

## **Impacto da Dengue na Saúde Pública em Santos: Uma Análise Exploratória dos Dados das Notificações de Dengue no Período de 2018 a 2023**

Luann Gagliardi  
luann.gagliardi@fatec.sp.gov.br

Paulo Henrique Tavares Delfino de Andrade  
paulo.andrade20@fatec.sp.gov.br

Raphael Imperator Cavicchia Fleury  
raphael.fleury@fatec.sp.gov.br

### **RESUMO**

**Palavras-chave:** Dengue em Santos; Análise de Dados; Saúde Pública.

A dengue é uma doença viral transmitida pelo mosquito *Aedes aegypti* e representa um desafio significativo para a saúde pública em regiões tropicais. Este estudo aborda a situação epidemiológica da dengue na cidade de Santos, São Paulo, entre 2018 e 2023, com foco nas notificações dentro do município (abrangendo pessoas que moram e não moram na cidade de Santos). O objetivo é identificar padrões e tendências de casos da doença, contribuindo para a formulação de políticas públicas mais eficazes. Os dados foram coletados do Sistema de Informação de Agravos de Notificação (SINAN) que estão armazenados no DATASUS. Foi utilizada a técnica de ETL (Extract, Transform, Load), a linguagem de programação Python e o Microsoft Power BI para tratar e visualizar os dados. A análise revelou um pico significativo de casos em 2021, com um aumento de quase 18 vezes em relação ao ano anterior. Os resultados destacam a importância do uso de ferramentas de ciência de dados na análise epidemiológica, permitindo uma compreensão mais aprofundada da proliferação da dengue, essa abordagem facilita a identificação de tendências e anomalias, essenciais para o planejamento e implementação de estratégias de combate à doença. O estudo enfatiza a necessidade de uma abordagem integrada e baseada em dados para enfrentar os desafios da dengue, promovendo ações de saúde pública mais efetivas em Santos.

### **ABSTRACT**

**KEYWORDS:** *Dengue in Santos; Data analysis; Public health.*

Dengue is a viral disease transmitted by the *Aedes aegypti* mosquito and represents a significant public health challenge in tropical regions. This study addresses the epidemiological situation of dengue in the city of Santos, São Paulo, between 2018 and 2023, focusing on notifications within the municipality (covering people who live and do not live in the city of Santos). The objective is to identify patterns and trends in cases of the disease, contributing to the formulation of more effective public policies. The data were collected from the Notifiable Diseases Information System (SINAN) which are stored in DATASUS. The ETL technique (Extract, Transform, Load), the Python programming language and Microsoft Power BI were used to process and visualize the data. The analysis revealed a significant spike in cases in 2021, with an increase of almost 18 times compared to the previous year. The results highlight the importance of using data science tools in epidemiological analysis, allowing a more in-depth understanding of the proliferation of dengue. This approach facilitates the identification of trends and anomalies, essential for planning and implementing strategies to combat the disease. The study emphasizes the need for an integrated, data-based approach to addressing the challenges of dengue, promoting more effective public health actions in Santos.

## 1. INTRODUÇÃO

A dengue é uma doença viral transmitida pela picada do mosquito *Aedes aegypti*, que tem se tornado um problema de saúde pública em diversas regiões tropicais e subtropicais ao redor do mundo. No Brasil, a dengue é uma preocupação constante devido à sua alta incidência e aos surtos frequentes que afetam a população. Ela se caracteriza como uma doença que apresenta sintomas semelhantes aos da gripe, como febre, dor de cabeça, dores musculares e articulares, além de náuseas e vômitos. No entanto, em casos mais graves, a dengue pode evoluir para a dengue hemorrágica, que pode levar à morte se não for tratada adequadamente. Dessa forma, a prevenção e o controle da doença são fundamentais para evitar complicações e reduzir a disseminação do vírus (Luz et al. 2024).

De acordo a Organização Pan-Americana da Saúde (OPAS, 2019), estima-se que cerca de 390 milhões de pessoas são infectadas com o vírus da dengue a cada ano em todo o mundo. Segundo a OPAS (2023), só no ano de 2002 foram registrados, no Brasil, aproximadamente 2.363.490 casos de dengue, representando o maior número de contaminações entre os países sul-americanos. A doença pode se manifestar de diferentes formas, incluindo a dengue clássica, a dengue hemorrágica e a síndrome de choque da dengue, esta última sendo a forma mais grave da doença e potencialmente fatal (Souza-Neto et al. 2019).

Em Santos, cidade litorânea do estado de São Paulo no Brasil, a dengue é uma preocupação constante das autoridades de saúde, devido à presença do mosquito transmissor em áreas urbanas e à alta densidade populacional, fatores que favorecem a propagação da doença. O município tem adotado diferentes estratégias para combater a dengue, como ações de controle do vetor, campanhas de conscientização da população, monitoramento dos casos e investimento em pesquisa e inovação na área da saúde. A dengue em Santos apresenta uma sazonalidade marcante, com picos de incidência durante os meses de verão e outono, quando as condições ambientais são mais propícias à proliferação do *Aedes aegypti*. Isso gera um aumento na demanda por serviços de saúde, sobrecarregando o sistema público e privado e impactando qualidade do atendimento prestado aos pacientes com dengue (Prefeitura de Santos, 2024).

Diante disto, o objetivo deste artigo é analisar a situação epidemiológica da doença na cidade de Santos, entre o período de 2018 e 2023, e seu impacto na saúde pública local. Através da coleta e análise de dados das notificações de casos de dengue, busca-se identificar padrões, tendências e possíveis fatores que contribuam para a disseminação do vírus. A análise exploratória dos dados das notificações de dengue é uma ferramenta fundamental para a compreensão da magnitude do problema na cidade.

A análise desses dados busca contribuir para a elaboração de políticas públicas mais eficazes no enfrentamento da dengue em Santos. A partir dos resultados da análise exploratória dos dados, serão possíveis o desenvolvimento de medidas de controle que possam reduzir a incidência da doença e minimizar seus impactos na saúde da população.

A realização desta pesquisa é fundamental para produzir conhecimento científico que possa subsidiar a tomada de decisões e a implementação de medidas efetivas no combate à Dengue na cidade de Santos. A análise detalhada dos dados disponíveis permitirá uma compreensão mais profunda da situação epidemiológica e contribuirá para a promoção da saúde pública. Para atingir este objetivo será necessário:

- a) Descrever o cenário da propagação de dengue em Santos, sendo os dados obtidos a partir da base de dados do SINAN (Sistema de Informação de Agravos de Notificação), armazenados através do DATASUS (Departamento de Informação e Informática do Sistema Único de Saúde), referindo-se aos dados registrados entre 2018 e 2023. Os dados obtidos incluem: município de notificação (neste caso, exclusivamente Santos), sexo, faixa etária, classificação final (casos confirmados ou não, e se foi dengue comum, alta gravidade de quadro clínico ou grave), evolução (cura ou óbito), sorotipo (variante DEN 1, DEN 2 ou DEN 3 do vírus) e se ocorreu hospitalização por conta da doença;
- b) Estudar técnicas e ferramentas de análises de dados focadas no uso de: ETL; linguagem de programação Python e suas 7 bibliotecas: “Pandas”, “numpy”, “scikitlearn.decomposition.PCA”, “matplotlib.pyplot”, “matplotlib.cm.get\_cmap”, “matplotlib.patches.Patch”, “sklearn.preprocessing.StandardScaler”. As bibliotecas Python apresentadas neste resumo oferecem um conjunto abrangente de ferramentas para análise e visualização de dados, tornando-se recursos valiosos para pesquisadores em diversos campos científicos.

Para executar os processos dos objetivos serão usadas as seguintes ferramentas da área de ciência de dados: linguagem de programação Python para extração e transformação dos dados na plataforma, juntamente com as bibliotecas do Python mencionadas anteriormente.

## 2 FUNDAMENTAÇÃO TEÓRICA

Por meio deste estudo, com finalidade de fornecer *insights* que possam informar políticas públicas de saúde, práticas de planejamento urbano (para impedir a propagação do *Aedes aegypti*) e iniciativas de saúde destinadas a proteger e conscientizar a população de Santos, é necessário compreender conceitos teóricos desta pesquisa e noções importantes na área de ciência de dados.

### 2.1 Complicações da Dengue

Uma das complicações mais comuns da dengue é a dengue grave, também conhecida como dengue hemorrágica. Essa forma mais grave da doença pode se desenvolver rapidamente e é caracterizada por sangramentos, queda abrupta da pressão arterial e danos nos órgãos. A dengue grave pode levar à síndrome de choque da dengue, que é uma condição potencialmente fatal que requer tratamento médico imediato. A síndrome de choque da dengue, ocorre quando o sistema imunológico do paciente reage de forma exagerada à infecção viral. Isso pode levar a um aumento da permeabilidade dos vasos sanguíneos, resultando em choque circulatório e falência de múltiplos órgãos. A síndrome de choque da dengue é uma emergência médica e requer cuidados intensivos para prevenir complicações graves (Leite, 2024).

Outra complicação associada à dengue é a encefalite viral, que ocorre quando o vírus da dengue invade o sistema nervoso central. A encefalite viral pode causar inflamação no cérebro e no sistema nervoso, resultando em sintomas como confusão, convulsões e até coma. Esta é uma complicação rara, mas grave, da dengue e pode levar a danos neurológicos permanentes se não for tratada adequadamente (Reimberg, 2023).

Além das complicações diretas da dengue, a doença também pode desencadear outras condições médicas, especialmente em pacientes com fatores de risco subjacentes. Por exemplo, a dengue pode agravar doenças cardiovasculares pré-existentes, como hipertensão e doença cardíaca, aumentando o risco de complicações cardiovasculares (Leite, 2024).

Além disso, a dengue tem sido associada a complicações respiratórias, como a síndrome da angústia respiratória aguda (SARA), que pode ocorrer em casos graves da doença. A SARA é uma condição grave que causa inflamação nos pulmões e dificuldade respiratória, podendo levar à insuficiência respiratória e morte. Esta complicação é mais comum em pacientes com dengue grave e requer suporte respiratório intensivo (Mendes et al. 2024).

A prevenção e o tratamento das complicações da dengue são fundamentais para reduzir a morbidade e a mortalidade associadas à doença. A vacinação é uma estratégia importante para prevenir a dengue e suas complicações, especialmente em áreas endêmicas, porém não há ainda uma vacina totalmente eficaz e promissora.

Considerando que as vacinas de partículas semelhantes a vírus (VLP) tem alcançado resultados positivos na prevenção de outras doenças contagiosas, é possível questionar se este tipo de vacina não teria uma boa eficácia se aplicada a prevenção da dengue (Reimberg, 2023).

Em suma, a dengue é uma doença viral potencialmente grave que pode levar a uma variedade de complicações, desde sintomas leves até condições que ameaçam a vida. As complicações mais comuns da dengue incluem a dengue grave, a síndrome de choque da dengue, a encefalite viral e as complicações cardiovasculares e respiratórias. A prevenção, o diagnóstico precoce e o tratamento adequado são essenciais para reduzir o impacto das complicações da dengue na saúde pública e na qualidade de vida dos pacientes.

### 2.1.2 Dengue e seus sorotipos

A Prefeitura de Porto Ferreira (2023) aponta, como apresentado na Figura 1, os 4 sorotipos distintos da dengue, sendo possível contrair a doença mais de uma vez pelas outras variantes com as quais a pessoa não teve contato:

**Figura 1 – Dengue: diferentes sorotipos**



Fonte: Prefeitura de Porto Ferreira (2023)

Segundo Barbosa (2022), existem quatro sorotipos diferentes do vírus da dengue, denominados DEN-1, DEN-2, DEN-3 e DEN-4, que podem causar desde formas assintomáticas até quadros graves da doença. Os quatro sorotipos do vírus do dengue são geneticamente semelhantes, porém distintos o suficiente para causar

reinfecções em indivíduos previamente infectados por um sorotipo diferente. Isso se deve à presença de antígenos específicos em cada sorotipo, que geram uma resposta imune adaptativa diferenciada. Dessa forma, a reinfecção por um sorotipo diferente aumenta o risco de desenvolvimento de formas graves da doença devido a uma reação imunológica exacerbada, conhecida como resposta imune mediada por anticorpos. O sorotipo DEN-2 é historicamente associado a formas mais severas de dengue, com complicações como a síndrome de choque do dengue e o sangramento grave. Por outro lado, o sorotipo DEN-4 costuma estar relacionado a formas mais brandas da doença, com menor incidência de complicações graves. O DEN-1 e o DEN-3 também podem causar quadros graves de dengue, porém com menor frequência em comparação com o DEN-2.

A dinâmica de transmissão da dengue é influenciada pela interação entre os diferentes sorotipos virais, os vetores mosquitos e os seres humanos. A coexistência de múltiplos sorotipos em uma mesma área geográfica aumenta o risco de surtos de dengue e complicações clínicas, devido à possibilidade de reinfecções e à diversidade genética dos vírus circulantes. Além disso, a presença de múltiplos sorotipos do vírus do dengue dificulta a implementação de estratégias eficazes de controle e prevenção da doença. As vacinas desenvolvidas para a imunização contra o dengue precisam ser capazes de conferir proteção contra os quatro sorotipos virais, o que representa um desafio considerável devido à alta variabilidade genética do vírus e à necessidade de garantir uma resposta imune equilibrada e duradoura (Reimberg, 2023).

### **2.1.3 Dengue em Santos**

No Brasil, os primeiros casos de dengue foram notificados na década de 1980 e, desde então, a doença tornou-se endêmica no país, causando surtos generalizados e representando uma séria ameaça à saúde pública. O município de Santos, no estado de São Paulo, tem sido particularmente afetado pela propagação da dengue por conta do clima quente e úmido da cidade proporcionar um ambiente ideal para o mosquito *Aedes aegypti* se reproduzir e prosperar, levando a uma alta prevalência da doença entre a população local. Além disso, a rápida urbanização da cidade e a infraestrutura sanitária inadequada contribuíram para a proliferação de criadouros de mosquitos, agravando ainda mais o problema (Fernandes et al., 2014).

Acredita-se que o vírus tenha sido introduzido no Brasil por meio da importação de mosquitos infectados ou pela movimentação de indivíduos infectados de outros países onde a doença já estava presente. Uma vez introduzido, o vírus se espalhou rapidamente por todo o país, encontrando um ambiente adequado para transmissão nas áreas urbanas densamente povoadas do Brasil (Mendonça; Souza; Dutra, 2009).

Segundo Fernandes et al. (2014), em Santos, os primeiros casos de dengue foram notificados no início da década de 1990 e, desde então, a cidade tem vivido vários surtos da doença, com milhares de casos notificados a cada ano. O impacto da dengue na população da cidade tem sido significativo, causando doenças generalizadas, hospitalizações e até mortes. O fardo econômico da doença também tem sido substancial, com o custo do tratamento da dengue e das medidas de prevenção sobrecarregando o sistema de saúde da cidade.

Para combater a propagação da dengue em Santos, as autoridades locais implementaram diversas estratégias destinadas a reduzir as populações de mosquitos e prevenir a transmissão do vírus. Estas estratégias incluem a fumigação dos criadouros do mosquito, a distribuição de inseticidas e campanhas de sensibilização pública para educar a população sobre os riscos da dengue e a importância das



medidas de proteção individual. Apesar desses esforços, a dengue continua sendo um grande problema de saúde pública em Santos, com a cidade enfrentando surtos frequentes da doença, representando uma ameaça ao bem-estar da população local. (Prefeitura de Santos, 2024)

## 2.2 Ferramentas da Área de Ciência de Dados

A ciência de dados, conforme explica Rautenberg e Do Carmo (2019), é uma área interdisciplinar que combina estatística, programação de computadores, matemática e conhecimento de domínio para extrair informações significativas e conhecimento útil a partir de grandes conjuntos de dados. Para o autor, é fundamental compreender que a ciência de dados vai além da simples análise estatística. Ela envolve a utilização de técnicas avançadas de processamento de dados, algoritmos de “*machine learning*” (em português, aprendizado de máquina) e inteligência artificial para identificar padrões, tendências e insights relevantes nos dados. O autor coloca ainda que este processo requer habilidades técnicas, como programação, modelagem estatística e pensamento analítico, bem como conhecimento do domínio em que os dados estão inseridos.

Para este estudo, as ferramentas e processos utilizados serão: ETL (do inglês Extract, Transform and Load); linguagem de programação Python em simultâneo com 7 bibliotecas que serão explicadas posteriormente.

### 2.2.1 ETL

O processo de extração, transformação e carga de dados, conhecido como ETL, desempenha um papel fundamental para as organizações que buscam otimizar a gestão e análise de informações. O ETL consiste em um conjunto de atividades que permitem a extração de dados de fontes diversas, a transformação desses dados em um formato adequado e a carga dessas informações em um destino final (Abreu, 2008).

**Figura 2 – Exemplo das etapas do processo realizado pelo ETL**



Fonte: Correia (2023)

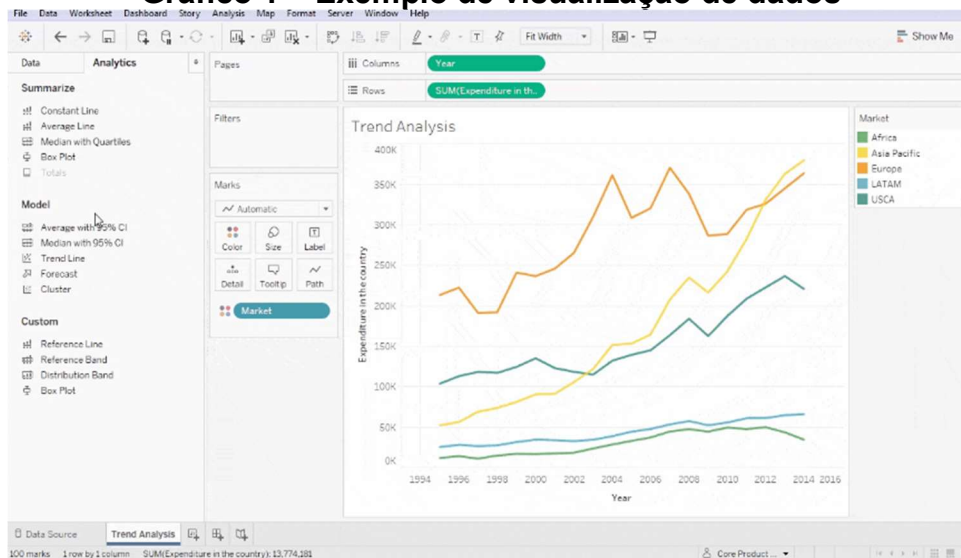
No exemplo apresentado na Figura 2, os dados foram extraídos da fonte, depois transformados para uma forma de análise mais precisa (se por acaso os dados tivessem sido extraídos do Brasil e a pessoa que for analisar e apresentar estes dados está nos Estados Unidos, ela transforma a forma de visualizar a data para deixar a análise mais clara possível) e após isso encaminha para o destino final.

### 2.2.2 Visualização de Dados

A visualização de dados consiste na representação gráfica de informações e dados. Utilizando elementos visuais, como diagramas, gráficos e mapas, essa técnica se torna uma forma acessível de visualizar e compreender exceções, tendências e padrões nos dados (Abreu, 2023).

Segundo Abreu (2023), a visualização de dados representa mais uma forma de expressão visual que desperta interesse e mantém o foco na mensagem que está sendo transmitida. Ao observar um gráfico, torna-se possível identificar tendências e anomalias de forma agradável. Todas as informações visuais disponíveis podem ser assimiladas rapidamente.

**Gráfico 1 – Exemplo de visualização de dados**



Fonte: Abreu (2023)

No modelo apresentado no Gráfico 1, é perceptível a clareza e a densidade de informações presentes em uma única imagem. O gráfico oferece uma análise dos gastos em centenas de dólares de um produto específico por continente. Através desta representação visual, é possível compreender de forma objetiva e concisa os dados apresentados.

### 2.2.3 Linguagem de Programação Python

Conforme explica Santiago et al. (2020), Python foi projetado com o objetivo de enfatizar a legibilidade e a eficiência do código. É conhecido por sua sintaxe limpa e fácil de aprender, tornando-o a escolha ideal tanto para programadores iniciantes quanto para profissionais experientes, sendo uma linguagem de programação ideal para a pesquisa. Um dos principais recursos que diferenciam o Python de outras linguagens de programação é sua capacidade de ser usado em uma ampla gama de aplicações. Segundo ao autor, desde o desenvolvimento web e análise de dados à inteligência artificial e aprendizado de máquina, o Python pode ser utilizado em vários domínios em diferentes setores. Sua extensa biblioteca padrão contém um rico conjunto de módulos e pacotes que fornecem aos desenvolvedores ferramentas para implementar facilmente funcionalidades complexas, economizando tempo e esforço no processo de desenvolvimento. Além disso, o forte suporte do Python para integração com outras linguagens e plataformas o torna uma escolha popular para a construção de aplicativos multiplataforma. Para ser possível toda a análise desta pesquisa, a linguagem Python ao mesmo tempo sua biblioteca foram essenciais. A seguir, as 7 bibliotecas usadas neste estudo:

**pandas:** Para manipulação e análise de dados em formato DataFrame.

**numpy:** Cálculos numéricos eficientes com arrays multidimensionais.

**scikit-learn.decomposition.PCA:** Redução da dimensionalidade de dados para visualização e análise.

**matplotlib.pyplot:** Criação de diversos gráficos e visualizações de dados.

**matplotlib.cm.get\_cmap:** Seleção de mapas de cores para realçar padrões em gráficos.

**matplotlib.patches.Patch:** Personalização de patches de legenda para melhor comunicação visual.

**sklearn.preprocessing.StandardScaler:** Normalização de dados numéricos para melhor desempenho em modelos de aprendizado de máquina.

Para realizar este estudo, é necessário utilizar uma linguagem de programação que pode lidar eficientemente com vastos conjuntos de dados, além de oferecer bibliotecas relevantes para essa finalidade, de maneira acessível e sem custos. Por essa razão, foi decidido por empregar a linguagem de programação Python nas etapas de desenvolvimento do processo de ETL.

### 3 PROCEDIMENTOS METODOLÓGICOS

Conforme explicam Lima e Miotto (2007), procedimentos metodológicos são um termo frequentemente utilizado no meio acadêmico para se referir ao conjunto de passos e etapas que um pesquisador deve seguir ao conduzir um estudo ou pesquisa. Os procedimentos englobam a coleta e análise de dados, sendo a precisão na análise dos dados essencial para garantir a validade e confiabilidade dos resultados obtidos, bem como para permitir que as conclusões da pesquisa sejam fundamentadas e passíveis de replicação.

Para assegurar a precisão na análise dos dados, serão usados: software para as análises dos dados (Power BI) e a linguagem Python (com a biblioteca Pandas) para processar e interpretar as informações coletadas. A combinação de análises descritivas, temporais e estatísticas é fundamental para uma avaliação completa da situação da dengue em Santos.

Os dados utilizados para este estudo foram coletados de fontes oficiais de saúde pública, especificamente dos registros de notificações de dengue fornecidos pelo Sistema de Informação de Agravos de Notificação, que de acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE, 2024), “o SINAN tem por objetivo o registro e processamento dos dados sobre agravos de notificação em todo o território nacional, fornecendo informações para análise do perfil da morbidade e contribuindo, desta forma, para a tomada de decisões”. Abrangendo o período de 2018 a 2023, as bases de dados contêm informações sobre os casos de dengue notificados, incluindo dados mensais, por sexo, faixa etária, classificação final, evolução dos casos, sorotipo e hospitalizações. Os procedimentos metodológicos para análise de dados da dengue



em Santos devem seguir uma série de etapas para garantir a precisão e confiabilidade das informações coletadas. A seguir, são descritos os procedimentos metodológicos adotados de forma eficaz:

1. **Coleta de Dados:** a primeira etapa consiste na coleta de dados relacionados aos casos de dengue em Santos. Estes dados são obtidos diretamente da base de dados do SINAN através do DATASUS.
2. **Organização dos Dados:** após a coleta, os dados são organizados e estruturados de forma adequada para facilitar a análise. Nesta etapa é utilizada a linguagem de programação Python, na plataforma Jupyter e suas bibliotecas: Pandas, numpy, scikitlearn.decomposition.PCA e sklearn.preprocessing.StandardScaler. Alinhando Dados para Análise PCA: Uma Transformação Essencial com Bibliotecas Python, reduzir a dimensionalidade de conjuntos de dados complexos, facilitando a análise de informações relevantes.
3. **Visualização dos Dados:** após a organização, os dados são utilizado a biblioteca: "matplotlib.pyplot", para
4. a criação de *gráficos e visualizações de dados* e tornar possível a visualização clara destes dados para fomentar a análise.
5. **Análise Temporal:** a análise temporal dos dados consiste em identificar variações sazonais e tendências ao longo do tempo. É importante acompanhar a evolução dos casos de dengue em Santos ao longo das estações do ano, a fim de antecipar prováveis surtos e adotar medidas preventivas adequadas.
6. **Modelagem Estatística:** Por fim, uma modelagem estatística dos dados da dengue em Santos permite prever a ocorrência de novos casos e avaliar a eficácia das estratégias de controle adotadas. A utilização de modelos matemáticos e estatísticos pode auxiliar na tomada de decisões fundamentadas em evidências científicas.

### 3.1 Base de Dados

Como apontado anteriormente, foi escolhida a base de dados do SINAN (Sistema de Informação de Agravos de Notificação). Os dados são recolhidos através das: Ficha Individual de Notificação (FIN), sendo preenchidas pelas unidades assistenciais de saúde; e Ficha Individual de Investigação (FII), sendo um meio que possibilita a identificação da fonte de infecção e os mecanismos de transmissão da doença (SINAN, 2017). Estes dados são armazenados pelo DATASUS (Departamento de Informação e Informática do Sistema Único de Saúde), sendo de responsabilidade do Ministério da Saúde (MS) do Governo Federal do Brasil (Ministério da Saúde, 2024a).

Os dados estão disponibilizados publicamente para acesso, sendo possível encontrá-los através da conexão disponibilizada no DATASUS (Ministério da Saúde, 2024b).

Os dados podem ser baixados tanto em arquivo CSV (do inglês, *comma-separated values*), que conforme explica Okada e Bezerra (2024), "de forma simples e objetiva podemos definir um arquivo CSV como sendo um arquivo de dados estruturados separados por vírgula e com a adição de algumas restrições"; quanto para tabwin, que segundo a Secretaria de Estado da Saúde (SES, 2011), é um tabulador de dados desenvolvido pelo DATASUS usado exclusivamente pelas bases de dados dos sistemas de informação da saúde pública do Brasil. Foi optado usar os

arquivos em CSV pela facilidade em todos os softwares conseguirem lê-lo sem precisar de sistema terceirizado, além pelo seu objeto em guardar informações de forma estruturada.

Um empecilho para esta pesquisa acadêmica é que o DATASUS não disponibilizou um *dataset* (do português, conjunto de dados) completo das informações, sendo elas dadas de formas separadas e com dimensões diferentes. Sendo assim, não é possível mesclar os dados do ano em si, apenas mesclar os dados com anos diferentes. Dessa forma, a análise exploratória deste estudo é focada pelas comparações dos dados pelo período de anos (2018-2023).

### 3.2 Tratamento dos Dados

Após os dados coletados por intermédio do DATASUS, é chegada a hora do processo de tratamento e organização destes dados, que de acordo com Gonçalves et al. (2018), estas são as etapas recomendadas para a finalidade desta pesquisa:

- a) **Transformação dos dados:** alterar e organizar os dados para facilitar a análise e compreensão, incluindo limpeza, organização e modificação dos dados conforme necessário;
- b) **Limpeza de dados:** é o processo de identificação e correção de problemas nos dados, como valores ausentes, outliers ou erros de formatação, para garantir que estejam consistentes, precisos e prontos para análise;
- c) **Integração de dados:** envolve a combinação de múltiplas fontes de dados em um único conjunto de dados. Isso pode incluir a fusão de conjuntos de dados separados, eliminação de duplicatas e a reconciliação de diferenças nos esquemas de dados;
- d) **Redução de dados:** é o processo de simplificar e resumir os dados para facilitar a análise e visualização. Isso pode incluir a seleção apenas das variáveis relevantes, agregação de dados para obter informações sumarizadas e a remoção de dados redundantes ou desnecessários.

A limpeza dos dados se deu principalmente pela substituição de valores ausentes: valores ausentes representados por '-' são substituídos por 0 em todos os *DataFrames* lidos, pois a interpretação é de zero paciente.

A integração dos dados foi importante para leitura e mesclagem dos conjuntos de dados: os dados são lidos a partir de vários arquivos CSV usando a função **read\_process\_data**, que lê e processa cada arquivo CSV. Os *DataFrames* resultantes são então mesclados em um único *DataFrame* usando a função **merge\_datasets**. Isso integra os diferentes conjuntos de dados em um único conjunto de dados, permitindo a análise conjunta de várias variáveis.

A redução dos dados: apenas as colunas relevantes para a análise são selecionadas em cada etapa, enquanto as colunas desnecessárias são descartadas. Isso é feito principalmente através da função **reshape\_data**, que seleciona e remodela os dados para cada variável de interesse. As colunas nomeadas de "Total" que aparecem nos bancos de dados foram todas removidas, pois são valores irrelevantes para a organização neste caso, sendo representados posteriormente em *dashboards*.

Renomeação de colunas: a renomeação de colunas foi necessária, pois "Ign/Branco" é presente em quase todos os bancos de dados, então para não ter conflito foi cada um renomeado para sua respectiva tabela, sendo usado o parâmetro

do nome de cada tabela para representá-lo. Então, por exemplo: “Ign/Branco” se torna “Ign/Branco\_Evolução” e assim por diante com os que tem esse tipo de dado disponível. As colunas referentes aos meses também foram alteradas para que não haja necessidade de manipulação destes dados quando transportados para o Power BI, então “Jan”, “Fev” e o restante foram transformados em numéricos de data “01/01/2018”, “01/02/2018” e assim por diante.

Remoção e adição de caracteres nas linhas: foi removido o código dado pelo DATASUS para a cidade de Santos (354850), pois ele é irrelevante para a análise, ficando apenas “SANTOS” na coluna de “Município de notificação”.

Conversão de Tipos de Dados: o código realiza a conversão de tipos de dados usando o método **astype()** do Pandas para garantir que os dados estejam no formato correto para análise.

A função **reshape\_data** foi utilizada para remodelar os dados de acordo com a necessidade de análise. Ela recebe um DataFrame **df** e os nomes das colunas que serão utilizadas como variáveis (**value\_vars**), o nome da coluna que conterá os nomes dessas variáveis (**var\_name**) e o nome da coluna que conterá os valores dessas variáveis (**value\_name**). A função utiliza o método **pd.melt** para derreter o *DataFrame*, transformando as colunas especificadas em variáveis e valores correspondentes. Isso é útil para converter dados de um formato mais amplo para um formato mais longo, facilitando a análise e a visualização. Após a leitura e processamento dos dados, várias manipulações diretas foram realizadas nos *DataFrames* para prepará-los para análise e visualização. Isso incluiu renomear colunas, remover colunas desnecessárias, substituir valores específicos (por exemplo, substituir '-' por 0), preencher valores ausentes, converter tipos de dados e criar novas colunas derivadas (por exemplo, calcular total de cada ano).

### 3.3 Carga dos Dados

Os dados foram importados para a plataforma do Google Colab através dos seguintes passos:

- a) a função **read\_process\_data** foi definida para ler e processar os dados dos arquivos CSV. Esta função utiliza a biblioteca “Pandas” para ler os arquivos CSV e realizar o processamento necessário, como a substituição de valores ausentes por zero;
- b) A função **merge\_datasets** é definida para mesclar os conjuntos de dados lidos de diferentes arquivos CSV. Ela lê e processa cada arquivo CSV usando a função **read\_process\_data**, e depois mescla esses conjuntos de dados em um único *DataFrame* utilizando a função **pd.merge**;
- c) Uma lista de caminhos de arquivo **file\_paths** é definida, contendo os caminhos para cada arquivo CSV que contém os dados a serem importados;
- d) Finalmente, a função **merge\_datasets** é chamada, passando a lista de caminhos de arquivo como argumento, para importar e mesclar os dados de todos os arquivos CSV em um único *DataFrame*, **df\_merged**.

A *análise dos dados* foram realizados pela importação dos dados para o Python em formato dos arquivos em CSV gerados nas transformações pelo uso da biblioteca “Pandas”. A biblioteca “matplotlib.pyplot” oferece uma série de recursos para transformar dados brutos em insights visuais, facilitando a compreensão e a

comunicação dos resultados do “scikit-learn.decomposition.PCA”: Utilizado para reduzir a dimensionalidade de dados para visualização e análise.

## 4 ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

Neste segmento, é realizada uma análise utilizando a Análise de Componentes Principais (PCA) para examinar detalhadamente os casos de dengue na cidade de Santos entre os anos de 2018 à 2023.

A PCA simplifica e interpreta dados multidimensionais de forma eficiente, extraindo componentes principais que capturam a variação nos casos de dengue em Santos. Isso permite uma compreensão clara das ocorrências, identificando padrões e variações distintas na cidade.

### 4.1 Dimensão dos Dados

Número de instâncias: Total de casos registrados de dengue.

Número de atributos: 6 atributos numéricos preditivos, correspondentes aos casos de dengue em seis anos consecutivos (2018, 2019, 2020, 2021, 2022, 2023).

### 4.2 Matriz de Covariância

A matriz de covariância é uma métrica estatística que descreve a interdependência entre as variáveis de um conjunto de dados, revelando as correlações entre suas características. A seguir, apresenta-se a Matriz Quadrada Reduzida da Matriz de Covariância:

$$\begin{bmatrix} 1.2 & 0.25398363 \\ 0.25398363 & 1.2 \end{bmatrix}$$

### 4.3 Autovalores e Autovetores

Os autovalores e autovetores, derivados da matriz de covariância, fornecem insights sobre as principais direções dos dados e suas relevâncias. Os autovetores delineiam as direções principais, ou componentes principais, enquanto os autovalores indicam a variância explicada por cada componente.

**Autovalores:**

$$[1.45398363 \quad 0.94601637]$$

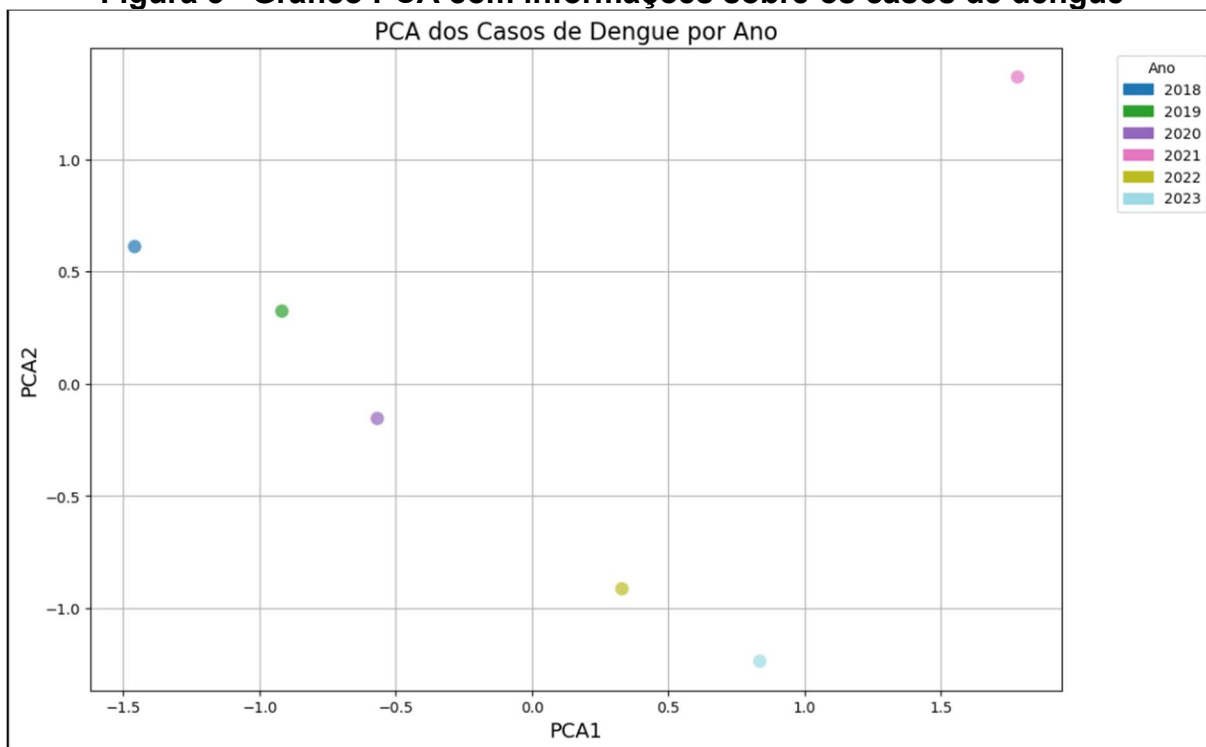
**Autovetores:**

$$\begin{bmatrix} 0.70710678 & -0.70710678 \\ 0.70710678 & 0.70710678 \end{bmatrix}$$

#### 4.4 Desempenho da Análise

Como ilustrado na Figura 3, a plotagem do PCA revela uma distribuição significativa de diversos pontos, indicando a variação dos casos de dengue por ano. Cada ponto representado no gráfico corresponde a um ano, distinguido por uma cor específica, conforme as características calculadas para diferenciar a quantidade anual de casos de dengue.

**Figura 3 –Gráfico PCA com informações sobre os casos de dengue**



Fonte: Elaborado no Python Jupyter pelos autores.

A análise utilizando a “Análise de Componentes Principais (PCA)” revelou padrões distintos nos casos de dengue em Santos ao longo dos anos, com uma diferença notável no padrão de casos no ano de 2021. A redução da dimensionalidade proporcionada pelo PCA facilita a interpretação dos dados, tornando-se uma ferramenta eficaz para a tomada de decisões no contexto da gestão e planejamento de saúde pública.

#### 5 Resultados e considerações finais

A análise dos dados de casos de dengue em Santos, utilizando a Análise de Componentes Principais (PCA), revelou padrões importantes e anomalias significativas entre os anos de 2018 e 2023. O ano de 2021 houve um aumento extraordinário no número de casos de dengue em 2021, evidenciado pela análise do PCA, que destacou uma variação significativa comparada aos outros anos. Esse surto anômalo sugere que fatores extraordinários contribuíram para a alta incidência. Após 2018, observou-se um aumento gradual de casos, culminando em 2021, seguido de uma redução em 2022 e um novo aumento em 2023. Esses dados indicam possíveis



ciclos de infestação e controle, além de eventos específicos que afetaram a dinâmica da dengue na região.

A aplicação do PCA simplificou a complexidade dos dados multidimensionais, permitindo uma visualização clara e interpretável dos padrões de variação anual. Esta técnica revelou-se eficiente para destacar as direções principais dos dados e a variância explicada por cada componente principal.

## REFERÊNCIAS

ABREU, Fábio. Desmistificando o Conceito de ETL. **Revista de Sistemas de Informação**, 2008. Disponível em: [https://www.fsma.edu.br/si/Artigos/V2\\_Artigo1.pdf](https://www.fsma.edu.br/si/Artigos/V2_Artigo1.pdf). Acesso em: 12 abr. 2024.

ABREU, Lima. Guia prático da visualização de dados: definição, exemplos e recursos de aprendizado. Disponível em: <https://www.tableau.com/pt-br/learn/articles/data-visualization>. Acesso em: 12 abr. 2024.

BARBOSA, Manuella de Mello. Diagnóstico diferencial dos sorotipos 1 e 4 do vírus Dengue em mosquitos *Aedes aegypti* por meio da espectroscopia no infravermelho próximo (NIRS). 2022. Tese de Doutorado. Disponível em: <https://www.arca.fiocruz.br/handle/icict/62638>. Acesso em: 05 maio 2024.

CORREIA, Lays Leal. ETL automatizado para sistema de monitoramento remoto de baterias de chumbo-ácido. 2023. Disponível em: <https://repositorio.ifpe.edu.br/xmlui/handle/123456789/1136>. Acesso em: 12 abr. 2024.

FERNANDES, Maria Angela Bellegarde; NATAL, Delsio; DE FÁTIMA DOMINGOS, Maria. ASPECTOS EPIDEMIOLÓGICOS DA TRANSMISSÃO DE DENGUE EM SANTOS, SÃO PAULO, NO PERÍODO DE 1997 A 2012. **Journal of Health & Biological Sciences**, v. 2, n. 1, p. 5-5, 2014. Disponível em: <https://periodicos.unichristus.edu.br/jhbs/article/view/45/45>. Acesso em: 05 maio 2024.

GONÇALVES, Tayná Costa; DA SILVA, Josenildo Costa; CORTES, Omar Andres Carmona. Técnicas de mineração de dados: um estudo de caso da evasão no ensino superior do Instituto Federal do Maranhão. **Revista Brasileira de Computação Aplicada**, v. 10, n. 3, p. 11-20, 2018. Disponível em: <https://seer.upf.br/index.php/rbca/article/view/8427/114114337>. Acesso em: 05 maio 2024.

Instituto Brasileiro de Geografia e Estatística – IBGE. Sistema de Informações de Agravos de Notificação – SINAN. 2024. Disponível em: <https://ces.ibge.gov.br/base-de-dados/metadados/ministerio-da-saude/sistema-de-informacoes-de-agravos-de-notificacao-sinan>. Acesso em: 05 maio 2024.

LEITE, André Matheus Carvalho Silva et al. REVISÃO DAS PRINCIPAIS COMPLICAÇÕES DA DENGUE. **Brazilian Journal of Implantology and Health Sciences**, v. 6, n. 3, p. 167-175, 2024. Disponível em: <https://bjih.emnuvens.com.br/bjih/article/view/1508>. Acesso em: 04 maio 2024.

LIMA, Telma Cristiane Sasso de; MIOTO, Regina Célia Tamasso. Procedimentos metodológicos na construção do conhecimento científico: a pesquisa bibliográfica. **Revista katálisis**, v. 10, p. 37-45, 2007. Disponível em: <https://www.scielo.br/j/rk/a/HSF5Ns7dkTNjQVpRyvhc8RR/?format=pdf&lang=pt>. Acesso em: 05 maio 2024.

LUZ, Dirce Ferreira et al. A educação ambiental e seu impacto no combate aos vetores transmissores de doenças. **Revista Pantaneira**, v. 23, p. 88-97, 2024. Disponível em: <https://periodicos.ufms.br/index.php/revpan/article/view/20276>. Acesso em: 04 maio 2024.

LOPES, Gesiel Rios et al. Introdução à análise exploratória de dados com python. **Minicursos ERCAS ENUCMPI**, v. 2019, p. 160-176, 2019. Disponível em: [https://www.researchgate.net/publication/336778766\\_Introducao\\_a\\_Analise\\_Exploratoria\\_de\\_Dados\\_com\\_Python](https://www.researchgate.net/publication/336778766_Introducao_a_Analise_Exploratoria_de_Dados_com_Python). Acesso em: 07 maio 2024.

MENDES, Marcela da Silva et al. Síndrome da angústia respiratória aguda associada à varicela em paciente adulto: exemplo de suporte respiratório extracorpóreo em doenças endêmicas brasileiras. **Revista brasileira de terapia intensiva**, v. 26, p. 410-415, 2014. Disponível em: <https://www.scielo.br/j/rbti/a/SmmKJ5zbsWqDg8FNLm3YWWx/#>. Acesso em: 04 maio 2024.

MENDONÇA, Francisco de Assis; SOUZA, Adilson Veiga; DUTRA, Denecir de Almeida. Saúde pública, urbanização e dengue no Brasil. **Sociedade & natureza**, v. 21, p. 257-269, 2009. Disponível em: <https://www.scielo.br/j/sn/a/tRqQNr3nLXBNvqV3MpZGvhP/?format=pdf&lang=pt>. Acesso em: 05 maio 2024.

Ministério da Saúde - MS. TabNet: Datasus. 2024a. Disponível em: <https://datasus.saude.gov.br/informacoes-de-saude-tabnet/>. Acesso em: 05 maio 2024.

Ministério da Saúde – MS. Dengue - Notificações Registradas No Sistema De Informação De Agravos De Notificação - São Paulo. 2024b. Disponível em: <http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sinannet/cnv/denguebsp.def>. Acesso em: 05 maio 2024.

OKADA, Renan Shindi; BEZERRA, Wesley Dos Reis. MANIPULADOR DE ARQUIVOS CSV NA LINGUAGEM JAVA, MONGODB E DOCKER. **Anais da Feira do Conhecimento Tecnológico e Científico**, n. 24, 2024. Disponível em: <https://publicacoes.ifc.edu.br/index.php/fetec/article/view/5142/4315>. Acesso em: 05 maio 2024.

Organização Pan-Americana da Saúde – OPAS. Dez ameaças à saúde que a OMS combaterá em 2019. 2019. Disponível em: <https://www.paho.org/pt/noticias/17-1-2019-dez-ameacas-saude-que-oms-combatera-em-2019>. Acesso em: 03 maio 2024.

Organização Pan-Americana da Saúde – OPAS. Actualización epidemiológica semanal para dengue, chikunguña y zika en 2022. 2023. Disponível em: [https://ais.paho.org/ha\\_viz/arbo/pdf/OPS%20Arbo%20Boletin%202022.pdf](https://ais.paho.org/ha_viz/arbo/pdf/OPS%20Arbo%20Boletin%202022.pdf). Acesso em: 03 maio 2024.

Prefeitura de Santos. Dengue, Zika e Chikungunya: o combate é o ano todo! 2024. Disponível em: <https://www.santos.sp.gov.br/?q=hotsite/dengue-zika-e-chikungunya>. Acesso em: 03 maio 2024.

Prefeitura de Porto Ferreira. Dengue: sorotipos diferentes circulam na região e solução é eliminar água parada. 2023. Disponível em: <https://www.portoferreira.sp.gov.br/noticias/saude/dengue-sorotipos-diferentes-circulam-na-regiao-e-solucao-e-eliminar-agua-parada>. Acesso em: 04 maio 2024.

RAUTENBERG, Sandro; DO CARMO, Paulo Ricardo Viviurka. Big data e ciência de dados: complementariedade conceitual no processo de tomada de decisão. **Brazilian Journal of Information Science**, v. 13, n. 1, p. 56-67, 2019. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=6983493>. Acesso em: 12 abr. 2024.

REIMBERG, Ariane Aparecida et al. Desenvolvimento de VLPs (vírus like particle) como potencial vacina recombinante contra o vírus da dengue (DENV). **Research, Society and Development**, v. 12, n. 13, p. e92121344384-e92121344384, 2023. Disponível em: <https://rsdjournal.org/index.php/rsd/article/view/44384>. Acesso em: 04 maio 2024.

SANTIAGO, Cynthia Pinheiro et al. Desenvolvimento de sistemas Web orientado a reuso com Python, Django e Bootstrap. **Sociedade Brasileira de Computação**, 2020. Disponível em: <https://sol.sbc.org.br/livros/index.php/sbc/catalog/download/48/219/457-1?inline=1>. Acesso em: 12 abr. 2024.

SANTOS, Iris Cristina dos. Plataforma Microsoft Power BI: Estudo de caso da utilização pela secretaria de saúde do estado do Espírito Santo para gestão da pandemia do COVID-19. 2020. Disponível em: [https://repositorio.ifes.edu.br/bitstream/handle/123456789/849/ARTIGO\\_PLATAFORMA\\_%20MICROSOFT\\_%20POWER%20\\_BI%20\\_ESTUDO%20\\_DE%20\\_CASO\\_%20ODA%20\\_UTILIZAÇÃO.pdf?sequence=1&isAllowed=y](https://repositorio.ifes.edu.br/bitstream/handle/123456789/849/ARTIGO_PLATAFORMA_%20MICROSOFT_%20POWER%20_BI%20_ESTUDO%20_DE%20_CASO_%20ODA%20_UTILIZAÇÃO.pdf?sequence=1&isAllowed=y). Acesso em: 12 abr. 2024.

Secretaria de Estado da Saúde - SES. Apresentando o tabwin e o tabnet. 2011. Disponível em: [https://www.saude.sp.gov.br/resources/ses/perfil/gestor/homepage/auditoria/reunioes/2011/tabwin\\_funcionalidades\\_acesso\\_e\\_uso\\_da\\_ferramenta.pdf](https://www.saude.sp.gov.br/resources/ses/perfil/gestor/homepage/auditoria/reunioes/2011/tabwin_funcionalidades_acesso_e_uso_da_ferramenta.pdf). Acesso em: 05 maio 2024.

Sistema de Informação de Agravos de Notificação – SINAN. Funcionamento. 2017. Disponível em: <https://portalsinan.saude.gov.br/funcionamentos>. Acesso em: 05 maio 2024.

SOUZA-NETO, Jayme A.; POWELL, Jeffrey R.; BONIZZONI, Mariangela. Aedes aegypti vector competence studies: A review. **Infection, genetics and evolution**, v. 67, p. 191-209, 2019. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1567134818307159?via%3Dihub>. Acesso em: 03 maio 2024.