

Universidade Federal do Paraná
Setor de Ciências Exatas
Departamento de Estatística
Programa de Especialização em *Data Science* e *Big Data*

Ulisses Fernandes Alves

**Estudo quantitativo de gargalos
operacionais em um sistema ferroviário**

Curitiba
2022

Ulisses Fernandes Alves

Estudo quantitativo de gargalos operacionais em um sistema ferroviário

Monografia apresentada ao Programa de Especialização em Data Science e Big Data da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Walmes Zeviani

Curitiba

2022

Estudo quantitativo de gargalos operacionais em um sistema ferroviário

Ulisses Fernandes Alves¹

Walmes Zeviani²

Resumo

O sistema ferroviário apresenta diversos eventos aleatórios que aumentam a complexidade do planejamento da circulação dos trens. Deste modo, é importante que os gestores das ferrovias conheçam os locais e eventos que mais contribuem para reduzir a capacidade de transporte na malha ferroviária e, com isso, identifiquem e forneçam soluções para os gargalos operacionais. Neste sentido, o presente trabalho desenvolveu uma aplicação *web* utilizando a linguagem de programação R com uso do pacote *Shiny* para elaborar um painel visual (*dashboard*) que contém informações, métricas e principais indicadores da circulação de trens para suportar o estudo e análise dos principais locais com gargalos operacionais. Além disso, foi utilizada a abordagem *K-Means* para definir o agrupamento de linhas singelas com características semelhantes de capacidade e fluxo diário de trens. A partir disto, foram estabelecidos 3 grupos dentre as 82 linhas singelas avaliadas, o que permitirá realizar ações estratégicas e específicas para o controle e manutenção de cada grupo.

Palavras-chave: Sistema ferroviário, *dashboard*, *shiny*.

Abstract

The railway system presents several random events that increase the complexity of planning the movement of trains. Thus, it is important that railroad managers know the places and events that most contribute to reducing the transport capacity in the rail network and, with that, identify and provide solutions for the operational bottlenecks. In this sense, the present work developed a web application using the R programming language and Shiny library to create a visual panel (dashboard) that contains information, metrics and main indicators of the circulation of trains to support the study and analysis of the main locations with operational bottlenecks. In addition, the K-Means approach was used to define the grouping of single tracks with similar characteristics of capacity and daily train flow. From this, 3 groups were established among the 82 single tracks evaluated, which will allow strategic and specific actions to be carried out for the control and maintenance of each group.

Keywords: Railroad system, *dashboard*, *shiny*.

I Introdução

Uma malha ferroviária de linha singela apresenta algumas especificidades para a circulação de trens pelo fato de que os segmentos de linha singela podem ser ocupados apenas por trens do mesmo sentido, enquanto que um outro trem no sentido oposto aguarda a desocupação da linha singela em um pátio com linhas desviadas que permitem o cruzamento dos trens.

Se ocorre uma parada não programada na via singela durante a circulação de um trem, os tempos de cruzamento para os trens em ambos sentidos podem aumentar de modo exponencial, pois ocorre um efeito cascata de parada em todos os trens que estão nas regiões adjacentes ao evento, semelhante ao que ocorre durante um congestionamento de tráfego rodoviário. Como exemplo de paradas não programadas, podem ser mencionadas as falhas de locomotivas, quebra de trilhos, atrasos durante a troca de jornada de maquinistas, falhas em vagões, vandalismos nos trens, dentre outras.

Estas anomalias atrasam de modo significativo o tempo de viagem dos trens e podem impedir o atendimento do volume de transporte planejado para o período. Isto ocorre por conta de restrições na quantidade disponível de ativos ferroviários (locomotivas e vagões) e também pela limitação na capacidade de veículos em circulação que a malha ferroviária suporta. Deste modo, quanto maior a duração na viagem dos trens, maior é o tempo gasto entre dois carregamentos sucessivos, o que implica na necessidade de uma quantidade maior de trens para atender o mesmo volume planejado. Contudo, ao se aproximar dos limites específicos na capacidade de circulação de trens na malha ferroviária, aumentam-se os tempos que os trens aguardam nos pátios de cruzamento.

Portanto, uma frequência alta de paradas não programadas pode trazer grandes prejuízos financeiros para a empresa pelo fato de impedir o atendimento dos contratos acordados para o transporte de volume ferroviário. Por conta disso, o presente trabalho visa contribuir com este problema ao apresentar um *dashboard* interativo que permite a visualização de informações que permitem realizar um diagnóstico dos locais da malha ferroviária que apresentam maior incidência de anomalias. Deste modo, são fornecidos subsídios para os gestores implementarem ações que tragam maior eficiência para a circulação dos trens em uma malha de linha singela.

¹Ulisses Fernandes Alves, ulissesalves@outlook.com.br.

²Professor do Departamento de Estatística DEST/UFPR e vice-coordenador da pós-graduação em Data Science e Big Data da UFPR, walmes@ufpr.br.

2 Materiais e Métodos

A malha ferroviária avaliada no presente trabalho é composta por quase 1400 quilômetros de linhas singelas e fica localizada entre as cidades de Rondonópolis/MT e Hortolândia/SP, sendo segmentada em 6 trechos distintos (trechos 2A, 2B, 3, 4, 5 e 6). É composta pelas concessões Rumo Malha Norte e Rumo Malha Paulista e durante o primeiro trimestre de 2022 apresentou a relevante parcela de quase 80% de toda tonelada-quilômetro útil transportada pela concessionária Rumo Logística [1].

Neste contexto, será desenvolvido um painel visual (*dashboard*) que contenha informações relevantes para identificar os locais com ineficiências operacionais. Foi utilizada a linguagem de programação R versão 4.1.2 (2021-11-01) em conjunto com os pacotes da biblioteca *Shiny* versão 1.7.1 no ambiente RStudio® versão 1.4.1 106 ("Tiger Daylily") para elaboração do *dashboard*.

Também foi realizado um estudo para o agrupamento das linhas singelas com a abordagem *K-Means* com o intuito de gerar conjuntos que apresentem características similares de capacidade de tráfego e fluxo diário de circulação de trens. A partir disso, é possível que os gestores definam estratégias específicas para o controle e manutenção de cada grupo de linhas singelas. O gráfico com os agrupamentos delimitados também está incluso no *dashboard* elaborado.

2.1 O conjunto de dados

O conjunto de dados contém registro das durações dos tempos de trânsito (*transit time*), condução, anomalias e propagações, cuja descrição é fornecida na Tabela 1:

Tabela 1: Dicionário do conjunto de dados.

Variável	Descrição
<i>Transit Time</i>	(contínua) Duração de tempo total para o trem percorrer o espaço avaliado (extensão de comprimento do pátio ou da linha singela)
E.g.	75 min, 30 min
Tempo de condução	(contínua) Duração de tempo total para o trem percorrer o espaço avaliado (extensão de comprimento do pátio ou da linha singela) desconsiderando os tempos de paradas (programadas ou não programadas)
E.g.	40 min, 22 min
Duração de Anomalias	(contínua) Duração de tempo em que o trem permanece parado em decorrência de alguma anomalia que ocorreu no próprio trem
E.g.	15 min, 10 min
Motivos de Anomalias	(discreta) Frequência de ocorrência de cada tipo específico de anomalia no período avaliado
E.g.	30 anomalias decorrentes de operação do pátio
Duração de Propagação	(contínua) Duração de tempo em que o trem permanece parado em decorrência de alguma anomalia que ocorreu em outro trem na malha
E.g.	60 min, 40 min
Motivos de Propagação	(discreta) Frequência de ocorrência de cada tipo específico de anomalia que ocasionou propagação nos demais trens no período avaliado
E.g.	4 propagações geradas pela parada de um trem

As informações das variáveis de interesse estão registradas ao longo de duas bases de dados que foram fornecidas pela Rumo Logística em formato de planilhas de Excel® para todos os dias que compreenderam o mês de maio de 2021. A primeira base (base 1) possui 626 mil e 326 registros, sendo que estes dados são gerados de forma automática por um sistema que registra, após o término da circulação dos trens em uma seção de bloqueio, a data e hora (Data Final) e também os valores realizados de *transit time* e tempo de condução. Cada trem possui uma identificação única (OS) e também um prefixo que permite inferir o seu sentido de deslocamento, sendo importação caso o prefixo termine com número ímpar ou exportação, no caso de terminar com número par. Os primeiros campos desta base podem ser visualizados na Figura 1:

Prefixo	Seção	Bloqueio	Transit Time	Condução	DataFinal	OS
M71	ZQBZBV		0:10:09	0:10:09	3/5/21 4:16	1774300
M71	ZBI1		0:15:05	0:09:26	3/5/21 4:31	1774300
M76	178181		0:15:21	0:15:21	3/5/21 20:47	1774346
M76	ZIQ178		0:05:05	0:05:05	3/5/21 20:52	1774346
M76	ZIQ1		0:26:59	0:26:59	3/5/21 21:19	1774346
M76	165ZIQ		0:06:58	0:06:58	3/5/21 21:26	1774346

Figura 1: Tabela com os registros de *transit time* e tempo de condução dos trens em cada seção de bloqueio.

A segunda base de dados (base 2) possui 11 mil e 160 registros para o período analisado, sendo alimentada manualmente por um técnico que registra as durações em minutos (MIN) e motivos (MOTIVO) das anomalias (A) e propagações (P) que ocorrem em cada local (LOCAL) durante as viagens dos trens. Na Figura 2 é possível visualizar os primeiros registros desta base de dados:

Prefixo	OS	MOTIVO	A/P	MIN	LOCAL	DESCRIÇÃO
C67	1773171	OPER PÁTIO	A	20	TGR	VG CM PORTA ABERTA
C67	1773171	FORA DE CHAVE	P	57	TSE	TRO NÃO RECEBE
I32	1773759	CCO INTERVALO	A	215	TSA	INTERVALO TSA
I32	1773759	OPER TRACÇÃO	P	21	TMJ	AG W70 CONF CAUDA
I32	1773759	OPER PÁTIO	A	25	TGR	MMB PÁTIO TGR
I32	1773759	EQUIPAGEM	A	21	TGR	AG TROCA DE MAQ
I32	1773759	EQUIPAGEM	A	22	TGR	AG TROCA DE MAQ
I34	1773760	CCO INTERVALO	P	129	TSA	INTERVALO E42
I38	1773762	CCO INTERVALO	P	53	TSA	INTERVALO E42
I38	1773762	FORA DE CHAVE	P	208	TBE	CONGESTIONAMENTO NA CIRCULAÇÃO
I38	1773762	OPER PÁTIO	A	26	TGR	MMB PÁTIO TGR
X11	1773488	VAGÃO	A	87	TGR	SUPORTE DE ENGATE QBRADO
I40	1773814	FORA DE CHAVE	P	133	TEP	CONGESTIONAMENTO NA CIRCULAÇÃO
I40	1773814	EQUIPAGEM	P	58	TBE	AG TROCA DE MAQ, AG TGR LIB
T52	1773935	FORA DE CHAVE	P	131	TMJ	CONGESTIONAMENTO NA CIRCULAÇÃO

Figura 2: Tabela com os registros dos motivos e durações em minutos das anomalias e propagações que acontecem em cada local da malha ferroviária.

As paradas não programadas são classificadas dentro de um dos seguintes motivos:

- Operação pátio: paradas ocasionadas por falhas de processos que são realizados em determinados pátios, como por exemplo, parada do trem por falta de combustível, carga dos vagões deslocada (pode ocasionar tombamento dos vagões), manobras não previstas na

operação do pátio, dentre outras.

- Fora de chave: paradas ou estacionamentos por conta de congestionamento nos trechos, pátios ou no porto.
- CCO Intervalo: paradas decorrentes de intervalos programados para manutenções corretivas como queda de barreira, reparos na via permanente ou acidentes.
- Operação tração: paradas ocasionadas por anomalias com os maquinistas, tais como falta do maquinista que assumiria o trem ou por conta de necessidades fisiológicas fora do programado (sono, dormir, etc.).
- Equipagem: atraso ou falha na convocação dos maquinistas
- Vagão: paradas ocasionadas por avarias nos vagões, como freio agarrado, emergência indesejada, etc.
- T.O.: falhas ocasionadas nos equipamentos de tecnologia operacional (T.O) como sinal semafórico ferroviário apagado, avaria no computador de bordo da locomotiva
- T.I.: falhas no sistema ferroviário decorrentes dos sistemas ou tecnologias da informação (T.I) por exemplo falha no sistema de licenciamento dos trens.
- Via: paradas ocasionadas por defeitos ou falhas na via permanente.
- Vandalismo: qualquer parada ocasionada pela ação de terceiros, tais como roubos de produto, obstrução intencional do gabarito da via com galhos, pedras, pneus ou outros materiais.
- Loco: paradas decorrentes de falhas ou anomalias nas locomotivas.
- Acidente: são caracterizados como acidentes todos os abalroamentos, descarrilamentos, atropelamentos e colisões graves.
- MRS: paradas nos trens da concessionária Rumo por consequência de paradas dos trens da concessionária MRS.
- Sistemp: paradas na circulação dos trens para realização de inspeção na via permanente em decorrência de fortes chuvas (possibilidade de alagamentos) ou por motivos térmicos da via permanente (temperaturas altas ou baixas).
- VLI: paradas nos trens da concessionária Rumo por consequência de paradas dos trens da concessionária VLI.
- CCO: atrasos nos trens ocasionados por falha do processo de operação do Centro de Controle Operacional (CCO) como cruzamentos errados, atrasos na digitação dos veículos ferroviários que compõem o trem, dentre outros.

2.2 Limpeza e preparo dos dados

Para cada base de dados foi elaborado um arquivo com código em linguagem R (*dados-tempos.R* e *dados-impactos.R*) para fazer a leitura, limpeza e tratamento das bases de dados.

Foram removidos todos os registros duplicados e também os casos em que, ou não haviam registros de OS para os trens, ou a OS tinha um valor igual a 0. Com a aplicação deste filtro, a base 1 apresentou 170 mil e 667 registros (redução de 73%), enquanto que a base 2 teve 8 mil e 212 registros (redução de 26%). Posteriormente, as durações dos tempos e as frequências de ocorrência foram agrupadas para cada pátio e linha singela que pertencem aos trechos da malha ferroviária.

Em cada um destes arquivos ainda foi implementado uma função para permitir a visualização dos gráficos do tipo *Boxplot*, conforme a aplicação de filtros na interface gráfica gerada pela implementação da biblioteca *Shiny*.

Para a análise de agrupamento *K-Means* também foi elaborado um arquivo específico em linguagem R (*dados-diarios.R*) a partir da primeira base de dados (base 1) para gerar a tabela (*dataframe*) que contém os valores médios mensais de capacidade de tráfego e fluxo diário de pares de trens para cada linha singela dos trechos avaliados.

O esquema apresentado na Figura 3 ilustra o fluxo de informação entre as bases, os arquivos em R e o pacote *Shiny* para gerar a visualização das informações do *dashboard*:



Figura 3: Esquema para o fluxo de informações entre as bases, os arquivos em R e o pacote *Shiny* para gerar o *dashboard*.

2.3 Implementação da interface gráfica dos indicadores

As aplicações desenvolvidas em *Shiny* são compostas pelos seguintes objetos:

- interface de usuário (*user interface - ui*): responsável por controlar o *layout* e os aspectos da aplicação
- função servidor (*server*): possui o encargo de

definir como o aplicativo funciona.

Estes dois objetos são passados como argumentos para a função *shinyApp* que cria o aplicativo a partir deste par de argumentos. Em complemento aos pacotes *shiny* e *shinydashboard* foram utilizadas as demais bibliotecas:

- *ggplot2*: geração dos gráficos
- *plotly*: elaboração de gráficos interativos
- *tidyverse*: manipulação e exploração dos dados
- *ggpubr*: aprimorar visualização dos gráficos
- *DT*: geração de tabelas no *dashboard*

As funções utilizadas na *user interface* (*ui*) e no *server* são detalhadas nas Tabelas 2 e 3, respectivamente:

Tabela 2: Funções utilizadas na *user interface* (*ui*)

Função	Finalidade
<i>dashboardPage</i>	Cria página de <i>dashboard</i>
<i>dashboardBody</i>	Permite incluir elementos no corpo do <i>dashboard</i>
<i>fluidRow</i>	Inclui um layout baseado em linhas
<i>box</i>	Realiza a inclusão de gráficos e conteúdo em uma caixa (<i>box</i>)
<i>tabItem</i>	Permite a criação de abas (<i>tabs</i>) para separar diferentes conteúdos no corpo do <i>dashboard</i>
<i>tabsetPanel</i>	Possibilita o uso de abas separadoras no corpo do <i>dashboard</i>
<i>tabPanel</i>	Cria um painel de abas que pode ser incluso dentro de uma função <i>tabsetPanel()</i> ou <i>navbarPage()</i>

Tabela 3: Funções utilizada no *server*

Função	Finalidade
<i>renderPlot</i>	Realiza a renderização de um gráfico reativo
<i>renderPlotly</i>	Realiza a renderização de um gráfico reativo e iterativo conforme possibilidades fornecidas pela biblioteca <i>plotly</i>
<i>renderDataTable</i>	Faz a renderização de uma tabela de dados

Deste modo, com a aplicação das bibliotecas e das funções mencionadas foi possível a elaboração do *dashboard* contendo informações sobre as variáveis mencionadas no dicionário de dados apresentado na Tabela 1. Os arquivos *dados-tempos.R*, *dados-impactos.R* e *dados-diarios.R* forneceram as tabelas de dados (*dataframes*) necessários para elaboração dos gráficos da interface do aplicativo.

Na barra lateral foram definidas duas abas principais para agrupar as categorias de indicadores:

- Trecho: Possui os seguintes subtópicos:
 - *BoxPlot* e Pareto: apresenta gráficos de *boxplot* para cada trecho selecionado e contém filtros para exposição dos

dados conforme o sentido de circulação dos trens e de acordo com o tipo de seção de bloqueio (linha singela ou pátio). Também apresenta gráficos de Pareto para os motivos de anomalias e propagações.

- Dispersão TT Diário: mostra gráficos de dispersão para a linha singela selecionada com o intuito de mostrar as correlações das seguintes variáveis com o *Transit Time* diário: tempo de condução, duração de anomalias, duração de propagação, fluxo diário em pares de trens.
- *Clustering*: apresenta o gráfico de cotovelo e o gráfico de agrupamento das linhas singelas, conforme resultado da aplicação da abordagem *K-Means*. Também apresenta uma tabela contendo a listagem das linhas singelas e seus respectivos grupos.
- *Alluvial*: mostra um par de diagramas aluviais (ou gráfico de Sankey) para mostrar as distribuições entre anomalias e propagações para cada tipo de motivo em cada trecho. O primeiro diagrama é elaborado para a frequência, enquanto que o segundo mostra a duração de cada anomalia.
- *BoxPlot* Op Norte: Apresenta gráficos de *boxplot* para todos pátios e linhas singelas, ordenados conforme a sequência de rota que os trens percorrem desde Rondonópolis/MT até Hortolândia/SP. Permite uma visão generalizada de como estão distribuídos os tempos ao longo da malha ferroviária. Contém um subtópico para cada tempo (*transit time*, tempo de condução, duração de anomalia, duração de propagação) além de apresentar separadamente em cada seção os indicadores para os sentidos de circulação exportação e importação.

2.4 Agrupamento de linhas singelas

As técnicas de agrupamento consistem em encontrar subgrupos de observações em uma base de dados. Neste trabalho foi aplicada a abordagem de agrupamento *K-Means* para determinar agrupamentos específicos de linhas singelas e a partir disso permitir a elaboração de estratégias de manutenção e controle da circulação destes grupos. Para tanto, são considerados os valores médios diários durante o mês de maio para as variáveis de capacidade de tráfego ferroviário e fluxo em pares de trens (PdT) em cada linha singela. Antes de explorar a metodologia *K-Means*, será abordado o conceito de capacidade de tráfego ferroviário e as particularidades

adotadas neste trabalho.

2.4.1 Capacidade de tráfego ferroviário

A capacidade de tráfego ferroviário é o número máximo de trens que percorrem a via permanente em um intervalo de tempo definido, sendo que neste trabalho foi adotado o período de tempo de 1 dia (24 horas). De acordo com a literatura [2], a capacidade de uma linha singela pode ser estimada a partir da fórmula apresentada na equação 1:

$$Capac = \left\lceil \frac{1440}{Ti + Tp + \theta} \right\rceil * K \quad (1)$$

em que:

Capac – Capacidade da seção em pares de trens por dia;
 1440 – Número de minutos no dia (min);
 Ti – Tempo de condução do trem no sentido importação (min);
 Tp – Tempo de condução do trem no sentido exportação (min);
 θ – Tempo de licenciamento considerando as características no sistema adotado de licenciamento e controle (min);
 K – Coeficiente ou fator redutor percentual, que varia de 60% a 80% conforme a eficiência de cada ferrovia (%).

No presente estudo, com o intuito de contabilizar a partir de dados realizados o efeito da eficiência descrita pelo coeficiente K, foram considerados os valores médios do período para os tempos de trânsito (*transit time*) dos trens no lugar dos tempos de condução. Com isso não é mais necessário multiplicar pelo coeficiente K, pois o *transit time* engloba além do tempo de condução, os tempos de ineficiências abrangidos pelo coeficiente K. Para o tempo de licenciamento adotou-se para todas as linhas singelas o valor de 7 minutos, conforme utilizado na Declaração de Rede da Rumo Logística no ano de 2022 [3].

Posteriormente, foram contabilizados para as linhas singelas, os valores médios diários da quantidade de pares de trens. Isso permitiu analisar o tráfego na linha singela em comparação com a sua capacidade estimada, o que pode gerar alguns agrupamentos conforme apresentados na Tabela 4:

Tabela 4: Características esperadas dos agrupamentos

Capacidade	x	Fluxo de tráfego
Alta		Alto
Baixa		Alto
Alta		Baixo
Baixa		Baixo

2.4.2 Algoritmo K-Means

O *K-Means* faz parte do aprendizado não supervisionado de *Machine Learning*, sendo um dos algoritmos de agrupamento não hierárquico mais conhecidos. Sua aplicação pode acontecer quando as variáveis de análise são quantitativas e a dissimilaridade é baseada na distância Euclidiana (equação 2) [4]:

$$d_{ij} = \sum_{j=1}^p (x_{ij} - x_{ij'})^2 = \|x_{ij} - x_{ij'}\|^2 \quad (2)$$

As etapas do algoritmo *K-Means* são resumidas nos passos a seguir:

- Passo 1: Para um dado codificador C , a variância total intra *cluster* é minimizada com relação a $m_1, m_2, \dots, m_k$, produzindo as médias da alocação atual.
- Passo 2: Dadas as médias atuais, a função objetivo é minimizada realocando cada observação ao cluster com a média mais próxima, ou seja, obedece a equação 3:

$$C(i) = \arg \min_{1 \leq k \leq K} \|x_i - m_k\|^2 \quad (3)$$

- Passo 3: Repetir os passos 1 e 2 até que não haja novas realocações.

Existem alguns métodos para se definir o número ótimo de *cluster* antes de se iniciar o algoritmo. No caso, o presente trabalho utilizou o método do Cotovelo (*Elbow*) que consiste em medir a variação total intra *cluster* (*total within-cluster sum of square – wss*) para distintos números de grupos. A partir disso, é possível gerar um gráfico que contém na abcissa o número de grupos e na ordenada os valores de *wss*, o que permite identificar o momento em que a adição de um grupo não melhora a variabilidade nos dados (usualmente o “cotovelo” do gráfico).

3 Resultados e Discussões

Os resultados foram divididos em três categorias, sendo a primeira relativa aos gráficos de *boxplot*, *pareto*, *dispersão* e *aluvial* disponíveis na aba “Trecho”, enquanto que na segunda categoria estão os resultados dos gráficos de *boxplot* da aba “BoxPlot Op Norte” e por fim, na terceira categoria, estão os agrupamentos das linhas singelas pela abordagem *K-Means*.

3.1 Trecho: *Boxplot*, Pareto, Aluvial e Dispersão

visualização dos pátios e linhas singelas do trecho 6 no sentido exportação.

A Figura 4 apresenta a imagem do *dashboard* com o conteúdo selecionado com aplicação dos filtros para

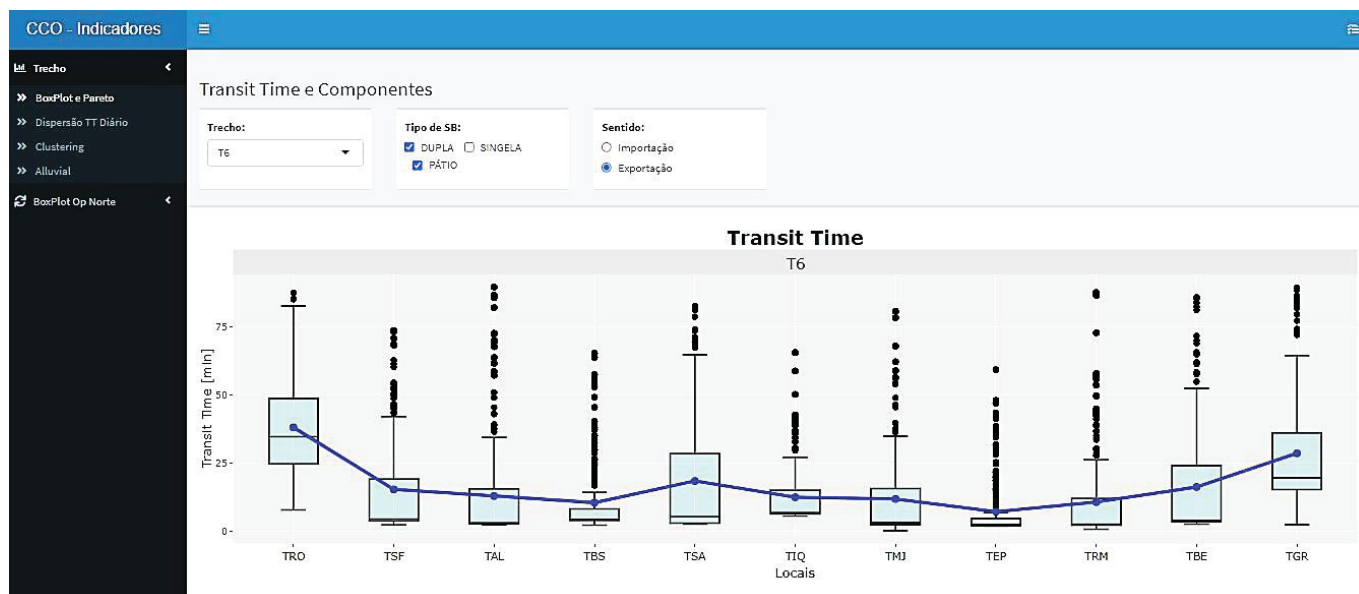


Figura 4: Imagem do *dashboard* para a aba da seção “Trecho”, a qual apresenta os gráficos *Boxplot* para a distribuição dos tempos de trânsito (*Transit Time*) ao longo dos pátios e linhas singelas do trecho 6.

Importante destacar que nos gráficos de *boxplot* existem linhas que conectam os valores médios dos tempos de cada local com o intuito de facilitar a análise e interpretação dos gráficos pelos analistas do indicador.

Também foram gerados gráficos para os tempos de

condução duração de anomalia, duração de propagação muito semelhantes ao que foi gerado para o *Transit Time* apresentado na Figura 4. Ademais, ao final deste subtópico ainda são apresentados os gráficos de Pareto para as anomalias e propagação, conforme ilustrado pelas Figura 6 e 7, respectivamente:

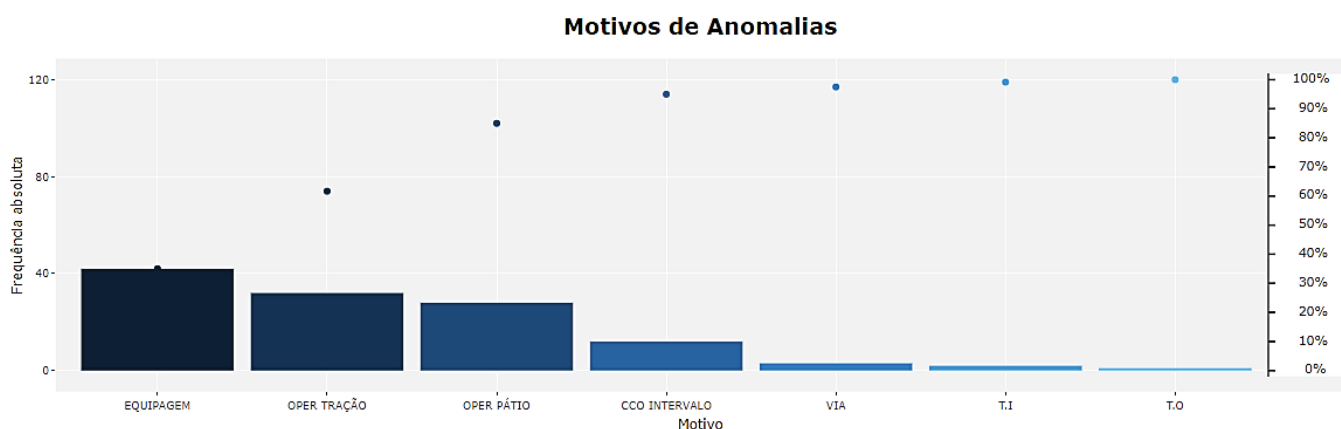


Figura 6: Imagem do gráfico de Pareto incluso na aba da seção “Trecho” no subtópico “BoxPlot e Pareto”, o qual apresenta a frequência de motivos de anomalias ao longo dos pátios e linhas singelas do trecho 6.

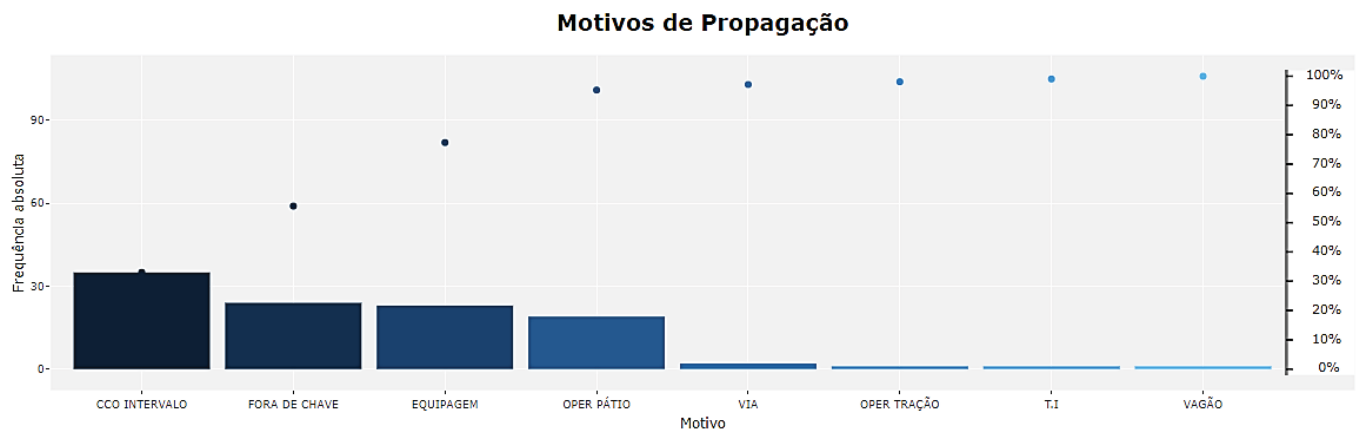


Figura 7: Imagem do gráfico de Pareto incluído na aba da seção “Trecho” no subtópico “BoxPlot e Pareto”, o qual apresenta a frequência de motivos de propagação ao longo dos pátios e linhas singelas do trecho 6.

Observa-se que no trecho 6 as maiores frequências de parada ocorrem por conta dos motivos de equipagem, operação tração e operação pátio (85% dos casos), enquanto que para a propagação deste trecho, os motivos que ocorrem com maior frequência são CCO intervalo, fora de chave e equipagem (77% dos casos).

Além disso, ainda na seção “Trecho” existe o subtópico “Alluvial” que contém diagramas aluviais que complementam as informações dos gráficos de Pareto e fornecem uma visão geral de como as frequências fluem entre os extratos de anomalia e propagação, motivos e trechos, conforme pode ser visualizado nas Figuras 8 e 9 para os valores de frequência e duração respectivamente:

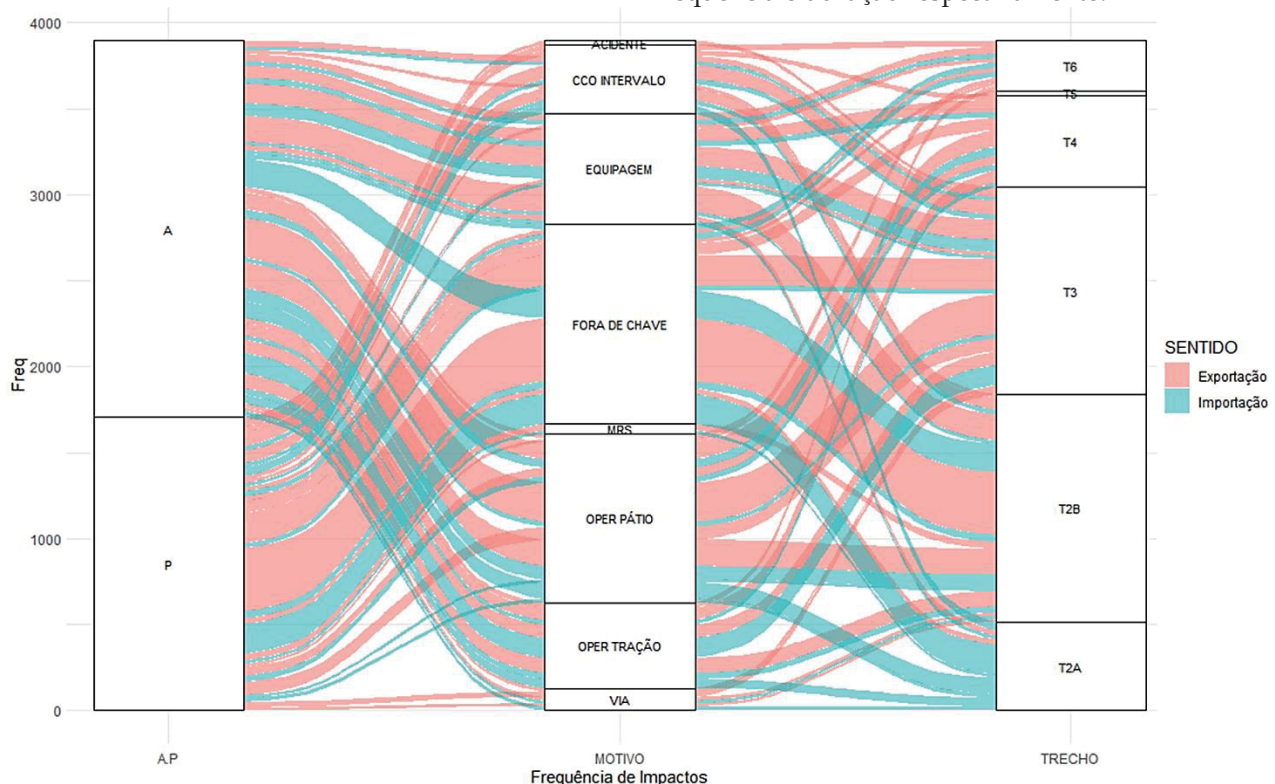


Figura 8: Imagem do diagrama aluvial de frequência de impactos com as estratificações de anomalia e propagação (A.P), motivo e trecho de ocorrência.

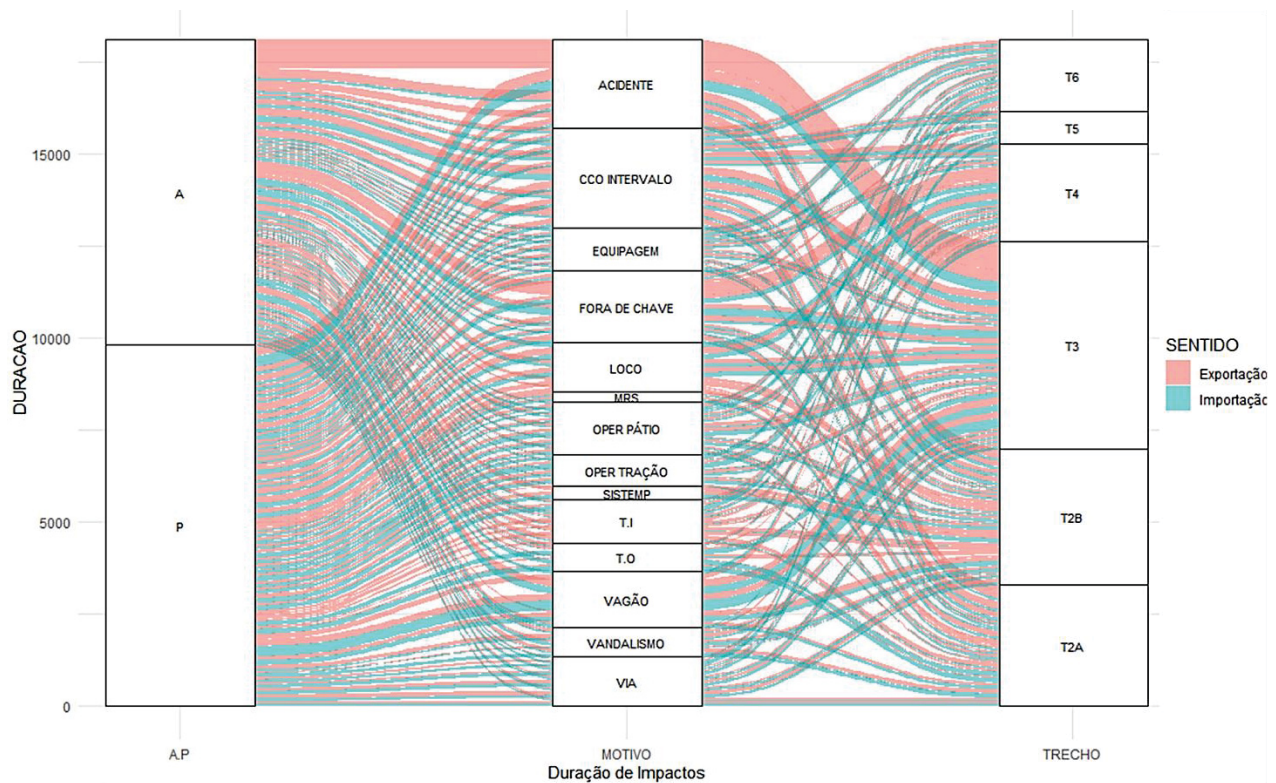


Figura 9: Imagem do diagrama aluvial de duração de impactos com as estratificações de anomalia e propagação (A.P), motivo e trecho de ocorrência.

No caso do diagrama aluvial para a frequência de impactos, pode-se observar a ocorrência de um estrato relevante de propagação com motivo fora de chave de trens no sentido exportação para o trecho 2B. Isto ocorre, porque nesta região são estacionados trens nos momentos em que o porto está congestionado e pode-se constatar que isso representou uma parcela significativa na frequência do período analisado.

Por outro lado, no diagrama aluvial para a duração dos impactos, é possível observar que a parcela de acidentes dos trens exportação no trecho 3 tem maior

contribuição em relação às demais parcelas.

Partindo para um outro tipo de análise, os gráficos apresentados no subtópico “Dispersão TT Diário” mostram as correlações com o *Transit Time* diário para as variáveis tempo de condução, duração de anomalia, duração de propagação e pares de trens por dia. Na Figura 10, é apresentado um gráfico para a via singela ZRUZGB mostrando a correlação com o tempo de condução.

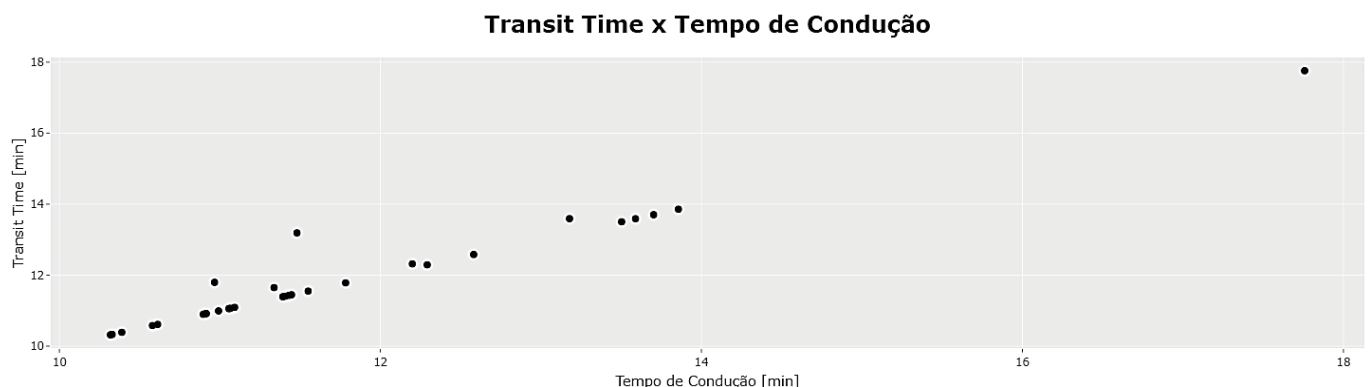


Figura 10: Imagem do gráfico de dispersão para apresentar a correlação entre o *Transit Time* e o Tempo de Condução na via singela ZRUZGB.

É possível constatar na Figura 10 uma relação com poucos desvios da linearidade entre o *transit time* e o tempo de condução, o que é positivo para a via singela, pois mostra que o tempo total gasto para o trem se deslocar nesta seção é eficiente e possui poucos impactos por paradas não programadas.

3.2 BoxPlot Op Norte

Para a aba “BoxPlot Op Norte” foram gerados gráficos do tipo boxplot para todos os trechos dos pátios analisadas no presente trabalho e para as variáveis *transit time*, tempo de condução, duração de anomalias e duração de propagação. A Figura 11 apresenta o caso para o *transit time* no sentido exportação:

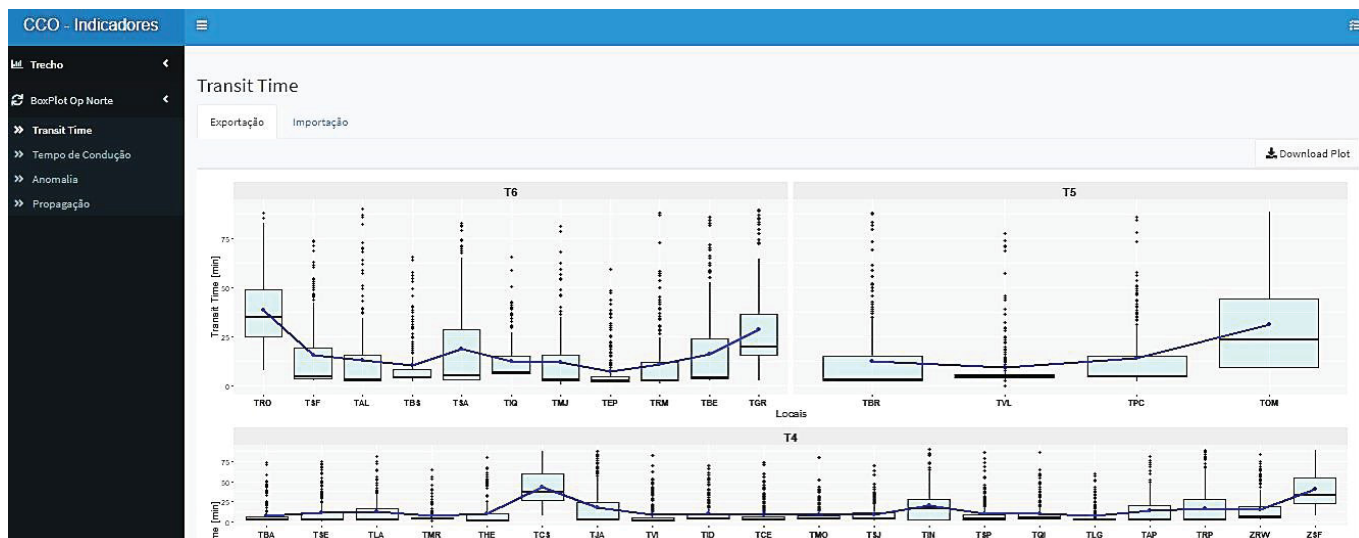


Figura 11: Imagem dos gráficos *Boxplot* para a distribuição dos tempos de trânsito (*Transit Time*) ao longo dos pátios e linhas singelas de todos os trechos de via singela nos trechos compreendidos pelo estudo.

Nota-se que existem uma grande quantidade de outliers nos *boxplots* dos trechos visualizados na Figura 11, além de também apresentarem grande variabilidade. Nota-se que os pátios em TRO, TSA, TOM e TCS apresentam maiores valores médios de *transit time*. Isto ocorre porque estes pátios ficam localizados nas adjacências de pontos de carregamentos de trens, o que mostra a interferência que estas operações causam na circulação dos trens na malha principal.

3.3 Agrupamento de linhas singelas com K-Means

Para definir o número ideal de clusters de linhas singelas, foi gerado o gráfico de cotovelo apresentado na Figura 12 e a partir dele foi possível definir que o número ideal seria igual a 3 *clusters*:

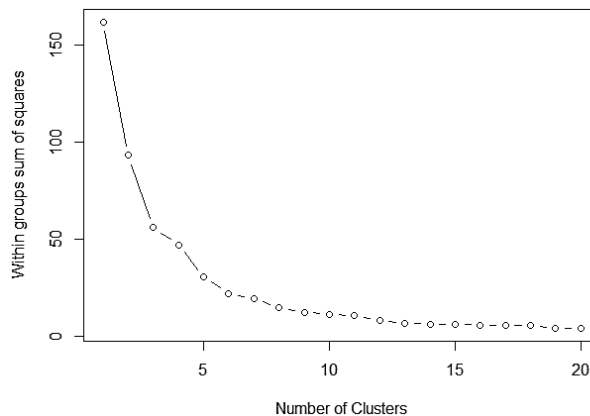


Figura 12: Esquema para o fluxo de informações entre as bases, os arquivos em R e o pacote Shiny para gerar o dashboard.

A partir disso, foi configurada a semente de número aleatório `set.seed(123)` para garantir a reprodutibilidade dos resultados. Após isso, foi utilizada a rotina *kmeans*, já disponível e implementada no R para executar o algoritmo *K-Means* que possibilitou gerar o gráfico apresentado na Figura 13 que mostra os agrupamentos de linhas singelas:

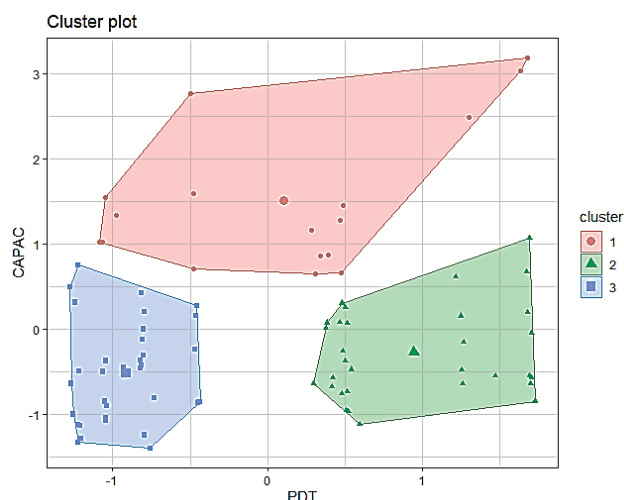


Figura 13: Esquema para o fluxo de informações entre as bases, os arquivos em R e o pacote Shiny para gerar o dashboard.

Os três grupos elencados pelo algoritmo estão dentro das características mencionadas na Tabela 3 e podem ser visualizados na Tabela 5:

Tabela 5: Características dos agrupamentos

Grupo (Cluster)	Capacidade	x	Fluxo de tráfego
1	Alta		Alto e Baixo
2	Baixa		Alto
3	Baixa		Baixo

A partir deste resultado, é possível o desenvolvimento de estratégias para manutenção e controle de anomalias em cada grupo. Por exemplo, o grupo 2 apresenta o caso mais crítico e que requer maior atenção, pois é composto por linhas singelas que apresentam baixa capacidade, mas com um fluxo de tráfego alto, sendo que pequenas anomalias nestes locais podem ocasionar um elevado impacto no *transit time* por conta da formação de filas.

O grupo 1 está em uma situação intermediária e possui situações de capacidade alta com fluxo de tráfego alto e também casos de capacidade alta com fluxo de tráfego baixo. Neste sentido, este grupo não requer tanto atenção quanto o grupo 2, mas que pode se tornar gargalo caso o fluxo de tráfego aumente ou a capacidade diminua.

Já o grupo 3 fica em uma posição que não requer tantos esforços para a contenção de anomalias, pois possui um tráfego baixo. Contudo, deve ser monitorado, pois caso o tráfego aumente ele pode se tornar gargalo.

4 Conclusões

No presente trabalho foi desenvolvida uma interface web para apresentação de um *dashboard* com indicadores de processos relevantes para a para

auxiliar a identificação de gargalos operacionais. Isso foi realizado com o apoio de gráficos *boxplot*, dispersão, Pareto e diagramas aluviais que evidenciaram locais com elevados tempos operacionais e também os motivos dos impactos de cada região. Além disso, foram definidos três grupos de linhas singelas para os quais é possível a aplicação de um conjunto de ações específicas. Portanto, o *dashboard* desenvolvido fornece importantes informações para os gestores ferroviários e contribui para a definição de importantes ações que garantam fluidez para o tráfego ferroviário.

Agradecimentos

Agradeço a Rumo Logística pelo patrocínio fornecido para a realização da especialização em *Data Science* e *Big Data* na Universidade Federal do Paraná. Também aos meus gestores diretos por sempre me fornecerem condições para o aprofundamento dos meus conhecimentos, seja com cursos de capacitação ou mesmo com disponibilização de tempo e dados para meus estudos.

Agradeço ao prof. Walmes pela orientação deste trabalho e pelas sugestões de indicadores operacionais a serem estudados.

Agradeço minha esposa Lucélia que me deu apoio durante o desenvolvimento deste trabalho e minha filha Lavínia para a qual desejo um futuro iluminado.

Referências

- [1] Rumo, "Relatório de Resultados," 2022. Acesso em: 12 de junho de 2022. Disponível em: <http://ri.rumolog.com/informacoes-financeiras/central-de-resultados/>.
- [2] T. de O. Victorino, G. Gonçalves de Matos, and J. Matsuo Shimoishi, "Capacidade de tráfego ferroviário: uma aplicação prática ao corredor ao corredor Rondonópolis Santos," in Congresso de Pesquisa e Ensino em Transporte da ANPET, 2018, pp. 681–692.
- [3] Agência Nacional de Transportes Terrestres (ANTT), "Declaração de rede 2022 - RMP," 2022. Acesso em: 12 de junho de 2022. Disponível em: <https://portal.antt.gov.br/documents/359178/0/Declaração+de+Rede+2022+-+RMP.xlsx/93a6ab8d-fd8c-b2ab-ca81-4687571cb764?t=1633695336350>.
- [4] C. Taconeli and W. Zeviani, "Análise de Agrupamento," 2020. Acesso em: 12 de junho de 2022. Disponível em: <http://cidamo.com.br/files/minicurso-cidweek-walmes.pdf>.