

# Estudo sobre alguns mitos na aviação civil brasileira

Hamilton Fonte II

Universidade de São Paulo (USP)

Escola Politécnica - Engenharia de Computação

Programa de Pós Graduação em Engenharia Elétrica

São Paulo, SP, Brasil

Email: hamiltonii@gmail.com

Su Yin he

Universidade de São Paulo (USP)

Escola Politécnica - Engenharia de Computação

Programa de Pós Graduação em Engenharia Elétrica

São Paulo, SP, Brasil

Email: suyinhe@gmail.com

**Abstract**—Ciência de dados tem atraído o interesse de várias instituições recentemente. A quantidade considerável de dados armazenados tem viabilizado análises e respondido perguntas que não eram possíveis com a quantidade de dados e as técnicas de análise anteriores. Entender o ciclo de vida dos dados e aplicar operações em datasets para responder perguntas é uma habilidade cuja demanda esta em ascensão. Este documento é o trabalho final dos estudantes Su e Hamilton para disciplina de Ciência de Dados (PCS-5031), nele analisamos os dados coletados em um período de 10 anos pelo Centro Nacional de Investigação e Prevenção de Acidentes Aeronauticos (CENIPAA) para confirmar ou refutar alguns mitos relacionados a aviação no Brasil.

## I. INTRODUÇÃO E OBJETIVOS

Este trabalho tem por objetivo confirmar ou refutar mitos relacionados as causas de acidentes e incidentes aeronauticos através da análise dos dados resultantes da investigação de acidentes e incidentes aeronauticos pelo CENIPA.

O Centro de Investigação e Prevenção de Acidentes Aeronauticos (CENIPA) foi criado em 1971 como órgão centralizador para o Sistema de Investigação e Prevenção de Acidentes Aeronauticos (SIPAER). O objetivo do CENIPA é o de promover a prevenção de acidentes aeronauticos através da reunião, análise de informações e obtenção de conclusões, incluindo a identificação dos fatores contribuintes para a ocorrência, visando a formulação de recomendações sobre a segurança. O Sistema de Investigação e Prevenção de Acidentes Aeronáuticos (SIPAER) não trabalha com "causa" de acidente, mas com fatores contribuintes.

Acidente Aeronáutico é toda ocorrência relacionada com a operação de uma aeronave, entre o período em que uma pessoa nela embarca com a intenção de realizar um voo, até o momento em que todas as pessoas tenham dela desembarcado e, durante o qual, pelo menos uma das situações abaixo ocorra:

- Uma pessoa sofra lesão grave ou morra como resultado de estar na aeronave
- A aeronave sofra dano ou falha estrutural que afete sua resistência estrutural, seu desempenho ou as suas características de voo
- A aeronave seja considerada desaparecida ou o local onde se encontre seja absolutamente inacessível.

## II. OS DATASETS

Para atingir os objetivos acima, utilizamos 3 datasets fornecidos no site do CENIPA

- 1) Aeronaves
- 2) Ocorrências
- 3) Fatores Contribuintes

O dataset "Aeronaves" contém informações sobre as aeronaves envolvidas no incidente, relevante a este estudo são as informações de **equipamento, fabricante, modelo, tipo motor, fase da operação, tipo da operação, nível de danos e quantidade de fatalidades**. O dataset "Ocorrências" contém informações sobre o ocorrido, relevantes a este estudo são as informações de **tipo, localidade, dia da ocorrência e horário UTC**. Finalmente o dataset de fatores contribuintes contem os dados de **fator, aspecto condicionante e área** dos incidentes ocorridos com as aeronaves.

### A. Planejamento

Como mencionado anteriormente, este estudo tem o objetivo de desmistificar ou confirmar alguns fatos relacionados a aviação no Brasil. Perguntas que estamos tentando responder são:

- Qual fabricante esta mais envolvido em incidente? Há uma percepção na população menos informada que os Airbus tem participado de incidentes aéreos com mais frequência que os Boeings.
- A maioria dos incidentes aconteceu na decolagem ou no pouso? Em várias entrevistas, especialistas em aviação concordam ao afirmar que essas duas fases são as mais perigosas.
- A maioria dos incidentes esta realmente relacionado a falha humana? Como no anterior, especialistas em aviação afirmam que, dado o nível de automatização e salvaguardas presente hoje nas aeronaves, incidentes são geralmente causados por intervenção humana inadequada em alguma fase da operação.

### B. Coleta dos dados e qualidade dos dados

O método investigativo do SIPAER utiliza probabilidades e hipóteses para esclarecer as causas e os fatores contribuintes de um incidente aéreo. O CENIPA mantém um cemitério de aeronaves que usa para estudos e como base histórica de

acidentes passados, combinado com a obtenção de declarações voluntárias de observadores e dados do gravador de voo da aeronave, formula-se um relatório que tem, entre outras informações, o que acredita serem as causas e os fatores contribuintes além de recomendações às companhias aéreas para evitar que um incidente parecido volte a ocorrer.

A coleção dos fatores contribuintes e suas recomendações, são armazenados nos 3 datasets usados neste estudo. As recomendações levam em consideração que a descrição voluntária de observadores esta sujeita a erros relacionados a percepção humana.

### C. Descrição dos dados

Aqui faremos uma breve descrição dos campos do dataset que consideramos relevantes para o estudo.

- Equipamento - VARCHAR(45): Campo discreto, descreve o tempo de aeronave envolvida no incidente, no dataset ele pode assumir os valores AVIÃO, HELICOPTERO, PLANADOR, TRIKE e ULTRALEVE
- Fabricante - VARCHAR(45): Campo discreto, descreve o nome do fabricante da aeronave envolvida no incidente.
- Modelo - VARCHAR(45): Campo discreto, descreve o modelo da aeronave envolvida no incidente. Este campo e os anteriores devem nos demonstrar alguma coincidência relacionada ao tipo de aeronave e fabricante, por exemplo um erro de construção ou de instrução pode aparecer na forma de vários incidentes com o mesmo modelo de aeronave.
- Tipo motor - VARCHAR(45): Campo discreto, descreve o tipo de motor usado na aeronave envolvida no incidente, no dataset ele pode assumir os valores, JATO, PISTÃO, SEM TRAÇÃO (típico de planadores), TURBO EIXO (usado em helicópteros), TURBO HÉLICE
- Fase da operação - VARCHAR(15): Campo discreto, descreve o momento da operação da aeronave em que ocorreu o incidente, dois exemplos de valores para este campo são DECOLAGEM e POUSO.
- Tipo - VARCHAR(80): Campo discreto, descreve de forma macro o que ocorreu com a aeronave no incidente, pode assumir valores como, COLISÃO COM PÁSSARO, SAÍDA DE PISTA entre muitos outros
- Localidade - VARCHAR(100): Campo discreto, descreve o município onde ocorreu o incidente. No caso de incidentes ocorridos em voo como colisões com pássaros ou perda de motor, a localidade descreve o município que a aeronave estava sobrevoando quando o incidente ocorreu.
- Aerodromo - VARCHAR(4): Campo discreto, descreve o aeródromo onde ocorreu o incidente. Em casos de incidentes ocorridos nas fases onde a aeronave esta fora do solo, este campo assume valores vazios (\*\*\*) exceto quando o incidente esta relacionado com as fases de aproximação para pouso ou decolagem.
- Fator - VARCHAR(45): Campo discreto, descreve a principal causa do incidente
- Aspecto - VARCHAR(45): Campo discreto, descreve a

que aspecto do fator causador o incidente esta relacionado.

- Condicionante - VARCHAR(45): Campo discreto, descreve sob que condições o aspecto se apresentou de modo tornar-se o fator principal no incidente.

### III. PREPARAÇÃO DOS DADOS

As seguintes atividades foram realizadas nos dados para viabilizar as descobertas.

- Na tabela **fator contribuinte**, substituímos o valor \*\*\*, os **espaços em branco** e **marcadores de fim de linha** por ;
- Ainda na tabela **fator contribuinte** aumentamos a quantidade de caracteres na coluna **area**
- Criar tabela **fator distinct** com as colunas **fator, aspecto, condicionante** e **area** com os fatores mapeados pelo CENIPA
- Na tabela **aeronave** substituir \*\*\*\* e \*\*\* por string vazia, aumentamos o numero de caracteres nas colunas **fase operacao** e **nivel dano**. Os valores NULL de **quantidade fatalidade** foram substituidos por 0
- Na tabela **ocorrências** substituir \*\*\*\*, \*\*\* e NULL por string vazia
- Preencher **data publicacao** em branco por 0000-00-00
- Modificar valores da coluna **relatorio publicado** de 1 e 0 para **sim** e **não**
- Modificar os valores de **saida de pista** de 1 e 0 para **sim** e **não**

### IV. DESCOBRIR

Pelo tamanho não muito grande do dataset optamos por uma ferramenta mais simples para realização das buscas, um banco mysql. Focando nos objetivos do trabalho, realizamos as seguintes buscas nos datasets do CENIPAA

- 1) Qual fabricante esta mais envolvido em incidente? Há uma percepção na população menos informada que os Airbus tem participado de incidentes aéreos com mais frequência que os Boeings?

Realizando buscas nos datasets chegamos aos valores relacionados na **table 1**. Podemos notar que no que tange quantidade de aeronaves envolvidas em incidentes, a AIRBUS não só não é o maior fabricante com aeronaves envolvidas em incidentes como também é a que menos se envolveu em incidentes em território brasileiro nos últimos 10 anos. Esta tabela também demonstra uma faceta não prevista do nosso estudo, apesar da AIRBUS ser o menor fabricante com incidentes é de longe a maior média de fatalidades por incidente chegando a ter uma média quase 3x maior que o segundo fabricante. Acreditamos que tal fato se deve a natureza do produto comercializado pelo fabricante. Olhando o catálogo da Airbus disponível na internet a mesma só comercializa aeronaves de grande porte, assim um incidente naturalmente tende a ter

várias fatalidades, os outros fabricantes como Boeing e Cessna tem uma variedade muito maior de aeronaves, desde monopostos de pouca capacidade de passageiros até jatos de transporte comercial de passageiros. A fabricante Neiva (Embraer) produz apenas monopostos para aplicações agrícolas.

- 2) A maioria dos incidentes aconteceu na decolagem ou no pouso? Em várias entrevistas, especialistas em aviação concordam ao afirmar que essas duas fases são as mais perigosas.

Em novas buscas nos datasets, chegamos aos valores relacionados na **table 2**. Os dados confirmam a percepção de que as fases em que mais ocorrem incidentes na aviação brasileira são realmente as fases de decolagem e pouso. A literatura aeronautica justifica esse fato relacionando-o a baixa altitude da aeronave durante essas fases de sua operação o que não deixa muita margem para recuperação como por exemplo em caso de pane em um ou todos os motores. Algo não previsto mas que demonstrou ser interessante na análise dos datasets é a quantidade de fatalidades por incidentes nas diversas fases de operação, a maior concentração de fatalidades não estava nem na decolagem nem no pouso mas em operação de cruzeiro. Acreditamos que tal fato se deve a definição de incidente utilizada pelo CENIPA, vamos relembrar "Uma pessoa sofra lesão grave ou morra como resultado de estar na aeronave". Esta definição engloba uma série de eventos como, infartos e lesões causadas por turbulência que, apesar de não expor a aeronave a uma situação de risco figuram como incidentes. Infelizmente o dataset não contém detalhes do incidente suficientes para conseguirmos averiguar os grupos de incidentes.

- 3) A maioria dos incidentes esta realmente relacionado a falha humana? Como no anterior, especialistas em aviação afirmam que, dado o nível de automatização e salvaguardas presente hoje nas aeronaves, incidentes são geralmente causados por intervenção humana inadequada em alguma fase da operação.

Para responder a esta pergunta foi necessário um agrupamento pois o dataset **Fatores Contribuintes** possui 4118 registros em que identificamos 63 fatores únicos os quais foram agrupados da seguinte forma:

- **Comunicação:** Falhas de comunicação entre a organização e o piloto ou fraseologias inadequadas. Por organização entende-se os controladores ou os tripulantes
- **Infraestrutura:** Problemas de infraestrutura em solo
- **Meio ambiente:** Condições no meio ambiente que dificulte a visibilidade como noite escura, reflexo do

TABLE I  
AERONAVES E FATALIDADES POR FABRICANTE

Fabricante	num fatalidades	num aeronaves	media fatalidades / aeronave
Airbus industries	199	17	11,7059
Being Company	154	32	4,8125
Neiva Industria Aeronautica	112	357	0,3137
Cessna	108	333	0,3243

TABLE II  
AERONAVES E FATALIDADES POR FASE DE OPERAÇÃO

Fase Operação	num fatalidades	num aeronaves	media fatalidades / aeronave
Pouso	216	345	0,6261
Decolagem	125	321	0,3894
Cruzeiro	238	225	1,5022
Subida	69	93	0,7419
Indeterminada	63	96	0,6563
Manobra	63	94	0,6702
Descida	51	62	0,8226
Aproximação Final	42	66	0,6364
Circuito de Tráfego	35	49	0,7143
Outra Fase	30	58	0,5172

sol, camuflagem.

- **Meteorologia adversa:** Condições meteorológicas adversas como tempestados, nevoeiro, nuvens baixas
- **Normas:** Operação em desacordo com normas
- **Organização:** Falha organizacional
- **Piloto:** Falha do piloto
- **Projeto:** Falha do projeto da aeronave, problemas no processo de fabricação
- **Treinamento:** Falta de treinamento inicial ou atualização

Nas buscas nos datasets chegamos aos valores relacionados na **table 3**. Quando colocamos o questionamento se a maioria das ocorrências estava relacionada a falha humana tínhamos em mente apenas erros do piloto. Embora a busca tenha confirmado esta percepção, é interessante notar que o CENIPA considera como falha humana um leque muito maior de fatores que apenas o erro do piloto, além deles também temos a falha de comunicação entre o piloto e a organização, falhas organizacionais e falhas por falta de treinamento inicial ou atualização. Voltando a questão original, é interessante notar que os fatores erro do piloto e questões organizacionais não só figuram entre os mais contribuintes a incidentes mas esta quase 4x mais presente em incidentes que o terceiro fator mais presente.

TABLE III  
FATORES E OCORRÊNCIAS

Classificação Fator Contribuinte	Fatores Ocorrências	Ocorrências Únicas
Piloto	1850	769
Organização	1374	714
Treinamento	234	192
Normas	206	198
Outros	132	132
Meteorologia Adversa	126	126
Infra Estrutura	79	78
Projeto	44	42
Meio Ambiente	38	38
Comunicação	35	33

## V. CONCLUSÃO

Acreditamos que um dos principais aprendizados deste trabalho foi entender que, em algumas situações, boa parte do controle do ciclo de vida dos dados esta sob responsabilidade de uma pessoa ou grupo de pessoas diferente daquele que vai realizar as descobertas sobre os dados, levando o grupo a algumas situações como as exemplificadas nos próximos parágrafos.

O dicionário de dados fornecido com o Dataset não explicava o significado de alguns dos dados como por exemplo o campo **operador**. O mesmo campo era preenchido por valores aos quais não conseguimos referências alguma o que impactou algumas questões que pretendíamos desvendar.

O dataset carecia de padronização quando os dados não estavam disponíveis, hora era uma string vazia, hora era NULL, hora era 0. Tinha falhas de preenchimento, 1013 ocorrências não tinha nenhum fator relacionado.

## REFERENCES

- [1] AMAZON. *Amazon Media Room Press Release*. [S.l.], 2016. 2 p.
- [2] KEYUN, R. et al. *Advances in Digital Forensics IV*. 7. ed. Orlando: [s.n.], 2011. 35–46 p. ISSN 1098-6596. ISBN 9788578110796.