



UNIVERSIDADE FEDERAL DE MINAS GERAIS

TP2 - Introdução a Banco de Dados

Ocorrências Aeronáuticos no Brasil com base nos dados do CENIPA

Integrantes:

Mateus Ryan de Castro Lima – 2021036752

Ana Clara Vasconcelos dos Santos – 2021080573

Emanuel Magalhães Figueiredo – 2024023821

Mateus Rabelo de Souza Melo – 2024023783

MG - BH

Junho de 2025

Sumário

1	Introdução	2
2	Análise Crítica da fonte - dificuldades, qualidade	2
3	Modelo ER	3
4	Dicionário de Dados	4
5	Conjunto de Metadados	5
6	Análise Exploratória	8
6.1	Preparação dos dados	8
6.2	Definição de Objetivos	9
6.3	Análise descritiva	10
6.3.1	Fatalidades por porte	10
6.3.2	Quantidades de Fatalidades por ano	11
6.3.3	Quantidades de fatalidades por estado	13
6.3.4	Fatores contribuintes para acidentes	14
6.3.5	Incidentes por tipos de motor e fase do voo	15
7	Conclusão	15
8	Referências	16

1 Introdução

A aviação é considerada um dos meios de transporte mais seguros, mas acidentes ainda acontecem e precisam ser analisados com atenção. Este trabalho tem como foco os acidentes aéreos ocorridos no Brasil, utilizando dados públicos fornecidos pelo CENIPA — órgão responsável pela investigação e prevenção dessas ocorrências.

A proposta é explorar essa base de dados para entender melhor os fatores envolvidos nos acidentes, além de organizar e analisar as informações por meio de modelagem e técnicas de análise exploratória. Com isso, buscamos não apenas identificar padrões, mas também mostrar como um banco de dados bem estruturado pode apoiar ações de prevenção e segurança na aviação.

2 Analise Critica da fonte - dificuldades, qualidade

A base de dados disponibilizada pelo CENIPA (Centro de Investigação e Prevenção de Acidentes Aeronáuticos) é, sem dúvida, uma das mais completas e relevantes no que diz respeito ao registro de incidentes e acidentes aéreos no Brasil. Ela oferece uma ampla variedade de informações que são de grande interesse para profissionais da aviação, pesquisadores e para o público em geral que busca compreender melhor o cenário da segurança operacional no país. No entanto, apesar de sua abrangência e valor, a base apresenta algumas limitações que dificultam seu uso direto para análises mais profundas.

Ao carregar os dados, foram encontrados diversos problemas de formatação, especialmente nos campos de latitude e longitude, que deveriam representar coordenadas geográficas, mas muitas vezes aparecem com formatos quebrados ou inconsistentes. Além disso, há uma quantidade significativa de campos com valores nulos ou vazios, inclusive em colunas que contêm informações essenciais, como número de fatalidades, tipo de motor e fase da operação. Em alguns registros, praticamente todas as colunas estão incompletas, o que compromete a integridade da informação. Esses fatores tornam necessário um tratamento prévio da base antes que ela possa ser utilizada de forma eficaz em qualquer tipo de análise automatizada ou estatística confiável.

Outro ponto que chama atenção é a ausência de uma documentação adequada. Muitos campos apresentam siglas ou termos técnicos da aviação que não são explicados em nenhum lugar da base, o que pode gerar confusão para usuários que não têm familiaridade com o setor. Um simples dicionário de dados com descrições claras dos campos e possíveis valores já seria suficiente para tornar a base mais acessível e compreensível.

Também foi observado que faltam alguns indicadores que poderiam enriquecer a análise. Por exemplo, não há um campo que indique se a aeronave era cargueira, embora esse dado possa ser parcialmente inferido pelo modelo da aeronave ou pela quantidade de assentos. A presença explícita dessa informação seria útil para diferenciar melhor os tipos de operação envolvidos em cada ocorrência. Além disso, nas tabelas que registram as recomendações geradas pelas investigações, não há nenhuma classificação ou categorização padronizada, o que dificulta a identificação de padrões ou a realização de análises qualitativas mais amplas. Embora seja compreensível que cada recomendação tenha um caráter específico, uma taxonomia mínima, mesmo que ampla, ajudaria a organizar melhor esses dados.

Em resumo, a base de dados do CENIPA é extremamente rica em conteúdo, mas carece de padronização, documentação e tratamento técnico adequado. Melhorias nessas áreas tornariam a base muito mais útil e acessível, potencializando seu papel como instrumento de pesquisa e prevenção de acidentes na aviação civil brasileira.

3 Modelo ER

Para representar de forma estruturada os dados dos acidentes aéreos, foi elaborado um Modelo Entidade-Relacionamento (ER), que mapeia as principais entidades, atributos e relacionamentos envolvidos.

O modelo foi desenvolvido com base na estrutura fornecida pelo conjunto de dados do CENIPA, buscando representar fielmente a organização lógica das informações, como ocorrências, aeronaves, fatores contribuintes e localizações.

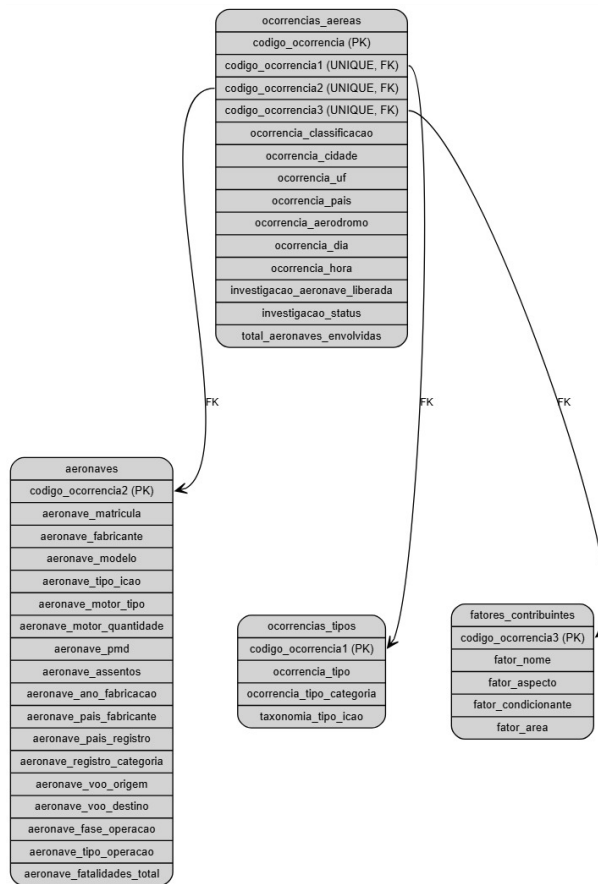


Figura 1: Diagrama Entidade-Relacionamento do banco de dados de acidentes aeronáuticos.

4 Dicionário de Dados

O dicionário de dados a seguir apresenta os campos das tabelas utilizadas no banco de dados, com seus respectivos nomes, tipos, chaves e descrições. Essa documentação é essencial para o entendimento da estrutura do banco e garante consistência nas análises.

RELAÇÃO	ATRIBUTO	TIPO/LARGURA	NULO?	ÚNICO?	VALORES PERMITIDOS	RESTRIÇÕES ADICIONAIS COMPORTAMENTO SE CHAVE ESTRANGEIRA
ocorrencias_aereas	codigo_ocorrendia	INTEGER	N	S		PK
ocorrencias_aereas	codigo_ocorrendia1	INTEGER	N	S		
ocorrencias_aereas	codigo_ocorrendia2	INTEGER	N	S		
ocorrencias_aereas	codigo_ocorrendia3	INTEGER	N	S		
ocorrencias_aereas	ocorrencia_classificacao	VARCHAR(50)	N	N	ACIDENTE;INCIDENTE; INCIDENTE GRAVE	
ocorrencias_aereas	ocorrencia_cidade	VARCHAR(100)	N	N		
ocorrencias_aereas	ocorrencia_uf	VARCHAR(2)	S	N		
ocorrencias_aereas	ocorrencia_pais	VARCHAR(50)	N	N		
ocorrencias_aereas	ocorrencia_aerodromo	VARCHAR(10)	N	N		
ocorrencias_aereas	ocorrencia_dia	DATE	N	N		
ocorrencias_aereas	ocorrencia_hora	TIME	S	N		
ocorrencias_aereas	investigacao_aeronave_liberada	VARCHAR(3)	N	N	SIM,NÃO	
ocorrencias_aereas	investigacao_status	VARCHAR(50)	S	N	FINALIZADA;ATIVA	
ocorrencias_aereas	total_aeronaves_envolvidas	INT	N	N		
ocorrencias_tipos	codigo_ocorrendia1	INT	N	S		PK;FK REFERENCES ocorrencias_aereas(codigo_ocorrendia1) ON DELETE CASCADE
ocorrencias_tipos	ocorrencia_tipo	VARCHAR(255)	N	N		
ocorrencias_tipos	ocorrencia_tipo_categoria	VARCHAR(255)	N	N		
ocorrencias_tipos	taxonomia_tipo_jcao	VARCHAR(20)	N	N		
aeronaves	codigo_ocorrendia2	INT	N	S		PK;FK REFERENCES ocorrencias_aereas(codigo_ocorrendia2) ON DELETE CASCADE
aeronaves	aeronave_matricula	TEXT	N	S		
aeronaves	aeronave_fabricante	TEXT	S	N		
aeronaves	aeronave_modelo	TEXT	S	N		
aeronaves	aeronave_tipo_jcao	TEXT	S	N		
aeronaves	aeronave_motor_tipo	TEXT	S	N		
aeronaves	aeronave_motor_quantidade	TEXT	S	N		
aeronaves	aeronave_pmd	INTEGER	N	N		
aeronaves	aeronave_assentos	INTEGER	S	N		
aeronaves	aeronave_ano_fabricacao	INTEGER	S	N		

Figura 2: Dicionário de Dados – Parte 1

RELAÇÃO	ATRIBUTO	TIPO/LARGURA	NULO?	ÚNICO?	VALORES PERMITIDOS	RESTRIÇÕES ADICIONAIS COMPORTAMENTO SE CHAVE ESTRANGEIRA
aeronaves	aeronave_pais_fabricante	TEXT	N	N		
aeronaves	aeronave_pais_registro	TEXT	N	N		
aeronaves	aeronave_registro_categoria	TEXT	S	N		
aeronaves	aeronave_voo_origem	TEXT	S	N		
aeronaves	aeronave_voo_destino	TEXT	S	N		
aeronaves	aeronave_fase_operacao	TEXT	S	N		
aeronaves	aeronave_tipo_operacao	TEXT	S	N		
aeronaves	aeronave_fatalidades_total	INTEGER	N	N		
fatores_contribuintes	codigo_ocorrendia3	INTEGER	N	S		PK;FK REFERENCES ocorrencias_aereas(codigo_ocorrendia3) ON DELETE CASCADE
fatores_contribuintes	fator_nome	TEXT	N	N		
fatores_contribuintes	fator_aspecto	TEXT	N	N		
fatores_contribuintes	fator_condicionante	TEXT	S	N		
fatores_contribuintes	fator_area	TEXT	N	N		

Figura 3: Dicionário de Dados – Parte 2

5 Conjunto de Metadados

A seguir são apresentados os metadados que descrevem a estrutura do banco de dados utilizado para a análise dos acidentes aeronáuticos. Cada tabela foi criada com base nos campos presentes no conjunto de dados original fornecido pelo CENIPA, respeitando os

relacionamentos lógicos entre os registros.

Tabela: aeronaves

```
CREATE TABLE aeronaves (  
    codigo_ocorrendia2 INT PRIMARY KEY,  
    aeronave_matricula TEXT,  
    aeronave_fabricante TEXT,  
    aeronave_modelo TEXT,  
    aeronave_tipo_icao TEXT,  
    aeronave_motor_tipo TEXT,  
    aeronave_motor_quantidade TEXT,  
    aeronave_pmd INTEGER,  
    aeronave_assentos INTEGER,  
    aeronave_ano_fabricacao INTEGER,  
    aeronave_pais_fabricante TEXT,  
    aeronave_pais_registro TEXT,  
    aeronave_registro_categoria TEXT,  
    aeronave_voo_origem TEXT,  
    aeronave_voo_destino TEXT,  
    aeronave_fase_operacao TEXT,  
    aeronave_tipo_operacao TEXT,  
    aeronave_fatalidades_total INTEGER,  
    FOREIGN KEY (codigo_ocorrendia2) REFERENCES  
        ocorrencias_aereas(codigo_ocorrendia2) ON DELETE CASCADE  
);
```

Tabela: ocorrencias_tipos

```
CREATE TABLE ocorrencias_tipos (  
    codigo_ocorrencia1 INT PRIMARY KEY,  
    ocorrencia_tipo VARCHAR(255),  
    ocorrencia_tipo_categoria VARCHAR(255),  
    taxonomia_tipo_icao VARCHAR(20),  
    FOREIGN KEY (codigo_ocorrencia1) REFERENCES  
        ocorrencias_aereas(codigo_ocorrencia1) ON DELETE CASCADE  
);
```

Tabela: ocorrencias_aereas

```
CREATE TABLE ocorrencias_aereas (  
    codigo_ocorrencia INT PRIMARY KEY,  
    codigo_ocorrencia1 INT UNIQUE,  
    codigo_ocorrencia2 INT UNIQUE,  
    codigo_ocorrencia3 INT UNIQUE,  
    ocorrencia_classificacao VARCHAR(50),  
    ocorrencia_cidade VARCHAR(100),  
    ocorrencia_uf VARCHAR(2),  
    ocorrencia_pais VARCHAR(50),  
    ocorrencia_aerodromo VARCHAR(10),  
    ocorrencia_dia DATE,  
    ocorrencia_hora TIME,  
    investigacao_aeronave_liberada VARCHAR(3),  
    investigacao_status VARCHAR(50),  
    total_aeronaves_envolvidas INT  
);
```


Tabela: fatores_contribuintes

```
CREATE TABLE fatores_contribuintes (  
    codigo_ocorrendia3 INT PRIMARY KEY,  
    fator_nome TEXT,  
    fator_aspecto TEXT,  
    fator_condicionante TEXT,  
    fator_area TEXT,  
    FOREIGN KEY (codigo_ocorrendia3) REFERENCES  
        ocorrencias_aereas(codigo_ocorrendia3) ON DELETE CASCADE  
);
```

6 Análise Exploratória

6.1 Preparação dos dados

Durante a importação dos dados utilizando o comando `COPY`, foram encontrados alguns erros comuns relacionados à codificação e ao tratamento de valores nulos. A seguir, descrevemos os principais problemas identificados e as soluções adotadas:

1. Erro de codificação UTF-8

Erro: sequência de bytes inválida para codificação UTF-8: 0xc3 0x4f

Contexto: `COPY aeronaves`, linha 2

Causa: O arquivo CSV estava salvo com codificação `LATIN1` (ISO-8859-1), comum em exportações do Excel ou sistemas brasileiros.

Tratamento aplicado: Foi especificado `ENCODING 'LATIN1'` no comando `COPY` para garantir a leitura correta dos caracteres especiais.

2. Valor "NULL" como string em colunas numéricas

Erro: sintaxe de entrada inválida para tipo `INTEGER`: "NULL"

Contexto: `COPY aeronaves`, linha 3, coluna `aeronave_assentos`: "NULL"

Causa: A string "NULL" foi interpretada literalmente como texto, e não como valor nulo, o que causou erro nas colunas numéricas.

Tratamento aplicado: Os campos "NULL" foram substituídos por um valor especial (como `''`), e o comando `COPY` foi ajustado com o parâmetro `NULL ''` para

interpretação correta de nulos.

3. Campos vazios em colunas numéricas

Erro: sintaxe de entrada inválida para tipo `INTEGER`:

Contexto: `COPY aeronaves`, linha 2126, coluna `aeronave_pmd`:

Causa: Campos vazios `()` não são automaticamente tratados como `NULL`, o que gera erro em campos numéricos.

Tratamento aplicado: Os campos vazios foram padronizados com o valor `'***'` e o comando `COPY` foi configurado com `NULL '*'` para permitir a conversão correta para nulo.

6.2 Definição de Objetivos

O principal objetivo desta análise é compreender os padrões, riscos e características associados aos acidentes aeronáuticos no Brasil entre os anos de 2007 e 2025, com base nos dados fornecidos pelo CENIPA. A partir disso, buscou-se:

- Avaliar a gravidade dos acidentes de acordo com o porte das aeronaves, identificando se há correlação entre tamanho da aeronave e severidade das ocorrências.
- Identificar tendências temporais, como o número de fatalidades por ano e a evolução do tipo de dano estrutural sofrido pelas aeronaves.
- Detectar padrões espaciais, analisando a distribuição geográfica das fatalidades por estado brasileiro.
- Investigar os principais fatores contribuintes para os acidentes, buscando evidenciar falhas humanas, técnicas ou organizacionais recorrentes.
- Explorar a relação entre tipo de motor e fase do voo, observando se certas combinações representam maior risco.

Esses objetivos orientaram a preparação e análise dos dados, garantindo foco na segurança operacional e fornecendo subsídios para tomadas de decisão e propostas de mitigação de riscos.

6.3 Análise descritiva

6.3.1 Fatalidades por porte

Fatalidades por Porte de Aeronave

Porte da Aeronave	Total de Ocorrências	Total de Fatalidades	Tatalidades per Ocorrência	Fatalidades por Assento
PEQUENO	6,930	1.246	0,18	0,0372
MÉDIO	1.243	131	0,11	0,0026
GRANDE	3.927	199	0,05	0,0004

Figura 4: Distribuição das fatalidades por porte da aeronave, considerando total de ocorrências, total de fatalidades, fatalidades por ocorrência e fatalidades por assento.

Para realizar a análise de segurança das aeronaves, foi adotada uma classificação por porte, utilizando preferencialmente o PMD (Peso Máximo de Decolagem) e, em casos onde esse valor estava ausente, foi utilizado o número de assentos da aeronave. As faixas foram definidas da seguinte forma:

- **Pequeno:** $\text{PMD} < 5.700 \text{ kg}$ ou número de assentos < 10
- **Médio:** $5.700 \text{ kg} \leq \text{PMD} < 27.000 \text{ kg}$ ou $10 \leq \text{número de assentos} < 100$
- **Grande:** $\text{PMD} \geq 27.000 \text{ kg}$ ou número de assentos ≥ 100

Com a classificação concluída, foram calculadas duas métricas principais para avaliação de segurança:

Fatalidades por ocorrência: obtida dividindo-se o total de fatalidades pelo número de ocorrências. Significado: representa a gravidade média dos acidentes de cada porte. Vantagens: simples de interpretar; útil para comparar a severidade dos acidentes. Desvantagens: não considera o tamanho da aeronave, podendo distorcer a percepção de risco real.

Fatalidades por assento: obtida dividindo-se o total de fatalidades pelo número total de assentos das aeronaves daquele porte. Significado: indica o risco proporcional de morte por passageiro transportado. Vantagens: normaliza a análise, permitindo comparações justas entre portes com capacidades muito diferentes. Desvantagens: depende da qualidade do preenchimento do número de assentos no banco de dados. (Supõe ocupação máxima)

A análise dos dados revelou que aeronaves de pequeno porte apresentaram tanto a maior taxa de fatalidade por ocorrência (0,18) quanto a maior taxa de fatalidade por assento (0,0372), indicando maior risco e gravidade nos acidentes. Já as aeronaves de grande porte apresentaram os índices mais baixos, com 0,05 fatalidades por ocorrência e apenas 0,0004 por assento, evidenciando que, apesar de transportarem mais pessoas, são substancialmente mais seguras.

6.3.2 Quantidades de Fatalidades por ano

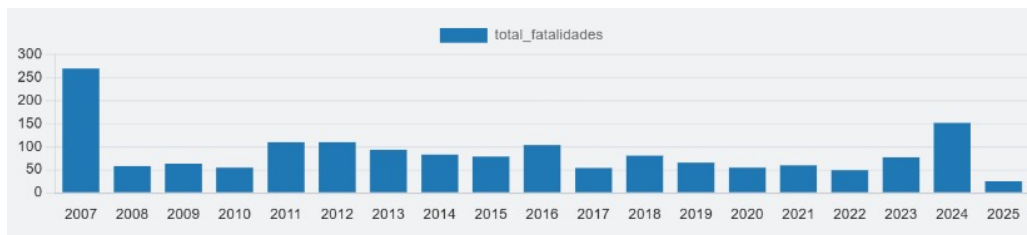


Figura 5: Evolução do total de fatalidades em acidentes aeronáuticos no Brasil entre 2007 e 2025.

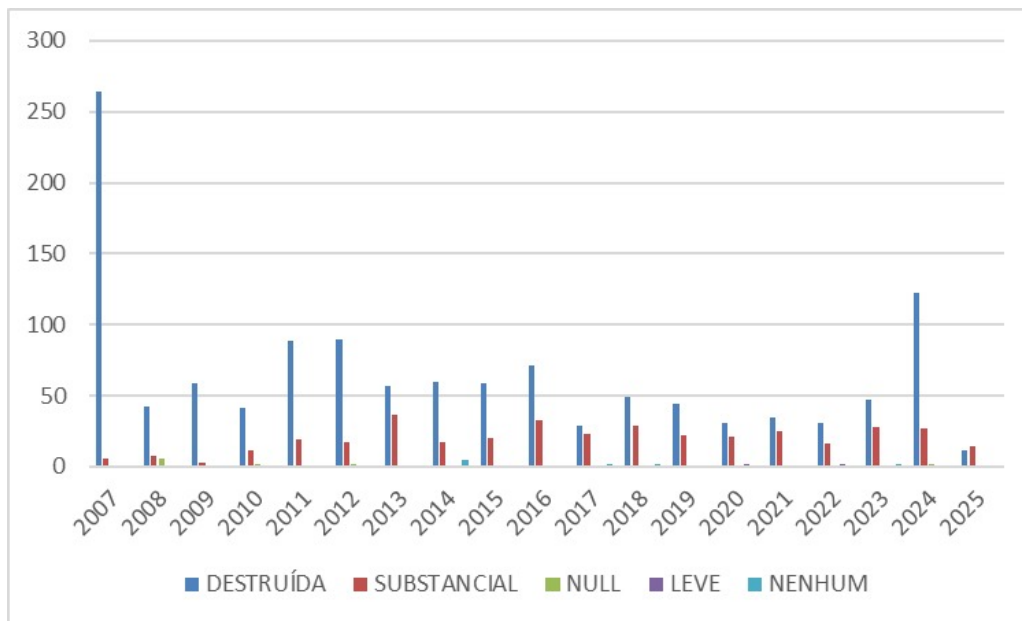


Figura 6: Distribuição dos tipos de dano sofridos pelas aeronaves em acidentes de 2007 a 2025.

Considerando o total de fatalidades por ano entre 2007 e 2025. Os registros foram agrupados por ano e somou-se o total de fatalidades em cada período. A visualização foi feita por meio de um gráfico de barras, que permite observar com clareza os anos com maior número de vítimas, como 2007, 2016 e 2024, além de destacar a tendência de redução ao longo dos anos. Além da análise de fatalidades, também foi avaliado o nível de dano sofrido pelas aeronaves envolvidas nos acidentes entre os anos de 2007 e 2025. Para isso, utilizamos os dados da tabela aeronaves, categorizando os danos em: Destruída, Substancial, Leve, Nenhum e valores ausentes (NULL). Os registros foram agrupados por ano e tipo de dano, permitindo observar a severidade dos acidentes ao longo do tempo. A categoria Destruída aparece de forma predominante em praticamente todos os anos, especialmente nos períodos com maior número de fatalidades, como 2007, 2011 e 2024, o que sugere correlação entre a destruição total da aeronave e o número elevado de mortes. Danos Substanciais também aparecem com frequência, enquanto os casos classificados como Leves ou Nenhum dano são bem menos comuns. Essa distribuição reforça a gravidade dos acidentes registrados e contribui para a compreensão do impacto estrutural nas aeronaves envolvidas.

6.3.3 Quantidades de fatalidades por estado

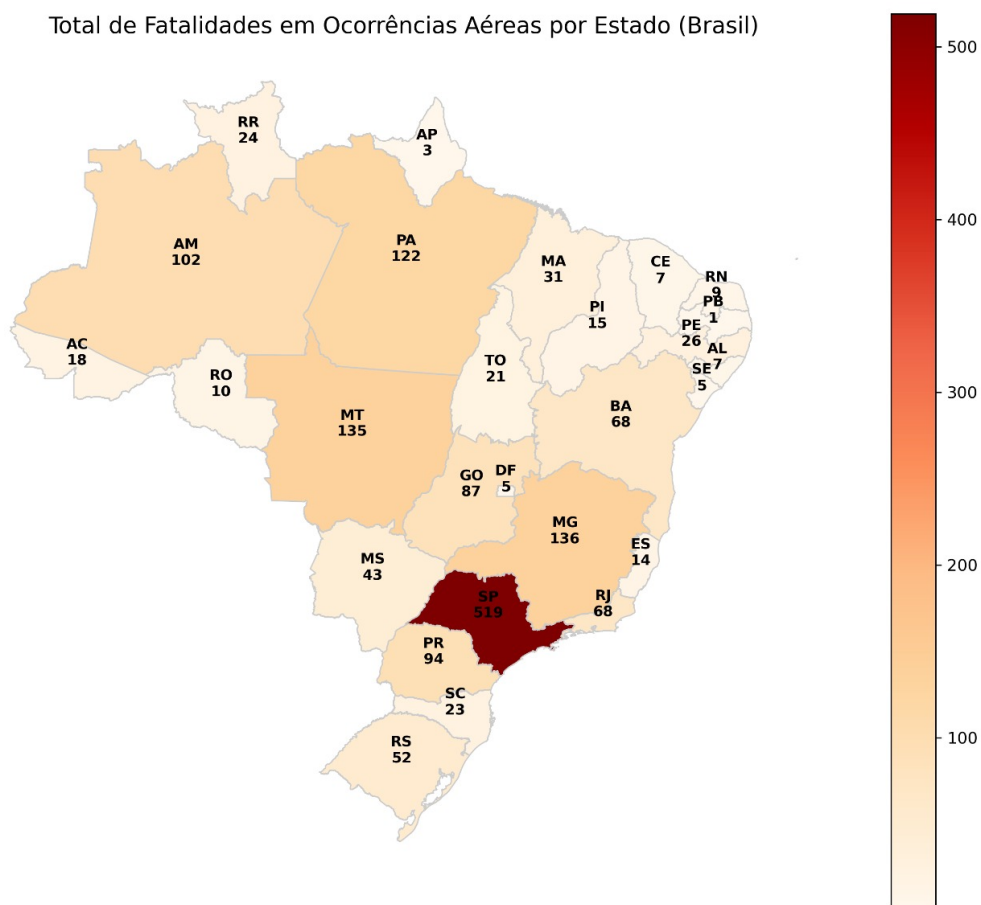


Figura 7: Distribuição geográfica das fatalidades por estado brasileiro.

O estado de São Paulo (SP) lidera com folga em número absoluto de fatalidades (519 mortes), seguido por Minas Gerais (MG), Mato Grosso (MT) e Pará (PA).

Estados do Norte e Centro-Oeste, como Amazonas (AM), Pará (PA) e Mato Grosso (MT) também registram números relevantes, sugerindo relação com voos regionais ou operacionais.

Estados com menor tráfego aéreo (ex: Roraima, Amapá, Acre) apresentam menos ocorrências fatais.

6.3.4 Fatores contribuintes para acidentes

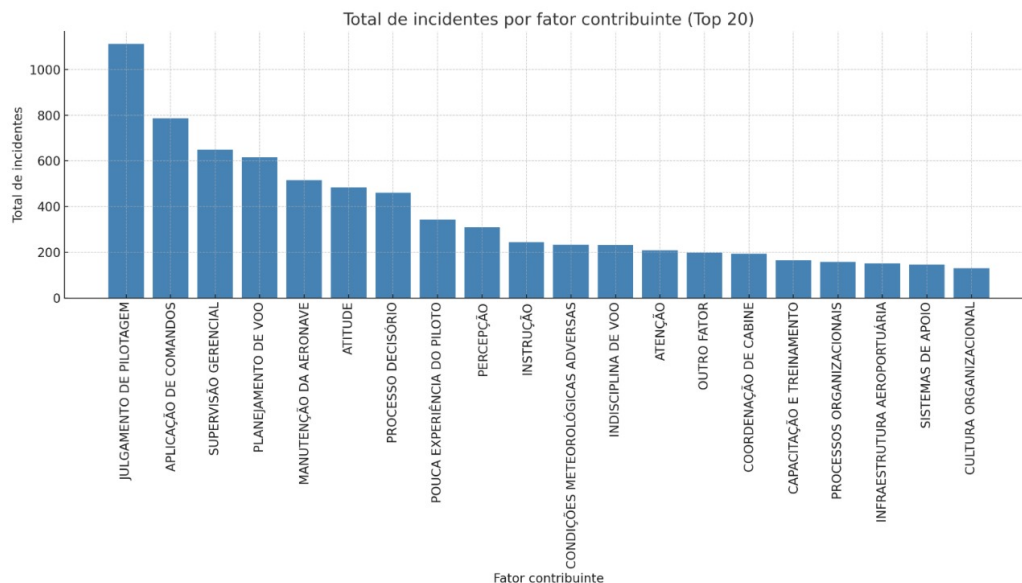


Figura 8: Top 20 fatores contribuintes mais frequentes em incidentes aéreos.

O objetivo principal da análise é entender o comportamento dos fatores contribuintes para incidentes, respondendo às seguintes perguntas: Quais fatores são mais frequentes? Existe concentração de incidentes em poucos fatores? É possível identificar padrões relacionados ao tipo de fator (humano, técnico, organizacional etc.)?

Esses dados mostram uma distribuição bastante desbalanceada, com poucos fatores concentrando a maior parte das ocorrências.

Existe uma forte concentração de ocorrências em poucos fatores. Fatores relacionados à atuação humana estão no topo da lista, o que reforça a importância da capacitação e do treinamento operacional. Falhas organizacionais, como supervisão e planejamento, também desempenham papel relevante.

6.3.5 Incidentes por tipos de motor e fase do voo

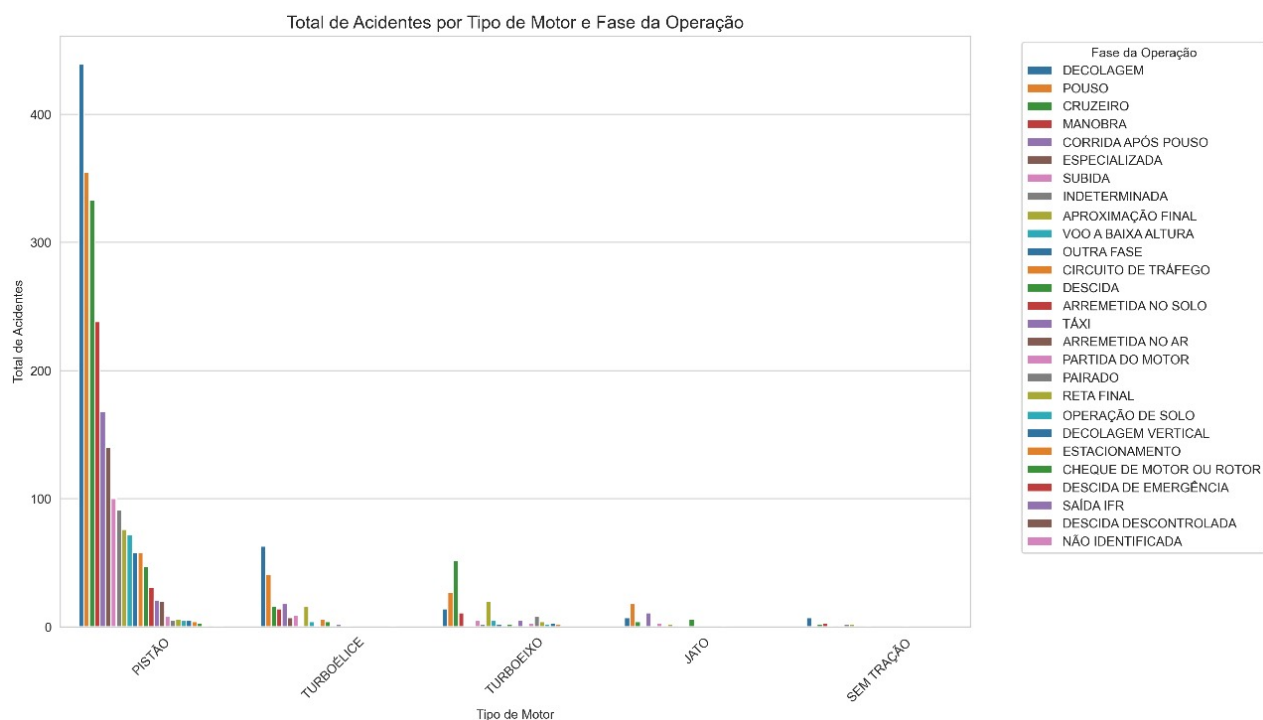


Figura 9: Relação entre o tipo de motor e as fases do voo com maior número de acidentes.

Para motores a pistão, a maioria dos acidentes ocorre em fases críticas como decolagem, pouso, cruzeiro e manobra, com dezenas de outras fases representadas.

Os motores turboélice e jato apresentam maior concentração em menos fases, indicando perfis operacionais mais estáveis ou controlados.

Algumas fases específicas como “subida”, “aproximação final” e “descida” também se destacam em tipos de motor mais avançados.

7 Conclusão

A análise dos dados de acidentes aeronáuticos no Brasil, com base nas informações do CENIPA, permitiu identificar padrões importantes sobre a gravidade, frequência e causas das ocorrências.

Aeronaves de pequeno porte apresentaram os maiores índices de fatalidade, tanto por ocorrência quanto por assento, destacando maior vulnerabilidade nesse tipo de operação.

No recorte temporal, observou-se uma tendência de redução nas fatalidades, embora com picos em anos específicos como 2007 e 2024.

Geograficamente, estados como São Paulo, Minas Gerais e Mato Grosso concentram a maior parte das fatalidades. Entre os fatores contribuintes, destaca-se a predominância de falhas humanas, o que reforça a necessidade de treinamento e melhoria na gestão operacional.

Por fim, a associação entre tipo de motor e fase do voo evidenciou que motores a pistão concentram mais incidentes em fases críticas, enquanto motores mais modernos tendem a operar de forma mais segura.

Esses resultados reforçam a importância de estratégias preventivas focadas em capacitação, manutenção e controle de risco, especialmente em operações regionais.

8 Referências

BRASIL. Centro de Investigação e Prevenção de Acidentes Aeronáuticos (CENIPA).

Base de Dados de Ocorrências Aeronáuticas da Aviação Civil Brasileira.

Disponível em: <https://dados.gov.br/dados/conjuntos-dados/>

ocorrencias-aeronauticas-da-aviacao-civil-brasileira. Acesso em: 20 jun. 2025.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS (ABNT). **NBR 6023:2023**

– **Informação e documentação – Referências – Elaboração.** Rio de Janeiro, 2023.

PANDAS DEVELOPMENT TEAM. **Pandas Documentation.** Disponível em:

<https://pandas.pydata.org/docs/>. Acesso em: 20 jun. 2025.