



Instituto Superior de Educação e Ciências

Trabalho de Projeto da Licenciatura em Ciências Aeronáuticas

Desenvolvimento de um sistema de tratamento de dados
a partir da OpenSky Network para a região de Lisboa

Autor: Miguel Barrameda Leite Salazar

Orientador: Professor Doutor Paulo Filipe Faria Machado

Data: Junho de 2023

Intencionalmente em Branco

Agradecimentos

Gostaria de agradecer, em primeiro lugar, à minha família que, durante toda a minha vida, esteve ao meu lado acreditando, incentivando e superando todos os desafios que envolvem, especialmente nos últimos 3 anos, uma mudança de país, escolhas de vida e decisões profissionais. O apoio incondicional garantiu a minha felicidade e estabilidade em meio à tanto.

Além disso, agradeço à minha namorada Manuela que, vendo de perto os desafios da vida acadêmica aliada ao forte interesse em aeronáutica, sempre se mostrou ouvinte, atenta e disposta a dizer o que era preciso ouvir e incentivar sempre quando era imperativo.

Agradeço também ao ISEC Lisboa por abrir suas portas em meio à um período pandémico complicado e, mesmo diante de todos os desafios e percalços ao longo dos últimos 3 anos, foi fundamental na minha caminhada dentro das Ciências Aeronáuticas. Agradeço aos meus colegas que também me receberam de braços abertos e sempre criaram um ambiente agradável e leve, mesmo com todas as dificuldades de uma faculdade.

Por fim, agradeço ao meu orientador Professor Paulo Machado que, desde as aulas lecionadas por ele bem como o desenvolvimento do trabalho, se mostrou disponível, interessado e capaz de transmitir os conhecimentos necessários e garantir meu grande interesse nos conteúdos das aulas. Um agradecimento também ao Coordenador e Professor Luís Santos que, apesar dos desafios dentro das Ciências Aeronáuticas, sempre se propôs a ajudar e a cobrar o melhor dos seus alunos.

Intencionalmente em Branco

Resumo

A obtenção de dados armazenados a partir da OpenSky Network aliada à programas capazes de processar *Big Data* permitem realizar análises e estudos para compreender o comportamento do fluxo de tráfego aéreo. Este trabalho, apoiando-se nesse contexto, visa criar uma ferramenta interativa e em *Real time*, no formato de *dashboard*, capaz de apresentar os dados advindos dessa base de dados para o ano de 2022 no Aeroporto Humberto Delgado (LPPT). O desenvolvimento dessa se deu por meio do R Studio, fazendo-se uso de livrerias como “*openskyr*” e “*rshiny*” que, aliadas aos conceitos de *Big Data*, capacitaram tal ferramenta a analisar o volume de chegadas e saídas, tanto no enquadramento do tempo quanto numa abordagem geográfica dos destinos e origens dos voos, além de uma abordagem às trajetórias das chegadas, tanto num perfil lateral como no vertical. Essa ferramenta, portanto, destinada à comunidade do ISEC Lisboa e à comunidade aeronáutica, é capaz de satisfazer a ausência de um intermediário entre *Big Data* e investigadores/estudantes, no que diz respeito a uma forma dinâmica de compreender tendências e padrões de tráfego aéreo.

Palavras-Chave: Open Sky Network, ADS-B, Air Traffic Management, Big Data, R Studio.

Intencionalmente em Branco

Índice

Agradecimentos	i
Resumo	iii
Índice de Figuras	vii
Índice de Tabelas	ix
Abreviaturas e Siglas	x
Definições	xii
1. Introdução	1
1.1. Generalidades	1
1.2. Âmbito	7
1.3. Motivação	7
1.4. Objetivos	8
1.5. Metodologia	9
1.6. Estrutura do Trabalho	9
2. Revisão Bibliográfica	11
3. Desenvolvimento do Tema	27
4. Análises do Resultado e Discussões	51
5. Conclusões e Recomendações	57
Bibliografia	59

Intencionalmente em Branco

Índice de Figuras

Figura 1 - Princípio de funcionamento do PSR (ICAO Asia 2007).	2
Figura 2 - Princípio de Funcionamento do SSR (ICAO Asia 2007).	3
Figura 3 - Protocolo ADS-B (Baek e Bang 2012).	4
Figura 4 - Princípio de Funcionamento do ADS-B (ICAO Asia 2007).	4
Figura 5 - Janela principal do Open Sky Network (Open Sky Network 2022).	6
Figura 6 - Janela selecionada de um voo rastreado pelo Open Sky Network (Open Sky Network, 2022).	6
Figura 7 - Perdas de mensagens do ADS-B (%) de acordo com a métrica (Strohmeier et al. 2014).	13
Figura 8 - VFE na TMA para o voo SAS410 com base nos dados da OSN e DDR2 (Polishchuk, Lemetti, e Sáez 2019).	18
Figura 9 - Perfil Lateral na TMA para o voo SAS 964 com base nos dados da OSN e DDR2 (Polishchuk, Lemetti, e Sáez 2019).	19
Figura 10 - Combustível consumido adicional por dia com base nos dados da OSN e DDR2 (Polishchuk, Lemetti, e Sáez 2019).	19
Figura 11 - Variação da Latitude e Longitude nas descolagens de LFBO. Dados do EUROCONTROL acima e da OSN abaixo (Perrichon, Gendre, e Klein 2022).	20
Figura 12 - Variação de Altitudes nas descolagens de LFBO. Dados do EUROCONTROL acima e OSN abaixo (Perrichon, Gendre, e Klein 2022).	21
Figura 13 - Perfil Lateral de chegadas na TMA de LFPO. Dados do ADS-B (Perrichon, Gendre, e Klein 2022).	21
Figura 14 - Perfil Lateral de chegadas na TMA de LFPO. Dados do EUROCONTROL (Perrichon, Gendre, e Klein 2022).	22
Figura 15 - Evolução de passageiro mundialmente entre 1945 e 2020 (ICAO 2021).	23
Figura 16 - Tráfego Regional e Mortes por COVID-19 no Brasil (Koelle e Barbosa 2021).	24
Figura 17 - Estrutura do package "openSkies" (Ayala et al. 2021)	25
Figura 18 - Camadas dos processos de Big Data.	27
Figura 19 - Código que busca as saídas de LPPT.	28
Figura 20 - Indicação dos dataframes LPPT_A e LPPT_D.	33
Figura 21 - Código que busca a trajetória da aeronave CS-PHC.	36

Figura 22 - Gráfico de volume de chegadas por hora do dia.....	40
Figura 23 - Mapa de volume de Destinos (saídas) para Europa e Norte de Africa.	41
Figura 24 - Mapa de volume de países de origem para Europa e Norte de Africa.....	42
Figura 25 - Mapa de volumes de países de origem mundialmente.	42
Figura 26 - Gráfico do Perfil Vertical de chegadas no Aeroporto de Lisboa.	45
Figura 27 - Mapa do Perfil Lateral de chegadas no Aeroporto de Lisboa.....	46
Figura 28 - Página Inicial do dashboard e Painel Lateral.....	47
Figura 29 - Página "Departures Volume/In Time".	48
Figura 30 - Página "Departures Volume/Geographically".	48
Figura 31 – Página “Arrival’s Routes”.....	49

Índice de Tabelas

Tabela 1 - Exemplo de informações providas pelo ADS-B e disponíveis publicamente (Schäfer, Lenders, e Martinovic 2013).....	14
Tabela 2 - Estrutura Inicial da função "get_airport_data"	29
Tabela 3 - Estrutura Inicial do API da OpenSky Network (OpenSky Network 2021) .	30
Tabela 4 – Dados do ficheiro "Airline Codes"	32
Tabela 5 – Dados do ficheiro "Aircraft Database"	33
Tabela 6 – Dados do ficheiro "Airport Codes"	33
Tabela 7 - Estrutura Final para o dataframe LPPT_A	34
Tabela 8 – Tipos de Dados para a Estrutura Final dos dataframes LPPT_D e LPPT_A	35
Tabela 9 - Estrutura da função "get_track_data"	36
Tabela 10 – Tipos de Dados para a Estrutura da função "get_track_data"	37
Tabela 11 - Recorte da Tabela de volume por tipo de aeronave	43
Tabela 12 - Recorte da Tabela de volume por Companhia Aérea	43
Tabela 13 - Recorte da Tabela de volume por País de Origem	43
Tabela 14 - Recorte da Tabela de volume por Aeroporto de Origem	44

Intencionalmente em Branco

Abreviaturas e Siglas

ACAS - Airborne Collision Avoidance System

ADS-B – Automatic Dependent Surveillance Broadcast

ATC – Air Traffic Control

ATCO – Air Traffic Control Operator

ATM – Air Traffic Management

CASCADE - Co-operative ATS through Surveillance and Communication Applications
Deployed in ECAC

FLARM – Flight Alarm

GLONASS – Global Navigational Satellite System

GNSS - Global Navigation Satellite System

GPS - Global Positioning System

ICAO – International Civil Aviation Organization

OSN – Open Sky Network

PSR – Primary Surveillance Radar

RADAR – Radio Detection and Ranging

SESAR - Single European Sky ATM Research

SSR – Secondary Surveillance Radar

TMA – Terminal Maneuvering Area

VANT – Veículo Aéreo Não Tripulado

WTC – Wake Turbulence Category

Intencionalmente em Branco

Definições

Bearing – Rumo de uma aeronave

Big Data – Conjunto de dados que reúnem grande variedade, volume e velocidade em seu tráfego e armazenamento.

Data Engineering – Atividade que permite a capacidade de um sistema captar e utilizar dados.

Data Science – Atividade estatística que faz uso de ferramentas da matemática e da ciência da computação para permitir a realização de análises e avaliações de dados tanto qualitativos quanto quantitativos.

Data Visualization – Forma de representar visualmente dados tratados de forma a obter rápido entendimento da informação transmitida.

Efeito Doppler – Fenômeno físico marcado pela mudança da frequência de uma onda emitida em uma fonte em movimento em relação a um observador.

Epoch Time – Também conhecido como Unix Time, é uma representação numérica que define o número de segundos decorridos desde o início de uma referência temporal conhecida como "época". Essa época geralmente é definida como 1º de janeiro de 1970 às 00:00:00 UTC (Coordinated Universal Time).

Modo S – Modo do Transponder usado no TCAS e no ACAS II

Security – Proteção da aviação contra atos ilícitos

Safety – Segurança Operacional

Intencionalmente em Branco

1. Introdução

Este trabalho visa relatar e descrever todo o embasamento teórico e o desenvolvimento prático que culmina na implementação de um sistema de tratamento de dados de ADS-B da rede OpenSky Network para a região de Lisboa. Traduz-se numa ferramenta capaz de utilizar as informações obtidas pela OSN (aeroportos de origem e destino, volumes de chegada e saída) para apresentá-las ao utilizador de forma que esse seja capaz de realizar estudos e análises. Utiliza como ponto de partida os conceitos de Data Engineering, Data Science e Data Visualization através da linguagem R e do R Studio.

1.1. Generalidades

O aumento do tráfego aéreo ao redor do mundo traz consigo a necessidade de desenvolver tecnologias capazes de dinamizar as operações aéreas. Essa dinamização, por sua vez, permite o aumento da capacidade de tráfego no mundo, e assim por diante (Schäfer et al. 2014). O rastreamento de aeronaves para a gestão e controlo do tráfego aéreo, apesar de não ser uma tecnologia recente, vêm se modernizando para atender à crescente demanda de tráfego aéreo, e tal modernização deu-se, principalmente, pelo aumento da precisão, quantidade e rapidez na transmissão da informação das aeronaves (Seymour 2018), que se traduz em uma operação extremamente confiável e precisa. Este capítulo se destina a apresentar tudo aquilo que atualmente se discute acerca do rastreamento de aeronaves, contemplando também estudos e casos que materializam toda a teoria discutida.

Levando-se em consideração a aviação moderna, podemos considerar que o principal sistema capaz de determinar posição é o RADAR que, essencialmente e na maioria dos casos, é composto por um Primary Surveillance Radar (PSR) e Secondary Surveillance Radar (SSR) (Abdulaziz et al. 2019; ICAO Asia 2007). O primeiro, PSR, é considerado o estado da arte do rastreamento de aeronaves. Desenvolvido nos anos 40, consiste na transmissão de sinais em alta frequência que, quando entram em contato com a aeronave, são capazes de determinar sua posição em relação a estação e seu *bearing* (Wesson, Humphreys, e Evans 2014; Gilbert 1973). É um sistema que funciona através do Efeito *Doppler*, ou seja, transmite a onda e o ato de “bater e voltar” é o que permite determinar posição e direção. Por isso, é considerado como um sistema independente e, apesar de

fornecer poucas informações e não garantir tamanha precisão, ainda é utilizado por gerar leituras confiáveis (Strohmeier et al. 2014; Miranda 2022a). Já o segundo, SSR, possui um funcionamento diferente, ao passo que a estação de solo emite “perguntas” às aeronaves dentro do seu alcance, numa frequência de 1030 MHz, e aguarda a resposta, emitida na frequência de 1090 MHz (sistema pergunta-reposta), para adquirir as informações de voo como a identificação da aeronave, posição, altitude. Por isso, depende de um dispositivo à bordo da aeronave – o Transponder – para compreender a pergunta feita pela estação e responder com as informações desejadas (Miranda 2022b). As imagens abaixo representam, de forma esquemática, o funcionamento do PSR (Figura 1) e SSR (Figura 2).

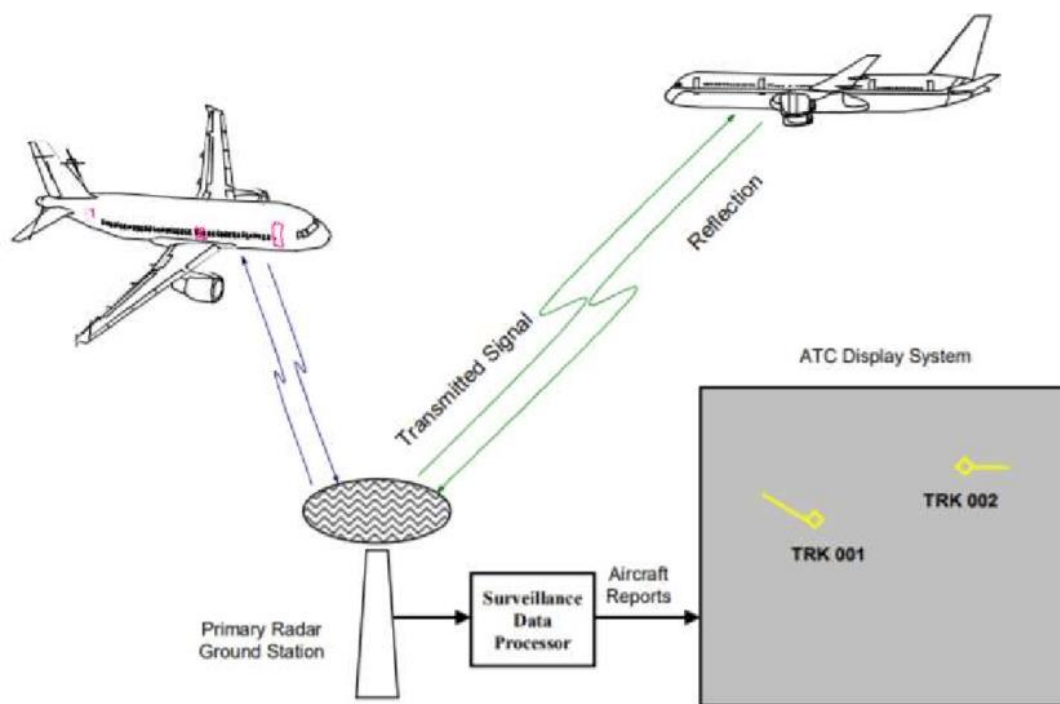


Figura 1 - Princípio de funcionamento do PSR (ICAO Asia 2007).

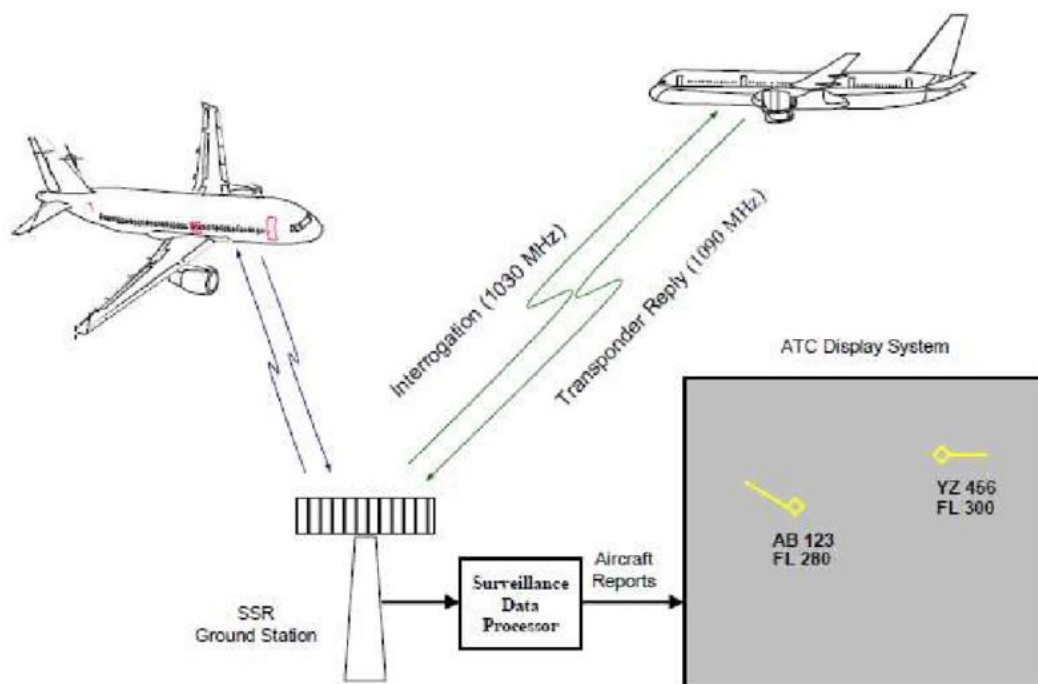


Figura 2 - Princípio de Funcionamento do SSR (ICAO Asia 2007).

Tendo em vista o aumento do fluxo de tráfego aéreo e limitações inerentes das tecnologias, o sistema de RADAR apresentado acima passou a se tornar sobrecarregado e sujeito à problemas, se tornando incapaz de atender à alta demanda. Por isso, o surgimento da tecnologia do Automatic Dependent Surveillance Broadcast (ADS-B) se dá com o intuito de satisfazer às necessidades do espaço aéreo (Abdulaziz et al. 2019). O ADS-B, de acordo com a ICAO, é “um meio pelo qual a aeronave, veículo de Aeródromo e outros objetos podem automaticamente transmitir e/ou receber dados como identificação, posição e dados adicionais, quando apropriado, em modo de transmissão por uma ligação de dados” (ICAO 2016). Trata-se, portanto de um sistema alimentado pelos dados de GNSS (GPS, Galileu, GLONASS, etc.) da aeronave para determinar a sua posição e gerar uma mensagem (Baek e Bang 2012). Estas mensagens, por sua vez, estão constantemente transmitidas para estações no solo (ADS-B OUT) e para outras aeronaves (ADS-B IN) (Rodrigues 2010). A Figura 3 mostra, através de um esquema, todo o protocolo para o sistema ADS-B.

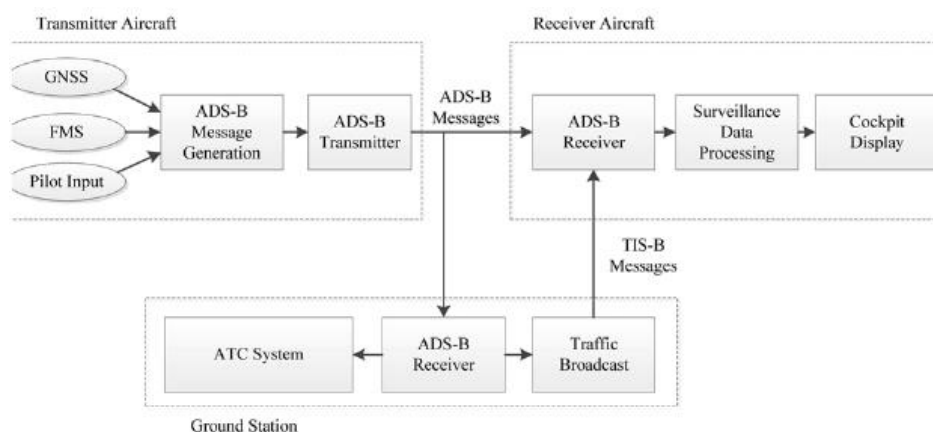


Figura 3 - Protocolo ADS-B (Baek e Bang 2012).

Portanto, seu propósito é determinar a posição da aeronave e, de forma automática e repetitiva, transmitir a informação juntamente com *callsign*, altitude, *heading* e identificação da aeronave. Uma das grandes vantagens do sistema ADS-B é que esse é compatível com o sistema *Mode S*, ao passo que a transmissão das informações também se dá na frequência de 1090 MHz, facilitando enormemente a aplicação desse novo método. A Figura 4 representa esquematicamente o princípio de funcionamento desse sistema.

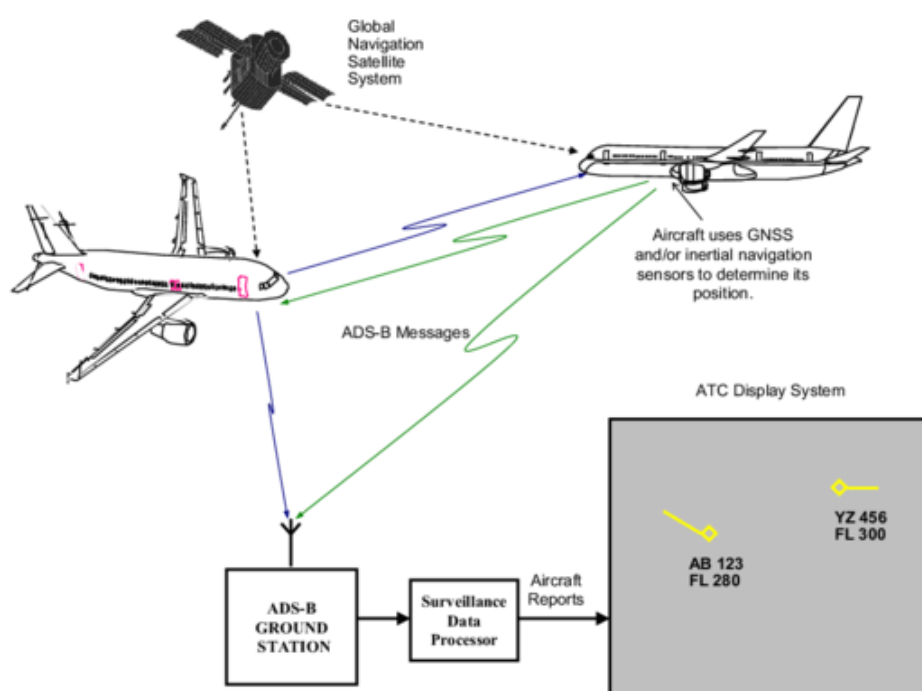


Figura 4 - Princípio de Funcionamento do ADS-B (ICAO Asia 2007).

O mecanismo do ADS-B, porém, possui uma série de questões relacionadas com *Security*, tendo em vista que foi concebida com um foco muito maior na *Safety* (Seymour 2018). O sistema possui uma série de vulnerabilidades, na medida em que as mensagens transmitidas não possuem nenhuma encriptação ou assinatura digital e, por isso, podem ser inseridas novas mensagens, ou essas podem ser modificadas e até mesmo apagadas (Seymour 2018; Strohmeier et al. 2014). As propostas para a criação de camadas de segurança para o ADS-B muitas vezes esbarram em uma série de questões regulamentais e técnicas. Além disso, qualquer modificação visando segurança cibernética no ADS-B obrigará uma diminuição na capacidade de transmitir mensagens e deixará o sistema sempre sob “vigilância” do RADAR, mesmo sendo o ADS-B o sistema que era suposto substituir os convencionais PSR e SSR (Wesson, Humphreys, e Evans 2014). Apesar de possuir questões de segurança para total aplicabilidade nas operações aeronáuticas, as mensagens trocadas a partir do sistema ADS-B são armazenadas em grandes bancos de dados e, tendo em vista a natureza das mensagens e a frequência de transmissões, possui um alto valor. Por essa mesma razão, diferentes organizações e projetos se utilizam dessa grande base de dados para desenvolver estudos e ferramentas para favorecer à comunidade aeronáutica, tanto na organização e disponibilização quanto na apresentação desses dados. É o caso da OpenSky Network, uma comunidade virtual que identifica e posiciona no mapa as aeronaves que transmitem a partir do sistema ADS-B (como também Modo S, TCAS e FLARM) em direto, e ainda fornece informações específicas de cada aeronave, como por exemplo a origem, destino, equipamento, fonte de transmissão, altitude (Open Sky Network 2022). Além de prover o mapa instantâneo, a OSN cria uma base de dados que armazena os mais variados dados acerca de voos realizados: hora de decolagem prevista e real, hora de aterragem prevista e real, atrasos, trajetória de voo, perfil de altitudes e velocidades ao longo do percurso, aeroportos, volume de tráfego. As duas imagens abaixo representam a funcionalidade de tempo real da rede, seja a análise mais holística (Figura 5) como também uma análise individualizada de determinado voo (Figura 6).

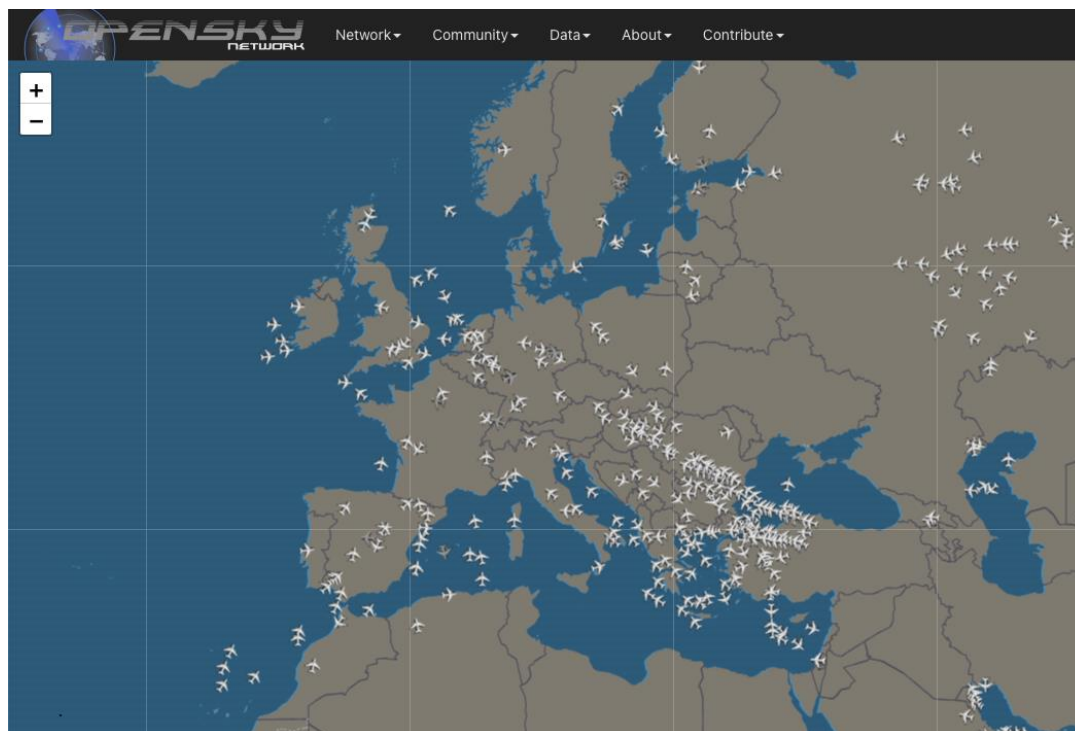


Figura 5 - Janela principal do Open Sky Network (Open Sky Network 2022).

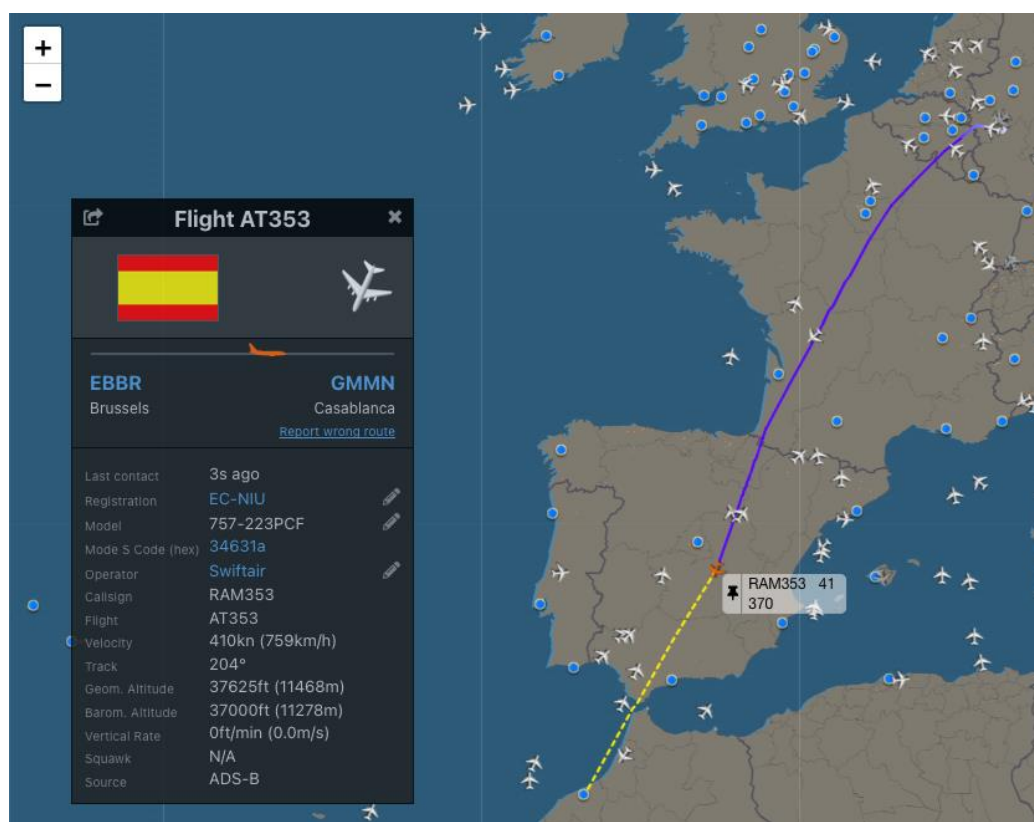


Figura 6 - Janela selecionada de um voo rastreado pelo Open Sky Network (Open Sky Network, 2022).

A grande quantidade de dados disponíveis, coletados a partir das transmissões ADS-B, para serem processadas e serem capazes de gerar análises e inferências à respeito dessas, necessitam de ferramentas computacionais, como por exemplo, a linguagem R e o programa R Studio (Comitz e Kersch 2016). De acordo com (Kumari e Verma 2019), vivemos em uma sociedade em que indivíduos e organizações se destacam de acordo com a sua capacidade de extrair melhor o conhecimento a partir da maior quantidade de informações possíveis, permitindo desenvolver uma visão holística e realizar previsões mais precisas.

O mesmo vale para *Big Data*, na medida em que, as análises feitas são mais precisas conforme o maior número de dados fornecidos. Dentro das inúmeras capacidades da linguagem R está a aptidão para lidar, ajustar e manipular grandes quantidades de dados para apresentá-los de maneira coerente e compreensível, tocando nas esferas de *Data Engineering*, *Data Science* e *Data Visualization*. Nesse sentido, a linguagem R se mostra não só capaz como ideal para o processamento dos dados armazenados pela OSN, ao passo que possui uma série de mecanismos dentro da própria estrutura da linguagem que facilita a interface com o usuário (Gandrud 2018; Kumari e Verma 2019). Além disso, a linguagem R permite a aplicação do conceito de *Reproducible Research*, ou seja, a capacidade de se estruturar realizar uma pesquisa que possa ser facilmente replicada em outro contexto, com outros parâmetros (Gandrud 2018).

1.2. Âmbito

Este trabalho reside no âmbito da Ciência de Dados, Air Traffic Management e do aspecto científico associado à pesquisa e desenvolvimento de critérios para a realização de análises e inferências.

1.3. Motivação

A realização deste trabalho possui como motivação principal a grande quantidade de dados coletados e reunidos ao longo de muito tempo, contendo informações que, se tratadas de certa forma, podem prover análises valiosas na busca pela dinamização do *Air Traffic Management*. A existência e desenvolvimento deste trabalho sugere que existe algo em falta dentro do nicho da aviação: atualmente existem inúmeras ferramentas capazes de coletar dados de voo e os apresentarem de diferentes maneiras – como por

exemplo OpenSky Network, FlightRadar24, Flight Aware. Este trabalho busca adquirir estas informações e utilizá-las para alimentar uma ferramenta capaz de fornecer, diferentemente das instituições listadas acima, análises e comportamentos retroativos com base no histórico existente, de forma a permitir a comunidade estudar e buscar padrões ou qualquer outra informação relevante.

A elaboração deste trabalho se sustenta a partir da formulação de determinadas perguntas, ao passo que o motivam, movem e direcionam: Como a comunidade do ISEC Lisboa busca atualmente informações sobre tráfego aéreo, e o quão fácil é obter o “raw data” para fazer as análises desejadas? Quais os benefícios que um membro de ISEC Lisboa terá com a existência deste trabalho? O que diferencia este trabalho dos outros que se utilizam da mesma base de dados?

A começar pela primeira pergunta, “Como a comunidade do ISEC Lisboa procura atualmente informações sobre tráfego aéreo, e o quão fácil é obter o “raw data” para fazer as análises desejadas?”. Qualquer aluno ou docente, nos dias de hoje, necessita de aceder a algum website como OpenSky Network, Flight Radar 24 ou Flight Aware, por exemplo, para aceder a um mapa contendo os voos em tempo real. Já a obtenção desses dados e dados retroativos podem ser feitos, porém com uma série de restrições e permissões académicas ou institucionais; em segundo: “Quais os benefícios que um membro de ISEC Lisboa terá com a existência deste trabalho?”. Por se tratar de uma ferramenta interna, qualquer um que faça parte da comunidade do ISEC Lisboa terá acesso aos dados fornecidos e poderá trabalhá-los da maneira que o convir e o auxiliar; por fim: “O que diferencia este trabalho dos outros que se utilizam da mesma base de dados?”. O que o diferencia é que, além de ser feito por alunos para a comunidade, será uma ferramenta interativa que se adapta às necessidades do investigador, e por isso faz dela muito mais versátil e capaz de atender qualquer ideia futura de outros projetos, ou até mesmo curiosidades.

1.4. Objetivos

O objetivo deste trabalho é criar uma ferramenta interativa capaz de analisar o volume de tráfego aéreo – tanto chegadas quanto saídas - no Aeroporto Humberto Delgado, em Lisboa, tendo como referência principal a otimização do tráfego aéreo para um Aeroporto de grande importância no cenário europeu, através da apresentação dos dados de uma

forma que seja possível estabelecer padrões relativamente ao tráfego. Assim, tal sistema servirá à comunidade aeronáutica ao passo que é capaz de transformar grandes quantidades de dados em inferências e informações valiosas. Essas podem ser utilizadas, à critério do utilizador, na busca pela otimização do Air Traffic Management já que, quanto maior o dinamismo nos processos e nas análises, maior a capacidade de determinado espaço aéreo comportar tráfego. Além disso, dentro dos objetivos da criação desta ferramenta, também busca-se estruturar todo o trabalho (linhas de código e métodos de visualização) de forma que obedeça aos conceitos de *Reproducible Research*, permitindo a extrapolação deste mesmo estudo aqui relatado para outros aeroportos, períodos e contextos. Cumprindo estes objetivos, será possível sanar uma carência não só dentro do ISEC Lisboa, mas também da comunidade aeronáutica no que diz respeito à habilidade de compreender comportamentos, tendências, históricos e fazer análises com base em dados de larga escala e alto valor informativo.

1.5. Metodologia

Para a realização deste trabalho, utilizar-se-á o programa R Studio para buscar os dados da OpenSky Network para o ano de 2022 no Aeroporto Humberto Delgado (LPPT). Com a escolha destes dados, estes serão modelados de forma que atendam e sejam capazes de gerar visualizações voltadas à análise dos volumes de chegadas e saídas, em função de parâmetros temporais ou geográficos. Por fim, dentro do R Studio, a livreria “rshiny” será responsável pela criação de um *dashboard* capaz de organizar e apresentar tudo aquilo que for gerado em termos de tabelas, gráficos e mapas.

Portanto, assume-se o método Dedutivo a partir da observação, tratamento e apresentação dos dados fornecidos pela OpenSky Network. As análises realizadas a partir da modelagem das informações no R Studio são extensivas, ao passo que consideram uma amostra relativa à região de Lisboa, observando todo o fluxo de tráfego aéreo desse espaço. Além disso, os resultados e conclusões obtidas podem ser consideradas do tipo híbrida, considerando que valores quantitativos (número de descolagens e aterragens, companhias e aeronaves) nos permitem obter conclusões qualitativas a respeito do comportamento estudado.

1.6. Estrutura do Trabalho

Este trabalho é estruturado a partir do Capítulo 2, referente à Revisão Bibliográfica, de suma importância para apresentar como que o trabalho possui relevância na atualidade com base em outros estudos; o Capítulo 3, Desenvolvimento, que apresenta tudo aquilo que há de disponível a ser utilizado, além da descrição dos processos que levaram à concretização dos objetivos propostos, além apresentação dos critérios estabelecidos no tratamento dos dados; seguido do Capítulo 4, Apresentação e Discussão dos Resultados, que visa avaliar se o desenrolar do trabalho fora capaz de atender aos objetivos e de que forma os resultados dialogam com as expectativas; e, por fim, o Capítulo 5, Conclusão, que discute e sacramenta a existência da ferramenta criada, além da importância de sua manutenção.

2. Revisão Bibliográfica

O segundo capítulo tem como objetivo desenvolver tudo aquilo que já foi discutido, estudado e apresentado dentro do universo acadêmico no âmbito dos principais temas deste trabalho, nomeadamente o sistema ADS-B, a Open Sky Network e a linguagem R, aliada ao R Studio e ao R Shiny. Para isso, recorre à diferentes artigos, livros e outras fontes académicas, abordando tais estudos à luz das premissas, hipóteses e resultados. Desta maneira, é possível melhor compreender não só a relevância deste tema, como também a importância da realização deste trabalho.

Em primeiro plano, é de extrema relevância referenciar o processo de implementação e consolidação do ADS-B como sistema *Next Generation* para rastreamento e controlo do tráfego aéreo. Como já abordado neste documento, o ADS-B tem como princípio de funcionamento a transmissão automática e constante de mensagens contendo informações de voo entre aeronaves num mesmo espaço aéreo. Este método reduz significativamente a intervenção do Air Traffic Control (ATC), além de reduzir o consumo de combustível e os custos, principalmente os associados à instalação e manutenção das estruturas que compõem o sistema RADAR.

De acordo com a Tese “ADS-B - Automatic Dependent Surveillance Broadcast: estudo do impacto em Portugal”, a implementação do ADS-B na Europa foi concretizada após uma série de medidas que visavam regular e unificar o espaço aéreo europeu como único. O lançamento do *Single European Sky*, em 2004, marca o ponto de partida nos alicerces para estabelecer o ADS-B. Juntamente com o Céu Único Europeu foi criada a *Single European Sky ATM Research* (SESAR), que é o viés tecnológico do Céu Único; e também o *Co-operative ATS through Surveillance & Communication Applications Deployed in ECAC* (CASCADE), uma divisão do SESAR focada na análise, segurança e sustentabilidade nos processos de homologação do ADS-B como ferramenta de vigilância aérea (Rodrigues 2010). Atualmente, o ADS-B possui larga cobertura ao longo da Europa e por isso, além de ser extremamente automatizado e barato, é também capaz de gerar grandes quantidade de dados e informações para análise, especialmente no âmbito académico dado a facilidade ao acesso desses (Rodrigues 2010; Schultz, Rosenow, e Olive 2022), mesmo que possua limitações no âmbito da *Security* e *Safety* para que seja integrado totalmente ao sistema *NextGen* de ATC (Strohmeier et al. 2014).

O artigo “Surveillance Radar System Limitations and the Advent of the Automatic Dependent Surveillance Broadcast system for Aircraft Monitoring” descreve não só o sistema que precedia o ADS-B, mas também aborda as vantagens, desvantagens e desafios associados à sua implementação. De acordo com Abdulaziz et al., 2019, as vantagens do ADS-B são a capacidade de dar cobertura de sinal em locais que o RADAR não era capaz de chegar, nomeadamente em navegações transoceânicas; aumento da eficiência e dinamismo nas operações; aumento da *Safety*; melhora na visibilidade, já que o sinal de transmissão não se deteriora com a distância (ou a aeronave está no alcance, ou não está; menor impacto ao ambiente, por conta da redução das instalações físicas. Assim, o sistema permite que o ATC tenha capacidade de diminuir a separação entre aterragens e descolagens sem que haja qualquer comprometimento à *Safety*, garantindo assim menores ocorrências de esperas para realização de aproximações, resultando na diminuição de custos associados (Abdulaziz et al. 2019). Por outro lado, o estudo considera também as desvantagens associadas ao sistema ADS-B: dependência dos transponders; possíveis perdas de posição GPS por conta de satélites indisponíveis; questões relacionadas à estabilidade, integridade e credibilidade relacionadas ao conteúdo do sistema ADS-B (Abdulaziz et al. 2019; Seymour 2018).

Tendo em vista o contexto que culminou na criação do ADS-B e sua posterior implementação como sistema de vigilância aéreo somado às suas características e barreiras, o documento “Realities and Challenges of NextGen Air Traffic Management: The Case of ADS-B” estuda a qualidade e confiabilidade das mensagens transmitidas pelo ADS-B à luz da *Security*, ou seja, se os métodos utilizados por esse sistema são capazes de proteger a operação de atos ilícitos (Seymour 2018). Para conseguir avaliar com precisão o canal de comunicação do ADS-B e suas mensagens, foram utilizados 53.626.642 registros de comunicação (Strohmeier et al. 2014). Foi constatado que existe uma grande perda de mensagens, quando consideradas as especificações e o comportamento das aeronaves rastreadas. Portanto, na tentativa de determinar possíveis fragilidades do sistema ADS-B, as perdas foram relacionadas com as métricas distância do recetor, hora do dia e quantidade de aeronaves dentro do mesmo alcance. A Figura 7 ilustra tais perdas em função dos parâmetros escolhidos: Em relação à distância (a), nota-se considerável perda até os 50 km de distância e também após os 280 km (a). Já no que diz respeito a hora do dia (b), as perdas são menores durante à manhã e à noite, ou seja, em horários fora do pico, quando há menos aeronaves atuando e transmitindo dentro de

um mesmo espaço aéreo. Por fim, quanto à quantidade de aeronaves a transmitir e receber ADS-B no mesmo espaço (c), as perdas crescem conforme o aumento do tráfego também cresce.

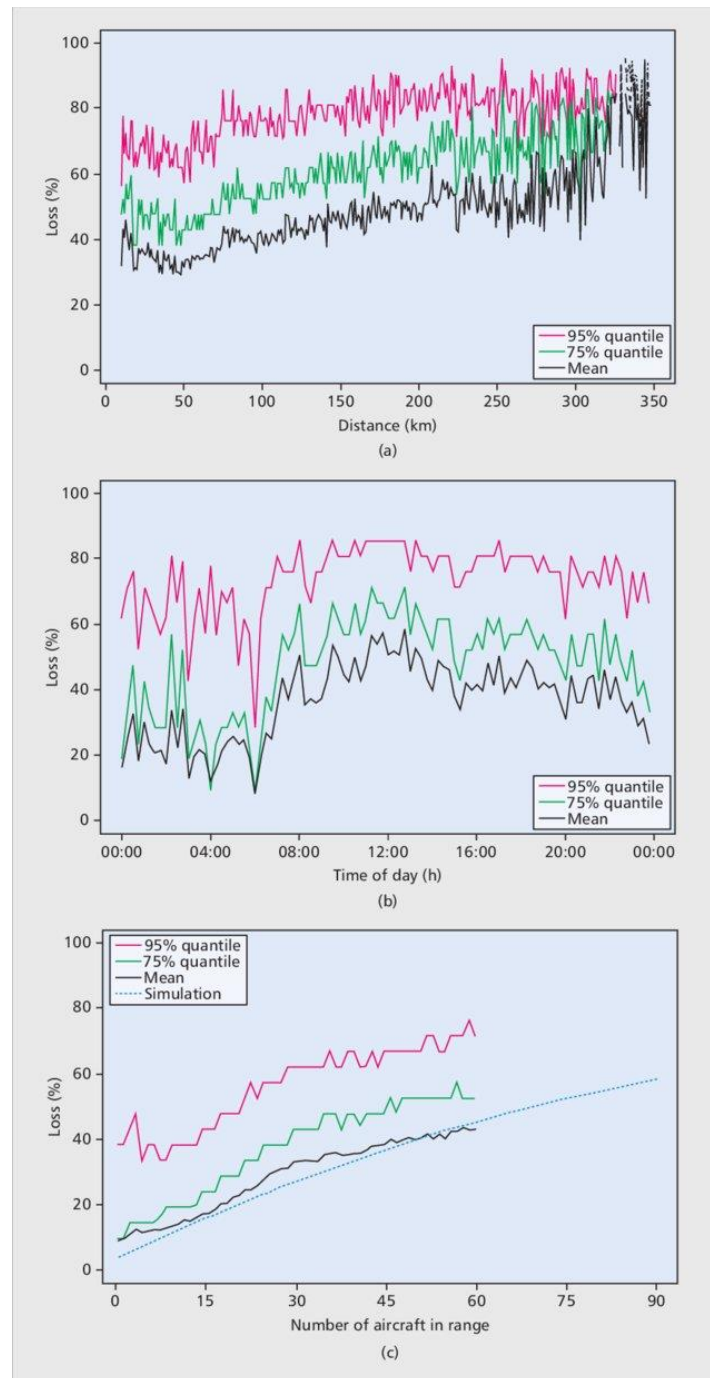


Figura 7 - Perdas de mensagens do ADS-B (%) de acordo com a métrica (Strohmeier et al. 2014).

As causas para a ocorrência dessas perdas, após estudos, é a Colisão de Mensagens (Strohmeier et al. 2014). Porém, não se trata só do encontro de mensagens ADS-B, já que

o Modo S, Modo A e Modo C operam na mesma banda de frequência (1090 MHz) e por isso influenciam na quantidade de informação a ser trocada em simultâneo. Por isso, o ADS-B se mostra muito instável especialmente em zonas terminais e horários de pico, o que deve ser levado em consideração no desenvolvimento dos protocolos de segurança e de operação, bem como na instalação de recetores (Strohmeier et al. 2014).

Levando em consideração as questões relacionadas à capacidade do sistema ADS-B de proteger-se de intervenções ilícitas e ciberataques, a esfera académica e as instituições de ATC têm se voltado às fragilidades do ADS-B para criação de medidas de mitigação em prol da *Security* (Seymour 2018). O relatório “Optimising the OpenSky Network” desenvolve que a principal ameaça à certo espaço aéreo com aeronaves a transmitir mensagens via ADS-B é de ataques no solo por grupos ativistas ou *hackers* (Seymour 2018). Tais ataques são capazes de transmitir mensagens falsas de ADS-B para o ATC, além de alterar mensagens, o que pode se traduzir, por exemplo, em mudar coordenadas de uma aeronave e fazer com que o ATC tenha em mente uma posição para a aeronave que não é real. O documento “Experimental Analysis of Attacks on Next Generation Air Traffic Communication” se debruça ainda mais acerca dos ataques ao sistema ADS-B, na medida em que os analisa, simula e mede os impactos que cada tipo de quebra de segurança pode trazer, através do tipo de informações que o sistema é capaz de prover, como mostra a Tabela 1. Inicialmente, esses ataques são divididos em passivos e ativos. Os primeiros são caracterizados e originados pela facilidade na obtenção dos dados transmitidos pelo ADS-B, que podem posteriormente serem utilizados para um ataque planeado e direcionado a partir das informações que os dados carregam, o que traz à tona discussões relativas à privacidade e, conseqüentemente, espionagem (Schäfer, Lenders, e Martinovic 2013).

Tabela 1 - Exemplo de informações providas pelo ADS-B e disponíveis publicamente (Schäfer, Lenders, e Martinovic 2013)

Call Sign	AAA	Flight No.	AAA
ICAO ID	AAA	Owner	AAA
Country	AAA	Start	AAA
Position	AAA	Destination	AAA
Altitude	37700 ft	Sched. Arrival	19:25
Heading	144°	Aircraft Model	Airbus A320-214
Speed	395 kn	Seats	126-168
Climbing rate	896 ft/min	Engine	CFM56-5B4/P

Por outro lado, os ataques do tipo ativo possuem gravidade mais severa ao passo que podem afetar diretamente à segurança do tráfego aéreo, nomeadamente na visualização do tráfego por parte do ATC bem como sistemas como o TCAS. Tais ataques estão sempre contidos no âmbito de injetar, deletar e modificar mensagens (Schäfer, Lenders, e Martinovic 2013). *Ghost Aircraft Injection*, ou Inserção de Aeronaves Fantasmas, é o ataque caracterizado por apresentar à qualquer recetor de ADS-B uma aeronave que não se existe mas que, em determinadas condições como baixa visibilidade, pode criar constrangimentos relacionados à aterragens, descolagens, separações e até mesmo no *ground Traffic* (tráfego no solo) (Schäfer, Lenders, e Martinovic 2013); *Ghost Aircraft Flooding* (“Inundação” de Aeronaves Fantasmas) consiste no mesmo princípio do anterior, porém múltiplas aeronaves em simultâneo, o que causa não só uma poluição visual para o ATC mas também a incapacidade do espaço aéreo de receber mais tráfego mesmo que, na realidade, não exista tal quantidade (Schäfer, Lenders, e Martinovic 2013; Strohmeier et al. 2014); *Virtual Trajectory Modification* (Modificação Virtual de Trajetória) são ataques que ou modificam a posição transmitida por uma aeronave, ou deleta a mensagem contendo a informação correta e insere uma nova mensagem com a informação falsa. Leva em consideração a imprecisão de um sistema por exemplo ADS-B+PSR e, mesmo que a mudança seja pequena, pode alterar as instruções do ATC para a aeronave (Schäfer, Lenders, e Martinovic 2013; Miranda 2022a); *False Alarm Attacks* (Ataques de Alarme Falso) consiste na modificação ou deletar e inserir uma transmissão com o intuito de forjar uma situação de perigo em determinada aeronave, por exemplo um sequestro (*hijacking*), de forma a desviar a atenção dos ATCOs e alterar com o fluxo de trabalho normal (Schäfer, Lenders, e Martinovic 2013); *Ground Station Flooding* (“Inundação” da Estação de Solo) é descrito por ATCOs como devastadores ao passo que podem, em áreas de demasiado tráfego, impedir que sistemas de alerta de colisão passem a ser ineficazes (Schäfer, Lenders, e Martinovic 2013); *Aircraft Disappearance*

(Desaparecimento de Aeronaves) consiste em deletar todas as mensagens transmitidas por determinada aeronave, o que impede o ATC, outras aeronaves ou o TCAS de levá-la em consideração (Schäfer, Lenders, e Martinovic 2013); *Aircraft Spoofing* (Falsificação da Aeronave) trata-se de falsificar o endereço 24bit ICAO para conseguir alterar os dados de uma aeronave (Schäfer, Lenders, e Martinovic 2013).

Tendo em vista a existência de diversos métodos para atacar as vulnerabilidades do sistema ADS-B, o relatório “Can Cryptography Secure Next Generation Air Traffic Surveillance?” busca apresentar formas de armar o ADS-B com defesas capazes de resistir aos ciberataques. É preciso levar em consideração que qualquer mudança ou renovação de protocolos de criptografia para proteger o ADS-B pode esbarrar em uma série de barreiras como: o ADS-B ser um protocolo internacional, e por isso deve ser compatível com diferentes políticas e disponibilidades tecnológicas; o ADS-B opera sob um espectro limitado de frequência; o ADS-B é limitado e sujeito à interferência por parte de transmissões do Modo S, por se propagarem em frequências similares, e qualquer modulação da mensagem do ADS-B pode resultar no aumento de interferências e na redução da capacidade operacional (Schäfer, Lenders, e Martinovic 2013; Wesson, Humphreys, e Evans 2014); o ADS-B opera num ambiente que, do ponto de vista criptográfico, não é confiável e, por isso, qualquer software de segurança estará disponível para acesso de qualquer parte (Wesson, Humphreys, e Evans 2014).

Tendo em vista a importância de se avaliar possíveis implementações de camadas de segurança no âmbito criptográfico, mas também levando em consideração as condições de fronteira descritas acima, encontra-se como caminhos possíveis a *Symmetric-Key Cryptography* (Criptografia por Chave Simétrica) e *Asymmetric-Key Cryptography* (Criptografia por Chave Assimétrica) (Wesson, Humphreys, e Evans 2014). A primeira parte da premissa de que o receptor e o transmissor possuam e compartilham de uma chave criptografada secreta, logo todas as mensagens e códigos de autorização de mensagens (MACs) são impossíveis de serem criados ou previstos. Já a segunda opção, *Asymmetric-Key Cryptography*, consiste na distribuição de pares de chaves público-privadas (*public-private keys*) em que cada usuário se conecta a sua chave através do *Certificate Authority* (ICAO ou FAA podem assumir esse papel, por exemplo. Nesse sentido, um usuário A é incapaz de forjar uma chave de um usuário B com sua própria chave. Logo, se alguma chave é comprometida, só se perde o par de chaves, e não todo o sistema, como é no caso da *Symmetric-Key Cryptography* (Wesson, Humphreys, e Evans 2014). Com tal avaliação

da segurança criptográfica do sistema ADS-B, esse é considerado, de acordo com Wesson et al., 2014, “inaceitavelmente incerto”, e qualquer tentativa de implementação de camadas de segurança esbarra em questões regulamentais e técnicas. Ainda assim, do ponto de vista da necessidade de inserção de criptografias para segurança do ADS-B, a mais viável seria a *Asymmetric-Key Cryptography*, mesmo que essa reduza a capacidade da frequência 1090 MHz e crie a necessidade da gestão das chaves, o que dificilmente seria aceite por agências reguladoras, companhias comerciais e entusiastas da aviação geral. No fim, o ADS-B continuará a depender das informações vindas do sistema RADAR para autenticação (Strohmeier et al. 2014; Abdulaziz et al. 2019; Wesson, Humphreys, e Evans 2014).

O advento do ADS-B facilitou a captação de dados para a criação de ferramentas capazes de apresentar as informações nos mais variados formatos. A Open Sky Network, sendo ela o segundo alicerce para a realização deste trabalho, é uma rede capaz de captar as mensagens ADS-B e apresentar as aeronaves em direto, além de criar bases de dados contendo informações como por exemplo o identificador do voo, aeronave, origem, destino, perfil de altitude e velocidade, trajetória (Open Sky Network 2022; Schäfer et al. 2014). Eventualmente, a quantidade de dados coletados pela OSN passou a ser utilizada como base para diferentes estudos relacionados à, por exemplo, eficiência operacional, pegada ecológica e emissões, estudo de trajetórias e até o impacto da pandemia da COVID-19 nos transportes aéreos (Perrichon, Gendre, e Klein 2022; Sun et al. 2022; Polishchuk, Lemetti, e Sáez 2019; Koelle e Barbosa 2021).

O primeiro estudo de caso a ser abordado foi apresentado no artigo “Evaluation of Flight Efficiency for Stockholm Arlanda Airport using Open Sky Network Data”, e trata da avaliação da eficiência do espaço aéreo do Aeroporto de Estocolmo à luz do consumo de combustível e perfil vertical de descida para aproximação, tendo em vista a operação do aeródromo e separação de tráfego (Polishchuk, Lemetti, e Sáez 2019; EASA 2022). Além disso, o estudo utiliza os dados da OSN bem como dados do EUROCONTROL *Demand Data Repository (DDR2)*, para também comparar a precisão e qualidade dos dados armazenados para as duas diferentes fontes. A realização dessa análise pressupõe a determinação e utilização de *Key Performance Indicators (KPIs)* para auxiliar na escolha das métricas a serem utilizadas. Dessa forma, foram utilizados o atraso médio, percentagem de voos atrasados, eficiência vertical de voo (VFE) em descida, tempo adicional em *Terminal Maneuvering Area (TMA)* e *Performance Indicator* à base de

combustível (Polishchuk, Lemetti, e Sáez 2019). Tanto os dados da OSN quanto os dados DDR2 fornecem perfis laterais e verticais, o que permite visualizar graficamente o VFE, como ilustra a Figura 8.

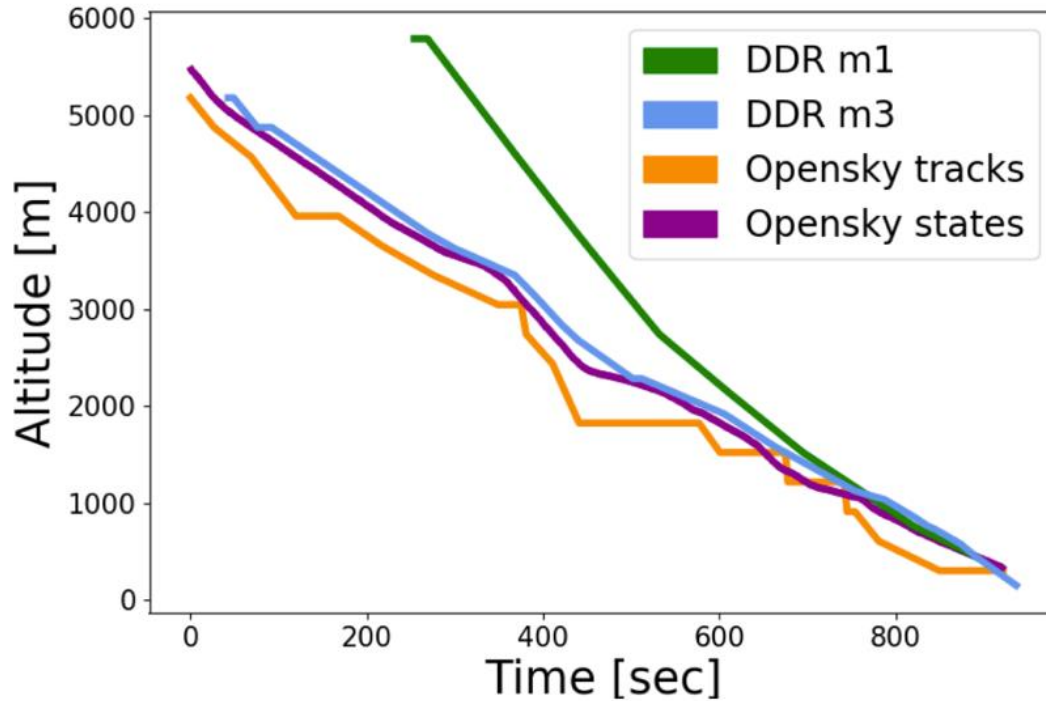


Figura 8 - VFE na TMA para o voo SAS410 com base nos dados da OSN e DDR2 (Polishchuk, Lemetti, e Sáez 2019).

A análise do consumo de combustível adicional, na Figura 10 é realizada a partir da diferença entre a trajetória de voo realizada e a trajetória de voo ideal para redução de consumo, com base no perfil lateral da aproximação na TMA, como mostra a Figura 9, e reflete um VFE baixo (Polishchuk, Lemetti, e Sáez 2019; Sun et al. 2022).

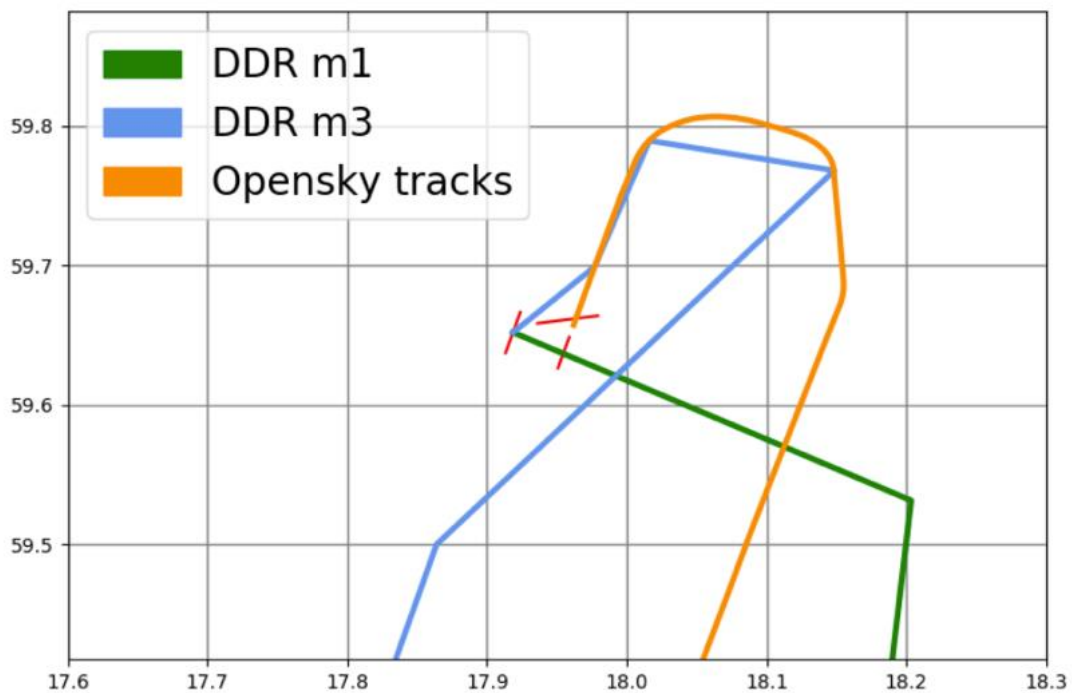


Figura 9 - Perfil Lateral na TMA para o voo SAS 964 com base nos dados da OSN e DDR2
(Polishchuk, Lemetti, e Sáez 2019).

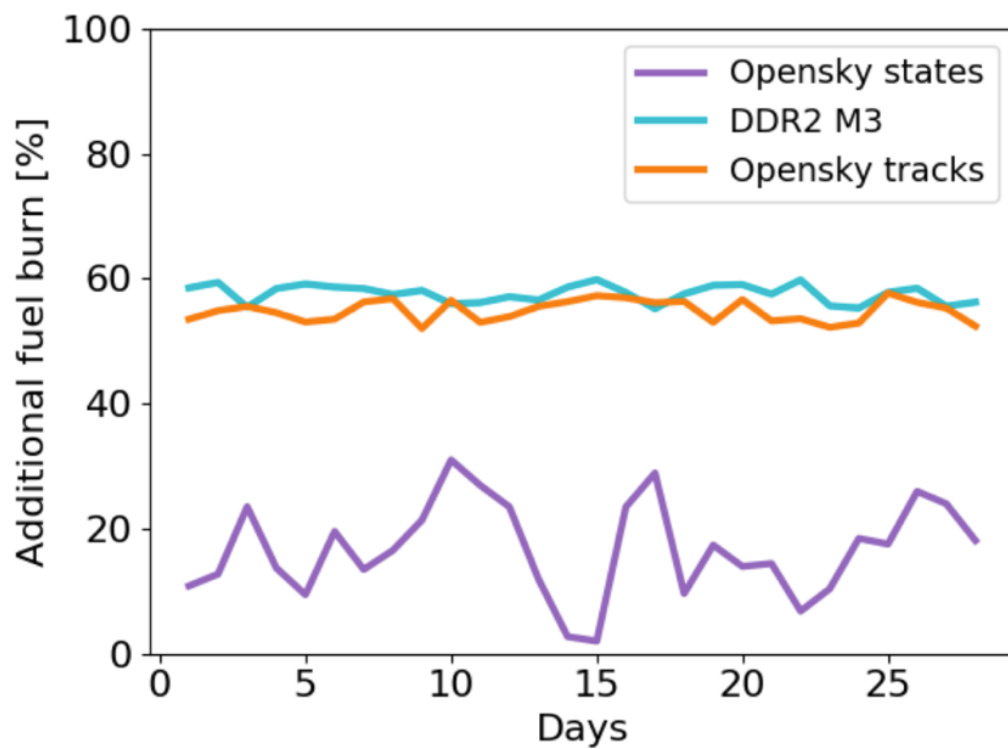


Figura 10 - Combustível consumido adicional por dia com base nos dados da OSN e DDR2
(Polishchuk, Lemetti, e Sáez 2019).

Portanto, a partir das Figuras 8, 9 e 10, acima, conclui-se que, apesar de possuir algumas lacunas nos dados, a OSN fornece uma precisão nos dados (registrados segundo a segundo) no âmbito vertical e lateral, além do grande volume de dados armazenados. Por isso, é uma fonte de dados recomendável para realização de análises como as apresentadas acima, mesmo sendo necessário recursos computacionais mais avançados (Polishchuk, Lemetti, e Sáez 2019; Open Sky Network 2022).

No ponto de vista das trajetórias de voo e a visualização dos dados da OSN como fonte de estudo, “A Geometric Approach to Study Aircraft Trajectories: The Benefits of OpenSky Network ADS-B Data” compara dados do EUROCONTROL e da OSN para comparar os dados e realizar análises. O enquadramento geométrico deste trabalho visa, primordialmente, comparar a diferença no formato das curvas e trajetórias entre as duas fontes de dados. Para isso, utiliza como referência o Aeroporto de Paris-Orly (LFPO) e o Aeroporto de Toulouse-Blagnac (LFBO). A Figura 11 e Figura 12 mostram, respectivamente, a primeira análise realizada, em LFBO, que utiliza o método de Interpolação Linear para traçar, a partir das variações de coordenadas e altitude, gráficos que permitem visualizar não só tais variações como também a diferença das fontes de dados (Perrichon, Gendre, e Klein 2022).

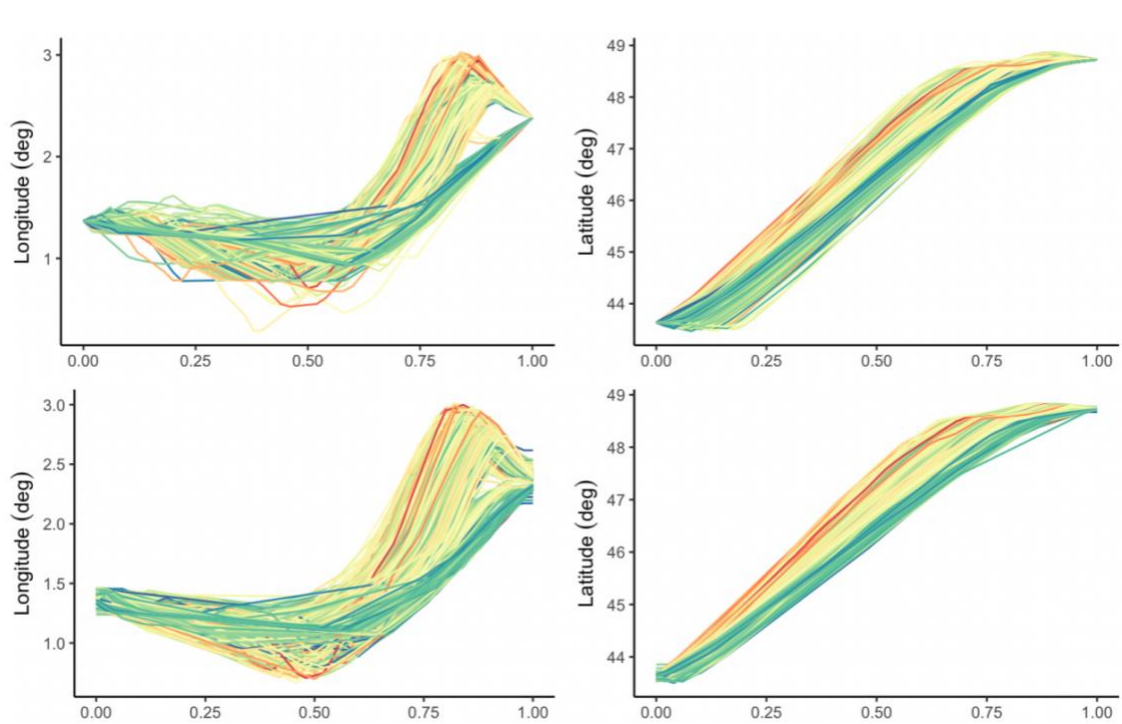


Figura 11 - Variação da Latitude e Longitude nas descolagens de LFBO. Dados do EUROCONTROL acima e da OSN abaixo (Perrichon, Gendre, e Klein 2022).

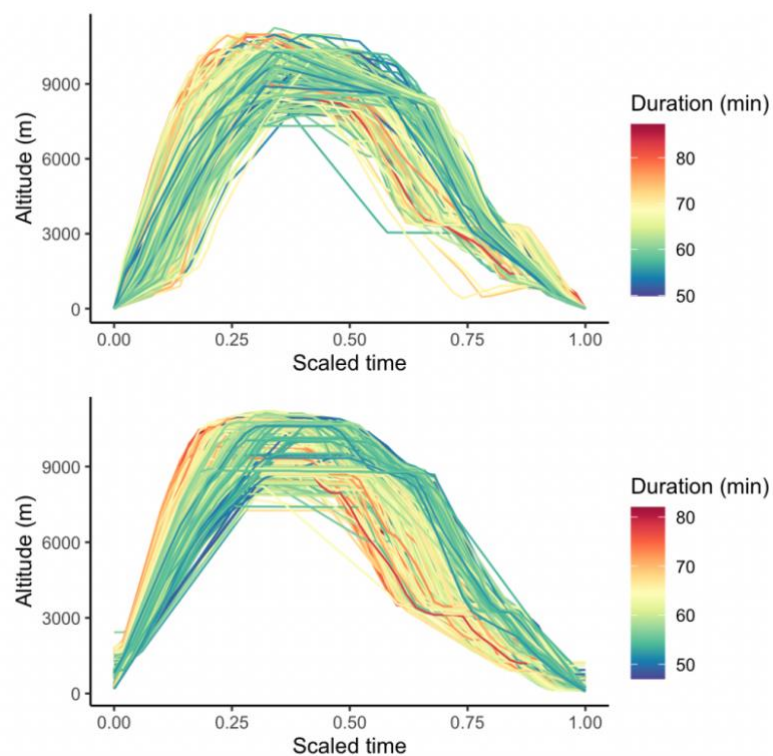


Figura 12 - Variação de Altitudes nas descolagens de LFBO. Dados do EUROCONTROL acima e OSN abaixo (Perrichon, Gendre, e Klein 2022).

Da mesma forma que foi feito para Toulouse, uma visualização das chegadas na TMA de LFPO também foi realizada a partir do perfil lateral das aeronaves, como mostram a Figura 13 e Figura 14, abaixo.

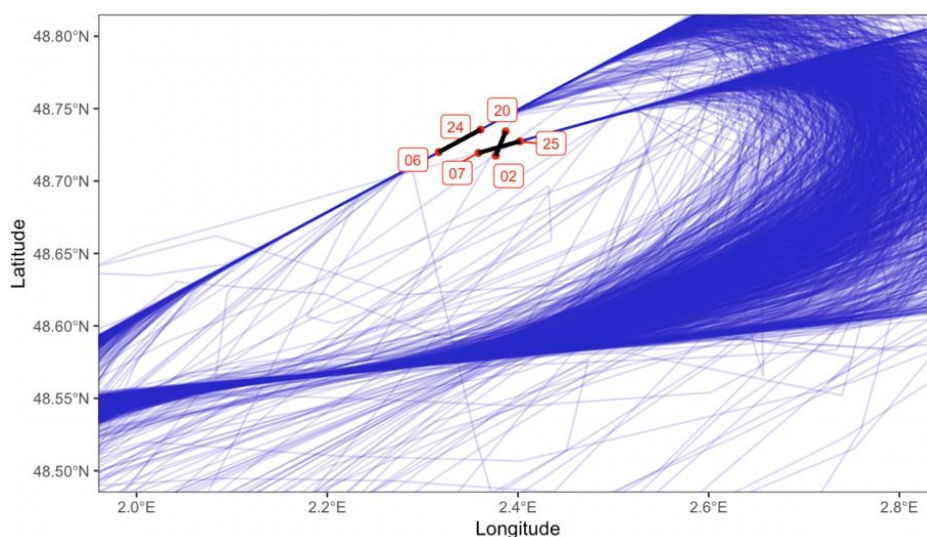


Figura 13 - Perfil Lateral de chegadas na TMA de LFPO. Dados do ADS-B (Perrichon, Gendre, e Klein 2022).

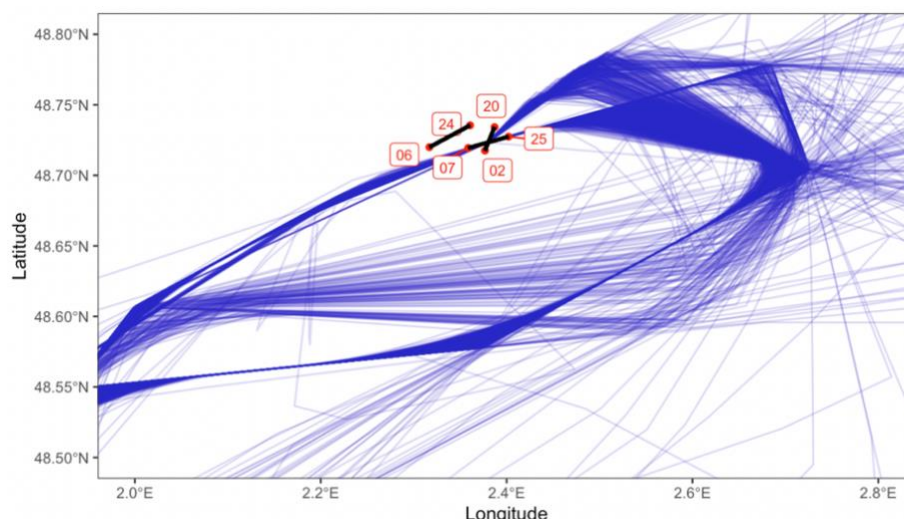


Figura 14 - Perfil Lateral de chegadas na TMA de LFPO. Dados do EUROCONTROL (Perrichon, Gendre, e Klein 2022).

É visível a diferença entre os dados do EUROCONTROL e OSN, ao passo que a fonte de dados da segunda – o ADS-B – registra informações num intervalo de tempo menor, por isso há maior precisão na informação apresentada (Perrichon, Gendre, e Klein 2022; Schäfer et al. 2018). Conclui-se, portanto, que o enquadramento geométrico para determinar não só a precisão dos dados, mas também retirar visualizações úteis como as vistas acima é de grande utilidade para compreensão do fluxo de tráfego aéreo, tendo aplicações em outras áreas como *Noise Abatement Procedures*, homologação e desenvolvimento de procedimentos de chegadas e saídas, otimização do consumo de combustível e dinamização do ATM, por exemplo.

No que diz respeito à relação entre a pandemia da COVID-19 e a análise de tráfego aéreo a partir da OSN, o *paper* “Assessing the Global COVID-19 Impact on Air Transport with Open Data” utiliza os dados retroativos coletados através do sistema ADS-B para avaliar o impacto que o período pandémico teve no fluxo de tráfego aéreo, como ilustra a Figura 15. Tal momento criou a necessidade de realizar estudos e buscar fontes capazes de determinar o real tamanho do problema, especialmente na aviação era imperativo a realização de estudos e análises acerca da capacidade de recuperação desta indústria (Koelle e Barbosa 2021; Strohmeier et al. 2021).

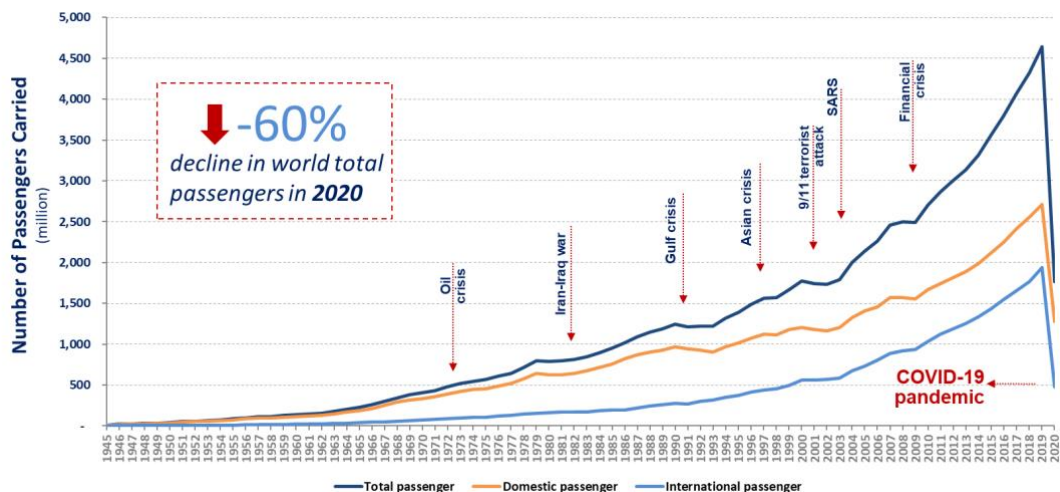


Figura 15 - Evolução de passageiro mundialmente entre 1945 e 2020 (ICAO 2021).

A partir deste cenário catastrófico, iniciativas como a Open Sky Network passaram a se popularizar não só porque forneciam dados em direto e utilizadores poderiam ver, naquele instante, a redução do tráfego aéreo, mas porque as mensagens do ADS-B eram armazenadas em *Data Sets*, que posteriormente serviram para determinar minuciosamente o processo de recuperação das operações em diferentes locais (Koelle e Barbosa 2021; Schäfer et al. 2014)

A capacidade da OSN de armazenamento mostrada permite ainda traçar linhas de investigação mais locais e relacionadas com indicadores e decisões sociopolíticas. Utilizando o caso do Brasil, por exemplo, que possui um total aproximado de 696.000 mortes por COVID-19 (World Health Organization 2022) e sempre esteve nos holofotes globais por conta da maneira como conduziu o período pandémico. É possível identificar que o aumento no tráfego no Brasil aumenta conforme as mortes aumentam, mas sofre forte redução de descolagens no início de 2021 por conta de um aumento abrupto na taxa de mortos, como mostra a Figura 16.

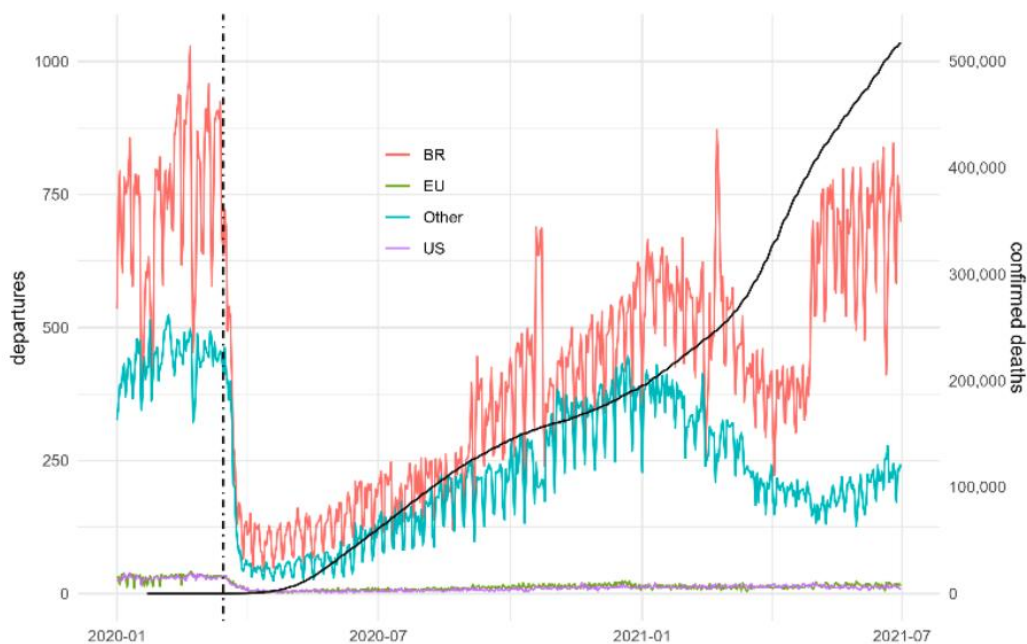


Figura 16 - Tráfego Regional e Mortes por COVID-19 no Brasil (Koelle e Barbosa 2021).

O terceiro alicerce fundamental que este trabalho se apoia sobre é a linguagem R, através do R Studio e livrarias como “rshiny”. A linguagem R possui grande relevância neste trabalho tendo em conta sua grande capacidade de gerir grandes quantidades de dados com alto volume de tráfego (Big Data) de uma maneira que as respostas dadas pelo programa (gráficos, tabelas, *dashboards*, por exemplo) são altamente personalizáveis e adaptáveis (Kumari e Verma 2019). Utilizando, portanto, o R Studio como meio para processar dados aeronáuticos da OSN, é imperativa a capacidade do programa de moldar e tratar as informações com o máximo de flexibilidade e interatividade.

Dentro do contexto da Open Sky Network e sua grande popularização após o período pandémico (em função da necessidade de estudar os comportamentos das operações aéreas pré, durante e pós COVID-19), diferentes autores da comunidade fizeram uso da programação aliada à bases de dados como a OSN para criar “pontes” – ou livrarias, denotação utilizada no R Studio – capazes de coletar os dados da rede e transformá-los no formato da linguagem R, permitindo que programadores façam uso desta ponte para moldar esses dados. Assim, autores como (Gasco 2020), desenvolveu e publicou a livreria “openskyr” que, através de 4 funções (*get_airport_data*, *get_flights_data*, *get_state_vectors* e *get_track_data*), busca na base de dados da OpenSky Network informações de determinado aeroporto, aeronave, trajetórias de voo, por exemplo, para um espaço de tempo. Outra livreria intitulada “openSkies”, de (Ayala et al. 2021), realiza

o mesmo processo, possuindo uma estrutura de funções como mostra a Figura 17. Além disso, a linguagem R vem sendo utilizada em diversos segmentos acadêmicos e de pesquisa, isso porque grande parte dessas contribuições se apoiam no conceito de *Reproducible Research* (Gandrud 2018). Nesse sentido, a estruturação das livrarias e deste trabalho se dão de forma que garantem a replicabilidade dos resultados, já que possui uma estrutura definida e permite a combinação de diferentes campos para a construção das mais variadas análises (Gandrud 2018; Kumari e Verma 2019).

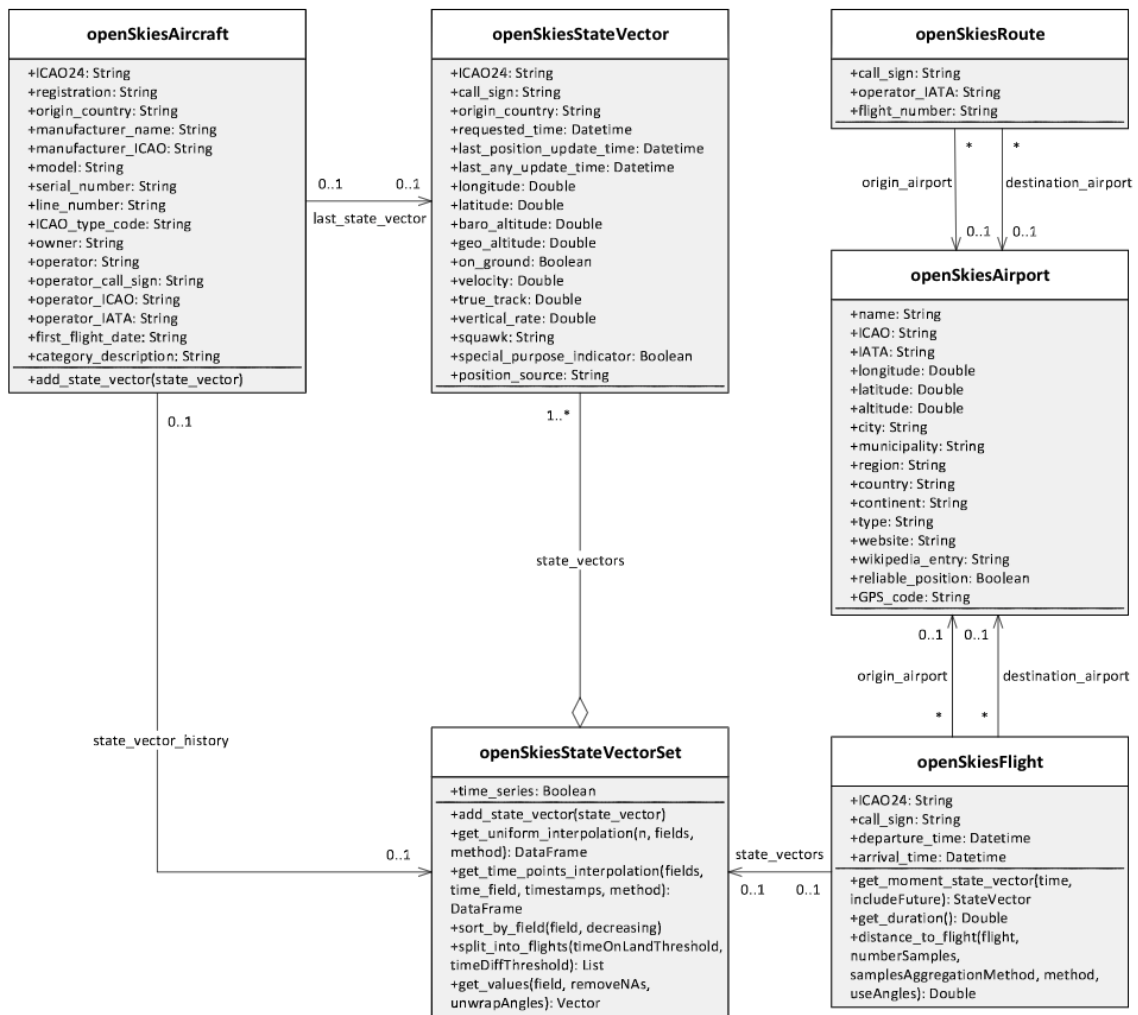


Figura 17 - Estrutura do package "openSkies" (Ayala et al. 2021)

Além dos trabalhos disponíveis e consolidados na comunidade científica enquanto fontes de informação e bases científicas para este projeto, é importante destacar a existência de projetos em desenvolvimento que possam vir a complementar e auxiliar o desenvolvimento deste. É o caso da tese “Projeto de uma estação ADS-B na rede

OpenSky Network”, que consiste no desenvolvimento e alimentação de uma base de dados a partir da instalação de um sensor capaz de captar as mensagens ADS-B transmitidas pelas aeronaves (Santos 2023). Teria a finalidade similar à OpenSky Network, por exemplo, porém o desenvolvimento deste projeto – que decorre também dentro do ISEC Lisboa – permite uma maior adaptabilidade e facilidade não só na obtenção dos dados, mas também na estruturação e apresentação desses, para futura utilização não só neste projeto, bem como outros dentro e para além da instituição.

Concluindo, este trabalho se apoia em três pilares principais: Automatic Dependent Surveillance Broadcast (ADS-B), Open Sky Network e a linguagem R. O primeiro é um sistema inovador que surgiu como o substituto do RADAR, mas que, por questões de segurança, acabou por se aliar ao suposto substituído e atualmente o ATM conta com a redundância de todas as informações, reduzindo assim o erro. Apesar da dificuldade de aplicação nas operações aéreas pela “facilidade” para sofrer ciberataques, é um sistema que alimenta o universo acadêmico com informações valiosas e que permitiu o surgimento de iniciativas como o segundo pilar deste projeto. A Open Sky Network é uma grande fonte e um grande objeto de estudo no mundo acadêmico por conta de sua capacidade de fornecer dados em direto bem como armazená-los e gerar um *Data Set* que permite a realização de diferentes pesquisas acerca do comportamento do tráfego aéreo. Por fim, a linguagem R funciona como o fio condutor de tudo, ao passo que tem a capacidade de receber, tratar, moldar e apresentar os dados que vão do ADS-B para a Open Sky Network.

3. Desenvolvimento do Tema

O terceiro capítulo deste relatório visa descrever os processos que culminaram no cumprimento dos objetivos apresentados inicialmente, apoiando-se nos conceitos e estudos discutidos. A descrição da realização desse trabalho será feita a partir das esferas principais de *Big Data*: *Data Engineering*, *Data Science* e *Data Visualization*, tendo em vista que todos os processos, passos e procedimentos que ocorreram atendem à um desses nichos, como esquematizado na Figura 18.



Figura 18 - Camadas dos processos de Big Data.

Em primeiro lugar, no campo da *Data Engineering* (Engenharia de Dados), foram feitas uma série de escolhas com o intuito de delimitar as capacidades da ferramenta a ser criada com base nos objetivos estipulados. Determinou-se o Aeroporto Humberto Delgado (LPPT) como o foco do estudo, tendo em vista que está imerso na realidade dos estudantes e docentes do ISEC Lisboa, além de se tratar de um Aeroporto que recebe uma grande quantidade de tráfego na escala europeia e mundial. Além do aeroporto, determinou-se o ano de 2022 a ser analisado, tendo em vista que é o ano completo mais recente e que não possui grande contaminação do tráfego em função da pandemia da COVID-19, o que garante maior fiabilidade às análises realizadas.

Delimitadas as características temporais e geográficas, dentro do R Studio, era imperativo realizar a escolha de uma livreria capaz de dar ao programa as habilidades necessárias para ter acesso e manipular os dados armazenados pela OpenSky Network. Nesse sentido, foi escolhida a livreria “*openskyr*”, tendo em vista a sua capacidade de obter os dados da OpenSky Network através de quatro funções diferentes: “*get_airport_data*”, “*get_flights_data*”, “*get_state_vectors*” e “*get_track_data*”. No âmbito deste trabalho, foi utilizada a primeira, *get_airport_data* e a última, *get_track_data*. A primeira liga-se à base de dados da OSN, gera *dataframes* (tabelas) para chegadas ou saídas de determinado aeroporto e determinado período temporal, conforme a Figura 19 apresenta.

```
dataframe_departures <- get_airport_data(username="your_username",password="your_password",  
                                         option="departures",airport="LPPT",  
                                         begin=1517227200,end=1517230800)
```

Figura 19 - Código que busca as saídas de LPPT.

A função “*get_airport_data*” associada aos parâmetros mostrados na Figura 19, gera um dataframe que contempla as saídas (“*departures*”) para o Aeroporto de Lisboa (“LPPT”) entre 1517227200 (29 Janeiro 2018 12:00:00) e 1517230800 (29 Janeiro 2018 13:00:00).

A Tabela 2 mostra o *dataframe* gerado no R Studio. Chamaremos a estrutura dessa tabela como Estrutura Inicial, na medida em que possui colunas que não têm utilidade no que tocam os objetivos deste trabalho, além ter em falta colunas necessárias, que serão futuramente adicionadas através de outros ficheiros de suporte. Mesmo se tratando de uma estrutura primária de apresentação dos dados, já é possível identificar o tipo de

informação e as capacidades que a livraria e a OpenSky possuem, como mostra a Tabela 3.

Tabela 2 - Estrutura Inicial da função "get_airport_data"

icao24	firstSeen	estDepartureAirport	lastSeen	estArrivalAirport	callsign
3423c4	151723057 3	LPPT	151723402 2	LEMD	IBE31ZD
4ca5dc	151723013 1	LPPT	151723120 8	NULL	RYR37ZK
491144	151722985 3	LPPT	151723154 0	NULL	TAP1140
4ca7c4	151722970 3	LPPT	151723976 6	EDDF	RYR7DE
495291	151722957 1	LPPT	151723079 7	NULL	TAP1689
4ca8ea	151722939 2	LPPT	151723773 6	EBCI	RYR3JB
EstDeparture AirportHorizDistance		EstDeparture AirportVertDistance		estArrivalAirportHoriz Distance	
6185		632		502	
6951		792		NULL	
7014		808		NULL	
6780		587		NULL	
7112		762		NULL	
6678		678		10599	
EstArrivalAirportDist ance		departureAirportCandidatesC ount		arrivalAirportCandidatesC ount	
182		2		3	
NULL		1		0	
NULL		1		0	
110		1		2	
NULL		1		0	
399		1		2	

Tabela 3 - Estrutura Inicial do API da OpenSky Network (OpenSky Network 2021).

Property	Type	Description
<i>icao24</i>	string	Unique ICAO 24-bit address of the transponder in hex string representation. All letters are lower case.
<i>firstSeen</i>	integer	Estimated time of departure for the flight as Unix time (seconds since epoch).
<i>estDepartureAirport</i>	string	ICAO code of the estimated departure airport. Can be null if the airport could not be identified.
<i>lastSeen</i>	integer	Estimated time of arrival for the flight as Unix time (seconds since epoch)
<i>estArrivalAirport</i>	string	ICAO code of the estimated arrival airport. Can be null if the airport could not be identified.
<i>callsign</i>	string	Callsign of the vehicle (8 chars). Can be null if no callsign has been received. If the vehicle transmits multiple callsigns during the flight, we take the one seen most frequently
<i>estDepartureAirportHorizDistance</i>	integer	Horizontal distance of the last received airborne position to the estimated departure airport in meters
<i>estDepartureAirportVertDistance</i>	integer	Vertical distance of the last received airborne position to the estimated departure airport in meters
<i>estArrivalAirportHorizDistance</i>	integer	Horizontal distance of the last received airborne position to the estimated arrival airport in meters
<i>estArrivalAirportVertDistance</i>	integer	Vertical distance of the last received airborne position to the estimated arrival airport in meters
<i>departureAirportCandidatesCount</i>	integer	Number of other possible departure airports. These are airports in short distance to estDepartureAirport.
<i>arrivalAirportCandidatesCount</i>	integer	Number of other possible departure airports. These are airports in short distance to estArrivalAirport.

A partir deste cenário inicial, ainda no contexto de *Data Engineering*, estipulou-se que seriam feitos dois dataframes, um para as saídas (“*departures*”), nomeado LPPT_D, e outro para as chegadas (“*arrivals*”), nomeado LPPT_A. Mesmo tendo o volume de

tráfego dividido em dois, as análises e, conseqüentemente, a apresentação dos dados, será feita de forma espelhada. No ambiente da linguagem R, era necessário se adaptar à uma limitação da função “*get_airport_data*”, na medida em que essa só era capaz de gerar um *dataframe* com os dados de um dia completo. Logo, não conseguiríamos, de imediato, colocar o instante inicial como 01/01/22 à 00:00:00 e o instante final como 31/12/22 às 23:59:59. Nesse cenário, foi realizada uma iteração, mecanismo que cria um código em forma de ciclo, que se repete dia a dia. Ou seja, o ciclo teria início no dia 01/01/22 à 00:00:00 e fim no dia 01/01/22 às 23:59:59 e buscaria somente dados do dia 1º de Janeiro de 2022. Completado o primeiro dia, faria o mesmo para o dia 02/01/2022 e assim sucessivamente, até chegar ao último dia, onde buscaria dados entre 31/12/2022 à 00:00:00 até 31/12/22 às 23:59:59. A realização deste movimento permitiu a criação dos *dataframes* LPPT_A e LPPT_D, ainda na Estrutura Inicial.

Finalizada a criação dos dois *dataframes*, que esboçavam o esqueleto do projeto no que toca a análise de volume de tráfego, era necessário manipular a Estrutura Inicial de forma que atendesse aos requisitos da linguagem R: remoção de colunas, mudança do tipo de dados de certas colunas, criação de novas colunas e, por fim, a adição de ficheiros externos que dão suporte à informação que existe inicialmente.

Foram removidas as colunas “*estDepartureAirportHorizDistance*”, “*estDepartureAirportVertDistance*”, “*estArrivalAirportHorizDistance*”, “*estArrivalAirportVertDistance*”, “*departureAirportCandidatesCount*” e “*arrivalAirportCandidatesCount*”, retirando informações que não seriam utilizadas para análise. Após essa remoção, foi realizada uma mudança no tipo de dados às colunas “*firstSeen*” e “*lastSeen*”: como a unidade de tempo emitida pela OpenSky Network é *Epoch Time*, era preciso trocar o tipo de dados para o formato “*data*”. Essa mudança permitia, posteriormente, através de uma função de separação de caracteres, a criação de colunas que continham apenas o dia, mês, ano, hora, período do dia, data e hora completa de cada voo registado. A partir desse movimento, obtém-se diferentes granularidades e escalas temporais para serem exploradas.

Após a formatação dos dados advindos da Estrutura Inicial, era importante fornecer aos dois *dataframes* uma maior qualidade e precisão às informações que já existiam. Por exemplo, a coluna “*icao24*” indicava a aeronave que operou o voo através do código de transponder – único para cada aeronave – porém, não era possível saber o tipo de aeronave operada (B737, A320, E195, por exemplo). O mesmo vale para a coluna “*callsign*” e a

coluna “*estDepartureAirport*” (ou “*estArrivalAirport*”, a depender do *dataframe*), ou seja, a informação disponibilizada precisa de um suporte externo para ter maior valor.

Por isso, foram buscados ficheiros externos ao R e à livreria “*openskyr*” para que seja possível relacionar um código de transponder (‘*icao24*’), um ‘*callsign*’ ou um código de aeródromo, por exemplo, à uma informação menos técnica e mais acessível, ou seja, o tipo de aeronave, a companhia que a opera e a localização do aeroporto. O primeiro ficheiro adicionado, intitulado “*Airline Codes*”, no formato .xlsx (Excel), possuía o código de uma companhia (“*Airline*”), associado ao nome completo (“*Airline Name*”) e o País de Registo (|| TAP Air Portugal || TAP || Portugal ||, por exemplo), como mostra a Tabela 4. No ambiente do R, era possível associar à coluna ‘*callsign*’ dos *dataframes* LPPT_A e LPPT_D à coluna ‘*Airline*’ do ficheiro externo, através da função “*merge*”, que utiliza termos em comum de dados diferentes para juntá-los. Assim, os *dataframes* de chegadas (LPPT_A) e saídas (LPPT_D) passaram a apresentar não só o código de três letras da companhia, mas também seu respetivo nome, por extenso.

Tabela 4 – Dados do ficheiro "Airline Codes"

Airline Name	Airline	Country
TAM Linhas Aéreas	TAM	Brazil
TAME – Linea Aérea del Ecuador	TAE	Ecuador
TAP Air Portugal	TAP	Portugal
TAROM	ROT	Romania
Tassili Airlines	DTH	Algeria
Thai Airways International	THA	Thailand

Além de “*Airline Codes*”, foi também incorporado o ficheiro “*Aircraft Database*”, também no formato .xlsx, que possui dados como o código “*icao24*”, “*Aircraft Type*” e “*Wake Turbulence Category*” (|| 3c70aa || A332 || Heavy ||, por exemplo), como mostra a Tabela 5. A partir da associação utilizando o termo comum “*icao24*”, é possível fornecer aos *dataframes* LPPT_A e LPPT_D a informação do tipo de aeronave, além da *Wake Turbulence Category*, um parâmetro de extrema importância no âmbito da separação de tráfego dentro do Air Traffic Management.

Tabela 5 – Dados do ficheiro "Aircraft Database"

icao24	Aircraft Type	Wake Turbulence Category
3c7021	A332	Heavy
3c70aa	A332	Heavy
3c71a4	A332	Heavy
3c7201	A332	Heavy

Por último, o ficheiro "Airport Codes", também no formato Excel foi adicionado de forma que associasse a coluna "ICAO Departure" (ou "ICAO Arrival", a depender do *dataframe*) ao país onde se encontra o aeroporto através do nome ("Destination Country" ou "Origin Country") e a latitude e longitude do aeroporto, para identificação geográfica. Tal estrutura é representada pela Tabela 6.

Tabela 6 – Dados do ficheiro "Airport Codes"

ICAO Airport	Country	Longitude	Latitude
LPPD	Portugal	-25.697901	37.74120
LPPI	Portugal	-28.441299	38.55430
LPPM	Portugal	-8.583960	37.14930
LPPR	Portugal	-8.681390	41.24810
LPPS	Portugal	-16.350000	33.07340
LPPT	Portugal	-9.135920	38.78130

A partir de todos os processos descritos que modificaram a Estrutura Inicial, obteve-se os *dataframes* finais LPPT_A e LPPT_D, com 21 colunas, como mostra a Figura 20, e a Tabela 7 mostra o resultado do *dataframe* LPPT_A, muito similar ao LPPT_D.



▶ LPPT_A	83543 obs. of 21 variables	
▶ LPPT_D	98721 obs. of 21 variables	

Figura 20 - Indicação dos *dataframes* LPPT_A e LPPT_D.

Tabela 7 - Estrutura Final para o dataframe LPPT_A

ICAO Departure		ICAO Arrival	Origin Country		Longitude	Latitude
EIDW		LPPT	Ireland		-6.270070	53.42130
EKCH		LPPT	Denmark		12.656000	55.61790
ELLX		LPPT	Luxembourg		6.204444	49.62333
ELLX		LPPT	Luxembourg		6.204444	49.62333
ELLX		LPPT	Luxembourg		6.204444	49.62333
ENGM		LPPT	Norway		11.100400	60.19390
callsign		Airline	Airline Name		Country	icao24
EIN4HR		EIN	Aer Lingus		Ireland	4ca5c8
TAP753A		TAP	TAP Air Portugal		Portugal	4952ca
LGL3759		LGL	Luxair		Luxembourg	4d00f1
EJU14FN		EJU	EasyJet Europe		Austria	440062
TAP2872		TAP	TAP Air Portugal		Portugal	495149
TAP761		TAP	TAP Air Portugal		Portugal	4952c3
Aircraft Type		WTC	Date Hour		Year	Month
A320		Medium	2022-01-02 18:03:29		2022	January
A20N		Medium	2022-01-02 08:44:15		2022	January
B737		Medium	2022-01-02 08:04:16		2022	January
A320		Medium	2022-01-02 22:40:33		2022	January
A21N		Medium	2022-01-02 15:42:13		2022	January
A20N		Medium	2022-01-02 22:35:34		2022	January
Day	Hour	Date	Time	Period of the Day		Weekday
02	18	2022-01-02	18:03:29	Afternoon		Sunday
02	8	2022-01-02	08:44:15	Morning		Sunday
02	8	2022-01-02	08:04:16	Morning		Sunday
02	22	2022-01-02	22:40:33	Night		Sunday
02	15	2022-01-02	15:42:13	Afternoon		Sunday
02	22	2022-01-02	22:35:34	Night		Sunday

Por fim, a Tabela 8 ilustra o tipo de dados para cada coluna, tanto do *dataframe* LPPT_D e LPPT_A.

Tabela 8 – Tipos de Dados para a Estrutura Final dos dataframes LPPT_D e LPPT_A

Coluna	Tipo	Descrição
'ICAO Arrival'	string	Aeródromo de chegada, no formato ICAO.
'ICAO Departure'	string	Aeródromo de partida, no formato ICAO.
'Destination/Origin Country'	string	País de origem ou destino do voo (LPPT_D e LPPT_A, respectivamente).
Longitude	numeric	Longitude do aeroporto de destino ou de origem (LPPT_D e LPPT_A, respectivamente).
Latitude	numeric	Latitude do aeroporto de destino ou de origem (LPPT_D e LPPT_A, respectivamente).
callsign	string	Identificador de voo emitido pela aeronave.
Airline	string	Primeiros três caracteres do <i>callsign</i> , que indicam a companhia aérea.
'Airline Name'	string	Nome completo da Companhia Aérea.
Country	string	País base da Companhia Aérea.
icao24	string	Identificador de transponder da aeronave.
'Aircraft Type'	string	Identificador do tipo de aeronave.
'Wake Turbulence Category'	string	Categoria de Esteira de Turbulência.
'Date Hour'	date	Data (AAAA/MM/DD) e Hora (HH:MM:SS) da partida ou saída do voo (LPPT_D e LPPT_A, respectivamente).
Year	string	Ano (AAAA), obtido a partir do campo 'Date Hour'.
Month	string	Mês, por extenso, obtido a partir do campo 'Date Hour'.
Day	string	Dia, obtido a partir do campo 'Date Hour'.
Hour	string	Hora do dia, obtido a partir do campo 'Date Hour'.
Date	date	Data (AAAA/MM/DD), obtido a partir do campo 'Date Hour'.
Time	string	Hora completa (HH:MM:SS), obtido a partir do campo 'Date Hour'.
'Period of the Day'	string	Período do dia (Dawn, Morning, Afternoon, Night).
Weekday	string	Dia da semana, por extenso.

Já a segunda função da livreria “openskyr”, *get_track_data*, obtém a trajetória de determinada aeronave a cumprir determinado voo a partir das mensagens ADS-B transmitidas pela aeronave. No ambiente do R, ilustra-se conforme ilustra a Figura 21. No exemplo, a função vai buscar dados de trajetória à aeronave identificada com o código “icao24” 494103 (um Embraer Phenom 300 de matrícula CS-PHC) no momento “1587126600” (17 de Abril de 2020, às 12:30:00). O retorno da função é um dataframe como ilustrado pela Tabela 9.

```
dataframe_track <- get_track_data(username=username,password="your_password",
                                  icao24="494103",time=1587126600)
```

Figura 21 - Código que busca a trajetória da aeronave CS-PHC.

Tabela 9 - Estrutura da função "get_track_data"

icao24	callsign	time	Latitude	Longitude	baro_altitude	true_track	on_ground
494103	NJE2SD	1587124694	46.9105	7.5017	304	319	FALSE
494103	NJE2SD	1587124708	46.9173	7.4934	609	321	FALSE
494103	NJE2SD	1587124709	46.9178	7.4928	609	320	FALSE
494103	NJE2SD	1587124725	46.926	7.4825	914	318	FALSE
494103	NJE2SD	1587124742	46.9351	7.4711	1219	320	FALSE
494103	NJE2SD	1587124759	46.9449	7.4596	1524	321	FALSE

É possível perceber que o retorno da função reúne os dados posicionais da aeronave em 4 componentes: latitude, longitude, altitude barométrica e trajetória. Ou seja, *get_track_data* é capaz de compilar, para cada linha, a posição, altitude e direção de uma mesma aeronave e, como as mensagens transmitidas pelo sistema ADS-B tem alta frequência, é possível traçar uma rota tanto no perfil lateral de deslocamentos como no horizontal quase que fidedigna, tendo em vista os intervalos de tempo mostrados na Tabela 9. A Tabela 10, por sua vez, identifica o tipo de dado e sua descrição para o dataframe gerado pela função *get_track_data*. A utilização desta função em larga escala, ou seja, aplicada automaticamente à uma série de aeronaves, permite a criação de uma base de dados capaz de localizar, através de latitude e longitude, a trajetória de muitas aeronaves em simultâneo.

Tabela 10 – Tipos de Dados para a Estrutura da função "get_track_data"

Coluna	Tipo	Descrição
'icao24'	string	Identificador de transponder da aeronave.
'Callsign'	string	Identificador de voo emitido pela aeronave.
'time'	string	Data e Hora, em Epoch Time, da transmissão da mensagem.
'longitude'	string	Longitude da aeronave no momento da transmissão.
'latitude'	string	Latitude da aeronave no momento da transmissão.
'baro_altitude'	string	Altitude barométrica da aeronave.
'true_track'	string	Direção da aeronave no momento da transmissão.
'on_ground'	string	Se o sistema identifica se a aeronave está em solo.

Tendo em mente a Estrutura acima, foi criado o *dataframe* LPPT_TRK_FLIGHTS, derivado do *dataframe* LPPT_A, para separar todas as aeronaves que serão sujeitas à função *get_track_data* para obtenção de trajetórias para as chegadas no Aeroporto de Lisboa. Essa filtragem também se deu por conta de limitações na função, que não é capaz de ligar-se à API da OpenSky Network para todas as aeronaves contidas no *dataframe* LPPT_A, que possui 83543 registos. Assim, foram filtrados somente os voos que chegaram ao Aeroporto de Lisboa no dia 15 de Setembro, durante a tarde, isto é, entre 12:00:00 e 19:00:00. O *dataframe* gerado possuía vinte e duas entradas, ou seja, vinte e dois voos aterraram em LPPT durante à tarde. Depois, tal *dataframe* fora submetido à função *get_track_data* através de uma iteração, ou seja, a função repete-se para cada identificador 'icao24' que apareça no *dataframe* "LPPT_TRK_FLIGHTS". O resultado desse ciclo gerou o *dataframe* "LPPT_TRK", contendo as trajetórias de cada aeronave contida na tabela inicial. Ou seja, para cada um dos vinte e dois voos obtidos pela função *get_track_data*, o novo *dataframe* criado continha toda a trajetória do primeiro desses

voos, seguido de toda a trajetória do segundo, e assim por diante. Por fim, “LPPT_TRK” seria subdividido entre “LPPT_TRK_VER”, que contém as informações de altitude dos voos obtidos (perfil vertical) e “LPPT_TRK_LAT”, contendo os dados de latitude e longitude dos voos (perfil lateral)

Tendo em conta uma fonte de dados estabelecida e capaz de prover informações e análises capazes de atender aos objetivos do trabalho, parte-se para o segundo plano do desenvolvimento, a Ciência de Dados (*Data Science*). No que toca esse campo de *Big Data*, no enquadramento da análise de volumes, parte-se das Estruturas dos *dataframes* “LPPT_A” e “LPPT_D” e “LPPT_TRK” para realizar aferições, ou seja, transformar uma grande quantidade de valores em informação que tenha importância dentro do âmbito do trabalho.

Na Tabela 8, é possível dividir as colunas de dados em quatro classificações: geográficas, da companhia, da aeronave e temporais. As colunas geográficas, nomeadamente ‘*ICAO Arrival*’, ‘*ICAO Departure*’, ‘*Destination/Origin Country*’, ‘*Longitude*’ e ‘*Latitude*’, localizam os voos que chegam ou que saem do Aeroporto Humberto Delgado a partir das coordenadas. A capacidade de localização de cada aeroporto de origem ou destino permite uma visualização mais direta da quantidade de voos que saem ou chegam ao Aeroporto de Lisboa, em um mapa, por exemplo.

As colunas que dão valores relativamente à companhia aérea, que são ‘*callsign*’, ‘*Airline*’ e ‘*Airline Name*’, permitem agregar os voos por companhia aérea. Tal agrupamento garante maior compreensão do comportamento das companhias em função das métricas temporais e geográficas (destino mais visitado, período do dia mais presente no Aeroporto, por exemplo).

As colunas relativas às aeronaves – ‘*icao24*’, ‘*Aircraft Type*’ e ‘*Wake Turbulence Category*’ – são capazes de agrupar as chegadas e saídas por tipo de aeronave e pela categoria da Esteira de Turbulência. Compreender o padrão de chegadas e saídas a partir das aeronaves tem um grande valor associado, na medida em que parte das operações dentro de um Aeroporto são limitadas a partir da Categoria da Esteira de Turbulência (*Wake Turbulence Category*), ao passo em que delimitam a separação entre as aeronaves, tanto na descolagem quanto na aterragem.

Por fim, os parâmetros temporais, nomeadamente ‘*Date Hour*’, ‘*Year*’, ‘*Month*’, ‘*Day*’, ‘*Hour*’, ‘*Date*’, ‘*Time*’, ‘*Period of the Day*’, ‘*Weekday*’, nos permitem analisar

padrões de tráfego em função das diferentes escalas de tempo obtidas a partir das granularidades apresentadas pela API da OpenSky Network.

Já num enquadramento das análises de rotas, o *dataframe* LPPT_TRK nos permite gerar uma visualização geográfica das trajetórias das aeronaves que chegam à Lisboa, tanto pela subdivisão “LPPT_TRK_LAT” quanto por “LPPT_TRK_VER”. A primeira forma de visualização nos permite compreender padrões de tráfego, esperas, encurtamentos e alongamentos de aproximação, ou seja, procedimentos normais numa TMA de grande tráfego. A segunda, por sua vez, permite observar perfis mais ou menos acentuados de descida para aproximação, além de possíveis procedimentos de espera na diminuição da altitude. Apesar de se tratar de uma amostra reduzida de voos, essa nos permite observar com maior detalhe alguns eventos de voo como os descritos acima.

Assim, é possível identificar as capacidades que os dados fornecidos pela OpenSky Network possuem no âmbito deste projeto. Tal identificação permite gerar formas de visualização desses dados em diferentes enquadramentos, porém com um aspeto condizente com suas propriedades. Nesse contexto se fazem valer os conceitos de Visualização de Dados (*Data Visualization*), isto é, o conjunto de processos que culmina numa visualização dos dados de maneira que transmita uma mensagem, mensagem esta que não possuiria o mesmo efeito se apresentada através de grandes bases de dados.

Tendo em mente a construção do *dashboard* como objetivo deste trabalho, era necessário construir, através da linguagem R, todas as visualizações que seriam apresentadas neste ambiente. Os *dataframes* LPPT_A e LPPT_D retornariam gráficos, tabelas e mapas relativamente ao volume de voos em função dos diferentes parâmetros disponíveis (temporais e geográficos). Já o *dataframe* LPPT_TRK retornaria uma visualização das chegadas no Aeroporto de Lisboa, num enquadramento lateral (rota) e vertical (perfil de descida).

Nesse sentido, partindo de LPPT_A e LPPT_D, foram utilizados os parâmetros temporais para construção de gráficos que permitissem representar uma granularidade temporal, ou seja, o volume de chegadas e saídas, porém em diferentes referências de tempo. Utilizando a livreria “ggplot2”, tal visualização se deu através de um gráfico de barras que, além da representação em função do tempo, também foi apresentada, para cada coluna, a divisão de tal volume em *Wake Turbulence Category*. Assim, era possível observar não só os padrões de tráfego em função do volume, mas também foi possível

qualificar a informação, provendo um outro eixo de visualização que, como já abordado, possui alto valor no que toca a gestão do tráfego aéreo em zonas terminais. A Figura 22 representa o gráfico de volume de chegadas no Aeroporto de Lisboa em função das horas do dia e com subdivisão dos volumes por WTC, exemplificando o processo descrito acima da construção de tais visualizações, que variam em função da granularidade temporal (Mês, Dia da Semana, Período do Dia e Hora do Dia). A granularidade temporal permitiu, então, a criação de quatro gráficos para as chegadas, bem como quatro gráficos para as saídas.

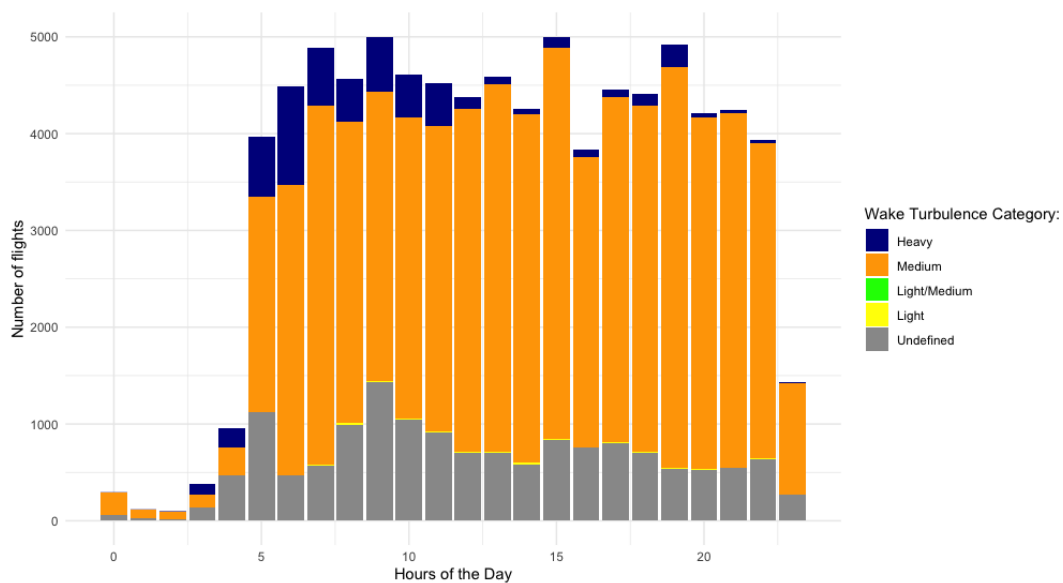


Figura 22 - Gráfico de volume de chegadas por hora do dia.

Além da visualização dos volumes através dos gráficos, a partir dos *dataframes* LPPT_A e LPPT_D, também era possível posicionar geograficamente a origem ou destino dos voos que passavam pelo Aeroporto de Lisboa. Para isso, foi escolhido a visualização através mapas, tanto para aeroportos quanto para países. Utilizando livrarias disponíveis no R como “ggplot2” e “rnatualearth”, a análise dos volumes por Aeroportos fora feita para Europa e Norte de África, enquanto o enquadramento por países fora feito também para a região apresentada, mas também numa escala mundial. Nesse sentido, foram desenvolvidos os mapas, tanto para saídas, ou seja, o volume de tráfego que recebe os voos que saem do Aeroporto de Lisboa, quanto para as chegadas, isto é, o volume de tráfego que manda voos a Lisboa, ambos posicionados geograficamente por países ou por

aeroportos. A Figura 23, Figura 24 e Figura 25 ilustram o resultado dos procedimentos apresentados acima.

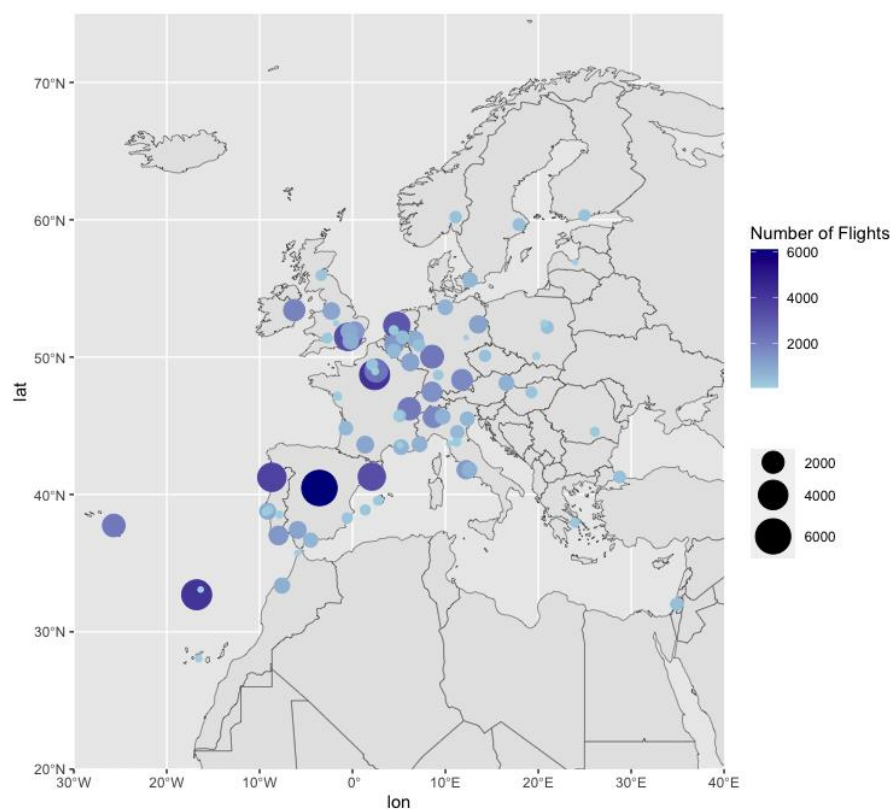


Figura 23 - Mapa de volume de Destinos (saídas) para Europa e Norte de Africa.

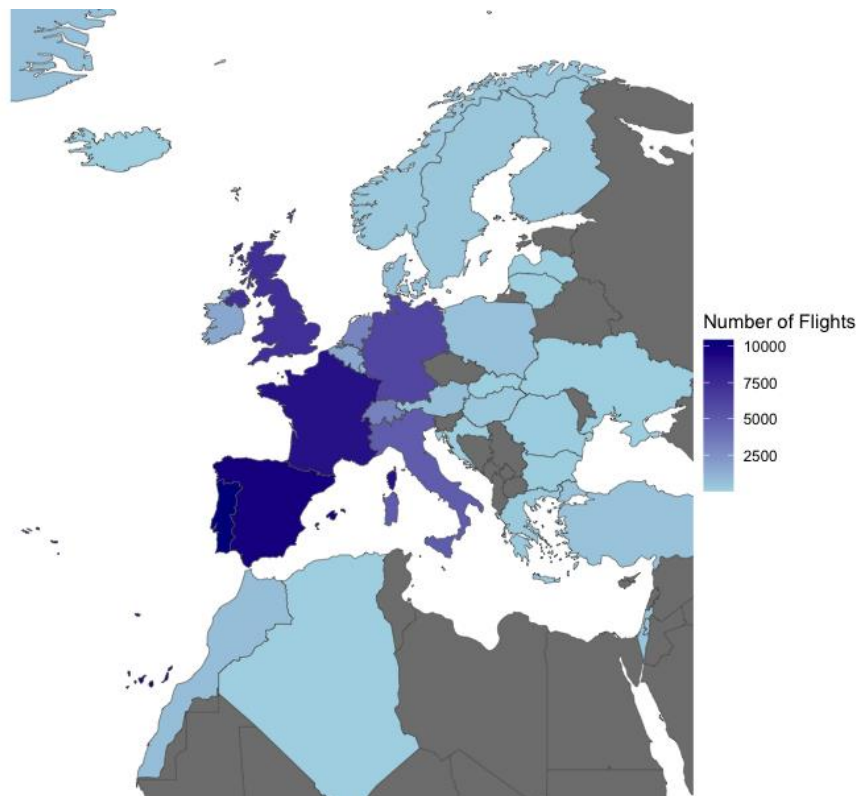


Figura 24 - Mapa de volume de países de origem para Europa e Norte de África.

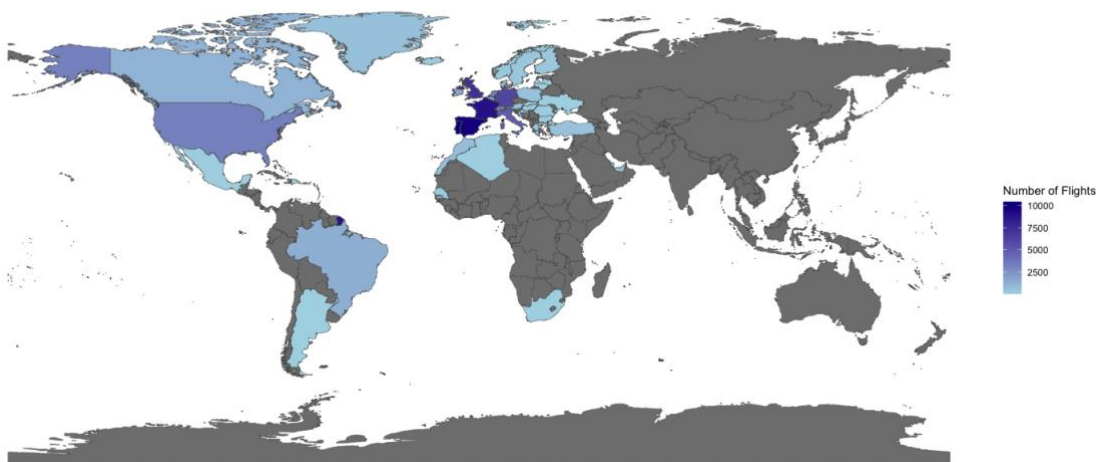


Figura 25 - Mapa de volumes de países de origem mundialmente.

Além da visualização por gráficos e mapas, também foram desenvolvidas tabelas a partir de LPPT_A e LPPT_D que, mesmo possuindo o mesmo formato e apresentação de grandes bases de dados como os *dataframes*, agrega os dados de forma que adquirissem maior capacidade de prover informação. Assim, as tabelas construídas eram também espelhadas para chegadas e saídas e indicavam o volume de voos em função do Tipo de Aeronave (*'Aircraft Type'*), Companhia Aérea (*'Airline'*), Aeroporto de Origem ou de

Destino (‘*ICAO Departure*’ ou ‘*ICAO Arrival*’) e País de Origem ou de Destino (‘*Origin Country*’ ou ‘*Destination Country*’). Além da indicação de volumes, foram adicionados às tabelas algumas modas, como por exemplo, aeronave mais comumente operada, país mais comum a ser operado, companhia aérea mais comum. Usar-se-á as tabelas geradas a partir do *dataframe* LPPT_A para ilustração, através das Tabela 11, Tabela 12, Tabela 13 e Tabela 14.

Tabela 11 - Recorte da Tabela de volume por tipo de aeronave

Aircraft Type	Number of Flights	Most common Airline	Most common Origin Country
A320	15373	TAP Air Portugal	Portugal
B738	10335	Ryanair	Netherlands
E190	6620	TAP Air Portugal	France

Tabela 12 - Recorte da Tabela de volume por Companhia Aérea

Aircraft Type	Number of Flights	Most common Aircraft Type	Most common Origin Country
TAP Air Portugal	41720	A320	Portugal
Ryanair	9208	B738	Italy
EasyJet Europe	5222	A320	France

Tabela 13 - Recorte da Tabela de volume por País de Origem

Origin Country	Number of Flights	Most common Airline	Most common operated Aircraft
Portugal	10435	TAP Air Portugal	A320
Spain	9815	TAP Air Portugal	A320
France	8921	TAP Air Portugal	E190

Tabela 14 - Recorte da Tabela de volume por Aeroporto de Origem

ICAO Departure	Number of Flights	Most common Airline	Most common operated Aircraft
LEMD	5097	TAP Air Portugal	A320
LFPO	3481	TAP Air Portugal	B738
LPMA	3461	TAP Air Portugal	A320

Para além dos *dataframes* LPPT_A e LPPT_D, que geraram as visualizações descritas acima, o *dataframe* LPPT_TRK, como previamente discutido, possui a capacidade e estrutura para apresentar as chegadas no Aeroporto de Lisboa no ponto de vista da trajetória das aeronaves. Considerando que a análise da trajetória para aterragem em um Aeroporto pode ser realizada a partir do ponto de vista do perfil lateral bem como o perfil vertical, LPPT_TRK foi dividido entre LPPT_TRK_VER, contendo as informações de variação de altitude para os voos; e LPPT_TRK_LAT, que apresenta os dados de variação de latitude e longitude ao longo do tempo para os mesmos voos.

Assim, para LPPT_TRK_VER, utilizando a livreria “ggplot2”, foi traçado um gráfico de linhas que ilustra o perfil de descida das aeronaves captadas, desde os últimos 10.000 pés (3.000 m) até o solo, separando cada perfil por cor para melhor identificação dos diferentes voos, e permitindo visualizar as diferentes trajetórias verticais, como mostra a Figura 26.

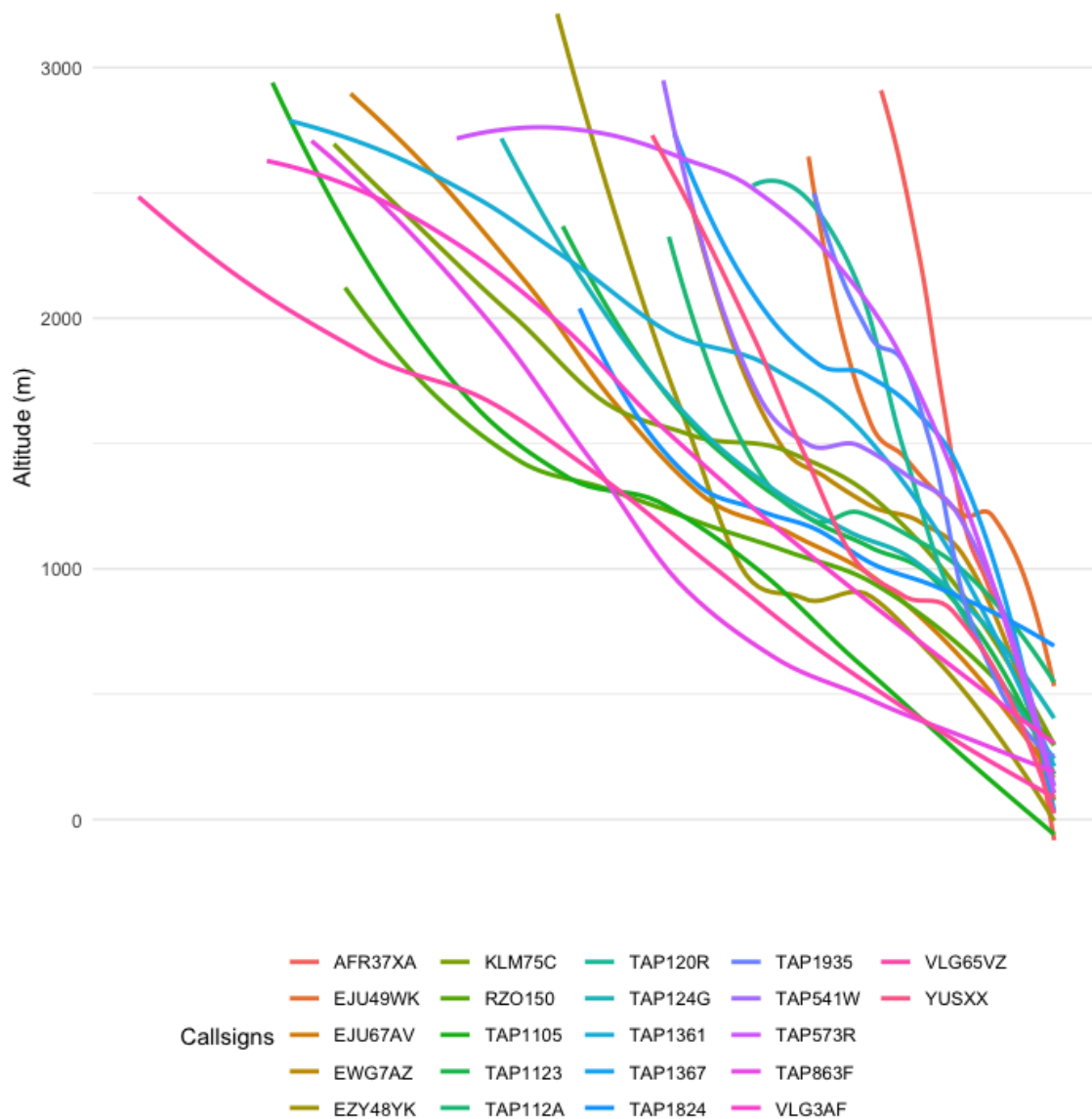


Figura 26 - Gráfico do Perfil Vertical de chegadas no Aeroporto de Lisboa.

Em segundo lugar, para o *dataframe* LPPT_TRK_LAT, utilizando a livreria “leaflet”, foi utilizado uma abordagem similar à abordagem ao LPPT_TRK_VER, na medida em que os dados de variação de, neste caso, latitude e longitude, seriam responsáveis por traçar as rotas de cada voo chegando à Lisboa. A livreria mencionada é responsável pela utilização de mapas sobrepostos com os dados desejados, mapa este que pode ser movido a partir da interação com o utilizador. Foi possível, portanto, desenvolver essa forma de visualização para representar a trajetória lateral para aproximação e aterragem no Aeroporto de Lisboa, como ilustra a Figura 27.

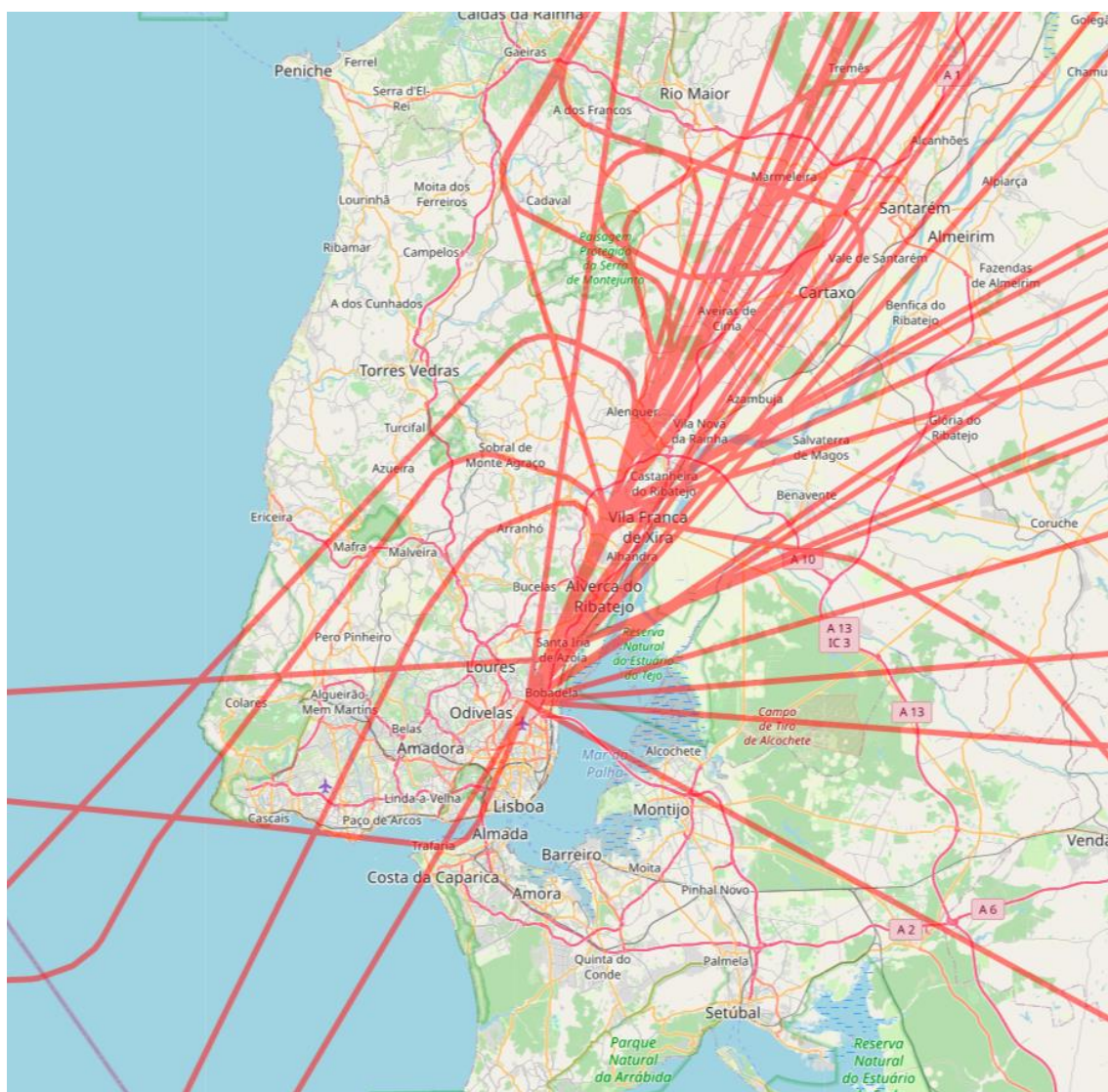


Figura 27 - Mapa do Perfil Lateral de chegadas no Aeroporto de Lisboa.

A partir da concretização do conteúdo responsável por preencher o *dashboard*, passa-se ao momento de estruturar e desenvolver tal ambiente interativo. Para a criação deste, foi utilizada as livrarias “rshiny” e “shinydashboard”, que associam a linguagem R à habilidade de construir ambientes interativos - *dashboards*. Todas as visualizações (gráficos, mapas e tabelas) necessitavam de ser alocados e apresentados de maneira que fossem condizentes com suas respetivas propriedades e tipo de dados.

Assim, o *dashboard* foi estruturado com um painel lateral que dividisse os conteúdos por chegadas (“LPPT_A”), saídas (“LPPT_D”) e rotas (“LPPT_TRK”), além de uma página inicial que apresenta o dashboard, ou seja, a criação de subdivisões que permitem uma navegação mais clara. Logo, o painel lateral é composto pela página “Home”, “Departures Volume”, “Arrivals Volume” e “Arrival’s Routes”. Além disso, as páginas

“Departures Volume” e “Arrivals Volume” dividiam-se em mais duas sub-páginas, “In Time” e “Geographically”, como ilustra a Figura 28.

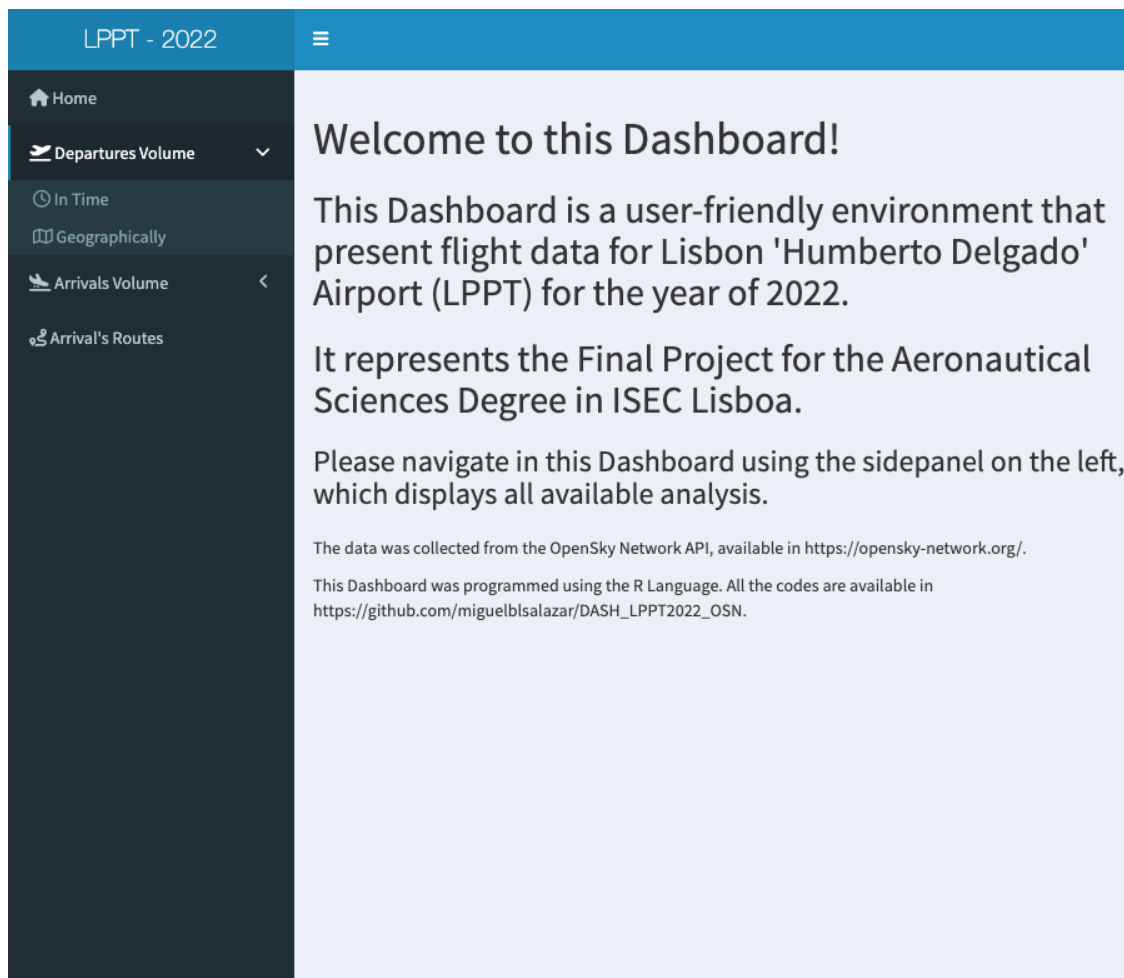


Figura 28 - Página Inicial do dashboard e Painel Lateral.

A partir do “esqueleto” que dá forma ao ambiente interativo, era necessário completar cada janela com as respetivas formas de visualização. Na aba “Departures Volume/In Time”, eram representados os gráficos de volumes de voos em função do tempo, em simultâneo, permitindo ao utilizador selecionar a granularidade temporal desejada. Tal janela possui a mesma estrutura que “Arrivals Volume/In Time”, e a Figura 29 ilustra tal aba.

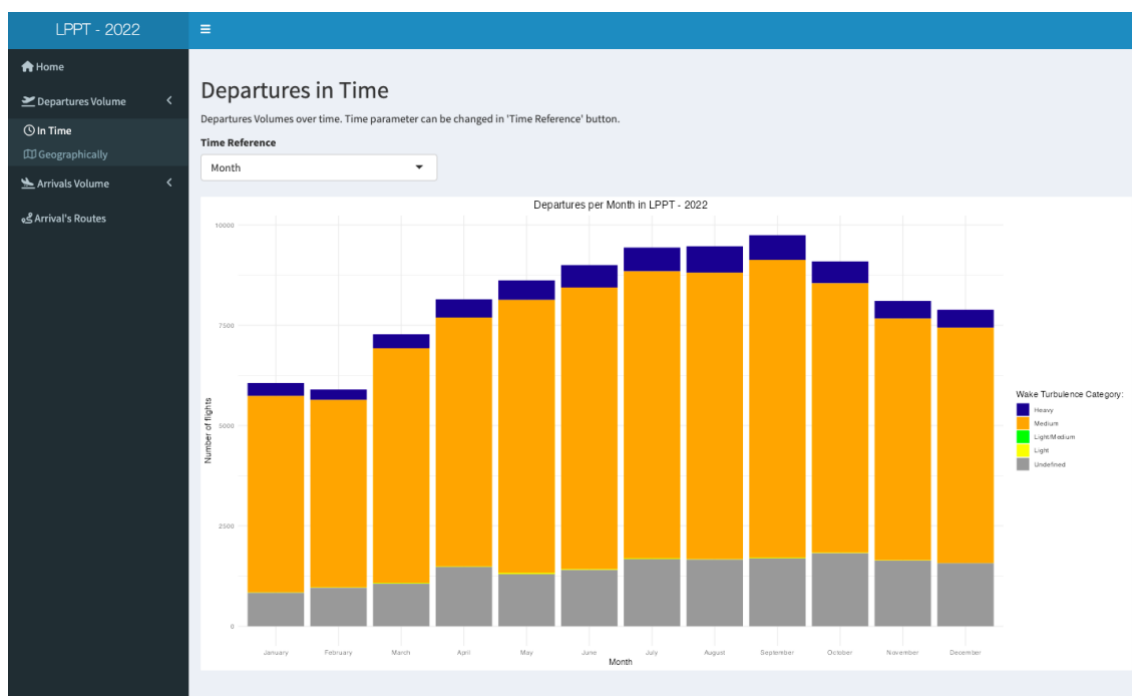


Figura 29 - Página "Departures Volume/In Time".

Para a janela “Departures Volume/Geographically”, que possui os mesmos conteúdos que “Arrivals Volume/Geographically”, é apresentado o mapa de volume em função de aeroporto ou país (escolha do utilizador). Além disso, ao lado, é apresentada a tabela em função dos parâmetros descritos pela Tabela 11, 12, 13 e 14, e a visualização de tais parâmetros também é de escolha do utilizador, como mostra a Figura 30

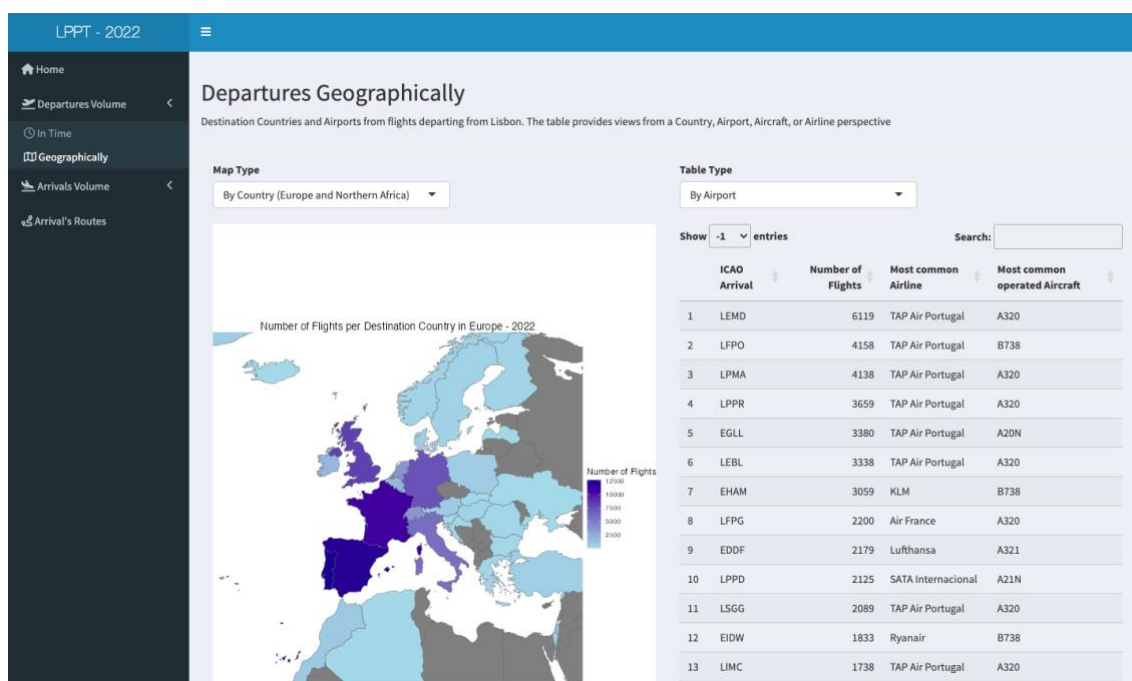


Figura 30 - Página "Departures Volume/Geographically".

A última aba do *dashboard*, “Arrival’s Routes”, apresenta, lado a lado, o mapa do perfil lateral de chegadas ilustrado na Figura 27 e o gráfico de perfil vertical apresentado pela Figura 26. A Figura 31 representa a janela descrita.

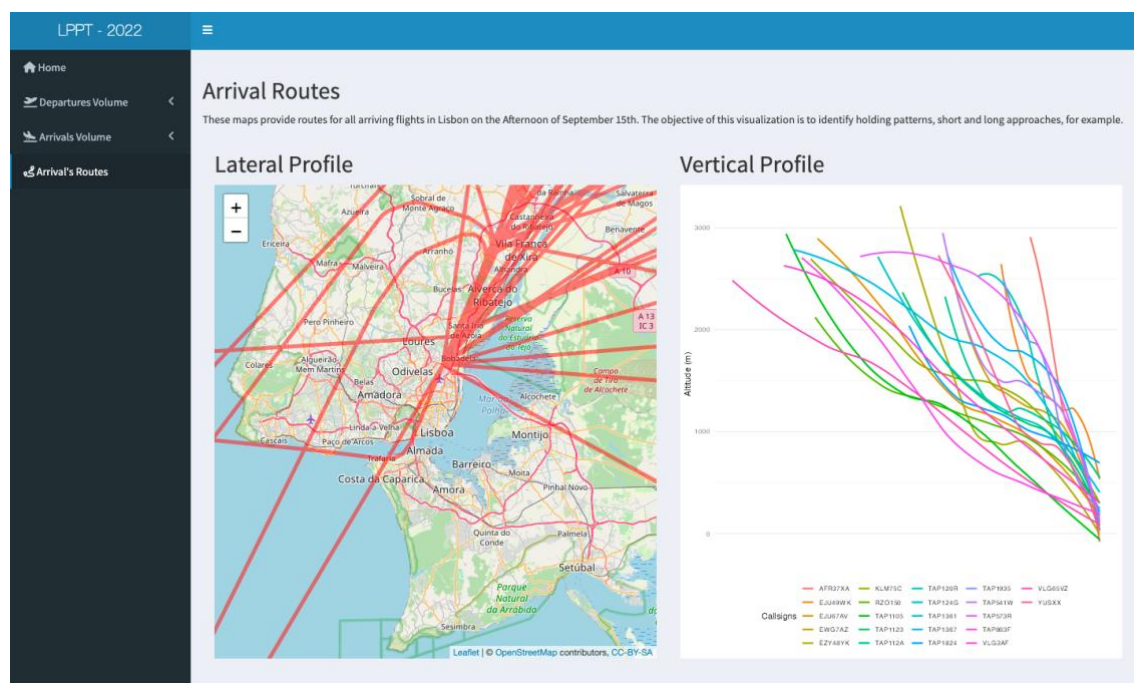


Figura 31 – Página “Arrival’s Routes”.

Portanto, o Desenvolvimento deste projeto, descrito ao longo do Capítulo 3, se deu a partir dos três conceitos principais de *Big Data*: *Data Engineering*, *Data Science* e *Data Visualization*. O primeiro, *Data Engineering*, consistiu na obtenção dos dados advindos do OpenSky Network API, seguido de um direcionamento (modelagem) dos dados em função dos objetivos deste projeto, sustentando-os em fontes de informação auxiliares que qualificaram os dados. A partir deste processo, foram obtidas as “matérias-primas” deste projeto: os *dataframes* “LPPT_A”, “LPPT_D” e “LPPT_TRK”. O segundo conceito, *Data Science*, consistiu no processo de entender as propriedades, características e capacidades de cada informação que os *dataframes* originais fornecem. Por último, *Data Visualization*, foram todos os procedimentos que visavam apresentar as informações desejadas de forma que a escolha das visualizações estivesse de acordo com aquilo que os dados selecionados são capazes de oferecer. A passagem por estas 3 etapas permitiu, por fim, o desenvolvimento do sistema de tratamento de dados a partir da OpenSky Network para a região de Lisboa, num formato de *dashboard* em linguagem R.

Intencionalmente em Branco

4. Análises do Resultado e Discussões

O Capítulo 4 deste projeto é referente à avaliação de tudo que foi obtido à luz dos objetivos estipulados e à literatura estudada. Contextualiza-se, mais uma vez, o cenário a qual este projeto está imerso, posicionando a ferramenta desenvolvida nesse mesmo cenário; avalia-se, também, a efetividade e profundidade dos processos que culminaram na concretização desta ferramenta, além da efetividade, performance e disponibilidade da ferramenta em si; discutem-se e desenvolvem-se as capacidades e limitações do *dashboard* à luz dos objetivos propostos; conecta o projeto à literatura apresentada ao longo do Capítulo 2, apontando ligações; discute conceitos-chave que deram forma à ferramenta; relaciona este projeto com outros em desenvolvimento e já consolidados, com o intuito de perpetuar a metodologia implementada para um escopo mais abrangente.

Em primeiro lugar, volta-se a referenciar que este projeto está imerso dentro do contexto e dos conceitos de *Big Data*. Ou seja, a existência de grande capacidade de coleta de informações relevantes em formato de dados sugere, portanto, uma semelhante capacidade de tratamento e processamento destes dados, transformando grandes bases em informação. No que toca este projeto, que consiste no Desenvolvimento de um sistema de tratamento de dados a partir da OpenSky Network para a região de Lisboa, a OSN possui a capacidade de coletar, posicionar e identificar aeronaves ao redor do globo. Já este projeto visa coletar tais dados e qualificá-los, isto é, dar-lhes a capacidade de transmitir uma mensagem. Tais mensagens são relativas à métricas e parâmetros que estão diretamente conectados ao *Air Traffic Management*, como o volume de aterragens e descolagens, além dos deslocamentos – verticais e laterais – em zonas terminais. A partir da qualificação dos dados em informação, realizou-se a apresentação de tais visualizações através do *dashboard* desenvolvido.

Em segundo lugar, é importante avaliar a eficiência e o passo a passo dos procedimentos que culminaram na criação do *dashboard*, nomeadamente *Data Engineering*, *Data Science* e *Data Visualization*. O processo de Engenharia de Dados, que permitiu a obtenção dos *dataframes* que representaram a matéria-prima deste trabalho – LPPT_A, LPPT_D e LPPT_TRK – foi marcado pela escolha da livreria “openskyr” (Gasco 2020) e subsequente imersão desta no ambiente R através das funções utilizadas (*get_airport_data* para os dois primeiros *dataframes* e *get_track_data*, para o último). O formato oferecido por pré definição do autor (Estrutura Inicial, referenciada na Tabela 3) aliados aos seguintes processos de modelagem e associação de dados com ficheiros

externos (Tabela 4, Tabela 5 e Tabela 6) permitiram às fontes de informação iniciais (*dataframes* “mães”) assumirem o formato ideal e alinhado com os objetivos do projeto, exemplificados nas Tabelas 7 e Tabela 9. No que toca a Ciência de Dados (*Data Science*), este processo permitiu obter uma maior clareza da capacidade dos dados fornecidos pela OSN associados às fontes externas de informação complementar. Consequentemente, a estruturação do pensamento voltado às visualizações que seriam geradas futuramente tornou-se muito mais orientado e objetivo. Por fim, o campo da Visualização de Dados (*Data Visualization*) nada mais foi do que a materialização dos processos anteriores, uma espécie de coroação de todo o trabalho feito no formato de gráficos, mapas e tabelas que estivessem, naturalmente, de acordo com as propriedades de cada informação e alinhados com os objetivos e pressupostos desta ferramenta.

Assim, podemos perceber não só a importância destes elementos dentro dos conceitos de *Big Data*, mas também perceber que, apesar da Figura 18 ilustrar um processo linear, a construção deste projeto sempre se mostrou como uma espécie de antevisão daquilo que se gostaria de visualizar em função das propriedades obtidas, e em função do que foi contruído como alicerce, inicialmente. Ou seja, a linearidade mostrada não se verifica, ao passo que, por exemplo, muito do que foi realizado na fase de *Data Engineering* tinha em mente também o “produto final”, isto é, a visualização, e o mesmo vale para o processo científico em função da engenharia bem como da visualização.

O resultado dos processos descritos acima se materializou no formato do *dashboard* ilustrado pelas Figuras 28, Figura 29, Figura 30 e Figura 31. Este ambiente interativo, já disponível para navegação, estudo e utilização, atende aos objetivos propostos por este projeto na medida em que fornece análises de volume (chegadas e saídas) bem como uma análise geográfica de rotas para o Aeroporto de Lisboa em 2022. A escolha do ano de 2022 – sendo um ano “concluído” e com pouca influência da COVID-19 – permitiu a realização de análises realistas ao longo de todo um ano, permitindo a compreensão de padrões de tráfego para um Aeroporto de grande importância no contexto europeu e mundial. É, portanto, uma ferramenta interativa, mas também com o caráter *Real time*, isto é, a extrapolação deste estudo a outro período ou outro aeroporto altera automaticamente as visualizações apresentadas no *dashboard*. De maneira a garantir o acesso completo à informação que sustenta este projeto, foi criado um repositório na plataforma Github, que armazena e apresenta os códigos utilizados e permite a utilização deste para novos estudos. O projeto está disponível através da

hiperligação “<https://github.com/pffmachado/flightir>” (Salazar e Filipe Faria Machado 2023).

A partir deste contexto, é possível avaliar as capacidades e limitações desta ferramenta. É um ambiente interativo na medida em que permite a ação do utilizador como selecionador dos parâmetros que alteram a visualização dos dados (escalas temporais, geográficas, parâmetros da companhia, aeronave, por exemplo), permitindo, assim, diferentes formas de observar os mesmos dados e até mesmo diferentes formas de mesclar dados de diferentes visualizações; As visualizações apresentadas são capazes de valorizar os dados que apresentam, na medida em que são condizentes com a mensagem que desejam transmitir. As diferentes formas de apresentar o volume de descolagens e aterragens para o Aeroporto de Lisboa, além de uma terceira abordagem voltada para as trajetórias, permitem uma visão mais holística das operações. Utilizando como exemplo o gráfico de Volume em função de escalas temporais, ilustrado na Figura 22, permite aferir padrões de tráfego nas mais variadas leituras de tempo (por mês, para compreender sazonalidade; por dia da semana, para entender fluxos semanais; e por período e hora do dia, para se compreender as operações com maior granularidade e precisão).

Por outro lado, é possível identificar limitações na ferramenta desenvolvida, como a discrepância considerável entre as chegadas registadas pelo *dataframe* LPPT_A e as partidas registadas pelo *dataframe* LPPT_D, como ilustra a Figura 20. A discrepância entre valores pode estar associada à forma como a OpenSky Network regista os Aeroportos de Origem e Destino: pela presença de outros aeródromos na proximidade do Aeroporto de Lisboa, a base de dados pode associar o destino como próximo à LPPT (LPCS – Aeródromo de Cascais, LPAR – Aeródromo de Alverca, LPST – Aeródromo de Sintra, por exemplo). Da mesma forma que voos que partem de Lisboa podem ter um código ICAO de aeroporto de destino não condizente com o aeródromo real de destino. Apesar desta discrepância, a análise realizada não foge do espectável e mantém-se uma coerência dos dados. Além disso, o API da OSN não provê uma forma direta para permitir a realização de uma análise dos atrasos, isto é, não existe uma comparação entre o horário previsto e o horário atual de chegada ou saída. Portanto, pode-se concluir que a ferramenta desenvolvida possui grande capacidades de processamento e apresentação de grandes bases de dados, porém está sujeita à algumas irregularidades e discrepâncias que estão relacionadas a fontes externas a este projeto.

Tendo em mente as características e propriedades de tal ferramenta, direciona-se o foco desta análise e discussão para a relação deste projeto com toda a literatura apresentada e desenvolvida ao longo do Capítulo 2. Apesar de toda a coletânea de documentos, relatórios, *papers* e livros dialogarem direta ou indiretamente, existem quatro que possuem uma conexão direta com a realização e concretização deste trabalho. “Assessing the Global COVID-19 Impact on Air Transport with Open Data” é um documento que, semelhante a este projeto, faz uso do ambiente promovido pela linguagem R e suas livrarias para alimentar análises e visualizações vindas da OpenSky Network (Koelle e Barbosa 2021). A procura pela compreensão dos impactos da pandemia da COVID-19 nos transportes aéreos direcionou o projeto para a base de dados da OSN, onde foi capaz de buscar volumes de voos em diferentes regiões do Globo, a fim de estabelecer comparações entre países, no que diz respeito a escolhas políticas, econômicas e sociais. O documento “Assessing the Global COVID-19 Impact on Air Transport with Open Data” relaciona-se, portanto, diretamente com este projeto tendo em vista a metodologia através do ambiente R, e relaciona-se mais especificamente com o desenvolvimento dos *dataframes* LPPT_A e LPPT_D, responsáveis pelo fornecimento de volumes de chegadas e saídas para o Aeroporto de Lisboa.

Por outro lado, mais voltado à análise de trajetórias que, neste projeto, fora abordada através do *dataframe* LPPT_TRK, é possível destacar dois documentos responsáveis por fornecerem referências e guiamento para a concretização da secção deste projeto. “A Geometric Approach to Study Aircraft Trajectories: The Benefits of OpenSky Network ADS-B Data” é um estudo que, apesar de possuir um enquadramento mais focado na geometria das trajetórias e comparação entre as fontes de dados do EUROCONTROL e do ADS-B (Perrichon, Gendre, e Klein 2022), fornece exemplos que dialogam com os objetivos deste projeto. Por exemplo, a visualização das chegadas dos voos que chegam de LFBO (Toulouse-Blagnac) em LFPO (Paris-Orly), como ilustrado na Figura 13, se relacionam diretamente com a análise apresentada no *dashboard* relativamente ao Perfil Lateral de chegadas ao Aeroporto de Lisboa, na medida em que apresentam trajetórias dos voos com um alto grau de precisão, fornecendo a possibilidade de visualizar as trajetórias em TMAs de alto tráfego.

Além da análise de trajetórias num perfil vertical, o documento “Evaluation of Flight Efficiency for Stockholm Arlanda Airport using Open Sky Network Data” provê uma análise de perfil vertical de determinada aeronave – também comparando fontes de dados

(OpenSky, EUROCONTROL, por exemplo) – numa ótica de melhor compreender o consumo de combustível, atrasos e esperas em TMA e falta de otimização de ATM para o Aeroporto de Estocolmo, Arlanda (Polishchuk, Lemetti, e Sáez 2019). A Figura 8 representa o perfil vertical para aproximação do voo SAS 410 e, a partir de modelos matemáticos capazes de compreender o consumo maior ou menor de combustível, é possível se estudar a eficiência de tal Aeroporto, inclusive de sua gestão de tráfego. Esta representação é uma das bases de apoio da análise realizada no *dashboard* para o perfil vertical, a partir dos voos que chegaram à Lisboa durante a tarde do dia 15 de Setembro de 2022, como ilustra a Figura 26.

Abordando um tema mais geral deste trabalho, que é a estruturação dos códigos, ficheiros e visualizações em função do conceito de Pesquisa Reprodutível, este pode ser visto como um dos grandes alicerces dentro do pilar do R Studio, tendo em vista a sua grande relevância no que diz respeito a propagar este estudo a outros Aeroportos, outros períodos do tempo e outras lentes. O livro “Reproducible Research with R and RStudio – Second Edition” busca apresentar este conceito, que nada mais é a capacidade de uma pesquisa, estudo ou análise, de ser independentemente reproduzida por outros indivíduos, de maneira que os resultados sejam semelhantes ou coerentes (Gandrud 2018). Tendo em vista este projeto, tanto a escrita dos códigos – de maneira coesa e coerente – com comentários explicativos, aliada à uma estrutura de ficheiros bem consolidada e disponível publicamente na plataforma “Github”, pode-se afirmar que este projeto caminha ao encontro dos conceitos explorados por Gandrud. Ou seja, a possibilidade de se criar um *dashboard* para o aeroporto do Porto (LPPR), Frankfurt (EDDF) ou Rio de Janeiro (SBGL), sem que a estrutura principal seja alterada, sendo possível analisar volumes de chegadas e saídas, trajetórias laterais e verticais.

Portanto, este projeto foi capaz de consolidar e formalizar a relação entre o API da OpenSky Network, utilizando um *dashboard* para o Aeroporto de Lisboa como materialização dos conceitos de Big Data, à luz de estudos prévios que, direta ou indiretamente, dialogam com os temas fulcrais aqui desenvolvidos. Além disso, a estruturação deste projeto passa também pela possibilidade de reproduzi-lo em outros cenários e modelos, mantendo uma estrutura base.

Intencionalmente em Branco

5. Conclusões e Recomendações

No quinto e último Capítulo, serão retomadas as premissas iniciais estabelecidas e serão apresentadas recomendações e possíveis projeções deste projeto. Portanto, este consistia no desenvolvimento de um sistema de tratamento de dados a partir da OpenSky Network para a região de Lisboa. Este ideal materializou-se em um *dashboard* responsável por apresentar o volume de chegadas, volume de partidas e análise de trajetórias para o Aeroporto de Lisboa em 2022. Dentre os objetivos estabelecidos além da criação desta ferramenta, estava a busca por sanar a falta de um intermediário entre o utilizador e grandes bases de dados como a OpenSky Network, FlightRadar24, por exemplo, além de prover ao ISEC Lisboa e à comunidade aeronáutica como um todo, um ambiente capaz apresentar informações relativas a padrões de tráfego aéreo para Lisboa.

A partir de tais condições iniciais, buscou-se no meio acadêmico fontes que se apoiassem nos mesmos temas e conceitos abordados neste projeto e que, conseqüentemente, auxiliaram na constituição das bases científicas do mesmo, garantindo não só credibilidade, mas também um guiamento voltado à conceitos estabelecidos na comunidade científica. As fontes buscadas asseguraram as premissas deste projeto, justificando a necessidade da criação de um sistema de tratamento de dados, de um ambiente interativo de visualização e da construção de todo este processo em função da reprodutibilidade de pesquisa.

Assim, o desenvolvimento deste projeto apoiou-se na comunidade e desenrolou-se a partir de três conceitos, ou camadas, dentro de *Big Data: Data Engineering, Data Science e Data Visualization*. O primeiro facilitou a criação das matérias-primas deste projeto: os *dataframes* LPPT_A, relativamente às chegadas no Aeroporto de Lisboa; LPPT_D, que apresenta as partidas do Aeroporto de Lisboa; e LPPT_TRK, que apresenta as trajetórias laterais e verticais de um recorte de voos que chegaram à Lisboa. Tal processo consistiu, principalmente, na utilização da linguagem R e associação, no ambiente de programação, entre a livreria “openskyr”, materializada nas funções *get_airport_data* e *get_track_data*, e ficheiros externos de órgãos aeronáuticos responsáveis por dar mais precisão aos dados apresentados. *Data Science*, por sua vez, fora responsável por organizar as matérias-primas adquiridas e compreender que tipo de informação e mensagem estas podem oferecer, ou seja, compreender a componente temporal, geográfica e os parâmetros voltados à Aeronave e à Companhia Aérea. Este processo também já estava voltado a esboçar e experimentar a combinação de diferentes componentes e métricas, garantindo

uma visualização coerente e com alto valor informacional. Por fim, a Visualização dos Dados consistiu na utilização do ambiente R para gerar gráficos, mapas e tabelas que seriam, posteriormente, apresentados em um ambiente interativo. Para tal, foram utilizadas as livrarias “rshiny” e “shinydashboard” para concretizar esta ferramenta no ambiente do R Studio. No fim, o resultado foi um *dashboard* que apresenta o volume de chegadas e saídas, bem como uma apresentação de trajetórias para voos que passaram pelo Aeroporto de Lisboa no ano de 2022.

No âmbito da perpetuação deste projeto para outros contextos em cenários, a publicação dos códigos das visualizações bem como a publicação do código do dashboard na plataforma Github, de acesso gratuito, auxilia na obtenção de tal informação e propagação desta abordagem às bases de dados atualmente disponíveis. Tendo em mente o caráter colaborativo deste universo de *Big Data*, deve-se notar que diferentes fontes de informação podem – e devem – ser utilizadas por essa plataforma interativa. Escalar esta ferramenta pode ser de grande utilidade para estudos similares em outros escopos.

Ainda dentro do conceito de colaboração, e levando-se em consideração o desenvolvimento do “Projeto de uma estação ADS-B na rede OpenSky Network” dentro do próprio ISEC Lisboa, o diálogo entre estes dois projetos pode gerar uma ferramenta mais precisa e que receba informações de maneira que seja de acordo com a demanda, sem que sejam necessárias conversões ou ficheiros externos. Tendo em mente que o projeto mencionado visa desenvolver uma base de dados similar à oferecida pela OpenSky Network, Flight Radar 24, por exemplo, há de existir, portanto, maior maleabilidade e adaptabilidade no que diz respeito a estruturação dos dados que alimentam este projeto e outros futuros.

Esta colaboração dentro do próprio ISEC Lisboa permite que os conceitos explorados neste e em outros projetos complementares extrapolem com mais força as paredes da instituição, atendendo à um dos principais objetivos não só deste, mas de parte de projetos académicos: a possibilidade de validação e reprodução em outro nicho, fornecendo à comunidade diferentes abordagens e ferramentas que auxiliam no aproveitamento de dados e informações armazenadas.

Bibliografia

- Abdulaziz, Abdulrazaq, Surajo Namadi, Abdulmalik Yaro, e Ashraf Ahmad. 2019. «Surveillance Radar System Limitations and the Advent of the Automatic Dependent Surveillance Broadcast system for Aircraft Monitoring». www.atbuftejoste.com.
- Ayala, Rafael, Daniel Ayala, Lara Sellés Vidal, e David Ruiz. 2021. «openSkies - Integration of Aviation Data into the R Ecosystem». *The R Journal* 13 (2): 590–99.
- Baek, Kwang-Yul, e Hyo-Choong Bang. 2012. «ADS-B based Trajectory Prediction and Conflict Detection for Air Traffic Management». *International Journal of Aeronautical and Space Sciences* 13 (3): 377–85. <https://doi.org/10.5139/ijass.2012.13.3.377>.
- Comitz, Paul, e Aaron Kersch. 2016. «Aviation analytics and the Internet of Things Network». Em *Integrated Communications Navigation and Surveillance (ICNS)*. IEEE. <https://doi.org/10.1109/ICNSURV.2016.7486327>.
- EASA. 2022. «Certification Specifications and Guidance Material for Aerodrome Design (CS-ADR-DSN)».
- Gandrud, Christopher. 2018. *Reproducible Research with R and RStudio*. Editado por Chapman and Hall/CRC. 2.^a ed.
- Gasco, Luis. 2020. «Luis Gasco - openskyr». 2020. <https://github.com/luisgasco/openskyr>.
- Gilbert, Glen A. 1973. «Historical Development of the Air Traffic Control System». *IEEE Transactions on Communications* 21 (5): 364–75. <https://doi.org/10.1109/TCOM.1973.1091699>.
- ICAO. 2016. *PANS-ATM Doc 4444*. ICAO.
- . 2021. «2020 passenger totals drop 60 percent as COVID-19 assault on International mobility continues». 2021.
- ICAO Asia. 2007. «Guidance material on surveillance Technology Comparison».

- Koelle, Rainer, e Fabio Lourenco Carneiro Barbosa. 2021. «Assessing the Global COVID-19 Impact on Air Transport with Open Data». Em *AIAA/IEEE Digital Avionics Systems Conference - Proceedings*. Vol. 2021-October. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/DASC52595.2021.9594337>.
- Kumari, Anita, e Neeraj Verma. 2019. «AN ANALYTICAL REVIEW STUDY ON BIG DATA ANALYSIS USING R STUDIO». *International Journal of Engineering Technologies and Management Research* 6 (6): 116–22. <https://doi.org/10.5281/zenodo.3266146>.
- Miranda, Miguel. 2022a. «Primary Radar». *Air Dream College*.
- . 2022b. «SSR & Transponder». *Air Dream College*.
- Open Sky Network. 2022. «About the OpenSky Network». 2022. <https://opensky-network.org/>.
- OpenSky Network. 2021. «OpenSky Rest API». 2021. <https://openskynetwork.github.io/opensky-api/rest.html#arrivals-by-airport>.
- Perrichon, Rémi, Xavier Gendre, e Thierry Klein. 2022. «A Geometric Approach to Study Aircraft Trajectories: The Benefits of OpenSky Network ADS-B Data». Em *Engineering Proceedings*. Basel Switzerland: MDPI. <https://doi.org/10.3390/engproc2022028006>.
- Polishchuk, Tatiana, Anastasia Lemetti, e Raúl Sáez. 2019. «Evaluation of Flight Efficiency for Stockholm Arlanda Airport using OpenSky Network Data». *Proceedings of the OpenSky Workshop 2019: EPIc series in computing* 67: 13–24.
- Rodrigues, Claudia Vanessa Coutinho. 2010. «ADS-B - Automatic Dependent Surveillance Broadcast. Estudo do Impacto em Portugal».
- Salazar, Miguel, e Paulo Filipe Faria Machado. 2023. «Flighir - Github Repository». 2023. <https://github.com/pffmachado/flightir>.
- Santos, João. 2023. «Projeto de uma estação ADS-B na rede OpenSky Network». ISEC Lisboa.

- Schäfer, Matthias, Vincent Lenders, e Ivan Martinovic. 2013. «Experimental Analysis of Attacks on Next Generation Air Traffic Communication». Em *International Conference on Applied Cryptography and Network Security*.
- Schäfer, Matthias, Martin Strohmeier, Vincent Lenders, Ivan Martinovic, Matthias Wilhelm, e T U Kaiserslautern. 2014. «Bringing up OpenSky: A Large-scale ADS-B Sensor Network for Research». Em *IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*, 83–94. <http://tinyurl.com/cwsyr2r>.
- Schäfer, Matthias, Martin Strohmeier, Matthew Smith, Markus Fuchs, Vincent Lenders, Ivan Martinovic, OpenSky Network, e Switzerland lastname. 2018. «OpenSky Report 2018: Assessing the Integrity of Crowdsourced Mode S and ADS-B Data». <http://opensky-network.org>.
- Schultz, Michael, Judith Rosenow, e Xavier Olive. 2022. «Data-driven airport management enabled by operational milestones derived from ADS-B messages». *Journal of Air Transport Management* 99: 102164. <https://doi.org/10.1016/j.jairtraman.2021.102164>.
- Seymour, William. 2018. «Optimising the OpenSky Network».
- Strohmeier, Martin, Xavier Olive, Jannis Lübke, Matthias Schäfer, e Vincent Lenders. 2021. «Crowdsourced air traffic data from the OpenSky Network 2019-2020». *Earth System Science Data* 13 (2): 357–66. <https://doi.org/10.5194/essd-13-357-2021>.
- Strohmeier, Martin, Matthias Schäfer, Vincent Lenders, e Ivan Martinovic. 2014. «Realities and Challenges of NextGen Air Traffic Management: The Case of ADS-B». *IEEE Communications Magazine*, 2014.
- Sun, Junzi, Luis Basora, Xavier Olive, Martin Strohmeier, Matthias Schäfer, Ivan Martinovic, e Vincent Lenders. 2022. «OpenSky Report 2022: Evaluating Aviation Emissions Using Crowdsourced Open Flight Data». <https://opensky-network.org>.
- Wesson, Kyle D, Todd E Humphreys, e Brian L Evans. 2014. «Can Cryptography Secure Next Generation Air Traffic Surveillance?» Em *IEEE SECURITY & PRIVACY*.
- World Health Organization. 2022. «WHO Coronavirus (COVID-19) Dashboard». 2022.