

# Projeto CSVac

## Equipe JHGMA

Membros da equipe:

- Gabriella Serrano Santana - 216698
- José Henrique Dioz Récio - 176622
- Maria Angélica Lobo Alves Paulino - 183465

# CSVac - Equipe JHGMA

Diante a pandemia de COVID-19 declarada em março de 2020 e que se estende até os dias atuais, surgiu a necessidade do gerenciamento dos dados de ocupação de leitos, crescimento da doença e vacinação pelo Ministério da Saúde. Eles são disponibilizados no DATASUS com a finalidade de compreender melhor a situação do país e otimizar a gestão de recursos.

Nosso projeto consiste em analisar, majoritariamente, os seguintes bancos de dados:

- Campanha Nacional de Vacinação contra Covid-19;
- Registro de Ocupação Hospitalar COVID-19;
- Base de dados do CNES;
- Sinopse do Censo Demográfico 2010 Sergipe.



SUS

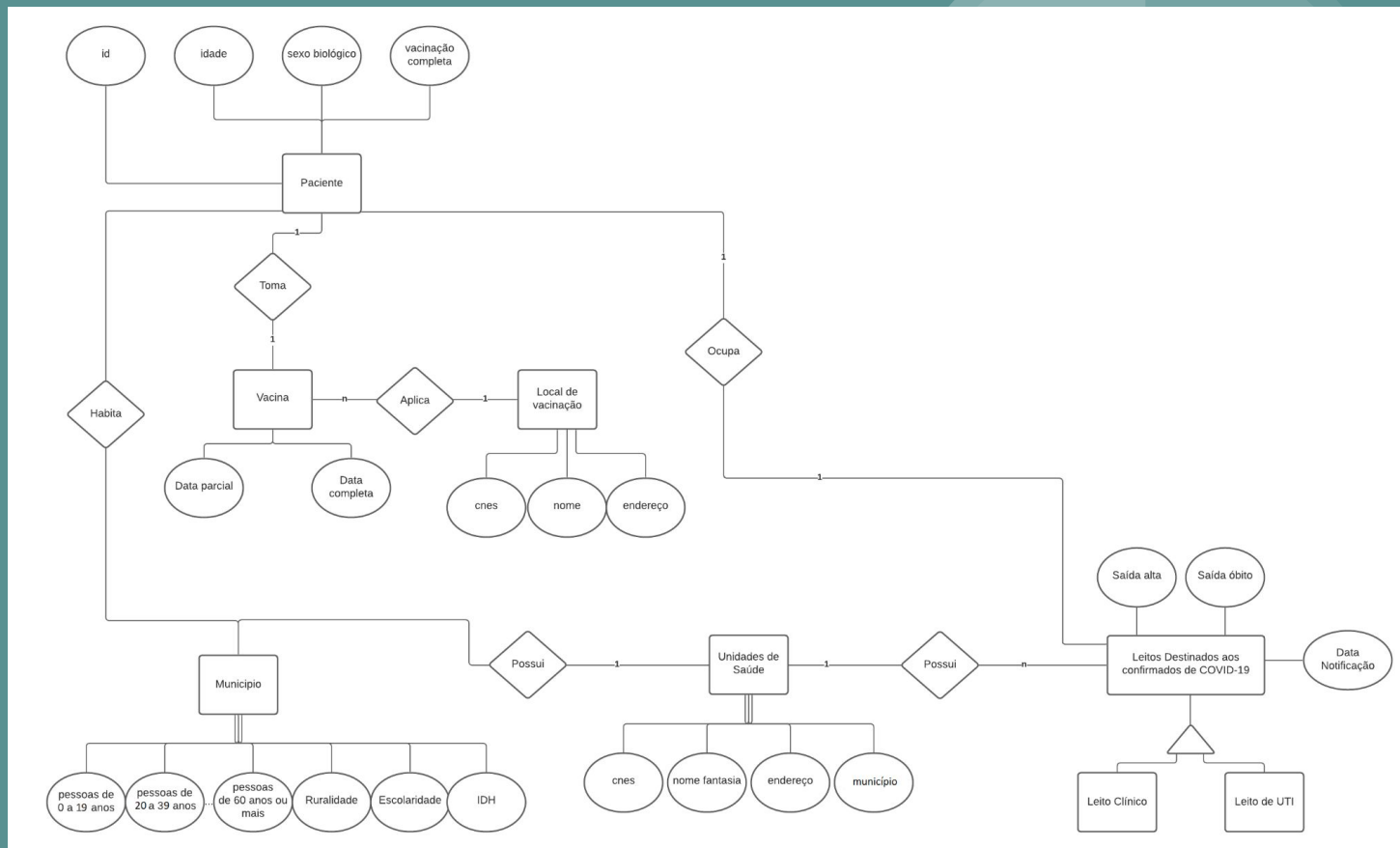
openDataSUS

# CSVac - Equipe JHGMMA

- Nossa proposta é relacionar a vacinação contra a COVID-19, a ocupação de leitos, escolaridade, ruralidade, entre outros aspectos que consideramos relevantes. Eles serão detalhados ao longo do projeto;
- Essa análise é muito importante para combater a desinformação e promover a ciência e seus impactos no bem estar da população;
- Faremos um recorte dessa análise para o estado de Sergipe.



# Modelo conceitual



# Modelo Lógico de Tabelas

Vacinação(*municipios*, *date*, vacinados\_parcial\_F\_19, vacinados\_parcial\_M\_19, vacinados\_completo\_F\_19, vacinados\_completo\_M\_19, vacinados\_parcial\_F\_39, vacinados\_parcial\_M\_39, vacinados\_completo\_F\_39, vacinados\_completo\_M\_39, vacinados\_parcial\_F\_59, vacinados\_parcial\_M\_59, vacinados\_completo\_F\_59, vacinados\_completo\_M\_59, vacinados\_parcial\_F\_200, vacinados\_parcial\_M\_200, vacinados\_completo\_F\_200, vacinados\_completo\_M\_200)

municipios chave estrangeira -> Município(nome)

Local\_de\_Vacinação(*nome*, cnes, cidade, qtd\_vacina)

Leitos(*cnes*, *mes*, ocupaçãoConfirmadoCli, ocupaçãoConfirmadoUti, saidaConfirmadaObitos, saidaConfirmadaAltas)

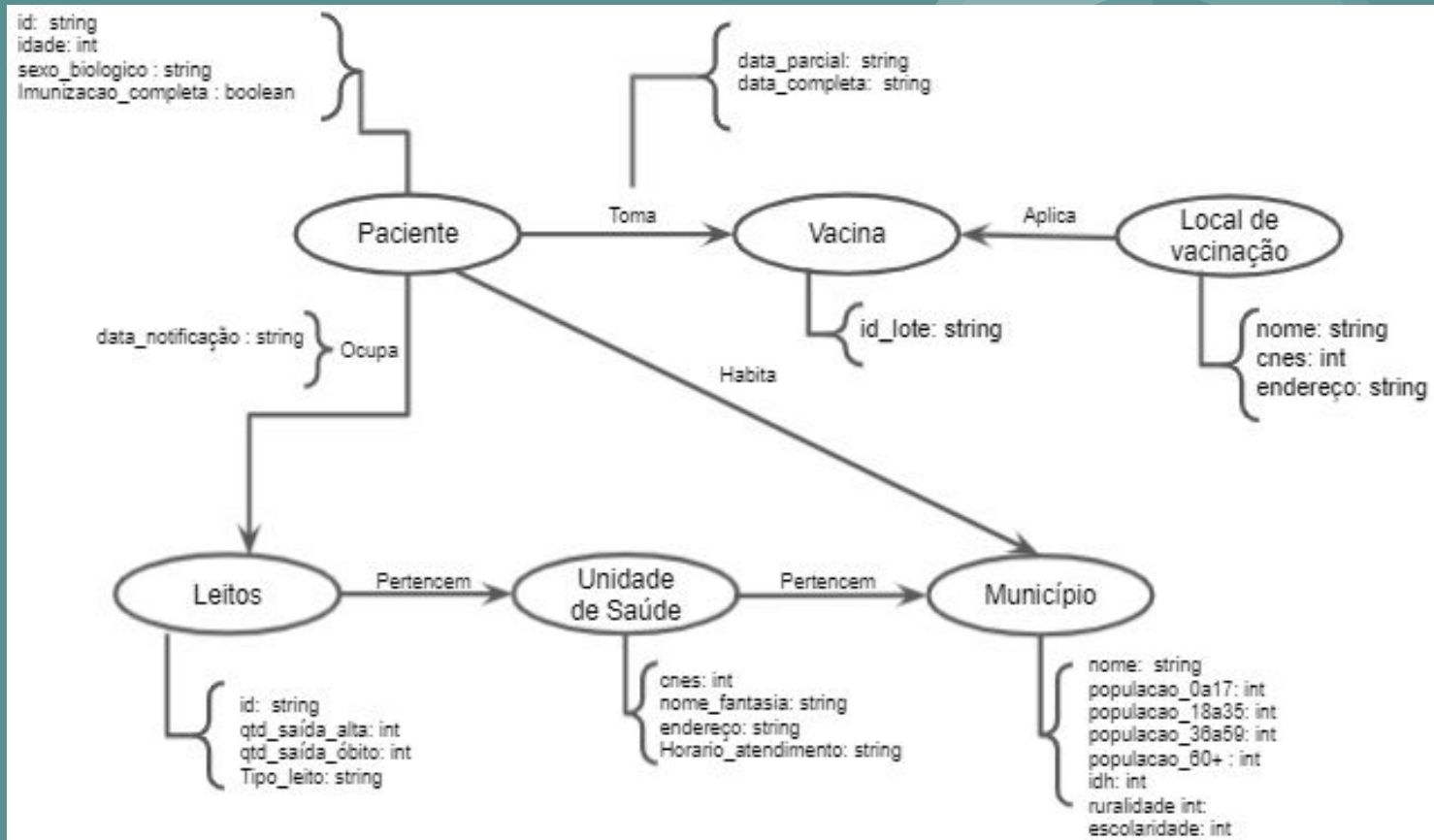
cnes chave estrangeira -> Unidade\_de\_saúde(cnes)

Unidade\_de\_saúde(*cnes*, cidade, nome\_fantasia, logradouro)

cidade chave estrangeira -> Município(nome)

Município(*nome*, população, idhm, escolarização, ruralidade, pop\_0a19anos, pop\_20a39anos, pop\_40a59anos, pop\_60+)

# Modelo Lógico de Grafos



# Dataset Publicado

## final\_vacinacao:

Esse arquivo modelo csv apresenta o número de homens e mulheres com vacinação completa e parcial de cada faixa etária para todas as cidade de Sergipe por mês.

## leitos\_final:

Esse arquivo modelo csv apresenta o número de leitos clínicos e de UTI ocupados por pacientes com COVID-19, além do número de óbitos e de altas registrados por mês em cada unidade de saúde de Sergipe.

## locais\_vacina:

Esse arquivo modelo csv apresenta a quantidade de vacinas aplicadas em cada local de vacinação e seu respectivo município.

## municipios\_final:

Esse arquivo modelo csv apresenta a população total e dividida por faixa etária, a ruralidade, o IDHM e a escolaridade de cada município de Sergipe.

## ubs-se:

Esse arquivo modelo csv apresenta o nome, o endereço, o cnes e a cidade de cada unidade de saúde de Sergipe.

# Fontes de dados

## Campanha Nacional de Vacinação contra Covid-19:

Essa base de dados registra todos os dados referentes a vacinação, sendo estas: informações do paciente, local de vacinação, data, marca da vacina, tipo de dose.

## Registro de Ocupação Hospitalar COVID-19:

Esse banco de dados controla os leitos hospitalares de COVID-19, registrando altas, óbitos e ocupação dos leitos em cada unidade de saúde do país.

## Base de dados do CNES:

Essa base de dados armazena o cadastro nacional de estabelecimentos de saúde e suas informações principais, como nome, cidade, endereço e horário de funcionamento.

## Censo Demográfico 2010 Sergipe via TABnet DATASUS:

Esse site permitiu gerar uma tabela (.csv) com os dados demográficos de Sergipe, com a quantidade de população por faixa etária em cada município do estado.



# Operações para a construção do dataset

- Integração de dados de múltiplas fontes:

## Construção de municipios.csv

A construção dessa tabela consistiu apenas no join de duas outras tabelas: municipio\_intermed.csv (df2) e população\_faixa\_etaria.csv (df).

```
result = df2.join(df[['0 a 19', '20 a 39', '40 a 59', '60+']])
```



# Operações para a construção do dataset

- Transformação de dados para facilitar análise e pesquisa:
  - Separação em vacinação parcial e completa;

```
df = df.assign(vacinacao_completa=(df.vacina_descricao_dose == '2ª Dose').where((df.vacina_fabricante_nome != 'JANSSEN'), True))
```

- Agrupamento de idades por data de nascimento;

```
age_groups = [(0, 19), (20, 39), (40, 59), (60, 200)]
```

- Agrupamento por sexo;

```
temp1_F = df[(df['paciente_idade'] > group[0]) &  
(df['paciente_idade'] < group[1]) &  
(df['vacina_dataaplicacao'] < date) &  
(df['vacinacao_completa'] == False) &  
(df['paciente_enumsexobiologico'] == 'F')].groupby(['estabelecimento_municipio_nome']).size().to_frame('vacinados_parcial_F_' + str(group[1]))  
df_ages = df_ages.merge(temp1_F, how='left', left_on='municipios', right_on='estabelecimento_municipio_nome').fillna(0)
```

# Operações para a construção do dataset

- Tratamento de dados;
  - por exemplo, tratamento de dados faltantes, incompletos e inconsistentes.

## Construção de leitos\_final.csv

Para a construção do .csv de leitos tivemos que importar o registro de leitos do dataset original disponibilizado no datasus, e utilizar métodos da biblioteca pandas para padronizar e selecionar somente o mês da coluna referente as datas de registros. Em seguida, selecionamos apenas as colunas desejadas agrupando por CNES e por mês, somando as ocupações, óbitos e altas, conforme abaixo.

```
df['dataNotificacao'] = pd.to_datetime(df['dataNotificacao'], format="%Y-%m-%dT%H:%M:%S.%f", errors='coerce')
df['mes'] = pd.to_datetime(df['dataNotificacao']).dt.month
leitos_final = df.groupby(['cnes', 'mes'], as_index=False)[['ocupacaoConfirmadoCli', 'ocupacaoConfirmadoUti', 'saidaConfirmadaObitos', 's
```

# Perguntas do modelo lógico de tabelas

- Municípios com boa escolaridade possuem maiores índices de vacinação? Tendo isso em vista, pode-se afirmar que uma boa educação ajuda na conscientização da necessidade de tomar vacinas?

```
SELECT
    fv.municipio,
    ("vacinados_completo_F_200" +
     "vacinados_completo_F_59" +
     "vacinados_completo_F_39" +
     "vacinados_completo_F_19" +
     "vacinados_completo_M_200" +
     "vacinados_completo_M_59" +
     "vacinados_completo_M_39" +
     "vacinados_completo_M_19")::decimal/mf.populacao AS "taxa de vacinados",
    escolarizacao
FROM final_vacinacao fv
JOIN municipios_final mf on fv.municipio = mf.municipio
WHERE date = '2021-12-01'
ORDER BY escolarizacao DESC;
```



# Perguntas do modelo lógico de tabelas

- O índice de ruralidade do município influência nos óbitos e na velocidade/aderência da vacinação?
  - Não encontramos relação entre os parâmetros avaliados. A query encontra-se no github do projeto.
- O IDHM (IDH do município) tem alguma relação com o índice de pessoas vacinadas? E de mortalidade?
  - Não encontramos relação entre os parâmetros avaliados. A query encontra-se no github do projeto.
- Qual o percentual por faixa etária de pessoas que foram vacinadas em um determinado período de tempo em determinada cidade? Com base nisso, qual deve ser o maior público alvo da campanha de vacinação durante esse período?
  - Podemos analisar o resultado da query, que encontra-se no github do projeto, para responder a essa pergunta.



# Perguntas do modelo lógico de tabelas

- Qual o percentual por faixa etária de pessoas que foram vacinadas em um determinado período de tempo em determinada cidade? Com base nisso, qual deve ser o maior público alvo da campanha de vacinação durante esse período?

município	"vacinados 60+"	"vacinados 40-59"	"vacinados 20-39"	"vacinados 0-19"
POCO REDONDO	1.0863993385696569	0.92107852784884865184	0.40727422171992191513	0.01196022529726722759
SIMAO DIAS	0.87283976957542137828	0.16963350785340314136	0.07745795601552393273	0.0007696075001749108
NOSSA SENHORA DO SOCORRO	1.5903914590747331	0.48508319835075835665	0.12551050050409254797	0.00113551518168242907
NOSSA SENHORA DA GLORIA	1.2615740740740741	1.0518355128407121	0.65585157829948122747	0.02445026593633404779
ITABAIANINHA	1.1804183614177804	0.84670608108108108108	0.42997698960564944854	0.00853651047104341952
ESTANCIA	1.2647651006711409	0.70343231777863478596	0.20701222314125539932	0.00297485635111455235
ITABAIANA	1.3460772104607721	1.00028607392150131594	0.52037708954256835558	0.02210116731517509728
RIBEIROPOLIS	1.1256855575868373	1.0169345579793341	0.45899948427024239299	0.00420450723175243861
CAPELA	1.0260898282694848	0.94956927494615936827	0.39898447417244409726	0.01114560942600111456
TOMAR DO GERU	1.047008547008547	0.74297012302284710018	0.26782692772540488277	0.00417061611374407583
TOBIAS BARRETO	1.1786046511627907	0.84590739975196362133	0.34956296536095823891	0.00793216630196936543
RIACHUELO	1.1001451378809869	0.98733413751507840772	0.45361753227259081357	0.02036021926389976507
RIACHAO DO DANTAS	1.1360116166505324	0.84766722878021360315	0.33182730923694779116	0.00460711543383670335
MOITA BONITA	1.1123066577000672	0.99458483754512635379	0.52125897294312534511	0.0096956638836520334
MONTE ALEGRE DE SERGIPE	1.1726618705035971	0.84294068504594820384	0.34983351831298557159	0.01215189873417721519

# Perguntas do modelo lógico de tabelas

- É possível observar queda nos óbitos mensais do estado a medida em que as pessoas são vacinadas?
  - A query completa encontra-se no github do projeto.

	"total de obitos por mes" ▾	"vacinados por mes" ▾	mes ▾
1	76947	44678	4
2	27434	849684	12
3	32073	849684	11
4	33003	25	2
5	31026	0	1
6	66785	715804	10
7	84895	133844	5
8	83875	176552	6
9	66763	19893	3
10	93350	219447	7
11	79534	466383	9
12	94944	323291	8



- A partir do mês 7 observa-se diminuição dos óbitos com o aumento contínuo da vacinação.



# Evolução do projeto

- Apenas relacionar os dados de vacinação e registro de leitos → Perguntas e análises muito limitada;
- Utilizar novos bancos externos e fatores (dados populacionais, IDH, escolarização, ruralidade, dados sobre as unidades de saúde e locais de vacinação) → Perguntas de análise mais complexas e relevantes;
- Grandes modificações nos modelos conceitual e lógico relacional:
  - Modelo conceitual → Inserção dos objetos Local de vacinação e Unidade de saúde, e suas respectivas propriedades e relações, além da adição de outras propriedades em objetos que já existiam;
  - Modelo lógico relacional --> Acréscimo de novas entidades e alteração das já existentes inserindo e excluindo atributos.
- Dificuldades:
  - Aumento do volume de dados → processamento mais lento, demora em testes da construção;
  - Dados desorganizados ou mal coletados → campos em formatos errados, dados faltantes;
  - Organização de quais dados seriam de fato essenciais para as queries.