

# Simulação de Dados de Hidrômetros Inteligentes

Este projeto tem como objetivo simular dados de hidrômetros inteligentes para fins de estudo e análise, com foco em IoT e Cidades Inteligentes. Os dados gerados buscam replicar cenários realistas, incluindo padrões de consumo variados, falhas e anomalias.

## Conteúdo do Dataset Simulado

O dataset simulado contém leituras horárias de 500 hidrômetros inteligentes, abrangendo um período de 12 meses (Janeiro a Dezembro de 2024). Os dados são separados por mês e por cidade, resultando em múltiplos arquivos para facilitar o gerenciamento e a análise.

### Variáveis Incluídas:

- `id_hidrometro` : Identificador único para cada hidrômetro.
- `timestamp` : Data e hora da leitura (formato horário).
- `consumo_m3` : Volume de água consumido em metros cúbicos desde a última leitura.
- `vazao_instantanea_m3h` : Vazão instantânea medida no momento da leitura.
- `status_hidrometro` : Estado operacional do medidor (ex: OK, BATERIA\_FRACA, FALHA\_COMUNICACAO, LEITURA\_INCORRETA, SUSPEITA\_FRAUDE).
- `alerta_vazamento` : Booleano indicando detecção de vazamento (True/False).
- `alerta_fraude` : Booleano indicando detecção de fraude (True/False).
- `temperatura_local_c` : Temperatura no local da instalação em graus Celsius.
- `pressao_rede_bar` : Pressão na rede de água em bares.
- `nome_cidade` : Nome da cidade onde o hidrômetro está localizado.
- `regiao_brasil` : Região geográfica do Brasil (Norte, Nordeste, Sudeste, Sul, Centro-Oeste).

- **bairro** : Nome do bairro real onde o hidrômetro está localizado, categorizado por perfil socioeconômico.
- **perfil\_socioeconomico** : Classificação socioeconômica do bairro (Alta, Média, Baixa).

## Cidades e Bairros Reais Incluídos:

Para cada cidade, foram selecionados bairros reais e categorizados por perfil socioeconômico (Alta, Média, Baixa) para tornar a simulação mais precisa:

- **Belém (Norte):**
  - **Alta:** Umarizal, Nazaré, Batista Campos, Reduto
  - **Média:** Marco, São Brás, Pedreira, Telégrafo
  - **Baixa:** Jurunas, Cremação, Guamá, Terra Firme, Cabanagem
- **Manaus (Norte):**
  - **Alta:** Adrianópolis, Ponta Negra, Nossa Senhora das Graças, Vieiralves, Parque Dez de Novembro
  - **Média:** Flores, Dom Pedro, Aleixo, São Geraldo
  - **Baixa:** Jorge Teixeira, Cidade Nova, Novo Aleixo, Japiim, Praça 14 de Janeiro, Alvorada, Coroadó
- **Recife (Nordeste):**
  - **Alta:** Boa Viagem, Jaqueira, Casa Forte, Pina
  - **Média:** Madalena, Boa Vista, Torre, Casa Amarela
  - **Baixa:** Várzea, Ibura, Cohab, Arruda
- **Salvador (Nordeste):**
  - **Alta:** Vitória, Alphaville, Horto Florestal, Graça, Barra, Ondina
  - **Média:** Pituba, Caminho das Árvores, Rio Vermelho, Imbuí, Itaigara, Brotas

- **Baixa:** São Caetano, Pirajá, Liberdade, Pernambués, Cabula, Valéria
- **São Paulo (Sudeste):**
  - **Alta:** Jardim Paulista, Moema, Itaim Bibi, Pinheiros, Vila Nova Conceição, Morumbi
  - **Média:** Vila Mariana, Perdizes, Tatuapé, Mooca, Santana, Ipiranga, Consolação, Bela Vista
  - **Baixa:** Grajaú, Jardim Ângela, Paraisópolis, Heliópolis, Cidade Tiradentes, Brasilândia
- **Rio de Janeiro (Sudeste):**
  - **Alta:** Leblon, Ipanema, Lagoa, Gávea, Jardim Botânico, Barra da Tijuca, São Conrado
  - **Média:** Copacabana, Botafogo, Flamengo, Laranjeiras, Tijuca, Vila Isabel, Méier
  - **Baixa:** Rocinha, Complexo do Alemão, Maré, Cidade de Deus, Campo Grande, Bangu, Santa Cruz
- **Porto Alegre (Sul):**
  - **Alta:** Moinhos de Vento, Bela Vista, Petrópolis, Três Figueiras
  - **Média:** Menino Deus, Cidade Baixa, Bom Fim, Rio Branco
  - **Baixa:** Restinga, Lomba do Pinheiro, Rubem Berta, Sarandi
- **Curitiba (Sul):**
  - **Alta:** Batel, Bigorrilho, Ecoville, Juvevê, Água Verde
  - **Média:** Portão, Santa Felicidade, Centro Cívico, Cristo Rei, Cabral
  - **Baixa:** Cidade Industrial de Curitiba (CIC), Sítio Cercado, Cajuru, Boqueirão, Tatuquara

## Falhas e Anomalias Simuladas:

O dataset inclui a simulação de diversas falhas e anomalias para tornar os dados mais realistas para análise de detecção de anomalias:

- **Falhas de Comunicação:** Períodos com dados nulos para consumo, vazão, temperatura e pressão.
- **Leituras Incorretas:** Valores absurdamente altos para consumo e vazão.
- **Bateria Fraca:** Períodos de consumo zero ou inconsistente, simulando falha do hidrômetro.
- **Vazamentos Lentos:** Pequeno consumo contínuo em horas de baixo uso (madrugada).
- **Fraudes:** Consumo zerado por longos períodos em hidrômetros que deveriam estar ativos.

## Estrutura dos Arquivos

Os arquivos são gerados na pasta `simulacao_hidrometros_por_mes_cidade/` e seguem o padrão de nomenclatura:

```
hidrometros_<nome_cidade_minusculo_sem_espacos>_<mes_minusculo>.csv
```

```
hidrometros_<nome_cidade_minusculo_sem_espacos>_<mes_minusculo>.parquet
```

Exemplo:

- `simulacao_hidrometros_por_mes_cidade/hidrometros_belem_january.csv`
- `simulacao_hidrometros_por_mes_cidade/hidrometros_belem_january.parquet`

## Como Utilizar os Dados

Você pode carregar esses arquivos em ferramentas de análise de dados como Pandas (Python) ou PySpark para realizar suas análises, visualizações e desenvolver modelos de Machine Learning para detecção de padrões, falhas e anomalias.

Python

```
import pandas as pd

# Exemplo de carregamento de um arquivo CSV
df_exemplo_csv =
```

```
pd.read_csv("simulacao_hidrometros_por_mes_cidade/hidrometros_belem_january.csv")
print(df_exemplo_csv.head())

# Exemplo de carregamento de um arquivo Parquet
df_exemplo_parquet =
pd.read_parquet("simulacao_hidrometros_por_mes_cidade/hidrometros_belem_january.parquet")
print(df_exemplo_parquet.head())
```

Esperamos que este dataset simulado seja uma ferramenta valiosa para seus estudos e projetos!