

**Universidade Federal do Rio de Janeiro**



**Disciplina: Tópicos Especiais Sist. Informação I**

**Alunos: ALESSANDRA PIERRO SARAIVA, FABIANE MONTEIRO CARVALHO,  
MASSAO OLIVEIRA IWANAGA**

## **PROJETO DO TRABALHO DA DISCIPLINA TÓP. ESP. SIST. INF. I**

### **1. Detalhamento e descrição do problema/questão de análise**

A humanidade vem enfrentando uma pandemia sem precedentes através da doença contagiosa chamada COVID19 (Coronavirus disease 2019). No contexto brasileiro, com o início e avanço, no ano de 2021, da vacinação contra essa enfermidade, fazem-se necessárias análises para avaliar o perfil dos vacinados nas diferentes regiões do país devido às suas desigualdades, principalmente a norte, com 18,7 milhões de habitantes aproximadamente, que foi muito afetada nos últimos meses por nova variante do coronavírus e também a região sudeste, por ser a mais populosa, com sua população em torno de 89 milhões de habitantes [3].

Verifica-se que essas análises poderão contribuir para um estudo mais detalhado sobre a cobertura da vacinação nessas regiões com seus quantitativos e perfis, permitindo tomadas de decisões através dos dados apresentados para aprimorar o processo de vacinação nesses locais, com informações sobre os estabelecimentos de saúde, de forma a diminuir a aglomeração de pessoas no atendimento e também comparar a distribuição desses estabelecimentos nas duas regiões do país.

### **2. Descrição do dataset e sua utilização**

Neste trabalho, serão usados os dois datasets listados a seguir, que podem ser obtidos no Portal Brasileiro de Dados Abertos, ferramenta disponibilizada pelo governo para que todos possam encontrar e utilizar os dados abertos e informações públicas<sup>1</sup>.

## 1. Dataset da Campanha Nacional de Vacinação contra Covid-19 [1]

Este dataset possui informações sobre os vacinados (idade, data de nascimento, sexo, raça e etc.), os estabelecimentos de saúde (nome, município, estado e etc.) onde ocorreu a vacinação, grupo de atendimento ao qual o vacinado pertence e da vacina (fabricante, lote e etc.), além de outras.

Esse dataset contém um dicionário de dados, conforme exibido a seguir, contendo os 32 campos e suas descrições.

Ordem	Campo	Descrição	Categoria
1	document_id	Identificador do documento	
2	paciente_id	Identificador do vacinado	
3	paciente_idade	Idade do vacinado	
4	paciente_dataNascimento	Data de nascimento do vacinado	
5	paciente_enumSexoBiologico	Sexo do vacinado	M = Masculino, F = Feminino
6	paciente_racaCor_codigo	Código da raça/cor do vacinado	1; 2; 3; 4; 99
7	paciente_racaCor_valor	Descrição da raça/cor do vacinado	1 = Branca; 2 = Preta; 3 = Parda; 4 = Amarela; 99 = Sem informação
8	paciente_endereco_colibgeMunicipio	Código IBGE do município de endereço do vacinado	
9	paciente_endereco_coPais	Código do país de endereço do vacinado	
10	paciente_endereco_nmMunicipio	Nome do município de endereço do vacinado	
11	paciente_endereco_nmPais	Nome do país de endereço do vacinado	
12	paciente_endereco_uf	Sigla da UF de endereço do vacinado	
13	paciente_endereco_cep	5 dígitos para anonimizado e 7 dígitos para identificado	
14	paciente_nacionalidade_enumNacionalidade	Nacionalidade do vacinado	
15	estabelecimento_valor	Código do CNES do estabelecimento que realizou a vacinação	
16	estabelecimento_razaoSocial	Nome/Razão Social do estabelecimento	
17	estabelecimento_noFantasia	Nome fantasia do estabelecimento	
18	estabelecimento_municipio_codigo	Código do município do estabelecimento	
19	estabelecimento_municipio_nome	Nome do município do estabelecimento	
20	estabelecimento_uf	Sigla da UF do estabelecimento	
21	vacina_grupo_atendimento_code	Código do grupo de atendimento ao qual pertence o vacinado	
22	vacina_grupo_atendimento_nome	Nome do grupo de atendimento ao qual pertence o vacinado	
23	vacina_categoria_code	Código da categoria	
24	vacina_categoria_nome	Descrição da categoria	
25	vacina_lote	Número do lote da vacina	
26	vacina_fabricante_nome	Nome do fabricante/fornecedor	
27	vacina_fabricante_referencia	CNPJ do fabricante/fornecedor	
28	vacina_dataAplicacao	Data de aplicação da vacina	
29	vacina_descricao_dose	Descrição da dose	
30	vacina_codigo	Código da vacina	
31	vacina_nome	Nome da vacina/produto	
32	sistema_origem	Nome do sistema de origem	

Figura 1 - Dicionário de dados do dataset da Campanha Nacional de Vacinação contra Covid-19

Os dados em si, dão um total de **8.958.579** registros, incluído o cabeçalho, e 3.6 GB de tamanho físico, disponibilizados em formato CSV (*Comma-separated values*).

<sup>1</sup> <https://dados.gov.br/>

## 2. Cadastro Nacional de Estabelecimentos de Saúde (CNES) - Ativo [2]

Este dataset possui informações sobre todos os estabelecimentos (código, nome, endereço, tipo e etc.) de saúde ativos no país, públicos ou não. Apesar de não haver disponível um dicionário de dados, criamos um, conforme exibido a seguir, a partir de informações contidas na página do dataset e no cabeçalho do arquivo.

Ordem	Campo	Descrição
1	co_cnes	Código CNES
2	co_ibge	Código IBGE
3	no_fantasia	Nome Fantasia
		Tipo de estabelecimento de saúde: Posto de Saúde, Centro de Saúde/Unidade Básica de Saúde, Policlínica, Hospital Geral, Hospital Especializado, Unidade Mista, Pronto Socorro Geral, Pronto Socorro Especializado, Consultório Isolado, Unidade Móvel Fluvial, Clínica Especializada/Amb. Especializado, Unidade de Serviço de Apoio de Diagnóstico e Terapia, Unidade Móvel Terrestre, Unidade Móvel de Nível Pré-hospitalar na Área de Urgência e Emergência: Farmácia: Unidade de Vigilância em Saúde, Cooperativa, Centro de Parto Normal Isolado, Hospital / Dia-Isolado, Central de Regulação de Serviços de Saúde, Laboratório Central de Saúde
4	ds_tipo_unidade	Pública – LACEN, Secretaria de Saúde
5	tp_gestao	Tipo de gestão: Estadual (E), Municipal (M) ou Dupla (D)
6	no_logradouro	Nome do Logradouro
7	nu_endereco	Número do Endereço
8	no_bairro	Bairro
9	co_cep	CEP
10	uf	UF do estabelecimento
11	municipio	Município
12	nu_telefone	Número de Telefone

Figura 2 - Dicionário de dados do dataset do Cadastro Nacional de Estabelecimentos de Saúde (CNES) - Ativo

Os dados em si, dão um total de 354.805 registros, incluído o cabeçalho, e 56MB de tamanho físico, disponibilizados em formato CSV.

## 3. Objetivos

### 3.1 Objetivos gerais

Aplicação de técnicas de data science que subsidiem a análise do problema como:

- Importação, limpeza e tratamento de dados dos datasets;
- Análise exploratória dos datasets;
- Visualização de dados que auxiliem na análise do problema proposto

Esse trabalho será implementado na linguagem python e deverá prever a reutilização do código, bem como informações de proveniência sobre o ambiente computacional e os dados utilizados.

### **3.2 Objetivos específicos**

Fornecer informações sobre os vacinados e também dos seus locais de vacinação por município e estados das regiões norte e sudeste.

Serão apresentados como resultados:

1 - O perfil dos vacinados (sexo, idade, raça, grupo) por estado/região nas duas regiões e quantitativos;

2 - O perfil dos vacinados (sexo, idade, raça, grupo) por tipo de estabelecimento de saúde nos estados/regiões nas duas regiões e seus quantitativos;

3 - O quantitativo dos estabelecimentos de saúde e seus tipos nos estados nas duas regiões por total de população;

4 - Os municípios de residência dos vacinados e dos estabelecimentos de saúde onde os mesmos se vacinaram para verificar se são iguais ou se os vacinados se deslocaram para outros municípios.

5- Visualização de % da população vacinada ao longo do tempo, nas duas regiões, permitindo uma análise comparativa da evolução da vacinação entre as regiões.

### **4. Métodos de data cleaning / tratamento de dados**

Será utilizada a biblioteca Pandas para importar os arquivos através do comando `read_csv` pois ambos os datasets estão nesse formato (.csv).

Como os dois datasets possuem informações sobre todos os estados brasileiros, visando reduzir a quantidade de dados a serem analisados e trazer eventuais ganhos de performance, será feita uma seleção inicial nos dataframes para utilizar apenas das informações dos estados das regiões norte e sudeste, que serão analisadas neste trabalho.

Será feita uma análise dos campos que são relevantes para o estudo. Para esses campos:

- Análise inicial dos dados para conhecer um pouco seu conteúdo e as características dos dados: tipos dos dados, identificar dados quantitativos e

qualitativos, gerar estatísticas básicas (valores mínimos e máximos, média, mediana)

- Identificação de valores nulos ou faltantes e definição de estratégia de tratamento (atribuição de valor ou eliminação da linha)
- Identificação de registros duplicados e definição de estratégia de tratamento.

Será feito um merge para que seja possível verificar o perfil dos vacinados por tipo de estabelecimento (hospital, centro de saúde, posto de saúde e etc.)

## **5. Proposta de modelo de extração de conhecimento e visualização dos dados**

Nesta etapa, será utilizada a biblioteca matplotlib para que seja possível construir os gráficos definidos no item 3.2 acima, utilizando os seus principais comandos (plot, show, grid e etc.)

## **6. Referências**

[1] OpenDataSus. **Campanha Nacional de Vacinação contra Covid-19 2021**. Disponível em: <<https://opendatasus.saude.gov.br/dataset/covid-19-vacinacao>> Acesso em: 05 de março de 2021.

[2] Ministério da Saúde. **Cadastro Nacional de Estabelecimentos de Saúde (CNES) 2018**. Disponível em: <[https://dados.gov.br/dataset/cnes\\_ativo](https://dados.gov.br/dataset/cnes_ativo)> Acesso em: 05 de março de 2021.

[3] SIDRA - Sistema IBGE de Recuperação Automática. População Residente Estimada - Ano 2020. Disponível em: <[sidra.ibge.gov.br](https://sidra.ibge.gov.br)>. Acesso em: 05 de março de 2021.