



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

GABRIEL MARINHO GODINHO

**Análise Exploratória das Tarifas Aéreas Praticadas no Brasil no Período de 2019 a 2022
e Perspectivas para o Serviço de Transporte por eVTOL**

BRASÍLIA - DF
2023
GABRIEL MARINHO GODINHO

**Análise Exploratória das Tarifas Aéreas Praticadas no Brasil no Período de 2019 a 2022
e Perspectivas para o Serviço de Transporte por 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.

BRASÍLIA – DF
2023
GABRIEL MARINHO GODINHO

**Análise Exploratória das Tarifas Aéreas Praticadas no Brasil no Período de 2019 a 2022
e Perspectivas para o Serviço de Transporte por eVTOL**

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

Gabriel Marinho Godinho

Professor Dr. Victor Rafael Rezende Celestino
Professor-Orientador

Profa. Dra., Silvia Araujo dos Reis,
Professor-Examinador

Profa. Dra., Olinda Maria Gomes
Lesses, Professor-Examinador

Brasília, 11 de dezembro de 2023.

RESUMO

O avanço das tecnologias para veículos elétricos (EVs) proporcionou o desenvolvimento acelerado do setor e o advento dos novos modelos de aeronaves de decolagem e pouso vertical elétrico (eVTOL). Os benefícios e potenciais dessas novas tecnologias prevê reaquecer o setor da Mobilidade Regional Aérea (RAM), tornando-o mais sustentável, acessível e lucrativo. O presente trabalho propõe-se a analisar e compreender as variáveis apresentadas pelo banco de dados da ANAC que influenciaram as tarifas dos voos tradicionais entre 2019 e 2022, e correlacionar as percepções obtidas com as análises desses dados aos serviços de transporte por eVTOL no modelo de Mobilidade Regional Aérea (RAM) brasileira. A pesquisa foi desenvolvida, com base na metodologia da Design Science Research (DSR), um paradigma derivado da Design Science (DS). Esta abordagem orienta o pesquisador a centrar sua pesquisa na resolução do problema apresentado no contexto estudado, através da identificação, projeto, avaliação e/ou criação de um ou mais artefatos que garantam a resolução do problema ou da melhoria desejada. Para o desenvolvimento do artefato proposto pela DSR, de maneira complementar, foi aplicada as etapas da metodologia da Análise Exploratória de Dados (AED), que prevê a coleta, organização, tratamento, análise e interpretação dos dados. O artefato consistiu no notebook do *Google Colaboratory*, utilizando linguagem de programação Python para organização, tratamento e análise dos dados. A partir dessa combinação, os resultados foram apresentados nos resultados da pesquisa. Por fim, foi possível constatar que as variáveis número de assentos ofertados nos voos, distância percorrida (km) e período do ano, devem influenciar as tarifas praticadas no futuro serviço de transporte por eVTOL, mas não serão as únicas. Para pesquisas futuras, sugere-se a investigação das demais variáveis que influenciarão o valor da tarifa praticada no serviço, e o aprimoramento da análise de correlação delas com a tarifa.

Palavras-chave: eVTOL, Análise Exploratória de Dados (AED), Mobilidade Aérea Regional (RAM), Mobilidade Aérea Urbana (UAM), Pesquisa Operacional (PO), *Design Science Research* (DSR).

Abstract

The advance of electric vehicle (EV) technologies has led to the accelerated development of the sector and the advent of new models of electric vertical take-off and landing (eVTOL) aircraft. The benefits and potential of these new technologies are expected to reignite the Regional Air Mobility (RAM) sector, making it more sustainable, accessible and profitable. This research aims to analyze and understand the variables presented by the ANAC database that influenced traditional flight fares between 2019 and 2022, and to correlate the perceptions obtained from the analysis of this data with eVTOL transport services in the Brazilian Regional Air Network (RAM) model. The research was carried out using the Design Science Research (DSR) methodology, a paradigm derived from Design Science (DS). This approach guides the researcher to focus their research on solving the problem presented in the context studied, through the identification, design, evaluation and/or creation of one or more artifacts that guarantee the resolution of the problem or the desired improvement. In order to develop the artifact proposed by the DSR, the steps of the Exploratory Data Analysis (EDA) methodology were applied in a complementary manner, which involves collecting, organizing, processing, analyzing and interpreting data. The artifact consisted of the Google Collaboratory notebook, using the Python programming language to organize, process and analyze the data. Based on this combination, the results of the exploratory data analysis were presented in the research results. Finally, it was found that the variables number of seats offered on flights, distance traveled (km) and time of year should influence the fares charged in the future eVTOL transport service, but they will not be the only ones. For future research, it is suggested to investigate the other variables that will influence the fare charged for the service and improving the analysis of their correlation with the fare.

Key-words: eVTOL, Exploratory Data Analysis (EDA), Regional Air Mobility (RAM), Urban Air Mobility (UAM), Operations Research (OR), Design Science Research (DSR).

LISTA DE ILUSTRAÇÕES

| | |
|---|----|
| Figura 1 - Subconjuntos da AAM..... | 14 |
| Figura 2 - Representação da AAM..... | 15 |
| Figura 3 - Comparação de diferentes conceitos de eVTOLs..... | 16 |
| Figura 4 - Modelo de aeronave eVTOL EVE..... | 17 |
| Figura 5 - Modelagem do cálculo ETA para o eVTOL..... | 20 |
| Figura 6 - Etapas da AED..... | 24 |
| Figura 7 - Etapas para a condução da DSR. | 27 |
| Figura 8 - Análise de Regressão Linear: Distância(km) vs Tarifa - 2019..... | 34 |
| Figura 9 - Análise de Regressão Linear: Distância(km) vs Tarifa - 2020..... | 35 |
| Figura 10 - Análise de Regressão Linear: Distância(km) vs Tarifa - 2021..... | 36 |
| Figura 11 - Análise de Regressão Linear: Distância(km) vs Tarifa - 2022..... | 37 |
| Figura 12 - Análise de Regressão Linear: Distância(km) vs Tarifa, por empresa - 2019..... | 38 |
| Figura 13 - Análise de Regressão Linear: Distância(km) vs Tarifa, por empresa - 2020..... | 39 |
| Figura 14 - Análise de Regressão Linear: Distância(km) vs Tarifa, por empresa - 2021..... | 40 |
| Figura 15 - Análise de Regressão Linear: Distância(km) vs Tarifa, por empresa - 2022..... | 40 |
| Figura 16 - Gráfico de Dispersão 3D: N° de Assentos Ofertados vs Distância(km) vs Tarifa (2019 a 2022)..... | 42 |
| Figura 17 - Lineplot: Tarifa Ponderada Média por Mês..... | 43 |
| Figura 18 - Análise de Regressão Linear: Distância(km) vs Tarifa - C208 - 2020..... | 44 |
| Figura 19 - Análise de Regressão Linear: Distância(km) vs Tarifa - C208 - 2021..... | 44 |
| Figura 20 - Análise de Regressão Linear: Distância(km) vs Tarifa - C208 - 2022..... | 45 |
| Figura 21 - Lineplot: Tarifa Ponderada Média por Mês - C208..... | 46 |

LISTA DE TABELAS

| | |
|-------------------------------------|----|
| Tabela 1 - Tipos de Artefatos | 22 |
|-------------------------------------|----|

LISTA DE ABREVIATURAS E SIGLAS

- AAM – *Advanced Air Mobility* / (Mobilidade Aérea Avançada)
- ConOps - *Concept of Operations* / (Conceito de Operações)
- DARP - *Dial-ARide Problem* / (Problema de Agendamento de Transporte Individual)
- DOF – *Desdegrees-of-freedom* / (Graus de Liberdade)
- DS – *Design Science*
- DSR – *Design Science Research*
- ETA – *Estimated Time of Arrival* / (Tempo Estimado de Chegada)
- EV – *Electric Vehicle* / (Veículo Elétrico)
- eVTOL – *Electric Vertical Take-Off and Landing* / (Decolagem e Pouso Vertical Elétrico)
- FAA – *Federal Aviation Administration* / (Administração Federal de Aviação)
- FAO – *Food and Agriculture Organization of the United Nations* / (Organização das Nações Unidas para Agricultura e Alimentação)
- IBGE - Instituto Brasileiro de Geografia e Estatística
- ITS – *Intelligent Transportation System* / (Sistemas de Transporte Inteligente)
- LOI - *Letter of Intent* / (Carta de Intenções)
- ML – *Machine Learning* / (Aprendizado de Máquina)
- NASA – *National Aeronautics and Space Administration* / (Administração Nacional da Aeronáutica e Espaço)
- ONU – Organização das Nações Unidas
- PO – Pesquisa Operacional
- RAM - *Regional Air Mobility* / (Mobilidade Aérea Regional)
- RAT – *Regional Air Transport* / (Transporte Aéreo Regional)
- RPK – Receita Passageiro por Quilômetro
- 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 | 9 |
| 1.3 Objetivo Geral | 10 |
| 1.4 Objetivos Específicos | 10 |
| 1.5 Justificativa | 10 |
| 2. REFERENCIAL TEÓRICO | 11 |
| 2.1 Definições no contexto | 12 |
| 2.1.1 Mobilidade Aérea Urbana (UAM) | 12 |
| 2.1.2 Mobilidade Aérea Regional (RAM) | 12 |
| 2.1.3 eVTOL (Electric Vertical Take-off and Landing) | 15 |
| 2.1.4 <i>Dynamic Pricing</i> - Precificação Dinâmica | 18 |
| 2.2 Modelagem do cálculo ETA para o eVTOL | 19 |
| 2.3 <i>Google Colaboratory</i> | 21 |
| 3. METODOLOGIA | 21 |
| 3.1 Tipologia e descrição geral dos métodos de pesquisa | 21 |
| 3.1.1 <i>Design Science Research</i> (DSR) | 21 |
| 3.1.2 Análise Exploratória de Dados (AED) | 23 |
| 3.1.2.1 Passo 1: Aeródromos e Distâncias entre Aeroportos | 24 |
| 3.1.2.2 Passo 2: Organização e Transformação dos Dados de Tarifas (2019 a 2022) | 25 |
| 3.1.2.3 Passo 3: Organização e Transformação dos Dados Estatísticos dos Voos (2019 a 2022) | 26 |
| 3.1.2.4 Passo 4 e 5: Transformação do Banco de Dados Final (2019 a 2022) e análise | 26 |
| 3.2 Caracterização e descrição dos instrumentos de pesquisa | 26 |
| 3.3 Caracterização da organização, setor e área | 30 |
| 3.3.1 Mobilidade Aérea Avançada (AAM) no Brasil | 30 |
| 3.3.2 Agência Nacional de Aviação Civil (ANAC) | 31 |
| 3.4 População e amostra da pesquisa | 31 |
| 3.5 Procedimentos de coleta de dados | 32 |
| 3.6 Análise de dados | 32 |
| 4. RESULTADOS E DISCUSSÕES | 33 |
| 4.1 Análise da Amostra do Mercado da Malha Regional Aérea (RAM e RAT) no Brasil | 33 |
| 4.1.1 Análise de Regressão Linear: Variáveis Distância e Tarifa | 33 |
| 4.1.2 Análise de Regressão Linear: Variáveis Distância e Tarifa por Empresa Aérea | 37 |

| | |
|--|-----------|
| 4.1.3 Gráfico de Dispersão 3D - Análise de Relação: Distância, Número de Assentos e Tarifas | 41 |
| 4.1.4 Lineplot: Tarifa Ponderada Média por Mês (2019 a 2022) | 42 |
| 4.2 Análise da amostra para a aeronave Cessna 208 Caravan | 43 |
| 4.2.1 Análise de Regressão Linear: Variáveis Distância e Tarifa..... | 43 |
| 4.2.2 Lineplot: Tarifa Ponderada Média por Mês (2019 a 2022) | 45 |
| 5. CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS..... | 47 |
| 5.1 Considerações Finais | 47 |
| 5.2 Objetivos Atendidos da Pesquisa | 48 |
| 5.3 Desafios e Limitações da Pesquisa | 49 |
| 5.4 Direcionamentos para Trabalhos Futuros | 50 |
| REFERÊNCIAS..... | 51 |
| APÊNDICE | 56 |
| Apêndice A – Notebook do <i>Google Colaboratory</i> : Análise Exploratória das Tarifas Aéreas praticadas no Brasil (2019 a 2022) | 57 |

1. INTRODUÇÃO

1.1 Contextualização

Segundo os primeiros resultados do censo IBGE 2022 (2023), a população brasileira chegou a 203,1 milhões em 2022, com um aumento de 6,5% em relação ao censo demográfico anterior, realizado em 2010. Em 2022, as concentrações urbanas abrigavam 124,1 milhões de pessoas, representando 61% da população. Esses dados indicam um adensamento populacional urbano significativo no Brasil. Além disso, quase metade dos municípios brasileiros (44,8%) tinham até 10 mil habitantes em 2022, mas apenas 12,8 milhões de pessoas, ou 6,3% da população do país, viviam em cidades desse porte.

Em vistas às tendências de adensamento populacional urbano, muitos problemas relacionados à infraestrutura das cidades surgem como desafios a serem enfrentados e solucionados. Dentre essas dificuldades, destacam-se a elevada incidência de gases do efeito estufa, referentes ao excesso de veículos movidos a combustíveis derivados do petróleo, e inflamação da infraestrutura de transporte das cidades.

Dessa forma, alternativas para solucionar os problemas relativos à infraestrutura de transportes e a poluição das cidades têm sido cada vez mais esperadas. Nesse sentido, o setor de Mobilidade Aérea Regional (RAM) e Urbana (UAM) entram em foco como opções cada vez mais viáveis para amenizar os desafios apresentados, por trazerem alternativas de otimização e desafogamento da infraestrutura de transporte urbano, viabilizando a criação de uma rede de transportes integrada e inteligente.

Ademais, as tecnologias do mercado de veículos elétricos avançaram consideravelmente nos últimos anos, viabilizando a criação das aeronaves do tipo eVTOLs (Decolagem e Pouso Vertical Elétrico) que estão sendo estudadas para implementação nos grandes centros urbanos, integradas às redes UAM e RAM (ambos pertencentes ao conceito de Mobilidade Aérea Avançada – AAM), e com previsões de operacionalização recentes. Essas aeronaves são consideradas alternativas sustentáveis por não emitirem gases de efeito estufa (elétricas), emitirem pouco ruído e, por não dependerem de combustíveis fósseis, conseguem operar com custos mais reduzidos, em relação aos helicópteros, por exemplo (EVE Air Mobility, 2022). Dessa forma, além de poderem tornar-se uma alternativa de locomoção acessível ao público dentro de alguns anos, possuem o potencial de desafogar o fluxo excessivo de veículos nos grandes centros urbanos em horários de picos.

1.2 Formulação do problema

Apesar das crescentes iniciativas e cooperações para o desenvolvimento e implementação da AAM no Brasil e do serviço de transporte de passageiros em eVTOL integrado a ela, muitos desafios, evidenciados no ConOps para o Rio de Janeiro, quanto a operacionalização do serviço ainda não estão evidentes, como regulamentação, aceitação do público, segurança do serviço, acessibilidade, dentre outros (EVE Air Mobility, 2022). Esse cenário se dá devido à recente emergência dessas tecnologias e, conseqüente, imaturidade do setor no mercado brasileiro e global, que – até o momento – não possui casos de serviços que foram operacionalizados.

Nesse sentido, um grande desafio para o futuro da implementação do serviço de transporte de passageiros por eVTOL é como operacionalizar esse serviço. Uma alternativa para entender esse problema é fragmentá-lo em diversas perspectivas, como - por exemplo - a perspectiva do modelo de tarifação. Ou seja, como calcular as tarifas para os serviços de transportes por eVTOLs? Para isso, vale entender como as tarifas são praticadas atualmente para os voos tradicionais e, com isso, entender como elas poderão ser praticadas para os voos com eVTOLs.

Dado esse cenário de incertezas e desafios com relação à operacionalização do serviço de transporte de passageiros em eVTOL, as questões norteadoras da pesquisa foram desenvolvidas: Como as tarifas dos voos tradicionais evoluíram entre 2019 e 2022, considerando as variáveis que possivelmente mais impactam seu valor? Como essas variáveis que influenciam as tarifas dos voos tradicionais podem ser correlacionadas ao contexto do transporte por eVTOL no modelo de Malha Regional Aérea (RAM) brasileira?

1.3 Objetivo Geral

O objetivo deste trabalho é analisar e compreender as variáveis apresentadas pelo banco de dados da ANAC que influenciaram as tarifas dos voos tradicionais entre 2019 e 2022, e correlacionar as percepções obtidas com as análises desses dados aos serviços de transporte por eVTOL no modelo de Malha Regional Aérea (RAM) brasileira.

1.4 Objetivos Específicos

Pretende-se alcançar os objetivos específicos:

1. Analisar a evolução das tarifas dos voos tradicionais ao longo dos anos de 2019 a 2022;
2. Identificar algumas das variáveis que têm impacto no valor das tarifas dos voos tradicionais;
3. Investigar a correlação entre a distância (km) percorrida nos voos tradicionais e o valor das tarifas;
4. Analisar a variação sazonal nas tarifas praticadas, com base no período dos voos tradicionais; e
5. Aplicar as percepções obtidas a partir da análise das tarifas dos voos tradicionais ao contexto do transporte por eVTOLs.

1.5 Justificativa

O avanço do setor de Mobilidade Aérea Regional e Urbana tem avançado substancialmente, paralelamente à proposta de operacionalização do serviço de transporte de passageiros em eVTOL. Dado o avanço do setor, a presente pesquisa visa trazer elementos que possibilizem potencializar a operacionalização das organizações que estão à frente das iniciativas e implementação dos serviços no Brasil.

Com a análise das tarifas dos voos tradicionais entre 2019 e 2022, a identificação das variáveis que estão correlacionadas à ela e entendendo como elas podem ser aplicadas ao contexto do transporte por eVTOLs, a pesquisa ganha potencial para gerar insumos que podem contribuir diretamente com a viabilização do serviço e, mais ainda, viabilizar às organizações atuantes no setor o desenvolvimento de um modelo de tarifação que pode vir a contribuir com a competitividade delas no mercado de mobilidade urbana.

Além das organizações, os usuários poderão ser beneficiados com os resultados dessa pesquisa, a partir do momento que ela gera uma perspectiva relacionada às variáveis que possivelmente virão a impactar os preços do serviço de eVTOL e gera *insights* que podem contribuir para a viabilização do serviço acessível ao usuário.

No âmbito acadêmico, o presente estudo está sendo realizado em um cenário emergente e com escassez de pesquisas do setor. Alinhado aos princípios do paradigma da *Design Science*, este trabalho visa contribuir com pesquisas futuras que desejem dar continuidade a essa ou pesquisar novas temáticas no mesmo grupo de problemas. Caso os objetivos deste trabalho sejam alcançados com sucesso, algumas lacunas existentes sobre o assunto serão preenchidas, abrindo caminho para novas pesquisas.

2. REFERENCIAL TEÓRICO

O referencial teórico desta pesquisa apresenta conceitos fundamentais para a imersão no contexto relacionado ao futuro serviço de transporte de passageiros por veículos elétricos voadores (eVTOL) e para fundamentar as posteriores análises sobre as perspectivas para esse serviço. Em um primeiro momento entende-se os conceitos do subconjunto da Mobilidade Aérea Avançada (AAM), Mobilidade Aérea Urbana (UAM) e Mobilidade Aérea Regional (RAM). A imersão sobre esses conceitos é fundamental para a compreensão da rede integrada de transporte que está sendo estudada de maneira complementar aos desenvolvimentos dos veículos eVTOLs (NASA, 2021).

Em seguida, aprofunda-se nos conceitos relacionados ao veículo elétrico voador, suas características e variedades de modelos desenvolvidos até o momento. Essas características são fundamentais para a realização de correlações entre as aeronaves estudadas na amostra de dados da pesquisa e os eVTOLs.

Outros conceitos estudados e fundamentais para a análise das perspectivas para o serviço de transporte por eVTOL, são: modelagem do cálculo ETA e precificação dinâmica. Ambos os conceitos refletem, como elucidado por Marcelo Ferreira (2023), possibilidades de composição da estrutura para o modelo de serviço de transporte. Dessa forma, a compreensão deles é essencial para embasar parte do desenvolvimento de perspectivas futuras para o serviço.

Por fim, entende-se o conceito do ambiente de desenvolvimento do artefato desta pesquisa, o *Google Colaboratory*.

2.1 Definições no contexto

2.1.1 Mobilidade Aérea Urbana (UAM)

O conceito de Mobilidade Aérea Urbana surgiu como uma alternativa para amenizar os impactos relacionadas ao adensamento populacional nas grandes cidades globais. Segundo relatório publicado pela Organização das Nações Unidas para Agricultura e Alimentação, FAO (2019), até 2050 cerca de dois terços (2/3) da população global viverão em áreas urbanas. Dentro dessa perspectiva, surge a preocupação com os impactos que o adensamento populacional pode causar nas grandes cidades, como o aumento da criminalidade, a sobrecarga dos sistemas de infraestrutura urbana e dos sistemas de transporte e locomoção, aumento da poluição e emissão de gases do efeito estufa, dentre outros. Dado o contexto apresentado, percebe-se que uma infraestrutura de transporte adequada, capaz de atender às pressões de crescimento populacional urbano, passa a ser um desafio.

A UAM (*Urban Air Mobility*) representa o conceito de expansão das redes de transporte para incluir voos curtos que transportam pessoas e mercadorias nas áreas metropolitanas. A UAM faz parte de uma mudança de paradigma mais ampla em direção à AAM (*Advanced Air Mobility*), em que novas tecnologias e modelos de negócios estão possibilitando aplicações transformadoras na aviação, permitindo que ela desempenhe um papel essencial no transporte regional e local. Nesse sentido, a AAM torna-se uma forma de aviação, segura, sustentável, econômica e acessível (Hill *et al.*, 2020).

Segundo a FAA (2023), a Mobilidade Aérea Urbana (UAM) refere-se à prestação de serviços de transporte aéreo altamente automatizados e cooperativos, tanto para passageiros como para carga, em áreas urbanas e seus arredores. A UAM é um subconjunto do conceito de Mobilidade Aérea Avançada (AAM), que está sendo desenvolvido em conjunto pela *Federal Aviation Administration* (FAA), a *National Aeronautics and Space Administration* (NASA) e o setor aeroespacial. Como parte da AAM, a UAM tem foco específico nas operações de transporte de pessoas e cargas em áreas metropolitanas e urbanas.

Uma das principais definições utilizadas para a AAM é a criada pela NASA e que será utilizada como base para a construção desta pesquisa:

“um sistema de transporte aéreo que transporta pessoas e cargas entre locais anteriormente não atendidos ou mal atendidos pela aviação - local, regional, intrarregional, urbano - usando novas aeronaves revolucionárias que só agora estão se tornando possíveis” (NASA, 2021, traduzido pelo autor).

2.1.2 Mobilidade Aérea Regional (RAM)

A mobilidade aérea regional (RAM) se refere ao transporte de passageiros em distâncias curtas utilizando aeronaves de pequeno porte. Essas operações enfrentam desafios devido ao baixo volume de passageiros, à competição com outros modos de transporte e aos altos custos operacionais. No entanto, avanços recentes na tecnologia de propulsão elétrica e autonomia estão proporcionando a promessa de mudanças significativas tanto em eficiência operacional quanto em sustentabilidade (Cedric; Alexia; Dimitri, 2022).

Kevin Antcliff *et al.* (2021) defende que a tecnologia RAM pode ser um complemento e um acelerador para a Mobilidade Aérea Avançada (AAM) e outras iniciativas que visam

transformar o espaço aéreo. Através de investimentos em tecnologias avançadas, como automação de aeronaves, modelos operacionais aprimorados e sistemas de propulsão mais eficientes, juntamente com a ampliação da utilização de energia renovável nos aeroportos, muitos dos quais já estão em andamento, a mobilidade aérea regional (RAM) tem o potencial de aumentar a segurança, a acessibilidade e a viabilidade econômica das viagens regionais.

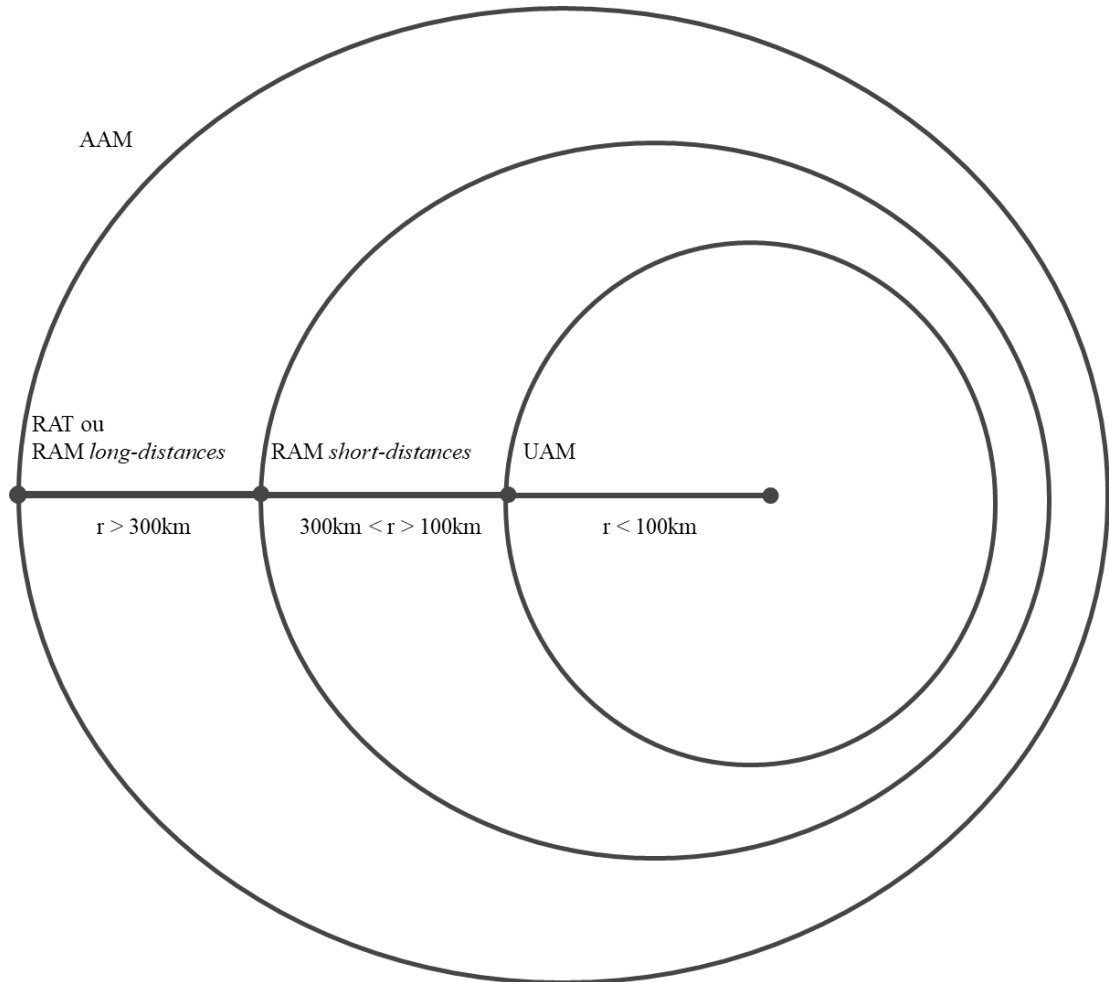
Entretanto, deve-se entender a principal diferença entre a UAM e a RAM. Segundo publicação realizada pela Lillium (2020), empresa aeroespacial, a principal diferença entre a mobilidade aérea urbana e a regional é a distância percorrida. A empresa citada, tem seu foco de atuação na RAM, com o alcance de voo de até 300 km (regional), ao contrário, a UAM percorre distâncias menores (<100km), focadas nas distâncias de cidades (urbanas).

Segundo Manfred Hader e Stephan Baur (2022, traduzido pelo autor), as diferenças dentre os subconjuntos de Mobilidade Aérea Avançada (AAM), são:

- Mobilidade Aérea Urbana (UAM): aeronaves com alcance inferior a 100km;
- Mobilidade Aérea Regional curtas-distâncias (RAM *short-distances*): aeronaves com alcance maior que 100km e menor que 300km;
- Mobilidade Aérea Regional longas-distâncias (RAM *long-distances*): aeronaves com alcance maior que 300km, mas com capacidade máxima de 19 passageiros; e
- Transporte Aéreo Regional (RAT): veículos de transporte aéreo com alcance maior que 300km e capacidade maior que 19 passageiros.

A Figura 1 a seguir ilustra as diferenças dentre os subconjuntos de AAM:

Figura 1 - Subconjuntos da AAM



Fonte: Elaboração própria (2023)

Jordan Cox *et al.* (2023) ressalta que o surgimento dos veículos de propulsão elétrica aumenta o potencial da exploração da RAM. Devido às características desses veículos, existe a possibilidade de obter-se o maior aproveitamento dos espaços nos aeroportos e de conectá-los com voos diretos a regiões mais curtas. Além disso, essas aeronaves podem reduzir os custos de viagens regionais (de curta distância), mitigando alguns dos desafios relacionados aos altos custos operacionais da RAM (Cedric; Alexia; Dimitri, 2022).

Figura 2 - Representação da AAM.



Fonte: NASA, 2021

2.1.3 eVTOL (Electric Vertical Take-off and Landing)

A fim de aprimorar o transporte de passageiros, a Mobilidade Aérea Urbana e Regional (UAM e RAM) buscam explorar conceitos de aeronaves elétricas de decolagem e aterrissagem vertical (e-VTOL). Essas aeronaves têm sido propostas como uma solução para viabilizar esse novo modo de transporte (Nathen *et al.* 2021).

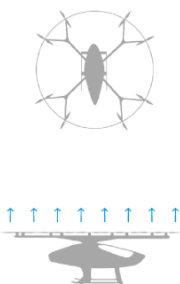
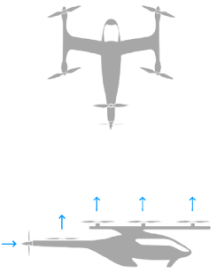

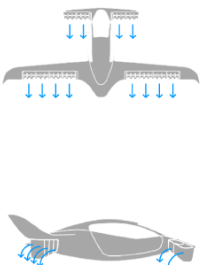
Ademais, os veículos elétricos de decolagem e pouso vertical são significativamente mais silenciosos, mais ecológicos e potencialmente mais baratos de operar do que os VTOLs (helicópteros) convencionais (Bris *et al.* 2022). Nesse sentido, operações em pequenos trajetos (regionais ou em cidades), demandam veículos com capacidade para atuar, atendendo as condições de espaço, custo e tempo. Dessa forma, os eVTOLs ganham espaço, atendendo a esses requisitos e a outros, como os supracitados.

Segundo o colégio de aeronáutica e tecnologia, *Spartan* (2021), a decolagem e pouso vertical elétrico (eVTOL) é uma inovação recente na indústria aeroespacial. Os veículos eVTOLs são elétricos e operam de maneira semelhante a um drone. Eles possuem grandes ventiladores omnidirecionais que auxiliam a aeronave a se mover em qualquer direção e a decolar verticalmente. Além disso, esses veículos são caracterizados por emitirem significativamente menos ruídos, quando comparado aos veículos VTOLs convencionais, por serem menos poluentes e operacionalmente mais econômicos.

Os modelos mais recentes de aeronaves eVTOLs para transporte de passageiros seguem, em linhas gerais, os chamados graus de liberdade (DOF), que são parâmetros para avaliar os seus *designs*, sendo eles a carga-útil, velocidade, alcance, ruído, simplicidade, segurança e custo, considerados essenciais para operacionalização nas cidades e escalabilidade do serviço de transporte (Nathen *et al.* 2021).

A Figura 3 a seguir ilustra os 4 modelos mais comuns das arquiteturas de aeronaves eVTOL, comparando-as com sua aplicação prevista nos mercados UAM e RAM.

Figura 3 - Comparação de diferentes conceitos de eVTOLs.

| | Multicopter | Lift + Cruise | Tilt Rotor | Ducted Vectored Thrust |
|----------|---|---|--|---|
| Aircraft |  |  |  |  |
| UAM | ✓ | ✓ | ✓ | ✓ |
| RAM | ✗ | ✗ | ✓ | ✓ |

Fonte: Nathen *et al.* 2021

A capacidade de assentos para passageiros de um eVTOL pode variar entre dois e seis (Biba, 2023) e os 4 modelos mais comuns de aeronaves eVTOLs são elucidados por Patrick Nathen *et al.* (2021) e apresentados aqui:

A primeira categoria de eVTOL é conhecida como arquitetura *Multicopter*. Essa configuração é considerada relativamente simples e apresenta alta eficiência durante a decolagem vertical, pouso e manutenção em voo estacionário, graças a sua baixa carga de disco. No entanto, devido à ausência de asas, os *multicopters* possuem baixa eficiência durante o voo de cruzeiro, o que restringe sua aplicação principalmente ao mercado de RAM.

A segunda categoria de eVTOL é conhecida como arquitetura de elevação e cruzeiro (*lift + cruise*). Essa categoria combina a funcionalidade de decolagem e pouso vertical dos multicópteros com uma aeronave convencional para o voo de cruzeiro. Essa combinação permite que a aeronave alcance tanto uma decolagem e aterrissagem vertical eficientes quanto um voo de cruzeiro eficiente. Para otimizar o desempenho desses conceitos, as hélices utilizadas para o VTOL são projetadas com menos pás e cordas mais curtas, a fim de reduzir o arrasto durante o voo de cruzeiro.

A terceira categoria de eVTOL é conhecida como arquitetura de rotor inclinado (*tilt rotor*), que envolve a inclinação das asas e das hélices, ou apenas das hélices, durante a transição do voo estacionário para o voo em frente. Essa inclinação permite que o eixo da hélice gire 90 graus conforme a aeronave faz a transição. Em geral, essa arquitetura permite a concepção de uma hélice mais otimizada do que seria possível em uma arquitetura de elevação e cruzeiro convencional. No entanto, isso vem com a complexidade técnica adicional e um aumento no tamanho e peso total, devido aos mecanismos de inclinação e passo variável.

A quarta categoria de eVTOL é conhecida como arquiteturas de *Ducted Vectored Thrust*. Uma vantagem significativa dos ventiladores com dutos em comparação com as hélices

sem dutos é a capacidade de reduzir consideravelmente o ruído, tanto o proveniente das pás quanto o de banda larga. Essa redução é alcançada devido à presença do duto, bem como aos revestimentos acústicos instalados dentro dele. Essa característica é particularmente importante quando a carga útil da aeronave é aumentada. Além disso, essa arquitetura traz vantagens de segurança, visto que é capaz de continuar funcionando mesmo com a perda de uma de suas hélices.

No Brasil, a subsidiária da EmbraerX, divisão da fabricante brasileira - Embraer - focada em projetos de “inovação disruptiva”, é a EVE Urban Air Mobility. Como evidenciado por Thiago Vinholes (2021), em artigo para a CNN, o objetivo da EVE é acelerar a criação de soluções para tornar o mercado de Mobilidade Aérea Urbana (UAM) uma realidade. A empresa participará do desenvolvimento de todas as etapas necessárias para habilitar o mercado UAM, incluindo a criação de bases de serviços e redes de suporte, pontos de embarque e desembarque de passageiros em centros urbanos, adaptações nos regulamentos de tráfego aéreo e a produção dos eVTOLs, também conhecidos como “carros voadores”. O protótipo do eVTOL da EVE tem uma configuração exótica com 10 motores elétricos a hélice, sendo oito para permitir a sustentação vertical e outros dois para o impulso horizontal, podendo seu *design* ser classificado na categoria *lift + cruise* (Nathen *et al.* 2021).

A Figura 4 a seguir representa o modelo de aeronave eVTOL da subsidiária brasileira, EVE Air Mobility:

Figura 4 - Modelo de aeronave eVTOL EVE.



Fonte: EVE Air Mobility, 2023.

EVE Air Mobility e Voar Aviation, empresa brasileira de serviços de aviação geral, assinaram uma Carta de Intenções (LOI) para impulsionar a UAM. A Voar fornecerá seus recursos para auxiliar a EVE na modelagem do mercado UAM em cidades selecionadas e apoiar a integração perfeita dos eVTOLs no ambiente urbano, aproveitando sua extensa infraestrutura e conhecimento em aviação. A Voar Aviation tem planos de expandir suas operações nas principais áreas metropolitanas e destinos turísticos populares, como São Paulo, Belo Horizonte, Brasília, Goiânia, Vitória, Florianópolis, Camboriú, Fortaleza, Natal, Recife e Salvador, reforçando o compromisso de ambas as empresas em abraçar novas oportunidades e fornecer soluções inovadoras, sustentáveis e acessíveis para comunidades em todo o Brasil (EVE Air Mobility, 2023).

2.1.4 *Dynamic Pricing* - Precificação Dinâmica

Segundo Arnoud V. den Boer (2015), a precificação dinâmica envolve a análise da identificação dos preços de venda mais vantajosos para produtos ou serviços em um contexto em que os preços podem ser modificados com rapidez e frequência.

A precificação dinâmica é um componente crucial dos Sistemas de Transporte Inteligente (ITS) contemporâneos e serve como solução para vários desafios, incluindo controle de congestionamento, redução de pico de carga e gerenciamento eficiente da mobilidade de veículos tradicionais e elétricos (EVs). Com a adoção de estratégias de preços dinâmicos, os sistemas de transporte podem lidar com esses problemas de forma eficaz e, ao mesmo tempo, garantir a relação custo-benefício. Além disso, a precificação dinâmica permite o planejamento otimizado de rotas para os veículos, contribuindo para o estabelecimento de um ambiente ecologicamente responsável (Saharan; Bawa; Kumar, 2019).

Sandeep Saharan, Seema Bawa e Neeraj Kumar (2019), através do levantamento bibliográfico realizado em sua pesquisa, identificaram quatro áreas da Precificação Dinâmica em ITS: Precificação de congestionamento; Precificação de Tarifas; Precificação de Estacionamento; e Precificação de Carregamento/Descarregamento para veículos elétricos. Para cada área, foram levantadas diversas técnicas de precificação dinâmica usadas para determinar preços dinâmicos. Dado o presente estudo, dar-se-á ênfase às técnicas de ‘Precificação de Tarifas’:

- *Mathematical analysis based techniques* (Técnicas baseadas em análise matemática, traduzido pelo autor): A determinação da tarifa ideal pode ser obtida por meio de várias técnicas, incluindo cálculo de variações, análise de dispersão e análise de árvore de decisão.
- *Optimization based techniques* (Técnicas baseadas em otimização, traduzido pelo autor): As técnicas de otimização desempenham um papel fundamental na determinação de preços de tarifas dinâmicas. Essas técnicas são essenciais para maximizar ou minimizar um ou vários parâmetros, respeitando várias restrições. No setor de companhias aéreas, por exemplo, os preços são definidos dinamicamente para maximizar o lucro, levando em conta vários fatores. Além disso, ao implementar a precificação dinâmica de tarifas, é importante considerar o comportamento dos viajantes, pois eles podem ser influenciados por alterações nas tarifas (sensibilidade). Ao incorporar a otimização e considerar o comportamento do viajante, a precificação dinâmica de tarifas pode ser implementada com eficácia para alcançar os resultados desejados.
- *Queuing theory based techniques* (Técnicas baseadas na teoria de filas, traduzido pelo autor): O modelo de fila M/M/c é usado para gerenciar as chegadas em uma única fila com "c" servidores. Na solução de um DARP (*Dial-ARide Problem*) não míope, por exemplo, a teoria de filas de vários servidores é aplicada com um horizonte infinito de antecipação. As tarifas são determinadas com base em parâmetros como atraso e duração da viagem (Liu; Yang *apud* Saharan; Bawa; Kumar, 2019).

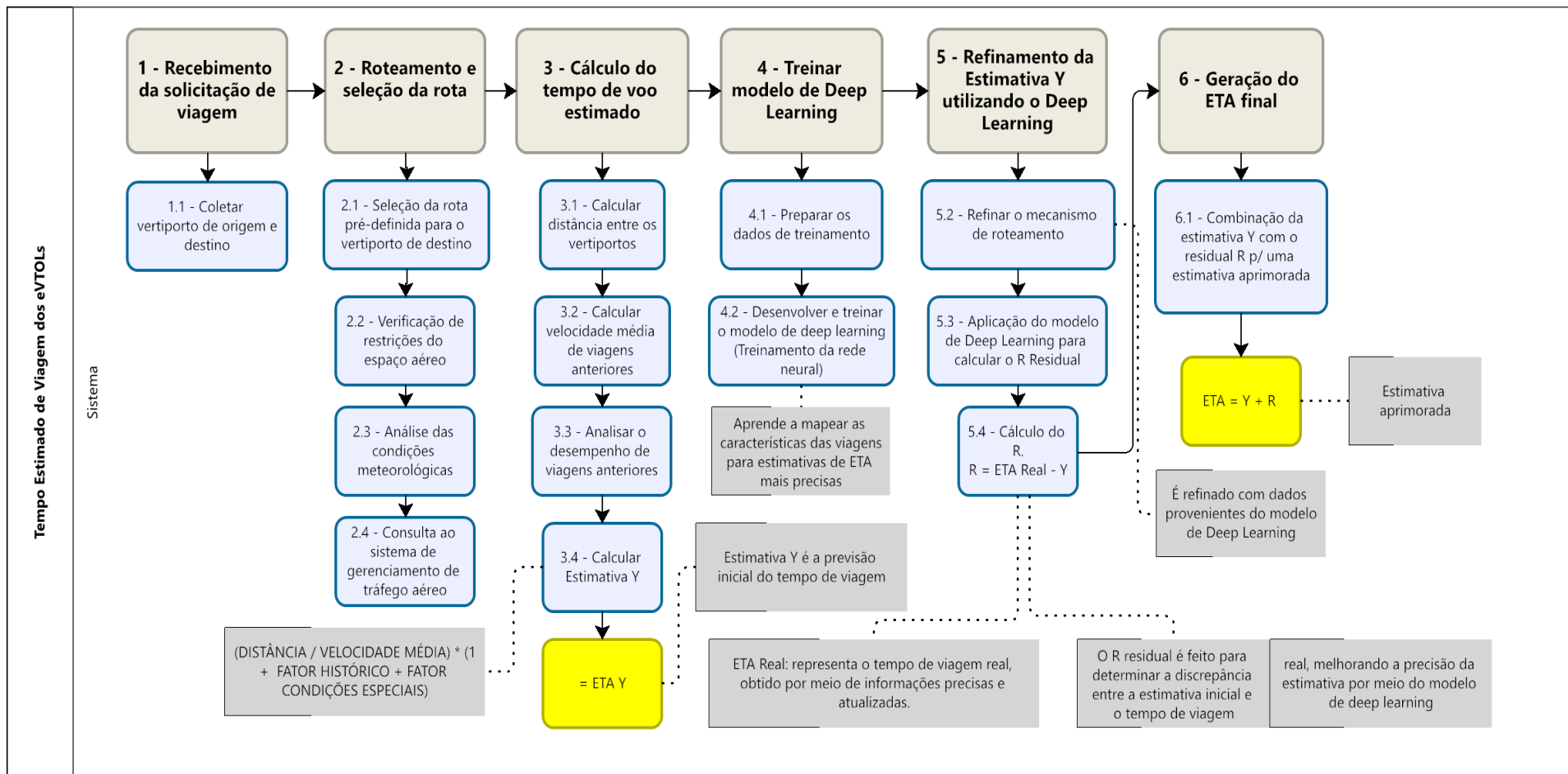
Entende-se que a formulação do modelo de precificação dinâmica de tarifas corrobora para a operacionalização do serviço de transporte por aeronaves do tipo eVTOL, visto que pode trazer eficiência ao cálculo de preços, ao mesmo tempo que o torna mais competitivo no

mercado. Além disso, os sistemas de AAM promovem a integração de sistemas e, portanto, a troca de dados em tempo real. Esses sistemas, se aplicados a um modelo de precificação dinâmica de tarifas eficiente, pode proporcionar sustentabilidade à operacionalização do serviço.

2.2 Modelagem do cálculo ETA para o eVTOL

Algumas pesquisas voltadas para a operacionalização do serviço de transporte de passageiros por eVTOL já começaram a emergir no contexto acadêmico Brasileiro. Dentre essas pesquisas, vale ressaltar a realizada por Marcelo Ferreira (2023) que realiza uma análise do processo de precificação do serviço de transportes da Uber, com o enfoque no cálculo do tempo estimado de chegada (ETA) do veículo, e o adapta para uma proposta de fluxo de processo para o serviço de transporte por eVTOL. A figura a seguir ilustra o resultado da adaptação do modelo.

Figura 5 - Modelagem do cálculo ETA para o eVTOL



Fonte: Ferreira, 2023

O processo evidenciado por Marcelo Ferreira (2023) resultou em uma simplificação do modelo de cálculo da Uber, dada às características da operacionalização dos eVTOLs, que devem trabalhar em rotas pré-definidas, entre vertiportos. Além disso, o processo prevê o acréscimo de algumas variáveis para o cálculo da taxa, como autonomia, velocidade e condições meteorológicas, que devem influir no cálculo do ETA e, conseqüentemente, na precificação dinâmica do serviço.

A partir do mapeamento realizado por Marcelo Ferreira (2023), foi possível compreender o funcionamento de um possível fluxo para o cálculo da tarifa do serviço de transporte por eVTOL e onde que algumas das variáveis que influenciam no cálculo da tarifa de um voo podem ser relacionadas. A referência nesse modelo pode auxiliar o desenvolvimento futuro do modelo matemático para a precificação do serviço, a partir do entendimento das variáveis e restrições que podem impactar no cálculo.

2.3 Google Colaboratory

O *Google Colaboratory*, também conhecido como Google Colab, é uma plataforma de pesquisa baseada em nuvem que permite aos usuários escrever e executar códigos Python. Ele fornece um ambiente de notebook *Jupyter* que não requer configuração para usar. Além disso, oferece acesso gratuito a recursos de computação, incluindo GPUs e TPUs, tornando-o uma ferramenta valiosa para a análise de dados, aprendizado de máquina e pesquisa em inteligência artificial (Santos, 2023).

No contexto deste trabalho, o *Google Colaboratory* foi utilizado como uma plataforma principal para a realização da Análise Exploratória de Dados (AED). Ele permitiu a manipulação eficiente dos dados da ANAC, que consistem em um grande conjunto de dados. As funcionalidades do Google Colab facilitaram a análise e visualização desses dados, permitindo a identificação de padrões e tendências, bem como a relação entre diferentes variáveis.

3. METODOLOGIA

3.1 Tipologia e descrição geral dos métodos de pesquisa

3.1.1 Design Science Research (DSR)

O método utilizado para o desenvolvimento desta pesquisa foi a *Design Science Research* (DSR), derivado do paradigma da *Design Science* (DS).

A pesquisa orientada pelo paradigma da DS possui natureza prática, orientada à resolução de problemas que podem melhorar sistemas existentes ou criar artefatos que promovam uma melhor performance humana em organizações e na sociedade. (Van Aken, 2005; Winter; Aier, 2016, *apud* Cauchick, 2019). Nesse sentido, a presente pesquisa analisa o contexto emergente do mercado do eVTOL e busca contribuir com o avanço nos estudos e soluções dos problemas derivados deste.

Segundo Paulo Cauchick (2019), o pesquisador orientado pela DS deve amparar sua pesquisa em torno da resolução do problema apresentado no contexto estudado. Para isso, deve-se identificar, projetar, avaliar e/ou criar um ou diversos artefatos que assegurem a resolução

do problema ou o aprimoramento da melhoria desejada tanto nas organizações quanto na sociedade.

No âmbito do projeto, ao qual a ciência da *Design Science* se ocupa, há um interesse na definição do que e como as coisas devem ser, envolvendo a concepção de artefatos que possam alcançar determinados objetivos. Dentro desse contexto, os artefatos são compreendidos como construções humanas ou objetos artificiais que podem ser descritos em termos de metas, funcionalidades e adaptações. Dessa forma, o cumprimento de um propósito ou a adaptação a um objetivo envolvem uma relação tripla entre o propósito ou objetivo, as características do artefato e o ambiente em que ele opera (Simon, 1996 *apud* Dresch *et al.*, 2015)

Ademais, entende-se que a função dos artefatos reside na sua capacidade de mediar a interação entre um ambiente interno e um ambiente externo, tendo como objetivo atender a uma finalidade específica (Simon, 1996, *apud* Cauchick, 2019). Ou seja, o ambiente interno configura a forma como o artefato deve funcionar para atender aos requisitos para a solução de um problema. Já o ambiente externo é o meio que traz as condições para o funcionamento do artefato. Dessa forma, uma alteração no ambiente externo pode ocasionar na mudança das condições de funcionamento do artefato, resultando na necessidade de alterações na configuração do mesmo para solucionar o problema determinado, ou seja, alterações no ambiente interno (Van Aken *et al.*, 2012, *apud* Cauchick, 2019).

Conforme citado por Paulo Cauchick (2019, p. 112), “os artefatos desenvolvidos sob o paradigma da *Design Science* podem ser classificados em: constructos, modelos, métodos, instanciacões (March; Smith, 1995) e *design propositions* (Venable, 2006)”.

Dessa forma, a Tabela 1, a seguir, evidencia os tipos de artefatos desenvolvidos sob o paradigma da *Design Science*, bem como as suas respectivas definições.

Tabela 1 - Tipos de Artefatos

| Tipo de artefato | Definição |
|---------------------------|--|
| Constructo | Conceitos utilizados para descrever problemas ou especificar as respectivas soluções. |
| Modelo | Conjunto de elementos e relações que representam a estrutura geral da realidade. |
| Método | Conjunto de elementos e relações que representam a estrutura geral da realidade. |
| Instanciação | Execução do(s) artefato (s) em seu ambiente real, evidenciando a viabilidade e eficácia dos artefatos. |
| <i>Design Proposition</i> | Regras tecnológicas ou regras de projeto, consideradas contribuições teóricas da <i>Design Science</i> . |

Fonte: Adaptado de Metodologia Científica para Engenharia (Cauchick, 2019)

O paradigma da *Design Science* é orientado pela busca da solução satisfatória do problema endereçado, não sendo fundamental a definição da solução ótima. Isso ocorre porque nem toda solução ótima é adequada ao contexto real no qual ocorre o problema. Nesse sentido, as soluções geradas por meio da pesquisa devem prezar pela adequação e utilidade no contexto para o qual foram projetadas (Cauchick, 2019). Segundo Dresch *et al.* (2015), a necessidade de estabelecer uma compreensão precisa dos resultados considerados satisfatórios no paradigma da DS pode ser atendida através de duas abordagens: por meio do consenso entre as partes

envolvidas no problema, e/ou através do avanço da solução atual em relação às soluções proporcionadas pelos artefatos anteriores.

As soluções contempladas com o desenvolvimento do artefato para um problema específico podem ser generalizadas para outros problemas similares. A partir disso, é possível levantar e agrupar diferentes classes de problemas e alocar grupos de artefatos específicos que podem ser capazes de solucionar determinados problemas em um contexto real. Com isso, pesquisadores e profissionais podem aplicar os conhecimentos gerados em contextos distintos, mas com problemas semelhantes (Cauchick, 2019).

Como elucidado por Dresch *et al.* (2015), a *Design Science* é a fundação teórica para a investigação do que é artificial. Por sua vez, a *Design Science Research* (DSR) é o método que fundamenta e opera a condução da pesquisa quando o objetivo a ser alcançado é um artefato ou uma prescrição. Assim, entende-se a DSR como o método ou, de maneira informal, como o “passo a passo” para o desenvolvimento da pesquisa, a qual deve embasar o aprimoramento ou a construção do artefato, contribuindo para o fortalecimento do conjunto de soluções existentes para o contexto.

Vale ressaltar que a condução bem-sucedida da pesquisa na DSR deve seguir dois fatores importantes: o rigor (científico) e a relevância (Dresch *et al.*, 2015). O rigor é fundamental para que a pesquisa seja considerada válida, confiável e possa contribuir para o avanço do conhecimento existente e a relevância (validade pragmática) relaciona-se com a pertinência da pesquisa para os atores diretamente relacionados ao contexto, sejam esses do setor, da organização ou da área estudada.

3.1.2 Análise Exploratória de Dados (AED)

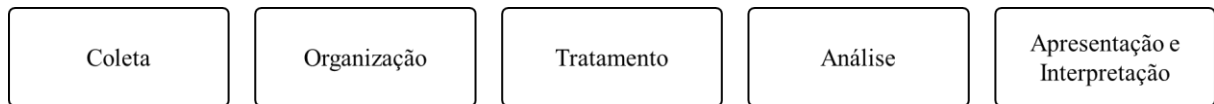
Tendo como base o método da *Design Science Research*, nessa pesquisa foi aplicada a metodologia da Análise Exploratória de Dados (AED), que orientou todo o processo de trabalho da grande quantidade de dados coletados no site institucional da ANAC sobre o mercado do transporte aéreo (ANAC, 2023a).

Atualmente, observa-se um aumento significativo no acesso a dados de diversos tipos e formatos. Em um projeto que envolve o trabalho com grande quantidade de dados, como a pesquisa em questão, é essencial processar e analisar esses dados até que se transformem em informações úteis. Quando essas informações são utilizadas corretamente, elas se transformam em conhecimento. A aplicação desse conhecimento para auxiliar na tomada de decisões é o que leva à sabedoria (Ferreira *et al.*, 2021). Nesse contexto, a pesquisa busca extrair conhecimentos dos dados da ANAC para entender as variáveis que influenciam as tarifas dos voos e aplicar essas percepções ao contexto do transporte por eVTOL.

Um dos grandes focos dessa pesquisa é entender a relação das variáveis do mercado de transporte aéreo com as tarifas praticadas ao longo dos 4 anos (2019 a 2022) e entender possíveis correlações entre elas. Nesse sentido a AED serve como uma ferramenta crucial para examinar dados. Ela permite um entendimento profundo dos dados coletados e, mais importante, das relações existentes entre as variáveis analisadas (Ferreira *et al.*, 2021).

A figura a seguir apresenta as etapas da AED, que foram percorridas ao longo desta pesquisa:

Figura 6 - Etapas da AED



Fonte: Adaptado de Ferreira et al., 2021, *apud* Bonat, Krainski e Mayer, 2020

Segundo Rafael G C. Ferreira (2021), na metodologia AED, os dados coletados são estruturados de maneira que permita a extração de informações úteis das características de interesse. Em seguida, esses dados são processados e ajustados para melhorar a qualidade e a precisão da análise subsequente.

A análise dos dados pode ser realizada de várias maneiras, incluindo análises univariada, multivariada e de correlações. Na análise univariada, cada variável é avaliada individualmente, enquanto na análise multivariada, são estabelecidas relações entre duas ou mais variáveis. As correlações são identificadas quando dois eventos, que não necessariamente têm uma relação de causa e efeito, tendem a ocorrer de forma sincronizada.

Por fim, os dados são apresentados e interpretados por meio de representações gráficas (ou tabulares) que devem ser autoexplicativas e de fácil compreensão. Existem diferentes tipos de gráficos para representar cada uma das variáveis. Essas etapas são fundamentais na pesquisa em questão para entender as variáveis que influenciam as tarifas dos voos e como essas relações podem ser aplicadas ao contexto do transporte por eVTOL.

Para a análise exploratória realizada nesta pesquisa, foram coletados, organizados e tratados quatro principais bancos de dados disponíveis no site institucional da ANAC, utilizando a linguagem de programação Python. Mais especificamente, os de aeródromos, tarifas e os microdados das estatísticas básicas e combinadas dos voos de 2019 a 2022. As etapas na íntegra para a organização e tratamento de cada banco de dados podem ser visualizadas no [Apêndice A](#), respectivamente, nos passos [1](#), [2](#) e [3](#). O apêndice evidencia o notebook do *Google Colaboratory*, no qual foram executadas as quatro primeiras etapas da AED (Ferreira *et al.*, 2021). Ao final do apêndice ([Passo 4](#)), os dados organizados e tratados foram concatenados em um único banco de dados, que foi analisado ([Passo 5](#)) por meio da plotagem de gráficos e visualização dos dados.

3.1.2.1 Passo 1: Aeródromos e Distâncias entre Aeroportos

O passo 1 consistiu na leitura, organização e transformação dos arquivos de aeródromos da ANAC (2023d) e criação do arquivo de banco de dados ‘distâncias’, originado do cálculo das distâncias entre as combinações dos aeroportos de origem e destino. Em suma, essas atividades foram feitas em quatro etapas.

Na primeira etapa, foi realizada a leitura do banco de dados, na qual foi possível identificar as suas variáveis e eliminar as que não contribuiriam - nesse primeiro momento - para a análise da pesquisa. Após, foram eliminados os dados que não possuíam valores registrados (N/A) e salvo o arquivo.

A segunda etapa, foi realizada com base no projeto 'airportsdata20231017' (Borsetti, 2022), 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 com a definição da função para o cálculo da distância do grande círculo entre dois pontos na superfície de uma esfera, denominada “*distance_on_unit_sphere*”. Para o cálculo da distância, é necessário converter a latitude e longitude em coordenadas esféricas.

Conversão de latitude do ponto A e do ponto B para coordenadas esféricas (conversão de graus para radianos):

$$\begin{aligned} \phi_1 &= (90 - \text{latitude1}) * \pi/180 \\ \phi_2 &= (90 - \text{latitude2}) * \pi/180 \end{aligned}$$

Conversão de longitude do ponto A e do ponto B para coordenadas esféricas (conversão de graus para radianos):

$$\begin{aligned} \theta_1 &= \text{longitude1} * \pi/180 \\ \theta_2 &= \text{longitude2} * \pi/180 \end{aligned}$$

Após a conversão da latitude e longitude, a função realiza o cálculo do comprimento do arco, que nada mais é que a distância esférica.:

$$\begin{aligned} \cos &= \sin(\phi_1) * \sin(\phi_2) * \cos(\theta_1 - \theta_2) + \cos(\phi_1) * \cos(\phi_2) \\ \text{arco} &= \arccos(\cos) \end{aligned}$$

Por fim, a função converte o comprimento do arco para distância, multiplicando-o pelo raio da Terra, que é - aproximadamente - 6.378,14 quilômetros:

$$\text{distância do grande círculo} = \text{arco} * 6378.14$$

Após a definição da função para o cálculo da distância entre dois pontos, a terceira 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). Na quarta etapa foi realizada a organização dos dados do arquivo de distâncias e transformação das suas variáveis quantitativas, de *strings* (texto) para *float* (número real).

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

O passo 2 consistiu na leitura, organização e transformação dos arquivos de tarifas da ANAC (2023c) para os anos de 2019 a 2022. Esses bancos de dados são disponibilizados separados por meses do ano. Dessa forma, para cada ano, esses arquivos foram concatenados para um único banco de dados, o qual foi organizado e transformado, convertendo as variáveis quantitativas para *float*, por exemplo.

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

O passo 3 consistiu na leitura, organização e transformação dos arquivos de estatísticas básicas e combinadas, disponibilizados pela ANAC (2023b), dos voos para os anos de 2019 a 2022. Esse passo, dividiu-se em duas grandes etapas, sendo a primeira a organização e o tratamento do banco de dados ‘estatísticas básicas’ e a segunda, a organização, o tratamento do banco de dados ‘estatísticas combinadas’ e, por fim, a correlação entre os dois bancos de dados.

O banco de dados das ‘estatísticas básicas’ foca na aeronave, acompanhando cada movimento desde a decolagem até o próximo pouso. Ele não leva em consideração onde ocorreu o embarque ou desembarque do objeto de transporte (pessoas e/ou cargas), mas sim o status da aeronave em cada etapa do voo, incluindo a movimentação de cargas e passageiros entre os aeródromos de origem e destino da aeronave. Por outro lado, o banco de dados ‘estatísticas combinadas’ se concentra no objeto de transporte (pessoas e/ou cargas). Ele identifica os pares de origem e destino, independentemente da existência de aeródromos intermediários atendidos pelo voo. Essa etapa fornece informações sobre a origem e o destino dos passageiros e cargas transportadas no voo, sem considerar as escalas realizadas. Portanto, enquanto o banco de dados ‘estatísticas básicas’ é centrado na aeronave e suas operações, o banco de dados ‘estatísticas combinadas’ é orientado para o passageiro e seu trajeto (ANAC, 2023e).

Dessa forma, o banco de dados mais adequado para ser estudado nesta pesquisa é o ‘estatísticas combinadas’, pois os dados elucidam os voos orientados aos passageiros e seus respectivos trajetos. Entretanto, algumas variáveis pertencentes ao ‘estatísticas básicas’, ‘tarifas’ e ‘distâncias’ são fundamentais para realizar as análises dos dados estatísticos dos voos entre 2019 e 2022, como a distância (km), o modelo de aeronave, a quantidade de assentos ofertados e tarifas praticadas. Por isso, no passo 3, esses bancos de dados foram correlacionados e essas variáveis foram adicionadas ao banco de dados de ‘estatísticas combinadas’ (2019 a 2022).

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

No passo 4, foi realizada a transformação final dos bancos de dados, onde foram retirados os dados vazios e os bancos de dados de cada ano foram concatenados em um único. Com esse banco de dados final, resultado da concatenação dos demais, foi feita a análise de dados por meio de representações gráficas.

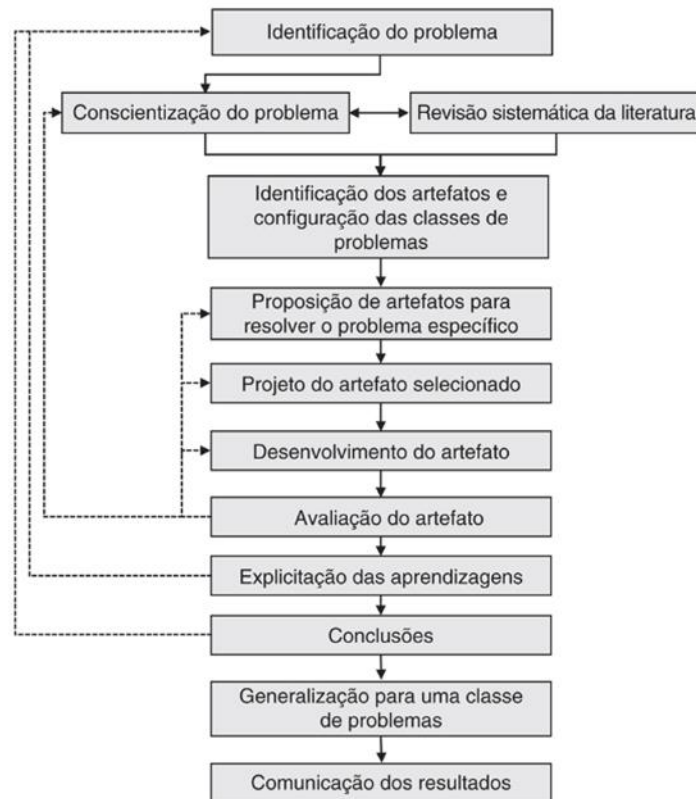
3.2 Caracterização e descrição dos instrumentos de pesquisa

Segundo Van Aken (*apud* Cauchick, 2019), a *Design Science Research* é um método que pode ser empregado para guiar a realização de investigações científicas em várias áreas do conhecimento. É um método recomendado quando o objetivo da pesquisa é prescrever soluções para um problema específico ou ainda desenvolver e/ou avaliar um artefato.

Com a aplicação do método da DSR na presente pesquisa, deve-se identificar as doze etapas propostas por Dresch *et al.* (2015) para a sua condução, as quais fundamentam-se nas

etapas consideradas básicas por diversos autores, destacando-se: (i) a definição do problema; (ii) proposta de sugestões para solucionar o problema; (iii) desenvolvimento do artefato; (iv) avaliação do artefato (Cauchick, 2019). As doze etapas para a condução da metodologia podem ser visualizadas na figura a seguir:

Figura 7 - Etapas para a condução da DSR.



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

I. Identificação do problema

A etapa inicial da DSR envolve a identificação e aprofundada compreensão do problema a ser investigado, abrangendo suas diferentes perspectivas e relações com o contexto no qual está inserido (Pidd, 1998, *apud* Cauchick, 2019).

Como evidenciado no tópico da formulação do problema de pesquisa, o problema estudado neste trabalho encontra-se no desafio de entender o comportamento das tarifas praticadas nos voos tradicionais, entre 2019 e 2022, a partir da correlação das variáveis que potencialmente impactam nesse valor e, com isso, entender como elas poderão ser praticadas para os voos com eVTOLs. Dessa forma, contribuindo com o que futuramente pode vir a ser um modelo de precificação dinâmica para o serviço de transporte de passageiros por eVTOL.

II. Conscientização do problema

Segundo Dresch *et al.* (2015), nessa segunda etapa do método, o problema deve ser aprofundado, buscando entender quais são suas principais causas, sua relação com o contexto que o cerca (interno e externo) e suas facetas. Além disso, nessa etapa, entende-se que desempenho é esperado obter com o artefato e como ele pode impactar na resolução do problema, seus requisitos de funcionamento e funcionalidades.

Entender o contexto que permeia o emergente serviço de transporte com eVTOL é um desafio complexo, visto que há uma grande quantidade de lacunas a serem esclarecidas quanto à implementação deste. Nesse sentido, a presente pesquisa imerge no contexto do mercado de transporte aéreo e busca contribuir para o esclarecimento de alguns desses desafios identificados, especialmente, no que se refere à tarifação do serviço de transporte com eVTOL.

Para isso, entender o contexto atual do mercado de transporte aéreo e estudar as variáveis que influenciam no cálculo das tarifas praticadas é fundamental para entender como será praticado a tarifação dos serviços no contexto mais específico do transporte com eVTOL.

III. Revisão da literatura

Segundo Paulo Cauchick (2019), a fim de obter uma compreensão mais aprofundada da situação problema e facilitar a identificação de potenciais artefatos existentes para a resolução do problema em questão, é recomendado que seja realizada uma revisão sistemática da literatura. Dresch *et al.* (2015) enfatiza que essa abordagem é fundamental para que se tenha acesso aos conhecimentos produzidos em outras pesquisas e que podem suplementar a que está sendo realizada. Além disso, outros artefatos produzidos podem ser utilizados como base para a formulação ou adaptação de um novo artefato, que – consequentemente – deve solucionar o problema em questão.

Nessa etapa, foram evidenciados estudos e materiais produzidos, relacionados ao problema de pesquisa, os quais foram apresentados no referencial teórico deste trabalho. Nesse ponto, pode-se dar destaque ao modelo elaborado por Marcelo Ferreira (2023), que desenvolve uma proposta de modelagem do fluxo do processo para o cálculo do ETA para o serviço de transporte por eVTOL, baseado no modelo de cálculo da Uber. Essa pesquisa é um exemplo de aplicação da metodologia da DSR, visto que seus resultados são complementares para os resultados das demais pesquisas no tema, em especial, para essa pesquisa, que possui o enfoque no estudo das tarifas aéreas praticadas entre 2019 e 2022 e das variáveis que estão correlacionadas.

IV. Identificação dos artefatos e configuração das classes de problemas

Nessa quarta etapa, tendo a conscientização do problema e a revisão sistemática da literatura, foram identificados os artefatos para o endereçamento do problema e das possíveis classes de problemas. Com a identificação desses artefatos existentes, pôde-se sugerir o desenvolvimento de outros artefatos para complementar a pesquisa com a situação problema (Cauchick, 2019). Dresch *et al.* (2015) destaca que a identificação de artefatos existentes, que podem estar endereçados para o mesmo grupo ao qual o problema de pesquisa se relaciona, pode dar assertividade à formulação dos novos artefatos.

Além do artefato desenvolvido por Marcelo Ferreira (2023), nessa etapa, foram identificados os bancos de dados da Agência Nacional de Aviação Civil (ANAC, 2023a), que possui dados do Mercado do Transporte Aéreo Nacional, os quais serão utilizados como insumos para realização de toda a análise exploratória de dados da pesquisa.

V. Proposição de artefatos para resolver o problema específico

Com os artefatos identificados, as classes de problemas estruturadas e as soluções satisfatórias formalizadas, inicia-se a quinta etapa da pesquisa, que consiste na proposição de artefatos para resolver o problema específico. Partindo do entendimento acumulado com os artefatos levantados anteriormente, que traziam a solução para uma perspectiva específica do problema estudado, deve-se adaptá-los à realidade em estudo, propondo o artefato que

considere a solução para o problema do contexto específico trabalhado na pesquisa (Dresch *et al.*, 2015).

Os artefatos propostos para a realização desta pesquisa são o banco de dados que será formulado a partir da combinação e correlação entre as bases da ANAC de “Demanda e Oferta” e “Tarifas” (ANAC, 2023a) e o *notebook* no Google Colab, o qual consolidará toda a organização, tratamento, análise dos dados extraídos do banco de dados, utilizando a linguagem de programação Python e que - neste documento - serão apresentados e interpretados. Dessa forma, conforme evidenciado por March; Smith (1995) e Venable (2006) (*apud* Cauchick, 2019) o tipo de artefato desenvolvido nesta pesquisa consiste em um constructo.

VI. Projeto do artefato selecionado

A sexta etapa prevista para o método da DSR consiste no projeto do artefato selecionado. Neste ponto, identifica-se qual será o artefato-foco das próximas etapas do método (Dresch *et al.*, (2015). Além disso, deve-se projetar como será a construção do artefato, considerando seus componentes, mecanismos internos, pontos de relacionamento com o ambiente e as restrições apresentadas pelo contexto estudado, evidenciando os requisitos que serão considerados para avaliar o alcance da solução satisfatória (Van Aken *et al.*, 2012, *apud* Cauchick, 2019).

Assim, os primeiros esforços da pesquisa concentraram-se na elaboração do banco de dados que será formulado a partir da coleta, organização e tratamento dos dados originados da ANAC. Em seguida, esses dados serão analisados, interpretados e apresentados neste documento.

VII. Desenvolvimento do artefato

A partir do desenvolvimento do projeto do artefato, inicia-se o seu desenvolvimento. Neste ponto, podem ser utilizadas diferentes técnicas, como algoritmos computacionais, representações gráficas, protótipos, maquetes, entre outras. Nessa fase, o pesquisador se dedica à criação do ambiente interno do artefato, conforme proposto por Simon (1996) (*apud* Dresch *et al.*, 2015).

A construção do artefato desta pesquisa se dará por meio de 5 passos derivados da aplicação prática das etapas da Análise Exploratória de Dados (AED), amparada pela linguagem de programação Python: os dados serão coletados dos arquivos disponibilizados pela ANAC, os quais serão organizados no notebook do *Google Colab* e tratados. Após o tratamento dos dados, eles serão analisados a partir da correção entre as variáveis com as tarifas praticadas, interpretados e apresentados neste documento.

VIII. Avaliação do artefato

A etapa de avaliação do artefato consiste na avaliação do comportamento do artefato desenvolvido ao atendimento dos requisitos identificados nas etapas relacionadas à conscientização do problema (etapas 1, 2 e 3). Nessa etapa, é essencial revisar os requisitos estabelecidos e, posteriormente, compará-los com os resultados obtidos para avaliar em que medida essas métricas são atendidas (Dresch *et al.*, 2015). Nesse sentido, nessa etapa será realizada a análise se os artefatos desenvolvidos atendem aos requisitos para contribuir com a operacionalização do serviço de transporte com eVTOL.

IX. Explicitação das aprendizagens

Após a conclusão da etapa de avaliação, na qual o artefato atinge os resultados desejados, é de extrema importância que sejam explicitadas as lições aprendidas ao longo do processo de pesquisa, destacando tanto os sucessos quanto os pontos de insucesso. Essa declaração de êxito e falhas é relevante para o avanço do conhecimento, conforme mencionado por Berends Van Aken e Van Der Bij (2012) e Cole *et al.* (2005) (*apud* Dresch *et al.*, 2015).

X. Conclusões

Como orientado por Dresch *et al.* (2015), as conclusões da pesquisa devem explicitar os desafios e aprendizados encontrados ao longo da elaboração do trabalho. Neste ponto, serão consolidados os resultados alcançados e dadas orientações para pesquisas futuras dentro do contexto trabalhado, renovando o ciclo proposto pela *Design Science Research*.

XI. Generalização para uma classe de problemas

Após a conclusão da pesquisa, é crucial que o artefato desenvolvido, juntamente com suas heurísticas de construção e contingenciais, possa ser generalizado para uma classe de problemas mais ampla. Isso significa que o artefato e suas abordagens podem ser aplicados em contextos semelhantes, com problemas similares (Gregor, 2009; Venable, 2006, *apud* Dresch *et al.*, 2015), ampliando assim o impacto e a utilidade do trabalho de pesquisa na *Design Science*.

XII. Comunicação dos resultados

A comunicação dos resultados da pesquisa consiste na última etapa da metodologia da DSR e é fundamental para que o ciclo de pesquisas na temática possa ser renovado por meio da ampliação do alcance dos conhecimentos gerados (Dresch *et al.*, 2015). Segundo Alturki *et al.* (2011) (*apud* Cauchick, 2019) é essencial que a pesquisa seja acessível tanto para a comunidade acadêmica quanto para os profissionais das organizações relacionadas aos problemas abordados pela *Design Science Research*.

3.3 Caracterização da organização, setor e área

3.3.1 Mobilidade Aérea Avançada (AAM) no Brasil

A presente pesquisa aborda os desafios emergentes do setor da Mobilidade Aérea Avançada (AAM), no que tange os desafios da operacionalização do serviço de transporte por eVTOL na malha de Mobilidade Aérea Regional (RAM) brasileira.

Recentemente, a startup EVE Air Mobility, empresa subsidiária da Embraer, iniciou o projeto para a viabilização do serviço de transporte com eVTOLs no Brasil, iniciando pelas principais metrópoles. A intenção dar-se-á mediante diversas parcerias com outras empresas, dentre elas a Voar Aviation, empresa brasileira de serviços de aviação geral, que recentemente assinou de uma Carta de Intenções (LOI) para avaliar a compra e venda de 70 aeronaves elétricas de decolagem e pouso vertical (eVTOL) para serem operadas em diversas localidades no Brasil (EVE Air Mobility, 2023).

A presente pesquisa pode contribuir para a operacionalização do serviço e está em confluência com os objetivos almejados por ambas as organizações, relacionados à sustentabilidade e acessibilidade da operacionalização do serviço de transporte no Brasil (EVE Air Mobility, 2023).

3.3.2 Agência Nacional de Aviação Civil (ANAC)

A Agência Nacional de Aviação Civil, é uma autarquia federal criada pela lei nº 11.182, de 27 de setembro de 2005. Segundo a Resolução Nº 381, de 14 de junho de 2016, a ANAC - em regime de colegiado - é responsável por analisar e decidir sobre questões relacionadas à aviação civil e à infraestrutura aeronáutica e aeroportuária. Suas competências incluem propor alterações regulatórias, assegurar o cumprimento das normas, orientar a atuação da Agência em negociações internacionais, conceder ou autorizar a exploração da infraestrutura aeronáutica e aeroportuária, exercer o poder normativo da Agência, e aprovar procedimentos administrativos de licitação. Além disso, a ANAC também é responsável por supervisionar a segurança operacional, a adoção de boas práticas e a disseminação da cultura de integridade, e aprovar o Plano de Gestão Anual da ANAC.

A ANAC, por meio das competências que lhe foram atribuídas, regulamenta a obrigatoriedade do fornecimento de alguns dados pelas empresas aéreas. Dentre eles, o registro das tarifas referentes aos serviços aéreos regulares domésticos (Resolução nº 140, de 9 de março de 2010), os dados estatísticos das empresas brasileiras de transporte aéreo público regular e irregular, com exceção das de táxi aéreo (Portaria nº 3.506, de 11 de novembro de 2019) e o cadastro obrigatório dos aeródromos civis (Art.30, Lei nº 7.565, de 19 de dezembro de 1986). Dada essas regulamentações, esses bancos de dados estão disponíveis no site da institucional da ANAC e serviram como fonte para coleta dos dados da pesquisa.

3.4 População e amostra da pesquisa

A população desta pesquisa abrange todas as aeronaves que podem voar em um alcance comparável aos dos eVTOLs, considerando os contextos de Mobilidade Aérea Urbana (UAM), Mobilidade Aérea Regional (RAM) *short-distances*, Mobilidade Aérea Regional (RAM) *long-distances* e Transporte Aéreo Regional (RAT) (Hader; Baur, 2022).

A amostra específica utilizada para esta pesquisa foi extraída do banco de dados da ANAC (2023a) e filtrada para incluir apenas as aeronaves e empresas aéreas de interesse, dentro do intervalo de 4 anos (2019 a 2022). A amostra resultante é composta pelos seguintes modelos de aeronaves:

- AT76 = AEROSPATIALE/ALENIA ATR 72-600/72-212A (600)
- AT75 = AEROSPATIALE/ALENIA ATR 72-500/72-212A (500)
- AT72 = AEROSPATIALE/ALENIA ATR 72 201/202
- AT45 = AEROSPATIALE/ALENIA ATR 42-500
- AT43 = AEROSPATIALE/ALENIA ATR-42-300/320
- C208 = CESSNA 208 CARAVAN

E as seguintes empresas aéreas, que possuem voos registrados com as aeronaves supracitadas:

- PTB = Passaredo
- AZU = Azul
- ACN = Azul Conecta
- PAM = MAP
- ABJ = Abaeté

A amostra resultante contém 717440 linhas de dados. Cada linha representa um voo realizado por uma das empresas aéreas listadas, utilizando um dos modelos de aeronaves mencionados.

Vale destacar que o critério utilizado para definir a amostra da pesquisa, levou em consideração os modelos de aeronaves que se enquadram dentro dos parâmetros elucidados por Manfred Hader; Stephan Baur (2022). Entretanto, o modelo que mais se assemelha ao das aeronaves eVTOLs é o C208, visto que ele possui o menor alcance entre as aeronaves estudadas e a menor capacidade de carregar passageiros e cargas.

3.5 Procedimentos de coleta de dados

Os dados coletados para a realização da pesquisa serão derivados de análise documental e de fontes secundárias da *web*. Dentre as fontes de coleta, destacam-se as Estatísticas Básicas e Combinadas de 2019 a 2022 (ANAC, 2023b), banco de dados das Tarifas praticadas de 2019 a 2022 (ANAC, 2023c) e Lista de aeródromos públicos V.2 (ANAC, 2023d).

3.6 Análise de dados

No decorrer deste estudo, a linguagem de programação Python emergiu como a ferramenta mais adequada para a organização, tratamento e análise dos dados. A decisão de utilizar Python foi baseada em diversos fatores.

A começar, Python é uma linguagem com uma sintaxe intuitiva que favorece a legibilidade, o que facilita a compreensão do código. Ademais, Python conta com uma comunidade de desenvolvedores ativa, o que se traduz em uma vasta quantidade de recursos para aprendizado e resolução de problemas, como o projeto desenvolvido por Mike Borsetti (2022) 'airportsdata20231017', que complementou a etapa de tratamento dos dados coletados.

Outro ponto importante é a existência de diversas bibliotecas, como NumPy © (2023), Pandas © (2023), Matplotlib © (2002–2012), Seaborn © (2012 - 2023) e SciPy © (2023), que oferecem ferramentas robustas para manipulação e análise de dados. Além disso, a ampla adoção de Python tanto na indústria quanto na academia torna o conhecimento desta linguagem uma habilidade relevante e transferível.

Com isso, a escolha de Python para este projeto se deu por razões práticas e estratégicas, permitindo uma análise de dados eficiente e o desenvolvimento de habilidades relevantes no campo da ciência de dados.

Dada as considerações supracitadas, a análise dos dados coletados será conduzida por meio de uma abordagem exploratória, pela plataforma do *Google Colaboratory*, utilizando a linguagem de programação Python, com a complementariedade dos pacotes Pandas © (2023), NumPy © (2023), Seaborn © (2012 - 2023), SciPy © (2023) e Matplotlib © (2002–2012) para o tratamento e visualização dos dados.

4. RESULTADOS E DISCUSSÕES

Como proposto pela pesquisa, os resultados derivados da análise exploratória de dados do Mercado do Transporte Aéreo (ANAC, 2023a) evidenciaram correlações entre algumas das variáveis disponíveis nos bancos de dados da ANAC com os valores de tarifas praticados para os anos de 2019 a 2022, na amostra trabalhada.

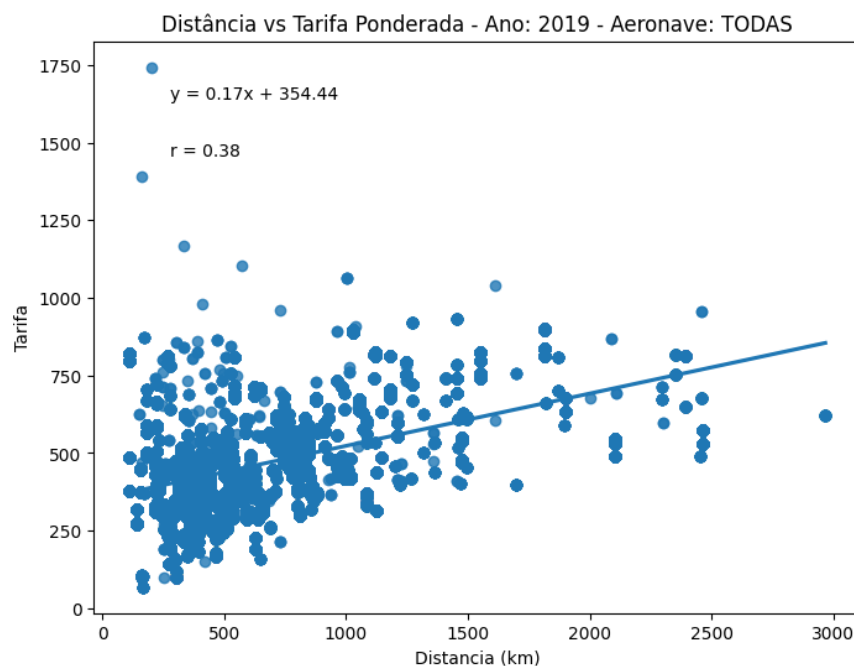
A análise dos dados da amostra se deu com base em duas perspectivas. A primeira leva em consideração toda as aeronaves da amostra de dados, pois buscou-se obter uma perspectiva mais abrangente do mercado da malha regional aérea (RAM e RAT) no Brasil. Em seguida, foram analisados os dados específicos dos voos com a aeronave Cessna 208 Caravan, que - como supracitado - é a aeronave da amostra que mais se assemelha ao que será o eVTOL, em termos de distância e número de assentos ofertados, com capacidade de até quatorze passageiros (tipicamente, nove passageiros) e alcance máximo de - aproximadamente - 1.981,64 quilômetros (Textron Aviation Inc., 2023).

4.1 Análise da Amostra do Mercado da Malha Regional Aérea (RAM e RAT) no Brasil

4.1.1 Análise de Regressão Linear: Variáveis Distância e Tarifa

A regressão linear entre as variáveis distância e tarifa foi realizada para entender a relação entre essas duas variáveis. A distância de um voo pode ser um fator importante na determinação da tarifa, pois voos mais longos podem exigir mais combustível, maior desgaste da aeronave e custos operacionais mais altos. Portanto, é razoável supor que haja uma relação entre as variáveis distância do voo e a tarifa cobrada. Os quatro gráficos a seguir apresentam as regressões lineares realizadas, separadas por ano. Note que em cada gráfico foi representada a equação da reta e o coeficiente de correlação de Person.

Figura 8 - Análise de Regressão Linear: Distância(km) vs Tarifa - 2019



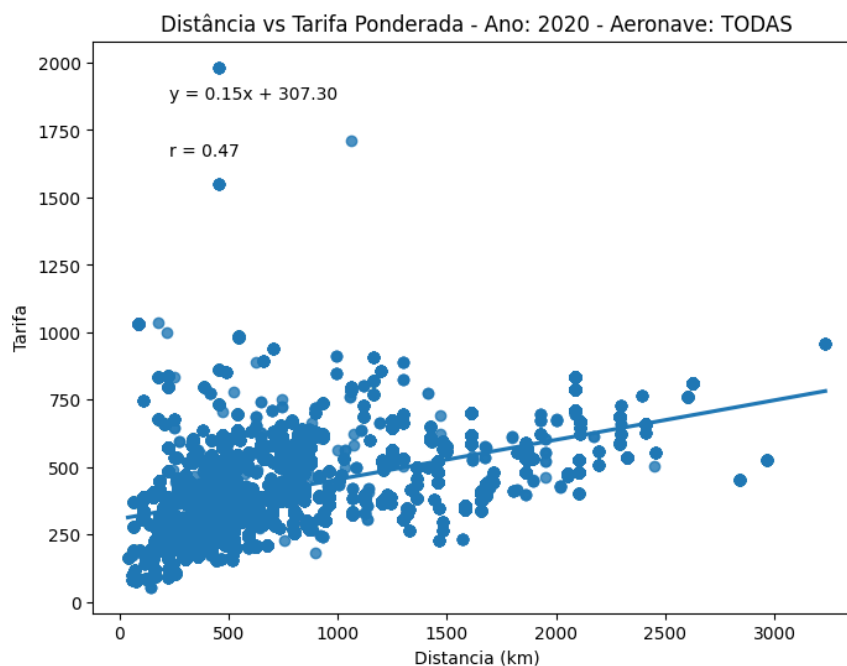
Fonte: elaboração própria.

A equação da reta para 2019, $y = 0,17x + 354,44$, representa a relação entre a distância do voo (x) e a tarifa (y) para o ano de 2019. O coeficiente angular de 0,17 indica que para cada aumento de 1 km na distância do voo, a tarifa aumenta em média R\$0,17. Em outras palavras, esse valor pode ser entendido como a sensibilidade da tarifa praticada para a variação da distância. Além disso, este é um indicativo de que há uma relação positiva entre a distância e a tarifa, ou seja, à medida que a distância do voo aumenta, a tarifa também tende a aumentar.

O valor 354,44 é o coeficiente linear, também conhecido como intercepto. Isso representa a tarifa base quando a distância do voo é zero em 2019. No entanto, na prática, um voo com distância zero não faz sentido. Portanto, este valor pode ser interpretado como a tarifa base ou mínima para um voo, independentemente da distância, no mercado da malha RAM brasileira.

O valor de “ r ” para o ano de 2019 é 0,38. Isso indica uma correlação positiva moderada entre a distância do voo e a tarifa. Consequentemente, o “ r^2 ” - para a regressão com essas variáveis - aponta uma correlação de 14,44%. Possivelmente, analisando a regressão com outras variáveis que potencialmente também impactam na variação da tarifa, o valor de “ r ” tende a aumentar.

Figura 9 - Análise de Regressão Linear: Distância(km) vs Tarifa - 2020



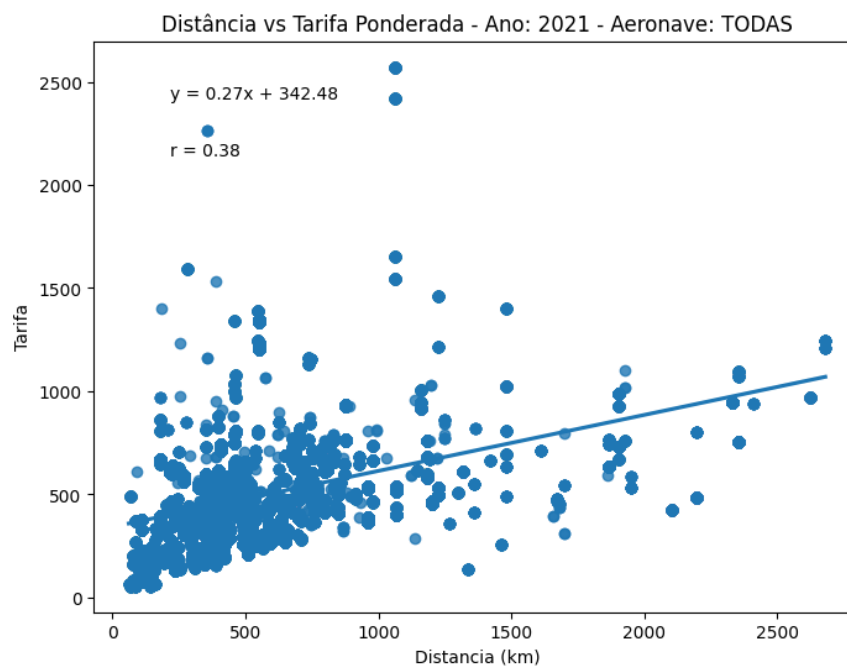
Fonte: elaboração própria.

A equação da reta para o ano de 2020 é $y = 0,15x + 307,30$, o que sugere que, em média, para cada aumento de 1 km na distância do voo, a tarifa aumenta em R\$0,15. O valor do intercepto, 307,30, pode ser interpretado como o custo base ou mínimo para um voo, independentemente da distância em 2020.

Comparando com a equação da reta de 2019, podemos ver que tanto a inclinação quanto o intercepto diminuíram em 2020. Isso pode indicar que, embora ainda haja uma relação positiva entre a distância e a tarifa, o aumento da tarifa por quilômetro voado foi menor em 2020 em comparação com 2019. Além disso, a tarifa base ou mínima para um voo também diminuiu. Essa redução do valor base praticado para tarifa pode estar atrelado a diversos fatores, como - por exemplo - a diminuição geral da demanda (RPK) por voos em 2020, igual a -48,7% em relação à 2019 (ANAC, 2021).

O coeficiente de correlação de Pearson para o ano de 2020 é 0,47. Isso é maior do que o valor de 2019 (0,38), indicando uma correlação positiva mais forte entre a distância e a tarifa em 2020. Em outras palavras, a tarifa está mais fortemente relacionada à distância do voo em 2020 do que em 2019, apontando uma correlação de 22,09% (r^2).

Figura 10 - Análise de Regressão Linear: Distância(km) vs Tarifa - 2021



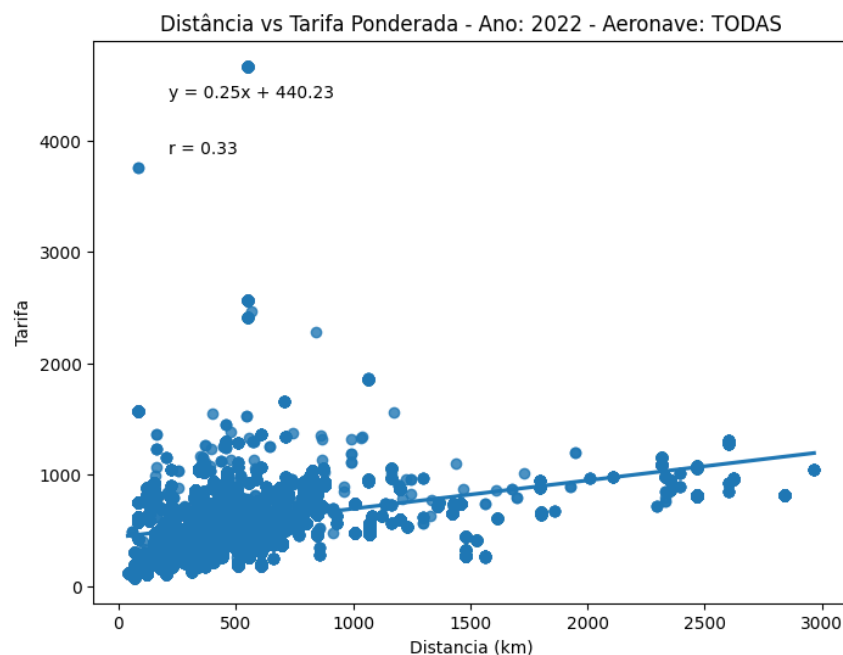
Fonte: elaboração própria.

A equação da reta para o ano de 2021 é $y = 0,27x + 342,48$, o que sugere que, em média, para cada aumento de 1 km na distância do voo, a tarifa aumenta em R\$0,27. O valor do intercepto - 342,48 - pode ser interpretado como o custo base ou mínimo para um voo, independentemente da distância.

Comparando com a equação da reta de 2020, podemos ver que a inclinação aumentou em 2021, enquanto o intercepto aumentou ligeiramente. Isso pode indicar que, embora ainda haja uma relação positiva entre a distância e a tarifa, o aumento da tarifa por quilômetro voado foi maior em 2021 em comparação com 2020. Além disso, a tarifa base ou mínima para um voo também aumentou.

O coeficiente de correlação de Pearson para o ano de 2021 é 0,38, o mesmo valor de 2019. Isso indica uma correlação positiva moderada entre a distância e a tarifa. Em outras palavras, à medida que a distância do voo aumenta, a tarifa também tende a aumentar.

Figura 11 - Análise de Regressão Linear: Distância(km) vs Tarifa - 2022



Fonte: elaboração própria.

A equação da reta para o ano de 2022 é $y = 0,25x + 440,23$, o que sugere que, em média, para cada aumento de 1 km na distância do voo, a tarifa aumenta em R\$0,25. O valor do intercepto, 440,23, pode ser interpretado como a tarifa base ou mínima para um voo, independentemente da distância.

Comparando com a equação da reta de 2021, podemos ver que há uma leve redução da inclinação e um aumento do intercepto em 2022. Isso pode indicar que a sensibilidade para a variação da tarifa praticada, em relação a distância, diminuiu ligeiramente em 2022, contudo a tarifa base aumentou. Esse aumento na tarifa base pode indicar uma tentativa de recuperação do mercado, dado o RPK apresentado em 2020, de -48,7% em relação à 2019 (ANAC, 2021).

O coeficiente de correlação de Pearson para o ano de 2022 é 0,33. Isso indica uma correlação positiva moderada entre a distância e a tarifa.

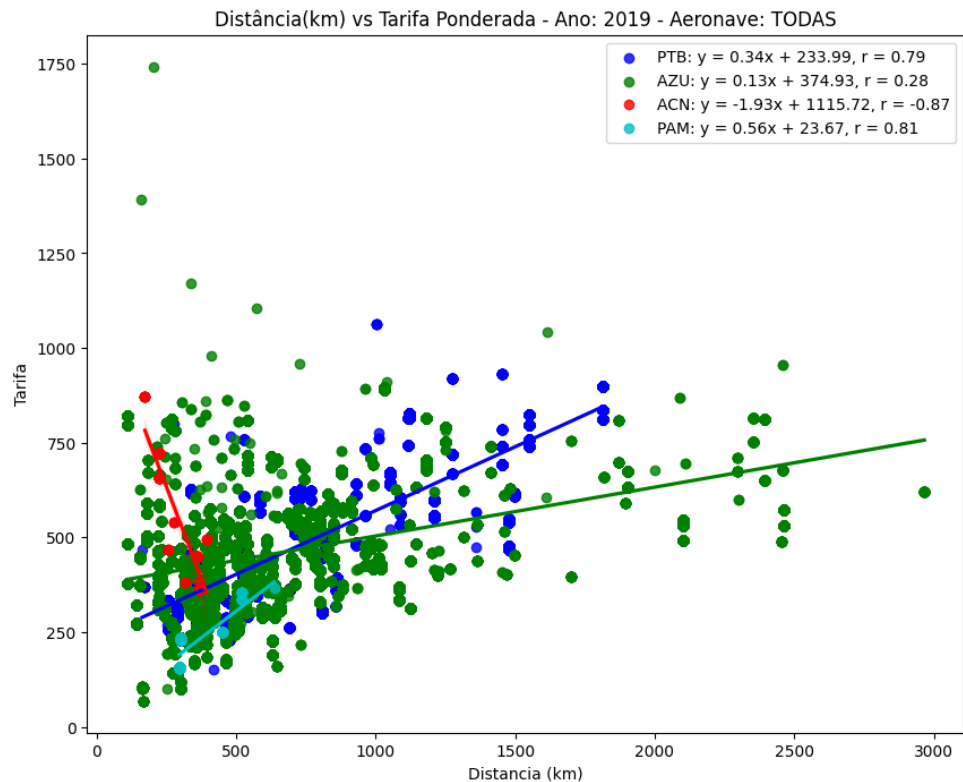
Analisando os gráficos, foi possível perceber que existe uma relação positiva entre a distância e a tarifa, como indicado pela equação da reta. Isso significa que, à medida que a distância do voo aumenta, a tarifa também tende a aumentar. No entanto, o valor de “r” (coeficiente de correlação de Pearson) não é muito alto, o que sugere que a relação entre a distância e a tarifa não é perfeitamente linear. Isso pode ser devido à presença de outras variáveis que também afetam a tarifa.

4.1.2 Análise de Regressão Linear: Variáveis Distância e Tarifa por Empresa Aérea

A regressão linear foi realizada para cada empresa aérea, buscando entender como a relação entre a distância do voo e a tarifa podem variar sobre a perspectiva de cada empresa. Isso é importante porque diferentes empresas aéreas podem ter diferentes estruturas de preços

e políticas de tarifas. Ao realizar uma regressão linear separada para cada empresa, é possível obter perspectivas mais precisas sobre como a distância do voo afeta a tarifa em cada empresa.

Figura 12 - Análise de Regressão Linear: Distância(km) vs Tarifa, por empresa - 2019



Fonte: elaboração própria.

Analisando o gráfico acima, vale destacar algumas percepções para cada empresa. De maneira geral, percebe-se que há uma correlação positiva entre a distância e a tarifa, sendo que para cada empresa, há diferentes níveis de sensibilidade, que variam de acordo com o coeficiente angular apresentado.

Para a empresa aérea MAP (PAM), a equação da reta é $y = 0,56x + 23,67$ e o coeficiente de correlação de Pearson é 0,81. Isso indica uma forte correlação positiva entre a distância e a tarifa, sugerindo que à medida que a distância do voo aumenta, a tarifa também tende a aumentar, ou seja, nessa análise, a variável distância é 65,61% (r^2) determinística para o valor da tarifa. A tarifa aumenta em média R\$0,56 para cada quilômetro adicional voado, o que evidencia a maior sensibilidade da tarifa à variação da distância entre todas as empresas analisadas em 2019.

Para a Passaredo (PTB), a equação da reta é $y = 0,34x + 233,99$ e o coeficiente de correlação de Pearson é 0,79. Isso indica uma forte correlação positiva entre a distância e a tarifa, sugerindo que à medida que a distância do voo aumenta, a tarifa também tende a aumentar. A tarifa aumenta em média R\$0,34 para cada quilômetro adicional voado, o que indica uma alta sensibilidade da tarifa à variação da distância. Além disso, a tarifa base praticada pela empresa para o ano de 2019 é indicada pelo intercepto de 233,99.

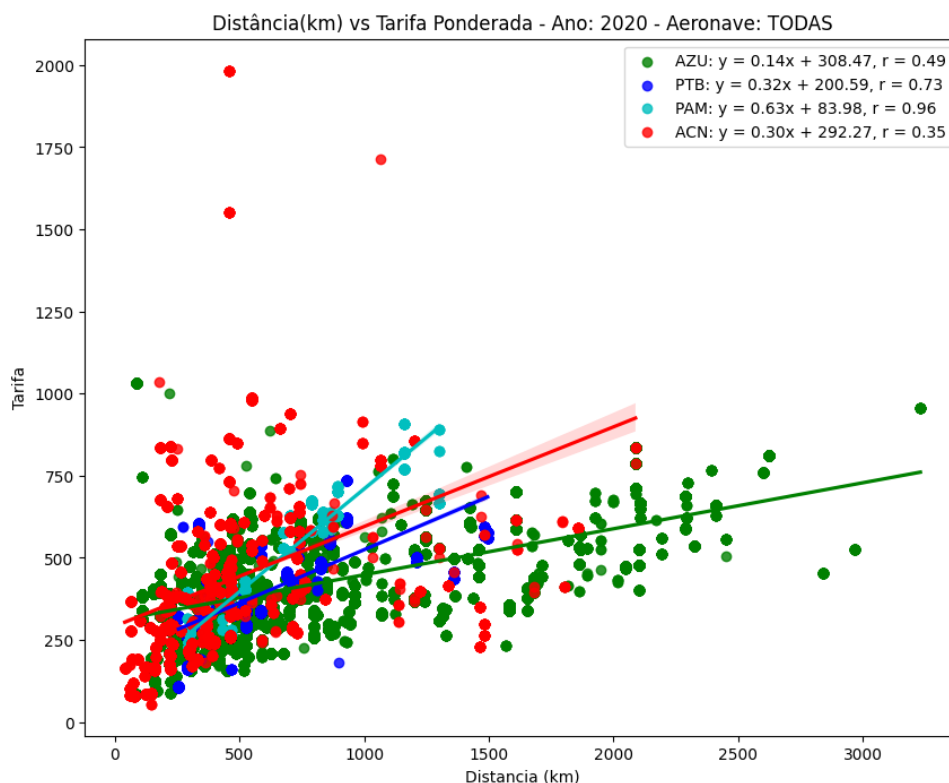
Observando os dados da Azul (AZU), a equação da reta é $y = 0,13x + 374,93$ e o coeficiente de correlação de Pearson é 0,28. Isso indica uma correlação positiva fraca entre a

distância e a tarifa, o que sugere que a relação entre a distância e a tarifa não é perfeitamente linear. Embora a tarifa tenda a aumentar com a distância, essa relação não é tão forte quanto na Passaredo, por exemplo. A tarifa aumenta em média R\$0,13 para cada quilômetro adicional voado. Isso pode ser devido à presença de outras variáveis que também afetam a tarifa, o que pode ser comum em uma empresa como a Azul que, comparativamente, possui uma estrutura mais robusta do que as demais empresas analisadas na amostra e a tarifa pode ser influenciada por outras variáveis decorrentes dessa estrutura.

Por fim, a última empresa observada nessa amostra para 2019 é a Azul Conecta (ACN), a qual apresenta a equação da reta sendo $y = -1,93x + 1115,72$ e o coeficiente de correlação de Pearson é $-0,87$. Isso indica uma forte correlação negativa entre a distância e a tarifa, sugerindo que à medida que a distância do voo aumenta, a tarifa tende a diminuir. Isso é incomum, quando comparado aos demais dados analisados, mas pode estar relacionado à aquisição pela Azul da empresa TwoFlex em 2020, originando a empresa Azul Conecta (ACN) (Basseto, 2020). De alguma forma, esses resultados podem ter sido impactados pela aquisição.

Os próximos três gráficos apresentam comportamentos dos dados semelhantes entre as empresas, com o enquadramento da Azul Conecta (ACN) - após a aquisição pela Azul (AZU) - dentro do comportamento observado pelas demais empresas.

Figura 13 - Análise de Regressão Linear: Distância(km) vs Tarifa, por empresa - 2020



Fonte: elaboração própria.

Figura 14 - Análise de Regressão Linear: Distância(km) vs Tarifa, por empresa - 2021

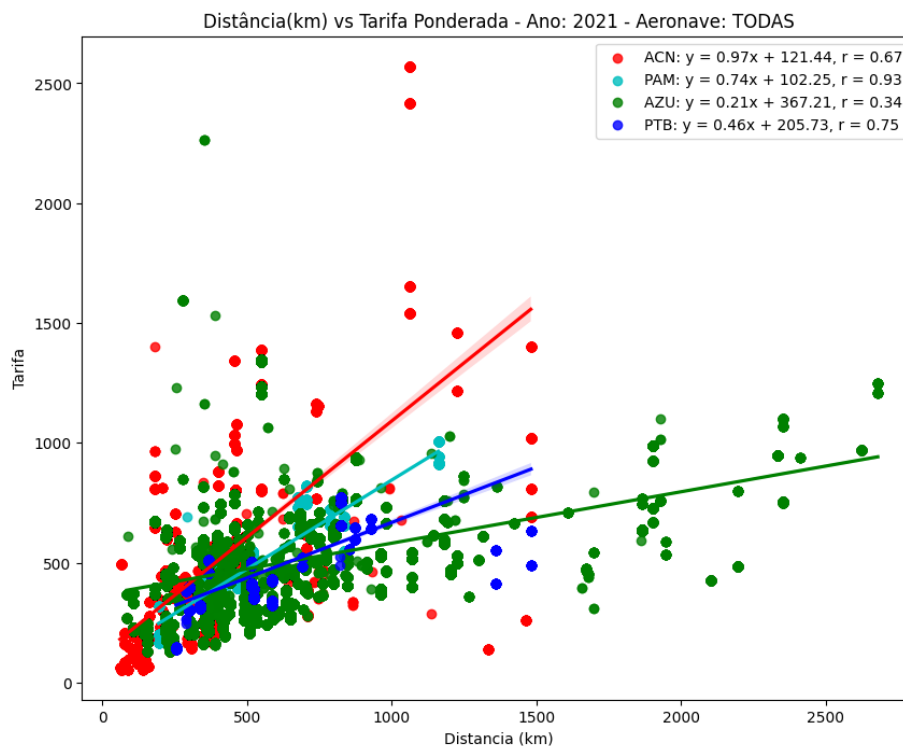
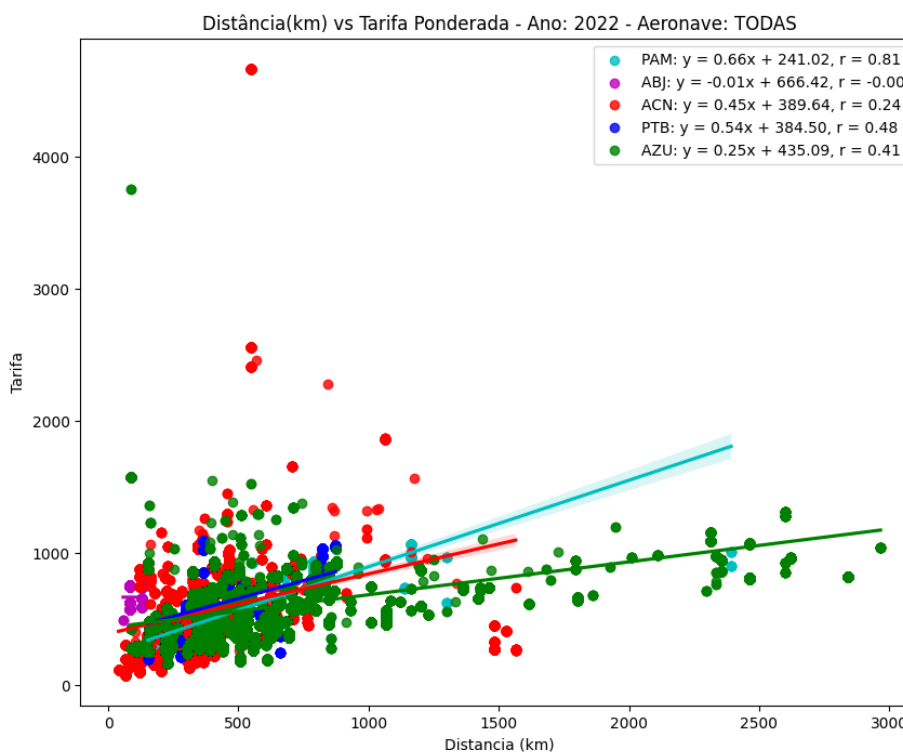


Figura 15 - Análise de Regressão Linear: Distância(km) vs Tarifa, por empresa - 2022



Analisando os gráficos, percebe-se que existe uma relação positiva entre a distância e a tarifa para todas as empresas, como indicado pelas linhas de regressão. No entanto, a força dessa relação, indicada pelo coeficiente de correlação de Pearson (r), varia entre as empresas. As empresas Passaredo (PTB) e Map (PAM) têm o “ r ” maior, indicando uma relação mais forte entre a distância e a tarifa, enquanto a Azul (AZU) e a Azul Conecta (ACN) têm o “ r ” menor, indicando uma relação mais fraca.

Além disso, a inclinação das linhas de regressão, que representa o aumento da tarifa por quilômetro voado, também varia entre as empresas. Algumas empresas apresentam o coeficiente angular maior, indicando maior sensibilidade da tarifa para a variação da distância, enquanto outras têm o coeficiente angular menor, indicando menor sensibilidade da tarifa para a variação da distância.

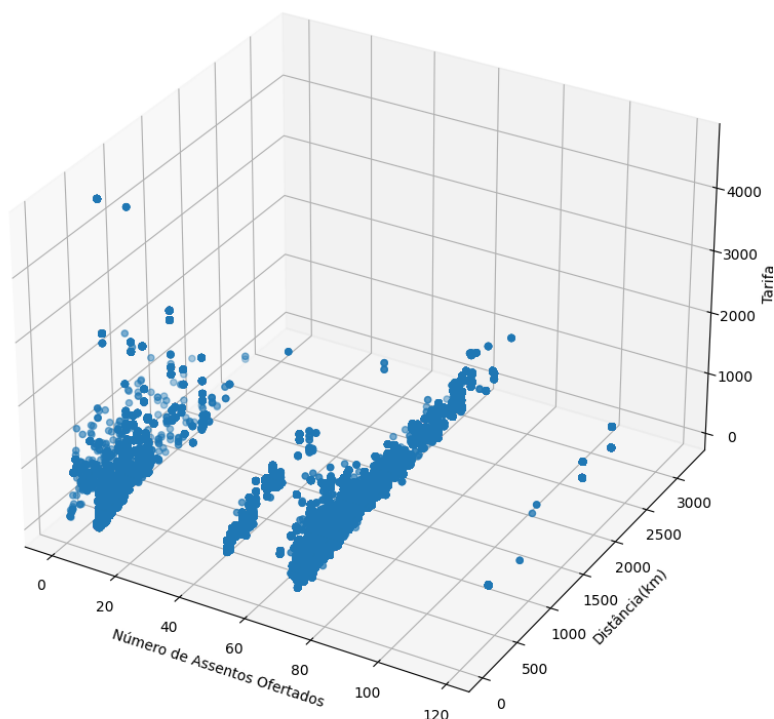
Essas diferenças supracitadas podem ser devidas a uma variedade de fatores, incluindo diferenças nas políticas de preços das empresas, custos operacionais, estratégias de mercado, entre outros. Esses fatores tendem a ser mais marcantes nas empresas com uma estrutura mais robusta e complexa, como a Azul, na qual percebe-se que a determinação da variável distância para o valor da tarifa (r^2) tende a diminuir, pois outras variáveis podem influenciar o valor da tarifa.

4.1.3 Gráfico de Dispersão 3D - Análise de Relação: Distância, Número de Assentos e Tarifas

O gráfico de dispersão 3D foi elaborado para analisar a relação entre as variáveis distância, número de assentos ofertados e tarifa. Como observado, a correlação da distância do voo com a tarifa tende a ser positiva (quando maior a distância, maior a tarifa). O número de assentos ofertados pode afetar a tarifa, pois - por uma relação de oferta e demanda - quanto menor a oferta para uma mesma demanda, espera-se que a tarifa aumente. O gráfico a seguir ilustra essa relação, observada nos anos de 2019 a 2022:

Figura 16 - Gráfico de Dispersão 3D: Nº de Assentos Ofertados vs Distância(km) vs Tarifa (2019 a 2022)

Relação: Nº Assentos Ofertados vs Distância(km) vs Tarifa (2019 a 2022) - Aeronave = TODAS



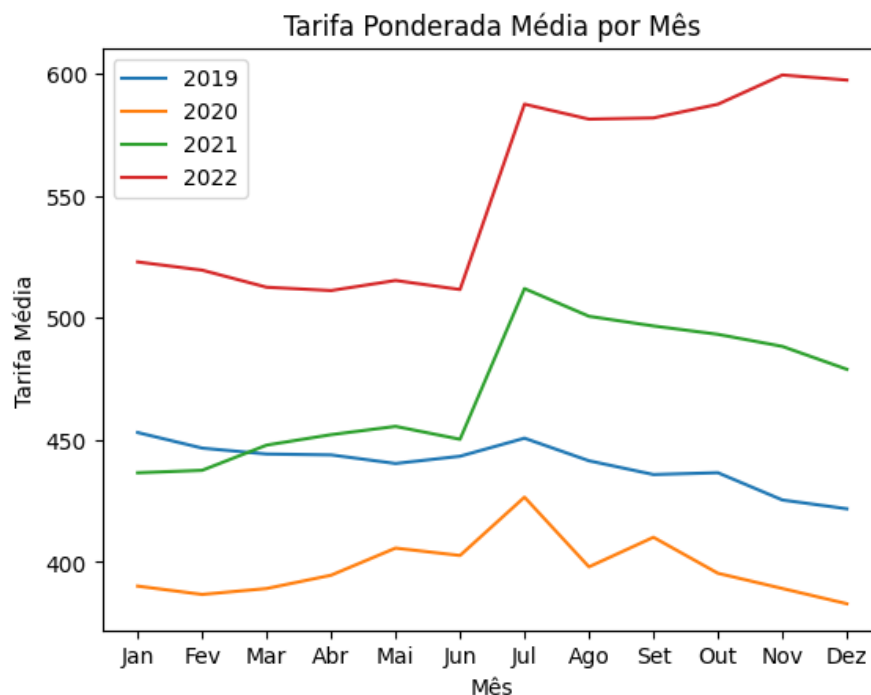
Fonte: elaboração própria.

Analisando o gráfico, percebe-se uma tendência de aumento no valor da tarifa conforme diminui o número de assentos ofertados. Isso pode indicar que voos com menos assentos tendem a ter tarifas mais altas. Uma possível explicação para isso é que voos com menos assentos podem ser operados por aeronaves menores, que podem ter custos operacionais mais altos por passageiro. Alternativamente, pode ser que voos com menos assentos sejam mais propensos a serem voos de nicho ou especializados, que podem cobrar tarifas mais altas.

4.1.4 Lineplot: Tarifa Ponderada Média por Mês (2019 a 2022)

O gráfico Lineplot foi elaborado para analisar o comportamento da variável tarifas ao longo dos meses para os anos de 2019 a 2022. Essa análise visa observar a variação da tarifa ao longo do ano, pois ao visualizar o comportamento das tarifas ao longo do tempo para cada ano, podemos identificar tendências e padrões que podem fornecer perspectivas sobre como as tarifas são determinadas.

Figura 17 - Lineplot: Tarifa Ponderada Média por Mês



Fonte: elaboração própria.

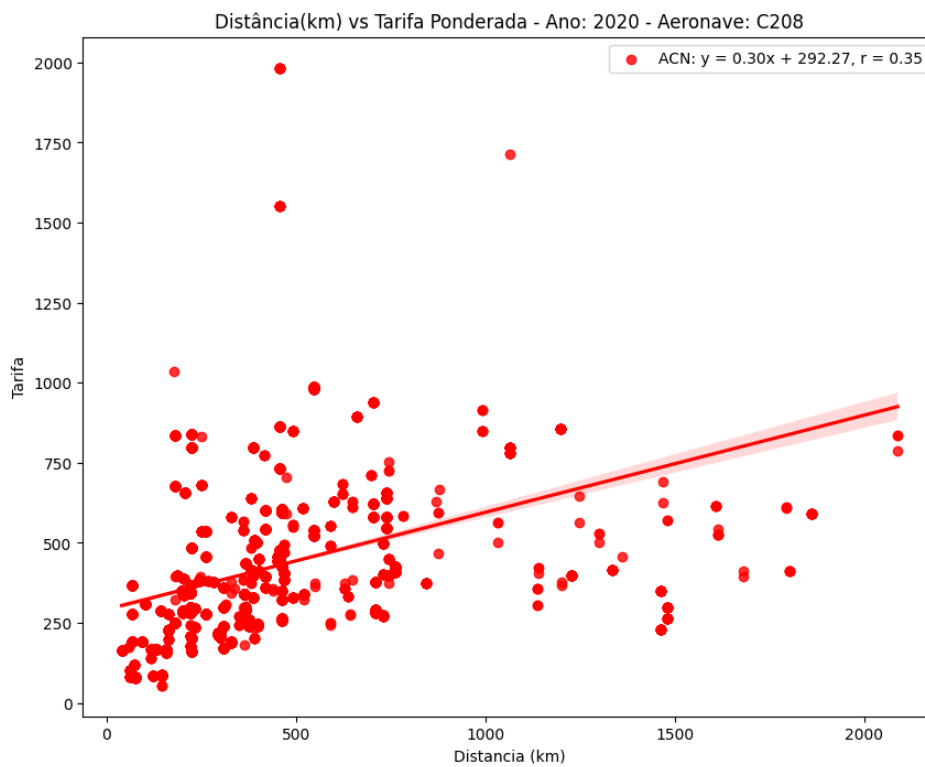
Observa-se que o comportamento das tarifas tende a seguir um padrão para cada ano, com um pico de aumento no meio do ano, decaindo até dezembro, sugerindo que existem fatores sazonais impactando os valores das tarifas. Além disso, observa-se também que houve uma tendência de crescimento da tarifa de 2019 para 2022, com exceção de 2020, ano da diminuição geral da demanda (RPK) por voos, igual a -48,7% em relação à 2019 (ANAC, 2021).

4.2 Análise da amostra para a aeronave Cessna 208 Caravan

4.2.1 Análise de Regressão Linear: Variáveis Distância e Tarifa

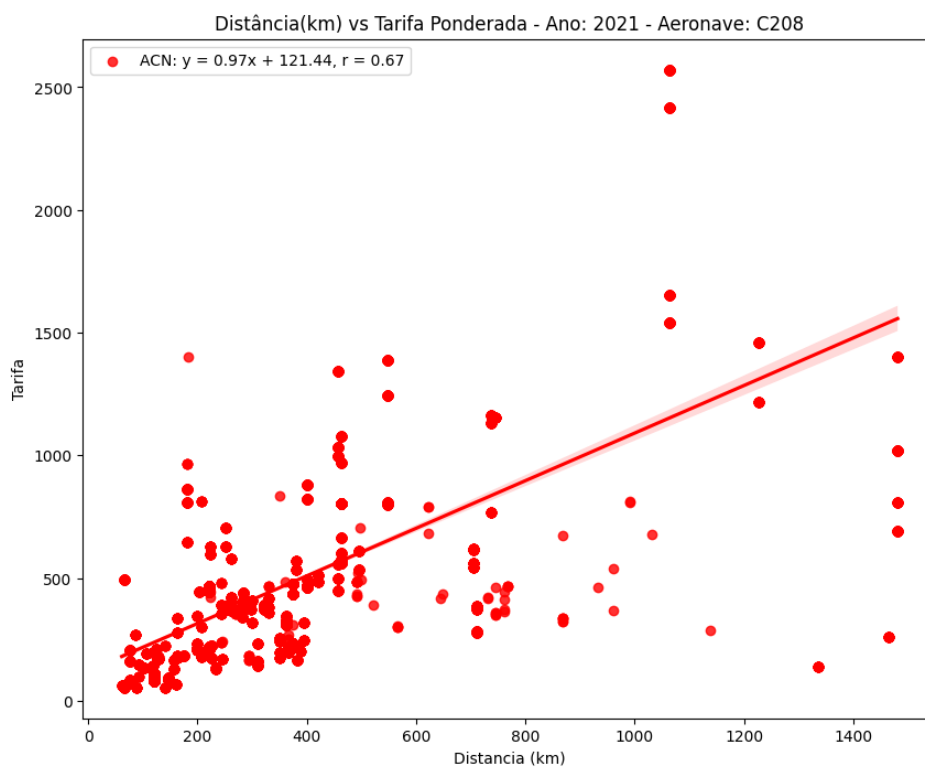
A regressão linear entre as variáveis distância e tarifa foi realizada para o modelo de aeronave Cessna 208 Caravan (C208) para entender a relação entre essas duas variáveis à aeronave que mais se assemelha ao eVTOL. Os gráficos a seguir representam essas análises, um para cada ano, 2019 a 2022. Contudo as análises para o ano de 2019 foram desconsideradas para o C208, visto que ela - por se tratar da empresa Azul Conecta (ACN) - foi afetada em 2019, dado o provável impacto da aquisição pela Azul da empresa TwoFlex em 2020 (Basseto, 2020).

Figura 18 - Análise de Regressão Linear: Distância(km) vs Tarifa - C208 - 2020



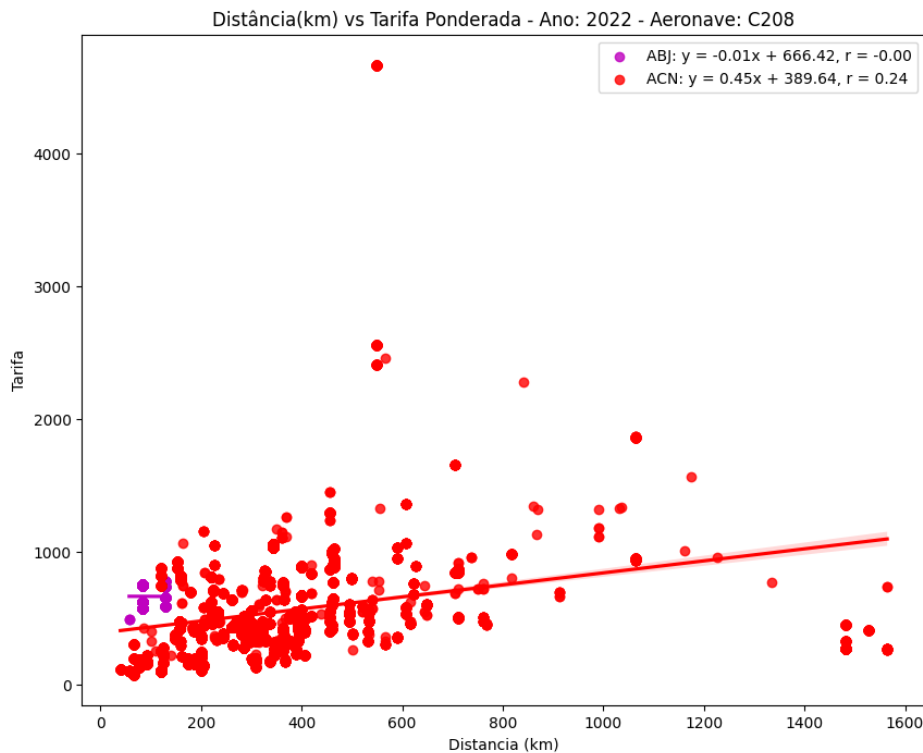
Fonte: elaboração própria.

Figura 19 - Análise de Regressão Linear: Distância(km) vs Tarifa - C208 - 2021



Fonte: elaboração própria.

Figura 20 - Análise de Regressão Linear: Distância(km) vs Tarifa - C208 - 2022



Fonte: elaboração própria.

Analisando os gráficos, foi possível perceber que existe uma relação positiva entre a distância e a tarifa, como indicado pela equação da reta. Isso significa que, à medida que a distância do voo aumenta, a tarifa também tende a aumentar. No entanto, o valor de “r” (coeficiente de correlação de Pearson) não é muito alto, o que sugere que a relação entre a distância e a tarifa não é perfeitamente linear. Isso pode ser devido à presença de outras variáveis que também afetam a tarifa.

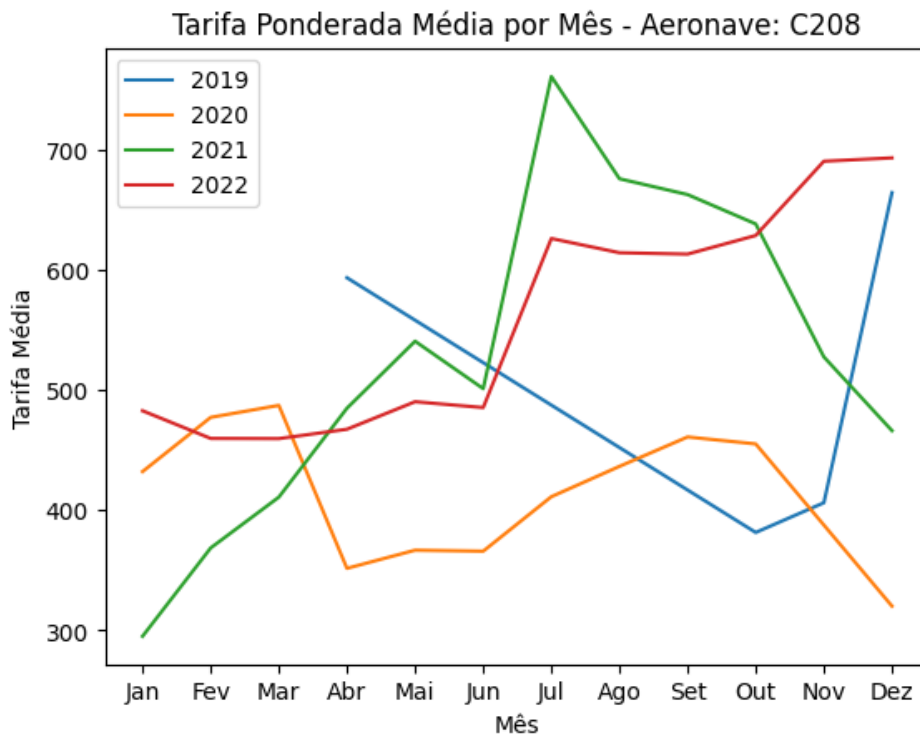
Nesse ponto, vale destacar que em 2022 houve amostra de voos com o C208 para a empresa Abaeté (ABJ), contudo a mesma não pôde ser analisada devido ao seu tamanho pouco representativo.

A análise desses gráficos leva a uma interpretação similar ao do comportamento analisado para o contexto geral do mercado da malha regional aérea (RAM e RAT) no contexto específico dos voos com o C208.

4.2.2 Lineplot: Tarifa Ponderada Média por Mês (2019 a 2022)

O gráfico Lineplot foi elaborado para entender o comportamento da variável tarifas dos voos com o modelo de aeronave Cessna 208 Caravan (C208) ao longo dos meses para os anos de 2019 a 2022. Ao visualizar o comportamento das tarifas ao longo dos meses para cada ano, podemos identificar tendências e padrões que podem nos dar perspectivas sobre como as tarifas poderão ser determinadas para os veículos eVTOLs.

Figura 21 - Lineplot: Tarifa Ponderada Média por Mês - C208



Fonte: elaboração própria.

A partir da análise do gráfico, é possível perceber que em 2021 e 2022 há picos de aumento e diminuição no valor das tarifas praticadas – especialmente – para os meses de maio, junho, julho e agosto. Entretanto, esse mesmo padrão não pôde ser observado para os anos de 2019 e 2020. Isso pode ocorrer, dado que a aquisição da Azul e a queda da demanda (RPK) por voos apresentados para esses anos, podem ter afetado o comportamento das tarifas para os voos com o C208.

Com essas análises, também foi possível observar uma tendência de sazonalidade das tarifas praticadas para o C208, de acordo com o período do ano. Contudo, esses dados não são conclusivos e podem ser influenciados por outras variáveis.

A análise das tarifas praticadas em 2019 a 2022 para os voos tradicionais suscitou percepções que podem ser aplicadas ao contexto do serviço de transporte por eVTOLs. Dentre elas, dá-se destaque a correlação positiva entre a tarifa e a distância, na qual quanto maior a distância, maior tende a ser a tarifa praticada. Essa percepção pode ser aplicada ao mesmo contexto do transporte por eVTOLs, no qual a maior distância de voo pode implicar tarifas maiores. Essa correlação é esperada, visto que distâncias maiores implicam também maiores custos ao voo. Vale ressaltar que a força dessa correlação pode variar dependendo de vários fatores, como o modelo do eVTOL, a eficiência energética, entre outros.

Outra percepção obtida foi que, nos voos tradicionais, há uma tendência de tarifas mais altas para voos com menos assentos. Isso pode ser ainda mais relevante para o transporte por eVTOLs, que geralmente têm uma capacidade de assentos menor em comparação com as aeronaves tradicionais. Portanto, as tarifas para o transporte por eVTOLs podem ser mais altas devido à menor capacidade de assentos.

Ademais, assim como nos voos tradicionais, as tarifas no transporte por eVTOLs podem variar ao longo do ano devido a fatores sazonais. Por exemplo, pode haver uma demanda maior e, portanto, tarifas mais altas durante certos períodos do ano, como feriados ou eventos especiais. Sabendo que o serviço de transporte por eVTOL deve ser operacionalizado em malhas aéreas regionais e urbanas, eventos sazonais regionais/locais (como datas comemorativas locais, feriados estaduais etc.) podem impactar de maneira mais pontual a sazonalidade das tarifas praticadas nesse tipo de serviço.

Por fim, sabe-se que provavelmente existem outras variáveis que impactaram nos valores das tarifas praticadas para os voos tradicionais em 2019 a 2022 e - da mesma forma - outras variáveis devem impactar os valores de tarifas praticados no serviço de transporte por eVTOLs. Como elucidado por Marcelo Ferreira (2023), o processo de 'Modelagem do Cálculo ETA para o eVTOL' prevê o acréscimo de algumas variáveis para o cálculo da taxa, como autonomia, velocidade e condições meteorológicas, que devem influir no cálculo do ETA e, conseqüentemente, na precificação dinâmica do serviço.

Como elucidado por Cedric Y. Justin, Alexia P. Payan e Dimitri N. Mavris (2022), na mobilidade aérea regional (RAM) o baixo volume de passageiros, a competição com outros modelos de transporte e os altos custos operacionais associados são fatores que corroboram para o seu encarecimento. A modalidade de transporte por eVTOL tende a aumentar a eficiência operacional da RAM, contudo ainda permeia o desafio do baixo volume de passageiros e a competição com outros modelos de transporte. Dessa forma, as tarifas praticadas tendem a diminuir, mas sem a garantia de que solucionará todos os desafios relacionados à operacionalização na RAM.

5. CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

5.1 Considerações Finais

Observou-se que há uma correlação positiva entre a variável distância e as tarifas praticadas pelas empresas aéreas. No entanto, essa correlação é mais determinística em algumas empresas do que em outras, sugerindo que a distância do voo é um fator relevante na determinação da tarifa, mas o grau de correlação varia entre as empresas.

Percebe-se que existe uma provável relação entre a complexidade e o tamanho da estrutura da empresa aérea com a quantidade de variáveis consideradas no cálculo das tarifas praticadas. Por exemplo, empresas com estruturas menos robustas tendem a ter uma correlação mais forte entre a distância e a tarifa, sugerindo que a distância é um fator mais determinante na definição das tarifas nessas empresas. Essa constatação é sustentada pela lógica de que, quando a estrutura da empresa é menos complexa, a estrutura de custos - por exemplo - também tende a ser, o que corrobora para a diminuição de variáveis que determinarão o cálculo da tarifa praticada.

Além disso, percebeu-se uma tendência inversamente proporcional ao número de assentos ofertados no voo com a tarifa praticada. Isso sugere que voos com menos assentos tendem a ter tarifas mais altas, o que pode ter implicações para o futuro serviço de transporte por veículos elétricos voadores (eVTOLs), que possuem uma capacidade que varia entre 2 e 6 assentos ofertados.

Constatou-se que, como esperado, outras variáveis, além das analisadas no banco de dados da ANAC, possivelmente impactaram as tarifas praticadas dos voos entre 2019 e 2022. Identificar essas variáveis é um grande passo para compreender como as empresas aéreas determinam as tarifas praticadas e como essas variáveis podem ser utilizadas para a definição de um modelo de cálculo das tarifas a serem praticadas no serviço de transporte por eVTOLs.

Outra constatação realizada é que eventos externos possivelmente impactam nas tarifas praticadas, visto que podem impelir restrições que podem impactar no cálculo da tarifa ótima a ser praticada pelas empresas aéreas.

Por fim, constatou-se que existe uma relação entre o período do ano (mês) e as tarifas praticadas, visto que foi observado um certo padrão entre o comportamento das tarifas entre os anos observados. Dessa forma, a sazonalidade é outro fator que impacta as tarifas praticadas e deve impactar o cálculo das tarifas que futuramente devem ser praticadas no serviço de transporte por eVTOLs.

Essas constatações fornecem insights valiosos sobre o comportamento das tarifas aéreas e podem ser úteis para futuros serviços de transporte por eVTOLs. No entanto, é importante lembrar que a correlação não implica causalidade, e existem muitos outros fatores que podem influenciar a tarifa de um voo. Portanto, embora essas análises forneçam uma visão geral da relação entre distância, número de assentos ofertados e tarifa, elas não capturam todas as nuances dessa relação.

Vale destacar que a utilização da linguagem de programação Python desempenhou um papel fundamental no alcance dos resultados desta pesquisa. A facilidade de uso, a eficiência e a flexibilidade oferecidas pela linguagem permitiram uma análise de dados mais aprofundada e precisa. Além disso, a ampla gama de bibliotecas disponíveis proporcionou uma maior capacidade de manipulação e visualização de dados, o que contribuiu para uma melhor compreensão dos resultados. Portanto, é altamente recomendável a utilização de Python em pesquisas futuras que envolvam o estudo e tratamento de dados, dada a sua comprovada eficácia neste projeto.

5.2 Objetivos Atendidos da Pesquisa

Dado o objetivo geral deste trabalho de analisar e compreender as variáveis apresentadas pelo banco de dados da ANAC que influenciaram as tarifas dos voos tradicionais entre 2019 e 2022, e correlacionar as percepções obtidas com as análises desses dados aos serviços de transporte por eVTOL no modelo de Malha Regional Aérea (RAM) brasileira, pode-se dizer que foi atendido. Contudo, percebe-se que existem outras variáveis, além das apresentadas nos bancos de dados da ANAC que determinam os preços praticados das tarifas nos voos da malha aérea regional brasileira.

Com relação aos objetivos específicos da pesquisa, conclui-se que esses também foram alcançados, pois - no que tange a análise da evolução das tarifas dos voos tradicionais ao longo dos anos de 2019 a 2022 - essa análise foi realizada e foi possível observar alguns padrões e comportamentos supracitados, apresentados nesse intervalo de anos.

Com relação à identificação de algumas das variáveis que têm impacto no valor das tarifas dos voos tradicionais, essas também foram identificadas, vide a distância (km) positivamente correlacionada ao aumento das tarifas praticadas no intervalo entre 2019 e 2022,

o número de assentos ofertados, apresentando uma tendência inversa às tarifas praticadas (ou seja, quanto menos assentos, maior tende ser a tarifa do voo) e - por fim - a tendência na sazonalidade observada das tarifas, relacionadas ao mês do ano. Vale ressaltar que todas as relações observadas não são conclusivas, visto que possivelmente existem outras variáveis que influenciam o valor das tarifas praticadas nesses anos e não foram contempladas nos bancos de dados analisados da ANAC.

Por fim, a partir da relação das variáveis analisadas com os valores das tarifas praticadas, foi possível levantar perspectivas de como as variáveis podem impactar o contexto do serviço de transporte por eVTOLs.

5.3 Desafios e Limitações da Pesquisa

Para a realização da pesquisa, alguns desafios e limitações foram enfrentados. Dentre eles, nota-se o desafio para realizar o tratamento dos dados trabalhados. Parte dos bancos de dados extraídos da ANAC, não trazem dados conclusivos e/ou não apresentam todos os dados desejados. Para contornar situações desse tipo, foi necessário utilizar os dados fornecidos para alcançar os desejados e, para isso, algumas variáveis foram calculadas, como a distância entre as diferentes combinações de aeroportos de origem e destino e a inclusão de aeródromos que não estavam contemplados no banco de dados ([Apêndice A – Inclusão de Aeroportos](#)).

Outro grande desafio da pesquisa foi correlacionar o banco de dados de tarifas com o de estatísticas combinadas. 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. Antes de realizar essa correlação, foi necessário adaptar o arquivo de tarifas, calculando - para o mesmo ano, mês, empresa, origem e destino do voo - as tarifas ponderadas correspondentes. ([Apêndice A – Inclusão da coluna de Tarifas nas Estatísticas Combinadas](#)).

Além dos desafios da pesquisa, algumas limitações da pesquisa originaram-se das correlações necessárias entre os bancos de dados. Dentre as limitações temos que, para a amostra do Cessna 208 Caravan (C208) não foi feita a análise entre as variáveis número de assentos ofertados, distância (km) e tarifa praticada, pois o número de assentos, para todos os voos foi único, igual a nove. Isso ocorre, pois a amostra evidencia os voos - na grande parte - da empresa Azul Conecta (ACN), a qual, conforme os dados analisados, ofertou voos com esse padrão, o que inviabilizou uma análise específica dessas variáveis para o modelo C208.

Outra limitação da pesquisa, ocorreu na correlação entre os bancos de dados de tarifas (2019 a 2022) e estatísticas combinadas (2019 a 2022). Nessa correlação, algumas amostras de voos realizados pela empresa Aerosul (ASO) com o Cessna 208 Caravan (C208) não foram para a amostra final, pois esses voos comercializados pela empresa não estavam contemplados nos bancos de dados de tarifas, diminuindo a amostra específica para os voos com o C208.

Por fim, após as análises gráficas realizadas, foram identificados os *outliers* da amostra – especificamente – os das variáveis tarifa e distância ([Apêndice A - Identificando Outliers](#)).

Com isso, foram identificados os 17.820 voos com outliers para a variável tarifas e 56.278 voos com outliers para a variável distância.

5.4 Direcionamentos para Trabalhos Futuros

Os primeiros direcionamentos para pesquisas futuras, estão relacionados ao refinamento do trabalho realizado nesta. Como esta pesquisa consistiu em uma análise exploratória dos dados fornecidos pela ANAC e uma investigação de possíveis correlações das variáveis apresentadas com as tarifas praticadas (2019 - 2022), não foi prioridade limpar os dados coletados, mitigar inconsistências ou dados contraditórios e eliminar *outliers*. Nesse sentido, o primeiro direcionamento é com relação a esse trabalho de refinamento dos dados.

Um outro direcionamento para pesquisas futuras é explorar mais as correlações das variáveis existentes nos bancos de dados públicos da ANAC e de outras organizações nacionais e internacionais. Desses bancos de dados, é possível extrair informações que podem trazer perspectivas muito úteis para os cenários futuros dos serviços com eVTOLs.

Entender as demais variáveis independentes que estão correlacionadas ao cálculo das tarifas praticadas nos voos tradicionais também é uma etapa para iniciar o desenvolvimento do modelo matemático de precificação do serviço de transporte por eVTOL. Para esse entendimento completo, sugere-se para pesquisas futuras com o foco no levantamento dessas variáveis e a exploração delas por análise de regressão linear multinível. Dessa forma, entende-se sobre uma perspectiva mais realista o comportamento da tarifa praticada nos voos tradicionais, sobre a influência de um conjunto de variáveis.

Após o refinamento de quais são o conjunto de variáveis consideradas para o cálculo das tarifas praticadas dos voos, sugere-se a exploração dos conceitos de Pesquisa Operacional, segundo Frederick S. Hillier e Gerald J. Lieberman (2013), para entender a melhor aplicação dessas variáveis, com vistas a otimizar o cálculo da tarifa. Para isso, será necessário entender as restrições que permeiam esse cálculo e definir uma função objetivo capaz de definir um valor ótimo de tarifa para um voo específico.

Somado ao conceito de Pesquisa Operacional, seguindo o desenvolvimento do modelo matemático para o cálculo do valor ótimo de tarifa, aplicam-se as técnicas de precificação dinâmica baseadas em otimização (Saharan; Bawa; Kumar, 2019), amparadas por um modelo de processo completo, como o apresentado - em parte - por Marcelo Ferreira (2023). Com isso, obtém-se um modelo otimizado para o cálculo das tarifas a serem praticadas no futuro serviço de transporte por eVTOL.

Contudo, para chegar nesse estágio de desenvolvimento, muitas pesquisas e trabalhos futuros ainda precisam ser desenvolvidos, dado que muitos desafios ainda cercam o contexto de desenvolvimento do serviço de transporte por eVTOL e a *Advanced Air Mobility* (AAM).

REFERÊNCIAS

AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL - 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. 2023e

AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL - ANAC. **Mercado do Transporte Aéreo**. Brasília, jan. 2023a. Disponível em: <<https://www.gov.br/anac/pt-br/assuntos/dados-e-estatisticas/mercado-do-transporte-aereo>>. Acesso em: 25 nov. 2023a.

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. 2023d

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. 2023b.

AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL - 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: 25 nov. 2023c.

AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL - ANAC. **Painel de Indicadores do Transporte Aéreo 2020**. Jan, 2021. Disponível em: <<https://www.gov.br/anac/pt-br/assuntos/dados-e-estatisticas/mercado-do-transporte-aereo/painel-de-indicadores-do-transporte-aereo/painel-de-indicadores-do-transporte-aereo-2020>>. Acesso em: 2 dez. 2023.

ANTCLIFF, Kevin *et al.* Regional Air Mobility: Leveraging our National Investments to Energize the American Travel Experience. [S.l]: NASA, abril 2021. Disponível em: <<https://sacd.larc.nasa.gov/ram/>>. Acesso em: 17 jul. 2023.

BASSETO, M. **Após aquisição pela Azul, TwoFlex revela que perderá seu nome**. Maio de 2020. Disponível em: <<https://aeroin.net/apos-aquisicao-azul-twoflex-perdera-nome-conecta/>>. Acesso em: 2 dez. 2023.

BIBA, J. **What Are eVTOLs? Are They the Future of Aviation?** Junho de 2023. Disponível em: <<https://builtin.com/transportation-tech/evtol-aircraft>>. Acesso em: 2 dez. 2023.

BOER, Arnoud V. Dynamic pricing and learning: Historical origins, current research, and new directions. *Surveys in Operations Research and Management Science*. Junho 2015, v. 20, p. 1-18. Disponível em: <<https://doi.org/10.1016/j.sorms.2015.03.001>>. Acesso em: 18 jul. 2023.

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

BRASIL. Lei nº 7.565, de 19 de dezembro de 1986. Diário Oficial da União, Brasília, DF, 23 dez. 1986. Seção 1, p. 19567.

BRASIL. Lei nº 11.182, de 27 de setembro de 2005. Diário Oficial da União, Brasília, DF, 28 set. 2005. Seção 1, p. 1.

BRASIL. Portaria nº 3.506, de 11 de novembro de 2019. Diário Oficial da União, Brasília, DF, 14 nov. 2019. Seção 1, p. 16.

BRASIL. Resolução nº 140, de 9 de março de 2010. Diário Oficial da União, Brasília, DF, 12 mar. 2010. Seção 1, p. 14

BRASIL. Resolução Nº 381, de 14 de junho de 2016. Dispões sobre o Regimento Interno da Agência Nacional de Aviação Civil – ANAC. Diário Oficial da União: seção 1, nº 113, p. 56, quarta-feira, 15 de junho de 2016. Disponível em: <<https://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?data=15/06/2016&jornal=1&pagina=56&totalArquivos=88>>. Acesso em: 25 nov. 2023.

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: 17 jul. 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: 22 abr. 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: 22 abr. 2023.

CEDRIC, Y. Justin; ALEXIA, P. Payan; DIMITRI N. Mavris. Integrated fleet assignment and scheduling for environmentally friendly electrified regional air mobility. Elsevier, Atlanta, GA, v. 138, p. 1 - 38, 19, março, 2022. Disponível em: <<https://doi.org/10.1016/j.trc.2022.103567>>. Acesso em: 17 jul. 2023.

COX, Jordan *et al.* Impacts of Regional Air Mobility and Electrified Aircraft on Airport Electricity Infrastructure and Demand. Golden, CO: National Renewable Energy Laboratory,

fevereiro 2023. Disponível em: <<https://www.nrel.gov/docs/fy23osti/84176.pdf>>. Acesso em: 17 jul. 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: 15 jul. 2023.

EVE AIR MOBILITY. Concept of Operations for Sustainable Urban Air Mobility in Rio de Janeiro. [S.l.]: abril 2022. Disponível em: <<https://eveairmobility.com/wp-content/uploads/2022/05/EVEConopsRJ.pdf>> Acesso em: 19 jul 2023.

EVE AIR MOBILITY. EVE and Voar Aviation Sign Letter of Intent to Drive Urban Air Mobility - Future Partnership to Revolutionize Urban Air Mobility in Brazil. [S.l.], 20 junho 2023. Disponível em: <<https://eveairmobility.com/eve-and-voar-aviation-sign-letter-of-intent-to-drive-urban-air-mobility/>>. Acesso em: 16 jul. 2023.

FAO. FAO framework for the Urban Food Agenda. Rome, Italy, 2019. p. 4 – 44. ISBN 978-92-5-131274-2. Disponível em: <<https://doi.org/10.4060/ca3151en>>. Acesso em: 16 jul. 2023.

FEDERAL AVIATION ADMINISTRATION - FAA. Urban Air Mobility (UAM) Concept of Operations. Version 2.0. Washington, DC: U.S. Department of Transportation, 26 abril 2023. Disponível em: <https://www.faa.gov/air-taxis/uam_blueprint>. Acesso em: 17 jul. 2023.

FERREIRA, M. Modelo de precificação de tarifa para a operacionalização do eVTOL: Um modelo baseado na Uber. Trabalho de Conclusão de Curso (TCC). 25 julho 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: 26 nov. 2023.

GROPANDAS. GeoPandas 0.13.2. [S.l.]. Disponível em: <<https://geopandas.org/en/stable/>>. Acesso em: 16 jul. 2023.

HADER, Manfred; BAUR, Stephan. Regional Air Mobility: How to unlock a New Generation of Mobility. ROLAND BERGER GMBH, BAUHAUS LUFTFAHRT E.V.: Germany, 21 junho 2022. Disponível em: <<https://www.rolandberger.com/en/Insights/Publications/Regional-Air-Mobility-How-to-unlock-a-new-era-of-aviation.html>>. Acesso em: 17 jul. 2023.

HILL, Brian P *et al.* Urban Air Mobility (UAM) Vision Concept of Operations (ConOps) UAM Maturity Level (UML)-4 Overview. [S. l.]: NASA, 2 dez. 2020. Disponível em: <<https://ntrs.nasa.gov/citations/20205011091>>. Acesso em: 17 July 2023.

HILLIER, Frederick S.; Lieberman, Gerald J. Introdução à pesquisa operacional. [S.l.]: Grupo A, 2013. E-book. ISBN 9788580551198. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9788580551198/>. Acesso em: 18 jul. 2023.

IBGE – Instituto Brasileiro de Geografia e Estatística. Censo Demográfico 2022. População e domicílios - Primeiros resultados. Rio de Janeiro: IBGE, 2023.

LILIUM. Why we're focusing on Regional Air Mobility. 24 jul. 2020. Disponível em: <https://lilium.com/newsroom-detail/why-regional-air-mobility>. Acesso em: 17 jul. 2023.

MATPLOTLIB. **Matplotlib documentation — Matplotlib 3.8.2 documentation**. Disponível em: <https://matplotlib.org/stable/>. Acesso em: 26 nov. 2023.

NASA. Advanced Air Mobility Mission Overview. nasa.gov, 2021. Disponível em: <https://www.nasa.gov/aam/overview/>. Acesso em: 17 jul. 2023.

NATHEN, Patrick *et al.* Architectural performance assessment of an electric vertical take-off and landing (e-VTOL) aircraft based on a ducted vectored thrust concept. [S. l.], 2021. Disponível em: <https://assets.evtol.com/wp-content/uploads/2021/03/Architectural-performance-assessment-of-an-e-VTOL-aircraft.pdf>. Acesso em: 17 jul. 2023.

NUMPY. **NumPy**. 16 set. 2023. Disponível em: <https://numpy.org/>. Acesso em: 26 nov. 2023.

PANDAS. **About pandas**. Disponível em: <https://pandas.pydata.org/about/index.html>. Acesso em: 26 nov. 2023.

SAHARAN, Sandeep; BAWA, Seema; KUMAR, Neeraj. Dynamic pricing techniques for Intelligent Transportation System in smart cities: A systematic review. **Computer Communications**, v. 150, p. 603-625, 2020. Disponível em: <https://doi.org/10.1016/j.comcom.2019.12.003>. Acesso em: 18 jul. 2023.

SANTOS, T. G. **Google Colab: o que é, tutorial de como usar e criar códigos**. Disponível em: <https://www.alura.com.br/artigos/google-colab-o-que-e-e-como-usar>. Acesso em: 26 nov. 2023.

SEABORN. **Seaborn: Statistical data visualization — seaborn 0.13.0 documentation**. Disponível em: <https://seaborn.pydata.org/>. Acesso em: 26 nov. 2023.

SCIPY. **SciPy 1.11.4 released!** Disponível em: <https://scipy.org/>. Acesso em: 2 dez. 2023.

TEXTRON AVIATION INC. **Cessna Caravan**. Disponível em: <https://cessna.txtav.com/en/turboprop/caravan>. Acesso em: 2 dez. 2023.

VINHOLES, Thiago. Conheça a EVE, a marca de ‘carros voadores’ da Embraer. CNN, São Paulo, 11 setembro 2021. Disponível em: <<https://www.cnnbrasil.com.br/economia/conheca-a-eve-a-marca-de-carros-voadores-da-embraer/>>. Acesso em: 18 jul. 2023.

What Is eVTOL?: Electric Vertical Take-off and Landing. SPARTAN College of Aeronautics and Technology, [S.l.], 8 fevereiro 2021. Disponível em: <<https://www.spartan.edu/news/what-is-evtol/>>. Acesso em: 17 jul. 2023.

APÊNDICE

Apêndice A – Notebook do *Google Colaboratory*: Análise Exploratória das Tarifas Aéreas praticadas no Brasil (2019 a 2022)

Análise Exploratória das Tarifas Aéreas praticadas no Brasil (2019 a 2022) ANAC - Estatísticas - Organização e Tratamento dos Dados

Versão 26/11/2023 →

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-presentacao-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 analisar e compreender as variáveis que influenciaram as tarifas dos voos tradicionais entre 2019 e 2022, e correlacionar essas percepções aos serviços de transporte por eVTOL no modelo de Malha Regional Aérea (RAM) brasileira.

Inicialização - Instalação de Pacotes Python

```
import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
import seaborn as sns
import scipy
```

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

```
%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.

Leitura do arquivo:

```
aeroportos = pd.read_csv('./Dados/Aerodromos/AerodromosPublicos.csv',
                        encoding='cp1252', delimiter = ';',skiprows=1)
```

Tamanho do arquivo:

```
aeroportos.shape
```

Visualização do arquivo:

```
aeroportos.head()
```

Variáveis do arquivo:

```
aeroportos.columns
```

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'])
```

Renomeando as variáveis do arquivo:

```
aer_coord.columns = ['OACI', 'Municipio', 'UF', 'Lat', 'Long']
```

Eliminando os dados vazios (NA):

```
aer_coord = aer_coord.dropna()
```

Salvando o arquivo "aeroportos2.csv":

```
aer_coord.to_csv('./Dados/Aerodromos/aeroportos2.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.

```

# Instalando Airportsdata:
%pip install -U airportsdata

import airportsdata

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

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

# 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).

Leitura do arquivo:

```
aerodromo = pd.read_csv('./Dados/Aerodromos/aerportos2.csv')
```

Visualização do arquivo:

```
aerodromo.head()
```

Criando uma lista com os dados da variável OACI (códigos dos aeroportos):

```
aerod = aerodromo.OACI.tolist()
```

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)
```

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)
```

Visualizando o dataframe distancias:

```
distancias.head()
```

Salvando o arquivo "distancias.csv":

```
distancias.to_csv('./Dados/Tarifas/distancias.csv', index=False)
```



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

Visualização do dataframe:

```
df.head()
```

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.

Leitura do arquivo:

```
dados_tarifas = pd.read_csv('./Dados/Tarifas/tarifas_2019.csv')
```

Tamanho do arquivo:

```
dados_tarifas.shape
```

Visualização do arquivo:

```
dados_tarifas.head()
```

Padronização dos dados respectivos à variável MES:

```
dados_tarifas.MES.unique()
```

Substituindo ',' por '.' na variável TARIFA:

```
dados_tarifas.TARIFA = dados_tarifas.TARIFA.str.replace(',','.')

```

Alterando o tipo de dado da variável TARIFA para float:

```
dados_tarifas.TARIFA = dados_tarifas.TARIFA.astype(float)
```

Salvando o arquivo "tarifas9.csv":

```
dados_tarifas.to_csv('./Dados/Tarifas/tarifas9.csv', index=False)
```

Etapa II. Tarifas - Transformação de Dados 2020

Etapa II.I. Leitura e agrupamento de arquivos de Tarifas 2020

Salvo em ./Dados/Tarifas/tarifas_2020.csv

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

Leitura do arquivo:

```
files = ['202001.CSV', '202002.CSV', '202003.CSV', '202004.CSV',
         '202005.CSV', '202006.CSV', '202007.CSV', '202008.CSV',
         '202009.CSV', '202010.CSV', '202011.CSV', '202012.CSV']
```

Estruturando o dataframe:

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

Agrupando os arquivos no dataframe estruturado:

for file in files:


```

# 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)

# Visualização do dataframe:
df.head()

# 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.

```

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

# Tamanho do arquivo:
dados_tarifas.shape

# Visualização do arquivo:
dados_tarifas.head()

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

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

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

# 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.

```

# 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']

```

```

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

# 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)

# Visualização do dataframe:
df.head()

# 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.

```

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

# Tamanho do arquivo:
dados_tarifas.shape

# Visualização do arquivo:
dados_tarifas.head()

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

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

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

# 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>. 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.

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']
```

Criação do dataframe:

```
df = pd.DataFrame()
```

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)
```

Visualização do dataframe:

```
df.head()
```

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.

Leitura do arquivo:

```
dados_estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas_2019.csv',
                                low_memory=False)
```

```

# Dimensões do arquivo:
dados_estatisticas.shape

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

dados_estatisticas.columns

# 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.

dados_estatisticas.info()

dados_estatisticas.isna().sum()

# 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']
)

# Dimensão do dataframe:
dados_estatisticas.shape

# 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.

```

# 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']

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

# 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)

# Visualização do dataframe:
df.head()

# 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.

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

# Dimensões do arquivo:
dados_estatisticas.shape

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

dados_estatisticas.columns

# 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')]

dados_estatisticas.info()

dados_estatisticas.isna().sum()

# 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']
)
```

```
# Dimensão do dataframe:
```

```
dados_estatisticas.shape
```

```
# 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.

```
# 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']
```

```
# Criação do dataframe:
```

```
df = pd.DataFrame()
```

```
# 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)
```

```
# Visualização do dataframe:
```

```
df.head()
```

```
# 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.

```
# Leitura do arquivo:
```

```
dados_estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas_2021.csv',
                                 low_memory=False)
```

```

# Dimensões do arquivo:
dados_estatisticas.shape

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

dados_estatisticas.columns

# 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')]

dados_estatisticas.info()

dados_estatisticas.isna().sum()

# 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']
)

# Dimensão do dataframe:
dados_estatisticas.shape

# 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.

```

# 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']

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

# 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)

# Visualização do dataframe:
df.head()

# 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.

```

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

# Dimensões do arquivo:
dados_estatisticas.shape

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

dados_estatisticas.columns

# 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')]

dados_estatisticas.info()

dados_estatisticas.isna().sum()

# 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']
)
```

```
# Dimensão do dataframe:
```

```
dados_estatisticas.shape
```

```
# 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.

```
# 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']
```

```
# Criando o dataframe:
```

```
df = pd.DataFrame()
```

```
# 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)

# Visualizando o dataframe:
df.head()

# 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.

```

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

```

```

# Dimensão do arquivo:
dados_estatisticas.shape

```

```

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

```

```

dados_estatisticas.columns

```

```

# 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.

```

```

dados_estatisticas.info()

```

```

dados_estatisticas.isna().sum()

```

```

# 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']
)
```

```
# Dimensão do dataframe:
```

```
dados_estatisticas.shape
```

```
# 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.

```
# Leitura dos arquivos:
```

```
estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas9c.csv')
```

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

```
# Criando um dicionário com as correspondências entre origem e destino, com as distancias (km):
```

```
correspondencias = distancia.set_index(['orig', 'dest'])['dist'].to_dict()
```

```
# 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)
```

```
# Visualizando o dataframe:
```

```
estatisticas.head()
```

```
# 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.

Leitura dos arquivos:

```
estatisticasc = pd.read_csv("./Dados/DemandaeOferta/estatisticas9c.csv")
estatisticasb = pd.read_csv("./Dados/DemandaeOferta/estatisticas9.csv")
```

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)]
```

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]
```

Identificando os números de voos com as respectivas aeronaves de interesse:

```
nr_voos = estatisticasb_filtrado['nr_voo'].unique()
print(nr_voos)
```

Filtrando os números de voos encontrados das estatísticas combinadas:

```
estatisticasc_filtrado = estatisticasc[estatisticasc['nr_voo'].isin(nr_voos)]
```

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

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()
```

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)
```

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()
```

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)
```

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()
```

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)

estatisticasc_filtrado.head()

estatisticasc_filtrado.shape

# 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.

```

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

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

#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)

estatisticas.head()

```

Adequação do arquivo tarifas9.csv

Descrição: Após a criação da variável semestre, foi necessária a criação da variável tarifa ponderada, que 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.

```

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

```

```
#Criando a coluna SEMESTRE
```

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

```
#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
```

```
# 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)
```

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

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

```
tarifas.head()
```

```
# 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()
```

```
# 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)
```

```
estatisticas.head()
```

```
#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))
```

```
# 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)
```

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

```
tarifas.head()
```

```
# 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()
```

```
# 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)
```

```
estatisticas.head()
```

```
# 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.

```
# 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']
```

```
# Criando o dataframe:
```

```
df = pd.DataFrame()
```

```
# 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)
```

```
# Visualizando o dataframe:
```

```
df.head()
```

```
# 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.

```
# Leitura do arquivo:
```

```
dados_estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas_2020_combinadas.csv',
low_memory=False)
```

```
# Dimensão do arquivo:
```

```
dados_estatisticas.shape
```

```
# Padronizando os dados respectivos ao mês:
```

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



```
dados_estatisticas.columns
```

```
# 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.
```

```
dados_estatisticas.info()
```

```
# Exclusão dos dados vazios (NA)
```

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

```
# 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']
)
```

```
# Dimensão do dataframe:
```

```
dados_estatisticas.shape
```

```
# 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.

```
# Leitura dos arquivos:
```

```
estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas0c.csv')
```

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

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

```
# 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)
```

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

```
# 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 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.

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

```
# 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)]
```

```
# 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]
```

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

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

```
# 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)]
```

```

estatisticasc_filtrado['sg_empresa_icao'].unique()

# 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()

# 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)

# 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()

# 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)

# 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()

# 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)

estatisticasc_filtrado.head()

estatisticasc_filtrado.shape

# 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.

```

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

```

```

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

#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)

estatisticas.head()

```

Adequação do arquivo tarifas0.csv

Descrição: Após a criação da variável semestre, foi necessária a criação da variável tarifa ponderada, que 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.

Leitura do arquivo:

```
tarifas = pd.read_csv('./Dados/Tarifas/tarifas0.csv')
```

#Criando a coluna SEMESTRE

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

#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

```

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)
```

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

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

```
tarifas.head()
```

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()
```

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)

estatisticas.head()

#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))

# 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)

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

tarifas.head()

# 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()

# 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)

estatisticas.head()

# 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.

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']

```

Criando o dataframe:

```

df = pd.DataFrame()

```

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)
```

Visualizando o dataframe:

```
df.head()
```

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.

Leitura do arquivo:

```
dados_estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas_2021_combinadas.csv',
                                low_memory=False)
```

Dimensão do arquivo:

```
dados_estatisticas.shape
```

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

```
dados_estatisticas.columns
```

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.

```
dados_estatisticas.info()
```

Exclusão dos dados vazios do dataframe (NA)

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

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']
)

# Dimensão do dataframe:
dados_estatisticas.shape

# 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.

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

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

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

# 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)

# Visualizando o dataframe:
estatisticas.head()

# 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 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.

Leitura dos arquivos:

```
estatisticasc = pd.read_csv("./Dados/DemandaeOferta/estatisticasc.csv")
```

```
estatisticasb = pd.read_csv("./Dados/DemandaeOferta/estatisticasb.csv")
```

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)]
```

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]
```

Identificando os números de voos com as respectivas aeronaves de interesse:

```
nr_voos = estatisticasb_filtrado['nr_voo'].unique()
```

```
print(nr_voos)
```

Filtrando os números de voos encontrados das estatísticas combinadas:

```
estatisticasc_filtrado = estatisticasc[estatisticasc['nr_voo'].isin(nr_voos)]
```

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)]
```

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

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()
```

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)
```

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()
```

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)
```

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()
```



```

# 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)

estatisticasc_filtrado.head()

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

# Salvando o arquivo "estatisticasc1c_filtered.csv":
estatisticasc_filtrado.to_csv('./Dados/DemandaeOferta/estatisticasc1c_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.

```

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

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

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

estatisticasc.head()

```

Adequação do arquivo tarifas1.csv

Descrição: Após a criação da variável semestre, foi necessária a criação da variável tarifa ponderada, que 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.

```

# Leitura do arquivo:
tarifasc = pd.read_csv('./Dados/Tarifas/tarifasc1.csv')

```

```
#Criando a coluna SEMESTRE
```

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

```
#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
```

```
# 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)
```

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

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

```
tarifas.head()
```

```
# 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()
```

```
# 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)
```

```
estatisticas.head()
```

```
#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))
```

```
# 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)
```

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

```
tarifas.head()
```

```
# 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()
```

```
# 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)
```

```
estatisticas.head()
```

```
# Salvando o arquivo "estatisticas1c_filtered.csv" com as novas variáveis:
```

```
estatisticas.to_csv('./Dados/DemandaeOferta/estatisticas1c_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.

```
# 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']
```

```
# Criando o dataframe:
```

```
df = pd.DataFrame()
```

```
# 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)
```

```
# Visualizando o dataframe:
```

```
df.head()
```

```
# 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.

```
# Leitura do arquivo:
```

```
dados_estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas_2022_combinadas.csv',
low_memory=False)
```

```
# Dimensão do arquivo:
```

```
dados_estatisticas.shape
```

```
# Padronizando os dados respectivos ao mês:
```

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

```
dados_estatisticas.columns
```

```
# 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.
```

```
dados_estatisticas.info()
```

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

```
# 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']
)
```

```
# Dimensão do dataframe:
```

```
dados_estatisticas.shape
```

```
# 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.

```
# Leitura dos arquivos:
```

```
estatisticas = pd.read_csv('./Dados/DemandaeOferta/estatisticas2c.csv')
```

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

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

```
# 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)
```

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

```
# 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 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.

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

```
# 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)]
```

```
# 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]
```

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

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

```
# 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)]
```

```

estatisticasc_filtrado['sg_empresa_icao'].unique()

# 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()

# 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)

# 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()

# 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)

# 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()

# 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)

estatisticasc_filtrado.head()

estatisticasc_filtrado.shape

# 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.

```

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

```

```

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

#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)

estatisticas.head()

```

Adequação do arquivo tarifas2.csv

Descrição: Após a criação da variável semestre, foi necessária a criação da variável tarifa ponderada, que 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.

```

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

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

#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

# 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)

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

tarifas.head()

# 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()

# 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)

estatisticas.head()

#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))

# 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)

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

tarifas.head()

# 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()

# 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)

estatisticas.head()

# 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

```
# 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")

# 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"])

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

estatisticas0['ano'] = 2020

estatisticas1['ano'] = 2021

estatisticas2['ano'] = 2022

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

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

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

estatisticas.to_csv('./Dados/DemandaeOferta/estatisticasc_filtered.csv', index=False)
estatisticas2.to_csv('./Dados/DemandaeOferta/estatisticas2c_filtered.csv', index=False)
estatisticas1.to_csv('./Dados/DemandaeOferta/estatisticas1c_filtered.csv', index=False)
estatisticas0.to_csv('./Dados/DemandaeOferta/estatisticas0c_filtered.csv', index=False)
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

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

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

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

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

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

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

# 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

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

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

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

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

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

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

# 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.

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.

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

```
# Importando o 'stats' para complementar o cálculo da regressão linear:
from scipy import stats
```

Análises - Aeronaves: TODAS

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

Análise de Regressão Linear: Distância (km_distancia) vs Tarifas (tarifa_ponderada):

```
for ano in [2019, 2020, 2021, 2022]:
```

```
    df_ano = estatisticas[(estatisticas['ano'] == ano)]
    x = df_ano['km_distancia']
    y = df_ano['tarifa_ponderada']
```

```
# Calculando a regressão linear
slope, intercept, r_value, p_value, std_err = stats.linregress(x, y)
```

```
# Criando o gráfico
plt.figure(figsize=(8, 6))
sns.regplot(x=x, y=y)
```

```
# Adicionando a equação da reta ao gráfico
plt.text(0.1, 0.9, f'y = {slope:.2f}x + {intercept:.2f}', transform=plt.gca().transAxes)
```

```
# Adicionando o coeficiente de correlação de Pearson ao gráfico
plt.text(0.1, 0.8, f'r = {r_value:.2f}', transform=plt.gca().transAxes)
```

```
plt.xlabel('Distancia (km)')
plt.ylabel('Tarifa')
plt.title(f'Distância vs Tarifa Ponderada - Ano: {ano} - Aeronave: TODAS')
plt.show()
```

Análise de Regressão Linear: Distância (km_distancia) vs Tarifas (tarifa_ponderada), por empresa:

```
# Personalizando as cores da legenda do gráfico:
```

```
cores = ['b', 'g', 'r', 'c', 'm', 'y', 'k']
mapa_cores = {'PTB': cores[0], 'AZU': cores[1], 'ACN': cores[2], 'PAM': cores[3], 'ABJ': cores[4]}
```

```
for ano in [2019, 2020, 2021, 2022]:
```

```
    df_ano = estatisticas[(estatisticas['ano'] == ano)]
```

```

# Criando uma nova figura:
plt.figure(figsize=(10, 8))

# Obtendo a lista de empresas aéreas:
empresas = df_ano['sg_empresa_icao'].unique()

for empresa in empresas:
    # Filtrando o dataframe para a empresa atual
    df_empresa = df_ano[df_ano['sg_empresa_icao'] == empresa]
    x = df_empresa['km_distancia']
    y = df_empresa['tarifa_ponderada']

    # Calcula a regressão linear
    slope, intercept, r_value, p_value, std_err = stats.linregress(x, y)

    # Cria o gráfico de dispersão e a linha de regressão
    sns.regplot(x=x, y=y, color=mapa_cores[empresa], label=f'{empresa}: y = {slope:.2f}x +
    {intercept:.2f}, r = {r_value:.2f}')

# Adiciona a legenda
plt.legend()

plt.xlabel('Distancia (km)')
plt.ylabel('Tarifa')
plt.title(f'Distância(km) vs Tarifa Ponderada - Ano: {ano} - Aeronave: TODAS')
plt.show()

```

Gráfico de Dispersão 3D - Análise de Relação: Distância(km_distancia) vs Número de Assentos (nr_assentos_ofertados) vs Tarifas (tarifa_ponderada):

```
from mpl_toolkits.mplot3d import Axes3D
```

```
fig = plt.figure(figsize=(10, 10))
ax = fig.add_subplot(111, projection='3d')
```

```
x = df_ano['nr_assentos_ofertados']
y = df_ano['km_distancia']
z = df_ano['tarifa_ponderada']
```

```
ax.scatter(x, y, z)
```

```
ax.set_xlabel('Número de Assentos Ofertados')
ax.set_ylabel('Distância(km)')
ax.set_zlabel('Tarifa')
```

```
plt.title(f'Relação: Nº Assentos Ofertados vs Distância(km) vs Tarifa (2019 a 2022) - Aeronave =
TODAS')
plt.show()
```

Lineplot: Tarifa Ponderada Média por Mês (2019 a 2022):

```
for ano in [2019, 2020, 2021, 2022]:
    df_ano = estatisticas[(estatisticas['ano'] == ano)]
```

```
# Calculando a média mensal
```

```

media_mensal = df_ano.groupby('nr_mes_partida_real')['tarifa_ponderada'].mean()

sns.lineplot(x=media_mensal.index, y=media_mensal.values, label=ano)

plt.xticks(np.arange(1, 13), ['Jan', 'Fev', 'Mar', 'Abr', 'Mai', 'Jun', 'Jul', 'Ago', 'Set', 'Out', 'Nov', 'Dez'])
plt.xlabel('Mês')
plt.ylabel('Tarifa Média')
plt.title('Tarifa Ponderada Média por Mês')
plt.legend()
plt.show()

```

Análises - Aeronave: C208

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

Análise de Regressão Linear: Distância (km_distancia) vs Tarifas (tarifa_ponderada):

for ano **in** [2019, 2020, 2021, 2022]:

```

df_ano = estatisticas[(estatisticas['ano'] == ano) & (estatisticas['sg_equipamento_icao'] == 'C208')]
x = df_ano['km_distancia']
y = df_ano['tarifa_ponderada']

```

Calculando a regressão linear

```
slope, intercept, r_value, p_value, std_err = stats.linregress(x, y)
```

Criando o gráfico

```

plt.figure(figsize=(8, 6))
sns.regplot(x=x, y=y)

```

Adicionando a equação da reta ao gráfico

```
plt.text(0.1, 0.9, f'y = {slope:.2f}x + {intercept:.2f}', transform=plt.gca().transAxes)
```

Adicionando o coeficiente de correlação de Pearson ao gráfico

```
plt.text(0.1, 0.8, f'r = {r_value:.2f}', transform=plt.gca().transAxes)
```

```
plt.xlabel('Distancia (km)')
```

```
plt.ylabel('Tarifa')
```

```
plt.title(f'Distância vs Tarifa Ponderada - Ano: {ano} - Aeronave: C208')
```

```
plt.show()
```

Análise de Regressão Linear: Distância (km_distancia) vs Tarifas (tarifa_ponderada), por empresa:

Personalizando as cores da legenda do gráfico:

```
cores = ['b', 'g', 'r', 'c', 'm', 'y', 'k']
```

```
mapa_cores = {'PTB': cores[0], 'AZU': cores[1], 'ACN': cores[2], 'PAM': cores[3], 'ABJ': cores[4]}
```

for ano **in** [2019, 2020, 2021, 2022]:

```
df_ano = estatisticas[(estatisticas['ano'] == ano) & (estatisticas['sg_equipamento_icao'] == 'C208')]
```

Criando uma nova figura:

```
plt.figure(figsize=(10, 8))
```

Obtendo a lista de empresas aéreas:

```
empresas = df_ano['sg_empresa_icao'].unique()
```

```

for empresa in empresas:
    # Filtrando o dataframe para a empresa atual
    df_empresa = df_ano[df_ano['sg_empresa_icao'] == empresa]
    x = df_empresa['km_distancia']
    y = df_empresa['tarifa_ponderada']

    # Calcula a regressão linear
    slope, intercept, r_value, p_value, std_err = stats.linregress(x, y)

    # Cria o gráfico de dispersão e a linha de regressão
    sns.regplot(x=x, y=y, color=mapa_cores[empresa], label=f'{empresa}: y = {slope:.2f}x +
    {intercept:.2f}, r = {r_value:.2f}')

    # Adiciona a legenda
    plt.legend()

    plt.xlabel('Distancia (km)')
    plt.ylabel('Tarifa')
    plt.title(f'Distância(km) vs Tarifa Ponderada - Ano: {ano} - Aeronave: C208')
    plt.show()

```

Lineplot: Tarifa Ponderada Média por Mês (2019 a 2022):

```

for ano in [2019, 2020, 2021, 2022]:
    df_ano = estatisticas[(estatisticas['ano'] == ano) & (estatisticas['sg_equipamento_icao'] == 'C208')]

    # Calculando a média mensal
    media_mensal = df_ano.groupby('nr_mes_partida_real')['tarifa_ponderada'].mean()

    sns.lineplot(x=media_mensal.index, y=media_mensal.values, label=ano)

    plt.xticks(np.arange(1, 13), ['Jan', 'Fev', 'Mar', 'Abr', 'Mai', 'Jun', 'Jul', 'Ago', 'Set', 'Out', 'Nov', 'Dez'])
    plt.xlabel('Mês')
    plt.ylabel('Tarifa Média')
    plt.title('Tarifa Ponderada Média por Mês - Aeronave: C208')
    plt.legend()
    plt.show()

```

Identificando Outliers (2019 - 2022)

Outliers da variável Tarifa:

```

# Identificando outliers de estatisticas

# Calculando Q1 and Q3:
Q1 = estatisticas['tarifa_ponderada'].quantile(0.25)
Q3 = estatisticas['tarifa_ponderada'].quantile(0.75)

# Faixa interquartil:
IQR = Q3 - Q1

# Encontrando os outliers
outliers = estatisticas[(estatisticas['tarifa_ponderada'] < (Q1 - 1.5 * IQR)) |

```

```
(estatisticas['tarifa_ponderada'] > (Q3 + 1.5 * IQR))]
```

```
outliers.shape
```

```
sns.boxplot(x = 'ano', y = 'tarifa_ponderada', data = estatisticas)
sns.stripplot(x = 'ano', y = 'tarifa_ponderada', data = estatisticas,
              jitter=True,
              marker='o',
              alpha=0.8,
              color="black")
plt.title('Boxplot of Tarifas by Ano')
plt.xticks(rotation=90) # Rotating x-axis labels for better readability
plt.show()
```

Outliers da variável Distância(km):

```
# Identificando outliers de estatisticas
```

```
# Calculando Q1 and Q3:
```

```
Q1 = estatisticas['km_distancia'].quantile(0.25)
```

```
Q3 = estatisticas['km_distancia'].quantile(0.75)
```

```
# Faixa interquartil:
```

```
IQR = Q3 - Q1
```

```
# Encontrando os outliers
```

```
outliers = estatisticas[(estatisticas['km_distancia'] < (Q1 - 1.5 * IQR)) | (estatisticas['km_distancia'] >
(Q3 + 1.5 * IQR))]
```

```
outliers.shape
```

```
sns.boxplot(x = 'ano', y = 'km_distancia', data = estatisticas)
sns.stripplot(x = 'ano', y = 'km_distancia', data = estatisticas,
              jitter=True,
              marker='o',
              alpha=0.8,
              color="black")
plt.title('Boxplot of Tarifas by Ano')
plt.xticks(rotation=90) # Rotating x-axis labels for better readability
plt.show()
```