

---

# Qualidade da água potável no Rio de Janeiro

Projeto de Introdução a Ciência de Dados

Caique Molina  
Felipe Metson  
Guilherme Dantas  
Thiago Nicácio



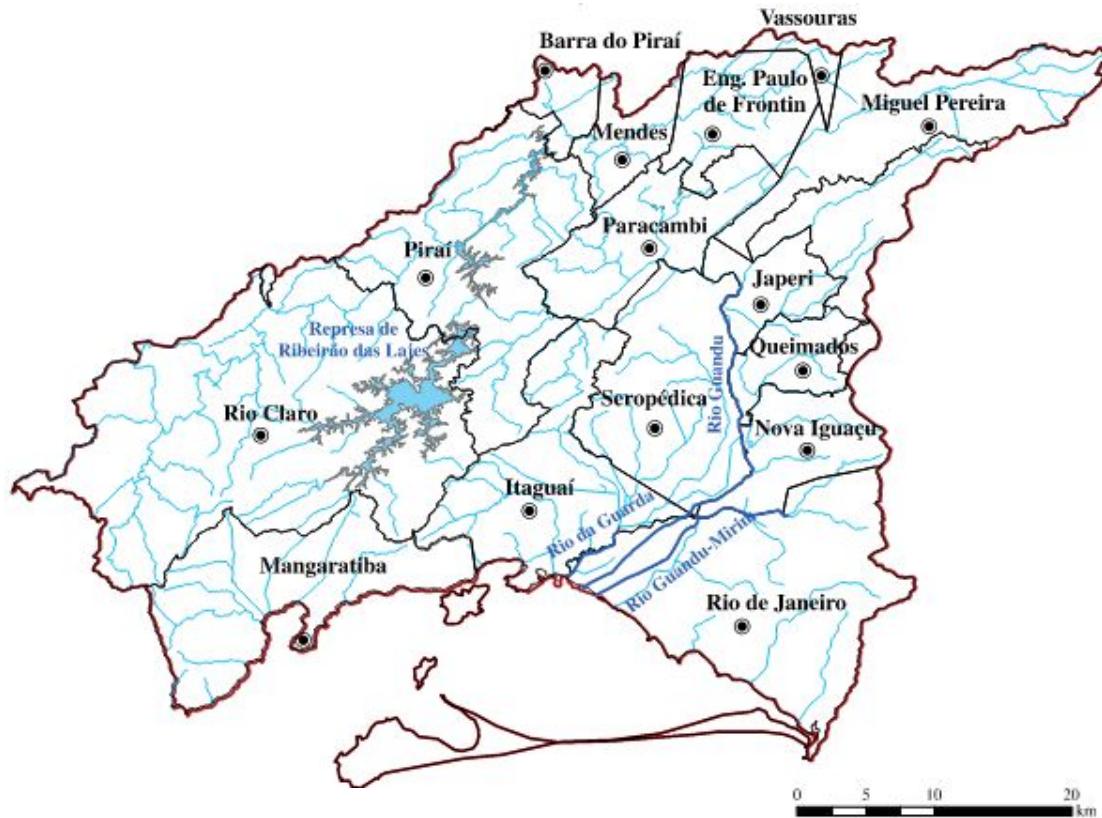
# O Sistema Guandu



## Extensão da Bacia do Rio Guandu pelos municípios do Rio de Janeiro



- Limite Municipal
- Hidrografia
- Represa
- RH II - Guandu
- Rios Principais
- Sedes Municipais



Fonte: CEDAE

## Estação de Tratamento de Água (ETA) de Guandu

- Maior ETA do Mundo, pelo Guinness
- Abastece 80% dos cariocas
- Lida com altos índices de poluição



Fonte: ZM Notícias



Fonte: Jovem Pan

## Questões de qualidade da água do Guandu na mídia

### Cedae vai aplicar argila perto do Guandu para evitar proliferação de algas

A aplicação será feita por meio de uma embarcação numa lagoa próxima à captação da estação de tratamento.

Fonte: G1 (29/01/2020)

### Acordo com Cedae para ressarcir consumidores prejudicados deve ficar pronto em dez dias

Fonte: O Globo (31/01/2020)

RIO • RADAR

### Estação do Guandu será desligada para colocação de bomba para tentar reduzir geosmina na água

Por causa da manobra, Rio e municípios da Baixada Fluminense podem ter problemas de abastecimento

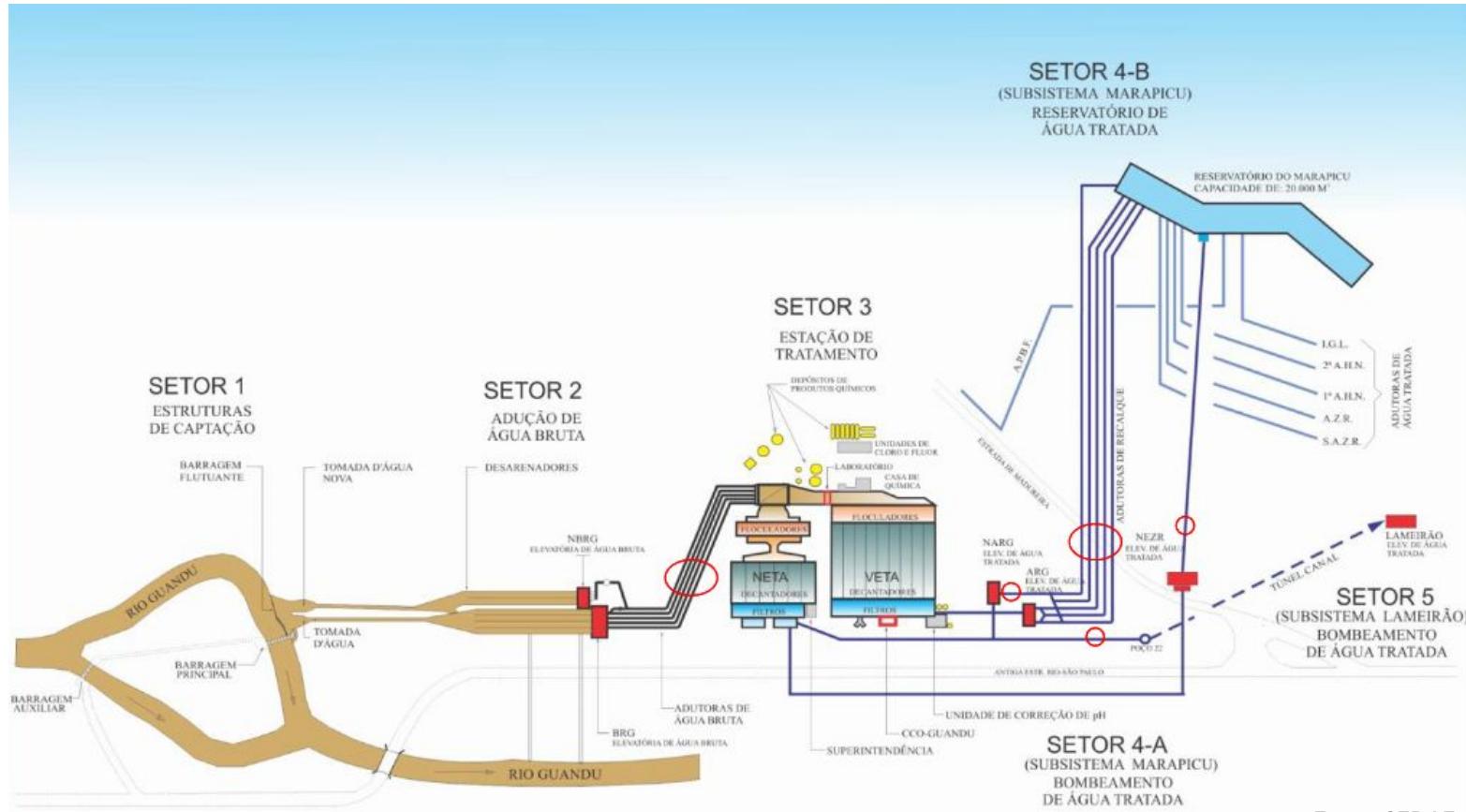
Fonte: O Globo (06/04/2021)

### Presidente da Cedae é demitido em meio à crise da água no Rio

*Decisão foi anunciada pelo governador Wilson Witzel*

Fonte: Agência Brasil (10/02/2020)

## Desenho esquemático do Sistema Guandu



# Extração e Limpeza de Dados do Sistema Guandu

---

# Relatórios de Qualidade da Água do Sistema Guandu

A CEDAE realiza rotineiramente mais de 100 coletas diárias de amostras para análise da qualidade da água ao longo de sua rede de distribuição, além de coletas nas Saídas de Tratamento das Estações de tratamento e mananciais. Além de dar publicidade aos laudos das análises laboratoriais segundo as determinações do Ministério da Saúde, a Companhia - visando levar mais informação ao público e buscando ainda mais transparência na sua relação com o consumidor - divulga os resultados de testes realizados ao longo de sua rede de distribuição e da Estação de Tratamento do Guandu.

**Qualidade da Água**

**Relatórios Guandu**

**2021**

**MONITORAMENTO - GOSTO E ODOR - REDE DE DISTRIBUIÇÃO**  
PDF | 513 KB

**MONITORAMENTO - GOSTO E ODOR - SAÍDA DE TRATAMENTO**  
PDF | 134 KB

**MONITORAMENTO - CIANOTOXINAS**  
PDF | 403 KB

**MONITORAMENTO - GEOSMINA / MIB**  
PDF | 157,0 KB

DATA DE COLETA	Análises Terceirizadas - Laboratório Oceanus - Contrato 138/2020 DSG			
	SAÍDA TRATAMENTO - NETA		SAÍDA TRATAMENTO - VETA	
	GOSTO	ODOR	GOSTO	ODOR
01/01/2021	2	1	2	1
02/01/2021	2	1	2	1
03/01/2021	2	1	2	2
04/01/2021	2	1	2	1
05/01/2021	2	1	1	1
06/01/2021	2	1	2	1
07/01/2021	2	1	2	1
08/01/2021	1	1	1	1
09/01/2021	1	1	1	1
10/01/2021	1	1	1	1
11/01/2021	1	1	1	1
12/01/2021	1	1	1	1
13/01/2021	1	1	1	1
14/01/2021	1	1	1	1
15/01/2021	1	1	1	1
16/01/2021	2	1	1	1
17/01/2021	2	1	2	1
18/01/2021	1	1	1	1
19/01/2021	2	1	3	1
20/01/2021	1	1	2	1
21/01/2021	1	1	1	1

Fonte: CEDAE

---

## Dados extraídos e tratados

- Concentração média de geosmina nas ETAs (26/01/2020-18/03/2020)
- Concentração de geosmina em cada ETA (23/10/2020-)
- Gosto e Odor em cada ETA (29/01/2020-)
- Gosto e Odor na rede de distribuição (13/02/2020-)
- Cor aparente, Turbidez, Cloro Residual Livre, Coliformes Totais, Coliformes Termotolerantes, E. Coli (05/2005-12/2020)

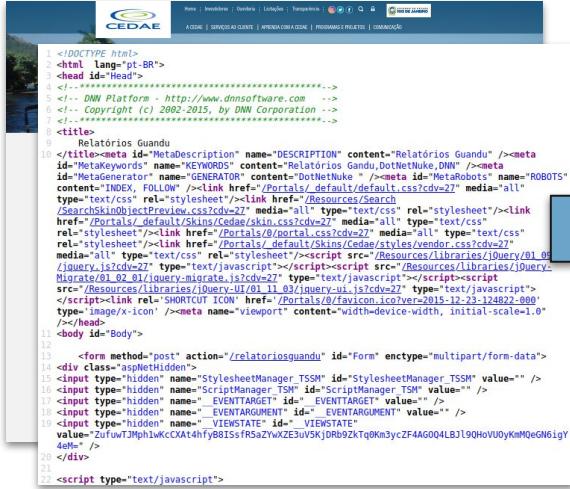
---

# Extração e tratamento dos dados do Guandu

1. Extraímos da página da CEDAE os relatórios em formato PDF
2. Baixamos os relatórios e os convertemos para tabelas pandas
3. Limpamos as tabelas pandas e as exportamos para arquivos CSV

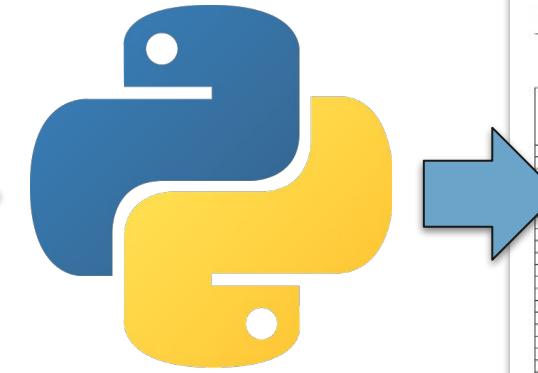


# Etapa 1a - Extração dos relatórios (automática)

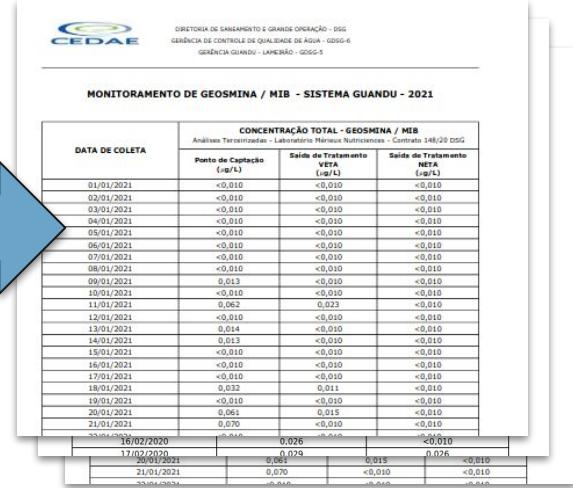


```
<!DOCTYPE html>
<html lang="pt-BR">
<head id="Head">
<!-- ***** HEAD INFORMATION ***** -->
<!-- DNN Platform --> http://www.dnnsoftware.com
<!-- Copyright (c) 2002-2015, by DNN Corporation -->
<!-- ***** HEAD INFORMATION ***** -->
<title>Relatório Guandu</title>
<meta id="MetaDescription" name="DESCRIPTION" content="Relatórios Guandu" /><meta id="MetaKeywords" name="KEYWORD" content="Guandu, CEDAE, Dnncache, DotNetNuke, DNN" /><meta id="MetaRobots" name="ROBOTS" content="INDEX, FOLLOW" /><link href="/Portals/_default/default.css?cv=27" media="all" type="text/css" rel="stylesheet" /><link href="/Portals/_default/Skins/Cedae/skin.css?cv=27" media="all" type="text/css" rel="stylesheet" /><link href="/Portals/_default/Portals/_skins?cv=27" media="all" type="text/css" rel="stylesheet" /><link href="/Portals/_default/Portals/_skins?cv=27" media="all" type="text/css" rel="stylesheet" /><script src="/Resources/libraries/jquery/jquery.js?cv=27" type="text/javascript"></script><script src="/Resources/libraries/jquery-Migrate/01_02_01/jquery-migrate.js?cv=27" type="text/javascript"></script><script src="/Resources/libraries/jquery-ui/01_11_03/jquery-ui.js?cv=27" type="text/javascript"></script><script rel="SHORTCUT ICON" href="/Portals/_default/favicon.ico?ver=2015-12-23-124922-000" type="image/x-icon" /><meta name="viewport" content="width=device-width, initial-scale=1.0" />
</head>
<body id="Body">
<form method="post" action="/relatoriosguandu" id="Form" enctype="multipart/form-data">
<div class="aspNetHidden">
<input type="hidden" name="StyleSheetManager_TSM" id="StyleSheetManager_TSM" value="" />
<input type="hidden" name="ScriptManager_TSM" id="ScriptManager_TSM" value="" />
<input type="hidden" name="EVENTTARGET" id="EVENTTARGET" value="" />
<input type="hidden" name="EVENTARGUMENT" id="EVENTARGUMENT" value="" />
<input type="hidden" name="VIEWSTATE" id="VIEWSTATE" value="ZfuW7Mph1wKcCXA4hf8y8B1SsfR5a2wXZE3uVSKjDRb6ZkTq0Km3ycZF4AG004LB3l90HuU0yKmM0eGn6ig4em=" />
</div>
</form>
<script type="text/javascript">
```

Página da CEDAE



Parser de HTML da  
biblioteca padrão



DATA DE COLETA	CONCENTRAÇÃO TOTAL - GEOSMINA / MIB		
	Ponto de Captura (g/L)	Saída de Tratamento VETA (g/L)	Saída de Tratamento NETO (g/L)
01/01/2021	<0,010	<0,010	<0,010
02/01/2021	<0,010	<0,010	<0,010
03/01/2021	<0,010	<0,010	<0,010
04/01/2021	<0,010	<0,010	<0,010
05/01/2021	<0,010	<0,010	<0,010
06/01/2021	<0,010	<0,010	<0,010
07/01/2021	<0,010	<0,010	<0,010
08/01/2021	<0,010	<0,010	<0,010
09/01/2021	0,013	<0,010	<0,010
10/01/2021	<0,010	<0,010	<0,010
11/01/2021	0,062	0,023	<0,010
12/01/2021	<0,010	<0,010	<0,010
13/01/2021	0,010	<0,010	<0,010
14/01/2021	0,013	<0,010	<0,010
15/01/2021	<0,010	<0,010	<0,010
16/01/2021	<0,010	<0,010	<0,010
17/01/2021	<0,010	<0,010	<0,010
18/01/2021	0,032	0,011	<0,010
19/01/2021	<0,010	<0,010	<0,010
20/01/2021	0,061	0,015	<0,010
21/01/2021	0,070	<0,010	<0,010
22/01/2021	<0,010	<0,010	<0,010
16/02/2020		0,026	
17/02/2020		0,079	0,026
18/02/2020		0,015	<0,010
19/02/2020		0,015	<0,010
20/02/2020		0,070	<0,010
21/02/2020		0,070	<0,010
22/02/2020		0,070	<0,010

Links para relatórios

## **Etapa 1b - Extração dos relatórios (manual)**



MONITORAMENTO DE GEOSMINA / MIB - SISTEMA GUANDU - 2021				
DATA DE COLETA	CONCENTRAÇÃO TOTAL - GEOSMINA / MIB			
	Análises Terceirizadas - Laboratório Márquez Nutriconexos - Contrato 14/20 DSQ	Ponto de Captação (ug/L)	Saída do Tratamento VETA (ug/L)	Saída do Tratamento NETA (ug/L)
01/01/2021		<0,010	<0,010	<0,010
02/01/2021		<0,010	<0,010	<0,010
03/01/2021		<0,010	<0,010	<0,010
04/01/2021		<0,010	<0,010	<0,010
05/01/2021		<0,010	<0,010	<0,010
06/01/2021		<0,010	<0,010	<0,010
07/01/2021		<0,010	<0,010	<0,010
08/01/2021		<0,010	<0,010	<0,010
09/01/2021		0,013	<0,010	<0,010
10/01/2021		<0,010	<0,010	<0,010
11/01/2021		0,062	0,023	<0,010
12/01/2021		<0,010	<0,010	<0,010
13/01/2021		0,014	<0,010	<0,010
14/01/2021		0,013	<0,010	<0,010
15/01/2021		<0,010	<0,010	<0,010
16/01/2021		<0,010	<0,010	<0,010
17/01/2021		<0,010	<0,010	<0,010
18/01/2021		0,032	0,011	<0,010
19/01/2021		<0,010	<0,010	<0,010
20/01/2021		0,061	0,015	<0,010
21/01/2021		0,070	<0,010	<0,010
22/01/2021		0,046	<0,010	<0,010
<b>18/02/2020</b>		<b>0,026</b>	<b>&lt;0,010</b>	<b>&lt;0,010</b>
17/02/2021		0,029	0,015	0,076
18/02/2021		0,026	0,015	<0,010
21/01/2021		0,070	<0,010	<0,010
22/04/2021		0,046	<0,010	<0,010

Página da CEDAE

## Relatórios em ZIP

# Etapa 2 - Extração das tabelas

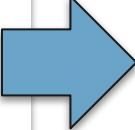
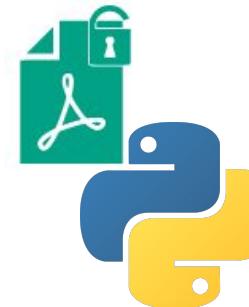


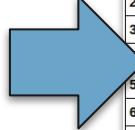
Diagram illustrating the process flow: A PDF document (Relatório em PDF) is processed by a Python parser (Parser de PDF da biblioteca tabula-py) to extract tables into pandas DataFrames (Tabelas pandas).

MONITORAMENTO DE GEOSMINA / MIB - SISTEMA GUANDU - 2021			
DATA DE COLETA	CONCENTRAÇÃO TOTAL - GEOSMINA / MIB		
	Análises Terceirizadas - Laboratório Mérieux Nutriesciences - Contrato 148/20 DSG	Ponto de Captação (µg/L)	Saída de Tratamento (µg/L)
01/01/2021	<0,010	<0,010	<0,010
02/01/2021	<0,010	<0,010	<0,010
03/01/2021	<0,010	<0,010	<0,010
04/01/2021	<0,010	<0,010	<0,010
05/01/2021	0,010	<0,010	<0,010
06/01/2021	<0,010	<0,010	<0,010
07/01/2021	<0,010	<0,010	<0,010
08/01/2021	<0,010	<0,010	<0,010
09/01/2021	0,013	<0,010	<0,010
10/01/2021	<0,010	<0,010	<0,010
11/01/2021	0,062	0,023	<0,010
12/01/2021	<0,010	<0,010	<0,010
13/01/2021	0,014	<0,010	<0,010
14/01/2021	0,013	<0,010	<0,010
15/01/2021	<0,010	<0,010	<0,010
16/01/2021	<0,010	<0,010	<0,010
17/01/2021	<0,010	<0,010	<0,010
18/01/2021	0,032	0,011	<0,010
19/01/2021	<0,010	<0,010	<0,010
20/01/2021	0,001	0,013	<0,010
21/01/2021	0,070	<0,010	<0,010
22/01/2021	<0,010	<0,010	<0,010

Relatório em PDF



Parser de PDF da  
biblioteca tabula-py



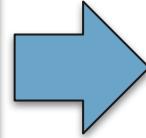
	Unnamed: 0	CONCENTRAÇÃO TOTAL - GEOSMINA / MIB
0	NaN	Análises Terceirizadas - Laboratório Mérieux N...
1	DATA DE COLETA	NaN
2	NaN	Saída de Tratamento Saída de Tratamento
3	NaN	Ponto de Captação
4	NaN	VETA NETA
5	NaN	(mg/L)
6	NaN	(mg/L) (mg/L)
7	01/01/2021	<0,010 <0,010 <0,010
8	02/01/2021	<0,010 <0,010 <0,010
9	03/01/2021	<0,010 <0,010 <0,010

Tabelas pandas

## Etapa 3 - Tratamento das tabelas

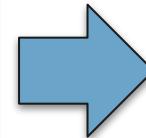
Unnamed: 0 CONCENTRAÇÃO TOTAL - GEOSMINA / MIB				
0	NaN	Análises Terceirizadas - Laboratório Mérieux N...		
1	DATA DE COLETA	NaN		
2	NaN	Saída de Tratamento	Saída de Tratamento	
3	NaN		Ponto de Captação	
4	NaN		VETA NETA	
5	NaN		(mg/L)	
6	NaN		(mg/L) (mg/L)	
7	01/01/2021	<0,010	<0,010 <0,010	
8	02/01/2021	<0,010	<0,010 <0,010	
9	03/01/2021	<0,010	<0,010 <0,010	

Tabela “crua”



	date	params
7	28/03/2021	0,077 0,092 0,020
8	29/03/2021	0,521 0,170 0,045
9	30/03/2021	0,495 0,109 0,018
10	31/03/2021	0,732 0,261 0,049
11	01/04/2021	0,726 0,194 0,052
12	02/04/2021	1,551 0,088 0,066
13	03/04/2021	1,166 0,516 0,080
14	04/04/2021	1,905 0,493 0,088
15	05/04/2021	0,537 0,308 0,075
16	06/04/2021	0,641 0,152 0,033

Removido o  
cabeçalho



	date	capture	veta	neta
7	23/10/2020	<0,010	<0,010	<0,010
8	24/10/2020	<0,010	<0,010	<0,010
9	25/10/2020	<0,010	<0,010	<0,010
10	26/10/2020	<0,010	<0,010	<0,010
11	27/10/2020	<0,010	<0,010	<0,010
12	28/10/2020	<0,010	<0,010	<0,010
13	29/10/2020	<0,010	<0,010	<0,010
14	30/10/2020	<0,010	<0,010	<0,010
15	31/10/2020	<0,010	<0,010	<0,010
16	01/11/2020	<0,010	<0,010	<0,010

Dados separados em  
colunas individuais

---

## Etapa 3 - Tratamento das tabelas

	date	capture	veta	neta
7	23/10/2020	<0,010	<0,010	<0,010
8	24/10/2020	<0,010	<0,010	<0,010
9	25/10/2020	<0,010	<0,010	<0,010
10	26/10/2020	<0,010	<0,010	<0,010
11	27/10/2020	<0,010	<0,010	<0,010
12	28/10/2020	<0,010	<0,010	<0,010
13	29/10/2020	<0,010	<0,010	<0,010
14	30/10/2020	<0,010	<0,010	<0,010
15	31/10/2020	<0,010	<0,010	<0,010
16	01/11/2020	<0,010	<0,010	<0,010



	date	capture	veta	neta	is_capture_ub	is_veta_ub	is_neta_ub
7	2021-01-01	0.010	0.01	0.01	True	True	True
8	2021-01-02	0.010	0.01	0.01	True	True	True
9	2021-01-03	0.010	0.01	0.01	True	True	True
10	2021-01-04	0.010	0.01	0.01	True	True	True
11	2021-01-05	0.010	0.01	0.01	True	True	True
12	2021-01-06	0.010	0.01	0.01	True	True	True
13	2021-01-07	0.010	0.01	0.01	True	True	True
14	2021-01-08	0.010	0.01	0.01	True	True	True
15	2021-01-09	0.013	0.01	0.01	False	True	True
16	2021-01-10	0.010	0.01	0.01	True	True	True

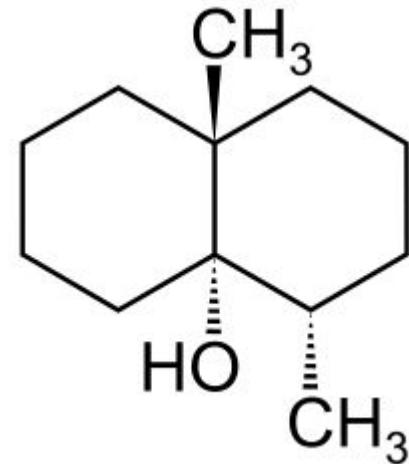
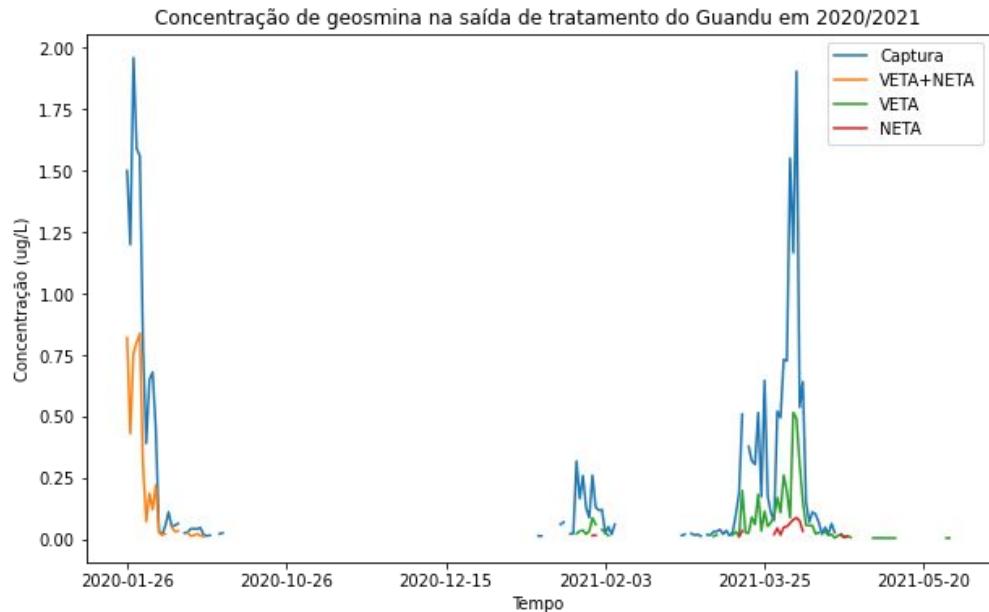
Dados separados em  
colunas individuais

Dados tipados e tratados

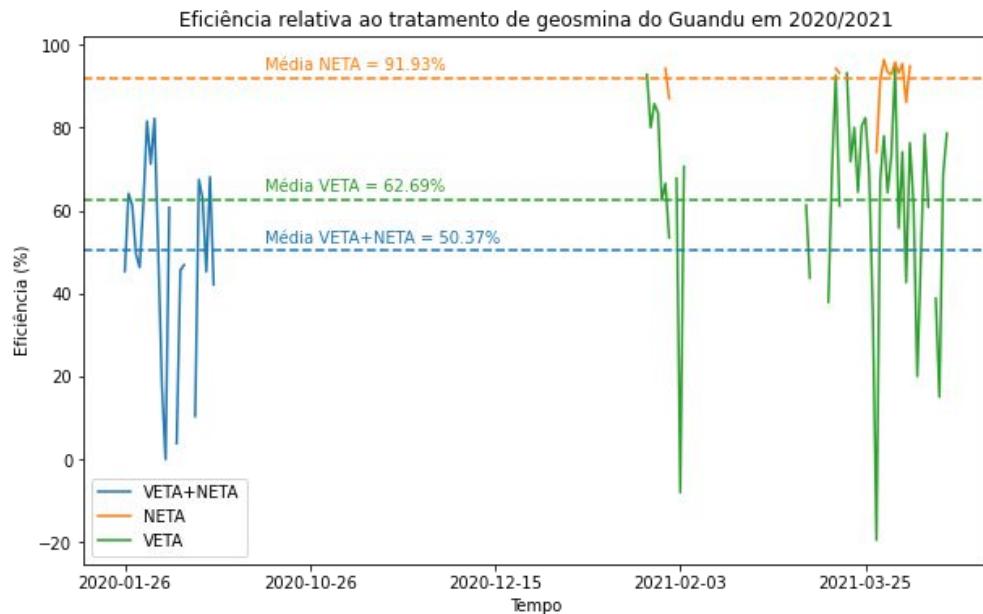
# Como a qualidade da água do Guandu variou com o tempo?

---

# Concentração de geosmina (26/01/2020-)



# Eficiência das ETAs (26/01/2020-)

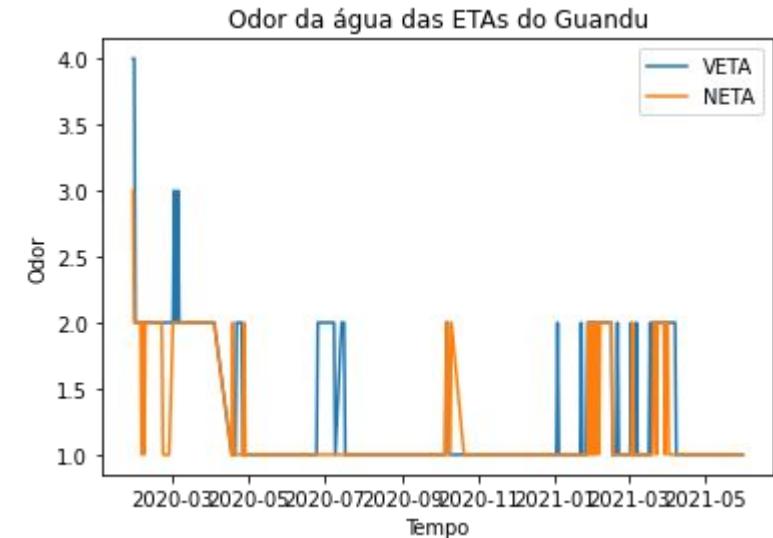
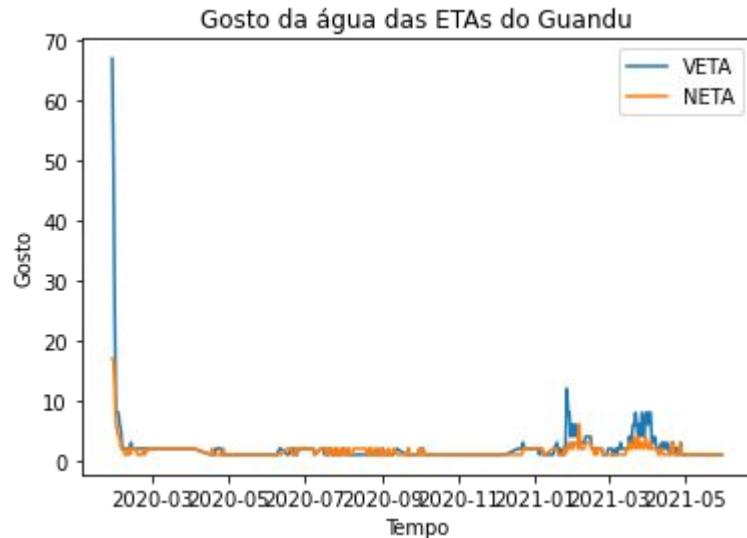


$$\eta = 1 - \frac{\text{Saida}}{\text{Entrada}}$$

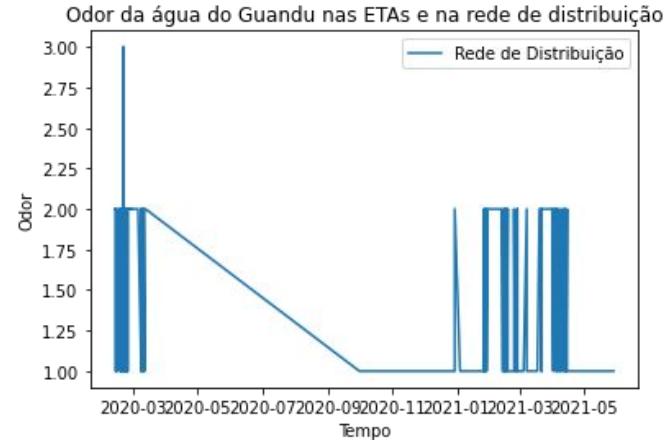
Limites de eficiência:

- $\eta = 1$  (Tratamento perfeito)
- $\eta = 0$  (Nenhum Tratamento)

## Gosto e odor na saída de tratamento (29/01/2020-)

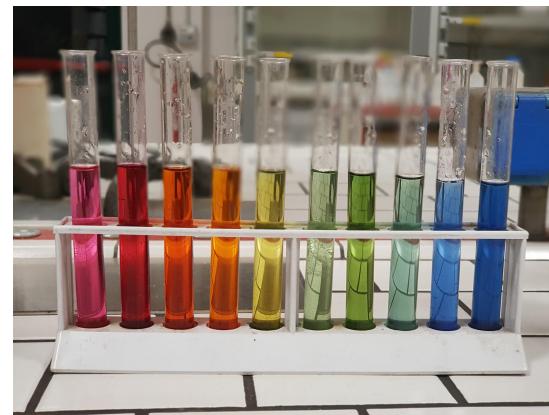
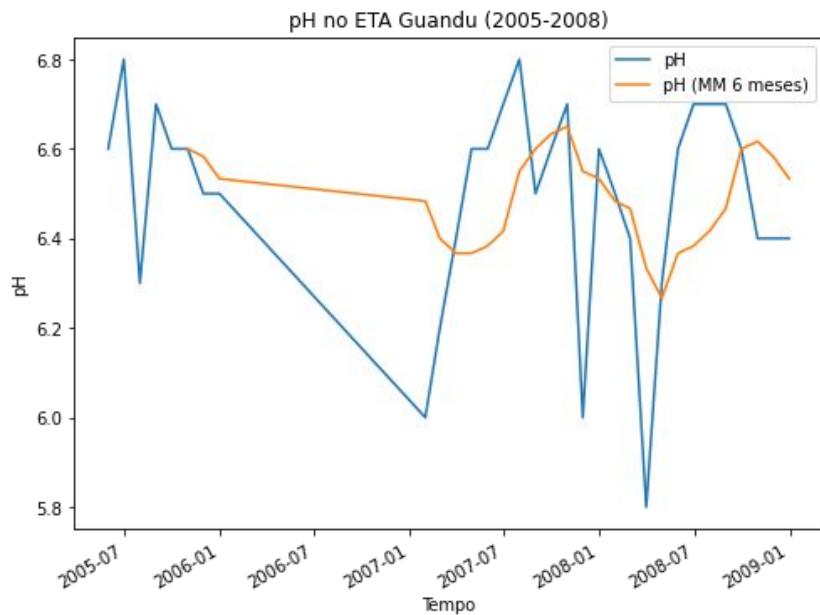


# Gosto e odor na rede de distribuição (13/02/2020-)



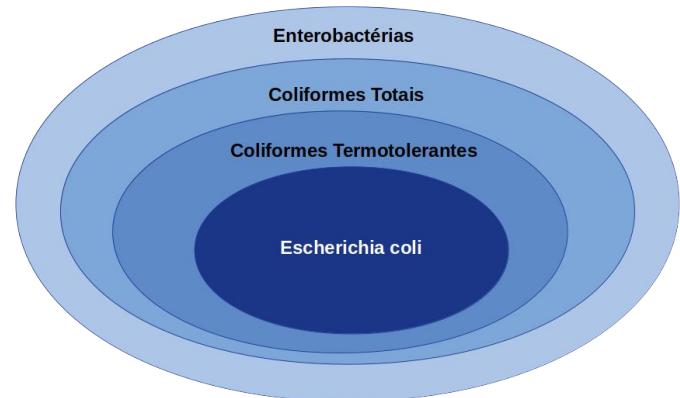
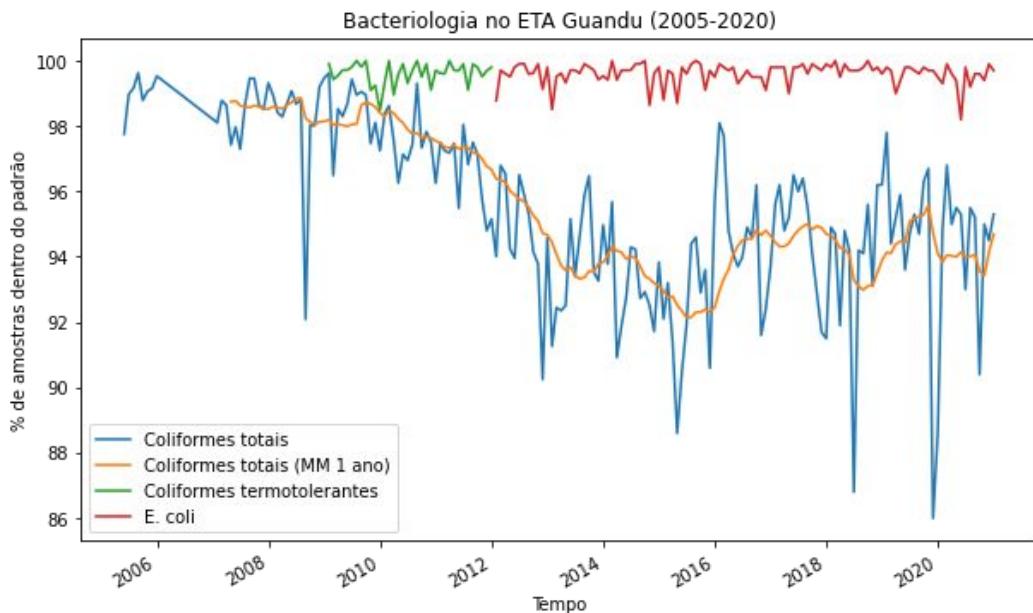
O quantitativo de amostras na rede de distribuição pode ter reduzido entre março e setembro 2020 devido à pandemia de COVID-19  
(Autorizado pela Secretaria de Estado de Saúde em resposta ao Ofício CEDAE/DSG N º 12/2020)

## pH (05/2005-12/2008)



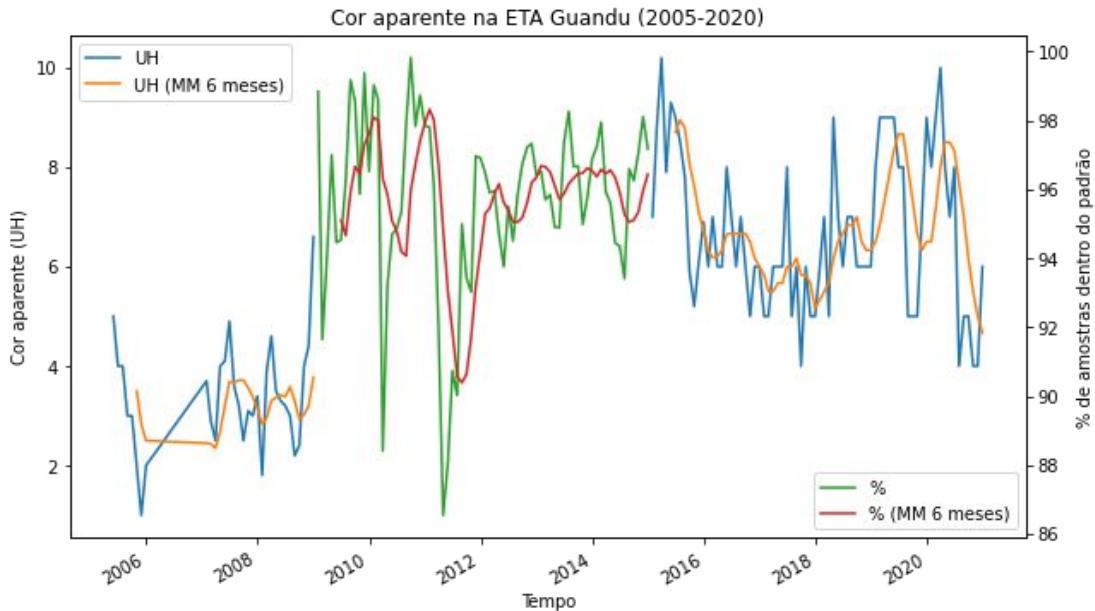
Potencial Hidrogeniônico ou pH

# Bacteriologia (05/2005-12/2020)



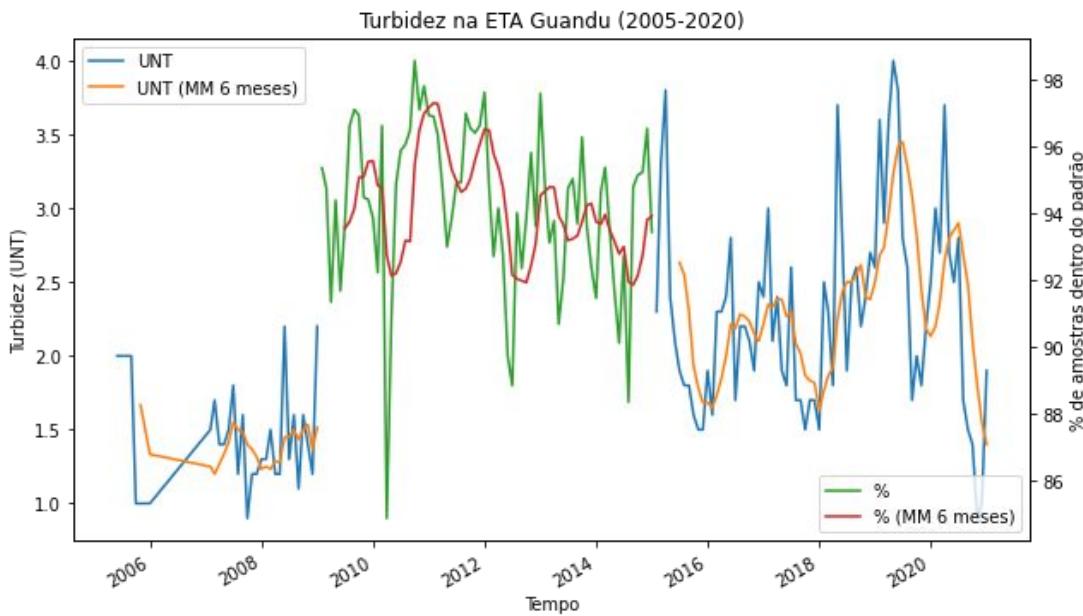
Taxonomia das bactérias coliformes

## Cor Aparente (05/2005-12/2020)



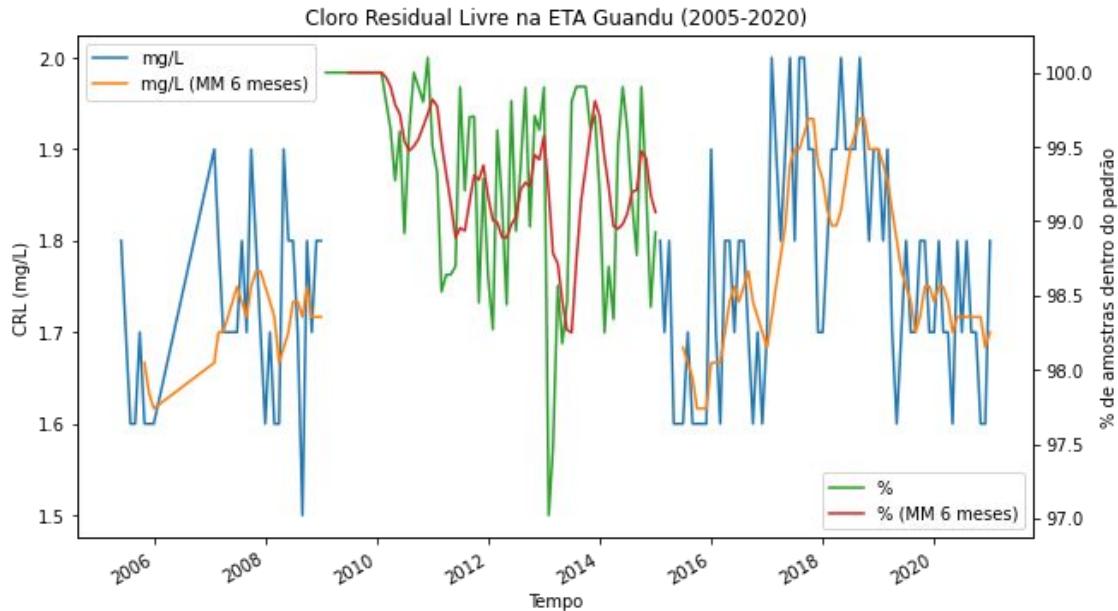
- **UH** = Unidade de Hazen
- Valor máximo permitido (VMP) de **15 UH**

## Turbidez (05/2005-12/2020)



- **UNT** = Unidades Nefelométricas de Turbidez
- Valor máximo permitido (VMP) de **5 UNT**

# Cloro Residual Livre (05/2005-12/2020)

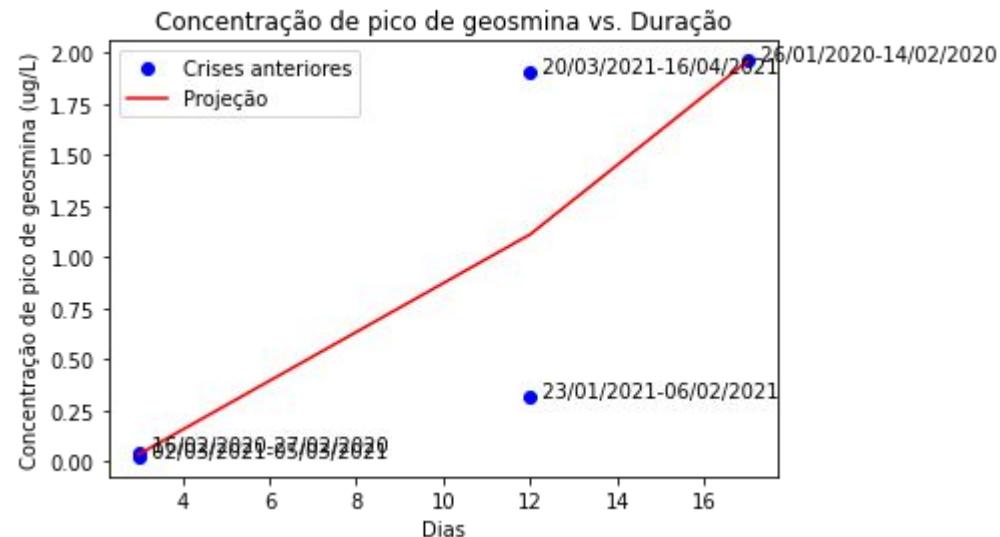
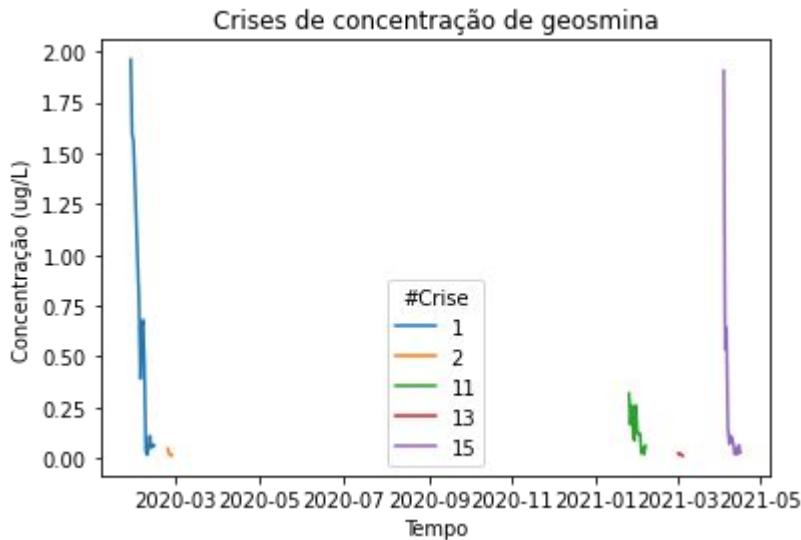


- Valor permitido entre 0,2 e 5,0 mg/L

**Quanto tempo dura uma crise  
de qualidade da água de uma  
dada magnitude no Guandu?**

---

## Concentração máxima x Duração de decaimento (26/01/2020-)

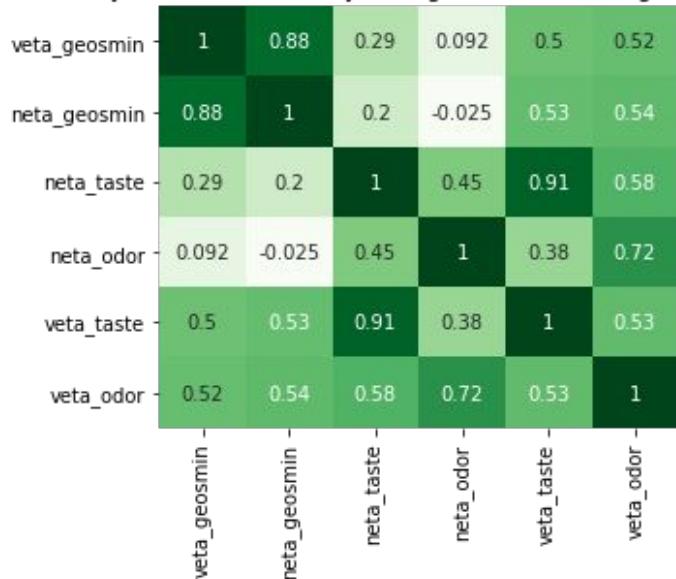


**Como os parâmetros de  
qualidade da água no Guandu  
se correlacionam?**

---

# Correlação de parâmetros 1 (29/01/2020-)

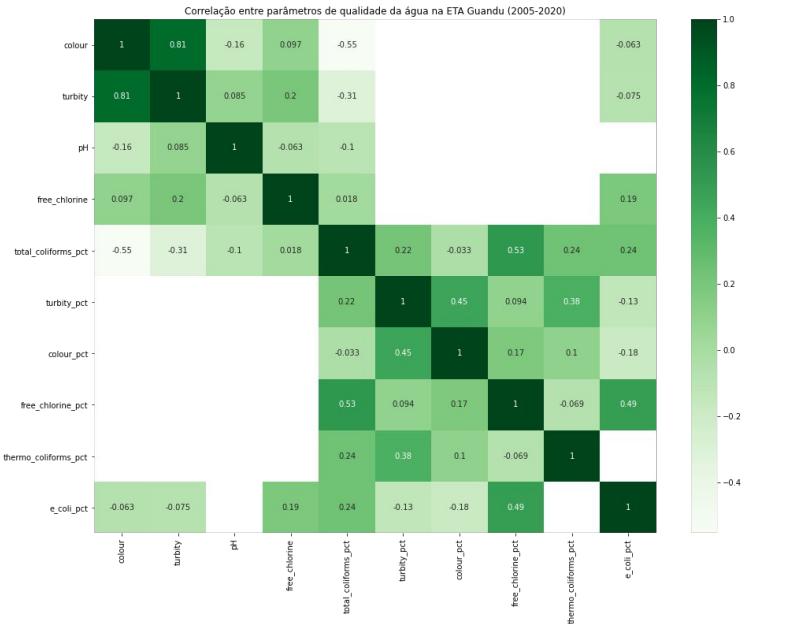
Correlação entre concentração de geosmina, odor e gosto nas ETAs



Número de amostras por parâmetro:

veta_geosmin	63
neta_geosmin	23
neta_taste	346
neta_odor	346
veta_taste	346
veta_odor	346

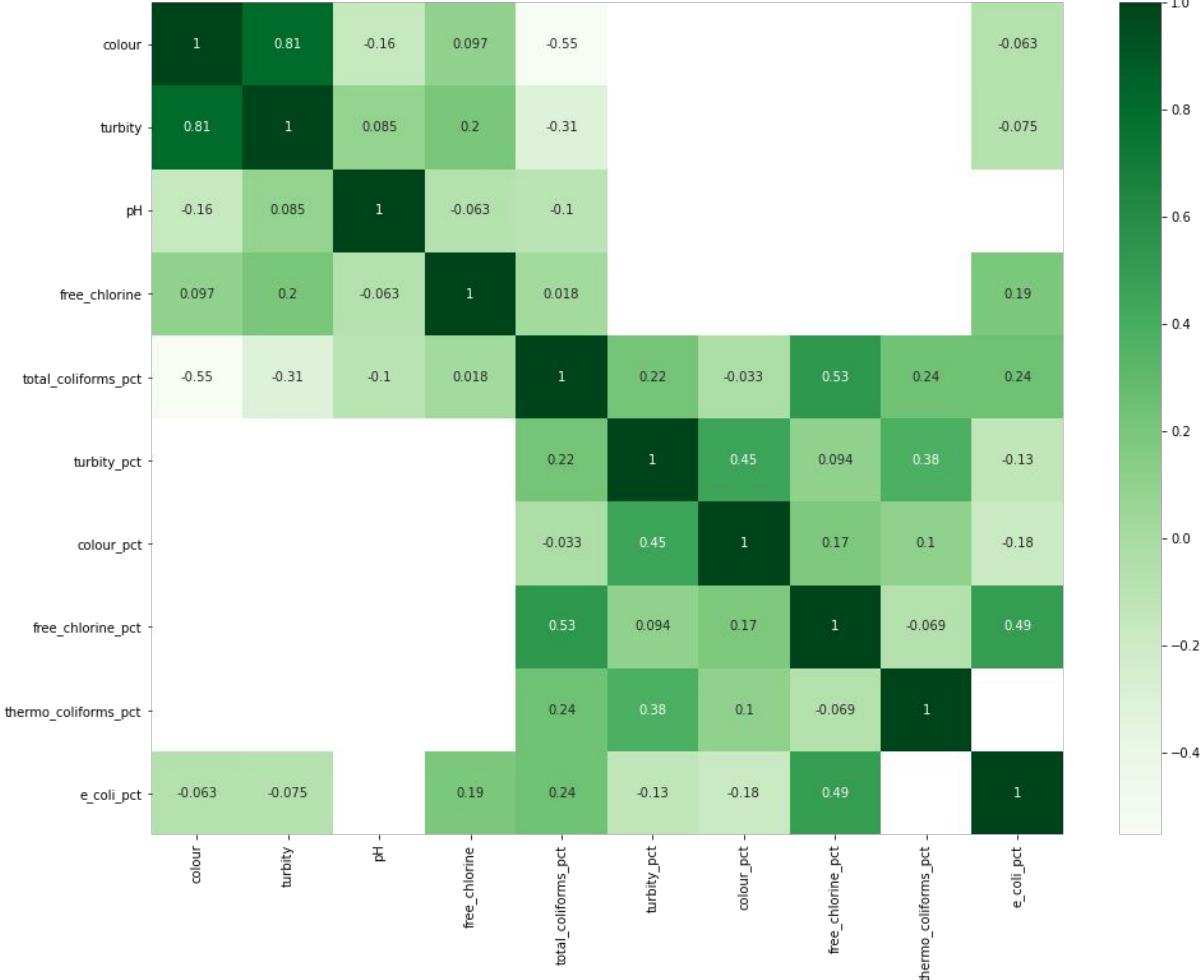
# Correlação entre parâmetros de qualidade da água na ETA Guandu (2005-2020)



Número de amostras por parâmetro:

colour	104
turbity	104
pH	32
free_chlorine	104
total_coliforms_pct	176
turbidity_pct	72
colour_pct	72
free_chlorine_pct	72
thermo_coliforms_pct	36
e_coli_pct	108

Correlação entre parâmetros de qualidade da água na ETA Guandu (2005-2020)



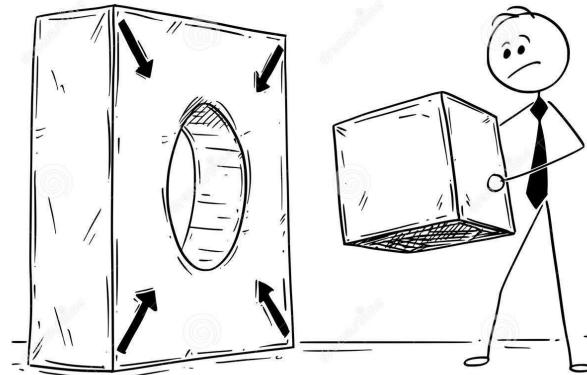
# Considerações Finais

---

---

# Problemas enfrentados

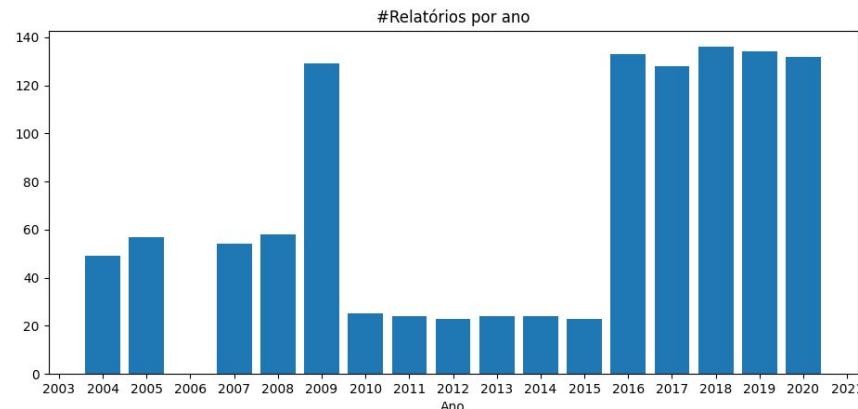
- Relatórios com formatos diferentes
  - Quase todo ano, o formato de relatório muda
  - Algumas tabelas, na verdade, são imagens
- Tabelas com dados diferentes
  - Medição da média x Medições individuais
- Dados insuficientes
  - Poucas crises de qualidade da água
  - Janelas de tempo sem medições
- Dados com pouca precisão
  - Limite superior
- Dados inconsistentes
  - Erros tipográficos



---

## Trabalhos futuros

- Realizar outras análises em cima dos dados já extraídos e tratados do Guandu
- Extrair e analisar dados de outros sistemas de abastecimento



---

## Código-fonte público no GitHub

<https://github.com/guidanoli/cedae>

- Todos os cadernos Jupyter
- Todos os relatórios anuais em PDF
- Todos os relatórios anuais do Guandu em CSV



**Obrigado pela atenção!  
Dúvidas?**



*Fonte: G1*