item scale for measuring consumer perc. **Journal of retailing**, Elsevier Limited, v. 64, n. 1, p. 12, 1988. Disponível em: <<https://www.proquest.com/openview/7d007e04d78261295e5524f15bef6837/1?pq-origsite=gscholar&cbl=41988>>. 30
- PARK, J.-J.; PARK, J.-W. Investigating the effects of service recovery quality elements on passengers' behavioral intention. **Journal of Air Transport Management**, v. 53, p. 235–241, 2016. ISSN 0969-6997. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0969699715300302>>. 22
- PEARSON (Ed.). **Artificial intelligence: a modern approach**. [S.l.: s.n.], 2021. 146
- PEDROSO, M. **Litigância predatória no setor aéreo: uma questão a ser enfrentada**. 2023. Disponível em: <<https://www.jota.info/coberturas-especiais/aviacao-desafios-da-retomada/litigancia-predatoria-no-setor-aereo-uma-questao-a-ser-enfrentada-31102023>>. 95
- PEFFERS, K. *et al.* A design science research methodology for information systems research. **Journal of Management Information Systems**, Routledge, v. 24, n. 3, p. 45–77, 2007. Disponível em: <<https://doi.org/10.2753/MIS0742-1222240302>>. 43
- PONT, T. R. D. *et al.* Regression applied to legal judgments to predict compensation for immaterial damage. **PeerJ Computer Science**, PeerJ Inc., v. 9, p. e1225, 2023. Disponível em: <<https://peerj.com/articles/cs-1225/>>. 36, 37, 38, 39
- PUCRS. **Demandas judiciais e morosidade da justiça civil: relatório final ajustado**. [S.l.], 2011. Disponível em: <<https://bibliotecadigital.cnj.jus.br/jspui/handle/123456789/119>>. 34
- RAGHAV, K. *et al.* Text and citations based cluster analysis of legal judgments. In: SPRINGER. **Mining Intelligence and Knowledge Exploration: Third International Conference, MIKE 2015, Hyderabad, India, December 9-11, 2015, Proceedings 3**. [S.l.], 2015. p. 449–459. 36
- RAGHUVVEER, K. *et al.* Legal documents clustering using latent dirichlet allocation. **IAES Int. J. Artif. Intell**, Citeseer, v. 2, n. 1, p. 34–37, 2012. 36
- RAJENDRAN, S.; SRINIVAS, S.; GRIMSHAW, T. Predicting demand for air taxi urban aviation services using machine learning algorithms. **Journal of Air Transport Management**, v. 92, p. 102043, 2021. ISSN 0969-6997. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0969699721000260>>. 70
- RITA, P.; MORO, S.; CAVALCANTI, G. The impact of covid-19 on tourism: Analysis of online reviews in the airlines sector. **Journal of Air Transport Management**, v. 104, p. 102277, 2022. ISSN 0969-6997. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0969699722000977>>. 18, 28
- ROBERTS, M. E. *et al.* Structural topic models for open-ended survey responses. **American journal of political science**, Wiley Online Library, v. 58, n. 4, p. 1064–1082, 2014. 24

- ROCHA, A. C. P. Dissertação de Mestrado Profissional em Computação Aplicada, **Mineração de textos para classificação de processos judiciais trabalhistas**. Brasília, Distrito Federal: [s.n.], 2020. Disponível em: <<http://www.realp.unb.br/jspui/handle/10482/37933>>. 70
- ROGERS, J. C.; WILLIAMS, T. G. Consumer personal values as antecedents to dyadic and third party public consumer complaining behavior: An exploratory study. **Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior**, v. 3, p. 71–81, 1990. 22
- SABO, I. C. *et al.* Clustering of brazilian legal judgments about failures in air transport service: an evaluation of different approaches. **Artificial Intelligence and Law**, Springer, v. 30, n. 1, p. 21–57, 2022. Disponível em: <<https://link.springer.com/article/10.1007/s10506-021-09287-3#citeas>>. 19, 36
- SALAMA, B.; CARLOTTI, D.; YEUNG, L. **As decisões da Justiça Trabalhista são imprevisíveis?** [S.l.], 2018. Disponível em: <<https://images.jota.info/wp-content/uploads/2019/07/3766fe23a027a8d593f98a85f29f1672.pdf>>. 34
- SAMOILI, S. *et al.* Ai watch. defining artificial intelligence 2.0. Publications Office of the European Union, Luxembourg (Luxembourg), n. KJ-NA-30873-EN-N (online), 2021. ISSN 1831-9424 (online). Disponível em: <<https://publications.jrc.ec.europa.eu/repository/handle/JRC126426>>. 146, 147
- SCARANELLO, B. B.; KLARMANN, J.; SILVA, P. de B. **Os impactos da cultura da judicialização no setor aéreo**. 2022. Disponível em: <<https://www.jota.info/coberturas-especiais/aviacao-desafios-da-retomada/os-impactos-da-cultura-da-judicializacao-no-setor-aereo-25112022#:~:text=De%20acordo%20com%20a%20ANAC%C2%B9,a%20aproximadamente%20R%24%20760%20milh%C3%B5es.>>>. 17, 33, 35
- SEZGEN, E.; MASON, K. J.; MAYER, R. Voice of airline passenger: A text mining approach to understand customer satisfaction. **Journal of Air Transport Management**, v. 77, p. 65–74, 2019. ISSN 0969-6997. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0969699718304873>>. 18, 23, 24
- SICSÚ, A. L.; SAMARTINI, A.; BARTH, N. L. **Técnicas de aprendizado de máquina**. [S.l.]: Editora Blucher, 2023. 36, 63, 71
- SILVEIRA, R. F. Dissertação de Mestrado em Direito, Justiça e Desenvolvimento, **Análise preditiva sobre o consumidor litigante A inteligência artificial pode prever o litígio entre o passageiro e a companhia aérea?** São Paulo, São Paulo: [s.n.], 2021. Disponível em: <<https://repositorio.idp.edu.br/handle/123456789/2976>>. 33
- SONG, C. *et al.* The adverse impact of flight delays on passenger satisfaction: An innovative prediction model utilizing wide deep learning. **Journal of Air Transport Management**, v. 114, p. 102511, 2024. ISSN 0969-6997. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0969699723001540>>. 18, 22, 25, 26
- STARLING, V. C. A. S.; VILLA, N. **A origem da judicialização desproporcional no mercado do transporte aéreo no Brasil**. 2022. Disponível em: <<https://www.jota.info/coberturas-especiais/aviacao-desafios-da-retomada/a-origem-da->

judicializacao-desproporcional-no-mercado-do-transporte-aereo-no-brasil-20062022>. 17, 32, 33

TING, S. *et al.* Is naive bayes a good classifier for document classification. **International Journal of Software Engineering and Its Applications**, v. 5, n. 3, p. 37–46, 2011. 128, 141

TOMPSON, J. J. *et al.* Joint training of a convolutional network and a graphical model for human pose estimation. In: GHAHRAMANI, Z. *et al.* (Ed.). **Advances in Neural Information Processing Systems**. Curran Associates, Inc., 2014. v. 27. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2014/file/e744f91c29ec99f0e662c9177946c627-Paper.pdf>. 25

TORRES, G. d. O.; GUTERRES, M. X.; CELESTINO, V. R. R. Legal actions in brazilian air transport: A machine learning and multinomial logistic regression analysis. **Frontiers in Future Transportation**, v. 4, 2023. ISSN 2673-5210. Disponível em: <<https://www.frontiersin.org/articles/10.3389/ffutr.2023.1070533>>. 18, 19, 20, 39, 63, 65, 68, 70, 128, 129, 130, 141, 143

TORRES, G. de O. Dissertação de Mestrado em Engenharia de Infraestrutura Aeronáutica, **Judicialização no Transporte Aéreo Brasileiro: uma Análise por meio de Aprendizado de Máquina e de Regressão Logística Multinomial**. São José dos Campos, São Paulo: [s.n.], 2022. Disponível em: <http://www.bdata.bibl.ita.br/tesesdigitais/lista_resumo.php?num_tese=78593>. 32, 70

TSAFARAKIS, S.; KOKOTAS, T.; PANTOUVAKIS, A. A multiple criteria approach for airline passenger satisfaction measurement and service quality improvement. **Journal of Air Transport Management**, v. 68, p. 61–75, 2018. ISSN 0969-6997. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0969699717300704>>. 22

TSANGARATOS, P.; ILIA, I. Comparison of a logistic regression and naïve bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size. **Catena**, Elsevier, v. 145, p. 164–179, 2016. 128, 141

ZHANG, H.; ZHOU, L. Similarity judgment of civil aviation regulations based on doc2vec deep learning algorithm. In: IEEE. **2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)**. [S.l.], 2019. p. 1–8. 36