

UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"

FACULDADE DE CIÊNCIAS - CAMPUS BAURU

DEPARTAMENTO DE COMPUTAÇÃO

BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

DANILO ALVES SIQUEIRA

**ANÁLISE DO IMPACTO DA SEGURANÇA VIÁRIA NA
MOBILIDADE ATIVA**

BAURU

Novembro/2023

DANILO ALVES SIQUEIRA

ANÁLISE DO IMPACTO DA SEGURANÇA VIÁRIA NA MOBILIDADE ATIVA

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/2023

S618a Siqueira, Danilo Alves
Análise do Impacto da Segurança Viária na Mobilidade Ativa
/ Danilo Alves Siqueira. -- Bauru, 2023
70 p. : il., tabs., mapas

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

1. Mobilidade Ativa. 2. Segurança Viária. 3. Ciência de
Dados. 4. Análise Geoespacial. 5. Visualização de Dados. I.
Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da
Faculdade de Ciências, Bauru. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Danilo Alves Siqueira

Análise do Impacto da Segurança Viária na Mobilidade Ativa

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 Ciência da Computação

Profa. Dra. Simone das Graças Domingues Prado

Universidade Estadual Paulista "Júlio de Mesquita Filho"

Faculdade de Ciências

Departamento de Ciência da Computação

Prof. Dr. Kelton Augusto Pontara da Costa

Universidade Estadual Paulista "Júlio de Mesquita Filho"

Faculdade de Ciências

Departamento de Ciência da Computação

Bauru, 17 de Novembro de 2023.

Agradecimentos

Agradeço, primeiramente, à Deus por me dar sua Graça, de poder amar e ser amado por Ele, e por abençoar minha vida.

Agradeço à toda minha família, meu pai Marcos, minha mãe Mara e meus irmãos Letícia e Leonardo por me amarem e me apoiarem em todos os momentos e me possibilitarem estar onde estou hoje

Agradeço à minha namorada e amor da minha vida Mel por sempre estar ao meu lado desde que nos conhecemos. Por me amar e me apoiar em tudo, por ser minha companheira e aliviar o estresse do dia a dia.

Agradeço ao meu orientador Prof. Dr. Higor Amario de Souza pela confiança nesse trabalho e por me orientar, e por todo auxílio e conhecimento passado durante o desenvolvimento dele.

Agradeço à todos os meus amigos que estiveram comigo nessa caminhada computacional em Bauru, e que foram fundamentais para mim. Meu amigo de apartamento e de vida, Gabriel Souza, e meus amigos mojo dojo Arthur, João Pedro, Nathan, Renato e Ronaldo.

Por fim, agradeço ao meu gato bípede Kadu por ser meu parceiro de madrugadas e ser o gato mais fofo do mundo.

"A ciência é mais que um corpo de conhecimento, é uma forma de pensar, uma forma cética de interrogar o universo, com pleno conhecimento da falibilidade humana."

Carl Sagan

Resumo

O presente trabalho tem como proposta aplicar técnicas de ciência de dados com *python* para analisar e visualizar dados tabulares e geolocalizados com o objetivo de entender o impacto do cenário atual da segurança viária sobre a mobilidade ativa na cidade de São Paulo. As análises buscam entender o panorama referente a quantidade de acidentes envolvendo ciclistas e pedestres, a distribuição desses, além de características dos acidentes e acidentados. Visa-se também entender a influência da segurança das infraestruturas cicloviárias sobre os acidentes e o gasto gerado ao sistema de saúde público para atendimento das vítimas.

Palavras-chave: Ciência de dados. Python. Segurança viária. Mobilidade ativa. Infraestrutura Cicloviária.

Abstract

The present work aims to apply data science techniques using python to analyze and visualize tabular and geolocated data with the goal of understanding the impact of the current road safety scenario on active mobility in the city of São Paulo. The analyzes seek to understand the panorama regarding the number of accidents involving cyclists and pedestrians, their distribution, as well as the characteristics of the accidents and victims. The aim is also to understand the influence of the safety on different cycling infrastructures over accidents and the expense generated by the public health system to care for victims.

Keywords: Data Science. Python. Road Safety. Active Mobility. Cycling Infrastructure.

Lista de figuras

Figura 1 – Ciclovía	17
Figura 2 – Ciclofaixa	17
Figura 3 – Ciclorrota	18
Figura 4 – Mapa Coroplético	20
Figura 5 – Mapa de Calor	21
Figura 6 – Fluxo de trabalho	24
Figura 7 – Importando as bases	27
Figura 8 – Amostra do <i>dataframe</i> de acidentes não fatais	28
Figura 9 – Amostra do <i>dataframe</i> de óbitos	28
Figura 10 – Amostra do <i>dataframe</i> de Custo com AIH	29
Figura 11 – Amostra do <i>dataframe</i> de Ciclovias e Ciclofaixas	30
Figura 12 – Amostra do <i>dataframe</i> de Ciclorrotas	30
Figura 13 – Amostra do <i>dataframe</i> de Distritos	31
Figura 14 – Amostra do <i>dataframe</i> de População por Distrito	31
Figura 15 – Aplicação de filtros	32
Figura 16 – Remoção de colunas desnecessárias	32
Figura 17 – Removendo coordenadas faltantes	33
Figura 18 – Substituindo valores nulos	33
Figura 19 – Abordagem para Análise do Problema Proposto	34
Figura 20 – Infraestrutura Ciclovária	38
Figura 21 – Infraestrutura Ciclovária Ampliada	38
Figura 22 – Cobertura Ciclovária	39
Figura 23 – População por distrito	40
Figura 24 – Total de acidentes com ciclistas por distrito	41
Figura 25 – Total de acidentes com pedestres por distrito	42
Figura 26 – Correlação pareada entre os valores de acidentes	42
Figura 27 – Acidentes com ciclistas para cada 1000 habitantes por distrito	43
Figura 28 – Acidentes com pedestres para cada 1000 habitantes por distrito	44
Figura 29 – Acidentes com ciclistas x Hora do dia	45
Figura 30 – Destino das viagens	45
Figura 31 – Acidentes com ciclistas x Mês	46
Figura 32 – Acidentes com ciclistas x Veículo envolvido	47
Figura 33 – Óbitos x Faixa Etária	48
Figura 34 – Idade dos ciclistas de São Paulo	48
Figura 35 – Óbitos x Sexo	49
Figura 36 – Principais melhorias para ciclistas	50

Figura 37 – Acidentes dentro e fora das infraestruturas cicloviárias - Amostra	51
Figura 38 – Acidentes dentro e fora das infraestruturas cicloviárias - Geral	52
Figura 39 – Acidentes dentro e fora das infraestruturas cicloviárias por km	53
Figura 40 – Índice de proporção entre os acidentes dentro e fora das infraestruturas	54
Figura 41 – Fluxo de Retro Alimentação para investimento em infraestrutura cicloviária	56

Lista de tabelas

Tabela 1 – Impactos do aumento do transporte individual e motorizado	15
Tabela 2 – Custo de Sinalização de Ciclovias e Ciclofaixas	54
Tabela 3 – Custo com AIHs de Acidentes com Ciclistas nos últimos 10 anos	55

Lista de abreviaturas e siglas

AIH	Autorização de Internação Hospitalar
CET	Companhia de Engenharia de Tráfego
SUS	Sistema Único de Saúde

Sumário

1	INTRODUÇÃO	15
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	Infraestrutura Ciclovária	17
2.2	Dados sobre saúde e acidentes	18
2.2.1	SUS	18
2.2.2	DATASUS	19
2.2.3	AIH	19
2.2.4	Valor total - AIH	19
2.2.5	Valor médio AIH	19
2.3	Ciência e Análise de Dados	19
2.3.1	Análise Geoespacial	20
2.3.2	Mapa Coroplético	20
2.3.3	Mapa de Calor - <i>Heatmap</i>	20
2.3.4	Análise Descritiva	21
2.3.5	Análise Diagnóstica	21
2.4	Ferramentas de análise de ciência de dados	22
2.4.1	Python	22
2.4.2	Jupyter Notebook	22
2.4.3	Pandas	22
2.4.4	Numpy	22
2.4.5	Matplotlib	22
2.4.6	Seaborn	23
2.4.7	Folium	23
2.4.8	Geopandas	23
2.4.9	ipywidgets	23
3	METODOLOGIA	24
3.1	Fluxo de trabalho	24
3.2	Conjunto de Dados - Extração	25
3.2.1	Dados de Acidentes	25
3.2.2	Dados de Custo de AIHs dos acidentes	25
3.2.3	Dados Geoespaciais - Infraestrutura Ciclovária	25
3.2.4	Dados de Custo de Sinalização de Ciclovia/Ciclofaixa	26
3.2.5	Dados Geoespaciais - Distritos do Município de São Paulo	26
3.2.6	População residente - Município de São Paulo	26

3.3	Conjunto de Dados - Importação	26
3.4	Características dos Dados	27
3.4.1	Acidentes Não Fatais	27
3.4.2	Óbitos	28
3.4.3	Dados de Custo de Sinalização de Ciclovia/Ciclofaixa	28
3.4.4	Dados de Custo de AIHs dos acidentes	29
3.4.5	Dados Geoespaciais - Infraestrutura Cicloviária	29
3.4.6	Dados Geoespaciais - Distritos do Município de São Paulo	30
3.4.7	População residente - Município de São Paulo	31
3.5	Tratamento, Limpeza e Manipulação dos Dados	32
3.5.1	Aplicação de filtros	32
3.5.2	Remoção de colunas desnecessárias	32
3.5.3	Dados Faltantes	33
3.5.4	Tratamento de Inconsistências	33
3.6	Criação das Análises	34
3.6.1	Análises Geoespaciais	34
3.6.2	Análises das Características	35
3.6.3	Análise da Segurança Viária	35
3.7	Disponibilização das Análises e Resultados	36
4	VISUALIZAÇÃO E DISCUSSÃO DOS RESULTADOS	37
4.1	Panorama Geral da Infraestrutura Cicloviária	37
4.2	Panorama Geral dos Acidentes	39
4.3	Características dos acidentes e acidentados	44
4.4	Segurança Viária na Infraestrutura Cicloviária	49
4.5	Custos de Sinalização de Ciclovia x Custos de AIHs de Acidentes	54
4.6	Considerações	55
5	CONCLUSÃO	57
5.1	Trabalhos Futuros	57
	REFERÊNCIAS	58
	APÊNDICE A – INFRAESTRUTURA CICLOVIÁRIA	60
	APÊNDICE B – COBERTURA CICLOVIÁRIA	61
	APÊNDICE C – JUNÇÃO DOS DADOS	62
	APÊNDICE D – VISUALIZAÇÃO DOS DADOS - MAPA CORO- PLÉTICO	63

APÊNDICE E – MAPA DE CALOR - ACIDENTES	65
APÊNDICE F – HORA E MÊS DOS ACIDENTES	66
APÊNDICE G – VEÍCULO ENVOLVIDO NOS ACIDENTES	67
APÊNDICE H – CRIAÇÃO DOS BUFFERS	69

1 Introdução

Um dos principais pontos observáveis no processo rápido e descontrolado de crescimento e urbanização da maioria das cidades pelo mundo é o intenso uso de transporte motorizado individual juntamente com a fraca ou inexistente presença de um planejamento urbano que leva em consideração a mobilidade urbana. Como consequência desse crescimento caótico e desorganizado, observam-se inúmeros pontos prejudiciais à população em diversos aspectos, sendo os principais, econômico, ambiental e social. Em contrapartida, a mobilidade ativa consiste na prática de se utilizar modais não-motorizados para se locomover, como bicicleta e a pé, e traz ganhos enormes à sociedade que combatem os danos causados pelos meios motorizados e individuais, indicados na Tabela 1 abaixo.

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 Habitações
Acidentes	Impacto sobre a saúde humana	Impactos Hidrológicos
Custos com Infraestrutura	Interação na Comunidade	
Custos para consumidores	Habitabilidade da Comunidade	
	Estética	

Fonte: Litman e Burwell (2006)

A capacidade de conectar a população a outros meios de transporte, recursos e oportunidades de forma eficiente, além de reduzir a emissão de gases de efeito estufa e contribuir para saúde de quem adere a esta prática, está fortemente ligada a criação de redes cicloviárias e à pé. O transporte ativo possui enorme potencial de exploração, visto que, em média nas grandes áreas urbanas, 35% das viagens feitas levam menos de 3 quilômetros e 50% levam menos de 10km o que faz com que sejam facilmente alcançadas pedalando e/ou caminhando (ADRIAZOLA-STEIL et al., 2021).

No entanto, estimular a mobilidade ativa requer cautela e planejamento, pois pedestres e ciclistas são os mais vulneráveis a acidentes de trânsito (PEDEN, 2004). Segundo WELLE et al. (2018), a cada ano, cerca de 1,25 milhão de pessoas morrem nas ruas das cidades, estradas e rodovias, levando acidentes no trânsito a ocupar a 10ª principal causa de morte no mundo. Relatório da OMS sobre a segurança do trânsito em 2018 aponta que pedestres e ciclistas representam 26% de todas as mortes no trânsito, número que chega a 44% no continente africano. O perigo devido à falta de segurança viária nas diferentes pistas, além de ser um dos principais fatores que causam mortes, também pode ser considerado o principal empecilho para a utilização de bicicletas em vias compartilhadas com transporte motorizado, conforme Fund

(2011). Dessa forma, a segurança viária é imprescindível para reter a parte da população já praticante da mobilidade ativa bem como para atrair novas pessoas à ela.

Neste trabalho, o objetivo foi usar técnicas de ciência de dados para analisar o impacto da segurança viária na mobilidade ativa. Entender quais são as características de locais com mais acidentes e os fluxos de ciclistas e pedestres pode colaborar para a proposição de políticas públicas para melhorar o sistema viário com base em evidências científicas.

Pretendeu-se, também, verificar a relação entre a presença de infraestrutura cicloviária e a quantidade de acidentes no seu entorno. Assim, foi possível investigar se os investimentos em melhoria da infraestrutura cicloviária podem reduzir os gastos de saúde relacionados com o atendimento das ocorrências de acidentes.

Para atingir os objetivos deste trabalho, foi realizado um levantamento bibliográfico sobre mobilidade ativa, segurança viária e técnicas de análise de dados, e um estudo as bibliotecas de ciência de dados da linguagem *Python*, tais como *Pandas*, *Geopandas*, *Matplotlib* e *Folium*. Em seguida, foi feita a coleta de dados públicos abertos do Sistema Único de Saúde, dados de acidentes de trânsito e dados da infraestrutura cicloviária do CET. Depois os dados foram processados e analisados por meio de análise exploratória usando técnicas de ciência de dados. Finalmente, foi feita a apresentação e a análise dos resultados obtidos.

Os resultados obtidos indicaram que há uma maior concentração de acidentes envolvendo ciclistas e pedestres por habitante nos distritos centrais e em parte da zona sul da capital paulista. Foi possível notar também que, de forma geral, os acidentes são maiores em horários de pico e os acidentados são em maioria homens e pessoas com menos de 50 anos. Finalmente, observou-se que as ciclovias garantem maior proteção, ao contrário de ciclorrotas; e, com o custo de internações aprovadas nos últimos 10 anos para acidentes com ciclistas, poderiam ser implementadas aproximadamente 2.696,02 km de ciclovias e 3.316,9 km de ciclofaixas. Outra informação importante descoberta com os resultados é que o custo de construção de 1 km de ciclovia equivale ao custo de internação de 222 ciclistas que sofreram algum acidente.

2 Fundamentação Teórica

2.1 Infraestrutura Ciclovitária

A Companhia de Engenharia de Tráfego (CET) da cidade de São Paulo define a Ciclovía como sendo “pista de uso exclusivo de bicicletas e outros ciclos, com segregação física do tráfego comum”, ou seja, há literalmente uma barreira física que segrega as bicicletas de modos motorizados. Sua estrutura pode ser vista na Figura 1

Figura 1 – Ciclovía



Fonte: Prefeitura de Maringá

A Ciclofaixa é uma faixa demarcada nas vias de tráfego de veículos motorizados, reservada para o uso de bicicletas. Sua principal diferença para a ciclovía está na ausência de segregação física entre as bicicletas e os veículos motorizados. A Figura 2 indica sua estrutura.

Figura 2 – Ciclofaixa



Fonte: Prefeitura Municipal de Curitiba

Já a ciclorrota - visualizada na Figura 3 - representa uma rota planejada para ciclistas

que utiliza vias urbanas e estradas existentes, mas não possui infraestrutura física específica, como ciclovia ou ciclofaixa. Ela é marcada com sinalização específica para orientar os ciclistas ao longo do percurso.

Figura 3 – Ciclorrota



Fonte: Prefeitura de Vitória

Do ponto de vista de segurança, a ciclovia apresenta maior segurança devido ao fato de estar separada fisicamente das pistas onde circula o tráfego motorizado, através de elementos como meio-fio, grade, canteiro, área verde e outros previstos na legislação. As ciclofaixas são um pouco menos seguras, visto que não estão segregadas fisicamente das vias, sendo somente uma faixa delimitada com pintura e sinalizadores. Já as ciclorrotas possuem pouca ou nenhuma segurança visto que compartilham as vias com os modos motorizados.

Já a cobertura ciclovária é uma métrica utilizada neste trabalho para indicar quantos metros de infraestrutura ciclovária existem para cada km² de área de determinado distrito da cidade de São Paulo. Será utilizado neste trabalho para entender a situação da malha ciclovária na capital paulista.

2.2 Dados sobre saúde e acidentes

2.2.1 SUS

O SUS, ou Sistema Único de Saúde, é, segundo o Ministério da Saúde, um dos sistemas de saúde pública mais extensos e intrincados globalmente, compreendendo serviços que variam desde consultas de rotina para medição da pressão arterial na Atenção Primária até a realização de transplantes de órgãos, assegurando que todos os cidadãos tenham acesso abrangente, universal e gratuito aos cuidados de saúde em todo o país.

2.2.2 DATASUS

O DATASUS representa o setor de tecnologia de informação do Sistema Único de Saúde brasileiro. É uma entidade vinculada à Secretaria de Gestão Estratégica e Participativa do Ministério da Saúde, encarregada de reunir, analisar e distribuir dados relativos à área da saúde. O DATASUS é responsável por disponibilizar dados de diferentes sistemas como SIH (Sistema de Informações Hospitalares) e SIA (Sistema de Informações Ambulatoriais), além de aplicativos como o TABNET que permite organizar dados de forma rápida, conforme a consulta que se deseja tabular.

2.2.3 AIH

A AIH - Autorização de Internação Hospitalar tem o objetivo de transcrever todos os atendimentos que provenientes de internações hospitalares que foram financiadas pelo SUS. Tal conceito é relevante a esse trabalho, pois serão levados em consideração somente gastos com internações que tiveram suas AIHs aprovadas.

2.2.4 Valor total - AIH

O Valor total, no sistema de informações hospitalares do SUS, corresponde ao valor gasto com AIH aprovadas no período escolhido. No entanto, conforme na nota técnica do SIH/SUS, este valor não obrigatoriamente corresponde ao valor repassado ao estabelecimento, pois, dependendo da situação das unidades, estes recebem recursos orçamentários ou pode haver retenções e pagamentos de incentivos, não aqui apresentados. Portanto, este valor deve ser considerado como o valor aprovado da produção.

2.2.5 Valor médio AIH

O valor médio AIH é simplesmente o Valor total, mencionado acima, dividido pela quantidade de AIH aprovada no período. Será uma métrica importante para análise dos gastos com segurança viária comparados aos de saúde.

2.3 Ciência e Análise de Dados

Segundo Provost e Fawcett (2013), a ciência de dados é, de maneira geral, um conjunto de princípios fundamentais que apoiam e orientam a extração de informações a partir de dados. Através desses princípios, processos e técnicas, é possível aplicar a análise de dados para compreender os fenômenos estudados. 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 da mobilidade ativa na cidade de São Paulo.

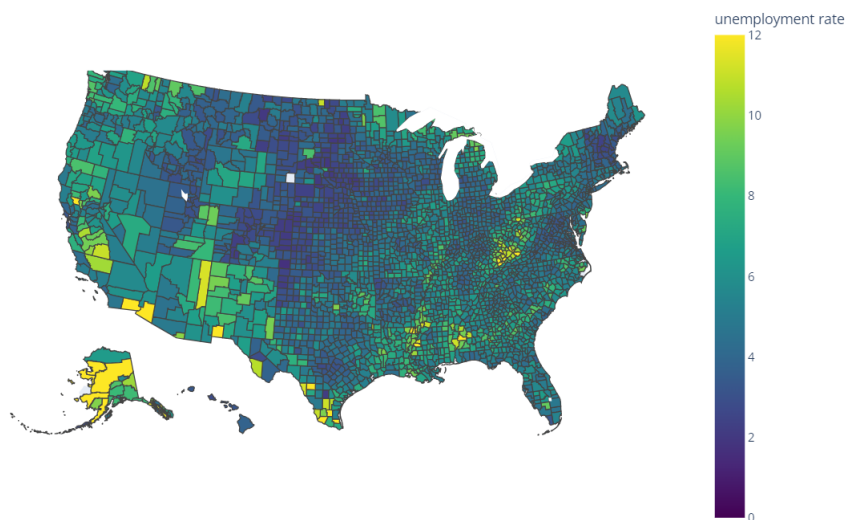
2.3.1 Análise Geoespacial

Segundo Câmara et al. (2004), "A ênfase da Análise Espacial é mensurar propriedades e relacionamentos, levando em conta a localização espacial do fenômeno em estudo de forma explícita". Os dados geoespaciais têm um grande número de aplicações no cotidiano, indo desde o entendimento e modelagem do comportamento urbano de pessoas, veículos e outros objetos móveis. Esse tipo de análise é sustentada, principalmente por visualizações em formas de mapa. Alguns utilizados nesse trabalho foram os citados abaixo.

2.3.2 Mapa Coroplético

Um mapa coroplético¹ tipicamente representa uma superfície estatística usando áreas que são simbolizadas com cores, sombreamentos ou padrões de acordo com uma escala que reflete a relação proporcional da variável estatística em questão, como, por exemplo, a densidade populacional ou a renda per capita. Os símbolos originados desse método coincidem com as regiões onde os dados foram coletados, criando a impressão de que os dados são uniformes dentro de cada região e que as mudanças ocorrem sempre nos limites dessas áreas. Um exemplo pode ser visto na Figura 4

Figura 4 – Mapa Coroplético



Fonte: <https://plotly.com/python/choropleth-maps/>

2.3.3 Mapa de Calor - *Heatmap*

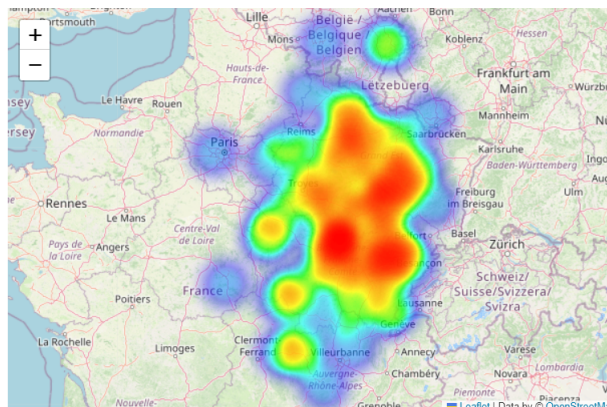
Um mapa de calor² - visto na Figura 5 - é um método de representação visual de dados que apresenta a magnitude de um evento utilizando cores em duas dimensões. A mudança

¹ <https://plotly.com/python/choropleth-maps/>

² https://python-visualization.github.io/folium/latest/user_guide/plugins/heatmap.html

na cor pode ocorrer devido à diferença no tom ou na intensidade, proporcionando indicações visuais claras ao observador sobre a distribuição ou variação espacial do fenômeno.

Figura 5 – Mapa de Calor



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

2.3.4 Análise Descritiva

Segundo Reis e Reis (2002), a etapa inicial de investigação dos dados coletados é conhecida como Análise Descritiva. Nesse estágio, emprega-se técnicas de Estatística Descritiva para estruturar, condensar e narrar os elementos cruciais presentes em um conjunto de características observadas, ou para realizar comparações entre essas características em dois ou mais conjuntos distintos. Ou seja, é uma técnica que visa identificar o que ocorreu baseados em fatos que são encontrados nos dados.

O presente trabalho visa utilizar dessa análise para descrever a situação dos acidentes envolvendo usuários de mobilidade ativa, e da segurança viária através de gráficos de barras e mapas, e medidas de síntese como porcentagens, índices e médias.

2.3.5 Análise Diagnóstica

A análise diagnóstica é definida como uma técnica de análise que visa entender por que algo aconteceu. No caso desse trabalho, é um método importante a ser aplicado para que se possa entender por que os acidentes ocorrem mais em certas infraestruturas cicloviárias. Não é evidente encontrar soluções para esse tipo de indagação, e, portanto, para identificar as potenciais razões, é interessante conduzir testes.

2.4 Ferramentas de análise de ciência de dados

2.4.1 Python

Segundo sua própria documentação, o *Python* é uma linguagem orientada a objeto, fácil de usar e com inúmeras poderosas bibliotecas que auxiliam nas mais diversas áreas de programação. Foi escolhida para este trabalho devido à facilidade de se trabalhar com dados geoespaciais e geo-localizados - tipos de dados predominantes para as análises necessárias.

2.4.2 Jupyter Notebook

Jupyter é um ambiente interativo e uma aplicação web que permite a criação e compartilhamento de *notebooks* - documentos interativos que podem conter código executável, texto explicativo, equações, gráficos e visualizações. É uma plataforma flexível para desenvolvimento, documentação e comunicação de projetos que envolvem código e análise de dados.

2.4.3 Pandas

O *Pandas* é uma biblioteca de *Python* amplamente utilizada para manipulação e análise de dados. Ele fornece estruturas de dados flexíveis e eficientes, bem como funções poderosas para trabalhar com dados tabulares. Facilita o entendimento, tratamento e limpeza dos dados com que se está trabalhando. Possui uma documentação bastante robusta, porém intuitiva.

2.4.4 Numpy

O *NumPy* é descrito em sua documentação como sendo a biblioteca fundamental em *Python* para computação científica e análise de dados. A biblioteca trabalha com o conceito de *arrays* multidimensionais - estruturas de dados eficientes para armazenar e manipular dados - e eficiência computacional, facilitando assim o trabalho com grandes *datasets*. Oferece recursos avançados para indexação e seleção de elementos em *arrays*, e inclui uma ampla variedade de funções matemáticas para realizar operações, facilitando a manipulação e análise de dados.

2.4.5 Matplotlib

O *Matplotlib* é uma biblioteca de *Python* amplamente usada para criação de gráficos e visualizações de dados. É uma biblioteca completa e fornece inúmeros tipos de visualização como gráficos de barra, linha, pizza e etc. Além disso, permite estilizar os elementos dos gráficos da maneira que achar melhor, cores, rótulos, tamanhos e é facilmente integrado à outras bibliotecas como *Pandas* e *Numpy*.

2.4.6 Seaborn

O *Seaborn* também é uma biblioteca para visualização de dados e foi construída sobre o *Matplotlib*. Suas principais características que auxiliam no desenvolvimento deste trabalho são a facilidade de uso com relação ao *Matplotlib*, paletas de cores agradáveis e padronizadas com destaque à paleta que favorece daltônicos a visualizarem melhor os gráficos.

2.4.7 Folium

O *Folium* é uma biblioteca *Python* construída sobre a biblioteca *JavaScript Leaflet* que simplifica a criação de mapas interativos que podem ser explorados em páginas web ou diretamente em notebooks como o *Jupyter*. Permite integrar dados geoespaciais, como *GeoJSON* ou *Shapefiles*, nos mapas criados com o *Folium*, adicionar camadas de mapa personalizadas, como imagens de satélite ou mapas personalizados, para enriquecer a visualização e representar dados em mapas por meio de cores, tamanhos de marcadores ou gráficos de calor para comunicar informações geoespaciais. Em geral, é uma ferramenta valiosa para a exploração de dados geográficos e a criação de visualizações de dados dinâmicas e interativas.

2.4.8 Geopandas

O *GeoPandas* é uma biblioteca de *Python* que estende as funcionalidades do *Pandas* para a manipulação de dados geoespaciais. Ela é projetada para simplificar a análise e a manipulação de dados geográficos, tornando mais fácil trabalhar com geometrias espaciais e atributos associados. Sua principal estrutura de dado é o *GeoDataFrame* - estrutura semelhante a um *DataFrame* do *Pandas*, mas inclui uma coluna especial que armazena geometrias espaciais, como pontos, linhas ou polígonos. Possibilita a manipulação de geometrias, projeção e transformação de coordenadas e análise espacial.

2.4.9 ipywidgets

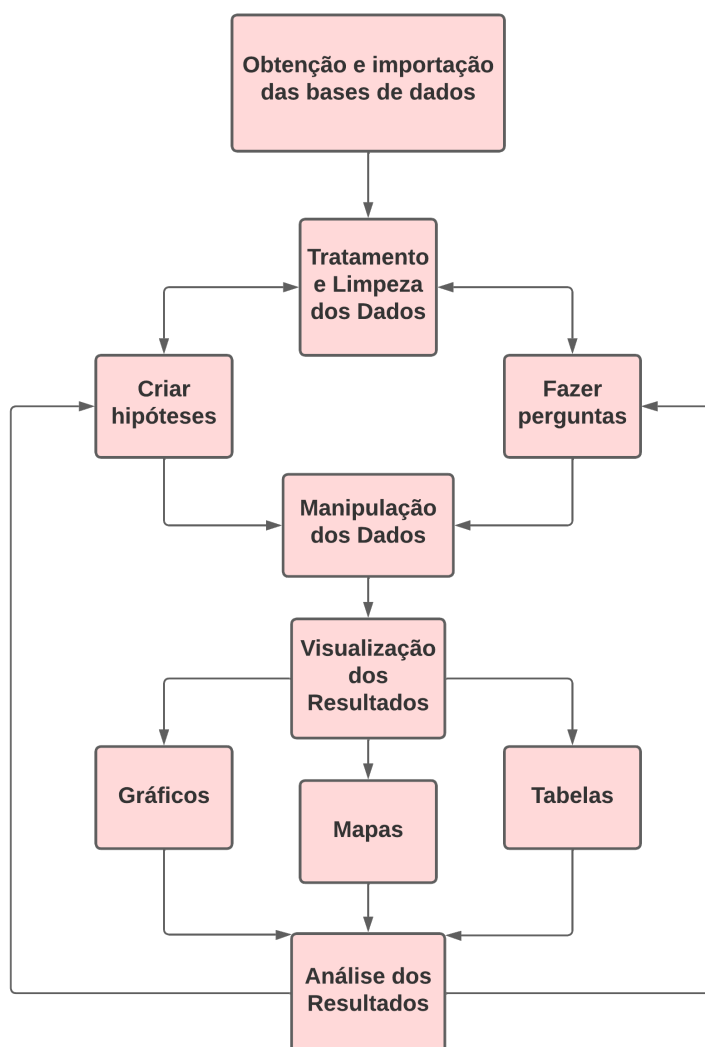
A biblioteca *ipywidgets* é uma parte do ecossistema *Jupyter*, que se concentra em criar *widgets* interativos para notebooks *Jupyter*, como *Jupyter Notebook* e *JupyterLab*. Ela é particularmente útil para criar interfaces de usuário interativas e dinâmicas para análises de dados, visualizações e demonstrações.

3 Metodologia

3.1 Fluxo de trabalho

Para a elaboração desse trabalho, foi necessário seguir etapas comuns à análise de dados. Para melhor entendimento da estrutura seguida, criou-se um diagrama com o fluxo de trabalho, visto na Figura 6

Figura 6 – Fluxo de trabalho



Fonte: Elaborada pelo Autor

As primeiras etapas descrevem a obtenção das bases de dados e sua importação para o ambiente onde serão feitas as análises. Em seguida, o tratamento e limpeza adequado dos

dados deve ser feito para que se possa começar a levantar hipóteses e fazer perguntas que guiarão as análises. Feito isso, deve-se manipular os dados de forma a obter as informações que se deseja visualizar e analisar. Criadas as visualizações, são feitas as análises que podem levar a novas perguntas e hipóteses e assim se segue, sucessivamente, até que se obtenha todas as informações necessárias.

3.2 Conjunto de Dados - Extração

3.2.1 Dados de Acidentes

A obtenção dos dados de acidentes foi feita através da Infosiga SP - Sistema de Informações Gerenciais de Acidentes de Trânsito do Estado de São Paulo, disponível dentro da plataforma Governo Aberto SP¹ - uma iniciativa do Governo do Estado de São Paulo, de disponibilização, através da internet, de documentos, informações e dados governamentais de domínio público para a livre utilização pela sociedade. A base de dados possui atualização mensal e informações sobre perfil do acidente, da vítima e localização geográfica.

3.2.2 Dados de Custo de AIHs dos acidentes

A obtenção dos dados de custo de AIHs foi feita através do Tabnet do DATASUS fazendo a filtrações necessárias:

- Linha: Município
- Conteúdo: AIH aprovadas, Valor total e Valor médio
- Períodos disponíveis: Jan/2014-Ago/2023
- Grupo de Causas: V10-V19 Ciclista traumatizado acid transporte

3.2.3 Dados Geoespaciais - Infraestrutura Ciclovária

Os dados que descrevem a infraestrutura ciclovária de São Paulo foram retirados do site da CET - Companhia de Engenharia de Tráfego, responsável por todas as atividades do Departamento de Operação do Sistema Viário como regulação do tráfego, planejamento urbano e mobilidade, segurança viária e fiscalização e educação no trânsito. Foram obtidos os dados sobre ciclovias, ciclofaixas e ciclorrotas. A base pode ser obtida no site da CET².

¹ <<http://catalogo.governoaberto.sp.gov.br/dataset?q=INFOSIGA>>

² <<http://www.cetesp.com.br/consultas/bicicleta/mapa-de-infraestrutura-ciclovitaria.aspx>>

3.2.4 Dados de Custo de Sinalização de Ciclovias/Ciclofaixas

Os dados de custo de sinalização de ciclovias/ciclofaixas foram obtidos da Superintendência de Planejamento e Projetos da CET.

3.2.5 Dados Geoespaciais - Distritos do Município de São Paulo

Visto que a granularidade espacial da análise são os distritos da cidade de São Paulo, foi-se, necessário, obter os dados geoespaciais com essas informações. A obtenção foi feita através do Portal de Dados Abertos da Prefeitura de São Paulo - plataforma essa que reúne conjuntos de dados de todas as secretarias, subprefeituras e empresas públicas municipais. O conjunto de dados veio da Secretaria Municipal de Urbanismo e Licenciamento - SMUL. A base pode ser obtida através do portal citado³.

3.2.6 População residente - Município de São Paulo

O conjunto de dados sobre a população residente do município de São Paulo é necessário para algumas análises em que seja interessante olhar valores proporcionais ao número de habitantes de cada distrito do município. O *dataset* foi obtido através da Fundação SEADE (Sistema Estadual de Análise de Dados), fundação vinculada à Secretaria da Fazenda e Planejamento do Governo do Estado de São Paulo. A base pode ser obtida através do SEADE⁴.

3.3 Conjunto de Dados - Importação

Obtidos os dados, o próximo passo é a importação desses para o notebook *Jupyter* para entender como estão formados e estruturados. Utilizando as bibliotecas *Pandas* e *GeoPandas* é possível ler todas as bases e visualizá-las como *dataframes/geodataframes*, como mostra a Figura 7

³ <http://dados.prefeitura.sp.gov.br/pt_PT/dataset/distritos>

⁴ <<https://repositorio.seade.gov.br/dataset/populacao-residente-municipio-de-sao-paulo-evolucao>>

Figura 7 – Importando as bases

Importação dos Dados Geoespaciais - Infraestrutura Ciclovária e Distritos

```
In [4]: # Ciclovias e Ciclofaixa
gdf_infra_ciclovias = gpd.read_file("../dados/SP/shapesInfraestruturaCiclovias/Ciclovias.shp")

# Separando Ciclovias de ciclofaixas
gdf_ciclofaixa = gdf_infra_ciclovias[gdf_infra_ciclovias['programa'] < 'CICLOVIA']

gdf_ciclovias = pd.merge(gdf_infra_ciclovias, gdf_ciclofaixa, indicator=True, how='outer').query('

# Ciclorrota
gdf_ciclorrota = gpd.read_file("../dados/SP/shapesInfraestruturaCiclovias/Ciclorrotas.shp")
gdf_ciclorrota = gdf_ciclorrota.set_crs(epsg='4326')

# Distrito
gdf_distrito = gpd.read_file("../dados/SP/LAYER_DISTrito/DEINFO_DISTrito.shp")
```

Importação dos Dados de Acidentes e População

```
In [9]: # Acidentes Não Fatais
nao_fatais = pd.read_csv("../dados/SP/acidentes_naofatais.csv", delimiter=

# Óbitos
obitos_sp = pd.read_csv("../dados/SP/obitos_publico.csv", delimiter=';',

# População
df_populacional = pd.read_csv("../dados/SP/evolucao_msp_crescimento2000_!
```

Fonte: Elaborada pelo autor

3.4 Características dos Dados

3.4.1 Acidentes Não Fatais

O conjunto de dados que descrevem acidentes não fatais vem no formato csv e possui 770184 registros. Além disso, possui 50 colunas e as datas dos acidentes variam de 2019 à 2023. Destacam-se as colunas de diferentes granularidades temporais (hora, dia, mês e ano), Latitude e Longitude dos acidentes, veículos envolvidos (bicicleta, automóvel, caminhão, ônibus, motocicleta e pedestre) e o município em que ocorreu tal acidente - importante, pois deve-se filtrar apenas pelo município de São Paulo. Uma amostra pode ser vista através da Figura 8

Figura 8 – Amostra do *dataframe* de acidentes não fatais

Ano do Acidente	Ano/Mês do Acidente	Dia da Semana	Hora do Acidente	Turno	Município	...	Serviço de Atendimento - PRF	Serviço de Atendimento - Radio Patrulha	Serviço de Atendimento - CPTRAN	Tipo de Via
2023	2023.04	DOMINGO	21:45	NOITE	SAO VICENTE	...	NÃO	SIM	NÃO	Vias Municipais
2023	2023.04	DOMINGO	20:47	NOITE	ADAMANTINA	...	NÃO	SIM	NÃO	Rodovias
2023	2023.04	DOMINGO	21:54	NOITE	GUARULHOS	...	NÃO	SIM	NÃO	Vias Municipais
2023	2023.04	DOMINGO	22:11	NOITE	ITAPEVA	...	NÃO	SIM	NÃO	Vias Municipais
2023	2023.04	DOMINGO	22:14	NOITE	MONTE APRAZIVEL	...	NÃO	SIM	NÃO	NAO DISPONIVEL

Fonte: Elaborada pelo autor

3.4.2 Óbitos

Já o conjunto de dados que descrevem os óbitos é, de certa forma, similar ao *dataset* dos não fatais. Está no formato csv, mas possui menos registros: 46174 registros e 34 colunas. No entanto, o intervalo de tempo disponível das datas dos acidentes é maior, abrangendo de 2011 à 2023. Além de apresentar informações sobre o acidente, o conjunto também indica características da vítima, como sexo, faixa etária, tempo entre acidente e óbito e local do óbito, o que permite fazer uma análise do perfil dos óbitos. Latitude e Longitude dos acidentes e o município são outras características relevantes do *dataframe*, que pode ser vista através da Figura 9

Figura 9 – Amostra do *dataframe* de óbitos

Mês Óbito Descrição	Ano do Óbito	...	Tipo do veículo da vítima	Tipo de vítima	Local do óbito	Tipo de acidente	Sub Tipo do Acidente	Sexo	Faixa etária	Idade da vítima
ABRIL	2023	...	MOTOCICLETA	CONDUTOR	VIA	CHOQUE	ÁRVORE	MASCULINO	18-24	22
ABRIL	2023	...	MOTOCICLETA	CONDUTOR	ESTABELECIMENTO DE SAUDE	COLISAO	FRONTAL	MASCULINO	18-24	18
ABRIL	2023	...	MOTOCICLETA	CONDUTOR	ESTABELECIMENTO DE SAUDE	ATROPELAMENTO	ANIMAL	MASCULINO	40-44	42
ABRIL	2023	...	MOTOCICLETA	CONDUTOR	VIA	COLISAO	LATERAL	MASCULINO	35-39	35
ABRIL	2023	...	MOTOCICLETA	CONDUTOR	ESTABELECIMENTO DE SAUDE	OUTROS	TOMBAMENTO VEÍCULO DE 2 RODAS	MASCULINO	18-24	20

Fonte: Elaborada pelo autor

3.4.3 Dados de Custo de Sinalização de Ciclovia/Ciclofaixa

A base com os dados de custo de sinalização de infraestrutura cicloviária está estruturada como um arquivo .xlsx e descreve qual a estrutura cicloviária, sua tipologia, extensão em metros

custo total de sinalização horizontal e vertical, e custo por km. Para esse trabalho, foi ignorada a estrutura ciclovária "Viaduto Bresser" devido às particularidades de sua implementação que acabou por gerar um custo por km muito acima do normal.

3.4.4 Dados de Custo de AIHs dos acidentes

Os dados obtidos do DATASUS de AIHs foram baixados no formato .csv, mas não foi preciso exportá-los ao notebook *Jupyter*. A base possui 4 colunas: Município, AIH_aprovas, Valor_total e Valor_médio_AIH, e uma só linha que descreve o município de São Paulo. Uma amostra pode ser vista através da Figura 10

Figura 10 – Amostra do *dataframe* de Custo com AIH

	A	B	C	D	E	F	G
1	Morbidade Hospitalar do SUS por Causas Externas - por local de internação - São Paulo						
2	AIH aprovadasValor totalValor médio AIH por Município						
3	Município: 355030 SAO PAULO						
4	Período:Jan/2014-Ago/2023						
5	Município	AIH_aprovas	Valor_total	Valor_médio_AIH			
6	355030 SAO PAULO	598146	876844484,2	1465,94			
7	Total	598146	876844484,2	1465,94			
8	Fonte: Ministério da Saúde - Sistema de Informações Hospitalares do SUS (SIH/SUS)						
9	Notas:						
10							
11	Situação da base de dados nacional em 29/04/2016.						
12	Dados de janeiro de 2015 até março de 2016 sujeitos a retificação.						
13							
14	Consulte o site da Secretaria Estadual de Saúde para mais informações.						
15							

Fonte: Elaborada pelo autor

3.4.5 Dados Geoespaciais - Infraestrutura Ciclovária

Os conjuntos de dados de infraestrutura ciclovária estão estruturados no formato *shapefile* (.shp) - formato de dados usado pelas bases de dados geoespaciais e vetoriais em sistemas de informações geográficas. São duas as bases de dados:

- Conjunto com os dados de ciclovias e ciclofaixas;
- Conjunto com dados sobre as ciclorrotas.

O primeiro, apresenta 1829 registros que misturam dados de ciclovias e ciclofaixas, sendo necessário tratar futuramente para que haja um *dataset* para cada infraestrutura. Além disso, possui somente 5 colunas sendo elas o nome do programa da ciclovias/ciclofaixa, sua inauguração, extensão do trecho, extensão total e sua geometria (no caso, em formato de linhas). No conjunto, um programa pode ter um ou mais trechos que o compõe, o que indica que cada linha representa um trecho da ciclovias/ciclofaixa. A estrutura dos dados de ciclorrotas

segue exatamente a mesma estrutura descrita acima possuindo apenas 117 linhas descrevendo as ciclorrotas. Ambos as bases podem ser vistas através das Figuras 11 e 12

Figura 11 – Amostra do *dataframe* de Ciclovias e Ciclofaixas

	programa	inauguracao	extensao_t	extensao_c	geometry
0	CICLOFAIXA AFONSO LOPES VIEIRA	2016-07-15	791	1182	LINESTRING (-46.66743 -23.46776, -46.66739 -23...
1	CICLOFAIXA AFONSO LOPES VIEIRA	2016-07-15	716	1182	LINESTRING (-46.66743 -23.46776, -46.66743 -23...
2	CICLOFAIXA AFONSO LOPES VIEIRA	2016-07-15	31	1182	LINESTRING (-46.66759 -23.46799, -46.66743 -23...
3	CICLOFAIXA AFONSO LOPES VIEIRA	2016-07-15	5	1182	LINESTRING (-46.66291 -23.46304, -46.66287 -23...
4	CICLOFAIXA AFONSO LOPES VIEIRA	2016-07-15	219	1182	LINESTRING (-46.67001 -23.46803, -46.66999 -23...
...	...	...	...	...	...
1824	CICLOFAIXA FRANCISCO MESQUITA	2018-09-08	827	3483	LINESTRING (-46.57685 -23.60044, -46.57697 -23...
1825	CICLOVIA RIO PINHEIROS - MARGEM OESTE	2013-12-17	106	3286	LINESTRING (-46.72144 -23.65807, -46.72215 -23...
1826	CICLOVIA RIO PINHEIROS - MARGEM OESTE	2013-12-17	1408	3286	LINESTRING (-46.72215 -23.65738, -46.72160 -23...
1827	CICLOVIA RIO PINHEIROS - MARGEM OESTE	2013-12-17	1616	3286	LINESTRING (-46.71185 -23.66522, -46.71189 -23...
1828	CICLOVIA RIO PINHEIROS - MARGEM OESTE	2013-12-17	156	3286	LINESTRING (-46.72215 -23.65738, -46.72206 -23...

Fonte: Elaborada pelo autor

Figura 12 – Amostra do *dataframe* de Ciclorrotas

	programa	inauguracao	extensao_t	extensao_c	geometry
0	CICLORROTA BATURITÉ/ DIAMANTE	2021-05-18	83	548	LINESTRING (-46.63046 -23.57150, -46.63047 -23...
1	CICLORROTA BATURITÉ/ DIAMANTE	2021-05-18	465	548	LINESTRING (-46.63046 -23.57150, -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 -23.63340, -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 (-46.64437 -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...

Fonte: Elaborada pelo autor

3.4.6 Dados Geoespaciais - Distritos do Município de São Paulo

Assim como os dados de infraestrutura cicloviária, o arquivo que contém os dados dos distritos do município de São Paulo está no formato denominado *shapefile*. Está estruturado em 96 linhas x 10 colunas, porém as únicas colunas que importam são duas, nome do distrito (NOME_DIST no *dataset*) e geometria (*geometry*). Uma amostra pode ser vista através da Figura 13

Figura 13 – Amostra do *dataframe* de Distritos

	CLASSID	FEATID	REVISIONNU	NOME_DIST	SIGLA_DIST	COD_DIST	COD_SUB	DATA_CRIAC	USUARIO_ID	geometry
0	4.0	8583485.0	1.0	JOSE BONIFACIO	JBO	47	27	2007-03-19	0.0	POLYGON ((-46.44623 -23.55651, -46.44610 -23.5...
1	4.0	8583484.0	1.0	JD SAO LUIS	JDS	46	18	2007-03-19	0.0	POLYGON ((-46.75838 -23.64859, -46.75835 -23.6...
2	4.0	8583445.0	1.0	ARTUR ALVIM	AAL	5	21	2007-03-19	0.0	POLYGON ((-46.47502 -23.52381, -46.47501 -23.5...
3	4.0	8583479.0	1.0	JAGUARA	JAG	40	08	2007-03-19	0.0	POLYGON ((-46.75600 -23.49516, -46.75599 -23.4...
4	4.0	8583437.0	1.0	VILA PRUDENTE	VPR	93	29	2007-03-19	0.0	POLYGON ((-46.58148 -23.57298, -46.58145 -23.5...
...	...	...	...	...	...	...	...	...	...	...
91	4.0	8583433.0	1.0	VILA MARIA	VMR	89	07	2007-03-19	0.0	POLYGON ((-46.58508 -23.52832, -46.58519 -23.5...
92	4.0	8583434.0	1.0	VILA MARIANA	VMN	90	12	2007-03-19	0.0	POLYGON ((-46.65081 -23.59868, -46.65080 -23.5...
93	4.0	8583435.0	1.0	VILA MATILDE	VMT	91	21	2007-03-19	0.0	POLYGON ((-46.54827 -23.53505, -46.54842 -23.5...
94	4.0	8583412.0	1.0	SANTA CECILIA	SCE	69	09	2007-03-19	0.0	POLYGON ((-46.66362 -23.53701, -46.66355 -23.5...
95	4.0	8583409.0	1.0	REPUBLICA	REP	66	09	2007-03-19	0.0	POLYGON ((-46.64631 -23.53909, -46.64607 -23.5...

Fonte: Elaborada pelo autor

3.4.7 População residente - Município de São Paulo

Por fim, a base de população residente em São Paulo encontra-se estruturado em um arquivo csv tal qual os acidentes. A base apresenta 96 linhas x 13 colunas sendo duas colunas para código e nome dos distritos, 6 colunas para população de 2000 a 2050 e 5 colunas para a taxa de crescimento populacional entre cada ano. Para esse trabalho, serão relevantes somente as colunas do nome do distrito e a população em 2020, mas uma amostra completa pode ser vista através da Figura 14

Figura 14 – Amostra do *dataframe* de População por Distrito

	codigo_distrito	distrito	populacao_2000	populacao_2010	populacao_2020	populacao_2030	populacao_2040	populacao_2050
0	80001	Água Rasa	85.992	84.971	82.564	79.135	74.878	69.962
1	80002	Alto de Pinheiros	44.513	43.128	40.962	38.177	34.741	30.824
2	80003	Anhanguera	38.037	65.561	84.719	96.506	103.239	105.420
3	80004	Aricanduva	94.845	89.664	85.868	84.307	81.563	77.429
4	80005	Artur Alvim	111.294	105.317	100.462	96.215	91.697	85.947
...	...	...	...	...	...	...	...	...
91	80092	Vila Medeiros	140.725	130.005	123.456	119.872	116.437	112.464
92	80093	Vila Prudente	102.227	104.225	104.686	103.240	99.764	94.710
93	80094	Vila Sônia	87.354	108.247	121.663	128.398	130.892	129.807
94	80095	São Domingos	82.724	84.825	86.403	87.080	86.399	83.718
95	80096	Lajeado	157.316	164.451	174.539	185.221	193.982	199.280

Fonte: Elaborada pelo autor

3.5 Tratamento, Limpeza e Manipulação dos Dados

Após obtidos e importados os dados, segue-se com a etapa fundamental à análise de dados que é a transformação desses. Essas três etapas juntas são conhecidas como ELT (do inglês *Extract, Load & Transform*), e é na etapa de transformação que se aplicam as regras necessárias para agregar valor, ou seja, tornar dados agregados em informações valiosas para análises e tomadas de decisão (HÜBNER, 2020).

3.5.1 Aplicação de filtros

Para começar a transformar os dados, o primeiro passo foi aplicar filtros nos *dataframes* de acidentes para eliminar linhas que possuíam informações de municípios que não eram o de São Paulo e de vítimas que não eram nem ciclistas, nem pedestres.

Uma simples manipulação de *dataframes* resolve essa etapa, trazendo apenas dados que condizem com a localização e as vítimas necessárias, como pode ser visto na Figura 15.

Figura 15 – Aplicação de filtros

```
In [27]: sp_nao_fatais = nao_fatais[nao_fatais['Município'] == "SAO PAULO"]

In [32]: acid_ciclistas = coord_acidentes_nao_fatais[coord_acidentes_nao_fatais['Veículos Envolvidos - Bicicleta'] == 1]
acid_pedestres = coord_acidentes_nao_fatais[coord_acidentes_nao_fatais['Veículos Envolvidos - Pedestre'] == 1]
```

Fonte: Elaborada pelo autor

3.5.2 Remoção de colunas desnecessárias

Algumas colunas podem ser consideradas desnecessárias para a análise, isso porque elas não possuem informações relevantes a respeito do que se quer descobrir. Dessa forma, ao olhar para as bases, pode-se ter uma ideia de quais serão as colunas irrelevantes para esse trabalho, removê-las e obter *datasets* mais limpos e eficientes. Vale ressaltar, no entanto, que é preciso ter certeza absoluta de que as colunas a serem excluídas não farão falta — e nem sempre pode-se ter essa certeza para todas. Assim sendo, só foram removidas as que certamente não serão utilizadas, como exemplifica a Figura 16.

Figura 16 – Remoção de colunas desnecessárias

```
In [95]: coord_obitos_ciclista_em_sp = obitos_ciclista_em_sp.drop(['Tipo de via', 'Administração', 'Jurisdição',
                                                                    'Conservação', 'Região Administrativa', 'Logradouro',
                                                                    'Numeral / KM', 'Tipo do local da ocorrência',
                                                                    'Sub Tipo do Acidente'], axis=1)

coord_obitos_pedestres_em_sp = obitos_pedestres_em_sp.drop(['Tipo de via', 'Administração', 'Jurisdição',
                                                             'Conservação', 'Região Administrativa', 'Logradouro',
                                                             'Numeral / KM', 'Tipo do local da ocorrência',
                                                             'Sub Tipo do Acidente'], axis=1)
```

Fonte: Elaborada pelo autor

3.5.3 Dados Faltantes

Dados faltantes podem prejudicar a análise e outras etapas que dependem dela e do pré-processamento. Nesse trabalho, para que seja possível criar visualizações em forma de mapas, não podem haver coordenadas nulas no conjunto. Além disso, algumas métricas foram geradas a partir de divisões em que é possível haver denominador 0, e, neste caso, o *Python* converterá o resultado para nulo. Portanto, é necessário removê-los ou substituir esses valores por outros.

No caso das coordenadas, foi-se estabelecido que, se os nulos representassem até 10% do *dataset*, poderiam simplesmente ser removidos. Ao olharmos à quantidade de coordenadas nulas, obtemos 13018 de 179245 registros do conjunto de acidentes não fatais, o que representa menos de 10%, e, portanto, podemos deletá-los. O processo pode ser visto na Figura 17

Figura 17 – Removendo coordenadas faltantes

```
In [31]: coord_acidentes_ao_fatais = coord_acidentes_ao_fatais.dropna()
```

Fonte: Elaborada pelo autor

O outro caso de dados faltantes para tratar foi em uma métrica gerada para cálculo do “Índice de Proporção de Acidentes Perto e Longe da Infraestrutura Ciclovária” que será explicado detalhadamente no capítulo 4, mas que possibilita denominadores com valor 0. Para isso, devemos interpretar que o valor do índice é igual a 0, logo, deve-se substituir os nulos por esse valor, como indica a Figura 18

Figura 18 – Substituindo valores nulos

```
df_final_indice_ciclovias['ind_prop_acid_perto_longe_infra_ciclovitaria'] = (df_final_indice_ciclovias['Qtd_dentro'] /  
df_final_indice_ciclovias['Qtd_Fora'])  
df_final_indice_ciclovias = df_final_indice_ciclovias.fillna(0)
```

Fonte: Elaborada pelo autor

3.5.4 Tratamento de Inconsistências

Por fim, para deixar os dados prontos para serem manipulados de forma a gerar as análises necessárias, foi preciso tratar certas inconsistências encontradas nas bases de dados. Exemplo são as coordenadas nas bases de acidentes não fatais e de óbitos que estão no formato de graus decimais, mas utilizando ',' ao invés de '.' como separador, o que gera erros ao tentar utilizá-las para plotar mapas com o *Folium* e o *GeoPandas*.

3.6 Criação das Análises

Com os dados limpos e tratados, o próximo passo foi começar a criar as análises. As análises foram realizadas seguindo uma abordagem de afunilamento de detalhes. Isto é, iniciou-se gerando visualizações que possibilitam ter uma ideia do panorama atual do problema proposto, e fazer análises mais gerais, seguidos de um nível mais detalhado de análises, que permitem tirar conclusões mais efetivas. O infográfico na Figura 19 indica tal abordagem.

Figura 19 – Abordagem para Análise do Problema Proposto



Fonte: Elaborada pelo Autor

3.6.1 Análises Geoespaciais

Para a criação das análises Geoespaciais foram utilizadas as técnicas citadas no capítulo 2, além de gerar um mapa com toda a infraestrutura cicloviária do município de São Paulo. A implementação dos mapas coropléticos, de calor e com a infraestrutura cicloviária pode ser vista nos apêndices A a E. Basicamente para cada visualização coroplética, foi preciso fazer junções espaciais, ou *spatial joins*, entre o conjunto de dados de distritos com os de acidentes de forma a associar cada acidente ou óbito a um distrito baseado nas colunas de coordenadas e geometria, respectivamente. Já a criação do mapa de calor ocorre de maneira mais simples, sendo necessário apenas instanciar um *Heatmap()* da biblioteca *folium* e passar as coordenadas dos acidentes/óbitos como dados que o compõe. Foram adicionadas camadas com a infraestrutura cicloviária também de forma a deixar o mapa de calor mais completo. Tais análises constituem um importante passo na compreensão do quadro atual de distribuição de

acidentes e óbitos envolvendo ciclistas e pedestres na cidade de São Paulo, além da disposição de ciclovias, ciclofaixas e ciclorrotas pelo município e sua cobertura por distrito.

3.6.2 Análises das Características

As análises também focaram em apresentar as características dos acidentados e dos acidentes como sexo, faixa etária, veículos envolvidos, hora e mês dos acidentes, além de informações relevantes como o custo de sinalização de ciclovias e ciclofaixas comparado ao custo com as AIHs aprovadas relacionadas aos acidentes envolvendo ciclistas.

Para as análises que caracterizam apenas os acidentados, como sexo e faixa etária, foi utilizado apenas a base de dados de óbitos, visto que a de acidentes não fatais não possui tais colunas. Nesse caso, foi necessário apenas produzir gráficos de barras utilizando as bibliotecas do *matplotlib* e *seaborn* a partir do *dataset* de óbitos já tratado e trazendo apenas a coluna desejada.

Já para criação das análises temporais e por veículos envolvidos foi necessário juntar os conjuntos de acidentes não fatais com o de óbitos. No caso das análises de hora e mês dos acidentes, filtrou-se os óbitos de 2019 em diante para que se mantenha na mesma faixa temporal dos acidentes não fatais. Em seguida, criou-se um *dataframe* através da concatenação das colunas de data e hora do acidente de ambos os conjuntos - disponível no anexo F. A partir disso, foi preciso gerar um gráfico utilizando cada coluna desse *dataset* concatenado. Utilizou-se também o gráfico de barras. O mesmo raciocínio foi seguido para análise dos acidentes por veículo envolvido, com exceção de que foi preciso manipular a um nível mais aprofundado, utilizando agrupamento dos dados e depois a concatenação entre as bases, como mostra o Anexo G

Vale ressaltar que para todos os gráficos gerados, foi utilizado a paleta de cores disponível pela própria biblioteca do *seaborn* denominada *colorblind* que utiliza cores de forma a favorecer também pessoas portadoras de daltonismo.

3.6.3 Análise da Segurança Viária

Por fim, buscou-se compreender o porquê dos acidentes acontecerem onde acontecem, através de visualizações geoespaciais e gráficas. Para isso, foi feita uma análise da estrutura e segurança viária com cada infraestrutura cicloviária e seu impacto nos acidentes com ciclistas.

Para criar tal análise foram aplicados dois *buffers* sobre as ciclovias, ciclofaixas e ciclorrotas: um *buffer* menor que será utilizado para abranger acidentes que serão considerados dentro da infraestrutura cicloviária e um *buffer* maior que abrange toda a via em que está localizada aquela infraestrutura que será utilizada para calcular os acidentes que ocorreram fora da infraestrutura. Assim, obtêm-se duas medidas:

- Acidentes na infraestrutura cicloviária = total de acidentes dentro do *buffer* menor aplicado
- Acidentes fora da infraestrutura cicloviária = total de acidentes dentro do *buffer* maior aplicado - total de acidentes dentro do *buffer* menor

Para determinação dos *buffers* foi preciso converter coordenadas decimais em metros. Com a utilização de uma calculadora geográfica obteve-se que 0.000018 graus = aproximadamente 2 metros. A seguir, são apresentados os *buffers* aplicados para cada infraestrutura:

- Ciclovia e Ciclofaixa: 5m para o *buffer* menor e 20m para o *buffer* maior visto que ambas ficam localizadas em vias mais largas como vias arteriais e coletoras
- Ciclorrota: 5m para o *buffer* menor e 10m para o *buffer* maior visto que ciclorrotas ficam em vias menores como vias locais.

Após a determinação dos *buffers*, cuja implementação se encontra no apêndice H, obteve-se a quantidade dos acidentes dentro e fora da infraestrutura cicloviária seguindo a lógica apresentada acima e foi criado um gráfico de barras empilhadas para cada infraestrutura cicloviária para comparar tais medidas. Em seguida, criou-se um **índice de proporção entre os acidentes dentro e fora das infraestruturas** que poderia ser transformado em uma paleta de cores linear para colorir as ciclovias, ciclofaixas e ciclorrotas de forma a indicar quais são as que possuem mais acidentes dentro quando comparados a quantidade de acidentes fora.

Por fim, obteve-se as informações de custos para sinalização de ciclovias e ciclofaixas e dos custos médios e totais com AIHs aprovadas para ciclistas acidentados nos últimos 10 anos. O objetivo é, a partir dos resultados obtidos com as análises acima, entender quantos km de ciclovias ou ciclofaixas poderiam ser implementadas a partir dos gastos com acidentados e, assim, melhorar a segurança viária e reduzir as quantidades de acidentes, e, consequentemente reduzir os gastos com AIHs.

3.7 Disponibilização das Análises e Resultados

Todas as etapas seguidas e descritas aqui neste capítulo estão disponíveis em um repositório no github⁵ permitindo que outras pessoas possam fazer outras análises ou darem continuidade ao trabalho.

⁵ https://github.com/danilo-siqueira/meu_tcc

4 Visualização e Discussão dos Resultados

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

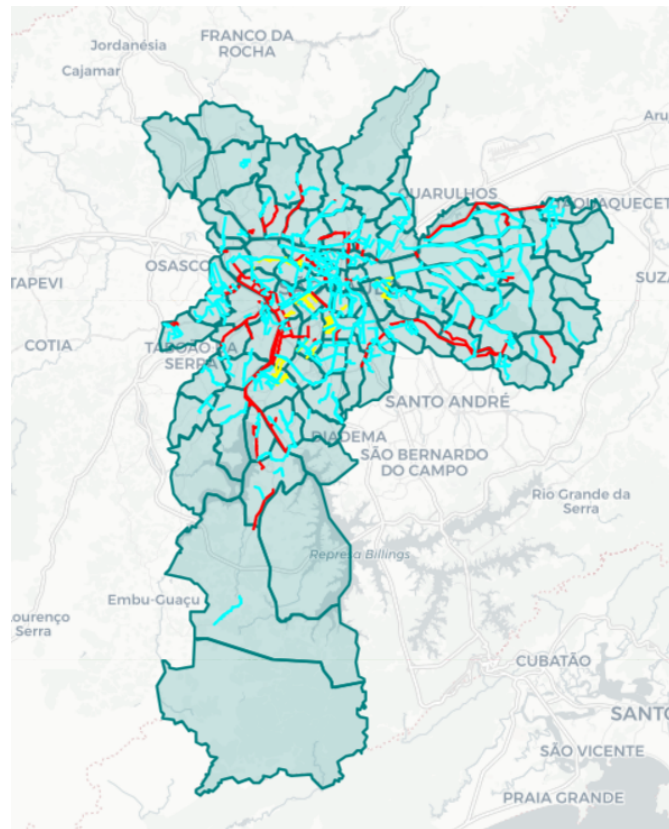
4.1 Panorama Geral da Infraestrutura Ciclovária

Inicialmente, é interessante ter uma ideia de como está distribuída a infraestrutura ciclovária. A cidade de São Paulo possui 722,1 km de vias com tratamento ciclovário permanente, sendo 690 km de Ciclovias/Ciclofaixas e 32,1 km de Ciclorrotas, de acordo com o site do CET¹. A Figura 20 indica visualmente essa distribuição, onde as ciclorrotas estão pintadas de amarelo, as ciclovias de vermelho e as ciclofaixas de azul claro. É possível notar uma concentração na faixa da região central da capital paulista.

A Figura 21 indica com mais detalhes a organização dessas infraestruturas e pode-se notar que as ciclofaixas fazem uma função de interligar as ciclovias e ciclorrotas, e os diferentes distritos e zonas de São Paulo.

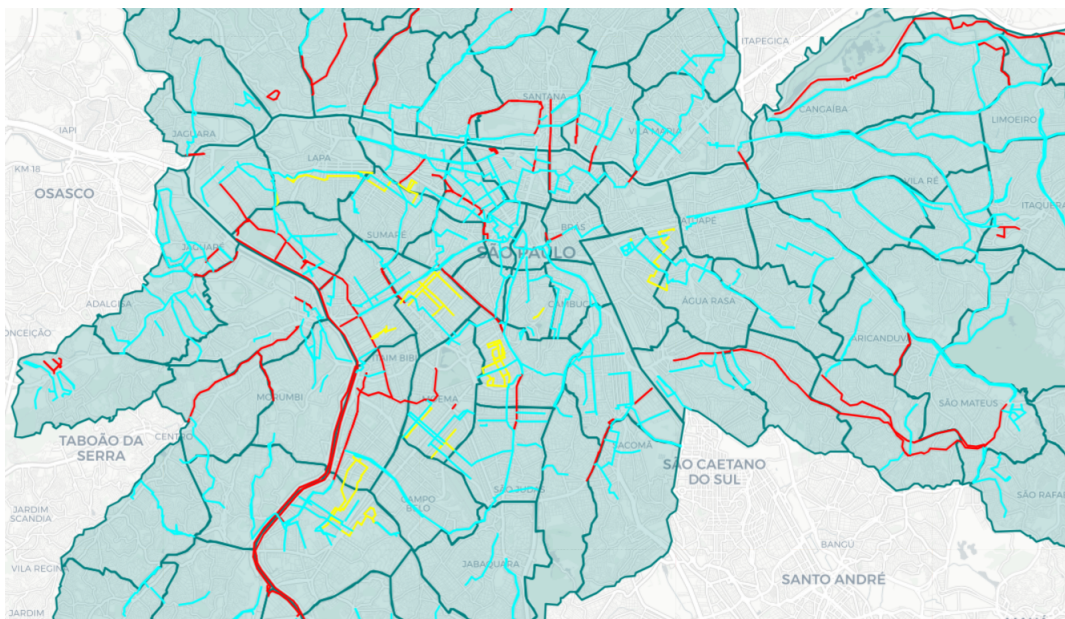
¹ <http://www.cetsp.com.br/consultas/bicicleta/mapa-de-infraestrutura-ciclovitaria.aspx>

Figura 20 – Infraestrutura Ciclovária



Fonte: Elaborada pelo autor

Figura 21 – Infraestrutura Ciclovária Ampliada

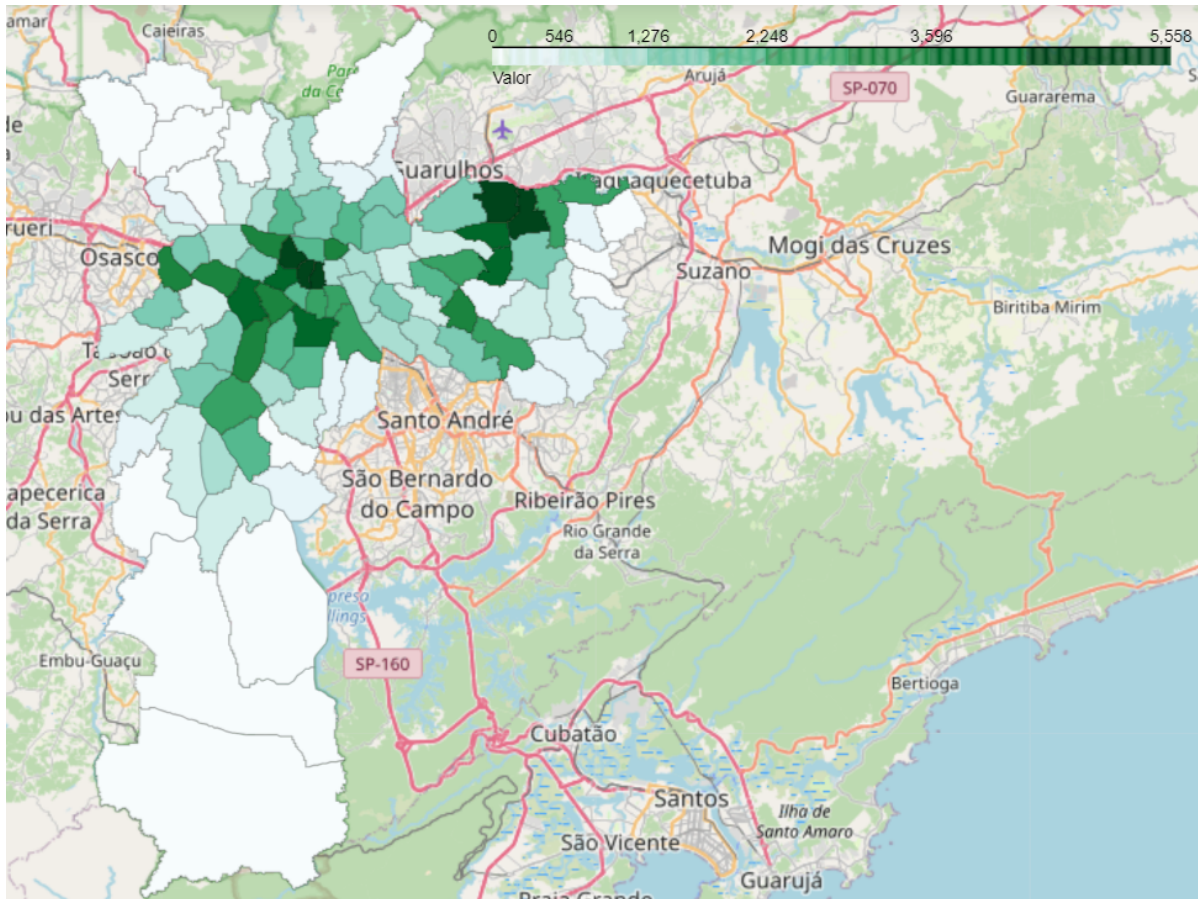


Fonte: Elaborada pelo autor

A visualização da cobertura ciclovária, na Figura 22, indica quantos metros de infraes-

estrutura ciclovária há para cada km² de determinado distrito. Esse mapa coroplético permite confirmar a maior concentração de ciclovias, ciclofaixas e ciclorrotas nos distritos centrais de São Paulo, com exceção de alguns distritos da Zona Leste e Sul.

Figura 22 – Cobertura Ciclovária

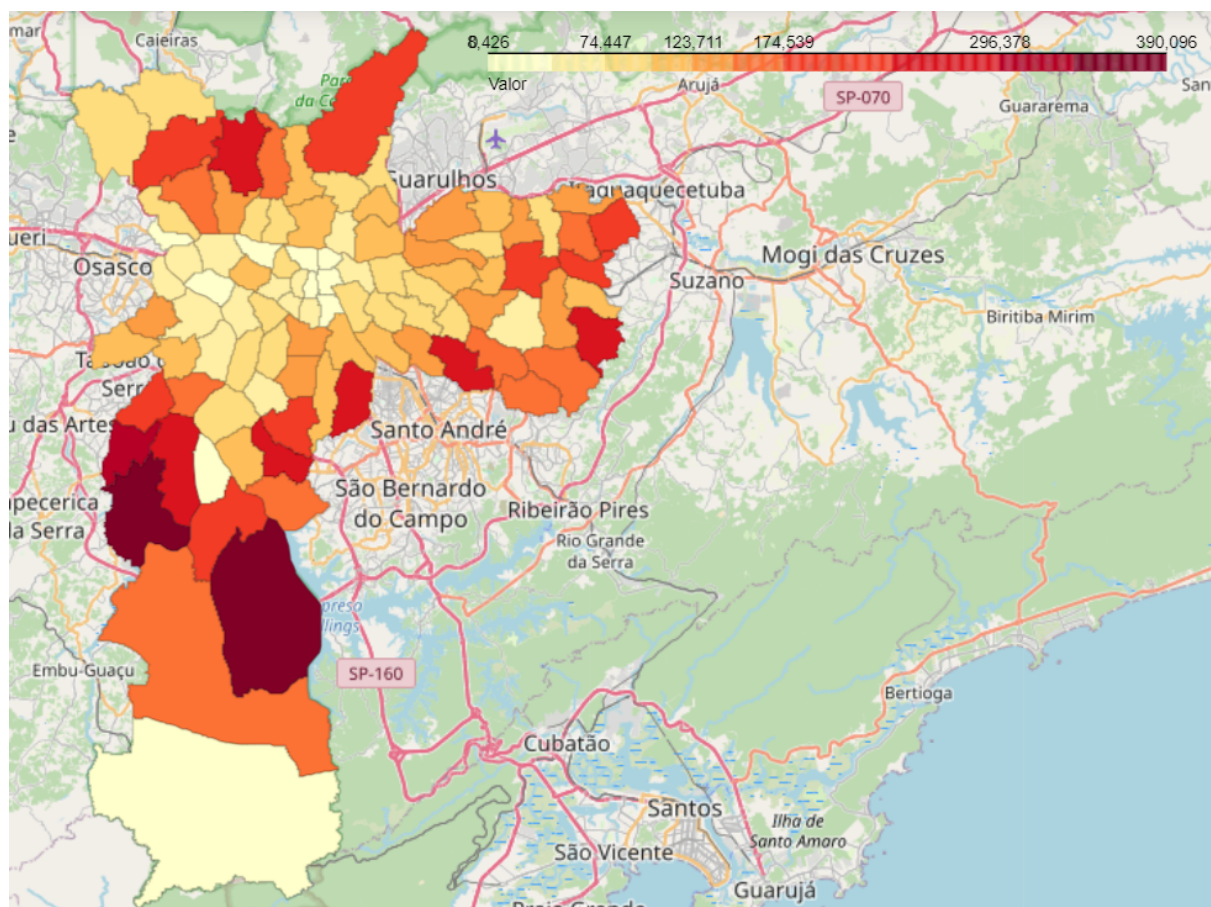


Fonte: Elaborada pelo autor

4.2 Panorama Geral dos Acidentes

Para entender o comportamento dos acidentes, é importante saber como está a distribuição populacional da cidade em seus distritos. Com a Figura 23, entende-se que os distritos mais populosos encontram-se às margens do município, e os distritos da zona central possuem baixa população.

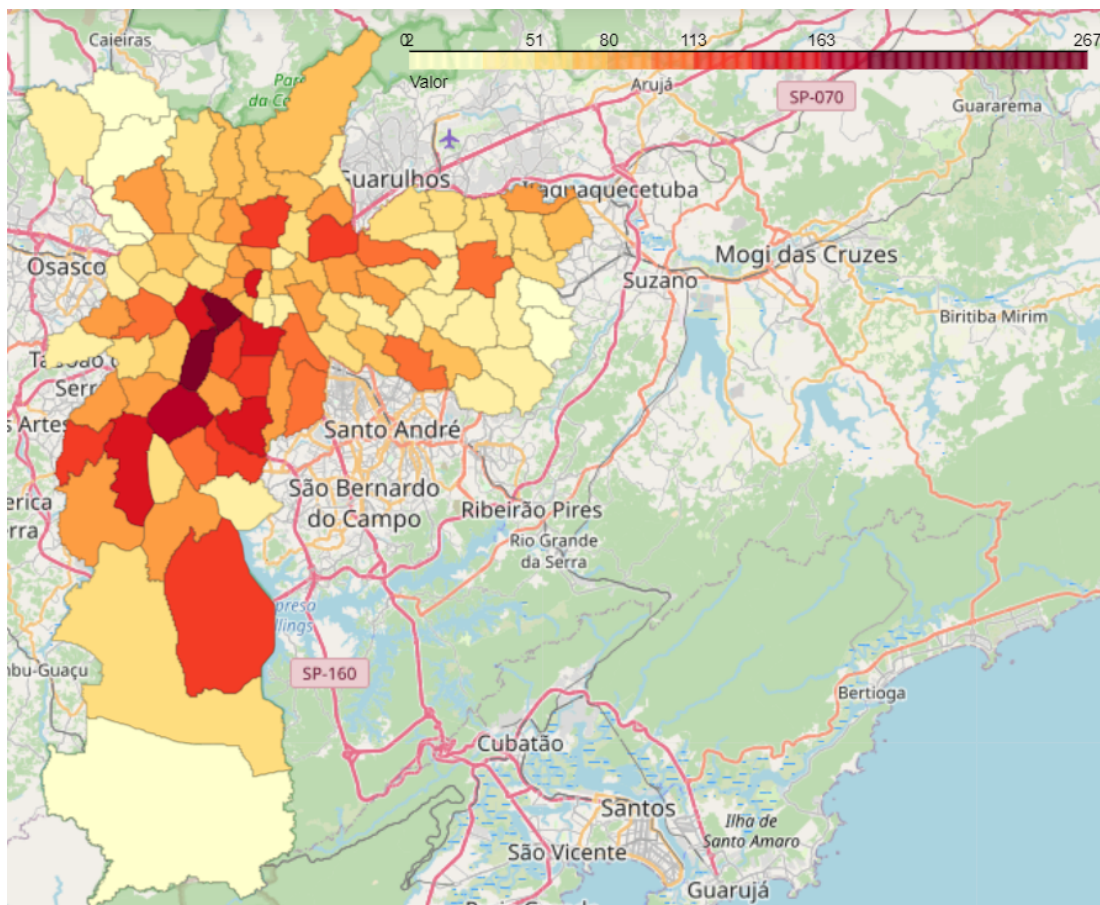
Figura 23 – População por distrito



Fonte: Elaborada pelo autor

O mapa coroplético com o total de acidentes com ciclistas por distrito, na Figura 24, traz poucas evidências. Visualmente, nota-se uma concentração na Zona Sul da capital paulista, e, por se estar olhando para valores absolutos, não há uma correlação efetiva existente com a distribuição populacional ou cobertura ciclovária.

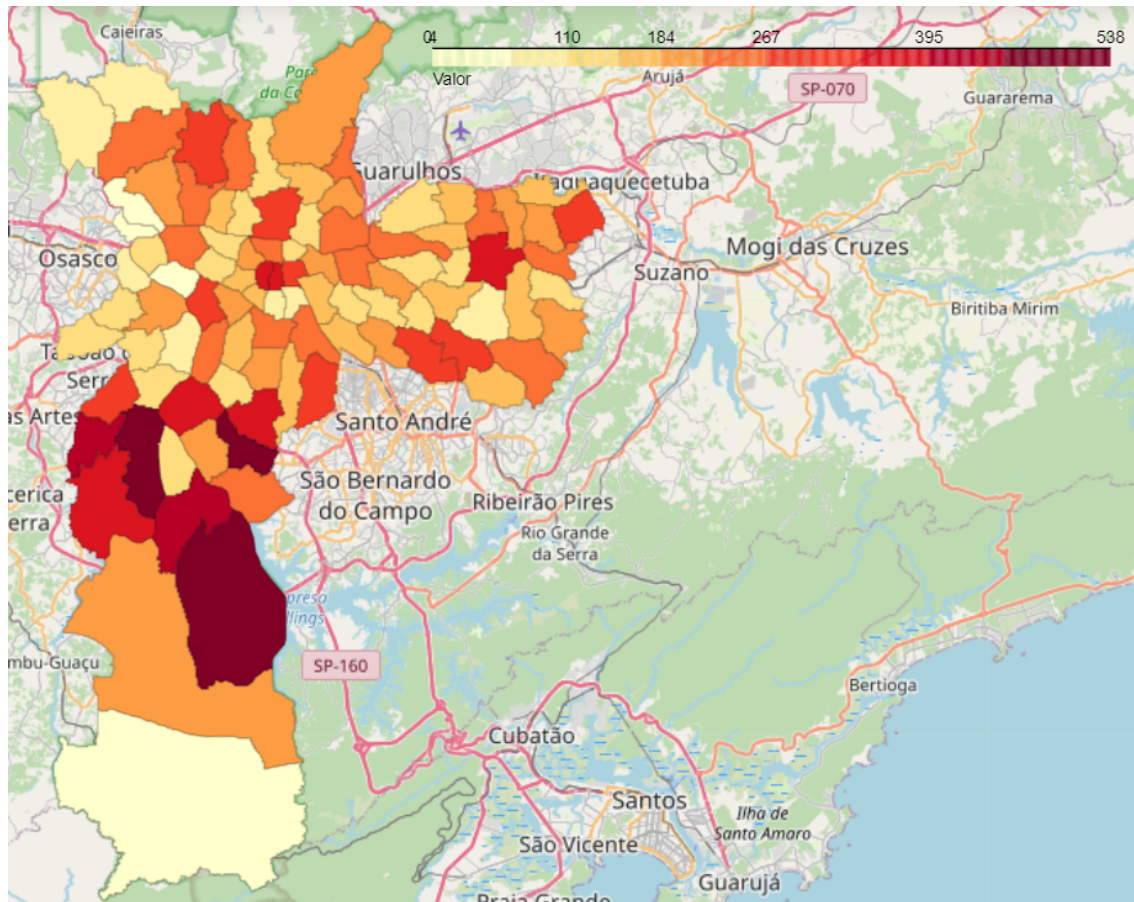
Figura 24 – Total de acidentes com ciclistas por distrito



Fonte: Elaborada pelo autor

Olhando para o total de acidentes com pedestres por distrito através da Figura 25, assim como pôde ser visto no mapa de acidentes com ciclistas, observa-se uma alta quantidade de acidentes ao sul do município. No entanto, também nota-se que os acidentes estão mais distribuídos pela cidade, quando comparado ao de ciclistas, e possuem uma correlação moderadamente forte com os valores de população, como pode ser visto na Figura 26

Figura 25 – Total de acidentes com pedestres por distrito



Fonte: Elaborada pelo autor

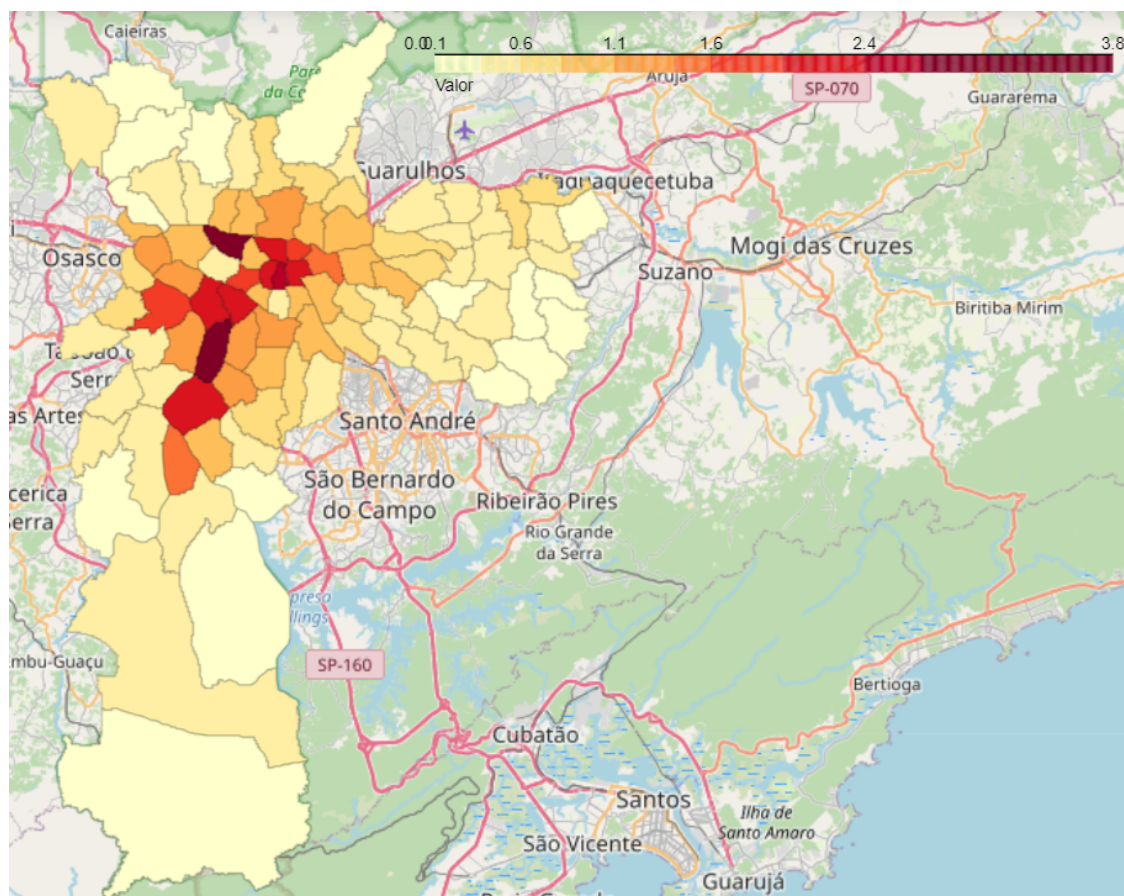
Figura 26 – Correlação pareada entre os valores de acidentes

	População	Acidente Ciclistas	Acidente Pedestres	Ciclistas/1000 Habitantes	Pedestres/1000 Habitantes
População	1.000000	0.236055	0.679739	-0.547671	-0.428688
Acidente Ciclistas	0.236055	1.000000	0.586170	0.450272	0.148906
Acidente Pedestres	0.679739	0.586170	1.000000	0.036035	0.204835
Ciclistas/1000 Habitantes	-0.547671	0.450272	0.036035	1.000000	0.804632
Pedestres/1000 Habitantes	-0.428688	0.148906	0.204835	0.804632	1.000000

Fonte: Elaborada pelo autor

Quando se observa os acidentes proporcionais à população de cada distrito, obtém-se uma visão completamente diferente do visto em valores absolutos. No caso de ciclistas, é visto que os distritos mais críticos estão na região central, enquanto o restante dos distritos apresentam valores relativamente baixos. Olhando novamente para a Figura 26, nota-se uma correlação moderada negativa entre os acidentes proporcionais à população - visto na Figura 27 - e os valores de população, o que indica que distritos com menos habitantes possuem mais acidentes com ciclistas por habitante.

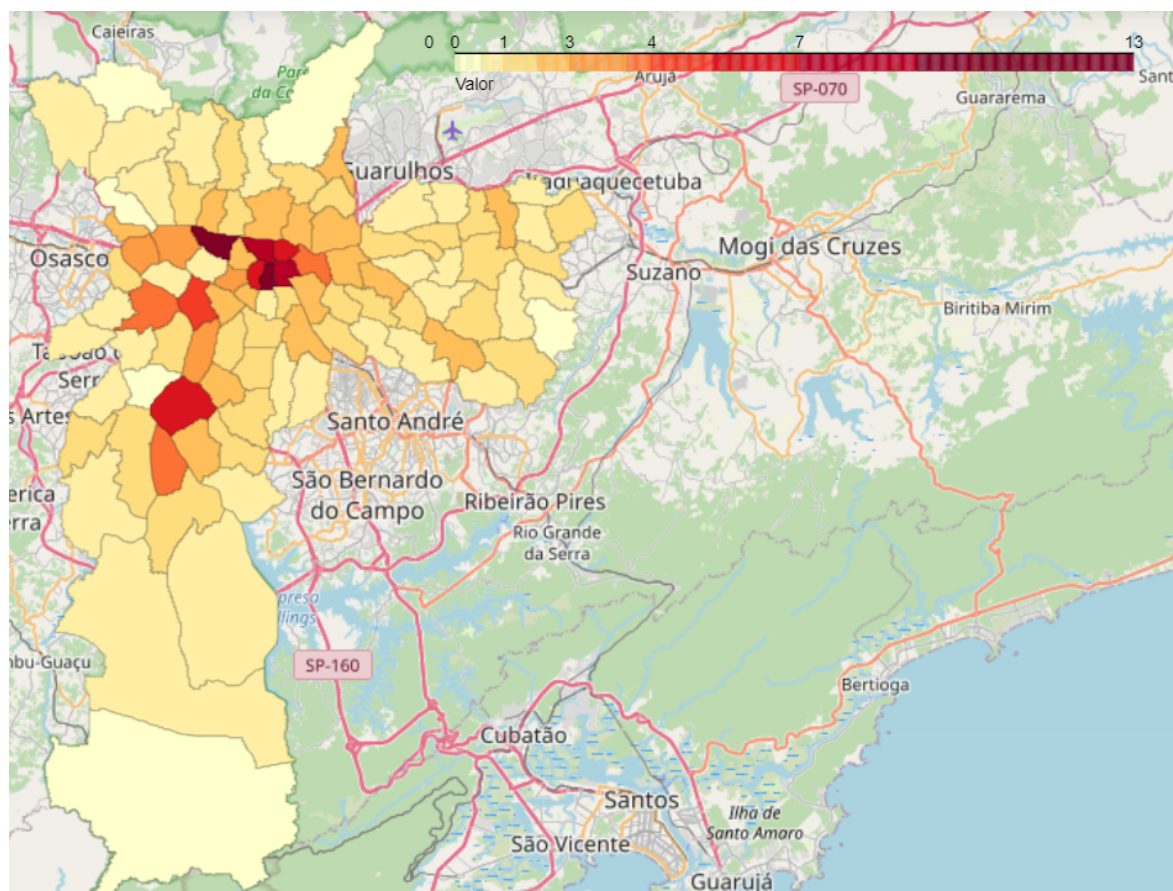
Figura 27 – Acidentes com ciclistas para cada 1000 habitantes por distrito



Fonte: Elaborada pelo autor

Já no caso de acidentes com pedestres proporcionais à população de distrito, na Figura 28, observa-se um comportamento bem semelhante ao de ciclistas (como pode ser visto na tabela de correlação pareada), tendo uma correlação negativa um pouco mais fraca com os dados de população.

Figura 28 – Acidentes com pedestres para cada 1000 habitantes por distrito



Fonte: Elaborada pelo autor

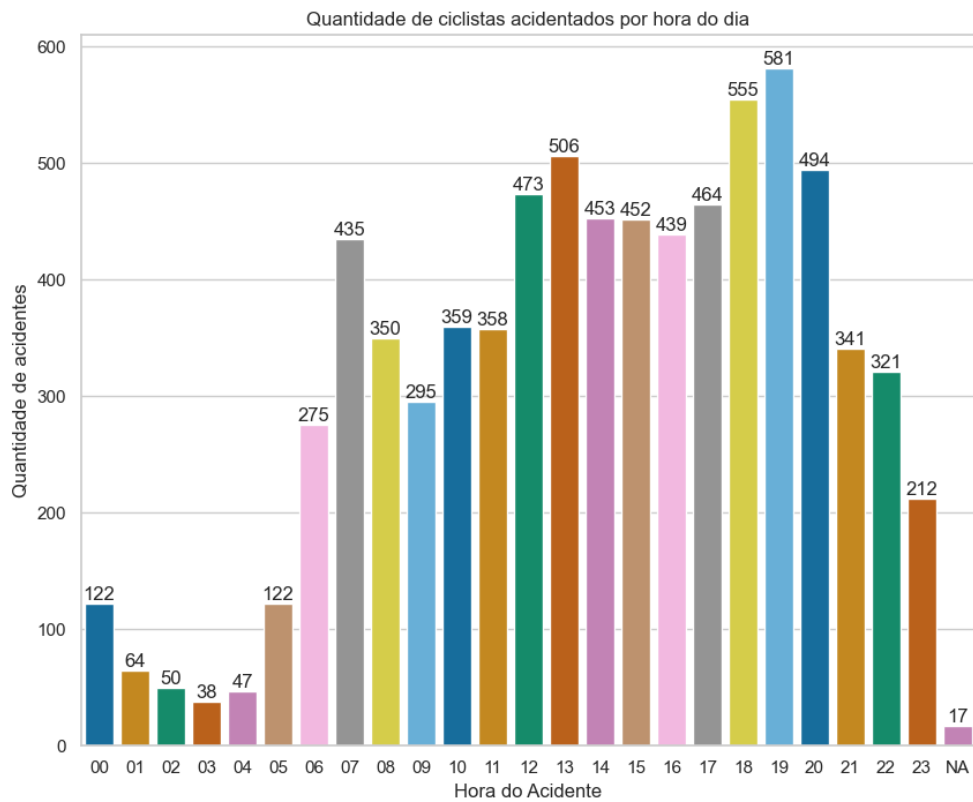
4.3 Características dos acidentes e acidentados

A Figura 29 indica a quantidade de acidentes com ciclistas para cada hora do dia. É interessante notar que, ao observar o gráfico, nota-se que os máximos locais (picos no gráfico que antecedem quedas) se encontram em horários de pico:

- **07:00 - 08:00** indica pessoas se deslocando para chegar ao local de trabalho;
- **12:00 - 13:00** indica o horário comum de almoço, onde há um fluxo de entregadores de aplicativo, além de pessoas que saem para comer;
- **18:00 - 19:00** indica o horário de fim de serviço onde as pessoas estão indo para suas casas, além disso o horário condiz com fim de tarde, ou seja, uma iluminação pior comparada à luz do dia.

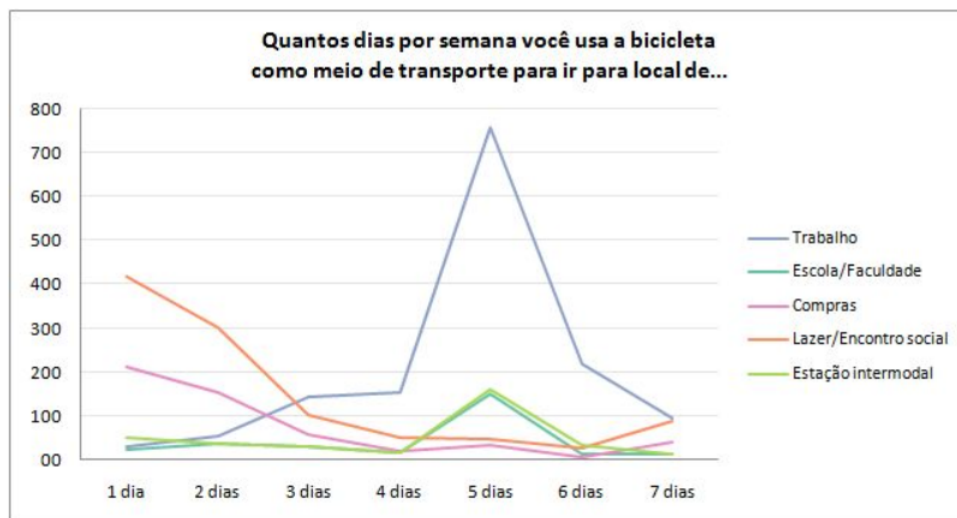
Ao comparar essa hipótese com a pesquisa de Freitas (2016) no quesito de uso das bicicletas de acordo com o destino da viagem, observável na Figura 30, percebe-se que tal análise casa com a realidade.

Figura 29 – Acidentes com ciclistas x Hora do dia



Fonte: Elaborada pelo autor

Figura 30 – Destino das viagens

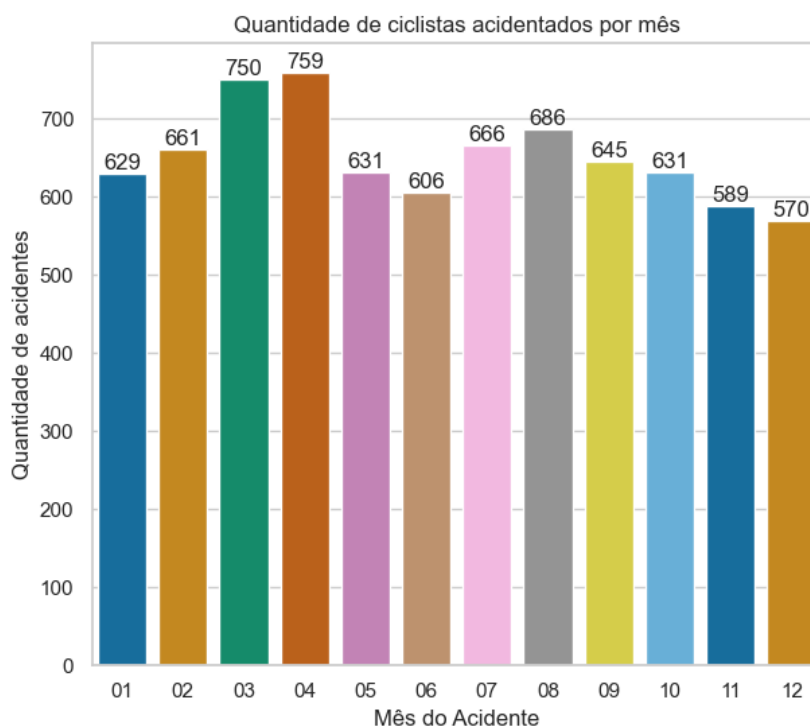


Fonte: (FREITAS, 2016)

Quando olha-se para os acidentes por mês, através do gráfico na Figura 31, nota-se

uma homogeneidade na distribuição, o que indica que não há um fator mensal que influencie em um maior fluxo de ciclistas pela cidade de São Paulo, e consequentemente, mais acidentes.

Figura 31 – Acidentes com ciclistas x Mês

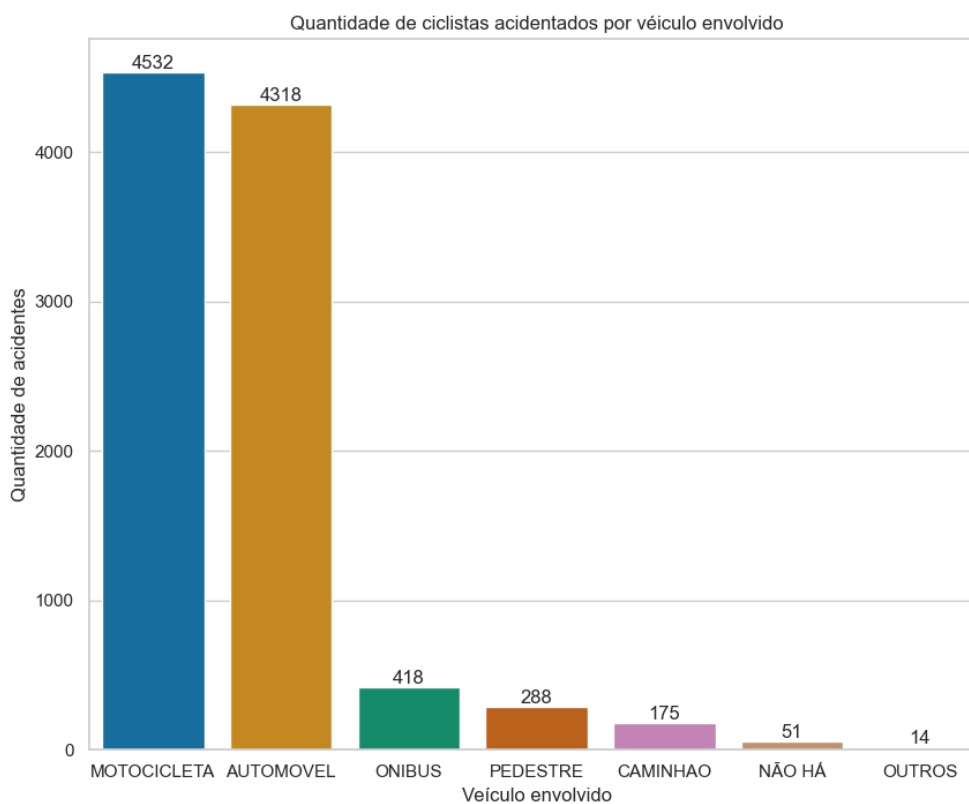


Fonte: Elaborada pelo autor

Olhando para os acidentes com ciclistas de acordo com o outro veículo envolvido, presente na Figura 32, percebe-se uma diferença discrepante entre motocicletas e automóveis para com o resto dos modais. Isso se explica pelo fato de o volume de carros e motos ser muito superior ao volume de ônibus, caminhões e pedestres. Estudo do IBGE² mostra que em 2022 a cidade de São Paulo possuía aproximadamente 9,2 milhões de veículos, sendo 6 milhões de automóveis e 1,1 milhão de motocicletas. Embora haja quase 6x mais carros que motos, ainda assim, o gráfico mostra que há mais motocicletas envolvidas em acidentes que carros, o que pode indicar uma maior imprudência de motoqueiros no trânsito.

² <https://cidades.ibge.gov.br/brasil/sp/sao-paulo/pesquisa/22/28120>

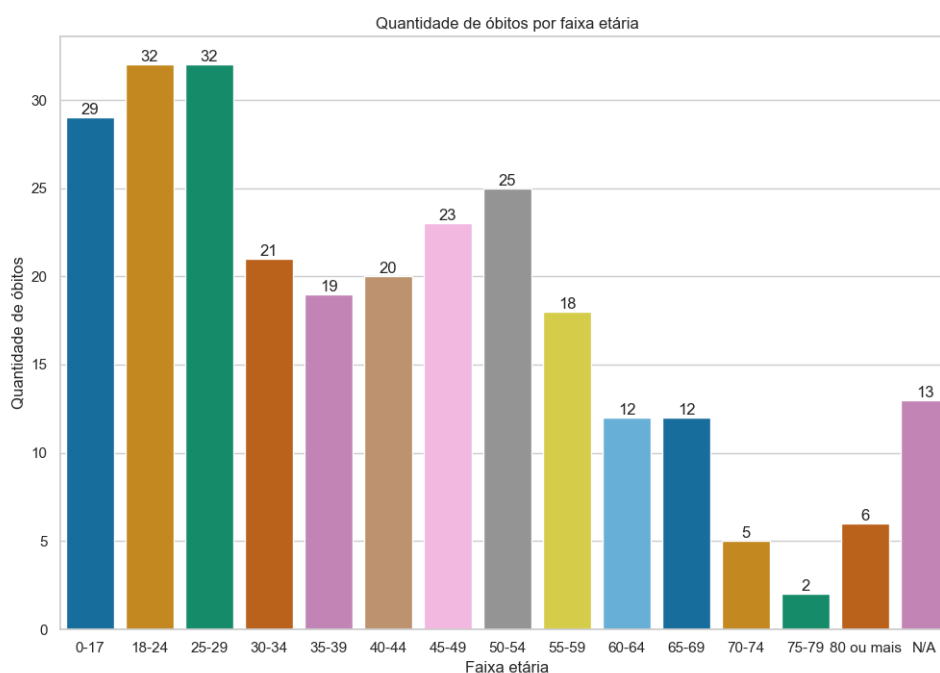
Figura 32 – Acidentes com ciclistas x Veículo envolvido



Fonte: Elaborada pelo autor

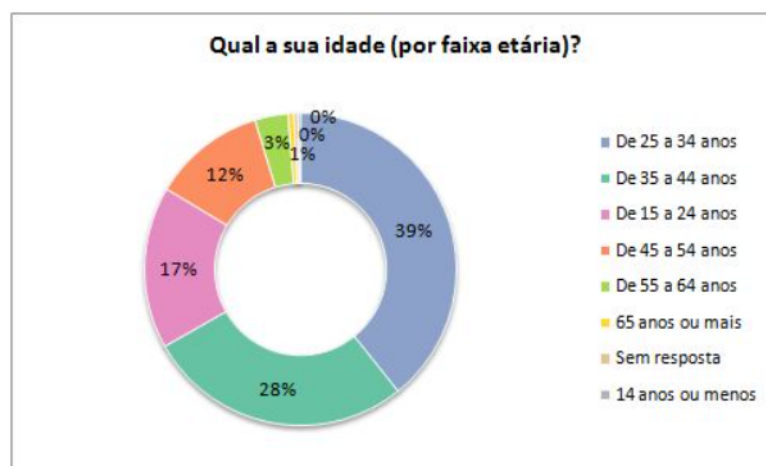
O gráfico de barras com óbitos por faixa etária na Figura 33 nos mostra uma tendência de queda conforme a idade avança. Pouco mais de 1/3 das vítimas tem menos de 30 anos, enquanto aproximadamente 50% possui menos de 40 anos. Levando em consideração a quantidade de ciclistas existentes na cidade de São Paulo por faixa etária, considerando os resultados de Freitas (2016) na Figura 34, é justo notar que os óbitos são relativamente proporcionais à quantidade de ciclistas para aquela faixa etária.

Figura 33 – Óbitos x Faixa Etária



Fonte: Elaborada pelo autor

Figura 34 – Idade dos ciclistas de São Paulo



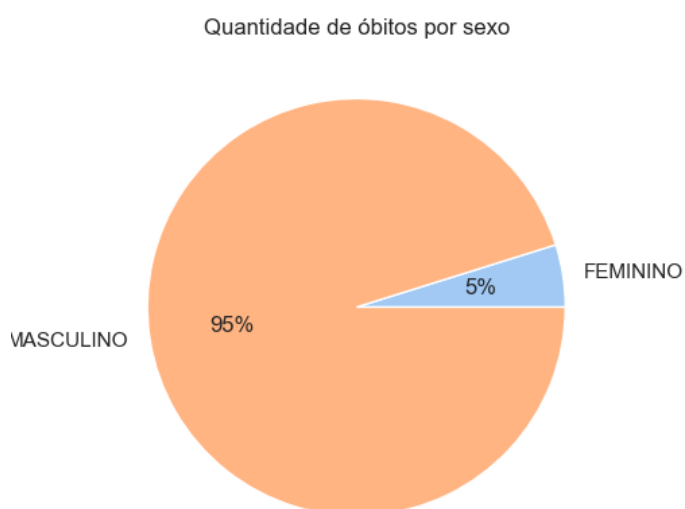
Fonte: (FREITAS, 2016)

A quantidade de óbitos por sexo no município de São Paulo apresenta uma discrepância elevada, onde ciclistas masculinos representam aproximadamente 95% dos óbitos, visto na Figura 35. Embora as mulheres representem pouco mais de 50% da população da capital paulistana, elas representam, em média 14% dos ciclistas da cidade. Segundo Freitas (2016), dados mostram uma alta concentração de uso de bicicleta entre mulheres com maiores níveis

de instrução e renda, sugerindo que mulheres andam mais de bicicleta ao terem maior renda e escolaridade. Isso evidencia que as barreiras para esse modo de transporte são maiores quando se trata de mulheres pobres e com menor escolaridade.

Dessa forma, os homens acabam por ser vasta maioria da composição de ciclistas, e, consequentemente, vítimas dos acidentes.

Figura 35 – Óbitos x Sexo



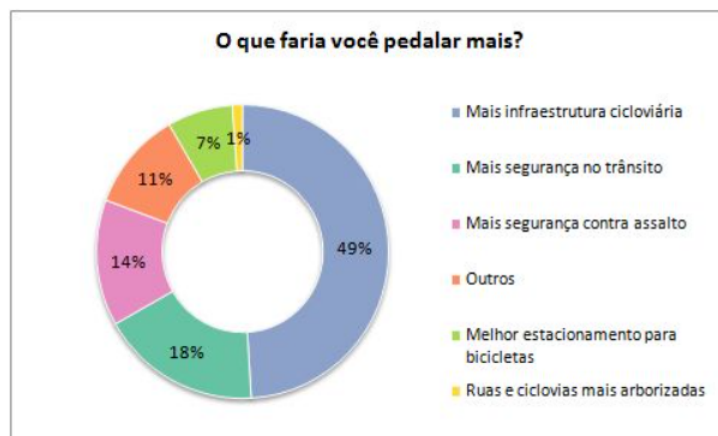
Fonte: Elaborada pelo autor

4.4 Segurança Viária na Infraestrutura Ciclovária

A análise da segurança viária e seu impacto nos acidentes é importante para tomadas de decisões no que diz respeito às futuras implementações de infraestrutura ciclovária.

A pesquisa de Freitas (2016) mostra que os principais motivos que fazem pessoas pedalarem mais são mais infraestrutura ciclovária e mais segurança no trânsito, como pode ser visto na Figura 36.

Figura 36 – Principais melhorias para ciclistas



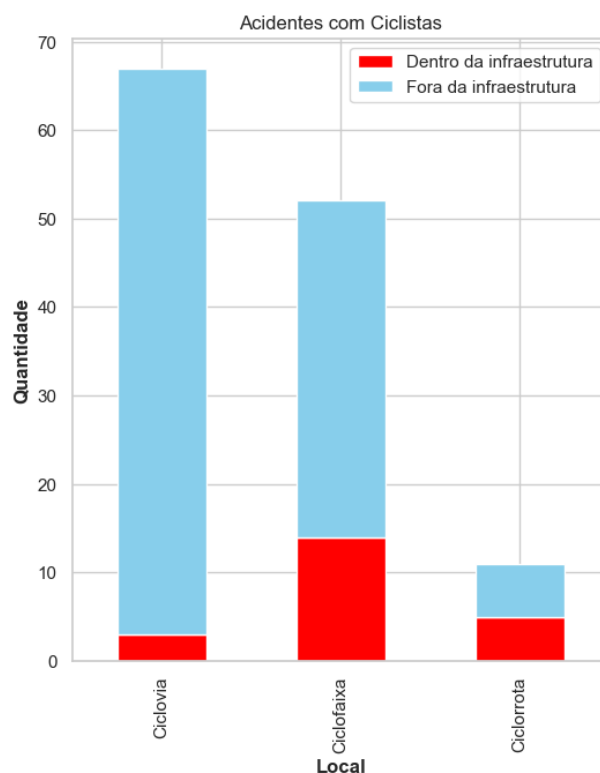
Fonte: Elaborada pelo autor

A Figura 37 indica, para o total de acidentes em cada infraestrutura cicloviária, quantos acidentes ocorreram dentro dela e quantos ocorreram fora dela, ou seja, na via em que ela se encontra. No gráfico abaixo, foi utilizado uma amostra de ciclovia, ciclofaixa e ciclorrota.

É possível notar que, em números absolutos, ocorrem mais acidentes em locais com ciclovias, seguidos de ciclofaixas e ciclorrotas. O motivo para tal se dá pelo fato de ciclovias se localizarem em vias com maior fluxo de veículos - vias arteriais, coletoras e expressas -, ciclofaixas em vias com fluxo médio, e ciclorrotas majoritariamente em vias locais.

Porém, embora o valor absoluto de acidentes diminui de ciclovias para ciclofaixas para ciclorrotas, a proporção de acidentes dentro das infraestruturas para acidentes fora da infraestrutura aumenta. Isso quer dizer que, embora as ciclovias fiquem em lugares mais movimentados, elas protegem muito mais, se comparadas às outras infraestruturas. Nota-se que, para ciclorrotas, cerca de metade dos acidentes ocorrem dentro dela, devido ao fato de ficar em meio às vias.

Figura 37 – Acidentes dentro e fora das infraestruturas cicloviárias - Amostra

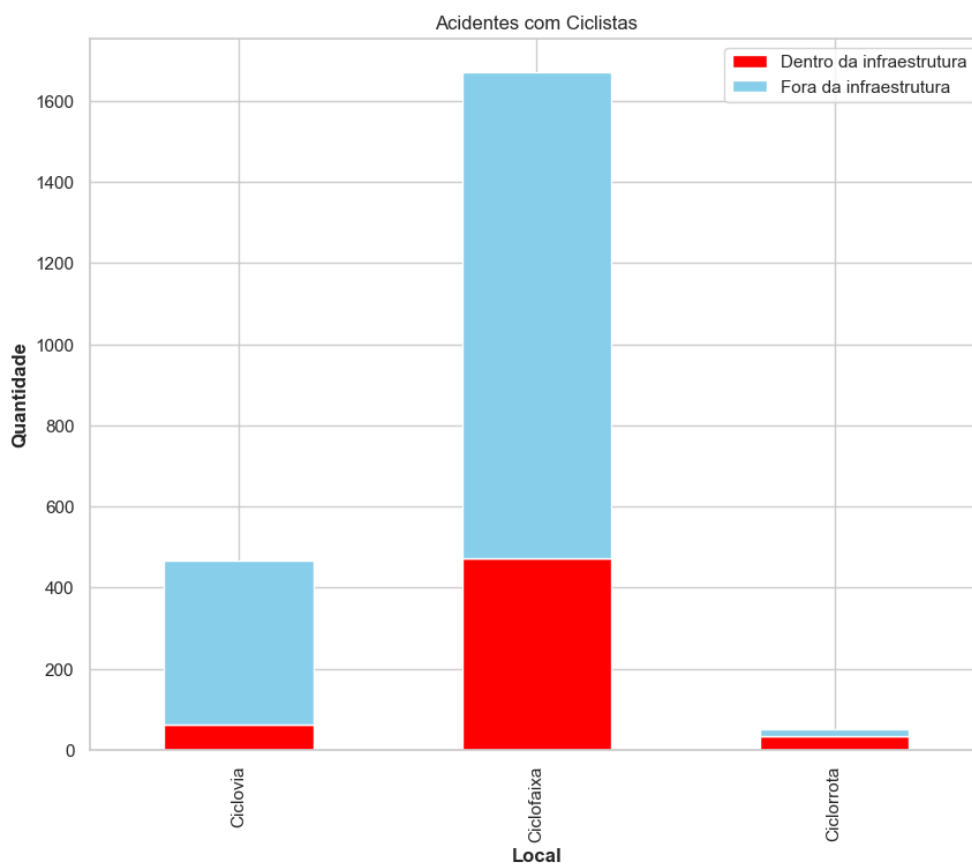


Fonte: Elaborada pelo autor

Saindo da amostra e analisando o resultado para todas as ciclovias, ciclofaixas e ciclorrotas, utilizando a Figura 38, obtém-se uma visão diferente sobre os valores totais. Observa-se que a quantidade de acidentes em ciclofaixas sobe drasticamente em relação às outras infraestruturas. Isso deve ao fato de haverem muito mais ciclofaixas - observado na Figura 20 - que ciclovias e ciclorrotas.

A Figura 38 acentua ainda mais a falta de segurança em ciclorrotas, visto que os acidentes dentro dela são maiores que fora. Já as proporções para ciclovias e ciclofaixas dentro e fora delas se mantêm praticamente inalteradas.

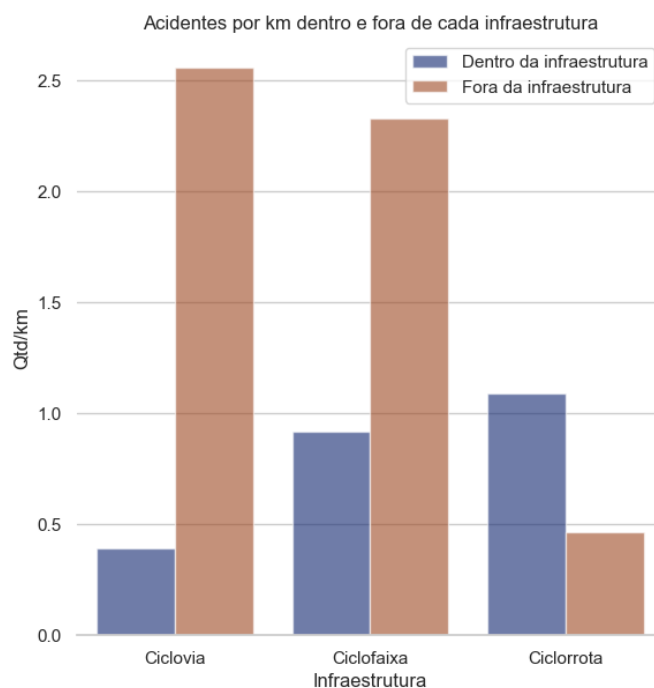
Figura 38 – Acidentes dentro e fora das infraestruturas ciclovárias - Geral



Fonte: Elaborada pelo autor

A análise dos acidentes dentro e fora das infraestruturas por km, apresentada na Figura 39, é importante para termos uma ideia da proporção de acidentes por km de cada infraestrutura. Se olharmos para as barras que representam os acidentes fora da infraestrutura, notamos que ciclovias e ciclofaixas indicam aproximadamente 2.5 acidentes por km de infraestrutura, porém as ciclovias comportam menos da metade dos acidentes dentro de sua infraestrutura comparadas às ciclofaixas, o que reforça a análise discutida sobre o gráfico da Figura 37. Já as ciclorrotas se diferem das outras duas infraestruturas com uma baixíssima segurança, tendo mais acidentes dentro do que fora delas.

Figura 39 – Acidentes dentro e fora das infraestruturas cicloviárias por km

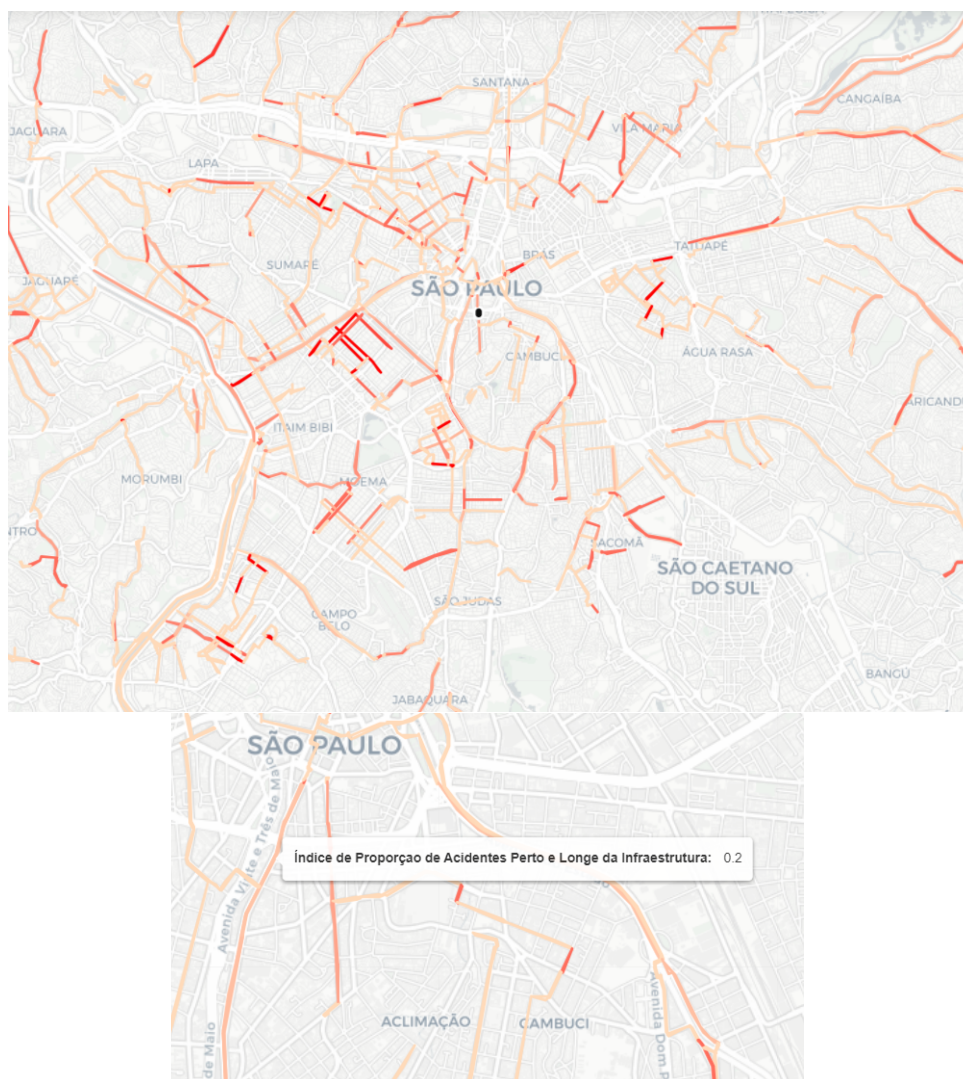


Fonte: Elaborada pelo autor

Por fim, gerou-se o **Índice de proporção entre os acidentes dentro e fora das infraestruturas** - explicado na subseção 3.6.3 e apresentada visualmente na Figura 40 - para cada trecho de ciclovia, ciclofaixa ou ciclorrota através da divisão da quantidade de acidentes dentro da infraestrutura pela quantidade de acidentes fora dela.

Essa visualização permite que autoridade competentes como a CET e a Prefeitura de São Paulo, possam utilizá-la para planejamento de futuras infraestruturas cicloviárias, além de medidas sobre as já existentes de forma a melhorar a segurança viária e estimular a mobilidade ativa.

Figura 40 – Índice de proporção entre os acidentes dentro e fora das infraestruturas



Fonte: Elaborada pelo autor

4.5 Custos de Sinalização de Ciclovia x Custos de AIHs de Acidentes

As análises sobre a segurança viária proporcionada por cada infraestrutura cicloviária nos permite perceber que o investimento em mais ciclovias, e em alguns casos ciclofaixas, pode auxiliar na redução de acidentes envolvendo ciclistas e, consequentemente, dos gastos do SUS com AIHs relacionadas à esses.

Tabela 2 – Custo de Sinalização de Ciclovia e Ciclofaixa

Tipologia	Custo Médio por KM
Ciclovia	R\$ 325.235,83
Ciclofaixa	R\$ 264.355,73

Fonte: Elaborada pelo autor

Tabela 3 – Custo com AIHs de Acidentes com Ciclistas nos últimos 10 anos

Município	AIHs Aprovadas	Valor Médio AIH	Valor Total
São Paulo	598.146	R\$ 1.465,94	R\$ 876.844.484,17

Fonte: Tabnet DATASUS - SIH/SUS

As tabelas 2 e 3 indicam o custo para implementação de ciclovias e ciclofaixas por km e o custo médio com uma AIH. Observando-as, é importante ressaltar que são necessários apenas 222 acidentes para que se possa construir 1km de ciclovia, sendo que apenas em janeiro e fevereiro de 2023 ocorreram 286 acidentes.

Se observado os últimos 10 anos, nota-se que foram 598.146 AIHs aprovadas para acidentes com ciclistas, gerando um gasto total de R\$ 876.844.484,17. Com esse valor, poderiam ser implementadas 2.696,02km de ciclovias e 3.316,9km de ciclofaixas.

4.6 Considerações

Os resultados acima mostram análises sobre o cenário da mobilidade ativa, focada principalmente no ciclismo, em diferentes níveis de detalhamento e abordagens.

As análises geoespaciais permitem ver de maneira mais geral, como está a distribuição dos acidentes pelos distritos do município de São Paulo. Percebe-se que, embora a zona central de SP comporte uma população menor (Figura 23), seus distritos possuem um intenso comércio como a Rua 25 de Março, o Mercado Municipal de São Paulo, e o Brás, e, portanto, um fluxo maior de pessoas e veículos, o que possibilita mais acidentes (Figuras 24 e 25).

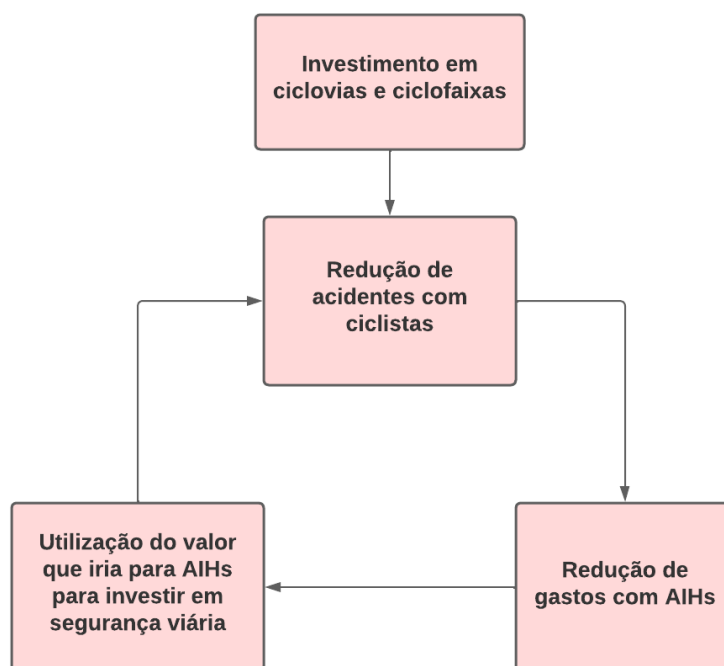
As análises sobre as características de cada acidente e dos acidentados permitem fazer reflexões socioeconômicas. O fato de a maioria dos acidentes acontecerem em horários de pico indica uma necessidade de revisão no controle de fluxo de veículos nesses horários. A altíssima diferença entre homens e mulheres ciclistas trazem à tona problemas relacionados à desigualdade de gênero:

"O que encontramos nas periferias paulistanas é, porém, fruto de interseccionalidades entre gênero, idade, raça, renda e escolaridade – além de, claro, conceitos e barreiras social e culturalmente construídos. Por exemplo, os dados da pesquisa mostram uma alta concentração de uso de bicicleta entre mulheres com maiores níveis de instrução e renda, sugerindo que mulheres andam mais de bicicleta ao terem maior renda e escolaridade. Isso evidencia que as barreiras para esse modo de transporte são maiores quando se trata de mulheres pobres e com menor escolaridade."(FREITAS, 2016).

Por fim, as análises de impacto da segurança viária de cada infraestrutura cicloviária trazem informações importantes sobre o porquê acontecem esses acidentes. Foi possível observar a diferença de proteção oferecidas por cada infraestrutura, destacando positivamente as ciclovias e ciclofaixas, e negativamente as ciclorrotas. Com essa informação, a dos custos

para implementar essas infraestruturas, e a do custo com AIHs, foi possível indicar um possível investimento na sinalização de ciclovias e ciclofaixas de forma a reduzir o custo com internações e, assim, sobrar mais dinheiro para investir novamente em infraestruturas cicloviárias, gerando uma espécie de retro-alimentação, como pode ser vista na Figura 41.

Figura 41 – Fluxo de Retro Alimentação para investimento em infraestrutura cicloviária



Fonte: Elaborada pelo autor

5 Conclusão

O presente trabalho propôs analisar os impactos da segurança viária sobre a mobilidade ativa no município de São Paulo e entender o panorama dos acidentes e da infraestrutura cicloviária em seus distritos utilizando técnicas de ciência e análise de dados descritivas e com auxílio de visualizações geoespaciais e gráficas. O aprofundamento de detalhamento das análises permite observar e analisar o problema proposto de vários ângulos diferentes.

A ciência de dados tem o poder de gerar tomadas de decisões baseadas em fatos. As técnicas aplicadas visam agregar o máximo de valor às visualizações e ao mesmo tempo mantê-las intuitivas para que pessoas interessadas também possam entender. Além disso, os resultados possibilitam que autoridades e organizações competentes como o CET, SUS, prefeitura e até o governo de São Paulo tomem decisões baseadas nas visualizações e métricas fornecidas, como o **Índice de Proporção Entre Acidentes Dentro e Fora das Infraestruturas**, importante para futuras implementações/manutenções. Medidas como redução de velocidades em vias coletoras e locais e implementação de mais radares de velocidade em vias mais rápidas podem ser levadas em consideração a partir do Índice citado acima. Estudo indicado por Neki et al. (2021) indica que a implementação de zonas de 30 km/h com medidas físicas para acalmar acidentes de trânsito resultaram em uma diminuição em acidentes envolvendo ciclistas de 48%.

5.1 Trabalhos Futuros

O *notebook* com todos os processos descritos nesse trabalho, todas as análises feitas e todos os resultados obtidos estão disponíveis no repositório indicado em 3.7 e permite que futuros trabalhos relacionados a essa área se apoiem nele como:

- Análise de vias e suas periculosidades relacionadas à mobilidade ativa;
- Análise de implementação de infraestrutura cicloviárias;
- Análise das fiscalizações voltadas para segurança de ciclistas e pedestres e seu impacto nos acidentes envolvendo eles.

Dessa forma, esse trabalho surge como importante no auxílio ao estímulo da mobilidade ativa, redução de acidentes e custos para o poder público.

Referências

- ADRIAZOLA-STEIL, C.; OHLUND, H.; SAMRA, S. E.; TARGA, F.; ZAYAS, G. Investing for momentum in active mobility. *World Bank Group*, 2021. Disponível em: <<https://thedocs.worldbank.org/en/doc/c6de7aa3e8b4b081029f639767c1486c-0190062021/original/TDI-paper-Investing-for-Momentum-in-Active-Mobility-October-2021.pdf>> . Acesso em: 20 de setembro de 2023.
- CÂMARA, G.; MONTEIRO, A. M. V.; DRUCK, S.; CARVALHO, M. S. Análise espacial e geoprocessamento. *Análise espacial de dados geográficos*. Brasília: EMBRAPA, 2004. Disponível em: <<https://www.dpi.inpe.br/gilberto/livro/analise/cap1-intro.pdf>> . Acesso em: 25 de outubro de 2023.
- FREITAS, F. Pesquisa perfil de quem usa bicicleta na cidade de são paulo – relatório completo. *CICLOCIDADE*, 2016. Disponível em: <https://www.ciclocidade.org.br/noticias/pesquisa-perfil-de-quem-usa-bicicleta-na-cidade-de-sao-paulo-relatorio-completo/#perfil_de_quem_usa> . Acesso em: 26 de outubro de 2023.
- FUND, C. P. Riding a bike for transportation. 2011. Disponível em: <<https://www.freestylecyclists.org/wp-content/uploads/2015/10/Cycling-Survey-2011-Riding-a-Bike-for-Transport.pdf>> . Acesso em: 20 de setembro de 2023.
- HÜBNER, A. T. Big data sob a ótica de modelagem de dados multidimensional para soluções de business intelligence. *Repositório Institucional UCS*, 2020. Disponível em: <<https://repositorio.ucs.br/xmlui/handle/11338/6327>> . Acesso em: 27 de outubro de 2023.
- LITMAN, T.; BURWELL, D. Issues in sustainable transportation. *Int. J. Global Environmental Issues*, v. 6, n. 4, 2006.
- NEKI, K.; LUMUMBA, M.; MITRA1, S.; JOB, R. S. Economic impact of 30km/h - benefits and costs of speeds in an urban environment. *World Bank Global Road Safety Facility*, 2021. Disponível em: <https://www.researchgate.net/publication/353710726_Economic_impact_of_30kmh_-_Benefits_and_Costs_of_Speeds_in_an_urban_environment> . Acesso em: 29 de outubro de 2023.
- PEDEN, M. *World report on road traffic injury prevention*. [S.l.]: World Health Organization, 2004. Disponível em: <<https://www.who.int/publications/i/item/world-report-on-road-traffic-injury-prevention>> . Acesso em: 06 de abril de 2023.
- PROVOST, F.; FAWCETT, T. Data science and its relationship to big data and data-driven decision making. *Big data*, 2013. Disponível em: <<http://doi.org/10.1089/big.2013.1508>> . Acesso em: 20 de setembro de 2023.
- REIS, I. A.; REIS, E. A. Análise descritiva de dados. *Relatório Técnico do Departamento de Estatística da UFMG*, 2002. Disponível em: <<https://www.est.ufmg.br/portal/wp-content/uploads/2023/01/RTE-02-2002.pdf>> . Acesso em: 27 de outubro de 2023.
- WELLE, B.; SHARPIN, A. B.; ADRIAZOLA-STEIL, C.; JOB, S.; SHOTTEN, M.; BOSE, D.; BHATT, A.; ALVEANO, S.; OBELHEIRO, M.; IMAMOGLU, T.

SUSTENTÁVEL E SEGURO Visão e Diretrizes para Zerar as Mortes no Trânsito.

[S.l.]: World Resources Institute, 2018. Disponível em: <<https://www.wri.org/research/sustainable-and-safe-vision-and-guidance-zero-road-deaths>> . Acesso em: 06 de abril de 2023.

APÊNDICE A – Infraestrutura Cicloviária

```
# Mapa de SP
# Latitude 23.5401 sul e longitude 46.6402 oeste
infra_cicloviaria = folium.Map(location=[-23.5401,-46.6402],
    zoom_start=11)

base_map = folium.FeatureGroup(name='Basemap', overlay=True, control
    =False)
folium.TileLayer(tiles='Stamen Terrain').add_to(base_map)
base_map.add_to(infra_cicloviaria)

layer1 = cria_e_adiciona_features('Distrito',infra_cicloviaria,True)
layer2 = cria_e_adiciona_features('Ciclovía',infra_cicloviaria,True)
layer3 = cria_e_adiciona_features('Ciclorrota',infra_cicloviaria,
    True)
layer4 = cria_e_adiciona_features('Ciclofaixa',infra_cicloviaria,
    True)

#Setando CRS pro shp das ciclovias:
gdf_ciclovía = gdf_ciclovía.set_crs(epsg=4326)
gdf_ciclofaixa = gdf_ciclofaixa.set_crs(epsg=4326)
gdf_distrito = gdf_distrito.to_crs(epsg='4326')

#Adicionando camadas de Distritos de SP e Ciclovias
adicionar_camadas(gdf_distrito,layer1,'distrito')
adicionar_camadas(gdf_ciclovía,layer2,'ciclovía')
adicionar_camadas(gdf_ciclorrota,layer3,'ciclorrota')
adicionar_camadas(gdf_ciclofaixa,layer4,'ciclofaixa')

folium.LayerControl(collapsed=False).add_to(infra_cicloviaria)
infra_cicloviaria
```

APÊNDICE B – Cobertura Cicloviária

```
# Mapa de SP
# Latitude 23.5401 sul e longitude 46.6402 oeste
infra_cicloviaria = folium.Map(location=[-23.5401,-46.6402],
                                zoom_start=11)

base_map = folium.FeatureGroup(name='Basemap', overlay=True, control
                                =False)
folium.TileLayer(tiles='cartodbpositron').add_to(base_map)
base_map.add_to(infra_cicloviaria)

layer1 = cria_e_adiciona_features('Distrito',infra_cicloviaria,True)
layer2 = cria_e_adiciona_features('Ciclovias',infra_cicloviaria,True)
layer3 = cria_e_adiciona_features('Ciclorrota',infra_cicloviaria,
                                True)
layer4 = cria_e_adiciona_features('Ciclofaixa',infra_cicloviaria,
                                True)

#Setando CRS pro shp das ciclovias:
gdf_ciclovias = gdf_ciclovias.set_crs(epsg=4326)
gdf_ciclofaixa = gdf_ciclofaixa.set_crs(epsg=4326)
gdf_distrito = gdf_distrito.to_crs(epsg='4326')

#Adicionando camadas de Distritos de SP e Ciclovias
adicionar_camadas(gdf_distrito,layer1,'distrito')
adicionar_camadas(gdf_ciclovias,layer2,'ciclovias')
adicionar_camadas(gdf_ciclorrota,layer3,'ciclorrota')
adicionar_camadas(gdf_ciclofaixa,layer4,'ciclofaixa')

folium.LayerControl(collapsed=False).add_to(infra_cicloviaria)
infra_cicloviaria
```

APÊNDICE C – Junção dos dados

Juntando dados de acidentes, óbitos e populacional

```
acidentes_populacao = pd.merge(df_populacional ,  
    pop_acidentes_ciclistas , left_on='distrito' , right_on='Distrito')  
acidentes_populacao = pd.merge(acidentes_populacao ,  
    pop_acidentes_pedestres , left_on='distrito' , right_on='Distrito')  
acidentes_populacao = pd.merge(acidentes_populacao ,  
    total_obitos_ciclistas , left_on='distrito' , right_on='Distrito')  
acidentes_populacao = pd.merge(acidentes_populacao ,  
    total_obitos_pedestres , left_on='distrito' , right_on='Distrito')
```

Limpando e tratando o novo dataframe unificado

```
acidentes_populacao = acidentes_populacao.drop(['distrito' ,  
    Distrito_y' ], axis=1)  
  
acidentes_populacao['Acidente Ciclistas'] = acidentes_populacao['  
    Acidente Ciclistas'] + acidentes_populacao['Qtd_Acidente_x']  
acidentes_populacao['Acidente Pedestres'] = acidentes_populacao['  
    Acidente Pedestres'] + acidentes_populacao['Qtd_Acidente_y']  
  
acidentes_populacao = acidentes_populacao.drop(['Qtd_Acidente_x' ,  
    Qtd_Acidente_y' ], axis=1)  
  
acidentes_populacao = acidentes_populacao.iloc[:, :-1]  
  
acidentes_populacao['populacao_2020'] = acidentes_populacao['  
    populacao_2020'] * 1000
```

Criação de métricas para análise

```
acidentes_populacao['Ciclistas/1000 Habitantes'] =  
    acidentes_populacao['Acidente Ciclistas'] * 1000 /  
    acidentes_populacao['populacao_2020']  
acidentes_populacao['Pedestres/1000 Habitantes'] =  
    acidentes_populacao['Acidente Pedestres'] * 1000 /  
    acidentes_populacao['populacao_2020']  
acidentes_populacao = acidentes_populacao.rename(columns={"  
    Distrito_x": "Distrito", "populacao_2020": "Popula o"})
```

Gerando o dataset de opção para criação da visualização

```
dados_panorama_acidentes = acidentes_populacao.rename(columns={"  
    populacao_2020": "Popula o"}).drop('Distrito' , axis=1)
```

APÊNDICE D – Visualização dos dados - Mapa Coroplético

```
opcoes = dados_panorama_acidentes

def gera_mapa(opcao):

    mapa_sp = folium.Map(location=[-23.7201,-46.6402],
                           attr= 'Heatmap',
                           zoom_start=10)

    choropleth_map = folium.Choropleth(
        geo_data=gdf_distrito ,
        data=acidentes_populacao ,
        columns=['Distrito',opcao],
        key_on='feature.properties.NOME_DIST',
        bins=10,
        use_jenks=True,
        fill_color='YlOrRd',
        fill_opacity=1,
        line_opacity=0.2,
        smooth_factor=0,
        Highlight= False,
        line_color = "Black",
        nan_fill_color = "White",
        legend_name= 'Valor'
    ).add_to(mapa_sp)

    dados_acidentes_heatmap_indexed = acidentes_populacao.set_index(
        'Distrito')

    for row in choropleth_map.geojson.data['features']:
        try:
            row['properties'][opcao] = str(
                dados_acidentes_heatmap_indexed.loc[row['properties']
                ]['NOME_DIST'],opcao))
        except KeyError:
            row['properties'][opcao] = 'No adherents'

    folium.GeoJsonTooltip(['NOME_DIST', opcao]).add_to(
        choropleth_map.geojson)

    display(mapa_sp)
```

```
opcao_widget = widgets.ToggleButtons(options=opcoes, layout={'width': 'max-content'}, description='Analise:')  
interact(gera_mapa, opcao=opcao_widget)
```

APÊNDICE E – Mapa de Calor - Acidentes

```
heatmap_acidentes_ciclistas = acid_ciclistas[['Latitude','Longitude']]

mapa_de_calor_teste = folium.Map(location=[-23.5901,-46.6702],
                                  tiles = 'cartodbpositron',
                                  attr= 'Heatmap',
                                  zoom_start=15)

layer5 = cria_e_adiciona_features('Distritos',mapa_de_calor_teste,
                                  True)
layer6 = cria_e_adiciona_features('Ciclovia',mapa_de_calor_teste,
                                  True)
layer7 = cria_e_adiciona_features('Ciclorrota',mapa_de_calor_teste,
                                  True)
layer8 = cria_e_adiciona_features('Ciclofaixa',mapa_de_calor_teste,
                                  True)

adicionar_camadas(gdf_distrito,layer5,'distrito')
adicionar_camadas(gdf_ciclovia,layer6,'ciclovia')
adicionar_camadas(gdf_ciclorrota,layer7,'ciclorrota')
adicionar_camadas(gdf_ciclofaixa,layer8,'ciclofaixa')

HeatMap(data=heatmap_acidentes_ciclistas).add_to(mapa_de_calor_teste)

mapa_de_calor_teste
```

APÊNDICE F – Hora e mês dos acidentes

```
obitos_ciclista_sp = obitos_sp[(obitos_sp['Tipo do veiculo da vitima'] == 'BICICLETA') & (obitos_sp['Ano do BO (RDO)'] > 2018) & (obitos_sp['Municipio'] == 'SAO PAULO')]
datas_obitos_ciclistas_sp = obitos_ciclista_sp[['Data do Acidente', 'Hora do Acidente']]
datas_obitos_ciclistas_sp['Data do Acidente'] =
    datas_obitos_ciclistas_sp['Data do Acidente'].str.slice(5, 7)
datas_obitos_ciclistas_sp['Hora do Acidente'] =
    datas_obitos_ciclistas_sp['Hora do Acidente'].str.slice(0, 2)

nao_fatais_ciclistas = nao_fatais[(nao_fatais['Veiculos Envolvidos – Bicicleta'] == 1) & (nao_fatais['Municipio'] == "SAO PAULO")]
datas_nao_fatais_ciclistas = nao_fatais_ciclistas[['Data do Acidente', 'Hora do Acidente']]
datas_nao_fatais_ciclistas['Data do Acidente'] =
    datas_nao_fatais_ciclistas['Data do Acidente'].str.slice(5, 7)
datas_nao_fatais_ciclistas['Hora do Acidente'] =
    datas_nao_fatais_ciclistas['Hora do Acidente'].str.slice(0, 2)

data_acidentes = pd.concat([datas_obitos_ciclistas_sp,
    datas_nao_fatais_ciclistas], axis=0)
data_acidentes = data_acidentes.sort_values(by= ['Data do Acidente', 'Hora do Acidente'], ascending=True)
data_acidentes = data_acidentes.rename(columns={"Data do Acidente": "Mes do Acidente"})
```

APÊNDICE G – Veículo Envolvido nos acidentes

```
def acidentes_por_veiculo(veiculo_envolvido, veiculo):
    df_acidente = nao_fatais_ciclistas.groupby(veiculo_envolvido,
        as_index=False)['ID'].count()
    df_acidente = df_acidente[df_acidente[veiculo_envolvido] == 1]
    df_acidente['Outro Veiculo Envolvido'] = veiculo
    df_acidente = df_acidente.drop(veiculo_envolvido, axis=1)
    df_acidente = df_acidente.rename(columns={"ID": "Quantidade"})

    return df_acidente

acidentes_caminhao = acidentes_por_veiculo('Veiculos Envolvidos - Caminhao', 'CAMINHAO')
acidentes_carro = acidentes_por_veiculo('Veiculos Envolvidos - Automovel', 'AUTOMOVEL')
acidentes_moto = acidentes_por_veiculo('Veiculos Envolvidos - Motocicleta', 'MOTOCICLETA')
acidentes_onibus = acidentes_por_veiculo('Veiculos Envolvidos - Onibus', 'ONIBUS')
acidentes_bicicleta = acidentes_por_veiculo('Veiculos Envolvidos - Bicicleta', 'BICICLETA')
acidentes_pedestre = acidentes_por_veiculo('Veiculos Envolvidos - Pedestre', 'PEDESTRE')

veiculos_obitos = obitos_ciclista_sp.groupby('Outro Veiculo Envolvido', as_index=False)['Id da Delegacia (RDO)'].count()
veiculos_obitos = veiculos_obitos.drop(labels=[1], axis=0)
veiculos_obitos = veiculos_obitos.rename(columns={"Id da Delegacia (RDO)": "Quantidade"})
veiculos_obitos['Outro Veiculo Envolvido'] = veiculos_obitos['Outro Veiculo Envolvido'].replace({'NAO DISPONIVEL': 'OUTROS'})

df_acidentes_por_veiculo = pd.concat([acidentes_caminhao,
    acidentes_carro, acidentes_moto,
    acidentes_onibus,
    acidentes_pedestre], axis=0)

df_acidentes_por_veiculo = pd.merge(df_acidentes_por_veiculo,
    veiculos_obitos, on=['Outro Veiculo Envolvido'], how='right')
df_acidentes_por_veiculo = df_acidentes_por_veiculo.fillna(0)
df_acidentes_por_veiculo['Quantidade'] = df_acidentes_por_veiculo['
```



```
Quantidade_x'] + df_acidentes_por_veiculo['Quantidade_y']  
df_acidentes_por_veiculo = df_acidentes_por_veiculo.drop(['  
    Quantidade_x', 'Quantidade_y'], axis=1)  
df_acidentes_por_veiculo
```

APÊNDICE H – Criação dos buffers

```
# 0.0306475 = 3411.664 metros —> 1 metro = 0.00000898 = aprox. 0.000009

# 0.000018 = aprox. 2 metros
# 0.000045 = 5 metros
# 0.000063 = 7 metros
# 0.000090 = 10 metros
# 0.00018 = 20 metros

# Ciclovia
buffer_ciclovia_geral = gdf_ciclovia[['programa', 'geometry']]

# Buffer para Acidentes Na Ciclovia
buffer_acidentes_na_ciclovia_geral = gpd.GeoDataFrame(geometry=gpd.
    GeoSeries(buffer_ciclovia_geral.buffer(0.000045)))

# Buffer para Acidentes Fora da Ciclovia
buffer_acidentes_fora_da_ciclovia_geral = gpd.GeoDataFrame(geometry=
    gpd.GeoSeries(buffer_ciclovia_geral.buffer(0.00018)))

# Junção ciclovias bufferizadas com acidentes
acidentes_na_ciclovia_geral = gpd.sjoin(left_df=
    buffer_acidentes_na_ciclovia_geral, right_df=
    gdf_acidentes_ciclistas_localizados, how='inner')
acidentes_fora_da_ciclovia_geral = gpd.sjoin(left_df=
    buffer_acidentes_fora_da_ciclovia_geral, right_df=
    gdf_acidentes_ciclistas_localizados, how='inner')

# Ciclofaixa
buffer_ciclofaixa_geral = gdf_ciclofaixa[['programa', 'geometry']]

# Buffer para Acidentes Na Ciclofaixa
buffer_acidentes_na_ciclofaixa_geral = gpd.GeoDataFrame(geometry=gpd.
    GeoSeries(buffer_ciclofaixa_geral.buffer(0.000045)))

# Buffer para Acidentes Fora da Ciclofaixa
buffer_acidentes_fora_da_ciclofaixa_geral = gpd.GeoDataFrame(
    geometry=gpd.GeoSeries(buffer_ciclofaixa_geral.buffer(0.00018)))

# Junção ciclofaixas bufferizadas com acidentes
acidentes_na_ciclofaixa_geral = gpd.sjoin(left_df=
    buffer_acidentes_na_ciclofaixa_geral, right_df=
    gdf_acidentes_ciclistas_localizados, how='inner')
acidentes_fora_da_ciclofaixa_geral = gpd.sjoin(left_df=
```

```
buffer_acidentes_fora_da_ciclofaixa_geral, right_df=
gdf_acidentes_ciclistas_localizados, how='inner')
```

Ciclorrota

```
buffer_ciclorrota_geral = gdf_ciclorrota[['programa', 'geometry']]
```

Buffer para Acidentes Na Ciclorrota

```
buffer_acidentes_na_ciclorrota_geral = gpd.GeoDataFrame(geometry=gpd
    .GeoSeries(buffer_ciclorrota_geral.buffer(0.000045)))
```

Buffer para Acidentes Fora da Ciclorrota

```
buffer_acidentes_fora_da_ciclorrota_geral = gpd.GeoDataFrame(
    geometry=gpd.GeoSeries(buffer_ciclorrota_geral.buffer(0.000090)))
```

Junção ciclorrota bufferizadas com acidentes

```
acidentes_na_ciclorrota_geral = gpd.sjoin(left_df=
    buffer_acidentes_na_ciclorrota_geral, right_df=
    gdf_acidentes_ciclistas_localizados, how='inner')
acidentes_fora_da_ciclorrota_geral = gpd.sjoin(left_df=
    buffer_acidentes_fora_da_ciclorrota_geral, right_df=
    gdf_acidentes_ciclistas_localizados, how='inner')
```