



COMPONENTE CURRICULAR:	Projeto Aplicado I
GRUPO	Grupo - Projeto Aplicado 18
ALUNOS	
Thaís Cristine de Andrade Gomes	10721642
Paulo Ricardo de Oliveira Ramos	10721464
Thabata Sabato Wiggers	

Análise da Mortalidade no Estado de São Paulo

Sumário

1. Glossário	3
2. Lista de Figuras	3
3. Lista de Tabelas	3
4. Introdução	4
5. Empresa e Problema de Pesquisa	4
6. Metadados	5
7. Objetivo	6
8. Análise Exploratória de Dados	7
9. Proposta Analítica	11
10. Link do Projeto no GitHub	13
11. Considerações Finais	13
12. Referências	14

Glossário

SIM - Sistema de Informação sobre Mortalidade, do Ministério da Saúde.

AED - Análise Exploratória de Dados, utilizada para conhecer e visualizar os dados.

DASHBOARD - Painel visual com indicadores e gráficos para tomada de decisão.

KPIs - Indicadores-chave de desempenho (Key Performance Indicators).

RAÇA/COR - Classificação utilizada pelo IBGE e Ministério da Saúde.

ASSISTÊNCIA MÉDICA - Atendimento médico realizado antes ou durante o óbito.

Lista de Figuras

- Figura 1: Distribuição de óbitos por sexo.
- Figura 2: Distribuição de óbitos por faixa etária.
- Figura 3: Distribuição de óbitos por raça/cor.
- Figura 4: Distribuição de óbitos por local de ocorrência.
- Figura 5: Distribuição de óbitos por tipo de óbito.
- Figura 6: Distribuição de óbitos por assistência médica.

Lista de Tabelas

- Tabela 1: Metadados da base SIM.

Introdução

Este documento apresenta a segunda etapa do Projeto Aplicado I, onde foca em desenvolver uma proposta de solução analítica com base em dados reais. O tema escolhido para o projeto na primeira etapa é a análise da mortalidade no estado de São Paulo no ano de 2023, utilizando dados públicos fornecidos pelo Sistema de Informação sobre Mortalidade (SIM), do Ministério da Saúde.

Na primeira etapa do projeto, foram abordados a apresentação da empresa responsável pela base de dados (o Ministério da Saúde), a contextualização do problema, o levantamento dos metadados e uma proposta inicial de solução baseada em dashboards e análises segmentadas. Nesta segunda etapa, aprofunda-se a análise com a aplicação de técnicas de Análise Exploratória de Dados (AED) em Python e o detalhamento do pipeline analítico.

A proposta visa identificar padrões, desigualdades regionais e características sociodemográficas relacionadas à mortalidade, contribuindo para a elaboração de estratégias públicas mais eficazes na área da saúde.

Esta segunda fase promove, além da descrição da metodologia utilizada, a apresentação dos scripts desenvolvidos em Python e das visualizações geradas a partir dos dados tratados, organizados em um repositório no GitHub para consulta e reprodução.

Empresa e Problema de Pesquisa

1. A organização.

A base de dados utilizada neste projeto foi extraída do Sistema de Informação sobre Mortalidade (SIM), mantido pelo Ministério da Saúde do Brasil. Este sistema é alimentado pelas Declarações de Óbito (DO) e tem como principal objetivo consolidar informações que possibilitem análises estatísticas sobre as causas e condições dos óbitos no país.

O SIM integra os sistemas de vigilância em saúde e é fundamental para o planejamento e a avaliação de políticas públicas. No caso do estado de São Paulo, a abrangência e detalhamento dos dados permitem análises específicas de caráter regional e sociodemográfico.

Missão: Promover a saúde da população brasileira por meio de políticas públicas eficazes, baseadas em dados e evidências.

Visão: Ser referência mundial em políticas públicas de saúde orientadas por dados e sustentadas por inovação e tecnologia.

Valores: Compromisso com a vida, equidade, transparência, ética e inovação.

2. Problema de pesquisa.

Apesar da robustez do sistema SIM, ainda existem desafios significativos sobre a mortalidade evitável, desigualdades no acesso aos serviços de saúde e ausência de assistência médica em boa parte dos óbitos registrados. O problema central deste projeto consiste em investigar:

- Existem desigualdades na mortalidade entre diferentes grupos sociodemográficos (sexo, faixa etária, raça/cor)?
- A ausência de assistência médica está concentrada em determinados perfis de óbitos ou regiões?
- Há padrões temporais (mês/estação) ou locais (tipo de estabelecimento) que se destacam?
- É possível identificar regiões prioritárias para atuação em políticas públicas?

A proposta utiliza o pensamento computacional para decompor o problema em partes menores e analisáveis, identificando padrões repetitivos, promovendo abstrações e possibilitando a criação de soluções automatizadas. A segmentação por variáveis como idade, sexo, raça/cor, tipo de óbito e local de ocorrência facilita desenvolver análises direcionadas que subsidiam tomadas de decisão na saúde pública.

Metadados

- Fonte: Sistema de Informação sobre Mortalidade (SIM) – Dados Abertos do Ministério da Saúde.
- Ano: 2023.
- Quantidade: +10 mil registros.

Variável	Tipo	Descrição
DTOBITO	Data	Data do óbito
ANO_OBITO	Numérica	Ano do óbito
MES_OBITO	Categórica	Mês do óbito
IDADE	Numérica	Idade do indivíduo falecido
SEXO_DESC	Categórica	Sexo (Masculino, Feminino)
RACACOR_DESC	Categórica	Raça/Cor (Branca, Parda, Preta, etc.)
UF_OCOR	Categórica	Unidade da federação da ocorrência do óbito
MUNICIPIO_OCOR	Categórica	Município de ocorrência do óbito
LOCOCOR_DESC	Categórica	Local de ocorrência (hospital, domicílio, etc.)
ASSISTMED_DESC	Categórica	Houve assistência médica (Sim/Não)
CIRCOBITO_DESC	Categórica	Circunstância do óbito (acidente, natural, etc.)

Objetivo

Realizar uma Análise Exploratória de Dados (AED) sobre os óbitos registrados em 2023 no estado de São Paulo. A proposta envolve a identificação de padrões com base em variáveis como sexo, faixa etária, raça/cor e local de ocorrência das mortes. Além disso, avaliar o acesso à assistência médica no momento do óbito, buscando e apontando possíveis desigualdades regionais ou sociodemográficas. Por fim, o projeto pretende propor uma solução analítica que possa auxiliar na tomada de decisão no que se refere às políticas públicas de saúde.

A proposta visa aplicar técnicas de Análise Exploratória de Dados (AED) e desenvolver uma solução analítica baseada em dashboards interativos, com o objetivo de apoiar gestores e

profissionais da saúde na formulação de políticas públicas mais assertivas e orientadas por evidências.

Objetivos específicos:

- Realizar a limpeza e tratamento da base de dados.
- Aplicar estatísticas descritivas e gerar visualizações gráficas que facilitem a compreensão dos dados e identificação de padrões relevantes.
- Identificar anomalias e possíveis indicadores críticos.
- Sugerir um pipeline analítico com ferramentas de visualização e suporte à tomada de decisão.

Análise Exploratória de Dados

A Análise Exploratória de Dados (AED) tem como objetivo examinar o comportamento das variáveis presentes na base de dados, buscando revelar padrões, anomalias, correlações e tendências relevantes para a compreensão do fenômeno da mortalidade no estado de São Paulo no ano de 2023.

A seguir, apresentamos os principais achados da AED com base na base tratada e estruturada em Python.

1. Dimensões e estrutura da base.
 - Total de registros: mais de 10 mil.
 - Total de colunas: 11
 - Principais variáveis: idade, sexo, raça/cor, tipo de óbito, local da ocorrência, assistência médica.
2. Dados ausentes.

Algumas variáveis apresentaram valores nulos ou não preenchidos, principalmente:

- 'ASSISTMED_DESC': registros com valor desconhecido ou vazio (~3%).
- 'CIRCOBITO_DESC': em menor escala, alguns registros ausentes (~1%).

Foram aplicadas técnicas simples de tratamento de dados: exclusão de registros incompletos ou substituição por categoria "Não informado".

3. Estatísticas descritivas.

A seguir, são apresentadas as estatísticas descritivas da variável numérica idade:

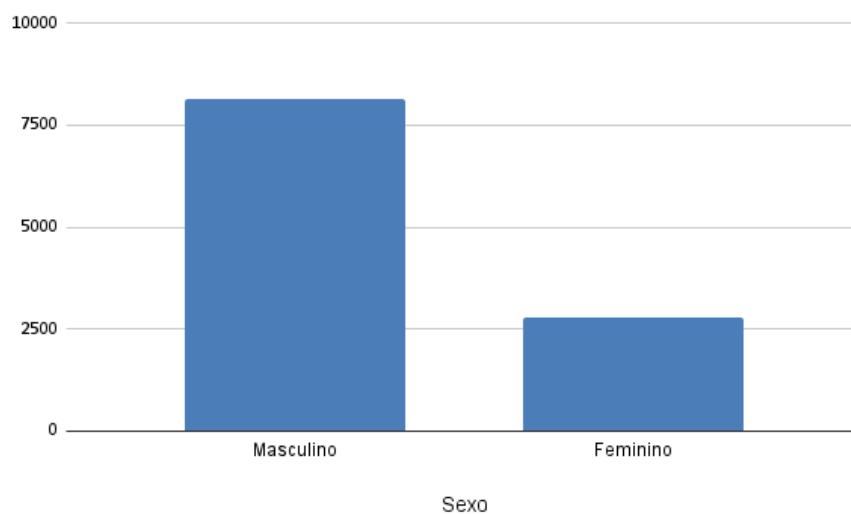
- Variável: Idade.
- Mínimo: 0.

- Máximo: 105.
- Média: 49,92.
- Mediana: 47.
- Desvio padrão: 23,21.

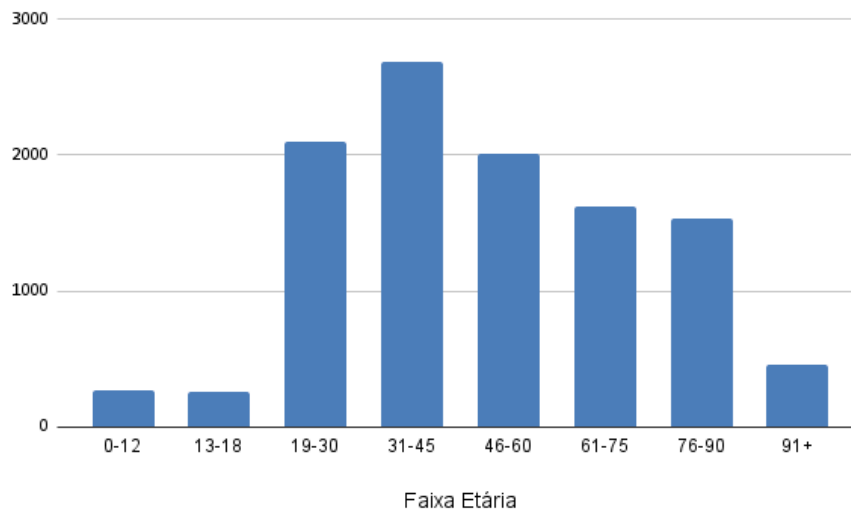
A idade média dos óbitos registrados no estado de São Paulo, em 2023, foi de 49,92 anos, enquanto a mediana foi de 47 anos, indicando uma distribuição assimétrica dos dados. A maior concentração de mortes ocorreu na faixa etária de 31 a 45 anos, evidenciando um padrão de mortalidade significativo entre adultos jovens. O desvio padrão de 23,21 anos reforça a ampla dispersão etária dos registros, com óbitos variando desde recém-nascidos até idosos com mais de 100 anos.

A Análise Exploratória de Dados (AED) identificou alguns padrões relevantes nos registros de óbitos ocorridos no estado de São Paulo em 2023. Esses padrões estão representados a seguir por meio de gráficos de colunas, que facilitam a visualização e interpretação dos principais indicadores observados.

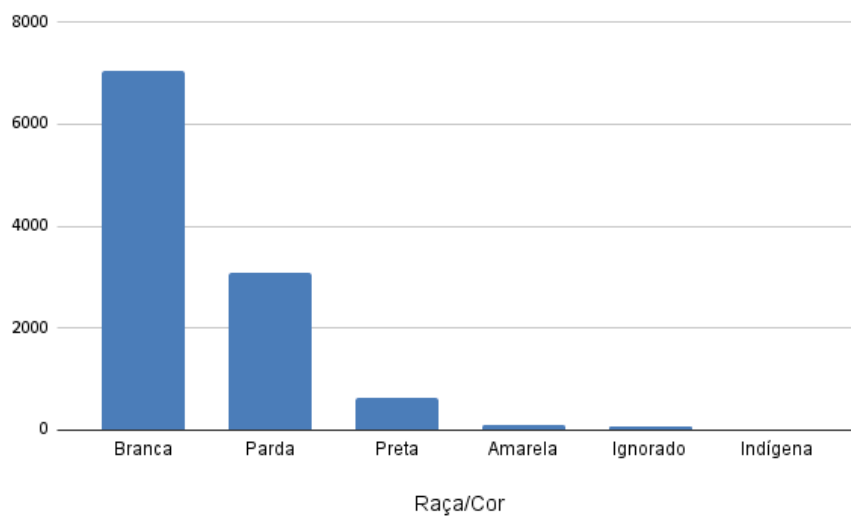
- Maior número de mortes entre pessoas do sexo masculino.



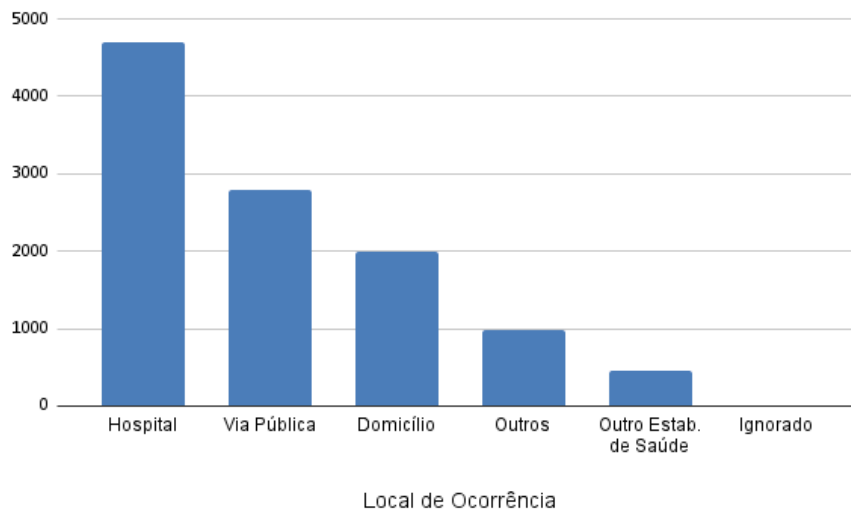
- A faixa etária com maior ocorrência de óbitos foi de 31 a 45 anos. Esse dado chama atenção por indicar uma elevada mortalidade em uma população considerada ainda economicamente ativa e, em tese, com maior acesso aos serviços de saúde.



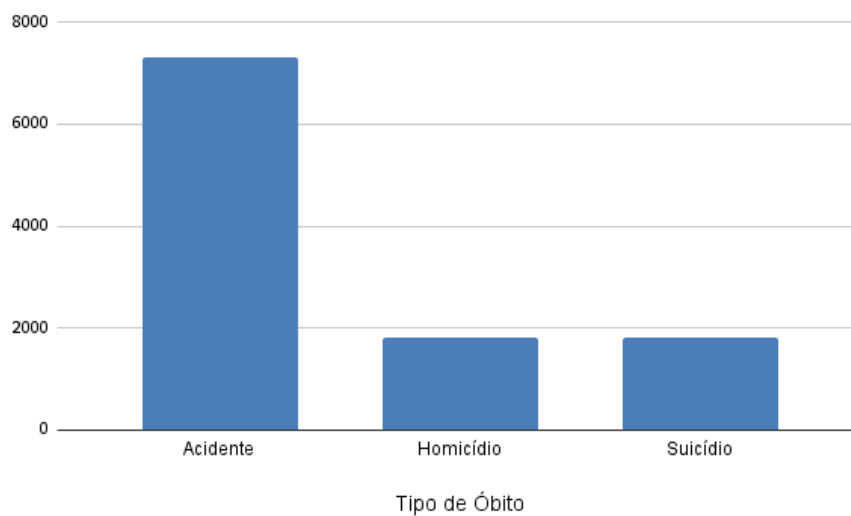
- A variável Raça/Cor com maior índice de mortalidade foram os indivíduos que se autodeclararam brancos, seguidos pela população parda.



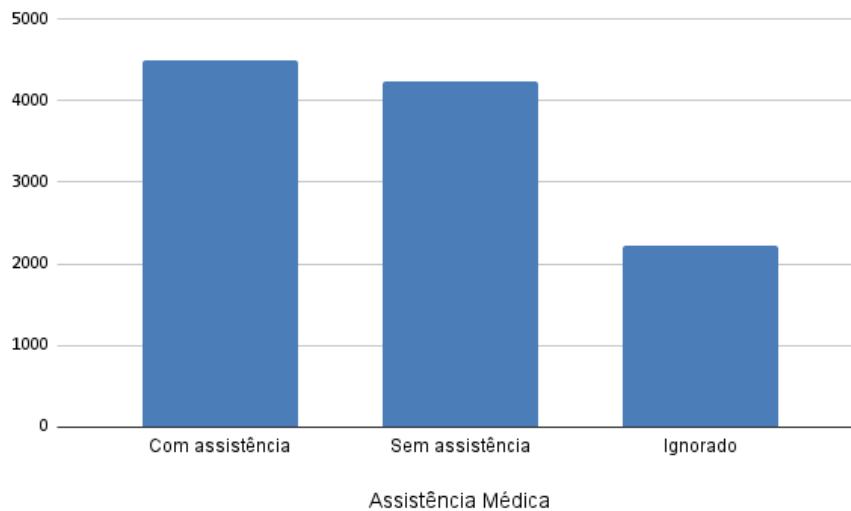
- A maioria dos óbitos foram registrados em hospitais, mas há grande quantidade em vias públicas e domicílios.



- A maioria dos homicídios registrados está relacionada a acidentes, sendo 60% dos casos.



- Outro ponto crítico identificado foi a ausência de assistência médica em parte significativa dos casos — embora a maioria das mortes tenham ocorrido com assistência, os números foram muito próximos dos casos sem assistência, o que evidencia possíveis falhas no acesso aos serviços de saúde.



Proposta Analítica

A proposta analítica foi construída a partir da base tratada com técnicas de ETL e exploração de dados no ambiente Python, por meio de um notebook Jupyter estruturado com as bibliotecas Pandas, Seaborn, Matplotlib e NumPy.

O objetivo principal foi identificar padrões de mortalidade no estado de São Paulo em 2023, considerando variáveis como sexo, raça/cor, faixa etária, local de ocorrência e existência de assistência médica.

➤ Etapas da análise.

O processo analítico seguiu as seguintes etapas:

- Importação das bases
 - Arquivo principal com dados do SIM no formato '.csv'.
 - Base auxiliar com códigos e nomes de municípios (formato '.xlsx').
- Tratamento dos dados (ETL).
 - Conversão e limpeza de colunas de data ('DTOBITO', 'DTNASC').
 - Cálculo da idade dos indivíduos no momento do óbito.
 - Padronização de categorias ('SEXO', 'RAÇACOR', 'LOCAL DE OCORRÊNCIA').
 - Junção da base de municípios para enriquecer os dados com nomes e UF.
 - Filtros para manter apenas registros do estado de São Paulo e dados válidos.
- Criação de novas variáveis.

- 'IDADE', 'ANO_NASC', 'ANO_OBITO', 'MES_OBITO', 'SEXO_DESC', 'RACACOR_DESC', entre outras.
- Estatísticas descritivas.
 - Aplicação do 'describe()' para a variável 'IDADE'.
 - Identificação da faixa etária predominante: 31 a 45 anos.
 - Verificação de outliers (idade 0 e acima de 100 anos).
- Exploração de variáveis categóricas.
 - 'value_counts()' para variáveis como 'SEXO_DESC', 'RACACOR_DESC', 'LOCOCOR_DESC', 'ASSISTMED_DESC'.
- Visualizações com Seaborn e Matplotlib.
 - Gráficos de barras para distribuição por sexo, faixa etária, raça/cor, local do óbito e tipo de assistência.
 - Gráficos salvos para futura inserção em dashboards ou relatórios.

➤ Pipeline de dados aplicado.

O pipeline de dados implementado no notebook foi estruturado conforme o fluxo abaixo, garantindo padronização e qualidade dos dados para a análise:

[1] Aquisição dos dados - Leitura de arquivos: BD_Completa_DATASUS.csv (óbitos) e BD_Municipios.xlsx (municípios) - Definição de colunas úteis e tipos apropriados.

[2] Pré-processamento (ETL) - Conversão de datas de nascimento e óbito - Cálculo da idade - Criação de colunas auxiliares: ano/mês de óbito, ano de nascimento - Mapeamento de códigos para categorias legíveis.

[3] Enriquecimento dos dados - Junção com base de municípios (para nome e UF) - Padronização de categorias: sexo, raça/cor, local, assistência médica, tipo de óbito.

[4] Filtragem dos dados - Foco exclusivo no estado de São Paulo (UF_OCOR = SP) - Remoção de registros com dados inválidos ou ignorados.

[5] Análise Exploratória de Dados (AED) - Estatísticas descritivas da idade - Frequência de variáveis categóricas - Verificação de outliers e valores ausentes.

[6] Visualização - Geração de gráficos de barras, pizza e histogramas com Seaborn e Matplotlib - Preparação para integração futura com dashboards.

[7] Organização e Documentação - Códigos comentados e estruturados em notebook Jupyter - Versão publicada no GitHub para reprodutibilidade.

Ferramentas utilizadas:

- Python 3.x.
 - Pandas para tratamento e manipulação dos dados.
 - Seaborn e matplotlib para visualizações.
 - Numpy para suporte estatístico.
- Jupyter Notebook.
 - Ambiente utilizado para documentação e execução da análise.
- GitHub.
 - Repositório para versionamento e entrega do notebook.

A proposta analítica se apoia integralmente no ambiente Python, explorando os dados com clareza e precisão. O notebook gerado serve como ponto de partida para futuras integrações com ferramentas como Power BI ou Google Data Studio, possibilitando a criação de dashboards e painéis interativos com base nos resultados da análise exploratória.

Link do Projeto no GitHub

O repositório do projeto foi criado para centralizar os materiais desenvolvidos, incluindo o notebook com a Análise Exploratória de Dados em Python. O link é: <https://github.com/PauloRicardo200419/ProjetoMackenzie>.

Considerações Finais

A proposta analítica desenvolvida neste trabalho teve como base os dados do Sistema de Informação sobre Mortalidade (SIM) e foi implementada por meio de um pipeline em Python, com foco na qualidade dos dados, na clareza das variáveis analisadas e na produção de conhecimento relevante sobre os padrões de mortalidade no estado de São Paulo ao longo do ano de 2023.

A Análise Exploratória de Dados (AED) revelou um cenário que contraria algumas percepções comuns sobre os perfis de mortalidade. A maior incidência de óbitos foi registrada na faixa etária de 31 a 45 anos, evidenciando a vulnerabilidade de uma parcela economicamente ativa da população. Além disso, a análise apontou uma maior concentração de mortes entre indivíduos do sexo masculino, com predominância da raça branca, seguida pela população parda.

Outro ponto crítico evidenciado foi a proporção expressiva de óbitos sem assistência médica, o que levanta questionamentos sobre a equidade no acesso aos serviços de saúde. Este dado, aliado à localização dos óbitos (hospitais, domicílios, vias públicas), fornece subsídios para identificar falhas estruturais no atendimento emergencial e preventivo.

A estruturação da análise em Python permitiu um alto nível de controle sobre o processo, desde a leitura e tratamento dos dados até a geração de visualizações significativas. O notebook desenvolvido pode ser utilizado como base para projetos futuros, além de permitir a reprodutibilidade e atualização com novas bases de dados.

A continuidade deste projeto poderá envolver a integração com ferramentas de BI como o Power BI ou Google Data Studio, ampliando o alcance da proposta por meio de dashboards interativos para uso de gestores públicos e pesquisadores.

Com isso, acredita-se que a solução analítica desenvolvida pode apoiar decisões mais estratégicas no âmbito da saúde pública, promovendo políticas mais eficazes e direcionadas a realidades regionais e demográficas específicas.

Referências

- BRASIL. Ministério da Saúde. Sistema de Informação sobre Mortalidade – SIM. Disponível em: <https://apidadosabertos.saude.gov.br/v1/>. Acesso em: abr. 2025.
- MACKENZIE. Slides e materiais da disciplina Projeto Aplicado I. Centro Universitário Presbiteriano Mackenzie, 2025.
- Python Software Foundation. Pandas – Python Data Analysis Library. Disponível em: <https://pandas.pydata.org/>.
- Python Software Foundation. Seaborn Statistical Data Visualization. Disponível em: <https://seaborn.pydata.org/>.
- Hunter, J. D. Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 2007.
- Python Software Foundation. NumPy: Fundamental package for scientific computing with Python. Disponível em: <https://numpy.org/>.



- Project Jupyter. Jupyter Notebook: An open-source web application. Disponível em: <https://jupyter.org/>.
- GitHub. Repositório do Projeto – Análise de Mortalidade no Estado de São Paulo. Disponível em: <https://github.com/PauloRicardo200419/ProjetoMackenzie>.
- Artigos e blogs sobre análise de dados, storytelling e visualização aplicados a projetos de dados abertos e políticas públicas (consultas complementares).