
Análise Exploratória de Dados de Acidentes Aeronáuticos no Brasil

João Paulo de Andrade Dantas ¹, Jelton Alexandre da Cunha ¹, Jamesson Lira Silva ¹, Alessandro Oliveira Arantes ¹, Vitor Conrado Faria Gomes ¹

¹ Instituto de Estudos Avançados – IEAv

BIOGRAFIA:

O Primeiro-Tenente Engenheiro João Paulo de Andrade Dantas concluiu em 2015 a graduação em Engenharia Mecânica-Aeronáutica pelo Instituto Tecnológico de Aeronáutica (ITA). Atualmente é aluno de mestrado no Programa de Pós-graduação em Engenharia Eletrônica e Computação (PG/EEC) do ITA e pesquisador do Instituto de Estudos Avançados (IEAv) no Grupo de Ciência de Dados da Divisão de C4ISR. Contato: dantas@ieav.cta.br, (12) 3947-5301.

O Primeiro-Tenente Engenheiro Jelton Alexandre da Cunha concluiu a graduação em Engenharia de Computação pela ETEP Faculdades, em 2011. Atualmente é aluno de mestrado no Programa de Pós-graduação em Ciências e Tecnologia Espaciais (PG/CTE) do Instituto Tecnológico de Aeronáutica (ITA), e pesquisador do Instituto de Estudos Avançados (IEAv), no grupo de Ciência de Dados da Divisão de C4ISR. Contato: jelton@ieav.cta.br, (12) 3947-5301.

O Primeiro-Tenente Fotógrafo Jamesson Lira Silva concluiu em 2015 a graduação em Fotointeligência pelo Centro de Instrução e Adaptação da Aeronáutica (CIAAR). Atualmente é aluno de mestrado no Programa de Pós-graduação em Ciências e Tecnologia Espaciais (PG/CTE) do Instituto Tecnológico de Aeronáutica (ITA), e pesquisador do Instituto de Estudos Avançados (IEAv), no grupo de Ciência de Dados da Divisão de C4ISR. Contato: jamessonls@ieav.cta.br, (12) 3947-5301.

O DSc. Alessandro Oliveira Arantes concluiu o mestrado em 2008 e o doutorado em 2016, ambos no curso de Computação Aplicada (CAP) pelo Instituto Nacional de Pesquisas Espaciais (INPE). Atualmente é tecnologista no Instituto de Estudos Avançados (IEAv), atuando no grupo de Ciência de Dados da Divisão de C4ISR. Contato: aarantes@ieav.cta.br, (12) 3947-5312.

O MSc. Vitor Conrado Faria Gomes concluiu o mestrado em 2012 e atualmente é aluno de doutorado no curso de Computação Aplicada (CAP) do Instituto Nacional de Pesquisas Espaciais (INPE). Atualmente é pesquisador no Instituto de Estudos Avançados (IEAv), no Grupo de Ciência de Dados da Divisão de C4ISR. Contato: vitor@ieav.cta.br, (12) 3947-5301.

RESUMO: O artigo apresenta um ambiente para análise exploratória dos dados abertos disponibilizados pelo Centro de Investigação e Prevenção de Acidentes Aeronáuticos (CENIPA), onde constam 6.238 ocorrências aeronáuticas notificadas entre os anos de 2006 e 2017. Utilizou-se, para tais estudos, tecnologias abertas que suportam o armazenamento, o acesso e a análise exploratória de dados. Além disso, o ambiente é estruturado de forma a permitir a realização de análise colaborativa e reprodutível. A partir da estruturação do ambiente, o processo ocorreu inicialmente com a leitura dos conjuntos de dados seguido da limpeza e adequação dos mesmos, possibilitando a análise por métodos computacionais. Por fim, foram realizadas visualizações gráficas dos dados, as quais permitiram responder questões acerca da progressão dos acidentes nos últimos anos, e ainda identificar os tipos de ocorrência mais frequentes. Foram verificados fatores como progressão temporal, localidades com maior concentração, dias e horários de maior incidência, fases de voo com momentos mais críticos e, ainda, a análise quantitativa de acidentes com fatalidades. Com isso, foi possível correlacionar essas informações e traçar um perfil de voo com mais frequência de acidentes. O ambiente pode ser utilizado em análises exploratórias de dados a fim do analista compreender melhor seus dados e identificar relações entre as variáveis observadas.

Palavras Chave: Acidentes Aeronáuticos, Ciência de Dados, Análise Exploratória de Dados, Ciência Reprodutível.

Exploratory Analysis of Data on Aeronautical Accidents in Brazil

ABSTRACT: This article presents an environment for exploratory analysis of open data made available by the Brazilian Aeronautical Accident Investigation and Prevention Center (CENIPA), where there are 6,238 aeronautical occurrences reported between 2006 and 2017. The studies have made use of open technologies which support storage, access and exploratory analysis of data. In addition, the environment is structured so as to allow collaborative and reproducible analysis. From the structuring of the environment, the process started with the reading of data sets followed by their cleaning and adequation, allowing analysis by computational methods. Finally, graphical visualizations of the data were carried out, allowing to answer questions about the progression of the accidents in the last years, and to identify the most frequent types of occurrence. The study verified factors such as temporal progression, locations with higher concentration, days and times of higher incidence, flight phases with more critical moments, and quantitative analysis of accidents with fatalities. Thus, it was possible to correlate the pieces of information, and to draw a flight profile with the most frequent accidents. The environment can be used in exploratory data analysis so that the analyst can better understand the data available, and identify relationships between the observed variables.

Keywords: Aeronautical Accidents. Data Science. Exploratory Data-Analysis. Reproducible Science.

Citação: Dantas, JPA, Cunha, JA, Silva, JL, Arantes, AO, Gomes, VCF. (2018) Análise Exploratória de Dados de Acidentes Aeronáuticos no Brasil. *Revista Conexão Sipaer*, Vol. 9, No. 2, pp. 106-127

1 INTRODUÇÃO

Nos últimos anos, a quantidade disponível de dados tem crescido, motivados por políticas públicas de disseminação de informação e impulsionados pelos avanços tecnológicos em instrumentos de aquisição e sistemas computacionais de armazenamento e difusão (GRUS, 2015; EMC, 2015).

A disponibilização dessas bases de dados tem grande valor para a compreensão de fenômenos naturais, sociais, econômicos e políticos. No entanto, a extração eficiente de informações desses dados não tem acompanhado o mesmo ritmo. Os desafios muitas vezes encontrados por usuários desses dados envolvem questões como capacidade computacional para armazenamento e processamento, ou a dificuldade em obter ou escolher ferramentas para a realização das análises. Frequentemente, o volume, a densidade, e a dimensionalidade também são fatores críticos que dificultam o processo de extração de informações pelo analista (Câmara et al., 2016).

A Análise Exploratória de Dados (EDA, do inglês *Exploratory Data Analysis*) é uma abordagem para analisar conjuntos de dados para extrair suas principais características, muitas vezes, por métodos visuais (EMC, 2015). O objetivo principal é fornecer ao cientista uma primeira compreensão dos dados e, possivelmente, fornecer meios para formular hipóteses ou avaliar a necessidade de coletas de novos dados.

Dada a diversidade de opções disponíveis, documentar a experiência na seleção de tecnologias, técnicas e métodos de visualização de dados ajuda pesquisadores a avaliar como diferentes propostas podem atender suas necessidades de análise de dados (Câmara et al., 2016).

Nesse sentido, esse trabalho apresenta a exploração de um conjunto de dados sobre ocorrências aéreas disponibilizado pelo Centro de Investigação e Prevenção de Acidentes Aeronáuticos (CENIPA). A exploração desses dados tem grande valor para a compreensão do panorama histórico desses eventos.

No restante deste trabalho, inicialmente é apresentada a motivação para o uso de EDA aplicada a dados de ocorrências aéreas. Na sequência, é apresentada a metodologia utilizada neste trabalho, onde são apresentados os dados e o ambiente utilizado para a realização da EDA. A seção seguinte é destinada para a apresentação da análise geral dos dados. Essa seção é subdividida em diferentes domínios utilizados para a exploração. Antes de finalizar esse documento, é apresentado um estudo de caso específico, onde são avaliadas informações a respeito do período do dia, dia da semana e porte da aeronave, para caracterizar um perfil de maior ocorrência de acidentes aéreos. Por fim, são realizadas as considerações finais.

1.1 Motivação

Para que se possa pensar em medidas que auxiliem a prevenção de acidentes aeronáuticos, é importante se conhecer o panorama histórico dos eventos ocorridos. A motivação deste trabalho é analisar diversos aspectos relacionados aos acidentes aéreos com o intuito de se identificar possíveis padrões e revelar cenários com maior frequência de ocorrências aeronáuticas, baseado nos dados históricos disponibilizados pelo CENIPA. Todavia, não é a pretensão deste trabalho elaborar qualquer recomendação ou norma visando a prevenção de acidentes com base nas informações inferidas pela EDA.

A motivação deste artigo é utilizar um conjunto de técnicas de Ciência de Dados para designar ferramentas, tecnologias, e métodos reprodutíveis capazes de analisar, visualizar, e reconhecer tendências escondidas nos dados que tenham influência na frequência dos acidentes aéreos, como, por exemplo, seus locais, principais tipos de ocorrências, horários, e até mesmo os tipos de aeronaves envolvidas.

2 METODOLOGIA

2.1 Ambiente para Análise Exploratória de Dados

Inicialmente, foi preparado um ambiente de análise para a realização da EDA. Visando a estruturação desse ambiente e a seleção das tecnologias, considerou-se a natureza dos dados e as necessidades dos cientistas. Esses profissionais são conservadores na escolha das ferramentas para a análise de dados. Eles preferem trabalhar com ferramentas simples que permitem novos métodos analíticos com a adição de novos pacotes (Câmara et al., 2016). Dessa forma, foram escolhidas ferramentas livres e que favorecessem a realização da análise de forma colaborativa e que pudessem ser facilmente reproduzidas por outros pesquisadores.

O ambiente de desenvolvimento R e a linguagem de alto nível Python são escolhas frequentes de analistas de dados, por contarem com uma variedade de ferramentas estatísticas, gráficas e de análise numérica. Para o ambiente utilizado no artigo, foi escolhida a linguagem Python. A escolha é motivada pela sua crescente popularidade na comunidade acadêmica (Wagner et al., 2017) e a disponibilidade de pacotes para computação científica (NumPy e SciPy), manipulação e análise de dados (Pandas), manipulação e análise de dados geoespaciais (GeoPandas), visualização de dados (Matplotlib e Seaborn) e produção de mapas de calor sobre o Google Maps (Gmplot).

Para o armazenamento e gerenciamento dos dados, foi utilizado o Sistema Gerenciador de Banco de Dados (SGBD) PostgreSQL, pelo fato de ser código livre e por suportar manipulação de informações geográficas mesmo em sua versão nativa. Para este trabalho, foram utilizados os módulos *psycopg2* para a conexão e leitura do banco de dados PostgreSQL e o GeoPandas para a conversão dos dados das tabelas do PostgreSQL em data frames georreferenciados.

Para a criação de gráficos, foram utilizados os módulos Seaborn e Matplotlib. Para a geração dos mapas de calor sobre o mapa do território brasileiro foi utilizado o módulo Gmplot.

Como ambiente de desenvolvimento para integrar essas tecnologias, foi selecionado o Jupyter Notebook. Essa plataforma é um ambiente de análise colaborativa, que permite a elaboração de programas que combinam programas, textos formatados e figuras.

2.2 Dados

Para as análises realizadas neste trabalho, foram utilizados conjuntos de dados disponibilizados pelo CENIPA no Portal Brasileiro de Dados Abertos¹. Esses dados contêm informações sobre as 6.238 ocorrências aeronáuticas que aconteceram no Brasil no período de 2006 a 2017. Essa base de dados possui quatro classes de dados: ocorrências, aeronaves, fatores e recomendações. As classes podem ser agrupadas a partir de um atributo em comum, e cada classe de dados é disponibilizada através de um arquivo CSV (*Comma-Separated Values*). Foram utilizados também os Anuários Estatísticos de Tráfego Aéreo confeccionados pelo CGNA (Centro de Gerenciamento de Navegação Aérea) e acessíveis no site do Instituto Brasileiro de Aviação².

Nesse estudo, foram analisados dados de ocorrências (informações sobre data, tipo de ocorrência, localidade, etc.) e sobre as aeronaves envolvidas nos eventos (informações de tipo de aeronave, fabricante, modelo, total de fatalidades, ano de fabricação, etc.). As Figs. 1 e 2 apresentam duas tabelas que listam os atributos contidos nas bases de dados de ocorrências e aeronaves, respectivamente, e disponíveis para análise. Ambas tabelas são formadas por três colunas: a primeira, que relaciona o nome dos atributos na base de dados; a segunda, que descreve o significado de cada um dos atributos no contexto do CENIPA; e a terceira, que especifica o tipo de dado contido.

| Ocorrências | | |
|--------------------------------|--|----------|
| Atributo | Descrição | Tipo |
| codigo_ocorrendia | Código da ocorrência | numérico |
| ocorrendia_classificacao | Classificação da ocorrência | texto |
| ocorrendia_tipo | Tipo da ocorrência | texto |
| ocorrendia_latitude | Latitude onde aconteceu a ocorrência | numérico |
| ocorrendia_longitude | Longitude onde aconteceu a ocorrência | numérico |
| ocorrendia_cidade | Cidade onde aconteceu a ocorrência | texto |
| ocorrendia_uf | Estado onde aconteceu a ocorrência | texto |
| ocorrendia_pais | País onde aconteceu a ocorrência | texto |
| ocorrendia_aerodromo | Código do aeródromo envolvido na ocorrência | texto |
| ocorrendia_dia | Data da ocorrência | data |
| ocorrendia_horario | Horário da ocorrência | hora |
| investigacao_aeronave_liberada | A aeronave foi liberada? | texto |
| investigacao_status | Status atual da investigação | texto |
| divulgacao_relatorio_numero | Número do relatório do CENIPA | texto |
| divulgacao_relatorio_publicado | O relatório do CENIPA foi publicado? | texto |
| divulgacao_dia_publicacao | Data da publicação do relatório do CENIPA | data |
| total_recomendacoes | Número total de recomendações | numérico |
| total_aeronaves_envolvidas | Número de aeronaves envolvidas na ocorrência | numérico |
| ocorrendia_saida_pista | Houve saída de pista na ocorrência? | texto |
| ocorrendia_dia_extracao | Data de extração desta base de dados | data |

Figura 1 – Atributos das ocorrências utilizados para análise.

¹ <http://dados.gov.br/>

² <http://www.institutoaviacao.org/>

| Aeronaves | | |
|-----------------------------|--|----------|
| Atributo | Descrição | Tipo |
| codigo_ocorrenca | Código da ocorrência | numérico |
| aeronave_matricula | Código de matrícula da aeronave | texto |
| aeronave_operador_categoria | Categoria de operação da aeronave | texto |
| aeronave_tipo_veiculo | Tipo de aeronave | texto |
| aeronave_fabricante | Fabricante da aeronave | texto |
| aeronave_modelo | Modelo da aeronave | texto |
| aeronave_tipo_icao | Código ICAO da aeronave | texto |
| aeronave_motor_tipo | Tipo de motor da aeronave | texto |
| aeronave_motor_quantidade | Quantidade de motores da aeronave | texto |
| aeronave_pmd | Peso máximo de decolagem da aeronave | numérico |
| aeronave_pmd_categoria | Categoria de peso da aeronave | texto |
| aeronave_assentos | Número de assentos da aeronave | numérico |
| aeronave_ano_fabricacao | Ano de fabricação da aeronave | numérico |
| aeronave_pais_fabricante | País de fabricação da aeronave | texto |
| aeronave_pais_registro | País que registrou a aeronave | texto |
| aeronave_registro_categoria | Categoria do registro da aeronave | texto |
| aeronave_registro_segmento | Segmento de registro da aeronave | texto |
| aeronave_voo_origem | Aeródromo de origem do voo | texto |
| aeronave_voo_destino | Aeródromo de destino do voo | texto |
| aeronave_fase_operacao | Fase de operação em que houve a ocorrência | texto |
| aeronave_tipo_operacao | Tipo de operação efetuado pela aeronave | texto |
| aeronave_nivel_dano | Nível de dano encontrado na aeronave | texto |
| total_fatalidades | Total de fatalidades resultantes da ocorrência | numérico |
| aeronave_dia_extracao | Data de extração desta base de dados | data |

Figura 2 – Atributos das aeronaves utilizados para análise.

A organização dos dados se iniciou a partir da carga dos arquivos no SGBD PostgreSQL. Com os dados carregados, foi realizada a criação de chaves e índices nos atributos correspondentes. A partir desse ponto, os dados já estavam disponíveis para serem acessados através do módulo *psycopg2* pelas aplicações de análise.

Os dados carregados através do *psycopg2* foram prontamente convertidos para estruturas de *GeoDataFrame* do *GeoPandas*, a fim de disponibilizar as funcionalidades dessa biblioteca para as análises realizadas nesse estudo.

Inicialmente, foram conduzidos processos de limpeza e extração de novas informações. A limpeza foi utilizada para remover atributos que não foram utilizados na análise, enquanto que novos atributos, como período do dia e dia da semana, foram produzidos a partir dos demais atributos existentes. Além desses, um novo atributo geográfico foi produzido a partir das colunas de latitude e longitude das ocorrências.

Os conjuntos de ocorrências e aeronaves foram unificados pelo atributo *codigo_ocorrenca*, utilizando a funcionalidade de *merge* disponível na estrutura do *GeoPandas*.

3 ANÁLISE EXPLORATÓRIA DE DADOS

Para realizar a EDA, foram estipulados domínios específicos para uma melhor visualização de inferências a respeito dos dados explorados. O processo de análise maximiza a obtenção de informações ocultas na estrutura de dados por meio de técnicas gráficas e quantitativas que detectam, por vezes, comportamentos não esperados e validam hipóteses assumidas, ou ainda justificam a escolha de modelos para determinado estudo de caso. Cada domínio específico de análise será apresentado nas subseções seguintes.

3.1 Classificação das ocorrências

Os dados de ocorrências disponibilizados pelo CENIPA são divididos em três classificações: Acidente, Incidente, Incidente Grave. De acordo com a NSCA 3-6 (Norma do Sistema do Comando da Aeronáutica), para classificar uma ocorrência como acidente ela precisa estar relacionada com a operação de uma aeronave que, no caso de aeronave tripulada, se dê entre o momento em que qualquer pessoa nela embarca com a intenção de realizar um voo, até o momento em que todas as pessoas tenham dela desembarcado, ou, no caso de uma aeronave não tripulada, se dê entre o momento em que a aeronave está pronta para se

movimentar com o propósito de voar, até o momento em que cessa a sua movimentação ao final do voo e seu sistema principal de propulsão é desligado na qual, pelo menos, uma das situações abaixo ocorra:

- Qualquer pessoa sofra lesão grave ou morra como resultado de estar na aeronave, ou em contato direto com qualquer uma de suas partes, incluindo aquelas que dela tenham se desprendido;
- A aeronave sofra dano ou falha estrutural que afete adversamente a resistência estrutural, o seu desempenho ou as suas características de voo e que normalmente exija a realização de grande reparo ou a substituição do componente afetado; e
- A aeronave seja considerada desaparecida, ou o local onde se encontre seja absolutamente inacessível.

Incidente, por sua vez, é toda ocorrência aeronáutica, relacionada com a operação de uma aeronave que não chegue a se caracterizar como um acidente aeronáutico grave, mas que afete ou possa afetar a segurança da operação.

Por fim, Incidente Grave trata-se de incidente aeronáutico envolvendo circunstâncias que indiquem que houve elevado potencial de risco de acidente relacionado à operação da aeronave e que:

- No caso de aeronave tripulada, tenha ocorrido entre o momento 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
- No caso de uma aeronave não tripulada, tenha ocorrido entre o momento que a aeronave está pronta para se movimentar, com a intenção de voo, até a sua inércia total pelo término do voo e seu sistema de propulsão tenha sido desligado.

É importante ressaltar, que a diferença entre o incidente grave e o acidente está apenas nas consequências. Há exemplos para melhor entendimento de incidentes graves e de acidentes na MCA 3-6 e no Manual de Notificação de Acidentes/Incidentes.

Na Fig. 3 é possível observar a distribuição das ocorrências de acordo com sua classificação. Consta-se que mais de dois terços das ocorrências são classificadas como incidentes ou incidentes graves. Somente 28,3% das ocorrências foram acidentes (equivalente a 1.764 ocorrências), mas conforme sua própria definição, o acidente representa justamente uma ocorrência mais significativa em termos de perdas materiais e humanas. Por esta razão, as análises apresentadas no restante deste trabalho consideram apenas os acidentes.

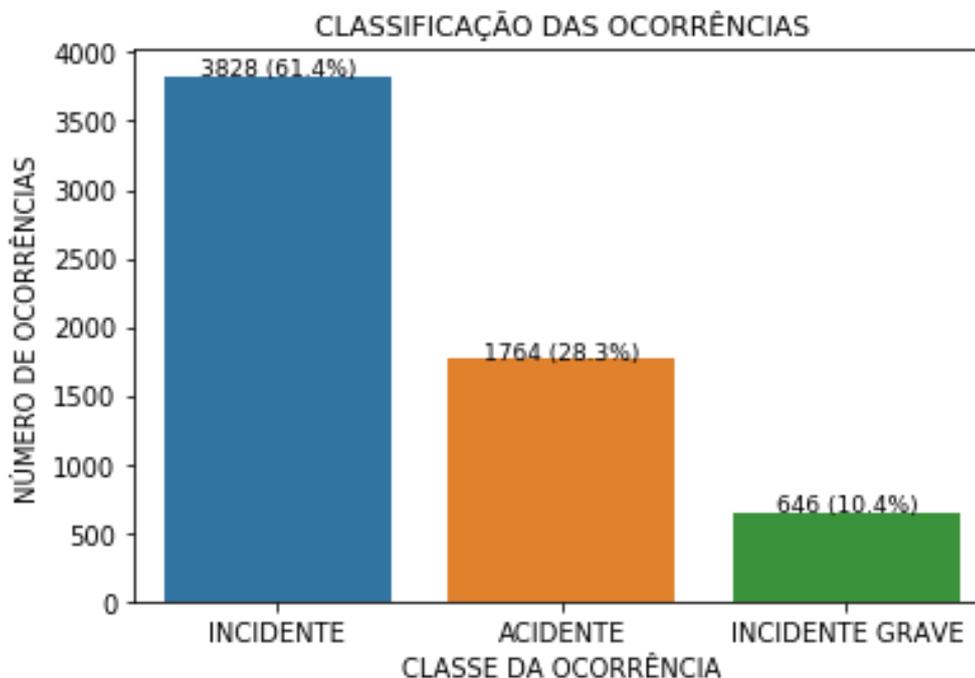


Figura 3 – Distribuição das ocorrências nas classificações de Incidente, Incidente Grave e Acidente.

3.2 Classificação dos Acidentes

De acordo com a estrutura dos dados adquiridos, um acidente pode ser classificado em diversos tipos. A Fig. 4 apresenta a distribuição dos acidentes, por tipo de ocorrência, que ocorrerem entre 2006 e 2017 e que se encontram na base de dados do CENIPA. Observa-se que Falha do Motor em Voo (20,2%), Perda de Controle em Voo (20,1%) e Perda de controle em Solo (13,5%) são os três principais tipos de acidentes registrados no período analisado.

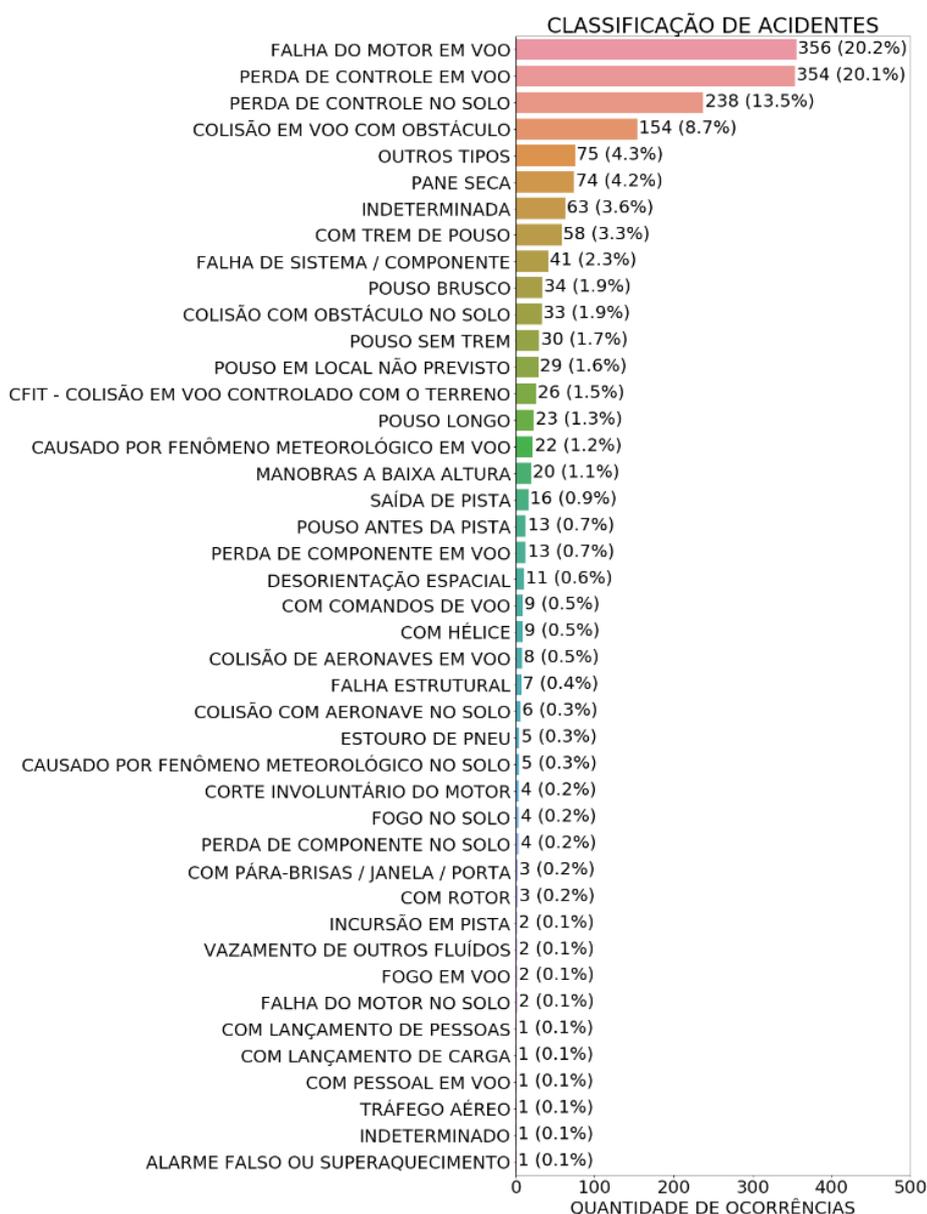


Figura 4 – Gráfico dos acidentes divididas por tipo e frequência.

As causas para tais ocorrências não necessariamente são em decorrência de uma falha humana na operação, como uma incompetência técnica por parte do piloto, por exemplo, pois a perda de controle em solo ou em voo pode ser devido a uma falha estrutural, mecânica, ou elétrica na aeronave. É importante mencionar que, apesar de haver apenas uma ocorrência de causa indeterminada, as análises relacionadas a classificação não são capazes de determinar com precisão o fator primário que desencadeou o acidente. Uma falha de motor em voo, por exemplo, pode ser classificada como a causa de um acidente, mas é uma condição que pode ser desencadeada por uma série de fatores como defeitos mecânicos, uso inadequado do equipamento, falta de manutenção, condições climáticas, entre outros. Estes fatores que determinam a origem dos acidentes não são tratados neste trabalho.

3.3 Progressão temporal dos acidentes

A Fig. 5 ilustra a progressão das ocorrências aéreas que cresceram, até 2012, de forma surpreendente. A partir de então os números estão retornando a patamares anteriores, sinalizando assim uma estabilização do número de acidentes no Brasil. Enquanto em 2006 houve menos que 80 acidentes, em 2012 esse número subiu para mais de 200. Já em 2017 esse valor está próximo de 140. É interessante observar que em 2006 e 2007 houveram os dois últimos grandes acidentes que mais marcaram a história do país, contudo, foram os anos com o menor número de acidentes no total. Por conseguinte, a maior quantidade de acidentes não necessariamente indica a maior quantidade de fatalidades.

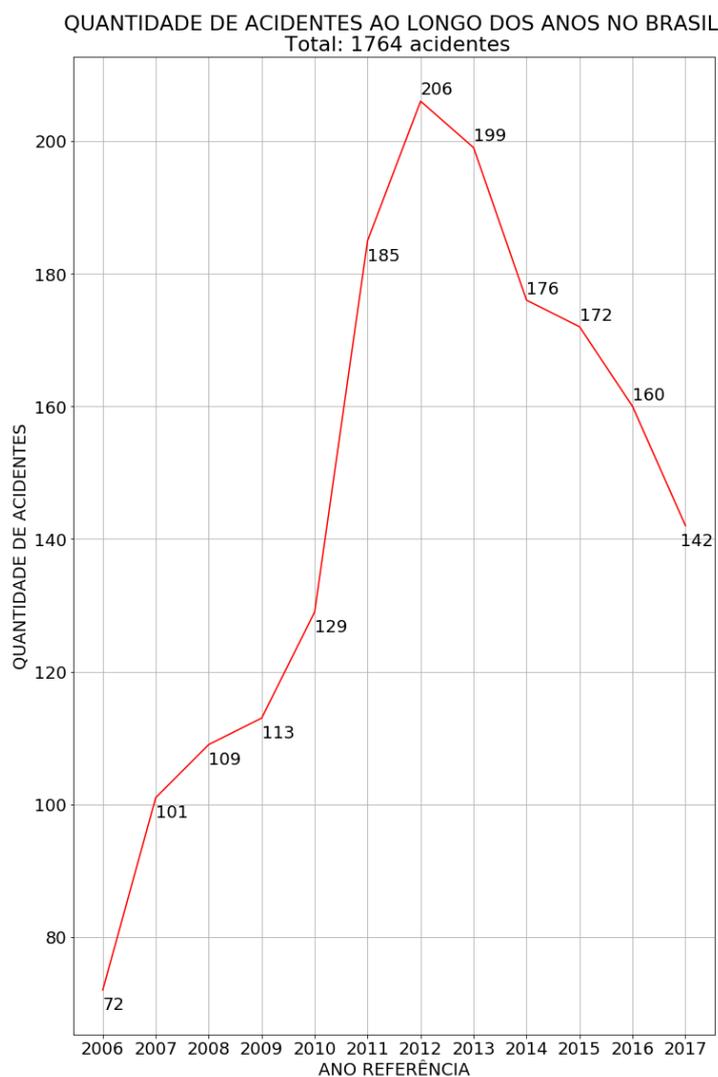


Figura 5 – Gráfico de progressão temporal das ocorrências.

3.4 Acidentes por Tipo de Aeronave

A Fig. 6 mostra que o avião é um tipo de aeronave mais frequente em ocorrências aéreas no Brasil. E, de fato, a quantidade total de ocorrências de aviões é maior do que a soma de todos os outros tipos de aeronaves.

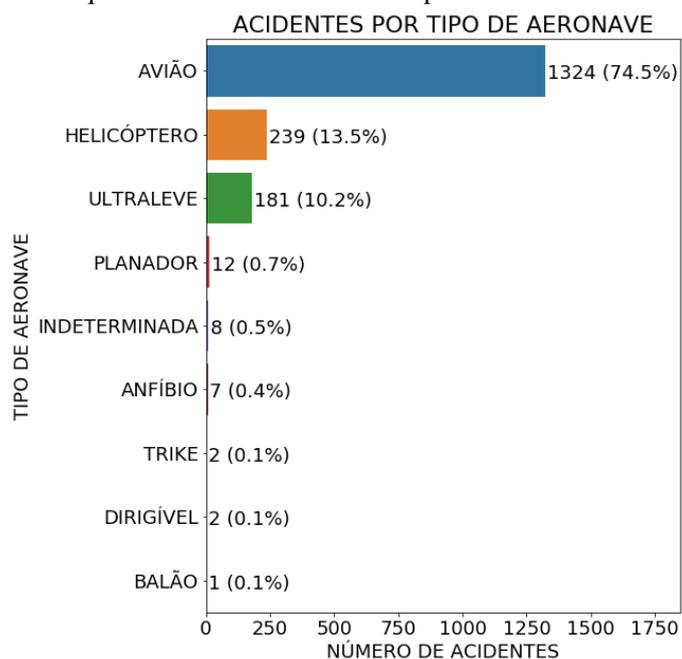


Figura 6 – Gráfico de acidentes por tipos de aeronaves.

3.5 Acidentes por Tipo de Motor

Os motores aeronáuticos são divididos em motores a pistão, a jato, turboélice e os turboeixo. É possível observar, no gráfico mostrado na Fig.7, que os motores a pistão são encontrados em maior frequência nas ocorrências reportadas. Isso provavelmente ocorre, pois, o motor a pistão é o tipo mais antigo e comum em aviões monomotores. Eles possuem um custo reduzido de fabricação e manutenção, sendo semelhantes aos motores dos carros. Porém, com o tratamento, padrão de qualidade, e confiabilidade da aeronáutica, funcionam muito bem em voos a baixas altitudes.

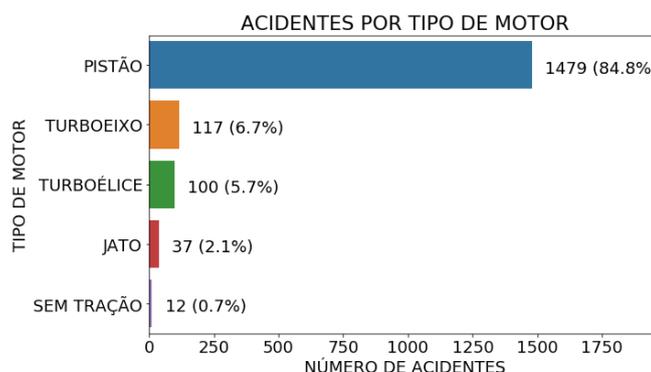


Figura 7 – Gráfico de ocorrências por tipos de motores.

Em seguida, os tipos de motores com mais ocorrências foram os turboélice e turboeixo. Assim como os motores a pistão, os turboélice também são caracterizados pela hélice que fica à frente do avião. Contudo, trata-se de um motor de reação mista, pois é basicamente um motor a jato acionando uma hélice através de um redutor de velocidade instalado entre o eixo da turbina e a hélice. Já os motores turboeixo utilizam a mesma mecânica dos turboélice, com a diferença de que eles atuam em um eixo vertical que sustenta asas rotativas. Então, na prática, turboélice e turboeixo diferem basicamente no fato de que o primeiro é usado em aviões, e o segundo em helicópteros.

3.6 Acidentes por Fase de Operação

O gráfico da Fig.8 mostra que as fases de aproximação, quando o avião começa a descer, até a corrida após o pouso, correspondem aos momentos mais críticos. Cerca de 546 ocorrências, das mais de 1.700 registradas, ocorreram nesta fase. Além disso, vale destacar também que o número de ocorrências durante a decolagem é consideravelmente próximo ao número de ocorrências durante o pouso, fazendo também da decolagem uma fase de operação delicada.

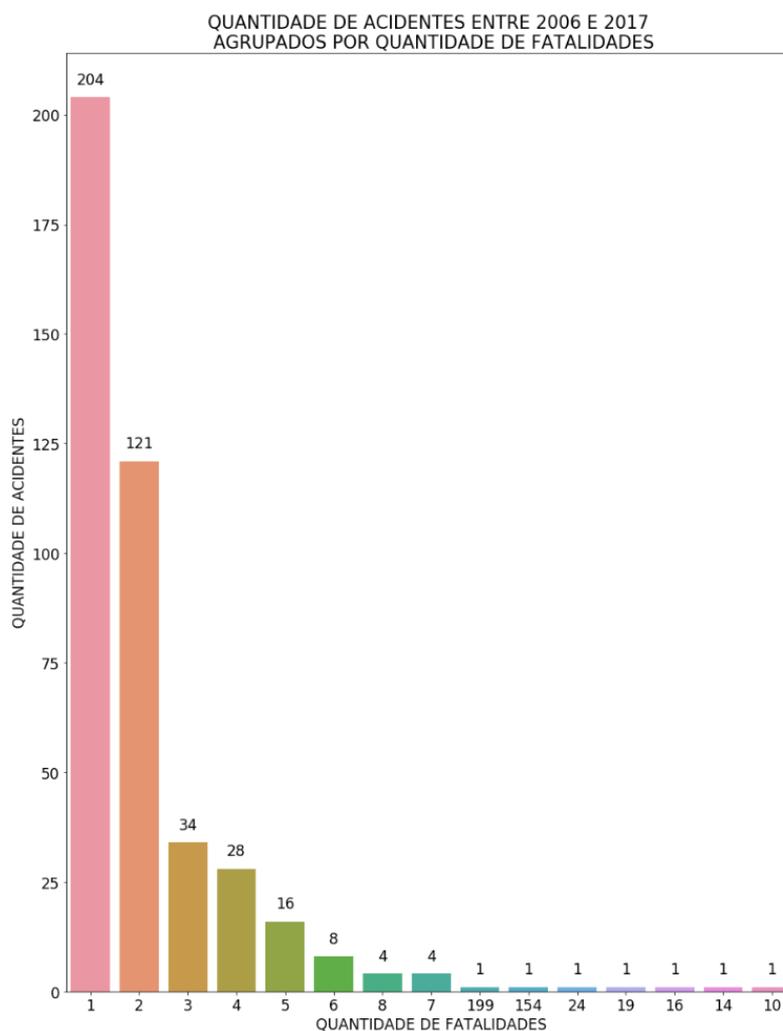
Isso ocorre porque durante essas fases, o avião está mais próximo do chão e mais vulnerável a falhas, sendo assim, qualquer falha humana ou mecânica pode dar início a uma sequência de erros que eventualmente causam um acidente fatal. Aliás, em geral espera-se que na hora do pouso as falhas mais comuns sejam humanas, já que a tripulação está sob maior pressão psicológica e estresse, e tem menos tempo de decisão para manobras de emergência. A decolagem também é considerada um momento crítico, pois na decolagem a aeronave opera em capacidade máxima de potência e sofre grande pressão estrutural. Isso ocorre, pois, o avião está mais pesado, cheio de combustível, ainda está ganhando velocidade, e subindo com uma capacidade de sustentação menor.



Figura 8 – Gráfico de ocorrências por Fase de Operação.

Como mencionado anteriormente, nem todo acidente resulta necessariamente na existência de fatalidades, pois uma aeronave seriamente avariada ou perdida também é caracterizada como acidente, mesmo que não haja sequer feridos. Uma análise mais detalhada especificamente das fatalidades ocorridas é mostrada a seguir.

3.7 Fatalidades

**Figura 9** – Gráfico de fatalidades por ocorrências.

É possível observar também que os dois grandes acidentes que marcaram o Brasil nos últimos anos se destacam no gráfico. Em 2006 um Boeing da Gol voo 1907 chocou-se contra um jato Legacy que se dirigia para os Estados Unidos. Como resultado do choque, a aeronave caiu em uma mata fechada na região de Peixoto de Azevedo, a 692 km ao Norte de Cuiabá. As 154 pessoas que estavam a bordo, entre tripulantes e passageiros, morreram na queda. O outro acidente, ocorreu em 2007 quando um Airbus A-320 que fazia o voo da TAM 3054 não conseguiu pousar na pista principal do aeródromo de Congonhas em São Paulo. Sob chuva, a aeronave ultrapassou os limites do aeródromo, atravessou a avenida Washington Luiz e acabou se chocando contra um prédio da TAM Express. Uma explosão e um grande incêndio seguiram a colisão que vitimou 199 pessoas dentre as que estavam no voo e as que morreram em solo.

3.8 Localidade

O gráfico de ocorrências por localidade, mostrado na Fig. 10, constata que o estado de São Paulo foi o que teve o maior número de ocorrências, superando inclusive a soma de todos os acidentes dos outros estados da região sudeste. São Paulo é o estado com mais aeródromos do Brasil, portanto, é de se esperar que como consequência tenha o maior número de voos, permitindo-nos cogitar que este é o motivo de tantas ocorrências. Outros estados, com menos aeródromos, como Goiás, Maranhão, Mato Grosso e Minas Gerais obtiveram um número de ocorrências consideravelmente menor que o estado de São Paulo. Observe também que dois acidentes não constam no gráfico, que computa 1.762 acidentes, pois foram ocorrências fora do território brasileiro, mas com aeronaves nacionais.

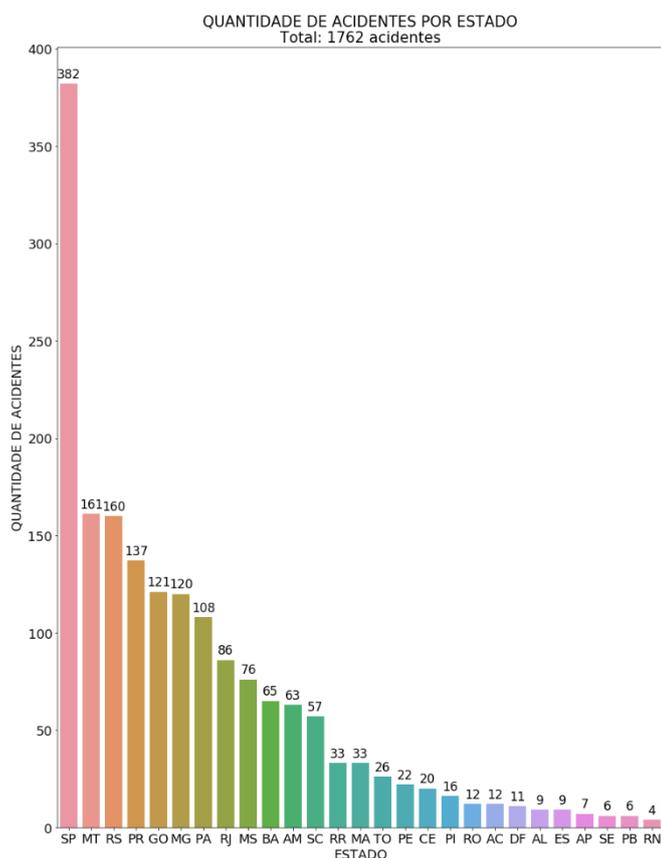


Figura 10 – Gráfico de acidentes por localidade.

Com o intuito de analisar a concentração das ocorrências em suas localidades, a Fig.11 mostra uma projeção dos dados com informações geográficas num mapa de calor sobre o território brasileiro visualizado no Google Maps. Mapas de calor são uma das melhores ferramentas de análise exploratória de dados para a visualização da densidade de ocorrências, pois torna fácil a identificação de ilhas onde existe uma elevada concentração de atividade (Choudhary et al., 2015).

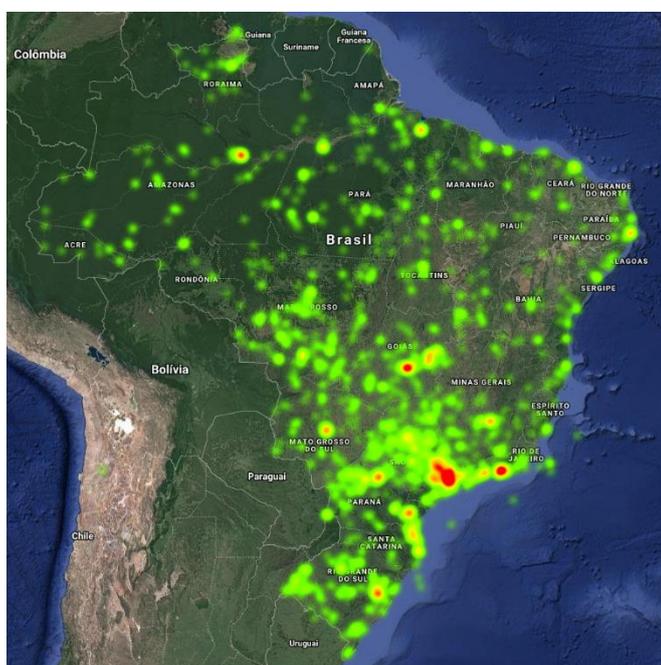


Figura 11 – Mapa de calor dos acidentes aeronáuticos ocorridos no Brasil.

Nota-se que a maioria dos acidentes ocorreu nas proximidades dos principais aeródromos do país, reforçando uma tendência já citada neste trabalho. É possível observar também que existem algumas ilhas de concentração onde a ocorrência de acidentes se destaca, como na região entre a grande São Paulo e Campinas, Rio de Janeiro e Goiânia. É elementar deduzir que em São Paulo e Rio de Janeiro a concentração de acidentes seja maior, pois os quatro principais aeródromos destas cidades estão entre os cinco mais movimentados do país, segundo a última atualização do ranking de aeródromos mais movimentados do Brasil feito

pela CGNA em 2016. No entanto, Goiânia causa certa surpresa em aparecer com tanto destaque, já que é apenas o décimo quarto colocado neste mesmo ranking.

De forma menos destacada, a figura mostra também outras ilhas em Porto Alegre, Curitiba, Londrina, Campo Grande, Belo Horizonte, Brasília, Manaus, Recife e Belém. E é interessante observar que Salvador, Cuiabá, Florianópolis e Vitória não aparenta ter muitas ocorrências, mesmo possuindo alguns aeródromos bem colocados no ranking. Por outro lado, Londrina concentra um número elevado de acidentes, tendo em vista que é apenas o vigésimo segundo lugar no ranking.

Outras regiões que se destacam em grau menor são: o litoral de Santa Catarina (região entre Florianópolis, Balneário Camboriú e Joinville), litoral do estado do Rio de Janeiro (região de Paraty, Angra dos Reis e Mangaratiba), Cuiabá, e interior de São Paulo (região entre São José do Rio Preto, Barretos e Ribeirão Preto).

3.9 Categoria da Aeronave

O gráfico da Fig.12 mostra a distribuição dos acidentes aeronáuticos no Brasil por categoria de aeronave. Essas definições são importantes para que se possa evidenciar as operações com maior ocorrência de acidentes com base no número absoluto de incidências.

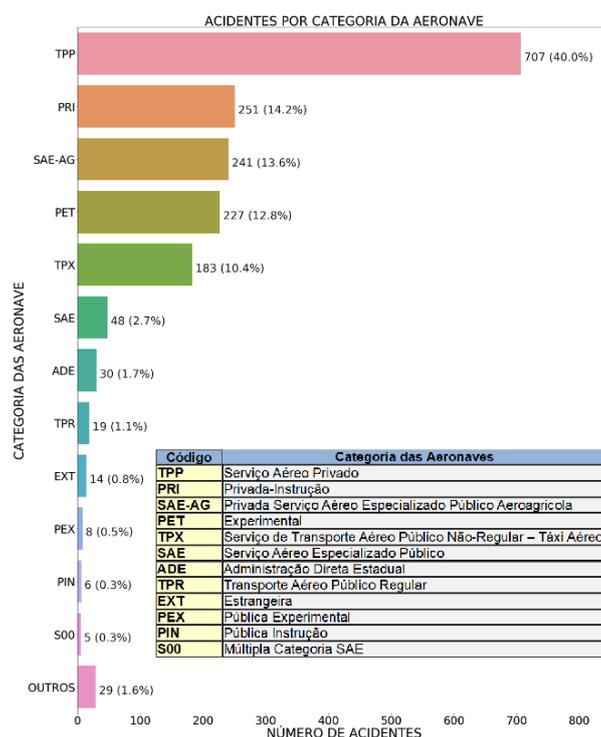


Figura 12 - Gráfico de acidentes por Categoria da Aeronave

É possível observar na Fig. 12 que a categoria de aeronave que sofreu o maior número de acidentes foi a de aeronaves de Serviços Aéreos Privados. Porém, é muito importante mencionar que, baseado na definição de voo privado que segue a Resolução nº 293 de 19 de novembro de 2013, da ANAC, a categoria das aeronaves privadas pode englobar várias subcategorias como aeronaves experimentais ou agrícolas, de forma que o valor observado nesta categoria pode representar um sentido dúbio em muitos casos, de acordo com a interpretação da pessoa que preencheu a informação no sistema. Ou seja, na prática, pode-se conjecturar que muitas aeronaves experimentais, agrícolas ou de instrução podem ter sido cadastradas como privadas na base de dados do CENIPA.

À exceção das aeronaves privadas, as categorias que se destacam com o maior número de acidentes são as aeronaves de instrução, agrícola, experimental e táxi-aéreo, respectivamente. As demais categorias apresentam números bem menos significativos.

3.10 Acidentes por horário do dia

A distribuição de acidentes por horário do dia está apresentada na Fig.13. Analisar as ocorrências ao longo do dia permite realizar inferências sobre os horários ou períodos específicos do dia com maior incidência de acidentes, o que pode contribuir para políticas de prevenção de acidentes. Analisando a Fig.13, pode-se observar uma maior concentração no número de ocorrências em períodos específicos do dia, como no início da tarde (12h-14h) e início da noite (18h-20h). A maior frequência de ocorrências nesses horários pode estar relacionada, por exemplo, com o maior número de voos nesses horários ou, ainda, com a iluminação solar. A fim de verificar a hipótese de que a maior ocorrência de acidentes no período das 18h às 20h está associada ao crepúsculo vespertino, foi produzido um mapa de calor, apresentado na Fig.14. Caso a hipótese for verdadeira, é esperado

que as ocorrências relacionadas ao pôr do Sol variem de horário durante o ano, uma vez que, no inverno, o Sol se põe mais cedo e, no verão, mais tarde. O mapa da Fig. 14 apresenta o número acumulado de acidentes aéreos para cada mês e cada hora do dia.

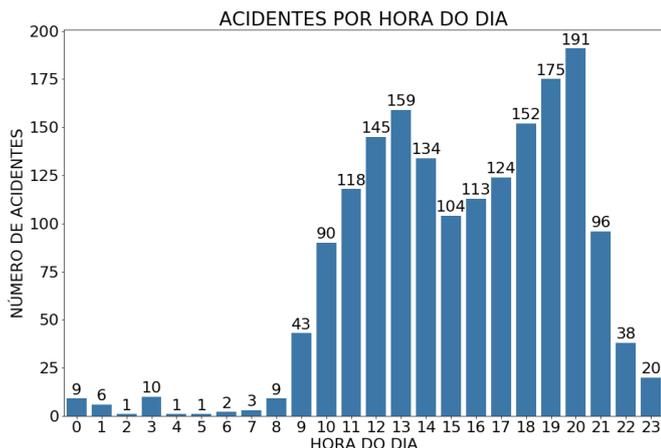


Figura 13 - Distribuição de acidentes por hora do dia

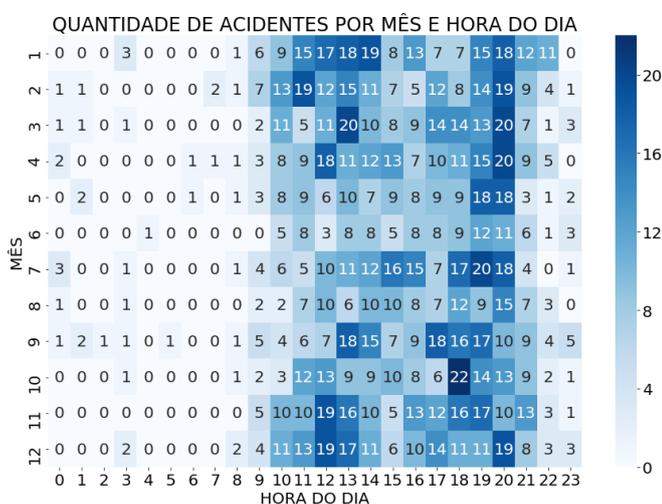


Figura 14 - Quantidade de acidentes por mês e hora do dia

Como pode-se observar, os picos de ocorrências no final da tarde não ocorrem com uma hora de antecedência nos meses de inverno, quando comparados aos meses de verão. Em diversos meses (janeiro, fevereiro, abril, março, setembro, novembro e dezembro) o pico de ocorrências no meio do dia (Sol próximo ao zênite) está muito próximo ao acumulado de ocorrências no final da tarde (pôr do Sol).

O panorama da distribuição de acidentes dividido por mês e hora do dia motivou os autores a discretizar períodos do dia de modo a tornar a análise mais perceptível. Para as análises a seguir, e para este trabalho, o dia foi dividido nos seguintes períodos: manhã (entre 6h e 12h), tarde (entre 12h e 18h), noite (entre 18h e 0h) e madrugada (entre 0h e 6h).

4 ESTUDO DE CASO

A seção anterior apresentou uma Análise Exploratória dos Dados de ocorrência aéreas no Brasil. Poderíamos observar separadamente cada domínio (variável) a fim de identificar as características que definem o perfil com maior frequência de acidente. Esse tipo de análise resultaria, por exemplo, na conclusão que a maior parte dos acidentes ocorrem com aviões de serviços aéreos privados (TPP), com falhas do motor (à pistão) em voo, durante a decolagem, no estado de SP, por volta das 20h aos sábados. Apesar de haver maior chance do perfil de ocorrência de acidentes possuir essas características, não é possível fazer essa afirmação sem, antes, fazer uma avaliação cruzada entre pares de variáveis, seguida por três, quatro, até o número total de variáveis menos um. Usando seis variáveis, por exemplo, seriam necessárias 56 avaliações cruzadas, que seriam as combinações dois a dois, três a três, quatro a quatro e cinco a cinco das variáveis para certificar o perfil de maior ocorrência.

A fim de apresentar o método da análise dos dados, esta seção apresenta um estudo de caso para identificação do perfil com maior probabilidade de envolvimento em acidente aéreo, considerando as variáveis dia da semana, período do dia e porte da aeronave. A redução do domínio de análise foi realizada a fim de simplificar a complexidade da análise, uma vez que o objetivo do trabalho está mais voltado a fornecer uma primeira compreensão dos dados e documentar tecnologias, técnicas e métodos em Ciência de Dados na análise de incidentes aéreos ocorridos no Brasil.

Em relação ao dia da semana, a distribuição demonstrou que o dia que ocorre mais acidentes é o **sábado**, com 300 ocorrências, e segunda-feira é o dia em que ocorre menos acidentes, apontando 207 ocorrências. A Fig.15 mostra a distribuição de acidentes por dia da semana, com 17,2% dos acidentes ocorridos no sábado.

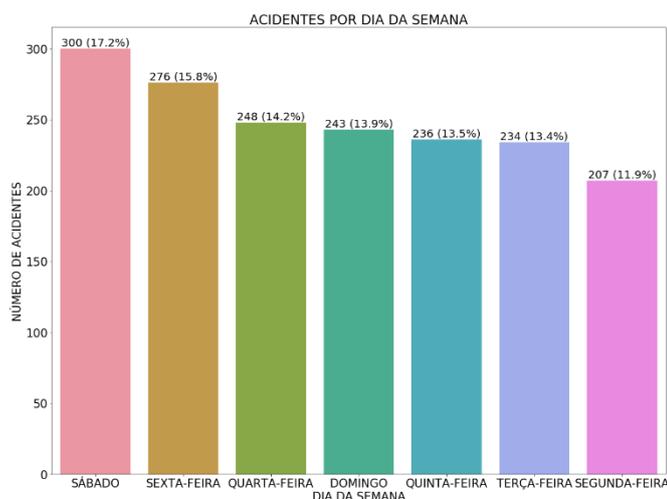


Figura 15 - Distribuição por dia da semana

Para isso, em relação ao período do dia, conforme convenção adotada neste artigo, observou-se uma maior incidência de ocorrências no período da **tarde**, ou seja, entre 12h e 18h. Foram um total de 779 ocorrências de acidentes neste período, o que corresponde a 44,6% das ocorrências. O gráfico da Fig.16 mostra a distribuição dos acidentes por período do dia.

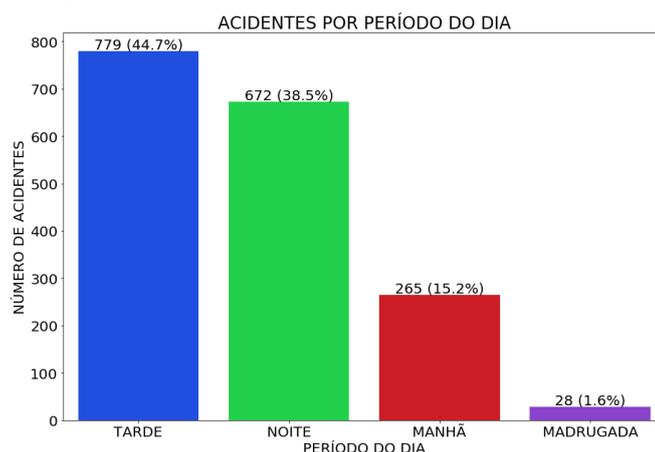


Figura 16 - Distribuição por período do dia

Já em relação ao porte da aeronave, o CENIPA (2006, p. 2) classifica as aeronaves em três categorias: leve (com até 2.250kg), média (entre 2.250kg e 5.700kg) e pesada (acima de 5.700kg). Classificando os acidentes de acordo com o porte da aeronave, chega-se a um predomínio dos **aparelhos de porte leve** com um total de 1.493 ocorrências, o que corresponde a 85,6% do total. O gráfico da Fig.17, mostra a distribuição de acidentes pelo porte da aeronave.

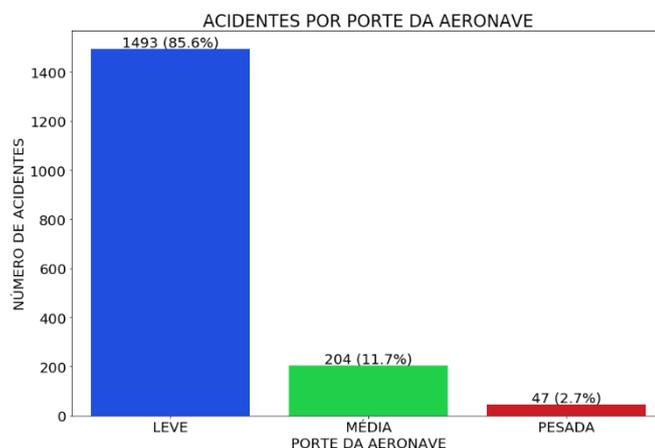


Figura 17 - Distribuição por dia da semana

Portanto, considerando as três distribuições anteriores, pode-se traçar um perfil com maior quantidade de acidentes, considerando os atributos mencionados: **aeronave de porte leve, cujo voo ocorre no sábado à tarde**.

4.1 Realizando as correlações

O próximo passo foi de confirmar o perfil de criticidade encontrada fazendo as correlações das características encontradas individualmente. Para isso, considerando apenas o sábado, tivemos um total de 138 acidentes no período da tarde, conforme a distribuição encontrada no gráfico da Fig. 18. Por isso, o percentual de ocorrências no período da tarde, para este caso é de 46%.

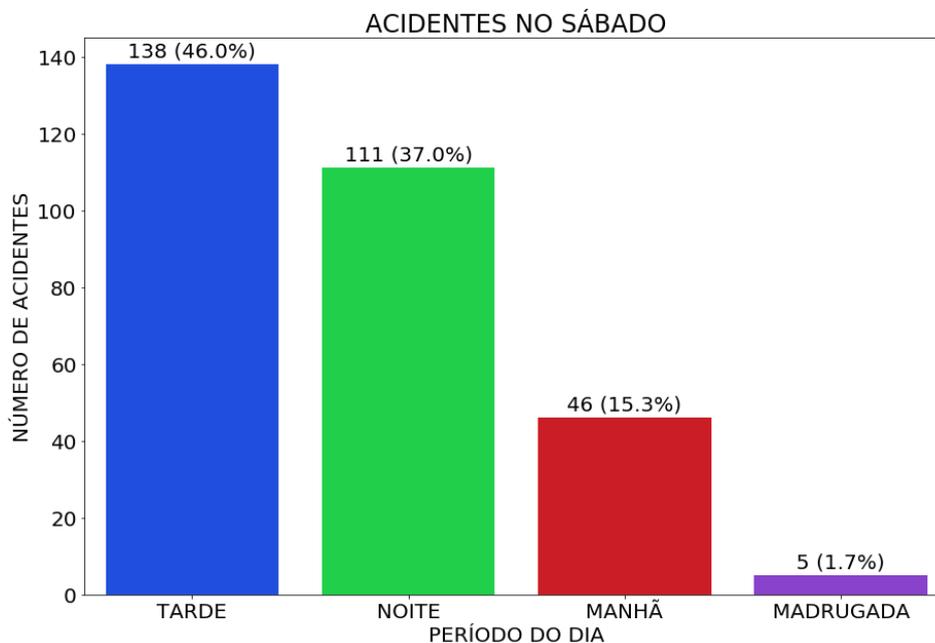


Figura 18 - Distribuição por período do dia no sábado

Ainda considerando os acidentes ocorridos no sábado, tivemos um total de 273 acidentes ocorridos com aviões de porte leve. A distribuição dos acidentes no sábado por porte da aeronave pode ser visualizada no gráfico da Fig.19. Destaca-se o significativo aumento do número de acidentes envolvendo as aeronaves do tipo leve aos sábados, representando 91% do total de ocorrências.

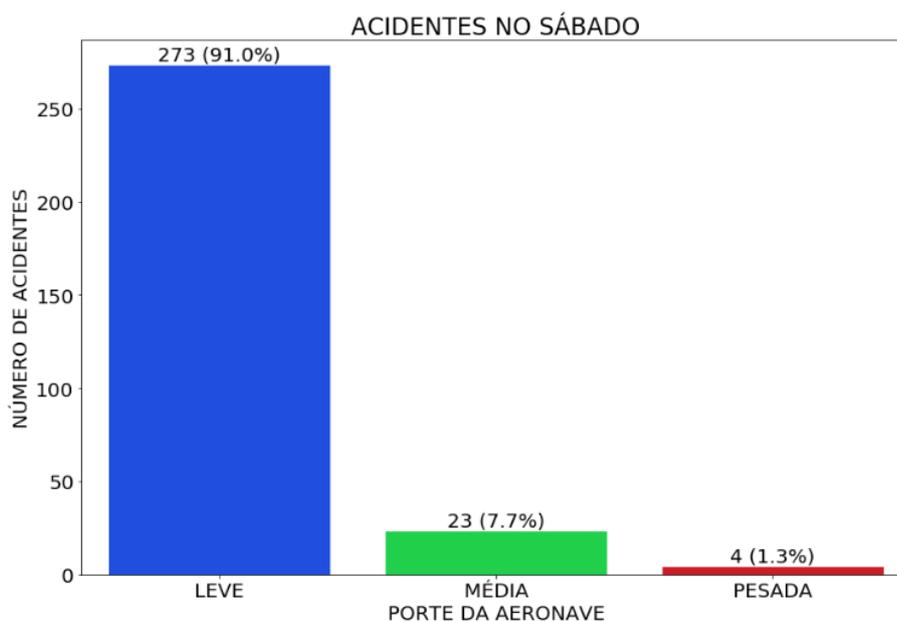


Figura 19 - Distribuição por porte de aeronave no sábado

Do mesmo modo, correlacionou-se os acidentes ocorridos no período da tarde com o porte da aeronave. Como resultado, tivemos um total 679 acidentes com aeronaves leves, ou 87,1% das ocorrências, cuja distribuição encontra-se no gráfico da Fig. 20.

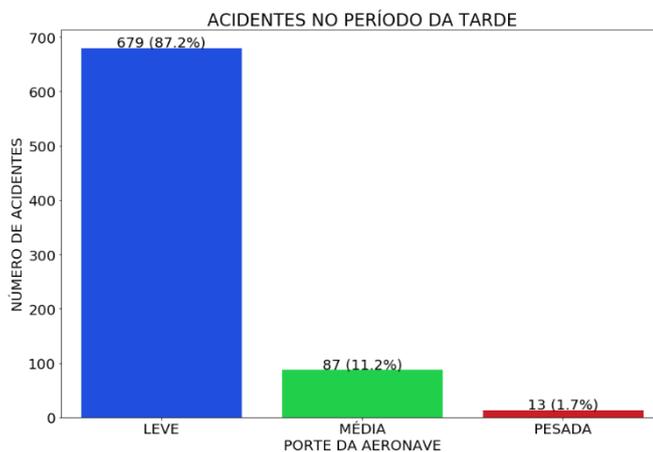


Figura 20 - Distribuição por porte de aeronave no período da tarde

Também no período da tarde, foi realizada a correlação considerando os dias da semana. Observando a distribuição das ocorrências, de acordo com a Fig. 21, é possível ratificar a informação de que o dia da semana que mais ocorre acidentes à tarde é sábado, com 17,7% dos eventos.

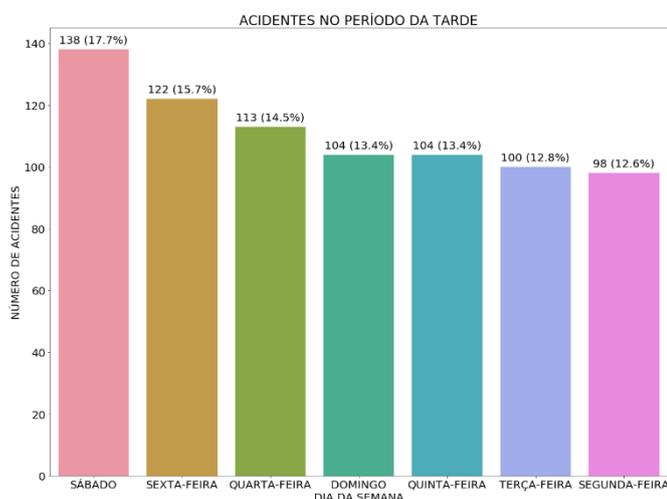


Figura 21 - Distribuição por dia da semana no período da tarde

Finalmente, efetuou-se as correlações de acidentes ocorridos com aeronaves de porte leve. A primeira correlação foi com os dias da semana. A distribuição dos dados, conforme mostra a Fig. 22, aponta, mais uma vez, o sábado como o dia da semana com mais ocorrências, com 273 registros, ou 18,3% dos acidentes.

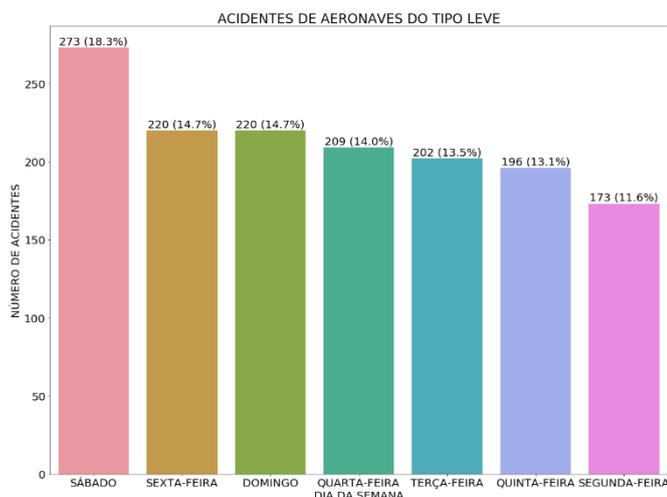


Figura 22 - Distribuição por dia da semana de acidentes de aeronave de porte leve

Em seguida, foi efetuada a correlação entre as aeronaves do tipo leve e o período de ocorrência dos acidentes deste tipo de aeronave. Mais uma vez, foi possível constatar que o período que mais ocorre acidentes é o da tarde, ou seja, entre 12h e 18h. Ocorreram um total de 679 ocorrências neste período, seguidos pelos períodos da noite, manhã e madrugada. A distribuição da

correlação é mostrada na Fig. 23. Ratificando as análises anteriores, o período do dia com mais eventos corresponde por 45,5% dos registros.

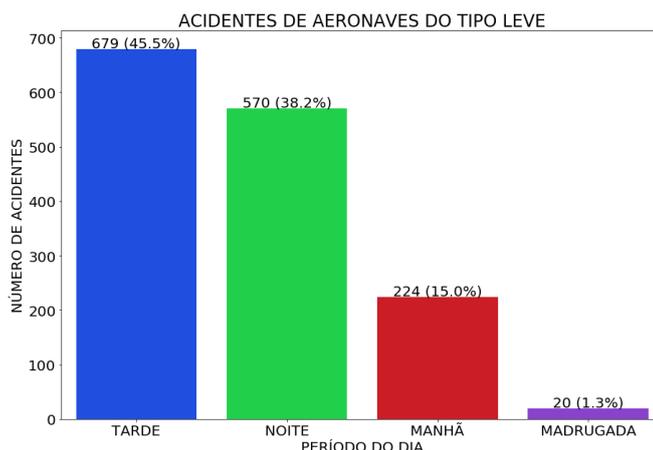


Figura 23 - Distribuição por período do dia de acidentes de aeronave de porte leve

Ainda em busca do perfil de voo mais suscetível a um acidente, foram efetuadas algumas outras correlações. Uma delas foi a correlação entre o tipo da aeronave que mais se acidenta no sábado, considerando apenas o período da tarde. O resultado foi um montante de 126 eventos com aviões do tipo leve, correspondente a 91,3% das ocorrências, conforme mostrado no gráfico da Fig. 24.

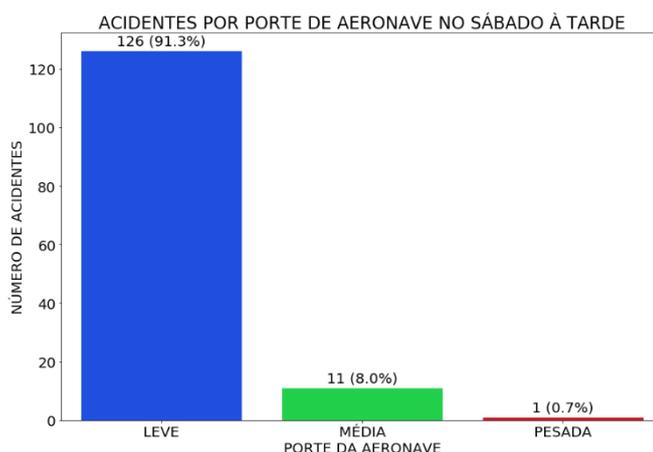


Figura 24 - Distribuição por porte da aeronave de acidentes ocorridos no sábado à tarde

A próxima correlação diz respeito ao período do dia dos acidentes envolvendo aviões de porte leve que se acidentam no sábado. O resultado da correlação foi um total de 126 ocorrências no período da tarde, ou 46,2% das ocorrências, conforme aponta a distribuição do gráfico da Fig. 25.

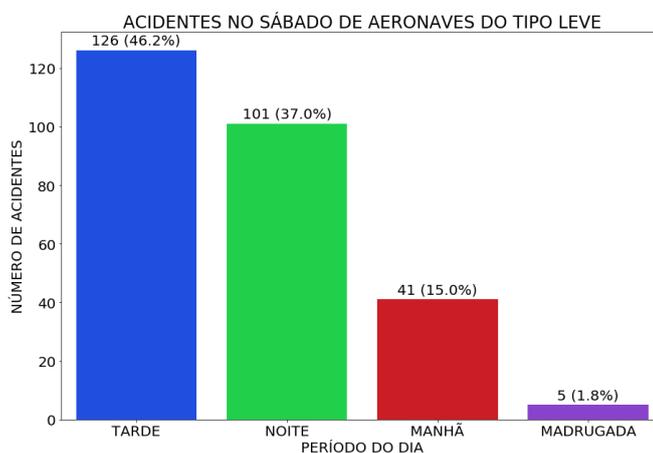


Figura 25 - Distribuição por período do dia de acidentes ocorridos no sábado à tarde

A última análise foi sobre o dia da semana com mais ocorrências considerando as aeronaves de porte leve acidentadas no período da tarde. O sábado aparece, novamente, como o dia da semana com mais ocorrências, com 126 eventos. Esse montante faz referência a 18,6% dos acidentes, conforme mostra o gráfico da Fig. 26.

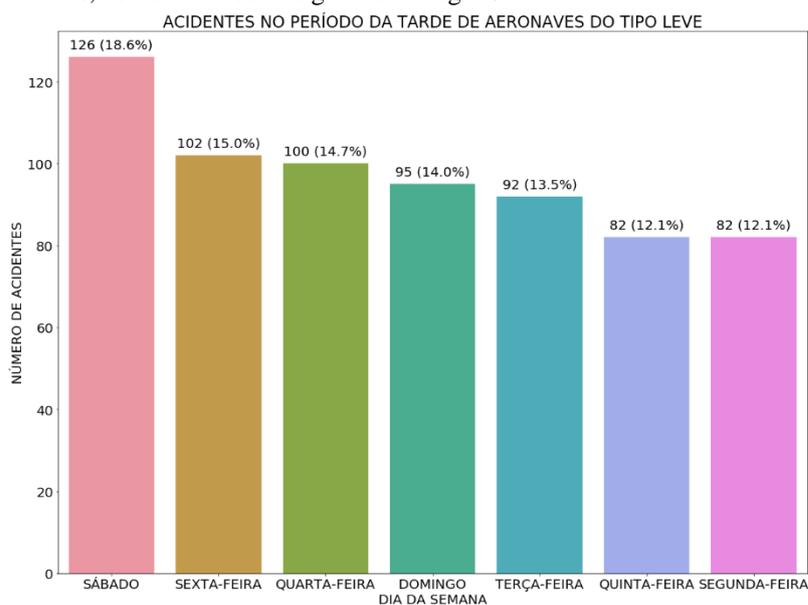


Figura 26 - Distribuição por dia da semana dos acidentes de aeronaves leves ocorridos no período da tarde

Portanto, por meio das diversas análises e correlações realizadas, foi possível obter e ratificar o perfil mais crítico de voo considerando as variáveis: dia da semana, período do dia e porte da aeronave. O resultado evidenciou que as **aeronaves de porte leve que voam no sábado no período da tarde** são mais suscetíveis a um acidente aeronáutico.

Não há dados que possam ser analisados para demonstrar essa tendência, mas pode-se conjecturar que o fato dos acidentes serem mais comuns em aeronaves de porte leve que voam no sábado à tarde, seja reflexo do tipo de uso que faz o público alvo dessas aeronaves, que em geral as utilizam como um hobby, para voos panorâmicos, instrução ou viagens. E tais eventos são mais comuns aos finais de semana.

De forma a complementar a análise, buscou-se identificar com informações geográficas os locais onde mais ocorrem acidentes com o perfil crítico supracitado. Como resultado de tal análise, obteve-se o mapa de calor mostrado na Fig. 27, onde é possível identificar alguns núcleos no território nacional com maior índice de acidentes relacionados ao perfil crítico destacado anteriormente.

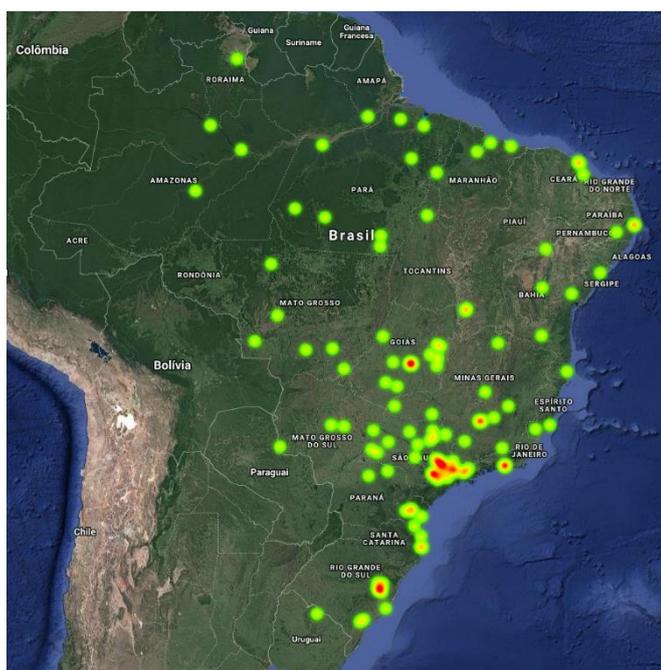


Figura 27 - Mapa de calor dos acidentes de aeronaves leves ocorridos no sábado à tarde

O estado de São Paulo, tanto a capital quanto o interior, destacam-se como uma das regiões com maior número de acidentes, conforme é mostrado de forma mais aproximada na Fig. 28.

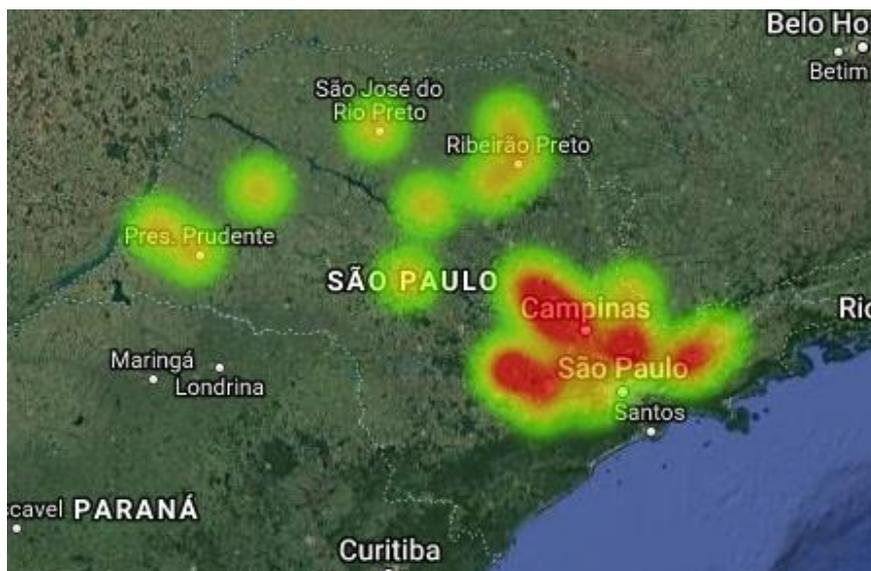


Figura 28 - Mapa de calor na região de São Paulo dos acidentes de aeronaves leves ocorridos no sábado à tarde

Os próximos passos deste trabalho serão no sentido de encontrar uma possível tendência ou padrões de comportamento dos acidentes aeronáuticos, mas não só em relação aos dias da semana e períodos do dia, como também em consideração a quantidade de movimentos de aeronaves em alguns aeródromos nestes períodos.

Como essas informações não constam na base de dados do CENIPA, para realizar estas análises foi necessário utilizar a base de dados publicados nos Anuários do Centro de Gerenciamento da Navegação Aérea (CGNA). Os dados destes anuários foram coletados tendo como base informações de Aeródromos, a partir de suas respectivas Torres de Controle (TWR) e de TMA (Área de Controle Terminal) e CTR (Zona de Controle), a partir de seus controles de aproximação, órgãos pertencentes ao SISCEAB (Sistema de Controle do Espaço Aéreo).

É preciso destacar que os anuários apresentam os dados de movimentação total pelo dia da semana e por período do dia, porém, não existem dados anteriores a 2012 e posteriores a 2016. Por isso as análises ficarão restritas a este período a partir deste ponto do trabalho.

É importante ressaltar que o movimento de aeronaves em um aeródromo abrange pousos, decolagens, cruzamentos e manobras *Touch and Go Landing* (TGL). Foram submetidas para a análise as TMA I e TMA II do estado de São Paulo (mostrados na Fig.29), por serem as regiões com mais ocorrências no período pesquisado.

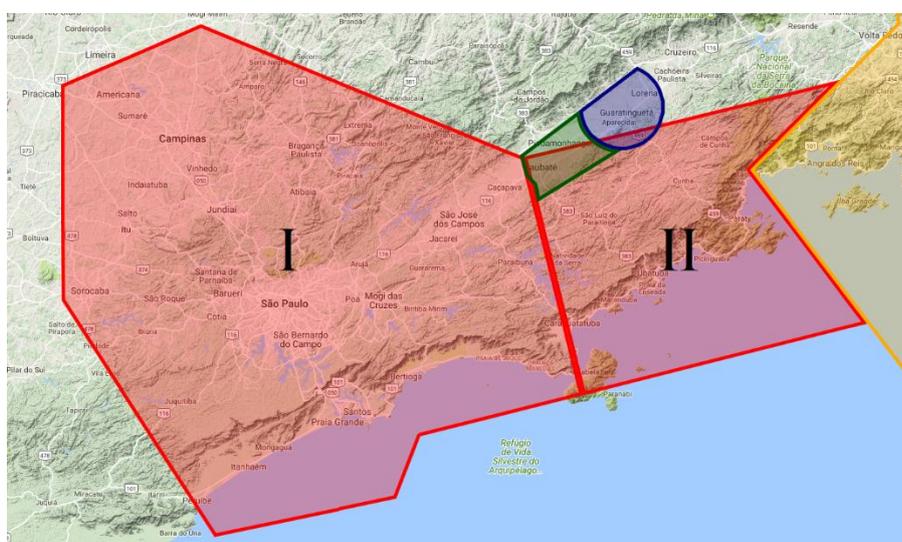


Figura 29 - TMA I e II (Áreas de Controle Terminal de São Paulo)

Entre 2012 e 2016, ocorreram quase quatro milhões de movimentos nos TMA I e TMA II do estado de São Paulo. Ao analisar as Figs. 30 e 31, é possível constatar que, apesar de pouco mais de 10% dos voos terem ocorrido aos sábados, este período

é responsável por quase 25% dos acidentes. No entanto, vemos que a terça-feira é o dia da semana com a menor incidência de acidentes (6.3%), mesmo sendo um dos dias com a maior quantidade de voos durante a semana (15.6%).

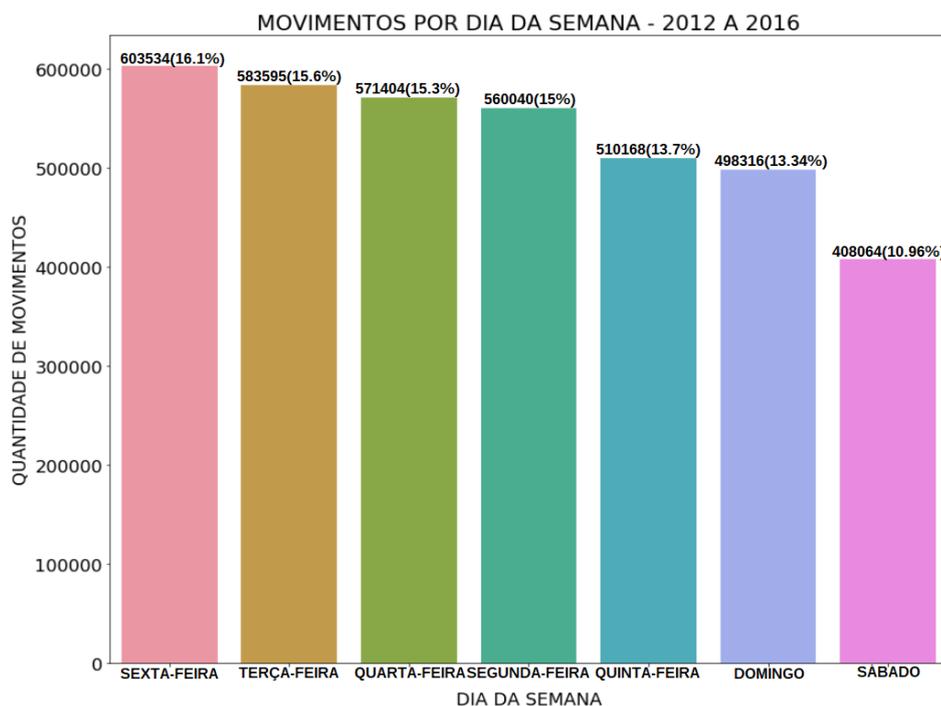


Figura 30 - Quantidade de voos por dia da semana

Assim como ocorreu em análises anteriores, mesmo levando-se em consideração o número proporcional de acidentes em relação ao número de movimentos de aeronaves nas TMA I e TMA II, é possível inferir que ocorreram mais acidentes nestas áreas aos sábados do que em outros dias da semana. Essa informação, isoladamente, tem pouca relevância no tocante à prevenção de acidentes aeronáuticos, contudo, quando associada com outras inferências no sentido de identificar perfis críticos, pode servir como um indicativo que auxilie na prevenção de acidentes.

Foi realizado uma segunda análise usando como base os dados do CGNA. Motivado pela ausência de dados anteriores, foi considerado os anos de 2014 a 2016 para a correlação entre acidentes e períodos do dia. Por convenção usamos os períodos madrugada (00h - 06h), manhã (06h - 12h), tarde (12h - 18h) e noite (18h - 24h).

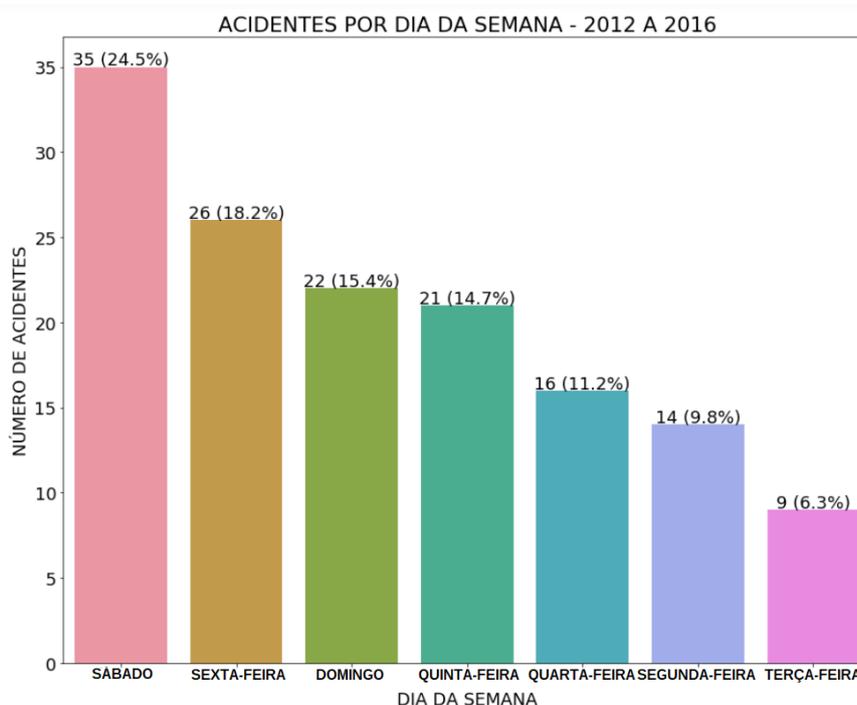


Figura 31 - Quantidade de acidentes por dia da semana entre 2012 e 2016

Restringindo novamente às TMA I e TMA II, foi possível constatar pela Fig. 32 que 26,91% dos voos ocorreram no período da noite, mas ao mesmo tempo, pela Fig. 33 mostra-se que 40% dos acidentes ocorrerem no período da noite. Isso revela uma

tendência de que, apesar de ter menos acidente do que o período da tarde em números absolutos, proporcionalmente o período da noite é mais crítico.

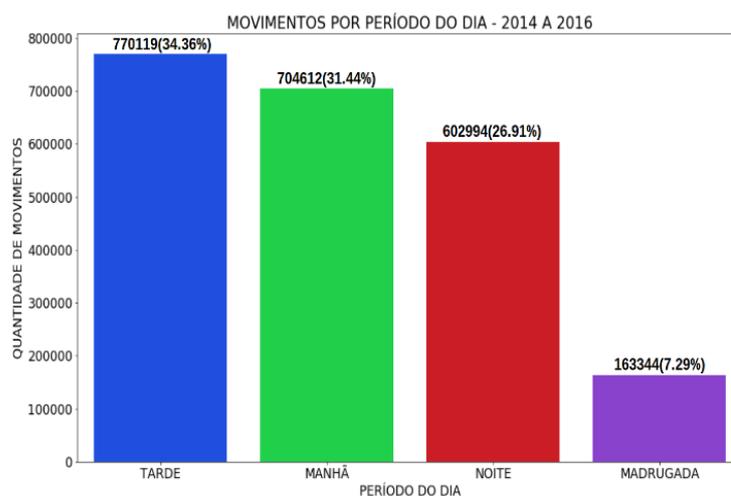


Figura 32 - Distribuição de voos pelo período do dia

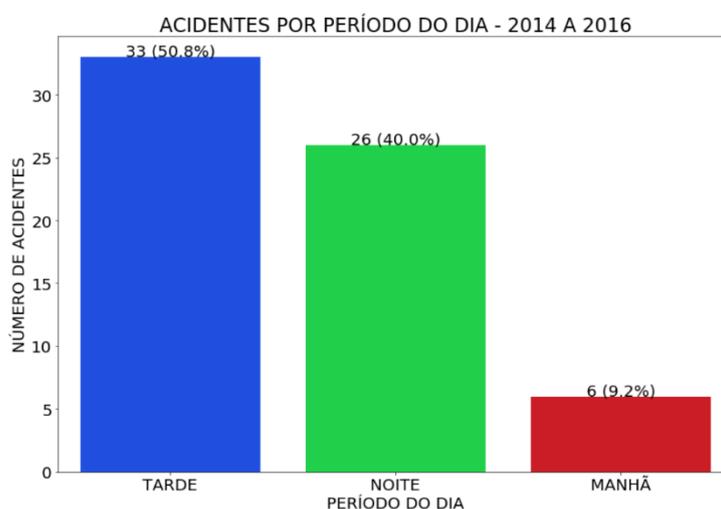


Figura 33 - Distribuição de acidentes pelo período do dia

Correlacionando os dados de movimento de aeronaves nas TMA I e TMA II com a quantidade de acidentes computados em cada dia da semana, tem-se a Fig.34 que mostra o sábado, mais uma vez, como o dia da semana mais crítico não só em termos absolutos como também em relação ao número de movimento de aeronaves.

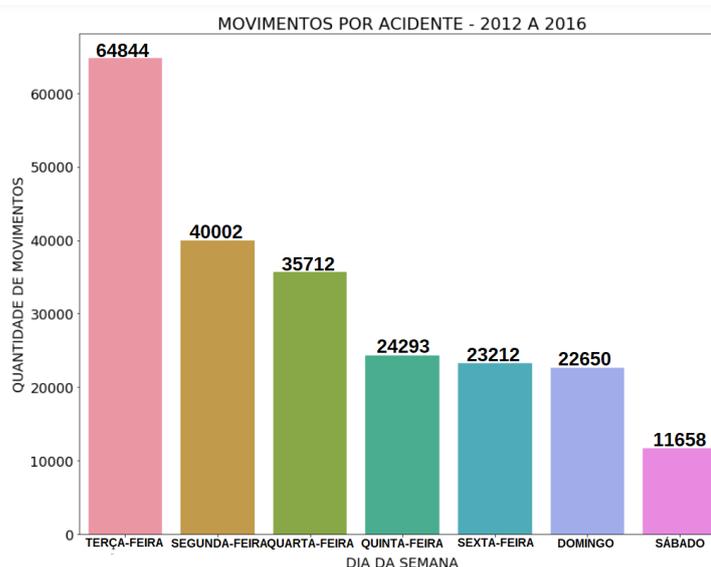


Figura 34 - Quantidade de acidentes por movimento de aeronaves nos dias da semana

A Fig. 35 mostra a correlação entre os dados de movimento de aeronaves nas TMA I e TMA II com a quantidade de acidentes computados em períodos do dia. Nota-se que durante o período da madrugada a ocorrência de acidentes é maior do que no restante do dia.

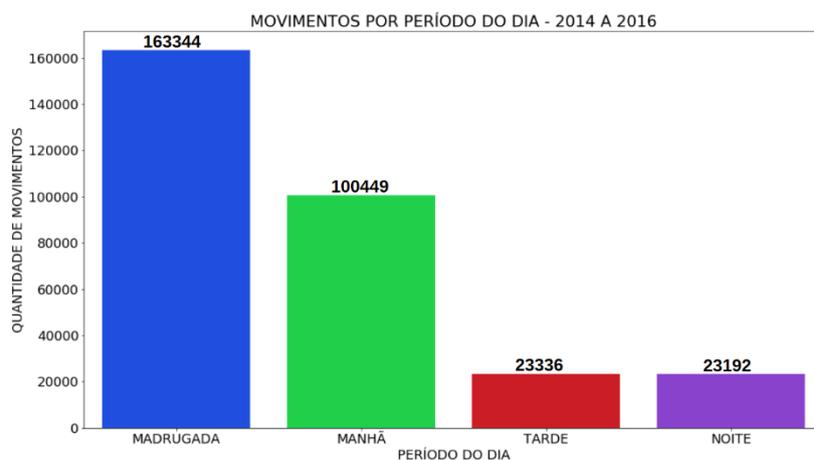


Figura 35 - Quantidade de acidentes por movimento de aeronaves nos dias da semana

5 CONSIDERAÇÕES FINAIS

Por meio da Análise Exploratória de Dados, foi possível extrair percepções dos acidentes aeronáuticos ocorridos no Brasil entre 2006 e 2017. As ferramentas utilizadas atenderam as necessidades dessa fase do ciclo de Ciência de Dados, sendo possível a realização da análise dos dados de forma colaborativa entre os pesquisadores.

Ao verificar os dados disponíveis na etapa de EDA, um conjunto de variáveis foi considerado para a visualização e análise. Como resultado da análise dos atributos dia da semana, período do dia e porte da aeronave, foi possível revelar, por meio das correlações realizadas, que o perfil mais suscetível a um acidente aeronáutico é o de uma aeronave de porte leve que voa no sábado à tarde.

Com o intuito de aprimorar o caso de estudo, e percebendo que o horário do dia pode ter alta influência sobre a ocorrência de acidentes, foi realizada uma análise que pudesse estimar o número exato de acidentes que ocorreram em cada intervalo de hora do dia. E por meio desta, foi possível constatar que o intervalo em que notadamente ocorreu mais acidentes foi no início da noite, e que houve também um acréscimo considerável no início da tarde. Entretanto, uma vez que os picos de ocorrências no início da noite não ocorrem mais cedo nos meses de Inverno, e nem mais tarde nos meses de Verão, não é sustentável afirmar que a variação da iluminação solar tem forte influência no aumento do número de ocorrências.

Identificou-se, também, por intermédio do perfil crítico encontrado, os locais do território nacional onde ocorreram mais eventos. A região do Estado de São Paulo destacou-se como uma das áreas com maior índice de acidentes, tanto na capital do Estado, como no interior. De posse desse resultado, procurou-se novos dados de modo a aprofundar ainda mais o estudo nessa região do país, e encontrou-se o número de movimentos de aeronaves nos principais aeródromos das TMA I e TMA II do estado de São Paulo publicados nos anuários do CGNA.

Não obstante, a ausência de dados apresentou-se como o maior fator limitante deste trabalho, pois sem o número exato de voos das aeronaves de todas as categorias e tipos, é impossível estabelecer uma relação proporcional da frequência de acidentes aéreos. A despeito dos dados do CGNA, estes dizem respeito apenas aos movimentos de aeronaves e não ao número de voos, além de constarem dados somente em relação aos principais aeródromos do país, excetuando-se aeródromos menores como os aeroclubes. Por conseguinte, uma vez que não foi possível nem sequer estabelecer uma relação entre a quantidade de acidentes em relação ao número de voos, não foi possível evidenciar informações com o grau de detalhamento requerido para depreender qualquer ilação que possa servir de base para uma mudança de planejamento dos profissionais que lidam com prevenção de acidentes.

Os resultados obtidos neste trabalho nos motivam a estabelecer uma continuidade visando ampliar as capacidades do ambiente e o aprofundamento da EDA. Assim sendo, para trabalhos futuros, projeta-se a exploração de dados referentes aos fatores mais determinantes para as ocorrências registradas, haja vista de que estes fatores podem demandar uma maior atenção por parte dos órgãos responsáveis, tanto pelo tráfego aéreo, como pela prevenção de acidentes aeronáuticos, visando mitigar o risco associado. Além disso, planeja-se fazer o cruzamento de tais informações com os registros de voo e de rotas aéreas no Brasil, visando uma análise proporcional dos eventos em relação às ocorrências e a identificação das rotas com maior frequência de ocorrências. Esse maior aprofundamento nas relações entre causas e efeitos das ocorrências aéreas, bem como a análise de casos mais críticos que, possivelmente, necessitam de uma maior priorização, permite uma antecipação de ações a serem tomadas e uma maior consciência situacional dos órgãos responsáveis por garantir a prevenção de acidentes aeronáuticos no Brasil.

Quanto às funcionalidades do ambiente para análise dos dados, além do ambiente desenvolvido e utilizado neste trabalho baseado na linguagem Python e seus módulos, foi possível vislumbrar que será vantajoso explorar novas ferramentas capazes de realizar diferentes análises e possivelmente novas inferências sobre os mesmos dados, como a ferramenta Orange³, Spyder⁴ ou a linguagem R.

AGRADECIMENTOS

Primeiramente, um especial agradecimento ao Instituto de Pesquisas e Ensaios em Voo (IPEV) pela oportunidade de divulgação desse artigo no 11º Simpósio de Segurança de Voo (SSV 2018). Destaca-se, ainda, a Divisão de C4ISR do Instituto de Estudos Avançados (IEAv), pelo o apoio de infraestrutura de laboratório e, em especial, a todos os professores, pesquisadores e colaboradores que contribuíram para o desenvolvimento do estudo demonstrado no artigo.

REFERÊNCIAS

- CENIPA, **FCA 58-1 Ocorrências Aeronáuticas: Panorama Estatístico da Aviação Brasileira**. 118p. Centro de Investigações e Prevenção de Acidentes Aeronáuticos. 2016.
- CAMARA, Gilberto; ASSIS, Luiz F.; QUEIROZ, Gilberto R.; FERREIRA, Karine R.; LLAPA, Eduardo; VINHAS, Lubia. **Big earth observation data analytics: matching requirements to system architectures**. Proceedings of the 5th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data. 2016.
- CHOUDHARY, Jayvant; OHRI, Anurag; KUMAR, Brind. **Identification of Road Accidents Hot Spots in Varanasi using QGIS**. Proceedings of National Conference on Open Source GIS: Opportunities and Challenges Department of Civil Engineering, IIT(BHU), 2015.
- EMC. **Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data**. Wiley, EMC Education Services, 1st Edition, 2015, 399 p.
- GRUS, Joel. **Data Science from Scratch: First Principles with Python**. O'Reilly Media, Inc., 1st ed., 2015, 330 p.
- WAGNER, Michael; LLORT, Gérman; MERCADAL, Estanislao; GIMÉNEZ, Judit; LABARTA, Jesús. **Performance Analysis of Parallel Python Applications**. Proceeding of International Conference on Computational Science, ICCS. 2017.

....

³ <https://orange.biolab.si>

⁴ <https://github.com/spyder-ide>