

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS



CATEGORIA 3

OTIMIZANDO A EFICIÊNCIA DO SISTEMA METROFERROVIÁRIO:

ANÁLISE OPERACIONAL COM CIÊNCIA DE DADOS

INTRODUÇÃO

Nos últimos anos, a crescente demanda por soluções tecnológicas que priorizam a análise de dados, tem se tornado indispensáveis para as empresas e ramos de todos os setores. Com o advento da tecnologia, trabalhos outrora extremamente manuais, estão cada vez mais autônomos, gerando um alto ganho de desempenho, reduzindo drasticamente o tempo de análise e gestão das informações, trazendo uma maior conformidade e precisão nos resultados obtidos, ajudando assim, na tomada de decisão. Ao longo deste trabalho, serão explorados os benefícios oriundos da análise de dados, tendo como modelo o perfil operacional de uma Linha metroriária que transporta em média 513 mil passageiros por dia útil, ligando a região central a zona sul da Capital do Estado de São Paulo.

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS



A Ciência de Dados é um campo interdisciplinar que utiliza estatística, matemática e computação para inferir informações valiosas de grandes conjuntos de dados. Ela envolve a coleta, processamento e análise de informações, identificando padrões e tendências que permitem tomar decisões informadas e resolver problemas complexos em diversos setores. É uma área essencial na era da informação, convertendo dados em percepções que impulsionam a inovação e o crescimento.

Para o sistema Metroferroviário a gestão das informações com aplicação de Ciência de Dados, pode auxiliar em uma simples mudança no tempo de parada nas estações, como até mesmo aquisição de uma nova frota de trens, construir uma nova estação e entre outras possibilidades.

O principal protagonista no sistema metroferroviário é o passageiro, então, tudo que é feito e pensado nos bastidores tem como premissa principal o atendimento na ponta, garantindo o melhor conforto possível, segurança e agilidade para o cliente. Neste sentido, aparecem os principais motores do sistema, que contribuem para o transporte das pessoas de um ponto a outro da cidade, são eles: Trem, Via Permanente, Rede Aérea, Sistemas das Estações e muitos outros.

Considere, se além do feedback dos clientes quanto a prestação do serviço, fosse possível também, “ouvir” o que o Trem e outros componentes do sistema “tem a dizer”. Por meio da análise dos dados que estes dispositivos emitem, é possível traçar um perfil e identificar, se existe alguma perda de desempenho no trajeto, qual o carro de uma

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS



composição é o mais carregado, qual a temperatura de cada salão, e mais uma infinidade de possibilidades.

Na Primeira Revolução Industrial, o trem era dotado basicamente de componentes mecânicos, no qual os operadores necessitavam identificar determinados padrões, ruídos e afins, para predizer se o equipamento iria falhar. Atualmente, com a Quarta Revolução Industrial, é possível identificar qual a temperatura exata de cada carro, se uma porta está prestes a apresentar falha, e ainda, saber quantos passageiros encontram-se embarcados. Tudo isso graças a comunicação constante dos equipamentos embarcados que trafegam milhares de dados por segundo, auxiliando outros dispositivos na tomada de decisão, como aplicar um freio, soar o gongo de porta, entre outros. O Frota P, um trem fabricado pela Espanhola CAF, que circula na Linha 5 Lilás, é dotado de mais de 17 mil variáveis contendo informações sobre diversos componentes.

Neste sentido, também é possível identificar informações importantes de outras partes do sistema, como as portas de plataforma, escadas rolantes, além de obter dados de bases já consolidadas e preenchidas diariamente pelas equipes com observações importantes sobre a operação, como a quantidade de clientes com mal súbito observada em um determinado dia e horário, intrusão indevida no trecho operacional e entre outros.

Ao centralizar as diversas informações operacionais em uma única ferramenta de análise, é viabilizada uma visão geral e ampliada do que de fato ocorre no sistema,

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS



trazendo um foco maior nos principais pontos a serem atacados para otimização de performance e conforto, além de possibilitar a identificação de qual o trem mais apresenta falha na operação, e nesta composição, qual o principal instigador do lapso, os momentos exatos para inserção e recolhimento dos trens para atender os passageiros quando eles mais precisam, e ainda, com base em todo o histórico, prever, qual o desgaste de trilho, o custo energético, quantas composições devem circular daqui 30, 60 ou mais dias, dentre outras análises futurísticas.

Pelo exposto, o principal objetivo deste trabalho é apresentar como aperfeiçoar ainda mais o serviço prestado com base no cruzamento de informações dos passageiros, e sistemas que contemplam a malha metroferroviária. Possibilitando ainda, a implementação de aprendizagem de máquina, para análise histórica das informações com prospecção para dias que estão por vir, traçando estatisticamente o melhor modelo a ser adotado, gerando ganho de desempenho, reduzindo possíveis desvios, aumentando a produtividade do sistema como um todo.

DIAGNÓSTICO

O trem é o ativo que transporta de fato o passageiro de um ponto a outro da linha, sendo ele, o lugar em que o cliente fica mais tempo dentro do sistema, então, “ouvir” o que a composição tem a dizer é importante para definir as melhores estratégias a serem adotadas para aumentar o conforto do usuário. Considere o exemplo a seguir:

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS



Figura 1 – Análise de Lotação Por Carro (Ultima Quinta-Feira de Junho)

Esta é uma análise baseada no carregamento do Pico Manhã, na Plataforma 2 de Vila das Belezas da Linha 5 Lilás. Nota-se que mesmo com o efeito da pandemia em 2021, até os dias atuais, o carro 4, é o mais carregado. Na média, as três análises ficam abaixo do limite permitido nesta linha, de 6 pass/m² atualmente e 2 pass/m² no período pandêmico, porém, enquanto os passageiros dos carros 1 e 2 desfrutam de um maior nível de conforto, o cliente no carro mais carregado experimenta uma sensação de conforto inferior.

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS

Olhando apenas pelo ponto de vista dos passageiros no carro 4, a sensação é de que todo o trem está altamente carregado e que é preciso aumentar a frota neste momento, porém, analisando o que a composição nos mostra, há espaço ainda.

A seguir, o mesmo exemplo observado, agora em um trem da Série 8500, que circula no sistema Ferroviário da Linha 9 Esmeralda.

Exemplo de Sensação de conforto

5,5 pass/m²



4 pass/m²



Figura 2 – Análise de Lotação Entre Carros de um Mesmo Trem.

Na imagem da esquerda, nota-se que todos os espaços foram preenchidos, para que o passageiro deixe a composição, o contato físico com os demais é impossível de não ocorrer. Já na imagem da direita, observa-se um distanciamento entre as pessoas, indicando que para desembarcar do trem, o contato físico pode ser evitado.

Com base nestas premissas, é possível adotar estratégias para melhorar a distribuição do fluxo nas plataformas, alterando o sentido de uma escada rolante, incluindo direcionadores entre outras possibilidades, fazendo com que os passageiros estejam bem distribuídos entre os carros do trem, garantindo um conforto melhor para todos.

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS

Para gerar essa única análise, foi necessária a coleta dos dados obtidos da bolsa de ar dos trens. Este equipamento auxilia a composição a manter o equilíbrio do carro principalmente em regiões de curvas acentuadas ao longo da via. Existe no interior da bolsa uma célula de carga, que mede o peso dos carros, este componente, auxilia o sistema do trem também na tomada de decisão para aceleração e frenagem, ou seja, quanto maior o peso, maior será a taxa de aceleração para partir da estação, bem como, a frenagem deverá ser antecipada garantindo a parada correta na plataforma.



Figura 3 – Exemplo de Bolsa de Ar

Para obtenção da quantidade de pessoas nos trens, ao angariar a informação de peso registrada, é necessário remover a “Tara do Trem” que é o valor da massa da composição sem passageiros, subtrair do valor total obtido do trem, e após, dividir pela média do peso das pessoas.

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS



$$\text{Passageiros} = \frac{\text{Peso Total do Trem} - \text{Tara do Trem}}{\text{Kg médio dos passageiros}}$$

Por fim, após obter o número médio de pessoas, para saber a lotação, basta subtrair o número de assentos do interior de todo trem e/ou de cada carro, e dividir pela área útil disponível para transporte.

$$\text{Pass}/\text{m}^2 = \frac{\text{Valor Médio de Pessoas no Trem} - \text{Assentos (Cadeiras)}}{\text{Área Útil Disponível Para Transporte em m}^2}$$

Este diagnóstico, foi apenas uma porção pequena, de toda a Ciência de Dados que é possível aplicar, considerando como premissa as informações que os sistemas transmitem o tempo todo.

ESTUDO DE CASO – Implantação do Sistema e Análise Operacional na
Linha 5 Lilás.

O ponto mais curioso de um dado, é que ele sempre estará lá, basta que seja desenvolvida uma forma de coletá-lo, e após, um método para trabalhá-lo e obter resultados. Em meados de 2019 na Linha 5 Lilás, a metodologia aplicada para obtenção do carregamento e nível de conforto era baseada em pesquisa de Origem e Destino interna e acompanhamento amostral em campo nas composições. Essa pesquisa era cruzada com os dados de bloqueios, que nesta linha para a maior parte das estações, não é em tempo real, a informação é obtida manualmente ao término da operação comercial, e disposta apenas como valor total consolidado do dia. Ou seja, também era

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA

10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO METROFERROVIÁRIOS

necessária uma pesquisa para identificar o horário de maior usabilidade do sistema o chamado “Pico”.

A seguir um exemplo de resultado obtido com a metodologia de 2019.

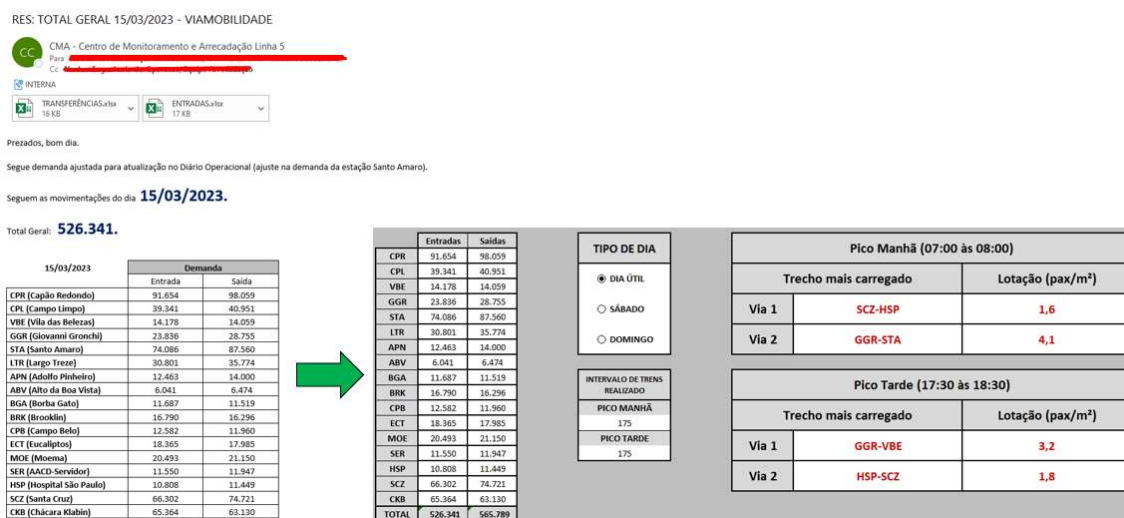


Figura 4 – Análise De Lotação Com a Metodologia de 2019

Basicamente, ao receber os dados de demanda das estações do dia anterior, as informações eram transferidas para uma Base do Excel, que calculava uma estimativa para os trechos mais carregados, sem evidenciar uma faixa horária precisa, e por se tratar de pesquisa, sempre estes mesmos trechos seriam os mais carregados.

Em 2020, com a deflagração da Pandemia de COVID-19, era necessário um refinamento das análises de lotação, para garantir o distanciamento social dentro das composições, e para tanto, era necessária uma análise por carro, então, passou-se a olhar para o trem, identificando que as composições já transmitiam informações relevantes a todo momento, bastando apenas desenvolver um método para colhê-las.

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS

Então, no ano de 2020, foi desenvolvida uma tecnologia capaz de obter e armazenar os dados do trem, já aproveitando os sistemas e recursos que a composição e seu fabricante dispuseram.

Foi desenvolvido um dispositivo que contém uma memória flash, que armazena variáveis pré-selecionadas e colhidas das composições. As informações são preenchidas pelo próprio trem em um arquivo de extensão .CSV com granularidade de no máximo 1 em 1 segundo. A tecnologia é composta por uma Aplicativo que colhe os dados deste dispositivo, sem a necessidade de acesso a cabine do trem.

Dispositivo Para Armazenamento do Dado



App Para Coleta do Dado



Figura 5 – Dispositivo de Coleta de Variáveis do Trem

As variáveis do trem, tem como objetivo o trato de manutenção do sistema, que requer uma análise técnica voltada para Material Rodante (MRO), uma vez que as informações servem para prever algum desvio, lapso, entre outras possibilidades.

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA 10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO METROFERROVIÁRIOS

Dentre as informações do trem, uma em especial, é amplamente utilizada para análise de carregamento: o peso das bolsas de ar, que mesmo sendo um dado para tomada de decisão do sistema interno do trem, se bem tratado, pode ser tão preciso quanto outras metodologias para obtenção do perfil de lotação do modelo. Então, por intermédio do dispositivo desenvolvido, passou-se a analisar as variáveis voltadas para o perfil da operação comercial, como por exemplo: Velocidade, Odometria, Data, Hora, Peso do Carro, retenção de portas, dentre outras.

A partir do tratamento das informações, foi possível entender o comportamento real da linha, observando curvas, e possibilidades de melhoria. A seguir um exemplo de carregamento de março de 2021.



Figura 6 – Análise de Lotação do Trem 12 em 25/03/2021

A partir daí, foi possível entender a distribuição por carro, bem como, avaliar qual o trecho mais carregado, auxiliando na reorganização do fluxo nas plataformas, garantindo o distanciamento social naquele momento.

Apenas com os dados dos trens já se observava um ganho expressivo na qualidade das análises que eram realizadas. Por exemplo, a seguir, é apresentada a diferença entre o

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS

método de 2019 por pesquisa e o método por peso do trem, já considerando dados de 2023.

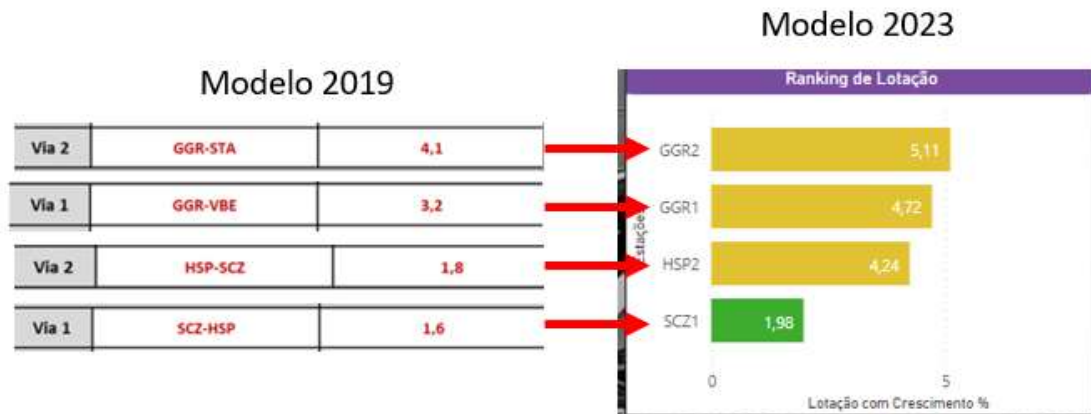


Figura 7 – Análise de 15/03/2023

Há desvios de até 2,44 pass/m², ou 422 passageiros entre as metodologias, isso ocorre, pois o peso do trem identifica o comportamento real, respeitando a sazonalidade do dia, o perfil do dia tipo dentre outras premissas, já a pesquisa, adota apenas o perfil estabelecido para o dia em que ela foi concebida, ou seja, se considerar uma pesquisa realizada em um sexta-feira, todos os parâmetros obtidos deste modelo, terão características daquele dia tipo pesquisado.

A seguir é apresentado na prática o impacto na mudança de metodologia do peso entre 2020 e 2022.

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA 10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO METROFERROVIÁRIOS



Figura 8 – Comparativo Entre Oferta e Demanda entre 2020 e 2022

Considerando as Ofertas de Picos e Vales, nota-se que em 2020, conforme a demanda aumentava, a oferta crescia na mesma proporção, isso ocorria, pois, no início da pandemia não se tinha a informação de distribuição dos passageiros pelos carros, então, uma simples foto, já seria um forte indício para o aumento de frota. A partir de abril de 2021, passou-se a ter noção da distribuição entre carros, o que possibilitou o ajuste da oferta, garantindo o distanciamento social, e em novembro de 2021, já se tinha a programação de trens fluando junto com a demanda, respeitando a necessidade real do cliente, e mesmo com o fim das restrições em março de 2022, observa-se a manutenção do novo modelo de oferta e demanda da Linha 5 Lilás.

A partir dos dados do trem, foi possível olhar outros sistemas e unificar as informações para entender o perfil operacional como um todo. A seguir uma análise que compõe os dados dos trens, demanda dos bloqueios e diário operacional.

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS

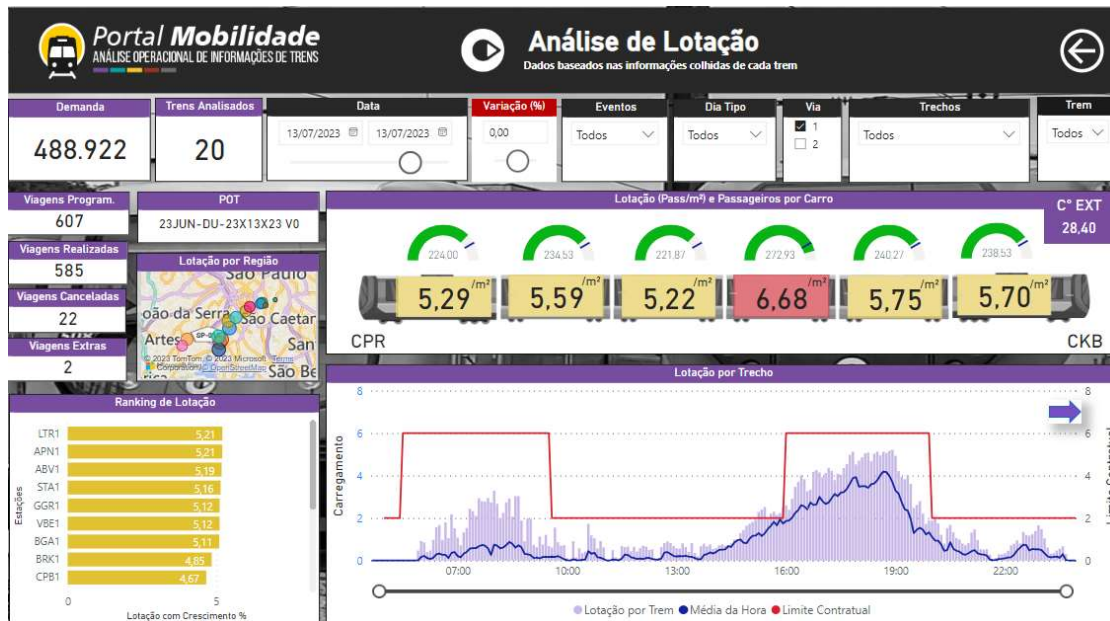


Figura 9 – Análise Operacional 13/07/2023

Na figura 9 nota-se que o trecho mais carregado do dia ocorreu entre Largo Treze e Santo Amaro, tendo sido a média máxima de lotação da composição mais densa de 5,21 pass/m², onde neste trem, houve um carro acima do 6 pass/m². Neste dia, observa-se ainda, o perfil de carregamento sazonal, tendo seu ápice as 18h55. Além dos dados do trem, nota-se que a demanda contabilizada nos bloqueios foi de 488.922, a oferta era de 23 trens nos picos e 13 no vale tarde, entre outras informações.

Dentre as análises com dados do trem, é possível observar ainda o perfil de velocidade da composição mais carregada no pico tarde do dia 13/07/23, notando, se houve algum lapso, bem como, analisar as velocidades máxima e média empregadas, além de comparar o comportamento com o perfil dos demais trechos e viagens anteriores.

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS

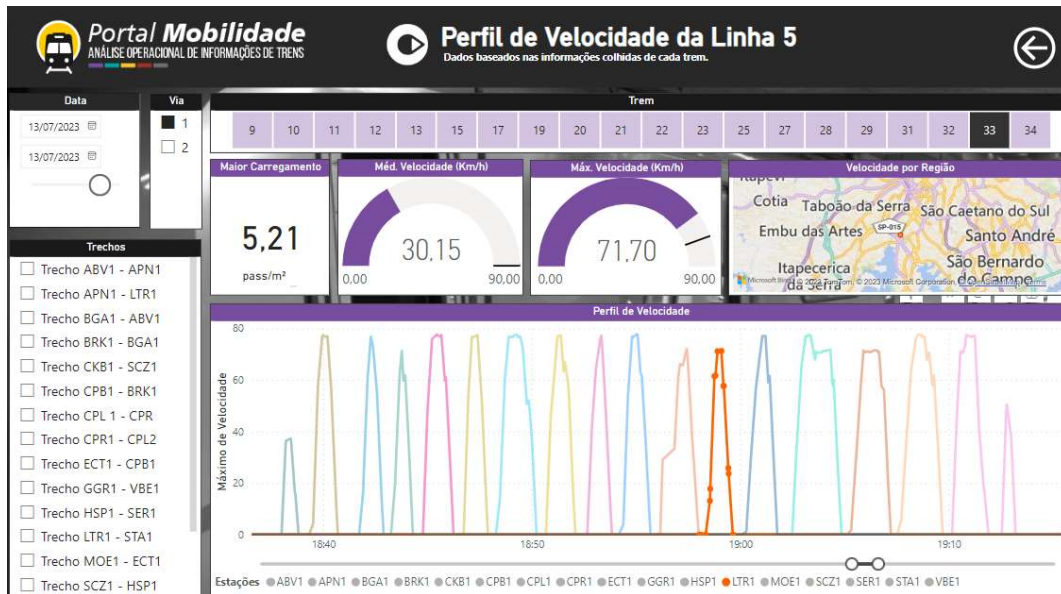


Figura 10 – Perfil de Velocidade do Trem Mais Carregado

Os dados do trem foram apenas o ponta pé inicial, após, passou-se a observar também outros sistemas, a fim de aperfeiçoar ainda mais o serviço prestado, com isso, vieram outras informações de forma centralizada numa única tela de análise, como, registros diários para eventos de Portas de Plataforma, Dispositivos de Via, Sinalização e entre outros, que anteriormente eram analisadas de forma separada e sem considerar a correlação entre algumas das variáveis. A seguir uma análise mais detalhada de todo o sistema para o dia 13/07/2023.

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS



Figura 11 – Relatório Operacional Do Dia 13/07/2023

Das viagens canceladas, o maior índice foi de Material Rodante, seguido pelas Portas de Plataforma. Nesta análise contabilizou-se também qual a maior incidência, se é de Operação, ou seja, relacionadas ao atendimento direto ao usuário, ou de Manutenção, que é voltada para lapsos pontuais de algum equipamento. É possível granular ainda mais a análise, sabendo qual a Porta de Plataforma, sua estação, podendo ainda, expandir a avaliação para o mês ou o ano inteiro.

Com todas essas informações compondo uma base histórica, é possível projetar estimativas para os dados e dias que estão por vir, observando qual a perspectiva de custo de energia de tração para o mês seguinte, qual a estimativa de desgaste de trilho esperado, quantos trens devem circular, quantos passageiros, entre outros. Abaixo mostra-se a projeção de tráfego para julho de 2023.

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS

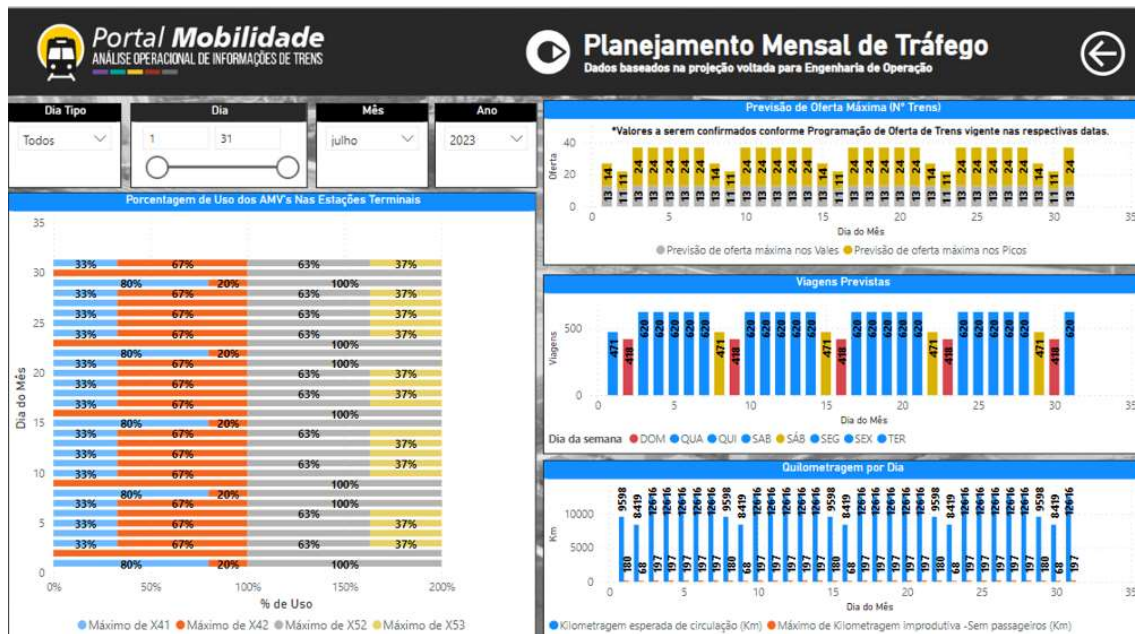


Figura 12 – Projeção de Tráfego Para julho de 2023.

Neste visual, já no início do mês, sabe-se qual a oferta de trens prevista para a última semana, quantas viagens são esperadas, qual o Aparelho de Mudança de Via (AMV) será mais utilizado e a rodagem por dia. Sendo possível ainda, filtrar mais períodos, sejam eles históricos ou futurísticos. Nos mesmos moldes, existem ainda a projeção das estimativas de Custos Variáveis como Insumos de Material Rodante, Desgaste de Trilho e Rede Aérea, Consumo de Energia de Tração, Custo Total de Produção do Carrossel, além das projeções de Demanda e Lotação, como Carregamento por Trecho, Demanda, Velocidade Média, Lotação Nos Picos e muitas outras projeções.

Com uma base tão robusta, é possível comparar os resultados realizados com as projeções, a fim de efetuar correções, fazendo o modelo ser ainda mais assertivo. Abaixo é apresentado o comparativo consolidado de junho de 2023 para viagens.

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS

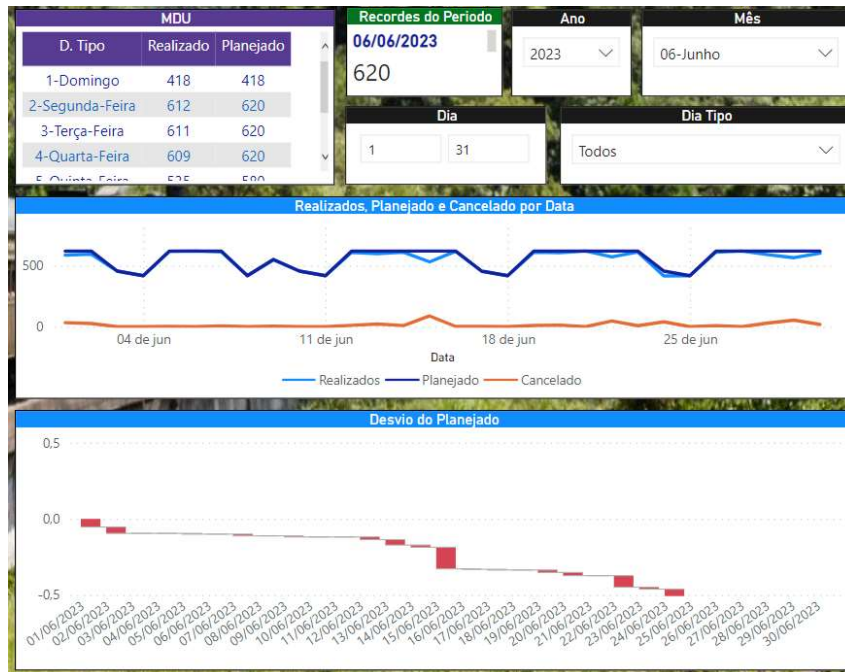


Figura 13 – Comparativo de Viagens para junho de 2023.

Com esta análise, é possível identificar o distanciamento, aplicando correções ao método, viabilizando uma assertividade maior para os próximos meses. A observa-se seguir como ficaram os ajustes aplicados ao mês de julho de 2023.



Figura 14 – Gráfico do Comparativo de Viagens para julho de 2023.

Nota-se que houve um acompanhamento quase igual entre o Realizado e o Planejado. Isso ocorreu pois, com base nas interferências de junho, o modelo identificou quais os

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS



possíveis impactos poderiam ocorrer, levando em conta o mês anterior e o perfil do ano de 2022, e corrigiu a projeção.

Até o presente momento, ao longo de um período de 2 anos e meio de estudo, houve êxito em reunir informações provenientes de diversas fontes e, por meio de uma ferramenta hábil, lograr a tarefa de integrá-las em bancos de dados consolidados. Essa integração permitiu a transformação de fontes de dados inicialmente desconexas em um conjunto de informações coerentes, as quais foram apresentadas de maneira visualmente envolvente e interativa.

ANÁLISE DOS RESULTADOS

Pelo exposto, nota-se que, com o levante das informações obtidas dos sistemas que compõem a malha metroviária da Linha 5 Lilás, foi possível cruzar os dados, para gerar análises mais precisas e detalhadas do comportamento do sistema como um todo. Essas ações auxiliam e muito nas tomadas de decisões, bem como, alimentam outras frentes, como por exemplo o estudo de fluxo e multidões, que através da simulação computacional, traz resultados análogos ao cenário real, possibilitando ajustes sem impactar negativamente a rotina do passageiro.

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS

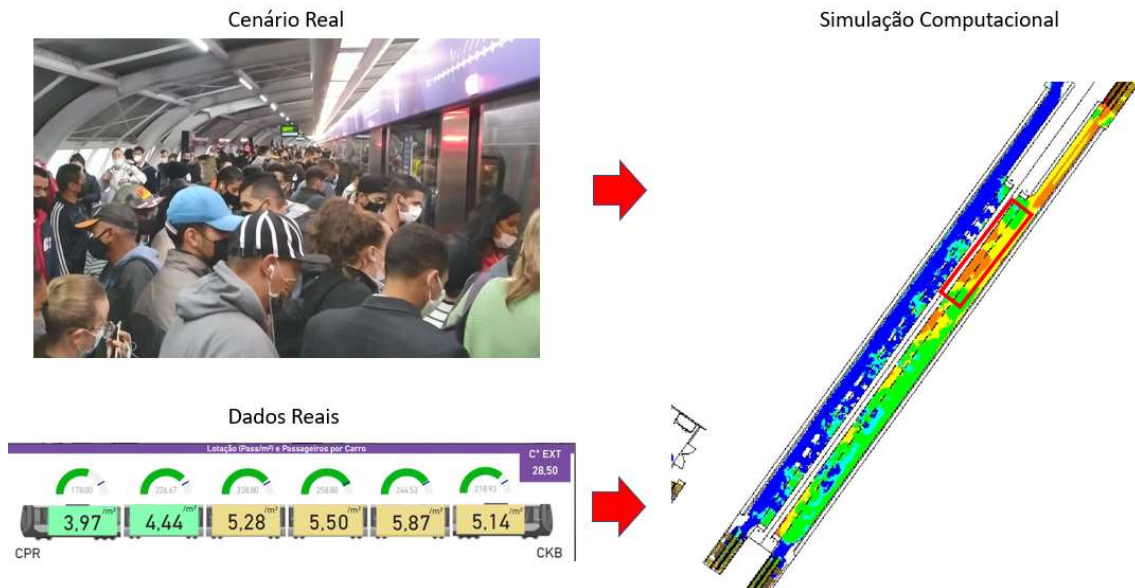


Figura 15 – Simulação Computacional Com Dados Reais Dos Trens

As informações obtidas não se limitam apenas aos estudos de fluxo; elas também têm utilidade em diversas outras áreas, como avaliação de Frota Futura, Estudos de Oferta e Demanda, Simulação de Cenários de Extensão da Linha, entre outras. Com o auxílio da análise desses dados, é possível projetar o crescimento ou a queda da lotação, o que desempenha um papel crucial na formulação de uma nova tabela horária. Isso, por sua vez, permite a elaboração da Programação de Oferta de Trens (POT) para dias específicos da semana. Abaixo, é apresentado um modelo aplicado na Linha 4 Amarela, considerando o mesmo método de análise mencionado neste documento.

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS

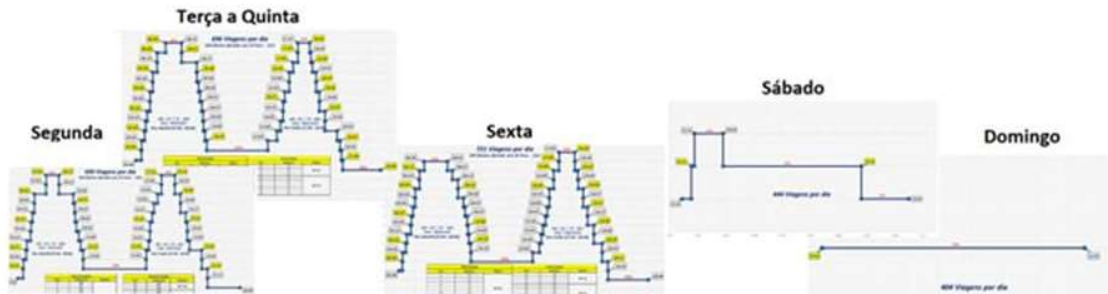


Figura 16 – Tabelas Horárias Por Dia Tipo Aplicadas na Linha 4 Amarela

O mesmo parâmetro foi aplicado nas linhas do Metrô Bahia e Linha 5 Lilás, possibilitando a identificação de padrões específicos para o uso do sistema ao longo da semana. Por meio de informações mais precisas e detalhadas, foi possível observar variações significativas nos padrões de demanda. Por exemplo, constatou-se que o sistema apresenta menor utilização nas segundas-feiras em comparação com os dias de terça a quinta-feira, apresentando um pico menor de atividade. Já às sextas-feiras, o pico é mais espaçado durante a manhã e adiantado no período da tarde.

Essa análise minuciosa foi viabilizada pelo cruzamento de dados de demanda, carregamento dos trens e ocorrências operacionais, entre outros fatores relevantes. A integração dessas informações possibilitou compreender de forma mais abrangente e precisa o comportamento do sistema de transporte em diferentes cenários, contribuindo para a otimização do planejamento das linhas mencionadas.

A DEFINIÇÃO DAS SÉRIES TEMPORAIS

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS



A observação cronológica da carga dos carros dos trens define uma série temporal. Uma série temporal pode ser pensada como uma lista de medidas indexadas pelo momento quando cada medida foi aferida.

A linha 5, atualmente, tem um total de 17 estações que vão de Capão Redondo até a Chácara Klabin. Define-se a via 2 como o sentido Capão Redondo-Chácara Klabin e a via 1 no sentido oposto Chácara Klabin-Capão Redondo. O trecho é o caminho entre duas estações subsequentes, por exemplo o trecho CPR2-CPL2 (Capão Redondo-Campo Limpo) na via 2 e o trecho oposto na via 1 CPL1-CPR1. Totalizando 34 trechos, sendo 17 entre as estações na via 1 mais 17 trechos entre as estações no sentido via 2.

A resolução das séries temporais foi determinada em intervalos de 5 minutos. Nesse intervalo mais de um trem pode passar por um único trecho. Agrupa-se então os pesos registrados em todos os trens que passaram por esse trecho nesse intervalo de tempo. Dessa forma a série temporal encerra informações da demanda daquele trecho específico nesse intervalo de 5 minutos. Além disso, tem-se uma série temporal para cada carro do trem, totalizando 6 séries temporais em cada trecho e 204 séries temporais em toda a linha 5.

Nessa resolução temporal um dia terá 288 pontos e um ano terá um total de 105.120 pontos para cada série temporal. Sendo assim o primeiro passo desse trabalho foi arquitetar uma lógica de centralização e padronização dessa quantidade de dados massiva para o projeto.

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS



A figura abaixo exibe uma amostra de 13 dias da série temporal do trecho GGR1-VBE1 (Giovanni Gronchi-Vila das Belezas).

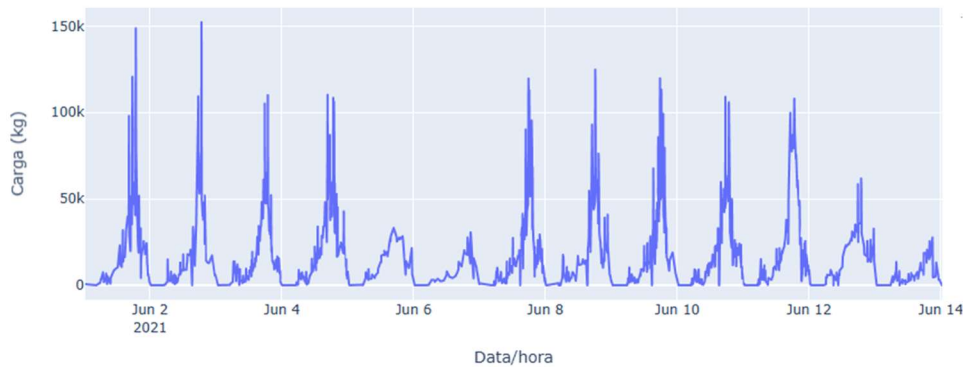


Figura 17 – Série temporal da carga do trem do trecho GGR1-VBE1.

ANÁLISE EXPLORATÓRIA

Com o objetivo de aplicar algoritmos de aprendizado de máquina (machine learning) para prever o comportamento futuro dessas séries temporais (forecasting), realizou-se a análise exploratória dos dados dessas séries. Essa análise das séries temporais é o esforço de extrair informação estatística significativa desses dados, tanto para diagnosticar o comportamento passado, quanto para prever o comportamento futuro. Informações que são essenciais para o ajuste dos modelos de forecasting.

Essa análise foi realizada em duas etapas: primeiramente usando métodos de análise exploratória familiares a todos os tipos de dados. Em seguida métodos exploratórios específicos de séries temporais.

Foi analisada a carga de passageiros na linha 5 entre 7 de outubro de 2022 e 19 de junho de 2023. A figura abaixo exibe o uso das estações em ambos os sentidos da linha 5.

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS



Observa-se que as estações terminais são menos usadas, assim como algumas estações são mais demandadas em um sentido do que no outro.

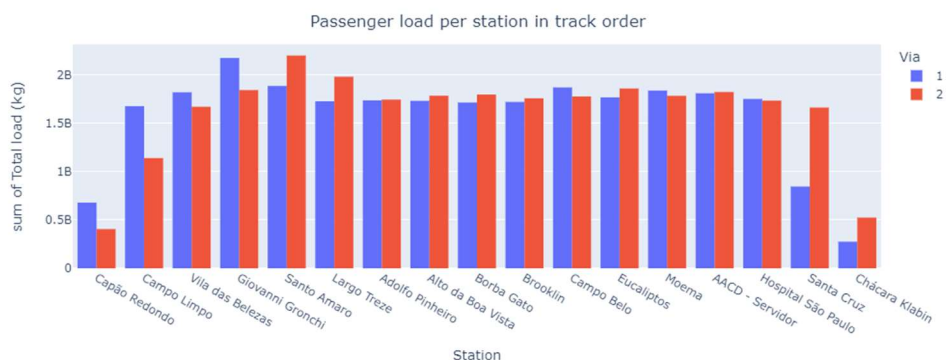


Figura 18 – Carga de passageiro por estação na ordem da via 1 e via 2.

Agrupando esses dados cronologicamente observa-se o quanto diminui a carga de passageiros durante os fins de semana (Figura 19).

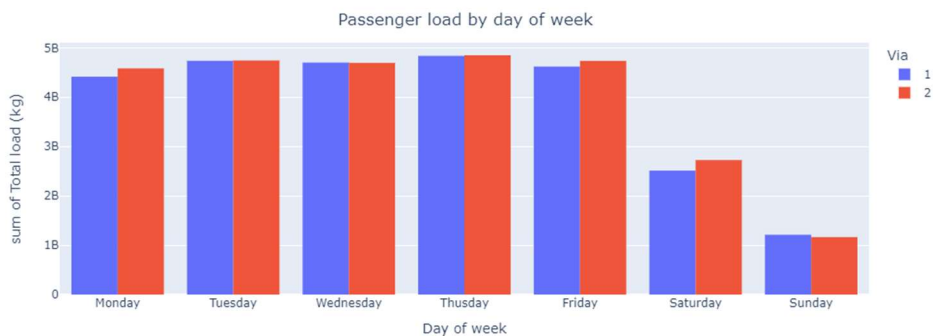


Figura 19 – Carga de passageiro acumulada nos dias da semana.

Quando se faz a amostragem horária observa-se que a via 1 é mais demandada no final da tarde enquanto a via 2 é mais demandada no início da manhã (Figura 20).

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS

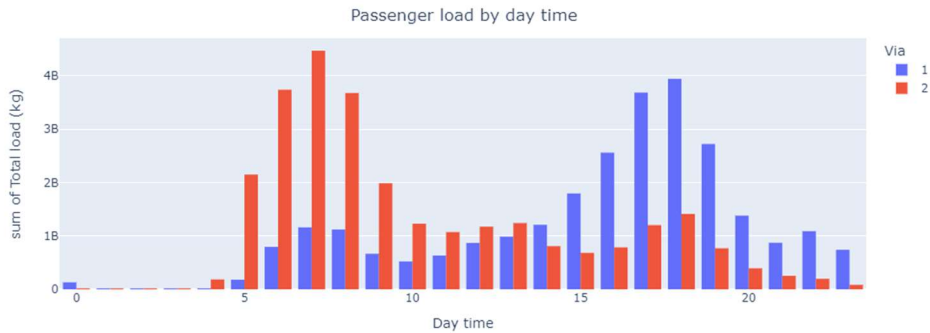


Figura 20– Carga de passageiro acumulada nas horas do dia.

A figura abaixo exibe uma sobreposição das séries temporais diárias de 2023 do trecho STA2-LTR2 agrupadas mensalmente. A sobreposição das séries mostra que há um padrão sazonal que se repete diariamente. Esse tipo de contemplação é essencial para a construção dos modelos de machine learning e devem ser investigados.

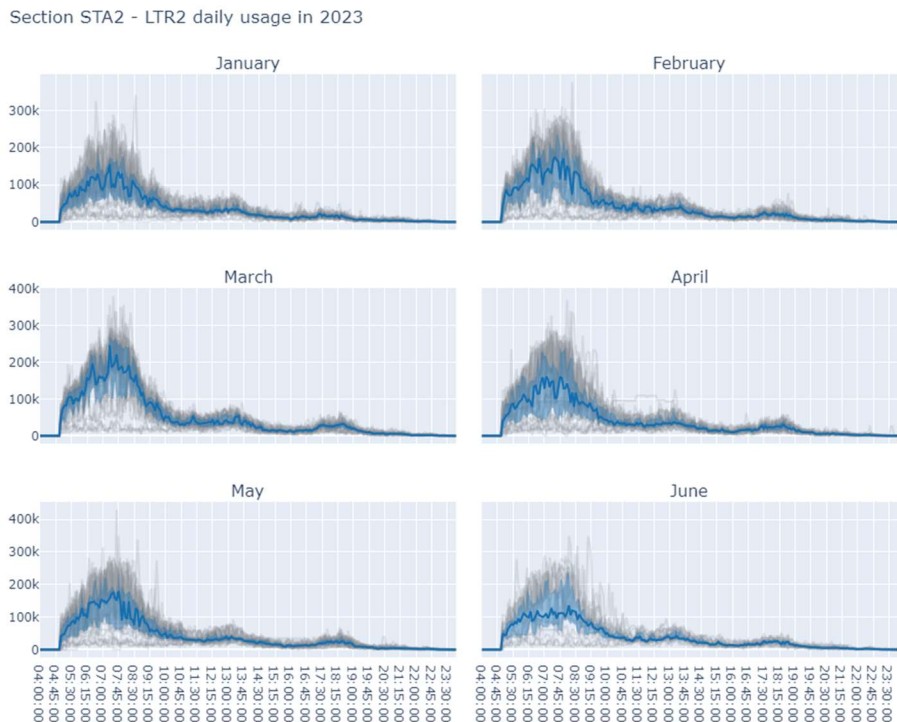


Figura 21 – Agrupamento diário das séries temporais de 2023 do trecho STA2-LTR2.

A análise seguiu realizando o teste de Dickey-Fuller que indica a presença de comportamento de tendência estocástica nas séries temporais. Juntamente foi realizado o teste de autocorrelação e autocorrelação parcial, empregados para verificar a dependência dos dados da série com dados defasados, isto é, o quanto que a ocupação de um trecho num instante depende das ocupações anteriores.

Realizou-se esses testes nas séries com resolução de 5 min, e agrupadas nas resoluções de 1 hora, 1 dia e 1 semana. A figura 22 exibe os resultados desses testes de uma série na resolução de 1 hora. Os resultados afirmam que as séries são estacionárias em torno de uma forte tendência. Os testes de autocorrelação indicaram que as séries têm autocorrelação na defasagem de 70 minutos verificada na série com resolução de 5 minutos, 24 horas na amostragem horária e de 7 dias na amostragem diária. Na amostragem semanal o teste não indicou nenhuma autocorrelação.

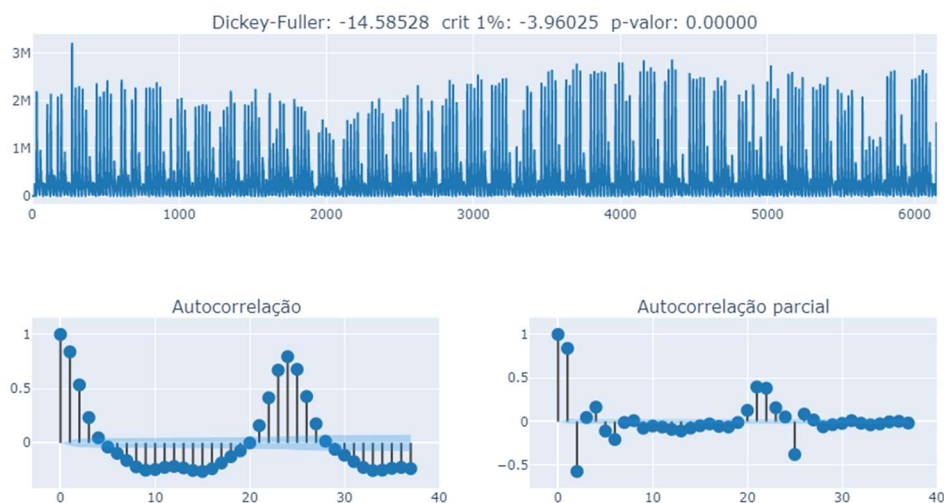


Figura 22 – Testes da série temporal do trecho STA2-LTR2 na amostragem horária.

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA 10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO METROFERROVIÁRIOS

Os dados de uma série temporal exibem padrões que normalmente só podem ser apreciados na decomposição dessa série. Usualmente decompõe-se uma série temporal em tendência mais componente cíclica, sazonalidade e ruído. A figura abaixo exhibe a decomposição da série do trecho STA2-LTR2.

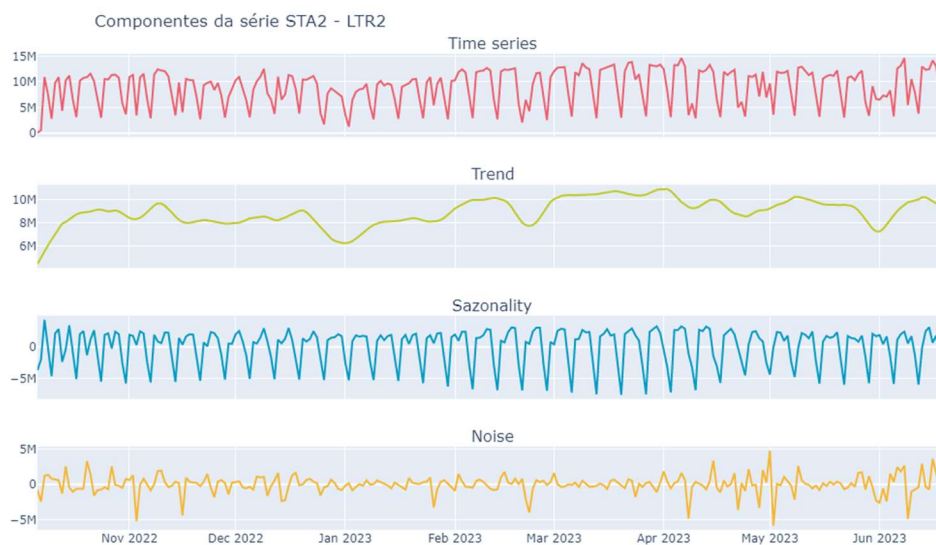


Figura 23 – Componentes da série temporal do trecho STA2-LTR2.

Essa decomposição permite ver bem o efeito da tendência, que é uma mudança do regime da série em longo prazo, além de facilitar a visualização da componente sazonal e do ruído. Essa decomposição indicou que a sazonalidade dessas séries é do tipo multiplicativa, informação essencial para os modelos de aprendizado.

CONSTRUÇÃO DO MODELO DE PREVISÃO DE DEMANDA DE TRENS

Na construção do modelo de previsão deve-se ter em mente algumas questões essenciais que vão indicar a previsibilidade dessas séries temporais:

- Os fatores são conhecidos?

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA 10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO METROFERROVIÁRIOS



- Há muitos dados disponíveis?
- O comportamento do passado é similar ao do presente?
- As previsões afetam o que se pretende prever?
- Quão distante é o horizonte a ser previsto?

É importante listar e dimensionar todos os fatores que compõem o modelo. Esses fatores alimentam o estudo na forma de séries temporais covariáveis. Então, se começa desenvolvendo um modelo de base que servirá de referência de desempenho para cenários futuros. Nele serão adicionadas covariáveis do calendário: dias da semana, finais de semana, feriados e horário de operação da linha. Futuramente é possível adicionar covariáveis externas ao calendário como eventos esportivos e culturais próximos a uma estação, clima etc.

Observou-se anteriormente que esses dados possuem uma forte tendência e sazonalidades bem definidas, que são indicativos fortes que o comportamento dessas séries no passado é similar ao comportamento no presente.

Deve-se ter em mente que no caso desse modelo as previsões vão afetar aquilo que se quer prever. No momento em que se colocar em prática essas previsões para ofertar a quantidade de trens necessárias na linha, pode ocorrer de não se observar mais uma quantidade demasiada de passageiros por metro quadrado prevista pelo modelo.

O horizonte de previsão será de 1 semana (7 dias), que parece ser um horizonte curto. Mas quando se considera a resolução de 5 minutos que fornece 288 pontos por dia, esse

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS



horizonte de previsão de 1 semana se torna um horizonte de 2016 pontos. O erro associado e o intervalo de confiança, com certeza vão aumentar ao longo desses 2016 pontos. Mas como é uma série de forte tendência e de sazonalidade marcante há artifícios que podem ser usados para atenuar o erro dessa previsão.

Para esse modelo de base foi utilizado o XGBoosting, uma biblioteca otimizada que implementa o framework de Gradient Boosting em seu algoritmo. O Gradient Boosting é uma técnica usada tanto para regressão quanto classificação onde agrega modelos mais fracos, tipicamente árvores de decisão para a construção de um modelo mais robusto e flexível.

Seguiu-se construindo a tabela de covariáveis essenciais para o modelo, usando os dados de calendário. Dividiu-se em dois grupos: covariáveis do passadas e covariáveis futuras. As covariáveis passadas são séries cronológicas cujos valores passados são conhecidos no momento da previsão. Essas séries contêm frequentemente valores que têm de ser observados para serem conhecidos. As covariáveis futuras são séries cronológicas cujos valores futuros são conhecidos no momento da previsão. Mais precisamente, para uma previsão efetuada no momento t para um horizonte de previsão n , os valores nos momentos $t + 1, \dots, t + n$ são conhecidos. Frequentemente, os valores passados das covariáveis futuras também são conhecidos. Para esse modelo de referência serão utilizadas as informações de calendário tanto para as covariáveis passadas e futuras: horário de funcionamento, dias da semana, fins de semana e feriados.

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS



Anteriormente apresentou-se os valores dos testes de autocorrelação que vão definir os parâmetros de defasagem do modelo (lag). Esses valores serão -13 (70 minutos), -288 (um dia) e -2016 (uma semana). Nas covariáveis passadas deve-se aplicar o lag de 1 dia e as covariáveis futuras usarão o valor no instante que se pretende prever, equivalente à defasagem igual a zero.

Para avaliar o desempenho do modelo foi realizada uma validação cruzada. Inicialmente definiu-se o período inicial para treinar o modelo de 180 dias (entre 7 de outubro de 2022 até 4 de abril de 2023). Serão previstas todas as semanas entre 4 de abril de 2023 e 19 de julho de 2023 e comparar os valores previsto com os valores reais.

A métrica usada será o erro médio absoluto de alcance relativo (Mean Absolute Ranged Relative Error, MARRE). Dada uma série de valores reais y_t e uma série de valores previstos \hat{y}_t , ambas com o mesmo comprimento T , o MARRE é um valor percentual dado por:

$$MARRE = 100 \cdot \frac{1}{T} \sum_{t=1}^T \left| \frac{y_t - \hat{y}_t}{\max y_t - \min y_t} \right|$$

O gráfico do desempenho do modelo está exibido na figura abaixo:

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS

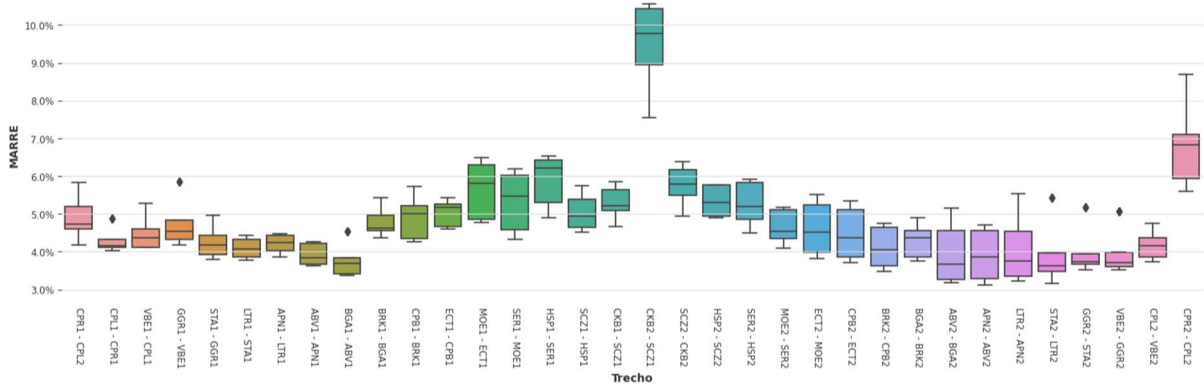


Figura 24 – Erro de validação do modelo para cada trecho.

Nota-se que o modelo teve um MARRE entre 4 e 6% para a maioria dos trechos. Alguns trechos o erro associado aumentou, entretanto, esses trechos são de estações terminais onde eventos de inversão de carro podem acontecer. Esse tipo de evento deve ser vetorizado e adicionado nas covariáveis dessas estações em versões futuras do modelo. A figura abaixo mostra a comparação entre as séries previstas e as séries reais da carga de passageiros para cada carro do trem no trecho STA1-GGR1 na semana entre os dias 7 e 14 de junho de 2023.

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS

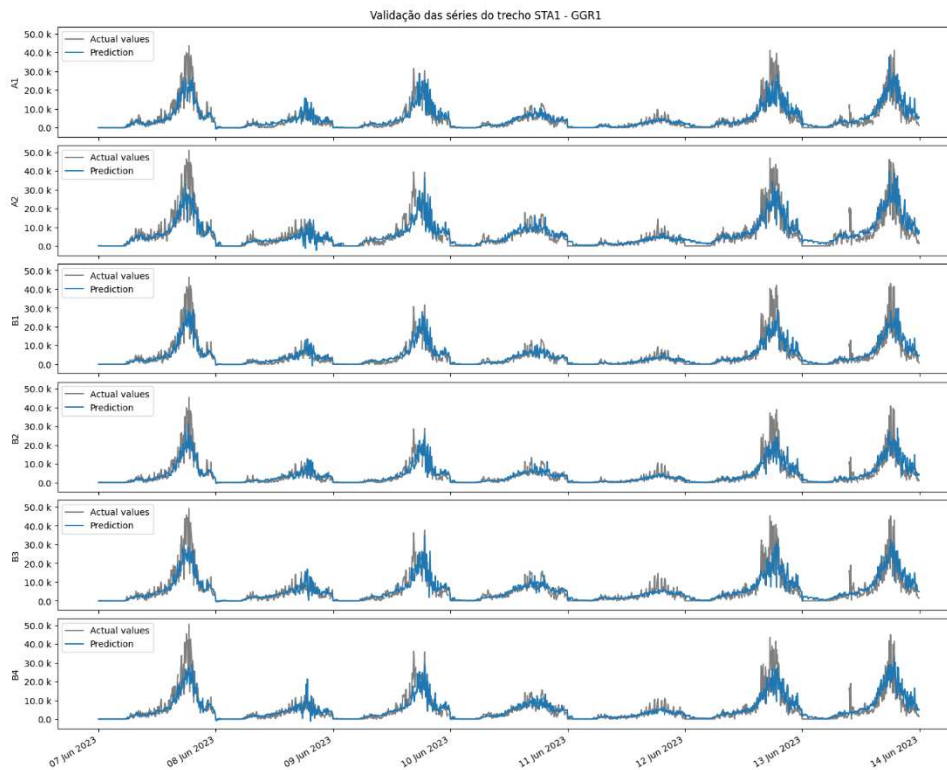


Figura 25 – Validação da previsão da carga por carro para um trecho da linha 5.

Observa-se na Figura 25 que mesmo para um conjunto com menos de 1 ano de dados o modelo conseguiu ajustar bem o feriado de Corpus Christi (quinta feira, 08 junho de 2023), assim como a demanda reduzida no feriado estendido da sexta feira. Mas mesmo com esse resultado satisfatório ainda há muito espaço para melhorias nesse modelo, como um ajuste fino do passo de aprendizado do algoritmo de boost de gradiente, uma melhor detecção de anomalias e a inserção de covariáveis externas ao calendário.

Uma vez que se tem a previsão de todos os carros, para todas as estações em ambos os sentidos da linha 5, pode-se calcular a densidade de passageiros por metro quadrado acumulado na hora sentido da linha e assim estimar a demanda de trens para cada horário. Na figura abaixo observa-se a densidade de passageiros calculada pelo

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS

agrupamento de todas as séries, onde cada ponto terá o valor da carga de passageiros do trecho mais demandado daquele instante.

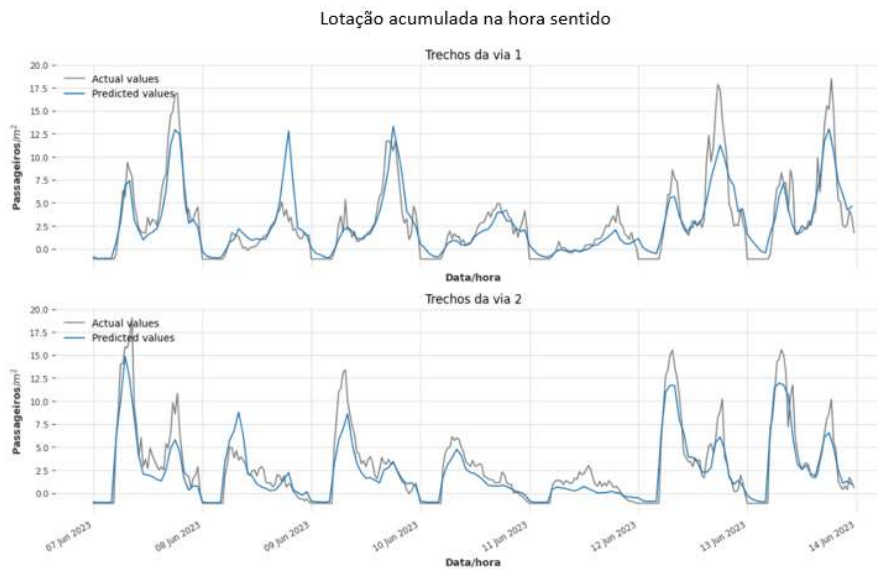


Figura 26 – Pass/m² acumulado por hora em ambos os sentidos da linha 5.

Este resultado é altamente promissor para um modelo de referência, que ainda passará por aprimoramentos em suas rodadas de ajustes finos de parâmetros. O sucesso dessa metodologia na previsão de demanda de passageiros, oferta de trens e outras possibilidades na Linha 5 Lilás destaca seu potencial para ser ampliado e aplicado em outras linhas da malha metroferroviária, como a Linha 4 Amarela, Linhas 8 Diamante e 9 Esmeralda, além das Linhas 1 Vermelha e 2 Azul do Metrô Bahia. Essa expansão, é claro, será conduzida com total respeito às particularidades de cada sistema.

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS



Figura 27 – Análise Operacional em Outras Linhas

CONCLUSÕES

A utilização da Ciência de Dados para otimizar a eficiência do sistema metroferroviário apresentou resultados notáveis nos últimos dois anos e meio de estudo. A coleta e análise de informações provenientes de diversas fontes permitiram a consolidação de bancos de dados, transformando dados não relacionados em insights coerentes e visualmente envolventes. Essa abordagem proporcionou uma tomada de decisão mais informada e precisa, trazendo benefícios significativos para a gestão operacional e o conforto dos passageiros.

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS



A aplicação da Ciência de Dados na análise operacional do sistema metroferroviário permitiu identificar pontos de melhoria, como a distribuição do fluxo nas plataformas, ajuste da oferta de trens com base na demanda real e previsões futuras, além da otimização do conforto dos passageiros durante a viagem. A obtenção de dados em tempo real diretamente dos trens trouxe maior precisão nas análises, em comparação com os métodos de pesquisa tradicionais. Isso possibilitou o desenvolvimento de estratégias mais eficazes para o sistema, considerando informações detalhadas sobre a lotação dos carros e o comportamento do trem em diferentes situações.

A interconexão das informações de diversos sistemas, como portas de plataforma, dispositivos de via e sinalização, proporcionou uma visão ampla e detalhada do funcionamento do sistema como um todo. Com base nessa abordagem, foram projetados cenários futuros, permitindo uma melhor tomada de decisão em relação à frota, demanda, estimativas de custos operacionais variáveis e outras informações importantes para o sistema metroferroviário.

Os resultados obtidos através da Ciência de Dados contribuíram para aprimorar a gestão operacional do sistema, garantindo uma oferta de transporte mais eficiente e ajustada à demanda real dos passageiros. Além disso, essa abordagem possibilitou a simulação de cenários de expansão da linha e estudos de fluxo e multidões, o que auxilia na previsão de impactos futuros e no planejamento estratégico do sistema.

Em conclusão, a aplicação da Ciência de Dados no sistema metroferroviário trouxe uma série de benefícios, que vão desde a melhoria na gestão operacional até a otimização do

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS



conforto e segurança dos passageiros. O uso inteligente e integrado das informações provenientes de diferentes fontes permitiu uma análise mais detalhada e precisa, impulsionando a eficiência e aprimorando a experiência dos usuários do transporte público. Essa abordagem pode servir como modelo para outras linhas e sistemas de transporte, resultando em avanços significativos na eficiência e qualidade dos serviços prestados, ao mesmo tempo em que contribui de forma relevante para uma operação cada vez mais sustentável.

REFERÊNCIAS BIBLIOGRÁFICAS

FRUIN, John J. (1970) – “Designing for Pedestrians: A Level of Service Concept” - polytechnic University of Brooklyn.

FRUIN, John J. (1971) – “Pedestrian Planning and Design” – Metropolitan Association of Urban Designers and Environmental Planners.

FTA (2013) – Federal Transit Administration, “Transit Capacity and Quality of Service Manual”.

CONCESSIONÁRIA DA LINHA 4 DO METRÔ DE SÃO PAULO S.A – VIAQUATRO. **Subsistema Central de Controle de Arrecadação e Passageiros (SCAP)**. 2023a. Acesso restrito.

CONCESSIONÁRIA DA LINHA 5 DO METRÔ DE SÃO PAULO S.A – VIAMOBILIDADE. **Subsistema Central de Controle de Arrecadação e Passageiros (SCAP)**. 2023a. Acesso restrito.

29ª SEMANA DE TECNOLOGIA METROFERROVIÁRIA
10º PRÊMIO TECNOLOGIA E DESENVOLVIMENTO
METROFERROVIÁRIOS



CONCESSIONÁRIA DA LINHA 5 DO METRÔ DE SÃO PAULO S.A – VIAMOBILIDADE. **Gestão de Documentos.** Disponível em:

<https://docnix.grupoccr.com.br/mobilidadesp/corporate/index.html#/maxdoc>. Acesso em: 03 jul. 2023^a. Acesso restrito.

COMPANHIA DO METRÔ BAHIA – CCR METRÔ BAHIA. **Subsistema Central de Controle de Arrecadação e Passageiros (SCAP).** 2023a. Acesso restrito.

CONCESSIONÁRIA DAS LINHAS 8 E 9 DO SISTEMA DE TRENS METROPOLITANOS DE SÃO PAULO – VIAMOBILIDADE. **Subsistema Central de Controle de Arrecadação e Passageiros (SCAP).** 2023a. Acesso restrito.

PORTAL MOBILIDADE – CCR ENGEMOB. **Análise Opr. De Trens.** Disponível em: <http://engemob.grupoccr.com.br/operacional/engenharia.asp#>. Acesso em: 10 jul. 2023a. Acesso restrito.

DOCUMENTAÇÃO POWER BI – MICROSOFT. **Visão Geral.** Disponível em: <https://learn.microsoft.com/pt-br/power-bi/transform-model/desktop-query-overview>. Acesso em: 12 jul. 2023a.

TRAIINTIC, CAF (2009) – COSMOS DIAGNOSIS SYSTEM, “Timon User Manual”.