

Pontualidade do Transporte Público de Curitiba: Uma Visão a Partir de Dados Abertos

Lucas L. da Silva¹, Mariana G. da Luz¹, Nádia P. Kozievitch¹, Rita Berardi¹

¹Universidade Tecnológica Federal do Paraná (UTFPR)
Av. Sete de Setembro, 3165 - Rebouças - CEP 80230-901 - Curitiba - PR

{lucsil.1997,marianagomesluz}@alunos.utfpr.edu.br

{nadiap,ritaberardi}@utfpr.edu.br

Abstract. *This work aims to analyze the punctuality of bus lines Interbairros II and Santa Cândida/Capão Raso from Curitiba as they are different types of bus lines and have different route lengths. For this, algorithms were adapted for detecting the direction of displacement of buses, mapping the passage times of vehicles on their route, and punctuality of the lines. The results obtained showed that the methodology can be generalized to other bus lines, with different bus stops and vehicles.*

Resumo. *Este trabalho tem como objetivo a análise da pontualidade das linhas de ônibus Interbairros II e Santa Cândida/Capão Raso de Curitiba por serem de diferentes tipos de linhas e possuírem diferentes extensões de itinerário. Para isso, são adaptados os algoritmos para detecção da direção de deslocamento dos ônibus, mapeamento dos horários de passagem dos veículos em sua rota e pontualidade das linhas. Os resultados obtidos mostraram que a metodologia pode ser generalizada outros tipos de linhas de ônibus, com pontos de ônibus e veículos diferenciados.*

1. Introdução

A partir de 2007 a área urbana passou a ser maior do que a área rural em termos populacionais no mundo [United Nations 2019]. O aumento demográfico representa não só uma alteração na dinâmica populacional dos países, mas também um desafio para que políticas de gestão pública absorvam e supram as demandas dessa população em crescimento.

No Brasil, a Lei Nº 12.527 ¹, de 18 de Novembro de 2011, trata da abertura dos dados públicos para acesso e utilização de qualquer pessoa física. Na cidade de Curitiba, a regulamentação foi feita através do Decreto Nº 1.135 ², de 30 de Julho de 2012, garantindo aos cidadãos o direito constitucional de acesso às informações públicas da gestão municipal. Para isso, a prefeitura criou dois portais principais, o Portal da Transparência ³ e a Base de Dados Abertos de Curitiba ⁴. Uma das bases de dados disponíveis no site é do transporte coletivo do município, de responsabilidade da mesma companhia que o gerencia: Urbanização de Curitiba S/A - mais conhecida pelo acrônimo URBS.

¹http://www.planalto.gov.br/ccivil_03/_Ato201120142011LeiL12527.htm

²<https://mid.curitiba.pr.gov.br/2018/00228941.pdf>

³<https://www.transparencia.curitiba.pr.gov.br>

⁴<https://www.curitiba.pr.gov.br/dadosabertos>

Em Curitiba, equipamentos de geolocalização (GPS) estão instalados nos ônibus desde 2012⁵ e dados históricos estão disponibilizados a partir de 2017 referentes às rotas, veículos, pontos de parada e deslocamentos podem ser acessados a partir da base aberta de Curitiba. Segundo uma pesquisa do Instituto de Pesquisa e Planejamento Urbano de Curitiba (IPPUC), das famílias com domicílio em Curitiba e cidades da região metropolitana, 36,4% das pessoas utilizava o transporte coletivo como forma de locomoção diária⁶. A imp pontualidade e imprevisibilidade das linhas pode não apenas gerar insatisfação dos usuários com o serviço, mas também afetar diretamente os compromissos diários da população que dependem desse serviço como, por exemplo, para deslocamento ao local de trabalho. Mesmo com o auxílio de aplicativos para acompanhamento de linhas de ônibus o uso do transporte público de Curitiba, ainda conta com a imp pontualidade em suas linhas, agravada pela falta de acompanhamento em tempo real para seus usuários.

Este trabalho tem como objetivo a análise da pontualidade das linhas de ônibus 020 - *Interbairros II* e 203 - *Santa Cândida/Capão Raso* da cidade de Curitiba - Paraná. Ambas possuem diferentes veículos, roteiros e extensões de itinerário. Para isso, são adaptados os algoritmos propostos por [Martins et al. 2022] para detecção da direção de deslocamento dos ônibus, mapeamento dos horários de passagem dos veículos em sua rota e pontualidade das linhas.

Este artigo está organizado da seguinte maneira: a Seção 2 apresenta trabalhos relacionados. A metodologia é apresentada na Seção 3. O desenvolvimento é apresentado na Seção 4 enquanto a Seção 5 conclui o artigo.

2. Trabalhos Relacionados

Ao analisar-se a movimentação de veículos com dados de GNSS (GPS), alguns problemas ocorrem, e são eles [Millard-Ball et al. 2019]:

- Falta de dados, quando o veículo não salvou ou não enviou dados em uma frequência adequada para se ter um entendimento correto da sua movimentação geográfica.
- Imprecisões do dispositivo GNSS (GPS). Como toda ferramenta de medição, os dispositivos de GPS sofrem imprecisões e suas causas podem ser: objetos no caminho (prédios, montanhas), interferências eletromagnéticas, falhas de hardware, hardware de baixa qualidade, falhas nos satélites.

Considerando estes problemas é possível dizer que as leituras de GPS de um carro em movimento nem sempre serão registrados no lado correto da rua ou até mesmo dentro dos limites da rua. Isso dificulta alguns tipos de análises onde essas imprecisões podem atrapalhar [Millard-Ball et al. 2019].

Em particular, o sistema transporte de público de Curitiba é complexo, possuindo vários tipos de linhas, veículos e itinerários. Previamente, uma das linhas (216 - *Cabral/Portão*), foi analisada [Martins et al. 2022], com dados de 01/05/2019 a 14/05/2019, sendo identificado inconsistências. O diferencial desta análise advém do uso de *Map Matching* para analisar a problemática das imprecisões das fontes de dados georreferenciadas

⁵<https://www.curitiba.pr.gov.br/noticias/site-da-urbs-mostra-deslocamento-do-onibus-em-tempo-real/30756>

⁶https://ippuc.etools.com.br/storage/uploads/175a0bce-18b1-49a4-8886efaccccd851/consolidacaode_dados.pdf

sobre o tempo de chegada do ônibus [Martins et al. 2022]. Assim, foi possível identificar uma tendência de parada adiantada durante o período da manhã e atrasada nos períodos tarde e noite para os dois pontos (Terminal do Cabral e Terminal do Portão).

Os mesmos dados abertos de mobilidade já foram utilizados em uma análise exploratória [Parcianello et al. 2018], observando aspectos como forma de disponibilização, integridade e acurácia. Os autores listaram os seguintes problemas nos dados: a) Granularidade: diferentes níveis de detalhes são utilizados pelas bases; b) Consistência: conflito nos tipos de dados; c) Acurácia: diferentes formas de armazenar um mesmo dado; e d) Diferença na quantidade e profundidade de dados nas bases sobre as mesmas entidades do mundo real.

De maneira resumida, os dados foram utilizados para análise exploratória [Parcianello et al. 2018] (revelando a grande quantidade de dados e a riqueza das informações que deles podem ser extraídos), origem-destino [Parcianello 2019] (analisando o deslocamento de cidadãos), a integração do ponto de vista de *Map Matching* [Martins et al. 2022], e clusterização [Vila 2016].

Nesta direção, podemos citar ainda o trabalho sobre destinos e origens mais comuns na cidade [Diniz Junior 2017], com a análise das velocidades dos ônibus. Um característica detectada pelo trabalho é que, na análise realizada com dados de Curitiba, a velocidade média é menor a medida que o veículo se aproxima do centro da cidade.

A visualização destas informações pode ser utilizada principalmente de duas formas diferentes: análise exploratória e análise explanatória, de acordo com [Knafllic 2015]. Os dados do tipo espaciais e temporais possuem algumas características próprias e que são comumente ignoradas nas análises. Essas especificidades são discutidas por [Keim et al. 2010], como dependência entre observações, incerteza, escala, e tempo.

Como é possível notar nos trabalhos apresentados, a disponibilidade de dados abertos do transporte público não é garantia de sucesso nos objetivos de um estudo sem antes realizar os devidos pré-processamentos para tratamento e limpeza dos dados, além da exploração da própria base para encontrar possíveis problemas durante sua manipulação e a utilização de algoritmos para enriquecimento da informação. Além disso, não foi encontrado nenhum trabalho anterior a [Martins et al. 2022], avaliando os atrasos dos ônibus com relação as suas tabelas horárias.

3. Metodologia

Este trabalho tem como objetivo a análise da pontualidade das linhas de ônibus “020” e “203” de Curitiba por serem de diferentes tipos de linhas e possuírem diferentes extensões de itinerário. As linhas da cidade possuem códigos únicos para sua identificação e determinam exatamente o itinerário pelo qual os veículos devem percorrer. Por exemplo, a linha “Interbairros II (horário)” é identificada unicamente pelo código “020” e define a rota circular pela cidade que pode ser observada em verde na Figura 1. Já a linha expressa “Santa Cândida/Capão Raso” é identificada pelo código único “203”, define a rota direta ligando dois terminais importantes da cidade (que dão o nome à linha) e também pode ser observada na Figura 1 em vermelho.

Para isso, são adaptados os algoritmos propostos por [Martins et al. 2022] para detecção da direção de deslocamento dos ônibus, mapeamento dos horários de passagem

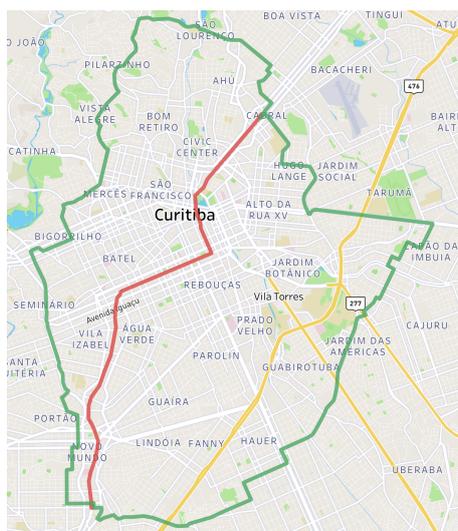


Figura 1. Itinerários da linhas “020 - Interbairros II (horário)” (em verde) e “203 - Sta Cândida/Capão Raso” (em vermelho)

dos veículos em sua rota e pontualidade das linhas. A metodologia é apresentada na Figura 2, com as seguintes etapas: mapeamento das áreas de conhecimento, análise de base de dados abertos, desenvolvimento do algoritmo de importação de dados, adaptação do algoritmo de [Martins et al. 2022] de forma a comparar diferentes linhas do transporte público a partir da mesma perspectiva, e a conclusão.

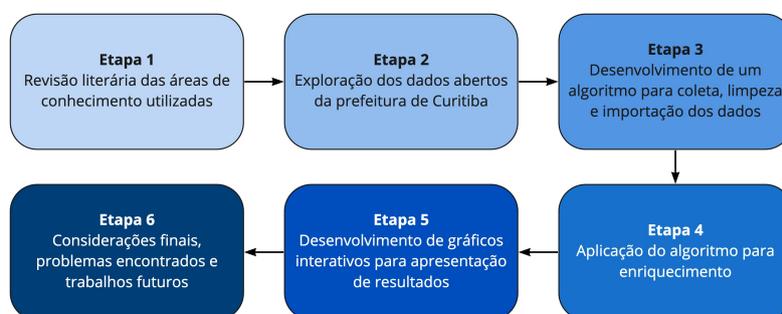


Figura 2. Metodologia.

As seguintes tecnologias foram utilizadas: *PostgreSQL* 14.2, *Python* 3.9, *Pandas* 1.5.0⁷, *Tableau*⁸. Os códigos desenvolvidos durante o projeto estão disponíveis em *GitHub*⁹.

4. Desenvolvimento

A base de dados históricos aberta é estruturada em oito arquivos diários diferentes, cada um contendo os dados do dia de sua respectiva tabela. Este trabalho utilizou os dados do período de 01/05/2019 a 14/05/2019.

⁷<https://pandas.pydata.org>

⁸<https://www.tableau.com/pt-br>

⁹[casluleal/urbs-open-data-analysis](https://github.com/casluleal/urbs-open-data-analysis)

A principal tabela desse trabalho está presente no arquivo “veiculos.json”. Ela armazena a posição (latitude e longitude) e *timestamp* (data e hora) dos veículos da frota a partir de seu equipamento GPS. Sua estrutura pode ser vista na Tabela 1.

Tabela 1. Estrutura da tabela *vehicle_position*, a partir do arquivo “veiculos.json”

Original	Remapeado	Tipo	Definição
VEIC	vehicle_id	text	Identificador do veículo
LAT	latitude	double precision	Latitude da posição no instante
LON	longitude	double precision	Longitude da posição no instante
DTHR	timestamp	timestamp	Data e hora do instante
COD_LINHA	bus_line_id	text	Identificador da linha de ônibus

O segundo arquivo de interesse é “linhas.json” (que armazena informações sobre cada uma das linhas de ônibus), seguido por “pontosLinha.json” (dados geográficos e outras informações sobre cada um dos pontos de ônibus), “shapeLinha.json” (caminhos exatos que os ônibus devem fazer em cada trecho de trajeto de uma linha), “tabelaVeiculo.json” (horário programado que cada veículo) e “tabelaLinha.json” (cronograma de passagem de algum ônibus da linha por determinado ponto).

Para analisar o período de 01/05/2019 a 14/05/2019 foram necessários oitenta e quatro arquivos. Por esse motivo, foi desenvolvido um algoritmo para o processo de *download*, descompressão, tratamento e importação dos dados.

O primeiro problema encontrado para o desenvolvimento do algoritmo de importação foi as estruturas diferentes dos arquivos disponíveis no servidor. Além disso, podemos citar, no arquivo “veiculos.json”, a falta de um caractere “{” no início do arquivo do dia 10/05/2019, tornando o objeto JSON incompreensível pelo interpretador da linguagem utilizada. O arquivo descompactado com falha (sem modificação do autor) é apresentado na Figura 3 e causou erro no programa de importação. Detalhes adicionais podem ser verificados em [De Lara and Stapenhorst Martins 2022].

```

> luuca > Projects > urbs-open-data-analysis > tmp > { } 2019_05_10_veiculos.json
1
2 "VEIC": "MB300", "LAT": "-25.424256", "LON": "-49.242276", "DTHR": "09\05\2019 19:13:43", "COD_LINHA": "021" }
3 {"VEIC": "MB300", "LAT": "-25.424531", "LON": "-49.242368", "DTHR": "09\05\2019 19:13:40", "COD_LINHA": "021"}
4 {"VEIC": "MB300", "LAT": "-25.424876", "LON": "-49.242428", "DTHR": "09\05\2019 19:13:37", "COD_LINHA": "021"}
5 {"VEIC": "MB300", "LAT": "-25.42542", "LON": "-49.242358", "DTHR": "09\05\2019 19:13:32", "COD_LINHA": "021"}
6 {"VEIC": "MB300", "LAT": "-25.425583", "LON": "-49.242311", "DTHR": "09\05\2019 19:13:30", "COD_LINHA": "021"}
7 {"VEIC": "MB300", "LAT": "-25.425881", "LON": "-49.242225", "DTHR": "09\05\2019 19:13:09", "COD_LINHA": "021"}

```

Figura 3. Arquivo “veiculos.json” do dia 10/05/2019 com falta de caractere do objeto JSON.

Dentre os problemas de importação, podemos citar arquivos vazios e dados faltantes, quantidade de veículos em operação diferem dos veículos programados, veículos de uma linha operando em outra linha, problemas com amostragem de GPS, comportamento de amostragem com ônibus “dando ré”, imprecisões no cadastro do ponto de ônibus e redundância de dados de colunas em arquivos diferentes [Martins et al. 2022].

4.1. Algoritmos de Enriquecimento de Dados Utilizados

Os algoritmos adaptados de [Martins et al. 2022] são compostos por três etapas: 1) Extração de Azimutes (ângulo ou direção de passagem de um ônibus) na apresentado na Figura 4; 2) Descoberta do Momento da Passagem do veículo no ponto de ônibus, apresentado na Figura 5 e 3) Comparação de Horário de Passagem Real do veículo e Programado.

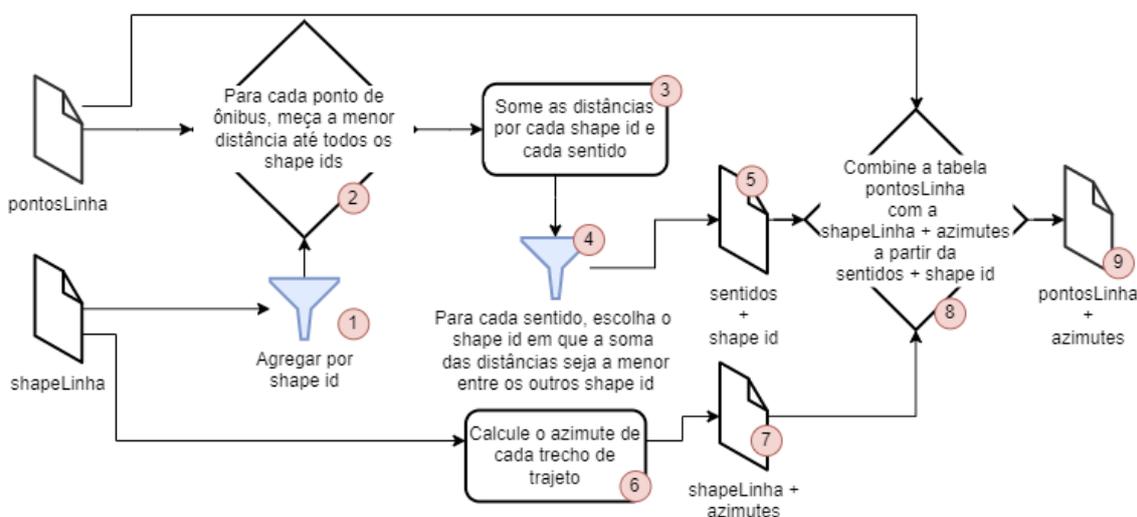


Figura 4. Algoritmo para a Extração de Azimutes (ângulo ou direção de passagem de um ônibus). Fonte [Martins et al. 2022]

Para utilizar o código do autor nas tabelas desse trabalho foram necessárias algumas adaptações, já que o nome das tabelas e colunas criadas são diferentes, ainda que possuam semelhanças. Além disso, a diferença na estrutura de importação impactou, uma vez que [Martins et al. 2022] utilizou tabelas individuais para cada arquivo e aqui foram utilizadas tabelas únicas para diferentes dias do mesmo arquivo.

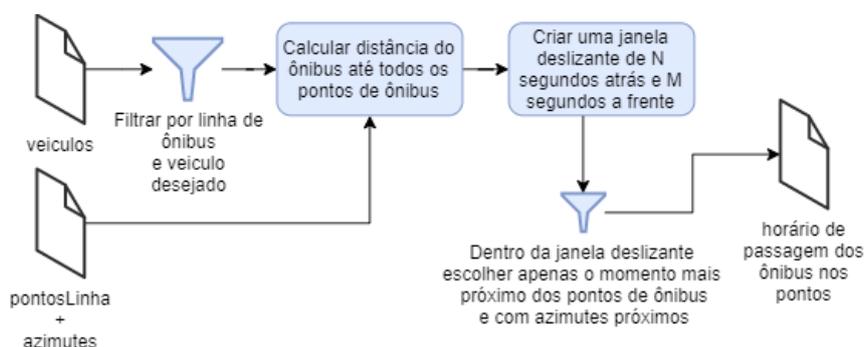


Figura 5. Algoritmo para identificação do momento de passagem dos veículos nos pontos. Fonte [Martins et al. 2022]

Como exemplo de alteração no código, podemos citar a utilização de *common table expressions* (CTE)¹⁰ para a manipulação e tratamento dos dados, criando diversas colunas virtuais intermediárias para serem consumidas na parte final do algoritmo.

¹⁰<https://www.postgresql.org/docs/current/queries-with.html>

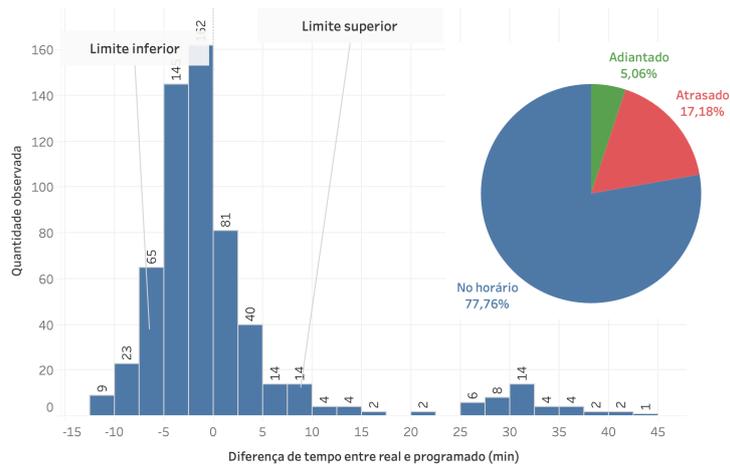


Figura 7. Diferenças entre horário real e programado para a linha “020”

Foi possível observar que a maior parte das passagens mapeadas da linha “020” estão no intervalo aceitável de 7,5 minutos de atraso e adiantamento: 77,8% no horário. Para a linha “216 - Cabral/Portão” (analisada no trabalho anterior [Martins et al. 2022]), o mesmo intervalo de tolerância da diferença representou 63,7% dos dados. Dessa forma, a linha “020” apresentou uma pontualidade maior do que a linha “216” em aproximadamente 13 pontos percentuais.

Já para a linha “203” o resultado foi muito mais disperso, como podemos observar na Figura 8. Temos muito mais resultados do algoritmo nessa linha, já que existem mais pontos com horário programado na URBS, porém existem também muito mais atrasos e adiantamentos do que as outras duas. Aqui temos 45,2% de passagens pontuais, uma queda significativa em relação às linhas “020” e “216” de aproximadamente 32,6 p.p. e 18,5 p.p, respectivamente. Além disso, os demais valores mapeados são distribuídos em 23% de adiantamentos e 31,8% de atrasos.

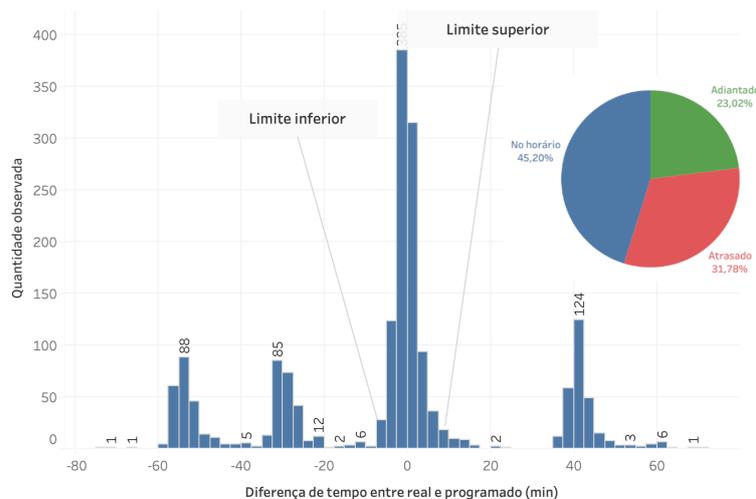


Figura 8. Diferenças entre horário real e programado para a linha “203”

Vale também ressaltar que a linha “020” possui um itinerário de aproximadamente 42 quilômetros de extensão, enquanto a linha “203” possui apenas 16 quilômetros de

extensão. De maneira resumida, durante o dia foi percebido que nos horários de pico além do aumento de veículos se tem um aumento no atraso. Também foi verificado que durante a manhã é mais provável um adiantamento e durante a noite é mais provável o atraso.

Dentre as limitações desse trabalho, podemos citar a falta de documentação, a presença de horários para somente alguns pontos de ônibus em algumas linhas, a acessibilidade dos dados no site, a qualidade da informação e até mesmo os formatos e estruturas dos arquivos disponibilizados.

5. Conclusão

Esse trabalho apresentou um estudo sobre a pontualidade do transporte público de Curitiba, baseando-se na pesquisa de [Martins et al. 2022], com o objetivo de analisar a pontualidade das linhas de ônibus. O método proposto foi aplicado em duas linhas de diferentes tipos: “020 - Interbairros II (horário)” e “203 - Santa Cândida/Capão Raso” para o mesmo período de 2019.

O trabalho realizou uma análise exploratória dos dados abertos e da pontualidade das linhas e como documentação da estrutura de alguns arquivos da base de dados abertos utilizada e apresentação de um método de importação automatizado desses dados para um banco de dados local.

Como trabalhos futuros, propõe-se a inclusão de outros dados e linhas de ônibus. Como recomendação às autoridades, a principal recomendação é a criação de uma documentação sobre a estrutura da base, descrevendo tabelas, colunas, o tipo de dados que armazenam, além de mais informações sobre os equipamentos utilizados, intervalos de coleta e cobertura dentro da frota de veículos da tabela.

6. Agradecimentos

Os autores gostariam de agradecer a Universidade Tecnológica Federal do Paraná (DIREC 10/2021), o IPPUC, URBS e a Prefeitura de Curitiba.

Referências

- De Lara, F. R. and Stapenhorst Martins, T. (2022). Pontualidade do Transporte Público de Curitiba: uma Visão a partir de Dados Abertos. Monografia (Bacharel em Informática), UTFPR.
- Diniz Junior, P. C. (2017). Serviços Telemáticos em uma Rede de Transporte Público Baseados em Veículos conectados e Dados Abertos. Master's thesis, UTFPR, Curitiba.
- Keim, D., Kohlhammer, J., Ellis, G., and Mansmann, F. (2010). *Mastering the information age: solving problems with visual analytics*. Goslar: Eurographics Association.
- Knaflic, C. N. (2015). *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons.
- Martins, T. S., Kozievitch, N. P., Gadda, T., Rosa, M., and Gutierrez, M. B. (2022). Map Matching: Uma análise de Dados Streaming de trajetórias de GPS no Transporte Público, SBSI 2022, Trilha Emergente de Cidades Inteligentes.

- Millard-Ball, A., Hampshire, R. C., and Weinberger, R. R. (2019). Map-matching poor-quality GPS data in urban environments: the pgMapMatch package. *Transportation Planning and Technology*, 42(6):539–553.
- Parcianello, Y. (2019). Análise de origem-destino do uso do sistema de transporte coletivo de Curitiba sob o ponto de vista de regions of interest. Master's thesis, UTFPR.
- Parcianello, Y., Kozevitch, N. P., Fonseca, K. V., Rosa, M. d. O., Gadda, T. M., and Malucelli, F. C. (2018). Transportation: An overview from open data approach. In *2018 IEEE International Smart Cities Conference (ISC2)*, pages 1–8. IEEE.
- United Nations (2019). *World Urbanization Prospects: The 2018 Revision*. United Nations, Department of Economic and Social Affairs, Population Division.
- Vila, J. J. F. R. (2016). Clusterização e Visualização Espaço-teporal de Dados Georreferenciados Adaptando o Algoritmo Marker Clusterer - um caso de uso em Curitiba. Master's thesis, UTFPR, Curitiba.