



## ANÁLISE DOS DADOS DE TRANSPORTE PÚBLICO E CORRELACIONAMENTO COM OS DADOS DEMOGRÁFICOS DA CIDADE DE SÃO PAULO

Jean M. M. dos SANTOS<sup>1</sup>; Jéssica R. Nogueira<sup>2</sup>.

### RESUMO

O crescimento horizontal e a periferização de São Paulo é uma realidade presente atualmente e, por esse motivo, é necessário saber se o transporte público é suficiente para atender à demanda de seus usuários. Considerando a carência de trabalhos na área, o presente estudo pretende preencher essa lacuna, através da análise de dados da base de transporte público GTFS, com informações da cidade de São Paulo, e correlacionar esses dados com os do Censo de 2022. Para isso, também serão construídos infográficos para apresentar os resultados sobre a distribuição do transporte público e essa informação será comparada com os dados socioeconômicos dos bairros. Por meio da ferramenta kepler.gl, serão criados mapas contendo as rotas de transporte público, apresentando as rotas, horários e paradas dos meios de transporte, fazendo a comparação, novamente, com os dados socioeconômicos dos bairros.

### Palavras-chave:

Transporte Público; Análise de Dados; Python; São Paulo.

### 1. INTRODUÇÃO

Até a chegada dos automóveis, cidades eram menores e limitadas a deslocamentos a pé. A popularização dos carros levou ao crescimento horizontal das cidades, especialmente no Brasil, onde políticas habitacionais focadas em baixa renda reforçam esse padrão (SOUSA; BRAGA, 2019). De acordo com Moura, Oliveira e Figueiredo (2016), São Paulo seguiu esse modelo, com investimentos em rodovias beneficiando os mais ricos, enquanto os de baixa renda ocupavam áreas periféricas. Para contrapor esse modelo de espraiamento e diminuir a dependência do carro criou-se o Desenvolvimento Orientado ao Transporte, buscando densidade e prédios mistos próximos ao transporte público (MARQUES, 2021), adotado pelo Plano Diretor Estratégico de 2014 (SÃO PAULO, 2021).

Para entender melhor essa situação em São Paulo, é crucial avaliar se o transporte público atende às necessidades da população e se o PDE é eficaz. O uso de ferramentas computacionais, como Python e a Especificação Geral de Feed de Trânsito, base de dados que contém informações sobre o transporte público, auxiliam na correlação de dados volumosos sobre transporte público com informações demográficas dos bairros obtidas no Censo 2022.

Essa abordagem se enquadra na Ciência de Dados, que engloba processos, modelos e tecnologias para explorar dados em seu ciclo completo: produção, armazenamento, transformação, análise e visualização, culminando no descarte após certo tempo. Contar histórias com dados é

<sup>1</sup>Discente do curso de Bacharelado em Ciência da Computação, IFSULDEMINAS – *Campus* Passos. E-mail: jean.santos@alunos.ifsuldeminas.edu.br.

<sup>2</sup>Orientador, IFSULDEMINAS – *Campus* Passos. E-mail: jessica.nogueira@ifsuldeminas.edu.br.

crucial, dada a crescente quantidade de dados e a importância das decisões baseadas neles. No presente estudo, serão criados dashboards e mapas usando a ferramenta kepler.gl para comunicar eficazmente os resultados, exibindo a distribuição populacional e das linhas de ônibus.

## **2. MATERIAL E MÉTODOS**

Inicialmente, será realizada a importação do dataset Especificação Geral de Feed de Trânsito (GTFS), utilizando o Jupyter Notebook. Com os dados importados, será feita a seleção dos dados necessários para criação do modelo de linhas de transporte público, com a linguagem de programação Python. Após a seleção, acontecerá a limpeza desses dados, de modo a analisar se todos os campos estão preenchidos corretamente.

Com os dados selecionados, será iniciada sua análise. O primeiro passo será a divisão das rotas por meios de transporte, em que cada meio estará em um grupo diferente. Em seguida, será feita a marcação do ponto de início e fim de cada rota, destacando as rotas de maior distância. Por fim, as rotas serão traçadas, marcando todos os pontos de parada.

Após a análise inicial dos dados de transporte público, será iniciada a análise dos dados do Censo 2010, para a cidade de São Paulo. Haverá o consumo da API (Interface de Programação de Aplicação) disponibilizada pelo Governo Federal. Nela, será possível dividir a cidade em bairros. Com os bairros divididos, serão armazenados os pontos de geolocalização utilizados para essa divisão, eles serão utilizados na futura fusão das informações do Censo com os da GTFS.

Com a API do Censo é possível coletar informações sociodemográficas de cada bairro. As informações coletadas serão as de distribuição de renda média, distribuição de idade, distribuição de sexo, distribuição de escolaridade e população total. Com elas, serão feitas divisões dos bairros para cada um dos pontos de interesse.

O passo seguinte será a junção das bases de dados, em que os bairros serão separados pela quantidade de paradas de ônibus. Será realizada também a análise dos pontos de início e fim das linhas de transporte, verificando quais bairros têm mais pontos de saída e quais têm mais pontos finais. Será realizada também a busca de padrões e a tentativa de correlacionar as características socioeconômicas e de densidade habitacional com a distribuição de transporte público.

Para facilitar a análise e disponibilização dos dados, serão utilizados gráficos e mapas. Os gráficos serão produzidos utilizando a biblioteca Matplotlib. Serão essenciais para a análise das diferenças entre bairros, explanando suas diferenças. Os mapas serão criados com Kepler.gl, com eles será mais fácil a visualização da distribuição das linhas de transporte público. A utilização da ferramenta de camadas será utilizada para uma melhor visualização entre as linhas e as características dos bairros

### 3. RESULTADOS E DISCUSSÃO

Espera-se, inicialmente, que seja possível realizar uma limpeza e análise dos dados do tipo Especificação Geral de Feed de Trânsito (GTFS), disponibilizados pela SPTrans, e dos dados do Censo 2010. É esperado também que se consiga fazer a união destes datasets e assim, fazer novas análises com os dados em conjunto. Após esse momento, com o auxílio da aplicação kepler.gl e de bibliotecas do Python, criar tabelas e mapas que facilitem o trabalho de análise desses dados. Por fim, almeja-se que os resultados possam ser apresentados com a utilização de técnicas de storytelling para melhorar sua compreensão.

Com os resultados obtidos no trabalho, almeja-se ampliar a compreensão da distribuição de transporte público da cidade de São Paulo, levando-se em consideração os dados socioeconômicos dos moradores paulistanos. Espera-se que esse estudo aperfeiçoe a distribuição da malha de transporte da capital paulista e auxilie em futuras ampliações dessa malha. Outro ponto esperado é que esses resultados incentivem novos trabalhos similares em outras cidades, melhorando a compreensão do transporte público de uma maneira geral.

### 4. CONCLUSÃO

O artigo tem como objetivo propiciar um melhor entendimento da distribuição da malha de transporte da capital paulista, contribuindo para possíveis alterações e melhorias no seu perfil. Entretanto alguns pontos poderão ser adicionados em futuros trabalhos para uma melhor compreensão deste perfil, entre eles é possível destacar uma futura utilização de bases que incluem a distribuição de empregos, formais e informais, pela cidade e bases com a distribuição de passageiros por linha.

### REFERÊNCIAS

MARQUES, Estela de Fátima Fernandes. **Crescimento e mobilidade urbana**: o vetor oeste da Região Metropolitana de São Paulo, período 2007-2017. 2021. Dissertação (Mestrado em Tecnologia da Arquitetura) - Faculdade de Arquitetura e Urbanismo, Universidade de São Paulo, São Paulo, 2021.

SÃO PAULO, Prefeitura Municipal de. Secretaria de Urbanismo e Licenciamento. Adensamento construtivo nos eixos e macroáreas a partir dos dados do IPTU. São Paulo, 2021.

SOUSA, I.C.N ; BRAGA, Roberto. Desenvolvimento Orientado ao Transporte Sustentável - DOTS na Análise da Inserção Urbana dos Empreendimentos do Programa Minha Casa Minha Vida - MCMV Faixa 1. XIII ENANPEGE, 2019. São Paulo, p. 128-140.