



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE TRANSPORTES
CURSO DE GRADUAÇÃO EM ENGENHARIA CIVIL

JOÃO VICTOR ABREU DA SILVA

UTILIZAÇÃO DE FERRAMENTAS DE VISUALIZAÇÃO DE DADOS PARA
AUXÍLIO NO PLANEJAMENTO E OPERAÇÃO DE SISTEMAS DE TRANSPORTE
PÚBLICO

FORTALEZA

2022

JOÃO VICTOR ABREU DA SILVA

UTILIZAÇÃO DE FERRAMENTAS DE VISUALIZAÇÃO DE DADOS PARA AUXÍLIO
NO PLANEJAMENTO E OPERAÇÃO DE SISTEMAS DE TRANSPORTE PÚBLICO

Projeto de graduação apresentado ao Curso de Graduação em Engenharia Civil do Centro de Tecnologia da Universidade Federal do Ceará, como requisito à obtenção do grau de Bacharel em Engenharia Civil.

Orientador: Prof. Dr. Francisco Moraes de Oliveira Neto

FORTALEZA

2022

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

S58u Silva, João Victor Abreu da.
Utilização de Ferramentas de Visualização de Dados para Auxílio no Planejamento e Operação de Sistemas de Transporte Público / João Victor Abreu da Silva. – 2022.
112 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Tecnologia, Curso de Engenharia Civil, Fortaleza, 2022.

Orientação: Prof. Dr. Francisco Moraes de Oliveira Neto.

1. Transporte público. 2. Ônibus. 3. AVL. 4. AFC. 5. Business intelligence. I. Título.

CDD 620

JOÃO VICTOR ABREU DA SILVA

UTILIZAÇÃO DE FERRAMENTAS DE VISUALIZAÇÃO DE DADOS PARA AUXÍLIO
NO PLANEJAMENTO E OPERAÇÃO DE SISTEMAS DE TRANSPORTE PÚBLICO

Projeto de graduação apresentado ao Curso de Graduação em Engenharia Civil do Centro de Tecnologia da Universidade Federal do Ceará, como requisito à obtenção do grau de Bacharel em Engenharia Civil.

Orientador: Prof. Dr. Francisco Moraes de Oliveira Neto

Aprovada em: 12 / 12 / 2022 .

BANCA EXAMINADORA

Prof. Francisco Moraes de Oliveira Neto, Ph.D. (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Manoel Mendonça de Castro Neto, Ph.D.
Universidade Federal do Ceará (UFC)

Vandeyberg Nogueira de Souza, M.Sc.
Universidade Federal do Ceará (UFC)

À minha avó, Neuza Sá,

Aos meus pais, Lirlandia e Benjamin.

AGRADECIMENTOS

Aos meus pais, Lirlândia e Benjamin, pelo apoio ininterrupto e incondicional, e pelo indiscutível impacto que tiveram no meu desenvolvimento por meio de seus esforços constantes durante toda a minha vida. Agradeço por sempre terem priorizado a minha educação e o meu desenvolvimento, e por terem dado tudo o que tinham, e até o que não tinham, para que eu pudesse crescer.

À Larissa Teles que, independentemente das circunstâncias, esteve ao meu lado durante a graduação, e que é sem dúvidas um dos motivos de eu poder escrever este texto. Obrigado pelo apoio nos pontos altos e nos pontos baixos, e espero que possamos continuar nos apoiando por muitos anos a vir. Estendo os agradecimentos aos seus pais, José Valdir e Ana Cláudia, que me acolheram, ajudaram e incentivaram durante esse período.

Ao professor Francisco Moraes e sua solicitude em me orientar neste trabalho, cujas contribuições foram fundamentais para a execução e finalização do projeto, sendo um grande guia durante o processo.

Estendo meus agradecimentos aos demais membros da banca examinadora, professor Manoel Mendonça, que instigou em mim o interesse pela área de transportes por meio de suas aulas, e Vandeyberg Nogueira, membro do Grupo de Pesquisa em Operação, Planejamento e Avaliação do Transporte Público.

À professora Verônica Castelo Branco, por ter acompanhado o projeto de graduação durante toda a sua construção. Seu impacto no trabalho, embora indireto, foi grandioso.

À professora Áurea Silva de Holanda, cujo apoio durante a graduação foi fundamental e condicional para a minha trajetória até aqui.

Aos meus amigos Lucas Mariano e Jonatha Farias, que estiveram comigo mesmo antes desta trajetória iniciar, e que perduram mesmo ao seu final. Agradeço fortemente pela presença deles na minha vida e espero que permaneçam por muito mais tempo.

Ao meu grande amigo Igor Lira, cuja participação nessa jornada foi de imensa significância. Seu impacto é multifacetado, incluindo ser para mim inspiração, dupla de trabalho e parceiro de reflexões, conversas e risadas. Nas suas palavras: demos nosso melhor na nossa pior fase. Muito obrigado!

Aos demais membros da Kombi: Ivan Acioli, Jonatas Cacau, Letícia Essabbá, Livia Maria e Luciana Nascimento. Vocês são implacáveis e são os melhores amigos que eu poderia ter tido nessa jornada pela universidade. Agradeço muito a cada um de vocês por terem participado da minha graduação, ela não teria sido a mesma sem vocês.

RESUMO

Sistemas de coleta automática de dados, como o sistema de localização automática de veículos e sistema de coleta de tarifa automatizada, têm a capacidade de gerar grandes volumes de informações referentes à operação de transportes públicos, como ônibus. Estes dados, usualmente utilizados para o acompanhamento de receita e monitoramento em tempo real da frota de ônibus, podem também ser armazenados em bancos de dados para análise futura. Por este prisma, sabendo que este grande volume de dados tem potencial de auxiliar o gerenciamento e o planejamento dos sistemas de transportes das cidades, o presente trabalho utiliza dados obtidos a partir desta tecnologia em Fortaleza para a identificação de indicadores de desempenho que podem ser calculados para a constante melhoria da confiabilidade dos serviços de ônibus da cidade. Ademais, o trabalho utiliza técnicas e ferramentas de visualização de dados e *business intelligence* com o intuito de acompanhar os indicadores determinados ao longo do tempo na cidade, facilitando a análise e o gerenciamento do sistema. No presente trabalho é exposto o método para atingir os objetivos referentes à determinação dos indicadores importantes, a avaliação e cálculo dos indicadores e a proposição da interface de visualização de dados que permite a consulta e análise das diversas linhas de ônibus da cidade ao longo do tempo em diferentes contextos. O trabalho contribui com o conhecimento científico para futuras pesquisas expondo que indicadores de desempenho podem ser calculados por meio destes dados, além de exibir considerações quanto à particularidade dos dados de Fortaleza. Por fim, foi criada uma interface de visualização das informações dos ônibus, tornando o gerenciamento do sistema mais baseado em dados, dirimindo o problema de ineficiência na qualidade e confiabilidade dos sistemas de ônibus do Brasil.

Palavras-chave: transporte público, ônibus, *AVL*, *AFC*, *business intelligence*, visualização de dados

ABSTRACT

Automated data collection systems, such as automatic vehicle location systems and automated fare collection systems, can generate large volumes of information related to the operation of public transports, such as buses. This data, usually used to take account of the revenue collected and the real time monitoring of bus fleets, can also be stored in databases for further analysis. Aware that those large volumes of data have the potential to assist the management and planning of metropolitan public transport systems, this paper utilizes the data obtained using this technology in Fortaleza to identify performance indicators that can be calculated aiming to consistently improve the bus service reliability in the city. Furthermore, the paper uses data visualization and business intelligence tools and techniques to visualize the selected indicators over time in the city, aiming to make the data analysis and management easier. This paper presents the methods used to fulfill the proposed objectives, such as the determination of important indicators, the evaluation and calculation of indicators and the interface proposition that'll allow querying and evaluating of transit routes in the city over time and at different contexts. This work contributes to future research by showing which performance indicators can be calculated through this data, while also presenting the particularities of the data that need to be addressed. Ultimately, a visualization interface was built for the bus information, further establishing a platform for a data driven management of the public transport system, thus tackling the inefficiency and lack of reliability of the bus service in Brazil.

Keywords: public transport, bus, AVL, AFC, business intelligence, data visualization

LISTA DE QUADROS

Quadro 1 - Perspectivas de Qualidade de Sistemas de Transporte Público	19
Quadro 2 - Lista de Indicadores de Qualidade	21
Quadro 3 - Lista de Indicadores da Operação	22
Quadro 4 - Lista de Indicadores de Eficiência	23
Quadro 5 - Horários de Chegada do Ônibus 30535 no Ponto de Controle 4169	48
Quadro 6 - Horários de Chegada na Parada Inicial e na Parada Final do Ônibus 30535 na Linha 75, Sentido Ida, no dia 22/10/2018	51
Quadro 7 - Chegadas ao Ponto de Controle 4169 entre 12:00 e 13:00, 23/10/2018	53
Quadro 8 - Total de Registros por Tabela do Banco de Dados	76
Quadro 9 - Distribuição dos Indicadores por Grupo de Cálculo e Classe	78

LISTA DE FIGURAS

Figura 1 - Criação de Tabela de Posição do Ônibus no <i>MySQL</i>	33
Figura 2 - Comando de Inserção de Dados de Arquivo <i>csv</i> em Tabela no <i>MySQL</i>	35
Figura 3 - Aproximações do Ônibus 30535 com o Ponto de Controle 4169	38
Figura 4 - Agrupamento das Aproximações do Ponto de Controle 4169 pelo Ônibus 30535 .	39
Figura 5 - Aproximação do Grupo 1 do Ônibus 30535 com o Ponto de Controle 4169	40
Figura 6 - Diagrama de Distância à Parada por Horário do Grupo 1 do Ônibus 30535 ao Ponto de Controle 4169	41
Figura 7 - Consulta de Criação de Tabela de Acompanhamento e Inserção de Dados	44
Figura 8 - Consulta de Determinação do Tempo de Ciclo do Ônibus 30535 no Ponto de Controle 4169	48
Figura 9 - Consulta de Determinação da Parada Inicial e Final da Linha 75 no Sentido Ida ...	50
Figura 10 - Consulta para Determinação do Horário de Chegada na Parada Inicial e na Parada Final do Ônibus 30535	50
Figura 11 - Consulta para Determinação do <i>Headway</i> para o Ponto de Controle 4169	52
Figura 12 - Consulta para Determinação de Frota de Ônibus na Linha 75, no dia 23/10/2018, entre 10:00 e 11:00	54
Figura 13 - Medida em DAX Referente ao Cálculo do Tempo em Viagem Médio.....	66
Figura 14 - Interface de Criação de Visuais do Power BI	67
Figura 15 - Gráfico de Barras de Headway Médio por Faixa Horária, sem tratamento visual	68
Figura 16 - Opção Publicar do Power BI Desktop	69
Figura 17 - Configuração para Publicar Relatório do Power BI na Web por meio de Link Público	69
Figura 18 - Amostra dos Dados Brutos de GPS	70
Figura 19 - Amostra dos Dados Brutos de Bilhetagem	71
Figura 20 - Amostra do Arquivo de Linhas do GTFS	73
Figura 21 - Esquematização das Tabelas Normalizadas e Seus Relacionamentos no Banco de Dados	75
Figura 22 - Amostra dos 5 Primeiros Registros da Tabela de Acompanhamento	78
Figura 23 - Aba de Seleção de Filtros da Interface de Consulta	81
Figura 24 - Cabeçalho, Rodapé e Aba de Filtros da Interface	82
Figura 25 - Página de Indicadores da Operação da Interface de Consulta	83
Figura 26 - Página de Indicadores de Eficiência da Interface de Consulta	84

Figura 27 - Página de Indicadores de Eficiência da Interface de Consulta, Gráfico Alternativo	85
Figura 28 - Página de Avaliação de Headway da Interface de Consulta	86
Figura 29 - Página de Avaliação de Headway da Interface de Consulta, faixa horária de 13h-16h	87
Figura 30 - Página de Tempos e Velocidades da Interface de Consulta	88
Figura 31 - Código QR para Acesso à Interface de Consulta	89

LISTA DE ABREVIATURAS E SIGLAS

AFC	Automated Fare Collection
APC	Automatic Passenger Counting
AVL	Automatic Vehicle Location
GTFS	General Transit Feed Specification
KPI	Key Performance Indicators
SIT-FOR	Sistema Integrado de Transporte de Fortaleza

SUMÁRIO

1	INTRODUÇÃO	13
2	PROBLEMA DE PESQUISA E QUESTÕES MOTIVADORAS	15
2.1	Justificativa	16
2.2	Questões Motivadoras	16
3	OBJETIVOS	17
3.1	Objetivo Geral	17
3.2	Objetivos Específicos	17
4	REVISÃO BIBLIOGRÁFICA	18
4.1	Indicadores de Desempenho	18
4.1.1	<i>Indicadores de Qualidade.....</i>	<i>19</i>
4.1.2	<i>Indicadores de Operação</i>	<i>22</i>
4.1.3	<i>Indicadores de Eficiência</i>	<i>22</i>
4.2	Sistemas Automatizados de Coleta de Dados.....	24
4.2.1	<i>Sistemas de AVL</i>	<i>24</i>
4.2.2	<i>Sistemas de AFC</i>	<i>25</i>
4.3	Visualização de Dados	26
4.4	Trabalhos Semelhantes e Resultados	28
4.5	Utilização de dados de <i>AVL</i> , <i>AFC</i> e <i>GTFS</i> em Fortaleza	30
5	MATERIAIS E MÉTODOS	32
5.1	Criação da Base de Dados.....	32
5.1.1	<i>Teoria da Normalização e Idealização do Formato da Base de Dados</i>	<i>32</i>
5.1.2	<i>Criação das Rotinas de Tratamento de Dados para o Formato Idealizado</i>	<i>34</i>
5.1.3	<i>População das Tabelas</i>	<i>35</i>
5.2	Modelagem dos Horários do Ônibus	36
5.2.1	<i>Aproximação Por Parada</i>	<i>36</i>
5.2.2	<i>Diagramas de Aproximação das Paradas</i>	<i>40</i>
5.2.3	<i>Correção do Sentido na Modelagem do Horário</i>	<i>42</i>
5.2.4	<i>Criação da Tabela de Acompanhamento</i>	<i>43</i>
5.3	Determinação dos Indicadores	44
5.3.1	<i>Compatibilização dos Indicadores</i>	<i>45</i>
5.3.2	<i>Considerações de Cálculo para Indicadores Indiretos</i>	<i>46</i>
5.3.3	<i>Cálculo dos Indicadores</i>	<i>47</i>

5.4	Interface de Consulta e Visualização de Dados	59
5.4.1	<i>Determinação das Visualizações</i>	59
5.4.2	<i>Criação da Interface</i>	64
5.4.3	<i>Implementação da Ferramenta Online</i>	68
6	RESULTADOS E DISCUSSÕES	70
6.1	Criação da Base de Dados.....	70
6.1.1	<i>Teoria da Normalização e Idealização do Formato da Base de Dados</i>	70
6.1.2	<i>Criação das Rotinas de Tratamento de Dados e População das Tabelas.....</i>	75
6.2	Modelagem dos Horários dos Ônibus	77
6.3	Determinação dos Indicadores	78
6.4	Interface de Consulta e Visualização de Dados	80
6.4.1	<i>Página de Indicadores da Operação</i>	82
6.4.2	<i>Página de Indicadores de Eficiência</i>	84
6.4.3	<i>Página de Avaliação do Headway</i>	86
6.4.4	<i>Página de Tempos e Velocidades</i>	87
7	CONSIDERAÇÕES FINAIS.....	90
7.1	Conclusão	90
7.2	Sugestões para Trabalhos Futuros	91
7.2.1	<i>Referente à Base de Dados</i>	91
7.2.2	<i>Referente à Modelagem do Horário nas Paradas</i>	92
7.2.3	<i>Referente aos Indicadores</i>	93
7.2.4	<i>Referente à Interface</i>	93
	REFERÊNCIAS	94
	APÊNDICE A – CONSULTAS DE CÁLCULOS DOS INDICADORES	97
	APÊNDICE B – CONSULTAS DE CRIAÇÃO DE TABELAS	101
	APÊNDICE C – ROTINA DE TRATAMENTO DOS DADOS	104

1 INTRODUÇÃO

Nas últimas décadas, há uma crescente representação dos desafios relacionados à sustentabilidade ecológica na sociedade. Frente a isso, há uma maior atenção ao sistema de transporte das cidades referentes a emissões climáticas, poluição do ar e congestionamento, o que gera maior enfoque ao uso de transportes públicos (PETTERSSON; SØRENSEN, 2020). Com isto, Pettersen e Sørensen (2020) afirmam que, dentre outras preocupações, a qualidade do transporte público deve ser assegurada, tornando-o mais atrativo para os usuários

A qualidade do transporte público urbano deve considerar a satisfação de todos os atores envolvidos no sistema, sendo estes os usuários, o governo, a comunidade, trabalhadores do setor e empresários do ramo (FERRAZ; TORRES, 2001). Neste cenário, frente ao crescimento da priorização do uso do transporte público nas metrópoles, faz-se necessária uma constatação recorrente da qualidade deste serviço para seu correto planejamento e gerenciamento.

Segundo Ferraz e Torres (2001), o gerenciamento do transporte público envolve fiscalização, definição de programação operacional, gestão de banco de dados, dentre outras atividades referentes à inteligência de dados. Estas competências se aliam diretamente à modernização das tecnologias de informação dos últimos anos. Tecnologias como coleta de dados de *Automatic Vehicle Location (AVL)*, que arquivam informações de localização dos veículos de transporte, ou sistemas *Automated Fare Collection (AFC)*, que arquivam a contagem de passageiros nos meios de transporte público que utilizam bilhetagem eletrônica, são ferramentas relativamente recentes que têm o potencial de auxiliar o gerenciamento e o planejamento a partir de uma correta análise e uso das informações coletadas.

Sistemas de *AVL* consistem na determinação da geolocalização dos veículos de transporte público por meio de *GPS*, e estas informações são armazenadas em uma base de dados para posterior uso. Este tipo de sistema está sendo cada vez mais usado em cidades ao redor do mundo, com o intuito de promover melhorias na eficiência, eficácia e qualidade do trânsito rodoviário, além de apoiar os operadores de trânsito tanto no gerenciamento e operação de frotas quanto na mensuração do desempenho dos sistemas e no planejamento da operação e dos serviços (BARABINO; DI FRANCESCO; MOZZONI, 2017). De forma análoga, sistemas de *AFC* determinam, nos veículos de transporte público, cada vez que ocorre uma instância de bilhetagem eletrônica automática; estas informações de bilhetagem são, então, armazenadas para o uso apropriado.

Outras tecnologias e setores de estudo de dados, como *Business Intelligence (BI)* e *Analytics* têm se tornado uma prioridade ao redor do mundo, por agregar valor às empresas, sobretudo gerencialmente, informando a tomada de decisões, conforme descrito por Moreno, Cavazzote e De Souza Carvalho (2020). *BI* é definido por Wixom e Watson (2010) como tecnologias, aplicações e processos de obtenção, armazenamento e análise de dados para ajudar seus usuários a tomarem melhores decisões (apud NAMVAR; CYBULSKI; PERERA, 2016). *BI* envolve também, principalmente para o usuário final, a utilização de visualização de dados, pois esta ferramenta torna a análise de informações mais fácil, e, portanto, tem grande impacto em decisões gerenciais (VALLURUPALLI; BOSE, 2018).

O uso destas tecnologias para melhoria do sistema de transportes é um fenômeno crescente, providenciando uma fonte rica de dados importantes para diversos usos em aplicações após seu devido processamento (HOUNSELL; SHRESTHA; WONG, 2012). Algumas aplicações descritas por Hounsell, Shrestha e Wong (2012) são a melhoria do monitoramento dos ônibus, enriquecimento das análises espaço-temporal e auxílio na operação e priorização do sistema. O *BI* e a criação de relatórios visuais e aplicações traz uma facilidade ao sistema, de acordo com Hanft *et al.* (2016), por dirimir a necessidade de um intermediário analítico no trabalho de planejamento dos sistemas de transporte. Hanft *et al.* (2016) conclui que o conhecimento estratégico adquirido a partir dos dados armazenados do sistema de transporte, ao serem passíveis de consulta a partir das aplicações e relatórios *web* criadas, aumentam a aplicabilidade, usabilidade e flexibilidade das análises, uma vez que as informações calculadas e apresentadas podem ser avaliadas mesmo sem conhecimento íntimo dos dados e das tecnologias utilizadas.

Visto que Fortaleza possui tanto sistemas de *AVL* como de *AFC* implementados em sua frota de ônibus, nota-se que as condições da infraestrutura de transportes da cidade são propícias ao uso destas ferramentas para a determinação de indicadores de qualidade, desempenho e operação dos ônibus e, devido a isso, será objeto de estudo deste trabalho.

O presente estudo busca, portanto, levantar mecanismos de averiguação da qualidade do transporte público por meio de indicadores de desempenho descritos na literatura e que se adequem aos dados disponíveis. As ferramentas de visualização de dados poderão, então, ser utilizadas para ilustrar esses dados de forma a facilitar a tomada de decisões referentes ao planejamento e gerenciamento do transporte público.

2 PROBLEMA DE PESQUISA E QUESTÕES MOTIVADORAS

O problema de pesquisa para qual este trabalho busca contribuir é, em maior escopo, a deficiência na qualidade e na confiabilidade do sistema de transporte público no Brasil.

Afunilando isto, o trabalho se motiva pela onerosidade do planejamento e gerenciamento do modo de transporte por ônibus, adentrando ainda na dificuldade do uso de ferramentas modernas de coleta e análise de dados para este fim.

Esta dificuldade é inerente aos sistemas de coleta automatizada. Mesmo em frotas totalmente equipadas com os equipamentos adequados de coleta, pode-se obter resultados que não refletem as experiências dos atores do sistema, sejam passageiros ou operadores. Isso se deve à diversos motivos, dentre eles falhas e anomalias encontradas nos dados arquivados, ausência de validação das informações, incorreta utilização ou nenhum ou somente parcial ajuste de anomalias (BARABINO; DI FRANCESCO; MOZZONI, 2017).

Além disso, avaliar o sistema de transportes nem sempre é possível diretamente a partir dos dados brutos de coleta automática. Muitas das informações relevantes para a compreensão do desempenho do sistema advêm de indicadores inferidos a partir dos dados, exigindo a implementação de ferramentas de consulta eficiente dos dados e de adequada visualização.

Alguns sistemas de coleta automática contam com detecção de pontos de controle quando os veículos atingem determinadas posições, como por exemplo as paradas de ônibus. A partir disso, a aferição de muitos indicadores importantes é possibilitada, sendo estes denominados de indicadores à nível de parada. Visto que os indicadores à nível de parada permitem a avaliação dos transportes numa maior granularidade, Ma (2014) conclui que esta configuração é ideal para a avaliação da qualidade, eficiência e confiabilidade do sistema.

É importante mencionar, ainda, que grande parte dos sistemas de transportes públicos no Brasil são sujeitos à iniciativa pública por meio de empresas públicas de gestão, fiscalização e gerenciamento. É notado por Henkel, Perjons e Sneiders (2017) que apesar das tecnologias modernas computacionais serem comuns no mercado privado, ainda há menor prevalência e falta de otimização delas no meio público, o que caracteriza um problema válido de pesquisa: pouca adoção de tecnologias modernas de análise nos sistemas de transporte no país.

2.1 Justificativa

A área de pesquisa relacionada ao uso de ferramentas modernas de coleta de dados automatizado e de *BI* para fins de acompanhamento de qualidade e desempenho é ainda pouco explorada, sobretudo quando tratando sobre seu uso na esfera pública. Atrelado a isso o fato de que há crescente demanda para a migração do transporte individual para o coletivo nas cidades, busca-se preencher a lacuna de aplicação de estudos que unam as duas temáticas num único trabalho.

Conforme dito por Hanft *et al.* (2016), a substituição dos dados manuais por dados automatizados melhora a confiança das informações sendo comunicadas, além de liberar o tempo dedicado à coleta em prol do uso e análise dos dados.

Com isso, o trabalho é justificado pela possibilidade de melhoria no sistema de transporte público por meio de ônibus advinda do uso destas ferramentas em prol da sociedade e da sustentabilidade.

2.2 Questões Motivadoras

As questões motivadoras deste trabalho são:

- a) Como é possível consultar os dados de *AVL* e *AFC* de forma eficiente?
- b) Como pode se determinar o instante em que um ônibus passa pela parada de ônibus?
- c) Que indicadores de desempenho podem ser calculados a partir dos dados de *AVL* e *AFC*?
- d) Como os indicadores podem ser visualizados de forma a auxiliar na compreensão dos sistemas de transporte?

3 OBJETIVOS

Nesta seção serão expostos os objetivos que se almeja atingir no decorrer deste projeto, aliados às questões de pesquisa previamente citadas.

3.1 Objetivo Geral

Este trabalho tem como objetivo geral utilizar ferramentas de visualização de dados em bases de dados coletadas automaticamente para o cálculo de indicadores de qualidade, eficiência e operação de ônibus em Fortaleza para auxílio no gerenciamento e planejamento do sistema de transportes da cidade.

3.2 Objetivos Específicos

Os objetivos específicos deste trabalho são:

- a) Criar banco de dados com informações consolidadas de geolocalização e bilhetagem dos ônibus do sistema de transporte;
- b) Modelar a chegada dos ônibus às paradas a partir das informações do banco de dados;
- c) Determinar que indicadores de desempenho podem ser calculados a partir do banco de dados;
- d) Propor uma interface que permita consulta, visualização e análise dos dados de sistema de transportes.

4 REVISÃO BIBLIOGRÁFICA

Nesta seção do trabalho será apresentada a revisão bibliográfica dos temas estudados neste projeto. Será dividida por temática abordada, servindo inclusive como fundamentação teórica para o método, tal como justificativa.

Haverá uma seção referente ao estudo de indicadores de desempenho, sobretudo dentro do contexto de sistemas de transportes de ônibus; uma seção sobre sistemas automáticos de coleta de dados, adentrando nas temáticas de *AVL* e *AFC*; uma seção sobre visualização de dados e *BI*; uma seção resumindo alguns trabalhos encontrados na literatura com enfoque semelhante, denotando suas conclusões referentes à indicadores e o uso das tecnologias abordadas no presente projeto e, por fim, uma seção referente à trabalhos que utilizam os dados de *AVL* e *AFC* de Fortaleza, de forma a servir de insumo para o desenvolvimento do projeto.

4.1 Indicadores de Desempenho

Indicadores de desempenho são ferramentas úteis para coletar informações e analisar a evolução de processos ou sistemas complexos (FRANCESCHINI; GALETTO; MAISANO, 2007). Ainda de acordo com Franceschini, Galetto e Maisano (2007), indicadores de desempenho devem ser capazes de representar aspectos de processos e, portanto, são modelos de representação para apoiar avaliações e decisões para estes processos.

Indicadores são sempre utilizados para monitorar o resultado de ações para melhorias e são vistas como ferramentas sistemáticas para nutrir melhorias contínuas em processos (FORTUIN, 1988).

Indicadores de desempenho são meios para um fim: eles providenciam à gerência uma ferramenta para comparar resultados atuais com alvos predeterminados, e para medir a extensão de qualquer desvio (FORTUIN, 1988, p. 1, tradução nossa).

De acordo com Fortuin (1988), indicadores de desempenho devem ser derivados dos objetivos a serem atingidos em questão, podendo ser aplicados em qualquer área, porém a implementação só é lógica se a organização responsável tem como intuito a melhoria ou monitoramento contínuo. Algumas recomendações importantes são, segundo Fortuin (1988):

- a) indicadores de desempenho devem ser bem definidos, simples, compreensíveis e prontamente disponíveis aos seus usuários;
- b) a apresentação dos indicadores de desempenho deve ser acompanhada por uma indicação do objetivo a ser atingido;

- c) indicadores de desempenho devem ser relevantes; referindo-se à parâmetros que podem ser controlados pelos usuários destes;
- d) ao implementar sistemas de indicadores, os responsáveis devem se concentrar em um número limitado, consistindo apenas dos mais importantes que consigam cobrir todos os aspectos relevantes de um serviço ou produto.

4.1.1 Indicadores de Qualidade

Se tratando de sistemas de transportes públicos, aponta-se um subconjunto de indicadores de desempenho referente à percepção dos usuários acerca da qualidade do serviço, que são os indicadores de qualidade. Indicadores de qualidade são importantes tanto para as empresas que disponibilizam os serviços para a apresentação e monitoramentos, como para as autoridades públicas de fiscalização (CASCETTA; CARTENÌ, 2014).

Segundo Cascetta e Cartenì (2014), para a avaliação da qualidade de serviços de transportes, são requeridos métodos para a definição de indicadores de padrões de qualidade e técnicas correlatas de mensuração de forma a traduzir as expectativas dos usuários e as percepções de qualidade dos diversos atores em parâmetros mensuráveis e gerenciáveis. O Quadro 1 define os conceitos de qualidade tanto do ponto de vista do fornecedor do serviço de transportes quanto do ponto de vista do usuário.

Quadro 1 - Perspectivas de Qualidade de Sistemas de Transporte Público

Tipo de Qualidade	Descrição
Qualidade de Serviço Alvo	O nível de serviço que a fornecedora do serviço busca atingir
Qualidade de Serviço Entregue	O nível de serviço que é de fato atingido (função do desempenho do sistema de transporte)
Qualidade de Serviço Percebida	A impressão do usuário do serviço acerca da qualidade de serviço entregue
Qualidade de Serviço Buscada	O nível de serviço requerido pelo usuário implicitamente ou explicitamente

Fonte: Adaptado de Cascetta e Cartenì (2014)

O fornecedor do serviço e/ou a autoridade pública estatal, por meio do planejamento de serviços, determina as características do transporte, como linhas de ônibus, quadro de horários e tipos de veículo por meio do equilíbrio entre os custos de produção e os alvos de qualidade. A condição de qualidade real do serviço de transporte público, englobando

congestão, regularidade do serviço e confiabilidade de horário, por exemplo, são mensuradas e descritas pela qualidade de serviço entregue (CASCETTA; CARTENÌ, 2014).

É devido a isto que o monitoramento do sistema de transportes é crucial; por meio deste é possível aferir o nível de satisfação dos clientes, suas escolhas de mobilidade e o real desempenho do sistema. Conforme Cascetta e Cartenì (2014), a estimativa e mensuração da qualidade entregue por meio de medidas objetivas, que são os indicadores de desempenho, permitem que a qualidade do serviço seja constantemente avaliada, e as informações obtidas devem influenciar o planejamento dos serviços num processo iterativo de evolução do sistema de transportes.

Uma consideração importante acerca da qualidade, sobretudo referente ao ponto de vista do usuário, que são a qualidade de serviço percebida e qualidade de serviço buscada no Quadro 1, é que estas não podem ser mensuradas de forma direta a partir de dados de operação do serviço, dada a sua subjetividade. Elas podem ser estimadas a partir de outras fontes de dados, como pesquisas de percepção dos usuários atendidos pelo serviço, onde são abordados indiretamente características dos desempenhos com os usuários, a fim de estimar a perspectiva geral do serviço.

Ferraz e Torres (2001) trazem sugestões de indicadores, considerações e padrões de qualidade para cada um dos atores envolvidos nos sistemas de transporte público, sendo eles os usuários do transporte, os empresários, os trabalhadores, a comunidade e o governo. Por meio de matrizes que possibilitam uma avaliação global da qualidade, Ferraz e Torres (2001) indicam que a qualidade é multifacetada, e envolve parâmetros para a avaliação que englobam desde comparações do sistema avaliado com outras cidades, indicadores de desempenho clássicos como a taxa de passageiros em pé nos veículos, até estudos e pesquisas específicas relacionadas à justiça social-financeira e satisfação dos atores, corroborando para a conclusão de que a qualidade, apesar de ter aspectos objetivos, é majoritariamente subjetiva.

Os principais indicadores de qualidade propostos por Ferraz e Torres (2001) são descritos no Quadro 1.

Quadro 2 - Lista de Indicadores de Qualidade

Indicador de Qualidade	Descrição do Indicador
Frequência de Atendimento	Quantidade de ônibus por intervalo de tempo. Caracterizado pela razão entre a quantidade de ônibus atendendo uma demanda e a faixa temporal avaliada.
Tempo de Viagem	Tempo médio que um passageiro tem que passar dentro do ônibus desde sua subida no veículo até o desembarque. Pode-se utilizar o tempo de <i>buffer</i> para avaliar também a variação.
Lotação	Refere-se ao percentual da capacidade do ônibus que está sendo utilizada. É, portanto, a razão entre o número de passageiros num determinado instante e a capacidade do veículo.
Confiabilidade	Tem relação com a variação da frequência de atendimento dos ônibus. Medido usualmente pelo coeficiente de variância da frequência.
Segurança	Compreende a quantidade de acidentes envolvendo os veículos da frota. Geralmente descrito pelo total de acidentes a cada 100 mil quilômetros percorridos.

Fonte: Adaptado de Ferraz e Torres (2001)

A confiabilidade, por si só, é alvo de estudo e discussão na literatura, e geralmente não é compreendida ou avaliada por um único indicador. Confiabilidade de trânsito é definido como a confiança em termos de tempo, volume de passageiros, segurança, informações e qualidade veicular num sistema de transportes (CEDER, 2007). Em contraponto, a falta de confiabilidade ou problemas de confiabilidade de serviços de transportes ocorrem quando algum destes aspectos é quebrado, seja por horários incertos dos ônibus, volumes acima dos esperados, informações errôneas ou *bus bunching*, por exemplo.

Ceder (2007) destaca que a confiabilidade é percebida pela variação dos atributos do serviço de transportes prestado. Quanto mais invariável são os atributos, maior confiabilidade tem os serviços; se o atraso dos ônibus, por exemplo, é sempre em torno de uma mesma pequena margem de tempo, considera-se que o serviço é mais confiável do que em um caso em que os serviços são prestados pontualmente na maioria das vezes, porém contando com grandes atrasos eventuais.

Por este prisma, as medidas de confiabilidade tendem a ser baseadas nas características da distribuição dos atributos, como média de tempo de espera e seu coeficiente de variação. Além disso, grande parte da percepção da confiabilidade do serviço advém da probabilidade de vários atributos de confiabilidade serem cumpridos de acordo com a expectativa dos usuários; por exemplo, o tempo de espera ser dentro da expectativa, em

sequência o volume de passageiros no transporte estar dentro do esperado, o tempo de viagem estar próximo do usual, e assim por diante.

4.1.2 Indicadores de Operação

Os indicadores de operação são aqueles utilizados para o planejamento e avaliação da operação de uma linha de ônibus. Estes indicadores servem como base para o planejamento, e utilizam primariamente informações de demanda e velocidades dos ônibus de uma determinada linha para a sua correta viabilização.

O Quadro 3 apresenta e descreve os principais indicadores da operação.

Quadro 3 - Lista de Indicadores da Operação

Indicador da Operação	Descrição do Indicador
Demanda Crítica	Se refere à maior demanda estimada no horário de pico de uma linha de ônibus. Calculada baseando-se no total de passageiros dentro do ônibus para cada trecho da linha.
Tempo de Ciclo	É o tempo desde o início de uma viagem do ônibus, até que este retorne ao mesmo ponto e inicie novamente a viagem. Em suma, considera todo o tempo de trajeto, mas também as paradas em terminais.
Tempo em Viagem	É o tempo desde o início de uma viagem do ônibus, até o seu final. Desconta-se do tempo de ciclo o tempo parado nos terminais.
Velocidade Comercial	É a velocidade do veículo considerando o tempo de ciclo. Ou seja, a velocidade média do ônibus na linha durante um ciclo.
Velocidade Operacional	Se refere à velocidade do ônibus enquanto este está em movimento, ou seja, utilizando o tempo em viagem.
Headway	Headway é definido como o inverso da frequência. Ou seja, é o tempo entre dois ônibus consecutivos passando pelo mesmo ponto. Ou seja, a defasagem temporal entre os veículos.
Frota	O total de ônibus atendendo uma determinada linha.

Fonte: Adaptado de Ferraz e Torres (2001)

4.1.3 Indicadores de Eficiência

A eficiência pode ser descrita como a característica de realizar atividades com maior nível de produtividade, minimizando o uso de recursos como tempo e dinheiro, mas ainda atingindo os objetivos; é geralmente expressa pela relação do volume de um serviço ou produto entregue e os insumos gastos. A eficiência de um sistema de transportes segue o mesmo

princípio; o produto geralmente são as viagens ofertadas, enquanto os insumos são os veículos, funcionários, combustível e afins, conforme Ferraz e Torres (2001).

A eficiência no ponto de vista do sistema de transportes afeta, prioritariamente, a agência que presta os serviços, visto que dela parte os insumos. Ferraz e Torres (2001) trazem três pontos que devem ser seguidos para a maximização da eficiência, sendo estes:

- a) reduzir ao mínimo a distância de transporte, assim trazendo uma diminuição do custo de combustível, pneus, salários, manutenção de veículos etc.;
- b) utilizar a maior velocidade possível, reduzindo o tempo de viagem e os gastos;
- c) usar veículos com a maior capacidade possível que atenda a demanda da linha, necessitando menor número de veículos.

Ferraz e Torres (2001) lista diversos indicadores normalmente utilizados para a avaliação da eficiência econômica de um sistema de transportes, e estes servirão de insumo para o trabalho. Alguns exemplos dentre os mais usuais são: índice de quilômetros por veículo; índice de aproveitamento da frota; índice de passageiros por quilômetro (IPK); custo por quilômetro e custo por passageiro.

O Quadro 4 lista os principais indicadores de eficiência.

Quadro 4 - Lista de Indicadores de Eficiência

Indicador de Eficiência	Descrição do Indicador
Distância Média entre Paradas	É meramente a distância média entre cada parada de uma linha. O cálculo é avaliado utilizando a distância sobre o trajeto projetado.
Índice de Quilômetros por Veículo	Medido em km / veículo / dia. Corresponde à relação entre o número total de quilômetros percorridos por todos os veículos e o total de veículos. Avalia a eficiência da utilização da frota.
Índice de Aproveitamento da Frota	Definido como a razão entre a frota de veículos máxima nos horários de pico utilizados e a frota total disponível da empresa.
Índice de Passageiros por Quilômetro	Corresponde à divisão entre o total de passageiros atendidos e a quantidade de quilômetros circulados por todos os ônibus. Reflete o grau de utilização do serviço.
Índice de Passageiros por Veículo	Obtém-se pela relação entre a quantidade de passageiros transportados por dia e o número de veículos atuando.
Custo por Passageiro	Mensurado em R\$ / km. Reflete o custo para o transporte de cada passageiro.
Razão Arrecadado por Custo	É definido como a razão entre toda a receita gerada pelas bilhetagens e o custo total da operação. Reflete a eficiência econômica absoluta do sistema.

Fonte: Adaptado de Ferraz e Torres (2001)

4.2 Sistemas Automatizados de Coleta de Dados

Sistemas automatizados de coleta de dados são ferramentas amplamente utilizadas na operação de sistemas de transportes e estão se tornando cada vez mais comuns em sistemas de trânsito ao redor do mundo (ZHAO; RAHBEE; WILSON, 2007). Estes sistemas consistem em mecanismos e ferramentas que coletam e armazenam informações de forma contínua e automática, sem necessidade da presença de pessoas, o que proporciona a possibilidade da coleta de volumes de dados muito maiores com custos posteriores à implementação menores.

A grande disponibilidade de sistema de coleta de dados automatizados no transporte público vem se tornando o facilitador de análises não possíveis no passado. Estes sistemas permitem, ainda, o monitoramento e avaliação de novas políticas e decisões no transporte, visto que os dados são coletados continuamente; assim, é possibilitada a comparação de comportamentos e indicadores de forma retroativa (HALVORSEN *et al.*, 2020).

Os sistemas de coleta de dados automatizados podem ser categorizados em três tipos: *APC* (*Automatic Passenger Counting*), *AVL* e *AFC* (ZHAO; RAHBEE; WILSON, 2007). A cidade de Fortaleza possui implementação de sistemas de *AVL* e *AFC*.

4.2.1 Sistemas de AVL

Segundo Zhao, Rahbee e Wilson (2007), sistemas de *AVL* são sistemas que gravam as informações de localização dos veículos, que podem ser utilizadas para manter os condutores informados de desvios de cronograma, informar as centrais de controle e afins. Muitas empresas estão utilizando estes sistemas para prover informações de localização em tempo real, que pode ser usada para controlar rotas e alimentar informações de passageiros (HORBURY, 1999).

Sistemas de *AVL* também geram um histórico de posição dos ônibus ao longo do dia em intervalos regulares, que os operadores podem utilizar posteriormente. Segundo Horbury (1999), estes dados podem ser utilizados para diversos propósitos, como:

- a) quadros de horários: os dados coletados do fluxo real do dia a dia de linhas de transporte podem ser usados para identificar área ao longo das rotas onde ônibus tendem a usualmente falhar a cumprir os horários previstos; melhores quadros podem ser desenvolvidos e testados para melhoria das linhas;
- b) propósitos estatísticos: os dados podem ser utilizados para análise estatísticas, como em confiabilidade e pontualidade dos veículos;

- c) redução de custos de operação: pode haver reduções em custos operacionais devido ao replanejamento, que pode reduzir tamanho da frota e eliminar quilometragem desnecessária.

Estabelece-se, ainda, que os dados armazenados de localização podem ser utilizados da velocidade dos ônibus com precisão comparável aos métodos manuais (HORBURY, 1999); os dados também possibilitam o cálculo de medidas de pontualidade (BARABINO; DI FRANCESCO; MOZZONI, 2015) e a determinação da qualidade dos serviços prestados (ZHAO; RAHBEE; WILSON, 2007).

Apesar dos diversos usos, os sistemas de *AVL* não disponibilizam dados totalmente precisos, nem refletem perfeitamente a experiência dos passageiros; os dados armazenados contêm anomalias que afetam a acurácia das medições de confiabilidade. Por isso, é fortemente recomendado que as anomalias nos dados sejam endereçadas, conforme descrito por Barabino, Di Francesco e Mozzoni (2017), pois isso contribui para a precisão das análises geradas.

Parte do potencial uso destes dados, porém, envolve o uso de outras tecnologias de coleta de dados em conjunto. Por exemplo, o uso de sistema de *APC* em conjunto permite o cálculo do volume de passageiros e a estimativa da taxa de chegada de passageiros nas paradas de ônibus (HORBURY, 1999), o que permite o cálculo da demanda e outros parâmetros importantes no planejamento de sistemas de transportes.

4.2.2 *Sistemas de AFC*

Sistemas de coleta automática de tarifa incluem uma vasta gama de tecnologias que gravam a localização e o horário de cada transação de tarifa de cada usuário (ZHAO; RAHBEE; WILSON, 2007). Sistemas de *AFC* utilizando *smart cards* (cartões de leitura automática) se tornaram populares durante os anos 90, e sua proliferação ao redor do mundo aconteceu durante os anos 2000. Estes dados de transação coletados têm muito valor para a pesquisa geográfica de transportes, uma vez que estas tecnologias disponibilizam informação temporal e espacial contínua de todos os usuários de transporte público de uma cidade ou região (GUTIÉRREZ *et al.*, 2020).

Sistemas de *AFC* se mostram particularmente vantajosos para análises de transportes, como estudos de demanda, pois estes facilitam um entendimento detalhado das características espaço-temporais de padrões de viagem individuais (HALVORSEN *et al.*, 2020). Dados mais detalhados permitem análises mais completas de condições existentes, tal como a criação de modelos para previsão de impactos de possíveis mudanças. Ainda conforme

Halvorsen *et al.* (2020), dados de *AFC* podem facilitar análises de painel de usuários para providenciar entendimento de comportamentos antes e após a introdução de novas estratégias de gerenciamento de transportes.

Algumas características do sistema *AFC* são, segundo Zhao, Rahbee e Wilson (2007):

- a) somente entrada: passageiros somente utilizam o cartão e tem bilhetagem gravada ao realizar o pagamento; com isso, não se tem informações referentes ao destino ou momento de entrada dos passageiros. Visto que as estruturas de *AFC* foram destinadas conceptualmente somente para a coleta de tarifas e contabilidade de receitas, foge do seu escopo a obtenção de informações de entrada e saída, apesar de que estas seriam informações muito valiosas para a análise de operação e planejamento de tráfego;
- b) descrição da gravação: um novo registro é gravado toda vez que há uma nova leitura dos cartões (*smart cards*), constando o tempo, localização, veículo e identificador único.

Tal como sistemas de *AVL*, os sistemas de *AFC* podem ter ainda maior impacto positivo no planejamento uma vez que estes sejam combinados. Conforme descrito por Zhao, Rahbee e Wilson (2007), esta análise conjunta pode permitir a obtenção de dados de desempenho operacional que facilitará planejamentos mais robustos e confiáveis para ônibus. Eventualmente deve ser possível, inclusive, compreender o comportamento dos usuários do transporte público bem melhor. Com os dados de *AVL*, se torna possível caracterizar a qualidade percebida dos serviços em qualquer instante no tempo; em contrapartida, com dados de *AFC*, pode-se observar o comportamento dos indivíduos ao longo do tempo e relacionar estas informações com a condição de qualidade oferecida a cada momento.

[Estes dados] eventualmente nos permitirão compreender melhor tanto o desempenho multidimensional dos serviços de transporte público quanto seu impacto no comportamento das viagens. Isto representa um grande mar de mudanças na forma que os gerentes e planejadores de transporte público podem entender seus sistemas e clientes, e podemos esperar mudanças significativas na eficiência e eficácia dos resultados de transporte público no longo prazo. (ZHAO; RAHBEE; WILSON, 2007, p. 386, tradução nossa).

4.3 Visualização de Dados

“Na era da informação, a habilidade de ler e construir visualizações de dados se torna tão importante quanto a habilidade de ler e escrever textos.” (BÖRNER; BUECKLE; GINDA, 2019).

Matzen *et al.* (2018) afirma que a maior parte da humanidade depende da visão para navegar e entender o mundo físico e, devido a esta dependência visual, visualizações são ferramentas comuns para auxiliar as pessoas na compreensão de informações.

Por este prisma, segundo Lee, Sips e Seidel (2013) se utiliza vastamente em diversas áreas do conhecimento representações gráficas de dados de forma a assistir as pessoas no reconhecimento de padrões, detecção de tendências e na identificação de relações escondidas em conjuntos de dados. Visualização de dados pode ainda ser definida, brandamente, como a representação visual de dados quantitativos para comunicação e análise de informações (NESTOROV *et al.*, 2019).

Usualmente, a análise de dados visualizados é feita sem um modelo formal ou atividade de teste de hipóteses; analistas rapidamente e facilmente identificam padrões, tendências e *outliers* a partir de gráficos e figuras sem descrevê-los formalmente. Uma vez que os padrões são identificados é que se utiliza métodos estatísticos ou computacionais para validação das hipóteses e testes (NESTOROV *et al.*, 2019).

Apesar disto, visualizações podem também ser confusas e errôneas, sobretudo quando se tratando de sistemas de informações complexos e multidimensionais que não dispõem de uma representação visual natural (MATZEN *et al.*, 2018).

Sobretudo nas áreas de gerência e planejamento, sistemas de *Business Intelligence* vêm atraindo bastante atenção de executivos e tomadores de decisão dada a habilidade da tecnologia de providenciar informações complexas e competitivas para os processos decisórios por meio de visualização de dados (AIN *et al.*, 2019). Segundo Vallurupalli e Bose (2018), com a crescente quantidade de dados armazenados nas organizações, se tornou necessário integrar dados de diversas fontes de forma a determinar uma “única verdade”; com isso, provendo a visualização e a maior velocidade de acesso aos dados, tal como suprimindo a necessidade de ferramentas avançadas de análise, o *BI* foi amplamente adotado ao redor do mundo.

O *BI* permite, segundo Ain *et al.* (2019):

- a) facilitação de agregação, integração sistêmica e gerenciamento de dados estruturados e dados não estruturados;
- b) permite o trabalho com conjuntos de dados muito grandes (como *Big Data*);
- c) providencia ao usuário final grandes capacidades de processamento para a descoberta de novo conhecimento;
- d) oferece soluções de análise, consultas *ad hoc*, previsão e apresentação dos dados.

A partir de ferramentas de *BI*, o usuário tem o poder de criar gráficos, figuras, *widgets*, relatórios e consultas de forma a compreender indicadores importantes das organizações, integrando a visualização de dados ao gerenciamento de sistemas e empresas (AIN *et al.*, 2019).

4.4 Trabalhos Semelhantes e Resultados

Esta seção busca mencionar trabalhos com temáticas semelhantes. Para tal, pesquisou-se as palavras-chave *AVL*, *AFC*, *Indicadores* e *KPIs* no periódico CAPES, na base de dados da Elsevier, e no Google Acadêmico. Espera-se mostrar os resultados encontrados, tal como indicadores e casos de uso que ofereçam conclusões importantes para o presente trabalho.

Ma, Ferreira e Mesbah (2013) propõe em seu trabalho uma estrutura de desenvolvimento de indicadores de confiabilidade de serviços de ônibus a partir de fontes distintas de dados. Dentre os referentes à utilização de dados de *AVL*, são apontados:

- a) desempenho no horário: determina-se a partir da localização dos ônibus se estes partiram e chegaram nos pontos de ônibus e terminais no horário previsto; calcula-se percentualmente, avaliando quantas viagens saíram m minutos antes ou n minutos após o horário esperado;
- b) regularidade dos *headways*: determina-se quanto tempo ocorre entre a passagem de ônibus em um determinado ponto; há diversas formas de cálculo, como por exemplo pela distribuição dos *headways* ao longo do tempo, ou em comparação com o horário esperado de cada ônibus;
- c) tempo de viagem: determina-se também utilizando distribuições de tempo de viagem, a partir dos dados de localização;
- d) tempo de espera: pode-se calcular, de forma semelhante, assumindo regularidade na chegada dos passageiros ou utilizando métodos estocásticos. Utiliza-se os dados de *AVL* para determinar quanto tempo se espera para a chegada de um ônibus específico num ponto;
- e) tempo de transferência: calcula-se avaliando paradas obrigatórias;
- f) tempo de *buffer*: determinado pelo tempo que o passageiro espera alocar para compensar demoras maiores numa viagem; utiliza-se os dados de *AVL* para se estimar uma margem segura de tempo necessário.

Santos *et al.* (2020) utiliza dados de *AVL* para calcular os indicadores citados a fim de criar uma estrutura de suporte às decisões para a melhoria da confiabilidade nos transportes.

É determinado ainda, que é possível inferir horários de saída dos veículos a partir da validação de *tickets*, ou seja, usando dados de *AFC* e *AVL*.

Referente à confiabilidade de tempo, Barabino, Di Francesco e Mozzoni (2017) utiliza dados de *AVL* da região metropolitana de Cagliari (Itália) para avaliar o impacto das anomalias dos dados na determinação do indicador, comparando formas distintas de validação dos dados. O trabalho conclui, a partir da análise de 6 cenários distintos de validação de anomalias dos dados, que ajustar as anomalias é fortemente recomendado para uma análise correta de indicadores de regularidade. A análise de pontualidade dos ônibus, entretanto, é pouco afetada pelo tratamento das anomalias.

Faroqi, Mesbah e Kim (2018) apontam, num estudo de revisão de literatura, diversos usos de dados de *AFC*, dentre eles a estimação de matrizes de origem e destino. No caso de sistemas de *AFC* com dados de descida dos ônibus, este processo é direto, entretanto em casos de sistemas de *AFC* somente com dados de pagamento, é exigida a utilização de dados além deste, e análises estatísticas mais robustas. Faroqi, Mesbah e Kim (2018) apontam, ainda, o uso para a mineração de padrões de viagem, detecção de propósito das viagens, modelagem de escolha de rota e impactos de mudança de políticas. Referente ao cálculo de indicadores de desempenho, aponta-se a utilização para:

- a) cálculo de indicadores de demanda e oferta;
- b) cálculo de indicadores de desempenho operacional, como o número de corridas por linha de ônibus, indicadores de veículo-quilômetro e veículo-horas;
- c) cálculo de indicadores de uso do sistema, como número de passageiros, passageiros-quilômetro e passageiros-hora.

Em maior escopo, Hanft *et al.* (2016) aponta usos de dados de *AVL* e *AFC* na cidade de Nova York. A priori, desenvolveu-se um modelo conectando os dados de ambas as fontes para estimar a localização de subida e descida de todos os passageiros de ônibus; para isso, utiliza-se o horário de pagamento das passagens e a localização dos ônibus para estimar a subida. Utiliza-se os dados de novas passagens de cada cartão para inferir as descidas e transferências, visto que o sistema de *AFC* de Nova York não possui rastreamento de descida. Com isso, não há garantia de que as jornadas modeladas são compatíveis com a realidade. Apesar disto, estes dados são utilizados para o processo de planejamento.

Em Nova York, também foi possível utilizar os dados para comunicar o desempenho do sistema; a partir de reclamações dos passageiros, foi possível averiguar o mau desempenho de linhas específicas (HANFT *et al.*, 2016). Outro uso documentado foi a determinação de tipos de passageiros a partir dos dados de *AFC*, usando informações de tipo de

cartão (viagem única, cartão mensal etc.), desconto e gratuidade (como nos casos de passageiros idosos ou pessoas com deficiência) ou inferindo a natureza da viagem (trabalho, turismo etc.). Utilizou-se os dados, ainda, para a estimação de novas rotas e tempos de ciclo, visto que os modelos gerados possibilitaram a previsão dos indicadores das novas linhas.

De forma semelhante, em Pequim, Ma (2014) utilizou dados de *AFC* e *AVL* para a inferência das matrizes de origem e destino da cidade. Os dados permitiram, ainda o cálculo de indicadores de desempenho do sistema como a velocidade comercial a nível de rede, a confiabilidade de tempo a nível de linha, a estimação das subidas e descidas nos ônibus a nível de ponto de ônibus e a variação dos *headways* a nível de ponto de ônibus.

A partir destes indicadores, foi criada uma interface de visualização de dados, onde foi possível visualizar o mapa de velocidades a nível de rede, e analisar a distribuição dos *headways* a nível de linha, dentre outras informações importantes.

4.5 Utilização de dados de *AVL*, *AFC* e *GTFS* em Fortaleza

O trabalho de Braga, Loureiro e Pereira (2020) tem premissas e objetivos semelhantes ao abordado no atual projeto, e por se alimentar das mesmas bases de dados de Fortaleza, encontra diversos problemas que possivelmente serão encontrados também aqui.

Apesar da temática específica do trabalho ser referente à acessibilidade, algumas considerações são inespecíficas; por exemplo, o tratamento dos dados brutos. Para a utilização dos dados, Braga, Loureiro e Pereira (2020) sugerem utilizar os dados de *AFC* em conjunto com os de *AVL* para a criação de uma esquematização de tempos de parada. Ou seja, busca-se recriar a tabela de paradas prevista nos dados de *GTFS* (*General Transit Feed Specification*, formato de dados de transporte público mais comum) disponibilizados pela prefeitura de Fortaleza, porém desta vez com os dados reais estimados de chegada.

No atual trabalho, será fundamental conseguir determinar o horário dos ônibus em cada parada de Fortaleza, tanto para a estimativa de confiabilidade de horários, como para a determinação de indicadores de desempenho. O *headway* por parada, por exemplo, necessita da informação dos horários. Medidas como a demanda por parada, ou mesmo velocidade de percurso por parada, também seriam calculados a partir dos horários dos ônibus em cada parada.

Para a determinação do horário que cada ônibus passa em cada parada, Braga, Loureiro e Pereira (2020) sugere uma estimativa por meio de interpolação linear sobre os dados de *GPS* e as distâncias percorridas acumuladas para cada ônibus e linha versus parada.

O trabalho discute, ainda, que aproximadamente 15% dos dados de *GPS* são descartados por ausência de informações referentes à linha atendida (275 veículos de 1800 / dia). Além disso, parte das linhas previstas pelo SIT-FOR (Sistema Integrado de Transporte de Fortaleza) não contém informações em *GTFS*, e, com isso, são por ora impossíveis de trabalhar, uma vez que não têm estruturadas as paradas e tabela; isso gera outro descarte de 5% dos dados.

Por fim, é explicitado que as informações estruturadas de viagens podem ser utilizadas para estimar diversos indicadores, mais notadamente referente à confiabilidade e acessibilidade, e propõe inclusive que sejam utilizados em avaliações *ex-ante* e *ex-post* de possíveis intervenções, trazendo então melhor qualidade de estimação e maior sensibilidade nas tomadas de decisões (BRAGA; LOUREIRO; PEREIRA, 2020).

5 MATERIAIS E MÉTODOS

A seção atual é destinada a descrever o método a ser utilizado para o atingimento dos objetivos descritos. Os materiais, incluindo *softwares*, ferramentas e dados serão descritos em suas respectivas etapas metodológicas.

Cada subcapítulo presente neste capítulo se refere a um dos objetivos descritos na seção 3.2, na mesma ordem, e estes, por sua vez, estarão separados aqui por suas respectivas etapas metodológicas.

5.1 Criação da Base de Dados

A criação da base de dados é ancorada na normalização de bases de dados relacionadas, que consiste em separar as informações em múltiplas tabelas com informações únicas e não redundantes, servindo para diminuir a velocidade de leitura e escrita dos dados, evitando também sobrecarga de armazenamento (KÖHLER; LINK, 2018).

Após a criação da base de dados e do ambiente de servidor, busca-se tratar os dados brutos de geolocalização e bilhetagem de Fortaleza, para posteriormente popular as tabelas da base de dados com as informações no formato idealizado.

5.1.1 Teoria da Normalização e Idealização do Formato da Base de Dados

A teoria da normalização presume dois tipos principais de tabela, que são as tabelas fato (ou primárias, ou de primeira forma normal) e tabelas dimensão (ou secundária, ou de segunda ou maior forma normal). A tabela fato tem como característica descrever o evento que está tendo os dados registrados, como as vendas de um supermercado ou as transações em um sistema bancário. No caso das tabelas dimensão, elas se referem às características do fato, como o produto comprado ou o tipo de transação bancária efetuada.

No caso do presente trabalho, busca-se criar uma base de dados referente à localização e bilhetagem dos ônibus; para isto deve-se identificar nos dados brutos o que configurariam as tabelas fato e o que configuraria as dimensões que descrevem estes fatos. Visto que a base de dados pretende armazenar os dados de localização e bilhetagem, as tabelas fato principais necessariamente devem conter eventos únicos de localização e bilhetagem, respectivamente. Qualquer dado necessário para a caracterização do evento da localização ou

bilhetagem, como o número do ônibus ou linha de ônibus da bilhetagem deve ser tratado como dimensão numa tabela separada.

Por fim, deve-se reduzir as tabelas aos seus respectivos menores formatos normais, transferindo as chaves secundárias das tabelas para suas próprias tabelas, e criar os relacionamentos e restrições necessários para a manutenção da base de dados. Os relacionamentos servem para manter a integridade relacional da tabela entre as chaves primárias e secundárias, no caso de ausência de dados específicos. Por exemplo, se um produto de supermercado não tem registro válido, a transação contendo esse produto deve ser invalidada, já que não é rastreável. O relacionamento de produto e transação aqui descritos seria equiparável a detectar uma bilhetagem em um ônibus não existente na base de dados. Nesse exemplo, a transação teve chave primária sendo uma identificação única da transação, e chave secundária sendo o produto sendo comprado; no caso da tabela de produto, a chave primária é o produto (HINGORANI; GITTENS; EDWARDS, 2017).

Uma vez portando a estrutura idealizada que descreve todos os eventos necessários na base de dados, tal como seus relacionamentos e restrições, será utilizado o *MySQL*, um *software* gratuito de gerenciamento de servidores de base de dados para a criação das tabelas.

O *MySQL* utiliza a linguagem *SQL* (*Structured Query Language*, linguagem estruturada de consulta, em tradução livre) para seus comandos. No caso desta etapa, utiliza-se principalmente o comando *CREATE TABLE* (ORACLE, 2022) para criar as tabelas. A Figura 1 exemplifica um comando de criação de tabela na interface do *software*.

Figura 1 - Criação de Tabela de Posição do Ônibus no *MySQL*

```

1  CREATE TABLE t_posicao (
2      ID_POSICAO int NOT NULL,
3      ID_TABELA int NOT NULL,
4      DT_REGISTRO datetime,
5      LAT float,
6      LON float,
7      PRIMARY KEY (ID_POSICAO),
8      FOREIGN KEY (ID_TABELA) REFERENCES t_tabela(ID_TABELA)
9  )

```

Fonte: Elaborado pelo autor.

Neste exemplo, está sendo criada uma tabela cujo nome é “t_posicao”, com os campos:

- a) ID_POSICAO, armazenado como número inteiro;
- b) ID_TABELA, armazenado como número inteiro;
- c) DT_REGISTRO, armazenada com formato de data e tempo;
- d) LAT, armazenado como número decimal;
- e) LON, armazenado como número decimal.

As linhas 7 e 8 referem-se às chaves primária e secundária da tabela. Como a tabela trata das posições, a chave primária é o ID_POSICAO; isso implica que não pode haver nenhuma outra linha na tabela com o mesmo ID_POSICAO, mantendo somente posições únicas. No caso da chave secundária (ou *FOREIGN*, traduzida livremente para estrangeira), ela se refere ao ID_TABELA de outra tabela chamada “t_tabela”; nesse caso, T_TABELA é uma dimensão que caracteriza a posição. Por definição, não pode haver nenhum ID_TABELA na tabela T_POSICAO que não conste na T_TABELA, devido à esta restrição.

Por meio deste mesmo tipo de comando, deve-se criar todas as tabelas que compõem a base de dados proposta, finalizando a estrutura idealizada. Por fim, avalia-se o esquema relacional gerado pelo *MySQL* para visualizar a estrutura e suas dependências.

5.1.2 Criação das Rotinas de Tratamento de Dados para o Formato Idealizado

O resultado da etapa 5.1.1 é a estrutura de dados idealizada para as tabelas do servidor, porém os arquivos brutos disponibilizados referentes à Fortaleza não estão ainda no formato adequado para popular o banco. Esta etapa, portanto, tem como intuito tratar os dados brutos para que se adequem ao proposto.

Os dados brutos de *AFC*, *AVL* e *GTFS* são referentes a todos os ônibus de Fortaleza nos dias 22/10/2018 à 26/10/2018, compreendendo todos os dias úteis da semana. Os arquivos foram disponibilizados pelo Departamento de Engenharia de Tráfego da Universidade Federal do Ceará, e estão em diversos formatos distintos. Os arquivos são descritos abaixo:

- a) arquivos de *AVL*: são arquivos *csv* (valores separados por vírgulas), separados por dia de registros dentre as datas disponibilizadas;
- b) arquivos de *AFC*: são arquivos *csv*, separados por dia de registros, tal como os dados de *AVL*;
- c) arquivos de *GTFS*: são uma série de arquivos em formato *txt* (arquivo de texto simples) e *shapes* (arquivos de geolocalização) descrevendo as rotas dos ônibus, as paradas de ônibus, formatos das linhas de ônibus, horários previstos e listagem de viagens no período.

A partir destes arquivos, cria-se os arquivos tratados no formato determinado por meio de rotinas em *Python*, uma das mais comuns linguagens de programação orientada a objetos. A ferramenta que será utilizada para escrever e compilar os códigos será o *Google Collab*, por ser gratuito e utilizar memória e discos virtuais, evitando o consumo de recursos pessoais. Como o tratamento dos dados envolve grandes volumes de dados, será utilizada a biblioteca *pandas* do *Python*, que permite a manipulação de tabelas por meio de *dataframes* (tabelas com índice e com tipos de dados assignados).

Os arquivos tratados terão formato *csv*, e serão nomeados a partir dos nomes das tabelas propostos na etapa 5.1.1, seguidos do nome da extensão do arquivo: “.csv”. No caso do exemplo de tabela descrito na Figura 1, o arquivo deve ser nomeado “t_posicao.csv”. A rotina de tratamento deve incluir a etapa de criação dos arquivos.

O código criado será disponibilizado e permitirá a introdução de novos dados, contanto que obedeçam a estrutura utilizada.

5.1.3 População das Tabelas

Esta etapa da criação do banco de dados envolve utilizar os arquivos que serão gerados a partir das rotinas da etapa 5.1.2 para preencher as tabelas da base de dados criada na etapa 5.1.1, finalizando o processo de criação da base.

Como o banco de dados criado utiliza a linguagem *SQL*, deve-se escrever os comandos que fazem inserção de dados em tabelas por meio de arquivos. Este comando é o *LOAD DATA*, conforme descrito no manual referencial do *software* (ORACLE, 2022). A Figura 2 mostra um exemplo de comando de carregamento de arquivos em tabelas no *MySQL*.

Figura 2 - Comando de Inserção de Dados de Arquivo *csv* em Tabela no *MySQL*

```
1 • LOAD DATA
2     INFILE 't_posicao.csv'
3     INTO TABLE t_posicao
4     FIELDS
5         TERMINATED BY ','
6     IGNORE 1 ROWS
```

Fonte: Elaborado pelo autor.

Neste caso, o comando carrega os dados do arquivo “t_posicao.csv” para a tabela “t_posicao” do banco de dados. As linhas 4 e 5 explicitam que os campos no arquivo são separados por vírgulas (por se tratar de arquivos *csv*), e a linha 6 determina que se deve ignorar 1 (uma) linha, visto que a primeira linha dos arquivos são o cabeçalho nomeando os campos.

Aos moldes deste exemplo, deve-se criar e executar estes comandos para todas as tabelas determinadas no banco de dados idealizados, assim preenchendo todas as tabelas com as informações necessárias. A própria ferramenta do *MySQL* aplica as restrições das chaves que foram criadas, mantendo a integridade relacional dos dados. Uma vez carregados todos os arquivos nas suas respectivas tabelas, estará finalizada a criação do banco de dados.

5.2 Modelagem dos Horários do Ônibus

A presente seção tem como intuito apresentar o método para atingimento do segundo objetivo específico do trabalho, referente à modelagem do horário que cada ônibus passa em cada ponto de sua respectiva linha. Os dados de *GTFS* apresentam os horários esperados de chegada de ônibus por parada, porém não o realmente realizado. A ideia daqui é, portanto, determinar, a partir dos dados de *AFC* e *AVL*, uma estimativa do horário realmente realizado por parada. Os dados de *AFC* são necessários para a estimativa pois apesar dos dados de *AVL* de Fortaleza conterem a identificação do ônibus que está sendo registrado, não se tem a informação da linha que está sendo atendida pelo veículo; por conta disso, a linha (e por consequência, os pontos de ônibus válidos) deve ser inferida a partir das bilhetagens.

Essa determinação é interessante pois permite determinar, por exemplo, o *headway* médio por parada e os tempos de ciclo e de viagem de forma fácil, pois todos dependem inerentemente do horário que um ônibus toca uma parada específica. No caso do *headway* médio, pode-se avaliar para uma única parada toda vez que um ônibus passou por ela; no caso do tempo de ciclo, bastaria avaliar quando um mesmo ônibus passou pela parada inicial novamente, e no caso do tempo de viagem, bastaria avaliar quando um ônibus passou pela primeira e pela última parada.

5.2.1 Aproximação Por Parada

Define-se aproximação como cada conjunto de posições que um ônibus descreve ao se aproximar de uma parada de ônibus válida. Nesta etapa inicial, utiliza-se os dados de

GTFS referentes às paradas que cada linha deve atender para avaliar, para cada parada, a aproximação dos ônibus que pertencem àquela linha.

Para tal, utiliza-se os dados tratados do banco de dados de posição dos ônibus, contendo a linha a qual ele pertence naquele instante e, cruzando com os dados de parada, determina-se todas as possíveis aproximações daquele ônibus com cada uma de suas paradas.

Para determinar se um ônibus está se aproximando de uma dada parada, filtra-se, dentre todas as posições do ônibus, aquelas em que o ônibus se encontra a pelo menos 500 metros de distância da parada específica. A distância de 500 metros foi selecionada pois, visto que se tem registros de posição dos ônibus a cada 30 segundos, caso o ônibus venha numa via reta, seriam detectadas ao menos 2 posições no caso de o ônibus estar trafegando a 60 km/h. Para determinar a distância de um ônibus para uma parada, portando tanto as posições (contendo latitude, longitude e horário da posição) e as paradas a qual ele deve atender (contendo latitude e longitude), utiliza-se a equação de Haversine, que descreve a distância sobre uma esfera (UTOMO; MAYLAWATI; ALAM, 2018), apresentada a seguir:

$$d = 2r \arcsin \left(\sqrt{\sin^2 \left(\frac{\varphi_2 - \varphi_1}{2} \right) + \cos \varphi_1 \cdot \cos \varphi_2 \cdot \sin^2 \left(\frac{\lambda_2 - \lambda_1}{2} \right)} \right) \quad (1)$$

Onde:

d é a distância entre dois pontos, arbitrariamente chamados de 1 e 2;

r é o raio da esfera, neste caso sendo o raio da Terra, valendo 6.371 km;

φ_1 e φ_2 são as latitudes dos pontos 1 e 2;

λ_1 e λ_2 são as longitudes dos pontos 1 e 2.

A partir da fórmula, pode-se checar, para cada posição de um ônibus, se ele está ao menos a 500 metros de uma parada específica. Caso esteja, guarda-se a informação, e pode-se avaliar visualmente por meio de algum *software* de visualização espacial. Visto que foi construído o banco de dados na seção 5.1, essa verificação das distâncias deve ser feita por meio de uma consulta às tabelas de posição e paradas, implementando a equação de Haversine em *SQL*.

Para a avaliação visual, utiliza-se neste trabalho o *software Power BI* para plotar as aproximações de uma parada. A Figura 3 exemplifica todas as aproximações de um ônibus (ônibus com ID_ONIBUS 30535 na tabela tratada) para a parada 4169 (ponto de ônibus na Av. Antônio Sales, 3510; 40ª parada de ida da linha 75).

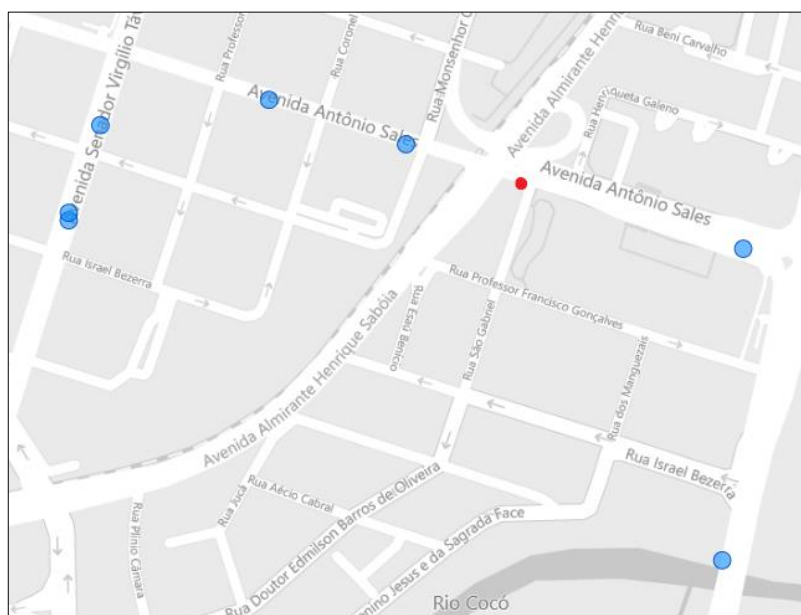
Figura 4 - Agrupamento das Aproximações do Ponto de Controle 4169 pelo Ônibus 30535

Posição ▾	Latitude ▾	Longitude ▾	Data do Registro ▾	Grupo ▾
1	-3,7489	-38,4970	22/10/2018 11:43	1
2	-3,7488	-38,4970	22/10/2018 11:43	1
3	-3,7480	-38,4967	22/10/2018 11:44	1
4	-3,7478	-38,4951	22/10/2018 11:44	1
5	-3,7482	-38,4938	22/10/2018 11:45	1
6	-3,7492	-38,4906	22/10/2018 11:45	1
7	-3,7521	-38,4908	22/10/2018 11:46	1
8	-3,7490	-38,4970	22/10/2018 18:20	2
9	-3,7483	-38,4968	22/10/2018 18:20	2
10	-3,7472	-38,4964	22/10/2018 18:21	2
11	-3,7478	-38,4948	22/10/2018 18:21	2
12	-3,7485	-38,4927	22/10/2018 18:22	2
13	-3,7491	-38,4907	22/10/2018 18:22	2
14	-3,7516	-38,4907	22/10/2018 18:23	2
15	-3,7488	-38,4969	22/10/2018 21:48	3
16	-3,7487	-38,4969	22/10/2018 21:49	3
17	-3,7474	-38,4965	22/10/2018 21:49	3
18	-3,7478	-38,4949	22/10/2018 21:50	3
19	-3,7483	-38,4931	22/10/2018 21:50	3
20	-3,7499	-38,4903	22/10/2018 21:51	3
21	-3,7515	-38,4906	22/10/2018 21:51	3
22	-3,7523	-38,4908	22/10/2018 21:52	3
23	-3,7489	-38,4970	23/10/2018 00:18	4

Fonte: Elaborado pelo autor.

Conforme descrito, os registros de posição 1 a 7 se referem ao mesmo grupo, pois todas as datas de registro caem dentro de um minuto da anterior. A posição 8, porém, acontece horas após a posição 7, indicando nova aproximação, e, portanto, recebe o grupo 2. Todas as observações seguintes, até a 14, mantém o grupo 2. A observação 15 já ocorre horas depois, indicando outra aproximação, e assim sucessivamente. De forma análoga, a Figura 5 mostra a aproximação do grupo 1 visualmente.

Figura 5 - Aproximação do Grupo 1 do Ônibus 30535 com o Ponto de Controle 4169



Fonte: Elaborado pelo autor.

Neste caso, observa-se que apenas são visíveis as 7 posições registradas do ônibus durante aquela passagem pela parada. Assim, pode-se avaliar individualmente cada aproximação e eventualmente determinar o momento de passagem na parada, a partir dos horários de cada posição.

Este agrupamento deve ser realizado para todas as aproximações de todos os ônibus com todas as paradas, permitindo avaliar cada aproximação única separadamente.

5.2.2 Diagramas de Aproximação das Paradas

Para a detecção do momento exato da passagem de um ônibus na parada, utiliza-se diagramas de espaço percorrido e tempo. Para cada aproximação, plota-se no eixo vertical a distância real a ser percorrida de cada posição até o ponto de ônibus avaliado, e no eixo horizontal, a variação temporal. Usualmente, um novo registro de posição ocorre a cada 30 segundos, então a variação temporal tende a ser fixa.

Espera-se que os diagramas exibam duas etapas:

- a) etapa de aproximação: o ônibus tende a se aproximar da parada a partir do momento que é detectado. Com isso, a primeira etapa deve mostrar queda na distância até a parada, aproximando-se da distância nula;

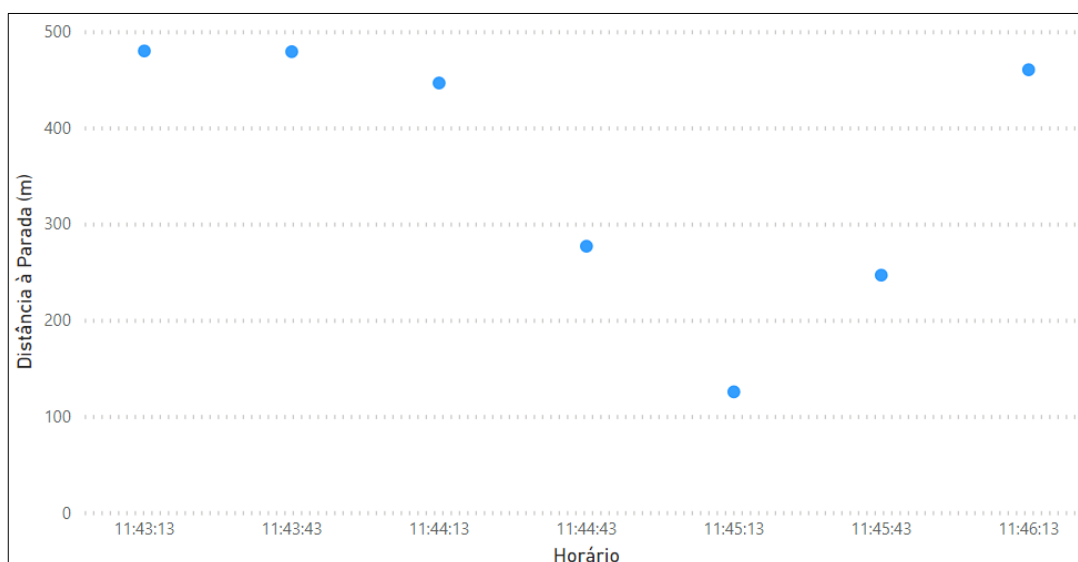
- b) etapa de distanciamento: o ônibus tende a se distanciar da parada após atingi-la, então a segunda etapa deve mostrar aumento na distância, até aproximar-se novamente de 500 m, onde a detecção é finalizada.

Para os diagramas que apresentem esse comportamento, assume-se que o ônibus realmente está fazendo uma aproximação com o ponto de ônibus. Caso o diagrama não siga esse comportamento, pode ter sido detectada alguma outra passagem do ônibus no raio de detecção do ponto de ônibus, mas que não configura uma aproximação. Essas amostras devem ser descartas, visto que o ônibus não passou realmente no ponto de ônibus. Isso ocorre, por exemplo, quando um ônibus está voltando à garagem da sua empresa, e por acaso passa próximo de uma parada.

No caso dos diagramas que seguem o comportamento usual, utiliza-se a velocidade média do ônibus na etapa (a) da aproximação para modelar o ponto onde a distância do ônibus à parada é nula, ou seja, o momento em que o ônibus chegou de fato no ponto de ônibus.

A Figura 6 exemplifica esta plotagem para a aproximação apresentada na Figura 5, mostrando todas as distâncias e horários respectivos.

Figura 6 - Diagrama de Distância à Parada por Horário do Grupo 1 do Ônibus 30535 ao Ponto de Controle 4169



Fonte: Elaborado pelo autor.

Neste exemplo, pode-se ver que a etapa (a) dessa aproximação se refere aos horários 11:43:13 até o horário 11:45:13, que é quando a distância do ônibus a parada vai diminuindo. A etapa (b) seria dos horários 11:45:13 até 11:46:13, que é o distanciamento do ônibus da parada até sair do raio de detecção. A partir deste tipo de diagrama, pode-se calcular a velocidade média do ônibus a partir da distância percorrida e do tempo gasto, e, com isso, pode-se

interpolando em que momento o ônibus chega na posição de distância 0, a partir dos pontos da etapa (a).

Este método é aplicado, por meio de rotinas computacionais e consultas em *SQL* ao banco de dados, em todas as aproximações, assim modelando os instantes estimados de chegada em todas as paradas.

5.2.3 *Correção do Sentido na Modelagem do Horário*

Visto que a direção do ônibus (sentido ida ou volta) não é informada a priori nos arquivos brutos e tabelas disponibilizadas de forma confiável, pode ocorrer de uma aproximação detectada ser referente ao sentido contrário da viagem. É especialmente comum em linhas que tem rota de ida e volta semelhante, passando pelas mesmas vias contendo pontos de ônibus nas duas direções.

O método aplicado na seção 5.2.2 não detecta esses erros, pois da mesma maneira nos sentidos de ida e volta, alguns ônibus se aproximam e afastam consistentemente das mesmas paradas. Para corrigir isso, esta etapa do método avalia o contexto da viagem de cada ônibus para averiguar se o ônibus está realmente passando no ponto de ônibus que faz sentido com sua viagem naquele instante.

Uma vez que é tido os resultados da modelagem inicial dos horários dos ônibus, basta avaliar, para cada instante de passagem na parada modelado, se anteriormente a este instante o ônibus estava vindo da parada anterior e se posteriormente a este instante o ônibus estava indo para a parada seguinte. Por exemplo, um ônibus que passou pela n -ésima parada necessariamente deve ter vindo da parada $n - 1$ e estar indo para a parada $n + 1$ em sequência.

Caso o ônibus obedeça ao padrão estabelecido, mantém-se a observação como válida, e em casos que não obedeçam, descarta-se do banco de dados aquelas aproximações. A partir disto, pode-se ainda determinar de forma confiável se um ônibus está no sentido de ida ou de volta, nas linhas que tenham esse tipo de separação.

No caso das paradas de ônibus de início e fim de linha, avalia-se somente as paradas seguintes e anteriores respectivamente, uma vez que não existe parada anterior à inicial, tampouco posterior à final. Ademais, o horário modelado para a parada inicial se refere sempre ao momento que o ônibus iniciou o trajeto (começou a sair da parada inicial), visto que não necessariamente o ônibus parte imediatamente, como nas paradas intermediárias da viagem.

A aplicação do método descrito nesta etapa é feita por consultas no banco de dados gerado, deletando linhas que quebram o padrão descrito. O comando utilizado, conforme no

manual referencial do *MySQL*, é o *DELETE*, condicionado por comparação particionada por ônibus e ponto de chegada, avaliando se o ponto de chegada atual é maior que o anterior e inferior ou próximo daquele mesmo ônibus (ORACLE, 2022).

5.2.4 Criação da Tabela de Acompanhamento

Para finalizar a modelagem, cria-se uma tabela de acompanhamento no banco de dados, com o intuito único de deixar mapeado e consultável todas as vezes que um ônibus passa numa parada de ônibus referente à sua linha e sentido.

Os campos da tabela devem conter as informações necessárias para os cálculos dos indicadores mais importantes, sendo elas:

- a) identificador único da passagem: seguindo os padrões de base relacionadas e normalização, é necessário um identificador único para cada passagem de ônibus em uma parada;
- b) identificador do ônibus: para cada passagem, necessita-se da identificação do ônibus. Esse valor é chave secundária da tabela de ônibus;
- c) identificador do ponto de controle: cada ponto de controle se refere a um ponto de ônibus de uma linha específica. A partir do ponto de controle, sabe-se a parada, linha e sentido daquele registro;
- d) data de passagem na parada: a data do registro, identificando de fato o horário da passagem.

A criação da tabela em si ocorre de maneira semelhante ao capítulo 5.1, onde utiliza-se o comando *CREATE TABLE*. Desta vez, porém, o carregamento dos dados advém das consultas criadas a partir das etapas de modelagem anteriores, vindo de dados internos ao próprio banco. Nesse caso, utiliza-se o comando *INSERT* (ORACLE, 2022). A Figura 7 exemplifica um comando de criação e inserção de dados, baseado no banco criado neste trabalho.

Figura 7 - Consulta de Criação de Tabela de Acompanhamento e Inserção de Dados

```

1  CREATE TABLE t_acompanhamento (
2      ID_ACOMPANHAMENTO int NOT NULL AUTO_INCREMENT,
3      ID_ONIBUS int NOT NULL,
4      ID_PONTO_DE_CONTROLE int NOT NULL,
5      DT_CHEGADA datetime,
6      PRIMARY KEY (ID_ACOMPANHAMENTO)
7  );
8
9  INSERT
10 INTO t_acompanhamento_real
11     (ID_ONIBUS, ID_PONTO_DE_CONTROLE, DT_CHEGADA)
12     SELECT
13         ID_ONIBUS,
14         ID_PONTO_DE_CONTROLE,
15         DT_CHEGADA
16     FROM
17         chegadas_na_parada_validas

```

Fonte: Elaborado pelo autor.

As linhas 1 a 7 compreendem a criação da tabela de acompanhamento, determinando ainda os campos descritos anteriormente nesta seção. Nas linhas subsequentes, apresenta-se a inserção dos dados de chegada nas paradas na tabela criada, a partir de uma seleção (*SELECT*) dos dados da tabela temporária “chegadas_na_parada_validas”, que por sua vez advém da etapa 5.2.3, já com o tratamento dos dados nos sentidos incorretos.

Fazendo este procedimento para todas as paradas modeladas, finaliza-se o processo de modelagem das chegadas e alimentação do banco de dados.

5.3 Determinação dos Indicadores

Objetivando responder ao terceiro objetivo específico, a presente seção do trabalho tem como intuito expor o método de determinação e cálculo dos indicadores importantes de planejamento e operação do sistema de transportes, além dos principais indicadores de qualidade e eficiência (FERRAZ; TORRES, 2001).

5.3.1 *Compatibilização dos Indicadores*

Esta etapa metodológica tem como intuito compatibilizar os principais indicadores, descritos na seção 4.1, com os dados tratados contidos no banco de dados criado no trabalho. Alguns indicadores não podem ser calculados diretamente a partir dos dados do banco, pois este carece de informações, por exemplo, de custo e tipo de ônibus. Essa etapa, portanto, busca fazer uma avaliação prévia de que indicadores os dados permitem calcular, e quais não é possível.

Abaixo, lista-se os indicadores principais levantados na revisão bibliográfica, separados por classe operacional, eficiência e qualidade:

- a) indicadores operacionais:
 - tempo de ciclo;
 - tempo em viagem;
 - velocidade operacional;
 - velocidade comercial;
 - *headway* médio;
 - frota;
 - demanda crítica;
- b) indicadores de eficiência:
 - distância média entre paradas;
 - índice de quilômetros por veículo;
 - índice de aproveitamento da frota;
 - índice de passageiros por quilômetro;
 - índice de passageiros por veículo;
 - custo por passageiro;
 - razão arrecadado sobre custo;
- c) indicadores de qualidade:
 - frequência de atendimento;
 - tempo de viagem médio;
 - lotação média;
 - coeficiente de variação do *headway*;
 - índice do tempo de buffer médio.

O método proposto de compatibilização presume que, para cada indicador, deve-se agrupá-los conforme a possibilidade do cálculo a partir dos dados tratados. Os grupos propostos são descritos abaixo:

- a) grupo 1: indicador pode ser calculado diretamente a partir dos dados do banco;
- b) grupo 2: indicador pode ser calculado de forma indireta ou estimada, mas necessita considerações de cálculo ou premissas simplificadoras;
- c) grupo 3: indicador não pode ser calculado, uma vez os dados são insuficientes para a avaliação, mesmo parcial.

Uma vez separados por grupos, descarta-se a possibilidade de cálculo, para o escopo deste trabalho, dos indicadores pertencentes ao grupo 3. Os indicadores do grupo 2 terão suas considerações pautadas na seção 5.3.2. Juntamente, tanto os indicadores do grupo 1 quanto do grupo 2 seguirão para o cálculo conforme proposta na seção 0.

5.3.2 Considerações de Cálculo para Indicadores Indiretos

Nesta etapa, busca-se formular considerações de cálculo para os indicadores do grupo 2, descritos na seção 5.3.1. Para tal, o método consiste em avaliar considerações simplificadoras na literatura ou propor novas premissas de cálculo, de forma a possibilitar o cálculo destes indicadores.

O método proposto é buscar na literatura citada na revisão bibliográfica, se há proposições que demonstram ser possível calcular os indicadores deste grupo a partir dos dados de *AVL* e *AFC*. No caso dos indicadores que já tenham proposta clara de estimativa, sugere-se que esta consideração de cálculo seja adotada e aplicada para estes indicadores na seção 0.

No caso de o indicador não ter proposição prévia na literatura, indica-se avaliar que informações próximas ao indicador desejado podem ser calculadas, até atingir o ponto limite de onde não se pode prosseguir com o cálculo. Neste momento, pode-se gerar hipóteses de cálculo ou proposições simplificadas para a estimativa. Os métodos adotados serão utilizados na seção seguinte para o cálculo dos indicadores.

Um exemplo possível de aplicação do método é o cálculo do índice de tempo de *buffer* de uma viagem numa determinada linha de ônibus numa determinada faixa horária. Como não se sabe o momento de descida dos usuários, não se pode determinar a priori o tempo de viagem por usuário. Entretanto, por saber o tempo de viagem total do ônibus, pode-se determinar um valor de índice de *buffer* geral, levando em conta todo o tempo de viagem da linha. Este valor por si só não representa necessariamente o índice de *buffer* de tempo, porém é

uma estimativa simplificada, assumindo que a variação do tempo de viagem individual equivale à variação do tempo de viagem total do ônibus.

5.3.3 Cálculo dos Indicadores

Nesta etapa, busca-se expor como serão calculados os indicadores a partir dos insumos advindos do restante deste capítulo. Cada subcapítulo abordará uma classe de indicador (operacional, eficiência ou qualidade), e cada subcapítulo será dividido por indicador, contendo sua formulação e o método de cálculo a partir do banco de dados gerado com as informações de *AVL* e *AFC*.

5.3.3.1 Indicadores Operacionais

Os indicadores operacionais tratam daqueles referentes ao que realmente ocorre na linha, desde a demanda real, ao tempo real de ciclo, viagem, e a velocidade dos veículos.

5.3.3.1.1 Tempo de Ciclo

O tempo de ciclo é o tempo que um ônibus demora para fazer todo o seu trajeto, incluindo as pausas nos terminais, até que chegue novamente na parada de onde iniciou a mensuração.

Para fins de cálculo, basta determinar o instante em que o ônibus passou em uma determinada parada, usualmente a inicial, e mede-se quando ele passa novamente na mesma parada. A diferença entre o tempo inicial e o tempo medido no final é o tempo de ciclo. O mesmo ônibus pode ter vários tempos de ciclo ao longo do dia, contanto que faça várias viagens.

Para uma parada p , e H_i sendo o horário que o ônibus passou pela parada pela i -ésima vez, o tempo de ciclo T_c pode ser calculado conforme a seguinte equação:

$$T_c(p, i) = H_{i+1} - H_i \quad (2)$$

O tempo de ciclo médio de um ônibus, assumindo n passagens pela parada, portanto, pode ser calculado pela equação:

$$\bar{T}_c(p) = \frac{\sum_{i=1}^n H_{i+1} - H_i}{n} \quad (3)$$

Isso pode ser traduzido em *SQL* para um ônibus e uma parada específica por meio de uma função chamada *LEAD* (ORACLE, 2022), que traz o valor seguinte de uma determinada partição. Por exemplo, no caso da tabela de acompanhamento de paradas, no banco de dados,

tem-se todas as vezes que todos os ônibus passaram por cada uma de suas paradas específicas. Nesse caso, pode-se particionar os dados por ônibus e parada, e avaliar, ordenando por horário do registro, todas as vezes que um ônibus passou em uma determinada parada. Um exemplo de consulta que faz isso é exibido na Figura 8.

Figura 8 - Consulta de Determinação do Tempo de Ciclo do Ônibus 30535 no Ponto de Controle 4169

```

1  *  SELECT
2      ID_ACOMPANHAMENTO,
3      ID_ONIBUS,
4      ID_PONTO_DE_CONTROLE,
5      DT_CHEGADA,
6      LEAD(DT_CHEGADA) OVER (ORDER BY DT_CHEGADA) `DT_CHEGADA_SEGUINTE`
7  FROM
8      t_acompanhamento
9  WHERE
10     ID_ONIBUS = 30535
11     AND ID_PONTO_DE_CONTROLE = 4169
12  ORDER BY
13     DT_CHEGADA
14  LIMIT 5

```

Fonte: Elaborado pelo autor.

A consulta seleciona, da tabela “t_acompanhamento”, os campos identificando o ônibus, o ponto de controle (parada de ônibus), a data de chegada e a data de chegada seguinte para o ônibus com identificação 30535 e ponto de controle com identificação 4169, ordenando por data de chegada na parada e limitando a somente os 5 primeiros resultados. A resposta da consulta é exibida no Quadro 5.

Quadro 5 - Horários de Chegada do Ônibus 30535 no Ponto de Controle 4169

Identificador	Ônibus	Parada	Data de Chegada	Chegada Seguinte
15514	30535	4169	22/10/2018 11:45	22/10/2018 14:26
15515	30535	4169	22/10/2018 14:26	22/10/2018 18:22
15516	30535	4169	22/10/2018 18:22	22/10/2018 21:50
15517	30535	4169	22/10/2018 21:50	22/10/2018 00:19
15518	30535	4169	22/10/2018 00:19	22/10/2018 10:03

Fonte: Elaborado pelo autor.

Para cada linha do quadro, pode-se ver o horário que o ônibus chegou na parada e o horário da próxima vez que este ônibus chegou na parada novamente. A diferença entre estes horários será, portanto, o tempo de ciclo deste ônibus para esta parada. Atenta-se que no quinto registro, o ônibus só passa novamente na parada horas depois. Isso se deve ao fato de que a viagem que teve data de chegada às 00:19 deve ter sido sua última, e, portanto, só passaria na parada novamente no dia seguinte.

Este cálculo deve ser feito para todos os ônibus, assim determinando todos os tempos de ciclo reais por faixa horária, ônibus, parada ou data.

5.3.3.1.2 Tempo em Viagem

De forma semelhante, o tempo em viagem é descrito pelo tempo que o ônibus esteve em movimento, descontando os tempos parados nos terminais.

No ponto de vista do banco de dados, que tem dados referentes às chegadas em cada parada, pode-se determinar o tempo de viagem observando os tempos de viagem entre as paradas iniciais e finais de cada sentido. Por exemplo, determina-se a parada inicial e final do sentido de ida de uma linha; o tempo em viagem na ida será, portanto, a diferença entre o horário de chegada na parada final e inicial. Nas linhas com ambos os sentidos, soma-se o tempo em viagem para os dois sentidos. Dessa forma, desconta-se o tempo que o ônibus fica parado entre os sentidos.

Matematicamente, o tempo em viagem T_v , para a n -ésima viagem de um ônibus em um trecho seria definido pela equação abaixo, sendo p_i a parada inicial do trecho, p_f a parada final, e $H(p)$ o horário que o ônibus passou pela parada p :

$$T_v(n) = H(p_f(n)) - H(p_i(n)) \quad (4)$$

Em *SQL*, isto pode ser traduzido por meio da junção de duas consultas, uma selecionando o tempo de chegada na parada inicial, e outra selecionando o tempo de chegada na parada final. A Figura 9 exemplifica uma consulta ao banco com este intuito, determinando inicialmente as paradas iniciais e finais de uma determinada rota (no exemplo, sentido ida da linha 75 – Campus do Pici / Unifor), e a Figura 10 calcula, para o ônibus 30535, todas as vezes que este passou pelas paradas iniciais e finais no dia 22/10/2018, juntando os resultados.

Figura 9 - Consulta de Determinação da Parada Inicial e Final da Linha 75 no Sentido Ida

```

1 * SELECT
2     MIN(ID_PONTO_DE_CONTROLE) `PONTO_INICIAL`,
3     MAX(ID_PONTO_DE_CONTROLE) `PONTO_FINAL`
4 FROM
5     t_ponto_de_controle PC
6     INNER JOIN t_rota R
7         ON PC.ID_ROTA = R.ID_ROTA
8         AND ID_LINHA = 75
9         AND FL_SENTIDO = 0

```

Fonte: Elaborado pelo autor.

Figura 10 - Consulta para Determinação do Horário de Chegada na Parada Inicial e na Parada Final do Ônibus 30535

```

1 * WITH
2     cte_parada_inicial AS (
3         SELECT
4             *,
5             ROW_NUMBER() OVER (ORDER BY DT_CHEGADA) `VIAGEM`
6         FROM t_acompanhamento_real
7         WHERE ID_ONIBUS = 30535
8             AND ID_PONTO_DE_CONTROLE = 4130
9             AND CONVERT(DT_CHEGADA, DATE) = '2018-10-22'
10        ORDER BY DT_CHEGADA),
11     cte_parada_final AS (
12         SELECT
13             *,
14             ROW_NUMBER() OVER (ORDER BY DT_CHEGADA) `VIAGEM`
15         FROM t_acompanhamento_real
16         WHERE ID_ONIBUS = 30535
17             AND ID_PONTO_DE_CONTROLE = 4180
18             AND CONVERT(DT_CHEGADA, DATE) = '2018-10-22'
19        ORDER BY DT_CHEGADA)
20 SELECT
21     cte1.ID_ONIBUS,
22     cte1.DT_CHEGADA,
23     cte2.DT_CHEGADA
24 FROM
25     cte_parada_inicial cte1
26     INNER JOIN cte_parada_final cte2
27         ON cte1.VIAGEM = cte2.VIAGEM

```

Fonte: Elaborado pelo autor.

Na consulta, cria-se duas *CTE* (*common table expressions*, ou expressões de tabela comuns, em tradução), uma (“cte_parada_inicial”) calculando todas as vezes que o ônibus 30535 passa pela parada inicial (4130) na data 22/10/2018 e outra (“cte_parada_final”) calculando o mesmo, mas para a parada final. Ambos são ordenados por data de chegada e enumerados, tal que o registro inicial seria a primeira vez que o ônibus passou pela parada inicial e pela final, o segundo seria a segunda vez etc. Por fim, junta-se por meio de um *INNER JOIN*, as duas tabelas, tal que estejam na mesma viagem.

Essa consulta retorna um quadro exibindo o ônibus, a data e hora que ele chegou na parada inicial e a data e hora que ele chegou na parada final, exemplificado no Quadro 6.

Quadro 6 - Horários de Chegada na Parada Inicial e na Parada Final do Ônibus 30535 na Linha 75, Sentido Ida, no dia 22/10/2018

Ônibus	Parada Inicial	Parada Final
30535	22/10/2018 10:47	22/10/2018 12:33
30535	22/10/2018 13:35	22/10/2018 16:20
30535	22/10/2018 17:26	22/10/2018 19:35
30535	22/10/2018 20:53	22/10/2018 22:25

Fonte: Elaborado pelo autor.

A partir do quadro gerado, pode-se determinar o tempo em viagem entre a parada inicial e parada final pela diferença entre os tempos, onde cada linha representa uma viagem. Este método deve ser aplicado para todos os ônibus e linhas, permitindo o mapeamento de todos os tempos em viagem.

5.3.3.1.3 Velocidade Operacional e Velocidade Comercial

A velocidade operacional é descrita pela razão entre a distância percorrida na viagem pelo ônibus e o tempo em viagem do ônibus. A velocidade comercial, por outro lado, é a razão entre a distância percorrida e o tempo de ciclo. Como já é tido o tempo em viagem e tempo de ciclo pelos métodos descritos na seção anterior, basta-se determinar um método para o cálculo da distância percorrida na viagem.

Nos dados de *GTFS* e, por conseguinte, no banco de dados criado, há uma tabela referente ao percurso de todas as linhas de ônibus. O arquivo original do *GTFS* é denominado “*shapes.txt*”, e a tabela correspondente tem o nome “*t_shape*” no banco de dados. Para cada ponto no *shape* das linhas, têm-se os pares de latitude e longitude. A partir destes dados, pode-

se calcular, para cada linha, o comprimento total das linhas e entre cada paradas por meio da equação de Haversine.

Portando as distâncias percorridas e os tempos, basta simplesmente calcular a razão entre os valores. O método deve ser aplicado para todo o banco de dados, contendo todas as viagens e ciclos.

5.3.3.1.4 *Headway* Médio

O *headway* é descrito pelo tempo entre a passagem de dois veículos pela mesma posição ou ponto de controle. No caso do sistema de transportes de ônibus aqui descrito, considera-se o *headway* o tempo entre a passagem de dois ônibus por uma mesma parada.

A formulação matemática para o *headway* h , sendo p a parada observada e H_n o horário da n -ésima vez que passou um ônibus esperado por ela é:

$$h(n, p) = H_n - H_{n-1} \quad (5)$$

Como já se tem modelado o horário de passagem dos ônibus por cada parada, para determinar todas as observações de *headway*, basta selecionar do banco de dados, para uma parada específica, todos os horários em que um ônibus passou, e ordenar por data do registro; por fim, basta calcular a diferença entre cada horário de passagem, obtendo assim, para passagens de ônibus, $n - 1$ observações de *headway*. A consulta na figura exemplifica esta seleção.

Figura 11 - Consulta para Determinação do *Headway* para o Ponto de Controle 4169

```

1  SELECT
2      *
3  FROM
4      t_acompanhamento
5  WHERE
6      ID_PONTO_DE_CONTROLE = 4169
7  ORDER BY
8      DT_CHEGADA

```

Fonte: Elaborado pelo autor.

O resultado desta consulta é uma tabela com todas as vezes que passou um ônibus no ponto de controle 4169. Um exemplo de resultado é exibido no Quadro 7.

Quadro 7 - Chegadas ao Ponto de Controle 4169 entre 12:00 e 13:00, 23/10/2018

Identificador	Ônibus	Ponto de Controle	Data de Chegada
15655	30717	4169	23/10/2018 12:01
15905	30203	4169	23/10/2018 12:04
15883	30708	4169	23/10/2018 12:10
15916	30707	4169	23/10/2018 12:34
15748	30715	4169	23/10/2018 12:40
15635	30716	4169	23/10/2018 12:51
15548	30701	4169	23/10/2018 12:55

Fonte: Elaborado pelo autor.

A partir deste quadro retornado pela consulta, pode-se calcular os *headways* pela simples subtração dos horários entre cada linha. Este método deve ser aplicado para de forma particionado para cada ponto de ônibus, obtendo assim todas as observações de *headway* do sistema no período avaliado.

5.3.3.1.5 Frota

A frota é descrita como a quantidade de ônibus que estão circulando numa determinada linha em dado período analisado. É importante no ponto de vista do planejamento, pois permite compreender quantos passageiros podem ser atendidos, e avaliar o consumo e outros indicadores importantes de eficiência.

A determinação da frota por meio dos dados contidos no banco é simples, bastando determinar quantos ônibus distintos passaram por algum ponto de ônibus (qualquer que seja este) no período selecionado. Por exemplo, para determinar a frota de ônibus de uma linha arbitrária numa faixa horária arbitrária, bastaria determinar quantos ônibus distintos passaram em algum ponto de ônibus da linha na faixa horária avaliada. A consulta que faz isso, em *SQL*, é exibida na Figura 12.

Figura 12 - Consulta para Determinação de Frota de Ônibus na Linha 75, no dia 23/10/2018, entre 10:00 e 11:00

```

1  SELECT
2      COUNT(DISTINCT ID_ONIBUS) `FROTA`
3  FROM
4      t_acompanhamento_real A
5      INNER JOIN t_ponto_de_controle pc
6          ON A.ID_PONTO_DE_CONTROLE = pc.ID_PONTO_DE_CONTROLE
7      INNER JOIN t_rota R
8          ON R.ID_ROTA = PC.ID_ROTA
9  WHERE
10     ID_LINHA = 75
11     AND DT_CHEGADA BETWEEN '2018-10-23 10:00:00' AND '2018-10-23 11:00:00'

```

Fonte: Elaborado pelo autor.

Em resumo, a consulta conta os ônibus distintos da tabela de acompanhamentos (que mapeia todas as passagens de ônibus por paradas), da linha 75, tal que a data de chegada destes ônibus na parada tenha sido entre 10:00 e 11:00 do dia 23/10/2018. O resultado da consulta é 23, que é a frota.

5.3.3.1.6 Bilhetagens Por Ponto de Ônibus

A demanda crítica não pode ser calculada diretamente por meio dos dados de *AFC* e *AVL*, visto que ela depende da quantidade de demandantes presentes no ônibus a cada trecho entre paradas numa faixa horária, enquanto os dados só englobam a bilhetagem.

O que é possível calcular, porém, é a bilhetagem estimada de passageiros por ponto de ônibus, por ônibus, e por faixa horária por meio da tabela “t_bilhetagem”. Para simplificar o processo, calcula-se por passagem do ônibus nas paradas, a quantidade de bilhetagens que ocorre entre o momento da passagem do ônibus na parada, e a passagem do mesmo ônibus na parada seguinte. Neste caso, assume-se que os passageiros que adentraram no veículo na primeira passagem, fizeram necessariamente a bilhetagem antes que o ônibus chegasse na parada posterior.

O procedimento de cálculo é:

- a) para cada registro da tabela “t_acompanhamento” (que guarda todas as vezes que um ônibus passa em uma parada), determina-se o horário que o ônibus referido no registro passa na parada posterior;

- b) portando o ônibus e o tempo entre a parada inicial e a parada posterior, calcula-se, para aquele ônibus, o volume de bilhetagens dentro daquele intervalo

Com isso, tem-se a bilhetagem estimada ao longo do tempo, das paradas e dos ônibus para cada aproximação do sistema.

5.3.3.2 Indicadores de Eficiência

Os indicadores de eficiência são de importância principalmente para o operador do sistema de transporte, e trata de eficiência de quilometragem, volume de passageiros e afins.

5.3.3.2.1 Distância Média entre Paradas

Para a determinação da distância média entre paradas, utiliza-se as informações do *shape* e de posição das paradas, disponibilizado nos dados de *GTFS*. Para cada ponto de ônibus, tem-se um ponto correspondente no arquivo de *shape* das linhas; dessa maneira, basta-se determinar a distância entre cada um dos pontos de *shape* intermediários entre paradas, utilizando a equação de Haversine.

Portando a distância entre todas as paradas de cada linha, basta-se calcular a média por linha, e será obtido o indicador.

5.3.3.2.2 Índice de Quilômetro por Veículo

Este indicador é descrito pela razão entre número total de quilômetros rodados e a frota total de veículos. Uma vez que se sabe, para cada ônibus, todos os horários que estes ônibus passaram por as paradas de suas linhas e, como dito na seção imediatamente anterior, sabe-se a distância entre cada parada por meio do arquivo de *shape*, pode-se inferir a distância percorrida por cada ônibus.

Conforme descrito, ainda, na seção 5.3.3.1.5, sabe-se o volume de veículos (frota) a qualquer momento a partir da contagem distinta de veículos que passaram por paradas. Portando então a frota e as distâncias percorridas por ônibus, o indicador pode ser calculado pela divisão dos dois termos, dividida pelo número de dias avaliados.

5.3.3.2.3 Índice de Passageiros por Quilômetro

De maneira semelhante, este indicador é descrito pela razão entre o total de passageiros atendidos, cujo cálculo é descrito na seção 5.3.3.1.6, e a quilometragem percorrida pelos veículos da frota analisada, que também pode ser calculada diretamente pelos dados de geolocalização e de passagens por paradas.

5.3.3.2.4 Índice de Passageiros por Veículo

Semelhantemente ao Índice de Quilômetro por Veículo, este indicador leva em conta o total de veículos da frota, porém o numerador da fração é o total de passageiros atendidos, ao invés da quilometragem percorrida. O total de passageiros está descrito na tabela de bilhetagens, e foi calculado anteriormente no trabalho.

O cálculo, portanto, presume a divisão direta entre o total de passageiros atendidos e o número de veículos da frota, dividido então pelo número de dias analisados.

5.3.3.2.5 Demais indicadores de eficiência

Os demais indicadores de eficiência descritos não são compatíveis com os dados de *AVL* e *AFC* contidos no banco de dados, pois envolvem informações não disponibilizadas nestes dados.

Os indicadores de custo, por exemplo, necessitam de informações de consumo dos veículos do sistema de transporte, tal como o custo administrativo das empresas operadores, como custo dos condutores, planejadores e afins. No caso do indicador de aproveitamento da frota, faz-se necessário, para o cálculo, a frota total de veículos das empresas operadoras.

Por esta incompatibilidade, estes indicadores não serão calculados neste trabalho.

5.3.3.3 Indicadores de Qualidade

Os indicadores de qualidade levam em conta, principalmente, a percepção do usuário frente ao serviço ofertado no sistema de transportes (FERRAZ; TORRES, 2001). Esta etapa busca expor o método de cálculo destes indicadores de qualidade selecionados.

5.3.3.3.1 Frequência de Atendimento

A determinação da frequência de atendimento é análoga à determinação do *headway* médio, descrito na seção 5.3.3.1.4, uma vez que a frequência é definida como o inverso do *headway*.

A formulação a seguir descreve a fórmula, e este será o método de cálculo adotado, adotando f como a frequência média dos ônibus para uma determinada parada p , e sendo h o *headway* médio para a mesma parada.

$$f(p) = \frac{1}{h(p)} \quad (6)$$

5.3.3.3.2 Tempo de Viagem Médio

Ferraz e Torres (2001) descrevem o tempo de viagem como o tempo médio gasto de um usuário desde o momento que este adentra no veículo, até o momento que este passageiro desembarca. Por este trabalho tratar de sistemas de *AFC*, não se sabe perfeitamente quando um passageiro adentra no ônibus, e não se sabe de forma alguma, diretamente, quando o passageiro desce do veículo.

Por este prisma, este indicador não será calculado neste trabalho, pois conforme descrito, exige dados a mais que não são contidos no atual sistema de coleta de dados automatizados de Fortaleza.

5.3.3.3.3 Lotação Média

De forma análoga ao Tempo de Viagem Médio, por não ser sabido o momento de descida dos usuários dos ônibus, não é possível determinar o volume de passageiros dentro do ônibus a cada instante. Pelo mesmo motivo, este indicador não será calculado neste trabalho.

5.3.3.3.4 Coeficiente de Variação do *Headway*

O coeficiente de variação do *headway* é um indicador que busca avaliar a confiabilidade do transporte público para o passageiro, no ponto de vista de variância da frequência dos ônibus.

É definido como a razão entre o desvio padrão do *headway* e sua média. Quanto maior o valor, maior a variação do *headway* e, portanto, menor a confiabilidade do sistema,

para o contexto analisado. A formulação matemática para o coeficiente de variação do *headway* é exibida a seguir.

$$CV_h = \frac{S_h^2}{\bar{h}} \quad (7)$$

Onde:

CV_h é o coeficiente de variação do *headway*;

S_h^2 é o desvio padrão observado do *headway*;

\bar{h} é a média observada do *headway*.

Para o cálculo do *headway* observado, utiliza-se o método descrito em 5.3.3.1.4, e então utiliza-se a formulação acima para calcular o coeficiente de variação para o sistema de transportes.

5.3.3.3.5 Índice do Tempo de *Buffer* Médio

O tempo de *buffer* médio descreve o tempo que o passageiro de um ônibus deve adicionar ao seu tempo de viagem esperado de forma a garantir que chegará, 95% das vezes, ao seu destino no horário previsto (HIGATANI *et al.*, 2009).

Para tal, deve-se determinar a distribuição do tempo de viagem do passageiro, e a partir da distribuição, calcula-se o tempo de viagem correspondente ao 95º percentil da distribuição. Portando a média e o 95º percentil, faz-se a subtração dos valores. O tempo excedente é o tempo extra que deve ser adicionado à viagem para garantir a pontualidade. O tempo de *buffer*, portanto, é maior quanto maior for a variabilidade do tempo de viagem.

O índice de tempo de *buffer*, por conseguinte, é a razão entre o tempo de *buffer* e o tempo de viagem, representando percentualmente quanto deve ser somado à viagem pelo passageiro para manter a confiabilidade das viagens. A formulação para o índice de tempo de *buffer* é mostrada a seguir (HIGATANI *et al.*, 2009):

$$\text{Índice de Tempo de Buffer} = \frac{\text{Tempo de Viagem}_{95^\circ \text{ percentil}} - \text{Tempo de Viagem}_{\text{médio}}}{\text{Tempo de Viagem}_{\text{médio}}} \quad (8)$$

Conforme discutido anteriormente, porém, não se tem nos dados de *AVL* e *AFC* as viagens dos passageiros, uma vez que o sistema carece da informação de subida e descida dos passageiros nos ônibus. Com isso, opta-se por adaptar a definição para avaliar o tempo de viagem total dos veículos. Dessa maneira, o índice de tempo de *buffer* levaria como tempos de viagem na formulação, os tempos de viagem de cada ônibus de uma respectiva linha e faixa horária. Assim, ele representaria uma versão generalista do tempo de *buffer*, assumindo que a

variação do tempo de viagem do ônibus é representativa da variação do tempo das viagens dos passageiros.

5.4 Interface de Consulta e Visualização de Dados

O presente subcapítulo apresentará o método de criação da interface de consulta aos dados do banco e análise dos indicadores calculados nas etapas anteriores, atendendo ao último objetivo específico do trabalho.

O intuito da criação da interface é que ela seja de acesso livre, fácil compreensão dos principais indicadores, permitindo filtragem por linha, sentido, ônibus, parada, data e faixa horária, de modo a possibilitar análise cruzada de diferentes cenários do sistema de transportes.

Além disso, é importante que a interface tenha código aberto, permitindo alterações aos dados e aos visuais, melhorando continuamente o acompanhamento do sistema por meios dos dados de *AVL* e *AFC*.

5.4.1 Determinação das Visualizações

Inicialmente, escolhe-se as visualizações a serem implementadas na interface. Nesta etapa, descreve-se que tipo de visualização será contida na interface, sendo gráficos, cartões informativos, mapas ou tabelas. Todas as páginas devem conter os filtros principais de análise cruzada dos dados: linha de ônibus, sentido da viagem, parada, ônibus, data e faixa horária.

5.4.1.1 Página de Indicadores da Operação

A página inicial é dedicada aos indicadores de operação, então deve constar visualizações claras que expliquem os resultados calculados referentes à operação, como velocidades comercial e operacional, tempos em viagem e de ciclo etc.

Para o resultado absoluto dos indicadores para o contexto de filtros selecionados, opta-se por exibir cartões informativos com o valor do indicador calculado e sua legenda, contendo um cartão para cada indicador operacional listado:

- a) tempo em viagem;
- b) velocidade operacional;
- c) tempo de ciclo;

- d) velocidade comercial;
- e) *headway* médio;
- f) bilhetagens;
- g) frota.

Como todos estes indicadores são calculados a partir das passagens de cada ônibus em cada parada modelada no trabalho, opta-se por exibir também uma tabela contendo informações de passagem dos ônibus nos pontos selecionados no contexto filtrado. Os campos da tabela serão:

- a) parada;
- b) ônibus;
- c) data;
- d) horário;
- e) *headway* referente ao ônibus seguinte naquela parada;
- f) bilhetes referentes àquele ônibus naquela parada;
- g) velocidade do ônibus ao chegar naquela parada.

Para determinar a velocidade do ônibus ao chegar na parada, utiliza-se a distância daquele ônibus à parada anterior e o tempo de trajeto. A velocidade então, pode ser traduzida na velocidade do trecho anterior.

Por fim, de modo a permitir a análise por parada dos indicadores operacionais selecionados, opta-se por implementar um mapa contendo todos os pontos de parada do contexto filtrado.

Cada parada de ônibus será representada por uma bolha, cuja coloração e tamanho varia de acordo com o indicador selecionado na visualização de mapa. Por exemplo, caso seja selecionado o indicador de *headway* médio, as paradas de ônibus do mapa serão maiores e mais azuladas em paradas com maior *headway* médio, e menores e alaranjadas quando a parada tem *headway* médio menor. Os indicadores passíveis de serem representados no mapa serão:

- a) velocidade na chegada (velocidade do trecho anterior);
- b) *headway* médio;
- c) bilhetagens;
- d) velocidade comercial;
- e) velocidade operacional.

5.4.1.2 *Página de Indicadores de Eficiência*

Mantendo o padrão da página inicial, a página referente aos indicadores de eficiência deve mostrar de forma clara os valores resultado dos indicadores de eficiência calculados para o contexto de filtragem selecionado na interface.

Os indicadores principais serão exibidos em cartões informativos, da mesma maneira que na página de indicadores de operação. Os cartões informativos mostrarão os seguintes indicadores e seus valores:

- a) bilhetagens;
- b) quilômetros rodados;
- c) frota;
- d) índice de passageiros por quilômetro;
- e) índice de passageiros por veículo;
- f) índice de quilômetros por veículo;
- g) distância média entre paradas.

Semelhantemente, será implementada uma tabela com as informações referentes aos indicadores de eficiência por ônibus. Dessa forma, pode-se avaliar e comparar o desempenho dos ônibus específicos referente aos indicadores calculados. Os campos da tabela serão:

- a) ônibus;
- b) bilhetagens;
- c) quilômetros rodados;
- d) índice de passageiros por quilômetro;
- e) velocidade operacional.

Para facilitar na percepção de demanda ao longo do tempo, será implementado nesta parte da interface um gráfico de barras exibindo o volume de bilhetagens que ocorreram para cada faixa horária. Desta forma, pode-se avaliar visualmente os horários com maior demanda de passageiros.

Em complemento ao gráfico de barras, cria-se um mapa contendo as paradas selecionadas, exibindo, para cada parada, o volume de bilhetagens. De maneira semelhante à página inicial, quanto maior a bolha de cada parada, maior o volume de bilhetagens. Caso necessário seja, opta-se por implementar um botão que alterna o mapa entre mapa e lista de paradas por bilhetagens. Assim, caso o utilizador queira ver as paradas com maior bilhetagem, é possível avaliar de forma ordenada.

5.4.1.3 Página de Avaliação do Headway

Esta página tem como ideia avaliar o comportamento do *headway* por faixa horária, averiguando ainda sua distribuição.

Para avaliação do *headway* por faixa horária, deve ser implementado um gráfico de barras com o *headway* médio por faixa horária selecionada. Dessa maneira, será possível ver que horários tem maior tempo de espera entre ônibus, possibilitando, ainda, avaliar se os *headways* estão conforme o planejado.

Para averiguar a distribuição do *headway*, cria-se um histograma, desta vez mostrando o volume de *headways* observados em cada faixa de *headway*, separadas de 5 em 5 minutos. Com isso, pode-se avaliar visualmente a forma da distribuição dos valores. Caso seja requerido, o usuário tem a opção de clicar numa faixa horária específica do gráfico anterior, assim fazendo o filtro cruzado na distribuição, mostrando a distribuição de *headways* para aquela faixa horária.

Alguns cartões informativos devem ser exibidos nesta página, com o intuito de mostrar informações pertinentes referentes ao *headway*. Os cartões informativos a serem implementados apresentam as seguintes informações:

- a) *headway* médio;
- b) desvio padrão do *headway*;
- c) coeficiente de variação do *headway*;
- d) tempo de espera.

O tempo de espera é calculado a partir da equação a seguir (MISHALANI; MCCORD, 2006):

$$T_e = \frac{\bar{h}}{2} \cdot (1 + CV_h^2) \quad (9)$$

Onde:

T_e é o tempo de espera estimado;

\bar{h} é o *headway* médio observado;

CV_h é o coeficiente de variação do *headway*.

Uma vez calculado o coeficiente de variação, opta-se por exibir também um gráfico de barras cujo eixo vertical é o coeficiente de variação e o eixo horizontal é a faixa horária.

Dessa forma, pode-se visualizar os horários com maior variação frente à média.

Por fim, implementa-se dois gráficos exibindo as paradas com maior tempo de espera estimado e com maior coeficiente de variação do *headway*, possibilitando a averiguação das paradas mais críticas no ponto de vista de variabilidade.

5.4.1.4 Página de Tempos e Velocidades

A página final da interface tem como intuito mostrar indicadores e visualizações referentes aos tempos de ciclo e viagem, tal como velocidades comerciais e operacionais do sistema.

Inicialmente, implementa-se um gráfico de barras agrupadas por faixa horária, mostrando tempos de ciclo e tempos de viagem. Dessa forma, pode-se avaliar ambos os indicadores por faixa horária, e ainda avaliar a diferença entre os dois indicadores. Idealiza-se que o usuário conseguirá ver os horários com viagens mais longas e, além disso, avaliar nesses horários, quanto tempo, em média, os ônibus ficam parados nos terminais nestes horários.

De forma análoga, cria-se um gráfico de barras agrupadas por faixa horária, porém desta vez avaliando as velocidades comerciais e operacionais. Da mesma forma como no caso dos tempos de viagem e de ciclo, pode-se avaliar os horários que permitem tráfego dos ônibus com maiores velocidades.

Como todas as páginas, implementa-se cartões informativos com os quatro indicadores descritos até agora:

- a) tempo de ciclo;
- b) tempo em viagem;
- c) velocidade comercial;
- d) velocidade operacional.

É nesta página que se implementa o indicador de tempo de *buffer* calculado na seção 5.3.3.3.5. Serão criados mais cartões informativos, desta vez exibindo:

- a) média do tempo de ciclo;
- b) mediana do tempo de ciclo;
- c) 95º percentil do tempo de ciclo;
- d) índice de tempo de *buffer* estimado.

O valor do índice de tempo de *buffer* estimado será exposto em outro gráfico de barras, mostrando seu valor para cada faixa horária selecionada. Assim, pode-se ver horários em que as viagens têm maior tempo extra necessário em relação à média.

5.4.2 Criação da Interface

A interface proposta será criada no *software Power BI*, por permitir a conexão com bancos de dados *MySQL*, ser gratuito e possibilitar a hospedagem da interface em uma página da *web* de forma nativa.

A criação é dividida em algumas etapas principais, sendo elas:

- a) subida das consultas ao banco para o relatório do *Power BI Desktop*;
- b) criação das medidas;
- c) criação das páginas e visuais;
- d) publicação do relatório no serviço *online* do *Power BI*.

Cada etapa será brevemente descrita a seguir.

5.4.2.1 Consultas ao Banco de Dados

Para a criação das páginas descritas na seção 5.4.1, é necessária a carga dos dados para um relatório do *Power BI Desktop*. A carga é feita por meio da interface do *Power Query*, que tem diversos conectores de dados disponíveis; no caso deste trabalho, utiliza-se o conector de bancos de dados *MySQL*.

Para a modelagem de dados dentro do *software*, separa-se, semelhantemente ao banco de dados, as tabelas em dimensões e fatos. No caso das tabelas dimensão, a consulta é meramente uma seleção de todas as linhas de todas as colunas de cada dimensão, uma vez que o modelo já é normalizado.

As dimensões do banco a serem carregadas no modelo de dados do *Power BI* são:

- a) tabela de ônibus;
- b) tabela de ponto de controle;
- c) tabela de parada de ônibus;
- d) tabela de rota;
- e) tabela de linha;
- f) tabela de sentido;
- g) tabela de faixa horária.

Essas dimensões serão utilizadas para filtragem das informações contidas nas tabelas de fato. As configurações relacionais entre as dimensões são idênticas às do banco de dados, com algumas pequenas mudanças para facilitar a utilização da interface. No caso, a alteração ocorre na relação entre pontos de controle e paradas de ônibus; opta-se por criar uma

relação bilateral, ao invés de uma relação unilateral onde a parada filtra o ponto de controle. Isso permite que a tabela de pontos de controle acesse informações da tabela de paradas, nomeadamente o par de latitude e longitude.

No caso das tabelas de fato, utiliza-se a tabela inteira de acompanhamento (que mostra quando um ônibus passa por cada parada), sendo também uma seleção simples de todas as linhas e todas as colunas da tabela. A tabela seria o suficiente para o cálculo de todos os indicadores selecionados e apresentados no método, porém o poder de processamento de dados necessário no *Power BI* é superior ao poder de processamento necessário para o cálculo no banco de dados *MySQL*.

Por este motivo, os indicadores mais onerosos, que devem ser calculados para cada registro da tabela de acompanhamento, serão operados por meio de consultas no *MySQL*, que serão subidas ao *Power BI* por meio do conector. Os indicadores a serem calculados no servidor ao invés de no *Power BI* são:

- a) tempo de ciclo;
- b) tempo de viagem;
- c) *headway*;
- d) velocidade no trecho;
- e) bilhetagens.

Para cada um destes indicadores, é feita uma consulta ao banco de dados que traz para cada registro de acompanhamento, o valor da observação daquele indicador. Como as consultas não são o principal ponto do trabalho, mas somente um meio para o fim, que é o cálculo dos indicadores, estas não serão apresentadas nessa seção do texto. A quem interessar, as consultas estão no APÊNDICE A – CONSULTAS DE CÁLCULOS DOS INDICADORES.

Uma vez inseridas as consultas no *Power BI*, relaciona-se os identificadores de acompanhamento para cada uma das tabelas de indicadores à tabela real de acompanhamento, adotando uma relação bidirecional um para um, permitindo que a tabela de fato tenha acesso aos valores dos indicadores e vice-versa.

5.4.2.2 Criação das Medidas

Tendo no modelo de dados do *Power BI* todas as tabelas necessárias para o cálculo, deve-se criar as medidas que calculam os indicadores por meio da linguagem *DAX*. A linguagem *DAX* é uma linguagem semelhante à utilizada no *Excel*, porém com arquitetura voltada ao cálculo em tabelas.

As medidas necessárias a serem calculadas são todas aquelas referidas na descrição dos visuais, na seção 5.4.1. O método de cálculo já foi descrito na seção, porém a criação das medidas exige que o método seja transcrito em *DAX*, obedecendo sua arquitetura e sintaxe.

Para o caso das medidas descritas na seção anterior, onde o cálculo foi feito dentro do servidor do *MySQL*, a criação das medidas é simples e exige somente a agregação dos valores por meio de alguma função do *DAX*, como por exemplo *AVERAGE*.

O cálculo do tempo de ciclo e tempo de viagem médios, por exemplo, pode ser resumido em uma medida que calcula o *AVERAGE* da coluna correspondente ao tempo de ciclo ou tempo de viagem respectivamente. A Figura 13 exemplifica a medida que calcula o tempo em viagem médio a partir das tabelas do modelo.

Figura 13 - Medida em DAX Referente ao Cálculo do Tempo em Viagem Médio

```
1 Tempo em Viagem Médio = AVERAGE(fTempoViagem[TEMPO_VIAGEM])
```

Fonte: Elaborado pelo autor.

O mesmo deve ser feito para as demais medidas que foram calculadas no próprio banco de dados.

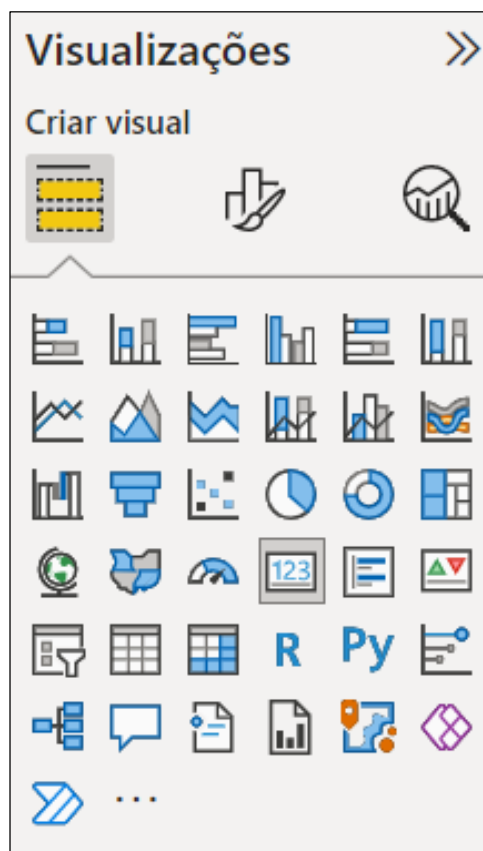
No caso das medidas que serão calculadas dentro do próprio *Power BI*, estas devem ser formuladas traduzindo as expressões matemáticas em *DAX*. O procedimento é simples, uma vez que todos os valores já estão tabelados.

5.4.2.3 Criação das Páginas e Visuais

Esta etapa consiste em criar os visuais descritos na seção 5.4.1 a partir das medidas criadas a partir do método descrito na seção 5.4.2.2.

O *Power BI* oferece uma interface de criação de visuais que será utilizada extensivamente nesta etapa. A Figura 14 mostra a interface de criação de visuais.

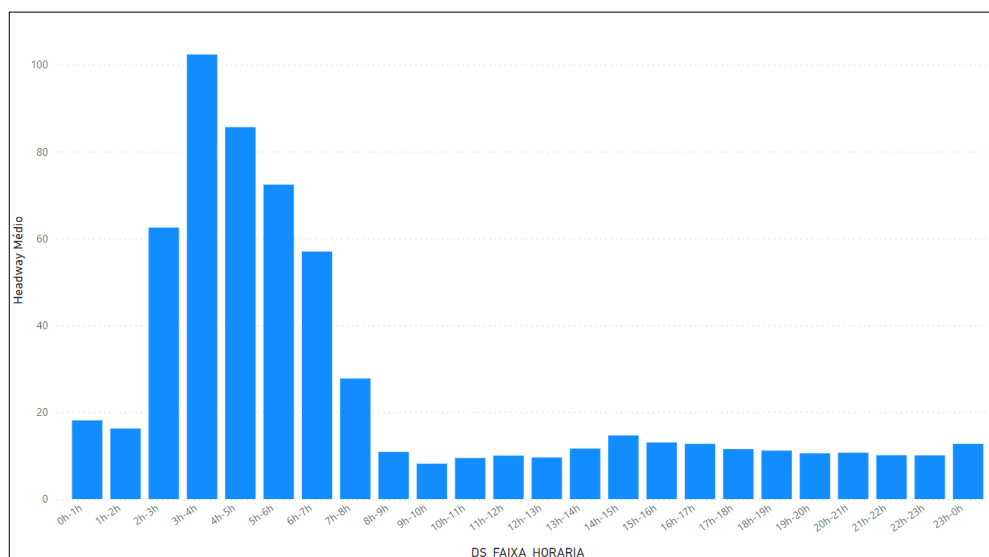
Figura 14 - Interface de Criação de Visuais do Power BI



Fonte: Elaborado pelo autor.

Nesta interface de seleção, o usuário deve selecionar que tipo de gráfico deve ser criado e, posteriormente, alimentar o visual com os dados a serem exibidos. No caso da criação dos visuais de gráficos de barra mostrando *headway* médio por faixa horária, por exemplo, deve-se selecionar o segundo visual da primeira linha, e alimentar o eixo horizontal com as faixas horárias e o eixo vertical com a medida criada referente ao *headway* médio. A Figura 15 mostra o resultado da ação descrita.

Figura 15 - Gráfico de Barras de Headway Médio por Faixa Horária, sem tratamento visual



Fonte: Elaborado pelo autor.

Como pode-se ver, o visual é formado, porém com os nomes brutos nos eixos, e não necessariamente ordenados. Por padrão, os valores nos eixos e os rótulos dos dados têm fontes pequenas e de difícil leitura, o que também não é ideal para a usabilidade. Além disso, o visual carece do contexto de filtros, pois não fora criado ainda. Por conta disso, faz-se necessário o tratamento do visual tanto no ponto de vista estético quanto no ponto de vista de clareza das informações; isso engloba renomear os eixos, mudar as fontes e seus tamanhos, dar títulos, criar rótulos, ordens e configurar a filtragem cruzada.

Além da criação dos visuais, deve-se criar a opção de filtragem para o usuário, para permitir a análise cruzada das informações. Os filtros visuais, chamados no *Power BI* de segmentação de dados, são como visuais comuns: deve-se selecioná-los dentre os visuais disponíveis, e alimentar o visual com os dados de dimensão que filtrarão os resultados.

O método desta etapa pode ser resumido, então, na criação dos diversos filtros e visuais descritos nas etapas metodológicas anteriores, separados em páginas e visualmente tratados. Ao fim desta etapa, espera-se que o relatório da interface já esteja pronto. Nesse caso, deve-se salvar o arquivo do relatório no formato padrão do *Power BI*: *pbix*.

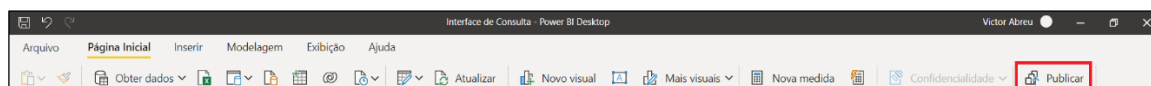
5.4.3 Implementação da Ferramenta Online

Uma vez finalizado o relatório, que será a interface de consulta aos dados, este deve ser implementado na *web* para acesso público. Para isto, o *Power BI* disponibiliza, gratuitamente, para as contas estudantis e corporativas, o serviço do *Power BI Online*. No caso

deste trabalho, é utilizado o e-mail estudantil do aluno (cuja extensão é @alu.ufc.br), e obtém-se acesso à área de trabalho (*workspace*) online.

O *workspace* pessoal vem naturalmente vazio, porém é possível publicar os relatórios criados no *Power BI Desktop* para o *Power BI Online* por meio da opção “Publicar”, mostrada na Figura 16.

Figura 16 - Opção Publicar do Power BI Desktop



Fonte: Elaborado pelo autor.

A partir desta opção, publica-se o relatório no *workspace* privado do *Power BI Online*. A partir deste momento, tem-se algumas opções a serem consideradas:

- a) atualização agendada;
- b) criação do *link* público.

A atualização agendada permite que o relatório da interface tenha seus dados constantemente atualizados, permitindo que a interface continue funcionando uma vez que sejam adicionados mais dados ao banco de dados. Neste momento do trabalho, ela não será configurada, uma vez que os dados ainda estão limitados ao escopo definido anteriormente.

No caso da criação do *link* público, isso é feito pelo proprietário do relatório, na seção de visualização. Para tal, deve-se seguir à opção “Arquivo”, seguido de “Inserir relatório” e seguir as instruções da janela “Publicar na Web (público)”, conforme visto na Figura 17.

Figura 17 - Configuração para Publicar Relatório do Power BI na Web por meio de Link Público



Fonte: Elaborado pelo autor.

Uma vez feito isso, é gerando um *link* de acesso ao relatório, que pode ser compartilhado livremente. Qualquer usuário que utilize o *link* poderá utilizar a interface de consulta dos dados.

Finalizando esta etapa, conclui-se a criação da interface *online* de visualização dos dados do sistema de transporte de ônibus de Fortaleza.

6 RESULTADOS E DISCUSSÕES

A presente seção do trabalho tem como fim a apresentação dos resultados obtidos a partir da execução das etapas metodológicas descritas anteriormente, tal como apresentar a discussão destes resultados.

6.1 Criação da Base de Dados

Referente ao primeiro objetivo específico do trabalho, esta etapa demonstra o produto do método descrito na seção 5.1, envolvendo desde a normalização das tabelas no ponto de vista relacional, até a criação e população do banco de dados em *MySQL*.

6.1.1 Teoria da Normalização e Idealização do Formato da Base de Dados

Partindo-se da teoria da normalização, sabe-se que as tabelas fatos principais que serão criadas no banco de dados são as tabelas de posição e bilhetagens. Com isso, conforme proposto no método, deve-se reduzir as tabelas de fato ao seu menor formato normal, levando qualquer informação descritiva para outras tabelas, relacionando as tabelas por meio das chaves primárias e secundárias.

A Figura 18 apresenta uma amostra dos dados brutos de *GPS*, exibindo seu formato, ao qual será aplicado a normalização.

Figura 18 - Amostra dos Dados Brutos de GPS

	A^B_C hora	A^B_C lon	A^B_C lat	A^B_C vehicleId
1	2018-10-22 23:59:57	-38.606093	-3.7192910000000006	39511
2	2018-10-22 23:59:38	-38.502733	-3.8031410000000005	32841
3	2018-10-22 23:59:33	-38.545281	-3.744476	32068
4	2018-10-22 23:59:13	-38.606095	-3.7192879999999997	39511
5	2018-10-22 23:59:08	-38.502788	-3.803146	32841

Fonte: Elaborado pelo autor.

No ponto de vista de normalização, a tabela já está no seu menor formato, uma vez que não há nenhuma informação adicional aos dados da posição. Ressalta-se, porém, que na análise exploratória dos dados, foi percebido que o campo *vehicleId* não se refere ao número do ônibus adotado nas demais tabelas disponibilizadas.

Seguindo a normalização, deve-se então manter um indicador único para cada objeto único, então o indicador do ônibus foi corrigido adotando os valores descritos nas demais

tabelas; essa correção é possível pois existe paridade 1 para 1 entre os *vehicleId* descritos nesta tabela e o *prefixo_onibus* apresentado nas tabelas referentes à bilhetagem.

Por fim, esta tabela carece de um identificador único para cada registro de posição, que se comportaria como a chave primária da tabela. Portanto, obedecendo a normalização, deve-se criar um campo índice, incrementando seu valor de um em um para cada registro da tabela.

Apesar da tabela de posições bruta já estar no seu formato mínimo, esta carece de informações importantes para a análise dos dados, como por exemplo, a linha que o ônibus opera e o sentido. Para manter a normalização, mas conter ainda assim os dados necessários, cria-se uma tabela dimensão que agrupe as informações de ônibus, linha e sentido em um único indicador, sendo a primeira tabela dimensão do sistema. Para a construção dessa tabela, utiliza-se os dados de bilhetagem. A Figura 19 mostra o formato dos dados brutos de bilhetagem.

Figura 19 - Amostra dos Dados Brutos de Bilhetagem

id	hora	linha	prefixo_carro	vehicleid	sentido_viagem	tipo_cartao	integracao	dia	momento
7508885	22/10/2018 14:35:39	77	30213	33055	Volta	Gratuidade	N	segunda-feira, 22 de outubro de 2018	17:35:39
7508885	22/10/2018 11:36:01	31	35838	33590	Ida	Gratuidade	N	segunda-feira, 22 de outubro de 2018	14:36:01
7508885	22/10/2018 14:54:02	913	30034	32868	Volta	Gratuidade	N	segunda-feira, 22 de outubro de 2018	17:54:02
7508871	22/10/2018 10:45:32	601	2404	32844	Ida	Gratuidade	N	segunda-feira, 22 de outubro de 2018	13:45:32
7508870	22/10/2018 10:46:41	45	42120	32344	Ida	Gratuidade	N	segunda-feira, 22 de outubro de 2018	13:46:41

Fonte: Elaborado pelo autor.

Como pode-se ver, o arquivo contém informações do *prefixo_carro* e *vehicleId*, conforme dito, que serão usadas para determinar o número do ônibus em cada posição. Além disso, as informações de linha e sentido serão utilizadas para a determinação da linha que o ônibus está percorrendo a cada instante.

Define-se então, que cada combinação de linha e sentido será uma chave única, denominada rota. Com isso, cria-se uma tabela de dimensão de rota, contendo a chave única, o número da linha (chamado *id_linha*) e o sentido da linha em forma binária (0 sendo ida e 1 sendo volta), ocupando menos espaço no servidor. Esse processo já é parte da normalização da tabela de bilhetagem, onde serão separados da tabela de bilhetagem as informações de linha e sentido, em prol de um único identificador (*id_rota*) que carrega por si só ambas as informações.

Como a tabela de bilhetagem conta ainda com os dois indicadores do ônibus, opta-se por criar outra tabela, referindo-se à combinação dos ônibus com as rotas definidas no parágrafo anterior. Ou seja, cada combinação de ônibus e rota (que por sua vez carrega a informação da linha e do sentido) será resumida a um único identificador, chamado de

identificador de tabela (*id_tabela*). Logo, uma tabela deve ser criada, carregando informações do ônibus e rota.

As demais informações da tabela de bilhetagem são referentes ao registro único de bilhetagem, porém ainda há redundância, uma vez que existe um campo de hora mostrando a data e hora do registro, um campo de dia mostrando somente a data e um campo de momento mostrando somente a hora. Opta-se por excluir os campos separados (dia e momento) a fim de economizar espaço no servidor. Os demais campos podem também ser reduzidos a informações menores, como o tipo de bilhetagem e integração. No caso do tipo de bilhetagem, mapeia-se os valores gratuidade, estudante e vale transporte para 0, 1 e 2 respectivamente, transformando em inteiro. No caso da integração, utiliza-se binários, onde integração é identificada por 1, e não integração por 0.

Vale ressaltar que os dados de latitude e longitude são estimados a partir das informações de *AVL*, e, portanto, foram descartados no momento. A tabela de bilhetagens normalizada consiste, portanto, nos campos:

- a) *ID_BILHETAGEM*, se referindo ao identificador único de cada bilhetagem no sistema;
- b) *ID_USUARIO*, que é identificador único do cartão do passageiro;
- c) *DT_BILHETAGEM*, que é a data e hora da bilhetagem;
- d) *ID_TABELA*, que é o identificador que armazena as informações de ônibus, sentido e linha;
- e) *DS_BILHETAGEM*, que descreve a bilhetagem daquele registro (gratuidade, estudante, vale transporte);
- f) *FL_INTEGRACAO*, um valor binário que afirma se houve ou não integração.

Por definição da normalização, o *ID_TABELA* deve ter sua própria tabela, e esta terá os campos:

- a) *ID_TABELA*, o identificador único;
- b) *ID_ONIBUS*, identificando o ônibus por meio do seu prefixo (nome impresso nos veículos);
- c) *ID_ROTA*, identificador único que representa a linha e o sentido naquele instante.

Novamente, por definição da normalização, é seguido para o terceiro grau da tabela, dessa vez referente ao *ID_ROTA*. A tabela deste identificador terá os campos:

- a) *ID_ROTA*, o identificador único;
- b) *ID_LINHA*, o identificador da linha, representado por um número inteiro;

- c) FL_SENTIDO, um valor binário que determina o sentido da viagem (0 é ida, 1 é volta).

Para o quarto grau da tabela de bilhetagens, têm-se ainda a tabela de linhas. Nesse caso, a tabela meramente determina o nome da linha para cada identificador. Esses dados estão presentes no arquivo de *GTFS* nomeado “*routes.txt*”. A Figura 20 mostra a estrutura deste arquivo.

Figura 20 - Amostra do Arquivo de Linhas do GTFS

	1 ² ₃ route_id	1 ² ₃ agency_id	1 ² ₃ route_short_name	A ^B _C route_long_name
1	4	1	4	Messejana/Papicu/Cambéba/TJ
2	11	1	11	Circular I
3	12	1	12	Circular II
4	13	1	13	Aguanambi I
5	14	1	14	Aguanambi II

Fonte: Elaborado pelo autor.

Os campos relevantes do arquivo são o *route_id* e *route_long_name*, que se referem justamente ao número da linha e seu nome. Para a criação da tabela de linhas, portanto, serão removidos os demais campos. As colunas da tabela de linhas serão:

- ID_LINHA, identificador único de cada linha;
- DS_LINHA, campo que descreve a linha. Nesse caso, exibe o nome completo da linha.

Uma vez tendo essas tabelas, pode-se determinar que os campos necessários para manter a forma mínima necessária da tabela de posições dos ônibus serão:

- ID_POSICAO, o identificador único da posição;
- ID_TABELA, identificando o ônibus, sua linha e sentido;
- LAT, identificando sua latitude;
- LONG, identificando sua longitude;
- DT_REGISTRO, identificando o momento daquela posição.

Dessa forma, já é tido todas as tabelas normalizadas referentes à posição e bilhetagens dos ônibus. Sabe-se, entretanto, que o banco de dados necessitará das informações de paradas de ônibus. Esses dados constam no *GTFS*, porém também precisam ser normalizados.

As paradas de ônibus de um sistema podem atender a diversas linhas diferentes, e, por conta disso, será feita uma distinção de parada e ponto de controle. Um ponto de controle se refere a uma parada e uma rota (que por sua vez se refere a uma linha e um sentido). Dessa forma, mesmo que uma parada atenda a mais de uma linha, um ponto de controle

necessariamente estará relacionado a uma linha específica; com isso, caso seja filtrada uma linha específica no banco de dados, só serão retornadas as paradas que pertencem àquela linha.

Os campos da tabela de pontos de controle são:

- a) *ID_PONTO_DE_CONTROLE*, sendo o identificador único;
- b) *ID_ROTA*, identificando qual linha e sentido é atendido por este ponto;
- c) *ID_PARADA*, identificando a parada em si;
- d) *ID_ORDEM*, exibindo a ordem sequencial da parada para aquela rota. Por exemplo, a primeira parada de uma rota específica tem valor 1, a segunda 2, e assim sucessivamente.

O *ID_ROTA* já foi contemplado com sua própria tabela anteriormente, porém necessita-se criar uma tabela referente a cada parada, identificada pelo *ID_PARADA*. A tabela de paradas terá campos:

- a) *ID_PARADA*, sendo o identificador único;
- b) *DS_PARADA*, sendo a descrição da parada, geralmente seu endereço;
- c) *LAT*, sendo a latitude;
- d) *LONG*, sendo a longitude.

Com isso, estima-se ter idealizados todas as tabelas por meio da teoria da normalização. Para finalização, deve-se criar as tabelas idealizadas no banco de dados, conforme mostrado na Figura 1. Todas as consultas de criação das tabelas estão exibidas no APÊNDICE B – CONSULTAS DE CRIAÇÃO DE TABELAS.

Finalmente, mostra-se na Figura 21 a esquematização das tabelas e suas relações no servidor do *MySQL*, já com as tabelas criadas.

- a) dentro do intervalo de tempo entre duas bilhetagens sequenciais no mesmo ônibus, linha e sentido, qualquer posição daquele ônibus detectada neste intervalo temporal necessariamente está atendendo àquela linha e sentido;
- b) dentro do intervalo de tempo entre duas bilhetagens sequenciais com o mesmo ônibus e linha, porém sentidos diferentes, adota-se que as posições dentro daquele intervalo temporal atendem ao sentido mais próximo daquele horário. Presume-se, nesse cenário, que a troca de sentido ocorreu entre as duas bilhetagens, e as posições são separadas por sentido baseando-se no sentido mais próximo detectado;
- c) para posições antes ou depois de qualquer bilhetagem, adota-se que o ônibus está atendendo à linha e sentido mais próximos.

Isso é implementado no código por meio da função *merge_asof*, da biblioteca *pandas*. A rotina completa está contida no APÊNDICE C – ROTINA DE TRATAMENTO DOS DADOS.

A população das tabelas é feita conforme descrito no método, utilizando a função *LOAD DATA*. Como resultado, mostra-se no Quadro 8 o volume de linhas em cada tabela da base de dados.

Quadro 8 - Total de Registros por Tabela do Banco de Dados

Tabela	Registros
t_bilhetagem	4.828.856
t_linha	301
t_onibus	1.836
t_parada	4.543
t_posicao	20.032.855
t_rota	588
t_tabela	7.446

Fonte: Elaborado pelo autor.

6.2 Modelagem dos Horários dos Ônibus

Esta etapa busca demonstrar os resultados obtidos a partir do método descrito na seção 5.2, referente à modelagem dos horários que os ônibus passam pelas paradas no sistema de transporte analisado.

Uma limitação encontrada nesta etapa do trabalho é advinda do poder e tempo de processamento necessário para a execução das rotinas propostas para geração dos registros dos horários. Pelo fato do servidor que armazena o banco de dados estar, até o momento, hospedado num computador pessoal, existe limitação do uso de disco e de memória computacional para a execução das atividades. Por conta disso, o escopo da modelagem foi reduzido à linha 75 – Campus do Pici / Unifor, por ser uma rota com importância no ponto de vista da universidade, sendo ainda uma linha com duplo sentido e grande demanda.

Cada execução de determinação de todas as distâncias entre as posições de um ônibus específico para uma parada específica nas datas analisadas, seguido do tratamento e estimativa da passagem nas paradas, demora, em média, 120 segundos. Considerando que a linha escolhida teve um total de 34 ônibus rodando nesse período e conta com 95 pontos de ônibus, estima-se que o tempo total gasto no processamento dos dados foi superior a 100 horas. Esse tempo pode ser um fator limitante, porém um servidor dedicado com mais disponibilidade de disco e memória deve diminuir consideravelmente o tempo de processamento.

Dito isso, o método descrito pode ser inviável se levar em conta que o sistema de ônibus de Fortaleza conta, hoje, com mais de 300 linhas diferentes; com isso, mesmo com maior poder de processamento, pode ocorrer de o gasto computacional inviabilizar o processamento.

Uma possível hipótese que explicaria o gasto computacional excessivo é a ausência de indexação das tabelas do banco de dados; é possível que a inserção da indexação nos campos chave melhore o desempenho das consultas.

Outra hipótese é que a implementação do cálculo das distâncias por meio da equação de Haversine pode utilizar muita capacidade computacional, isso porque envolve diversas funções trigonométricas, que tem custo de avaliação maior que funções lineares e afins, além de que o cálculo deve ser feito para cada posição de cada ônibus contra cada parada daquela linha específica. A equação pode ser substituída por simplificações, porém sacrificando a precisão; dito isso, como as distâncias calculadas são em pontos próximos (dentro da mesma cidade), o impacto da simplificação não necessariamente implicaria em diferenças significativas na precisão (KARNEY, 2013).

Ademais, o método foi aplicado no escopo reduzido à linha 75, e assim foi criada e alimentada a tabela “t_acompanhamento” no banco de dados, onde cada registro se refere a uma passagem de ônibus por ponto de controle. O total de registros de acompanhamento foram de 46.453 linhas. A Figura 22 apresenta uma pequena amostra da tabela, retirada do *MySQL*.

Figura 22 - Amostra dos 5 Primeiros Registros da Tabela de Acompanhamento

	ID_ACOMPANHAMENTO	ID_TABELA	ID_PONTO_DE_CONTROLE	DT_CHEGADA
▶	1057	978	4131	2018-10-22 17:15:53
	1058	977	4131	2018-10-22 13:43:22
	1059	977	4131	2018-10-22 20:26:11
	1060	1869	4131	2018-10-22 11:24:14
	1061	1870	4131	2018-10-22 19:20:48

Fonte: Elaborado pelo autor.

6.3 Determinação dos Indicadores

Nesta etapa, busca-se determinar os indicadores que serão calculados no trabalho, e calculá-los, sempre separados em indicadores de operação, eficiência e qualidade.

Como descrito no método, separa-se os indicadores listados em três grupos distintos, a depender da forma de cálculo destes indicadores. O Quadro 9 exibe a listagem de indicadores de cada grupo para cada classe.

Quadro 9 - Distribuição dos Indicadores por Grupo de Cálculo e Classe

Grupo / Classe	Operação	Eficiência	Qualidade
Grupo 1	Tempo de ciclo, tempo em viagem, velocidade operacional, velocidade comercial, <i>headway</i> médio, frota	Distância média entre paradas, índice de quilômetros por veículo, índice de passageiros por quilômetro, índice de passageiros por veículo	Frequência de atendimento, coeficiente de variação do <i>headway</i>
Grupo 2	Embarques		Índice de tempo de <i>buffer</i> médio
Grupo 3		Índice de aproveitamento da frota, custo por passageiro, razão da receita arrecadada pelo custo	Lotação média, tempo médio em viagem

Fonte: Elaborado pelo autor.

Conforme descrito, os indicadores do grupo 3 serão descartados do cálculo, pois não há dados os suficientes que permitam seus levantamentos. Para os indicadores do grupo 2, deve-se avaliar hipóteses simplificadoras que permitam seu cálculo.

No caso da demanda crítica, sabe-se que é necessário determinar o volume de passageiros presentes nos ônibus a cada trecho entre paradas em uma determinada faixa horária. Como os dados de *AFC* carecem da informação de volumetria de passageiros dentro do ônibus, opta-se por calcular, ao invés, para cada faixa horária e ponto de controle da linha, o volume de bilhetagens. Por conta desta limitação, opta-se por chamar o indicador de embarques, uma vez que melhor descreve o que está sendo calculado. O método de cálculo é descrito na seção 5.3.3.1.6.

A bilhetagem por ponto de controle não equivale à demanda crítica, porém é válido mencionar que é um passo mais próximo da determinação deste indicador. Uma possível hipótese que permite o cálculo da demanda crítica utilizando a bilhetagem por ponto de controle, é a execução de pesquisas de desembarques dos ônibus por parada; dessa maneira, os custos de levantamento da demanda crítica seriam reduzidos, já que não seria necessário acompanhar os embarques.

Alternativamente, pode-se inferir a demanda crítica utilizando os dados de bilhetagem crítica a partir do levantamento do Fator de Renovação numa determinada faixa horária; isso porque já se tem o total de passageiros atendidos. A equação a seguir exhibe a formulação da demanda crítica ($D_{m\acute{a}x}$) a partir da bilhetagem total (B) e do Fator de Renovação (FS).

$$D_{m\acute{a}x} = \frac{B}{FS} \quad (10)$$

No caso do índice de tempo de *buffer*, o indicador a priori não pode ser calculado diretamente pelo mesmo motivo: não se tem informações de desembarque dos passageiros. Dessa maneira, não é possível criar uma distribuição do tempo de viagem real dos passageiros.

Uma possível estratégia de cálculo é a inferência, a partir dos dados de identificação dos cartões únicos de passagem, das viagens individuais, ou, em outras palavras, dos desembarques a partir das demais utilizações dos cartões. Por exemplo, um passageiro fez duas viagens em um dia, do ponto A ao ponto B, e de ponto B ao ponto C; nesse caso, o passageiro teria dois registros de passagem, uma no ponto A e outro no ponto B. Assume-se, então, que a primeira viagem teve desembarque na parada mais próxima à B, assim mapeando a viagem. A partir disso, pode-se estimar a distribuição dos tempos das viagens (HANFT *et al.*, 2016). Essa abordagem, porém, está fora do escopo do atual trabalho, uma vez que exige robusta modelagem de viagens individuais.

De forma simplificada, então, opta-se por assumir que a distribuição dos tempos de ciclo dos ônibus na linha tem variação congruente à distribuição dos tempos de viagens

individuais. Dessa forma, pode-se calcular a distribuição dos tempos de ciclo, suas variações e por conseguinte, seu tempo de *buffer*, e, assumindo que a premissa é consistente, o índice de tempo de *buffer* seria equivalente. O tempo de *buffer* absoluto, porém, não seria equivalente, afinal dificilmente uma viagem de um passageiro é o ciclo inteiro da linha. Essa abordagem está descrita na seção 5.3.3.3.5 do trabalho.

Os demais indicadores, pertencentes ao grupo 1, podem ser calculados diretamente a partir dos dados do banco, e assim serão feitos. O procedimento dos cálculos foi descrito na seção 5.3.3, e foi aplicado aqui.

Uma consideração relevante referente ao cálculo dos indicadores do grupo 3 é que apesar de não ser possível calculá-los neste trabalho, suas formulações não seriam difíceis se houvesse mapeamento das viagens individuais ou detecção de desembarque no sistema. Um sistema de *APC* (contagem automática de passageiros) resolveria o problema dos indicadores de lotação e tempo médio em viagem, e a literatura também já demonstra a utilização de estimativas de matrizes origem-destino para este fim (GUTIÉRREZ *et al.*, 2020).

No caso dos indicadores de aproveitamento da frota, custo por passageiro e lucro, faz-se necessário ter no banco de dados informações referentes à frota total dos ônibus do sistema por empresa, além de informações dimensionais dos ônibus, como sua capacidade e eficiência do combustível por volume de passageiros. Portando estes dados, o cálculo sairia diretamente, uma vez que, conforme discutido, seria possível determinar o volume de passageiros a cada trecho das viagens, e os valores tarifários são fixos.

6.4 Interface de Consulta e Visualização de Dados

Nesta etapa serão apresentados os resultados referentes ao último objetivo específico trabalho, cuja temática é a criação de uma interface de consulta dos dados do sistema de ônibus de Fortaleza, permitindo a visualização dos indicadores calculados na seção anterior, tal como a análise cruzada das informações do banco de dados.

Na seção 5.4.1 são definidos os filtros, as páginas e os visuais a serem apresentados na interface. Iniciando pelos filtros, a figura abaixo mostra a aba lateral criada para a seleção das linhas, sentidos, ônibus, paradas, datas e faixas horárias a serem analisadas.

Figura 23 - Aba de Seleção de Filtros da Interface de Consulta

Aba de Seleção de Filtros da Interface de Consulta

Linha:
Campus do Pici/Unifor

Sentido:
Todos

Parada:
Todos

Ônibus:
Todos

Data:
22/10/2018 27/10/2018

Faixa Horária
Seleções múltiplas

Fonte: Elaborado pelo autor.

A aba de filtros deve estar implementada em todas as páginas da interface, e mantendo a sincronia, de forma a facilitar a análise pelos usuários. Conforme mencionado anteriormente, o escopo do trabalho foi reduzido à linha 75 – Campus do Pici / Unifor, e, portanto, é a única opção selecionável na segmentação de linha.

Os filtros afetam todos os visuais da página, então qualquer seleção aqui afetará os indicadores, gráficos e mapas da interface, inclusive das outras páginas, já que estão sincronizadas.

Consistente também em todas as páginas da interface são o cabeçalho e o rodapé da página. No cabeçalho constam o logotipo do Departamento de Engenharia de Transportes da Universidade Federal do Ceará, o título da interface, e nomes do aluno e professor orientador responsáveis por este trabalho. O rodapé tem função de seleção da página da interface, dentre as quatro propostas no método. Para alterar a página, basta clicar na página desejada, e a interface mudará para as informações correspondentes.

A Figura 24 abaixo mostra os filtros, rodapé e cabeçalho da interface.

Figura 24 - Cabeçalho, Rodapé e Aba de Filtros da Interface

Cabeçalho:

- Logo: Detran
- Título: Interface de Consulta de Indicadores e Visualização de Dados Sistema de Transporte Público de Ônibus de Fortaleza
- Aluno: João Vítor Abreu da Silva
- Orientador: Prof. Francisco Moraes
- Universidade Federal do Ceará

Indicadores da Operação

Filtros:

- Linha: Campus do Pici/Unifor
- Sentido: Todos
- Parada: Todos
- Ônibus: Todos
- Data: 22/10/2018 a 27/10/2018
- Faixa Horária: Seleções múltiplas

Rodapé:

- Indicadores da Operação
- Indicadores de Eficiência
- Avaliação de Headway
- Tempos e Velocidades

Fonte: Elaborado pelo autor.

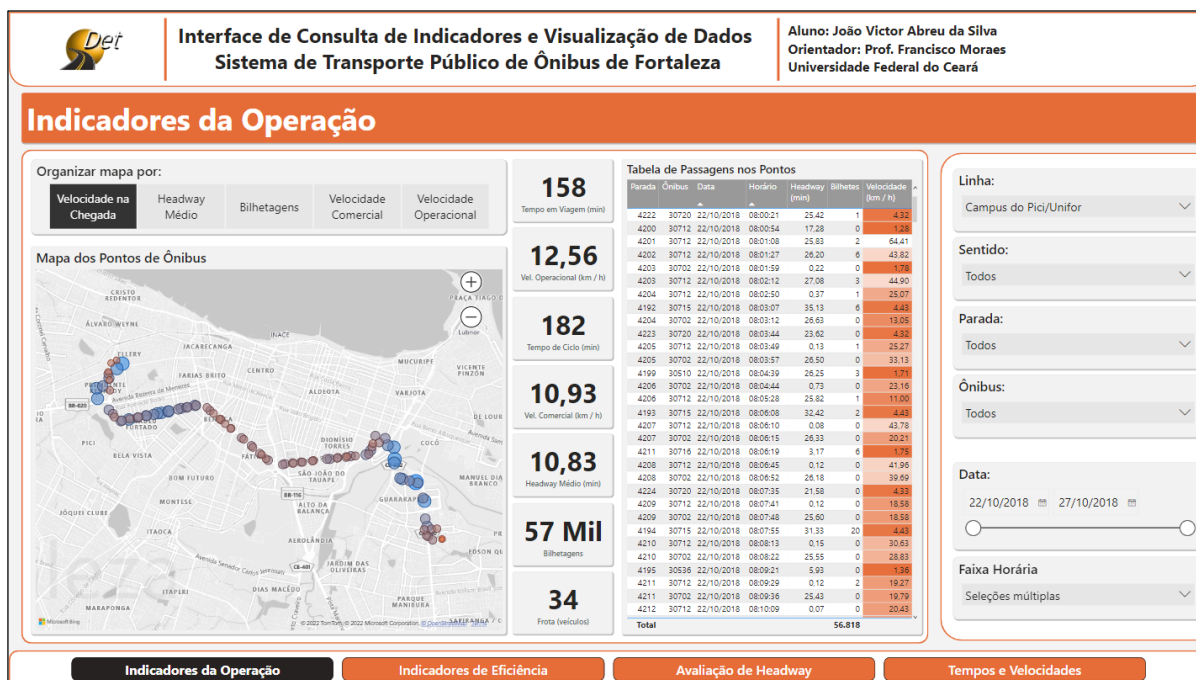
O espaço vazio no centro da interface de consulta é onde são implementados os visuais propostos de cada página. Em laranja, abaixo do cabeçalho, mostra-se o título da página.

6.4.1 Página de Indicadores da Operação

Conforme descrito, na página inicial, referente aos indicadores da operação, foram implementados os cartões informativos com os valores dos indicadores, a tabela com as informações de passagens de ônibus em todas as paradas (sempre obedecendo os filtros selecionados) e o mapa com as paradas de ônibus cuja coloração e tamanho varia com os indicadores de velocidade na chegada, *headway* médio, bilhetagens, velocidade comercial e velocidade operacional, a depender do que for selecionado.

A Figura 25 apresenta a página de Indicadores da Operação com os visuais implementados.

Figura 25 - Página de Indicadores da Operação da Interface de Consulta



Fonte: Elaborado pelo autor.

O menu superior ao *Mapa dos Pontos de Ônibus* permite a seleção do indicador que por sua vez afeta o mapa. No estado de seleção da figura, tem-se o mapa organizado por *Velocidade na Chegada*; bolinhas maiores e mais azuladas representam paradas com maior velocidade, enquanto menores e alaranjadas representam menor velocidade. Selecionando outros indicadores acima do mapa trará um comportamento semelhante de organização das paradas da linha, mudando apenas que a avaliação ocorrerá agora com referência ao indicador selecionado.

Ao lado do mapa tem-se o valor absoluto dos indicadores para o contexto de filtragem selecionado, permitindo que o usuário já perceba rapidamente a situação operacional daquele sistema de transportes. No período filtrado, por exemplo, tiveram 57 mil bilhetagens, e velocidades operacionais e comerciais pequenas, abaixo de 15 km / h.

A tabela mostra, para cada vez que um ônibus passa em uma parada de ônibus, as informações pertinentes àquele registro: qual a parada, qual o ônibus, qual a data e horário, quantas bilhetagens advém daquela parada, o *headway* até o veículo posterior passar e a velocidade estimada do veículo na chegada.

A qualquer momento, o usuário pode mudar o contexto de filtros tanto clicando nos filtros laterais, como pode clicar em uma parada específica do mapa que queira avaliar. Clicando numa parada, os indicadores são recalculados, tal como a tabela.

6.4.2 Página de Indicadores de Eficiência

Quanto à página de Indicadores de Eficiência, foram implementados cartões informativos dos indicadores de bilhetagens, quilômetros rodados, frota, índice de passageiros por quilômetro, índice de passageiros por veículo, índice de quilômetros por veículo e a distância média entre paradas.

Para facilitar a visualização da demanda ao longo do tempo e do espaço, foram implementados também os gráficos de bilhetagem por faixa horária e o mapa de bilhetagem por parada, que pode ser alternado para uma lista de paradas por bilhetagem. O usuário pode ainda clicar numa faixa horária do gráfico e automaticamente o mapa será filtrado para refletir somente informações do horário selecionado.

A tabela com as informações por ônibus também foi implementada, permitindo avaliar e comparar o desempenho de cada um.

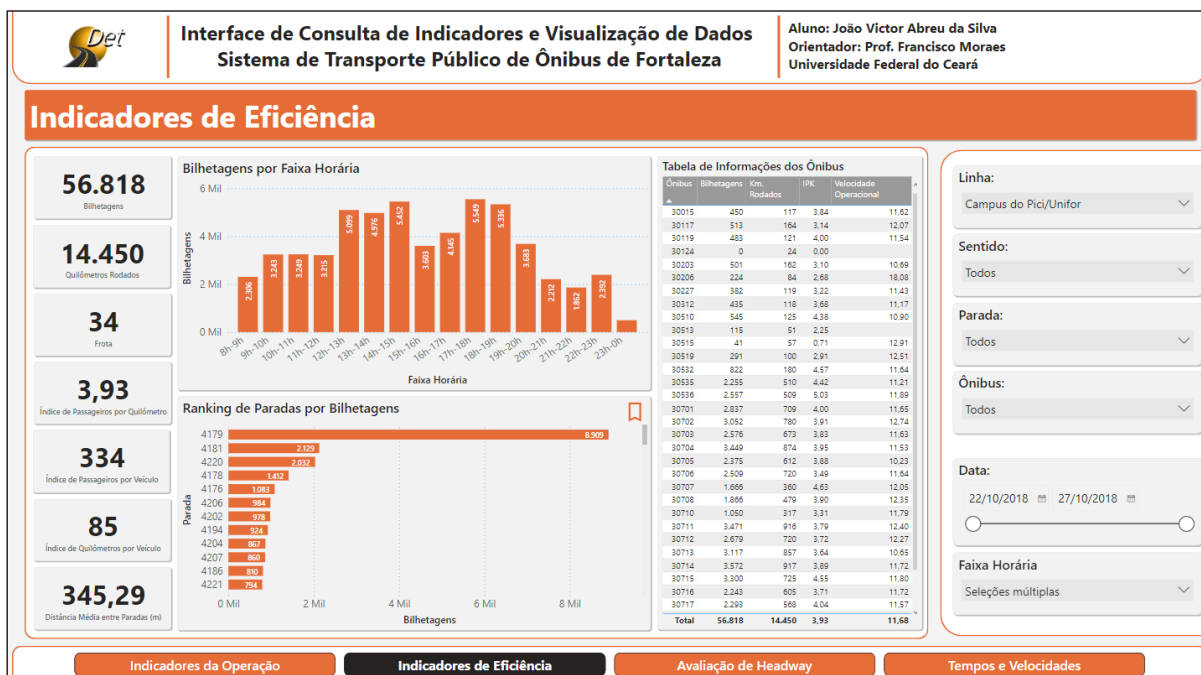
A Figura 26 mostra a página de indicadores de eficiência na interface, e a Figura 27 mostra a mesma página, porém com a opção de alternar o mapa para lista.

Figura 26 - Página de Indicadores de Eficiência da Interface de Consulta



Fonte: Elaborado pelo autor.

Figura 27 - Página de Indicadores de Eficiência da Interface de Consulta, Gráfico Alternativo



Fonte: Elaborado pelo autor.

De forma análoga à página anterior, o usuário pode ver os resultados absolutos dos indicadores calculados na fileira de cartões informativos, para o contexto de filtro selecionado, conseguindo tirar informações brevemente ao olhar para os valores. O IPK, por exemplo, nesse contexto, é inferior a 4, mesmo tendo um número considerável de bilhetagens, de quase 57 mil.

O gráfico de bilhetagens por faixa horária permite ver que horários estão apresentando maior demanda de passageiros, e, logo abaixo, pode-se avaliar as paradas com maior demanda. Ao selecionar uma faixa horária, caso queira, no gráfico superior, o contexto é recalculado observando somente essa faixa, permitindo, por exemplo, ver a parada com maior bilhetagem deste horário.

A tabela do lado direito permite a visualização comparativa dos ônibus no contexto selecionado, mostrando a bilhetagem por ônibus, sua quilometragem, IPK individual e ainda sua velocidade operacional. Ao fim da tabela, tem-se os valores totais destes indicadores, que condizem com a fileira da esquerda.

Essa página tem como principal usuário idealizado o operador da linha, uma vez que este pode averiguar rapidamente os valores dos principais indicadores, avaliar as faixas e paradas mais críticas, e comparar os resultados por ônibus.

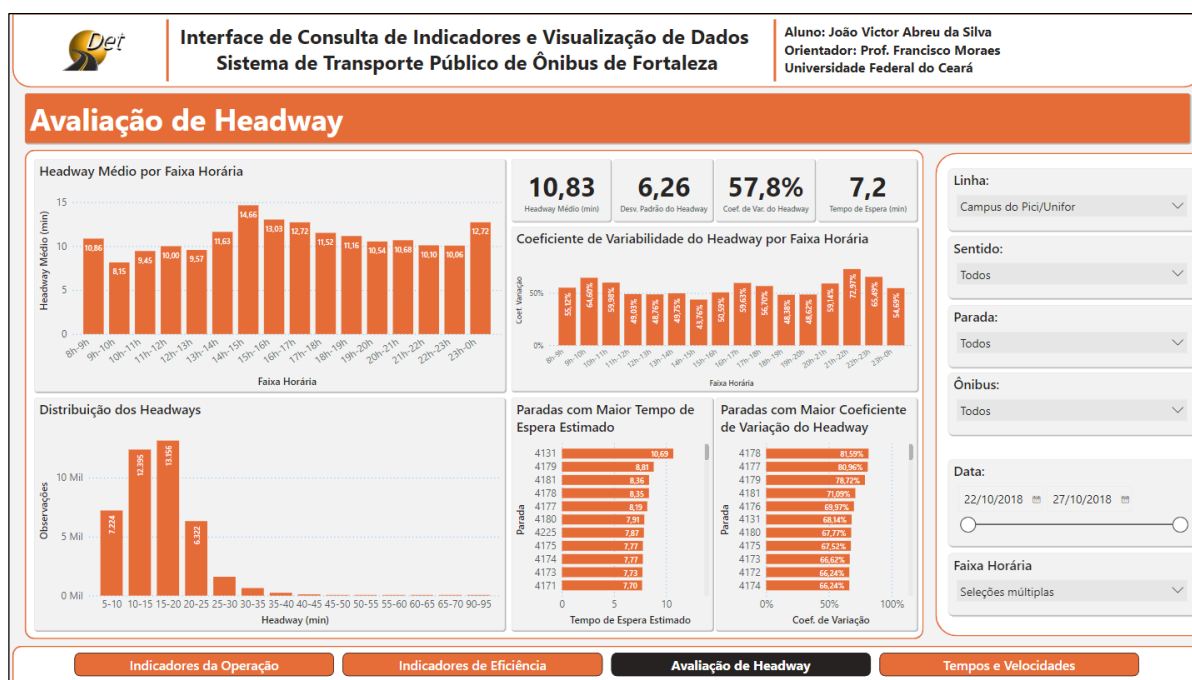
6.4.3 Página de Avaliação do Headway

Esta página tem como intuito principalmente a análise do *headway* por horário, a visualização da sua distribuição e a visualização dos indicadores de *headway* por parada de ônibus.

Como descrito na seção 5.4.1.3, as visualizações necessárias são um gráfico de *headway* médio por faixa horária, um gráfico de distribuição dos *headways*, cartões informativos referentes ao *headway* médio total, desvio padrão, coeficiente de variação e o tempo de espera. Por fim, descreve-se um gráfico com o coeficiente de variação por faixa horária e duas listas ordenadas mostrando as paradas com maior tempo de espera estimado e de maior variação do *headway*.

A Figura 28 apresenta a página implementada na interface, atendendo ao método.

Figura 28 - Página de Avaliação de Headway da Interface de Consulta

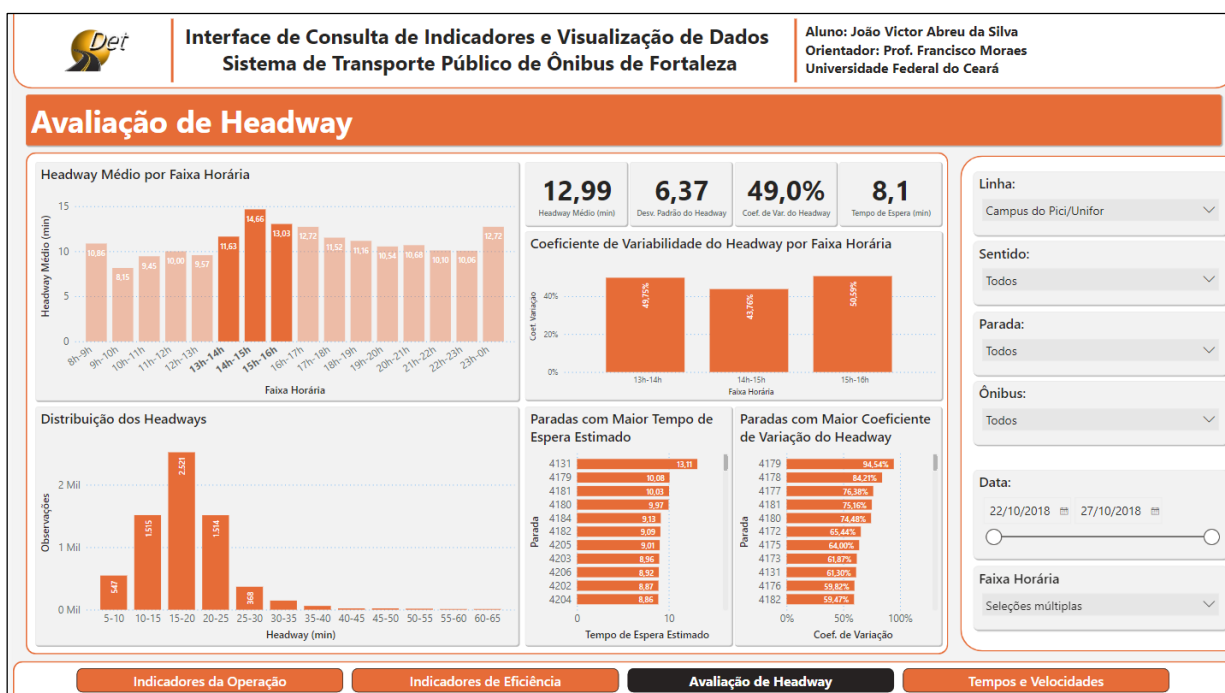


Fonte: Elaborado pelo autor.

É possível tirar informações pertinentes referentes ao *headway* a partir desta página. Por exemplo, é possível ver que o horário mais crítico de *headway* no período foi das 14h às 15h, e que as piores paradas no ponto de vista do tempo de espera e coeficiente de variação são as paradas 4131 e 4178, respectivamente. A partir disto, já pode-se pensar em análises futuras.

Um exemplo de uso é a seleção de uma ou mais faixas horárias no gráfico inicial, o que recalcula os resultados para este contexto. A Figura 29 apresenta o resultado quando selecionadas as faixas horárias de 13h-16h.

Figura 29 - Página de Avaliação de Headway da Interface de Consulta, faixa horária de 13h-16h



Fonte: Elaborado pelo autor.

Pode-se ver nítida diferença na distribuição dos *headways*, tal como aumento do *headway* médio, do desvio padrão e do tempo de espera estimado. A parada com maior tempo de espera permanece a 4131, porém a parada com maior coeficiente de variação é a 4179.

A ideia é que o planejador possa perceber, a partir desta página da interface, os horários mais críticos no ponto de vista de tempo de espera, além de poder averiguar as distribuições dos *headways* para diferentes contextos. Acredita-se que, por este prisma, os resultados condizem com os objetivos.

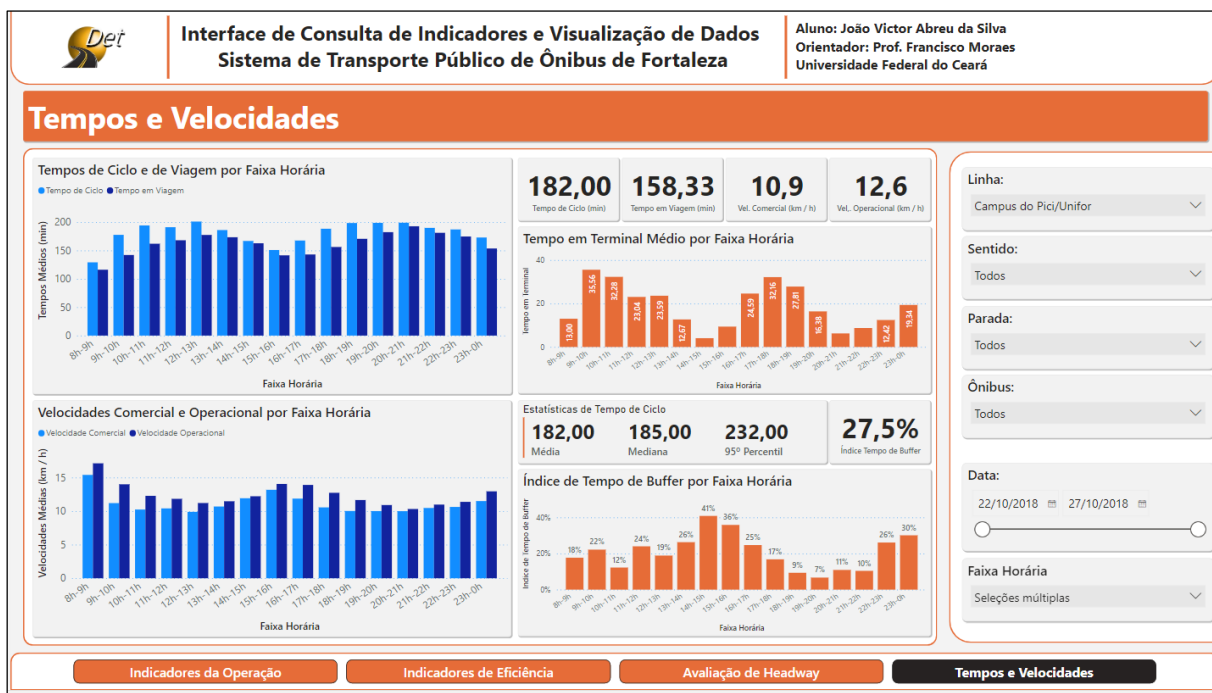
6.4.4 Página de Tempos e Velocidades

Com o intuito de poder avaliar os tempos de ciclo e de viagem, tal como as velocidades comerciais e operacionais, esta página deve conter os visuais propostos na etapa 5.4.1.4. Os visuais são, primariamente dois gráficos de barras agrupadas exibindo no seu eixo horizontal a faixa horária avaliada, e no eixo horizontal os valores agrupados de tempos e velocidades, permitindo a comparação.

Os cartões mostrando os valores gerais dos indicadores foram implementados ao lado destes gráficos, e abaixo dos cartões fora implementada a avaliação do índice de tempo de *buffer* estimado, cujo cálculo está descrito na seção 5.3.3.3.5, mantendo as premissas levantadas.

A Figura 30 apresenta o resultado da criação desta página.

Figura 30 - Página de Tempos e Velocidades da Interface de Consulta



Fonte: Elaborado pelo autor.

Os gráficos do lado esquerdo permitem averiguar, por faixa horária, as discrepâncias entre tempos de ciclo e tempos de viagem no contexto de filtros selecionados. É possível ver, por exemplo, que a velocidade é mais alta nas faixas horárias de 8h-10h, e que a discrepância entre as velocidades operacionais e comerciais não é tão grande, neste contexto.

No lado direito, conforme proposto, os cartões informativos mostram os valores dos quatro principais indicadores desta página. O gráfico abaixo dos cartões permite ver que horários tem maior pausa em terminais, em média. Os horários coincidem com os horários usuais de lanche adotados pelas empresas.

Logo abaixo, o usuário pode visualizar a avaliação do índice de tempo de *buffer* estimado. Mostra-se num cartão único os três valores estatísticos principais para o cálculo do índice, sendo estes a mediana, média e o 95º percentil do tempo de ciclo observado. A partir destes dados, calcula-se ao lado o índice de tempo de *buffer*, que para o contexto selecionado é de 27,5%, ou seja: deve-se fazer um acréscimo de 27,5% no tempo médio previsto da viagem

para garantir que uma observação de tempo de ciclo esteja dentro do intervalo de tempo esperado 95% das vezes.

Finalizando a avaliação visual do índice de tempo de *buffer*, o gráfico abaixo dos cartões mostra o valor do indicador para cada faixa de horário selecionada, permitindo ao usuário perceber em que horários deve ser acrescentado maior margem para atender à pontualidade prevista. Os horários com maior índice de tempo de *buffer* são necessariamente aqueles com maior variabilidade de tempo de ciclo, chamando atenção para a operação do sistema nestes horários específicos.

Acredita-se que os resultados podem ser utilizados tanto no ponto de vista da operação, para perceber inconsistências ou dificuldades no sistema, quanto para o usuário do sistema, que pode ter uma melhor capacidade de avaliar os tempos necessários para fazer viagens específicas, dependendo do horário. Isso é consistente com a proposta pela literatura, onde as interfaces de consulta de dados são utilizadas para avaliação do desempenho do sistema e mesmo de linhas específicas (HANFT *et al.*, 2016).

O código QR referente à interface é exibido na Figura 31, e pode ser utilizado por qualquer leitor para acessá-la.

Figura 31 - Código QR para Acesso à Interface de Consulta



Fonte: Elaborado pelo autor.

7 CONSIDERAÇÕES FINAIS

Frente aos objetivos propostos no início do trabalho, considera-se que o projeto teve sucesso na sua execução. O objetivo geral do trabalho é encapsulado no uso da interface criada, permitindo, conforme proposto, a utilização de ferramentas de visualização de dados para o auxílio na compreensão de um sistema de transportes portando coleta de dados automatizada de bilhetagem e geolocalização.

7.1 Conclusão

O servidor e banco de dados criado permite o acesso aos dados tratados de forma simples, rápida e normalizada, utilizando um banco de dados com linguagem estruturada em *SQL*, que é ainda o padrão mais utilizado no mercado de tecnologia de dados. A utilização do banco de dados para consultas permite o trabalho com os dados em diversos *softwares* e ferramentas distintas, o que não é o caso para os dados não tratados, que consiste em múltiplos arquivos separados e sem normalização. A implementação do banco de dados confere aos dados, ainda, maior democratização, uma vez que qualquer usuário com conexão permitida ao banco é capaz de utilizar os dados para análise própria, ou mesmo na elaboração de novos trabalhos acadêmicos. Além disso, o banco de dados tem suas consultas e códigos disponibilizados, permitindo melhorias futuras.

A modelagem da passagem dos ônibus em cada parada por meio da utilização dos diagramas de distância por horário possibilita o cálculo de diversos tipos de indicadores e parâmetros referente ao sistema de transportes de ônibus, conforme foi demonstrado nos resultados. O fato dessa modelagem ter sido implementada no banco de dados também facilita o acesso à informação, permitindo que a abordagem seja refinada e servindo de insumo para comparação de outras técnicas de interpolação do momento da passagem dos ônibus nas paradas.

Conforme foi descrito ao longo do trabalho, foi possível calcular muitos dos indicadores importantes da operação, de eficiência e qualidade. O cálculo dos tempos de ciclo e de viagem, *headway* e frota a partir dos dados permitem averiguar se o sistema está atendendo ao planejamento proposto, identificando possíveis melhorias e oportunidades de replanejamento da operação. O cálculo dos indicadores permite ainda avaliações *ex-post* e *ex-ante* do sistema, percebendo seu comportamento antes de alterações na operação, por exemplo, ou servindo de insumo para a proposição dessas variações futuras.

Além disso, apesar do cálculo dos indicadores ter sido restringido pela carência de dados de desembarque, principalmente, o cálculo da bilhetagem por ônibus, parada, linha, faixa horária e afins permite que as pesquisas físicas de demandas, comuns em sistemas de transporte público, sejam significativamente diminuídas, uma vez que as demandas críticas podem ser determinadas pelo levantamento do fator de renovação por si só, ou ainda, fazendo pesquisas somente de desembarque e de demandas nos terminais, pois são as variáveis que não são contempladas nos dados. Essas considerações de cálculo permitiram, ainda, trazer maior clareza às limitações dos sistemas de *AVL* e *AFC* no ponto de vista de determinação de indicadores.

A interface criada trouxe resultados satisfatórios de usabilidade, o que permite a dinamização da análise dos dados do sistema de transporte de ônibus. Os visuais implementados possibilitam uma análise cruzada das informações por linha, sentido, faixa horária e diversas outras variações, permitindo que o interlocutor tenha uma compreensão mais profunda do sistema analisado, porém sem sacrificar a clareza dos resultados.

Pelo fato de a interface ter seu código disponibilizado, também é possível implementar mais visualizações que atendam às demandas necessárias a partir dos dados do banco, ou mesmo inserir mais dados além dos que foram disponibilizados. A implementação da interface *online* também traz a vantagem de acessibilidade, uma vez que qualquer usuário pode ver os dados e fazer sua própria análise, independente do seu objetivo.

Ademais, conclui-se que o trabalho foi capaz de atingir os objetivos propostos, possibilitando a melhoria, mesmo que sucinta, da operação e da compreensão dos sistemas de transporte frente às tecnologias de dados mais robustas.

7.2 Sugestões para Trabalhos Futuros

Ao longo do projeto foi possível perceber algumas lacunas ou alternativas que podem ser exploradas no futuro, tanto para o avanço do que foi feito neste trabalho, quanto para estudos independentes que tenham objetivos ou métodos semelhantes. Essa seção descreve sucintamente estas sugestões, agrupadas pelos objetivos deste trabalho.

7.2.1 Referente à Base de Dados

No que se refere à criação da base de dados, traz-se as seguintes sugestões:

- a) criação do banco de dados por meio de uma estruturação de bases não relacionadas, como o *mongoDB*. Nesse caso, as informações não serão

estruturadas por meio de tabelas, mas sim objetos que se adequam à necessidade dos dados. Essa sugestão se contrapõe à estrutura usada no presente trabalho, que usa tabelas relacionadas e normalização destas tabelas. O objetivo da criação do banco por essa estrutura seria comparar o desempenho das duas estruturas no ponto de vista de velocidade de consulta, velocidade de carga de dados e facilidade de utilização;

- b) modelar, a partir da tabela de bilhetagens, as viagens individuais, consistindo dos embarques e desembarques, utilizando o identificador único do cartão de bilhetagem. O método de modelagem utilizaria as informações de bilhetagens particionadas por este identificador, e assim seria mapeado toda vez que este cartão fez uma bilhetagem; dessa forma, pode-se estimar de forma aproximada os desembarques e as matrizes origem / destino do sistema. O ganho dessa abordagem está em eliminar algumas limitações dos dados de *AVL* e *AFC*, permitindo, por exemplo, a inferência direta da demanda crítica.

7.2.2 Referente à Modelagem do Horário nas Paradas

No que se refere à modelagem do horário que cada ônibus passa em cada ponto de controle, traz-se duas sugestões:

- a) modelagem do horário da chegada de cada ônibus ao ponto de controle utilizando as informações de horário e histórico de densidade de tráfego; dessa forma, pode ser possível encontrar uma modelagem mais precisa que a utilizada nesse trabalho. A sugestão inicial é que se utilize as informações de densidade de tráfego armazenadas pelo *Google Maps* como variável para uma função de velocidade de trajeto no trecho avaliado. A partir da calibração da função, pode-se implementar o modelo nos dados e comparar com a interpolação feita neste trabalho;
- b) utilização do método descrito neste trabalho para a modelagem em sistema de transporte de ônibus que tenham maior granularidade referente às posições (*AVL*). Os dados de posição deste trabalho têm granularidade de 30 segundos, o que muitas vezes implica que o ônibus pode passar por duas paradas próximas sem ter um registro detectado, a depender de sua velocidade e a distância das paradas. Um sistema com maior granularidade, que tenha observações a cada 10 segundos, por exemplo, pode trazer grandes diferenças nos resultados. Essas

diferenças podem ser mensuradas e comparadas, determinando se o aumento do armazenamento de dados é justificável, ou se maior granularidade não oferece maior precisão nesse tipo de modelagem.

7.2.3 Referente aos Indicadores

Quanto à determinação e os cálculos dos indicadores propostos, são sugeridas três possibilidades de estudos futuros:

- a) Utilização dos dados de viagens individuais para inferir a lotação e a demanda crítica dos ônibus. Essa sugestão é uma proposição consequente da sugestão (b) da seção 7.2.1. Portando os dados de viagens e matrizes de origem e destino, o cálculo dos indicadores de lotação e demanda crítica, que são extremamente importantes para a aferição da qualidade e para o planejamento do sistema, respectivamente, é muito facilitado. Caso seja possível, os resultados deste objetivo complementariam os resultados deste trabalho e suas conclusões, pois diminuiria as limitações destes sistemas de coleta automática de dados (*AVL* e *AFC*);
- b) Coleta das características dos ônibus. É uma sugestão operacional, mas o uso dos dados referentes às características dos veículos, como a capacidade de passageiros e consumo de combustível pelo veículo, ajuda na descrição do sistema de transportes. Portando estes dados e os de demanda crítica inferida, é possível mapear todo o processo de planejamento da linha, considerando inclusive os custos.

7.2.4 Referente à Interface

No que se refere à interface de consulta desenvolvida, apresenta-se somente uma sugestão, que é elaboração e aplicação de pesquisas qualitativas com possíveis usuários da interface, como passageiros do sistema, operadores e as agências responsáveis em busca de determinar informações que cada um dos usuários necessita. Partindo dessas informações e sugestões, pode-se aplicar as mudanças à interface e apresentar novamente aos usuários, averiguando sua aceitação antes e após as alterações. Esse processo iterativo traria melhorias expressivas à interface e sua utilização, contribuindo para difusão da compreensão dos sistemas de transportes de ônibus.

REFERÊNCIAS

- AIN, N. U. *et al.* Two decades of research on business intelligence system adoption, utilization and success – A systematic literature review. **Decision Support Systems**, v. 125, 1 out. 2019.
- BARABINO, B.; DI FRANCESCO, M.; MOZZONI, S. Rethinking bus punctuality by integrating Automatic Vehicle Location data and passenger patterns. **Transportation Research Part A: Policy and Practice**, v. 75, p. 84–95, 1 maio 2015.
- BARABINO, B.; DI FRANCESCO, M.; MOZZONI, S. Time Reliability Measures in bus Transport Services from the Accurate use of Automatic Vehicle Location raw Data. **Quality and Reliability Engineering International**, v. 33, n. 5, p. 969–978, 1 jul. 2017.
- BÖRNER, K.; BUECKLE, A.; GINDA, M. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. **Proceedings of the National Academy of Sciences of the United States of America**, v. 116, n. 6, p. 1857–1864, 5 fev. 2019.
- BRAGA, C. K. V.; LOUREIRO, C. F. G.; PEREIRA, R. H. M. Analisando a variabilidade de estimativas de acessibilidade por transporte público a partir de dados de GPS. **TRANSPORTES**, v. 28, n. 5, p. 169–184, 15 dez. 2020.
- CASCETTA, E.; CARTENÌ, A. A Quality-Based Approach to Public Transportation Planning: Theory and a Case Study. **International Journal of Sustainable Transportation**, v. 8, n. 1, p. 84–106, 2014.
- CEDER, A. **Public Transit Planning and Operation: Theory, Modeling and Practice**. 1^a ed. Haifa, Israel: Butterworth-Heinemann, 2007.
- FAROQI, H.; MESBAH, M.; KIM, J. Applications of transit smart cards beyond a fare collection tool: A literature review. **Advances in Transportation Studies**, v. 45, p. 107–122, 1 jul. 2018.
- FERRAZ, A. C. “COCA” P.; TORRES, I. G. E. **Transporte Público Urbano**.^a 2^a ed. São Carlos, SP: RiMa Editora, 2001.
- FORTUIN, L. Performance indicators-Why, where and how? **European Journal of Operational Research**, v. 34, p. 1–9, 1988.
- FRANCESCHINI, F.; GALETTO, M.; MAISANO, D. **Management by Measurement: Designing Key Indicators and Performance Measurement Systems**. 1. ed. Leipzig: Springer, 2007.
- GUTIÉRREZ, A. *et al.* Profiling tourists’ use of public transport through smart travel card data. **Journal of Transport Geography**, v. 88, 1 out. 2020.
- HALVORSEN, A. *et al.* Demand management of congested public transport systems: a conceptual framework and application using smart card data. **Transportation**, v. 47, n. 5, p. 2337–2365, 1 out. 2020.
- HANFT, J. *et al.* Transforming Bus Service Planning Using Integrated Electronic Data Sources at NYC Transit. **Journal of Public Transportation**, v. 19, n. 2, p. 89–108, 2016.

HENKEL, M.; PERJONS, E.; SNEIDERS, E. Examining the potential of language technologies in public organizations by means of a business and IT architecture model. **International Journal of Information Management**, v. 37, n. 1, p. 1507–1516, 1 fev. 2017.

HIGATANI, A. *et al.* Empirical analysis of travel time reliability measures in hanshin expressway network. **Journal of Intelligent Transportation Systems: Technology, Planning, and Operations**, v. 13, n. 1, p. 28–38, jan. 2009.

HINGORANI, K.; GITTENS, D.; EDWARDS, N. REINFORCING DATABASE CONCEPTS BY USING ENTITY RELATIONSHIPS DIAGRAMS (ERD) AND NORMALIZATION TOGETHER FOR DESIGNING ROBUST DATABASES. **Issues In Information Systems**, v. 18, n. 1, p. 148–155, 2017.

HORBURY, A. X. Using non-real-time Automatic Vehicle Location data to improve bus services. **Transproation Research Part B**, n. 33, p. 559–579, 1999.

HOUNSELL, N. B.; SHRESTHA, B. P.; WONG, A. Data management and applications in a world-leading bus fleet. **Transportation Research Part C: Emerging Technologies**, v. 22, p. 76–87, 2012.

KARNEY, C. F. F. Algorithms for geodesics. **Journal of Geodesy**, v. 87, n. 1, p. 43–55, 1 jan. 2013.

KÖHLER, H.; LINK, S. SQL schema design: foundations, normal forms, and normalization. **Information Systems**, v. 76, p. 88–113, 1 jul. 2018.

LEE, S.; SIPS, M.; SEIDEL, H. P. Perceptually driven visibility optimization for categorical data visualization. **IEEE Transactions on Visualization and Computer Graphics**, v. 19, n. 10, p. 1746–1757, 2013.

MA, X. Development of a data-driven platform for transit performance measures using smart card and GPS data. **Journal of Transportation Engineering**, v. 140, n. 12, 1 dez. 2014.

MA, Z.; FERREIRA, L.; MESBAH, M. A Framework for the Development of Bus Service Reliability Measures. **Australasian Transport Research Forum**, 2013.

MATZEN, L. E. *et al.* Data Visualization Saliency Model: A Tool for Evaluating Abstract Data Visualizations. **IEEE Transactions on Visualization and Computer Graphics**, v. 24, n. 1, p. 563–573, 1 jan. 2018.

MISHALANI, R. G.; MCCORD, M. M. **Passenger Wait Time Perceptions at Bus Stops** **Passenger Wait Time Perceptions at Bus Stops: Empirical Results and Impact on Evaluating Real-Time Bus Arrival Information Motivation and Hypothesis** **Journal of Public Transportation**. [s.l: s.n.].

MORENO, V.; CAVAZOTTE, F.; DE SOUZA CARVALHO, W. Business intelligence and analytics as a driver of dynamic and operational capabilities in times of intense macroeconomic turbulence. **Journal of High Technology Management Research**, v. 31, n. 2, 1 nov. 2020.

NAMVAR, M.; CYBULSKI, J. L.; PERERA, L. Using business intelligence to support the process of organizational sensemaking. **Communications of the Association for Information Systems**, v. 38, n. 1, p. 330–352, 1 mar. 2016.

NESTOROV, S. *et al.* Generating insights through data preparation, visualization, and analysis: Framework for combining clustering and data visualization techniques for low-cardinality sequential data. **Decision Support Systems**, v. 125, 1 out. 2019.

ORACLE. **MySQL 8.0 Reference Manual Including MySQL NDB Cluster 8.0**. Oracle, , 2022.

PETTERSSON, F.; SØRENSEN, C. H. Why do cities invest in bus priority measures? Policy, polity, and politics in Stockholm and Copenhagen. **Transport Policy**, v. 98, p. 178–185, 1 nov. 2020.

SANTOS, V. *et al.* Decision support for reliability improvement of multimodal public transport services. **Transport Research Arena**, v. 2020, 27 abr. 2020.

UTOMO, R. G.; MAYLAWATI, D. S.; ALAM, C. N. Implementasi Algoritma Cheapest Insertion Heuristic (CIH) dalam Penyelesaian Travelling Salesman Problem (TSP). **Jurnal Online Informatika**, v. 3, n. 1, p. 61, 30 jun. 2018.

VALLURUPALLI, V.; BOSE, I. Business intelligence for performance measurement: A case based analysis. **Decision Support Systems**, v. 111, p. 72–85, 1 jul. 2018.

ZHAO, J.; RAHBEE, A.; WILSON, N. H. M. Estimating a Rail Passenger Trip Origin-Destination Matrix Using Automatic Data Collection Systems. **Computer-Aided Civil and Infrastructure Engineering**, v. 22, p. 376–387, 2007.

APÊNDICE A – CONSULTAS DE CÁLCULOS DOS INDICADORES

a) Cálculo dos *Headways*:

```

1 SELECT
2     *,
3     CEILING(HEADWAY / 5.0) * 5 ID_BUCKET_HEADWAY
4 FROM
5     (
6         SELECT
7             ROW_NUMBER() OVER (ORDER BY ID_PONTO_DE_CONTROLE, DT_CHEGADA) `ID_HEADWAY`,
8             ID_ACOMPANHAMENTO,
9             ID_ONIBUS,
10            ID_PONTO_DE_CONTROLE,
11            DT_CHEGADA,
12            LEAD(DT_CHEGADA) OVER (PARTITION BY ID_PONTO_DE_CONTROLE ORDER BY DT_CHEGADA) `DT_CHEGADA_SEGUINTE`,
13            TIMESTAMPDIFF(second, DT_CHEGADA, LEAD(DT_CHEGADA) OVER (PARTITION BY ID_PONTO_DE_CONTROLE ORDER BY DT_CHEGADA)) / 60 `HEADWAY`
14        FROM
15            t_acompanhamento_real A
16    ) `HEADWAYS`
17 WHERE
18     HEADWAY IS NOT NULL AND HEADWAY <> 0

```

b) Cálculo das Bilhetagens:

```

1 WITH
2     cte_acompanhamento AS (
3         SELECT
4             ID_ACOMPANHAMENTO,
5             ID_ONIBUS,
6             ID_PONTO_DE_CONTROLE,
7             DT_CHEGADA,
8             LEAD(DT_CHEGADA) OVER (PARTITION BY ID_ONIBUS ORDER BY DT_CHEGADA) `DT_CHEGADA_PARADA_SEGUINTE`
9         FROM
10            t_acompanhamento_real A
11     )
12
13 SELECT
14     ID_ACOMPANHAMENTO,
15     COUNT(ID_BILHETAGEM) BILHETAGENS
16 FROM
17     cte_acompanhamento A
18     LEFT JOIN T_BILHETAGEM B
19         ON A.ID_ONIBUS = B.ID_ONIBUS
20         AND DT_BILHETAGEM BETWEEN DT_CHEGADA AND DT_CHEGADA_PARADA_SEGUINTE
21 GROUP BY
22     ID_ACOMPANHAMENTO

```

c) Cálculo do Tempo de Ciclo:

```

1  SELECT
2      ID_ACOMPANHAMENTO,
3      TIMESTAMPDIFF(minute, DT_CHEGADA, PASSAGEM_SEGUINTE) TEMPO_CICLO,
4      CEILING(TIMESTAMPDIFF(minute, DT_CHEGADA, PASSAGEM_SEGUINTE)/ 5.0) * 5 ID_BUCKET_TEMPO_CICLO
5
6  FROM
7      (
8      SELECT
9          *,
10         LEAD(DT_CHEGADA) OVER (PARTITION BY ID_ONIBUS, ID_PONTO_DE_CONTROLE ORDER BY DT_CHEGADA) `PASSAGEM_SEGUINTE`
11      FROM
12          T_ACOMPANHAMENTO_REAL
13      ) A
14  WHERE
15      TIMESTAMPDIFF(minute, DT_CHEGADA, PASSAGEM_SEGUINTE) IS NOT NULL

```

d) Cálculo do Tempo em Viagem:

```

1  with
2  cte_1 AS (
3      SELECT
4          ID_ACOMPANHAMENTO,
5          ID_ONIBUS,
6          ID_PONTO_DE_CONTROLE,
7          DT_CHEGADA,
8          DT_CHEGADA_FIM
9      FROM
10         (
11         SELECT
12             A1.ID_ACOMPANHAMENTO,
13             A1.ID_ONIBUS,
14             A1.ID_PONTO_DE_CONTROLE,
15             A1.DT_CHEGADA,
16             A2.DT_CHEGADA DT_CHEGADA_FIM,
17             ROW_NUMBER() OVER (PARTITION BY A1.ID_ACOMPANHAMENTO ORDER BY TIMESTAMPDIFF(second, A1.DT_CHEGADA, A2.DT_CHEGADA)) ORDEM
18         FROM
19             T_ACOMPANHAMENTO_REAL A1
20         INNER JOIN T_ACOMPANHAMENTO_REAL A2
21             ON A1.ID_ONIBUS = A2.ID_ONIBUS
22             AND (
23                 IF(
24                     A1.ID_PONTO_DE_CONTROLE = 4130,
25                     A2.DT_CHEGADA > A1.DT_CHEGADA,
26                     A2.DT_CHEGADA >= A1.DT_CHEGADA
27                 )
28             )
29         AND A2.ID_PONTO_DE_CONTROLE = 4225

```



```

30     ) A
31     WHERE
32         ORDEM = 1
33     ORDER BY
34         ID_PONTO_DE_CONTROLE
35 )
36 SELECT
37     ID_ACOMPANHAMENTO,
38     TIMESTAMPDIFF(minute, DT_CHEGADA_INICIO, DT_CHEGADA_FIM) `TEMPO_VIAGEM`
39 FROM
40     (
41         select *
42         from
43             (
44                 SELECT
45                     A1.ID_ACOMPANHAMENTO,
46                     A1.ID_ONIBUS,
47                     A1.ID_PONTO_DE_CONTROLE,
48                     A1.DT_CHEGADA,
49                     A1.DT_CHEGADA_FIM,
50                     A2.DT_CHEGADA DT_CHEGADA_INICIO,
51                     ROW_NUMBER() OVER (PARTITION BY A1.ID_ACOMPANHAMENTO ORDER BY TIMESTAMPDIFF(second, A2.DT_CHEGADA, A1.DT_CHEGADA)) ORDEM
52                 FROM
53                     cte_1 A1
54                     INNER JOIN T_ACOMPANHAMENTO_REAL A2
55                         ON A1.ID_ONIBUS = A2.ID_ONIBUS
56                         AND A2.ID_PONTO_DE_CONTROLE = 4130
57                         AND (
58                             IF(

```

```

59                     A1.ID_PONTO_DE_CONTROLE = 4225,
60                     A2.DT_CHEGADA < A1.DT_CHEGADA,
61                     A2.DT_CHEGADA <= A1.DT_CHEGADA
62                 )
63             )
64         ORDER BY
65             A1.ID_PONTO_DE_CONTROLE
66     ) A
67     WHERE
68         ORDEM = 1
69 ) A
70 WHERE
71     TIMESTAMPDIFF(minute, DT_CHEGADA_INICIO, DT_CHEGADA_FIM) <= 240

```

e) Cálculo da Velocidade:

```
1 • SELECT
2     ID_ACOMPANHAMENTO,
3     (DISTANCIA / TEMPO) * 3.6 `VELOCIDADE`
4 FROM
5     (
6     SELECT
7         ID_ACOMPANHAMENTO,
8         timestampdiff(second, DT_ANTERIOR, DT_CHEGADA) `TEMPO`,
9         SUM(DISTANCIA_PERCORRIDA) DISTANCIA
10    FROM
11        (
12        SELECT
13            *,
14            LAG(ID_PONTO_DE_CONTROLE) OVER (PARTITION BY ID_ONIBUS ORDER BY DT_CHEGADA) `PONTO_ANTERIOR`,
15            LAG(DT_CHEGADA) OVER (PARTITION BY ID_ONIBUS ORDER BY DT_CHEGADA) `DT_ANTERIOR`
16        FROM
17            T_ACOMPANHAMENTO_REAL
18        ) A1
19    LEFT JOIN T_PONTO_DE_CONTROLE PC
20        ON PC.ID_PONTO_DE_CONTROLE BETWEEN A1.PONTO_ANTERIOR + 1 AND A1.ID_PONTO_DE_CONTROLE
21        AND ID_ROTA IN (86, 87)
22    GROUP BY
23        ID_ACOMPANHAMENTO,
24        timestampdiff(second, DT_ANTERIOR, DT_CHEGADA)
25    ) A
26 WHERE
27     (DISTANCIA / TEMPO) * 3.6 < 65
```

APÊNDICE B – CONSULTAS DE CRIAÇÃO DE TABELAS

a) Criação da Tabela de Ônibus:

```
1 * CREATE TABLE t_onibus (  
2     ID_ONIBUS int PRIMARY KEY,  
3     ID_VEICULO int  
4 )
```

b) Criação da Tabela de Paradas:

```
1 * CREATE TABLE t_parada (  
2     ID_PARADA int PRIMARY KEY,  
3     DS_PARADA VARCHAR(255),  
4     LAT float,  
5     LON float  
6 )
```

c) Criação da Tabela de Linhas:

```
1 * CREATE TABLE t_linha (  
2     ID_LINHA int PRIMARY KEY,  
3     DS_LINHA varchar(255)  
4 )
```

d) Criação da Tabela de Rotas:

```
1 * CREATE TABLE t_rota (  
2     ID_ROTA int PRIMARY KEY,  
3     ID_LINHA int,  
4     FL_SENTIDO tinyint(1),  
5     FOREIGN KEY (ID_LINHA) REFERENCES t_linha(ID_LINHA)  
6 )
```

e) Criação da Tabela de Shapes:

```
1  CREATE TABLE t_shape (  
2      ID_SHAPE int PRIMARY KEY,  
3      ID_ROTA int,  
4      SEQUENCIA int,  
5      LAT float,  
6      LON float,  
7      FOREIGN KEY (ID_ROTA) REFERENCES t_rota(ID_ROTA)  
8  )
```

f) Criação da Tabela de Pontos de Controle:

```
1  CREATE TABLE t_ponto_de_controle (  
2      ID_PONTO_DE_CONTROLE int PRIMARY KEY,  
3      ID_ROTA int,  
4      ID_PARADA int,  
5      ID_ORDEM int,  
6      FOREIGN KEY (ID_ROTA) REFERENCES t_rota(ID_ROTA),  
7      FOREIGN KEY (ID_PARADA) REFERENCES t_parada(ID_PARADA)  
8  )
```

g) Criação da Tabela de Tabelas:

```
1  CREATE TABLE t_tabela (  
2      ID_TABELA int PRIMARY KEY,  
3      ID_ONIBUS int,  
4      ID_ROTA int,  
5      FOREIGN KEY (ID_ONIBUS) REFERENCES t_onibus(ID_ONIBUS),  
6      FOREIGN KEY (ID_ROTA) REFERENCES t_rota(ID_ROTA)  
7  )
```

h) Criação da Tabela de Posições:

```
1 • CREATE TABLE t_posicao (  
2     ID_POSICAO int PRIMARY KEY,  
3     ID_TABELA int,  
4     DT_REGISTRO datetime,  
5     LAT float,  
6     LON float,  
7     FOREIGN KEY (ID_TABELA) REFERENCES t_tabela(ID_TABELA)  
8 )
```

i) Criação da Tabela de Bilhetagens:

```
1 • CREATE TABLE t_bilhetagem (  
2     ID_BILHETAGEM int PRIMARY KEY,  
3     ID_USUARIO int,  
4     DT_BILHETAGEM datetime,  
5     ID_TABELA int,  
6     TIPO_BILHETAGEM varchar(255),  
7     FL_INTEGRACAO tinyint(1),  
8     FOREIGN KEY (ID_TABELA) REFERENCES t_tabela(ID_TABELA)  
9 )
```

j) Criação da Tabela de Acompanhamento:

```
1 • CREATE TABLE t_acompanhamento (  
2     ID_ACOMPANHAMENTO int PRIMARY KEY,  
3     ID_TABELA int,  
4     ID_PONTO_DE_CONTROLE int,  
5     DT_CHEGADA datetime,  
6     FOREIGN KEY (ID_TABELA) REFERENCES t_tabela(ID_TABELA),  
7     FOREIGN KEY (ID_PONTO_DE_CONTROLE) REFERENCES t_ponto_de_controle(ID_PONTO_DE_CONTROLE)  
8 )
```

APÊNDICE C – ROTINA DE TRATAMENTO DOS DADOS

▼ Configurações Iniciais

```
# Importando bibliotecas:
import pandas as pd
from os import listdir

# Importando e montando o Drive:
from google.colab import drive
drive.mount('/content/drive')
```

▼ Criação dos Dataframes

Essa etapa busca criar os arquivos brutos que serão importados para a base de dados SQL para uso das informações.

```
# O arquivo com a estrutura da base de dados se encontra em:
/content/drive/MyDrive/11. Joao Victor/Scripts/Estrutura de Dados Idealizada.txt
```

Serão usados os arquivos de GPS, GTFS e Bilhetagem para a construção da base de dados. O script atual será comentado para a criação de cada um dos dataframes referente a cada tabela descrita no arquivo.

▼ Tabela 1: T_ONIBUS

A tabela T_ONIBUS busca simplesmente interligar as informações de **vehicleid** dos arquivos de GPS com a informação de **prefixo_onibus** contido nos arquivos de bilhetagem.

```
# Importando o arquivo de Bilhetagem em um dataframe:
caminho = '/content/drive/MyDrive/11. Joao Victor/Dados/Bilhetagem'

arquivos = listdir(caminho)

lista = []

for arquivo in arquivos:
    df = pd.read_csv(caminho + '/' + arquivo, index_col = None, header = 0,
                    parse_dates = ["hora"])
    lista.append(df)

df_bilhetagem = pd.concat(lista, axis = 0, ignore_index = True)

# Determinando associação de vehicleid para prefixo_carro:
```

```

distintos = df_bilhetagem[["vehicleid", "prefixo_carro"]].drop_duplicates()
df_onibus = distintos[(distintos["prefixo_carro"].notnull()) &
                      (distintos["vehicleid"].notnull())]
# Isso assume que não há repetição de vehicleid para prefixo_carro distintos.
# No dataset disponibilizado, não há repetição. Foram tirados os nulos também.

# Finalizando a tabela:
df_onibus = df_onibus.astype({"vehicleid": int, "prefixo_carro": int})
df_onibus = df_onibus.rename(columns = {"vehicleid": "id_veiculo",
                                       "prefixo_carro": "id_onibus"})
df_onibus = df_onibus.set_index("id_onibus")

```

▼ Tabela 2: T_LINHA

A tabela T_LINHA tem como intuito armazenar o identificador de cada linha de ônibus de Fortaleza e seus nomes. Para isso, usa-se os dados do GTFS.

```

# Importando o arquivo de GTFS em um dataframe:
caminho = '/content/drive/MyDrive/11. Joao Victor/Dados/GTFS/routes.txt'

df_linhas = pd.read_csv(caminho)

# Tratando e finalizando a tabela:
df_linhas = df_linhas[["route_id", "route_long_name"]]
df_linhas = df_linhas.rename(columns = {"route_id": "id_linha",
                                       "route_long_name": "ds_linha"})
df_linhas = df_linhas.set_index("id_linha")

```

▼ Tabela 3: T_ROTAS

A tabela T_ROTAS tem como intuito separar cada linha em dois sentidos, trazendo um identificador único para cada sentido. Isso é necessário pois muitos dados de GTFS estão separados por ida e volta, tal como as viagens, bilhetagem, etc.

```

# É necessário que sejam usados os dados de GTFS visto que algumas linhas não
# têm ambos os sentidos. O GTFS tem os dados de todas as viagens e, portanto,
# possui as informações necessárias para o identificador único.

# Importando os dados de viagens do GTFS:
caminho = '/content/drive/MyDrive/11. Joao Victor/Dados/GTFS/trips.txt'

df_rotas = pd.read_csv(caminho)

# Tratando os dados para obter as informações necessárias (sentido, linha):
df_rotas = df_rotas[["route_id", "trip_id"]]
df_rotas["trip_id"] = df_rotas["trip_id"].str[-1:]
df_rotas = df_rotas.drop_duplicates()

```

```
df_rotas["trip_id"] = df_rotas["trip_id"].map({"I": 0, "V": 1})

# Finalizando a tabela:
df_rotas.rename(columns = {"route_id": "id_linha", "trip_id": "fl_sentido"},
                inplace = True)
df_rotas = df_rotas.reset_index()
df_rotas = df_rotas[["id_linha", "fl_sentido"]]
df_rotas.index.name = "id_rota"
```

▼ Tabela 4: T_TABELA

A tabela T_TABELA tem como intuito criar uma chave única para cada combinação de ônibus e rota. Assim, será possível analisar separadamente cada ônibus de cada linha em cada direção.

Os dados da tabela vêm da tabela de bilhetagem, pois esta mostra cada bilhetagem em cada ônibus em cada linha, em cada sentido.

```
# Analisando a tabela de bilhetagem:
df_tabela = df_bilhetagem[["prefixo_carro", "linha", "sentido_viagem"]]
df_tabela = df_tabela[df_tabela["prefixo_carro"].notnull()].drop_duplicates()
df_tabela.rename(columns = {"prefixo_carro": "id_onibus", "linha": "id_linha",
                           "sentido_viagem": "fl_sentido"}, inplace = True)
df_tabela["fl_sentido"] = df_tabela["fl_sentido"].map({"Ida": 0, "Volta": 1})
df_tabela = df_tabela.astype({"id_onibus": int, "id_linha": int,
                              "fl_sentido": int})

# Fazendo o merge para pegar a chave única do id_rota para a tabela:
df_rotas_right = df_rotas.copy()
df_rotas_right["id_novo"] = df_rotas_right.index
df_rotas_right

df_tabela = df_tabela.merge(df_rotas_right, how = "left",
                           on = ["id_linha", "fl_sentido"])

# Tratando e finalizando a tabela:
df_tabela = df_tabela[["id_onibus", "id_novo"]].rename(columns =
                                                    {"id_novo": "id_rota"})
df_tabela = df_tabela[df_tabela["id_rota"].notnull()]
df_tabela = df_tabela.reset_index()
df_tabela = df_tabela[["id_onibus", "id_rota"]].astype({"id_onibus": int,
                                                         "id_rota": int})
df_tabela.index.name = "id_tabela"
```

▼ Tabela 5: T_PARADAS

A tabela T_PARADAS simplesmente busca mostrar a posição geográfica de cada uma das paradas registradas no GTFS, para posterior uso.


```
# Importando a tabela os dados de paradas do GTFS:
caminho = '/content/drive/MyDrive/11. Joao Victor/Dados/GTFS/stops.txt'

df_paradas = pd.read_csv(caminho)

# Tratando e finalizando a tabela:
df_paradas = df_paradas[["stop_id", "stop_name", "stop_lat", "stop_lon"]]
df_paradas.rename(columns = {"stop_id": "id_parada", "stop_name": "ds_parada",
                             "stop_lat": "lat", "stop_lon": "long"}, inplace = True)
df_paradas.set_index("id_parada", inplace = True)
```

▼ Tabela 6: T_PONTO_DE_CONTROLE

A tabela T_PONTO_DE_CONTROLE tem como intuito designar cada parada que pertence a cada rota de ônibus. Por exemplo, quais as paradas que estão na linha 075 no sentido da ida? Para isso, filtraria-se por id_rota e se descobria os id_parada daquela rota.

Os dados de parada por rota advém dos arquivos de GTFS.

```
# Importando o arquivo de tempos de parada:
caminho = '/content/drive/MyDrive/11. Joao Victor/Dados/GTFS/stop_times.txt'

df_tempo_parada = pd.read_csv(caminho)

# Separando os dados de linha e sentido:
df_tempo_parada = df_tempo_parada[["trip_id", "stop_id", "stop_sequence"]]
df_tempo_parada["linha"] = df_tempo_parada["trip_id"].str[1:4]
df_tempo_parada["sentido"] = df_tempo_parada["trip_id"].str[-1:]

# Tratamento preliminar dos dados:
df_tempo_parada = df_tempo_parada.astype({"stop_id": int, "stop_sequence": int,
                                           "linha": int})
df_tempo_parada = df_tempo_parada.rename(columns = {"stop_id": "id_parada",
                                                    "stop_sequence": "ordem",
                                                    "linha": "id_linha",
                                                    "sentido": "fl_sentido"})
df_tempo_parada["fl_sentido"] = df_tempo_parada["fl_sentido"].map({"I": 0,
                                                                    "V": 1})

# Fazendo o join para trazer o id_rota:
df_tempo_parada = df_tempo_parada.merge(df_rotas_right, how = "left",
                                         on = ["id_linha", "fl_sentido"])
df_tempo_parada = df_tempo_parada[["trip_id", "id_parada", "ordem",
                                   "id_novo"]]
df_tempo_parada = df_tempo_parada.rename(columns = {"id_novo": "id_rota"})

# Descobrindo a trip_id com o maior número de paradas para cada id_rota:
```

```

idx = df_tempo_parada.groupby("id_rota")["ordem"].transform(max) == df_tempo_parada["ordem"]
df_trip_maior_paradas = df_tempo_parada[idx].groupby('id_rota').first().reset_index()[["id

# Tratando e finalizando a tabela:
df_ponto_de_controle = df_tempo_parada.merge(df_trip_maior_paradas,
                                             how = "inner",
                                             on = ["trip_id", "id_rota"])
df_ponto_de_controle = df_ponto_de_controle[["id_rota", "id_parada", "ordem"]]
df_ponto_de_controle.index.name = "id_ponto_de_controle"
df_ponto_de_controle = df_ponto_de_controle.rename(columns =
                                                    {"ordem": "id_ordem"})

```

▼ Tabela 7: T_POSICAO

A tabela T_POSICAO intuita armazenar todos os registros de posição de cada ônibus por rota. Busca-se, ainda trazer uma flag que determina se o ônibus está num ponto de controle ou não.

Para a determinação da flag, será checado se a posição daquele ônibus é a pelo menos 50 metros de distância de um dos seus pontos de controle mais próximos. Caso seja, considera-se que o ônibus está no ponto de controle. Caso contrário, não.

Essa flag é somente heurística, e será refinada posteriormente, de modo a ser possível determinar o momento exato em que um ônibus chega e sai de um ponto de controle, que será o conteúdo da Tabela 8.

```

# Importando os arquivos de GPS em um dataframe:
caminho = '/content/drive/MyDrive/11. Joao Victor/Dados/GPS'

arquivos = listdir(caminho)
arquivos.remove('NumLinhas-Veiculos.csv')
arquivos.remove('Readme.txt')

lista = []

for arquivo in arquivos:
    df = pd.read_csv(caminho + '/' + arquivo, index_col = None, header = 0,
                    parse_dates = ["hora"])
    lista.append(df)

df_gps = pd.concat(lista, axis = 0, ignore_index = True)

# Fazendo o join para determinar qual o id_onibus de cada registro:

# Tratamento inicial do dataframe:
df_gps.rename(columns = {"vehicleid": "id_veiculo", "lon": "long",
                        "hora": "dt_registro"}, inplace = True)

# Tratamento inicial do df_onibus:
df_onibus_right = df_onibus.copy()
df_onibus_right.reset_index(inplace = True)

```

```

# Join das tabelas:
df_gps = df_gps.merge(df_onibus_right, on = "id_veiculo", how = "inner")
df_gps = df_gps[["id_onibus", "dt_registro", "lat", "long"]]
df_gps = df_gps.sort_values(by = "dt_registro")

# Determinando qual a linha e sentido para cada registro:

# Para determinar a linha do ônibus, utilizar-se-á a tabela de bilhetagem.
# Se a bilhetagem mais próxima tiver sido de uma linha específica, assume-se
# que o ônibus está operando naquela linha e naquele sentido.

# Tratando a tabela de bilhetagem, visto que contem as mesmas informações:
df_bilhete = df_bilhetagem.copy()
df_bilhete = df_bilhete[["prefixo_carro", "hora", "lat", "lon", "linha",
                        "sentido_viagem"]]
df_bilhete = df_bilhete[df_bilhete["prefixo_carro"].notnull()]
df_bilhete = df_bilhete.rename(columns = {"prefixo_carro": "id_onibus",
                                          "hora": "dt_registro",
                                          "linha": "id_linha",
                                          "sentido_viagem": "fl_sentido",
                                          "lon": "long"})
df_bilhete = df_bilhete.astype({"id_onibus": int})
df_bilhete["fl_sentido"] = df_bilhete["fl_sentido"].map({"Ida": 0,
                                                         "Volta": 1})
df_bilhete["dt_registro"] = pd.to_datetime(df_bilhete["dt_registro"]).dt.tz_convert(None)
df_bilhete = df_bilhete.sort_values(by = "dt_registro")

# Inferindo o id_linha e fl_sentido a partir da bilhetagem mais próxima:
df_gps = pd.merge_asof(df_gps, df_bilhete, on = "dt_registro",
                      by = "id_onibus", direction = "nearest")
df_gps = df_gps[["id_onibus", "dt_registro", "lat_x",
                  "long_x", "id_linha", "fl_sentido"]]
df_gps = df_gps.rename(columns = {"lat_x": "lat",
                                  "long_x": "long"})

# Performando uma concatenação de ambas as tabelas:
df_posicao = pd.concat([df_gps, df_bilhete], ignore_index = True)
df_posicao = df_posicao[df_posicao["lat"].notnull() &
                      df_posicao["long"].notnull()]

# Conseguindo o id_rota para cada registro:
df_rotas_right = df_rotas.copy()
df_rotas_right = df_rotas_right.reset_index()

df_posicao = df_posicao.merge(df_rotas_right, how = "left",
                           on = ["id_linha", "fl_sentido"])
df_posicao = df_posicao[["id_onibus", "dt_registro", "lat", "long", "id_rota"]]

# Conseguindo o id_tabela para cada registro:
df_tabela_right = df_tabela.copy()
df_tabela_right = df_tabela_right.reset_index()

```

```

df_posicao = df_posicao.merge(df_tabela_right, how = "left",
                             on = ["id_onibus", "id_rota"])
df_posicao = df_posicao[["id_tabela", "dt_registro", "lat", "long"]]

# Tratando e finalizando a tabela:
df_posicao = df_posicao[df_posicao["id_tabela"].notnull()]
df_posicao = df_posicao.drop_duplicates()
df_posicao = df_posicao.reset_index()
df_posicao = df_posicao[["id_tabela", "dt_registro", "lat", "long"]]
df_posicao = df_posicao.astype({"id_tabela": int})
df_posicao.index.name = "id_posicao"

```

▼ Tabela 8: T_SHAPE

A tabela T_SHAPE busca armazenar os dados das linhas de ônibus de Fortaleza. Cada linha é composta por diversos pontos que, ao serem ligados sequencialmente, formam a linha no mapa.

Aqui serão tratados os dados para ficarem num formato adequado ao MySQL, contendo ID_SHAPE, ID_LINHA, FL_SENTIDO, SEQUENCIA, LAT e LONG.

```

# Importando os arquivos de GTFS em um dataframe:
caminho = '/content/drive/MyDrive/11. Joao Victor/Dados/GTFS/shapes.txt'

df_shape = pd.read_csv(caminho)

# Organizando as colunas e tratando:
df_shape["ID_LINHA"] = df_shape["shape_id"].str[5:8]
df_shape["FL_SENTIDO"] = df_shape["shape_id"].str[9].map({"I": 0,
                                                           "V": 1})

df_shape = df_shape.rename(columns = {"shape_pt_sequence": "SEQUENCIA",
                                     "shape_pt_lat": "LAT",
                                     "shape_pt_lon": "LONG"})

df_shape = df_shape.astype({"ID_LINHA": int})
df_shape = df_shape[["ID_LINHA", "FL_SENTIDO", "SEQUENCIA", "LAT", "LONG"]]
df_shape.index.name = "ID_SHAPE"

```

▼ Tabela 9: T_BILHETAGEM

A tabela T_BILHETAGEM busca armazenar as informações de bilhetagens que ocorreram nos ônibus de Fortaleza. Os dados que constam nela são: ID_BILHETAGEM, ID_USUARIO, ID_ONIBUS, DT_BILHETAGEM e TIPO_BILHETAGEM. Os dados podem ser usados para determinar a demanda, volume de passageiros, valor gerado em passagens, etc.

```

# Tratando os dados:

```

```

df_bilhetagem = df_bilhetagem[["id", "hora", "prefixo_carro", "tipo_cartao",
                                "integracao"]]
df_bilhetagem = df_bilhetagem[(df_bilhetagem["prefixo_carro"].notnull()) &
                                (df_bilhetagem["id"].notnull())]
df_bilhetagem = df_bilhetagem.rename(columns = {"id": "ID_USUARIO",
                                                "hora": "DT_BILHETAGEM",
                                                "prefixo_carro": "ID_ONIBUS",
                                                "tipo_cartao": "TIPO_BILHETAGEM",
                                                "integracao": "FL_INTEGRACAO"})
df_bilhetagem["FL_INTEGRACAO"] = df_bilhetagem["FL_INTEGRACAO"].map({"S": 1,
                                                                      "N": 0})
df_bilhetagem["DT_BILHETAGEM"] = pd.to_datetime(df_bilhetagem["DT_BILHETAGEM"]).dt.tz_conv
df_bilhetagem = df_bilhetagem.astype({"ID_USUARIO": int,
                                       "ID_ONIBUS": int})
df_bilhetagem.index.name = "ID_BILHETAGEM"

```

▼ Subindo as tabelas para o MySQL

A presente etapa busca subir os dataframes criados para tabelas em um servidor local do MySQL, onde será finalizado o tratamento dos dados e armazenados os dados finais.

Para isso, serão gerados .csv para cada tabela, e será subido manualmente no MySQL.

▼ Gerando arquivos .csv

```

# Gerando os arquivos das tabelas:

# Tabela 1: T_ONIBUS
df_onibus.to_csv('t_onibus.csv')
# Tabela 2: T_LINHA
df_linhas.to_csv('t_linha.csv')
# Tabela 3: T_ROTA
df_rotas.to_csv('t_rota.csv')
# Tabela 4: T_TABELA
df_tabela.to_csv('t_tabela.csv')
# Tabela 5: T_PARADA
df_paradas.to_csv('t_parada.csv')
# Tabela 6: T_PONTO_DE_CONTROLE
df_ponto_de_controle.to_csv('t_ponto_de_controle.csv')
# Tabela 7: T_POSICAO
df_posicao.to_csv('t_posicao.csv')
# Tabela 8: T_SHAPE
df_shape.to_csv('t_shape.csv')
# Tabela 8: T_BILHETAGEM
df_bilhetagem.to_csv('t_bilhetagem.csv')

# Importando a biblioteca de arquivos do Colab:
from google.colab import files

# Uploadando os arquivos para o Colab:

```

```
# importando os arquivos para a maquina:
lista = listdir()
lista_final = []
for arquivo in lista:
    if arquivo[-4:] == '.csv':
        lista_final.append(arquivo)

for arquivo in lista_final:
    files.download(arquivo)
```