

Análise dos dados de vacinação da COVID-19 e seus estabelecimentos de saúde nas regiões norte e sudeste do Brasil

Alessandra Pierro Saraiva, Fabiane Monteiro Carvalho e Massao Oliveira Iwanaga

Universidade Federal do Rio de Janeiro
Av. Pedro Calmon no 50, 2o andar, Ilha do Fundão - Rio de Janeiro, RJ, 21941-596, Brasil
alessandrapsaraiva@gmail.com, famcrj@gmail.com,
massaoiwanaga@ufrj.br

Resumo. A humanidade vem enfrentando uma pandemia sem precedentes através da doença contagiosa chamada COVID19 (Coronavirus disease 2019). No contexto brasileiro, fazem-se necessárias análises sobre a vacinação da COVID19 para avaliar o perfil dos vacinados e de seus estabelecimentos de saúde. O presente artigo apresenta o resultado de um estudo da vacinação nas regiões norte e sudeste do país, duas regiões antagônicas no que se refere a desenvolvimento humano. Foram aplicadas técnicas de Data Science (DS) nos *datasets* de vacinação e de estabelecimentos de saúde ativos no Brasil. Este trabalho foi implementado através do Google Colaboratory, um serviço de nuvem gratuito hospedado pela Google, sem nenhuma configuração local necessária e com acesso gratuito a GPUs (Unidades de Processamento Gráfico). A proveniência dos dados, através de técnicas implementadas nesta pesquisa, habilita também a reprodutibilidade deste experimento. Ao final, serão fornecidas as visualizações das informações necessárias para tomadas de decisão.

Palavras-chave: Data Science, Analytics, Covid19 Vaccination, Provenance.

1 Introdução

A humanidade vem enfrentando uma pandemia sem precedentes através do surto da doença contagiosa chamada COVID-19 (*Coronavirus disease* 2019). No contexto brasileiro, duas regiões foram particularmente muito afetadas. Em primeiro lugar, a região norte¹, com população de aproximadamente 18,6 milhões de habitantes², e, em

¹

<https://revistagalileu.globo.com/Ciencia/Saude/noticia/2020/06/pesquisa-aponta-regiao-norte-como-epicentro-da-covid-19-no-pais.html>. Acesso em 07/05/2021.

² <https://sidra.ibge.gov.br/home/pms/brasil>. Acesso em: 07/05/2021.

segundo, a região sudeste³, a região mais populosa, com número de habitantes em torno de 87,7 milhões⁴.

Com o início e avanço, no ano de 2021, da vacinação contra essa enfermidade em todo o país, fazem-se necessárias análises a partir dos dados dos vacinados e dos estabelecimentos de saúde para avaliar diversas questões como: perfis dos vacinados e dos estabelecimentos de saúde, se os sistemas de vacinações locais suprem a carência de sua população local ou se ela precisa se vacinar em outro município, discrepâncias e/ou desigualdades entre as regiões, velocidade de vacinação, entre outras.

Este trabalho, através de três *datasets* disponibilizados publicamente pelo governo brasileiro, busca analisar comparativamente os dados das regiões norte e sudeste do Brasil. Em termos de tecnologia, foram usados para chegar em tal fim: a linguagem de programação Python, o ambiente de desenvolvimento Google Colab, três *datasets* no formato CSV (Comma-Separated Values), o versionador de arquivos GIT e o serviço de hospedagem em nuvem Google Drive. Foram implementados também os conceitos de proveniência (utilizando-se a biblioteca PROV) e reprodutibilidade, como descrito ulteriormente.

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.

O trabalho está organizado em seis seções, já considerando esta introdução. Na segunda seção, serão abordadas sucintamente as referências bibliográficas consultadas para o referencial teórico. Na terceira, serão apresentados os *datasets* usados na implementação. Na quarta seção, será descrito o estudo de caso em si, que através da manipulação dos *datasets*, gerou os gráficos, as análises e também as informações de proveniência. Na quinta, apresenta-se uma discussão acerca dos resultados obtidos na seção anterior. Para finalizar, o trabalho se encerra com a conclusão, onde é feito o fechamento das ideias discutidas.

1.1 Objetivos gerais

Este trabalho almeja aplicar técnicas de Data Science (DS) que subsidiem análises do problema, tais como:

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

Além do listado anteriormente, a implementação prevê a reutilização do código, bem como informações de proveniência sobre os dados utilizados.

3

<http://associacaopaulistamedicina.org.br/noticia/covid-19-sudeste-e-a-regiao-mais-afetada-do-pais>. Acesso em 07/05/2021.

⁴ <https://sidra.ibge.gov.br/home/pms/brasil>. Acesso em: 07/05/2021.

1.2 Objetivos específicos

Quanto aos objetivos específicos, o presente artigo busca fornecer informações sobre os vacinados e também dos seus locais de vacinação por município, estados, regiões metropolitanas e interioranas das regiões norte e sudeste, da seguinte forma:

1. Perfil dos vacinados (sexo, idade, raça, grupo) e quantitativos nas duas regiões, dividido entre regiões metropolitanas e regiões interioranas;
2. Quantitativos de vacinados por tipo de gestão e por tipo de unidade de estabelecimento de saúde nas duas regiões metropolitanas e interioranas; e
3. Quantitativos e percentual de vacinados que tomaram a vacina em município e/ou estado diferente daquele de sua residência;

2 Trabalhos Relacionados

Em [5], Marcelo Ribeiro contribui analiticamente com o debate acerca das desigualdades de renda nas regiões metropolitanas brasileiras que, normalmente, tendem a considerar somente as características dos indivíduos como cor/raça e sexo. Marcelo defende que o grau de escolaridade tem uma correlação importante e deve ser uma variável a ser analisada no contexto de desigualdade.

Já Luiza Garnelo e Sully Sampaio, em [6], discutem as bases sócio-culturais do controle social em saúde indígena, considerando o risco negativo de se terceirizar a saúde indígena para o mercado privado. O trabalho é importante devido a quantidade proporcionalmente grande de indígenas vivendo na região norte do Brasil. Os indígenas são etnias que tiveram prioridade na vacinação contra o COVID-19.

O artigo de Maria Donalisio et. al [7] disserta, no contexto da vacinação contra influenza em idosos em municípios do sudeste, sobre os fatores associados à essa vacinação. Ela objetiva defender que, mesmo gratuita e disponível no Brasil desde 1999, a cobertura vacinal contra a influenza é inadequada em diversos municípios do país. Através dessa exposição, buscou estimular a cobertura vacinal e identificar fatores relacionados à vacinação contra a influenza em idosos. Apesar da disparidade de gravidade entre a influenza e o COVID-19, todo o aprendizado obtido através do estudo do surto da influenza e de como foi e é gerida sua vacinação, pode ser útil para ser aprimorado no combate a pandemia na qual vivemos.

3 Datasets

Neste trabalho, serão usados três *datasets* conforme descritos em seguida. Eles foram obtidos através de duas fontes: o Portal Brasileiro de Dados Abertos⁵, ferramenta disponibilizada pelo governo para que todos os brasileiros possam encontrar e utilizar os dados abertos nacionais e informações públicas; e o site do Instituto Brasileiro de Geografia e Estatística (IBGE).

⁵ <https://dados.gov.br/>. Acesso em 17/05/2021.

3.1 Dataset da Campanha Nacional de Vacinação contra COVID-19 [1]

O primeiro *dataset* utilizado é o maior e possui informações sobre os vacinados (idade, data de nascimento, sexo, raça e etc.) contra a COVID-19, 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. Ele possui relacionado um dicionário de dados, conforme exibido a seguir, contendo os 32 campos, suas descrições e categorias.

| 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_colbgeMunicipio | 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_codigo | 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 | |

Fig. 1. Dicionário de dados do *dataset* da Campanha Nacional de Vacinação contra COVID-19.

É importante notar que, como a vacinação contra o COVID-19 continua em andamento, o *dataset* é atualizado diariamente no portal. O trabalho foi realizado com a versão do *dataset* disponibilizada no dia 05/03/2021 e que possui dados de vacinação até o dia 03/03/2021. Os dados em si, são em um total de 7.908.950 registros, incluído o cabeçalho, e 3,95 GB de tamanho físico, disponibilizados em formato CSV.

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

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

| 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 Diagnose 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 |

Fig. 2. Dicionário de dados do *dataset* do Cadastro Nacional de Estabelecimentos de Saúde (CNES) - Ativo.

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

3.3 Regiões Metropolitanas das Regiões Sudeste e Norte do Brasil

O terceiro e último *dataset* foi gerado manualmente a partir do *dataset* de regiões metropolitanas do Brasil disponibilizado pelo Instituto Brasileiro de Geografia e Pesquisa [3]. Foram filtrados os municípios componentes dos estados das regiões sudeste e norte do Brasil que fossem considerados de região metropolitana. O dicionário de dados apresentado a seguir, também criado, contém a lista dos campos do arquivo e suas descrições.

| Ordem | Campo | Descrição |
|-------|---------------|-------------------------------------|
| 1 | REGIAO | Região do Brasil (NORTE ou SUDESTE) |
| 2 | UF | Código da Unidade Federativa |
| 3 | MUNICIPIO | Nome do Município |
| 4 | COD_MUNICIPIO | Código IBGE do Município |

Fig. 3. Dicionário de dados do *dataset* de Regiões Metropolitanas das Regiões Sudeste e Norte do Brasil.

É importante frisar que o estado do Acre (AC) oficialmente não possui região metropolitana⁶, sendo assim, no escopo deste trabalho, para não haver a desconsideração de um estado, foi decidido que o município de Rio Branco, a capital, fosse considerado como região metropolitana do Acre. Essa linha foi incluída no *dataset*.

4 Estudo de Caso

A implementação deste trabalho ocorreu através do Google Colaboratoty, ou simplesmente, Colab⁷. Os *datasets* foram baixados dos repositórios conforme indicado na seção 3 e em seguida colocados no Google Drive, de onde foram acessados pelo *notebook* do Colab.

Algumas bibliotecas foram usadas no projeto, como por exemplo a Pandas, que é uma biblioteca de software escrita para a linguagem de programação Python com fins de manipulação e análise de dados. Em particular, ela oferece estruturas de dados e operações para manipular tabelas numéricas e séries temporais⁸. Ela é a responsável por importar os arquivos, uma vez que ambos *datasets* estão no formato CSV. Outra biblioteca usada é a NumPy, útil para realização de cálculos numéricos em *arrays* multidimensionais. E, por último, vale ressaltar a utilização da biblioteca *matplotlib.pyplot* para a confecção dos gráficos analíticos.

Alguns *dataframes* foram criados para possibilitar as análises. Primeiramente, foi feito o *merge* entre os dados de vacinação e os dados de CNES através do *join* pelo campo “Código CNES”. A partir de então, foram gerados os *dataframes* das regiões norte e sudeste através do filtro de suas UFs correspondentes. Derivados dos dados do norte e sudeste, foram gerados os *dataframes* de suas respectivas regiões metropolitanas e interioranas. Para isso, tomou-se como premissa que o interior de um estado é toda a região não pertencente à região metropolitana.

⁶ <http://prvl.org.br/regioes-metropolitanas/regioes-metropolitanas/> Acesso em: 07/05/2021.

⁷ Serviço de nuvem gratuito hospedado pela Google para, dentre outros fins, escrever código Python no navegador sem nenhuma configuração local necessária e com acesso gratuito a GPUs (Unidades de Processamento Gráfico) <https://colab.research.google.com/>. Acesso em 16/05/2021.

⁸ [https://en.wikipedia.org/wiki/Pandas_\(software\)](https://en.wikipedia.org/wiki/Pandas_(software)). Acesso em 16/05/2021.

4.1 Análise e seleção dos dados

Os sete estados da região norte e os quatro da região sudeste, assim como suas respectivas siglas, estão listados a seguir nas Tabelas 1 e 2.

Tabela 1. Estados da região norte do Brasil.

| | | |
|----------------|---------------|--------------|
| Amazonas (AM) | Acre (AC) | Amapá (AP) |
| Pará (PA) | Rondônia (RO) | Roraima (RR) |
| Tocantins (TO) | | |

Tabela 2. Estados da região sudeste do Brasil.

| | |
|---------------------|-------------------|
| Espírito Santo (ES) | Minas Gerais (MG) |
| Rio de Janeiro (RJ) | São Paulo (SP) |

Como os *datasets* de vacinação e de CNES possuem informações sobre todos os estados brasileiros, com o intuito de reduzir a quantidade de dados a serem analisados e trazer eventuais ganhos de performance, foi feita uma seleção inicial nos *dataframes* para utilizar apenas as informações dos estados das regiões norte e sudeste, escopo deste trabalho. Ou seja, no *dataset* de vacinação, apenas foram considerados os registros cujo campo “estabelecimento_uf” contivessem um dos seguintes valores: 'RJ', 'SP', 'MG', 'ES', 'AM', 'AC', 'AP', 'PA', 'RO', 'RR', 'TO'.

Os campos que não seriam utilizados para atingir os objetivos desta pesquisa foram eliminados dos *datasets*, também visando ganhos de performance. Por exemplo: No *dataset* dos vacinados, optou-se por excluir os seguintes campos: “CEP do Paciente”, “Lote da Vacina”, “Código da Vacina”, “Razão Social do Estabelecimento”, “Nome Fantasia do Estabelecimento”, “Referência do Fabricante da Vacina” e “Sistema Origem”, conforme destacado na Figura 4 a seguir.

| Ordem | Descrição | Categoria | 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_colbgeMunicipio | 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_racaoSocial | Nome/Raça 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_grupoAtendimento_codigo | Código do grupo de atendimento ao qual pertence o vacinado | |
| 22 | vacina_grupoAtendimento_nome | Nome do grupo de atendimento ao qual pertence o vacinado | |
| 23 | vacina_categoria_codigo | 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 | |

Fig. 4. Colunas (em vermelho) excluídas do *dataset* de Vacinação.

No *dataset* de CNES, os campos eliminados foram: “Nome do Logradouro”, “Código IBGE”, “Número do Endereço”, “Bairro” e “CEP”, conforme destacado na Figura 5.

| Ordem | Campo | Descrição |
|-------|-----------------|---|
| 1 | co_cnes | Código CNES |
| 2 | co_ibge | Código IBGE |
| 3 | no_fantasia | Nome Fantasia |
| 4 | ds_tipo_unidade | 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 Diagnose 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 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 |

Fig. 5. Colunas (em vermelho) excluídas do dataset do CNES.

No contexto deste trabalho, apenas são relevantes os registros de vacinação da primeira dose. Todos os registros de vacinação da segunda dose foram desconsiderados. O campo usado nesse filtro é o “Descrição Dose Vacina”.

Após a eliminação dos registros de vacinação que não serão analisados neste estudo, o dataset de vacinação ficou com 3.529.642 registros.

4.2 Tratamento dos dados / data cleaning

A etapa de tratamento e limpeza de dados tem um papel importante no contexto da Ciência de Dados. Ela tem por objetivo, por exemplo, a identificação de valores nulos e/ou faltantes e a definição de estratégia de tratamento (atribuição de valor ou eliminação da linha) a serem utilizadas nos *datasets*. Neste trabalho, para o campo sexo com valor diferente de ([M|F]), o mesmo será descartado por não ser possível saber o seu significado. No caso de raça, caso o valor presente seja diferente de “1-branca”, “2-preta”, “3-parda”, “4-amarela”, “99-sem informação”, foi definido o valor para o campo como sendo “99-sem informação” (valor sentinela). Quanto à identificação nos *datasets* de vacinação e de estabelecimentos de saúde da existência de registros duplicados, foi estipulado a eliminação dos que tiverem as maiores data/hora, ou seja, foi mantido o registro de menor data/hora no *dataset*.

Após a eliminação dos registros duplicados, o dataset de vacinação ficou com 3.497.840 registros.

4.3 Proposta de modelo de extração de conhecimento e visualização dos dados

Nesta etapa, foi utilizada a biblioteca Pandas que permite que seja feito o merge entre os *datasets*, para se possa fazer uma ligação entre as informações contidas neles e extrair o conhecimento necessário para gerar as visualizações dos dados. Para isso também foi utilizada a biblioteca Matplotlib para que fosse possível construir os gráficos propostos por meio de seus principais comandos (*plot*, *show*, *grid*, etc.).

4.4 Coleta de metadados de proveniência dos experimentos

Nesta etapa, foram coletadas as informações dos *datasets* que indicam a origem dos dados, ou seja, de onde foram baixados, de que forma foram tratados e as transformações a que foram submetidos para análises futuras e posterior reuso.

Foi usada a biblioteca PROV, uma implementação em Python do Modelo de Dados de Proveniência da W3C⁹. Ela suporta importação/exportação dos formatos PROV-O (RDF), PROV-XML e PROV-JSON.

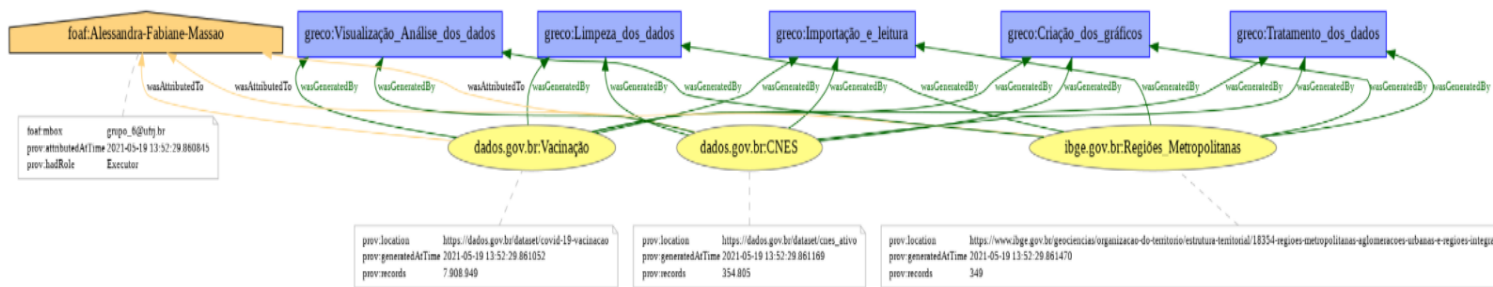


Fig. 6. Exemplo de grafo de proveniência gerado.

Como exemplificado na Figura 6, foram coletadas as seguintes informações de proveniência:

- ☐ O agente executor da atividade;
- ☐ O nome da atividade executada;
- ☐ O *dataset* consultado/consumido/analísado;
- ☐ Informações adicionais do agente executor:
 - ☐ E-mail;
 - ☐ Data e hora (*timestamp*) da atribuição do perfil ao agente;
 - ☐ Perfil do agente;

⁹ World Wide Web Consortium (W3C) é uma comunidade internacional que desenvolve padrões abertos com o intuito de garantir o crescimento da Web a longo prazo. Link: <https://www.w3.org/>. Acesso em: 16/05/2021.

- ❑ Data e hora (*timestamp*) da geração dos dados de proveniência.

4.5 Projeto de reprodutibilidade do experimento

Um experimento reprodutível é aquele em que se é possível recriar o mesmo resultado usando código e dados originais, mesmo quando sendo feito por equipes, sistemas de medição e localidades diferentes [4]. Há de se atentar para não se confundir com o conceito de replicabilidade, que é a capacidade de recriar o mesmo resultado, usando, porém, novos dados, a partir do mesmo plano de experimento. Pode ocorrer do programa não ser reprodutível devido a diferentes razões, como: versão antiga do compilador, métodos ou funções descontinuados pela linguagem, erro de programação, etc.

Com o objetivo de se reproduzir o mesmo resultado deste trabalho, é necessário que se use o mesmo código, ambiente computacional e *datasets* usados no experimento. Como existe a possibilidade dos *datasets* serem modificados/atualizados em seus repositórios originários de armazenamento, foram disponibilizados publicamente para *download*, as versões dos *datasets* usadas na experimentação¹⁰, assim como o código fonte do programa¹¹ em linguagem Python.

Especificamente em relação a experimentos computacionais, garantir a reprodutibilidade por um longo período é um desafio, dada a evolução da tecnologia e a eventual descontinuidade de recursos utilizados. Nesse caso, busca-se descrever o ambiente computacional com o máximo de detalhe para que seja facilmente identificável qualquer ajuste necessário quando não for possível utilizar exatamente o mesmo ambiente computacional.

Dessa forma, as seguintes medidas foram tomadas no próprio código-fonte do experimento visando a reprodutibilidade do mesmo:

1. Identificação em detalhes do ambiente computacional;
2. Inclusão de comentários descrevendo as ações realizadas;
3. Utilização de funções ou recursos da versão mais atual do Python; e
4. Auto-verificação do código.

5 Discussão

Havia sido planejado que para o campo raça/cor do paciente, caso o código estivesse diferente dos valores contidos no dicionário de dados, ou seja, diferente de “1”, “2”, “3”, “4” e “99”, que o valor seria alterado sumariamente para “99”, que significa “não especificado”. Porém, na implementação dessa limpeza, verificou-se que existiam registros com o código “5”, cujo nome da raça/cor era “indígena”. Diante desse dilema, decidimos não excluir essas linhas e considerar que houve uma falha na criação do dicionário de dados do *dataset*. Além do “5”, não houve nenhuma ocorrência de valor diferente dos listados.

¹⁰

<https://drive.google.com/drive/folders/16LuaTdN1iDFMuEOTJhDhiGNgVglDhi5r?usp=sharing>. Acesso em: 17/05/2021.

¹¹ <https://github.com/massaoiwanaga/topespsi>. Acesso em: 17/05/2021.

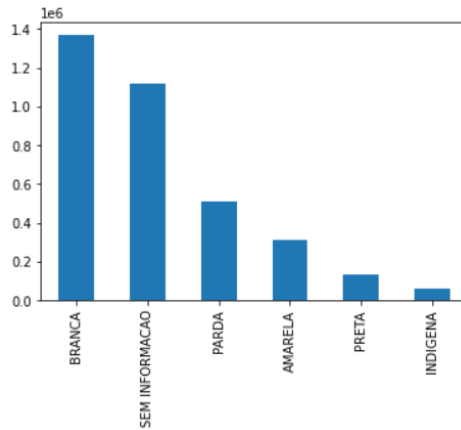


Fig. 7. Raça/cor dos vacinados.

Encontramos no *dataset* diversos casos de pacientes com mais de 2 registros de vacinação para a primeira dose da vacina, situação que não deveria acontecer. A decisão tomada com relação a esses casos foi de manter um único registro para cada paciente (*paciente_id*) com a menor data de vacinação, ou seja, o mais antigo foi mantido e o restante excluído.

| | document_id | paciente_id | paciente_idade | paciente_dataNascimento | paciente_enumSexoBiologico | pa |
|---------|---|---|----------------|-------------------------|----------------------------|----|
| 4828100 | 56834bab-40c1-466b-b309-d06076eb4b3b-i0b0 | 0c3390b4d0b9a535cb22799c2618053106a61def30b85... | 71.0 | 1949-06-14 | M | 99 |
| 3670249 | 4e09a541-fa26-4daa-8983-2ab84668a1e2-i0b0 | 0c3390b4d0b9a535cb22799c2618053106a61def30b85... | 71.0 | 1949-06-14 | M | 99 |
| 356327 | ae5f43ff-7839-4851-b3f2-2183cb62d729-i0b0 | 0c3390b4d0b9a535cb22799c2618053106a61def30b85... | 71.0 | 1949-06-14 | M | 99 |
| 4773242 | 7f53f141-1d9c-4bcd-8d24-504b67d4ea50-i0b0 | 0fc334df6e1b34ea25366709ab808fd731e4f9f1c02347... | 33.0 | 1987-08-15 | M | 99 |
| 1842150 | 11e92ee3-af7c-46c1-9096-e25235864e32-i0b0 | 0fc334df6e1b34ea25366709ab808fd731e4f9f1c02347... | 33.0 | 1987-08-15 | M | 99 |
| 6830377 | 7d66212c-9a74-4ebe-ba99-4be3e2a79a43-i0b0 | 0fc334df6e1b34ea25366709ab808fd731e4f9f1c02347... | 33.0 | 1987-08-15 | M | 99 |

Fig. 8. Registros com mais de um *paciente_id*.

Analisando a idade dos pacientes que tomaram a vacina, verificamos que a idade média é de 57 anos, a idade máxima é de 221 anos e a mínima é de 0 anos de vida. Notamos também que 1.376 vacinados têm menos de 18 anos de idade, o que

contraria a recomendação de idade para a vacina. Além disso, 1.986 vacinados possuem idade maior que 110 anos.

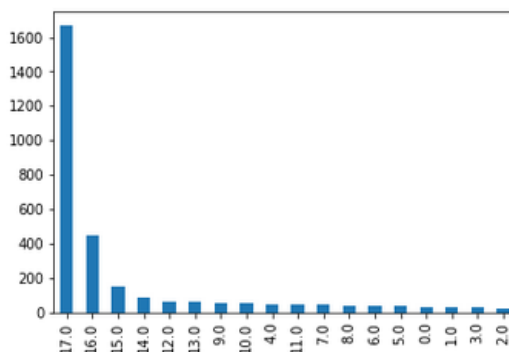


Fig. 9. Números de vacinados com menos de 18 anos de idade.

Para os casos de idade de vacinação abaixo de 18 anos ou acima de 110, a idade foi calculada a partir da data de nascimento, mas foram obtidos os mesmos valores dos originalmente informados no dataset, o que não permitiu qualquer ajuste em eventual erro de cadastro. Assim, esses valores de idade serão desconsiderados nas análises relacionadas à idade.

Verificou-se também que o sexo dos pacientes apresentou o valor “I”, além dos “F” de feminino e “M” de masculino. Como representavam menos de 2% do total de registros, ficou decidido mantê-los no *dataset* mas retirá-los da visualização por serem irrelevantes no resultado final.

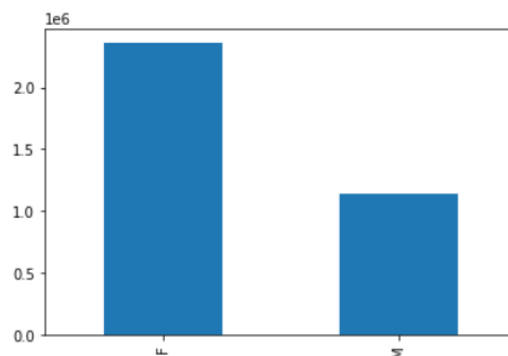


Fig. 10. Sexo biológico dos vacinados.

5.1 Perfil dos vacinados (sexo, idade, raça, grupo) e quantitativos nas duas regiões, dividido entre regiões metropolitanas e regiões interioranas

Quantitativo de vacinados nas regiões metropolitanas da região sudeste e norte por UF:

Foi possível verificar através das análises abaixo que a maior quantidade de vacinados na região sudeste está localizada nas três regiões metropolitanas mais populosas do Brasil. No caso da região norte, a maior parte dos vacinados também encontra-se nas duas maiores regiões metropolitanas mais populosas, as de Manaus e Belém.

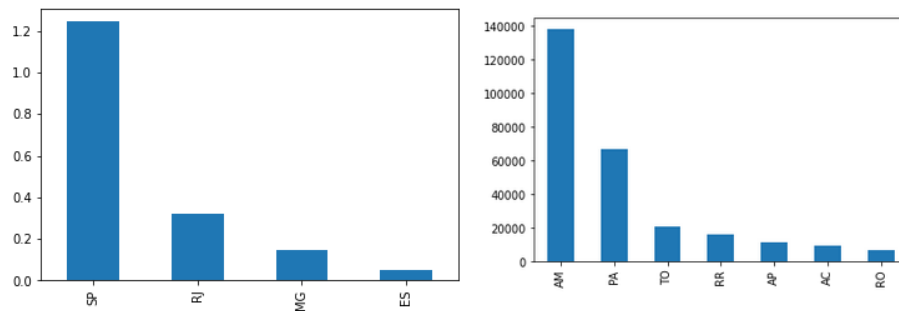


Fig. 11. Quantidade de vacinados por região metropolitana (sudeste e norte).

Quantitativo de vacinados nas regiões metropolitanas da região sudeste e norte por categoria:

Nos gráficos abaixo foi verificado que as duas categorias com maior quantidade de vacinados nas duas regiões metropolitanas foram: “Trabalhadores da Saúde” e “Faixa Etária”. Em terceiro lugar, na região metropolitana do sudeste (gráfico à esquerda) aparece a categoria “Pessoa de 60 anos ou mais institucionalizadas”, e na região metropolitana do norte (gráfico à direita) aparecem “Povos Indígenas”.

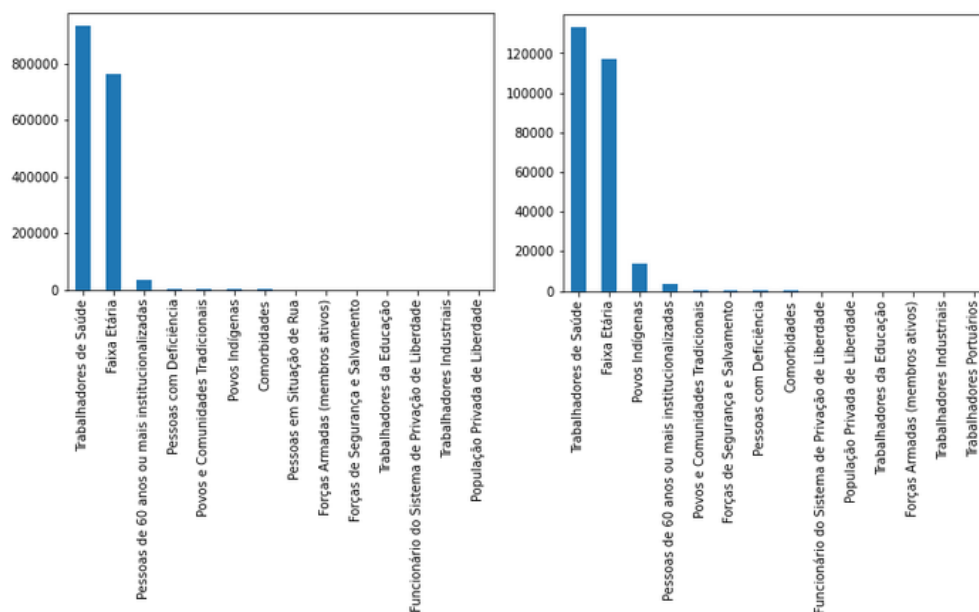


Fig. 12. Quantidade de vacinados das regiões metropolitanas (sudeste à esquerda e norte à direita) por categoria.

Quantitativo de vacinados nas regiões metropolitanas da região sudeste e norte por sexo biológico:

Com os gráficos abaixo, nota-se que a maior parte dos vacinados nas duas regiões metropolitanas são do sexo Feminino e em segundo lugar os do sexo Masculino.

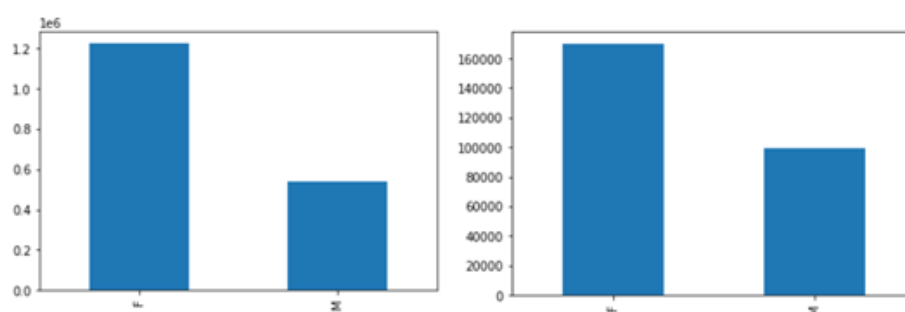


Fig. 13. Quantidade de vacinados das regiões metropolitanas (sudeste à esquerda e norte à direita) por sexo biológico.

Quantitativo de vacinados nas regiões metropolitanas da região sudeste e norte por raça:

Foi possível analisar com os gráficos abaixo que na região metropolitana do sudeste (gráfico à esquerda), a maior quantidade de vacinados está “Sem informação” para a raça e em ordem decrescente aparecem a branca, a parda, a amarela, a preta e a indígena. Para a região metropolitana do norte (gráfico à direita), a maior parte dos vacinados é da raça parda, seguidos dos “Sem informação”, e da amarela, branca, indígena e preta.

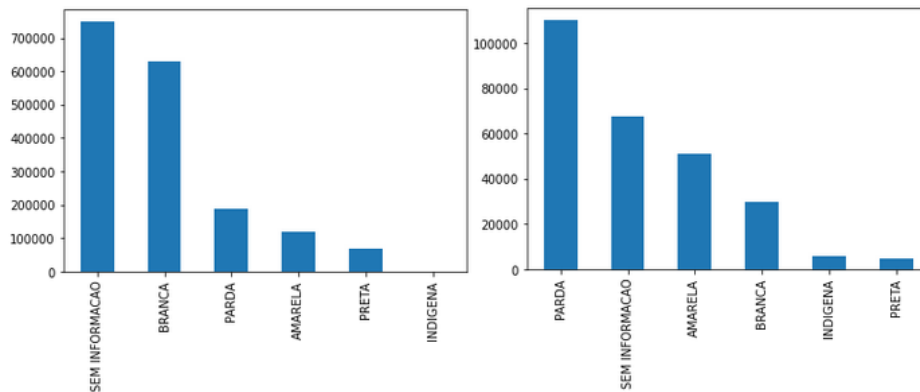


Fig. 14. Quantidade de vacinados das regiões metropolitanas (sudeste à esquerda e norte à direita) por raça.

Quantitativo de vacinados nas regiões metropolitanas da região sudeste e norte por nome de vacina:

Através dos gráficos abaixo, nota-se que nas duas regiões metropolitanas da região sudeste (gráfico à esquerda) e da região norte (gráfico à direita), a maior parte dos vacinados tomaram a vacina Coronavac e em segundo lugar, a Covishield (AstraZeneca).

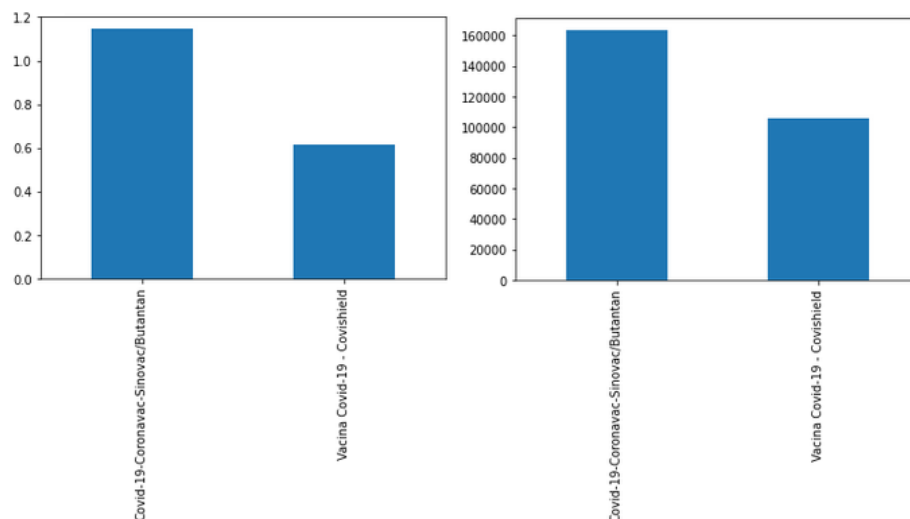


Fig. 15. Quantidade de vacinados das regiões metropolitanas (sudeste à esquerda e norte à direita) por nome de vacina.

Quantitativo de vacinados nas regiões metropolitanas da região sudeste e norte por idade:

Analisando os gráficos abaixo, na região metropolitana do sudeste (gráfico à direita), nota-se que a maior quantidade de vacinados pertence a faixa etária de 80 a 90 anos, em segundo lugar aparece a de 30 a 40 anos e em terceiro lugar, a de 40 a 50 anos. Na região metropolitana do norte (gráfico à esquerda), a maior parte dos vacinados encontra-se na faixa de etária de 70 a 80 anos, em segundo lugar a de 80 a 90 anos e em terceiro lugar a de 30 a 40 anos. Percebe-se que a expectativa de vida da região metropolitana do sudeste é maior do que a do norte.

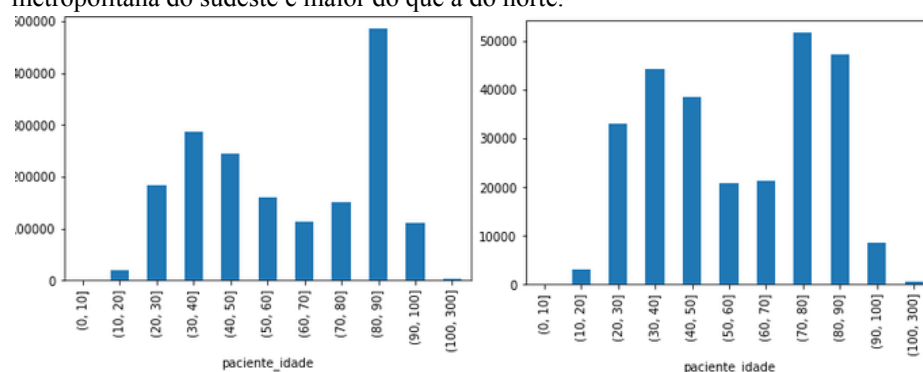


Fig. 16. Quantidade de vacinados das regiões metropolitanas (sudeste à esquerda e norte à direita) por faixa etária..

Quantitativo de vacinados nas regiões interioranas da região sudeste e norte por UF:

Verifica-se através dos gráficos abaixo, que não ocorreram muitas mudanças nas regiões interioranas do sudeste (gráfico à esquerda) e do norte (gráfico à direita) em relação à quantidade de vacinados das regiões metropolitanas sudeste e norte. O terceiro lugar ocupado pelo estado de Rondônia na região interiorana do norte mudou em relação ao estado do Tocantins na região metropolitana do norte, assim como as outras posições da região interiorana do norte (TO, AC, AP e RR) em relação à metropolitana do norte (RR, AP, AC e RO).

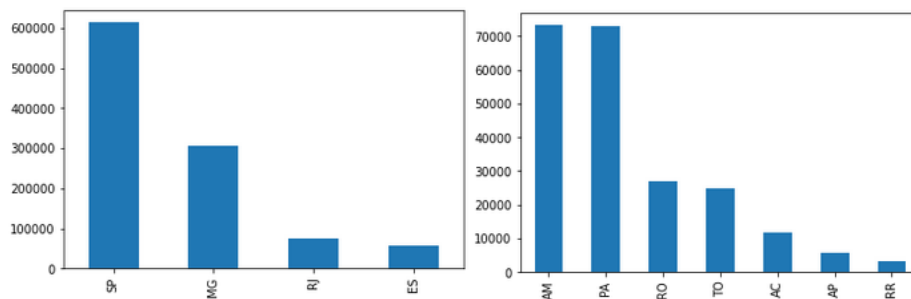


Fig. 17. Quantidade de vacinados das regiões interioranas (sudeste à esquerda e norte à direita) por UF.

Quantitativo de vacinados nas regiões interioranas da região sudeste e norte por categoria:

Os gráficos abaixo mostram que as regiões interioranas das regiões sudeste (gráfico à direita) e norte (gráfico à esquerda) não apresentaram mudanças em relação à categoria das regiões metropolitanas do sudeste e do norte.

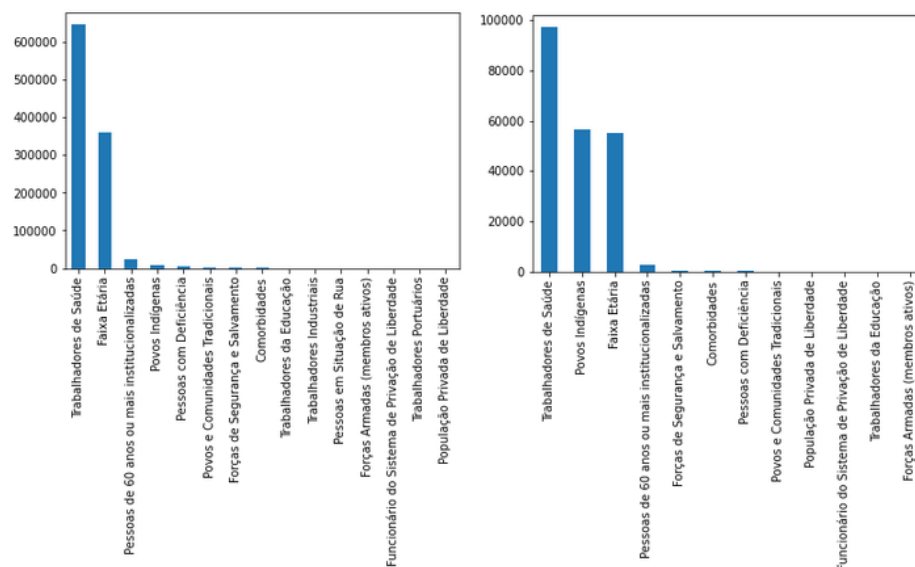


Fig. 18. Quantidade de vacinados das regiões interioranas (sudeste à esquerda e norte à direita) por categoria.

Quantitativo de vacinados nas regiões interioranas da região sudeste e norte por sexo biológico:

Os gráficos abaixo mostram que as regiões interioranas das regiões sudeste (gráfico à direita) e norte (gráfico à esquerda) não apresentaram mudanças em relação ao sexo biológico das regiões metropolitanas do sudeste e do norte.

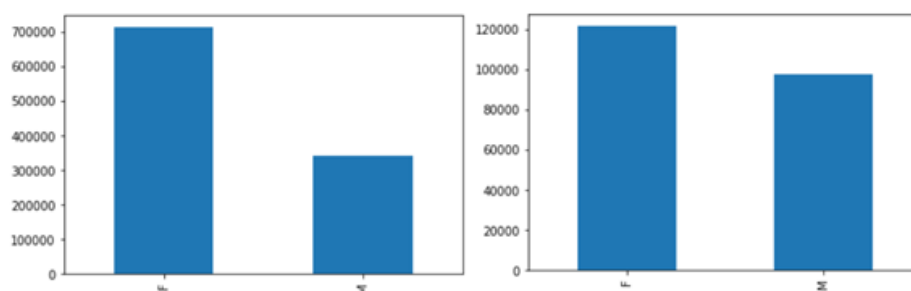


Fig. 19. Quantidade de vacinados das regiões interioranas (sudeste à esquerda e norte à direita) por sexo biológico.

Quantitativo de vacinados nas regiões interioranas da região sudeste e norte por raça:

Foi possível analisar com os gráficos abaixo que na região interiorana do sudeste (gráfico à esquerda), a maior quantidade de vacinados são da raça branca e em segundo lugar aparecem os “Sem informação”. Para o restante das outras raças, não houve mudança em relação à região metropolitana em ordem decrescente (parda, amarela, preta e indígena). Para a região interiorana do norte (gráfico à direita), a maior parte dos vacinados é da raça parda igual ao da região metropolitana do norte, seguidos dos da amarela, indígena, “Sem informação”, branca e preta.

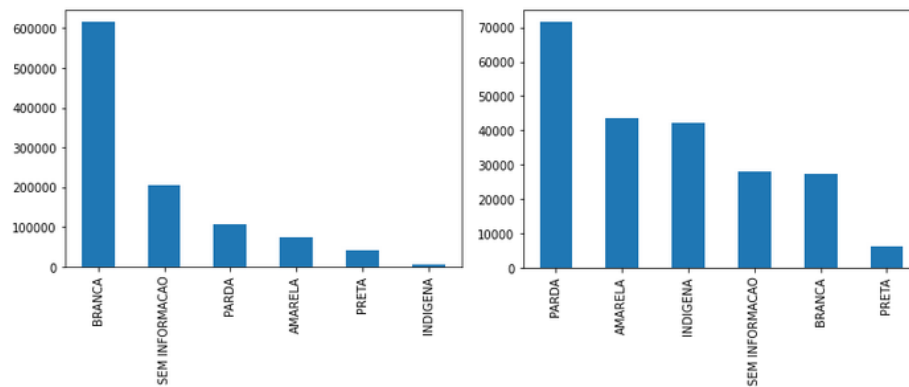


Fig. 20. Quantidade de vacinados das regiões interioranas (sudeste à esquerda e norte à direita) por raça.

Quantitativo de vacinados nas regiões interioranas da região sudeste e norte por nome de vacina:

Através dos gráficos abaixo, nota-se que nas duas regiões interioranas da região sudeste (gráfico à esquerda) e da região norte (gráfico à direita), a maior parte dos vacinados tomaram a vacina Coronavac e em segundo lugar, a Covishield (AstraZeneca) assim como nas suas respectivas regiões metropolitanas. A diferença apresentada foi que em terceiro lugar aparece a vacina Biontech Pfizer que não aparece nas duas regiões metropolitanas.

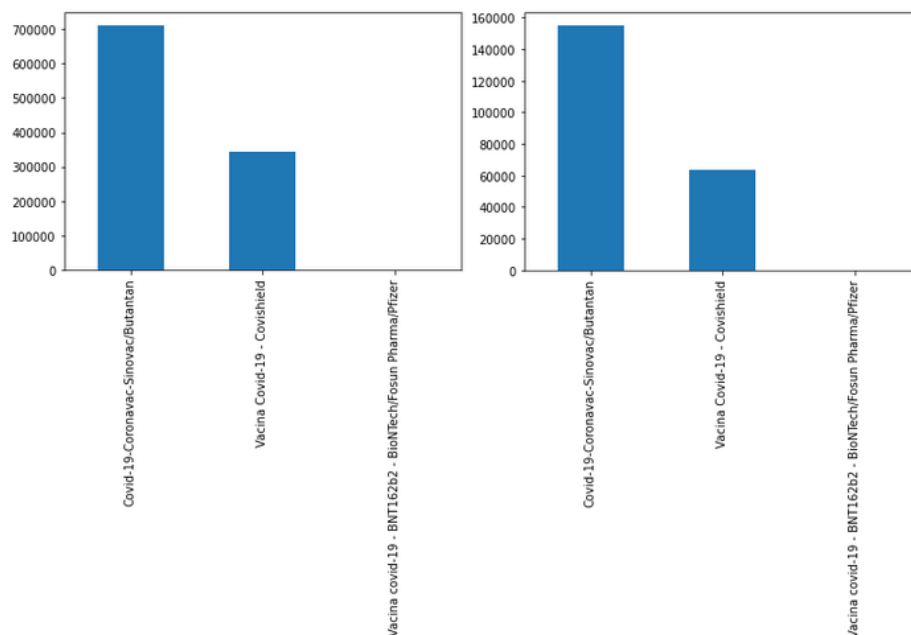


Fig. 21. Quantidade de vacinados das regiões interioranas (sudeste à esquerda e norte à direita) por nome de vacina.

Quantitativo de vacinados nas regiões interioranas da região sudeste e norte por idade:

Analisando os gráficos abaixo, na região interiorana do sudeste (gráfico à direita), nota-se que o gráfico não se modificou em relação à região metropolitana do sudeste. Na região interiorana do norte (gráfico à direita), verifica-se que houve uma mudança em relação à região metropolitana do norte, pois a maior parte dos vacinados encontra-se em faixa etária mais jovem, em primeiro lugar a de 30 a 40 anos, em segundo lugar a de 20 a 30 anos e em terceiro lugar a de 40 a 50 anos.

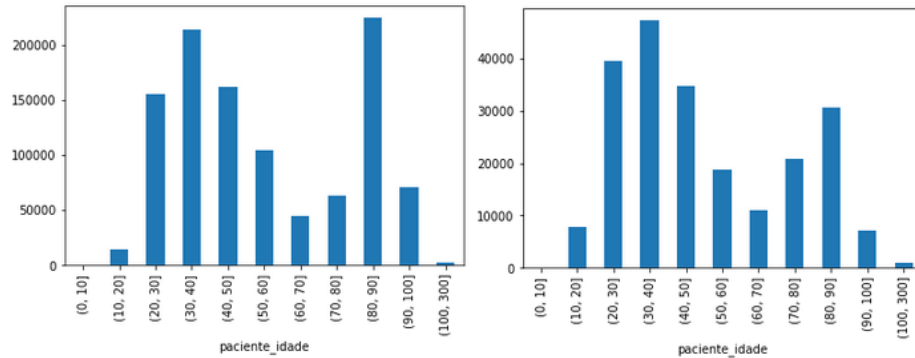


Fig. 22. Quantidade de vacinados das regiões interioranas (sudeste à esquerda e norte à direita) por faixa etária.

5.2 Quantitativos de vacinados por tipo de gestão e por tipo de unidade de estabelecimento de saúde nas duas regiões metropolitanas e interioranas

Foi observado que no dataset CNES ativos não foram encontrados registros duplicados e os nulos no campo no_fantasia não foram tratados pois o mesmo é irrelevante para a análise e será descartado.

Quantitativo de vacinados nas regiões metropolitanas da região sudeste e norte por tipo de gestão de estabelecimento de saúde (M - Municipal, E - Estadual, D - Dupla):

Verifica-se nos gráficos abaixo que a maior parte dos vacinados se vacinaram em estabelecimentos de saúde que possuem gestão municipal nas duas regiões metropolitanas, seguidos pela estadual e dupla.

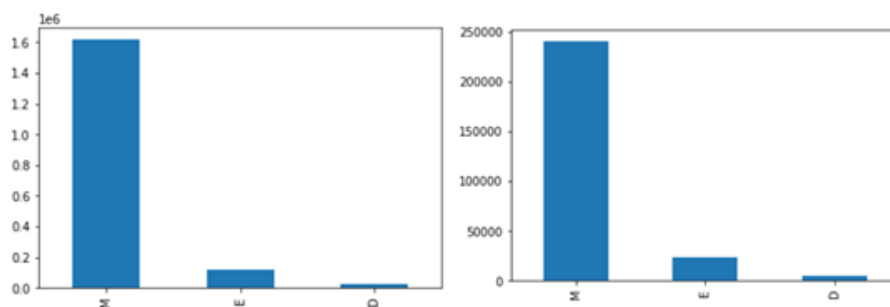


Fig. 23. Quantidade de vacinados das regiões metropolitanas (sudeste à esquerda e norte à direita) por tipo de gestão.

Quantitativo de vacinados nas regiões metropolitanas da região sudeste e norte por tipo de unidade:

É possível analisar através dos gráficos abaixo, que nas regiões metropolitanas do sudeste e do norte, a maior quantidade de vacinados tomou a vacina em “Centros de Saúde/Unidades Básicas”. Na região metropolitana do norte, em quarto lugar aparece a “Unidade de Atenção à Saúde Indígena”.

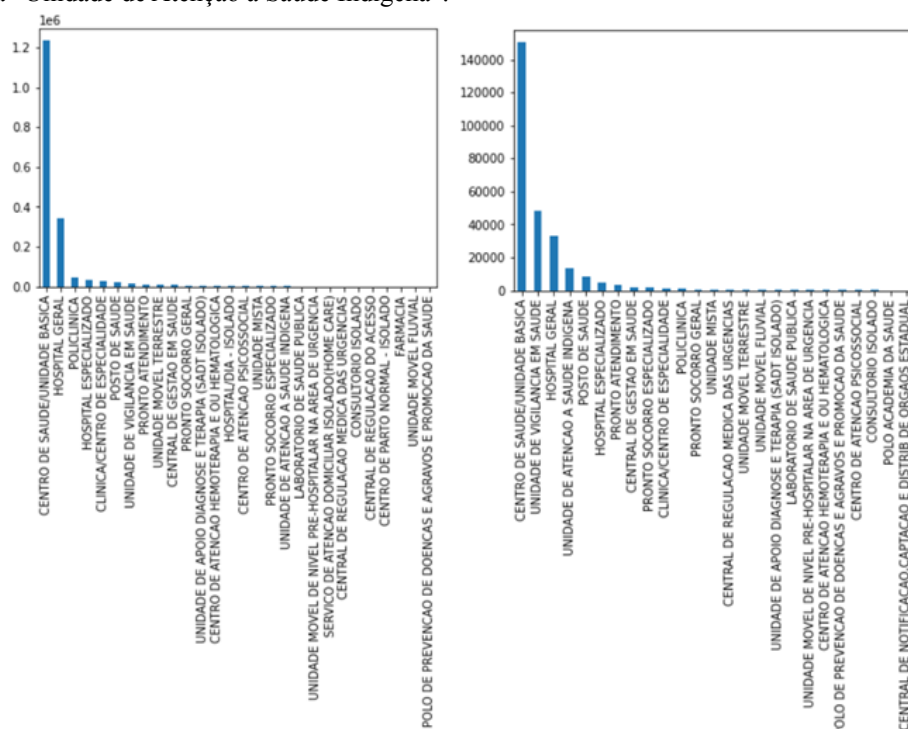


Fig. 24. Quantidade de vacinados das regiões metropolitanas (sudeste à esquerda e norte à direita) por tipo de unidade.

Quantitativo de vacinados nas regiões interioranas da região sudeste e norte por tipo de gestão de estabelecimento de saúde (M - Municipal, E - Estadual, D - Dupla):

Observando os gráficos abaixo, nota-se que a maior parte dos vacinados das duas regiões interioranas se vacinaram em estabelecimentos de saúde de gestão municipal, assim como os das regiões metropolitanas. Em segundo lugar, observa-se uma mudança em relação às regiões metropolitanas pois os estabelecimentos de saúde

nestes gráficos são de gestão dupla e não estadual como aparece nas regiões metropolitanas.

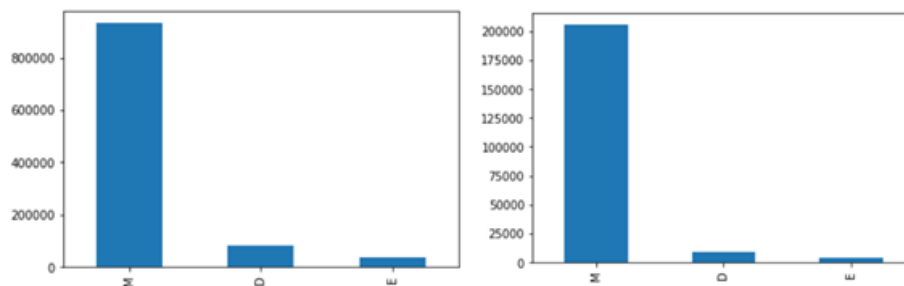


Fig. 25. Quantidade de vacinados das regiões interioranas (sudeste à esquerda e norte à direita) por tipo de gestão.

Quantitativo de vacinados nas regiões interioranas da região sudeste e norte por tipo de unidade:

É possível analisar através dos gráficos abaixo, que nas regiões interioranas do sudeste e do norte, a maior quantidade de vacinados tomou a vacina em “Centros de Saúde/Unidades Básicas” assim como nas regiões metropolitanas. Na região interiorana do norte, em segundo lugar aparece a “Unidade de Atenção à Saúde Indígena” ao invés do quarto lugar na região metropolitana. Percebe-se uma maior quantidade de indígenas na região interiorana do norte.

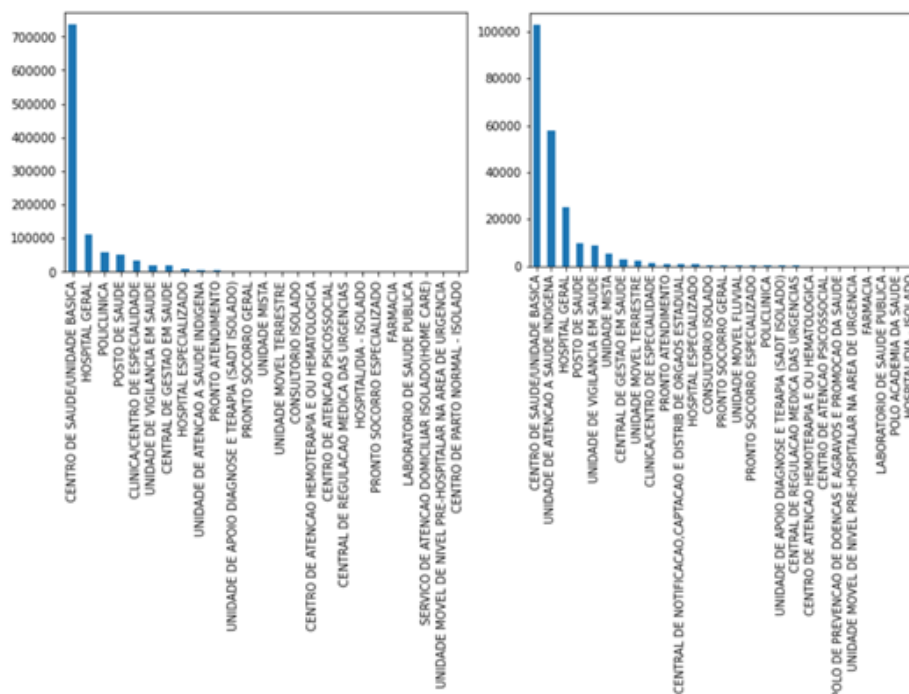


Fig. 26. Quantidade de vacinados das regiões interioranas (sudeste à esquerda e norte à direita) por tipo de unidade.

5.3 Quantitativos e percentual de vacinados que tomaram a vacina em município diferente daquele de sua residência

Com quantidade limitada de vacinas a serem aplicadas, a estratégia adotada pelo governo federal foi a distribuição de doses de vacinas aos estados, que por sua vez foram responsáveis pela distribuição entre os municípios.

Um dos critérios para a distribuição de doses entre os municípios é sua população estimada. Se, dentre os vacinados em um município, grande parte destes residem em outro município, isso pode levar a uma escassez de doses neste município. Sendo assim, a avaliação do município de residência dos vacinados torna-se relevante, pois pode contribuir para uma melhor distribuição de doses entre os municípios, frente à real demanda de cada um.

Do conjunto de dados trabalhados neste experimento, aproximadamente 0,5% dos vacinados não possuía dados informados sobre o município de residência (18.055) ou possuía dados inválidos (631). Não seria possível inferir essa informação com os dados disponíveis nesse estudo, dessa forma, a análise realizada, referente aos vacinados fora de seu município de residência, desconsiderou esses registros. A

avaliação foi realizada com base em 3.479.154 vacinados das regiões norte e sudeste que possuíam informações válidas sobre seu município de residência.

Inicialmente foi realizada a avaliação de algumas métricas sobre a vacinação de pessoas fora de seu município de residência, dentre as quais podemos destacar:

- 1) A quantidade média de vacinados fora de seu município de residência é aproximadamente 354 pessoas e a média do percentual de vacinados fora do município é de 19,10%. Ou seja, em média, 19,10% de vacinados não residem no município onde foram vacinados.
- 2) O município que percentualmente mais vacinou pessoas que residem em outro município, vacinou 59,23% de pessoas que residem em outro município.
- 3) No último quartil, ou 25%, dos municípios das regiões norte e sudeste vacinaram mais de 23,74% de pessoas que residem em outros municípios.
- 4) Em seguida, identificamos os 20 municípios com maior quantidade de vacinados fora do município em números absolutos e percentualmente em relação à quantidade total de vacinados. A relação de cada um é mostrada a seguir:

Em seguida, foram identificados os 20 municípios com maior quantidade de vacinados fora do município em números absolutos e percentualmente em relação à quantidade total de vacinados. A relação para cada um dos critérios é apresentada nas tabelas 3 e 4 abaixo.

Com relação aos municípios com maior quantidade de vacinados que residem fora do município de residência em números absolutos, pode-se observar que destacam-se as capitais do sudeste e grandes cidades das regiões metropolitanas. Os únicos municípios da região norte que aparecem são as duas maiores capitais, Manaus e Belém.

Já para os municípios com maior percentual de vacinados residentes fora do município em que se vacinaram, podemos notar que a relação muda completamente. Dentre os 20 municípios listados, temos apenas três da região norte, todos do estado do Pará.

Podemos observar também que 11 municípios vacinaram mais de 50% de pessoas residentes em outro município. Além disso, dentre os 20 destacados, podemos perceber percentuais acima de 47,8%.

Tabela 3. Municípios com maior quantidade de vacinados em números absolutos:

| Município | UF | Qtde Vacinados Fora do Município |
|-----------------------|----|----------------------------------|
| SÃO PAULO | SP | 111874 |
| RIO DE JANEIRO | RJ | 27942 |
| BELO HORIZONTE | MG | 25315 |
| SÃO BERNARDO DO CAMPO | SP | 14723 |
| CAMPINAS | SP | 14645 |
| MANAUS | AM | 13852 |
| SANTO ANDRÉ | SP | 13041 |
| SANTOS | SP | 12469 |
| VITÓRIA | ES | 10884 |
| GUARULHOS | SP | 9971 |
| SÃO JOSÉ DO RIO PRETO | SP | 8756 |
| BELÉM | PA | 8158 |
| SÃO JOSÉ DOS CAMPOS | SP | 8144 |
| SÃO CAETANO DO SUL | SP | 6939 |
| RIBEIRÃO PRETO | SP | 6800 |
| OSASCO | SP | 6786 |
| SERRA | ES | 6410 |
| BARUERI | SP | 6377 |
| TABOÃO DA SERRA | SP | 5295 |
| VILA VELHA | ES | 5042 |

Tabela 4. Municípios com maior percentual de vacinados em relação à quantidade total:

| Município | UF | % Vacinados de Fora do Município |
|-----------------------|----|----------------------------------|
| SENADOR JOSE PORFIRIO | PA | 59,27% |
| BARUERI | SP | 57,01% |
| MESQUITA | RJ | 54,91% |
| ANAPU | PA | 54,82% |
| NOVA LIMA | MG | 54,57% |
| CLARAVAL | MG | 54,10% |
| ITAPECERICA DA SERRA | SP | 52,75% |
| SANTANA DE PARNAÍBA | SP | 52,64% |
| VESPASIANO | MG | 52,00% |
| VOTORANTIM | SP | 51,66% |
| ITAGUAÍ | RJ | 50,11% |
| MACAÉ | RJ | 49,43% |
| PRAIA GRANDE | SP | 49,37% |
| VITÓRIA DO XINGU | PA | 49,26% |
| COTIA | SP | 49,25% |
| ITAQUAQUECETUBA | SP | 49,11% |
| FERRAZ DE VASCONCELOS | SP | 48,87% |
| ITATIAIA | RJ | 48,86% |
| ITAPEVI | SP | 47,97% |
| HOLAMBRA | SP | 47,91% |

Foram realizadas duas análises mais detalhadas em relação aos vacinados fora do município de residência: uma sobre quais destes se tratam de municípios de outros estados e outra sobre as categorias de vacinados.

Avaliação com relação a residentes em outros estados

Com relação à população vacinada total dos estados das regiões norte e nordeste que residem em outros municípios, quais destes se tratam de municípios localizados em outros estados? Esse tipo de análise poderia permitir identificar indícios de turismo de vacina ou maior incidência de casos em municípios de fronteira, por exemplo.

Residentes no mesmo estado? (N e SE)

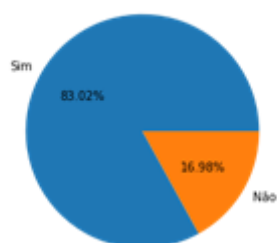


Fig. 27. Vacinados residentes fora do município de vacinação (todos os municípios N e SE)

A mesma análise foi realizada para os cinco municípios com maior percentual de vacinados residentes fora do município em que se vacinaram:

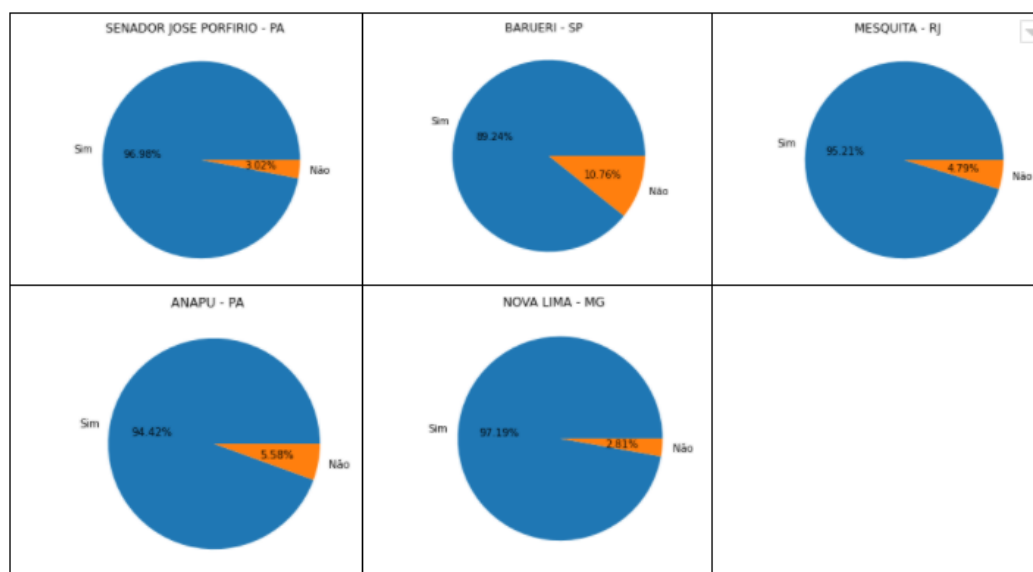


Fig. 28. Vacinados residentes fora do município de vacinação (cinco municípios com maior percentual de vacinados residentes fora do município)

Podemos observar que, apesar do número para o total de municípios das regiões norte e sudeste indicarem que 17% dos vacinados que residem em outros municípios, residem inclusive em municípios de outros estados, esse percentual se mostrou mais reduzido para os cinco municípios analisados. Dos cinco, o único município que

mostrou um percentual um pouco maior foi Barueri, que indicou quase 11% dos vacinados como residentes em outro estado.

Avaliação com relação às categorias de vacinados

Outra análise realizada sobre os vacinados que não residiam no município em que foram vacinados foi em relação à distribuição de categorias de vacinados.

A distribuição de categorias, para todos os municípios das regiões norte e nordeste, conforme pode ser visto no gráfico abaixo, foram: “trabalhadores de saúde”, “Faixa Etária”, “Povos Indígenas” e “Pessoas de 60 anos ou mais institucionalizadas”.

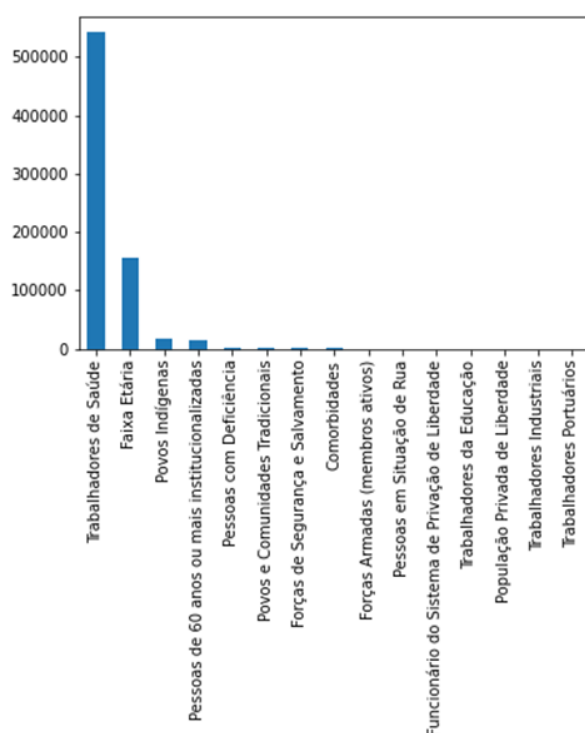


Fig. 29. Categorias de vacinados fora do município de residência dos municípios das regiões norte e sudeste.

Foi analisada então quais seriam as principais categorias de vacinados para os cinco municípios com maior percentual de vacinados fora do município em relação à população total.

Observa-se que algumas categorias, para estes cinco municípios, tinham uma quantidade de vacinados muito baixa (menor do que 15), o que aumentava a quantidade de informações no gráfico sem contribuir de forma relevante na análise.

Desta forma, essas categorias foram retiradas da visualização, resultando no gráfico a seguir:

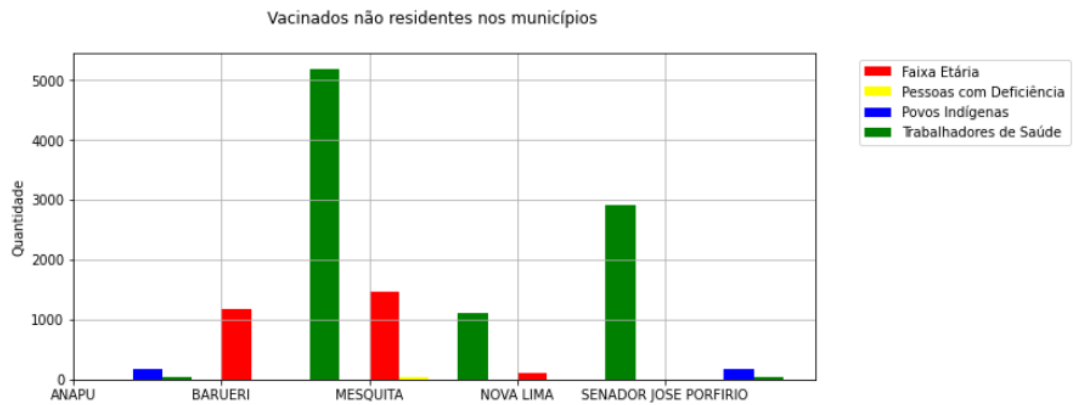


Fig. 30. Cinco municípios com maior percentual de vacinados residentes fora do município por categoria.

O gráfico mostra que nos municípios de Barueri e Nova Lima, as principais categorias de vacinados são “Trabalhadores de saúde” e “Faixa Etária”, as mesmas que os municípios das regiões norte e nordeste de forma geral.

No entanto, para o município de Mesquita, podemos observar um padrão diferente: os vacinados neste município e que residem em outro município concentram-se em maior quantidade na categoria de “Faixa Etária”, seguida pela categoria “Trabalhadores de saúde”.

Os dois municípios do estado do Pará, Anapu e Senador José Porfírio, possuem como maior quantidade de vacinados que não residem no município a categoria “Povos Indígenas”. Isso explica por que esses dois municípios do Pará ficaram entre os municípios de maior percentual de vacinados fora do município.

6 Conclusão

Com este trabalho, foi possível verificar as diferenças no perfil dos vacinados nas duas regiões, norte e sudeste, analisando também suas regiões metropolitanas e interioranas, assim como os estabelecimentos de saúde utilizados pelos vacinados nessas regiões.

Verificou-se que os resultados dos gráficos da região sudeste são mais uniformes dos que os da região norte, que apresentam mais diferenças. Como a região sudeste apresenta índices sociais mais elevados, por possuir a segunda maior qualidade de vida do país, observado pelo seu IDH (Índice de Desenvolvimento Humano) [9], deduz-se que essa seja uma das explicações pela uniformidade dos seus resultados.

Nas duas regiões a maioria dos vacinados por sexo biológico é feminina. Sobre a vacina, a maioria foi vacinada com a Coronavac, seguida da Covishield (Oxford) e no

interior houve registro de doses da BioNtech/Pfizer. Com relação aos estabelecimentos de saúde utilizados pelos vacinados, a maioria deles é de gestão municipal, seguida de estadual e dupla (municipal e estadual) nas regiões metropolitanas e no interior, a maioria continua sendo a municipal, seguida da dupla e estadual. O tipo de estabelecimentos de saúde com mais vacinados foi o “Centro de Saúde/Unidade Básica”.

Com relação aos vacinados fora de seu município de residência, neste trabalho foram realizadas algumas análises, com foco nos municípios das regiões norte e sudeste. Avaliamos um conjunto mais detalhado de cinco municípios, mas ainda num contexto mais amplo dessas duas regiões, mas que permitiu identificar algumas características interessantes sobre a população vacinada.

Como trabalhos futuros, em primeiro lugar, como continuidade desta pesquisa, seria importante repetir as análises realizadas com dados mais avançados da vacinação no Brasil, com mais meses ou anos do seu início, para que se tenha um espectro maior de massa de dados. Com certeza, para este fim, será necessária uma infraestrutura mais robusta de ambiente de desenvolvimento que permita a importação de *datasets* com tamanhos grandes (maiores que 10 GB). Pode-se também expandir o escopo das análises para os imunizados, ou seja, os que já tomaram a segunda dose da vacina, inclusive. Nesse caso caberá a decisão de como lidar com casos de pacientes que tomaram as duas doses de vacinas diferentes. Esse é um tema também interessante de ser aprofundado, pois foram encontrados casos assim. Quanto ao recorte, faz-se necessário também analisar as demais regiões do Brasil, como nordeste, sul e centro-oeste ou municípios específicos.

Ainda como trabalhos futuros é possível realizar avaliações sobre os vacinados fora de seu município de residência com foco em determinado estado ou pequenos grupos de municípios com características específicas, aliadas às análises de outras informações de vacinação, que podem permitir identificação de padrões ou características da população vacinada, fornecendo informações relevantes para contribuir na calibração de doses a serem distribuídas aos municípios, de acordo com os grupos prioritários e o plano nacional de imunização.

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> Acesso em: 05 de março de 2021.
3. Regiões Metropolitanas, Aglomerações Urbanas e Regiões Integradas de Desenvolvimento. Disponível em: <<https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial/18354-regioes-metropolitanas-aglomeracoes-urbanas-e-regioes-integradas-de-desenvolvimento.html?=&t=acesso-ao-produto>> Acesso em: 07 de maio de 2021.
4. Re-run, Repeat, Reproduce, Reuse, Replicate: Transforming Code into Scientific Contributions Fabien C. Y. Benureau and Nicolas P. Rougier.

5. RIBEIRO, Marcelo Gomes. Território e Desigualdades de Renda em Regiões Metropolitanas do Brasil. **Dados**, Rio de Janeiro, v. 58, n. 4, p. 913-950, Dez. 2015. Disponível em:
<http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0011-52582015000400913&lng=en&nrm=iso>. acesso em 07 Maio de 2021.
<http://dx.doi.org/10.1590/00115258201562>.
6. GARNELO, Luiza; SAMPAIO, Sully. Bases sócio-culturais do controle social em saúde indígena: problemas e questões na Região Norte do Brasil. Cad. Saúde Pública, Rio de Janeiro, v. 19, n. 1, p. 311-317, Fev. 2003. Disponível em
<http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2003000100035&lng=en&nrm=iso>. Acesso em 07 Maio 2021.
<http://dx.doi.org/10.1590/S0102-311X2003000100035>.
7. Donalisio, Maria Rita, Ruiz, Tânia, Cordeiro, Ricardo Fatores associados à vacinação contra influenza em idosos em município do Sudeste do Brasil. Revista de Saúde Pública [en linea]. 2006, 40(1), 115-119 [data de Consulta 7 de Mayo de 2021]. ISSN: 0034-8910. Disponível em: <https://www.redalyc.org/articulo.oa?id=67240151018>
8. LNCS Homepage, <http://www.springer.com/lncs>, last accessed 2016/11/21.
9. IPEA -Instituto de Pesquisa Econômica Aplicada. Evolução do IDHM e de Seus Índices Componentes no período de 2012 a 2017. Disponível em https://www.ipea.gov.br/portal/images/stories/PDFs/livros/livros/190416_rada_IDHM.pdf. Acesso em 14 Maio de 2021.