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

LUCA MELO MUNEKATA

ANÁLISE DE MOBILIDADE URBANA UTILIZANDO DADOS DE VIAGENS DE BICICLETA

BAURU
Novembro/2024

LUCA MELO MUNEKATA

ANÁLISE DE MOBILIDADE URBANA UTILIZANDO DADOS DE VIAGENS DE BICICLETA

Trabalho de Conclusão de Curso do curso de Bacharelado em Ciência da Computação da Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Ciências, Campus Bauru.
Orientador: Prof. Dr. Higor Amario de Souza

BAURU
Novembro/2024

M965a

Munekata, Luca Melo
Análise de mobilidade urbana utilizando dados de viagens de bicicleta / Luca Melo Munekata. -- Bauru, 2024
60 p. : il., mapas

Trabalho de conclusão de curso (Bacharelado - Ciência da Computação) - Universidade Estadual Paulista (UNESP), Faculdade de Ciências, Bauru
Orientador: Higor Amario de Souza

1. Ciência de dados. 2. Mobilidade ativa. 3. Ciclistas. 4. Infraestrutura cicloviária. 5. Políticas públicas. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Dados fornecidos pelo autor(a).

Luca Melo Munekata

Análise de mobilidade urbana utilizando dados de viagens de bicicleta

Trabalho de Conclusão de Curso do curso de Bacharelado em Ciência da Computação 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

Profa. Dra. Simone das Graças Domingues Prado
Universidade Estadual Paulista "Júlio de Mesquita Filho"
Faculdade de Ciências
Departamento de Computação

Profa. Dra. Juliana da Costa Feitosa
Universidade Estadual Paulista "Júlio de Mesquita Filho"
Faculdade de Ciências
Departamento de Computação

Bauru, 14 de Novembro de 2024.

Agradecimentos

Utilizo este espaço para agradecer àqueles que de alguma forma me auxiliaram ao longo da minha vida.

Aos meus pais, minha mais profunda gratidão, por tudo que fizeram e continuam por fazer por mim, todos os dias. Por abdicarem de tantas coisas para que eu pudesse estar onde estou hoje. Ao meu irmão pelo companheirismo de todos esses anos. Aos meus demais familiares, por todo apoio e amor desde sempre.

Aos meus amigos, pelos momentos e alegrias. Àqueles que mesmo de longe estiveram comigo ao longo dos anos, e àqueles que passaram por todos os desafios da graduação ao meu lado e que em pouco tempo se tornaram pessoas muito especiais para mim.

Ao meu orientador, Prof. Dr. Higor Amario de Souza, pelo auxílio e ensinamentos ao longo de todo o ano. Aos docentes que de alguma forma impactaram minha formação, em especial aos presentes na banca.

A todos vocês meus mais sinceros agradecimentos.

Resumo

A mobilidade ativa tem ganho relevância em discussões sobre o planejamento urbano, se mostrando um meio de transporte que traz benefícios para a qualidade de vida nas cidades. Entretanto, ainda existem empecilhos para que o ciclismo e outros métodos de mobilidade ativa tenham seu incentivo justificado, como limitações na infraestrutura urbana. Além disso, pesquisas que fornecem dados para estudos na área levam um longo período de tempo para serem atualizadas. Nesse contexto, o presente trabalho tem como objetivo aplicar técnicas de Ciência de Dados no desenvolvimento de uma ferramenta que permita a realização de análises estatísticas e geoespaciais sobre dados de ciclistas de um aplicativo rastreador de viagens na cidade de São Paulo. Visando a implementação de políticas públicas, as análises em questão envolvem a distribuição de viagens de ciclistas pela cidade, as características de viagens e ciclistas, e a relação delas com a infraestrutura dedicada, com a possibilidade de atualizações de dados em intervalos temporais menores. A ferramenta foi desenvolvida utilizando linguagem Python, em *notebooks* Jupyter.

Palavras-chave: Ciência de dados. Mobilidade ativa. Ciclistas. Infraestrutura cicloviária. Políticas públicas.

Abstract

Active mobility has been gaining relevance in discussions on urban planning, as a mean of transportation that benefits quality of life in cities. Nevertheless, there are still obstacles to fully justifying the promotion of cycling and other active mobility methods, such as limitations in urban infrastructure. Furthermore, research that provides data for studies in the field takes a long time to be updated. In this context, the current work aims to apply Data Science techniques in the development of a tool capable of performing statistical and geospatial analyses on cyclists data from a trip-tracking application, in the city of São Paulo. With the goal of supporting public policymaking, these analyses involve examining on the spread of cycling trips throughout the city, the characteristics of trips and cyclists, and their relationship with dedicated infrastructure, with the possibility of data updates at shorter time intervals. The tool was developed using Python in Jupyter notebooks.

Keywords: Data Science. Active mobility. Cyclists. Cycling Infrastructure. Public Policies.

Lista de figuras

Lista de abreviaturas e siglas

CET	Companhia de Engenharia de Tráfego
CSV	*Comma-Separated Values*
DBF	*Database File*
IIQ	Intervalo Interquartil
OD	Pesquisa Origem e Destino
OSM	OpenStreetMap
PPBEs	Políticas Públicas Baseadas em Evidências
PRJ	*Projection File*
RMSP	Região Metropolitana de São Paulo
SHP	*Shapefile*
SHX	*Shape Index*
SIG	Sistema de Informação Geográfica

Sumário

1 Introdução

A questão da mobilidade urbana ativa é um tópico que tem ganho relevância em discussões ao longo dos últimos anos. Além da importância como método de locomoção viável, o transporte ativo também se mostra influente nos âmbitos do planejamento urbano e do desenvolvimento econômico (GERIKE et al., 2019), bem como em indicadores urbanos, como o trânsito, a poluição do ar e a qualidade de vida, fazendo com que as grandes cidades busquem incentivar o aumento do número de ciclistas e pedestres (SARAGIOTTO, 2020).

Aliado ao aumento dos esforços voltados à área, a popularização do uso de dispositivos móveis, como *smartphones*, *laptops* e *smartwatches*, e a difusão de serviços *mobile* de rastreamento baseado em GPS, tem remodelado a experiência de deslocamento e/ou prática de atividades físicas (HAFERMALZ et al., 2020). Aplicativos que monitoram dados de viagens de bicicleta, por exemplo, são sincronizáveis com os mais diversos tipos de dispositivos eletrônicos, permitindo que os usuários monitorem seus treinos, deslocamentos diários a trabalho ou a passeio, e interajam com pessoas de todas as partes do mundo, compartilhando rotas, treinos e métricas de desempenho, agindo como uma rede social para atletas e praticantes casuais. As atividades de cada usuário geram um grande volume de dados de mobilidade, os quais podem ser utilizados para realizar análises sobre a mobilidade urbana ativa (FISCHER; NELSON; WINTERS, 2022).

Neste contexto, o trabalho proposto tem como objetivo utilizar bases de dados de viagens de bicicleta provenientes de um aplicativo de rastreamento de mobilidade ativa[1] e de outras fontes abertas, e, por meio de conceitos e recursos de Ciência de Dados, desenvolver uma ferramenta capaz de analisar e reconhecer padrões e tendências de mobilidade entre ciclistas, sob a ótica da cidade de São Paulo.

1.1 Problemática

Com o aumento das discussões sobre mobilidade ativa urbana e seus conhecidos benefícios a indicadores urbanos, cidades como São Paulo têm buscado incentivar o aumento do número de ciclistas e pedestres. Entretanto, ainda existem empecilhos para que este incentivo seja justificado, como limitações na infraestrutura para tais atividades, qualidade e extensão das ciclovias, condições das calçadas e sinalização de trânsito não adequada, além da questão da segurança (SARAGIOTTO, 2020).

Para tal, informações sobre a circulação de ciclistas, nas mais diferentes localidades da

[1] Durante o texto, iremos nos referir a esses dados como "dados de viagens de bicicleta" ou "aplicativo", por questões de confidencialidade no contrato de acesso a esses dados.

cidade, se mostram de extrema valia para a implementação de Políticas Públicas Baseadas em Evidências (PPBEs), que serão exploradas com mais detalhes na Seção 2.4.

Outro fator é a disponibilidade de dados frequentemente atualizados. Pesquisas como a Origem e Destino (OD), abordada na Seção 2.3, levam um extenso período de tempo para serem concluídas e terem seus resultados publicados (dez anos, no caso da pesquisa citada). Caso relevantes, os resultados obtidos a partir de dados de viagens de bicicleta podem trazer evidências em menores intervalos temporais, mesmo que em menor completude, para a atualização frequente de estudos.

1.2 Justificativa

Tendo em vista a problemática descrita anteriormente, aplicativos com dados de viagens de bicicleta podem fornecer dados muito importantes para o estudo da mobilidade ativa urbana. Devido ao grande volume, variedade e variabilidade destas informações, tarefas como o tratamento, processamento, integração, visualização e análise em tempo real de dados, são cada vez mais importantes (TORRE-BASTIDA et al., 2018).

Para realizar estas tarefas de maneira mais eficiente, a Ciência de Dados se mostra uma solução fundamental, auxiliando no reconhecimento de padrões e extração de outras informações úteis (PROVOST; FAWCETT, 2013). Além disso, é possível aplicar análise de dados geolocalizados, já que os dados de viagens de bicicleta permitem relacionar os comportamentos dos usuários com a localidade em que eles ocorrem, possibilitando uma análise mais rica.

Desse modo, o desenvolvimento de um ambiente que possibilite a análise das atividades/viagens de bicicleta, inseridas no cenário da cidade de São Paulo, se mostra uma opção válida para aplicar técnicas de Ciência de Dados. Utilizando este ambiente, um indivíduo com acesso aos dados do aplicativo pode obter resultados personalizados de acordo com suas necessidades, frequentemente atualizados. Cabe ressaltar que a Secretaria de Mobilidade e Trânsito da Cidade de São Paulo (SMT) também possui acesso a esses dados.

1.3 Objetivos

1.3.1 Objetivo Geral

Desenvolver uma ferramenta capaz de analisar padrões e tendências de mobilidade entre ciclistas usuários do aplicativo de dados de viagens de bicicleta na cidade de São Paulo, utilizando recursos de Ciência de Dados, com potencial para colaborar com a implantação de políticas públicas para mobilidade ativa.

1.3.2 Objetivos Específicos

O presente trabalho tem como objetivos específicos:

- Analisar e identificar e locais com maior e menor circulação de ciclistas;
- Caracterização de usuários do aplicativo de dados de viagens de bicicleta e suas atividades, na cidade de São Paulo;
- Analisar particularidades na mobilidade ativa urbana, em diferentes granularidades espaciais (zonas, distritos, subprefeituras).
- Identificar relações entre padrões de deslocamento e características locais da cidade de São Paulo;
- Comparar padrões de deslocamento atuais encontrados com pesquisas anteriores realizadas na cidade; e
- Identificar locais nos quais a melhoria da infraestrutura urbana pode beneficiar a popularização do uso de bicicletas.

2 Fundamentação Teórica

Neste capítulo, serão introduzidos conceitos sobre mobilidade ativa, infraestrutura cicloviária, políticas públicas baseadas em evidências e de ciência de dados que serão abordados ao longo do presente trabalho.

2.1 Mobilidade Ativa Urbana

Mobilidade ativa é definida como utilizar caminhada e ciclismo como transporte exclusivamente ou em combinação com formas de transporte público. Além de modo de transporte, a mobilidade ativa também se mostra conveniente como forma de exercício físico e para o planejamento urbano, sendo relevante para a qualidade de vida nas cidades (GERIKE et al., 2016).

2.2 Sistema Cicloviário

O Sistema Cicloviário, ou Infraestrutura Cicloviária, é definido como o conjunto de infraestruturas necessárias para a circulação segura de ciclistas e das políticas de incentivo ao uso da bicicleta (Companhia de Engenharia de Tráfego - CET, 2016). Entre as estruturas em questão se encontram:

- Ciclovia: pista de uso exclusivo de bicicletas e outros ciclos, com segregação física do tráfego comum;
- Ciclofaixa: parte da pista de rolamento, calçada ou canteiro destinada à circulação exclusiva de ciclos, delimitada por sinalização específica;
- Ciclorrota: via com velocidade máxima reduzida, características de volume de tráfego baixo e com sinalização específica, indicando o compartilhamento do espaço viário entre veículos motorizados e bicicletas;
- Bicicletário: estacionamento de bicicletas em área pública ou privada com vigilância presencial ou eletrônica; e
- Paraciclo: suporte para a fixação de bicicletas em área pública ou privada.

2.3 Pesquisa Origem Destino

A OD, pesquisa realizada pela Companhia do Metropolitano de São Paulo, o Metrô, é uma pesquisa atualizada a cada 10 anos desde 1967, sendo essencial para o planejamento urbano e organização dos fluxos de viagens da capital paulista.

A edição mais recente, de 2017, utilizada como base deste trabalho, marcou o quinquagésimo aniversário da pesquisa, que contou com financiamento do Banco Mundial. É constituída de duas grandes partes, a Pesquisa Domiciliar, que analisa viagens realizadas dentro da Região Metropolitana de São Paulo (RMSP), e a Pesquisa na Linha de Contorno, que analisa as viagens que têm origem ou destino fora da RMSP ou que a atravessam (COMPANHIA DO METROPOLITANO DE SÃO PAULO - METRÔ, 2019).

De modo geral, a OD tem como objetivos:

- Quantificar e caracterizar o padrão das viagens na RMSP;
- Obter informações para projeções de viagens futuras; e
- Obter subsídios para planejamentos futuros.

2.4 Políticas Públicas Baseadas em Evidências

A implementação de PPBEs é uma filosofia que preza pelo uso de pesquisa, avaliação, análise e métodos científicos para auxiliar o processo de tomada de decisão (LUM; KOPER, 2014). Elas são parte importante na atuação de governos ao redor do mundo, como forma de justificar suas ações (COSTA; DA SILVA, 2016).

Os dados sobre a circulação de ciclistas e pedestres, nas mais diferentes localidades da cidade, se mostram de extrema valia para a implementação de PPBEs. Por meio delas, possibilita-se que os recursos sejam alocados de maneira mais eficiente, priorizando reforçar a infraestrutura em rotas mais populares e incentivando regiões menos participativas.

2.5 Ciência de Dados e Medidas Estatísticas

Ciência de Dados é um conjunto de princípios fundamentais que dão suporte e orientam a extração de conhecimento e informações relevantes a partir de dados (PROVOST; FAWCETT, 2013). Com a combinação de técnicas estatísticas, matemáticas e computacionais, pode-se analisar dados e obter *insights* valiosos para as mais diversas áreas de interesse. Desse modo, a ciência de dados se mostra a base deste trabalho, o qual busca analisar e aprofundar conhecimentos quanto a mobilidade urbana ativa, em especial o ciclismo, na cidade de São Paulo.

2.5.1 Intervalo Interquartil

O Intervalo Interquartil (IIQ) ou Amplitude Interquartil (AIQ) é uma técnica de avaliação da dispersão de dados distribuídos continuamente. Neste método, o conjunto de dados é ordenado de maneira crescente e dividido em quatro partes iguais (quartis) (VINUTHA; POORNIMA; SAGAR, 2018). Define-se o IIQ como a diferença entre o terceiro (Q3) e o primeiro (Q1) quartis (Equação 2.1).

$$IIQ = Q3 - Q1 \tag{2.1}$$

Encontrado o IIQ, pode-se definir os limites inferior e superior e, a partir destes limites, encontrar e tratar os valores discrepantes, também conhecidos como *outliers*.

2.5.2 Correlação de Pearson

O coeficiente de correlação de Pearson, representado por r, é definido como uma medida de associação linear entre variáveis. Na estatística, uma associação se dá quando variáveis guardam semelhanças na distribuição de seus escores. A correlação de Pearson exige um compartilhamento de variância, sendo essa distribuída linearmente (FIGUEIREDO FILHO; SILVA JÚNIOR, 2009).

A equação da correlação de Pearson se dá pela equação 2.2.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{2.2}$$

Onde:

- r representa o coeficiente de correlação de Pearson.
- x_i e y_i representam valores medidos para as variáveis x e y, respectivamente.
- $\bar{x}$ e $\bar{y}$ representam as médias das variáveis x e y, respectivamente
- n representa o número de valores medidos.

O coeficiente de Pearson varia de -1 a 1, e quanto maior a proximidade do coeficiente ao valores limite (independente do sinal), maior o grau de dependência linear entre as variáveis. Consequentemente, quanto mais próximo de 0, menor o grau de dependência. Já o sinal é o indicador da direção do relacionamento entre as variáveis.

2.5.3 Análise de Dados Espaciais

O foco da análise espacial é avaliar propriedades e relacionamentos, levando em conta o espaço em que o objeto de estudo está inserido, ou seja, incorporar o espaço à análise (CÂMARA et al., 2004).

No estudo sobre a mobilidade ativa de uma cidade, e consequentemente do seu espaço urbano, os dados geoespaciais se mostram de suma importância para enriquecer as análises. A visualização destes dados geralmente se dá utilizando Sistemas de Informação Geográfica (SIG) e por meio de mapas.

2.5.4 Formatos de Dados

2.5.4.1 Dados Tabulares - *Comma-Separated Values* (CSV)

Arquivos em formato CSV são utilizados para armazenar dados tabulares em formato de texto simples, de maneira rápida e eficiente. Neles, cada linha corresponde a um registro e os valores de cada coluna são separados por vírgulas.

Em Ciência de Dados, manipular arquivos CSV é uma tarefa recorrente, sendo facilitada por recursos como a linguagem Python e bibliotecas adequadas para tal, como pandas. Estes recursos permitem realizar operações como leitura, escrita, junções, intersecções, agrupamentos, filtragens, manipulação de colunas, entre outras.

2.5.4.2 Dados Geospaciais - *Shapefiles*

O formato *shapefile* é utilizado para armazenar dados geoespaciais vetoriais, voltados principalmente para SIGs. Desenvolvido e regulado pela *Environmental Systems Research Institute* (Esri), é um formato amplamente utilizado pela comunidade da área.

Um *shapefile* geralmente é associado à forma geométrica em si, mas na realidade é composto por um conjunto de arquivos. O arquivo *shape* (SHP) individualmente não é suficiente para representar um conjunto de dados. A seguir são descritos os arquivos que geralemente compõem um *shapefile*:

- *Shape* (SHP): armazena as formas geométricas reais (pontos, linhas, polígonos) dos dados espaciais. Presença obrigatória para utilização do *shapefile*;
- *Shape Index* (SHX): arquivo de indexação que permite acesso rápido às informações do arquivo SHP. Presença obrigatória para utilização do *shapefile*;
- *DataBase File* (DBF): armazena os atributos das formas de maneira tabular (como um banco de dados). Presença obrigatória para utilização do *shapefile*; e

- *Projection File* (PRJ): armazena as informações de projeção e do sistema de coordenadas utilizado.

3 Metodologia

O presente capítulo aborda a metodologia e as bases de dados utilizadas ao longo do trabalho.

3.1 Métodos de pesquisa

Para a realização deste trabalho, utilizou-se estatística descritiva em conjunto com técnicas de análise espacial e geoprocessamento.

Segundo Guedes et al. (2005), a estatística descritiva é uma das três áreas da Estatística, cujo objetivo básico, como próprio nome diz, é descrever uma série de dados de mesma natureza, possibilitando a organização e obtenção de um panorama global de variação destes valores.

Já a análise espacial, é constituída por "um conjunto de procedimentos encadeados cuja finalidade é a escolha de um modelo inferencial que considere explicitamente o relacionamento espacial presente no fenômeno" (CÂMARA et al., 2004).

3.2 Tecnologias e Ferramentas

Nesta seção serão introduzidas as ferramentas utilizadas durante o desenvolvimento do projeto.

3.2.1 Python

Python é uma linguagem de programação de alto nível, interpretada e orientada a objetos, utilizada nas mais diversas áreas de interesse, como Ciência de Dados, desenvolvimento web e Inteligência Artificial. Optou-se por utilizar Python por se tratar de uma linguagem intuitiva, versátil, de fácil utilização e alta adaptabilidade, com uma ampla gama de bibliotecas e *frameworks à disposição*, em especial para Ciência de Dados.

3.2.2 Jupyter Notebooks

Jupyter Notebook é uma plataforma web que fornece ao usuário um ambiente interativo, no qual é possível criar, editar e compartilhar *notebooks*. Um *notebook* é um documento que combina códigos executáveis, textos descritivos, dados e ricas opções de visualização (gráficos, tabelas, figuras, modelos 3D, *widgets*, entre outros).

É amplamente utilizado em projetos de Ciência de Dados por sua flexibilidade, leveza e fácil manipulação e recursos de visualização.

Foi escolhido como ambiente para desenvolvimento da ferramenta de análise, utilizando a linguagem Python (Seção 3.2.3)

3.2.3 Pandas

Pandas é uma biblioteca de código aberto disponível para Python, amplamente utilizada para análise e manipulação de dados, devido à sua simplicidade e intuitividade, somadas à estruturas de dados eficientes, fácil integração com outras bibliotecas e uma comunidade ativa e engajada.

3.2.4 GeoPandas

GeoPandas é uma biblioteca de código aberto disponível para Python, amplamente utilizada para análise e manipulação de dados geoespaciais, geralmente encontrados em formatos como *shapefile*, GeoJSON e KML. A biblioteca expande os tipos de dados utilizados pelo Pandas (Series e DataFrame) adicionando uma coluna referente à geometria, o que possibilita a realização de operações espaciais que exigiriam um SIG. Permite também a projeção dos dados e suas respectivas geometrias, a partir do sistema de coordenadas adotado.

3.2.5 Folium

Folium é uma biblioteca de código aberto disponível para Python, baseada no poder de mapeamento da biblioteca Leaflet.js, que possibilita a criação de mapas interativos e sua integração a ambientes web, como *notebooks* Jupyter. Os mapas podem ser combinados com dados geoespaciais, os quais são projetados sobre os recursos cartográficos como camadas, enriquecendo ainda mais as análises e visualizações.

3.2.6 Matplotlib

Matplotlib é uma biblioteca de código aberto disponível para Python, utilizada para criar gráficos e outras formas de visualização de dados. A biblioteca permite a criação de visualizações estáticas, animadas e interativas e a personalização de componentes, por meio de cores, rótulos, estilos, entre outros. Possui fácil integração com ambientes como Jupyter e outras bibliotecas como pandas.

3.2.7 ipywidgets

Ipywidgets é uma biblioteca de código aberto disponível para Python, utilizada para adicionar *widgets* interativos ao ambiente Jupyter. Eles permitem com que os usuários modifiquem variáveis e visualizem dados de maneira dinâmica, sem a necessidade de executar o

código novamente. Entre esses *widgets* se encontram *sliders*, *dropdowns*, *radio buttons*, botões, barras de progresso, caixas de texto, entre outros.

3.3 Conjuntos de Dados

Nesta seção são apresentados os conjuntos de dados utilizados, suas características e forma de obtenção.

3.3.1 Dados de Viagens de Bicicleta

Os dados foram obtidos do aplicativo de viagens de bicicleta, que contém vastos conjuntos de dados sobre mobilidade ativa, especialmente ciclistas e pedestres. Conforme os usuários utilizam o aplicativo para acompanhar suas atividades (treinos, viagens, deslocamentos), o aplicativo agrega e contextualiza os dados obtidos gerando o conjunto em questão. Para obtenção dos dados, deve-se buscar uma parceria com a referida empresa.

Para manter a privacidade dos usuários, o aplicativo remove qualquer tipo de identificador ao agregar os dados, além de desconsiderar atividades ligadas a perfis privados. Ainda como medida de anonimização, as informações disponíveis são agrupadas em múltiplos de cinco. Os arredondamentos realizados levaram à pequenas perdas de dados, em situações onde o agrupamento em cinco não foi possível.

3.3.1.1 Viagens por Arestas

Dados horários referentes à viagens de ciclistas. Uma viagem é definida como a travessia de uma determinada aresta pelos ciclistas em suas atividades, sendo que uma atividade pode conter múltiplas viagens por uma mesma aresta. Já uma aresta é um segmento de uma rua (ou avenida, rodovia, entre outros).

Os dados são disponibilizados de forma separada por mês, de janeiro de 2019 até agosto de 2024. Para cada um deles, estão disponíveis uma série de arquivos relacionados à geolocalização das arestas (formato *shapefile*) e um arquivo em formato CSV, onde cada linha é vinculada a uma aresta e a um determinado intervalo. As colunas de viagens estão separadas nos sentidos de ida e volta das travessias definidas pelo OpenStreetMap (padrão e reverso, respectivamente), que são indiferentes para o presente trabalho. Os atributos disponíveis podem ser vistos a seguir:

- Identificador da aresta;
- Tipo de atividade (ciclistas, pedestres);
- Data e período (formato AAAA-MM-DD-HH);

- Total de viagens;
- Total de viagens por sentido;
- Total de pessoas únicas;
- Total de viagens utilitárias;
- Total de viagens de lazer;
- Total de pessoas por gênero;
- Total de pessoas faixa etária;
- Velocidade média em metros por segundo;
- Total de viagens com bicicletas tradicionais; e
- Total de viagens com bicicletas elétricas.

3.3.1.2 Origens e Destinos

Dados referentes às origens e destinos das atividades dos ciclistas, sendo que uma atividade só é considerada quando definida como completa pelo usuário. As áreas de origens e destinos são formatadas como hexágonos.

Os dados foram solicitados separados mês a mês, de janeiro de 2019 até agosto de 2024. Para cada um deles, estão disponíveis uma série de arquivos relacionados à geolocalização dos hexágonos (formato *shapefile*) e dois arquivos em formato CSV, um contendo as origens e outro contendo os destinos. Os atributos disponíveis podem ser vistos a seguir:

- Identificador do hexágono;
- Mês e ano correspondente (formato AAAA-MM);
- Tipo (origem ou destino);
- Tipo de atividade (ciclistas, pedestres);
- Total de atividades se iniciando ou terminando no hexágono;
- Total de atividades utilitárias se iniciando ou terminando no hexágono;
- Total de atividades de lazer se iniciando ou terminando no hexágono;
- Total de atividades por faixa horária se iniciando ou terminando no hexágono;
- Total de atividades em dias úteis se iniciando ou terminando no hexágono; e
- Total de atividades em fins de semana se iniciando ou terminando no hexágono.

3.3.2 Dados Populacionais de São Paulo

Os dados populacionais da cidade de São Paulo foram obtidos através do repositório da Fundação Sistema Estadual de Análise de Dados Estatísticos (SEADE), órgão da Secretaria de Planejamento e Gestão do Governo do Estado de São Paulo.

O arquivo está disponível em formato CSV e contém a população de cada distrito da capital paulista em 2021 (Figura 1). Os dados possuem os campos:

- CodIBGE (int64): Código definido pelo IBGE para distrito;
- Municípios (object): Nome do Distrito;
- Ano (int64): Ano de referência;
- População (int64): População segundo município de residência e ano de referência;
- Porte Populacional (object): Classes definidas conforme o volume populacional; e
- Codigo_Porte (int64): Código do porte populacional.

Figura 1 – Amostra do *dataframe* relativo à população por distrito

	CodIBGE	Municípios	Ano	População	Porte Populacional	Codigo_Porte
0	1	Água Rasa	2021	82264	De 50 a 100 mil	2
1	2	Alto de Pinheiros	2021	40708	Até 50 mil	1
2	3	Anhanguera	2021	86020	De 50 a 100 mil	2
3	4	Aricanduva	2021	85747	De 50 a 100 mil	2
4	5	Artur Alvim	2021	100040	De 100 a 200 mil	3
...	...	...	...	...	...	...
92	93	Vila Prudente	2021	104616	De 100 a 200 mil	2
93	94	Vila Sônia	2021	122464	De 100 a 200 mil	3
94	95	São Domingos	2021	86504	De 50 a 100 mil	3
95	96	Lajeado	2021	175632	De 100 a 200 mil	3

Fonte: Elaborada pelo autor.

3.3.3 Dados da Infraestrutura Cicloviária

Os dados geoespaciais sobre a infraestrutura cicloviária da cidade são fornecidos pela Companhia de Engenharia de Tráfego (CET), vinculada à Secretaria Municipal de Mobilidade da prefeitura de São Paulo, responsável pelo gerenciamento, operação e fiscalização do sistema viário da cidade.

No *site* da CET[1] estão disponíveis para *download* os arquivos da infraestrutura cicloviária,

[1] Disponível em: <https://www.cetsp.com.br/consultas/bicicleta/mapa-de-infraestrutura-cicloviaria.aspx>>

em formato *shapefile*. Neles estão contidos dados geoespaciais sobre infraestrturas como ciclovias, ciclofaixas, ciclorrotas, conexões, bicicletários e paraciclos espalhados pela cidade de São Paulo.

Os arquivos relativos às ciclovias e ciclofaixas (Figura 2) e ciclorrotas (Figura 3), nos quais cada linha corresponde a um trecho das mesmas, possuem as colunas:

- programa (object): Programa da ciclovia/ciclofaixa/ciclorrota;
- inaguracao (datetime64[ms]): Inauguração do trecho;
- extensao_t (int64): Extensão do trecho em quilômetros;
- extensao_c (int64): Extensão total da ciclovia/ciclofaixa/ciclorrota em quilômetros; e
- geometry (geometry): Geometria do trecho em formato de linha (LINESTRING).

Figura 2 – Amostra do *dataframe* relativo às ciclovias e ciclofaixas

	programa	inauguracao	extensao_t	extensao_c	geometry
0	CICLOFAIXA AFONSO LOPES VIEIRA	2016-07-15	791	1197	LINESTRING (-46.66744 -23.46776, -46.66739 -23...
1	CICLOFAIXA AFONSO LOPES VIEIRA	2016-07-15	716	1197	LINESTRING (-46.66744 -23.46776, -46.66744 -23...
2	CICLOFAIXA AFONSO LOPES VIEIRA	2016-07-15	31	1197	LINESTRING (-46.6676 -23.46799, -46.66744 -23....
3	CICLOFAIXA AFONSO LOPES VIEIRA	2016-07-15	5	1197	LINESTRING (-46.66291 -23.46304, -46.66287 -23...
4	CICLOFAIXA AFONSO LOPES VIEIRA	2016-07-15	219	1197	LINESTRING (-46.67001 -23.46803, -46.66999 -23...
...	...	...	...	...	...
1961	CICLOFAIXA JAIR RIBEIRO	2016-06-23	16	776	LINESTRING (-46.68352 -23.69508, -46.68359 -23...
1962	CICLOFAIXA JAIR RIBEIRO	2016-06-23	869	776	LINESTRING (-46.68553 -23.70303, -46.68549 -23...
1963	CICLOFAIXA JAIR RIBEIRO	2016-06-23	895	776	LINESTRING (-46.68553 -23.70303, -46.68551 -23...
1964	CICLOFAIXA JAIR RIBEIRO	2016-06-23	252	776	LINESTRING (-46.68373 -23.69461, -46.68373 -23...
1965	CICLOFAIXA JAIR RIBEIRO	2016-06-23	41	776	LINESTRING (-46.68352 -23.69508, -46.68358 -23...

1966 rows × 5 columns

Fonte: Elaborada pelo autor.

Já os arquivos relativos aos bicicletários e paraciclos (Figura 4), nos quais cada linha corresponde a um equipamento, possuem as colunas:

- LOCAL (object): Local de instalação do equipamento;
- EQUIPAMENT (object): Tipo de equipamento, bicicletário ou paraciclo;
- VAGAS (int64): Número de vagas do equipamento;
- RESPONS (object): Responsável pelo equipamento; e
- geometry (geometry): Geometria do trecho em formato de ponto (POINT).

Figura 3 – Amostra do *dataframe* relativo às ciclorrotas

	programa	inauguracao	extensao_t	extensao_c	geometry
0	CICLORROTA BATURITÉ/ DIAMANTE	2021-05-18	83	548	LINESTRING (-46.63046 -23.5715, -46.63048 -23....
1	CICLORROTA BATURITÉ/ DIAMANTE	2021-05-18	465	548	LINESTRING (-46.63046 -23.5715, -46.63044 -23....
2	CICLORROTA BROOKLIN	2011-07-20	911	5950	LINESTRING (-46.69809 -23.63173, -46.69858 -23...
3	CICLORROTA BROOKLIN	2011-07-20	279	5950	LINESTRING (-46.70033 -23.6334, -46.69984 -23....
4	CICLORROTA BROOKLIN	2011-07-20	227	5950	LINESTRING (-46.69886 -23.63128, -46.69809 -23...
...	...	...	...	...	...
112	CICLORROTA VILA MARIANA	2012-05-24	540	4953	LINESTRING (-46.63956 -23.59337, -46.64018 -23...
113	CICLORROTA VILA MARIANA	2012-05-24	170	4953	LINESTRING (-46.64483 -23.59293, -46.64617 -23...
114	CICLORROTA VILA MARIANA	2012-05-24	121	4953	LINESTRING (-46.64438 -23.57932, -46.64397 -23...
115	CICLORROTA VILA MARIANA	2012-05-24	103	4953	LINESTRING (-46.64243 -23.57866, -46.64204 -23...
116	CICLORROTA VILA MARIANA	2012-05-24	389	4953	LINESTRING (-46.64204 -23.57952, -46.64165 -23...

117 rows × 5 columns

Fonte: Elaborada pelo autor.

Figura 4 – Amostra do *dataframe* relativo aos bicicletários e paraciclos

	LOCAL	EQUIPAMENT	VAGAS	RESPONS	geometry
0	ESTACAO TUCURUVI	PARACICLO	8	METRO	POINT (-46.60347 -23.48014)
1	ESTACAO PARADA INGLESA	PARACICLO	8	METRO	POINT (-46.60891 -23.48715)
2	ESTACAO JARDIM SAO PAULO	PARACICLO	16	METRO	POINT (-46.61699 -23.49273)
3	ESTACAO JAPAO - LIBERDADE	PARACICLO	8	METRO	POINT (-46.63554 -23.55504)
4	ESTACAO SANTA CRUZ	PARACICLO	8	METRO	POINT (-46.6368 -23.59862)
...	...	...	...	...	...
116	TERMINAL JARDIM BRITANIA	BICICLETARIO	12	SPTRANS	POINT (-46.78723 -23.43156)
117	ESTACAO DE TRANSFERENCIA AGUA ESPRA	BICICLETARIO	84	SPTRANS	POINT (-46.6959 -23.6135)
118	ESTACAO DE TRANSFERENCIA AGUA ESPRA	PARACICLO	4	SPTRANS	POINT (-46.6959 -23.6135)
119	ESTAÇÃO DE TRANSFERENCIA ITAQUERA	BICICLETARIO	20	SPTRANS	POINT (-46.45344 -23.5348)
120	LARGO DA BATATA	BICICLETARIO	100	TEMBICI	POINT (-46.69382 -23.56718)

121 rows × 5 columns

Fonte: Elaborada pelo autor.

3.3.4 Limites Administrativos

Para enriquecer e especificar as análises, decidiu-se por possibilitar a visualização e filtragem do dados por diferentes divisões administrativas da cidade de São Paulo. Foram escolhidos três níveis de granularidade: zonas (divididas pela pesquisa OD2017), distritos e subprefeituras.

A pesquisa OD dividiu a Região Metropolitana de São Paulo e o município de São Paulo em 517 e 342 zonas de pesquisa, respectivamente. Chamadas de Zonas OD, cada uma delas é "a menor unidade geográfica a partir da qual está assegurada a representatividade estatística dos dados". Os arquivos *shapefile* correspondentes aos limites administrativos a essas zonas, assim como a síntese da pesquisa, estão disponíveis no Portal da Transparência

do Metrô de São Paulo[2].

O arquivo relativo às Zonas OD (Figura 5) possui os atributos:

- NumeroZona (int64): Identificador numérico da zona;
- NomeZona (object): Nome da zona;
- NumeroMuni (int64): Identificador numérico do município a qual a zona pertence;
- NomeMunici (object): Nome do município a qual a zona pertence;
- NumDistrit (int64): Identificador numérico do distrito a qual a zona pertence;
- NomeDistri (object): Nome do distrito a qual a zona pertence;
- Area_ha_2 (float64): Área da zona, em hectares quadrados; e
- geometry (geometry): Geometria da zona em formato de polígono (POLYGON).

Figura 5 – Amostra do *dataframe* relativo às Zonas OD

	NumeroZona	NomeZona	NumeroMuni	NomeMunici	NumDistrit	NomeDistri	Area_ha_2	geometry
0	1	Sé	36	São Paulo	80	Sé	57.10	POLYGON Z ((333739.415 7394619.838 0, 333792.4...
1	2	Parque Dom Pedro	36	São Paulo	80	Sé	113.64	POLYGON Z ((333106.146 7395425.48 0, 333120.09...
2	3	Praça João Mendes	36	São Paulo	80	Sé	47.75	POLYGON Z ((333353.211 7393933.156 0, 333238.0...
3	4	Ladeira da Memória	36	São Paulo	67	República	75.11	POLYGON Z ((332742.619 7394795.328 0, 332742.5...
4	5	República	36	São Paulo	67	República	74.95	POLYGON Z ((332983.962 7395262.578 0, 333004.2...
...	...	...	...	...	...	...	...	...
512	513	Quatro Encruzilhadas	17	Itapevi	113	Itapevi	2101.43	POLYGON Z ((300336.368 7389696.621 0, 300324.3...
513	514	Itapevi	17	Itapevi	113	Itapevi	1070.03	POLYGON Z ((303326.146 7397176.35 0, 303326.14...
514	515	Amador Bueno	17	Itapevi	113	Itapevi	5113.23	POLYGON Z ((303292.377 7397174.79 0, 303326.14...
515	516	Santana de Parnaíba	31	Santana de Parnaíba	127	Santana de Parnaíba	18034.76	POLYGON Z ((296271.998 7398946.927 0, 296196.3...
516	517	Pirapora do Bom Jesus	25	Pirapora do Bom Jesus	121	Pirapora do Bom Jesus	10876.89	POLYGON Z ((291856.851 7406803.921 0, 291853.6...

Fonte: Elaborada pelo autor.

Já os arquivos dos limites administrativos referentes aos distritos e subprefeituras também foram obtidos em formato *shapefile*, mas através da plataforma GeoSampa[3], mantida pela Secretaria Municipal de Urbanismo e Licenciamento (SMUL) da cidade de São Paulo. São considerados 96 distritos e 32 subprefeituras na cidade.

O arquivo relativo aos distritos (Figura 6), contém os atributos:

[2] Disponível em: <https://transparencia.metrosp.com.br/dataset/pesquisa-origem-e-destino>
[3] Disponível em: <https://geosampa.prefeitura.sp.gov.br/PaginasPublicas/_SBC.aspx>

- ds_nome (object): Nome do distrito;
- ds_codigo (object): Código numérico do distrito;
- ds_cd_sub (object): Código numérico da subprefeitura a qual o distrito pertence;
- ds_subpref (object): Nome da subprefeitura a qual o distrito pertence;
- ds_sigla (object): Sigla do distrito;
- ds_areamt (float64): Área do distrito, em metros quadrados;
- ds_areakm (float64): Área do distrito, em quilômetros quadrados; e
- geometry (geometry): Geometria do distrito em formato de polígono (POLYGON).

Figura 6 – Amostra do *dataframe* relativo aos Distritos

	ds_nome	ds_codigo	ds_cd_sub	ds_subpref	ds_sigla	ds_areamt	ds_areakm	geometry
0	MANDAQUI	51	05	SANTANA-TUCURUVI	MAN	13249456.11	13.249	POLYGON ((333079.583 7408102.398, 333077.161 7...
1	MOEMA	32	12	VILA MARIANA	MOE	9079516.47	9.080	POLYGON ((331290.13 7392111.284, 331324.217 73...
2	ARTUR ALVIM	5	21	PENHA	AAL	6505750.23	6.506	POLYGON ((349420.638 7397694.618, 349423.468 7...
3	IGUATEMI	33	30	SAO MATEUS	IGU	19434636.54	19.435	POLYGON ((350874.784 7389641.837, 350875.63 73...
4	ITAIM BIBI	35	11	PINHEIROS	IBI	10026327.68	10.026	POLYGON ((327871.085 7386152.364, 327865.295 7...
...	...	...	...	...	...	...	...	...
91	BUTANTA	12	10	BUTANTA	BUT	12952520.76	12.953	POLYGON ((322140.524 7393133.295, 322137.103 7...
92	CACHOEIRINHA	13	04	CASA VERDE-CACHOEIRINHA	CAC	13571736.73	13.572	POLYGON ((330933.585 7402181.544, 330910.867 7...
93	CAMBUCI	14	09	SE	CMB	3924676.72	3.925	POLYGON ((334806.666 7391501.308, 334783.054 7...
94	CAMPO GRANDE	16	14	SANTO AMARO	CGR	12936654.38	12.937	POLYGON ((325612.963 7381815.938, 325634.233 7...
95	CAMPO LIMPO	17	17	CAMPO LIMPO	CLM	12594802.91	12.595	POLYGON ((321408.128 7387296.041, 321422.038 7...

Fonte: Elaborada pelo autor.

O arquivo relativo às subprefeituras (Figura 7), contém os atributos:

- sp_nome (object): Nome da subprefeitura;
- sp_codigo (object): Código numérico da subprefeitura;
- sp_id (float64): Identificador numérico da subprefeitura;
- sp_sigla (object): Sigla da subprefeitura;
- sp_areamt (float64): Área da subprefeitura, em metros quadrados;
- sp_areakmt (float64): Área da subprefeitura, em quilômetros quadrados; e
- geometry (geometry): Geometria da subprefeitura em formato de polígono (POLYGON).

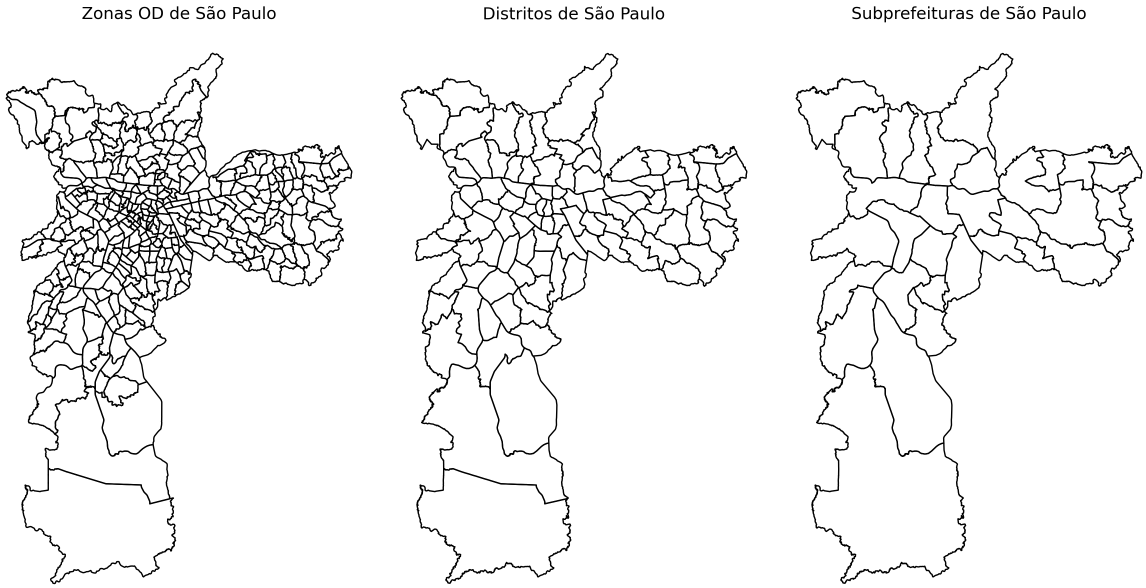
Figura 7 – Amostra do *dataframe* relativo às Subprefeituras

	sp_nome	sp_codigo	sp_id	sp_sigla	sp_areamt	sp_areakmt	geometry
0	FREGUESIA-BRASILANDIA	03	2.0	FO	3.198020e+07	32.0	POLYGON ((327340.628 7399133.313, 327331.514 7...
1	CASA VERDE-CACHOEIRINHA	04	3.0	CV	2.723234e+07	27.0	POLYGON ((329084.795 7402363.669, 329086.123 7...
2	LAPA	08	13.0	LA	4.063870e+07	41.0	POLYGON ((321633.729 7393535.365, 321633.031 7...
3	PERUS	01	31.0	PR	5.688931e+07	57.0	POLYGON ((325464.932 7409329.344, 325464.513 7...
4	SE	09	14.0	SE	2.666547e+07	27.0	POLYGON ((330197.017 7396087.885, 330211.849 7...
5	SANTANA-TUCURUVI	05	4.0	ST	3.578252e+07	36.0	POLYGON ((334076.366 7398045.594, 334074.986 7...
6	PINHEIROS	11	22.0	PI	3.199353e+07	32.0	POLYGON ((327871.085 7386152.364, 327865.295 7...
7	JACANA-TREMEMBE	06	5.0	JT	6.511566e+07	65.0	POLYGON ((335167.648 7404409.048, 335167.247 7...
8	VILA MARIA-VILA GUILHERME	07	6.0	MG	2.689922e+07	27.0	POLYGON ((336762.078 7401144.267, 336794.867 7...
9	MOOCA	25	15.0	MO	3.604691e+07	36.0	POLYGON ((334126.109 7396389.068, 334125.831 7...
10	IPIRANGA	13	19.0	IP	3.749732e+07	37.0	POLYGON ((335254.612 7391150.376, 335253.982 7...
11	VILA MARIANA	12	20.0	VM	2.698441e+07	27.0	POLYGON ((332031.435 7385561.331, 332030.94 73...
12	PENHA	21	7.0	PE	4.044975e+07	40.0	POLYGON ((346801.065 7402842.892, 346800.103 7...
13	ERMELINO MATARAZZO	22	8.0	EM	1.594588e+07	16.0	POLYGON ((349116.316 7399473.221, 349133.905 7...
14	SAO MIGUEL	23	9.0	MP	2.615923e+07	26.0	POLYGON ((352480.323 7397515.871, 352477.052 7...

Fonte: Elaborada pelo autor.

A figura 8 demonstra a cidade de São Paulo dividida nos diferentes níveis de granularidade (zonas OD, distritos e subprefeituras, respectivamente).

Figura 8 – Divisões Administrativas de São Paulo



Fonte: Elaborada pelo autor.

3.4 Tratamento de dados

Após a obtenção dos conjuntos de dados, se inicia o processo de tratamento dos mesmos. Tratar os dados garante com que sejam mais consistentes e consequentemente mais confiáveis para apoiar decisões.

3.4.1 Identificação e Tratamento de Valores Discrepantes

Após uma observação dos dados de viagens de bicicleta, notou-se a presença de alguns valores discrepantes nos campos correspondentes a velocidade média em algumas arestas.

Para tratá-los utilizou a técnica de IIQ, descrita na Subseção 2.5.1. Iniciou-se dividindo os dados em quartis, calculando o IIQ e definindo os limites inferior e superior (Figura 9).

Figura 9 – Cálculo do IIQ e dos limites

```
# Cálculo do Q1 e Q3
q1 = pd.Series(all_non_zero_speeds).quantile(0.25)
q3 = pd.Series(all_non_zero_speeds).quantile(0.75)
iiq = q3 - q1

# Cálculo dos limites inferior e superior
lower_bound = q1 - 1.5 * iiq
upper_bound = q3 + 1.5 * iiq
```

Fonte: Elaborada pelo autor.

Depois, filtrou-se os dados válidos (dados não nulos e não discrepantes) e calculou-se a velocidade média dos mesmos (Figura 10).

Figura 10 – Filtro de dados válidos

```
# Filtragem de dados válidos/outliers e média dos dados válidos
valid_data = [value for value in all_non_zero_speeds if lower_bound <= value <= upper_bound]
average_valid_data = sum(valid_data) / len(valid_data) if valid_data else 0
```

Fonte: Elaborada pelo autor.

Por fim, chegou-se aos resultados observados na Figura 11. A partir deste números, manipulou-se os arquivos originais para tratar os *outliers*. A solução escolhida foi substituí-los pela geral das amostras.

3.4.2 Manipulação e Remoção de Colunas Desnecessárias

Após a verificação de quais informações são realmente relevantes para o estudo, remover e manipular colunas desnecessárias é uma tarefa válida para polir e simplificar conjuntos de dados.

Figura 11 – Resultados obtidos aplicando o Intervalo Interquartil

```
IIQ: 2.915
Limite Inferior: -0.1125
Limite Superior: 11.5475
% de Outliers: 1.5237%
Média dos valores válidos: 5.580371021315769
```

Fonte: Elaborada pelo autor.

No caso dos dados de viagens de bicicleta, os tempos de leitura tornaram-se uma preocupação, já que cada mês de dados utiliza mais de 200 *megabytes* (mb) de armazenamento. Com isso em mente, também observou-se a já comentada presença de uma divisão das viagens nos dois sentidos definidos pelo OSM (*forward* e *reverse*). Neste trabalho, essa divisão se mostra irrelevante, já que a simples identificação de uma viagem na aresta já é o suficiente para as análises. Desse modo, optou-se por somar as colunas *forward* e *reverse*.

Como exceções, se destacam as colunas com a velocidade média nos dois sentidos, já que viu-se necessário o cálculo de uma média ponderada quanto as velocidades médias, além de convertê-las de metros por segundo (m/s) para quilômetros por hora (km/h). Por fim, foram removidas as colunas que não seriam mais utilizadas ao longo do trabalho.

Ao fim dessa etapa, realizou-se uma análise para verificar o ganho de armazenamento em *megabytes* e de tempo de leitura em segundos (s) com a remoção dessas colunas, utilizando uma amostra de dez arquivos mensais CSV selecionados aleatoriamente. Foi observado uma redução de cerca de 43% no tamanho e de cerca de 60% nos tempos de leitura dos arquivos (Figuras 12 e 13, respectivamente). Para os tempos de leitura, foram executadas dez iterações e calculadas as médias.

Figura 12 – Antes e Depois - Tamanho dos arquivos

Tamanho dos arquivos
(em megabytes)

	Antes	Depois	Redução (%)
0	268.91	153.28	43.00
1	203.74	116.14	42.99
2	241.19	137.45	43.01
3	240.38	136.92	43.04
4	257.93	146.99	43.01
5	247.92	141.21	43.04
6	227.22	129.32	43.08
7	224.62	127.86	43.08
8	221.54	126.05	43.10
9	267.82	152.47	43.07

Fonte: Elaborada pelo autor.

Figura 13 – Antes e Depois - Tempo de leitura dos arquivos

Tempo de leitura dos arquivos
(em segundos)

	Antes	Depois	Redução (%)
0	5.22	2.09	59.97
1	3.97	1.49	62.48
2	4.31	1.75	59.35
3	4.21	1.72	59.10
4	4.58	1.86	59.38
5	4.42	1.80	59.16
6	3.98	1.65	58.44
7	4.13	1.59	61.45
8	3.87	1.59	58.85
9	4.94	1.98	59.98

Fonte: Elaborada pelo autor.

3.4.3 Aplicação de Filtros

Previamente ao início das análises, também se viu necessário filtrar as Zonas OD que pertenciam ao município de São Paulo, já que os dados do aplicativo não abrangem toda a RMSP.

No segmento de código a seguir (Figura 14), realiza-se a seleção apenas das Zonas OD cujo campo 'NumeroMuni' é igual a 36 (código associado ao município de São Paulo).

Figura 14 – Filtro das Zonas OD do município de São Paulo

```
zonas_od = gpd.read_file('OD-2017/Mapas-OD2017/Shape-OD2017/Zonas_2017_region.shp')
zonas_sp = zonas_od[zonas_od['NumeroMuni'] == 36]
```

Fonte: Elaborada pelo autor.

3.5 Desenvolvimento das Análises

As análises descritas neste capítulo estão disponíveis em repositório no github[4], permitindo acesso de outras pessoas interessadas por assuntos na área.

3.5.1 Análises das Características das Viagens

Para obter um panorama geral sobre as características das viagens realizadas pelos ciclistas, realizou-se análises utilizando os atributos disponíveis, como gênero, idade, período do dia, tipo de viagem, entre outros.

[4] <https://github.com/LucaMunekata/tcc->

No conjunto de dados que divide as viagens por arestas, desenvolveu-se gráficos de linha para demonstrar a variação da quantidade de viagens e pessoas únicas (de acordo com uma determinada característica) ao longo de um período escolhido pelo usuário. Também foram desenvolvidos gráficos de setores para demonstrar as proporções encontradas para cada característica. Foi utilizado um subconjunto dos *dataframes* originais, agregando todos os dados mensais em uma única linha, ao invés da divisão de hora em hora. Já para o conjunto de dados que abrange as origens e destinos das viagens, os mesmos tipos de gráficos foram utilizados mas desta vez para analisar o período do dia em que as atividades ocorrem.

Os gráficos foram desenvolvidos utilizando a biblioteca *matplotlib* e seus recursos para proporcionar interatividade aos usuários.

3.5.2 Análises Geoespaciais

Nestas análises, busca-se aprofundar o estudo dos dados, adicionando a localização geográfica como um novo elemento de interesse. Utilizando de técnicas de mapeamento, é possível buscar *insights* específicos e particularidades de cada região de interesse e geolocalizar padrões encontrados previamente.

A partir do conjunto dados de viagens de bicicleta, foi possível desenvolver mapas coropléticos que destacam as arestas com mais viagens dentro de cada região administrativa. Por meio de junções espaciais (função *sjoin()* do *geopandas*) entre os *shapefiles* das Zonas/Distritos/Subprefeituras de São Paulo com as arestas (Figura 15), pode-se definir a qual subdivisão administrativa pertencem cada uma delas e utilizar os identificadores das mesmas para agrupar os dados relacionados (em caso de uma aresta pertencer a duas subdivisões diferentes, considerou-a para os cálculos de ambas). Desse modo, introduziu-se ao usuário a possibilidade de filtrar a visualização dos dados por região administrativa. A filtragem por gênero ou idade também é um recurso disponível.

Outra possível análise, foi o mapeamento da quantidade total de viagens por zona OD/distrito/subprefeitura. O mesmo se aplica para os arquivos relacionados às origens e destinos (hexágonos), mas com filtros quanto ao tipo e horários das viagens.

Figura 15 – Realização de uma junção espacial entre arestas e limites das Zonas OD

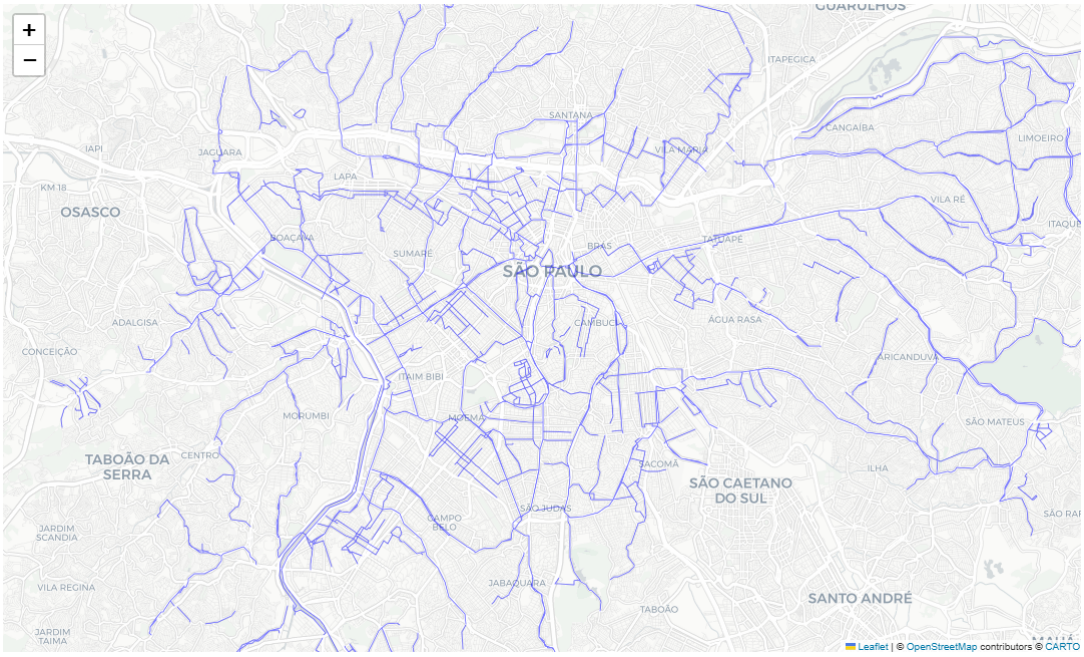
```
edges_in_zones = gpd.sjoin(edges, zonas_sp, how="inner", predicate="intersects")
df_merged_zones = pd.merge(df, edges_in_zones[['edgeUID', 'NumeroZona', 'NomeZona']], left_on='edge_uid', right_on='edgeUID', how='inner')
```

Fonte: Elaborada pelo autor.

Também realizou-se análises quanto a relação entre a estrutura cicloviária da cidade (Figura 16) e as viagens de bicicleta. Foram definidos *buffers* (uma geometria que considera todos os pontos dentro de uma certa distância de um outro objeto geométrico) de 100 metros (Figura 17), para abranger as arestas que se encontram dentro da malha cicloviária. A partir dos *buffers* pode-se obter informações relevantes como a proporção de viagens dentro e fora da

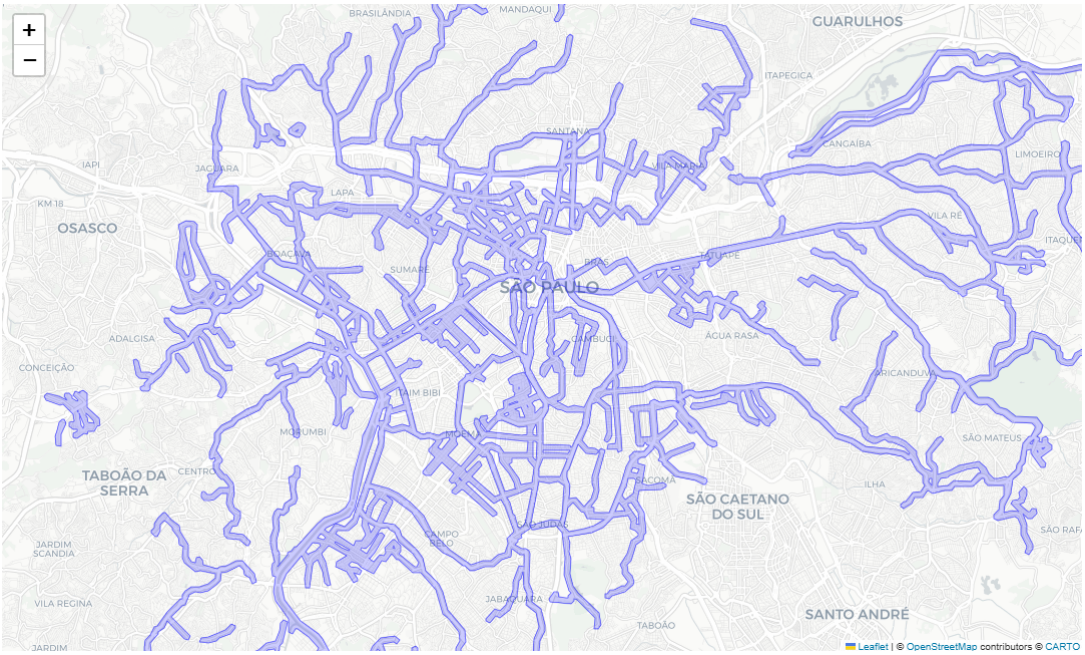
infraestrutura cicloviária, a relação entre velocidade e localidade das viagens, além de possíveis lacunas na malha.

Figura 16 – Recorte da estrutura cicloviária de São Paulo



Fonte: Elaborada pelo autor.

Figura 17 – Recorte do *buffer* da estrutura cicloviária de São Paulo



Fonte: Elaborada pelo autor.

4 Resultados

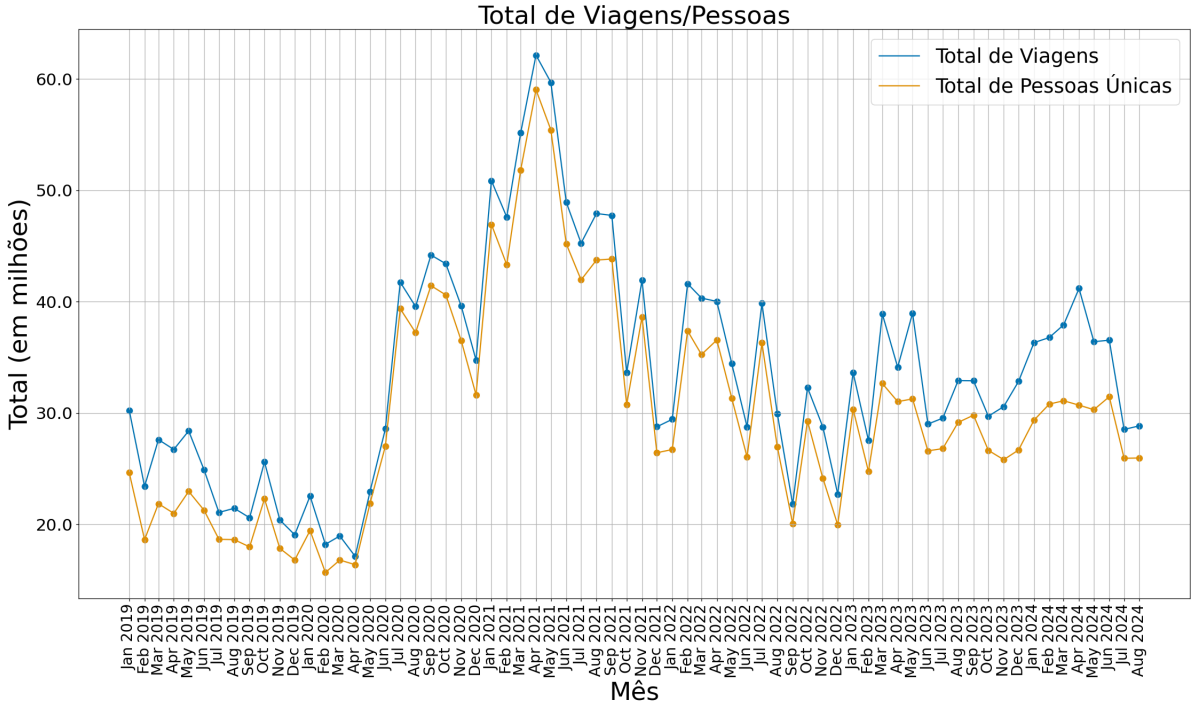
No presente capítulo, serão apresentados os resultados obtidos através das análises descritas anteriormente.

4.1 Visão Geral das Viagens de Bicicleta

Após o desenvolvimento das análises e agregação dos dados de todo o período de estudo, pôde-se identificar características relevantes sobre as viagens de bicicleta em São Paulo.

Como pode ser observado no gráfico de série temporal exposto na Figura 18, há o início de um aumento acentuado no número de viagens entre abril e maio de 2020, curiosamente, os primeiros meses do período de isolamento devido à pandemia da Covid-19. O número de pessoas únicas segue tendências similares ao total de viagens.

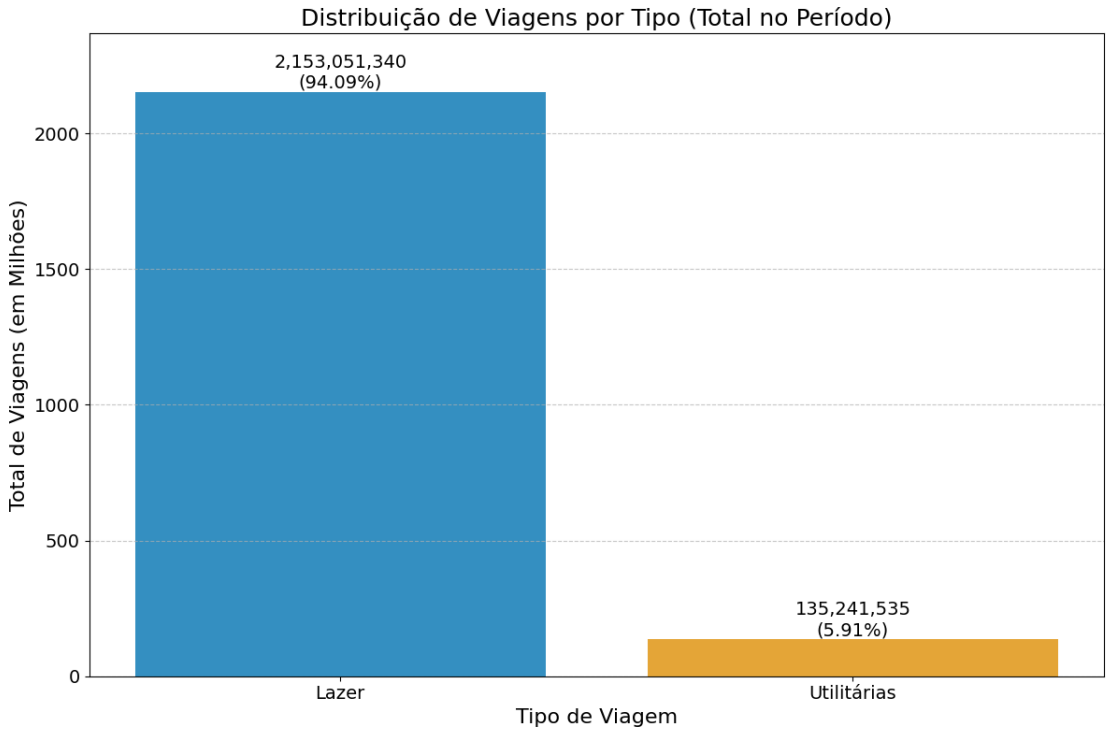
Figura 18 – Total de viagens e pessoas únicas ao longo do período



Fonte: Elaborada pelo autor.

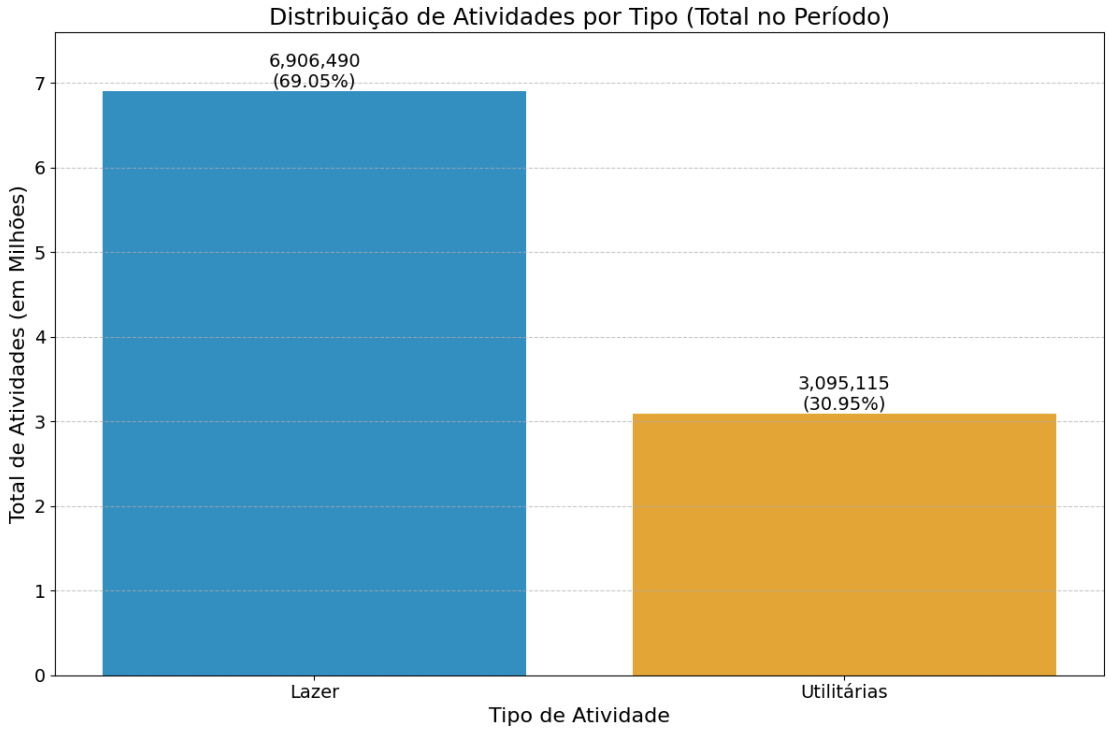
Quanto aos tipos das viagens, nota-se um amplo predomínio das viagens/atividades de lazer sobre as utilitárias, como pode-se observar nos gráficos de setores das Figuras 19 e 20 a seguir.

Figura 19 – Distribuição de viagens por tipo



Fonte: Elaborada pelo autor.

Figura 20 – Distribuição de atividades por tipo



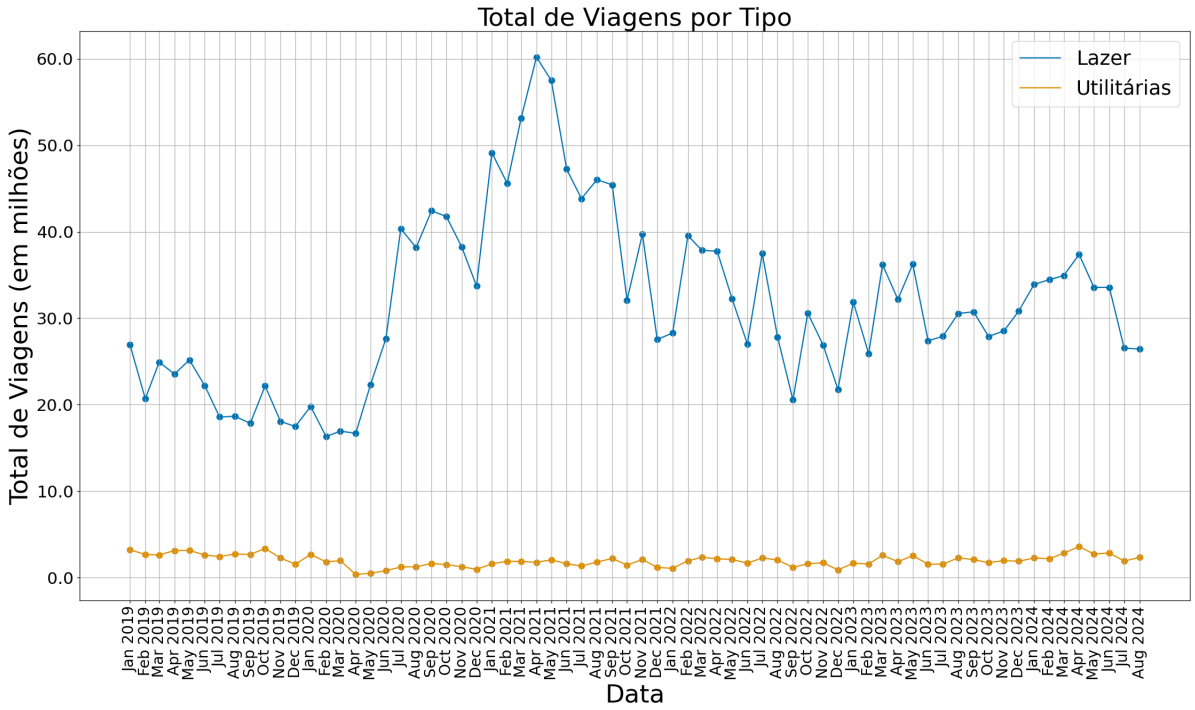
Fonte: Elaborada pelo autor.

Entretanto, nota-se diferenças consideráveis entre a proporção recreativas/utilitárias, quando comparadas viagens e atividades. Essa diferença pode ser explicada por meio da definição de viagem e atividade já apresentada na Subseção 3.3.1.1. O aplicativo considera uma viagem como a travessia de uma determinada aresta por usuários em suas atividades, sendo que uma atividade pode conter múltiplas viagens por uma mesma aresta. É razoável afirmar que atividades de lazer podem conter rotas cíclicas, explicando o aumento do número de viagens em detrimento de atividades.

Verificando agora as séries temporais contida nas Figuras 21 e 22, em conjunto com as matrizes de correlação de Pearson (conforme apresentada na Seção 2.5.2), contida nas Figuras 23 e 24, nota-se comportamentos diferentes para viagens/atividades recreativas e utilitárias, ao longo de todo o período dos dados.

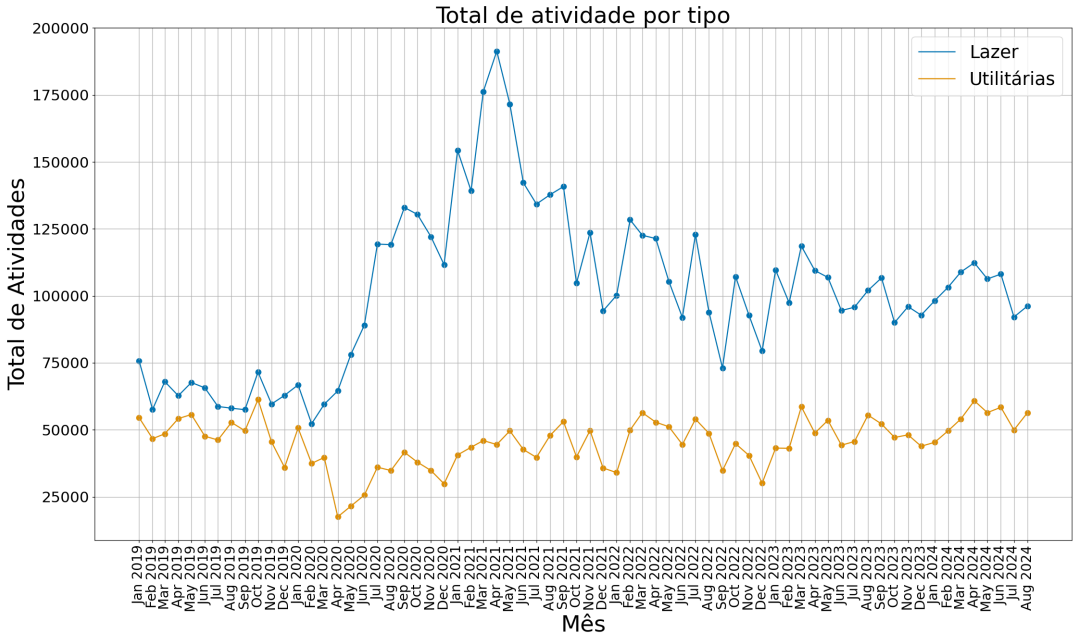
Como maioria, dadas as devidas proporções, as atividades/viagens de lazer acabam por modelar a variação de atividades/viagens totais, observando uma correlação praticamente perfeita. Enquanto isso, as atividades/viagens utilitárias não possuem o mesmo comportamento ao longo do período, com coeficientes de correlação muito baixos, comparadas às atividades/viagens recreativas e totais.

Figura 21 – Total de viagens por tipo ao longo do período



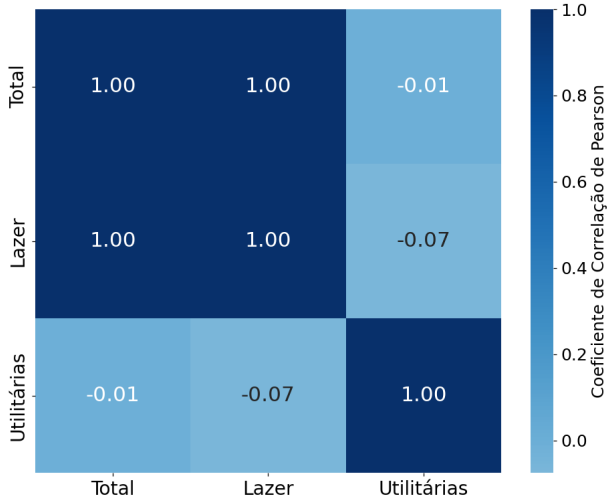
Fonte: Elaborada pelo autor.

Figura 22 – Total de atividades por tipo ao longo do período



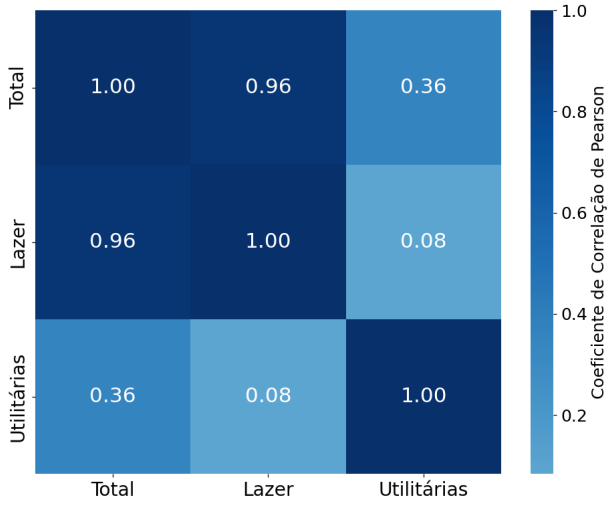
Fonte: Elaborada pelo autor.

Figura 23 – Matriz de Correlação de Pearson - Viagens por Tipo



Fonte: Elaborada pelo autor.

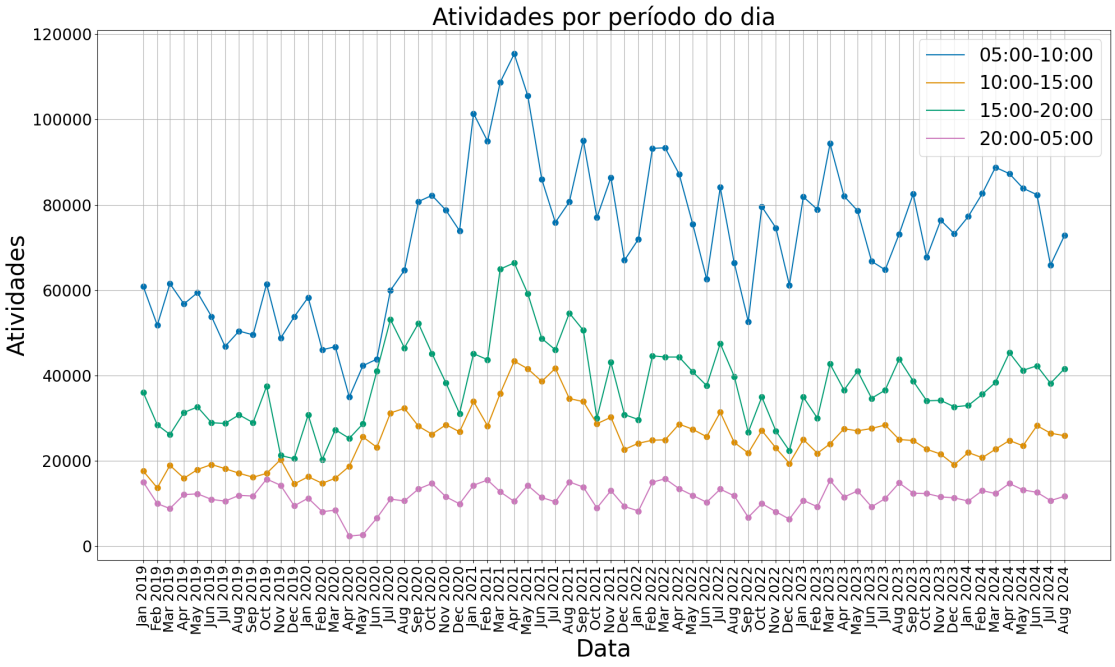
Figura 24 – Matriz de Correlação de Pearson - Atividades por Tipo



Fonte: Elaborada pelo autor.

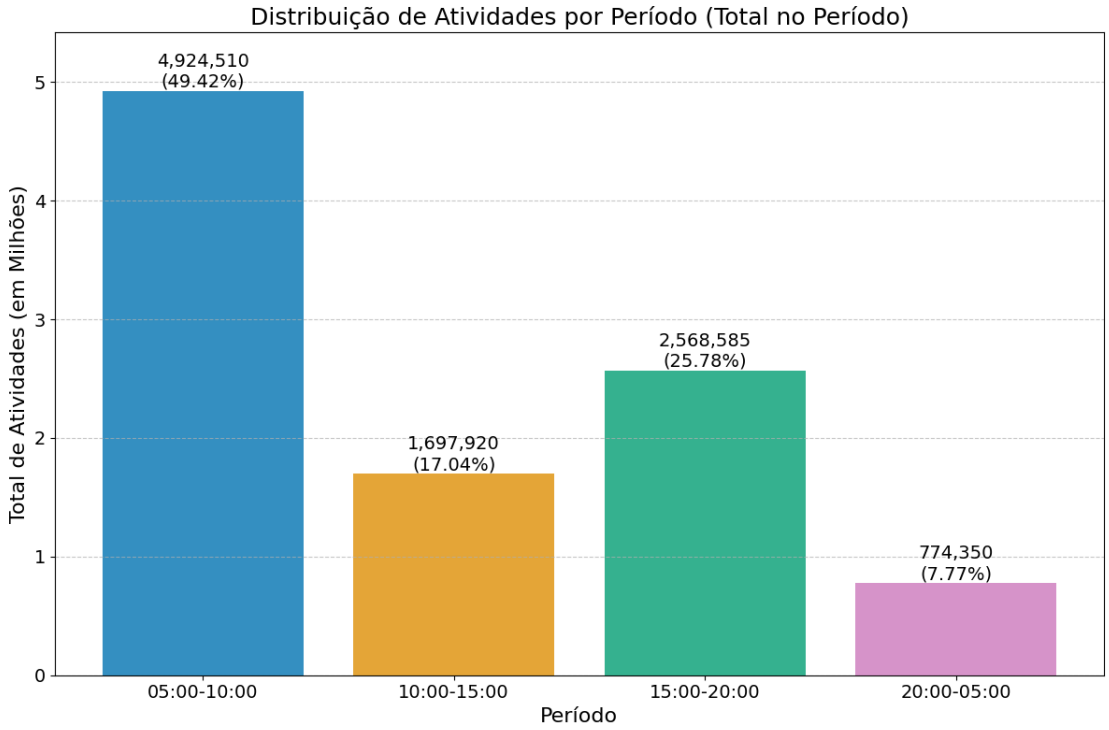
Outra questão a ser analisada é o espalhamento das atividades ao longo do dia. Pode-se observar os padrões horários ao longo do período de estudo representados pela Figura 25 e a distribuição representada pelo gráfico da Figura 26. De maneira geral, as atividades estão bem distribuídas ao longo do dia, com o período predominante sendo entre 05:00 e 10:00 (período este definido como '*morning*'), contendo praticamente metade das atividades totais.

Figura 25 – Total de atividades por período do dia, ao longo do período de estudo



Fonte: Elaborada pelo autor.

Figura 26 – Distribuição das atividades por período do dia

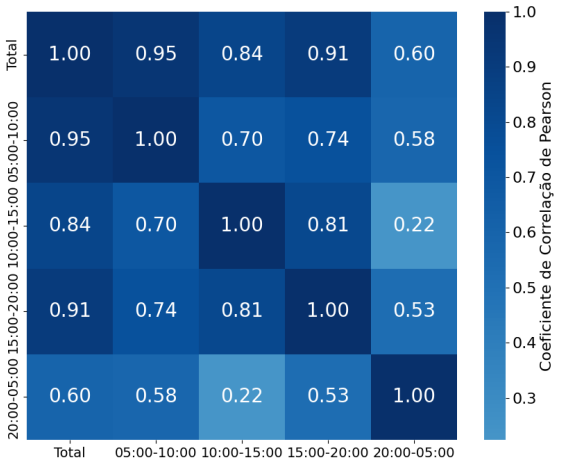


Fonte: Elaborada pelo autor.

Apoiando-se no caráter recreativo observado nas viagens de bicicleta, e levando em consideração os dois períodos mais populares do dia, é razoável inferir que os períodos pré e pós jornada de trabalho são comumente escolhidos pelos usuários do aplicativo, para prática do ciclismo. Já o período entre 20:00 e 05:00 (definido como '*overnight*'), obteve os menores valores encontrados. A escuridão, principalmente em trechos pouco iluminados e a preocupação com a segurança neste período, podem ser citadas como possíveis explicações para a ocorrência de tal fenômeno.

Com relação à possíveis mudanças de padrões horários ao longo do período de estudo, notou-se um considerável grau de correlação entre os valores das atividades totais, com os três grupos de horários mais populares (05:00-10:00, 10:00-15:00, 15:00-20:00), como apoiado pela matriz de correlação de Pearson da Figura 27. Por outro lado, as atividades entre 20:00 e 05:00, não seguem as mesmas tendências dos demais horários, possivelmente por se tratarem de atividades menos rotineiras.

Figura 27 – Matriz de Correlação de Pearson - Período do dia

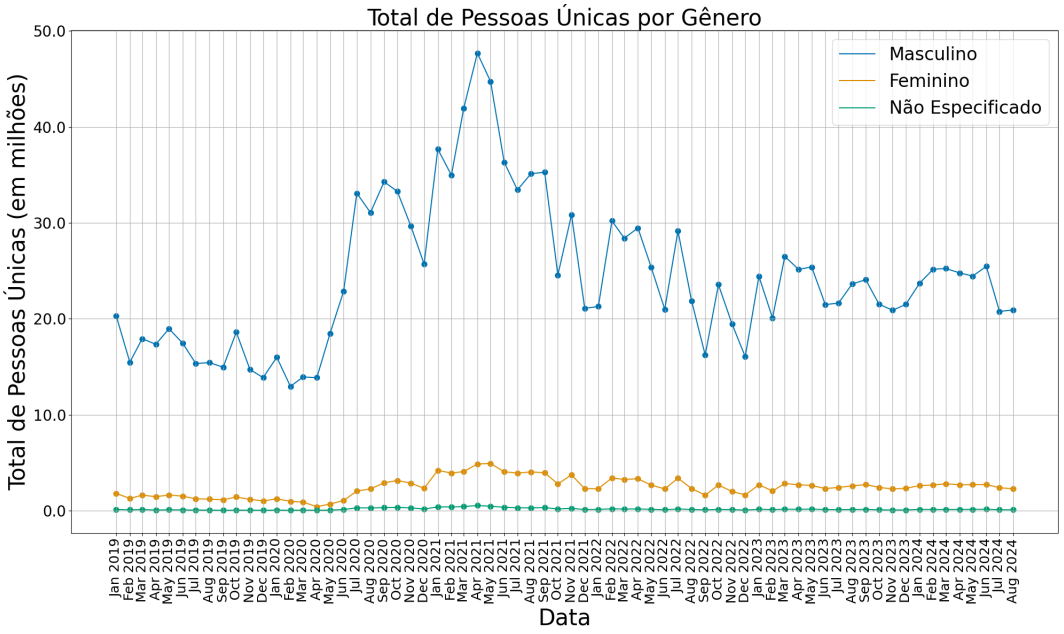


Fonte: Elaborada pelo autor.

4.2 Características dos Ciclistas

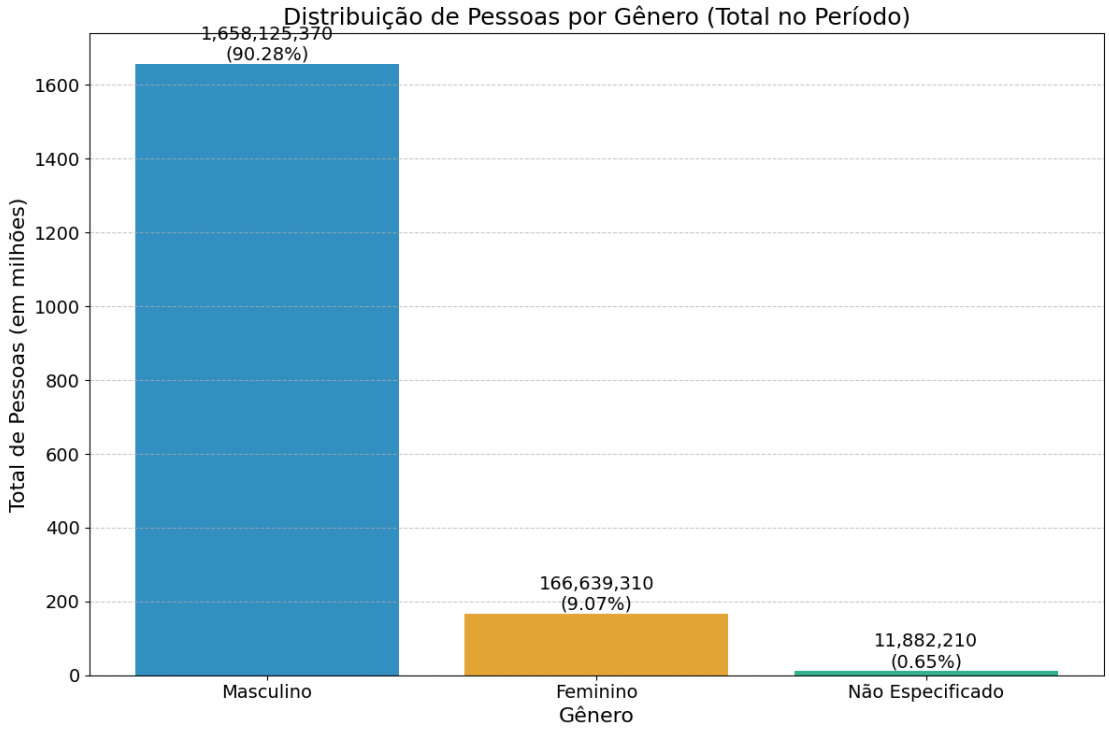
Analisando os resultados obtidos, pôde-se notar algumas características dominantes dos ciclistas em São Paulo. A influência do gênero na variação do número de pessoas ao longo do tempo, é demonstrada pelo gráfico de série temporal contido na Figura 28. Já sobre a proporção de pessoas por gênero, verifica-se um grande predomínio de usuários do sexo masculino, como é possível visualizar no gráfico da Figura 29.

Figura 28 – Total de pessoas únicas por gênero ao longo do período



Fonte: Elaborada pelo autor.

Figura 29 – Distribuição de usuários únicos por gênero



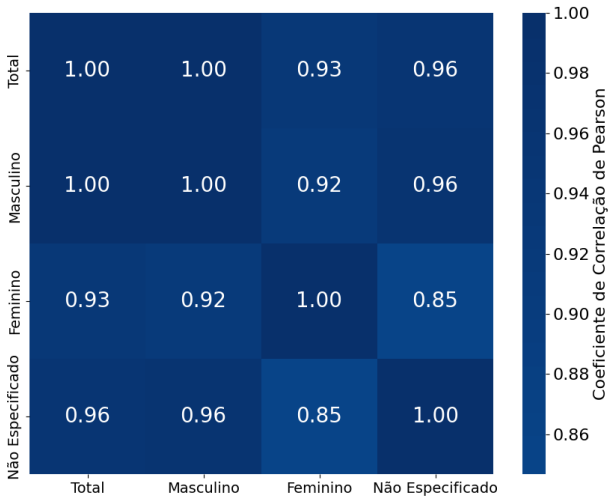
Fonte: Elaborada pelo autor.

Embora correspondam a uma ligeira maioria na população de São Paulo, mulheres compõem apenas cerca de 9% dos usuários ciclistas do aplicativo na capital paulista. As motivações para tal passam muito pelo contexto da sociedade paulistana e mundial. Esse número é similar aos dados de viagens de bicicleta apresentados pela pesquisa OD 2017 (VIANNA JR.; SOUZA; MÜLFARTH, 2024) realizada pelo Metrô de São Paulo.

O estudo de Harkot (2018) busca entender características da cidade de São Paulo e da sociedade que influenciem a prática ou não prática de ciclismo pelo público feminino. A pesquisa mostrou que grande parte das mulheres ciclistas sofreram alguma tentativa de violência, seja ela motivada por trânsito ou gênero. A autora também cita questões culturais, como a menor propensão de mulheres a utilizarem bicicletas quando jovens, quando comparadas aos homens na mesma idade, e questões socioeconômicas como a maior popularidade do ciclismo entre mulheres de maior renda, como possíveis explicações para a desigualdade de gênero neste tipo de mobilidade.

Analisando novamente o gráfico da Figura 28, agora em conjunto com a matriz de correlação de Pearson contida na Figura 30, pode-se notar que as subidas e descidas na quantidade de usuários não estão atreladas a alguma questão de gênero, pois as curvas correspondentes às porções masculina, feminina e não especificada possuem comportamentos similares ao longo de todo o período. O mesmo ocorre, analisando a relação da variação dos gêneros comparado com o valor total de pessoas únicas.

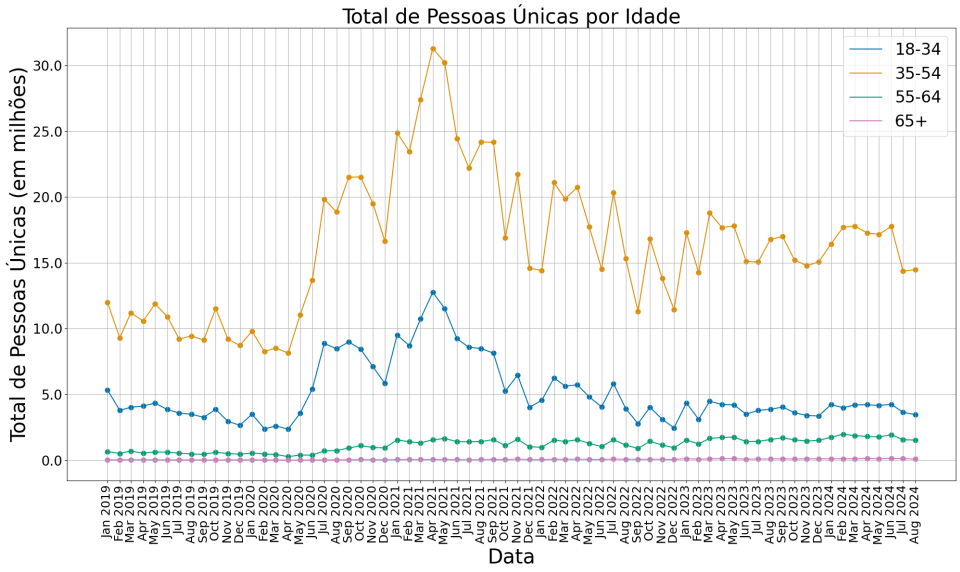
Figura 30 – Matriz de Correlação de Pearson - Gênero



Fonte: Elaborada pelo autor.

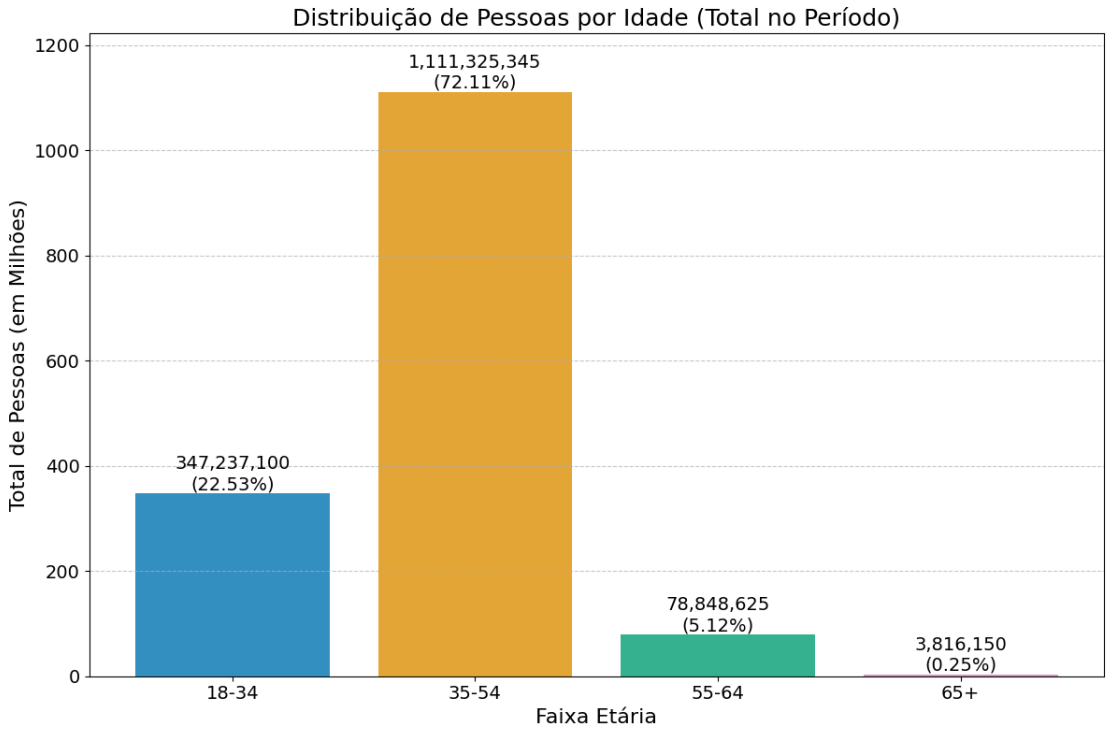
Analisando agora a questão etária, verifica-se um predomínio das faixas etárias mais jovens, como apoiado pela série temporal da Figura 31 e pelo gráfico de barras da Figura 32. Pessoas entre 35 e 54 anos correspondem a grande maioria, contemplando cerca de 72% dos usuários únicos, seguido pelo grupo etário entre 18 e 34 anos, com mais de 22%. Enquanto isso, indivíduos com 55 anos ou mais correspondem a menos de 6% das pessoas únicas. Esta distribuição é consideravelmente similar à obtida pela pesquisa OD 2017, embora as faixas etárias tenham sido divididas de maneira diferente (Figura 33).

Figura 31 – Total de pessoas únicas por faixa etária ao longo do período



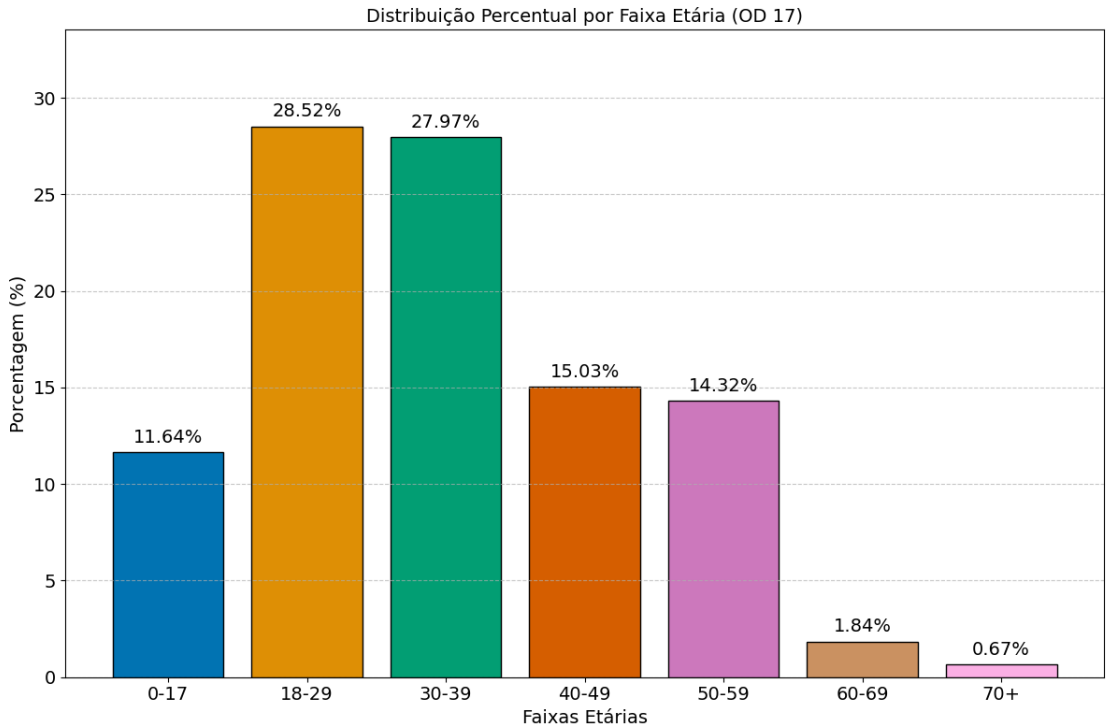
Fonte: Elaborada pelo autor.

Figura 32 – Distribuição de usuários únicos por faixa etária



Fonte: Elaborada pelo autor.

Figura 33 – Distribuição percentual por faixa etária (OD 2017)

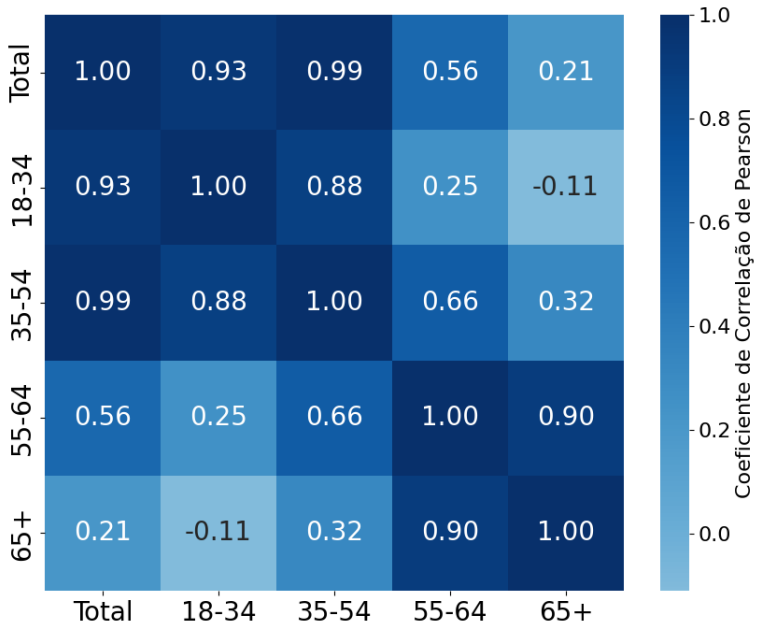


Fonte: Elaborada pelo autor.

A perda de vigor físico conforme o avanço de idade é uma causa razoável para explicar a diminuição de praticantes de ciclismo nas faixas etária mais velhas. O processo de envelhecimento é caracterizado por perdas de desempenho quanto a coordenação, flexibilidade, força, velocidade e resistência (HOLLMANN et al., 2007). Além disso, fatores estruturais também afetam a popularidade da prática do ciclismo entre adultos mais velhos, entre eles, o tipo de infraestrutura disponível, separação adequada do tráfego motorizado, densidade do tráfego, uniformidade e condições das ciclovias, entre outros (VAN CAUWENBERG et al., 2019). A menor familiaridade desses indivíduos com dispositivos móveis, também se mostra uma possível explicação.

Enfim, buscando identificar uma possível influência do fator idade na variação da quantidade de usuários únicos ao longo do período, realizou-se o mesmo processo da análise por gênero. Levando em consideração a série temporal contida na Figura 31 e matriz de correlação contida na Figura 34, percebe-se que, diferentemente do caso anterior, nem todas as curvas referentes às faixas etárias possuem um elevado grau de dependência.

Figura 34 – Matriz de Correlação de Pearson - Idade



Fonte: Elaborada pelo autor.

É possível notar, que os grupos etários predominantes (18-34 e 35-54 anos) são consideravelmente correlatos entre si e comparados ao número total de pessoas. Da mesma forma, os grupos etários mais velhos (55-64 e 65+ anos) também possuem alto grau de correlação entre si. Entretanto, o mesmo não se aplica ao correlacionar estes dois grupos maiores. A partir disso, pode-se concluir que o número de praticantes mais jovens segue tendências diferentes do número de praticantes mais velhos e que motivações para o aumento/diminuição

de número de usuários de um grupo não necessariamente valem para o outro.

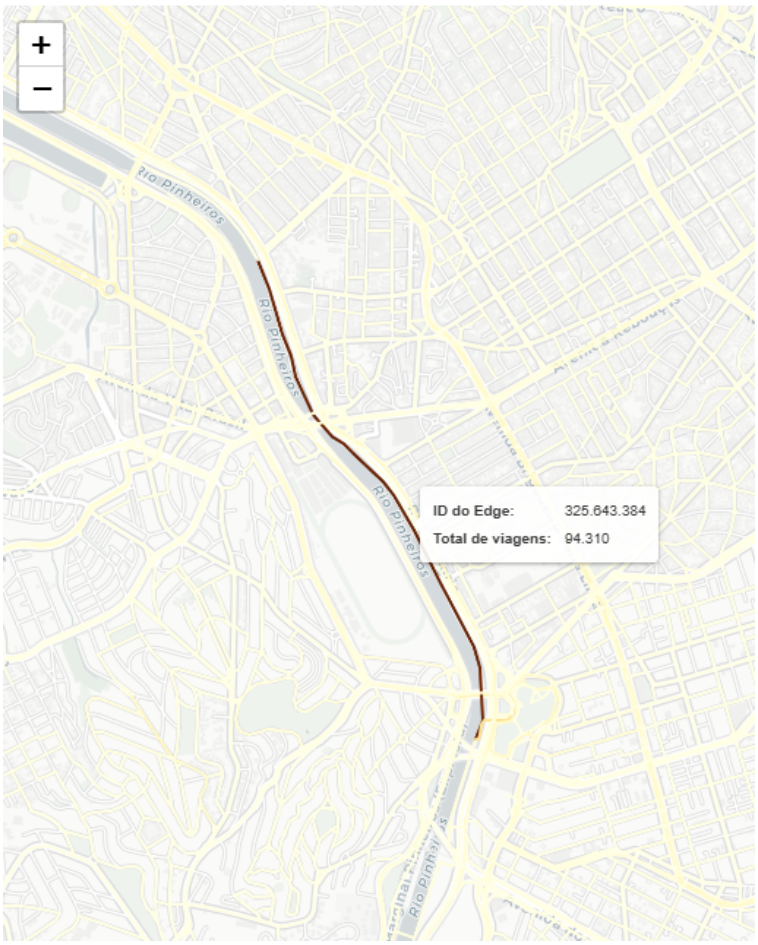
4.3 Resultados Geoespaciais

A partir dos resultados obtidos das análises geoespaciais, pôde-se identificar alguns padrões de deslocamento e localidades de destaque na cidade de São Paulo, além de entender a distribuição das mesmas.

Os maiores fluxos de ciclistas de acordo com os dados de viagens de bicicleta na capital paulista ocorrem nas imediações da Marginal Pinheiros, e, consequentemente, as zonas OD, distritos e subprefeituras com maior concentração de viagens, são aquelas pelas quais a Marginal Pinheiros se estende. Este comportamento também é observado na concentração das origens/destinos das atividades dos usuários.

Como é possível observar na Figura 35, ao utilizar os números de viagens brutos para elaborar as escalas, a diferença na popularidade desse trecho, comparado às demais localidades de cidade é tanta, que a visualização acaba por dizer pouco sobre o resto da cidade.

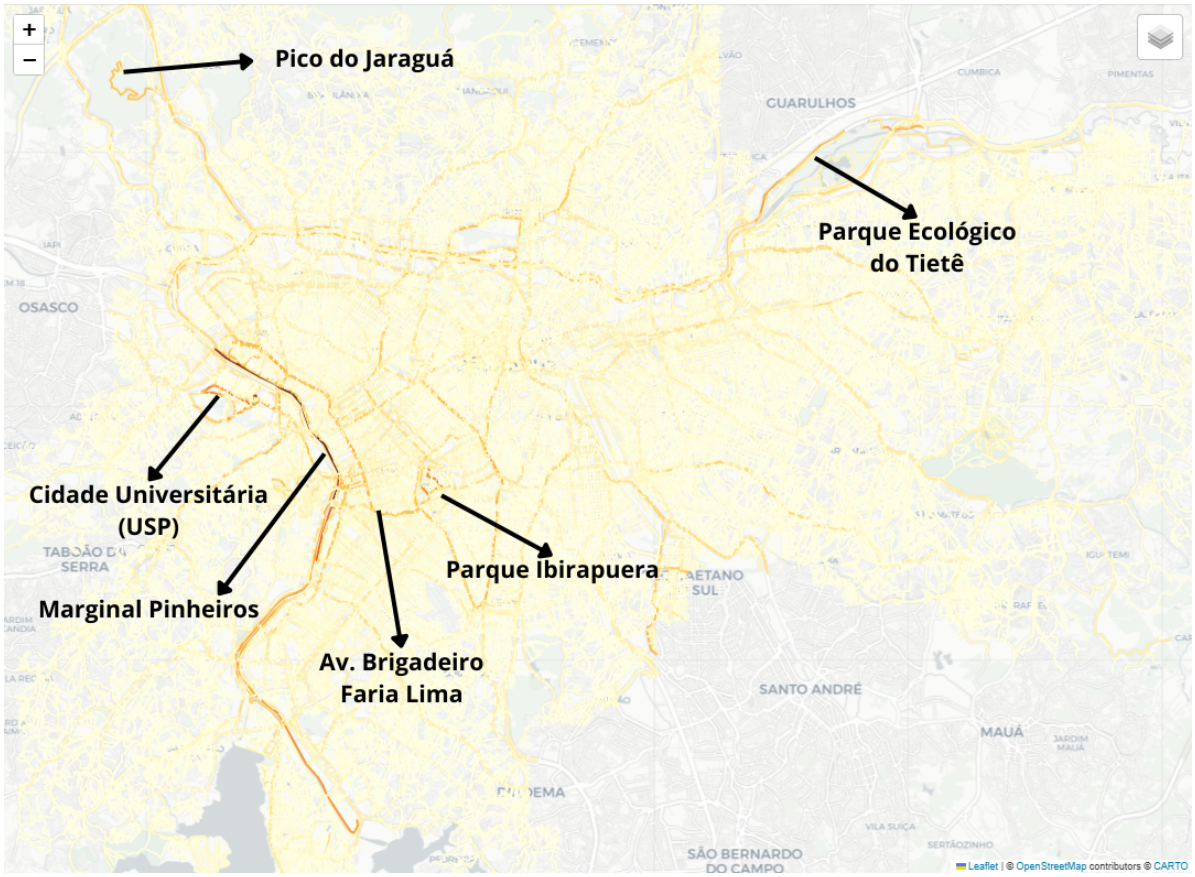
Figura 35 – Concentração de viagens nas proximidades da Marginal Pinheiros



Fonte: Elaborada pelo autor.

Uma solução encontrada foi aplicar uma escala de raiz cúbica na visualização, para suavizar dados discrepantes e manter uma melhor relação entre eles, adotando um caráter menos agressivo do que outras escalas, como as logarítmicas. Aplicando a escala de raiz cúbica, conseguiu-se notar outras localidades populares, como o Parque Ecológico do Tietê, o Parque Ibirapuera, a Cidade Universitária da Universidade de São Paulo (USP), as imediações da Avenida Brigadeiro Faria Lima, rotas próximas ao Pico do Jaraguá, entre outras (Figura 36).

Figura 36 – Remapeamento utilizando escala de raiz cúbica

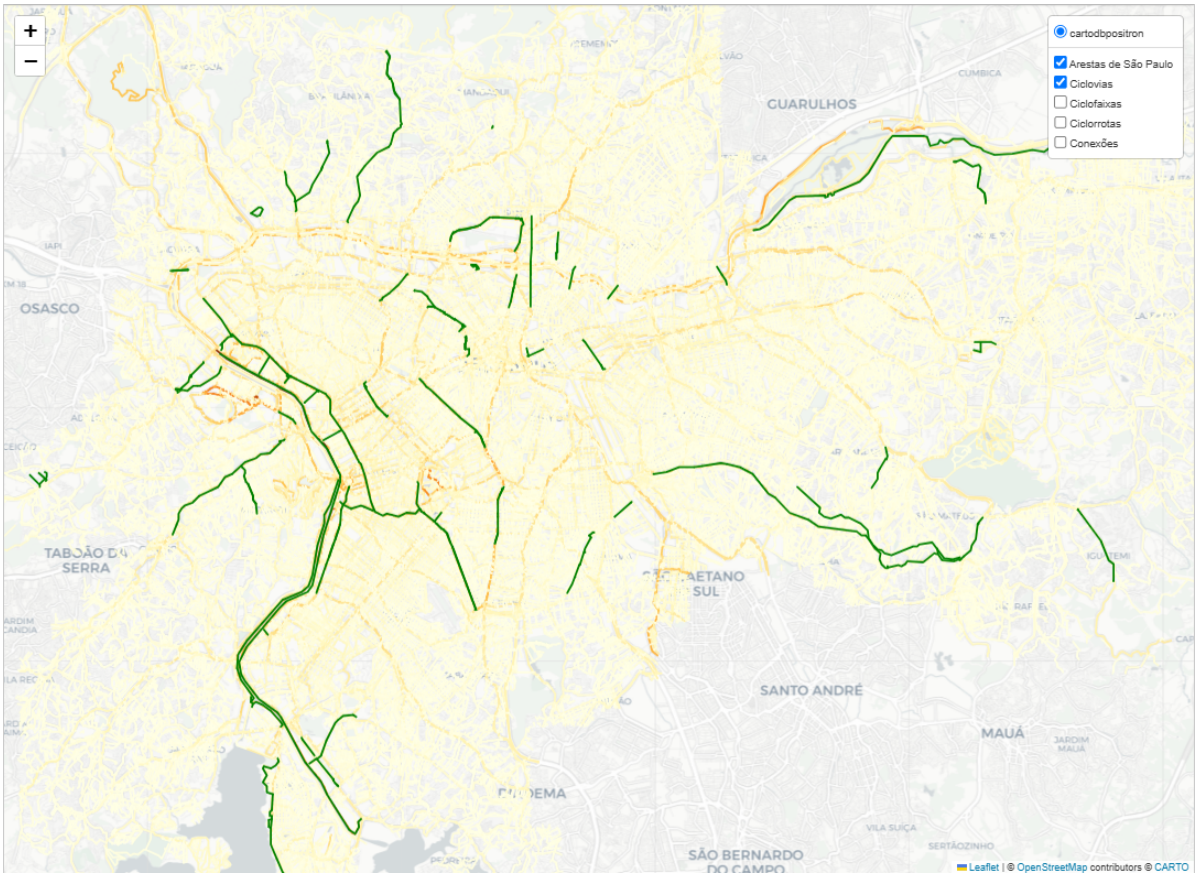


Fonte: Elaborada pelo autor.

Ao analisar a popularidade da Marginal Pinheiros em conjunto com a cobertura da infraestrutura cicloviária da área, conforme a Figura 37, pode-se observar um potencial motivo para tal. A presença de um extenso trecho de ciclovias, e outras estruturas dedicadas, às margens do Rio Pinheiros, facilita a locomoção de ciclistas e explica a grande concentração de usuários naquela região.

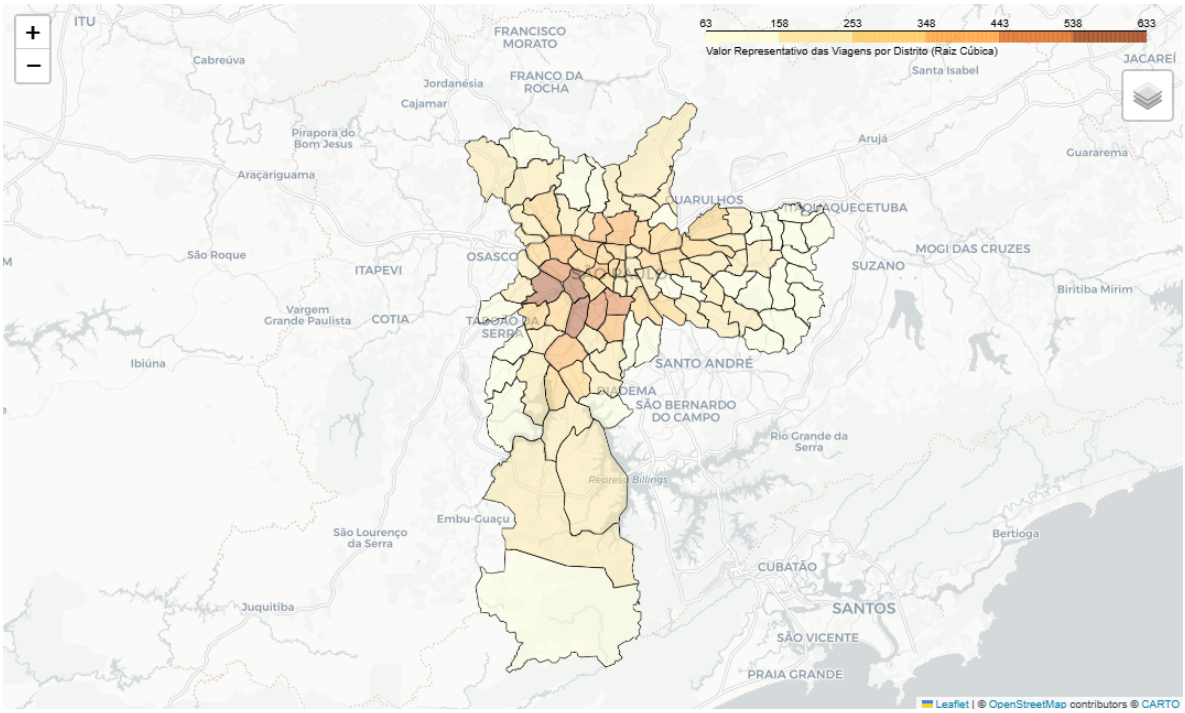
Investigando uma possível correlação entre a população por distrito e o total de viagens por distrito, utilizou-se novamente o coeficiente de correlação de Pearson. Verificou-se que a quantidade de viagens de bicicleta em um distrito (Figura 38) não possui correlação direta com a população do mesmo (em 2021, dados do SEADE), como apoiado pela matriz de correlação de Pearson da Figura 39.

Figura 37 – Presença de Infraestrutura Urbana próxima a Marginal Pinheiros



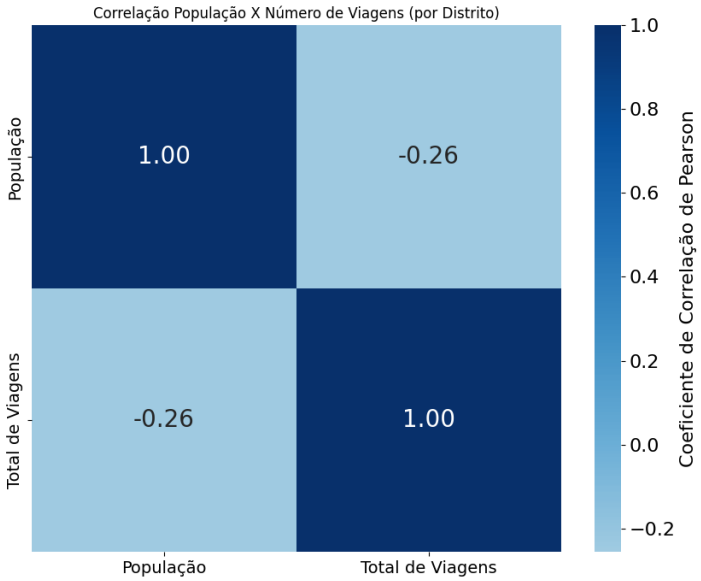
Fonte: Elaborada pelo autor.

Figura 38 – Distribuição de Viagens por Distrito



Fonte: Elaborada pelo autor.

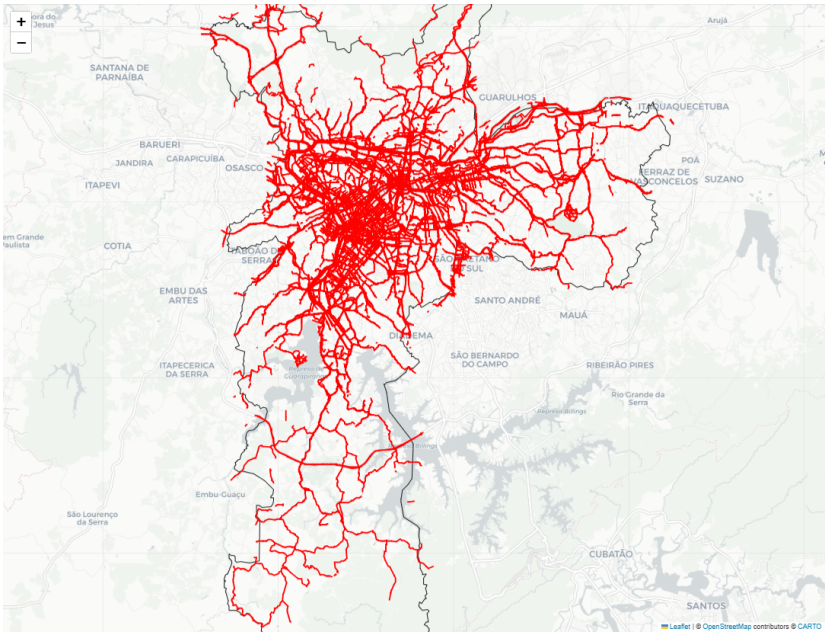
Figura 39 – Matriz de Correlação de Pearson - Viagens por Distrito X População por Distrito



Fonte: Elaborada pelo autor.

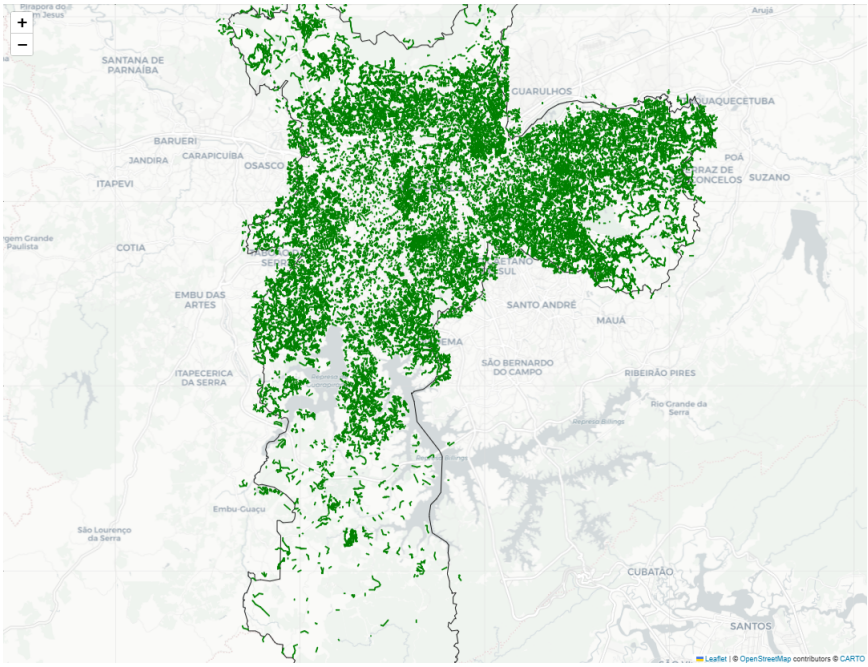
Verificando a divisão das arestas em quartis, baseada na quantidade de viagens em cada uma delas, notou-se que as arestas com mais viagens são em sua maioria contínuas (Figura 40), formando trechos maiores, enquanto as com menos viagens (Figura 41), formam menos caminhos contínuos. Esse comportamento é proporcional nos quartis intermediários.

Figura 40 – Q4 das arestas por quantidade viagens



Fonte: Elaborada pelo autor.

Figura 41 – Q1 das arestas por quantidade viagens



Fonte: Elaborada pelo autor.

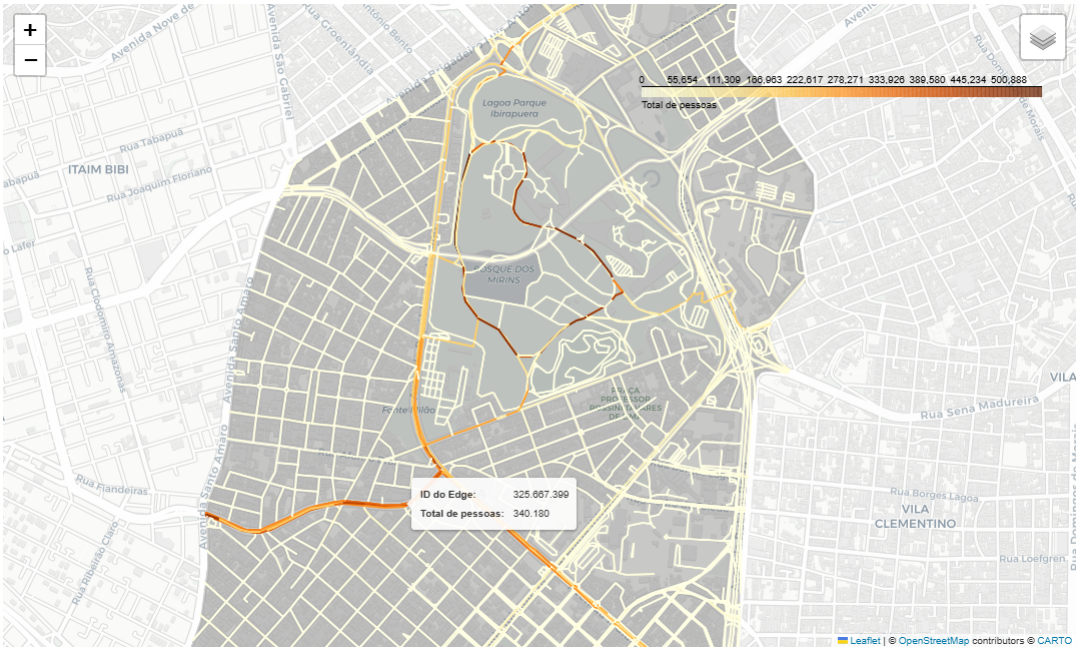
Além destes resultados mais generalizados, também foi possível observar o comportamento e a concentração dos ciclistas em óticas mais específicas, ao filtrar os dados do aplicativo por região administrativa (zonas OD, distritos, subprefeituras), gênero, faixa etária e período do dia.

Desse modo, é possível atender às necessidades específicas de alguma região, já que cada uma delas possui suas próprias particularidades. Um indivíduo interessado em uma localidade específica, sob parâmetros específicos, pode obter *insights* personalizados, de maneira completamente isolada.

Essa personalização pode ser observada nas Figuras 42 e 44 e seus respectivos filtros (Figuras 43 e 45). Neste exemplo, pode-se notar que a distribuição de ciclistas em Moema aplicando o filtro de idade é similar à distribuição de ciclistas geral no distrito, tendo o Parque do Ibirapuera e suas vias de acesso em destaque (arestas mais escuras). O número de pessoas em cada aresta é naturalmente diferente, como demonstrado pela escala e a caixa de informações de uma aresta exemplo.

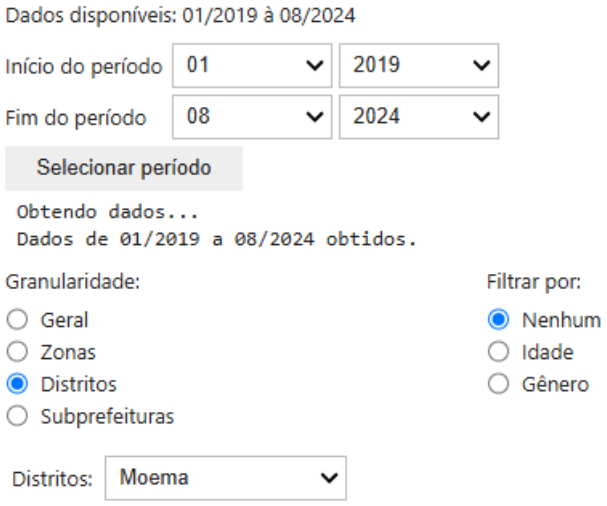
Para outro exemplo, analisando agora a subprefeitura do Butantã, pode-se notar uma grande diferença da distribuição das viagens na Cidade Universitária da Universidade de São Paulo, comparando os períodos de 2019 (Figura 46) e 2020 (Figura 48) e seus respectivos filtros (Figuras 47 e 49, respectivamente). Ao observar os mapas, pode-se perceber que em 2020 (ano pandêmico), outras localidades da subprefeitura receberam mais viagens em comparação com a Cidade Universitária, comportamento diferente do visto em 2019.

Figura 42 – Exemplo de consulta 1 - Pessoas totais no distrito de Moema



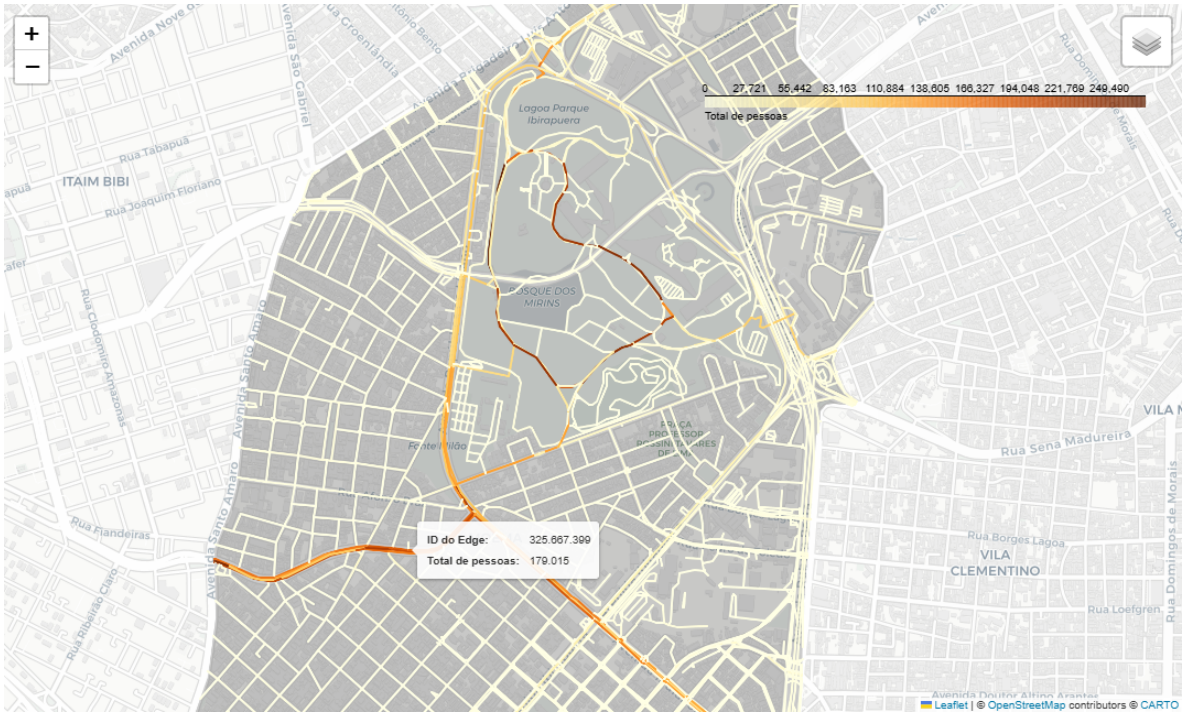
Fonte: Elaborada pelo autor.

Figura 43 – Filtros usados para o mapa da Figura 42



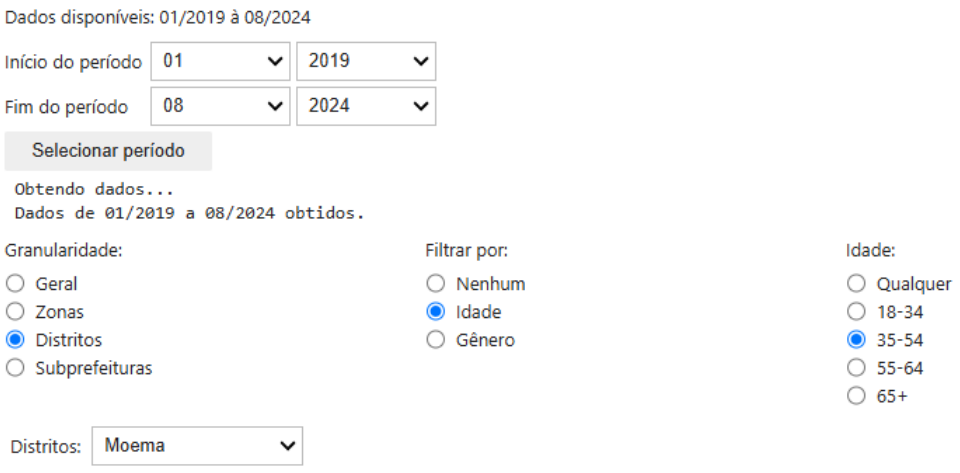
Fonte: Elaborada pelo autor.

Figura 44 – Exemplo de consulta 2 - Pessoas de idade entre 35-54 no distrito de Moema



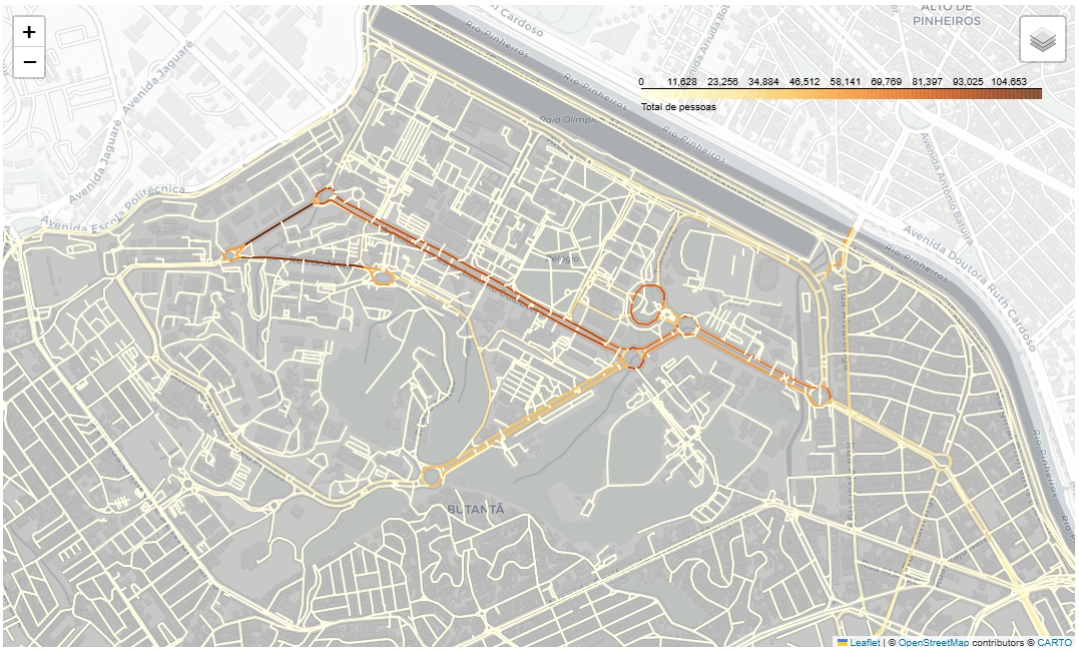
Fonte: Elaborada pelo autor.

Figura 45 – Filtros usados para o mapa da Figura 44



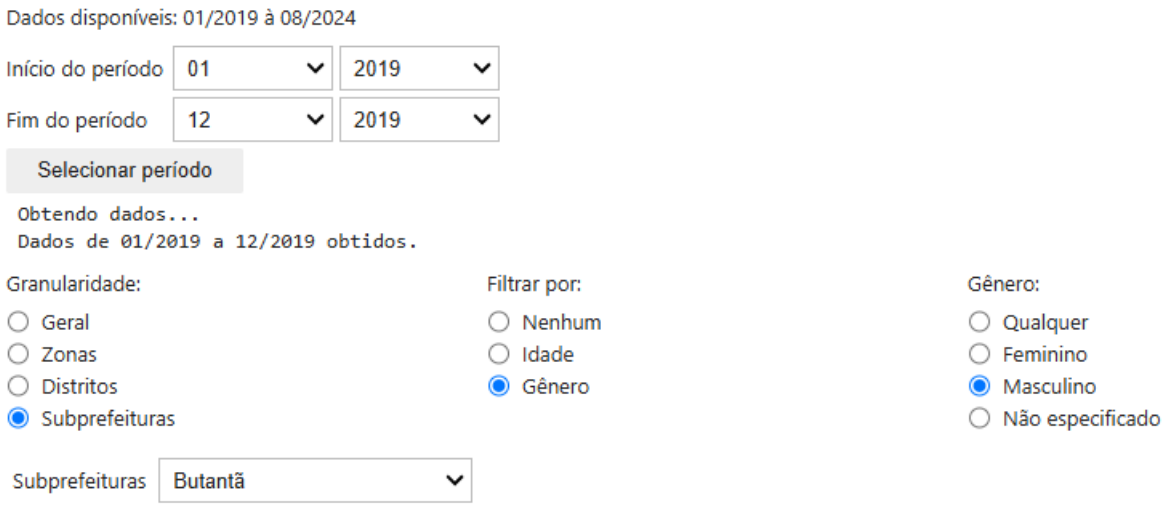
Fonte: Elaborada pelo autor.

Figura 46 – Exemplo de consulta 3 - Pessoas do gênero masculino na subprefeitura do Butantã em 2019



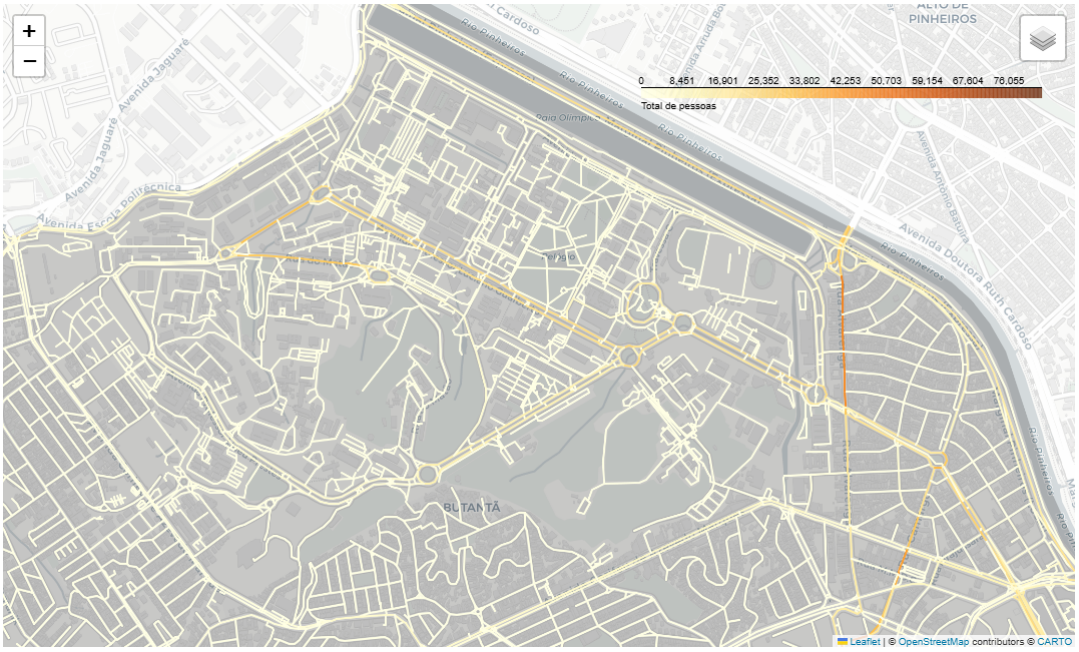
Fonte: Elaborada pelo autor.

Figura 47 – Filtros usados para o mapa da Figura 46



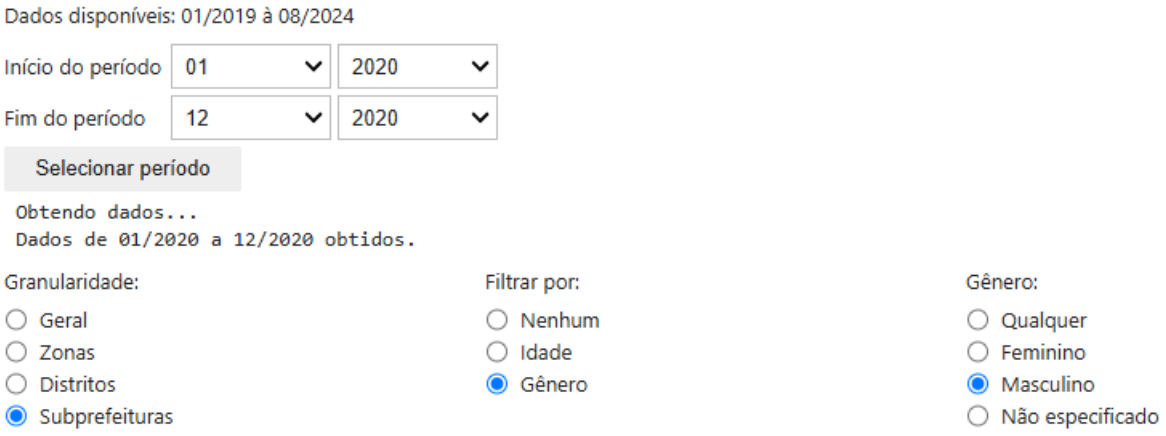
Fonte: Elaborada pelo autor.

Figura 48 – Exemplo de consulta 4 - Pessoas do gênero masculino na subprefeitura do Butantã em 2020 (ano pandêmico)



Fonte: Elaborada pelo autor.

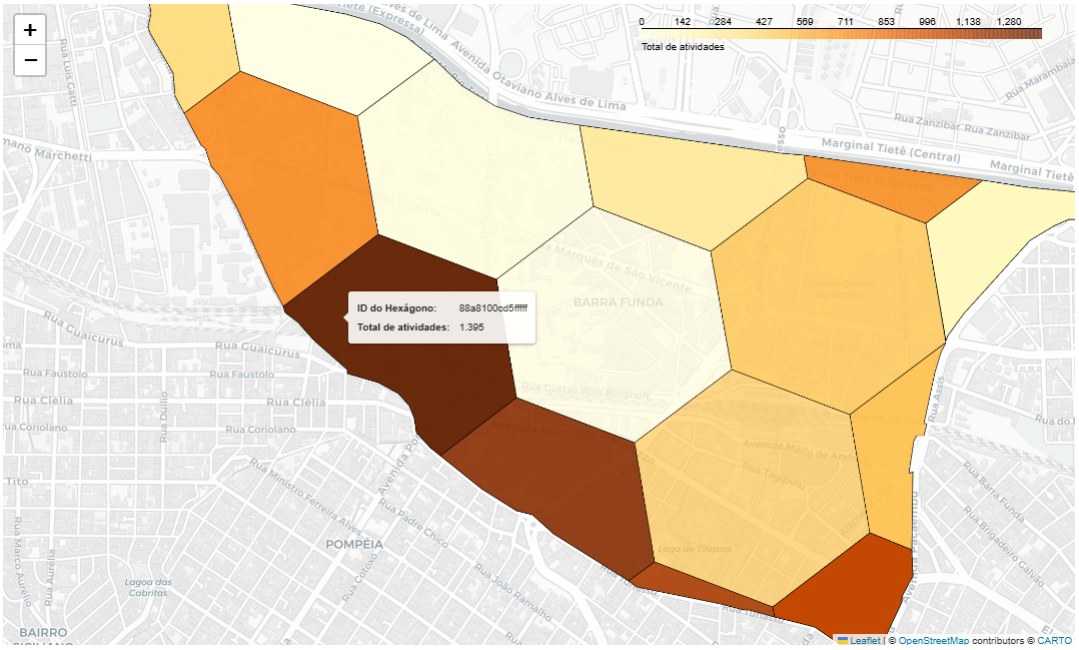
Figura 49 – Filtros usados para o mapa da Figura 48



Fonte: Elaborada pelo autor.

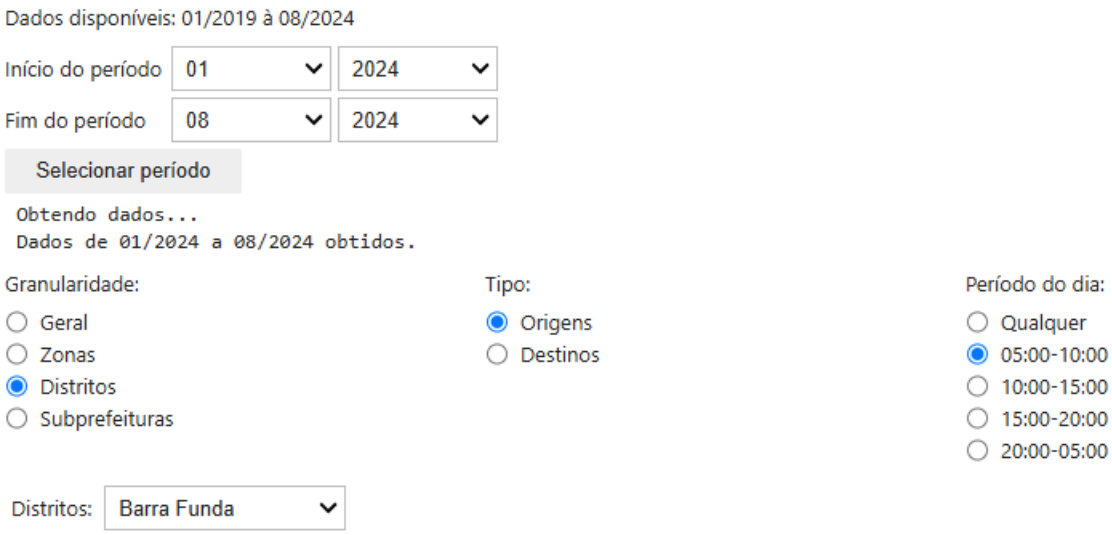
Para a análise das Origens e Destinos das atividades mais populares, também é possível realizar filtragens, como mostram as Figuras 50 e 52, e seus filtros (Figuras 51 e 53, respectivamente). Nestes exemplos, considerou-se o distrito da Barra Funda, e é nítida a diferença entre a distribuições das origens (onde se iniciou uma atividade) quando comparados dois períodos diferentes do dia (05:00-10:00 e 15:00-20:00).

Figura 50 – Exemplo de consulta 5 - Origens das atividades no distrito da Barra Funda entre 05:00 e 10:00



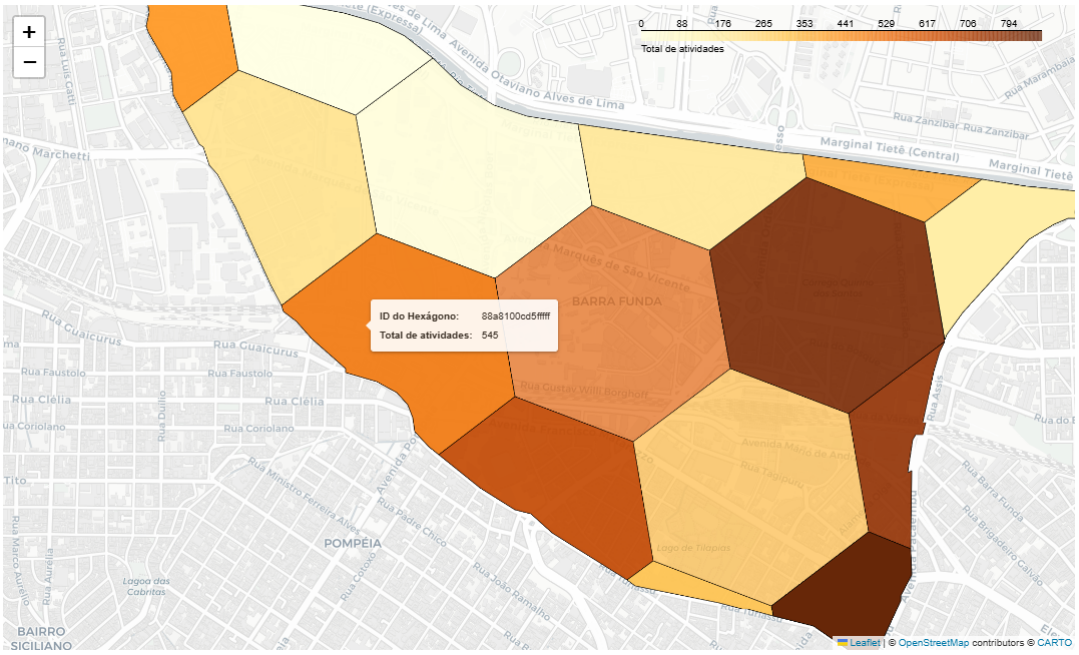
Fonte: Elaborada pelo autor.

Figura 51 – Filtros usados para o mapa da Figura 50



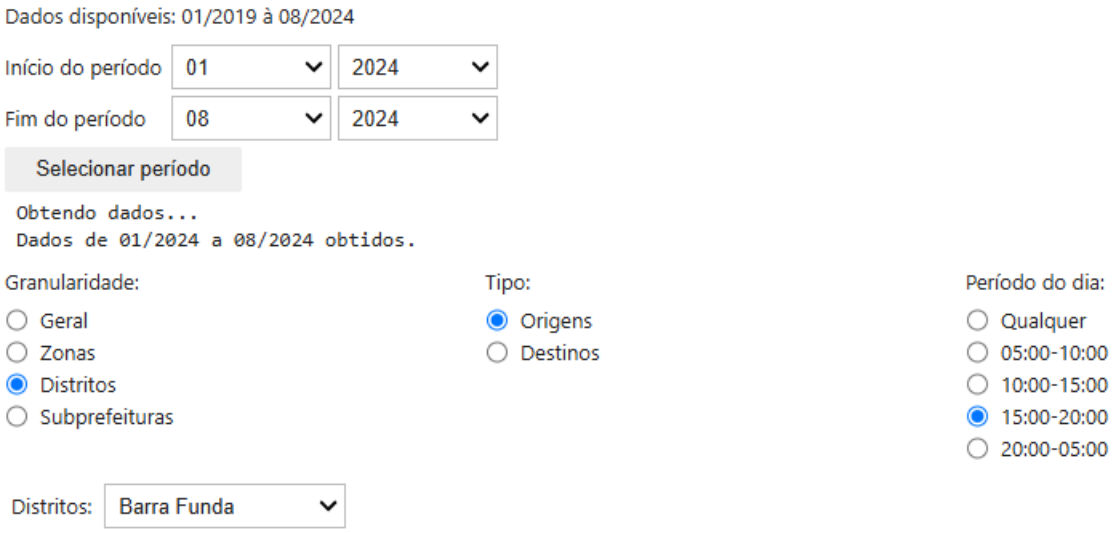
Fonte: Elaborada pelo autor.

Figura 52 – Exemplo de consulta 6 - Origens das atividades no distrito da Barra Funda entre 15:00 e 20:00



Fonte: Elaborada pelo autor.

Figura 53 – Filtros usados para o mapa da Figura 52



Fonte: Elaborada pelo autor.

4.4 Considerações

Os resultados apresentados neste capítulo expõem análises quanto a ciclistas na cidade de São Paulo.

Por meio das análises de características das atividades/viagens e dos usuários pode-se traçar um perfil sobre como a bicicleta é utilizada na capital paulista e quem a utiliza (quanto ao gênero e idade). Foi possível perceber nitidamente a questão da desigualdade de gênero na mobilidade, a diferenciação da faixa etária dos indivíduos. Também notou-se os objetivos gerais dos ciclistas, com a predominância de viagens de lazer (práticas de atividade física, preocupações com a saúde).

Por fim, utilizando as análises geoespaciais, pôde-se notar a distribuição das viagens ao redor da cidade. A concentração de grande parte de viagens às margens do Rio Pinheiros, e a presença de uma infraestrutura cicloviária adequada na região, ajuda a verificar a relação entre o ciclismo e as condições para a prática de tal. Também mostrou-se possível a personalização das análises, para atender as necessidades de estudo em granularidades espaciais diferentes.

5 Conclusão

O presente trabalho propôs analisar dados de ciclistas usuários de aplicativo de viagens de bicicleta na cidade de São Paulo e buscar identificar padrões e tendências no comportamento dos mesmos, por meio de técnicas de Ciência de Dados. Para isso, utilizou-se recursos computacionais, estatísticos e geoespaciais.

Utilizando as técnicas e conceitos apresentados ao longo deste documento, buscou-se desenvolver uma ferramenta que possibilite a realização de análises ricas e customizáveis, que agreguem valor aos estudos e gerem resultados claros e intuitivos, que possam ser utilizados com evidências, no processo de implementação de PPBEs, por parte de órgãos públicos da cidade de São Paulo, como a CET e o Metrô de São Paulo, por exemplo, nas áreas de infraestrutura cicloviária e de incentivo ao uso de mobilidade ativa.

Os ambientes desenvolvidos em forma de *notebooks*, as análises realizadas e os resultados foram disponibilizados em repositório referenciado na Seção **??**, permitindo com que indivíduos e instituições com acesso aos dados de viagens de bicicleta possam usufruir deles para auxiliar estudos futuros sobre o assunto.

Pode-se recomendar como trabalhos futuros na área, utilizando a ferramenta desenvolvida:

- Análises para implementação de infraestrutura cicloviária em regiões específicas. Identificação de lacunas no sistema cicloviário, verificação de localidades pouco exploradas por ciclistas, entre outras;
- Integração das análises com outras bases de dados e parâmetros, como dados de segurança viária, indicadores de tráfego, dados socioeconômicos, entre outros; e
- Aprofundamento no relacionamento entre as análises desenvolvidas e pesquisas já realizadas, como a OD.

Referências

CÂMARA, G.; MONTEIRO, A. M.; FUCKS, S. D.; CARVALHO, M. S. Análise espacial e geoprocessamento. *Análise espacial de dados geográficos. Brasília: EMBRAPA*, p. 21–54, 2004. Disponível em: urlhttps://portalidea.com.br/cursos/bsico-em–anlise-espacial-de-dados-geogrficos-apostila02.pdf . Acesso em: 23 out. 2024.

Companhia de Engenharia de Tráfego - CET. *Sistema Cicloviário*. 2016. Disponível em: <https://www.cetsp.com.br/consultas/bicicleta/sistema-cicloviario.aspx> . Acesso em: 24 out. 2024.

COMPANHIA DO METROPOLITANO DE SÃO PAULO - METRÔ. *Relatório Síntese*. São Paulo: [s.n.], 2019. Disponível em: <https://transparencia.metrosp.com.br/dataset/pesquisa-origem-e-destino/resource/b3d93105-f91e-43c6-b4c0-8d9c617a27fc> . Acesso em: 24 out. 2024.

COSTA, C. G. F.; DA SILVA, E. V. O que realmente importa no processo de tomada de decisão considerando políticas públicas baseadas em evidência. *Revista Administração em Diálogo-RAD*, v. 18, n. 2, p. 124–143, 2016. Disponível em: <https://revistas.pucsp.br/index.php/rad/article/view/rad.v18i2.20315> . Acesso em: 25 out. 2024.

FIGUEIREDO FILHO, D. B.; SILVA JÚNIOR, J. A. Desvendando os mistérios do coeficiente de correlação de pearson (r). *Revista Política Hoje*, v. 18, n. 1, p. 115–146, 2009. Disponível em: <https://dirin.s3.amazonaws.com/drive_materias/1666287394.pdf> . Acesso em 24 out. 2024.

FISCHER, J.; NELSON, T.; WINTERS, M. Riding through the pandemic: Using strava data to monitor the impacts of covid-19 on spatial patterns of bicycling. *Transportation research interdisciplinary perspectives*, v. 15, p. 100667, 2022. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2590198222001270> . Acesso em: 17 out. 2024.

GERIKE, R.; NAZELLE, A. de; NIEUWENHUIJSEN, M.; PANIS, L. I.; ANAYA, E.; AVILA-PALENCIA, I.; BOSCHETTI, F.; BRAND, C.; COLE-HUNTER, T.; DONS, E.; ERIKSSON, U.; GAUPP-BERGHAUSEN, M.; KAHLMEIER, S.; LAEREMANS, M.; MUELLER, N.; ORJUELA, J. P.; RACIOPPI, F.; RASER, E.; ROJAS-RUEDA, D.; SCHWEIZER, C.; STANDAERT, A.; UHLMANN, T.; WEGENER, S.; GÖTSCHI, T. Physical activity through sustainable transport approaches (pasta): a study protocol for a multicentre project. *BMJ open*, v. 6, n. 1, p. 2, 2016. Disponível em: <https://bmjopen.bmj.com/content/6/1/e009924.abstract> . Acesso em: 28 set. 2024.

GERIKE, R.; NAZELLE, A. de; WITTWER, R.; PARKIN, J. Special issue "walking and cycling for better transport, health and the environment". *Transportation research Part A: Policy and practice*, v. 123, p. 1–6, 2019. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0965856419302873> . Acesso em: 17 out. 2024.

GUEDES, T. A.; MARTINS, A. B. T.; ACORSI, C. R. L.; JANEIRO, V. Estatística descritiva. *Projeto de ensino aprender fazendo estatística*, sn, p. 1–49, 2005. Disponível em: <https://www.ime.usp.br/~rvicente/Guedes_etal_Estatistica_Descritiva.pdf>. Acesso em 31 out. 2024.

HAFERMALZ, E.; JOHNSON, R. B.; HOVORKA, D. S.; RIEMER, K. Beyond 'mobility': A new understanding of moving with technology. *Information Systems Journal*, v. 30, n. 4, p. 762–786, 2020. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/isj.12283> . Acesso em: 17 out. 2024.

HARKOT, M. K. *A bicicleta e as mulheres: mobilidade ativa, gênero e desigualdades socioterritoriais em São Paulo*. Tese (Doutorado) — Universidade de São Paulo, 2018. Disponível em: <https://www.teses.usp.br/teses/disponiveis/16/16139/tde-17092018-153511/publico/MEmarinakohlerharkot_rev.pdf?utm_medium=website&utm_source=archdaily.com.br> . Acesso em 24 out. 2024.

HOLLMANN, W.; STRÜDER, H. K.; TAGARAKIS, C. V.; KING, G. Physical activity and the elderly. *European Journal of Preventive Cardiology*, Oxford University Press, v. 14, n. 6, p. 730–739, 2007. Disponível em: <https://academic.oup.com/eurjpc/article/14/6/730/5933610> .

LUM, C.; KOPER, C. S. Evidence-based policing. In: BRUINSMA, G.; WEISBURD, D. (Ed.). *Encyclopedia of Criminology and Criminal Justice*. Nova Iorque: Springer New York, 2014. p. 1426—-1437. Disponível em: <https://link.springer.com/referencework/10.1007/978-1-4614-5690-2> . Acesso em: 25 out. 2024.

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. Disponível em: <https://www.liebertpub.com/doi/full/10.1089/big.2013.1508> . Acesso em: 28 set. 2024.

SARAGIOTTO, D. *Mobilidade ativa como meio de transporte em São Paulo*. São Paulo: [s.n.], 2020. Mobilidade. Disponível em: <https://mobilidade.estadao.com.br/meios-de-transporte/mobilidade-ativa-como-meio-de-transporte-em-sao-paulo/> . Acesso em: 17 out. 2024.

TORRE-BASTIDA, A. I.; SER, J. del; LAÑA, I.; ILARDIA, M.; BILBAO, M. N.; CAMPOS-CORDOBÉS, S. Big data for transportation and mobility: recent advances, trends and challenges. *IET Intelligent Transport Systems*, v. 12, n. 8, p. 742–755, 2018. Disponível em: <https://ietresearch.onlinelibrary.wiley.com/doi/full/10.1049/iet-its.2018.5188> . Acesso em: 17 out. 2024.

VAN CAUWENBERG, J.; BOURDEAUDHUIJ, I. D.; CLARYS, P.; GEUS, B. D.; DEFORCHE, B. Older adults' environmental preferences for transportation cycling. *Journal of transport & health*, Elsevier, v. 13, p. 185–199, 2019. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S2214140518304705> .

VIANNA JR., E. de O.; SOUZA, H. A. de; MÜLFARTH, R. C. K. Mobilidade urbana sustentável e modos ativos: soluções para o incremento de viagens de bicicletas em São Paulo. *Submetido para publicação*, 2024.

VINUTHA, H.; POORNIMA, B.; SAGAR, B. Detection of outliers using interquartile range technique from intrusion dataset. In: SPRINGER. *Information and decision sciences: Proceedings of the 6th international conference on ficta*. [S.l.], 2018. p. 511–518. Disponível em: <https://link.springer.com/chapter/10.1007/978-981-10-7563-6_53> . Acesso em 24 out. 2024.