



Universidade de Brasília  
Faculdade de Economia, Administração, Contabilidade e  
Gestão de Políticas Públicas  
Departamento de Administração

JOÃO VITOR DA SILVA MELO

**Análise exploratória dos dados de reclamações do setor de aviação  
civil entre os anos de 2019 a 2021: Uma aplicação da Linguagem  
de Programação Python**

Brasília - DF  
2023

JOÃO VITOR DA SILVA MELO

**Análise exploratória dos dados de reclamações do setor de aviação civil entre os anos de 2019 a 2021: Uma aplicação da Linguagem de Programação Python**

Monografia apresentada ao Departamento de Administração como requisito parcial à obtenção do título de Bacharel em Administração.

Professor Orientador: Professor Dr., Victor Rafael Rezende Celestino

Brasília – DF  
2023

JOÃO VITOR DA SILVA MELO

**Análise exploratória dos dados de reclamações do setor de aviação civil, entre os anos de 2019 à 2021: Uma aplicação da linguagem de programação Python**

A Comissão Examinadora, abaixo identificada, aprova o Trabalho de Conclusão do Curso de Administração da Universidade de Brasília do aluno  
**João Vitor da Silva Melo**

Professor Dr., Victor Rafael Rezende Celestino  
Professor-Orientador

Profa. Dra., Silvia Araujo dos Reis,  
Professora-Examinadora

Profa. Dra., Vanessa Cabral Gomes  
Professora-Examinadora

Brasília, 14 de fevereiro de 2023.

## **AGRADECIMENTOS**

Agradeço primeiramente a minha família que me apoiou em toda a minha busca por conhecimento, minha namorada pela motivação diária e a meu orientador Professor Dr. Victor por compartilhar seus conhecimentos com tanta diligência. A busca por objetivos dá sentido à vida, o propósito do conhecimento pleno, sendo inalcançável, nos proporciona sentido constante. Portanto agradeço a busca pelo saber.

The world's most valuable resource is no longer oil, but data.

The Economist

## RESUMO

Compreender as reclamações dos consumidores constitui relevante fonte de informações para as empresas e órgãos reguladores. Este trabalho objetiva efetuar uma análise exploratória para extrair e modelar os dados de reclamações dos consumidores do setor de aviação civil, referente aos anos de 2019 à 2021, registradas na plataforma Consumidor.gov e disponibilizadas pela Agência Nacional de Aviação Civil (ANAC) gerando informações úteis aos stakeholders. Utilizando como ferramenta a linguagem de Programação Python e métodos de estatística descritiva, o trabalho busca desenvolver um artefato eficiente e capaz de gerar informações úteis à tomada de decisão. A metodologia utilizada é a de *Design Science* e o período de análise coincide com o da pandemia do coronavírus em que o setor de aviação foi fortemente impactado. Portanto, investigou-se os possíveis efeitos da crise de Covid-19 nas reclamações. As análises realizadas permitiram maior entendimento sobre as características dos dados e demonstraram que o Python é uma ferramenta útil na análise exploratória de dados.

Palavras-chave: análise exploratória de dados, Python, aviação civil, *Design Science*, pandemia do coronavírus.

## **ABSTRACT**

The understanding of consumer complaints is an important source of information for companies and regulatory agencies. This work aims to realize an exploratory analyze to extract and model data about the consumers complaints civil aviation sector, referring to the years 2019 to 2021, registered on the Consumidor.gov platform and made available by the National Civil Aviation Agency (ANAC), generating useful information to stakeholders. Using Python programming language and descriptive statistics methods, this work seeks to develop an efficient artifact capable of generating useful information for the decision making. The methodology used is Design Science and the period of analysis match the coronavirus pandemic which the aviation sector was strongly. Therefore, this work investigated possible effects of the Covid-19 crisis on claims. The analyzes performed using Python allowed a greater understanding of the data characteristics and generated insights about the sector's complaints.

Keywords: exploratory data analyze, python, civil aviation, Design Science, coronavirus pandemic.

## LISTA DE FIGURAS

Figura 1 - Passageiros Pagos: embarques + desembarques .....	18
Figura 2 - Crescimento projetado de dados estruturados e não estruturados.....	19
Figura 3 - Prazo para Análise do Gestor.....	35
Figura 4 - Tempo de Resposta.....	35
Figura 5 - Sexo dos Reclamantes.....	36
Figura 6 - Faixa Etária dos Reclamantes.....	37
Figura 7 - Número de Reclamações por Ano.....	38
Figura 8 - Reclamações por canal de origem.....	42
Figura 9 - Registro de reclamações por mês.....	43
Figura 10 - Registro de Reclamações por Área.....	44
Figura 11 - Registro de Reclamações por grupo.....	45
Figura 12 - Histograma para Tempo de Resposta.....	46
Figura 13 - Resolução das Reclamações.....	47
Figura 14 - Notas dos Reclamantes.....	48
Figura 15 - Histograma para Notas dos Consumidores.....	49

## LISTA DE TABELAS

Tabela 1 - Abordagens estatísticas e estratégia utilizadas.....	20
Tabela 2 - Tipos de Artefatos.....	23
Tabela 3 - Quantidade de Voos s domésticos x Quantidade de passageiros.....	28
Tabela 4 - Metadados.....	30
Tabela 5 - Tipos de variável no Pyhton.....	34
Tabela 6 - Registros em branco na Base de dados.....	39
Tabela 7 - Registro de Reclamações por Estados entre 2019 e 2021.....	40
Tabela 8 - Número de Reclamações por Assunto.....	44

## SUMÁRIO

1.	INTRODUÇÃO.....	12
1.1	Contextualização .....	12
1.2	Formulação do problema .....	13
1.3	Objetivo Geral .....	13
1.4	Objetivos Específicos .....	14
1.5	Justificativa.....	14
2.	REVISÃO TEÓRICA.....	16
2.1	Reclamações .....	16
2.2	Efeitos da covid-19 no setor de aviação .....	18
2.3	Ciência de Dados .....	19
2.4	Análise Exploratória de Dados .....	21
2.5	Bibliotecas do Python.....	23
3.	MÉTODOS E TÉCNICAS DE PESQUISA .....	24
3.1.	Tipologia e descrição geral dos métodos de pesquisa.....	23
3.2.	Caracterização e descrição dos instrumentos de pesquisa.....	23
3.3	Caracterização da Organização, setor e área .....	26
3.3.1	Agência Nacional de Aviação Civil (ANAC).....	27
3.3.2	Consumidor.gov.Br .....	27
3.4	População e amostra da pesquisa.....	28
3.5	Procedimento de coleta de dados .....	29
3.6	Análise de dados .....	30
4.	RESULTADO E DISCUSSÃO.....	31
4.1.	Avaliação Inicial e limpeza dos dados (Data Cleaning).....	31
4.2.	Perfil dos reclamantes.....	37
4.3.	Registro de Reclamações por Empresas.....	38

4.4. Quantidade de Reclamações Registradas .....	39
4.5. Registros Incompletos .....	40
4.6. Análise por Regiões.....	41
4.7. Registro de Reclamações por Canal de Origem .....	43
4.8. Períodos com o maior número de reclamações .....	43
4.9. Registro de Reclamações por Classificação .....	44
4.10. Tempo de Resposta .....	46
4.11. Resolução das reclamações pelas empresas .....	477
4.12. Nota para o Atendimento da Empresa.....	48
5. CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS .....	51
REFERÊNCIAS BIBLIOGRÁFICAS .....	54
APÊNDICES .....	55
APÊNDICE A - Notebook Análise de Reclamações – Parte I.....	56
APÊNDICE B - Notebook Análise de Reclamações – Parte II.....	57
APÊNDICE C - Notebook Análise de Reclamações – Parte III .....	59

# 1. INTRODUÇÃO

## 1.1 Contextualização

O estudo das reclamações vem ganhando destaque nas organizações, as empresas passaram a considerar a insatisfação dos clientes como uma oportunidade de identificar pontos de melhoria em seus produtos e serviços, compreender os principais problemas enfrentados pelos clientes e modificar a percepção de um indivíduo até então insatisfeito, através de ações de *service recovery* (SCHWEIKHART, 1993). O entendimento sobre as reclamações tem, portanto, o potencial de fornecer importantes *insights* sobre os negócios, além de constituir fonte de tomada de decisões estratégicas.

Com o advento da Tecnologia da Informação e Comunicação (TIC), o volume de dados obtidos e armazenados pelas corporações aumentou significativamente (DHAR, 2013). Nesse contexto, os dados sobre reclamações também se tornaram mais robustos, levando diversas empresas a empreenderem esforços relacionados a compreensão desses dados, objetivando obter as vantagens supracitadas. O avanço das ferramentas de ciência de dados impulsionou esse fenômeno de maior interesse das empresas por compreender os dados acerca das características e comportamentos de seus consumidores.

No setor aéreo comercial de transporte regular de passageiros, aqui chamado apenas de setor de aviação civil, esse fenômeno também pode ser observado, considerando o contexto de que as empresas buscam melhorar constantemente a qualidade de seus serviços. No setor de aviação civil, o principal produto consumido é o serviço de transporte aéreo, geralmente os clientes que consomem esse serviço precisam viajar longas distâncias em curtos períodos de tempo, optando por voar em detrimento de outros meios de transporte mais baratos, porém mais lentos, como ônibus ou carro. Consumidores do transporte aéreo, geralmente utilizam o serviço como um meio para se alcançar determinado fim, ou seja, os clientes precisam se deslocar para realizar alguma atividade, por exemplo: uma reunião de trabalho, visitar um amigo ou viagem a lazer, conseqüentemente problemas no deslocamento geralmente afetam ainda a atividade fim que gerou o consumo do serviço. Portanto, problemas na prestação dos serviços de transporte aéreo, como atrasos e cancelamentos, costumam impactar fortemente a percepção de qualidade dos consumidores.

No setor aéreo comercial de transporte regular de passageiros uma grande fonte de informações é a Agência Nacional de Aviação Civil (ANAC), que fornece diversos relatórios

e compilados de dados sobre do desempenho do setor anualmente. Esses dados são uma oportunidade para se aprofundarem os estudos acerca do setor tanto pelo viés acadêmico, quanto das empresas atuantes no setor.

Nesse contexto, esse trabalho busca explorar especificamente os dados referentes as reclamações, no setor de aviação civil brasileiro, relacionados ao transporte regular de passageiros entre os anos de 2019 a 2021. Os dados de reclamações são disponibilizados anualmente pela ANAC e descritos pela agência como “dados relativos às reclamações registradas na plataforma Consumidor.gov.br ([www.consumidor.gov.br](http://www.consumidor.gov.br)) contabilizam os atendimentos realizados pelas empresas aéreas em atenção às demandas dos passageiros.”. A análise desses dados é uma oportunidade de compreender os principais problemas enfrentados pelos usuários do transporte aéreo, além de ser favorável às empresas a obterem insights sobre seus processos e serviços.

## **1.2 Formulação do problema**

A ANAC disponibiliza os dados de reclamações para download em arquivos segmentados por anos a partir do ano de 2019, em dois formatos, csv e json, estruturados em formato de tabela. Para visualizar e editar essas tabelas é necessário que o usuário possua um software compatível e uma máquina com capacidade computacional para trabalhar com esses dados. Cada tabela possui em média 28 colunas e mais de 40 mil linhas. Diante dessas características nota-se que o formato em que os dados são disponibilizados dificultam uma análise eficiente, sem que essas informações sejam devidamente tratadas.

Diante da importância de compreender os dados acerca das reclamações do setor de aviação, entendimento esse que pode gerar informações e insights para um conjunto de empresas do setor, para os clientes e órgãos reguladores, como a própria ANAC, especialmente considerando os dados referentes aos anos mais recentes disponibilizados no site. Nesse sentido a questão desta pesquisa é “Como extrair e modelar o grande conjunto de dados de reclamações dos consumidores do setor de aviação civil, disponibilizados pela ANAC, possibilitando a retirada de informações úteis aos stakeholders?”.

## **1.3 Objetivo Geral**

Esse trabalho tem como objetivo uma análise exploratória para extrair e modelar os dados de reclamações dos consumidores do setor aéreo comercial de transporte regular de passageiros, chamado aqui de setor de aviação civil, referente aos anos de 2019 à 2021, disponibilizados pela ANAC, gerando informações úteis aos stakeholders. Para tal busca desenvolver uma ferramenta capaz de realizar a extração, tratamento e análise dessas informações, através da linguagem de programação Python.

#### **1.4 Objetivos Específicos**

Pretende-se alcançar quatro principais objetivos:

1. Compor um modelo de notebook em Python que extraia os dados do portal da ANAC;
2. Desenvolver um notebook que realize a análise exploratória dos dados extraídos;
3. Interpretar os resultados obtidos a partir da análise dos dados; e
4. Investigar os possíveis efeitos da pandemia nas reclamações, considerando que o período de análise coincide com o da pandemia mundial de covid-19.

#### **1.5 Justificativa**

A compreensão das reclamações registradas acerca do setor de aviação contribui para a sociedade reivindicar seus direitos como consumidores. Apenas uma pequena parte da população tem acesso a ferramentas e o conhecimento necessário para realização de análises mais robustas. Além de ser um importante informativo sobre como as empresas do setor tratam as reclamações de seus clientes, constituindo fator possivelmente relevante a escolha e tomada de decisão de compra de passagens aéreas.

O mercado de aviação também pode se beneficiar através dos insights retirados dos dados. Trazer luz a esses dados pode influenciar as empresas a desempenharem maiores esforços nas resoluções dos problemas dos consumidores, elevando a qualidade dos serviços. O uso da linguagem Python, em especial da biblioteca Pandas, para modelagem de dados abre um leque de possibilidades, considerando a flexibilidade dessa linguagem de programação e o factível uso de outras bibliotecas em complemento ao Pandas. O artefato criado, com algumas adaptações, pode ser utilizado na análise de outros dados que possuam estrutura semelhante, ou ainda, servir de ponto de partida para o desenvolvimento de futuras tecnologias

No contexto acadêmico este trabalho objetiva ocupar o gap existente devido a pequena quantidade de trabalhos que se valem da Análise Exploratória de dados, para investigar o tema das reclamações, sobretudo no setor de aviação.

## 2. REVISÃO TEÓRICA

Esta revisão fornece o embasamento teórico fundamental a compreensão do processo de análise adotado, destacando os principais conceitos envolvidos. Cabe reiterar que a revisão teórica corresponde à terceira etapa do Design Science, Revisão da Literatura, paradigma utilizado como metodologia neste trabalho, e descrito no tópico 3.1 Tipologia e Descrição Geral dos Métodos de Pesquisa.

Retomando o objeto deste trabalho que é modelar e compreender os dados de reclamação do setor de aviação, é primordial compreender o que são essas reclamações e qual seu papel no ambiente de negócios. Além de considerar possíveis efeitos da pandemia causada pela COVID – 19 no setor. Nesse sentido esses são os dois primeiros tópicos desta revisão.

Considerando que esse trabalho busca realizar uma análise exploratória dos dados, a partir de um grande conjunto de dados, ele perpassa pelo campo da ciência de dados, comumente chamada de *Data Science*, para esclarecer esse conceito foi dedicado a ele um tópico. Em seguida retomo a análise exploratória de dados, que utiliza a estatística como meio para compreensão dos dados.

### 2.1 Reclamações

Reclamações geralmente partem de uma insatisfação com o produto adquirido ou serviço prestado. De acordo com a tipologia definida por Jagdip Singh (apud HALSTEAD, 2002) existem três dimensões acerca do comportamento de reclamação de clientes: *voice complaints* (reclamações feitas diretamente ao vendedor), *private complaints* (reclamações feitas a amigos e parentes) e *third party complaints* (reclamações realizadas a organizações independentes de mídias, grupos de consumo ou agências legais em busca de reparação). Singh argumenta que os tipos de reclamação são ações independentes e que pesquisas futuras devem abordar cada tipo individualmente. Esse trabalho segue tal orientação, considerando que as reclamações investigadas neste trabalho são notadamente do tipo *third party complaints*, como esclarecido a seguir.

No contexto deste trabalho reclamações são entendidas como queixas registradas por consumidores na plataforma Consumidor.gov.br, um ambiente público e digital disponibilizado pelo governo federal para que os clientes possam se comunicar diretamente com as empresas para registrar suas reclamações e buscarem reparação pela perda ou dano percebidos na relação

de consumo. Caso não obtenham podem ainda recorrer aos órgãos oficiais de Defesa do Consumidor, pertencentes ao Sistema Nacional de Defesa do Consumidor, que é uma estrutura de âmbito federal, e nos Sistemas Estadual e Municipal de Defesa do Consumidor. O órgão que materializa o principal canal para registros das reclamações é o PROCON - Programa de Proteção e Defesa do Consumidor. De acordo com a Lei nº 8.078 de 11 de setembro de 1990, os Procons estaduais, municipais e demais entidades que visam a defesa do consumidor, estão aparados e pertencem ao Sistema Nacional de Defesa do Consumidor

O gerenciamento das reclamações ganhou especial atenção na literatura a partir dos anos 80, com o aumento do interesse pelo conceito de *service recovery*, em língua portuguesa recuperação de serviços, que está diretamente ligado a satisfação e lealdade dos clientes. Sharon B. Schweikhart define *service recovery* como parte do programa da administração da qualidade designado a alterar a percepção negativa de um consumidor insatisfeito e em última instância manter o relacionamento de consumo entre o consumidor e a empresa (SCHWEIKHART, 1993). Empresas capazes de responder a insatisfações e reclamações com uma estratégia adequada de *service recovery*, entregando uma resolução satisfatória ao problema, podem transformar um cliente insatisfeito em um satisfeito, impactando positivamente a imagem da companhia e as taxas de recompra (HALSTEAD, 2002).

A interação da sociedade com o meio digital impactou significativamente as relações de consumo. Os clientes possuem agora uma infinidade de fontes de informação para embasarem sua tomada de decisão, como vídeos, avaliações do produto, blogs, sites de reclamações, etc. Para esse trabalho destacamos os sites especializados em registrar reclamações, no Brasil os principais são o Reclame Aqui e o Bondfaro, essas plataformas permitem que os usuários registrem abertamente suas reclamações sobre a compra de produtos ou serviços que geraram insatisfação, com o objetivo de compartilharem suas experiências de consumo. Os usuários dessas plataformas, antes de realizar a compra de um produto ou contratação de um serviço, costumam acessá-las para obter informações sobre os problemas enfrentados pelos demais consumidores daquele produto, como forma de obter informações que subsidiem sua decisão de realizar ou não a compra ou contratação.

No contexto deste trabalho a ANAC disponibiliza o registro das reclamações realizadas por clientes da aviação civil no Consumidor.gov.br, a maioria das reclamações é realizada através de plataformas digitais. Diferente das plataformas não oficiais como o Reclame Aqui, nos canais oficiais as reclamações não são disponibilizadas abertamente ao público, apenas o reclamante pode fazer o acompanhamento da reclamação. Outra diferença é que nos canais não oficiais o consumidor está buscando compartilhar sua experiência com determinado produto ou

serviço com os demais usuários, porém ao fazer uma reclamação em um canal oficial do Consumidor.gov.br geralmente o cliente está buscando ressarcimento pelo dano sofrido, caracterizando reclamações do tipo *third party complaints*.

O Estudo das reclamações caracterizadas como *third party complaints* é uma oportunidade para as empresas compreenderem os principais problemas em seus produtos e serviços identificados pelos próprios clientes, além de fonte de insights para a implementação do *service recovery*. Portanto a compreensão dessas insatisfações poder gerar diversas possibilidades de melhorias no contexto organizacional.

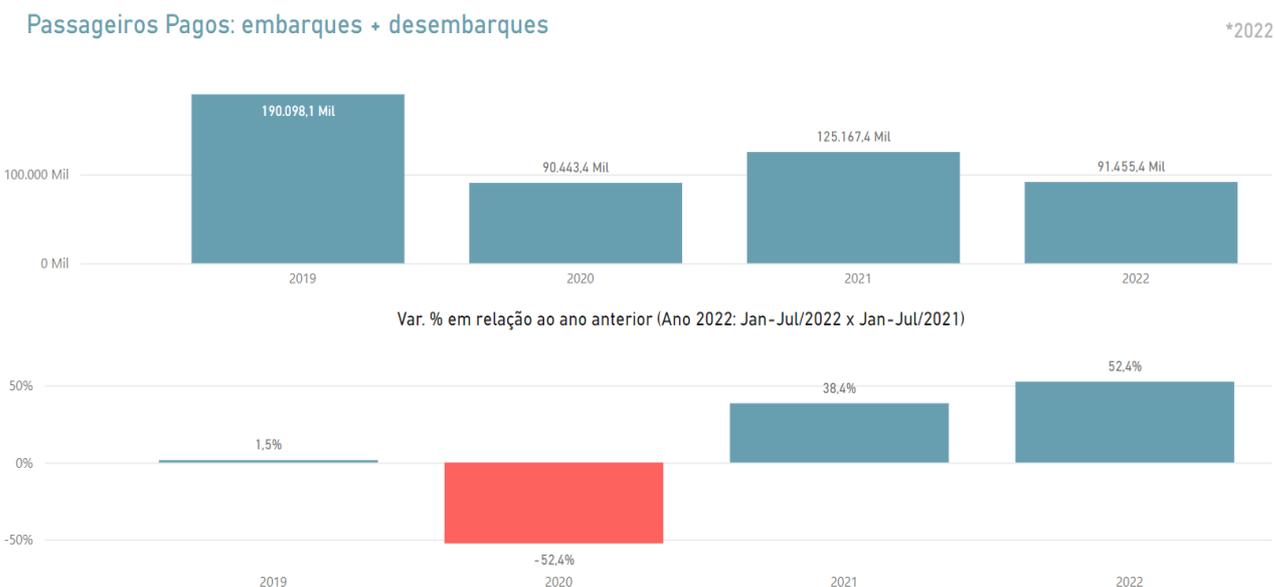
A falta de ferramentas de extração e análise dessas reclamações constitui uma lacuna tanto acadêmica como de mercado, diante dos benefícios já explicitados. Nesse sentido esse trabalho busca preencher essa lacuna.

## **2.2 Efeitos da covid-19 no setor de aviação**

No ano de 2020 teve início uma pandemia global causada pelo vírus SARS–COV–2, conhecido popularmente como coronavírus ou vírus da Covid-19. A busca pela contenção do vírus levou diversos países a imporem severas restrições de diversas naturezas, como: fechamento de fronteiras, restrição de transito interno, fechamento de comércios, toques de recolher, entre outras medidas sanitárias e de distanciamento social. Cabe destacar que uma das principais medidas ao redor do globo foi o fechamento de diversos aeroportos e restrições de transito de pessoas, tornando o setor de aviação um dos mais afetados.

As alterações no cenário económico mundial impostas pela pandemia do coronavírus causaram a maior crise da história do setor aéreo internacional (RODRIGUES, 2020). O setor no Brasil também foi fortemente impactado, o governo brasileiro impôs diversas medidas de distanciamento social que impactaram diretamente a aviação comercial. A Figura 1 se e refere ao número de passageiros pagos que embarcaram e desembarcaram, em voos nacionais, segmentado por anos, entre 2019 e 2022, na parte superior da Figura. Na parte inferior são apresentadas as variações percentuais em relação ao ano anterior.

FIGURA 1 – Passageiros Pagos: embarques + desembarques



Fonte: ANAC, Consulta Interativa – Indicadores do Mercado de Transporte Aéreo (ANAC, 2022).

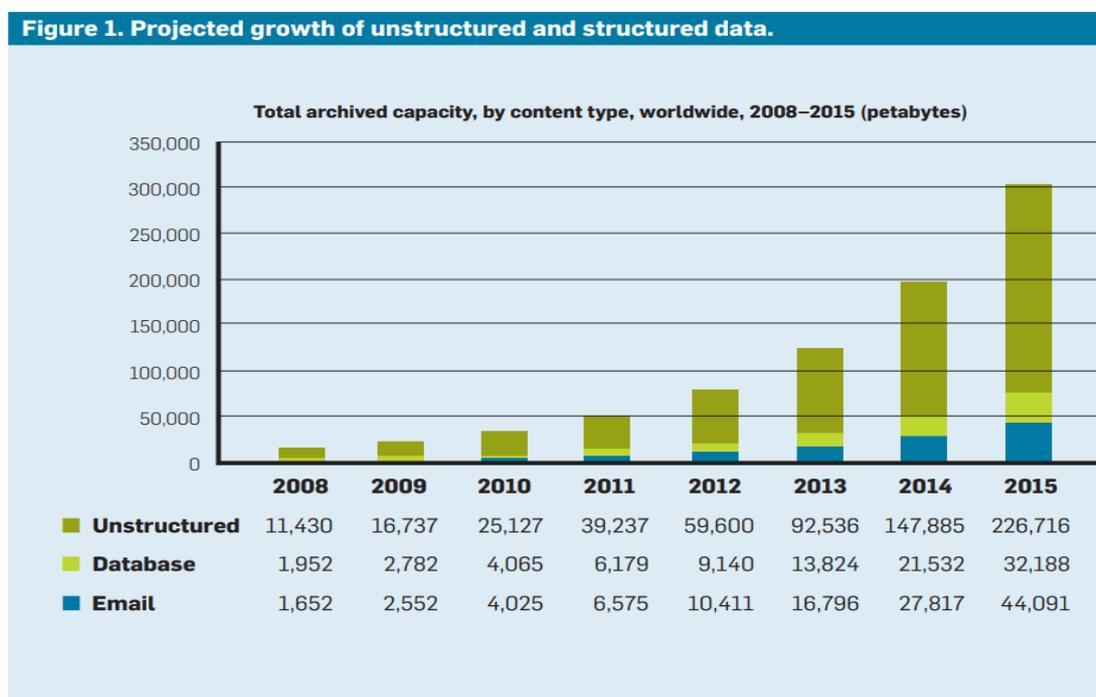
Nota-se que de 2019 para 2020 temos uma forte queda de 52,4% no número de voos, é consenso que a redução se deu em razão das medidas de restrição impostas pela pandemia. No ano de 2021 é possível notar uma significativa retomada, porém com resultados ainda distantes do período pré-pandemia. O ano de 2022 apresenta valor menor que o de 2021, pois é o ano corrente que ainda não se findou, portanto ainda não existe um valor total do número de voos, entretanto cabe destacar que o ano de 2022 já supera o ano de 2019, mesmo antes de seu compilado total. Demonstrando um processo de recuperação, impulsionado pela distribuição da vacina que gerou uma gradual retomada.

### 2.3 Ciência de Dados

Definir a Ciência de Dados é um desafio, Vasant Dhar aponta em seu artigo *Data Science and Prediction* que o termo ciência remete a construção de conhecimento através de estudos sistemático, portanto Ciência de Dados implica um foco científico no estudo dos dados (DHAR, 2013). Esse ramo de estudo tem se fortalecido em importância e volume, no Brasil, em 2019, o percentual de brasileiros com mais de 10 anos de idade que utilizam a internet chegou a 78,3%, um aumento de mais de 3 pontos percentuais em relação ao ano anterior, segundo dados da Pesquisa Nacional por Amostra de Domicílios Contínua de 2019 (PNAD 2019), realizada pelo Instituto Brasileiro de Geografia e Estatística – IBGE.

A capacidade da humanidade de armazenar informações também se desenvolveu. O volume de dados armazenados disparou em todo o globo, subsidiado por tecnologias mais baratas e eficientes de armazenamento, associadas ao citado aumento do uso de internet. Como mostra a Figura 2, retirada do artigo *Data Science and Prediction* (DHAR, 2013), que apresenta o volume projetado de dados estruturados e não estruturados entre 2008 e 2015, é possível notar um forte crescimento principalmente no volume de dados não estruturados. Se temos tal crescimento é lógico esperar que a necessidade de profissionais e ferramentas capazes de extrair, estruturar e gerar informações úteis a partir desses dados também tenha aumentado ao longo dos anos.

FIGURA 2 – Crescimento projetado de dados estruturados e não estruturado



Fonte: Artigo Data Science and Prediction (DHAR, 2013).

Esse contexto de maior relevância dos dados na sociedade evoca o conceito de Data Science (DS) ou ciência de dados, o campo interdisciplinar que estuda a extração de informações e conhecimento a partir de dados. A ciência de dados se vale métodos estatísticos, matemáticos e da ciência da computação na busca desses insights.

A ciência de dados é um campo de estudo multidisciplinar, no que aceita ferramentas advindas de diversos outros campos da ciência e no que também fornece diversas contribuições para outras áreas. A articulação da ciência de dados com os demais campos da ciência se mostra

promissor, ao passo que suas ferramentas tem a capacidade de impulsionar pesquisas e estudos que envolvam dados, enquanto expandem seu próprio campo do saber.

## 2.4 Análise Exploratória de Dados

Neste trabalho serão utilizados métodos estatísticos, mais especificamente a Análise Exploratória de Dados (AED), cuja finalidade “é verificar dados para qualquer aplicação estatística. Com isso, é possível obter entendimento sobre os dados coletados e, principalmente, sobre as relações existentes entre as variáveis analisadas” (FERREIRA et al., 2021, p 15).

Segundo Ferreira et al. (2021), na AED a estratégia utilizada segue a seguinte ordem: formular um problema, obter os dados, realizar a análise dos dados e por fim indicar um modelo. Diferente de outras estratégias como a clássica e a bayesiana que primeiro geram um modelo para posteriormente realizam a análise dos dados. A Tabela 1. abaixo apresenta uma síntese das três estratégias utilizadas em distintas abordagens estatísticas.

TABELA 1 – Abordagens estatísticas e estratégias utilizadas

Quadro 1. Abordagens estatísticas e estratégias utilizadas	
Abordagem	Estratégia
Estatística clássica	Problema → Dados → Modelo → Análise
Estatística bayesiana	Problema → Dados → Modelo Priori → Análise
AED	Problema → Dados → Análise → Modelo

Fonte: Adaptado de Capítulo... ([2011]).

Fonte: Preparação e Análise Exploratória de Dados (FERREIRA et al., 2021, p 15)

Em uma EAD os dados são considerados como variáveis, por exemplo peso, altura e idade. Essas variáveis podem representar diferentes valores, como numéricos e valores não numéricos. As variáveis do tipo numérico, são conhecidas como valores quantitativos e demandam um tratamento diferente das variáveis do tipo não numérico, também chamadas de valores qualitativos. No contexto estatístico essas variáveis são chamadas de variáveis estatísticas quantitativas e qualitativas.

As variáveis estatísticas quantitativas podem ser de dois tipos: discretas ou contínuas. As variáveis estatísticas discretas são aquelas que só podem tomar um número inteiro de valores, por isso também são chamadas de valores discretos, por exemplo idade, número de filhos, quantidade de produtos em uma prateleira. Já as variáveis estatísticas contínuas podem tomar um valor numérico ou real em qualquer intervalo, a exemplo a altura de um indivíduo, seu peso, nível de colesterol, entre outros (FERREIRA et al., 2021, p 43).

Variáveis estatísticas qualitativas podem também podem ser de duas categorias: nominais ou ordinais. São nominais quando suas categorias de dados não seguem uma ordem, por exemplo valores relacionados ao nome, cor e identificação de um indivíduo. São do tipo ordinais quando suas categorias são valores que podem ser ordenados ou ranqueados como a patente em uma organização militar (soldado, cabo, sargento) ou classe social (baixa, média e alta), sendo chamadas de variáveis estatísticas qualitativas ordinais (FERREIRA et al., 2021, p 43).

Perceba que os valores qualitativos, tanto nominais quanto ordinais, não permite sua utilização em operações matemáticas ou estatísticas, portanto não é possível somar o nome de duas pessoas ou calcular a média de um grupo de cores. Por outro lado, valores do tipo numérico ou quantitativos permitem sua utilização em tais operações, seria possível por exemplo realizar o cálculo da média de idade entre duas pessoas ou somar o peso de um grupo de indivíduos.

Essa distinção entre os tipos de variáveis estatísticas é especialmente relevante ao tratarmos de estatística descritiva, um componente essencial a AED. A estatística descritiva consiste em interpretar os dados, visualizá-los e por fim relacioná-los com as variáveis (FERREIRA et al., 2021, p 17). Nesse sentido são aplicados cálculos estatísticos como média amostral, desvio padrão, variância, entres outros a depender da amostra e dos objetivos da análise. A visualização conjunta e a interpretação dessas medidas fornecem conhecimentos acerca das características do conjunto de dados. A AED pode ser dividida em diversas etapas, geralmente: coleta, organização, tratamento, análise, apresentação e interpretação dos dados. Todas essas etapas, com exceção da coleta, utilizam métodos da estatística descritiva, o que explicita sua forte relação com a AED.

O exposto corrobora o apontamento de Gabriel Stankevix (apud FERREIRA et al., 2021, p 17) no que defende “a análise exploratória dos dados é considerada a principal etapa da análise dos dados. Ela inclui um processo de organizar, resumir, aplicar alguns cálculos e resumir as informações em um formato adequado à visualização.”. Portanto a AED, subsidiada pela estatística descritiva, é um componente capaz de trazer importantes contribuições em um processo de análise de dados. Além de garantir a qualidade dos dados.

## 2.5 Bibliotecas do Python

A extração dos dados pode ser construída de acordo com a forma na qual os dados de interesse são capturados de um banco de dados, site ou fonte, em um formato que possibilite a manipulação e extração de informações. Podendo ser disponibilizados de diversas maneiras e formatos, em muitos casos é interessante desenvolver uma ferramenta capaz de realizar a extração e interpretação dessas informações. A linguagem de Programação Python é uma das mais utilizadas no mundo para o desenvolvimento deste tipo de ferramenta. No contexto deste trabalho foram utilizadas diversas bibliotecas que possibilitaram a realização das análises de maneira mais simples e rápida, as principais foram Selenium, Pandas, Matplotlib e Seaborn.

A biblioteca Selenium foi utilizada na criação da função de download automático dos dados disponibilizados no site da ANAC. O Selenium é uma biblioteca com diversos métodos e funções de integração e automação web, portanto extremamente utilizado para construir a comunicação entre as aplicações em Python e páginas web.

Pandas é uma biblioteca Python que oferece estruturas de dados de alto nível e funções, projetadas para fazer com que trabalhar com dados estruturados ou tabulares seja rápido, fácil e expressivo. (MCKINNEY, 2018, cap 1.3). Extremamente útil para realização de análises de dados de formas simples e rápida, o Pandas é adequado para diferentes tipos de dados como: dados tabulares, dados de séries temporais ordenados e não ordenados, matriz arbitrária e quase qualquer tipo de dados estatísticos (FERREIRA et al., 2021, p 28).

A biblioteca Matplotlib é uma das mais populares ferramentas de visualização de dados do mundo, desenvolvida por John Hunter com a contribuição de diversos desenvolvedores. Sua função principal é a de plotagem de gráficos estatísticos, de diversos tipos, de maneira simples e rápida. A biblioteca Seaborn complementa a Matplotlib, no sentido em que foi construída com base nela, e também é integrada às estruturas de dados do Pandas. Portanto as três bibliotecas podem ser utilizadas de maneira complementar para gerar análises mais completas e visualmente eficientes.

### 3. MÉTODOS E TÉCNICAS DE PESQUISA

#### 3.1. Tipologia e descrição geral dos métodos de pesquisa

Este trabalho é uma pesquisa quantitativa de abordagem pura, elaborada com base no paradigma Design Science (DS), utilizando a metodologia Design Science Research (DSR). Conforme Dresch, Lacerda e Cauchick-Miguel (2019, v. 1, p. 109) “A DS/DSR pode ser aplicada para a formalização, prescrição e o projeto de soluções satisfatórias para as organizações”.

O Design Science é assentado na natureza prática de solução de problemas, como afirma Simon (DRESCH et al., 2019), seus fundamentos são: i) natureza da pesquisa; ii) os artefatos produzidos; iii) as soluções satisfatórias geradas; iv) a validade pragmática da solução proposta; e v) as classes de problemas que organizam a trajetória do conhecimento no âmbito da Design Science. Para cumprir sua incumbência de resolução de problemas a Design Science se vale da criação de artefatos, que são definidos como uma interface entre um ambiente interno e um ambiente externo e são projetados para atender a um determinado propósito (DRESCH et al., 2015). Recebem a classificação de: constructos, modelos, métodos e instanciações, na Tabela 2 são apresentadas as definições de cada tipo de artefato.

TABELA 2 - Tipos de Artefatos

Tipo de artefato	Definição	Exemplo de artefato desenvolvido	Referências dos exemplos
Constructo <sup>1</sup>	Conceitos utilizados para descrever problemas ou especificar as respectivas soluções	Constructos para descrever os fluxos de informação de um processo de negócio	<a href="#">Rosenkranz e Holten (2011)</a>
Modelo	Conjunto de elementos e relações que representam a estrutura geral da realidade	Modelo de controle integrado da produção e da qualidade para construção civil	<a href="#">Leão, Isatto e Formoso (2016)</a>
Método	Conjunto de passos lógicos necessários para a efetivação de determinada atividade	Método para resolver desafios concernentes à arquitetura organizacional	<a href="#">Nakakawa, Van Bommel e Proper (2011)</a>
Instanciação	Execução do(s) artefato (s) em seu ambiente real, evidenciando a viabilidade e eficácia dos artefatos	Instanciação de um método para desenvolver habilidades de resiliência em eletricitistas	<a href="#">Saurin et al. (2014)</a>
<i>Design Proposition</i>	Regras tecnológicas ou regras de projeto, consideradas contribuições teóricas da <i>Design Science</i> .	<i>Design proposition</i> para uma abordagem de inovação aberta para micro e pequenas empresas	<a href="#">Krause e Schutte (2016)</a>

Fonte: Metodologia Científica para Engenharia, capítulo 5 (Dresch et al., 2019).

#### 3.2. Caracterização e descrição dos instrumentos de pesquisa

A Design Science Research possui 12 etapas para sua execução, de acordo com Dresch et al. (2015). Abaixo são apresentadas cada uma das etapas, com seus respectivos objetivos e breve descrição de como serão desenvolvidas como instrumento de pesquisa neste trabalho:

### 1. Identificação do problema:

Entender exatamente a problemática sobre a qual o trabalho será desenvolvido. Este trabalho busca sanar o problema da falta de informações úteis a tomada de decisão acerca das reclamações dos usuários da aviação civil doméstica do Brasil.

### 2. Conscientização do problema

Identificação dos principais elementos que contribuem para a existência do problema. Apesar da ANAC divulgar os dados acerca das reclamações, esses são disponibilizados em extensas tabelas no formato CSV e JSON. A capacidade computacional necessária para análise dos dados é substancial. Nesse sentido, apesar de ser um ponto positivo a divulgação dos dados, a extração, tratamento e análise sobre os dados se mostra necessária.

### 3. Revisão da literatura

Estudo do conteúdo já produzido sobre ou relacionado ao problema de pesquisa. No presente trabalho esse tópico foi desenvolvido durante a revisão teórica da literatura.

### 4. Identificação de artefatos e configuração das classes de problemas

Identificação dos artefatos criados pelos pesquisadores durante a revisão da literatura. Artefatos são artificialidades ou artifícios criados, ou seja, não naturais, para o atingimento de determinado objetivo do projeto. Os artefatos utilizados neste trabalho são métodos, modelos e instâncias como instrumentos de análise e tratamento de dados com base em estatística e linguagem de programação Python.

Este trabalho é caracterizado por instrumentos estatísticos aplicados através da linguagem de programação Python, trazendo uma abordagem moderna ao instrumento de análise. Os códigos desenvolvidos constituem um uso aplicado a análise de dados a linguagem de programação.

### 5. Proposição de artefatos para resolver o problema específico

Ainda com base na revisão da literatura, propor mecanismos não naturais para a resolução do problema de pesquisa. O mecanismo proposto neste trabalho é o desenvolvimento de um Notebook em Python capaz de realizar a extração e análise dos dados.

### 6. Projeto do artefato selecionado

Neste trabalho os artefatos escolhidos foram notebooks, criados na linguagem Python. Estes notebooks implementam um método e constituem-se em um modelo aplicado. A instanciação ocorre quando os dados são inseridos para resolução do modelo. O notebook foi desenvolvido para acessar o site da ANAC, onde são disponibilizados os dados, e fazer o download dos dados referentes aos anos de interesse, importando-os para diretório do Google

Colab. Com dados importados, foi criado um notebook para realizar a análise e o tratamento desses dados.

#### 7. Desenvolvimento do artefato

Os artefatos foram desenvolvidos na linguagem de programação Python, com a utilização dos pacotes Pandas, PuLP, Matplotlib, Seaborn e Selenium. Os códigos desenvolvidos encontram-se nos Apêndices.

#### 8. Avaliação do artefato

A etapa de avaliação do artefato busca mensurar a efetividade do artefato desenvolvido quanto a resolução do problema e objetivos do projeto. Espera-se que o artefato seja capaz de atender aos objetivos gerais e possivelmente os específicos do trabalho.

#### 9. Explicitação das aprendizagens

Os aprendizados conquistados com a aplicação do artefato devem ser explicados e documentados, de forma a contribuir com a comunidade acadêmica e a sociedade.

#### 10. Conclusões

Na etapa de conclusão são apresentadas as contribuições do estudo, o artefato desenvolvido, a desempenho do artefato e os resultados obtidos. Cabe também destacar as limitações do estudo e propor sugestões para futuras pesquisas que possam desenvolver ainda mais o tema.

#### 11. Generalização para uma classe de problemas

Se cabível, deve se apresentar as possíveis generalizações do artefato para a solução de problemas de mesma natureza, ou semelhante, a apresentada no trabalho. Tornando o artefato um instrumento útil a sociedade e um facilitador em futuros estudos sobre o tema.

#### 12. Comunicação dos resultados

Comunicar os resultados obtidos para a comunidade acadêmica e para profissionais que possuam relação com o problema que foi desenvolvido. É importante a divulgação deste trabalho em canais que facilitem o acesso, além do uso de uma linguagem, que sem prejuízo a qualidade técnica, permita a compreensão do maior número de leitores, contribuído para a democratização da ciência.

### **3.3 Caracterização da Organização, setor e área**

O setor a ser analisado neste estudo é setor de aviação civil brasileiro. As informações

analisadas se referem as reclamações dos passageiros acerca dos voos domésticos, com base nos dados coletados pela Agência Nacional de Aviação Civil através da plataforma Consumidor.gov.br.

### **3.3.1 Agência Nacional de Aviação Civil (ANAC)**

A Agência Nacional de Aviação Civil (ANAC) foi criada pela Lei nº 11.182, de 27 de setembro de 2005, a qual define a ANAC como entidade integrante da Administração Pública Federal indireta, submetida a regime autárquico especial, vinculada ao Ministério da Defesa, Em seu Art. 4º caracteriza que a Agência possui independência administrativa, autonomia financeira, ausência de subordinação hierárquica e mandato fixo de seus dirigentes. Destacando no artigo subsequente a ANAC como autoridade da aviação civil (BRASIL, 2005).

A lei de criação da ANAC define seus objetivos como os de regular e fiscalizar as atividades de aviação civil e de infraestrutura aeronáutica e aeroportuária (BRASIL, 2005). Portanto a instituição é agência reguladora que fiscaliza os voos domésticos no Brasil, para além a agência disponibiliza dados e relatórios anualmente sobre o cenário de aviação no país. Dados esses que serão utilizados como subsídio neste trabalho.

Cabe destacar que as atividades de investigação e prevenção de acidentes aéreos, proteção e controle de tráfego são de responsabilidade do Centro de Investigação e Prevenção de Acidentes Aeronáuticos (CENIPA) e do Departamento de Controle do Espaço Aéreo (DECEA).

Os dados estatísticos do transporte aéreo do Brasil são disponibilizados pela Gerência de Acompanhamento de Mercado (GEAC), que compõem a estrutura da Superintendência de Acompanhamento de Serviços Aéreos (SAS). A organização da SAS está descrita na Portaria da ANAC Nº 4.211, de 9 de fevereiro de 2021, que em seu Art. 9º delega a GEAC a responsabilidade do recebimento, fiscalização, tratamento e à disponibilização dos dados estatísticos de voos (ANAC, 2021).

### **3.3.2 Consumidor.gov.Br**

Cabe descrever a plataforma pela qual os usuários registram suas reclamações e o processo de registro para melhor compreensão do conjunto de dados. O Consumidor.gov.br de

acordo com o site de mesmo endereço “é um serviço público e gratuito que permite a interlocução direta entre consumidores e empresas para solução alternativa de conflitos de consumo pela internet”. A plataforma não substitui ou impede que os reclamantes recorram aos Órgãos oficiais de defesa do Consumidor. A principal inovação do Consumidor.gov.br está em possibilitar o contato direto entre consumidores e empresas, em um ambiente público, dispensada a intervenção do Poder Público na tratativa individual. Entretanto a participação das empresas é voluntária e se dar por um processo formal, de maneira que os consumidores só podem registrar sua reclamação caso a empresa esteja registrada. No setor de aviação a ANAC optou pelo uso da plataforma para registro das reclamações e as três principais empresas de aviação do Brasil, Azul Linhas Aéreas, Latam Airlines e Gol Linhas Aéreas, além de diversas outras.

O processo de registro pode ser realizado através do site Consumidor.gov.br ou aplicativo disponível para Android e IOS, após o registro a empresa tem 10 dias para analisar e responder a reclamação, a partir da data de reposta o reclamante tem até 20 dias para comentar e avaliar a resposta da empresa, informando se sua reclamação foi resolvida ou não resolvida, e ainda indicar seu nível de satisfação com o atendimento recebido, dando uma nota entre de 1 a 5. Ainda de acordo com a plataforma “Os dados das reclamações registradas no Consumidor.gov.br alimentam uma base de dados pública, com informações sobre as empresas que obtiveram os melhores índices de solução e satisfação no tratamento das reclamações”, essa base é utilizada como fonte de dados neste trabalho.

### **3.4 População e amostra da pesquisa**

A população analisada são todas reclamações realizadas por passageiras do setor de transporte aéreo doméstico de passageiros registradas na plataforma Consumidor.gov.br. O conjunto de todas reclamações configura imenso grupo de dados que precisa ser organizado e analisado para que gere informações úteis a tomada de decisão. Portanto para a análise dos dados foram escolhidos os anos mais recentes disponibilizados pela ANAC, 2019, 2020 e 2021, afim de se obter uma amostra de conveniência mais recente. Esse período também foi escolhido por contemplar o ano anterior a pandemia mundial de covid-19, 2019, e os dois anos em que pandemia esteve em seu auge, 2020 e 2021, possibilitando considerar se a crise gerou impactos nas reclamações do setor.

Os dados foram obtidos através do sítio eletrônico da ANAC, e são descritos pela própria Agência como:

“Os dados relativos às reclamações registradas na plataforma Consumidor.gov.br (www.consumidor.gov.br) contabilizam os atendimentos realizados pelas empresas aéreas em atenção às demandas dos passageiros.” (ANAC, 2020)

A ANAC disponibiliza quadro que descreve os dados que são objeto desse estudo 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>>. Abaixo segue Tabela 3 que trata da quantidade de voos e de passageiros transportados no mercado doméstico nos anos de interesse.

TABELA 3 - Quantidade de Voos domésticos x Quantidade de passageiros

Ano	2019	2020	2021
Quantidade de Voos domésticos	802.400	406.335	546.580
Quantidade de Passageiros transportados	95.049.029	45.249.089	62.583.158

Fonte: Dados retirados do Anuário do Transporte Aéreo de 2021 - Planilhário do Transporte Aéreo – Gráficos e Tabelas – ANAC.

É notável a redução de quase 50%, em 2020 em relação a 2019, tanto na quantidade de voos como no número de passageiros transportados. Uma abordagem empírica sugere que a diminuição se deve às restrições causadas pela pandemia de covid-19 que impactou fortemente o setor de transportes aéreos em 2020. Em 2021 houve um aumento de 34,5% no número de voos domésticos realizados, também a observação empírica sugere que esse aumento se deu em função do início da vacinação contra o coronavírus e consequente relaxamento das medidas de restrição. Entretanto quando comparamos o número de voos domésticos em 2021 com o período inicial da pandemia, 2019, a quantidade de voos ainda é 46,68% menor, demonstrando que em 2021 o mercado de aviação começou a se recuperar, porém ainda distante dos números antes da pandemia.

### 3.5 Procedimento de coleta de dados

A coleta de dados se deu através do sítio eletrônico da ANAC, na área de Dados do Consumidor.gov. Os dados estão disponíveis no formato csv e json, ficando a critério do usuário a escolha do formato para download.

Foi desenvolvido um notebook em Python para realizar o download dos dados e importá-los para o diretório do Google Colab (Apêndice A). No código é possível definir os anos de interesse para a realização do download, através do link <<https://sistemas.anac.gov.br/dadosabertos/Voos%20e%20opera%C3%A7%C3%B5es%20a%C3%A9reas/Dados%20do%20consumidor.gov>> considerando que anos de interesse deste trabalho são de 2019 à 2021, as planilhas de interesse formam baixadas. A partir do conjunto de dados importados foi possível realizar análises utilizando a linguagem Python.

### **3.6 Análise de dados**

A partir do conjunto de dados foram realizadas as análises utilizando a linguagem de programação Python e posteriormente apresentadas as informações obtidas em formato de tabela ou gráfico para facilitar a visualização. Pela facilidade de edição de gráficos de setores, comumente chamados de gráficos de pizza, no Microsoft Excel, alguns dados, após analisados no Python, foram transferidos para o Excel apenas para a construção do gráfico. Importante reiterar que apenas os gráficos de pizza foram construídos no Excel, os demais gráficos foram todos originados através do Python. Para possibilitar as análises e apresentações gráficas foram utilizadas as bibliotecas Pandas, Seaborn e Matplotlib. Para orientar a análise dos dados foram criadas algumas perguntas a serem respondidas:

- Quais estados do Brasil apresentam o maior número de reclamações?
- Qual o prazo médio de reposta por parte das empresas acerca das reclamações?
- Existe um aumento no número de reclamações nos períodos de alta temporada?
- Qual a porcentagem de reclamações resolvidas de acordo com a classificação dos consumidores?
- Qual a nota média dada pelos consumidores para a resolução das reclamações?

A criação de perguntas é importante para fornecer objetivos que possam orientar a análises dos dados, sem prejuízo a análises complementares que possam trazer outras informações úteis não contempladas pelas perguntas, mas relevantes ao contexto do trabalho.

## 4. RESULTADO E DISCUSSÃO

### 4.1. Avaliação Inicial e limpeza dos dados (*Data Cleaning*)

A primeira parte deste tópico se refere a avaliação inicial dos dados, para o contexto deste trabalho consiste no entendimento acerca do conjunto de dados e sua estrutura, portanto é apresentado a descrição e a concatenação dos dados referentes aos três anos que são escopo desse trabalho. A segunda parte se dá com a limpeza de dados, que seria o tratamento inicial dado ao conjunto de dados para eliminar inconsistências que possam vir a atrapalhar as demais etapas da análise exploratória. Nesse sentido a limpeza seguiu a seguinte ordem: verificação acerca da existência de registros incompletos ou duplicados, exclusão de colunas irrelevantes, avaliação do tipo das variáveis e busca por outliers.

Uma avaliação inicial dos dados demonstrou que as planilhas de cada ano possuíam a mesma estrutura, com 38 colunas. A descrição de cada coluna, como informado no tópico 3.3 está disponível no sitio da ANAC e é apresentada também no Tabela 4 – Metadados, na coluna “campo” é apresentado com nome de cada coluna e na coluna “Descrição” sua respectiva descrição:

TABELA 4 - Metadados

(Continua)

Campo	Descrição
Gestor	Nome da entidade de defesa do consumidor, responsável pela gestão da reclamação do consumidor.
Canal de Origem	Origem do registro no Consumidor.gov.br.
Região	Sigla da região geográfica do consumidor.
UF	Sigla do estado do consumidor.

TABELA 4 – Metadados (Cont.)

(Continuação)

<b>Campo</b>	<b>Descrição</b>
Cidade	Município do consumidor.
Sexo	Sigla do sexo do consumidor.
Faixa Etária	Faixa etária 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.
Hora Abertura	Hora de abertura da reclamação pelo consumidor.
Data Resposta	Data de resposta da reclamação pela empresa.
Hora Resposta	Hora de resposta da reclamação pela empresa.
Data Análise	Data da análise da recusa da empresa pelo Gestor da reclamação.
Hora Análise	Hora da análise da recusa da empresa pelo Gestor da reclamação.
Data Recusa	Data de registro da recusa pela empresa e envio da reclamação para análise do Gestor.
Hora Recusa	Hora de registro da recusa pela empresa e envio da reclamação para análise do Gestor.
Data Finalização	Data de finalização da reclamação.
Hora Finalização	Hora 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	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.

TABELA 4 – Metadados (Cont.)

(Continuação)

<b>Campo</b>	<b>Descrição</b>
Tempo Resposta	Número de dias para a resposta da reclamação, entre a Data de Resposta e a Data de Abertura, desconsiderado 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.
Segmento de Mercado	Principal segmento de mercado da empresa participante.
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.
Como Comprou Contratou	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?”.
Respondida	Sigla que indica se a empresa respondeu à reclamação ou não.
Situação	Situação atual da reclamação no sistema.
Avaliação Reclamação	Classificação atribuída pelo consumidor sobre o desfecho da reclamação.
Nota do Consumidor	Número da nota de 1 a 5 atribuída pelo consumidor ao atendimento da empresa.
Análise da Recusa	Resultado da Análise da Recusa pelo Gestor, se procedente (aceita) ou improcedente (indeferida ou não aceita).
Edição de Conteúdo	Sigla que indica se houve ou não edição de conteúdo da reclamação pelo Órgão Gestor.

TABELA 4 – Metadados (Cont.)

(Conclusão)

Campo	Descrição
Interação do Gestor	Sigla que indica se houve ou não interação do Gestor com a empresa ou com o consumidor ao longo da reclamação.
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 <a href="#">Portaria nº 867/SAS, de 21 de março de 2019</a> .

Fonte: ANAC, Metadados do conjunto de dados: Dados do consumidor.gov (ANAC, 2022).

Na coluna “Campo” é apresentado o nome da coluna dentro das planilhas disponibilizadas e na coluna “Descrição” da respectiva coluna, portanto informa que tipo de informação está sendo armazenada, essa tabela é uma importante fonte para a compreensão dos dados. Além das 38 colunas as planilhas, principalmente após serem concatenadas, possuem extenso número de linhas, mais de 200 mil, que apresentam diversos detalhes acerca de cada reclamação. Uma tabela com tantas células demanda grande capacidade de processamento computacional, portanto realizar o tratamento e análise dos dados utilizando Python se mostra uma alternativa frente a softwares como Excel ou Google Sheet, que teriam grande dificuldade de processar operações com esse volume de dados.

A primeira etapa da avaliação inicial dos dados foi concatenar as planilhas referentes a cada ano em uma única, de maneira a possibilitar análises referentes aos três anos: 2019, 2020 e 2021, de maneira a formar um único conjunto de dados. A segunda etapa foi verificar se haviam registros incompletos, ou seja, com dados em branco, foram encontrados diversos registros incompletos que geraram informações relevantes, que estão apresentadas no tópico 4.5 Registros Incompletos. A terceira etapa foi verificar a existência de registros duplicados, o que não foi encontrado, atestando a consistência das planilhas.

A etapa seguinte foi excluir colunas não relevantes para as análises pretendidas. As colunas excluídas foram: `AnáliseDaRecusa`, `EdiçãoDeConteúdo`, `InteraçãoDoGestor`, `CódigoClassificadorANAC`, `Cidade`, `HoraAbertura`, `HoraAnálise`, `HoraFinalização`, `InteraçãoDoGestor`, `CódigoClassificadorANAC`, `EdiçãoDeConteúdo`. O novo número de colunas é na ordem de 30.

Para avaliar quais dados eram passíveis de análises estatísticas, utilizando o Python foi verificado o tipo atribuído a cada variável, apresentado a seguir na Tabela 5:

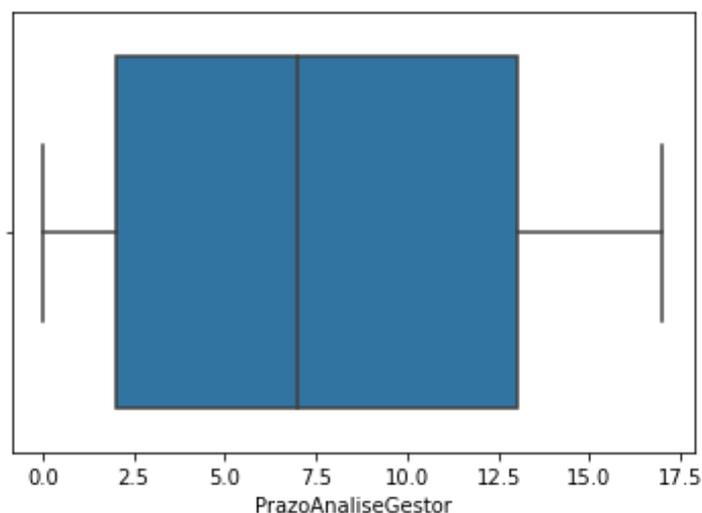
TABELA 5 - Tipos de variável no Python

<b>Nome da coluna</b>	<b>tipo de dado</b>
Gestor	object
CanalDeOrigem	object
Região	object
UF	object
Sexo	object
FaixaEtária	object
AnoAbertura	int64
MêsAbertura	int64
DataAbertura	object
DataResposta	object
HoraResposta	object
DataAnálise	object
DataRecusa	object
HoraRecusa	object
DataFinalização	object
PrazoResposta	object
PrazoAnáliseGestor	float64
TempoResposta	float64
NomeFantasia	object
SegmentoDeMercado	object
Área	object
Assunto	object
GrupoProblema	object
Problema	object
ComoComprouContratou	object
ProcurouEmpresa	object
Respondida	object
Situação	object
AvaliaçãoReclamação	object
NotaDoConsumidor	float64

Portanto as colunas que possuem dados que permitem operações matemáticas e estatísticas são: AnoAbertura, MêsAbertura, PrazoAnáliseGestor, TempoResposta e NotaDoConsumidor. A partir da identificação das variáveis numéricas é importante analisar se existem outliers no conjunto de dados. Contudo, não faz sentido buscar outliers nas colunas AnoAbertura e MêsAbertura, pois essas registram datas, também não é relevante buscar nos dados da coluna NotaDoConsumidor, afinal esta registra notas limitadas pelo intervalo entre 1 e 5. As duas colunas, portanto, em que existe sentido em realizar a análise de outliers são

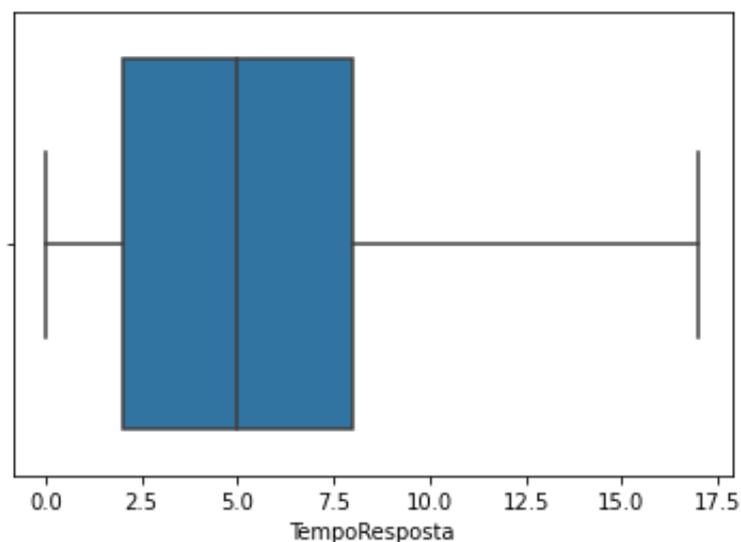
PrazoAnaliseGestor e TempoResposta. Para tal foi criado um gráfico boxplot utilizando a biblioteca Seaborn, apresentado a seguir:

FIGURA 3 - Prazo para Análise do Gestor



A Figura 3 demonstra que não existem outliers na coluna PrazoAnaliseGestor, isso significa que os órgãos gestores das reclamações quando provocados a analisarem a recusa de uma reclamação o fazem no prazo de até 17 dias, não tendo havido casos que superaram esse prazo entre 2019 e 2021. Já as reclamações não recusadas são respondidas pelas empresas, e para avaliar o se existiam outliers acerca desse prazo foi criado o gráfico a seguir:

FIGURA 4 - Tempo de Resposta

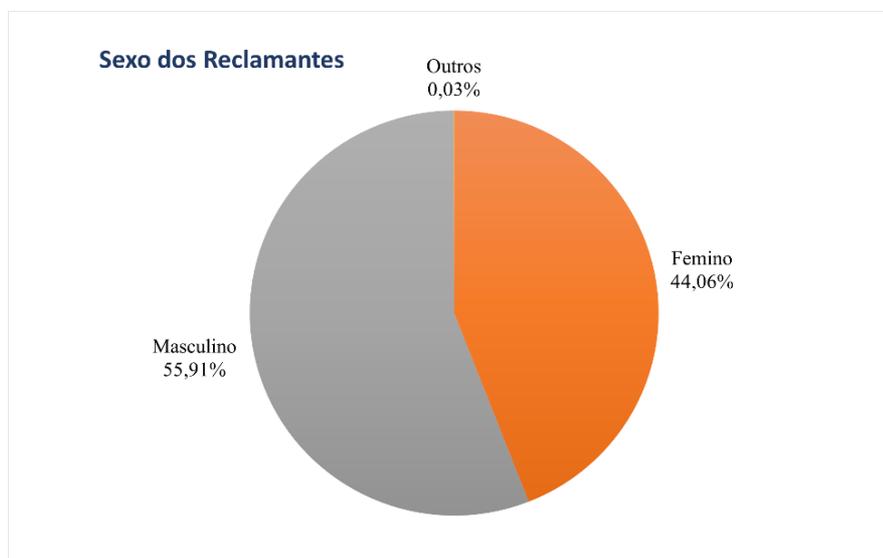


A Figura 4 se refere ao tempo de Resposta também não apresenta outliers, o prazo máximo de reposta para uma reclamação fica em torno de 17 dias, e o tempo médio em 5 dias. Outro aspecto interessante é que o primeiro e terceiro quartis estão muito mais próximos do limite inferior do que do limite superior, e o valor máximo do terceiro quartil é inferior a 8, apontando que 75% das repostas as reclamações são dadas em até 8 dias.

#### 4.2. Perfil dos reclamantes

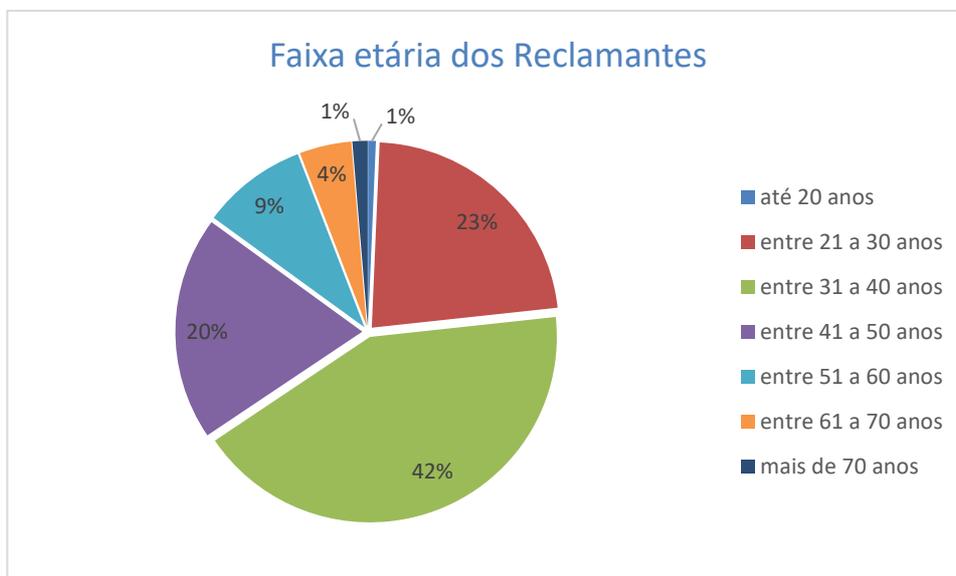
A segunda etapa da análise buscou entender o perfil dos reclamantes, informação especialmente interessante para que as empresas possam compreender quais públicos demandam maior atenção em relação aos serviços prestados. Infelizmente o conjunto de dados objeto desse estudo possui apenas duas variáveis relacionadas ao perfil dos reclamantes, sexo e faixa etária de idade, portanto não é possível traçar um perfil mais completo dos indivíduos. Em relação ao sexo dos reclamantes as porcentagens em relação ao total de reclamações são apresentadas a seguir na Figura 5:

FIGURA 5 – Sexo dos Reclamantes



A Figura 5 demonstra que o sexo dos reclamantes é dividido em três categorias: masculino, feminino e outros. Um pouco mais da metade dos indivíduos é do sexo masculino, enquanto menos de 1% pertencem ao grupo outros. Já em relação a idade dos consumidores são apresentadas suas faixas etárias abaixo, na Figura 6:

FIGURA 6 – Faixa Etária dos Reclamantes



A faixa etária dos indivíduos é dividida em classes como apresentado na legenda do Figura 6. Cabe destacar que cerca de 85% dos reclamantes possuem entre 21 e 50 anos, com maior concentração para aqueles que possuem entre 31 e 40 anos. Logo o perfil dos reclamantes é majoritariamente formado por homens com idade entre 31 e 40 anos.

#### 4.3. Registro de Reclamações por Empresas

De acordo com os dados do Anuário do Transporte Aéreo de 2021, disponibilizado pela ANAC, as três empresas com maior participação no mercado doméstico de passageiros são respectivamente: Azul Linhas Aéreas (41,7% de participação no número de voos), Latam Airlines (26,1% de participação no número de voos), Gol Linhas Aéreas (24,4% de participação no número de voos). Essas também são as empresas que registram o maior número de reclamações no intervalo analisado neste trabalho. A empresa que mais recebeu reclamações foi a Latam Airlines, com 75.785 registros, seguida pela Gol Linhas Aéreas com 65.934 e em terceiro lugar a Azul Linhas Aéreas com 30.236 reclamações. Cabe destacar que a Azul Linhas Aéreas apesar de ser a empresa que possui maior participação no mercado brasileiro é apenas a terceira colocada em número de reclamações e com menos da metade de registros da segunda

colocada a Gol Linhas Aéreas. As demais reclamações foram direcionadas a diversas outras companhias que de maneira pulverizada.

#### 4.4. Quantidade de Reclamações Registradas

Para analisar a quantidade de reclamações, primeiro se analisou a quantidade de registros em cada arquivo referente os seus respectivos anos:

FIGURA 7 – Número de Reclamações por Ano



Em 2019 foram 49.866 reclamações, em 2020 foram registradas 66.956 e em 2021 chegou-se a 107.609. É notável o padrão de crescimento das reclamações ao longo dos anos, de 2019 para 2021 o aumento foi de 215,79%, portanto o número de reclamações mais que dobrou. A ANAC aponta em um artigo disponível em seu site com o seguinte título ‘ANAC divulga Boletim de Monitoramento do Consumidor.gov.br anual de 2021’ dois fatores que possivelmente contribuíram para o crescimento no número de reclamações: o aumento do uso do Consumidor.gov.br, como canal para registro de reclamações e a pandemia causada pelo vírus da covid-19.

O primeiro fator de acordo com a ANAC foi o aumento da divulgação do canal de reclamações disponibilizado para os passageiros, o que teria proporcionado maior número de registros, como fica claro no seguinte trecho do artigo “Ressalta-se ainda que, em certa medida,

a própria adoção e reforço do Consumidor.gov.br, como canal da ANAC para reclamações, tem contribuído para o conhecimento e crescimento da utilização da plataforma pelos consumidores nos últimos anos”. Ressalta-se, portanto, que esse fator não está ligado ao aumento no número de problemas enfrentados pelos passageiros, mas ao aumento do registro dessas intercorrências.

O segundo fator pode ser inferido a partir do trecho também retirado do mesmo artigo que aponta que “Os efeitos da pandemia da covid-19 permaneceram como desafio para o setor de transporte aéreo no ano de 2021, com demanda por voos e planejamento da malha aérea ainda impactados. Diante desse contexto, medidas emergenciais foram adotadas com o objetivo de mitigar os reflexos negativos sobre o setor. Portanto, na análise dos dados, deve-se levar em consideração tal cenário, com impactos nos índices de todas as empresas aéreas e no resultado geral.”. Apesar da dificuldade em precisar a proporção dos **impactos da pandemia no número de reclamações**, empiricamente é sabido que foi muito significativo, como fica claro no trecho supracitado. Sendo um forte indício que o número de reclamações apresentou o expressivo aumento por intercorrências causadas pela pandemia de covid-19. Outro argumento que reforça essa tese foi apresentado no tópico 2.2 Efeitos da Covid-19 no setor de aviação, de 2019 para 2020 houve uma forte queda de 52,4% no número de voos, em função das medidas de restrição impostas pela pandemia. Em períodos normais o esperado com a redução no número de voos seria também uma redução no número de reclamações, afinal haveria menos viajantes, porém os dados indicam que mesmo com a queda nos voos o número de reclamações aumentou consideravelmente.

Ademais o número total de reclamações registradas pelos Canais monitorados pela ANAC durante os três anos de interesse deste trabalho foi de 224.431, esse número reforça a necessidade computacional necessária para se realizarem análises acerca desse grande volume de dados.

#### 4.5. Registros Incompletos

A etapa de limpeza de dados demonstrou que alguns registros não estão completos e possuem algumas informações em branco. As colunas que possuíam informações em branco e o respectivo número de dados em branco são apresentados a seguir na Tabela 6:

TABELA 6 - Registros em branco na Base de dados

(Continua)

Colunas com registro de dados em branco	Nº de dados em branco
---	-----------------------

DataResposta	6.549
HoraResposta	6.549
DataAnálise	219.581
TABELA 6 - Registros em branco na Base de dados (Cont.)	
	(Conclusão)
HoraAnálise	<b>219.581</b>
DataRecusa	219.536
HoraRecusa	219.536
PrazoAnaliseGestor	219.581
TempoResposta	6.549
AvaliaçãoReclamação	2.337
NotaDoConsumidor	93.445
AnáliseDaRecusa	219.541
CódigoClassificadorANAC	2.338

A Tabela 6 permite retirar os primeiros insights acerca do registro de reclamações. A princípio faz sentido que as colunas: DataAnálise, HoraAnálise, DataRecusa, HoraRecusa, PrazoAnaliseGestor e AnáliseDaRecusa possuam muitos registros em branco, afinal esses dados só são preenchidos se reclamação registrada for julgada improcedente pela empresa.

As colunas DataResposta, HoraResposta, TempoResposta registram o mesmo número de dados nulos, 6.549, naturalmente, afinal se não foi preenchida a data de resposta e nem a hora de resposta, torna-se impossível calcular o prazo decorrido para o reclamante receber uma resposta acerca de sua reclamação. Portanto, essa análise apresenta um insight preocupante, pois avaliar o prazo de resposta de uma empresa é essencial para entender o nível de eficiências das empresas na resolução das reclamações. Ademais de acordo com a Resolução da ANAC Nº 400, de 13 de dezembro de 2016 “Art. 38. As informações solicitadas pelo usuário deverão ser prestadas imediatamente e suas reclamações resolvidas no prazo máximo de 10 (dez) dias a contar do registro, ressalvados os prazos específicos contidos nesta Resolução.”. Nesse sentido o registro dessas informações também é imprescindível para que a ANAC consiga realizar a fiscalização acerca do cumprimento da referida resolução.

Outro ponto relevante é o grande número de dados em branco na coluna NotaDoConsumidor, de acordo com o site da ANAC essa coluna se refere “Número da nota de 1 a 5 atribuída pelo consumidor ao atendimento da empresa.”. Depreende-se que 41,63% dos consumidores não avaliam o atendimento das empresas, informação importante para atestar o nível de satisfação de um consumidor quanto a resolução de um problema por ele enfrentado. Tal nível de satisfação tem especial relevância no processo de *service recovery*.

#### 4.6. Análise por Regiões

As planilhas disponibilizadas pela ANAC possuem informações acerca do local onde foi registrada a reclamação, porém como o número de cidades no Brasil é muito extenso, foi preferível realizar as análises apenas considerando os estados. Abaixo é apresentada a Tabela 7 que demonstra o número de reclamações registradas por estado no período entre 2019 e 2021.

TABELA 7 - Registro de Reclamações por Estados entre 2019 e 2021

<b>Estado</b>	<b>Nº de Reclamações Registradas</b>
SP	53818
MG	23918
DF	21078
RJ	20346
PR	15929
RS	11353
SC	11142
BA	9489
CE	7352
GO	6586
ES	6054
PE	6009
MT	3988
PA	3705
AM	3198
MA	2824
RN	2460
MS	2456
PB	2447
PI	1995
RO	1884
AL	1601
SE	1560
TO	1055
AC	860
RR	741
AP	583

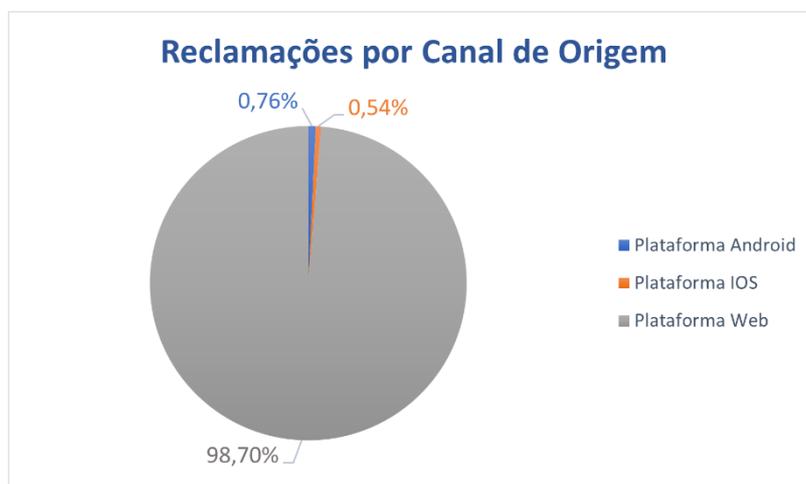
O estado que mais registrou reclamações foi São Paulo, com mais que o dobro de registros que o segundo colocado Minas Gerais. Essa informação tem sentido, principalmente, quando levamos em consideração que SP é o estado mais populoso do Brasil, com cerca de 46 milhões de habitantes, e possui mais moradores do que as regiões Norte, Sul e Centro-Oeste do Brasil juntas de acordo com a prévia do Censo Demográfico 2022 divulgado pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Ademais SP também abriga os dois aeroportos mais movimentados do Brasil, Guarulhos e Congonhas. Os dados continuam condizentes

quando avaliamos o segundo estado com o maior número de reclamações, Minas Gerais, que também é o segundo estado com o maior número de habitantes no país, com cerca de 20 milhões de pessoas. Para o terceiro estado que registra o maior número de reclamações, Distrito Federal, a aparente correlação entre densidade populacional e número de reclamações se enfraquece, pois, terceiro estado mais populoso do país é o Rio de Janeiro. Já o estado que registrou menos reclamações foi o Amapá, com apenas 583 registros, o que equivale apenas 1% dos registros do estado de São Paulo.

#### 4.7. Registro de Reclamações por Canal de Origem

O canal pelo qual o reclamante realizou o registro da reclamação é chamado de Canal de Origem, os canais são divididos em três. A Plataforma Android e Plataforma IOS, se referem aos canais de registro via aplicativo mobile “Consumidor.gov.br”, e existe também a Plataforma Web que se refere aos registros feitos através do site Consumidor.gov.br. A Figura 8 destacada abaixo apresenta a porcentagem de reclamações em cada canal:

FIGURA 8 – Reclamações por canal de origem

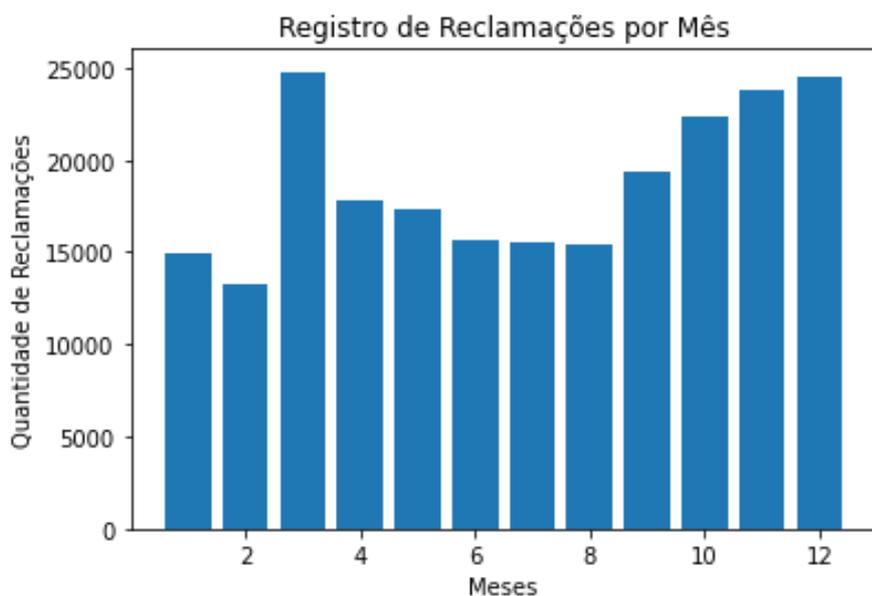


A grande maioria das reclamações é registrada através da Plataforma Web, 98,7%, o que demonstra que o uso do aplicativo Consumidor.gov.br ainda é muito baixo em comparação com o site.

#### 4.8. Períodos com o maior número de reclamações

Este tópico é dedicado a analisar os meses com maior número de reclamações, para tal foi desenvolvido através do Python a Figura 9, que demonstra a quantidade de reclamações registrada em cada mês.

FIGURA 9 – Registro de reclamações por mês

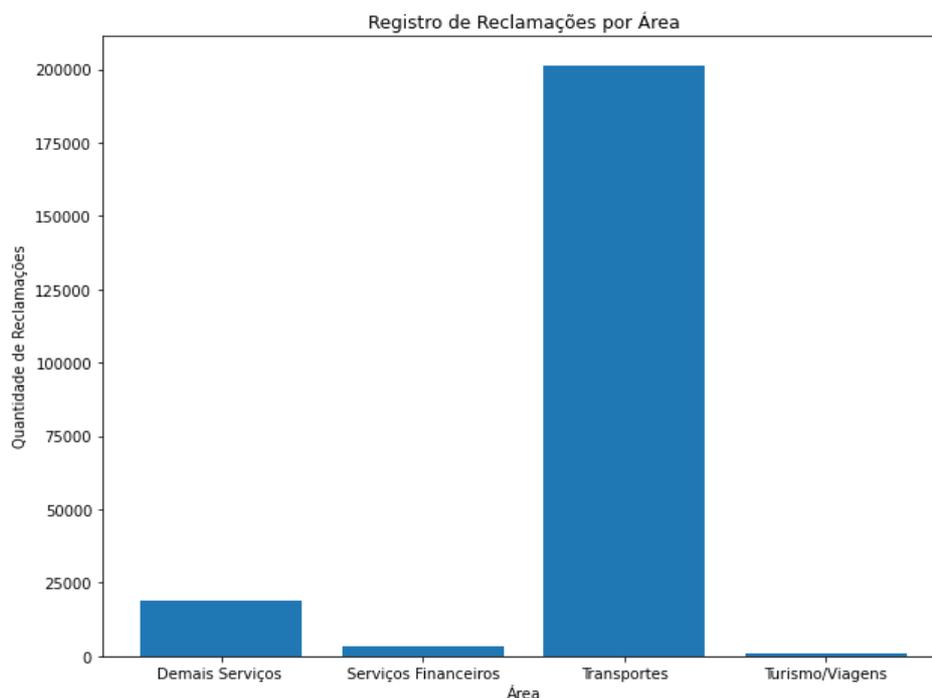


Observa-se que o mês com o maior número de reclamações é o mês de março, seguido por dezembro e novembro. A Figura 9 também demonstra que não existe um pico de reclamações nos meses de alta temporada, que são aqueles que coincidem o período de férias escolares no Brasil, historicamente janeiro, julho e dezembro.

#### 4.9. Registro de Reclamações por Classificação

Em nossa fonte de dados existem algumas classificações de reclamações, como pode ser visto na tabela 4, elas são divididas em: área, assunto, grupo e problema. A Figura 10 a seguir esclarece a quantidade de reclamações por área:

FIGURA 10 – Registro de Reclamações por Área



A Figura 10 apresenta as áreas e a quantidade de reclamações registrada em cada uma delas, como o conjunto de dados se refere ao setor de aviação civil é natural que a área transportes concentre o maior número de reclamações, aproximadamente 89,77%. Quanto ao Assunto das reclamações os dados são apresentados na tabela 8 para melhor visualização:

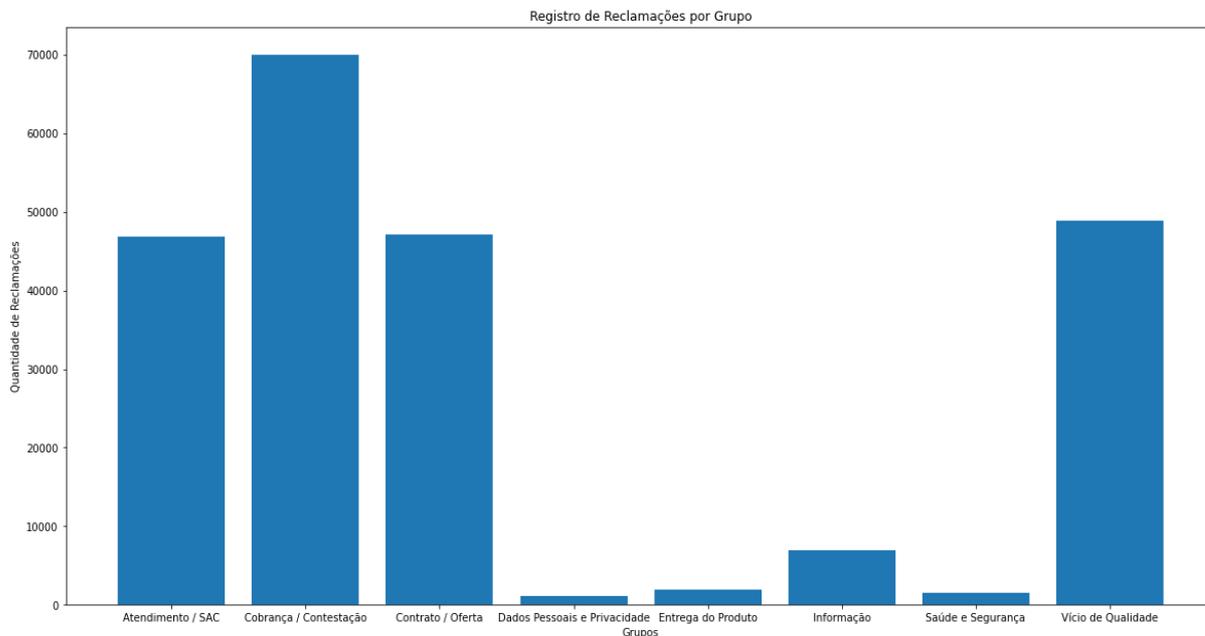
TABELA 8 - Número de Reclamações por Assunto

Assunto	Número de Reclamações	Porcentagem em relação ao total
Aéreo	199547	88,912%
Programas de Fidelidade / Benefícios (pontos, milhagem etc)	18887	8,416%
Seguro Viagem	3385	1,508%
Cargas e Encomendas	1938	0,864%
Pacote / excursão / agência de turismo	397	0,177%
Transfer / Receptivo	130	0,058%
Cultura, Esporte, Lazer e Eventos	119	0,053%
Hospedagem	25	0,011%
Locação/ Aluguel (carros, roupas,etc)	3	0,001%

Seguindo a lógica da área de reclamação faz sentido que o assunto “Aéreo” concentre o maior número de reclamações. Cabe destacar que o assunto “Programas de Fidelidade / Benefícios” é o segundo maior em número de reclamações, um insight principalmente para as

empresas do setor, que demonstra que além do próprio serviço de voo muitos clientes enfrentam problemas com os programas de fidelidade e benefícios. O terceiro tipo de classificação, grupo, de reclamações é analisado na Figura 11 a seguir:

FIGURA 11 – Registro de Reclamações por grupo



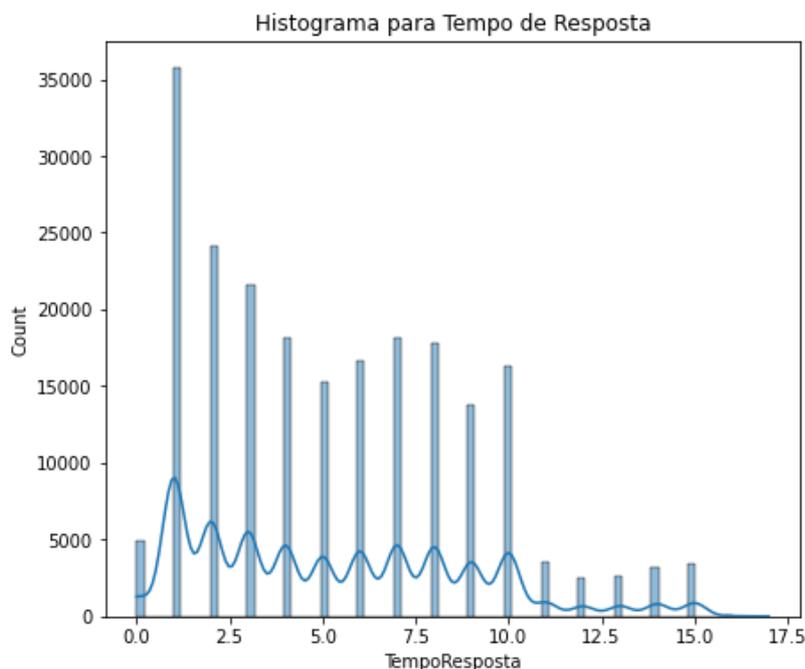
A Figura 11 referente as reclamações por grupo apresentam menor concentração que as classificações anteriormente analisadas. O grupo “Cobrança/Contestação” apresenta maior quantidade de registros, infere-se dos dados que a maioria dos problemas enfrentados pelos consumidores do setor de aviação civil está ligado a questões financeiras relacionadas ao transporte aéreo. O segundo grupo é “Vício de Qualidade”, ligado a problemas com próprio serviço de táxi aéreo, como atrasos e cancelamentos de voo. Seguido de perto por problemas de “Atendimento / SAC”. Os dados podem fornecer para as empresas de aviação insights acerca dos pontos que mais demandam melhorias no setor. O último grupo de classificação “problema” permite que cada cliente descreva seu problema específico, como são milhares de reclamações e com diversas especificidades, essa classe não será analisada.

#### 4.10. Tempo de Resposta

Foi calculado através do notebook em Python o tempo médio que as empresas levam para responder a uma reclamação, cabe lembrar que a proposta do Consumidor.gov.br é de que as empresas respondam em até 10 dias. Esse dado já havia sido apresentado no tópico

“Avaliação Inicial e limpeza dos dados (*Data Cleaning*)” quando analisamos a existência de outliers utilizando o gráfico boxplot que também apresenta a média do conjunto de dados. Entretanto a média foi calculada novamente para se obter um valor exato utilizando uma função específica do Python para cálculo da média, o resultado foi de 5,30, confirmando que o prazo médio de resposta é de 5 dias, portanto, dentro do prazo proposto. Havíamos concluído também que 75% das reclamações são respondidas em até 8 dias. Uma nova informação surgiu com a avaliação da moda desse conjunto, que é 1, portanto a maioria das reclamações é respondida em apenas 1 dia, mais especificamente 35.732, ou seja 15,92% dos registros. Para melhor visualização desse insight foi criado um histograma para avaliar a distribuição do conjunto de dados de tempo de resposta:

FIGURA 12 – Histograma para Tempo de Resposta



O histograma mostra que esse conjunto de dados não segue uma distribuição normal, além de confirmar a informação apontada pela moda de que a maioria das reclamações é respondida em 1 um dia. Foi calculada também a variância que foi de 13.33 e o Desvio Padrão 3.65, que corroboram com interpretação do gráfico gerado, nesse sentido é possível concluir que os dados desse conjunto estão concentrados entre 1 e 10 dias.

#### 4.11. Resolução das reclamações pelas empresas

Após o reclamante registrar sua queixa e a empresa responder, o consumidor tem o prazo de até 20 dias para informar se o problema foi resolvido e classificar o processo. Foi desenvolvido uma Figura que demonstra a percentagem de reclamações classificadas como resolvidas:

FIGURA 13 – Resolução das Reclamações



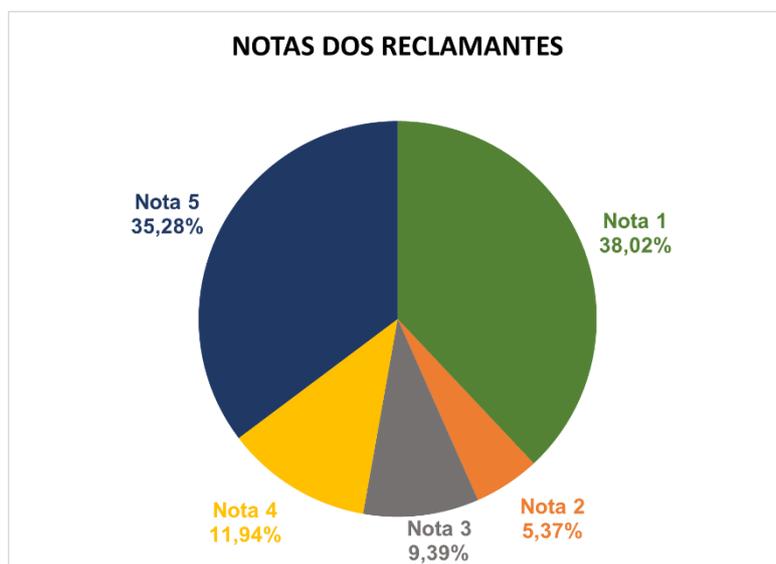
A Figura 13 aponta que a maior parte das reclamações não são avaliadas, cerca de 41%, demonstrando que ainda é necessário maior esforço da plataforma Consumidor.gov.br e das empresas do setor para incentivarem os reclamantes a apontarem seus feedbacks acerca da resolução dos problemas. É importante lembrar que no tópico “Registros Incompletos” identificamos 2.237 registros que sem nenhuma classificação quanto a resolução das reclamações, esses dados em branco deveriam em último caso serem somados a classificação de reclamações não avaliadas, afinal a percentagem de reclamações não resolvidas é significativa, um ponto de atenção para os stakeholders.

#### **4.12. Nota para o Atendimento da Empresa**

A última etapa do processo atribuição de uma nota pelo consumidor para o atendimento da empresa quanto a resolução ou não de sua reclamação, entre 1 e 5 de acordo com seu grau de satisfação, quanto mais satisfeito maior deve ser a nota. Relevante destacar que dentro o total de reclamações, 93.445 consumidores não registraram uma nota, ou seja, 41,63%, quantidade significativa. Para as empresas e para a própria plataforma o registro da nota é importante para

atestar a satisfação ou não, dos consumidores quanto ao processo de resolução do problema. Dentre os consumidores que registraram, a distribuição foi a seguinte:

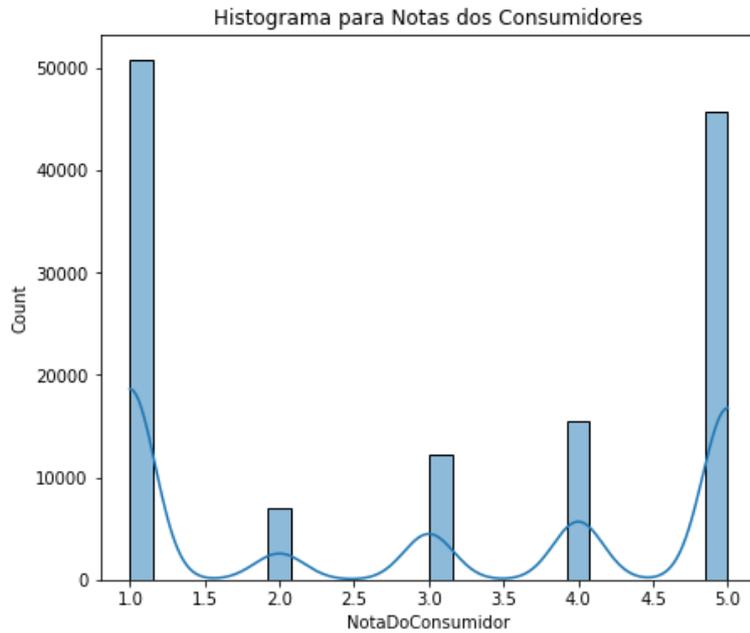
FIGURA 14 – Notas dos Reclamantes



A maior parte dos consumidores, cerca de 38% classificou a resolução de sua reclamação com nota 1, demonstrando que boa parte dos consumidores não ficam satisfeitos com o atendimento de suas reclamações, um ponto de extrema atenção para as empresas pois demonstra que as iniciativas de *service recovery* demandam atenção. A segunda maior parte dos reclamantes, cerca de 35%, atribui nota 5 a resolução. Logo a maioria dos reclamantes ou fica muito satisfeita ou muito insatisfeita com a resposta acerca de sua reclamação.

Somando os dois maiores grupos notamos que eles representam 73% do conjunto de dados, para entender melhor essa distribuição foi plotada utilizando Python um histograma:

FIGURA 15 – Histograma para Notas dos Consumidores



O histograma deixa claro que distribuição do conjunto de dados não segue a distribuição normal, e confirma a avaliação de que os dados estão concentrados próximos aos limites do conjunto. Foram realizados alguns cálculos utilizando funções específicas para confirmar os dados apresentados nos gráficos supracitados. A média calculada foi de 2,98 e a moda 1. O desvio padrão foi de 1,76, confirmando que os dados estão agrupados distantes da média. Nesse sentido a média não é uma medida representativa acerca das notas dos consumidores.

## 5. CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

O objetivo geral deste trabalho de realização da extração e análise exploratória dos dados de reclamações dos consumidores do setor de aviação civil referentes aos anos de 2019, 2020 e 2021, registradas no Consumidor.gov.br e disponibilizadas no site da ANAC foi atingido, através do desenvolvimento da ferramenta em Python, que se encontra nos Apêndices A, B e C.

O código foi dividido em três partes para facilitar sua compreensão. A Parte I está disponível no APÊNDICE A – Notebook Análise de Reclamações – Parte I: Montagem do Google Drive e Instalação de Pacotes Necessários, que se refere a preparação do ambiente do Google Colab para o recebimento dos dados, conexão do ambiente com o Google Drive e instalação dos pacotes necessários a execução das análises. A segunda parte do artefato desenvolvido se refere ao primeiro objetivo específico deste trabalho, a lembrar “Compor um modelo de notebook em Python que extraia os dados do portal ANAC”, foi atingido e o código criado se encontra no Apêndice B.

Para tornar possível a continuação das análises, a terceira parte do código, disponível no Apêndice C, parte do pressuposto que as planilhas foram baixadas manualmente para o diretório do Google Drive e segue com a análise dos dados. Nesse sentido o segundo objetivo específico “Desenvolver um notebook que realize a análise exploratória dos dados extraídos” foi contemplado. A partir das figuras e informações extraídas através dessa última parte do artefato foi possível realizar a interpretação acerca dos gráficos e informações obtidas com as análises, contemplando o terceiro objetivo específico “Interpretar os resultados obtidos a partir da análise dos dados”.

Aprimorando a estrutura do conjunto de dados e plotando as figuras foi possível aprofundar o conhecimento acerca das reclamações, a interpretação gerou diversos insights, os principais são apresentados a seguir:

- A maior parte dos reclamantes é do sexo masculino e possui entre 31 e 40 anos;
- Cerca de 75% das reclamações são respondidas em até 8 dias;
- A empresa Azul Linhas Aéreas apesar de ser a empresa que possui maior participação no mercado brasileiro é apenas a terceira colocada em número de reclamações;
- 41,63% dos consumidores não avaliam o atendimento das empresas, o que pode dificultar a avaliação quanto a satisfação dos consumidores quanto a solução de suas reclamações;

- O maior número de reclamações, aproximadamente 24%, foram registradas no estado de São Paulo;
- 98,7% das reclamações são registradas através da Plataforma Web;
- Não existe um pico de reclamações nos meses de alta temporada;
- A maior parte dos problemas enfrentados pelos consumidores do setor de aviação civil está ligado a questões financeiras;
- Apenas 33% das reclamações são classificadas como resolvidas pelos consumidores.
- A nota para resolução das empresas por parte dos consumidores se concentra em dois grupos. O grupo com maior número de registros, 38%, classificou a resolução com nota 1, portanto a maior parte se mostrou insatisfeita. Por outro lado 35%, dos reclamantes atribuiu nota 5, demonstrando ainda que parte considerável ficou satisfeita com o atendimento.
- O número de reclamações de entre 2019 e 2021 cresceu de 215,79%, o que de acordo com a ANAC pode ser explicado pelo aumento do uso da plataforma Consumidor.gov.br e por efeitos decorrentes da Pandemia de Covid-19;

O último tópico, ao considerar os **efeitos da pandemia do coronavírus**, contribui com o atingimento do quarto o objetivo específico deste trabalho. Entretanto, cabe destacar as limitações acerca deste tema. Considerar o impacto da pandemia nas reclamações do setor de aviação é um tema complexo, pois sofre influência de diversos fatores não contemplados nos dados. Portanto a conclusão é de que houve um aumento significativo no número de reclamações, apesar de uma forte redução do número de voos nos anos de 2020 e 2021, e a pandemia é um dos motivadores, de acordo com avaliação empírica da ANAC.

As perguntas apresentadas no tópico 3.5 Análise dos Dados demonstraram ser úteis no sentido de orientar as análises e foram respondidas como apresentado nos insights acima alcançando o último objetivo específico.

No decorrer do desenvolvimento desta pesquisa houve algumas limitações. Dentre elas o pacote Selenium utilizado para baixar as planilhas diretamente do site da ANAC passou por uma atualização, que gerou uma inconsistência em seu processo de instalação no Python, o que impediu a execução da funcionalidade de download dos dados, que até então executava o processo perfeitamente, levando a necessidade de baixar os dados manualmente. O código é apresentado no Apêndice B - Notebook Análise de Reclamações – Parte II: Extração de Dados do site da ANAC. Espera-se que os desenvolvedores do pacote Selenium e comunidade de

desenvolvedores encontrem uma forma de solucionar o problema, tornando essa funcionalidade do Notebook útil novamente. Outra limitação foi que, no desenvolvimento deste trabalho, nenhum colaborador da ANAC, empresas aéreas ou reclamante foi entrevistado.

Como sugestão para trabalhos posteriores, a explicação para o comportamento de alguns dados poderia ser elucidada através de investigações futuras que utilizem tanto análises de dados quanto de entrevistas, proporcionando uma visão mais profunda e ampla acerca do contexto de reclamações do setor de aviação. A partir dos insights obtidos neste trabalho, pesquisas futuras sobre o tema podem ainda buscar compreender o comportamento das reclamações ao longo dos meses e quais fatores externos ou internos podem influenciar em variações em seu cenário.

## REFERÊNCIAS BIBLIOGRÁFICAS

Agência Nacional de Aviação Civil. **Consulta Interativa – Indicadores do Mercado de Transporte Aéreo**. ANAC, 2022, Disponível em: <<https://www.gov.br/anac/pt-br/assuntos/dados-e-estatisticas/mercado-de-transporte-aereo/consulta-interativa/demanda-e-oferta-origem-destino>>. Acesso em 23 de setembro de 2022.

Agência Nacional de Aviação Civil. **Metadados do conjunto de dados: Dados do consumidor.gov**. ANAC, 2022. Disponível em: <<https://www.anac.gov.br/aceso-a-informacao/dados-abertos/areas-de-atuacao/voos-e-operacoes-aereas/dados-do-consumidor-gov/52-dados-do-consumidor-gov>> Acesso em 22 de setembro de 2022.

Agência Nacional de Aviação Civil. **ANAC divulga Boletim de Monitoramento do Consumidor.gov.br anual de 2021**. ANAC, 2022. Disponível em: <<https://www.gov.br/anac/pt-br/noticias/2022/anac-divulga-boletim-de-monitoramento-do-consumidor-gov-br-anual-de-2021>> Acesso em 17 de janeiro de 2023.

Agência Nacional de Aviação Civil, ANAC, Portaria nº 4.211, de 9 de fevereiro de 2021. Portaria de Organização Interna da Superintendência de Acompanhamento de Serviços Aéreos. Brasília, DF: Boletim de Pessoal e Serviço - V.16 Nº 37. 2021.

BRASIL. Lei nº 11.182, de 27 de setembro de 2005. Cria a Agência Nacional de Aviação Civil – ANAC, e dá outras providências. Brasília, DF: Diário Oficial da União, 2005.

DRESCH, Aline. et al. Capítulo 5: Fundamentos da Design Science e da Design Science Research para a engenharia. In: CAUCHICK, Paulo (Org.). **Metodologia Científica para Engenharia**. 1ª ed. Rio de Janeiro, Elsevier, 2019. p. 108-128. Disponível em: <[https://integrada.minhabiblioteca.com.br/reader/books/9788595150805/epubcfi/6/2\[%3Bvnd.vst.idref%3Dcover\]!/4/4/2\[cover01\]/2%4051:1](https://integrada.minhabiblioteca.com.br/reader/books/9788595150805/epubcfi/6/2[%3Bvnd.vst.idref%3Dcover]!/4/4/2[cover01]/2%4051:1)>. Acesso em 25 de setembro de 2022.

DHAR, Vasant. **Data science and prediction**. Communications of the ACM. Vol. 56, p. 64, dezembro de 2013. Disponível em:<<https://dl.acm.org/doi/abs/10.1145/2500499>>. Acesso em 25 de setembro de 2022.

FERREIRA, Rafael G C.; MIRANDA, Leandro B. A D.; PINTO, Rafael A.; et al. **Preparação e Análise Exploratória de Dados**. Grupo A, 2021. E-book. ISBN 9786556902890. Disponível em: <<https://integrada.minhabiblioteca.com.br/#/books/9786556902890/>>. Acesso em 18 setembro 2022.

HALSTEAD, Diane. **Negative Word of Mouth: Substitute for Supplement to Consumer Complaints?**. Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior, vol. 15, 2022. Disponível em: <<https://jcsdcb.com/index.php/JCSDCB/article/view/89>>. Acesso em 22 de setembro de 2022.

Instituto Brasileiro de Geografia e Estatística. **Tabelas - Prévia da População dos Municípios com base nos dados do Censo Demográfico 2022 coletados até 25/12/2022**. IBGE, 2022. Disponível em: <<https://www.ibge.gov.br/estatisticas/sociais/saude/22827-censo-demografico-2022.html?=&t=resultados>> Acesso em 17 de janeiro de 2023.

K. JARROD, Millman; MICHAEL, Aivazis. **Python for Scientists and Engineers**. Computing in Science & Engineering, vol. 13, p. 9-12, abril de 2011. Disponível em: <<https://aip.scitation.org/doi/10.1109/mcse.2011.36>>. Acesso em 20 de setembro de 2022.

MCKINNEY, Wes. **Python para Análise de Dados**. 1ª ed., Novatec, 2018.

RODRIGUES, Lucas. **Transporte Aéreo de Passageiros e o Avanço da Covid-19 no Brasil**. Hygeia. Revista Brasileira de Geografia Médica e da Saúde, [S. l.], p. 193–201, 2020. Disponível em: <<https://seer.ufu.br/index.php/hygeia/article/view/54407>>. Acesso em 24 de setembro de 2022.

SCHWEIKHART, Sharon B; STRASSER, Stephen; KENNEDY, Melissa R. **Service Recovery in Health Services Organizations**. Hospital & Health Services Administration, vol. 38, p.1, Spring 1993.

SENNA, Viviane; SOUZA, Adriano. **Consequências da pandemia de covid-19 para a aviação civil no Brasil**. Exacta, maio de 2011. Disponível em: <<https://doi.org/10.5585/exactaep.2021.19744>>. Acesso em 22 de setembro de 2022.

## APÊNDICE A

### Notebook Análise de Reclamações – Parte I: Montagem do Google Drive e Instalação de Pacotes Necessários

#### Montagem do Google Drive

```
[ ]
from google.colab import drive
drive.mount('/content/drive')
[ ]
%cd /content/drive/My Drive/Colab Notebooks/
[ ]
!pwd
[ ]
from google.colab import data_table
data_table.enable_dataframe_formatter()
```

#### Instalação de Pacotes do Python

```
[ ]
from time import sleep
import pandas as pd
import time
import os
import sys
from zipfile import ZipFile
[ ]
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
```

## APÊNDICE B

### Notebook Análise de Reclamações – Parte II: Extração de Dados do site da ANAC

#### Instalação do Selium

```
[ ]
!pip install -U selenium

[ ]
from selenium.webdriver.common.desired_capabilities import DesiredCapabilities
from selenium.webdriver.common.action_chains import ActionChains
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support.ui import Select
from selenium.webdriver.common.by import By
from selenium.webdriver.common.keys import Keys
from selenium import webdriver

[ ]
!apt-get update # to update ubuntu to correctly run apt install
!apt install chromium-chromedriver
!cp /usr/lib/chromium-browser/chromedriver /usr/bin

[ ]
sys.path.insert(0, '/usr/lib/chromium-browser/chromedriver')
```

#### Download dos dados diretamente do site da ANAC

```
In[ ]
wd.get('https://sistemas.anac.gov.br/dadosabertos/Voos%20e%20opera%C3%A7%C3%B5es%20a%C3%A9reas/Dados%20do%20consumidor.gov/')

[ ]
#Encontrar elementos na página - https://selenium-python.readthedocs.io/locating-elements.html
button1.click()

[ ]
dados2019 = pd.read_csv('DadosDoConsumidor2019.csv', skiprows=1, delimiter=';')

[ ]
#Encontrar elementos na página - https://selenium-python.readthedocs.io/locating-elements.html
button2 = wd.find_element(By.LINK_TEXT, 'DadosDoConsumidor2020.csv')
button2.click()

[ ]
```

```
dados2020 = pd.read_csv('DadosDoConsumidor2020.csv', skiprows=1, delimi  
ter=';')
```

```
[]
```

```
#Encontrar elementos na página - https://selenium-  
python.readthedocs.io/locating-elements.html
```

```
button3 = wd.find_element(By.LINK_TEXT, 'DadosDoConsumidor2021.csv')  
button3.click()
```

```
[]
```

```
dados2021 = pd.read_csv('DadosDoConsumidor2021.csv', skiprows=1, delimi  
ter=';')
```

## APÊNDICE C

### Notebook Análise de Reclamações – Parte III: Análise dos Dados

#### Lendo lanilhas Salvas e Concatenando

```
[ ]
#Lendo Planilhas
dados2019 = pd.read_csv('DadosDoConsumidor2019.csv', skiprows=1, delimi
ter=';')
dados2020 = pd.read_csv('DadosDoConsumidor2020.csv', skiprows=1, delimi
ter=';')
dados2021 = pd.read_csv('DadosDoConsumidor2021.csv', skiprows=1, delimi
ter=';')

[ ]
dados_final = [dados2019, dados2020, dados2021]

[ ]
dados = pd.concat(dados_final, axis=0)

[ ]
#Confirmando se os dados foram concatenados corretamente
dados.head()

[ ]
dados.tail()
```

#### Análise Inicial dos dados

```
[ ]
#Verificando tamanho dos dados de 2019
dados2019.shape

[ ]
#Verificando tamanho dos dados de 2020
dados2020.shape

[ ]
#Verificando tamanho dos dados de 2021
dados2021.shape

[ ]
#Verificando tamanho do conjunto de dados concatenado
dados.shape

[ ]
#Verificando Colunas
```

```
dados.columns
```

```
[]  
#Analisando dados em branco no conjunto de dados  
dados.isnull().sum()
```

## Limpeza dos dados (Data Cleaning)

```
[]  
#Deletando Colunas  
dados_limpos = dados.drop(columns = ["AnáliseDaRecusa", "EdiçãoDeConteúdo", "InteraçãoDoGestor", "CódigoClassificadorANAC", "Cidade", "HoraAbertura", "HoraAnálise", "HoraFinalização", "InteraçãoDoGestor", "CódigoClassificadorANAC", "EdiçãoDeConteúdo"], axis = 0)
```

```
[]  
dados_limpos.shape
```

```
[]  
#verificando a existência de dados duplicados  
dados_limpos.duplicated().sum
```

```
[]  
#verificando a existência de Outliers  
#verificando a Existência de Outliers em PrazoAnaliseGestor e criando  
boxplot  
plt.figure(figsize=(6,4))  
sns.boxplot(x=dados_limpos['PrazoAnaliseGestor'])  
plt.title('Prazo de Análise do Gestor')  
plt.xlabel('Número de dias')
```

```
[]  
#verificando a Existência de Outliers em TempoResposta e criando  
boxplot  
plt.figure(figsize=(6,4))  
sns.boxplot(x=dados_limpos['TempoResposta'])  
plt.xlabel('Número de dias')
```

## Análise dos dados

### Por Regiões

```
[]  
#Qual estado registrou o maior número de reclamações?  
n_reclamação = dados_limpos.groupby('UF').count().AnoAbertura.sort_values(ascending=False)  
n_reclamação  
[]  
#Confirmando se a contagem está correta  
n_reclamação.sum()
```

## Por Canal de Origem

```
[  
#Número de reclamações por canal de origem  
graf_CanalDeOrigem = dados_limpos.groupby('CanalDeOrigem').count().Gest  
or  
graf_CanalDeOrigem  
  
[  
#Plotagem do Gráfico  
plt.bar(graf_CanalDeOrigem.index, graf_CanalDeOrigem.values)  
plt.title('Quantidade de Reclamações por Canal')
```

## Por idade dos Respondentes

```
[  
graf_idade = dados_limpos.groupby('FaixaEtária').count().Gestor  
graf_idade
```

## Por períodos

```
[  
graf_meses = dados_limpos.groupby('MêsAbertura').count().Gestor  
graf_meses  
  
[  
plt.figure(figsize=(10,8))  
plt.bar(graf_meses.index, graf_meses.values)  
plt.title('Registro de Reclamações por Mês')  
plt.xlabel('Meses')  
plt.ylabel('Quantidade de Reclamações')
```

## Quantidade de reclamações improcedentes

```
[  
graf_improcedente = dados.groupby('AnáliseDaRecusa').count().Gestor  
graf_improcedente
```

## Contagem de reclamações por classificação

```
[  
#Reclamções por área  
graf_rec_area = dados_limpos.groupby('Área').count().Gestor  
graf_rec_area  
  
[  
plt.figure(figsize=(10,8))  
plt.bar(graf_rec_area.index, graf_rec_area.values)  
plt.title('Registro de Reclamações por Área')
```

```

plt.xlabel('Área')
plt.ylabel('Quantidade de Reclamações')

[]
#Reclamações por Assunto
graf_rec_assunto = dados_limpos.groupby('Assunto').count().Gestor.sort_
values(ascending=False)
graf_rec_assunto

[]
plt.figure(figsize=(20,8))
plt.bar(graf_rec_assunto.index, graf_rec_assunto.values)
plt.title('Registro de Reclamações por Assunto')
plt.xlabel('Assunto')
plt.ylabel('Quantidade de Reclamações')

[]
#Reclamações por Problema
graf_rec_problema = dados_limpos.groupby('Problema').count().Gestor
graf_rec_problema

[]
#Reclamação por Grupo do Problema
graf_rec_grupo = dados_limpos.groupby('GrupoProblema').count().Gestor
graf_rec_grupo

[]
plt.figure(figsize=(20,10))
plt.bar(graf_rec_grupo.index, graf_rec_grupo.values)
plt.title('Registro de Reclamações por Grupo')
plt.xlabel('Grupos')
plt.ylabel('Quantidade de Reclamações')

```

## Por Tempo de Resposta

```

[]
#Média do tempo de reposta das reclamações
dados_limpos[['TempoResposta']].mean()

[]
#Moda do tempo de reposta das reclamações
dados_limpos[['TempoResposta']].mode()

[]
#Confirmando Moda
graf_contagem = dados_limpos.groupby('TempoResposta').count().Gestor.so
rt_values(ascending=False)
graf_contagem

```

```

[]
#Porcentagem da moda sobre o total de reclamações
print(35732/224431)

[]
#Variância do tempo de reposta das reclamações
dados_limpos['TempoResposta'].var()

[]
#Devio Padrão do tempo de reposta das reclamações
dados_limpos['TempoResposta'].std()

[]
#Medidas de dispersão do tempo de reposta das reclamações
dados_limpos['TempoResposta'].describe()

[]
plt.figure(figsize=(7,6))
sns.histplot(data=dados_limpos.TempoResposta, kde=True)
plt.title('Histograma para Tempo de Resposta')

```

## Resolução por Empresas

```

[]
graf_empresas = dados_limpos.groupby('NomeFantasia').count().TempoRespo
sta.sort_values(ascending=False)
graf_empresas

[]
graf_resolução = dados_limpos.groupby('AvaliaçãoReclamação').count().Te
mpoResposta.sort_values(ascending=False)
graf_resolução

```

## Por Notas

```

[]
graf_notas = dados_limpos.groupby('NotaDoConsumidor').count().TempoResp
osta.sort_values(ascending=False)
graf_notas

[]
plt.figure(figsize=(8,6))
sns.boxplot(x=dados_limpos['NotaDoConsumidor'])
plt.title('Notas dos Reclamantes')
plt.xlabel('nota')

[]
#verificando a quantidade de dados nulos nessa coluna

```

```
graf_notas_nulas = dados_limpos[['NotaDoConsumidor']].isnull().count()
graf_notas_nulas

[]
#Variância
dados_limpos['NotaDoConsumidor'].var()

[]
dados_limpos['NotaDoConsumidor'].describe()

[]
print(93445/224431)

[]
plt.figure(figsize=(7,6))
sns.histplot(data=dados_limpos.NotaDoConsumidor, kde=True)
plt.title('Histograma para Notas dos Consumidores')
```