



Ministério da Educação
Universidade Tecnológica Federal do Paraná
Pró-Reitoria de Pesquisa e Pós-Graduação

Relatório Final de Atividades de Iniciação Científica

Identificação do perfil de direção de motorista no transporte público de Curitiba

vinculado ao projeto

OTIMIZAÇÃO DE PROBLEMAS USANDO TÉCNICAS DE APRENDIZAGEM DE MÁQUINA DE ALGORITMOS EVOLUCIONÁRIOS

Felipe Snitynski Camillo  

Sistemas de Informação

Data de ingresso no programa: 01/11/2023

Prof(a). Dr(a). Marcella Scoczynski Ribeiro Martins  

Área do Conhecimento: 1.03.03.00-6 Metodologia e Técnicas da Computação

CÂMPUS PONTA GROSSA, 2024

FELIPE SNITYNSKI CAMILLO
MARCELLA SCOCZYNSKI RIBEIRO MARTINS

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

Relatório de Pesquisa do Programa de Iniciação
Científica da Universidade Tecnológica Federal
do Paraná.

PONTA GROSSA, 2024

SUMÁRIO

INTRODUÇÃO 3

METODOLOGIA 3

CONCLUSÕES 4

AGRADECIMENTOS 5

INTRODUÇÃO

O transporte público é uma peça fundamental para a mobilidade urbana, especialmente em grandes cidades como Curitiba. Com a crescente demanda por eficiência e pontualidade, a análise de padrões de funcionamento, como atrasos nas rotas de ônibus, tornou-se um ponto central para melhorar a qualidade do serviço. A identificação de pontos críticos, onde os atrasos são mais frequentes, pode auxiliar no desenvolvimento de políticas que visam otimizar o sistema e aumentar a satisfação dos usuários.

Neste trabalho, utilizamos dados em tempo real fornecidos pela URBS (Urbanização de Curitiba S/A), responsável pela operação do transporte público da cidade, para investigar as ocorrências de atrasos ao longo das rotas de ônibus. Foram extraídos dados de localização e estado operacional dos veículos, e essas informações foram analisadas utilizando técnicas de aprendizado de máquina. O foco principal foi a identificação de padrões de atraso e a clusterização dos locais onde esses atrasos ocorrem com maior frequência.

A metodologia empregada inclui a coleta automatizada de dados ao longo de um mês, seguida da aplicação de algoritmos de clusterização, como o k-means++. Essa abordagem permitiu identificar os principais pontos da cidade onde ocorrem os atrasos, fornecendo uma base sólida para futuras melhorias no sistema de transporte.

METODOLOGIA

Foi utilizado a linguagem Python, através do software Jupyter para extrair os dados fornecidos pela URBS, que estavam disponíveis em formato JSON em uma página web, essas informações, geralmente atualizadas a cada minuto, eram exibidas diretamente na página HTML. Para acessar os dados de uma linha específica, era preciso alterar um argumento na URL da página.

Armazenamos esses dados em um DataFrame da biblioteca Pandas, semelhante a uma tabela do Excel, foi criado um algoritmo personalizado para facilitar a extração dos dados de cada linha. Assim, ao inserir o número da linha desejada, o algoritmo retornava um DataFrame com informações em tempo real dessa linha, contendo os seguintes detalhes:

- **COD:** Prefixo do veículo
- **REFRESH:** Hora da Atualização (HH24:MI)
- **LAT:** Latitude (ponto flutuante)
- **LON:** Longitude (ponto flutuante)
- **CODIGOLINHA:** Prefixo da Linha (3 caracteres)
- **TABELA:** Tabela que o veículo está executando
- **SITUACAO:** Situação processada do veículo

Essas informações eram ainda insuficientes para realizar uma análise robusta dos locais onde mais ocorriam atrasados ao longo do dia, por isso foi necessário criar algoritmos para acrescentar dados no dataframe, sendo: algoritmo para localizar a estação mais próxima do ônibus e sua distância da estação, algoritmo para atribuir a sequência de estações da linha, começando com SEQ em 1 para a primeira estação da linha e indo até SEQ-MAX para a estação final da linha

Para coletar os dados ao longo de um dia inteiro foi necessário utilizar uma máquina virtual do cluster PROFIAP disponibilizado pela UTFPR de Ponta Grossa, foi necessário usar do comando screen, o qual cria uma janela no Linux que permite programas rodarem em background mesmo após fechar a conexão SSH para conectar-se a máquina virtual.

O algoritmo coletou dados ao longo de um mês, gerando uma base de dados robusta para iniciar o processo de clusterização, para uma análise exploratória dos locais onde mais ocorrem atrasos ao longo da linha de Ônibus.

Foi utilizado o método k-means++ para criação do cluster, sendo selecionado na base de dados apenas os ônibus com a situação atrasado e sendo selecionado sobre essa base um algoritmo para filtrar as posição inicial de atraso do ônibus, ou seja, em qual localidade o ônibus saiu do argumento no horário para atrasado, passando suas respectivas coordenadas para iniciar o processo de inicialização do cluster. Para determinar o número de centroides do cluster, foi utilizado o método do cotovelo, o qual nos gera um gráfico de distorção entre o número de clusters e sua distorção dos dados.

CONCLUSÕES

Através da discussão feita na seção anterior, foi possível gerar o seguinte mapa contendo os centroides(locais) com as maiores ocorrências de atrasos:

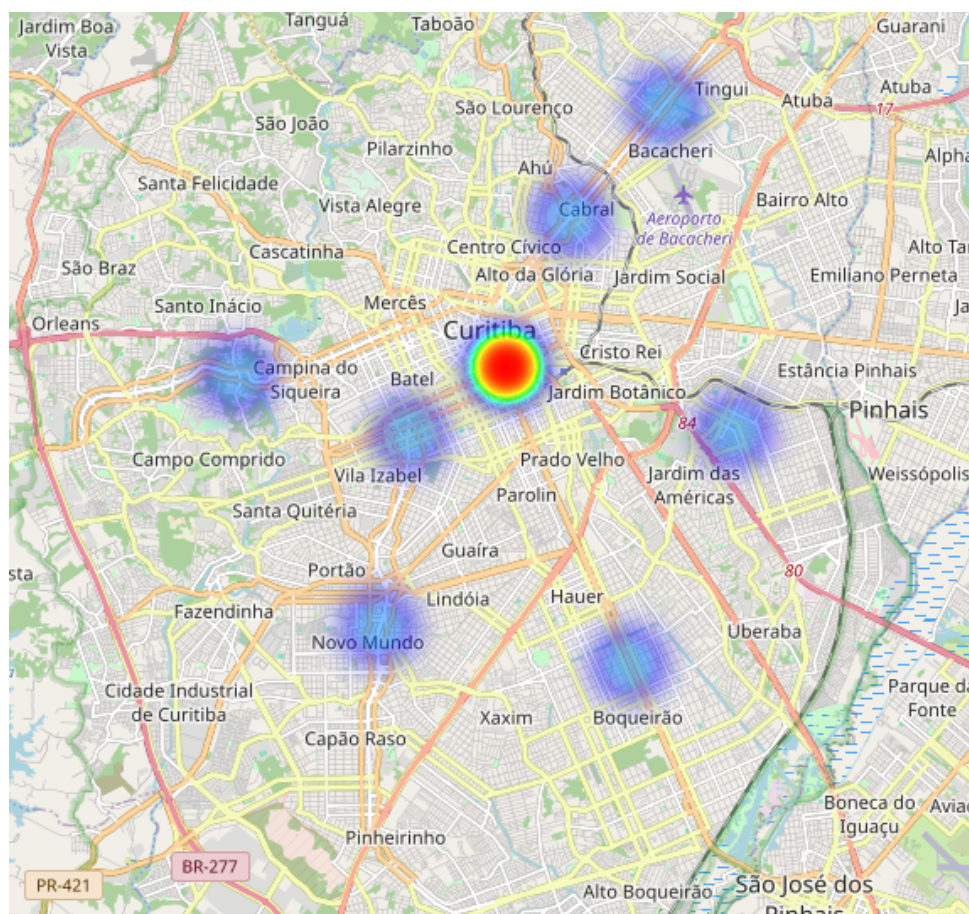


Figura 1. Mapa gerado com dados de transporte público de Curitiba.

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. É possível concluir, a partir das informações construídas, que a maioria dos atrasos durante os trechos itinerários de algumas linhas expressas ocorrem em locais onde há grande demanda de passageiros, como em terminais e na região central da cidade.

AGRADECIMENTOS

Meus agradecimentos a professora Marcella, pela dedicação e empenho em orientar os processos da iniciação científica e ao co-autor e colega da pesquisa Matheus Scheibel.

Marcella Scoczynski Ribeiro Martins

DocuSigned by:
Felipe Snitynski Camillo
5D3C0A5994D64B2...

Felipe Snitynski Camillo