



UNIVERSIDADE DE BRASÍLIA
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, CONTABILIDADE
E GESTÃO DE POLÍTICAS PÚBLICAS - FACE
DEPARTAMENTO DE ADMINISTRAÇÃO

IGOR CAVALCANTE ARAUJO

**Análise Exploratória do Consumo Operacional da Aviação Regional no Brasil:
Perspectivas para a Operação de aeronaves do tipo eVTOL**

BRASÍLIA - DF
2024

**Análise Exploratória do Consumo Operacional da Aviação Regional no Brasil:
Perspectivas para a Operação de aeronaves do tipo eVTOL**

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.

Análise Exploratória do Consumo Operacional Direto de Combustível e a correlação para o Serviço de Transporte por eVTOL na Malha Aérea do Brasil

A Comissão Examinadora, abaixo identificada, aprova o Trabalho de Conclusão do Curso de Administração da Universidade de Brasília do (a) aluno (a)

Igor Cavalcante Araujo

Professor Dr. Victor Rafael Rezende Celestino
Professor-Orientador

Professora Polliana Candida Oliveira Martins
Professora-Examinadora

Professora Silvia Araujo dos Reis
Professora-Examinadora

Brasília, 01 de Julho de 2024.

RESUMO

Este trabalho de pesquisa em administração tem como objetivo realizar uma análise exploratória do consumo operacional direto de combustível e uma avaliação das perspectivas de operacionalização do serviço de transporte por eVTOL na malha aérea do Brasil. Utilizando as metodologias de Design Science Research e Análise Exploratória de Dados (AED), esta pesquisa investiga os custos operacionais diretos associados ao uso de combustíveis na aviação tradicional. A pesquisa coleta dados sobre custos de tarifas de voos, estatísticas operacionais, demanda e ofertas potenciais, além de informações adicionais de consumo. Adicionalmente, examinam-se as características geográficas e operacionais da malha aérea brasileira para determinar a viabilidade econômica e os benefícios potenciais dos eVTOLs, incluindo sua sustentabilidade e acessibilidade. Com base nesses dados, serão estabelecidos insights que permitam ao serviço de transporte por eVTOL competir com os modelos de aviação tradicionais. Esses resultados são importantes para compreender os desafios e oportunidades associados à adoção de eVTOLs na mobilidade aérea do Brasil, impulsionando o desenvolvimento desse mercado.

Palavras-chave: eVTOL, mobilidade aérea, Custos Operacionais Diretos, Design Science Research (DSR), análise de viabilidade.

ABSTRACT

This research project in administration aims to conduct an exploratory analysis of direct operational fuel consumption and an evaluation of the operationalization prospects of eVTOL transportation the services in Brazil's air network. Utilizing Design Science Research and Exploratory Data Analysis (EDA) methodologies, this research investigates the direct operational costs associated with fuel usage in traditional aviation. The research collects data on flight fare costs, operational statistics, demand and potential supply, in addition to additional consumption information. Additionally, the geographical and operational characteristics of the Brazilian air network are examined to determine the economic viability and potential benefits of eVTOLs, including their sustainability and accessibility. Based on these data, insights will be established to enable the eVTOL transportation service to compete with traditional aviation models. These results are important for understanding the challenges and opportunities associated with the adoption of eVTOLs in Brazil's air mobility, driving the development of this market.

Keywords: eVTOL, air mobility, Direct Operational Costs, Design Science Research (DSR), feasibility analysis.

LISTA DE ILUSTRAÇÕES

Figura 1 - Modelo de Aeronave eVTOL.....	12
Figura 2 - Comparativo de Modelos.....	13
Figura 3 - Exemplo da AAM.....	14
Figura 4 - Etapas da Condução da DSR.....	21
Figura 5 - Mapa de Rotas da Aviação Brasileira.....	26
Figura 6 - CESSNA 208 CARAVAN.....	27
Figura 7 - Análise de Consumo de Combustível: Cessna 208 Caravan.....	29
Figura 8 - Análise de Consumo de Combustível: AEROSPATIALE/ALENIA ATR 72 201/202.....	30
Figura 9 - Análise de Consumo de Combustível: AEROSPATIALE/ALENIA ATR 72 500/72-212A (500).....	30
Figura 10 - Análise de Consumo de Combustível: AEROSPATIALE/ALENIA ATR 72-600/72-212A (600).....	31
Figura 11 - Análise de CASK e Valor do Combustível: CESSNA 208 Caravan.....	34
Figura 12 - Análise de CASK e Valor do Combustível: AEROSPATIALE/ALENIA ATR 72 201/202.....	35
Figura 13 - Análise de CASK e Valor do Combustível: AEROSPATIALE/ALENIA ATR 72 500/72-212A (500).....	35
Figura 14 - Análise de CASK e Valor do Combustível: AEROSPATIALE/ALENIA ATR 72 600/72-212A (600).....	36
Figura 15 - Análise de DOC por Distância: CESSNA 208 Caravan.....	38
Figura 16 - Análise de DOC por Distância: AEROSPATIALE/ALENIA ATR 72 201/202.....	39
Figura 17 - Análise de DOC por Distância: AEROSPATIALE/ALENIA ATR 72 500/72-212A (500).....	39
Figura 18 - Análise de DOC por Distância: AEROSPATIALE/ALENIA ATR 72 600/72-212A (600).....	40

LISTA DE ABREVIATURAS E SIGLAS

AED - Análise Exploratória dos Dados

AAM – *Advanced Air Mobility* / (Mobilidade Aérea Avançada)

CASK - *Cost Available per Seat-Kilometer* / (Custo Médio por Assento-Quilômetro)

DOC - *Direct Operating Costs* / (Custos Operacionais Diretos)

DS – *Design Science*

DSR – *Design Science Research*

eVTOL – *Electric Vertical Take-Off and Landing* / (Decolagem e Pouso Vertical Elétrico)

FAA – *Federal Aviation Administration* / (Administração Federal de Aviação)

NASA – *National Aeronautics and Space Administration* / (Administração Nacional da Aeronáutica e Espaço)

PO – Pesquisa Operacional

RAM - *Regional Air Mobility* / (Mobilidade Aérea Regional)

UAM – *Urban Air Mobility* / (Mobilidade Aérea Urbana)

VTOL – *Vertical Take-Off and Landing* / (Decolagem e Pouso Vertical)

SUMÁRIO

1. INTRODUÇÃO.....	9
1.1 Contextualização.....	9
1.2 Formulação do problema.....	10
1.3 Objetivo Geral.....	10
1.4 Objetivos Específicos.....	10
1.5 Justificativa.....	10
2. REFERENCIAL TEÓRICO.....	11
2.1 Introdução.....	11
2.2 Definições no contexto.....	12
2.2.1 eVTOL (Electric Vertical Take-off and Landing).....	12
2.2.2 Mobilidade Aérea.....	13
2.3 Google Colaboratory.....	14
2.4 Custos Diretos da Operação.....	15
2.4.1 Custos Operacionais Diretos (DOC).....	15
2.5 Pesquisa Operacional.....	16
3. METODOLOGIA.....	17
3.1 Tipologia e descrição geral dos métodos de pesquisa.....	17
3.1.1 Design Science Research (DSR).....	17
3.1.2 Análise Exploratória dos Dados.....	18
3.1.2.1 Passo 1: Aeródromos e Distâncias entre Aeroportos.....	19
3.1.2.2 Passo 2: Organização e Transformação dos Dados de Tarifas (2019 a 2022)	19
3.1.2.3 Passo 3: Organização e Transformação dos Dados Estatísticos dos Voos	
(2019 a 2022).....	20
3.1.2.4 Passos 4 e 5: Transformação do Banco de Dados Final (2019 a 2022) e	
Análise.....	20
3.1.2.5 Conclusão.....	20
3.2 Caracterização e descrição dos instrumentos de pesquisa.....	20
3.3 Caracterização da organização, setor e área.....	25
3.4 População e amostra da pesquisa.....	25
3.5 Procedimentos de coleta de dados.....	27
3.6 Análise de dados.....	28
4. RESULTADOS E DISCUSSÕES.....	28
4.1 Análise de Consumo de Combustível: Consumo de Combustível e Distância.....	29
4.1.1 Introdução.....	29
4.1.2 Análise Geral.....	29
4.1.3 Aeronave: Cessna 208 Caravan.....	31
4.1.4 Aeronave: AEROSPATIALE/ALENIA ATR 72 201/202.....	32

4.1.5 Aeronave: AEROSPATIALE/ALENIA ATR 72 500/72-212A (500).....	32
4.1.6 Aeronave: AEROSPATIALE/ALENIA ATR 72-600.....	33
4.1.7 Síntese.....	33
4.2 Análise de CASK e Valor do Combustível: CASK Médio por Trimestre e Valor Médio do Combustível por Trimestre.....	34
4.2.1 Introdução.....	34
4.2.2 Análise Geral.....	34
4.2.3 Aeronave: Cessna 208 Caravan.....	36
4.2.4 Aeronave: AEROSPATIALE/ALENIA ATR 72 201/202.....	37
4.2.5 Aeronave: AEROSPATIALE/ALENIA ATR 72-500/72-212A (500).....	37
4.2.6 Aeronave: AEROSPATIALE/ALENIA ATR 72-600/72-212A (600).....	37
4.2.7 Síntese.....	37
4.3 Análise de DOC: Consumo de Combustível e Valor Médio do Combustível por Distância.....	38
4.3.1 Introdução.....	38
4.3.2 Análise Geral.....	38
4.3.3 Aeronave: Cessna 208 Caravan.....	41
4.3.4 Aeronave: AEROSPATIALE/ALENIA ATR 72 201/202.....	41
4.3.5 Aeronave: AEROSPATIALE/ALENIA ATR 72-500/72-212A (500).....	41
4.3.6 Aeronave: AEROSPATIALE/ALENIA ATR 72-600/72-212A (600).....	42
4.3.7 Síntese.....	42
4.4 Análise da Operacionalização do Serviço de Transporte por eVTOL.....	42
4.4.1 Introdução.....	42
4.4.2 Comparação com o CESSNA 208 Caravan.....	42
4.4.3 Requisitos para a Operacionalização dos eVTOLs.....	43
4.4.4 Síntese.....	43
5.1 Considerações Finais.....	44
5.2 Objetivos da Pesquisa.....	44
5.3 Desafios da Pesquisa.....	45
5.4 Direcionamentos para trabalhos futuros.....	45
REFERÊNCIAS.....	46
APÊNDICE.....	50

1. INTRODUÇÃO

1.1 Contextualização

A adoção dos veículos de decolagem e pouso vertical elétrico (eVTOLs) no Brasil apresenta oportunidades significativas para aprimorar a mobilidade aérea e enfrentar desafios relacionados à infraestrutura e sustentabilidade. Com uma população de aproximadamente 203 milhões de pessoas, de acordo com o IBGE, o país registrou mais de 82,2 milhões de passageiros somente em 2022, segundo dados da Agência Nacional de Aviação (ANAC). Entretanto, o Brasil enfrenta diversos problemas que impactam negativamente a mobilidade, como o congestionamento urbano, a conectividade regional limitada e a dificuldade de acesso a áreas remotas.

A implementação de eVTOLs no país traz benefícios tanto em termos de sustentabilidade ambiental quanto de eficiência operacional. Os eVTOLs são menos poluentes em comparação com aeronaves tradicionais, uma vez que utilizarão sistemas de propulsão elétrica, reduzindo as emissões de gases de efeito estufa e melhorando a qualidade do ar nas áreas urbanas, além disso, a adoção desse modelo de transporte pode reduzir a dependência de combustíveis fósseis, alinhando-se com os objetivos de transição energética e mitigação das mudanças climáticas, tornando-se um meio de transporte mais ecológico (Bris, Gaël et al. 2022).

Em termos de eficiência operacional, os eVTOLs têm a capacidade de decolar e pousar verticalmente, o que elimina a necessidade de infraestruturas aeroportuárias tradicionais e permite a operação em áreas urbanas densamente povoadas ou em locais remotos com acesso limitado. Isso pode resultar em uma maior conectividade regional, encurtando tempos de viagem e facilitando o acesso a áreas de difícil alcance, como regiões rurais e comunidades isoladas. Além disso, a implementação de serviços de transporte por eVTOLs pode ter um impacto econômico significativo. Estudos demonstram que a mobilidade aérea urbana pode gerar novos mercados, criar empregos diretos e indiretos e impulsionar o desenvolvimento de negócios em setores como turismo, comércio e serviços. Além disso, a redução dos custos operacionais e a otimização das rotas podem tornar o transporte aéreo mais acessível a um número maior de pessoas, democratizando o acesso a regiões antes subatendidas (Cocca et al., 2020).

No entanto, a implementação bem-sucedida de eVTOLs no Brasil requer uma análise cuidadosa de viabilidade e um modelo de precificação adequado. É essencial considerar fatores como custos operacionais, demanda potencial, regulamentações governamentais e aspectos técnicos relacionados à infraestrutura. A pesquisa acadêmica sobre os custos operacionais diretos da aviação tradicional contribuirá para preencher essa lacuna de conhecimento e fornecer informações valiosas para a tomada de decisões estratégicas no setor de Serviços de Transporte por eVTOL na Malha Aérea do Brasil.

Em suma, a adoção de eVTOLs no Brasil apresenta oportunidades para melhorar a mobilidade aérea, reduzir a emissão de poluentes e promover o desenvolvimento econômico. A pesquisa acadêmica nesse campo é fundamental para avaliar a viabilidade, destacar as dificuldades e desenvolver modelos de operacionalização adequados, considerando as

características específicas do país e contribuindo para o avanço da mobilidade aérea sustentável no Brasil.

1.2 Formulação do problema

Considerando o crescente interesse no desenvolvimento dos serviços de transporte por veículos elétricos de decolagem e pouso vertical (eVTOL) no Brasil, surge a necessidade de investigar a viabilidade da implementação de um modelo de operacionalização para esse serviço competir com os voos tradicionais na malha aérea do país. Diante disso, o problema de pesquisa é: É viável implementar um modelo para o serviço de transporte por eVTOL na malha aérea do Brasil, considerando aspectos de custos operacionais da aviação regional?

1.3 Objetivo Geral

O objetivo deste trabalho é compreender e comparar os custos operacionais diretos de combustível da aviação regional nas principais malhas aéreas do Brasil, garantindo a viabilidade da operacionalização do serviço de transporte por eVTOL. Essa análise tem como finalidade contribuir para a implementação e operacionalização eficiente desse modelo de transporte no país.

1.4 Objetivos Específicos

Pretende-se alcançar os seguintes objetivos específicos:

1. Identificar a correlação entre o consumo de combustível e a distância para as quatro aeronaves abordadas na pesquisa;
2. Investigar a relação do custo médio por assento-quilômetro e do valor de combustível ao longo dos anos de 2019 a 2022;
3. Comparar os dados de custos operacionais diretos de combustível dos voos tradicionais;
4. Avaliar qualitativamente a viabilidade da operacionalização do serviço de transporte por eVTOL.

1.5 Justificativa

O avanço do setor de Mobilidade Aérea Regional e Urbana tem impulsionado o desenvolvimento dos serviços de transporte por veículos elétricos de decolagem e pouso vertical (eVTOL) no Brasil. No entanto, a operacionalização desses serviços ainda é um desafio devido à complexidade e particularidades desse novo modelo de transporte. Portanto, é crucial compreender e comparar os custos operacionais diretos dos serviços de eVTOL em relação aos voos tradicionais, a fim de promover uma implementação eficiente desse meio de transporte no país.

O primeiro objetivo busca identificar a correlação entre o consumo de combustível e a distância percorrida. Este objetivo é essencial para entender a eficiência das aeronaves em

diferentes rotas e como o consumo de combustível varia conforme a distância. Compreender essa relação é fundamental para planejar rotas mais eficientes e reduzir os custos operacionais.

Outro objetivo específico é investigar a relação do custo médio por assento-quilômetro (CASK) e do valor de combustível ao longo dos anos de pesquisa. Essa análise permitirá avaliar como os custos variam ao longo do tempo e quais fatores influenciam essas variações, contribuindo para um planejamento financeiro mais preciso e sustentável.

Além disso, comparar o modelo de custos operacionais diretos de combustível dos voos tradicionais com o modelo proposto para os eVTOLs é um passo importante. Essa comparação ajudará a identificar as vantagens e desvantagens de cada modelo, permitindo avaliar a competitividade dos eVTOLs no mercado aéreo.

O quarto objetivo específico é realizar uma avaliação qualitativa da viabilidade da operacionalização do serviço de transporte por eVTOL. Utilizando a aviação regional como um norteador para a integração dos eVTOLs. A análise qualitativa fornecerá insights valiosos sobre o comportamento dos custos operacionais e permitirá adaptar as estratégias de implementação dos eVTOLs de maneira mais eficaz e realista.

Em suma, a pesquisa proposta tem como objetivo geral contribuir para a implementação e operacionalização eficiente dos serviços de eVTOL no Brasil. Os objetivos específicos são fundamentais para alcançar esse propósito, permitindo uma compreensão aprofundada dos custos operacionais diretos de combustível, a comparação com os voos tradicionais e a análise de sua viabilidade operacional. Os resultados deste estudo poderão beneficiar tanto as organizações atuantes no setor quanto os clientes, além de preencher lacunas de pesquisa e abrir caminho para estudos futuros na área de eVTOL e Mobilidade Aérea Regional e Urbana.

2. REFERENCIAL TEÓRICO

2.1 Introdução

O referencial teórico é uma base essencial para a compreensão e contextualização do estudo sobre a viabilidade e a operacionalização do serviço de transporte por veículos do tipo eVTOL na malha aérea do Brasil. Este capítulo tem como objetivo fornecer uma fundamentação teórica abrangente, abordando conceitos, definições e estudos relevantes que embasam a pesquisa.

A justificativa para incluir um referencial teórico detalhado reside na necessidade de contextualizar o avanço dos eVTOLs, a mobilidade aérea, os custos operacionais diretos e a aplicação da pesquisa operacional. Este embasamento teórico não só clarifica os conceitos chave, mas também orienta a análise dos dados e a formulação de estratégias eficientes para a implementação dos eVTOLs no Brasil. Ao explorar definições e contextos variados, o capítulo busca oferecer uma visão consolidada que sustente as análises e discussões subsequentes, garantindo uma abordagem rigorosa e bem fundamentada.

2.2 Definições no contexto

2.2.1 eVTOL (*Electric Vertical Take-off and Landing*)

O surgimento dos Veículos Elétricos de Decolagem e Pouso Vertical (eVTOLs) é resultado do avanço tecnológico no setor aeroespacial e da busca por soluções mais eficientes e sustentáveis no transporte aéreo. Inicialmente impulsionado pelo interesse em veículos aéreos não tripulados (drones), o desenvolvimento de sistemas de propulsão elétrica e aprimoramentos na aerodinâmica permitiram a evolução dos eVTOLs para aeronaves capazes de transportar passageiros. Os eVTOLs têm o potencial de revolucionar o transporte aéreo urbano, oferecendo uma alternativa mais sustentável e eficiente em termos de energia. Essas aeronaves elétricas possuem potencial de redução das emissões de carbono e o impacto ambiental associado ao transporte aéreo convencional.

Os eVTOLs apresentam vantagens significativas em termos de flexibilidade operacional e acessibilidade. A capacidade de decolar e pousar verticalmente elimina a necessidade de infraestruturas aeroportuárias tradicionais, permitindo que essas aeronaves atendam a áreas urbanas congestionadas, onde o espaço para helipontos convencionais é limitado. Além disso, a propulsão elétrica dessas aeronaves reduz o ruído e as emissões, tornando-as mais adequadas para operações urbanas e áreas sensíveis ao ruído. A figura 1 a seguir mostra um exemplo de aeronave do tipo eVTOL.

Figura 1 - Modelo de aeronave eVTOL.



Fonte: EVE Air Mobility, 2023.

A viabilidade de implementar um serviço de transporte por eVTOL na malha aérea do Brasil requer uma análise criteriosa de diversos aspectos, incluindo a viabilidade técnica, econômica e regulatória. A precificação desse serviço desempenha um papel fundamental na determinação da atratividade e da sustentabilidade financeira do projeto. Considerando as características singulares dos eVTOLs, como a capacidade reduzida de passageiros e a menor necessidade de investimentos em infraestrutura de pouso e decolagem, é essencial desenvolver um modelo de operacionalização adequado e competitivo. Estudos e pesquisas sobre modelos de operacionalização para serviços de transporte aéreo têm sido realizados no contexto de companhias aéreas tradicionais, porém, é necessário considerar as particularidades inerentes aos serviços de transporte por eVTOL. A análise cuidadosa dos custos operacionais, demanda esperada, concorrência, regulamentações vigentes e características do mercado brasileiro são aspectos fundamentais para desenvolver um modelo

eficiente e viável para o serviço de transporte por eVTOL na malha aérea do Brasil. A figura 2 demonstra um comparativo entre helicópteros, eVTOLs e aviões elétricos.

Figura 2 - Comparativo de modelos.



Fonte: G1, 2023.

2.2.2 Mobilidade Aérea

A mobilidade aérea é um conceito que se refere à capacidade de deslocamento de pessoas e mercadorias no espaço aéreo utilizando aeronaves. Enquanto a aviação comercial tradicional é dominada por grandes aeroportos e aeronaves de grande porte, a mobilidade aérea busca trazer soluções mais flexíveis e acessíveis, permitindo voos em áreas urbanas e suburbanas, com a utilização de aeronaves menores e mais versáteis, como exemplo os eVTOLs.

A NASA (*National Aeronautics and Space Administration*) tem se envolvido na criação de padrões e diretrizes para a indústria de mobilidade aérea. A agência trabalha em colaboração com órgãos reguladores, como a Administração Federal de Aviação dos Estados Unidos (FAA), e com as empresas do setor aeroespacial para desenvolver regulamentações e políticas que permitam a operação segura e eficiente das aeronaves de mobilidade aérea.

Dentro desse contexto, surgem alguns termos importantes relacionados à mobilidade aérea:

1. **AAM (Advanced Air Mobility):** AAM refere-se à próxima geração de transporte aéreo, que inclui aeronaves elétricas de decolagem e pouso vertical (eVTOL) e outros veículos aéreos autônomos. Essas aeronaves têm a capacidade de decolar e pousar verticalmente, dispensando a necessidade de pistas de pouso tradicionais. O objetivo do AAM é proporcionar um transporte aéreo mais eficiente, para pessoas e cargas, seguro e sustentável em áreas urbanas e suburbanas.
2. **UAM (Urban Air Mobility):** UAM é um termo frequentemente usado para descrever a mobilidade aérea específica nas áreas urbanas. Envolve o uso de aeronaves elétricas de decolagem e pouso vertical (eVTOL) para transporte de passageiros e mercadorias dentro das cidades. O UAM tem como objetivo aliviar o tráfego terrestre congestionado, oferecendo viagens rápidas e eficientes entre locais urbanos.
3. **RAM (Regional Air Mobility):** RAM é um conceito que se refere à mobilidade aérea regional, abrangendo voos de média distância entre cidades e áreas adjacentes. A utilização de aeronaves elétricas de decolagem e pouso vertical (eVTOL) também pode desempenhar um papel importante na expansão do RAM, proporcionando conexões rápidas e diretas entre regiões sem a necessidade de grandes aeroportos.

Figura 3 - Exemplo da AAM.



Fonte: NASA, 2021.

2.3 Google Colaboratory

O Google Colaboratory, também conhecido como Colab, emerge como um ambiente de desenvolvimento de notebooks Python na nuvem fornecido pelo Google. Notabiliza-se por possibilitar a redação, execução e compartilhamento de código Python diretamente no

navegador, sem a necessidade de configurar um ambiente de desenvolvimento local. O Colab tem conquistado amplo reconhecimento em projetos de aprendizado de máquina, análise de dados e pesquisa. Sua infraestrutura disponibiliza acesso a recursos computacionais robustos, como GPUs e TPUs, e se destaca pela interatividade de notebooks, bibliotecas pré-instaladas, colaboração e compartilhamento de projetos, além da integração harmoniosa com outros serviços do Google.

Nesse contexto, o Google Colaboratory assume uma posição central neste trabalho, onde se evidencia como a plataforma primordial para a condução da análise exploratória de dados. Sua utilização será essencial para manipular os dados em questão, permitindo a execução de todo o processo de tratamento, análise e visualização dos dados em um único ambiente. Ademais, o Colab oferece a possibilidade de cooperação com trabalhos anteriores, os quais servirão de base para o presente estudo, e facilita a interação com o orientador designado. Essa escolha fundamenta-se na flexibilidade e facilidade de uso proporcionadas pelo Google Colab, as quais contribuirão para a obtenção de *insights* valiosos e embasamento sólido para as conclusões deste trabalho acadêmico.

2.4 Custos Diretos da Operação

Embora o eVTOL seja amplamente considerado uma solução altamente promissora e viável para abordar os desafios da mobilidade urbana, devido à sua capacidade de operar em espaços reduzidos e sua aplicabilidade em diversas áreas geográficas, é imprescindível destacar a existência dos custos operacionais diretos e indiretos inerentes ao serviço oferecido por esse sistema de transporte. Somente por meio de uma análise abrangente desses custos é possível obter uma compreensão precisa do valor real do serviço e avaliar a relação entre os benefícios de transporte para os clientes e a margem operacional das empresas envolvidas. A partir dessa análise criteriosa, é possível determinar a viabilidade da proposta de desenvolver o eVTOL como um meio de transporte economicamente acessível em comparação aos voos tradicionais.

A introdução do serviço de eVTOL na malha aérea urbana (UAM) é justificada pela sua adequação a viagens de distâncias razoáveis, como apontado por Wang, Li e Qu (2023). Além disso, há uma preferência por implantá-lo em grandes centros urbanos, impulsionada pela redução dos custos operacionais, o que, por sua vez, atrai um maior número de clientes.

Os custos de operação podem ser categorizados em dois tipos: Custos Operacionais Diretos (DOC) e Custos Operacionais Indiretos (DOI).

2.4.1 Custos Operacionais Diretos (DOC)

Os DOC são despesas diretamente associadas à operação do eVTOL, incluindo:

1. **Consumo de Combustível/Energia:** Um dos principais componentes dos DOC é o consumo de combustível ou energia elétrica, no caso dos eVTOLs. A eficiência energética dos eVTOLs e a capacidade de utilizar fontes de energia renováveis podem representar uma vantagem competitiva significativa em relação aos aviões tradicionais. Para análises mais detalhadas, serão utilizadas métricas como o CASK,

que mede o custo operacional por assento disponível por quilômetro, permitindo uma avaliação comparativa entre diferentes modalidades de transporte aéreo.

2. **Manutenção e Reparos:** A manutenção regular e reparos necessários para garantir a segurança e a eficiência operacional dos eVTOLs são custos que devem ser considerados. A frequência e o custo dessas manutenções podem variar dependendo da tecnologia utilizada e da intensidade de uso das aeronaves.
3. **Tripulação e Treinamento:** Os custos com a tripulação, incluindo salários e benefícios, além dos custos com o treinamento necessário para operar eVTOLs, são componentes importantes dos DOC. Embora os eVTOLs possam ser projetados para operar com uma tripulação reduzida, o treinamento especializado ainda representa um investimento significativo.
4. **Seguros:** Os custos de seguros para cobrir potenciais riscos operacionais e de responsabilidade civil são outro componente crucial dos DOC. A avaliação de risco e os custos associados podem variar dependendo da implementação e das regulamentações locais.

A análise dos custos operacionais diretos é fundamental para determinar a viabilidade econômica do uso de eVTOLs como uma alternativa ao transporte aéreo tradicional. Ao entender detalhadamente esses custos, é possível avaliar a sustentabilidade financeira do serviço de eVTOLs e sua competitividade no mercado de transporte aéreo urbano. Esta análise será essencial para futuras decisões estratégicas e para a expansão do uso de eVTOLs na malha aérea do Brasil. No contexto deste trabalho, é de particular importância considerar o DOC, que está diretamente relacionado ao voo da aeronave. Isso inclui elementos como o consumo de combustível, que desempenha um papel fundamental na condução das análises. O consumo de combustível será usado no decorrer do trabalho para eventuais análises das curvas da operação em distância atrelada ao consumo de combustível nesse período, além disso também poderá aparecer na forma de CASK (Custo Operacional por Assento-Quilômetro).

2.5 Pesquisa Operacional

A Pesquisa Operacional (PO) é uma abordagem científica que utiliza métodos quantitativos para resolver problemas complexos relacionados à tomada de decisão em organizações. Ela surgiu durante a Segunda Guerra Mundial, quando os militares precisavam otimizar o uso de recursos escassos, como mão de obra, materiais e tempo, para alcançar objetivos estratégicos. A PO se baseia em técnicas matemáticas e estatísticas, modelagem e simulação para analisar e resolver problemas de forma eficiente e racional (Hillier, F. S., & Lieberman, G. J., 2014).

A PO funciona em etapas distintas, começando pela identificação e formulação do problema, seguida pela construção de modelos matemáticos que representem a situação real, considerando as restrições e os objetivos envolvidos. Esses modelos são então solucionados utilizando técnicas de otimização e simulação, que visam encontrar a melhor solução possível dentro das limitações existentes. Por fim, os resultados obtidos são interpretados e comunicados aos tomadores de decisão, auxiliando-os na escolha das melhores alternativas (Taha, H. A., 2017).

A pesquisa operacional é amplamente utilizada em diversos setores, como logística, produção, finanças, saúde e transportes, entre outros. No contexto do estudo dos eVTOLS, a pesquisa operacional pode ser aplicada para otimizar a operacionalização do serviço, de acordo com o planejamento de rotas, frequência e horários de voo, a gestão de recursos e a análise de viabilidade econômica. Ao utilizar modelos matemáticos e técnicas de otimização, a PO pode ajudar a maximizar a eficiência operacional dessas aeronaves, considerando os recursos disponíveis, as demandas de transporte e os objetivos específicos das operações de eVTOL.

3. METODOLOGIA

3.1 Tipologia e descrição geral dos métodos de pesquisa

3.1.1 Design Science Research (DSR)

O presente estudo adota a metodologia de Design Science Research (DSR), baseada no paradigma da Design Science (DS), para abordar a análise do consumo operacional direto de combustível e a viabilidade do serviço de transporte por eVTOL na malha aérea do Brasil. A DSR é uma abordagem prática que visa resolver problemas existentes, melhorar sistemas e criar artefatos que promovam um desempenho aprimorado em organizações e na sociedade (Hevner et al., 2004; Gregor et al., 2013).

Neste contexto, a pesquisa busca compreender a dinâmica do mercado emergente de eVTOL no Brasil e contribuir para a solução de problemas relacionados à viabilidade e eficiência do consumo de combustível desse serviço. De acordo com Hevner et al. (2004), a DSR requer a identificação, o desenvolvimento, a avaliação e a demonstração de artefatos para resolver problemas complexos.

Os artefatos desenvolvidos nesta pesquisa podem ser classificados da seguinte forma:

1. **Constructos:** Os constructos utilizados neste estudo descrevem os conceitos fundamentais relacionados ao consumo operacional direto de combustível e a viabilidade do serviço de transporte por eVTOL na malha aérea do Brasil. Esses constructos são essenciais para compreender os desafios e as necessidades do mercado, bem como para identificar estratégias eficazes de operacionalização (Peffer et al., 2007).
2. **Modelos:** Os modelos desenvolvidos neste trabalho representam a estrutura geral da realidade do serviço de transporte por eVTOL na malha aérea do Brasil. Esses modelos podem incluir modelos econômicos, de demanda, de consumo de combustível e de custos, fornecendo uma base sólida para a formulação de estratégias de viabilidade e eficiência operacional (Winter et al., 2008).
3. **Métodos:** Os métodos desenvolvidos nesta pesquisa referem-se a abordagens sistemáticas para lidar com os desafios do consumo operacional direto de combustível e da viabilidade do serviço de transporte por eVTOL na malha aérea do Brasil. Esses métodos podem envolver análises de viabilidade econômica, pesquisas de mercado e

avaliações de custos, auxiliando na tomada de decisões informadas sobre a operacionalização do serviço (Peffer et al., 2007; Hevner et al., 2004).

4. **Instanciação:** A instanciação dos artefatos desenvolvidos neste estudo envolverá a implementação e execução dos mesmos em um ambiente real, evidenciando a viabilidade e a eficácia das estratégias de consumo operacional direto de combustível propostas. Essa etapa permitirá verificar se as soluções propostas são adequadas ao contexto da malha aérea do Brasil e se podem contribuir para a melhoria do serviço de transporte por eVTOL (Peffer et al., 2007).
5. **Design Propositions:** As design propositions neste estudo consistem em regras tecnológicas ou de projeto que são consideradas contribuições teóricas da Design Science. Essas proposições podem ser formuladas com base nos resultados obtidos com os artefatos desenvolvidos, fornecendo diretrizes e orientações para pesquisas futuras e práticas relacionadas à viabilidade e eficiência do serviço de transporte por eVTOL na malha aérea do Brasil (Hevner et al., 2004).

Em suma, este estudo, embasado na DSR, busca desenvolver artefatos que contribuam para a resolução dos desafios de viabilidade e eficiência do consumo operacional direto de combustível do serviço de transporte por eVTOL na malha aérea do Brasil. Por meio da identificação de constructos, criação de modelos, aplicação de métodos, instanciação dos artefatos e formulação de design propositions, espera-se obter insights valiosos para o aprimoramento das estratégias de operacionalização no contexto do eVTOL no Brasil.

3.1.2 Análise Exploratória dos Dados

Esta seção foi desenvolvida em colaboração com o trabalho “Análise Exploratória das Tarifas Aéreas praticadas no Brasil no Período de 2019 a 2022 e Perspectivas para o Serviço de Transporte por eVTOL” (GODINHO, 2023). Embora as tarifas aéreas tenham sido o foco principal do trabalho citado, a análise deste trabalho se direciona especificamente para os Custos Operacionais Diretos de Combustível, utilizando uma metodologia similar de Análise Exploratória de Dados (AED) para examinar os dados coletados da ANAC.

Em um cenário de crescente disponibilidade de dados, especialmente no setor aéreo, a capacidade de processar e transformar grandes volumes de dados em informações úteis é crucial. Quando essas informações são aplicadas de forma eficaz, elas se tornam conhecimento, o que, por sua vez, pode ser utilizado para uma tomada de decisões mais informada e estratégica (Ferreira, Rafael G. C. et al., 2021). A pesquisa aqui desenvolvida visa extrair insights dos dados da ANAC para entender melhor os fatores que influenciam os custos operacionais de combustível no transporte aéreo e aplicar essas descobertas ao contexto emergente do transporte por eVTOL.

O objetivo central é explorar como as variáveis do mercado de transporte aéreo, especificamente os custos de combustível, variaram de 2019 a 2022, e identificar possíveis correlações entre elas. A Análise Exploratória de Dados (AED) é uma ferramenta vital nesse processo, permitindo-nos desvendar padrões e relações complexas nos dados (Ferreira, Rafael G. C. et al., 2021).

Conforme Ferreira, Rafael G. C. et al. (2021), a AED estrutura os dados coletados de maneira a maximizar a extração de informações relevantes. Os dados são então refinados para melhorar sua qualidade e precisão, essencial para análises subsequentes robustas.

A análise pode ser univariada, focando em uma variável por vez, ou multivariada, examinando relações entre múltiplas variáveis. A identificação de correlações, que indicam quando dois eventos tendem a ocorrer simultaneamente, também é uma parte crucial do processo.

Os resultados da AED são frequentemente apresentados em formas gráficas ou tabulares, que devem ser claras e intuitivas para facilitar a interpretação. A variedade de gráficos disponíveis permite a visualização adequada de cada tipo de variável, o que é fundamental para compreender as influências no custo do combustível e como essas podem ser aplicadas ao transporte por eVTOL.

Para esta análise exploratória, foi organizado e tratado os quatro principais bancos de dados da ANAC, referentes a aeródromos, tarifas, e estatísticas de voos de 2019 a 2022. A metodologia detalhada para a organização e tratamento desses dados está descrita no Apêndice A. Este apêndice também documenta o uso do Google Colaboratory para executar as primeiras etapas da AED (Ferreira, Rafael G. C. et al., 2021). No final do processo, os dados foram integrados em um único banco de dados, que foi então analisado utilizando técnicas de visualização de dados.

3.1.2.1 Passo 1: Aeródromos e Distâncias entre Aeroportos

Inicialmente, foi realizada a leitura, organização e transformação dos dados de aeródromos da ANAC. A criação do banco de dados de distâncias entre aeroportos foi efetuada calculando-se as distâncias entre combinações de aeroportos de origem e destino. Este processo envolveu quatro etapas principais:

1. **Leitura e Limpeza dos Dados:** Os dados dos aeródromos foram lidos, variáveis irrelevantes foram eliminadas, e dados ausentes foram removidos.
2. **Coleta de Coordenadas Geográficas:** Com base no projeto 'airportsdata20231017' (Borsetti, M., 2022), as latitudes e longitudes dos aeroportos foram obtidas, possibilitando o cálculo das distâncias usando a fórmula do grande círculo.
3. **Cálculo das Distâncias:** Utilizando a função `distance_on_unit_sphere`, todas as combinações de aeroportos foram processadas para calcular as distâncias entre eles.
4. **Organização dos Dados:** As distâncias calculadas foram organizadas e convertidas de strings para floats, completando assim a criação do banco de dados de distâncias.

3.1.2.2 Passo 2: Organização e Transformação dos Dados de Tarifas (2019 a 2022)

No segundo passo, foi realizado a leitura, organização e transformação dos dados de voos da ANAC para os anos de 2019 a 2022. Os arquivos mensais foram concatenados em um único banco de dados anual, e as variáveis quantitativas foram convertidas para o formato float para análise posterior.

3.1.2.3 Passo 3: Organização e Transformação dos Dados Estatísticos dos Voos (2019 a 2022)

O terceiro passo consistiu na leitura e organização dos arquivos de estatísticas básicas e combinadas dos voos. Os dados de 'estatísticas básicas', focados nos movimentos das aeronaves, e de 'estatísticas combinadas', orientados para os trajetos dos passageiros, foram tratados separadamente e posteriormente integrados.

A integração dos bancos de dados de 'estatísticas básicas', 'tarifas' e 'distâncias' com o banco de dados de 'estatísticas combinadas' foi essencial para fornecer uma visão completa e correlacionada dos dados de 2019 a 2022. Variáveis como distância, modelo de aeronave, assentos ofertados e tarifas praticadas foram incorporadas ao banco de dados final.

3.1.2.4 Passos 4 e 5: Transformação do Banco de Dados Final (2019 a 2022) e Análise

No quarto passo, realizamos a transformação final, removendo dados vazios e consolidando os dados anuais em um único banco de dados. Este banco consolidado foi então analisado utilizando representações gráficas, que facilitaram a interpretação das relações entre as variáveis.

A visualização gráfica permitiu identificar padrões e correlações que influenciam os custos operacionais de combustível. Esses insights são valiosos para a aplicação no contexto do transporte por eVTOL, proporcionando uma base sólida para futuras pesquisas e planejamento estratégico.

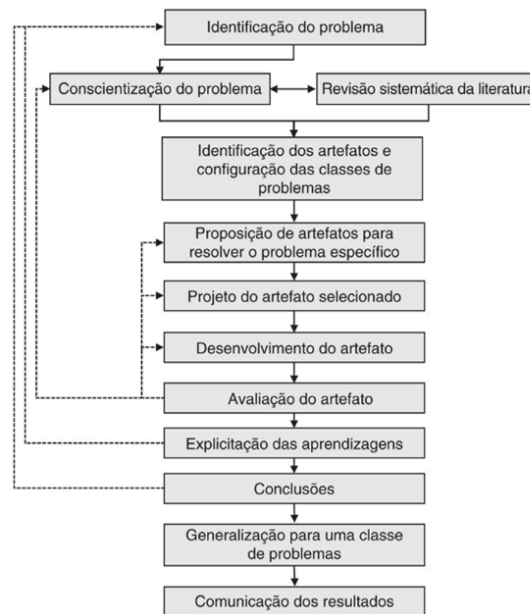
3.1.2.5 Conclusão

Em resumo, a aplicação da Análise Exploratória de Dados (AED) nos dados coletados da ANAC possibilitou uma compreensão aprofundada dos fatores que influenciam os custos operacionais de combustível no setor aéreo. A colaboração com pacotes do *python* foi crucial para o desenvolvimento desta análise, permitindo a estruturação e a extração eficiente de informações dos dados. Os resultados fornecem uma base robusta para avaliar como essas variáveis podem ser aplicadas ao contexto do transporte por eVTOL, contribuindo para o avanço e a inovação nesse setor emergente.

3.2 Caracterização e descrição dos instrumentos de pesquisa

A Design Science Research proporciona uma abordagem única, direcionada para a geração de resultados concretos e soluções práticas. Seu valor é especialmente evidente quando a pesquisa busca prescrever soluções para problemas específicos ou quando se propõe a desenvolver e/ou avaliar artefatos. Portanto, surge como um método dinâmico e inovador capaz de orientar investigações científicas nas mais diversas áreas do conhecimento, de acordo com Van Aken (apud Cauchick, Paulo, 2019). Tal raciocínio de pesquisa terá por objetivo identificar todas as doze etapas para a condução da DSR (Dresch et al., 2015). Tais etapas de direção da pesquisa podem ser visualizadas a seguir:

Figura 4 - Etapas para a condução da DSR.



Fonte: Metodologia Científica para Engenharia (Cauchick, Paulo, 2019, *apud* Dresch et al. 2015, p. 125)

1. Identificação do problema

Neste trabalho, o desafio em questão é a análise dos custos diretos de combustível das aeronaves tradicionais para servir de insumo para os veículos elétricos de decolagem e pouso vertical (eVTOL). Esta etapa inicial do Design Science Research (DSR) é fundamental, pois permite uma compreensão aprofundada do problema e a análise de suas implicações no contexto da aviação (Pidd, 1998, *apud* Cauchick, Paulo, 2019).

2. Conscientização do problema

Na segunda etapa do método, é essencial aprofundar a compreensão do problema, identificando suas principais causas e sua relação com o contexto interno e externo. Também é importante considerar o desempenho esperado do artefato e como ele pode impactar a resolução do problema, levando em conta seus requisitos de funcionamento e funcionalidades (Dresch et al., 2015).

No caso específico do emergente serviço de transporte com eVTOL, compreender o contexto que o envolve é um desafio complexo, devido às muitas incertezas relacionadas à sua implementação. Nesse sentido, a proposta de combinar os modelos de pesquisa de Souza, Juliana (2023), intitulado "Uma proposta de estimativa de CASK de aeronaves do tipo eVTOL baseada em modelo híbrido de desempenho de aeronaves de asa rotativa e asa fixa", e de Godinho, Gabriel (2023), intitulado "Análise Exploratória das Tarifas Aéreas praticadas no Brasil no Período de 2019 a 2022 e Perspectivas para o Serviço de Transporte por eVTOL", busca preencher algumas dessas lacunas e avaliar a viabilidade da operacionalização desse serviço, com base nas condições propostas por esses dois modelos e desenvolvidas nesta pesquisa.

Para alcançar esse objetivo, é imprescindível aprofundar a compreensão dos pré-requisitos e funcionalidades fundamentais necessárias para a integração do estudo. Além disso, é de extrema importância obter um entendimento abrangente do contexto no qual esses modelos serão implementados, levando em consideração cuidadosamente as restrições e oportunidades disponíveis para o serviço de transporte por eVTOL.

3. Revisão da literatura

Durante esta etapa, foram examinados estudos e materiais relevantes relacionados ao cerne da pesquisa. Conforme as considerações apresentadas por Cauchick, Paulo (2019), é essencial realizar uma revisão sistemática da literatura para obter uma compreensão aprofundada da situação-problema e facilitar a identificação de artefatos existentes que possam contribuir para sua resolução.

Destacando-se novamente o modelo proposto por Souza, Juliana (2023), que aborda a estimativa de CASK de aeronaves do tipo eVTOL, e com base na Análise Exploratória de Dados realizada por Godinho, Gabriel (2023), busca-se desenvolver um modelo de cálculo de tarifa dinâmica para o serviço de transporte com eVTOL na malha aérea brasileira. A partir dos conhecimentos adquiridos com as pesquisas citadas e demais conteúdos apresentados no referencial teórico, o objetivo é criar um artefato que possa solucionar efetivamente o problema em questão.

Esses recursos, aliados aos apresentados no referencial teórico, serão a base para a pesquisa. Dresch et al. (2015) destacam a importância dessa abordagem, pois permite acessar o conhecimento produzido em pesquisas anteriores, que podem complementar o estudo atual.

4. Identificação dos artefatos e configuração das classes de problemas

Durante esta quarta etapa do projeto, que segue a conscientização do problema e a realização da revisão sistemática da literatura, foram identificados artefatos relevantes que podem contribuir para abordar o problema em questão, assim como suas possíveis categorias. É importante destacar, conforme ressaltado por Dresch et al. (2015), a relevância de identificar artefatos existentes que possam estar direcionados ao mesmo grupo de problemas relacionados à pesquisa. Isso auxilia na formulação precisa dos novos artefatos.

Nesta etapa, dois artefatos principais foram identificados como suporte para o desenvolvimento do próprio artefato nesta pesquisa. O primeiro é o estudo "Uma proposta de estimativa de CASK de aeronaves do tipo eVTOL baseada em modelo híbrido de desempenho de aeronaves de asa rotativa e asa fixa" (Souza, Juliana, 2023). O segundo é a pesquisa "Análise Exploratória das Tarifas Aéreas praticadas no Brasil e perspectivas para o Serviço de Transporte por eVTOL" (Godinho, Gabriel, 2023). Além disso, será feito uso dos dados fornecidos pela ANAC como suporte para as análises e investigações realizadas.

Esses artefatos e dados serão fundamentais para embasar e enriquecer as análises, permitindo uma compreensão mais profunda do problema e orientando o desenvolvimento do próprio artefato.

5. Proposição de artefatos para resolver o problema específico

A quinta fase desta pesquisa consistirá na formalização das etapas anteriores, a fim de iniciar a proposição do próprio artefato. Este artefato, fundamentado nos estudos mencionados anteriormente e na correlação com o banco de dados fornecido pela ANAC nos últimos quatro anos completos, permitirá propor o desenvolvimento de uma análise de custos operacionais

diretos para o serviço de transporte por eVTOL e compreender sua viabilidade no contexto brasileiro. Além disso, essa etapa tem como objetivo fornecer uma compreensão específica do problema, permitindo a proposição de artefatos que ofereçam soluções eficazes para os desafios abordados (Dresch et al., 2015).

6. Projeto do artefato selecionado

Neste estágio da pesquisa, será identificado o artefato a ser utilizado nas etapas subsequentes da metodologia Design Science Research (DSR). O projeto incluirá a construção de um modelo para análise dos custos operacionais diretos do serviço de aviação tradicional, com foco nas variáveis de combustível. Após um estudo aprofundado da realidade emergente dos eVTOLs, será possível compreender quais requisitos avaliados são de suma importância para o desenvolvimento de um modelo de operacionalização adequado, capaz de atender às necessidades da pesquisa e, conseqüentemente, do artefato em questão. Este processo garantirá que o modelo desenvolvido seja robusto e aplicável ao contexto específico da aviação no Brasil, fornecendo insights valiosos para a implementação e gestão eficiente do serviço de transporte por eVTOL.

7. Desenvolvimento do artefato

A etapa de desenvolvimento do artefato, conforme Dresch et al. (2015), representa um momento crucial em que todas as etapas anteriores convergem para a construção e produção do produto final. Nessa fase, as especificações técnicas e o projeto detalhado são transformados em um artefato tangível. É fundamental seguir procedimentos e normas de qualidade, garantindo que o produto atenda aos requisitos estabelecidos, seja seguro, tenha alto desempenho e resolva os problemas específicos propostos.

Durante o desenvolvimento do artefato, a viabilidade econômica do projeto deve ser cuidadosamente considerada, buscando soluções que sejam financeiramente viáveis e que minimizem o impacto ambiental do serviço de transporte por eVTOL. Esta etapa também envolve a realização de testes rigorosos para garantir a conformidade dos dados do modelo, além de identificar possíveis ajustes ou melhorias. O feedback obtido a partir desses testes é essencial para aprimorar o produto antes da entrega final.

Em resumo, a etapa de desenvolvimento do artefato é quando todo o trabalho anterior se materializa, resultando em um produto concreto e funcional que atende aos requisitos técnicos, de qualidade e de sustentabilidade estabelecidos. Para isso, pode-se utilizar a interface do Google Colab juntamente com a linguagem Python para o tratamento dos dados e análise gráfica. Essa abordagem garante a precisão e a eficiência necessárias para desenvolver um artefato robusto e aplicável ao contexto da aviação e do transporte por eVTOL no Brasil.

8. Avaliação do artefato

Nesta etapa, será conduzida uma revisão detalhada dos objetivos para verificar se os resultados obtidos estão alinhados com o objetivo central da pesquisa (Dresch et al., 2015). O processo de avaliação do artefato assume um papel central, pois é o momento em que se poderá avaliar o desenvolvimento do modelo de análise dos custos operacionais diretos de combustível para a viabilidade do serviço de transporte por eVTOL.

9. Explicação das aprendizagens

Para impulsionar o avanço do conhecimento, é fundamental expor de maneira explícita os principais elementos da pesquisa após a etapa de avaliação. Nesse sentido, é crucial abordar os pontos positivos e negativos do estudo, além de fornecer uma descrição clara do processo de aprendizado adotado e das lições aprendidas durante a pesquisa. Essa abordagem visa facilitar a progressão do conhecimento, conforme sugerido por Dresch et al. (2015).

10. Conclusões

A abordagem proposta por Dresch et al. (2015) destaca a importância de explicitar os desafios enfrentados, os aprendizados adquiridos e os resultados alcançados ao longo da pesquisa. Esses elementos são essenciais para fortalecer o conhecimento existente e orientar pesquisas futuras. Ao compartilhar os aprendizados, como insights teóricos, descobertas empíricas e lições metodológicas, contribui-se para o avanço da comunidade científica. É fundamental apresentar os resultados de forma clara e objetiva, relacionando-os aos objetivos da pesquisa, para que outros pesquisadores possam compreender seu significado e contribuição para o campo de estudo. Além disso, fornecer orientações para pesquisas futuras, como extensões possíveis, questões a serem investigadas, abordagens alternativas e lacunas a serem exploradas, renova o ciclo proposto pela Design Science Research, estimulando a continuidade da pesquisa e o desenvolvimento de soluções mais eficazes e inovadoras.

11. Generalização para uma classe de problemas

A etapa em questão assume um papel crucial na garantia de que o artefato resultante da pesquisa seja altamente eficaz e aplicável a uma ampla gama de problemas. Isso requer a criação de modelos que sejam específicos para o serviço de transporte por eVTOL, com a capacidade de serem abrangentes, adaptáveis e utilizáveis em situações semelhantes provenientes de problemas similares (Purao, 2003).

12. Comunicação dos resultados

A comunicação dos resultados da pesquisa desempenha um papel fundamental na Metodologia da Pesquisa de Design Science Research, conforme defendido por Dresch et al. (2015). Por meio dessa etapa, os pesquisadores têm a oportunidade de compartilhar suas descobertas, contribuindo para o avanço do conhecimento na área. Além disso, a comunicação eficaz dos resultados permite que outros pesquisadores possam se inspirar e expandir as investigações em temas relacionados, promovendo a renovação do ciclo de pesquisas.

Segundo Cauchick (2019), a acessibilidade dos resultados da pesquisa é uma preocupação crucial. Ao tornar a pesquisa acessível tanto para a comunidade acadêmica quanto para os profissionais das organizações envolvidas com os problemas abordados pela DSR, os pesquisadores podem garantir que os conhecimentos gerados sejam aplicados de forma prática e relevante. Isso implica em apresentar os resultados de maneira clara, utilizando linguagem acessível e utilizando formatos de comunicação adequados aos diferentes públicos-alvo. Dessa forma, a comunicação dos resultados não apenas amplia o alcance dos conhecimentos gerados, mas também promove a sua aplicação efetiva, contribuindo para a solução de problemas reais enfrentados pelas organizações e impulsionando a inovação e o progresso na área.

3.3 Caracterização da organização, setor e área

A presente pesquisa consiste na análise do setor de aviação, com maior foco no conceito emergente de eVTOLs.

A aviação é um setor fundamental que engloba todas as atividades relacionadas às aeronaves e seu uso, incluindo o transporte de passageiros e carga, turismo aéreo, aviação executiva, entre outros. No contexto brasileiro, a ANAC desempenha um papel crucial na regulamentação e supervisão da aviação civil, sendo responsável por estabelecer normas e regulamentos para garantir a segurança, eficiência e qualidade dos serviços aéreos no Brasil.

Uma categoria emergente no setor da aviação é a do eVTOL, essas aeronaves elétricas se destacam e têm recebido crescente atenção e investimento devido ao seu potencial para revolucionar o transporte aéreo urbano e promover a mobilidade aérea urbana.

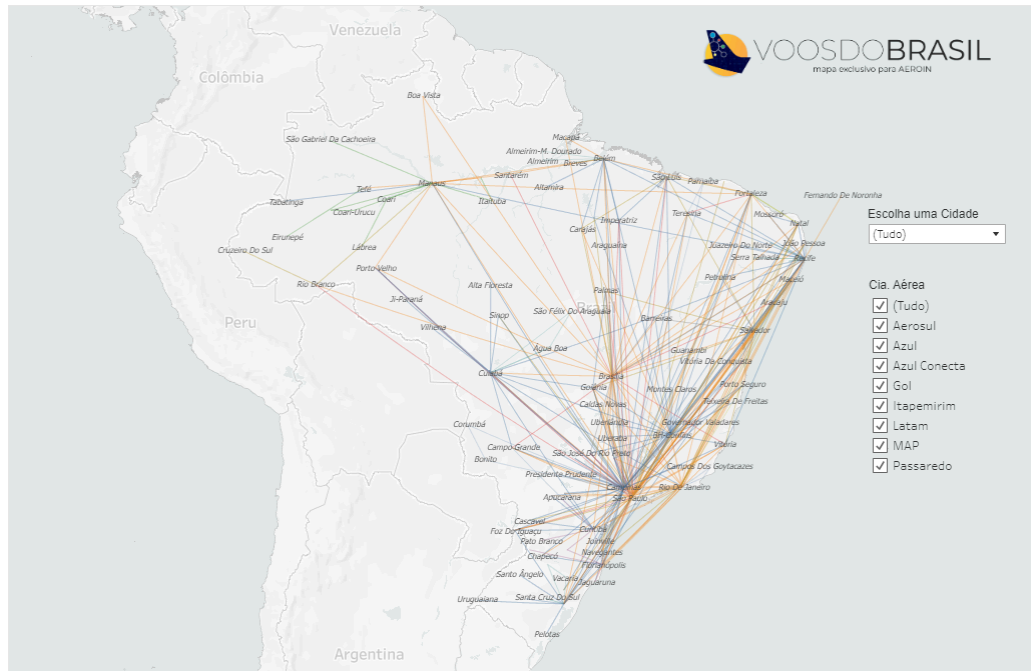
Assim, a malha aérea se correlaciona com a presente pesquisa por ser a rede de rotas e conexões estabelecidas pelas companhias aéreas para operar seus voos. Essa rede complexa de rotas conecta diversos destinos, tanto nacionais quanto internacionais, permitindo a mobilidade de passageiros e o transporte de carga. A concepção da malha aérea leva em consideração fatores como demanda de passageiros, viabilidade econômica e conectividade entre diferentes regiões e cidades.

A compreensão desses conceitos é essencial para a caracterização da organização, setor e área, proporcionando uma visão abrangente do contexto em que a empresa ou instituição está inserida. Essa análise contribui para a formulação de estratégias e tomada de decisões adequadas, levando em consideração as particularidades do setor da aviação, as regulamentações da ANAC, o potencial disruptivo dos eVTOLs e o impacto da malha aérea na conectividade e acessibilidade.

3.4 População e amostra da pesquisa

O presente estudo terá como amostra de pesquisa os voos realizados no Brasil durante 2019 e 2022, a qual será utilizado como base os dados disponibilizados pela Agência Nacional de Aviação (ANAC), tais como origem, destino, valor da tarifa e os voos. A partir disso, será possível elaborar os demais dados necessários para uma análise concisa dos custos operacionais diretos, tais como distância entre aeroportos, número de passageiros por voo, quantidade de assentos ofertados, modelo de aeronave, companhia aérea, entre outros. A figura 5 demonstra um exemplo do mapa de rotas da aviação brasileira.

Figura 5 - Mapa de Rotas da Aviação Brasileira.



Fonte: Aeroin, 2021

Para detalhar a análise, foram escolhidos modelos de aeronaves representativas de diferentes categorias de transporte aéreo. A seleção incluiu os seguintes modelos e suas respectivas capacidades máximas de passageiros e consumo médio de combustível por hora, conforme especificado nos manuais das aeronaves:

1. AEROSPATIALE/ALENIA ATR 42-300 / 320

Capacidade máxima de passageiros: 48

Consumo médio de combustível por hora: 710 litros

2. AEROSPATIALE/ALENIA ATR 42-500

Capacidade máxima de passageiros: 47

Consumo médio de combustível por hora: 1013,75 litros

3. AEROSPATIALE/ALENIA ATR 72-201/202

Capacidade máxima de passageiros: 118

Consumo médio de combustível por hora: 900 litros

4. AEROSPATIALE/ALENIA ATR 72-500/72-212A (500)

Capacidade máxima de passageiros: 72

Consumo médio de combustível por hora: 952,50 litros

5. AEROSPATIALE/ALENIA ATR 72-600/72-212A (600)

Capacidade máxima de passageiros: 72
Consumo médio de combustível por hora: 812,50 litros

6. CESSNA 208 CARAVAN

Capacidade máxima de passageiros: 9
Consumo médio de combustível por hora: 233,75 litros

Os dados foram organizados em um *dataframe* que contém 717.440 linhas e 50 colunas, permitindo uma análise detalhada e abrangente dos custos operacionais diretos de combustível das aeronaves selecionadas.

Dentre as aeronaves analisadas, o CESSNA 208 CARAVAN será o foco principal da pesquisa devido à sua menor capacidade de passageiros e à sua utilização predominante em voos de curta distância. A escolha do CESSNA 208 CARAVAN se justifica pela relevância desse modelo em operações regionais e sua representatividade em termos de eficiência de consumo de combustível em rotas curtas, proporcionando insights valiosos para a otimização de custos operacionais em contextos similares.

Figura 6 - CESSNA 208 CARAVAN



Fonte: Textron Aviation, 2024.

As aeronaves ATR 42-300 / 320 e ATR 42-500 foram desconsideradas das análises devido à insuficiência de frequência de voos, o que compromete a robustez estatística necessária para uma análise significativa. Focar nos modelos com dados suficientes garante a precisão e a relevância dos resultados obtidos nesta pesquisa.

3.5 Procedimentos de coleta de dados

Os dados coletados para a realização desta pesquisa serão obtidos através da análise de documentos e fontes secundárias disponíveis na *web*. Entre as principais fontes de coleta, destaca-se a Agência Nacional de Aviação Civil (ANAC), que disponibiliza publicamente

informações detalhadas sobre demanda, oferta, aeronaves, tarifas e voos. Além disso, serão utilizados dados fornecidos pela Petrobras, especialmente aqueles relacionados aos preços do querosene de aviação.

3.6 Análise de dados

A análise dos dados coletados será conduzida por meio de uma abordagem exploratória detalhada, utilizando a linguagem de programação Python como base. Para auxiliar no processamento, tratamento e visualização dos dados, serão utilizados diversos pacotes complementares amplamente reconhecidos e utilizados na comunidade científica e de análise de dados.

Um dos pacotes essenciais será o GeoPandas © (2013-2022), que permitirá realizar operações geoespaciais e manipulação de dados geográficos de forma eficiente. Com ele, será possível realizar análises espaciais, como a identificação de áreas de maior demanda por transporte eVTOL ou a visualização de rotas otimizadas com base em dados geográficos.

Para a visualização dos resultados e a criação de gráficos, serão utilizados o pacote Matplotlib © (2002-2023) e o Seaborn © (2012-2023). Esses pacotes oferecem uma ampla variedade de opções de visualização, desde gráficos simples até plotagens mais complexas, que permitirão uma melhor compreensão e interpretação dos dados obtidos.

Na etapa de análise, será crucial utilizar o pacote Numpy © (2005-2023), que fornece suporte para cálculos numéricos eficientes e manipulação de *arrays* multidimensionais. Ele será usado para realizar operações matemáticas e estatísticas nos dados, como cálculos de média, desvio padrão e correlação.

Por fim, o pacote PuLP © (2023) será empregado para solucionar problemas matemáticos, permitindo a formulação e resolução de modelos matemáticos que visam maximizar a eficiência das análises e, conseqüentemente, a qualidade do serviço de transporte eVTOL.

Dessa forma, a combinação desses pacotes e suas respectivas funcionalidades auxiliará na análise aprofundada dos dados coletados, na visualização dos resultados e na formulação de estratégias para melhorar o serviço de transporte por eVTOL.

4. RESULTADOS E DISCUSSÕES

Nesta seção, são apresentadas as correlações derivadas do banco de dados trabalhado na pesquisa, com ênfase nos Custos Operacionais Diretos de Combustível das aeronaves selecionadas. A análise compreende gráficos que elucidam a relação entre o consumo de combustível e as demais variáveis de distância, tempo e data para cada modelo de aeronave. Este estudo abrange o período de 2019 a 2022, permitindo uma visão detalhada das tendências e flutuações nos custos de combustível ao longo dos anos. A metodologia aplicada permite estudar o CASK Fuel, proporcionando uma base sólida para a compreensão dos fatores operacionais impactados pelo consumo de combustível.

Adicionalmente, a seção explora as comparações entre aeronaves tradicionais e modelos emergentes de eVTOL, enfatizando as diferenças em eficiência e sustentabilidade.

As subseções subsequentes oferecem uma análise detalhada de cada aeronave, identificando padrões específicos e discutindo as implicações dessas descobertas para o futuro do transporte aéreo. Esta abordagem híbrida não só permite uma avaliação crítica dos dados históricos, mas também abre caminho para considerações futuras no planejamento e operação de aeronaves, especialmente no contexto da introdução de tecnologias mais sustentáveis e inovadoras como os eVTOLs.

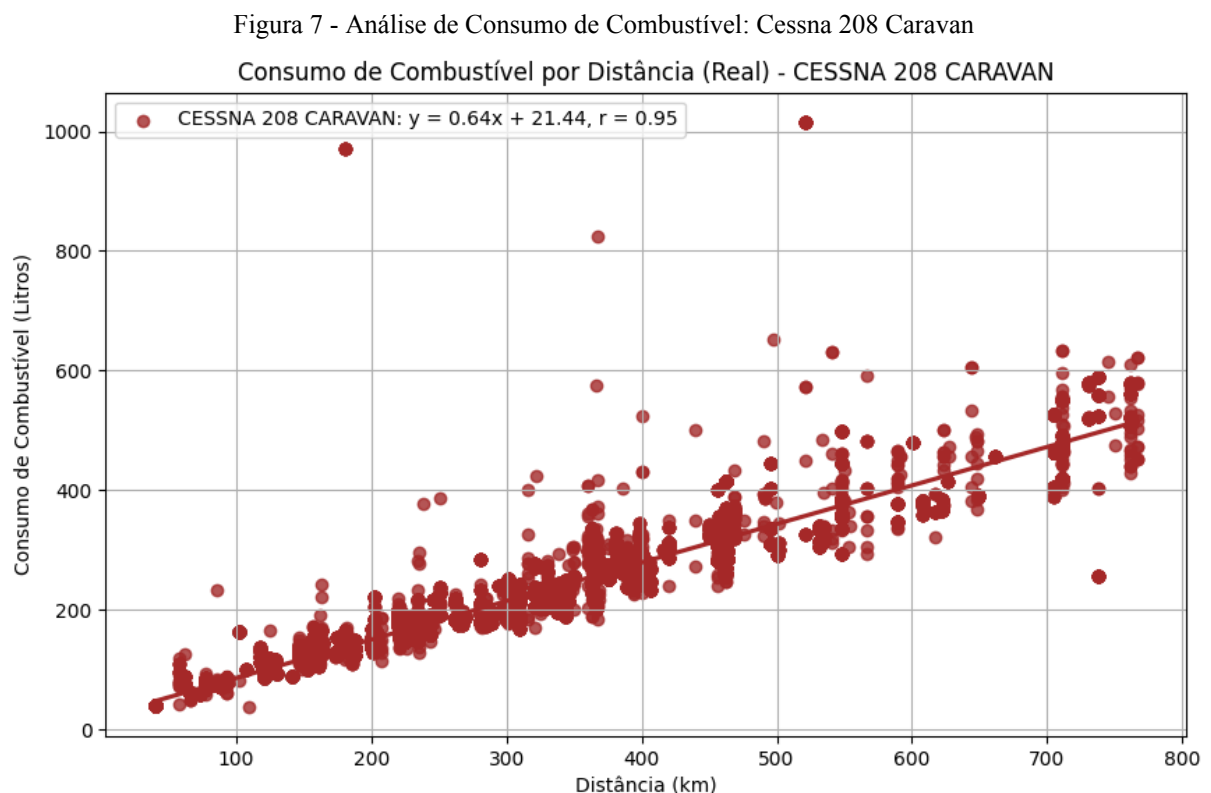
4.1 Análise de Consumo de Combustível: Consumo de Combustível e Distância

4.1.1 Introdução

Nesta seção, analisamos o consumo de combustível em relação à distância percorrida para quatro modelos de aeronaves: CESSNA 208 CARAVAN, AEROSPATIALE/ALENIA ATR 72-201/202, ATR 72-500/72-212A (500) e ATR 72-600/72-212A (600). A análise inclui gráficos de dispersão e regressão linear, fornecendo uma visão detalhada sobre a relação entre a distância percorrida e o consumo de combustível.

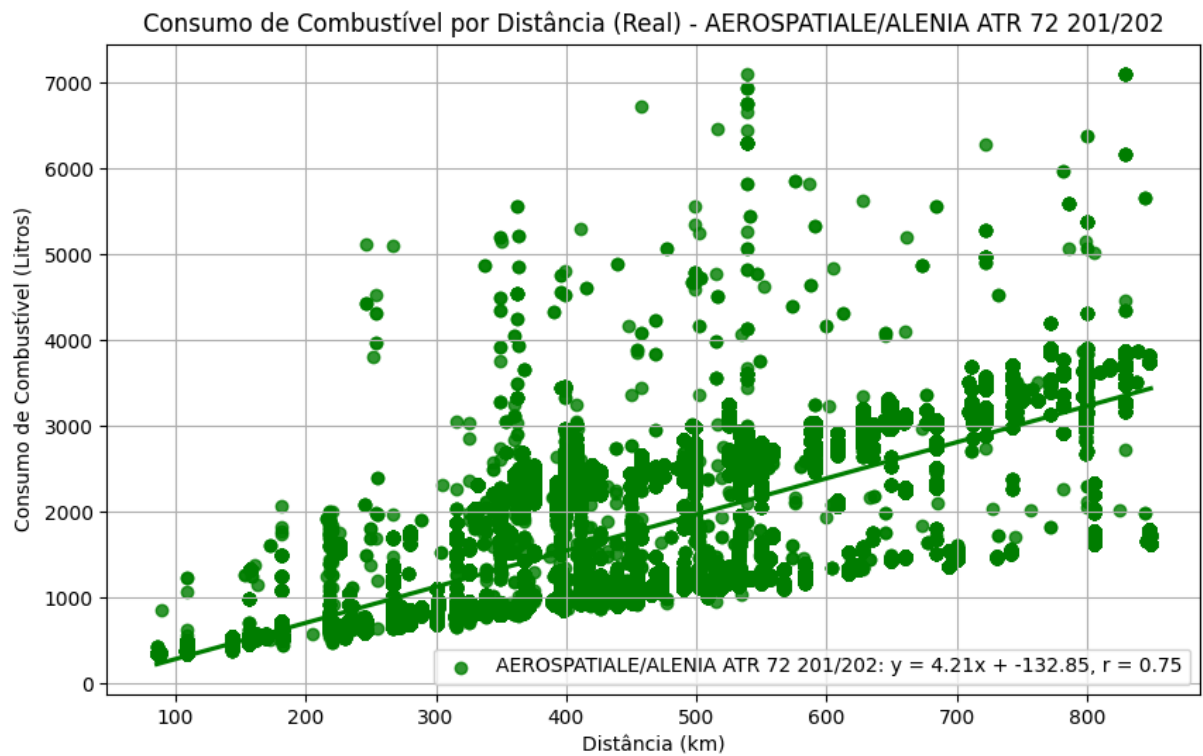
4.1.2 Análise Geral

Os gráficos a seguir mostram os gráficos de dispersão e de regressão linear do consumo de combustível pela distância percorrida para as quatro aeronaves usadas na pesquisa:



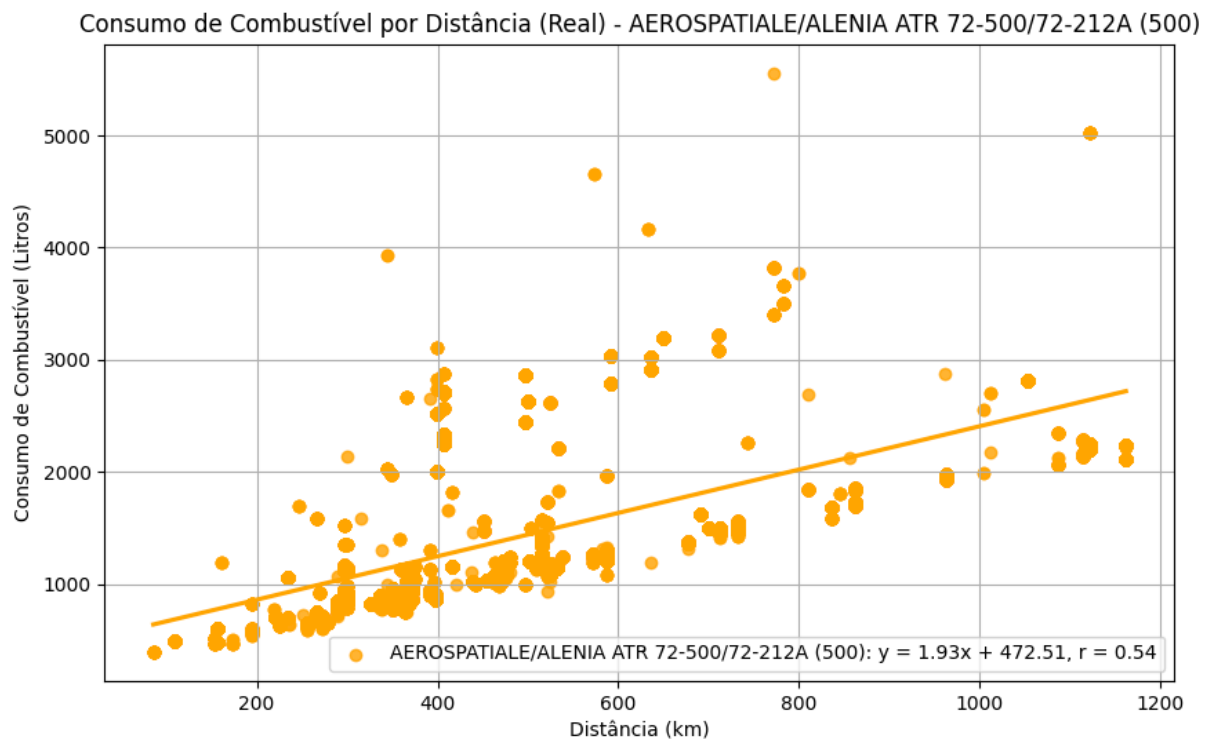
Fonte: elaboração Própria.

Figura 8 - Análise de Consumo de Combustível: AEROSPATIALE/ALENIA ATR 72 201/202



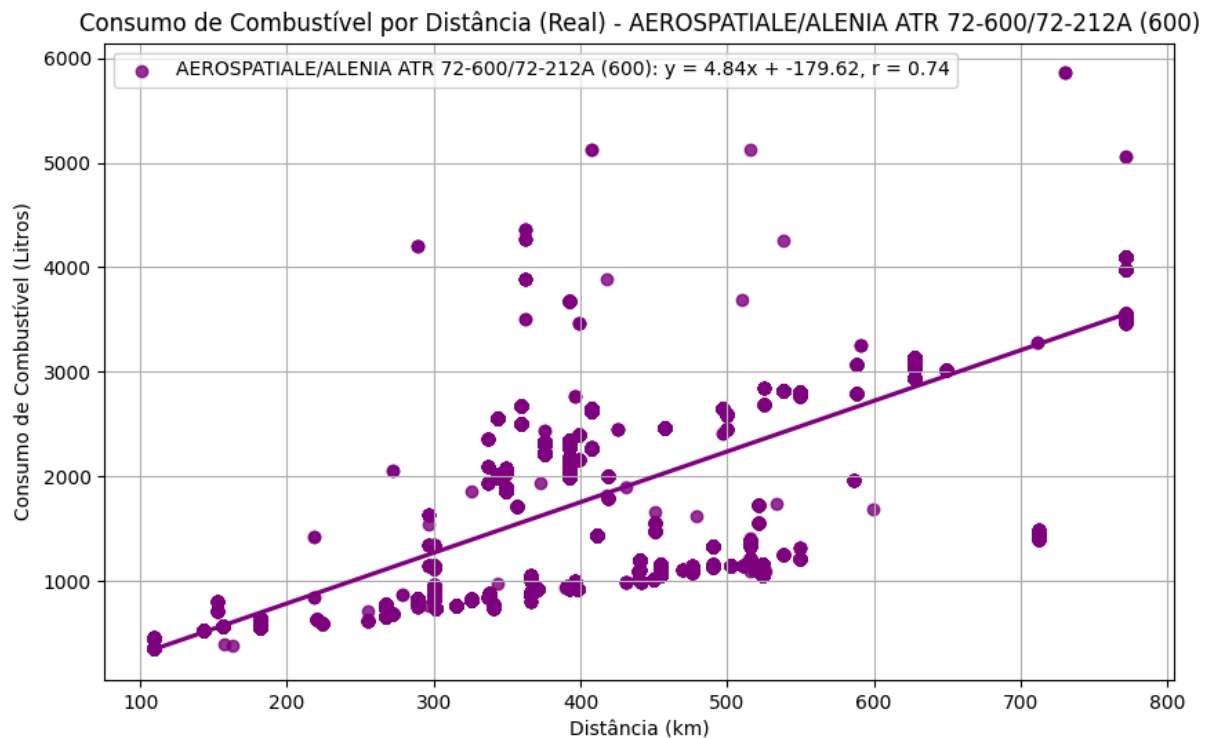
Fonte: elaboração Própria.

Figura 9 - Análise de Consumo de Combustível: AEROSPATIALE/ALENIA ATR 72 500/72-212A (500)



Fonte: elaboração Própria.

Figura 10 - Análise de Consumo de Combustível: AEROSPATIALE/ALENIA ATR 72-600/72-212A (600)



Fonte: elaboração Própria.

Os gráficos de dispersão e regressão linear apresentados para os modelos de aeronaves CESSNA 208 CARAVAN, AEROSPATIALE/ALENIA ATR 72-201/202, ATR 72-500/72-212A (500) e ATR 72-600/72-212A (600) fornecem uma visão detalhada sobre a relação entre a distância percorrida e o consumo de combustível. Cada modelo apresenta características distintas em termos de capacidade de passageiros e padrões de consumo de combustível, refletindo suas aplicações e eficiências operacionais.

4.1.3 Aeronave: Cessna 208 Caravan

O CESSNA 208 CARAVAN com capacidade para até 09 passageiros tem a linha de regressão linear ajustada aos dados representada pela equação $y = 0.64x + 21.44$, onde y é o consumo de combustível e x é a distância percorrida. O coeficiente de correlação $r = 0.95$ indica uma forte correlação positiva entre as duas variáveis, sugerindo uma relação linear robusta entre a distância do voo e o consumo de combustível.

O coeficiente angular de 0.64 indica que, em média, para cada quilômetro adicional percorrido, o consumo de combustível aumenta em 0.64 litros, desconsiderando os fatores externos que impactam no consumo, como condições climáticas e peso de carga da aeronave. Este coeficiente representa a sensibilidade do consumo de combustível em relação à distância percorrida, evidenciando que distâncias maiores demandam de mais combustível. O intercepto de 21.44, embora não tenha um significado prático direto (já que um voo com distância zero não é aplicável), pode ser interpretado como o consumo inicial necessário para realizar qualquer voo, refletindo os custos fixos de consumo de combustível inerentes à operação de decolagem e outros preparativos de voo.

O valor do coeficiente de correlação $r = 0.95$ indica uma correlação positiva muito forte entre a distância percorrida e o consumo de combustível. Este valor próximo de 1 sugere que a distância é um fator determinante no consumo de combustível do CESSNA 208 CARAVAN, permitindo prever o consumo de combustível com alta precisão baseando-se na distância percorrida. A linearidade evidenciada pela regressão reforça a consistência desta relação.

4.1.4 Aeronave: AEROSPATIALE/ALENIA ATR 72 201/202

O ATR 72-201/202 com capacidade para até 118 passageiros tem a linha de regressão linear ajustada aos dados representada pela equação $y = 4.21x - 132.85$, onde y é o consumo de combustível e x é a distância percorrida. O coeficiente de correlação $r = 0.75$ indica uma correlação positiva moderada a forte entre as duas variáveis, sugerindo que a distância percorrida é um fator importante no consumo de combustível.

O coeficiente angular de 4.21 indica que, em média, para cada quilômetro adicional percorrido, o consumo de combustível aumenta em 4.21 litros. Este coeficiente reflete a taxa de variação do consumo em função da distância, evidenciando que voos mais longos requerem proporcionalmente mais combustível. O intercepto de -132.85, embora não tenha uma interpretação prática direta (já que um consumo negativo não é possível), pode ser visto como um ajuste estatístico necessário para a linha de regressão.

O valor do coeficiente de correlação $r = 0.75$ sugere uma correlação positiva moderada a forte, indicando que a distância percorrida é um fator significativo, mas não o único, que afeta o consumo de combustível do ATR 72 201/202. A dispersão dos pontos ao redor da linha de regressão e a presença de diversos outliers, especialmente em distâncias maiores, indicam variabilidade no consumo de combustível que pode ser atribuída a outros fatores operacionais, como condições meteorológicas, peso da carga e altitude do voo.

4.1.5 Aeronave: AEROSPATIALE/ALENIA ATR 72 500/72-212A (500)

O ATR 72-500/72-212A (500) com capacidade para até 72 passageiros tem a linha de regressão linear ajustada aos dados representada pela equação $y = 1.93x + 472.51$, onde y é o consumo de combustível e x é a distância percorrida. O coeficiente de correlação $r = 0.54$ indica uma correlação positiva moderada entre as duas variáveis.

O coeficiente angular de 1.93 sugere que, em média, para cada quilômetro adicional percorrido, o consumo de combustível aumenta em 1.93 litros. Este coeficiente reflete a taxa de variação do consumo de combustível em função da distância percorrida, indicando que voos mais longos requerem proporcionalmente mais combustível. O intercepto de 472.51 litros, embora não tenha uma interpretação prática direta (já que um consumo inicial tão elevado não é realista), pode ser visto como um ajuste estatístico necessário para a linha de regressão.

O valor do coeficiente de correlação $r = 0.54$ indica uma correlação positiva moderada, sugerindo que a distância percorrida é um fator significativo no consumo de combustível do ATR 72-500/72-212A (500), mas não o único. A dispersão dos pontos ao redor da linha de regressão e a presença de diversos outliers, especialmente em distâncias maiores, indicam que

outros fatores operacionais, como condições meteorológicas, peso da carga e altitude do voo, também influenciam o consumo de combustível.

4.1.6 Aeronave: AEROSPATIALE/ALENIA ATR 72-600

O ATR 72-600/72-212A (600), com a capacidade idêntica ao modelo anterior, tem a linha de regressão linear ajustada aos dados representada pela equação $y=4.84x-179.62$, onde y é o consumo de combustível e x é a distância percorrida. O coeficiente de correlação $r=0.74$ indica uma correlação positiva moderada a forte entre as duas variáveis.

O coeficiente angular de 4.84 indica que, em média, para cada quilômetro adicional percorrido, o consumo de combustível aumenta em 4.84 litros. Este coeficiente representa a sensibilidade do consumo de combustível em relação à distância percorrida, destacando que voos mais longos demandam proporcionalmente mais combustível. O intercepto de -179.62 litros, embora não tenha uma interpretação prática direta (já que um consumo negativo não é realista), pode ser considerado um ajuste estatístico necessário para a linha de regressão.

O valor do coeficiente de correlação $r=0.74$ indica uma correlação positiva moderada a forte, sugerindo que a distância percorrida é um fator significativo no consumo de combustível do ATR 72-600/72-212A (600). A dispersão dos pontos ao redor da linha de regressão e a presença de diversos outliers, especialmente em distâncias maiores, indicam que outros fatores operacionais, como condições meteorológicas, peso da carga e altitude do voo, também influenciam o consumo de combustível.

4.1.7 Síntese

Os eVTOLs, projetados para acomodar comumente poucos passageiros em distâncias curtas, apresentam uma alternativa inovadora e potencialmente mais sustentável para o transporte aéreo, especialmente em contextos urbanos e de curta distância. Ao comparar os modelos de aeronaves tradicionais com os futuros eVTOLs, várias considerações importantes emergem, como sendo elétricos, eliminam a necessidade de combustível fóssil, reduzindo significativamente as emissões de carbono e o impacto ambiental. Em contraste, os modelos tradicionais analisados apresentam consumo de combustível que aumenta linearmente com a distância percorrida, com variações significativas dependendo da capacidade de passageiros e do modelo da aeronave.

Enquanto os modelos ATR, com capacidades máximas entre 72 e 118 passageiros, são adequados para voos de distâncias médias e longas, os eVTOLs são projetados para atender nichos de transporte mais específicos, como deslocamentos urbanos rápidos, com menor capacidade de passageiros. O CESSNA 208 CARAVAN, com capacidade para até 09 passageiros, é mais comparável em termos de aplicação para voos curtos, mas ainda depende de combustível fóssil e tem enorme correlação positiva entre combustível e distância. Além disso, os eVTOLs oferecem maior flexibilidade em termos de locais de decolagem e pouso, devido à sua capacidade de operar em espaços reduzidos para decolagens e pousos verticais.

A análise comparativa revela que, conforme a capacidade de passageiros aumenta, há um aumento proporcional no consumo de combustível por quilômetro nos modelos de aeronaves tradicionais. O CESSNA 208 CARAVAN, sendo o menor e mais eficiente em voos

curtos, apresenta alta correlação de aumento do consumo de combustível por quilômetro. Em contraste, os modelos ATR, com maior capacidade de passageiros, demonstram aumentos mais significativos no consumo de combustível por quilômetro, refletindo suas aplicações para distâncias maiores e a capacidade de transportar mais passageiros. Os futuros eVTOLs, com menor capacidade de passageiros e propulsão elétrica, apresentam uma promissora alternativa sustentável e eficiente para o transporte aéreo urbano de curta distância.

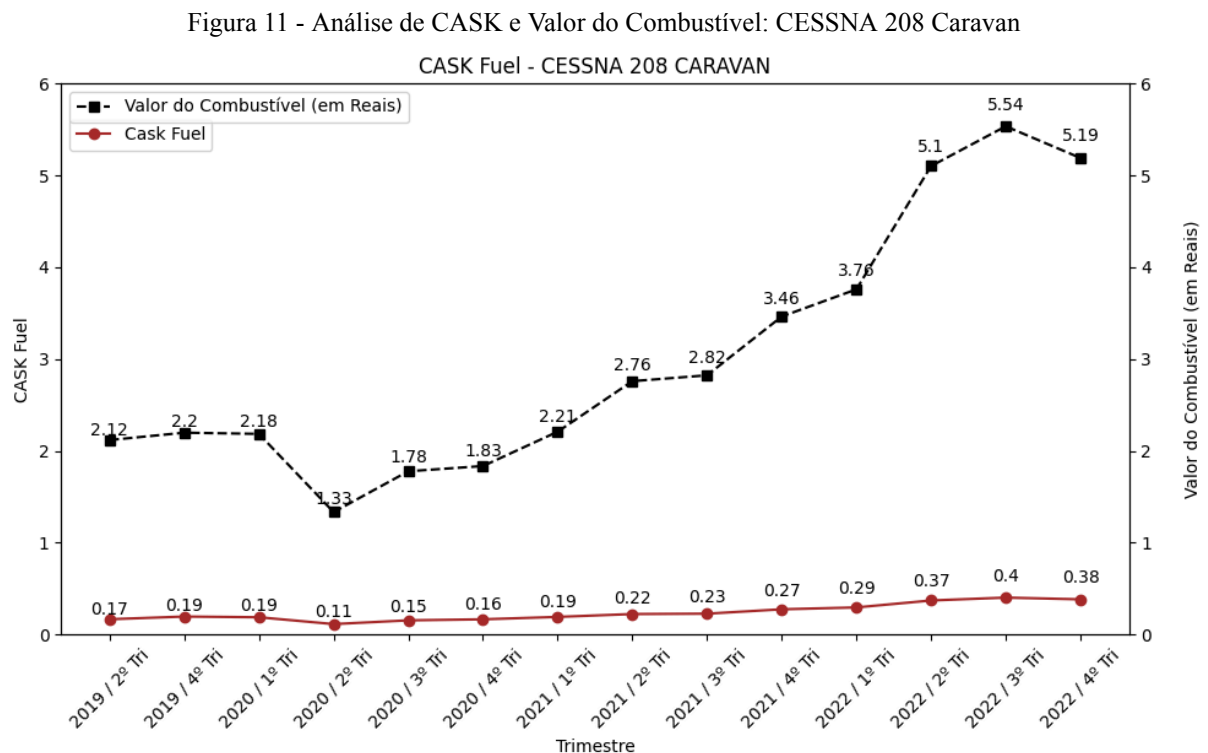
4.2 Análise de CASK e Valor do Combustível: CASK Médio por Trimestre e Valor Médio do Combustível por Trimestre

4.2.1 Introdução

A análise do CASK Fuel (*Cost Available per Seat-Kilometer*) e do valor do combustível médio é essencial para compreender a eficiência operacional e os custos das aeronaves. Neste estudo, foram analisados os gráficos de quatro modelos de aeronaves: CESSNA 208 CARAVAN, ATR 72-201/202, ATR 72-500/72-212A (500) e ATR 72-600/72-212A (600), considerando os trimestres de 2019 a 2022. Os gráficos permitem observar as tendências de consumo de combustível e suas correlações com os preços ao longo do tempo.

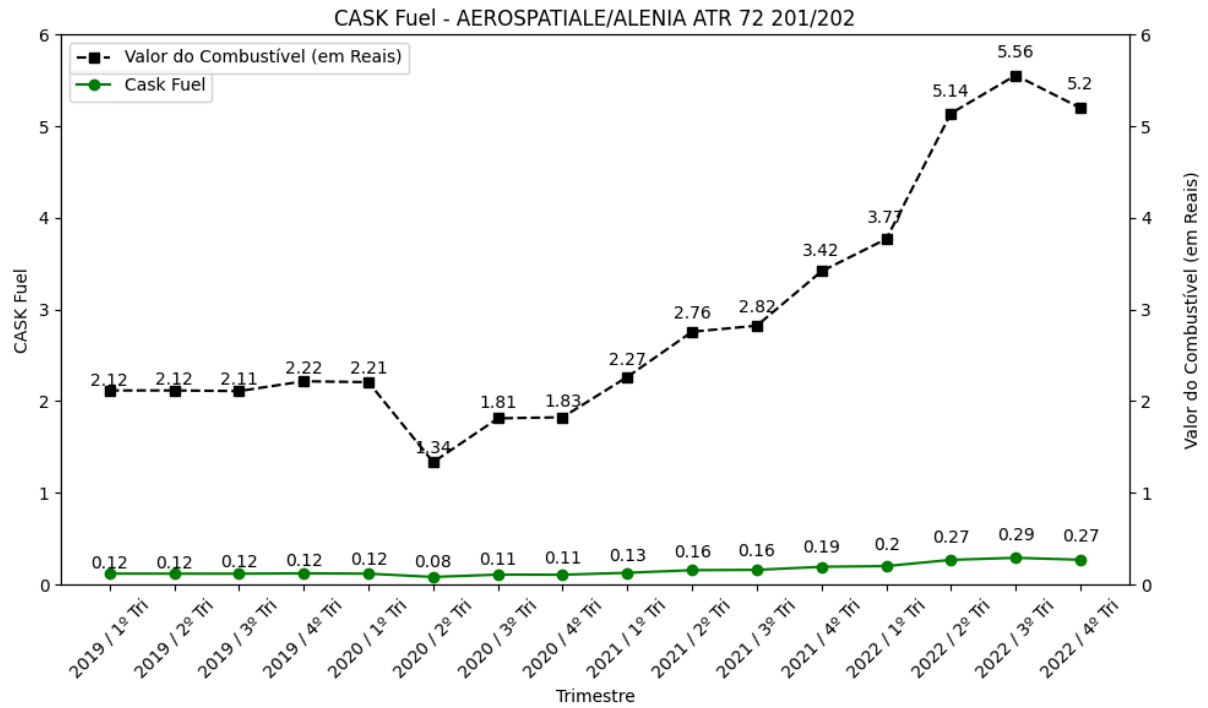
4.2.2 Análise Geral

Os gráficos a seguir mostram as tendências de CASK Fuel e valor do combustível médio para as quatro aeronaves analisadas:



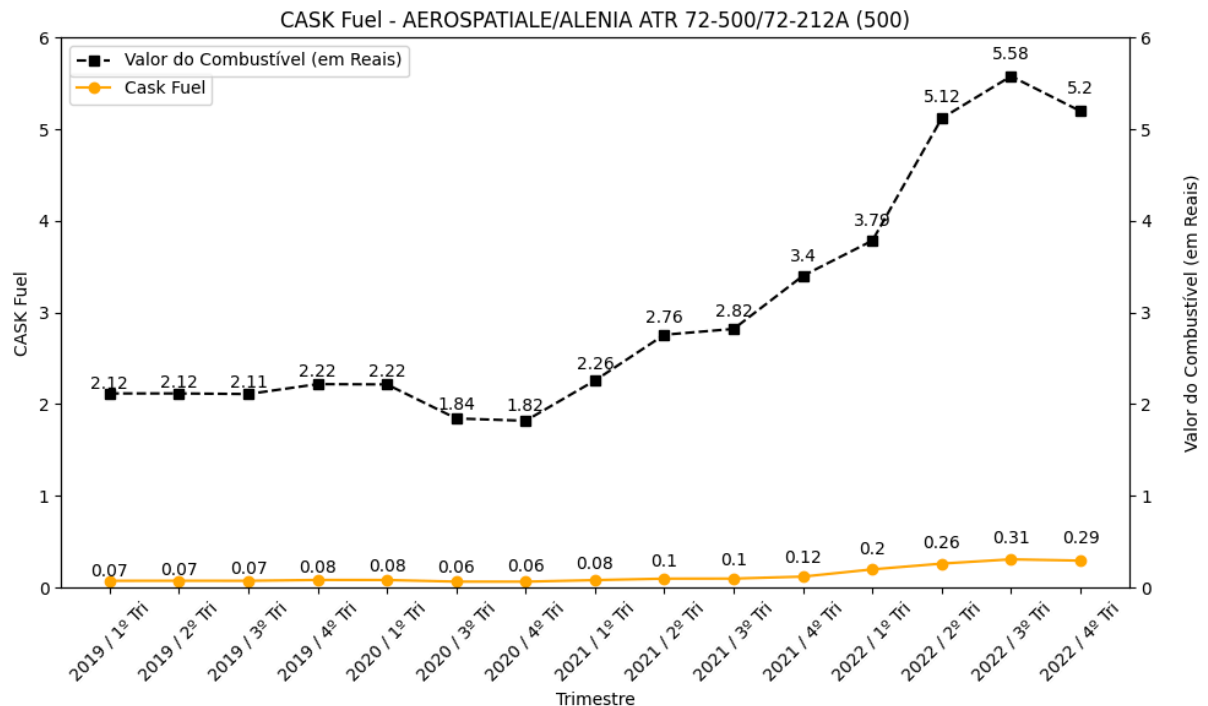
Fonte: elaboração Própria.

Figura 12 - Análise de CASK e Valor do Combustível: AEROSPATIALE/ALENIA ATR 72 201/202



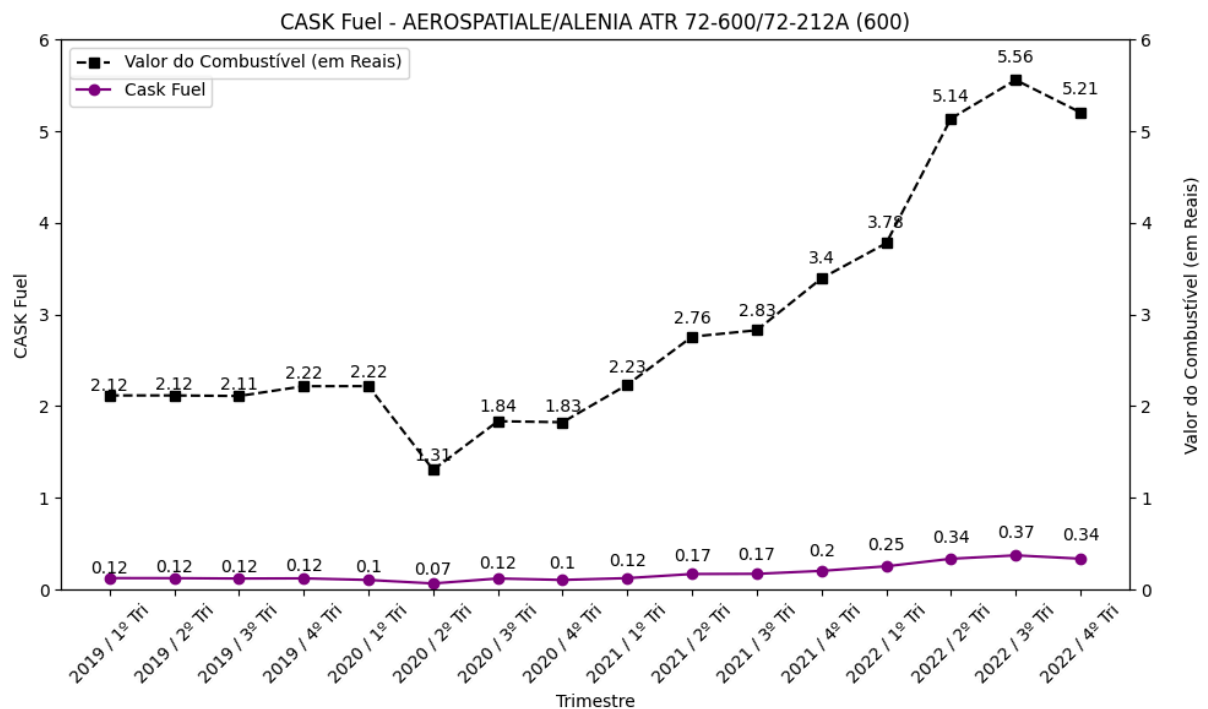
Fonte: elaboração Própria.

Figura 13 - Análise de CASK e Valor do Combustível: AEROSPATIALE/ALENIA ATR 72 500/72-212A (500)



Fonte: elaboração Própria.

Figura 14 - Análise de CASK e Valor do Combustível: AEROSPATIALE/ALENIA ATR 72 600/72-212A (600)



Fonte: elaboração Própria.

Os gráficos de CASK Fuel e valor do combustível médio mostram algumas tendências comuns entre as diferentes aeronaves. Inicialmente, o CASK Fuel manteve-se relativamente estável de 2019 a 2020 para todas as aeronaves analisadas. Este período foi marcado por uma demanda reduzida devido à pandemia de COVID-19, o que contribuiu para a estabilidade do CASK Fuel, apesar das variações nos preços do combustível.

No 2º trimestre de 2020, observou-se uma queda significativa no valor do combustível, refletindo a baixa demanda global. Essa queda foi seguida por uma recuperação gradual dos preços, com picos notáveis no início de 2022. A partir do 2º trimestre de 2021, uma correlação mais forte entre o aumento do valor do combustível e o CASK Fuel tornou-se aparente. Esse aumento sugere que as companhias aéreas começaram a sentir mais fortemente os efeitos das variações nos preços do combustível em seus custos operacionais com a recuperação econômica global e o aumento da demanda por viagens aéreas.

4.2.3 Aeronave: Cessna 208 Caravan

O CASK Fuel para o CESSNA 208 CARAVAN manteve-se notavelmente estável entre 2019 e 2020, variando entre 0,11 e 0,19. No entanto, a partir do 2º trimestre de 2021, observou-se um aumento gradual, atingindo 0,38 no 4º trimestre de 2022. Durante o 2º trimestre de 2020, o valor do combustível sofreu uma queda acentuada, passando de R\$2,20 para R\$1,33, seguido por uma recuperação com picos de R\$5,10 no 1º trimestre de 2022 e R\$5,54 no 3º trimestre de 2022. Apesar das grandes flutuações no valor do combustível, o CASK Fuel manteve-se inicialmente estável, sugerindo eficiência no uso do combustível, possivelmente devido à baixa demanda. No entanto, desde o 2º trimestre de 2021, uma correlação mais acentuada entre o aumento do valor do combustível e o CASK Fuel tornou-se evidente.

4.2.4 Aeronave: AEROSPATIALE/ALENIA ATR 72 201/202

O CASK Fuel para o ATR 72-201/202 manteve-se relativamente estável de 2019 a 2020, variando entre 0.08 e 0.13. A partir do 2º trimestre de 2021, houve um aumento gradual, atingindo 0.29 no 3º trimestre de 2022, antes de cair ligeiramente para 0.27 no 4º trimestre de 2022. O valor do combustível caiu significativamente no 2º trimestre de 2020, de R\$2,22 para R\$1,34, seguido por uma recuperação gradual com picos de R\$5,14 no 1º trimestre de 2022 e R\$5,56 no 3º trimestre de 2022. Apesar das variações no valor do combustível, o CASK Fuel manteve-se estável inicialmente, provavelmente devido à baixa demanda de voos, sugerindo eficiência no uso do combustível. No entanto, a partir do 2º trimestre de 2021, uma correlação mais forte entre o aumento do valor do combustível e o CASK Fuel tornou-se aparente.

4.2.5 Aeronave: AEROSPATIALE/ALENIA ATR 72-500/72-212A (500)

O CASK Fuel para o ATR 72-500/72-212A (500) apresentou estabilidade entre 2019 e 2020, variando entre 0.06 e 0.08. A partir do 2º trimestre de 2021, houve um aumento gradual, alcançando 0.31 no 3º trimestre de 2022, seguido por uma leve redução para 0.29 no 4º trimestre de 2022. O valor do combustível caiu significativamente no 2º trimestre de 2020, de R\$2,22 para R\$1,84, com uma subsequente recuperação gradual, atingindo picos de R\$5,12 no 1º trimestre de 2022 e R\$5,58 no 3º trimestre de 2022. Inicialmente, o CASK Fuel manteve-se estável, indicando eficiência no uso do combustível. No entanto, a partir do 2º trimestre de 2021, uma correlação mais forte entre o aumento do valor do combustível e o CASK Fuel tornou-se evidente, sublinhando a importância de estratégias de gestão de combustível e eficiência operacional.

4.2.6 Aeronave: AEROSPATIALE/ALENIA ATR 72-600/72-212A (600)

O CASK Fuel para o ATR 72-600/72-212A (600) manteve-se estável entre 2019 e 2020, variando entre 0.07 e 0.12. A partir do 2º trimestre de 2021, observou-se um aumento gradual, atingindo 0.37 no 3º trimestre de 2022, seguido por uma leve redução para 0.34 no 4º trimestre de 2022. Durante o 2º trimestre de 2020, o valor do combustível caiu drasticamente de R\$2,22 para R\$1,31, com uma subsequente recuperação gradual, atingindo picos de R\$5,14 no 1º trimestre de 2022 e R\$5,56 no 3º trimestre de 2022. Apesar das variações significativas no valor do combustível, o CASK Fuel permaneceu inicialmente estável, provavelmente devido à baixa demanda de voos durante esse período. A partir do 2º trimestre de 2021, uma correlação mais forte entre o aumento do valor do combustível e o CASK Fuel tornou-se evidente, ressaltando a importância de uma gestão eficiente dos custos de combustível.

4.2.7 Síntese

A análise dos gráficos de CASK Fuel e valor do combustível médio para as aeronaves CESSNA 208 CARAVAN, ATR 72-201/202, ATR 72-500/72-212A (500) e ATR

72-600/72-212A (600) revela que, apesar das variações significativas nos preços do combustível, todas as aeronaves mantiveram um CASK Fuel relativamente estável até 2020, início da pandemia de COVID-19. A partir do 2º trimestre de 2021, uma correlação mais forte entre o aumento do valor do combustível e o CASK Fuel tornou-se aparente, evidenciando a importância de estratégias de gestão de combustível e eficiência operacional para mitigar os impactos dos preços voláteis do combustível. A recuperação econômica global e o aumento da demanda por viagens aéreas intensificaram o uso das aeronaves, reforçando a necessidade de uma gestão eficiente dos custos operacionais diretos relacionados aos combustíveis.

4.3 Análise de DOC: Consumo de Combustível e Valor Médio do Combustível por Distância

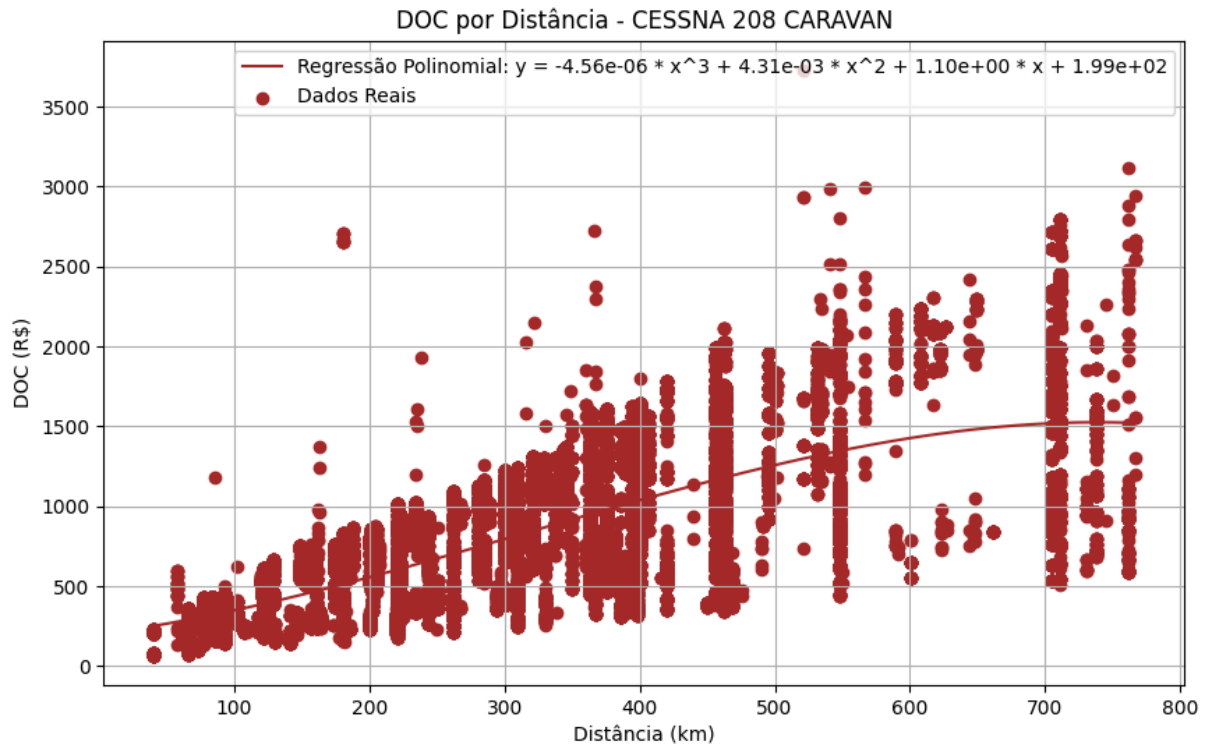
4.3.1 Introdução

Nesta seção, apresentamos as correlações derivadas dos dados de consumo operacional direto de combustível (DOC) e as distâncias percorridas pelas aeronaves estudadas. Utilizando gráficos de regressão polinomial, analisamos o comportamento dos DOCs em relação às distâncias percorridas para os modelos de aeronaves AEROSPATIALE/ALENIA ATR 72-201/202, ATR 72-500/72-212A (500), ATR 72-600/72-212A (600) e CESSNA 208 CARAVAN. Esta análise é essencial para identificar padrões e tendências que podem contribuir significativamente para o desenvolvimento deste trabalho, no entanto, é importante ressaltar que alguns aspectos relevantes, como a densidade do ar, o peso da carga de combustível, a rota aérea e os desvios climáticos, não foram considerados nesta análise.

4.3.2 Análise Geral

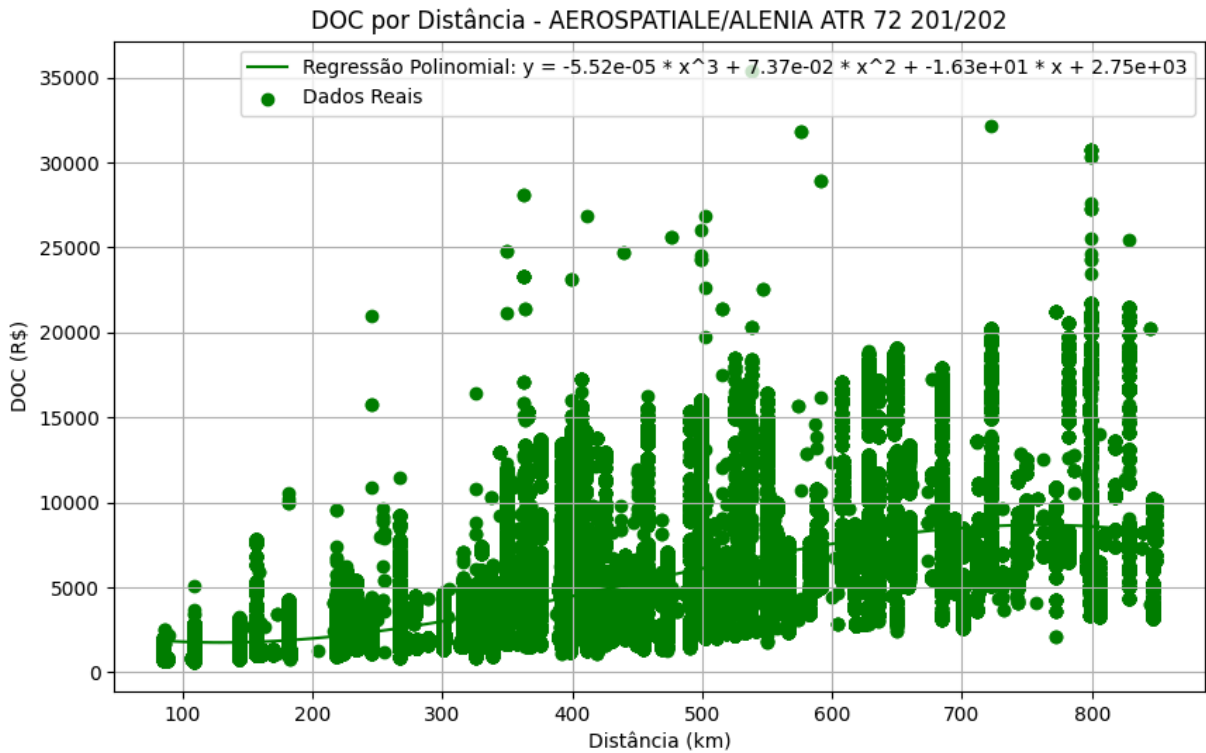
Os gráficos a seguir mostram as regressões polinomiais de DOC Combustível por distância para as quatro aeronaves analisadas:

Figura 15 - Análise de DOC por Distância: CESSNA 208 Caravan



Fonte: elaboração Própria.

Figura 16 - Análise de DOC por Distância: AEROSPATIALE/ALENIA ATR 72 201/202



Fonte: elaboração Própria.

Figura 17 - Análise de DOC por Distância: AEROSPATIALE/ALENIA ATR 72 500/72-212A (500)

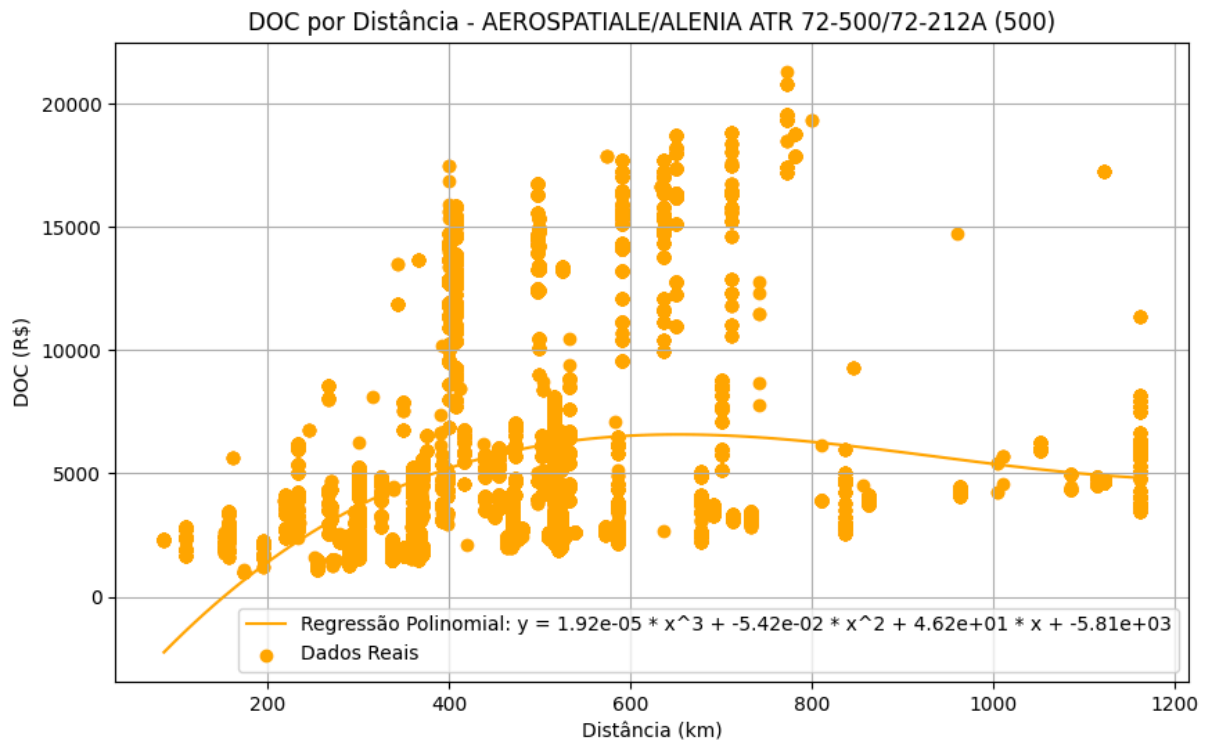
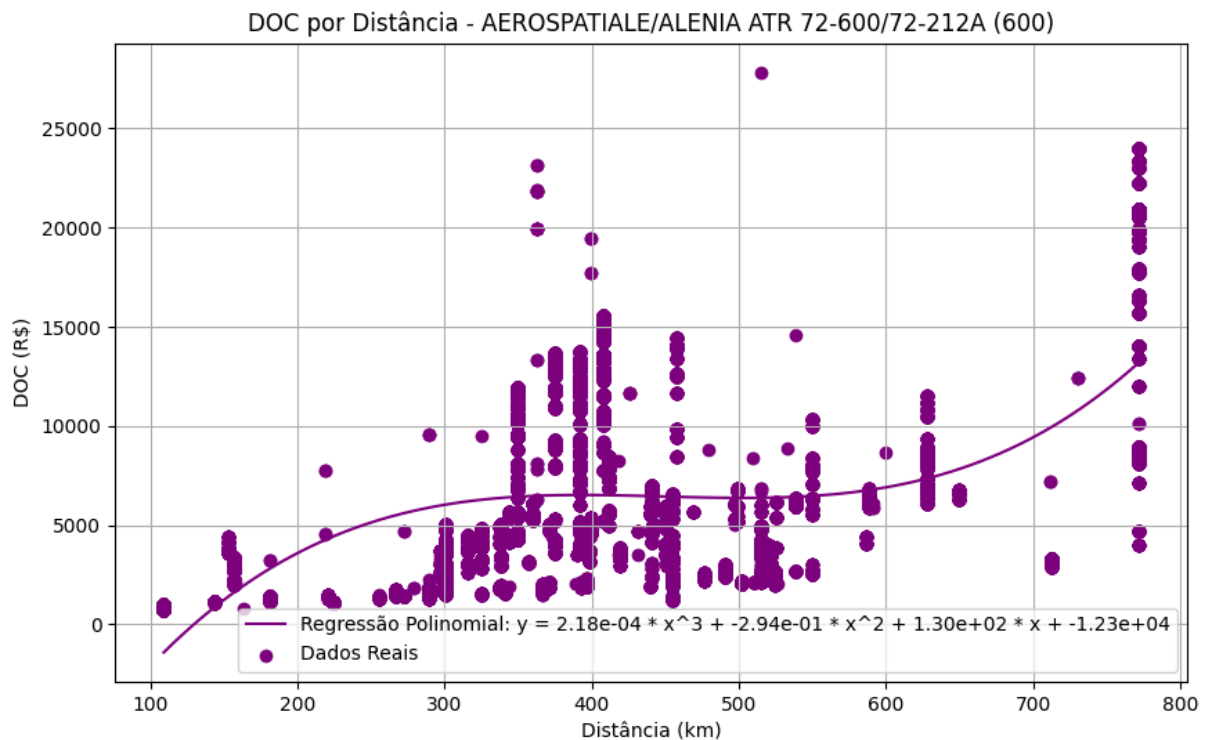


Figura 18 - Análise de DOC por Distância: AEROSPATIALE/ALENIA ATR 72 600/72-212A (600)



Os gráficos de regressão polinomial apresentados fornecem uma visão abrangente sobre a relação entre a distância percorrida e os Custos Operacionais Diretos de Combustível (DOC) para diferentes modelos de aeronaves. A escolha de uma regressão polinomial de grau

3 permite capturar a complexidade das variações de consumo de combustível, oferecendo uma modelagem mais precisa em comparação com regressões lineares ou de menor grau. Cada gráfico ilustra como o DOC varia com a distância percorrida, destacando tanto padrões lineares quanto não lineares nos dados.

Os modelos de aeronaves analisados demonstram comportamentos distintos em termos de consumo de combustível, refletindo suas capacidades de passageiros e eficiências operacionais. As curvas polinomiais mostram que, para distâncias menores, algumas aeronaves mantêm um aumento quase linear no consumo de combustível, enquanto outras exibem uma curvatura significativa que sugere um aumento exponencial ou desaceleração na taxa de crescimento do DOC.

Em geral, os resultados indicam que, à medida que a distância aumenta, o DOC tende a crescer, mas a taxa desse crescimento varia entre os diferentes modelos de aeronaves. Esta variação pode ser atribuída a diversos fatores, incluindo a capacidade de passageiros, o design aerodinâmico, a eficiência dos motores e as práticas operacionais específicas de cada aeronave.

4.3.3 Aeronave: Cessna 208 Caravan

O gráfico de regressão polinomial para o CESSNA 208 CARAVAN revela uma relação quase linear entre a distância e o DOC em distâncias menores, com uma ligeira curvatura positiva em distâncias maiores. A equação ajustada mostra que o consumo de combustível aumenta de forma consistente com a distância, refletindo a eficiência desse modelo para voos de curta a média distância. A tendência indica que, apesar de ser uma aeronave menor, o CESSNA 208 CARAVAN mantém um perfil de consumo de combustível previsível e gerenciável, essencial para operações regulares em rotas regionais.

4.3.4 Aeronave: AEROSPATIALE/ALENIA ATR 72 201/202

Para o ATR 72-201/202, a regressão polinomial mostra uma curva mais acentuada em comparação com o CESSNA 208 CARAVAN. A equação polinomial indica um aumento significativo no DOC à medida que a distância percorre aumenta, o que é esperado devido à maior capacidade de passageiros e à utilização em rotas mais longas. A curva sugere que, além do aumento linear, há um componente exponencial no consumo de combustível, destacando a importância de otimizar a operação dessas aeronaves para maximizar a eficiência.

4.3.5 Aeronave: AEROSPATIALE/ALENIA ATR 72-500/72-212A (500)

A análise do ATR 72-500/72-212A (500) mostra uma curva de regressão com um coeficiente angular mais baixo inicialmente, mas que se acentua em distâncias maiores. Isso indica que o consumo de combustível é relativamente baixo em voos curtos, mas aumenta significativamente à medida que a distância cresce. A equação ajustada reflete essa variação, sugerindo que a aeronave é eficiente para distâncias curtas a médias, mas pode se tornar

menos competitiva em rotas muito longas sem uma gestão adequada do consumo de combustível.

4.3.6 Aeronave: AEROSPATIALE/ALENIA ATR 72-600/72-212A (600)

O ATR 72-600/72-212A (600) apresenta uma curva de regressão polinomial que combina elementos de aumento linear e exponencial do DOC com a distância. A equação ajustada mostra que, enquanto a aeronave é eficiente em distâncias curtas, o consumo de combustível aumenta de forma mais pronunciada em distâncias maiores. Isso sugere que, embora o ATR 72-600 seja uma escolha robusta para uma variedade de rotas, sua eficiência pode ser melhorada através de estratégias de otimização de consumo de combustível, especialmente em voos mais longos.

4.3.7 Síntese

A análise dos modelos de aeronaves mostra que, enquanto todas as aeronaves apresentam um aumento no DOC com a distância, a taxa desse aumento varia significativamente entre os modelos. Os gráficos de regressão polinomial fornecem uma compreensão clara das tendências de consumo de combustível, destacando a necessidade de estratégias de gestão de combustível para maximizar a eficiência operacional. Em comparação com os modelos tradicionais, os eVTOLs prometem uma maior eficiência, principalmente em rotas curtas e urbanas, devido ao seu uso de propulsão elétrica. A análise detalhada dos modelos de aeronaves fornece uma base sólida para futuras pesquisas e planejamento estratégico no setor de transporte aéreo.

4.4 Análise da Operacionalização do Serviço de Transporte por eVTOL

4.4.1 Introdução

A análise qualitativa da viabilidade da operacionalização do serviço de transporte por eVTOL envolve uma avaliação detalhada dos custos operacionais diretos (DOC) e do custo médio por assento-quilômetro (CASK) das aeronaves tradicionais em comparação com os eVTOLs. Nesta seção, exploramos os requisitos necessários para que os eVTOLs possam ser implementados de forma eficiente e competitiva no mercado brasileiro. A análise é fundamentada nos dados de consumo de combustível e características operacionais dos modelos de aeronaves tradicionais, com foco no CESSNA 208 Caravan, devido à sua relevância como aeronave de voos curtos e baixa capacidade de passageiros.

4.4.2 Comparação com o CESSNA 208 Caravan

A análise dos gráficos dispostos nas seções 4.1, 4.2 e 4.3 da aeronave CESSNA 208 Caravan revela os seguintes pontos:

1. **CASK Fuel:** O gráfico de CASK Fuel mostra que o valor de combustível para o CESSNA 208 Caravan aumentou significativamente ao longo dos anos, refletindo a volatilidade dos preços do querosene de aviação. Em contrapartida, o CASK Fuel

apresentou um aumento gradual, indicando que os custos operacionais por assento-quilômetro também cresceram com o tempo. Os eVTOLs, por outro lado, dependem de energia elétrica, que pode ser gerada a partir de fontes renováveis, reduzindo a dependência de combustíveis fósseis e a volatilidade associada aos preços do querosene de aviação. Isso pode proporcionar uma maior estabilidade nos custos operacionais e contribuir para a sustentabilidade ambiental.

2. **DOC por Distância:** O gráfico de DOC por distância mostra uma relação linear entre a distância percorrida e o custo de combustível. O CESSNA 208 Caravan apresenta um aumento contínuo no DOC à medida que a distância aumenta, com uma taxa de crescimento proporcional. Este comportamento linear sugere que a aeronave é eficiente em distâncias curtas a médias, mas os custos aumentam significativamente em rotas mais longas. Em comparação, os eVTOLs têm o potencial de serem mais eficientes em distâncias curtas devido à sua propulsão elétrica e capacidade de decolagem e pouso vertical, eliminando a necessidade de longas pistas de decolagem e aterrissagem. Isso pode resultar em menores custos operacionais diretos para rotas curtas e urbanas, onde a infraestrutura aeroportuária tradicional é limitada ou inexistente.

4.4.3 Requisitos para a Operacionalização dos eVTOLs

Para que os eVTOLs sejam operacionais de forma eficiente e competitiva, é crucial que atendam aos seguintes requisitos:

1. **Menor CASK:** Os eVTOLs devem apresentar um CASK significativamente menor em comparação com o CESSNA 208 Caravan. Isso pode ser alcançado através da eficiência energética da propulsão elétrica e da redução dos custos operacionais.
2. **Menor DOC:** O DOC dos eVTOLs precisa ser competitivo, especialmente em distâncias curtas, onde o CESSNA 208 Caravan mostra maior eficiência. A capacidade de manter baixos custos operacionais diretos em voos curtos será um diferencial crucial.
3. **Sustentabilidade e Eficiência:** Além dos custos, os eVTOLs devem ser mais sustentáveis, com menor emissão de poluentes e menor impacto ambiental. A eficiência energética e a capacidade de operar em áreas urbanas densamente povoadas sem causar ruído excessivo são vantagens importantes.
4. **Flexibilidade Operacional:** A capacidade de decolar e pousar verticalmente, eliminando a necessidade de infraestrutura aeroportuária extensa, é um dos maiores benefícios dos eVTOLs. Esta flexibilidade pode resultar em maior conectividade regional e acesso a áreas de difícil alcance.

4.4.4 Síntese

A presente análise qualitativa demonstra que, para competir efetivamente com aeronaves tradicionais como o CESSNA 208 Caravan, os eVTOLs devem não apenas igualar,

mas superar esses modelos em termos de custos operacionais e eficiência. A implementação bem-sucedida dos eVTOLs no Brasil dependerá da capacidade de atender a esses requisitos e proporcionar vantagens competitivas significativas. Esta avaliação servirá como norteadora para as estratégias de operacionalização e implementação dos serviços de transporte por eVTOL, promovendo um avanço sustentável e eficiente na mobilidade aérea urbana e regional do país.

5. CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

5.1 Considerações Finais

A presente pesquisa teve como objetivo principal compreender e comparar os custos operacionais diretos de combustível dos serviços de transporte por veículos elétricos de decolagem e pouso vertical (eVTOL) nas principais malhas aéreas do Brasil. A análise dos dados revelou padrões significativos na relação entre a distância percorrida e o consumo de combustível para diferentes modelos de aeronaves. Este estudo se alinha com as tendências globais de sustentabilidade e inovação tecnológica no setor de aviação, conforme discutido por Roland Berger (2018), que enfatiza a importância de novas tecnologias para reduzir custos e emissões.

Os achados desta pesquisa são fundamentais para a implementação eficiente dos serviços de eVTOL, contribuindo para uma compreensão mais profunda dos custos operacionais e promovendo a competitividade e sustentabilidade do setor de transporte aéreo no Brasil. A identificação de padrões no consumo de combustível em relação à distância percorrida permite otimizar as operações e planejar rotas mais eficientes, potencialmente reduzindo os custos operacionais.

Além disso, a análise revelou que os eVTOLs possuem um potencial significativo para transformar a mobilidade urbana, oferecendo uma solução inovadora e eficiente para o transporte aéreo em áreas metropolitanas densamente povoadas. De acordo com Thipphavong et al. (2018), a integração de eVTOLs no espaço aéreo urbano pode melhorar a eficiência do transporte, reduzir os tempos de viagem e aliviar a congestão no trânsito terrestre.

Os resultados obtidos destacam a relevância dos eVTOLs não apenas como uma alternativa viável aos modelos tradicionais de aviação, mas também como uma solução que pode atender às necessidades emergentes de mobilidade urbana sustentável. A adoção de eVTOLs pode promover a competitividade no setor de transporte aéreo, ao mesmo tempo em que contribui para a redução das emissões de carbono, alinhando-se com as metas globais de sustentabilidade ambiental (Roland Berger, 2018).

5.2 Objetivos da Pesquisa

Os objetivos desta pesquisa foram amplamente alcançados. Foi possível compreender as diferentes variáveis que impactam os custos de operações aéreas, identificar a correlação entre o consumo de combustível e a distância, e investigar a relação do custo médio por assento-quilômetro (CASK) e do valor de combustível ao longo dos anos analisados. A comparação entre os modelos de custos operacionais diretos de combustível dos voos tradicionais e dos eVTOLs proporcionou insights valiosos para avaliar a viabilidade

econômica e operacional desses novos modelos de transporte. A análise confirmou a relevância e a aplicabilidade dos eVTOLs no contexto atual e futuro da mobilidade aérea no Brasil, mostrando que, com uma gestão eficiente, eles podem se tornar uma alternativa viável e competitiva em relação aos modelos tradicionais, cujo custo de combustível está intrinsecamente ligado à sua atividade operacional.

5.3 Desafios da Pesquisa

Durante a realização desta pesquisa, foram enfrentados diversos desafios. A coleta e o tratamento de dados precisos e abrangentes constituíram uma tarefa complexa, devido à limitação de dados disponíveis e à necessidade de calcular e integrar informações de diferentes fontes. A comparação dos custos operacionais diretos de combustível exigiu a adaptação de metodologias para se adequar às especificidades dos eVTOLs e à falta de dados históricos desse novo modelo ainda em desenvolvimento. Além disso, a presença de outliers nos dados coletados representou um desafio significativo, pois poderiam distorcer as análises e os resultados. Para mitigar esse problema, utilizamos técnicas eficientes de tratamento e análise de dados para identificar e tratar os outliers, garantindo a integridade e a validade das conclusões obtidas.

Esses obstáculos foram superados com uma abordagem metodológica rigorosa e um esforço colaborativo contínuo, embora tenham impactado o tempo e os recursos disponíveis para o estudo. A experiência adquirida durante o processo destacou a importância de um planejamento detalhado e da flexibilidade para adaptar metodologias conforme necessário. A revisão da literatura reforçou a necessidade de adotar práticas rigorosas na coleta e tratamento de dados, conforme discutido por Osborne e Overbay (2004), que destacam a importância de identificar e tratar outliers para melhorar a precisão das análises estatísticas.

5.4 Direcionamentos para trabalhos futuros

Os resultados desta pesquisa abrem várias possibilidades para estudos futuros. Investigações adicionais podem focar na análise de outros componentes dos custos operacionais tanto diretos quanto indiretos para eVTOLs e aeronaves tradicionais. Além disso, a análise de viabilidade econômica pode ser aprofundada com a inclusão de variáveis para precificação, políticas de subsídios governamentais, impactos ambientais, entre outros. Estudos comparativos entre diferentes tecnologias de propulsão e modelos de negócios também são recomendados para fornecer uma visão mais abrangente sobre o futuro do transporte aéreo urbano e regional. A integração de dados de simulação com dados reais, conforme a tecnologia de eVTOLs avança, poderá oferecer insights ainda mais precisos. Futuras pesquisas poderão explorar a percepção dos usuários e a aceitação do mercado para serviços de eVTOL, contribuindo para a construção de um ecossistema robusto e sustentável para a mobilidade aérea urbana.

REFERÊNCIAS

AARONS, Richard N. Special Report: Cessna's Caravans. July 1988. Disponível em: <https://d16bsf97ryvc45.cloudfront.net/Media/2012/10/caravan.pdf>. Acesso em: 05 mai. 2024.

AEROIN. Disponível em: <<https://aeroin.net/rotas-domestico/>>. Acesso em 08 dez. 2023.

ALENCAR, V. A. et al. Characterizing client usage patterns and service demand for car-sharing systems. Information Systems, 2019. DOI: 10.1016/j.is.2019.101448. Disponível em: <<https://doi.org/10.1016/j.is.2019.101448/>>. Acesso em: 09 dez. 2023.

ATR. AEROSPATIALE/ALENIA ATR 72-600/72-212A (600). 2022. Disponível em: https://www.atr-aircraft.com/wp-content/uploads/2022/06/ATR_Fiche72-600-3.pdf?_gl=1*a19r3q*_up*MQ..*_ga*MjExMzAxMTQyNy4xNzE1MDM0ODE4*_ga_C5YPS486CN*MTcxNTAzNDgxNy4xLjAuMTcxNTAzNDgxNy4wLjAuMA... Acesso em: 05 mai. 2024

ATR. AEROSPATIALE/ALENIA ATR 72-500/72-212A (500). 2020. Disponível em: <https://www.atr-aircraft.com/wp-content/uploads/2020/07/72-500.pdf>. Acesso em: 05 mai. 2024

ATR. AEROSPATIALE/ALENIA ATR 42-500. 2020. Disponível em: <https://www.atr-aircraft.com/wp-content/uploads/2020/07/42-500.pdf>. Acesso em: 05 mai. 2024

ATR. AEROSPATIALE/ALENIA ATR 42-300 / 320. 2020. Disponível em: <https://www.atr-aircraft.com/wp-content/uploads/2020/07/42-300.pdf>. Acesso em: 05 mai. 2024

ATR. AEROSPATIALE/ALENIA ATR 72-201 / 202. 2020. Disponível em: <https://www.atr-aircraft.com/wp-content/uploads/2020/07/72-200.pdf>. Acesso em: 05 mai. 2024

BRIS, Gaël et al. Preparing Your Airport for Electric Aircraft and Hydrogen Technologies. National Academies of Sciences, Engineering, and Medicine. Washington, DC: The National Academies Press, 2022. Ebook. Disponível em: <<https://doi.org/10.17226/26512>>. Acesso em: 24 nov. 2023.

CAUCHICK, Paulo. Metodologia Científica para Engenharia. 1ª edição. Rio de Janeiro: Elsevier Editora Ltda, 2019. E-book. ISBN 978-85-352-9071-4. Disponível em: <<https://integrada.minhabiblioteca.com.br/#/books/9788595150805/>>. Acesso em: 24 nov. 2023.

CAUCHICK, Paulo. Metodologia de Pesquisa em Engenharia de Produção e Gestão de Operações. 3ª edição. Rio de Janeiro: Elsevier Editora Ltda, 2018. E-book. ISBN 978-85-352-9135-3. Disponível em: <<https://integrada.minhabiblioteca.com.br/#/books/9788595150805/>>. Acesso em: 24 nov. 2023.

DRESCH, Aline; LACERDA, Daniel P.; JÚNIOR, José A. V A. Design science research: método de pesquisa para avanço da ciência e tecnologia. Grupo A, 2015. E-book. ISBN 9788582605530. Disponível em: <<https://integrada.minhabiblioteca.com.br/#/books/9788582605530/>>. Acesso em: 24 nov. 2023.

FERREIRA, Rafael G C.; MIRANDA, Leandro B. A de; PINTO, Rafael A.; et al. Preparação e Análise Exploratória de Dados. Porto Alegre: Grupo A, 2021. E-book. ISBN 9786556902890. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9786556902890/>. Acesso em: 15 jun. 2023.

G1. 'Carro voador': veja projetos de eVTOLs que podem ganhar os céus nos próximos anos. G1, São Paulo, 17 de dezembro 2023. Disponível em: <<https://g1.globo.com/inovacao/noticia/2023/12/17/carro-voador-veja-projetos-de-evtols-que-podem-ganhar-os-ceus-nos-proximos-anos.ghtml/>>. Acesso em 18 dez. 2023.

GODINHO, Gabriel. Análise Exploratória das Tarifas Aéreas praticadas no Brasil no Período de 2019 a 2022 e Perspectivas para o Serviço de Transporte por eVTOL. Trabalho de Conclusão de Curso (TCC). 11 dezembro 2023.

GEOPANDAS. Disponível em: <<https://geopandas.org/en/stable/>>. Acesso em: 08 dez. 2023.

GREGOR, Shirley; HEVNER, Alan. Positioning and Presenting Design Science Research for Maximum Impact. MIS Quarterly, v. 37, p. 337-356, 2013. DOI: 10.25300/MISQ/2013/37.2.01. Disponível em: https://www.researchgate.net/publication/262350911_Positioning_and_Presenting_Design_Science_Research_for_Maximum_Impact/. Acesso em: 16 dez. 2023.

HEVNER, Alan; MARCH, Salvatore; PARK, Jinsoo; RAM, Sudha. Design Science in Information Systems Research. Management Information Systems Quarterly, v. 28, p. 75- , 2004. Disponível em: https://www.researchgate.net/publication/201168946_Design_Science_in_Information_Systems_Research/. Acesso em: 16 dez. 2023.

HILLIER, F. S.; LIEBERMAN, G. J. Introduction to Operations Research. McGraw-Hill Education, 2014.

IBGE - Instituto Brasileiro de Geografia e Estatística. Censo Demográfico 2022. Panorama da População. Disponível em: <<https://censo2022.ibge.gov.br/panorama/>>. Acesso em 08. dez. 2023

MATPLOTLIB. Disponível em: <<https://matplotlib.org/>>. Acesso em: 08 dez. 2023.

MINISTÉRIO DO TURISMO. Mais de 82,2 milhões de passageiros voaram pelo país em 2022, segundo ANAC. Disponível em: <<https://www.gov.br/turismo/pt-br/assuntos/noticias/mais-de-82-2-milhoes-de-passageiros-voaram-pelo-pais-em-2022-segundo-anac#:~:text=Mais%20de%2082%2C2%20milh%C3%B5es,segundo%20ANAC%20%E2%80%94%20Minist%C3%A9rio%20do%20Turismo/>>. Acesso em: 08 dez. 2023

NAGLE, T. T.; HOGAN, J. E.; ZALE, J. Z. The Strategy and Tactics of Pricing: A Guide to Growing More Profitably. Routledge, 2018.

NASA. Advanced Air Mobility Mission Overview. nasa.gov, 2021. Disponível em: <<https://www.nasa.gov/aam/overview/>>. Acesso em: 09 dez. 2023.

NETWORKX. NetworkX – Network Analysis in Python. [S.l.]. Disponível em: <<https://networkx.org/>>. Acesso em: 08 dez. 2023.

NUMPY. Disponível em: <<https://numpy.org/>>. Acesso em: 08 dez. 2023.

OSBORNE, Jason W.; OVERBAY, Amy. The power of outliers (and why researchers should always check for them). *Practical Assessment, Research, and Evaluation*, v. 9, n. 1, p. 6, 2004. Disponível em: <https://doi.org/10.7275/qq69-7k43>. Acesso em: 25 jun. 2024.

PEFFERS, Ken et al. A design science research methodology for information systems research. *Journal of Management Information Systems*, v. 24, p. 45-77, 2007. Disponível em: https://www.researchgate.net/publication/284503626_A_design_science_research_methodology_for_information_systems_research/. Acesso em: 14 dez. 2023.

PULP. Disponível em: <https://pypi.org/project/PuLP/>. Acesso em: 08 dez. 2023.

PURAO, S.; STOREY, V. C.; HAN, T. D. Improving Analysis Pattern Reuse in Conceptual Design: Augmenting Automated Processes with Supervised Learning. *Information Systems Research*, p. 269-290. Disponível em: <https://doi.org/10.1287/isre.14.3.269.16559/>. Acesso em: 14 dez. 2023.

ROLAND BERGER. *Urban Air Mobility - The rise of a new mode of transportation*. 2018. Disponível em: https://www.rolandberger.com/publications/publication_pdf/roland_berger_urban_air_mobility.pdf. Acesso em: 25 jun. 2024.

SEABORN. Disponível em: <https://seaborn.pydata.org/>. Acesso em: 08 dez. 2023.

SOUZA, Juliana. Uma proposta de estimativa de CASK de aeronaves do tipo eVTOL baseada em modelo híbrido de desempenho de aeronaves de asa rotativa e asa fixa. Trabalho de Conclusão de Curso (TCC). 11 dezembro 2023.

TAHA, H. A. *Pesquisa Operacional*. Pearson Brasil, 2017.

TEXTRON AVIATION. Cessna Caravan. Disponível em: <https://cessna.txtav.com/en/turboprop/caravan>. Acesso em: 13 jun. 2024.

THIPPHAVONG, David P.; KOPARDIKAR, Parimal V.; NGO, Thomas. *Urban Air Mobility Airspace Integration Concepts and Considerations*. NASA Technical Reports, 2018. Disponível em: <https://ntrs.nasa.gov/api/citations/20180005218/downloads/20180005218.pdf>. Acesso em: 25 jun. 2024.

VOEGOL. Significado de CASK. Disponível em: <https://ri.voegol.com.br/informacoes-aos-investidores/glossario/#:~:text=CUSTO%20OPERACIONAL%20POR%20ASSENTO%20DISPON%C3%8DVEL,total%20de%20assentos%20quil%C3%B4metro%20oferecidos>. Acesso em: 13 jun. 2024.

WINTER, Robert; BROCKE, Jan vom. Teaching Design Science Research. 2021. Disponível em:

https://www.researchgate.net/publication/355826960_Teaching_Design_Science_Research/.

Acesso em: 16 dez. 2023.

APÊNDICE

ANAC - Estatísticas - Organização e Tratamento dos Dados

Versão 01/07/2024 →

Introdução

Este notebook consolida o trabalho de organização e tratamento dos dados coletados nas bases da ANAC, no intervalo entre os anos de 2019 e 2022. Inicialmente, foi necessário estudar os bancos de dados para extrair os dados desejados para a nossa pesquisa e fazer as devidas correlações entre eles. Além disso, alguns dados serviram como insumo inicial para o cálculo de outras variáveis, como será apresentado a seguir no *notebook*.

Os bancos de dados que originalmente serão organizados e tratados neste notebook são:

- Estatísticas Básicas e Combinadas (2019 - 2022)

- Fonte: AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL - ANAC. Microdados. Disponível em:
<https://www.gov.br/anac/pt-br/assuntos/regulados/empresas-aereas/Instrucoes-para-a-elaboracao-e-apresentacao-das-demonstracoes-contabeis/envio-de-informacoes/microdados>. Acesso em: 25 nov. 2023.
 - Lista de Aeródromos Públicos V2
 - Fonte: AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL - ANAC. Metadados do conjunto de dados: Lista de aeródromos públicos V2. Disponível em:
<https://www.anac.gov.br/acesso-a-informacao/dados-abertos/areas-de-atuacao/aerodromos/lista-de-aerodromos-publicos-v2/70-lista-de-aerodromos-publicos-v2>. Acesso em: 25 nov. 2023
-

Objetivo Geral da Pesquisa

O objetivo deste trabalho é compreender e comparar os custos operacionais diretos dos serviços de transporte por veículos elétricos de decolagem e pouso vertical (eVTOL) nas principais malhas aéreas do Brasil garantindo a viabilidade em relação aos voos tradicionais. Essa análise tem como finalidade contribuir para a implementação e operacionalização eficiente desse modelo de transporte no país.

Inicialização - Instalação de Pacotes Python

```
python
#Importando todas as bibliotecas necessárias
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import scipy
from scipy import stats
from datetime import datetime
from scipy.optimize import curve_fit
```

```
python
from google.colab import drive
drive.mount('/gdrive')
```

```
python
%cd /gdrive/MyDrive/Colab Notebooks/ANAC
```

Passo 1 - Inicialização - Aeródromos e Distâncias entre Aeroportos

Descrição: leitura, organização e transformação dos arquivos de aeródromos da ANAC e criação do arquivo de distâncias, originado do cálculo das distâncias entre as combinações dos aeroportos de origem e destino.

Etapa I. Acessando 'Lista de Aeródromos Públicos V2

Construção inicial do arquivo salvo em ./Dados/Aerodromos/aeroportos2.csv

Descrição: Nesta etapa foi realizada a leitura do arquivo da ANAC, organização e transformação do arquivo, excluindo as variáveis que não são de interesse para a pesquisa.

```
python
# Leitura do arquivo:
aeroportos = pd.read_csv('./Dados/Aerodromos/AerodromosPublicos.csv',
                        encoding='cp1252', delimiter = ';', skiprows=1)
```

```
python
# Tamanho do arquivo:
aeroportos.shape
```

```
python
# Visualização do arquivo:
aeroportos.head()
```

```
python
# Variáveis do arquivo:
aeroportos.columns
```

```
python
# Exclusão das variáveis que não são de interesse para a pesquisa:
aer_coord = aeroportos.drop(columns=['CIAD', 'Nome', 'Município', 'UF',
    'Latitude', 'Longitude', 'Altitude', 'Operação Diurna',
    'Operação Noturna', 'Designação 1', 'Comprimento 1', 'Largura 1',
    'Resistência 1', 'Superfície 1', 'Designação 2', 'Comprimento 2',
    'Largura 2', 'Resistência 2', 'Superfície 2', 'Situação',
    'Validade do Registro', 'Portaria de Registro', 'Link Portaria'])
```

```
python
# Renomeando as variáveis do arquivo:
aer_coord.columns = ['OACI', 'Município', 'UF', 'Lat', 'Long']
```

```
python
# Eliminando os dados vazios (NA):
aer_coord = aer_coord.dropna()
```

```
python
# Salvando o arquivo "aerportos2.csv":
aer_coord.to_csv('./Dados/Aerodromos/aerportos2.csv', index=False)
```

Etapa II. Definindo a Função para o Cálculo da Distância entre Aeródromos

Descrição: Esta etapa foi realizada com base no projeto 'airportsdata20231017', que faz o levantamento da latitude e longitude para um conjunto de aeroportos. Através da latitude e longitude de um aeroporto de origem para um aeroporto de destino, foi possível calcular a distância entre eles.

Fonte: BORSETTI, M. Airportsdata 20231017. 2022. Disponível em: <https://pypi.org/project/airportsdata/>. Acesso em: 26 nov. 2023.

```
python
# Instalando Airportsdata:
%pip install -U airportsdata
```

```
python
import airportsdata
```

```
python
# Buscando por código ICAO:
airports = airportsdata.load() # key is ICAO code (default)
print(airports['SBBV'])
```

```
python
#Buscando por código IATA
airports = airportsdata.load('IATA') # key is IATA code, not the default
print(airports['BSB'])
```

```
python
# Função para calcular distância de grande círculo entre aeroportos:
# Observação: Distância de grande círculo é a menor distância entre dois pontos em uma
esfera, medida ao longo da superfície da esfera.

import math

def distance_on_unit_sphere(lat1, long1, lat2, long2):
```

```

# Converta a latitude e a longitude em coordenadas esféricas em radianos:
degrees_to_radians = math.pi/180.0

# phi = 90 - latitude:
phi1 = (90.0 - lat1)*degrees_to_radians
phi2 = (90.0 - lat2)*degrees_to_radians

# theta = longitude:
theta1 = long1*degrees_to_radians
theta2 = long2*degrees_to_radians

# Calcular a distância esférica a partir de coordenadas esféricas

# Para dois locais em coordenadas esféricas
# (1, theta, phi) and (1, theta, phi)
# cosine( arc length ) =
#   sin phi sin phi' cos(theta-theta') + cos phi cos phi'
# distance = rho * arc length

cos = (math.sin(phi1)*math.sin(phi2)*math.cos(theta1 - theta2) +
       math.cos(phi1)*math.cos(phi2))
arc = math.acos( cos )

# Multiplique o arco pelo raio da Terra em seu conjunto de unidades desejado (nesse
caso, KM) para obter o comprimento.
arc = arc * 6378.14 #raio da Terra em Quilômetros

return arc

```

Etapa III. Criando arquivo com distância entre aeroportos

Construção inicial do arquivo salvo em `./Dados/Tarifas/distancias.csv`

Descrição: Esta etapa foi realizada para criar o arquivo de distâncias entre os aeroportos de origem e destino. Para isso, foram feitas todas as combinações possíveis entre aeroportos de origem e destino e, com base na latitude e longitude de cada um, foi aplicada a função `distance_on_unit_sphere` para calcular a distância entre eles (KM).

```

python
# Leitura do arquivo:
aerodromo = pd.read_csv('./Dados/Aerodromos/aeroportos2.csv')

```

```

python
# Visualização do arquivo:
aerodromo.head()

```

```

python
# Criando uma lista com os dados da variável OACI (códigos dos aeroportos):
aerod = aerodromo.OACI.tolist()

```

```
python
```

```
# 1. Criando as combinações possíveis de aeroportos (OACIs) de origem e destino e aplicando
a função 'distance_on_unit_sphere' para as suas respectivas latitudes e longitudes de origem
e destino.
# 2. Append de 'i' na lista orig_aer, de 'j' na lista dest_aer e de 'distancia' na lista
dist_aer:
orig_aer = []
dest_aer = []
dist_aer = []
for i in aerod:
    for j in aerod:
        if i != j:
            o = aerodromo[aerodromo.OACI == i]
            d = aerodromo[aerodromo.OACI == j]
            lat1 = o['Lat'].values[0]
            lon1 = o['Long'].values[0]
            lat2 = d['Lat'].values[0]
            lon2 = d['Long'].values[0]
            distancia = distance_on_unit_sphere(lat1, lon1, lat2, lon2)
        else:
            distancia = 0.0
        orig_aer.append(i)
        dest_aer.append(j)
        dist_aer.append(distancia)
```

```
python
```

```
# Criando o dataframe distancias:
orig_series = pd.Series(orig_aer)
dest_series = pd.Series(dest_aer)
dist_series = pd.Series(dist_aer)
frame2 = {'orig': orig_series,
          'dest': dest_series,
          'dist': dist_series}
distancias = pd.DataFrame(frame2)
```

```
python
```

```
# Visualizando o dataframe distancias:
distancias.head()
```

```
python
```

```
# Salvando o arquivo "distancias.csv":
distancias.to_csv('./Dados/Tarifas/distancias.csv', index=False)
```

Etapa IV. Leitura do arquivo "distancias.csv" e transformação inicial para float

Salvo em ./Dados/Tarifas/distancias2.csv

Descrição: Esta etapa foi realizada para transformar o tipo do dado da variável distância de *string* para *float*.

```
python
# Leitura do arquivo:
dados_distancias = pd.read_csv('./Dados/Tarifas/distancias.csv')
```

```
python
# Visualização do arquivo:
dados_distancias.head()
```

```
python
# Substituindo ',' por '.':
dados_distancias.dist = dados_distancias.dist.str.replace(',', '.')
```

```
python
# Transformando tipo de dado para 'float' (número real):
dados_distancias.dist = dados_distancias.dist.astype(float)
```

```
python
# Salvando o arquivo "distancias2.csv":
dados_distancias.to_csv('./Dados/Aerodromos/distancias2.csv', index=False)
```

Passo 2 - Tarifas - Transformação de Dados

Descrição: leitura, organização e transformação dos arquivos de tarifas disponibilizados pela ANAC, entre 2019 e 2022.

Descrição das Variáveis: ANAC. Microdados de tarifas aéreas comercializadas. Disponível em:

<https://www.gov.br/anac/pt-br/assuntos/dados-e-estatisticas/microdados-de-tarifas-aereas-comercializadas>. Acesso em: 26 nov. 2023.

Atenção: os arquivos baixados são .CSV, mas no formato Microsoft Excel, *comma separated values*, portanto, não tem o encoding utf-8. Além disso, o Excel em português salva o CSV separado com ponto e vírgula (;').

Etapa I. Tarifas - Transformação de Dados 2019

Etapa I.I. Leitura e agrupamento de arquivos de Tarifas 2019

Salvo em ./Dados/Tarifas/tarifas_2019.csv

Descrição: Nesta etapa foi realizada a leitura e agrupamento dos arquivos de tarifas disponibilizados pela ANAC para 2019.

```
python
# Leitura do arquivo:
files = ['201901.CSV', '201902.CSV', '201903.CSV', '201904.CSV',
        '201905.CSV', '201906.CSV', '201907.CSV', '201908.CSV',
        '201909.CSV', '201910.CSV', '201911.CSV', '201912.CSV']
```

```
python
# Estruturando o dataframe:
df = pd.DataFrame(columns=['ANO', 'MES', 'EMPRESA', 'ORIGEM',
                          'DESTINO', 'TARIFA', 'ASSENTOS'])
```

```
python
# Agrupando os arquivos no dataframe estruturado:
for file in files:
    data = pd.read_csv(f'./Dados/Tarifas/Tarifas_2019/{file}',
                      encoding='cp1252', delimiter = ';')
    data.columns = ['ANO', 'MES', 'EMPRESA', 'ORIGEM',
                   'DESTINO', 'TARIFA', 'ASSENTOS']
    df = pd.concat([df, data], axis=0, ignore_index=True)
```

```
python
# Visualização do dataframe:
df.head()
```

```
python
# Salvando o arquivo "tarifas_2019.csv":
df.to_csv('./Dados/Tarifas/tarifas_2019.csv', index=False)
```

Etapa I.II. Obtenção das Tarifas em 2019 e transformação inicial

Salvos em ./Dados/Tarifas/tarifas9.csv

Descrição: Nesta etapa foi realizada a transformação dos dados do arquivo tarifas_2019, o adequando para - posteriormente - ser analisado.

```
python
# Leitura do arquivo:
dados_tarifas = pd.read_csv('./Dados/Tarifas/tarifas_2019.csv')
```

```
python
# Tamanho do arquivo:
dados_tarifas.shape
```



```
data.columns = ['ANO', 'MES', 'EMPRESA', 'ORIGEM',
                'DESTINO', 'TARIFA', 'ASSENTOS']
df = pd.concat([df, data], axis=0, ignore_index=True)
```

```
python
```

```
# Visualização do dataframe:
df.head()
```

```
python
```

```
# Salvando o arquivo "tarifas_2020.csv":
df.to_csv('./Dados/Tarifas/tarifas_2020.csv', index=False)
```

Etapa II.II. Obtenção das Tarifas em 2020 e transformação inicial

Salvos em ./Dados/Tarifas/tarifas0.csv

Descrição: Nesta etapa foi realizada a transformação dos dados do arquivo tarifas_2020, o adequando para - posteriormente - ser analisado.

```
python
```

```
# Leitura do arquivo:
dados_tarifas = pd.read_csv('./Dados/Tarifas/tarifas_2020.csv')
```

```
python
```

```
# Tamanho do arquivo:
dados_tarifas.shape
```

```
python
```

```
# Visualização do arquivo:
dados_tarifas.head()
```

```
python
```

```
# Padronização dos dados respecivos à variável MES:
dados_tarifas.MES.unique()
```

```
python
```

```
# Substituindo ',' por '.' na variável TARIFA:
dados_tarifas.TARIFA = dados_tarifas.TARIFA.str.replace(',', '.')
```

```
python
```

```
# Alterando o tipo de dado da variável TARIFA para float:
dados_tarifas.TARIFA = dados_tarifas.TARIFA.astype(float)
```

```
python
```

```
# Salvando o arquivo "tarifas0.csv":
dados_tarifas.to_csv('./Dados/Tarifas/tarifas0.csv', index=False)
```

Etapa III. Tarifas - Transformação de Dados 2021

Etapa III.I. Leitura e agrupamento de arquivos de Tarifas 2021

Salvo em ./Dados/Tarifas/tarifas_2021.csv

Descrição: Nesta etapa foi realizada a leitura e agrupamento dos arquivos de tarifas disponibilizados pela ANAC para 2021.

```
python
# Leitura do arquivo:
files = ['202101.CSV', '202102.CSV', '202103.CSV', '202104.CSV',
        '202105.CSV', '202106.CSV', '202107.CSV', '202108.CSV',
        '202109.CSV', '202110.CSV', '202111.CSV', '202112.CSV']
```

```
python
# Estruturando o dataframe:
df = pd.DataFrame(columns=['ANO', 'MES', 'EMPRESA', 'ORIGEM',
                          'DESTINO', 'TARIFA', 'ASSENTOS'])
```

```
python
# Agrupando os arquivos no dataframe estruturado:
for file in files:
    data = pd.read_csv(f'./Dados/Tarifas/Tarifas_2021/{file}',
                      encoding='cp1252', delimiter = ';')
    if file == '202112.CSV':
        data = data.iloc[:,1:]
    data.columns = ['ANO', 'MES', 'EMPRESA', 'ORIGEM',
                   'DESTINO', 'TARIFA', 'ASSENTOS']
    df = pd.concat([df, data], axis=0, ignore_index=True)
```

```
python
# Visualização do dataframe:
df.head()
```

```
python
# Salvando o arquivo "tarifas_2021.csv":
df.to_csv('./Dados/Tarifas/tarifas_2021.csv', index=False)
```

Etapa III.II. Obtenção das Tarifas em 2021 e transformação inicial

Salvos em ./Dados/Tarifas/tarifas1.csv

Descrição: Nesta etapa foi realizada a transformação dos dados do arquivo `tarifas_2021`, o adequando para - posteriormente - ser analisado.

```
python
# Leitura do arquivo:
dados_tarifas = pd.read_csv('./Dados/Tarifas/tarifas_2021.csv')
```

```
python
# Tamanho do arquivo:
dados_tarifas.shape
```

```
python
# Visualização do arquivo:
dados_tarifas.head()
```

```
python
# Padronização dos dados respectivos à variável MES:
dados_tarifas.MES.unique()
```

```
python
# Substituindo ',' por '.' na variável TARIFA:
dados_tarifas.TARIFA = dados_tarifas.TARIFA.str.replace(',', '.')
```

```
python
# Alterando o tipo de dado da variável TARIFA para float:
dados_tarifas.TARIFA = dados_tarifas.TARIFA.astype(float)
```

```
python
# Salvando o arquivo "tarifas1.csv":
dados_tarifas.to_csv('./Dados/Tarifas/tarifas1.csv', index=False)
```

Etapa IV. Tarifas - Transformação de Dados 2022

Etapa IV.I. Leitura e agrupamento de arquivos de Tarifas 2022

Salvo em `./Dados/Tarifas/tarifas_2022.csv`

Descrição: Nesta etapa foi realizada a leitura e agrupamento dos arquivos de tarifas disponibilizados pela ANAC para 2022.

```
python
# Leitura do arquivo:
files = ['202201.CSV', '202202.CSV', '202203.CSV', '202204.CSV',
         '202205.CSV', '202206.CSV', '202207.CSV', '202208.CSV',
```

```
'202209.CSV', '202210.CSV', '202211.CSV', '202212.CSV']
```

```
python
```

```
# Estruturando o dataframe:
df = pd.DataFrame(columns=['ANO', 'MES', 'EMPRESA', 'ORIGEM',
                          'DESTINO', 'TARIFA', 'ASSENTOS'])
```

```
python
```

```
# Agrupando os arquivos no dataframe estruturado:
for file in files:
    data = pd.read_csv(f'./Dados/Tarifas/Tarifas_2022/{file}',
                      encoding='cp1252', delimiter = ';')
    if file == '202204.CSV':
        data = data.iloc[:,1:]
    data.columns = ['ANO', 'MES', 'EMPRESA', 'ORIGEM',
                   'DESTINO', 'TARIFA', 'ASSENTOS']
    df = pd.concat([df, data], axis=0, ignore_index=True)
```

```
python
```

```
# Visualização do dataframe:
df.head()
```

```
python
```

```
# Salvando o arquivo "tarifas_2022.csv":
df.to_csv('./Dados/Tarifas/tarifas_2022.csv', index=False)
```

Etapa IV.II. Obtenção das Tarifas em 2022 e transformação inicial

Salvos em ./Dados/Tarifas/tarifas2.csv

Descrição: Nesta etapa foi realizada a transformação dos dados do arquivo tarifas_2022, o adequando para - posteriormente - ser analisado.

```
python
```

```
# Leitura do arquivo:
dados_tarifas = pd.read_csv('./Dados/Tarifas/tarifas_2022.csv')
```

```
python
```

```
# Tamanho do arquivo:
dados_tarifas.shape
```

```
python
```

```
# Visualização do arquivo:
dados_tarifas.head()
```

```
python
```

```
# Padronização dos dados respectivos à variável MES:
dados_tarifas.MES.unique()
```

```
python
```

```
# Substituindo ',' por '.' na variável TARIFA:
dados_tarifas.TARIFA = dados_tarifas.TARIFA.str.replace(',','.')
```

```
python
```

```
# Alterando o tipo de dado da variável TARIFA para float:
dados_tarifas.TARIFA = dados_tarifas.TARIFA.astype(float)
```

```
python
```

```
# Salvando o arquivo "tarifas2.csv":
dados_tarifas.to_csv('./Dados/Tarifas/tarifas2.csv',index=False)
```

Passo 3 - Demanda e Oferta - Transformação de Dados

Descrição: leitura, organização e transformação dos arquivos de estatísticas disponibilizados pela ANAC, respectivos aos voos realizados entre 2019 e 2022.

Descrição das Variáveis: ANAC. Descrição das variáveis (microdados). Disponível em:

<https://www.gov.br/anac/pt-br/assuntos/dados-e-estatisticas/mercado-do-transporte-aereo/demanda-e-oferta/descricao-das-variaveis-microdados>. Acesso em: 26 nov. 2023.

Atenção: os arquivos baixados são .CSV, mas no formato Microsoft Excel, *comma separated values*, portanto, não tem o encoding utf-8. Além disso, o Excel em português salva o CSV separado com ponto e vírgula (;').

Etapa I. Demanda e Oferta - Básica

Etapa Básica (flight stage): As etapas básicas são aquelas realizadas pela aeronave desde a sua decolagem até o próximo pouso, independentemente de onde tenha ocorrido o embarque ou o desembarque do objeto de transporte. Os dados estatísticos das etapas básicas representam o status da aeronave em cada etapa do voo, apresentando a movimentação de cargas e passageiros entre os aeródromos de origem e destino da aeronave. É a operação de uma aeronave entre uma decolagem e o próximo pouso, a ligação entre dois aeródromos.

Fonte: ANAC. Descrição das variáveis (microdados). Disponível em:

<https://www.gov.br/anac/pt-br/assuntos/dados-e-estatisticas/mercado-do-transporte-aereo/demanda-e-oferta/descricao-das-variaveis-microdados>

[ere/demanda-e-oferta/descricao-das-variaveis-microdados](#). Acesso em: 26 nov. 2023

Etapa I.I. Leitura e agrupamento de arquivos de estatísticas básicas 2019

Salvo em `./Dados/DemandaeOferta/estatisticas_2019.csv`

Descrição: Nesta etapa foi realizada a leitura e agrupamento dos arquivos de estatísticas disponibilizados pela ANAC para 2019.

```
python
# Leitura do arquivo:
files = ['basica2019-01.zip', 'basica2019-02.zip', 'basica2019-03.zip',
         'basica2019-04.zip', 'basica2019-05.zip', 'basica2019-06.zip',
         'basica2019-07.zip', 'basica2019-08.zip', 'basica2019-09.zip',
         'basica2019-10.zip', 'basica2019-11.zip', 'basica2019-12.zip']
```

```
python
# Criação do dataframe:
df = pd.DataFrame()
```

```
python
# Concatenação dos arquivos no dataframe:
for file in files:
    print(file)
    data = pd.read_csv(f'./Dados/DemandaeOferta/2019_basica/{file}',
                      encoding='cp1252', delimiter = ';', low_memory=False)
    df = pd.concat([df, data], axis=0, ignore_index=True)
```

```
python
# Visualização do dataframe:
df.head()
```

```
python
# Salvando o arquivo "estatisticas_2019.csv":
df.to_csv('./Dados/DemandaeOferta/estatisticas_2019.csv', index=False)
```

Etapa I.II. Obtenção das estatísticas básicas de Demanda e Oferta em 2019

Salvo em `/Dados/Estatisticas/estatisticas9.csv`

Descrição: Nesta etapa foi realizada a transformação do arquivo `estatisticas_2019`.

```
python
# Leitura do arquivo:
dados_estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas_2019.csv',
                                  low_memory=False)
```

```
python
```

```
# Dimensões do arquivo:
dados_estatisticas.shape
```

```
python
```

```
# Padronização da variável mês:
dados_estatisticas['nr_mes_referencia'].unique()
```

```
python
```

```
dados_estatisticas.columns
```

```
python
```

```
# Filtrando o dataframe com a natureza da etapa de voo = doméstica, serviço referente ao
tipo de linha = passageiro e tipo de Voo = regular:
dados_estatisticas = dados_estatisticas[
    (dados_estatisticas['ds_natureza_etapa'] == 'DOMÉSTICA') & \
    (dados_estatisticas['ds_servico_tipo_linha'] == 'PASSAGEIRO') & \
    (dados_estatisticas['ds_grupo_di'] == 'REGULAR')]
# Com isso, excluem-se os dados que não são do interesse da pesquisa.
```

```
python
```

```
dados_estatisticas.info()
```

```
python
```

```
dados_estatisticas.isna().sum()
```

```
python
```

```
# Exclusão das variáveis que não são do interesse para a pesquisa:
dados_estatisticas = dados_estatisticas.drop(columns=
    ['id_basica', 'id_empresa', 'sg_empresa_iata', 'nm_pais', 'ds_tipo_empresa',
    'nr_singular', 'id_di', 'cd_di', 'ds_di', 'ds_grupo_di', 'dt_referencia',
    'nr_ano_referencia', 'nr_semestre_referencia', 'nm_semestre_referencia',
    'nr_trimestre_referencia', 'nm_trimestre_referencia', 'nr_mes_referencia',
    'nm_mes_referencia', 'nr_semana_referencia', 'nm_dia_semana_referencia',
    'nr_dia_referencia', 'nr_ano_mes_referencia', 'id_tipo_linha',
    'cd_tipo_linha', 'ds_tipo_linha', 'ds_natureza_tipo_linha', 'ds_servico_tipo_linha',
    'ds_natureza_etapa', 'nr_ano_partida_real', 'nr_semestre_partida_real',
    'nm_semestre_partida_real', 'nr_trimestre_partida_real', 'nm_trimestre_partida_real',
    'nm_mes_partida_real', 'nr_ano_mes_partida_real', 'id_aerodromo_origem',
    'nm_aerodromo_origem', 'nm_municipio_origem', 'nm_pais_origem', 'nm_continente_origem',
    'nr_ano_chegada_real', 'nr_semestre_chegada_real', 'nm_semestre_chegada_real',
    'nr_trimestre_chegada_real', 'nm_trimestre_chegada_real', 'nm_mes_chegada_real',
    'nr_ano_mes_chegada_real', 'id_aerodromo_destino', 'nm_aerodromo_destino',
    'nm_municipio_destino', 'nm_regiao_destino', 'nm_pais_destino', 'nm_continente_destino',
    'nr_escala_destino', 'id_arquivo', 'nm_arquivo', 'nr_linha', 'dt_sistema']
)
```

```
python
```

```
# Dimensão do dataframe:
dados_estatisticas.shape
```

```
python
# Salvando o arquivo "estatisticas9.csv":
dados_estatisticas.to_csv('./Dados/DemandaeOferta/estatisticas9.csv',
                          index=False)
```

Etapa I.III. Leitura e agrupamento de arquivos de estatísticas básicas 2020

Salvo em ./Dados/DemandaeOferta/estatisticas_2020.csv

Descrição: Nesta etapa foi realizada a leitura e agrupamento dos arquivos de estatísticas disponibilizados pela ANAC para 2020.

```
python
# Leitura do arquivo:
files = ['basica2020-01.zip', 'basica2020-02.zip', 'basica2020-03.zip',
         'basica2020-04.zip', 'basica2020-05.zip', 'basica2020-06.zip',
         'basica2020-07.zip', 'basica2020-08.zip', 'basica2020-09.zip',
         'basica2020-10.zip', 'basica2020-11.zip', 'basica2020-12.zip']
```

```
python
# Criação do dataframe:
df = pd.DataFrame()
```

```
python
# Concatenação dos arquivos no dataframe:
for file in files:
    print(file)
    data = pd.read_csv(f'./Dados/DemandaeOferta/2020_basica/{file}',
                      encoding='cp1252', delimiter = ';', low_memory=False)
    df = pd.concat([df, data], axis=0, ignore_index=True)
```

```
python
# Visualização do dataframe:
df.head()
```

```
python
# Salvando o arquivo "estatisticas_2020.csv":
df.to_csv('./Dados/DemandaeOferta/estatisticas_2020.csv', index=False)
```

Etapa I.IV. Obtenção das estatísticas básicas de Demanda e Oferta em 2020

Salvo em /Dados/Estatisticas/estatisticas0.csv

Descrição: Nesta etapa foi realizada a transformação do arquivo estatisticas_2020.

```
python
```

```
# Leitura do arquivo:
dados_estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas_2020.csv',
                                  low_memory=False)
```

```
python
```

```
# Dimensões do arquivo:
dados_estatisticas.shape
```

```
python
```

```
# Padronização da variável mês:
dados_estatisticas['nr_mes_referencia'].unique()
```

```
python
```

```
dados_estatisticas.columns
```

```
python
```

```
# Filtrando o dataframe com a natureza da etapa de voo = doméstica, serviço referente ao
tipo de linha = passageiro e tipo de Voo = regular:
dados_estatisticas = dados_estatisticas[
    (dados_estatisticas['ds_natureza_etapa'] == 'DOMÉSTICA') & \
    (dados_estatisticas['ds_servico_tipo_linha'] == 'PASSAGEIRO') & \
    (dados_estatisticas['ds_grupo_di'] == 'REGULAR')]
```

```
python
```

```
dados_estatisticas.info()
```

```
python
```

```
dados_estatisticas.isna().sum()
```

```
python
```

```
# Exclusão das variáveis que não são do interesse para a pesquisa:
dados_estatisticas = dados_estatisticas.drop(columns=
    ['id_basica', 'id_empresa', 'sg_empresa_iata', 'nm_pais', 'ds_tipo_empresa',
     'nr_singular', 'id_di', 'cd_di', 'ds_di', 'ds_grupo_di', 'dt_referencia',
     'nr_ano_referencia', 'nr_semestre_referencia', 'nm_semestre_referencia',
     'nr_trimestre_referencia', 'nm_trimestre_referencia', 'nr_mes_referencia',
     'nm_mes_referencia', 'nr_semana_referencia', 'nm_dia_semana_referencia',
     'nr_dia_referencia', 'nr_ano_mes_referencia', 'id_tipo_linha',
     'cd_tipo_linha', 'ds_tipo_linha', 'ds_natureza_tipo_linha', 'ds_servico_tipo_linha',
     'ds_natureza_etapa', 'nr_ano_partida_real', 'nr_semestre_partida_real',
     'nm_semestre_partida_real', 'nr_trimestre_partida_real', 'nm_trimestre_partida_real',
     'nm_mes_partida_real', 'nr_ano_mes_partida_real', 'id_aerodromo_origem',
     'nm_aerodromo_origem', 'nm_municipio_origem', 'nm_pais_origem', 'nm_continente_origem',
     'nr_ano_chegada_real', 'nr_semestre_chegada_real', 'nm_semestre_chegada_real',
     'nr_trimestre_chegada_real', 'nm_trimestre_chegada_real', 'nm_mes_chegada_real',
```

```
'nr_ano_mes_chegada_real', 'id_aerodromo_destino', 'nm_aerodromo_destino',
'nm_municipio_destino', 'nm_regiao_destino', 'nm_pais_destino', 'nm_continente_destino',
'nr_escala_destino', 'id_arquivo', 'nm_arquivo', 'nr_linha', 'dt_sistema']
)
```

```
python
```

```
# Dimensão do dataframe:
dados_estatisticas.shape
```

```
python
```

```
# Salvando o arquivo "estatisticas0.csv":
dados_estatisticas.to_csv('./Dados/DemandaeOferta/estatisticas0.csv',
                           index=False)
```

Etapa I.V. Leitura e agrupamento de arquivos de estatísticas básicas 2021

Salvo em ./Dados/DemandaeOferta/estatisticas_2021.csv

Descrição: Nesta etapa foi realizada a leitura e agrupamento dos arquivos de estatísticas disponibilizados pela ANAC para 2021.

```
python
```

```
# Leitura do arquivo:
files = ['basica2021-01.zip', 'basica2021-02.zip', 'basica2021-03.zip',
         'basica2021-04.zip', 'basica2021-05.zip', 'basica2021-06.zip',
         'basica2021-07.zip', 'basica2021-08.zip', 'basica2021-09.zip',
         'basica2021-10.zip', 'basica2021-11.zip', 'basica2021-12.zip']
```

```
python
```

```
# Criação do dataframe:
df = pd.DataFrame()
```

```
python
```

```
# Concatenação dos arquivos no dataframe:
for file in files:
    print(file)
    data = pd.read_csv(f'./Dados/DemandaeOferta/2021_basica/{file}',
                       encoding='cp1252', delimiter = ';', low_memory=False)
    df = pd.concat([df, data], axis=0, ignore_index=True)
```

```
python
```

```
# Visualização do dataframe:
df.head()
```

```
python
```

```
# Salvando o arquivo "estatisticas_2021.csv"
df.to_csv('./Dados/DemandaeOferta/estatisticas_2021.csv', index=False)
```

Etapa I.VI. Obtenção das estatísticas básicas de Demanda e Oferta em 2021

Salvo em /Dados/Estatisticas/estatisticas1.csv

Descrição: Nesta etapa foi realizada a transformação do arquivo estatisticas_2021.

```
python
# Leitura do arquivo:
dados_estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas_2021.csv',
                                low_memory=False)
```

```
python
# Dimensões do arquivo:
dados_estatisticas.shape
```

```
python
# Padronização da variável mês:
dados_estatisticas['nr_mes_referencia'].unique()
```

```
python
dados_estatisticas.columns
```

```
python
# Filtrando o dataframe com a natureza da etapa de voo = doméstica, serviço referente ao
tipo de linha = passageiro e tipo de Voo = regular:
dados_estatisticas = dados_estatisticas[
    (dados_estatisticas['ds_natureza_etapa'] == 'DOMÉSTICA') & \
    (dados_estatisticas['ds_servico_tipo_linha'] == 'PASSAGEIRO') & \
    (dados_estatisticas['ds_grupo_di'] == 'REGULAR')]
```

```
python
dados_estatisticas.info()
```

```
python
dados_estatisticas.isna().sum()
```

```
python
# Exclusão das variáveis que não são do interesse para a pesquisa:
dados_estatisticas = dados_estatisticas.drop(columns=
    ['id_basica', 'id_empresa', 'sg_empresa_iata', 'nm_pais', 'ds_tipo_empresa',
     'nr_singular', 'id_di', 'cd_di', 'ds_di', 'ds_grupo_di', 'dt_referencia',
     'nr_ano_referencia', 'nr_semestre_referencia', 'nm_semestre_referencia',
     'nr_trimestre_referencia', 'nm_trimestre_referencia', 'nr_mes_referencia',
     'nm_mes_referencia', 'nr_semana_referencia', 'nm_dia_semana_referencia',
     'nr_dia_referencia', 'nr_ano_mes_referencia', 'id_tipo_linha',
```

```
'cd_tipo_linha','ds_tipo_linha','ds_natureza_tipo_linha','ds_servico_tipo_linha',
'ds_natureza_etapa','nr_ano_partida_real','nr_semestre_partida_real',
'nm_semestre_partida_real','nr_trimestre_partida_real','nm_trimestre_partida_real',
'nm_mes_partida_real','nr_ano_mes_partida_real','id_aerodromo_origem',
'nm_aerodromo_origem','nm_municipio_origem','nm_pais_origem','nm_continente_origem',
'nr_ano_chegada_real','nr_semestre_chegada_real','nm_semestre_chegada_real',
'nr_trimestre_chegada_real','nm_trimestre_chegada_real','nm_mes_chegada_real',
'nr_ano_mes_chegada_real','id_aerodromo_destino','nm_aerodromo_destino',
'nm_municipio_destino','nm_regiao_destino','nm_pais_destino','nm_continente_destino',
'nr_escala_destino','id_arquivo','nm_arquivo','nr_linha','dt_sistema']
)
```

```
python
```

```
# Dimensão do dataframe:
dados_estatisticas.shape
```

```
python
```

```
# Salvando o arquivo "estatisticas1.csv":
dados_estatisticas.to_csv('./Dados/DemandaeOferta/estatisticas1.csv',
                          index=False)
```

Etapa I.VII. Leitura e agrupamento de arquivos de estatísticas básicas 2022

Salvo em `./Dados/DemandaeOferta/estatisticas_2022.csv`

Descrição: Nesta etapa foi realizada a leitura e agrupamento dos arquivos de estatísticas disponibilizados pela ANAC para 2021.

```
python
```

```
# Leitura do arquivo:
files = ['basica2022-01.zip', 'basica2022-02.zip', 'basica2022-03.zip',
        'basica2022-04.zip', 'basica2022-05.zip', 'basica2022-06.zip',
        'basica2022-07.zip', 'basica2022-08.zip', 'basica2022-09.zip',
        'basica2022-10.zip', 'basica2022-11.zip', 'basica2022-12.zip']
```

```
python
```

```
# Criação do dataframe:
df = pd.DataFrame()
```

```
python
```

```
# Concatenação dos arquivos no dataframe:
for file in files:
    print(file)
    data = pd.read_csv(f'./Dados/DemandaeOferta/2022_basica/{file}',
                      encoding='cp1252', delimiter = ';', low_memory=False)
    df = pd.concat([df, data], axis=0, ignore_index=True)
```

```
python
```

```
# Visualização do dataframe:
```

```
df.head()
```

```
python
# Salvando o arquivo "estatisticas_2022.csv":
df.to_csv('./Dados/Estatisticas/estatisticas_2022.csv', index=False)
```

Etapa I.VIII. Obtenção das estatísticas básicas de Demanda e Oferta em 2022

Salvo em /Dados/Estatisticas/estatisticas2.csv

Descrição: Nesta etapa foi realizada a transformação do arquivo estatisticas_2022.

```
python
# Leitura do arquivo:
dados_estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas_2022.csv',
                                low_memory=False)
```

```
python
# Dimensões do arquivo:
dados_estatisticas.shape
```

```
python
# Padronização da variável mês:
dados_estatisticas['nr_mes_referencia'].unique()
```

```
python
dados_estatisticas.columns
```

```
python
# Filtrando o dataframe com a natureza da etapa de voo = doméstica, serviço referente ao
tipo de linha = passageiro e tipo de Voo = regular:
dados_estatisticas = dados_estatisticas[
    (dados_estatisticas['ds_natureza_etapa'] == 'DOMÉSTICA') & \
    (dados_estatisticas['ds_servico_tipo_linha'] == 'PASSAGEIRO') & \
    (dados_estatisticas['ds_grupo_di'] == 'REGULAR')]
```

```
python
dados_estatisticas.info()
```

```
python
dados_estatisticas.isna().sum()
```

```
python
# Exclusão das variáveis que não são do interesse para a pesquisa:
```



```

dados_estatisticas = dados_estatisticas.drop(columns=
    ['id_basica', 'id_empresa', 'sg_empresa_iata', 'nm_pais', 'ds_tipo_empresa',
     'nr_singular', 'id_di', 'cd_di', 'ds_di', 'ds_grupo_di', 'dt_referencia',
     'nr_ano_referencia', 'nr_semestre_referencia', 'nm_semestre_referencia',
     'nr_trimestre_referencia', 'nm_trimestre_referencia', 'nr_mes_referencia',
     'nm_mes_referencia', 'nr_semana_referencia', 'nm_dia_semana_referencia',
     'nr_dia_referencia', 'nr_ano_mes_referencia', 'id_tipo_linha',
     'cd_tipo_linha', 'ds_tipo_linha', 'ds_natureza_tipo_linha', 'ds_servico_tipo_linha',
     'ds_natureza_etapa', 'nr_ano_partida_real', 'nr_semestre_partida_real',
     'nm_semestre_partida_real', 'nr_trimestre_partida_real', 'nm_trimestre_partida_real',
     'nm_mes_partida_real', 'nr_ano_mes_partida_real', 'id_aerodromo_origem',
     'nm_aerodromo_origem', 'nm_municipio_origem', 'nm_pais_origem', 'nm_continente_origem',
     'nr_ano_chegada_real', 'nr_semestre_chegada_real', 'nm_semestre_chegada_real',
     'nr_trimestre_chegada_real', 'nm_trimestre_chegada_real', 'nm_mes_chegada_real',
     'nr_ano_mes_chegada_real', 'id_aerodromo_destino', 'nm_aerodromo_destino',
     'nm_municipio_destino', 'nm_regiao_destino', 'nm_pais_destino', 'nm_continente_destino',
     'nr_escala_destino', 'id_arquivo', 'nm_arquivo', 'nr_linha', 'dt_sistema']
)

```

```
python
```

```

# Dimensão do dataframe:
dados_estatisticas.shape

```

```
python
```

```

# Salvando o arquivo "estatisticas2.csv":
dados_estatisticas.to_csv('./Dados/Estatisticas/estatisticas2.csv',
                          index=False)

```

Etapa II. Demanda e Oferta - Combinada

Etapa Combinada (On flight origin and destination - OFOD): As etapas combinadas identificam os pares de origem, onde ocorreu o embarque do objeto de transporte, e destino, onde ocorreu o desembarque do objeto de transporte, independentemente da existência de aeródromos intermediários, atendidos por determinado voo. É a etapa de voo vista com foco no objeto de transporte (pessoas e/ou cargas), com base no embarque e no desembarque nos aeródromos relacionados. Os dados estatísticos da etapa combinada informam a origem e destino dos passageiros e cargas transportadas no voo, independentemente das escalas realizadas.

Fonte: ANAC. Descrição das variáveis (microdados). Disponível em:

<https://www.gov.br/anac/pt-br/assuntos/dados-e-estatisticas/mercado-do-transporte-aereo/demanda-e-oferta/descricao-das-variaveis-microdados>. Acesso em: 26 nov. 2023

Etapa II.I. Organização e Transformação dos Dados - Estatísticas Combinadas 2019

1. Leitura e agrupamento de arquivos de estatísticas combinadas 2019

Salvo em ./Dados/DemandaeOferta/2019_combinada.csv

Descrição: Nesta etapa foi realizada a leitura e agrupamento dos arquivos de estatísticas combinadas disponibilizados pela ANAC para 2019.

```
python
# Leitura do arquivo:
files = ['combinada2019-01.zip', 'combinada2019-02.zip', 'combinada2019-03.zip',
        'combinada2019-04.zip', 'combinada2019-05.zip', 'combinada2019-06.zip',
        'combinada2019-07.zip', 'combinada2019-08.zip', 'combinada2019-09.zip',
        'combinada2019-10.zip', 'combinada2019-11.zip', 'combinada2019-12.zip']
```

```
python
# Criando o dataframe:
df = pd.DataFrame()
```

```
python
# Concatenando os arquivos no dataframe:
for file in files:
    print(file)
    data = pd.read_csv(f'./Dados/DemandaeOferta/2019_combinada/{file}',
                      encoding='cp1252', delimiter = ';', low_memory=False)
    df = pd.concat([df, data], axis=0, ignore_index=True)
```

```
python
# Visualizando o dataframe:
df.head()
```

```
python
# Salvando o arquivo "estatisticas_2019_combinadas.csv":
df.to_csv('./Dados/DemandaeOferta/estatisticas_2019_combinadas.csv', index=False)
```

2. Obtenção das estatísticas combinadas de Demanda e Oferta em 2019

Salvo em /Dados/Estatisticas/estatisticas9c.csv

Descrição: Nesta etapa foi realizada a transformação do arquivo estatisticas_2019_combinadas.csv.

```
python
# Leitura do arquivo:
dados_estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas_2019_combinadas.csv',
                                low_memory=False)
```

```
python
# Dimensão do arquivo:
dados_estatisticas.shape
```

```
python
```

```
# Padronizando os dados respectivos ao mês:
dados_estatisticas['nr_mes_referencia'].unique()
```

```
python
```

```
dados_estatisticas.columns
```

```
python
```

```
# Filtrando o dataframe com a natureza da etapa de voo = doméstica, serviço referente ao
tipo de linha = passageiro e tipo de Voo = regular:
dados_estatisticas = dados_estatisticas[
    (dados_estatisticas['ds_natureza_etapa'] == 'DOMÉSTICA') & \
    (dados_estatisticas['ds_servico_tipo_linha'] == 'PASSAGEIRO') & \
    (dados_estatisticas['ds_grupo_di'] == 'REGULAR')]
# Com isso, foram excluídos os dados que não serão trabalhados nessa pesquisa.
```

```
python
```

```
dados_estatisticas.info()
```

```
python
```

```
dados_estatisticas.isna().sum()
```

```
python
```

```
# Exclusão das variáveis que não serão estudadas na pesquisa:
dados_estatisticas = dados_estatisticas.drop(columns=
    ['id_combinada', 'id_empresa', 'sg_empresa_iata', 'nm_pais', 'ds_tipo_empresa',
    'nr_singular', 'id_di', 'cd_di', 'ds_di', 'ds_grupo_di', 'dt_referencia',
    'nr_ano_referencia', 'nr_semestre_referencia', 'nm_semestre_referencia',
    'nr_trimestre_referencia', 'nm_trimestre_referencia', 'nr_mes_referencia',
    'nm_mes_referencia', 'nr_semana_referencia', 'nm_dia_semana_referencia',
    'nr_dia_referencia', 'nr_ano_mes_referencia', 'id_tipo_linha',
    'cd_tipo_linha', 'ds_tipo_linha', 'ds_natureza_tipo_linha', 'ds_servico_tipo_linha',
    'ds_natureza_etapa', 'nr_ano_partida_real', 'nr_semestre_partida_real',
    'nm_semestre_partida_real', 'nr_trimestre_partida_real', 'nm_trimestre_partida_real',
    'nm_mes_partida_real', 'nr_ano_mes_partida_real', 'id_aerodromo_origem',
    'nm_aerodromo_origem', 'nm_municipio_origem', 'nm_pais_origem', 'nm_continente_origem',
    'nr_ano_chegada_real', 'nr_semestre_chegada_real', 'nm_semestre_chegada_real',
    'nr_trimestre_chegada_real', 'nm_trimestre_chegada_real', 'nm_mes_chegada_real',
    'nr_ano_mes_chegada_real', 'id_aerodromo_destino', 'nm_aerodromo_destino',
    'nm_municipio_destino', 'nm_regiao_destino', 'nm_pais_destino', 'nm_continente_destino',
    'nr_escala_destino', 'id_arquivo', 'nm_arquivo', 'nr_linha', 'dt_sistema']
)
```

```
python
```

```
# Dimensão do dataframe:
dados_estatisticas.shape
```

```
python
```

```
# Salvando o arquivo "estatisticas9c.csv":
dados_estatisticas.to_csv('./Dados/DemandaeOferta/estatisticas9c.csv',
                          index=False)
```

3. Inclusão de coluna de distancia nas Estatísticas Combinadas 2019

Salvos em ./Dados/DemandaeOferta/estatisticas9c.csv

Descrição: Para a análise do dataframe, é interessante ter a variável distância correlacionada às estatísticas combinadas. Para isso, foram identificadas as variáveis correspondentes aos aeroportos de origem e destino de cada voo no arquivo estatisticas9c e correlacionadas às distâncias do arquivo distancias2.

```
python
# Leitura dos arquivos:
estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas9c.csv')
```

```
python
distancia = pd.read_csv('./Dados/Aerodromos/distancias2.csv')
```

```
python
# Criando um dicionário com as correspondências entre origem e destino, com as distancias
(km):
correspondencias = distancia.set_index(['orig', 'dest'])['dist'].to_dict()
```

```
python
# Criando a variável km_distancia e mapeando os valores de orig/sg_icao_origem e
dest/sg_icao_destino para obter as distâncias correspondentes:
estatisticas['km_distancia'] = estatisticas[['sg_icao_origem',
'sg_icao_destino']].apply(lambda row: correspondencias.get((row['sg_icao_origem'],
row['sg_icao_destino'])), axis=1)
```

```
python
# Visualizando o dataframe:
estatisticas.head()
```

```
python
# Salvando o arquivo "estatisticas9c.csv" com a variável distância inclusa.
estatisticas.to_csv('./Dados/DemandaeOferta/estatisticas9c.csv', index=False)
```

4. Inclusão de coluna de Modelo de Aeronave e Número de Assentos nas Estatísticas Combinadas 2019

Salvos em ./Dados/DemandaeOferta/estatisticas9_filtered.csv

Descrição: Apesar dos bancos de dados de estatísticas combinadas dos voos serem mais adequados para a pesquisa, por representarem a origem e destino dos passageiros no voo - independentemente das escalas realizadas - um grande limitador desses bancos foi a carência de algumas variáveis essenciais para as análises da pesquisa, como o Modelo da Aeronave, Número de Assentos e - principalmente - das Tarifas praticadas para cada voo.

As etapas subsequentes foram realizadas para a inclusão dessas variáveis nos arquivos de estatísticas combinadas, a partir da correlação entre os arquivos de estatísticas básicas e tarifas. As correlações foram realizadas a partir das combinações entre as variáveis-chaves dos arquivos.

```
python
# Leitura dos arquivos:
estatisticasc = pd.read_csv("../Dados/DemandaeOferta/estatisticas9c.csv")
estatisticasb = pd.read_csv("../Dados/DemandaeOferta/estatisticas9.csv")
```

```
python
# Filtrando o dataframe estatísticas básicas por empresas desejadas:
empresas_desejadas = ["ACN", "PAM", "ASO", "AZU", "PTB", "ABJ"]
estatisticasb_filtrado =
estatisticasb[estatisticasb['sg_empresa_icao'].isin(empresas_desejadas)]
```

```
python
# Filtrando o dataframe estatísticas básicas (filtrado) por aeronaves indesejadas:
aeronaves_indesejadas = ["A20N", "A21N", "A319", "A320", "A321", "A332", "A339", "A359",
"B38M", "B737", "B738", "B763", "B77W", "B789", "E195", "E295", "E190"]
condicao = estatisticasb_filtrado['sg Equipamento_icao'].isin(aeronaves_indesejadas)
estatisticasb_filtrado = estatisticasb_filtrado[~condicao]
```

```
python
# Identificando os números de voos com as respectivas aeronaves de interesse:
nr_voos = estatisticasb_filtrado['nr_voo'].unique()
print(nr_voos)
```

```
python
# Filtrando os números de voos encontrados das estatísticas combinadas:
estatisticasc_filtrado = estatisticasc[estatisticasc['nr_voo'].isin(nr_voos)]
```

```
python
# Filtrando o dataframe estatísticas combinadas (filtrado) por empresas desejadas:
empresas_desejadas = ['PAM', 'ABJ', 'ACN', 'PTB', 'AZU', 'ASO']
estatisticasc_filtrado =
estatisticasc_filtrado[estatisticasc_filtrado['sg_empresa_icao'].isin(empresas_desejadas)]
```

```
python
```

```
estatisticasc_filtrado['sg_empresa_icao'].unique()
```

```
python
```

```
# Incluindo a coluna ICAO de aeronave na estatistica combinada
# Criando um dicionário com as correspondências entre nr_voo/sg_empresa_origem e
sg Equipamento_icao
correspondencias = estatisticasb_filtrado.set_index(['nr_voo',
'sg_empresa_icao'])['sg Equipamento_icao'].to_dict()
```

```
python
```

```
# Mapeando os valores de nr_voo/sg_empresa_origem e sg_empresa_icao para obter os modelos
de aeronaves correspondentes
estatisticasc_filtrado['sg Equipamento_icao'] = estatisticasc_filtrado[['nr_voo',
'sg_empresa_icao']].apply(lambda row: correspondencias.get((row['nr_voo'],
row['sg_empresa_icao'])), axis=1)
```

```
python
```

```
# Incluindo a coluna modelo de aeronave na estatistica combinada
# Criando um dicionário com as correspondências entre sg Equipamento_icao e ds_modelo
correspondencias2 =
estatisticasb_filtrado.set_index(['sg Equipamento_icao'])['ds_modelo'].to_dict()
```

```
python
```

```
# Mapeando os valores de sg_empresa_icao e ds_modelo para obter os modelos de aeronaves
correspondentes
estatisticasc_filtrado['ds_modelo'] =
estatisticasc_filtrado[['sg Equipamento_icao']].apply(lambda row:
correspondencias2.get((row['sg Equipamento_icao'])), axis=1)
```

```
python
```

```
# Incluindo a coluna assentos ofertados na estatistica combinada
# Criando um dicionário com as correspondências entre
nr_voo/sg_empresa_icao/sg Equipamento_icao e nr Assentos_ofertados
correspondencias3 = estatisticasb_filtrado.set_index(['nr_voo', 'sg_empresa_icao',
'sg Equipamento_icao'])['nr Assentos_ofertados'].to_dict()
```

```
python
```

```
# Mapeando os valores de nr_voo/sg_empresa_origem e ds_modelo para obter os modelos de
aeronaves correspondentes
estatisticasc_filtrado['nr Assentos_ofertados'] = estatisticasc_filtrado[['nr_voo',
'sg_empresa_icao', 'sg Equipamento_icao']].apply(lambda row:
correspondencias3.get((row['nr_voo'], row['sg_empresa_icao'], row['sg Equipamento_icao'])),
axis=1)
```

```
python
```

```
estatisticasc_filtrado.head()
```

```
python
```

```
estatisticasc_filtrado.shape
```

```
python
# Salvando o arquivo "estatisticas9c_filtered.csv":
estatisticasc_filtrado.to_csv('./Dados/DemandaeOferta/estatisticas9c_filtered.csv',
index=False)
```

5. Inclusão da coluna de Tarifas nas Estatísticas Combinadas 2019

Salvos em ./Dados/DemandaeOferta/estatisticas9c_filtered.csv

Descrição: Para incluir a variável Tarifas nos arquivos de estatísticas combinadas, foi necessário criar uma variável de semestre, pois a variável mês do arquivo tarifas não corresponde à variável mês do arquivo estatísticas combinadas. Isso ocorre, porque o mês de compra da passagem não corresponde necessariamente ao mês do voo.

Dessa forma, a criação da variável semestre foi uma tentativa de aproximação das variáveis para correlacionar ambos os bancos de dados e obter a variável tarifa para cada voo no arquivo estatísticas combinadas.

```
python
# Leitura do arquivo:
estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas9c_filtered.csv')
```

```
python
#Criando a coluna SEMESTRE
estatisticas['SEMESTRE'] = pd.cut(estatisticas['nr_mes_partida_real'], bins=[0, 6, 12],
labels=[1, 2])
```

```
python
#Incluindo a coluna das empresas que comercializam as passagens (sg_comercializa_icao)
empresas = {
    "PAM": "PTB",
    "PTB": "PTB",
    "ACN": "AZU",
    "AZU": "AZU",
    "ABJ": "ABJ"
}
estatisticas['sg_comercializa_icao'] = estatisticas['sg_empresa_icao'].map(empresas)
```

```
python
estatisticas.head()
```

Adequação do arquivo tarifas9.csv

Descrição: Após a criação da variável semestre, foi necessária a criação da variável tarifa ponderada, qua consiste na média ponderada das tarifas praticadas para o

mesmo grupo ('ANO', 'SEMESTRE', 'EMPRESA', 'ORIGEM', 'DESTINO'), ponderadas pela quantidade de assentos.

Após a criação da variável tarifa ponderada, foram calculados os desvios padrões das tarifas ponderadas.

```
python
# Leitura do arquivo:
tarifas = pd.read_csv('./Dados/Tarifas/tarifas9.csv')
```

```
python
#Criando a coluna SEMESTRE
tarifas['SEMESTRE'] = pd.cut(tarifas['MES'], bins=[0, 6, 12], labels=[1, 2])
```

```
python
#Criando a coluna média ponderada das Tarifas
tarifas['produto'] = tarifas['TARIFA'] * tarifas['ASSENTOS']
soma_ponderada = tarifas.groupby(['ANO', 'SEMESTRE', 'EMPRESA', 'ORIGEM', 'DESTINO'])['produto'].sum()
soma_assentos = tarifas.groupby(['ANO', 'SEMESTRE', 'EMPRESA', 'ORIGEM', 'DESTINO'])['ASSENTOS'].sum()
TARIFA_PONDERADA = soma_ponderada / soma_assentos
```

```
python
# Adicionando a variável TARIFA_PONDERADA ao dataframe:
tarifas = tarifas.merge(TARIFA_PONDERADA.rename('TARIFA_PONDERADA'), left_on=['ANO', 'SEMESTRE', 'EMPRESA', 'ORIGEM', 'DESTINO'], right_index=True)
```

```
python
# Excluindo a variável criada provisoriamente para o cálculo da média ponderada da tarifa:
tarifas = tarifas.drop(columns='produto')
```

```
python
tarifas.head()
```

```
python
# Incluindo a coluna tarifas na estatística combinada
# Criando um dicionário com as correspondências entre MES/EMPRESA/ORIGEM/DESTINO e TARIFA
correspondencias = tarifas.set_index(['SEMESTRE', 'EMPRESA', 'ORIGEM', 'DESTINO'])['TARIFA_PONDERADA'].to_dict()
```

```
python
# Mapeando os valores de nr_mes_partida_real/sg_empresa_icao/sg_icao_origem/sg_icao_destino e tarifa para obter os modelos de aeronaves correspondentes
estatisticas['tarifa_ponderada'] = estatisticas[['SEMESTRE', 'sg_comercializa_icao', 'sg_icao_origem', 'sg_icao_destino']].apply(lambda row:
```



```
correspondencias.get((row['SEMESTRE'], row['sg_comercializa_icao'], row['sg_icao_origem'],
row['sg_icao_destino'])), axis=1)
```

```
python
```

```
estatisticas.head()
```

```
python
```

```
#Calculando o Desvio Padrão
tarifas['desvio'] = tarifas['ASSENTOS'] * (tarifas['TARIFA'] -
tarifas['TARIFA_PONDERADA'])**2
soma_desvio = tarifas.groupby(['ANO', 'SEMESTRE', 'EMPRESA', 'ORIGEM',
'DESTINO'])['desvio'].sum()
soma_assentos = tarifas.groupby(['ANO', 'SEMESTRE', 'EMPRESA', 'ORIGEM',
'DESTINO'])['ASSENTOS'].sum()
desvio_padrao = np.sqrt(soma_desvio / (soma_assentos - 1))
```

```
python
```

```
# Agora, vamos adicionar essa série ao dataframe original
tarifas = tarifas.merge(desvio_padrao.rename('DESVIO_PADRAO'), left_on=['ANO', 'SEMESTRE',
'EMPRESA', 'ORIGEM', 'DESTINO'], right_index=True)
```

```
python
```

```
tarifas = tarifas.drop(columns='desvio')
```

```
python
```

```
tarifas.head()
```

```
python
```

```
# Incluindo a coluna DESVIO_PADRAO na estatística combinada
# Criando um dicionário com as correspondências entre
SEMESTRE/EMPRESA/ORIGEM/DESTINO/TARIFA_PONDERADA e DESVIO_PADRAO
correspondencias = tarifas.set_index(['SEMESTRE', 'EMPRESA', 'ORIGEM', 'DESTINO',
'TARIFA_PONDERADA'])['DESVIO_PADRAO'].to_dict()
```

```
python
```

```
# Mapeando os valores de SEMESTRE/sg_empresa_icao/sg_icao_origem/sg_icao_destino/tarifa
para obter o desvio_padrao correspondente
estatisticas['desvio_padrao'] = estatisticas[['SEMESTRE', 'sg_comercializa_icao',
'sg_icao_origem', 'sg_icao_destino', 'tarifa_ponderada']].apply(lambda row:
correspondencias.get((row['SEMESTRE'], row['sg_comercializa_icao'], row['sg_icao_origem'],
row['sg_icao_destino'], row['tarifa_ponderada'])), axis=1)
```

```
python
```

```
estatisticas.head()
```

```
python
```

```
# Salvando o arquivo "estatisticas9c_filtered.csv" com as novas variáveis:
estatisticas.to_csv('./Dados/DemandaeOferta/estatisticas9c_filtered.csv', index=False)
```

Etapa II.II. Organização e Transformação dos Dados - Estatísticas Combinadas 2020

1. Leitura e agrupamento de arquivos de estatísticas combinadas 2020

Salvo em `./Dados/DemandaeOferta/2020_combinada.csv`

Descrição: Nesta etapa foi realizada a leitura e agrupamento dos arquivos de estatísticas combinadas disponibilizados pela ANAC para 2020.

```
python
# Leitura do arquivo:
files = ['combinada2020-01.zip', 'combinada2020-02.zip', 'combinada2020-03.zip',
        'combinada2020-04.zip', 'combinada2020-05.zip', 'combinada2020-06.zip',
        'combinada2020-07.zip', 'combinada2020-08.zip', 'combinada2020-09.zip',
        'combinada2020-10.zip', 'combinada2020-11.zip', 'combinada2020-12.zip']
```

```
python
# Criando o dataframe:
df = pd.DataFrame()
```

```
python
# Concatenando os arquivos no dataframe:
for file in files:
    print(file)
    data = pd.read_csv(f'./Dados/DemandaeOferta/2020_combinada/{file}',
                      encoding='cp1252', delimiter = ';', low_memory=False)
    df = pd.concat([df, data], axis=0, ignore_index=True)
```

```
python
# Visualizando o dataframe:
df.head()
```

```
python
# Salvando o arquivo "estatisticas_2020_combinadas.csv":
df.to_csv('./Dados/DemandaeOferta/estatisticas_2020_combinadas.csv', index=False)
```

2. Obtenção das estatísticas combinadas de Demanda e Oferta em 2020

Salvo em `/Dados/Estatisticas/estatisticas0c.csv`

Descrição: Nesta etapa foi realizada a transformação do arquivo `estatisticas_2020_combinadas.csv`.

```
python
```

```
# Leitura do arquivo:
dados_estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas_2020_combinadas.csv',
                                  low_memory=False)
```

```
python
```

```
# Dimensão do arquivo:
dados_estatisticas.shape
```

```
python
```

```
# Padronizando os dados respectivos ao mês:
dados_estatisticas['nr_mes_referencia'].unique()
```

```
python
```

```
dados_estatisticas.columns
```

```
python
```

```
# Filtrando o dataframe com a natureza da etapa de voo = doméstica, serviço referente ao
tipo de linha = passageiro e tipo de Voo = regular:
dados_estatisticas = dados_estatisticas[
    (dados_estatisticas['ds_natureza_etapa'] == 'DOMÉSTICA') & \
    (dados_estatisticas['ds_servico_tipo_linha'] == 'PASSAGEIRO') & \
    (dados_estatisticas['ds_grupo_di'] == 'REGULAR')]
# Com isso, foram excluídos os dados que não serão trabalhados nessa pesquisa.
```

```
python
```

```
dados_estatisticas.info()
```

```
python
```

```
# Exclusão dos dados vazios (NA)
dados_estatisticas.isna().sum()
```

```
python
```

```
# Exclusão das variáveis que não serão estudadas na pesquisa:
dados_estatisticas = dados_estatisticas.drop(columns=
    ['id_combinada', 'id_empresa', 'sg_empresa_iata', 'nm_pais', 'ds_tipo_empresa',
    'nr_singular', 'id_di', 'cd_di', 'ds_di', 'ds_grupo_di', 'dt_referencia',
    'nr_ano_referencia', 'nr_semestre_referencia', 'nm_semestre_referencia',
    'nr_trimestre_referencia', 'nm_trimestre_referencia', 'nr_mes_referencia',
    'nm_mes_referencia', 'nr_semana_referencia', 'nm_dia_semana_referencia',
    'nr_dia_referencia', 'nr_ano_mes_referencia', 'id_tipo_linha',
    'cd_tipo_linha', 'ds_tipo_linha', 'ds_natureza_tipo_linha', 'ds_servico_tipo_linha',
    'ds_natureza_etapa', 'nr_ano_partida_real', 'nr_semestre_partida_real',
    'nm_semestre_partida_real', 'nr_trimestre_partida_real', 'nm_trimestre_partida_real',
    'nm_mes_partida_real', 'nr_ano_mes_partida_real', 'id_aerodromo_origem',
    'nm_aerodromo_origem', 'nm_municipio_origem', 'nm_pais_origem', 'nm_continente_origem',
    'nr_ano_chegada_real', 'nr_semestre_chegada_real', 'nm_semestre_chegada_real',
    'nr_trimestre_chegada_real', 'nm_trimestre_chegada_real', 'nm_mes_chegada_real',
```

```
'nr_ano_mes_chegada_real', 'id_aerodromo_destino', 'nm_aerodromo_destino',
'nm_municipio_destino', 'nm_regiao_destino', 'nm_pais_destino', 'nm_continente_destino',
'nr_escala_destino', 'id_arquivo', 'nm_arquivo', 'nr_linha', 'dt_sistema']
)
```

```
python
# Dimensão do dataframe:
dados_estatisticas.shape
```

```
python
# Salvando o arquivo "estatisticas0c.csv":
dados_estatisticas.to_csv('./Dados/DemandaeOferta/estatisticas0c.csv',
                          index=False)
```

3. Inclusão de coluna de distancia nas Estatísticas Combinadas 2020

Salvos em ./Dados/DemandaeOferta/estatisticas0c.csv

Descrição: Para a análise do dataframe, é interessante ter a variável distância correlacionada às estatísticas combinadas. Para isso, foram identificadas as variáveis correspondentes aos aeroportos de origem e destino de cada voo no arquivo estatisticas0c e correlacionadas às distâncias do arquivo distancias2.

```
python
# Leitura dos arquivos:
estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas0c.csv')
```

```
python
distancia = pd.read_csv('./Dados/Aerodromos/distancias2.csv')
```

```
python
# Criando um dicionário com as correspondências entre origem e destino, com as distancias
(km):
correspondencias = distancia.set_index(['orig', 'dest'])['dist'].to_dict()
```

```
python
# Criando a variável km_distancia e mapeando os valores de orig/sg_icao_origem e
dest/sg_icao_destino para obter as distâncias correspondentes:
estatisticas['km_distancia'] = estatisticas[['sg_icao_origem',
'sg_icao_destino']].apply(lambda row: correspondencias.get((row['sg_icao_origem'],
row['sg_icao_destino'])), axis=1)
```

```
python
# Visualizando o dataframe:
estatisticas.head()
```

```
python
# Salvando o arquivo "estatisticas0c.csv" com a variável distância inclusa.
estatisticas.to_csv('./Dados/DemandaeOferta/estatisticas0c.csv', index=False)
```

4. Inclusão de coluna de Modelo de Aeronave e Número de Assentos nas Estatísticas Combinadas 2020

Salvos em ./Dados/DemandaeOferta/estatisticas0c_filtered.csv

Descrição: Apesar dos bancos de dados de estatísticas combinadas dos voos serem mais adequados para a pesquisa, por representarem a origem e destino dos passageiros no voo - independentemente das escalas realizadas - um grande limitador desses bancos foi a carência de algumas variáveis essenciais para as análises da pesquisa, como o Modelo da Aeronave, Número de Assentos e - principalmente - das Tarifas praticadas para cada voo.

As etapas subseqüentes foram realizadas para a inclusão dessas variáveis nos arquivos de estatísticas combinadas, a partir da correlação entre os arquivos de estatísticas básicas e tarifas. As correlações foram realizadas a partir das combinações entre as variáveis-chaves dos arquivos.

```
python
# Leitura dos arquivos:
estatisticasc = pd.read_csv('./Dados/DemandaeOferta/estatisticas0c.csv')
estatisticasb = pd.read_csv('./Dados/DemandaeOferta/estatisticas0.csv')
```

```
python
# Filtrando o dataframe estatísticas básicas por empresas desejadas:
empresas_desejadas = ["ACN", "PAM", "ASO", "AZU", "PTB", "ABJ"]
estatisticasb_filtrado =
estatisticasb[estatisticasb['sg_empresa_icao'].isin(empresas_desejadas)]
```

```
python
# Filtrando o dataframe estatísticas básicas (filtrado) por aeronaves indesejadas:
aeronaves_indesejadas = ["A20N", "A21N", "A319", "A320", "A321", "A332", "A339", "A359",
"B38M", "B737", "B738", "B763", "B77W", "B789", "E195", "E295", "E190"]
condicao = estatisticasb_filtrado['sg Equipamento_icao'].isin(aeronaves_indesejadas)
estatisticasb_filtrado = estatisticasb_filtrado[~condicao]
```

```
python
# Identificando os números de voos com as respectivas aeronaves de interesse:
nr_voos = estatisticasb_filtrado['nr_voo'].unique()
print(nr_voos)
```

```
python
# Filtrando os números de voos encontrados das estatísticas combinadas:
estatisticasc_filtrado = estatisticasc[estatisticasc['nr_voo'].isin(nr_voos)]
```

```
python
# Filtrando o dataframe estatisticas combinadas (filtrado) por empresas desejadas:
empresas_desejadas = ['PAM', 'ABJ', 'ACN', 'PTB', 'AZU', 'ASO']
estatisticasc_filtrado =
estatisticasc_filtrado[estatisticasc_filtrado['sg_empresa_icao'].isin(empresas_desejadas)]
```

```
python
estatisticasc_filtrado['sg_empresa_icao'].unique()
```

```
python
# Incluindo a coluna ICAO de aeronave na estatistica combinada
# Criando um dicionário com as correspondências entre nr_voo/sg_empresa_origem e
sg Equipamento_icao
correspondencias = estatisticascb_filtrado.set_index(['nr_voo',
'sg_empresa_icao'])['sg Equipamento_icao'].to_dict()
```

```
python
# Mapeando os valores de nr_voo/sg_empresa_origem e sg_empresa_icao para obter os modelos
de aeronaves correspondentes
estatisticasc_filtrado['sg Equipamento_icao'] = estatisticasc_filtrado[['nr_voo',
'sg_empresa_icao']].apply(lambda row: correspondencias.get((row['nr_voo'],
row['sg_empresa_icao'])), axis=1)
```

```
python
# Incluindo a coluna modelo de aeronave na estatistica combinada
# Criando um dicionário com as correspondências entre sg Equipamento_icao e ds_modelo
correspondencias2 =
estatisticascb_filtrado.set_index(['sg Equipamento_icao'])['ds_modelo'].to_dict()
```

```
python
# Mapeando os valores de sg_empresa_icao e ds_modelo para obter os modelos de aeronaves
correspondentes
estatisticasc_filtrado['ds_modelo'] =
estatisticasc_filtrado[['sg Equipamento_icao']].apply(lambda row:
correspondencias2.get((row['sg Equipamento_icao'])), axis=1)
```

```
python
# Incluindo a coluna assentos ofertados na estatistica combinada
# Criando um dicionário com as correspondências entre
nr_voo/sg_empresa_icao/sg Equipamento_icao e nr Assentos_ofertados
correspondencias3 = estatisticascb_filtrado.set_index(['nr_voo', 'sg_empresa_icao',
'sg Equipamento_icao'])['nr Assentos_ofertados'].to_dict()
```

```
python
# Mapear os valores de nr_voo/sg_empresa_origem e ds_modelo para obter os modelos de
aeronaves correspondentes
estatisticasc_filtrado['nr Assentos_ofertados'] = estatisticasc_filtrado[['nr_voo',
'sg_empresa_icao', 'sg Equipamento_icao']].apply(lambda row:
```

```
correspondencias3.get((row['nr_voo'], row['sg_empresa_icao'], row['sg Equipamento_icao'])),
axis=1)
```

```
python
```

```
estatisticasc_filtrado.head()
```

```
python
```

```
estatisticasc_filtrado.shape
```

```
python
```

```
# Salvando o arquivo "estatisticas0c_filtered.csv":
estatisticasc_filtrado.to_csv('./Dados/DemandaeOferta/estatisticas0c_filtered.csv',
index=False)
```

5. Inclusão da coluna de Tarifas nas Estatísticas Combinadas 2020

Salvos em `./Dados/DemandaeOferta/estatisticas0c_filtered.csv`

Descrição: Para incluir a variável Tarifas nos arquivos de estatísticas combinadas, foi necessário criar uma variável de semestre, pois a variável mês do arquivo tarifas não corresponde à variável mês do arquivo estatísticas combinadas. Isso ocorre, porque o mês de compra da passagem não corresponde necessariamente ao mês do voo.

Dessa forma, a criação da variável semestre foi uma tentativa de aproximação das variáveis para correlacionar ambos os bancos de dados e obter a variável tarifa para cada voo no arquivo estatísticas combinadas.

```
python
```

```
# Leitura do arquivo:
estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas0c_filtered.csv')
```

```
python
```

```
#Criando a coluna SEMESTRE
estatisticas['SEMESTRE'] = pd.cut(estatisticas['nr_mes_partida_real'], bins=[0, 6, 12],
labels=[1, 2])
```

```
python
```

```
#Incluindo a coluna das empresas que comercializam as passagens (sg_comercializa_icao)
empresas = {
    "PAM": "PTB",
    "PTB": "PTB",
    "ACN": "AZU",
    "AZU": "AZU",
    "ABJ": "ABJ"
}
estatisticas['sg_comercializa_icao'] = estatisticas['sg_empresa_icao'].map(empresas)
```

```
python
estatisticas.head()
```

Adequação do arquivo tarifas0.csv

Descrição: Após a criação da variável semestre, foi necessária a criação da variável tarifa ponderada, qua consiste na média ponderada das tarifas praticadas para o mesmo grupo ('ANO', 'SEMESTRE', 'EMPRESA', 'ORIGEM', 'DESTINO'), ponderadas pela quantidade de assentos.

Após a criação da variável tarifa ponderada, foram calculados os desvios padrões das tarifas ponderadas.

```
python
# Leitura do arquivo:
tarifas = pd.read_csv('./Dados/Tarifas/tarifas0.csv')
```

```
python
#Criando a coluna SEMESTRE
tarifas['SEMESTRE'] = pd.cut(tarifas['MES'], bins=[0, 6, 12], labels=[1, 2])
```

```
python
#Criando a coluna média ponderada das Tarifas
tarifas['produto'] = tarifas['TARIFA'] * tarifas['ASSENTOS']
soma_ponderada = tarifas.groupby(['ANO', 'SEMESTRE', 'EMPRESA', 'ORIGEM',
'DESTINO'])['produto'].sum()
soma_assentos = tarifas.groupby(['ANO', 'SEMESTRE', 'EMPRESA', 'ORIGEM',
'DESTINO'])['ASSENTOS'].sum()
TARIFA_PONDERADA = soma_ponderada / soma_assentos
```

```
python
# Adicionando a variável TARIFA_PONDERADA ao dataframe:
tarifas = tarifas.merge(TARIFA_PONDERADA.rename('TARIFA_PONDERADA'), left_on=['ANO',
'SEMESTRE', 'EMPRESA', 'ORIGEM', 'DESTINO'], right_index=True)
```

```
python
# Excluindo a variável criada provisoriamente para o cálculo da média ponderada da tarifa:
tarifas = tarifas.drop(columns='produto')
```

```
python
tarifas.head()
```

```
python
# Incluindo a coluna tarifas na estatistica combinada
# Criando um dicionário com as correspondências entre MES/EMPRESA/ORIGEM/DESTINO e TARIFA
```



```
correspondencias = tarifas.set_index(['SEMESTRE', 'EMPRESA', 'ORIGEM',
'DESTINO'])['TARIFA_PONDERADA'].to_dict()
```

```
python
```

```
# Mapeando os valores de nr_mes_partida_real/sg_empresa_icao/sg_icao_origem/sg_icao_destino
e tarifa para obter os modelos de aeronaves correspondentes
estatisticas['tarifa_ponderada'] = estatisticas[['SEMESTRE', 'sg_comercializa_icao',
'sg_icao_origem', 'sg_icao_destino']].apply(lambda row:
correspondencias.get((row['SEMESTRE'], row['sg_comercializa_icao'], row['sg_icao_origem'],
row['sg_icao_destino'])), axis=1)
```

```
python
```

```
estatisticas.head()
```

```
python
```

```
#Calculando o Desvio Padrão
tarifas['desvio'] = tarifas['ASSENTOS'] * (tarifas['TARIFA'] -
tarifas['TARIFA_PONDERADA'])*2
soma_desvio = tarifas.groupby(['ANO', 'SEMESTRE', 'EMPRESA', 'ORIGEM',
'DESTINO'])['desvio'].sum()
soma_assentos = tarifas.groupby(['ANO', 'SEMESTRE', 'EMPRESA', 'ORIGEM',
'DESTINO'])['ASSENTOS'].sum()
desvio_padrao = np.sqrt(soma_desvio / (soma_assentos - 1))
```

```
python
```

```
# Agora, vamos adicionar essa série ao dataframe original
tarifas = tarifas.merge(desvio_padrao.rename('DESVIO_PADRAO'), left_on=['ANO', 'SEMESTRE',
'EMPRESA', 'ORIGEM', 'DESTINO'], right_index=True)
```

```
python
```

```
tarifas = tarifas.drop(columns='desvio')
```

```
python
```

```
tarifas.head()
```

```
python
```

```
# Incluindo a coluna DESVIO_PADRAO na estatística combinada
# Criando um dicionário com as correspondências entre
SEMESTRE/EMPRESA/ORIGEM/DESTINO/TARIFA_PONDERADA e DESVIO_PADRAO
correspondencias = tarifas.set_index(['SEMESTRE', 'EMPRESA', 'ORIGEM', 'DESTINO',
'TARIFA_PONDERADA'])['DESVIO_PADRAO'].to_dict()
```

```
python
```

```
# Mapeando os valores de SEMESTRE/sg_empresa_icao/sg_icao_origem/sg_icao_destino/tarifa
para obter o desvio_padrao correspondente
estatisticas['desvio_padrao'] = estatisticas[['SEMESTRE', 'sg_comercializa_icao',
'sg_icao_origem', 'sg_icao_destino', 'tarifa_ponderada']].apply(lambda row:
```

```
correspondencias.get((row['SEMESTRE'], row['sg_comercializa_icao'], row['sg_icao_origem'],
row['sg_icao_destino'], row['tarifa_ponderada'])), axis=1)
```

```
python
```

```
estatisticas.head()
```

```
python
```

```
# Salvando o arquivo "estatisticas0c_filtered.csv" com as novas variáveis:
estatisticas.to_csv('./Dados/DemandaeOferta/estatisticas0c_filtered.csv', index=False)
```

Etapa II.III. Organização e Transformação dos Dados - Estatísticas Combinadas 2021

1. Leitura e agrupamento de arquivos de estatísticas combinadas 2021

Salvo em `./Dados/DemandaeOferta/2021_combinada.csv`

Descrição: Nesta etapa foi realizada a leitura e agrupamento dos arquivos de estatísticas combinadas disponibilizados pela ANAC para 2021.

```
python
```

```
# Leitura do arquivo:
files = ['combinada2021-01.zip', 'combinada2021-02.zip', 'combinada2021-03.zip',
         'combinada2021-04.zip', 'combinada2021-05.zip', 'combinada2021-06.zip',
         'combinada2021-07.zip', 'combinada2021-08.zip', 'combinada2021-09.zip',
         'combinada2021-10.zip', 'combinada2021-11.zip', 'combinada2021-12.zip']
```

```
python
```

```
# Criando o dataframe:
df = pd.DataFrame()
```

```
python
```

```
# Concatenando os arquivos no dataframe:
for file in files:
    print(file)
    data = pd.read_csv(f'./Dados/DemandaeOferta/2021_combinada/{file}',
                      encoding='cp1252', delimiter = ';', low_memory=False)
    df = pd.concat([df, data], axis=0, ignore_index=True)
```

```
python
```

```
# Visualizando o dataframe:
df.head()
```

```
python
```

```
# Salvando o arquivo "estatisticas_2021_combinadas.csv":
df.to_csv('./Dados/DemandaeOferta/estatisticas_2021_combinadas.csv', index=False)
```

2. Obtenção das estatísticas combinadas de Demanda e Oferta em 2021

Salvo em /Dados/Estatisticas/estatisticas1c.csv

Descrição: Nesta etapa foi realizada a transformação do arquivo estatisticas_2021_combinadas.csv.

```
python
# Leitura do arquivo:
dados_estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas_2021_combinadas.csv',
                                  low_memory=False)
```

```
python
# Dimensão do arquivo:
dados_estatisticas.shape
```

```
python
dados_estatisticas['nr_mes_referencia'].unique()
```

```
python
dados_estatisticas.columns
```

```
python
# Filtrando o dataframe com a natureza da etapa de voo = doméstica, serviço referente ao
# tipo de linha = passageiro e tipo de Voo = regular:
dados_estatisticas = dados_estatisticas[
    (dados_estatisticas['ds_natureza_etapa'] == 'DOMÉSTICA') & \
    (dados_estatisticas['ds_servico_tipo_linha'] == 'PASSAGEIRO') & \
    (dados_estatisticas['ds_grupo_di'] == 'REGULAR')]
# Com isso, foram excluídos os dados que não serão trabalhados nessa pesquisa.
```

```
python
dados_estatisticas.info()
```

```
python
# Exclusão dos dados vazios do dataframe (NA)
dados_estatisticas.isna().sum()
```

```
python
# Exclusão das variáveis que não serão estudadas na pesquisa:
dados_estatisticas = dados_estatisticas.drop(columns=
    ['id_combinada', 'id_empresa', 'sg_empresa_iata', 'nm_pais', 'ds_tipo_empresa',
     'nr_singular', 'id_di', 'cd_di', 'ds_di', 'ds_grupo_di', 'dt_referencia',
     'nr_ano_referencia', 'nr_semestre_referencia', 'nm_semestre_referencia',
     'nr_trimestre_referencia', 'nm_trimestre_referencia', 'nr_mes_referencia',
```

```
'nm_mes_referencia','nr_semana_referencia','nm_dia_semana_referencia',
'nr_dia_referencia','nr_ano_mes_referencia','id_tipo_linha',
'cd_tipo_linha','ds_tipo_linha','ds_natureza_tipo_linha','ds_servico_tipo_linha',
'ds_natureza_etapa','nr_ano_partida_real','nr_semestre_partida_real',
'nm_semestre_partida_real','nr_trimestre_partida_real','nm_trimestre_partida_real',
'nm_mes_partida_real','nr_ano_mes_partida_real','id_aerodromo_origem',
'nm_aerodromo_origem','nm_municipio_origem','nm_pais_origem','nm_continente_origem',
'nr_ano_chegada_real','nr_semestre_chegada_real','nm_semestre_chegada_real',
'nr_trimestre_chegada_real','nm_trimestre_chegada_real','nm_mes_chegada_real',
'nr_ano_mes_chegada_real','id_aerodromo_destino','nm_aerodromo_destino',
'nm_municipio_destino','nm_regiao_destino','nm_pais_destino','nm_continente_destino',
'nr_escala_destino','id_arquivo','nm_arquivo','nr_linha','dt_sistema']
)
```

```
python
```

```
# Dimensão do dataframe:
dados_estatisticas.shape
```

```
python
```

```
# Salvando o arquivo "estatisticas1c.csv":
dados_estatisticas.to_csv('./Dados/DemandaeOferta/estatisticas1c.csv',
                          index=False)
```

3. Inclusão de coluna de distancia nas Estatísticas Combinadas 2021

Salvos em ./Dados/DemandaeOferta/estatisticas1c.csv

Descrição: Para a análise do dataframe, é interessante ter a variável distância correlacionada às estatísticas combinadas. Para isso, foram identificadas as variáveis correspondentes aos aeroportos de origem e destino de cada voo no arquivo estatisticas1c e correlacionadas às distâncias do arquivo distancias2.

```
python
```

```
# Leitura dos arquivos:
estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas1c.csv')
```

```
python
```

```
distancia = pd.read_csv('./Dados/Aerodromos/distancias2.csv')
```

```
python
```

```
# Criando um dicionário com as correspondências entre origem e destino, com as distancias (km):
correspondencias = distancia.set_index(['orig', 'dest'])['dist'].to_dict()
```

```
python
```

```
# Criando a variável km_distancia e mapeando os valores de orig/sg_icao_origem e dest/sg_icao_destino para obter as distâncias correspondentes:
```

```
estatisticas['km_distancia'] = estatisticas[['sg_icao_origem',
'sg_icao_destino']].apply(lambda row: correspondencias.get((row['sg_icao_origem'],
row['sg_icao_destino'])), axis=1)
```

```
python
```

```
# Visualizando o dataframe:
estatisticas.head()
```

```
python
```

```
# Salvando o arquivo "estatisticas9c.csv" com a variável distância inclusa.
estatisticas.to_csv('./Dados/DemandaeOferta/estatisticas1c.csv', index=False)
```

4. Inclusão de coluna de Modelo de Aeronave e Número de Assentos nas Estatísticas Combinadas 2021

Salvos em ./Dados/DemandaeOferta/estatisticas1c_filtered.csv

Descrição: Apesar dos bancos de dados de estatísticas combinadas dos voos serem mais adequados para a pesquisa, por representarem a origem e destino dos passageiros no voo - independentemente das escalas realizadas - um grande limitador desses bancos foi a carência de algumas variáveis essenciais para as análises da pesquisa, como o Modelo da Aeronave, Número de Assentos e - principalmente - das Tarifas praticadas para cada voo.

As etapas subsequentes foram realizadas para a inclusão dessas variáveis nos arquivos de estatísticas combinadas, a partir da correlação entre os arquivos de estatísticas básicas e tarifas. As correlações foram realizadas a partir das combinações entre as variáveis-chaves dos arquivos.

```
python
```

```
# Leitura dos arquivos:
estatisticasc = pd.read_csv("./Dados/DemandaeOferta/estatisticas1c.csv")
estatisticasb = pd.read_csv("./Dados/DemandaeOferta/estatisticas1.csv")
```

```
python
```

```
# Filtrando o dataframe estatisticas básicas por empresas desejadas:
empresas_desejadas = ["ACN", "PAM", "ASO", "AZU", "PTB", "ABJ"]
estatisticasb_filtrado =
estatisticasb[estatisticasb['sg_empresa_icao'].isin(empresas_desejadas)]
```

```
python
```

```
# Filtrando o dataframe estatisticas básicas (filtrado) por aeronaves indesejadas:
aeronaves_indesejadas = ["A20N", "A21N", "A319", "A320", "A321", "A332", "A339", "A359",
"B38M", "B737", "B738", "B763", "B77W", "B789", "E195", "E295", "E190"]
condicao = estatisticasb_filtrado['sg Equipamento_icao'].isin(aeronaves_indesejadas)
estatisticasb_filtrado = estatisticasb_filtrado[~condicao]
```

```
python
```

```
# Identificando os números de voos com as respectivas aeronaves de interesse:
nr_voos = estatisticasb_filtrado['nr_voo'].unique()
print(nr_voos)
```

```
python
```

```
# Filtrando os números de voos encontrados das estatísticas combinadas:
estatisticasc_filtrado = estatisticasc[estatisticasc['nr_voo'].isin(nr_voos)]
```

```
python
```

```
# Filtrando o dataframe estatísticas combinadas (filtrado) por empresas desejadas:
empresas_desejadas = ['PAM', 'ABJ', 'ACN', 'PTB', 'AZU', 'ASO']
estatisticasc_filtrado =
estatisticasc_filtrado[estatisticasc_filtrado['sg_empresa_icao'].isin(empresas_desejadas)]
```

```
python
```

```
estatisticasc_filtrado['sg_empresa_icao'].unique()
```

```
python
```

```
# Incluindo a coluna ICAO de aeronave na estatística combinada
# Criando um dicionário com as correspondências entre nr_voo/sg_empresa_origem e
sg Equipamento_icao
correspondencias = estatisticasb_filtrado.set_index(['nr_voo',
'sg_empresa_icao'])['sg Equipamento_icao'].to_dict()
```

```
python
```

```
# Mapeando os valores de nr_voo/sg_empresa_origem e sg_empresa_icao para obter os modelos
de aeronaves correspondentes
estatisticasc_filtrado['sg Equipamento_icao'] = estatisticasc_filtrado[['nr_voo',
'sg_empresa_icao']].apply(lambda row: correspondencias.get((row['nr_voo'],
row['sg_empresa_icao'])), axis=1)
```

```
python
```

```
# Incluindo a coluna modelo de aeronave na estatística combinada
# Criando um dicionário com as correspondências entre sg Equipamento_icao e ds_modelo
correspondencias2 =
estatisticasb_filtrado.set_index(['sg Equipamento_icao'])['ds_modelo'].to_dict()
```

```
python
```

```
# Criando os valores de sg_empresa_icao e ds_modelo para obter os modelos de aeronaves
correspondentes
estatisticasc_filtrado['ds_modelo'] =
estatisticasc_filtrado[['sg Equipamento_icao']].apply(lambda row:
correspondencias2.get((row['sg Equipamento_icao'])), axis=1)
```

```
python
```

```
# Incluindo a coluna assentos ofertados na estatística combinada
```

```
# Criando um dicionário com as correspondências entre
nr_voo/sg_empresa_icao/sg Equipamento_icao e nr Assentos_ofertados
correspondencias3 = estatisticasc_filtrado.set_index(['nr_voo', 'sg_empresa_icao',
'sg Equipamento_icao'])['nr Assentos_ofertados'].to_dict()
```

```
python
```

```
# Mapeando os valores de nr_voo/sg_empresa_origem e ds_modelo para obter os modelos de
aeronaves correspondentes
estatisticasc_filtrado['nr Assentos_ofertados'] = estatisticasc_filtrado[['nr_voo',
'sg_empresa_icao', 'sg Equipamento_icao']].apply(lambda row:
correspondencias3.get((row['nr_voo'], row['sg_empresa_icao'], row['sg Equipamento_icao'])),
axis=1)
```

```
python
```

```
estatisticasc_filtrado.head()
```

```
python
```

```
estatisticasc_filtrado[estatisticasc_filtrado["sg_empresa_icao"] == "ABJ"]
```

```
python
```

```
# Salvando o arquivo "estatisticasc_filtered.csv":
estatisticasc_filtrado.to_csv('./Dados/DemandaeOferta/estatisticasc_filtered.csv',
index=False)
```

5. Inclusão da coluna de Tarifas nas Estatísticas Combinadas 2021

Salvos em ./Dados/DemandaeOferta/estatisticasc1c_filtered.csv

Descrição: Para incluir a variável Tarifas nos arquivos de estatísticas combinadas, foi necessário criar uma variável de semestre, pois a variável mês do arquivo tarifas não corresponde à variável mês do arquivo estatísticas combinadas. Isso ocorre, porque o mês de compra da passagem não corresponde necessariamente ao mês do voo.

Dessa forma, a criação da variável semestre foi uma tentativa de aproximação das variáveis para correlacionar ambos os bancos de dados e obter a variável tarifa para cada voo no arquivo estatísticas combinadas.

```
python
```

```
# Leitura do arquivo:
estatisticasc = pd.read_csv('./Dados/DemandaeOferta/estatisticasc1c_filtered.csv')
```

```
python
```

```
#Criando a coluna SEMESTRE
estatisticasc['SEMESTRE'] = pd.cut(estatisticasc['nr_mes_partida_real'], bins=[0, 6, 12],
labels=[1, 2])
```

```
python
#Incluindo a coluna das empresas que comercializam as passagens (sg_comercializa_icao)
empresas = {
    "PAM": "PTB",
    "PTB": "PTB",
    "ACN": "AZU",
    "AZU": "AZU",
    "ABJ": "ABJ"
}
estatisticas['sg_comercializa_icao'] = estatisticas['sg_empresa_icao'].map(empresas)
```

```
python
estatisticas.head()
```

Adequação do arquivo tarifas1.csv

Descrição: Após a criação da variável semestre, foi necessária a criação da variável tarifa ponderada, qua consiste na média ponderada das tarifas praticadas para o mesmo grupo ('ANO', 'SEMESTRE', 'EMPRESA', 'ORIGEM', 'DESTINO'), ponderadas pela quantidade de assentos.

Após a criação da variável tarifa ponderada, foram calculados os desvios padrões das tarifas ponderadas.

```
python
# Leitura do arquivo:
tarifas = pd.read_csv('./Dados/Tarifas/tarifas1.csv')
```

```
python
#Criando a coluna SEMESTRE
tarifas['SEMESTRE'] = pd.cut(tarifas['MES'], bins=[0, 6, 12], labels=[1, 2])
```

```
python
#Criando a coluna média ponderada das Tarifas
tarifas['produto'] = tarifas['TARIFA'] * tarifas['ASSENTOS']
soma_ponderada = tarifas.groupby(['ANO', 'SEMESTRE', 'EMPRESA', 'ORIGEM', 'DESTINO'])['produto'].sum()
soma_assentos = tarifas.groupby(['ANO', 'SEMESTRE', 'EMPRESA', 'ORIGEM', 'DESTINO'])['ASSENTOS'].sum()
TARIFA_PONDERADA = soma_ponderada / soma_assentos
```

```
python
# Adicionando a variável TARIFA_PONDERADA ao dataframe:
tarifas = tarifas.merge(TARIFA_PONDERADA.rename('TARIFA_PONDERADA'), left_on=['ANO', 'SEMESTRE', 'EMPRESA', 'ORIGEM', 'DESTINO'], right_index=True)
```

```
python
# Excluindo a variável criada provisoriamente para o cálculo da média ponderada da tarifa:
```



```
tarifas = tarifas.drop(columns='produto')
```

```
python
```

```
tarifas.head()
```

```
python
```

```
# Incluindo a coluna tarifas na estatística combinada
# Criando um dicionário com as correspondências entre MES/EMPRESA/ORIGEM/DESTINO e TARIFA
correspondencias = tarifas.set_index(['SEMESTRE', 'EMPRESA', 'ORIGEM',
'DESTINO'])['TARIFA_PONDERADA'].to_dict()
```

```
python
```

```
# Mapeando os valores de nr_mes_partida_real/sg_empresa_icao/sg_icao_origem/sg_icao_destino
e tarifa para obter os modelos de aeronaves correspondentes
estatisticas['tarifa_ponderada'] = estatisticas[['SEMESTRE', 'sg_comercializa_icao',
'sg_icao_origem', 'sg_icao_destino']].apply(lambda row:
correspondencias.get((row['SEMESTRE'], row['sg_comercializa_icao'], row['sg_icao_origem'],
row['sg_icao_destino'])), axis=1)
```

```
python
```

```
estatisticas.head()
```

```
python
```

```
#Calculando o Desvio Padrão
tarifas['desvio'] = tarifas['ASSENTOS'] * (tarifas['TARIFA'] -
tarifas['TARIFA_PONDERADA'])**2
soma_desvio = tarifas.groupby(['ANO', 'SEMESTRE', 'EMPRESA', 'ORIGEM',
'DESTINO'])['desvio'].sum()
soma_assentos = tarifas.groupby(['ANO', 'SEMESTRE', 'EMPRESA', 'ORIGEM',
'DESTINO'])['ASSENTOS'].sum()
desvio_padrao = np.sqrt(soma_desvio / (soma_assentos - 1))
```

```
python
```

```
# Agora, vamos adicionar essa série ao dataframe original
tarifas = tarifas.merge(desvio_padrao.rename('DESVIO_PADRAO'), left_on=['ANO', 'SEMESTRE',
'EMPRESA', 'ORIGEM', 'DESTINO'], right_index=True)
```

```
python
```

```
tarifas = tarifas.drop(columns='desvio')
```

```
python
```

```
tarifas.head()
```

```
python
```

```
# Incluindo a coluna DESVIO_PADRAO na estatística combinada
```

```
# Criando um dicionário com as correspondências entre
SEMESTRE/EMPRESA/ORIGEM/DESTINO/TARIFA_PONDERADA e DESVIO_PADRAO
correspondencias = tarifas.set_index(['SEMESTRE', 'EMPRESA', 'ORIGEM', 'DESTINO',
'TARIFA_PONDERADA'])['DESVIO_PADRAO'].to_dict()
```

```
python
# Mapeando os valores de SEMESTRE/sg_empresa_icao/sg_icao_origem/sg_icao_destino/tarifa
para obter o desvio_padrao correspondente
estatisticas['desvio_padrao'] = estatisticas[['SEMESTRE', 'sg_comercializa_icao',
'sg_icao_origem', 'sg_icao_destino', 'tarifa_ponderada']].apply(lambda row:
correspondencias.get((row['SEMESTRE'], row['sg_comercializa_icao'], row['sg_icao_origem'],
row['sg_icao_destino'], row['tarifa_ponderada'])), axis=1)
```

```
python
estatisticas.head()
```

```
python
# Salvando o arquivo "estatisticaslc_filtered.csv" com as novas variáveis:
estatisticas.to_csv('./Dados/DemandaeOferta/estatisticaslc_filtered.csv', index=False)
```

Etapa II.IV. Organização e Transformação dos Dados - Estatísticas Combinadas 2022

1. Leitura e agrupamento de arquivos de estatísticas combinadas 2022

Salvo em `./Dados/DemandaeOferta/2022_combinada.csv`

Descrição: Nesta etapa foi realizada a leitura e agrupamento dos arquivos de estatísticas combinadas disponibilizados pela ANAC para 2022.

```
python
# Leitura do arquivo:
files = ['combinada2022-01.zip', 'combinada2022-02.zip', 'combinada2022-03.zip',
'combinada2022-04.zip', 'combinada2022-05.zip', 'combinada2022-06.zip',
'combinada2022-07.zip', 'combinada2022-08.zip', 'combinada2022-09.zip',
'combinada2022-10.zip', 'combinada2022-11.zip', 'combinada2022-12.zip']
```

```
python
# Criando o dataframe:
df = pd.DataFrame()
```

```
python
# Concatenando os arquivos no dataframe:
for file in files:
    print(file)
    data = pd.read_csv(f'./Dados/DemandaeOferta/2022_combinada/{file}',
encoding='cp1252', delimiter = ';', low_memory=False)
    df = pd.concat([df, data], axis=0, ignore_index=True)
```

```
python
```

```
# Visualizando o dataframe:
df.head()
```

```
python
```

```
# Salvando o arquivo "estatisticas_2022_combinadas.csv":
df.to_csv('./Dados/DemandaeOferta/estatisticas_2022_combinadas.csv', index=False)
```

2. Obtenção das estatísticas combinadas de Demanda e Oferta em 2022

Salvo em /Dados/Estatisticas/estatisticas2c.csv

Descrição: Nesta etapa foi realizada a transformação do arquivo estatisticas_2022_combinadas.csv.

```
python
```

```
# Leitura do arquivo:
dados_estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas_2022_combinadas.csv',
                                  low_memory=False)
```

```
python
```

```
# Dimensão do arquivo:
dados_estatisticas.shape
```

```
python
```

```
# Padronizando os dados respectivos ao mês:
dados_estatisticas['nr_mes_referencia'].unique()
```

```
python
```

```
dados_estatisticas.columns
```

```
python
```

```
# Filtrando o dataframe com a natureza da etapa de voo = doméstica, serviço referente ao
tipo de linha = passageiro e tipo de Voo = regular:
dados_estatisticas = dados_estatisticas[
    (dados_estatisticas['ds_natureza_etapa'] == 'DOMÉSTICA') & \
    (dados_estatisticas['ds_servico_tipo_linha'] == 'PASSAGEIRO') & \
    (dados_estatisticas['ds_grupo_di'] == 'REGULAR')]
# Com isso, foram excluídos os dados que não serão trabalhados nessa pesquisa.
```

```
python
```

```
dados_estatisticas.info()
```

```
python
```

```
dados_estatisticas.isna().sum()
```

```
python
```

```
# Exclusão das variáveis que não serão estudadas na pesquisa:
dados_estatisticas = dados_estatisticas.drop(columns=
    ['id_combinada', 'id_empresa', 'sg_empresa_iata', 'nm_pais', 'ds_tipo_empresa',
     'nr_singular', 'id_di', 'cd_di', 'ds_di', 'ds_grupo_di', 'dt_referencia',
     'nr_ano_referencia', 'nr_semestre_referencia', 'nm_semestre_referencia',
     'nr_trimestre_referencia', 'nm_trimestre_referencia', 'nr_mes_referencia',
     'nm_mes_referencia', 'nr_semana_referencia', 'nm_dia_semana_referencia',
     'nr_dia_referencia', 'nr_ano_mes_referencia', 'id_tipo_linha',
     'cd_tipo_linha', 'ds_tipo_linha', 'ds_natureza_tipo_linha', 'ds_servico_tipo_linha',
     'ds_natureza_etapa', 'nr_ano_partida_real', 'nr_semestre_partida_real',
     'nm_semestre_partida_real', 'nr_trimestre_partida_real', 'nm_trimestre_partida_real',
     'nm_mes_partida_real', 'nr_ano_mes_partida_real', 'id_aerodromo_origem',
     'nm_aerodromo_origem', 'nm_municipio_origem', 'nm_pais_origem', 'nm_continente_origem',
     'nr_ano_chegada_real', 'nr_semestre_chegada_real', 'nm_semestre_chegada_real',
     'nr_trimestre_chegada_real', 'nm_trimestre_chegada_real', 'nm_mes_chegada_real',
     'nr_ano_mes_chegada_real', 'id_aerodromo_destino', 'nm_aerodromo_destino',
     'nm_municipio_destino', 'nm_regiao_destino', 'nm_pais_destino', 'nm_continente_destino',
     'nr_escala_destino', 'id_arquivo', 'nm_arquivo', 'nr_linha', 'dt_sistema']
)
```

```
python
```

```
# Dimensão do dataframe:
dados_estatisticas.shape
```

```
python
```

```
# Salvando o arquivo "estatisticas2c.csv":
dados_estatisticas.to_csv('./Dados/DemandaeOferta/estatisticas2c.csv',
                          index=False)
```

3. Inclusão de coluna de distancia nas Estatísticas Combinadas 2022

Salvos em ./Dados/DemandaeOferta/estatisticas2c.csv

Descrição: Para a análise do dataframe, é interessante ter a variável distância correlacionada às estatísticas combinadas. Para isso, foram identificadas as variáveis correspondentes aos aeroportos de origem e destino de cada voo no arquivo estatisticas2c e correlacionadas às distâncias do arquivo distancias2.

```
python
```

```
# Leitura dos arquivos:
estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas2c.csv')
```

```
python
```

```
distancia = pd.read_csv('./Dados/Aerodromos/distancias2.csv')
```

```
python
```

```
# Criando um dicionário com as correspondências entre origem e destino, com as distancias
(km):
correspondencias = distancia.set_index(['orig', 'dest'])['dist'].to_dict()
```

```
python
```

```
# Criando a variável km_distancia e mapeando os valores de orig/sg_icao_origem e
dest/sg_icao_destino para obter as distâncias correspondentes:
estatisticas['km_distancia'] = estatisticas[['sg_icao_origem',
'sg_icao_destino']].apply(lambda row: correspondencias.get((row['sg_icao_origem'],
row['sg_icao_destino'])), axis=1)
```

```
python
```

```
# Visualizando o dataframe:
estatisticas.head()
```

```
python
```

```
# Salvando o arquivo "estatisticas2c.csv" com a variável distância inclusa.
estatisticas.to_csv('./Dados/DemandaeOferta/estatisticas2c.csv', index=False)
```

4. Inclusão de coluna de Modelo de Aeronave e Número de Assentos nas Estatísticas Combinadas 2022

Salvos em ./Dados/DemandaeOferta/estatisticas2c_filtered.csv

Descrição: Apesar dos banco de dados de estatísticas combinadas dos voos serem mais adequados para a pesquisa, por representarem a origem e destino dos passageiros no voo - independentemente das escalas realizadas - um grande limitador desses bancos foi a carência de algumas variáveis essenciais para as análises da pesquisa, como o Modelo da Aeronave, Número de Assentos e - principalmente - das Tarifas praticadas para cada voo.

As etapas subsequentes foram realizadas para a inclusão dessas variáveis nos arquivos de estatísticas combinadas, a partir da correlação entre os arquivos de estatísticas básicas e tarifas. As correlações foram realizadas a partir das combinações entre as variáveis-chaves dos arquivos.

```
python
```

```
# Leitura dos arquivos:
estatisticasc = pd.read_csv("./Dados/DemandaeOferta/estatisticas2c.csv")
estatisticasb = pd.read_csv("./Dados/DemandaeOferta/estatisticas2.csv")
```

```
python
```

```
# Filtrando o dataframe estatisticas básicas por empresas desejadas:
empresas_desejadas = ["ACN", "PAM", "ASO", "AZU", "PTB", "ABJ"]
estatisticasb_filtrado =
estatisticasb[estatisticasb['sg_empresa_icao'].isin(empresas_desejadas)]
```

```
python
```

```
# Filtrando o dataframe estatísticas básicas (filtrado) por aeronaves indesejadas:
aeronaves_indesejadas = ["A20N", "A21N", "A319", "A320", "A321", "A332", "A339", "A359",
"B38M", "B737", "B738", "B763", "B77W", "B789", "E195", "E295", "E190"]
condição = estatisticasb_filtrado['sg_equipamento_icao'].isin(aeronaves_indesejadas)
estatisticasb_filtrado = estatisticasb_filtrado[~condição]
```

```
python
```

```
# Identificando os números de voos com as respectivas aeronaves de interesse:
nr_voos = estatisticasb_filtrado['nr_voo'].unique()
print(nr_voos)
```

```
python
```

```
# Filtrando os números de voos encontrados das estatísticas combinadas:
estatisticasc_filtrado = estatisticasc[estatisticasc['nr_voo'].isin(nr_voos)]
```

```
python
```

```
# Filtrando o dataframe estatísticas combinadas (filtrado) por empresas desejadas:
empresas_desejadas = ['PAM', 'ABJ', 'ACN', 'PTB', 'AZU', 'ASO']
estatisticasc_filtrado =
estatisticasc_filtrado[estatisticasc_filtrado['sg_empresa_icao'].isin(empresas_desejadas)]
```

```
python
```

```
estatisticasc_filtrado['sg_empresa_icao'].unique()
```

```
python
```

```
# Incluindo a coluna ICAO de aeronave na estatística combinada
# Criando um dicionário com as correspondências entre nr_voo/sg_empresa_origem e
sg_equipamento_icao
correspondencias = estatisticasb_filtrado.set_index(['nr_voo',
'sg_empresa_icao'])['sg_equipamento_icao'].to_dict()
```

```
python
```

```
# Mapeando os valores de nr_voo/sg_empresa_origem e sg_empresa_icao para obter os modelos
de aeronaves correspondentes
estatisticasc_filtrado['sg_equipamento_icao'] = estatisticasc_filtrado[['nr_voo',
'sg_empresa_icao']].apply(lambda row: correspondencias.get((row['nr_voo'],
row['sg_empresa_icao'])), axis=1)
```

```
python
```

```
# Incluindo a coluna modelo de aeronave na estatística combinada
# Criando um dicionário com as correspondências entre sg_equipamento_icao e ds_modelo
correspondencias2 =
estatisticasb_filtrado.set_index(['sg_equipamento_icao'])['ds_modelo'].to_dict()
```

```
python
```

```
# Mapeando os valores de sg_empresa_icao e ds_modelo para obter os modelos de aeronaves
correspondentes
```

```

estatisticasc_filtrado['ds_modelo'] =
estatisticasc_filtrado[['sg Equipamento Icao']].apply(lambda row:
correspondencias2.get((row['sg Equipamento Icao'])), axis=1)

```

```
python
```

```

# Incluindo a coluna assentos ofertados na estatística combinada
# Criando um dicionário com as correspondências entre
nr_voo/sg_empresa_icao/sg Equipamento Icao e nr Assentos Ofertados
correspondencias3 = estatisticasb_filtrado.set_index(['nr_voo', 'sg_empresa_icao',
'sg Equipamento Icao'])['nr Assentos Ofertados'].to_dict()

```

```
python
```

```

# Mapeando os valores de nr_voo/sg_empresa_origem e ds_modelo para obter os modelos de
aeronaves correspondentes
estatisticasc_filtrado['nr Assentos Ofertados'] = estatisticasc_filtrado[['nr_voo',
'sg_empresa_icao', 'sg Equipamento Icao']].apply(lambda row:
correspondencias3.get((row['nr_voo'], row['sg_empresa_icao'], row['sg Equipamento Icao'])),
axis=1)

```

```
python
```

```
estatisticasc_filtrado.head()
```

```
python
```

```
estatisticasc_filtrado.shape
```

```
python
```

```

# Salvando o arquivo "estatisticas2c_filtered.csv":
estatisticasc_filtrado.to_csv('./Dados/DemandaeOferta/estatisticas2c_filtered.csv',
index=False)

```

5. Inclusão da coluna de Tarifas nas Estatísticas Combinadas 2022

Salvos em ./Dados/DemandaeOferta/estatisticas2c_filtered.csv

Descrição: Para incluir a variável Tarifas nos arquivos de estatísticas combinadas, foi necessário criar uma variável de semestre, pois a variável mês do arquivo tarifas não corresponde à variável mês do arquivo estatísticas combinadas. Isso ocorre, porque o mês de compra da passagem não corresponde necessariamente ao mês do voo.

Dessa forma, a criação da variável semestre foi uma tentativa de aproximação das variáveis para correlacionar ambos os bancos de dados e obter a variável tarifa para cada voo no arquivo estatísticas combinadas.

```
python
```

```

# Leitura do arquivo:
estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas2c_filtered.csv')

```

```
python
#Criando a coluna SEMESTRE
estatisticas['SEMESTRE'] = pd.cut(estatisticas['nr_mes_partida_real'], bins=[0, 6, 12],
labels=[1, 2])
```

```
python
#Incluindo a coluna das empresas que comercializam as passagens (sg_comercializa_icao)
empresas = {
    "PAM": "PTB",
    "PTB": "PTB",
    "ACN": "AZU",
    "AZU": "AZU",
    "ABJ": "ABJ"
}
estatisticas['sg_comercializa_icao'] = estatisticas['sg_empresa_icao'].map(empresas)
```

```
python
estatisticas.head()
```

Adequação do arquivo tarifas2.csv

Descrição: Após a criação da variável semestre, foi necessária a criação da variável tarifa ponderada, qua consiste na média ponderada das tarifas praticadas para o mesmo grupo ('ANO', 'SEMESTRE', 'EMPRESA', 'ORIGEM', 'DESTINO'), ponderadas pela quantidade de assentos.

Após a criação da variável tarifa ponderada, foram calculados os desvios padrões das tarifas ponderadas.

```
python
# Leitura do arquivo:
tarifas = pd.read_csv('./Dados/Tarifas/tarifas2.csv')
```

```
python
#Criando a coluna SEMESTRE
tarifas['SEMESTRE'] = pd.cut(tarifas['MES'], bins=[0, 6, 12], labels=[1, 2])
```

```
python
#Criando a coluna média ponderada das Tarifas
tarifas['produto'] = tarifas['TARIFA'] * tarifas['ASSENTOS']
soma_ponderada = tarifas.groupby(['ANO', 'SEMESTRE', 'EMPRESA', 'ORIGEM',
'DESTINO'])['produto'].sum()
soma_assentos = tarifas.groupby(['ANO', 'SEMESTRE', 'EMPRESA', 'ORIGEM',
'DESTINO'])['ASSENTOS'].sum()
TARIFA_PONDERADA = soma_ponderada / soma_assentos
```

```
python
# Adicionando a variável TARIFA_PONDERADA ao dataframe:
```



```
tarifas = tarifas.merge(TARIFA_PONDERADA.rename('TARIFA_PONDERADA'), left_on=['ANO',
'SEMESTRE', 'EMPRESA', 'ORIGEM', 'DESTINO'], right_index=True)
```

```
python
```

```
# Excluindo a variável criada provisoriamente para o cálculo da média ponderada da tarifa:
tarifas = tarifas.drop(columns='produto')
```

```
python
```

```
tarifas.head()
```

```
python
```

```
# Incluindo a coluna tarifas na estatística combinada
# Criando um dicionário com as correspondências entre MES/EMPRESA/ORIGEM/DESTINO e TARIFA
correspondencias = tarifas.set_index(['SEMESTRE', 'EMPRESA', 'ORIGEM',
'DESTINO'])['TARIFA_PONDERADA'].to_dict()
```

```
python
```

```
# Mapeando os valores de nr_mes_partida_real/sg_empresa_icao/sg_icao_origem/sg_icao_destino
e tarifa para obter os modelos de aeronaves correspondentes
estatisticas['tarifa_ponderada'] = estatisticas[['SEMESTRE', 'sg_comercializa_icao',
'sg_icao_origem', 'sg_icao_destino']].apply(lambda row:
correspondencias.get((row['SEMESTRE'], row['sg_comercializa_icao'], row['sg_icao_origem'],
row['sg_icao_destino'])), axis=1)
```

```
python
```

```
estatisticas.head()
```

```
python
```

```
#Calculando o Desvio Padrão
tarifas['desvio'] = tarifas['ASSENTOS'] * (tarifas['TARIFA'] -
tarifas['TARIFA_PONDERADA'])**2
soma_desvio = tarifas.groupby(['ANO', 'SEMESTRE', 'EMPRESA', 'ORIGEM',
'DESTINO'])['desvio'].sum()
soma_assentos = tarifas.groupby(['ANO', 'SEMESTRE', 'EMPRESA', 'ORIGEM',
'DESTINO'])['ASSENTOS'].sum()
desvio_padrao = np.sqrt(soma_desvio / (soma_assentos - 1))
```

```
python
```

```
# Agora, vamos adicionar essa série ao dataframe original
tarifas = tarifas.merge(desvio_padrao.rename('DESVIO_PADRAO'), left_on=['ANO', 'SEMESTRE',
'EMPRESA', 'ORIGEM', 'DESTINO'], right_index=True)
```

```
python
```

```
tarifas = tarifas.drop(columns='desvio')
```

```
python
```

```
tarifas.head()
```

```
python
```

```
# Incluindo a coluna DESVIO_PADRAO na estatística combinada
# Criando um dicionário com as correspondências entre
SEMESTRE/EMPRESA/ORIGEM/DESTINO/TARIFA_PONDERADA e DESVIO_PADRAO
correspondencias = tarifas.set_index(['SEMESTRE', 'EMPRESA', 'ORIGEM', 'DESTINO',
'TARIFA_PONDERADA'])['DESVIO_PADRAO'].to_dict()
```

```
python
```

```
# Mapeando os valores de SEMESTRE/sg_empresa_icao/sg_icao_origem/sg_icao_destino/tarifa
para obter o desvio_padrao correspondente
estatisticas['desvio_padrao'] = estatisticas[['SEMESTRE', 'sg_comercializa_icao',
'sg_icao_origem', 'sg_icao_destino', 'tarifa_ponderada']].apply(lambda row:
correspondencias.get((row['SEMESTRE'], row['sg_comercializa_icao'], row['sg_icao_origem'],
row['sg_icao_destino'], row['tarifa_ponderada'])), axis=1)
```

```
python
```

```
estatisticas.head()
```

```
python
```

```
# Salvando o arquivo "estatisticas2c_filtered.csv" com as novas variáveis:
estatisticas.to_csv('./Dados/DemandaeOferta/estatisticas2c_filtered.csv', index=False)
```

Passo 4 - Banco de Dados Final

Descrição: concatenação final dos bancos de dados tratados neste notebook em um último, que será analisado.

Atenção: os arquivos baixados são .CSV, mas no formato Microsoft Excel, *comma separated values*, portanto, não tem o encoding utf-8. Além disso, o Excel em português salva o CSV separado com ponto e vírgula (;).

Concatenando os arquivos estatisticas2c_filtered | estatisticas1c_filtered | estatisticas0c_filtered | estatisticas9c_filtered

Salvo em ./Dados/DemandaeOferta/estatisticasc_filtered.csv

```
python
```

```
# Leitura dos arquivos:
estatisticas2 = pd.read_csv("./Dados/DemandaeOferta/estatisticas2c_filtered.csv")
estatisticas1 = pd.read_csv("./Dados/DemandaeOferta/estatisticas1c_filtered.csv")
estatisticas0 = pd.read_csv("./Dados/DemandaeOferta/estatisticas0c_filtered.csv")
estatisticas9 = pd.read_csv("./Dados/DemandaeOferta/estatisticas9c_filtered.csv")
```

```
python
```

```
# Eliminação dos dados em branco (NA):
estatisticas2.dropna(subset=["tarifa_ponderada","km_distancia","sg Equipamento_icao"])
estatisticas1.dropna(subset=["tarifa_ponderada","km_distancia","sg Equipamento_icao"])
estatisticas0.dropna(subset=["tarifa_ponderada","km_distancia","sg Equipamento_icao"])
estatisticas9.dropna(subset=["tarifa_ponderada","km_distancia","sg Equipamento_icao"])
```

```
python
```

```
# Incluindo a variável "ano" em cada arquivo:
estatisticas9['ano'] = 2019
```

```
python
```

```
estatisticas0['ano'] = 2020
```

```
python
```

```
estatisticas1['ano'] = 2021
```

```
python
```

```
estatisticas2['ano'] = 2022
```

```
python
```

```
# Concatenando os dataframes em um único arquivo:
estatisticas = pd.concat([estatisticas9, estatisticas0, estatisticas1, estatisticas2])
```

```
python
```

```
# Reordenando a variável "ano":
cols = list(estatisticas.columns)
cols = [cols[-1]] + cols[:-1]
estatisticas = estatisticas[cols]
```

```
python
```

```
# Eliminando as linhas com dados vazios:
estatisticas =
estatisticas.dropna(subset=["tarifa_ponderada","km_distancia","sg Equipamento_icao"])
estatisticas.shape
```

```
python
```

```
estatisticas.to_csv('./Dados/DemandaeOferta/estatisticasc_filtered.csv', index=False)
```

```
python
```

```
estatisticas2.to_csv('./Dados/DemandaeOferta/estatisticas2c_filtered.csv', index=False)
```

```
python
```

```
estatisticas1.to_csv('./Dados/DemandaeOferta/estatisticasl1c_filtered.csv', index=False)
```

```
python
estatisticas0.to_csv('./Dados/DemandaeOferta/estatisticas0c_filtered.csv', index=False)
```

```
python
estatisticas9.to_csv('./Dados/DemandaeOferta/estatisticas9c_filtered.csv', index=False)
```

Inclusão de Aeroportos no Banco de Dados ✈

Descrição: esse tópico do notebook evidencia alguns dos aeroportos que não estavam contemplados nos dados extraídos da ANAC e que, por tanto, tiveram que ser incluídos para realizar a correlação correta entre o arquivo de distancias e estatísticas.

Foi perceptível a falta desses aeroportos, porque - no momento de correlacionar os arquivos - o código apontou erro e, a partir disso, as variáveis aeroportos foram correlacionadas para identificar os aeroportos que não estavam contemplados no arquivo distancias.

Após identificados os aeroportos, foram identificados a UF e município do aeroporto, o código OACI e a latitude e longitude. Com isso, esses dados foram inclusos no arquivo aeroportos2, viabilizando a inclusão do aeroporto no arquivo distancias2.

Acrescentando OACI == 'SBRG' em aeroportos2.csv

```
python
# Leitura dos arquivos:
aeroportos = pd.read_csv('./Dados/Aerodromos/aeroportos2.csv')
distancias = pd.read_csv('./Dados/Aerodromos/distancias2.csv')
```

```
python
# Criando um dataframe provisório, com os dados do novo aeroporto:
novo_aeroporto = []
novo_aeroporto.append(['SBRG', "Rio Grande", "RS", -32.01667, -52.08333])
```

```
python
# Foi necessário retirar uma coluna excedente do arquivo 'aeroportos2'
aeroportos = aeroportos.iloc[:,0:5]
aeroportos.head()
```

```
python
# Criando um dataframe com as mesmas colunas de aeroportos2 e os dados do novo aeroporto
novo_aeroporto = pd.DataFrame(novo_aeroporto, columns=aeroportos.columns)
```

```
python
# Concatenando o dataframe do novo aeroporto ao de aeroportos2
aeroportos = pd.concat([aeroportos, novo_aeroporto], axis = 0, ignore_index = True)
```

```
python
# Pesquisando o novo aeroporto no dataframe aeroportos2
aeroportos[aeroportos.OACI == "SBRG"].drop_duplicates()
```

```
python
# Salvando o arquivo com os novos dados:
aeroportos.to_csv('./Dados/Aerodromos/aeroportos2.csv', index=False)
```

✓ Agora basta incluir o novo aeroporto no arquivo `distancias2`, repetindo as Etapas III e IV do Passo 1.

Acrescentando OACI == 'SBGS' em aeroportos2.csv

```
python
# Leitura dos arquivos:
aeroportos = pd.read_csv('./Dados/Aerodromos/aeroportos2.csv')
distancias = pd.read_csv('./Dados/Aerodromos/distancias2.csv')
```

```
python
# Criando um dataframe provisório, com os dados do novo aeroporto:
novo_aeroporto = []
novo_aeroporto.append(['SBGS', "Ponta Grossa", "PR", -25.184444, -50.143889])
```

```
python
# Foi necessário retirar uma coluna excedente do arquivo 'aeroportos2'
aeroportos = aeroportos.iloc[:,0:5]
aeroportos.head()
```

```
python
# Criando um dataframe com as mesmas colunas de aeroportos2 e os dados do novo aeroporto
novo_aeroporto = pd.DataFrame(novo_aeroporto, columns=aeroportos.columns)
```

```
python
# Concatenando o dataframe do novo aeroporto ao de aeroportos2
aeroportos = pd.concat([aeroportos, novo_aeroporto], axis = 0, ignore_index = True)
```

```
python
# Pesquisando o novo aeroporto no dataframe aeroportos2
aeroportos[aeroportos.OACI == "SBGS"].drop_duplicates()
```

```
python
```

```
# Salvando o arquivo com os novos dados:
aeroportos.to_csv('./Dados/Aerodromos/aeroportos2.csv', index=False)
```

✓ Agora basta incluir o novo aeroporto no arquivo `distancias2`, repetindo as Etapas III e IV do Passo 1.

Inclusão do Valor do Combustível e Consumo de Referência das Aeronaves no Banco de Dados ✈

Descrição: Este tópico do notebook evidencia a inclusão dos valores dos preços de combustíveis, que foram extraídos do site da Petrobras, assim como os valores de consumo de referência fornecidos pelos fabricantes das aeronaves. A necessidade dessa inclusão surgiu durante a análise dos dados, quando se percebeu que a ausência desses valores estava impactando a precisão das correlações e análises.

Inclusão do valor médio do Querosene de Aviação do mês em que o voo foi realizado

```
python
```

```
# Ler o arquivo CSV com os dados
estatisticas = pd.read_csv('./Dados/Estatisticas/estatisticasc_filtered.csv')

# Ler o arquivo CSV com os valores de QAV e preparar os dados
QAV = pd.read_csv('./Dados/Estatisticas/QAV_Valor.csv', encoding='latin-1',
usecols=['Valor', 'Mês', 'Ano'], delimiter=';')
QAV['Valor'] = QAV['Valor'].str.replace(',', '.').astype(float) / 1000
QAV['Mês'] = QAV['Mês'].astype(float)

# Mesclar os dataframes
estatisticas = estatisticas.merge(QAV, left_on=['ano', 'nr_mes_chegada_real'],
right_on=['Ano', 'Mês'], how='left')

# Renomear a coluna 'Valor' para 'valor_combustivel'
estatisticas = estatisticas.rename(columns={'Valor': 'valor_combustivel'})

# Preencher valores NaN na coluna 'Valor' com 0
estatisticas['valor_combustivel'] = estatisticas['valor_combustivel'].fillna(0)

# Remover as colunas 'Mês' e 'Ano' do dataframe
estatisticas = estatisticas.drop(['Mês', 'Ano'], axis=1)
```

Inclusão do Consumo de Referência de cada aeronave disponibilizado pelas suas fabricantes

```
python
```

```
# Ler o arquivo CSV e preparar os dados de consumo médio
Consumo_Médio = pd.read_csv('./Dados/Estatisticas/Consumo_Aeronaves.csv',
encoding='latin-1', usecols=['Modelo', 'Consumo em kg', 'Consumo em Litros'],
delimiter=';')
Consumo_Médio['Modelo'] = Consumo_Médio['Modelo'].str.strip()
```

```

Consumo_Médio['Consumo em Litros'] = Consumo_Médio['Consumo em Litros'].str.replace(',',
'.').astype(float)

# Remover espaços em branco na coluna 'ds_modelo' do dataframe 'estatisticas'
estatisticas['ds_modelo'] = estatisticas['ds_modelo'].str.strip()

# Mesclar os dataframes para adicionar a coluna 'Consumo em Litros' à esquerda
estatisticas = estatisticas.merge(Consumo_Médio, left_on='ds_modelo', right_on='Modelo',
how='left')

# Preencher valores NaN na coluna 'Consumo em Litros' com 0
estatisticas['Consumo em Litros'] = estatisticas['Consumo em Litros'].fillna(0)

# Remover a coluna 'Modelo' do dataframe após a mesclagem
estatisticas = estatisticas.drop(['Modelo'], axis=1)

```

Inclusão do Consumo Total de Combustível de Referência

Descrição: Este tópico do notebook evidencia a inclusão dos valores de consumo de referência totais de acordo com o tempo de voo e os dados de consumo fornecidos pelos fabricantes das aeronaves.

Cálculo da Fração de Horas que a duração do voo teve

```

python

# Função para calcular a diferença em minutos entre duas datas e horários
def calcular_diferenca(row):
    partida_completa = f"{row['dt_partida_real']} {row['hr_partida_real']}"
    chegada_completa = f"{row['dt_chegada_real']} {row['hr_chegada_real']}"

    try:
        partida = datetime.strptime(partida_completa, "%Y-%m-%d %H:%M")
        chegada = datetime.strptime(chegada_completa, "%Y-%m-%d %H:%M")
    except ValueError:
        # Tratar formato de hora inconsistente
        partida = datetime.strptime(partida_completa, "%Y-%m-%d %H:%M:%S")
        chegada = datetime.strptime(chegada_completa, "%Y-%m-%d %H:%M:%S")

    diferenca = (chegada - partida).total_seconds() / 60
    return diferenca

# Aplicar a função calcular_diferenca a todas as linhas do DataFrame
estatisticas['diferenca_minutos'] = estatisticas.apply(calcular_diferenca, axis=1)

# Adicionar a nova coluna representando a fração em relação a uma hora
estatisticas['fracao_em_horas'] = estatisticas['diferenca_minutos'] / 60

```

Cálculo do Consumo em Litros Total de acordo com os valores de referência

```

python

# Cálculo do Consumo em Litros Total
estatisticas['Consumo Total em Litros'] = estatisticas['fracao_em_horas'] *
estatisticas['Consumo em Litros']

```

Inclusão do Consumo Total de Combustível Real ✈

Descrição: Este tópico do notebook evidencia a inclusão dos valores de consumos reais disponibilizados pela ANAC.

Leitura e Concatenar os arquivos de cada ano

```
python
# Leitura dos arquivos:
estatisticas2 = pd.read_csv("../Dados/Estatisticas/estatisticas2.csv")
estatisticas1 = pd.read_csv("../Dados/Estatisticas/estatisticas1.csv")
estatisticas0 = pd.read_csv("../Dados/Estatisticas/estatisticas0.csv")
estatisticas9 = pd.read_csv("../Dados/Estatisticas/estatisticas9.csv")
```

```
python
# Eliminação dos dados em branco (NA):
estatisticas2.dropna(subset=["sg Equipamento_icao"])
estatisticas1.dropna(subset=["sg Equipamento_icao"])
estatisticas0.dropna(subset=["sg Equipamento_icao"])
estatisticas9.dropna(subset=["sg Equipamento_icao"])
```

```
python
# Incluindo a variável "ano" em cada arquivo:
estatisticas9['ano'] = 2019
```

```
python
estatisticas0['ano'] = 2020
```

```
python
estatisticas1['ano'] = 2021
```

```
python
estatisticas2['ano'] = 2022
```

```
python
# Concatenando os dataframes em um único arquivo:
estatisticasbasicas = pd.concat([estatisticas9, estatisticas0, estatisticas1,
estatisticas2])
```

```
python
# Reordenando a variável "ano":
cols = list(estatisticasbasicas.columns)
cols = [cols[-1]] + cols[:-1]
```



```
estatisticasbasicas = estatisticasbasicas[cols]
```

```
python
```

```
# Eliminando as linhas com dados vazios:
estatisticasbasicas = estatisticasbasicas.dropna(subset=["sg_equipamento_icao"])
estatisticasbasicas.shape
```

```
python
```

```
# Filtrando o dataframe estatísticas básicas por empresas desejadas:
empresas_desejadas = ["ACN", "PAM", "ASO", "AZU", "PTB", "ABJ"]
estatisticasbasicas =
estatisticasbasicas[estatisticasbasicas['sg_empresa_icao'].isin(empresas_desejadas)]
```

Mesclar os Dataframes para obter o combustível real

```
python
```

```
# Leitura do Arquivo que contém o consumo de combustível Real
estatisticasbasicas.to_csv('./Dados/Estatisticas/estatisticasbasicas.csv', index=False)

#Mesclar os Dataframes
correspondencias = estatisticasbasicas.groupby(['nr_voo', 'sg_icao_destino',
'sg_icao_origem'])['lt_combustivel'].mean().to_dict()
estatisticas['lt_combustivel'] = estatisticas.apply(lambda row:
correspondencias.get((row['nr_voo'], row['sg_icao_destino'], row['sg_icao_origem']),
axis=1))

# Salvando o arquivo "estatisticas.csv":
estatisticas.to_csv('./Dados/DemandaeOferta/estatisticas.csv',
index=False)
```

Passo 5 - Análise Gráfica dos Dados (2022 / 2021 / 2020 / 2019)

Descrição: Nessa etapa foi realizada a análise gráfica do banco de dados trabalhado ao longo da pesquisa.

```
python
```

```
# Leitura dos bancos de dados por ano e do concatenado com todos os anos:
estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas.csv')

# Definir valor zero quando nr_passag_pagos for igual a zero, caso contrário, calcular normalmente
estatisticas['CASK_Fuel'] = np.where(estatisticas['nr_assentos_ofertados'] == 0, 0,
estatisticas['valor_combustivel'] * estatisticas['lt_combustivel'] /
(estatisticas['nr_assentos_ofertados'] * estatisticas['km_distancia']))
# Definir valor zero quando diferenca_minutos for igual a zero, caso contrário, calcular normalmente
estatisticas['valor_gasto_por_hora'] = np.where(estatisticas['diferenca_minutos'] == 0, 0,
estatisticas['valor_combustivel'] * estatisticas['lt_combustivel'] /
(estatisticas['diferenca_minutos'] / 60))
```

```
python
```

```
pd.set_option('display.max_columns', None)
estatisticas.head()
```

Análises - Aeronaves: TODAS

Descrição: análise dos dados de todas as aeronaves, observando-se a perspectiva geral da amostra.

Análise de Boxplot: Distância (km_distancia) vs Ano:

```
python
```

```
# Modelos de aeronave
modelos_aeronave = ['AEROSPATIALE/ALENIA ATR 42-300 / 320',
                    'AEROSPATIALE/ALENIA ATR 42-500',
                    'AEROSPATIALE/ALENIA ATR 72 201/202',
                    'AEROSPATIALE/ALENIA ATR 72-500/72-212A (500)',
                    'AEROSPATIALE/ALENIA ATR 72-600/72-212A (600)',
                    'CESSNA 208 CARAVAN']

# Loop para calcular e remover os outliers separadamente para cada modelo de aeronave
for modelo in modelos_aeronave:
    # Filtrar os dados apenas para o modelo de aeronave atual
    dados_modelo = estatisticas[estatisticas['ds_modelo'] == modelo]

    # Calculando Q1 e Q3 para o modelo de aeronave atual
    Q1 = dados_modelo['km_distancia'].quantile(0.25)
    Q3 = dados_modelo['km_distancia'].quantile(0.75)

    # Faixa interquartil para o modelo de aeronave atual
    IQR = Q3 - Q1

    # Encontrando os outliers para o modelo de aeronave atual
    outliers = dados_modelo[(dados_modelo['km_distancia'] < (Q1 - 1.5 * IQR)) |
                             (dados_modelo['km_distancia'] > (Q3 + 1.5 * IQR))]

    # Remover os outliers do DataFrame estatisticas para o modelo de aeronave atual
    estatisticas = estatisticas.drop(outliers.index)

    # Plotar o boxplot e stripplot sem os outliers para o modelo de aeronave atual
    sns.boxplot(x='ano', y='km_distancia', data=estatisticas[estatisticas['ds_modelo'] ==
                                                             modelo])
    sns.stripplot(x='ano', y='km_distancia', data=estatisticas[estatisticas['ds_modelo'] ==
                                                                modelo],
                  jitter=True,
                  marker='o',
                  alpha=0.8,
                  color="black")
    plt.title(f'Boxplot of Distancia (km) by Ano - {modelo}')
    plt.xlabel('Ano')
    plt.ylabel('Distancia (km)')
    plt.xticks(rotation=90)
    plt.show()
```

Análise de Combustível: Consumo de Combustível (lt_combustivel) vs Distância (km_distancia):

```
python
# Configurar os modelos de aeronave
modelos = ['AEROSPATIALE/ALENIA ATR 42-300 / 320',
           'AEROSPATIALE/ALENIA ATR 42-500',
           'AEROSPATIALE/ALENIA ATR 72 201/202',
           'AEROSPATIALE/ALENIA ATR 72-500/72-212A (500)',
           'AEROSPATIALE/ALENIA ATR 72-600/72-212A (600)',
           'CESSNA 208 CARAVAN']

# Configurar as cores e estilos das linhas
cores = ['blue', 'red', 'green', 'orange', 'purple', 'brown'] # Adicione mais cores se necessário

# Iterar sobre os modelos de aeronave
for i, modelo in enumerate(modelos):
    # Filtrar os dados para o modelo de aeronave atual e remover valores ausentes
    dados_modelo = estatisticas[estatisticas['ds_modelo'] == modelo].copy()
    dados_modelo.dropna(subset=['km_distancia', 'lt_combustivel'], inplace=True)

    # Extrair os dados para a regressão linear
    x = dados_modelo['km_distancia']
    y = dados_modelo['lt_combustivel']

    # Calculando a regressão linear
    slope, intercept, r_value, p_value, std_err = stats.linregress(x, y)

    # Criar o gráfico
    plt.figure(figsize=(10, 6))
    sns.regplot(x=x, y=y, color=cores[i], label=f'{modelo}: y = {slope:.2f}x + {intercept:.2f}, r = {r_value:.2f}')

    plt.xlabel('Distância (km)')
    plt.ylabel('Consumo de Combustível (Litros)')
    plt.title('Consumo de Combustível por Distância (Real) - ' + modelo)
    plt.grid(True)
    plt.legend()
    plt.show()
```

Análise de CASK e Valor do Combustível: CASK (Cask_Fuel) vs Trimestre | Valor do Combustível (valor_combustivel) vs Trimestre:

```
python
# Lista das aeronaves específicas
aeronaves = ['AEROSPATIALE/ALENIA ATR 42-300 / 320',
             'AEROSPATIALE/ALENIA ATR 42-500',
             'AEROSPATIALE/ALENIA ATR 72 201/202',
             'AEROSPATIALE/ALENIA ATR 72-500/72-212A (500)',
             'AEROSPATIALE/ALENIA ATR 72-600/72-212A (600)',
             'CESSNA 208 CARAVAN']

# Configurar as cores e estilos das linhas
cores = ['blue', 'red', 'green', 'orange', 'purple', 'brown'] # Adicione mais cores se necessário

# Filtrar o DataFrame apenas para as aeronaves e anos desejados
```

```

filtro = estatisticas['ds_modelo'].isin(aeronaves) & estatisticas['ano'].isin([2019, 2020,
2021, 2022])
dados_filtrados = estatisticas.loc[filtro]

# Extrair o trimestre da coluna 'nr_mes_partida_real'
dados_filtrados['trimestre'] = ((dados_filtrados['nr_mes_partida_real'] - 1) // 3 +
1).astype(int)

# Calcular a média do CASK_Fuel para cada aeronave e trimestre
media_cask_fuel = dados_filtrados.groupby(['ds_modelo', 'ano',
'trimestre'])['CASK_Fuel'].mean().reset_index()

# Calcular a média do valor de combustível para cada aeronave e trimestre
media_valor_combustivel = dados_filtrados.groupby(['ds_modelo', 'ano',
'trimestre'])['valor_combustivel'].mean().reset_index()

# Criar um gráfico de linhas separado para cada aeronave
for i, aeronave in enumerate(aeronaves):
    dados_aeronave_cask = media_cask_fuel[media_cask_fuel['ds_modelo'] == aeronave]
    dados_aeronave_valor_combustivel =
media_valor_combustivel[media_valor_combustivel['ds_modelo'] == aeronave]

    plt.figure(figsize=(10, 6))

    # Plotar os dados no eixo esquerdo (eixo Y padrão)
    ax1 = plt.gca()
    trimestres = dados_aeronave_cask['ano'].astype(str) + " / " +
dados_aeronave_cask['trimestre'].astype(str) + "° Tri"
    trimestres_pos = range(len(trimestres))

    ax1.plot(trimestres_pos, dados_aeronave_cask['CASK_Fuel'], marker='o', color=cores[i])
    ax1.set_xticks(trimestres_pos)
    ax1.set_xticklabels(trimestres, rotation=45)
    ax1.set_xlabel('Trimestre')
    ax1.set_ylabel('CASK Fuel')
    ax1.set_title(f'CASK Fuel - {aeronave}')
    ax1.set_ylim(0, 6) # Define o intervalo do eixo y de 0 a 6

    # Adicionar rótulos de dados acima de cada ponto, ajustando a posição vertical
    for j, valor in enumerate(dados_aeronave_cask['CASK_Fuel']):
        ax1.text(trimestres_pos[j], valor + 0.05 + j * 0.01, round(valor, 2), ha='center')

    # Plotar a média do valor de combustível para cada trimestre
    ax2 = ax1.twinx()
    trimestres_pos_combustivel = range(len(dados_aeronave_valor_combustivel))
    ax2.plot(trimestres_pos_combustivel,
            dados_aeronave_valor_combustivel['valor_combustivel'], marker='s',
linestyle='--', color='black')
    ax2.set_ylabel('Valor do Combustível (em Reais)', labelpad=20)
    ax2.set_ylim(0, 6) # Define o intervalo do eixo y de 0 a 6

    # Adicionar rótulos de dados acima de cada ponto no eixo ax2, ajustando a posição
vertical
    for j, valor in enumerate(dados_aeronave_valor_combustivel['valor_combustivel']):
        ax1.text(trimestres_pos_combustivel[j], valor + 0.05 + j * 0.01, round(valor, 2),
ha='center')

    # Adicionar legenda para o eixo ax2
    legenda1 = ['Valor do Combustível (em Reais)']
    ax2.legend(legenda1, loc='upper left')

    # Adicionar legenda para o eixo ax1 abaixo do eixo ax2
    legenda2 = ['Cask Fuel']

```

```
ax1.legend(legenda2, loc='upper left', bbox_to_anchor=(0, 0.95))
```

```
plt.tight_layout()
plt.show()
```

Análise de COD Fuel: Consumo de Combustível real (lt_combustivel) x Valor do Combustível (valor_combustivel) vs Distância

```
python
```

```
# Função de regressão polinomial (grau 3)
def func_poly(x, a, b, c, d):
    return a * x**3 + b * x**2 + c * x + d

# Configurar os modelos de aeronave
modelos = ['AEROSPATIALE/ALENIA ATR 42-300 / 320',
           'AEROSPATIALE/ALENIA ATR 42-500',
           'AEROSPATIALE/ALENIA ATR 72 201/202',
           'AEROSPATIALE/ALENIA ATR 72-500/72-212A (500)',
           'AEROSPATIALE/ALENIA ATR 72-600/72-212A (600)',
           'CESSNA 208 CARAVAN']

# Configurar as cores e estilos das linhas
cores = ['blue', 'red', 'green', 'orange', 'purple', 'brown'] # Adicione mais cores se
necessário

# Iterar sobre os modelos de aeronave
for i, modelo in enumerate(modelos):
    # Filtrar os dados para o modelo de aeronave atual e remover valores ausentes
    dados_modelo = estatisticas[estatisticas['ds_modelo'] == modelo].copy()
    dados_modelo.dropna(subset=['km_distancia', 'lt_combustivel', 'valor_combustivel'],
inplace=True)

    # Calcular o DOC
    dados_modelo['DOC'] = dados_modelo['lt_combustivel'] *
dados_modelo['valor_combustivel']

    # Extrair os dados para a regressão
    x = dados_modelo['km_distancia']
    y = dados_modelo['DOC']

    # Ajustar a curva polinomial
    params_poly, _ = curve_fit(func_poly, x, y)

    # Criar o gráfico
    plt.figure(figsize=(10, 6))

    # Plotar a linha de regressão polinomial
    x_vals = np.linspace(min(x), max(x), 100)
    y_vals_poly = func_poly(x_vals, *params_poly)
    plt.plot(x_vals, y_vals_poly, color=cores[i], linestyle='-', label=f'Regressão
Polinomial: y = {params_poly[0]:.2e} * x^3 + {params_poly[1]:.2e} * x^2 +
{params_poly[2]:.2e} * x + {params_poly[3]:.2e}')

    # Plotar os pontos de dados reais
    plt.scatter(x, y, color=cores[i], label='Dados Reais')

    plt.xlabel('Distância (km)')
    plt.ylabel('DOC (R$)')
    plt.title('DOC por Distância - ' + modelo)
    plt.grid(True)
```

```
plt.legend()
```

```
plt.show()
```