**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS**

**PUC-GOIÁS**

**Ciência de Dados e Inteligência Artificial**

**RELATÓRIO DESAFIO VI CIÊNCIAS DE DADOS**

**Alunos**

Anna Carolina Figuccio Alves

Daniel Oliveira Araujo

Marcus Pablo Rezende Teixeira Manso

**Goiânia**

**2024**

Anna Carolina Figuccio Alves

Daniel Oliveira Araujo

Marcus Pablo Rezende Teixeira Manso

**RELATÓRIO DESAFIO VI CIÊNCIAS DE DADOS**

Relatório apresentado ao curso de Ciências de Dados e Inteligência Artificial, como parte dos requisitos necessários à conclusão do desafio VI Ciências de Dados, Segurança na Aviação.

Orientadora: Maria Jose Pereira Santos

**Goiânia**

**2024**

**Resumo**

Ao decorrer da última década o Brasil tem passado por vários acidentes aéreos, sendo relatado 7567acidentes entre 2010 e 2021. O seguinte relatório irá trazer a resolução do desafio proposto aos alunos da Pontifica Universidade Católica de Goiás do curso de Ciência de Dados e Inteligência Artificial. O desafio tem como finalidade mostrar os acidentes e incidentes ocorridos no período, características das aeronaves envolvidas, os climas dos locais dos episódios e outras informações, onde possam associar os dados para tentar gerar uma relação entre as ocorrências e outros fatores externos. Para a conclusão do trabalho os alunos vão destacar possíveis relações dos acontecimentos com outros fatores, utilizar gráficos para cálculos de probabilidade e associação dos gráficos entre si para observar se há coerência a probabilidade aumentar ou diminuir de acordo com a relação, processos de machine learning, e modelagem de dados, tudo a partir de programas feitos na linguagem Python criado pelos alunos.

Palavras-Chaves: Acidentes aéreos, Python, Machine Learning, Modelagem de Dados

**Abstract**

Over the last decade, Brazil has experienced several aviation accidents, with 7567 accidents reported between 2010 and 2021. The following report will resolve the challenge proposed to students at the Pontifica Universidade Católica de Goiás in the Data Science and Artificial Intelligence course. The challenge aims to show the accidents and incidents that occurred during the period, characteristics of the aircraft involved, the climate of the locations of the episodes and other information, where they can associate the data to try to generate a relationship between the occurrences and other external factors. To complete the work, students will highlight possible relationships between events and other factors, use graphs to calculate probability and associate the graphs with each other to observe whether there is coherence and the probability increases or decreases according to the relationship, machine learning processes, and data modeling, all from programs made in the Python language created by the students.

Keywords: Air accidents, Python, Machine Learning, Data Modeling**Lista de Abreviaturas e Siglas**

CENIPA - Centro de Investigação e Prevenção de Acidentes Aeronáuticos do Brasil

INMET - Instituto Nacional de Meteorologia

OACI – Organização de Aviação Civil Internacional

**Lista de Gráficos e Tabelas**

**Figuras**

* Figura 3.1 Planilha apresentado após tratamento de dados
* Figura 3.2 Mapa das Ocorrências dos Acidentes Aéreos
* Figura 3.3 Cidades com Maior Ocorrências
* Figura 3.4 Histograma de Ocorrências x Estados
* Figura 3.5 Histograma Fatalidades x Estados
* Figura 3.6 Gráfico de pizza com distribuição dos tipos de ocorrências
* Figura 3.7 Histograma de Fatalidades x Fase de Operação
* Figura 3.8 Gráfico de Dispersão de Fatalidades x Número de Assentos
* Figura 3.9 Distribuição de Acidentes Aéreos por ano

**Sumario**

**1. INTRODUÇÃO.....................................................................................7**

**2. DESENVOLVIMENTO.........................................................................8**

* **2.1 OBTENÇÃO DOS DADOS...................................................8**
* **2.2 TRATAMENTO DOS DADOS..............................................8**
* **2.3 CÓDIGO** **REALIZADO..........................................................8**

**3. CONCLUSÃO....................................................................................10**

#### **4. REFERÊNCIAS.................................................................................19**

**1. Introdução**

O domínio do céu através da construção e aperfeiçoamento dos aviões foi um dos grandes avanços para a humanidade, trazendo uma grande evolução em relação a viagens, cortando o tempo necessário para ir de um local para outro.

O avião foi criado no início do século XX, iniciou-se com os irmãos Wright, Wilbur e Orville, situados nos Estados Unidos. Começaram com os experimentos dos irmãos, onde suas investigações puderam contribuir para o desenvolvimento do primeiro avião. Em 23 de outubro de 1906, brasileiro Alberto Santos Drummond, conseguiu voar 220m com sua nova aeronave 14-Bis, e logo em seguida, 1908, ele se uniu com Clement-Bayard para desenvolver o primeiro avião produzido em massa a Demoiselle 19.

Com o avanço tecnológico a aviação passou por uma série de mudanças, adaptando a nova era digital, com os computadores desenvolvendo aeronaves ainda mais complexas e ainda mais seguras, criou-se um dos métodos mais seguros da viagem contemporânea, o avião comercial.

Apesar dos métodos de segurança das aeronaves, e todas as precauções tomadas pelos pilotos e a OACI, ainda há a possibilidade de ocorrer um acidente aéreo. No Brasil durante o período de 2010-2021 ocorreu 7567 acidentes aéreos de acordo com os relatos do CENIPA, estendendo-se por volta de todo o país. Sob esse prisma, foi lançando um desafio aos integrantes da Universidade, PUC-GO, com o intuito de tentar determinar as ocorrências dos incidentes.

O desafio, proposto pela professora de probabilidade e estatística, Maria Jose Pereira Santos, foi para seus alunos do curso de Ciências de Dados e Inteligência Artificial, onde tem como objetivo testar os conhecimentos de seus alunos, e introduzir eles a trabalhos com modelagem de dados e banco de dados com probabilidade.

**2. Desenvolvimento**

A partir dos dados distribuídos pelo CENIPA a respeito dos acidentes aéreos, os dados a respeito dos climas distribuídos pelo INMET e pesquisas feito pelo Google para a determinação da latitude e longitude das cidades onde os incidentes ocorreram, foi possível o cumprimento do desafio elaborado.

**2.1 Obtenção dos Dados**

Em primeiro plano, foi utilizado a plataforma Excel para o agrupamento dos dados dos acasos, onde pelo CENIPA foi distribuído a grande maioria dos fatos com os incidentes, sendo eles, número de assentos, ano de fabricação da aeronave, número de mortes, cidade onde ocorreu, dentre outros. Apesar dos dados recebidos, foi necessário criar outras planilhas para ser utilizado no código, trazendo novas informações como número de mortos por estado e uma planilha apenas com a latitude e longitude média de cada cidade devido a erros que deparamos com os dados do CENIPA relacionados as suas coordenadas.

**2.2 Tratamento dos Dados**

Em segundo plano, foi importe a limpeza dos dados para a utilização deles em qualquer trabalho, então a partir de trabalho manual e comandos utilizados pelo Python, os dados foram tratados. O trabalho manual consistiu em olhar os dados e observar se havia algum erro neles, foi encontrado erro na parte das coordenadas, onde foi necessário criar a planilha com a latitude e longitude separado devido a todos os números distribuídos na primeira planilha estarem sem as casas decimais, resultando em coordenadas fora do alcance da Terra. Também teve erro por parte dos números de assentos e ano de fabricação da aeronave, em que seu erro foi de ter um 0 a mais em todos seus números, todavia para a resolução do problema, retirou-se o 0 de todos os dados de forma manual para o conserto de seus números. Através a biblioteca Pandas do Python foi utilizado o comando Dropna() para retirar as linhas em que todos os dados não foram distribuídos, o comando procura em nosso Dataset (planilha do Excel importado para o trabalho) se há alguma linha com ausência de dados para retirar a linha inteira, pois a ausência de dados pode interferir com o código e dar algum erro indesejado.

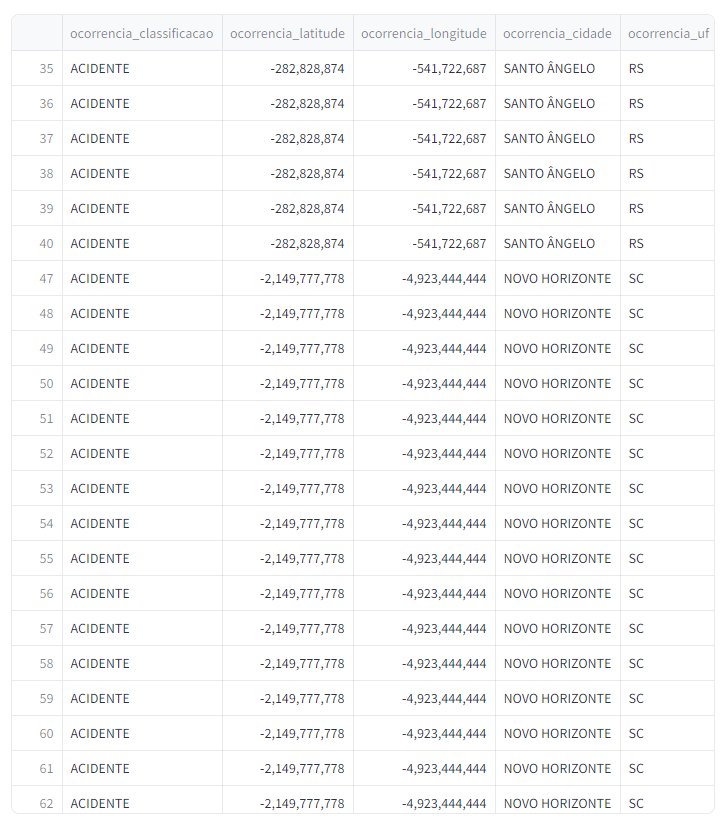
**2.3 Código realizado**

Sendo assim, com os dados nas planilhas do Excel foi criado um programa na linguagem Python a partir do Visual Studio Code com o intuito de gerar um modelo interativo, para isso foi utilizado a biblioteca StreamLit, onde sua função é criar uma página na internet com nossos gráficos, mapas e listas com os dados importados do Excel através da biblioteca Pandas, citado anteriormente junto com a ajuda do Pathlib. Após importação dos dados e criação da página, utilizou-se Matplotlib.pyplot, Seaborn e Plotly.express, cuja sua função foi a geração de gráficos como de dispersão, histograma, de linhas e de barras. Além dos gráficos foi criado um mapa interativo, em que mostra todas as cidades com as ocorrências dos acidentes e incidentes, para este feito foi usado Folium e Folium.map. Finalmente para criação de operações matemáticas e Machine Learning, como média e regressão linear foi usado as bibliotecas Numpy e Scikit-learn.

**3. Conclusão**

Através do código realizado, o primeiro dado mostrado será o gráfico tratado dos acidentes do CENIPA, onde mostrara todos os dados utilizados no código. (Figura 3.1)

Figura 3.1 Planilha apresentado após tratamento de dados



Fonte: Autores da pesquisa, 2024

Logo após a ser introduzido aos dados, foi apresentado um mapa, cuja sua funcionalidadeé ser interativo para poder dar zoom em regiões do Brasil para melhor entendimento das localidades dos acidentes. (Figura 3.2)

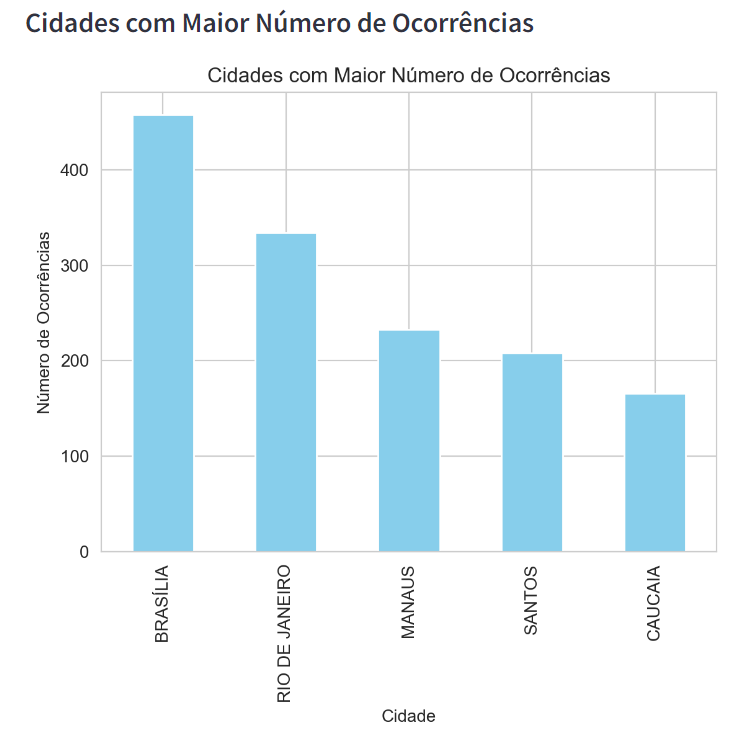
Figura 3.2 Mapa das Ocorrências dos Acidentes Aéreos



Fonte: Autores da pesquisa, 2024

Após análise do mapa pode-se observar que maior parte dos acidentes ocorrem nas regiões Sul e Sudeste, locais aonde o fluxo de aviação é maior relacionado ao resto do país. Então para melhor compreendimento o próximo gráfico apresentado é um de barras. (Figura 3.3)

Figura 3.3 Cidades com Maior Ocorrências



Fonte: Autores da pesquisa, 2024

Pode-se observar que, a cidade líder em ocorrências é Brasília, isso ocorre devido ao Distrito Federal não possuir diversas cidades com ocorrências, centralizando-se todas suas ocorrências a uma única cidade, Brasília.

A relação relatada pode ser confusa, porém com o próximo gráfico, mostrara que a região Sudeste ainda é líder em acidentes, e Brasília possui maior número de ocorrências por causa da ausência de cidades com aeroportos. (Figura 3.4)

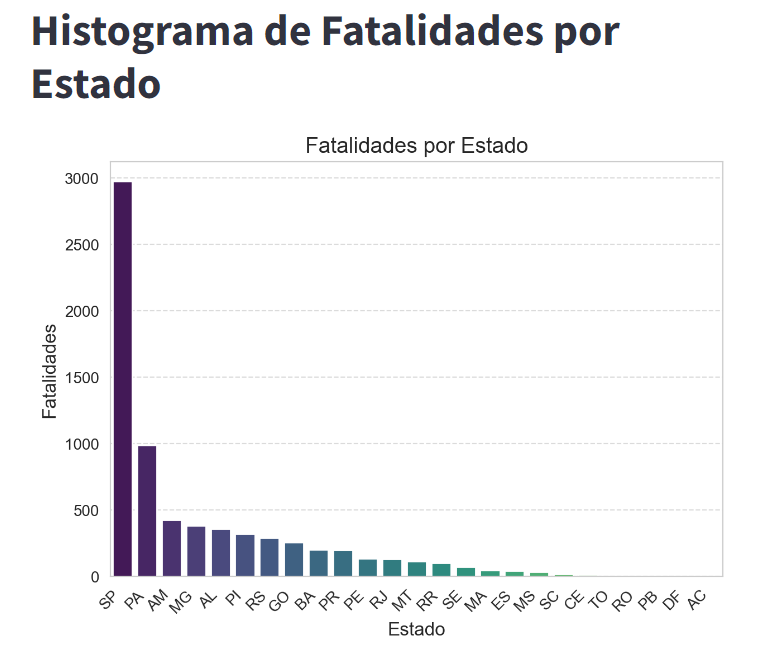
Figura 3.4 Histograma de Ocorrências x Estados



Fonte: Autores da pesquisa, 2024

Através do histograma, pode se observar que há um grande volume de ocorrências no estado de São Paulo, o estado com maior número de influxo de aviação no Brasil, porém, isso traz a pergunta se essa sequência de estados tem relação com as fatalidades ocorridas, no entanto, segue-se o histograma de fatalidades x estados. (Figura 3.5)

Figura 3.5 Histograma Fatalidades x Estados

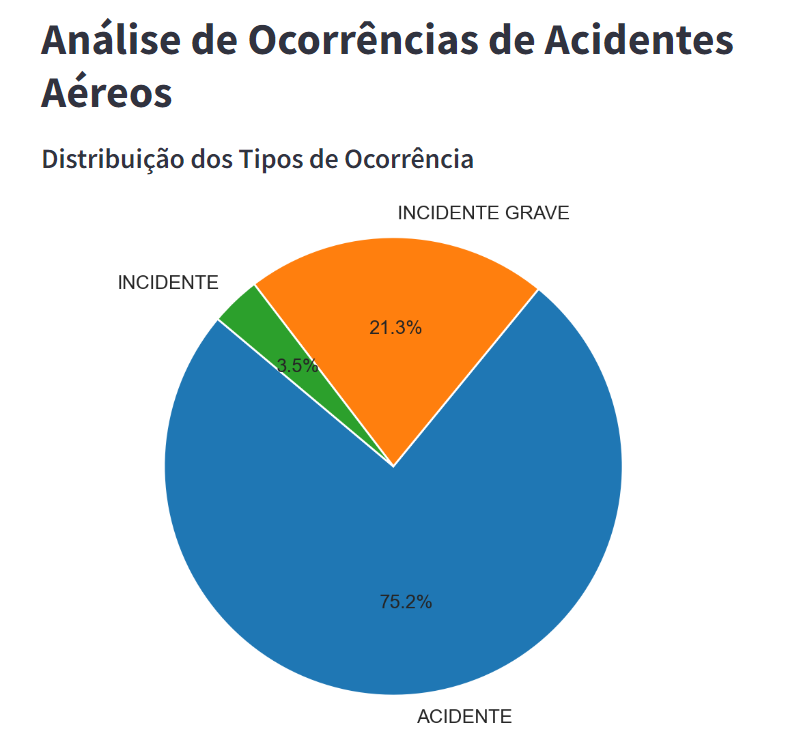


Fonte: Autores da pesquisa, 2024

O gráfico tem uma pequena relação ao anterior, mostra que o estado com maior número de fatalidades foi o de São Paulo, também líder de ocorrências, porém o líder de ocorrências por cidades, Brasília, localizado em Distrito Federal, não houve fatalidades, isso ocorre pois no Distrito Federal, suas ocorrências consistem na maioria como Acidentes e não Incidentes e Incidentes Graves em que houve fatalidades.

O próximo gráfico irá relatar as porcentagens das ocorrências, pegando o número de casos específicos (Acidentes, Incidentes ou Incidentes Graves) e dividindo ele sobre o total de ocorrências, gerando as porcentagens necessárias para o gráfico de pizza a seguir. (Figura 3.6)

Figura 3.6 Gráfico de pizza com distribuição dos tipos de ocorrências

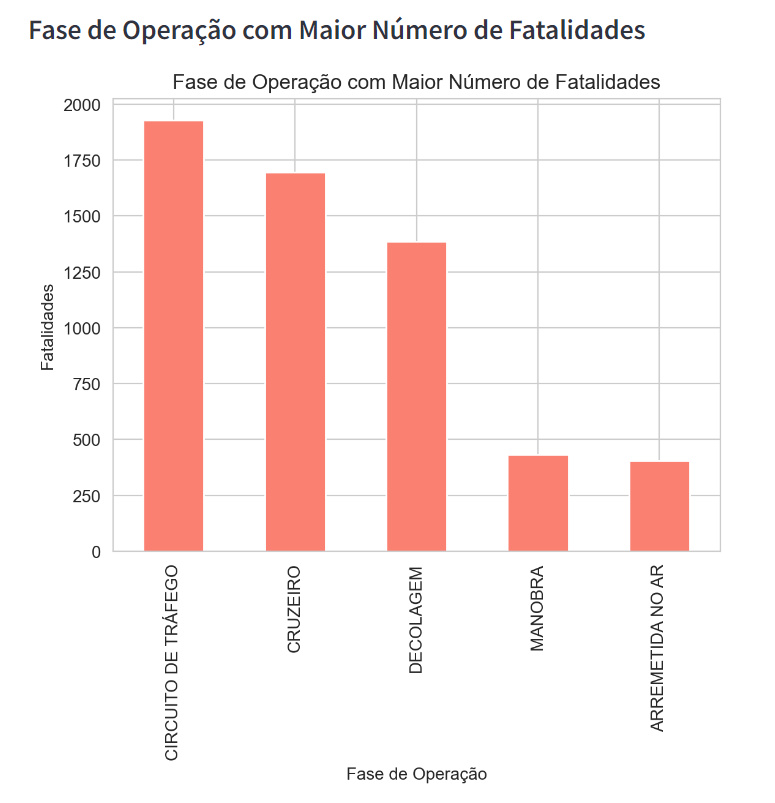


Fonte: Autores da pesquisa, 2024

O código logo após a amostra do gráfico de pizza relata que a ocorrência mais comum em que se ocorreu foi PERDA DE CONTROLE EM VOO, com 1432 ocorrências tendo perda de controle como raiz do problema.

Em seguida relata-se o período da fase de operação em houve maior número de fatalidades para tentar determinar o ponto em que há maior risco para verificar se há algo que possa realizar para diminuir as fatalidades. (Figura 3.7)

Figura 3.7 Histograma de Fatalidades x Fase de Operação

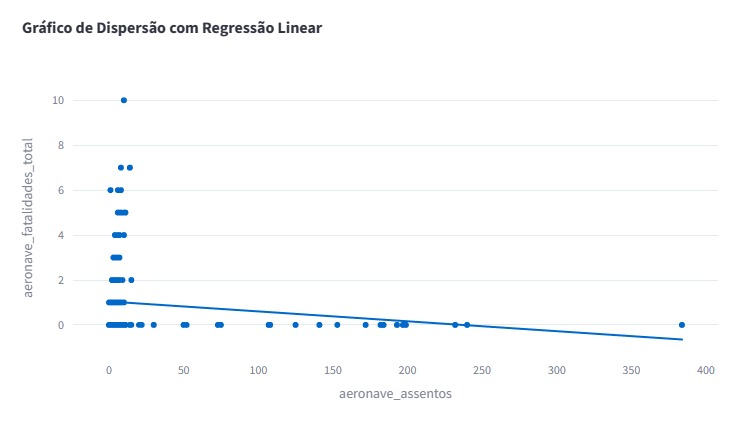


Fonte: Autores da pesquisa, 2024

Depois de inspecionar o histograma, torna-se evidente que a fase de operação mais perigosa é o CIRCUITO DE TRÁFEGO, onde seu objetivo é organizar os tráfegos visuais decolando e pousando, e junto com a fase de decolagem, torna-se evidente que os momentos mais perigosos da aviação seria a decolagem/pouso das aeronaves.

Após a determinação do ponto mais perigoso da aviação, geramos um gráfico interativo em que pode selecionar seus tipos de dados da planilha para gerar um gráfico de dispersão com regressão linear, para a figura foi selecionado os números de assentos e números de fatalidades para determinar se quanto maior o avião mais fatalidades ocorre. (Figura 3.8)

Figura 3.8 Gráfico de Dispersão de Fatalidades x Número de Assentos

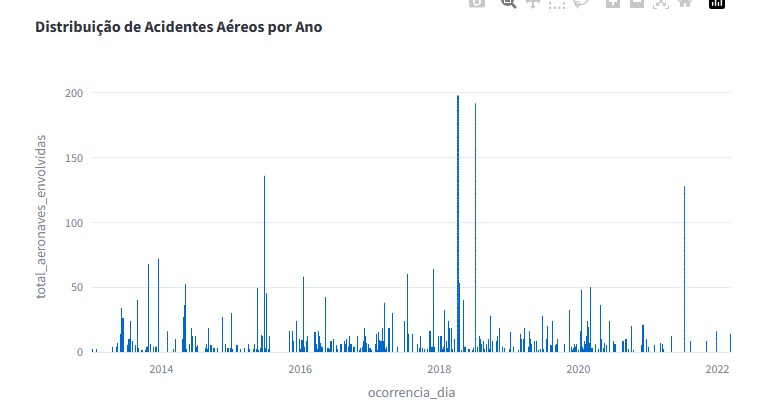


Fonte: Autores da pesquisa, 2024

Com a informação do gráfico e da reta da regressão linear, pode se concluir que aviões de pequeno porte tem mais índice de fatalidades.

E finalmente, conclui-se o código com um último gráfico mostrando a distribuição das aeronaves envolvidas em acidentes com o tempo, trazendo um tendencia em que maior parte das ocorrências ocorre por volta dos períodos de férias. (Figura 3.9)

Figura 3.9 Distribuição de Acidentes Aéreos por ano



Fonte: Autores da pesquisa, 2024

Por meio do trabalho apresentado determina-se que o maior número de acidentes ocorre por causa de perda de controle durante voo, onde o momento mais perigoso é a decolagem/pouso, aviões de pequeno porte são mais suscetíveis a incidentes com fatalidades. Uma medida que pode ser adotada para melhor segurança poderia melhor comunicação com a torre e a aeronave durante a fase de circuito de tráfego para melhor entendimento da região/clima para o piloto talvez adiar seu pouso ou decolagem para melhor proteção de sua vida, assim como as dos possíveis passageiros.

**Referências**

**Spartan College of Aeronautics and Technology,** 2021.

Disponível em: https://www.spartan.edu/news/history-of-aviation/#:~:text=The%20history%20of%20aviation%20dates,four%20pairs%20of%20glider%20wings. Acesso em: 25 abr. 2024.

Circuito de Tráfego. **IVAO Brasil,** 2024. Disponível em: https://wiki.br.ivao.aero/treinamento/materiais/Circuito-de-trafego. Acesso em: 05 maio 2024.

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS** **(PUC GOIÁS)**. VI Desafio em Ciências de Dados: Modelagem de Dados para Prevenir Acidentes Aéreos, Tecnólogo em Ciências de Dados e Inteligência Artificial