

UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
FACULDADE DE CIÊNCIAS - CAMPUS BAURU
DEPARTAMENTO DE COMPUTAÇÃO
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

DIOGO MEDEIROS DE ALMEIDA

**ANÁLISE DO PADRÃO DE COMPORTAMENTO DE PEDESTRES
EM SÃO PAULO USANDO DADOS DE UM APLICATIVO DE
RASTREAMENTO DE ATIVIDADES FÍSICAS**

BAURU
Novembro/2024

DIOGO MEDEIROS DE ALMEIDA

**ANÁLISE DO PADRÃO DE COMPORTAMENTO DE PEDESTRES
EM SÃO PAULO USANDO DADOS DE UM APLICATIVO DE
RASTREAMENTO DE ATIVIDADES FÍSICAS**

Trabalho de Conclusão de Curso do Curso
de Sistemas de Informação da Universidade
Estadual Paulista “Júlio de Mesquita Filho”,
Faculdade de Ciências, Campus Bauru.
Orientador: Prof. Dr. Higor Amario de Souza

BAURU
Novembro/2024

Diogo Medeiros de Almeida ANÁLISE DO PADRÃO DE COMPORTAMENTO DE PEDESTRES EM SÃO PAULO USANDO DADOS DE UM APLICATIVO DE RASTREAMENTO DE ATIVIDADES FÍSICAS/ Diogo Medeiros de Almeida. – Bauru, Novembro/2024– 51 p. : il. (algumas color.) ; 30 cm.
Orientador: Prof. Dr. Higor Amario de Souza
Trabalho de Conclusão de Curso – Universidade Estadual Paulista “Júlio de Mesquita Filho”
Faculdade de Ciências
Sistemas de Informação, Novembro/2024.
1. Mobilidade urbana 2. Pedestres 3. Análise geoespacial 4. Análise de dados 5. Aplicativo de Atividades Físicas 6. São Paulo 7. Mapas de calor

Diogo Medeiros de Almeida

ANÁLISE DO PADRÃO DE COMPORTAMENTO DE PEDESTRES EM SÃO PAULO USANDO DADOS DE UM APLICATIVO DE RASTREAMENTO DE ATIVIDADES FÍSICAS

Trabalho de Conclusão de Curso do Curso de Sis-
temas de Informação da Universidade Estadual
Paulista "Júlio de Mesquita Filho", Faculdade
de Ciências, Campus Bauru.

Banca Examinadora

Prof. Dr. Higor Amario de Souza

Orientador

Universidade Estadual Paulista "Júlio de
Mesquita Filho"

Faculdade de Ciências

Departamento de Computação

Prof. Dr. Jose Remo Ferreira Brega

Universidade Estadual Paulista "Júlio de
Mesquita Filho"

Faculdade de Ciências

Departamento de Computação

**Prof. Dr. Kelton Augusto Pontara da
Costa**

Universidade Estadual Paulista "Júlio de
Mesquita Filho"

Faculdade de Ciências

Departamento de Computação

Bauru, 25 de Novembro de 2024.

Agradecimentos

Primeiramente, agradeço a Deus, por ser minha fortaleza e guia em todos os momentos, proporcionando-me forças, sabedoria e perseverança para concluir mais esta etapa da minha vida.

Aos meus pais, João e Célia, os quais são meu maior exemplo de amor, dedicação e apoio incondicional. Suas palavras de encorajamento e presença constante me deram confiança para superar os desafios e continuar acreditando nos meus sonhos.

Ao meu orientador, Prof. Dr. Higor Amario de Souza, pela paciência, orientação precisa e incentivo ao longo de todo o processo. Sua dedicação e conhecimentos foram fundamentais para o desenvolvimento deste trabalho, e sua confiança em mim foi um estímulo constante.

Aos meus amigos, que estiveram ao meu lado em todos os momentos, compartilhando risadas, incentivando em dias difíceis e celebrando cada conquista comigo. Vocês são uma parte especial dessa jornada e sempre levarei comigo as memórias que construímos juntos.

Aos meus colegas de curso, que tornaram essa caminhada mais leve com companheirismo e trocas de aprendizado. Cada conversa e discussão agregaram muito ao meu crescimento acadêmico e pessoal.

Aos professores que, ao longo da minha formação, compartilharam seu conhecimento e incentivaram meu desenvolvimento. Cada ensinamento deixou marcas importantes para minha trajetória.

Por fim, dedico este trabalho a todos que, de alguma forma, contribuíram para a minha formação, direta ou indiretamente. Agradeço por acreditarem em mim e me ajudarem a chegar até aqui. Este momento é fruto de muito esforço e apoio, e sou eternamente grato por isso.

*Às vezes são as pessoas que ninguém imaginou nada que fazem as coisas que ninguém pode
imaginar*

— Alan Turing

Resumo

Este trabalho mostra uma análise do padrão de comportamento de pedestres na cidade de São Paulo, utilizando dados obtidos de um aplicativo de rastreamento de atividades físicas e pela plataforma GeoSampa. Este trabalho apresenta um estudo que tem em vista compreender como os pedestres se deslocam, identificar fatores que influenciam suas rotas e sugerir melhorias na infraestrutura urbana, alinhando-se aos princípios da Política Nacional de Mobilidade Urbana. A metodologia inclui análises geoespaciais, descritivas e exploratórias, com o uso de mapas de calor para visualizar fluxos e trajetos. Os resultados destacam diferenças entre viagens de lazer e deslocamento, mudanças no comportamento durante a pandemia da COVID-19 e padrões relacionados a gênero e faixa etária. As conclusões apontam para a importância de políticas públicas que promovam a mobilidade ativa, contribuindo para cidades mais sustentáveis e acessíveis.

Palavras-chave: Mobilidade urbana, Pedestres, Análise geoespacial, Análise de dados, Aplicativo de Atividades Físicas, São Paulo, Mapas de calor.

Abstract

This paper presents an analysis of pedestrian behavior patterns in the city of São Paulo, using data obtained from a physical activity tracking application and the GeoSampa platform. This paper presents a study that seeks to understand how pedestrians move, identify factors that influence their routes, and suggest improvements in urban infrastructure, in line with the principles of the National Urban Mobility Policy. The methodology includes geospatial, descriptive, and exploratory analyses, using heat maps to visualize flows and routes. The results highlight differences between leisure and commuting trips, changes in behavior during the COVID-19 pandemic, and patterns related to gender and age group. The conclusions point to the importance of public policies that promote active mobility, contributing to more sustainable and accessible cities.

Keywords: Urban mobility, Pedestrians, Geospatial analysis, Data analysis, Tracking of Physical Activities, São Paulo, Heatmaps.

Lista de figuras

Figura 1 – Exemplo de um heatmap tradicional	17
Figura 2 – Exemplo de heatmap usado na análise	17
Figura 3 – Fluxo de trabalho	19
Figura 4 – HeatMap de todas as viagens na cidade de São Paulo-2019 a 2024	28
Figura 5 – Mapa Ampliado	29
Figura 6 – Viagens de lazer na cidade de São Paulo-2019 a 2024	32
Figura 7 – Viagens de deslocamentos na cidade de São Paulo-2019 a 2024	33
Figura 8 – Distritos com maior número de viagens de lazer-2019 a 2024	33
Figura 9 – Distritos com maior número de viagens de deslocamento-2019 a 2024	34
Figura 10 – Comparação dos mapas de lazer e deslocamento de Moema-2019 a 2024	35
Figura 11 – Comparação dos mapas de lazer e deslocamento de Vila Mariana-2019 a 2024	36
Figura 12 – Viagens na Pandemia por Semestre-2019 a 2024	37
Figura 13 – Viagens na Pandemia - 2020	37
Figura 14 – Viagens Deslocamento na Pandemia-Fevereiro 2020	39
Figura 15 – Viagens Deslocamento na Pandemia-Março 2020	40
Figura 16 – Viagens Deslocamento na Pandemia-Abril 2020	40
Figura 17 – Viagens Deslocamento na Pandemia-Fevereiro 2021	41
Figura 18 – Viagens Deslocamento na Pandemia-Março 2021	41
Figura 19 – Viagens Deslocamento na Pandemia-Abril 2021	42
Figura 20 – Viagens Horário-2019 a 2024	43
Figura 21 – Viagens por Faixa Etária por Horário-2019 a 2024	44
Figura 22 – Viagens por Dia da Semana-2019 a 2024	45
Figura 23 – Viagens das Faixas Etárias por Dia da Semana-2019 a 2024	46
Figura 24 – Viagens por Meses-Todos Gêneros e Todas as Faixas Etárias-2019 a 2024	47

Lista de tabelas

Tabela 1 – Impactos do aumento do transporte individual e motorizado.	13
---	----

Lista de abreviaturas e siglas

API	Interface de Programação de Aplicações (Application Programming Interface).
CPU	Unidade Central de Processamento (Central Processing Unit).
CSV	Valores Separados por Vírgula (Comma-Separated Values).
ELT	Extração, Carga e Transformação (Extract, Load and Transform).
GPU	Unidade de Processamento Gráfico (Graphics Processing Unit).
GPS	Sistema de Posicionamento Global (Global Positioning System).
OSM	OpenStreetMap.
PPBE	Políticas Públicas Baseadas em Evidências.
SHP	Formato de arquivo para dados espaciais.
SIG	Sistemas de Informações Geográficas.
WebGL	Biblioteca Gráfica para a Web (Web Graphics Library).

Sumário

1	INTRODUÇÃO	13
2	FUNDAMENTAÇÃO TEÓRICA	15
2.1	Mobilidade ativa	15
2.2	Políticas públicas baseadas em evidências	15
2.3	Ciência e Análise de Dados	15
2.3.1	Análise Geoespacial	16
2.3.1.1	Mapa de Arestas (HeatMap)	16
2.3.2	Análise Descritiva	18
2.3.3	Análise Exploratória	18
3	METODOLOGIA	19
3.1	Ferramentas de análise de dados	20
3.1.1	Python	20
3.1.2	Jupyter Notebook	20
3.1.3	Pandas	20
3.1.4	Geopandas	21
3.1.5	NumPy	21
3.1.6	Keplergl	21
3.1.7	Matplotlib	21
3.1.8	Seaborn	22
3.1.9	Ipywidgets	22
3.2	Dados Usados	22
3.2.1	Aplicativo de rastreamento de atividades físicas	22
3.2.2	GeoSampa	23
3.3	Obtenção dos dados	23
3.3.1	Dados de Viagens	23
3.3.2	Dados Geoespaciais	23
3.4	Importação dos dados	24
3.5	Características dos Dados	24
3.5.1	OpenStreetMap	25
3.5.2	Aplicativo de Rastreamento de Atividades Físicas	25
3.5.3	GeoSampa	25
3.6	Tratamento, limpeza e manipulação de dados	25
3.6.1	Transformação dos dados	26
3.6.2	Aplicação dos Filtros	26

3.6.3	Tratamento de Inconsistências	28
3.7	Criação das Análises	28
3.7.1	Análises Geoespaciais	28
3.7.2	Análise da mobilidade durante a pandemia da COVID-19	31
3.7.3	Análises das Características das Viagens de Pedestres	31
3.8	Disponibilização das Análises e Resultados	31
4	RESULTADOS E ANÁLISES	32
4.1	Análise Geoespacial	32
4.2	Análise da mobilidade durante a pandemia da COVID-19	36
4.3	Análises das Características das Viagens de Pedestres	42
4.4	Desafios Enfrentados	47
4.5	Considerações	48
5	CONCLUSÃO	50
5.1	Trabalhos futuros	50
	REFERÊNCIAS	51

1 Introdução

O crescimento urbano e a priorização do transporte motorizado têm gerado diversos impactos nas grandes cidades, incluindo congestionamento, poluição do ar, ruído e acidentes fatais (BONGIORNO et al., 2019). Esses problemas são particularmente agravantes em metrópoles como São Paulo, onde a mobilidade urbana sustentável emerge como uma prioridade para melhorar a saúde e qualidade de vida da população (CRUZ; PAULINO, 2019). Dentre as alternativas de transporte, a mobilidade ativa, especialmente a caminhada, surge como uma forma eficiente e saudável de deslocamento para curtas e médias distâncias, contribuindo para a redução da emissão de gases poluentes e promovendo o bem-estar físico dos cidadãos (PUCHER et al., 2011; FAN; LOO, 2021).

Estudos indicam que o aumento de espaços urbanos adaptados para pedestres e ciclistas pode não apenas mitigar os efeitos nocivos do uso excessivo de veículos, mas também transformar as cidades em ambientes mais inclusivos e seguros (DEGREAS, 2022). Além de trazer ganhos enormes à sociedade que combatem os danos causados pelos meios motorizados e individuais, indicados na Tabela 1.

Tabela 1 – Impactos do aumento do transporte individual e motorizado.

Econômicos	Sociais	Ambientais
Trânsito e Congestionamento	Distribuição desigual dos impactos	Poluição do ar, água e sonora
Barreiras de Mobilidade	Desvantagens de Mobilidade	Perda de Habitacões
Acidentes	Impacto sobre a saúde humana	Impactos Hidrológicos
Custos com Infraestrutura	Interação na Comunidade	–
Custos para consumidores	Habitabilidade da Comunidade	–
–	Estética	–

Fonte: (LITMAN; BURWELL, 2006).

A Política Nacional de Mobilidade Urbana do Brasil, estabelecida pela Lei 12.587/2012, destaca a prioridade dos modos de transporte não motorizados e promove a integração entre diferentes formas de deslocamento, visando à acessibilidade universal e ao desenvolvimento sustentável dos espaços urbanos.

Esse trabalho visa analisar principalmente o padrão de comportamento de pedestres na cidade de São Paulo, a partir de dados gerados por um aplicativo de rastreamento de atividades físicas. Busca-se compreender como os pedestres se deslocam em áreas específicas da cidade e identificar fatores que influenciam suas rotas e preferências de trajeto. Além disso, o trabalho verifica quais regiões necessitam de melhorias na infraestrutura para garantir maior segurança

e acessibilidade, alinhando-se aos princípios da Política Nacional de Mobilidade Urbana, que promove o uso de modos de transporte não motorizados.

Para alcançar esse objetivo, foram utilizados dados geoespaciais fornecidos pelo aplicativo, que rastreia as rotas percorridas por pedestres e ciclistas por meio de dispositivos de GPS. Estes dados foram analisados utilizando técnicas de análise espacial e de dados, integradas ao uso de mapas de calor, para visualizar a intensidade de tráfego de pedestres em diferentes regiões da cidade de São Paulo.

Este estudo apresenta uma ferramenta de análise que permite que especialistas examinem os padrões de comportamento de pedestres utilizando dados reais da cidade. Essa análise pode colaborar para o desenvolvimento de políticas públicas de mobilidade ativa, promovendo um ambiente urbano seguro e acessível para todos. (GALANIS; BOTZORIS; ELIOU, 2017).

2 Fundamentação Teórica

2.1 Mobilidade ativa

A mobilidade ativa é um conceito que engloba todos os deslocamentos realizados por meios não motorizados, como caminhar, pedalar ou o uso de dispositivos auxiliares, como cadeiras de rodas. É caracterizada pela autonomia dos usuários e pelo impacto positivo na sustentabilidade urbana, por promover a redução de emissões de gases poluentes, melhorar a qualidade do espaço público e incentivar um estilo de vida mais saudável. Além disso, a mobilidade ativa desempenha um papel importante na inclusão social, ao ampliar o acesso à cidade e às suas oportunidades, reforçando a equidade no uso do espaço urbano. (CRUZ; PAULINO, 2019)

2.2 Políticas públicas baseadas em evidências

Políticas públicas baseadas em evidências (PPBE) são iniciativas governamentais ou sociais que utilizam dados empíricos e análises científicas para orientar a formulação, implementação e avaliação de políticas públicas. Esse modelo se fundamenta na ideia de que as decisões políticas devem ser fundamentadas em “o que funciona”, sendo apoiadas por pesquisas anteriores, análises rigorosas e evidências sistemáticas que garantam eficácia e eficiência.

A PPBE enfatiza a importância de revisões sistemáticas, como a meta-análise e as análises narrativas, que examinam intervenções passadas para identificar boas práticas e formular novos programas que possam ser adaptados aos contextos específicos. No entanto, debates metodológicos, como os apresentados por Pawson (2002), destacam que não existe um “padrão-ouro” para avaliação. Ele argumenta que é essencial considerar mecanismos, contextos e variações de resultados ao transferir aprendizados de programas anteriores para novos cenários.

Em suma, as políticas públicas baseadas em evidências buscam reduzir erros, orientar a implementação e promover intervenções informadas, favorecendo uma gestão mais responsável e eficaz dos recursos públicos (PAWSON, 2002).

2.3 Ciência e Análise de Dados

Segundo Provost e Fawcett (2013), a Ciência de Dados é um conjunto de princípios fundamentais que sustentam e orientam a extração de informações e conhecimento a partir de dados. Através da aplicação de métodos e técnicas, a ciência de dados permite a compreensão de fenômenos e a otimização da tomada de decisões. A ciência de dados é a base deste

trabalho, onde, através do entendimento, da manipulação, da análise e da visualização dos dados, busca-se entender o comportamento das viagens realizadas na cidade de São Paulo.

2.3.1 Análise Geoespacial

A análise geoespacial é utilizada para adicionar tempo e localização aos tipos tradicionais de dados e construir visualizações como mapas, gráficos e cartogramas, mostrando alterações históricas e tendências atuais (IBM, 2024). Ela permite entender padrões e tendências em relação à localização, distância e interação espacial. Ferramentas de Sistemas de Informações Geográficas (SIG) são comumente utilizadas para processar, analisar e visualizar dados geoespaciais.

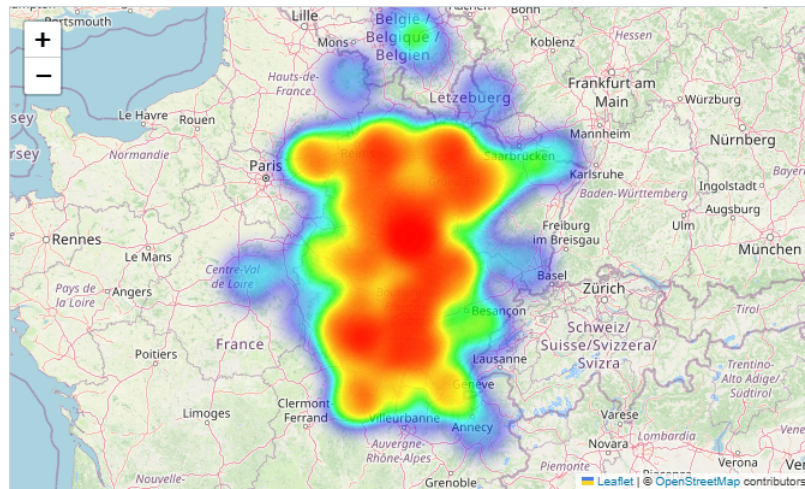
2.3.1.1 Mapa de Arestas (HeatMap)

O *heatmap*, ou mapa de calor, é uma representação gráfica de dados onde os valores individuais contidos em uma matriz ou uma área são representados por cores (WILKINSON; FRIENDLY, 2009). Tradicionalmente, os heatmaps utilizam uma escala de cores para indicar intensidades, sendo amplamente utilizados em análises estatísticas, visualizações geográficas e representações de dados em diversas áreas. A intensidade da cor reflete a densidade ou frequência dos dados em determinada área, proporcionando uma compreensão visual imediata das tendências.

No contexto desta análise, optou-se por um mapa de arestas como heatmap, o que permite visualizar de maneira mais clara as viagens realizadas. Em vez de apenas pontos ou manchas de densidade, o mapa de arestas destaca os trajetos efetivamente percorridos, enfatizando os corredores de maior fluxo e possibilitando uma compreensão espacial detalhada dos padrões de deslocamento.

Um exemplo tradicional de heatmap pode ser observado na Figura 1.

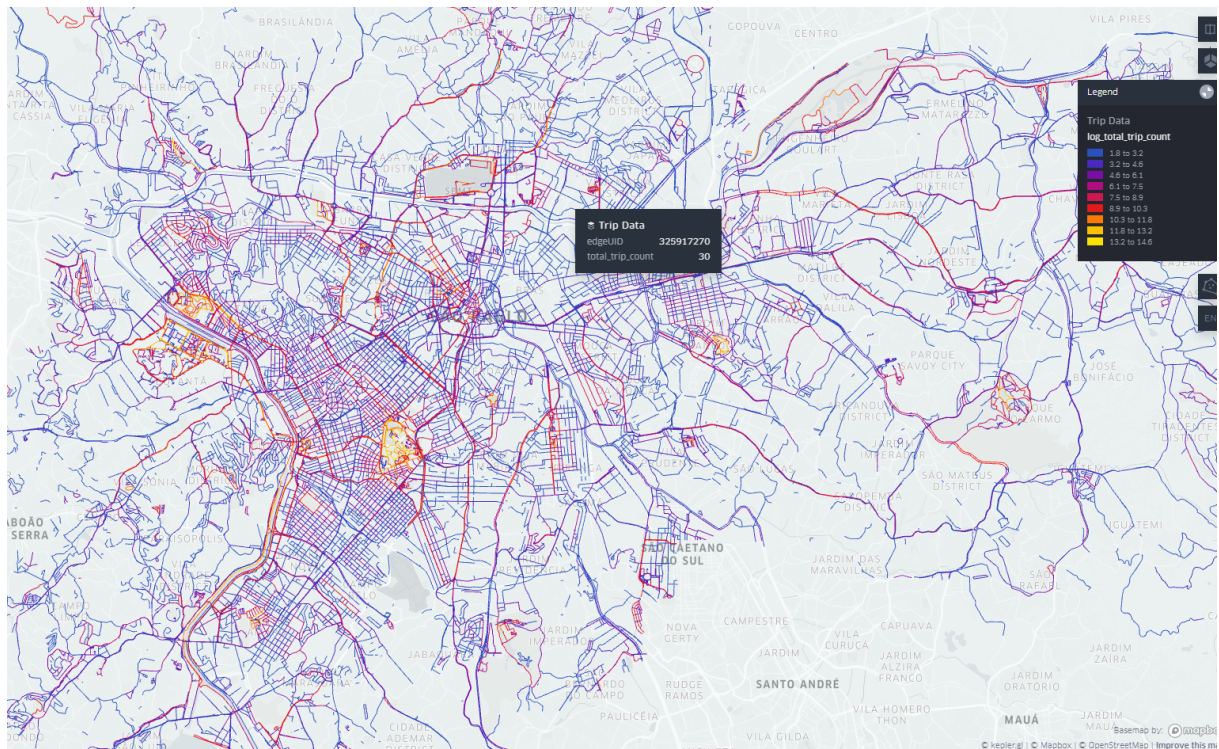
Figura 1 – Exemplo de um heatmap tradicional



Fonte: <https://python-visualization.github.io/folium>.

Um exemplo específico do mapa de arestas utilizado nesta análise pode ser visto na Figura 2.

Figura 2 – Exemplo de heatmap usado na análise



Fonte: Elaborada pelo autor.

2.3.2 Análise Descritiva

A Análise Descritiva concentra-se em descrever e resumir dados existentes, proporcionando uma visão geral e sintetizada do conjunto de informações disponíveis. Seu foco principal é identificar padrões básicos, frequências e medidas de tendência central. Essa abordagem é essencial para compreender as características fundamentais dos dados antes de avançar para análises mais complexas. O objetivo da Análise Descritiva é fornecer insights iniciais sobre o comportamento dos dados, facilitando a identificação de tendências gerais e discrepâncias. As ferramentas utilizadas nessa análise incluem gráficos simples, como histogramas e gráficos de barras, que permitem visualizar a distribuição e a frequência dos dados intuitivamente. (REIS; REIS, 2002)

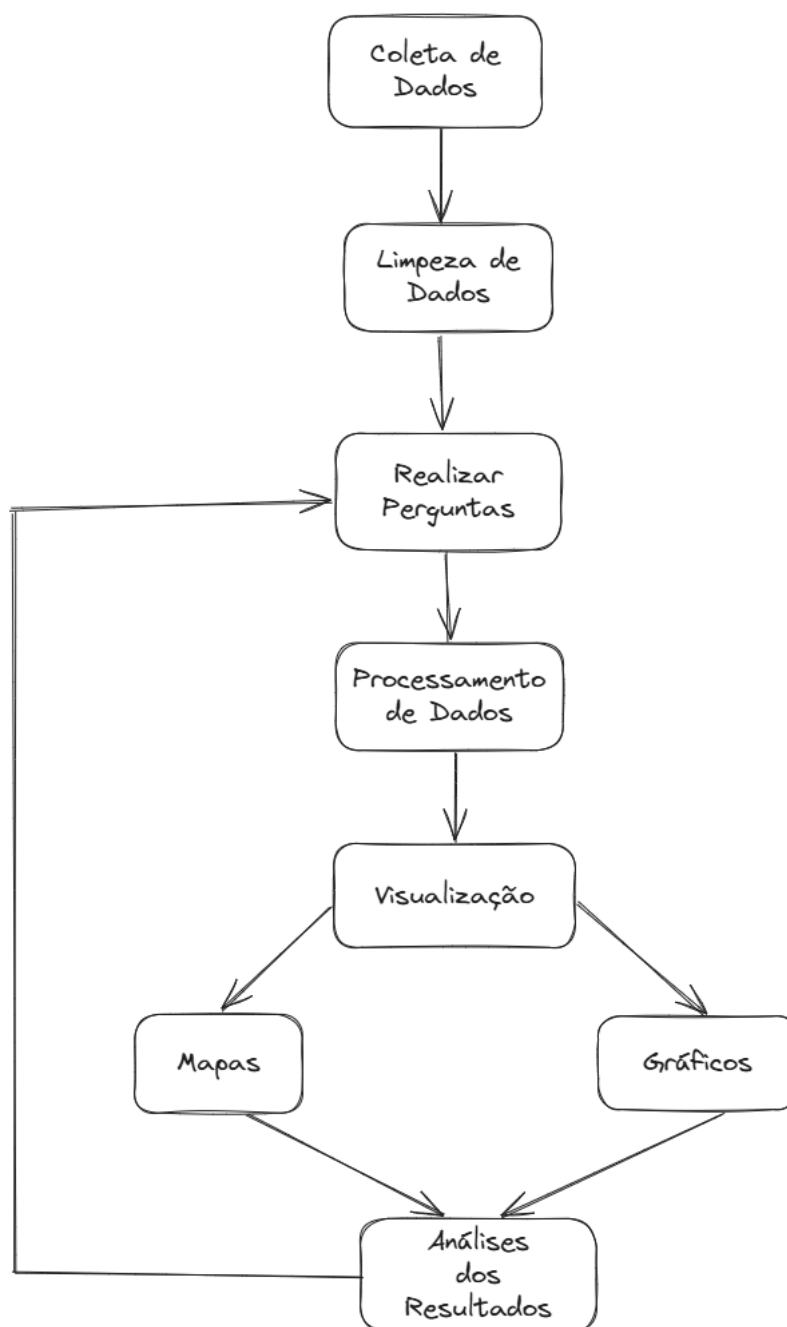
2.3.3 Análise Exploratória

A Análise Exploratória vai além da descrição básica, buscando explorar os dados em profundidade para descobrir padrões ocultos, anomalias, tendências e relacionamentos entre variáveis. O foco dessa abordagem é investigar os dados de maneira mais detalhada, preparando o terreno para análises mais sofisticadas e modelagens preditivas. O objetivo da análise exploratória é identificar insights significativos e formular questões relevantes que possam orientar pesquisas futuras ou estratégias de negócio. (LEMOS, 2022)

3 Metodologia

Durante a elaboração do trabalho, foi seguida uma sequência de etapas comuns na análise de dados. A estrutura desse processo está apresentada no diagrama de fluxo de trabalho na Figura 3, proporcionando uma visão mais clara do desenvolvimento e das etapas da análise.

Figura 3 – Fluxo de trabalho



Fonte: Elaborada pelo autor.

As primeiras etapas concentram-se na obtenção e limpeza dos dados, o que é fundamental para garantir a qualidade das análises subsequentes. Com os dados devidamente preparados, podemos começar a formular perguntas relevantes sobre eles. Em seguida, processamos os dados utilizando diversas técnicas para extrair as informações necessárias para as análises desejadas. Após criar visualizações dos dados, que facilitem a compreensão dos padrões e tendências, realizamos análises aprofundadas. Esse processo pode levar ao surgimento de novas perguntas e hipóteses, impulsionando um ciclo contínuo de investigação e refinamento até que todas as informações desejadas sejam obtidas e compreendidas em sua totalidade.

3.1 Ferramentas de análise de dados

3.1.1 Python

Python ¹ é uma linguagem poderosa e de fácil aprendizado, com estruturas de dados de alto nível que permitem uma programação eficiente. Com uma sintaxe clara e tipagem dinâmica, combinadas com sua natureza interpretada, Python é ideal para desenvolver rapidamente aplicações em diversas áreas e em múltiplas plataformas. Além disso, possui um extenso ecossistema de bibliotecas que facilitam o trabalho com diversos tipos de dados. Para este trabalho, Python foi escolhida por sua versatilidade e pela facilidade em lidar com dados geoespaciais e geolocalizados, essenciais para as análises exigidas.

3.1.2 Jupyter Notebook

O Jupyter Notebook ² é uma aplicação web de código aberto que permite criar e compartilhar documentos interativos contendo código, visualizações e textos explicativos. Ele suporta várias linguagens de programação, incluindo Python, e é amplamente utilizado em ciência de dados, pesquisa acadêmica e educação. Com o Jupyter, você pode explorar dados interativamente, experimentar com diferentes técnicas de análise e documentar todo o processo em um único ambiente.

3.1.3 Pandas

O Pandas ³ é uma biblioteca de software desenvolvida para a linguagem Python, focada na manipulação e análise de dados. Ela oferece estruturas de dados flexíveis e eficientes, além de funções poderosas para trabalhar com dados em formato tabular. No contexto do projeto, desempenha um papel fundamental na preparação dos dados para análise, permitindo a organização e transformação dos dados brutos em formatos mais claros e úteis. A biblioteca

¹ <<https://www.python.org/>>

² <<https://jupyter.org/>>

³ <<https://pandas.pydata.org/>>

facilita o entendimento, tratamento e limpeza dos dados em uso. Sua documentação é extensa e detalhada, mas intuitiva e de fácil acesso.

3.1.4 Geopandas

GeoPandas ⁴ é uma biblioteca Python que expande as funcionalidades do Pandas para manipulação de dados geoespaciais, simplificando a análise e o trabalho com dados geográficos. Sua principal estrutura, o GeoDataFrame, é semelhante ao DataFrame do Pandas, mas inclui uma coluna especial para armazenar geometrias espaciais, como pontos, arestas e polígonos. Com o GeoPandas, é possível realizar manipulações de geometrias, transformações e projeções de coordenadas, além de análises espaciais diversas. Essa biblioteca integra operações espaciais a tipos geométricos, permitindo explorar e analisar dados espaciais eficientemente, aproveitando todo o potencial das funcionalidades do Pandas.

3.1.5 NumPy

A documentação do NumPy ⁵ o descreve como a principal biblioteca em Python para computação científica e análise de dados. Essa biblioteca se baseia em arrays multidimensionais, os quais são estruturas de dados eficientes para armazenamento e manipulação de grandes volumes de informações. Graças à sua eficiência computacional, o NumPy permite trabalhar com datasets extensos de forma otimizada. Além disso, oferece recursos avançados para indexação e seleção de elementos em arrays, bem como uma ampla gama de funções matemáticas para operações diversas, simplificando o processo de manipulação e análise de dados.

3.1.6 Kepler.gl

Segundo sua documentação, o Kepler.gl ⁶ é uma poderosa ferramenta de código aberto para análise geoespacial, projetada para visualizar grandes conjuntos de dados de localização diretamente no navegador. Construído sobre o MapLibreGL e o deck.gl, ele utiliza o WebGL para renderizar rapidamente milhões de pontos, permitindo a exploração interativa de tendências espaciais e temporais. Além disso, o Kepler.gl é um componente React que pode ser integrado a aplicações webs, oferecendo uma interface intuitiva para adicionar dados, aplicar filtros e personalizar mapas, facilitando insights profundos em dados geoespaciais.

3.1.7 Matplotlib

O Matplotlib ⁷ é uma biblioteca amplamente utilizada em Python para criação de gráficos e visualizações de dados. Com uma vasta gama de estilos e tipos de gráficos, como

⁴ <<https://geopandas.org/>>

⁵ <<https://numpy.org/>>

⁶ <<https://kepler.gl/>>

⁷ <<https://matplotlib.org/>>

gráficos de barras, arestas, pizza e dispersão, permite que os usuários personalizem cada aspecto visual, como cores, rótulos e tamanhos, para tornar a apresentação de dados mais clara e atrativa. Além disso, o Matplotlib se integra facilmente com outras bibliotecas, como Pandas e NumPy, e oferece uma API orientada a objetos que facilita a incorporação de gráficos em aplicativos e interfaces.

3.1.8 Seaborn

O Seaborn ⁸ é uma biblioteca Python para visualização de dados que expande as funcionalidades do Matplotlib, oferecendo uma interface de alto nível para a criação de gráficos estatísticos esteticamente agradáveis e informativos. Suas principais vantagens incluem facilidade de uso, paletas de cores padronizadas e opções otimizadas para acessibilidade, como uma paleta específica que favorece a visualização por pessoas daltônicas. O Seaborn é especialmente útil para visualização de dados univariados e bivariados, visualização de séries temporais e na resolução de problemas de sobreposição de dados, além de melhorar a aparência dos gráficos.

3.1.9 Ipywidgets

A biblioteca ipywidgets ⁹ faz parte do ecossistema Jupyter e é voltada para a criação de widgets interativos em notebooks, como o Jupyter Notebook e o JupyterLab. Ela é especialmente útil para desenvolver interfaces de usuário interativas e dinâmicas, facilitando análises de dados, visualizações e demonstrações.

3.2 Dados Usados

Nesta seção, são apresentados os principais conjuntos de dados utilizados na elaboração deste trabalho. A escolha dessas fontes foi fundamentada na relevância e na diversidade das informações disponibilizadas, garantindo uma análise robusta e abrangente. Os dados contemplam tanto aspectos relacionados à mobilidade urbana e comportamento físico dos usuários quanto informações geoespaciais da cidade de São Paulo, promovendo uma integração que possibilita explorar múltiplas dimensões do tema estudado. A seguir, detalham-se as principais fontes e suas características.

3.2.1 Aplicativo de rastreamento de atividades físicas

O aplicativo é uma plataforma digital que permite aos usuários registrar e compartilhar atividades físicas, como corridas e pedaladas, utilizando dados de GPS. Além de monitorar o desempenho, o aplicativo oferece recursos sociais que possibilitam a interação entre os

⁸ <<https://seaborn.pydata.org/>>

⁹ <<https://ipywidgets.readthedocs.io/en/stable/>>

atletas. A plataforma disponibiliza uma API pública que permite a desenvolvedores acessar dados de atividades, segmentos e clubes, possibilitando a criação de aplicativos e serviços que complementam a experiência dos usuários.

3.2.2 GeoSampa

O GeoSampa é o mapa digital oficial da cidade de São Paulo, desenvolvido pela Prefeitura Municipal. Essa plataforma integra dados geoespaciais de diversas secretarias municipais, oferecendo informações detalhadas sobre zoneamento, equipamentos públicos, infraestrutura urbana e outros aspectos territoriais da cidade. Os dados são disponibilizados em formato aberto, permitindo que cidadãos, pesquisadores e desenvolvedores os utilizem para análises, estudos e desenvolvimento de aplicações que promovam a transparência e a participação cidadã.

3.3 Obtenção dos dados

A obtenção dos dados é uma etapa crucial para o desenvolvimento de análises robustas e fundamentadas. Para este trabalho, foi adotada uma abordagem que integra diferentes fontes de informações, para explorar múltiplas dimensões relacionadas à mobilidade urbana e ao comportamento de pedestres na cidade de São Paulo.

Os dados utilizados foram obtidos de fontes confiáveis, que oferecem informações detalhadas e diversificadas. Além de promoverem uma base sólida para análises, essas fontes garantem a possibilidade de explorar aspectos espaciais e comportamentais integradamente. A seguir, detalha-se como cada conjunto de dados foi acessado, destacando suas características e relevância para o estudo em questão.

3.3.1 Dados de Viagens

Os dados de viagem utilizados neste estudo foram obtidos de um aplicativo de rastreamento de atividades físicas e correspondem a deslocamentos de pedestres na cidade de São Paulo. O aplicativo possui uma plataforma que fornece dados anônimos e agregados de atividades de ciclismo e corrida. Esses dados são valiosos para planejadores urbanos e autoridades municipais, por oferecerem insights detalhados sobre os padrões de mobilidade das pessoas, permitindo aprimorar a infraestrutura urbana e promover deslocamentos mais sustentáveis na cidade.

3.3.2 Dados Geoespaciais

Os dados geoespaciais necessários para realizar análises mais detalhadas por distrito foram obtidos por meio do GeoSampa¹⁰, o portal oficial de mapas da cidade de São Paulo.

¹⁰ <<https://geosampa.prefeitura.sp.gov.br/Default.aspx>>

Código 1 – Importação dos CSVs e shapefiles

```

1      # Carregar shapefile de edges do aplicativo
2      gdf_edges = gpd.read_file(shapefile_path)
3
4      # Carregar shapefile de distritos GeoSampa
5      districts_gdf = gpd.read_file(districts_shapefile_path)
6      districts_gdf = districts_gdf.set_crs(epsg=31983)
7      district_names = districts_gdf['ds_nome'].unique().tolist()
8
9      # Garantir que os CRS são iguais
10     gdf_edges = gdf_edges.to_crs(districts_gdf.crs)
11
12     # Carregar e processar dados CSV
13     caminho_pasta = "csvs"
14     arquivos_csv = glob.glob(os.path.join(caminho_pasta, "*.csv"))
15
16     lista_dataframes = []
17     for arquivo in arquivos_csv:
18         df_final = pd.read_csv(arquivo)
19         lista_dataframes.append(df_final)
20
21     df_final = pd.concat(lista_dataframes, ignore_index=True)

```

Essa plataforma disponibiliza uma ampla variedade de informações geográficas, permitindo acesso a dados precisos e atualizados sobre os diferentes distritos da cidade. Com isso, é possível conduzir estudos aprofundados e tomar decisões informadas em diversas áreas, como planejamento urbano, gestão ambiental e desenvolvimento socioeconômico.

3.4 Importação dos dados

Após a coleta dos dados necessários, o próximo passo é importá-los para o Jupyter para podermos compreendê-los melhor. Para isso, utilizaremos as bibliotecas Pandas e GeoPandas, como mostra o Código 1, que nos permitirão manipular dados tabulares e geoespaciais, respectivamente. A biblioteca Pandas facilitará o trabalho com dados em formato de tabelas, como CSV, enquanto o GeoPandas será essencial para lidar com informações geográficas, como coordenadas.

3.5 Características dos Dados

Nesta seção, apresentamos as principais características dos conjuntos de dados utilizados neste estudo. Compreender a estrutura e o conteúdo desses dados é essencial para realizar

análises precisas e obter insights significativos sobre a mobilidade urbana e o comportamento dos pedestres na cidade de São Paulo. Exploraremos as especificidades de cada fonte de dados — OpenStreetMap, aplicativo de rastreamento de atividades físicas e GeoSampa — destacando como elas se complementam e contribuem para uma compreensão abrangente do cenário urbano analisado. Ao detalhar os formatos, as informações disponíveis e como os dados se inter-relacionam, visamos fornecer uma base sólida para as etapas subsequentes de processamento e análise.

3.5.1 OpenStreetMap

OpenStreetMap (OSM) é um projeto colaborativo que fornece dados geoespaciais abertos e gratuitos de todo o mundo. O OSM utiliza uma estrutura de dados composta por três elementos fundamentais: *nós*, *arestas* (também conhecidas como *ways*) e *relações*. **Nós** são pontos geográficos definidos por coordenadas de latitude e longitude. **Arestas** são sequências ordenadas de nós que formam linhas (como estradas e rios) ou polígonos fechados (como parques e edifícios). **Relações** são coleções de nós e arestas que definem objetos geográficos mais complexos.

3.5.2 Aplicativo de Rastreamento de Atividades Físicas

Os dados de viagens vêm em formato CSV e possuem um arquivo Shapefile acompanhando com dados geoespaciais. As colunas usadas nas análises incluem *id*, viagens de lazer e deslocamento, data da viagem, gênero (Masculino, Feminino, Não Especificado) e faixa etária (18–34, 35–54, 55–64, 65+). O Shapefile contém o *id* necessário para vincular com os dados do CSV e as informações geoespaciais. Para mapear os dados de viagens e distritos no contexto geográfico fornecido pelo OSM, utilizamos os nós e arestas do OSM como base cartográfica. Os Shapefiles dos distritos de São Paulo contêm coordenadas que correspondem aos nós no OSM.

3.5.3 GeoSampa

Os conjuntos de dados dos distritos do município de São Paulo estão no formato denominado Shapefile (*.shp*)—formato de dados usado pelas bases de dados geoespaciais. Dentre as colunas que o arquivo apresenta, os únicos que nos interessam são os que possuem o nome dos distritos, o nome das subprefeituras e as coordenadas que os mesmos ocupam.

3.6 Tratamento, limpeza e manipulação de dados

Após obter e importar os dados, chegamos a uma etapa crucial da análise: a transformação dos dados. Esse processo faz parte das etapas conhecidas como ELT (do inglês Extract,

Código 2 – Transformação de colunas

```

1     df_final['commute_trip_count'] = df_final[['
        forward_commute_trip_count', 'reverse_commute_trip_count'
    ]].sum(axis=1)
2     df_final['leisure_trip_count'] = df_final[['
        forward_leisure_trip_count', 'reverse_leisure_trip_count'
    ]].sum(axis=1)
3
4     df_final['male_people_count'] = df_final['
        forward_male_people_count'] + df_final['
        reverse_male_people_count']
5     df_final['female_people_count'] = df_final['
        forward_female_people_count'] + df_final['
        reverse_female_people_count']
6     df_final['unspecified_people_count'] = df_final['
        forward_unspecified_people_count'] + df_final['
        reverse_unspecified_people_count']
7     df_final['total_people_count'] = df_final[['male_people_count',
        'female_people_count', 'unspecified_people_count']].sum(
        axis=1)

```

Load and Transform), onde, na fase de transformação, aplicamos as regras necessárias para agregar valor aos dados. É nesse momento que eles se tornam informações valiosas, prontas para análises mais aprofundadas e tomadas de decisão (HübNER, 2020).

3.6.1 Transformação dos dados

O primeiro passo para o processamento dos dados é agrupar as colunas em outras mais úteis para facilitar a análise de dados, uma vez que não precisamos dessa granularidade de informações. Um exemplo disso é mostrado no Código 2. A unificação dos dados nessas colunas torna o dataframe mais leve ao mesmo tempo que facilita o processo de filtragem.

3.6.2 Aplicação dos Filtros

Com o dataframe pronto, podemos começar a aplicar os filtros selecionados ao mesmo. Na primeira execução, como não possui filtros selecionados, ele carrega todos os dados; uma vez selecionado um filtro, ele é aplicado ao dataframe. Esse passo é necessário para trazer apenas os dados desejados na análise. Um exemplo de manipulação de dataframes é mostrado no Código 3.

Código 3 – Aplicação de filtros

```

1      if ano:
2          df = df[df['year'].isin(ano)]
3      if mes:
4          df = df[df['month'].isin(mes)]
5      if faixa_etaria != 'Todos':
6          if activity_type == 'Todos':
7              df = df[
8                  (df[f'{map_genero[genero]}_people_count'] > 0) &
9                  (df[f'forward_{map_faixa_etaria[faixa_etaria]}_people_count'] + df[f'reverse_{map_faixa_etaria[faixa_etaria]}_people_count'] > 0) &
10                 (df['total_trip_count'] == df[['commute_trip_count', 'leisure_trip_count']].sum(axis=1))
11             ]
12          else:
13              df = df[
14                  (df[f'{map_genero[genero]}_people_count'] > 0) &
15                  (df[f'forward_{map_faixa_etaria[faixa_etaria]}_people_count'] + df[f'reverse_{map_faixa_etaria[faixa_etaria]}_people_count'] > 0) &
16                  (df['total_trip_count'] == df[f'{map_atividade[activity_type]}_trip_count'])
17             ]
18      else:
19          if activity_type == 'Todos':
20              df = df[
21                  (df[f'{map_genero[genero]}_people_count'] > 0) &
22                  (df['total_trip_count'] == df[['commute_trip_count', 'leisure_trip_count']].sum(axis=1))
23             ]
24          else:
25              df = df[
26                  (df[f'{map_genero[genero]}_people_count'] > 0) &
27                  (df['total_trip_count'] == df[f'{map_atividade[activity_type]}_trip_count'])
28             ]
29      if genero == 'Homem':
30          df['total_people_count'] = df['male_people_count']
31      elif genero == 'Mulher':
32          df['total_people_count'] = df['female_people_count']
33      elif genero == 'Não especificado':
34          df['total_people_count'] = df['unspecified_people_count']

```


3.6.3 Tratamento de Inconsistências

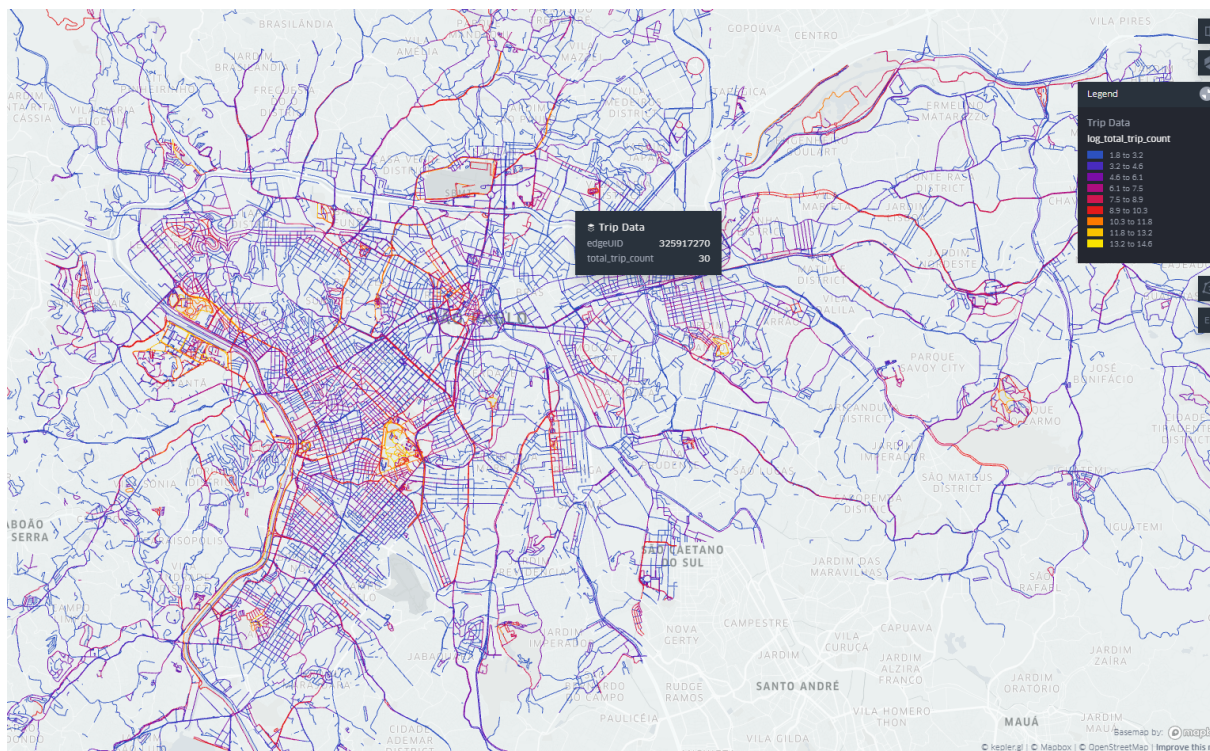
Por fim, para deixar os dados prontos para serem manipulados e gerar as análises necessárias, foi preciso tratar certas inconsistências encontradas nas bases de dados. Um exemplo são as coordenadas nos dados do GeoSampa, que vêm em formato de metros, enquanto as do aplicativo de rastreamento de atividades físicas estão em graus, o que as torna incompatíveis ao se tentar realizar uma junção. Para resolver esse problema, foi necessária a conversão do formato dos dados do GeoSampa para graus, tornando-os compatíveis com os dados do aplicativo.

3.7 Criação das Análises

3.7.1 Análises Geoespaciais

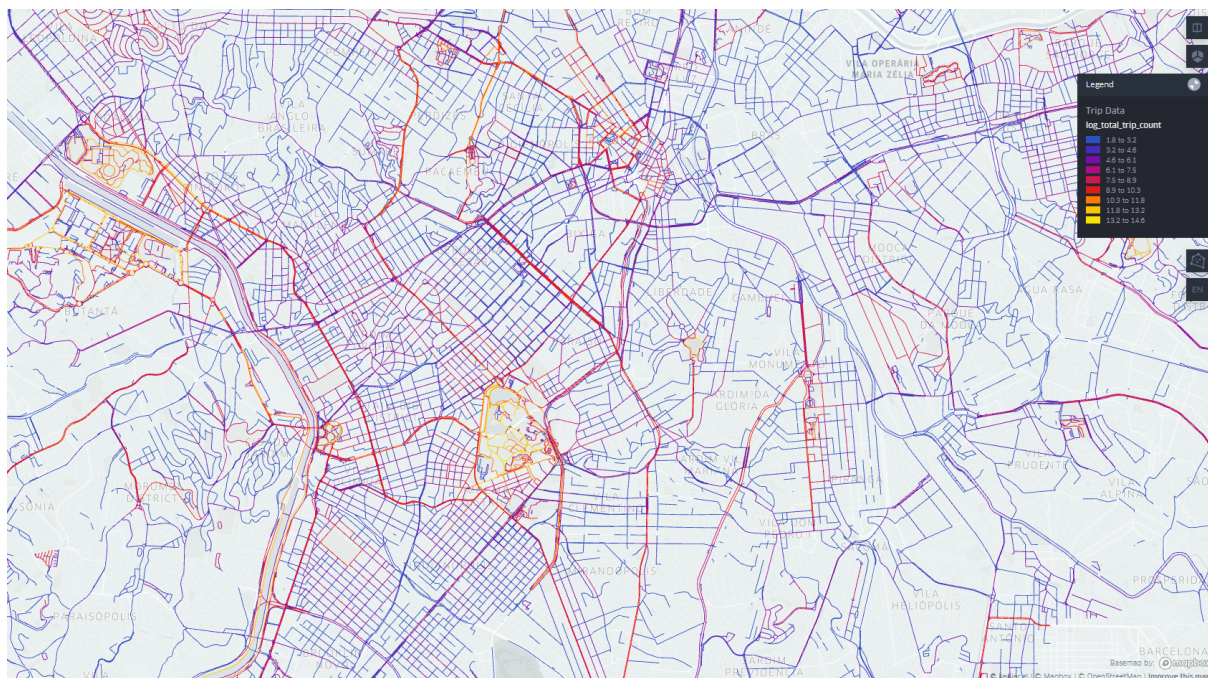
Para a criação do mapa completo da cidade de São Paulo, foram necessários apenas os dados do aplicativo de rastreamento de atividades físicas, o que tornou sua implementação trivial, exigindo apenas a aplicação dos filtros selecionados e a instanciação do mapa. Esse mapa pode ser visualizado na Figura 4.

Figura 4 – HeatMap de todas as viagens na cidade de São Paulo-2019 a 2024



Fonte: Elaborada pelo autor.

Figura 5 – Mapa Ampliado



Fonte: Elaborada pelo autor.

Devido à natureza dos dados coletados, observou-se um acúmulo significativo de viagens em determinadas áreas da cidade. Esse acúmulo ocorre porque certas regiões, como centros comerciais, áreas com grande oferta de transporte público ou pontos turísticos, naturalmente atraem mais deslocamentos. Quando representamos esses dados em um mapa de calor ou em gráficos, essa concentração elevada pode dominar a visualização, dificultando identificar padrões em áreas com menor volume de viagens.

Para resolver esse problema e proporcionar uma visualização mais equilibrada e informativa, foi necessário aplicar uma transformação logarítmica à coluna que registra o total de viagens. A transformação logarítmica é uma técnica estatística que reduz a amplitude dos dados, comprimindo os valores mais altos e expandindo os mais baixos. Isso é particularmente útil quando os dados apresentam uma distribuição assimétrica ou quando há uma grande disparidade entre os valores mínimos e máximos.

Ao transformar os dados para uma escala logarítmica, conseguimos atenuar o impacto visual das áreas com altíssimos volumes de viagens, permitindo que regiões com menos deslocamentos também sejam visualizadas e analisadas com mais clareza. Essa abordagem facilita a identificação de padrões e tendências que poderiam ser ocultados em uma escala linear devido ao domínio visual das áreas de alta concentração.

No contexto deste trabalho, a aplicação da transformação logarítmica permitiu uma representação mais equilibrada dos dados de viagens. Com isso, foi possível não apenas destacar

as regiões com maior fluxo de pedestres, mas também revelar insights em áreas que poderiam ter sido negligenciadas. O Código 4 mostra como a transformação logarítmica foi usada nessa análise.

Código 4 – Transformação em Log

```
1 merged_gdf['log_total_trip_count'] = np.log1p(merged_gdf['  
    total_trip_count'])
```

A função `np.log1p` (logaritmo natural de 1 mais o valor) é amplamente utilizada na análise de dados, especialmente ao lidar com valores que incluem zeros ou são muito pequenos. Essa função é matematicamente válida para valores não negativos, já que adicionar 1 antes do cálculo do logaritmo permite transformar valores iguais a zero seguramente. Além disso, a transformação logarítmica atenua variações extremas ao comprimir a escala, reduzindo a influência de valores muito altos (outliers) enquanto mantém as relações relativas entre os dados. Essa abordagem é particularmente útil em dados de mobilidade, onde algumas áreas ou viagens podem apresentar volumes desproporcionalmente elevados. Como resultado, o conjunto de dados transformado reflete de forma mais precisa as variações nos padrões de deslocamento, destacando nuances que poderiam ser ofuscadas em uma escala linear. Isso facilita análises mais robustas e eficazes na identificação de tendências e padrões, permitindo a elaboração de estratégias de intervenção mais assertivas.

Para a criação dos mapas por distrito e subprefeitura, o processo foi um pouco mais complexo, por envolver o uso simultâneo de duas bases de dados: os dados de viagens provenientes do aplicativo de rastreamento de atividades físicas e os dados geoespaciais do GeoSampa. O GeoSampa, como mapa digital oficial da cidade de São Paulo, forneceu os limites geográficos precisos dos distritos e subprefeituras, essenciais para uma análise espacial detalhada.

Para associar cada viagem a um distrito ou subprefeitura específico, foi necessário realizar junções espaciais entre os conjuntos de dados. Essa técnica permitiu vincular as coordenadas geográficas das viagens aos polígonos que delimitam os distritos e subprefeituras definidos pelo GeoSampa. Com base nas colunas de coordenadas de cada base, foi possível determinar em qual distrito ou subprefeitura cada viagem ocorreu.

A integração dos dados do GeoSampa à análise possibilitou identificar e mapear o volume de viagens em cada região administrativa da cidade. Isso não apenas viabilizou uma visualização detalhada dos padrões de mobilidade por área, mas também permitiu identificar quais distritos e subprefeituras registraram maior volume de viagens.

3.7.2 Análise da mobilidade durante a pandemia da COVID-19

Por fim, este estudo visou compreender o comportamento dos pedestres ao longo da pandemia por meio de uma análise detalhada com o uso de gráficos. A visualização dos dados permitiu observar variações no comportamento ao longo dos diferentes períodos da pandemia, oferecendo uma visão clara das mudanças ao longo do tempo. Foram utilizados gráficos mensais que cobrem os anos de 2019 a 2024, abrangendo o período imediatamente anterior ao início da pandemia, o período de maior restrição e impacto, e o momento de transição para o pós-pandemia.

Para facilitar a análise comparativa, foram desenvolvidos três gráficos específicos. O primeiro representa o comportamento dos pedestres antes do início da pandemia, capturando um contexto mais “normal” e sem as influências das medidas de distanciamento social. O segundo gráfico foca no período durante a pandemia, evidenciando as adaptações e possíveis restrições que afetaram diretamente os pedestres. Por fim, o terceiro gráfico ilustra o período pós-pandemia, permitindo observar as eventuais mudanças no comportamento e os ajustes feitos pelos pedestres ao retomarem uma rotina próxima à normalidade.

3.7.3 Análises das Características das Viagens de Pedestres

As análises realizadas neste estudo buscaram compreender as características predominantes nas viagens de pedestres, considerando, aspectos como sexo e faixa etária ao longo do tempo. Com isso, foi possível explorar os perfis mais comuns de pedestres e suas variações em diferentes contextos.

Para aprofundar essa investigação, as análises foram segmentadas por períodos específicos, como faixas horárias, dias da semana e meses de maior atividade. Esse recorte temporal permitiu observar variações nas características das viagens e identificar os períodos de maior movimentação.

Após a aplicação de filtros, foram gerados gráficos de barras com as bibliotecas Matplotlib e Seaborn, que facilitaram a visualização das tendências nos dados. Esse processo possibilitou uma interpretação mais clara dos padrões ao longo do tempo e uma visão mais detalhada dos fatores observados.

3.8 Disponibilização das Análises e Resultados

Todas as etapas seguidas e descritas aqui neste capítulo estão disponíveis em um repositório no GitHub ¹¹ permitindo que outras pessoas possam fazer outras análises ou darem continuidade ao trabalho.

¹¹ <https://github.com/DiogoM-Almeida/meu_tcc>

4 Resultados e análises

Neste capítulo, serão apresentados e discutidos os resultados obtidos das análises feitas, descritas no Capítulo 3.

4.1 Análise Geoespacial

A Figura 4 apresenta as viagens a lazer realizadas por pedestres em São Paulo. As vias com cores amarelas representam as que possuem maior concentração de viagens, enquanto as azuis indicam menor quantidade de viagens. Já as vias em tons de cinza indicam aquelas sem registro de viagens. O mapa destaca a área sudoeste de São Paulo, com o Rio Pinheiros e a Cidade Universitária à esquerda, o Parque do Ibirapuera mais abaixo e o Centro da cidade no canto superior direito. A maioria das áreas com cores mais claras está localizada ao redor de parques, o que se explica pelo fato de a maioria das viagens registradas no aplicativo estar associada a atividade de lazer. Se forem separados os tipos das viagens, nas Figuras 6 e 7 pode-se ver isso claramente.

Figura 6 – Viagens de lazer na cidade de São Paulo-2019 a 2024



Fonte: Elaborada pelo autor.

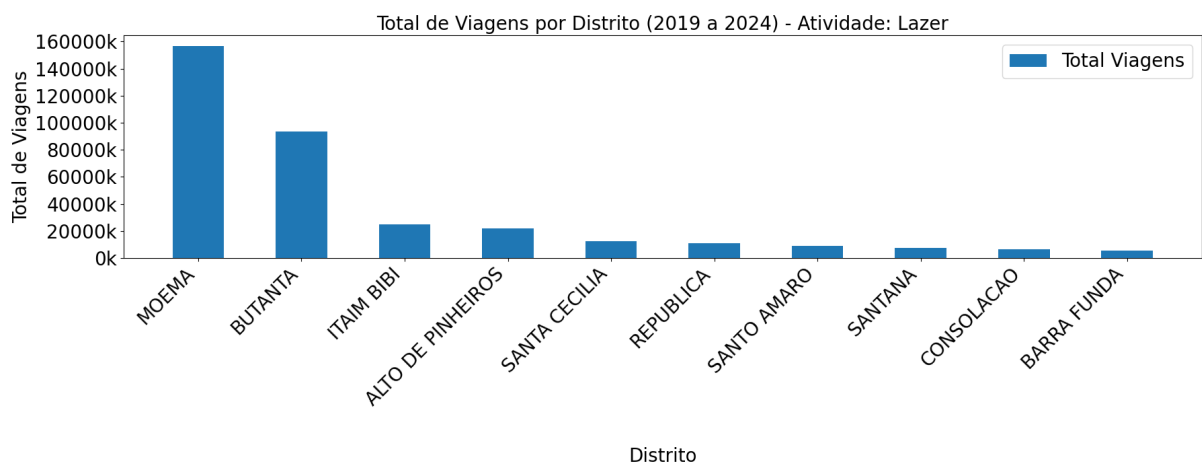
Figura 7 – Viagens de deslocamentos na cidade de São Paulo-2019 a 2024



Fonte: Elaborada pelo autor.

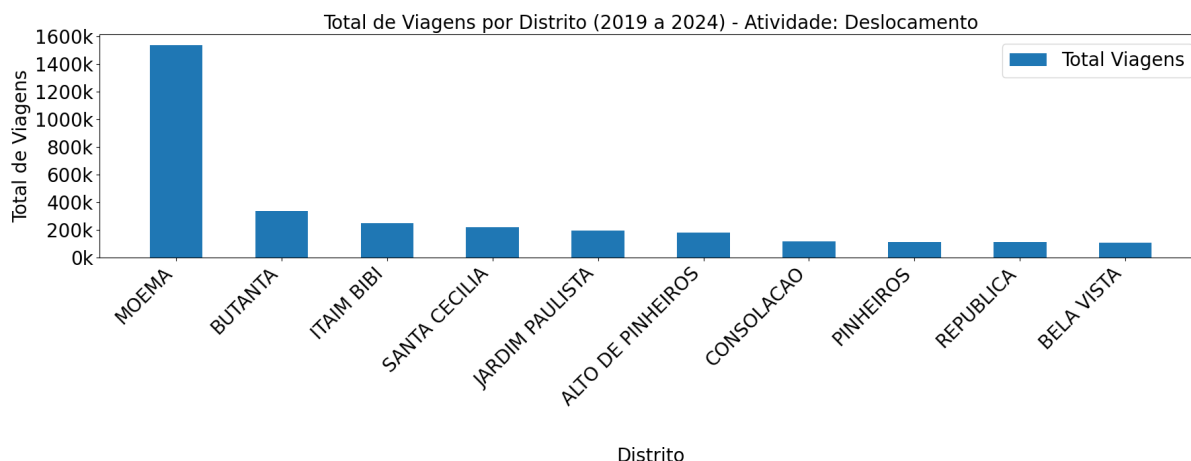
Consegue-se agora ver nas Figuras 6 e 7 que ao contrário das viagens a lazer realizadas geralmente perto de parques, as de deslocamento se concentram principalmente em avenidas. Para poder-se ter um melhor entendimento, também é importante saber como é a distribuição por distrito.

Figura 8 – Distritos com maior número de viagens de lazer-2019 a 2024



Fonte: Elaborada pelo autor.

Figura 9 – Distritos com maior número de viagens de deslocamento-2019 a 2024



Fonte: Elaborada pelo autor.

As Figuras 8 e 9 mostram os distritos com mais viagens de lazer e deslocamento, respectivamente. Apesar da diminuição do número de viagens, o distrito com mais viagens permaneceu o mesmo, sendo ele o distrito de Moema com maior número de viagens, sejam elas de lazer ou deslocamento. Pode-se observar também uma distribuição mais homogênea na Figura 9, em comparação a Figura 8, além de uma mudança nos distritos apresentados.

Se for plotado o mapa do distrito de Moema, Figura 11(a) e 11(b), pode-se ver em mais detalhes as observações anteriores no mapa, com o maior número de viagens a lazer no parque, Ibirapuera com uma somatória de 900.000 viagens, e o maior número de deslocamento sendo em suas avenidas, com a avenida Indianópolis com uma somatória de 205 viagens.

A partir dessa análise, podemos inferir que a Avenida Indianópolis, devido ao seu alto volume de tráfego pedonal, deveria ser priorizada para a manutenção e melhoria das calçadas. A manutenção adequada das calçadas é fundamental para garantir a segurança e o conforto dos pedestres, reduzindo riscos de acidentes e incentivando o deslocamento a pé. Além disso, calçadas bem conservadas promovem a acessibilidade para pessoas com mobilidade reduzida, contribuindo para uma cidade mais inclusiva.

Figura 10 – Comparação dos mapas de lazer e deslocamento de Moema-2019 a 2024



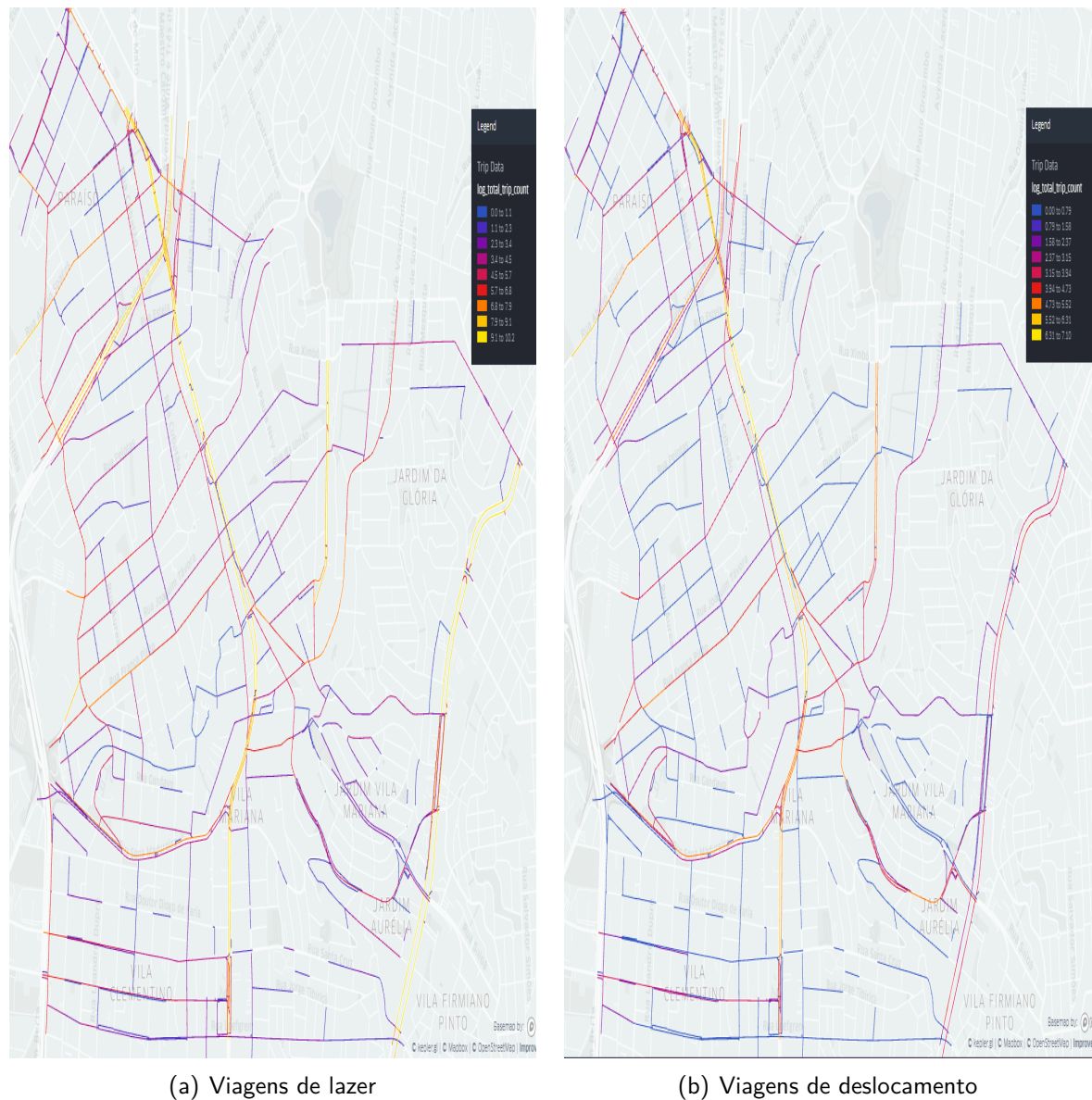
Fonte: Elaborada pelo autor.

Se for plotado o mapa do distrito da Vila Mariana, Figura 12(a) e 12(b), pode-se ver em mais detalhes as observações anteriores no mapa, com o maior número de viagens a lazer na Avenida Bernardino de Campos, totalizando 11.000 viagens, e o maior número de deslocamento sendo na Rua Vergueiro, com uma somatória de 310 viagens.

A partir dessa análise, pode-se inferir que a Rua Vergueiro, devido ao seu alto volume de tráfego pedonal, deveria ser priorizada para a manutenção e melhoria das calçadas. A manutenção adequada das calçadas é fundamental para garantir a segurança e o conforto dos pedestres, reduzindo riscos de acidentes e incentivando o deslocamento a pé. Além disso, calçadas bem conservadas promovem a acessibilidade para pessoas com mobilidade reduzida, contribuindo para uma cidade mais inclusiva.

Pode-se observar que a Avenida Bernardino de Campos é um importante polo de lazer no distrito, atraindo muitas pessoas para atividades recreativas. Portanto, investir em infraestrutura de lazer, como áreas verdes, mobiliário urbano e segurança, é essencial para melhorar a experiência dos frequentadores. Além disso, garantir calçadas amplas e acessíveis na avenida contribui para a mobilidade dos pedestres e para o incentivo de atividades ao ar livre.

Figura 11 – Comparação dos mapas de lazer e deslocamento de Vila Mariana-2019 a 2024



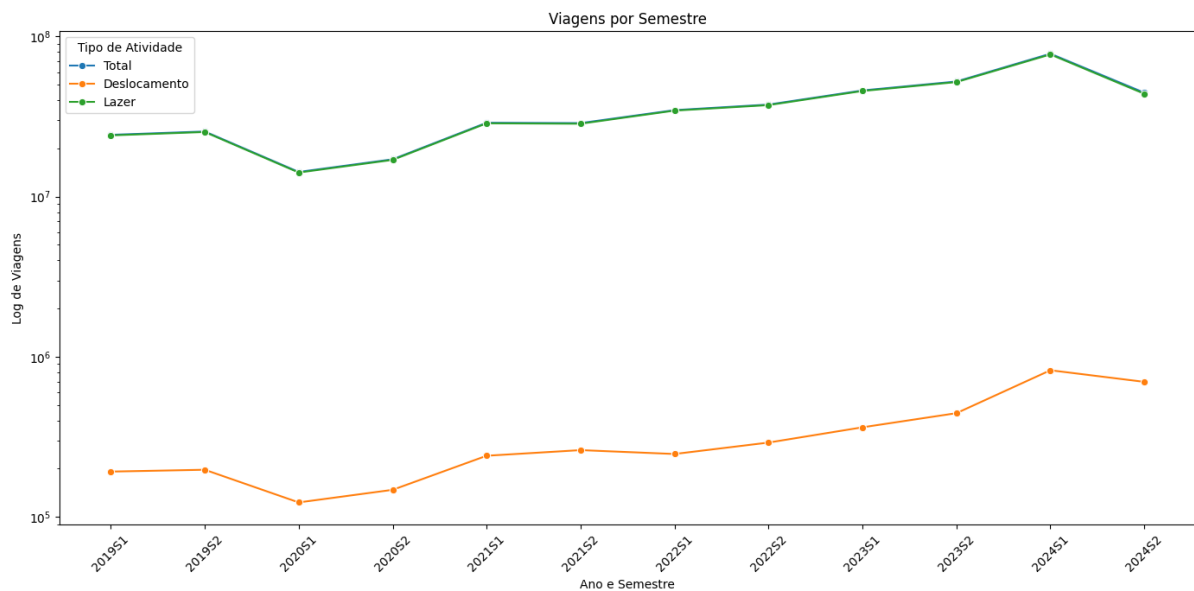
Fonte: Elaborada pelo autor.

4.2 Análise da mobilidade durante a pandemia da COVID-19

Os dados do aplicativo de rastreamento de atividades físicas, abrangendo o período de 2019 a 2024, permitem uma análise detalhada do comportamento dos pedestres, considerando

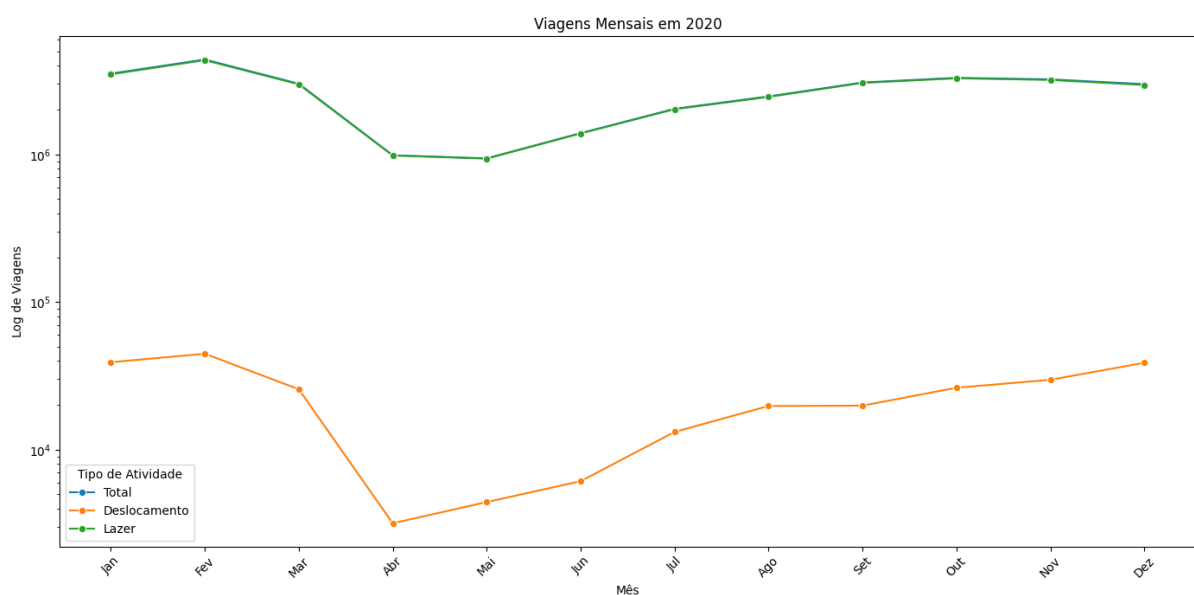
as mudanças antes, durante e após a pandemia. Essa abordagem possibilita avaliar a evolução nas práticas de mobilidade e os impactos das políticas públicas implementadas no contexto da pandemia, oferecendo insights valiosos sobre a eficácia dessas medidas e as mudanças nos padrões de deslocamento e atividade física ao longo desses anos.

Figura 12 – Viagens na Pandemia por Semestre-2019 a 2024



Fonte: Elaborada pelo autor.

Figura 13 – Viagens na Pandemia - 2020



Fonte: Elaborada pelo autor.

A pandemia de COVID-19 no Brasil iniciou-se oficialmente em fevereiro de 2020, quando os primeiros casos foram confirmados. Em março, o país implementou medidas de isolamento social, incluindo o fechamento de escolas, comércios e a restrição de atividades não essenciais, na tentativa de conter a propagação do vírus. O auge das mortes ocorreu entre maio e julho de 2020, e novamente no início de 2021, representando um dos períodos mais críticos para o sistema de saúde brasileiro. A partir do segundo semestre de 2021, com o avanço da vacinação, observou-se uma diminuição nos casos e mortes, permitindo uma gradual retomada das atividades econômicas e sociais.

A Figura 13 e a 12 mostra que, antes da pandemia, o número de viagens se manteve estável, com variações concentradas nos meses de férias, como junho e dezembro. No entanto, em 2020, durante a pandemia, observa-se uma queda acentuada no número de viagens em março, atingindo seu ponto mais baixo em abril. Em seguida, os dados indicam uma retomada gradual das viagens, sugerindo uma adesão parcial às políticas públicas de restrição daquele período. Nos anos subsequentes, o padrão de viagens retorna ao observado no período pré-pandemia, com um aumento notável no último ano, refletindo uma possível retomada e até crescimento no uso de mobilidade ativa.

Os meses selecionados — fevereiro, março e abril de 2020 e de 2021, Figuras 14 a 19 respectivamente — oferecem uma perspectiva interessante sobre o impacto da pandemia de COVID-19 na mobilidade das pessoas no Brasil.

Em fevereiro de 2020, antes das medidas restritivas, o número de viagens mantinha-se nos padrões habituais, refletindo uma normalidade pré-pandêmica. As atividades econômicas e sociais ocorriam sem grandes interferências, e a mobilidade não apresentava variações significativas.

Com a chegada de março de 2020, iniciam-se as medidas de isolamento social, como o fechamento de escolas e comércios. Isso resultou em uma redução acentuada no número de viagens, à medida que a população começou a aderir às restrições impostas. A queda mais significativa é observada em abril de 2020, quando as restrições estavam plenamente vigentes e a mobilidade atingiu seu ponto mais baixo. Esse período reflete uma alta adesão às políticas públicas e um esforço coletivo para conter a propagação do vírus.

Ao avançarmos para fevereiro de 2021, notamos uma retomada gradual das viagens em comparação com abril de 2020. Apesar de a pandemia ainda estar em curso, houve uma flexibilização das restrições e um aumento na mobilidade, possivelmente devido ao cansaço pandêmico e à necessidade econômica. Em março e abril de 2021, embora o país enfrentasse uma nova onda de casos, as viagens não diminuíram tanto quanto no mesmo período do ano anterior. Isso sugere uma menor adesão às medidas restritivas ou uma adaptação da população às circunstâncias, buscando equilibrar prevenção e manutenção das atividades diárias.

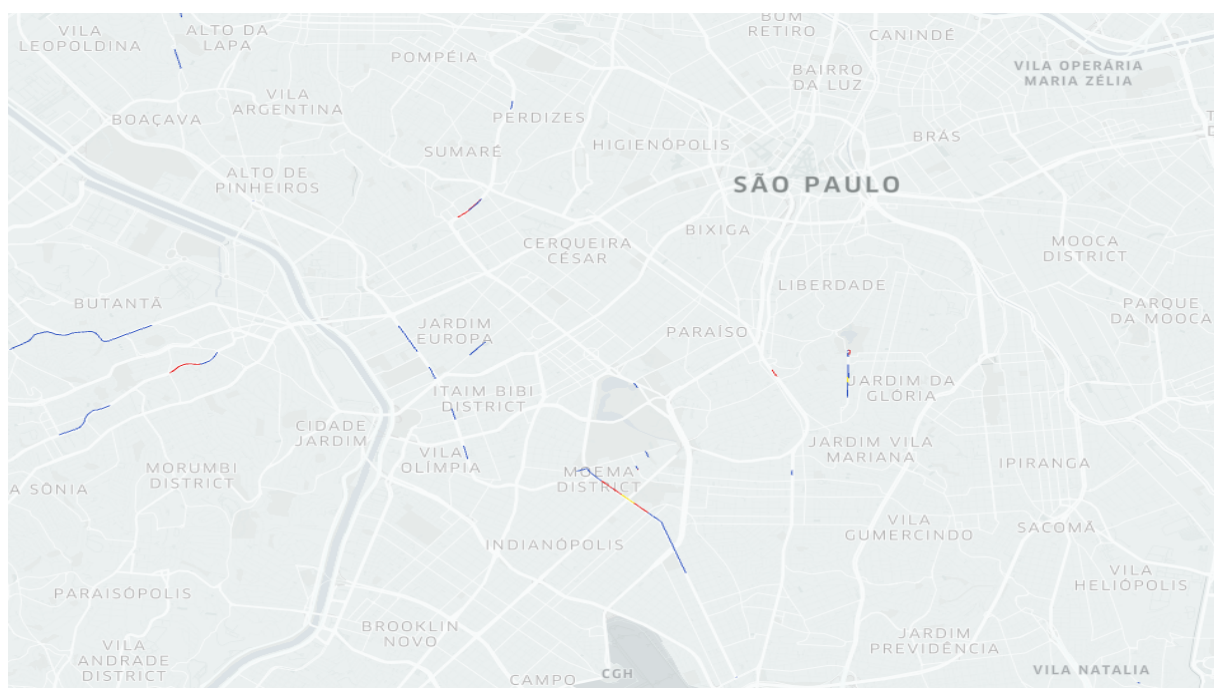
As semelhanças entre os meses de março e abril de ambos os anos residem na tentativa

Figura 15 – Viagens Deslocamento na Pandemia-Março 2020



Fonte: Elaborada pelo autor.

Figura 16 – Viagens Deslocamento na Pandemia-Abril 2020



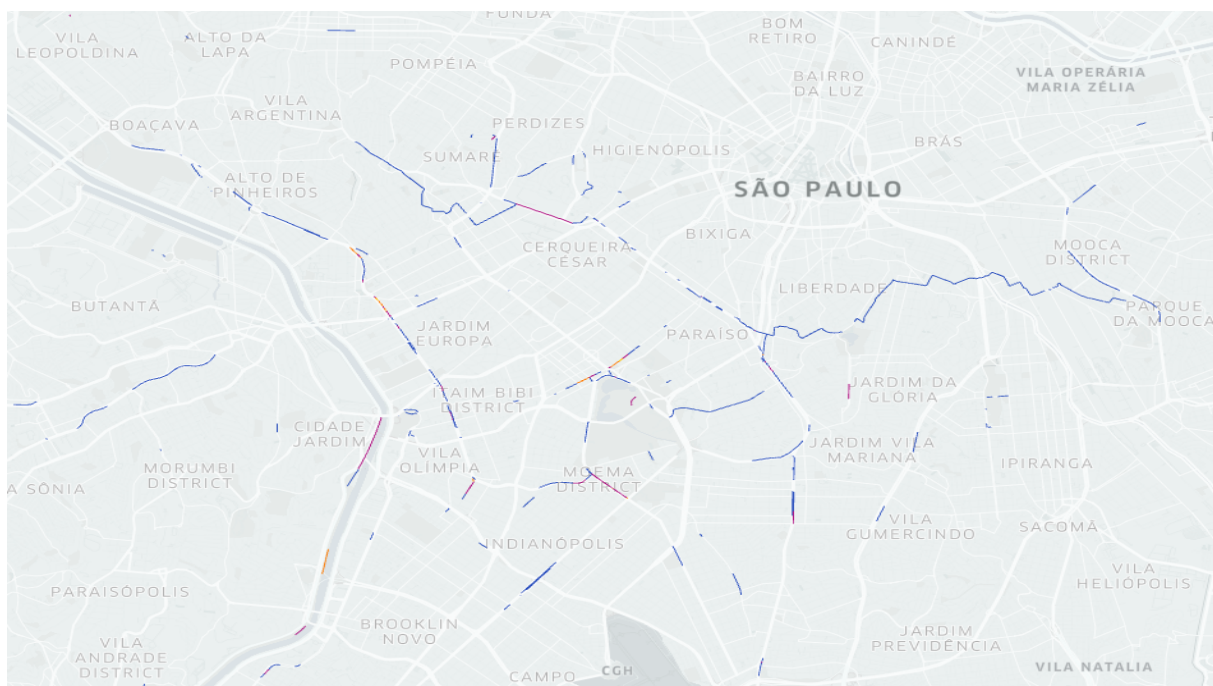
Fonte: Elaborada pelo autor.

Figura 17 – Viagens Deslocamento na Pandemia-Fevereiro 2021



Fonte: Elaborada pelo autor.

Figura 18 – Viagens Deslocamento na Pandemia-Março 2021



Fonte: Elaborada pelo autor.

Figura 19 – Viagens Deslocamento na Pandemia-Abril 2021



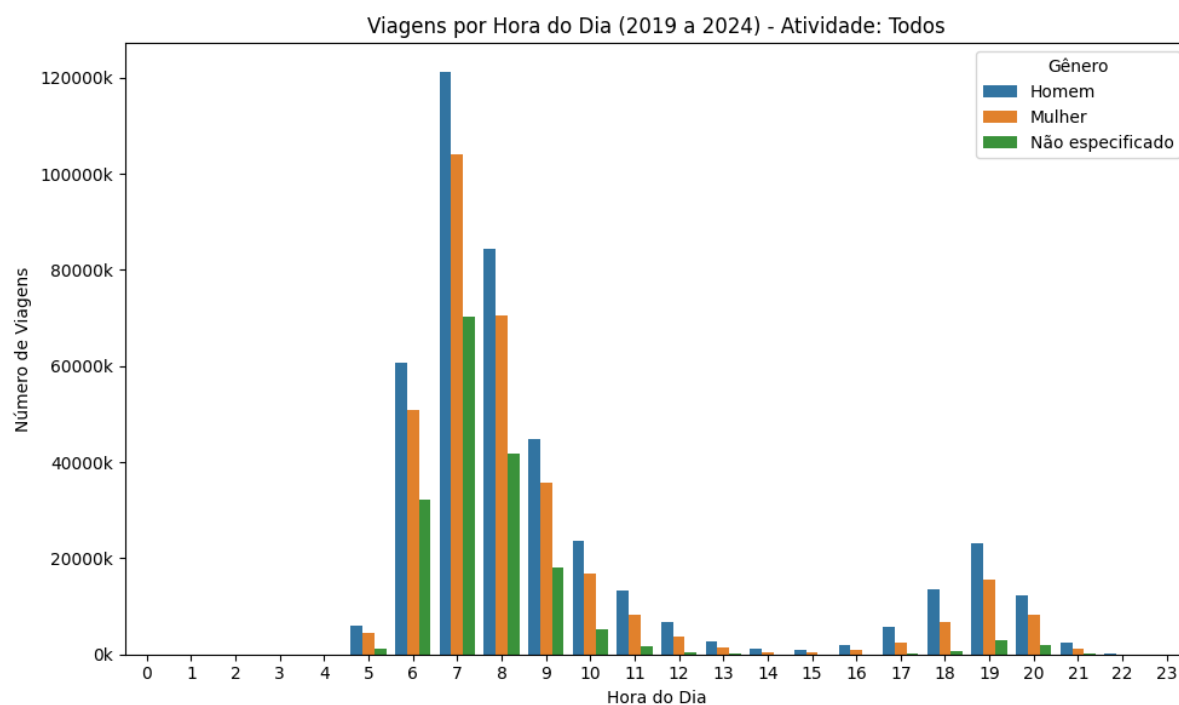
Fonte: Elaborada pelo autor.

4.3 Análises das Características das Viagens de Pedestres

Neste capítulo, exploram-se as características das viagens de pedestres com base em diferentes parâmetros. A análise abrange aspectos como gênero (homens e mulheres), faixa etária (18 a 34 anos, 35 a 54 anos, 55 a 64 anos e 65 anos ou mais), além de considerações sobre os horários do dia, dias da semana e meses do ano em que as viagens ocorrem. Busca-se identificar padrões e tendências nos deslocamentos dos pedestres, compreendendo como variáveis demográficas e temporais influenciam o uso da plataforma. Essas informações são fundamentais para aprimorar serviços, planejar infraestruturas e desenvolver políticas que atendam às necessidades específicas dos usuários em diferentes contextos.

Como se pode observar na Figura 20, os horários de pico de viagens ocorrem principalmente durante os períodos de ida e volta ao trabalho, com maior volume de viagens entre as 07h e 08h da manhã e entre as 18h e 19h. Ao analisarmos os dados separadamente por gênero, vemos que a tendência se mantém, mas com um número maior de viagens realizadas por homens, indicando um volume superior de homens utilizando a plataforma em comparação às mulheres.

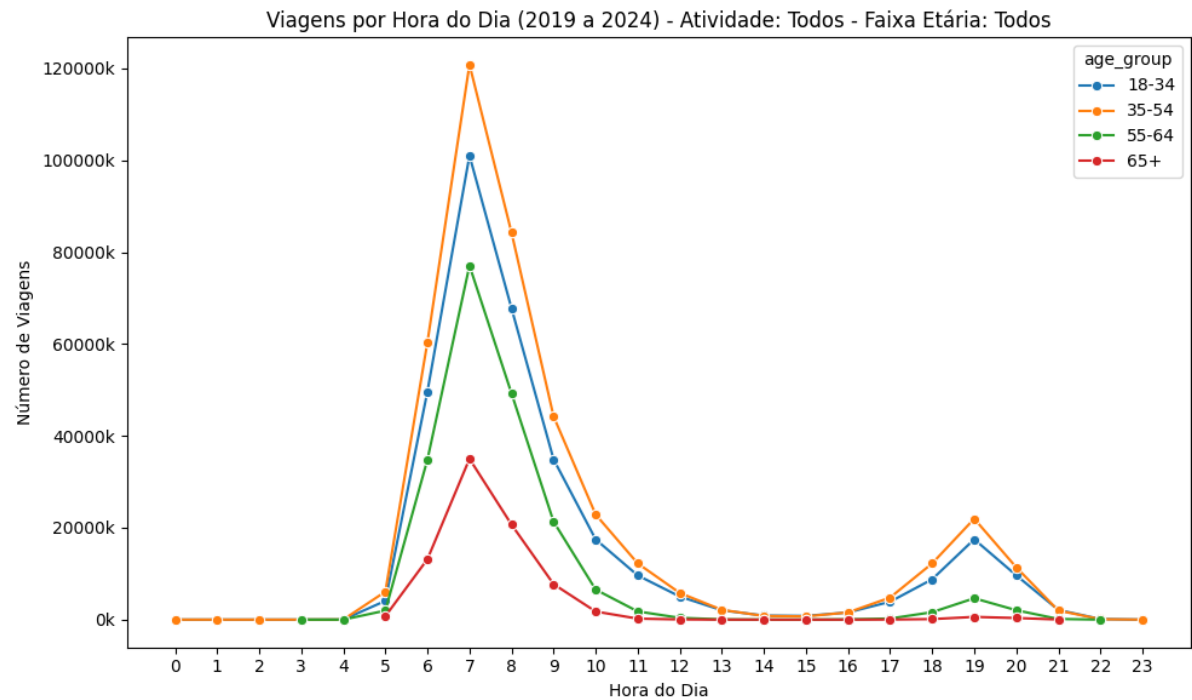
Figura 20 – Viagens Horário-2019 a 2024



Fonte: Elaborada pelo autor.

Ao se observarem as viagens por faixa etária, na Figura 21, o padrão de pico nos horários de 07h-08h e 18h-19h também é evidente. A faixa etária de 35 a 54 anos registra o maior número de viagens, enquanto a faixa de 65 anos ou mais apresenta o menor volume, com essas viagens ocorrendo principalmente pela manhã.

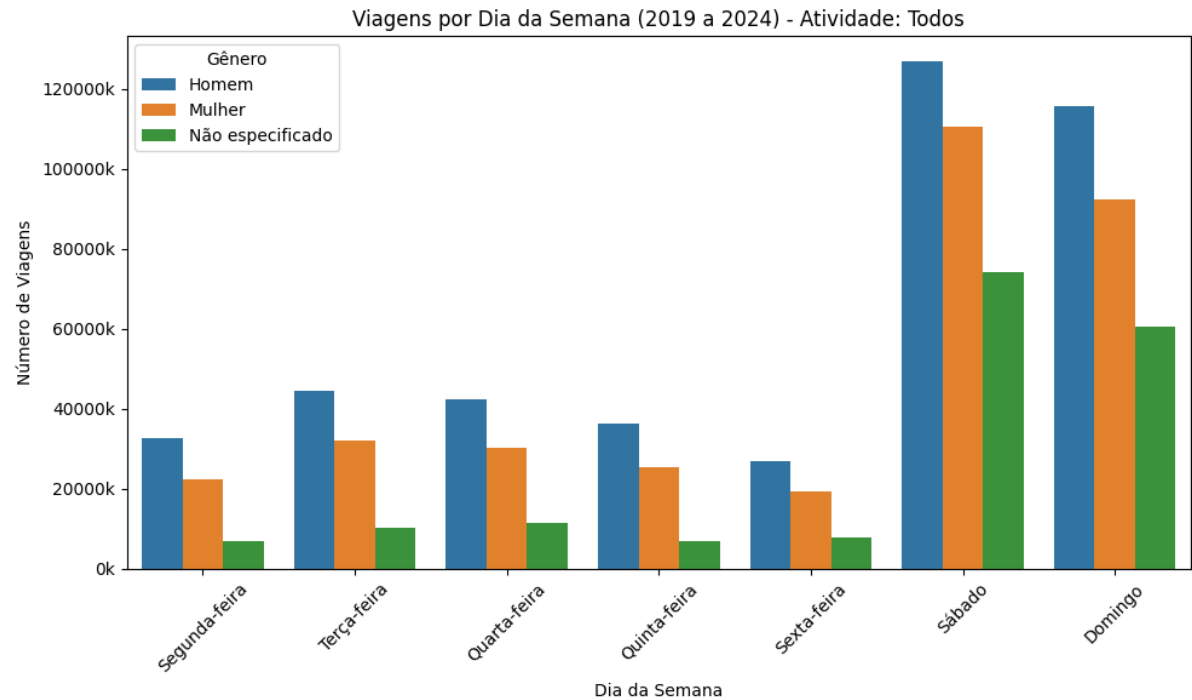
Figura 21 – Viagens por Faixa Etária por Horário-2019 a 2024



Fonte: Elaborada pelo autor.

Quanto à distribuição das viagens ao longo da semana, é possível notar na Figura 22 que os fins de semana concentram um número significativamente maior de viagens em comparação aos dias úteis, que têm uma distribuição mais uniforme ao longo dos dias. Quando a análise é feita por gênero, observa-se que os homens realizam mais viagens em todos os dias da semana, em relação às mulheres.

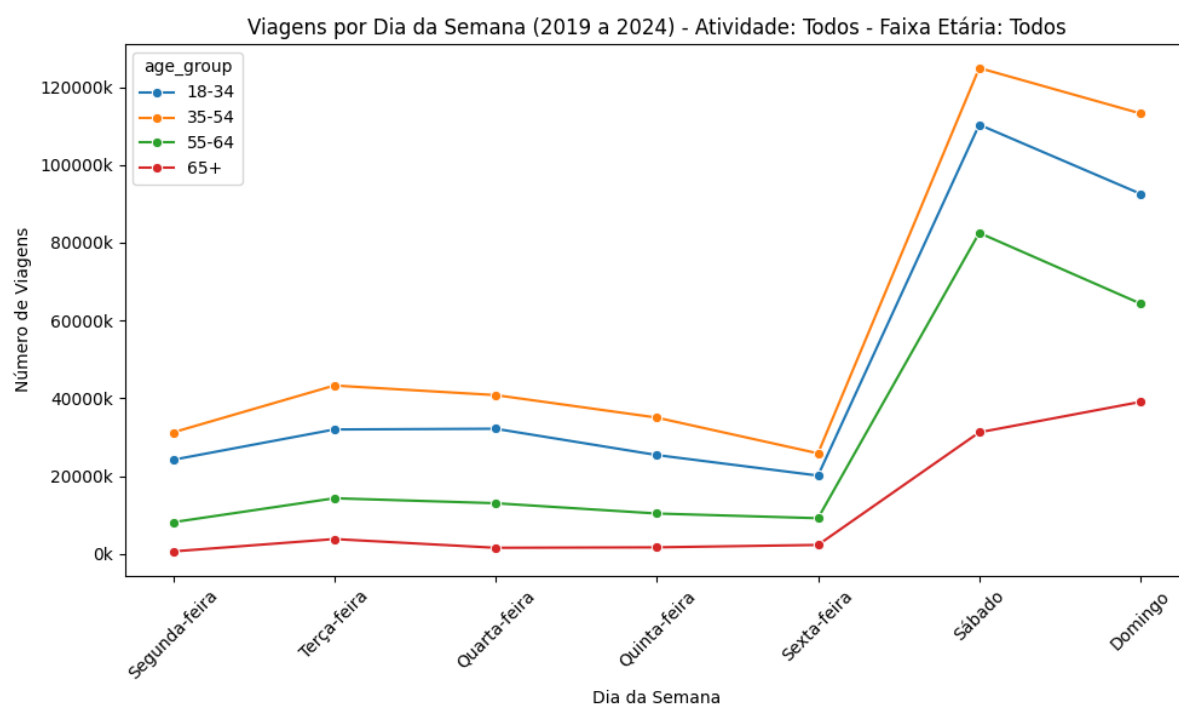
Figura 22 – Viagens por Dia da Semana-2019 a 2024



Fonte: Elaborada pelo autor.

Ao segmentar as viagens por faixa etária, podemos ver na Figura 23, que o padrão se mantém, com os fins de semana apresentando mais viagens, especialmente na faixa de 35 a 54 anos, e a faixa etária de 65+ registrando o menor número de viagens, principalmente durante os dias úteis.

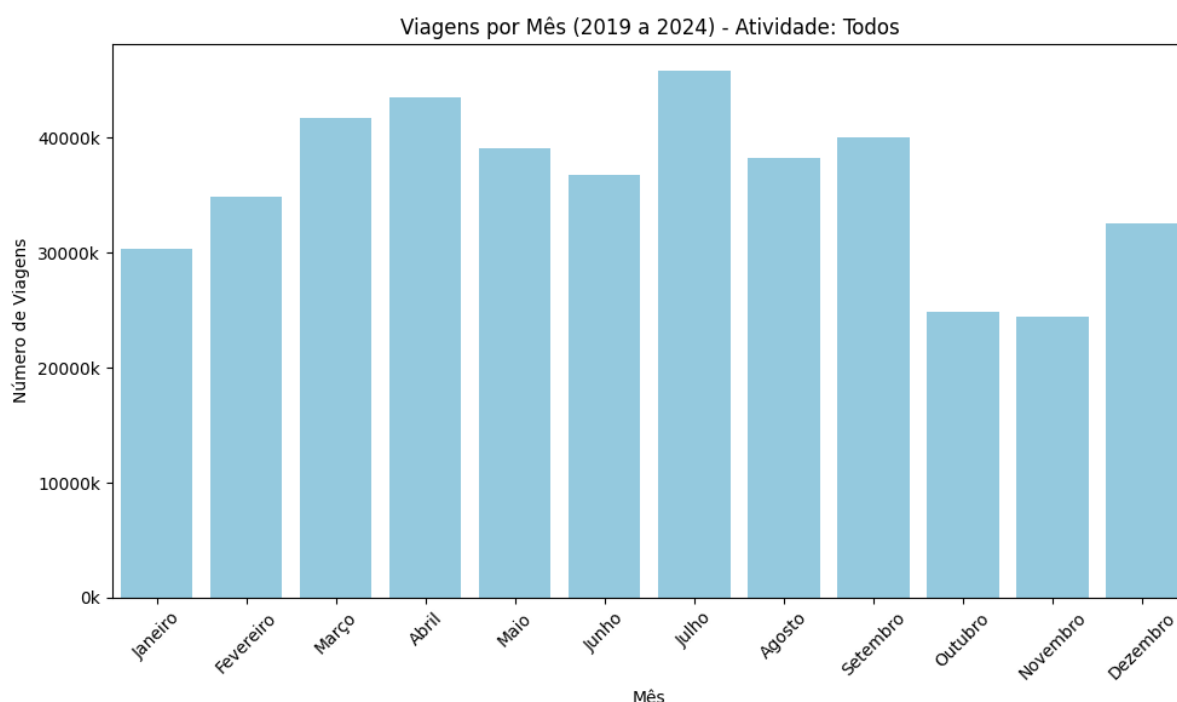
Figura 23 – Viagens das Faixas Etárias por Dia da Semana-2019 a 2024



Fonte: Elaborada pelo autor.

Finalmente, ao analisar os dados mensais na Figura 24, percebe-se uma distribuição relativamente homogênea, com a maioria das viagens ocorrendo na primeira metade do ano. A partir de junho, no entanto, observa-se uma queda no número de viagens, seguida por um leve aumento nos meses de férias, como julho e dezembro.

Figura 24 – Viagens por Meses-Todos Gêneros e Todas as Faixas Etárias-2019 a 2024



Fonte: Elaborada pelo autor.

4.4 Desafios Enfrentados

Ao longo do desenvolvimento deste trabalho, diversos desafios foram enfrentados no manuseio e análise do conjunto de dados devido ao seu grande volume e complexidade. Dois aspectos principais se destacaram: problemas relacionados à memória e limitações de ferramentas de visualização geoespacial.

O primeiro desafio significativo foi o gerenciamento eficiente da memória do sistema. O conjunto de dados utilizado era extremamente volumoso, contendo milhões de registros que exigiam processamento e análise detalhados. Operações que, em condições normais, seriam executadas rapidamente tornaram-se lentas e, em alguns casos, inviáveis devido ao consumo excessivo de memória RAM.

Apesar de diversas tentativas para mitigar este problema, como a implementação de técnicas de processamento em blocos (chunking) e a utilização de estruturas de dados mais eficientes, o desafio de memória permaneceu não solucionado. As limitações de hardware do sistema utilizado não permitiram o processamento completo do conjunto de dados, restringindo a capacidade de análise e impactando potencialmente a profundidade dos insights obtidos. Este obstáculo evidenciou a necessidade de recursos computacionais mais robustos para lidar com big data, bem como a importância de considerar essas limitações durante o planejamento do

projeto.

Outro desafio crítico enfrentado foi na etapa de visualização dos dados geoespaciais. Inicialmente, a biblioteca Folium foi escolhida para a plotagem de mapas interativos devido à sua simplicidade e integração com o Python. No entanto, ao tentar plotar um número elevado de arestas, a ferramenta mostrou-se inadequada. Por ser baseada em renderização via CPU, o Folium não conseguiu lidar eficientemente com a quantidade massiva de elementos gráficos, resultando em tempos de carregamento elevados e, frequentemente, em travamentos do sistema.

Essa limitação impediu a exploração visual dos dados de maneira eficaz, o que é crucial para a identificação de padrões e insights relevantes. A necessidade de uma solução mais robusta tornou-se evidente.

Para superar as limitações impostas pelo Folium, optou-se por migrar para a biblioteca Kepler.gl. Diferentemente do Folium, o Kepler.gl é construído sobre a tecnologia WebGL, permitindo a renderização gráfica utilizando a GPU em vez da CPU. Essa mudança foi fundamental para lidar com o grande volume de arestas, pois a GPU é otimizada para processar operações gráficas complexas de maneira paralela e eficiente.

A adoção do Kepler.gl proporcionou uma experiência de visualização fluida e responsiva, mesmo com conjuntos de dados extensos. Além disso, a ferramenta ofereceu recursos avançados de interação e personalização dos mapas, enriquecendo a análise e interpretação dos resultados.

4.5 Considerações

Os dados das análises realizados acima mostram o cenário da mobilidade ativa focada no pedestre. As análises geoespaciais permitem observar a distribuição espacial das atividades de lazer e deslocamento dos usuários do aplicativo na cidade de São Paulo, com destaque para a concentração de viagens de lazer próximas a parques e viagens de deslocamento ao longo de grandes avenidas. Essas distribuições evidenciam o uso diferenciado do espaço urbano, com atividades de lazer predominando em áreas recreativas e os deslocamentos focados em vias principais. Ao observarmos a distribuição por distrito, notamos padrões específicos, como o distrito de Moema, que se destaca tanto nas viagens de lazer quanto nas de deslocamento, sugerindo seu papel central na mobilidade urbana e sua proximidade a importantes áreas de lazer como o parque, Ibirapuera.

A análise da pandemia permite visualizar as mudanças nos padrões de mobilidade ao longo de 2019 a 2024, refletindo o impacto das restrições de mobilidade e o efeito das políticas públicas sobre a circulação dos pedestres. Durante o período inicial da pandemia, observou-se uma queda acentuada nas viagens, seguida de uma retomada gradual, indicando uma adaptação parcial dos usuários às restrições. Nos anos seguintes, o aumento gradual no número de viagens

sinaliza uma recuperação no uso da mobilidade ativa, sugerindo que as atividades físicas ao ar livre, especialmente em áreas de lazer, podem ter ganhado popularidade como uma alternativa segura de recreação.

A análise das características dos usuários demonstra padrões comportamentais que refletem preferências e rotinas distintas entre gêneros, faixas etárias e períodos do dia. Os picos de viagens nos horários de ida e volta ao trabalho (07h-08h e 18h-19h) indicam um padrão de deslocamento bem definido, com a faixa etária de 35 a 54 anos registrando o maior volume de viagens. Nos fins de semana, a maior concentração de atividades aponta para um aumento das viagens de lazer, com os homens realizando mais viagens em todos os dias da semana. Esse comportamento revela diferenças no uso da plataforma e na mobilidade urbana entre os diferentes perfis demográficos, com um aumento considerável nas viagens de lazer durante os fins de semana e um uso mais intensivo entre adultos de meia-idade.

Essas análises contribuem para entender melhor como diferentes fatores — como a geografia da cidade, a pandemia e as características demográficas — influenciam a mobilidade urbana e a interação dos pedestres com o ambiente, oferecendo insights relevantes para o planejamento urbano e a promoção de políticas públicas que incentivem o uso de transporte ativo e sustentável.

Com base nos resultados obtidos, é possível direcionar ações específicas para melhorar a infraestrutura urbana, especialmente no que diz respeito à manutenção de calçadas. A identificação de vias com alto volume de tráfego pedonal, como a Avenida Indianópolis no distrito de Moema e a Rua Vergueiro na Vila Mariana, permite que as autoridades priorizem intervenções nesses locais. Ao concentrar esforços na manutenção e melhoria das calçadas em áreas de maior uso pelos pedestres, não apenas se aumenta a segurança e o conforto dos transeuntes, mas também se promove a acessibilidade e incentiva-se o uso contínuo de transporte ativo. Dessa forma, os insights gerados pelas análises podem orientar políticas públicas mais eficientes, garantindo que os recursos sejam aplicados onde são mais necessários e beneficiando diretamente a mobilidade urbana sustentável.

5 Conclusão

O presente trabalho visa apresentar a análise da distribuição de viagens com base nos dados fornecidos pelo aplicativo de rastreamento de atividades físicas, buscando compreender a mobilidade ativa e o panorama das viagens na cidade de São Paulo. Para isso, foram utilizadas técnicas de análise de dados geoespaciais, descritivas e exploratórias, complementadas por visualizações gráficas e mapas gerados pela ferramenta desenvolvida.

Essa análise de dados possibilita a tomada de decisões fundamentadas, permitindo que autoridades competentes, como a prefeitura, identifiquem os locais prioritários para manutenção de calçadas, avaliem a eficácia de políticas públicas e acompanhem o impacto das intervenções, com base em visualizações e métricas fornecidas pela pesquisa.

5.1 Trabalhos futuros

É importante ressaltar que a presente pesquisa serviu como uma base inicial para a compreensão da mobilidade ativa em São Paulo. Os resultados obtidos evidenciam a necessidade de aprofundar a análise em diferentes aspectos, visando contribuir de forma mais significativa para o planejamento urbano e para a melhoria da infraestrutura destinada aos pedestres. O notebook com todos os processos descritos nesse trabalho e todas as análises feitas estão disponíveis no repositório indicado em 3.8 e permite que futuros trabalhos relacionados a essa área se apoiem nele como:

- Comparar as análises deste trabalho com os dados da Pesquisa Origem-Destino do Metrô de São Paulo¹.
- Usar dados de acidentes envolvendo pedestres, para assim geolocalizar os acidentes e relacioná-los com áreas de maior fluxo de pedestres.
- Coletar dados sobre a qualidade e características das calçadas em diferentes locais e cruzar esses dados com as contagens de pedestres e dados de acidentes para identificar correlações.
- Obter as rotas mais comuns para pedestres.
- Dashboard que possibilite alterar escolher as cores e escalas dos valores dos gráficos.

¹ <<https://transparencia.metrosp.com.br/dataset/pesquisa-origem-e-destino/resource/4362eaa3-c0aa-410a-a32b-37355c091075>>

Referências

- BONGIORNO, C.; SANTUCCI, D.; KON, F.; SANTI, P.; RATTI, C. Comparing bicycling and pedestrian mobility: Patterns of non-motorized human mobility in Greater Boston. *Journal of Transport Geography*, v. 80, n. C, 2019. Disponível em: <<https://ideas.repec.org/a/eee/jotrge/v80y2019ics0966692319300870.html>>.
- CRUZ, S. S.; PAULINO, S. R. Desafios da mobilidade ativa na perspectiva dos serviços públicos: experiências na cidade de são paulo. *urbe. Revista Brasileira de Gestão Urbana*, Pontifícia Universidade Católica do Paraná, v. 11, p. e20190026, 2019. ISSN 2175-3369. Disponível em: <<https://doi.org/10.1590/2175-3369.011.e20190026>>.
- DEGREAS, H. *Deficientes são as cidades, não os seus cidadãos – Jovem Pan*. 2022. Disponível em: <<https://jovempan.com.br/opiniao-jovem-pan/comentaristas/helena-degreas/deficientes-sao-as-cidades-nao-os-seus-cidadaos.html>>.
- FAN, Z.; LOO, B. P. Street life and pedestrian activities in smart cities: opportunities and challenges for computational urban science. *Computational Urban Science*, v. 1, n. 1, p. 26, nov. 2021. ISSN 2730-6852. Disponível em: <<https://doi.org/10.1007/s43762-021-00024-9>>.
- GALANIS, A.; BOTZORIS, G.; ELIOU, N. Pedestrian road safety in relation to urban road type and traffic flow. *Transportation Research Procedia*, v. 24, p. 220–227, 12 2017.
- HÜBNER, A. T. Big data sob a ótica de modelagem de dados multidimensional para soluções de business intelligence. 2020. Disponível em: <<https://repositorio.ucs.br/xmlui/handle/11338/6327>>.
- IBM. *O que são dados geoespaciais?* 2024. Acesso em: 20 nov. 2024. Disponível em: <<https://www.ibm.com/br-pt/topics/geospatial-data>>.
- LEMOS, E. D. Análise exploratória de dados: prerrogativas de uma inferência estatística eficiente. 2022.
- LITMAN, T.; BURWELL, D. Issues in sustainable transportation. *International Journal of Global Environmental Issues*, v. 6, p. 331–347, 02 2006.
- PAWSON, R. Evidence-based policy: In search of a method. *Evaluation*, v. 8, n. 2, p. 157–181, 2002. Disponível em: <<https://doi.org/10.1177/1358902002008002512>>.
- PROVOST, F.; FAWCETT, T. Data science and its relationship to big data and data-driven decision making. *Big Data*, v. 1, n. 1, p. 51–59, 2013. PMID: 27447038. Disponível em: <<https://doi.org/10.1089/big.2013.1508>>.
- PUCHER, J.; BUEHLER, R.; MEROM, D.; BAUMAN, A. Walking and cycling in the united states, 2001-2009: Evidence from the national household travel surveys. *American journal of public health*, v. 101 Suppl 1, p. S310–7, 05 2011.
- REIS, E. A.; REIS, I. A. Análise descritiva de dados. *Relatório Técnico do Departamento de Estatística da UFMG*, v. 1, 2002.
- WILKINSON, L.; FRIENDLY, M. *The Grammar of Graphics*. 2nd. ed. Springer Science & Business Media, 2009. Disponível em: <<https://doi.org/10.1007/978-0-387-24544-1>>.