



UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO

CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA

ESCOLA DE INFORMÁTICA APLICADA

Geração de uma Base de Dados Aberta para Apoio a Análise dos Registros de Voos das
Autoridades Operados pela Força Aérea Brasileira

Lucas Swinerd & Davi Alves

Orientador

Prof. Kate Cerqueira Revoredo, D.Sc.

RIO DE JANEIRO, RJ – BRASIL

DEZEMBRO DE 2016

Geração de uma Base de Dados Aberta para Apoio a Análise dos Registros de Voos das
Autoridades Operados pela Força Aérea Brasileira

LUCAS SWINERD & DAVI ALVES

Projeto de Graduação apresentado à Escola de
Informática Aplicada da Universidade Federal do Estado
do Rio de Janeiro (UNIRIO) para obtenção do título de
Bacharel em Sistemas de Informação.

Aprovada por:

KATE CERQUEIRA REVOREDO

FERNANDA ARAÚJO BAIÃO

CLÁUDIA CAPELLI

RIO DE JANEIRO, RJ – BRASIL. DEZEMBRO DE 2016

Agradecimentos

Eu, Lucas,

Gostaria primeiramente de agradecer meus pais, João Alfredo e Ana Maria, pois sem a ajuda deles nada disso seria possível. Apesar das desavenças, sempre me apoiaram e suportaram minhas decisões.

Aos meus amigos, agradeço a ajuda em minha jornada, mesmo que indiretamente auxiliando a percorrer meu caminho, sempre seguindo em frente. Um agradecimento especial ao meu amigo de trabalho, e de vida, Davi Alves, o qual apresentou sua ideia de projeto e a dividiu comigo, para que trabalhássemos em conjunto no desenvolvimento deste projeto.

Agradeço, também, a minha linda e compreensível namorada, Bruna Comanche, pela paciência e as inúmeras horas de séries sozinha enquanto foi preciso finalizar este trabalho.

Eu, Davi,

Agradeço aos meus pais por me darem a base para trilhar os meus próprios caminhos, me ajudando nos meus estudos no que fosse possível.

Agradeço, também, a minha esposa e minha filha pelo apoio que me deram, não estaria aonde estou hoje se não por elas, por entender o sacrifício de algumas horas com elas e dedicadas a este trabalho.

Ambos,

Agradecemos nossa orientadora Kate Revoredo, pela disposição e atenção, apesar dos horários apertados, em nos ajudar na construção deste trabalho. Por fim, agradecemos a todos os mestres por quais tivemos o prazer de dividir uma sala de aula e nos ajudaram a completar este objetivo.

RESUMO

A indignação popular com os gastos governamentais está se tornando uma forte marca da era da transparência. Com tantas leis, decretos e complementos criados com o intuito de facilitar o acesso as informações desses gastos, é preciso estar atento a como esta informação é divulgada. Tentando aproximar o cidadão do governo, leis de transparência foram criadas, permitindo a fiscalização dos processos de elaboração de planos, lei de diretrizes orçamentárias e orçamentos e, também, tentando minimizar fraudes e superfaturamento de produtos, sistemas ou projetos implementados. Os acessos a essas informações fortalecem a capacidade dos cidadãos de participar mais efetivamente do processo de tomada de decisões que os afetam. De acordo com as leis, todos os gastos do governo devem ser disponibilizados em seus sites oficiais. Neste trabalho, será posto em pauta as viagens aéreas realizadas por entidades políticas através da Força Aérea Brasileira. Este é um direito concedido a poucas entidades, como os ministros. A Força Aérea Brasileira disponibiliza algumas informações sobre estes voos, porém, não é fácil realizar qualquer tipo de análise, devido ao formato do arquivo utilizado e a escolha das informações divulgadas tornarem-se um limitador, pois realizar tais análises seria extremamente custoso sem qualquer tipo de manipulação ou enriquecimento de dados. Assim, se inicia a proposta deste trabalho, a qual visa reunir a maior quantidade de informação útil sobre estes voos, em um único lugar, utilizando métodos de Sistemas de Informação que facilitarão a extração, manipulação, divulgação de uma base de dados aberta que apoiará análises.

Palavras-chave: FAB, Força Aérea Brasileira, Transparência, Gastos Governamentais, Dados Abertos.

ABSTRACT

Popular outrage over government spending is becoming a strong pattern in this new era of transparency. With so many laws, decrees and law supplements that are created to facilitate information access, it becomes necessary to raise awareness about how this information is diffused. With the purpose of reaching its citizens, transparency laws were created, allowing the fiscalization of the process of plan elaboration, budget guideline laws and to minimize fraud, overbilling of products, systems and implemented projects. Access to this information strengthens citizens' ability to participate more effectively in the decision-making process that affects them. Without a doubt, the world-wide web is the most effective way to reach these objectives. All government spending should be made available on their official websites, according to Brazilians laws. The use of flights by political entities through Brazilian Air Force will be put into questioning in this work. In Brazil's political system, this is a right granted to only some entities, such as ministers. The format and disclosed information chosen by Brazilian Air Force, doesn't facilitate any kind of analysis of the flights. To do it without any kind of manipulation would be extremely exhaustive. Thus, the proposal of this work begins, which its objective is to create a unified base with all the useful data about these flights, through applying Information Systems methods that will facilitate extraction, manipulation, diffusion of this unified open base to assist on analysis of the data.

Keywords: Brazilian Air Force, Government Spending, Transparency, Open Data.

ÍNDICE

1. Introdução	12
1.1. Motivação	12
1.2. Objetivos	14
1.3. Organização do texto	14
2. Ferramentas	15
2.1. ParseHub	15
2.2. Bonita Studio / BPM Engine – Bonitasoft	17
2.3. Google BigQuery	19
2.4. Tableau.....	19
3. Execução do Trabalho	21
3.1. Escopo do Trabalho.....	21
3.2. Coletar Dados.....	22
3.2.1. Extrair Links.....	24
3.2.2. Download dos Arquivos.....	26
3.2.3. Converter os Arquivos para XML.....	26
3.3. Enriquecer Dados.....	27
3.3.2. Dados.....	27
3.4. Criar Base de Dados.....	28
3.4.1. Definir Modelo.....	28
3.5. Definição Formal dos Atributos.....	29
3.5.1. Preparar Dados	31
3.6. Disponibilizar Dados	32
3.7. Importar Base.....	34
3.7.1. Instruções	34

4.	Análise	36
4. 1.	Relação Quantidade de Voos por Período.....	36
4. 2.	Relação de Voo por Autoridade	37
4. 2. 1.	Boxplot de Voo por Autoridade ao Ano	38
4. 2. 2.	Número de Voos por Autoridade de julho a dezembro de 2013.....	39
4. 2. 3.	Número de Voos por Autoridade de 2014	40
4. 2. 4.	Número de Voos por Autoridade de 2015	42
4. 2. 5.	Número de Voo por Autoridade de 1º de janeiro a 12 de maio de 2016 – Referente ao Governo da Presidente Dilma Rousseff	44
4. 2. 6.	Número de Voo por Autoridade de 12 de maio a 10 de novembro de 2016 – Referente ao Governo do Presidente Michel Temer	45
4. 3.	Análise de Voos do Governo da Presidente Dilma Rousseff em relação ao Governo do Presidente Michel Temer	47
4. 3. 1.	Análise do Período TD e TT	47
4. 4.	Boxplot – Máximo de Passageiros por Autoridade.....	49
4. 5.	Total de Passageiros por Autoridade ao Ano	50
4. 5. 1.	Comparativo entre 2014, 2015 e 2016.....	51
4. 6.	Relação do Número de Voos por Tempo de Mandato	52
5.	Conclusão	56
5.1.	Considerações Finais	56
5.2.	Limitações do Projeto	57
5.2.1.	Associar aeronave ao voo.....	57
5.2.2.	Fonte descentralizada de informações.....	57
5.2.3.	Valorar os gastos públicos.....	58
5.3.	Trabalhos Futuros	58

ÍNDICE DE TABELAS

Tabela 1 - Passageiro.....	29
----------------------------	----

Tabela 2 - <i>Autoridade</i>	30
Tabela 3 - Gabinete.....	30
Tabela 4 - Voo_Passageiro.....	30
Tabela 5 - Voo.....	30

ÍNDICE DE FIGURAS

Figura 1 - Diagrama da Integração das Ferramentas com as Etapas do Processo	15
Figura 2 - Exemplo de projeto no ParseHub	17
Figura 3 - Exemplo de Definição de Processo no Bonita Studio	18
Figura 4 - Exemplo de gráfico no Tableau	20
Figura 5 - Definição do Processo	21
Figura 6 - Especificação do Macroprocesso de Coleta de Dados.....	22
Figura 7 - Informações contidas no PDF da Força Aérea Brasileira.....	23
Figura 8 – Definição da seção “flight extract” no ParseHub.....	24
Figura 9 - Definição da seção “select flight year” no ParseHub	24
Figura 10 – Exemplo do resultado final gerado pelo ParseHub em JSON.....	25
Figura 11 – Exemplo de seleção no site da FAB.....	25
Figura 12 – Exemplo da estrutura de arquivos gerada pelo script.....	26
Figura 13 – Lista de parâmetros do “pdftohtml”	26
Figura 14 - Especificação dos subprocessos de Criar Base de dados.....	28
Figura 15 - Modelo lógico entidade-relacionamento da base de dados	29
Figura 16 - Passo 1 da Disponibilização pelo BigQuery	33
Figura 17 - Passo 2 da Disponibilização pelo BigQuery	33
Figura 18 - Passo 1 da Importação de Base no BigQuery	34
Figura 19 - Passo 2 da Importação de Base no BigQuery	35
Figura 20 - Gráfico de Linha - Número de Voos por Ano e sua Média - Sem 2013	37
Figura 21 - Gráfico de Linha - Número de Voos por Ano e sua Média.....	37
Figura 22 – Gráfico de Boxplot – Voo por Autoridade ao Ano	39

Figura 23 – Gráfico de Barra – Voos por Autoridade no período de 2013	40
Figura 24 - Gráfico de Barra – Voo por Autoridade no periodo de 2014	41
Figura 25 - Gráfico de Barra - Voos por Autoridade no período de 2015	43
Figura 26 - Gráfico de Linha - Comparação do Número de Voos dos anos de 2014 a 2016	44
Figura 27 – Gráfico de Barra – Voo por Autoridade no período pré-impeachment da Presidente.....	45
Figura 28 – Gráfico de Barra – Voo por Autoridade no período pós-impeachment da Presidente.....	46
Figura 29 - Gráfico de Linha - Comparativo entre períodos do governo Dilma e Temer.	48
Figura 30 - Gráfico de Boxplot - Valor Máximo de passageiros em um Voo por Autoridade	50
Figura 31 - Gráfico de Barra - Total de Passageiros por Autoridade	51
Figura 32 - Exemplos de Autoridades que Refutam a Hipótese.....	52
Figura 33 - Relação Voo x Vigência - Ricardo Leyser	53
Figura 34 - Relação Voo x Vigência - Arthur Chioro	53
Figura 35 - Relação Voo x Vigência - Ministério da saúde	53
Figura 36 - Relação Voo x Vigência - Ministério das Cidades	54
Figura 37 - Gráfico de Linha - Comparação entre os períodos das Autoridades do Ministério das Cidades	55

LISTA DE SIGLAS, ABREVIATURAS E SÍMBOLOS

API	Application Programming Interface
BPM	Business Process Management
BPMN	Business Process Modeling Notation
FAB	Força Aérea Brasileira
JSON	JavaScript Object Notation
PDF	Portable Document Format
TB	Terabyte
URL	Uniform Resource Locator
XML	Extensible Markup Language

1. Introdução

1.1. Motivação

A divulgação de informações governamentais, como por exemplo, contratos ou licitações, se tornou uma prática frequente a partir dos anos 2000. Leis foram criadas com o objetivo de tornar qualquer receita e gasto governamental público e de fácil acesso, como por exemplo, a lei 12.527[3] e a lei complementar 131[2]. Essa transparência tenta aproximar o cidadão de seu governo incentivando-o a participar dos processos de elaboração de planos, leis de diretrizes orçamentárias e orçamentos e, também, tentando minimizar fraudes e o superfaturamento de produtos, sistemas implementados ou projetos.

Todas informações devem ser providas de acordo com o decreto de 7.185[4]. Neste decreto, podemos enxergar a preocupação com a qualidade de divulgação, como visto no artigo 7º, o que deve estar contido e o quão pormenorizado deve ser esse dado. Pelo decreto, toda informação deve ser disponibilizada de modo consolidado, permitir armazenamento, importação e exportação e, também, possuir mecanismos de integridade, confiabilidade e disponibilidade. Já na lei 12.527[3], verificamos a atenção com o que deve ser exposto, onde e como. Conforme o item II do 3º parágrafo do Artigo 8 do capítulo 2, desta mesma lei, a informação disponibilizada deve *“possibilitar a gravação de relatórios em diversos formatos eletrônicos, inclusive abertos e não proprietários, tais como planilhas e texto, de modo a facilitar a análise das informações”*. Portanto, a escolha do formato de disponibilização da informação é tão importante quanto a informação em si, na questão de facilitar a análise dos dados e sua manipulação.

Em nossa instância, serão utilizados os dados providos pela Força Aérea Brasileira, referente aos voos realizados por autoridades políticas. Segundo o decreto presidencial 4.244[17] definido em seu primeiro parágrafo, somente as autoridades listadas a seguir podem requerer voos pela FAB,

“I - Vice-Presidente da República;

II - Presidentes do Senado Federal, da Câmara dos Deputados e do Supremo Tribunal Federal;

III - Ministros de Estado e demais ocupantes de cargo público com prerrogativas de Ministro de Estado;

~~IV – Comandantes das Forças Armadas.~~

IV - Comandantes das Forças Armadas e Chefe do Estado-Maior Conjunto das Forças Armadas. ”

Essas informações estão disponibilizadas no formato PDF, não atendendo ao item II do 3º parágrafo do Artigo 8 do capítulo 2 da Lei 12.527. Além disso, para realizar análises, sejam elas triviais ou não, seria extremamente custoso devido ao formato utilizado e a falta de mais informações, como, por exemplo, nome do ministro que realizou o voo e aeronave utilizada. Portanto, a facilidade para análises, em cima do formato e forma de divulgação no trecho destacado da lei 12.527, “*possibilitar a gravação de relatórios em diversos formatos eletrônicos, inclusive abertos e não proprietários, tais como planilhas e texto, de modo a facilitar a análise das informações*”, não é alcançada.

Segundo o Kit de Dados Abertos¹: A publicação de dados abertos é a nova tendência presente na estratégia de governo digital do Brasil e de diversos países do mundo. Seu compromisso foi assumido pela presidente Dilma no momento da assinatura da declaração² da Parceria para Governo Aberto (OGP³) e pelos diversos órgãos do executivo nos dois Planos de Ação da Parceria. Além disso, uma vantagem de aplicar esta prática é a criação de oportunidades na cobertura positiva de iniciativas de dados abertos e o impacto que estes dados possuem na sociedade.

Assim, surge a necessidade de transformar estes dados, enriquecê-los com informações externas, as quais a FAB não divulga, criar uma base consolidada a partir dos dados coletados e disponibilizá-la para que entre de acordo com as definições de dados abertos e, por fim, realizar análises em cima dessa base.

¹ <http://kit.dados.gov.br/>

² http://www.cgu.gov.br/governoaberto/a-ogp/o_que_e_Governo_Aberto.html

³ <http://www.opengovpartnership.org/>

1.2. Objetivos

Considerando-se a dificuldade em realizar análises em cima do formato dos arquivos disponibilizados pela FAB, este trabalho de conclusão de curso tem como principal objetivo a criação e utilização de um processo em Sistemas de Informação, o qual crie e utilize ferramentas, que possibilitem e auxiliem na extração, manipulação e disponibilização de uma base unificada de dados que auxilie em análises.

1.3. Organização do texto

O presente trabalho está estruturado em capítulos e, além desta introdução, será desenvolvido da seguinte forma:

- Capítulo II: Introdução e descrição das ferramentas utilizadas ao longo do projeto. Como algumas ferramentas, não são de conhecimento comum, foi identificada a necessidade de descrevê-las para que o leitor possa entender seu objetivo, e porque elas foram utilizadas.
- Capítulo III: Aborda as etapas realizadas na execução do projeto.
- Capítulo IV: Destinado a análises e resultados dos dados coletados.
- Capítulo V: Aborda a conclusão, reunindo as considerações finais, assinalando as contribuições da base criada e sugere possibilidades de aprofundamento posterior.

2. Ferramentas

Neste capítulo serão abordadas as ferramentas utilizadas no processo, que não são de conhecimento comum, e que foram bastante importantes para o desenvolvimento do projeto. As ferramentas utilizadas no processo são descritas de acordo com seu objetivo, sendo elas a Bonita BPM da BonitaSoft, ParseHub, BigQuery do Goole e Tableau da Tableau Software. Conforme a *Figura 1* demonstrando a integração das ferramentas com as etapas do processo.

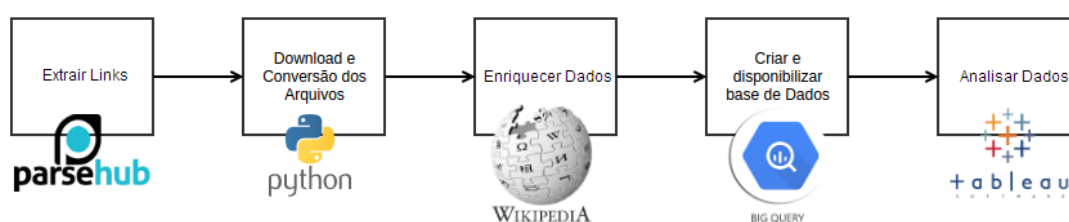


Figura 1 - Diagrama da Integração das Ferramentas com as Etapas do Processo

2.1. ParseHub

Em busca de minimizar o tempo de desenvolvimento na extração de dados, foram analisadas duas ferramentas que auxiliam nessa tarefa: ParseHub⁴ e Scrapy⁵.

O ParseHub possui um software para PC com uma interface própria, permitindo identificar os elementos a serem extraídos de um site utilizando somente o mouse e um pseudocódigo, obtendo o resultado em JSON ou CSV. A outra abordagem seria utilizar o Scrapy, onde é necessária a criação de scripts que extraíam programaticamente os dados do site desejado. Dessa forma, o ParseHub atende melhor as necessidades do projeto, pela facilidade de uso, tornando-o uma boa escolha.

⁴ <https://www.parsehub.com/>

⁵ <https://scrapy.org/>

Uma das fundadoras do ParseHub, Angelina Fomina, publicou um post sobre algumas possibilidades da ferramenta[6]. Destacamos o seguinte trecho, pois foi com esse objetivo que esta ferramenta foi utilizada no projeto, *“Developing software that automates the monotonous task of gathering data, is the next step that will give data scientists more freedom to focus on creative aspects of their job. ‘Data scientists...spend from 50 percent to 80 percent of their time mired in this more mundane labor of collecting and preparing unruly digital data, before it can be explored for useful nuggets’ - Steve Lohr, for New York Times.”* Conforme apontado pela referência usada por Fomina, grande parte do tempo utilizado em projetos, que necessitam de mineração de dados, é gasto na extração.

Fomina também exalta em seu post o seguinte aspecto, *“At the end of the day it is less about the data and more about what you do with it. Let’s speed through the data aggregation process, now, so you can go on to explore the results.”*. Ou seja, o ParseHub busca diminuir o tempo necessário para extração dos dados, com o objetivo de enriquecer as análises realizadas em cima deles.

Exemplo: O objetivo é obter um formato de fácil manipulação e alguns destes formatos, são: JSON e XML. No exemplo citado por Fomina, a página⁶ contém as informações da distribuição da população dos Estados Unidos da América em função das cidades. Como podemos ver na *Figura 2*, abaixo, a tabela em sua página possui as seguintes informações:

- 1- O rank das cidades em relação à população, conforme estimativa da Agência de Censo⁷, realizada em 1º de julho de 2015;
- 2- O nome da cidade;
- 3- Estado;
- 4- Estimativa da população realizada pela Agência de Censo, em 1º de julho de 2015;
- 5- A população da cidade, enumerada pelo Censo Americano⁸, em 1º de abril de 2010;

⁶ https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population

⁷ <https://www.census.gov/>

⁸ <http://www.census.gov/2010census/>

- 6- Percentual da mudança populacional entre abril de 2010 e julho de 2015;
- 7- Área da cidade;
- 8- Densidade da população, conforme censo realizado em julho de 2015 (Residentes por unidade de área);

Ainda na *Figura 2*, podemos verificar a caixa contendo as funções criadas, no canto direito, as quais resultarão no JSON, já este, se encontra no canto inferior esquerdo.

2014 rank	City	State ^[5]	2014 estimate	2010 Census	Change	2014 land area	2010 population density
1	New York City ^[6]	New York	8,491,079	8,175,133	+3.86%	302.6 sq mi 783.8 km ²	27,012 per sq mi 10,430 km ⁻²
2	Los Angeles	California	3,928,864	3,792,621	+3.59%	468.7 sq mi 1,213.9 km ²	8,092 per sq mi 3,124 km ⁻²
3	Chicago	Illinois	2,722,389	2,695,598	+0.99%	227.6 sq mi 589.6 km ²	11,842 per sq mi 4,572 km ⁻²
4	Houston ^[7]	Texas	2,239,558	2,100,263	+6.63%	599.6 sq mi 1,552.9 km ²	3,501 per sq mi 1,352 km ⁻²
5	Philadelphia ^[8]	Pennsylvania	1,560,297	1,526,006	+2.25%	134.1 sq mi 347.3 km ²	11,379 per sq mi 4,394 km ⁻²
6	Phoenix	Arizona	1,537,058	1,445,632	+6.32%	516.7 sq mi 1,338.3 km ²	2,798 per sq mi 1,080 km ⁻²
7	San Antonio	Texas	1,436,697	1,327,407	+8.23%	460.9 sq mi	2,880 per sq mi

```

{
  "cities": [
    {
      "city": "New York City",
      "state": "New York",
      "location": "New York City, New York"
    },
    {
      "city": "Los Angeles",
      "state": "California",
      "location": "Los Angeles, California"
    },
    {
      "city": "Chicago",
      "state": "Illinois",
      "location": "Chicago, Illinois"
    },
    {
      "city": "Houston",
      "state": "Texas",
      "location": "Houston, Texas"
    },
    {
      "city": "Philadelphia",
      "state": "Pennsylvania",
      "location": "Philadelphia, Pennsylvania"
    },
    {
      "city": "Phoenix",
      "state": "Arizona",
      "location": "Phoenix, Arizona"
    },
    {
      "city": "San Antonio",
      "state": "Texas",
      "location": "San Antonio, Texas"
    }
  ]
}

```

Figura 2 - Exemplo de projeto no ParseHub

2.2. Bonita Studio / BPM Engine – Bonitasoft

O Bonita Studio⁹ é uma ferramenta que permite ao usuário modificar graficamente os processos de negócios utilizando o padrão BPMN¹⁰. O usuário desta ferramenta pode, também, conectar os processos a outras partes do sistema. O objetivo do Bonita Studio é gerar um software empresarial autônomo e acessível como formulário web. O Bonita Studio pode, também, utilizar sua BPM Engine para executar definições de processos.

⁹ <http://www.bonitasoft.com/>

¹⁰ <http://www.bpmn.org/>

Através do Bonita Studio, o usuário pode, através de meios de desenvolvimento, interagir com seu processo.

Em nossa instância, o Bonita Studio foi utilizado somente para design do processo de negócio. Como a ferramenta já possui suporte ao padrão BPMN e possuía-se conhecimentos prévios sobre suas funcionalidades, ela foi utilizada para facilitar e criar os processos dentro do padrão BPMN.

Exemplo: Na *Figura 3* podemos ver um exemplo da definição de um processo simples no Bonita Studio, utilizando o padrão BPMN. Neste processo, estão definidos uma requisição de férias, feita por um empregado, a validação do pedido e sua aprovação pelo gerente de RH, a notificação, caso seja aprovada e seu término.

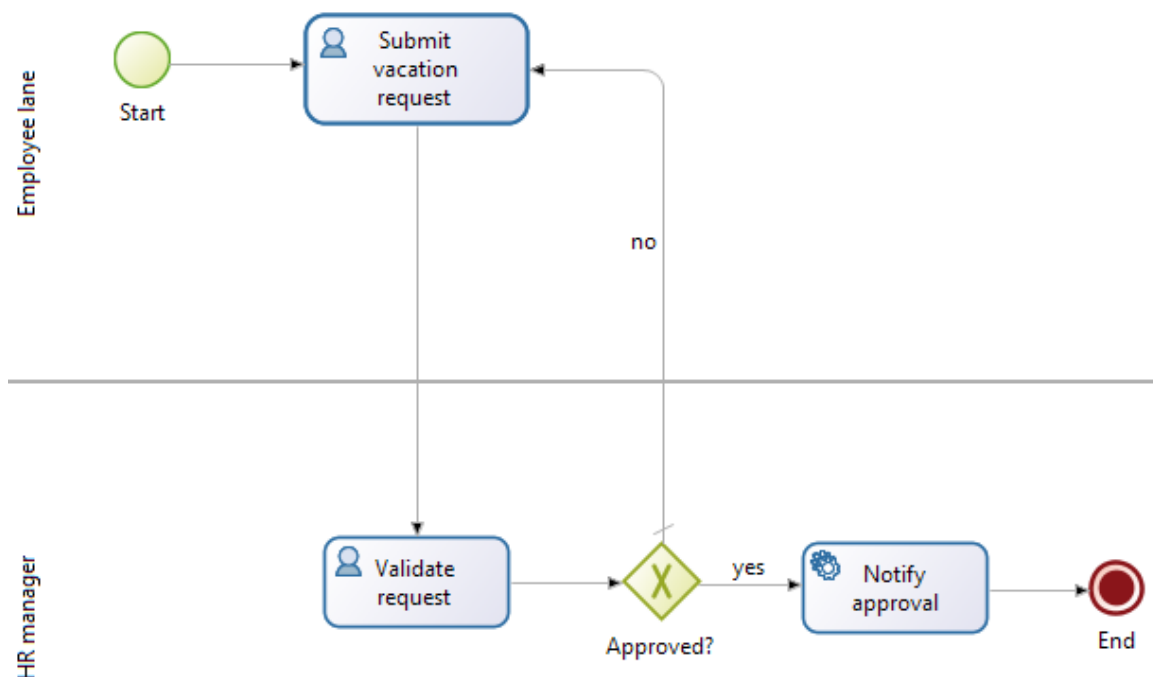


Figura 3 - Exemplo de Definição de Processo no Bonita Studio

2.3. Google BigQuery

O Google BigQuery¹¹ visa enfrentar um dos maiores problemas de lidar com grandes massas de dados, que é a necessidade de manter e monitorar a infraestrutura utilizada para guardar e consultar essas informações, sem contar com os custos financeiros e o tempo realizado nessas tarefas. O Google BigQuery foi escolhido por ser uma solução que utiliza a infraestrutura do Google, sem necessidade de configuração dos servidores, fazendo assim com que o usuário só precise se preocupar em gerar e consultar os seus dados. O BigQuery controla todos os aspectos técnicos de armazenagem dos dados, incluindo compressão, criptografia, replicação, ajustes de desempenho e escalonamento.

Esta ferramenta foi escolhida pelos também pelo fato de ser possível compartilhar a base de dados, tornando-a pública de maneira fácil. Com o compartilhamento, qualquer pessoa pode ter acesso e executar as consultas que desejar, criando outras visualizações ou até mesmo confirmar análises do projeto.

2.4. Tableau

A visualização de dados é, simplesmente, o processo de descrever informações através de imagens. A visualização permite o entendimento universal e imediato dos dados, se aproveitando do poderoso sistema de processamento visual do ser humano.

Os avanços da tecnologia ajudaram a deixar a visualização de dados mais prevalente e eficaz do que antigamente, aumentando a sua importância na inteligência de negócios.

No processo de análise de dados, foram analisadas algumas soluções, como o Redash¹², o Google Data Studio¹³ e o Tableau¹⁴. A busca foi por uma ferramenta que possuísse uma fácil integração com fontes de dados e que conseguisse gerar gráficos de

¹¹ <https://cloud.google.com/bigquery/>

¹² <https://redash.io/>

¹³ <https://datastudio.google.com/>

¹⁴ <http://www.tableau.com/>

3. Execução do Trabalho

Neste capítulo, será abordada a execução da ideia e dos conhecimentos apresentados nos capítulos 1 e 2. Ao final deste capítulo, o indivíduo terá conhecimento suficiente para executar todos os conceitos apresentados previamente e, também, obter um resultado aplicado em um caso real.

3. 1. Escopo do Trabalho

Este trabalho tem como principal meta disponibilizar uma base para que seja possível efetuar análises estatísticas em cima dos voos realizados pela FAB para transporte de entidades governamentais. O processo irá extrair estas informações, disponibilizadas em PDF no próprio site da FAB, referentes aos anos de 2013 a 2016, transformar os arquivos em XML, enriquecer os dados dos voos, definir e criar o modelo da base a ser utilizado, disponibilizá-la e, por fim, realizar análises em cima deste conjunto. Conforme descrito, este procedimento foi dividido em 5 processos, dentre eles 2 macroprocessos, conforme a *Figura 5*:

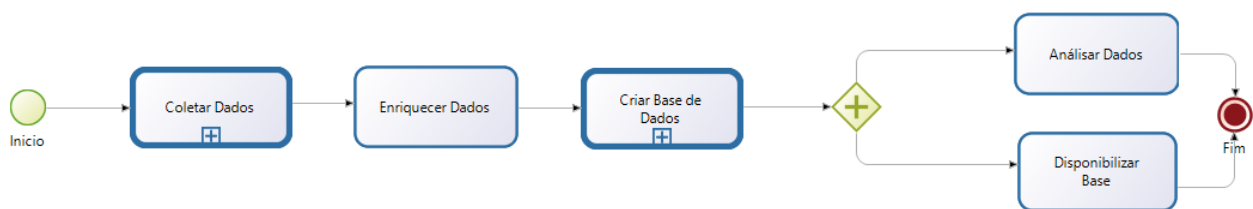


Figura 5 - Definição do Processo

Estes processos serão detalhados conforme as sessões a seguir.

3.2. Coletar Dados

Tendo como objetivo a extração de dados, foi utilizado o ParseHub, citado no Capítulo 2. Através do projeto criado nele conseguimos extrair os links para todos os PDFs contidos no site da FAB. Os links disponibilizados pela FAB são referentes ao período de 13 de junho de 2013 até a data corrente.

Neste trabalho, o conjunto de dados utilizado refere-se ao período de 13 de junho de 2013 até 10 de novembro de 2016. Todo o processo abordado nesta seção é realizado de forma automatizada. Entretanto, houve a necessidade de escolher uma data limite, pois o processo de enriquecimento é realizado de forma manual devido à ausência de informações necessárias para a análise dos PDFs, dificultando a automatização total do processo. Os subprocessos da coleta de dados, conforme a *Figura 6*, serão detalhados nos tópicos a seguir.

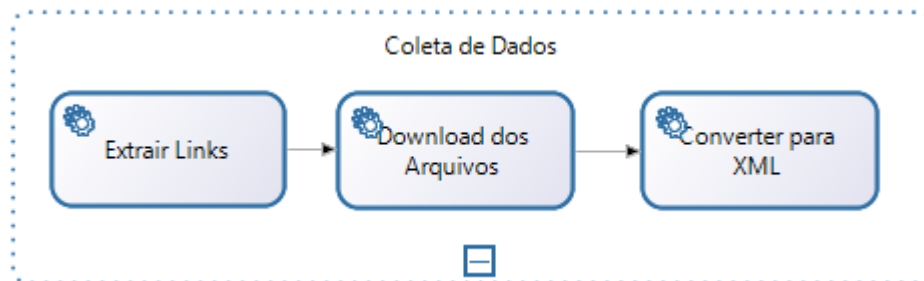


Figura 6 - Especificação do Macroprocesso de Coleta de Dados

No PDF disponibilizado pela FAB pode-se extrair as seguintes informações sobre os voos: Autoridades Apoiadas, Origem do Voo, Hora local da decolagem, Destino do Voo, Hora local do pouso, motivo e previsão de passageiros, conforme a *Figura 7*.

**MINISTÉRIO DA DEFESA
COMANDO DA AERONÁUTICA
GABINETE DO COMANDANTE DA AERONÁUTICA**

**VIAGENS CONCLUÍDAS
31 DE OUTUBRO DE 2016**

31 DE OUTUBRO DE 2016

AUTORIDADES APOIADAS	ORIGEM	DECOLAGEM (H. LOCAL)	DESTINO	POUSO (H. LOCAL)	MOTIVO	PREVISÃO DE PASSAGEIROS
Presidente da Câmara dos Deputados <i>Em Exercício</i>	São Luís	31/10/2016 - 09:10	Brasília	31/10/2016 - 12:30	Serviço	9
Ministro-chefe da Casa Civil da Presidência da República	Porto Alegre	31/10/2016 - 07:00	Brasília	31/10/2016 - 09:10	Segurança	1
Ministro da Justiça e Cidadania	São Paulo (Congonhas)	31/10/2016 - 13:15	Brasília	31/10/2016 - 14:35	Serviço	5
Ministro do Desenvolvimento Social e Agrário	Brasília	31/10/2016 - 06:20	Caxias do Sul	31/10/2016 - 08:35	Serviço	8
Ministro do Desenvolvimento Social e Agrário	Caxias do Sul	31/10/2016 - 15:20	Brasília	31/10/2016 - 17:20	Serviço	5
Ministro das Cidades	Recife	31/10/2016 - 08:45	João Pessoa	31/10/2016 - 09:10	Serviço	10
Ministro das Cidades	João Pessoa	31/10/2016 - 13:25	Campina Grande	31/10/2016 - 13:45	Serviço	10
Ministro das Cidades	Campina Grande	31/10/2016 - 16:40	Brasília	31/10/2016 - 20:10	Serviço	10

Figura 7 - Informações contidas no PDF da Força Aérea Brasileira

3.2.1. Extrair Links

Nesta etapa, utilizamos o ParseHub para gerar o JSON dos links referente aos arquivos. O primeiro passo é a criação do projeto na ferramenta. Em nossa instância, foram criados 4 projetos, cada um referente a um ano. Cada projeto possui duas seções, sendo uma referente à seleção e outra à extração, conforme a *Figura 8* e *Figura 9* abaixo:

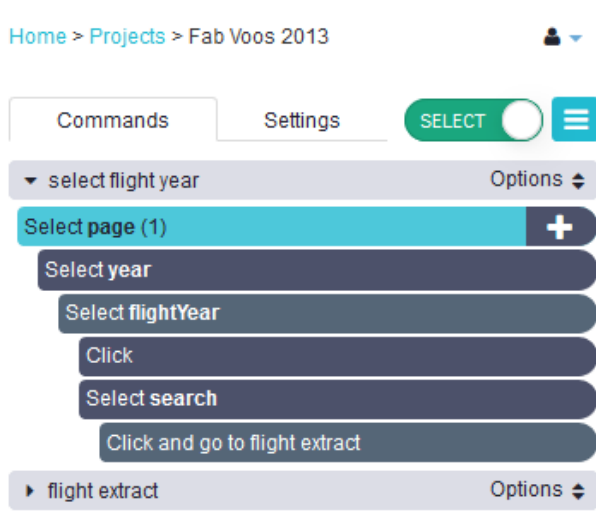


Figura 9 - Definição da seção “select flight year” no ParseHub

1st body

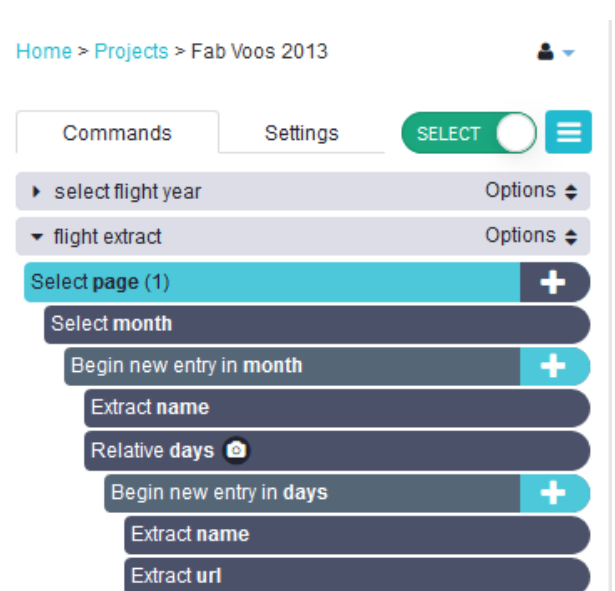


Figura 8 – Definição da seção “flight extract” no ParseHub

1st body

Na seção “select flight year”, utilizamos os módulos da ferramenta para selecionarmos o ano requerido, no caso do exemplo 2013, conforme a *Figura 11*, abaixo. Assim, o script irá selecionar o ano na área de busca e clicar no botão “pesquisar”, concluindo sua seção e chamando a seção “flight extract”.

REGISTRO DE VOOS

Ano 2013

PESQUISAR

Janeiro

D	S	T	Q	Q	S	S
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26

Fevereiro

D	S	T	Q	Q	S	S
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23

Março

D	S	T	Q	Q	S	S
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23

Figura 11 – Exemplo de seleção no site da FAB

Na execução da “flight extract”, o script irá selecionar o objeto “mês” e extrair seu nome, em seguida, irá selecionar todos os objetos “dia”, do mês selecionado, e extrair seu nome e URL que referência ao link do arquivo PDF daquele dia. Esta etapa será executada recursivamente, até que todos objetos “mês” daquele ano sejam coletados. Ao final da execução pode-se obter o resultado como JSON, conforme a *Figura 10*:

```
{
  "month": [
    {
      "days": [
        {
          "name": "5",
          "url": "http://www.fab.mil.br/cabine/voos/20150105_182310.pdf"
        },
        {
          "name": "30",
          "url": "http://www.fab.mil.br/cabine/voos/20150130_134152.pdf"
        }
      ],
      "name": "Janeiro"
    }
  ]
}
```

Figura 10 – Exemplo do resultado final gerado pelo ParseHub em JSON

3.2.2. Download dos Arquivos

Após a disponibilização dos links no formato JSON, criamos um script em Python que integra a API do ParseHub, a qual obtém JSON gerado pela ferramenta, e realiza o download de cada PDF.

O script em Python, separa cada PDF na seguinte estrutura de pastas, de acordo com o exemplo da *Figura 12*:

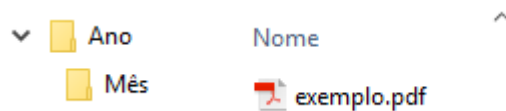


Figura 12 – Exemplo da estrutura de arquivos gerada pelo script

Ou seja, ao final da extração teremos 4 pastas referentes aos anos de 2013 a 2016, cada ano contendo pastas referentes aos seus meses e, finalmente, os arquivos com as informações dos voos no formato PDF.

3.2.3. Converter os Arquivos para XML

Nesta etapa, é realizada a conversão pela ferramenta “pdftohtml”, encontrada em algumas distribuições do Linux. Este passo pode ser realizado de diversas maneiras, não há obrigatoriedade de utilização desta ferramenta. A escolha dela foi baseada em sua qualidade de conversão, na qual o XML gerado é a mais fiel ao arquivo PDF.

O script de conversão realiza uma busca nas pastas criadas, na etapa anterior, pelos arquivos com terminação “.pdf”, e realiza a chamada para a “pdftohtml” com os seguintes parâmetros, conforme a *Figura 13*:

```
command = ['pdftohtml', '-nomerge', '-xml', '-s', '-noframes', pdfPath, outputFile]
```

Figura 13 – Lista de parâmetros do “pdftohtml”

3.3. Enriquecer Dados

Ao analisarmos as informações disponibilizadas pelo PDF da FAB verificamos a falta de algumas informações, como por exemplo: nome do ministro e seu partido, para que fosse possível ligar um voo a uma pessoa ou seu partido. Com isso, foi identificada a necessidade de busca por outras fontes de dados.

3.3.1. Pesquisa de Fontes

Houve grande dificuldade para encontrar uma fonte oficial de dados que pudesse agregar as informações já coletadas. A busca pelas entidades por trás de cada gabinete, em um determinado período, nos sites oficiais do governo, não foi possível, pois esta informação não se relacionava ao tempo de vigência. Em conjunto, a falta de padronização, destes sites, se configurou como um desafio para a criação de uma extração automática, de forma similar ao script de extração dos voos da FAB.

Assim, a busca por estes dados se deu de forma manual, visto que o custo de tempo da geração de uma ferramenta, ou sistema, que pudesse auxiliar nesta extração era muito alto.

3.3.2. Dados

Os dados coletados foram todos extraídos do site Wikipedia¹⁵ [7]. As informações adicionadas a base, foram:

- Nome do Ministro;
- Vigência do Ministro;
- Partido do Ministro;

¹⁵ https://pt.wikipedia.org/wiki/Ministérios_do_Brasil

3.4. Criar Base de Dados

Este processo, assim como a coleta de dados, foi dividido em subprocessos. Os quais estão definidos na *Figura 14*.

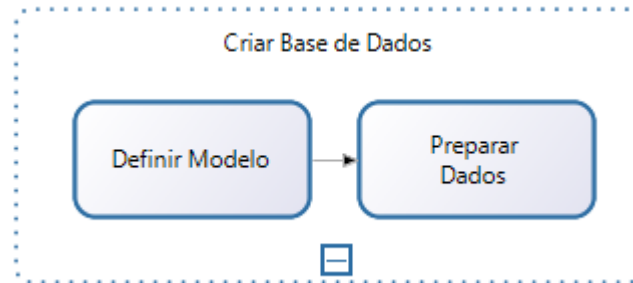


Figura 14 - Especificação dos subprocessos de Criar Base de dados

3.4.1. Definir Modelo

O primeiro passo foi a definição e do modelo de dados que será utilizado por nossa base. O modelo foi definido a partir dos seguintes objetos:

- Gabinete - Ofício ao qual a autoridade foi nomeada ou ocupa;
- Autoridade - Pessoa que ocupa ou ocupou um cargo;
- Voo - Informações disponibilizadas pela FAB sobre os voos;
- Voo_Passageiro - Tabela de associação entre o voo, passageiro e o gabinete responsável
- Passageiro - Entidade que estava presente no voo

A base gerada; suas tabelas; atributos e relações, foram definidos conforme o modelo lógico da *Figura 15*:

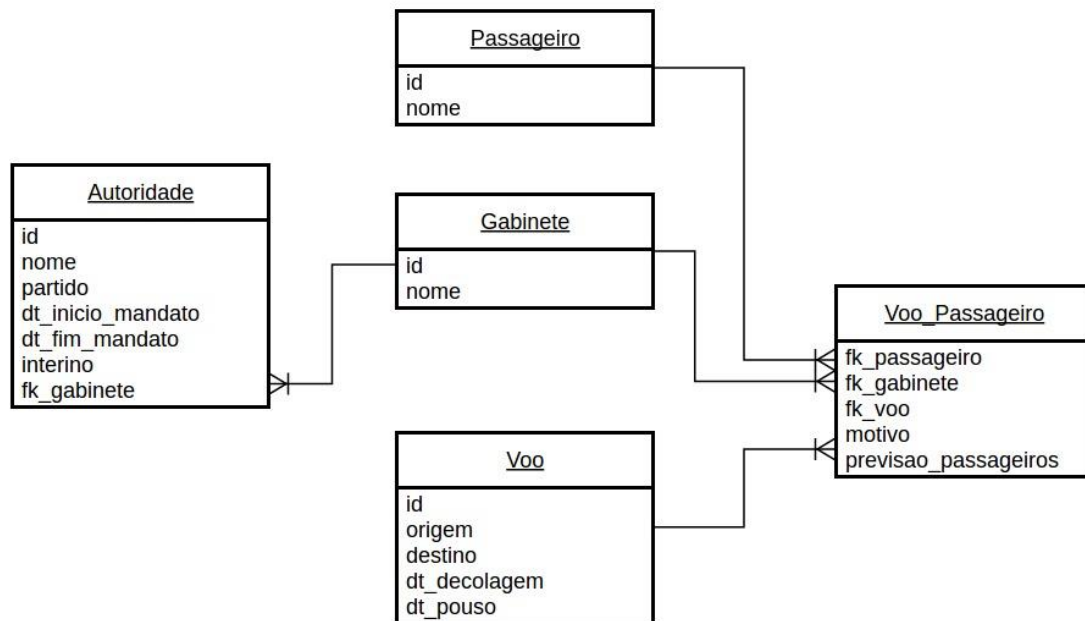


Figura 15 - Modelo lógico entidade-relacionamento da base de dados

3.5. Definição Formal dos Atributos

Os atributos das entidades foram definidos conforme as tabelas abaixo:

Tabela 1 - Passageiro

Campo	Tipo	Descrição
passageiro_id	int	chave primária da tabela, contador
nome	varchar(255)	nome da autoridade apoiada, retirada do PDF

Tabela 2 - Autoridade

Campo	Tipo	Descrição
autoridade_id	int	chave primária da tabela, contador
nome	varchar(255)	nome da autoridade do gabinete, por período
partido	varchar(255)	partido da autoridade do gabinete
dt_inicio_mandato	datetime	data do inicio do mandato da autoridade
dt_fim_mandato	datetime	data do termino do mandato. nulo se o mandato for atual
interino	boolean	identifica se a autoridade é interina
fk_gabinete_id	int	id do gabinete da autoridade

Tabela 3 - Gabinete

Campo	Tipo	Descrição
gabinete_id	int	chave primária da tabela, contador
nome	varchar(255)	nome do gabinete da autoridade apoiada

Tabela 4 - Voo_Passageiro

Campo	Tipo	Descrição
fk_passageiro_id	int	id do passageiro desse voo
fk_gabinete_id	int	id do gabinete que requereu o voo
fk_voo_id	int	id do voo
motivo	varchar(255)	motivo da viagem
previsao_passageiros	int	previsao de passageiros

Tabela 5 - Voo

Campo	Tipo	Descrição
voo_id	int	chave primária da tabela, contador

origem	varchar(255)	cidade de origem do voo
destino	varchar(255)	cidade de destino do voo
dt_decolagem	datetime	data da decolagem do voo
dt_pouso	datetime	data do pouso do voo

3.5.1. Preparar Dados

O próximo passo executado após a coleta da base, seu enriquecimento e a definição do modelo, é a preparação dos dados. Varremos a base de dados procurando por inconsistências, erros de digitação, erros de preenchimento ou desvio do padrão adotado. Esse passo é de suma importância, pois caso esses erros fossem computados por nossas análises, poderíamos obter respostas imprecisas, o que poderia acarretar em conclusões equivocadas.

Como os dados informados pela FAB, em sua maioria não são sistemáticos, ou seja, dependem de preenchimento humano, foram encontradas inúmeras inconsistências, erros de digitação, erros de preenchimento e desvio do padrão adotado. Alguns exemplos foram:

- Cargo/Função descrito com definição de gênero – O padrão adotado é que todo e qualquer cargo deve aparecer sem referência de gênero. Entretanto, podemos encontrar exemplos de ministros e entidades com seus cargos/ocupações descritas com tais definições. **Exemplo:** Os seguintes dados foram alterados de “*Ministra-chefe da Secretaria de Direitos Humanos da PR*” e “*Ministro-chefe da Secretaria de Direitos Humanos da PR*” para “*Secretaria de Direitos Humanos*”; “*Presidente da Câmara de Deputados*” para “*Presidência da Câmara de Deputados*”.
- Não padronização na definição de interino – No conjunto, a definição de interino não está padronizada. **Exemplo:** Ministério do Desenvolvimento Agrário (Interino) e Ministério da Saúde – Interino.

Neste caso, foi utilizado um atributo boolean, na tabela “autoridade”, para definir se ela era interina ou não.

- Coluna excedente – No arquivo referente ao ano de 2013, é possível constatar uma coluna a mais, diferente dos outros anos. Esta coluna foi removida, pois apresenta-se como uma inconsistência ao formato.
- Inconsistência de nomes dos Ministérios – Foram observadas algumas inconsistências nos nomes dos Ministérios. **Exemplo:** Ministério do Desenvolvimento da Indústria e Comércio Exterior
O nome deste Ministério está incorreto, o correto, segundo o site oficial do ministério é “Ministério do Desenvolvimento, Indústria e Comércio Exterior”. Este Ministério posteriormente, no governo do Pres. Michel Temer, teve seu nome alterado para “Ministério da Indústria, Comércio Exterior e Serviços”
- Abreviações indevidas e inconsistentes – Neste caso, na entrada de nome das autoridades, foram adotadas abreviações desnecessárias e não identificadas em nenhum outro dado, portanto, resolvemos removendo as abreviações ou as expandindo. **Exemplo:** Abreviação de Deputado para “Deput”; Abreviação de “Forças Armadas” para “FA”; Abreviação de “Presidência da República” para “PR”; Abreviação de “Gabinete” para “Gab.”
- Definição de “em exercício” e “afastado” sem padronização – A definição da autoridade em exercício ou afastada é utilizada duas vezes nos dados. A fim de facilitar e padronizar, esta informação foi removida e foi utilizado o campo “dt_início_mandato”, referente a data de início do mandato e “dt_fim_mandato”, referente a data na qual a autoridade deixou sua função.

3.6. Disponibilizar Dados

Os conjuntos de dados do projeto foram disponibilizados no Google BigQuery, devido a facilidade que seu serviço fornece em compartilhar massas de dados e à ampla integração com outros serviços e sistemas que auxiliam na análise de dados.

A forma de compartilhar dados no BigQuery é feita de forma bem simples, apenas selecionando que se deseja compartilhar o conjunto de dados com *"All Authenticated Users"*, permitindo que o endereço ou nome do projeto se torne importável. Conforme a *Figura 16* e *Figura 17*.

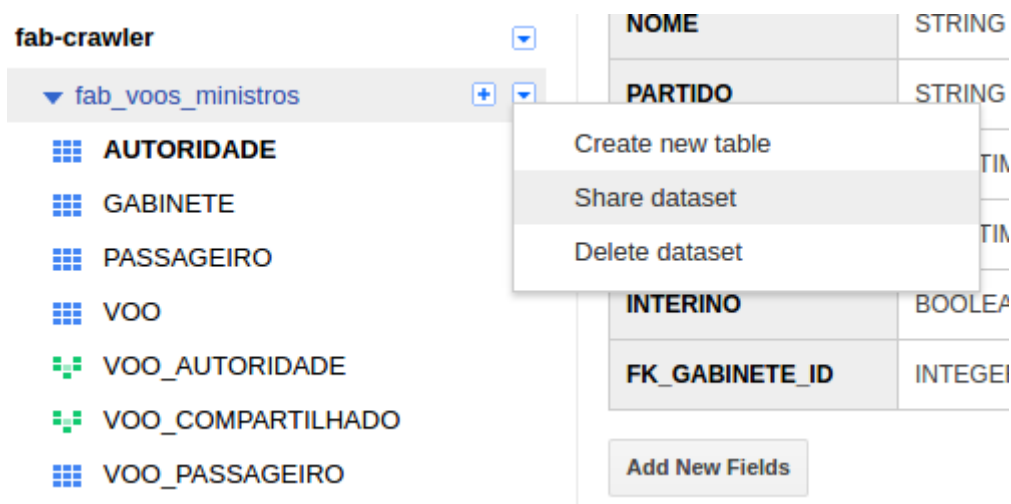


Figura 16 - Passo 1 da Disponibilização pelo BigQuery

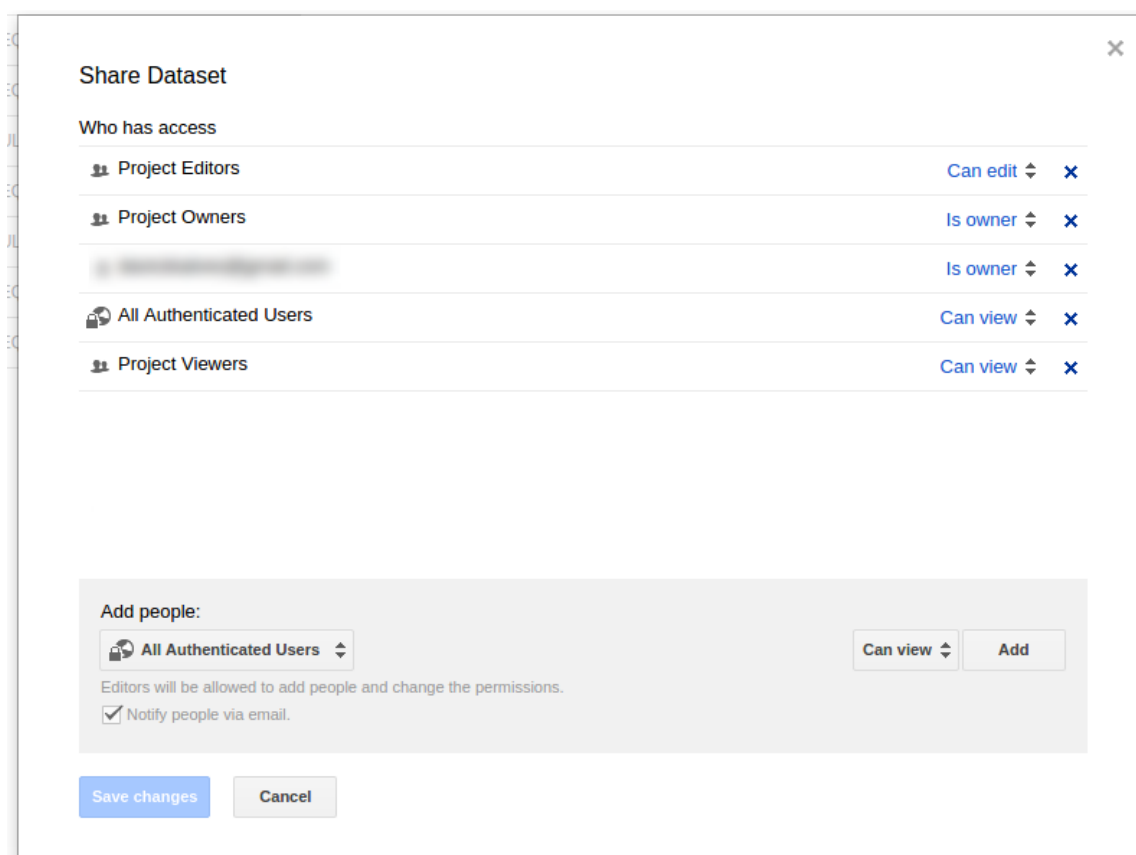


Figura 17 - Passo 2 da Disponibilização pelo BigQuery

Para acessar os dados compartilhados, é necessário ter uma conta no Google Cloud e criar um projeto, porém sem necessidade de pagamento, lembrando que o BigQuery permite a execução de consultas de até 1TB por mês de graça.

Após criar um projeto e adicionar a API do Google BigQuery ao projeto, será necessário acessar a URL¹⁶ e adicionar o conjunto de dados compartilhado.

3.7. Importar Base

Esta etapa não faz parte do processo, entretanto foi julgado importante descrever essa seção para informar a possibilidade de importação da base já consolidada neste projeto.

3.7.1. Instruções

A documentação do BigQuery descreve essa etapa da seguinte forma: é possível que um usuário compartilhe um conjunto de dados, porém sem adicionar pessoas ao projeto. Assim, para utilizar esse conjunto de dados, é necessário adicioná-lo manualmente em sua conta.

Para importar os dados, deve-se selecionar a opção de *"Display Project"*, e inserir *"fab-crawler"* nome do projeto no *"Project Id"*, conforme a *Figura 18* e *Figura 19* abaixo. Após esse processo, será possível executar query's em cima do conjunto de dados do projeto.

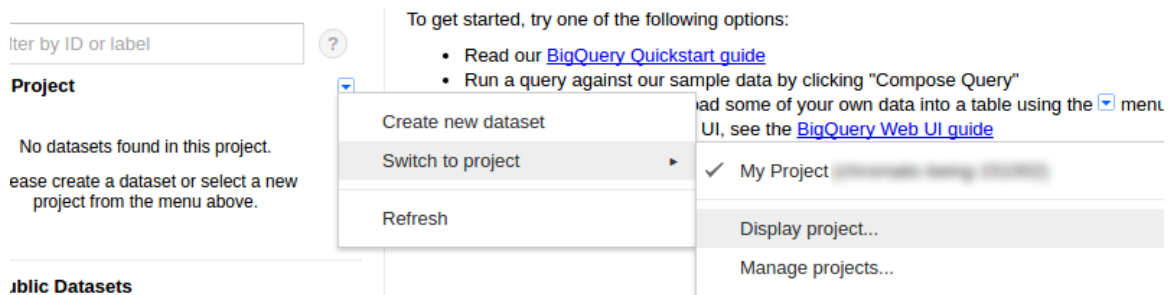


Figura 18 - Passo 1 da Importação de Base no BigQuery

¹⁶ <http://bigquery.cloud.google.com/>

×

Display Project

Project ID

fab-crawler

?

☒ Display project in navigation panel

☐ Make this my current project

OK

Cancel

Figura 19 - Passo 2 da Importação de Base no BigQuery

4. Análise

Neste Capítulo, os dados coletados e consolidados no Capítulo anterior serão abordados e analisados com o objetivo de apresentar e discutir resultados.

A principal busca é conseguir identificar instâncias que desviam do padrão encontrado na base, como por exemplo uma autoridade com números de voos acima da média em um período. Foi gerada a hipótese de que quanto maior o período de vigência, maior será a quantidade de voos realizados pela autoridade, esta hipótese será respondida ao longo deste capítulo.

4. 1. Relação Quantidade de Voos por Período

Nesta análise, serão relacionados os períodos com o número de voos em cada. No gráfico da *Figura 21*, é possível verificar que o ano de 2014 foi o ano com maior número de voos realizados, em contrapartida, o ano de 2013 possui o menor número, entretanto, este ano possui somente 6 meses. Assim, realizamos a média ponderada, levando em consideração o número de meses em cada ano, sendo 2013 com peso 6, 2014 e 2015 com peso 12 e 2016 com peso 11. Apesar desta medida, o resultado poderia causar estranheza, visto o aumento significativo de 2013 para 2014, portanto foi optado por cortar o ano de 2013 desta análise. Com esta restrição, foi gerado o gráfico da *Figura 20*, onde pode-se ter uma melhor noção da progressão do número de voos realizados ao ano.

Conforme a *Figura 20*, é possível inferir que houve uma diminuição do número de voos ao longo dos anos, sendo 2014, o ano com o maior número, contendo 3.467 voos e o ano de 2016, até o corrente mês de novembro, com 1.861 voos, bem abaixo da média estipulada de 2.688 voos ao ano.

Voos por Período

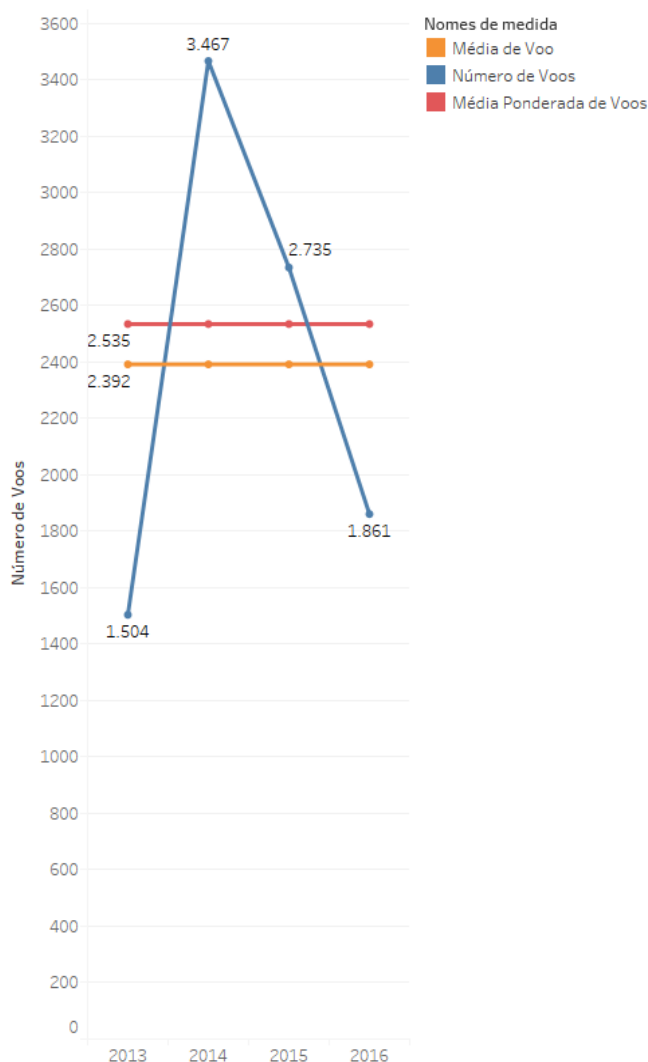


Figura 21 - Gráfico de Linha - Número de Voos por Ano e sua Média

Voos por Período

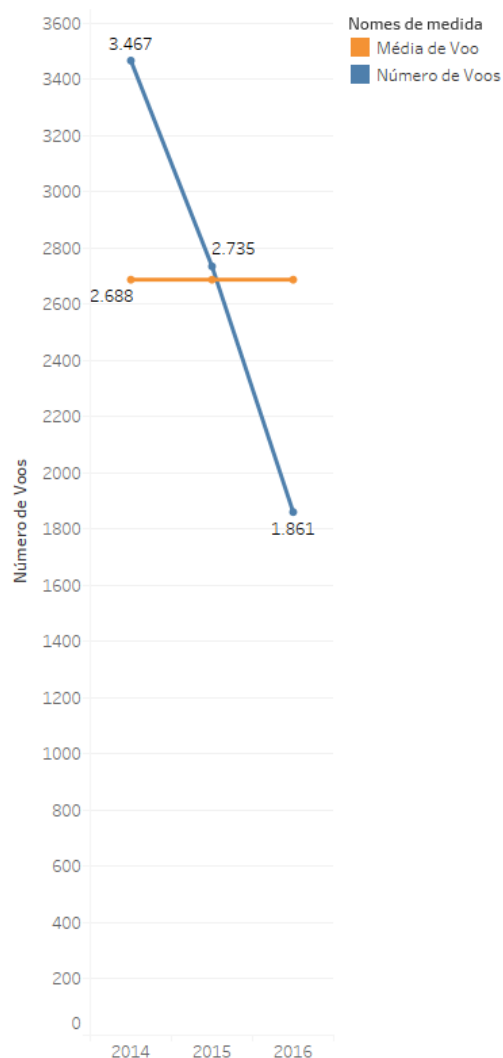


Figura 20 - Gráfico de Linha - Número de Voos por Ano e sua Média - Sem 2013

4. 2. Relação de Voo por Autoridade

Nesta seção, foram gerados gráficos, os quais foram obtidos a partir do agrupamento de voos por autoridade no período. O período definido foi anual, resultando em 5 gráficos, os quais correspondem ao segundo semestre de 2013, aos anos de 2014, 2015 e ao ano de 2016, este último que foi dividido em 2 gráficos, pois houve a mudança de Presidente devido ao Impeachment.

A relação Autoridade-Voo não leva em consideração o tempo de vigência. O gráfico utilizado é um gráfico de barra, com uma linha alaranjada destacando a média do período.

Ao longo do período, julho de 2013 a novembro de 2016, foi identificada uma média de 43 voos por autoridade ao ano, ou seja, estima-se que uma autoridade realize, em média 43 viagens utilizando aeronaves da FAB no período de um ano.

4. 2. 1. Boxplot de Voo por Autoridade ao Ano

Antes de serem retirados os dados discriminados, foi extraído um boxplot, conforme a *Figura 22*, para verificar se haviam pontos fora da curva. Estes pontos foram identificados, para posteriormente serem analisados especificamente. São eles:

- **Referente ao ano de 2015**

A média deste ano foi de 53 voos por Autoridade

- Aldo Rebelo
161 Voos
- Gilberto Kassab
223 Voos

- **Referente ao ano de 2016**

A média deste ano foi de 27 voos por Autoridade

- Gilberto Kassab
108 Voos
- Alexandre de Moraes
96 Voos

Boxplot - Voo x Autoridade - Anual

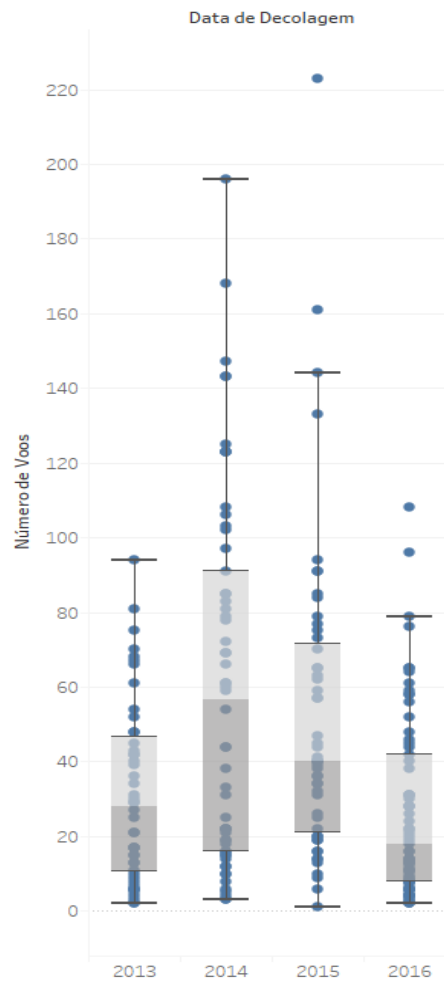


Figura 22 – Gráfico de Boxplot – Voo por Autoridade ao Ano

4. 2. 2. Número de Voos por Autoridade de julho a dezembro de 2013

Neste período, a média de voo encontrada foi de 31 voos por autoridade. Apesar do período restrito, pois as informações divulgadas pela FAB iniciaram somente em julho de 2013, totalizando 6 meses para este ano, esta análise será realizada devido a quantidade limitada de dados obtidos.

Voo por Autoridade

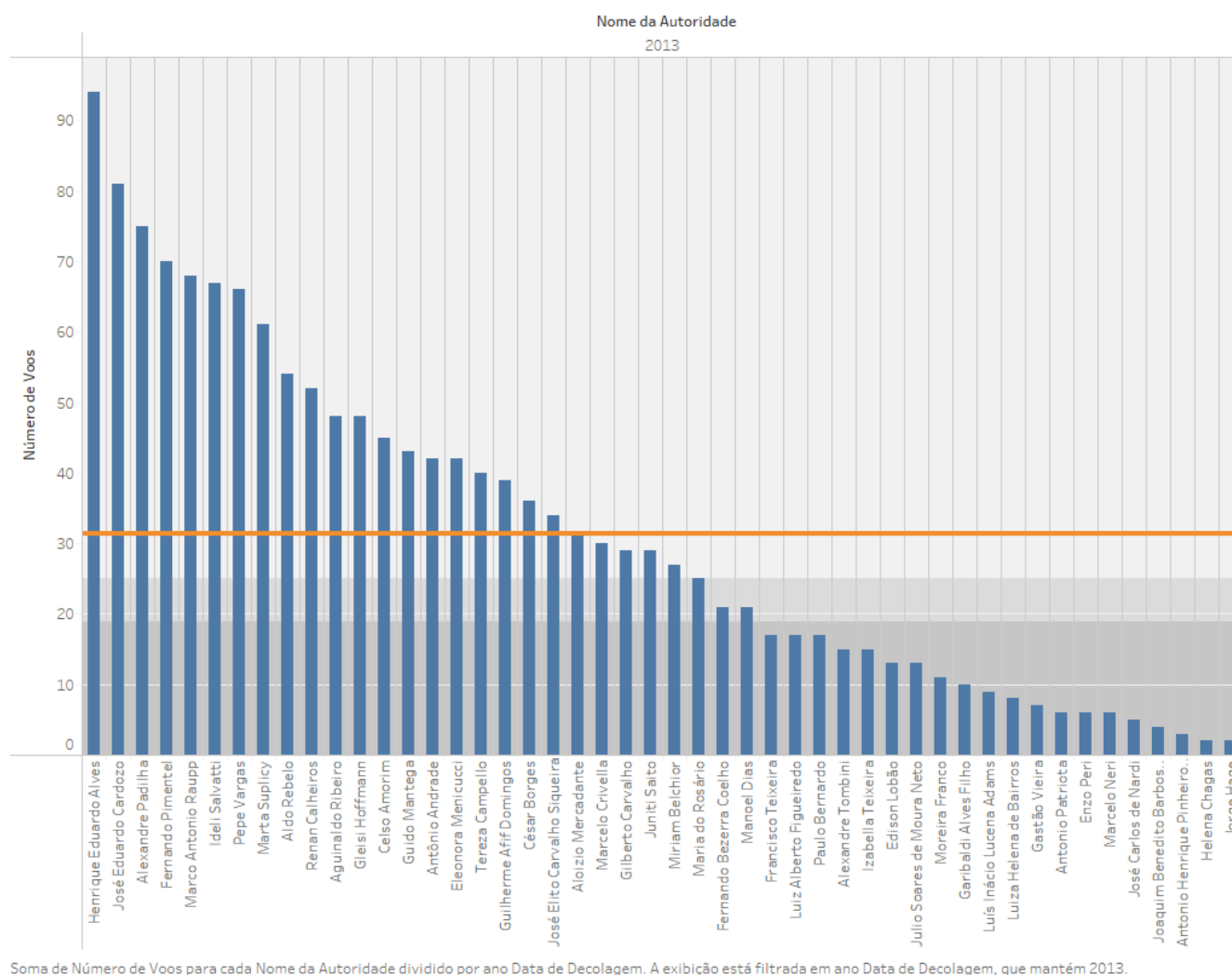


Figura 23 – Gráfico de Barra – Voos por Autoridade no período de 2013

4. 2. 3. Número de Voos por Autoridade de 2014

Este período foi identificado como o maior em número de voos, no conjunto de dados analisado, nesta associação. Como pode-se notar, há um grande salto em relação aos voos realizados, se comparado aos outros períodos, tendo mais da metade das autoridades ultrapassado a média de voo estipulada para o ano.

Ressaltamos as seguintes autoridades, pois estas, obtiveram mais que o dobro do valor da média, 60, conforme a *Figura 24*. Ressaltamos que apesar delas serem destaques neste ano, elas não foram identificadas como pontos fora da curva no boxplot da *Figura 22*.

- Arthur Chioro
196 voos;
- José Eduardo Cardozo
168 voos;
- Aldo Rebelo
147 voos;
- Ideli Salvatti
143 voos;
- Neri Geller
143 voos;

Voo por Autoridade

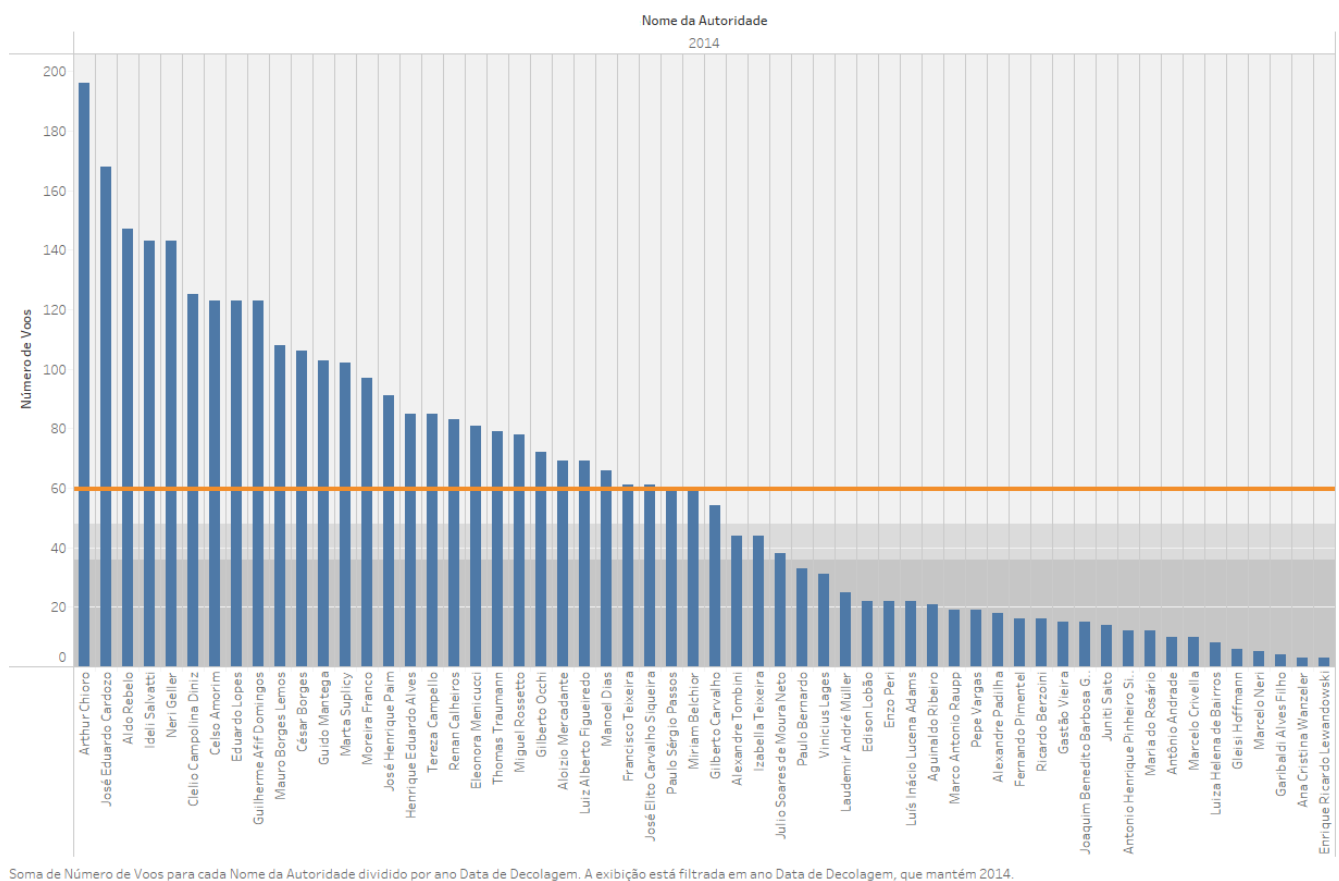


Figura 24 - Gráfico de Barra – Voo por Autoridade no período de 2014

4. 2. 4. Número de Voos por Autoridade de 2015

No período de 2015, mais precisamente em meados de abril, houve uma medida aprovada pela Presidente Dilma[9], onde visava-se reduzir os gastos com utilização de voos da FAB por motivo de moradia. Ou seja, deixou ser permitido a utilização de voos por motivo de residência. A princípio, nota-se uma diminuição significativa na quantidade de autoridades que ultrapassaram a média, apesar de ainda existirem casos que a ultrapassaram, conforme a *Figura 25*.

Para comparação, foi realizado um gráfico em linha contendo os dados de 2014, 2015, 2016 e suas médias. O objetivo é, analisar os anos em um mesmo período de uma forma mais clara, podendo, assim, extrair conclusões se o decreto foi efetivo ou não.

Com o resultado desta análise, podemos inferir que a medida causou um impacto na quantidade de voos realizados, pois nos meses subsequentes a aprovação do decreto, abril de 2015, se inicia uma grande diferença entre os anos. A média de um ano para o outro, reduziu de 289 voos em 2014, para 228 em 2015. Esta redução continuou no ano subsequente, passando de 228 em 2015 para 169 em 2016. Assim, pode-se dizer que houve um recuo de 42% do número de voos da FAB de 2014 a 2016. No gráfico, conforme a *Figura 26*, estão exibidas as linhas dos anos comparando seus valores e suas médias.

Conforme a *Figura 25*, destacamos as seguintes autoridades no período de 2015, devido a sua alta quantidade de voos realizados em relação à média do ano, duas delas já destacadas no boxplot da *Figura 22*:

- Gilberto Kassab
223 voos;
- Aldo Rebelo
161 voos;
- Eduardo Cunha
144 voos;
- George Hilton
133 voos;

Voo por Autoridade

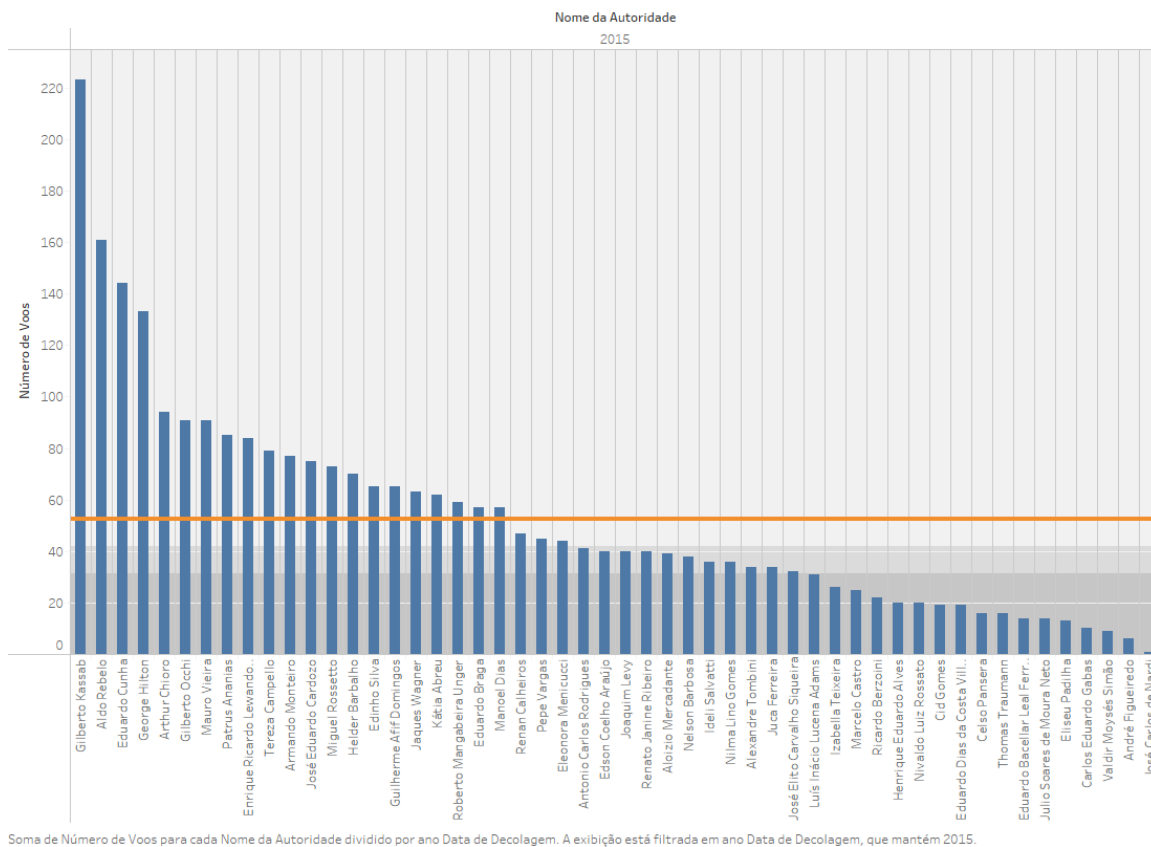


Figura 25 - Gráfico de Barra - Voos por Autoridade no período de 2015

Voos por Período 2014 x 2015 x 2016



Figura 26 - Gráfico de Linha - Comparação do Número de Voos dos anos de 2014 a 2016

4. 2. 5. Número de Voo por Autoridade de 1º de janeiro a 12 de maio de 2016 – Referente ao Governo da Presidente Dilma Rousseff

Mantendo a tendência após o decreto, no gráfico da *Figura 27* podemos verificar que somente algumas autoridades ultrapassaram a média de 16 voos no período. Como o período se tornou muito pequeno, destacaremos somente as autoridades mais que dobraram a média, na sua quantidade de voos realizados.

- Aldo Rebelo
58 voos;

- Gilberto Kassab
40 voos;
- Edinho Silva
38 voos;

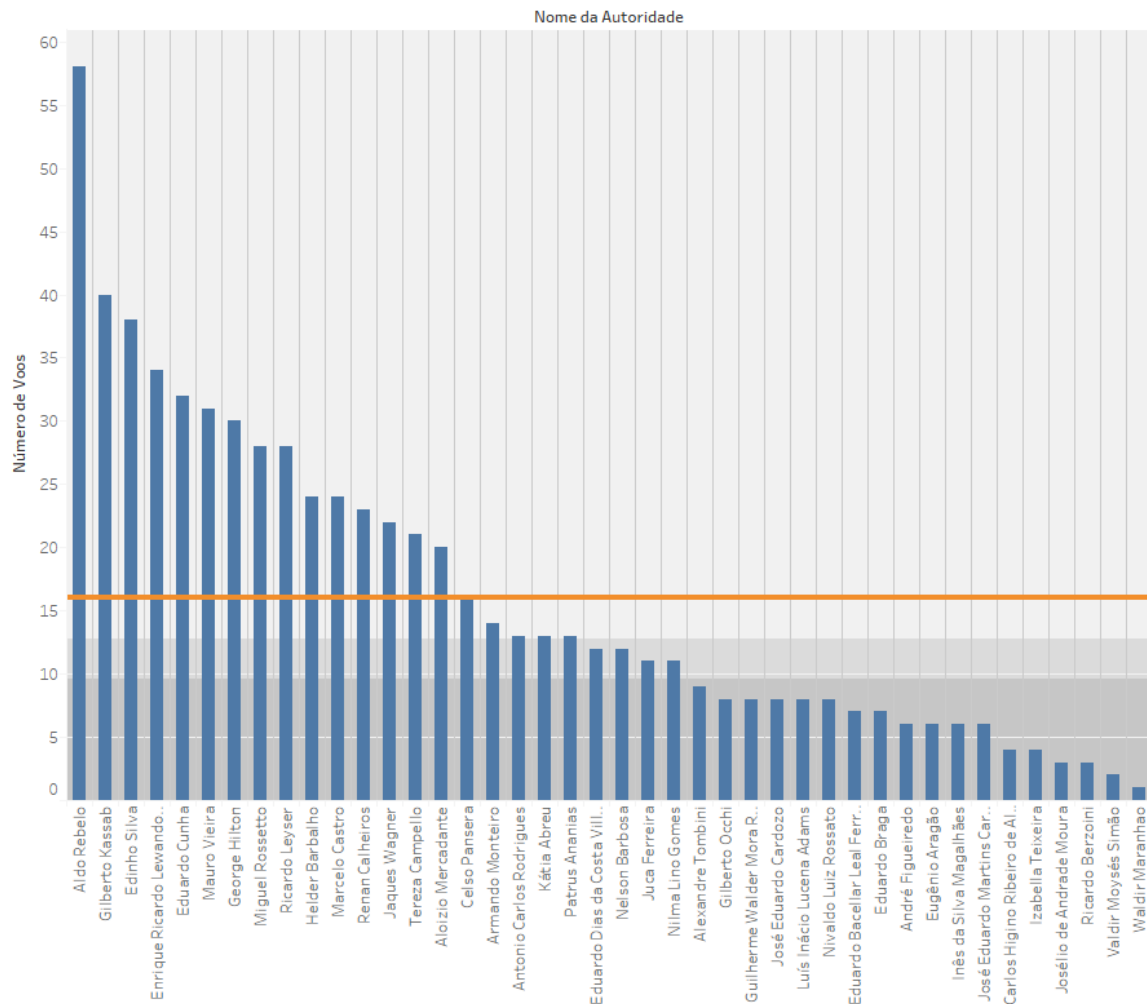


Figura 27 – Gráfico de Barra – Voo por Autoridade no período pré-impeachment da Presidente

4. 2. 6. Número de Voo por Autoridade de 12 de maior a 10 de novembro de 2016 – Referente ao Governo do Presidente Michel Temer

Assim como nos gráficos anteriores, foi notória a diferença com os períodos passados, entretanto, devemos destacar novamente, as autoridades que ultrapassaram a média estipulada para este intervalo, conforme a *Figura 28*. No segundo semestre de 2016,

foi identificada uma certa semelhança com o semestre anterior, onde um grande número de autoridades se manteve abaixo da média e outras ultrapassaram, com destaque para algumas, a quais mais do que duplicaram a média estipulada para o período. A média do período foi de 32 voos por autoridade. Esta média aumentou, se comparada ao semestre anterior que foi de 15 voos por autoridade. Autoridades destacadas:

- Alexandre de Moraes
92 voos;
- José Serra
79 voos;

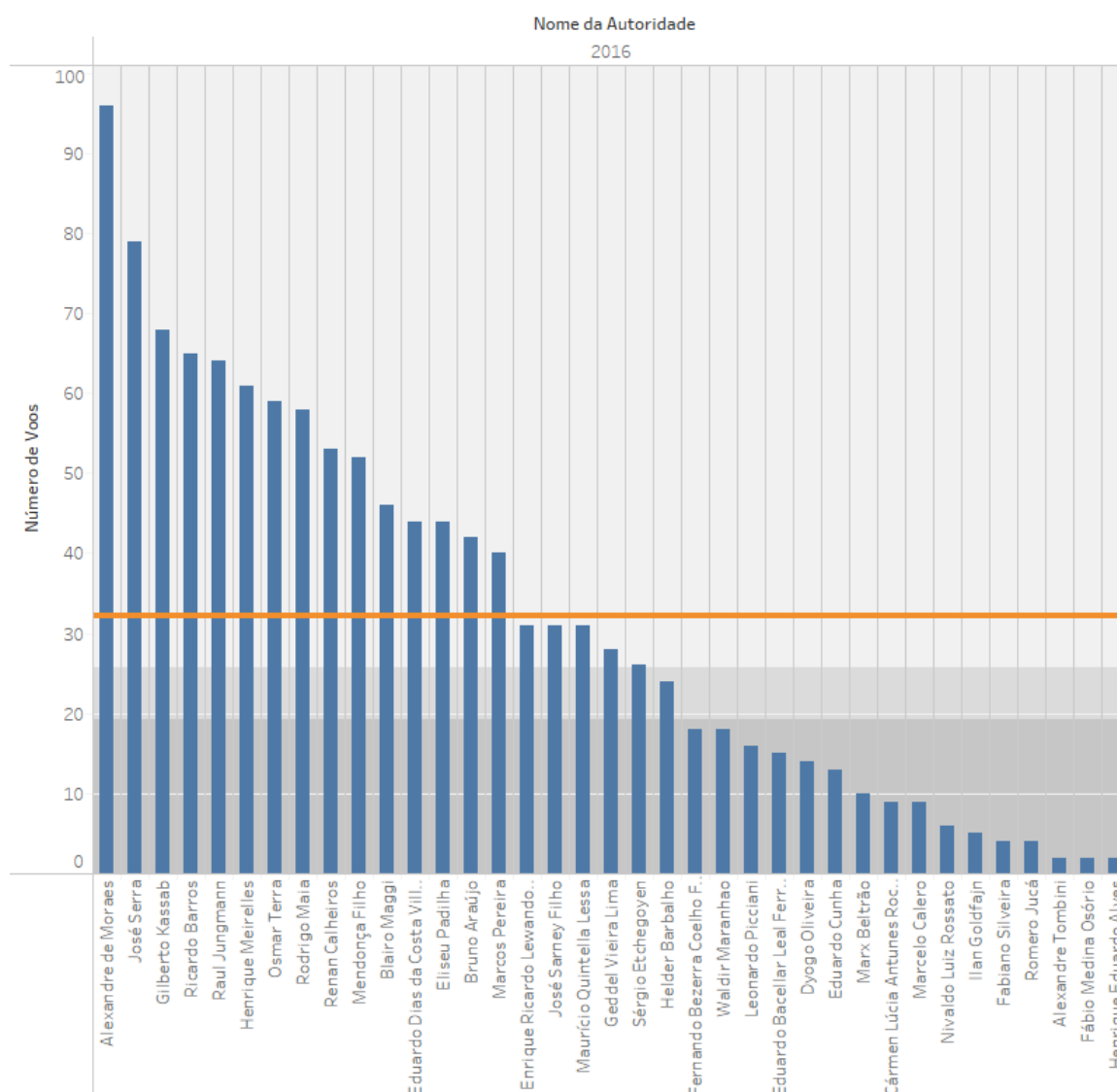


Figura 28 – Gráfico de Barra – Voo por Autoridade no período pós-impeachment da Presidente

4. 3. Análise de Voos do Governo da Presidente Dilma Rousseff em relação ao Governo do Presidente Michel Temer

Esta análise tem como principal objetivo a comparação dos períodos entre 12 de maio à 10 de novembro dos anos de 2015 e 2016. Para facilitar a nomeação, este período no ano de 2015 será tratado como *TD*, referente ao governo da ex-Presidente Dilma Rousseff. No ano de 2016 este mesmo período será tratado como *TT*, referente ao governo do Presidente Michel Temer. Esta divisão foi realizada para tentar comparar os dois governos, visto que a data inicial do conjunto de dados, junho de 2013, até 12 de maio de 2016, data do impeachment da ex-presidente, temos somente um governo vigente. Por isso, o intervalo *TD e TT*, serão utilizados na comparação.

Como não há muitos dados referente a outro governo, que não seja o da ex-Presidente Dilma Rousseff, tornou-se difícil realizar comparações. Portanto, a análise se restringirá a comparar as quantidades neste pequeno período e concluir se há alguma discrepância notável.

4. 3. 1. Análise do Período *TD e TT*

Conforme a *Figura 29*, é possível verificar que houve um decréscimo no consumo de voos da FAB, entre *TD e TT*, mais precisamente, a média entre os anos diminuiu de 204 voos para 168. Entretanto, os gráficos são muito parecidos neste período, tendo curvas similares, principalmente entre os meses de maio e setembro, possivelmente configurando-se um padrão. Apesar disso, seriam necessários mais dados para se conseguir confirmar tal norma.

Ainda com o gráfico, é possível inferir que o decreto proposto pela ex-Presidente Dilma Rousseff, em abril de 2015, foi melhor executado após sua saída. Todavia, serão necessários mais dados, relativos aos anos anteriores a 2013, para que esta afirmação seja ratificada.

Voos do Governo Dilma e Temer (12 mai ~ 11 nov)



Figura 29 - Gráfico de Linha - Comparativo entre períodos do governo Dilma e Temer

4. 4. Boxplot – Máximo de Passageiros por Autoridade

Neste gráfico, o objetivo é obter as autoridades que se identificam como pontos fora da curva, no quesito quantidade máxima de passageiros.

Foram identificados os seguintes pontos fora da curva na *Figura 30*:

- Eduardo Cunha
51 passageiros em 2015;
- José Eduardo Cardozo
50 passageiros em 2014;
- Marcos Pereira
45 passageiros em 2016;
- Aldo Rebelo
40 passageiros em 2016;
- Celso Amorim
35 passageiros em 2013;
- Enzo Peri
35 passageiros em 2014;
- Blairo Maggi
35 passageiros em 2016;
- Gilberto Carvalho
31 passageiros em 2014;
- Ideli Salvatti
28 passageiros em 2014;
- Gleisi Hoffman
27 passageiros em 2014;

Maximo de Passageiros em um Voo por Autoridade

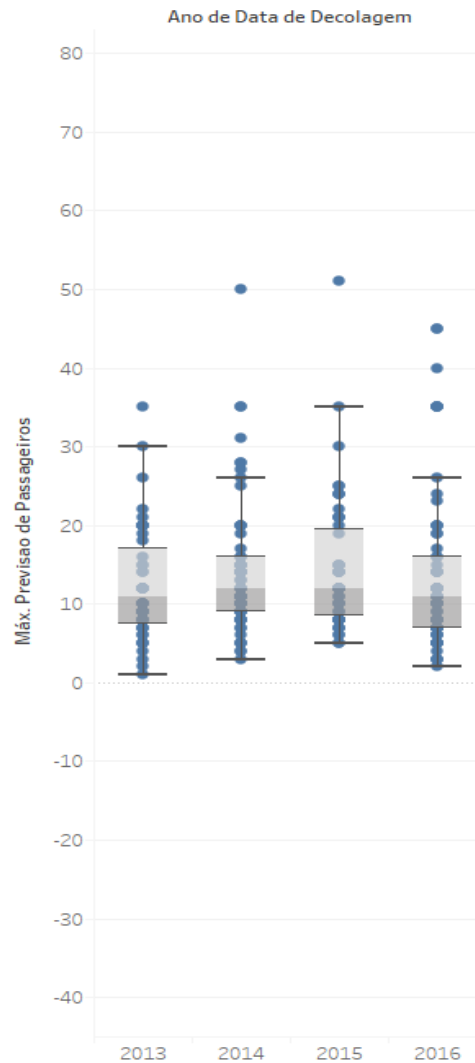


Figura 30 - Gráfico de Boxplot -
Valor Máximo de passageiros em um
Voo por Autoridade

4. 5. Total de Passageiros por Autoridade ao Ano

No gráfico de Total de Passageiros por Autoridade, é demonstrada a visualização do somatório de passageiros por autoridade, para os anos de 2013 a 2016, conforme a *Figura 31*. Ao lado de cada gráfico de barra, observa-se a média do total de passageiros

para o ano. Foram destacadas as autoridades que possuem pelo menos o dobro da média de cada ano.

Total de Passageiros por Autoridade

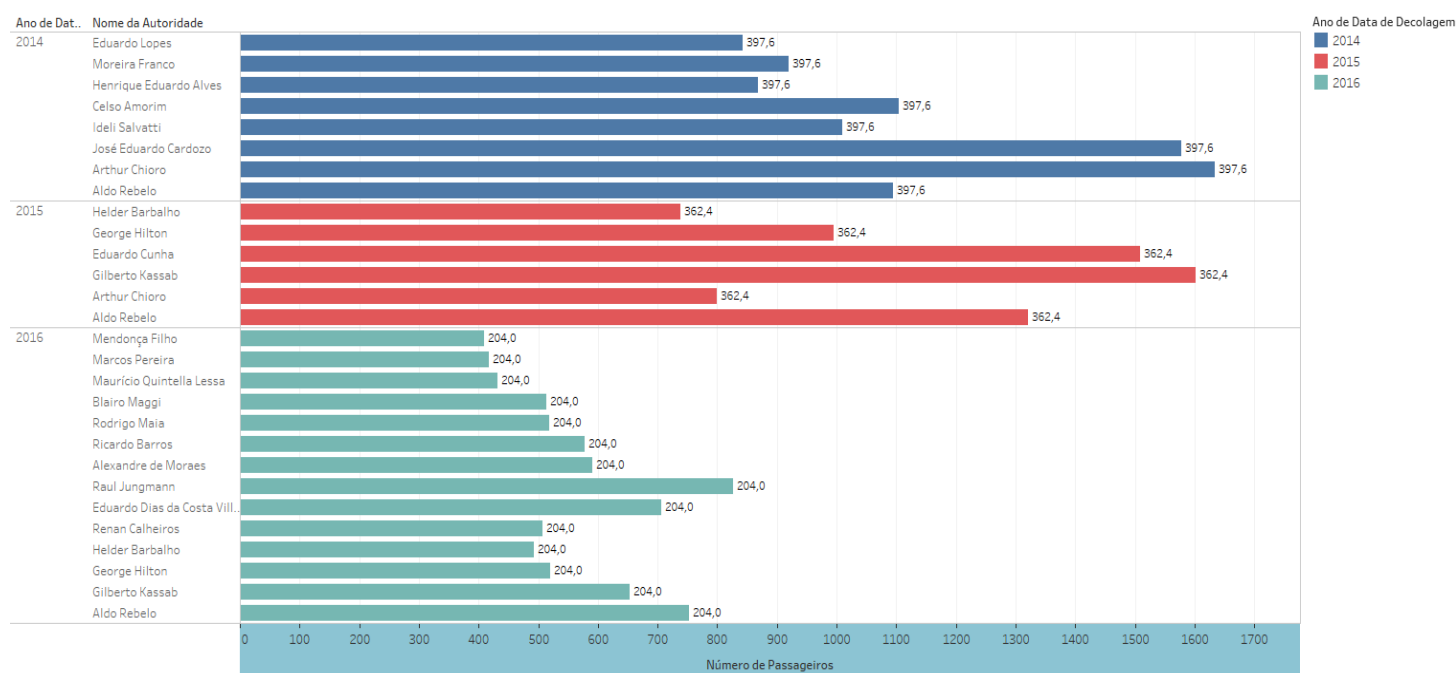


Figura 31 - Gráfico de Barra - Total de Passageiros por Autoridade

4. 5. 1. Comparativo entre 2014, 2015 e 2016

No ano de 2014, destacam-se os Ministros José Eduardo Cardozo e Arthur Chioro, os quais obtiveram os maiores números de passageiros transportados consigo mesmo, com um total de 1.634 e 1.578, respectivamente, nas suas viagens ao longo do ano. Sendo a média para este ano de 397 passageiros.

Nota-se o mesmo padrão para o ano de 2015, mantendo-se praticamente a mesma média. Destacam-se o Presidente da Câmara dos Deputados, Eduardo Cunha e o Ministro Gilberto Kassab, os quais obtiveram 1.508 e 1.602 passageiros transportados consigo mesmo, respectivamente. Sendo a média para este ano de 362 passageiros.

Assim como notado o decréscimo do número de voos entre 2014 e 2016, percebe-se uma grande diminuição do número de passageiros transportados no ano de 2016, sendo o Ministro Henrique Alves a autoridade que teve o maior número de passageiros

transportados consigo mesmo, com um total de 826 passageiros nas suas viagens. A média para este ano foi de 204 passageiros.

Ao longo do período de 1º de janeiro de 2014 a 10 de novembro de 2016, a previsão de passageiros transportados foi de 55.983, gerando uma média de 18.661 passageiros ao ano.

4. 6. Relação do Número de Voos por Tempo de Mandato

Nesta análise, o objetivo é associar o número de voos realizados pela autoridade com seu tempo de vigência, assim conseguindo traçar uma proporção para cada autoridade.

O valor máximo encontrado nesta análise foi de 66,67% de um mandato de 42 dias. O valor mínimo foi de 0,17% para um mandato de 1.152 dias. Com este último valor, quebramos a hipótese criada no início deste Capítulo, onde foi dito que quanto maior a vigência, maior seria o número de voos realizados. Entretanto, há inúmeros exemplos que refutam esta hipótese, como os listados na *Figura 32*.

Nome da Autoridade	Gabinete	Número de Voos	Vigencia Mandato	% Voo x Mandato
Fernando Bezerra Coelho	Ministério da Integração Nacional	21	1.004	2,09%
Enzo Peri	Comando do Exército	28	1.496	1,87%
Izabella Teixeira	Ministério do Meio Ambiente	89	1.958	4,55%
Julio Soares de Moura Neto	Comando da Marinha	65	1.498	4,34%
Luís Inácio Lucena Adams	Advocacia-geral da União	70	1.888	3,71%
Alexandre Padilha	Ministério da Saúde	93	1.129	8,24%

Figura 32 - Exemplos de Autoridades que Refutam a Hipótese

Nas análises a seguir, foram selecionadas algumas autoridades julgadas mais interessantes para o estudo. Antes das análises serem realizadas, é preciso enfatizar que todas as hipóteses levantadas nesta seção são meramente conjecturais, e que elas podem ser ratificadas através do cruzamento das informações disponíveis neste trabalho com as agendas das autoridades. Entretanto, conforme a reportagem do Jornal Estadão[10], estas agendas não são uma fonte 100% confiável, visto que nem todas as autoridades as mantêm atualizadas, apesar de ser obrigatória a divulgação de compromissos conforme lei[11]. Na reportagem, também foi destacada a limitação de que, para muitos ministros, os compromissos não estavam listados.

- Ricardo Leyser

Através da análise crua da informação da *Figura 33* é fácil concluir que o ministro do esporte teve uma frequência de voo muito alta. Entretanto, ao analisarmos seu período de mandato, de março a maio de 2016, podemos inferir que a alta frequência se deve a Olimpíada.

Nome da Autoridade	Gabinete	Número de Voos	Vigencia Mandato	% Voo x Mandato
Ricardo Leyser	Ministério do Esporte	28	42	66,67%

Figura 33 - Relação Voo x Vigência - Ricardo Leyser

- Arthur Chioro

Outro exemplo de uma frequência alta de voos, pelo período do mandato, é identificada nesta autoridade, conforme a *Figura 34*.

Se for comparada sua frequência com a de outras autoridades do mesmo gabinete, mas em períodos diferentes, conforme a *Figura 35*, é possível inferir que houve desvio do padrão no número de voos utilizados.

Nome da Autoridade	Gabinete	Número de Voos	Vigencia Mandato	% Voo x Mandato
Arthur Chioro	Ministério da Saúde	290	606	47,85%

Figura 34 - Relação Voo x Vigência - Arthur Chioro

Porcentagem de Voo pelo Tempo de Mandato

Nome da Autoridade	Gabinete	Número de Voos	Vigencia Mandato	% Voo x Mandato
Arthur Chioro	Ministério da Saúde	290	606	47,85%
Ricardo Barros	Ministério da Saúde	65	183	35,52%
Marcelo Castro	Ministério da Saúde	49	208	23,56%
Alexandre Padilha	Ministério da Saúde	93	1.129	8,24%

Figura 35 - Relação Voo x Vigência - Ministério da saúde

- Gilberto Kassab

Assim como a autoridade acima, nota-se um certo exagero na frequência de voos, se comparado a outras autoridades do mesmo gabinete, conforme a *Figura 36*.

Porcentagem de Voo pelo Tempo de Mandato

Nome da Autoridade	Gabinete	Número de Voos	Vigencia Mandato	% Voo x Mandato
Gilberto Kassab	Ministério das Cidades	263,0	470,0	55,96%
Gilberto Occhi	Ministério das Cidades	72,0	290,0	24,83%
Bruno Araújo	Ministério das Cidades	42,0	183,0	22,95%
Inês da Silva Magalhães	Ministério das Cidades	6,0	27,0	22,22%
Aguinaldo Ribeiro	Ministério das Cidades	69,0	769,0	8,97%

Figura 36 - Relação Voo x Vigência - Ministério das Cidades

No gráfico da *Figura 37*, pode-se visualizar através das linhas, a comparação do número de voos entre as autoridades, em períodos diferentes. Kassab assumiu em janeiro de 2015, e o término de seu mandato foi em abril de 2016. Já Occhi, teve seu mandato iniciado em março de 2014 e término em janeiro de 2015. Com isso, pode-se inferir que Kassab obteve uma frequência maior de voos, devido a seu período de mandato, visto que era véspera das Olimpíadas.



Figura 37 - Gráfico de Linha - Comparação entre os períodos das Autoridades do Ministério das Cidades

5. Conclusão

Ao final de todos os passos abordados anteriormente, a saída do processo consiste em o compartilhamento de uma base de dados consolidada sobre os voos realizados pela FAB.

Este capítulo reserve-se a: descrever limitações encontradas ao longo do projeto; analisar o produto final encontrado das análises e almejar futuras progressões no âmbito da análise e coleta desta categoria de dados.

5.1. Considerações Finais

Ao final deste trabalho, podemos afirmar que a proposta do governo de aproximar o cidadão através da transparência de seus dados, em muitos casos se torna apenas a divulgação de uma informação. Esta divulgação não garante que ela é transparente e entendível ao cidadão, o que não atrai esta proximidade.

Através deste entendimento, o objetivo deste trabalho foi a criação de um processo, no qual se possa extrair as informações divulgadas, traduzi-las para um formato de fácil manipulação, enriquecê-las com mais dados, divulgá-las através de uma base de dados e inferir análises em cima delas. Assim, ao término dele, será possível executar este processo, traduzindo as informações divulgadas pela Força Aérea Brasileira sobre os voos realizados, em uma base de dados consolidada, onde há o enriquecimento através de fontes externas, sua divulgação e análises.

Ao longo da execução deste processo, foram abordadas inúmeras dificuldades encontradas, como por exemplo a falta de padrão nas informações divulgadas. As soluções para estas dificuldades, sejam por utilização ou criação de ferramentas, foram descritas e detalhadas.

Nas análises, algumas inferências foram realizadas. Uma delas, é que o decreto criado pela Presidente Dilma, em abril de 2015, surtiu um efeito imediato no número de

voos realizados pela FAB, assim, possivelmente diminuindo os gastos do governo com este setor.

Assim, a conclusão foi, com as informações e o formato de divulgação utilizado pela FAB, não é possível realizar análises concretas. Portanto, fez-se necessária a criação deste trabalho de conclusão de curso, para que tais informações não fossem simplesmente divulgadas, mas sim, tornarem-se passíveis de análises e enriquecimento. Criando, de fato, o objetivo proposto pela lei de transparência, que é o incentivo à participação da população nos processos de elaboração de planos, leis e controle de gastos do governo.

5.2. Limitações do Projeto

Nesta seção serão citadas e discutidas as limitações encontradas ao longo do desenvolvimento do projeto.

5.2.1. Associar aeronave ao voo

Por questões de segurança a FAB não informa quais aeronaves foram utilizadas nos voos listados. Desta maneira, não foi possível realizar a associação, o que limitou bastante a análise de dados, no quesito valoração, visto que esta é uma informação crucial para este tipo de análise.

5.2.2. Fonte descentralizada de informações

Por se tratar de informações governamentais, e toda base de transparência proposta pelo governo, a expectativa era conseguir encontrar a grande maioria delas em um único local. Entretanto, não é o caso. As informações sobre as pessoas, por trás de cada cargo, e associada ao tempo de vigência, não foi encontrada em nenhum site oficial. Assim, foi preciso utilizar uma fonte alternativa na etapa do enriquecimento, que conforme sua seção descreveu, foi realizada manualmente.

5.2.3. Valorar os gastos públicos

Um dos principais objetivos que almejávamos, antes do início do projeto, era a valoração dos gastos de cada autoridade e gabinete com voos realizados. Ou seja, após a análise, iríamos extrair medidas de gastos para cada gabinete e para cada autoridade, tirando, assim, conclusões sobre consumos que fujam ao padrão.. Entretanto, foi identificado que, por questões de segurança, a FAB não divulga quais aeronaves foram utilizadas em seus voos. Conforme reportagem do UOL[5], *“O combustível consumido pelas aeronaves oficiais entra na conta geral da FAB, que negocia os contratos por meio de licitações com fornecedores anualmente, com uma quantidade já pré-estabelecida para cada ano e não há desconto de impostos.*

O Comando da Aeronáutica, entretanto, não divulga os valores dessas transações nem quanto querosene é consumido pelos aparelhos de transporte, tampouco os gastos com manutenção. A questão é tratada como ‘informação estratégica’.”

Desta forma, torna-se difícil realizar a associação aeronave-voo, que resultaria em valoração do consumo de cada gabinete e/ou autoridade. Uma possível solução para este tópico será abordada na seção 5.3 deste projeto.

5.3. Trabalhos Futuros

Neste tópico, serão abordadas sugestões para continuação do vigente trabalho. Assim, pode-se enumerá-las:

- **Automatização do Processo de Enriquecimento**

Conforme descrito na seção 3.3 do Capítulo 3, o enriquecimento de dados é realizado de forma manual, visto a falta de um sítio único e consolidado das informações necessárias, como nome, vigência e partido da autoridade. Nesta sugestão, será necessário ou criação ou utilização de uma única fonte de dados confiável, que possa prover estas informações.

- **Associar Aeronave ao Voo e obter Valores dos Voos**

Devido às limitações das informações divulgadas, não se consegue associar a aeronave utilizada no voo. Desta forma, não é possível valorar os voos utilizados. Esta

sugestão tem como objetivo valorar os gastos públicos com os voos realizados pela FAB, assim, podendo obter uma análise mais valiosa dos dados.

- **Associar Tableau com o R**

O R é uma ferramenta open-source de estatística muito poderosa, podendo obter-se inúmeros resultados estatísticos os quais o Tableau não possui suporte, como por exemplo uma análise de correspondência. Nesta integração, é possível exibir a matriz de correspondência no Tableau, através dos resultados gerados no R. Assim, sugerimos para trabalhos futuros, que seja realizada uma análise mais profunda dos dados, utilizando esta integração com o R.

Melhoria na normalização

O processo de normalização da informação sobre a autoridade está muito engessada, o ideal seria criar um padrão que conseguisse identificar os termos e fizesse a normalização de forma mais robusta.

Referências Bibliográficas

- [1] BRASIL. Lei complementar nº 101, de 4 de maio de 2000. Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências. Disponível em:
http://www.planalto.gov.br/ccivil_03/leis/LCP/Lcp101.htm
- [2] BRASIL. Lei complementar nº 131, de 27 de maio de 2009. Acrescenta dispositivos à Lei[1], a fim de determinar a disponibilização, em tempo real, de informações pormenorizadas sobre a execução orçamentária e financeira da União, dos Estados, do Distrito Federal e dos Municípios. Disponível em:
https://www.planalto.gov.br/ccivil_03/leis/LCP/Lcp131.htm
- [3] BRASIL. Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei no 8.112, de 11 de dezembro de 1990; revoga a Lei no 11.111, de 5 de maio de 2005, e dispositivos da Lei no 8.159, de 8 de janeiro de 1991; e dá outras providências. Disponível em:
http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm.
- [4] BRASIL. Decreto nº 7.185, de 27 de maio de 2010. Dispõe sobre o padrão mínimo de qualidade do sistema integrado de administração financeira e controle, no âmbito de cada ente da Federação, nos termos do art. 48, parágrafo único, inciso III, da Lei[1], e dá outras providências. Disponível em:
http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2010/decreto/d7185.htm
- [5] Vinholes, Thiago TV. 23 de outubro de 2015 “Conheça os aviões executivos da FAB”. <http://airway.uol.com.br/conheca-os-avioes-executivos-da-fab/>. Acesso em outubro de 2016.
- [6] Fomina, Angelina AF. Julho de 2015. “ParseHub gives Data Scientists a better, faster way to collect data”. <http://www.kdnuggets.com/2015/07/parsehub-better-faster-way-collect-data.html>. Acesso em outubro 2016.
- [7] Wikipedia. Lista dos Ministérios do Brasil. Disponível em:
https://pt.wikipedia.org/wiki/Ministérios_do_Brasil. Acesso em outubro de 2016.
- [8] BRASIL. Decreto nº 4.244, de 22 de maio de 2002. Dispõe sobre o transporte aéreo, no País, de autoridades em aeronave do Comando da Aeronáutica. Disponível em:
https://www.planalto.gov.br/ccivil_03/decreto/2002/D4244.htm
- [9] BRASIL. Decreto nº 8.432, de 9 de abril de 2015. Restringe o uso de aeronaves do Comando da Aeronáutica em deslocamentos para o local de domicílio. Disponível em:
https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/decreto/d8432.htm.
- [10] Araújo, Carla CA e Peron, Isadora IP. O Estado de S.Paulo. 07 de novembro 2016. “Ministros ignoram normas em 238 viagens pela FAB”. Disponível em:
<http://politica.estadao.com.br/noticias/geral,ministros-ignoram-normas-em-238-viagens-pela-fab,10000086781> Acesso em novembro de 2016.

[11] BRASIL. lei nº 12.813, de 16 de maio de 2013. Dispõe sobre o conflito de interesses no exercício de cargo ou emprego do Poder Executivo federal e impedimentos posteriores ao exercício do cargo ou emprego; e revoga dispositivos da Lei nº 9.986, de 18 de julho de 2000, e das Medidas Provisórias nºs 2.216-37, de 31 de agosto de 2001, e 2.225-45, de 4 de setembro de 2001. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2013/Lei/L12813.htm.