



**CENTRO UNIVERSITÁRIO INTERNACIONAL UNINTER
ESCOLA SUPERIOR POLITÉCNICA**

**CURSO SUPERIOR DE TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE
SISTEMAS**

REGINALDO GALLI DOS SANTOS

**ANÁLISE DE DADOS DA VACINAÇÃO CONTRA COVID-19 NO
MUNICÍPIO DE ATIBAIA-SP**

ATIBAIA— SÃO PAULO

2021

RESUMO

O presente material tem como objetivo apresentar um relatório detalhando todas as etapas envolvendo a pesquisa e análise de dados da vacinação contra Covid-19 no município de Atibaia-SP. Adotou-se como metodologia para a análise de dados os conceitos apresentados no livro **Pandas Python: Data Wrangling para Ciência de Dados**. Rio de Janeiro: Casa do Código, 2020, de Eduardo Correa. A pesquisa proporcionou a aquisição de diversos conhecimentos na área de programação e análise de dados, possibilitando também atingir o objetivo proposto como tema central deste trabalho. Todos os códigos desenvolvidos estão disponíveis para consulta e utilização no site: <https://github.com/reginaldogalli/analiseDadosVacinacao>.

Palavras-chave: vacinação, Covid-19, análise de dados, linguagem de programação, Python.

ABSTRACT

This material aims to present a report detailing all the steps involved in the research and data analysis of vaccination against Covid-19 in the city of Atibaia-SP. It was adopted as a methodology for the analysis of the concepts presented in the book **Pandas Python: Data Wrangling for Data Science**. Rio de Janeiro: Casa do Código, 2020 by Eduardo Correa. The research provided the acquisition of diverse knowledge in the area of programming and data analysis, also enabling the achievement of the proposed objective as the central theme of this work. All codes developed are available for consultation and use on the website: <https://github.com/reginaldogalli/analiseDadosVacinacao>.

Keywords: vaccination, Covid-19, data analysis, programming language, Python.

LISTA DE FIGURAS

FIGURA 1 - BAIXAR ARQUIVO CSV.....	9
FIGURA 2 - IMPORTAR BIBLIOTECA PANDAS.....	9
FIGURA 3 - PARSE ERROR.....	10
FIGURA 4 - LEITURA POR CHUNKS.....	10
FIGURA 5 - LEITURA DATA FRAME.....	11
FIGURA 6 - RENOMEANDO VACINAS.....	11
FIGURA 7 - CONVERTENDO TIPOS DE DADOS.....	12
FIGURA 8 - ABRINDO ARQUIVO SALVO.....	12
FIGURA 9 - IMPORTANDO A BIBLIOTECA MATPLOTLIB.....	13
FIGURA 10 - CÓDIGO PARA GERAR GRÁFICO INDICANDO A EVOLUÇÃO DA VACINAÇÃO.....	13
FIGURA 11 - GRÁFICO INDICANDO A EVOLUÇÃO DA VACINAÇÃO.....	14
FIGURA 12 - CÓDIGO PARA GERAR GRÁFICO INDICANDO A QUANTIDADE DE DOSES APLICADAS POR FABRICANTE.....	14
FIGURA 13 - GRÁFICO INDICANDO A QUANTIDADE DE DOSES APLICADAS POR FABRICANTE.....	15
FIGURA 14 - CÓDIGO PARA GERAR GRÁFICO INDICANDO A QUANTIDADE DE DOSES APLICADAS DE ACORDO COM O SEXO DA POPULAÇÃO.....	15
FIGURA 15 - GRÁFICO INDICANDO A QUANTIDADE DE DOSES APLICADAS DE ACORDO COM O SEXO DA POPULAÇÃO.....	16
FIGURA 16 - CÓDIGO PARA GERAR GRÁFICO INDICANDO A QUANTIDADE DE DOSES APLICADAS DE ACORDO COM A RAÇA DA POPULAÇÃO.....	16
FIGURA 17 - GRÁFICO INDICANDO A QUANTIDADE DE DOSES APLICADAS DE ACORDO COM A RAÇA DA POPULAÇÃO.....	17
FIGURA 18 - CÓDIGO PARA GERAR GRÁFICO INDICANDO A QUANTIDADE DE DOSES APLICADAS DE ACORDO COM O POSTO DE VACINAÇÃO.....	17
FIGURA 19 - GRÁFICO INDICANDO A QUANTIDADE DE DOSES APLICADAS DE ACORDO COM O POSTO DE VACINAÇÃO.....	18
FIGURA 20 - CÓDIGO PARA GERAR GRÁFICO INDICANDO A QUANTIDADE DE DOSES APLICADAS EM DOSE ÚNICA 1ª E 2ª DOSE.....	19
FIGURA 21 - GRÁFICO INDICANDO A QUANTIDADE DE DOSES APLICADAS EM DOSE ÚNICA 1ª E 2ª DOSE.....	19

LISTA DE TABELAS

TABELA 1 – CRONOGRAMA.....	6
----------------------------	---

SUMÁRIO

1.INTRODUÇÃO.....	5
2.CRONOGRAMA.....	6
3.PLANEJAMENTO INICIAL/METODOLOGIA.....	7
4.DESENVOLVIMENTO.....	8
5.RESULTADOS.....	13
6.CONCLUSÃO.....	20
7.REFERÊNCIAS.....	21

1.INTRODUÇÃO

Devido à pandemia de Covid-19, cientistas espalhados por todo o mundo trabalharam incansavelmente durante o ano de 2020 a fim de desenvolver vacinas eficazes contra a doença. Ao final de 2020 iniciou-se a vacinação contra Covid-19 em alguns países. No Brasil, a vacinação teve início nos primeiros dias de 2021. Tendo em vista os constantes avanços de informatização nas mais diversas áreas, não poderia ser diferente com relação à área da saúde.

Todos os dados referentes à vacinação no país estão sendo registrados, possibilitando o acompanhamento diário da vacinação no país. A disponibilização dos dados oferece aos governantes a possibilidade de análises detalhadas que podem servir de base para tomadas de decisão.

Os dados, porém, existem em grande volume e são difíceis de serem analisados, sendo necessário convertê-los em informações. Conforme Baltzan e Philips (2012), os dados são fatos brutos que descrevem um evento, enquanto a informação é o produto da conversão dos dados em um contexto significativo e útil.

O presente relatório tem como objetivo documentar todo o processo de pesquisa, coleta, filtragem e processamento dos dados, obtendo-se informações relevantes sobre a vacinação contra Covid-19 no município de Atibaia-SP.

2.CRONOGRAMA

Tabela 1 - Cronograma

CRONOGRAMA - ANÁLISE DE DADOS DA VACINAÇÃO CONTRA COVID-19 NO MUNICÍPIO DE ATIBAIA-SP																		
ATIVIDADE	JUNHO			JULHO				AGOSTO					SETEMBRO				OUTUBRO	
	14/6	21/6	28/6	5/7	12/7	19/7	26/7	2/8	9/8	16/8	23/8	30/8	6/9	13/9	20/9	27/9	4/10	14/10
Escolha do tema a ser desenvolvido																		
Pesquisa e coleta de dados para o trabalho																		
Desenvolvimento do trabalho																		
Elaboração do relatório																		
Elaboração dos demais tópicos do trabalho																		
Formatação segundo norma abnt																		
Revisão e finalização do trabalho																		
Entrega do trabalho																		

Fonte: elaborado pelo autor

3. PLANEJAMENTO INICIAL/METODOLOGIA

Após a definição do tema, iniciaram-se as pesquisas na literatura para fundamentação teórica no desenvolvimento do trabalho.

Conforme Corrêa (2020),

Em geral um processo de ciência de dados possui diferentes etapas de execução conforme as etapas listadas a seguir:

- Definição do problema: nesta etapa, o objetivo é realizar a modelagem do problema que se pretende resolver, utilizando a perspectiva da ciência de dados.
- Pré-processamento dos dados: é subdividida em duas fases: (i) seleção de dados, (ii) limpeza e transformação dos dados. De acordo com a literatura, a fase de pré-processamento de dados costuma ser a mais trabalhosa em qualquer projeto relacionado à ciência de dados, ocupando tipicamente 80% do tempo consumido. (<https://www.infoworld.com/article/3228245/the-80-20-data-science-dilemma.html>)
- Criação do modelo: trata-se da etapa "nobre" do processo, onde um algoritmo é aplicado sobre os dados que foram selecionados, limpos e transformados na etapa anterior.
- Interpretação dos resultados: nesta etapa os especialistas avaliam o modelo extraído na etapa anterior.

Este trabalho teve como foco as três etapas iniciais, pois considerou-se que no caso dos dados analisados, a interpretação dos resultados obtidos, deveria ser feita por profissionais da área da saúde ou da gestão do município.

4.DESENVOLVIMENTO

Para a análise dos dados da vacinação contra Covid-19, buscando utilizar os conhecimentos adquiridos durante os primeiros meses do curso de Análise e Desenvolvimento de Sistemas, optou-se por utilizar a linguagem de programação Python, pois conforme Banin (2018) trata-se de uma linguagem simples e intuitiva por um lado, poderosa e robusta por outro.

Ainda conforme Banin (2018)

Por ser simples e intuitiva, ela atende bem ao propósito de ser uma linguagem inicial utilizada por estudantes de programação que precisam de uma ferramenta para implementar seus primeiros algoritmos. Por ser poderosa e robusta, além de contar com uma grande e variada gama de bibliotecas aplicáveis a várias áreas da tecnologia da informação, ela pode ser adotada por profissionais de programação que necessitem de uma ferramenta que traga produtividade e qualidade aos projetos de software.

Foi possível encontrar dados detalhados sobre a vacinação através do site open DataSUS (<https://opendatasus.saude.gov.br/dataset/covid-19-vacinacao>), através do qual foi obtida a seguinte informação: O Ministério da Saúde, por meio do Sistema de Informação do Programa Nacional de Imunizações (SI-PNI), disponibiliza nesta página dados referentes à Campanha Nacional de Vacinação contra Covid-19 para análise de instituições interessadas e do público.

No site são disponibilizados arquivos do tipo CSV (Comma Separated Values), contendo os dados da vacinação por estado.

O primeiro grande problema encontrado foi o fato do arquivo com os dados da vacinação do estado de São Paulo, à época do estudo, conter aproximadamente 12 GB. Considerando-se um cenário real, onde seria feita uma análise diária ou semanal, isso acabaria consumindo muitos recursos de armazenamento e de internet local, além de exigir uma boa velocidade de conexão para realizar o download do arquivo.

Optou-se então por utilizar a plataforma de edição de códigos Colaboratory, ou “Colab” do Google. A vantagem desta ferramenta é a integração ao Google Drive, permitindo assim, realizar a leitura de arquivos armazenados na nuvem, eliminando a necessidade do download.

Foi utilizado o código abaixo para salvar uma cópia do arquivo diretamente no Google Drive:

Figura 1 - Baixar arquivo CSV

```
# baixa baixa um arquivo csv da internet direto para a pasta no google drive  
!wget -P '/content/drive/MyDrive/Colab Datasets/vacina_sp' https://s3-sa-east-1.amazonaws.com/ckan.saude.gov.br/
```

Fonte: elaborado pelo autor

Para realizar o processo de análise de dados optou-se por utilizar a biblioteca Pandas. Segundo o Correa (2020), a biblioteca pandas, *Python Data Analysis Library* (<https://pandas.pydata.org>), é um software livre, do tipo *open source*, que ao longo dos últimos anos se consolidou como a biblioteca para ciência de dados mais utilizada no ambiente Python. Ainda segundo o autor, a biblioteca pandas torna o processo de manipulação de dados tabulares mais rápido, simples e eficaz, acrescentando ao Python duas estruturas de dados extremamente poderosas e flexíveis: Series e Data Frame, sendo que a segunda serve para representar dados tabulares similares aos de uma planilha Excel ou de uma tabela de banco de dados relacional.

O código abaixo realiza a importação da biblioteca pandas para o “Colab”:

Figura 2 - Importar biblioteca pandas

```
# importa a biblioteca pandas como pd  
import pandas as pd
```

Fonte: elaborado pelo autor

Ao tentar realizar a leitura da maneira mais tradicional deparou-se com outro problema devido ao tamanho do arquivo. Ao tentar ler o arquivo e atribuir a um *data frame* para iniciar o processo de filtragem dos dados, ocorreu um erro conforme a imagem abaixo:

Figura 3 - Parse error

```

df = pd.read_csv('/content/drive/MyDrive/Colab_Datasets/vacina_sp/part-00000-f537b1aa-bf85-43a3-93e6-839ed89e4258.c000.csv')

ParserError                                Traceback (most recent call last)
<ipython-input-9-a5f87fe3b833> in <module>()
----> 1 df = pd.read_csv('/content/drive/MyDrive/Colab_Datasets/vacina_sp/part-00000-f537b1aa-bf85-43a3-93e6-839ed89e4258.c000.csv')

3 frames
/usr/local/lib/python3.7/dist-packages/pandas/io/parsers.py in read(self, nrows)
    2155     def read(self, nrows=None):
    2156         try:
-> 2157             data = self._reader.read(nrows)
    2158         except StopIteration:
    2159             if self._first_chunk:

pandas/_libs/parsers.pyx in pandas._libs.parsers.TextReader.read()

pandas/_libs/parsers.pyx in pandas._libs.parsers.TextReader._read_low_memory()

pandas/_libs/parsers.pyx in pandas._libs.parsers.TextReader._read_rows()

pandas/_libs/parsers.pyx in pandas._libs.parsers.TextReader._tokenize_rows()

pandas/_libs/parsers.pyx in pandas._libs.parsers.raise_parser_error()

ParserError: Error tokenizing data. C error: Expected 1 fields in line 3148869, saw 2

```

Fonte: elaborado pelo autor

Foi necessário então pesquisar uma nova solução para tornar possível a leitura de arquivos grandes. Após alguma pesquisa descobriu-se a possibilidade de realizar a leitura de arquivos grandes através de *chunks* (lotes), minimizando assim o impacto sobre o hardware local para realizar a leitura. Outra necessidade foi a de realizar uma filtragem, pois a análise de dados seria realizada somente referente ao município de Atibaia-SP e o arquivo contava com os dados da vacinação do estado inteiro. A solução implementada através do código abaixo resolvia os dois problemas, fazendo a leitura em lotes e fazendo a filtragem dos dados somente para as vacinas que foram aplicadas no município de Atibaia. Segue o código utilizado para realizar a leitura do arquivo:

Figura 4 - Leitura por chunks

```

# este código lê o arquivo csv de lotes de 10000 linhas e vai salvando na lista temporária filtro_df. Ao final salva na variável df
filtro_df = []
for chunk in pd.read_csv('/content/drive/MyDrive/Colab_Datasets/vacina_sp/part-00000-f537b1aa-bf85-43a3-93e6-839ed89e4258.c000.csv', sep=';', chunksize=10000):
    temp_df = chunk.loc[chunk['estabelecimento_municipio_nome'] == 'ATIBAIA']
    filtro_df.append(temp_df)
df = pd.concat(filtro_df, ignore_index=True)

```

Fonte: elaborado pelo autor

Após a utilização do código acima se observou uma redução no tamanho do arquivo de 12 GB para pouco mais de 100 MB e também já foi possível visualizar os dados contidos no arquivo CSV, conforme imagem abaixo:

Figura 5 - Leitura Data Frame

	document_id	paciente_id	paciente_idade	paciente_datanascimento	paciente_en
0	9ae11125-a272-40fd-b091-78d1403fa91f-i0b0	3b2e20e659e4928aac7e29c93c0d4c1da70e35bdafe024...	22	1998-07-30	
1	ccbe1a85-e5f3-4445-83e5-66a312507536-i0b0	8e9678d0ca1651548ae99aebeee3326704a97339d36341...	31	1989-07-03	
2	3a6dbc44-69b1-4497-8d4f-182a4c71a973-i0b0	d7710a9d5b926b8f38711a2790c92b0d67af383d837d8e...	26	1995-01-14	

Fonte: elaborado pelo autor

Fazendo a leitura dos dados, percebeu-se a necessidade de realizar uma melhor filtragem e organização dos dados para facilitar a futura análise dos dados.

Optou-se por simplificar a nomenclatura das vacinas para facilitar a análise e compreensão dos dados. Para isso foram utilizados os códigos abaixo:

Figura 6 - Renomeando vacinas

```
# muda o nome da vacina
novo_df = df.loc[df['vacina_nome'] == 'Vacina covid-19 - Ad26.COV2.S - Janssen-Cilag', 'vacina_nome'] = 'Janssen/Cilag'

novo_df = df.loc[df['vacina_nome'] == 'Vacina covid-19 - BNT162b2 - BioNTech/Fosun Pharma/Pfizer', 'vacina_nome'] = 'Pfizer/BioNTech'

novo_df = df.loc[df['vacina_nome'] == 'Vacina Covid-19 - Covishield', 'vacina_nome'] = 'Covishield/AstraZeneca'

novo_df = df.loc[df['vacina_nome'] == 'Covid-19-Coronavac-Sinovac/Butantan', 'vacina_nome'] = 'Coronavac/Butantan'

novo_df
```

Fonte: elaborado pelo autor

Em seguida foi necessário fazer a conversão do tipo *object* para o tipo *datetime* dos dados contidos na coluna referente à data de aplicação da vacina, para que fosse feita a ordenação dos dados por data de aplicação. Também foi criada uma nova coluna chamada *vacina_semanaaplicacao* para utilização no gráfico de evolução da vacina por semana do ano. Por último foi utilizado um comando para salvar o arquivo com as alterações realizadas, conforme os códigos abaixo:

Figura 7 - Convertendo tipos de dados

```
# converte os dados da data de aplicacao para o tipo data
df.vacina_dataaplicacao = pd.to_datetime(df.vacina_dataaplicacao)

# ordena as linhas por data de aplicacao
df.sort_values('vacina_dataaplicacao', inplace=True)

# cria a coluna semana do ano em que foi aplicada a vacina
df['vacina_semanaaplicacao'] = df.vacina_dataaplicacao.dt.isocalendar().week

# salvar os dados filtrados em um novo csv eliminando o indice original
df.to_csv ('/content/drive/MyDrive/Colab Datasets/vacina_sp/vacina_atibaia2.csv', index = False, sep=';')
```

Fonte: elaborado pelo autor

Para abrir o arquivo salvo foi utilizado o seguinte código:

Figura 8 - Abrindo arquivo salvo

```
# abre o arquivo salvo na biblioteca
df = pd.read_csv('/content/drive/MyDrive/Colab Datasets/vacina_sp/vacina_atibaia2.csv', sep=';')
```

Fonte: elaborado pelo autor

5.RESULTADOS

Após a etapa de pré-processamento dos dados, iniciou-se a geração dos gráficos contendo as informações.

Para enriquecer a análise dos dados através de gráficos, foi utilizada a biblioteca `matplotlib` com documentação disponível em: <https://matplotlib.org/stable/contents.html>

O código abaixo realiza a importação da biblioteca `pandas` para o “Colab”:

Figura 9 - Importando a biblioteca `matplotlib`

```
# importa a biblioteca matplotlib como plt
import matplotlib.pyplot as plt
```

Fonte: elaborado pelo autor

O código abaixo foi utilizado para gerar um gráfico que demonstra a evolução da quantidade de doses aplicadas por semana do ano. Para gerar este gráfico foi necessário buscar conhecimentos relacionados à parte de formatação do gráfico para que as informações fossem apresentadas de maneira clara.

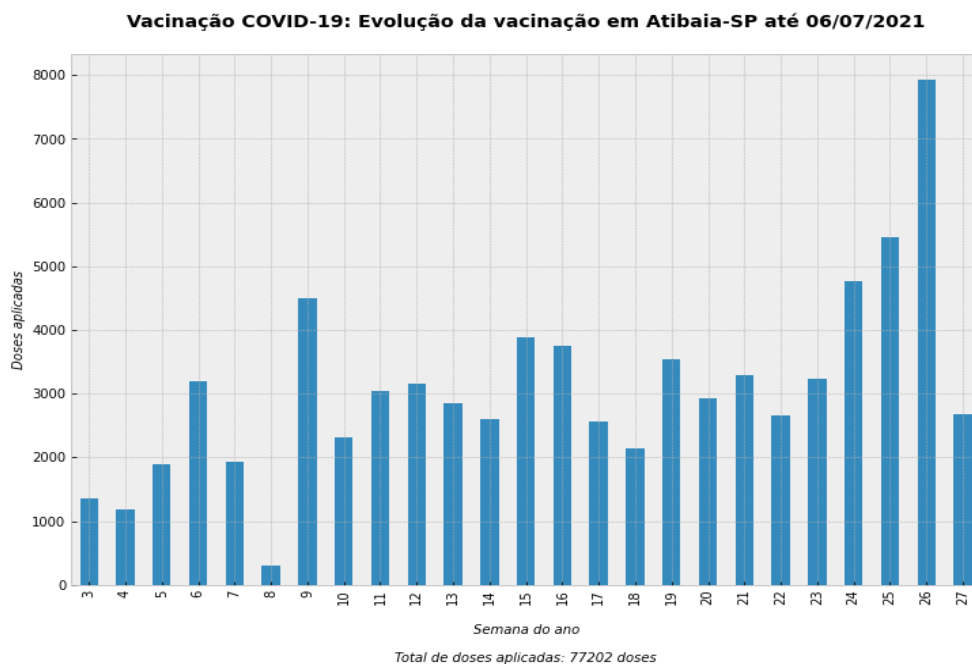
Figura 10 - Código para gerar gráfico indicando a evolução da vacinação

```
#Exibe o grafico de doses aplicadas de acordo com a semana do ano
df.groupby(['vacina_semanaaplicacao']).agg({'vacina_semanaaplicacao':'count'}).plot(kind='bar', figsize=(12,8), legend=False)
plt.style.use('bmh')
plt.title('Vacinação COVID-19: Evolução da vacinação em Atibaia-SP até 06/07/2021\n', fontsize=14, fontweight='bold', fontfamily='sans-serif')
plt.xlabel('\nSemana do ano\nTotal de doses aplicadas: 77202 doses', fontsize=10, fontstyle='italic', fontfamily='sans-serif')
plt.ylabel('Doses aplicadas', fontsize=10, fontstyle='italic', fontfamily='sans-serif')
plt.show()
```

Fonte: elaborado pelo autor

Para o gráfico em questão, foi escolhido o gráfico do tipo barras, pois considerou-se o mais indicado para esta aplicação. Resultado Obtido:

Figura 11 - Gráfico indicando a evolução da vacinação



Fonte: elaborado pelo autor com base nos dados do open DataSUS

O código abaixo foi utilizado para demonstrar a quantidade de doses aplicadas de acordo com o fabricante da vacina.

Figura 12 - Código para gerar gráfico indicando a quantidade de doses aplicadas por fabricante

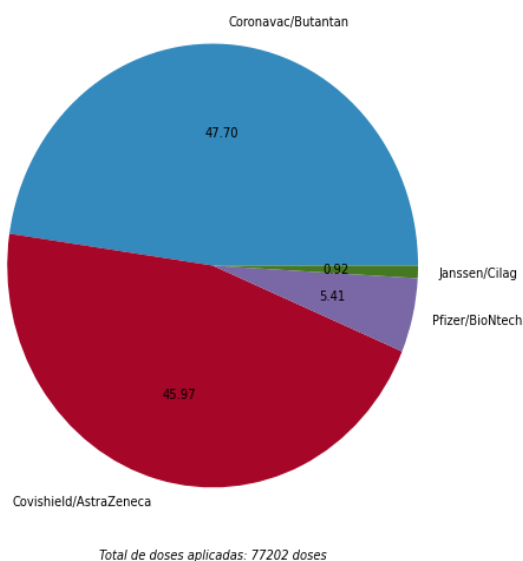
```
#Exibe a quantidade de vacinas aplicadas de acordo com o fabricante
df.vacina_nome.value_counts(normalize=True).plot(kind='pie', autopct='%2f', figsize=(10,7))
plt.style.use('bmh')
plt.title('Vacinação COVID-19: Porcentagem de doses por fabricante aplicadas em Atibaia-SP até 06/07/2021\n', fontsize=14, fontweight='bold', fontfamily='sans-serif')
plt.ylabel('')
plt.xlabel('Total de doses aplicadas: 77202 doses', fontsize=10, fontstyle='italic', fontfamily='sans-serif')
plt.tight_layout(w_pad=0.5, h_pad=0.5)
```

Fonte: elaborado pelo autor

Para o gráfico abaixo, foi escolhido o gráfico do tipo pizza (*pie*), pois considerou-se que este tipo facilitou a compreensão das informações. Resultado Obtido:

Figura 13 - Gráfico indicando a quantidade de doses aplicadas por fabricante

Vacinação COVID-19: Porcentagem de doses por fabricante aplicadas em Atibaia-SP até 06/07/2021



Fonte: elaborado pelo autor com base nos dados do open DataSUS

O código abaixo foi utilizado para demonstrar a quantidade de doses aplicadas em homens e mulheres.

Figura 14 - Código para gerar gráfico indicando a quantidade de doses aplicadas de acordo com o sexo da população

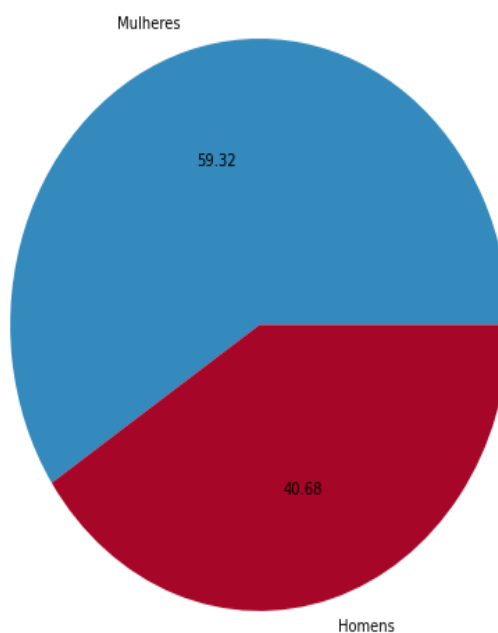
```
#Exibe a quantidade de vacinas aplicadas de acordo com o sexo
df.paciente_enumsexobiologico.value_counts(normalize=True).plot(kind='pie', autopct='%2f', labels=['Mulheres', 'Homens'], figsize=(10,7))
plt.style.use('bmh')
plt.title('Vacinação COVID-19: Proporção de homens e mulheres vacinados\n em relação ao total de doses aplicadas em Atibaia SP até 06/07/2021', fontsize=14, fontweight='bold', fontfamily='sans-serif')
plt.ylabel('')
plt.tight_layout(w_pad=0.5, h_pad=0.5)
```

Fonte: elaborado pelo autor

Para o gráfico abaixo, foi escolhido o gráfico do tipo pizza (*pie*), pois considerou-se de fácil interpretação neste caso. Resultado Obtido:

Figura 15 - Gráfico indicando a quantidade de doses aplicadas de acordo com o sexo da população

Vacinação COVID-19: Proporção de homens e mulheres vacinados em relação ao total de doses aplicadas em Atibaia SP até 06/07/2021



Fonte: elaborado pelo autor com base nos dados do open DataSUS

O seguinte código foi utilizado para demonstrar a quantidade de doses aplicadas de acordo com a raça declarada do vacinado.

Figura 16 - Código para gerar gráfico indicando a quantidade de doses aplicadas de acordo com a raça da população

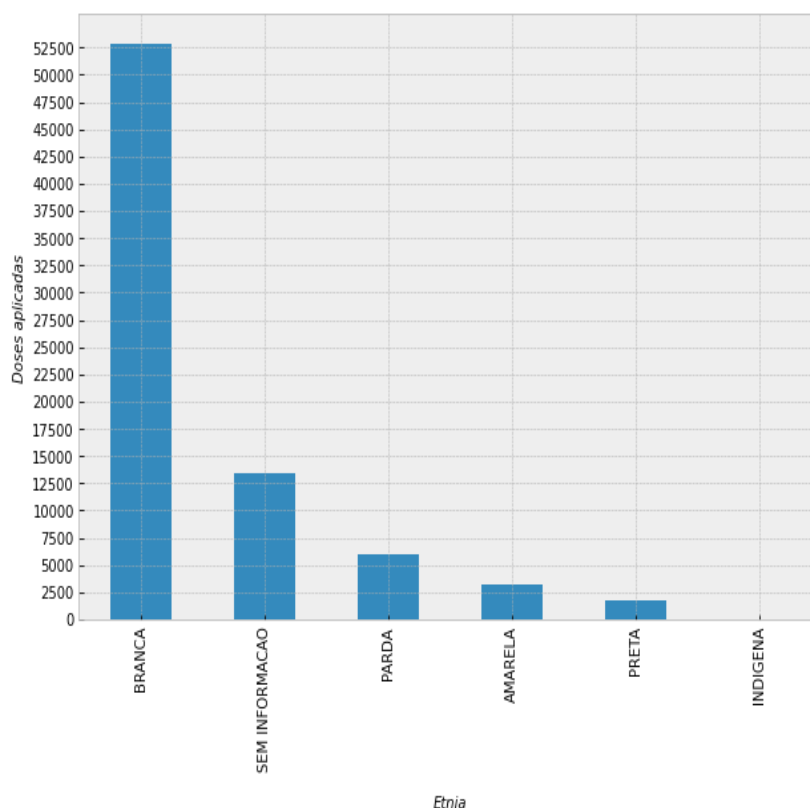
```
#Exibe a porcentagem de vacinados de acordo com o raça
df.paciente_racacor_valor.value_counts().plot(kind='bar', figsize=(10,7))
plt.style.use('bmh')
plt.title('Vacinação COVID-19: Quantidade de vacinados por etnia em Atibaia-SP até 06/07/2021\n', fontsize=14, fontweight='bold', fontfamily='sans-serif')
plt.xlabel('\nEtnia', fontsize=10, fontstyle='italic', fontfamily='sans-serif')
plt.ylabel('Doses aplicadas', fontsize=10, fontstyle='italic', fontfamily='sans-serif')
bins = range(0,55000,2500)
plt.yticks(bins)
```

Fonte: elaborado pelo autor

Para o gráfico em questão, foi escolhido o gráfico do tipo barras, pois considerou-se o mais indicado para esta aplicação. Resultado Obtido:

Figura 17 - Gráfico indicando a quantidade de doses aplicadas de acordo com a raça da população

Vacinação COVID-19: Quantidade de vacinados por etnia em Atibaia-SP até 06/07/2021



Fonte: elaborado pelo autor com base nos dados do open DataSUS

O próximo código foi utilizado para demonstrar a quantidade de doses aplicadas de acordo com o posto de vacinação.

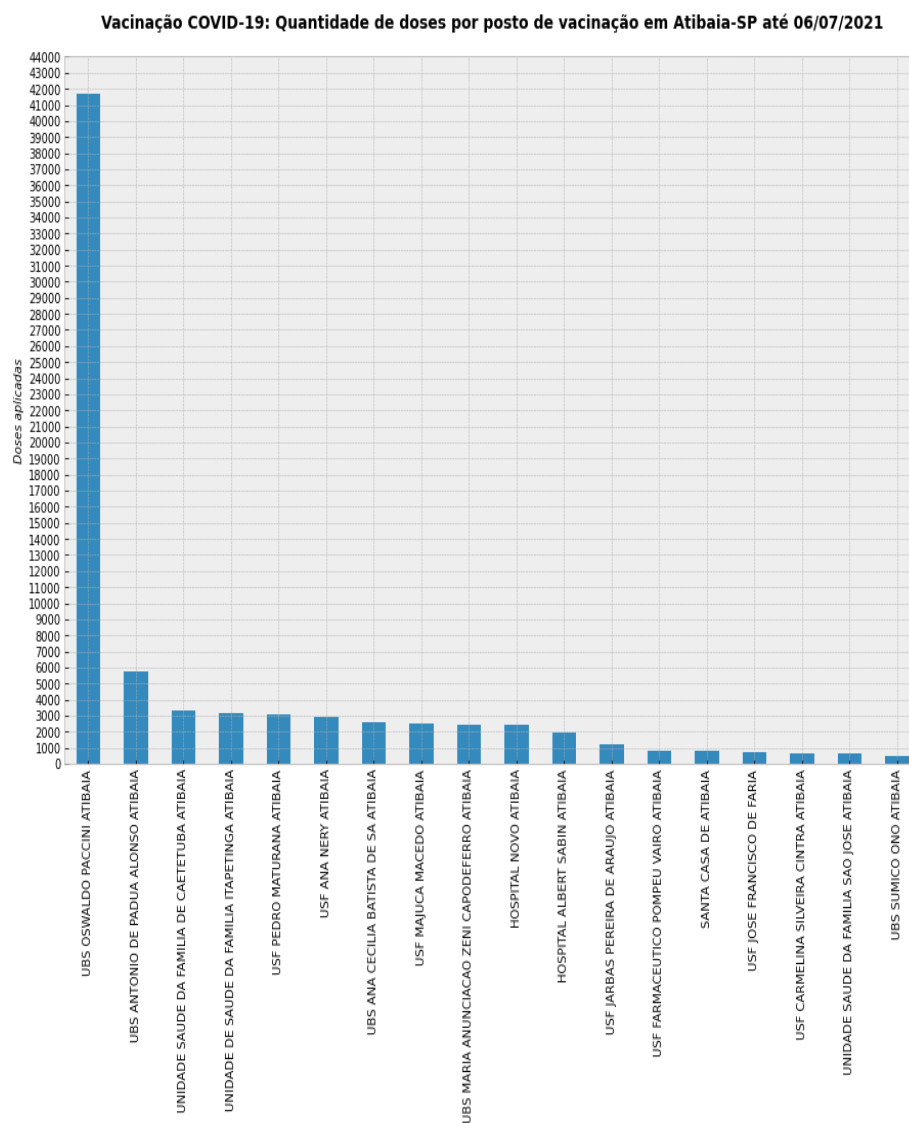
Figura 18 - Código para gerar gráfico indicando a quantidade de doses aplicadas de acordo com o posto de vacinação

```
#Exibe a quantidade de vacinas aplicadas por localidade
df.estalecimento_nofantasia.value_counts().plot(kind='bar', figsize=(15,10))
plt.style.use('bmh')
plt.title('Vacinação COVID-19: Quantidade de doses por posto de vacinação em Atibaia-SP até 06/07/2021\n', fontsize=14, fontweight='bold', fontfamily='sans-serif')
plt.ylabel('Doses aplicadas', fontsize=10, fontstyle='italic', fontfamily='sans-serif')
bins = range(0,45000,1000)
plt.yticks(bins)
```

Fonte: elaborado pelo autor

Para o gráfico em questão, foi escolhido o gráfico do tipo barras, pois considerou-se o mais indicado para esta aplicação. Resultado Obtido:

Figura 19 - Gráfico indicando a quantidade de doses aplicadas de acordo com o posto de vacinação



Fonte: elaborado pelo autor com base nos dados do open DataSUS

O código abaixo foi utilizado para demonstrar a quantidade de doses aplicadas em homens e mulheres.

Figura 20 - Código para gerar gráfico indicando a quantidade de doses aplicadas em dose única 1ª e 2ª dose

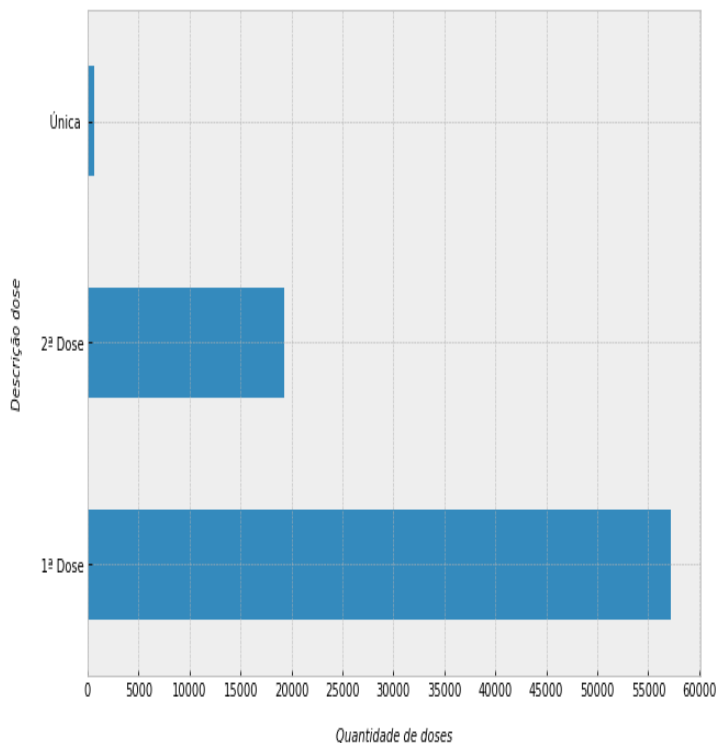
```
#Exibe a quantidade de vacinas aplicadas na primeira, segunda e dose unica
df.vacina_descricao_dose.value_counts().plot(kind='barh', figsize=(10,7))
plt.style.use('bmh')
plt.title('Vacinação COVID-19: Quantidade de doses aplicadas em dose única, 1ª e 2ª dose em Atibaia-SP ate 06/07/2021', fontsize=14, fontweight='bold', fontfamily='sans-serif')
plt.xlabel('\nQuantidade de doses', fontsize=10, fontstyle='italic', fontfamily='sans-serif')
plt.ylabel('Descrição dose', fontsize=10, fontstyle='italic', fontfamily='sans-serif')
bins = range(0,65000,5000)
plt.xticks(bins)
```

Fonte: elaborado pelo autor

Para o gráfico em questão, foi testado o gráfico do tipo barras horizontais. Resultado Obtido:

Figura 21 - Gráfico indicando a quantidade de doses aplicadas em dose única 1ª e 2ª dose

Vacinação COVID-19: Quantidade de doses aplicadas em dose única, 1ª e 2ª dose em Atibaia-SP ate 06/07/2021



Fonte: elaborado pelo autor com base nos dados do open DataSUS

6.CONCLUSÃO

O estudo realizado proporcionou a oportunidade de colocar em prática vários conhecimentos adquiridos sobre a linguagem de programação Python. Também foram realizadas pesquisas na área de análise de dados, possibilitando a aquisição de conhecimentos nessa área. Ao final do estudo foram obtidos diversos gráficos detalhando a vacinação contra Covid-19 no município de Atibaia-SP.

Visando a contribuição com a sociedade, os códigos desenvolvidos serão disponibilizados através do link: <https://github.com/reginaldogalli/analiseDadosVacinacao>. Através dos códigos disponibilizados, é possível replicar a mesma solução para qualquer município realizando pequenas alterações no código. As informações obtidas através dos gráficos gerados podem servir como informação e até mesmo como base para tomada de decisão por parte de profissionais da saúde dos municípios.

O estudo teve como propósito realizar uma análise simplificada com base nos dados obtidos, sendo considerado que o objetivo do estudo foi atingido.

Considera-se, porém, que as possibilidades na área de análise de dados são inúmeras, não descartando a possibilidade de aprofundamento e aprimoramento nos resultados obtidos realizando, por exemplo, cruzamento de dados de duas ou mais planilhas diferentes.

Para concluir, pode-se dizer que através do estudo realizado foi possível demonstrar uma das inúmeras aplicações da linguagem de programação Python, proporcionando grande enriquecimento de conhecimento e crescimento pessoal.

7.REFERÊNCIAS

CORRÊA, Eduardo. **Pandas Python: Data Wrangling para Ciência de Dados**. Rio de Janeiro: Casa do Código, 2020.

BALTZAN, P; PHILLIPS, A. **Sistemas de informação**. Porto Alegre: AMGH, 2012.

RUIZ, Armand. **The 80/20 data science dilemma**. Disponível em: (<https://www.infoworld.com/article/3228245/the-80-20-data-science-dilemma.html>). Acesso em: 16 jun. 2021.

BANIN, Sérgio Luiz. **Python 3: conceitos e aplicações: uma abordagem didática** / Sérgio Luiz Banin. São Paulo: Érica, 2018

OPEN DATASUS, **Campanha Nacional de Vacinação contra Covid-19**. Disponível em: (<https://opendatasus.saude.gov.br/dataset/covid-19-vacinacao>). Acesso em: 16 jun. 2021.

COLABORATORY. **Perguntas frequentes**. Disponível em: (<https://research.google.com/colaboratory/intl/pt-BR/faq.html>). Acesso em: 16 jun. 2021.

PANDAS, Disponível em: (<https://pandas.pydata.org/>). Acesso em: 16 jun. 2021.

MATPLOTLIB, **Overview**. Disponível em: (<https://matplotlib.org/stable/contents.html>). Acesso em: 16 jun. 2021.

SILVA, Tiago. **Python de A a Z: guia de primeiros passos com a linguagem**. São Paulo: Sguerra Desing, 2019.