



ANÁLISE EXPLORATÓRIA DE DADOS RELATIVOS A **ACIDENTES AÉREOS**

Alunos: Ricardo Süffert, Guilherme Stein e Rafaela Varão.
Professor: Daniel Callegari.
Semestre: 2023/1.

1. DESCRIÇÃO DO DATASET ORIGINAL:

O *dataset* escolhido contém informações sobre acidentes aéreos ocorridos em território nacional e reportados às autoridades aeronáuticas brasileiras — no caso, a Agência Nacional de Aviação Civil (ANAC) e a Força Aérea Brasileira (FAB). A versão original do *dataset* foi extraída do Portal de Dados Abertos¹ do Governo Federal do Brasil, com dados coletados e disponibilizados pela ANAC e pela FAB. A caracterização oficial do *dataset* e o dicionário de dados estão disponíveis no site da ANAC e podem ser acessados por meio [deste link](#). Há uma variedade de acidentes, datando desde a década de 1950 até os dias atuais, mas os registros de acidentes estão restritos apenas àqueles ocorridos em território nacional e, novamente, reportados às autoridades aeronáuticas brasileiras.

No geral, o *dataset* parece estar bem construído; contudo, algumas inconsistências foram notadas durante o processo de transformação e manipulação dos dados, com relação, por exemplo, às coordenadas geográficas do acidente. Foi notado, ao plotar em um mapa os acidentes com base nas suas coordenadas de latitude e longitude, acidentes marcados como acontecidos no estado do Amazonas, por exemplo, foram plotados no continente africano. Além disso, algumas ocorrências reportavam a operadora do voo como “AZUL”, enquanto outras reportavam como “AZUL LINHAS AÉREAS BRASILEIRAS S.A.”, denotando falta de padronização dos dados dessa coluna. Inconsistências como essa em colunas que serão utilizadas para as análises realizadas nesse trabalho terão que ser tratadas durante a fase de pré-processamento — ou fase de transformação do processo de *extract, transform and load* (ETL) — dos dados, explicada na seção 3.

2. QUESTÕES DE NEGÓCIO:

Dadas as características do *dataset* utilizado, que contém apenas acidentes ocorridos em território nacional e reportados às autoridades regulatórias, serão considerados para responder às perguntas de negócio abaixo exclusivamente acidentes com essas características. Além disso, deseja-se considerar apenas voos comerciais regulamentados — isto é, aqueles voos tripulados realizados por companhias aéreas ou empresas de táxi aéreo registradas junto às autoridades aeronáuticas brasileiras. Então, para responder às duas perguntas abaixo, voos privados, por exemplo, serão removidos do *dataset* no passo de transformação e pré-processamento dos dados. As perguntas definidas para serem respondidas são:

I. Quais estratégias e incentivos podem/deveriam ser adotados por parte das autoridades regulatórias para prevenir acidentes aéreos?

- Para responder a essa pergunta, será necessário descobrir (1) quais os tipos/causas mais comuns de acidentes aéreos no *dataset* e (2) quais as companhias aéreas que mais acidentam, para saber em quais companhias as operações de prevenção devem ser focadas. Então, com base nessas informações, obter sugestões (idealmente junto a especialistas da área) para evitar que essas situações/problemas aconteçam.

¹ Portal de Dados Abertos: <https://dados.gov.br/dados/conjuntos-dados/ocorrencias-aeronauticas>

- II. **Qual a tendência, ao longo dos anos, para a frequência de acidentes aéreos?** Uma tendência de queda pode indicar a eficácia das políticas de prevenção de acidentes aéreos em vigência, enquanto uma tendência de alta, por outro lado, pode ser mais um indicativo de que novas políticas podem ser necessárias.

→ Para responder a essa pergunta, será necessário descobrir, no *dataset*, quantos acidentes aconteceram a cada ano para, então, colocá-los em um gráfico e analisar a evolução ao longo dos anos.

Outras questões foram levantadas e consideradas para análise; entretanto, as duas acima se sobressaíram. Uma outra pergunta interessante poderia ser tentar relacionar o modelo, porte e/ou fabricante da aeronave à frequência de acidentes, mas percebe-se que não seria uma análise muito assertiva, pois a responsabilidade pela manutenção das aeronaves é das companhias aéreas e, ao fazer uma análise baseada na fabricante, não há como saber as condições em que a companhia aérea mantinha as aeronaves. De fato, de acordo com reportagem de setembro de 2021 do portal Terra Notícias² com informações da Agência Nacional de Aviação Civil (ANAC), a falta de manutenção nas aeronaves, que é um dever que recai sobre as companhias aéreas, é uma das principais causas de acidentes aéreos. Desse modo, foi escolhido tentar estabelecer uma relação entre as próprias companhias aéreas e a frequência dos acidentes.

3. **PROCESSO DE ETL E O DATASET FINAL (MODIFICADO):**

Conforme explicado na seção 1, os dados foram extraídos do Portal de Dados Abertos¹ do Governo Federal do Brasil. O *dataset* original foi obtido no formato *comma-separated values* (CSV). Dito isso, para o pré-processamento dos dados, foi implementado um *script* — utilizando a biblioteca *pandas* do Python para análise e visualização de dados — para executar os seguintes passos:

- I. Filtrar as colunas do *dataset* original e manter apenas aquelas necessárias para responder às perguntas de negócio definidas na seção 2:
 - A. Para ambas as questões de negócio, a coluna *Operacao*, para determinar quais os acidentes que envolveram aviões comerciais;
 - B. Para a primeira questão de negócio, colunas *Descricao_do_Tipo* e *Operador_Padronizado*;
 - C. Para a segunda questão de negócio, coluna *Data_da_Ocorrencia*.
- II. Alterar os valores de algumas células, quando necessário. As alterações incluem: (1) remover a acentuação e caracteres especiais de todo o *dataset*, a fim de evitar problemas de codificação, inclusive nas próximas etapas do processo de tratamento dos dados, e (2) transformar a coluna *Data_da_Ocorrencia* em *Ano_Ocorrencia*, mantendo apenas o ano em que ocorreram os acidentes;

² Reportagem do Terra Notícias e ANAC:

<https://www.terra.com.br/noticias/falta-ou-ma-manutencao-periodica-das-aeronaves-ainda-e-uma-das-principais-causas-de-acidentes-aereos.af4c0d2cf9ae7d906af4ebd718d81b9c52toff41.html>

- III. Remover, com base na coluna Operacao, aqueles acidentes aéreos que não envolviam voos regulares/comerciais (operados por companhias aéreas regulamentadas) e nem operações de táxi aéreo (por também se tratarem de voos comerciais regulamentados);
- IV. Padronizar os valores que não estão padronizados (ex.: “AZUL” e “AZUL LINHAS AÉREAS BRASILEIRAS S.A.”);
- V. Criar um novo arquivo CSV com o número de acidentes ocorridos para cada ano;
- VI. Fazer uma visualização dos dados e verificação sobre a padronização das colunas mantidas. Por exemplo, para verificar que as colunas Descricao_do_Tipo (que contém a descrição do tipo de acidente), Operacao (que contém o tipo de voo — privado, regular etc.) e Operador_Padronizado (que contém a companhia que operava o voo que se acidentou) estão, de fato, padronizadas.

Por fim, os dados tratados foram carregados para o Microsoft Power BI para a montagem do *dashboard* apresentado na seção 4.

4. ***DASHBOARD*** MONTADO E RESPOSTAS ÀS PERGUNTAS DE NEGÓCIO:

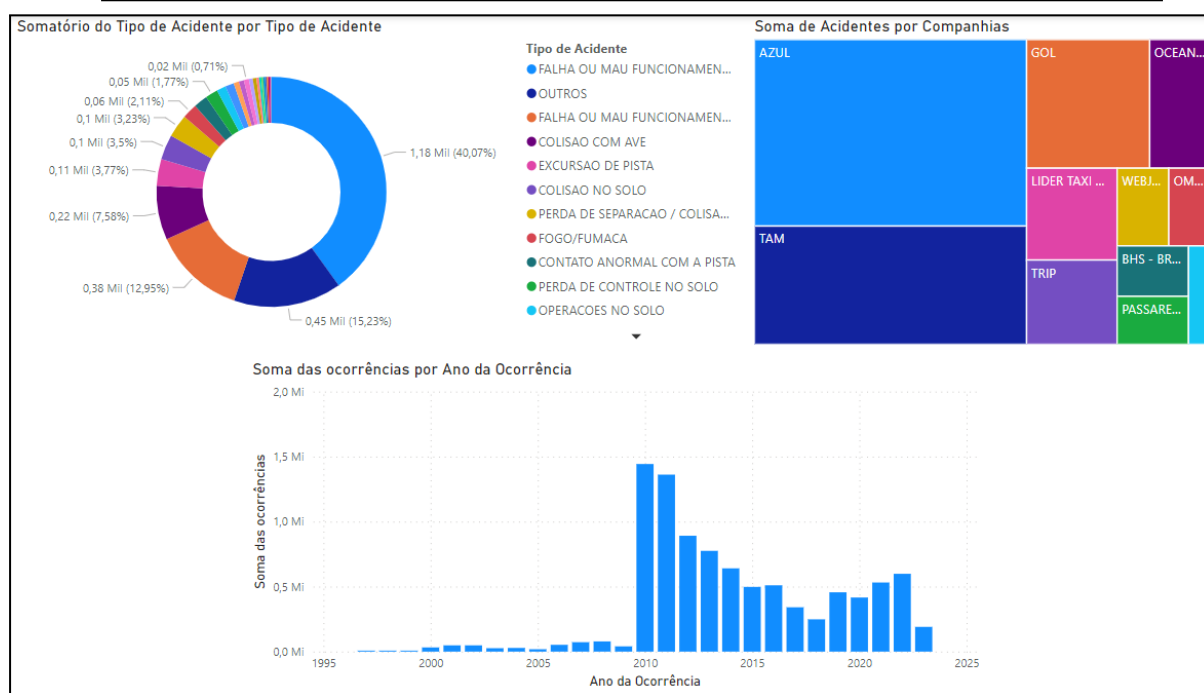


Figura 1: *dashboard* montado para responder às perguntas de negócio da seção 3.

A primeira pergunta levantada para este trabalho visa propor estratégias para reduzir a incidência de acidentes aéreos em território nacional. A partir da análise do primeiro infográfico (canto superior esquerdo), destaca-se a categoria dos acidentes causados por “falha ou mau funcionamento de sistema/componente”. A segunda causa mais frequente é a de “falha ou mau funcionamento do motor”. Juntas, essas duas categorias representam mais de 50% das causas de acidentes aéreos ocorridos no país. A outra metade dos acidentes aéreos têm causas bastante variadas e bem distribuídas. Tudo isso indica que **os acidentes aéreos estão muito mais relacionados com falhas mecânicas do que com falhas humanas**

e, portanto, investir mais em treinamento de pessoal, por exemplo, não parece ser a solução mais eficiente para reduzir a quantidade de acidentes atuais. Além disso, ao analisar o infográfico do canto superior direito, percebe-se que, **juntas, a Azul e a Tam (também conhecida como LATAM) correspondem a quase 60% dos acidentes aéreos** ocorridos em território nacional, com a Azul liderando com 36% do total. Isso indica que, para tomar as medidas preventivas da forma mais eficiente possível, devem ser conduzidas ações que busquem aprimorar os sistemas e componentes das aeronaves operadas, especialmente, por essas duas companhias. Como essas duas se destacam com quantidade de acidentes muito superior à da Gol — que é um outra companhia que opera uma quantidade muito grande de voos, semelhantemente às duas anteriores — talvez seja válido, inclusive, realizar uma investigação sobre a origem dos componentes e sistemas das aeronaves da Azul e da Tam, comparando-os com os da Gol para entender melhor o que pode causar tamanha disparidade entre a quantidade de acidentes de companhias aéreas de portes semelhantes.

Por outro lado, a segunda pergunta que este trabalho objetiva responder diz respeito à evolução da quantidade de acidentes aéreos ao longo dos anos. A partir da análise do infográfico mais abaixo na imagem supracitada, percebe-se que, em 2010, aconteceu um *boom* na quantidade de acidentes registrados no *dataset*. Isso pode ter sido causado pela falta de informações relativas aos acidentes dos anos anteriores, mas, se considerarmos a parte do gráfico **a partir do ano de 2010, percebe-se, até 2018, uma tendência forte de queda na quantidade de acidentes aéreos registrada. Contudo, a partir de 2019, essa tendência torna-se de crescimento**, o que é curioso, visto que, nesse período, grande parte do tráfego aéreo havia sido suspenso por conta da pandemia de Covid-19. Sobre esse fato, também seria válida uma investigação conjunta com as autoridades aeronáuticas brasileiras para entender melhor o que pode estar causando essa tendência de alta nos acidentes aéreos depois de tantos anos em queda e se isso está relacionado, de alguma forma, com a pandemia de coronavírus enfrentada pelo mundo nos últimos anos.

5. AUTOAVALIAÇÃO:

O grupo considera que, ao final do trabalho, conseguiu definir duas questões de negócio interessantes e pertinentes para o contexto do problema abordado (prevenção de acidentes aéreos), que se trata de uma questão de utilidade pública. Ambas as perguntas de negócio ajudam a caminhar na direção de uma solução para o problema e foram respondidas de forma satisfatória. No entanto, para ir mais a fundo na investigação, seria necessário trabalhar com as autoridades para conduzir a investigação sobre a origem dos componentes eletrônicos e dos sistemas dos aviões da Azul e da Tam mencionada na seção anterior e para entender o motivo da tendência de ocorrência de acidentes aéreos ter se tornado de alta a partir de 2019, e pode-se dizer que não poder seguir adiante com essa investigação foi uma dificuldade encontrada no trabalho. Entretanto, ainda assim, foi possível determinar que:

1. Como resposta à questão de negócio I, para prevenir os acidentes aéreos de forma eficiente — isto é, atacar o problema em sua raiz — uma boa abordagem seria **investigar a origem dos componentes eletrônicos** (já que são eles que causam a maioria dos acidentes) **das companhias Azul e Tam** (já que elas concentram a maior quantidade de acidentes) e **comparar essas informações com aquelas relativas à**

Gol — que é uma companhia de porte similar que, no entanto, possui uma quantidade de acidentes registrados muito menor — a fim de entender o que há de diferente entre essas companhias para fazer a Gol ter um desempenho melhor nesse quesito.

2. Como resposta à questão de negócio II, **de 2010 a 2018, a tendência dos acidentes aéreos era de queda, mas, curiosamente, a partir de 2019** (ano em que começou a pandemia do coronavírus), **a tendência passou a ser o aumento no número de acidentes.**

Ademais, o grupo ficou satisfeito com a solução encontrada e considera seu desempenho digno de uma boa avaliação. Em uma média de todos os quesitos, atribui nota 9,5 a seu desempenho no presente trabalho.