



VACINAÇÃO CONTRA A COVID-19 NO ESTADO DA PARAÍBA: ANÁLISE QUANTO AO PERFIL DA POPULAÇÃO VACINADA COM AS DOSES DE REFORÇO

Aluna: Ana Paula de Alexandria Nudes¹

Orientador: Messias Rafael Batista²

RESUMO

Com o surgimento da pandemia causada pela COVID-19, o país, assim como o estado da Paraíba, teve de aderir a medidas protetivas para conter a proliferação do vírus. No entanto, com o surgimento da vacina, a necessidade de ações como o *lockdown* diminuiu permitindo uma maior liberdade no dia a dia dos brasileiros e, por consequência, dos paraibanos. Porém, a manutenção de uma rotina sem demais restrições está relacionada à adesão ao programa de vacinação efetuado pelo Sistema Único de Saúde (SUS), pois o processo de imunização necessita ser mantido a partir de doses de reforço. Todavia, dados apontam que o número de adeptos a vacina pertencente à categoria de doses adicionais é menor em comparação a primeira e a segunda dose. Tal comportamento sugere a importância de um estudo sobre o tema, já que as doses de reforço contribuem para a saúde coletiva da população tendo em vista que sua aplicabilidade permite a diminuição da proliferação de variantes do SARS-COV-2. Isto posto, o presente trabalho pretende traçar um perfil da população paraibana que não tomou a dose de reforço com o intuito de auxiliar em trabalhos futuros sobre a temática, assim como, contribuir para a formulação de políticas públicas que incentivem a adesão à campanha de vacinação contra a COVID-19 disponibilizada pelo Governo do Estado da Paraíba.

Palavras-chaves: COVID-19; Vacinas; Dados; Paraíba

1 Graduada em Sistemas para internet, email de contato: anapaula.jp.br@gmail.com.

2 Mestrando em Engenharia da comunicação (UPE), Mestre em Relações internacionais (UEPB), Especializando em Engenharia e computação de dados (IESP), Especialista em Administração em Banco de Dados (ESTÁCIO/PB), Graduado em Sistemas para Internet (UNIPÊ). Docente no IESP/PB. Email de contato: mrafaelbatista@gmail.com

ABSTRACT

With the pandemic emergence caused by COVID-19, Brazil, as well as the state of Paraíba, had to adhere to protective measures to contain the virus spread. However, as the number of vaccine grows up, the need for actions such as the lockdown decreased, allowing more freedom for the Brazilians daily lives and, consequently, for the Paraíba state. Nevertheless, maintaining a routine without further restrictions is related to the adherence of the vaccination program carried out by the Sistema Único de Saúde (SUS), as the immunization process needs to be maintained from booster doses. The data indicates that the number of adherents to the vaccine belonging to the category of additional doses is smaller when compared to the first and second doses. Such behavior suggests the importance of a study on that subject, since the booster doses contribute to the collective health of the population, considering that their applicability allows the reduction of SARS-COV-2 variants proliferation. That said, the present work has the intend of draw a profile for the population from Paraíba state that did not take any additional dose in order to assist in a future work on that subject, as well as contribute to the formulation of public policies that encourage adherence to the vaccination campaign against COVID-19 made available by the Governo do Estado da Paraíba.

Keyword: COVID-19; Vaccines; Data; Paraíba

1 - INTRODUÇÃO

A pandemia, impulsionada pelo Coronavírus, transformou os hábitos em todo o mundo. Recomendadas pela Organização Mundial da Saúde (OMS), foram adotadas no Brasil, medidas de prevenção ao vírus, tais como o uso de máscaras, a higienização de produtos e artefatos, assim como, o uso de álcool em gel, etílico 70%, para a assepsia das mãos. No entanto, ao longo das descobertas quanto à letalidade e a capacidade de proliferação do SARS-COV-2, ações como o isolamento social alteraram, durante cerca de dois anos, segundo Rocha *et al.* (2021), drasticamente a vida de pessoas pelo mundo a partir da execução do *lockdown*. Com essa nova realidade posta, o surgimento da vacina tornou-se o elemento de maior importância global já que a partir da sua aplicação nas populações dos diversos países haveria uma redução nas medidas protetivas, em especial o isolamento social,

e um retorno aos hábitos já consolidados nos anos que antecederam a pandemia.

No decorrer dos anos, com o surgimento da vacina contra o COVID-19, os países foram se mobilizando para imunizar seus cidadãos. No entanto, faz-se importante frisar que o acesso a vacinação refletiu, em níveis globais, as desigualdades tanto externas, ou seja, entre os países, quanto internas, entre a população dos dados países. De acordo com Souza e Buss (2021), essa discrepância entre os países com maior poder de compra em comparação com os países mais pobres impediu tanto a equidade, no que tange ao acesso à vacina, quanto em linhas mais precisas, o maior controle da proliferação do vírus. Logo, esse desnível de acesso às doses presente entre os países, a depender da sua economia, ou seja, entre os mais ricos e os mais pobres, também se reflete internamente de modo que o cidadão com maior poder aquisitivo pode utilizar os seus recursos para ter acesso mais rápido à vacina. Sendo assim, esse comportamento se contrapõe a ideia de uma distribuição igualitária de um recurso tão necessário a população mundial.

No Brasil, o Programa Nacional de Imunização (PNI), criado em 1975, atua em conjunto com empresas privadas, no fornecimento gratuito de vacinas ao povo brasileiro por meio do SUS. Essa dinâmica de parceria apresenta alguns problemas já que deixa o país à mercê de ditames externos para desenvolver e fabricar suas próprias vacinas, como aponta Peres *et al.* (2021), pois ainda existe a necessidade de obtenção de insumos e transferência de tecnologias a partir de fontes não nacionais. Sendo assim, embora haja ao longo dos anos uma efetivação do programa de imunização faz-se necessária a busca pela independência nessa área tão importante para a saúde da população brasileira.

Diante disso, no contexto da pandemia, esse quadro não sofreu modificações, de acordo com Peres *et al.* (2021), de modo que instituições públicas brasileiras, como o Butantan, em parceria com instituições privadas desenvolveram um imunizante contra a COVID-19. Porém a partir de tecnologias e insumos fornecidos por outros países. Além da vacina citada anteriormente, a Coronavac, o Brasil teve acesso a outras como Pfizer, Janssen, Fiocruz e Astrazeneca. No entanto, faz-se importante frisar que a aplicação da primeira dose da vacina não é suficiente para garantir uma proteção contra o vírus ao longo do tempo. Tendo isso em vista é preciso realizar a tomada da dose de reforço após o final do ciclo vacinal de cada imunizante.

Isto posto, dados apontam uma redução no número de vacinados quando se trata da dose de reforço. Tendo em vista a importância desse procedimento para a manutenção da saúde da população brasileira como um todo, o presente trabalho tem por objetivo traçar o perfil dos paraibanos - baseado em dados como idade, sexo, cidade na qual a dose foi tomada,

etc - que tomaram as doses de reforço. Sendo assim, entendemos que os resultados que serão apresentados neste trabalho podem auxiliar na elaboração de políticas que incentivem um determinado grupo de pessoas - delimitado por idade, raça ou qualquer outro atributo -, apontado a partir dos resultados da análise dos dados, a tomarem as devidas doses e assim fortalecer o campo imunológico mais frágil da população paraibana até o presente momento.

Para tanto, o trabalho está dividido em cinco seções cuja apresentação é realizada a partir de uma sequência lógica de escrita. A primeira seção abordará a introdução. Na segunda, constará três subseções nas quais estão fundamentados, com base em outros estudos, os temas referentes a COVID-19 no Brasil e na Paraíba, a aplicação das vacinas no país e no território paraibano e, por fim, uma explanação sobre análise de dados. Para a terceira seção, reservou-se a metodologia onde é apresentado o método utilizado para o desenvolvimento deste trabalho. A quarta seção é destinada a apresentar os processos utilizados para o tratamento dos dados adquiridos, assim como, discutir os resultados obtidos diante da formulação dos gráficos desenvolvidos. Para esta finalidade, a seção quatro é dividida em três partes: uma subseção que trata da aquisição dos dados, uma que aborda o processo de tratamento desses dados e, por fim, uma subseção que expõe os gráficos formulados e os analisa. Por último, discorre-se na conclusão sobre os resultados obtidos e as suas importâncias para o objetivo deste trabalho.

2- FUNDAMENTAÇÃO TEÓRICA

Com o intuito de embasar os resultados obtidos a partir da aquisição, tratamento e exposição dos dados aos quais este trabalho se refere, entende-se que é importante tratar sobre temas voltados para: a COVID-19 no Brasil e, mais especificamente, no estado da Paraíba, a vacina contra o SARS-COV-2 e sua tomada e distribuição no território nacional e paraibano e, por fim, a análise de dados, sua definição e importância.

2.1 A COVID-19 NO BRASIL E NA PARAÍBA

A humanidade ao longo da história lida com eventos em que doenças são transmitidas em uma grande escala que podem vir a se tornar global. A esse cenário dá-se o nome de pandemia. Com o passar dos anos, a interação entre as populações dos diversos países do mundo se tornou cada vez mais intensa de modo que a probabilidade no aumento do contágio de uma doença aumentou na mesma proporção. Logo, a sociedade passou por

momentos históricos como o surgimento da Praga de Justiniano, da Peste Negra e da Varíola, por exemplo, apesar de serem eventos de grande escala na história humana, devido aos aspectos destrutivos que cada um causou a população que os vivenciou. Atualmente eventos como esses tendem a ser mais danosos à sociedade devido a facilidade e rapidez no contágio.

A pandemia causada pela COVID-19 é um exemplo do quão rápido uma doença, com possibilidade de progresso letal, pode se espalhar pelo mundo já que, de acordo com Rocha *et al* (2021), até o início do ano de 2021, a pandemia havia atingido mais de 188 países e afetado mais de 106 milhões de pessoas causando mais de dois milhões de mortes.

De acordo com a Organização Pan-Americana de Saúde (OPAS), há sete cepas do coronavírus que foram identificados em seres humanos, de modo que algumas dessas variantes são responsáveis por resfriados comuns e, algumas outras, causam síndromes gripais graves como, por exemplo, o SARS-COV. A COVID-19 é causada pelo SARS-CoV-2, ou seja, por uma variante do SARS-COV, logo, pode ser letal. As primeiras manifestações desse novo coronavírus surgiram na província de Hubei, mais especificamente na cidade de Wuhan, localizada na República Popular da China. A sua proliferação, como se sabe, foi rápida o que culminou, em cerca de três meses após a detecção do vírus, na declaração da Organização Mundial da Saúde cujo teor foi um alerta emergencial de surto pandêmico, como sugere Araújo (2021):

A Organização Mundial da Saúde declarou a disseminação da COVID-19 como pandemia em março de 2020, desde então, o seu combate tornou-se um desafio na comunidade científica em elucidar a progressão clínica da doença, terapias adequadas e o desenvolvimento de vacinas. (ARAÚJO *et al*, 2021, p. 2).

No Brasil, assim como na maior parte dos países do mundo, os efeitos da pandemia se refletiram aos poucos no dia a dia da população. Restrições como o uso de máscaras, a utilizações de álcool em gel, etílico 70%, para a higienização das mãos, de artefatos e de superfícies modificaram a vida de milhões de brasileiros. Os impactos não foram sentidos apenas no cotidiano das pessoas, mas surgiram mudanças na economia e em setores de serviços como, por exemplo, instituições educacionais que, devido às novas restrições, tiveram de reestruturar sua forma de funcionamento. De acordo com Malta *et al.*(2020), a adesão a esses novos comportamentos pela população brasileira provocou, especificamente no público adulto, uma redução nas horas destinadas à atividade física, um aumento no consumo de alimentos ultra processados, assim como, um aumento no tempo dedicado a dispositivos de tela como TVs e celulares. Por consequência, o conjunto de hábitos listados anteriormente afetaram consideravelmente a saúde de modo a aumentarem a propulsão de doenças cardiovasculares, assim como, a obesidade, entre outros danos ao corpo.

Diante desse cenário, consoante ao que demonstra Costa, Tavares e Clemente (2020), entre 6 de março e primeiro de abril de 2020, o nordeste brasileiro foi a segunda região mais afetada no país, já que os índices de casos confirmados foram mais altos em comparação a outros pontos do território nacional. Nesse contexto, a Paraíba teve o seu primeiro caso de COVID identificado na cidade de João Pessoa, capital do estado, no dia 18 de março de 2020. Ao longo do primeiro semestre de 2020, o número de casos referentes à doença cresceu no estado de modo que até o dia 16 de maio, do ano referenciado, conforme também aponta Costa (2020), foram registrados 4.063 casos confirmados de infecção pelo SARS-COV-2.

Esses casos estavam distribuídos em 135 municípios, dentre os 223 que compõem o estado, porém, foi na Região Metropolitana de João Pessoa que se concentrou cerca de 56% das ocorrências. Seguindo as consequências sofridas por todo o país, o território paraibano também teve de lidar com medidas protetivas para controlar o número de infectados, assim como, atuar com ações voltadas para tratar os contaminados. Logo, foi elaborado, pela Secretaria de Estado da Saúde, um plano de contingência para o combate à pandemia: o Plano de Contingência Estadual para Infecção Humana pelo Coronavírus (COVID-19).

2.2 AS VACINAS NO BRASIL E NA PARAÍBA

Ao longo dos anos, o desenvolvimento tecnológico proporcionou ao homem um aumento na expectativa de vida. Dentre os instrumentos que permitiram esse avanço, a vacina é um deles. Há muitas doenças que impuseram a humanidade grandes perdas populacionais, dentre as quais podemos citar a Poliomielite, o Sarampo e a Varíola. Durante muito tempo, a humanidade foi acometida pela varíola que matou uma quantidade significativa da população por mais de 2000 anos, cerca de 400 mil mortes por ano nos países europeus e mais de 3 milhões de mortes nas Américas, de acordo com o Dr. Guido Carlos Levi (VARELLA, 2021). No século XVIII, segundo Riedel (2005), o cientista britânico Edward Jenner, com o intuito de amenizar os danos causados pela varíola, foi o pioneiro no tratamento da doença por meio de um processo denominado como vacinação.

No Brasil, a erradicação da varíola se deu após dez anos da aplicação de campanhas pontuais para a supressão desta doença. Desse modo, as campanhas se iniciaram em 1962, em postos que se concentravam em sua maioria em centros urbanos e os resultados foram incisivos já que em 1972 a varíola já havia sido exterminada. Esse processo se deu antes da criação do Programa Nacional de Imunização (PNI) cuja importância na vida da população

brasileira, consoante o Dr. Gabriel Oselka (VARELLA, 2021), se dá devido ao fato de que por meio desse programa há uma maior abrangência na aplicação de vacinas, ademais, estima-se que o país vacine cerca de 300 milhões de pessoas por ano por meio do PNI. Sendo assim, o reflexo dessa cobertura vacinal está presente no aumento da expectativa de vida do brasileiro, assim como, no crescimento da qualidade de vida da população já que, com a erradicação, não há mais a necessidade de lidar com as sequelas causadas pela doença.

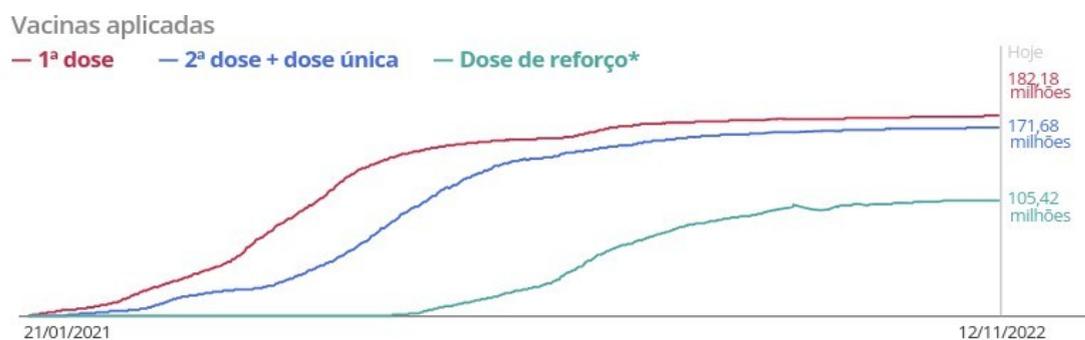
Apesar dos resultados efetivos proporcionados a partir da criação do PNI, em 1973, é preciso colocar em pauta alguns pontos concernentes ao modo como ele é estruturado. A instrumentalização do fornecimento de vacinas no país é composto por dois pilares: empresas públicas e privadas. Essa organização é formada com a participação complementar por parte das instituições privadas e com a participação de base da rede pública. Diante dessa estrutura, de acordo com Peres *et al.* (2021), a rede privada detinha, no período de 2004 a 2018, 80,6% dos registros de vacinas ativos, de modo que uma pequena porcentagem estava concentrada na rede pública, em especial entre três empresas: Fundação Ataulpho de Paiva, Fundação Ezequiel Dias (Funed), Fundação Oswaldo Cruz e Instituto Butantan. Além disso, ainda de acordo com Peres *et al.* (2021), apesar do histórico de ações efetivas por parte do PNI na saúde da população brasileira, a dependência do programa no que se refere aos insumos e as transferências de tecnologias estrangeiras deixa a mercê de pautas exteriores, e consequentemente diplomacias e interesses secundários, um elemento crucial para a proteção e saúde do país.

Como em todo território nacional, o nordeste brasileiro também foi acometido com um crescente número de infectados e de óbitos, especialmente em momentos de pico ao longo dos meses que se seguiram desde o início da pandemia. Dentre as medidas para a contenção da COVID-19, a imunização da população presente no nordeste foi uma grande aliada no combate a diminuição no número de infectados e no percentual de óbitos. Em conformidade com Silva (2022), o processo de vacinação dessa região teve início em 18 de janeiro de 2021, no entanto, o estado da Paraíba iniciou suas atividades de imunização no dia seguinte, ou seja, no dia 19 do mesmo mês e ano. Ainda de acordo com Silva (2022), seguindo o padrão aplicado em diversas regiões do país, a campanha de vacinação na Paraíba foi organizada em fases de modo que a primeira fase priorizou grupos considerados mais vulneráveis como: a população indígena que vivia em aldeias, trabalhadores voltados para a área da saúde, pessoas que residiam em asilos com idade de 60 anos ou mais e pessoas com deficiência. A aplicabilidade dessa campanha permitiu uma diminuição em três aspectos centrais para o combate ao COVID-19 no estado: diminuição do número de internos nas

UTIs, atenuação nas taxas de óbitos e redução na transmissão do vírus.

Outrossim, faz-se importante salientar que, mesmo com a aplicação da vacina, há, com o passar dos meses, uma redução na proteção contra o vírus de modo que haja a necessidade da tomada da dose de reforço. De acordo com Figueira e Fonseca (2021), de um modo geral, a diminuição da cobertura contra o vírus no corpo ocorre a partir do quinto mês após a aplicação da dose anterior. Ademais, uma população com uma baixa taxa de imunização permite um crescimento no número da proliferação de variantes do SARS-Cov-2. Todavia, dados fornecidos pelo G1 e pelo Governo da Paraíba, indicam, respectivamente, uma diminuição no número de vacinados com as doses de reforço tanto no que tange a população brasileira como no que se refere aos paraibanos. Desse modo, tal comportamento é um dado preocupante, tendo em vista que a dose de reforço é um importante instrumento contra os danos provocados pela COVID-19, conforme podem ser observados nas figuras 1 e 2.

Figura 1 - Indicativo do número de doses tomadas pela população brasileira a partir das categorias de doses.



(Fonte: website G1, 2022)

Figura 2- Recorte dos indicadores do número de aplicações de vacinas no estado da Paraíba inclusos em um dashboard (Painel de vacinação).



Fonte: Governo da Paraíba, 12/11/2022

2.3 ANÁLISE DE DADOS

Com o advento da internet e o surgimento de instrumentos como redes sociais e acesso a vídeos por *streaming*, por exemplo, a quantidade de dados em todo o mundo teve um crescimento exponencial. Com base nessas transformações, segundo Castro e Ferrari (2016), em um mundo em que o acesso às informações e aos serviços é fácil e, por consequência, de consumo cada vez mais rápido. O desafio se faz presente no desenvolvimento da capacidade de gerenciar, armazenar, processar e extrair conhecimentos a partir das informações obtidas. Diante disso, termos como “*Big data*” e “mineração de dados” foram cunhados a partir desse novo acesso aos dados em grande escala que, em diversos ramos de comercialização, são utilizados para impulsionar os índices de consumo e para prever comportamentos.

Ademais, para tratarmos o tema de modo assertivo, faz-se necessário o esclarecimento de alguns vocábulos que fazem parte do escopo do presente trabalho: dados, informação e conhecimento. De acordo com Belfiore e Fávero (2017), há uma hierarquia entre os termos citados, de modo que dado é uma unidade cujo tratamento resulta em informação. Por outro lado, o uso da informação, ou seja, a sua aplicabilidade, é o que define o termo conhecimento. Sendo assim, o fluxo da hierarquização apontada é a base para o entendimento do conceito de mineração de dados, de modo que, o processo de transformar dados em informação e, por conseguinte, informação em conhecimento, é o escopo para a finalidade da atividade de minerar dados.

Além disso, a aplicabilidade das informações adquiridas com a mineração de dados pode constituir uma análise de informações voltada apenas para a descrição ou ainda para a predição. A abordagem preditiva está vinculada a uma área de estudo denominada

Inteligência Artificial que usa o método de Aprendizagem de Máquina como ferramenta para o desenvolvimento de seus estudos. Com base nessa utilidade, e dentro do contexto pandêmico que ainda está sendo vivenciado, de acordo com Rocha *et al.*(2021), o uso da inteligência artificial é um instrumento que pode ser utilizado para a distribuição equitativa no que tange ao acesso a vacinas de modo a proporcionar ao PNI uma estratégia inovadora para lidar com os percalços advindos da tarefa de imunizar uma população em situações habitacionais e sociais tão diversas.

Isto posto, o uso de ferramentas que manipulam grandes quantidades de dados disponíveis atualmente pode auxiliar em pesquisas e na tomada de decisões para diversos campos do conhecimento humano. Em vista disso, utilizaremos neste trabalho a análise de dados para, por meio do tratamento dos dados, fornecer informações e conhecimentos que possam auxiliar na tomada de decisões sobre como impulsionar o crescimento do número de vacinados população paraibana com as doses de reforço.

3- METODOLOGIA

O presente trabalho se propõe a tratar do tema em questão por meio da análise de uma base secundária de dados de acesso público, de modo que não houve a necessidade de avaliação e aprovação do Comitê de Ética. Sendo assim, serão abordados dados nos quais estão presentes informações acerca da idade, do município, do sexo, da raça, da classe de grupos de vacinados nos quais o paciente se enquadra, do mês e ano de aplicação da dose, da categoria de doses e do número de doses tomadas pelas pessoas vacinadas.

Esses dados foram tratados e relacionados entre si com o auxílio de ferramentas, tais como o *Apache Spark*, o *Google Colaboratory* e o *Power BI*, respectivamente, de modo que o resultado desse processo tenha proporcionado uma análise concisa e efetiva entre o número de doses de reforço aplicadas e os atributos da população paraibana vacinada. A explanação desses dados será realizada a partir de um conjunto de ações que foram tomadas com base em três etapas: obtenção, tratamento e análise de dados referenciais. É importante salientar que a base secundária utilizada é, até o presente momento, atualizada diariamente, no entanto, tendo em vista uma maior precisão nos resultados, optamos por reduzir nosso escopo não só a partir do ponto de vista territorial, ou seja, com o foco no estado da Paraíba, mas também com um recorte temporal. Desse modo, os dados tratados neste trabalho se limitam a apresentar informações que vão até o dia em que foram colhidos: dia 20 de outubro de 2022.

Consequente, destacamos que o objetivo do nosso trabalho é explanar as características referentes a população paraibana que tomou as doses de reforço destinadas a

imunização contra a COVID-19 e com isso esperamos contribuir com o desenvolvimento de trabalhos voltados para este tema, assim como, para a aplicação de políticas públicas que visem interferir nos números de adeptos ao programa de vacinação através de incentivos que proporcionem uma maior adesão ao programa, em especial por parte dos grupos e perfis que são apontados a partir dos resultados obtidos.

4- RESULTADO E DISCUSSÃO

4.1- AQUISIÇÃO DOS DADOS

A princípio, a base de dados a ser utilizada para a análise das informações sobre o perfil dos paraibanos que se vacinaram dentro do grupo de aplicação das doses de reforço foi disponibilizada pela organização não-governamental Base dos Dados que, de modo participativo, tem por objetivo democratizar o acesso a dados públicos que são adquiridos pela organização, tratados por seus colaboradores e cedidos à população. No entanto, tendo em vista a dificuldade de se fazer o recorte territorial, já que o modo como os dados são disponibilizados pela organização inviabilizou a delimitação do escopo por conta da grande quantidade de dados ofertada para o acesso, foi utilizada outra fonte para a obtenção de uma base substancial de registros.

Diante desse impasse, este trabalho usa o *dataset* disponibilizado pelo Ministério da Saúde por meio do DATASUS cujos dados já estavam, na data de aquisição, separados por estado. A extração desses dados foi realizada a partir do download de três arquivos de extensão CSV que foram retirados da *web* através do link <https://bityli.com/uDVgMOMxBY>. Os documentos foram armazenados no *Google Drive* e, com o auxílio do *Google Colaboratory*, foram transformados, cada um, em *dataframes* com o uso da ferramenta *Apache Spark*. Conforme pode ser observado na **figura 3**.

Figura 3: Dataframe criado a partir do primeiro arquivo CSV obtido.

```
[ ] df_1 = spark.read.option("delimiter", ";").csv("/content/drive/MyDrive/partel1_pb.csv", inferSchema=True, header=True)
```

Contagem de dados do DF_1:

```
[ ] df_1.count()
```

3076885

```
df_1.show()
```

document_id	paciente_id	paciente_idade	paciente_dataNascimento	paciente_enumSexoBiologico	paciente_racaCor_codigo	paciente_racaCor_valor	paciente_enc
ab81633f-8345-43e...	421b19d8ff456b4da...	18	2003-05-06	F	3	PARDA	
d247e1cf-bfb8-466...	206443dca6140b1f0...	48	1973-04-10	M	3	PARDA	
ab82a604-4406-411...	891c673b201c3a8ba...	18	2003-11-07	F	3	PARDA	
d24d7164-3d96-46d...	0a7f28c6a5c6944fa...	17	2004-08-31	M	99	SEM INFORMACAO	
ab858bd6-3d45-4b9...	34e913a250aae195b...	52	1969-11-10	F	3	PARDA	
d25aa8a0-efd4-452...	6173d9c962fd2bd56...	73	1948-03-03	M	1	BRANCA	
ab889ac9-8a18-4a4...	e953760c1abc582bf...	18	2002-11-28	F	99	SEM INFORMACAO	
d25bbabd-fe9-493...	74a6fa51784f19608...	25	1996-02-10	M	3	PARDA	
ab8e68f5-e925-484...	c139cf9e3bce637f2...	53	1968-10-07	M	1	BRANCA	
d25bbc44-6403-45d...	273ae7939292e2d6d...	45	1975-09-10	F	4	AMARELA	
ab919eb1-83d8-4a2...	7ceb7dde647f34f2f...	39	1982-02-25	M	3	PARDA	
d25c9f27-2c03-449...	f94e74700766720e7...	33	1989-02-11	F	99	SEM INFORMACAO	
ab9a78e2-0263-4e6...	845e7a95b9a3b1bdb...	68	1953-03-04	F	4	AMARELA	
d25cfb9d-c37c-437...	95b92b9eb9378b78a...	23	1998-10-31	F	4	AMARELA	
aba51b2c-ffe3-43a...	414c511949553366b...	81	1940-04-18	F	1	BRANCA	
d2634f1d-63cc-4a0...	f8a1c5b85064a77a4...	83	1937-04-06	M	99	SEM INFORMACAO	
abb4d614-913a-46d...	4922243b32b6414f3...	14	2007-06-15	M	3	PARDA	

Fonte: (Elaborado pela autora, 2022)

4.2- TRATAMENTO DOS DADOS

Tendo em vista que o recorte de dados utilizado contempla apenas uma única tabela, já que todos os registros destinados a análise estavam presentes em um único *dataset*, o tratamento dos dados foi realizado no *Google Colaboratory* a partir de ferramentas presentes no *Apache Spark*. A quantidade total de registros adquiridos foi de 9.239.879 milhões. A estrutura da tabela inicialmente obtida é composta de 32 colunas cujo tipo de dados apresentados são unicamente dois: *String* e *Integer*. Com base nisso, foram aplicados os seguintes procedimentos para que os dados pudessem ser enviados para o Power BI com o intuito de formular os gráficos que serão apresentados adiante (figuras 4 a 10):

- a) União entre os *dataframes* gerados a partir da leitura dos arquivos no formato CSV adquiridos.

Figura 4: União dos dataframes.

```
UNIÃO DOS DATAS:

[ ] uniao_1 = df_1.union(df_2)

[ ] uniao_1.count()

6154255

[ ] df_uniao_total = uniao_1.union(df_3)

[ ] df_uniao_total.count()

9239879

df_uniao_total.show(1000)
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| document_id | paciente_id | paciente_idade | paciente_dataNascimento | paciente_enumSexoBiologico | paciente_racaCor_codigo | paciente_racaCor_valor | paciente_endereco_colibgeMunicipio | paciente_endereco_coP |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| [abb81633f-8345-43e-... | 1421b19d8ff456b4da... | 18 | 2003-05-06 | F | 3 | PARDA | 256150 |
| [d247e1cf-bf88-466-... | 286443dca6140b1f0... | 48 | 1973-04-10 | M | 3 | PARDA | 256960 |
| [ab823604-4486-411-... | 891c073b201c3a8ba... | 18 | 2003-11-07 | F | 3 | PARDA | 256750 |
| [2447164-3096-46d-... | 0b7f28c6cc0944fa... | 17 | 2004-08-31 | M | 99 | SEM INFORMACAO | 251210 |
| [ab858bd6-3d45-4b9-... | 34e913a250aae195b... | 52 | 1969-11-10 | F | 3 | PARDA | 251250 |
| [d25aa8a0-e1dd-452-... | 617309c962fd2b056... | 73 | 1948-03-03 | M | 1 | BRANCA | 256790 |
| [ab899a3c-9a18-404-... | e09376c1cab382bf... | 18 | 2002-11-28 | F | 99 | SEM INFORMACAO | 251310 |
| [d25bbabd-ef9-493-... | 74a6fa51784f19608... | 25 | 1996-02-10 | M | 3 | PARDA | 251430 |
| [ab8e88f5-e925-484-... | 139c75e3c6e37f2... | 53 | 1968-10-07 | M | 1 | BRANCA | 251290 |
| [d25bcb44-e483-45d-... | 273ae793929c206d... | 45 | 1975-09-19 | F | 4 | AMARELA | 256750 |
```

b) Fonte: (Elaborado pela autora, 2022)

c) Seleção das colunas cujos dados serão utilizados para formular um perfil de vacinados. Com esse intuito, e a partir desse recorte, o número de colunas resultou em treze.

Figura 5: Colunas selecionadas para o dataset utilizado.

```
[ ] df_select = df_uniao_total.select('paciente_id', 'paciente_idade', 'paciente_enumSexoBiologico', 'paciente_racaCor_valor', 'paciente_endereco_nmMunicipio',
'estabelecimento_razaosocial', 'estabelecimento_noFantasia', 'estabelecimento_municipio_nome',
'vacina_grupoAtendimento_nome', 'vacina_categoria_nome', 'vacina_fabricante_nome', 'vacina_dataAplicacao', 'vacina_descricao_dose')

[ ] df_select.show()

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| paciente_id | paciente_idade | paciente_enumSexoBiologico | paciente_racaCor_valor | paciente_endereco_nmMunicipio | estabelecimento_razaosocial | estabelecimento_noFantasia | estabelecimento_municipio_nome | vacina_grupo |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 421b19d8ff456b4da... | 18 | F | PARDA | BANANEIRAS | PREFEITURA MUNICI... | SECRETARIA MUNICI... | BANANEIRAS | Pessi |
| 286443dca6140b1f0... | 48 | M | PARDA | MONTE HOREBE | PREFEITURA MUNICI... | CENTRO DE SAUDE J... | MONTE HOREBE | Hipe |
| 891c073b201c3a8ba... | 18 | F | PARDA | JOAO PESSOA | SECRETARIA DE SAUDE | CARTAO SUS DE JOA... | JOAO PESSOA | Pessi |
| 0b7f28c6cc0944fa... | 17 | M | SEM INFORMACAO | POMBAL | POMBAL PREFEITURA | CENTRO DE SAUDE A... | POMBAL | Pessi |
| 34e913a250aae195b... | 52 | F | PARDA | QUEIMADAS | PREFEITURA MUNICI... | SECRETARIA MUNICI... | QUEIMADAS | Pessi |
| 617309c962fd2b056... | 73 | M | BRANCA | JURUPIRANGA | MUNICIPIO DE JURI... | UNIDADE MISTA DE ... | JURUPIRANGA | Pessi |
| e09376c1cab382bf... | 18 | F | SEM INFORMACAO | SAO SEBASTIAO DE ... | MUNICIPIO DE SAO ... | UBSF SANTA TEREZIMA | SAO SEBASTIAO DE ... | Pessi |
| 74a6fa51784f19608... | 25 | M | PARDA | SAO JOSE DE CAIANA | PREFEITURA MUNICI... | UNIDADE MISTA DE ... | SAO JOSE DE CAIANA | Pessi |
| 139c75e3c6e37f2... | 53 | M | BRANCA | RIO TINTO | MUNICIPIO DE RIO ... | UNIDADE DE SAUDE ... | RIO TINTO | Hipe |
| 273ae793929c206d... | 45 | F | AMARELA | JOAO PESSOA | SECRETARIA DE SAUDE | CARTAO SUS DE JOA... | JOAO PESSOA | Hipe |
| 74c07d66e47f34f2f... | 39 | M | PARDA | JOAO PESSOA | PREFEITURA MUNICI... | POLICLINICA MUNIC... | BAYEUX | Pessi |
| f94e74700766720e7... | 33 | F | SEM INFORMACAO | CAJAZEIRAS | PREFEITURA MUNICI... | UNIDADE MUNICIPAL... | CAJAZEIRAS | Func. |
| 845e7a95b9a3b1bbd... | 68 | F | AMARELA | JOAO PESSOA | SECRETARIA DE SAUDE | CARTAO SUS DE JOA... | JOAO PESSOA | Hipe |
| 9539299a959378078... | 23 | F | AMARELA | ARACAJI | PREFEITURA MUNICI... | UBS PSF I SANTO A... | ARACAJI | Pessi |
| 414c511949553366b... | 81 | F | BRANCA | CUITI | MUNICIPIO DE CAMP... | CENTRO DE SAUDE F... | CAMPINA GRANDE | Pessi |
| f8a1c3b8964a77a4... | 83 | M | SEM INFORMACAO | CAICARA | PREFEITURA MUNICI... | UBS REGINA MOREI... | CAICARA | Pessi |
| 4922243026414f3... | 14 | M | PARDA | SANTA HELENA | PREFEITURA MUNICI... | UNIDADE DE SAUDE ... | TRILINHO | Pessi |
| 7d830d68ed2b7fbd... | 75 | M | SEM INFORMACAO | VERTENTES | PREFEITURA MUNICI... | UNIDADE BASICA DE... | ALCANTIL | Pessi |
| ef08650f6f1ed1ef... | 12 | F | PARDA | TAPEROA | PREFEITURA MUNICI... | ESF VII JOSE MELQ... | TAPEROA | Pessi |
| 18cc03ebf9bfa0b5a... | 62 | F | PARDA | CAMPINA GRANDE | MUNICIPIO DE CAMP... | CENTRO DE SAUDE D... | CAMPINA GRANDE | Pessi |
```

Fonte: (Elaborado pela autora, 2022)

d) Modificação do nome das colunas para, a partir de uma nomenclatura mais clara, tornar mais fácil o entendimento dos dados na construção dos gráficos.

Figura 6: Código utilizado para modificar o nome das colunas.

```
[ ] df_select_rename = df_select.toDF(("Identificador", "Idade", "Sexo", "Raça", "Município", "Estabelecimento_razão_social", "Estabelecimento_nome_fantasia",
"Estabelecimento_município", "Vacina_grupo_de_atendimento", "Vacina_categoria_de_vacinados", "Vacina_nome_de_fabricante",
"Vacina_data_de_aplicação", "Vacina_doses"))

[ ] df_select_rename.show()

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Identificador|Idade|Sexo| Raça| Município|Estabelecimento_razão_social|Estabelecimento_nome_fantasia|Estabelecimento_município|Vacina_grupo_de_atendimento|Vacina_categoria_de_vacinados|V
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 421b19d8ff456b4da...| 18 | F | PARD| BANANEIRAS| PREFEITURA MUNICI...| SECRETARIA MUNICI...| BANANEIRAS| Pessoas de 18 a 6...| Faixa Etária|
| 206443c8c140b1f0...| 48 | M | PARD| MONTE MOREJE| PREFEITURA MUNICI...| CENTRO DE SAUDE J...| MONTE MOREJE| Hipertensão de d1...| Comorbidades|
| 891c673b2b1c3a8ba...| 18 | F | PARD| JOAO PESSOA| SECRETARIA DE SAUDE| CARTAO SUS DE JOA...| JOAO PESSOA| Pessoas de 18 a 6...| Faixa Etária|
| 0a07728c6a5c6944fa...| 17 | M|SEM INFORMACAO| POMBAL| POMBAL PREFEITURA| CENTRO DE SAUDE A...| POMBAL| Pessoas de 12 a 1...| Faixa Etária|
| 34e913a250a6c193b...| 52 | F | PARD| QUEIMADAS| PREFEITURA MUNICI...| SECRETARIA MUNICI...| QUEIMADAS| Pessoas de 18 a 6...| Faixa Etária|
| 6173d9c962f02bd56...| 73 | M | BRANCA| JURUPIRANGA| MUNICIPIO DE JUR...| UNIDADE MISTA DE ...| JURUPIRANGA| Pessoas de 70 a 7...| Faixa Etária|
| e953766c1abc582bf...| 18 | F|SEM INFORMACAO|SAO SEBASTIAO DE ...| MUNICIPIO DE SAO ...| UBSF SANTA TEREZINHA| SAO SEBASTIAO DE ...| Pessoas de 18 a 6...| Faixa Etária|
| 74a6fa51784f39008...| 25 | M | PARD| SAO JOSE DE CALANA| PREFEITURA MUNICI...| UNIDADE MISTA DE ...| SAO JOSE DE CALANA| Pessoas de 18 a 6...| Faixa Etária|
| c139cf9a3b8c4c37f2...| 53 | M | BRANCA| RIO TINTO| MUNICIPIO DE RIO ...| UNIDADE DE SAUDE ...| RIO TINTO| Hipertensão de d1...| Comorbidades|
| 273ae7939292e2d6d...| 45 | F | AMARELA| JOAO PESSOA| SECRETARIA DE SAUDE| CARTAO SUS DE JOA...| JOAO PESSOA| Pessoas de 60 nos...| Pessoas de 60 ano...|
| 7ceb7d0e647f34f2f...| 39 | M | PARD| JOAO PESSOA| PREFEITURA MUNICI...| POLICLINICA MUNIC...| BAYEUX| Pessoas de 18 a 6...| Faixa Etária|
| f94e74700f607267...| 33 | F|SEM INFORMACAO| CAJAZEIRAS| PREFEITURA MUNICI...| UNIDADE MUNICIPAL...| CAJAZEIRAS| Funcionário do S1...|
| 845e7a5b9a3b1bb...| 68 | F | AMARELA| JOAO PESSOA| SECRETARIA DE SAUDE| CARTAO SUS DE JOA...| JOAO PESSOA| Hipertensão de d1...| Comorbidades|
| 9509209e9378b78a...| 23 | M | AMARELA| ARACAGI| PREFEITURA MUNICI...| UBS PSF I SANTO A...| ARACAGI| Pessoas de 18 a 6...| Faixa Etária|
| 414c51194955366b...| 81 | F | BRANCA| CUITEI| MUNICIPIO DE CAMP...| CENTRO DE SAUDE F...| CAMPINA GRANDE| Pessoas de 80 ano...| Faixa Etária|
| f8a1c58564a7744...| 83 | M|SEM INFORMACAO| CAICARA| PREFEITURA MUNICI...| UBS REGINA MOREIR...| CAICARA| Pessoas de 80 ano...| Faixa Etária|
| 4922243b2b6414f3...| 14 | M | PARD| SANTA HELENA| PREFEITURA MUNICI...| UNIDADE DE SAUDE ...| TRIUNFO| Pessoas de 12 a 1...| Faixa Etária|
| 7f83068e8b2b1dd...| 75 | M|SEM INFORMACAO| VERTENTES| PREFEITURA MUNICI...| UNIDADE BASICA DE...| ALCANTIL| Pessoas de 70 a 7...| Faixa Etária|
| ef8065f9f3e1ef...| 12 | F | PARD| TAPEIRA| PREFEITURA MUNICI...| ESF VIL JOSE MELQ...| TAPEIRA| Pessoas de 12 a 1...| Faixa Etária|
| 18c6c5ebf9bfa6b5a...| 62 | F | PARD| CAMPINA GRANDE| MUNICIPIO DE CAMP...| CENTRO DE SAUDE D...| CAMPINA GRANDE| Pessoas de 18 a 6...| Faixa Etária|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
only showing top 20 rows
```

Fonte: (Elaborado pela autora, 2022)

e) Filtragem dos dados tendo como base apenas os registros referentes a população que foi vacinada com doses de reforço. Essa delimitação resultou em um número de 2.562.349 milhões de dados.

Figura 7: Filtragem dos tipos de doses para a construção do dataset utilizado.

```
[ ] df_select_dose = df_select_rename.filter((df_select_rename.Vacina_doses=='Reforço') | (df_select_rename.Vacina_doses=='2º Reforço')
| (df_select_rename.Vacina_doses=='3º Reforço') | (df_select_rename.Vacina_doses=='3ª Dose') | (df_select_rename.Vacina_doses=='4ª Dose')
| (df_select_rename.Vacina_doses=='Dose Adicional'))

[ ] df_select_dose.select("Vacina_doses").distinct().show()

+-----+
| Vacina_doses|
+-----+
| Reforço|
| 2º Reforço|
| 3ª Dose|
| 4ª Dose|
| Dose Adicional|
| 3º Reforço|
+-----+
```

Fonte: (Elaborado pela autora, 2022)

f) Criação de duas colunas a partir da coluna referente a data de aplicação da dose. As colunas criadas estão relacionadas ao mês e ao ano de realização da vacinação. Tal modificação resultou em um *dataframe* composto por 15 colunas.

Figura 8: Criação de duas colunas cujo conteúdo é referente ao ano e ao mês de aplicação da dose.

```
Criando colunas como ano_de_aplicação e mês_de_aplicação:

[] df_separacao_data = df_select_dose.withColumn('Vacina_ano_de_aplicação', split(df_select_dose['Vacina_data_de_aplicação'], '-').getItem(0))
.withColumn('Vacina_número_do_mês_de_aplicação', split(df_select_dose['Vacina_data_de_aplicação'], '-').getItem(1))

[] df_separacao_data.show()
```

Identificador	Idade	Sexo	Raça	Município	Estabelecimento_razão_social	Estabelecimento_nome_fantasia	Estabelecimento_município	Vacina_grupo_de_atendimento	Vacina_categoria_de_vacinados	Vacina
34e913a258bae1959...	52	F	PARDA	QUEIMADAS	PREFEITURA MUNICI...	SECRETARIA MUNICI...	QUEIMADAS	Pessoas de 18 a 6...	Faixa Etária	
c139c9f5c39e4e3772...	53	M	BRANCA	RIO TINTO	MUNICIPIO DE RIO...	UNIDADE DE SAUDE...	RIO TINTO	Hipertensão de di...	Comorbidades	
f94e74709766720e7...	33	F	M SEM INFORMACAO	CAJAZEIRAS	PREFEITURA MUNICI...	UNIDADE MUNICIPAL...	CAJAZEIRAS	Funcionário do Si...	Funcionário do Si...	
414c511949553366b...	81	F	BRANCA	CUITE	MUNICIPIO DE CAMP...	CENTRO DE SAUDE F...	CAMPINA GRANDE	Pessoas de 80 ano...	Faixa Etária	
492224383286414f5...	14	M	PARDA	SANTA HELENA	PREFEITURA MUNICI...	UNIDADE DE SAUDE...	TRUNFOI	Pessoas de 12 a 1...	Faixa Etária	
c8e7e01d6737e498a...	31	M SEM INFORMACAO	CAMPINA GRANDE	CAMPINA GRANDE	MUNICIPIO DE CAMP...	CENTRO DE SAUDE F...	CAMPINA GRANDE	Pessoas de 18 a 6...	Faixa Etária	
e2d3b4c7474dc9203...	65	F	BRANCA	LAGOA DE DENTRO	PREFEITURA MUNICI...	UNIDADE SAUDE DA...	LAGOA DE DENTRO	Pessoas de 18 a 6...	Faixa Etária	
8e653bca1340f0b2f...	50	M	AMARELA	MARI	PREFEITURA MUNICI...	POLO DE ATENDIMEN...	MARI	Hipertensão de di...	Comorbidades	
c7f892a71f134c2e6...	58	F	BRANCA	PIANCO	MUNICIPIO DE PIANCO	USF FERNANDO VIEI...	PIANCO	Pessoas de 18 a 6...	Faixa Etária	
a1fb261882a728757...	65	M	PARDA	TAVARES	MUNICIPIO DE TAVARES	ESF UBS V MARIA D...	TAVARES	Pessoas de 18 a 6...	Faixa Etária	
624a50917889322a2...	78	F	PARDA	CUITE	MUNICIPIO DE CUITE	UBSF SERRA DO BUM...	CUITE	Pessoas de 75 a 7...	Faixa Etária	
b9b1c34e24801a995...	25	F SEM INFORMACAO	MARITII	MUNICIPIO DE CAMP...	MUNICIPIO DE CAMP...	SECRETARIA DE SAU...	CAMPINA GRANDE	Pessoas de 18 a 6...	Faixa Etária	
153b6e763a9dd6a21...	31	F	AMARELA	JOAO PESSOA	SECRETARIA DE SAUDE	CARTAO SUS DE JOA...	JOAO PESSOA	Pessoas de 18 a 6...	Faixa Etária	
7fe20553e4457e92b...	78	M	PRETA	JURUPIRANGA	MUNICIPIO DE JURI...	UNIDADE MISTA DE...	JURUPIRANGA	Pessoas de 75 a 7...	Faixa Etária	
6a8ec43db0f7e190e...	64	F	BRANCA	CABEDELO	PREFEITURA MUNICP...	HOSPITAL E WATERW...	CABEDELO	Pessoas de 18 a 6...	Faixa Etária	
e494d699d9380bc02...	31	F	PARDA	JOAO PESSOA	SECRETARIA DE SAUDE	CARTAO SUS DE JOA...	JOAO PESSOA	Pessoas de 18 a 6...	Faixa Etária	
88694d138d7ed58e2...	26	F	PARDA	PRINCESA ISABEL	PREFEITURA MUNICI...	ESF PADRE IBIAPINA	PRINCESA ISABEL	Pessoas de 18 a 6...	Faixa Etária	
126f6899798402172...	26	F	BRANCA	JOAO PESSOA	SECRETARIA DE SAUDE	CARTAO SUS DE JOA...	JOAO PESSOA	Ensino Básico	Trabalhadores da ...	
c7e945f9c27386b5...	49	F SEM INFORMACAO	OLHO D'AGUA	OLHO D'AGUA	ESF I PARTEIRA MA...	ESF I PARTEIRA MA...	OLHO D'AGUA	Hipertensão de di...	Comorbidades	
204488a057b75b29...	33	M SEM INFORMACAO	CAMPINA GRANDE	CAMPINA GRANDE	MUNICIPIO DE CAMP...	SECRETARIA DE SAU...	CAMPINA GRANDE	Pessoas de 18 a 6...	Faixa Etária	

g) **Fonte:** (Elaborado pela autora, 2022)

h) Criação de mais uma coluna, com base na coluna relacionada ao mês de aplicação da vacina, em que ao invés de ter o número como identificador tem-se o nome do mês registrado. Com isso, o *dataset* utilizado para o desenvolvimento deste trabalho totaliza 16 colunas.

Figura 9: Criação da coluna que traz o nome do mês de aplicação da vacina.

```
Arcestando a coluna Vacina_nome_do_mês_de_aplicação com base na coluna Vacina_número_do_mês_de_aplicação:

[] df_tcc_final = df_separacao_data.withColumn('Vacina_nome_do_mês_de_aplicação', when(df_separacao_data.Vacina_número_do_mês_de_aplicação.contains("01"), "Janeiro")
.when(df_separacao_data.Vacina_número_do_mês_de_aplicação.contains("02"), "Fevereiro")
.when(df_separacao_data.Vacina_número_do_mês_de_aplicação.contains("03"), "Março")
.when(df_separacao_data.Vacina_número_do_mês_de_aplicação.contains("04"), "Abril")
.when(df_separacao_data.Vacina_número_do_mês_de_aplicação.contains("05"), "Maio")
.when(df_separacao_data.Vacina_número_do_mês_de_aplicação.contains("06"), "Junho")
.when(df_separacao_data.Vacina_número_do_mês_de_aplicação.contains("07"), "Julho")
.when(df_separacao_data.Vacina_número_do_mês_de_aplicação.contains("08"), "Agosto")
.when(df_separacao_data.Vacina_número_do_mês_de_aplicação.contains("09"), "Setembro")
.when(df_separacao_data.Vacina_número_do_mês_de_aplicação.contains("10"), "Outubro")
.when(df_separacao_data.Vacina_número_do_mês_de_aplicação.contains("11"), "Novembro")
.when(df_separacao_data.Vacina_número_do_mês_de_aplicação.contains("12"), "Dezembro")
.otherwise(""))
```

Fonte: (Elaborado pela autora, 2022)

f) Transformação do *dataframe* em um único arquivo no formato CSV.

Figura 10: Armazenamento do dataframe em um único arquivo CSV.

```
Criando um csv com os dados acima para serem utilizados no Power BI:

[] df_tcc_final.repartition(1).write.options(header='True', delimiter=',', encoding='utf-8').csv('/content/drive/MyDrive/dados_tratados_tcc_2_revisto.csv')
```

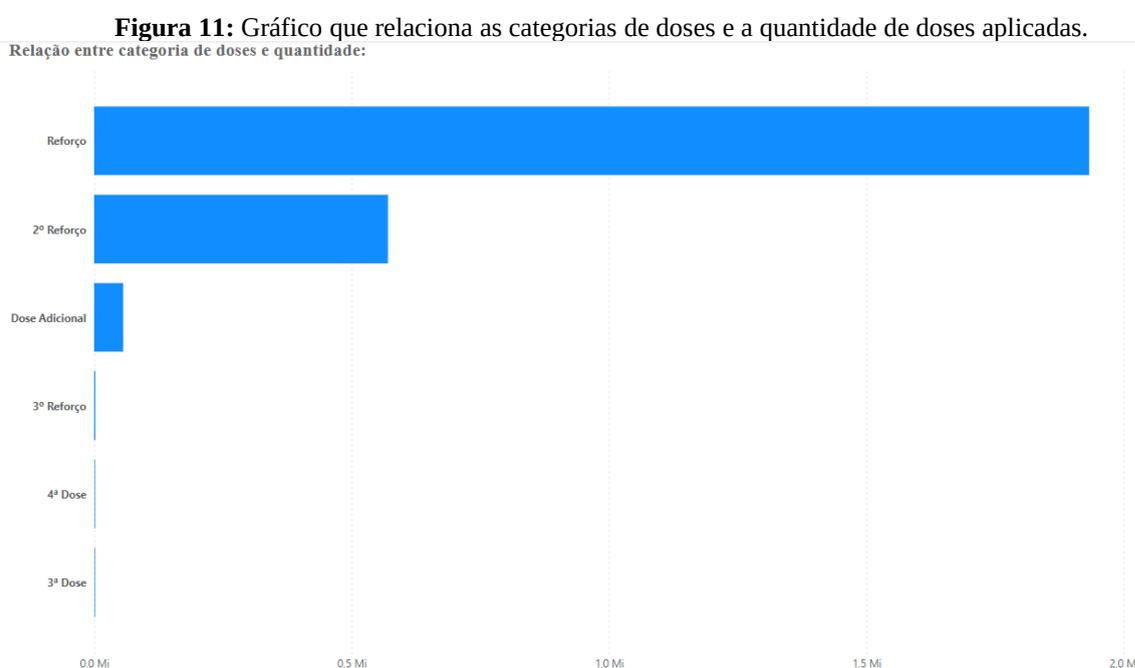
Fonte: (Elaborado pela autora, 2022)

Posto isto, faz-se necessário um esclarecimento quanto ao que foi apontado no item *d* desta subseção. As doses estipuladas, que constam na base de dados secundária e pública, cujos dados originalmente foram adquiridos para o desenvolvimento deste trabalho, foram categorizadas em 16 grupos: reforço, revacinação, 2ª dose, 3ª dose revacinação, 1ª dose, única, 2ª reforço, dose inicial, 4ª dose revacinação, 3ª dose, 4ª dose, dose, dose adicional, 3ª

reforço, 1ª dose revacinação e 2ª dose revacinação. No entanto, levando em consideração a definição da dose de reforço adotada neste trabalho, optou-se por se fazer um recorte nos registros referentes às categorias de doses aplicadas na Paraíba. Logo, entende-se que as doses de reforço com maior relevância para o estudo estão no escopo das seguintes terminologias: reforço, 2º reforço, 3º dose, 4º dose, dose adicional e 3º reforço. Partindo desse delineamento, utilizou-se registros direcionados a 6 grupos de doses.

4.3- APRECIACÃO DOS DADOS

Tendo em vista o que foi exposto nas seções anteriores, nesta serão abordados os gráficos que foram desenvolvidos com o uso do *Power BI*. Esses gráficos serão apresentados com o intuito de demonstrar o perfil da população paraibana que recebeu a dose de reforço disponibilizada na campanha de vacinação contra a COVID. Para tanto, adotou-se uma sequência lógica que partiu dos atributos mais amplos para os mais específicos e todos estão relacionados ao campo **dose** do *dataset* em questão.

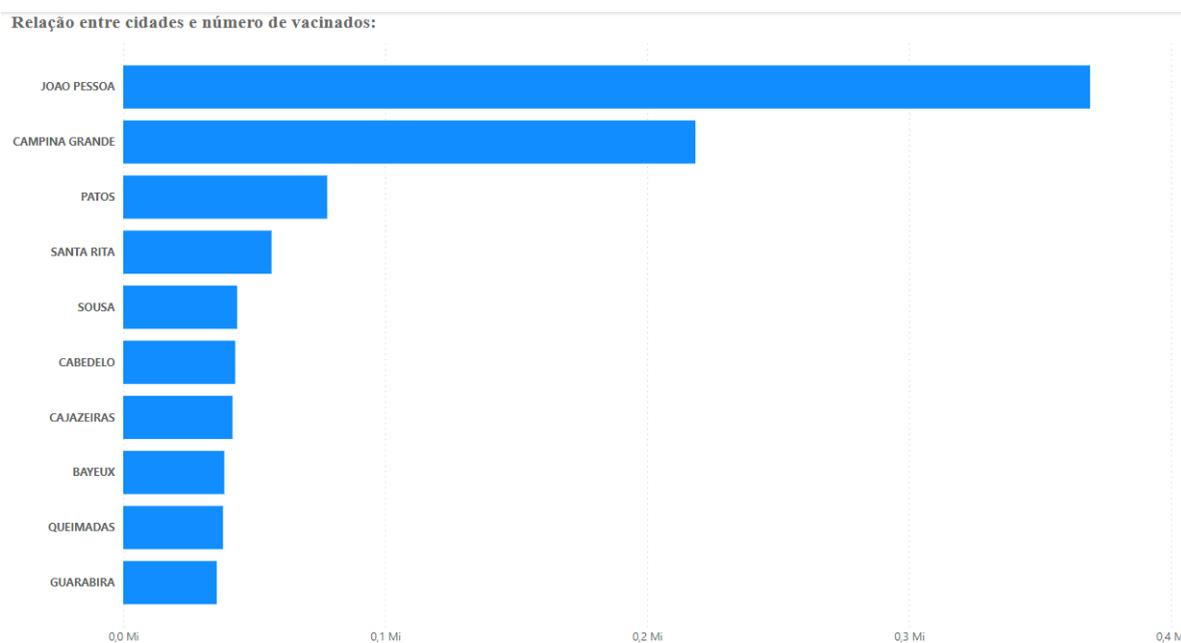


Fonte: (Elaborado pela autora, 2022)

Na relação entre quantidade de doses e categoria de doses aplicadas, exposta na **Figura 11**, o grupo de vacinas denominado como **reforço** teve a maior quantidade de amostras. Logo, essa categoria contabilizou 75,43% do percentual de doses. A partir do resultado indicado, há uma variação entre as categorias que se estabelece com uma amostra

de 1.932.853 de vacinados ante 4 vacinados com as doses agrupadas, respectivamente, como reforço e 3ª dose. É deduzível que, possivelmente, a discrepância entre as contagens demonstradas tenha relação com a forma de categorização já que a 4ª dose foi disponibilizada, por exemplo, no dia 29 de março, logo, um pouco mais de 6 meses anteriores à coleta de dados utilizados neste trabalho.

Figura 12: Gráfico que relaciona as 10 cidades que mais aplicaram doses.



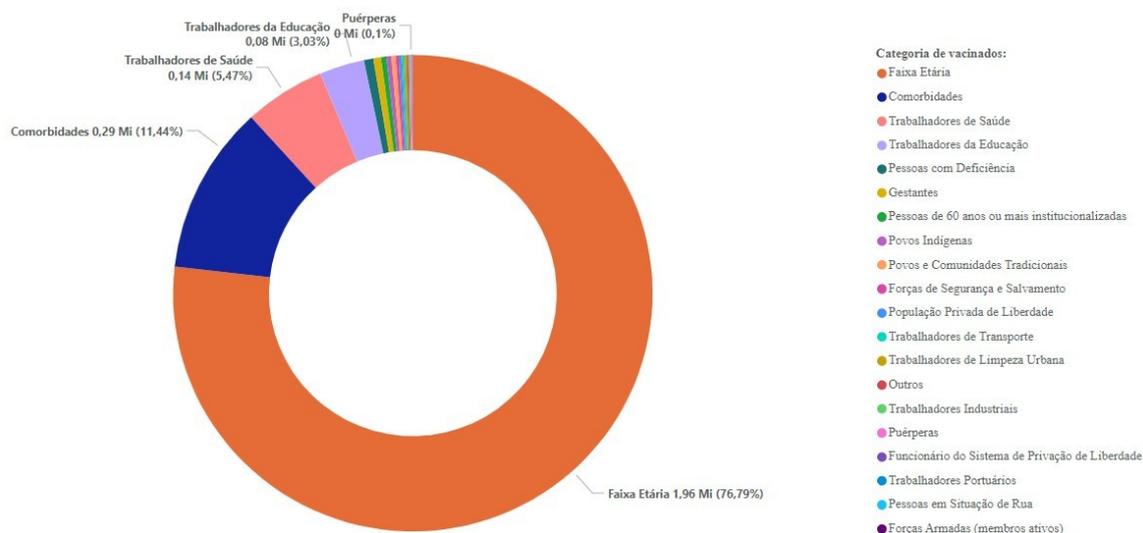
Fonte: (Elaborado pela autora, 2022)

Diante do resultado apontado na quantidade de vacinas de acordo com a categoria, é importante indicar também a relação entre o número de doses e as 10 cidades com maior percentual de pessoas imunizadas a partir da vacina de reforço. Os dados, como um todo, são

referentes a 223 municípios. Dentre estes, os 10 que apresentaram maior número de aplicações estão, possivelmente, relacionados com as maiores densidades demográficas do estado. A capital, João Pessoa, apresentou maior percentual com 38,35% em comparação as demais cidades. A cidade de Guarabira foi a que teve o menor número de pessoas imunizadas, contando com 35.710 vacinados.

Figura 13: Gráfico que relaciona a categoria de vacinados com a quantidade de doses aplicadas.

Relação entre categoria de vacinados e doses aplicadas:

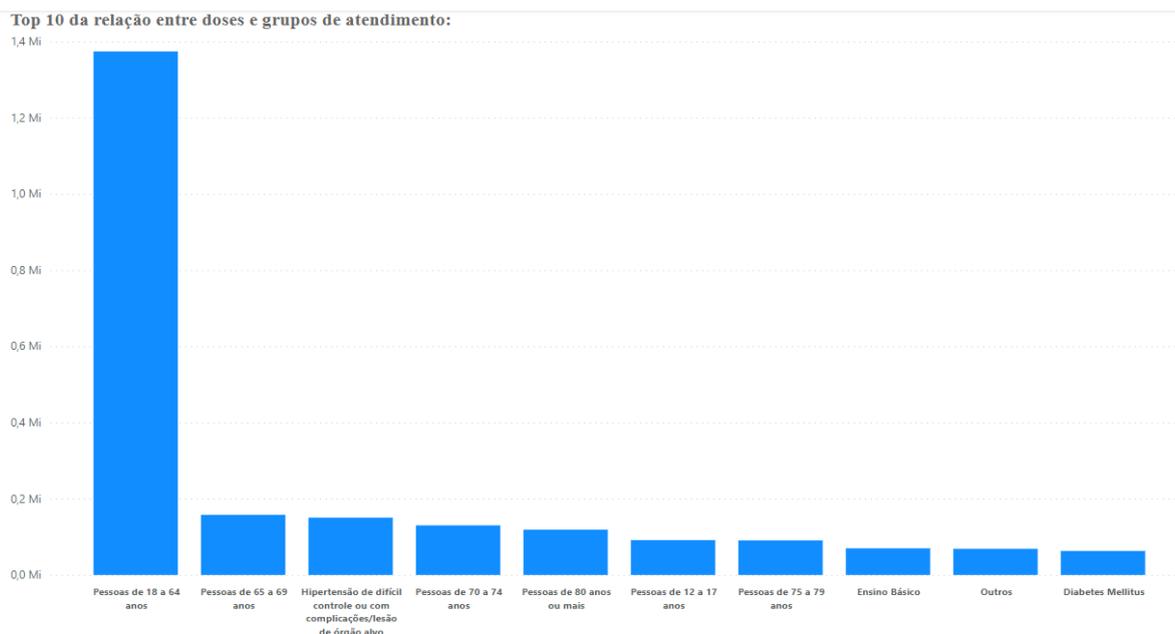


Fonte: (Elaborado pela autora, 2022)

Dentre as 20 categorias devidamente descritas, identificou-se que a **faixa etária** corresponde a 76,79% do número total de vacinados. Entende-se que esse maior volume na

categoria em questão se dá por conta da forma como o programa de vacinação foi aplicado no estado, já que os grupos mais específicos foram vacinados em uma primeira fase da campanha que, ao longo dos meses, adotou a idade da população como principal critério de divisão para a aplicação da vacina.

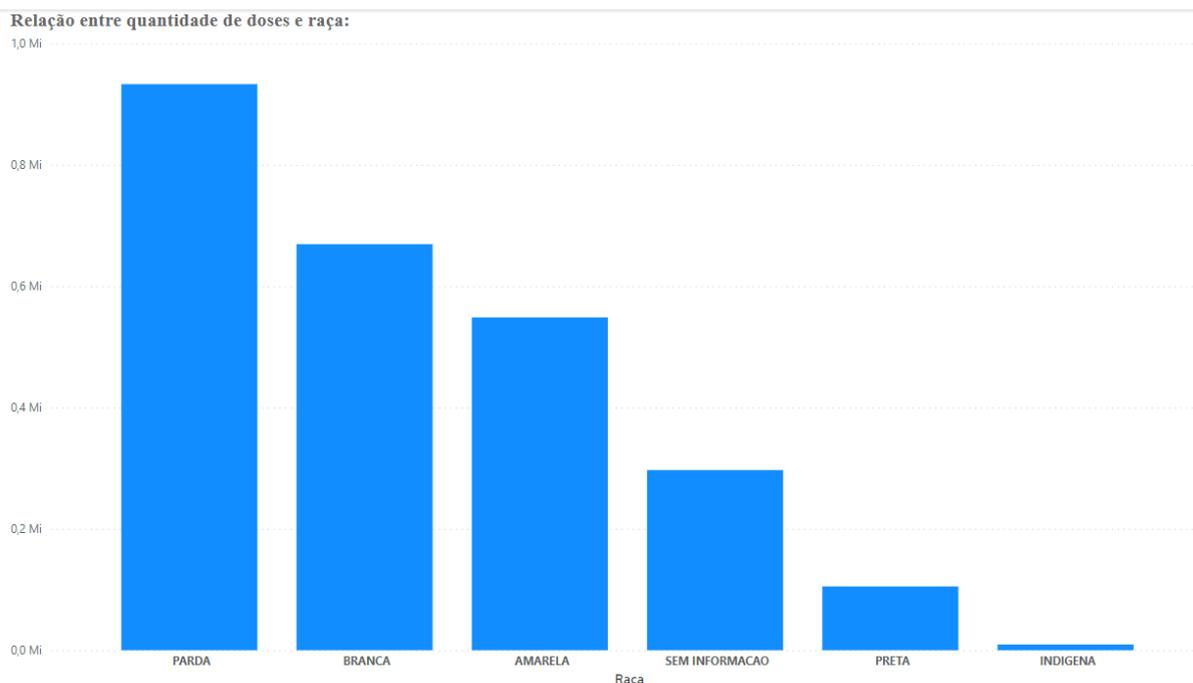
Figura 14: Gráfico que relaciona os 10 maiores grupos de atendimento com a quantidade de doses.



Fonte: (Elaborado pela autora, 2022)

Tendo em vista o que foi apresentado na Figura 13, entende-se que os grupos de atendimento, cujo os dez maiores estão elencados na Figura 14, representam um subconjunto das categorias de vacinados. Partindo disso, infere-se que sendo a **faixa etária** a categoria de vacinados com um maior percentual dentre as outras categorias, a quantidade de grupos de atendimento voltados para o intervalo de idade dos vacinados reforça a faixa etária como principal meio de identificação para caracterizar a população imunizada da Paraíba. Diante dos 93 grupos presentes no *dataset* utilizado, realizou-se um recorte abrangendo os 10 maiores. A partir dessa delimitação, identificou-se que o número de pessoas entre 18 e 64 anos apresentou o maior percentual, 59,37%, em comparação aos demais grupos. O agrupamento com menor quantidade de imunizados foi o destinado a portadores de *Diabetes Mellitus* computando 62.966 vacinados.

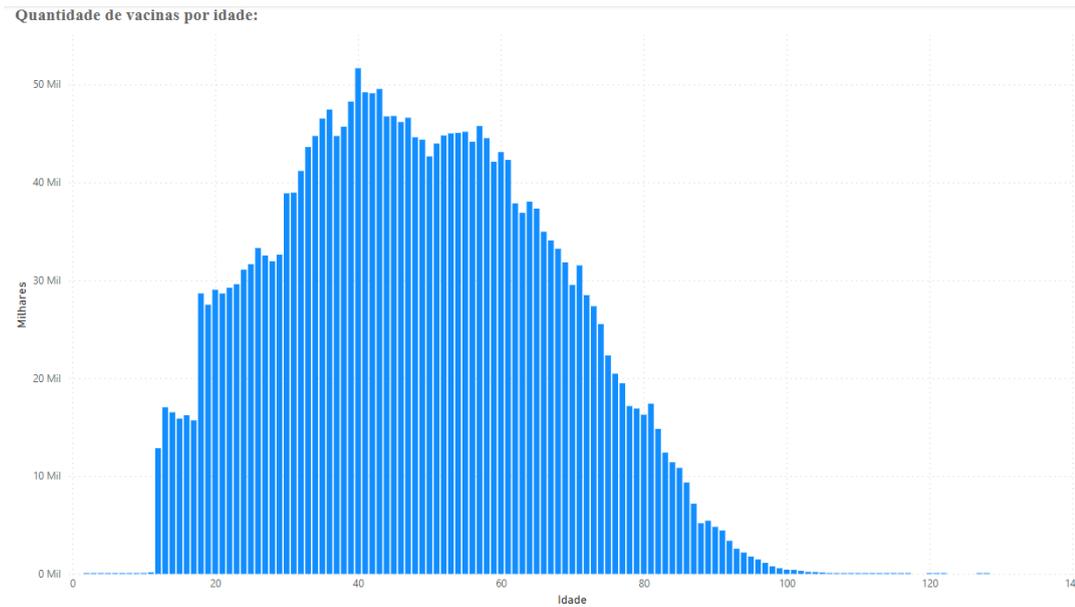
Figura 15: Relação entre quantidade de doses e raça.



Fonte: (Elaborado pela autora, 2022)

Partindo do princípio que a questão racial da população paraibana que recebeu a dose de reforço pode servir como um afinamento no processo de traçar um perfil dos imunizados, tendo em vista que critérios mais abrangentes, como grupos de vacinados, já foram abordados, entende-se que o acesso a esses dados pode contribuir para uma análise socioeconômica que relacione raça, desigualdade racial e acesso as doses. Isto posto, identifica-se que pessoas Pardas contabilizaram 36,41% da contagem. A variação entre os grupos de maior e menor quantidade foi de 933.042 para a população Parda e 9.419 para a população Indígena, respectivamente.

Figura 16: Gráfico que relaciona idade com o número de doses tomadas.

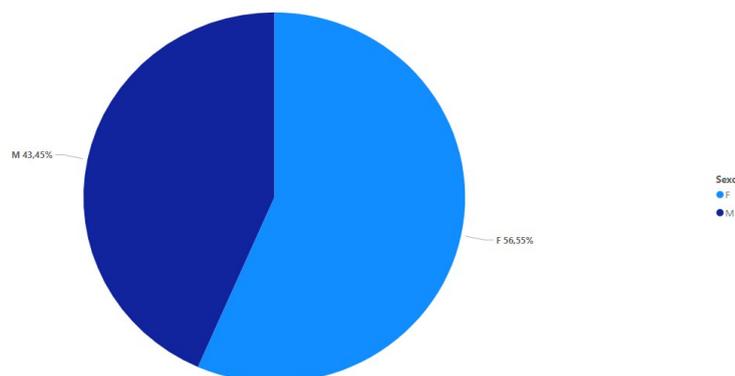


Fonte: (Elaborado pela autora, 2022)

Considerando que a idade é um dos principais atributos adotados para caracterizar os grupos de pessoas vacinadas na Paraíba, compreende-se que analisar mais especificamente esse tipo de registro disponibilizado no *dataset* contribui para uma melhor compreensão do perfil de vacinados no estado. Em consonância com o gráfico disponibilizado na Figura 14, verifica-se que a idade com maior incidência foi a de 40 anos, com um percentual de 2,02%. De um modo geral, essa faixa etária, por volta dos 40 anos, teve o maior número de pessoas imunizadas. É importante ressaltar que no *dataset* havia o registro de idades negativas, no entanto, para este estudo, foi realizado um filtro para manter apenas os registros com valores positivos.

Figura 17: Gráfico que relaciona o sexo e a quantidade de doses tomadas.

Percentual de doses por sexo:



Fonte: (Elaborado pela autora, 2022)

Por fim, na Figura 17 tem-se a demonstração do percentual do número de vacinados separados a partir do sexo. O *dataset* aponta mais duas categorizações além da nomenclatura feminino e masculino, no entanto, no gráfico pontuou-se apenas esses principais identificadores. A diferença entre o número de homens e mulheres vacinados com a dose de reforço é de 335.886 onde o percentual são, respectivamente, 43,45% e 56,55%.

5- CONSIDERAÇÕES FINAIS

Os hábitos adotados a partir das exigências de medidas para a contenção da COVID-19 aos poucos deixaram de fazer parte do dia a dia da população brasileira e, por conseguinte, da paraibana. Essa modificação tem como principal motivo o acesso gratuito e universal das vacinas disponibilizadas pelo Sistema Único de Saúde (SUS) a todos os brasileiros. No entanto, para a manutenção de uma saúde coletiva livre dos percalços que a contaminação pela COVID-19 pode causar, é de extrema importância a continuidade da aplicação dos imunizantes disponibilizados, ou seja, das doses de reforço.

5.1 LIMITAÇÕES DO TRABALHO

Os resultados alcançados tendo como base o desenvolvimento deste trabalho não foram suficientes para abarcar de modo mais detalhado um perfil dos grupos que não tomaram as doses de reforço. Tal impossibilidade tem como principal fator as idiosincrasias que perpassam por cada um dos grupos, tais como: condições socioeconômicas, posicionamento político e acesso aos meios pelos quais as vacinas foram disponibilizadas.

Ademais, cabe outros fatores que corroboram com a insuficiência de aspectos que possibilitem a formulação de uma *persona* que não se vacinou com as doses adicionais, no entanto, entende-se que a inconsistência dos dados, diante dos registros que os compõem, é o principal componente que contribui para essa falta não só neste trabalho mas, inclusive, em trabalhos futuros. Porém, há perguntas que podem ser respondidas a partir de outras bases e pesquisas.

5.2 TRABALHOS FUTUROS

Sendo assim, diante do que foi exposto, cabe um estudo mais aprofundado sobre o

perfil dos paraibanos que não se vacinaram com as doses de reforço. Tal aprofundamento deve ter por intuito discorrer mais precisamente sobre os aspectos que determinaram, por exemplo, o motivo pelo qual houve um menor número de vacinados entre pessoas com faixa etária inferior e superior ao intervalo de idades entre 40 anos. Faz-se importante também investigar as relações existentes entre a adesão às doses de reforço e o sexo dos vacinados tendo em vista que a população feminina obteve um maior percentual no que tange ao número de adeptos as doses. Outro ponto que merece destaque para uma abordagem mais minuciosa é a relação entre a quantidade de pessoas vacinadas e a população das dez cidades destacadas neste trabalho.

Em face do que foi dito, entende-se que as informações adquiridas a partir da análise elaborada neste estudo não são suficientes para traçar um perfil da população paraibana que foi vacinada com as doses em questão, pois, para tanto, acredita-se haver a necessidade de uma investigação mais apurada quanto aos motivos étnicos e sociais que determinam a tomada de decisão em aderir ou não a vacinação de reforço. No entanto, infere-se que este trabalho contribui para o conhecimento sobre os atributos que podem nortear futuros estudos sobre o tema e auxiliar na elaboração de medidas que visem incentivar determinados grupos a se vacinarem.

7- REFERÊNCIAS

- BELFIORE, Patrícia ; FÁVERO, Paulo Luiz. **Manual de análise de dados**. Rio de Janeiro: 1º edição, Elsevier, 2017.
- BUTANTAN, Instituto. **A parceria tecnológica que fez da CoronaVac a vacina do Brasil**. In: Instituto Butantan, 18 de jan. de 2021. Disponível em: <<https://butantan.gov.br/noticias/a-parceria-tecnologica-que-fez-da-coronavac-a-vacina-do-brasil>>. Acesso em: 6 de out. de 2022.
- CANAL DO PIRULA, **Vacinas valem a pena?**, 2014. 1 vídeo (32 min). Disponível em: <https://www.youtube.com/watch?v=hUvHKz3ugOg>. Acesso em: 07 de outubro de 2022.
- CASTRO, Leandro N.; FERRARI, Daniel Gomes. **Introdução à Mineração de Dados: conceitos básicos, algoritmos e aplicações**. São Paulo: Editora Saraiva, 2016.
- COSTA, Emilie O. ; TAVARES, Max O. ; CLEMENTE, Clarissa Maria. **A COVID-19 na Paraíba: análise epidemiológica e social**. In: Revista Diálogos em Saúde, 2020.
- EDWAR JENNER. In: WIKIPEDIA, a enciclopédia livre. Flórida: Wikimedia Foundation, 2022. Disponível em: <https://pt.wikipedia.org/w/index.php?title=Edward_Jenner&oldid=63311710>. Acesso em: 7 de out. de 2022.
- FIGUEIRA, Ana Cristina; FONSECA, Gabriel Cavalcanti da. **CoronaFatos: #75 - Dose de Reforço**. Apresentação: Ana Cristina Figueira e Gabriel Fonseca. Rio de Janeiro: Canal Saúde, 26 nov. 2021. Podcast. (11m08s). Disponível em: <https://www.arca.fiocruz.br/handle/icict/50085> . Acesso em: 10 de out. de 2022.
- GOVERNO DA PARAÍBA, **Plano de Contingência Estadual para Infecção Humana pelo Coronavírus (COVID-19)**. Paraíba. Secretaria de Estado da Saúde, 2020. Disponível em: <https://paraiba.pb.gov.br/diretas/saude/coronavirus/arquivos/plano-de-contingencia-estadual-para-infe-cao-humana-pelo-coronavirus-pb-atualizado-21-06-2021.pdf> . Acesso em: 10 out. de 2022.

OLÁ, CIÊNCIA!, **As 5 maiores pandemias da história**, 2021. 1 vídeo (10 min). Disponível em: <https://www.youtube.com/watch?v=MdSTOal9dxA>. Acesso em: 07 de out de 2022.

OPAS, **Histórico da pandemia de COVID-19**. In: OPAS - Organização Pan-Americana da Saúde. Disponível em: <<https://www.paho.org/pt/covid19/historico-da-pandemia-covid-19>>. Acesso em: 6 de out. de 2022.

OPAS, **Doença causada pelo novo coronavírus (COVID-19)**. In: OPAS - Organização Pan-Americana da Saúde. Disponível em: <<https://www.paho.org/pt/doenca-causada-pelo-novo-coronavirus-covid-19>>. Acesso em: 6 de out. de 2022.

PERES, K. C. *et al.* **Vacinas no Brasil: análise histórica do registro sanitário e a disponibilização no Sistema de Saúde**. Santa Catarina, 2021.

RIEDEL, Stefan. Edward Jenner and the history of smallpox and vaccination. **National Library of Medicine - National Center for Biotechnology Information**. Estados Unidos, jan. de 2005.

Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1200696/>.. Acesso em: 7 de outubro de 2022.

ROCHA, T. A. H. *et al.* **Plano nacional de vacinação contra a COVID-19: uso de inteligência artificial espacial para superação de desafios**. 2021.

SILVA, Paloma Cristina Melo. **Relação entre os dados epidemiológicos da COVID-19 e as políticas de imunização e contenção no nordeste do Brasil**. Orientador: Alex Bruno Ferreira Marques Nascimento. 2022. 40 f. TCC(Graduação) - Curso Superior de Tecnologia em Gestão Pública, Universidade Federal de Campina Grande, Campina Grande, 2022. Disponível em: <http://dspace.sti.ufcg.edu.br:8080/xmlui/handle/riufcg/27426>. Acesso em: 7 de out. de 2022.

SOUZA, Luís E. P. Fernandes ; BUSS, Paulo Marchiori. **Desafios globais para o acesso equitativo à vacinação contra a COVID-19**. In: Cadernos de Saúde Pública (CSP), 2021.

VARELLA, Drauzio. **Revolução das Vacinas: Como era o mundo sem vacinação | Episódio 1**, 2021, 1 vídeo (12 min). Disponível em: <<https://www.youtube.com/watch?v=BylqsLsYQ-k>> . Acesso em: 07 de out de 2022.