

TCC RMB Saneamento+Saúde (MVP COP30)

Projeto: Console de decisão para priorização de investimentos em água, esgoto e saúde na Região Metropolitana de Belém (RMB)

Equipe: (inserir nomes e papéis)

Versão: 1.0

Data: 27/10/2025

1) Sumário Executivo

Este dossiê consolida a proposta do TCC “**RMB Saneamento+Saúde**”, um **MVP** (produto mínimo viável) orientado à tomada de decisão pública, que integra dados oficiais (SISAGUA, SNIS/SINISA, SIH/DATASUS, INMET e IBGE) para **prever, simular e priorizar** investimentos com potencial de reduzir **internações por doenças infecciosas intestinais (CID-10 A00–A09)**. O projeto inclui **wireframe de dashboard, catálogo de dados públicos e plano de engenharia de dados (ETL bronze→silver→gold)**, com **rastreabilidade às ODS 6/3/11**.

2) Contexto e Motivação

- **COP30/Belém:** obras de saneamento (canais) e a **ETE Una** inaugurada em 2025 criam uma janela de priorização baseada em evidências.
 - **Novo arranjo institucional:** separação entre produção (COSANPA/Complexo Bolonha) e distribuição (Águas do Pará/Aegea) demanda ferramentas de coordenação e transparência.
 - **Sustentabilidade e Saúde:** água e esgoto adequados reduzem agravos (A00–A09), gerando ganhos sociais e econômicos.
-

3) Problema e Oportunidade

- **Problema:** cobertura de esgoto insuficiente, perdas elevadas de água e não conformidades de qualidade, somados à sazonalidade climática, resultam em internações evitáveis.

- **Oportunidade:** combinar dados públicos e técnicas de IA/ML para quantificar **impacto sanitário** e **priorizar** ações de maior retorno.
-

4) Objetivos

1. **Integrar** dados abertos (RMB) em um **painel preditivo e prescritivo**.
 2. **Estimar** o efeito de melhorias operacionais (coleta, tratamento, perdas, qualidade) sobre **A00–A09/100 mil**.
 3. **Simular** cenários “what-if” e **ranquear** a “próxima melhor ação” por município/sistema.
 4. **Rastrear ODS** 6/3/11 com indicadores e metas verificáveis.
 5. **Entregar** um **MVP navegável** e documentação reproduzível (ETL + dicionário + scripts).
-

5) Escopo e Recortes

- **Âmbito:** municípios da **RMB** (Belém, Ananindeua, Marituba, Benevides, Santa Bárbara do Pará, Santa Izabel do Pará, Castanhal, Barcarena).
 - **Período:** 2014–2025 (ajustável por fonte); **Censo 2022** como base estrutural.
 - **Granularidade:** municipal e/ou por sistema (quando disponível no SISAGUA).
 - **Desfecho principal:** **A00–A09/100 mil** (mensal).
 - **Restrições:** foco em séries temporais e painéis; **sem geoprocessamento pesado**.
-

6) Alinhamento às ODS (Quadro de Rastreabilidade)

ODS 6 — Água Limpa e Saneamento

- **6.1** Acesso à água: **% domicílios com rede de água** (IBGE SIDRA t/6803)
- **6.2** Saneamento: **% domicílios com coleta de esgoto** (IBGE SIDRA t/6805; apoio SNIS/SINISA)
- **6.3** Qualidade da água: **% amostras fora do padrão** (SISAGUA – cloro, turbidez, microbiologia); **% esgoto tratado** (SNIS/SINISA)
- **6.4** Eficiência: **Perdas na distribuição** (SNIS/SINISA)

ODS 3 — Saúde e Bem-Estar

- **3.3/3.9** Doenças transmissíveis / riscos ambientais: **Taxa A00–A09 por 100 mil** (SIH/TABNET; denominador: Estimativas Populacionais IBGE)

ODS 11 — Cidades Sustentáveis (complementar)

- **11.6** Impacto ambiental urbano: **% domicílios com coleta/destino do lixo** (IBGE SIDRA t/6892)

A matriz ODS→KPI→Fonte será exibida no **ODS Tracker** do dashboard e reproduzida no relatório técnico.

7) Perguntas de Negócio

1. Quais **ações** (aumentar coleta, elevar % tratado, reduzir perdas, melhorar conformidade de qualidade) **reduzem mais** as internações **A00–A09** na RMB?
2. Quais **municípios/sistemas** geram o **maior retorno sanitário** por ponto percentual de melhoria?
3. Como a **chuva** (e seus *lags*) modula o risco ao longo do ano?
4. Qual o **ranking de priorização** de investimentos para os próximos 12 meses?

5. Como medir e reportar o progresso nas ODS 6/3/11?

8) Bases de Dados Públicas (acesso livre)

Todos os links são públicos e acessíveis a qualquer cidadão. As coletas serão automatizadas onde houver API/CKAN.

8.1 Saneamento — operação e cobertura

SINISA / SNIS – Água e Esgotos (Ministério das Cidades)

- **Finalidade:** cobertura de água/esgoto; % esgoto tratado; **perdas**; volumes; indicadores econômico-financeiros.
- **Cobertura:** Brasil; **anual**; séries por município/prestador.
- **Campos-chave:** `%_atendimento_agua`, `%_coleta_esgoto`,
`%_esgoto_tratado`, `perdas_%`, `volume_produzido/faturado`.
- **Acesso:**
 - Portal SNIS (diagnósticos, “Baixar Tabelas”):
<https://www.gov.br/cidades/pt-br/acesso-a-informacao/acoes-e-programas/saneamento/snis>
 - Diagnósticos anteriores (exemplo 2020):
<https://www.gov.br/cidades/pt-br/acesso-a-informacao/acoes-e-programas/saneamento/snis/diagnosticos-anteriores-do-snis/agua-e-esgotos-1/2020>
 - SINISA (transição/novo portal):
<https://www.gov.br/cidades/pt-br/acesso-a-informacao/acoes-e-programas/saneamento/sinisa/sinisa-1/sinisa>

8.2 Qualidade da água (vigilância e controle)

SISAGUA / VIGIÁGUA – OpenDataSUS (CKAN)

- **Finalidade:** controle e vigilância da **qualidade da água**; população abastecida; captação; tratamento.

- **Cobertura:** Brasil; granularidade municipal/sistemas; **mensal/semanal/anual.**
- **Campos-chave:** cloro_residual, turbidez, cor, pH, E_coli, %_amostras_fora_do_padrao, pontos_captacao, tipo_tratamento, populacao_abastecida.
- **Acesso (principais conjuntos):**
 - Controle Mensal – Parâmetros básicos:
<https://opendatasus.saude.gov.br/dataset/sisagua-controle-mensal-parmetros-basicos>
 - Controle Mensal – Demais parâmetros:
<https://opendatasus.saude.gov.br/dataset/sisagua-controle-mensal-demais-parametros>
 - Vigilância – Parâmetros básicos:
<https://opendatasus.saude.gov.br/dataset/sisagua-vigilancia-parametros-basicos>
 - Controle Semestral:
<https://opendatasus.saude.gov.br/dataset/sisagua-controle-semestral>
 - Pontos de captação:
<https://opendatasus.saude.gov.br/dataset/sisagua-pontos-de-captacao>
 - Tratamento de água (cadastro):
<https://opendatasus.saude.gov.br/dataset/sisagua-tratamento-de-agua>
 - População abastecida (cadastro):
<https://opendatasus.saude.gov.br/dataset/sisagua-populacao-abastecida>

8.3 Saúde — desfechos e capacidade

DATASUS – TABNET (SIH/SUS – Morbidade Hospitalar)

- **Finalidade:** contagens mensais de internações por **CID-10** (usar **A00–A09**).
- **Acesso:**

- Portal SIH/SUS:
<https://datasus.saude.gov.br/acesso-a-informacao/morbidade-hospitalar-do-sus-sih-sus/>
 - Tabnet (entrada):
<https://datasus.saude.gov.br/informacoes-de-saude-tabnet/>
- **Observação:** para **automação** preferir microdados DBC via scripts (PySUS/microdatasus ou CLI datasus-fetcher).

OpenDataSUS – CNES / UBS

- **Finalidade:** cadastro e leitos por município (covariável).
- **Acesso:** perfil “Dados Abertos” (lista de conjuntos):
<https://opendatasus.saude.gov.br/user/dadosabertos>

8.4 Demografia e condições domiciliares

IBGE – Censo 2022 (SIDRA)

- **Acesso:**
 - Tabela 6803 (abastecimento de água):
<https://sidra.ibge.gov.br/tabela/6803>
 - Tabela 6805 (tipo de esgotamento):
<https://sidra.ibge.gov.br/tabela/6805>
 - Tabela 6892 (destino do lixo): <https://sidra.ibge.gov.br/tabela/6892>

IBGE – Estimativas Populacionais

- **Acesso:**
<https://www.ibge.gov.br/estatisticas/sociais/populacao/9103%20-estimativas-de-populacao.html>

8.5 Clima

INMET – BDMEP

- **Portal:** <https://bdmep.inmet.gov.br/>
Base dos Dados (espelho público)
- **Catálogo:** <https://basedosdados.org/dataset/br-inmet-bdmep>

8.6 Orçamento em saúde

SIOPS – Finanças públicas em saúde

- **Portal:** <https://www.gov.br/saude/pt-br/acesso-a-informacao/siops>
- **OpenDataSUS (datasets):** <https://opendatasus.saude.gov.br/dataset/siops>

8.7 ODS — referências institucionais

- **ANA – ODS 6:**
<https://www.gov.br/ana/pt-br/centrais-de-conteudos/publicacoes/ods6>
 - **Plataforma ODS – IBGE:**
<https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/20945-nova-plataforma-digital-permite-acompanhar-indicadores-da-agenda-2030>
-

9) Suficiência de Dados (para ML/IA)

- **SISAGUA (mensal):** alto volume de registros e **muitas variáveis** (físico-químicas/microbiológicas) → bom para **classificação** (conformidade) e **séries temporais**.
- **SNIS/SINISA (anual):** ampla cobertura e múltiplas métricas operacionais → robusto para **efeitos estruturais e comparações**.
- **SIH (mensal):** microdados ricos (AIH) → ideal para **taxas municipais** e análises de **defasagem climática**.
- **INMET (diário):** longa série (agregação mensal) → **features climáticas** com **lags**.
- **IBGE (Censo/Estimativas):** denominadores e covariáveis domiciliares.
Conclusão: há **volume e variáveis** suficientes para treinar modelos de

regressão/classificação, com avaliação temporal e explicabilidade.

10) Arquitetura de Dados (ETL) e Governança

Repositório e camadas

- **GitHub** (código, issues, wiki) + **DVC** (versionamento de dados) com **Google Drive** como *remote*.
- Estrutura:

```
/data/bronze/ # dumps crus (CSV, ZIP, DBC)
/data/silver/ # padronizados (Parquet/CSV)
/data/gold/   # métricas finais e tabelas analíticas
/src/etl/     # scripts: sisagua.py, sih.py, snis.py, sidra.py, inmet.py, siops.py
/notebooks/   # EDA e validações
/dash/        # app do dashboard (streamlit/react)
/docs/        # dicionário de dados + ODS tracker
/tests/       # testes (pandera/great_expectations)
```

Pipelines

- **Makefile/Poetry** para tarefas (`make pull_sisagua`, `make pull_sih`, `make build_gold`).
 - **CI (GitHub Actions)** com **cron semanal** e *dispatch* manual → publica artefatos `/data/gold` para o dashboard.
 - **Qualidade**: testes de esquema e *row counts* mínimos por período; *flags* de imputação; **glossário** (mudanças SNIS→SINISA).
-

11) Modelagem (aderente ao curso)

Alvo principal (contínuo): A00–A09/100 mil (mensal).

- **Baselines:** Regressão Linear Múltipla, Ridge/Lasso; Random Forest Regressor (não-linearidades).
- **Validação:** *TimeSeriesSplit* / backtesting, MAE/RMSE.

Alvo auxiliar (binário): não conformidade (SISAGUA) ou “município acima de limiar de risco”.

- **Baselines:** Regressão Logística, Árvore/Random Forest, SVM; métricas: F1, ROC-AUC, matriz de confusão.

Extensões opcionais (não obrigatórias):

- **GLM Poisson/NegBin** para contagens;
 - **ITS (Interrupted Time Series)** para marcos (canais/ETE).
 - **SHAP** para interpretabilidade.
-

12) Simulador de Cenários e Priorização

- **Entradas (sliders):** $+\Delta$ coleta, $+\Delta$ % tratado, $-\Delta$ perdas, $+\Delta$ conformidade (cloro/turbidez).
 - **Saídas:** Δ previsto em A00–A09/100 mil e ranking por município/sistema.
 - **Racional:** aplicar variações controladas nos preditores e estimar impacto marginal com base no modelo treinado (baseline linear e/ou RF).
-

13) Dashboard (Wireframe)

Páginas:

1. **Visão Executiva**: KPIs (A00–A09; % coleta; % tratado; perdas; % fora do padrão; chuva), alertas e **termômetro ODS**.
2. **Saúde & Clima**: séries mensais A00–A09 × chuva (com *lags*) e previsão 3–6m.
3. **Qualidade da Água**: cloro, turbidez, E. coli; % não conformidades; tendência e sazonalidade.
4. **Operação & Cobertura**: perdas, coleta e % tratado; evolução por município.
5. **Priorizar & Simular**: metas (sliders) → **impacto previsto + ranking**.
6. **ODS Tracker**: matriz KPI→Meta ODS→Fonte (métodos e fórmulas).

O **wireframe funcional** já está disponível (preview) e será conectado aos dados /data/gold após a primeira execução do ETL.

14) Critérios de Avaliação e Métricas de Sucesso

- **Impacto previsto**: Δ A00–A09/100 mil (12 meses) e economia proxy (custo por internação evitada).
 - **Originalidade aplicada**: **previsão + simulação + priorização** (não apenas painel descritivo).
 - **Viabilidade**: dados reais carregados; demo navegável; scripts e dicionário públicos.
 - **ODS**: rastreabilidade clara e metas.
 - **Piloto**: plano de teste com 2 municípios (Belém e Ananindeua).
-

15) Cronograma

1. **S1–S2 – Coleta & Staging:** SISAGUA, SIH, SNIS/SINISA, INMET, IBGE, SIOPS; padronização IBGE.
 2. **S3–S4 – Limpeza & Features:** dicionário, *lags* climáticos, indicadores de qualidade/saneamento.
 3. **S5–S6 – Modelagem & Simulador:** baseline (Linear/Ridge/Lasso/RF); ITS opcional; ranking e cenários.
 4. **S7 – Dashboard & Relatório:** publicação do demo; sumário executivo e ODS.
-

16) Piloto Proposto

- **Municípios foco: Belém e Ananindeua.**
 - **Roteiro:** rodar baseline, calibrar pesos do ranking com especialistas, testar 2–3 cenários e recolher feedback operacional.
-

17) Riscos e Mitigações

- **Defasagem entre fontes (mensal × anual)** → harmonização com *flags* e interpolação conservadora.
 - **Cobertura heterogênea no SISAGUA** → médias móveis e checagem de completude.
 - **Mudanças SNIS→SINISA** → glossário e documentação de indicadores.
 - **Eventos de intervenção (canais/ETE)** → **ITS** com controle climático.
-

18) Aderência ao Curso (Encontros 16–21)

- **Processo:** CRISP-DM e pipeline.
 - **Pré-processamento:** escala/codificação e cuidados com desbalanceamento.
 - **Algoritmos: Regressão** (Linear/Ridge/Lasso/RF) e **Classificação** (Logística/Árvore/RF/SVM).
 - **Avaliação:** hold-out + validação cruzada; métricas apropriadas.
 - **Automação (opcional):** agentes/orquestração de tarefas para ETL e testes.
-

19) Papéis e Responsabilidades (proposta)

- **Engenharia de Dados (líder):** setup GitHub/DVC/Drive; ETL de SISAGUA/SIH/SNIS/SIDRA/INMET/SIOPS; qualidade e versionamento.
 - **Modelagem:** definição de features; baseline; ITS opcional; interpretabilidade (SHAP).
 - **Visualização/Produto:** dashboard; UX; narrativa executiva; ODS Tracker.
 - **Relações Institucionais:** contatos com órgãos locais e validação do piloto.
-

20) Plano de Reunião (Amanhã à noite)

1. **Alinhamento rápido** (10 min): objetivo, escopo e ODS.
2. **Demonstração do wireframe** (10 min).
3. **Arquitetura e ETL** (15 min): aprovação da estrutura e tarefas.
4. **Divisão de papéis e cronograma** (15 min).
5. **Próximos passos** (10 min): catálogo de dados, scripts `sisagua.py` e `sih.py`, primeira carga `/gold`.

21) Próximos Passos Imediatos

- Criar repositório GitHub + DVC/Drive com a estrutura proposta.
 - Implementar (CKAN) e (microdados DBC).
 - Publicar **gold inicial** (RMB 2019–2025) com KPIs: A00–A09/100 mil; % coleta; % tratado; perdas; % fora do padrão; chuva.
 - Conectar o dashboard ao /data/gold e liberar o demo para o grupo.
-

22) Anexos — Matriz ODS + Glossário + Guia de Execução (ETL)

1) Matriz ODS (detalhada) — Fórmulas, Fontes e Observações

1.1 ODS 6 — Água Limpa e Saneamento

KPI 6.1 — Acesso à água

- **Indicador:** % de domicílios com **abastecimento de água por rede geral**.
- **Fórmula:** $\text{cobertura_agua\%} = (\text{domicílios_com_rede_geral_agua} / \text{domicílios_totais}) * 100$.
- **Fonte (pública): IBGE/SIDRA — Tabela 6803 —**
<https://sidra.ibge.gov.br/tabela/6803>
- **Periodicidade/Granularidade:** Censo 2022 (municipal).
- **Observações:** Para o recorte municipal, usar o código IBGE (n6). Para automação, utilizar a **API do SIDRA** (/values/t/6803/...).

KPI 6.2 — Saneamento (coleta de esgoto)

- **Indicador:** % de domicílios com **coleta de esgoto por rede geral ou pluvial**.
- **Fórmula:** $\text{coleta_esgoto\%} = (\text{domicílios_com_rede_geral_ou_pluvial} /$

`domicílios_totais) * 100.`

- **Fonte (pública): IBGE/SIDRA — Tabela 6805 —**
<https://sidra.ibge.gov.br/tabela/6805>
- **Periodicidade/Granularidade:** Censo 2022 (municipal).
- **Observações:** Documentar o critério de “coleta adequada”. Em painéis anuais, complementar com SNIS/SINISA.

KPI 6.3a — Qualidade da água: % de amostras fora do padrão

- **Indicador:** proporção de amostras **não conformes** (cloro residual livre, turbidez, microbiologia etc.) no período.
- **Fórmula:** `%_fora_do_padrão = (n amostras_não_conformes / n amostras_totais) * 100.`
- **Fonte (pública): SISAGUA/Op e nDataSUS (CKAN)** — Conjuntos de “Controle Mensal” e “Vigilância”
 - Controle Mensal — Parâmetros básicos:
<https://opendatasus.saude.gov.br/dataset/sisagua-controle-mensal-parametros-basicos>
 - Controle Mensal — Demais parâmetros:
<https://opendatasus.saude.gov.br/dataset/sisagua-controle-mensal-demais-parametros>
 - Vigilância — Parâmetros básicos:
<https://opendatasus.saude.gov.br/dataset/sisagua-vigilancia-parametros-basicos>
- **Periodicidade/Granularidade:** Mensal (por sistema/municipal).
- **Observações:** As regras de conformidade seguem a regulação vigente (SISAGUA já agrega por parâmetro).

KPI 6.3b — % do esgoto coletado que é tratado

- **Indicador:** proporção do volume **tratado** sobre o **coletado**.

- **Fórmula:** `tratado_% = (volume_esgoto_tratado / volume_esgoto_coletado) * 100.`
- **Fonte (pública): SNIS/SINISA — Água e Esgotos** (Diagnósticos, “Baixar Tabelas”)
 - Portal SNIS:
<https://www.gov.br/cidades/pt-br/acesso-a-informacao/acoes-e-programas/saneamento/snis>
 - SINISA:
<https://www.gov.br/cidades/pt-br/acesso-a-informacao/acoes-e-programas/saneamento/sinisa/sinisa-1/sinisa>
- **Periodicidade/Granularidade:** Anual (municipal/prestador).
- **Observações:** Harmonizar códigos de município/prestador e anos-base.

KPI 6.4 — Eficiência: perdas na distribuição

- **Indicador:** Índice de perdas na distribuição (oficial SNIS/SINISA).
 - **Fórmula (referência):** utilizar o indicador disponibilizado pelo SNIS/SINISA (quando necessário: aproximar por volumes produzidos × faturados).
 - **Fonte (pública): SNIS/SINISA — Água e Esgotos** (Diagnósticos, “Baixar Tabelas”).
 - **Periodicidade/Granularidade:** Anual (municipal/prestador).
-

1.2 ODS 3 — Saúde e Bem-Estar

KPI 3.3/3.9 — Taxa de internações A00–A09 por 100 mil hab.

- **Indicador:** taxa mensal de internações por **CID-10 A00–A09** (doenças infecciosas intestinais).
- **Fórmula:** `taxa_A00A09 = (internações_A00A09 / população) * 100000.`

- Numerador: internações mensais por município (SIH/SUS).
 - Denominador: **IBGE — Estimativas Populacionais** (anual), distribuídas por mês (critério proporcional simples).
- **Fonte (pública):**
 - **SIH/TABNET:**
<https://datasus.saude.gov.br/informacoes-de-saude-tabnet/>
 - **IBGE — Estimativas:**
<https://www.ibge.gov.br/estatisticas/sociais/populacao/9103%20-estimativas-de-populacao.html>
 - **Periodicidade/Granularidade:** Mensal (internações); Municipal.
 - **Observações:** Aplicar janelas de média móvel e/ou *backtesting* para estabilidade.
-

1.3 ODS 11 — Cidades e Comunidades Sustentáveis (complementar)

KPI 11.6 — Coleta e destino do lixo domiciliar

- **Indicador:** % de domicílios com coleta/destino adequado.
 - **Fórmula:** `coleta_lixo_% = (domicílios_com_coleta/destino_adequado / domicílios_totais) * 100.`
 - **Fonte (pública): IBGE/SIDRA — Tabela 6892 —**
<https://sidra.ibge.gov.br/tabela/6892>
 - **Periodicidade/Granularidade:** Censo 2022 (municipal).
 - **Observações:** Indicador de contexto urbano; manter como painel complementar.
-

2) Glossário de Variáveis — por Fonte de Dados

Nomes são indicativos. Ajustar aos rótulos reais após a primeira coleta/inspeção.

2.1 SISAGUA (Controle/Vigilância)

- `data_coleta` — data de coleta da amostra.
- `municipio_ibge, municipio_nome` — identificadores municipais.
- `sistema_abastecimento` — código/nome do sistema (quando disponível).
- `cloro_residual_livre_mgL` — mg/L.
- `turbidez_NTU` — NTU.
- `cor_uH, pH` — unidades do padrão.
- `e_coli` — presença/ausência ou NMP/100 mL.
- `resultado_conformidade` — conforme / não conforme.
- `amostras_totais, amostras_nao_conformes` — contagens agregadas (se fornecidas).
- **Cadastros associados:** `pontos_captacao, tipo_tratamento, populacao_abastecida`.

2.2 SNIS/SINISA (Água e Esgoto)

- `ano, municipio_ibge, prestador` — chaves de série.
- `atendimento_agua_%` — cobertura de água.
- `coleta_esgoto_%` — cobertura de coleta.
- `esgoto_tratado_%` — proporção tratada/volume coletado.
- `perdas_%` — índice de perdas na distribuição.

- `volume_produzido_m3, volume_faturado_m3` — volumetrias.
- `receita_operacional, despesa_operacional` — dimensões econômicas (quando disponíveis).

2.3 SIH/SUS (Morbidade Hospitalar) — microdados/tabulações

- `competencia` — ano/mês da internação.
- `municipio_residencia_ibge, municipio_internacao_ibge`.
- `cid10_principal` — diagnóstico principal (filtrar A00–A09).
- `sexo, idade, faixa_etaria`.
- `dias_permanencia, valor_total` — severidade/custo (opcional).
- **Agregados usados:** `internacoes_A00A09_mensal` por município.

2.4 IBGE/SIDRA — Censo 2022

- `domicilios_total`.
- `domicilios_rede_agua` (t/6803).
- `domicilios_rede_esgoto_pluvial` (t/6805).
- `domicilios_coleta_lixo` (t/6892).
- **Derivados:** `cobertura_agua_%`, `coleta_esgoto_%`, `coleta_lixo_%`.

2.5 IBGE — Estimativas Populacionais

- `populacao_municipal_anual` — denominador para taxas (distribuição mensal proporcional).
- **Derivado:** `taxa_A00A09_100k`.

2.6 INMET/BDMEP

- `estacao_codigo, lat, lon.`
- `data, precipitacao_total_mm, temperatura_media_c,`
`umidade_relativa_%.`
- **Agregados:** `chuva_mensal_mm` e *lags* `chuva_lag1, chuva_lag2.`

2.7 SIOPS (Finanças em Saúde)

- `despesa_saude_total, despesa_saude_pc` (per capita).
 - `receita_corrente_liquida, %gasto_saude_sobre_RCL.`
 - **Uso:** covariável de capacidade fiscal/assistencial.
-

3) Guia de Execução — ETL

3.1 Estrutura de Pastas

/data/bronze/ # dumps crus (CSV, ZIP, DBC)
/data/silver/ # padronizados (Parquet/CSV)
/data/gold/ # tabelas analíticas e KPIs
/src/etl/ # sisagua.py, sih.py, snis.py, sidra.py, inmet.py, siops.py
/notebooks/ # EDA e validações
/docs/ # dicionário de dados + ODS tracker
/tests/ # validações (pandera/great_expectations)
