

## **IDENTIFICAÇÃO DO PERFIL DE DIREÇÃO DE MOTORISTA NO TRANSPORTE PÚBLICO DE CURITIBA**

felipe snitynski camillo<sup>1</sup>, matheus scheibel<sup>3</sup>, marcella scoczynski martins<sup>2</sup>

**RESUMO:** O presente estudo visa identificar padrões e desafios no transporte público de Curitiba, focando nas linhas do sistema *Bus Rapid Transit*. Curitiba é referência mundial em mobilidade urbana sustentável, e seu sistema é um modelo amplamente replicado em outras cidades. Contudo, como em qualquer sistema de transporte de grande escala, a cidade enfrenta problemas relacionados à pontualidade e à eficiência das linhas, especialmente nos horários de pico. A pesquisa utilizou técnicas de aprendizado de máquina, com ênfase em algoritmos de clusterização, para analisar dados de cinco linhas principais. Dessas, três eram do tipo expressa (303 - Centenário/Campo Comprido, 503 - Boqueirão e 203 - Santa Cândida/Capão Raso) e duas do tipo Ligeirões (250 - Ligeirão Norte/Sul e 500 - Ligeirão Boqueirão/Carlos Gomes). As linhas foram selecionadas pela sua importância na conexão de diferentes regiões ao centro da cidade. A metodologia incluiu os processos de extração de dados e sua aplicação. Com a obtenção dos dados, foi realizada uma filtragem para remover ruídos e informações irrelevantes, como viagens sem atrasos significativos. O algoritmo K-Means foi aplicado para identificar aglomerações de atrasos e possíveis padrões que pudessem estar contribuindo para esses atrasos. Os resultados revelaram que os atrasos se concentram majoritariamente na região central de Curitiba, além das saídas dos terminais, como Boqueirão e Santa Cândida. A análise visual, utilizando mapas sobrepostos com a biblioteca Folium, permitiu identificar os principais pontos críticos, facilitando o planejamento de intervenções. Áreas densamente povoadas, com maior fluxo de passageiros, apresentaram os maiores desafios operacionais. A utilização de aprendizado de máquina e análise de dados foi essencial para identificar padrões ocultos e fornecer insights sobre como otimizar o transporte público da cidade.

**Palavras-chave:** Aprendizado de Máquina, Transporte Público, Análise de padrões.

---

1 Autor, UTFPR,, felipesnitynski@alunos.utfpr.edu.br

2 Orientadora, UTFPR, marcella@utfpr.edu.br

3 Co-autor, UTFPR, matheusscheibel@alunos.utfpr.edu.br

### **1 INTRODUÇÃO**

O BRT de Curitiba é caracterizado por corredores exclusivos para ônibus, estações-tubo que facilitam o embarque e desembarque de passageiros, e uma operação que prioriza a rapidez e a eficiência dos deslocamentos. Esse sistema não só melhorou a qualidade de vida dos moradores, reduzindo os tempos de viagem, mas também serviu de modelo para inúmeras outras cidades ao redor do mundo. Além do BRT (*Bus Rapid Transit*), o

transporte público de Curitiba é composto por linhas de ônibus convencionais, alimentadoras e interbairros, que garantem a cobertura de praticamente toda a cidade. A integração tarifária permite que os usuários façam transferências entre diferentes linhas sem custo adicional, tornando o sistema mais econômico e conveniente.

Dado as dificuldades relacionadas a extração dos dados, foram analisadas apenas cinco linhas de ônibus, sendo três do tipo expressa (303 - Centenário/Campo Comprido, 503 - Boqueirão e 203 - Santa Cândida/Capão Raso) e duas do tipo expresso ligeirão (250 - Ligeirão Norte/Sul e 500 - Ligeirão Boqueirão/Carlos Gomes). Essas linhas foram selecionadas, pois integram diferentes regiões ao centro da cidade, como se observa na figura 1, contendo os respectivos trechos itinerários analisados na pesquisa.

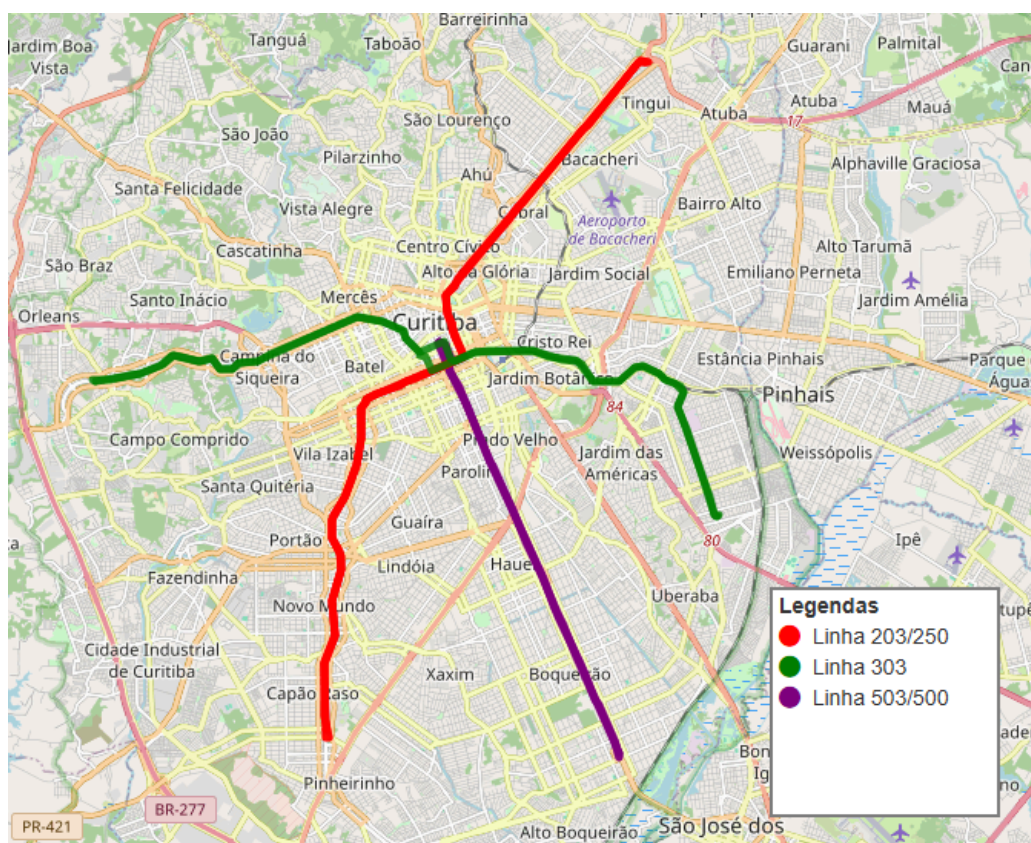


Figura 1: Mapa dos trechos itinerários analisados.

## 1.1 Objetivos

O objetivo desta pesquisa é identificar padrões no transporte público de Curitiba. Para isso, foram utilizadas ferramentas computacionais para a extração e modelagem de dados, além de processos de aprendizado de máquina para detectar padrões relevantes em diferentes linhas de ônibus.

## 1.2 Justificativa

Em um ambiente urbano com crescente densidade populacional, a pontualidade do transporte público é crucial para reduzir congestionamentos e emissões de CO<sub>2</sub>. Segundo o Instituto de Pesquisa Econômica Aplicada (Ipea), investimentos em transporte coletivo são fundamentais para mitigar a emissão de gases de efeito estufa e controlar os níveis de poluição, já que um usuário de automóvel emite significativamente mais CO<sub>2</sub> do que um usuário de transporte público.

## 2 METODOLOGIA

Nesta seção, são descritos a metodologia e o processo de desenvolvimento. Primeiramente, detalha-se a extração dos dados, seguida pela filtragem e modelagem. Por fim, apresenta-se a aplicação de técnicas de *machine learning* para a busca de padrões.

### 2.1 Extração dos dados

Para a extração dos dados, utilizou-se a linguagem Python como principal ferramenta. Os dados foram fornecidos pela URBS (Urbanização de Curitiba S/A) por meio de uma URL específica, previamente dada por cadastramento no Webservice da URBS. Ao acessar a URL com o identificador da linha requisitada, eram retornados dados detalhados no formato JSON (JavaScript Object Notation) sobre os veículos desta linha.

Após a etapa de coleta dos dados, foram desenvolvidos algoritmos específicos para acrescentar as informações associadas às linhas de ônibus. Esses algoritmos adicionaram dados complementares, tais como:

- **HORA:** Hora no formato ano-mês-dia hh:mm:ss;
- **PONTO MAIS PRÓXIMO:** Ponto de parada do respectivo ônibus mais próximo do veículo no momento da captura;
- **DISTANCIA MINIMA:** Distância entre o veículo e o ponto de parada mais próximo.
- **SEQ:** Número inteiro que representa a sequência de pontos
- **SEQ MAX:** Número inteiro que representa o número máximo de sequência.

Para executar o arquivo contendo os algoritmos para a coleta de dados, foi disponibilizada uma máquina virtual do cluster PROFIAP de Ponta Grossa, acessada via

terminal utilizando o software PuTTY, com a conexão sendo estabelecida por meio de um SSH (*Secure Shell*) por VPN (*Virtual Private Network*), a VPN foi conectada através do FortiClient VPN. A máquina virtual rodava um sistema Linux com Python 3 pré-instalado. O script foi executado em segundo plano (*background*) utilizando o comando *screen* durante um mês. Após a coleta dos dados, foi utilizado o software WinSCP para realizar a troca de dados entre a máquina local e a virtual, transferindo assim o arquivo no formato CSV (*Comma-Separated Values*) contendo os dados de ônibus, a tabela 1 resume os softwares e suas funções na coleta e transformação dos dados .

Tabela 1 – Softwares utilizados durante a pesquisa

Jupyter	Software da linguagem Python
FortiClient VPN	Software para conexão de VPN
PuTTY	Software para acessar um máquina virtual via SSH
WinSCP	Software para transferência de arquivos entre uma máquina local e uma virtual via SSH

Fonte: Autoria própria.

## 2.2 Filtragem dos dados

Após a finalização da coleta dos dados, é realizada uma filtragem dos dados. Foi criado um novo *data frame* com apenas os ônibus que estavam em situação de atraso. Em seguida, foi aplicado um algoritmo que filtra apenas o primeiro momento em que o ônibus atrasou. Essa etapa se destaca por possibilitar uma análise posterior sobre os ônibus atrasados.

## 2.3 Aplicação do método K-Means

De acordo com Morettin (2020), em seu livro Estatística e Ciência de Dados, o processo de clusterização K-means, tem como objetivo agrupar pontos em K grupos, minimizando a soma dos quadrados das distâncias entre os pontos e os centros dos clusters. Esse método, pertencente à classe de algoritmos de partição, utiliza centróides e exige que o espaço de dados seja euclidiano. No ambiente Python, a função `kmeans()` do pacote `sklearn`, baseada no algoritmo de Hartigan e Wong (1979), foi utilizada para



criar os agrupamentos. A essência do algoritmo é minimizar a variação dentro de cada grupo, o que, geralmente, corresponde à soma dos quadrados das distâncias euclidianas entre os pontos de um grupo e seu centróide. O processo segue alguns passos:

- i) Definir K e escolher K pontos iniciais;
- ii) Considerar esses pontos como centróides;
- iii) Atribuir os dados ao grupo com o centróide mais próximo;
- iv) Recalcular os centróides para cada grupo após a atribuição;
- v) Iterar o processo até que os centróides não mudem significativamente.

Após a filtragem dos dados, é possível obter localizações iniciais dos atrasos de cada veículo. Com esses dados, é possível aplicar o algoritmo de clusterização K-Means para determinar os centróides. A partir das médias de latitude e longitude de cada cluster, é possível sobrepor esses centróides em um mapa utilizando a biblioteca Folium, visualizando assim os pontos com maiores concentrações de atrasos. O mapa gerado destaca as áreas com os índices mais altos de atrasos.

### 3 RESULTADOS E DISCUSSÃO

A partir da discussão da seção anterior, foi gerado a figura 2, sendo o mapa contendo os locais com maiores índices de atrasos:

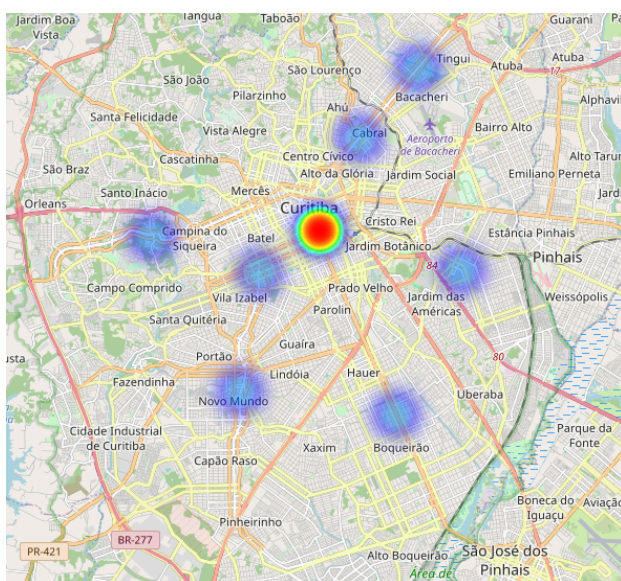


Figura 2: Mapa dos locais com mais atrasos das linhas analisadas

A partir do mapa gerado pela clusterização, é possível notar que a região central da cidade concentra a maioria dos atrasos. Além disso, é possível observar que os atrasos fora da região central ocorrem em sua maioria próximos a terminais.

#### **4 CONSIDERAÇÕES FINAIS**

A busca por padrões no Sistema BRT de Curitiba, através da clusterização, baseada em dados históricos e mapeamento dos atrasos, mostrou os principais gargalos no transporte público da cidade. As áreas com maior incidência de atrasos foram identificadas principalmente nas saídas dos terminais e na região central, destacando-se como pontos críticos que necessitam de intervenções para a melhoria do sistema.

A utilização de sistemas de informação e tecnologias de análise de dados mostrou-se essencial para identificar padrões de atraso e entender os locais que contribuem para a falta de pontualidade.

#### **AGRADECIMENTOS**

Meus agradecimentos a UTFPR e CNPq.

#### **REFERÊNCIAS**

MORETTIN, Pedro Alberto. *Estatística e ciência de dados*. São Paulo: Toaz.info, 2020.