

Análise de dados
Ocorrências Aeronáuticas na Aviação Civil Brasileira

Tabela de ocorrências

Nome das colunas	Tipo de dados	Chave primária	Chave estrangeira	Código de identificação da ocorrência
codigo_ocorrendia	INT(11)	x	-	Código de identificação da ocorrência
classificacao	VARCHAR(30)	-	-	Classificação da ocorrência
tipo	VARCHAR(30)	-	-	Tipo da ocorrência
localidade	VARCHAR(100)	-	-	Cidade / município onde houve a ocorrência
uf	VARCHAR(2)	-	-	Estado / província onde houve a ocorrência
pais	VARCHAR(30)	-	-	País onde houve a ocorrência
aerodromo	VARCHAR(4)	-	-	Código ICAO do aeródromo onde houve a ocorrência
dia_ocorrendia	DATE	-	-	Data da ocorrência
horario_utc	TIME	-	-	Horário da ocorrência no padrão UTC
sera_investigada	VARCHAR(5)	-	-	Informação se a ocorrência será ou não investigada
comando_investigador	VARCHAR(15)	-	-	Comando investigador responsável pela ocorrência
status_investigacao	VARCHAR(10)	-	-	Informação se a investigação está ativa ou finalizada
numero_relatorio	VARCHAR(30)	-	-	Número de identificação do relatório final de investigação
relatorio_publicado	VARCHAR(5)	-	-	Informação se o relatório final de investigação foi ou não publicado
dia_publicacao	DATE	-	-	Data da publicação do relatório final de investigação
quantidade_recomendacoes	INT(11)	-	-	Quantidade de recomendações de segurança emitidas
aeronaves_envolvidas	INT(11)	-	-	Quantidade de aeronaves envolvidas na ocorrência
saida_pista	INT(11)	-	-	Informação se houve ou não saída de pista na ocorrência
dia_extracao	DATE	-	-	Data da extração dos dados na base de dados do CENIPA

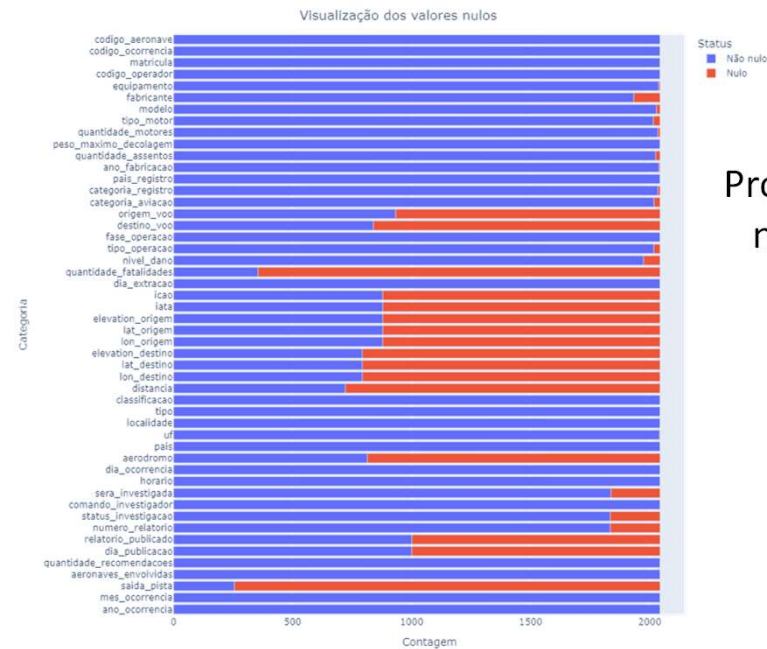
Tabela de aeronaves

Nome das colunas	Tipo de dados	Chave primária	Chave estrangeira	Comentários
codigo_aeronave	INT(11)	x	-	Código de identificação de aeronave
codigo_ocorrendia	INT(11)	-	x	Código de identificação de ocorrência
matricula	VARCHAR(10)	-	-	Matrícula da aeronave
codigo_operador	INT(11)	-	-	Código de identificação do operador
equipamento	VARCHAR(45)	-	-	Tipo da aeronave
fabricante	VARCHAR(45)	-	-	Fabricante da aeronave
modelo	VARCHAR(45)	-	-	Modelo da aeronave
tipo_motor	VARCHAR(45)	-	-	Tipo de motor da aeronave
quantidade_motores	INT(11)	-	-	Quantidade de motores da aeronave
peso_maximo_decolagem	FLOAT	-	-	Peso máximo para decolagem
quantidade_assentos	INT(11)	-	-	Quantidade de assentos na aeronave
ano_fabricacao	INT(11)	-	-	Ano de fabricação da aeronave
pais_registro	VARCHAR(30)	-	-	País de registro da aeronave
categoria_registro	VARCHAR(4)	-	-	Categoria de registro da aeronave no momento da ocorrência
categoria_aviao	VARCHAR(15)	-	-	Categoria de registro da aeronave no momento da ocorrência
origem_voo	VARCHAR(4)	-	-	Origem de voo da aeronave
destino_voo	VARCHAR(4)	-	-	Destino de voo da aeronave
fase_operacao	VARCHAR(15)	-	-	Fase de operação da aeronave no momento da ocorrência
tipo_operacao	VARCHAR(20)	-	-	Tipo de operação da aeronave no momento da ocorrência
nivel_dano	VARCHAR(10)	-	-	Nível do dano da aeronave
quantidade_fatalidades	INT(11)	-	-	Quantidade de fatalidades (mortos) na aeronave
dia_extracao	DATE	-	-	Data da extração dos dados na base de dados do CENIPA

Dados dos aeroportos

<https://github.com/mwgg/Airports>

Os dados dos aeroportos foram coletados e incluídos na base de dados para obter elevação, longitude e latitude.



Proporção de valores nulos e não nulos do conjunto de dados

O que é IATA?

IATA é a sigla inglesa de International Air Transport Association ou Associação Internacional de Transportes Aéreos, em português. A IATA foi criada há mais de 60 anos por um grupo de companhias aéreas, com o objetivo de representá-las em todos os assuntos relacionados à aviação.

O que é ICAO?

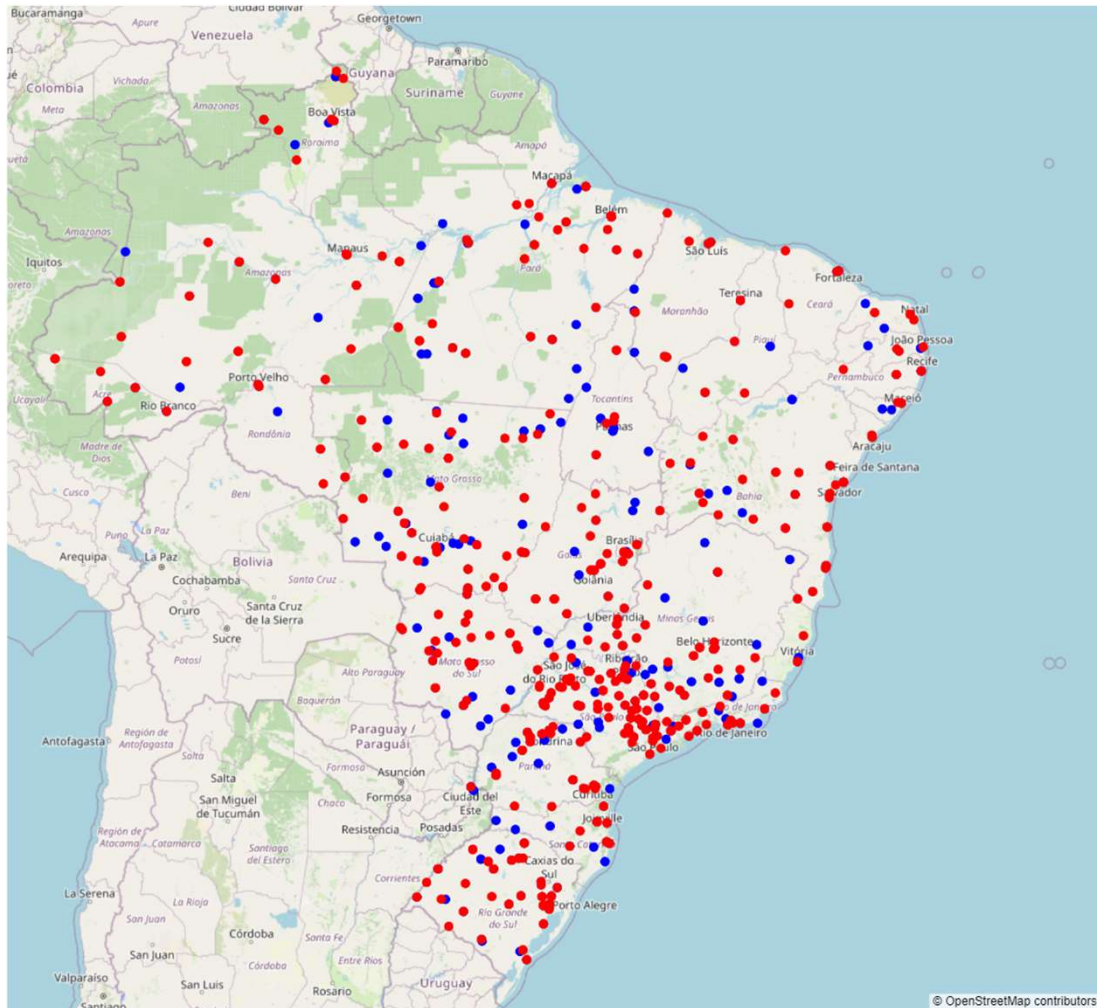
A ICAO (sigla em inglês para Organização da Aviação Civil Internacional) criou um código de 4 letras exclusivo para cada aeroporto, visando padronizar e facilitar a identificação. A 1ª letra designa a região ICAO (S para a América do Sul), a 2ª informa o país (B/D/I/J/N/S/W estão ### reservadas para o Brasil) e as outras duas letras completam o código do aeroporto propriamente dito. Todos os aeroportos com código começado por SB produzem METAR – informe meteorológico regular. A IATA (sigla em inglês para Associação Internacional de Transportes Aéreos) também criou um ### código de 3 letras com a mesma finalidade (nesse caso, representando destinos de companhias áreas membros da associação).

Os dados foram incluídos utilizando a sigla ICAO que é referente a cada aeroporto. Essa sigla pode ser encontrada na tabela aeronaves nas colunas origem_voo e destino_voo.

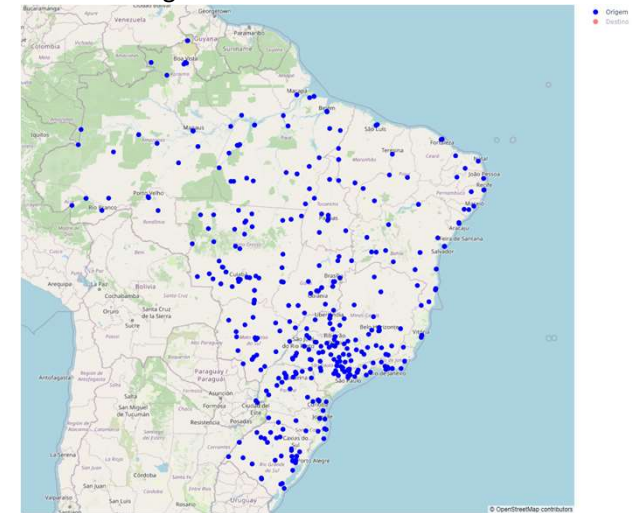
	origem_voo	destino_voo
	SDPW	SDPW I
	SBBP	SBBP
	****	****

Aeroportos de origem e destino dos incidentes do conjunto de dados

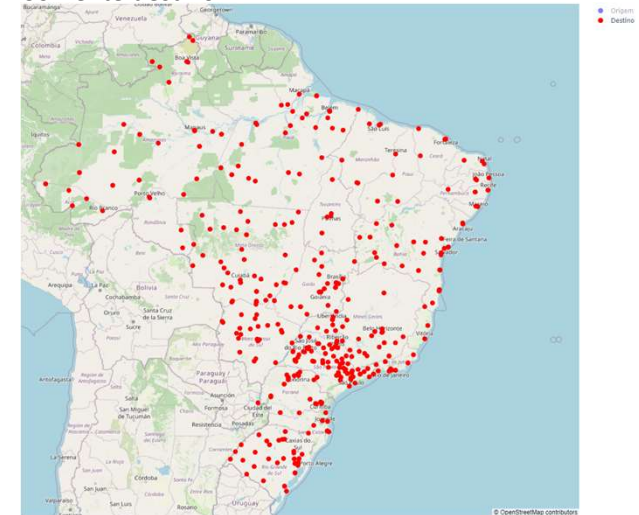
É possível verificar um comportamento esperado onde a região sudeste e sul possuem maiores concentrações. Os aeroportos são mais próximos devido ao menor espaço territorial comparado ao norte e nordeste. Contudo, nenhuma grande concentração fora do normal foi detectado visualmente.



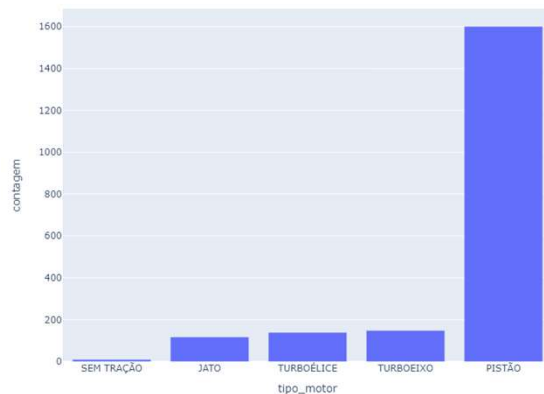
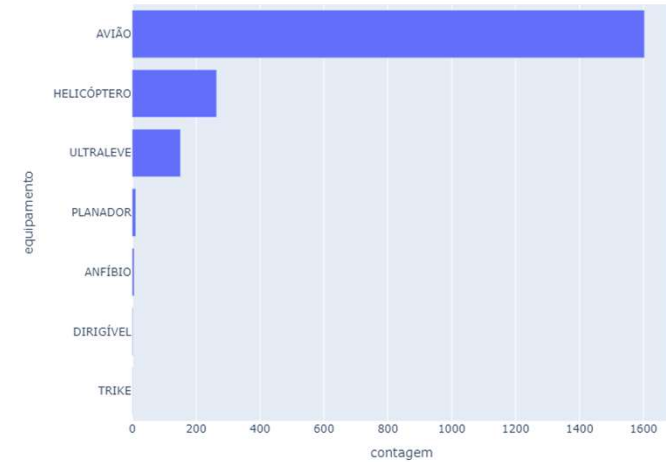
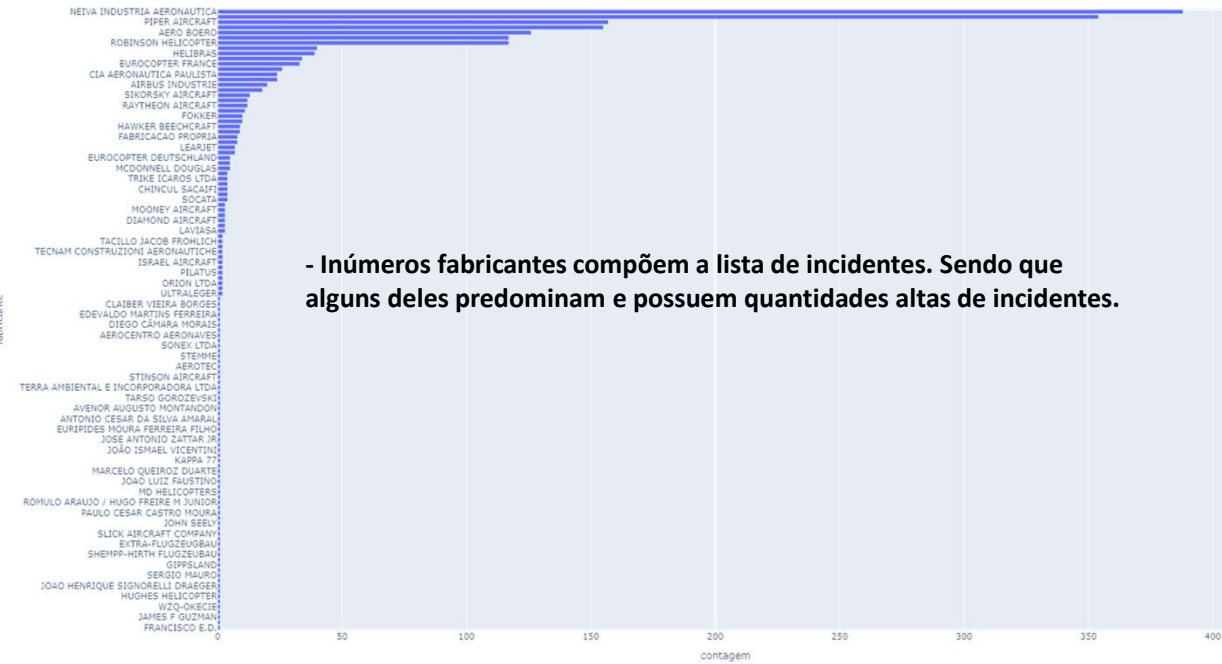
Somente origem



Somente destino



Análise Univariada



Análise Univariada



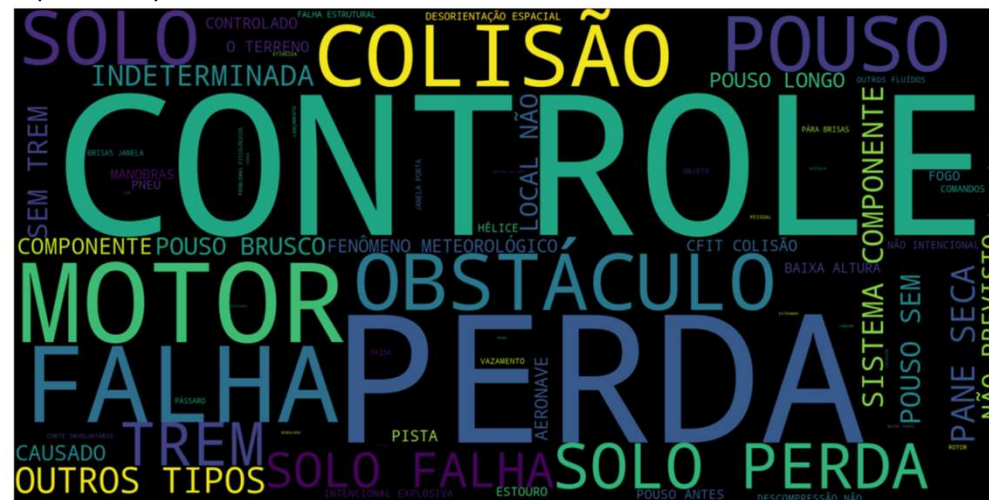
Se somarmos as 3 principais causas de acidente chegamos a praticamente 50% das causas de acidentes

- Falha do motor = 18.45%
- Perda de controle no solo = 15.91%
- Perda de controle em voo = 15.22%

Total = 49.58%

FALHA DO MOTOR EM VOO	18.45
PERDA DE CONTROLE NO SOLO	15.91
PERDA DE CONTROLE EM VOO	15.22
COLISÃO EM VOO COM OBSTÁCULO	6.95
COM TIPO DE POUSO	6.36
OUTROS TIPOS	4.26
PARA SECA	3.62
FALHA DE SISTEMA / COMPONENTE	2.99
INDETERMINADA	2.79
POUSO SEM TIPO	2.44
POUSO EM LOCAL NÃO PREVISTO	2.13
COLISÃO COM OBSTÁCULO NO SOLO	1.96
POUSO BRUSCO	1.96
POUSO LENTO	1.97
CAUSADO POR FENÔMENO METEOROLÓGICO EM VOO	1.52
PERDA DE COMPONENTE EM VOO	1.32
CFIT - COLISÃO EM VOO CONTROLADO COM O TERRENO	1.40
MANOBRAS A BAIXA ALTURA	0.78
POUSO ANTES DA PISTA	0.73
ESTOURO DE PNEU	0.69
COLISÃO COM AERONAVE NO SOLO	0.69
TRÁFEGO AÉREO	0.64
COLISÃO DE AERONAVES EM VOO	0.59
COM HÉLICE	0.54
DESORIENTAÇÃO ESPACIAL	0.49
DESCOMPRESSIONAMENTO INTENCIONAL / EXPLOSIVA	0.44
COM COMANDOS DE VOO	0.39
POUSO EM VOO	0.39
FALHA ESTRUTURAL	0.34
POUSO NO SOLO	0.34
VAZAMENTO DE OUTROS FLUIDOS	0.29
PERDA DE COMPONENTE NO SOLO	0.29
COM PARA-BRISAS / JANELA / PORTA	0.24
SÁDIA DE PISTA	0.24
COLISÃO COM PASSAGEIRO	0.24
COM ROTOR	0.15
DIRIGIR INDEBIDAMENTE NO MOTOR	0.15
CAUSADO POR FENÔMENO METEOROLÓGICO NO SOLO	0.15
FALHA DO MOTOR NO SOLO	0.15
AERONAVE ATINGIDA POR OBJETO	0.10
COM PESSOAL EM VOO	0.10
INCIDENTES FÍSICO-OLÍTICOS	0.10
COM LANÇAMENTO DE PESSOAS	0.05
PARA-CHOQUE	0.05
QTD - DANO CAUSADO POR OBJETO ESTRANHO	0.05
COM LANÇAMENTO DE CORREIA	0.05
COLISÃO EM VOO COM OBJETO REBOCADOR	0.05
INCUSÃO EM PISTA	0.05
COLISÃO DE MÉTODOS COM AERONAVE	0.05

Uma nuvem de palavras foi gerada para analisar os tipos de falhas. As falhas são descritas com um texto curto. As palavras controle, perda, falha, motor, colisão e obstáculo se destacaram entre todas as palavras que fazem parte da nuvem.



Análise Bivariada

Verificando-se as falhas de motor em voo, observou-se casos que consta a quantidade de motores igual a zero. Quando olha-se esses casos e através de uma pesquisa na internet, pode-se detectar que são aviões com assentos. Alguns exemplos foram coletados para fazer essa verificação. Isso caracteriza-se como uma inconsistência existente na base de dados.

Cessna 210 é terceiro modelo de avião que mais se envolve em acidentes

Por Rafael Ignácio 10/11/2016 18h01. Atualizado em 10/11/2016 18h01

Recomendado

Aviões monomotores do modelo Cessna 210 foram responsáveis por 5,2% dos acidentes aéreos no Brasil nos últimos 10 anos. O modelo é o mesmo da aeronave que caiu nesta sexta-feira (30) na zona norte de São Paulo.

Extra EA-300	
	
Extra 300 pilotado por Péter Besenyi.	
Descrição	
Tipo / Missão	Aeronave de acrobacias
País de origem	 Alemanha
Fabricante	Extra Aircraft
Primeiro voo em	maio de 1988 (34 anos)
Variantes	300S 330SX 300SP 300SHF 300SR 300L 300LP 330SC 330LX 330LT
Tripulação	1/2
Carga útil	270 kg (595 lb)

Airbus A340	
	
Airbus A340-300 da Lufthansa	
Descrição	
País de origem	 Alemanha  França  Reino Unido  Espanha
Fabricante	Airbus Construcciones Aeronáuticas S.A.
Período de produção	1991–2011
Quantidade produzida	377
Custo unitário	A340-300: US\$ 211 milhões A340-500: US\$ 233 milhões A340-600: US\$ 245 milhões
Desenvolvido de	Airbus A300
Primeiro voo em	25 de outubro de 1991 (31 anos)
Introduzido em	15 de março de 1993 (30 anos) com a Air France
Tripulação	2 (piloto e co-piloto)
Passageiros	380
Número de classes	3 classe(s)
Especificações	
Dimensões	

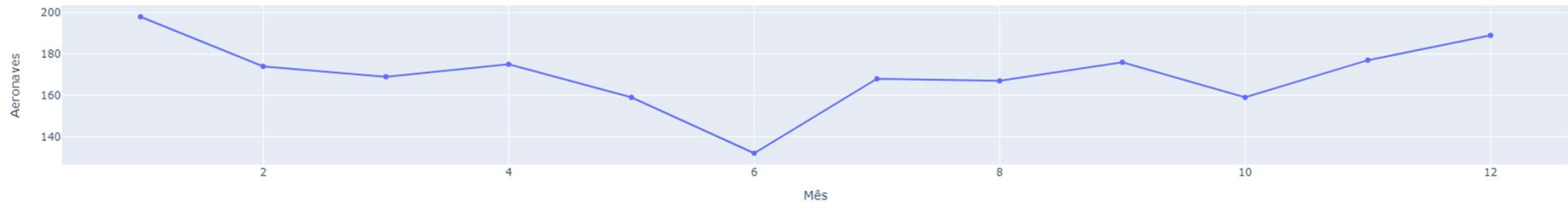
ULTRALEVE	BRAVO 700	AEROBRAVO LTDA	1
	CONQUEST 180 LSA	INPAER INDUSTRIA PAULISTA E AERONÁUTICA	1
	FOX V5 SUPER	FABRICACAO PROPRIA	1
	TRIKE	FABRICACAO PROPRIA	1
AVIÃO	MD11	MCDONNELL DOUGLAS	1
	100	BEECH AIRCRAFT	1
	EMB145	EMBRAER	1
	767-328	BOEING COMPANY	1
	150M	CESSNA AIRCRAFT	1
	152	CESSNA AIRCRAFT	1
	182A	CESSNA AIRCRAFT	1
	182T	CESSNA AIRCRAFT	1
	210L	CESSNA AIRCRAFT	1
	210M	CESSNA AIRCRAFT	1
	210N	CESSNA AIRCRAFT	1
	402B	CESSNA AIRCRAFT	1
	525	CESSNA AIRCRAFT	1
	747-312M	BOEING COMPANY	1
	767-3P6	BOEING COMPANY	1
	114	ROCKWELL	1
	95-B55	BEECH AIRCRAFT	1
	A-37B	CESSNA AIRCRAFT	1
	A330-243	AIRBUS INDUSTRIE	1
	A340	AIRBUS INDUSTRIE	1
	A35	BEECH AIRCRAFT	1
	A36	BEECH AIRCRAFT	1
	C206	CESSNA AIRCRAFT	1
	CMA	FABRICACAO PROPRIA	1
	DC-10-30	MCDONNELL DOUGLAS	1
	EA 300/L	EXTRA-FLUGZEUGBAU	1
	EMB-195	EMBRAER	1

equipamento	modelo	fabricante	
ULTRALEVE	ULTRALEVE	MAULE AIRCRAFT	3
	EXPERIMENT	FABRICACAO PROPRIA	3
	210	CESSNA AIRCRAFT	3
	EMB-505	EMBRAER	2
	172N	CESSNA AIRCRAFT	2
	767-322	BOEING COMPANY	2
	A340-541	AIRBUS INDUSTRIE	2
	58	BEECH AIRCRAFT	2
	PA-WNE	PIPER AIRCRAFT	1
	PA34-200	PIPER AIRCRAFT	1
AVIÃO	PA34	PIPER AIRCRAFT	1
	PA-34-220T	PIPER AIRCRAFT	1
	PA-42	PIPER AIRCRAFT	1
	RANS S-10-AVIAO	FABRICACAO PROPRIA	1
	PA-34-200	PIPER AIRCRAFT	1
	PA-23	PIPER AIRCRAFT	1
	MD82	MCDONNELL DOUGLAS	1
	R182	CESSNA AIRCRAFT	1
	T206H	CESSNA AIRCRAFT	1
	RV-10	FLYER INDUSTRIA AERONAUTICA LTDA	1
HELICÓPTERO	RV6	FLYER INDUSTRIA AERONAUTICA LTDA	1
	G58	BEECH AIRCRAFT	1
	T210L	CESSNA AIRCRAFT	1
	TU206G	CESSNA AIRCRAFT	1
	206	BELL HELICOPTER	1
	EC 225 LP	EUROCOPTER FRANCE	1
	NIMBUS 3D PLAN.	SHEMPPP-HIRTH FLUGZEUBAU	1
	BRAVO 700	AEROBRAVO LTDA	1
	CONQUEST 180 LSA	INPAER INDUSTRIA PAULISTA E AERONÁUTICA	1
	FOX V5 SUPER	FABRICACAO PROPRIA	1

Análise Bivariada

Os meses da baixa temporada com passagens mais baratas são março, abril e agosto. Já a alta temporada ocorre em janeiro, julho e dezembro devido as férias escolares. Entre setembro e novembro, é possível encontrar voos baratos também. O mês de junho apresenta o menor patamar de incidentes que possivelmente se deve ao número de voos neste mês.

Série temporal - Aeronaves x Mês de ocorrência

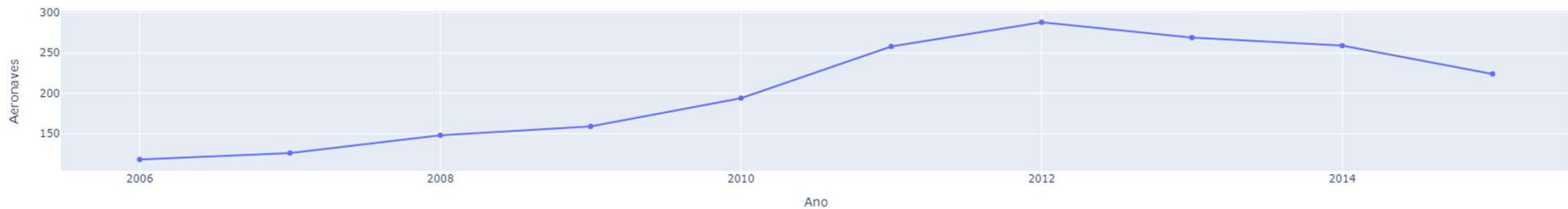


Número de passageiros aéreos cresceu 210% de 2000 a 2014, diz CNT

https://www.correiobraziliense.com.br/app/noticia/turismo/2015/11/11/interna_turismo,506012/numero-de-passageiros-aereos-cresceu-210-de-2000-a-2014-diz-cnt.shtml

Baseado nisso pode-se avaliar que o crescimento de passageiros e de aeronaves tenha gerado um crescimento nos acidentes também. A série temporal mostra um crescimento entre 2006 (início dos dados) e 2012, posteriormente um declínio até 2015.

Série temporal - Aeronaves x Ano de ocorrência



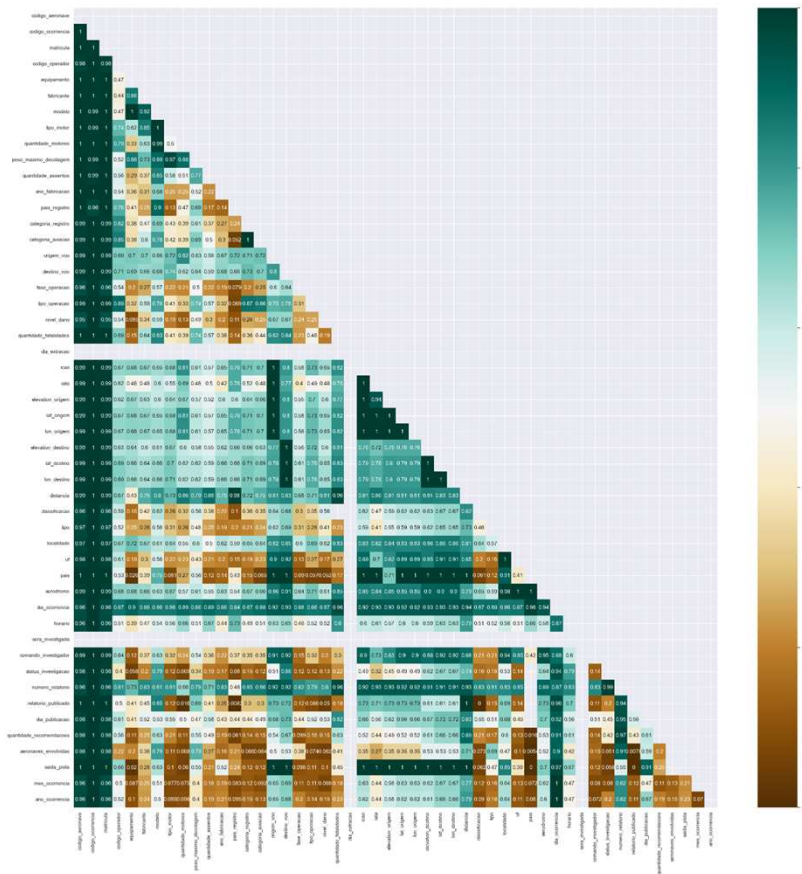
Análise Bivariada

O nível de dano dos incidentes foi analisado em relação a elevação, longitude e latitude dos aeroportos e nenhum padrão foi detectado. Mesmo considerando outras variáveis, esses valores estão nos mesmos intervalos. Talvez se tivéssemos essas informações referentes ao momento do incidente poderiam ser mais úteis do que a dos aeroportos.



Análise Multivariada

V de crammer



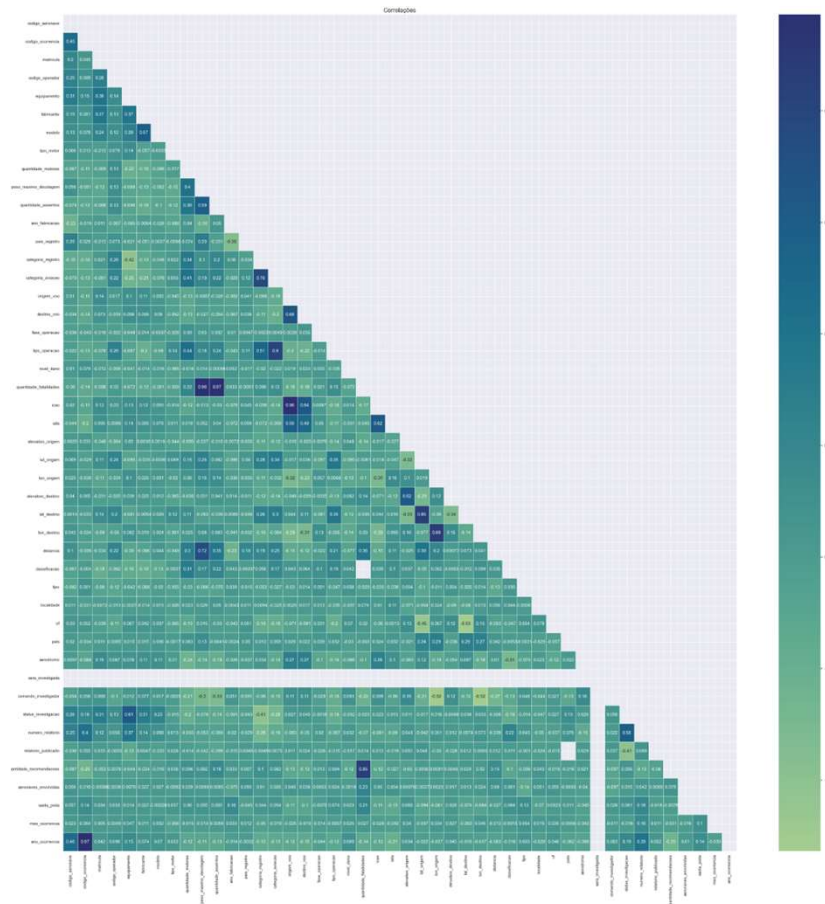
Utilizou-se o método V de crammer para avaliar a correlação das variáveis. Os valores não foram satisfatórios. Contudo, é importante reportar que isto foi feito.

Embora a correlação V de Cramer seja uma métrica útil para medir a associação entre variáveis categóricas, há situações em que ela pode não ser adequada ou precisa.

- Tamanho da amostra: se a amostra for muito pequena, a correlação V de Cramer pode não ser confiável. É importante lembrar que a correlação é baseada em frequências observadas, e se as frequências forem muito baixas, a correlação pode ser afetada. A base de dados possui variáveis com categorias com baixa frequência.
- Variáveis com muitas categorias: se as variáveis categóricas tiverem muitas categorias, a correlação V de Cramer pode não ser tão informativa. Isso ocorre porque a matriz de contingência pode ficar muito grande e as células com poucas observações podem prejudicar a precisão da medida. A base de dados possui muitas variáveis com grande número de categorias.

Análise Multivariada

Correlação de Pearson



Foi realizado o encoding das variáveis categóricas e posteriormente verificada a correlação de Pearson.

Os resultados se mostraram mais coerentes, contudo, alguns testes de hipóteses foram realizados posteriormente para avaliar algumas situações.

A correlação de Spearman foi verificada e apresentou coeficientes um pouco mais altos para as correlações.

Testes de hipóteses

Como não foi fornecida a informação se a base de dados trata-se de toda a população de incidentes de aeronaves ou se é uma amostra, optou-se por realizar os testes de hipóteses. Optou-se por esse teste por ser recomendado para variáveis categóricas, a base de dados possui variáveis categóricas em grande parte.

Hipóteses

- **Primeira hipótese: Existe associação entre categoria e a classificação do acidente.**
- **Segunda hipótese: Existe associação entre o ano de fabricação e a classificação do acidente.**
- **Terceira hipótese: Existe associação entre o tipo de operação e o tipo de dano.**
- **Quarta hipótese: Existe associação entre o tipo de operação e o nível de dano.**
- **Quinta hipótese: Existe associação entre o fabricante e o nível de dano.**
- **Sexta hipótese: Existe associação entre o fabricante e a classificação do acidente.**

Foi utilizado o teste qui-quadrado para estas hipóteses.

chi2: 248.9196748335234
Valor de p: 4.703422789056981e-47
A hipótese nula foi rejeitada
A categoria de aviação possui influência na classificação do acidente

chi2: 94.43149108541286
Valor de p: 0.05485450079751537
A hipótese alternativa foi rejeitada
Não existe associação entre o ano de fabricação a a classificação do acidente.

chi2: 1308.5202241345419
Valor de p: 3.5432081501226214e-101
A hipótese nula foi rejeitada
Existe associação entre o tipo de operação e o tipo de dano.

chi2: 371.7280993827219
Valor de p: 4.640908314436517e-64
A hipótese nula foi rejeitada
Existe associação entre o tipo de operação e o nível de dano.

chi2: 663.9552805104454
Valor de p: 3.695231225195525e-21
A hipótese nula foi rejeitada
Existe associação entre o fabricante e o nível de dano.

chi2: 333.8182364016943
Valor de p: 1.6916751859575533e-22
A hipótese nula foi rejeitada
Existe associação entre o fabricante e a classificação do acidente.

Conclusões

Inconsistências:

- Existe aeronaves de pequeno e grande porte que constam como não possuírem assentos. Ao pesquisar alguns desses modelos na internet, comprovou-se que isso não é verdade.

Constatações:

- As 3 principais causas de acidente somadas chegam a praticamente 50% dos incidentes.
- Ao longo dos anos a quantidade de viagens de avião cresceu e por consequência os desastres.
- O mês de junho possui queda nas viagens segundo informações coletadas e isso se refletiu no gráfico de séries temporais dos incidentes.
- A distribuição de acidentes considerando os locais de destino e origem dos voos, não demonstraram nenhuma alta concentração em local específico no mapa.

Resumo das hipóteses:

- Primeira hipótese: Existe associação entre categoria e a classificação do acidente. (CONFIRMADA)
- Segunda hipótese: Existe associação entre o ano de fabricação e a classificação do acidente. (REJEITADA)
- Terceira hipótese: Existe associação entre o tipo de operação e o tipo de dano. (CONFIRMADA)
- Quarta hipótese: Existe associação entre o tipo de operação e o nível de dano. (CONFIRMADA)
- Quinta hipótese: Existe associação entre o fabricante e o nível de dano. (CONFIRMADA)
- Sexta hipótese: Existe associação entre o fabricante e a classificação do acidente. (CONFIRMADA)

Mesmo as variáveis tendo apresentado baixa correlação de Pearson ou Spearman, elas ainda podem apresentar associação significativa de acordo com o teste do qui-quadrado. Isso ocorre porque o teste do qui-quadrado leva em consideração a frequência de ocorrência conjunta das variáveis em questão, enquanto a correlação de Pearson ou Spearman avalia apenas a relação linear e não-linear entre elas.

Informações relacionadas a manutenção das aeronaves e sobre os voos que não sofreram acidente poderiam contribuir para enriquecer a análise dos dados.