



UFOP

Universidade Federal
de Ouro Preto

**Universidade Federal de Ouro Preto
Instituto de Ciências Exatas e Aplicadas
Departamento de Computação e Sistemas**

**DashMobi: Dashboard para
Visualização de Dados de Mobilidade
Urbana do Município de Belo
Horizonte**

Vinícius Correa Nobre Borges

**TRABALHO DE
CONCLUSÃO DE CURSO**

ORIENTAÇÃO:

Rafael Frederico Alexandre

COORIENTAÇÃO:

Tiago Franca Melo de Lima

**Setembro, 2025
João Monlevade–MG**

Vinícius Correa Nobre Borges

**DashMobi: Dashboard para Visualização de
Dados de Mobilidade Urbana do Município de
Belo Horizonte**

Orientador: Rafael Frederico Alexandre

Coorientador: Tiago Franca Melo de Lima

Monografia apresentada ao curso de Engenharia de Computação do Instituto de Ciências Exatas e Aplicadas, da Universidade Federal de Ouro Preto, como requisito parcial para aprovação na Disciplina “Trabalho de Conclusão de Curso II”.

Universidade Federal de Ouro Preto

João Monlevade

Setembro de 2025



FOLHA DE APROVAÇÃO

Vinícius Correa Nobre Borges

DashMobi: Dashboard para Visualização de Dados de Mobilidade Urbana do Município de Belo Horizonte

Monografia apresentada ao Curso de Engenharia da Computação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Engenharia da Computação

Aprovada em 05 de setembro de 2025

Membros da banca

Doutor - Rafael Frederico Alexandre - Orientador (Universidade Federal de Ouro Preto)
Doutor - Tiago Franca Melo de Lima - Coorientador - (Universidade Federal de Ouro Preto)
Doutor - Clodoveu Augusto Davis Junior - (Universidade Federal de Minas Gerais)
Doutora - Helen de Cassia Sousa da Costa Lima - (Universidade Federal de Ouro Preto)

Rafael Frederico Alexandre, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 25/09/2025



Documento assinado eletronicamente por **Rafael Frederico Alexandre, PROFESSOR DE MAGISTERIO SUPERIOR**, em 25/09/2025, às 11:58, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0984805** e o código CRC **80102930**.

Agradecimentos

A conclusão deste trabalho marca o fim de uma etapa importante em minha vida, que não teria sido possível trilhar sozinho. Expresso aqui minha mais profunda gratidão a todos que, de alguma forma, fizeram parte desta conquista.

Agradeço, em primeiro lugar, à minha família. À minha mãe, Thaís, e ao meu pai, Ricardo, pelo apoio incondicional, pelos valores ensinados e por terem me dado todas as condições necessárias para que esta realização se tornasse possível. Ao meu irmão, Daniel, pela amizade e cumplicidade. À minha namorada, Maria Eduarda, por sua parceria e pelo apoio em todas as situações.

Aos meus orientadores, Rafael e Tiago, pelos direcionamentos e orientação no desenvolvimento deste trabalho. Um agradecimento especial ao professor Rafael, que desde meus primeiros passos na UFOP, me ofereceu oportunidades e conselhos que se mostraram muito valiosos para a minha formação.

Por fim, agradeço aos meus colegas de curso, companheiros de uma longa jornada. As dificuldades foram muitas, mas as vitórias compartilhadas tornaram tudo mais gratificante.

A todos vocês, meu muito obrigado.

“The only way to do great work is to love what you do.”

— Steve Jobs (1955 – 2011),
in: Commencement Address at Stanford University, 2005.

Resumo

A complexidade da mobilidade urbana em grandes metrópoles como Belo Horizonte, aliada ao crescente volume de dados abertos, demanda ferramentas analíticas que transformem informações brutas em *insights* para a gestão pública e o controle social. Este trabalho teve como objetivo principal o desenvolvimento do *DashMobi*, um *dashboard* interativo para a visualização e análise de dados do transporte coletivo rodoviário do município. A metodologia adotada envolveu a construção de uma pipeline de Extração, Transformação e Carga (ETL) para o tratamento de um vasto conjunto de dados operacionais, abrangendo o período de 2015 a 2024. Os dados foram estruturados em um *data warehouse* com modelo estrela e otimizados para performance através de tabelas agregadas e *views* materializadas. A solução foi implementada seguindo uma arquitetura de três camadas, com um banco de dados PostgreSQL/PostGIS, uma API *RESTful* desenvolvida em Python com FastAPI, e uma interface web reativa construída com React. Como resultado, foi entregue uma ferramenta funcional que permite a exploração de dados através de módulos temáticos (Linhas, Ocorrências, Bairros) e de estudos de caso guiados, como a análise de eficiência das linhas e a distribuição de falhas mecânicas da frota. O *DashMobi* valida a eficácia da arquitetura proposta ao prover uma plataforma performática e intuitiva, contribuindo com um modelo de como os dados abertos podem ser utilizados para fomentar uma gestão mais eficiente e transparente da mobilidade urbana.

Palavras-chaves: Visualização de Dados. Mobilidade Urbana. *Dashboard*. Dados Abertos. *Business Intelligence*.

Abstract

The complexity of urban mobility in large metropolises such as Belo Horizonte, coupled with the growing volume of open data, demands analytical tools that transform raw information into insights for public management and social oversight. The main objective of this work was the development of DashMobi, an interactive dashboard for the visualization and analysis of the city's public bus transportation data. The methodology involved the construction of an Extract, Transform, Load (ETL) pipeline to process a vast set of operational data spanning from 2015 to 2024. The data was structured in a star-schema data warehouse and optimized for performance using aggregate tables and materialized views. The solution was implemented following a three-tier architecture, featuring a PostgreSQL/PostGIS database, a RESTful API developed in Python with FastAPI, and a reactive web interface built with React. The result is a functional tool that allows data exploration through thematic modules (Lines, Incidents, Neighborhoods) and guided case studies, such as the analysis of line efficiency and the distribution of fleet mechanical failures. DashMobi validates the effectiveness of the proposed architecture by providing a performant and intuitive platform, contributing a model for how open data can be leveraged to foster more efficient and transparent urban mobility management.

Key-words: Data Visualization. Urban Mobility. Dashboard. Open Data. Business Intelligence.

Lista de ilustrações

Figura 1 – Arquitetura de três camadas do sistema <i>DashMobi</i>	27
Figura 2 – Fluxograma da Pipeline de ETL para o dataset MCO Passageiros	29
Figura 3 – Fluxograma da Pipeline de ETL para o dataset de Pontos de Ônibus	29
Figura 4 – Fluxograma da Pipeline de ETL para o dataset de Bairros	30
Figura 5 – Fluxograma da arquitetura de dados, desde a origem até a modelagem do <i>Data Warehouse</i>	37
Figura 6 – Esboço de baixa fidelidade explorando os tipos de visualizações	41
Figura 7 – Esboço de média fidelidade dos componentes visuais	41
Figura 8 – Protótipo inicial do <i>dashboard</i>	42
Figura 9 – Protótipo final do <i>dashboard</i>	42
Figura 10 – Tela principal do <i>DashMobi</i> , apresentando a Visão Geral do sistema	46
Figura 11 – Tela de análise geral do módulo “Linhas”	48
Figura 12 – Modal de seleção para a Análise Individual de Linha	49
Figura 13 – Tela de análise individual do módulo “Linhas”	49
Figura 14 – Visão geral da página do estudo de caso “Eficiência das Linhas”, destacando os KPIs, a pergunta-chave e as métricas utilizadas	50
Figura 15 – Gráfico de Análise de Quadrantes, classificando as linhas por eficiência espacial e temporal	51
Figura 16 – <i>Cards</i> interpretativos de cada quadrante	52
Figura 17 – Recursos do <i>dashboard</i> para exibir as informações das linhas	53
Figura 18 – Rankings de “Destaques da Análise”, que sumarizam os principais achados do estudo de caso	53
Figura 19 – Etapa 1 da análise, mostrando o ranking de empresas por taxa de falhas e a distribuição dos tipos de falha	54
Figura 20 – Etapa 2 da análise, com o gráfico de bolhas correlacionando quilometragem, falhas e idade dos veículos	55
Figura 21 – Painel de filtros interativos do gráfico de bolhas da Etapa 2	55
Figura 22 – Etapa 3, com o ranking de linhas mais afetadas, e a seção final de “Destaques da Análise”	56
Figura 23 – Visão geral da página do estudo de caso “Horários de Utilização”, com os filtros de análise e os KPIs de horários de pico	57
Figura 24 – Gráficos de tendência temporal para o volume de passageiros e ocorrências ao longo do dia	58
Figura 25 – Gráficos de tendência temporal para a duração média das viagens ao longo do dia	58

Lista de tabelas

Tabela 1 – Comparação entre Portais de Dados Abertos	20
Tabela 2 – Quadro Comparativo de Soluções de Visualização de Mobilidade	25
Tabela 3 – Dicionário dos dados: <i>Dataset</i> MCO	32
Tabela 4 – Dicionário dos dados: <i>Dataset</i> de pontos de ônibus	33
Tabela 5 – Dicionário dos dados: <i>Dataset</i> de bairros	33

Lista de abreviaturas e siglas

API	Application Programming Interface
ETL	Extract, Transform, Load
GTFS	General Transit Feed Specification
JSON	JavaScript Object Notation
KPI	Key Performance Indicators
MaaS	Mobility-as-a-Service
MCO	Mapa de Controle Operacional
SPA	Single-Page Application
UFOP	Universidade Federal de Ouro Preto

Sumário

1	INTRODUÇÃO	13
1.1	Justificativa	14
1.2	Objetivos	15
1.3	Metodologia	16
1.4	Organização do trabalho	17
2	REVISÃO BIBLIOGRÁFICA	18
2.1	Visualização de Dados e Dashboards Analíticos	18
2.2	Cenário dos Dados Abertos de Mobilidade no Brasil	19
2.2.1	A Política de Dados Abertos e o Portal de Belo Horizonte	19
2.2.2	Análise Comparativa de Portais de Dados	20
2.3	Trabalhos Relacionados	22
2.3.1	Análise de Ferramentas Existentes	22
2.3.2	Comparativo com Outras Soluções	24
3	DESENVOLVIMENTO	27
3.1	Arquitetura da Solução	27
3.1.1	Visão Geral da Arquitetura	27
3.1.2	Pipeline dos Dados	28
3.2	Construção do Ambiente de Dados	30
3.2.1	Fonte de Dados	30
3.2.2	Pré-Processamento e Carga (ETL)	33
3.2.2.1	Procedimentos Comuns	34
3.2.2.2	<i>Dataset</i> MCO	34
3.2.2.3	<i>Dataset</i> de Pontos de Ônibus	35
3.2.2.4	<i>Dataset</i> de Bairros	36
3.2.3	Modelagem do Banco de Dados Analítico	36
3.2.3.1	Tabelas de Dimensão	37
3.2.3.2	Tabela de Fatos	38
3.2.3.3	Estratégias de Carga e Otimização	39
3.2.3.4	Tabelas Agregadas e Views Materializadas	39
3.3	Projeto da Interface e Experiência do Usuário	40
3.3.1	Prototipação da Interface	40
3.3.2	Módulos de Visualização e Análise	43
3.3.3	Estudos de Caso Guiados	44

4	RESULTADOS	46
4.1	Apresentação da Ferramenta <i>DashMobi</i>	46
4.1.1	Componentes Visuais do Dashboard	47
4.1.2	Apresentação dos Módulos Temáticos	48
4.2	Análise dos Dados e Validação dos Estudos de Caso	49
4.2.1	Estudo de Caso: Eficiência das Linhas	50
4.2.2	Estudo de Caso: Distribuição das Falhas Mecânicas	54
4.2.3	Estudo de Caso: Horários de Utilização	56
4.3	Validação da Solução	58
4.3.1	Performance da API	59
4.3.2	Usabilidade da Interface	59
4.3.3	Atendimento aos Objetivos do Projeto	60
5	CONCLUSÃO	61
5.1	Trabalhos Futuros	62
	REFERÊNCIAS	64

1 Introdução

O processo de expansão dos centros urbanos brasileiros ocorreu de forma acelerada e desorganizada, com o aumento desenfreado da população nas cidades, que buscavam melhores condições de vida (PAULO, 2018). Como consequência deste processo desordenado, o crescimento populacional nas cidades não foi acompanhado pela infraestrutura urbana necessária para atender todos os cidadãos, afetando negativamente, em diferentes áreas, uma parcela considerável da sociedade (ANDRADE; GALVÃO et al., 2016).

A mobilidade urbana, em particular, emergiu como um dos maiores desafios enfrentados por grandes centros urbanos como Belo Horizonte, que lida diariamente com a complexa tarefa de oferecer um transporte coletivo eficiente, acessível e alinhado às demandas da população. Diante dessa complexidade, torna-se essencial ampliar a atenção dedicada à identificação de problemas e à implementação de soluções eficazes (CARVALHO, 2016). Mais especificamente, os desafios de mobilidade urbana no Brasil apresentam características particulares, visto que estudos indicam que as consequências dos problemas de transporte não se distribuem uniformemente entre a população, afetando especialmente grupos mais vulneráveis (PERO; STEFANELLI, 2015). A qualidade do transporte público em cidades brasileiras tem sido caracterizada como mediana ou péssima em diversos atributos, evidenciando a necessidade de ferramentas que auxiliem na identificação de problemas específicos e no monitoramento contínuo da qualidade dos serviços (SUCENA; CURY, 2022).

Diante desse complexo cenário, duas grandes transformações criaram um novo paradigma para a gestão urbana. A primeira é a evolução das políticas de transparência, como a Lei de Acesso à Informação (LAI) e a Política de Dados Abertos, que tornaram público um volume sem precedentes de dados governamentais (BRAGA et al., 2008). A segunda é a ascensão do conceito de cidades inteligentes (*smart cities*), que se baseiam em tecnologias emergentes para otimizar recursos e compreender os fluxos urbanos a partir do crescimento exponencial de dados gerados por sensores e sistemas de transporte (MARTINS et al., 2022; BONETTE; REIS, 2021).

Essa confluência entre a disponibilidade de dados e a necessidade de soluções mais inteligentes abriu caminho para uma gestão pública mais moderna. Dessa forma, a capacidade de transformar grandes volumes de dados em informações estratégicas tem se mostrado fundamental para auxiliar gestores urbanos na tomada de decisões mais precisas e baseadas em evidências (FERREIRA; FERRETTO; DUARTE, 2023). Ferramentas de análise de dados e *Business Intelligence* permitem não apenas identificar padrões e tendências nos sistemas de transporte, mas também prever cenários futuros e avaliar o

impacto de políticas públicas.

Nesse contexto, o uso de tecnologias que consolidam e apresentam informações complexas de maneira gráfica e intuitiva surge como uma solução promissora. Ferramentas como *dashboards* interativos possibilitam a transformação de grandes volumes de dados brutos em informações compreensíveis (FEW, 2006). Além dos benefícios para a gestão, a disponibilização desses dados de forma acessível fortalece a participação e o controle social. À vista disso, a democratização do acesso à informação, facilitada por interfaces intuitivas, permite que cidadãos, pesquisadores e organizações da sociedade civil acompanhem e avaliem a qualidade dos serviços de transporte público (CORDEIRO et al., 2012; MEIRELES et al., 2018).

Diante do exposto, o desenvolvimento de um *dashboard* para visualização de dados de mobilidade urbana de Belo Horizonte representa uma contribuição significativa para a modernização da gestão pública municipal. A aplicação de tecnologias como sistemas de transporte inteligentes (STIs) e análise de dados ao contexto específico do transporte público rodoviário de Belo Horizonte pode não apenas otimizar o sistema local, mas também servir como modelo para outras cidades brasileiras que enfrentam desafios similares (SOUZA; VILLAS, 2017; SANTOS; VASCONCELOS,).

1.1 Justificativa

A crescente disponibilidade de dados sobre mobilidade urbana, embora positiva, impõe um desafio substancial: a sua complexidade e volume bruto representam barreiras para a geração de conhecimento que possa dar suporte à tomada de decisão. Por isso, formatos tradicionais como planilhas e arquivos CSV, apesar de conterem um alto grau de detalhamento, são insuficientes para a identificação intuitiva de padrões, a correlação entre múltiplas variáveis e a visualização de tendências geoespaciais e temporais. É precisamente para superar essa lacuna entre dados brutos e inteligência estratégica que o presente trabalho se justifica.

A relevância do *DashMobi* reside, portanto, em sua capacidade de traduzir esse grande volume de dados em uma ferramenta analítica interativa e acessível. A pertinência de tal solução se manifesta em múltiplas dimensões, atendendo a diferentes atores do ecossistema urbano:

- **Para Gestores Públicos:** Oferece uma visão macro e micro do sistema, permitindo a identificação de linhas com baixa eficiência, a análise de concentração de ocorrências e o monitoramento de indicadores-chave de desempenho. Tais funcionalidades são essenciais para embasar decisões estratégicas sobre alocação de recursos, otimização de rotas e planejamento operacional.

- **Para Planejadores Urbanos e Pesquisadores:** Possibilita a correlação entre dados operacionais do transporte e variáveis socioeconômicas e demográficas. A ferramenta permite investigar questões críticas sobre equidade no serviço, como a correspondência entre a densidade populacional de uma região e a qualidade da oferta de transporte, ou a distribuição do tempo de viagem entre diferentes estratos sociais.
- **Para Cidadãos e Sociedade Civil:** Promove a transparência e fortalece o controle social ao permitir que a população monitore a qualidade do serviço, compreenda o desempenho das linhas que utiliza diariamente e participe de forma qualificada nos debates sobre a mobilidade urbana, munida de evidências concretas.

Dessa forma, este trabalho transcende a mera criação de um visualizador de dados. A proposta é desenvolver um observatório de mobilidade que emprega o conceito de *Data Storytelling* como pilar central. Ao aplicar técnicas de modelagem de *Business Intelligence* e projetar visualizações que constroem narrativas lógicas e compreensíveis, o *DashMobi* busca transformar dados em conhecimento acionável, contribuindo para uma gestão mais eficiente, equitativa e transparente do transporte coletivo em Belo Horizonte.

1.2 Objetivos

O objetivo geral deste trabalho é projetar e desenvolver o *DashMobi*, uma plataforma de visualização de dados em formato de *dashboard*, concebida para transformar dados brutos do transporte coletivo rodoviário de Belo Horizonte em informações estratégicas, apoiando tanto a gestão pública quanto o controle social.

Para alcançar esse propósito, foram definidos os seguintes objetivos específicos:

- Modelar e implementar um banco de dados analítico em modelo estrela (*star schema*), projetado para otimizar o desempenho de consultas complexas sobre grandes volumes de dados de mobilidade.
- Desenvolver uma API (*Application Programming Interface*) no *back-end* para realizar o processamento, a agregação e a disponibilização segura e escalável dos dados, servindo como a ponte entre o banco de dados e a interface do usuário.
- Construir uma interface de usuário (*UI*) responsiva e intuitiva, aplicando princípios de Visualização de Informação para apresentar os dados por meio de componentes interativos, como gráficos, tabelas dinâmicas e mapas geoespaciais.
- Implementar funcionalidades de análise que permitam ao usuário a exploração interativa de indicadores-chave, como frequência e ocupação das linhas, padrões de

ocorrências e a correlação entre a oferta de transporte e as características demográficas dos bairros.

1.3 Metodologia

A metodologia adotada para o desenvolvimento deste trabalho foi estruturada em cinco fases sequenciais, projetadas para garantir o rigor técnico e a relevância analítica do resultado final. Essas fases são detalhadas a seguir:

1. **Levantamento do Estado da Arte:** Realização de uma revisão sistemática da literatura, com a análise crítica de artigos científicos, trabalhos acadêmicos e soluções de mercado correlatas. O objetivo desta fase foi identificar as melhores práticas em visualização de dados para mobilidade urbana, compreender as lacunas existentes e embasar as decisões de arquitetura e design da solução proposta.
2. **Coleta e Modelagem dos Dados:** Esta fase compreendeu a identificação, a seleção e o tratamento dos conjuntos de dados públicos pertinentes. Foi desenvolvida uma pipeline de Extração, Transformação e Carga (ETL) para automatizar a limpeza, a padronização e a integração dos dados. Subsequentemente, foi projetado e implementado o banco de dados analítico, utilizando a modelagem em esquema estrela para garantir a performance nas consultas, com a definição de tabelas fato, dimensão e pontes.
3. **Desenvolvimento da Solução Tecnológica:** Consistiu na implementação dos componentes de software do *DashMobi*. O desenvolvimento do *back-end* focou na construção de uma API com *endpoints* otimizados para as consultas analíticas. O desenvolvimento do *front-end* se concentrou na criação da interface do usuário (UI), na implementação dos componentes visuais (gráficos, tabelas, mapas) e na codificação das funcionalidades de interatividade para a exploração dos dados.
4. **Validação e Testes:** Execução de um processo de verificação em duas frentes. A primeira, a **validação de dados**, focou em assegurar a consistência, a acurácia e a integridade das informações após o processo de ETL e a carga no banco de dados. A segunda, os **testes de software**, envolveu testes funcionais e de integração para garantir que as visualizações, os filtros e as interações da aplicação se comportassem conforme o especificado e refletissem com precisão os dados consultados.
5. **Análise dos Resultados e Discussão:** Utilização da plataforma *DashMobi* como instrumento de análise para explorar os estudos de caso definidos. A discussão dos resultados foi centrada na análise crítica dos *insights* gerados pelas visualizações, correlacionando-os com os desafios da mobilidade urbana discutidos na literatura e avaliando a aderência da ferramenta aos objetivos propostos.

1.4 Organização do trabalho

Para além deste capítulo introdutório, a estrutura deste documento está organizada como explicado abaixo.

O Capítulo 2 é dedicado à revisão da literatura, onde se examinam criticamente os trabalhos académicos e as soluções de mercado relacionados à visualização de dados para mobilidade urbana, fundamentando o estado da arte da área.

O Capítulo 3 detalha a concepção e o desenvolvimento da plataforma *DashMobi*. Nesta seção são abordados todos os aspectos técnicos do projeto, incluindo a arquitetura da solução, a modelagem do banco de dados, as tecnologias empregadas e os processos de implementação do *back-end* e do *front-end*.

O Capítulo 4 apresenta os resultados e as análises obtidas. Por meio de estudos de caso, esta seção demonstra a aplicação prática da ferramenta, discutindo os *insights* gerados pelas visualizações e validando a capacidade do *dashboard* em responder a questões relevantes sobre o transporte coletivo.

Finalmente, o Capítulo 5 sintetiza as conclusões do trabalho, reitera suas principais contribuições para a área, discute as limitações do estudo e propõe direções para trabalhos futuros, visando a continuidade e a expansão da pesquisa.

2 Revisão bibliográfica

Este capítulo estabelece a fundamentação teórica que sustenta o presente trabalho. Para isso, são explorados os conceitos centrais de Visualização de Dados e *Dashboards* Analíticos, as tecnologias aplicadas no desenvolvimento da solução e, por fim, uma análise de trabalhos correlatos, situando o *DashMobi* no panorama atual da área.

2.1 Visualização de Dados e Dashboards Analíticos

A Visualização de Dados é considerada por muitas áreas como a versão contemporânea da comunicação visual. Essa disciplina abrange tanto a criação quanto o estudo de representações gráficas de dados, que consistem na tradução de informações abstratas em um formato esquemático, codificando visualmente os atributos e as variáveis de um conjunto de dados (CARD, 2009).

O principal objetivo da Visualização de Dados é transmitir informações de forma clara e eficiente por meio de gráficos estatísticos, infográficos, tabelas e representações visuais selecionadas. Uma visualização eficaz auxilia os usuários na análise e interpretação dos dados, facilitando o raciocínio e a tomada de decisões. Além de tornar conjuntos de dados complexos mais acessíveis e compreensíveis, ela permite a realização de tarefas analíticas, como comparações, identificação de padrões e compreensão das relações entre os dados (SATO, 2017).

Dentro deste campo, os *dashboards* consolidaram-se como ferramentas relevantes para a exploração de dados e a resolução de problemas em diversas áreas, incluindo *business intelligence*, educação e, de forma crescente, no planejamento urbano. Sua função é converter dados complexos em painéis visuais interativos e compreensíveis, aprimorando a capacidade do usuário de extrair *insights* e tomar decisões informadas (SANTORO; SPIEGEL; JÚNIOR, 2024).

Segundo Few (2006), um *dashboard* é uma apresentação visual das informações mais importantes, necessárias para alcançar um ou mais objetivos de negócios. Essas informações são consolidadas e ajustadas em uma única tela para que possam ser monitoradas de forma rápida. O autor propõe uma analogia com o painel de um carro, que em um único e limitado espaço são disponibilizadas informações essenciais para o condutor, que tem a seu dispor indicadores, gráficos e ponteiros, facilitando uma análise prática e contribuindo com a direção.

2.2 Cenário dos Dados Abertos de Mobilidade no Brasil

A eficácia de dashboards analíticos, como o proposto neste trabalho, depende fundamentalmente da disponibilidade e da qualidade dos dados. Nesse sentido, esta seção explora o panorama dos dados abertos de mobilidade no Brasil, com foco na política que os rege e na infraestrutura que os disponibiliza, contextualizando a escolha de Belo Horizonte como objeto de estudo.

2.2.1 A Política de Dados Abertos e o Portal de Belo Horizonte

A disponibilização de dados governamentais no Brasil é um pilar da modernização da administração pública, impulsionada por marcos legais como a Lei de Acesso à Informação (LAI), Lei nº 12.527/2011, e a Política Nacional de Dados Abertos (Decreto nº 8.777/2016). Essa legislação estabelece que as informações produzidas pelo setor público devem ser transparentes, acessíveis e gratuitas. O principal instrumento para a efetivação dessa política são os Portais de Dados Abertos: plataformas digitais onde os dados são publicados em formatos padronizados, não proprietários e processáveis por máquina, fomentando a inovação e a participação cidadã (FERRER-SAPENA; PESET; ALEIXANDRE-BENAVENT, 2011).

Esses portais funcionam como repositórios centrais de dados brutos, abrangendo áreas estratégicas como saúde, educação, orçamento e, crucialmente para este trabalho, o transporte. A premissa é que, ao oferecer os dados sem interpretações prévias, permite-se que pesquisadores, desenvolvedores, jornalistas e a sociedade civil os utilizem para análises independentes, fiscalização e criação de novos serviços (DINIZ, 2010).

No contexto da mobilidade urbana, a abertura de dados operacionais do transporte público é um passo fundamental. Informações sobre itinerários, horários de operação, localização de veículos em tempo real e dados de bilhetagem eletrônica, quando disponibilizadas, permitem a realização de estudos aprofundados sobre a eficiência da rede, a criação de aplicativos de rota e o desenvolvimento de políticas de transporte mais equitativas.

Alinhada a essa diretriz nacional, a Prefeitura de Belo Horizonte mantém o seu Portal de Dados Abertos (dados.pbh.gov.br). Esta plataforma centraliza e disponibiliza um vasto conjunto de informações sobre o município, incluindo dados essenciais do sistema de transporte coletivo. A existência desse portal é, portanto, a condição fundamental que viabiliza o desenvolvimento de ferramentas analíticas como o *DashMobi*, que se propõe a consumir, processar e transformar esses dados brutos em conhecimento estratégico. A importância de tais portais transcende a transparência, atuando como catalisadores para a inovação e como infraestrutura basilar para a construção de cidades inteligentes, onde a gestão pública se torna mais orientada por dados e a sociedade civil participa ativamente na solução de problemas urbanos.

2.2.2 Análise Comparativa de Portais de Dados

Embora a Política de Dados Abertos seja uma diretriz nacional, a sua implementação e, principalmente, a disponibilidade e a qualidade dos dados de mobilidade urbana variam significativamente entre os municípios brasileiros. Para contextualizar a escolha de Belo Horizonte como objeto de estudo deste trabalho e justificar a viabilidade das análises propostas, foi realizada uma análise comparativa de seu portal de dados com os de outras três grandes capitais brasileiras, reconhecidas pela complexidade de seus sistemas de transporte: Curitiba, Rio de Janeiro e São Paulo.

Tabela 1 – Comparação entre Portais de Dados Abertos

Aspecto Comparado	B. Horizonte	São Paulo	R. de Janeiro	Curitiba
Dados Operacionais	MCO detalhado de viagens (2015-2024)	API em tempo real, sem histórico	GPS em tempo real, dados de trajeto	Acesso controlado, dados básicos
Granularidade	Muito alta: viagem individual com 18 atributos	Média: posição atual dos veículos	Alta: GPS por minuto, trajetos	Baixa: linhas e pontos
Dados Geográficos	Excelente: pontos e bairros georreferenciados	Básico: coordenadas dos veículos	Muito bom: GPS e pontos de parada	Limitado: latitude/longitude básica
Frequência de Atualização	Mensal para históricos, alguns em tempo real	Tempo real para posições	GPS atualizado a cada minuto	Semanal, via GTFS
Formato e Qualidade	Excelentes formatos, documentação completa	Boa API JSON, foco em tempo real	Bom, inconsistências pontuais	Médio, documentação limitada
Documentação	Muito boa, dicionários completos	Boa com exemplos de uso	Regular, metadados incompletos	Limitada, acesso burocrático

A avaliação foi pautada em critérios essenciais para o desenvolvimento de uma ferramenta de *business intelligence* como o *DashMobi*, focando não apenas na existência dos dados, mas na sua granularidade, acessibilidade e adequação para análises operacionais e geográficas profundas. Os aspectos avaliados incluem a disponibilidade de históricos de viagens, a riqueza de atributos em cada registro, a qualidade dos dados geográficos e a clareza da documentação. A tabela 1 resume os resultados desta análise, evidenciando as

particularidades de cada portal.

A análise comparativa revela diferenças significativas na abordagem e qualidade dos dados abertos de mobilidade entre as quatro capitais estudadas. Belo Horizonte se destaca como referência nacional na abertura de dados de transporte público, como confirmado pelo primeiro lugar no ranking do Índice de Dados Abertos para Cidades (ODI) de 2023 ([Prefeitura de Belo Horizonte, 2024](#)). Sua principal vantagem reside na disponibilização do Mapa de Controle Operacional (MCO) Consolidado, um dataset excepcionalmente rico que documenta cada viagem individual realizada no sistema desde 2015, com 18 atributos detalhados incluindo dados de passageiros, ocorrências operacionais e falhas mecânicas.

Com relação à São Paulo, apesar de ser a maior metrópole do país, apresenta uma abordagem focada exclusivamente no tempo real através da API Olho Vivo. Embora tecnicamente robusta para monitoramento instantâneo da frota, essa estratégia limita severamente as possibilidades de análises históricas e identificação de padrões operacionais de longo prazo. A ausência de dados históricos granulares compromete estudos de eficiência, planejamento de rotas e avaliação de políticas públicas de mobilidade ([São Paulo Transporte S/A, 2024](#)).

O Rio de Janeiro ocupa uma posição intermediária, oferecendo dados GPS detalhados atualizados por minuto através do portal “data.rio”. A cidade pioneira em iniciativas de dados abertos no Brasil possui um volume considerável de informações, mas enfrenta problemas de consistência e atualização. Alguns *datasets* importantes, como os pontos de parada das linhas, não recebem atualizações regulares, comprometendo a confiabilidade das análises. Além disso, a qualidade da documentação é irregular, dificultando o uso efetivo dos dados por desenvolvedores e pesquisadores ([Governo do Estado do Rio de Janeiro, 2024](#)).

Curitiba, tradicionalmente reconhecida como modelo em transporte público urbano, surpreende negativamente no quesito transparência de dados. O município adota uma política restritiva que exige cadastro formal e aprovação para acesso aos dados operacionais. Embora disponibilize GTFS e alguns dados básicos, a ausência de dados operacionais detalhados e o processo burocrático para obtenção de informações mais granulares representam uma barreira significativa para análises aprofundadas e controle social ([Prefeitura Municipal de Curitiba, 2024](#)).

Esta disparidade reflete diferentes filosofias de transparência e gestão pública. Enquanto Belo Horizonte adota uma abordagem proativa de abertura de dados, priorizando a granularidade e a qualidade das informações, outras capitais ainda enfrentam desafios técnicos, políticos ou organizacionais para implementar uma política robusta de dados abertos. A liderança de Belo Horizonte neste aspecto não apenas facilita o controle social e a pesquisa acadêmica, mas também cria um ambiente propício para o desenvolvimento de soluções inovadoras em mobilidade urbana, como o próprio *DashMobi* desenvolvido

neste trabalho.

2.3 Trabalhos Relacionados

Com o crescente volume de dados abertos de mobilidade, diversas iniciativas governamentais e acadêmicas têm surgido para analisá-los e visualizá-los. A seguir, são apresentadas algumas das principais ferramentas existentes, estabelecendo um comparativo que contextualiza a contribuição e a originalidade da solução DashMobi.

2.3.1 Análise de Ferramentas Existentes

Para contextualizar a contribuição do *DashMobi*, analisaram-se ferramentas de visualização de dados de mobilidade urbana de referência no Brasil. Dentre as plataformas governamentais e acadêmicas, destacam-se duas iniciativas principais: o Painel do Sistema Nacional de Informações em Mobilidade Urbana (SIMU), mantido pelo Governo Federal, e o MobiliDADOS, desenvolvido pelo Instituto de Políticas de Transporte e Desenvolvimento (ITDP Brasil).

O Painel SIMU tem como objetivo monitorar a Política Nacional de Mobilidade Urbana (PNMU) por meio de indicadores de alto nível, oferecendo um panorama sobre temas como transporte público coletivo, mobilidade ativa, acessibilidade e desenvolvimento urbano em escala nacional, regional e, em alguns casos, municipal. A plataforma agrega dados de diversas fontes, como a Pesquisa de Informações Básicas Municipais (MUNIC) do IBGE e o Sistema de Informações da Mobilidade Urbana da ANTP, para construir seus indicadores ([Ministério das Cidades, Governo Federal do Brasil, 2024](#)).

O MobiliDADOS, por sua vez, é uma iniciativa do ITDP Brasil que disponibiliza indicadores e dados abertos para apoiar a elaboração e monitoramento de políticas públicas de mobilidade urbana. A plataforma abrange as 27 capitais e 9 maiores regiões metropolitanas do país, disponibilizando dados sobre divisão modal, infraestrutura de transporte, acessibilidade, segurança viária e aspectos socioeconômicos da mobilidade ([União de Ciclistas do Brasil \(UCB\), 2024](#)).

A ferramenta organiza seus dados em categorias como taxa de motorização, tempo médio de deslocamento casa-trabalho, percentual da população próxima às estações de transporte de média e alta capacidade, emissões de CO₂ por habitante, e indicadores de segurança no trânsito segmentados por modalidade (pedestres, ciclistas, motociclistas). Um diferencial da plataforma é sua abordagem de equidade, apresentando dados desagregados por faixas de renda e recortes de gênero e raça, permitindo análises sobre disparidades no acesso ao transporte.

Apesar da inegável relevância dessas plataformas para o monitoramento de políticas

públicas em escala macro, uma análise aprofundada dessas plataformas revela lacunas significativas quando o objetivo é a análise operacional e detalhada de um sistema de transporte municipal específico, como o de Belo Horizonte. As principais limitações identificadas, que justificam a necessidade de uma ferramenta como o *DashMobi*, são:

- **Granularidade Baixa dos Dados:** Tanto o SIMU quanto o MobiliDADOS trabalham com dados altamente agregados. O SIMU apresenta indicadores consolidados anualmente em níveis estaduais ou para grandes agrupamentos de municípios, enquanto o MobiliDADOS oferece indicadores metropolitanos e por capitais, sem granularidade temporal ou espacial suficiente para análises operacionais. Ambas as plataformas carecem da granularidade diária, horária e por viagem que é essencial para entender a dinâmica operacional do transporte, como os padrões de horários de pico ou a performance de linhas individuais.
- **Natureza dos Dados:** Nenhuma das plataformas provê acesso a dados operacionais brutos, como o MCO de Belo Horizonte. Consequentemente, é impossível analisar indicadores de desempenho como passageiros por viagem, quilometragem por veículo, ou a eficiência financeira de uma rota específica. A análise se restringe a indicadores de resultado de políticas, não de processo operacional.
- **Foco em Política vs. Operação:** O design das ferramentas é orientado à verificação de metas de políticas públicas (ex: cobertura da rede, emissões de CO₂, mortalidade no trânsito). Elas respondem se uma cidade está alinhada a diretrizes nacionais de sustentabilidade e acesso, mas não se o sistema de transporte está funcionando de forma eficiente no dia a dia.
- **Interatividade para Análise Local:** Embora interativas, as plataformas não foram projetadas para o tipo de análise exploratória e de *drill-down* que o *DashMobi* propõe. Um gestor de Belo Horizonte não poderia utilizá-las para identificar qual linha específica sofre mais com falhas mecânicas ou qual bairro apresenta a maior defasagem entre oferta e demanda.
- **Defasagem Temporal:** Os dados utilizados, especialmente em plataformas que dependem de pesquisas censitárias, possuem um hiato de atualização significativo (em alguns casos, de vários anos). O *DashMobi*, ao contrário, utiliza dados operacionais com atualização mensal, permitindo análises sobre o comportamento recente da mobilidade urbana.

Em suma, o Painel SIMU e o MobiliDADOS são observatórios de políticas públicas de mobilidade, cumprindo o papel de monitoramento em escala nacional e metropolitana. Contudo, eles não atendem à demanda por uma ferramenta de *business intelligence* voltada

para a gestão operacional em nível municipal. É precisamente esta lacuna—entre a análise macro de políticas e a inteligência micro de operações—que o *DashMobi* se propõe a preencher, transformando o rico volume de dados operacionais de Belo Horizonte em conhecimento acionável para a otimização do sistema de transporte local.

2.3.2 Comparativo com Outras Soluções

Para contextualizar a contribuição e a originalidade do *DashMobi*, é fundamental compará-lo com outras soluções acadêmicas e aplicadas que também exploram a visualização de dados de mobilidade urbana. A análise a seguir aborda quatro trabalhos — MetroViz, BusInRio, Dashboard Salvador e UrbanFlow Milano — que, embora compartilhem o objetivo geral de analisar o transporte público, diferem significativamente em escopo, tecnologia e profundidade analítica.

O MetroViz (DU et al., 2015), desenvolvido para a cidade de Blacksburg (EUA), foca na análise de duas métricas operacionais principais: a adesão aos horários e o volume de passageiros (*ridership*). Sua principal contribuição é um modelo de dados de três níveis (parada, rota e viagem) e uma interface que combina visualizações geográficas e temporais, como o “*calendar view*”, para explorar dados históricos. Embora eficaz para a análise da qualidade do serviço em nível de rota, o MetroViz não explora a integração com dados demográficos nem adota uma arquitetura de *Business Intelligence* robusta para consultas complexas.

O BusInRio (CRUZ; ANDRADE; SAMPAIO, 2016) diferencia-se por sua ênfase no uso de dados abertos em tempo real, provenientes dos GPS dos ônibus do Rio de Janeiro. Sua arquitetura distribuída e baseada em nuvem visa fornecer ao usuário final informações dinâmicas sobre a localização dos veículos e as condições do trânsito. O foco da solução é o monitoramento operacional instantâneo, sendo menos voltado para a análise histórica aprofundada, modelagem de dados analíticos ou a correlação com fatores socioeconômicos do espaço urbano.

O Dashboard para Dados sobre a Mobilidade Urbana em Salvador (FILHO; FLORENTINO, 2022) representa uma iniciativa de análise exploratória a partir de fontes de dados não estruturadas, como os Anuários de Transportes em formato PDF. O trabalho demonstra o desafio e o valor da engenharia de dados para extrair e organizar informações históricas sobre a frota, o número de passageiros e a relação com indicadores de renda. No entanto, sua natureza é mais descritiva e limitada pela baixa frequência e agregação dos dados de origem, sem a granularidade necessária para uma análise operacional detalhada.

Por fim, o UrbanFlow Milano (DELFINI; SPAHIU; VIZZARI, 2024) é uma ferramenta moderna focada na análise de dados de mobilidade compartilhada (*Mobility-as-a-Service* - MaaS) em Milão, Itália. Destaca-se pelo uso de técnicas de visualização

avançadas, como mapas de fluxo Origem-Destino (OD), fluxos de trajetória e diagramas de corda, e por uma forte ênfase no design centrado no usuário. Seu escopo, contudo, é específico para um novo paradigma de mobilidade, não abordando o transporte público coletivo tradicional com a mesma profundidade operacional e demográfica que o *DashMobi* propõe.

Tabela 2 – Quadro Comparativo de Soluções de Visualização de Mobilidade

Característica	DashMobi (2025)	MetroViz (2015)	BusInRio (2017)	Dashboard Salvador (2022)	UrbanFlow (2024)
Escopo de Análise	Operacional, geográfico e demográfico integrado	Foco em adesão a horários e volume de passageiros	Monitoramento em tempo real da frota e trânsito	Análise histórica de relatórios anuais	Foco em mobilidade compartilhada (MaaS)
Granularidade Geográfica	Nível de bairro, linha e ponto de ônibus	Nível de rota e ponto de ônibus	Localização do veículo em tempo real	Agregado por município	Nível de distrito (NIL) e trajetórias individuais
Modelagem de Dados	Data Warehouse com Modelo Estrela e agregações	Modelo conceitual de 3 níveis (parada, rota, viagem)	Bases NoSQL para dados em tempo real	Arquivos CSV extraídos de PDFs	Foco na visualização de fluxos (OD e trajetória)
Análise Demográfica	Sim, correlação com dados populacionais dos bairros	Não abordado	Não abordado	Relação com renda per capita (agregado)	Não abordado
Período de Análise	Histórico extenso (2015-2024)	Análise de dados históricos (3 anos)	Foco em dados instantâneos (tempo real)	Histórico baseado em anuários disponíveis	Análise de dados de um período específico (2023)
Aspectos BI	Foco em Data Storytelling e estudos de caso guiados	Ferramenta de exploração visual interativa	Serviço de informação ao usuário final	Dashboard descritivo e exploratório	Visual Analytics com foco em usabilidade

A Tabela 2 sintetiza a comparação entre essas soluções e o *DashMobi*, destacando

as características distintivas de cada uma.

A análise comparativa evidencia que o *DashMobi* se posiciona como uma solução singular ao preencher lacunas importantes deixadas pelas demais ferramentas. Sua principal inovação reside na integração sinérgica de três dimensões de análise: a operacional (eficiência das linhas, ocorrências), a geográfica (distribuição por bairros) e a demográfica (equidade do serviço).

Além disso, a ênfase em *Data Storytelling*, materializada nos estudos de caso guiados, eleva o *DashMobi* de um simples visualizador de dados para uma verdadeira ferramenta de geração de *insights*, permitindo que gestores e cidadãos não apenas “vejam” os dados, mas também “entendam as histórias” que eles contam sobre a eficiência, os desafios e a equidade da mobilidade urbana em Belo Horizonte.

3 Desenvolvimento

Este capítulo detalha o processo de construção da solução *DashMobi*, abordando desde a arquitetura de software e a modelagem do banco de dados até as tecnologias e os processos de implementação de cada componente que compõe a plataforma.

3.1 Arquitetura da Solução

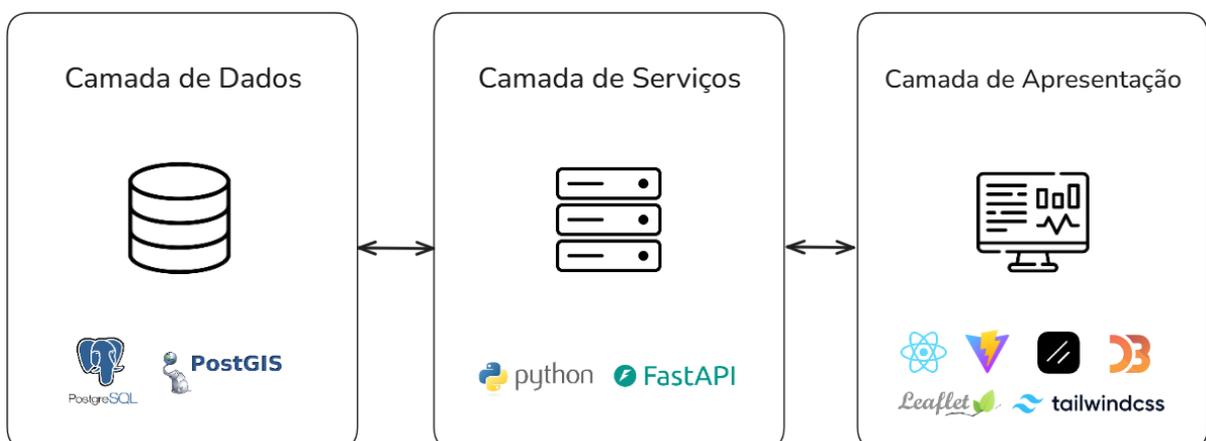
A materialização dos objetivos do projeto exige uma estrutura de software robusta, capaz de gerenciar grandes volumes de dados e, ao mesmo tempo, oferecer uma experiência de usuário fluida e interativa. Para atender a esses requisitos, a arquitetura do *DashMobi* foi cuidadosamente planejada, conforme detalhado a seguir.

3.1.1 Visão Geral da Arquitetura

Para atender aos objetivos de performance, escalabilidade e manutenibilidade, a solução *DashMobi* foi projetada seguindo uma arquitetura de três camadas (*three-tier architecture*), um padrão consolidado para o desenvolvimento de aplicações web. Essa abordagem promove o desacoplamento das responsabilidades do sistema em camadas lógicas e independentes: a camada de dados, a camada de serviços e a camada de apresentação. A principal vantagem desse modelo é permitir que cada camada seja desenvolvida, otimizada e escalada de forma autônoma.

A Figura 1 ilustra a interação entre essas camadas e as tecnologias escolhidas para cada uma delas. A seguir, cada componente da arquitetura é detalhado.

Figura 1 – Arquitetura de três camadas do sistema *DashMobi*



Fonte: Produzido pelo autor

1. **Camada de Dados (*Data Layer*):** Fundamento da arquitetura, esta camada é responsável pelo armazenamento persistente e gerenciamento otimizado dos dados. Foi empregado o Sistema de Gerenciamento de Banco de Dados (SGBD) **PostgreSQL**, estendido com o módulo **PostGIS** para o tratamento eficiente de dados geoespaciais. Conforme detalhado na Seção 3.2.3, os dados foram estruturados em um modelo estrela para maximizar a performance de consultas analíticas complexas.
2. **Camada de Serviço (*Service Layer*):** Esta camada é responsável por intermediar a comunicação entre o banco de dados e a interface do usuário. Foi desenvolvida uma API RESTful utilizando **Python** e o *framework* de alta performance **FastAPI**. A API encapsula toda a lógica de negócio, expondo *endpoints* específicos que executam consultas SQL, processam e agregam os dados, e os retornam em formato JSON para a camada de apresentação.
3. **Camada de Apresentação (*Presentation Layer*):** Esta é a interface com a qual o usuário interage diretamente. Foi desenvolvida como uma *Single-Page Application* (SPA) utilizando o framework **React**, com o ferramental **Vite**. A camada de apresentação consome os dados da API e os renderiza em componentes visuais interativos. Para isso, foram utilizadas as seguintes bibliotecas: **Shadcn UI** para a composição da interface, **Leaflet** para os mapas, **D3.js** para os gráficos e **Tailwind CSS** para a estilização.

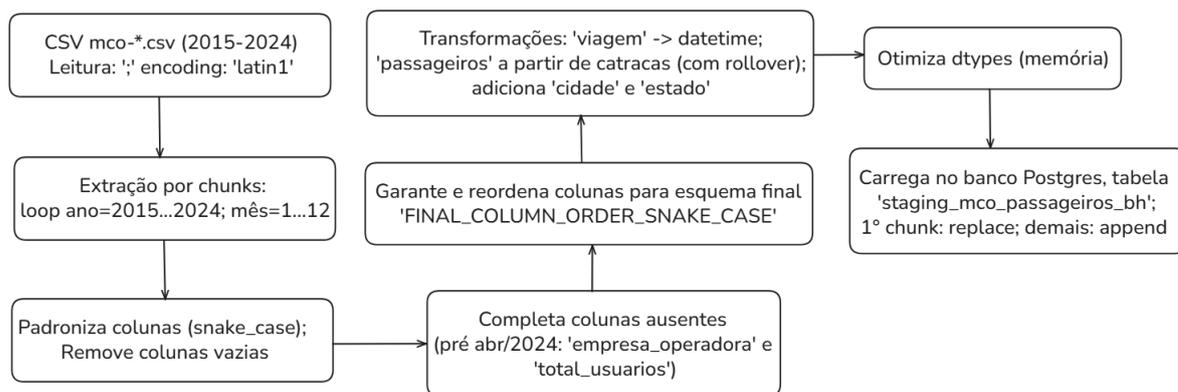
3.1.2 Pipeline dos Dados

Para transformar os dados brutos, provenientes de múltiplos arquivos CSV com estruturas e períodos distintos, em um conjunto de informações limpo, padronizado e otimizado para análise, foi desenvolvida uma pipeline de Extração, Transformação e Carga (ETL). Este processo foi estruturado em *scripts* Python, utilizando a biblioteca Pandas, com um fluxo específico para cada uma das três fontes de dados, conforme ilustrado nas Figuras 2, 3 e 4. O objetivo central desta pipeline foi automatizar o tratamento dos dados e carregá-los em tabelas de *staging* no banco de dados PostgreSQL, preparando-os para a modelagem analítica.

A pipeline para o *dataset* de MCO Passageiros, o mais volumoso e complexo, foi projetada para lidar com milhões de registros de forma eficiente. A extração é realizada em blocos para evitar o consumo excessivo de memória, lendo iterativamente os 120 arquivos CSV mensais (2015-2024). O processo de transformação inclui diversas etapas críticas: a padronização dos nomes das colunas para o formato *snake_case*; o preenchimento de colunas ausentes em arquivos anteriores a abril de 2022 (*total_usuarios* e *empresa_operadora*) para garantir a consistência estrutural; a conversão da coluna viagem para o tipo *datetime*; e o cálculo da métrica passageiros a partir dos dados de catraca, incluindo a lógica para tratar

o *rollover* do contador. Ao final, após a otimização dos tipos de dados para reduzir o uso de memória, os blocos processados são carregados na tabela *staging_mco_passageiros_bh*, utilizando uma estratégia de substituição para o primeiro bloco e anexação para os subsequentes.

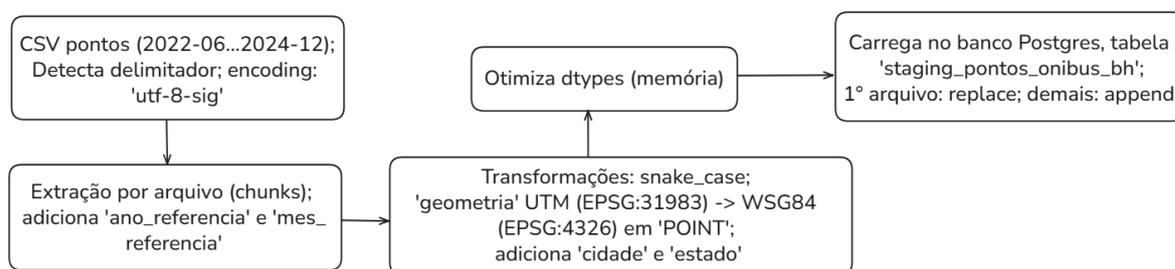
Figura 2 – Fluxograma da Pipeline de ETL para o dataset MCO Passageiros



Fonte: Produzido pelo autor

Para o *dataset* de Pontos de Ônibus, a pipeline também realiza a extração em blocos a partir dos arquivos mensais (2022-2024), detectando automaticamente o delimitador de cada arquivo. A principal etapa de transformação, além da padronização dos nomes das colunas, é o tratamento da coluna de geometria. As coordenadas, originalmente no sistema UTM (EPSG:31983), são convertidas para o sistema de referência geoespacial padrão WGS 84 (EPSG:4326), utilizado em aplicações web de mapeamento. Adicionalmente, colunas de contexto temporal (*ano_referencia* e *mes_referencia*) são adicionadas a partir dos nomes dos arquivos. Após a otimização dos tipos, os dados são carregados na tabela *staging_pontos_onibus_bh*.

Figura 3 – Fluxograma da Pipeline de ETL para o dataset de Pontos de Ônibus

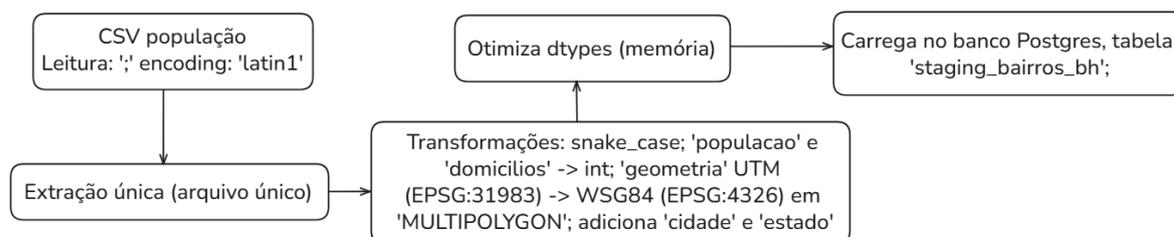


Fonte: Produzido pelo autor

A pipeline para o *dataset* de Bairros é mais simples, pois opera sobre um único arquivo. A extração lê o arquivo CSV por completo. Na fase de transformação, além da padronização e conversão de tipos das colunas *populacao* e *domicilios* para inteiro, o foco principal é a conversão da coluna geometria do formato *SDO_GEOMETRY* (nativo

do Oracle) para o formato aberto WKT (*Well-Known Text*), permitindo sua posterior utilização no PostGIS. Após a otimização de tipos, o *DataFrame* completo é carregado na tabela *staging_bairros_bh* com uma estratégia de substituição total, adequada para dados cadastrais.

Figura 4 – Fluxograma da Pipeline de ETL para o dataset de Bairros



Fonte: Produzido pelo autor

3.2 Construção do Ambiente de Dados

A construção de um *dashboard* analítico eficaz começa com a estruturação cuidadosa dos dados que o alimentarão. Esta seção detalha as etapas fundamentais para a criação do ambiente de dados do *DashMobi*. O processo inicia-se com a identificação e descrição das fontes de dados brutas, seguido pelo detalhamento do pipeline de pré-processamento e carga (ETL), e culmina na modelagem do banco de dados analítico, que servirá de base para todas as visualizações e análises.

3.2.1 Fonte de Dados

A base de dados para este trabalho origina-se do sistema de transporte público de Belo Horizonte, com informações disponibilizadas pela Superintendência de Mobilidade (BHTrans) no Portal de Dados Abertos do município. A escolha deste acervo foi justificada na Seção 2.2.2, dada a sua riqueza, granularidade e disponibilidade, que o tornam ideal para uma análise multifacetada. Foram selecionados três conjuntos de dados principais, que se complementam para formar a base analítica do *DashMobi*.

1. Mapa de Controle Operacional (MCO) Consolidado

Este é o *dataset* central do estudo. Ele contém registros detalhados de cada viagem de ônibus realizada, incluindo informações cruciais como a linha, o veículo, os horários de partida e chegada (programados e realizados), dados de catraca e ocorrências operacionais (desvios, falhas mecânicas). O MCO também classifica o tipo de dia (útil, sábado, domingo/feriado) e registra a extensão da viagem. Notavelmente, a partir de abril de 2022, o conjunto de dados foi enriquecido com o total de usuários transportados e a

empresa operadora, ampliando o potencial analítico. Para este trabalho, foram processados os dados de janeiro de 2015 a dezembro de 2024.

O material original está organizado em arquivos CSV mensais (padrão mco-MM-AAAA.csv), totalizando 120 arquivos. O volume consolidado para o período analisado é de aproximadamente 94,4 milhões de registros e 18 colunas, ocupando 7,95 GB de armazenamento. Cada registro corresponde a uma viagem ou a um segmento desta. Seus atributos principais são: VIAGEM, LINHA, SUBLINHA, PC, CONCESSIONARIA, SAIDA, VEICULO, CHEGADA, CATRACA SAIDA, CATRACA CHEGADA, OCORRENCIA, JUSTIFICATIVA, TIPO DIA, EXTENSAO, FALHA MECANICA, EVENTO INSEGURO, INDICADOR FECHAMENTO, DATA FECHAMENTO, TOTAL USUARIOS, e EMPRESA OPERADORA. O dicionário de dados detalhado para o MCO é referenciado na Tabela 3.

2. Dados de Pontos de Ônibus

Este conjunto de dados estabelece a dimensão espacial da rede, contendo informações sobre os pontos de parada da cidade e associando cada um às linhas que o atendem. Sua característica mais importante é a inclusão das coordenadas geográficas de cada ponto, representadas por uma geometria do tipo *PONTO*. Foram utilizados os arquivos mais recentes e consistentes, de junho de 2022 a dezembro de 2024, para garantir a representação atual da infraestrutura. O conjunto é composto por 31 arquivos CSV (padrão AAAA-MM_ponto_onibus.csv). Os atributos principais são: ID_PONTO_ONIBUS_LINHA, COD_LINHA, NOME_LINHA, NOME_SUB_LINHA, ORIGEM, IDENTIFICADOR_PONTO_ONIBUS e GEOMETRIA. O dicionário de dados é apresentado na Tabela 4.

3. Dados de Bairros de Belo Horizonte

Para a camada de análise demográfica, utilizou-se o *dataset* de bairros do município, contendo informações cadastrais e geométricas baseadas no censo de 2022. Ele inclui o nome do bairro, dados populacionais, número de domicílios, área (km²) e a geometria poligonal que define seus limites. Este conjunto é composto por um único arquivo CSV. O dicionário de dados pode ser observado na Tabela 5.

Apesar da alta qualidade dos dados, a análise exploratória inicial revelou desafios inerentes a projetos de engenharia de dados, como a necessidade de padronização de formatos, tratamento de valores ausentes e otimização de tipos de dados para performance. Os procedimentos adotados para mitigar essas questões são detalhados na seção subsequente.

Tabela 3 – Dicionário dos dados: *Dataset MCO*

Nome	Tipo	Descrição
VIAGEM	ALFANUMÉRICO	Data e hora de início da viagem
LINHA	ALFANUMÉRICO	Número da linha em que a viagem foi realizada
SUBLINHA	NUMÉRICO	Número da sublinha em que a viagem foi realizada
PC	NUMÉRICO	Número do Ponto de Controle de origem em que a viagem foi iniciada
CONCESSIONÁRIA	NUMÉRICO	Número da concessionária ao qual a linha está vinculada
SAÍDA	ALFANUMÉRICO	Hora de saída da viagem do PC de origem
VEÍCULO	NUMÉRICO	Número de ordem do veículo
CHEGADA	ALFANUMÉRICO	Hora de chegada da viagem do PC de destino
CATRACA SAÍDA	NUMÉRICO	Catraca registrada no início da viagem com 5 dígitos
CATRACA CHEGADA	NUMÉRICO	Catraca registrada no fim da viagem com 5 dígitos
OCORRÊNCIA	ALFANUMÉRICO	Indicador se houver interrupção de viagem
JUSTIFICATIVA	ALFANUMÉRICO	Indicador do tipo de justificativa da ocorrência
TIPO DIA	NUMÉRICO	Tipo do dia em que a viagem foi realizada
EXTENSÃO	NUMÉRICO	Extensão da viagem realizada, em metros
FALHA MECÂNICA	ALFANUMÉRICO	Indicador se houve falha mecânica durante a viagem
EVENTO INSEGURO	ALFANUMÉRICO	Indicador se houve evento inseguro durante a viagem
INDICADOR FECHAMENTO	ALFANUMÉRICO	Indicador se a viagem foi fechada
DATA FECHAMENTO	ALFANUMÉRICO	Data e hora do fechamento da viagem
TOTAL USUÁRIOS	ALFANUMÉRICO	Número total de usuários da viagem (a partir de abr/2022)
EMPRESA OPERADORA	ALFANUMÉRICO	Código da empresa operadora do veículo que realizou a viagem (a partir de abr/2022)

Tabela 4 – Dicionário dos dados: *Dataset* de pontos de ônibus

Nome	Tipo	Descrição
ID_PONTO_ONIBUS_LINHA	NUMBER	Código identificador automático da camada
COD_LINHA	VARCHAR2	Código da linha de ônibus que possui embarque e desembarque de passageiros em cada ponto de ônibus
NOME_LINHA	VARCHAR2	Nome da linha de ônibus
NOME_SUB_LINHA	VARCHAR2	Nome da sub linha de ônibus
ORIGEM	VARCHAR2	Origem da saída da linha de ônibus
IDENTIFICADOR_PONTO_ONIBUS	NUMBER	Identificador do ponto de ônibus, esse atributo é igual para todos os pontos que possuem a mesma localização
GEOMETRIA	SDO_GEOMETRY	Representação da geometria do ponto de ônibus, é do tipo Ponto

Tabela 5 – Dicionário dos dados: *Dataset* de bairros

Nome	Tipo	Descrição
ID_POP_DOMIC_BAIRRO_2022	NUMBER	Estrutura sequencial que identifica unicamente um registro
NUM_BAIRRO	NUMBER	Código identificador do bairro
BAIRRO	VARCHAR2	Nome do bairro
POPULACAO	NUMBER	Quantidade de pessoas por bairro
DOMICILIOS	NUMBER	Quantidade de domicílios por bairro
AREA_KM	FLOAT	Área em km ² do bairro
GEOMETRIA	SDO_GEOMETRY	Geometria polígono que representa o bairro

3.2.2 Pré-Processamento e Carga (ETL)

O pré-processamento dos dados constituiu uma etapa fundamental no desenvolvimento deste trabalho, visando garantir a qualidade, consistência e usabilidade das informações para as análises subsequentes e para a construção do *dashboard*. Para esta fase, foi utilizada a linguagem de programação Python, especificamente na versão 3.11.4, com destaque para a biblioteca *pandas*, amplamente reconhecida por suas capacidades de manipulação e transformação de dados. O processo foi estruturado em *scripts* de Extração, Transformação e Carga (ETL) dedicados a cada um dos três conjuntos de dados (MCO, Pontos de Ônibus e Bairros), conforme detalhado a seguir.

3.2.2.1 Procedimentos Comuns

Foram aplicados procedimentos comuns de pré-processamento de forma adaptada a todos os *datasets*. Primeiramente, realizou-se a renomeação de colunas: todos os nomes de colunas foram padronizados para o formato *snake_case* (letras minúsculas, com palavras separadas por *underscore*). Este procedimento, implementado através de uma função utilitária denominada *to_snake_case*, tem como objetivo melhorar a legibilidade do código e facilitar a manipulação programática das colunas (por exemplo, a coluna “CATRACA SAIDA” foi transformada em “catraca_saida”).

Em segundo lugar, procedeu-se à otimização de tipos de dados. Com o intuito de reduzir o consumo de memória e acelerar as operações de processamento, os tipos de dados de cada coluna foram otimizados com o auxílio da função *utils.optimize_df_dtypes*. Esta função opera da seguinte maneira: converte colunas do tipo *object* para um tipo numérico, se tal conversão for possível; caso contrário, se a cardinalidade da coluna for baixa (isto é, menos de 50% de valores únicos), ela é convertida para o tipo *category*, que é mais eficiente em termos de uso de memória. Adicionalmente, a função ajusta colunas numéricas inteiras para o menor tipo possível que ainda comporte todos os seus valores (por exemplo, *int8*, *uint16*, *int32*), levando em consideração se os valores são exclusivamente positivos ou se podem assumir valores negativos. Colunas do tipo *float64* são convertidas para *float32* sempre que essa conversão não implicar perda de precisão.

Como terceira medida comum, realizou-se a adição de colunas contextuais: as colunas *cidade* (com o valor constante *belo_horizonte*) e *estado* (com o valor *mg*) foram adicionadas a todos os *DataFrames*. Esta padronização visa facilitar a eventual integração com outros sistemas ou contextos geográficos.

Por fim, implementou-se a leitura incremental (*chunking*). Para os *datasets* mais volumosos, especificamente o MCO e o de Pontos de Ônibus, a leitura dos arquivos CSV foi realizada em blocos (*chunks*). Esta abordagem, que utiliza a capacidade da biblioteca *pandas* e a estrutura de geradores em Python, evitou o carregamento integral dos arquivos na memória RAM, permitindo o processamento de grandes volumes de dados de forma eficiente e prevenindo problemas de estouro de memória.

3.2.2.2 Dataset MCO

No que concerne ao pré-processamento específico do *dataset* MCO, a etapa de extração envolveu a leitura iterativa dos arquivos CSV mensais, organizados por ano (de 2015 a 2024) e mês. Foi especificado o delimitador “;” e a codificação *latin1* para a correta interpretação dos arquivos. Durante o tratamento de formatos, identificou-se uma diferença estrutural entre os arquivos antigos e os mais recentes: arquivos anteriores a abril de 2022 não possuíam as colunas *total_usuarios* e *empresa_operadora*. Para sanar

essa divergência, essas colunas foram adicionadas programaticamente aos *DataFrames* correspondentes aos arquivos mais antigos, sendo preenchidas com valores *pd.NA* (*Not Available*), de forma a manter a consistência estrutural com os arquivos mais recentes. A consistência estrutural geral foi assegurada através de uma lista pré-definida de colunas na ordem esperada, garantindo não apenas a ordem final das colunas, mas também que todas as colunas esperadas estivessem presentes (preenchendo as ausentes com *pd.NA*). Colunas que se apresentaram totalmente vazias ou com nomes vazios foram removidas.

Diversas conversões de tipos específicos foram realizadas. A coluna *viagem*, originalmente uma *string*, foi convertida para o tipo *datetime*, especificando-se o parâmetro *dayfirst=True* para correta interpretação do formato brasileiro de datas. Um importante cálculo foi o da coluna *passageiros*, derivada das colunas *catraca_chegada* e *catraca_saida*. Ambas as colunas de catraca foram inicialmente convertidas para tipo numérico, tratando-se quaisquer erros de conversão (valores não numéricos) como *pd.NA*. A lógica de cálculo de passageiros considerou o fenômeno de *rollover* da catraca (quando a contagem reinicia após atingir seu limite máximo, definido como 99999). Assim, se o valor de *catraca_chegada* fosse menor que o de *catraca_saida* (indicando um *rollover*), a fórmula aplicada foi: $(99999 - \text{catraca_saida}) + \text{catraca_chegada} + 1$. Caso contrário, utilizou-se a subtração direta ($\text{catraca_chegada} - \text{catraca_saida}$). Valores *NA* resultantes na coluna *passageiros* (devido a valores *NA* nas colunas de catraca originais) foram preenchidos com 0, e o tipo final da coluna *passageiros* foi definido como inteiro.

Finalmente, na etapa de carga, os *DataFrames* processados por *chunk* foram carregados na tabela *staging_mco_passageiros_bh* no banco de dados PostgreSQL. A primeira carga de dados para esta tabela utilizou a estratégia *if_exists='replace'*, enquanto as cargas subsequentes de novos *chunks* utilizaram *if_exists='append'*, anexando os novos dados à tabela existente.

3.2.2.3 Dataset de Pontos de Ônibus

Para o pré-processamento específico do *dataset* de Pontos de Ônibus, a extração consistiu na leitura iterativa dos arquivos CSV mensais, abrangendo o período de 2022 a 2024. Uma função auxiliar (*detect_delimiter*) foi implementada para identificar automaticamente se o delimitador utilizado no arquivo era “;” ou “,”, através da leitura da primeira linha do arquivo. Foi utilizada a codificação *utf-8-sig* para lidar corretamente com eventuais *Byte Order Marks* (BOM) que pudessem estar presentes nos arquivos. Na fase de transformação, as colunas *ano_referencia* e *mes_referencia* foram adicionadas a cada *DataFrame*. Estas colunas foram preenchidas com base no ano e mês extraídos do nome do arquivo de origem, fornecendo um contexto temporal para os registros.

Na etapa de carga, os *chunks* processados foram carregados na tabela *staging_pontos_onibus_bh* no PostgreSQL, seguindo a mesma lógica de *replace* para a primeira carga e

append para as subsequentes, adotada para o *dataset* MCO.

3.2.2.4 *Dataset* de Bairros

Por fim, o pré-processamento específico do *dataset* de Bairros iniciou-se com a extração do arquivo CSV único, que foi lido utilizando o delimitador “;” e a codificação *latin1*. Durante a transformação de tipos específicos, as colunas *populacao* e *domicilios* foram convertidas para o tipo numérico, tratando erros de conversão como *NaN* (*Not a Number*); os valores *NaN* resultantes foram então preenchidos com 0 antes da conversão final para o tipo inteiro. A coluna *area_km* também foi convertida para tipo numérico.

Um tratamento notável foi aplicado à coluna geometria. Conforme o dicionário de dados de origem (Tabela 5), este campo é fornecido no formato textual *SDO_GEOMETRY*, nativo do banco de dados Oracle. Para viabilizar seu uso em um ambiente PostGIS, o script de ETL em Python foi responsável por analisar esta string e convertê-la para o formato aberto WKT (*Well-Known Text*). Esta decisão permitiu que a coluna fosse carregada como *TEXT* no PostgreSQL, facilitando sua posterior conversão para o tipo espacial nativo *GEOMETRY* diretamente no banco de dados para a realização de análises geoespaciais.

Na etapa de carga, o *DataFrame* completo dos bairros foi carregado na tabela *staging_bairros_bh* no PostgreSQL. Para este *dataset*, foi utilizada a estratégia *if_exists='replace'*, pois trata-se de um *dataset* cadastral que geralmente é substituído integralmente em atualizações, em vez de ter novos dados anexados.

Ao final desta abrangente etapa de pré-processamento, os três *datasets* (MCO, Pontos de Ônibus e Bairros) encontravam-se limpos, padronizados, com seus tipos de dados otimizados para performance e armazenamento, e devidamente carregados em tabelas de *staging* no banco de dados PostgreSQL. Desta forma, os dados estavam prontos para as etapas subsequentes de modelagem de dados, visando a construção do *dashboard* e a realização das análises propostas neste trabalho.

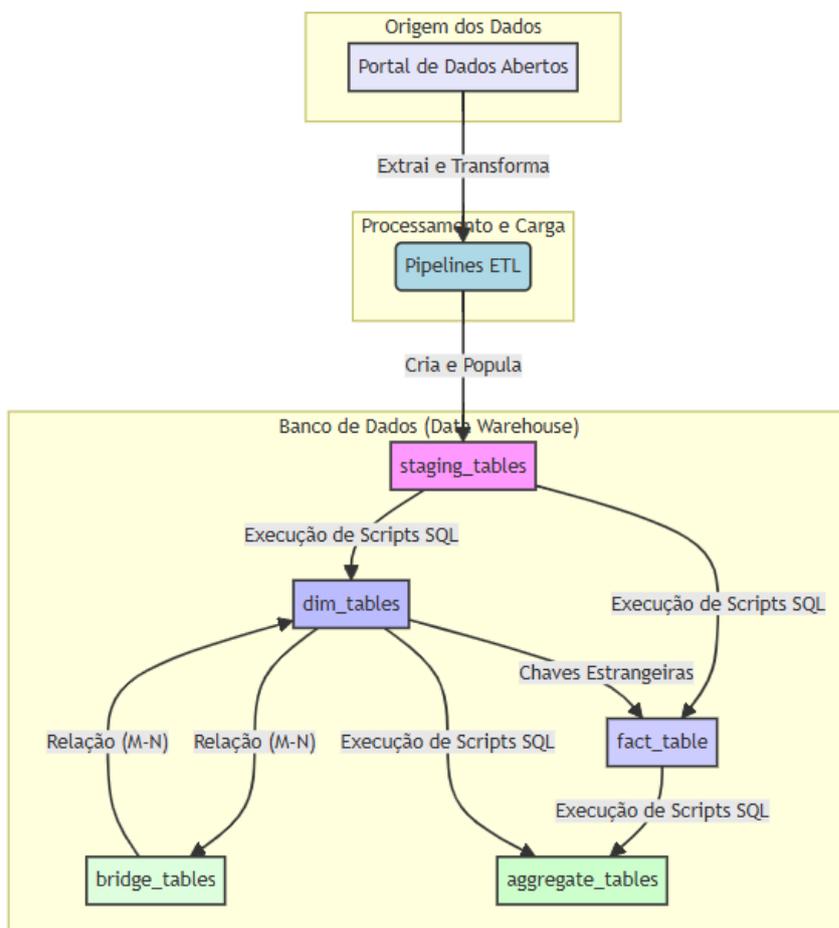
3.2.3 Modelagem do Banco de Dados Analítico

Após a etapa de pré-processamento e carga dos dados brutos em tabelas de *staging*, foi necessário estruturar um ambiente analítico otimizado para as consultas do *dashboard*. A abordagem de consultar diretamente as tabelas de *staging*, que contêm milhões de registros, seria ineficiente para uma ferramenta interativa que exige respostas rápidas. Para este fim, optou-se pela implementação de um modelo estrela (*star schema*).

O modelo estrela é uma abordagem de modelagem de dados para *data warehouses* que organiza as informações em uma tabela de fatos central, contendo as métricas quantitativas, e um conjunto de tabelas de dimensão, que armazenam os atributos descritivos e contextuais. Essa estrutura simplifica as consultas e melhora significativamente a

performance de agregações (somadas, contagens, médias), sendo ideal para ferramentas de *Business Intelligence* e visualização de dados como o *dashboard* proposto.

Figura 5 – Fluxograma da arquitetura de dados, desde a origem até a modelagem do *Data Warehouse*



Fonte: Produzido pelo autor

Com o conceito de modelo estrela definido, a Figura 5 ilustra a arquitetura completa do fluxo de dados implementada no projeto. O diagrama detalha visualmente a jornada da informação: desde sua origem no Portal de Dados Abertos, passando pelo processamento na Pipeline ETL que alimenta as tabelas de *staging*, até a execução dos *scripts* SQL que constroem o modelo analítico. Fica evidente a centralidade do modelo estrela (composto pelas *dim_tables* e a *fact_table*), bem como sua relação com as *bridge_tables* e a camada final de otimização com as *aggregate_tables*.

3.2.3.1 Tabelas de Dimensão

As tabelas de dimensão são a base para os filtros e agrupamentos do *dashboard*. Elas armazenam os atributos de “quem, o quê, onde e quando” dos eventos. Foram criadas as seguintes dimensões principais:

- **dim_data**: Uma tabela de dimensão de tempo, essencial para análises temporais. Ela é populada a partir da coluna viagem do *dataset* MCO e contém atributos como ano, mês, dia e dia da semana. Para tratar casos em que uma mesma data possuía múltiplos *tipo_dia* nos dados de origem (e.g., um dia útil que também era um evento especial), foi utilizada a função de agregação *MAX()* durante a carga para garantir a unicidade de cada data, conforme a restrição da tabela.
- **dim_linha**: Consolida os atributos únicos de cada linha de ônibus. Sua população exigiu a junção da tabela *staging_mco_passageiros_bh*, para obter a lista de linhas efetivamente operantes, com a *staging_pontos_onibus_bh*, para extrair os campos *nome_linha* e *origem*. Para o campo *extensao_km*, adotou-se o cálculo da mediana (*PERCENTILE_CONT(0.5)*) para obter um valor central robusto e imune a *outliers* presentes nos dados.
- **dim_bairro**: Armazena as informações cadastrais e geográficas dos bairros, oriundas da *staging_bairros_bh*. Para otimizar as consultas espaciais, a coluna *geom*, do tipo *GEOMETRY* nativo do PostGIS, foi priorizada em detrimento da sua representação textual. Adicionalmente, a coluna *densidade_demografica* foi calculada e materializada nesta tabela, a partir da divisão da *populacao* pela *area_km*.
- **dim_veiculo**: Contém os atributos de cada veículo. Para esta dimensão, foram calculadas a primeira e a última data de viagem de cada veículo, permitindo derivar sua idade em anos completos e os meses adicionais restantes, utilizando a função *age()* do PostgreSQL para maior precisão.
- **dim_empresa** e **dim_concessionaria**: Tabelas criadas para normalizar os códigos e nomes das empresas e concessionárias, populadas a partir dos valores distintos do *dataset* MCO e de mapeamentos externos.
- **dim_ocorrencia** e **dim_justificativa**: A *dim_ocorrencia* armazena os tipos gerais de evento (e.g., “VIAGEM INTERROMPIDA”), enquanto a *dim_justificativa* detalha a causa específica (e.g., “PNEU FURADO”), mantendo uma relação hierárquica através de uma chave estrangeira. Esta abordagem normalizada elimina redundância e aumenta a flexibilidade das análises.

3.2.3.2 Tabela de Fatos

O núcleo do modelo é a *fact_viagens*, uma tabela de fatos cujo “grão” é cada viagem individual registrada no MCO. Ela armazena as principais métricas a serem analisadas e as chaves estrangeiras que a conectam a todas as tabelas de dimensão.

Seu principal fato numérico é a métrica *passageiros*, calculada a partir da lógica de *rollover* da catraca. Além disso, foram adicionadas métricas calculadas para enriquecer a

análise, como *extensao_realizada_km* e *duracao_realizada_minutos*, que considera viagens que cruzam a meia-noite. Para facilitar as agregações no *dashboard*, foram criadas colunas de indicadores no formato *flag* (com valores 0 ou 1), como *flag_viagem_realizada* e *flag_possui_ocorrencia*.

3.2.3.3 Estratégias de Carga e Otimização

A população do modelo estrela a partir das tabelas de *staging* foi realizada por meio de *scripts* SQL, empregando técnicas específicas para lidar com a complexidade e o volume dos dados.

A carga da *fact_viagens*, por ser a operação mais pesada, foi estruturada para ser executada em lotes, processando um mês de dados por vez para evitar transações de longa duração e o consumo excessivo de memória. Para maximizar a performance dos *JOINS*, foi criada e indexada uma coluna auxiliar, *data_viagem*, na tabela *staging_mco_passageiros_bh*, eliminando a necessidade de conversão de tipo em tempo de execução.

Adicionalmente, para viabilizar análises que cruzam informações de linhas e bairros, foram criadas duas tabelas de ponte. A população destas tabelas foi otimizada para executar a custosa consulta espacial com a função *ST_Contains* do PostGIS apenas uma vez, para criar a *bridge_ponto_bairro*. O resultado desta primeira carga foi então reutilizado para popular a *bridge_linha_bairro* através de *JOINS* padrão, uma abordagem significativamente mais performática.

3.2.3.4 Tabelas Agregadas e Views Materializadas

Para garantir a performance necessária para um *dashboard* interativo, a abordagem de consultar diretamente a tabela de fatos, com seus quase 100 milhões de registros, seria inviável para a maioria das visualizações que exigem totalizações em tempo real. A latência de tais consultas comprometeria a experiência do usuário.

Para mitigar este desafio e assegurar respostas na ordem de milissegundos, foi implementada uma camada de agregação no *data warehouse*. Esta camada consiste em tabelas pré-calculadas e *views* materializadas que sumarizam os dados da tabela de fatos em diferentes níveis de granularidade. Esta estratégia de otimização é um pilar da arquitetura da solução. As principais estruturas criadas foram:

- **Tabelas Agregadas Diárias** (*agg_metricas_*_diarias*): Foi criado um conjunto de tabelas que sumarizam as principais métricas (total de viagens, passageiros, ocorrências, etc.) por dia para cada uma das principais dimensões de análise: *agg_metricas_linhas_diarias*, *agg_metricas_empresas_diarias*, *agg_metricas_concessionarias_diarias* e *agg_metricas_veiculos_diarias*. Ao consultar estas tabelas,

a API consegue calcular KPIs e rankings para longos períodos de tempo de forma extremamente eficiente, lendo um volume de dados drasticamente menor.

- **Tabela Agregada Horária** (*agg_metricas_horarias*): Para atender às necessidades do estudo de caso “Horários de Utilização”, foi criada uma tabela agregada ainda mais granular. Ela sumariza as métricas por hora, permitindo que as análises de horários de pico também sejam executadas com alta performance, sem a necessidade de processar a tabela de fatos em tempo real.
- **View Materializada** (*mv_empresa_principal_veiculo*): Para resolver uma consulta específica e computacionalmente muito custosa — a identificação da empresa que mais operou cada veículo —, optou-se pelo uso de uma *view* materializada. Esta estrutura armazena o resultado pré-calculado desta consulta complexa, transformando uma operação que levaria minutos em uma busca instantânea, similar a consultar uma tabela indexada. Como o conjunto de dados do projeto é estático, a atualização desta *view* não se faz necessária, tornando-a uma “tabela de atalho” permanente e eficiente.

A utilização estratégica dessas estruturas de agregação foi a decisão arquitetural chave para garantir que o *DashMobi* atendesse aos requisitos de interatividade e velocidade, validando a eficácia do modelo de dados implementado.

3.3 Projeto da Interface e Experiência do Usuário

Com o ambiente de dados analítico devidamente estruturado e populado, a etapa subsequente do desenvolvimento concentra-se na construção da aplicação *DashMobi*, que consumirá esses dados e os apresentará de forma interativa. Esta seção descreve a arquitetura planejada para o sistema e detalha as funcionalidades de visualização e análise que serão implementadas.

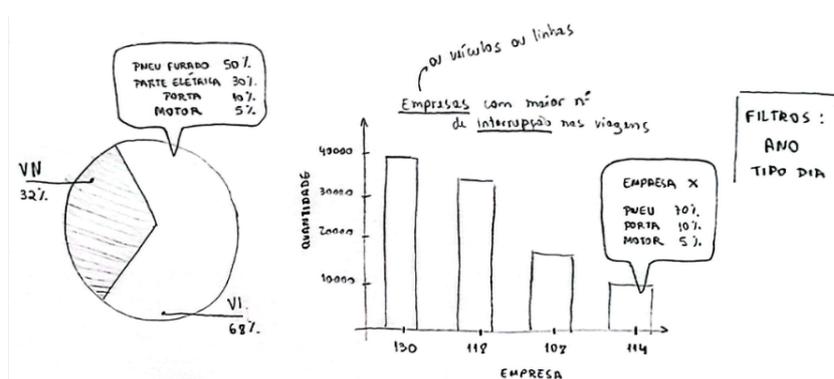
3.3.1 Prototipação da Interface

Antes de iniciar a implementação do *front-end*, foi realizada uma etapa de prototipação da interface, um processo fundamental para a concepção e validação do design da ferramenta. Seguindo práticas de desenvolvimento de sistemas interativos, a prototipação foi dividida em duas fases: a criação de protótipos de baixa fidelidade para a exploração inicial de ideias e, posteriormente, o desenvolvimento de protótipos de média fidelidade para o refinamento do layout e da experiência do usuário.

A fase inicial consistiu na elaboração de esboços à mão, ou protótipos de baixa fidelidade, que serviram como uma ferramenta de *brainstorming* para definir os principais

componentes analíticos do *dashboard*. Conforme ilustrado na Figura 6, esses rascunhos permitiram explorar diferentes formas de visualizar os dados, como a relação entre tipos de ocorrência (Viagem Não Realizada - VN, e Viagem Interrompida - VI), as justificativas mais comuns (e.g., Pneu Furado, Motor) e o ranking de entidades (empresas) mais afetadas. Esta abordagem ágil e de baixo custo foi crucial para materializar as ideias iniciais e validar os conceitos centrais de análise antes de avançar para uma ferramenta de design digital.

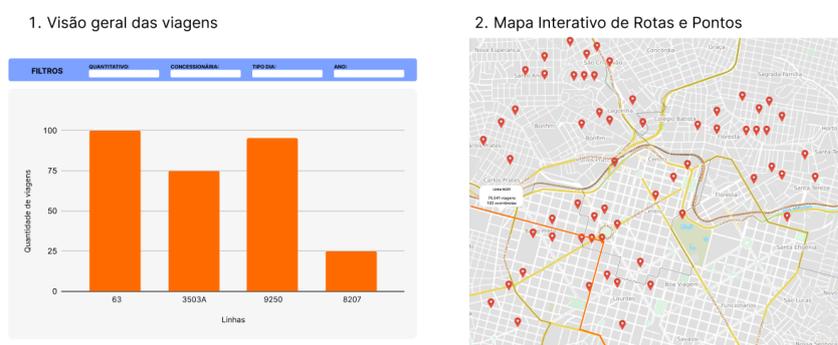
Figura 6 – Esboço de baixa fidelidade explorando os tipos de visualizações



Fonte: Produzido pelo autor

Após a consolidação dos conceitos iniciais, a segunda fase envolveu a criação de protótipos de média fidelidade utilizando a ferramenta de design Figma. Esta etapa foi marcada por um processo iterativo de refinamento, que evoluiu de componentes isolados para um layout de página consolidado. Inicialmente, o foco foi na concepção dos elementos visuais centrais, como os gráficos de barras e o mapa interativo, para definir sua aparência e funcionalidade básica (Figura 7).

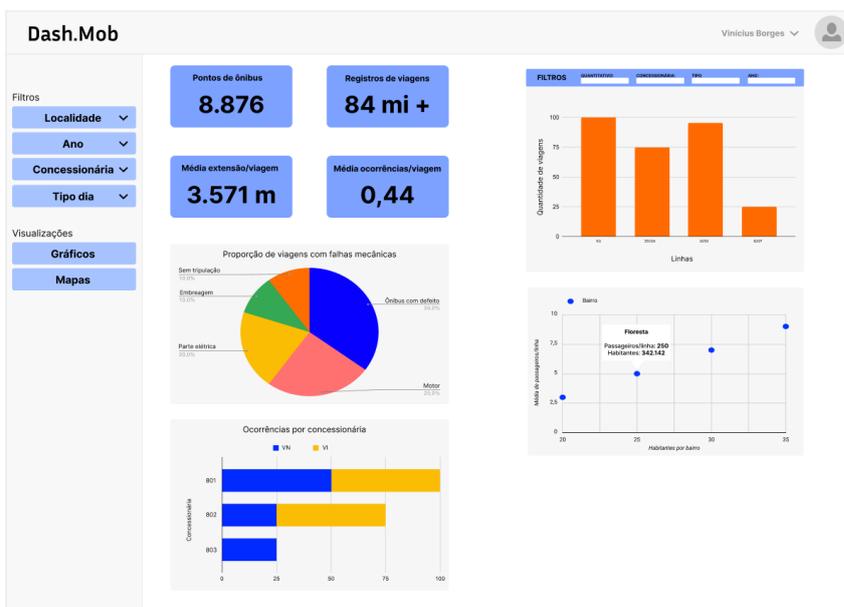
Figura 7 – Esboço de média fidelidade dos componentes visuais



Fonte: Produzido pelo autor

Em seguida, o design avançou para a estruturação de uma página temática completa, explorando como os KPIs, gráficos e a navegação poderiam ser organizados em um único painel (Figura 8). Esta versão inicial ajudou a definir a hierarquia visual e a disposição geral dos elementos.

Figura 8 – Protótipo inicial do *dashboard*



Fonte: Produzido pelo autor

Por fim, esses aprendizados foram consolidados em um protótipo mais maduro e estruturado (Figura 9), que representa uma versão próxima da interface final implementada. Neste design, foi definida a estrutura padrão da interface, com o menu de navegação lateral, a barra de filtros globais no topo e a organização do conteúdo principal em *cards* e grades. O processo evolutivo de prototipação no Figma foi fundamental para antecipar desafios de usabilidade, equilibrar a densidade de informações com a clareza visual e servir como um guia preciso para a etapa de desenvolvimento do *front-end*, otimizando o tempo e reduzindo a necessidade de retrabalho.

Figura 9 – Protótipo final do *dashboard*



Fonte: Produzido pelo autor

3.3.2 Módulos de Visualização e Análise

A interface do *DashMobi* foi projetada para organizar as análises em sete módulos temáticos, cada um focado em uma dimensão específica do sistema de transporte, permitindo uma exploração de dados tanto guiada quanto intuitiva.

- **Visão Geral:** Tela principal do *dashboard*, dedicada a fornecer uma visão macro do sistema. Apresentará KPIs para o total de passageiros, viagens e ocorrências. O destaque será um mapa de calor interativo de Belo Horizonte para visualizar a concentração de passageiros, ocorrências ou linhas por bairro.
- **Linhas:** Focado no desempenho das linhas. Permitirá a criação de rankings dinâmicos por número de passageiros, ocorrências ou viagens. Ao selecionar uma linha, a interface exibirá um perfil completo, incluindo sua extensão (*extensao_km*), empresa e concessionária, utilizando os dados da *dim_linha*. A análise de pontos será feita em conjunto com o módulo de bairros.
- **Ocorrências:** Destinado a entender as interrupções do serviço. Beneficiando-se da estrutura normalizada das dimensões *dim_ocorrencia* e *dim_justificativa*, a interface permitirá ao usuário filtrar por tipo de ocorrência (e.g., “Viagem Interrompida”) e visualizar um gráfico com as justificativas mais comuns (e.g., “Pneu Furado”), cruzando esses dados por linha, empresa ou bairro.
- **Bairros:** Central para análises territoriais. O módulo correlacionará dados demográficos (população, densidade) da *dim_bairro* com a oferta de transporte, como a quantidade de linhas e pontos de ônibus que atendem a região (calculada via *bridge_linha_bairro* e *bridge_ponto_bairro*), identificando potenciais desequilíbrios entre oferta e demanda.
- **Concessionárias:** Permitirá a análise comparativa entre os consórcios. Serão exibidos rankings e gráficos comparando as concessionárias por métricas como total de passageiros transportados, número de linhas gerenciadas, total de viagens realizadas e frequência de ocorrências, utilizando a *dim_concessionaria* para agrupar os dados da *fact_viagens*.
- **Empresas:** Semelhante ao módulo de concessionárias, esta seção focará na análise individual das empresas operadoras. O usuário poderá comparar o desempenho das empresas em relação à pontualidade (implícito nas ocorrências de atraso), número de falhas mecânicas, e eficiência operacional (passageiros por quilômetro), utilizando a *dim_empresa*.
- **Veículos:** Este módulo será dedicado à análise da frota. A interface permitirá visualizar a distribuição da idade da frota, utilizando o campo *idade_veiculo_anos*

da *dim_veiculo*. Será possível também filtrar por empresa para analisar a idade média da frota de cada uma e correlacionar a idade dos veículos com a quantidade de falhas mecânicas registradas.

3.3.3 Estudos de Caso Guiados

Para demonstrar a capacidade analítica da ferramenta e aplicar os princípios de *Data Storytelling*, o *DashMobi* conta com uma seção dedicada a estudos de caso. Cada estudo de caso é apresentado em uma página própria, com uma narrativa visual que guia o usuário desde uma pergunta de alto nível até a exploração de *insights* detalhados. Foram desenvolvidos três estudos de caso principais:

- **Eficiência das Linhas:** Este estudo de caso investiga a eficiência operacional e financeira das linhas de ônibus do sistema. A análise central é um gráfico de quadrantes que classifica cada linha com base em duas métricas: eficiência por distância (Passageiros/Km) e eficiência por tempo (Passageiros/Minuto). A interface permite que o usuário explore os quatro perfis de desempenho resultantes (“As Melhores”, “Lentas e Lotadas”, “Rápidas e Vazias” e “Alvos de Otimização”). A análise é enriquecida com a possibilidade de o usuário simular cenários financeiros, ajustando parâmetros como o custo operacional por quilômetro e o valor da tarifa, e visualizar o impacto dessas mudanças no resultado líquido (lucro ou subsídio) de cada linha através de *tooltips* e modais interativos.
- **Distribuição das Falhas Mecânicas:** Este estudo aborda o problema das falhas mecânicas, seguindo uma narrativa em três etapas para identificar suas causas e consequências. A primeira etapa apresenta um panorama geral, com rankings de empresas por taxa de falhas e um gráfico de pizza com os tipos de problema mais comuns. A segunda etapa aprofunda a análise com um gráfico de bolhas interativo, que correlaciona o desgaste por uso (Total de Km Percorridos) de cada veículo com o número de falhas, enquanto o tamanho da bolha representa o desgaste por tempo (Idade do Veículo). A terceira etapa foca no impacto geográfico, exibindo o ranking das linhas que mais sofrem com falhas. A interface permite ao usuário filtrar os dados e explorar os destaques da análise, como os veículos e as empresas mais problemáticos.
- **Horários de Utilização:** O terceiro estudo de caso analisa a dimensão temporal da operação, investigando como o “pulso” da cidade varia ao longo das 24 horas do dia. A página apresenta três gráficos de linha sincronizados que exibem o volume de passageiros, o volume de ocorrências e a duração média das viagens para cada hora. A principal funcionalidade desta análise é a sua capacidade de filtragem avançada, permitindo que o usuário segmente os dados não apenas pelo período, mas também

por Linha, Empresa e Tipo de Dia (Dia Útil, Sábado, Domingo/Feriado), revelando padrões de utilização específicos para diferentes contextos operacionais.

A implementação destes módulos e funcionalidades no *back-end* e *front-end* constitui a fase final do desenvolvimento, culminando na ferramenta *DashMobi* plenamente funcional, cujos resultados e validação serão apresentados no capítulo seguinte.

4 Resultados

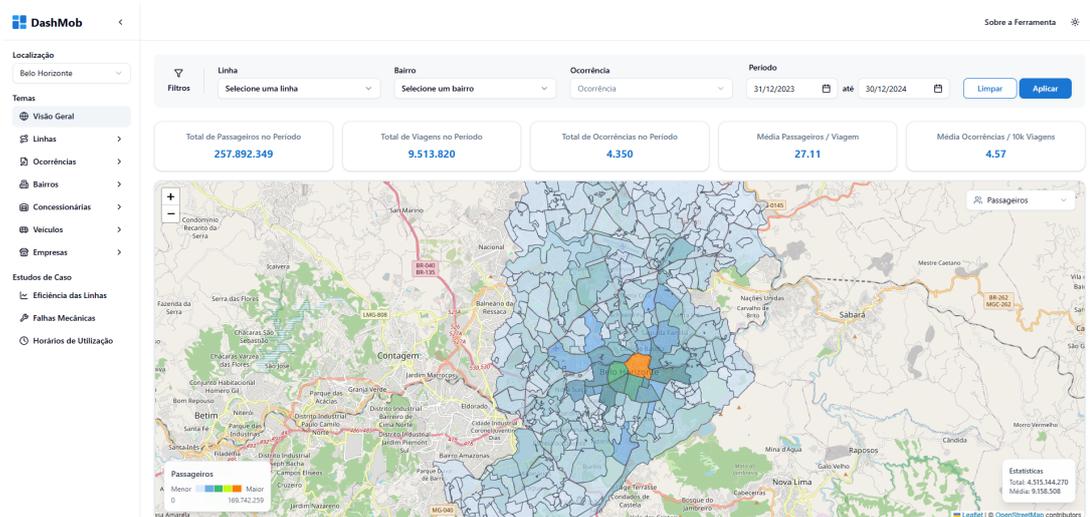
Neste capítulo, são apresentados os resultados obtidos com o desenvolvimento da ferramenta *DashMobi*. A exposição será dividida em três etapas principais: primeiramente, uma apresentação da interface geral e de seus componentes visuais fundamentais; em seguida, uma demonstração da aplicação desses componentes nos módulos temáticos, utilizando o módulo “Linhas” como exemplo representativo; e, por fim, a validação dos estudos de caso propostos, que atestam a capacidade analítica da solução.

4.1 Apresentação da Ferramenta *DashMobi*

O resultado deste trabalho é o *DashMobi*, um *dashboard* analítico interativo para a exploração de dados da mobilidade urbana de Belo Horizonte. A ferramenta foi projetada com foco em uma experiência de usuário intuitiva, organizando as análises em módulos temáticos e estudos de caso, acessíveis através de um menu de navegação lateral.

A página principal, intitulada “Visão Geral” (Figura 10), serve como ponto de partida para a análise, oferecendo um panorama do sistema de transporte. Conforme ilustrado, a interface é composta por uma barra de filtros globais no topo, seguida pelos principais componentes visuais: uma fileira de Indicadores-Chave de Desempenho (KPIs) e um mapa geográfico interativo. A barra de filtros permite ao usuário segmentar os dados por Linha, Bairro, Ocorrência e Período. Ao clicar no botão “Aplicar”, todas as visualizações da página são atualizadas dinamicamente para refletir a seleção, permitindo uma análise focada e contextualizada.

Figura 10 – Tela principal do *DashMobi*, apresentando a Visão Geral do sistema



Fonte: Captura de tela realizada pelo autor

4.1.1 Componentes Visuais do Dashboard

A interface do *DashMobi* é construída a partir de um conjunto de componentes visuais padronizados, cada um selecionado para cumprir uma função analítica específica. A combinação desses elementos permite a exploração dos dados em diferentes níveis de profundidade.

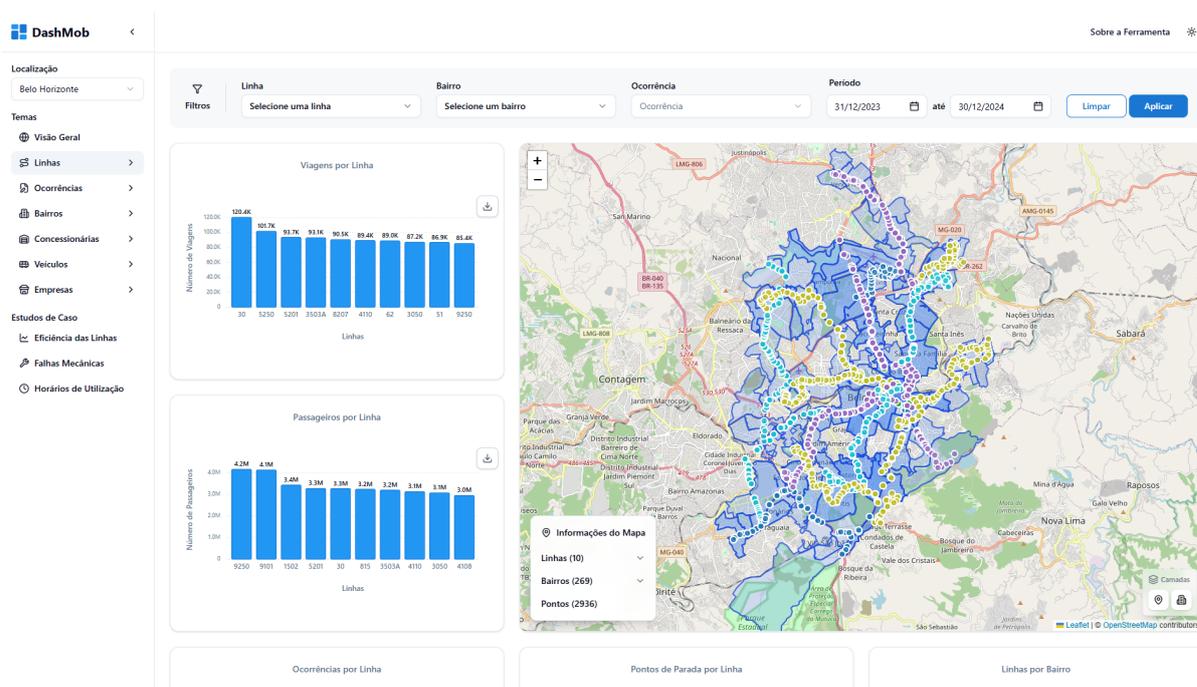
- **Indicadores-Chave de Desempenho (KPIs):** Posicionados com destaque nas telas de visão geral e de análise individual, os KPIs têm a função de fornecer um resumo quantitativo e imediato sobre o desempenho do sistema. Eles apresentam métricas essenciais como o total de passageiros e viagens no período, além de indicadores de eficiência, como a “Média de Passageiros por Viagem”. Por serem o primeiro ponto de contato do usuário com os dados, seu design é minimalista, focando na clareza e legibilidade dos números.
- **Mapas Interativos:** Sendo a dimensão geográfica um pilar deste trabalho, os mapas são um componente fundamental. Foram implementados dois tipos principais de visualização: mapas coropléticos, utilizados na “Visão Geral” para exibir dados agregados por área (Figura 10), onde os polígonos dos bairros são coloridos com tonalidades que variam de acordo com o valor da métrica (e.g., volume de passageiros); e mapas de pontos e polígonos, utilizados nas telas de análise individual para exibir a localização exata dos pontos de ônibus e os contornos dos bairros atendidos, fornecendo um contexto geográfico preciso para a análise operacional.
- **Gráficos de Barras:** Este componente é o principal recurso para a criação de rankings e comparações diretas entre diferentes entidades (linhas, veículos, empresas, etc.). Conforme demonstrado na Figura 7, os gráficos de barras permitem uma identificação rápida dos itens com maior e menor desempenho em uma determinada métrica, como o total de ocorrências, sendo essenciais para as telas de análise comparativa dos módulos temáticos.
- **Gráficos de Pizza:** Este tipo de gráfico é empregado para exibir a composição proporcional de uma métrica, respondendo à pergunta “qual a participação de cada categoria no todo?”. No *DashMobi*, é utilizado nas telas de análise individual para detalhar a natureza das ocorrências, dividindo o total de eventos de uma linha, por exemplo, em suas justificativas mais comuns (e.g., “Motor”, “Pneu Furado”).
- **Gráficos de Tendência Temporal:** Para analisar a evolução de um indicador ao longo do tempo, foram utilizados gráficos de linha (ou de área). Este componente é central no módulo de “Ocorrências”, onde exibe a tendência mensal do número de eventos, permitindo ao usuário identificar padrões de sazonalidade, picos ou quedas abruptas na série histórica.

4.1.2 Apresentação dos Módulos Temáticos

Para aprofundar a análise, o usuário pode navegar pelos módulos temáticos. O módulo “Linhas”, por ser o mais completo, serve como um exemplo representativo da estrutura e capacidade da ferramenta. Ele é dividido em duas visualizações principais: uma análise geral comparativa e uma análise individual detalhada.

A tela de **Análise Geral de Linhas** (Figura 11) utiliza múltiplos gráficos de barras para apresentar rankings comparativos de desempenho. O mapa interativo, por sua vez, exibe a totalidade da malha de transporte, permitindo uma visão geográfica de toda a rede. Além disso, implementou-se um recurso de camadas, permitindo ao usuário exibir ou ocultar os bairros e os pontos de ônibus ao selecionar as camadas no canto inferior direito do mapa. Esta combinação de rankings quantitativos e visualização espacial permite ao usuário identificar padrões tanto operacionais quanto geográficos em todo o sistema.

Figura 11 – Tela de análise geral do módulo “Linhas”



Fonte: Captura de tela realizada pelo autor

Ao optar pela **Análise Individual**, o usuário é apresentado a um modal de seleção (Figura 12) e, após escolher uma rota, é direcionado para um dashboard dedicado (Figura 13). Esta tela oferece um “mergulho profundo” nos dados da linha selecionada. Um card de “Informações da Linha” exibe estatísticas detalhadas, como a empresa operadora, totais de viagens, extensão da rota e médias de passageiros. O mapa, nesta visualização, foca exclusivamente no itinerário da linha, exibindo seu traçado, a localização de cada ponto de ônibus e os polígonos dos bairros que ela atende. Adicionalmente, gráficos de desempenho, como a “Média de passageiros por dia da semana”, revelam os padrões de uso específicos

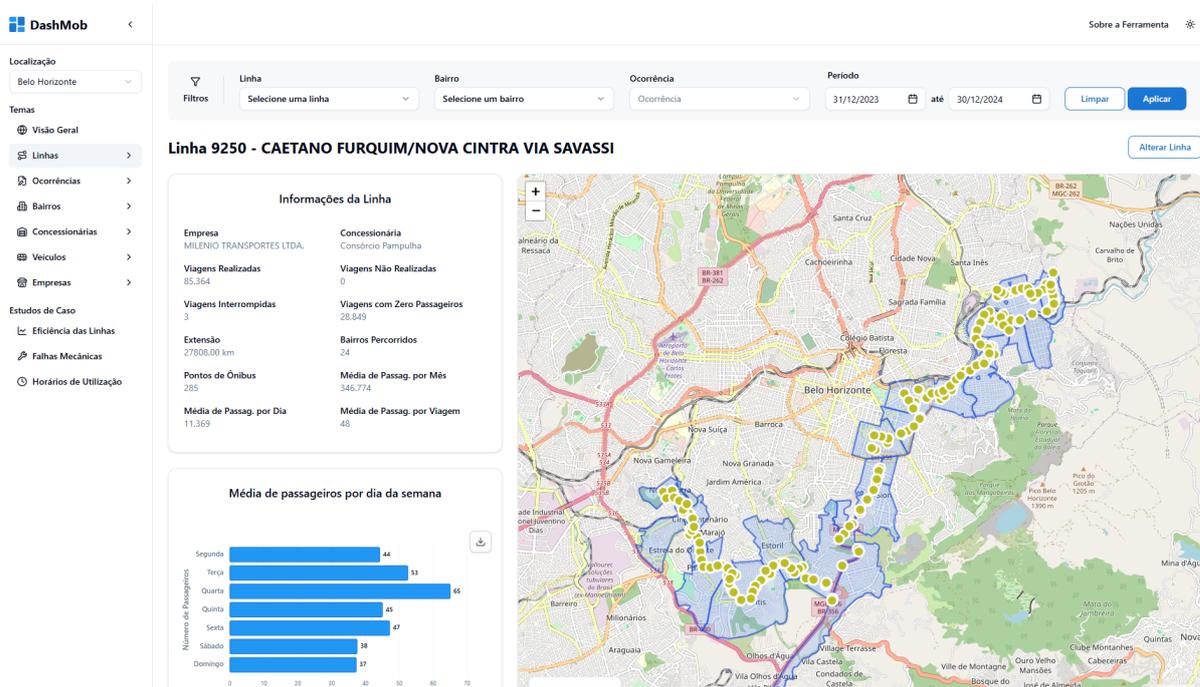
da rota, transformando os dados brutos em um perfil completo e multifacetado de cada linha do sistema.

Figura 12 – Modal de seleção para a Análise Individual de Linha



Fonte: Captura de tela realizada pelo autor

Figura 13 – Tela de análise individual do módulo “Linhas”



Fonte: Captura de tela realizada pelo autor

4.2 Análise dos Dados e Validação dos Estudos de Caso

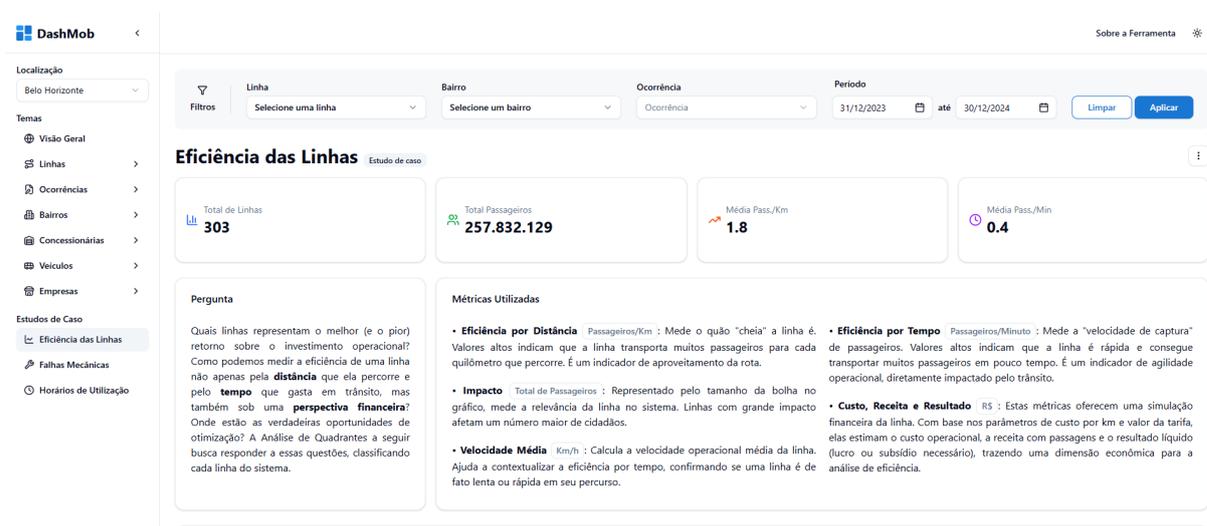
Para validar a capacidade analítica da ferramenta, esta seção apresenta a execução dos estudos de caso projetados. Utilizando as visualizações e os recursos interativos do

dashboard, são demonstrados os resultados e discutidos os *insights* gerados, atestando a eficácia do sistema em responder a perguntas complexas sobre a mobilidade urbana em Belo Horizonte.

4.2.1 Estudo de Caso: Eficiência das Linhas

O primeiro estudo de caso investiga a eficiência operacional e financeira das linhas de ônibus do sistema. A página dedicada a esta análise, ilustrada na Figura 14, foi projetada para responder à pergunta central: “Quais linhas representam o melhor (e o pior) retorno sobre o investimento operacional?”. Para isso, a tela integra múltiplos componentes que permitem uma análise em camadas.

Figura 14 – Visão geral da página do estudo de caso “Eficiência das Linhas”, destacando os KPIs, a pergunta-chave e as métricas utilizadas

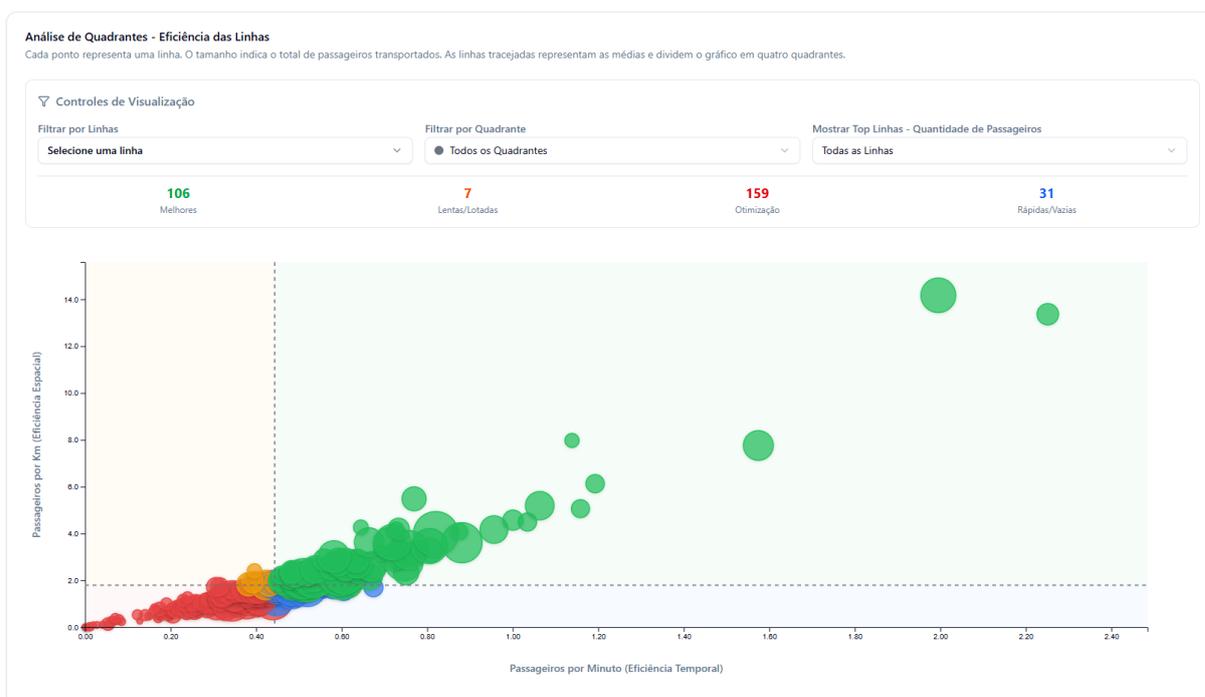


A seção superior da página apresenta KPIs que sumarizam o desempenho de todas as linhas no período selecionado, como o total de linhas analisadas e o total de passageiros transportados por elas. *Cards* de texto contextualizam a análise, explicando tanto a pergunta que guia o estudo quanto as métricas utilizadas no gráfico principal. Além disso, a interface permite que o usuário ajuste os parâmetros de simulação financeira — o custo operacional por quilômetro e o valor da tarifa — através de um modal (não exibido na imagem), tornando a análise financeira totalmente dinâmica e adaptável a diferentes cenários.

O componente visual central é a “Análise de Quadrantes” (Figura 15), um gráfico de dispersão que posiciona cada linha do sistema com base em duas métricas de eficiência: passageiros por quilômetro no eixo Y, caracterizando a eficiência espacial, e passageiros por minuto no eixo X, sendo esta a eficiência temporal. O tamanho de cada bolha é proporcional ao impacto da linha (total de passageiros). Linhas tracejadas, que representam

as médias do sistema, dividem o gráfico em quatro quadrantes distintos, cada um com um significado analítico, conforme detalhado nos *cards* de “Interpretação dos Quadrantes”. Esta visualização permite a classificação imediata de todas as linhas em perfis de desempenho: “As Melhores”, “Lentas e Lotadas”, “Rápidas e Vazias” e “Alvos de Otimização”.

Figura 15 – Gráfico de Análise de Quadrantes, classificando as linhas por eficiência espacial e temporal



Fonte: Captura de tela realizada pelo autor

Para aprofundar a capacidade de exploração do gráfico de quadrantes, a interface oferece um painel de “Controles de Visualização” que permite ao usuário refinar os dados exibidos. O filtro por linhas permite selecionar múltiplas linhas específicas, restringindo a exibição do gráfico apenas a elas e favorecendo análises comparativas mais detalhadas. Já o filtro por quadrante possibilita isolar um dos quatro perfis de desempenho, como “As Melhores” ou “Alvos de Otimização”, focando a análise em um grupo específico de linhas. Adicionalmente, o controle “Mostrar Top Linhas” permite reduzir a sobrecarga visual ao exibir apenas as linhas mais relevantes com base no total de passageiros transportados (e.g., Top 10, Top 25). Esses mecanismos, combinados com os filtros globais da página, conferem ao usuário um alto grau de controle sobre a análise, facilitando a identificação de padrões e exceções.

A divisão do gráfico em quatro quadrantes permite interpretar o desempenho das linhas de acordo com perfis específicos, articulando tanto a perspectiva da gestão pública quanto a do usuário do transporte:

- **As Melhores (Quadrante Superior Direito):** Caracterizam-se por alta eficiência tanto no eixo espacial (passageiros por quilômetro) quanto no temporal (passageiros

Figura 16 – Cards interpretativos de cada quadrante



Fonte: Captura de tela realizada pelo autor

por minuto). São as linhas mais performáticas, pois transportam grandes volumes de passageiros em trajetos relativamente curtos e rápidos. Do ponto de vista da gestão, representam o melhor retorno sobre o investimento, otimizando a utilização da frota e servindo como modelos a serem estudados e replicados.

- **Lentas e Lotadas (Quadrante Superior Esquerdo):** Apresentam alta eficiência espacial, mas baixa eficiência temporal. Essas linhas transportam muitos passageiros, mas sofrem com velocidades reduzidas, geralmente devido a gargalos de infraestrutura viária ou tráfego intenso. Para os gestores, são “linhas de batalha” do sistema, essenciais para a rede, mas que exigem intervenções externas (como faixas exclusivas ou prioridade semaforica) para alcançar melhor desempenho. Na perspectiva do usuário, o tempo de viagem elevado compromete a qualidade do serviço, mesmo com a ampla disponibilidade da linha.
- **Rápidas e Vazias (Quadrante Inferior Direito):** São linhas com alta eficiência temporal, mas baixa eficiência espacial. Realizam trajetos rapidamente, mas transportam relativamente poucos passageiros. Do ponto de vista da gestão, podem indicar subutilização de recursos ou desenho inadequado da rede, já que o custo de operação pode não se justificar frente à baixa demanda. Para o usuário, no entanto, representam linhas ágeis e confortáveis, ainda que pouco representativas no sistema como um todo, podendo atender nichos específicos ou rotas mal integradas.
- **Alvos de Otimização (Quadrante Inferior Esquerdo):** Concentrando baixa eficiência nos dois eixos, essas linhas apresentam baixo aproveitamento operacional e reduzida atratividade para o usuário. São candidatas prioritárias para reavaliação

de rota, frequência ou mesmo descontinuidade, de modo a realocar recursos em áreas de maior necessidade. Para a gestão, representam desperdício de capacidade; para o usuário, são pouco funcionais, não atendendo de maneira satisfatória às suas demandas de mobilidade.

(a) Tooltip com resumo da Linha 202



(b) Modal com análise financeira detalhada da mesma linha

Figura 17 – Recursos do *dashboard* para exibir as informações das linhas

A interatividade é um elemento chave nesta análise. Ao passar o mouse sobre uma linha no gráfico, um *tooltip* exibe suas principais métricas de desempenho e o resultado financeiro estimado (Figura 17a). Ao clicar em uma linha, um modal com os detalhes da linha é aberto (Figura 17b), apresentando um dossiê completo com todos os indicadores, incluindo o custo operacional, a receita estimada e o resultado líquido (lucro ou subsídio), permitindo uma investigação aprofundada.

Figura 18 – Rankings de “Destaques da Análise”, que sumarizam os principais achados do estudo de caso



Fonte: Captura de tela realizada pelo autor

Finalmente, a seção destaques da análise (Figura 18) sintetiza os resultados, apresentando rankings das cinco principais linhas em categorias como as mais eficientes por distância, maior impacto e principal alvo de otimização. Essa sumarização guia o usuário

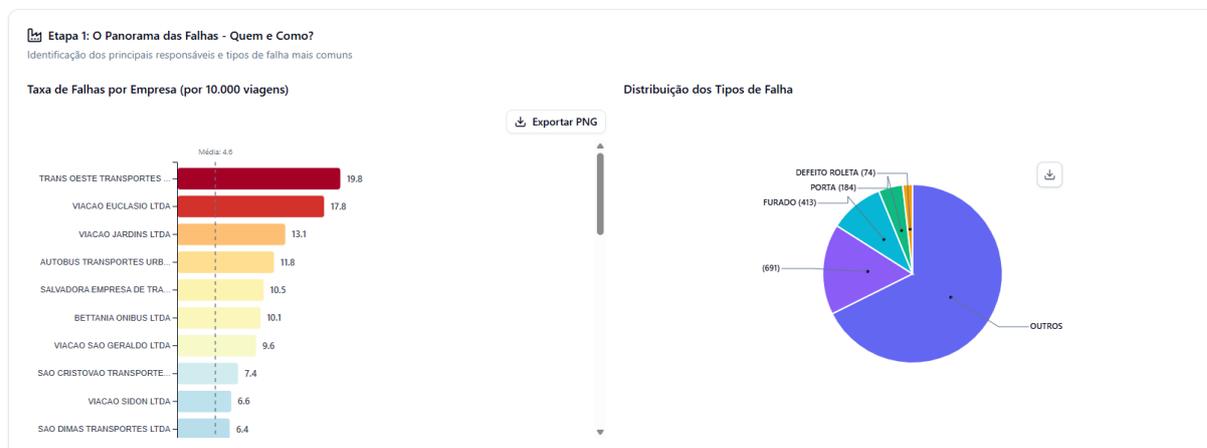
para os *insights* mais relevantes gerados pela ferramenta, concluindo a narrativa do estudo de caso sem impor uma única conclusão, mas sim destacando as evidências encontradas nos dados.

4.2.2 Estudo de Caso: Distribuição das Falhas Mecânicas

O segundo estudo de caso aborda um problema crítico para a operação do transporte público: as falhas mecânicas. O objetivo desta análise é investigar se tais eventos seguem padrões identificáveis, correlacionando-os com as empresas, a idade e o uso da frota, e as rotas operadas. A página dedicada a este estudo, ilustrada nas figuras a seguir, foi estruturada em três etapas sequenciais para guiar o usuário desde um panorama geral até uma análise de causas e consequências.

A seção superior da página contextualiza a análise, assim como no estudo de caso “Eficiência das Linhas”. Ela apresenta os KPIs gerais para o período selecionado, como o total de falhas e o total de veículos analisados, e utiliza *cards* de texto para expor a pergunta de pesquisa e detalhar as métricas utilizadas na investigação, como o total de quilômetros percorridos e a taxa de falhas a cada 10.000 km.

Figura 19 – Etapa 1 da análise, mostrando o ranking de empresas por taxa de falhas e a distribuição dos tipos de falha

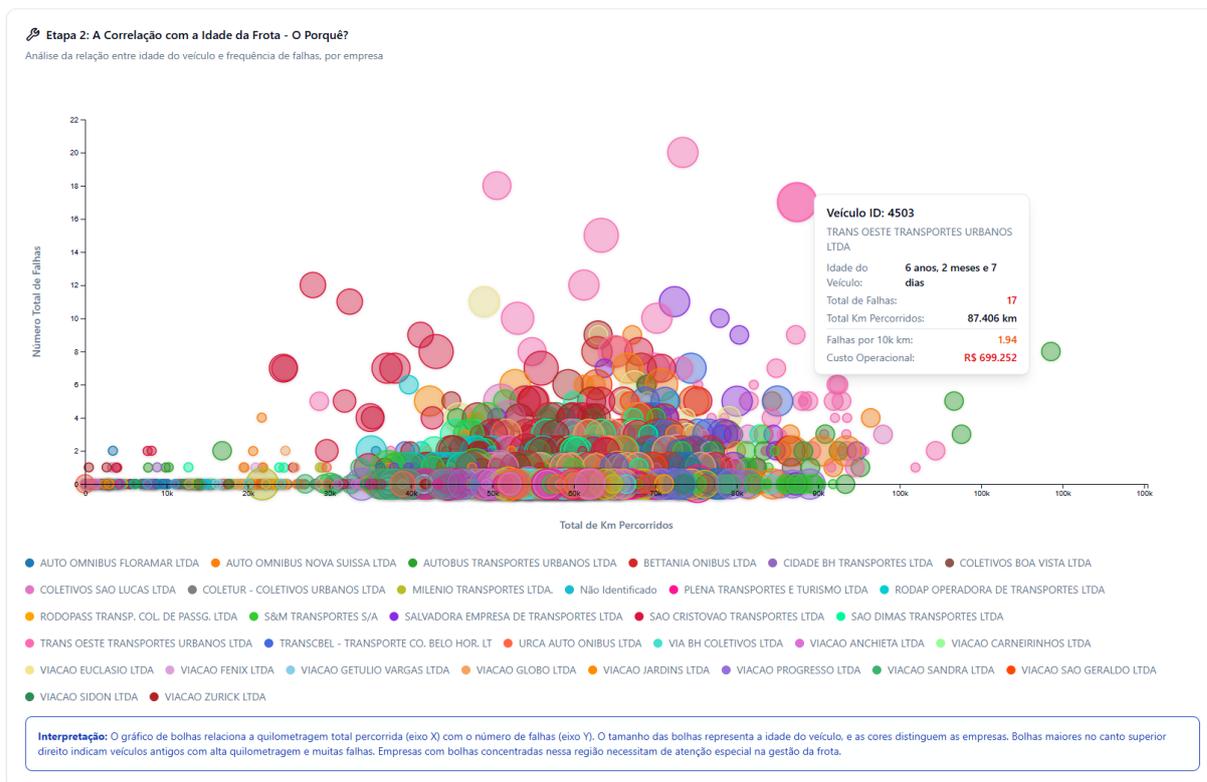


Fonte: Captura de tela realizada pelo autor

A primeira etapa, denominada “O Panorama das Falhas”, ilustrada pela figura 19, foca em responder “Quem e Como?”. Um gráfico de barras horizontais exhibe a taxa de falhas por empresa (por 10.000 viagens), permitindo uma comparação justa entre as operadoras e a identificação daquelas com maior incidência proporcional de problemas. Ao lado, um gráfico de pizza detalha a distribuição dos tipos de falha, mostrando as causas mais comuns, como problemas de motor ou pneus, para todas as ocorrências no período.

A segunda etapa, “A Correlação com a Idade da Frota” (Figura 20), aprofunda a investigação sobre o “Porquê?” das falhas. O componente central é um gráfico de bolhas

Figura 20 – Etapa 2 da análise, com o gráfico de bolhas correlacionando quilometragem, falhas e idade dos veículos



Fonte: Captura de tela realizada pelo autor

interativo que correlaciona o desgaste dos veículos com a frequência de problemas. Dessa forma, o eixo X representa o total de quilômetros percorridos, caracterizando o desgaste por uso, o eixo Y o número total de falhas, e o tamanho de cada bolha representa a idade do veículo, sendo este o desgaste por tempo. Esta visualização permite identificar padrões complexos, como veículos novos que falham muito (bolhas pequenas no canto superior esquerdo) ou veículos antigos e robustos que falham pouco (bolhas grandes no canto inferior direito).

Além disso, ao passar o mouse sobre uma bolha, um *tooltip* exibe informações detalhadas do veículo, incluindo seu custo operacional estimado, complementando a análise visual.

Figura 21 – Painel de filtros interativos do gráfico de bolhas da Etapa 2

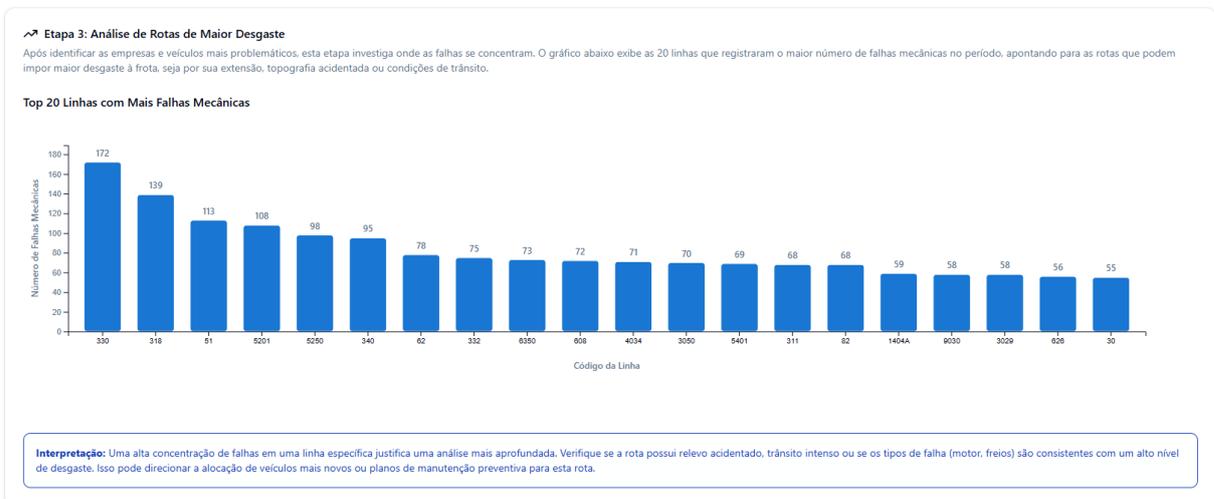


Fonte: Captura de tela realizada pelo autor

Para permitir um estudo mais detalhado e a investigação de cenários específicos, a ferramenta dispõe de um painel de filtros interativos, apresentado na Figura 21. Por

meio desses controles, o usuário pode segmentar a visualização por empresas específicas, definir uma idade mínima para os veículos analisados e ajustar dinamicamente as faixas de “Quantidade de Falhas” e “Km Percorridos” através de seletores de intervalo (*sliders*). Adicionalmente, opções de seleção permitem ocultar do gráfico os veículos com a informação “Não Identificado” ou aqueles com zero falhas, focando a análise nos ativos problemáticos. A combinação destes filtros possibilita ao analista isolar variáveis e aprofundar a investigação sobre as causas dos problemas, testando hipóteses como, por exemplo, se a frota de uma determinada empresa com quilometragem elevada apresenta uma frequência de falhas acima da média.

Figura 22 – Etapa 3, com o ranking de linhas mais afetadas, e a seção final de “Destaques da Análise”



Fonte: Captura de tela realizada pelo autor

Por fim, a terceira etapa, denominada “Análise de Rotas de Maior Desgaste” (Figura 22), focaliza a dimensão espacial da análise. Após a identificação dos veículos e empresas com pior desempenho, esta seção apresenta um gráfico de barras com as vinte linhas com mais falhas mecânicas. A visualização aponta para as rotas que podem impor maior estresse à frota, seja por sua topografia, extensão ou condições de trânsito. Um *card* de interpretação guia o usuário a correlacionar essa informação com os tipos de falha, sugerindo ações como a alocação de veículos mais novos para essas rotas. A página conclui com os destaques da análise, que sumarizam os principais *insights* encontrados, como as empresas mais problemáticas e os veículos mais afetados.

4.2.3 Estudo de Caso: Horários de Utilização

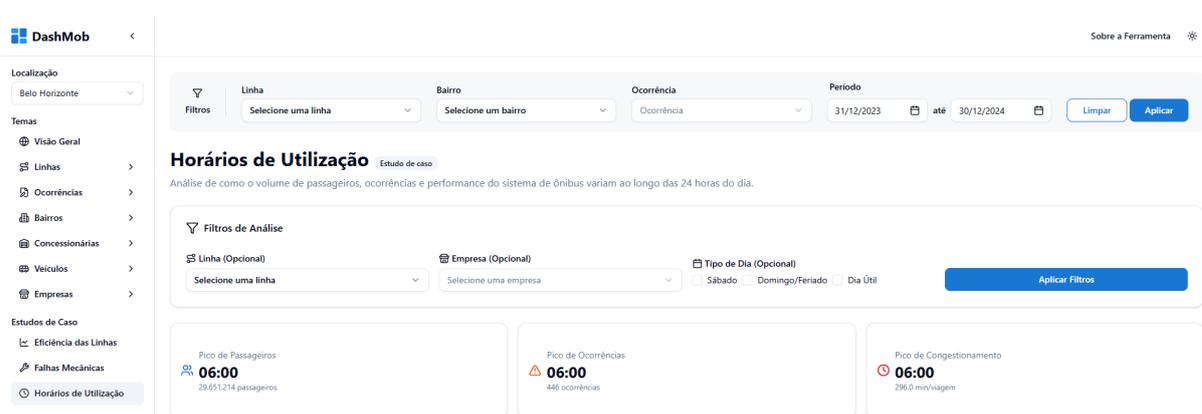
O terceiro e último estudo de caso foca na análise da dimensão temporal do sistema de transporte, investigando como o “pulso” da cidade varia ao longo das 24 horas do dia. O objetivo desta análise é responder à pergunta: “Como o volume de passageiros, a

incidência de ocorrências e a performance do sistema se comportam nos diferentes horários, e como esses padrões mudam de acordo com o tipo de dia, linha ou empresa?”.

A página dedicada a este estudo (Figura 23) foi projetada para oferecer uma visão clara e comparativa dos padrões diários. A interface dispõe de filtros próprios que permitem ao usuário segmentar os dados não apenas pelo período, mas também opcionalmente por Linha, Empresa e Tipo de Dia (Dia Útil, Sábado, Domingo/Feriado). Essa flexibilidade permite desde uma análise macro de todo o sistema até um *drill-down* focado no comportamento de uma única linha em dias úteis, por exemplo.

No topo da análise, são apresentados três KPIs que identificam automaticamente os horários críticos do período selecionado: o “Pico de Passageiros”, o “Pico de Ocorrências” e o “Pico de Congestionamento”, este último medido pela maior duração média de viagem.

Figura 23 – Visão geral da página do estudo de caso “Horários de Utilização”, com os filtros de análise e os KPIs de horários de pico

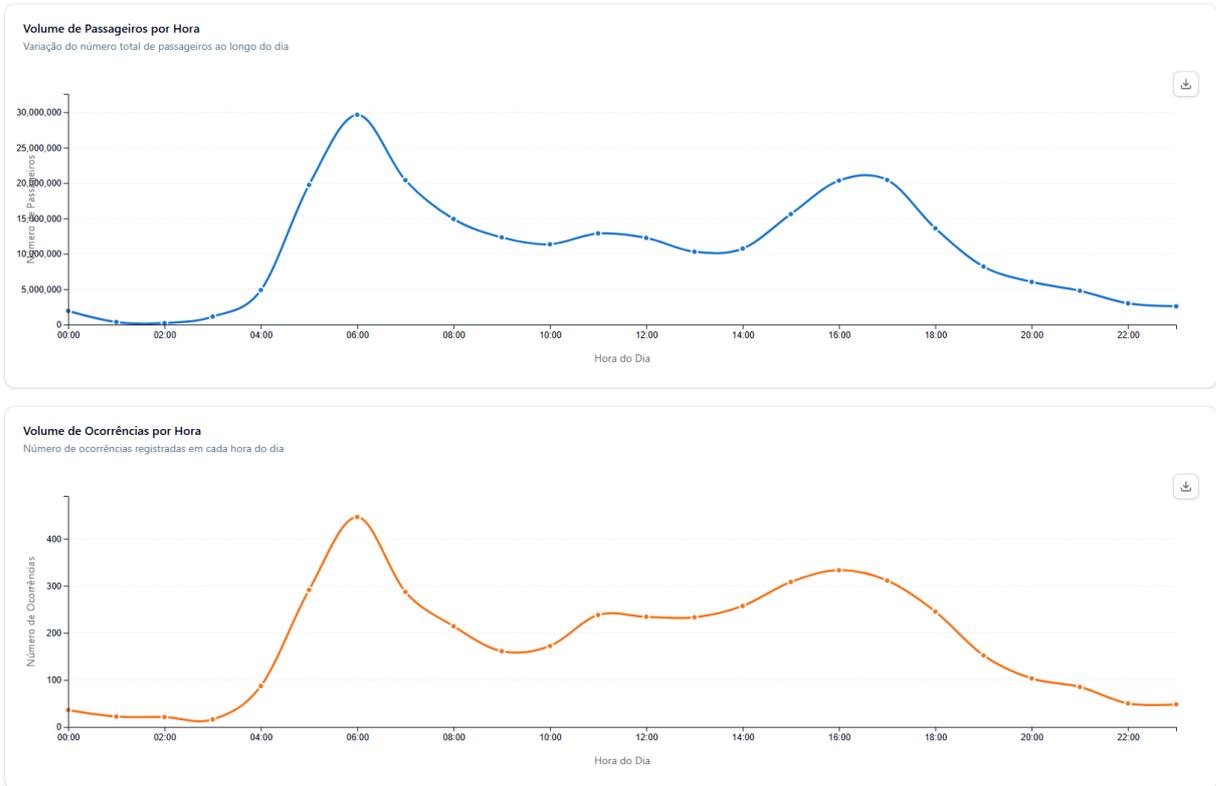


Fonte: Captura de tela realizada pelo autor

O corpo principal da análise é composto por três gráficos de linha que detalham a variação horária das principais métricas operacionais (Figura 24 e 25). O primeiro gráfico, “Volume de Passageiros por Hora”, exibe a curva de demanda do sistema, permitindo a clara identificação dos horários de pico da manhã e da tarde. O segundo, “Volume de Ocorrências por Hora”, plota a frequência de problemas ao longo do dia, possibilitando a correlação entre o aumento do estresse no sistema e a incidência de falhas.

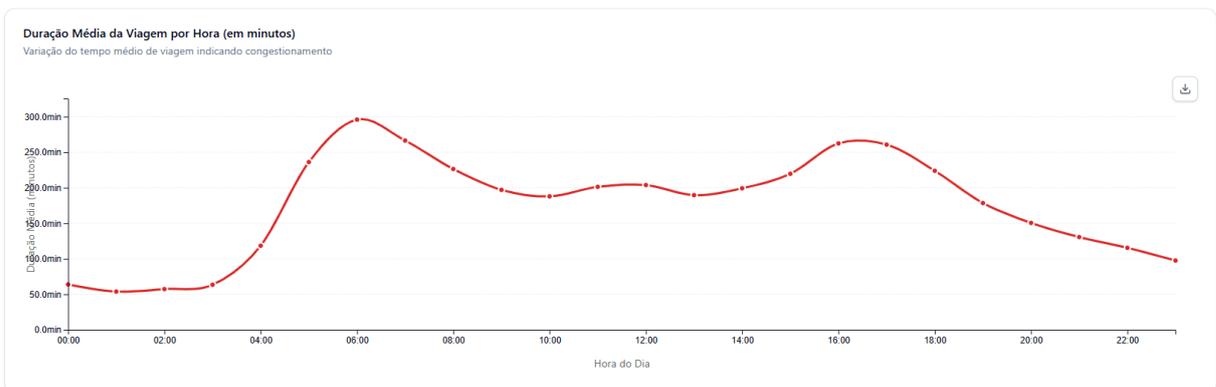
Finalmente, o terceiro gráfico, “Duração Média da Viagem por Hora”, funciona como um indicador direto do impacto do congestionamento no trânsito. A variação desta métrica ao longo do dia expõe os períodos em que as viagens se tornam mais lentas, afetando a pontualidade e a eficiência do transporte. A interatividade dos gráficos, com *tooltips* que exibem os valores exatos de cada hora, permite uma exploração detalhada dos dados. Em conjunto, essas visualizações fornecem um diagnóstico completo do ritmo operacional do sistema de transporte de Belo Horizonte.

Figura 24 – Gráficos de tendência temporal para o volume de passageiros e ocorrências ao longo do dia



Fonte: Captura de tela realizada pelo autor

Figura 25 – Gráficos de tendência temporal para a duração média das viagens ao longo do dia



Fonte: Captura de tela realizada pelo autor

4.3 Validação da Solução

A validação da solução *DashMobi* foi realizada sob três perspectivas: a performance técnica da API, a usabilidade da interface e o atendimento aos objetivos específicos definidos para o projeto. O objetivo desta seção é demonstrar a eficácia da arquitetura e do design propostos na criação de uma ferramenta analítica robusta e funcional.

4.3.1 Performance da API

Um dos principais desafios técnicos deste trabalho era garantir que uma aplicação interativa pudesse consultar um volume de dados de quase 100 milhões de registros com tempos de resposta baixos. A abordagem de consultar diretamente a tabela de fatos (`fact_viagens`) se mostrou inviável durante os testes iniciais, com *endpoints* críticos, como o de correlação entre idade e falhas de veículos, levando vários minutos para retornar uma resposta.

A solução, conforme descrito na Seção 3.2.3, foi a criação de um conjunto de tabelas agregadas e views materializadas. Para validar a eficácia desta abordagem, foram medidos os tempos de resposta dos principais endpoints da API em um ambiente de desenvolvimento local, com o cache do navegador desabilitado. Os resultados confirmam a performance da solução:

- O endpoint da Visão Geral (`/api/v1/geral/kpis-periodo`), que resume os dados de todo o sistema para um período, apresentou um tempo de resposta médio de 120ms.
- A consulta para o mapa de calor (`/api/v1/geral/heatmap-bairros`), que agrega dados por bairro, respondeu em aproximadamente 600ms.
- O endpoint mais complexo, da análise de correlação de falhas (`/api/v1/estudos/falhas-mecanicas/correlacao-idade-veiculo`), que antes era o principal gargalo, agora apresenta um tempo de resposta médio de 700ms.

Esses tempos, na ordem de milissegundos, são adequados para uma aplicação web interativa e validam a decisão arquitetural de pré-agregar os dados, que se mostrou crucial para a viabilidade do projeto.

4.3.2 Usabilidade da Interface

A validação da usabilidade foi realizada através de uma análise qualitativa da interface final, focada na clareza, intuitividade e eficácia para a exploração dos dados. A organização do *dashboard* em módulos temáticos, acessíveis por um menu lateral, permite uma navegação clara entre os diferentes contextos de análise. A padronização da barra de filtros globais no topo de todas as páginas garante uma experiência de usuário consistente.

Os componentes visuais, como os KPIs e os gráficos, foram projetados com foco na legibilidade, utilizando títulos e legendas claras. A interatividade, como o uso de *tooltips* para exibir detalhes sob demanda (Figura 17a) e modais para análises aprofundadas (Figura 17b), permite que o usuário acesse camadas mais profundas de informação sem sobrecarregar a tela principal. A resposta visual da ferramenta à aplicação de filtros é praticamente instantânea, consequência direta da performance da API.

4.3.3 Atendimento aos Objetivos do Projeto

Finalmente, realiza-se a validação do trabalho em relação aos objetivos específicos delineados na Seção 1.2, atestando o êxito de cada etapa planejada:

- **Estruturar um banco de dados analítico:** Alcançado. A implementação do data warehouse com modelo estrela, complementado por tabelas de ponte e uma robusta camada de agregação (detalhada no Capítulo 3), atende plenamente a este objetivo, garantindo a performance necessária para as consultas analíticas.
- **Desenvolver um *back-end* seguro e escalável:** Alcançado. A API RESTful, desenvolvida com FastAPI, cumpre seu papel ao prover *endpoints* performáticos e bem-definidos que abstraem com sucesso a complexidade do banco de dados e entregam os dados de forma otimizada para a interface.
- **Criar uma interface clara e intuitiva:** Alcançado. Conforme evidenciado pelas telas apresentadas neste capítulo (Figuras 9 a 20), a interface do usuário foi construída com componentes visuais padronizados que traduzem dados complexos em uma apresentação clara, objetiva e interativa.
- **Proporcionar ferramentas de análise interativa:** Alcançado. A sinergia entre os filtros globais, os módulos temáticos e, principalmente, os estudos de caso guiados, materializa este objetivo. A plataforma oferece ao usuário um ambiente rico para a investigação de questões complexas, desde a eficiência das linhas até a correlação entre a idade da frota e as falhas mecânicas.

A eficácia da arquitetura de dados e o design da interface validam, portanto, a concepção do *DashMobi* não apenas como um visualizador, mas como uma plataforma de *business intelligence* funcional, performática e capaz de gerar *insights* analíticos sobre o sistema de transporte.

5 Conclusão

Este trabalho se propôs a enfrentar o desafio de transformar o vasto e complexo volume de dados abertos sobre mobilidade urbana de Belo Horizonte em uma ferramenta de análise visual, intuitiva e acionável. O resultado deste esforço é o *DashMobi*, um *dashboard* interativo que não apenas centraliza e exibe informações, mas também capacita gestores públicos, planejadores urbanos e cidadãos a extrair conhecimento e *insights* significativos sobre o sistema de transporte coletivo da cidade.

A principal contribuição deste projeto reside na concepção e implementação de uma solução de *Business Intelligence* de ponta a ponta. O processo abrangeu desde a rigorosa engenharia de dados, com a construção de uma pipeline de ETL para o tratamento de quase 100 milhões de registros, até a modelagem de um *data warehouse* analítico em *star schema*. A decisão arquitetural de utilizar múltiplas camadas de agregação, como as tabelas *agg_* e as *views* materializadas, mostrou-se fundamental para o sucesso do projeto, resolvendo os gargalos de performance e permitindo que a API respondesse a consultas complexas em milissegundos, validando a eficácia da arquitetura proposta.

No âmbito da visualização de dados, o *DashMobi* demonstrou como a aplicação de princípios de *Data Storytelling*, através dos estudos de caso guiados, pode transformar a análise de dados em uma narrativa investigativa. A ferramenta provou ser capaz de responder a perguntas complexas sobre a eficiência operacional das linhas, a distribuição de falhas mecânicas e os padrões de utilização do sistema ao longo do dia, transformando dados brutos em evidências visuais para a tomada de decisão.

É fundamental, contudo, reconhecer as limitações inerentes ao estudo, que derivam principalmente do escopo e da granularidade dos dados de origem disponibilizados. Primeiramente, a ausência de dados sobre a geometria dos itinerários e a sequência ordenada dos pontos de parada impede a representação visual exata das rotas em um mapa; a análise espacial fica, portanto, restrita à distribuição dos pontos, sem a capacidade de visualizar o trajeto completo das linhas. Adicionalmente, as “ocorrências” (como falhas mecânicas ou desvios) são registradas por viagem, mas carecem de informações temporais e geográficas precisas, o que impede a identificação de “*hotspots*” de problemas. Outra carência notável reside na contagem de passageiros, pois não há dados consolidados sobre o volume de embarques em cada ponto de ônibus individualmente, restringindo análises de demanda em nível micro. Por fim, o detalhamento cadastral dos veículos é limitado, sem informações como ano de fabricação ou quilometragem total, o que torna indireta a análise da relação entre a idade da frota e seu desempenho.

Conclui-se, portanto, que os objetivos propostos foram plenamente alcançados.

O *DashMobi* se firma como uma contribuição relevante para a modernização da gestão pública em Belo Horizonte, oferecendo um modelo de como os dados abertos podem ser potencializados por meio de tecnologia para promover um transporte mais eficiente, transparente e alinhado às necessidades da população.

5.1 Trabalhos Futuros

Embora funcional e robusto, o *DashMobi* foi projetado como uma plataforma com grande potencial de evolução. As lições aprendidas durante o desenvolvimento e a análise dos dados permitem delinear as seguintes direções para trabalhos futuros:

- **Aprofundamento da Análise de Equidade:** Desenvolver novos estudos de caso para investigar outras questões pertinentes, como a correlação entre a densidade populacional dos bairros e a oferta de transporte, ou a relação entre a renda per capita e a qualidade do serviço.
- **Evolução para Controle Operacional em Tempo Real:** Integrar a plataforma com APIs de dados em tempo real da BHTrans. Isso transformaria o *DashMobi* de uma ferramenta primariamente estratégica (análise histórica) para uma ferramenta também tática, permitindo o monitoramento ao vivo da frota, o controle de pontualidade e a gestão de desvios.
- **Implementação de Modelos Preditivos:** Utilizar o rico histórico de dados da *fact_viagens* para treinar modelos de *machine learning*. Seria possível desenvolver modelos de previsão de séries temporais para a demanda de passageiros ou modelos de classificação para estimar a probabilidade de falhas mecânicas em veículos, com base em sua idade e histórico de uso.
- **Enriquecimento da Análise Geoespacial:** Adicionar novas camadas de dados geográficos, como informações de relevo, uso do solo ou infraestrutura cicloviária. Além disso, a integração com APIs externas de mapas permitiria agregar dados dinâmicos sobre congestionamentos, obras na via e acidentes, enriquecendo a análise de eventos operacionais. Isso possibilitaria análises mais profundas sobre como o ambiente urbano impacta a operação, o consumo de combustível e a integração entre diferentes modais de transporte.
- **Integração com Dados de *Crowdsourcing*:** Desenvolver um aplicativo móvel complementar que permita aos usuários contribuir com informações em tempo real. Através deste canal, seria possível coletar dados sobre a percepção da qualidade do serviço, como lotação dos veículos, condições de limpeza, pontualidade e segurança nos pontos de parada. Essa camada de dados qualitativos enriqueceria as análises

operacionais, oferecendo uma visão mais completa e centrada na experiência do passageiro.

- **Módulo de Simulação e Planejamento Interativo:** Criar uma seção na ferramenta que permita a gestores e planejadores urbanos simular o impacto de alterações nas rotas. Os usuários poderiam propor modificações, como a inclusão ou remoção de pontos de parada ou a alteração de um itinerário, e a plataforma calcularia os efeitos dessas mudanças em indicadores-chave, como a cobertura populacional, o tempo de viagem estimado e a demanda potencial. Isso transformaria o *DashMobi* de uma ferramenta de diagnóstico para uma plataforma de apoio ao planejamento participativo.

Essas sugestões representam caminhos naturais para a evolução do *DashMobi*, solidificando-o ainda mais como um observatório completo e indispensável para a mobilidade urbana de Belo Horizonte.

Referências

- ANDRADE, J. N.; GALVÃO, D. C. et al. O conceito de smart cities aliado à mobilidade urbana. *Revista Hum@nae*, v. 10, n. 1, 2016. Citado na página 13.
- BONETTE, L. R.; REIS, J. G. M. d. Composição, concentração e classificação do transporte urbano de massa inteligente e sustentável em quatro cidades inteligentes. *Research, Society and Development*, Research, Society and Development, v. 10, n. 13, p. e37101320634, Oct 2021. Disponível em: <<https://rsdjournal.org/index.php/rsd/article/view/20634>>. Citado na página 13.
- BRAGA, L. V. et al. O papel do governo eletrônico no fortalecimento da governança do setor público. *Revista do Serviço Público*, v. 59, n. 1, p. 05–21, 2008. Citado na página 13.
- CARD, S. Information visualization. In: *Human-Computer Interaction*. [S.l.]: CRC press, 2009. p. 199–234. Citado na página 18.
- CARVALHO, C. H. R. de. *Desafios da mobilidade urbana no Brasil*. [S.l.], 2016. Citado na página 13.
- CORDEIRO, A. et al. Governo eletrônico e redes sociais: informação, participação e interação. *Reciis*, v. 6, p. 1001, 2012. Disponível em: <<https://api.semanticscholar.org/CorpusID:142957273>>. Citado na página 14.
- CRUZ, S. M. S. da; ANDRADE, L. S.; SAMPAIO, J. O. Explorando dados abertos governamentais sobre a mobilidade urbana na cidade do rio de janeiro. In: SBC. *Seminário Integrado de Software e Hardware (SEMISH)*. [S.l.], 2016. p. 1645–1655. Citado na página 24.
- DELFINI, L.; SPAHIU, B.; VIZZARI, G. Visual analytics for sustainable mobility: Usability evaluation and knowledge acquisition for mobility-as-a-service (maas) data exploration. *Digital*, MDPI, v. 4, n. 4, p. 821–845, 2024. Citado na página 24.
- DINIZ, V. Como conseguir dados governamentais abertos. 2010. Citado na página 19.
- DU, F. et al. Metroviz: Visual analysis of public transportation data. *arXiv preprint arXiv:1507.05215*, 2015. Citado na página 24.
- FERREIRA, L. S.; FERRETTO, D.; DUARTE, D. H. S. Mapeando a morfologia urbana: aplicações do método local climate zones (lcz) em diferentes escalas do projeto. *PosFAUUSP*, Universidade de São Paulo. Agência de Bibliotecas e Coleções Digitais, v. 30, n. 56, p. e197563, Jan 2023. Disponível em: <<https://revistas.usp.br/posfau/article/view/197563>>. Citado na página 13.
- FERRER-SAPENA, A.; PESET, F.; ALEIXANDRE-BENAVENT, R. Acceso a los datos públicos y su reutilización: open data y open government. *Profesional de la información*, v. 20, n. 3, p. 260–269, 2011. Citado na página 19.
- FEW, S. *Information dashboard design: The effective visual communication of data*. [S.l.]: O'Reilly Media, Inc., 2006. Citado 2 vezes nas páginas 14 e 18.

FILHO, L. B.; FLORENTINO, P. V. *Dashboard para Dados sobre a Mobilidade Urbana em Salvador*. Salvador, Brasil: [s.n.], 2022. Citado na página 24.

Governo do Estado do Rio de Janeiro. *Portal de Dados Abertos do Estado do Rio de Janeiro*. 2024. <<https://dadosabertos.rj.gov.br/>>. Acesso em: 18 de ago. de 2024. Citado na página 21.

MARTINS, T. et al. Map matching: Uma análise de dados streaming de trajetórias de gps no transporte público. In: *Anais Estendidos do XVIII Simpósio Brasileiro de Sistemas de Informação*. Porto Alegre, RS, Brasil: SBC, 2022. p. 294–301. ISSN 0000-0000. Disponível em: <https://sol.sbc.org.br/index.php/sbsi_estendido/article/view/21604>. Citado na página 13.

MEIRELES, A. V. et al. Democracia digital: consultas públicas interativas. *Design e Tecnologia*, v. 8, n. 15, p. 81–95, 2018. Citado na página 14.

Ministério das Cidades, Governo Federal do Brasil. *SIMU - Sistema Nacional de Informações em Mobilidade Urbana*. 2024. <<https://simu.cidades.gov.br/>>. Acesso em: 18 de ago. de 2024. Citado na página 22.

PAULO, R. F. *Crescimento urbano desordenado: o papel do Estado e da sociedade diante dos impactos socioambientais*. [S.l.]: Editora Fi, 2018. Citado na página 13.

PERO, V.; STEFANELLI, V. A questão da mobilidade urbana nas metrópoles brasileiras. *Revista de economia contemporânea*, SciELO Brasil, v. 19, p. 366–402, 2015. Citado na página 13.

Prefeitura de Belo Horizonte. *Belo Horizonte é primeira no ranking do Índice de Dados Abertos para Cidades (ODI)*. 2024. <<https://prefeitura.pbh.gov.br/politica-urbana/informes-tecnicos/belo-horizonte-e-primeira-no-ranking-do-indice-de-dados-abertos-para-cidades-odi>>. Acesso em: 18 de ago. de 2024. Citado na página 21.

Prefeitura Municipal de Curitiba. *Dados Abertos de Curitiba*. 2024. <<https://dadosabertos.curitiba.pr.gov.br/>>. Acesso em: 18 de ago. de 2024. Citado na página 21.

São Paulo Transporte S/A. *API do Olho Vivo – Guia de Referência*. 2024. <<https://www.sptrans.com.br/desenvolvedores/api-do-olho-vivo-guia-de-referencia/>>. Acesso em: 18 de ago. de 2024. Citado na página 21.

SANTORO, F. de O.; SPIEGEL, T.; JÚNIOR, D. S. da S. O uso de modelos e técnicas de ux na criação de dashboards de bi: Uma revisão sistemática da literatura. *LUMEN ET VIRTUS*, v. 15, n. 42, p. 7017–7035, 2024. Citado na página 18.

SANTOS, R. M. dos; VASCONCELOS, R. O. A atuação dos sistemas de transportes inteligentes no transporte público via ônibus impulsionado pelo 5g. Citado na página 14.

SATO, S. N. *A infografia na divulgação científica: um estudo de caso da revista Pesquisa FAPESP*. Tese (Doutorado) — Universidade de São Paulo, 2017. Citado na página 18.

SOUZA, A. M. de; VILLAS, L. A. Controle de congestionamento de veículos utilizando sistemas de transporte inteligentes. In: SBC. *Concurso de Teses e Dissertações (CTD)*. [S.l.], 2017. p. 2379–2384. Citado na página 14.

SUCENA, M. P.; CURY, M. V. Q. Relações fuzzy subsidiando o diagnóstico do transporte público pelo iqmu. *Revista Brasileira de Planejamento e Desenvolvimento*, v. 11, n. 2, p. 556–577, 2022. Citado na página 13.

União de Ciclistas do Brasil (UCB). *Mobilidados: Dados sobre Mobilidade Ativa no Brasil*. 2024. <<https://mobilidados.org.br/>>. Acesso em: 18 de ago. de 2024. Citado na página 22.