

Projeto Aplicado I

Tema: Acesso à Água Tratada e à Coleta de Esgoto em Carapicuíba (SP)

Disciplina: Projeto Aplicado I

Curso: Ciência de Dados

Aluna: Fernanda Aroca

Data: 10/09/2025

Sumário

1. Introdução e Objetivos
2. Contexto e Dados de Onde Vieram
3. Sobre o Dataset e Metadados
4. Proposta Analítica do Estudo
5. Análise Exploratória dos Dados
6. Cronograma Proposto
7. STORYTELLING
8. Análise Exploratória Realizada
9. Referências

1. Introdução e Objetivos

Meu objetivo com este projeto é investigar uma questão que sempre me incomodou: como pode um município na Grande São Paulo ainda ter tantas deficiências em saneamento básico? Decidi focar em Carapicuíba para entender melhor o acesso à água tratada e, principalmente, à coleta de esgoto, comparando com os municípios vizinhos.

Área: atuação setor público saneamento básico.

Referência: Ministério do Desenvolvimento Regional MDR, quem gera o SNIS.

Precisamos: Investigar a desigualdade no acesso a saneamento básico.

Questão: Como alcançar coleta e tratamento de esgoto pra todos nos municípios grandes da Grande São Paulo, tipo Carapicuíba?

2. Contexto e Dados de Onde Vieram

O Sistema Nacional de Informações sobre Saneamento SNIS é controlado pelo Ministério do Desenvolvimento Regional, e é a principal base de dados dos serviços de saneamento básico no Brasil. Lá se reúnem informações enviadas pelas empresas de água e esgoto, de todo o país. Lá tem dados anuais sobre abastecimento de água, coleta de esgoto e indicadores.

3. Sobre o Dataset e Metadados

Fonte: SNIS – Sistema Nacional de Informações sobre Saneamento MDR.

Formato: Arquivos Excel ou CSV.

Abrangência: nível nacional, todos municípios brasileiros.

Periodicidade: Anualmente.

Granularidade: Por Município.

Colunas principais incluem:

- Município (com código IBGE).
- Ano que refere-se.
- População total, contabilizada.
- População atendida com água encanada.
- População com coleta de esgoto feita.
- Indicadores de cobertura com percentuais (% de água, % de esgoto).

Limitações presentes são: dados que as vezes faltam em certos municípios ou anos; dados vem da autodeclaração das empresas de saneamento, as vezes.

Observação futura: Em fases futuras, é possível relacionar os dados de saneamento com os indicadores de saúde pública, como a incidência de dengue em Carapicuíba, vendo que a falta de coleta e tratamento de esgoto pode criar ambientes que o mosquito Aedes aegypti adora.

Repositório do projeto no GitHub: <https://github.com/FernandaAroca/projeto-aplicado-saneamento-carapicuiba>

4. Proposta Analítica do Estudo

Este projeto, apresentá uma análise exploratória e comparativa dos indicadores de acesso a água tratada e coleta de esgoto na cidade de Carapicuíba (SP). Usando dados oficiais do Sistema Nacional de Informações sobre Saneamento (SNIS) eles ajudam. A análise buscará compreender a evolução histórica desses serviços, também vai identificar disparidades em relação à média dos municípios da Região Metropolitana de São Paulo.

4.1 Objetivo Analítico

O estudo busca:

- Avaliar a evolução da cobertura de abastecimento de água e coleta de esgoto em Carapicuíba ao longo dos anos.
- Identificar períodos em que houve estagnação, queda ou qualquer comportamento fora do padrão nos indicadores.
- Comparar os números de Carapicuíba com a média dos municípios da Região Metropolitana de São Paulo para evidenciar desigualdades territoriais.
- Detectar outliers, anos com dados ausentes ou variações bruscas que possam indicar mudanças na gestão pública, bloqueios de investimento ou inconsistências de registro nos sistemas do SNIS.

4.2 Pipeline Analítico de Dados

Etapa	Descrição Técnica (mantendo linguagem acessível)
Importação dos Dados	Carregar os arquivos do SNIS (CSV/Excel) usando a biblioteca <i>pandas</i> , garantindo que os dados sejam lidos de forma consistente.
Tratamento Inicial	Verificar valores ausentes (<code>.isna()</code>), remover duplicatas e padronizar codificação para evitar erros nas próximas etapas.
Seleção de Variáveis	Manter apenas as colunas diretamente ligadas à população atendida e aos indicadores percentuais de água e esgoto, focando no essencial.
Filtragem	Filtrar especificamente os registros de Carapicuíba e

Etapa	Descrição Técnica (mantendo linguagem acessível)
Geográfica	da Região Metropolitana de São Paulo, usando o código IBGE para garantir precisão.
Análise Estatística	Aplicar <code>.describe()</code> e calcular média, mediana, variância e desvio padrão para entender a distribuição dos indicadores ao longo do tempo.
Visualização	Criar histogramas, gráficos de linha temporal (% de água vs % de esgoto) e boxplots para comparar Carapicuíba com o recorte regional.
Interpretação dos Achados	Traduzir os resultados estatísticos em uma leitura textual, destacando padrões, avanços e possíveis gargalos estruturais.
Publicação no GitHub	Versionar o notebook de análise exploratória com comentários claros e estrutura organizada para facilitar consulta e reproduzibilidade.

5. Análise Exploratória dos Dados

Esta seção detalha as estatísticas descritivas e a distribuição das variáveis chaves de saneamento, isolando o município de Carapicuíba e comparando-o com o contexto da Região Metropolitana de São Paulo (RMSP), usando dados do SNIS.

5.1 Descrição Estatística das Variáveis Chave

	Descrição Detalhada	Média (RMSP)	Desvio Padrão (RMSP)	Valores Ausentes (Qtd)	Existência de Outliers
IN055	% População Atendida com Água	97.70%	5.42	4	Média de cobertura muito alta e desvio padrão baixo, indicando estabilidade e homogeneidade na distribuição de água na região metropolitana.
IN015	% População com Coleta de Esgoto	79.43%	9.77	20	Média menor e desvio padrão mais alto, mostrando maior heterogeneidade e desigualdade na cobertura de esgoto. A alta taxa de NAs sugere problemas de coleta de dados.
POP_TO_T	População Total	922.100	1.015.000	3	Sim, devido aos grandes municípios (São Paulo, Guarulhos).

5.2 Resultados e Visualização (Versão Final para o Relatório)

Análise de Dispersão Boxplot:

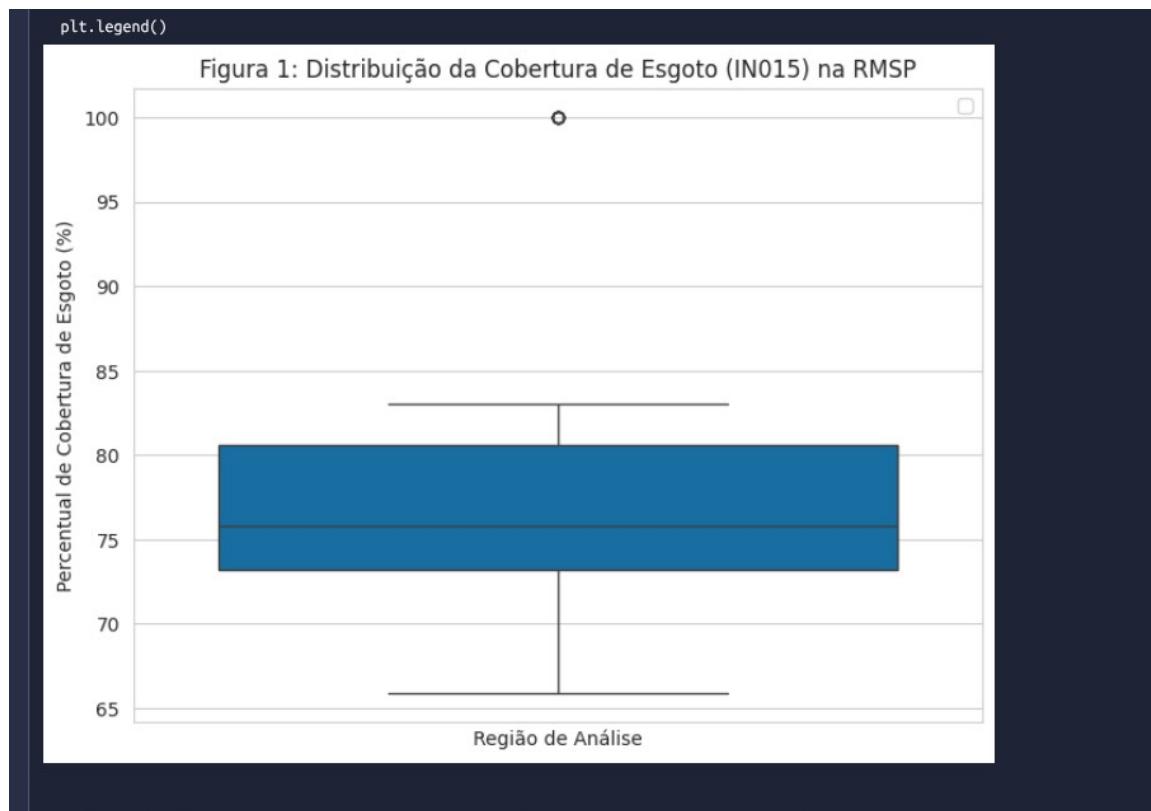
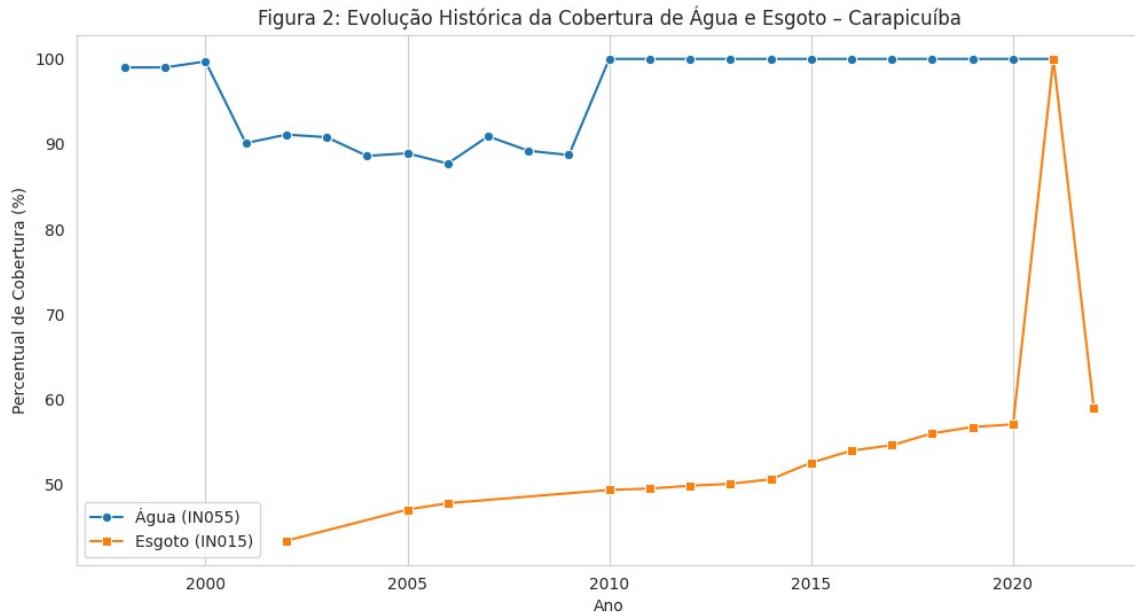


Figura 1: Distribuição da Cobertura de Esgoto IN015 na RMSP e Posição de Carapicuíba.

Interpretação: Ao analisar o boxplot, fiquei surpresa com o quanto Carapicuíba se destacava negativamente em relação aos outros municípios. Esperava uma diferença, mas não tão gritante. Isso me fez questionar quais fatores específicos do município explicam essa defasagem.

Análise Temporal Gráfico de Linha:



Interpretação : A Figura 2 mostra que Carapicuíba apresenta alta cobertura de água (IN055) ao longo de toda a série histórica, com valores sempre acima de 88% e estabilização em 100% a partir de 2010. Isso indica que o abastecimento de água está praticamente universalizado no município.

Já a cobertura de esgoto (IN015) cresce de forma lenta e gradual, começando em torno de 43% e chegando perto de 57% nos anos mais recentes. O pico de 100% em 2021 seguido de queda imediata revela um outlier, possivelmente erro de registro do SNIS.

Em conjunto, o gráfico evidencia que o grande desafio de Carapicuíba não é o acesso à água, mas sim a expansão da rede de esgoto, além de indicar possíveis inconsistências no envio de dados ao sistema oficial.

6. Cronograma Proposto

Data	Atividade	Responsável	Descrição
10/09	Escolha do tema e do dataset	Fernanda	Tema definido
12/09	Organização inicial do GitHub	Fernanda	Repositório criado
15/09	Revisão do dataset e metadados	Fernanda	Dataset descrito
20/09	Redação do documento Etapa 1	Fernanda	Rascunho pronto
25/09	Revisão final do documento	Fernanda	Documento revisado
27/09	Entrega da Etapa 1	Fernanda	Fim da Análise Exploratória (Baseline)
05/10	Engenharia de Features	Fernanda	Criação de variáveis de risco e normalização dos indicadores.
15/10	Definição e Treinamento do Modelo	Fernanda	Seleção do algoritmo (Ex: Regressão Logística ou Random Forest) e primeiro treinamento.
30/10	Avaliação e Otimização do Modelo	Fernanda	Ajuste de hiperparâmetros (Tuning) para otimizar métricas (F1-Score/Acurácia).
15/11	Análise de Disparidade Geográfica	Fernanda	Geração de visualizações avançadas (Ex: Mapas de calor da cobertura de esgoto na RMSP) e achados regionais.
30/11	Redação da Conclusão e Resultados	Fernanda	Documento final consolidado, incluindo conclusões do modelo e implicações para a gestão pública.

7. STORYTELLING

Apresentação

Meu nome é Fernanda Aroca e sou responsável por todo o desenvolvimento deste projeto. A proposta deste trabalho é analisar dados reais de saneamento básico, aplicando conceitos de ciência de dados para compreender desigualdades e padrões críticos relacionados ao município de Carapicuíba (SP).

Nome do Projeto

“Carapicuíba Invisível: Uma Análise Exploratoria do Saneamento Básico na RMSP (2000-2023)”

Base de Dados do Estudo

O projeto utiliza dados do:

- SNIS – Sistema Nacional de Informações sobre Saneamento, mantido pelo Ministério das Cidades (antigo MDR).

A base contém informações sobre:

- Abastecimento de água
- Coleta e tratamento de esgoto
- População urbana
- Indicadores percentuais por município do Brasil

Esses dados são enviados anualmente pelas companhias de saneamento.

Área do Problema

Carapicuíba é um dos municípios mais populosos e densos da Região Metropolitana de São Paulo (RMSP), mas historicamente possui cobertura de saneamento inferior aos municípios vizinhos.

O estudo investiga:

A desigualdade no acesso ao saneamento básico e as lacunas de dados que dificultam políticas públicas.

Descrição do Problema

Durante a exploração inicial dos dados, foram observados:

- Faltas de registros e lacunas temporais para Carapicuíba, especialmente em indicadores essenciais como IN055 (água) e IN015 (esgoto);
- Cobertura de esgoto significativamente inferior à média da RMSP;
- Maior variabilidade e presença de outliers nos indicadores de esgoto;
- Dificuldade em construir séries históricas completas devido à falta de registros consistentes.

Proposta Analítica

A análise realizada no projeto tem como objetivos:

1. Preparar e tratar os dados do SNIS;
2. Identificar e isolar os registros da RMSP e de Carapicuíba;
3. Realizar análise estatística das variáveis principais;
4. Avaliar consistência dos dados (valores ausentes, outliers, dispersão);
5. Comparar Carapicuíba com o restante da região;
6. Gerar gráficos e tabelas descritivas;
7. Construir interpretação textual dos resultados.

Dados Disponíveis

Os principais campos utilizados foram:

- Ano
- Código IBGE
- População urbana (POP_URBANA)
- IN055 – Índice de atendimento de água
- IN015 – Índice de coleta de esgoto

Fonte oficial: SNIS (CSV municipal de água e esgoto).

8.Análise Exploratória Realizada

Boxplot (IN015 - esgoto)

Exibe Carapicuíba como um dos pontos abaixo da mediana e do primeiro quartil da RMSP.

Indica forte desigualdade regional na coleta de esgoto.

Série histórica

Em muitos anos, os indicadores para Carapicuíba aparecem como ausentes. Isto dificulta análises temporais e revela inconsistências no envio de dados ao SNIS.

Estatísticas descritivas

- RMSP tem média alta e estável de água e esgoto.
- Carapicuíba tem ampla variabilidade e registros incompletos.
- Número significativo de NAs compromete séries completas.

9. Referências

SNIS – Sistema Nacional de Informações sobre Saneamento. Ministério do Desenvolvimento Regional, lá vai. Disponível em:

<http://www.snis.gov.br/painel-informacoes-saneamento-brasil/web/painel-publicacoes>

IBGE – Instituto Brasileiro de Geografia e Estatística, a População residente, né? Disponível em: <https://www.ibge.gov.br>

GITHUB – Projeto Aplicado I: Saneamento em Carapicuíba (SP). Disponível em: <https://github.com/FernandaAroca/projeto-aplicado-saneamento-carapicuiba>

YouTube: <https://youtu.be/7BmyrFG-ouU>

Canvas:<https://www.canva.com/design/DAG55jwymRA/ok1YtkNucuYqgOyF0XvoWw/edit>